A VARIATIONAL METHOD FOR BAYESIAN BLIND IMAGE DECONVOLUTION

Aristidis Likas and Nikolas P. Galatsanos
Department of Computer Science
University of Ioannina
Ioannina, Greece 45110
{arly,galatsanos}@cs.uoi.gr

ABSTRACT

In this paper the blind image deconvolution (BID) problem is solved using the Bayesian framework. In order to find the parameters of the proposed Bayesian model we present a new generalization of the expectation maximization (EM) algorithm based on the variational approximation methodology. The proposed variational-based algorithm for BID can be derived in closed form and can be implemented in the discrete Fourier domain. Thus, it is very efficient even for very large images. We demonstrate with numerical experiments that the algorithm which was derived by the variational methodology yields promising improvements as compared to previous Bayesian algorithms for BID. Furthermore, the methodology presented here is very general with potential applications to other Bayesian models for this and other imaging problems.

1. INTRODUCTION

The blind image deconvolution (BID) problem is a difficult and challenging problem because the observed data does not define uniquely the convolved signals. In many applications the observed images have been blurred by an unknown or a partially known point spread function (PSF). Such examples include astronomy and remote sensing where the atmospheric turbulence cannot be exactly measured, medical imaging where the PSF of different instruments is measured and thus is subject to errors, in photography where the camera settings might be unknown etc.

A plethora of methods has been proposed to solve this problem, see [1] for a six year old survey. Since in BID the observed data are not sufficient to specify the convolved functions, most recent methods attempt to incorporate in the BID algorithm some prior knowledge about these functions. Since it is very hard to track the properties of the PSF and the image simultaneously most recent BID methods include a loop in which image and PSF are estimated in an alternating manner keeping the other constant; see for example [2, 3]. Prior knowledge in the form of convex sets and regularization with anisotropic diffusion functionals were used in [2, 3]. In [4] a Bayesian framework was used to tackle this problem.

The Bayesian approach provides a very powerful and flexible methodology for estimation problems including BID. It provides a structured framework to include prior knowledge concerning the quantities to be estimated. Nevertheless, in most Bayesian estimation problems of interest in order to obtain the final solution we are faced either with a hard optimization problem or a very difficult integral that cannot be computed analytically. In [4, 11] a Laplace approximation of the integral that appears in the Bayesian formulation of the BID problem was used. In spite of this, it was reported in [4] that the estimates of the statistics of the errors in the PSF and the image could be in error by orders of magnitude depending on their initialization. Thus, using the Bayesian approach in [4], it is impossible to obtain accurate restorations unless accurate prior knowledge about either the statistics of the error in the PSF or the image is available in the form of hyper-priors [11].

In what follows we present a new Bayesian model of the BID problem and a new variational framework which was applied to solve for it. This results in an algorithm that finds iteratively the parameters of this model. We present numerical experiments where we observe that the proposed algorithm provides simultaneously reasonable estimates of the errors in *both* the PSF and the image. Thus, a good restoration of the degraded image is possible even when very little is known a priori about the PSF and the image. In addition, the proposed methodology is very general and can be applied to many other Bayesian models for both the BID and other imaging problems.

2. BACKGROUND ON VARIATIONAL METHODS

The variational framework constitutes a generalization of the well-known Expectation Maximization (EM) algorithm for likelihood maximization in Bayesian estimation problems with "hidden variables". The EM algorithm has been proved a valuable tool for many problems, since it provides an elegant approach to bypass difficult optimization and integrations in Bayesian estimation problems. In order to efficiently apply the EM algorithm two requirements should be fulfilled [5]: i) In the E-step we should be able to compute the probability density function (PDF) of the "hidden variables" given the observation data. ii) In the M-step, it is highly preferable to have analytical formulas for the update equations of the parameters. Nevertheless, in many problems it is not possible to meet the above requirements and several variants of the basic EM algorithm have emerged. For example, a variant of the EM called the "generalized EM" (GEM) proposes a partial M step in which the likelihood always improves. In many cases partial implementation of the E step is also natural. An algorithm along such lines was investigated in [6].

The most difficult situation for applying the EM algorithm emerges when it is not possible to specify the conditional PDF of the hidden variables given the observed data that is required in the E-step. In such cases the implementation of the EM is not possible. This significantly restricts the range of problems where the EM can be applied. To overcome this serious shortcoming of the EM, the variational methodology was developed [12]. In addition, it can be shown that the EM naturally arises as a special case of variational methodology.

Assume that x and s are the observed and hidden variables, respectively, and that θ are the unknown parameters to be estimated. All PDFs are parameterized by the parameters, ie. $p(x;\theta)$, $p(s,x;\theta)$ and $p(s/x;\theta)$ and we omit θ for brevity. By selecting an *arbitrary PDF* $q(s)$ of the hidden variables s it is easy to show that:

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

where E_q denotes the expectation with respect to $q(s)$.

