VARIATIONAL BAYES AND MEAN FIELD APPROXIMATIONS FOR MARKOV FIELD UNSUPERVISED ESTIMATION

Ali Mohammad-Djafari and Hacheme AYASSO

Laboratoire des Signaux et Systmes, (UMR8506 CNRS-SUPELEC-UNIV PARIS SUD 11)
3 rue Joliot-Curie, F-91192 Gif-sur-Yvette cedex, France

ABSTRACT

We consider the problem of parameter estimation of Markovian models where the exact computation of the partition function is not possible or computationally too expensive with MCMC methods. The main idea is then to approximate the expression of the likelihood by a simpler one where we can either have an analytical expression or compute it more efficiently. We consider two approaches: Variational Bayes Approximation (VBA) and Mean Field Approximation (MFA) and study the properties of such approximations and their effects on the estimation of the parameters.

1. INTRODUCTION

Markovian models have gained a great interest in many domains, especially in Bayesian framework for inverse problems in imaging systems [1, 2, 3, 4], [6], for their capacity to represent the local spatial dependencies between neighbouring sites (pixels). Markovian models are described either as a collection of conditional probability laws: $p(x_i|x_j, j \in \mathcal{V}(i)),\ \forall i \in \mathcal{I}$, where $\mathcal{I}$ represents a set (of pixel positions for example) and $\mathcal{V}(i)$ represents the neighbors of i, or as a global joint probability law (Gibbs measure):

$$p(\boldsymbol{x}|\lambda) = \frac{1}{Z_p(\lambda)} \exp\left(-\lambda \mathcal{E}(\boldsymbol{x})\right), \tag{1}$$

where $\boldsymbol{x} = \{x_i, \forall i \in \mathcal{I}\} \in \mathcal{X}$ and

$$Z_p(\lambda) = \int_{\mathcal{X}} \exp\left(-\lambda \mathcal{E}(\boldsymbol{x})\right)\, \mathrm{d}\boldsymbol{x} \tag{2}$$

is the partition function, $\mathcal{E}(.)$ is a Hamiltonian (energy function): $\mathcal{E}(\boldsymbol{x}) = \sum_{c\in\mathcal{C}} \Phi_c(\boldsymbol{x}_c)$, $\mathcal{C}$ is the set of cliques defined over the set $\mathcal{I}$ with the neighbourhood system $\mathcal{V}(i)$ and $\Phi_c(.)$ are their associated potential functions.

If we consider each pixel i of an image as a particle and its gray level x_i as the state of that particle, then $p(\boldsymbol{x}|\lambda)$ can be interpreted as the *Boltzmann's Law* with $\lambda = 1/T$, where T is the temperature of the system. Then $F_{\text{Helmoltz}} = -\ln Z_p(\lambda)$ is the *Helmotz free energy* of the system. This quantity is a fundamentally important in statistical mechanics and a great amount of works in Physics are devoted for computing it, and, because its direct and exact computation is often too expensive, there has been great amount of works devoted to developing methods to obtain good approximations to it.

One important technique is based on a variational approach [7] where a *trial* distribution $q(\boldsymbol{x})$ is proposed in place of $p(\boldsymbol{x})$ and one defines a *variational free energy*

$$F(q) = U(q) - H(q), \tag{3}$$

where

$$U(q) = < \mathcal{E}(\boldsymbol{x}) >_q = \int_{\mathcal{X}} q(\boldsymbol{x})\, \mathcal{E}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{4}$$

is the *variational average energy* and

$$H(q) = < -\ln q >_q = -\int_{\mathcal{X}} q(\boldsymbol{x}) \ln q(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{5}$$

is *variational entropy*.

Noting that $U(q) = -\ln Z(\lambda) + < \ln p >_q$ and using the definition of the Kullback-Leibler divergence

$$\mathrm{KL}(q|p) = < -\ln \frac{q}{p} >_q = -\int_{\mathcal{X}} q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})}\, \mathrm{d}\boldsymbol{x}, \tag{6}$$

it follows directly that:

$$F(q) = -\ln Z(\lambda) + \mathrm{KL}(q|p) = F_{\text{Helmoltz}} + \mathrm{KL}(q|p). \tag{7}$$

Since $\mathrm{KL}(q|p)$ is always non-negative and zero only if $q = p$, we see that $F(q) \geq F_{\text{Helmoltz}}$, with equality when $q = p$. Thus, minimizing the variational free energy $F(q)$ is a good way to compute $F_{\text{Helmoltz}} = -\ln Z$ and use it where necessary.

