Bayesian Inverse Quantum Theory*

Jörg C. Lemm**, Joerg Uhlig

Institut für Theoretische Physik I, Wilhelm–Klemm–Str.9, D–48149 Münster, Germany

Abstract. A Bayesian approach is developed to determine quantum mechanical potentials from empirical data. Bayesian methods, combining empirical measurements and *a priori* information, provide flexible tools for such empirical learning problems. The paper presents the basic theory, concentrating in particular on measurements of particle coordinates in quantum mechanical systems at finite temperature. The computational feasibility of the approach is demonstrated by numerical case studies. Finally, it is shown how the approach can be generalized to such many–body and few–body systems for which a mean field description is appropriate. This is done by means of a Bayesian inverse Hartree–Fock approximation.

Contents

---

*Dedicated to Professor Achim Weiguny on the occasion of his retirement

**lemm@uni-muenster.de

1 Introduction

The problem addressed in this paper is the reconstruction of Hamiltonians of quantum systems from observational data. Finding such "causes" or "laws" from a finite number of observations constitutes an *inverse problem* and is typically *ill–posed* in the sense of Hadamard [1–8].

Two research fields deal in particular with the reconstruction of potentials from spectral data (energy measurements): *inverse spectral theory* and *inverse scattering theory*. Inverse spectral theory characterizes the kind of data necessary, in addition to a given spectrum, to determine the potential [7, 9–13]. (See also Sect. 3.2.1.) Inverse scattering theory, in particular, considers, in addition to the spectrum, boundary data obtained 'far away' from the scatterer. Those can be, for example, phase shifts obtained from scattering experiments [12, 14, 15].

In this paper, contrasting those two approaches, we will not exclusively be interested in spectral data, but will develop a formalism which allows to extract information from quite heterogeneous empirical data. In particular, we will consider in more detail the situation where the position of a quantum mechanical particle has been measured a finite number of times.

Due to increasing computational resources, the last decade has also seen a rapidly growing interest in applied empirical learning problems. They appear as density estimation, regression or classification problems and include, just to name a few, image reconstruction, speech recognition, time series prediction, and object recognition. Many disciplines, like applied statistics, artificial intelligence, computational and statistical learning theory, statistical physics, and also psychology and biology, contributed in developing a variety of learning algorithms, including for example smoothing splines [16], regularization and kernel approaches [4], support vector machines [17, 18], generalized additive models [19], projection pursuit regression [20], expert systems and decision trees [21], neural networks [22], and graphical models [23].

Recently, their has been much work devoted to the comparison and unification of methods arising from different disciplines. (For an overview and comparison of methods see for example [24].) Hereby, especially the *Bayesian approach* to statistics proved to be useful as a unifying framework for empirical learning [22, 25–36].Bayesian approaches put special emphasis on *a priori* information which always has to accompany empirical data to allow successful learning.

The present paper is written from a Bayesian perspective. In particular, *a priori* information will be implemented in form of stochastic processes [37]. Compared to parametric techniques this has the advantage, that *a priori* information can typically be controlled more explicitly. Technically, this approach is intimately related to the well known Tikhonov regularization [2, 3]. For an outline of the basic principles see also [38].

The paper is organized as follows: Sect. 2 gives a short introduction to Bayesian statistics. Sect. 3 applies the Bayesian approach to quantum mechanics and quantum statistics, with Sect. 3.1 concentrating on the treatment of empirical data for quantum systems and Sect. 3.2 discussing the implementation of *a priori* information. Sect. 3.3 presents two numerical case studies, the first dealing with the approximation of approximately periodic potentials, the second with inverse two–body problems. Sect. 4 shows how the approach can be applied to many–body systems, including the fundamentals of an inverse version of Hartree–Fock theory. Finally, Sect. 5 concludes the paper.

2 The Bayesian approach

2.1 Basic notations

A Bayesian approach is based upon two main ingredients:

1. A model of Nature, i.e., a space $\mathcal{H}$ of hypotheses h defined by their *likelihood* functions $p(x|c,h)$. Likelihood functions specify the probability density for producing outcome x (measured value or dependent visible variable, assumed to be directly observable) under hypothesis h (possible state of Nature or hidden variable, assumed to be not directly observable) and condition c (measurement device parameters or independent visible variable, assumed to be directly observable).

2. A *prior density* $p_0(h) = p(h|D_0)$ defined over the space $\mathcal{H}$ of hypotheses, D_0 denoting collectively all available *a priori* information.

Now assume (new) *training data* $D_T = (x_T, c_T) = \{(x_i, c_i)|1 \le i \le n\}$ become available, consisting of pairs of measured values x_i under known conditions c_i (and unknown h). Then *Bayes' theorem*

$$p(h|D) = \frac{p(x_T|c_T,h)\, p_0(h)}{p_0(x_T|c_T)}, \tag{1}$$

is used to update the prior density $p_0(h) = p(h|D_0)$ to get the (new) *posterior density* $p(h|D) = p(h|D_T, D_0)$. Here we wrote $D = (D_T, D_0)$ to denote both, training data and *a priori* information. Assuming i.i.d. training data D_T the likelihoods factorize $p(x_T|c_T,h) = \prod_i^n p(x_i|c_i,h)$. Note that the denominator which appears in Eq. (1), i.e., $p_0(x_T|c_T) = \int dh\, p(x_T|c_T,h)\, p_0(h)$, is h–independent. It plays the role of a normalization factor, also known as evidence. Thus, the terms in Eq. (1) are named as follows,

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}. \tag{2}$$

To make predictions, a Bayesian approach aims at calculating the *predictive density*

$$p(x|c, D) = \int dh\, p(x|c, h)\, p(h|D), \tag{3}$$

which is a likelihood average weighted by their posterior probability. The h–integral can be extremely high dimensional, and often, like in the case we are focusing on here, even be a functional integral [39, 40] over a space of likelihood functions $\mathcal{H}$. In as far as an analytical integration is not possible, one has to treat the integral, for example, by Monte Carlo methods [30, 41–44] or in saddle point approximation [26, 30, 45, 46]. Assuming the likelihood term $p(x|c, h)$ to be slowly varying at the stationary point the latter is also known as *maximum posterior approximation*. In this approximation the h–integration is effectively replaced by a maximization of the posterior, meaning the predictive density is approximated by

$$p(x|c, D) \approx p(x|c, h^*), \tag{4}$$

where

$$h^* = \text{argmax}_{h\in\mathcal{H}} p(h|D) = \text{argmin}_{h\in\mathcal{H}}[-\ln p(h|D)]. \tag{5}$$

The term $-\ln p(h|D)$ is also often referred to as (regularized) *error functional* and indeed, a maximum posterior approximation is technically equivalent to minimizing error functionals with Tikhonov regularization [2–4, 47]. The difference between the Bayesian approach and the classical Tikhonov regularization is the interpretation of the extra term as costs or as *a priori* information, respectively.

Within a *maximum likelihood approach* an optimal hypothesis h is obtained by maximizing only its training likelihood $p(x_T|c_T, h)$ instead of its complete posterior. This is equivalent to a maximum posterior approximation with uniform prior density. A maximum likelihood approach can be used for hypotheses $h = h(\xi)$, parameterized by (vectors) ξ. A maximum likelihood approach is possible if that parameterization is restrictive enough (and well enough adapted to the problem), so no additional prior is required to allow generalization from training to non–training data. For completely flexible nonparametric approaches, however, the prior term is necessary to provide the necessary information linking training data and (future) non–training data. Indeed, if every number $p(x|c, h)$ is considered as a single degree of freedom [restricted only by the positivity constraint $p(x|c, h) > 0$ and the normalization over x] then, without *a priori* information, training data contain no information about non–training data.

2.2 An Example: Regression

Before applying the Bayesian framework to quantum theory, we shortly present one of its standard applications: the case of (Gaussian) regression. (For more details see for example [34].) This also provides an example for the relation between the Bayesian maximum posterior approximation and the minimization of regularized error functionals.

A regression model is a model with Gaussian likelihoods,

$$p(x|c,h) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-h(c))^2}{2\sigma^2}}, \tag{6}$$

with fixed variance σ^2. The function $h(c)$ is known as regression function. (In regression, one often writes y for the dependent variable, which is x in our notation, and x for the "condition" c.) Our aim is to determine an approximation for $h(c)$ using observational data $D = \{(x_i, c_i)|1 \le i \le n\}$. Within a parametric approach one searches for an optimal approximation in a space of parameterized regression functions $h(c;\xi)$. For example, in the simple cases of a constant or a linear regression such a parameterization would be $h(c;\xi) = \xi$ or $h(c;\xi_1,\xi_2) = \xi_2 c + \xi_1$, respectively. If the parameterization is restrictive enough then a prior term $p_0(h)$ is not needed and maximizing the likelihood over all data D is thus equivalent to minimizing the squared error,

$$E_{\rm sq}(\xi) = \sum_i^n \Big(x_i - h(c_i;\xi) \Big)^2. \tag{7}$$

There are, however, also very flexible parametric approaches, which usually do require additional *a priori* information. An example of such a nonlinear one–parameter family has been given by Vapnik and is shown in Fig. 1. Without additional *a priori* information, which may for example restrict the number of oscillations, such functions can in most cases not be expected to lead to useful predictions. Nonparametric approaches, which treat the numbers $h(c)$ as single degrees of freedom, are even more flexible and do always require a prior $p_0(h)$. For nonparametric approaches such a prior can be formulated in terms of the function values $h(c)$. A technically very convenient choice is a Gaussian process prior [48, 49],

$$p_0(h) = \left(\det \frac{\lambda}{2\pi} \mathbf{K}_0 \right)^{\frac{1}{2}} e^{-\frac{\lambda}{2} < h-h_0 \,|\, \mathbf{K}_0 \,|\, h-h_0 >}, \tag{8}$$

with mean h_0, representing a reference or template for the regression function h, and inverse covariance $\lambda\mathbf{K}_0$ given by a real symmetric, positive (semi–)definite operator scaled by λ and acting on functions h. The operator $\mathbf{K}_0$ defines the scalar product,

$$< h - h_0 \,|\, \mathbf{K}_0 \,|\, h - h_0 > = \int dc\, dc' \, [h(c) - h_0(c)] \, \mathbf{K}_0(c,c') \, [h(c') - h_0(c')]. \tag{9}$$

Typical priors enforce the regression function h to be smooth. Such smoothness priors are implemented by choosing differential operators for $\mathbf{K}_0$. For example, taking for $\mathbf{K}_0$ the negative Laplacian and choosing a zero mean $h_0 = 0$, yields

$$< h - h_0 \,|\, \mathbf{K}_0 \,|\, h - h_0 > = -\int dc\, h(c) \frac{\partial^2}{\partial c^2} h(c) = \int dc \left(\frac{\partial h(c)}{\partial c} \right)^2, \tag{10}$$

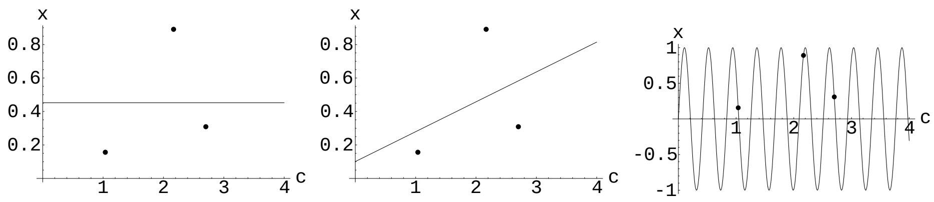

Figure 1. Examples of parametric regression functions with increasing flexibility (with 3 data points). L.h.s: A fitted constant $h(c) = \xi$. Middle: A linear $h(c) = \xi_2 c + \xi_1$. R.h.s: The function $h(c) = \sin(\xi c)$ can fit an arbitrary number of data points (with different $c_i \neq 0$ and $|x_i| \leq 1$) well [17]. Additional *a priori* information becomes especially important for flexible approaches.

where we integrated by parts assuming vanishing boundary terms. In statistics one often prefers inverse covariance operators with higher derivatives to obtain smoother regression functions [16, 50–55]. An example of such a prior with higher derivatives is a "Radial Basis Functions" prior with the pseudo-differential operator $\mathbf{K}_0 = \exp\left(-\sigma^2_{\rm RBF}\Delta/2\right)$ as inverse covariance.

