Automated Experiment:

… as a scientist…

Bayesian optimization:

1. Works only in low-dimensional spaces
2. The correlations are defined by the kernel function (very limiting)
3. We do not use any knowledge about physics of the system
4. We do not use cheap information available during the experiment (proxies)

GP Augmented with Structural model

Define a probabilistic model:

$$\mathbf{y} \sim MVNormal(\mathbf{m}, \mathbf{K})$$

$$K_{ij} = \sigma^2 \exp(0.5(x_i - x_j)^2 / l^2$$

$$\sigma \sim LogNormal(0, s_1)$$

$$l \sim LogNormal(0, s_2)$$

Prediction on new data X_*:

$$\mathbf{f}_*^i \sim MVNormal\left(\mu_{\boldsymbol{\theta}^i}^{\text{post}}, \Sigma_{\boldsymbol{\theta}^i}^{\text{post}}\right)$$

replaced with

$$\mu_{\boldsymbol{\theta}^i}^{\text{post}} = \mathbf{m}(X_*) + \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}(X, X|\boldsymbol{\theta}^i)^{-1}(\mathbf{y} - \mathbf{m}(X)) \rightarrow \mu_{\boldsymbol{\Omega}^i}^{\text{post}} = \mathbf{m}(X_*|\phi^i) + \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}(X, X|\boldsymbol{\theta}^i)^{-1}(\mathbf{y} - \mathbf{m}(X|\phi^i))$$

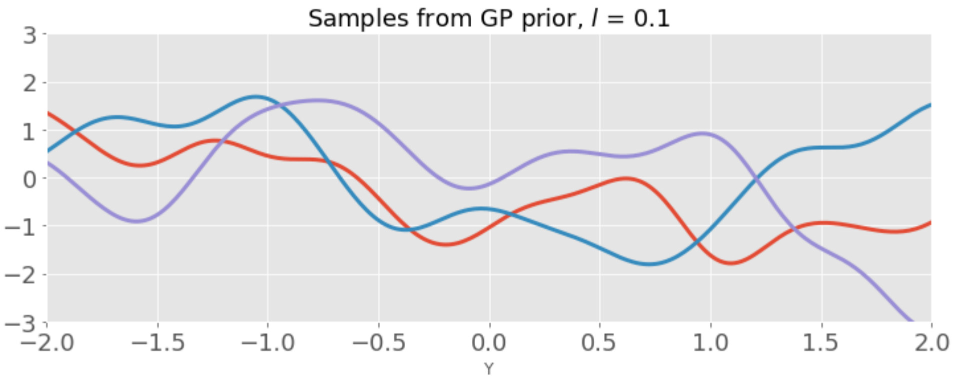
$$\Sigma_{\boldsymbol{\theta}^i}^{\text{post}} = \mathbf{K}(X_*, X_*|\boldsymbol{\theta}^i) - \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}(X, X|\boldsymbol{\theta}^i)^{-1}\mathbf{K}(X, X_*|\boldsymbol{\theta}^i)$$

$\boldsymbol{\Omega}^{\boldsymbol{i}} = \{\phi^i, \boldsymbol{\theta}^i\}$ is a single HMC posterior sample with the kernel and prob model parameters

- We substitute a constant GP prior mean function **m** with a structured probabilistic model of the expected system's behavior.
- This probabilistic model reflects our prior knowledge about the system, but it does not have to be precise.
- The model parameters are inferred together with the kernel parameters via the Hamiltonian Monte Carlo.
- The fully Bayesian treatment of the model allows additional control over the optimization via the selection of priors for the model parameters.

GP Augmented with Structural model

Prior predictive distribution: GP



Probabilistic model

$$m = y_0 - \sum_{n=1}^{N} L_n \quad (N=2)$$

$$y_0 \sim Uniform(-10, 10)$$

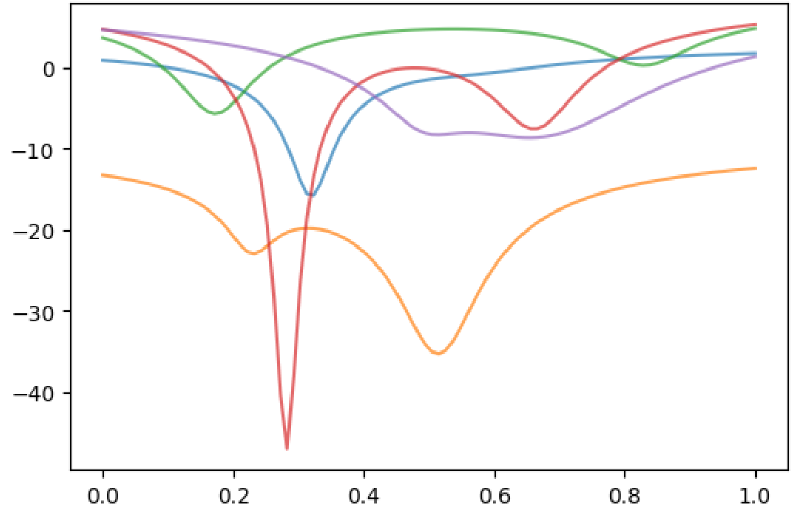
$$L_n \sim \frac{A_n}{\sqrt{(x - x_n^0)^2 + w_n^2}}$$