In imaging systems, these models are, in general, used as prior models where $\boldsymbol{x} = \{x(\boldsymbol{r}_i), \boldsymbol{r}_i \in \mathcal{R}\}$ represent the pixels of an image $x(\boldsymbol{r})$ where $\boldsymbol{r}_i$ is the spatial position (in a plane for 2D case or in the space for 3D case) of the pixel or voxel number i and $\mathcal{R}$ is either the surface of the image or the volume of the scene. When we use such a model for a class of images, one of the problems is estimating λ.

The classical maximum likelihood (ML) method is:

$$\begin{aligned} \widehat{\lambda}_{\text{MV}} &= \arg\max_{\lambda} \left\{\ln p(\boldsymbol{x}|\lambda)\right\} \\ &= \arg\max_{\lambda} \left\{-\ln Z(\lambda) - \lambda \mathcal{E}(\boldsymbol{x})\right\}, \end{aligned} \tag{8}$$

which needs the computation of $Z(\lambda)$. This solution satisfies :

$$\frac{-\partial \ln Z(\lambda)}{\partial \lambda} = \mathcal{E}(\boldsymbol{x}). \quad (9)$$

In inverse problems in imaging systems, these models are used as prior models for images. In these inverse problems, we do not observe directly the images. If we note by $\boldsymbol{y}$ the observed data in these systems where, in general, we know the forward mathematical model $\boldsymbol{y} = \mathcal{A}(\boldsymbol{x}) + \boldsymbol{\epsilon}$ where $\mathcal{A}$ represents the response of the observation system and $\boldsymbol{\epsilon}$ represents the errors (modeling and measurement noise). Thus, in inverse problems related to these imaging systems, we know the expression of the likelihood $p(\boldsymbol{y}|\boldsymbol{x})$ and using the Bayes rule, we obtain the expression of the posterior law $p(\boldsymbol{x}|\boldsymbol{y};\lambda) \propto p(\boldsymbol{y}|\boldsymbol{x})\ p(\boldsymbol{x}|\lambda)$. From this expression, we see that, when the forward operator $\mathcal{A}$ is not mixing ($\mathcal{A} = \boldsymbol{I}$) or is a local support operator such as a convolution, then using a Markov model as a prior results also to a Markov model as the posterior law. From now then, we do not distinguish between the two cases, because both prior or the posterior laws are Markov models, but only their neighbourhood sizes differ. However, in the following, we consider two cases:

a) Direct modelling of images (training step), where the main problem is then the estimation of the parameter λ from a set of direct observations $\boldsymbol{x}$. This estimation can be done either by ML approach or through the Bayesian MAP criterion which is

$$\begin{aligned} \widehat{\lambda} &= \arg\max_{\lambda} \{\ln p(\lambda|\boldsymbol{x})\} \\ &= \arg\max_{\lambda} \{-\ln Z(\lambda) - \lambda \mathcal{E}(\boldsymbol{x}) + \ln \pi(\lambda)\}. \end{aligned} \quad (10)$$

b) Infering on λ in an unsupervised Bayesian approach where we have some data $\boldsymbol{y}$ related to the unknowns $\boldsymbol{x}$ through a forward model giving the expression of likelihood $p(\boldsymbol{y}|\boldsymbol{x})$ and using (1) as a prior for $\boldsymbol{x}$ and a prior for $\pi(\lambda)$ for the parameter λ which results to the joint posterior

$$p(\boldsymbol{x}, \lambda|\boldsymbol{y}) \propto p(\boldsymbol{x}|\boldsymbol{y};\lambda)\, \pi(\lambda) \propto p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x}|\lambda)\, \pi(\lambda), \quad (11)$$

which is then used to infer on $\boldsymbol{x}$ and on λ.