Maximizing the posterior $p(x|c, D)$ for a Gaussian prior (8) is equivalent to minimizing the regularized error functional

$$E_{\rm reg}(h) = \frac{1}{2}\sum_i^n \Big(x_i - h(c_i)\Big)^2 + \frac{\lambda'}{2} < h - h_0 \,|\, \mathbf{K}_0 \,|\, h - h_0 > . \tag{11}$$

The "regularization" parameter $\lambda' = \lambda\sigma^2$, representing a so called *hyperparameter*, controls the balance between empirical data and *a priori* information. In a Bayesian framework one can include a hyperprior $p(\lambda)$ and either integrate over λ or determine an optimal λ in maximum posterior approximation [22, 26]. Alternative ways to determine λ are crossvalidation techniques[16], the discrepancy and the self–consistent method [56]. For example in the case of a smoothness prior, a larger λ' will result in a smoother regression function h^*. It is typical for the case of regression that the regularized error functional $E_{\rm reg}(h)$ is quadratic in h. It is therefore easily minimized by setting the functional derivative with respect to h to zero, i.e., $\delta E_{\rm reg}/\delta h = \delta_h E_{\rm reg} = 0$. This stationarity equation is then linear in h and thus has a unique solution h^*. (This is equivalent to so called kernel methods with kernel $\mathbf{K}_0^{-1}$. It is specific for regression with Gaussian prior that, given $\mathbf{K}_0^{-1}$, only an n–dimensional equation has to be solved to obtain h^*.) As the resulting maximum posterior solution $p(x|c, h^*)$ is Gaussian by definition, we find for its mean

$$\int dx\, x\, p(x|c, h^*) = h^*(c). \tag{12}$$

It is not difficult to check that, for regression with a Gaussian prior, $h^*(c)$ is also equal to the mean $\int dx\, x\, p(x|c, D)$ of the exact predictive density (3). Furthermore it can be shown that, in order to minimize the squared error $[x - h(c)]^2$ for (future) test data, it is optimal to predict outcome $x = h(c)$ for situation c.

In the following sections we will apply the Bayesian formalism to quantum theory. Hence, training data x_i will represent the results of measurements on quantum systems and conditions c_i will describe the kind of measurements performed. Being interested in the determination of quantum potentials our hypotheses h will in the following represent potentials v.

3 Inverse quantum statistics

3.1 The likelihood model of quantum theory

3.1.1 Measurements in quantum theory

The state of a quantum mechanical system is characterized by its density operator ρ. In particular, the probability of measuring value x for observable O in a state described by ρ is known to be [57, 58]

$$p(x|O,\rho) = \mathrm{Tr}\Big(P_O(x)\,\rho\Big). \tag{13}$$

This defines the likelihood model of quantum theory. The observable O, represented by a hermitian operator, corresponds to the condition c of the previous section. The projector $P_O(x) = \sum_l |\, x_l \!><\! x_l \,|$ projects on the space of eigenfunctions $|\, x_l\!>$ of O with eigenvalue x, i.e., for which $O|\, x_l\!> = x|\, x_l\!>$. For non–degenerate eigenvalues $P_O(x) = |\, x \!><\! x \,|$.

To be specific, we will consider the measurement of particle positions, i.e., the case $O = \hat{x}$ with $\hat{x}$ being the multiplication operator in coordinate space. However, the formalism we will develop does not depend on the particular kind of measured observable. It would be possible, for example, to mix measurements of position and momentum (see, for example, Section 3.2.5).

For the sake of simplicity, we will assume that no *classical noise* is added by the measurement process. It is straightforward, however, to include a classical noise factor in the likelihood function. If, for example, the classical noise is, conditioned on x_i, independent of quantum system then

$$p(\bar{x}_i|O,\rho) = \int\! dx_i\, p(\bar{x}_i|x_i)\, p(x_i|O,\rho), \tag{14}$$

where we denoted the 'true' coordinates by x_i and the corresponding noisy output by $\bar{x}_i$. A simple model for $p(\bar{x}_i|x_i)$ could be a Gaussian.

In contrast to the (ideal) measurement of a classical system, the measurement of a quantum system changes the state of the system. In particular, the measurement process acts as projection $P_O(x)$ to the space of eigenfunctions of operator O with eigenvalues consistent with the measurement result. Thus, performing multiple measurements under the assumption of a constant density operator ρ requires special care to ensure the correct preparation of the quantum system before each measurement. In particular, considering a quantum statistical system at finite temperature, as we will do in the the next section, the time between two consecutive measurements should be large enough to allow thermalization of the system.

3.1.2 Likelihood in the canonical ensemble

From now on we will consider a quantum mechanical canonical ensemble at temperature $1/\beta$. Such a system is described by a density operator

$$\rho = \frac{e^{-\beta H}}{\mathrm{Tr}\, e^{-\beta H}}, \tag{15}$$

H denoting the Hamiltonian of the system. Specifically, we will focus on repeated measurements of a single particle in a heat bath of temperature $1/\beta$ with sufficient time between measurements to allow the heat bath to reestablish the canonical density operator. For (non–degenerated) particle coordinates x_i the likelihood for ρ becomes the thermal expectation

$$p(x_i|\hat{x},\rho) = \mathrm{Tr}\,(P_{\hat{x}}(x_i)\rho) = \sum_\alpha p_\alpha |\phi_\alpha(x_i)|^2 = < |\phi(x_i)|^2 >, \tag{16}$$

$< \cdots >$ denoting the thermal expectation with probabilities

$$p_\alpha = \frac{e^{-\beta E_\alpha}}{Z}, \quad Z = \sum_\alpha e^{-\beta E_\alpha}, \tag{17}$$

and energies and orthonormalized eigenstates

$$H|\,\phi_\alpha > = E_\alpha |\,\phi_\alpha >. \tag{18}$$

In particular, we will consider a hermitian Hamiltonian of the standard form $H = T + V$, with kinetic energy T, being $1/(2m)$ times the negative Laplacian $-\Delta$ for a particle with mass m (setting $\hbar = 1$), and local potential $V(x,x') = v(x)\delta(x-x')$. Thus, in one dimension

$$H(x,x') = \left(-\frac{1}{2m}\frac{\partial^2}{\partial x^2} + v(x) \right) \delta(x-x'), \tag{19}$$

where the δ–functional is usually skipped.

For n independent position measurements x_i the likelihood for $\rho(v)$, and thus for v, becomes (writing now $p(x_i|\hat{x},\rho) = p(x_i|\hat{x},v)$)

$$p(x_T|\hat{x},v) = \prod_{i=1}^n p(x_i|\hat{x},v) = \prod_{i=1}^n < |\phi(x_i)|^2 > . \tag{20}$$

We remark that it is straightforward to allow β to vary between measurements.

3.1.3 Maximum likelihood approximation

A maximum likelihood approach selects the potential v with maximal likelihood $p(x_T|\hat{x},v)$ under the training data. Beginning with a discussion of the *parametric* approach we consider a potential $v(\xi,x)$ parameterized by a parameter vector ξ with components ξ_l. To find the parameter vector which maximizes the training likelihood we have to solve the stationarity equation

$$0 = \partial_\xi p(x_T|\hat{x},v), \tag{21}$$

$\partial_\xi = \partial/\partial\xi$ denoting the gradient operator with components ∂_{ξ_l}. Obtaining from Eq. (20)

$$\begin{aligned} \partial_\xi p(x_i|\hat{x},v) &= <(\partial_\xi \phi^*(x_i))\phi(x_i)> + <\phi^*(x_i)(\partial_\xi \phi(x_i))> \\ &- \beta\left(<|\phi(x_i)|^2 \partial_\xi E> - <|\phi(x_i)|^2><\partial_\xi E>\right), \end{aligned} \tag{22}$$

we see that to solve Eq. (21) we have to calculate the derivatives of the eigenvalues $\partial_\xi E_\alpha$ and of the eigenfunctions at the data points $\partial_\xi \phi_\alpha(x_i)$. Those are implicitly defined by the eigenvalue equation for $H = H(v)$. To proceed we take the derivative of the eigenvalue equation (18)

$$(\partial_\xi H)\,|\,\phi_\alpha> + H|\,\partial_\xi\phi_\alpha> = (\partial_\xi E_\alpha)\,|\,\phi_\alpha> + E_\alpha|\,\partial_\xi\phi_\alpha>. \tag{23}$$

Projecting onto $<\phi_\alpha\,|$ yields, using $\partial_\xi H = \partial_\xi v$ and the hermitian conjugate of Eq. (18) we arrive at

$$\partial_\xi E_\alpha = \frac{<\phi_\alpha\,|\,\partial_\xi v\,|\,\phi_\alpha>}{<\phi_\alpha\,|\,\phi_\alpha>}, \tag{24}$$

$$(E_\alpha - H)|\,\partial_\xi\phi_\alpha> = (\partial_\xi v - \partial_\xi E_\alpha)\,|\,\phi_\alpha>. \tag{25}$$

Because all orbitals with energy E_α (which may be more than one if E_α is degenerated) are in the null space of the operator $(E_\alpha - H)$, Eq. (25) alone does not determine $\partial_\xi\phi_\alpha$ uniquely. We also notice, that because the left hand side of Eq. (25) vanishes if projected on a eigenfunction ϕ_γ with $E_\gamma = E_\alpha$ we find for degenerate eigenfunctions $<\phi_\gamma\,|\,\partial_\xi H\,|\,\phi_\alpha> = 0$, if we choose $<\phi_\alpha\,|\,\phi_\gamma> = \delta_{\alpha,\gamma}$. A unique solution for $\partial_\xi\phi_\alpha$ can be obtained be setting $<\phi_\gamma\,|\,\partial_\xi\phi_\alpha> = 0$ for eigenfunctions ϕ_γ with $E_\alpha = E_\gamma$. This corresponds to fixing normalization and phase of eigenfunctions and, in case of degenerate eigenvalues, uses the freedom to work with arbitrary, orthonormal linear combinations of the corresponding eigenfunctions. Because the operator $(E_\alpha - H)$ is invertible in the space spanned by all eigenfunctions ϕ_γ with different energy $E_\gamma \neq E_\alpha$, this yields, using orthonormal eigenfunctions,

$$|\,\partial_\xi\phi_\alpha> = \sum_{\substack{\gamma \\ E_\gamma \neq E_\alpha}} \frac{1}{E_\alpha - E_\gamma}|\,\phi_\gamma><\phi_\gamma\,|\,\partial_\xi v\,|\,\phi_\alpha>. \tag{26}$$

For nondegenerated energies the sum becomes $\sum_{\gamma\neq\alpha}$. The stationarity equation (21) can now be solved iteratively by starting from an initial guess v^0 for v, calculating $E_\alpha(v)$ and $\phi_\alpha(v)$ to obtain $\partial_\xi E_\alpha$ and $\partial_\xi\phi_\alpha(x_i)$ from Eqs. (24,25) and thus $\partial_\xi p(x_i|\hat{x},v)$ from Eq. (22). Then a new guess for v is calculated (switching to log–likelihoods)

$$v^{\text{new}} = v^{\text{old}} + \eta A^{-1}\sum_i \partial_\xi \ln p(x_i|\hat{x},v^{\text{old}}), \tag{27}$$

with some step width η and some positive definite operator A (approximating for example the Hessian of $\ln p(x_T|\hat{x},v)$). This procedure is now iterated till convergence.

