Published in *Phys. Rev. Lett.* **84** (2000), 2068

A Bayesian Approach to Inverse Quantum Statistics

J. C. Lemm, J. Uhlig, and A. Weiguny

Institut für Theoretische Physik I, Universität Münster, 48149 Münster, Germany

(June 15, 2000)

A nonparametric Bayesian approach is developed to determine quantum potentials from empirical data for quantum systems at finite temperature. The approach combines the likelihood model of quantum mechanics with a priori information on potentials implemented in form of stochastic processes. Its specific advantages are the possibilities to deal with heterogeneous data and to express a priori information explicitly in terms of the potential of interest. A numerical solution in maximum a posteriori approximation is obtained for one–dimensional problems. The number of measurements being small compared to the degrees of freedom of a nonparametric estimate, the results depend strongly on the implemented a priori information.

05.30.-d, 02.50.Rj, 02.50.Wp

The last decade has seen a rapidly growing interest in learning from empirical data. Increasing computational resources enabled successful applications of empirical learning algorithms in many different areas including, for example, time series prediction, image reconstruction, speech recognition, and many more regression, classification, and density estimation problems. Empirical learning, i.e., the problem of finding underlying general laws from observations, represents a typical inverse problem and is usually ill–posed in the sense of Hadamard [1–3]. It is well known that a successful solution of such problems requires additional a priori information. In empirical learning it is a priori information which controls the generalization ability of a learning system by providing the link between available empirical "training" data and unknown outcome in future "test" situations.

The empirical learning problem we study in this Letter is the reconstruction of potentials from measuring quantum systems at finite temperature, i.e., the problem of inverse quantum statistics. Two classical research fields dealing with the determination of potentials are inverse scattering theory [4] and inverse spectral theory [5,6]. They characterize the kind of data which are necessary, in addition to a given spectrum, to identify a potential uniquely. For example, such data can be a second complete spectrum for different boundary conditions, knowledge of the potential on a half interval, or the phase shifts as a function of energy. However, neither a complete spectrum nor specific values of potentials or phase shifts for all energies can be determined empirically by a finite number of measurements. Hence, any practical algorithm for reconstructing potentials from data must rely on additional a priori assumptions, if not explicitly then implicitly. Furthermore, besides energy, other observables like particle coordinates or momenta may have been measured for a quantum system. Therefore, the approach we study in this Letter is designed to deal with arbitrary data and to treat situation specific a priori information in a flexible and explicit manner.

Many disciplines have contributed empirical learning algorithms, some of the most widely spread being decision trees, neural networks, projection pursuit techniques, various spline methods, regularization approaches, graphical models, support vector machines, and, becoming especially popular recently, nonparametric Bayesian methods [2,7–11]. Motivated by the clear and general framework it provides, the approach we will rely on is that of Bayesian statistics [12,13] which can easily be adapted to inverse quantum statistics. Computationally, however, its application to quantum systems turns out to be more demanding than, for example, typical applications to regression problems.

A Bayesian approach is based on two probability densities: 1. a likelihood model $p(x|O,v)$, quantifying the probability of outcome x when measuring observable O given a (not directly observable) potential v and 2. a prior density $p_0(v) = p(v|D_0)$ defined over a space $\mathcal{V}$ of possible potentials assuming a priori information D_0. Further, let $D_T = (x_T, O_T) = \{(x_i, O_i)|1 \le i \le n\}$ denote available training data and $D = D_T \cup D_0$ the union of training data and a priori information. To make predictions we aim at calculating the predictive density for given data D