The above equation can be written as,

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

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

Rearranging the previous equation we obtain:

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

where $H(q)$ is the entropy of *q(s)*. From Eq. (1) it is clear that $\mathbf{F}(q,\theta)$ provides a lower bound for the likelihood of θ parameterized by the family of PDFs $q(s)$, since $KL(q(s)//p(s/x))\geq 0$. When $q^*(s)=p(s/x;\theta)$, then the lower bound becomes exact: $F(q^*,\theta)=L(\theta)$. Using this framework the EM can then be viewed as a special case when $q^*(s)=p(s/x;\theta)$.

However, the previous framework allows based on Eq. (1) to find a local maximum of $L(\theta)$ *using an arbitrary* PDF *q(s)*. This is a very useful generalization because it bypasses one of the main restrictions of the EM that of exactly knowing *p(s/x)*. The variational method works to maximize the lower bound of $\mathbf{F}(q,\theta)$ with respect to both θ and *q*. This is justified by a theorem in [6] stating that, if $\mathbf{F}(q,\theta)$ has a local maximum at $q^*(s)$ and θ^*, then $L(\theta)$ has a local maximum at θ^*. Furthermore, if $\mathbf{F}(q,\theta)$ has a global maximum at $q^*(s)$ and θ then $L(\theta)$ has a global maximum at θ^*. Consequently the variational EM approach can be described as follows:

$$\text{E-step: } q^{(t+1)}=\arg\max_q F(q,\theta^{(t)})$$

$$\text{M-step: } \theta^{(t+1)}=\arg\max_\theta F(q^{(t+1)},\theta)$$

This iterative approach increases at each step $(t+1)$ the value of the bound $\mathbf{F}(q,\theta)$ until a local maximum is attained.

3. VARIATIONAL BLIND DECONVOLUTION

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

$$g=h*f+w=H\cdot f+w=F\cdot h+w \quad (2)$$

and we assume the $N\times 1$ vector g to be the observed variables, the $N\times 1$ vectors f and h are the hidden variables and w is Gaussian noise. For this problem we assume Gaussian PDFs for the priors of f and h. In other words, we assume

$p(f)=N(\mu_f,\Sigma_f)$, $p(h)=N(\mu_h,\Sigma_h)$ and $p(w)=N(0,\Sigma_w)$.

Thus, the parameters are $\theta=[\mu_f,\Sigma_f,\mu_h,\Sigma_h,\Sigma_w]^t$. The key difficulty here is that the PDF $p(f,h/g;\theta)$ *is unknown*. This does not allow the direct application of the EM. However, with the variational approximation we can bypass this difficulty. More specifically, we *select* $q(s)$ to be $q(s)=q(h,f)=q(h)\cdot q(f)=p(h/g;f)\cdot p(f/g;h)$. (3)

Since the observed data can be seen both as $g=\begin{bmatrix}H & I\end{bmatrix}\begin{bmatrix}f\\w\end{bmatrix}=\begin{bmatrix}F & I\end{bmatrix}\begin{bmatrix}h\\w\end{bmatrix}$ we have that $p(f/g;h)=N(m_{f/g},C_{f/g})$ and $p(h/g;f)=N(m_{h/g},C_{h/g})$ with the means and covariances known [9]. The above choice of q is justified by the fact that it seems a reasonable approximation of $p(f,h/g;\theta)$ that also leads to a *tractable* variational formulation.

From the RHS of Equation (3) we have

$$\mathbf{F}(q,\theta)=E_q(\log p(x,s))+H(q) \quad (4)$$

where, $p(x,s)=p(g,f,h)=p(g/f,h)\cdot p(f)\cdot p(h)$ with $p(g/h,f)=N(h*f,\Sigma_w)$. The variational approach requires the computation of the expectation (Gaussian integral) in Eq. (4). The term in the expectation of the first part of Eq. (4) is given by

$$\log p(g,f,h) = \log p(g/f,h) + \log p(f) + \log p(h) =$$

$$K_1 - \frac{1}{2}\Big\{\log|\Sigma_w| + (g-h*f)^t\,\Sigma_w^{-1}(g-h*f) + \log|\Sigma_f| + (f-\mu_f)^t\,\Sigma_f^{-1}(f-\mu_f)$$

$$+\log|\Sigma_h| + (h-\mu_h)^t\,\Sigma_h^{-1}(h-\mu_h)\Big\}.$$

where K is a constant. In order to facilitate computations for large images, we will assume circulant convolutions in (2) and that matrices $\Sigma_f, \Sigma_h, \Sigma_w$, $C_{f/g}$,and $C_{f/g}$ are circulant. This allows an easy implementation of matrix-vector equations in the Discrete Fourier Domain (DFT). Computing this expectation and the entropy of $q(s)$ in (4) we can write the result in the DFT domain as

