TOTAL VARIATION BLIND DECONVOLUTION USING A VARIATIONAL APPROACH TO PARAMETER, IMAGE, AND BLUR ESTIMATION

S. Derin Babacan, Rafael Molina, Aggelos K. Katsaggelos*

Department of Electrical Engineering
and Computer Science
Northwestern University, Evanston, IL 60208, USA
sdb@northwestern.edu, aggk@eecs.northwestern.edu

*Departamento de Ciencias
de la Computación e I.A.
Universidad de Granada, 18071 Granada, Spain
rms@decsai.ugr.es

ABSTRACT

In this paper we propose novel algorithms for total variation (TV) based blind deconvolution and parameter estimation utilizing a variational framework. Within a hierarchical Bayesian formulation, the reconstructed image, the blur and the unknown hyperparameters for the image prior, the blur prior and the image degradation noise are simultaneously estimated. We develop two algorithms resulting from this formulation which provide approximations to the posterior distributions of the latent variables. Different values can be drawn from these distributions as estimates to the latent variables and the uncertainty of these estimates can be measured. Experimental results are provided to demonstrate the performance of the algorithms.

1. INTRODUCTION

The general discrete model for a linear degradation caused by blurring and additive noise can be expressed in matrix-vector form as

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

where $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{n}$ represent the original image, the observed image, and the noise, respectively, all ordered lexicographically. The general objective of blind deconvolution is to estimate $\mathbf{x}$ and $\mathbf{H}$ based on $\mathbf{y}$ and prior knowledge about these unknown quantities and the noise.

Approaches to the blind deconvolution problem can in general be classified into two categories based on the stage where the blur is identified [1]. In the first category, that of *a priori* blur identification methods, the blur is identified separately from the image, and later used in one of the classical image restoration algorithms to obtain estimates for the image. The majority of the existing methods fall in the second category, consisting of joint blur identification and image restoration methods. Most methods in this category incorporate prior knowledge about the image and blur in a deterministic or stochastic formulation [1].

Methods based on the Bayesian formulation are of the most commonly used methods in the blind deconvolution literature. Such methods introduce prior models on the image, blur, and their model parameters, which impose constraints on the estimates and act as regularizers. Simultaneous Autoregression (SAR), Conditional Autoregression (CAR), and Gaussian models are some of the commonly used priors for the image and blur.

Recently there has been an interest in applying variational methods to the blind deconvolution problem. These methods attempt to obtain approximations to the posterior distributions on the unknowns with the use of the Kullback-Leibner cross entropy [2]. Miskin and Mackay [3], Adami [4], Likas and Galatsanos [5], and Molina *et. al.* [6] employ this variational methodology to the blind deconvolution problem in a Bayesian formulation.

In this paper we also apply variational methods to the blind deconvolution problem, by proposing to use a Total Variation (TV) function as the image prior, and a SAR model for the blur. Although the TV model has been used in blind deconvolution before (see, for example, [7]), to our knowledge no work has been reported on the simultaneous estimation of the model parameters, image, and blur in TV-based variational blind deconvolution. We develop two new variational methods based on a hierarchical Bayesian formulation, and provide approximations to the posterior distributions of the image, blur, and model parameters, which allow us to efficiently estimate the unknowns and also analyze their uncertainties.

This paper is organized as follows: In Sec. 2 we present the hierarchical Bayesian model and the priors on the unknown quantities. Section 3 describes the variational approximation method utilized in the Bayesian inference. We present our experimental results in Sec. 4 and conclusions are drawn in Sec. 5.

