Contribution to the Twentieth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2000), CNRS, Gif sur Yvette, France, July 8-13, 2000

BAYESIAN FIELD THEORY AND APPROXIMATE SYMMETRIES

Applications to Density Estimation, Regression, and Inverse Problems in Quantum Mechanics

J.C. LEMM
Institut für Theoretische Physik
Wilhelm–Klemm–Str.9
D-48149 Münster†

Abstract. Nonparametric Bayesian approaches to density estimation ("Bayesian field theories") have typically to be solved numerically on a lattice. This is often numerically quite expensive. The Paper wants to show that such numerical calculations are nowadays feasible for some interesting problem classes. In particular, Bayesian field theories are defined by 1. a likelihood model, being a probabilistic description of the measurement process of observational data. and 2. a prior model, determining the generalization behavior of the theory by implementing available *a priori* knowledge. In this Paper some variations of prior and likelihood models are discussed: First, the implementation of approximate symmetries with Gaussian process priors is demonstrated for approximate periodic and for approximate fractal functions. Second, besides a discussion of the classical likelihood models of general density estimation and regression, special emphasis is put on the likelihood model of quantum theory to treat the inverse problem of reconstructing potentials of quantum systems from a finite number of observational data.

Key words: Nonparametric Bayesian Inference, Approximate Symmetries, Inverse Quantum Theory

1. Introduction to Bayesian Field Theory

A Bayesian approach to empirical learning problems requires the definition of two probabilistic models [1,2]: 1. a (conditional) *likelihood model* $p(y|x,\phi)$ representing the probability of measuring y given visible variables x (conditions) and a hidden state of Nature ϕ. 2. a *prior model* $p(\phi)$ representing the *a priori* information available on ϕ. One can choose the visible variables x and y so that the joint probability factorizes according to

$$p(x,y,\phi) = p(y|x,\phi)\, p(x)\, p(\phi). \tag{1}$$

†Email: lemm@uni-muenster.de

We are interested in empirical learning based on n independent (training) data $D = \{(x_i, y_i)|1 \le i \le n\}$ with $p(y_D|x_D, \phi) = \prod_{i=1}^n p(y_i|x_i, \phi)$ and $p(x) = \prod_{i=1}^n p(x_i)$ where y_D (x_D) denotes the vector of training data y_i (x_i). In *nonparametric* approaches, the function values $P(x, y) = p(y|x, \phi)$ itself, specifying a state of Nature ϕ, are considered as primary degrees of freedom, their only constraint being the positivity and normalization conditions for $p(y|x, \phi)$. In general, likelihoods $P(x, y)$ can be considered which are functionals of functions (or *fields*) $\phi(x, y)$, i.e., $P = P(\phi)$.

Different classes of likelihood functions define different problem types. The most general problem type, which we will discuss first, is that of density estimation[3]. The other problem types listed in Tab. 1 can be seen as special cases thereof. In regression problems, for instance, the likelihood model is assumed Gaussian.

Examples of functionals $P(\phi)$ are listed in Tab. 2 where the first and third row show the 'local' functionals $P(\phi)$ corresponding to $\phi(x, y) = P(x, y)$ and $\phi(x, y) = \ln P(x, y)$, the other functionals are nonlocal in y. Depending on the choice of the functional $P(\phi)$ the normalization and positivity constraint for P take different forms for ϕ. For example, the normalization condition is automatically fulfilled in rows two and four of Tab. 2. In the last row, where $\phi(x, y) = \int_{-\infty}^{y} dy'\, P(x, y')$, the normalization constraint for P becomes the boundary condition $\phi(x, \infty) = 1$, and positivity of P means monotony of $\phi(x, \cdot)$.

In this paper, our aim will be to reconstruct the field ϕ from observational data D in *Maximum A Posteriori Approximation* (MAP), i.e, by maximizing

$$p(\phi|D) \propto p(y_D|x_D, \phi)\, p(\phi), \tag{2}$$

with respect to ϕ. [In many cases one can alternatively use Monte–Carlo methods to approximate the Bayesian predictive density $p(y|x, D)$.] If not automatically satisfied by the choice of ϕ, the normalization conditions on P (for all x) can be implemented by adding a Lagrange multiplier function $\Lambda_X(x)$. Similarly the positivity constraint can be fulfilled. For nonzero probabilities, however, positivity is not an active constraint at the stationary point, so a corresponding Lagrange multiplier term is often not necessary. Introducing a Gaussian process prior term [4] for ϕ (with zero mean and inverse covariance $\mathbf{K}_0$, real symmetric, positive definite), maximizing the posterior (2) is equivalent to minimizing the error or energy functional,

$$E_\phi = -<\ln P(\phi)\,|\,N> + \frac{1}{2}<\phi\,|\,\mathbf{K}_0\phi> + <P(\phi)\,|\,\Lambda_X>. \tag{3}$$

Here N denotes the empirical density, i.e., $N(x, y) = \sum_{i=1}^n \delta(x - x_i)\delta(y - y_i)$, and the bra–ket notation $<\cdot\,|\,\cdot>$, denoting scalar products, stands for integration over x and y. (Prior terms can be made more flexible by including hyperparameters [1,5,6]. Notice that a prior which is Gaussian in terms of the field ϕ is in general non–Gaussian in terms of the conditional likelihood P.