In the first case, the main problem is the estimation of the parameter λ from a set of direct observation of $\boldsymbol{x}$ (training set, for prior model parameter estimation). In the second case, a first problem is to provide a point estimator for the unknown $\boldsymbol{x}$ such as the Maximum A posteriori (MAP) or the Posterior Mean (PM) from a set of data and knowing the prior parameters λ. The second problem, in an unsupervised Bayesian framework, is also to estimate λ (which is called hyperparameter in that context) either from $\boldsymbol{x}$ computed in a previous iteration or directly from the observed data $\boldsymbol{y}$.

Both problems are, in general, intractable or needs high computational cost, because $Z(\lambda)$ has not an explicit form except for a few simple cases, for example, when $\mathcal{E}$ is quadratic (Gauss-Markov models). In all other cases, an approximation method should be used in order to obtain a scalable algorithm for real applications. Two classes of methods are proposed in literature:

i) Numerical approximation methods such as MCMC which compute numerically the desired MAP or PM estimators, and

ii) Analytical approximation methods which try, in a first step provide an analytical simpler approximation $q(\boldsymbol{x})$ for $p(\boldsymbol{x}|\lambda)$ and then use it to do the necessary computations.

In this work, we propose to use the second approach. This paper is then organized as follows, In the next section, first we present the particular cases of Markov models we will use in practical imaging systems. In section 3, the basic ideas of Variational Bayesian (VB) and Mean Fields (MF) approximation methods are presented. In section 4 and 5, we drive the details of these approximations for the proposed markovian models. In section 6, we will show some simulation results. In section 7, we describe the main domain of application of this paper which is to provide algorithms to estimate the parameter λ of the proposed models and compare their relative performances with their estimates without approximations.

2. PROPOSED MARKOVIAN MODELS

In this paper, we consider a particular case where

$$\mathcal{E}(\boldsymbol{x}) = \sum_i \sum_{j \in \mathcal{V}i} \Phi_i(x_i, x_j),$$

where

$$\Phi_i(x_i, x_j) = \Phi(x_i - x_j)\ \forall i. \quad (12)$$

With this notation we have

$$p(\boldsymbol{x}|\lambda) = \prod_i p\left(x_i | x_j, j \in \mathcal{V}(i)\right)$$

with

$$p(x_i | x_j, \in \mathcal{V}) \propto \exp\left(-\lambda \sum_{j \in v(i)} \Phi(x_i - x_j)\right) \quad (13)$$

We discuss in this paper three different Markov models that covers a broad spectrum of applications.

2.1. Generalized Gaussian Markov models

This model is used to account for mutual dependence for continuous variable with convex potential function [8, 9]. Its associated energy function is:

$$\mathcal{E}(\boldsymbol{x}) = \sum_i \sum_{j \in \mathcal{V}(i)} |x_i - x_j|^{\beta} \quad (14)$$

2.2. Entropic Markov models (I-Distribution family)

This family of probability distributions are based on I-divergence, well known in information theory community [10, 11, 12]. Since I-divergence is not symmetric, we can define two kind of distributions depending on the way we use the distance. Their associated energy functions are as follows:

First kind I-distribution:

$$\mathcal{E}(\boldsymbol{x}) = \sum_i \sum_{j \in \mathcal{V}(i)} x_j \ln \frac{x_j}{x_i} - (x_j - x_i). \tag{15}$$

Second kind I-distribution:

$$\mathcal{E}(\boldsymbol{x}) = \sum_i \sum_{j \in \mathcal{V}(i)} x_i \ln \frac{x_i}{x_j} - (x_i - x_j). \tag{16}$$

2.3. Potts and Ising models

This model is well known in segmentation application, where it is used to account for the spatial dependence between elements of class discrete variables $x_j \in \{1, ..., K\}$ [5, 13]. Its associated energy function is given by:

$$\mathcal{E}(\boldsymbol{x}) = -\sum_i \sum_{j \in \mathcal{V}} \delta(x_i - x_j), \tag{17}$$

where δ is the Kronecker function. In a particular case, Ising model is defined for $x_j \in \{0, 1\}$.

3. BASICS OF VBA AND MFA

As we mentionned in the introduction, using directly these markovian models, for any estimation or inference, is, in general, intractable or needs high computational cost, because $Z(\lambda)$ has not an explicit form except for a few simple cases, for example where $\mathcal{E}$ quadratic (Gauss-Markov models). In all other cases, an approximation method should be used in order to obtain a scalable algorithm for real applications.

