TRIUMF

Extrapolating Nuclear Many-Body Calculations with Constrained Gaussian Processes

Peter Gysbers

M. Gennari, W. Fedorko, P. Navrátil

TRIUMF-Helmholtz Working Group

Nov 13, 2019

Discovery, accelerated

Context: *Ab Initio* Nuclear Theory

Goal: solve the nuclear eigenvalue problem

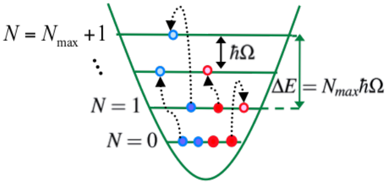
$$H|\Psi_k\rangle = E_k|\Psi_k\rangle \text{, where } H = \sum_i^A T_i + \sum_{i<j} V_{ij} + \sum_{i<j<f} V_{ijf} + \cdots$$

with nucleons as the degrees of freedom

The No-core Shell Model

Expand in anti-symmetrized products of harmonic oscillator single-particle states

$$|\Psi_k\rangle = \sum_{N=0}^{N_{max}} \sum_j c_{Nj}^k |\Phi_{Nj}\rangle$$

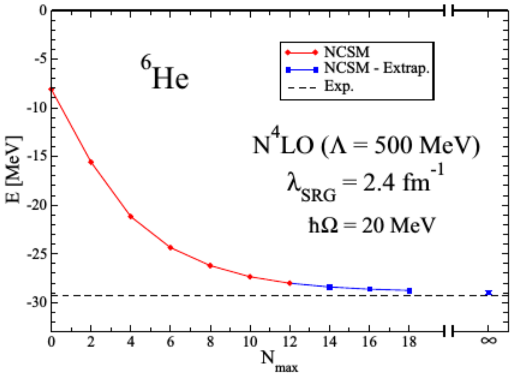


Calculations should converge to the exact value as $N_{max} \to \infty$

Motivation

- Computational complexity grows exponentially with basis size parameter N_{max}
- The functional form of convergence curve is not known
- Ad hoc extrapolation:

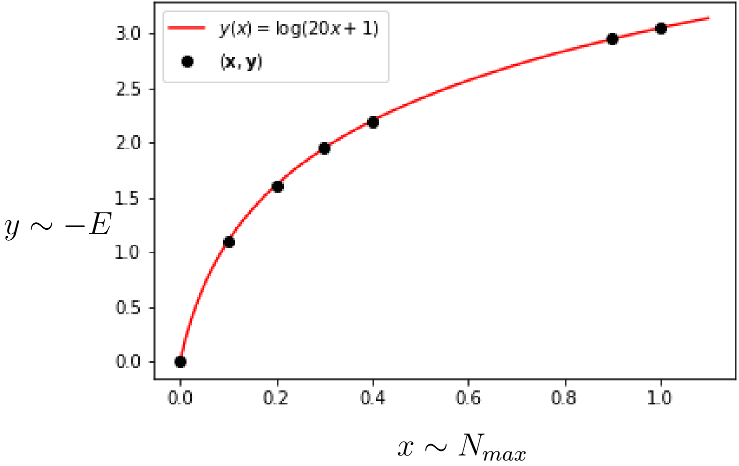
$$E = E_{\infty} + \alpha \exp(-bN_{max})$$



Goal: predict value at $N_{max} \to \infty$ with a meaningful error bar

Problem Statement

Given some data $\mathbf{y} = y(\mathbf{x})$, find the underlying function $y(x)$, i.e. predict $\mathbf{y}^* = y(\mathbf{x}^*)$



Gaussian Processes

Key Assumption (Prior):

$\mathbf{y}$ and $\mathbf{y}^*$ are drawn from a joint Gaussian distribution

$$p\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{y}^*\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\mu\\ \mu_*\end{bmatrix}, \begin{bmatrix}C & C_*\\ C_*^T & C_{**}\end{bmatrix}\right)$$

$$\begin{aligned} C &= C[\mathbf{y}, \mathbf{y}] \\ &= r(\mathbf{x}, \mathbf{x}) \\ C_* &= C[\mathbf{y}, \mathbf{y}^*] \\ &= r(\mathbf{x}, \mathbf{x}^*) \\ C_{**} &= r(\mathbf{x}^*, \mathbf{x}^*) \end{aligned}$$

