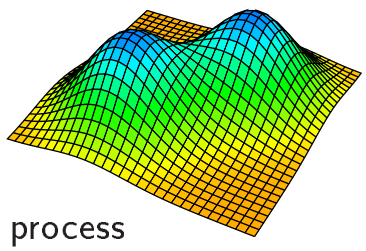
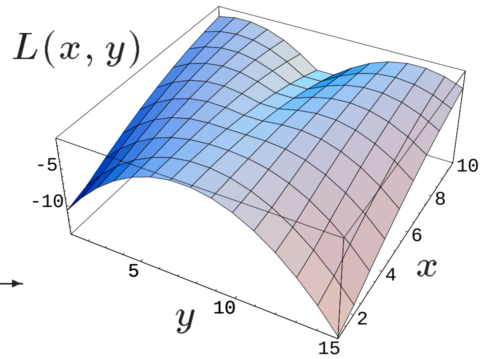
Bayesian density estimation

prior density $p(L)$

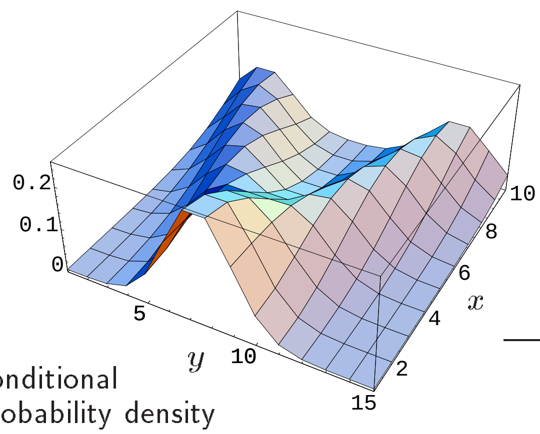


stochastic process
(e.g., mixture of Gaussians,
constraint on norm and positivity)



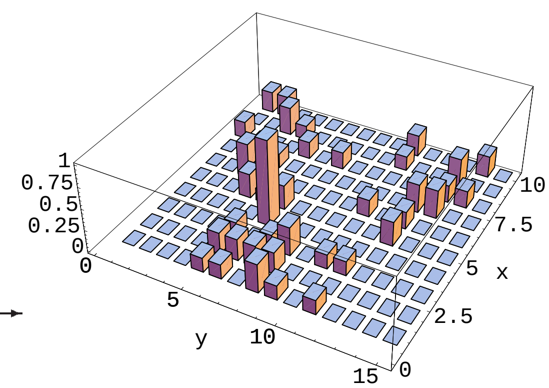
example
$L(x, y) = \ln p(y|x, L)$

likelihood model $p(y|x, L)$

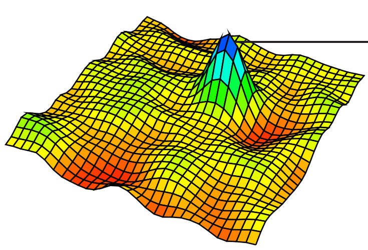


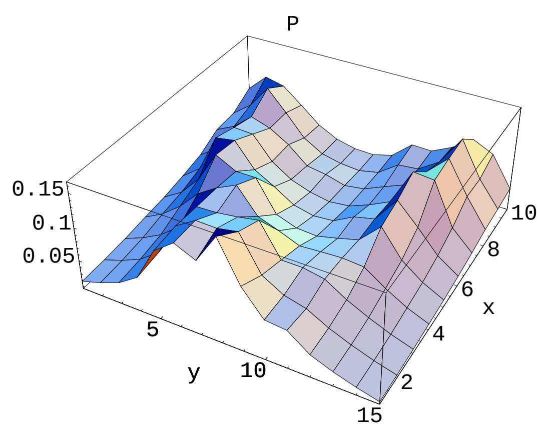
conditional
probability density

(training) data D



posterior density $p(L|D)$





maximum posterior solution
for $p(y|x, L)$

Nonparametric density estimation

The stationarity equations for density estimation are nonlinear from the beginning due to a non–quadratic data term $\sum_i \ln P_i$ and the normalization and non–negativity condition.

$$\boxed{-\beta E_\phi = \langle \ln P[\phi] | N \rangle - \frac{1}{2} \left\langle \phi \left| \mathbf{K} \right| \phi \right\rangle - \langle P[\phi] | \Lambda_X \rangle \,,}$$

$P[\phi]$ denotes a (possibly nonlinear) mapping from $\phi(x,y)$ to $P(x,y) = p(y|x,\phi)$

Λ_X is a Lagrange multiplier function

$N(x,y) = \sum_i \delta(x-x_i)\delta(y-y_i)$ represents the data

$$\boxed{\begin{aligned} 0 &= \mathbf{P}'[\phi] \left(\mathbf{P}^{-1}[\phi] N - \Lambda_X \right) - \mathbf{K}\phi \\ \Lambda_X &= \mathbf{I}_X \left(N - \mathbf{P}\mathbf{P}'^{-1} \mathbf{K}\,\phi \right) \end{aligned}}$$

$\mathbf{P}(x,y;x',y') = \delta(x-x')\delta(y-y') P[\phi](x,y)$

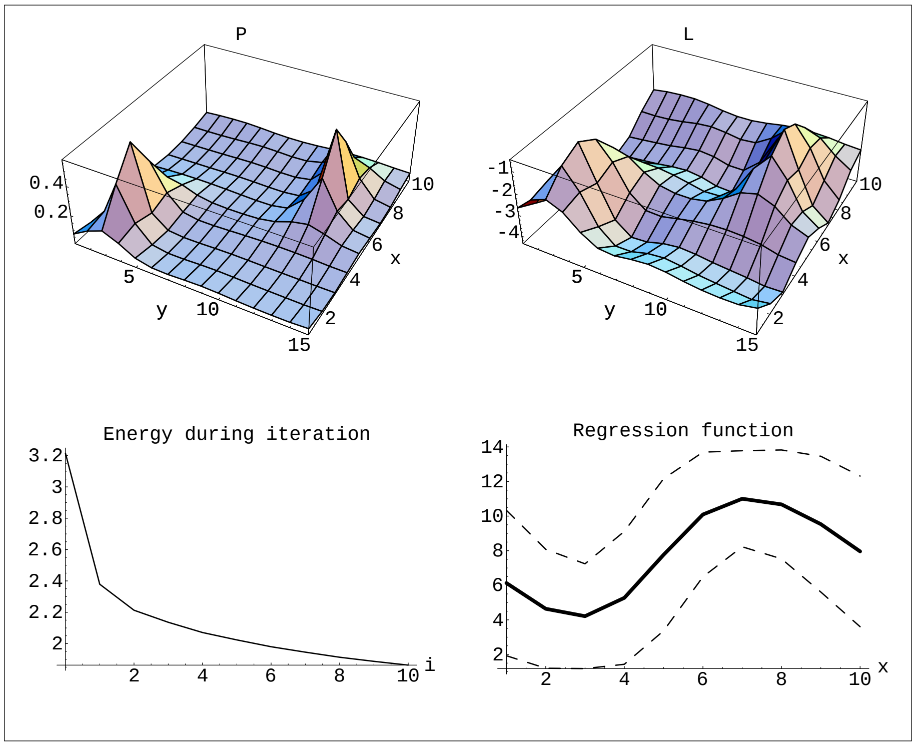
$\mathbf{P}'[\phi](x,y;x',y') = \delta P[\phi](x',y') / \delta\phi(x,y)$

$\mathbf{I}_X(x,y;x',y') = \delta(x-x')$

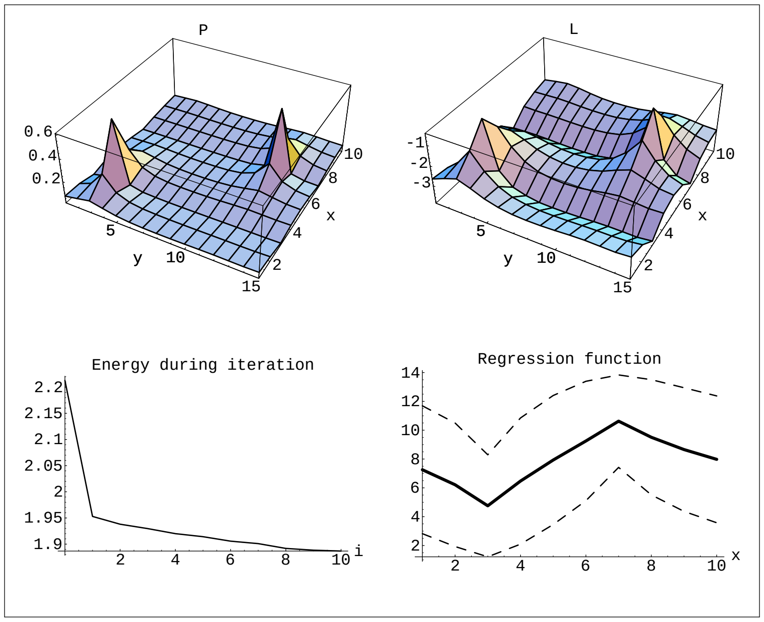
Solving by iteration according to $\phi^{(r+1)} = \phi^{(r)} + \eta \mathbf{A}^{-1} \left[\left(\mathbf{I} - \mathbf{P}^{(r)} \mathbf{I}_X \right) \left(N - \mathbf{P}^{(r)} \left(\mathbf{P}'^{(r)} \right)^{-1} \mathbf{K}\,\phi^{(r)} \right) \right]$

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

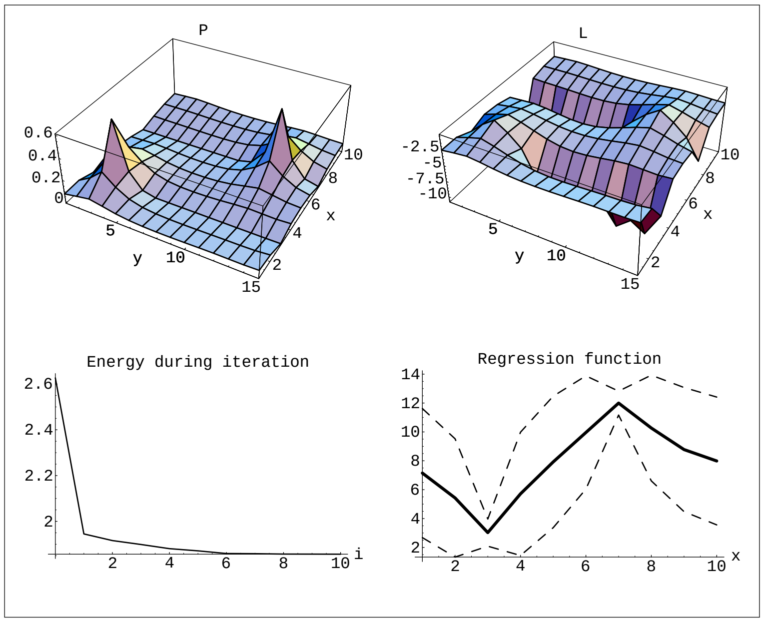
Constraints in density estimation for some specific choices of the field ϕ.



