SPLINE-BASED ADAPTIVE RESOLUTION IMAGE RECONSTRUCTION

W. VON DER LINDEN, V. DOSE, AND R. FISCHER
Max-Planck-Institut für Plasmaphysik,
EURATOM Association
D-85740 Garching b. München, Germany ¶

Abstract. Bayesian probability theory allows to infer an image given data constraints, prior knowledge, and background information. One ingredient of the background information is usually paid little attention, namely the image-grid, i.e. the points on which the desired image is reconstructed. In many problems an equidistant mesh is used. The choice of the image-grid can, however, strongly influence the reconstruction: if the grid is too coarse accuracy is wasted, if it is too fine artificial structures due to ringing and noise-fitting can show up. In order to achieve the best resolution supported by the data we include the grid into the Bayesian analysis and allowed for locally varying resolution. We applied our procedure to one dimensional problems. The image is reconstructed at the image-grid and interpolated in the interstitial regions by cubic splines. The bayesian analysis contains two competing tendencies: the data-constraints tend towards a fine grid as it allows to reduce the misfit, while Ockham's factor favors a coarse grid to keep the image "simple". The Bayesian solution represents a trade-off between the two trends and leads to results which are significantly improved over those obtained by fixed-grid approaches: over-fitting is eliminated and ringing is strongly suppressed, while the sharp structures are improved considerably. We applied the adaptive resolution idea to different types of problems such as deconvolution and density estimation. For both applications we present a representative physical problem.

Key words: ringing,adaptive kernels,adaptive resolution,over-fitting

1. Introduction

We have shown in previous papers [1, 2] that the image reconstructed by maxent, even if no approximations are invoked, suffers from ringing and over-fitting. We recall the results of a maxent deconvolution of a spectral density in fig.1. Details will be deferred to a later section. The reconstructed image is depicted in the lower panel which clearly suffers from noise fitting and ringing. These unpleasant features are, however, not a shortcoming of the entropic prior only but are also present for many other priors. The reason as discussed in [2] are the low-lying eigenvalues of the apparatus function. It has been pointed out that ringing is inevitable even for exact data constraints if the apparatus function is sufficiently broad and has correspondingly small eigenvalues.

What went wrong in the Bayesian analysis of theses problems? The fact that we are sure that the result has artificial structures implies that we have some strong prior knowledge about the solution which has not yet been included into the

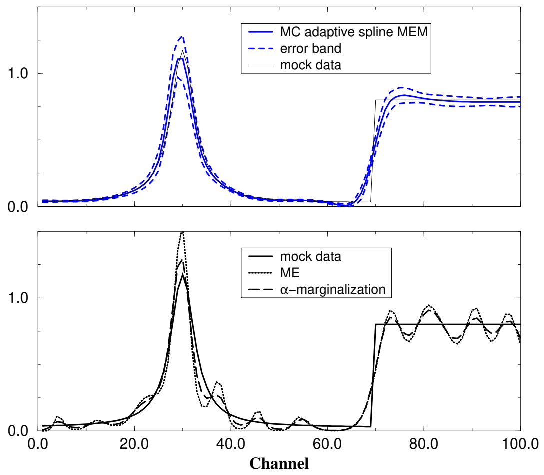

Figure 1: *Deconvolution problem for a spectral density consisting of a quasiparticle peak and the onset of continuum excitations. The lower panel depicts maxent results for a equidistant image-grid. As described in the legend the result of the evidence approximation and that of the full Bayesian analysis is shown. The upper panel provides a comparison between the exact spectrum and deconvolution with the present approach.*

prior. In most applications there are two features which the solution should have: local smoothness and structures only if they are supported by the data. In early publications it had been claimed that the maxent solution has these properties but that is obviously not the case, the reason being that the entropic prior contains no correlation among the image points. The fact that any permutation of the image has the same entropy and hence the same prior probability is certainly not appropriate for image or spectrum reconstruction, although it might be desirable for other applications. In essence the reconstruction is done with a resolution which is at least locally to high. Accordingly, the grid has to be included into the Bayesian analysis, either by choosing the best grid via model selection or by marginalizing over all possible grids. To achieve local smoothness on a coarse grid, we interpolate between the design points by cubic splines. The advantage of interpolating with splines instead of step-functions or straight lines is obvious, the number of degrees of freedom to represent physical structures is small while the noise requires very many design points. In this sense the noise is well separated from the signal and can unambiguously be erased by Ockham's factor.

