USC–SIPI REPORT #285

Approximate Maximum Likelihood Hyperparameter Estimation for Gibbs Priors

by

Zhenyu Zhou and Richard M. Leahy

June 1995

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

Approximate Maximum Likelihood Hyperparameter Estimation for Gibbs Priors*

Zhenyu Zhou, *Student Member, IEEE*, and Richard M. Leahy, *Member, IEEE*
Signal and Image Processing Institute,
University of Southern California,
Los Angeles, CA90089-2564.

Abstract

The parameters of the prior, the *hyperparameters*, play a critical role in Bayesian image estimation. Of particular importance for the case of Gibbs priors is the global hyperparameter, β, which multiplies the Hamiltonian. Here we consider maximum likelihood (ML) estimation of β from *incomplete* data, i.e. problems in which the image, which is drawn from the Gibbs distribution, is observed indirectly through some degradation or blurring process. Important applications include image restoration and image reconstruction from projections. Exact ML estimation of β from incomplete data is intractable for most image processing. Here we present an approximate ML estimator which is computed simultaneously with a maximum *a posteriori* (MAP) image estimate. The algorithm is based on a mean field approximation technique through which multidimensional Gibbs distributions are approximated by a separable function equal to a product of one dimensional densities. We show how this approach can be used to simplify the ML estimation problem. We also show how the Gibbs-Bogoliubov-Feynman bound can be used to optimize the approximation for a restricted class of problems. We present the results of a Monte-Carlo study that examines the bias and variance of this estimator when applied to image restoration.

1 Introduction

Bayesian approaches to inverse problems in image processing typically involve computing a point estimate of an unknown image $\boldsymbol{x} \in \mathcal{X}$ from a set of data $\boldsymbol{y} \in \mathcal{Y}$. We assume that the two quantities

*This work was supported by the National Cancer Institute under Grant No. R01-CA59794 and by the National Institute of Mental Health under Grant No. R01-MH53213.

are related by a known conditional probability, $P(\boldsymbol{y}|\boldsymbol{x})$. This conditional probability or likelihood function is dependent on the imaging modality and is problem-specific. The estimate of $\boldsymbol{x}$ is computed as a function of the posterior density $P(\boldsymbol{x}|\boldsymbol{y})$, which requires the specification of a prior density $P(\boldsymbol{x})$ in addition to the likelihood function. In the context of Bayesian image estimation, the parameters of the prior are referred to as *hyperparameters*. In this paper we address the problem of estimating these parameters in the case where it is not possible to observe the true image $\boldsymbol{x}$ directly. We describe a practical method for estimating hyperparameters from observations of the data $\boldsymbol{y} \in \mathcal{Y}$. We begin by briefly reviewing two models for image restoration and reconstruction problems for which this method is applicable.

One of the most widely addressed models in image restoration and reconstruction is the linear Gaussian model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{y}$ is the observed data, $\boldsymbol{x}$ is the underlying image, $\boldsymbol{n}$ is zero-mean Gaussian noise with covariance matrix $\boldsymbol{C}$ and matrix $\boldsymbol{A}$ is a linear degradation operator. Then,

$$P(\boldsymbol{y}|\boldsymbol{x}) = (2\pi)^{-N/2}|\boldsymbol{C}|^{-N/2} \exp\left[-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^T \boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})\right]. \tag{1}$$

A second common model is the linear Poisson model which arises in problems where the data acquisition system is photon limited, e.g. emission tomography, gamma-ray astronomy, and fluorescence microscopy. In this model, the mean of $\boldsymbol{y}$ is related to the image $\boldsymbol{x}$ by a linear operator $\boldsymbol{A}$, i.e., $E[\boldsymbol{y}] = \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{y}$ follows a joint Poisson distribution:

$$P(\boldsymbol{y}|\boldsymbol{x}) = \prod_i \frac{(\sum_j A_{ij}x_j)^{y_i}}{(y_i)!} \exp\left(-\sum_j A_{ij}x_j\right). \tag{2}$$

The objective of the inverse problems of interest here, is to obtain a point estimate of $\boldsymbol{x}$ from the observation $\boldsymbol{y}$. Since $\boldsymbol{A}$ is often ill-conditioned, direct inversion based on maximizing the likelihood function does not always provide a unique and stable solution. Bayesian methods solve this type of ill-posed inverse problem by combining information contained in the observed data with prior information concerning the relative probabilities of possible solutions. The unknown image can then be estimated by maximizing over the posterior density $P(\boldsymbol{x}|\boldsymbol{y})$ to form a *maximum a posteriori* (MAP) estimate. The posterior density is proportional to the product of the likelihood function $P(\boldsymbol{y}|\boldsymbol{x})$, and a prior on the image, $P(\boldsymbol{x}|\beta)$. Usually the prior reflects an expectation that images are locally smooth. Markov random fields (MRFs) [3, 9, 22] have been widely used to model local smoothness in images and will be the class of priors considered here. The joint density for the MRF

has the form of a Gibbs distribution:

$$P(\boldsymbol{x}|\beta) = \frac{1}{Z}\exp\{-\beta U(\boldsymbol{x})\} \tag{3}$$

where $U(\boldsymbol{x})$ is the Gibbs energy function, Z is the partition function, and β is the global hyperparameter.

Image estimation using MRF priors has proven to be a powerful approach to restoration and reconstruction of high quality images. However, a major problem limiting its utility is the lack of a practical and robust method for selecting the parameters of the prior. Of particular importance for the case of homogeneous isotropic MRFs is the global hyperparameter β which multiplies the Gibbs energy function. MAP estimates of the image $\boldsymbol{x}$ are clearly functions of β, which plays a critical role by controlling the balance of influence of the Gibbs prior and that of the likelihood. If β is too large, the prior will tend to have an over-smoothing effect on the solution. Conversely, if it is too small, the MAP estimate may be unstable, reducing to the ML solution as β goes to zero.

To illustrate the effect of the hyperparameter on the MAP estimate, we show two curves in Figure1 computed for a typical application to image restoration. In Figure 1(a), we have a typical L-curve [13, 35], which is a plot of the value of the Gibbs energy $U(\boldsymbol{x})$ versus the log-likelihood $lnP(\boldsymbol{y}|\boldsymbol{x})$, computed at the MAP solution for a range of values of β. We observe two characteristic parts on the curve, namely a "flat" part where the MAP solution is dominated by the prior, and an almost "vertical" part, where the solution is dominated by the likelihood function. Heuristically, the region between these two characteristic parts, i.e. the "corner", corresponds to a good balance between fidelity to the data and smoothness of the solution. Figure 1(b) shows a curve of the squared error in the MAP estimate for a range of β values. Here it is clear that an appropriate choice of β is necessary to achieve a small error. Furthermore, we have observed that the corner of the L-curve corresponds to values of the hyperparameter β that are close to that which minimizes the squared error for the MAP estimation problem described here (both points are indicated by $*$). Similar observations were made in [13] concerning the more general regularization problem.

A truly Bayesian formulation requires either that the hyperparameters are known or that we specify a "hyperprior" density. However, in practice the hyperparameters are often unknown because the true images can never be observed directly, and little evidence exists to justify an informative hyperprior density. Even if β is known, problems can arise if there is an unknown gain factor in the transfer function $\boldsymbol{A}$ in (2) or an unknown noise variance in (1). These problems can

be avoided if the hyperparameters are estimated directly from the the observed data. Data-driven selection of the hyperparameter is often performed in an *ad hoc* fashion through visual inspection of the resulting images. There are two basic approaches for choosing β in a more principled manner: (i) treating β as a regularization parameter and applying techniques such as generalized cross validation, the L-curve, and χ^2 goodness of fit tests; (ii) estimation theoretic approaches such as maximum likelihood (ML).

The generalized cross-validation (GCV) method [8] has been applied in Bayesian image restoration and reconstruction [16]. Several difficulties are associated with this method: the GCV function is often very flat and its minimum is difficult to locate numerically [30]. Also the method may fail to select the correct hyperparameter when measurement noise is highly correlated [31]. For problems of large dimensionality, this method may be impractical due to the amount of computation required. Hansen and Leary's L-curve is based on the empirical observation that the corner of the curve, illustrated in Figure 1(a), corresponds to a good choice of β in terms of other validation measures [13]. The L-curve has similar performance to GCV for uncorrelated measurement errors, however, the L-curve criterion is also able to work, under certain restrictions, for correlated errors [13]. We have used the L-curve to select the hyperparameter in MAP image reconstruction [35]. The corner of the L-curve is difficult to find without multiple evaluations of the MAP solution for different hyperparameter values. Thus the computation cost is again very high. χ^2 statistics have been widely used to choose the regularization parameter [29]. For MAP image estimation, Hebert *et al* [15] developed an adaptive scheme based on a χ^2 statistic to select β. Since the image is estimated from the data, the degrees of freedom of the test should be reduced accordingly. This presents a problem when the data and image are of similar dimension. It is also commonly agreed that χ^2 methods tend to over-smooth the solution [29].

As an alternative to the regularization based methods discussed above, a well grounded approach to selection of the hyperparameter is to apply ML estimation. The image $\boldsymbol{x}$, which is drawn from the *complete data* sample space $\mathcal{X}$ characterized by the parameter β, is not observed directly. Instead, we observe a second process $\boldsymbol{y}$ which is drawn from the *incomplete data* sample space $\mathcal{Y}$. The ML estimate of the hyperparameter corresponds to the maximizer of the incomplete data likelihood function $P(\boldsymbol{y}|\beta)$, which is found by marginalization of the joint probability density for the complete and incomplete data, $P(\boldsymbol{x}, \boldsymbol{y}|\beta)$, over the complete data sample space. Selection of the hyperparameter can therefore be viewed as a ML estimation problem in an incomplete/complete data framework and is a natural candidate for the EM algorithm [7]. However, in most imaging

applications, the high dimensionality of the densities involved make the EM approach impractical. Geman and McClure [10] propose using a stochastic relaxation technique, such as a Gibbs sampler, to evaluate the E-step of the EM algorithm. While this approach provides a means of overcoming the intractability of the true EM algorithm, the computational cost remains extremely high. Other estimation methods have been studied which do not share the desirable properties of true MLE but have much lower computational cost. Several generalized maximum likelihood approaches have been described [3, 17, 24] that make the simplifying approximation that the ML estimate of β and the MAP estimate of the image $\boldsymbol{x}$ can be found simultaneously as the joint maximizers of the joint density of $\boldsymbol{x}$ and $\boldsymbol{y}$. This approach works well in some situations, but the crudeness of the approximation results in poor performance in general. The method of moments (MOM) [10, 23] defines a statistical moment of the incomplete data that is ideally chosen to reflect the variability in the unobserved image and to establish a one-to-one correspondence between the moment value and the global hyperparameter. Initial computational costs for this method are very large, but the moment vs. hyperparameter curve is independent of the observed data and can be computed off-line. For each new data set the hyperparameter is determined by simply comparing the computed statistic with the precomputed curve. The major limitation in using this method is in finding a statistic with sufficient slope that the hyperparameter can be reliably determined. In practice it has been observed that the method performs well only for relatively small values of β [23]. Finally, a variational method is described in [1]. This approach leads to a procedure similar to, but simpler than, the EM algorithm. However, the computational cost remains high, and few validation or experimental results have been published for this method.

Here we return to the ML approach but develop an approximation that results in a reasonable computational cost. The major difficulty in computing a true ML estimate of β is in evaluating the multi-dimensional integrals of the highly coupled joint density function $P(\boldsymbol{x}, \boldsymbol{y}|\beta)$ over the complete data sample space $\mathcal{X}$. For a Gibbs prior and the Gaussian or Poisson likelihood functions defined in (1) and (2) respectively, the posterior density $P(\boldsymbol{x}|\boldsymbol{y}, \beta)$ can be written as a Gibbs distribution, albeit with a much larger neighborhood than the prior. We approximate this Gibbs distribution with a simple and separable density so that the multidimensional integral to compute the marginal density becomes a product of one dimensional integrals (one per pixel). This approximation renders the ML approach tractable. The approximation is closely related to the mean field approximation of statistical mechanics. In the mean field approach, the separable approximation is achieved by replacing the statistical influence of the neighbors of each pixel with their estimated means. In our

work, we use a mode-field rather than a mean-field approximation, where the mode of the posterior density is computed using a MAP image estimation algorithm. We use a sequential updating scheme to estimate both the image and the hyperparameter. Successive iterates of a MAP image estimation algorithm are substituted in the mode-field approximation, which in turn is used to update the hyperparameter estimate.