Density estimation with two data points and Gaussian prior $-\Delta^3$ on $L = \ln P$



Density estimation with two data points and Laplacian smoothness prior $-\Delta$ on $L = \ln P$



Density estimation with two data points and Laplacian smoothness prior $-\Delta$ on P

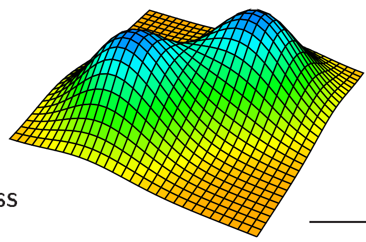
Some possible parameterizations of the likelihood

Ansatz	Functional form	to be optimized
linear ansatz	$\phi(z) = \sum_l \xi_l B_l(z)$	ξ_l
linear model with interaction	$\phi(z) = \xi_0 + \sum_l \xi_l z_l + \sum_{mn} \xi_{mn} z_m z_n + \cdots$	ξ_0, ξ_l $\xi_{mn}, \cdots$
mixture model	$\phi(z) = \sum \xi_{0,l} B_l(\xi_l, z)$	$\xi_{0,l}$, ξ_l
additive model with interaction	$\phi(z) = \sum_l \phi_l(z_l) + \sum_{mn} \phi_{mn}(z_m z_n) + \cdots$	$\phi_l(z_l)$ $\phi_{mn}(z_m z_n), \cdots$
product ansatz	$\phi(z) = \prod_l \phi_l(z_l)$	$\phi_l(z_l)$
decision trees	$\phi(z) = \sum_l \xi_l \prod_k \Theta(z_{\xi_{lk}} - \xi_{0,lk})$	ξ_l, $\xi_{0,lk}$, ξ_{lk}
projection pursuit	$\phi(z) = \xi_0 + \sum_l \phi_l(\xi_{0,l} + \sum_l \xi_l z_l)$	ϕ_l, ξ_0, $\xi_{0,l}$, ξ_l
neural net (2 lay.)	$\phi(z) = \sigma(\sum_l \xi_l\, \sigma(\sum_k \xi_{lk} z_k))$	ξ_l, ξ_{lk}

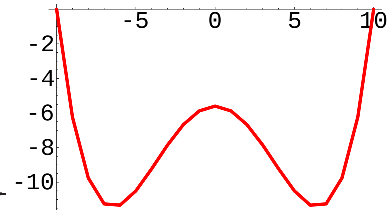
J.L., Uhlig, Weiguny, PRL 84 (2000)

Inverse Quantum Statistics à la Bayes

prior density $p(v)$



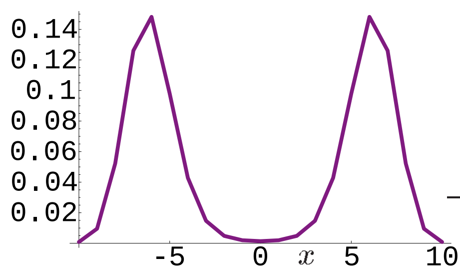
stochastic process
(e.g., mixture of
Gaussians)



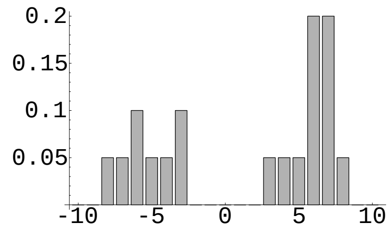
example of a
local potential $v(x)$

likelihood model $p(x|\hat{x}, v)$

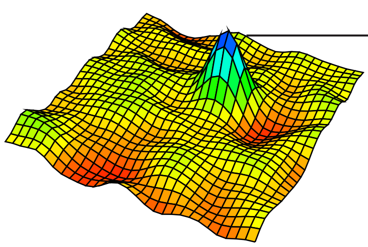
a possible
realization of
(training) data D

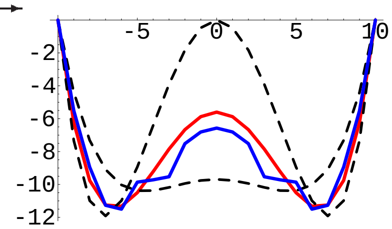


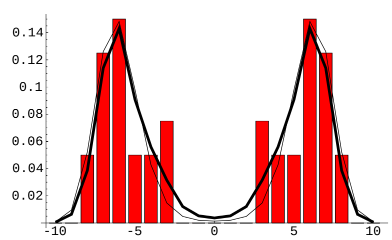
likelihood
for $\beta = 1$



posterior density $p(v|D)$

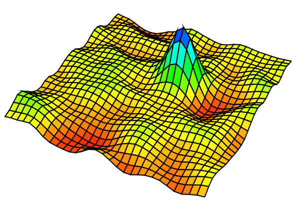




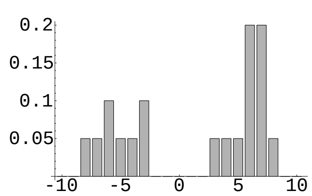


maximum posterior solution

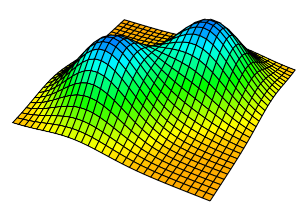
with



$\propto$



$\times$



The likelihood model of quantum statistics

Quantum system with Hamiltonian

$$H = T + V$$

e.g., with local potential

$$V(x, x') = \delta(x - x') \underbrace{v(x)}_{\text{to find!}}$$

The probability to find value x when measuring observable O at a quantum system in a state described by density operator ρ is given by

$$p(x|O, v) = \mathrm{Tr}\left[\pi(x, O)\ \rho(v)\right]$$

where $\pi(x, O) = \sum_\xi |x, \xi\rangle\langle x, \xi|$ denotes the projector in the space of eigenfunctions $|x, \xi\rangle$ of O with eigenvalues x. (The label ξ distinguishes degenerate eigenfunctions.)

For a canonical ensemble the density operator is given by

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

For position measurements $O = \hat{x}$

$$p(x|\hat{x}, v) = \sum_\alpha p_\alpha |\phi_\alpha(x)|^2 = \left\langle |\phi(x)|^2 \right\rangle$$

with $H\,|\phi_\alpha\rangle = E_\alpha\,|\phi_\alpha\rangle$, $\quad p_\alpha = e^{-\beta E_\alpha}/Z$, $Z = \sum_\alpha e^{-\beta E_\alpha}$

J.L., Uhlig, Weiguny, PRL 84 (2000)

Functional derivative of likelihood

Functional derivative of likelihood with respect to $v(x)$

$$\delta_{v(x)} p(x_i|\hat{x},v) = \frac{\delta}{\delta v(x)} \left(\frac{\sum_\alpha e^{-\beta E_\alpha} |\phi_\alpha(x_i)|^2}{\sum_{\alpha'} e^{-\beta E_{\alpha'}}} \right)$$

$$= \langle \phi^*(x_i) \delta_{v(x)} \phi(x_i) \rangle + \langle \delta_{v(x)} \phi^*(x_i) \phi(x_i) \rangle$$
$$-\beta \left[\left\langle \delta_{v(x)} E \, |\phi(x_i)|^2 \right\rangle - \langle \delta_{v(x)} E \rangle \left\langle |\phi(x_i)|^2 \right\rangle \right]$$

From the Schrödinger equation $H\,|\phi_\alpha\rangle = E_\alpha\,|\phi_\alpha\rangle$ follows

$$\left(\delta_{v(x)} H\right) |\phi_\alpha\rangle + H \delta_{v(x)} |\phi_\alpha\rangle$$

$$= \left(\delta_{v(x)} E_\alpha\right) |\phi_\alpha\rangle + E_\alpha \delta_{v(x)} |\phi_\alpha\rangle$$

and thus (for orthonormal, nondegenerate eigenfunctions)

$$\begin{aligned} \delta_{v(x)} E_\alpha &= \left\langle \phi_\alpha \left| \delta_{v(x)} H \right| \phi_\alpha \right\rangle = |\phi_\alpha(x)|^2 \\ \delta_{v(x)} \phi_\alpha(x') &= \sum_{\gamma \neq \alpha} \frac{1}{E_\alpha - E_\gamma} \phi_\gamma(x') \phi_\gamma^*(x) \phi_\alpha(x) \end{aligned}$$

with

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

Functional derivative of Gaussian processes

For a Gaussian process

$$p(v) = \left(\det \frac{\lambda \mathbf{K}_0}{2\pi}\right)^{\frac{1}{2}} e^{-\dfrac{\lambda}{2} \langle v - v_0 \,| \mathbf{K}_0 |\, v - v_0 \rangle}$$

the functional derivative with respect to the potential v is

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

The equation is also valid for mixtures of Gaussian processes

$$p(v) = \sum_{k=1}^{M} p(k)\, p(v|k)$$

where

$$\begin{aligned} \mathbf{K}_0 &= \sum_k p(k|v)\, \mathbf{K}_k \\ v_0 &= \mathbf{K}_0^{-1} \sum_k p(k|v)\, \mathbf{K}_k v_k \end{aligned}$$

with $\quad p(k|v) = p(v|k) p(k) / p(v)$.

