

GAUSSIAN PROCESS EMULATORS WITH ACTIVE SUBSPACE LEARNING

with Margarida Companys and Achim Schwenk

Tom Plies



MOTIVATION

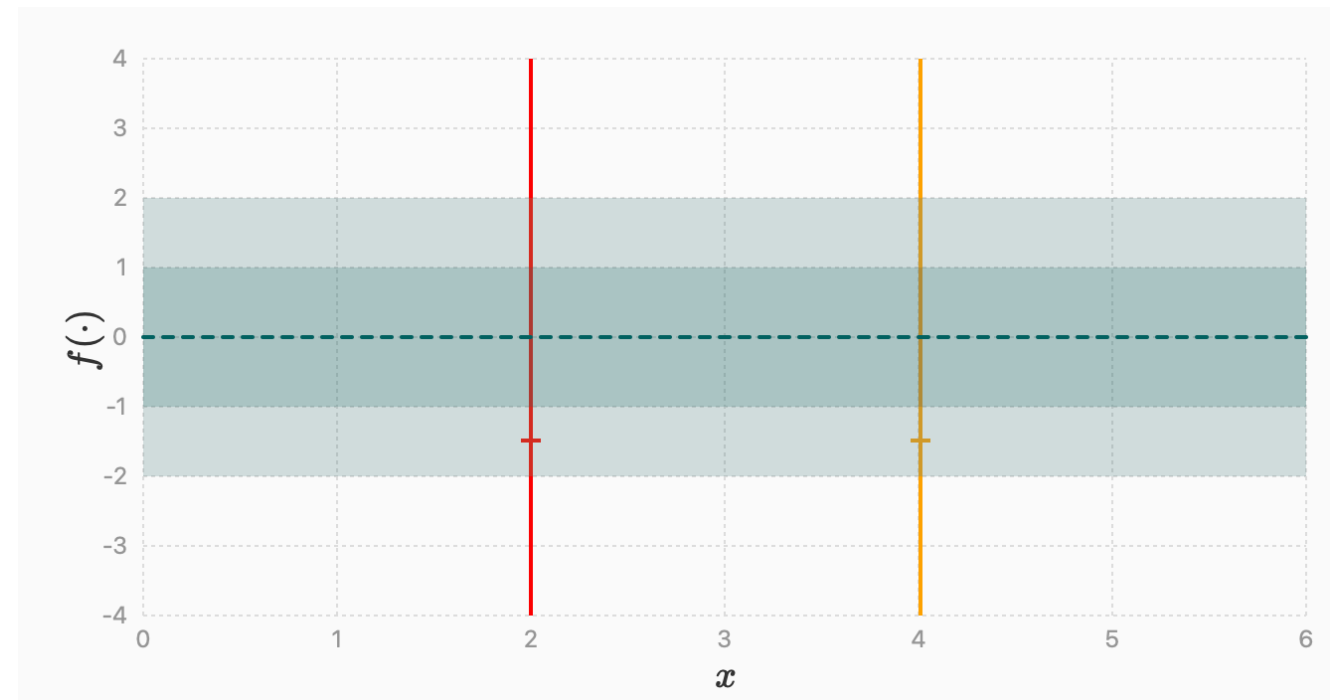
- Bayesian LEC inference could be improved through medium-mass constraints
 - Full IMSRG calculations are significantly too expensive
 - Emulators can be constructed based on Gaussian Processes (GPs)
 - Regular GP emulators are too inefficient to approximate IMSRG
-
- **Goal: Emulator for nuclear structure observables with Gaussian Processes**

GAUSSIAN PROCESSES

- **Multivariate normal prior over functions**
- $(f(x_1), \dots, f(x_n)) \sim \mathcal{N}(m(X), K(X, X'))$
- Correlation induced through Kernel K and determined by correlation length ℓ
- **Condition on data**
- $p(f_* | \mathbf{y}) = \mathcal{N}(\mu_*, \sigma_*^2)$
- **Learn hyperparameters**
- $\theta^* = \arg \max_{\theta} \log p(\mathbf{y} | \theta)$
- GP effectively an interpolator (for large enough training space)

Unconditioned GP (prior)