Two classes of methods are proposed in literature:
i) Numerical approximation methods such as MCMC which compute numerically the desired MAP or PM estimators, and
ii) Analytical approximation methods which try, in a first step to provide an analytical simpler approximation $q(\boldsymbol{x})$ for $p(\boldsymbol{x}|\lambda)$ and then use it to do the necessary computations.

In this work, we propose to use the second approach. This consists in, choosing an appropriate class of probability laws $\mathcal{Q}$ and a distance measure (KL or any other divergence measure) and find an approximate probability law $q(\boldsymbol{x})$ in that class which minimizes that distance or divergence measure:

$$q(\boldsymbol{x}) = \arg\min_{q \in \mathcal{Q}} KL(q : p). \tag{18}$$

Then, this simpler probability law $q(\boldsymbol{x})$ can be used to do any bayesian computation. This general scheme is presented in Fig. 1.

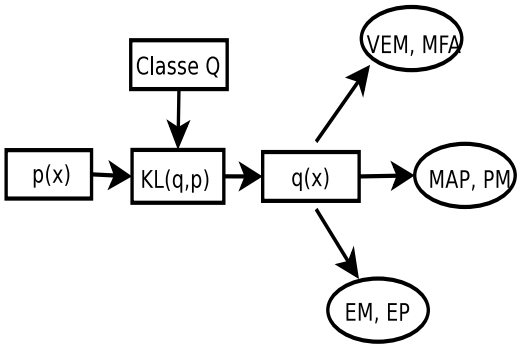


Fig. 1. Minimizing KL leads to different Variational Bayesian Approximation (VBA) inference algorithms: Maximum A posteriori (MAP), Posterior Mean (PM), Expectation-Maximization (EM), Expectation-Propagation (EP), Variational EM, Mean Field Approximation (MFA), etc.

In equation (18), the KL divergence is given by:

$$\begin{aligned} \mathrm{KL}(q : p) &= \int q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x}|\lambda)} \, \mathrm{d}\boldsymbol{x} \\ &= \langle \ln q(\boldsymbol{x}) \rangle_{q(\boldsymbol{x})} - \langle \ln p(\boldsymbol{x}|\lambda) \rangle_{q(\boldsymbol{x})} \\ &= \sum_j \langle \ln q(x_j) \rangle_{q(x_j)} - \langle \ln p(\boldsymbol{x}|\lambda) \rangle_{q(\boldsymbol{x})}, \end{aligned} \tag{19}$$

where $\langle z \rangle_q$ is the expectation of z w.r.t q. So the problem is to find the optimal value of q that minimises $\mathrm{KL}(q : p)$ under the constraint that q is normalised.

In this work, we propose the estimation of Markov model with its parameter in an unsupervised Bayesian framework using the following methods:

a) Variational Bayes Approximation (VBA), where we look for a free form separable $q(\boldsymbol{x}) = \prod_j q(x_j)$ approximating distribution which minimizes $\mathrm{KL}(q : p)$ with respect to $q \in \mathcal{Q} = \{q \ : \ \int q = 1\}$. The expression of $q_i(x_i)$ in this case is given by

$$q_i(x_i) = \frac{1}{Z_i(\lambda)} \exp\left(-\lambda \left\langle \mathcal{E}(\boldsymbol{x}) \right\rangle_{\prod_{j \neq i} q_j}\right) \tag{20}$$

which needs the expression of $\left\langle \mathcal{E}(\boldsymbol{x}) \right\rangle_{\prod_{j \neq i} q_j}$.

b) Mean Field Approximation (MFA), where we impose the form of the approximating distribution to be in a particular parametric family. For example, when $p(\boldsymbol{x}|\lambda) = \prod_i p\left(x_i | x_j, j \in \mathcal{V}(i)\right)$ with

$$p(x_i | x_j, \in \mathcal{V}) = \frac{1}{Z_i(\lambda)} \exp\left(-\lambda \sum_{j \in v(i)} \Phi(x_i - x_j)\right), \tag{21}$$

we choose

$$q(\boldsymbol{x}|\lambda) = \prod_j p\left(x_i|\bar{x}_j, j \in \mathcal{V}(i)\right), \tag{22}$$

where $\bar{x}_j$ becomes a parameter to be estimated such that $\mathrm{KL}(q:p)$ be minimized.

