Blind Deconvolution Using a Variational Approach to Parameter, Image, and Blur Estimation

Rafael Molina, Javier Mateos, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—Following the hierarchical Bayesian framework for blind deconvolution problems, in this paper, we propose the use of simultaneous autoregressions as prior distributions for both the image and blur, and gamma distributions for the unknown parameters (hyperparameters) of the priors and the image formation noise. We show how the gamma distributions on the unknown hyperparameters can be used to prevent the proposed blind deconvolution method from converging to undesirable image and blur estimates and also how these distributions can be inferred in realistic situations. We apply variational methods to approximate the posterior probability of the unknown image, blur, and hyperparameters and propose two different approximations of the posterior distribution. One of these approximations coincides with a classical blind deconvolution method. The proposed algorithms are tested experimentally and compared with existing blind deconvolution methods.**

***Index Terms*—Bayesian framework, blind deconvolution, parameter estimation, variational methods.**

I. Introduction

BLIND deconvolution refers to a class of problems of the form

$$g(x) = h(x) * f(x) + n(x), \quad x = (x_1, x_2) \in I \tag{1}$$

where $I \subset \mathcal{R}^2$ is the support of the image, and $f(x)$, $g(x)$, $h(x)$, and $n(x)$ represent, respectively, the unknown original image, the observed image, the unknown impulse response or point spread function (PSF) of the blurring system, and the observation noise. The operator $(*)$ in (1) denotes 2-D convolution, given by

$$(h * f)(x) = \sum_{s \in \mathcal{S}} h(s) f(x - s) \tag{2}$$

where $\mathcal{S} \subset \mathcal{R}^2$ is the support of the PSF.

Equation (1) can be written in matrix-vector form as

$$g = Hf + n \tag{3}$$

by lexicographically ordering g, f, and n. Matrix H is a block-Toeplitz matrix which is approximated by a block-circulant matrix.

In classical image restoration, the blurring function is assumed to be known, and the degradation process is inverted using one of the many existing restoration algorithms. Various restoration approaches have appeared in the literature which depend on the particular degradation and image models used (see, for example, [1]–[3] for details).

The objective of blind deconvolution methods is to obtain estimates of h and f based on g and prior knowledge about the unknown quantities and the noise. There are two main approaches to the blind deconvolution problem [4], [5]. With the first one, the blur PSF is identified separately from the original image and later used in combination with one of the known image restoration algorithms, while with the second one the blur identification step is incorporated into the restoration procedure.

Two types of approaches, an experimental and a theoretical one, have been reported in the literature for identifying the PSF separately from the original image. With the experimental approach, the images of one or more point sources are collected and used to obtain the PSF [6]. With the theoretical approach, the PSF is mathematically modelled, usually assuming a particular type of degradation, like out-of-focus blur [7] or Gaussian blur [8], [9], or by considering a particular imaging application like microscopy [10], medical ultrasound [11], remote sensing [12], or astronomy (see, for instance, Tiny Tim [13], a program to simulate the PSF of the Hubble Space Telescope).

When the PSF estimation is performed jointly with the restoration process, most methods address the blind (or semi-blind) deconvolution problem by incorporating prior knowledge about the image and blur into the deconvolution process. This knowledge can be expressed, for instance, in the form of convex sets and regularization techniques (see, for example, [14]–[17]) or with the use of the Bayesian paradigm with prior models on the problem unknowns [18], [19]. In this paper, we will use the Bayesian paradigm to jointly estimate the image, blur, and unknown hyperparameters in the blind deconvolution problem.

Our goal will be, first, to define a joint distribution $\mathrm{p}(\Omega, f, h, g)$ of the observation, g, the unknown image, f, the blur, h, and the hyperparameters, Ω, describing their distributions. Then, we will calculate the posterior distribution of the unknowns given the observed image $\mathrm{p}(\Omega, f, h|g)$ and use this posterior distribution to estimate the image and blur. Bayesian modeling and inference is based on building $\mathrm{p}(\Omega, f, h, g)$ to later perform inference based on $\mathrm{p}(\Omega, f, h|g)$.

To model the joint distribution, we utilize in this paper the hierarchical Bayesian paradigm (see, for example, [20]). This

Manuscript received April 26, 2005; revised March 23, 2006. This work was supported in part by the Comisión Nacional de Ciencia y Tecnología under Contract TIC2003-00880, in part by the Greece-Spain Integrated Action HG2004-0014, and in part by the "Instituto de Salud Carlos III" project FIS G03/185. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stanley J. Reeves.

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

A. K. Katsaggelos is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208-3118 USA (e-mail: aggk@ece.northwestern.edu).

Digital Object Identifier 10.1109/TIP.2006.881972

paradigm has been applied to various areas of research. For instance, Molina *et al.* [20] applied this paradigm to image restoration, Mateos *et al.* [21] to removing blocking artifacts in compressed images, and Galatsanos *et al.* [19] in deconvolution problems partially known blurs.

In the hierarchical approach to blind deconvolution, we have at least two stages. In the first stage, knowledge about the structural form of the observation noise and the structural behavior of the image and PSF is used in forming $\mathrm{p}(g|f,h,\Omega)$, $\mathrm{p}(f|\Omega)$, and $\mathrm{p}(h|\Omega)$, respectively. These noise, image, and blur models depend on the unknown hyperparameters Ω. In the second stage, a hyperprior on the hyperparameters is defined, thus allowing the incorporation of information about these hyperparameters into the process. We note here that each of the three above mentioned conditional distributions will depend only on a subset of Ω, but we use this more general notation until we precisely describe the parameters that define Ω.

For Ω, f, h, and g, the following joint distribution is defined

$$\mathrm{p}(\Omega, f, h, g) = \mathrm{p}(\Omega)\mathrm{p}(f|\Omega)\mathrm{p}(h|\Omega)\mathrm{p}(g|f,h,\Omega) \tag{4}$$

and inference is based on $\mathrm{p}(\Omega, f, h|g)$.

At least three crucial questions have to be addressed when modeling and performing inference for blind deconvolution problems using the hierarchical Bayesian paradigm.

The first one relates to the definition of $\mathrm{p}(\Omega)$. Blind deconvolution is an ill-posed problem, which in a very simplistic way and without considering the fact that the PSF values add to one, consists of estimating two numbers whose product is known. Clearly, there are a number of pairs of numbers whose product is the same. Consequently, the more information we add to the solution process the more accurate the estimates of the unknown parameters will be.

The second crucial problem to be considered is to decide how inference will be carried out. A commonly used approach consists of estimating the hyperparameters in Ω by using

$$\hat{\Omega} = \arg\max_{\Omega} \mathrm{p}(\Omega|g) = \int_f \int_h \mathrm{p}(\Omega, f, h, g)\, df\, dh \tag{5}$$

and then estimating the image and blur by solving

$$\hat{f}, \hat{h} = \arg\max_{f,h} \mathrm{p}(f, h|\hat{\Omega}, g). \tag{6}$$

This inference procedure aims at optimizing a given function and not at obtaining posterior distributions that can be simulated to obtain additional information on the quality of the estimates. The solution of the above equations for estimates of the elements of Ω, the image, and the blur can be viewed as the approximation of posterior distributions by delta functions. Instead of having a distribution over all possible values of the parameters, image, and blur, the above inference procedure chooses a specific set of values. This means that we have neglected many other interpretations of the data. If the posterior is sharply peaked, other values of the hyperparameters, image, and blur will have a much lower posterior probability but, if the posterior is broad, choosing a unique value will neglect many other choices of them with similar posterior probabilities. This is relevant to our blind deconvolution problem, where the choice of broad priors on the unknown hyperparameters leads to broad posterior distributions. Note that when the hyperprior on the hyperparameters is given by $p(\Omega) = \text{constant}$, the solution of (5) is the maximum likelihood estimate of the hyperparameters given the observations g (see, for instance, [22]–[24] for the use of this model).

The third crucial problem to be solved when using the Bayesian paradigm on blind deconvolution problems is to decide how to calculate $\mathrm{p}(\Omega, f, h|g)$. The Laplace approximation of distributions has been used in problems where the blur is partially known [19], [25] or in order to calculate $\hat{f}$, $\hat{h}$ when the parameters of the distributions for the image, blur, and noise are assumed known [16], [26]. An alternative method is provided by variational distribution approximation. This approximation can be thought of as being between the Laplace approximation (see, for instance, [19] and [25]) and sampling methods [27]. The basic underlying idea is to approximate $p(\Omega, f, h|g)$ with a simpler distribution, usually one which assumes that f, h, and the hyperparameters are independent given the data (see [28, Ch. II] for an excellent introduction to variational methods and their relationships to other inference approaches).

The last few years have seen a growing interest in the application of variational methods [29], [30] to inference problems. These methods attempt to approximate posterior distributions with the use of the Kullback-Leibler cross-entropy [31]. Application of variational methods to Bayesian inference problems include graphical models and neural networks [29], independent component analysis [30], mixtures of factor analyzers, linear dynamic systems, hidden Markov models [28], and support vector machines [32].

Variational methods have been recently applied to blind deconvolution problems. Miskin and MacKay [33] use gamma priors on the image and blur (see, also, [34]), which do not enforce any spatial relationship between neighboring pixels in the image or blur, and gamma distributions as hyperpriors for the unknown hyperparameters of the priors. Likas and Galatsanos [24] use normal distributions on the unknown blur and image and an improper hyperprior $(p(\Omega) = \text{constant})$ for the hyperparameters. These models and the corresponding variational posterior distribution approximations will be described and commented in the following sections when we justify the priors and hyperpriors proposed in our work and their corresponding variational posterior distribution approximations.

In this paper, we propose the use of simultaneous autoregressions as prior distributions for the image and blur, and gamma distributions for the unknown parameters (hyperparameters) of the priors and the image formation noise. We show how the gamma distributions on the unknown hyperparameters can be used to prevent the proposed blind deconvolution method from converging to undesirable image and blur estimates and also how they can be inferred in realistic situations. We apply variational methods to approximate the posterior probability of the unknown image, blur, and hyperparameters and propose two different approximations of the posterior distribution.

The rest of the paper is organized as follows. The hyperpriors, priors, and observation models proposed in this paper are described and compared to other models used in the blind deconvolution literature in Section II. Section III describes the variational approach to distribution approximation for the blind deconvolution problem, as well as, how inference is performed. We propose different approximations of the posterior distribu-

tion of the image and the blurring function, as well as, the unknown hyperparameters, based on the variational approach for the blind deconvolution problem and compare them to other approaches reported in the literature. Finally, in Section IV, experimental results and comparisons with other methods on synthetic and real images are shown and Section V concludes the paper.