From Eq. (3) the MAP stationarity equation to be solved is found by setting the functional derivative of (3) with respect to $\phi(x, y)$ to zero, yielding,

$$0 = \mathbf{P}'(\phi)\mathbf{P}^{-1}(\phi)N - \mathbf{K}_0\phi - \mathbf{P}'(\phi)\Lambda_X, \tag{4}$$

where $\mathbf{P}'(x,y;x',y') = \frac{\delta P(x',y')}{\delta\phi(x,y)}$ and $\mathbf{P}(x,y;x',y') = \delta(x-x')\delta(y-y')P(x,y)$. For existing $\mathbf{P}\mathbf{P}'^{-1}$ =$(\mathbf{P}'\mathbf{P}^{-1})^{-1}$ this can be written

$$0 = N - \mathbf{P}\mathbf{P}'^{-1}\mathbf{K}_0\,\phi - \mathbf{P}\Lambda_X. \tag{5}$$

From the normalization condition $\mathbf{I}_X P = I$ [where $\mathbf{I}_X(x,y;x',y') = \delta(x-x')$ and $I(x,y) = 1$], it follows for $\Lambda_X(x,y) = \Lambda_X(x) \neq 0$ that $\mathbf{I}_X\mathbf{P}\Lambda_X = \Lambda_X$, and thus, multiplying Eq. (5) by $\mathbf{I}_X$,

$$\Lambda_X = \mathbf{I}_X\left(N - \mathbf{P}\mathbf{P}'^{-1}\mathbf{K}_0\,\phi\right) = N_X - \mathbf{I}_X\mathbf{P}\mathbf{P}'^{-1}\mathbf{K}_0\,\phi, \tag{6}$$

where $N_X(x,y) = N_X(x) = \sum_i \delta(x-x_i)$. Λ_X can now be eliminated by inserting Eq. (6) into Eq. (5)

$$0 = (\mathbf{I} - \mathbf{P}\mathbf{I}_X)\left(N - \mathbf{P}\mathbf{P}'^{-1}\mathbf{K}_0\,\phi\right), \tag{7}$$

where $\mathbf{I}$ stands for the identity. In general, the solution of Eq. (7) cannot be found in a finite dimensional space defined by the training data D as it is the case for regression problems with Gaussian process priors [7] (where MAP is exact) and similarly for classification with a specific choice of $P(\phi)$ [8], Thus, Eq. (7) has to be solved by discretization, analogously, for example, to solving field theories on a lattice [9]. Starting with an initial guess $\phi^{(0)}$, a possible iteration scheme is given by

$$\phi^{(r+1)} = \phi^{(r)} + \eta\mathbf{A}^{-1}\left[\left(\mathbf{I} - \mathbf{P}^{(r)}\mathbf{I}_X\right)\left(N - \mathbf{P}^{(r)}\left(\mathbf{P}'^{(r)}\right)^{-1}\mathbf{K}_0\,\phi^{(r)}\right)\right], \tag{8}$$

with some positive definite matrix $\mathbf{A}$, and a number $\eta > 0$, both also possibly changing during the iteration. A convenient r–independent choice for $\mathbf{A}$ is often the inverse covariance $\mathbf{K}_0$. (Quasi–Newton methods, for example, use an r–dependent $\mathbf{A}^{(r)}$, the standard gradient algorithm corresponds to choosing for $\mathbf{A}$ the identity matrix.) At each iteration step the factor η can be determined by a one–dimensional line search algorithm. An iteration scheme similar to Eq. (8), can be obtained using Eq. (4), which then requires to adapt the Lagrange multiplier function $\Lambda_X(x)$ during iteration. An example of the numerical reconstruction of a density $P(x,y)$ with a smoothness prior on $\ln P$ is shown in Fig. 1.

Without further approximations a direct discretization is only feasible for low dimensional x and y variables (say one– or two–dimensional, like time series or images). For higher dimensional x or y variables additional approximations are necessary like for instance mean field approaches [10].

