Blind deconvolution of images corrupted by Gaussian noise using Expectation Propagation

Abdullah Abdulaziz, Dan Yao, Yoann Altmann and Stephen McLaughlin

School of Engineering and Physical Sciences, Heriot-Watt University,
Edinburgh EH14 4AS, United Kingdom

***Abstract*—Blind image deconvolution consists of inferring an image from its blurry and noisy version when the blur is unknown. To solve this highly ill-posed inverse problem, Expectation Maximization (EM)-based algorithms can be adopted. In several previous studies, Variational Bayes (VB) approaches were deployed to approximate the intractable conditional probability distribution of the image that appears in the E-step of the traditional EM algorithm. In this paper, we propose to use an Expectation Propagation (EP) algorithm to derive an alternative approximation of the conditional probability distribution. The simulations conducted show that the resulting EP-EM approach can provide more reliable approximations, reflected by better image estimates and more reliable uncertainty maps than VB-EM for a comparable computational time.**

***Index Terms*—image restoration, blind image deconvolution, Expectation-Propagation, approximate Bayesian inference.**

I. Introduction

In many imaging applications such as commercial photography, microscopy and remote sensing, images can be degraded during the observation process by blur, noise, and other degradations due to imperfections in the sensing system. Image deconvolution techniques aim to recover the underlying, usually sharper, image from its blurry and noisy observations. Non-blind image deconvolution algorithms assume that the blur is known or can be estimated prior to the deconvolution process. However, this assumption might not hold in practice or can produce reconstruction artefacts. Conversely, Blind Image Deconvolution (BID) seeks to recover a sharp image without knowing the blur in advance, leading to one of the most challenging problems in the image processing community. The image degradation model for linear shift-invariant blurring effects can be written as [1], [2]

$$\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{h} + \boldsymbol{w} = \mathbf{H}\boldsymbol{x} + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the observed blurry image composed of size N pixels, $\boldsymbol{x} \in \mathbb{R}^N$ is the original sharp image and $\boldsymbol{h} \in \mathbb{R}^k$ is a blur kernel whose support k (*i.e.* the number of pixels affected by the intensity of a given pixel in the original image) is small compared to the image size. The matrix $\mathbf{H}$ represents the block-circulant matrix constructed from the kernel $\boldsymbol{h}$, $*$ is the convolution operator and the observation noise $\boldsymbol{w}$ is usually assumed to be i.i.d. Gaussian noise with known variance σ^2.

Many methods have been proposed to address the BID problem and reviews of the major approaches can be found in [1]–[5]. The work herein adopts an approximate Bayesian approach and an Expectation Maximization (EM)-based algorithm to solve the BID problem. The proposed approach consists of estimating $\boldsymbol{h}$ via Marginal Maximum A Posteriori (MMAP) estimation, and the estimated blur is then used to estimate $\boldsymbol{x}$ via conditional MMSE estimation. This estimation strategy can be implemented efficiently using an EM-based algorithm [6]. The main limitation of the traditional EM is that the conditional probability distribution of the image given the observed data and kernel parameters, required in the E-step, is usually intractable. While Variational Bayes (VB) methods have been widely deployed in the literature to approximate the conditional probability distribution of interest (variational EM), in this paper we propose to use an Expectation Propagation (EP) algorithm to approximate this conditional probability distribution and show that it can provide more reliable marginal variances than VB.

The remainder of the paper is structured as follows. In Section II, we formulate the BID problem as a Bayesian inverse problem and describe several estimation strategies available. The EP algorithm adopted for approximate Bayesian inference is presented in Section III. Simulation and results are presented in Section IV. Conclusions are finally reported in Section V.

II. Bayesian estimation

The BID problem consists of inferring the original image and the blur kernel given the blurred image. In this context, the observation model (1) defines the likelihood

$$\boldsymbol{y} |\ \boldsymbol{x}, \boldsymbol{h} \sim \mathcal{N}(\boldsymbol{y}|\mathbf{H}\boldsymbol{x}, \sigma^2\mathbf{I}). \tag{2}$$

The BID problem is severely ill-posed in the sense that there exist an infinite set of $(\boldsymbol{x}, \boldsymbol{h})$ that can explain the observed data. To discard unwanted solutions, Bayesian methods incorporate prior information about the original

This work was supported by the Royal Academy of Engineering under the Research Fellowship scheme RF201617/16/31 and by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/T00097X/1.

