Bayesian Optimization Techniques for Accelerator Control and Characterization

Ryan Roussel

11/29/2023

rroussel@slac.stanford.edu

U.S. DEPARTMENT OF ENERGY | Stanford University

SLAC NATIONAL ACCELERATOR LABORATORY

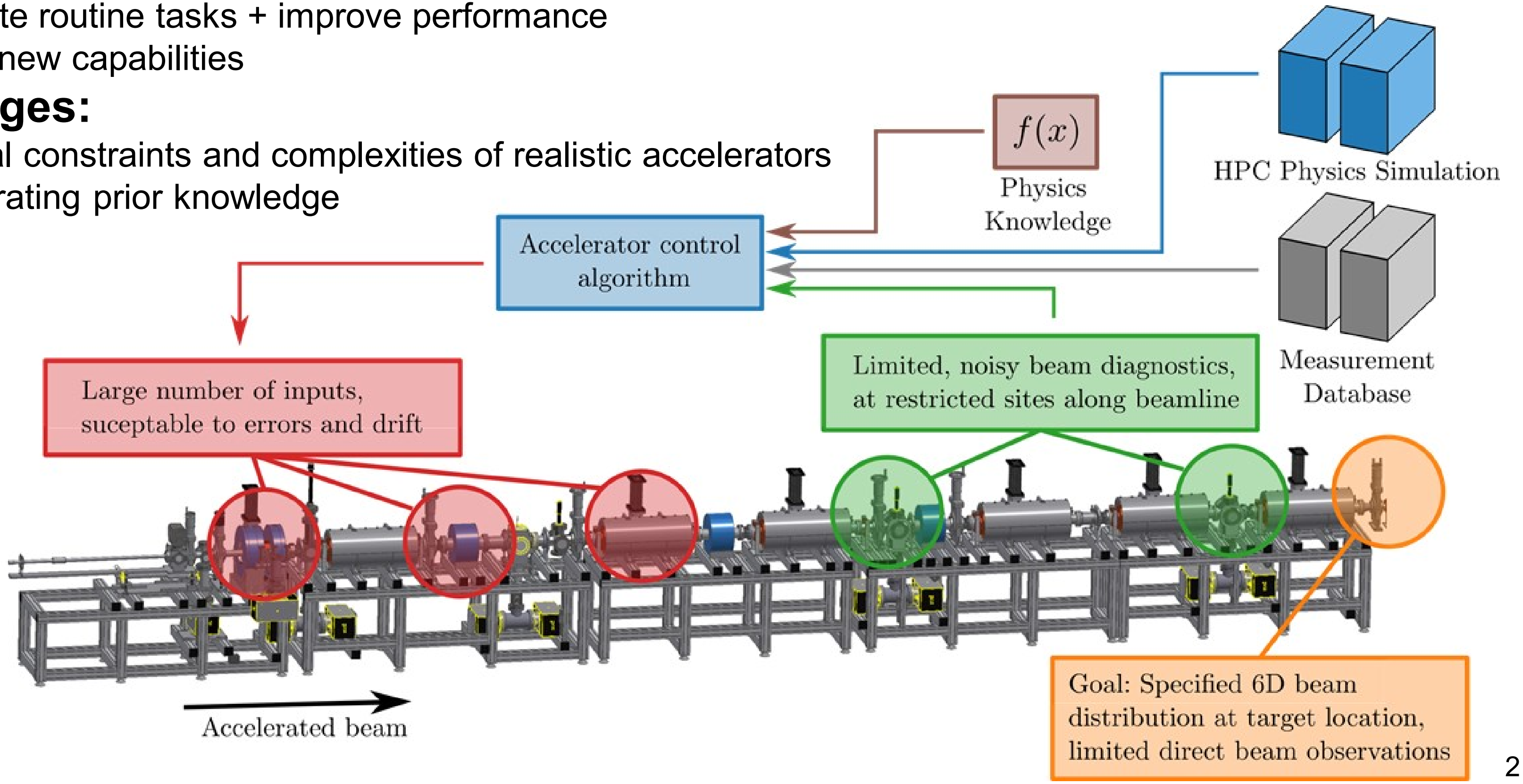
Machine Learning Based Accelerator Control

Goals:

- Automate routine tasks + improve performance
- Enable new capabilities

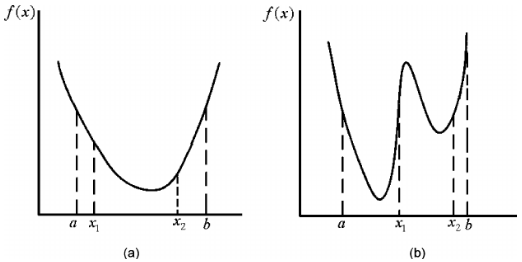
Challenges:

- Practical constraints and complexities of realistic accelerators
- Incorporating prior knowledge
- Scaling

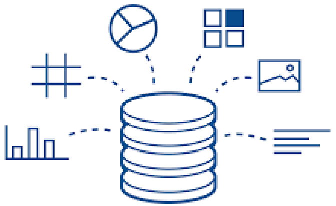


Optimization Considerations

Problem complexity
how difficult is the problem to solve?

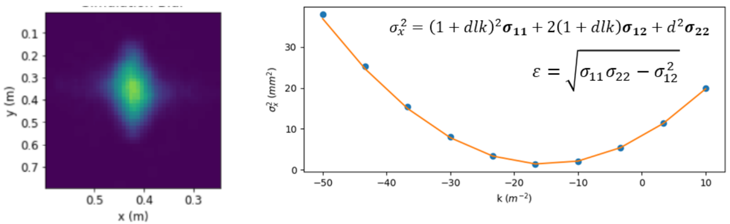


Overhead
how expensive is it to prepare for optimization?



Optimizer cost
how expensive is it to make decisions?

Evaluation cost
how expensive is it to evaluate objectives/constraints?

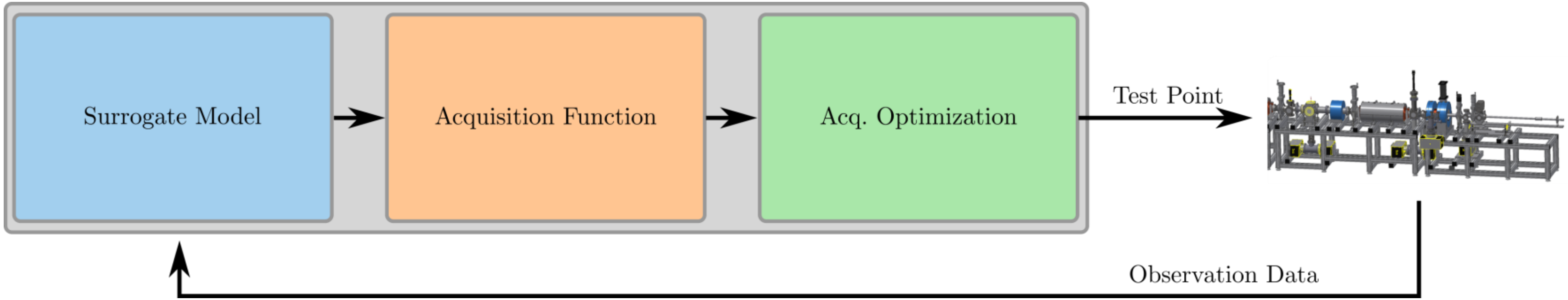


Bayesian Optimization Algorithms

Create a computational model of the system

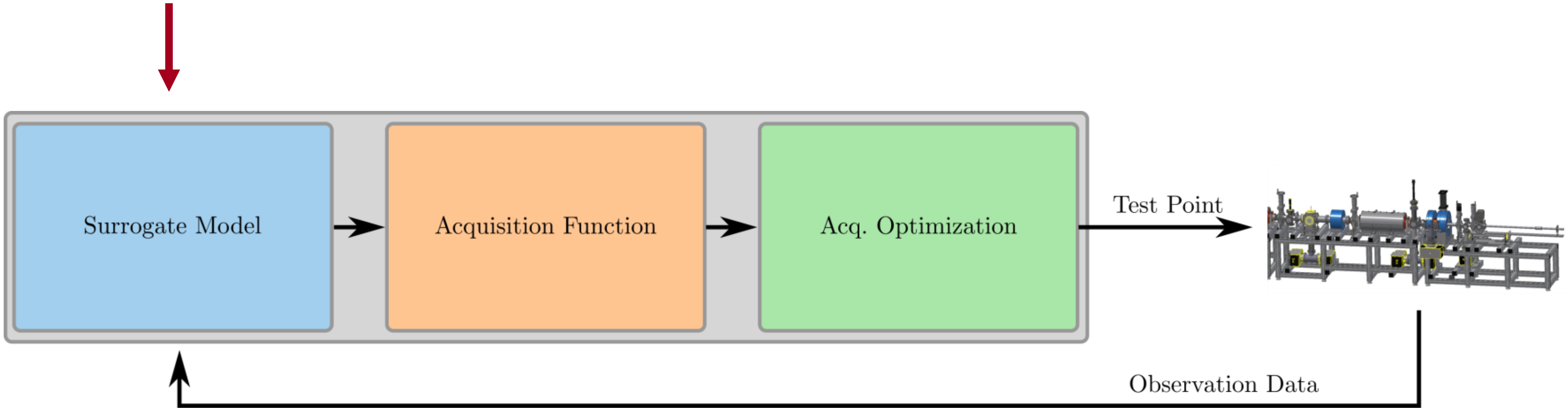
Determine the value of potential future measurements