Another way to incorporate local smoothness into the Bayesian image reconstruction has been proposed by Silver et al. [3]. The present form of this approach does however not consider the adaptive resolution concept. An approach to treat adaptive resolution has been suggested by Puetter et al. [4] in their pixon scheme. This approach has a couple of ad hoc elements and rests on a particular form of prior information. An approach closely related to the one by Puetter et al. which uses adaptive kernels and which uses consistently Bayesian probability theory is given somewhere else in these proceeding[5].

2. The Adaptive Spline Basis Approach

The goal of the present paper is to add local smoothness and the adaptive resolution without changing the prior otherwise. The adaptiveness can easily be achieved upon allowing the grid to vary locally. The number of design points "n" can vary between 2 and N_g where a reasonable upper limit N_g is given by the number of data points N_d. The position of the design points ξ_i can be anywhere within the image bounds, say $\xi_i \in [a, b]$. We will see that the Ockham factor in Bayesian probability theory will keep the number of design points as small as consistent with the data constraints, following Ockham's maxim: entities should not be multiplied without necessity. How do we incorporate local smoothness? A measure for smoothness is the global second derivative of the image $\int_a^b f''(x)^2 dx$. If the image is supported at the design points and the global second derivative is minimized additionally than the solution is a cubic spline. For details about spline-smoothing and related ideas see [6, 7, 8]. Our procedure is as follows: given the number "n", the position of the design points $\vec{\xi}$, and the respective value of the image at the design points, say $\vec{y}$, the desired image is the cubic spline:

$$f(x) = f^s(x|n, \vec{\xi}, \vec{y}) \tag{1}$$

which passes through the points (ξ_i, y_i) and which is twice continuously differentiable in the interval $[a, b]$. As usual a cubic spline is not unique. Here we enforce zero slope at both ends $f'(a) = f'(b) = 0$. Other applications may need different boundary conditions, like in natural splines where the second derivative vanishes at the boundaries $f"(a) = f"(b) = 0$. What follows is closely related to the analysis of Bretthorst used for spectrum analysis and parameter estimation [9]. The model (grid) $\mathcal{M}$ is characterized by the set $n, \vec{\xi}$. The Bayesian approach consists of two steps. First model selection based upon the probability for the model. Second, given the most probable model the associated image is determined. Or alternatively, if the probability for the optimal model is not strongly peaked a summation over model space is carried out.

We start with the probability for the model given the data and background knowledge.

$$p(\mathcal{M}|\mathcal{D}, \mathcal{I}) = \frac{p(\mathcal{M}|\mathcal{I}) \int p(\mathcal{D}|\mathcal{M}, \vec{y}, \mathcal{I}) p(\vec{y}|\mathcal{M}, \mathcal{I}) d^n y}{p(\mathcal{D}|\mathcal{I})} \quad . \tag{2}$$

Let us discuss the various factors. The likelihood term $p(\mathcal{D}|\mathcal{M}, \vec{y}, \mathcal{I}) = p(\mathcal{D}|f^s, \mathcal{I})$ describes as usual the error statistics of the experiment (the information about the errors is contained in I) assuming the true image is given as $f(x) = f^s(x|n, \vec{\xi}, \vec{y})$. The normalization $p(\mathcal{D}|\mathcal{I})$ is unimportant for the present purpose. The prior for the model $p(\mathcal{M}|\mathcal{I})$ factors into the prior for the model order $p(n|\mathcal{I})$ and the prior for the design points $p(\vec{\xi}|n,\mathcal{I})$. The uninformative priors in both cases are

