Bayesian reconstruction of free energy profiles from umbrella samples

Thomas Stecher
with Gábor Csányi and Noam Bernstein

21st November 2012

Motivation and Goals

- Many conventional methods implicitly assume a smooth free energy
  - Thermodynamic Integration
  - Umbrella Integration
- Improve by making **smoothness assumption explicit**?
- Aim to develop an efficient technique to analyse data
- Try and use as much information as possible
- Alternative to WHAM, UI
- Bayesian framework: consistent way to analyse noisy data

Umbrella sampling in multiple windows

- Free energy in collective variable $u = s(\mathbf{x})$:

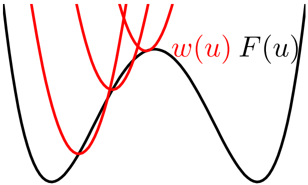
$$F(u) = -\frac{1}{\beta} \ln P(u) \quad P(u) = \frac{1}{Z} \int e^{-\beta V(\mathbf{x})} \delta(s(\mathbf{x}) - u) d\mathbf{x}$$

- Thermal barriers make direct sampling impossible

- Torrie and Valleau (1974): add umbrella potential(s) $w(u)$ to sample specific region:

$$F(u) = F^b(u) - w(u) + C$$

- Repeat
- Complication: C will vary from window to window



Bayes' Theorem

- Tackles **inverse probability** problems: given some data, what are the model parameters?
- One-line derivation:

$$P(\mathbf{f}|\mathbf{y})P(\mathbf{y}) = P(\mathbf{y}|\mathbf{f})P(\mathbf{f}) = P(\mathbf{y},\mathbf{f})$$

$$P(\mathbf{f}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{f})P(\mathbf{f})}{P(\mathbf{y})} \qquad \text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

- $P(\mathbf{y}) = \int P(\mathbf{y}|\mathbf{f})P(\mathbf{f})d\mathbf{f}$ is a normalisation constant

Gaussian Process Prior

- Prior distribution over functions
- "A collection of random variables, any finite number of which have a joint Gaussian distribution"
- Think: infinite-dimensional multivariate Gaussian
- Defined by a mean function,

$$m(x) = \langle f(x) \rangle$$

and a covariance function,

$$k(x_1, x_2) = \langle (f(x_1) - m(x_1))(f(x_2) - m(x_2)) \rangle$$

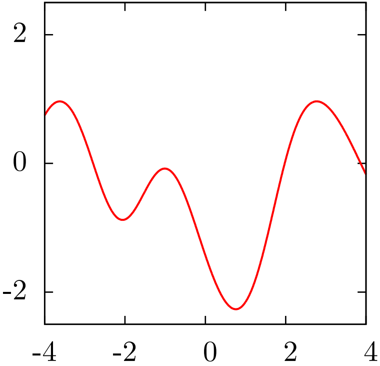
Samples from a Gaussian Process Prior

$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2/l^2)$$

Samples from a Gaussian Process Prior

$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2/l^2)$$

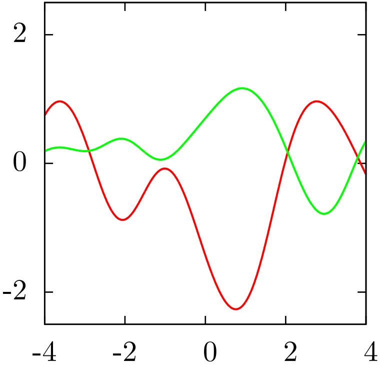
$l = 1.0$, $\sigma_f = 1.0$



Samples from a Gaussian Process Prior

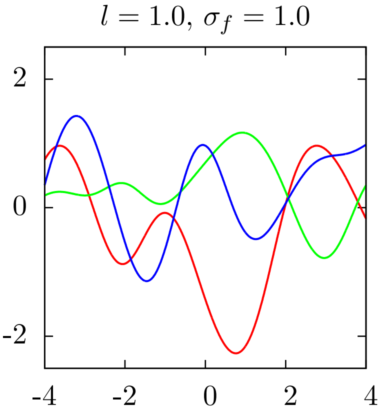
$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2/l^2)$$

$l = 1.0$, $\sigma_f = 1.0$



Samples from a Gaussian Process Prior

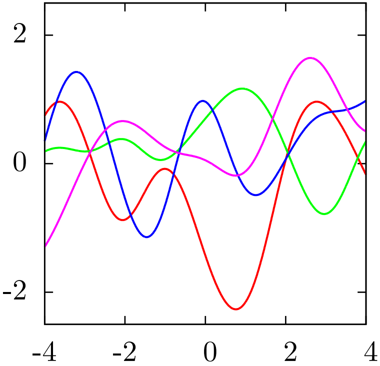
$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2 / l^2)$$