We present a brief summary of the problem formulation for ML estimation of the hyperparameter from incomplete data in Section 2. We then describe our mean/mode field approach to parameter estimation in Section 3. We also describe how an optimal approximation can be found in special cases using the Gibbs-Bogoliubov-Feynman bound [6]. We then describe the application of this method to the problem of image restoration from Gaussian or Poisson data in Section 4. Finally, in Section 5, we present the results of extensive Monte-Carlo studies that examine the bias and variance of this estimator for cases where the true value of β is known.

2 Background

2.1 Gibbs Priors for Image Estimation

We will assume a homogeneous isotropic MRF model for the image, $\boldsymbol{x}$, characterized by the Gibbs distribution:

$$P(\boldsymbol{x}|\beta) = \frac{1}{Z(\beta)} \exp\{-\beta U(\boldsymbol{x})\} \tag{4}$$

with Gibbs energy $U(\boldsymbol{x})$, partition function $Z(\beta)$ and global hyperparameter β. The partition function is the scaling constant:

$$Z(\beta) = \int_{\mathcal{X}} \exp\{-\beta U(\boldsymbol{x})\} d\boldsymbol{x}. \tag{5}$$

Here we indicate explicitly that partition function is dependent on the hyperparameter β.

We restrict the Gibbs energy to pairwise interactions on a second order (eight nearest neighbor) system:

$$U(\boldsymbol{x}) = \sum_i \sum_{j>i, j\in N_i} \kappa_{ij} V(x_i, x_j), \tag{6}$$

where N_i denotes the set of eight nearest neighbors of pixel i with κ_{ij} unity for horizontal and vertical neighbors and $1/\sqrt{2}$ for diagonal neighbors. The image sample space is $\mathcal{X} = [0, x_{max}]^N$ where N is the total number of pixels in the image.

The defining feature of a MRF is $P(x_i|x_j, \forall j \neq i) = P(x_i|x_j, j \in N_i)$. For the pairwise interaction model above, the conditional density has the special form [22]:

$$P(x_i|x_j, j \in N_i) = \frac{\exp\{-\beta \sum_{j \in N_i} \kappa_{ij} V(x_i, x_j)\}}{\int_{x_i} \exp\{-\beta \sum_{j \in N_i} \kappa_{ij} V(x_i, x_j)\} dx_i}. \tag{7}$$

The specific potential functions, $V(\cdot, \cdot)$, used in this work are as follows:

$$\begin{aligned}
&\text{Quadratic function:} && V_1(x_i, x_j) = (x_i - x_j)^2 \\
&\text{Huber function:} && V_2(x_i, x_j) = \begin{cases} \frac{1}{2\delta}(x_i - x_j)^2, & \text{if } |x_i - x_j| \leq \delta \\ |x_i - x_j| - \frac{\delta}{2}, & \text{otherwise} \end{cases} \\
&\text{Log-quadratic:} && V_3(x_i, x_j) = \ln(1 + \frac{(x_i - x_j)^2}{\delta^2}) \\
&\text{Saturated-quadratic:} && V_4(x_i, x_j) = \frac{(x_i - x_j)^2}{\delta^2 + (x_i - x_j)^2}.
\end{aligned} \tag{8}$$

These four functions are representatives of three major categories for image priors: strictly convex (V_1), semi-convex (V_2), and non-convex (V_3, V_4) potential functions. They are illustrated in Figure 2. All four can be used to model smooth images. However, the quadratic function $V_1(\cdot)$ penalizes the differences of neighboring pixels at an increasing rate, which tends to force the image to be smooth everywhere. The Huber function $V_2(\cdot)$ behaves as a quadratic function when the difference of the neighboring pairs are small, but applies a linear penalty when the differences are large, i.e. the rate of the penalty applied to intensity differences does not change beyond the threshold δ. Therefore, this prior does not differentiate substantially between slow monotonic changes and abrupt changes and consequently does not penalize the presence of edges or boundaries in the image. The function $V_3(\cdot)$ was introduced in [15] and $V_4(\cdot)$ in [10]. Both have saturating properties which actually decrease the rate of penalty applied to intensity differences beyond a threshold determined by δ. Consequently they positively favor the presence of edges in the image. However, $V_3(\cdot)$ and $V_4(\cdot)$ are nonconvex which presents difficulties in computing global MAP estimates.

2.2 Maximum Likelihood Hyperparameter Estimation from Incomplete Data

Given the observed incomplete data, $\boldsymbol{y}$, a maximum likelihood estimate of β can be found from the maximizer of the marginalized likelihood function [7]:

$$\begin{aligned} P(\boldsymbol{y}|\beta) &= \int_{\mathcal{X}} P(\boldsymbol{y},\boldsymbol{x}|\beta)d\boldsymbol{x} = \int_{\mathcal{X}} P(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x}|\beta)d\boldsymbol{x} \\ &= \frac{\int_{\mathcal{X}} \exp\{\ln P(\boldsymbol{y}|\boldsymbol{x}) - \beta U(\boldsymbol{x})\}d\boldsymbol{x}}{\int_{\mathcal{X}} \exp\{-\beta U(\boldsymbol{x})\}d\boldsymbol{x}} = \frac{Z(\boldsymbol{y},\beta)}{Z(\beta)}, \end{aligned} \tag{9}$$

where

$$Z(\boldsymbol{y},\beta) = \int_{\mathcal{X}} \exp\{\ln P(\boldsymbol{y}|\boldsymbol{x}) - \beta U(\boldsymbol{x})\}d\boldsymbol{x} \tag{10}$$

is the partition function of the posterior density, $P(\boldsymbol{x}|\boldsymbol{y},\beta)$. Therefore:

$$\ln P(\boldsymbol{y}|\beta) = \ln Z(\boldsymbol{y},\beta) - \ln Z(\beta), \tag{11}$$

and the ML estimator of the hyperparameter is a root of the equation

$$0 = \frac{\partial \ln P(\boldsymbol{y}|\beta)}{\partial\beta} \leftrightarrow \frac{\partial \ln Z(\boldsymbol{y},\beta)}{\partial\beta} = \frac{\partial \ln Z(\beta)}{\partial\beta}. \tag{12}$$

It is straightforward to verify that

$$\frac{\partial \ln Z(\beta)}{\partial\beta} = -\frac{\int_{\mathcal{X}} U(\boldsymbol{x}) \exp\{-\beta U(\boldsymbol{x})\}d\boldsymbol{x}}{\int_{\mathcal{X}} \exp\{-\beta U(\boldsymbol{x})\}d\boldsymbol{x}} = -E[U(\boldsymbol{x})|\beta] \tag{13}$$

$$\frac{\partial \ln Z(\boldsymbol{y},\beta)}{\partial\beta} = -\frac{\int_{\mathcal{X}} U(\boldsymbol{x}) \exp\{\ln P(\boldsymbol{y}|\boldsymbol{x}) - \beta U(\boldsymbol{x})\}d\boldsymbol{x}}{\int_{\mathcal{X}} \exp\{\ln P(\boldsymbol{y}|\boldsymbol{x}) - \beta U(\boldsymbol{x})\}d\boldsymbol{x}} = -E[U(\boldsymbol{x})|\boldsymbol{y},\beta], \tag{14}$$

where $E[\cdot|\beta]$ and $E[\cdot|\boldsymbol{y},\beta]$ denote expectation with respect to the prior and posterior densities respectively. It follows from (12),(13) and (14) that the ML estimate of β from $\boldsymbol{y}$ is a root of the likelihood equation:

$$E[U(\boldsymbol{x})|\boldsymbol{y},\beta] = E[U(\boldsymbol{x})|\beta]. \tag{15}$$

This equation can in principle be solved using an EM algorithm [7, 10] as follows:

E-Step : Estimate the complete-data sufficient statistic $U(\boldsymbol{x})$ by finding $U^{(k)}(\boldsymbol{x}) = E[U(\boldsymbol{x})|\boldsymbol{y},\beta^{(k)}]$.

M-Step : Determine $\beta^{(k+1)}$ as the solution of the equation

$$E[U(\boldsymbol{x})|\beta] = U^{(k)}(\boldsymbol{x}).$$

Exact solution of this EM problem is impractical. Geman and McClure [10] note that a solution can be found using stochastic sampling from the posterior and prior densities to approximate the expectations. Due to the complexity of sampling from the posterior, the computation cost remains unacceptable. Therefore, in [10], a second estimation method is also described. This method of moments (MOM) simply requires the computation of a statistic $M(\boldsymbol{y})$ of the data. The parameter β is then chosen as a root of the equation:

$$M(\boldsymbol{y}) = E[M(\boldsymbol{y})|\beta] \tag{16}$$

where the moment curve $E[M(\boldsymbol{y})|\beta]$ is precomputed for a large range of β and should be monotonic with respect to β to ensure identifibility of the hyperparameter from the moment curve. Unfortunately, this method tends to perform poorly, at least for statistics which we have considered, due to small gradients in the moment curve which result in high variance estimates of β.

In [3, 17, 24] an alternative simplified approach is taken whereby, instead of maximizing with respect to β over the marginalized density (9), β is computed with $\boldsymbol{x}$ as the pair $\{\beta, \boldsymbol{x}\}$ that jointly maximize $P(\boldsymbol{y}, \boldsymbol{x}|\beta) = P(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x}|\beta)$:

$$\begin{aligned} \{\boldsymbol{x}, \beta\} &= \arg\max_{\boldsymbol{x},\beta} \ln P(\boldsymbol{y}, \boldsymbol{x}|\beta) \\ &= \arg\max_{\boldsymbol{x}} \{\ln P(\boldsymbol{y}|\boldsymbol{x}) + \max_{\beta} \ln P(\boldsymbol{x}|\beta)\}. \end{aligned} \tag{17}$$

Some authors term this the generalized ML (GML) method [24]. The optimization can be performed in a two-step algorithm,

$$\hat{\boldsymbol{x}}^{(k)} = \arg\max_{\boldsymbol{x}} P(\boldsymbol{y}, \boldsymbol{x}|\hat{\beta}^{(k)}) \tag{18}$$

$$\hat{\beta}^{(k+1)} = \arg\max_{\beta} P(\boldsymbol{x}^{(k)}|\beta). \tag{19}$$

Note that the first step is actually the MAP estimate of $\boldsymbol{x}$ given the current choice of β, and the second step is the maximum likelihood estimate of β using the current MAP estimate of $\boldsymbol{x}$ as a direct (complete data) observation of $\boldsymbol{x}$. It is straightforward to show that the second step is equivalent to solving the equation $U(\boldsymbol{x}^{(k)}) = E[U(\boldsymbol{x})|\beta]$. From the viewpoint of statistical mechanics, GML gives an approximate solution to the likelihood equation (15) which uses a saddle point approximation to

evaluate the posterior expectation $E[U(\boldsymbol{x})|\boldsymbol{y},\beta]$. Saddle point approximation neglects *all* statistical fluctuations in the field $\boldsymbol{x}$ and considers only the contribution of the maximum term to integrals with respect to a Gibbs distribution [11]. As we shall see below, the mean field approach is far less restrictive than that the saddle point approximation, which translates into significantly improved estimates of β when compared to GML.

In practice, direct computation of the GML estimate is still difficult as the second step requires evaluation of the partition function of the prior. This step is usually approximated using maximum pseudo-likelihood (MPL) [3, 17], i.e., we replace the second step with

$$\hat{\beta}^{(k+1)} = \arg\max_{\beta} \prod_i P(x_i|\hat{x}_j^k, j \in N_i, \beta). \tag{20}$$

We refer to this as the generalized maximum pseudo-likelihood (GMPL) method in the following.

3 Maximum Likelihood Hyperparameter Estimation Using Mean and Mode Field Approximation

True ML estimation of β is difficult because of the complexity and dimensionality of the joint density $P(\boldsymbol{y},\boldsymbol{x}|\beta)$. It is essentially impossible to compute the marginal density in (9) for each new data set $\boldsymbol{y}$. One approach to simplifying this problem is to approximate the multidimensional densities with separable joint density functions equal to a product of one dimensional probability densities. The multi-dimensional integrals involved in computing marginal densities, partition functions, or moments, can then be approximated with a product or sum of one dimensional integrals with respect to these one dimensional pixel-wise densities.