While a parametric approach restricts the space of possible potentials v, a *nonparametric* approach treats each function value $v(x)$ itself as individual degree of freedom, not restricting the space of potentials. The corresponding nonparametric stationarity equation is obtained analogous to the parametric stationarity equation (21) replacing partial derivatives ∂_ξ with the functional derivative operator $\delta_v = \delta/\delta v$ with components $\delta_{v(x)} = \delta/\left(\delta v(x)\right)$ [59]. Because the functional derivative of H is simply

$$\delta_{v(x)} H(x', x'') = \delta_{v(x)} V(x', x'') = \delta(x - x')\delta(x' - x''), \tag{28}$$

we get, using the same arguments leading to Eq. (26)

$$\delta_{v(x)} E_\alpha = \frac{<\phi_\alpha \mid \delta_{v(x)} H \mid \phi_\alpha>}{<\phi_\alpha \mid \phi_\alpha>} = |\phi_\alpha(x)|^2, \tag{29}$$

$$\delta_{v(x)} \phi_\alpha(x') = \sum_{\substack{\gamma \\ E_\gamma \neq E_\alpha}} \frac{1}{E_\alpha - E_\gamma} \phi_\gamma(x')\phi^*_\gamma(x)\phi_\alpha(x), \tag{30}$$

and therefore

$$\begin{aligned} \delta_{v(x)} p(x_i|\hat{x}, v) &= < \left(\delta_{v(x)} \phi^*(x_i)\right) \phi(x_i) > + < \phi^*(x_i)\delta_{v(x)}\phi(x_i) > \\ &- \beta\left(< |\phi(x_i)|^2 |\phi(x)|^2 > - < |\phi(x_i)|^2 >< |\phi(x)|^2 >\right). \end{aligned} \tag{31}$$

(The partial derivative with respect to parameters ξ and the functional derivative with respect to $v(x)$ are related according to the chain rule $\partial_{\xi_l} p(x_i|\hat{x}, v) = \int dx \left(\partial_{\xi_l} v(x)\right) \delta_{v(x)} p(x_i|\hat{x}, v) = \mathbf{v}_\xi \delta_v p(x_i|\hat{x}, v)$ with operator $\mathbf{v}_\xi(l, x) = \partial_{\xi_l} v(\xi, x)$.)

The large flexibility of the nonparametric approach allows an optimal adaption of v to the available training data. However, as it is well known in the context of learning it is the same flexibility which makes a satisfactory generalization to non–training data (e.g., in the future) impossible, leading, for example, to 'pathological', δ-functional like solutions. Nonparametric approaches require therefore additional restrictions in form of *a priori* information. In the next section we will include *a priori* information in form of stochastic processes, similarly to Bayesian statistics with Gaussian processes [16, 37, 44, 48, 60–63] or to classical regularization theory [2, 4, 16]. In particular, *a priori* information will be implemented *explicitly*, by which we mean it will be expressed directly in terms of the function values $v(x)$ itself. This is a great advantage over parametric methods where *a priori* information is *implicit* in the chosen parameterization, thus typically difficult or impossible to analyze and not easily adapted to the situation under study. Indeed, because it is only *a priori* knowledge which relates training to non–training data, its explicit control is essential for any successful learning.

3.2 Prior models for potentials

3.2.1 The need for *a priori* information

Typical results of inverse spectral theory show that, for example, a one–dimensional local potential can be reconstructed if a set of two complete spectra

$\{E_\alpha^{(1)}\}_1^\infty$, $\{E_\alpha^{(2)}\}_1^\infty$ is given for two different boundary conditions for ϕ_α [7, 12]. Alternatively, a single spectrum is sufficient, if either a complete set of norming constants $u_\alpha = \int \phi_\alpha dx$ is given (for a certain normalization of ϕ_α which fixes the values of ϕ_α on the boundary) [9] or the potential is already known on half of the interval [64]. Results from inverse scattering theory show under which circumstances a potential can be reconstructed from, e.g., a complete spectrum and the phase shifts as function of energy [12, 14, 15]. In practice, however, the number of actual measurements can only be finite. Thus, even if noiseless measurement devices would be available, an empirical determination of a complete spectrum, or of phase shifts as function of energy, is impossible. Therefore, to reconstruct a potential from experimental data in practice, inverse spectral or inverse scattering theory has to be combined with additional *a priori* information. If such *a priori* information is not made explicit — as we try to do in the following — it nevertheless enters any algorithm at least implicitly.

We address in this paper the measurement of arbitrary quantum mechanical observables, not restricted to spectral or scattering data. In particular, we have considered the measurement of particle positions. However, measuring particle positions only can usually not determine a quantum mechanical potential completely. For example, consider the ideal case of an infinite data limit $n \to \infty$ for a discrete x variable (so derivatives with respect to x have to be understood as differences) at zero temperature (i.e., $\beta \to \infty$). This, at least, would allow to obtain $p(x|\hat{x}, v) = |\phi_0(x)|^2$ to any desired precision. But even when we restrict to the case of a local potential, we would also need, for example, the ground state energy E_0 and $\phi_0^*(x)\phi_0''(x)$ to determine $v(x)$ from the eigenvalue equation of H

$$v(x) = E_0 + \frac{1}{2m} \frac{\phi_0^*(x)\phi_0''(x)}{|\phi_0(x)|^2}, \tag{32}$$

where $\phi_0'' = \partial^2 \phi_\alpha(x)/\partial x^2$ (or a discretized version thereof). For finite data, a nonlocal potential, continuous x, or finite temperature the situation is obviously even worse. In the high temperature limit, for example, $p(x|\hat{x}, v)$ becomes uniform and independent from the potential. Summarizing, even in the ideal case where the complete true likelihood $p(x|\hat{x}, v)$ is assumed to be known, the problem of reconstructing potential can still be ill–posed. (The corresponding time–dependent problem, i.e., the reconstruction of a potential v given the complete time–dependent likelihood, is treated in [65]. A Bayesian approach for time–dependent systems, based on finite data, can be found in [66].) Hence, while *a priori* information is crucial for every learning problem [32, 33, 67], the reconstruction of a quantum mechanical potential is particularly sensitive to the implemented *a priori* information.

3.2.2 Gaussian processes and smooth potentials

In this section we include, in addition to the likelihood terms, *a priori* information in form of a prior density $p_0(v)$. Having specified $p_0(v)$ a Bayesian approach aims at calculating the predictive density (3). Within a maximum posterior approximation the functional integral in Eq. (3) can be calculated by Monte Carlo methods or, as we will do in the following, in saddle point approximation, i.e., by selecting the potential with maximal posterior. The posterior density of v is according to Eq. (1) proportional to the product of training likelihood and prior

$$p(v|D) \propto p_0(v) \prod_i < |\phi(x_i)|^2 > . \tag{33}$$

Hence, the maximum likelihood approximation we have discussed in the last section is equivalent to a maximum posterior approximation under the assumption of a uniform prior.

Technically the most convenient priors are *Gaussian processes* which we already have introduced in (8) for regression models. Such priors read for v,

$$p_0(v) = \left(\det \frac{\lambda}{2\pi}\mathbf{K}_0\right)^{\frac{1}{2}} e^{-\frac{\lambda}{2}<v-v_0\,|\,\mathbf{K}_0\,|\,v-v_0>}, \tag{34}$$

with mean v_0, representing a reference potential or template for v, and real symmetric, positive (semi–)definite covariance operator $(1/\lambda)\mathbf{K}_0^{-1}$, acting on potentials v and not on wave functions ϕ_α. The operator $\mathbf{K}_0$ defines a scalar product and thus a distance measuring the deviation of v from v_0. The most common priors are *smoothness priors* where $\mathbf{K}_0$ is taken as differential operator. (In that case $\mathbf{K}_0$ defines a Sobolev distance.) Examples of smoothness related inverse prior covariances are the negative Laplacian $\mathbf{K}_0 = -\Delta$, which we have already met in Eq. (10), or operators with higher derivatives like a "Radial Basis Functions" prior with pseudo-differential operator $\mathbf{K}_0 = \exp\left(-\sigma^2_{\rm RBF}\Delta/2\right)$.

Finally, we want to mention that also the prior density can be parameterized, making it more flexible. Parameters of the prior density, also known as *hyperparameters*, are in a Bayesian framework included as integration variables in Eq. (3), or, in maximum posterior approximation, in the maximization of Eq. (5) [22, 26]. Hyperparameters allow to transform the point–like maxima of Gaussian priors to submanifolds of optimal solutions. For a Gaussian process prior, for example, the mean or reference potential v_0 and the covariance $\mathbf{K}_0^{-1}/\lambda$ can so be adapted to the data [32].

3.2.3 Approximate symmetries

To be more general let us consider *a priori* information related to some *approximate symmetry* [67]. In contrast to an exact symmetry where it is sufficient to restrict v to be symmetric, approximate symmetries require the definition of a distance measuring the deviation from exact symmetry. In particular, consider a unitary symmetry operation S, i.e., $S^\dagger = S^{-1}$, $S^\dagger$ denoting the hermitian conjugate of S. Further, define an operator $\mathbf{S}$ acting on (local or nonlocal) potentials V, by $\mathbf{S}V = S^\dagger V S$. In case of an exact symmetry V commutes with S, i.e., $[V, S] = 0$ and thus $\mathbf{S}V = S^\dagger V S = V$. In case of an approximate symmetry we may choose a prior

$$p_0(V) \propto e^{-E_S}, \tag{35}$$

with 'symmetry energy' or 'symmetry error'

$$E_S = \frac{1}{2} <V - \mathbf{S}V\,|\,\mathbf{K}_S\,|\,V - \mathbf{S}V> = \frac{1}{2} <V\,|\,\mathbf{K}_0\,|\,V>, \tag{36}$$

some positive (semi–)definite $\mathbf{K}_S$, hence positive semi–definite $\mathbf{K}_0 = (\mathbf{I} - \mathbf{S})^\dagger \mathbf{K}_S (\mathbf{I} - \mathbf{S})$, $\mathbf{I}$ denoting the identity. (Symmetric V are within the Null space of $\mathbf{K}_S$.) If $\mathbf{S}$ belongs to a Lie group it can be expressed by a Lie group parameter θ and the generator $\mathbf{s}$ of the corresponding infinitesimal symmetry operation as $\mathbf{S}(\theta) = \exp(\theta \mathbf{s})$. Hence, we can define an error with respect to the infinitesimal operation $\mathbf{s}$ with, say, $\mathbf{K}_S = \mathbf{I}$,

$$E_s = \lim_{\theta \to 0} \frac{1}{2} < \tfrac{V - \mathbf{S}(\theta)V}{\theta} \,|\, \tfrac{V - \mathbf{S}(\theta)V}{\theta} > = \frac{1}{2} < V \,|\, \mathbf{s}^\dagger \mathbf{s} \,|\, V > . \tag{37}$$

Choosing, for instance, $\mathbf{s}$ as the derivative operator (for vanishing or periodic boundary terms) results in the typical Laplacian smoothness prior which measures the degree of symmetry of v under infinitesimal translations.

Another possibility to implement approximate symmetries is given by a prior with symmetric reference potential $V_S = \mathbf{S} V_S$

$$E_{V_S} = \frac{1}{2} < V - V_S \,|\, V - V_S > . \tag{38}$$

In contrast to Eq. (36) which is minimized by any symmetric V, this term is minimized only by $V = V_S$. Note, that also in Eq. (36) an explicit non–zero reference potential V_0 can be included, meaning that not V but the difference $V - V_0$ is expected to be approximately symmetric.

Finally, a certain deviation κ from exact symmetry might even be expected. This can be implemented by including 'generalized data terms' [32]

$$E_{S,\kappa} = \frac{1}{2}(E_S(V) - \kappa)^2 = \frac{1}{2}(\frac{1}{2} < V - \mathbf{S}V \,|\, V - \mathbf{S}V > - \kappa)^2, \tag{39}$$

similar to the usual mean squared error terms used in regression.

3.2.4 Mixtures of Gaussian process priors

Stochastic process priors have, compared to priors over parameters ξ, the advantage of implementing *a priori* knowledge explicitly in terms of the function values $v(x)$. Gaussian processes, in particular, always correspond to simple quadratic error surfaces, i.e., concave densities. Being technically very convenient, this is, on the other hand, a strong restriction. Arbitrary prior processes, however, can easily be built by using mixtures of Gaussian processes without loosing the advantage of an explicit prior implementation [32, 33, 63, 67]. (We want to point out that using a mixture of Gaussian process priors does not restrict v to a mixture of Gaussians.)