Pick the point that maximizes value



Gaussian Process Modeling

Gaussian processes (GPs)

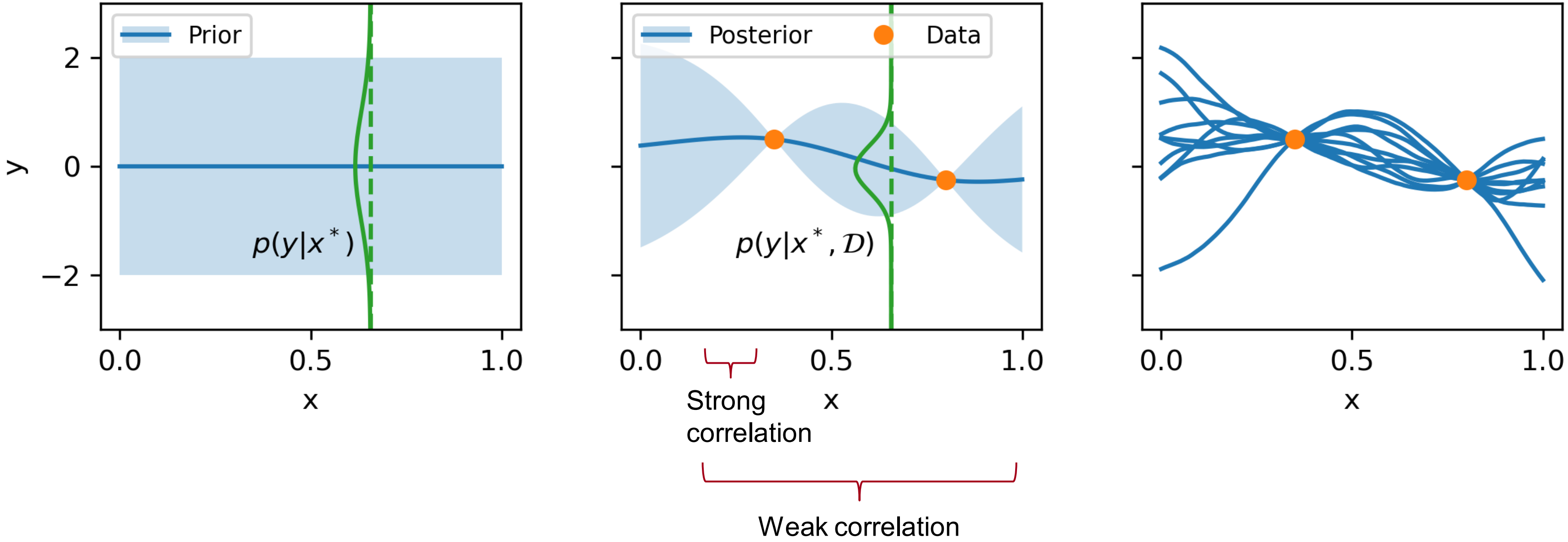


Why?

- Extracts a lot of information from a small number of data points → efficient
- Inherently accounts for noise and sources of uncertainty → ideal for accelerators + global optimization

Gaussian Process Modeling

- Assume a Normal distribution of function values at prediction points x^*
- Use correlations between function values at different locations in input space to make predictions



Fitting Gaussian Processes to Data

We specify a **kernel** that specifies function value covariances at two points x, x' → controls the overall function behavior. It is parameterized by **hyperparameters** which are fit to the data.