Approximating Gibbs distributions using separable joint density functions is the basis for the mean field theory in classical statistical mechanics [6]. The mean field theory was originally developed as a statistical mechanics tool for the analysis of many body systems through approximation as a set of single body systems. The basic idea is to focus on one particular particle (in our case a pixel site) in the system and assume that the role of the neighboring particles (pixels) can be approximated by an average field which acts on the tagged particle. This approach, therefore, neglects the effects of statistical fluctuations in all pixels other than current tagged one. The corresponding joint description is simply the product of that for each individual particle or pixel. Mean field approximation has previously been applied in the image processing field to surface reconstruc-

tion [11], image segmentation [32] and motion estimation [33]. However, we believe this is the first application of this approach to parameter estimation in image processing.

In this section, we focus first on a restricted class of Gibbs distribution for which we develop an optimal mean field approximation. We use the Gibbs-Bogoliubov-Feynman bound to select the mean field approximation which leads to an optimal approximation of the partition function. Using this result we describe an "optimal" ML hyperparameter estimator focusing on the problem of image restoration from Gaussian data with a quadratic Gibbs prior. Unfortunately, this optimized approximation is not applicable to the general problem. For the general case, we provide a heuristic development of an alternative approximation which can be applied to problems with Gibbs priors for any of the four potential functions in (8) with either the Gaussian or Poisson likelihood functions.

3.1 Optimal Approximation of the Partition Function

We can see from (12) that the true ML estimate of β is completely determined by the prior and posterior partition functions. Therefore, for the purposes of computing an accurate ML estimate of β, the mean field approximations of the prior and posterior Gibbs distributions should be chosen to give the best approximations of their respective partition functions. We begin by describing this optimization procedure for a restricted class of Gibbs distributions. We then apply this to approximation of the prior and posterior distributions to develop the mean field approximated ML estimator of β. The development below is based on that in [6] in several places. We emphasize that it is the application of this approximation to parameter estimation, rather than the approximation itself, that is novel.

The approximation involves replacing the true Gibbs distribution, $P(\boldsymbol{x})$, with a *mean field reference* distribution, $P_{MF}(\boldsymbol{x})$, which is a separable function in $\boldsymbol{x}$:

$$P(\boldsymbol{x}) \approx P_{MF}(\boldsymbol{x}) = \prod_i P_i^{mf}(x_i), \tag{21}$$

i.e. the pixels are modeled as independent random variables. The choice of the mean field reference is based on the following result:

Theorem 1 Gibbs-Bogoliubov-Feynman bound [6]

For a Gibbs distribution with partition function Z and Gibbs energy E, and any other Gibbs dis-

tribution with partition function Z_{MF} and Gibbs energy E_{MF}, we have the following inequality

$$Z \geq Z_{MF} \exp\{- < E - E_{MF} >_{MF}\} \tag{22}$$

where

$$< \ldots >_{MF} \stackrel{\text{def}}{=} Z_{MF}^{-1} \int_{\mathcal{X}} [\ldots] \exp(-E_{MF}) d\mathbf{x}. \tag{23}$$

Theorem 1 states that if we use *any* Gibbs distribution to approximate the partition function Z of the original Gibbs distribution, the quantity $Z_{MF} \exp\{- < E - E_{MF} >_{MF}\}$ will never exceed the original Z. Consequently, the mean field reference distribution which leads to the best approximation of the original partition function, can be found by maximizing the quantity on the right-side of the GBF bound.

Proposition 1 *The partition function Z can be best approximated through a mean field reference distribution with partition function Z_{MF} and Gibbs energy E_{MF} as*

$$Z \approx Z_{MF} \exp\{- < E - E_{MF} >_{MF}\} \tag{24}$$

where E_{MF} maximizes $Z_{MF} \exp\{- < E - E_{MF} >_{MF}\}$.

Unfortunately, a closed form solution to this optimization problem exists only for a restricted class of Gibbs distributions. This includes the class of continuous state *auto-models* [2], to which we now apply Proposition 1. The auto-models have the form $P(\boldsymbol{x}) = Z^{-1} \exp\{-E(\boldsymbol{x})\}$ where

$$E(\boldsymbol{x}) = \sum_i \left[x_i G_i(x_i) + \frac{1}{2} \sum_{j \in N_i} b_{ij} x_i x_j \right], \tag{25}$$

with $b_{ij} = b_{ji}$ and the single pixels sample space $x_i \in [0, x_{max}]$. The mean field reference $P_{MF}(\boldsymbol{x})$ is chosen in this case as a separable Gibbs distribution with mean field energy $E_{MF}(\boldsymbol{x})$ of the form:

$$E_{MF}(\boldsymbol{x}) = \sum_i x_i G_i(x_i) + \sum_i \Delta H_i x_i. \tag{26}$$

This reference distribution approximates the influence of neighboring pixels $\{x_j, j \in N_i\}$ by a constant ΔH_i. We now develop an optimal reference in the sense of choosing ΔH_i to maximize the right side of the GBF bound.

Since the reference field is separable, i.e. $P_{MF}(\boldsymbol{x}) = \prod_i P_i^{mf}(x_i)$, we consider first the local mean field reference density:

$$P_i^{mf}(x_i) = \frac{1}{Z_i^{mf}} \exp\{-(x_i G_i(x_i) + \Delta H_i x_i)\} \tag{27}$$

with

$$Z_i^{mf} = \int_{x_i} \exp\{-(x_i G_i(x_i) + \Delta H_i x_i)\} dx_i \tag{28}$$

the corresponding local mean field partition function. The joint mean field partition function Z_{MF} is then:

$$Z_{MF} = \prod_i Z_i^{mf}. \tag{29}$$

As a direct result of (27) and (28), the mean of the reference field is:

$$< x_i >^{mf} = \frac{1}{Z_i^{mf}} \int_{x_i} x_i \exp[-(x_i G_i(x_i) + \Delta H_i x_i)] dx_i = -\frac{\partial \ln Z_{MF}}{\partial \Delta H_i}, \quad \forall i. \tag{30}$$

Notice that since the reference mean field density is separable, $< x_i >^{mf}$ is a function of ΔH_i, while $< x_j >^{mf}, j \neq i$, is not. As we discussed above, the value of ΔH_i should be chosen to maximize the right side of the GBF bound. Thus ΔH_i must satisfy:

$$0 = \frac{\partial}{\partial \Delta H_i} \ln \left(Z_{MF} \exp\{- < E - E_{MF} >_{MF}\} \right) = \frac{\partial \ln Z_{MF}}{\partial \Delta H_i} - \frac{\partial}{\partial \Delta H_i} \{< E - E_{MF} >_{MF}\}, \quad \forall i. \tag{31}$$

We proceed with

$$< E - E_{MF} >_{MF} = \frac{1}{2} \sum_i \sum_{j \in N_i} b_{ij} < x_i >^{mf} < x_j >^{mf} - \sum_i \Delta H_i < x_i >^{mf} \tag{32}$$

where we again use the independence of pixels in the reference field to simplify $< x_i x_j >^{mf} = < x_i >^{mf} < x_j >^{mf}$ for $i \neq j$. By combining (30) and (32) in (31), noting that each pair $< x_i >^{mf} < x_j >^{mf}$ appears twice in the summation, and that $b_{ij} = b_{ji}$, we get

$$0 = - < x_i >^{mf} - \frac{\partial < x_i >^{mf}}{\partial \Delta H_i} \sum_{j \in N_i} b_{ij} < x_j >^{mf} + < x_i >^{mf} + \Delta H_i \frac{\partial < x_i >^{mf}}{\partial \Delta H_i}. \tag{33}$$

Solving this gives:

$$\Delta H_i = \sum_{j \in N_i} b_{ij} < x_j >^{mf} . \tag{34}$$

This is the value of the constant ΔH_i which maximizes the right side of the GBF bound over the set of separable Gibbs distribution with energies of the form of (26). Substituting gives

$$E_{MF}(\boldsymbol{x}) = \sum_i x_i G_i(x_i) + \sum_i \sum_{j \in N_i} b_{ij} < x_j >^{mf} x_i, \tag{35}$$

and

$$\begin{aligned} < E - E_{MF} >_{MF} &= < \frac{1}{2} \sum_i \sum_{j \in N_i} b_{ij} x_i x_j - \sum_i \sum_{j \in N_i} b_{ij} < x_j >^{mf} < x_i >^{mf} \\ &= -\frac{1}{2} \sum_i \sum_{j \in N_i} b_{ij} < x_i >^{mf} < x_j >^{mf} . \end{aligned} \tag{36}$$

The optimal approximation of the partition function Z is then

$$\begin{aligned} Z &\approx Z_{MF} \exp\{- < E - E_{MF} >_{MF}\} \\ &= \left(\prod_i^N \int_{x_i} \exp[-(x_i G_i(x_i) + x_i \sum_{j \in N_i} b_{ij} < x_j >^{mf})] dx_i \right) \left(\exp[\frac{1}{2} \sum_i \sum_{j \in N_i} b_{ij} < x_i >^{mf} < x_j >^{mf}] \right) . \end{aligned} \tag{37}$$

3.2 Hyperparameter Estimation using an Optimal Approximation

The optimal partition function approximation can be directly applied to the image restoration and reconstruction problems with the Gaussian likelihood function (1) and the quadratic Gibbs prior, $V_1(.,.)$ in (8). We can write the Gibbs energies of the prior and posterior densities, respectively, as:

$$\begin{aligned} E^{(PR)}(\boldsymbol{x}) &= \beta U(\boldsymbol{x}) = \beta \sum_i \sum_{j \in N_i^{PR}, j > i} \kappa_{ij} (x_i - x_j)^2 \\ &= \beta (\sum_i \sum_{j \in N_i^{PR}} \kappa_{ij} x_i^2 - \sum_i \sum_{j \in N_i^{PR}} \kappa_{ij} x_i x_j), \end{aligned} \tag{38}$$

and

$$\begin{aligned} E^{(PO)}(\boldsymbol{x}) &= -\ln P(\boldsymbol{y}|x) + \beta U(\boldsymbol{x}) \\ &= \frac{1}{2} (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^T \boldsymbol{C}^{-1} (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \beta \sum_i \sum_{j \in N_i^{PR}, j > i} \kappa_{ij} (x_i - x_j)^2 + K_1 \end{aligned} \tag{39}$$

$$= \sum_i F_i(x_i) - \sum_i \sum_{j \in N_i^{PO}} \gamma_{ij} x_i x_j + \beta \left(\sum_i \sum_{j \in N_i^{PR}} \kappa_{ij} x_i^2 - \sum_i \sum_{j \in N_i^{PR}} \kappa_{ij} x_i x_j \right) + K_2,$$

where

$$\gamma_{ij} = \frac{1}{2} \left[\boldsymbol{A}^T \boldsymbol{C}^{-1} \boldsymbol{A} \right]_{ij} \tag{40}$$

and

$$F_i(x_i) = - \left[\boldsymbol{A}^T \boldsymbol{C}^{-1} \boldsymbol{y} \right]_i x_i + \frac{1}{2} \left[\boldsymbol{A}^T \boldsymbol{C}^{-1} \boldsymbol{A} \right]_{ii} x_i^2. \tag{41}$$

The superscripts PR and PO denote prior and posterior, respectively. The constant terms K_1 and K_2 are independent of $\boldsymbol{x}$ and β and do not affect the choice of ΔH_i or estimation of β. N_i^{PR} and N_i^{PO} denote the prior and posterior neighborhoods of pixel i. The posterior neighborhood for pixel i is the set $N_i^{PO} = \{j : i \neq j, \left[\boldsymbol{A}^T \boldsymbol{C}^{-1} \boldsymbol{A} \right]_{ij} \neq 0\} \cup N_i^{PR}$.