As we may note that the second solution is a suboptimal one compared to the first one. However, as we will see, it is interesting to compare the relative complexities of these two approximations.

In both cases, the main idea is to use $q(\boldsymbol{x})$ in place of $p(\boldsymbol{x})$ to make inference, for example to estimate the parameter λ. In the following, for each of the proposed markovian models in previous section, first, we give the expression of obtained $q(\boldsymbol{x})$, or more precisely $q(x_j)$. Then, we give the equation to be solved to obtain the parameter λ.

4. VARIATIONAL BAYES APPROACH (VBA)

4.1. Generalized Gaussian Markov models

Unfortunately, for this general case, we could not yet obtain a usable solution. However, the special case of $\beta = 2$ is easy and we use it just to show the way to follow to obtain the VB approximation.

4.1.1. Gaussian case $\beta = 2$

This is the simplest case since the partition function can be found in an explicit way. However, it is always interesting to compare the result of approximation for this case. From (eq.20), we can write,

$$\begin{aligned}
\ln\left(q_i(x_i)\right) &\propto -\frac{\lambda}{2}\sum_{j\in\mathcal{V}(i)}\int_{x_j}\left(x_i - x_j\right)^2 q_j(x_j)\,\mathrm{d}x_j \\
&\propto -\frac{\lambda}{2}\sum_{j\in\mathcal{V}(i)}\left[x_i^2 - 2x_i\tilde{\mu}_j + \tilde{\mu}_j^2 + \tilde{v}_j\right] \\
\Rightarrow q_i(x_i) &= \mathcal{N}(\tilde{\mu}_i, \tilde{v}_i) \\
\text{with}\quad \tilde{\mu}_i &= \frac{1}{|\mathcal{V}|}\sum_{j\in\mathcal{V}(i)}\tilde{\mu}_j \ \text{ and } \ \tilde{v}_i = \frac{1}{|\mathcal{V}|\lambda}.
\end{aligned} \tag{23}$$

The bayesian estimation of λ can be done easily by assigning a Gamma distribution for the prior (i.e. $\pi(\lambda) = \Gamma(a, b)$), and by using conjugacy, the posterior is given by

$$p(\lambda|\boldsymbol{x}) = \Gamma(\hat{a}, \hat{b}) \tag{24}$$

where $\hat{a} = \left[\frac{1}{a} + \sum_{i\in\mathcal{R}}(x_i - \tilde{\mu}_i)^2\right]^{-1}$ and $\hat{b} = \frac{|\mathcal{R}|}{2} + b$.

4.2. Entropic Markov distribution

4.2.1. First kind I-distribution

Here, the computations can be done easily and we obtain:

$$\begin{aligned}
\ln(q(x_i)) &\propto -\lambda\sum_{j\in\mathcal{V}(i)}\int\left[x_j\ln\tfrac{x_j}{x_i} - (x_j - x_i)\right]q(x_j)\,\mathrm{d}x_j \\
&\propto -\lambda\sum_{j\in\mathcal{V}(i)}\langle x_j\rangle_{q_j}\ln\frac{\langle x_j\rangle_{q_j}}{x_i} - x_i + \langle x_j\rangle_{q_j}
\end{aligned}$$

This is interesting since we stay in the same family and the dependence on neighbours transforms to a mean value one. Moreover, moments of this distribution are easily accessible. The partition function can be obtained as:

$$Z_i(\lambda) = e^{-\lambda\tilde{\mu}_i\ln\tilde{\mu}_i + \lambda\tilde{\mu}_i}\lambda^{-\lambda\tilde{\mu}_i - 1}\Gamma(\lambda\tilde{\mu}_i + 1) \tag{25}$$

with $\tilde{\mu}_i = \sum_{j\in\mathcal{V}(i)}\langle x_j\rangle_{q_j}$.