II. Hyperpriors, Priors, and Observation Models Used in Blind Deconvolution

In this section, we describe the prior models for the image and blur and the observation model we propose for the first stage of the hierarchical Bayesian paradigm in blind deconvolution problems. Then, since these prior and observation models depend on unknown hyperparameters, we proceed to explain the hyperprior distributions on these hyperparameters we use.

A. First Stage: Prior Models on Images and Blurs

Our prior knowledge about the smoothness of the object luminosity distribution makes it possible to model the distribution of f by a simultaneous autoregression (SAR) [35], that is

$$\mathrm{p}(f|\alpha_{\mathrm{im}}) \propto \alpha_{\mathrm{im}}^{N/2} \exp\left\{-\frac{1}{2}\alpha_{\mathrm{im}}\|Cf\|^2\right\} \tag{7}$$

where C denotes the Laplacian operator, $N = P \times Q$ is the size of the column vector denoting the lexicographically ordered $P \times Q$ image by rows, and $\alpha_{\mathrm{im}}^{-1}$ is the variance of the Gaussian distribution. To be precise, we should use $N-1$ instead of N in (7), since the Gaussian distribution we are using for f is singular, that is $Cf = 0$, when $f = \mathrm{const} \times \mathbf{1}$, for all $\mathrm{const} \in \mathcal{R}$. This priori model has also been used in [24].

We use the same model for the PSF, that is

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

where C denotes again the Laplacian operator, $M = U \times V$ is the size of the support of the blur, h is a column vector of size $N = P \times Q$ formed by lexicographically ordering the blur by rows (this vector has all its components equal to zero outside the region of support of the blur), and $\alpha_{\mathrm{bl}}^{-1}$ is the variance of the Gaussian distribution.

Instead of the prior blur model defined in (8), the blur model used in [24] is

$$h \sim \mathcal{N}\left(m_h, \alpha_h^{-1} I\right) \tag{9}$$

where m_h is the unknown vector mean and $1/\alpha_h$ is the unknown variance of the multidimensional normal distribution. Note that the components of h are assumed statistically independent and the number of unknowns in this distribution equals the size of the support of the blur plus one (the variance).

Let us denote by u either the image or blur. At a higher level of complexity, we can model the distribution of u by

$$u \sim \mathcal{N}(m_u, \Sigma_u) \tag{10}$$

where m_u and Σ_u denote the unknown vector mean and covariance matrix of the normal distribution. One of the problems with the use of this model is that unless the vector mean and covariance matrix are known its use leads to the simultaneous estimation of a very large number of hyperparameters.

B. First Stage. Observation Model

By assuming that the observation noise in (1) or (3) is Gaussian with zero mean and variance equal to β^{-1}, the probability of the observed image g, if f and h were respectively the "true" image and blur, is equal to

$$\mathrm{p}(g|f,h,\beta) \propto \beta^{N/2} \exp\left[-\frac{1}{2}\beta\|g - Hf\|^2\right]. \tag{11}$$

Similarly, we can use f to form the $N \times N$ convolution matrix F and rewrite (11) as

$$\mathrm{p}(g|f,h,\beta) \propto \beta^{N/2} \exp\left[-\frac{1}{2}\beta\|g - Fh\|^2\right]. \tag{12}$$

C. Second Stage: Hyperprior on the Hyperparameters

An important problem is the estimation of the parameters α_{im}, α_{bl}, and β in (7), (8), and (11), respectively, when they are unknown. To deal with this estimation problem the hierarchical Bayesian paradigm introduces a second stage (the first stage consisting again of the formulation of $\mathrm{p}(f|\alpha_{\mathrm{im}})$, $\mathrm{p}(h|\alpha_{\mathrm{bl}})$, and $\mathrm{p}(g|f,h,\beta)$). In this stage, the hyperprior $\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta)$ is also formulated, resulting in the joint global distribution

$$\begin{aligned}&\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h, g)\\ &\quad = \mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta)\mathrm{p}(f|\alpha_{\mathrm{im}})\mathrm{p}(h|\alpha_{\mathrm{bl}})\mathrm{p}(g|f,h\beta).\end{aligned} \tag{13}$$

A large part of the Bayesian literature is devoted to finding hyperprior distributions $\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta)$ for which $\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h|g)$ can be calculated in a straightforward way or be approximated. These are the so-called conjugate priors [36], which were developed extensively in Raiffa and Schlaifer [37].

Besides providing for easy calculation or approximations of $\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h|g)$, conjugate priors have, as we will see later, the intuitive feature of allowing one to begin with a certain functional form for the prior and end up with a posterior of the same functional form, but with the parameters updated by the sample information.

Taking the above considerations about conjugate priors into account, we will assume that each of the hyperparameters has as hyperprior the gamma distribution, $\Gamma(\omega|a_\omega^o, b_\omega^o)$, defined by

$$\mathrm{p}(\omega) = \Gamma\left(\omega|a_\omega^o, b_\omega^o\right) = \frac{\left(b_\omega^o\right)^{a_\omega^o}}{\Gamma\left(a_\omega^o\right)}\omega^{a_\omega^o - 1}\exp\left[-b_\omega^o\omega\right] \quad (14)$$

where $\omega > 0$ denotes a hyperparameter, $b_\omega^o > 0$ is the scale parameter, and $a_\omega^o > 0$ is the shape parameter. These parameters are assumed known. We will show how they can be calculated in the experimental section. The gamma distribution has the following mean, variance and mode

$$E[\omega] = \frac{a_\omega^o}{b_\omega^o}, \quad Var[\omega] = \frac{a_\omega^o}{\left(b_\omega^o\right)^2}, \quad \mathrm{Mode}[\omega] = \frac{a_\omega^o - 1}{b_\omega^o}. \quad (15)$$

Note that the mode does not exist when $a_\omega^o \leq 1$ and that mean and mode do not coincide.

We note here that the model proposed in [24] has as hyperparameters α_{im}, m_h, α_h, and β defined in (7), (9), and (11) and uses as hyperprior on these hyperparameters

$$\mathrm{p}(\alpha_{\mathrm{im}}, m_h, \alpha_h, \beta) \propto \mathrm{const}. \quad (16)$$

The problem with this hyperprior is that, as we will see in the experimental section, the estimation process relies exclusively on the observations, and, therefore, it is very sensitive to the amount of observational noise, as well as the initial estimates of the hyperparameters.

We note here that, for the components of the vector mean m_u in (9), the corresponding conjugate prior is a normal distribution. Furthermore, if we want to use the prior model in (10), the hyperprior for Σ_u is given by an inverse Wishart distribution (see [38]).

III. Bayesian Inference and Variational Approximation of the Posterior Distribution for Blind Deconvolution Problems

For our selection of hyperparameters in the previous section, the set of all hyperparameters Ω introduced in Section I is given by

$$\Omega = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta) \quad (17)$$

and the set of all unknown Θ is given by

$$\Theta = (\Omega, f, h) = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h). \quad (18)$$

As already known, the Bayesian paradigm dictates that inference on θ should be based on

$$\mathrm{p}(\Theta|g) = \mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h|g) = \frac{\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h, g)}{\mathrm{p}(g)} \quad (19)$$

where $\mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h, g)$ is given by (13).

Once $\mathrm{p}(\Theta|g)$ has been calculated, f and h can be integrated out to obtain $\mathrm{p}(\Omega|g) = \mathrm{p}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta|g)$. This distribution is then used to simulate or select the hyperparameters. If a point estimate, $\hat{\alpha}_{\mathrm{im}}$, $\hat{\alpha}_{\mathrm{bl}}$, $\hat{\beta}$, is required, then the mode or the mean of this posterior distribution can be used. Finally, a point estimate of the original image and blur, $\hat{f}$ and $\hat{h}$, can be obtained by maximizing $\mathrm{p}(f, h|g, \hat{\alpha}_{\mathrm{im}}, \hat{\alpha}_{\mathrm{bl}}, \hat{\beta})$. Alternatively, the mean value of this posterior distribution can be selected as the estimate of the image and blur.

From the above discussion, it is clear that, in order to perform inference, we need to either calculate or approximate the posterior distribution $\mathrm{p}(\Theta|g)$. Since $\mathrm{p}(\Theta|g)$ can not be found in closed form, we will apply variational methods to approximate this distribution by the distribution $\mathrm{q}(\Theta)$.

The variational criterion used to find $\mathrm{q}(\Theta)$ is the minimization of the Kullback–Leibler divergence, given by [31], [39]

$$\begin{aligned} C_{KL}\left(\mathrm{q}(\Theta)||\mathrm{p}(\Theta|g)\right) &= \int_\Theta \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta|g)}\right)d\Theta \\ &= \int_\Theta \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta, g)}\right)d\Theta \\ &\quad + \mathrm{const} \end{aligned} \quad (20)$$

which is always non negative and equal to zero only when $\mathrm{q}(\Theta) = \mathrm{p}(\Theta|g)$. We note in passing that the term Ensemble Learning has also been used to denote the variational approximation of distributions (see [30, p. 20]).

We choose to approximate the posterior distribution $\mathrm{p}(\Theta|g)$ by the distribution

$$\mathrm{q}(\Theta) = \mathrm{q}(\Omega)\mathrm{q}(f)\mathrm{q}(h) \quad (21)$$

where $\mathrm{q}(f)$ and $\mathrm{q}(h)$ denote distributions on f and h, respectively, and $\mathrm{q}(\Omega)$ is given by

$$\mathrm{q}(\Omega) = \mathrm{q}(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta) = \mathrm{q}(\alpha_{\mathrm{im}})\mathrm{q}(\alpha_{\mathrm{bl}})\mathrm{q}(\beta). \quad (22)$$

We now proceed to find the best of these distributions in the divergence sense.