image and the blur as prior distributions, *i.e.* $\mathrm{p}(\boldsymbol{x})$ and $\mathrm{p}(\boldsymbol{h})$. This yields the following joint distribution

$$\mathrm{p}(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{y}) = \mathrm{p}(\boldsymbol{y}|\ \boldsymbol{x}, \boldsymbol{h})\ \mathrm{p}(\boldsymbol{x})\ \mathrm{p}(\boldsymbol{h}). \tag{3}$$

A. Prior models

a) Image prior models: The most common prior models are Gaussian prior models which are used to promote smoothness [7], [8] as they naturally lead to simple inference schemes in the presence of Gaussian noise. Given the fact that most natural images can be sparsely represented in some domain, sparsity-inducing models have also been considered in the BID literature. The authors in [9] proposed to use a mixture of Gaussians model to approximate the distribution of the gradient magnitude and promote sparsity. Other sparsity-promoting priors were also used, such as the Student-t prior [10] and the ℓ_p prior [11]. It is worth noting that a general representation for sparse priors was proposed in [12] using super Gaussian distributions. Another traditional choice for promoting spatial correlation between neighboring pixels are Markov random fields (MRFs). The total-variation (TV), which promotes piece-wise constant images, is a popular example of MRF prior. TV was frequently used in BID in the aim to preserve edges while reducing noise in flat regions [13], [14]. The ℓ_1-norm based TV prior can be defined as

$$\mathrm{p}_1(\boldsymbol{x}) \propto \exp^{-\lambda\ \mathrm{TV}(\boldsymbol{x})}, \tag{4}$$

with $\mathrm{TV}(\boldsymbol{x}) = \sum_{i=1}^{N} |\Delta_i^h \boldsymbol{x}| + |\Delta_i^v \boldsymbol{x}|$, and where $\Delta_i^h \boldsymbol{x}$ and $\Delta_i^v \boldsymbol{x}$ denote the horizontal and vertical first order differences corresponding to the left and top neighbors of pixel i. The hyperparameter $\lambda \geq 0$ controls the amount of prior smoothness.

In this work, the ℓ_1-based TV prior is adopted as the image prior model along with an image non-negativity constraint (if appropriate) to satisfy physical constraints. Using the non-differentiable TV prior makes the computation of the E-step in the EM algorithm intractable. However, we show in the following section that approximate Bayesian methods such as VB and EP can lead to closed-form update. By including the non-negativity constraint, the image prior model used here becomes $\mathrm{p}(\boldsymbol{x}) \propto \mathrm{p}_1(\boldsymbol{x})\ \mathrm{p}_2(\boldsymbol{x})$, with $\mathrm{p}_2(\boldsymbol{x}) \propto \iota_{\mathbb{R}_+^N}(\boldsymbol{x})$ is the indicator function of the non-negative orthant in $\mathbb{R}^N$.

b) Blur model: The BID literature contains many specific blur models. For instance, the authors in [15] proposed a Dirichlet blur model that offers flexibility in incorporating vague or precise knowledge about the blur. To estimate motion blur, the authors in [16] promoted sparsity of the kernel in a curvelet dictionary. In this work, we restrict ourselves to the use of a flat prior for the blur with non-negativity and unit-norm constraints. These constraints have shown to give good estimate of the blur given the fact that usually $k \ll N$ [1], [12]. The resulting prior reduces to $\mathrm{p}(\boldsymbol{h}) \propto \iota_{\mathcal{C}}(\boldsymbol{x})$, with $\mathcal{C} = \{\boldsymbol{h}|\boldsymbol{h} \in \mathbb{R}_+^k \text{and } \mathbf{1}_k^T \boldsymbol{h} = 1\}$, where $\mathbf{1}_k$ stands for the vector of size k with all coefficients equal to 1, and ^{T} stands for the transpose operator.

B. Estimation strategies

Using the joint probability distribution (3), our goal is to infer the unknown variables $(\boldsymbol{x}, \boldsymbol{h})$ given the observations $\boldsymbol{y}$. This can be achieved in different ways. Stochastic sampling methods such as Markov chain Monte Carlo (MCMC) generate a sequence of samples from the joint posterior distribution $\mathrm{p}(\boldsymbol{x}, \boldsymbol{h}|\boldsymbol{y})$ [17]. These samples can then be used to approximate the joint MMSE estimator of $(\boldsymbol{x}, \boldsymbol{h})$. Hyperparameters such as σ^2 can also be included via an augmented posterior distribution [2]. However, such methods usually require many samples to accurately explore the posterior distribution which makes them very intensive for high dimensional problems.