$$\mathbf{F}(q,\theta) = \text{C-}\frac{1}{2}\sum_{k=0}^{N-1}\Big(\log\Lambda_w(k) + \log\Lambda_f(k) + \log\Lambda_h(k) + \log S_{f/g}(k) + \log S_{h/g}(k)\Big)$$

$$-\frac{1}{2}\sum_{k=0}^{N-1}\frac{\frac{1}{N}\Big(|G(k)|^2 - 2\,\mathrm{Re}\big\{M_{f/g}(k)M_{h/g}(k)G^*(k)\big\}\Big)}{\Lambda_w(k)} \qquad (5)$$

$$-\frac{1}{2}N\sum_{k=0}^{N-1}\frac{\Big(S_{f/g}(k) + \frac{1}{N}\big|M_{f/g}(k)\big|^2\Big)\Big(S_{h/g}(k) + \frac{1}{N}\big|M_{h/g}(k)\big|^2\Big)}{\Lambda_w(k)}$$

$$-\frac{1}{2}\sum_{k=0}^{N-1}\frac{\Big(S_{f/g}(k) + \frac{1}{N}\big|M_{f/g}(k)\big|^2\Big) + \frac{1}{N}\Big(|M_f(k)|^2 - 2\,\mathrm{Re}\big\{M_f^*(k)M_{f/g}(k)\big\}\Big)}{\Lambda_f(k)}$$

$$-\frac{1}{2}\sum_{k=0}^{N-1}\frac{\Big(S_{h/g}(k) + \frac{1}{N}\big|M_{h/g}(k)\big|^2\Big) + \frac{1}{N}\Big(|M_h(k)|^2 - 2\,\mathrm{Re}\big\{M_h^*(k)M_{h/g}(k)\big\}\Big)}{\Lambda_h(k)}$$

where $S_{f/g}(k)$, $S_{h/g}(k)$, $\Lambda_f(k)$, $\Lambda_h(k)$, and $\Lambda_w(k)$, the eigenvalues of the $N \times N$ circulant covariance matrices $C_{f/g}$, $C_{h/g}$, Σ_f, Σ_h and Σ_w, respectively. Also $G(k)$, $M_{f/g}(k)$ and $M_{h/g}(k)$ are the DFT coefficients of the vectors g, $m_{f/g}$ and $m_{h/g}$, respectively.

In the E-step, Eq. (5) is minimized with respect to the parameters of *q(s)*, which are the $M_{f/g}(k)$, $C_{f/g}(k)$, $M_{h/g}(k)$ and $C_{h/g}(k)$. It can be easily shown that this maximization can be performed analytically, thus providing the updated values for the conditional means and variances.

In the M-step, the parameters of $q(s)$ are considered fixed and Eq. (5) is maximized with respect to the parameters θ leading to the following update equations for θ:

$$M_f(k) = M_{f/g}(k),\; M_h(k) = M_{h/g}(k),$$

$$\Lambda_f(k) = S_{f/g}(k)\;\; \Lambda_h(k) = S_{h/g}(k).$$

4. NUMERICAL EXPERIMENTS

In our experiments we used a simultaneously autoregressive model [8] for the image, in other words we assumed $p(f) = N\Big(0, \big(\alpha Q^T Q\big)^{-1}\Big)$, for h we assumed $C_h = \sigma_h^2 I$ and $p(w) = N\big(0, \sigma^2 I\big)$, where Q the circulant matrix that represents the convolution with the Laplacian operator. In this paper we present two experiments where the PSF was a sum of a random component and a deterministic component as $h = h_0 + \Delta h$. For the first experiment the deterministic component h_0 was a Gaussian kernel with variance $\sigma_h^2 = 20$. The random component was $\Delta h \sim N\big(0, \beta I\big)$ with $\beta = .001$. Gaussian noise w with $\sigma_w^2 = .001$ was also added to generate the observations in Fig. 1(a). The iterative Wiener filter [9] was used to restore this image (Fig. 1(b)) and the variational approach presented in this paper (Fig. 1(c)). As metric of performance the improvement in signal-to-noise ratio (ISNR) was used. This metric is defined as $ISNR = \log 10 \frac{\|f-g\|^2}{\|f-\hat{f}\|^2}$, where $\hat{f}$ the restored image. In the second experiment a Gaussian PSF with variance $\sigma_h^2 = 20$ was used for blurring to generate the observations while a Gaussian PSF with $\sigma_h^2 = 12$ was given to the restoration algorithm. The variance of the additive noise in the data w was the same as before. As the *initial estimate* of the mean of the PSF provided to the proposed variational algorithm and as the PSF of the iterative Wiener filter [9] the Gaussian with $\sigma_h^2 = 12$ was given . The results of this experiment are shown in Fig.s 2. The variational algorithm in both experiments showed the ability to estimate simultaneously the variances of the errors in both the observation and the PSF. For example, in the first experiment the noise in the PSF was estimated as $\hat{\beta} = .0007$ and the noise in the data as $\hat{\sigma}_W^2 = .0015$. This is an important improvement over the hierarchical Bayesian approach of [4] where the estimates could be in error by orders of magnitude. Furthermore, the proposed approach does not rely on the use of hyper-priors as in [11].