2. Approximate symmetries

2.1. GENERAL CASE

Often *a priori* information can be formulated as approximate invariance under certain symmetry operations. Smoothness, for instance, means approximate symmetry under infinitesimal translations. Similarly, one can consider approximate

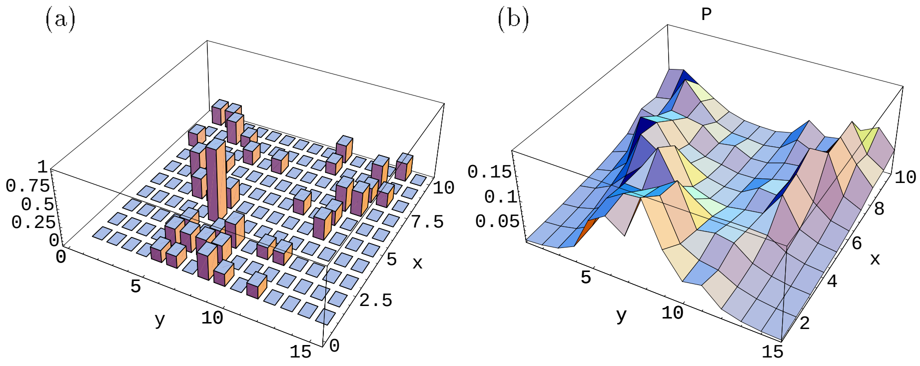

Figure 1. Reconstruction of a conditional likelihood using a Gaussian smoothness prior on $\phi(x,y) = \ln P(x,y)$ with Laplacian inverse covariance in x and y: $\mathbf{K}_0 = -\Delta_x - 0.5\Delta_y$ with $\Delta_x = \partial^2/\partial x^2$ and $\Delta_y = \partial^2/\partial y^2$. (a) 50 data points shown as conditional empirical density $P_{\rm emp} = \sum_{i=1}^n \delta(x-x_i)\delta(y-y_i)/\sum_{i=1}^n \delta(x-x_i)$ (b) Reconstructed conditional density $P(x,y) = p(y|x,\phi)$. (Details of the calculation can be found in [6].)

problem type	(conditional) likelihood $p(y\vert x,\phi)$
density estimation	of general form
classification	discrete y
regression	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[y-\phi(x)]^2}{2\sigma^2}}$ (Gaussian, σ fixed)
clustering	$\sum_k \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{[y-\phi_k(x)]^2}{2\sigma_k^2}}$ (mixture of Gaussians)
inverse quantum mechanics	$\text{Trace}(\rho(\phi)\,\pi_{(y,x)})$

TABLE 1. Special cases of density estimation. For classification and regression see [7,8,11,6] and references therein, for clustering see [12], and for Bayesian approaches to inverse quantum mechanics aiming for example in the reconstruction of potentials from observational data, see Section 3.

symmetry under rotations, approximate periodicity, or approximate scale invariance. In this Section we will show how general approximate symmetries can be implemented by choosing a Gaussian process prior with appropriate covariance operator.

While exact invariance of ϕ under a symmetry operation $\mathbf{S}$ means $\phi = \mathbf{S}\phi$, we

$P(\phi)$	constraints	
$P(x,y)=\phi(x,y)$	norm	positivity
$P(x,y)=\frac{\phi(x,y)}{\int \phi(x,y')\,dy'}$	—	positivity
$P(x,y)=e^{\phi(x,y)}$	norm	—
$P(x,y)=\frac{e^{\phi(x,y)}}{\int e^{\phi(x,y')}\,dy'}$	—	—
$P(x,y)=\frac{d\phi(x,y)}{dy}$	boundary	monotony

TABLE 2. Constraints in density estimation for some specific choices of the field ϕ.

are in the following interested in the case of an approximate invariance

$$\phi \approx \mathbf{S}\phi. \tag{9}$$

To find a prior which generates such functions, consider the inverse covariance

$$\mathbf{K}_S = (\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S}), \tag{10}$$

where $\mathbf{I}$ denotes the identity operator and the superscript ^{T} the transpose. Clearly, all functions with exact symmetry under $\mathbf{S}$ are zero modes of $\mathbf{K}_S$. Furthermore, a prior energy

$$E_S = \frac{1}{2}\!<\phi-\mathbf{S}\phi\,|\,\phi-\mathbf{S}\phi\!> = \frac{1}{2} <\phi\,|\,\mathbf{K}_S\,|\,\phi>, \tag{11}$$

is small for approximately symmetric functions. Thus, a Gaussian prior

$$p(\phi) \propto \exp\{-E_S(\phi)\} \tag{12}$$

can be used to generate functions which are approximately invariant under $\mathbf{S}$.

Similarly, continuously parameterized transformations correspond to Lie groups $\mathbf{S}(\theta) = \exp(\theta\mathbf{s})$, where θ is a real parameter (vector) and $\mathbf{s} = -\mathbf{s}^T$ an antisymmetric (vector) operator representing the generator(s) of the infinitesimal transformation(s) of the Lie–group. A prior energy as an error measure with respect to an infinitesimal transformation can be defined as,

$$E_S = \frac{1}{2}\!<\frac{\phi-(1+\theta\mathbf{s})\phi}{\theta}\,|\,\frac{\phi-(1+\theta\mathbf{s})\phi}{\theta}\!> = \frac{1}{2} <\phi\,|\,\mathbf{s}^T\mathbf{s}\,|\,\phi>, \tag{13}$$

yielding an inverse prior covariance $\mathbf{K}_s = \mathbf{s}^T\mathbf{s}$. For example, for infinitesimal translations where $\mathbf{s} = \nabla$ this yields, assuming vanishing boundary terms, the common Laplacian smoothness term $-\frac{1}{2} <\phi\,|\,\Delta\,|\,\phi>$.