2. HIERARCHICAL BAYESIAN MODELING

In the first stage of the Bayesian formulation, we model the observation process, the image, and the blur. Assuming the degradation noise is additive and Gaussian, the probability distribution of the observation in Eq. (1) can be expressed as

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

where β^{-1} is the noise variance and h is the impulse response of the linear space-invariant degradation system, which is used in forming the block-circulant matrix $\mathbf{H}$. We adopt the TV function for the image prior, that is,

$$\mathrm{p}(\mathbf{x}|\alpha_{im}) \propto \frac{1}{Z_{\mathrm{TV}}(\alpha_{im})} \exp\left[-\alpha_{im}\mathrm{TV}(\mathbf{x})\right], \tag{3}$$

where $Z_{\mathrm{TV}}(\alpha_{im})$ is the partition function and $\alpha_{\mathrm{im}}^{-1}$ is the image variance. The TV function is defined as

$$\mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2}, \tag{4}$$

where the operators $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ correspond to, respectively, the horizontal and vertical first order differences, at pixel i, that is, $\Delta_i^h(\mathbf{x}) = x_i - x_{l(i)}$ and $\Delta_i^v(\mathbf{x}) = x_i - x_{a(i)}$, with $l(i)$ and $a(i)$ denoting the nearest neighbors of i, to the left and above, respectively. As shown in [8], this prior can be approximated by

$$\mathrm{p}(\mathbf{x}|\alpha_{im}) = \text{const} \times \alpha_{im}^{N/2} \exp\left[-\alpha_{im}\mathrm{TV}(\mathbf{x})\right], \qquad (5)$$

where N is the total number of pixels in the image. We use the SAR model for the prior on blur, that is,

$$\mathrm{p}(h|\alpha_{\mathrm{bl}}) \propto \alpha_{\mathrm{bl}}^{M/2} \exp\{-\frac{1}{2}\alpha_{\mathrm{bl}} \parallel \mathbf{C}h \parallel^2\}, \qquad (6)$$

where $\mathbf{C}$ denotes the Laplacian operator, $\alpha_{\mathrm{bl}}^{-1}$ is the variance of the Gaussian distribution, and M is the support of the blur, which is assumed to be the same as the image support.

The model parameters α_{im}, α_{bl}, and β are referred to as the *hyperparameters*, and are important in determining the performance of the algorithms to a great extent. They are in general not known, and thus we introduce a second stage in the Bayesian formulation and assume that each of these hyperparameters has a hyperprior. In this paper we utilize the gamma distribution as their hyperprior, defined by

$$\mathrm{p}(\omega) = \Gamma(\omega|a_\omega^o, b_\omega^o) = \frac{(b_\omega^o)^{-a_\omega^o}}{\Gamma(a_\omega^o)} \omega^{a_\omega^o - 1} \exp\left[-\frac{\omega}{b_\omega^o}\right], \qquad (7)$$

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

$$E[\omega] = a_\omega^o b_\omega^o,\ Var[\omega] = a_\omega^o (b_\omega^o)^2, \quad \mathrm{Mode}[\omega] = (a_\omega^o - 1) b_\omega^o. \qquad (8)$$

3. BAYESIAN INFERENCE AND VARIATIONAL APPROXIMATION OF THE POSTERIOR DISTRIBUTIONS

We denote the set of all the hyperparameters introduced in the previous section by $\Omega = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta)$ and the set of all unknowns by $\Theta = (\Omega, \mathbf{x}, h) = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, h)$. Combining the first and second stage of the Bayesian model we obtain the following joint distribution

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

The inference on $(\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, h)$ should be based on

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

Since it is not possible to directly evaluate $\mathrm{p}(\Theta|\mathbf{y})$, following the variational methodology, it is approximated by $\mathrm{q}(\Theta)$, which can be found by minimizing the Kullback-Leibler divergence, given by [9, 2]

$$C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|\mathbf{y})) = \int_\Theta \mathrm{q}(\Theta) \log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta|\mathbf{y})}\right) d\Theta = \int_\Theta \mathrm{q}(\Theta) \log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta, \mathbf{y})}\right) d\Theta + \text{const}. \qquad (9)$$

Assuming statistical independence of the latent variables, we have that $\mathrm{q}(\Theta) = \mathrm{q}(\Omega)\mathrm{q}(\mathbf{x})\mathrm{q}(h)$, where $\mathrm{q}(\Omega) = \mathrm{q}(\alpha_{\mathrm{im}})\mathrm{q}(\alpha_{\mathrm{bl}})\mathrm{q}(\beta)$.