The joint MAP estimator of $(\boldsymbol{x}, \boldsymbol{h})$ can be approximated via alternating optimization, which consists of maximizing the posterior distribution with respect to each unknown while holding the other one fixed. MAP approaches are computationally very efficient but they do not allow uncertainty quantification for assessing the reliability of the estimated solution.

A third approach consists of estimating $\boldsymbol{h}$ via MMAP estimation, i.e., by maximizing $\mathrm{p}(\boldsymbol{h}|\boldsymbol{y}) = \int \mathrm{p}(\boldsymbol{x}, \boldsymbol{h}|\boldsymbol{y})\ \mathrm{d}\boldsymbol{x}$. The resulting estimated blur $\hat{\boldsymbol{h}}_{MMAP}$ can then be used to estimate $\boldsymbol{x}$ via MMSE or MAP estimation using $\mathrm{p}(\boldsymbol{x}|\boldsymbol{y}, \hat{\boldsymbol{h}}_{MMAP})$. This estimation strategy can be implemented using MCMC [18] or EM-based algorithms [6] whose basic principle is recalled below.

The standard EM algorithm consists of two steps, the expectation step (E-step) and the maximization step (M-step), which are repeated until convergence.

- E-step: $F(\boldsymbol{h}, \boldsymbol{h}^{(t-1)}) = \mathrm{E}_{\mathrm{p}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{h}^{(t-1)})}\left[\ln\left(\mathrm{p}(\boldsymbol{x}, \boldsymbol{h}|\boldsymbol{y})\right)\right]$
- M-step: $\boldsymbol{h}^{(t)} = \mathrm{argmax}_{\boldsymbol{h} \in \mathbb{R}^k}\ F(\boldsymbol{h}, \boldsymbol{h}^{(t-1)})$

It has been shown for different imaging problems that estimating the blur kernel first, via MMAP estimation, often delivers better images than joint MMSE. This can be explained by the fact that typically $\mathrm{p}(\boldsymbol{h}|\boldsymbol{y})$ has its maximum at a good value for $\boldsymbol{h}$, while its mass spreads across a wider range of $\boldsymbol{h}$ and hence $\hat{\boldsymbol{h}}_{MMSE}$ can be suboptimal [18]. In many cases however, the standard EM algorithm cannot be applied directly because the expectation involved in the E-step is intractable, e.g., when the likelihood and $\mathrm{p}(\boldsymbol{x})$ are not conjugate. To overcome this shortcoming, VB methods replace the conditional distribution $\mathrm{p}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{h}^{(t-1)})$ by a tractable approximate distribution $\mathrm{q}(\boldsymbol{x})$ (usually Gaussian, denoted by $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}|\boldsymbol{m}_{\boldsymbol{x}}, \mathbf{C}_{\boldsymbol{x}})$), which makes the expectation tractable. The resulting Variational EM (VEM) algorithm leads to the following steps

- E-step: $F(\boldsymbol{h}, \boldsymbol{h}^{(t-1)}) = \mathrm{E}_{\mathrm{q}(\boldsymbol{x})}\left[\ln\left(\mathrm{p}(\boldsymbol{x}, \boldsymbol{h}|\boldsymbol{y})\right)\right]$
- M-step: $$\boldsymbol{h}^{(t)} = \underset{\boldsymbol{h} \in \mathbb{R}^k}{\mathrm{argmin}}\ \|\boldsymbol{y} - \boldsymbol{m}_{\boldsymbol{x}} * \boldsymbol{h}\|^2 + \boldsymbol{h}^\dagger \mathbf{D}_{\boldsymbol{x}} \boldsymbol{h}$$ $$\text{subject to } \begin{cases} \mathbf{1}_k^T \boldsymbol{h} = 1, \\ \boldsymbol{h} \in \mathbb{R}_+^k, \end{cases} \tag{5}$$

where the matrix $\mathbf{D}_{\boldsymbol{x}} \in \mathbb{R}^{k\times k}$ represents the covariance of all $\sqrt{k}\times\sqrt{k}$ windows in $\boldsymbol{x}$. Using VB [1], the approximate density $\mathrm{q}(\boldsymbol{x})$ is obtained by minimizing the the Kullback–Leibler (KL) divergence between the approximate and the exact distribution, i.e., $\mathrm{KL}(\mathrm{q}(\boldsymbol{x})||\mathrm{p}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{h}^{(t-1)}))$. Conversely, here we propose to use EP [19] to build $\mathrm{q}(\boldsymbol{x})$. The details of EP, which targets the reverse KL divergence $\mathrm{KL}(\mathrm{p}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{h}^{(t-1)})||\mathrm{q}(\boldsymbol{x}))$, are presented in the following section.