2.2. APPROXIMATE PERIODICITY

Analogously to smoothness, i.e., approximate invariance under infinitesimal translations, one can implement approximate periodicity, i.e., approximate invariance under a specific finite translation,

$$\mathbf{S}_\theta \phi(x) = \phi(x-\theta) \approx \phi(x). \tag{14}$$

The corresponding prior energy is

$$E_S = \frac{1}{2}\int dx\, |v(x) - v(x+\theta)|^2 = \frac{1}{2} < \nabla_\theta^R \phi \,|\, \nabla_\theta^R \phi > = -\frac{1}{2} < \phi \,|\, \Delta_\theta \,|\, \phi >, \tag{15}$$

which, assuming periodic boundary conditions, has been written using a "finite Laplacian"

$$\mathbf{K}_\theta = -\Delta_\theta = -\nabla_\theta^L \nabla_\theta^R = (\nabla_\theta^R)^T \nabla_\theta^R, \tag{16}$$

defined in terms of the difference operators

$$\left(\nabla_\theta^R \phi\right)(x) = \phi(x+\theta) - \phi(x), \quad \left(\nabla_\theta^L \phi\right)(x) = \phi(x) - \phi(x-\theta), \tag{17}$$

with $(\nabla_\theta^L)^T = -\nabla_\theta^R$. For instance, on a one–dimensional lattice (with six points and periodic boundary conditions) the "finite Laplacian" (16) takes the form

$$-\Delta_\theta = \begin{pmatrix} 2 & 0 & -1 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 & 0 & -1 \\ -1 & 0 & 2 & 0 & -1 & 0 \\ 0 & -1 & 0 & 2 & 0 & -1 \\ -1 & 0 & -1 & 0 & 2 & 0 \\ 0 & -1 & 0 & -1 & 0 & 2 \end{pmatrix}.$$

We will demonstrate the reconstruction of functions with approximate periodicity for Gaussian regression problems. In that case the likelihood $p(y|x,\phi)$ is assumed to be a Gaussian (with unknown mean or regression function $\phi(x)$ and fixed variance) and a Gaussian process prior is imposed on the regression function $\phi(x)$. The posterior energy $E(\phi|D) = -\ln p(\phi|D)$ reads for such problems

$$E(\phi|D) = \frac{1}{2\sigma^2}\sum_{i=1}^n |y_i - \phi(x_i)|^2 + \frac{\lambda}{2} < \phi \,|\, \mathbf{K}_0 \,|\, \phi > . \tag{18}$$

It is convenient to rewrite the quadratic error term

$$\frac{1}{2\sigma^2}\sum_{i=1}^n |y_i - \phi(x_i)|^2 = \frac{1}{2\sigma^2} < \phi - t_D \,|\, \mathbf{K}_D \,|\, \phi - t_D > \tag{19}$$

with inverse data covariance $\mathbf{K}_D(x,x') = \delta(x-x')\sum_{i=1}^n \delta(x-x_i)$ and mean data function t_D, representing the mean of data measured at x, i.e., $t_D(x) = \frac{1}{n_x}\sum_{i=1}^n y_i \delta(x-x_i)$ if $n_x = \sum_{i=1}^n \delta(x-x_i) \neq 0$ and $t_D(x)$ arbitrary if no data point is available for x, i.e., if $n_x = 0$.

As sum of the two quadratic functionals in (18) we obtain again a quadratic functional

$$E(\phi|D) = \frac{\lambda}{2} < \phi - t \,|\, \mathbf{K} \,|\, \phi - t > + c, \tag{20}$$

where c is a ϕ–independent constant. Hence, the regression function with maximal posterior probability (which in this Gaussian case corresponds to the mean of the predictive density) is given by

$$t = \frac{1}{\sigma^2}\mathbf{K}^{-1}\mathbf{K}_D t_D = \frac{1}{\sigma^2}\mathbf{K}_0^{-1}\mathbf{K}_{t_D} t_D, \tag{21}$$

with inverse posterior covariance $\mathbf{K} = \mathbf{K}_D/\sigma^2 + \lambda\mathbf{K}_0$, and $\mathbf{K}_{t_D} = \left(\mathbf{K}_D^{-1} + \mathbf{K}_{0,D}^{-1}\right)^{-1}$ where $\mathbf{K}_{0,D}$ represents the projection of $\mathbf{K}_0$ into that subspace where data are available.

Fig. 2a shows the numerical reconstruction of a smooth and approximately periodic regression function. Therefore 40 data points have been sampled from a true regression function $\phi_{\text{true}}(x)$ which has been drawn before from a Gaussian process with inverse covariance

$$\mathbf{K}_0 = (1-\lambda')\mathbf{K}_\theta + \lambda'\mathbf{K}_{\text{RBF}}. \tag{22}$$

Here

$$\mathbf{K}_{\text{RBF}} = \sum_{k=0}^{3}\frac{1}{k!}\left(-\frac{\sigma_0^2}{2}\Delta\right)^k \approx \exp\left(-\frac{\sigma_0^2}{2}\Delta\right) \tag{23}$$

is the inverse covariance of a truncated Radial Basis Function smoothness prior which produces smoother functions than for example a negative Laplacian [13]. Fig. 3a shows the corresponding prior covariance $\mathbf{K}_0^{-1}$.