A mixture of M Gaussian processes with component means v_k and inverse component covariances $\lambda \mathbf{K}_k$ reads

$$p_0(v) = \sum_k^M p(v,k) = \sum_k^M p(k)\, p_0(v|k) = \sum_k^M \frac{p(k)}{Z_k} e^{-\frac{\lambda}{2} < v - v_k \,|\, \mathbf{K}_k \,|\, v - v_k >} \tag{40}$$

with $Z_k = \left(\det \frac{\lambda}{2\pi} \mathbf{K}_k\right)^{-\frac{1}{2}}$ and mixture probabilities $p(k)$. The parameter λ plays the role of an inverse mixture temperature. Analogous to annealing techniques changing λ allows to control the degree of concavity of the mixture [33, 63].

3.2.5 Average energy

Using a standard Gaussian smoothness prior as in Eq. (34) with zero reference potential $v_0 \equiv 0$ (and, say, zero boundary conditions for v) leads to flat potentials for large smoothness factors λ. Especially in such cases it turned out to be useful to include besides smoothness also *a priori* information which determines the depth of the potential. One such possibility is to include information about the *average energy*

$$U = \sum_\alpha p_\alpha E_\alpha = < E > . \tag{41}$$

We may remark, that for fixed boundary values of v a certain average energy cannot be obtained by simply adding a constant to the potential. The average energy can, however, be set to a value κ by including a Lagrange multiplier term

$$E_U = \mu(U - \kappa), \tag{42}$$

Similarly, and technically sometimes easier, one can include noisy 'energy data' of the form

$$p_U \propto e^{-E_U}, \quad E_U = \frac{\mu}{2}(U - \kappa)^2. \tag{43}$$

For $\mu \to \infty$ this results in $U \to \kappa$ so both approaches coincide.

3.2.6 Maximum posterior approximation

Let us consider prior densities being a product of a Gaussian prior p_0 as in Eq. (34), or more general a mixture of Gaussian processes as in Eq. (40), and a non–Gaussian energy prior p_U of the form of Eq. (43). In that case, the stationarity equation we have to solve to maximize the posterior density of Eq. (33) reads

$$0 = \delta_{v(x)} \ln p_0(v) + \delta_{v(x)} \ln p_U(v) + \sum_i \delta_{v(x)} \ln p(x_i|\hat{x}, v). \tag{44}$$

While $\delta_{v(x)} p(x_i|\hat{x}, v)$ has already been calculated in Sect. 3.1.3, we now need also $\delta_{v(x)} p_0(v)$ and $\delta_{v(x)} p_U(v)$ For a Gaussian $p_0(v)$ the functional derivative is easily found to be

$$\delta_v \ln p_0 = \frac{\delta_v p_0}{p_0} = -\lambda \mathbf{K}_0 (v - v_0). \tag{45}$$

Similarly, for a mixture of Gaussian processes it is not difficult to show that

$$\delta_v \ln p_0 = -\lambda \sum_k^M p_0(k|v) \mathbf{K}_k (v - v_k) \tag{46}$$

where $p_0(k|v) = p_0(v, k)/p_0(v)$.

To get the functional derivative of the non–Gaussian p_U we calculate first

$$\delta_{v(x)} U = < \delta_{v(x)} E > -\beta < E\, \delta_{v(x)} E > +\beta < E >< \delta_{v(x)} E > . \tag{47}$$

As $\delta_{v(x)} E_\alpha$ has been found in Eq. (29) this yields

$$\delta_{v(x)} E_U = \mu\,(U - \kappa) \left(< |\phi(x)|^2 > -\beta \left(< E\, |\phi(x)|^2 > -U < |\phi(x)|^2 > \right) \right) . \tag{48}$$

Collecting all terms, we can now solve the stationarity equation (44) by iteration

$$v^{\text{new}} = v^{\text{old}} + \eta \mathbf{A}^{-1} \left(\lambda \mathbf{K}_0 (v_0 - v^{\text{old}}) + \sum_i \delta_x \ln p(x_i|\hat{x}, v^{\text{old}}) - \delta_x E_U \right), \tag{49}$$

where we introduced $\mathbf{K}_0 = \sum_k p_0(k|v)\, \mathbf{K}_k$ and $v_0 = \mathbf{K}_0^{-1} \sum_k p_0(k|v)\, \mathbf{K}_k v_k$ and a step width η and positive definite iteration matrix $\mathbf{A}$ has to be selected. Choosing $\mathbf{A}$ as the identity matrix means moving in the direction of the gradient of the posterior. Taking for $\mathbf{A}$ the Hessian one obtains the Newton method. Quasi–Newton methods, like the DFP (Davidon–Fletcher–Powell) or BFGS (Broyden–Fletcher–Goldfarb–Shanno) variable metric methods, approximate the Hessian iteratively [68–72] (For the case of solving for continuous functions see [73].)

A simple and useful choice in our case is $\mathbf{A} = \mathbf{K}_0$ which approximates the Hessian. For a single Gaussian prior this choice does not depend on v and has thus not to be recalculated during iteration. Eq. (49) then becomes

$$v^{\text{new}} = (1-\eta) v^{\text{old}} + \eta \left(v_0 + (\lambda \mathbf{K}_0)^{-1} \Big(\sum_i \delta_x \ln p(x_i|\hat{x}, v^{\text{old}}) - \delta_x E_U \Big) \right). \tag{50}$$

Due to the nonparametric approach for the potential combined with *a priori* information implemented as stochastic process, the Bayesian approach formulated in the previous sections is clearly computationally demanding. The situation for inverse quantum theory is worse than, e.g., for Gaussian process priors in regression problems (i.e., for a Gaussian likelihood, local in the regression function) where it is only necessary to work with matrices having a dimension equal to the number of training data [16, 48]. In our case, where the likelihood is nonlocal in the potential and also non–Gaussian prior terms may occur, the stationarity equation has to be solved be discretizing the problem.

The following section demonstrates that at least one–dimensional problems can be solved numerically without further approximation. Higher dimensional problems, however, e.g., for many–body systems, require additional approximations. In such higher dimensional situations, the potential may be parameterized (without skipping necessarily the prior terms) or the problem has to be divided in lower dimensional subproblems, e.g., by restricting v to certain (typically, additive or multiplicative) combinations of lower dimensional functions. (Similarly, for example to additive models [19] projection pursuit [20] or neural network like [22] approaches.)

3.3 Numerical case studies in inverse quantum statistics

3.3.1 Approximately periodic potentials

As a first numerical application we discuss the reconstruction of a one–dimensional periodic potential. For example, such a potential may represent a one–dimensional solid surface. To be specific, assume we expect the potential to be periodic, or even more, to be similar to a certain periodic reference potential v_0. However, we do not want to restrict the approximation to a parametric

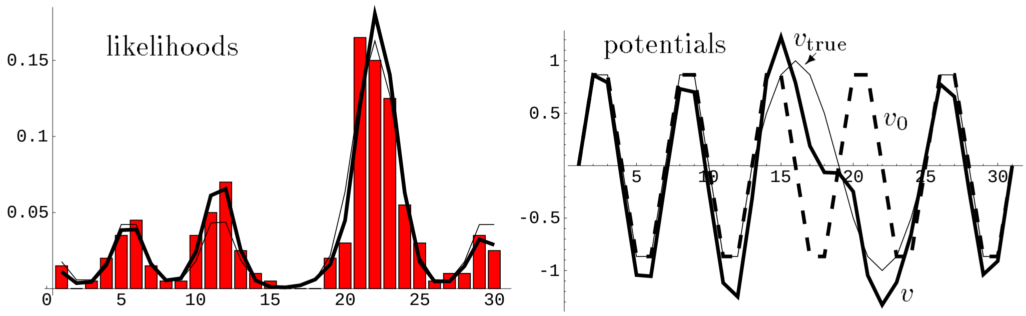

Figure 2. Reconstruction of an approximately periodic potential from empirical data. The left hand side shows likelihoods and the right hand side potentials: Original likelihood and potential v_{true} (thin lines), approximated likelihood and potential v (thick lines), empirical density (bars), reference potential v_0 (dashed). The parameters used are: 200 data points for a particle with mass $m = 0.25$, inverse physical temperature $\beta = 4$, inverse Laplacian covariance with $\lambda = 0.2$. (Average energy $U(v_{\text{true}}) = -0.3539$ for original potential and $U(v) = -0.5521$ for the reconstructed potential.) Notice, that the reconstructed potential v shows clearly the deviation from the strictly periodic reference potential v_0.

form but want to keep the approximating potential flexible, so it can adapt to arbitrary deviations from the periodic reference potential as indicated by the data. For example, such deviations may be caused by impurities on an otherwise regular surface. Assuming the deviations from the reference to be smooth on a scale defined by λ, these assumptions can be implemented by a Gaussian smoothness prior with mean v_0, and, say, Laplacian inverse covariance. Including the likelihood terms for the empirical data, and possibly a term p_U adapting the average energy, we end up with an error functional (negative log–posterior) to be minimized

$$-\ln p(v|D) = -\sum_i \ln p(x_i|\hat{x}, v) - \frac{\lambda}{2} < v - v_0 \,|\, \Delta \,|\, v - v_0 > + E_U. \tag{51}$$

Fig. 2 shows representative examples of numerical results for functional (51) without energy penalty term E_U and with a periodic reference potential (dashed line), $v_0(x) = \sin(\pi x/3)$, on a one–dimensional grid with 30 points. Data have been sampled according to a likelihood function derived from a 'true' or original potential v_{true} (shown as thin line) under periodic boundary conditions for ϕ_α. The reconstructed potential (thick line) has been obtained by minimizing Eq. (51) iterating according to Eq. (49) with $\mathbf{A} = -\lambda\Delta$ and zero boundary conditions for v (so $\mathbf{A}$ is invertible) and initial guess $v^{(0)} = v_0$.

Notice, that the distortion of the underlying 'true' potential has been clearly identified. On the other hand the reconstructed potential coincides well with the periodic reference potential at locations where supported by data.

We want to stress two phenomena which are typical for the reconstruction of potentials from empirical data and can also be seen in the figures. Firstly, the approximation of the likelihood function is usually better than the approximation of the potential. This is due to the fact that quite different potentials can

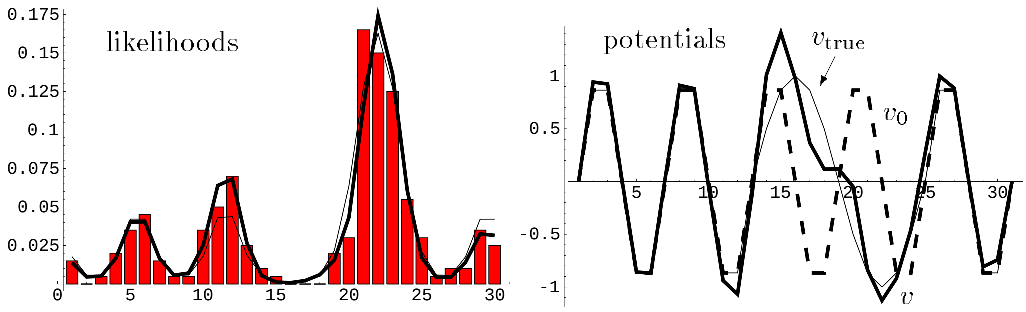

Figure 3. Same data and parameter as for Fig. 2, except for a nonzero energy penalty term E_U with $\mu = 1000$ and $\kappa = -0.3539 = U(v_{\rm true})$. While there are, compared to Fig. 2, only slight modifications of the likelihood the average energy of the reconstructed potential, $U(v) = -0.3532$, is now nearly the same as that of the original potential.

produce similar likelihoods. This emphasizes the relevance of *a priori* information for reconstructing potentials. Secondly, especially in low data regions, i.e., at high potentials, the potential is not well determined. Thus, empirical data mainly contribute to the approximation of regions with low potential, while *a priori* information becomes especially important in regions where the potential is large. More data will can be obtained for high potential regions when the temperature is increased which spreads the data over a wider area. At the same time, however, the likelihood becomes more uniform at large temperatures, making an identification of v more difficult.