$$p(n|\mathcal{I}) = \begin{cases} 1/(N_g-1) & \text{for } 2 \le n \le N_g \\ 0 & \text{otherwise} \end{cases} \quad , \tag{3}$$

and

$$p(\vec{\xi}|n,\mathcal{I}) = const = \frac{1}{(b-a)^n} \quad . \tag{4}$$

Hence

$$p(\mathcal{M}|\mathcal{I}) = \frac{1}{(N_g-1)(b-a)^n} \tag{5}$$

Later on we will discretize the x-values on a fine grid with a sufficiently large number N of grid points. This is a convenient but by no means necessary restriction. In this case the design points can take the values $\xi_i \in \{1, 2, \ldots, N\}$ and we can order them $\xi_1 < \xi_2 < \cdots < \xi_n$. In the discretized case we have

$$p(\mathcal{M}|\mathcal{I}) = \frac{1}{(N_g-1)\binom{N}{n}} \tag{6}$$

In the spline basis $f(x)$ depends linearly upon the spline coefficients

$$f(x) = \sum_{\nu=1}^{n} \mathcal{S}_{\nu(x,\mathcal{M})} y_\nu \quad . \tag{7}$$

In the following we consider only the discrete case and therefore $f(x) \to f_i = f(x_i)$ and $\mathcal{S}_\nu(x,\mathcal{M}) \to S_{i\nu}(\mathcal{M})$ leading in matrix notation to

$$\vec{f} = \mathcal{S}(\mathcal{M})\vec{y} \quad . \tag{8}$$

Finally, we need the prior for $\vec{y}$. Here the connection to the original prior $p_f(\vec{f})$ for the image $\vec{f}$ enters via $p(\vec{y}|\mathcal{M},\mathcal{I}) \propto p_f(\vec{f}(\vec{y}))$ which is typically of the form

$$p(\vec{y}|\mathcal{M},\mathcal{I}) = \frac{e^{\phi(\vec{f}(\vec{y}))}}{Z(n,\vec{\xi})} \quad , \tag{9}$$

where $Z(n,\vec{\xi})$ represents the normalization which depends explicitly upon the number of design points. The normalization of $Z(n,\vec{\xi})$ along with the n-dependence of $p(\mathcal{M}|\mathcal{I})$ forms the Ockham factor which tends to keep n as small as possible. We can now return to Eq.2 and determine the most probable model $\mathcal{M}$, i.e. the model order and the design points. This is more complicated as it looks at first glance as all model orders and for given model order and all design point arrangements have to be scanned to find the most probable model. The number of different models is about 2^N. We used the simulated annealing technique (for reference see e.g. [10]) to find the optimum solution in this high dimensional space. The annealing was performed at a rate $T_{n+1} = 0.95 T_n$. We chose the following elementary moves a) shift of one design point or b) creation, annihilation of a design point. Moves of type a) don't change the number of design points while those of type b) lead to a change by ± 1. The number of equilibration steps per temperature was chosen equal to the average number of models that could be reached by the moves. Once the most probable model $\mathcal{M}$ is determined, we can evaluate the posterior probability for the image $p(\vec{f}|\mathcal{D},\mathcal{M},\mathcal{I})$ given data and model. To this end we introduce the spline coefficients $\vec{y}$ in the sum rule.

$$p(\vec{f}|\mathcal{D},\mathcal{M},\mathcal{I}) = \int p(\vec{f}|\mathcal{D},\mathcal{M},\vec{y},\mathcal{I})p(\vec{y}|\mathcal{D},\mathcal{M},\mathcal{I})d^N f \quad . \tag{10}$$