2.3. APPROXIMATE FRACTALS

Nice looking fractal images can be produced by Multiple Reduction Copy Machines (MRCM) being examples of so called Iterated Function Systems [14,15]. Each iteration step of a MRCM consists in placing M reduced affine copies of the whole image on specific regions of the original image. A fractal image is then produced by iterating this multiple affine copy operation.

For instance, consider a one–dimensional function defined on an interval L and a multiple reduced copying operation which places M down–scaled copies of the whole interval L in M (nonoverlapping) subintervals $A_k = [a_k, b_k]$, thus reducing the interval L by scaling factors $|A_k|/|L| = \theta_k < 1$. Such an operation can be written as

$$\mathbf{C} = (\mathbf{I} - \sum_{k=1}^{M}\pi_k) + \sum_{k=1}^{M}\mathbf{C}_k\pi_k \tag{24}$$

where π_k is the projector on region A_k and each $\mathbf{C}_k$ represents a reduced copy operation of the interval L into one of the A_k, i.e.,

$$\phi'(x) = \mathbf{C}_k\phi(x) = \phi(\frac{x-a_k}{\theta_k}). \tag{25}$$

(In Eq.(24) we have chosen to leave the function unchanged in regions not contained in any of the sets A_k.) Approximate invariance under m repeated applications of the multiple reduced copy operation $\mathbf{C}$ is according to Eq. (10) implemented by the inverse covariance

$$\mathbf{K}_{\mathrm{F1}}^{(m)} = (\mathbf{I} - \mathbf{C}^m)^T (\mathbf{I} - \mathbf{C}^m). \tag{26}$$

To create a perfect fractal the operation $\mathbf{C}$ has to be iterated an infinite number of times. Clearly, on a discrete lattice a reduced copy operation can only be performed a finite number of times. In particular, down scaling a function on a finite lattice means loss of information (so the discrete version of the transformation $\mathbf{C}$ is not invertible). Thus, an effective operation $\mathbf{C}$ can be implemented in the form

$$\mathbf{C} = (\mathbf{I} - \sum_{k=1}^{M} \pi_k) + \sum_{k=1}^{M} \mathbf{C}_k \mathbf{A}_k \pi_k \tag{27}$$

where $\mathbf{A}_k$ stands for an averaging operation $\mathbf{A}_k \phi(x) = \sum_{x'} K_k(x - x')\phi(x')$ with some real, positive and normalized kernel K_k. Taking integer $1/\theta_k$, we can for example uniformly average over all the $1/\theta_k$ points of ϕ which contribute to the same point of the down–scaled function ϕ'

$$\mathbf{A}_k \phi(x) = \theta_k \sum_{x'=x}^{x-1+1/\theta_k} \phi(x'). \tag{28}$$

Due to the averaging in Eq. (27) the classical one–dimensional Cantor set (considered as binary function of x), for instance, is not invariant under $\mathbf{C}^m$ on a discrete lattice. It is however invariant under the following "bottom up" operation

$$\mathbf{K}_{\mathrm{F2}}^{(m)} = \sum_{k=1}^{M} (\mathbf{I} - \mathbf{C}^m \mathbf{B}_k)^T (\mathbf{I} - \mathbf{C}^m \mathbf{B}_k) \tag{29}$$

where $\mathbf{B}_k$ denotes the "blowing up" operation which maps the subinterval A_k to the whole interval L

$$\phi'(x) = \mathbf{B}_k\ \phi(x) = \phi(\theta_k x + a_k). \tag{30}$$

In a discrete implementation of (30) the function $\phi(\theta_k x + a_k)$ for noninteger x' $= \theta_k x + a_k$ has either to be interpolated or replaced by its value at, e.g., the nearest or next lower integer. An inverse covariance of the form (29) implements the expectation to find the same small scale structures in all subintervals A_k. To allow a stronger influence of lower iteration levels, inverse covariances of the form (29) can also be added, e.g., for all levels $1 \le l \le m$,

$$\mathbf{K}_{\mathrm{F3}}^{(m)} = \sum_{l=1}^{m} \mathbf{K}_{\mathrm{F2}}^{(l)} = \sum_{l=1}^{m} \sum_{k=1}^{M^l} (\mathbf{I} - \mathbf{C}^l \mathbf{B}_k)^T (\mathbf{I} - \mathbf{C}^l \mathbf{B}_k). \tag{31}$$

Fig. 2b shows the reconstruction of an approximately fractal regression function, where the true regression function ϕ_{true}, unknown to the reconstruction algorithm,