The use of the TV prior makes the integral in Eq. (9) difficult to evaluate. Therefore, we utilize a minorization of the TV prior which renders this integral easier to evaluate. Let us define the following functional $\mathrm{M}(\alpha_{im}, \mathbf{x}, \mathbf{v})$, for α, $\mathbf{x}$, and an $N-$dimensional vector $\mathbf{v} \in (R^+)^N$

$$\mathrm{M}(\alpha_{im}, \mathbf{x}, \mathbf{v}) = \text{const} \times \alpha_{im}^{N/2} \exp\left[-\frac{\alpha_{im}}{2}\sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + v_i}{\sqrt{v_i}}\right]. \qquad (10)$$

Using the following inequality in [8] for $u \geq 0$ and $v > 0$

$$\sqrt{u} \leq \sqrt{v} + \frac{1}{2\sqrt{v}}(u - v). \qquad (11)$$

we have that

$$\exp[-\alpha_{\mathrm{im}}\mathrm{TV}(\mathbf{x})] = \exp\left[-\alpha_{\mathrm{im}}\sum_i \sqrt{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2}\right] \geq \exp\left[-\frac{\alpha_{im}}{2}\sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + v_i}{\sqrt{v_i}}\right]. \qquad (12)$$

Combining Eqs. (5), (10), and (12) we obtain the following lower bound for the image prior

$$\mathrm{p}(\mathbf{x}|\alpha_{im}) \geq \text{const} \times \mathrm{M}(\alpha_{im}, \mathbf{x}, \mathbf{v}), \qquad (13)$$

and the following lower bound for the joint probability distribution

$$\mathrm{p}(\Theta, \mathbf{y}) \geq \mathrm{p}(\Omega)\mathrm{M}(\alpha_{im}, \mathbf{x}, \mathbf{v})\mathrm{p}(h|\alpha_{bl})\mathrm{p}(\mathbf{y}|\mathbf{x}, h, \beta) = \mathrm{F}(\Theta, \mathbf{v}, \mathbf{y}). \qquad (14)$$

For $\theta \in \{\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, \mathbf{x}, h\}$ let us denote by Θ_θ the subset of Θ with θ removed; for instance, if $\theta = \mathbf{x}$, $\Theta_{\mathbf{x}} = (\alpha_{\mathrm{im}}, \alpha_{\mathrm{bl}}, \beta, h)$. Then, utilizing Eq. (14), Eq. (9) can be written as

$$C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|\mathbf{y})) \leq C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{F}(\Theta, \mathbf{v}, \mathbf{y})) = \int_\theta \mathrm{q}(\theta)\left(\int_{\Theta_\theta} \mathrm{q}(\Theta_\theta)\log\left(\frac{\mathrm{q}(\theta)\mathrm{q}(\Theta_\theta)}{\mathrm{F}(\Theta, \mathbf{v}, \mathbf{y})}\right) d\Theta_\theta\right) d\theta. \qquad (15)$$

We can utilize this upper bound to find estimates of the posterior distributions in an alternating fashion, that is, given $\mathrm{q}(\Theta_\theta)$, the posterior $\mathrm{q}(\theta)$ can be computed by solving

$$\mathrm{q}(\theta) = \arg\min_{\mathrm{q}(\theta)} C_{KL}(\mathrm{q}(\Theta_\theta)\mathrm{q}(\theta) \parallel \mathrm{F}(\Theta, \mathbf{v}, \mathbf{y})). \qquad (16)$$

Differentiation of the integral on the right hand side in Eq. (15) with respect to $\mathrm{q}(\theta)$ results in (see Eq. (2.28) in [10]),

$$\hat{\mathrm{q}}(\theta) = \text{const} \times \exp\Big(\mathrm{E}[\,\log\mathrm{F}(\Theta, \mathbf{v}, \mathbf{y})\,]_{\mathrm{q}(\Theta_\theta)}\Big), \qquad (17)$$

where

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

Applying this minimization to each unknown in an alternating way we obtain the following iterative procedure to find $q(\Theta)$:

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

1. *Find*

$$q^k(\mathbf{x}) = \arg\min_{q(\mathbf{x})} \int_{\mathbf{x}} \int_{\Theta_{\mathbf{x}}} q^k(\Theta_{\mathbf{x}}) q(\mathbf{x}) \times \log\left(\frac{q^k(\Theta_{\mathbf{x}})q(\mathbf{x})}{\mathrm{F}(\Theta^k_{\mathbf{x}}, \mathbf{x}, \mathbf{v}^k, \mathbf{y})}\right) d\Theta_{\mathbf{x}} d\mathbf{x} \quad (18)$$

2. *Find*

$$q^{k+1}(h) = \arg\min_{q(h)} \int_h \int_{\Theta_h} q^k(\Theta_h) q(h) \times \log\left(\frac{q^k(\Theta_h)q(h)}{\mathrm{F}(\Theta^k_h, h, \mathbf{v}^k, \mathbf{y})}\right) d\Theta_h dh \quad (19)$$

3. *Find*

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \int_{\Theta} q^k(\Theta_h) q^{k+1}(h) \times \log\left(\frac{q^k(\Theta_h)q^{k+1}(h)}{\mathrm{F}(\Theta^k_h, h^{k+1}, \mathbf{v}, \mathbf{y})}\right) d\Theta \quad (20)$$

4. *Find*

$$q^{k+1}(\Omega) = \arg\min_{q(\Omega)} \int_{\Omega} \int_{\Theta_\Omega} q^k(\Theta_\Omega) q(\Omega) \times \log\left(\frac{q^k(\Theta_\Omega)q(\Omega)}{\mathrm{F}(\Theta^k_\Omega, \Omega, \mathbf{v}^k, \mathbf{y})}\right) d\Theta_\Omega d\Omega \quad (21)$$

Now we proceed to state the solutions at each step of the algorithm (Eqs. (18)-(21)) explicitly. In estimating $q(\mathbf{x})$ and $q(h)$ we assume that the hyperparameters Ω are known. From Eq. (17) it is clear that $q^k(\mathbf{x})$ is an N-dimensional Gaussian distribution, rewritten as,

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

The covariance and mean of this normal distribution can be calculated from Eq. (18) as

$$\mathrm{cov}_{q^k(\mathbf{x})}[\mathbf{x}] = \left(\beta^k \mathrm{E}^k(\mathbf{H})^t E^k(\mathbf{H}) + \beta^k cov^k(h) + \alpha^k_{\rm im}(\Delta^h)^t W(\mathbf{v}^k)(\Delta^h) + \alpha^k_{\rm im}(\Delta^v)^t W(\mathbf{v}^k)(\Delta^v)\right)^{-1}, \quad (22)$$

$$\mathrm{E}_{q^k(\mathbf{x})}[\mathbf{x}] = \mathrm{cov}_{q^k(\mathbf{x})}[\mathbf{x}]\beta^k E^k(H)^t \mathbf{y}, \quad (23)$$

where $W(\mathbf{v})$ is the $N \times N$ diagonal matrix of the form

$$W(\mathbf{v}) = diag\left(\frac{1}{\sqrt{v^k_i}}\right), \; i = 1,\ldots,N \quad (24)$$

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

$$q^{k+1}(h) = \mathcal{N}\left(h \mid E^{k+1}(h), cov^{k+1}(h)\right), \quad (25)$$

with

$$cov^{k+1}(h) = \left(\alpha^k_{\rm bl} C^t C + \beta^k \mathrm{E}_{q^k(\mathbf{x})}[\mathbf{x}]^t \mathrm{E}_{q^k(\mathbf{x})}[\mathbf{x}] + \beta^k \mathrm{cov}_{q^k(\mathbf{x})}[\mathbf{x}]\right)^{-1}, \quad (26)$$

and

$$E^{k+1}(h) = cov^{k+1}(h)\beta^k \mathrm{E}_{q^k(\mathbf{x})}[\mathbf{x}]^t \mathbf{y}, \quad (27)$$

Next we find $\mathbf{v}^{k+1}$ at step 4 of the algorithm as

$$\mathbf{v}^{k+1}_i = \mathrm{E}_{q^k(\mathbf{x})}[(\Delta^h_i(\mathbf{x}))^2 + (\Delta^v_i(\mathbf{x}))^2], \; i = 1,\ldots,N. \quad (28)$$