According to the background knowledge I, knowing $\mathcal{M}$ and $\vec{y}$ implies that we know $\vec{f} = \vec{f}^s(n,\vec{\xi},\vec{y})$, i.e. $p(f|\mathcal{D},\mathcal{M},\vec{y},\mathcal{I}) = \delta(\vec{f} - \mathcal{S}(\mathcal{M})\vec{y})$. In other words, the image is restricted to the subspace spanned by cubic splines supported at the design points. Moreover, we employ Bayes theorem to rewrite the second factor in Eq.10 and obtain

$$p(\vec{f}|\mathcal{D},\mathcal{M},\mathcal{I}) = \int \delta(\vec{f} - \vec{f}^s(\mathcal{M}\vec{y})\frac{p(\mathcal{D}|\vec{f}^s(\mathcal{M},\vec{y}),\mathcal{I})p(\vec{y}|\mathcal{M},\mathcal{I})}{p(\mathcal{D}|\mathcal{M},\mathcal{I})}d^n y \quad . \tag{11}$$

The three factors besides the delta term are known by now, namely the likelihood term, the prior for the spline coefficients, and the evidence for the model.

Two limiting cases can be easily derived: if the data constraints are much stronger than Ockham's factor then $n \to N_d$ and $\vec{f} \to \vec{y}$ and the introduction of the spline basis has no influence at all. In the opposite limit of very weak data constraints, n will become 2 and the image will be flat. The real world applications are somewhere in between.

The Bayesian analysis yields the entire probability distribution which is however somewhat complicated due to the presence of the delta function. This is of minor importance since our interest usually focuses on expectation values of some functionals of $\vec{f}$, say $g(\vec{f})$. From Eq.10 we obtain

$$\begin{aligned} E(g(\vec{f})|\mathcal{DM}I) &= \int g(\vec{f})p(\vec{f}|\mathcal{D},\mathcal{M},\mathcal{I})d^N f & (12)\\ &= \int g(\vec{f}^s(\mathcal{M}\vec{y}))\frac{p(\mathcal{D}|\vec{f}^s(\mathcal{M},\vec{y}),\mathcal{I})p(\vec{y}|\mathcal{M},\mathcal{I})}{p(\mathcal{D}|\mathcal{M},\mathcal{I})}d^n y & (13)\\ &= \frac{\int g(\vec{f}^s(\mathcal{M},\vec{y}))p(\mathcal{D}|\vec{f}^s(\mathcal{M},\vec{y}),\mathcal{I})p(\vec{y}|\mathcal{M},\mathcal{I})d^n y}{\int p(\mathcal{D}|\vec{f}^s(\mathcal{M},\vec{y}),\mathcal{I})p(\vec{y}|\mathcal{M},\mathcal{I})d^n y} \quad . & (14) \end{aligned}$$

Before we continue we comment on possible forms of ignorant priors $p(\vec{y}|\mathcal{M},\mathcal{I})$. The prior probability is generally assigned by the maximum entropy (ME) principle [11]. Typically one has testable information for the image in form of expectation values $E(\phi(f)) = \int p(f|\mathcal{I})\phi(f)\mathcal{D}f = c$. Examples are the average signal power [9] $\phi(f) = \int f(x)^2 dx$, global curvature $\phi(f) = \int f(x)''^2 dx$, entropy $\phi(f) = \int f(x) - m(x) - f(x)\log(f(x))dx$, or Fisher information $\phi(f) = \int f'(x)^2/f(x)dx$. The ME assignment [11] of the prior probabilities yields in all cases

$$p(f|I\lambda) = \frac{1}{Z(\lambda)} e^{-\lambda\phi(f)} \qquad , \tag{15}$$

where the Lagrange parameter λ is adjusted such as to fulfill the constraint given by the testable information. In practice, however, this constraint is only known in principle. It could be measured, which makes it a testable information and it is a characteristic quantity of the image, but we do not know its value, which makes it a hyper-parameter of the theory. The correct way to cope with such nuisance parameters is to integrate them out yielding the desired parameter-free prior

$$p(f|\mathcal{I}) = \int p(f|\lambda, \mathcal{I}) p(\lambda|\mathcal{I}) d\lambda \tag{16}$$