Because the reference potential v_0 has the same average energy U as the underlying original potential the results are already reasonable without energy penalty term E_U. Indeed, Fig. 3 shows the relatively small influence of an additional energy penalty term with quite large κ on the likelihood function. Thus, the approximated probability for empirical data is not much altered. The presence of an E_U term is better visible for the potential. In particular its minima fit now better that of the original. In the next section, where we will work with a zero reference potential $v_0 \equiv 0$, the energy penalty term E_U will be more important.

3.3.2 Inverse two–body problems

As a second example we study the reconstruction of a two–body potential by measuring inter–particle distances x_r. Consider the two–body problem

$$\frac{P_1^2}{2m_1} + \frac{P_2^2}{2m_2} + v(x_1 - x_2)\delta(x_1 - x_2 - x_1' + x_2')\delta(x_1 + x_2 - x_1' - x_2') \qquad (52)$$

with single particle momenta $P_i = -i\partial/\partial x_i$. The problem is transformed to a one–body problem in the relative coordinates in the usual way by introducing i.e., $x_r = x_1 - x_2$, $P_r = (m_1 P_1 - m_2 P_2)/(m_1 + m_2)$, $x_c = (m_1 x_1 + m_2 x_2)/(m_1 + m_2)$, P_c

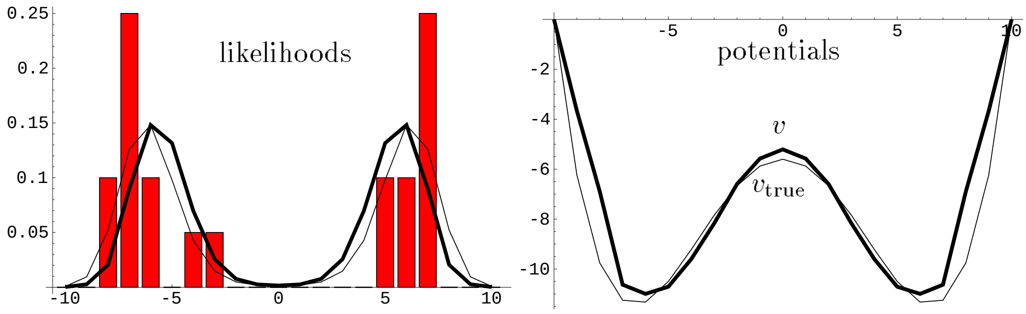

Figure 4. Approximation of symmetric potential. Shown are likelihoods (left hand side) and potentials (right hand side): Original likelihood and potential (thin lines), approximated likelihood and potential (thick lines), empirical density (bars), The parameters used are: 20 data points for a particle with $m = 0.1$, truncated RBF covariances as in Eq. (54) with $\sigma_{\rm RBF} = 7$, $\lambda = 0.001$, energy penalty term E_U with $\mu = 20$ and reference value $\kappa = -9.66 = U(v_{\rm true})$ (average energy $U(v) = -9.33$ for the approximated v, ground state energy $E_0(v) = -9.52$) inverse physical temperature $\beta = 1$, and a potential fulfilling $v(x) = v(-x)$ and $v = 0$ at the boundaries.

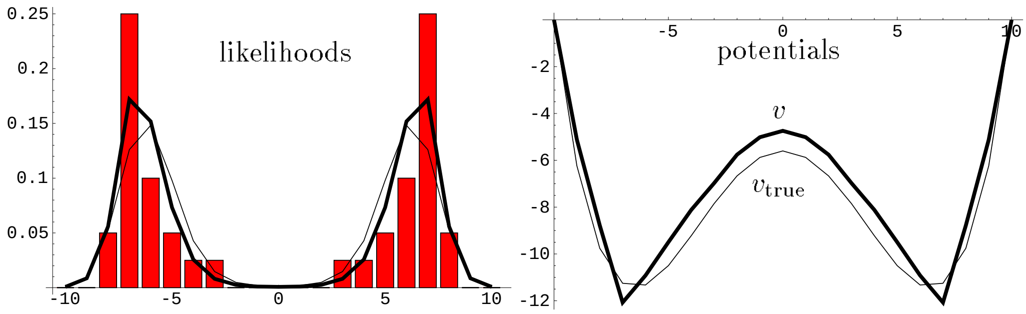

Figure 5. Same data and parameter as for Fig. 4 with the exception of $\sigma_{\rm RBF} = 4$, that means with a smaller smoothness constraint, and $\mu = 5$. ($U(v) = -9.46$.) To allow an easier comparison with the reconstructed likelihood the figure shows the symmetrized empirical density $P_{\rm sym} = (P_{\rm emp}(x) + P_{\rm emp}(-x))/2$.

$= P_1 + P_2$, $m = (m_1 m_2)/(m_1 + m_2)$, and $M = m_1 + m_2$ resulting in

$$\left(\frac{P_r^2}{2m} + v(x_r)\right)\psi_\alpha(x_r) = E_\alpha \psi_\alpha(x_r). \tag{53}$$

The total energy is additive $E_\alpha^{\rm total.}(P_c) = E_\alpha + P_c^2/(2M)$ so the thermal probabilities $p^{\rm total}$ factorize and integrating out the center of mass motion leaves $p_\alpha = e^{-\beta E_\alpha}/Z$, with E_α being the eigenvalues of Eq. (53).

Figs. 4 – 6 show typical results for the numerical reconstruction of a one–dimensional, strictly symmetric potential, fulfilling $v(x) = v(-x)$ and set to zero at the boundaries. Training data have been sampled from a 'true' likelihood function (thin lines), resulting in an empirical density $P_{\rm emp}(x) = n(x)/n$ (shown by

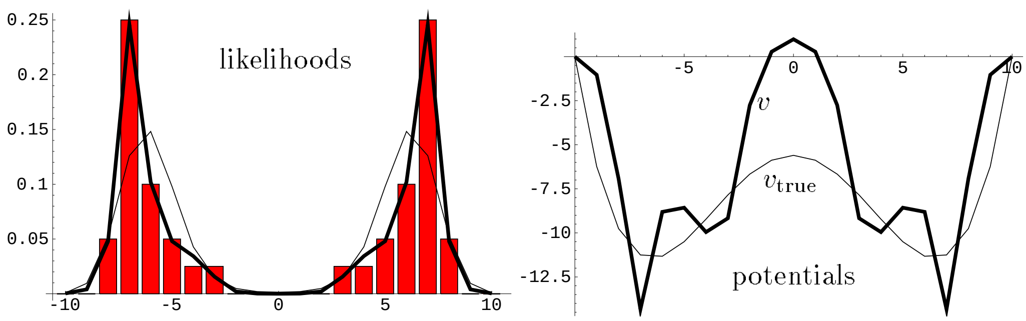

Figure 6. Same data and parameter as for Fig. 5 but with even smaller smoothness constraint $\sigma_{\rm RBF} = 1$. ($\mu = 5$, empirical density symmetrized, $U(v) = -9.59$.) Compared with Figs. 4 and 5, the empirical density is better approximated but not the original potential and its likelihood function.

bars), where $n(x)$ denotes the number of times the value x occurs in the training data. The 'true' likelihood has been constructed from a 'true' potential (thin lines) choosing periodic boundary conditions for the wavefunctions. In contrast to Sect. 3.3.1 a zero reference potential $v_0 \equiv 0$ and a truncated Radial Basis Function (RBF) prior [32] has been used

$$\mathbf{K}_0 = \sum_{k=0}^{3} \frac{\sigma_{\rm RBF}^{2k}}{k!2^k}(-1)^k \Delta^k, \tag{54}$$

(Δ^k denoting the kth iterated Laplacian) which includes, compared to a Laplacian prior, higher derivatives, hence producing a rounder reconstructed potential (cmp. Sect. 3.2.2). The approximated potentials (thick lines) have been obtained by iterating Eq. (49), including a term E_U adapting the thermal energy average to that of the original potential. As iteration matrix we used $\mathbf{A} = \lambda \mathbf{K}_0$ together with an adaptive step size η. An initial guess for the potential has been obtained by adding negative δ–peaks on the data points (except for data on the boundary), i.e., $v^{(0)} = -\sum_i \delta_{x,x_i}$. The number of iterations necessary to obtain convergence has been typically between 50 and 100.

Comparing Figs. 4 – 6 one sees that a smaller smoothness leads to a better fit of the empirical density. A larger smoothness, on the other hand, leads to better fit in regions where smoothness is an adequate prior. Near the boundaries, however, where the original is relatively steep, a higher smoothness leads to a poorer approximation. A remedy would be, for example, an adapted reference potential v_0.

Fig. 7 presents an application of a mixture of Gaussian process priors as given in Eq. (40). Such mixture priors can in principle be used to construct an arbitrary prior density, adapted to the situation under study. For the numerical example a two component mixture has been chosen with equal component variances $\mathbf{K}_k = \mathbf{K}_0$ of the form of Eq. (54) and two reference potentials v_i (shown as dashed lines) with the same average energy U. In the special situation shown in the figure both

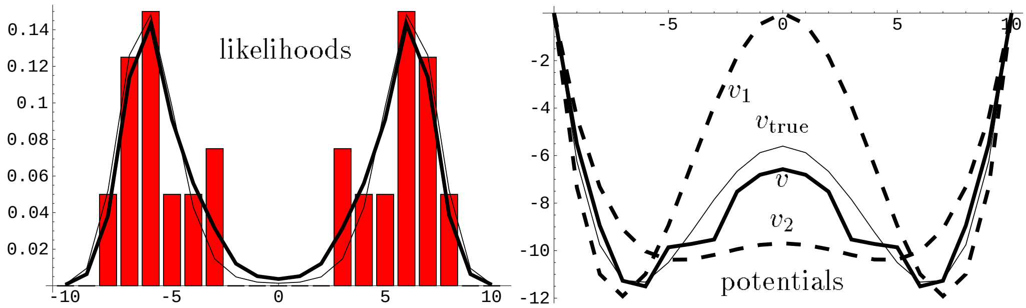

Figure 7. Approximation of symmetric potential with mixture of Gaussian process priors. The left hand side shows likelihoods and the right hand side potentials: Original likelihood and potential (thin lines), approximated likelihood and potential (thick lines), symmetrized empirical density (bars), and the two reference potentials v_1,v_2 (dashed, v_2 deeper in the middle). The parameters used are: 20 data points for a particle with $m = 0.1$, inverse physical temperature $\beta = 1$, $\mathbf{K}_0 = -\Delta$, inverse mixture temperature $\lambda = 0.1$, energy penalty factor $\mu = 10$ for average energy $\kappa = -9.66 = U(v_{\rm true})$ (and $U(v) = -9.55$, $E_0(v) = -9.82$) $v(x) = v(-x)$ symmetric, and $v = 0$ at the boundaries. Because the data support both reference potentials v_1 and v_2, the approximated v is in regions with no data essentially a smoothed mixture between v_1 and v_2 with mixture coefficients for prior components $p_0(1|v) = 0.3$, $p_0(2|v) = 0.7$.

reference functions v_i fit similarly well to the empirical data. (The final mixture coefficients for v_1 and v_2 are $p_0(1|v) = 0.3$ and $p_0(2|v) = 0.7$.) Hence, in the no–data region the approximated potential v becomes a smoothed, weighted average of v_1 and v_2. Because both reference potentials coincide also relatively well with the original v near the boundaries, the approximation in Fig. 7 is better than in Figs. 4 – 6.

In conclusion, the two one–dimensional examples show that a direct numerical solution of the presented Bayesian approach to inverse quantum theory can be feasible.

3.4 Classical approximation

Before discussing a possible approximation for many–body systems we will first study the classical limit of inverse quantum statistics. The classical limit is much easier to solve than the full quantum mechanical problem and may, for example for large masses, already give a useful approximation.