Radial Basis Function:

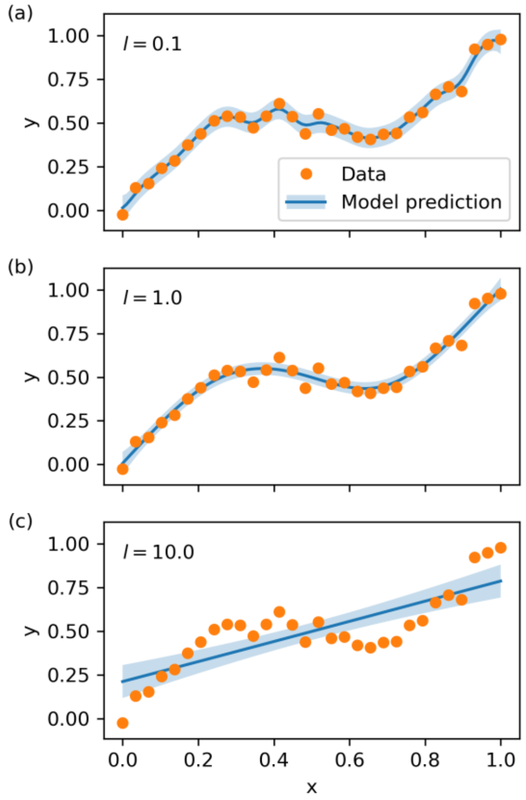
$$k(x, x') = \boldsymbol{\sigma_f^2} \exp\left(-\frac{1}{2\boldsymbol{l}^2}(x - x')^2\right) + \boldsymbol{\sigma_n^2}\delta_{xx'}$$

Kernel amplitude

Kernel length scale

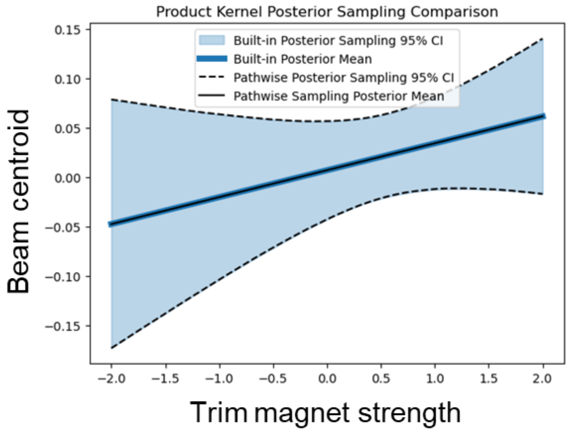
Noise

We learn **low dimensional structure** of the objective function during optimization.

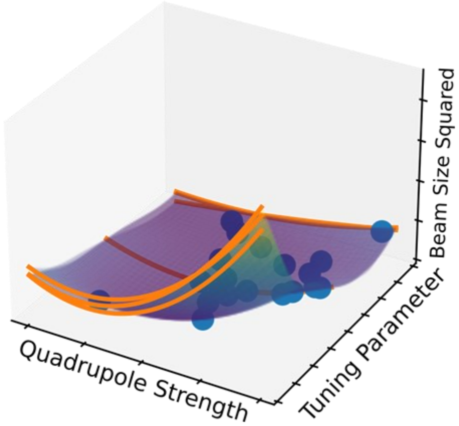


Incorporating Physics Information into Kernels

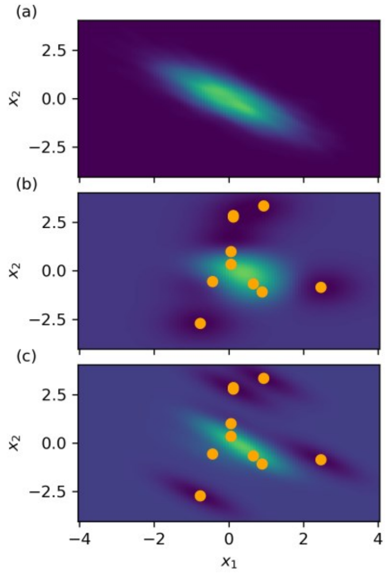
Enforce linear centroid response to steering magnets