III. Expectation Propagation

A. General principle

EP [19] approximates the probability distribution $\mathrm{p}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{h}^{(t-1)})$ by a simpler Gaussian distribution $\mathrm{q}(\boldsymbol{x})$ such that $\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{x}|\boldsymbol{m}_{\boldsymbol{x}},\mathbf{C}_{\boldsymbol{x}})$ (with $\boldsymbol{y}$ known). Using the block-formulation of EP, the joint probability distribution $\mathrm{p}(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{h}^{(t-1)})$ can be factorized into three different factors f_0, f_1, f_2 such that

$$\mathrm{p}(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{h}^{(t-1)}) \propto \prod_{i=0}^{2} f_i(\boldsymbol{x}), \tag{6}$$

where $f_0(\boldsymbol{x}) \propto \mathrm{p}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}^{(t-1)})$, $f_1(\boldsymbol{x}) \propto \mathrm{p}_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x}) \propto \mathrm{p}_2(\boldsymbol{x})$. EP approximates each $f_i(\boldsymbol{x})$ by a simpler Gaussian distribution $\tilde{f}_i(\boldsymbol{x}), \boldsymbol{x}\sim\mathcal{N}(\boldsymbol{x}|\boldsymbol{m}_i,\mathbf{C}_i)$ such that

$$\mathrm{q}(\boldsymbol{x}) \propto \prod_{i=0}^{2} \tilde{f}_i(\boldsymbol{x}). \tag{7}$$

At each iteration of EP, the approximation factors $\tilde{f}_i(\boldsymbol{x})$ are refined sequentially by minimizing the KL divergence between the tilted distribution $\tilde{\mathrm{q}}_i(\boldsymbol{x}) = f_i(\boldsymbol{x})\ \mathrm{q}_{\backslash i}(\boldsymbol{x})$ and $\mathrm{q}(\boldsymbol{x})$ for $i=0,1,2$, where $\mathrm{q}_{\backslash i}(\boldsymbol{x}),\ \boldsymbol{x}\sim\mathcal{N}(\boldsymbol{x}|\boldsymbol{m}_{\backslash i},\mathbf{C}_{\backslash i})$ is the i-th cavity distribution, *i.e.* the product of all but the i-th approximation factor, and is given by

$$\mathrm{q}_{\backslash i}(\boldsymbol{x}) \propto \mathrm{q}(\boldsymbol{x})/\tilde{f}_i(\boldsymbol{x}). \tag{8}$$

The solution to the KL minimization problem is found by matching the moments of the tilted distribution $\tilde{\mathrm{q}}_i(\boldsymbol{x})$ to the EP approximation $\mathrm{q}(\boldsymbol{x})$. Then the corresponding approximation factor $\tilde{f}_i(\boldsymbol{x})$ is refined using $\mathrm{q}(\boldsymbol{x})/\mathrm{q}_{\backslash i}(\boldsymbol{x})$. While EP is not guaranteed to converge in general and might oscillate [19], damping strategies can be implemented [20] to mitigate convergence issues.

A computationally attractive implementation of EP can be achieved by considering diagonal approximate covariance matrices while moment matching, i.e., $\mathbf{C}_i = \mathrm{diag}(\boldsymbol{v}_i)$ and $\mathbf{C}_{\boldsymbol{x}} = \mathrm{diag}(\boldsymbol{v}_{\boldsymbol{x}})$. Note that this approximation implies that the images pixels are assumed a posteriori independent in $\mathrm{q}(\boldsymbol{x})$, as in VB.

The update of the approximation factors $\tilde{f}_i(\boldsymbol{x})$ depends on computing the moments of the corresponding tilted distributions $\tilde{\mathrm{q}}_i(\boldsymbol{x})$, where $\tilde{\mathrm{q}}_0(\boldsymbol{x}) \propto \exp^{-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\mathbf{H}\boldsymbol{x}\|_2^2}\mathrm{q}_{\backslash 0}(\boldsymbol{x})$, $\tilde{\mathrm{q}}_1(\boldsymbol{x}) \propto \exp^{-\lambda\mathrm{TV}(\boldsymbol{x})}\mathrm{q}_{\backslash 1}(\boldsymbol{x})$ and $\tilde{\mathrm{q}}_2(\boldsymbol{x}) \propto \iota_{\mathbb{R}_+^N}(\boldsymbol{x})\mathrm{q}_{\backslash 2}(\boldsymbol{x})$. Considering that the covariances of $\mathrm{q}_{\backslash i}(\boldsymbol{x})$ are kept diagonal, the marginal means and variances of $\tilde{\mathrm{q}}_i(\boldsymbol{x})$ can be obtained in closed form [20], [21]. The update of $\tilde{f}_0$ is briefly recalled in the next paragraph. The update of $\tilde{f}_1$ is not detailed here to due space constraints but can be derived using the MRF properties of f_1.