$$p(x|O,D) = \int dv\, p(x|O,v)\, p(v|D), \tag{1}$$

assuming $p(x|O,v,D) = p(x|O,v)$, $p(v|O,D) = p(v|D)$. The posterior density $p(v|D)$ appearing in that formula is related to prior density and likelihood through Bayes' theorem $p(v|D) = p(x_T|O_T,v)\, p_0(v)/p_0(x_T|O_T)$, where the likelihood factorizes for independent data $p(x_T|O_T,v) = \prod_i p(x_i|O_i,v)$ and the denominator is v–independent. The v–integral stands for an integral over parameters if we choose a parameterized space $\mathcal{V}$, or for a functional integration if $\mathcal{V}$ is an infinite dimensional function space. Because we will in most cases not be able to solve that integral exactly we treat it in maximum a posteriori approximation, i.e., in saddle point approximation if the likelihood varies slowly at the stationary point. Then, assuming $p(x|O,D) \approx p(x|O,v^*)$ with $v^* = \text{argmax}_{v\in\mathcal{V}} p(v|D)$, integration is replaced by maximization.

According to the axioms of quantum mechanics, observables O are represented by hermitian operators and the probability of finding outcome x measuring observable O is given by

$$p(x|O,v) = \text{Tr}\Big(P_O(x)\,\rho(v)\Big). \qquad (2)$$

Here ρ denotes the density operator of the quantum mechanical system and $P_O(x) = \sum_\xi |x,\xi><x,\xi|$ is the projector on the space spanned by the orthonormalized eigenfunctions $|x,\xi>$ of O with eigenvalue x. 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, we will consider a canonical ensemble of quantum systems at temperature $1/\beta$ (setting Boltzmann's constant to 1) characterized by the density operator $\rho = \exp(-\beta H)/(\text{Tr}\exp(-\beta H))$. Furthermore, we assume a Hamiltonian $H = T + V$ being the sum of a kinetic energy term T for a particle with mass m and an unknown local potential $V(x,x') = v(x)\delta(x-x')$ which we want to reconstruct from data. In case of repeated measurements in a canonical ensemble one has to wait with the next measurement until thermal equilibrium is reached again. We will in the following focus on measurements of particle coordinates in a single particle system in a heat bath with temperature $1/\beta$. In that case the x_i represent particle positions corresponding to measurements of the observable $O_i = \hat{x}$ with $\hat{x}|x_i> = x_i|x_i>$. The likelihood becomes

$$p(x_i|\hat{x},v) = \sum_\alpha p_\alpha |\phi_\alpha(x_i)|^2 = <|\phi(x_i)|^2>, \qquad (3)$$

for $H|\phi_\alpha> = E_\alpha|\phi_\alpha>$ and with $<\cdots>$ denoting a thermal expectation value with probabilities $p_\alpha = \exp(-\beta E_\alpha)/Z$ and $Z = \sum_\alpha \exp(-\beta E_\alpha)$. One may also add a classical noise process $p(\bar{x}_i|x_i)$ on top of x_i in which case an additional integration is necessary to get the density $p(\bar{x}_i|\hat{x},v) = \int dx_i\, p(\bar{x}_i|x_i)\, p(x_i|\hat{x},v)$ of the noisy output $\bar{x}_i$.

Already at zero temperature, even complete knowledge of the true likelihood would just determine the modulus of the ground state and thus not be sufficient to determine a potential uniquely. The situation is still worse in practice, where only a finite number of probabilistic measurements is available, and at finite temperatures, as the likelihood becomes uniform in the infinite temperature limit. Hence, in addition to Eq.(2) giving the likelihood model of quantum mechanics, it is essential to include a prior density over a space of possible potentials v. To be able to formulate a priori information explicitly in terms of the function values $v(x)$ we use a stochastic process. Technically convenient is a Gaussian process prior density

$$p_0(v) = \left(\det\frac{\lambda \mathbf{K}_0}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\lambda}{2}<v-v_0|\mathbf{K}_0|v-v_0>}, \qquad (4)$$

with mean v_0, representing a reference potential or template for v, and a real symmetric, positive (semi–)definite covariance operator $(\lambda\mathbf{K}_0)^{-1}$, acting on potentials v rather than on wave functions and measuring the distance between v and v_0. Here λ is technically equivalent to a Tikhonov regularization parameter. If allowed to vary, λ is an example of a so called hyperparameter, parameterizing the prior density. Such parameters have to be included in the integration of Eq.(1) and are often also treated in maximum a posteriori approximation.