The correct prior for the scale parameter λ is according to Jeffreys $p(\lambda|\mathcal{I}) = c/\lambda$. Using the transformation given in Eq.7 along with the correct normalization we obtain $p(\vec{y}|\mathcal{M}, \mathcal{I}) \propto \tilde{\phi}(\vec{y})^{n/2}$. We see that the prior introduces a divergence if ϕ becomes zero for some $\vec{y}$, which is the case for the above mentioned examples. This divergence originates from the fact that the integral over λ contains infinitely strong prior information, which can never be over-ruled by noisy data constraints [12]. This divergence is usually regularized tacitly upon employing the steepest descent, evidence, or similar approximations. Here we avoid this problem by the ubiquitous steepest descent approximation[13], in which the $\vec{y}$-dependence of the integrand in Eq.14 is approximated by a multi-variate normal distribution

$$\begin{aligned} P &:= p(\mathcal{D}|\vec{f}^s(\mathcal{M}, \vec{y}), \mathcal{I}) p(\vec{y}|\mathcal{M}, \mathcal{I}) \\ &= p(\mathcal{D}|\vec{f}^s(\mathcal{M}, \vec{y}^*), \mathcal{I}) p(\vec{y}^*|\mathcal{M}, \mathcal{I}) \cdot e^{-\frac{1}{2}\Delta\vec{y}^T H \Delta\vec{y}} \end{aligned} \tag{17}$$

which best describes the maximum, i.e. $\vec{y}^*$ is the position of the maximum of P, $\Delta\vec{y} = \vec{y} - \vec{y}^*$, and the Hessian $H_{ij} = -\frac{\partial^2 \log(P)}{\partial y_i \partial y_j}$. The integral in Eq.14 can now be evaluated easily yielding

$$E(g(\vec{f})|\mathcal{D}\mathcal{M}I) = \frac{\int g(\vec{f}^s(\mathcal{M}\vec{y})) e^{-\frac{1}{2}\Delta\vec{y}^T H \Delta\vec{y}} d^n y}{(2\pi)^{n/2} \det(H)^{-1/2}} \tag{18}$$

Also the probability for the model (Eq.2) simplifies considerably

$$p(\mathcal{M}|\mathcal{D}, \mathcal{I}) = \frac{(2\pi)^{n/2} p(\mathcal{M}|\mathcal{I}) p(\mathcal{D}|\mathcal{M}, \vec{y}^*, \mathcal{I}) p(\vec{y}^*|\mathcal{M}, \mathcal{I})}{p(\mathcal{D}|\mathcal{I}) \det(H)^{1/2}} \qquad . \tag{19}$$

In the following example, we summarize the results by the posterior mean and the posterior variance

$$E(f_i|\mathcal{M}\mathcal{D}I) = \sum_l \mathcal{S}_{il} E(y_l|\mathcal{M}\mathcal{D}I) = \sum_l \mathcal{S}_{il} y_l^* \tag{20}$$

$$E(\Delta^2 f_i|\mathcal{M}\mathcal{D}I) = \sum_{ll'} \mathcal{S}_{il}\mathcal{S}_{il'} E(\Delta y_l \Delta y_l'|\mathcal{M}\mathcal{D}I) = \sum_{ll'} \mathcal{S}_{il}\mathcal{S}_{il'} H_{ll'} \tag{21}$$

This completes the second step of the Bayesian approach. If the probability for the model $\mathcal{M}$ in Eq.19 is not sharply peaked at one particular model, we have to do the full summation over all models with the corresponding probability

$$E(g(\vec{f})|\mathcal{D}, I) = \sum_{\mathcal{M}} p(\mathcal{M}|\mathcal{D}, \mathcal{I}) \cdot E(g(\vec{f})|\mathcal{D}, \mathcal{M}, I) \tag{22}$$

In this case the most effective way to perform the summation is by Monte Carlo importance sampling (a detailed discussion is given in [10]). The same elementary moves are used for the random walk as in simulated annealing. The moves were chosen such as to ensured detailed balance.