Make predictions by conditioning on data:

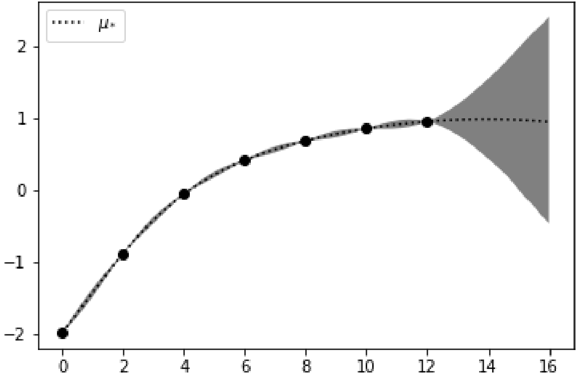
$$p(\mathbf{y}^*|\mathbf{y}) = \mathcal{N}\left(\mu_*, \Sigma_*\right)$$

where

$$\begin{aligned} \mu_* &= C_*^T C^{-1}\mathbf{y} \\ \Sigma_* &= C_{**} - C_*^T C^{-1} C_* \end{aligned}$$

Gaussian Processes give a distribution of predictions (within error band)

Problem: Error bars blow up outside of data!



Idea: Use information about derivatives

Constraints on Derivatives

Weight probability of samples:

$$p(\mathbf{y}^*|\mathbf{y}) \sim \mathcal{N}(\mu_*, \Sigma_*) \times m(\mathbf{y}') \times n(\mathbf{y}'')$$

based on criteria:

$$m(\mathbf{y}') = \sum_i \left(m(y_i') = \begin{cases} 1 & \text{if } y_i' > 0 \\ 0 & \text{otherwise} \end{cases} \right)$$

$$n(\mathbf{y}'') = \sum_i \left(n(y_i'') = \begin{cases} 1 & \text{if } y_i'' < 0 \\ 0 & \text{otherwise} \end{cases} \right)$$

Using Derivatives

- The derivative of a Gaussian process is a Gaussian process
- i.e. $\mathbf{y'}_i \equiv \frac{\mathrm{d}y}{\mathrm{d}x}|_{x=x'_i}$ is also jointly Gaussian distributed (as is $\mathbf{y''}$)

$$p\left(\begin{bmatrix}\mathbf{y}^*\\ \mathbf{y'}\\ \mathbf{y''}\end{bmatrix}\middle|\mathbf{y}\right) = \mathcal{N}(\nu, \Sigma)$$

(ν and Σ are more complicated (see Extra Slides))

We want the posterior distribution:

$$p\left(\begin{bmatrix}\mathbf{y}^*\\ \mathbf{y'}\\ \mathbf{y''}\end{bmatrix}\middle|\mathbf{y}\right) = \mathcal{N}(\nu, \Sigma) \times m(\mathbf{y'}) \times n(\mathbf{y''})$$

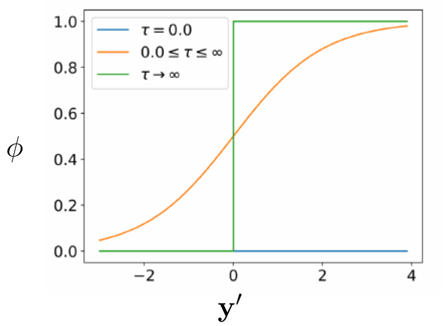
Use SMC!