Noisy energy measurements

Measurements of energy levels or the expectation of the energy at finite temperatures

$$U = \langle E \rangle = \sum_\alpha p_\alpha E_\alpha$$

can be useful to determine better the depth of a potential. For instance, a noisy measurement κ of U corresponds to

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

where $U \to \kappa$ for $\mu \to \infty$. With

$$\delta_{v(x)} U = \left\langle \delta_{v(x)} E \right\rangle - \beta \left\langle E \ \delta_{v(x)} E \right\rangle + \beta \langle E \rangle \left\langle \delta_{v(x)} E \right\rangle$$

one finds for the functional derivative of E_U

$$\delta_{v(x)} E_U = \mu \, (U - \kappa) \left\langle |\phi(x)|^2 \left[1 - \beta \, (E \ - U)\right] \right\rangle$$

Functional derivative of the posterior density

The stationarity equation for a posterior with Gaussian process prior and noisy energy measurement

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

$$= \delta_{v(x)} \left(\sum_i \ln p(x_i|\hat{x}, v) - \frac{\lambda}{2} \langle v - v_0 \, |\mathbf{K}_0| \, v - v_0 \rangle - E_U \right)$$

can be obtained by collecting:

1. the functional derivative of the likelihood

$$\begin{aligned} &\delta_{v(x)} p(x_i|\hat{x}, v) = \\ &\quad \left\langle \left[\delta_{v(x)} \phi^*(x_i) \right] \phi(x_i) \right\rangle + \left\langle \phi^*(x_i) \delta_{v(x)} \phi(x_i) \right\rangle \\ &-\beta \left[\left\langle |\phi(x_i)|^2 |\phi(x)|^2 \right\rangle - \left\langle |\phi(x_i)|^2 \right\rangle \left\langle |\phi(x)|^2 \right\rangle \right] \end{aligned}$$

2. the functional derivative of the prior

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

3. the functional derivative of the energy term

$$\delta_{v(x)} E_U = \mu \, (U - \kappa) \left\langle |\phi(x)|^2 \left[1 - \beta \, (E - U) \right] \right\rangle$$

Iteration methods = learning algorithms

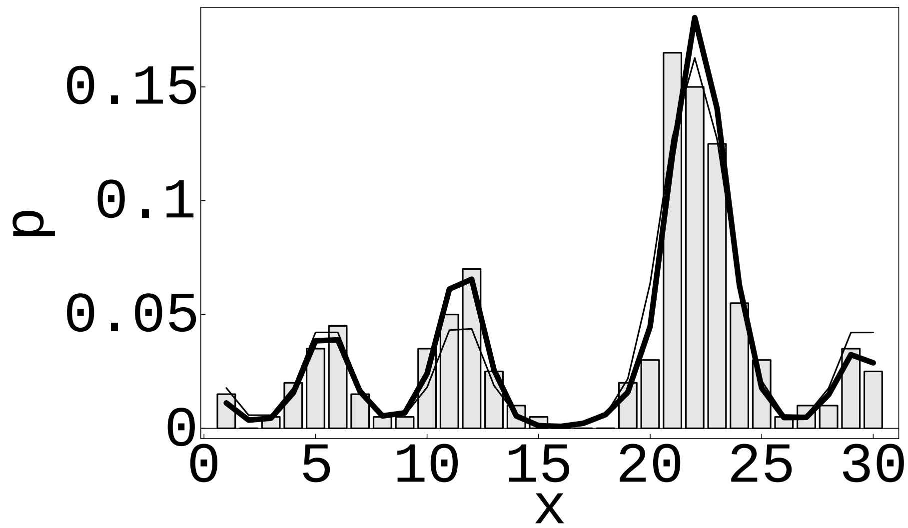
Choosing an initial guess $v^{(0)}$, step width η, and a positive definite matrix $\mathbf{A}$ the MAP stationarity equation can be solved by iteration

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

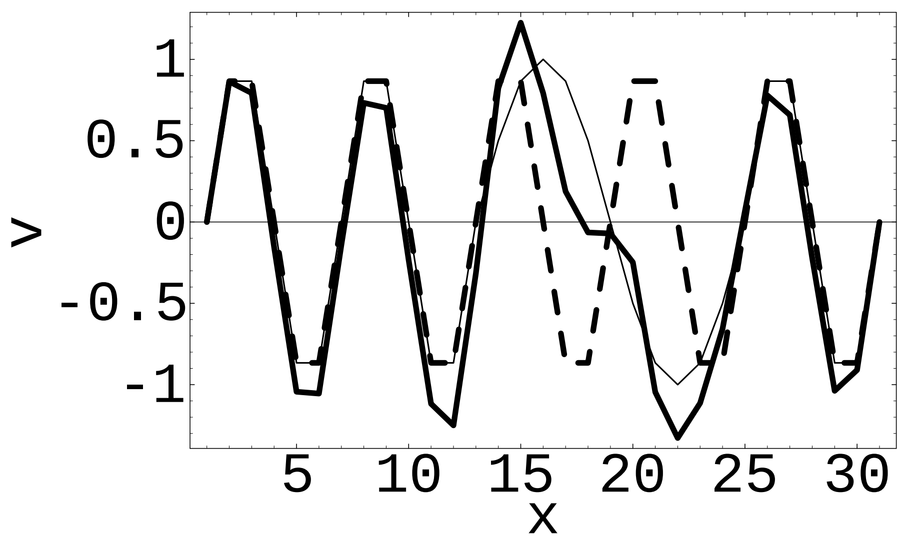
Typically, η and posssibly $\mathbf{A}$ will be adapted during iteration The simple choice $\mathbf{A} = \mathbf{K}_0$ often proved to be useful. Different initial guesses $v^{(0)}$ can lead to different results.

Reconstruction of approximately periodic potentials

(a) LIKELIHOODS



(b) POTENTIALS

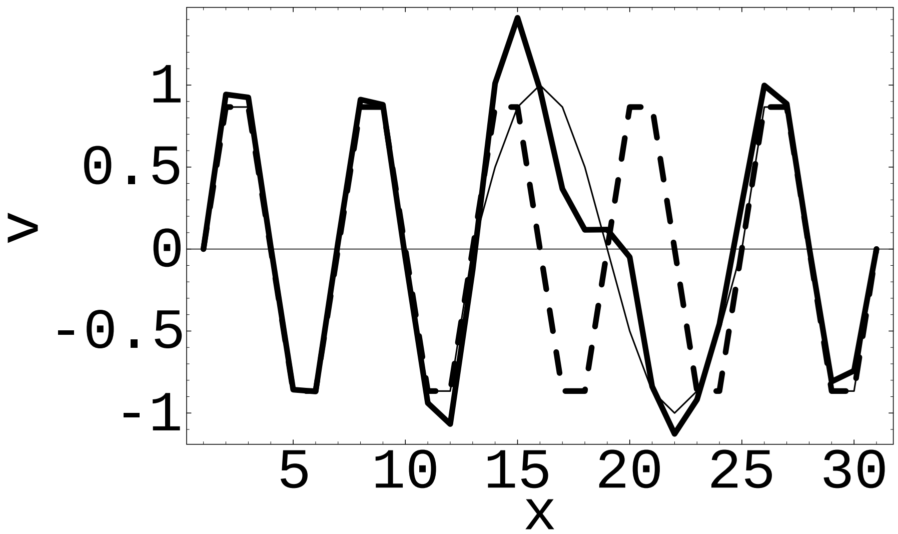


(a) relative frequencies of observations (bars), true (thin) and reconstructed (thick) likelihoods. (b) True potential v_{true} (thin), reference potential $v_0 = \sin(\pi x/3)$ (dashed), and reconstructed potential v (thick). (200 data points, mesh with 30 points, $m = 0.25$ with $\hbar = 1$, $\beta = 4$, $\mathbf{K}_0 = -\Delta$, $\lambda = 0.2$, $\mu = 0$, $U(v_{\mathrm{true}}) = -0.354$, $U(v) = -0.552$.)

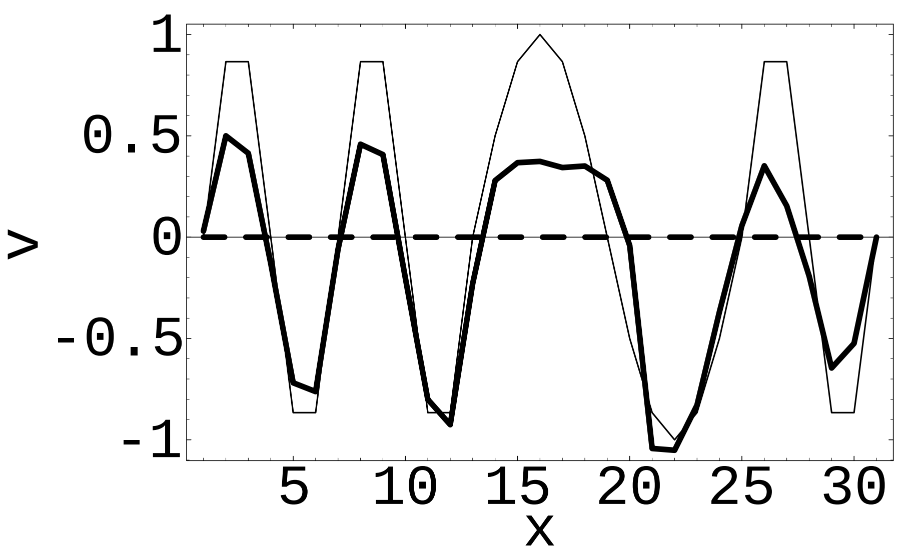
J.L., Uhlig, Weiguny, PRL 84 (2000)

Reconstruction of approximately periodic potentials

(a) WITH PERIODIC REFERENCE POTENTIAL

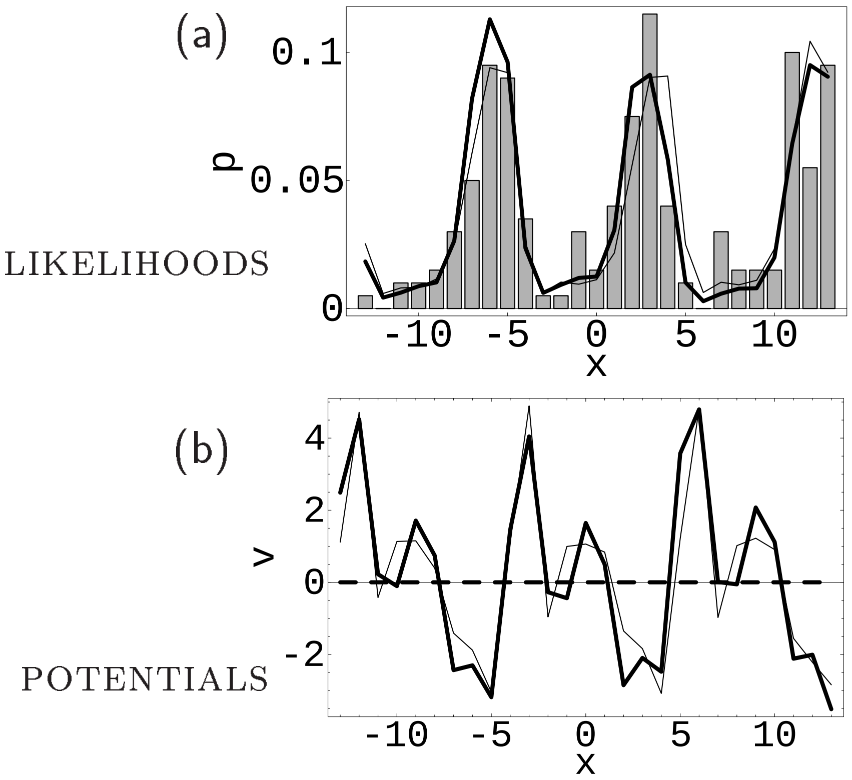


(b) WITH PERIODIC PRIOR COVARIANCE



Reconstruction including E_U with $\kappa = U(v_{\rm true}) = -0.354$. (a) $\mathbf{K}_0 = -\Delta$, $v_0(x) = \sin(\pi x/3)$, $\mu = 1000$, $U(v) = -0.353$. (b) $\mathbf{K}_0 = -\Delta - \gamma\Delta_\theta$, $v_0 \equiv 0$, $\theta = 6$, $\gamma = 0.12$, $\lambda = 0.05$, $\mu = 10$, $U(v) = -0.351$.