	Strategy (I)	Strategy (II)	Strategy (III)
$\widetilde{\mathbf{C}}_{\boldsymbol{x}}$	Expensive-$\widetilde{\mathbf{C}}_{\boldsymbol{x}}$	Fast-$\widetilde{\mathbf{C}}_{\boldsymbol{x}}$	RBMC-$\widetilde{\mathbf{C}}_{\boldsymbol{x}}$
$\mathbf{D}_{\boldsymbol{x}}$	Expensive-$\mathbf{D}_{\boldsymbol{x}}$	Fast-$\mathbf{D}_{\boldsymbol{x}}$	RBMC-$\mathbf{D}_{\boldsymbol{x}}$

TABLE I: Acronyms of the different computation and approximations strategies presented in Section III-B for the covariance matrix $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ and the covariance matrix $\mathbf{D}_{\boldsymbol{x}}$.

B. Covariance matrix computation and approximations

The bottleneck of EP arises during the update of $\tilde{f}_0$, when matching the moments of $\tilde{\mathrm{q}}_0(\boldsymbol{x})$ and $\mathrm{q}(\boldsymbol{x})$. The updated moments of $\mathrm{q}(\boldsymbol{x})$ are given by

$$\mathbf{C}_{\boldsymbol{x}}^{-1} = \mathbf{C}_{\backslash 0}^{-1} + \sigma^{-2}\mathbf{H}^T\mathbf{H}, \tag{9}$$

$$\boldsymbol{m}_{\boldsymbol{x}}^{\mathrm{new}} = \mathbf{C}_{\boldsymbol{x}}(\mathbf{C}_{\backslash 0}^{-1}\boldsymbol{m}_{\backslash 0} + \sigma^{-2}\mathbf{H}^T\boldsymbol{y}), \tag{10}$$

$$\boldsymbol{v}_{\boldsymbol{x}}^{\mathrm{new}} = \mathrm{diag}(\widetilde{\mathbf{C}}_{\boldsymbol{x}}), \tag{11}$$

where $\mathbf{C}_{\backslash 0} = \mathrm{diag}(\boldsymbol{v}_{\backslash 0})$. Although we are only interested in the diagonal of $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ to update $\boldsymbol{v}_{\boldsymbol{x}}^{\mathrm{new}}$ (11), this requires the costly inversion of matrix $\mathbf{C}_{\boldsymbol{x}}^{-1} \in \mathbb{R}^{N\times N}$. This inversion problem is common in all Bayesian approximation methods such as VB (see Eq.24 in [1]). In addition, the covariance matrix $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ is used to compute the matrix $\mathbf{D}_{\boldsymbol{x}}$, which is required in the M-step, $\mathbf{D}_{\boldsymbol{x}}(i,j) = \sum_{n=1}^{N}\widetilde{\mathbf{C}}_{\boldsymbol{x}}(i+n,j+n)$ [1]. Different strategies can be adopted in EP and VB to compute $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$:

- Strategy (I): $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ is a full matrix computed as the inverse of $\mathbf{C}_{\boldsymbol{x}}^{-1}$.
- Strategy (II): $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ is a diagonal matrix whose diagonal is the inverse of the diagonal of $\mathbf{C}_{\boldsymbol{x}}^{-1}$.
- Strategy (III): $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ is a diagonal matrix whose diagonal is approximated using an efficient covariance approximation method, dubbed Rao–Blackwellized Monte Carlo (RMBC) [22].

While EP uses only the diagonal of $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ while iterating (11), we can use (9) at the last iteration such the off-diagonal terms can be incorporated in $\mathbf{D}_{\boldsymbol{x}}$ in the M-step. A similar strategy can be used for VB (Eq.24 in [1]). This allows running EP and VB with a certain covariance approximation while estimating the kernel with a different covariance approximation. We present in TABLE I the adopted acronyms of the different computation and approximations strategies for the covariance matrices $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ and $\mathbf{D}_{\boldsymbol{x}}$.