Enforce quadratic beam size squared response to quadrupole magnets



Add cross correlations between magnet parameters

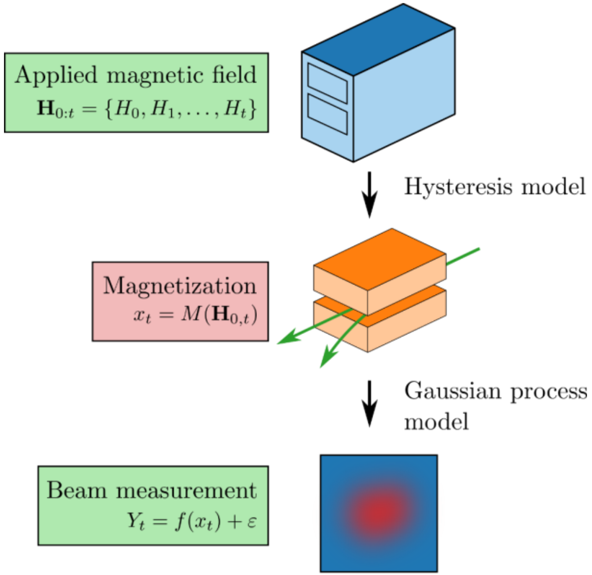


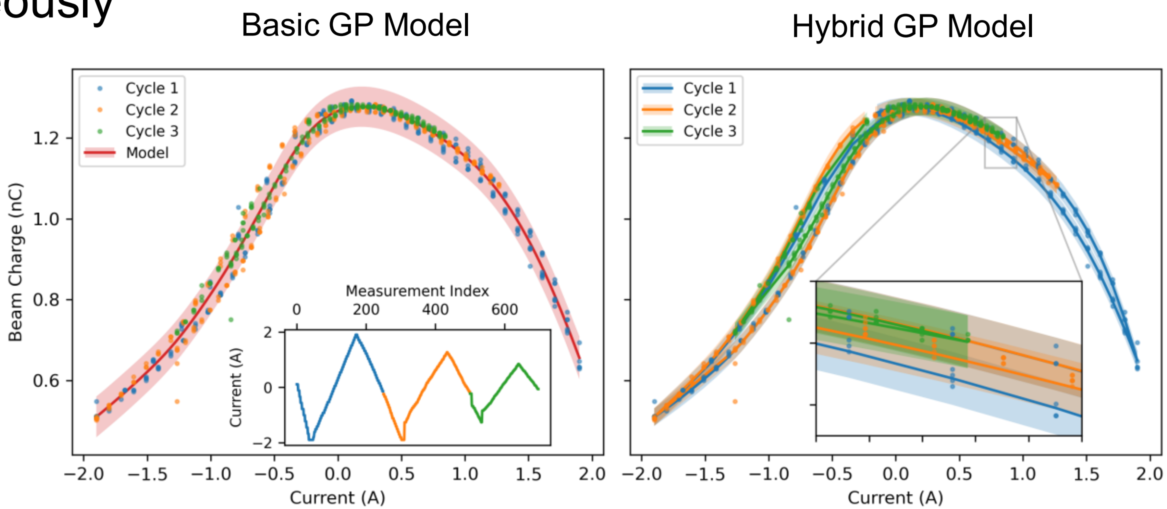
Duris et. al., PRL, 2020

Incorporating physics information into GP models improves accuracy
→ Enables better decision making → faster convergence to optimum

Modeling Complex Physical Processes

Learn both hysteresis properties and beam response simultaneously using two step modeling





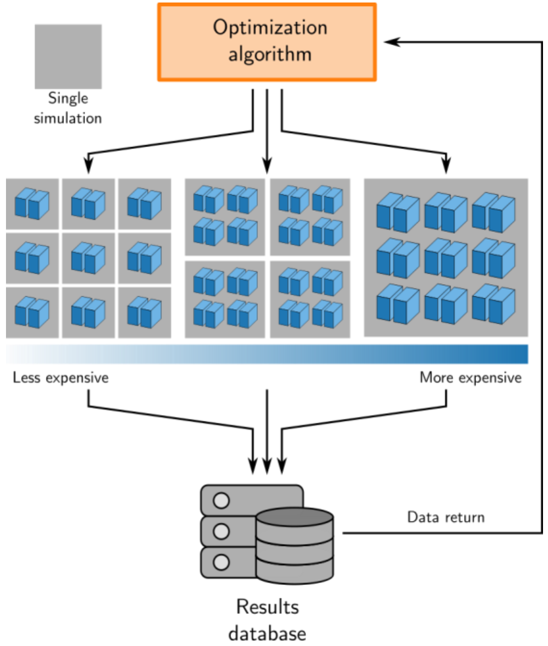
More accurate modeling → improved optimization performance

