TOTAL VARIATION IMAGE RESTORATION AND PARAMETER ESTIMATION USING VARIATIONAL POSTERIOR DISTRIBUTION APPROXIMATION

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 image restoration and parameter estimation utilizing variational distribution approximations. By following the hierarchical Bayesian framework, we simultaneously estimate the reconstructed image and the unknown hyperparameters for both the image prior and the image degradation noise. Our algorithms provide an approximation to the posterior distributions of the unknowns so that both the uncertainty of the estimates can be measured and different values from these distributions can be used for the estimates. We also show that some of the current approaches to TV-based image restoration are special cases of our variational framework. Experimental results show that the proposed approaches provide competitive performance without any assumptions about unknown hyperparameters and clearly outperform existing methods when additional information is included.

Index Terms— Image restoration, total variation, variational methods, parameter estimation, Bayesian methods.

1. INTRODUCTION

The image degradation process is often represented by a linear model 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 matrix $\mathbf{H}$ represents the blurring matrix, which is assumed to be known. It is also assumed that $\mathbf{n}$ is sampled from a zero-mean independent Gaussian random process with variance β^{-1}.

The image restoration problem is to find an estimate of $\mathbf{x}$ from $\mathbf{y}$ and $\mathbf{H}$ using prior knowledge about $\mathbf{n}$ and $\mathbf{x}$. The literature on image restoration is rich (a review and classification of the major approaches can be found for example in [1]).

Methods based on Bayesian formulations are of the most commonly used methods in the image restoration literature. Such methods introduce an image model on $\mathbf{x}$ which is used to incorporate prior knowledge and to impose constraints on the image estimate, acting as a regularizer in the estimation process. A number of different prior models are introduced, such as, Simultaneous Autoregression (SAR), Conditional Autoregression (CAR), or Total Variation (TV). The prior models typically depend on a parameter which is related to the global variance of the image. In this paper, we utilize a TV function as the image prior in the estimation process.

Additionally, the observation model involves a parameter which is related to the variance of the degradation noise. These parameters, also called *hyperparameters*, determine the performance of the algorithm to a great extent, and have to be determined carefully, or estimated by the algorithm. To our knowledge, not much work has been reported on the simultaneous estimation of the hyperparameters and the image in TV-based image restoration. Rudin *et al.* [2] consider the constrained minimization of the image prior and then proceed to estimate both the image and the Lagrange multiplier associated with the prior. Bertalmio *el al.* [3] make the Lagrange multiplier region dependent. Bioucas-Dias *et al.* [4] using their majorization-minimization approach [5] propose a Bayesian method to estimate the original image and the hyperparameter of the image model assuming that an estimate of the noise variance is available.

In this paper we adopt a hierarchical Bayesian formulation (see, for example, [6, 7]) to jointly estimate the image and the hyperpriors using a TV image prior. A variational approach is utilized to derive the distributions of both the image and the hyperparameters, which allows us to derive different estimations for the unknown variables and also to analyze the uncertainty of those estimates.

This paper is organized as follows. In Sec. 2 we present the hierarchical Bayesian model. Section 3 describes the variational inference methods and the derivation of the proposed methods. We present the experimental results in Sec. 4 and conclude in Sec. 5.

2. BAYESIAN MODELING

Utilizing a hierarchical Bayesian model, the image and the observation noise are modeled in the first stage using some unknown hyperparameters, and the hyperprior distributions of the hyperparameters are modeled in the second stage.

2.1. First stage: prior models on image and observation

Given the observation model in Eq. (1), the corresponding probability distribution can be stated as

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

We use the TV prior as the image model, given by

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

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

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

where the operators $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ correspond to, respectively, horizontal and vertical first order differences, at pixel i, that is, $\Delta_i^h(\mathbf{x}) =$

$x_i - x_{l(i)}$ and $\Delta_i^v(\mathbf{x}) = x_i - x_{a(i)}$, where $l(i)$ and $a(i)$ denote the nearest neighbors of i, to the left and above, respectively. We can approximate the partition function $Z_{\mathrm{TV}}(\alpha)$ using

$$\int_u \int_v \exp\left[-\alpha\sqrt{u^2+v^2}\right] du dv = 2\pi/\alpha^2, \tag{5}$$

as proposed in [5], to obtain

$$\mathrm{p}(\mathbf{x}) \propto \alpha^{N/2} \exp\left[-\alpha \mathrm{TV}(\mathbf{x})\right], \tag{6}$$

where N is the size of the original image $\mathbf{x}$.