After finding estimates of the posterior distributions of the image and blur, we find the estimates for the hyperpriors at the last step of the algorithm. For $\omega \in \{\alpha_{\rm im}, \alpha_{\rm bl}, \beta\}$, evaluating Eq. (21) using Eq. (17) results in

$$q^{k+1}(\omega) \propto \exp \mathrm{E}_{q^k(\mathbf{x})q^{k+1}(h)q(\Omega_\omega)}[\log \mathrm{F}(\Omega^k_\omega, \omega, \mathbf{x}^k, h^{k+1}, \mathbf{v}^{k+1}, \mathbf{y})].$$

Evaluating this explicitly we obtain

$$\begin{aligned} E\left[\log \mathrm{F}(\Theta)\right]_{q^k(\mathbf{x})q^{k+1}(h)} &= \mathrm{const} \\ &+ \sum_{\omega \in \{\alpha_{\rm im}, \alpha_{\rm bl}, \beta\}} ((a^o_\omega - 1)\log\omega - \omega/b^o_\omega) \\ &+ \frac{N}{2}\log\alpha_{\rm im} + \frac{M}{2}\log\alpha_{\rm bl} + \frac{N}{2}\log\beta \\ &- \frac{1}{2}\alpha_{\rm im} E\left[\sum_i \frac{(\Delta^h_i(\mathbf{x}))^2 + (\Delta^v_i(\mathbf{x}))^2 + v_i}{\sqrt{v_i}}\right]_{q^k(\mathbf{x})} \\ &- \frac{1}{2}\alpha_{\rm bl} E\left[\| Ch \|^2\right]_{q^{k+1}(h)} - \frac{1}{2}\beta E\left[\| \mathbf{y} - \mathbf{H}\mathbf{x} \|^2\right]_{q^k(\mathbf{x})q^{k+1}(h)}, \end{aligned} \quad (29)$$

where

$$E\left[\sum_i \frac{(\Delta^h_i(\mathbf{x}))^2 + (\Delta^v_i(\mathbf{x}))^2 + v_i}{\sqrt{v_i}}\right]_{q^k(\mathbf{x})} = 2\sum_i \sqrt{v^{k+1}_i},$$

$$E\left[\| Ch \|^2\right]_{q^{k+1}(h)} = \| CE^{k+1}(h) \|^2 + \mathrm{trace}(C^t C cov^k(h)),$$

and

$$\begin{aligned} E\left[\| \mathbf{y} - \mathbf{H}\mathbf{x} \|^2\right]_{q^k(\mathbf{x})q^{k+1}(h)} &= \| \mathbf{y} - E^{k+1}(h)E^k(\mathbf{x}) \|^2 \\ &+ \mathrm{trace}(cov^k(\mathbf{x})cov^{k+1}(h)) + \mathrm{trace}(E^k(\mathbf{x})^t E^k(\mathbf{x}) cov^{k+1}(h)) \\ &+ \mathrm{trace}(E^{k+1}(H)^t E^{k+1}(H) cov^k(\mathbf{x})). \end{aligned}$$

It can be seen from Eq. (29) that all hyperparameters have gamma distributions, given by

$$q^{k+1}(\alpha_{\rm im}) \propto \alpha_{\rm im}^{N/2 + a^o_{\alpha_{\rm im}} - 1} \times \exp\left[-\alpha_{\rm im}(1/b^o_{\alpha_{\rm im}} + \sum_i \sqrt{v^{k+1}_i})\right],$$

$$\mathrm{q}^{k+1}(\alpha_{\mathrm{bl}}) \propto \alpha_{\mathrm{bl}}^{M/2+a^o_{\alpha_{\mathrm{bl}}}-1} \times \exp\left[-\alpha_{\mathrm{bl}}(1/b^o_{\alpha_{\mathrm{bl}}} + \frac{E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}(h)}}{2})\right],$$

$$\mathrm{q}^{k+1}(\beta) \propto \beta^{N/2+a^o_{\beta}-1} \times \exp\left[-\beta(1/b^o_{\beta} + \frac{E\left[\| \mathbf{y}-\mathbf{Hx} \|^2\right]_{\mathrm{q}^k(\mathbf{x})\mathrm{q}^{k+1}(h)}}{2})\right],$$