For $\theta \in \{\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, f, h\}$, let us denote by Θ_θ the subset of Θ with θ removed; for instance, if $\theta = f$, $\Theta_f = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, h)$. Then, (20) can be written as

$$\begin{aligned} &C_{KL}\left(\mathrm{q}(\Theta)||\mathrm{p}(\Theta|g)\right) \\ &\quad = C_{KL}\left(\mathrm{q}(\theta)\mathrm{q}(\Theta_\theta)||\mathrm{p}(\Theta|g)\right) \\ &\quad = \mathrm{const} + \int_\theta \mathrm{q}(\theta) \\ &\qquad \times \left(\int_{\Theta_\theta} \mathrm{q}(\Theta_\theta)\log\left(\frac{\mathrm{q}(\theta)\mathrm{q}(\Theta_\theta)}{\mathrm{p}(\theta, \Theta_\theta, g)}\right)d\Theta_\theta\right)d\theta. \end{aligned} \quad (23)$$

Now, given $\mathrm{q}(\Theta_\theta) = \prod_{\rho\neq\theta}\mathrm{q}(\rho)$ (if, for instance, $\theta = f$ then $\mathrm{q}(\Theta_f) = \mathrm{q}(\alpha_{\mathrm{im}})\mathrm{q}(\alpha_{\mathrm{bl}})\mathrm{q}(\beta)\mathrm{q}(h)$), an estimate of $\mathrm{q}(\theta)$ is obtained as

$$\hat{\mathrm{q}}(\theta) = \arg\min_{\mathrm{q}(\theta)} C_{KL}\left(\mathrm{q}\left(\theta\right)\mathrm{q}(\Theta_\theta)\right)||\mathrm{p}(\Theta|g))\,.$$

The differentiation of (23) with respect to $\mathrm{q}(\theta)$ results in (see [30, Eq. 2.28])

$$\hat{\mathrm{q}}(\theta) = \mathrm{const} \times \exp\left(\mathrm{E}\left[\log \mathrm{p}(\Theta)\mathrm{p}(g|\Theta)\right]_{\mathrm{q}(\Theta_\theta)}\right) \quad (24)$$

where

$$\mathrm{E}\left[\log \mathrm{p}(\Theta)\mathrm{p}(g|\Theta)\right]_{\mathrm{q}(\Theta_\theta)} = \int \log \mathrm{p}(\Theta)\mathrm{p}(g|\Theta)\mathrm{q}(\Theta_\theta)d\Theta_\theta.$$

The above equations lead to the following iterative procedure to find $\mathrm{q}(\Theta)$.

Algorithm 1

Given $\mathrm{q}^1(h)$, $\mathrm{q}^1(\alpha_{\mathrm{im}})$, $\mathrm{q}^1(\alpha_{\mathrm{bl}})$, and $\mathrm{q}^1(\beta)$ the initial estimates of the distributions $\mathrm{q}(h)$, $\mathrm{q}(\alpha_{\mathrm{im}})$, $\mathrm{q}(\alpha_{\mathrm{bl}})$, and $\mathrm{q}(\beta)$ for $k = 1, 2, \ldots$ until a stopping criterion is met.

1) Find

$$\mathrm{q}^k(f) = \arg\min_{\mathrm{q}(f)} \times C_{KL}\left(\mathrm{q}^k(\alpha_{\mathrm{im}})\mathrm{q}^k(\alpha_{\mathrm{bl}})\mathrm{q}^k(\beta)\mathrm{q}(f)\mathrm{q}^k(h)\|\mathrm{p}(\Theta|g)\right). \quad (25)$$

2) Find

$$\mathrm{q}^{k+1}(h) = \arg\min_{\mathrm{q}(h)} \times C_{KL}\left(\mathrm{q}^k(\alpha_{\mathrm{im}})\mathrm{q}^k(\alpha_{\mathrm{bl}})\mathrm{q}^k(\beta)\mathrm{q}^k(f)\mathrm{q}(h)\|\mathrm{p}(\Theta|g)\right). \quad (26)$$

3) Find

$$\mathrm{q}^{k+1}(\alpha_{\mathrm{im}}) = \arg\min_{\mathrm{q}(\alpha_{\mathrm{im}})} \times C_{KL}\left(\mathrm{q}(\alpha_{\mathrm{im}})\mathrm{q}^k(\alpha_{\mathrm{bl}})\mathrm{q}^k(\beta)\mathrm{q}^k(f) \times \mathrm{q}^{k+1}(h)\|\mathrm{p}(\Theta|g)\right) \quad (27)$$

$$\mathrm{q}^{k+1}(\alpha_{\mathrm{bl}}) = \arg\min_{\mathrm{q}(\alpha_{\mathrm{bl}})} \times C_{KL}\left(\mathrm{q}^k(\alpha_{\mathrm{im}})\mathrm{q}(\alpha_{\mathrm{bl}})\mathrm{q}^k(\beta)\mathrm{q}^k(f) \times \mathrm{q}^{k+1}(h)\|\mathrm{p}(\Theta|g)\right) \quad (28)$$

$$\mathrm{q}^{k+1}(\beta) = \arg\min_{\mathrm{q}(\beta)} \times C_{KL}\left(\mathrm{q}^k(\alpha_{\mathrm{im}})\mathrm{q}^k(\alpha_{\mathrm{bl}})\mathrm{q}(\beta)\mathrm{q}^k(f) \times \mathrm{q}^{k+1}(h)\|\mathrm{p}(\Theta|g)\right). \quad (29)$$

We note here that the distributions of the hyperparameters are updated in parallel in the above algorithm. The same distributions would have been obtained if the updating had been done sequentially since $\log p(\Theta|g)$ does not contain terms involving pairs of hyperparameters. As stopping criterion of the above iterations, the convergence of the parameters defining the distributions $\mathrm{q}^k(f)$, $\mathrm{q}^{k+1}(h)$, $\mathrm{q}^{k+1}(\alpha_{\mathrm{im}})$, $\mathrm{q}^{k+1}(\alpha_{\mathrm{bl}})$, and $\mathrm{q}^{k+1}(\beta)$ can be used. In order to simplify the above criterion, $\|E[f]_{\mathrm{q}^k(f)} - E[f]_{\mathrm{q}^{k-1}(f)}\|^2 / \|E[f]_{\mathrm{q}^{k-1}(f)}\|^2 < \epsilon$, where ϵ is a prescribed bound, can also be used for terminating algorithm 1. Note that this is a convergence criterion over the image but it normally implies convergence on the posterior hyperparameter and blur distributions, since their convergence is required for the convergence of the posterior distribution of the image.

Regarding the convergence of the algorithm we first note that, by construction, at every iteration of the distributions of the image, blur, and hyperparameters the value of the Kullback–Leibler divergence decreases. To gain further insight into the above algorithm, let us consider a degenerate distribution, $\mathrm{q}(\Omega)$, that is

$$\mathrm{q}(\Omega) = \begin{cases} 1, & \text{if } \Omega = \underline{\Omega} \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

and use $\mathrm{q}^*(f,h) = \mathrm{p}(f,h|g,\Omega)$, a conditional distribution which can not be calculated for our problem but we use it to illustrate how algorithm 1 works.

If at the kth iteration of algorithm 1, $\mathrm{q}^k(\Omega)$ is a degenerate distribution on $\underline{\Omega}_k$, then the step of algorithm 1 to update the image and blur produces

$$\mathrm{q}^{*k}(f,h) = \mathrm{p}(f,h|g,\underline{\Omega}_k) \quad (31)$$

and the step in algorithm 1 to update the degenerate distribution on the hyperparameters produces

$$\Omega^{k+1} = \arg\max_{\Omega} E\left[\log(p(\Omega,f,h,g)\right]_{\mathrm{q}^{*k}(f,h)}. \quad (32)$$

Interestingly, this is the EM formulation of the maximum *a posteriori* (MAP) estimation of the hyperparameters (see [40]) for our blind deconvolution problem. What algorithm 1 does is to replace $\mathrm{q}^{*k}(f,h)$ by a distribution easier to calculate and also to replace the search for just one hyperparameter by the search for the best distribution on the hyperparameters.

A. Optimal Random Distributions for $\mathrm{q}(f)$ and $\mathrm{q}(h)$

We now proceed to explicitly calculate the distributions $\mathrm{q}^k(f)$, $\mathrm{q}^{k+1}(h)$, $\mathrm{q}^{k+1}(\alpha_{\mathrm{im}})$, $\mathrm{q}^{k+1}(\alpha_{\mathrm{bl}})$ and $\mathrm{q}^{k+1}(\beta)$ in the above algorithm. Let us now assume that at the kth iteration step of the above algorithm the distribution of h has mean vector and covariance matrix given by

$$E[h]_{q^k(h)} = E^k(h), cov[h]_{q^k(h)} = cov^k(h) \quad (33)$$

and for the distribution of the hyperparameters, we have

$$E[\alpha_{\mathrm{im}}]_{q^k(\alpha_{\mathrm{im}})} = \alpha_{\mathrm{im}}^k,\ E[\alpha_{\mathrm{bl}}]_{q^k(\alpha_{\mathrm{bl}})} = \alpha_{\mathrm{bl}}^k,\ E[\beta]_{q^k(\beta)} = \beta^k. \quad (34)$$

Then, from (24), we have that the best estimate of the *a posteriori* conditional distribution of the real image given the observation is given by the distribution $\mathrm{q}^k(f)$ satisfying

$$-2\log \mathrm{q}^k(f) = \mathrm{const} + \alpha_{\mathrm{im}}^k\|Cf\|^2 + \beta^k E\left[\|g - Hf\|^2\right]_{\mathrm{q}^k(h)} \quad (35)$$

and, thus, we have

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

The mean of the normal distribution is the solution of

$$\frac{\partial 2\log \mathrm{q}^k(f)}{\partial f} = 0$$

while the covariance is given by

$$-\frac{\partial^2 2\log \mathrm{q}^k(f)}{\partial f^2} = \left[cov^k(f)\right]^{-1}.$$

From these two equations, we obtain

$$E^k(f) = \left(M^k(f)\right)^{-1} \beta^k E^k(H)^t g \tag{36}$$

$$cov^k(f) = \left(M^k(f)\right)^{-1} \tag{37}$$

with

$$M^k(f) = \alpha_{\text{im}}^k C^t C + \beta^k E^k(H)^t E^k(H) + \beta^k cov^k(h). \tag{38}$$

Once $\text{q}^k(f)$ has been calculated, following the same steps we obtain from (26) that the solution of (26) is

$$\text{q}^{k+1}(h) = \mathcal{N}\left(h|E^{k+1}(h), cov^{k+1}(h)\right) \tag{39}$$

with

$$M^k(h) = \alpha_{\text{bl}}^k C^t C + \beta^k E^k(F)^t E^k(F) + \beta^k cov^k(f) \tag{40}$$

$$E^{k+1}(h) = \left(M^k(h)\right)^{-1} \beta^k E^k(F)^t g \tag{41}$$

$$cov^{k+1}(h) = \left(M^k(h)\right)^{-1}. \tag{42}$$

An important observation based on (38) and (40) is that, in order to be able to calculate the covariance of the distributions $\text{q}^k(f)$ and $\text{q}^{k+1}(h)$ using the discrete Fourier transform (DFT), we only need to utilize a circulant covariance matrix in the distribution $\text{q}^1(h)$. This will guarantee that the covariances of the estimates of the distributions of f and h at the kth iteration of algorithm 1 can be easily calculated using the DFT.