$\ell = 1$

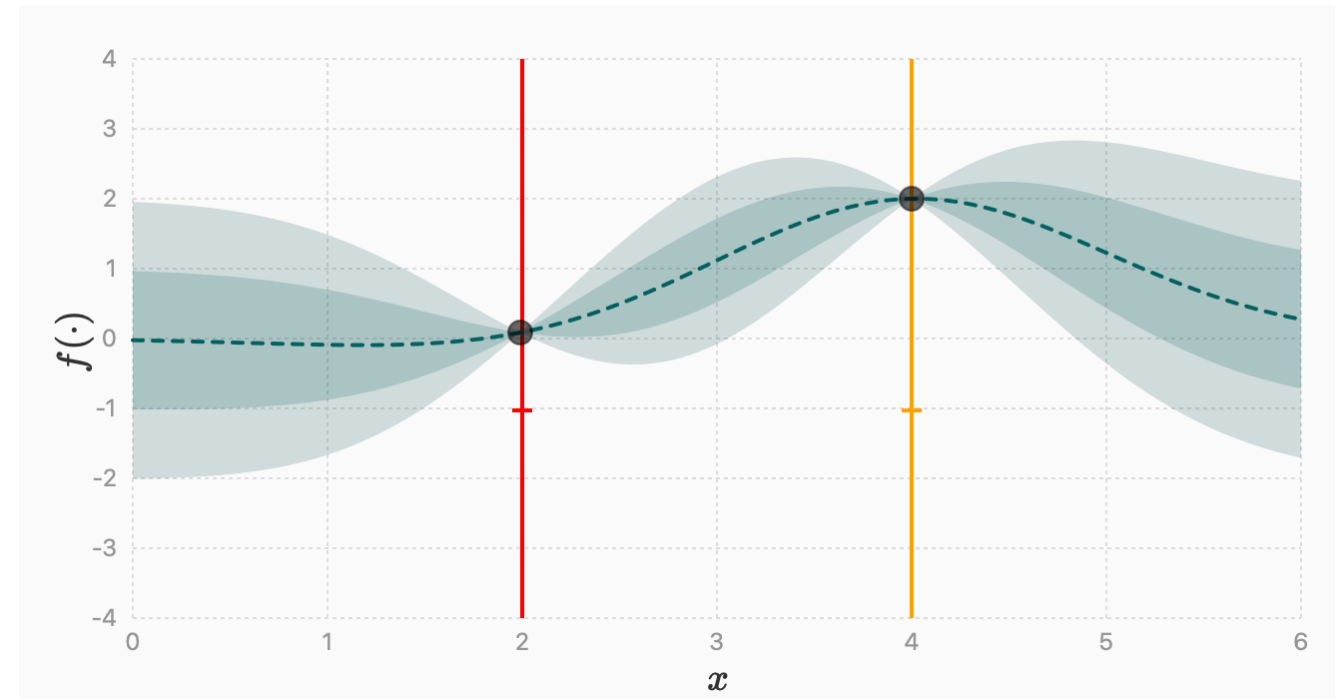


<https://www.infinitecuriosity.org/vizgp/>

GAUSSIAN PROCESSES

- **Multivariate normal prior over functions**
- $(f(x_1), \dots, f(x_n)) \sim \mathcal{N}(m(X), K(X, X'))$
- Correlation induced through Kernel K and determined by correlation length ℓ
- **Condition on data**
- $p(f_* | \mathbf{y}) = \mathcal{N}(\mu_*, \sigma_*^2)$
- **Learn hyperparameters**
- $\theta^* = \arg \max_{\theta} \log p(\mathbf{y} | \theta)$
- GP effectively an interpolator (for large enough training space)