2.2. Second stage: hyperprior on the hyperparameters

In this work we use flat improper hyperpriors on α and β, that is, we utilize

$$\mathrm{p}(\alpha) \propto \text{const}, \qquad \mathrm{p}(\beta) \propto \text{const}. \tag{7}$$

Note that with this choice of the hyperpriors the observation $\mathbf{y}$ is fully responsible for the estimation of the image and parameters α and β.

Finally, combining the first and second stage of the problem modeling we propose the following global distribution

$$\begin{aligned}&\mathrm{p}(\alpha,\beta,\mathbf{x},\mathbf{y}) = \mathrm{p}(\alpha)\mathrm{p}(\beta)\mathrm{p}(\mathbf{x}|\alpha)\mathrm{p}(\mathbf{y}|\mathbf{x},\beta)\\ &\propto \alpha^{N/2}\beta^{N/2}\exp\left[-\alpha\mathrm{TV}(\mathbf{x})\right]\exp\left[-\frac{\beta}{2}\parallel \mathbf{y}-\mathbf{H}\mathbf{x}\parallel^2\right].\end{aligned} \tag{8}$$

3. BAYESIAN INFERENCE AND VARIATIONAL APPROXIMATION

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

$$\mathrm{p}(\alpha,\beta,\mathbf{x}\mid\mathbf{y}) = \frac{\mathrm{p}(\alpha,\beta,\mathbf{x},\mathbf{y})}{\mathrm{p}(\mathbf{y})}. \tag{9}$$

Since the distribution $\mathrm{p}(\alpha,\beta,\mathbf{x}\mid\mathbf{y})$ is difficult to evaluate, we apply variational methods to approximate it by $\mathrm{q}(\alpha,\beta,\mathbf{x}) = \mathrm{q}(\alpha,\beta)\mathrm{q}(\mathbf{x})$.

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

$$\begin{aligned}&C_{KL}(\mathrm{q}(\alpha,\beta,\mathbf{x}) \parallel \mathrm{p}(\alpha,\beta,\mathbf{x}|\mathbf{y}))\\ &= \int_\alpha\int_\beta\int_\mathbf{x} \mathrm{q}(\alpha,\beta,\mathbf{x})\log\left(\frac{\mathrm{q}(\alpha,\beta,\mathbf{x})}{\mathrm{p}(\alpha,\beta,\mathbf{x}|\mathbf{y})}\right) d\alpha d\beta d\mathbf{x}\\ &= \int_\alpha\int_\beta\int_\mathbf{x} \mathrm{q}(\alpha,\beta,\mathbf{x})\log\left(\frac{\mathrm{q}(\alpha,\beta,\mathbf{x})}{\mathrm{p}(\alpha,\beta,\mathbf{x},\mathbf{y})}\right) d\alpha d\beta d\mathbf{x} + \text{const},\end{aligned} \tag{10}$$

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

Due to the form of the TV prior, it is difficult to evaluate the integral in Eq. (10). We therefore utilize a minorization of the TV prior. Let us define, for α, $\mathbf{x}$ and any $N-$dimensional vector $\mathbf{v} \in (R^+)^N$, with components $v_i,\ i = 1,\ldots,N$, the following functional

$$\mathrm{M}(\alpha,\mathbf{x},\mathbf{v}) = \exp\left[-\frac{\alpha}{2}\sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + v_i}{\sqrt{v_i}}\right] \tag{11}$$

Using the inequality in [5], for $u \geq 0$ and $v > 0$

$$\sqrt{u} \leq \sqrt{v} + \frac{1}{2\sqrt{v}}(u-v), \tag{12}$$

we have that

$$\mathrm{p}(\mathbf{x}|\alpha) \geq \text{const}\cdot\ \alpha^{N/2}\cdot\exp\mathrm{M}(\alpha,\mathbf{x},\mathbf{v}), \tag{13}$$

which leads to the following lower bound of the joint probability distribution

$$\begin{aligned}\mathrm{p}(\alpha,\beta,\mathbf{x},\mathbf{y}) &\geq \mathrm{p}(\alpha)\mathrm{p}(\beta)\mathrm{M}(\alpha,\mathbf{x},\mathbf{v})\mathrm{p}(\mathbf{y}|\mathbf{x},\beta) &(14)\\ &= \mathrm{F}(\alpha,\beta,\mathbf{x},\mathbf{v},\mathbf{y}). &(15)\end{aligned}$$