4.2.2. Second kind I-distribution

In a similar way, we obtain:

$$\begin{aligned}
\ln(q(x_i)) &\propto -\lambda\sum_{j\in\mathcal{V}(i)}\int\left[x_i\ln\tfrac{x_i}{x_j} - (x_i - x_j)\right]q(x_j)\,\mathrm{d}x_j \\
&\propto -\lambda\sum_{j\in\mathcal{V}(i)} x_i\ln\frac{x_i}{e^{-\langle\ln x_j\rangle_{q_j}}} + x_i - \langle x_j\rangle_{q_j}
\end{aligned}$$

Here, the dependency is on the logarithmic moment $b_i = \exp\left[-\langle\ln x_j\rangle_{q_j}\right], \forall x_j \in \mathcal{V}(i)$ w.r.t the approximating distribution q_j which we are able to write it analytically.

For the partition function, we obtain

$$Z_i(\lambda) = \frac{e^{-\lambda\tilde{\mu}_i}}{\lambda}\alpha(\lambda\tilde{\mu}_i) \tag{26}$$

with:
$\tilde{\mu}_i = e^{-\sum_{j\in\mathcal{V}(i)}\langle\ln x_j\rangle_{q_j}}$ and $\alpha(a) = \int_0^\infty a^y e^{-y\ln y + y}\,\mathrm{d}y$.

4.3. Potts and Ising models

Here, the expression can be obtained in a straightforward way and is given by:

$$q_i(x_i = k) = \frac{\exp\left(-\lambda\sum_{j\in\mathcal{V}(i)} q_j(x_j = k)\right)}{\sum_{k=1}^K \mathrm{e}^{-\lambda\sum_{j\in\mathcal{V}(i)} q_j(x_j = k)}} \tag{27}$$

$x_j \in \{1, ..., K\}$ is the general case and $x_j \in \{0, 1\}$ is the Ising model case.

5. MEAN FIELD APPROXIMATION (MFA)

Interestingly, for the Potts model and the Entropic models, the VBA and MFA will give the same results. This is not the case for Generalized Gaussian case where we have started to do, but not yet obtained really usable results.

6. SIMULATIONS

Here, we will give some comparison results of computational cost and performances of the proposed approximation methods with respect to the optimal method of the estimation of the parameters λ for the proposed Markov models. The main protocol is, for each class of models, choose a *true* value of λ, generate samples from the models, and then