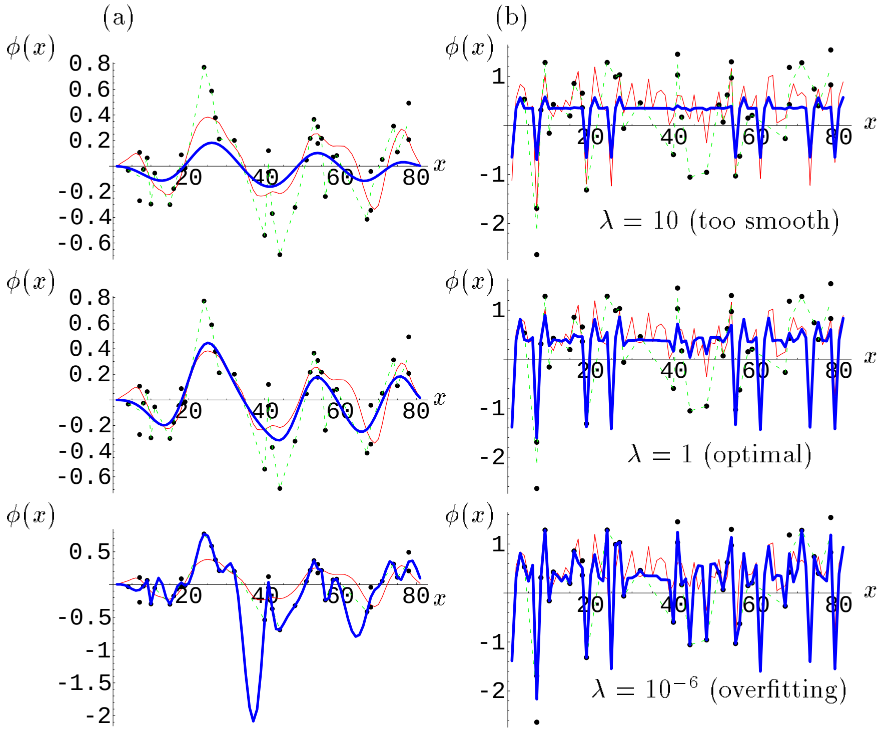

Figure 2. Reconstruction results with varying regularization factor λ [see Eq. (18)] of (a) a smooth and approximately periodic regression function, where $\mathbf{K}_0 = (1-\lambda')\mathbf{K}_\theta + \lambda'\mathbf{K}_{\rm RBF}$ as in (22) and (23) with $\lambda' = 0.1$, $\sigma_0 = 10$, and $\sigma = 0.1$, and (b) a smooth and approximately fractal regression function where $\mathbf{K}_0 = (1-\lambda')\mathbf{K}_{F3} - \lambda'\Delta$ as in (32) with $\lambda' = 0.1$. The figures show the true regression functions $\phi_{\tt true}(x)$ (thin lines, sampled from the corresponding prior covariance and unknown to the learning algorithm) and the reconstructed regression functions $\phi(x)$ (thick lines) obtained from using 40 data points (dots). (The thin dashed line depicts the piecewise linear interpolation of the data points.)

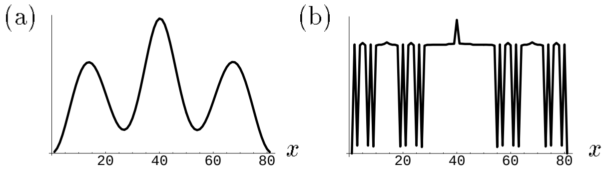

Figure 3. (a) Prior covariance $K_0^{-1}(x, x_0)$ (for $x_0 = 40$) used for Fig. 2a generating smooth and approximate periodic functions. (b) Prior covariance $K_0^{-1}(x, x_0)$ (for $x_0 = 40$) used for Fig. 2b generating smooth and approximate fractal functions.

has been sampled from a Gaussian process with an inverse prior covariance of the form

$$\mathbf{K}_0 = (1-\lambda')\mathbf{K}_{F3} - \lambda'\Delta. \tag{32}$$

The corresponding prior covariance $\mathbf{K}_0^{-1}$ is shown in Fig. 3b.

3. Inverse Quantum Theory

A special density estimation problem is the reconstruction of potentials from a finite number of observations on a quantum mechanical system. To apply the Bayesian framework to such inverse problems of quantum theory we have to recall the likelihood model of quantum theory: The probability to find value x when measuring observable O for a quantum system in a state described by density operator ρ is given by[1]

$$p(x|O,v) = \mathrm{Tr}\Big(P_O(x)\,\rho(v)\Big), \tag{33}$$

where $P_O(x) = \sum_\xi |x,\xi><x,\xi|$ is the projector on the space spanned by the orthonormalized eigenfunctions $|x,\xi>$ of O with eigenvalue x and the label ξ distinguishes eigenfunctions which are degenerate with respect to O. If the system is not in an eigenstate of the observable O, a quantum mechanical measurement changes the state of the system. Hence, to perform repeated measurements under the same ρ requires the restoration of ρ before each measurement.