In order to find $q^{k+1}(\omega)$, $\omega \in \{\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta\}$ in step 3) of algorithm 1, we have to calculate the corresponding mean value $E[\log \text{p}(g|\Theta)\text{p}(\Theta)]_{\text{q}^k(f)\text{q}^{k+1}(h)}$ in (24). After some straightforward calculations, we obtain

$$\begin{aligned} &E\left[\log \text{p}(g|\Theta)\text{p}(\Theta)\right]_{\text{q}^k(f)\text{q}^{k+1}(h)} \\ &\quad = \text{const} + \sum_{\omega \in \{\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta\}} \left((a_\omega^o - 1)\log\omega - \omega b_\omega^o\right) \\ &\quad + \frac{N}{2}\log\alpha_{\text{im}} + \frac{M}{2}\alpha_{\text{bl}} + \frac{N}{2}\log\beta \\ &\quad - \frac{1}{2}\alpha_{\text{im}} E\left[\|Cf\|^2\right]_{\text{q}^k(f)} - \frac{1}{2}\alpha_{\text{bl}} E\left[\|Ch\|^2\right]_{\text{q}^{k+1}(h)} \\ &\quad - \frac{1}{2}\beta E\left[\|g - Hf\|^2\right]_{\text{q}^k(f)\text{q}^{k+1}(h)} \end{aligned} \tag{43}$$

with

$$\begin{aligned} &E\left[\|Cf\|^2\right]_{\text{q}^k(f)} \\ &\quad = \left\|CE^k(f)\right\|^2 + \text{trace}\left(C^t C cov^k(f)\right) \end{aligned} \tag{44}$$

$$\begin{aligned} &E\left[\|Ch\|^2\right]_{\text{q}^{k+1}(h)} \\ &\quad = \left\|CE^{k+1}(h)\right\|^2 + \text{trace}\left(C^t C cov^k(h)\right) \end{aligned} \tag{45}$$

$$\begin{aligned} &E\left[\|g - Hf\|^2\right]_{\text{q}^k(f)\text{q}^{k+1}(h)} \\ &\quad = \left\|g - E^{k+1}(h)E^k(f)\right\|^2 \\ &\qquad + \text{trace}\left(cov^k(f)cov^{k+1}(h)\right) \end{aligned} \tag{46}$$

$$+ \text{trace}\left(E^k(F)^t E^k(F) cov^{k+1}(h)\right) \tag{47}$$

$$+ \text{trace}\left(E^{k+1}(H)^t E^{k+1}(H) cov^k(f)\right) \tag{48}$$

where $E^k(f)$, $cov^k(f)$, $E^{k+1}(h)$, and $cov^{k+1}(h)$ have been defined in (36), (37), (41), and (42), respectively.

From (43), we have

$$\text{q}^{k+1}(\omega) = \Gamma\left(\omega|a_\omega^{k+1}, b_\omega^{k+1}\right)$$

where the parameters a_ω^{k+1} and b_ω^{k+1} are given by

$$a_{\alpha_{\text{im}}}^{k+1} = a_{\alpha_{\text{im}}}^o + \frac{N}{2} \tag{49}$$

$$b_{\alpha_{\text{im}}}^{k+1} = b_{\alpha_{\text{im}}}^o + \frac{1}{2}E\left[\|Cf\|^2\right]_{\text{q}^k(f)} \tag{50}$$

$$a_{\alpha_{\text{bl}}}^{k+1} = a_{\alpha_{\text{bl}}}^o + \frac{M}{2} \tag{51}$$

$$b_{\alpha_{\text{bl}}}^{k+1} = b_{\alpha_{\text{bl}}}^o + \frac{1}{2}E\left[\|Ch\|^2\right]_{\text{q}^{k+1}(h)} \tag{52}$$

$$a_\beta^{k+1} = a_\beta^o + \frac{N}{2} \tag{53}$$

$$b_\beta^{k+1} = b_\beta^o + \frac{1}{2}E\left[\|g - Hf\|^2\right]_{\text{q}^k(f)\text{q}^{k+1}(h)}. \tag{54}$$

These distributions have the following means:

$$E[\alpha_{\text{im}}]_{q^{k+1}(\alpha_{\text{im}})} = \frac{a_{\alpha_{\text{im}}}^o + \frac{N}{2}}{b_{\alpha_{\text{im}}}^o + \frac{1}{2}E\left[\|Cf\|^2\right]_{\text{q}^k(f)}} \tag{55}$$

$$E[\alpha_{\text{bl}}]_{q^{k+1}(\alpha_{\text{bl}})} = \frac{a_{\alpha_{\text{bl}}}^o + \frac{M}{2}}{b_{\alpha_{\text{bl}}}^o + \frac{1}{2}E\left[\|Ch\|^2\right]_{\text{q}^{k+1}(h)}} \tag{56}$$

$$E[\beta]_{q^{k+1}(\beta)} = \frac{a_\beta^o + \frac{N}{2}}{b_\beta^o + \frac{1}{2}E\left[\|g - Hf\|^2\right]_{\text{q}^k(f)\text{q}^{k+1}(h)}}. \tag{57}$$

which are then used to recalculated the distributions of f and h in algorithm 1.

We provide an interpretation of (55), (56), and (57) by rewriting them as

$$\begin{aligned} E[\alpha_{\text{im}}]_{q^{k+1}(\alpha_{\text{im}})}^{-1} &= \gamma_{\alpha_{\text{im}}} \frac{1}{\overline{\alpha}_{\text{im}}^o} + (1 - \gamma_{\alpha_{\text{im}}}) \\ &\quad \times \frac{E\left[\|Cf\|^2\right]_{\text{q}^k(f)}}{N} \end{aligned} \tag{58}$$

$$\begin{aligned} E[\alpha_{\text{bl}}]_{q^{k+1}(\alpha_{\text{bl}})}^{-1} &= \gamma_{\alpha_{\text{bl}}} \frac{1}{\overline{\alpha}_{\text{bl}}^o} + (1 - \gamma_{\alpha_{\text{bl}}}) \\ &\quad \times \frac{E\left[\|Ch\|^2\right]_{\text{q}^{k+1}(h)}}{M} \end{aligned} \tag{59}$$

$$\begin{aligned} E[\beta]_{q^{k+1}(\beta)}^{-1} &= \gamma_\beta \frac{1}{\overline{\beta}^o} + (1 - \gamma_\beta) \\ &\quad \times \frac{E\left[\|g - Hf\|^2\right]_{\text{q}^k(f)\text{q}^{k+1}(h)}}{N} \end{aligned} \tag{60}$$

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

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

The above equations indicate that $\gamma_{\alpha_{\text{im}}}$, $\gamma_{\alpha_{\text{bl}}}$, and γ_β can be understood as normalized confidence parameters. They take values in the interval [0,1). That is, when they are zero no confidence is placed on the given parameters $\overline{\alpha}_{\text{im}}^o$, $\overline{\alpha}_{\text{bl}}^o$, and $\overline{\beta}^o$, while when the corresponding normalized confidence parameter is asymptotically equal to one it fully enforces the prior knowledge of the mean (no estimation of the hyperparameters is performed).

A particularly interesting case corresponds to

$$a^o_{\alpha_{\rm im}} = a^o_{\alpha_{\rm bl}} = a_\beta = 1 \text{ and } b^o_{\alpha_{\rm im}} = b^o_{\alpha_{\rm bl}} = b^o_\beta = 0 \qquad (61)$$

which corresponds to the hyperprior model in (16). This type of hyperprior modeling (used, for instance, in [24]) makes the observation responsible for the whole estimation process. The performance of the algorithm in this case heavily depends on the level of the observation noise and the initial distributions used in the iterative process, as will also be verified experimentally.

B. Optimal Degenerate Distributions for $q(f)$ and $q(h)$

In algorithm 1, we have presented, and explicitly calculated later, the best possible approximation of the posterior distribution given its chosen factorization. However, nothing prevents us from using (23) and selecting random distributions for the image, blur, or hyperparameters, which are suboptimal, in the sense that they decrease the value of the KL divergence but not by the maximum possible amount, as was the case with the algorithm presented in the previous section. This is, in a way, similar to the use of the generalized EM (GEM) algorithms [40] instead of the EM algorithm. Notice that the GEM algorithms have to be utilized, for instance, when the unknowns that globally maximize the M-step in the EM formulation can not be easily found. In [24], $q(f)$ and $q(h)$ are assumed to be Gaussian and the distribution on the hyperparameters is assumed degenerate (assigning probability one to one value of the hyperparameters). Other alternatives are also possible.

Another suboptimal choice is to assume that $\mathrm{q}(f)$ and $\mathrm{q}(h)$ are both degenerate distributions (we will use the subscript BD to denote this approximation). Given $E^k_{\rm BD}(h)$, the current blur estimate where we assume that the degenerate distribution $\mathrm{q}^k_{\rm BD}(h)$ is located, we proceed to find $E^k_{\rm BD}(f)$.

Taking into account that the distributions on f and h are degenerate, we have in algorithm 1

$$\begin{aligned} M^k_{\rm BD}(f) &= \alpha^k_{\rm im} C^t C + \beta^k E^k_{\rm BD}(H)^t E^k_{\rm BD}(H) \\ M^k_{\rm BD}(h) &= \alpha^k_{\rm bl} C^t C + \beta^k E^k_{\rm BD}(F)^t E^k_{\rm BD}(F) \\ E^k_{\rm BD}(f) &= \left(M^k_{\rm BD}(f)\right)^{-1} \beta^k E^k_{\rm BD}(H)^t g \qquad (62) \\ E^{k+1}_{\rm BD}(h) &= \left(M^k_{\rm BD}(h)\right)^{-1} \beta^k E^k_{\rm BD}(F)^t g. \qquad (63) \end{aligned}$$

Note that the above iterative procedure is equivalent to solving

$$E^k_{\rm BD}(f) = \arg\min_f \left\{ \alpha^k_{\rm im} \|Cf\|^2 + \beta^k \left\| g - E^k_{\rm BD}(H) f \right\|^2 \right\} \qquad (64)$$

and then

$$E^{k+1}_{\rm BD}(h) = \arg\min_h \left\{ \alpha^k_{\rm bl} \|Ch\|^2 + \beta^k \left\| g - E^k_{\rm BD}(F) h \right\|^2 \right\}. \qquad (65)$$

We mention here that fixing the unknown hyperparameters and not updating them, the above iterative procedure on f and h is the same as the one proposed in [16] to jointly estimate the image and blurring functions in blind deconvolution problems.