3. Applications

We start out with the mock data discussed in the introduction which describe a situation encountered in spectroscopy, a signal composed of a single particle peak and sharp onset (step) of some continuum. The latter has been exaggerated to study the worst conceivable case. The spectral density is blurred by a Gaussian apparatus function and corrupted by noise. A Gaussian likelihood term was used. As already discussed, the lower panel of fig.1 depicts the standard ME result using all standard approximations, namely steepest descent to determine the probability for the regularization parameter, entering the evidence approximation. In addition, the same panel contains the full Bayesian result without any approximations obtained upon marginalizing over the regularization parameter as proposed by Strauss et al. [14] and elaborated by us [2]. The key message of fig.1 is that the standard approximations tend to overfit the data and enhance ringing. But even the full Bayesian evaluation of ME has similar unpleasant features. For details we refer the reader to [2]. As pointed out before this is due to the fact that a rigid image-grid was used which contains more degrees of freedom than the information content of the data can honestly assign. Including the grid into the Bayesian analysis reveals that the reliable number of degrees of freedom is merely about 10 as compared to 100 used before. The upper panel of fig. 1 illustrates that the adaptive spline approach entirely eliminates the over-fitting, i.e. the result is robust against noise corruption. The reason is that many data points contribute to the area in one grid interval improving the statistics considerably. Also the ringing, which is generally observed in inverse problems at sudden structures, has been suppressed strongly by the adaptive spline approach. This example is certainly the worst case ever to be encountered in spectrum restoration since real spectra don't exhibit such strong discontinuities. The inversion has been performed using the entropic prior for $\vec{y}$ derived from the entropic prior for $\vec{f}$. We also used the Gaussian and global curvature prior and found only little differences in the images. The reason is that in this example the effective number of degrees of freedom is small as compared to the number of data, which strongly reduces the effective noise level and amplifies the importance of likelihood term over the prior.

As another example we applied our approach to the Old Faithful density estimation problem which has become the standard problem to test density estimation

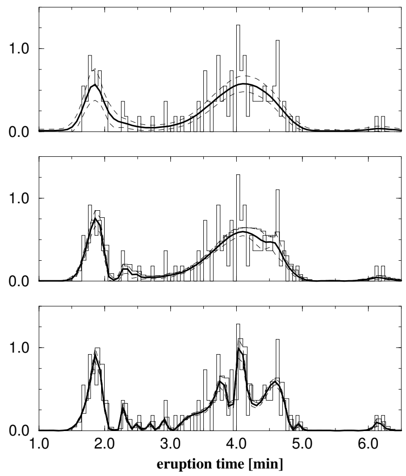

Figure 2: *Density estimation of Old Faithful eruption times. The histogram represents the original data, full curves represent the result of the adaptive spline approach along with the error band (dashed line). The upper panel contains the result for the standard data set as listed in [??]. The total number of counts is 109. The middle (lower) panel illustrates how the result changes if the number of counts is increase by a factor 5 (10) while retaining the original histogram.*

schemes [15, 16, 17]. It should be noted that the essential point of our approach is the adaptive grid and local smoothness which can be added to any prior one considers adequate for the problem under consideration. Here we used the Gaussian prior. As elaborated upon by Sibisi and Skilling[17] the prior for the density estimation problem is correctly described by the Dirichlet process. But as pointed out in the first example, the results depend only slightly upon the prior once the adaptive grid is used. To ensure positivity in this case, the density $\rho(t)$ for the eruption times t is expressed in terms of an auxiliary function $f(t)$ as $\rho(t) = f^2(t)$ and the latter is expressed in the spline basis.

The present example is one of the rare cases where the likelihood is not a Gaussian but rather given by

$$p(D|\rho, \mathcal{I}) = \prod_{i=1}^{L} p(t_i|\rho, \mathcal{I}) = e^{L \int \rho_0(t) \log(\rho(t)) dt} \qquad . \tag{23}$$

Uncorrelated measurements are assumed. The maximum likelihood solution subject to the normalization constraint $\int \rho(t) dt = 1$ is

$$\rho(x) = \rho_0(t) = \frac{1}{L} \sum_{i=1}^{L} \delta(t - t_i) \qquad . \tag{24}$$

In this example, the smoothness constraint is particularly important. Fig.2 shows the density estimation result for Old Faithful eruption times using the adaptive spline approach. According to Eq.23 the weight of the likelihood term is determined by the total counts L. To illustrate the interplay between data-constraint and prior, we modified the total counts while keeping the histogram fixed. We can see nicely how increasing the total counts and hence the data-constraints pulls the Bayesian solution towards the maximum likelihood solution.