Samples from a Gaussian Process Prior

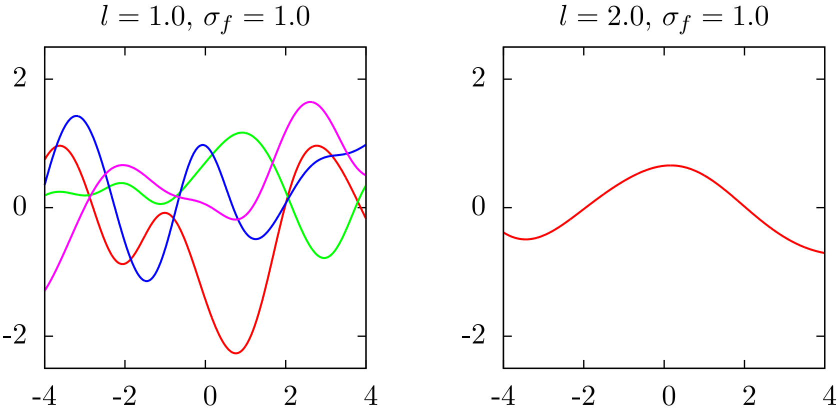
$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2 / l^2)$$

$l = 1.0$, $\sigma_f = 1.0$



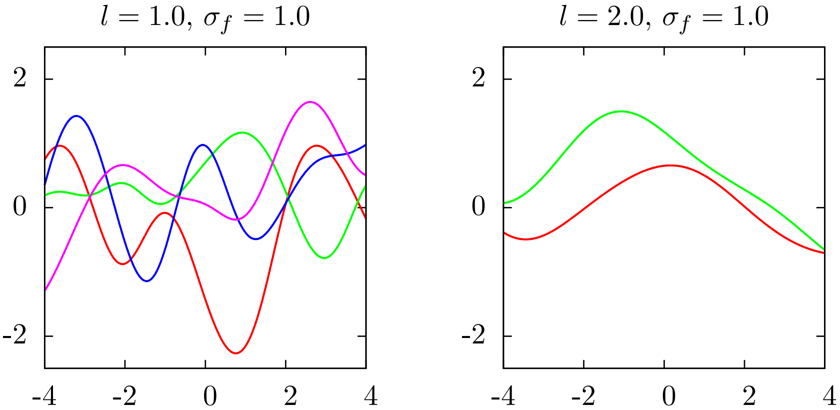
Samples from a Gaussian Process Prior

$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2/l^2)$$



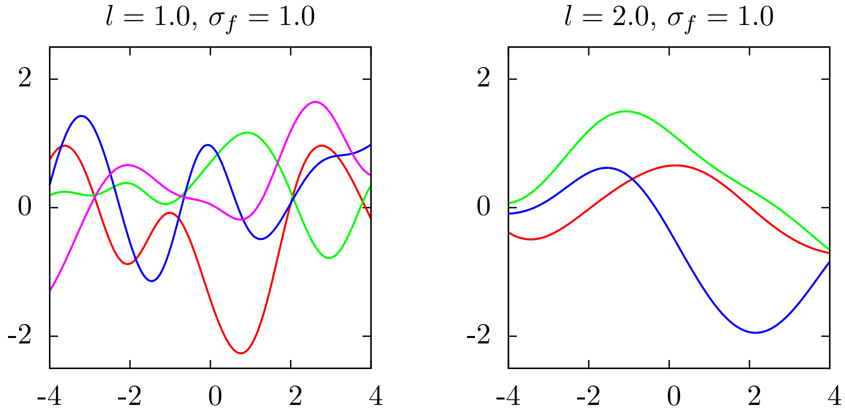
Samples from a Gaussian Process Prior

$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2/l^2)$$



Samples from a Gaussian Process Prior

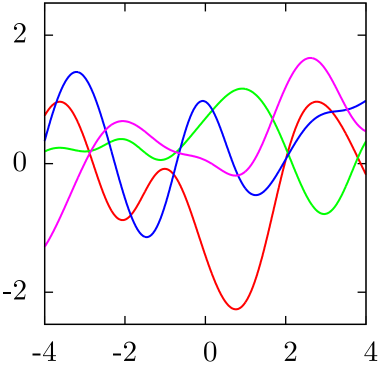
$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2/l^2)$$



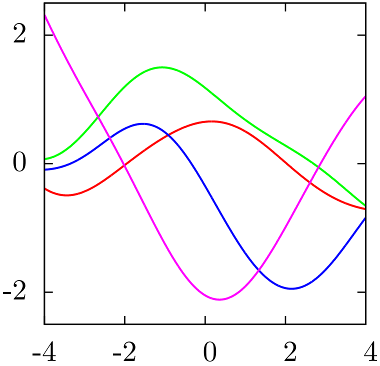
Samples from a Gaussian Process Prior

$$m(x) = 0 \qquad k(x_1, x_2) = \sigma_f^2 \exp(-\tfrac{1}{2}(x_1 - x_2)^2/l^2)$$