Typical choices for $\mathbf{K}_0$ implementing smoothness priors are the negative Laplacian $\mathbf{K}_0 = -\Delta$, e.g., in one dimension $-\Delta = -d^2/dx^2$, or a Radial Basis Function prior $\mathbf{K}_0 = \exp\left(-\sigma^2_{\text{RBF}}\Delta/2\right)$ [14]. Gaussian process priors can, for example, be related to approximate symmetries. Assume we expect the potential to commute approximately with a unitary symmetry operation S. Then $V \approx S^\dagger V S = \mathbf{S}V$ defines an operator $\mathbf{S}$ acting on potentials V. In that case a natural prior would be $p_0 \propto \exp(-E_S)$ with $E_S = <V - \mathbf{S}V|V - \mathbf{S}V>/2 = <V|\mathbf{K}_0|V>/2$ for $\mathbf{K}_0 = (\mathbf{I}-\mathbf{S})^\dagger(\mathbf{I}-\mathbf{S})$ and $\mathbf{I}$ denoting the identity. Note that symmetric potentials are in the null space of such a $\mathbf{K}_0$, hence another prior has to be included unless the combination with training data does determine the potential. Similarly, for a Lie group $\mathbf{S}(\theta) = \exp(\theta\mathbf{s})$ an approximate infinitesimal symmetry is implemented by $\mathbf{K}_0 = \mathbf{s}^\dagger\mathbf{s}$. In particular, a negative Laplacian smoothness prior enforces approximate symmetry under infinitesimal translation. Alternatively, a more explicit prior implementing an approximate symmetry can be obtained by choosing a symmetric reference potential $V_0 = S^\dagger V_0 S$ and $E_S = <V - V_0|V - V_0>/2$.

While a Gaussian process prior is only able to model a unimodal, concave prior density, more general prior densities can be arbitrarily well approximated by mixtures of Gaussian process priors [15]

$$p_0(v) = \sum_{k=1}^M p_0(k)\, p_0(v|k), \qquad (5)$$

with mixture probabilities $p_0(k)$ and Gaussian prior processes $p_0(v|k)$ having means v_k and covariances $(\lambda\mathbf{K}_k)^{-1}$.

To find the potential with maximal posterior we maximize for independent data, following Bayes' theorem,

$$p(v|D) \propto p_0(v) \prod_{i=1}^n p(x_i|O_i,v). \qquad (6)$$

This is done by requiring the functional derivatives of the log–posterior (technically often more convenient than the posterior) with respect to $v(x)$ to be zero,

$$\delta_{v(x)} \ln p(v|D) = 0 \tag{7}$$

where $\delta_{v(x)}$ stands for $\delta/(\delta v(x))$. To calculate the functional derivative of the likelihoods in Eq.(6) we need $\delta_{v(x)} E_\alpha$ as well as $\delta_{v(x)}\phi_\alpha(x')$. These quantities can be found by taking the functional derivative of the eigenvalue equation of the Hamiltonian yielding, for orthonormal, nondegenerate eigenfunctions,

$$\delta_{v(x)} E_\alpha = <\phi_\alpha|\,\delta_{v(x)} H\,|\phi_\alpha> = |\phi_\alpha(x)|^2, \tag{8}$$

$$\delta_{v(x)}\phi_\alpha(x') = \sum_{\gamma\neq\alpha} \frac{1}{E_\alpha - E_\gamma}\phi_\gamma(x')\phi_\gamma^*(x)\phi_\alpha(x), \tag{9}$$

using $\delta_{v(x)} H(x',x'') = \delta_{v(x)} V(x',x'') = \delta(x-x')\delta(x'-x'')$ and taking $<\phi_\alpha|\delta_{v(x)}\phi_\alpha> = 0$. Now it is straightforward to calculate the functional derivative of the likelihood

$$\begin{aligned}&\delta_{v(x)} p(x_i|\hat{x}, v) = \qquad\qquad\qquad\qquad (10)\\ &\quad < \left(\delta_{v(x)}\phi^*(x_i)\right)\phi(x_i) > + < \phi^*(x_i)\delta_{v(x)}\phi(x_i) > \\ &-\beta\left(< |\phi(x_i)|^2|\phi(x)|^2 > - < |\phi(x_i)|^2 >< |\phi(x)|^2 >\right).\end{aligned}$$