where the shape and scale parameters a^{k+1}_{ω} and b^{k+1}_{ω} of the gamma distributions are given by (see Eq. (7))

$$a^{k+1}_{\alpha_{\mathrm{im}}} = a^o_{\alpha_{\mathrm{im}}} + \frac{N}{2}, \quad (30)$$

$$(b^{k+1}_{\alpha_{\mathrm{im}}})^{-1} = \frac{1}{b^o_{\alpha_{\mathrm{im}}}} + \sum_i \sqrt{v_i^{k+1}}, \quad (31)$$

$$a^{k+1}_{\alpha_{\mathrm{bl}}} = a^o_{\alpha_{\mathrm{bl}}} + \frac{M}{2}, \quad (32)$$

$$(b^{k+1}_{\alpha_{\mathrm{bl}}})^{-1} = \frac{1}{b^o_{\alpha_{\mathrm{bl}}}} + \frac{E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}(h)}}{2}, \quad (33)$$

$$a^{k+1}_{\beta} = a^o_{\beta} + \frac{N}{2}, \quad (34)$$

$$(b^{k+1}_{\beta})^{-1} = \frac{1}{b^o_{\beta}} + \frac{E\left[\| \mathbf{y}-\mathbf{Hx} \|^2\right]_{\mathrm{q}^k(\mathbf{x})\mathrm{q}^{k+1}(h)}}{2}. \quad (35)$$

The means of these gamma distributions can be found using Eq. (8). However, we choose to represent them as follows

$$(E[\alpha_{\mathrm{im}}]_{q^{k+1}(\alpha_{\mathrm{im}})})^{-1} = \gamma_{\alpha_{\mathrm{im}}} \frac{1}{\overline{\alpha}^o_{\mathrm{im}}} + (1-\gamma_{\alpha_{\mathrm{im}}}) \frac{\sum_i \sqrt{v_i^{k+1}}}{N/2}, \quad (36)$$

$$E[\alpha_{\mathrm{bl}}]^{-1}_{q^{k+1}(\alpha_{\mathrm{bl}})} = \gamma_{\alpha_{\mathrm{bl}}} \frac{1}{\overline{\alpha}^o_{\mathrm{bl}}} + (1-\gamma_{\alpha_{\mathrm{bl}}}) \frac{E\left[\| Ch \|^2\right]_{\mathrm{q}^{k+1}(h)}}{M}, \quad (37)$$

$$(E[\beta]^{-1}_{q^{k+1}(\beta)} = \gamma_{\beta} \frac{1}{\overline{\beta}^o} + (1-\gamma_{\beta}) \frac{E\left[\| \mathbf{y}-\mathbf{Hx} \|^2\right]_{\mathrm{q}^k(\mathbf{x})\mathrm{q}^{k+1}(h)}}{N}, \quad (38)$$

where $\overline{\alpha}^o_{\mathrm{im}} = a^o_{\alpha_{\mathrm{im}}}/b^o_{\alpha_{\mathrm{im}}}$, $\overline{\alpha}_{\mathrm{bl}} = a^o_{\alpha_{\mathrm{bl}}}/b^o_{\alpha_{\mathrm{bl}}}$ and $\overline{\beta}^o = a^o_{\beta}/b^o_{\beta}$ and

$$\gamma_{\alpha_{\mathrm{im}}} = \frac{a^o_{\alpha_{\mathrm{im}}}}{a^o_{\alpha_{\mathrm{im}}} + \frac{N}{2}}, \quad \gamma_{\alpha_{\mathrm{bl}}} = \frac{a^o_{\alpha_{\mathrm{bl}}}}{a^o_{\alpha_{\mathrm{bl}}} + \frac{M}{2}}, \quad \gamma_{\beta} = \frac{a^o_{\beta}}{a^o_{\beta} + \frac{N}{2}}.$$

The parameters $\gamma_{\alpha_{\mathrm{im}}}, \gamma_{\alpha_{\mathrm{bl}}}$, and γ_{β} can be understood as normalized confidence parameters, taking values in the interval $(0,1)$. When they are asymptotically equal to zero no confidence is placed on the initial values of the hyperparameters, whereas a value asymptotically equal to one will result in no update on the hyperparameters, so that the algorithm will fully rely on the given initial parameters.