$l = 1.0,\ \sigma_f = 1.0$



$l = 2.0,\ \sigma_f = 1.0$



Likelihood and posterior distribution

- Observations with **Gaussian noise**:

$$\mathbf{y}|\mathbf{f}(\mathbf{x}) \sim N(\mathbf{f}(\mathbf{x}), \Sigma_y)$$

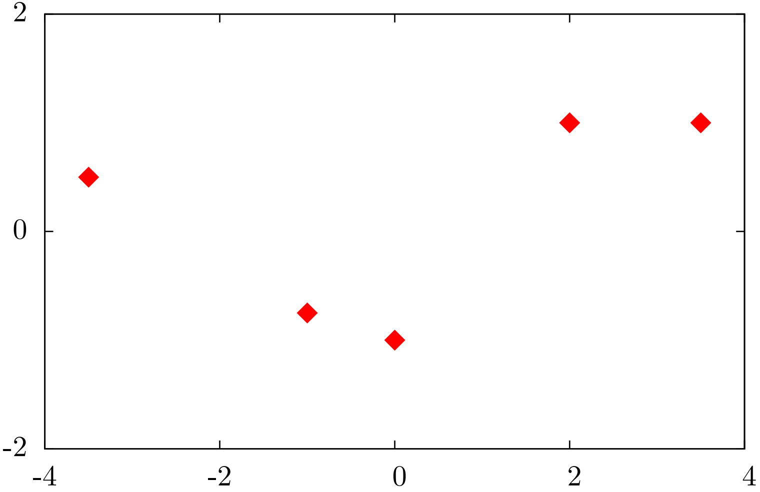
- Results in Gaussian process posterior:

$$\bar{f}(x^*) = \mathbf{k}^T(x^*)(K + \Sigma_y)^{-1}\mathbf{y}$$
$$\operatorname{cov}(f(x_1^*), f(x_2^*)) = k(x_1^*, x_2^*) - \mathbf{k}^T(x_1^*)(K + \Sigma_y)^{-1}\mathbf{k}(x_2^*)$$