The phase space density of a classical canonical ensemble is given by

$$p(x, p_{\rm cl}|v) = Z^{-1} e^{-\beta\left(\frac{p_{\rm cl}^2}{2m}+v(x)\right)}, \tag{55}$$

with

$$Z = \int dp_{\rm cl}\, dx\, e^{-\beta\left(\frac{p_{\rm cl}^2}{2m}+v(x)\right)}. \tag{56}$$

Here we used $p_{\rm cl}$ to denote the classical momentum to distinguish it from a density p. The probability $p(x|v)$ for measuring x [to simplify the notation we abstain in this context from denoting the observable O explicitly] is then obtained by integrating over $p_{\rm cl}$,

$$p(x|v) = \int dp_{\rm cl}\, p(x,p_{\rm cl}|v) = Z_x^{-1} e^{-\beta v(x)}, \tag{57}$$

where

$$Z_x = \int dx\, e^{-\beta v(x)}. \tag{58}$$

Notice, that the classical $p(x|v)$ is mass independent, and, most important, that it can be obtained directly from $v(x)$ without having to solve an eigenvalue problem like in the quantum case.

Analogously to the quantum mechanical approach the classical likelihood model (57) for position measurements can now be combined with a prior model for potentials v, leading to a posterior density $p(v|D)$. In particular, adding a Gaussian process prior the log-posterior becomes

$$\ln p(v|D) = -\beta \sum_{i=1}^{n} v(x_i) - \frac{\lambda}{2} < v - v_0 \,|\, \mathbf{K}_0 \,|\, v - v_0 > -n \ln Z_x - \det\left(\frac{\lambda}{2\pi}\mathbf{K}_0\right)^{\frac{1}{2}}. \tag{59}$$

Again, we may refer to a maximum posterior approximation and consider the potential which maximizes the posterior as the solution of our reconstruction problem. The corresponding stationarity equation is found by setting the functional derivative of the log-posterior with respect to $v(x)$ to zero,

$$0 = \delta_v \ln p(v|D) = -\beta N - \lambda \mathbf{K}_0 (v - v_0) + n\beta p(x|v). \tag{60}$$

Choosing an initial guess $v^{(0)}$ Eq. (60) can be solved by straightforward iteration. The results of a classical calculation (with parameters and data as in Fig. 4, but without energy penalty term) are shown in Fig. 8.

4 Inverse many–body theory

4.1 Systems of Fermions

In this section the Bayesian approach for inverse problems will be applied to many–body systems. To be specific, we will study the simultaneous measurement of the positions of N particles. We assume the measurement result to be given as a vector x_i consisting of N single particle coordinates $x_{i,j}$. The treatment can easily be generalized to partial measurements of x_i by including an integration over components which have not been observed. The likelihood for v, if measuring a vector x_i of coordinates, becomes for a many–body system

$$p(x_i|\hat{x},v) = \mathrm{Tr}\Big(|\, x_{i,1}, \cdots x_{i,N} >< x_{i,1}, \cdots x_{i,N} \,|\, \rho\Big), \tag{61}$$

which is now a thermal expectation with respect to many–body energies E_α

$$p(x_i|\hat{x},v) = \sum_\alpha p_\alpha |\psi_\alpha^{(N)}(x_{i,1}, \cdots x_{i,N})|^2 =< |\psi^{(N)}(x_{i,1}, \cdots x_{i,N})|^2 > . \tag{62}$$

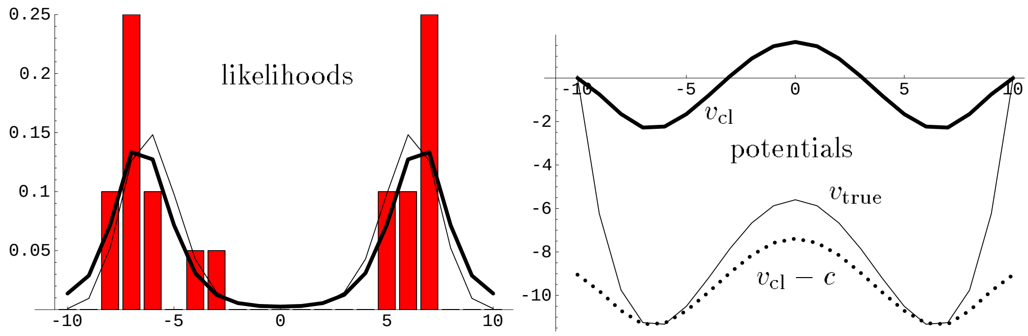

Figure 8. Classical approximation of symmetric potential. Shown are likelihoods (left hand side) and potentials (right hand side): Original likelihood and potential (thin lines), approximated likelihood and potential $v_{\rm cl}$ (thick lines), empirical density (bars). The dotted line shows $v_{\rm cl} - c$ with constant $c = \min[v_{\rm cl}] - \min[v_{\rm true}]$. Except for the fact that no energy penalty term has been used for this classical calculation the parameters and data are the same as in Fig. 4. (20 data points, sampled from the true quantum mechanical likelihood, truncated RBF covariances (54) with $\sigma_{\rm RBF} = 7$, $\lambda = 0.001$, inverse physical temperature $\beta = 1$, $v(x) = v(-x)$ and $v = 0$ at the boundaries.)

In particular, we will be interested in *fermions* for which the wave functions ψ_α and $|\, x_{i,1}, \cdots x_{i,N} >$ have to be antisymmetric. Considering a canonical ensemble, the density operator ρ has still the form of Eq. (15), but with H replaced now by a many–body Hamiltonian. For fermions, it is convenient to express the many–body Hamiltonian in second quantization, i.e., in terms of creation and annihilation operators [76, 77]. A Hamiltonian with one–body part T, e.g., $T = -(1/2m)\mathit{\Delta}$, and two–body potential V can so be written

$$H = T + V = \sum_{ij} T_{ij}\, a_i^\dagger a_j + \frac{1}{4}\sum_{ijkl} V_{ijkl}\, a_i^\dagger a_j^\dagger a_l a_k, \tag{63}$$

with antisymmetrized matrix elements V_{ijkl}. Hereby, $a_\alpha a_\gamma^\dagger + a_\gamma^\dagger a_\alpha = <\varphi_\alpha \,|\, \varphi_\gamma>$ is equal to the overlap of the one–body orbitals $|\, \varphi_\alpha> = a_\alpha^\dagger |\, 0>$ and $|\, \varphi_\gamma> = a_\gamma^\dagger |\, 0>$ which are created or destroyed by the operators $a_\gamma^\dagger$ or a_α, respectively. Furthermore, $a_\alpha^\dagger a_\gamma^\dagger + a_\gamma^\dagger a_\alpha^\dagger = 0$, $a_\alpha a_\gamma + a_\gamma a_\alpha = 0$. A two–body eigenfunction of the Hamiltonian (63) can for example be expanded as follows

$$|\, \psi_\alpha^{(2)}> = \sum_{\alpha,\gamma} c_{\alpha,\gamma} |\, \varphi_\alpha, \varphi_\gamma>, \tag{64}$$

where $|\, \varphi_\alpha, \varphi_\gamma> = a_\alpha^\dagger a_\gamma^\dagger |\, 0>$ denotes a Slater determinant being an antisymmetrized wavefunction.

The symmetrized version of a potential, local in relative coordinates, is

$$V_{x_1 x_2 x_1' x_2'} = v(|x_1 - x_2|)\Big(\delta(x_1 - x_1')\delta(x_2 - x_2') - \delta(x_1 - x_2')\delta(x_2 - x_1')\Big). \tag{65}$$

Here we can always choose $v(0) = 0$. Now, assume we are interested in the reconstruction of $v(x)$ for $x > 0$. Solving the stationarity equation of the maximum posterior approximation analogous to Eq. (44) of Sect. 3, the prior terms remains unchanged and only the likelihood terms have to be adapted. Using

$$\delta_{v(x)} v(|x_1 - x_2|) = \delta(x - |x_1 - x_2|), \tag{66}$$

we find

$$\delta_{v(x)} H = \frac{1}{2} \sum_{x_1} a^\dagger_{x_1} a^\dagger_{x_1 - x} a_{x_1 - x} a_{x_1} + \frac{1}{2} \sum_{x_1} a^\dagger_{x_1} a^\dagger_{x_1 + x} a_{x_1 + x} a_{x_1}, \tag{67}$$

where $x > 0$, and can write, similar to the one–body case,

$$\delta_{v(x)} E_\alpha = \frac{<\psi_\alpha \,|\, \delta_{v(x)} H \,|\, \psi_\alpha>}{<\psi_\alpha \,|\, \psi_\alpha>}, \tag{68}$$

$$|\, \delta_{v(x)} \psi_\alpha> = \sum_{\substack{\gamma \\ E_\gamma \neq E_\alpha}} \frac{1}{E_\alpha - E_\gamma} |\, \psi_\gamma><\psi_\gamma \,| \delta_{v(x)} H |\, \psi_\alpha>. \tag{69}$$

From this the functional derivatives of the likelihoods, $\delta_{v(x)} p(x_i|\hat{x}, v)$, can be obtained. However, a direct numerical or analytical solution of the full inverse many–body equations is usually not feasible. To deal with this problem, a mean field approach will be developed in the next section.

4.2 Inverse Hartree–Fock theory

To tackle the inverse many–body problem we will treat it in Hartree–Fock approximation [74–77]. Thus, we replace the full many–body Hamiltonian H by a one–body Hartree–Fock Hamiltonian $H^{HF} = \sum_{kl} h_{kl} a^\dagger_k a_l$ with matrix elements h defined, for example in coordinate representation, as

$$h_{xx'} = T_{xx'} + \sum_k^N <x \varphi_k \,|\, V \,|\, x' \varphi_k>, \tag{70}$$

the φ_k being the N–lowest (orthonormalized) eigenstates of h. The corresponding eigenvalue equation

$$h \varphi_k = \epsilon_k \varphi_k, \tag{71}$$

is nonlinear, due to the φ_k–dependent definition (70) of h, and has to be solved by iteration. The Hartree–Fock ground state is given by the Slater determinant $|\,\Phi_0> = \det\{|\,\varphi_k>\}$ made from the N–lowest orbitals, and has energy $E_0^{HF} = \sum_k^N t_{kk} + \frac{1}{2} \sum_{kl}^N v_{klkl} = \sum_k^N \epsilon_k - \frac{1}{2} \sum_{kl}^N v_{klkl}$. Considering now the case of zero temperature, the many–body likelihood for the true ground state ψ_0,

$$p(x_i|\hat{x}, \rho(v)) = <\psi_0 \,|\, x_i><x_i \,|\, \psi_0>, \tag{72}$$

becomes in Hartree–Fock approximation

$$p(x_i|\hat{x}, \rho_{HF}(v)) = <\Phi_0 \,|\, x_i><x_i \,|\, \Phi_0>. \tag{73}$$

The scalar product of the Hartree–Fock ground state Φ_0 and the many–body position eigenfunction $|\, x_i >$ corresponding to the measured vector x_i is a determinant and can be expanded in its cofactors $M_{kl;i}$

$$< x_i \,|\, \Phi_0 > = \det\{< x_{i,l} \,|\, \varphi_k >\} = \det B_i = \sum_l^N M_{kl;i} B_{kl;i}, \tag{74}$$

B_i being the matrix of overlaps with elements $B_{kl;i} = < x_{i,l} \,|\, \varphi_k > = \varphi_k(x_{i,l})$. (For the generalization to non–hermitian h see for example [76, 78].)