Conditioned GP (posterior) $\ell = 1$

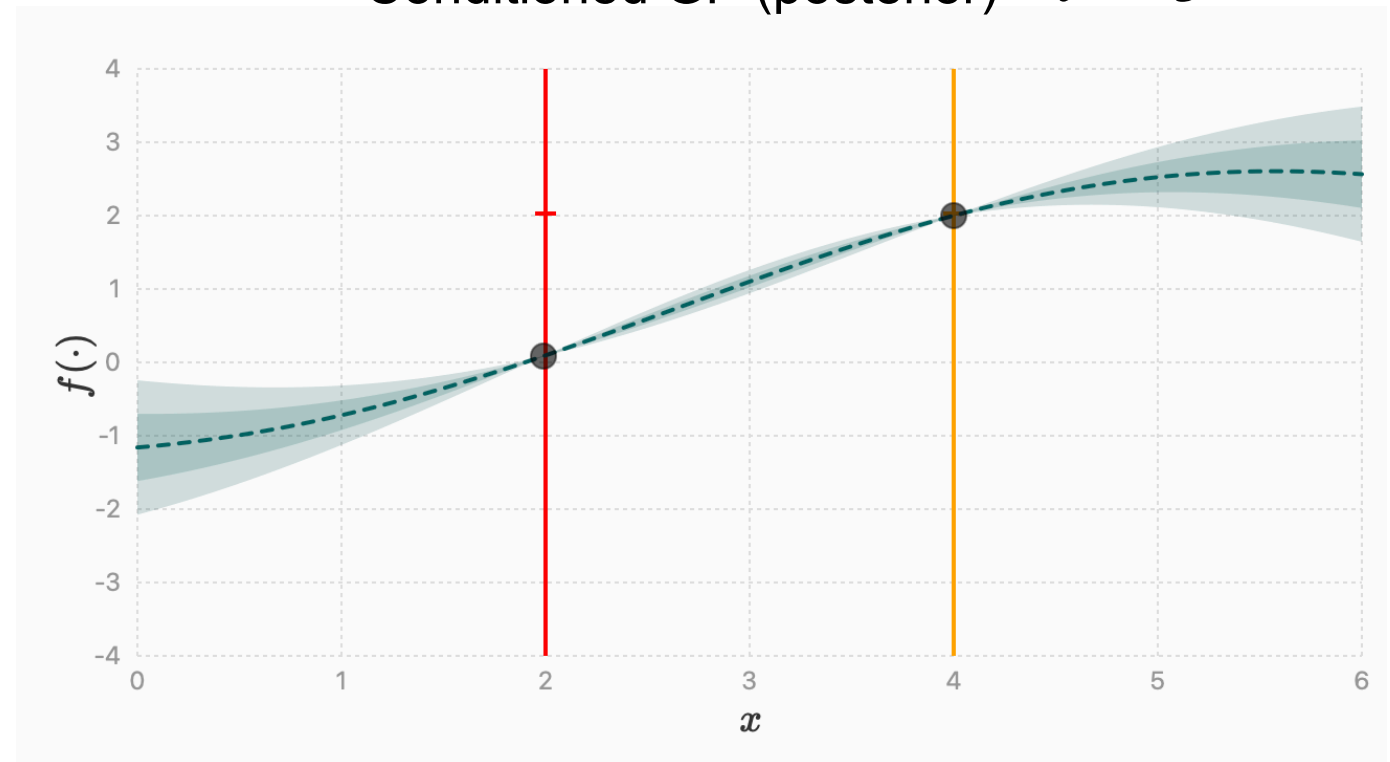


<https://www.infinitecuriosity.org/vizgp/>

GAUSSIAN PROCESSES

- **Multivariate normal prior over functions**
- $(f(x_1), \dots, f(x_n)) \sim \mathcal{N}(m(X), K(X, X'))$
- Correlation induced through Kernel K and determined by correlation length ℓ
- **Condition on data**
- $p(f_* | \mathbf{y}) = \mathcal{N}(\mu_*, \sigma_*^2)$
- **Learn hyperparameters**
- $\theta^* = \arg \max_{\theta} \log p(\mathbf{y} | \theta)$
- GP effectively an interpolator (for large enough training space)

Conditioned GP (posterior) $\ell = 3$



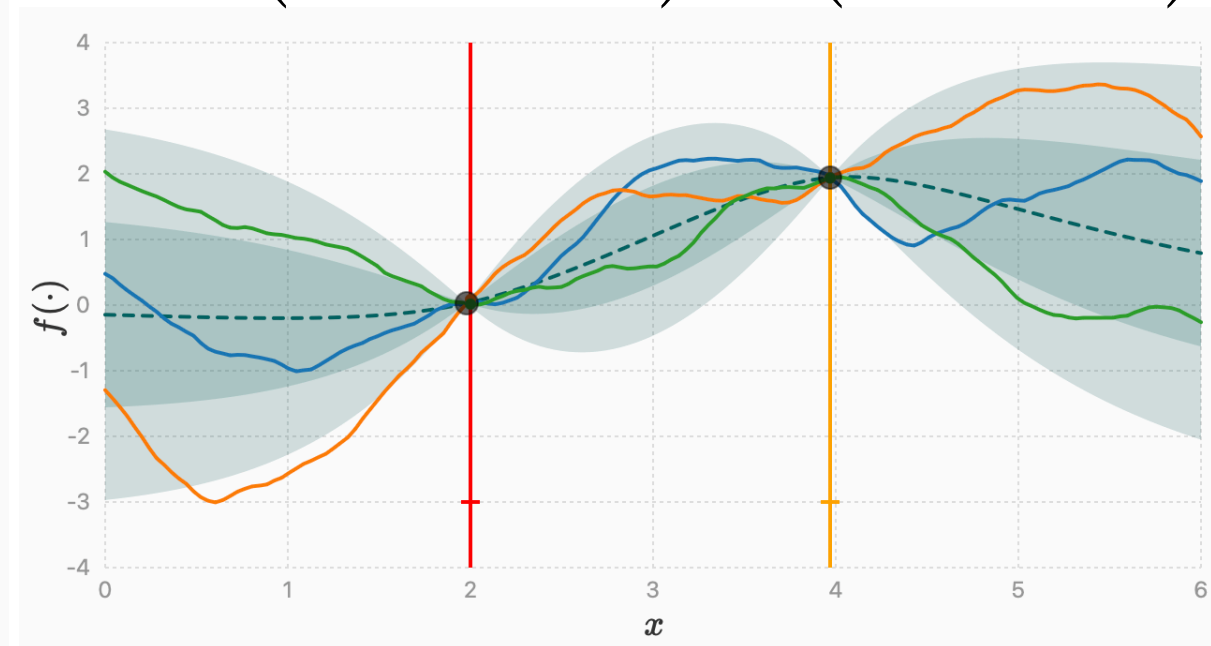
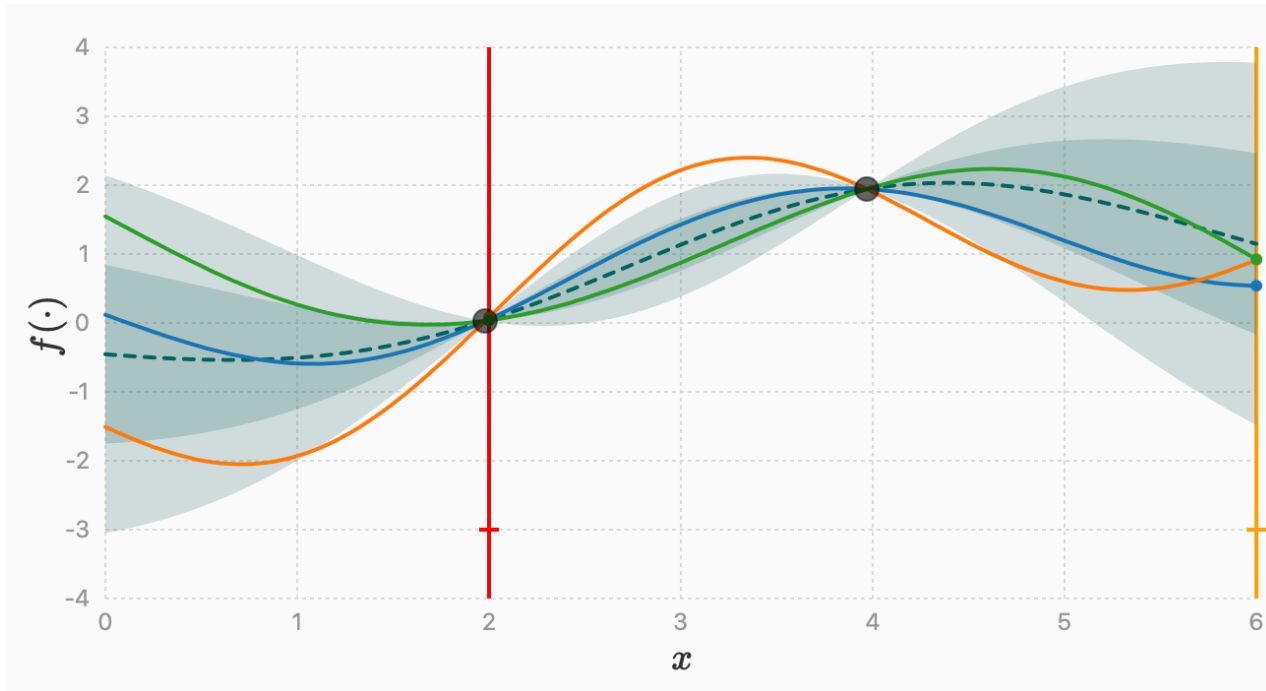
<https://www.infinitecuriosity.org/vizgp/>