Sequential Monte-Carlo / Particle Filter

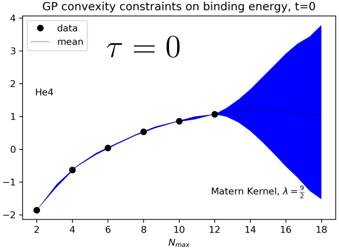
Draw N samples ("particles": $\begin{bmatrix}\mathbf{y}^* \\ \mathbf{y}' \\ \mathbf{y}''\end{bmatrix}$) from a GP

for τ_1, τ_2 from 0 to ∞:

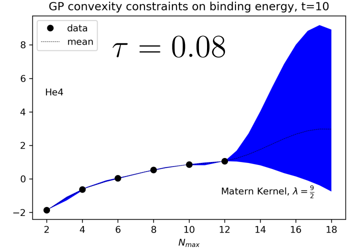
- for each particle:
  - propose new $\begin{bmatrix}\mathbf{y}^* \\ \mathbf{y}' \\ \mathbf{y}''\end{bmatrix}$ values "nearby" old values
  - accept or reject according to $p\left(\begin{bmatrix}\mathbf{y}^* \\ \mathbf{y}' \\ \mathbf{y}''\end{bmatrix} \middle| \mathbf{y}, \tau_1, \tau_2\right) = \mathcal{N}(\nu, \Sigma) \times \phi(\tau_1 \mathbf{y}') \times \phi(-\tau_2 \mathbf{y}'')$
- resample: throw away "bad" particles and keep multiple copies of "good" particles (weighted by constraints)



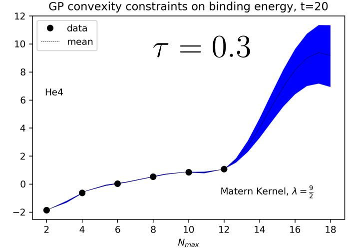
Results: He^4



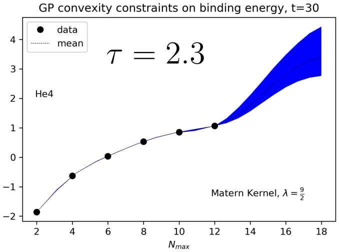
GP convexity constraints on binding energy, t=0
data
mean
$\tau = 0$
He4
Matern Kernel, $\lambda = \frac{9}{2}$
N_{max}



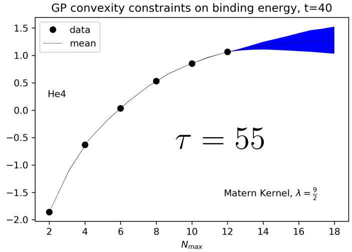
GP convexity constraints on binding energy, t=10
data
mean
$\tau = 0.08$
He4
Matern Kernel, $\lambda = \frac{9}{2}$
N_{max}



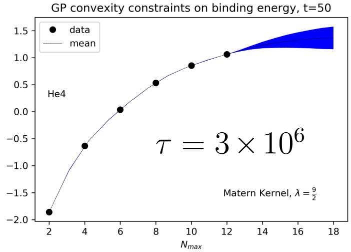
GP convexity constraints on binding energy, t=20
data
mean
$\tau = 0.3$
He4
Matern Kernel, $\lambda = \frac{9}{2}$
N_{max}



GP convexity constraints on binding energy, t=30
data
mean
$\tau = 2.3$
He4
Matern Kernel, $\lambda = \frac{9}{2}$
N_{max}

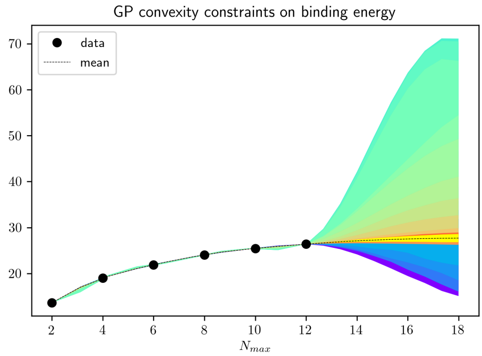


GP convexity constraints on binding energy, t=40
data
mean
He4
$\tau = 55$
Matern Kernel, $\lambda = \frac{9}{2}$
N_{max}