Finally, to update the distribution of the hyperparameters in (50), (52), and (54) when using degenerate distributions on f and h, we have

$$\begin{aligned} E\left[\|Cf\|^2\right]_{\mathrm{q}^k_{\rm BD}(f)} &= \left\| C E^k_{\rm BD}(f) \right\|^2 \\ E\left[\|Ch\|^2\right]_{\mathrm{q}^{k+1}_{\rm BD}(h)} &= \left\| C E^{k+1}_{\rm BD}(h) \right\|^2 \\ E\left[\|g - Hf\|^2\right]_{\mathrm{q}^k_{\rm BD}(f)\mathrm{q}^{k+1}_{\rm BD}(h)} &= \left\| g - E^{k+1}_{\rm BD}(H) E^k_{\rm BD}(f) \right\|^2 \end{aligned}$$

where $E^k_{\rm BD}(f)$ and $E^{k+1}_{\rm BD}(h)$ have been defined in (62) and (63), respectively.

Two very important problems to be commented on. The selection of the parameters of the hyperpriors and the quality of the approximation of $\mathrm{p}(\Theta|g)$ by $\mathrm{q}(\Theta)$. The discussion on the selection of the parameters will be postponed to the experimental section.

The goodness of the approximation of $\mathrm{p}(\Theta|g)$ by $\mathrm{q}(\Theta)$ is still an open question. However, insightful comments on when the variational approximation may be tight can be found in [29] (see, also, [41]). Related to this problem is the selection of the type of probability distributions defining $\mathrm{q}(\Theta)$ and, in particular, $\mathrm{q}(f)$, and $\mathrm{q}(h)$. We believe that there is work to be done, for instance, on the modeling of the distributions of f and h by mixtures of Gaussian distributions. These mixtures will, in general, still be tractable when gamma distributions are used on the hyperparameters. Furthermore, the use of mixtures of Gaussian distributions will lead naturally to the problem of model selection by the use of Bayes factors (see, for instance, [42]–[44]).

IV. Experimental Results

A number of experiments have been performed with the proposed methods using several synthetically degraded and real astronomical images and PSFs, some of which are presented here. Henceforth, we are refering to the proposed methods as BR (both distributions of f and h are random) and BD (both distributions of f and h are degenerate). They are compared with the approach VAR1 in [24] (denoted by LG) and the method in [45], which assumes that the blur is known. The latter method (denoted by MOL), uses the prior model in (7) and the degradation model in (11), and simultaneously provides maximum likelihood estimates of the hyperparameters $\alpha_{\rm im}$ and β and the MAP estimate of the image given the estimated hyperparameters and the observed image. Since this method assumes exact knowledge of the PSF, it provides an upper bound of the achievable quality by the blind deconvolution methods.

As an objective measure of the quality of the restored image, we use the improvement in signal-to-noise ratio (ISNR) defined as $\mathrm{ISNR} = 10\log_{10}(\|f-g\|^2/\|f-\hat{f}\|^2)$, where f, g and $\hat{f}$ are respectively the original, observed, and estimated images. For all experiments, the criterion $\|E^k(f) - E^{k-1}(f)\|^2/\|E^{k-1}(f)\|^2 < 10^{-4}$ was used for terminating algorithm 1.

For the first set of experiments, the "Lena" image was blurred with a Gaussian shaped PSF with variance 9. Gaussian noise was then added to this blurred image at two noise levels, one with variance $\beta^{-1} = 0.23$ [SNR = 40 dB, Fig. 1(a)], and a second one with variance $\beta^{-1} = 16$ [SNR = 20 dB, Fig. 1(b)].

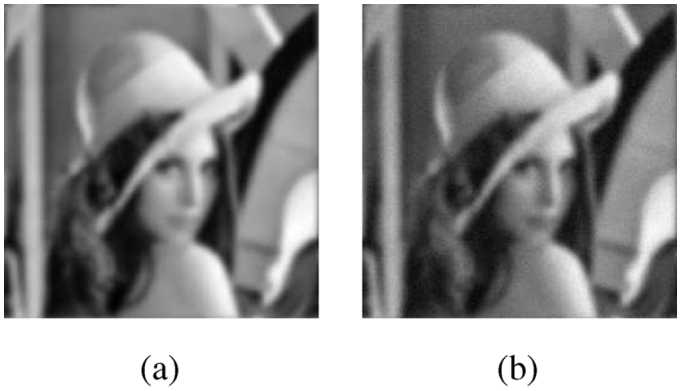


Fig. 1. Images degraded by a Gaussian shaped PSF with variance 9 and Gaussian noise of variance (a) 0.23 ($\mathrm{SNR} = 40$ dB), (b) 16 ($\mathrm{SNR} = 20$ dB).

TABLE I
ISNR VALUES, AND NUMBER OF ITERATIONS FOR THE LENA IMAGE USING $\gamma_{\alpha_{\mathrm{im}}} = \gamma_{\alpha_{\mathrm{bl}}} = \gamma_{\beta} = 0$

Noise (SNR)	Method	ISNR (dB)	iterations
40dB	*MOL*	3.83	77
	LG	2.32	150
	BD	2.60	15
	BR	2.38	96
20dB	*MOL*	2.48	68
	LG	1.10	1500
	BD	-8.31	18
	BR	1.80	25

The initial values in Algorithm 1 were chosen as follows: The observed image was used as initial estimate for $E^0(f)$. Various starting points were used for the PSF $\mathrm{q}^1(h)$, as reported below, all providing similar restoration results. The initial values $E^1[\beta]$, $E^1[\alpha_{\mathrm{im}}]$, and $E^1[\alpha_{\mathrm{bl}}]$ were then chosen according to (58)–(60), assuming a BD approximation. Note that, except for the initial value for $q^1[h]$ which has been manually fixed, all other initial parameters are automatically chosen from the available data.

For the first experiment in this set, the four methods (BR, BD, LG, and MOL) are compared when no prior information on the hyperparameters is included, that is, $\gamma_{\alpha_{\mathrm{im}}} = \gamma_{\alpha_{\mathrm{bl}}} = \gamma_{\beta} = 0$. In this case, the observations are fully responsible for the whole estimation process. For the LG method, the initial PSF is chosen to be Gaussian with variance 4, that is, an initial value close to the real one, since the method is quite sensitive to the initial parameters. The rest of the initial parameters were also empirically chosen to ensure the method produces the best results.

Table I shows the resulting ISNR and number of iterations for the BD, BR, LG and MOL methods when the initial PSF is chosen to be Gaussian with variance 4. Their corresponding restorations are displayed in Figs. 2 and 3 for the 40- and 20-dB SNR images, respectively.

As expected, all blind deconvolution methods perform worse than the MOL (PSF is known). There are, however, differences among the blind deconvolution methods. For the 40-dB SNR observed image all three blind deconvolution methods provide

Fig. 2. Restorations of 40-dB SNR Lena image using (a) the MOL method, (b) the LG method, (c) the BR method with $\gamma_{\omega} = 0, \omega \in \{\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta\}$, and (d) the BD method with $\gamma_{\omega} = 0, \omega \in \{\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta\}$.

Fig. 3. Restorations of 20-dB SNR Lena image using (a) the MOL method, (b) the LG method, (c) the BR method with $\gamma_{\omega} = 0, \omega \in \{\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta\}$, and (d) the BD method with $\gamma_{\omega} = 0, \omega \in \{\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta\}$.

similar results although the BR and, especially, the BD methods produce better ISNR values than the LG method. For the 20-dB

TABLE II
POSTERIOR MEANS OF THE DISTRIBUTIONS OF THE HYPERPARAMETERS, ISNR, AND NUMBER OF ITERATIONS FOR THE LENA IMAGE WITH 40 dB SNR USING $\overline{\beta}^o = 1/0.22, \overline{\alpha}^o_{\rm im} = 1/93.6, \overline{\alpha}^o_{\rm bl} = 1 \times 10^6$, FOR DIFFERENT VALUES OF $\gamma_\beta, \gamma_{\alpha_{\rm im}}$, AND $\gamma_{\alpha_{\rm bl}}$

Approx.	γ_β	$\gamma_{\alpha_{\rm im}}$	$\gamma_{\alpha_{\rm bl}}$	$E[\beta]^{-1}$	$E[\alpha_{\rm im}]^{-1}$	$E[\alpha_{\rm bl}]$	ISNR (dB)	iterations
BD	1.0	0.0	0.0	0.22	19.37	4.2×10^9	2.61	13
	0.0	0.0	0.7	0.21	73.72	3.3×10^9	2.95	26
	0.4	0.0	0.7	0.21	73.66	3.4×10^9	2.96	26
BR	0.8	0.0	0.0	0.22	166.9	3.5×10^8	2.21	81
	0.0	0.0	0.7	0.22	97.3	4.6×10^8	2.32	85
	1.0	0.0	0.6	0.22	99.2	4.6×10^8	2.34	70

SNR case, the BD approximation usually converges to the trivial solution, $E(f) = \text{const}$ and $E(h) = \text{const}$, that is, a flat image and a uniform PSF. It is important to note that the BD approximation provided this solution independently of the initial PSF and parameter values. The BR method always converges to a meaningful solution with higher ISNR than the one by the LG method, although both restorations are rather noisy and not all the blur has been successfully removed.

We also note here that when the BD and BR methods initialize the iteration with a Gaussian shaped PSF with variance 0.009 (a PSF close to a delta function, thus allowing the deconvolution method to make it "grow"), their resulting ISNRs are similar to the ones reported in Table I. For the 40-dB SNR example, the obtained ISNRs were 2.56 and 2.19 dB for the BD and BR methods, respectively, and for the image with 20-dB SNR, the ISNRs were -8.36 and 1.58 dB for the BD and BR methods, respectively. These values demonstrate the robustness of the proposed methods to parameter initialization.