To maximize the posterior, we have to calculate the functional derivative of the Hartree–Fock likelihood with respect to the potential [79]

$$\delta_{v(x)} p(x_i|\hat{x}, \rho_{\rm HF}(v)) = < \delta_{v(x)}\Phi_0 \,|\, x_i >< x_i \,|\, \Phi_0 > + < \Phi_0 \,|\, x_i >< x_i \,|\, \delta_{v(x)}\Phi_0 >. \tag{75}$$

Here the factors

$$< x_i \,|\, \delta_{v(x)}\Phi_0 > = \sum_{kl}^N M_{kl;i} < x_{i,l} \,|\, \delta_{v(x)}\varphi_k > = \sum_{kl}^N M_{kl;i}\, \Delta_{kl;i}(x), \tag{76}$$

can be expressed by single particle derivatives $\Delta_{kl;i}(x) = < x_{i,l} \,|\, \delta_{v(x)}\varphi_k > = \delta_{v(x)}\varphi_k(x_{i,l})$. Analogously to Sect. 3 the functional derivatives $\delta_{v(x)}\varphi_k$ can be obtained from the functional derivative of Eq. (71)

$$(\delta_{v(x)} h)\, \varphi_k + h\, \delta_{v(x)}\varphi_k = (\delta_{v(x)}\epsilon_k)\, \varphi_k + \epsilon_k\, \delta_{v(x)}\varphi_k. \tag{77}$$

Projecting onto $< \varphi_k \,|$ and using the hermitian conjugate of Eq. (71) we find the Hartree–Fock version of Eqs. (68) and (69)

$$\delta_{v(x)}\epsilon_k = \frac{< \varphi_k \,|\, \delta_{v(x)} h \,|\, \varphi_k >}{< \varphi_k \,|\, \varphi_k >}, \tag{78}$$

$$|\, \delta_{v(x)}\varphi_k > = \sum_{\substack{l \\ \epsilon_l \neq \epsilon_k}} \frac{1}{\epsilon_k - \epsilon_l} \,|\, \varphi_l >< \varphi_l \,|\delta_{v(x)} h|\, \varphi_k >, \tag{79}$$

where we, as done before, have fixed orthonormalization and phases by choosing $< \delta_{v(x)}\varphi_k \,|\, \varphi_l > = 0$ for orbitals with equal energy. In contrast to Sect. 3, however, h, and thus $\delta_{v(x)} h$, now obey a nonlinear equation. Indeed, from Eq. (70) it follows

$$\begin{aligned} \delta_{v(x)} h_{x'x''} = \sum_k^N \Big(& < x'\, \varphi_k \,|\, \delta_{v(x)} V \,|\, x''\, \varphi_k > \\ & + < x'\, \delta_{v(x)}\varphi_k \,|\, V \,|\, x''\, \varphi_k > + < x'\, \varphi_k \,|\, V \,|\, x''\, \delta_{v(x)}\varphi_k > \Big). \end{aligned} \tag{80}$$

Inserting Eq. (78) and Eq. (80) into Eq. (79), we obtain the inverse Hartree–Fock equation for $\delta_{v(x)}\varphi_j$

$$\begin{aligned} \delta_{v(x)}\varphi_k(x') = & \sum_{\substack{l \\ \epsilon_l \neq \epsilon_k}} \frac{1}{\epsilon_k - \epsilon_l}\, \varphi_l(x') \sum_j^N \Big(< \varphi_l \varphi_j \,|\, \delta_{v(x)} V \,|\, \varphi_k \varphi_j > \\ + & < \varphi_l \delta_{v(x)}\varphi_j \,|\, V \,|\, \varphi_k \varphi_j > + < \varphi_l \varphi_j \,|\, V \,|\, \varphi_k \delta_{v(x)}\varphi_j > \Big). \end{aligned} \tag{81}$$

Recalling the definition of the antisymmetric matrix elements of V we finally arrive at

$$\delta_{v(x)}\varphi_k(x') = \sum_{\substack{l\\ \epsilon_l\neq\epsilon_k}} \frac{1}{\epsilon_k-\epsilon_l}\,\varphi_l(x')\,\sum_j^N \times \tag{82}$$
$$\Bigg(\int\! dz\,\varphi_l^*(z)\varphi_j^*(z-x)\Big(\varphi_k(z)\varphi_j(z-x)-\varphi_k(z-x)\varphi_j(z)\Big)$$
$$+\int\! dz\,\varphi_l^*(z)\varphi_j^*(z+x)\Big(\varphi_k(z)\varphi_j(z+x)-\varphi_k(z+x)\varphi_j(z)\Big)$$
$$+\int\! dz\,dz'\,\varphi_l^*(z)\Big(\delta_{v(x)}\varphi_j^*(z')\Big)v(|z-z'|)\Big(\varphi_k(z)\varphi_j(z')-\varphi_k(z')\varphi_j(z)\Big)$$
$$+\int\! dz\,dz'\,\Big(\varphi_l^*(z)\varphi_j^*(z')-\varphi_l^*(z')\varphi_j^*(z)\Big)v(|z-z'|)\varphi_k(z)\delta_{v(x)}\varphi_j(z')\Bigg).$$

This linear equation can be solved directly (where for Hamiltonian with real matrix elements in coordinate space the orbitals, and thus their functional derivatives, can be chosen real) or, quite effectively, by iteration, starting for example with initial guess $\delta_{v(x)}\varphi_j(z')=0$. As the $\delta_{v(x)}\varphi_k(x')$, which are only required for the N lowest orbitals, depend on two position variables x, x', Eq. (82) has essentially the dimension of a two–body equation. Having calculated $\delta_{v(x)}\varphi_k(x_{i,l}) = \Delta_{kl;x,i}$ (for $x>0$, $1\le k\le N$, $1\le l\le N$, $1\le i\le n$) from Eq. (82) the likelihood terms in the stationarity equation (44) follow as

$$\begin{aligned}\delta_{v(x)}\ln p(x_i|\hat{x},\rho_{\rm HF}(v)) &= \frac{\sum_{kl}^N M_{kl;i}\Delta_{kl;i}(x)}{\det B_i}+\frac{\sum_{kl}^N M_{kl;i}^\dagger\Delta_{kl;i}^\dagger(x)}{\det B_i^\dagger}\\ &= {\rm Tr}(B_i^{-1}\Delta_i(x))+{\rm Tr}({B_i^\dagger}^{-1}\Delta_i^\dagger(x)),\end{aligned} \tag{83}$$

recalling that $M_{kl,i}=(B_i)_{lk}^{-1}\det B_i$ and defining analogously to B_i the matrix $\Delta_i(x)$ with elements $\Delta_{kl;i}(x)$. The freedom to linearly rearrange orbitals within the Slater determinants $\det\{|\,x_{i,l}\!>\}=\det\{|\,\tilde{x}_{i,l}\!>\}$ (for each data point i, analogously for $\det\{|\,\varphi_k\!>\}$), makes it possible to diagonalize the matrix of overlaps $<\!\varphi_k\,|\,\tilde{x}_{i,l}\!>$ in new orbitals $|\,\tilde{x}_{i,l}\!>$, which are then linear combinations of the $|\,x_{i,l}\!>$ [80, 78].

4.3 Numerical example of an inverse Hartree–Fock calculation

To test the numerical implementation of an inverse Hartree–Fock approach, we study a two–body problem, defined by the Hamiltonian

$$H=-\frac{1}{2m}\Delta+V_1+V_{\rm true}. \tag{84}$$

Herein we assume the local one–body potential $V_1(x,x')=\delta(x-x')v_1(x)$ to be given and the two–body potential V to be unknown, but local as in Eq. (65). Hence, our aim is to approximate the function $v(|x-x'|)$, defining the matrix

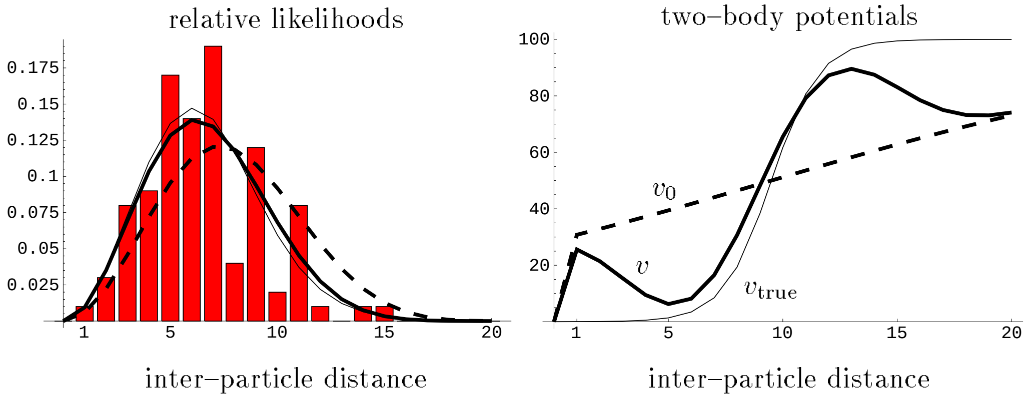

Figure 9. Inverse Hartree–Fock Approximation: The exact two–body likelihood has been calculated for two one–dimensional particles with many–body Hamiltonian H, given in Eq. (84), a local one–body potential $V_1(x,x') = \delta(x-x')a(x/10)^2$, $a = 10^{-3}$ breaking translational symmetry, mass $m = 10^{-3}$, and a given local two–body potential $V_{\rm true}$ of form (65) with $v_{\rm true}(|x-x'|)$ $= b/(1+e^{-2\gamma(x-k/2)/k})$, $b = 100$, $\gamma = 10$, $k = 21$, (thin line on the r.h.s.). As training data 100 pairs $\{(x_{i,1},x_{i,2})|1 \le i \le 100\}$ have been sampled according to that exact likelihood. The corresponding exact (thin line) and empirical (bars) likelihoods for inter–particle distances $|x_{i,1}-x_{i,2}|$ are shown on the l.h.s. of the figure. To reconstruct the potential a Gaussian prior has been used with $\lambda\mathbf{K}_0 = \lambda(\mathbf{I}-\Delta)/2$, $\lambda = 0.5\ 10^{-3}$, and reference potential (dashed on r.h.s) $v_0(|x-x'|) =$ $b/(1+e^{-2\gamma(x-k/2)/k})$, $b = 100$, $\gamma = 1$, $v_0(0) = 0$. The related reference likelihood of inter–particle distances is shown on the l.h.s.(dashed). The reconstructed potential v has been obtained by iterating with $\mathbf{A} = \mathbf{K}_0$ according to Eq. (50) and solving Eqs. (71) and (82) within each iteration step. The problem has been studied at zero temperature, v fulfilling the boundary conditions $v(0) = 0$ and $v =$ constant beyond the right boundary. No energy penalty term E_U had to be included. Note, that the number of data is not only small for large inter–particle distances where the potential is large, but also for small distances. This effect is due to antisymmetry which does not allow particles to be at the same place. Thus, the reconstructed potential v is nearly equal to the reference potential v_0 for large and for small distances.

elements of V, by using empirical data in combination with appropriate *a priori* information.

Fig. 9 shows the results of a corresponding inverse Hartree–Fock calculation. (The prior process and parameters are given in the figure caption, computational details will be presented elsewhere). For this two–body problem it is possible to calculate the exact solution and corresponding likelihood numerically. Hence, we was able to sample training data using the exact likelihood. Note that, besides the problem of simulating realistic data, an inverse Hartree–Fock calculation for more than two particles is not much more complex than for two particles. It only requires to add one single particle orbital for every additional particle. Thus, an analogous inverse Hartree–Fock calculation is clearly computationally feasible for many–body systems with three or more particles.

We have already discussed in previous sections that, in regions where the potential is large, the reconstruction of a potential is essentially based on *a priori* information. Training data are less important in such regions, because finding a particle there is very unlikely. In Fig. 9, for example, *a priori* information is thus especially important for large distances. A new, similar phenomenon occurs now when dealing with fermions: The antisymmetry, we have to require for fermions, forbids different particles to be at the same location. Hence, antisymmetry reduces the number of training data for small distances, and *a priori* information becomes especially important. This effect can clearly be seen in the figure, where the reconstructed potential v is influenced by the data mainly for medium distances. For large, but also for small inter–particle distances, the reconstructed potential is quite similar to the reference potential.