Substitution of the approximation (26) with the optimal choice of ΔH_i as in (34) in (38) and (39) gives the following mean field energy functions:

$$E_{MF}^{(PO)} = \sum_i \{ L_i^{PO}(x_i) + \beta U_i^{PO}(x_i) \} \tag{42}$$

$$E_{MF}^{(PR)} = \beta \sum_i U_i^{PR}(x_i). \tag{43}$$

where

$$L_i^{PO}(x_i) = F_i(x_i) - 2 \sum_{j \in N_i^{PO}} \gamma_{ij} x_i < x_j >_{mf}^{PO} \tag{44}$$

$$U_i^{PO}(x_i) = \sum_{j \in N_i^{PR}} \kappa_{ij} x_i^2 - 2 \sum_{j \in N_i^{PR}} \kappa_{ij} x_i < x_j >_{mf}^{PO} \tag{45}$$

$$U_i^{PR}(x_i) = \sum_{j \in N_i^{PR}} \kappa_{ij} x_i^2 - 2 \sum_{j \in N_i^{PR}} \kappa_{ij} x_i < x_j >_{mf}^{PR}. \tag{46}$$

Here $< x_j >_{mf}^{PO}$ and $< x_j >_{mf}^{PR}$ denote $< x_j >^{mf}$ with respect to posterior and prior densities. Having developed the optimal mean field reference densities, it remains to use these to compute the partition function approximations:

$$Z(\boldsymbol{y}, \beta) \approx Z_{MF}^{(PO)} \exp\{ - < E^{(PO)} - E_{MF}^{(PO)} >_{MF}^{PO} \} \tag{47}$$

$$Z(\beta) \approx Z_{MF}^{(PR)} \exp\{ - < E^{(PR)} - E_{MF}^{(PR)} >_{MF}^{PR} \} \tag{48}$$

and substitute these in the likelihood equation (12) to compute the mean field approximated ML estimate of β. After substituting in (12) we have the following approximate likelihood equation:

$$\frac{\partial \ln Z_{MF}^{(PO)}}{\partial \beta} - \frac{\partial < E^{(PO)} - E_{MF}^{(PO)} >_{MF}^{PO}}{\partial \beta} = \frac{\partial \ln Z_{MF}^{(PR)}}{\partial \beta} - \frac{\partial < E^{(PR)} - E_{MF}^{(PR)} >_{MF}^{PR}}{\partial \beta}. \tag{49}$$

This can be rewritten as

$$\begin{aligned} &\sum_i \frac{\int_{x_i} U_i^{PO}(x_i) \exp\{-L_i^{PO}(x_i) - \beta U_i^{PO}(x_i)\} dx_i}{\int_{x_i} \exp\{-L_i^{PO}(x_i) - \beta U_i^{PO}(x_i)\} dx_i} - \sum_i \sum_{j \in N_i^{PR}} \kappa_{ij} < x_i >_{mf}^{PO} < x_j >_{mf}^{PO} \\ = &\sum_i \frac{\int_{x_i} U_i^{PR}(x_i) \exp\{-\beta U_i^{PR}(x_i)\} dx_i}{\int_{x_i} \exp\{-\beta U_i^{PR}(x_i)\} dx_i} - \sum_i \sum_{j \in N_i^{PR}} \kappa_{ij} < x_i >_{mf}^{PR} < x_j >_{mf}^{PR}, \end{aligned} \tag{50}$$

or equivalently:

$$\begin{aligned} &\sum_i E[U_i^{PO}(x_i) | < x_j >_{mf}^{PO}, j \in N_i^{PO}; \boldsymbol{y}, \beta] - \sum_i \sum_{j \in N_i^{PR}} \kappa_{ij} < x_i >_{mf}^{PO} < x_j >_{mf}^{PO} \\ = &\sum_i E[U_i^{PR}(x_i) | < x_j >_{mf}^{PR}, j \in N_i^{PR}, \beta] - \sum_i \sum_{j \in N_i^{PR}} \kappa_{ij} < x_i >_{mf}^{PR} < x_j >_{mf}^{PR}. \end{aligned} \tag{51}$$

This is the basic equation that must be solved to estimate β. For a given mean field $< x_i >^{mf}, \beta$ can be computed by finding a root of this equation. Since the mean field $< x_i >^{mf}$ is itself dependent on the value β, a recursive procedure which alternates between computation of $< x_i >^{mf}$ using the current value of β and vice versa, is required. We return to the problem of computing the solution in Section 3.4

3.3 Hyperparameter Estimation using a Generalized Approximation

The preceding development works only for the restricted class of auto-Gibbs distributions of the form (25). We now consider the more general case, and develop a sub-optimal mean field reference that can be applied to both Poisson and Gaussian likelihoods with any of the four potential functions in (8). Consider the general Gibbs distribution which is to be approximated:

$$P(\boldsymbol{x}) = \frac{1}{Z} \exp\{-E(\boldsymbol{x})\} \tag{52}$$

with conditional density:

$$P(x_i|x_{S\setminus i}) = \frac{1}{Z_i}\exp\{-E_i(x_i; x_{S\setminus i})\} \tag{53}$$

where $E_i(x_i; x_{S\setminus i})$ is the sum over all potential functions in $E(\boldsymbol{x})$ that include site i and $S\setminus i$ denotes the set of all pixel sites excluding i. We again use a separable mean field approximation:

$$P_{MF}(\boldsymbol{x}) = \frac{1}{Z_{MF}}\exp\{-E_{MF}(\boldsymbol{x})\} \equiv \prod_j P_i^{mf}(x_i), \tag{54}$$

where we define the local mean field densities, $P_i^{mf}(x_i)$, to be equal to the conditional density for each site, conditioned on the mean field of their neighbors, i.e.

$$\begin{aligned} P_i^{mf}(x_i) &= P(x_i|x_{S\setminus i})|_{x_{S\setminus i}=<x_{S\setminus i}>^{mf}} & (55)\\ &= \frac{1}{Z_i^{mf}}\exp\{-E_i^{mf}(x_i; < x_{S\setminus i} >^{mf})\} & (56) \end{aligned}$$

The corresponding partition function Z_i^{mf} is then given by:

$$Z_i^{mf} = \int_{x_i} \exp\{-E_i^{mf}(x_i; < x_{S\setminus i} >^{mf})\}dx_i. \tag{57}$$

Combining the local energy and partition functions gives the overall *mean field energy function* $E_{MF}(\boldsymbol{x})$, and *mean field partition function*, Z_{MF}:

$$E_{MF}(\boldsymbol{x}) = \sum_i E_i^{mf}(x_i; < x_{S\setminus i} >^{mf}) \qquad Z_{MF} = \prod_i Z_i^{mf}. \tag{58}$$

This mean field approximation can be applied to either the prior, $P(\boldsymbol{x}|\beta)$, or posterior, $P(\boldsymbol{x}|\boldsymbol{y},\beta)$, densities in Bayesian inverse problems provided the densities are written in the form of a Gibbs distribution. There are two alternative paths we can follow. One is to approximate both the prior and posterior partition functions and equate their derivatives to obtain a mean field approximation to the likelihood equation (12). Alternatively, we can use a mean field approximation of the posterior density only, and then substitute this in (9) and parallel the subsequent development in Section 2.2. In practice, we have observed little difference in the relative performance of the two approaches. Here we describe only the second of these two methods.

To find the posterior mean field approximation, we first form the conditional posterior densities. For the Gaussian and Poisson likelihoods, and the Gibbs priors with pairwise interactions as defined

in (4), we can write:

$$P(x_i|x_{S\setminus i}, \boldsymbol{y}) = \frac{P(\boldsymbol{y}|x_i; x_{S\setminus i})P(x_i|x_{S\setminus i})}{P(\boldsymbol{y}|x_{S\setminus i})} \tag{59}$$

$$= \frac{1}{Z_i}\exp\{-E_i(x_i; x_{S\setminus i})\} \tag{60}$$

$$= \frac{1}{Z_i}\exp\{\ln P(\boldsymbol{y}|x_i; x_{S\setminus i}) - \beta \sum_{j\in N_i^{PR}} V(x_i, x_j)\} \tag{61}$$

We can then write the local mean field densities in terms of these conditionals as:

$$P_i^{mf} = \frac{1}{Z_i^{mf}}\exp\{-E_i^{PO}(x_i; < x_{S\setminus i} >_{mf}^{PO})\} \tag{62}$$

where

$$E_i^{PO}(x_i; < x_{S\setminus i} >_{mf}^{PO}) = -\ln P(\boldsymbol{y}|x_i, < x_{S\setminus i} >_{mf}^{PO}) + \beta \sum_{j\in N_i^{PR}} V(x_i, < x_j >_{mf}^{PO}) \tag{63}$$

where $< x >_{mf}^{PO}$ denotes expectation with respect to the posterior mean field.

The joint mean field approximation of the marginalized likelihood (9) is then found by combining the P_i^{mf} to form the posterior mean field approximation and integrating over the sample space $\mathcal{X}$:

$$P(\boldsymbol{y}|\beta) \approx P_{MF}(\boldsymbol{y}|\beta) = \frac{1}{C}\prod_i \int_{x_i} exp\{\ln P(\boldsymbol{y}|x_i; < x_{S\setminus i} >_{mf}^{PO}) - \beta \sum_{j\in N_i^{PR}} V(x_i, < x_j >_{mf}^{PO})\}dx_i \tag{64}$$

where C contains terms which are independent of the data $\boldsymbol{y}$. We chose this normalizing constant so that $\int_{\boldsymbol{y}} P(\boldsymbol{y}|\beta)d\boldsymbol{y} = 1$:

$$P_{MF}(\boldsymbol{y}|\beta) = \frac{\prod_i \int_{x_i} exp\{\ln P(\boldsymbol{y}|x_i, < x_{S\setminus i} >_{mf}^{PO}) - \beta \sum_{j\in N_i^{PR}} V(x_i, < x_j >_{mf}^{PO})\}dx_i}{\prod_i \int_{x_i} \exp\{-\beta \sum_{j\in N_i^{PR}} V(x_i, < x_j >_{mf}^{PO})\}dx_i} \tag{65}$$

Starting from this approximation, we can now parallel the development in section 2.2, by setting the derivative of the log of the approximated marginal density in (65) to zero, to arrive at the approximate likelihood equation:

$$\sum_i \frac{\int_{x_i} U_i^{PO}(x_i)\exp\{\ln P(\boldsymbol{y}|x_i, < x_{S\setminus i} >_{mf}^{PO}) - \beta U_i^{PO}(x_i)\}dx_i}{\int_{x_i}\exp\{\ln P(\boldsymbol{y}|x_i, < x_{S\setminus i} >_{mf}^{PO}) - \beta U_i^{PO}(x_i)\}dx_i} \tag{66}$$

$$= \sum_j \frac{\int_{x_i} U_i^{PO}(x_i)\exp\{-\beta U_i^{PO}(x_i)\}dx_i}{\int_{x_i}\exp\{-\beta U_i^{PO}(x_i)\}dx_i}. \tag{67}$$

where

$$U_i^{PO}(x_i) = \sum_{j \in N_i^{PR}} V(x_i, < x_j >_{mf}^{PO}) \tag{68}$$

Finally, we can also rewrite this equation as

$$\sum_i E[U_i^{PO}(x_i) | < x_{S\setminus i} >_{mf}^{PO}; \boldsymbol{y}, \beta] = \sum_i E[U_i^{PO}(x_i) | < x_{S\setminus i} >_{mf}^{PO}, \beta] \tag{69}$$

which can be interpreted as a mean field approximation of the likelihood equation (15). Note that this version of the mean field approximated ML estimator is slightly different from that derived using the GBF bound i.e. equation (51). As we see below, methods which use the GBF bound outperform those based on (69). This is not surprising given the optimal nature of the first and heuristic nature of the second method. However, in cases where the optimal approximation cannot be found, the second method still performs exceptionally well in comparison to other well known methods.

3.4 Mean vs. Mode Field Approximations

In many imaging applications, we are more interested in computing a MAP estimate of the image than a minimum mean square error (MMSE) estimate. These correspond respectively, to the mode and the mean of the posterior densities. Therefore, rather than also computing the mean field of the posterior reference field, we replace the mean field with a *mode-field.* This mode is computed using an iterative MAP estimation procedure. Note that using the separable approximations described above, the mode of the original and reference fields are identical. In cases where the posterior density is unimodal and symmetric, mean and mode fields are equivalent. Such is the case for Gaussian data with the convex potential functions V_1 and V_2 defined in (8). For Poisson data, the Poisson likelihood is asymmetric and the two methods are not equivalent. However, for relatively high mean value, the Poisson likelihood is well approximated by a symmetric Gaussian function (i.e. the log-likelihood function $\ln P(\boldsymbol{y}|\boldsymbol{x})$ is approximately symmetric). We therefore anticipate only minor differences between the mode-field and mean-field approximations in this case.

We refer to the parameter estimation methods described above as *mode field approximated maximum likelihood* (MFAML). To distinguish the two approximations in section 3.1 and 3.2, we refer to them as MFAML-Opt and MFAML-Gen respectively.

4 Numerical Methods

4.1 Combined MAP Image Estimation and ML Hyperparameter Estimation

Using the approximations described above, the MAP estimate of the image and the ML estimate of the hyperparameter can be jointly computed using a two step iteration:

[1] Initialize the image $\boldsymbol{x}^k = \boldsymbol{x}^0$ and hyperparameter $\beta^k = \beta^0$. Set $k = 0$.

[2] Maximize $P(\boldsymbol{x}|\boldsymbol{y};\beta^k)$ to find $\boldsymbol{x}^{k+1}$.