KERNELS

- Squared-exponential kernel vs Matern 3/2

$$K(x, x') = \sigma^2 \exp \left(-(x - x')^T(x - x')/2\ell^2 \right)$$

$$K(x, x') = \sigma^2 \left(1 + \frac{\sqrt{3} |x - x'|}{\ell} \right) \exp \left(-\frac{\sqrt{3} |x - x'|}{\ell} \right)$$



<https://www.infinitecuriosity.org/vizgp/>

MULTIDIMENSIONAL INPUT GAUSSIAN PROCESSES

- Example: Variation of 2 LECs \longrightarrow Prediction of binding energies

$$K(\mathbf{x}, \mathbf{x}') = \bar{c}^2 \exp\left(-(\mathbf{x} - \mathbf{x}')^T L^{-1}(\mathbf{x} - \mathbf{x}')\right) \quad \text{with } \mathbf{x} = (c_D, c_E)$$

$$L = \begin{pmatrix} \ell_{c_D}^2 & \ell_{c_D, c_E}^2 \\ \ell_{c_E, c_D}^2 & \ell_{c_E}^2 \end{pmatrix}$$

Number of ℓ parameters scales as $\frac{n_{\text{LECs}} \times (n_{\text{LECs}} - 1)}{2}$

MULTIDIMENSIONAL INPUT GAUSSIAN PROCESSES

- Example: Variation of 2 LECs \longrightarrow Prediction of binding energies

$$K(\mathbf{x}, \mathbf{x}') = \bar{c}^2 \exp\left(-(\mathbf{x} - \mathbf{x}')^T L^{-1}(\mathbf{x} - \mathbf{x}')\right) \quad \text{with } \mathbf{x} = (c_D, c_E)$$

$$L = \begin{pmatrix} \ell_{c_D}^2 & \ell_{c_D, c_E}^2 \\ \ell_{c_E, c_D}^2 & \ell_{c_E}^2 \end{pmatrix} \longrightarrow L' = \begin{pmatrix} \ell_A^2 & 0 \\ 0 & \ell_B^2 \end{pmatrix}$$

Number of ℓ parameters scales as $\frac{n_{\text{LECs}} \times (n_{\text{LECs}} - 1)}{2}$

Goal: Transform parameter space into orthogonal parameter directions

ACTIVE SUBSPACE LEARNING

Constantine et al., SIAM J. Sci. Comput. 36 (2014)

- **Goal: Identify directions in LEC space along which the observable is most sensitive**

$$C = \mathbb{E}[\nabla f(x) \nabla f(x)^T]$$

$$C = W\Lambda W^T$$

$$C \approx \tilde{C} = \frac{1}{M} \sum_{j=1}^M (\nabla_x f(x_j)) (\nabla_x f(x_j))^T$$

- Have to estimate the gradient for each sample

$$\tilde{C} = GG^T \quad G = \frac{1}{\sqrt{M}} [\nabla_x f_1 \quad \dots \quad \nabla_x f_M]$$