IV. Simulations and results

In this section, we showcase the performance of the proposed approach on simulations utilizing the Cameraman image, with a peak value normalized to 1. Simulations conducted on different test images are provided in [23] (available at https://researchportal.hw.ac.uk/en/persons/

yoann-altmann). The original image is degraded with a linear motion blur of length 5 and angle 40°, and corrupted with i.i.d. Gaussian noise with variance $\sigma^2 = 0.0025$, corresponding to SNR = 20 dB. We start by evaluating the performance of EP in terms of reconstruction quality and computational time assuming the kernel is known. EP with TV prior and non-negativity constraint, denoted by EP (TV+), is compared to EP (TV) (*i.e.* no non-negativity constraint is imposed) and VB with TV prior, denoted by VB (TV) [1]. We also report the results of the proximal MCMC algorithm [24] with TV prior and non-negativity constraint, denoted by MCMC (TV+), and with only TV prior, MCMC (TV). The regularization parameter associated with the TV prior is set to $\lambda = 6.7$ in all the methods. For EP and VB, the covariance matrix $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ is approximated as the inverse of the diagonal of $\mathbf{C}_{\boldsymbol{x}}^{-1}$. The MCMC chain is of length 2×10^5 samples including a 1000-sample burn-in period. Fig. 1, (b)-(f) displays the deconvolution results (MMSE estimators) obtained by different methods with image size $N = 128^2$ pixels. The figure clearly shows comparable mean images for all the methods. Interestingly, the images of marginal variances of EP (TV+) and EP (TV) (Fig. 1, (b) and (d), second row) are very close to those obtained with MCMC (Fig. 1 (c) and (e), second row), confirming the improved reliability of the estimated variances of EP over VB here. In addition, the computational time of EP is around 4 orders of magnitudes lower than that for MCMC. We also notice that VB (TV) seems to under-estimate the variance (Fig. 1, (f), second row). In addition, incorporating the non-negativity constraint in the prior image model tends to decrease the uncertainty in the low-intensity regions (see Fig. 1, (b) and (c), second row).

We then assess the performance of EP when the kernel is unknown. The proposed BID approach EP-EM with TV prior and non-negativity constraint, EP-EM (TV+), and with only TV prior, EP-EM (TV) is compared to VB-EM with TV prior, denoted by VB-EM (TV) [1]. Tests are carried out with different image sizes $N = \{64^2, 128^2, 256^2, 512^2\}$. Fig. 2 shows the image PNSR and compute time scores corresponding to running EP-EM (TV+), EP-EM (TV) and VB-EM (TV) with the different covariance computation strategies explained in Section III-B. Fig. 2, first row shows a comparable computational time between EP-EM (TV+) and VB-EM (TV) for all strategies while EP-EM (TV) takes more time to converge in some scenarios (see Fig. 2, (c)). Fig. 2, the second row suggests that running EP and VB with Fast-$\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ and RBMC-$\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ gives PSNR values very close to those obtained when running with Expensive-$\widetilde{\mathbf{C}}_{\boldsymbol{x}}$. However, the computational time can be drastically reduced as shown in Fig. 2, first row. Finally, we notice that both EP-EM (TV+) and EP-EM (TV) maintain better PSNR scores (around 1 dB enhancement) for all image sizes. This supports the observation that VB can be less accurate than EP for some applications as stated in [20].

V. Conclusions

In this paper, we presented a blind deconvolution method leveraging EP within an EM-based algorithm. As opposed to other works in the literature where the VEM was deployed to approximate the intractable E-step of the EM algorithm, we propose the use of the EP algorithm as better approximate method. We showed using controlled simulation that the proposed EP-EM algorithm can provide better supervised and blind deconvolution results reflected by higher PSNR and more reliable uncertainty maps than VB-EM for a comparable computational time, when using the same original model. Future work will include a more detailed analysis of the EP-EM for different blur sizes and more complex image prior models.

References

[1] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind image deconvolution: A review," *Digit. Signal Process.*, vol. 47, pp. 116–127, 2015.

[2] P. Campisi and K. Egiazarian, *Blind image deconvolution: theory and applications.* CRC press, 2017.

[3] T. E. Bishop, S. D. Babacan, B. Amizic, A. K. Katsaggelos, T. Chan, and R. Molina, "Blind image deconvolution: problem formulation and existing approaches," pp. 1–41, 2007.

[4] S. Chaudhuri, R. Velmurugan, and R. Rameshan, *Blind image deconvolution.* Springer, 2016.

[5] P. Satish, M. Srikantaswamy, and N. K. Ramaswamy, "A comprehensive review of blind deconvolution techniques for image deblurring." *Trait. Signal*, vol. 37, no. 3, 2020.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J R Stat Soc Series B*, vol. 39, no. 1, pp. 1–22, 1977.

[7] A. C. Likas and N. P. Galatsanos, "A variational approach for bayesian blind image deconvolution," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2222–2233, 2004.