[3] Compute a new hyperparameter value β^{k+1} by solving the approximated likelihood equation (51) or (69) using $\boldsymbol{x}^{k+1}$ as the current mode field.

[4] Set $k = k + 1$, goto step [2].

In practice, neither steps 2 or 3 need be iterated to convergence before moving to the next step. We have no convergence proof for this method. However, in running the method for a wide range conditions, we have never observed a case in which the method does not converge.

4.2 Computing the MAP Image Estimate

For a Gibbs prior of the form (5), the MAP estimate is found by maximizing over the log posterior density:

$$\bar{\boldsymbol{x}}(\beta) = \arg\max_{\boldsymbol{x}}\{-\frac{1}{2}(\boldsymbol{y}-A\boldsymbol{x})^T\mathbf{C_n}^{-1}(\boldsymbol{y}-A\boldsymbol{x}) - \beta\sum_j \sum_{k>j,k\in N_j} \kappa_{jk}V(x_j,x_k)\} \tag{70}$$

for the Gaussian likelihood, and

$$\bar{\boldsymbol{x}}(\beta) = \arg\max_{\boldsymbol{x}}\{\sum_i(-\sum_j A_{ij}x_j + y_i \ln(\sum_j A_{ij}x_j)) - \beta\sum_j \sum_{k>j,k\in N_j} \kappa_{jk}V(x_j,x_k)\} \tag{71}$$

for the Poisson likelihood.

These functions are concave for V_1 and V_2 but not for V_3 and V_4. Gradient based optimization will therefore lead only to local maxima for the last two potential functions. However, it is widely accepted that for most practical applications a local optimum is acceptable. We therefore restrict attention here to local search methods, although the MFAML method described above can be combined with any numerical procedure for computing a MAP image estimate. Many computational methods for solving large inverse problems in image processing have been studied in recent years. These include Gauss-Siedel procedures (sequential coordinate descent algorithm) [5],

conjugate gradient methods [27, 25], the method of iterated conditional modes [3], iterated conditional average (ICA) [23, 16] and generalized EM-methods. The performance of these algorithms in terms of computation cost and convergence rate is highly problem dependent. We have previously found that conjugate gradient methods produce favorable performance for image restoration and reconstruction problems [25] and will use this approach in the results presented below.

A special problem involved with gradient based MAP estimation when using a non-negativity constraint, is how to enforce non-negativity of the pixel value. We can write the MAP estimation problem in the general form:

$$\begin{aligned} &\max_{\boldsymbol{x}} f(\boldsymbol{x}) \\ &\text{subject to} \quad \boldsymbol{x} > \mathbf{o} \end{aligned} \tag{72}$$

where $f(\boldsymbol{x})$ is the log-posterior density. A penalty function is introduced to convert the problem to an unconstrained one:

$$\max_{\boldsymbol{x}} \; q(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{1}{\gamma} G(\boldsymbol{x}) \tag{73}$$

where γ is a positive constant and we use

$$\frac{1}{\gamma} G(\boldsymbol{x}) = \sum_j \left(\frac{x_j}{\gamma}\right)^2 u(-x_j) \tag{74}$$

where $u(.)$ is the unit step function. In principle, a sequence of solutions to this unconstrained problem, corresponding to a decreasing sequence in the parameter γ, should be generated. This sequence converges to a solution of the original constrained problem. In practice we find that provided an appropriate value of γ is used it can be held constant throughout the iteration process without either significantly reducing the convergence rate of the algorithm or resulting in significant negative pixel values. The use of conjugate gradient methods in conjunction with a penalty function for image reconstruction is described in detail in [25].

4.3 Computing the Hyperparameter Value

The method that we use to implement Step 3 is an EM-like algorithm. We adopted this approach after finding problems with numerical stability when using a standard Newton-Raphson procedure. For hyperparameter estimation using the mean field approximation based on the GBF bound, we

perform step 3 as follows:

[3a] Compute the mean field approximated statistic $U_{MF}^{(k+1)}(\boldsymbol{x})$ defined as the current left hand side of the mean field likelihood equation (50):

$$U_{MF}^{(k+1)}(\boldsymbol{x}) = \sum_i E[U_i^{PO}(x_i)|x_{S\setminus i}^{k+1}, j \in N_i^P; \boldsymbol{y}, \beta] - \sum_i \sum_{j \in N_i} \kappa_{ij} x_i^{k+1} x_j^{k+1}. \tag{75}$$

[3b] Compute the new hyperparameter value β^{k+1} by solving the equation:

$$\sum_i E[U_i^{PR}(x_i)| < x_j >_{mf}^{PR}, j \in N_i; \beta] - \sum_i \sum_{j \in N_i} \kappa_{ij} < x_i >_{mf}^{PR} < x_j >_{mf}^{PR} = U_{MF}^{(k+1)}(\boldsymbol{x}), \tag{76}$$

For the general approximation, we use the following method to solve step 3:

[3a] Compute the mean field approximated statistic $U_{MF}^{(k+1)}(\boldsymbol{x})$ defined as the current left hand side of the mean field likelihood equation (66):

$$U_{MF}^{(k+1)}(\boldsymbol{x}) = \sum_i E[U_i^{PO}(x_i)|x_{S\setminus i}^{k+1}; \boldsymbol{y}, \beta] \tag{77}$$

[3b] Compute the new hyperparameter value β^{k+1} by solving the equation

$$\sum_i E[U_i^{PO}(x_i)|x_{S\setminus i}^{k+1}; \beta] = U_{MF}^{(k+1)}(\boldsymbol{x}) \tag{78}$$

In Step 3b of this EM-like algorithm, the new hyperparameter value is computed using one or more iterations of a Newton-Raphson procedure. All integrals encountered were computed numerically using an adaptive quadrature method [26]. We also use a scaling procedure to ensure that the single pixel sample space is approximately $[0, 1]$. This can be achieved by a corresponding inverse scaling of the elements of the A operator in the likelihood function. This has the effect of avoiding large numerical errors when computing integrals containing integrands of the form $\exp\{-\beta \sum_{j \in N_i} V(x_i, x_j)\}$.

4.4 Computational Cost

The computational cost of the algorithm we describe above is highly problem dependent. We usually run 5 to 10 iterations of the conjugate gradient algorithm to update the MAP image estimate for a given value of β and then use one or two Newton Raphson iterations to update

the value of β. We typically repeat this procedure 10-20 times to achieve effective convergence in β. We have observed that the number of iterations required increases with both the degree of blurring and the variance of the additive noise. For image restoration with local blurring only, the dominant computational cost is associated with the Newton Raphson iterations for updating the hyperparameter. On a SunSPARC20/61 workstation, each iteration of the conjugate gradient MAP algorithm for a 256×256 pixel image requires only a few seconds. Each iteration of the Newton Raphson algorithm can take from several seconds to several minutes. This is because each Newton Raphson iterations requires with $3 \times 256 \times 256$ numerical integrations. For problems with Gaussian likelihoods and quadratic priors, we can replace the integrals using an error function look up table, thus reducing the per iteration cost to a few seconds.

5 Performance Studies

We have applied the mode field approximated maximum likelihood (MFAML) method to image restoration and reconstruction. We present the results for image restoration below. Application of this method to parameter estimation in positron emission tomography (PET) is described in [36]. We simply note here that we have observed similar performance for the PET problem to that described below for image restoration.

5.1 Estimator Bias and Variance using Stochastic Sampling

We used extensive Monte Carlo simulations to evaluate the performance of the new MFAML hyperparameter estimators in the problem of image restoration from blurred data with additive Gaussian noise. We have compared the performance of the MFAML methods described above with generalized maximum pseudolikelihood (GMPL) and the method of moments (MOM), for which the statistic $M(\boldsymbol{y})$ takes the same form as the Gibbs energy function of the prior, computed over the noisy image $\boldsymbol{y}$ with eight nearest neighbor interactions.

We performed Monte Carlo studies for image restoration as follows. For each value of the hyperparameter, fifty sample images were drawn from a specific prior using the Metropolis algorithm [22]. Each sampled image was then blurred by one of the following 3×3 kernels:

True β	0.0004	0.0010	0.0040	0.0100	0.0400	0.100
GMPL Mean	4.134e-4	1.093e-3	7.742e-3	*	*	*
GMPL Bias (%)	3.35%	9.30%	93.6%	*	*	*
GMPL STD (%)	1.74%	1.60%	6.49%	*	*	*
MOM Mean	4.039e-4	1.012e-3	4.175e-3	1.154e-2	0.0775	0.3565
MOM Bias (%)	0.97%	1.20%	4.37%	15.4%	93.7%	257 %
MOM STD (%)	2.13%	3.21%	10.9%	31.0%	340%	380 %
MFAML-Gen Mean	4.010e-4	1.009e-3	4.114-3	1.071e-2	0.0472	0.1241
MFAML-Gen Bias (%)	0.25%	0.89%	2.84%	7.11%	17.7%	24.1%
MFAML-Gen STD (%)	1.64%	1.35%	1.63%	2.23%	3.98%	8.81%
MFAML-Opt Mean	4.153e-4	1.004e-3	3.977e-3	9.526e-3	0.0357	0.07842
MFAML-Opt Bias (%)	3.8%	0.42%	-0.572%	-4.74%	-10.7%	-21.58%
MFAML-Opt STD (%)	1.21%	1.37%	1.27%	2.51%	4.02%	9.02%

Table 1: Monte Carlo Test for hyperparameter estimation comparing performance of generalized maximum pseudo-likelihood (GMPL), the method of moments (MOM), and the three mode field approximated ML methods (MFAML) described in Section 2.3. Mean, percentage bias and variance were computed using 50 independent images drawn from the prior using a Gibbs sampler and then blurred and contaminated with Gaussian noise $N(0, 16)$ The prior had a quadratic Hamiltonian defined on a second order neighborhood. (* indicates algorithm fails to reach a solution. STD=Standard Deviation)

$$\text{Opt 1:} \begin{pmatrix} 0.001 & 0.028 & 0.001 \\ 0.028 & 0.884 & 0.028 \\ 0.001 & 0.028 & 0.001 \end{pmatrix}, \text{Opt 2:} \begin{pmatrix} 1/16 & 1/8 & 1/16 \\ 1/8 & 1/4 & 1/8 \\ 1/16 & 1/8 & 1/16 \end{pmatrix}$$

and

$$\text{Opt 3:} \begin{pmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{pmatrix}.$$

Note that the degree of smoothing increases from Opt 1 to Opt 3. Pseudo-random Gaussian noise with known variance σ was generated to contaminate each of the resulting blurred images. The likelihood function for these noisy data take the form of (1). The hyperparameters were estimated for each method of interest for each of the fifty noisy images. Since the original images are sampled from specific priors with known hyperparameter values, we were able to calculate bias and variance across the fifty resulting estimates.

A comparison of the performance of the various methods for a range of values of β is shown in Table 1. The original images were generated using the Metropolis algorithm with the the quadratic prior with the single pixel sample space $[0, 100]$. These were then blurred using Opt 1

Table 2: Monte Carlo test of MFAML-opt and MFAML-gen performance as a function of additive noise variance. different noise variance

True β	σ^2	MFAML-Gen			MFAML-Opt		
		mean	Bias (%)	Var (%)	mean	Bias (%)	Var (%)
0.001	4	1.009e-3	0.9%	1.35%	1.004e-3	0.42%	1.37%
	16	1.053e-3	52.8%	1.39%	1.005e-3	0.47%	1.68%
	36	1.111e-3	11.1%	1.85%	9.863e-4	-1.37%	1.99%
	100	1.232e-3	23.2%	3.83%	8.844e-4	-11.5%	2.19%
	400	1.348e-3	34.8%	10.4%	7.450e-4	-25.5%	12.1%

and contaminated by zero mean Gaussian noise with variance $\sigma^2 = 4$. All methods perform best when β is small and deteriorate as β increases and the images become smoother. The GMPL method works only for the smaller values of β. As β increases, the two step method, which iterates between MAP estimation of the image x and estimation of β, fails to converge. The MOM method performs better in general, but as β increases, the slope of the moment curve decreases, leading to increased bias and variance. In all cases, both the general and optimal forms of MFAML outperform both of the other techniques. The differences are very clear for the cases where β is large, which corresponds to the case of very smooth images. For these larger β values, MFAML shows approximately a ten fold reduction in bias and variance relative to the MOM method. The optimal form of MFAML exhibits lower bias than the general form, with slightly larger variance with overall superior performance. However, in practice these differences are small and lead to little noticeable difference in image quality when applied to real images.