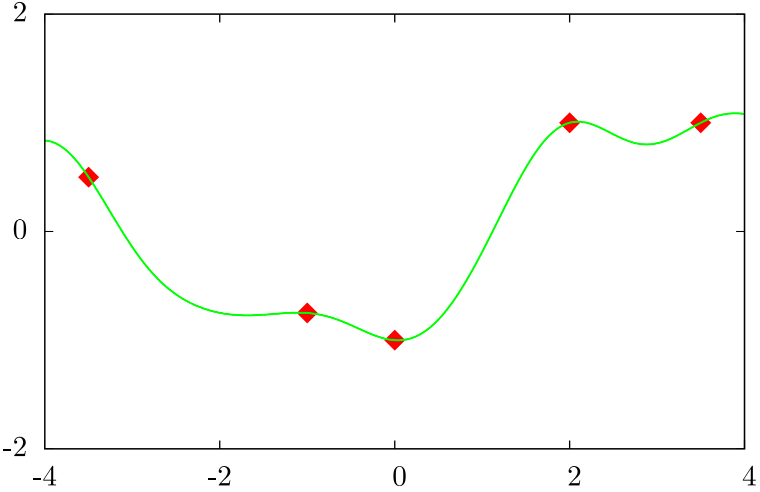
where

$$(\mathbf{k}(x^*))_i = k(x^*, x_i) \quad K_{ij} = k(x_i, x_j)$$

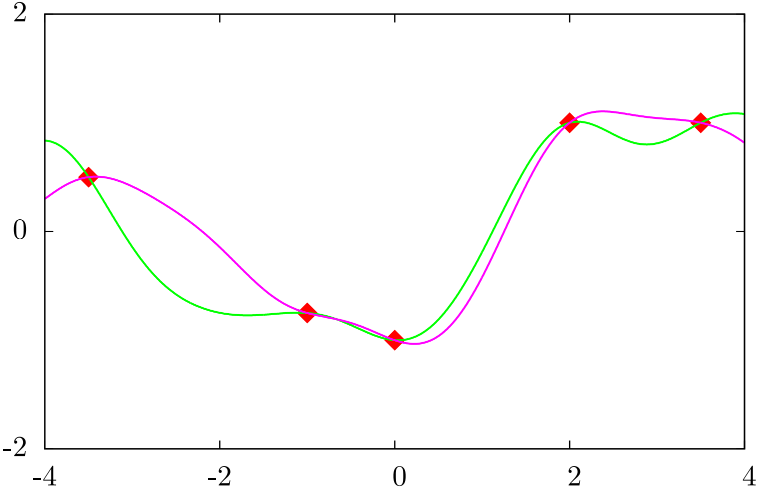
Samples from the posterior



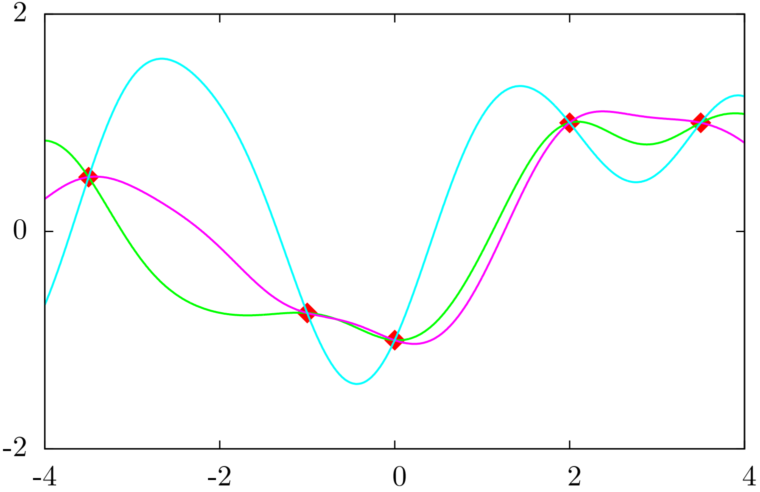
Samples from the posterior



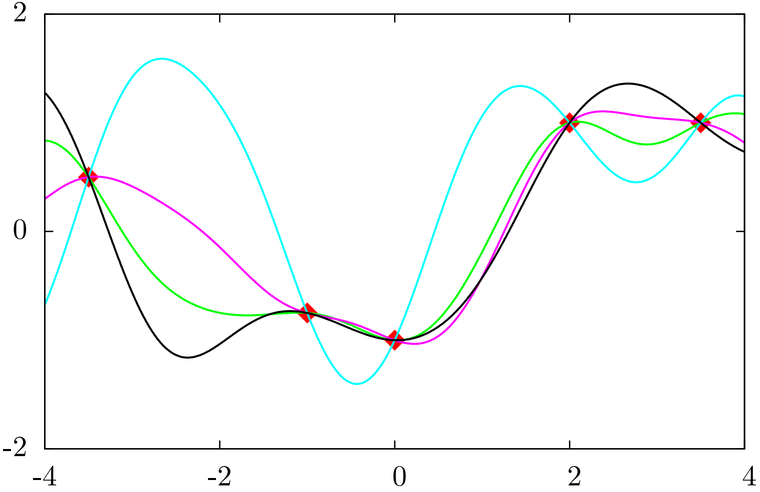
Samples from the posterior



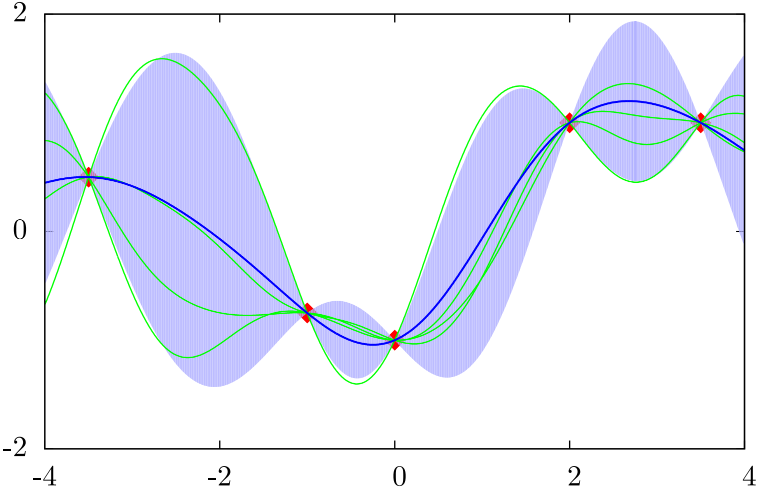
Samples from the posterior



Samples from the posterior



Samples from the posterior



Learning from umbrella sampled histograms

- Bin probabilities → biased free energies → unbiased free energies

$$F(u) = F^b(u) - w(u) + C$$

- Use unbiased free energies in GPR. Two complications:

1. True **likelihood** (noise) not Gaussian: approximate:
   - Multinomial bin probabilities → ≈ Gaussian → propagate to **first order** to free energies
2. Need to **account for unknown constants** C

Accounting for the unknown free energy offsets

Two equivalent ideas:

1. **Model unknown constants explicitly** with a flat (uninformative) prior:

$$\begin{aligned} \bar{f}(x^*) &= \mathbf{k}^T(x^*) K_y^{-1}(\mathbf{y} - H^T \bar{\boldsymbol{\beta}}) \\ \mathrm{cov}(f(x_1^*), f(x_2^*)) &= k(x_1^*, x_2^*) - \mathbf{k}^T(x_1^*) K_y^{-1} \mathbf{k}(x_2^*) \\ &\quad + \mathbf{k}^T(x_1^*) K_y^{-1} H^T [H K_y^{-1} H^T]^{-1} H K_y^{-1} \mathbf{k}(x_2^*) \end{aligned}$$

   where

$$\bar{\boldsymbol{\beta}} = [H K_y^{-1} H^T]^{-1} H K_y^{-1} \mathbf{y}$$

2. Use free energy **differences** as input
   - Linearity of difference operation crucial

Derivative information (UI)

- Mean forces from either constraint (TI) or restraint simulations (UI: Kästner and Thiel, 2005)
- Harmonic umbrella:

$$\left.\frac{\partial F(u)}{\partial u}\right|_{\bar{u}_i} \approx -\kappa_i(\bar{u}_i - u_i)$$

- Derivative (linear) of Gaussian process is another Gaussian process:

$$\begin{aligned} k_{f,f'}(x_1,x_2) &= \left\langle f(x_1)\frac{\partial}{\partial x_2} f(x_2)\right\rangle \\ &= \frac{\partial}{\partial x_2}\langle f(x_1)f(x_2)\rangle = \frac{\partial}{\partial x_2} k(x_1,x_2) \end{aligned}$$

$$k_{f',f'}(x_1,x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} k(x_1,x_2)$$

Derivative information (UI)

$$k_{f,f'}(x_1, x_2) = \frac{\partial}{\partial x_2} k(x_1, x_2)$$
$$k_{f',f'}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} k(x_1, x_2)$$

Results in posterior:

$$\bar{f}(x^*) = \mathbf{k}_{f,f'}^T(x^*)(K_{f',f'} + \Sigma_{y'})^{-1}\mathbf{y}'$$
$$\mathrm{cov}(f(x_1^*), f(x_2^*)) = k(x_1^*, x_2^*) - \mathbf{k}_{f,f'}^T(x_1^*)(K_{f',f'} + \Sigma_{y'})^{-1}\mathbf{k}_{f,f'}(x_2^*)$$

where

$$(\mathbf{k}_{f,f'}(x^*))_i = k_{f,f'}(x^*, x_i) \quad (K_{f',f'})_{ij} = k_{f',f'}(x_i, x_j)$$

Using derivative information in conjunction with histograms

- Define a joint Gaussian process of function values and its derivatives

$$K_{f\oplus f'} = \begin{pmatrix} K(\mathbf{x}_y, \mathbf{x}_y) & K_{ff'}(\mathbf{x}_y, \mathbf{x}_{y'}) \\ K^T_{ff'}(\mathbf{x}_y, \mathbf{x}_{y'}) & K_{f'f'}(\mathbf{x}_{y'}, \mathbf{x}_{y'}) \end{pmatrix}$$

- Assume no correlation between function and derivative noise:

$$\Sigma_{y\oplus y'} = \Sigma_y \oplus \Sigma_{y'}$$

Using derivative information in conjunction with histograms

- Define a joint Gaussian process of function values and its derivatives

$$K_{f\oplus f'} = \begin{pmatrix} K(\mathbf{x}_y, \mathbf{x}_y) & K_{ff'}(\mathbf{x}_y, \mathbf{x}_{y'}) \\ K_{ff'}^T(\mathbf{x}_y, \mathbf{x}_{y'}) & K_{f'f'}(\mathbf{x}_{y'}, \mathbf{x}_{y'}) \end{pmatrix}$$

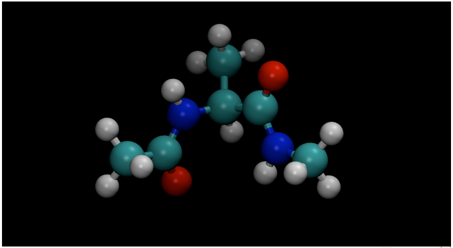
- Assume no correlation between function and derivative noise:

$$\Sigma_{y\oplus y'} = \Sigma_y \oplus \Sigma_{y'}$$

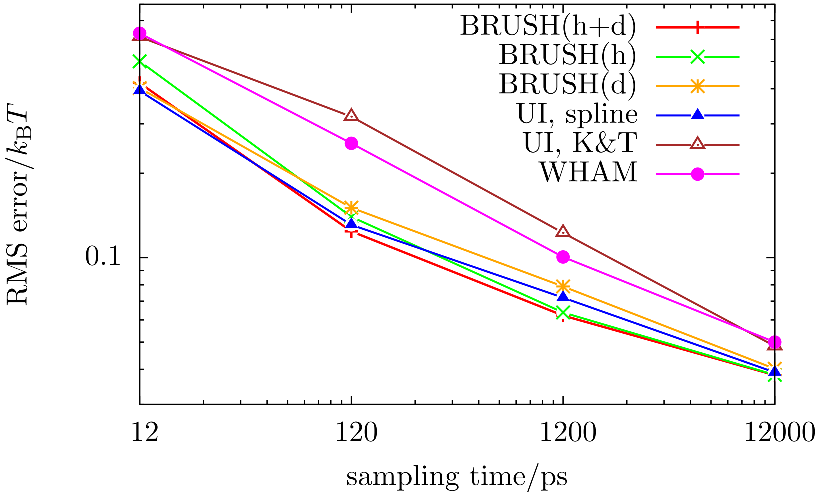
- BRUSH: Bayesian Reconstruction of free energy from Umbrella Samples using Histograms
  - BRUSH(h)
  - BRUSH(d)
  - BRUSH(h+d)

1D performance

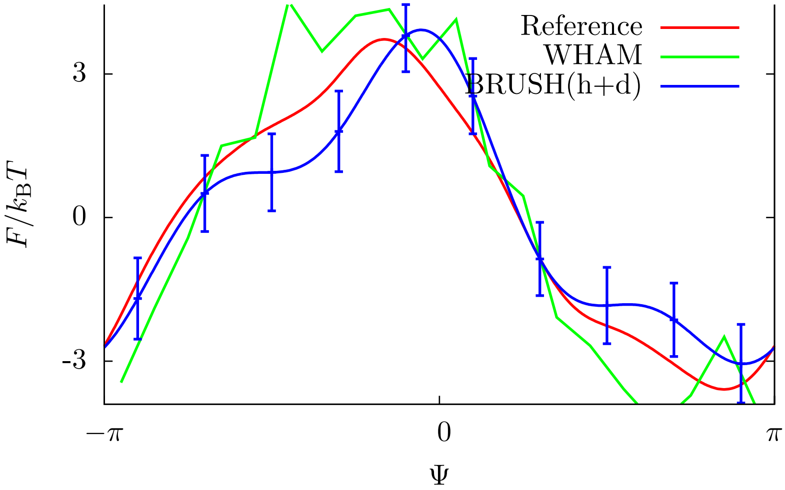
- Slice of alanine dipeptide as test system
- Reference reconstruction:
  - 50 x 50 ns MD
  - umbrella strength: 100 kcal/mol
- Compare methods by comparing **average RMS errors** (over 100 reconstructions) for an **'optimised' set of parameters**:
  - Umbrella strength: 5, 15, 25, 50, 100 kcal/mol
  - Number of windows: 24, 12, 8
  - Number of bins: 20, 40, 80 (WHAM); 2-10 (GP)