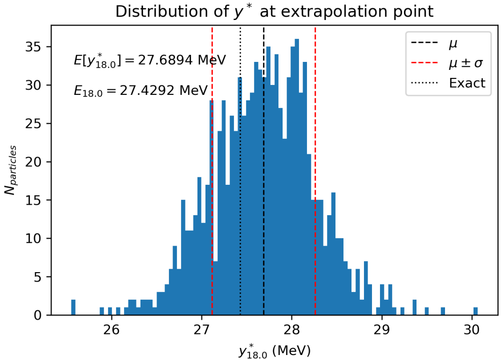


GP convexity constraints on binding energy, t=50
data
mean
He4
$\tau = 3 \times 10^6$
Matern Kernel, $\lambda = \frac{9}{2}$
N_{max}

Results: He^4

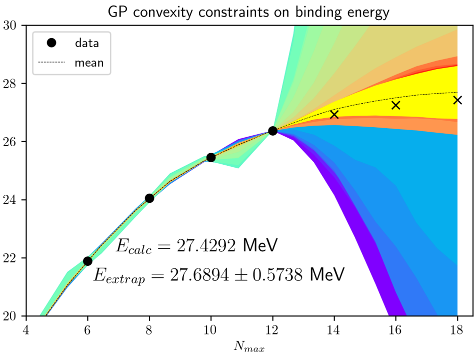


GP convexity constraints on binding energy
data
mean
70
60
50
40
30
20
2
4
6
8
10
12
14
16
18
N_{max}

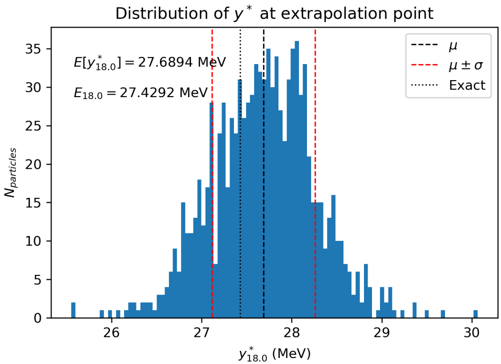


Distribution of y^* at extrapolation point
$E[y^*_{18.0}] = 27.6894$ MeV
$E_{18.0} = 27.4292$ MeV
μ
$\mu \pm \sigma$
Exact
$N_{particles}$
35
30
25
20
15
10
5
0
26
27
28
29
30
$y^*_{18.0}$ (MeV)

Results: He^4

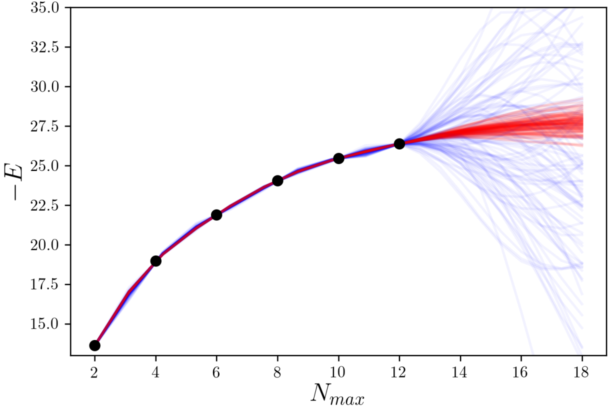


GP convexity constraints on binding energy
data
mean
$E_{calc} = 27.4292$ MeV
$E_{extrap} = 27.6894 \pm 0.5738$ MeV
N_{max}
30
28
26
24
22
20
4
6
8
10
12
14
16
18

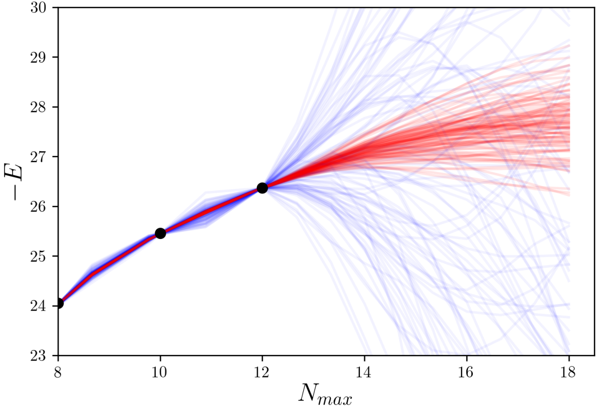


Distribution of y^* at extrapolation point
$E[y^*_{18.0}] = 27.6894$ MeV
$E_{18.0} = 27.4292$ MeV
μ
$\mu \pm \sigma$
Exact
$N_{particles}$
$y^*_{18.0}$ (MeV)
35
30
25
20
15
10
5
0
26
27
28
29
30