To test the robustness of the MFAML methods to noise, we used the same setup as in the comparative studies above and generated data for a range of additive noise variances. As before, ensemble statistics were computed to determine the effects of different noise levels on the bias and variance of β. We summarize these results in Table 2. Although we do observe deterioration in the performance when noise variance increases, both MFAML methods appear to perform well and are stable even for very large additive noise variances.

The conditioning of the likelihood affects the degree of ill-posedness of the inverse problem, i.e. the conditioning of the operator $\boldsymbol{A}$ determines our ability to recover the image x from the blurred data, which in turn affects our ability to accurately estimate β. Results in Table 3 show that as the degree of blurring increases and the inverse problem becomes more ill-posed, performance of the MFAML methods deteriorates. The bias in the estimator appears to be more affected than variance

Table 3: Robustness of MFAML-gen and MFAML-opt to different smoothing operators

True β	3×3 operator	MFAML-Gen			MFAML-Opt		
		mean	Bias (%)	Var (%)	mean	Bias (%)	Var (%)
0.004	Opt 1	4.114e-3	2.84%	1.63%	3.977e-3	-0.572%	1.27%
	Opt 2	9.821e-3	145%	3.78%	5.867e-3	46.6%	2.07%
	Opt 3	1.079e-2	170%	3.31%	5.840e-3	46.0%	2.14%
0.01	Opt 1	1.071e-2	7.11%	2.23%	9.526e-3	-4.74%	2.51%
	Opt 2	3.401e-2	240%	5.88%	1.060e-2	5.96%	1.97%
	Opt 3	3.628e-2	262%	10.2%	1.085e-3	8.48%	2.01%
0.04	Opt 1	0.0472	17.7%	3.89%	0.0357	-10.7%	4.02%
	Opt 2	0.1030	157%	8.7%	0.0215	-46.2%	1.26%
	Opt 3	0.1055	164%	9.71%	0.0225	44.1%	5.21%

by changes in the degree of blurring. Note also that in this example, there are more substantial differences in performance between the general and optimal MFAML methods than was seen in Table 1. For the Opt-2 and Opt-3 blurring kernels, GMPL does not converge and MOM is unable to identify the parameter due to the flatness of the moment curve.

Finally, we note that the bias in several cases in the tables presented above is often very large. While in many problems, bias of more than a few percent may be unacceptable, we will see below that MAP estimation is fairly robust to errors in β. Even mis-estimating β by a factor of 2 or 3 may not lead to gross errors in the associated MAP image estimate.

5.2 Applications and Validations with Real Images

In this experiment, we used the 3×3 blurring mask Opt 2 to blur two 256×256 pixel images ("Boat" and "Moon"). The single pixel sample space of the Boat image is $[0, 255]$, and that of the Moon image is $[0, 128]$. Then we generated Gaussian noise with a variance of 100 to contaminate the resulting blurred Boat image, and used a pseudo-random Poisson generator to make the blurred Moon image Poisson. The images were then restored using MAP estimation for each of the four potential functions in (8) and the appropriate likelihood function. Images were reconstructed for a range of fixed values of β and the total squared error between the original and restored image calculated. The images were then reconstructed again with simultaneous MFAML estimation of β. For the case of Gaussian noise and the quadratic prior we use both MFAML-Gen and MFAML-Opt estimator. In all other cases we use just the MFAML-gen method.

The restored images for the cases where β is estimated are shown in Figure 3 and 6. The

corresponding curves showing the restored image error as a function of hyperparameter are shown in Figure 4 and 5 for the Boat image, and Figure 7 and 8 for the Moon image. Note the log-scale on the β axis. We show the location of the MFAML-gen and MFAML-opt estimate of the hyperparameter on the curves using "*" and "O" respectively. Also shown in these figures is the corresponding L-curve, again with the location of the estimated hyperparameter indicated. These results show that estimated hyperparameter gives close the the minimum squared error in all cases, and is located close to the knee of the L-curve. Notice also the robustness to errors in β as evidenced by the need to use a log scale in the β axis.

Conclusion

We have described a general method for estimating the hyperparameter of Gibbs priors from incomplete data. This method is based on a mean-field like approximation of the Gibbs distributions involved. The result provides a balance between the over-simplified model implicit in the generalized ML methods and the intractability of a true ML estimator. While computational costs are significant, we anticipate they will be acceptable in practical situations. Convergence of the method by which the solution is computed simultaneously with a MAP image estimate has not been shown, however we have not encountered any problems with convergence in the many cases we have run.

The results presented indicate that good performance is achieved over a range of conditions when applied to image restoration. We have also observed similar behavior in applications to positron emission tomography [36]. We do observe that the estimator degrades as the degree of blurring increases. This is inevitable in the sense that the ultimate performance of the method is limited by the slope of the likelihood function $p(\boldsymbol{y}|\beta)$. The method described here is not limited to estimation of a single parameter nor to the specific problems described. It appears straightforward to modify this approach to the estimation of multiple parameters, and also to estimation of the hyperparameters of discrete spatial processes such as those used for image segmentation and labelling.

References

[1] M. Almeida and G. Gidas, "A Variational Method for Estimating the Parameters of MRF from Complete or Incomplete Data," Technical Report, Brown University, 1989.

[2] J. Besag, " Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of Royal Statistical Society, B,* vol. 36, pp. 192-326, 1974

[3] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, B,* vol. 48(3), pp. 259-302, 1986

[4] A. Blake and A. Zisserman, *Visual Reconstruction*, Artificial Intelligence, MIT Press, Cambridge, MA, 1987

[5] C. Bouman and K. Sauer, "Fast Numerical Methods for Emission and Transmission Tomographic Reconstruction," *In Proc. Conf. Info. Sci. Sys., Johns Hopkins,* 1993

[6] D. Chandler, *Introduction to Modern Statistical Mechanics.* Oxford University Press, 1987.

[7] A. Dempster, N. Laird and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society,* Vol. 29, pp. 1-38, 1977

[8] P. Craven and G. Wahba, "Smoothing Noisy Data with Spline Function," *Numerische Mathematik,* vol. 31, pp. 377-403, 1979

[9] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Trans. Patt. Anal. and Machine Intell.,* vol. PAMI-6, pp. 721-741, Nov. 1984.

[10] S. Geman and D. McClure, "Statistical Methods For Tomographic Image Reconstruction," *In Proc. of the 46th Session of the ISI, Bulletin of the ISI,* Vol.52, (1987)

[11] D. Geiger and F. Girosi,"Parallel and Deterministic Algorithms for MRFs: Surface Reconstruction and Integration," *IEEE Trans.Patt. Anal. and Machine Intell.,* Vol. PAMI-12, pp 401-412, May 1991

[12] G. G. Potamianos and J. K.Goutsias, "Patition Function Estimation of Gibbs Random Field Images Using Monte Carlo Simulations," *IEEE Trans. Inform. Theory,* vol. 39, pp. 1322-1332, July 1993

[13] P. C. Hansen,"Analysis of Discrete Ill-posed Problems by Means of the L-curve," *SIAM Review* vol. 34, No.4,pp.561-580, 1992

[14] P.C. Hansen and D.P.O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems,"Report UMIACS-TR-91-142, Dept. of Computer Science, University of Maryland, College Park, MD.

[15] T. J.Hebert and R. Leahy,"Statistic-Based MAP Image Reconstruction from Poisson Data Using Gibbs Priors," *In IEEE Tran. on Signal Proc.* vol. 40, No. 9, pp. 2290-2303, Sept. 1992

[16] V. E.Johnson, W. H.Wong, X. Hu, and C. Chen," Image Restoration Using Gibbs Priors: Boundary Modeling, Treatment of Blurring, and Selection of Hyperparameter," *IEEE Trans. Patt. Anal. and Machine Intell.* vol. PAMI-13, No. 5, pp. 413-425, May 1991

[17] S. Lakshmanan and H. Derin, "Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated Annealing,"*IEEE Trans. on Patt. Anal. and Mchine Intell.* vol. PAMI-11, pp. 799-813,1989

[18] R. Leahy and X. Yan, "Incorporation of Anotomical MR data for improved Functional Imaging with PET", In A. C. F. Colchester and D. J.Hawkes, editors, *Information Processing in Medical Imaging,* pp. 105-120, Springer-Verlag, 1991.

[19] P. J. M.van Laarhoven and E. H. L.Arts, *Simulated Annealing: Theory and Applications,* D.Reidel Publishing Company, 1987

[20] D. Luenberger, *Linear and Nonlinear Programming.* Addison-Wesley Inc., Menlo Park, CA, 1989

[21] J. Mathews, R. L.Walker, *Mathematical Methods of Physics*, The Benjamin/Cummings Inc. Menlo Park, CA, 1970

[22] J. Marroquin,"*Probabilistic Solution of Inverse Problems,*" Ph.D. thesis, MIT. Cambridge, Sept. 1985

[23] K. M. Manbeck, "*Bayesian Statistical Methods Applied to Emission Tomography with Physical Phantom and Patient Data,*" Ph.D. thesis, Brown University, 1990.

[24] A. Mohammad-Djafari, "On the Estimation of Hyperparameters in Bayesian Approach of Solving Inverse Problems," *In Proc. ICASSP-93* pp. V495-498.

[25] E. Mumcuoglu, R. Leahy, S. Cherry, Z. Zhou, "Fast Gradient-Based Methods for Bayesian Reconstruction of Transmission and Emission PET Images," *To appear IEEE Trans. on Medical Imaging, Dec., 1994*

[26] H. R. Schwarz, J. Waldvogel, *Numerical Analysis- A Comprehensive Introduction*, pp. 339-342, J. Wiley&Sons, New York.

[27] A. Rangarajan, "*Representation and Recovery of Discontinuities in Some Early Vision Problems*", Ph.D. thesis, University of Southern California, Nov. 1990

[28] G. Gindi, M. Lee, A. Rangarajan, et al, "A Continuation Method for Emission Tomography", *In Proc. 1992 Nucl. Sci. Symp and Med. Imag. Conf.,* pp 1204-1206, 1992.

[29] A. M. Thompson, J. C. Brown, et al,"A Study of Methods of Choosing the Smoothing Parameter in Image Restoration by Regularization," *In IEEE Tran. on Patt. Anal. and Machine Intell.* vol. 13, No. 4, pp. 326-339, April 1991

[30] J. Varah, "Pitfalls in the numerical solution of ill-posed problems," *SIAM J. Sci. Statist. Comput*, 4 pp.164-176, 1983

[31] G. Wahba, "Spline Models for Observational Data," *CBMS-NSF Reginal Conference Series in Applied Mathematics, Vol.59, Society for Industrial and Applied Mathematics,* Philadelphia, PA, 1990

[32] J. Zhang, "The Mean Field Theory in EM Procedures for Markov Random Fields,"*IEEE Trans. on Signal Processing.* vol. 40. No. 10. Oct. 1992

[33] J. Zhang and J. Hanauer, "The Mean Field Theory for Image Motion Estimation", In *Proc. ICASSP, 1993*, pp.V197-V200.

[34] J. Zhang, J. W.Modestino, and D. A.Langan,"Maximum-Likelihood Parameter Estimation for Unsupervised Stochastic Model-Based Image Segmentation,"*IEEE Trans. Image Proc..* vol. 3, pp. 404-420, July, 1994

[35] Z. Zhou, R. Leahy, E. Mumguoclu, "A Comparative Study of Using Anatomical Boundary in PET Reconstruction," In *IEEE Proc. Nucl. Sci. Symp and Med. Imag. Conf. 1993*

[36] Z. Zhou, R. Leahy, E. Mumguoclu,"Maximum Likelihood Hyperparameter Estimation for Gibbs Priors from Incomplete Data with Applications to Positron Emission Tomography" *In Proc. XIVth International Conference on Information Processing in Medical Imaging,* France, 1995

[37] Z. Zhou, "*Maximum Likelihood Hyperparameter Estimation for Gibbs Priors from Incomplete Data with Applications in Image Processing,*" Ph.D Thesis, University of Southern California, 1994.