Utilizing this lower bound, we obtain

$$\begin{aligned}&\int_\alpha\int_\beta\int_\mathbf{x} \mathrm{q}(\alpha,\beta,\mathbf{x})\log\left(\frac{\mathrm{q}(\alpha,\beta,\mathbf{x})}{\mathrm{p}(\alpha,\beta,\mathbf{x}|\mathbf{y})}\right) d\alpha d\beta d\mathbf{x}\\ &\leq \int_\alpha\int_\beta\int_\mathbf{x} \mathrm{q}(\alpha,\beta,\mathbf{x})\log\left(\frac{\mathrm{q}(\alpha,\beta,\mathbf{x})}{\mathrm{F}(\alpha,\beta,\mathbf{x},\mathbf{v},\mathbf{y})}\right) d\alpha d\beta d\mathbf{x}\end{aligned} \tag{16}$$

The right-hand side of Eq. (16) is easier to evaluate than the left-hand side. It can therefore be used in the following algorithm for evaluating the approximating posteriors $\mathrm{q}(\mathbf{x})$ and $\mathrm{q}(\alpha,\beta)$.

Algorithm 1 *Posterior parameter and image distributions estimation in TV restoration using* $q(\alpha,\beta,\mathbf{x}) = q(\alpha,\beta)q(\mathbf{x})$.
Given $q^1(\alpha,\beta)$, *an initial estimate of the distribution* $q(\alpha,\beta)$, *and* $\mathbf{v}^1 \in (R^+)^N$,
For $k = 1,2,\ldots$ *until a stopping criterion is met:*

1. *Find*

$$\begin{aligned}q^k(\mathbf{x}) = \arg\min_{q(\mathbf{x})}&\int_\mathbf{x}\int_\alpha\int_\beta q^k(\alpha,\beta)q(\mathbf{x})\\ &\times\log\left(\frac{q^k(\alpha,\beta)q(\mathbf{x})}{\mathrm{F}(\alpha,\beta,\mathbf{x},\mathbf{v}^k,\mathbf{y})}\right) d\alpha d\beta d\mathbf{x}\end{aligned} \tag{17}$$

2. *Find*

$$\begin{aligned}\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}}&\int_\alpha\int_\beta\int_\mathbf{x} q^k(\alpha,\beta)q^k(\mathbf{x})\\ &\times\log\left(\frac{q^k(\alpha,\beta)q^k(\mathbf{x})}{\mathrm{F}(\alpha,\beta,\mathbf{x},\mathbf{v},\mathbf{y})}\right) d\alpha d\beta d\mathbf{x}\end{aligned} \tag{18}$$

3. *Find*

$$\begin{aligned}q^{k+1}(\alpha,\beta) = \arg\min_{q(\alpha,\beta)}&\int_\alpha\int_\beta\int_\mathbf{x} q(\alpha,\beta)q^k(\mathbf{x})\\ &\times\log\left(\frac{q(\alpha,\beta)q^k(\mathbf{x})}{\mathrm{F}(\alpha,\beta,\mathbf{x},\mathbf{v}^{k+1},\mathbf{y})}\right) d\alpha d\beta d\mathbf{x}\end{aligned} \tag{19}$$

Set

$$q(\alpha,\beta) = \lim_{k\to\infty} q^k(\alpha,\beta),\quad q(\mathbf{x}) = \lim_{k\to\infty} q^k(\mathbf{x}). \tag{20}$$

In order to find $\mathrm{q}(\mathbf{x})$, we differentiate the integral on the right-hand side of Eq. (17) with respect to $\mathrm{q}(\mathbf{x})$ and set it equal to zero to obtain

$$\mathrm{q}^k(\mathbf{x}) \propto \exp\mathrm{E}_{\mathbf{q}^k(\alpha,\beta)}[\ln\mathrm{F}(\alpha,\beta,\mathbf{x},\mathbf{v}^k)]. \tag{21}$$