[8] R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, 2006.

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

[10] A. Mohammad-Djafari, "Bayesian blind deconvolution of images comparing jmap, em and bva with a student-t a priori model," in *International Workshops on Electrical Computer Engineering Subfields*, 2014, pp. 98–103.

[11] X. Zhou, R. Molina, F. Zhou, and A. Katsaggelos, "Fast iteratively reweighted least squares for lp regularized image deconvolution and reconstruction," in *IEEE ICIP*, 2014, pp. 1783–1787.

[12] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *IEEE ECCV*. Springer, 2012, pp. 341–355.

[13] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, pp. 370–375, 1998.

[14] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind deconvolution using a total variation prior," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 12–26, 2008.

[15] R. Molina, A. K. Katsaggelos, J. Abad, and J. Mateos, "A bayesian approach to blind deconvolution based on dirichlet distributions," in *IEEE ICASSP*, vol. 4, 1997, pp. 2809–2812.

[16] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in *IEEE CVPR*, 2009, pp. 104–111.

[17] D. Ge, J. Idier, and E. Le Carpentier, "A new mcmc algorithm for blind bernoulli-gaussian deconvolution," in *IEEE EUSIPCO*, 2008, pp. 1–5.

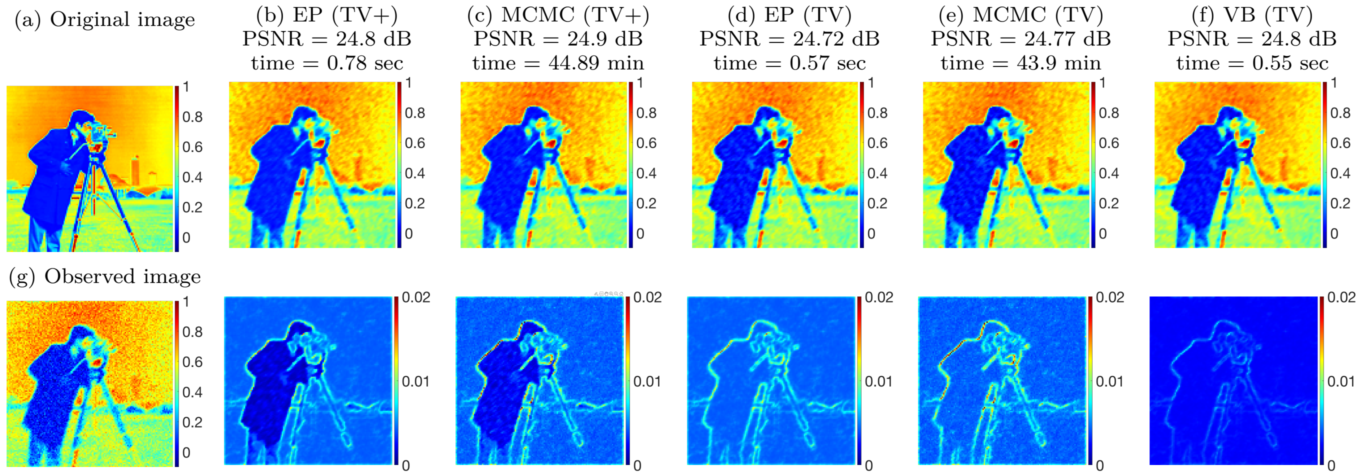

Fig. 1: Supervised deconvolution: original (a) and degraded (g) Cameraman image. (b)-(f) deconvolution mean images (top) and variance images (bottom) obtained with the diffent approaches.