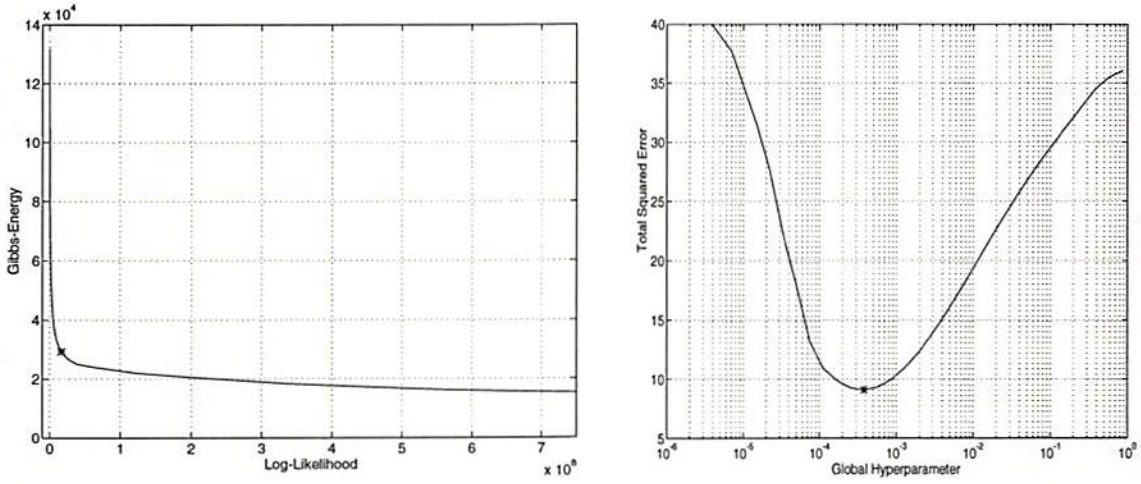


Figure 1: Illustration of the quantitative effect of the global hyperparameter (a) L-curve, (b) Global mean-squared-error of restored image versus β.

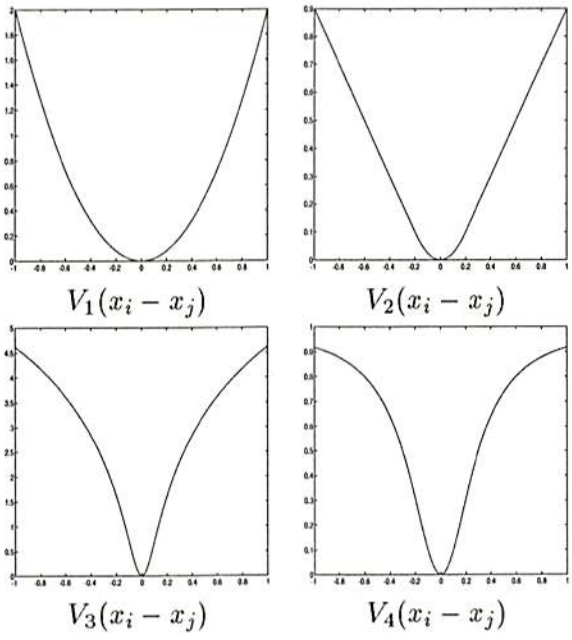


Figure 2: Illustration of potential functions of Gibbs prior

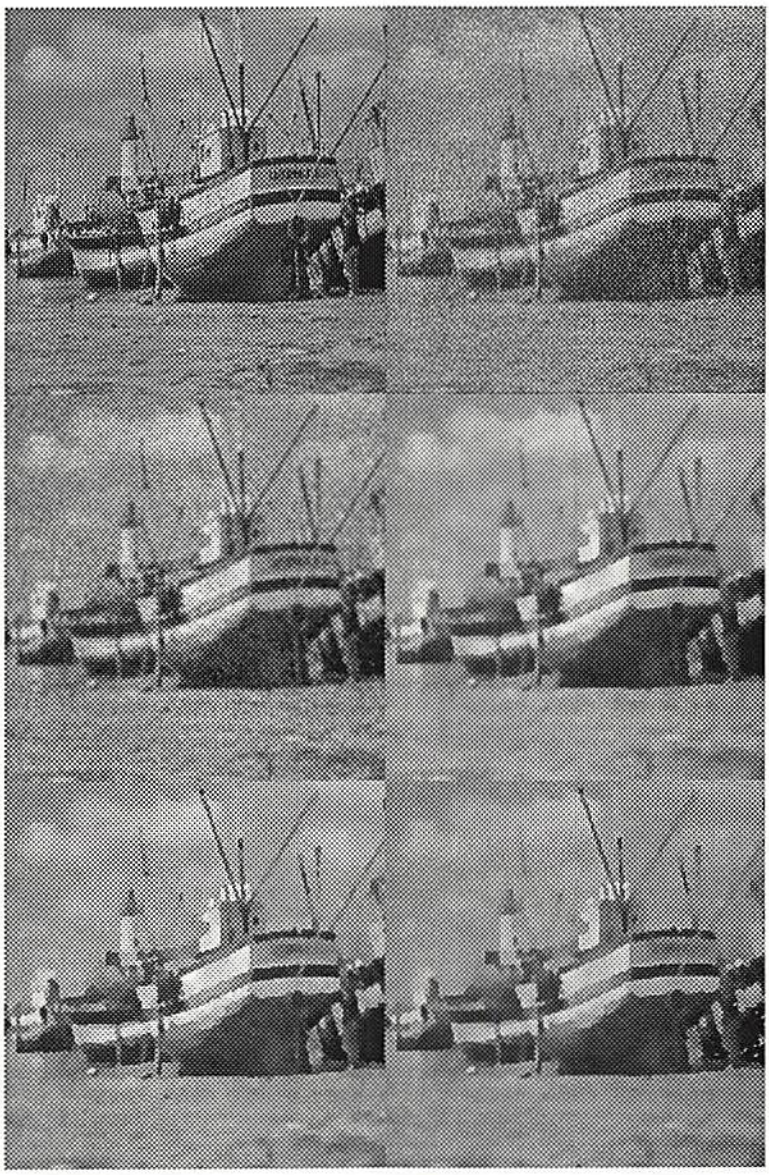
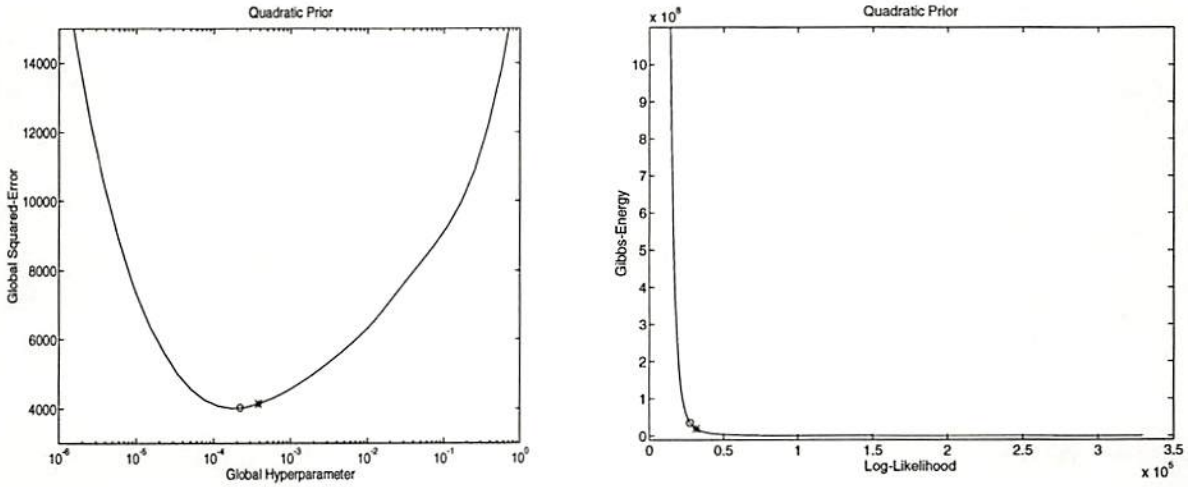
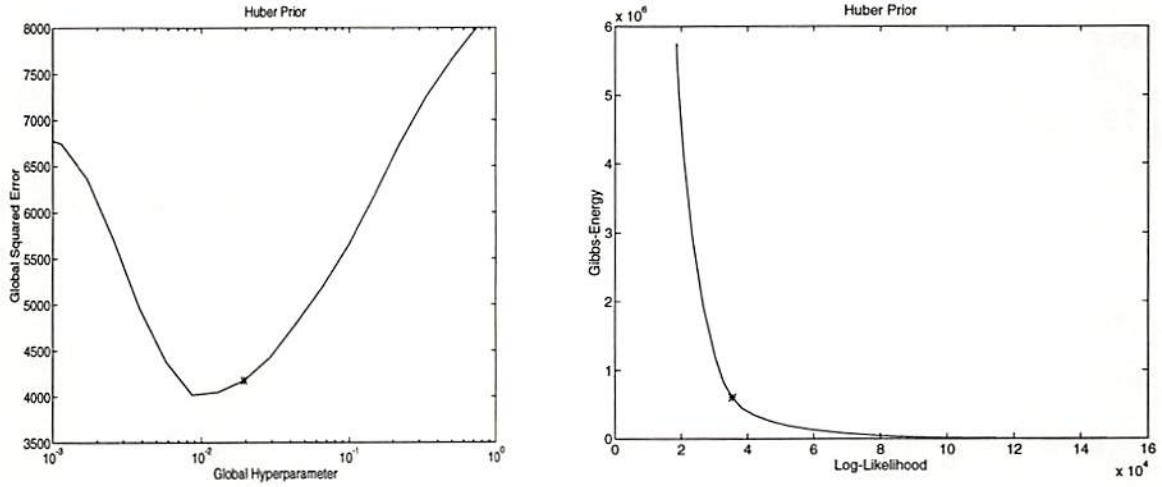


Figure 3: Experiment for image restoration from Gaussian data, $\sigma^2 = 100$, with blurring kernel Opt-2: Top row: left=original, right=noisy, blurred data; middle row: left - MAP with Quadratic prior, right - MAP with Huber prior; bottom row: left - MAP with log-quadratic prior, right - MAP with saturated-quadratic prior. All images shown above correspond to the estimated β use MFAML-gen.

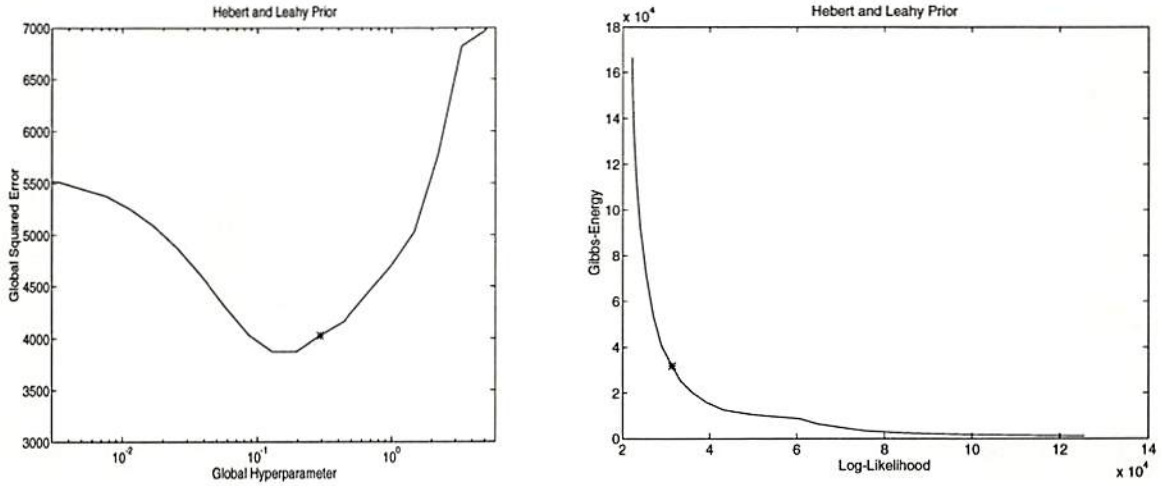


MAP with quadratic prior

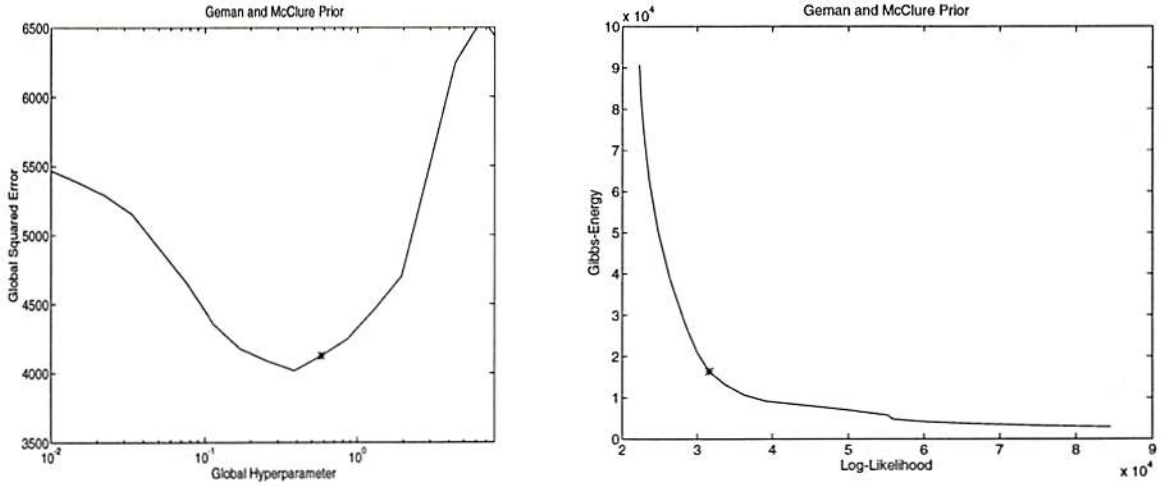


MAP with Huber prior

Figure 4: Left: plots of total squared error in MAP estimates versus hyperparameter for the boat image restored from Gaussian data; right: L-curves. '*' indicates β value estimated using MFAML-gen, 'o' indicates β value estimated using MFAML-opt.