We now examine how the introduction of additional information on the unknown hyperparameters leads to improved ISNRs for the BD and BR methods. As we have already shown when no information about the values of $\overline{\beta}^o$, $\overline{\alpha}^o_{\rm im}$, and $\overline{\alpha}^o_{\rm bl}$ is available, we can select $\gamma_\omega = 0, \omega \in \{\alpha_{\rm im}, \alpha_{\rm bl}, \beta\}$, making the observed data responsible for the estimation of the parameters. However, we usually have, at least, some information on those parameters. For instance, if we have access to the camera used to observe the scene we can estimate the observation noise by observing a uniformly flat colored object and calculating the variance of this image. The image prior variance is more difficult to estimate since it depends on the image. However, an estimation of this value could be obtained from images with the same characteristics as the image being processed. For the blur prior variance estimation, methods developed for estimating the image prior variance can also be used, due to the exchangeability between image and PSF estimation in our formulation. You and Kaveh [16] established that the image and blur parameters should follow the relation

$$\frac{\overline{\alpha}^o_{\rm im}}{\overline{\alpha}^o_{bl}} \simeq \sum_{x \in \mathcal{R}^2} \hat{f}(x) \max_{x \in \mathcal{R}^2} \hat{f}(x) \tag{66}$$

and this relationship can be used to estimate the blur prior variance.

In this paper, we propose a method to estimate $\overline{\beta}^o$, $\overline{\alpha}^o_{\rm im}$, and $\overline{\alpha}^o_{\rm bl}$ based on the use of the MOL method. The method automatically estimates these parameters, given the degraded image and an initial PSF. Unfortunately, there is no way to fix *a priori* this PSF, and a small number of trial and error experiments have to be carried out. We suggest to use a few PSFs (for our experiments we used Gaussian shaped PSFs with variance 1, 4, and 16) and chose the one that produced the "best" final restored image, based on visual observation. We want to point out that our simulations demonstrate that the proposed procedure is not very sensitive to the selection of the trial and error PSFs. By running the MOL method on the observed image using a Gaussian shaped PSF with variance equal to 4 we obtained $\overline{\beta}^o = 1/0.22$ and $\overline{\alpha}^o_{\rm im} = 1/93.6$ and $\overline{\beta}^o = 1/15.7$ and $\overline{\alpha}^o_{\rm im} = 1/206$, for the 40- and 20-dB SNR experiments, respectively. Due to the symmetry between image and blur estimation, the resulting restored image was used in the MOL method as the "true" PSF to restore the observed image; the output of the MOL is now an estimate of the PSF, the noise, and the prior variances. The values $\overline{\alpha}^o_{\rm bl} = 1 \times 10^6$ and $\overline{\alpha}^o_{\rm bl} = 2.15 \times 10^7$ for the 40- and 20-dB SNR experiments, respectively, were then obtained. Note that the relation between $\overline{\alpha}^o_{\rm im}$ and $\overline{\alpha}^o_{\rm bl}$ is similar to the one obtained using (66). Note also that we are using only the image, blur, and observation variances provided by running the MOL method twice. We have now to select the confidence parameters, $\gamma_{\alpha_{\rm im}}$, $\gamma_{\alpha_{\rm bl}}$, and γ_β. In our experiments, we initially chose a set of eleven values ranging from 0 to 1 for the confidence parameters. Once we have presented the results we will further discuss the selection of these confidence parameters.

Tables II and III show for the 40- and 20-dB SNR experiments, respectively, the means of the posterior distributions of the hyperparameters, ISNR, and the number of iterations for some selected values of the confidence parameters. The confidence parameters γ_β, $\gamma_{\alpha_{\rm im}}$, and $\gamma_{\alpha_{\rm bl}}$ are chosen to provide the maximum ISNR in the following cases: 1) when we only include information on the expected value of the noise variance, β; 2) when we only include information on the expected value of the image prior variance $\alpha_{\rm im}$ and blur prior variance $\alpha_{\rm bl}$; and 3) when information about the value of all three hyperparameters is available. From these tables the BR approximation provides a good solution when $\gamma_{\alpha_{\rm im}} = 0$ and $\gamma_{\alpha_{\rm bl}} = 0$, although better results are obtained when some information about the expected value of $E[\alpha_{\rm im}]$ and, especially, $E[\alpha_{\rm bl}]$ is provided. The noise parameter $E[\beta]$ is always accurately estimated regardless of the confidence on the parameter values.

The BD approximation again converges to the trivial solution, $E(f) = \text{const}$ and $E(h) = \text{const}$, when the noise level is high and no information about the hyperparameters $\overline{\alpha}^o_{\rm im}$ and $\overline{\alpha}^o_{\rm bl}$ is provided (see Table III), while the BR approximation

TABLE III
POSTERIOR MEANS OF THE DISTRIBUTIONS OF THE HYPERPARAMETERS, ISNR, AND NUMBER OF ITERATIONS FOR THE LENA IMAGE WITH 20 dB SNR USING $\overline{\beta}^o = 1/15.7$, $\overline{\alpha}^o_{\rm im} = 1/206$, $\overline{\alpha}^o_{\rm bl} = 2.15 \times 10^7$, FOR DIFFERENT VALUES OF γ_β, $\gamma_{\alpha_{\rm im}}$, AND $\gamma_{\alpha_{\rm bl}}$

Approx.	γ_β	$\gamma_{\alpha_{\rm im}}$	$\gamma_{\alpha_{\rm bl}}$	$E[\beta]^{-1}$	$E[\alpha_{\rm im}]^{-1}$	$E[\alpha_{\rm bl}]$	ISNR (dB)	iterations
BD	1.0	0.0	0.0	15.7	24.15	7.2×10^{10}	-0.28	66
	0.0	0.9	1.0	15.4	188.1	2.15×10^7	2.03	27
	0.3	0.9	1.0	15.7	188.1	2.15×10^7	2.04	29
BR	0.0	0.0	0.0	16.5	475	3.87×10^8	1.80	25
	0.0	0.5	1.0	16.6	226	2.15×10^8	2.02	30
	1.0	0.4	1.0	15.7	237.1	2.15×10^8	2.06	31

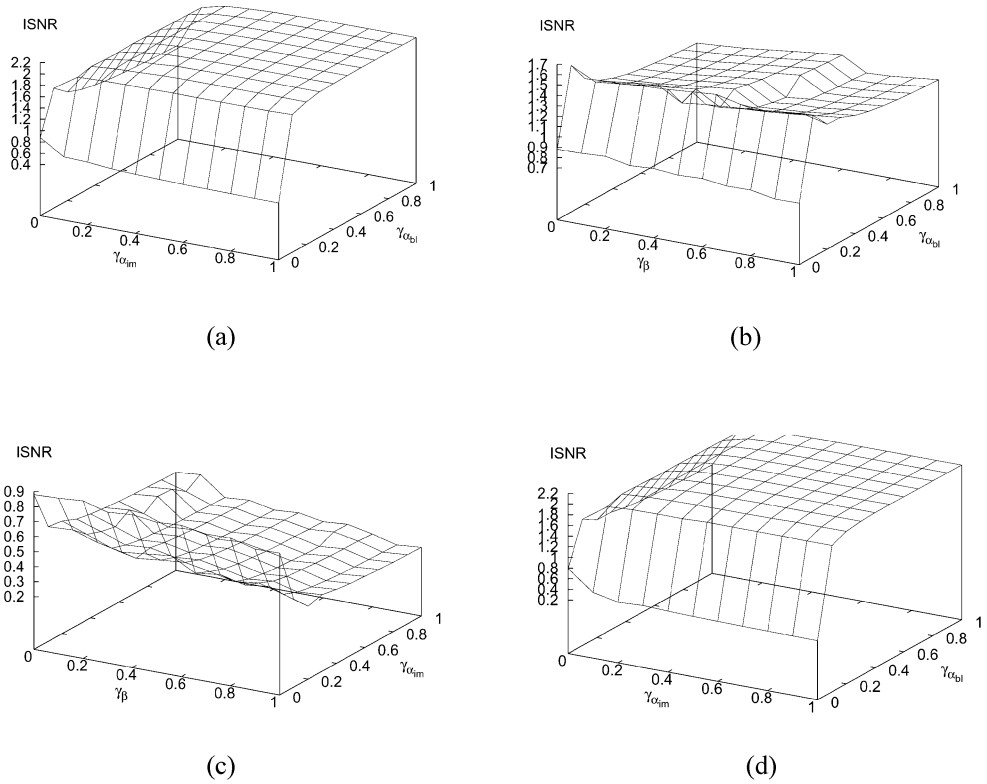


Fig. 4. ISNR evolution for different values of γ_ω, $\omega \in \{\beta, \alpha_{\rm im}, \alpha_{\rm bl}\}$ for the BR approximation applied to the Lena image with 20-dB SNR. (a) For fixed $\gamma_\beta = 0.0$, (b) for fixed $\gamma_{\alpha_{\rm im}} = 0.0$, (c) for fixed $\gamma_{\alpha_{\rm bl}} = 0.0$, and (d) for fixed $\gamma_\beta = 1.0$.

does not exhibit this behavior. When the confidence on $\overline{\alpha}^o_{\rm im}$ and $\overline{\alpha}^o_{\rm bl}$ is greater than zero, both proposed methods give useful and very similar solutions in terms of ISNR and visual quality. Note that this experiment justifies the claim that as the observation noise increases more information has to be provided to solve the blind deconvolution problem. When the noise level is low (see Table II) both BD and BR approximations provide good solutions, although better solutions are obtained by the BD method.

Fig. 4 depicts the evolution of the ISNR for a range of values of the confidence parameters for the BR approximation on the 20-dB SNR Lena image. Similar ISNR evolution is obtained for the BD approximations so their corresponding plots are not displayed. From this figure it is clear that there is almost no ISNR variation when the noise parameter confidence, γ_β, changes from 0 to 1, while increasing the value of the confidence parameters on $\overline{\alpha}^o_{\rm im}$ and especially on $\overline{\alpha}^o_{\rm bl}$ also increases the ISNR.

Restorations which provide the maximum ISNR for the Lena image with 40- and 20-dB SNRs are depicted in Figs. 5 and 6, respectively. From the displayed images it is clear that both approximations provide very similar restorations, visually and with respect to their ISNR values, when the parameters are selected so that the maximum ISNR is achieved. Fig. 7 depicts a slice through the center of the real and estimated PSFs. The estimated PSFs approximate quite well the real PSF.