Reconstruction of approximately periodic potentials



Reconstruction of an approximately periodic potential with Gaussian prior from position measurements (without energy constraint). **(a)** shows empirical density (bars), true (thin line) and reconstructed likelihood (thick line). **(b)** shows the true potential (thin line), the zero reference potential $v_0 \equiv 0$ (dashed line), and a solution for the reconstructed potential v (thick line).

200 data points, mesh with 27 points, (200 data points, mesh with 27 points, mass $m =$ 0.1, inverse temperature $\beta = 1$, inverse prior covariance $\lambda \mathbf{K}_0 = (1 - \lambda')\mathbf{K}_\theta - \lambda'\Delta$, with $\lambda = 10$, $\lambda' = 0.001$, $\theta = 9$, and periodic boundary conditions for v as well as for the wave functions. The free additive constant of v has been chosen so that its mean coincides with the mean of the original function.

Prior energy variants

Gaussian prior

$$E(v) = \frac{1}{2}\left\langle v - v_0 \left| \mathbf{K}_0 \right| v - v_0 \right\rangle = \frac{1}{2}|\omega_0|^2$$

with inverse covariance $\mathbf{K}_i = \mathbf{W}_i^T \mathbf{W}_i$
and $\omega_i(x) = \int dx' \, \mathbf{W}_i(x,x')[v(x') - v_i(x')]$

Prior with switching or mixing hyperparameter θ

$$E(v|\theta) = \frac{1-\theta}{2}|\omega_1|^2 + \frac{\theta}{2}|\omega_2|^2$$

$$= \frac{1-\theta}{2}\left\langle v - v_1 \left| \mathbf{K}_1 \right| v - v_1 \right\rangle + \frac{\theta}{2}\left\langle v - v_2 \left| \mathbf{K}_2 \right| v - v_2 \right\rangle$$

Prior with local hyperfield $\theta(x)$

$$E(v|\theta) = \frac{1}{2}\int dx \, \left([1-\theta(x)]|\omega_1(x)|^2 \right.$$
$$\left. + \theta(x)|\omega_2(x)|^2\right) + \ln Z_{\mathcal{V}}(\theta)$$

or $E(v|\theta) = \frac{1}{2}\left\langle v - \tilde{v}_0(\theta) | v - \tilde{v}_0(\theta)\right\rangle + \frac{1}{2}\left\langle v \left| \mathbf{K} \right| v \right\rangle$

Non–Gaussian prior with auxiliary field $B(x;v)$

$$E(v) = \frac{1}{2}\int dx \, \left([1 - B(x)]|\omega_1(x)|^2 + B(x)|\omega_2(x)|^2\right)$$

with $B(x) = \sigma\left(u(x) - \vartheta\right)$ for sigmoidal function σ and
$u(x) = |\omega_1(x)|^2 - |\omega_2(x)|^2$

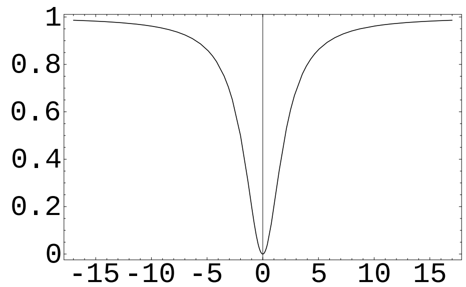
Typically hyperpriors for hyperfields or additional priors for auxiliary fields have to be added, e.g. smoothness constraints.

Nonquadratic potentials

Non–Gaussian prior

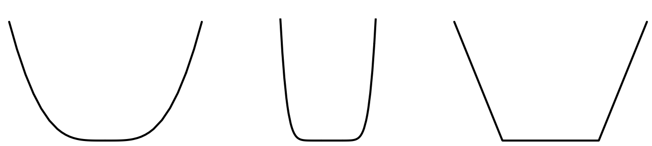
$$p(v) \propto e^{-\frac{1}{2}\int dx\, \psi[\omega(x)]},$$

where ψ is a non–quadratic function (potential) and $\omega(x) = \int dx'\, \mathbf{W}(x,x')[v(x') - v_0(x')]$ for $\mathbf{K} = \mathbf{W}^T\mathbf{W}$



Example of a non–quadratic "cup"–function $\psi(x) = a(1.0 - 1/(1 + (|x - x_0|/b)^\gamma))$ with a= 5, b = 10, γ = 0.7, x_0 = 0 for which one large step is cheaper than many small ones, allowing jumps in the reconstructed functions and being less sensitive to outliers

Similarly, "robust" error functions are insensitive to small errors



Left: Logarithm of mixture with two Gaussians with equal variance nd different means. Middle: Logarithm of mixture with 11 Gaussians with equal variance and different means. Right: ϵ–insensitive error.

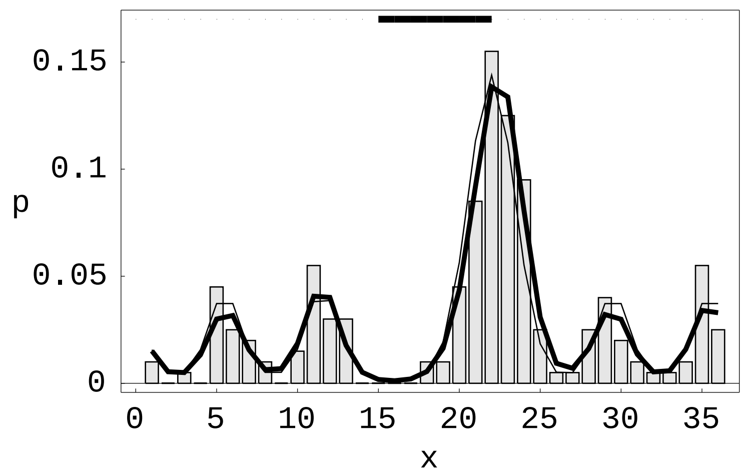
Reconstruction with local hyperfield

Local switching between different reference potentials $v_0(x;\theta) = [1-\theta(x)]v_1(x) + \theta(x)v_2(x)$ using a binary hyperfield $\theta(x)$

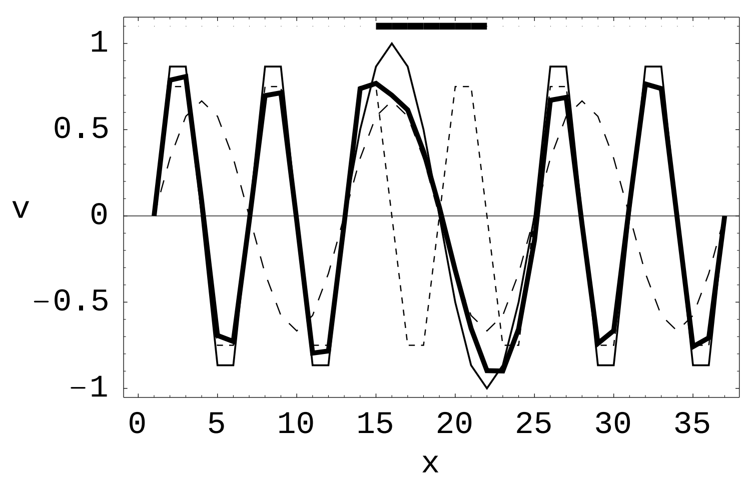
$$E_0(v,\theta) = \frac{\lambda_1}{2}\langle v - v_0(\theta)|v - v_0(\theta)\rangle - \frac{\lambda_2}{2}\langle v\,|\Delta|\,v\rangle - \ln p(\theta),$$

Additional hyperprior $p(\theta)$ to control number of discontinuities of $\theta(x)$. Hyperfield optimized by "simulated annealing".

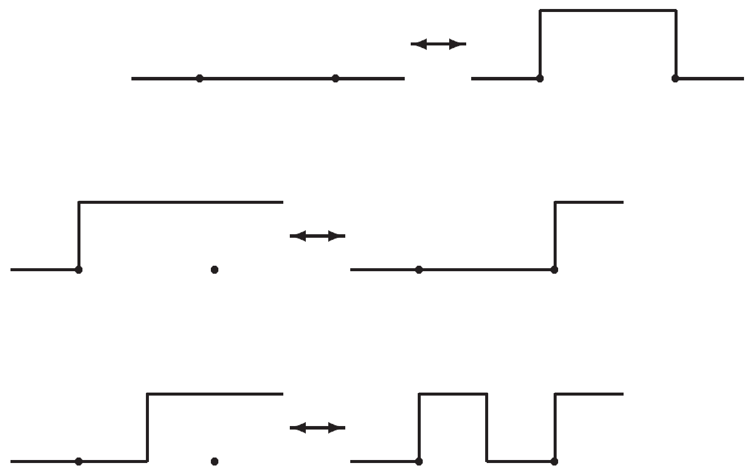
(a) LIKELIHOODS



(b) POTENTIALS



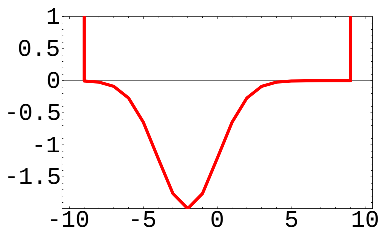
Some operations on switching fields

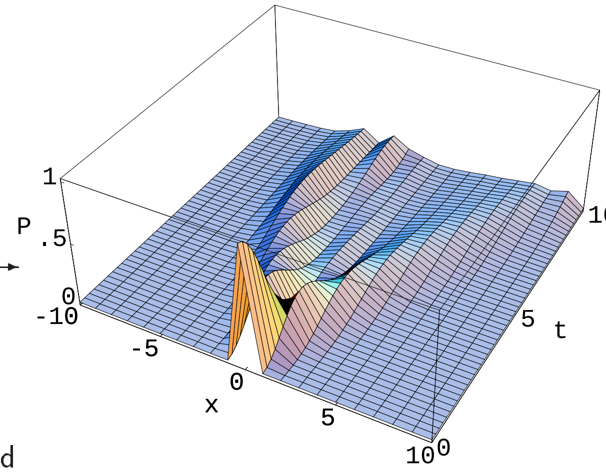


Generation of new trial configurations for simulated annealing by selecting two points randomly and exchanging the values zero and one of the binary function in between. Those mechanism can be used to optimize binary (switching) functions.