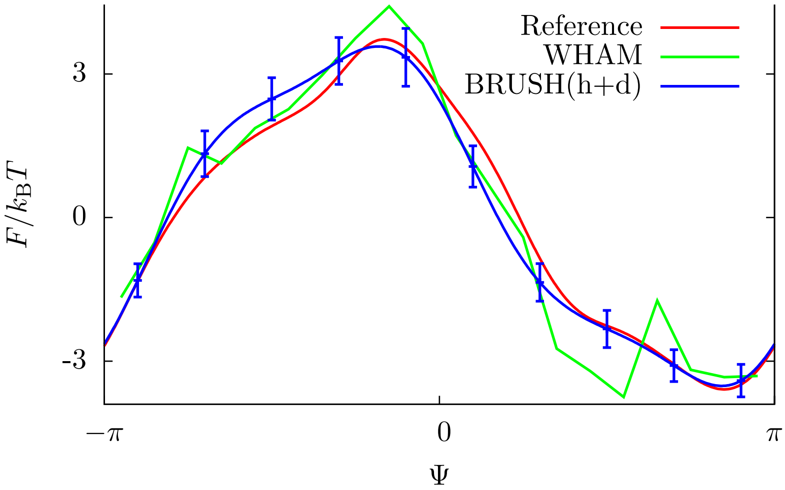
1D performance



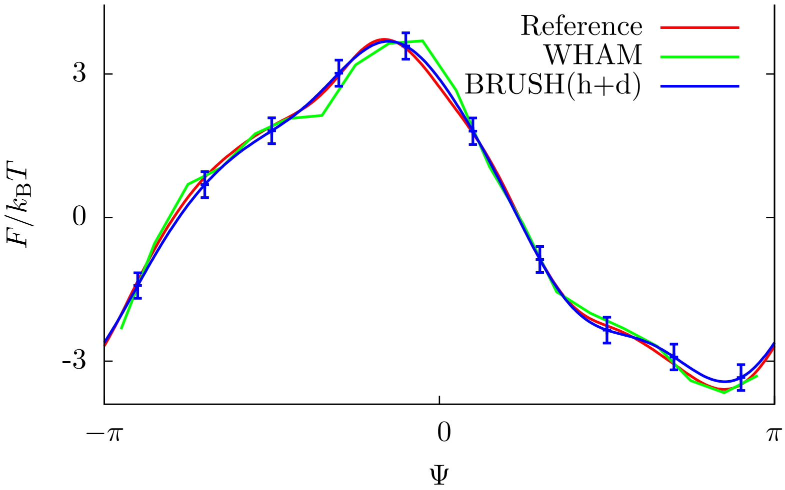
Representative reconstructions (6 ps)



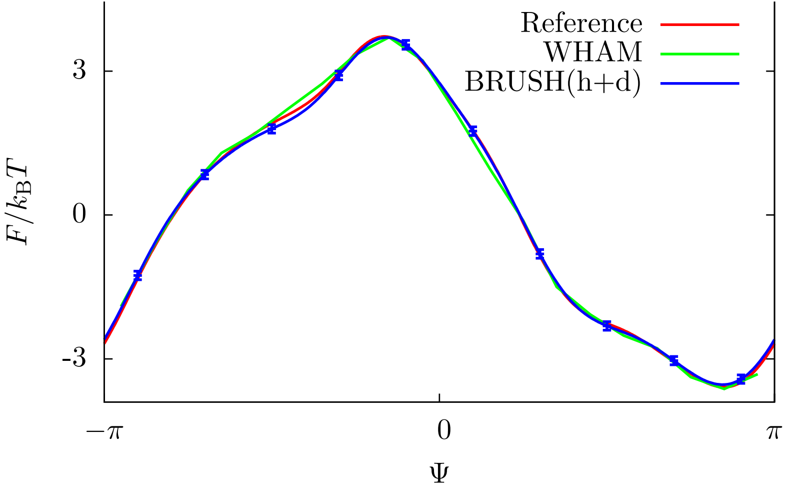
Representative reconstructions (12 ps)



Representative reconstructions (120 ps)



Representative reconstructions (1200 ps)

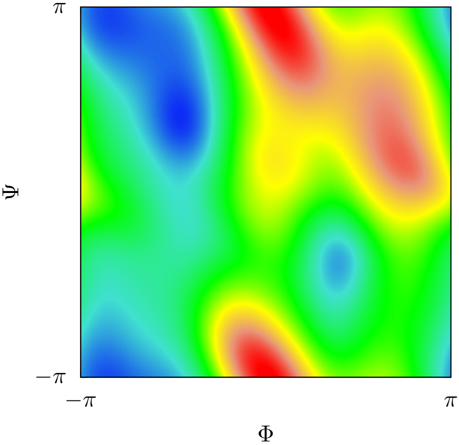


1D Conclusions

- Ouperform conventional methods such as WHAM and UI
- Can use different types of data systematically
- Uncertainty estimates intrinsic to the method and accurate
- Smooth reconstructions

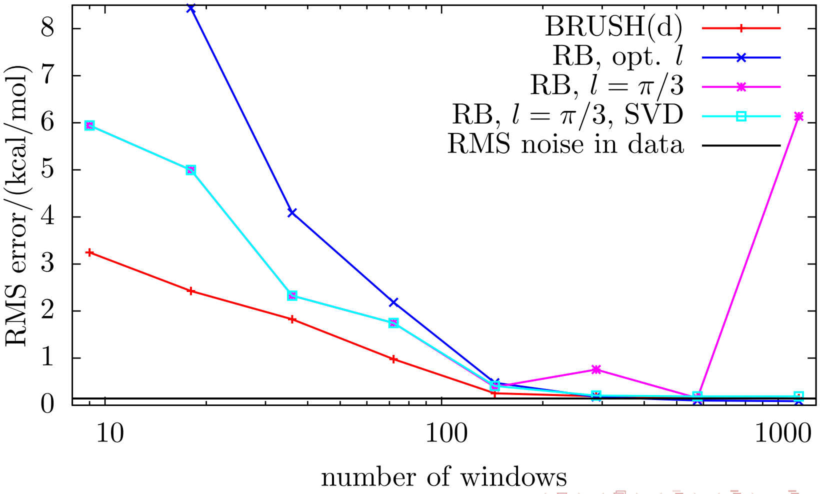
Two and more collective variables

- Want to reconstruct from derivative information
- Conventional approach: least-squares fit of of **radial basis (RB)** functions
- Known problems with high condition numbers