From the results, we can see that the proposed methods accurately estimate the noise parameter β and, therefore, including prior information about it does not significantly increase the quality of the restorations. However, including information on the value of $\alpha_{\rm im}$ and especially $\alpha_{\rm bl}$, helps to increase the quality of the restorations. The best value for $\gamma_{\rm im}$ and $\gamma_{\rm bl}$ depends on the accuracy of the values $\overline{\alpha}^o_{\rm im}$ and $\overline{\alpha}^o_{\rm bl}$, respectively. From our experience and the results of the experiments, we suggest to use a value around 0.6–0.8 for $\alpha_{\rm bl}$ and a value close to 0.0 for $\alpha_{\rm im}$.

(a) (b)

Fig. 5. Best restorations of the image with 40-dB SNR in Fig. 1(a) for: (a) BD approximation for $\gamma_\beta = 0.4$, $\gamma_{\alpha_{\rm im}} = 0.0$, and $\gamma_{\alpha_{\rm bl}} = 0.7$ (ISNR = 2.96 dB); (b) BR approximation for $\gamma_\beta = 1.0$, $\gamma_{\alpha_{\rm im}} = 0.0$, and $\gamma_{\alpha_{\rm bl}} = 0.6$ (ISNR = 2.34 dB).

(a) (b)

Fig. 6. Best restorations of the image with 20-dB SNR in Fig. 1(b) for: (a) BD approximation for $\gamma_\beta = 0.3$, $\gamma_{\alpha_{\rm im}} = 0.9$, and $\gamma_{\alpha_{\rm bl}} = 1.0$ (ISNR = 2.04 dB); (b) BR approximation for $\gamma_\beta = 1.0$, $\gamma_{\alpha_{\rm im}} = 0.4$, and $\gamma_{\alpha_{\rm bl}} = 1.0$ (ISNR = 2.06 dB).

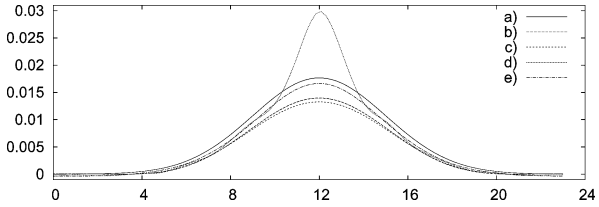
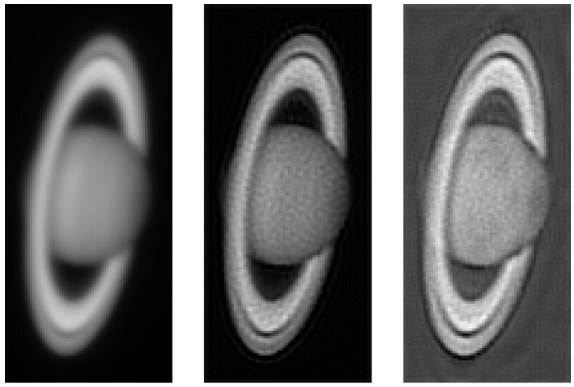


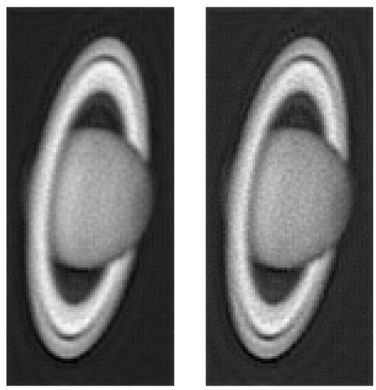
Fig. 7. One-dimensional slice through the origin of the original and estimated PSFs: a) real PSF; b) estimated BR PSF with $\gamma_{\rm im} = 1.0$, $\gamma_{\rm bl} = 0.4$, and $\gamma_\beta = 1.0$, for 20-dB SNR; c) estimated BD PSF with $\gamma_{\rm im} = 0.3$, $\gamma_{\rm bl} = 0.9$, and $\gamma_\beta = 1.0$, for 20-dB SNR; d) estimated BR PSF with $\gamma_{\rm im} = 1.0$, $\gamma_{\rm bl} = 0.0$, and $\gamma_\beta = 0.6$, for 40-dB SNR; e) estimated BD PSF with $\gamma_{\rm im} = 0.4$, $\gamma_{\rm bl} = 0.0$, and $\gamma_\beta = 0.7$, for 40-dB SNR.

Regarding the convergence of the proposed methods, both the BD and BR methods typically require only 25–50 iterations to reach convergence. This is a small number of iterations but also, since all calculations can be performed in the Fourier domain, each iteration takes only about 0.043 CPU seconds on a Xeon 3200 processor.

As a general comment, as expected, both BD and BR approximations with additional information achieve much better ISNRs than the LG method and the BD and BR methods without the inclusion of additional information on the hyperparameters (see results in Tables I–III). When introducing additional information on the hyperparameters, the resulting restorations are sharper and most of the blur is removed. The results are close



(a) (b) (c)



(d) (e)

Fig. 8. (a) Observed Saturn image. Restoration with the method (b) in [45]; (c)BR with $\gamma_\beta = \gamma_{\alpha_{\rm im}} = \gamma_{\alpha_{\rm bl}} = 0.0$; (d) BD with $\gamma_\beta = 0.8$, $\gamma_{\rm im} = 0.1$, and $\gamma_{\rm bl} = 0.8$; (e) BR with $\gamma_\beta = 0.8$, $\gamma_{\rm im} = 0.1$, and $\gamma_{\rm bl} = 0.8$.

to the results by the MOL method, both visually and in terms of the ISNR values.

For the second experiment, the 96 × 250 image of Saturn taken at the Cassegrain f/8 focus of the 1.52-m telescope at the Calar Alto Observatory (Spain) in July 1991, depicted in Fig. 8(a), was used. The image was taken through a narrow-band interference filter centered at the wavelength of 9500 Å with values ranging from 100 to 6100.

Although there is no exact expression describing the shape of the PSF for images taken from ground based telescopes, previous studies [46], [47] have suggested the following radially symmetric approximation for the PSF

$$h(r) \propto \left(1 + \frac{r^2}{R^2}\right)^{-\delta} \tag{67}$$

where the parameters δ and R were estimated from the intensity profiles of satellites of Saturn that were recorded simultaneously with the planet and from stars that were recorded very close in time and airmass to the planetary images. The estimates we obtained are $\delta \sim 3$ and $R \sim 3.4$ pixels.

The values for the distribution parameters were estimated using the MOL method with the PSF in (67) being the "true" PSF, thus obtaining $\overline{\beta}^o = 1/150$ and $\overline{\alpha}^o_{\rm im} = 1/99000$. As in the previous experiment, the resulting restoration was provided to MOL method as "true" PSF to restore the observed image (note that this restoration process provides a blur estimate). This

TABLE IV
OBTAINED VALUE OF THE PARAMETERS AND NEEDED NUMBER OF ITERATIONS FOR DIFFERENT VALUES OF $\gamma_\omega, \omega \in \{\beta, \alpha_{\rm im}, \alpha_{\rm bl}\}$ FOR THE SATURN IMAGE FOR $\overline{\beta}^o = 1/150, \overline{\alpha}_{\rm im} = 1/99000$ AND $\overline{\alpha}_{\rm bl} = 3 \times 10^7$

γ_β	$\gamma_{\alpha_{\rm im}}$	$\gamma_{\alpha_{\rm bl}}$	Method	$E[\beta]$	$E[\alpha_{\rm im}]$	$E[\alpha_{\rm bl}]$	iters.
0.0	0.0	0.0	*BD*	1/520	1/49572	1.08×10^8	134
			BR	1/185	1/184445	9.81×10^8	76
0.8	0.1	0.8	*BD*	1/129	1/36197	2.63×10^7	78
			BR	1/155	1/88734	2.9×10^7	62

resulted in a value of $\overline{\alpha}^o_{\rm bl} = 3 \times 10^7$. For comparison purposes, we present the restoration obtained by the MOL method in Fig. 8(b).

Our experiments show that, again, the BD approximation converges to the trivial solution when the values of the confidence parameters are close to zero, while the BR approximation produces the restoration shown in Fig. 8(c). In order to obtain better restorations, we have to include information about the parameter values. Following the approach described in the previous experiments we have selected $\gamma_\beta = 0.8$, $\gamma_{\alpha_{\rm im}} = 0.1$, and $\gamma_{\alpha_{\rm bl}} = 0.8$. Using these parameters the restorations depicted in Fig. 8(d) and 8(e) for the BD, and BR approximations, respectively, are obtained.

The estimated parameters, as well as the required number of iterations to reach convergence are shown in Table IV. This table shows that the BR approximation always obtains accurate estimates of the noise parameter β even when no information about the value of the parameter is provided, while the BD approximation needs some information about the parameter values to provide useful estimations.

In all restored images, the improvement in spatial resolution is evident. In particular, the ring light contribution has been successfully removed from equatorial regions close to the actual location of the rings and amongst the rings of Saturn, the Cassini division is enhanced in contrast, and the Encke division appears on the ansae of the rings in most of the deconvolved images. The restorations provided by the proposed methods are almost indistinguishable and not as noisy as the estimation provided by the MOL method. This may be due to the fact that the PSF approximation in (67) does not take into account small atmospheric turbulence that distorts the theoretical blur. In order to compare the estimated PSFs, Fig. 9 depicts an 1-D slice through the origin of the theoretical PSF in (67) and the estimated ones. This plot shows that the estimated PSF by BR approximation with $\gamma_\beta = \gamma_{\alpha_{\rm im}} = \gamma_{\alpha_{\rm bl}} = 0.0$ is flatter and not as accurate as the one obtained with $\gamma_\beta = 0.8$, $\gamma_{\rm im} = 0.1$, and $\gamma_{\rm bl} = 0.8$. The PSFs obtained by the proposed methods for this case are close to the theoretical PSF.

V. CONCLUSION

New methods for the simultaneous estimation of the image, blur, and unknown hyperparameters in blind deconvolution problems have been proposed, based on the variational approach to distribution approximation. Using this approach, we can approximate the posterior distribution of the image and blurring function, as well as, the unknown hyperparameters. The proposed methods have been analyzed, validated, and compared experimentally with synthetic and real data. Useful recommendations are provided regarding initial conditions and the values of the confidence parameters.