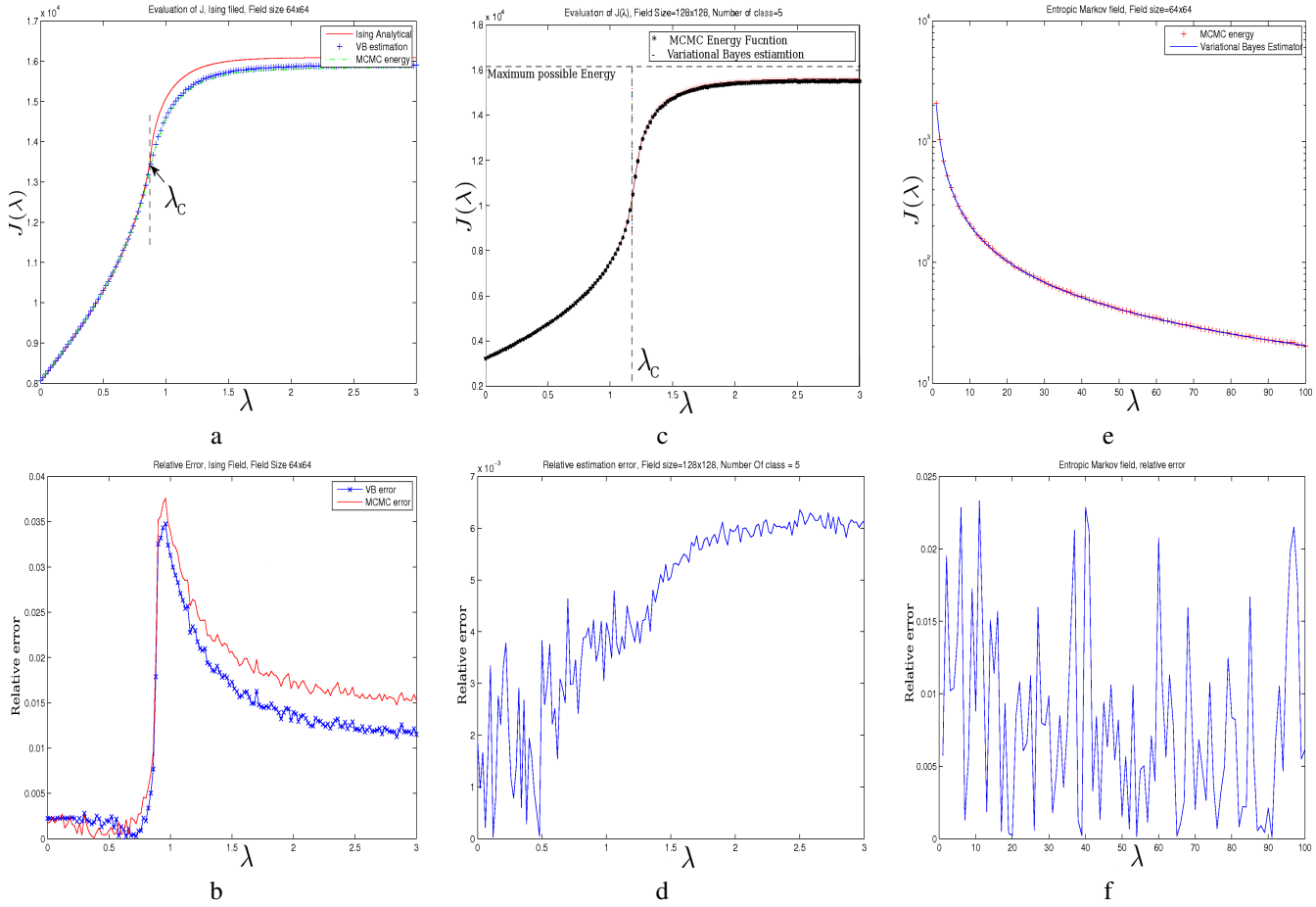


Fig. 2. Comparison between different simulation results for different field types: Ising, Potts, entropic: a)J vs λ for analytical, VB, MCMC b)relative error between analytical function and VBA or MCMC vs λ c)J vs λ for VBA and MCMC for 5 class Potts field d) relative error between VBA and MCMC vs λ for Potts field e) J vs λ for entropic field f) relative error between VBA and MCMC vs λ

estimate this parameter, either in an optimal way or by VBA or by MCMC, and finaly, compare these results.

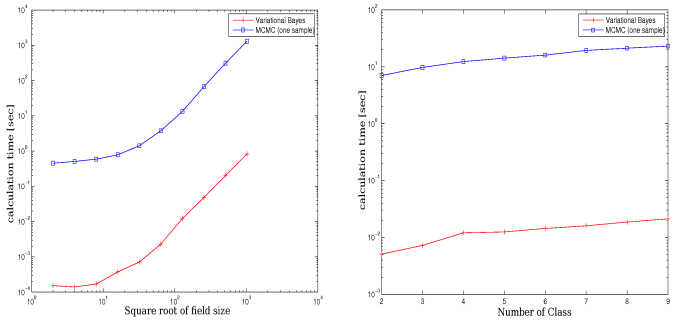
For this purpose, we evaluate $J(\lambda) = \frac{-\partial \ln Z(\lambda)}{\partial \lambda}$, since the optimal value of λ using (eq.9) for the value corresponding to field energy $\mathcal{E}(\boldsymbol{x})$. We first compare the true value of this function for the case of the Ising model, where we have an analytical expression of partition function [14, 15], with the VB solution. We note that for $\lambda < \lambda_C \approx 0.88$, estimation error is very small and it becomes important for higher values. Looking at the energy of generated samples, we find the same error with analytical solution, which suggest that the generated samples did not converge yet. However, we should revise our Gibbs sampling process, known for its slow convergence for high values of λ. Figure(2) shows the results for Ising model.