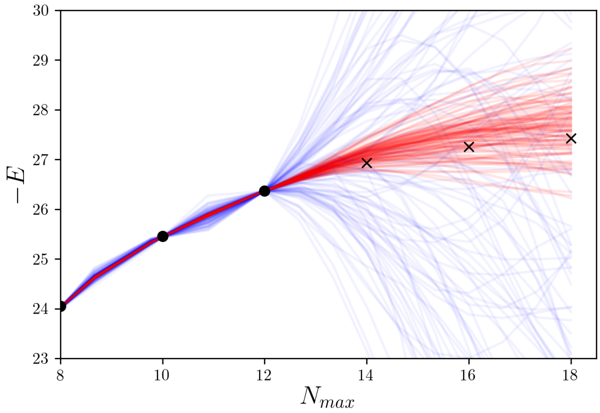
Results: He^4



Results: He^4



Results: He^4



Summary

- Constrained Gaussian processes are not yet competitive with state-of-the-art
- Predictions far from data are difficult

Outlook

- Add $y' \to 0$ constraint at very large N_{max}
- Try adaptive constraint schedules
- Try ”log kernels”
- Re-factored code to be shared

TRIUMF

Thank you
Merci

Discovery, accelerated

Extra: Gaussian Processes

Key Assumption (Prior):

$\mathbf{y}$ values are drawn from a multivariate Gaussian distribution

$$p\left(\begin{bmatrix} y_i \\ y_j \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \mu[y_i] \\ \mu[y_j] \end{bmatrix}, \begin{bmatrix} C[y_i, y_i] & C[y_i, y_j] \\ C[y_j, y_i] & C[y_j, y_j] \end{bmatrix}\right)$$

where C is the Covariance function defined by a *kernel* function e.g. Gaussian:

$$C[y_1, y_2] = r(x_1, x_2) = \sigma^2 \exp\left(-\frac{(x_1 - x_2)^2}{2\ell^2}\right)$$

In other words:

Assumption on function space: nearby inputs have nearby outputs
(i.e. if $|x_1 - x_2| \sim \ell$ then $|y_1 - y_2| > \sigma$ is unlikely)

Extra slide: Including Derivatives

$$p\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{y}^*\\ \mathbf{y}'\\ \mathbf{y}''\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\mu\\ \mu_*\\ \mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}C & C_* & C_1 & C_2\\ C_*^T & C_{**} & C_{1*} & C_{2*}\\ C_1^T & C_{*1} & C_{11} & C_{12}\\ C_2^T & C_{*2} & C_{21} & C_{22}\end{bmatrix}\right)$$

then

$$p\left(\begin{bmatrix}\mathbf{y}^*\\ \mathbf{y}'\\ \mathbf{y}''\end{bmatrix} |\mathbf{y}\right) = \mathcal{N}(\nu, \Sigma)$$

where

$$\nu = [C_*, C_1, C_2]C^{-1}\mathbf{y}$$

$$\Sigma = \begin{bmatrix}C_{**} & C_{1*} & C_{2*}\\ C_{*1} & C_{11} & C_{12}\\ C_{*2} & C_{21} & C_{22}\end{bmatrix} - [C_*, C_1, C_2]C^{-1}\begin{bmatrix}C_*\\ C_1\\ C_2\end{bmatrix}$$

$$\begin{aligned} C_{*1} &= C[\mathbf{y}^*, \mathbf{y}'] \\ &= \frac{\partial}{\partial x_j} r(\mathbf{x}^*, \mathbf{x}') \\ &\vdots \\ C_{22} &= C[\mathbf{y}'', \mathbf{y}''] \\ &= \frac{\partial^2}{\partial x_i^2}\frac{\partial^2}{\partial x_j^2} r(\mathbf{x}'', \mathbf{x}'') \end{aligned}$$