In particular, consider canonical ensembles of quantum systems where

$$\rho = \frac{\exp(-\beta H)}{\mathrm{Tr}\,\exp(-\beta H)} \tag{34}$$

with Hamiltonians $H = T + V$ consisting of a kinetic energy $T = -(1/2m)\Delta$ (with mass m) and an unknown local potential $V(x,x') = v(x)\delta(x-x')$. Similarly to the Bayesian reconstruction of density operators [16–18] our aim in the following will be the reconstruction of the function $v(x)$ from observational data. For position measurements, for example, the likelihood (33) in a canonical ensemble (34) becomes

$$p(x_i|\hat{x},v) = \sum_\alpha p_\alpha |\phi_\alpha(x_i)|^2 = <|\phi(x_i)|^2>, \tag{35}$$

where $H|\phi_\alpha> = E_\alpha|\phi_\alpha>$ and $<\cdots>$ denotes the thermal expectation value with probabilities $p_\alpha = \exp(-\beta E_\alpha)/Z$ and $Z = \sum_\alpha \exp(-\beta E_\alpha)$. Like in regression models a Gaussian process prior can be used

$$p(v) = \left(\det\frac{\lambda\mathbf{K}_0}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda}{2} < v - v_0|\mathbf{K}_0|v - v_0 >\right), \tag{36}$$

where the mean function v_0 represents a reference potential or *a priori* guess for v. Similarly, more flexible prior processes can be constructed be using mixtures of Gaussian process priors [11]. To find the potential with maximal posterior probability we have according to Bayes' theorem to maximize

$$p(v|D) \propto p(v)\prod_{i=1}^n p(x_i|O_i,v). \tag{37}$$

[1]The correspondences to the notations of the previous sections are as follows $\phi \to v$, $x \to O$, $y \to x$.

This maximization can be done by setting the functional derivative of the (logarithm of the) posterior with respect to the potential $v(x)$ to zero

$$\delta_{v(x)} \ln p(v|D) = 0. \tag{38}$$

where $\delta_{v(x)}$ stands for $\delta/\delta v(x)$. While the functional derivative of the log–prior is simply

$$\delta_v \ln p(v) = -\lambda \mathbf{K}_0 (v - v_0), \tag{39}$$

the functional derivative of the likelihood requires the functional derivatives of energies $\delta_{v(x)} E_\alpha$ as well as the functional derivatives of the eigenfunctions $\delta_{v(x)}\phi_\alpha(x')$. These matrices can be obtained by calculating the functional derivative of the Schrödinger equation

$$\delta_{v(x)} E_\alpha = <\phi_\alpha|\, \delta_{v(x)} H \,|\phi_\alpha>= |\phi_\alpha(x)|^2, \tag{40}$$

$$\delta_{v(x)}\phi_\alpha(x') = \sum_{\gamma\neq\alpha} \frac{1}{E_\alpha - E_\gamma} \phi_\gamma(x')\phi_\gamma^*(x)\phi_\alpha(x), \tag{41}$$

using $\delta_{v(x)} H(x',x'') = \delta_{v(x)} V(x',x'') = \delta(x-x')\delta(x'-x'')$ and taking $<\phi_\alpha|\delta_{v(x)}\phi_\alpha>$ $= 0$. For the likelihood of position measurements in a canonical ensemble this yields [19,20]

$$\begin{aligned}\delta_{v(x)} p(x_i|\hat{x}, v) = {} & < \left(\delta_{v(x)}\phi^*(x_i)\right)\phi(x_i) > + < \phi^*(x_i)\delta_{v(x)}\phi(x_i) > \\ & -\beta\left(< |\phi(x_i)|^2 |\phi(x)|^2 > - < |\phi(x_i)|^2 >< |\phi(x)|^2 >\right).\end{aligned} \tag{42}$$

Similarly, measurements of particle momenta or energies can be included. Using (39) and (42) the stationarity Eq. (38) can be solved by iteration, analogously to (8),

$$v^{(r+1)} = v^{(r)} + \eta \mathbf{A}^{-1}\Big(\lambda \mathbf{K}_0(v_0 - v^{(r)}) + \sum_i \delta_{v(x)} \ln p(x_i|\hat{x}, v^{(r)})\Big). \tag{43}$$

Fig. 4 shows the reconstruction of an approximate periodic potential from position measurements in a canonical ensemble. A true potential $v_{\rm true}$, unknown to the reconstruction algorithm has been drawn from a Gaussian process with inverse covariance $\lambda\mathbf{K}_0$, where $\mathbf{K}_0 = (1-\lambda')\mathbf{K}_\theta - \lambda'\Delta$, which generates smooth and approximately periodic potentials $v(x)$. Then, sampling data at finite temperature from the quantum system with potential $v_{\rm true}$, a reconstruction of the potential has been obtained by iterating according to Eq. (43) with $\mathbf{A} = \mathbf{K}_0$.