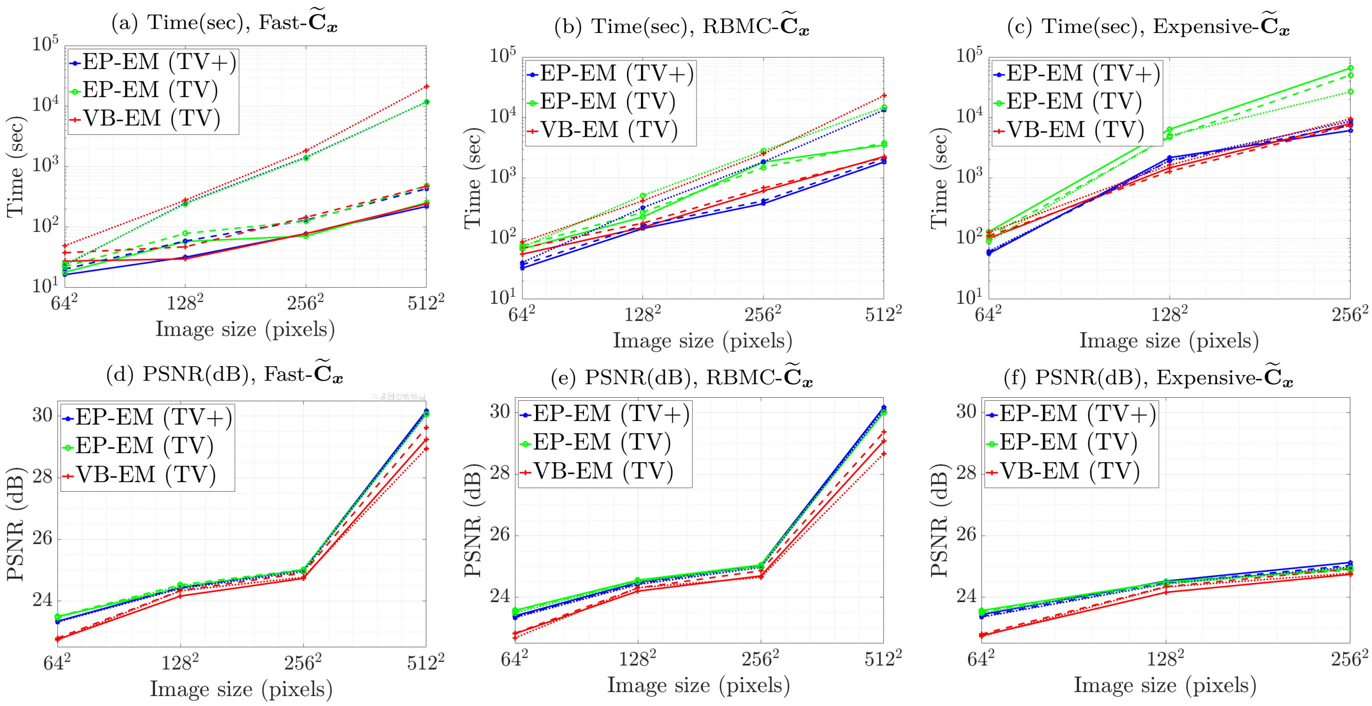

Fig. 2: Blind deconvolution: computational time (top) and image PSNR (bottom) as a function of the image size. The different panels show the performance of running EP and VB with different types of $\widetilde{\mathbf{C}}_{\boldsymbol{x}}$ covariance approximations. Each panel displays the results of estimating the kernel with different types of $\mathbf{D}_{\boldsymbol{x}}$ covariance approximations: solid lines using Fast-$\mathbf{D}_{\boldsymbol{x}}$, dashed lines using RBMC-$\mathbf{D}_{\boldsymbol{x}}$, and dotted lines using Expensive-$\mathbf{D}_{\boldsymbol{x}}$.

[18] A. F. Vidal, V. De Bortoli, M. Pereyra, and A. Durmus, "Maximum likelihood estimation of regularization parameters in high-dimensional inverse problems: An empirical bayesian approach part i: Methodology and experiments," *SIAM J. Imag. Sci.*, vol. 13, no. 4, pp. 1945–1989, 2020.

[19] T. P. Minka, "A family of algorithms for approximate bayesian inference," Ph.D. dissertation, MIT, 2001.

[20] J. M. Hernández-Lobato, D. Hernández-Lobato, and A. Suárez, "Expectation propagation in linear regression models with spike-and-slab priors," *Mach. Learn.*, vol. 99, no. 3, pp. 437–487, 2015.

[21] Y. Altmann, D. Yao, S. McLaughlin, and M. E. Davies, "Robust linear regression and anomaly detection in the presence of poisson noise using expectation-propagation," in *WCCM*, 2019.

[22] P. Sidén, F. Lindgren, D. Bolin, and M. Villani, "Efficient covariance approximations for large sparse precision matrices," *J. Comput. Graph. Stat.*, vol. 27, no. 4, pp. 898–909, 2018.

[23] A. Abdulaziz, D. Yao, Y. Altmann, and S. McLaughlin, "Blind deconvolution of images corrupted by Gaussian noise using expectation propagation: Supplementary material," Heriot-Watt University, Edinburgh, United Kingdom, Tech. Rep., May 2021.

[24] A. Durmus, E. Moulines, and M. Pereyra, "Efficient bayesian computation by proximal markov chain monte carlo: when langevin meets moreau," *SIAM J. Imag. Sci.*, vol. 11, no. 1, pp. 473–506, 2018.