Therefore $\mathrm{q}^k(\mathbf{x})$ is an N-dimensional Gaussian distribution with parameters

$$\begin{aligned}\mathrm{cov}_{\mathbf{q}^k(\mathbf{x})}[\mathbf{x}] &= \big(\mathrm{E}_{\mathbf{q}^k(\beta)}[\beta]\mathbf{H}^t\mathbf{H} + \mathrm{E}_{\mathbf{q}^k(\alpha)}[\alpha](\Delta^h)^t W(\mathbf{v}^k)(\Delta^h)\\ &\quad + \mathrm{E}_{\mathbf{q}^k(\alpha)}[\alpha](\Delta^v)^t W(\mathbf{v}^k)(\Delta^v)\big)^{-1}\\ &= [\mathbf{C}^k(\mathbf{v}^k)]^{-1},\end{aligned} \tag{22}$$

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

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

$$W(\mathbf{v}) = diag\left(\frac{1}{\sqrt{v_i^k}}\right),\ i = 1, \ldots, N \tag{24}$$

Similarly, we have from Eq. (18)

$$\mathbf{v}_i^{k+1} = \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2],\ \ i = 1, \ldots, N. \tag{25}$$

where

$$\begin{aligned}\mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2] &= (\Delta_i^h(\mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]))^2 \\ + ((\Delta_i^v(\mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]))^2 &+ \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[(\Delta_i^h(\mathbf{x} - \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]))^2] \\ &+ \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[(\Delta_i^v(\mathbf{x} - \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]))^2]\end{aligned} \tag{26}$$

Finally we find $\mathrm{q}^{k+1}(\alpha, \beta)$ by differentiating the integral on the right hand side of Eq. (19) with respect to $\mathrm{q}(\alpha, \beta)$ and setting it equal to zero to obtain

$$\mathrm{q}^{k+1}(\alpha, \beta) = \mathrm{q}^{k+1}(\alpha)\mathrm{q}^{k+1}(\beta) \tag{27}$$

where $\mathrm{q}^{k+1}(\alpha)$ is the gamma distribution given by

$$\mathrm{q}^{k+1}(\alpha) \propto \alpha^{N/2} \exp\left[-\alpha \sum_i \sqrt{v_i^{k+1}}\right] \tag{28}$$

and $\mathrm{q}^{k+1}(\beta)$ is the gamma distribution given by

$$\mathrm{q}^{k+1}(\beta) \propto \beta^{N/2} \exp\left[-\beta \frac{\mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2}{2}\right] \tag{29}$$

The mean and mode of these distributions are given by

$$\mathrm{E}_{\mathrm{q}^{k+1}(\alpha)}[\alpha] = \frac{N/2 + 1}{\sum_i \sqrt{v_i^{k+1}}},\ \mathrm{Mode}_{\mathrm{q}^{k+1}(\alpha)}[\alpha] = \frac{N/2}{\sum_i \sqrt{v_i^{k+1}}} \tag{30}$$

and

$$\mathrm{E}_{\mathrm{q}^{k+1}(\beta)}[\beta] = \frac{N + 2}{\mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2}, \tag{31a}$$

$$\mathrm{Mode}_{\mathrm{q}^{k+1}(\beta)}[\beta] = \frac{N}{\mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2}. \tag{31b}$$

Note that $\mathrm{E}_{\mathrm{q}^k(\mathbf{x})} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2$ in Eqs. (29)-(31b) can be stated as

$$\begin{aligned}\mathrm{E}_{\mathrm{q}^k(\mathbf{x})}\left[\parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2\right] =& \parallel \mathbf{y} - \mathbf{H}\mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}] \parallel^2 \\ &+ \mathrm{trace}\left(\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]\mathbf{H}^t\mathbf{H}\right).\end{aligned} \tag{32}$$

In this proposed algorithm for estimating the posterior distribution of the image and the unknown parameters, no assumptions were made about the posterior approximation. Alternatively, we can assume $\mathrm{q}(\mathbf{x})$ being a degenerate distribution, that is, a distribution which takes one value with probability one and the rest with probability zero. We propose next an algorithm where $\mathrm{q}^k(\mathbf{x})$ takes the value $\mathbf{x}^k$ with probability one.