Finally, using

$$\delta_{v(x)} \ln p_0 = -\lambda \mathbf{K}_0 (v - v_0), \tag{11}$$

the functional derivative of the posterior can be calculated. Formula (11) is also valid for Gaussian mixture models of the form (5) provided we understand $\mathbf{K}_0 = \sum_k p_0(k|v)\,\mathbf{K}_k$ and $v_0 = \mathbf{K}_0^{-1}\sum_k p_0(k|v)\,\mathbf{K}_k v_k$ where $p_0(k|v) = p_0(v|k)p_0(k)/p_0(v)$.

It is straightforward to include also other kinds of data or a priori information. For example, a Gaussian smoothness prior as in Eq.(4) with zero reference potential $v_0 \equiv 0$ and $v(x) = 0$ at the boundaries tends to lead to flat potentials when the regularization parameters λ becomes large. For such cases it is useful to include besides smoothness also a priori information or data which are related more directly to the depth of the potential. One such possibility is to include information about the average energy $U = < E > = \sum_\alpha p_\alpha E_\alpha$. The average energy can then be controled by introducing a Lagrange multiplier term $\mu(U-\kappa)$, or, technically sometimes easier, by a term representing noisy energy data κ,

$$p_U \propto e^{-E_U}, \quad E_U = \frac{\mu}{2}(U-\kappa)^2, \tag{12}$$

so that $U \to \kappa$ for $\mu \to \infty$. Using $\delta_{v(x)} U = <\delta_{v(x)} E> -\beta < E\ \delta_{v(x)} E > + \beta < E >< \delta_{v(x)} E >$, we find for the functional derivative of E_U

$$\delta_{v(x)} E_U = \mu\,(U - \kappa) < |\phi(x)|^2\,(1 - \beta\,(E\ - U)) > . \tag{13}$$

Collecting all terms we are now able to solve the stationarity Eq.(7) by iteration. Starting with an initial guess $v^{(0)}$, choosing a step width η and a positive definite matrix $\mathbf{A}$ we can iterate according to

$$\begin{aligned} v^{(r+1)} &= v^{(r)} + \eta \mathbf{A}^{-1}\Big(\mathbf{K}_0^{(r)}(v_0 - v^{(r)}) \qquad (14)\\ &\quad + \sum_i \delta_{v(x)} \ln p(x_i|\hat{x}, v^{(r)}) - \delta_{v(x)} E_U^{(r)}\Big). \end{aligned}$$

Here we included an E_U term which depends on v, like $\mathbf{K}_0$ for mixture models, and thus changes during iteration. Typically, η and often also $\mathbf{A}$ are adapted during iteration. In the numerical examples we have studied, $\mathbf{A} = \lambda \mathbf{K}_0$ proved to be a good choice. Note that in general different initial guesses $v^{(0)}$ can yield different solutions.

The numerical difficulties of the nonparametric Bayesian approach arise from the fact that the quantum mechanical likelihood (2) is non–Gaussian and non–local in the potential $v(x)$. Similar to general density estimation problems, even for Gaussian priors none of the $v(x)$–integrations in (1) can be carried out analytically [16]. In contrast, for example, Gaussian regression problems have a likelihood being Gaussian and local in the function of interest, and an analogous nonparametric Bayesian approach with a Gaussian process prior requires only to deal with matrices with dimension not larger than the number of training data [11]. The following examples will show, however, that a direct numerical solution of Eq.(7) by discretization is feasible for one–dimensional problems. Higher dimensional problems, on the other hand, require further approximations. For example, work on inverse many–body problems on the basis of a Hartree–Fock approximation is in progress.