Non–stationary quantum systems

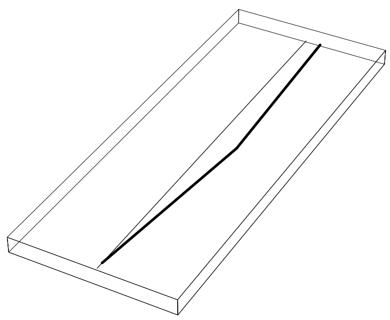
local potential $v(x)$

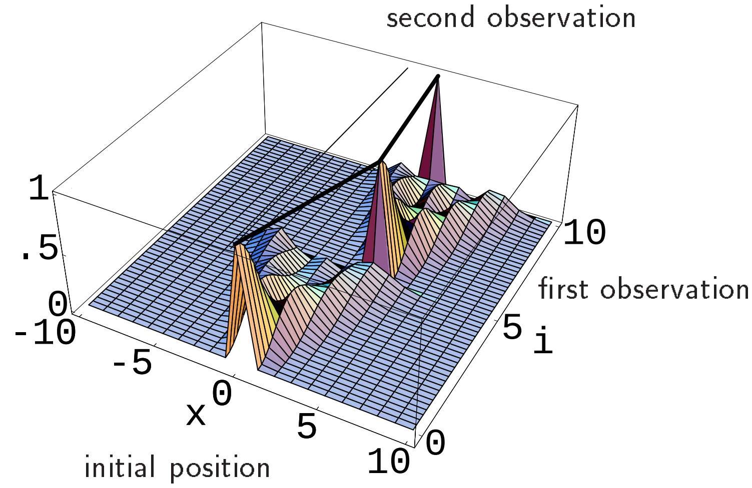




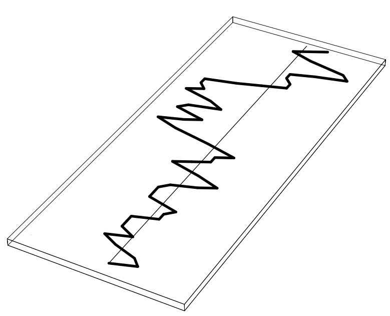
time evolution of likelihood
for a particle with initial position $x_0 = 0$

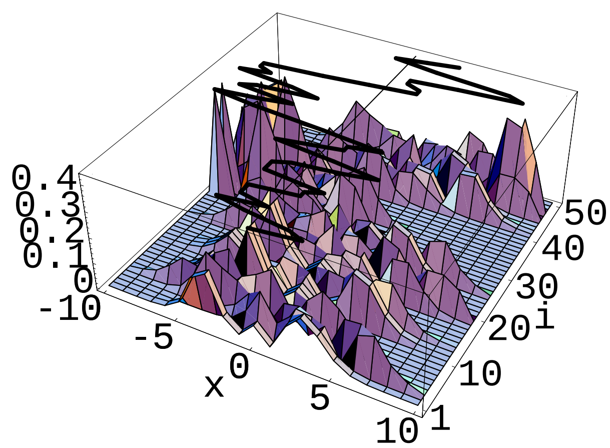
measurements of particle position
with time intervals $\Delta t = 5$.





observed time series





Likelihoods for time–dependent systems

Probability to observe position x_i for a particle which have been at position x_{i-1} at time $t_i - \Delta_i$

$$p_i = p(x_i|\Delta_i, x_{i-1}, v) = |\phi_i(x_i)|^2$$

where $\phi_i(x_i) = \langle x_i|\phi_i\rangle = \langle x_i|U_i\, x_{i-1}\rangle$ and v–dependent time–evolution operator:

$$U_i = \exp\left(-i\Delta_i H\right)$$

With $H\psi_\alpha = E_\alpha\psi_\alpha$ and

$$\begin{aligned} \delta_{v(x)} E_\alpha &= \left\langle \psi_\alpha \,\middle|\, \delta_{v(x)} H \,\middle|\, \psi_\alpha \right\rangle = |\psi_\alpha(x)|^2, \\ \delta_{v(x)} \psi_\alpha(x') &= \sum_{\gamma\neq\alpha} \frac{1}{E_\alpha - E_\gamma}\, \psi_\gamma(x')\psi_\gamma^*(x)\psi_\alpha(x) \end{aligned}$$

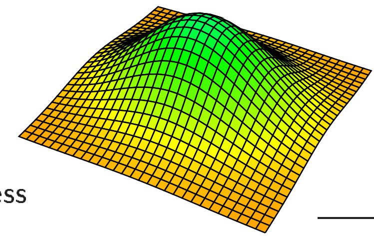
it follows

$$\begin{aligned} \delta_{v(x)}\phi_i(x_i) &= \delta_{v(x)}\, \langle x_i|U_i\, x_{i-1}\rangle = \\ & \sum_\alpha e^{-i\Delta_i E_\alpha}\Big[\left(-i\Delta_i|\psi_\alpha(x)|^2\right) \psi_\alpha(x_i)\psi_\alpha^*(x_{i-1}) \\ & + \sum_{\gamma\neq\alpha} \frac{1}{E_\alpha - E_\gamma}\psi_\gamma(x_i)\psi_\gamma^*(x)\psi_\alpha(x)\psi_\alpha^*(x_{i-1}) \\ & + \sum_{\gamma\neq\alpha} \frac{1}{E_\alpha - E_\gamma}\psi_\gamma^*(x_{i-1})\psi_\gamma(x)\psi_\alpha^*(x)\psi_\alpha(x_i)\Big] \end{aligned}$$

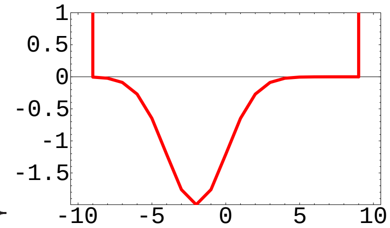
J.L., PLA 276 (2000)

Time–dependent inverse quantum theory

prior density $p(v)$

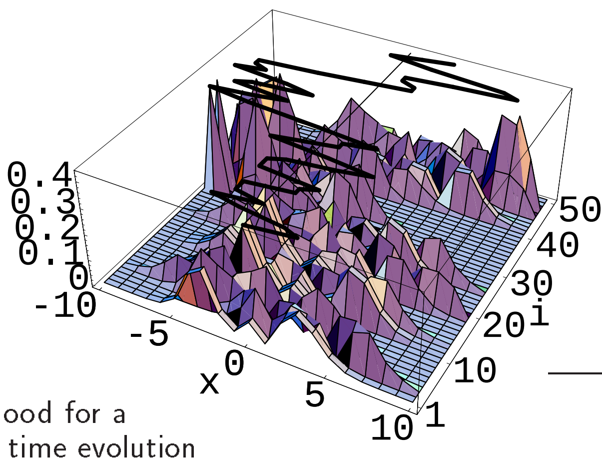


stochastic process



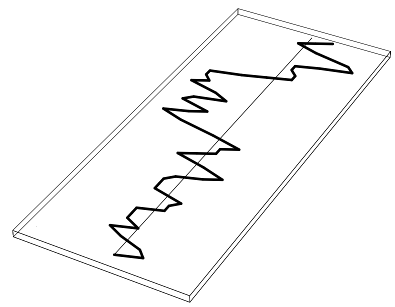
example of a
local potential $v(x)$

likelihood model $p(x|t_i, x_{i-1}, v)$

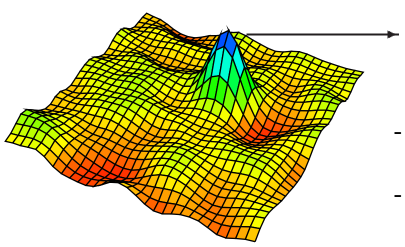


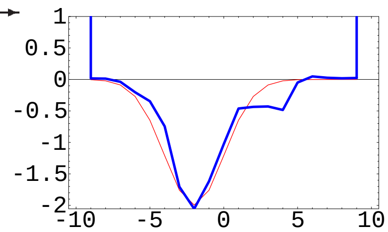
likelihood for a
given time evolution

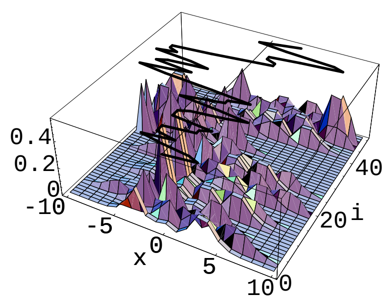
a possible
realization of
(training) data D



posterior density $p(v|D)$

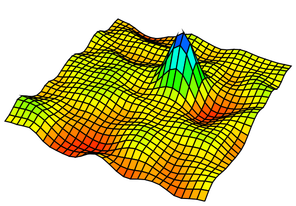
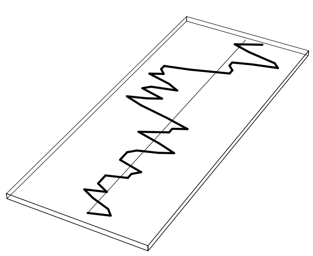
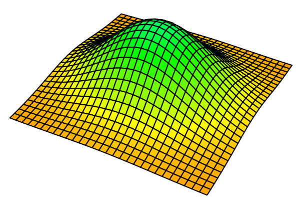