Finally we want to mention that both examples have also been studied in the frame of the adaptive Kernel approach [5]. The results are remarkable agreement.

4. Summary

We presented a Bayesian analysis for image reconstruction which augments the usual approaches by the concept of adaptive resolution and local smoothness. We can now honestly use the statement, which was deceivingly used for maxent before, namely that the Bayesian approach allowing for adaptive resolution and local smoothness yields the most uncommittal reconstruction compatible with the data. All structures in the image are supported by the data constraints. This has been demonstrated by two examples of the field of spectral deconvolution and density estimation.

References

1. W. von der Linden, R. Fischer, and V. Dose, "Evidence integrals," in *Maximum Entropy and Bayesian Methods*, R. Silver and K. Hanson, eds., Kluwer Academic Publishers, Dordrecht, 1996. to be published.
2. R. Fischer, W. von der Linden, and V. Dose, "On the importance of α marginalization in maximum entropy," in *Maximum Entropy and Bayesian Methods*, R. Silver and K. Hanson, eds., Kluwer Academic Publishers, Dordrecht, 1996. to be published.
3. R. Silver, "Quantum statistical inference," in *Maximum Entropy and Bayesian Methods*, G. Heidbreder, ed., Kluwer Academic Publishers, Dordrecht, 1996. to be published.
4. R. Puetter, "Pixon-based multiresolution image reconstruction and the quantification of picture information content," *Int. J. Image Sys. & Tech.*, **6 (4)**, p. 314, 1995.
5. R. Fischer, W. von der Linden, and V. Dose, "Adaptive kernals and occam's razor in inversion problems," in *Maximum Entropy and Bayesian Methods (1996)*, J. S. S. Sibisi, ed., Kluwer Academic Publishers, Dordrecht, 1997. to be published.
6. B. W. Silverman, "Some aspects of the spline smoothing approach to non-parametric regression curve fitting," *J.R.Statist.Soc.B*, **47**, pp. 1–52, 1985.
7. C. Gu, "Smoothing, spline density estimation," *J. Amer. Statist. Ass.*, **88**, pp. 495–504, 1993.
8. D. Keren and M. Werman, "A full bayesian approach to curve and surface reconstruction," in *Maximum Entropy and Bayesian Methods 1995*, R. Silver and K. Hanson, eds., Kluwer Academic, Dordrecht, to appear.
9. G. Bretthorst, "Excerpts from bayesian spectrum analysis and parameter estimation," in *Maximum Entropy and Bayesian Methods in Science and Engineering*,

G. Erickson and C. Smith, eds., vol. 1, p. 75, Kluwer Academic Publishers, Dordrecht, 1988.
10. J. Ruanaidh and W. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Springer, New York, 1996.
11. R. D. Rosenkrantz, ed., *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, 1983.
12. S. F. Gull, "Bayesian inductive inference," in *Maximum Entropy and Bayesian Methods*, ?, ed., p. ?, Kluwer Academic Publishers, Dordrecht, ?
13. S. Gull, "Developments in maximum entropy data analysis," in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., pp. 53–71, Kluwer Academic Publishers, Dordrecht, 1989.
14. C. Strauss, D. Wolpert, and D. Wolf, "Alpha, evidence, and the entropic prior," in *Maximum Entropy and Bayesian Methods*, G. Heidbreder, ed., Kluwer Academic Publishers, Dordrecht, 1996. to be published.
15. B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.
16. R. Silver, H. Martz, and T. Wallstrom, "Quantum statistical inference for density estimation," *J. Amer. Statist. Ass.*, p. x, 1993.
17. S. Sibisi and J. Skilling, "Bayesian density estimation," *preprint*, 1994.