$$\text{SVD: } G = \tilde{W} \sqrt{\tilde{\Lambda}} V^T$$

- Divide into active and inactive directions

$$\mathbf{x} = W_1 \mathbf{y} + W_2 \mathbf{z}$$

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \quad W = [W_1 \quad W_2]$$

- Transform input into active subspace

$$\mathbf{y} = \tilde{W}_1^T \cdot \mathbf{x}$$

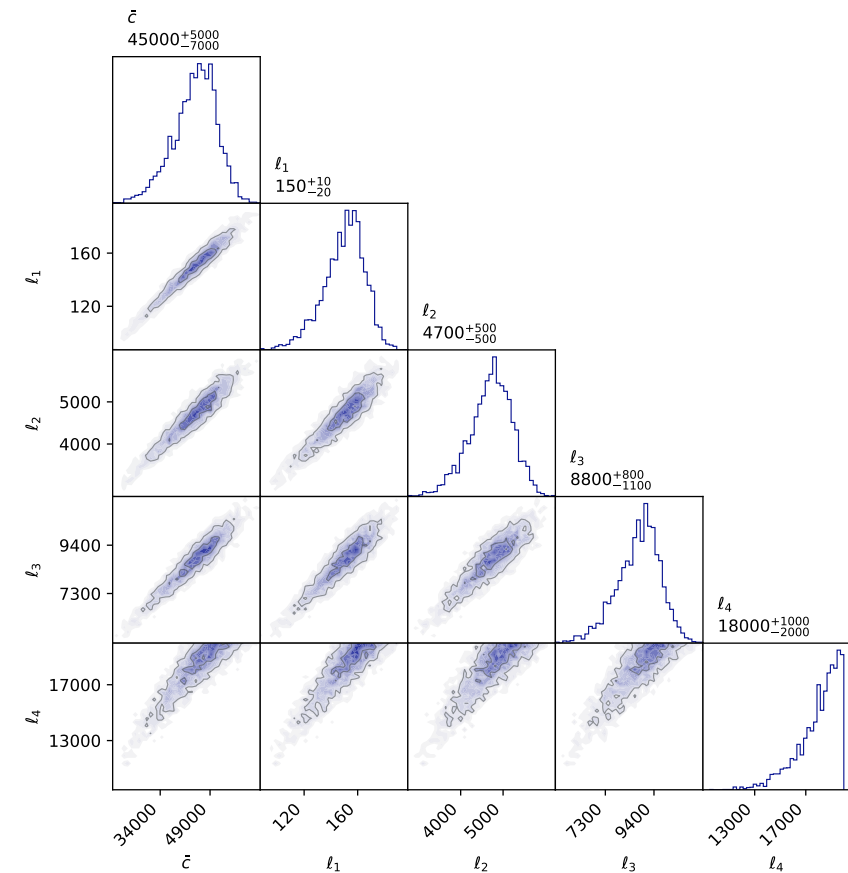
- Discard inactive directions

$$\mathbf{z} = \tilde{W}_2^T \cdot \mathbf{x}$$

APPLICATION TO BINDING ENERGIES

Companys et al., arxiv:2510.08362 (2025)

- ^{48}Ca binding energies from Hartree-Fock emulator with 4 LECs : $\mathbf{x} = (c_3, c_4, c_D, c_E)$, $f(\mathbf{x}) = E_{\text{gs}}$
- Transform data into active subspace: $\mathbf{y} = \tilde{W}_1^T \cdot \mathbf{x}$
- Sample hyperparameters: $\bar{c}, \ell_1, \ell_2, \ell_3, \ell_4$ with MCMC
- Use maximum a posteriori (MAP) parameter values
- Higher-order directions could potentially be discarded (indicated by large ℓ)

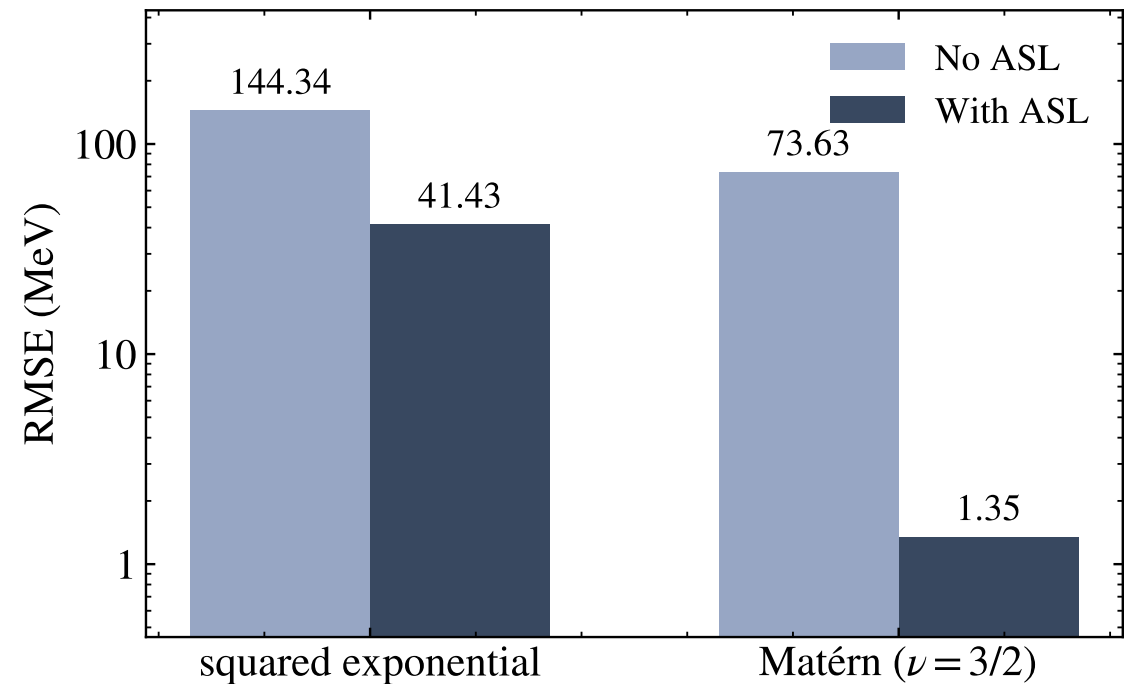


RESULTS

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{\text{GP},i})^2}$$

- Active subspace learning (ASL) and Matern instead of squared-exponential kernel combined: 100x smaller RMSE
- Overall scale: $E_{\text{HF}}(^{48}\text{Ca}) \sim -300 \text{ MeV}$
- Good average performance with few outliers
- Identify source of outliers
- Saturation at around 800 training points with $\sim 0.1\%$ deviation
- Poor empirical coverage \rightarrow overestimated uncertainties