maximum posterior solution

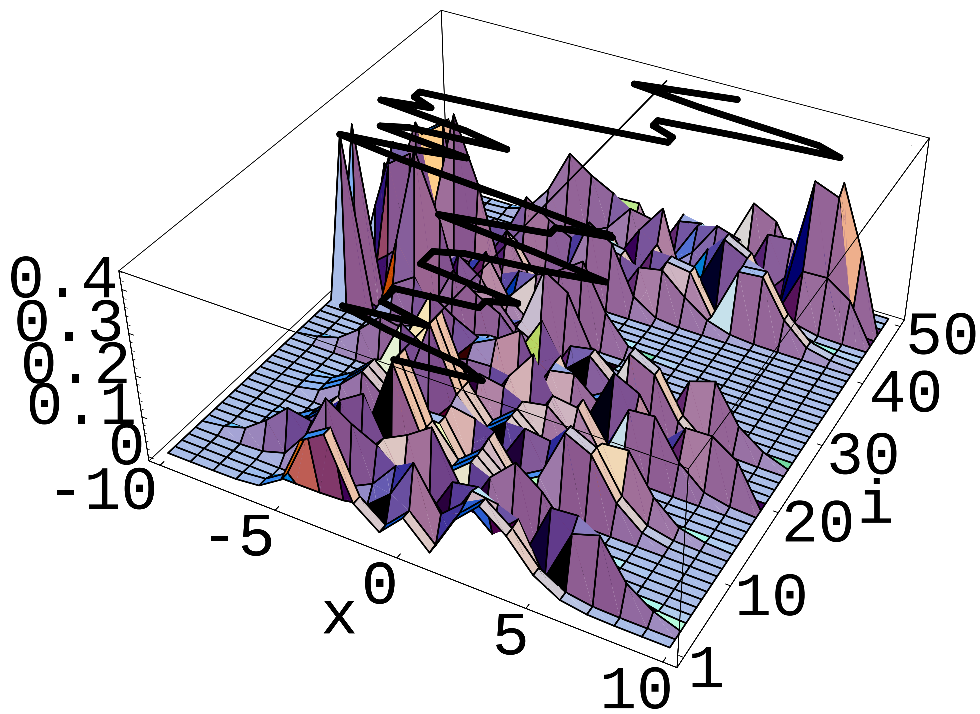
where

 $\propto$  $\times$ 

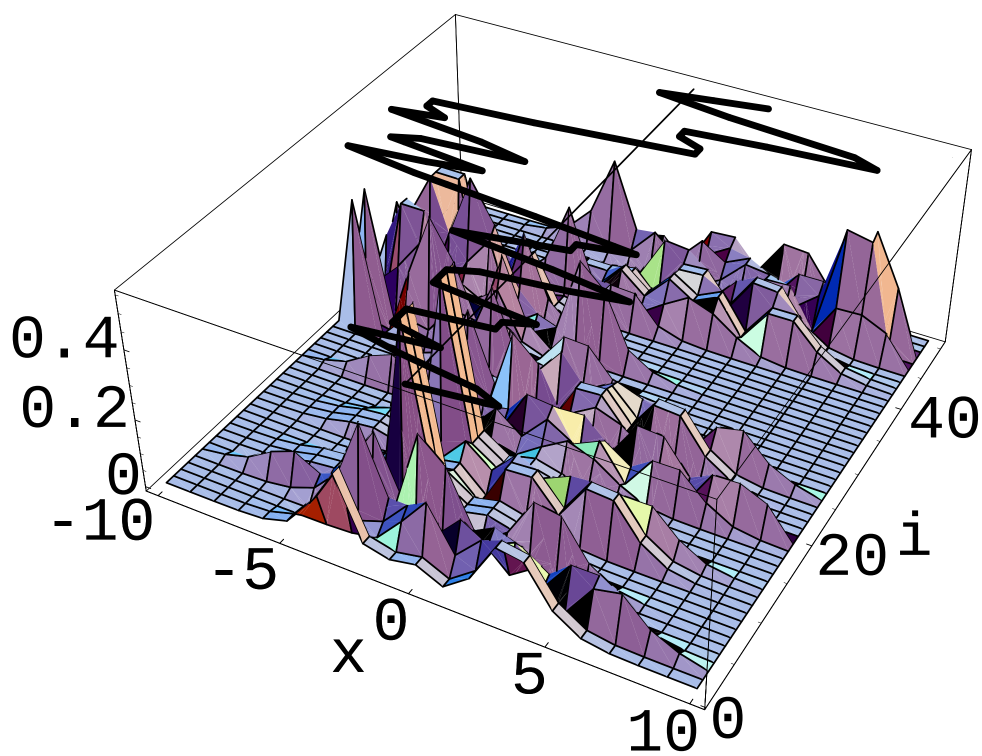
Time–dependent inverse quantum theory:

Time series data

original likelihoods



reconstructed likelihoods

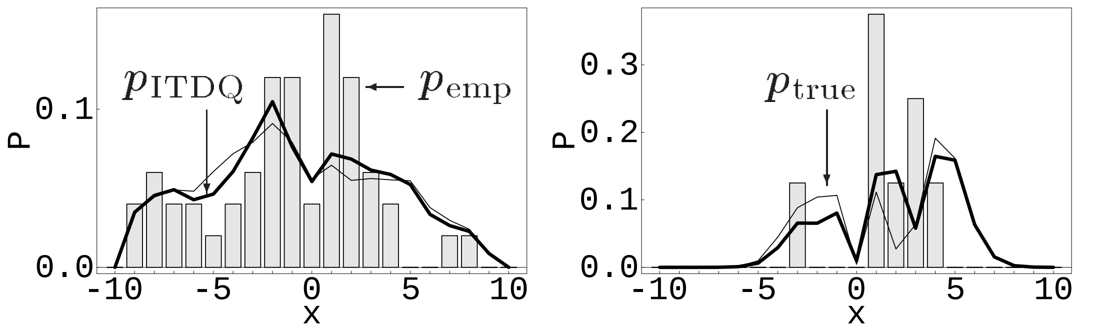


particle distance

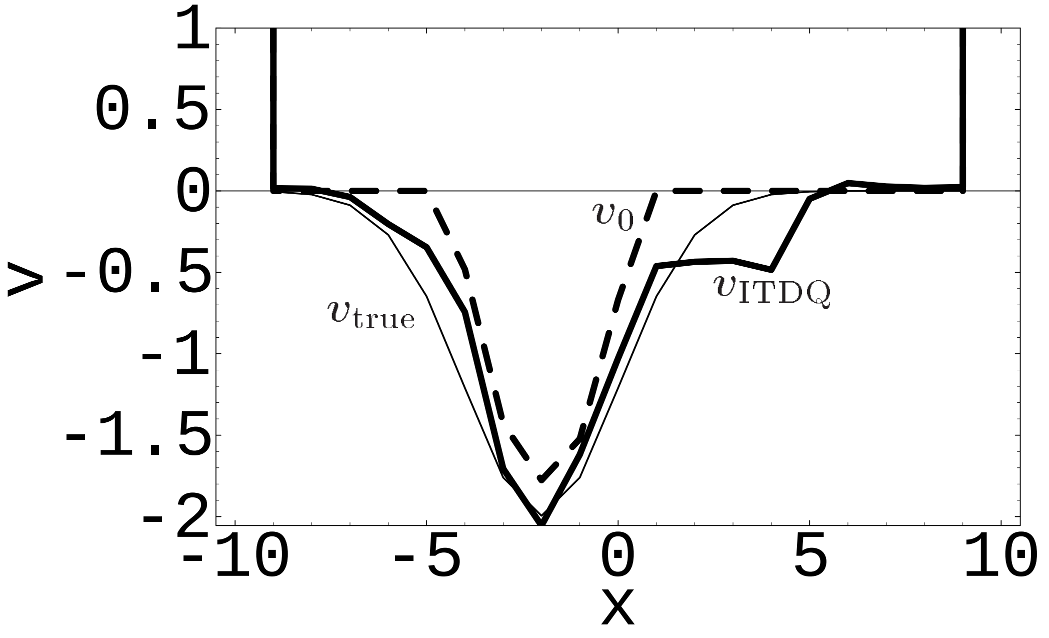
Time–dependent inverse quantum theory:

Time series data

transition probabilities



sum over all x_{i-1} for $x_{i-1} = 1$



Many–body likelihood

Fermionic many–body Hamiltonian

$$H = \sum_{\alpha\beta} T_{\alpha\beta}\, a^\dagger_\alpha a_\beta + \frac{1}{4}\sum_{\alpha\beta\gamma\delta} V_{\alpha\beta\gamma\delta}\, a^\dagger_\alpha a^\dagger_\beta a_\delta a_\gamma$$

with local two–body–potential in $z = |x_{(1)} - x_{(2)}|$

$$V_{x_{(1)}x_{(2)}x'_{(1)}x'_{(2)}} = v(z)[\delta_{x_{(1)},x'_{(1)}}\delta_{x_{(2)},x'_{(2)}} - \delta_{x_{(1)},x'_{(2)}}\delta_{x_{(2)},x'_{(1)}}]$$

Hartree–Fock approximation: $h\varphi_k = \epsilon_k\varphi_k$ with

$$h_{xx'} = T_{xx'} + \sum_{k=1}^{N} \left\langle x\varphi_k \middle| V \middle| x'\varphi_k \right\rangle$$

HF ground state: $|\Phi_0\rangle = |\varphi_1, \cdots, \varphi_N\rangle$

Hartree–Fock likelihoods:
(A) 2–particle data: $p_{\rm HF}(\vec{x}_i|\hat{x}, v) = |\Phi_0(\vec{x}_i)|^2$
(B) 1–particle data: $p_{\rm HF}(x_i|\hat{x}, v) = \sum_{k=1}^{N} |\varphi_k(x_i)|^2$

Inverse HF equation:

$$\underbrace{\left|\delta_{v(z)}\varphi_k\right\rangle}_{\substack{\text{two}--\text{body}\\ \text{equation}}} = \sum_{l\neq k} \frac{1}{\epsilon_k - \epsilon_l} |\varphi_l\rangle \left\langle \varphi_l \middle| \delta_{v(z)} h \middle| \varphi_k \right\rangle$$

$$\delta_{v(z)} h_{xx'} = \sum_{j=1}^{N} \Big[\left\langle x\,\varphi_j \middle| \delta_{v(z)} V \middle| x'\,\varphi_j \right\rangle + \left\langle x\,\delta_{v(z)}\varphi_j \middle| V \middle| x'\,\varphi_j \right\rangle + \left\langle x\,\varphi_j \middle| V \middle| x'\,\delta_{v(z)}\varphi_j \right\rangle \Big]$$

$\Rightarrow$ 3 nested iteration procedures for v, $\delta_v\varphi_k$ and h

The Bayesian inverse Hartree–Fock algorithm

The reconstruction of a potential from observational data by means of the inverse Hartree–Fock approaches requires

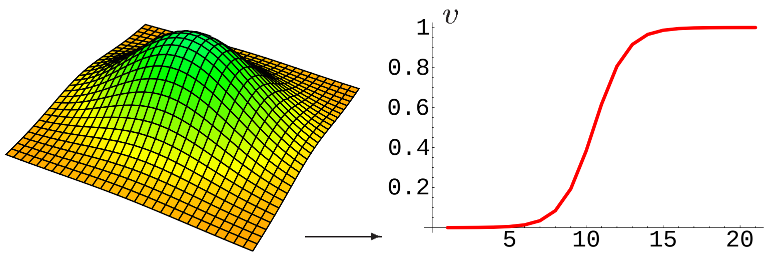
1. the iterative solution of the Maximum–A–Posteriori stationarity equation for the Potential v. Each iteration step requires
2. the functional derivatives of the likelihoods, e.g., $\delta_{v(x)} \ln p(\vec{x}_i|\hat{x}, v)$ or $\delta_{v(x)} \ln p(x_i|\hat{x}, v)$. Those are obtained by solving the (two–particle–like) inverse Hartree–Fock equation for given
3. one–particle–orbitals φ_i, defined as solution of the self–consitent (one–body–)Hartree–Fock equation.