3.1. CONCLUSIONS

Some density estimation problems have been presented where numerical solutions of nonparametric Bayesian approaches are feasible. In particular, the reconstruction of approximately periodic or approximately fractal regression functions and the reconstruction of approximately periodic potentials of quantum systems at finite temperature from position measurements has been discussed in more detail.

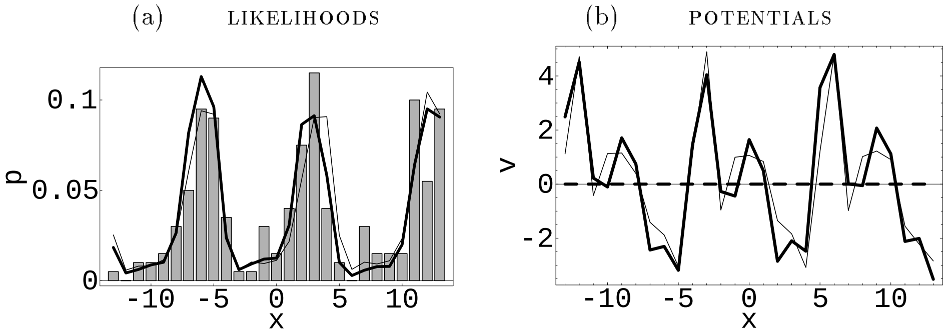

Figure 4. Reconstruction of approximately periodic potential with Gaussian prior from position measurements (without energy constraint). (a) shows empirical density (bars), true (thin line) and reconstructed likelihood (thick line). (b) shows the true potential (thin line), the zero reference potential $v_0 \equiv 0$ (dashed line), and a solution for the reconstructed potential v (thick line). (200 data points, mesh with 27 points, mass $m = 0.1$, inverse temperature $\beta = 1$, inverse prior covariance $\lambda \mathbf{K}_0 = (1 - \lambda')\mathbf{K}_\theta - \lambda'\Delta$, with $\lambda = 10$, $\lambda' = 0.001$, $\theta = 9$, and periodic boundary conditions for v as well as for the wave functions. The free additive constant of v has been chosen so that its mean coincides with the mean of the original function.)

Acknowledgments

The author wants to thank B. Giraud, J. Uhlig, and A. Weiguny for fruitful discussions.

References

1. J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag, New York, 1980.
2. D. Sivia, *Data Analysis: A Bayesian Tutorial*, Oxford University Press, Oxford, 1996.
3. B. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London, 1986.
4. I.J.Good and R. Gaskins, "Nonparametric roughness penalties for probability densities," *Biometrika*, **58**, pp. 255–277, 1971.
5. D. MacKay, "Hyperparameters: optimise or integrate out?," in *Maximum Entropy and Bayesian Methods, Santa Barbara 1993*, G. Heidbreder, ed., (Dordrecht), Kluwer, 1994.
6. J. Lemm, "Bayesian field theory," Tech. Rep. MS-TP1-99-1, Univ. of Münster, 1999.
7. C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems 8*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, eds., (Cambridge, MA), pp. 514–520, The MIT Press, 1996.
8. C.K.I.Williams and D. Barber, "Bayesian classification with gaussian processes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **20**, (12), pp. 1342–1351, 1998.
9. I. Montvay and G. Münster, *Quantum Fields on a Lattice*, Cambridge University Press, Cambridge, 1994.
10. J. Lemm and J. Uhlig, "Hartree–Fock approximation for inverse many-body problems," *Phys. Rev. Lett.*, **84**, pp. 4517–4520, 2000.
11. J. Lemm, "Mixtures of gaussian process priors," in *ICANN 99. Proceedings of the 9th International Conference on Artificial Neural Networks, Edinburgh, UK, 7–10 September 1999*, , ed., (London), IEEE Conference Publication, 1999.
12. K. Rose, E. Gurewitz, and G. Fox, "Statistical mechanics and phase transitions in clustering," *Phys. Rev. Lett.*, **65**, pp. 945–948, 1990.
13. F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural Computation*, **7**, (2), pp. 219–269, 1995.
14. H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals*, Springer–Verlag, New York, 1992.

15. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman and Company, New York, 1977.
16. R.Derka, V. Buzek, G.Adam, and P.L.Knight, "From quantum bayesian inference to quantum tomography," *Journal of Fine Mechanics and Optics*, **11–12**, p. 341, 1996.
17. M. Tan, "An inverse problem approach to optical homodyne tomography," *J. Mod. Opt.*, **44**, p. 2233, 1997.
18. V. Buzek, G. Drobný, R. Derka, G. Adam, and H. Wiedemann, "Quantum state reconstruction from incomplete data." arXiv:quant-ph/9805020, 1998.
19. J. Lemm, J. Uhlig, and A. Weiguny, "Bayesian approach to inverse quantum statistics," *Phys. Rev. Lett.*, **84**, pp. 2068–2071, 2000.
20. J. Lemm and J. Uhlig, "Bayesian inverse quantum theory," *Few Body Systems*, **29**, pp. 25–52, 2000.