5. CONCLUSIONS AND FUTURE RESEARCH

In this paper a new framework for estimation problems was presented. This framework is an extension of the popular EM algorithm and was applied to a Bayesian formulation of the blind image deconvolution problem. We demonstrated with numerical experiments that this algorithm showed promise as compared to previous Bayesian solutions to this problem. We plan to test this algorithm with more numerical experiments and derive other variational algorithms for this problem using different $q(s)$.

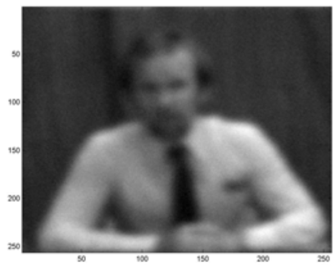

Fig. 1(a): Observed image

Fig. 1(b): Iterative Wiener (ISNR: 1.1dB)

Fig1 (c): Variational (ISNR: 2.1dB)

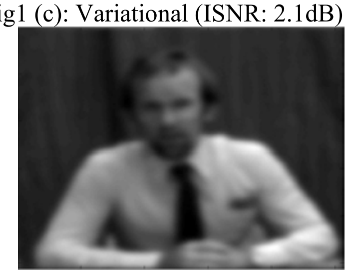

Fig. 2(a): Observed Image

Fig. 2(b): Iterative Wiener (ISNR2.1dB)

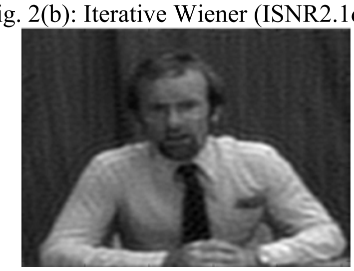

Fig. 2 (c): Variational (ISNR=3.2dB)

6. REFERENCES

[1] D. Kundur and D. Hatzinakos, "Blind Image Deconvolution," *IEEE Signal Processing Magazine,* pp. 43-64, May 1996.

[2] Y. Yang, N. P. Galatsanos, and H. Stark, "Projection Based Blind Deconvolution," *Journal of the Optical Society of America-A*, Vol. 11, No. 9, pp. 2401-2409, September 1994.

[3] Y.L. You, and M. Kaveh, "Blind Image Restoration by Anisotropic Regularization," *IEEE Tran. on Image Processing*, Vol. 8, pp. 396-407, March 1999.

[4] N. P. Galatsanos, V. N. Mesarovic, R. M. Molina and A. K. Katsaggelos, "Hierarchical Bayesian Image Restoration from Partially-Known Blurs," *IEEE Trans. on Image Processing*, Vol. 9, No. 10, pp. 1784-1797, October 2000.

[5] A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the E-M algorithm," *J. Roy. Stat. Soc.*, vol. B39, pp 1-37, 1977.

[6] R. M. Neal, and G. E. Hinton, "A View of the E-M Algorithm that Justifies Incremental, Sparse and Other Variants", In M. I. Jordan (ed.) *Learning in Graphical Models*, pp. 355-368. Cambridge, MA: MIT Press 1998.

[7] 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" (editor M.I. Jordan), pp. 105-162, Cambridge, MA: MIT Press 1998.

[8] R. Molina, B. D. Ripley, "Using spatial models as priors in astronomical images analysis", *J. Appl. Stat.*, vol.16, pp.193-206, 1989.

[9] A. K. Katsaggelos, ed., *Digital Image Restoration*, Ch. 6, New York: Springer-Verlag, 1991.

[10] T. S. Jaakkola, Variational methods for inference and learning in graphical models, Ph.D. thesis, MIT, 1997.

[11] N. P. Galatsanos, V. N. Mesarovic, R. M. Molina, J. Mateos, and A. K. Katsaggelos, "Hyper-parameter Estimation Using Gamma Hyper-priors in Image Restoration from Partially-Known Blurs," *Optical Engineering*, 41(8), pp. 1845-1854, August 2002.