Algorithm 2 *Posterior parameter and image distributions estimation in TV restoration using $q(\alpha, \beta, \mathbf{x}) = q(\alpha, \beta)q(\mathbf{x})$ where $q(\mathbf{x})$ is a degenerate distribution.*

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

1. *Calculate*
$$\mathbf{x}^k = [\mathbf{C}^k(\mathbf{v}^k)]^{-1}\mathbf{H}^t\mathbf{y}, \tag{33}$$
2. *Calculate*
$$\mathbf{v}_i^{k+1} = (\Delta_i^h(\mathbf{x}^k))^2 + (\Delta_i^v(\mathbf{x}^k))^2],\ \ i = 1, \ldots, N. \tag{34}$$
3. *Calculate*
$$q^{k+1}(\alpha) \propto \alpha^{N/2} \exp\left[-\alpha \sum_i \sqrt{v_i^{k+1}}\right] \tag{35}$$
$$q^{k+1}(\beta) \propto \beta^{N/2} \exp\left[-\beta \frac{\parallel \mathbf{y} - \mathbf{H}\mathbf{x}^k \parallel^2}{2}\right] \tag{36}$$

The mean and mode of these distributions are given by

$$\mathrm{E}_{\mathrm{q}^{k+1}(\alpha)}[\alpha] = \frac{N/2 + 1}{\sum_i \sqrt{v_i^{k+1}}},\ \mathrm{Mode}_{\mathrm{q}^{k+1}(\alpha)}[\alpha] = \frac{N/2}{\sum_i \sqrt{v_i^{k+1}}} \tag{37}$$

and

$$\mathrm{E}_{\mathrm{q}^{k+1}(\beta)}[\beta] = \frac{N + 2}{\parallel \mathbf{y} - \mathbf{H}\mathbf{x}^k \parallel^2}, \tag{38a}$$

$$\mathrm{Mode}_{\mathrm{q}^{k+1}(\beta)}[\beta] = \frac{N}{\parallel \mathbf{y} - \mathbf{H}\mathbf{x}^k \parallel^2}. \tag{38b}$$

The estimates of the image in Eqs. (23) and (33) are computed iteratively, with the use of a conjugate gradient or gradient descent algorithm thus avoiding the inversion of $\mathbf{C}^k(\mathbf{v}^k)$. However, $\mathrm{cov}_{\mathrm{q}^k(\mathbf{x})}[\mathbf{x}]$ is explicitly needed to evaluate Eq. (32). We propose the approximation $W(\mathbf{v}^k) \approx z(\mathbf{v}^k)\mathbf{I}$, where

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

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

In this section we present experimental results obtained by the use of the proposed algorithms on two images with two different degradation functions, i.e., a Gaussian blur with variance 9 and a uniform blur of size 9x9. In all cases, white Gaussian noise is added to the blurred images to obtain degraded images with blurred-signal-to-noise (BSNR) ratios of 20, 30, and 40dB. We use "Lena" and "Cameraman" as our test images.

Table 1 shows the quantitative results corresponding to the experiments, where ISNR is defined as $10\log_{10}(\parallel x - y \parallel^2 / \parallel x - \hat{x} \parallel^2)$, where x, y and $\hat{x}$ are the original, observed, and estimated images, respectively. Algorithms 1 and 2 are denoted by *ALG1* and *ALG2*, respectively, and *BFO1* represents the method in [5] while *BFO2* represents the method in [4]. Note that both of these methods are based on TV-priors, where *BFO1* uses a hand-tuned empirical value for the hyperparameters and *BFO2* incorporates an adaptive scheme to estimate the hyperparameter α. Both algorithms assume that the noise variance is available. It is important to note that both

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

Fig. 1. (a) Image degraded by a Gaussian shaped PSF with variance 9 and Gaussian noise of variance 0.16 (BSNR=40dB), (b) Restored image using *ALG1* (ISNR = 4.84dB), (c) Restored image using *ALG2* (ISNR = 4.64dB), (d) Restored image using *ALG1-True* (ISNR = 6.95dB).

BFO1 and *BFO2* can be derived as special cases of our formulation. The *BFO1* method can be obtained by assuming both hyperparameters are known, and *BFO2* can be obtained by assuming β is known and by replacing $\mathrm{E}_{\mathrm{q}^k(\alpha)}[\alpha]$ by $\mathrm{Mode}_{\mathrm{q}^k(\alpha)}[\alpha]$ in Eq. (33).