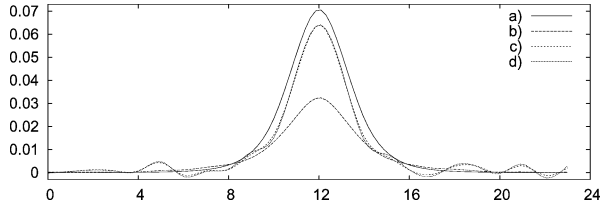


Fig. 9. One-dimensional slice through the origin of the original and estimated PSFs for the Saturn image: (a) theoretical PSF; b) estimated BR PSF with $\gamma_\beta = \gamma_{\rm im} = \gamma_{\rm bl} = 0.0$; c) estimated BR PSF with $\gamma_\beta = 0.8$, $\gamma_{\rm im} = 0.1$, and $\gamma_{\rm bl} = 0.8$; d) estimated BD PSF with $\gamma_\beta = 0.8$, $\gamma_{\rm im} = 0.1$, and $\gamma_{\rm bl} = 0.8$.

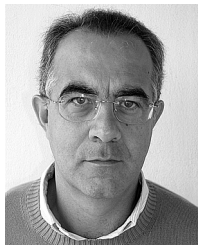
ACKNOWLEDGMENT

The authors would like to thank Dr. A. C. Likas and Dr. N. Galatsanos in the Department of Computer Science, University of Ioannina, Greece, for their fruitful discussions in the course of this work and for providing the estimated images and blurs by their VAR1 method in [24].

REFERENCES

[1] A. K. Katsaggelos, Ed., *Digital Image Restoration*. New York: Springer-Verlag, 1991.
[2] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Proces. Mag.*, vol. 14, no. 2, pp. 24–41, Feb. 1997.
[3] R. Molina, J. Núñez, F. J. Cortijo, and J. Mateos, "Image restoration in Astronomy. A Bayesian perspective," *IEEE Signal Process. Mag.*, vol. 18, no. 2, pp. 11–29, Feb. 2001.
[4] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, Mar. 1996.
[5] ——, "Blind image deconvolution revisited," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 61–63, Jun. 1996.
[6] J. Markham and J. A. Conchello, "Parametric blind deconvolution of microscopic images: further results," in *Proc. SPIE Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing V*, C. J. Cogswell, J. A. Conchello, and T. Wilson, Eds., 1998, vol. 3261, pp. 38–49.
[7] A. Savakis and H. J. Trussell, "Blur identification by residual spectral matching," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 141–151, Apr. 1993.
[8] J. Elder and S. Zucker, "Local scale control for edge detection and blur estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 7, pp. 699–716, Jul. 1998.
[9] F. Rooms, M. Ronsse, A. Pizurica, and W. Philips, "PSF estimation with applications in autofocus and image restoration," in *Proc. 3rd IEEE Benelux Signal Processing Symp.*, 2002, pp. 13–16.
[10] F. S. Gibson and F. Lanni, "Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy," *J. Opt. Soc. Amer. A*, vol. 8, pp. 1601–1613, 1991.
[11] O. Michailovich and D. Adam, "A novel approach to the 2-D blind deconvolution problem in medical ultrasound," *IEEE Trans. Med. Imag.*, vol. 24, no. 1, pp. 86–104, Jan. 2005.
[12] T. Bretschneider, P. Bones, S. McNeill, and D. Pairman, "Image-based quality assessment of SPOT data," presented at the American Society for Photogrammetry and Remote Sensing, 2001.
[13] J. Krist, "Simulation of HST PSFs using Tiny Tim," in *Astronomical Data Analysis Software and Systems IV*, R. A. Shaw, H. E. Payne, and J. J. E. Hayes, Eds. San Francisco, CA: Astronomical Soc. of the Pacific, 1995, pp. 349–353.
[14] Y. Yang, N. P. Galatsanos, and H. Stark, "Proyection based blind deconvolution," *J. Opt. Soc. Amer. A*, vol. 11, no. 9, pp. 2410–2409, 1994.
[15] A. K. Katsaggelos and K. T. Lay, "Image identification and image restoration based on the expectation-maximization algorithm," *Opt. Eng.*, vol. 29, no. 5, pp. 436–445, 1990.
[16] 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.
[17] ——, "Blind image restoration by anisotropic regularization," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 396–407, Mar. 1999.
[18] B. D. Jeffs and J. C. Christou, "Blind Bayesian restoration of adaptive optics telescope images using generalized Gaussian Markov random field models," in *Proc. SPIE. Adaptive Optical System Technologies*, D. Bonaccini and R. K. Tyson, Eds., 1998, vol. 3353, pp. 1006–1013.

[19] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, and J. Mateos, "Hyperparameter estimation in image restoration problems with partially-known blurs," *Opt. Eng.*, vol. 41, no. 8, pp. 1845–1854, 2002.
[20] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, Feb. 1999.
[21] J. Mateos, A. Katsaggelos, and R. Molina, "A Bayesian approach to estimate and transmit regularization parameters for reducing blocking artifacts," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1200–1215, 2000.
[22] A. K. Katsaggelos and K. T. Lay, "Maximum likelihood blur identification and image restoration using the EM algorithm," *IEEE Trans. Signal Process.*, vol. 39, no. 3, pp. 729–733, Mar. 1991.
[23] ——, "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.
[24] A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2222–2233, Aug. 2004.
[25] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, and A. K. Katsaggelos, "Hierarchical Bayesian image restoration for partially-known blur," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1784–1797, Oct. 2000.
[26] R. Molina, A. K. Katsaggelos, J. Abad, and J. Mateos, "A Bayesian approach to blind deconvolution based on Dirichlet distributions," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Munich, Germany, 1997, vol. IV, pp. 2809–2812.
[27] C. Andrieu, N. de Freitras, A. Doucet, and M. Jordan, "An introduction to MCMC for machine learning," *Mach. Learn.*, vol. 50, pp. 5–43, 2003.
[28] M. Beal, "Variational Algorithms for Approximate Bayesian Inference," Ph.D. dissertation, Gatsby Comput. Neurosci. Unit, Univ. College London, London, U.K., 2003.
[29] 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*. Cambridge, MA: MIT Press, 1998, pp. 105–162.
[30] J. Miskin, "Ensemble Learning for Independent Component Analysis," Ph.D. dissertation, Astrophysics Group, Univ. Cambridge, Cambridge, U.K., 2000.
[31] S. Kullback*, Information Theory and Statistics*. New York: Dover, 1959.
[32] C. Bishop and M. Tipping, "Variational relevance vector machine," in *Proc. 16th Conf. Uncertainty in Articial Intelligence*, 2000, pp. 46–53.
[33] 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.
[34] K. Z. Adami, "Variational methods in Bayesian deconvolution," in *Proc. PHYSTAT ECONF*, 2003, vol. CO30908, p. TUGT002.
[35] B. D. Ripley*, Spatial Statistics*. New York: Wiley, 1981, pp. 88–90.
[36] J. O. Berger*, Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag, 1985, ch. 3–4.
[37] H. Raiffa and R. Schlaifer*, Applied Statistical Decision Theory*. Boston, MA: Div. Res., Graduate School of Business, Harvard Univ., 1961.
[38] A. Gelman, J. B. Carlin, H. S. Stern, and D. R. Rubin*, Bayesian Data Analysis*. New York: Chapman & Hall, 2003.
[39] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.
[40] G. J. McLachlan and T. Krishnan*, The EM Algorithm and Extensions*. New York: Wiley, 1997.
[41] 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.
[42] R. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Statist. Soc.*, vol. 90, pp. 773–795, 1995.
[43] D. Stanford and A. Raftery, "Approximate Bayes factors for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 11, pp. 1517–1520, Nov. 2002.
[44] F. Murtagh, A. E. Raftery, and J.-L. Starck, "Bayesian inference for multiband image segmentation via model-based cluster trees," *Image Vis. Comput.*, vol. 23, p. 587596, 2005.
[45] R. Molina, "On the hierarchical Bayesian approach to image restoration. Applications to astronomical images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 11, pp. 1122–1128, Nov. 1994.
[46] 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.
[47] R. Molina and B. D. Ripley, "Using spatial models as priors in astronomical image analysis," *J. Appl. Statist.*, vol. 16, pp. 193–206, 1989.



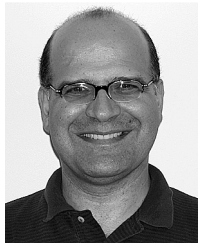
Rafael Molina was born in 1957. He received the degree in mathematics (statistics) in 1979 and the Ph.D. degree in optimal design in linear models in 1983.

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

Javier Mateos was born in Granada, Spain, in 1968. He received the degree in computer science in 1991 and the Ph.D. degree in computer science in 1998, both from the University of Granada.

He was an Assistant Professor with the Department of Computer Science and Artificial Intelligence, University of Granada, from 1992 to 2001, and then he became a permanent Associate Professor. He is coauthor of *Superresolution of Images and Video* (Claypool, 2006). He is conducting research on image and video processing, including image restoration, image, and video recovery and super-resolution from (compressed) stills and video sequences.

Prof. Mateos is a member of the Asociación Española de Reconocimento de Formas y Análisis de Imágenes (AERFAI) and International Association for Pattern Recognition (IAPR).



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

In 1985, he joined the Department of Electrical and Computer Engineering at Northwestern University, Evanston, IL, where he is currently a Professor, holding the Ameritech Chair of Information Technology. He is also the Director of the Motorola Center for Seamless Communications and a member of the Academic Affiliate Staff, Department of Medicine, at Evanston Hospital. He is the editor of *Digital Image Restoration* (New York: Springer-Verlag, 1991), coauthor of *Rate-Distortion Based Video Compression* (Norwell, MA: Kluwer, 1997) and *Superresolution of Images and Video* (Claypool, 2006), and co-editor of *Recovery Techniques for Image and Video Compression and Transmission* (Norwell, MA: Kluwer, 1998), and the co-inventor of eight international patents

Dr. Katsaggelos is a member of the Publication Board of the IEEE PROCEEDINGS, the IEEE Technical Committees on Visual Signal Processing and Communications, and Multimedia Signal Processing, the Editorial Board of Academic Press, Marcel Dekker: Signal Processing Series, *Applied Signal Processing*, and *Computer Journal*. He has served as Editor-in-Chief of the *IEEE Signal Processing Magazine* (1997–2002), member of the Publication Boards of the IEEE Signal Processing Society, the IEEE TAB Magazine Committee, Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992), Area Editor for the journal*Graphical Models and Image Processing* (1992–1995), member of the Steering Committees of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1992–1997) and the IEEE TRANSACTIONS ON MEDICAL IMAGING (1990–1999), member of the IEEE Technical Committee on Image and Multi-Dimensional Signal Processing (1992–1998), 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), and an IEEE ICME Best Paper Award (2006).