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

As a final remark, we would like to note that the estimate of the image in Eq. (23) is computed iteratively, with the use of a conjugate gradient or gradient descent method. However, $\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]$ is explicitly needed in the estimation of the blur (Eq. (26)) and the hyperparameters. We propose the approximation $W(\mathbf{v}^k) \approx z(\mathbf{v}^k)\mathbf{I}$, where

$$z(\mathbf{v}^k) = \frac{1}{N} \sum_i \frac{1}{\sqrt{v_i^k}}. \quad (39)$$

We can therefore obtain a form of $\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]$ that can be represented by a block circulant matrix with circulant blocks (BCCB), whose inverse can be computed in the Fourier domain.

4. EXPERIMENTAL RESULTS

A number of experiments have been performed with the proposed methods. We will denote algorithm 1 as *TV1*, and the second algorithm, where the distributions $\mathrm{q}(\mathbf{x})$ and $\mathrm{q}(h)$ are both degenerate, as *TV2*. We present two sets of experiments with different selections of the confidence parameters $\gamma_{\alpha_{\mathrm{im}}}$, $\gamma_{\alpha_{\mathrm{bl}}}$, γ_{β}.

For our experiments, the image "Lena" is blurred with a Gaussian-shaped function and white Gaussian noise is added to obtain degraded images with blurred-signal-to-noise ratios (BSNR) of 20, 30 and 40dB. For comparison, we include the results from the non-blind versions of our algorithms, where the blur function is known and only the image and the hyperparameters are estimated during iterations. These algorithms will be denoted as *TV1-NB* and *TV2-NB*. We also compare our algorithms to another blind deconvolution method based on variational approximations, which uses SAR-models for both the image and the blur (see [6] for details). We will denote these algorithms as *SAR1* and *SAR2*.

The initial values for the *TV1* and *TV2* algorithms are chosen as follows: The observed image is used as the initial estimate of $\mathrm{E}_{\mathrm{q}^1(\mathbf{x})}[\mathbf{x}]$, and as for the initial estimate of $\mathrm{E}_{\mathrm{q}^1(h)}[h]$ we chose a Gaussian function with variance 4. The covariance matrices $\mathrm{cov}^1(h)$ and $\mathrm{cov}^1(\mathbf{x})$ are set equal to zero. The initial values $E^1[\beta]$, $E^1[\alpha_{\mathrm{im}}]$, and $E^1[\alpha_{\mathrm{bl}}]$ are calculated according to Eqs. (36)–(38), assuming degenerate distributions. Note that, except for the initial values of the image and blur, all parameters are automatically estimated from the observed image. For the *SAR1* and *SAR2* algorithms, the same initial blur is used, and other parameters are found also automatically from the observed image [6].

In the first set of experiments, we set all confidence parameters equal to zero, i.e, the observation is made fully responsible for the estimation process. The quantitative results are shown in Table 1, where ISNR is defined as $10\log_{10}(\| x-y \|^2 / \| x-\hat{x} \|^2)$, where x, y and $\hat{x}$ are the original, observed, and estimated images, respectively. As expected, the *TV1-NB* and *TV2-NB* algorithms result in higher ISNR values since the blur is assumed to be known. The proposed algorithms result in higher ISNR values than *SAR1* and *SAR2*,

Table 1: ISNR values, and the number of iterations obtained by the proposed algorithms compared with other methods.

	BSNR = 20dB		BSNR = 30dB		BSNR = 40dB	
Method	ISNR (dB)	iterations	ISNR (dB)	iterations	ISNR (dB)	iterations
TV1	1.58	26	1.9	20	1.99	9
TV2	-7.16	17	1.55	18	1.95	7
SAR1	0.88	22	1.44	8	1.29	9
SAR2	-8.36	15	-0.16	19	1.13	8
TV1-NB	2.61	30	3.72	17	4.47	20
TV2-NB	2.61	30	3.71	17	4.45	23

Table 2: Experimental results with different confidence parameters using $\overline{\alpha}^o_{\rm im} = 0.06$, $\overline{\alpha}^o_{\rm bl} = 5\times 10^6$, $\overline{\beta}^o = 1/16$ for the BSNR = 40dB case.