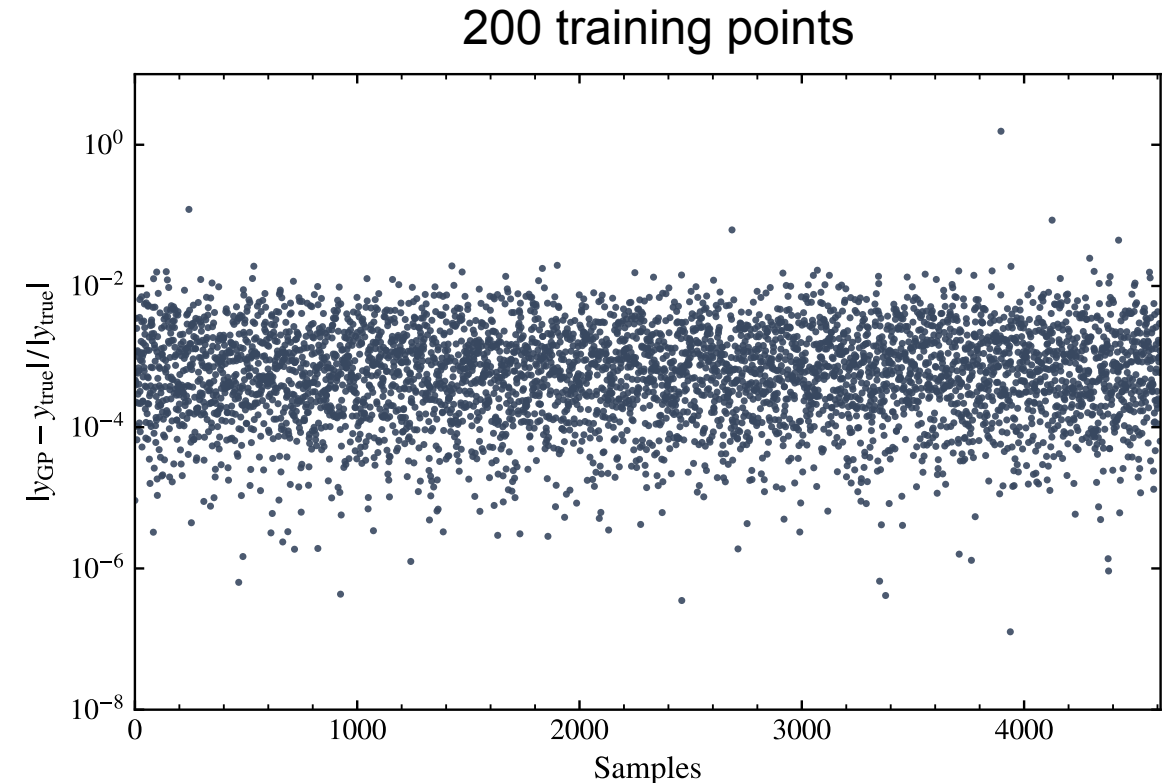
200 training points



RESULTS

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{\text{GP},i})^2}$$

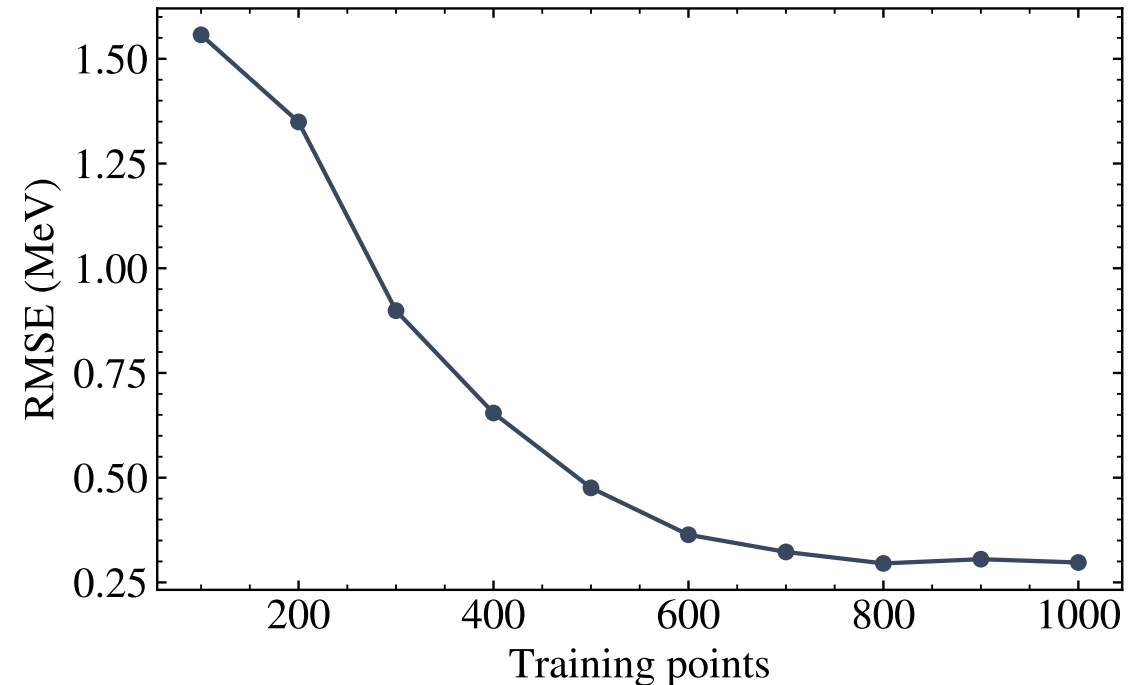
- Active subspace learning (ASL) and Matern instead of squared-exponential kernel combined: 100x smaller RMSE
- Overall scale: $E_{\text{HF}}(^{48}\text{Ca}) \sim -300 \text{ MeV}$
- Good average performance with few outliers
- Identify source of outliers
- Saturation at around 800 training points with $\sim 0.1\%$ deviation
- Poor empirical coverage \rightarrow overestimated uncertainties