$$A_n \sim LogNormal(0, 1)$$

$$w_n \sim HalfNormal(.1)$$

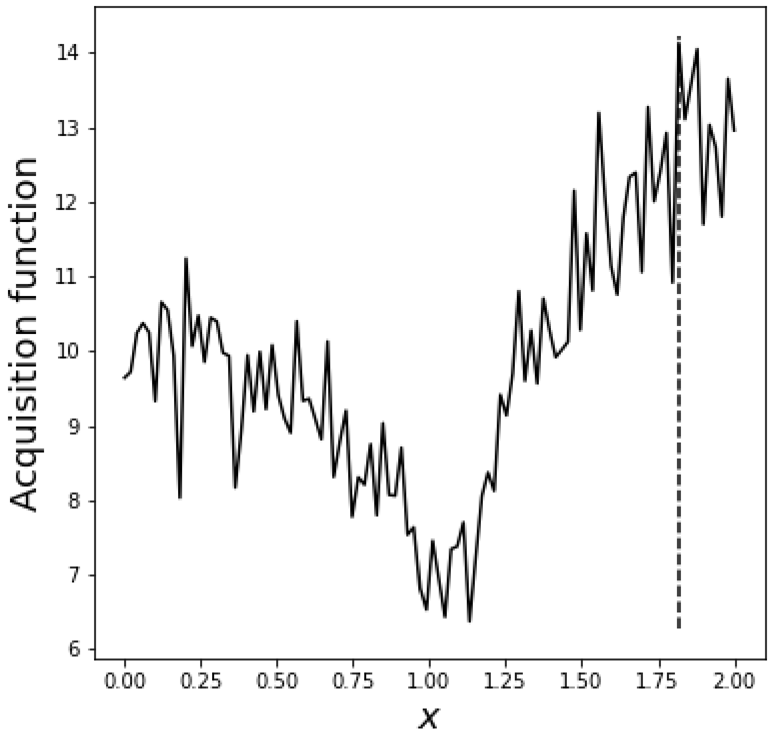
$$x_n^0 \sim Uniform(0, 1)$$

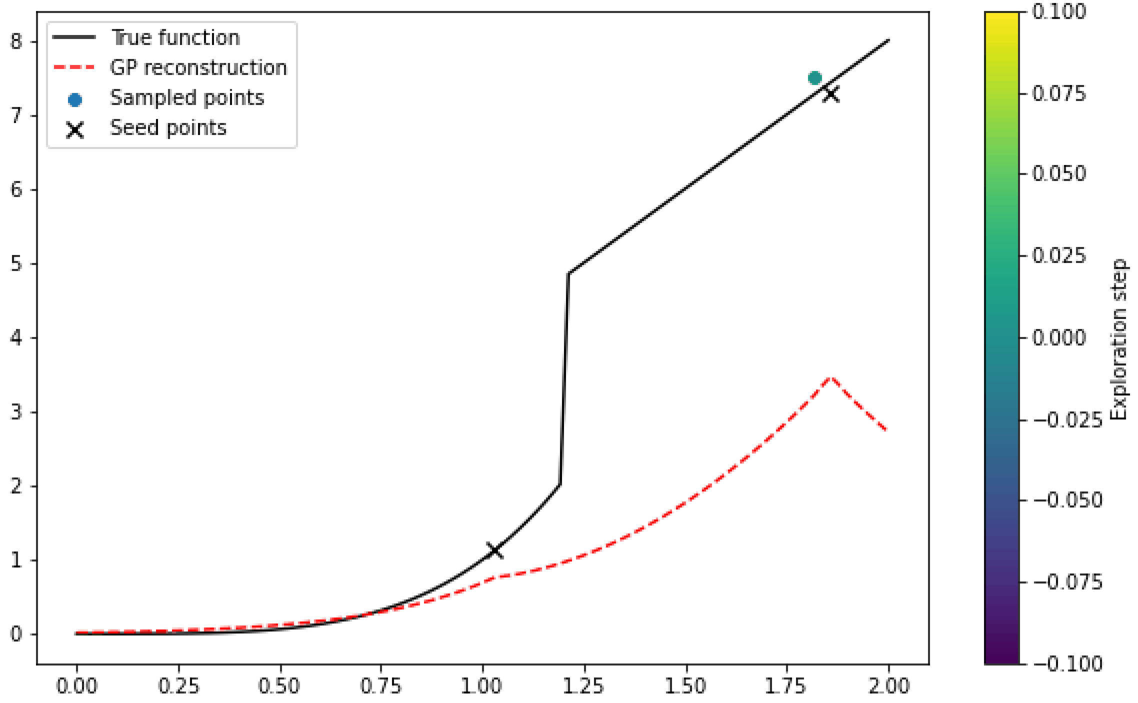
Prior predictive distribution: sGP



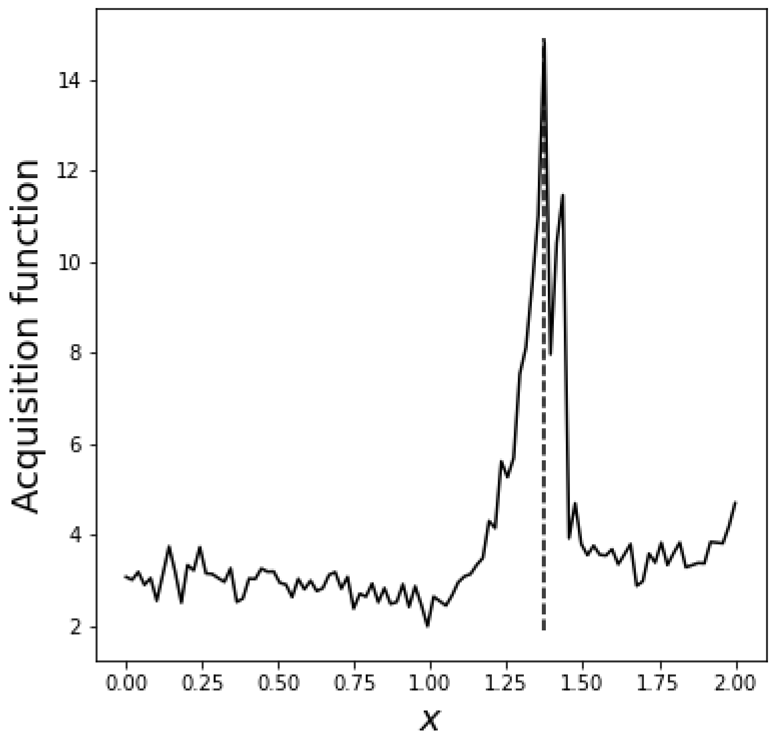
This model simply tells us that there are two minima in our data but does not assume to