J.L., Uhlig, PRL 84 (2000)

Hartree–Fock approximation for inverse problems

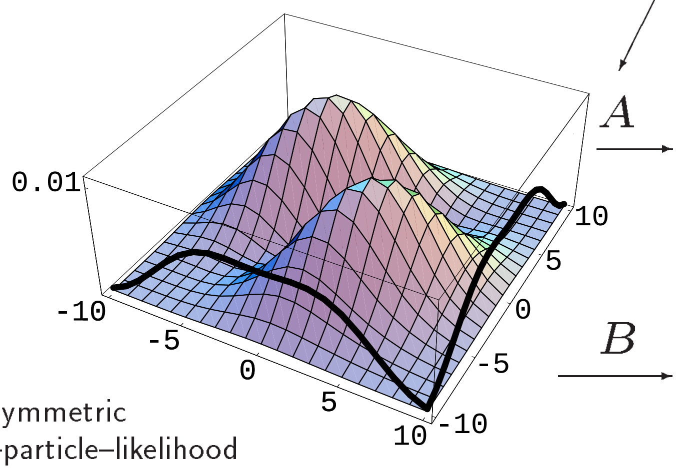
priori density $p(v)$

stochastic prozes

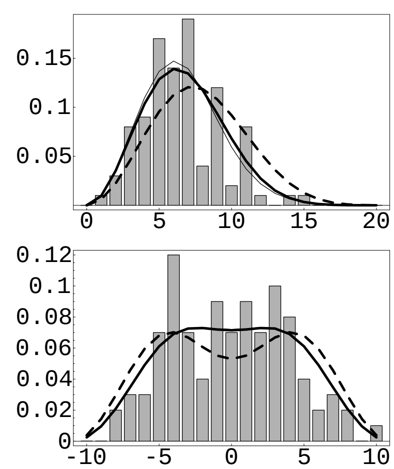


example of a
local potential $v(|x_{(1)} - x_{(2)}|)$

likelihood model $p(x_{(1)}, x_{(2)}|\hat{x}, v)$

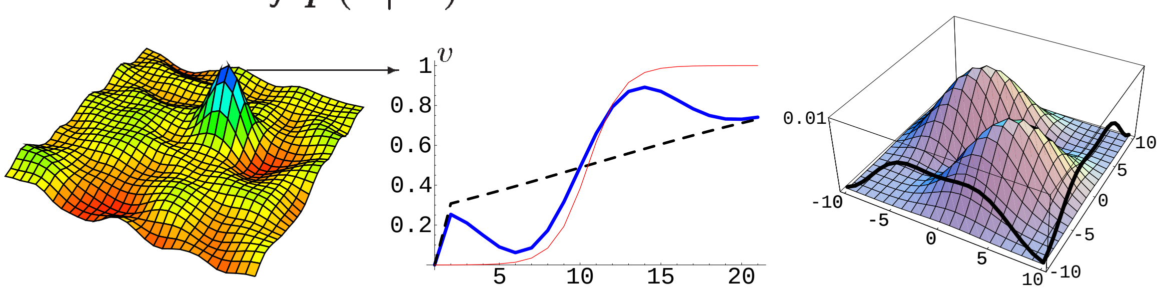


(training) data D



antisymmetric
two–particle–likelihood
and one–particle–likelihood/10

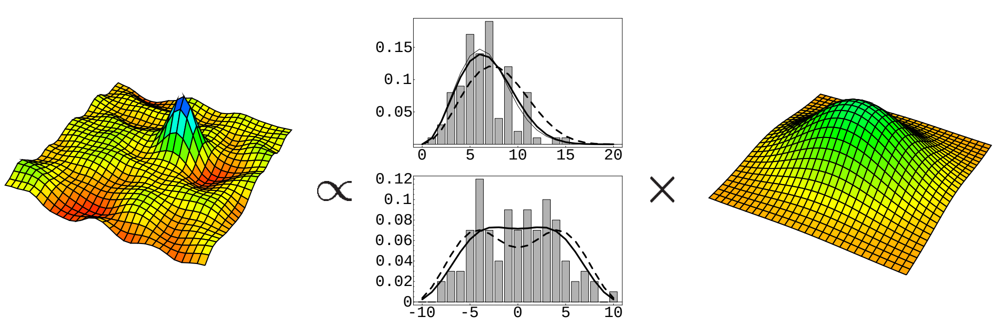
posterior density $p(v|D)$



maximum–posterior solution

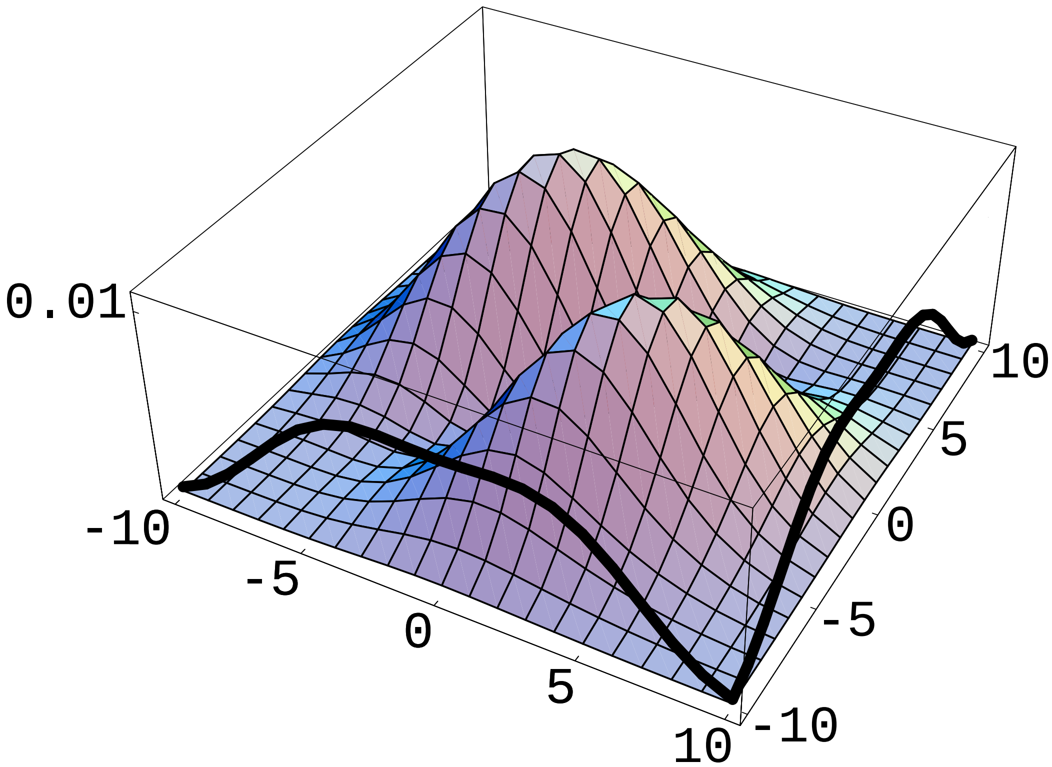
Hartree–Fock likelihood

where

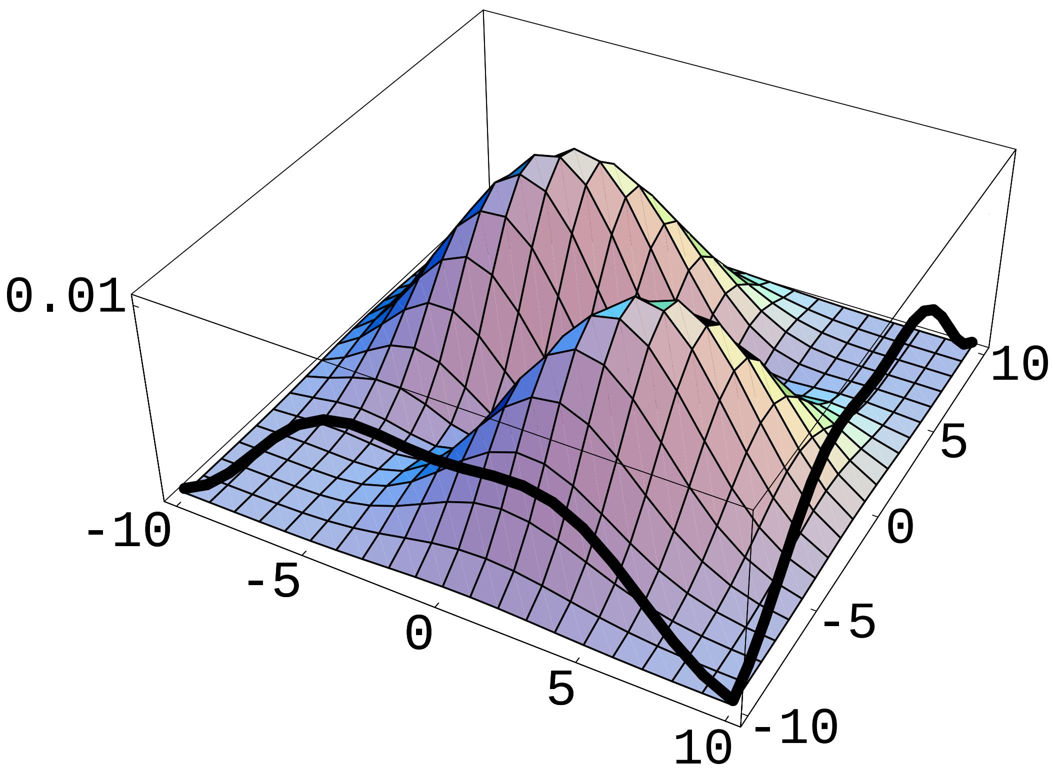


Bayesian inverse Hartree–Fock theory

original likelihoods

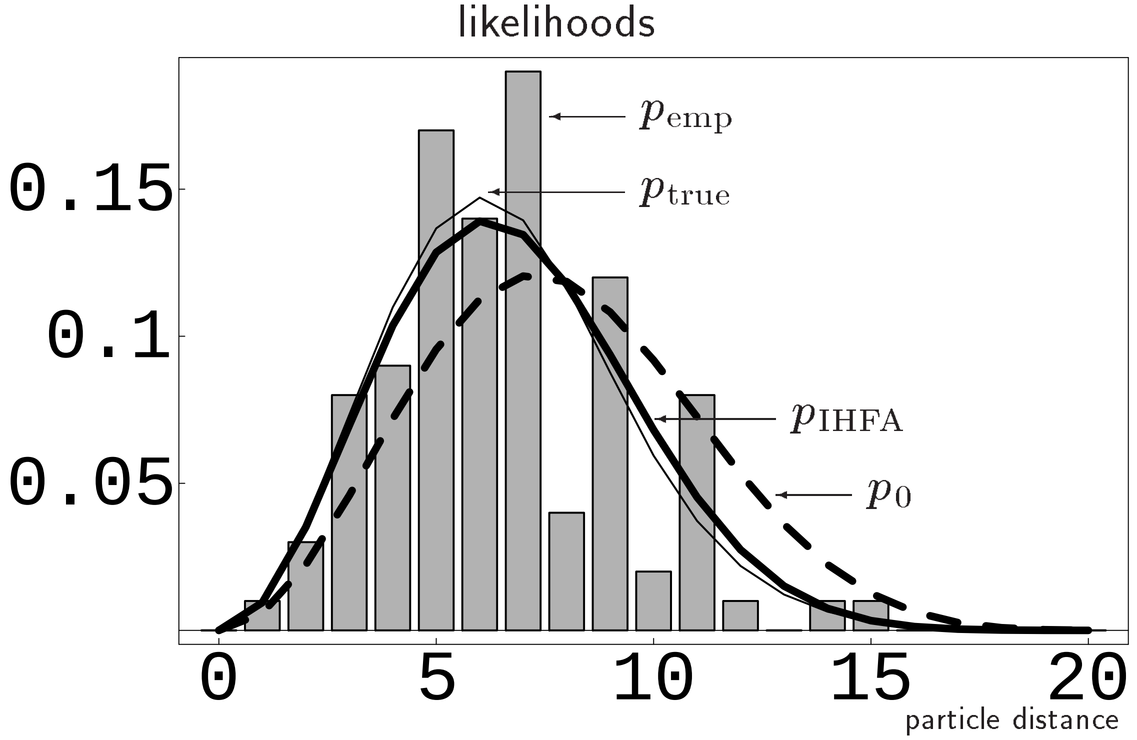


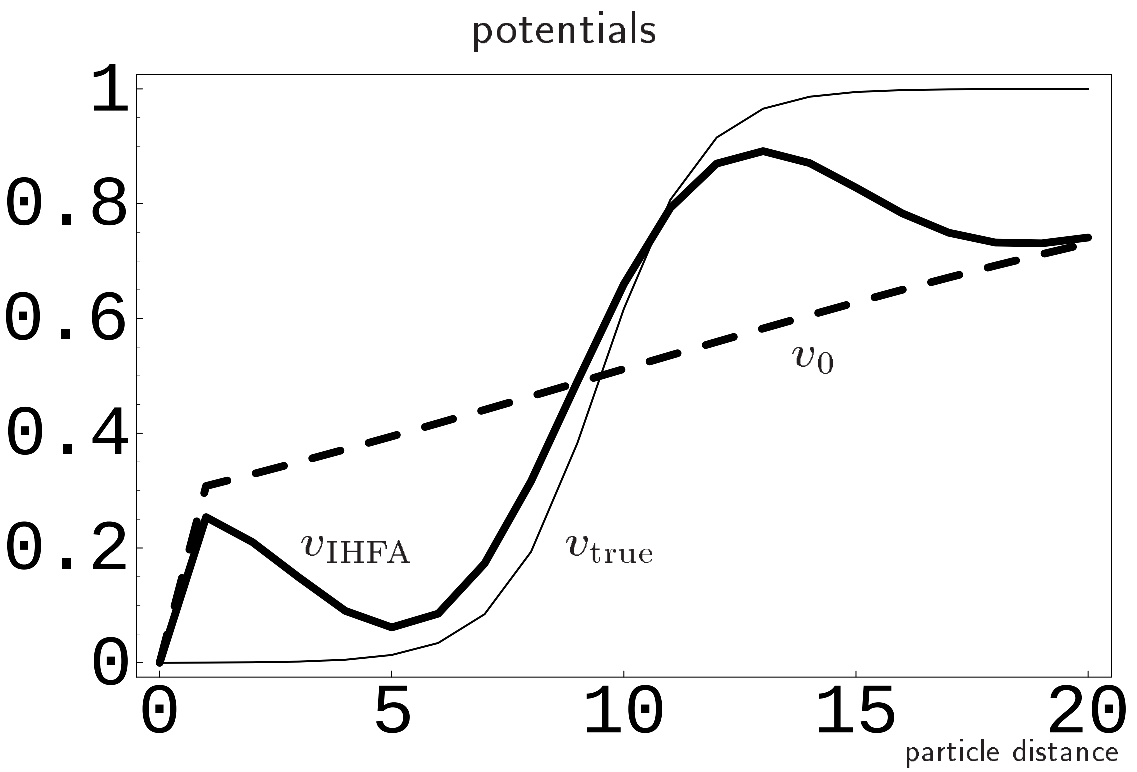
reconstructed likelihoods



Bayesian inverse Hartree–Fock theory:

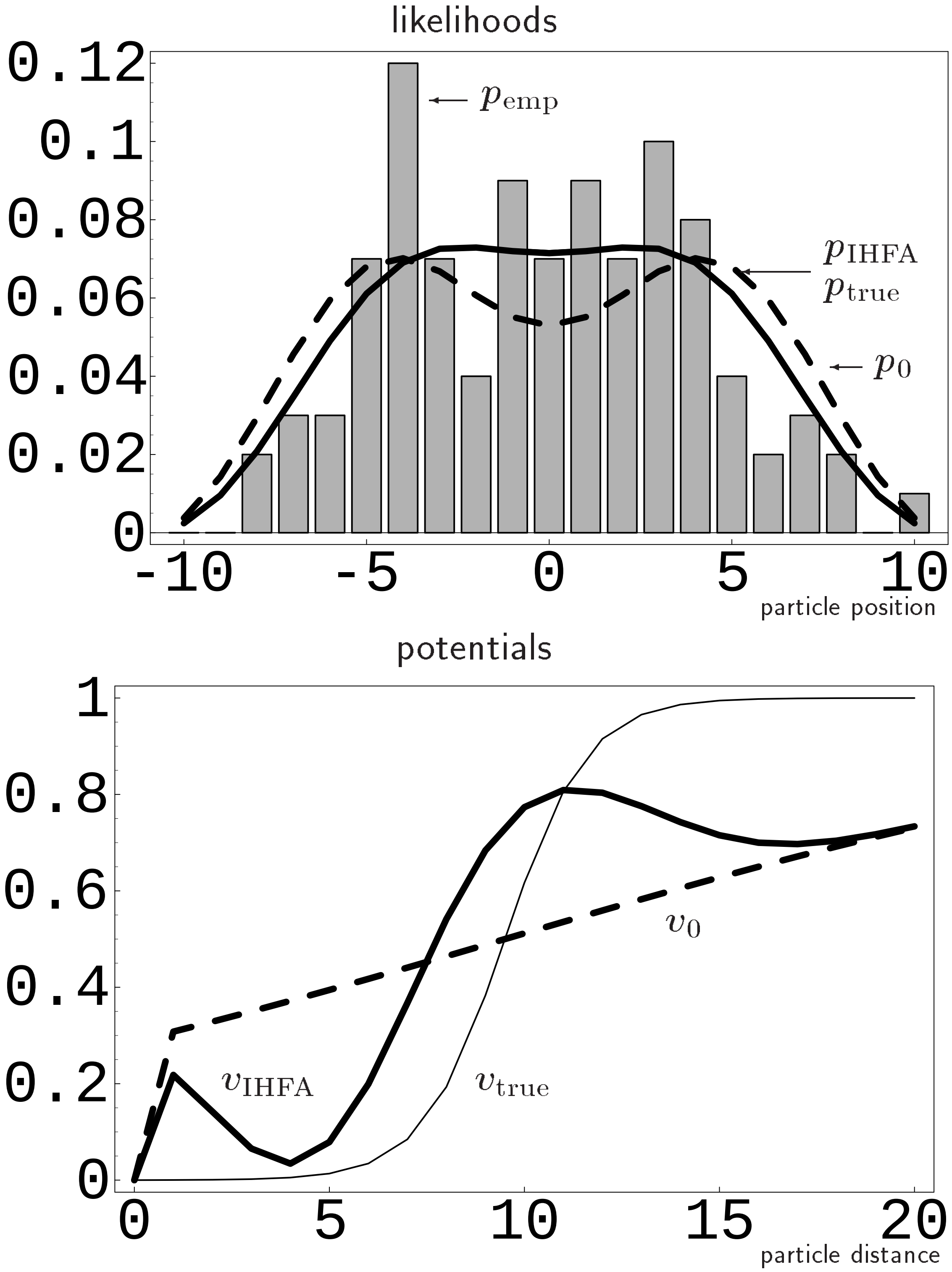
two–particle data





Bayesian inverse Hartree–Fock theory:

one–particle data



“Bayesian Field Theories”

likelihood $p(y\|x,h)$	problem type