MAP with log-quadratic prior



MAP with saturated-quadratic prior

Figure 5: Left: plots of total squared error in MAP estimates versus hyperparameter for the boat image restored from Gaussian data; right: L-curves. '*' indicates β value estimated using MFAML-gen.

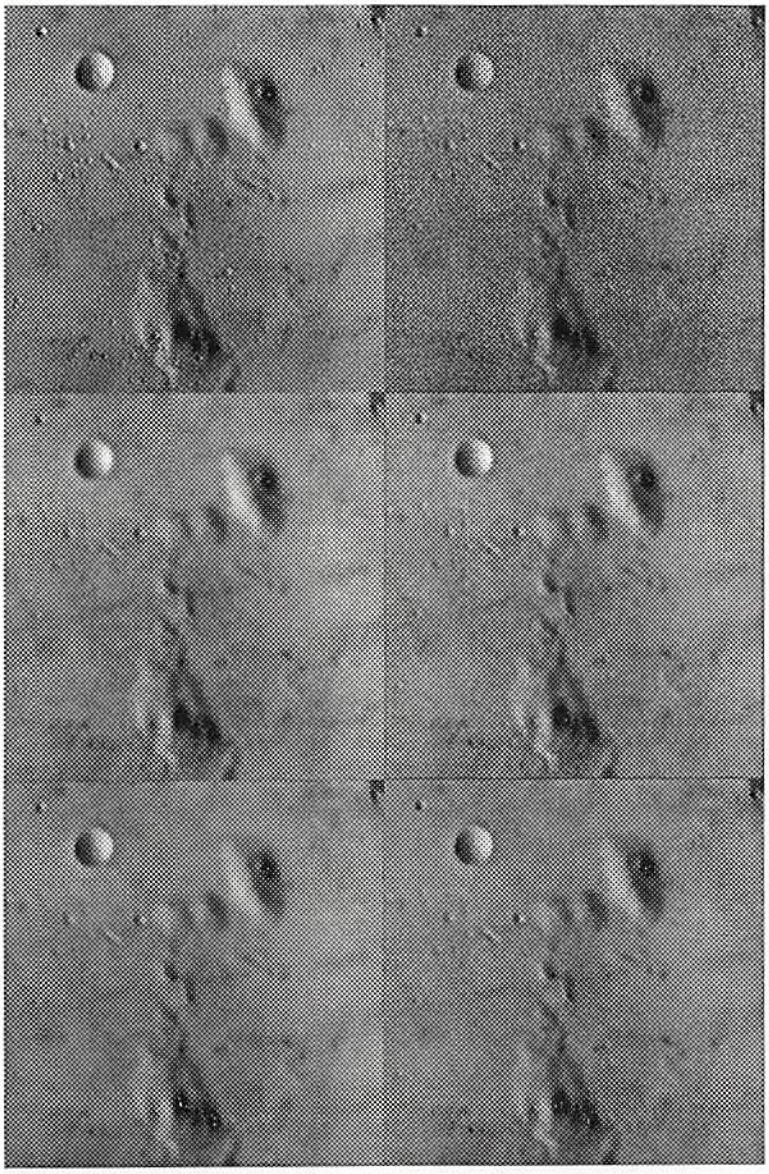
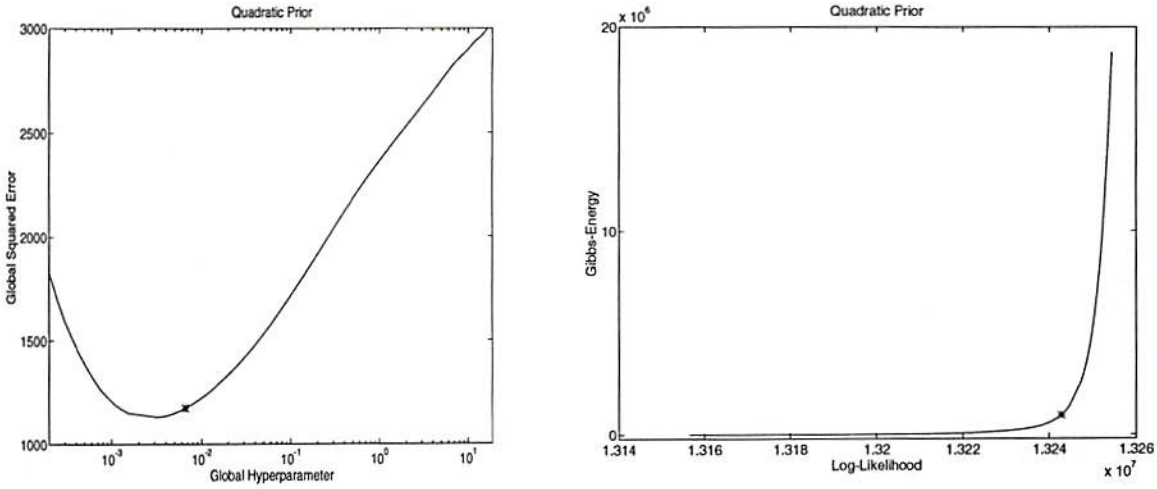
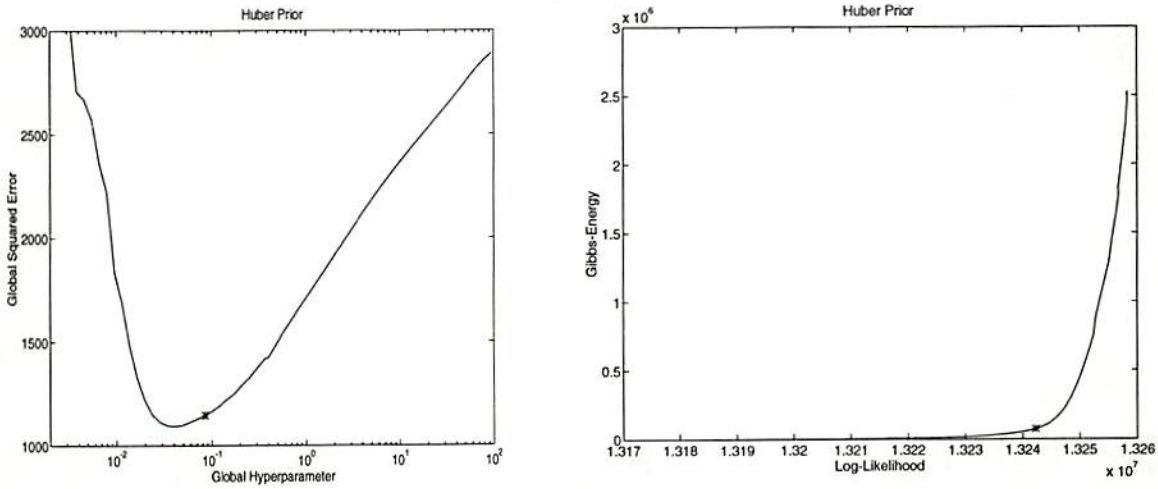


Figure 6: Experiment for image restoration from Poisson data with blurring kernel Opt-2: Top row: left=original, right=noisy, blurred data; middle row: left - MAP with Quadratic prior, right - MAP with Huber prior; bottom row: left - MAP with log-quadratic prior, right - MAP with saturated-quadratic prior. All images shown above correspond to the estimated β use MFAML-gen.

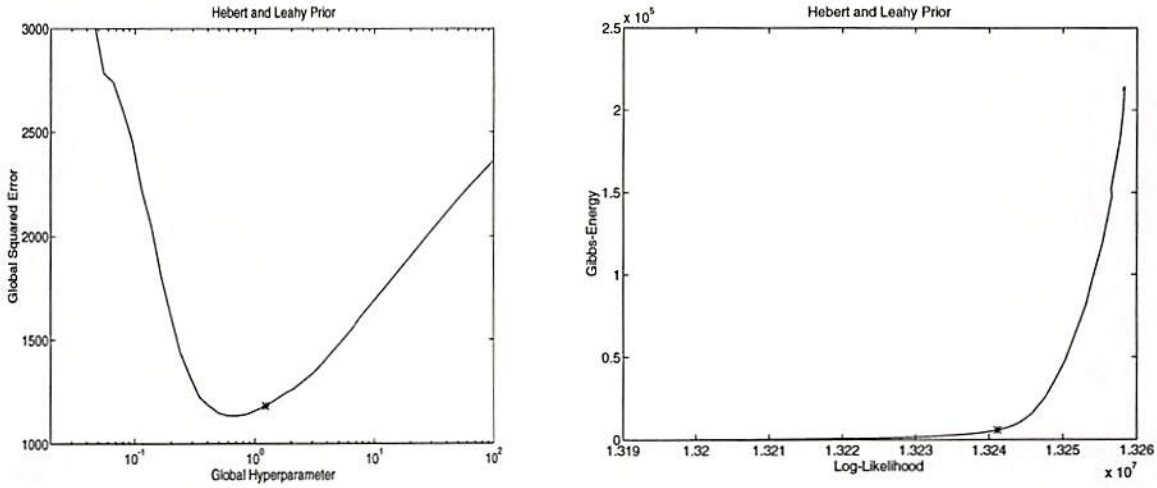


MAP with quadratic prior

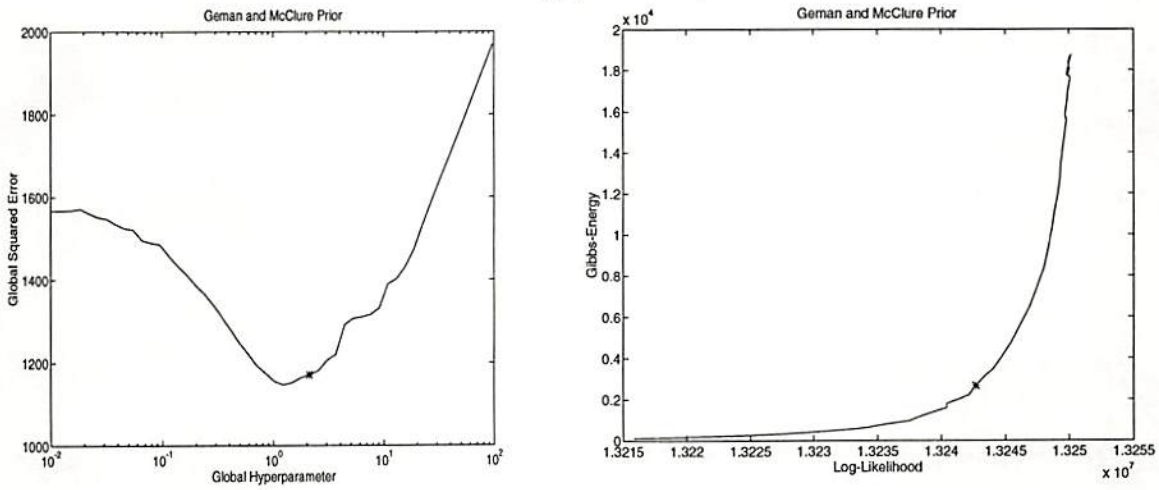


MAP with Huber prior

Figure 7: Left: plots of total squared error in MAP estimates versus hyperparameter for the moon image restored from Poisson data; right: L-curves. '*' indicates β value estimated using MFAML-gen



MAP with log-quadratic prior



MAP with saturated-quadratic prior

Figure 8: Left: plots of total squared error in MAP estimates versus hyperparameter for the moon image restored from Poisson data; right: L-curves. '*' indicates β value estimated using MFAML-gen.