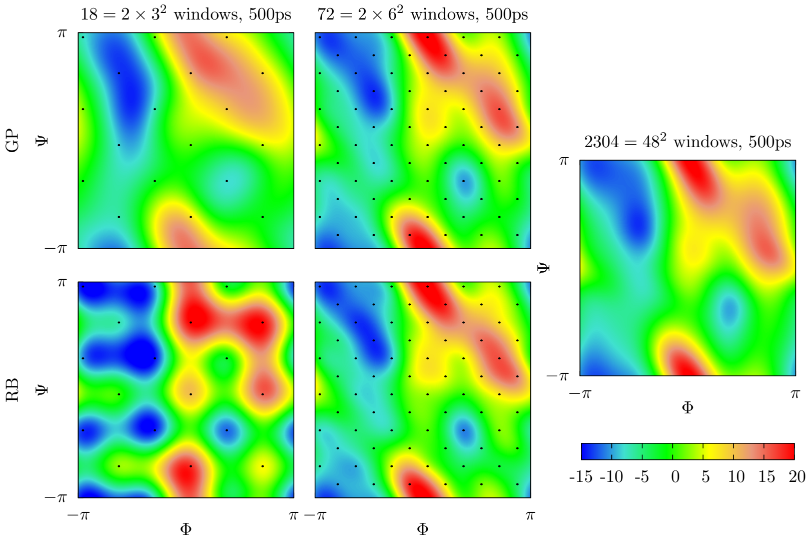


Varying the number of windows (derivative readings)

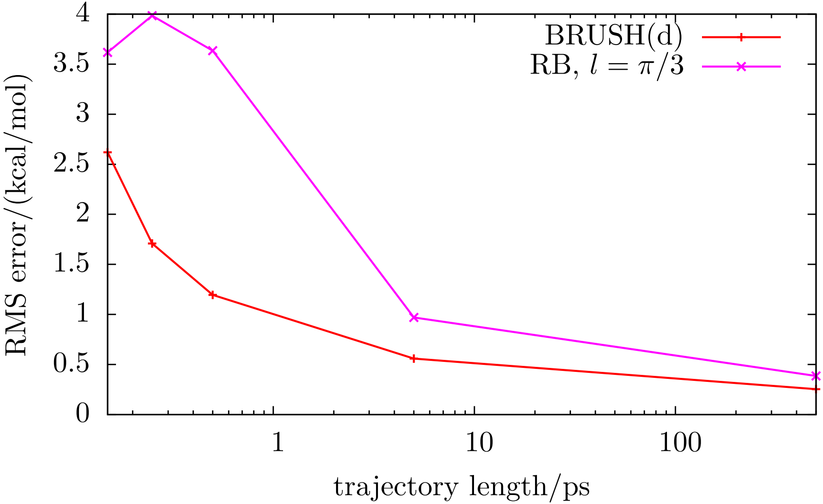
500 ps/window



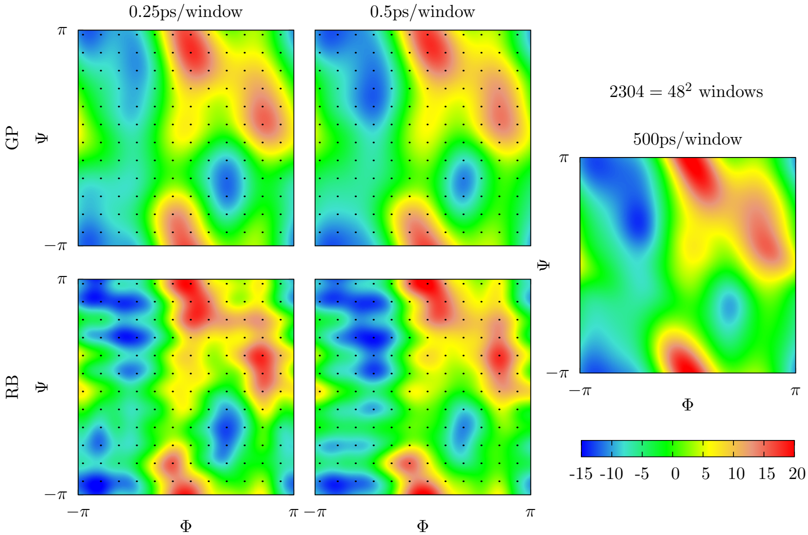
Varying the number of windows (derivative readings)



Varying the simulation length



Varying the simulation length



Conclusions and Outlook

- More efficient use of data in Bayesian framework
- Concurrent use of different information possible
  - Histograms
  - Derivatives
- Outperforms WHAM and UI
- Outperforms radial basis functions in several regimes

Conclusions and Outlook

- More efficient use of data in Bayesian framework
- Concurrent use of different information possible
  - Histograms
  - Derivatives
- Outperforms WHAM and UI
- Outperforms radial basis functions in several regimes

- Go beyond Gaussian noise approximation?
- Learn from MD trajectory directly? (Density estimation)