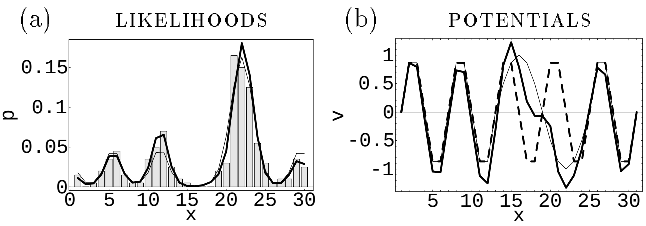

FIG. 1. Reconstruction of an approximately periodic potential from coordinate measurements. Shown are (a) empirical density, i.e., relative frequencies of sampled data (bars), true (thin line) and reconstructed likelihood (thick line) (b) $v_{\rm true}$ (thin line), reference potential $v_0 = \sin(\pi x/3)$ (dashed), and reconstructed potential v (thick line). (200 data points, mesh with 30 points, $m = 0.25$ for $\hbar = 1$, $\beta = 4$, $\mathbf{K}_0 = -\Delta$, $\lambda = 0.2$, $\mu = 0$, $U(v_{\rm true}) = -0.354$, $U(v) = -0.552$.)

In the following numerical examples we discuss the reconstruction of an approximately periodic, one–dimensional potential v, with the reference potential v_0 chosen periodic. The potential v may describe a one–dimensional surface, deviating from exact periodicity due to localized defects. To enforce the deviation from v_0 to be smooth we take as prior on v a negative Laplacian covariance, i.e., $\mathbf{K}_0 = -\Delta$. Fig.1 shows representative numerical results for a grid with 30 points and 200 data sampled from the likelihood $p(x|\hat{x}, v_{\rm true})$ for some chosen potential $v_{\rm true}$. The reconstructed potential v has been found by iterating without energy penalty term E_U according to Eq.(14) with $\mathbf{A} = \mathbf{K}_0$. We took zero boundary conditions for v, so $\mathbf{K}_0$ becomes invertible, and, consistently, periodic boundary conditions for the eigenfunctions ϕ_α. Note that the data have been sufficient to identify clearly the deviation from the periodic reference potential. Fig.2a shows the same example with an energy penalty term E_U with $\kappa = U(v_{\rm true})$. While the reconstructed likelihood (not shown) is not much altered, the true potential is now better approximated in regions where it is small. As a rule, similar likelihoods do not necessarily imply similar potentials, and vice versa.

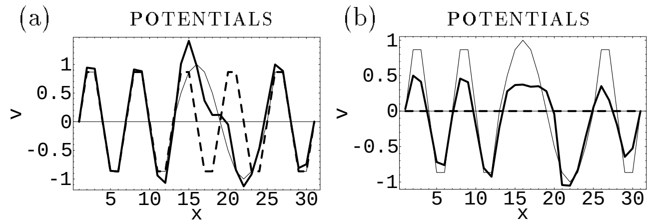

FIG. 2. Reconstruction with energy penalty term E_U with $\kappa = U(v_{\rm true}) = -0.354$. Shown are true (thin lines), reconstructed (thick lines), and reference potential (dashed). (a) $\mathbf{K}_0 = -\Delta$, $v_0(x) = \sin(\pi x/3)$, $\mu = 1000$, $U(v) = -0.353$. (b) $\mathbf{K}_0 = -\Delta - \gamma\Delta_\theta$, $v_0 \equiv 0$, $\theta = 6$, $\gamma = 0.12$, $\lambda = 0.05$, $\mu = 10$, $U(v) = -0.351$. All other parameters as for Fig.1.