TV1

$\gamma_{\alpha_{\rm im}}$	$\gamma_{\alpha_{\rm bl}}$	γ_β	$E[\alpha_{\rm im}]$	$E[\alpha_{\rm bl}]$	$E[\beta]$	ISNR (dB)
0.0	0.0	0.0	0.09	3.2×10^{10}	1/21.48	1.99
1.0	0.8	0.0	0.06	3.9×10^{9}	1/21.44	2.11
1.0	0.8	1.0	0.06	3.5×10^{9}	1/16	2.62

TV2

$\gamma_{\alpha_{\rm im}}$	$\gamma_{\alpha_{\rm bl}}$	γ_β	$E[\alpha_{\rm im}]$	$E[\alpha_{\rm bl}]$	$E[\beta]$	ISNR (dB)
0.0	0.0	0.0	0.06	3.9×10^{9}	1/26.1	1.95
1.0	0.8	0.0	0.06	4.9×10^{9}	1/16.73	2.13
1.0	0.8	1.0	0.06	4.4×10^{9}	1/16	2.65

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

Figure 1: (a) Degraded Lena image (BSNR=40dB), (b) Restored image using *SAR1* (ISNR = 1.29dB), (c) Restored image using *TV1* (ISNR = 1.99dB), (d) Restored image using *TV2* (ISNR = 1.95dB).

and the blur is better removed as shown in Fig. (1), although the reconstructed images have more ringing artifacts.

We examine the effect of prior information on the performance of the proposed algorithms in the second set of experiments. The results with different confidence parameters are summarized in Table 2 for the BSNR = 40dB case. As the results indicate, if some information on the hyperparameters is available, biasing the algorithm towards these hyperparameters leads to improved ISNR values.

5. CONCLUSIONS

A novel total variation based blind deconvolution methodology has been proposed which simultaneously estimates the reconstructed image, the blur, and the hyperparameters of the Bayesian formulation. We have adopted a variational approach to approximate the posterior distributions of the unknown parameters, so that the uncertainty of the estimates can be evaluated and different values from these distributions can be used in the restoration process. Two algorithms are provided resulting from this approach. We have shown that the unknown parameters of the Bayesian formulation can be calculated automatically using only the observation or with different confidence values to improve the performance of the algorithms. Experimental results demonstrated the improved performance of the proposed algorithms.

REFERENCES

[1] D. Kundur and D. Hatzinakos, “Blind image deconvolution,” *IEEE Signal Processing Mag.*, vol. 13, no. 3, pp. 43–64, 1996.

[2] S. Kullback, *Information Theory and Statistics*, New York, Dover Publications, 1959.

[3] 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. Springer-Verlag, July 2000.

[4] K. Z. Adami, “Variational methods in Bayesian deconvolution,” *PHYSTAT2003 ECONF*, vol. C030908, pp. TUGT002, 2003.

[5] A. C. Likas and N. P. Galatsanos, “A variational approach for Bayesian blind image deconvolution,” *IEEE Trans. on Sig. Proc.*, vol. 52, no. 8, pp. 2222–2233, Sept. 2004.

[6] R. Molina, J. Mateos, and A.K. Katsaggelos, “Blind deconvolution using a variational approach to parameter, image, and blur estimation,” *IEEE Trans. on Img Proc.*, vol. 15, no. 12, pp. 3715–3727, Dec. 2006.

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

[8] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, “Total-variation image deconvolution: A majorization-minimization approach,” in *Proc. ICASSP’2006*, 2006.

[9] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Annals of Math. Stat.*, vol. 22, pp. 79–86, 1951.

[10] J. Miskin, *Ensemble Learning for Independent Component Analysis*, Ph.D. thesis, Astrophysics Group, University of Cambridge, 2000.