RESULTS

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{\text{GP},i})^2}$$

- Active subspace learning (ASL) and Matern instead of squared-exponential kernel combined: 100x smaller RMSE
- Overall scale: $E_{\text{HF}}(^{48}\text{Ca}) \sim -300 \text{ MeV}$
- Good average performance with few outliers
- Identify source of outliers
- Saturation at around 800 training points with $\sim 0.1\%$ deviation
- Poor empirical coverage \rightarrow overestimated uncertainties



OUTLOOK

- Built wave-function-free multi-input emulator
- Extend to include NN parameters
- Potentially discard inactive directions
- Optimize selection of training points
- **Apply to IMSRG**

OUTLOOK

- Built wave-function-free multi-input emulator
- Extend to include NN parameters
- Potentially discard inactive directions
- Optimize selection of training points
- Apply to IMSRG



Gaussian Process Emulators with Active Subspace Learning

Tom Plies with Margarida Compagny and Achim Schwenk



Motivation

- Bayesian LEC inference could be improved through medium-mass constraints [1]
- Full IMSRG calculations are significantly too expensive
- No fast IMSRG emulators yet
- Emulators can be constructed based on Gaussian Processes (GPs) [2]
- Default GP emulators are too inefficient to approximate IMSRG

Goal: Emulator for nuclear structure observables for improved LEC inference with Gaussian Processes

- Optimize Gaussian Process through transformation of the parameter space and kernel choice

Gaussian Processes

Probability distribution over functions such that any finite set of functions is jointly Gaussian [3]

$$(f(x_1), \dots, f(x_n)) \sim \mathcal{N}(m(X), K(X, X'))$$

- Correlation induced through covariance kernel K and determined by L

Squared-exponential kernel:
 $K_{SE}(x, x') = \sigma^2 \exp(-r^2)$

Matérn ($\nu = 3/2$) kernel:
 $K_{SE}(x, x') = \sigma^2 (1 + \sqrt{3}r) \exp(-\sqrt{3}r) \quad r = \sqrt{(x-x')^T L^{-1}(x-x')}$

Usual assumption for multidimensional input data:

- Principal axes of variation align with parameter axes
- No cross-coupling between parameters

$$L = \text{diag}\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_D^2}\right)$$

- Full L matrix is more accurate but requires too many new parameters
- Instead: transform input space to make L approximately diagonal

- Prior over functions**
 $f \sim \mathcal{GP}(m, k)$
- Condition on data**
 $p(f, L | y) = \mathcal{N}(y | \mu, \sigma^2)$
- Learn hyperparameters**
 $\theta^* = \arg \max_{\theta} \log p(y | \theta)$

Active subspace learning (ASL)

Goal: Identify directions in LEC space along which the observable is most sensitive [5]

- Find average global gradient to identify active directions of data

$$C = E[\nabla f(x) \nabla f(x)^T]$$

$$C = W \Lambda W^T$$

- Approximate C with finite points:

$$C \approx \hat{C} = \frac{1}{M} \sum_{j=1}^M (\nabla_x f(x_j)) (\nabla_x f(x_j))^T$$

- Singular value decomposition

$$G = \hat{C} \Lambda^{-1/2} V^T$$

Divide into active and inactive directions

$$x = W_1 y + W_2 z$$

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \quad W = [W_1 \quad W_2]$$

- Transform input into active subspace $y = W_1^T x$
- Discard inactive directions $z = W_2^T x$

Estimate gradients through nearest neighbours

$$f(x) \approx f(x_i) + \nabla f(x_i)^T (x - x_i)$$

$$\begin{pmatrix} (x_1 - x_i)^T \\ \vdots \\ (x_k - x_i)^T \end{pmatrix} \nabla f(x_i) \approx \begin{pmatrix} f(x_1) - f(x_i) \\ \vdots \\ f(x_k) - f(x_i) \end{pmatrix}$$

$$A_i z \approx b$$

Minimize residuals $g^* = \arg \min_g \|A_i g - b\|_2$

⁴⁸Ca Hartree-Fock results

- Dataset: 10,000 samples of $E_{\text{gr}}^{\text{HF}}$ for variations of $(c_2, c_3, c_4, c_D, c_E)$ [6]
- 200-1000 training points, 5000 test points, 15 nearest neighbours for local gradients
- Sample posterior with MCMC and use maximum a posteriori values for hyperparameters $(\ell, \ell_1, \ell_2, \ell_3, \ell_4)$
- Optimal correlation lengths ordered by magnitude after ASL transformation