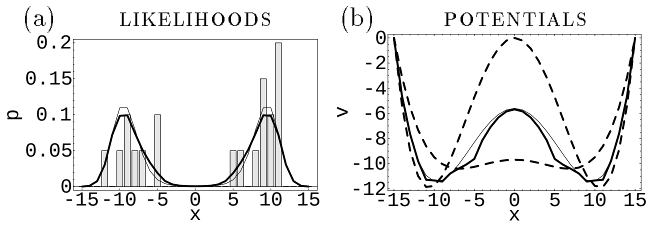

FIG. 3. Reconstruction of symmetric potential with mixture of Gaussian priors. (a) shows empirical density (bars), true (thin line) and reconstructed likelihood (thick line). (b) shows the true potential (thin line), the two reference potentials v_1 and v_2 (dashed lines), and a solution for the reconstructed potential v (thick line). Because both reference potentials are partially supported by the data the approximated v is a mixture of the solutions for v_1 and v_2. (Final mixture coefficients $p_0(1|v) = 0.57$, $p_0(2|v) = 0.43$, 20 data points, mesh with 31 points, $v(x) = v(-x)$, boundary condition $v(\pm 15) = 0$, and $m = 0.1$, $\beta = 1$, $\mathbf{K}_0 = -\Delta$, $\lambda = 0.5$, $\mu = 10$, $\kappa = -10, 01 = U(v_{\rm true})$ resulting in $U(v) = -10.1$.)

Fig.2b shows the implementation of approximate periodicity by an operator Δ_θ, defined by $< v|\Delta_\theta|v > = \int dx |v(x) - v(x+\theta)|^2$ for periodic boundary conditions on v, thus measuring the difference between the potential $v(x)$ and the potential translated by θ. To find smooth solutions we added a negative Laplacian term with zero reference potential, i.e., we used $\mathbf{K}_0 = -\Delta - \gamma\Delta_\theta$. To have an invertible matrix for periodic boundary conditions on v we iterated this time with $\mathbf{A} = \mathbf{K}_0 + 0.1\mathbf{I}$. The implementation of approximate periodicity by Δ_θ instead of a periodic v_0 is more general in so far as it annihilates arbitrary functions with period θ. As, however, the reference function of the Laplacian term does not fit the true potential very well, the reconstruction is poorer in regions where the potential is large and thus no data are available. In these regions a priori information is of special importance. Finally, Fig.3 shows the implementation of a mixture of Gaussian prior processes.

In conclusion, we have applied a nonparametric Bayesian approach to inverse quantum statistics and shown its numerical feasibility for one–dimensional examples.

---

[1] A. Tikhonov and V. Arsenin, *Solution of Ill–posed Problems* (Wiley, New York, 1977).

[2] V. Vapnik, *Statistical Learning Theory* (Wiley, New York, 1998).

[3] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer Verlag, New York, 1996).

[4] R. Newton, *Inverse Schrödinger Scattering in Three Dimensions* (Springer Verlag, Berlin, 1989).

[5] K. Chadan, D. Colton, L. Päivärinta, and W. Rundell, *An Introduction to Inverse Scattering and Inverse Spectral Problems* (SIAM, Philadelphia, 1997).

[6] B. Zakhariev and V. Chabanov, Inverse Problems **13**, R47 (1997).

[7] G. Wahba, *Spline Models for Observational Data* (SIAM, Philadelphia, 1990).

[8] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation* (Addison–Wesley, Redwood City, CA, 1991).

[9] *Machine Learning, Neural and Statistical Classification*, edited by D. Michie, D. Spiegelhalter, and C. Taylor (Ellis Horwood, New York, 1994).

[10] S. Lauritzen, *Graphical Models* (Oxford University Press, Oxford, 1996).

[11] C. K. I. Williams and C. E. Rasmussen, in *NIPS 8*, edited by D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (The MIT Press, Cambridge, MA, 1996), pp. 514–520.

[12] J. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer Verlag, New York, 1980).

[13] D. Sivia, *Data Analysis: A Bayesian Tutorial* (Oxford University Press, Oxford, 1996).

[14] F. Girosi, M. Jones, and T. Poggio, Neural Computation **7**, 219 (1995).

[15] J. Lemm, in *Proceedings of ICANN 99* (IEEE Conference Publication, London, 1999).

[16] J. Lemm, Technical Report No. MS-TP1-99-1, Univ. of Münster, physics/9912005.