For higher dimension of Potts field, where the calculation of true function is not possible in a reasonable time, we compare the VB function with an approximated value based on an MCMC method: For each λ we generate a number of samples that we use their energy function to approximate $J(\lambda)$ using (eq.9). The results are encouraging: the error of estimation is very small for small values of λ. This is very important since better estimation is needed for λ around its critical value, and for higher values estimation error of λ becomes less significant since the energy change is less important.

We have performed the same study for the case of entropic Markov model and again we have a very good estimation results over the whole axe of λ.

For computational cost comparison, the first motivation of our work, we studied calculus time for the variational method and the needed time for one sample of Potts field using Gibbs sampler. We have excluded the comparison with true function since the complexity is $O(K^S)$ with K is the

number of classes and S the size of the field. To give an idea this complexity, we need to evaluate an exponential function $\approx 3.4 \times 10^{30}$ times. Interestingly our method needed way less time than the sampling method.



a b

Fig. 3. Comparison of computational costs for VB and MCMC, a)calculation time for different field dimensions b) calculation time for different number of classes

7. APPLICATIONS

The main application of these results will be:
a) in image segmentation where a Hidden Potts model is used to model the image;
b) in image denoising, restoration and tomographic image reconstruction inverse problems, where a hierarchical Gauss-Markov-Potts model is used to model the unknown images.

8. CONCLUSION

We considered the problem of parameter estimation of Markovian models where the exact computation of the partition function is not possible or computationally too expensive. The main idea is to approximate the expression of the likelihood by a simpler one where we can either have analytical expression or compute it more efficiently. We considered two approaches: Variational Bayes Approximation (VBA) and Mean Field Approximation (MFA) and studied the properties of such approximations. We studied the relative performances of these approximations in two aspects: computational cost and estimation error as a function of the size of the images.

9. ACKNOWLEDGEMENT

The authors would like to thank Jean-Franois Giovannelli for helpful discussions on Ising model.

10. REFERENCES

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, November 1984.

[2] S. Brette, J. Idier, and A. Mohammad-Djafari, "Scale invariant Markov models for Bayesian inversion of linear inverse problems," Maximum Entropy and Bayesian Methods, (Cambridge, UK), pp. 199–212, Kluwer Academic Publ., 1994.

[3] X. Descombes, *Champs Markoviens en analyse d'image*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, Paris, France, December 1993.

[4] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise gaussian mrf," *IEEE Trans. on Image Processing*, vol. 7, pp. 571–585, April 1998.

[5] A. Mohammad-Djafari, "Gauss-markov-potts priors for images in computer tomography resulting to joint optimal reconstruction and segmentation," *International Journal of Tomography & Statistics*, vol. 11, no. W09, pp. 76–92, 2008.

[6] H. Ayasso and A. Mohammad-Djafari, "Variational bayes with gauss-markov-potts prior models for joint image restoration and segmentation," Visapp Proceedings, (Funchal, Madaira, Portugal), Int. Conf. on Computer Vision and Applications, 2008.

[7] K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny, "Variational free energy and the laplace approximation," *Neuroimage*, no. 2006.08.035, 2006. Available Online.

[8] C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, vol. 2, pp. 296–310, July 1993.

[9] M. Ichir and A. Mohammad-Djafari, "A mean field approximation approach to blind source separation with lp priors," in *Eusipco 2005, Antalya, Turkey, September 2005*, Eusipco 2005, Antalya, Turkey, September 2005, September 2005.

[10] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *AP*, vol. 3, no. 1, pp. 146–158, 1975.

[11] L. K. Jones and C. L. Byrne, "General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis," *IEEE Transactions on Information Theory*, vol. 36, pp. 23–30, January 1990.

[12] S. Brette, J. Idier, and A. Mohammad-Djafari, "Scale invariant Bayesian estimator for inversion of noisy linear system," in *Fifth Valencia Int. Meeting on Bayesian Statistics*, (Spain), June 1994.

[13] H. Ayasso and A. Mohammad-Djafari, "Joint image restoration and segmentation using gauss-markov-potts prior models and variational bayesian computation," *Submitted to IEEE Image Processing*, Febuary, 2009.

[14] J. Giovannelli, "Estimation of the Ising field parameter using the partition function", Research report, Bordeaux1 univercity, 2009.

[15] L. Onsager, "Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition," Physical Review, vol. 65, Fab. 1944, p. 117.