It is clear from Table 1 that *ALG1* and *ALG2* both almost always perform better than *BFO2* although the latter assumes a known noise variance whereas the proposed algorithms estimate this parameter simultaneously with the image. The performance of the algorithms is better than even *BFO1* in Gaussian blur case. It is also clear that both *ALG1* and *ALG2* perform better at restoring images blurred with Gaussian PSFs.

The performance difference between *ALG1* and *ALG2* is very small, and they give slightly lower values for ISNR than *BFO1*. This is however expected, since *BFO1* assumes knowledge of the noise variance and picks a hand-tuned parameter α. To evaluate the performance of the proposed methods in this case, we calculated the hyperparameters α and β using $\alpha = \frac{N/2+1}{\sum_i \sqrt{v_i^{\mathrm{org}}}}$ and $\beta = \frac{N+2}{\|\mathbf{y}-\mathbf{H}\mathbf{x}^{\mathrm{org}}\|^2}$, where v_i^{org} is obtained from the original image $\mathbf{x}^{\mathrm{org}}$, and run our algorithms by keeping these parameters fixed as the image is estimated. The corresponding ISNR values are shown in Table 1, denoted by *ALG1-True* and *ALG2-True*. It is obvious that in this case the proposed algorithms provide a superior performance over the rest of the algorithms.

The restoration results of the Lena image in the case of Gaussian blur with 40dB BSNR as well as the degraded version are shown in Fig. (1). Considering that the parameters of both algorithms are estimated automatically using the degraded observation without requiring any prior knowledge about the noise, it can be stated that the restored images are of high quality.

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

	Cameraman with a 9x9 uniform blur			Lena with a Gaussian blur of variance 9		
BSNR	Method	ISNR (dB)	iterations	Method	ISNR (dB)	iterations
40dB	*BFO1*	8.55	8	*BFO1*	4.72	20
	BFO2	8.25	12	*BFO2*	4.50	19
	ALG1	6.76	24	*ALG1*	4.84	10
	ALG2	8.29	21	*ALG2*	4.64	16
	ALG1-True	11.34	9	*ALG1-True*	6.95	3
	ALG2-True	11.33	13	*ALG2-True*	6.95	4
30dB	*BFO1*	5.68	10	*BFO1*	3.87	24
	BFO2	4.65	14	*BFO2*	3.56	21
	ALG1	5.41	19	*ALG1*	4.03	16
	ALG2	4.39	17	*ALG2*	3.67	21
	ALG1-True	8.26	5	*ALG1-True*	6.01	2
	ALG2-True	8.26	6	*ALG2-True*	6.01	2
20dB	*BFO1*	3.31	14	*BFO1*	3.02	20
	BFO2	2.12	20	*BFO2*	2.47	22
	ALG1	2.46	22	*ALG1*	3.06	23
	ALG2	2.12	28	*ALG2*	2.58	34
	ALG1-True	5.33	5	*ALG1-True*	5.07	2
	ALG2-True	5.33	6	*ALG2-True*	5.07	2

5. CONCLUSIONS

In this paper we represented a novel TV-based image restoration methodology which simultaneously estimates the reconstructed image and the hyperparameters of the Bayesian formulation. We adopt a variational approximation approach to estimate the posterior distributions of the image and the hyperparameters, so that the uncertainty of these parameters can be evaluated and different values from these distributions can be used in the restoration process to obtain better performance. We have provided two different algorithms that resulted from this formulation, and have shown that their performance is competitive to other TV-based methods which assume knowledge about the image degradation process whereas our algorithms work automatically without any assumptions. We have also shown that if additional information about the image formation and degradation processes is provided, the proposed algorithms provide superior performance over the existing methods.

6. REFERENCES

[1] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Mag.*, vol. 14, no. 2, pp. 24–41, Mar 1997.

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

[3] M. Bertalmio, V. Caselles, B. Rougé, and A. Solé, "Tv based image restoration with local constraints," *Journal of Scientific Computing*, vol. 19, pp. 95–122, 2003.

[4] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Adaptive bayesian/total-variation image deconvolution: A majorization-minimization approach," in *EUSIPCO'2006*, 2006.

[5] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Total-variation image deconvolution: A majorization-minimization approach," in *ICASSP'2006*, 2006.

[6] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, and J. Mateos, "Hyperparameter estimation in image restoration problems with partially-known blurs," *Optical Eng.*, vol. 41, no. 8, pp. 1845–1854, 2002.

[7] 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.