have any prior knowledge about their relative depth, width, or distance

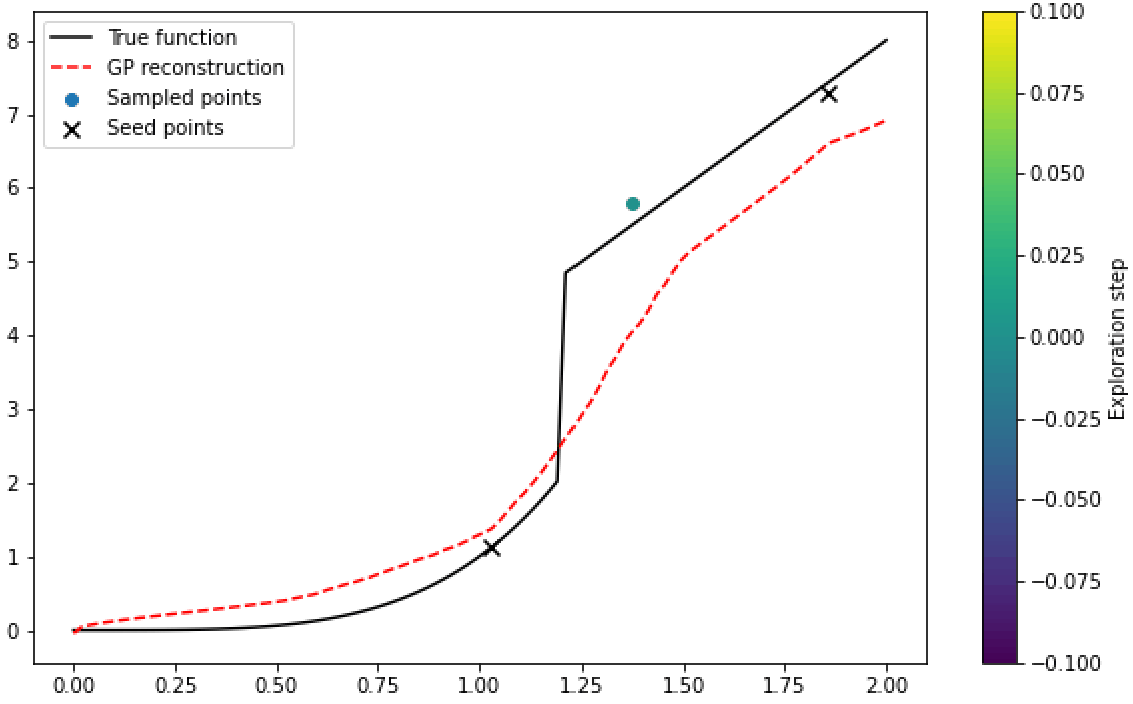
Simple GP search





Structured GP search





Application to Ising model

Probabilistic model

$$A/\tanh\left(\frac{f(J_1)+f(J_2)}{w}\right)$$

where $f(J)$ is a third-degree polynomial with normal priors on its parameters