Summarizing, we note that for inverse Hartree–Fock problems in addition to the direct Hartree–Fock Eq. (71) a second equation (81) has to be solved determining the change of Hartree–Fock orbitals under a change of the potential. Despite this complication it was possible to solve the inverse Hartree–Fock equations numerically for the example problem considered in this section.

5 Conclusions

We have studied the inverse problem of reconstructing a quantum mechanical potential from empirical measurements. The approach presented in this paper is based on Bayesian statistics which has already been applied successfully to many empirical learning problems. For quantum mechanical systems, empirical data enter the formalism through the likelihood function as defined by the axioms of quantum mechanics. Additional *a priori* information is implemented in form of stochastic processes. The reconstructed potential is then found by maximizing the Bayesian posterior density.

The specific advantage of this new, nonparametric Bayesian approach to inverse quantum theory is the possibility to combine heterogeneous data, resulting from arbitrary quantum mechanical measurements, with a flexible and explicit implementation of *a priori* information.

Two numerical examples — the reconstruction of an approximately periodic potential and of a strictly symmetric potential — have demonstrated the computational feasibility of the Bayesian approach for one–dimensional systems. While a direct numerical solution is thus possible for one–dimensional problems, it becomes computationally demanding for two– or three dimensional problems.

As a possible approximation scheme for many–body systems an inverse Hartree–Fock approach has been proposed. An implementation of a corresponding reconstruction algorithm has been tested for a system of fermions, for which we were able to solve the inverse Hartree–Fock equation numerically.

Finally, we want to emphasize the flexibility of the Bayesian approach which can be easily adapted to a variety of different empirical learning situations. This includes, as we have seen, inverse problems in quantum theory at zero and at finite temperature, for single particles as well as for few– or many–body systems.

Acknowledgements

We are very grateful to A. Weiguny for many stimulating discussions.

References

1. Hadamard, J.: Lectures on the Cauchy Problem in Linear Partial Differential Equations. New Haven: Yale University Press 1923
2. Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. Soviet Math. Dokl. **4**, 1035–1038 (1963)
3. Tikhonov, A.N., Arsenin, V.: Solution of Ill–posed Problems. New York: Wiley 1977
4. Vapnik, V.N.: Estimation of dependencies based on empirical data. New York: Springer 1982
5. Keller, J.B.: Ann. Math. Mon. **83**, 107–118 (1976)
6. Louis, A.K.: Inverse und schlecht gestellte Probleme. Stuttgart: Teubner 1989
7. Kirsch, A.: An Introduction to the Mathematical Theory of Inverse Problems. New York: Springer 1996
8. Hofmann, B.: Mathematik inverser Probleme. Leibzig: Teubner 1999
9. Gel'fand, I.M., Levitan, B.M.: Trans. Amer. Soc. **1**, 253–302 (1951)
10. Kac, M.: Am. Math. Mon. **73**, 1–23 (1966)
11. Marchenko, V.A.: Sturm–Liouville Operators and Applications. Basel: Birkhauser 1986
12. Chadan, K., Colton, D., Päivärinta, L., Rundell, W.: An Introduction to Inverse Scattering and Inverse Spectral Problems. Philadelphia: SIAM, 1997
13. Zakhariev, B.N., Chabanov, V.M.: Inverse Problems. **13**, R47–R79 (1997)
14. Newton, R.G.: Inverse Schrödinger Scattering in Three Dimensions. New York: Springer 1989
15. Chadan, K., Sabatier, P.C.: Inverse Problems in Quantum Scattering Theory. Berlin: Springer 1989
16. Wahba, G.: Spline Models for Observational Data. Philadelphia: SIAM 1990
17. Vapnik, V.N.: The Nature of Statistical Learning Theory. New York: Springer 1995
18. Vapnik, V.N.: Statistical Learning Theory. New York: Wiley 1998
19. Hastie,T.J., Tibshirani, R.J.: Generalized Additive Models. London: Chapman & Hall 1990
20. Huber, P-J.: Ann. Statist. **13**(2), 435–475 (1985)

21. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees, New York: Chapman & Hall 1993

22. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford: Oxford University Press 1995

23. Lauritzen, S.L.: Graphical Models. Oxford: Clarendon Press, 1996

24. Michie, D., Spiegelhalter, D.J., Taylor, C.C. (eds.): Machine Learning, Neural and Statistical Classification. New York: Ellis Horwood 1994

25. Box, G.E.P., Tiao, G.C.: Bayesian Inference in Statistical Analysis. New York: Wiley 1992 (Originally published in 1973 by Addison–Wesley, Reading, MA)

26. Berger, J.O.: Statistical Decision Theory and Bayesian Analysis. New York: Springer–Verlag 1980

27. Loredo T.: From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics. In Fougère, P.F. (ed.) Maximum-Entropy and Bayesian Methods, Dartmouth, 1989, 81–142. Dordrecht: Kluwer 1990. Available at `http://bayes.wustl.edu/gregory/gregory.html`.

28. Bernado, J.M., Smith, A.F.: Bayesian Theory. New York: John Wiley 1994

29. Robert, C.P.: The Bayesian Choice. New York: Springer 1994

30. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis. New York: Chapman & Hall 1995

31. Sivia, D.S.: Data Analysis: A Bayesian Tutorial. Oxford: Oxford University Press 1996

32. Lemm, J.C.: Prior Information and Generalized Questions. A.I.Memo No. 1598, C.B.C.L. Paper No. 141, Massachusetts Institute of Technology 1996. Available at `http://pauli.uni-muenster.de/~lemm`.

33. Lemm, J.C.: How to Implement A Priori Information: A Statistical Mechanics Approach. Technical Report, MS-TP1-98-12, Münster University 1998, `arXiv:cond-mat/9808039`.

34. Lemm, J.C.: Bayesian Field Theory. Nonparametric approaches to density estimation, regression, classification, and inverse quantum problems. Technical Report, MS-TP1-99-1, Münster University 1999, `arXiv: physics/9911077`.

35. Jeffrey, R.: Probabilistic Thinking. 1999. Available at `http://www.princeton.edu/~bayesway/`.

36. Jaynes, E.T.: Probability Theory: The Logic Of Science. (In preparation) Available at `http://bayes.wustl.edu/etj/prob.html`.

37. Doob, J.L: Stochastic Processes. New York: Wiley 1953 (New edition 1990)

38. Lemm, J.C., Uhlig, J., Weiguny, A.: Phys. Rev. Lett. **84**, 2068 (2000)

39. Schulman, L.S.: Techniques and Applications of Path Integration. New York: Wiley 1981

40. Glimm, J., Jaffe, A.: Quantum Physics. A Functional Integral Point of View. (2nd ed.) New York: Springer 1987

41. Hammersley, J.M., Handscomb, D.C.: Monte Carlo Methods. London: Chapman & Hall 1964

42. Binder, K. (ed.): The Monte Carlo Method in Condensed Matter Physics. Berlin: Springer 1992

43. Winkler, G.: Image Analysis, Random Fields and Dynamic Monte Carlo Methods. Berlin: Springer Verlag 1995

44. Neal, R.M.: Technical Report No. 9702, Dept. of Statistics, Univ. of Toronto, Canada 1997

45. de Bruijn, N.G.: Asymptotic Methods in Analysis. Amsterdam: North–Holland, 1961.

46. Bleistein, N. , Handelsman, N.: Asymptotic Expansions of Integrals. New York: Dover 1986 (Originally published in 1975 by Holt, Rinehart and Winston, New York)

47. Honerkamp, J: Statistical Physics. Berlin: Springer–Verlag 1998

48. Williams, C.K.I., Rasmussen, C.E.: Gaussian Processes for Regression. In Advances in Neural Information Processing Systems **8**, D.S. Touretzky et al (eds.), 515–520, Cambridge, MA: MIT Press 1996

49. MacKay, D.J.C.: Introduction to Gaussian processes. In Bishop, C., (ed.) Neural Networks and Machine Learning. NATO Asi Series. Series F, Computer and Systems Sciences, Vol. 168, 1998

50. Whittaker, E.T.: Proc. Edinborough Math. Assoc. **78**, 81-89 (1923)

51. Shiller, R.: Econometrica **41**, 775–778 (1973)

52. Akaike, H.: In Bayesian Statistics. J.M. Bernanda, M.H. De Groot, D.V. Lindley, A.F.M. Smith (eds.), 143–166, Valencia: Valencia University Press 1980

53. Green, P.J., Silverman, B.W.: Nonparametric Regression and Generalized Linear Models. A roughness penalty approach. London: Chapman & Hall 1994

54. Girosi, F., Jones, M., Poggio, T.: Neural Computation **7** (2), 219–269 (1995)

55. Kitagawa, G., Gersch, W.: Smoothness Priors Analysis of Time Series. New York: Springer 1996

56. Honerkamp, J., Weese J.: Cont. Mech. Thermodyn. **2**, 17–30 (1990)

57. Messiah, A.: Quantum Mechanics. Amsterdam: North–Holland, 1961

58. Balian, R.: From Microphysics to Macrophysics. Vol. I. Berlin: Springer 1991

59. Choquet–Bruhat, Y., DeWitt–Morette, C., Dillard–Bleick, M.: Analysis, Manifolds and Physics. (rev. ed.) Amsterdam: North–Holland 1982

60. Lifshits, M.A.: Gaussian Random Functions. Kluwer Academic Publ. 1995

61. Neal, R.M.: Bayesian Learning for Neural Networks. New York: Springer 1996

62. Williams, C.K.I., Barber, D.: IEEE Trans. on Pattern Analysis and Machine Intelligence. **20**(12), 1342–1351 (1998)

63. Lemm, J.C.: Mixtures of Gaussian Process Priors. In Proceedings of the Ninth International Conference on Artificial Neural Networks (ICANN99), IEEE Conference Publication No. 470. London: Institution of Electrical Engineers 1999

64. Hochstadt, H., Lieberman, B.: SIAM J. Appl. Math. **34**, 676–680 (1976)

65. Zhu, W., Rabitz, H.: J. Chem. Phys. **111**, 472–480 (1999)

66. Lemm, J.C.: Inverse Time–Dependent Quantum Mechanics. Technical Report, MS-TP1-00-1, Münster University 2000, `arXiv:quant-ph/0002010`.

67. Lemm, J.C.: Quadratic Concepts. In: Niklasson; L., Bodén, M., Ziemke, T. (eds.) Proceedings of the 8th International Conference on Artificial Neural Networks, Skövde, Sweden, 2–4 September 1998., 579–584, London: Springer 1998

68. Pierre, D.A.: Optimization Theory with Applications. New York: Dover 1986. (Original edition Wiley, 1969).

69. Fletcher, R.: Practical Methods of Optimization. New York: Wiley 1987

70. Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P.: Numerical Recipes in C. Cambridge: Cambridge University Press 1992

71. Bazaraa, M.S., Sherali, H.D., & Shetty, C.M.: Nonlinear Programming. (2nd ed.) New York: Wiley 1993

72. Bertsekas, D.P.: Nonlinear Programming. Belmont, MA: Athena Scientific 1995

73. Airapetyan, R.G., Puzynin, I.V.: Comp. Phys. Comm. **102**, 97–108 (1997)

74. Eisenberg, J.M., Greiner, W.: Microscopic Theory of the Nucleus. Amsterdam: North–Holland 1972

75. Ring, P., Schuck, P.: The Nuclear Many–Body Problem. New York: Springer Verlag 1980

76. Blaizot, J.–P., Ripka, G.: Quantum Theory of Finite Systems. Cambridge, MA: The MIT Press 1986

77. Negele, J.W., Orland, H.: Quantum Many–Particle Systems. Frontiers In Physics Series, Vol. 68, Redwood City, CA: Addison–Wesley 1988

78. Lemm, J. C.: Annals of Physics **244** (1), 136–200 (1995)

79. Lemm, J.C., Uhlig, J.: Phys. Rev. Lett. **84**, 4517 (2000)

80. Lemm, J. C., Giraud, B.G., Weiguny A.: Phys. Rev. Lett. **73**, 420–423 (1994)