Parameter	Range
c_2	$[-6.2, -0.2] \text{ GeV}^{-1}$
c_3	$[0.4, 10.0] \text{ GeV}^{-1}$
c_4	$[-5.0, 10.0]$
c_D	$[-2.0, 2.0]$

- First active subspace direction almost identical for radii and energies

$$w_{1,LE} = \begin{pmatrix} -0.4 \\ -0.19 \\ 0.23 \\ -0.86 \end{pmatrix} \quad w_{1,R} = \begin{pmatrix} -0.38 \\ -0.17 \\ 0.22 \\ -0.88 \end{pmatrix}$$

Median: $E_{\text{gr}}^{\text{HF}}(^{48}\text{Ca}) \sim -272 \text{ MeV}$

Median: $R_{\text{gr}}^{\text{HF}}(^{48}\text{Ca}) \sim 3.37 \text{ fm}$

RMSE = $\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$

- Matérn significantly more effective than squared-exponential
- Up to 50x improvement in RMSE with ASL
- Radii more accurate than energies

Results converge at ~800 training points to 0.1% error

Outlook

So far: wave-function free 4-dimensional input Hartree-Fock emulator for energies and radii with $\sim 0.5\%$ error for 200 training points

- Include NN parameters
- Apply to IMSRG results
- Investigate training point convergence in higher dimensions
- Explore importance of higher-order directions
- Optimize training point selection

References

[1] Waselowski, Svensson, Ekström, Fornäs, Furnstahl, Melendez, Phillips, *PRC* **104** (2021)

[2] Higdon, McDonnell, Scurlock, Sarich, *Wid. J. Phys.* **G 42** (2019)

[3] Rasmussen, Williams *MIT Press*, Cambridge (2006)

[4] <https://www.infocoursecity.org/cgpf/>

[5] Constantine, Dow, Wang, *SIAM J. Sci. Comput.* **36**, A1500 (2014)

[6] Compagny, Tolm, Heide, Schwenk, *arxiv:22.10.08338* (2022)

Thank you for your attention

LOCAL GRADIENTS

- Assume approximately local linear behavior: $f(\mathbf{x}_j) \approx f(\mathbf{x}_i) + \nabla f(\mathbf{x}_i)^T (\mathbf{x}_j - \mathbf{x}_i)$

For one point:

$$\begin{pmatrix} x_{j1} - x_{i1} \\ x_{j2} - x_{i2} \end{pmatrix} \cdot \nabla f(\mathbf{x}_i) = f(\mathbf{x}_j) - f(\mathbf{x}_i) \quad \nabla f(\mathbf{x}_i) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_i) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}_i) \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$

For k points:

$$\begin{pmatrix} (\mathbf{x}_1 - \mathbf{x}_i)^T \\ \vdots \\ (\mathbf{x}_k - \mathbf{x}_i)^T \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \approx \begin{pmatrix} f(\mathbf{x}_1) - f(\mathbf{x}_i) \\ \vdots \\ f(\mathbf{x}_k) - f(\mathbf{x}_i) \end{pmatrix} \quad \xrightarrow{\text{Residual } r = Ag - b}$$

$Ag \approx b$

Minimize residuals $g^* = \arg \min_g \|Ag - b\|_2^2$

- Compute differences in f and \mathbf{x} between all **k nearest neighbours** of x_i with $j \in \{1, \dots, k\}$
- Minimize residuals with least-square fit to find compatible common gradient $\nabla f(\mathbf{x}_i)$

GAUSSIAN PROCESSES

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')), \quad m(x) = \mathbb{E}[f(x)], \quad k(x, x') = \text{Cov}(f(x), f(x')).$$

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{f}_* \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} K + \sigma_n^2 I & K_*^\top \\ K_* & K_{**} \end{bmatrix}\right),$$

$$\mathbf{f}_* | X, \mathbf{y}, X_* \sim \mathcal{N}(\bar{\mathbf{f}}_*, \text{Cov}(\mathbf{f}_*)),$$

$$\bar{\mathbf{f}}_* = \mathbf{m}_* + K_* (K + \sigma_n^2 I)^{-1} (\mathbf{y} - \mathbf{m}),$$

$$\text{Cov}(\mathbf{f}_*) = K_{**} - K_* (K + \sigma_n^2 I)^{-1} K_*^\top,$$

$$[K_*]_{ij} = k(x_i, x_{*j}), \quad [K_{**}]_{ij} = k(x_{*i}, x_{*j}).$$

SENSITIVITY ANALYSIS

- Identify most sensitive LEC directions
- c_E is most important
- Energies and radii have almost identical dominant directions

$$w_{1,E} = \begin{pmatrix} -0.4 \\ -0.19 \\ 0.23 \\ -0.86 \end{pmatrix} \quad w_{1,R} = \begin{pmatrix} -0.38 \\ -0.17 \\ 0.22 \\ -0.88 \end{pmatrix}$$

RADII RESULTS

- $R(^{48}\text{Ca}) \sim 3.37 \text{ fm}$
- 0.1% deviation at 200 training points