$\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{[y-h(x)]^2}{2\sigma^2}\right]$ (Gaussian, σ fixed)	regression J.L., ICANN99 (image completion)
$\sum_k \frac{1}{\sigma_k\sqrt{2\pi}}\exp\left[-\frac{[y-h_k(x)]^2}{2\sigma_k^2}\right]$ (mixture of Gaussians)	clustering
discrete y	classification
of general form	density estimation J.L., physics/9912005, MaxEnt2000
Trace$[\rho(h)\,\pi(x,y)]$	inverse quantum statistics J.L., Uhlig, Weiguny, PRL 84 (2000) inverse Hartree–Fock Theory J.L., Uhlig, PRL 84 (2000) inverse time–dependent QM J.L., PLA 276 (2000)
$\langle\delta(y-x[x_{\rm cl},p_{\rm cl}])\rangle_{\rm cl}$	classical approx. of inv. QS J.L., Uhlig, to app. in Few Body Sys.

Nonparametric Bayesian methods

Advantages:

1. Based on realistic data assumptions: Only a finite number of noisy observational data is used. No expectation values or whole functions are needed as input.

2. Explicit and situation specific implementation of "a–priori" information is possible (mixture of Gausian prior processes, hyperparameter and hyperfields).

Disadvantage:

Numerically expensive