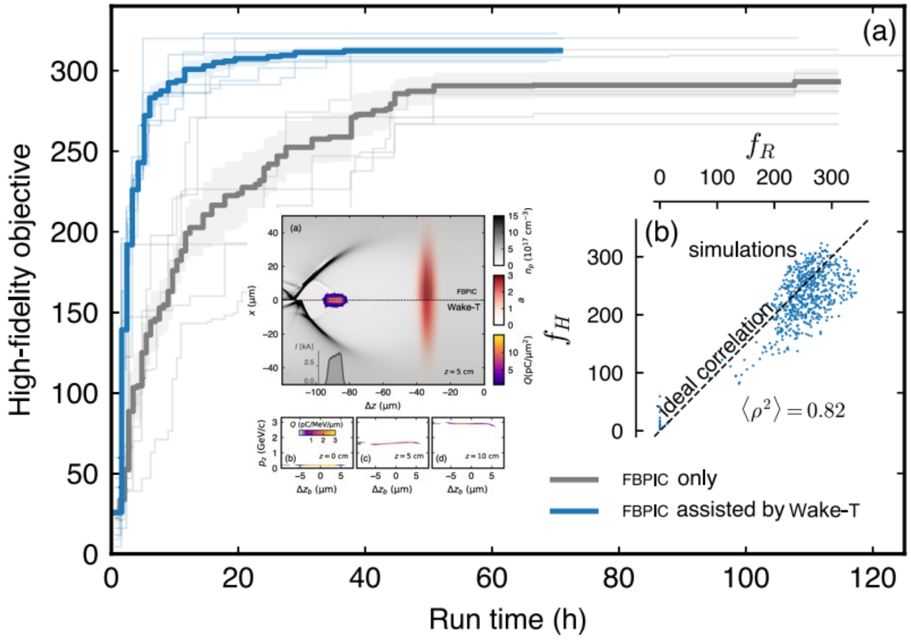
R. Roussel, et. al. Phys. Rev. Lett. **128**, 204801

Multi-Fidelity Modeling

Use **low fidelity approximations** to inform optimization at **high fidelities**

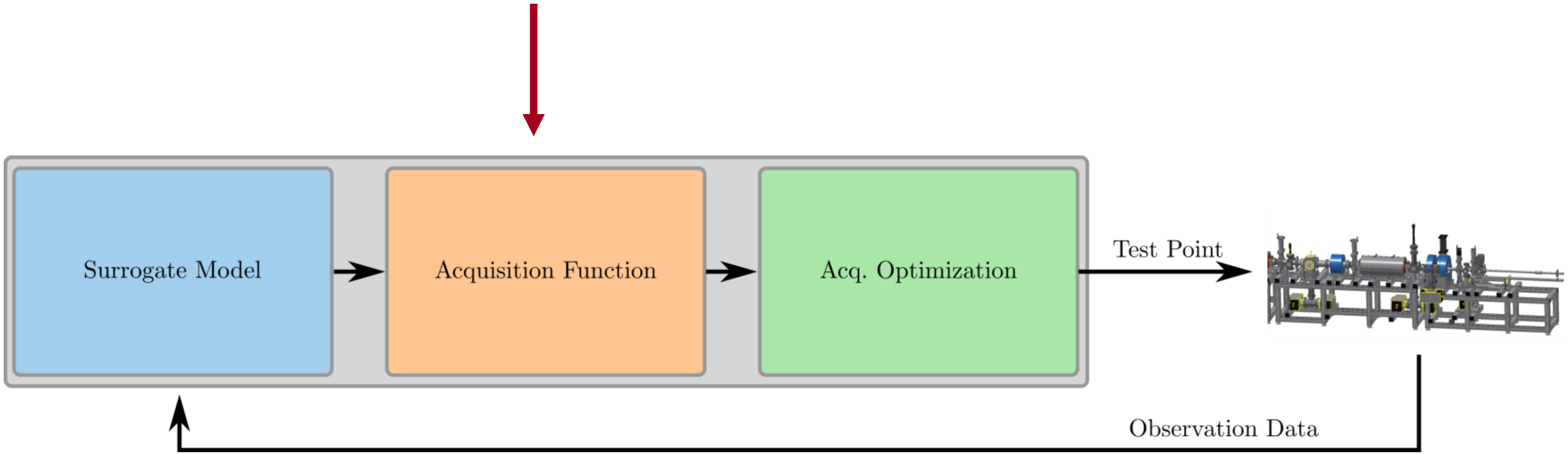
Model kernel: $k(x, x', s, s') = k(x, x') \times k(s, s')$





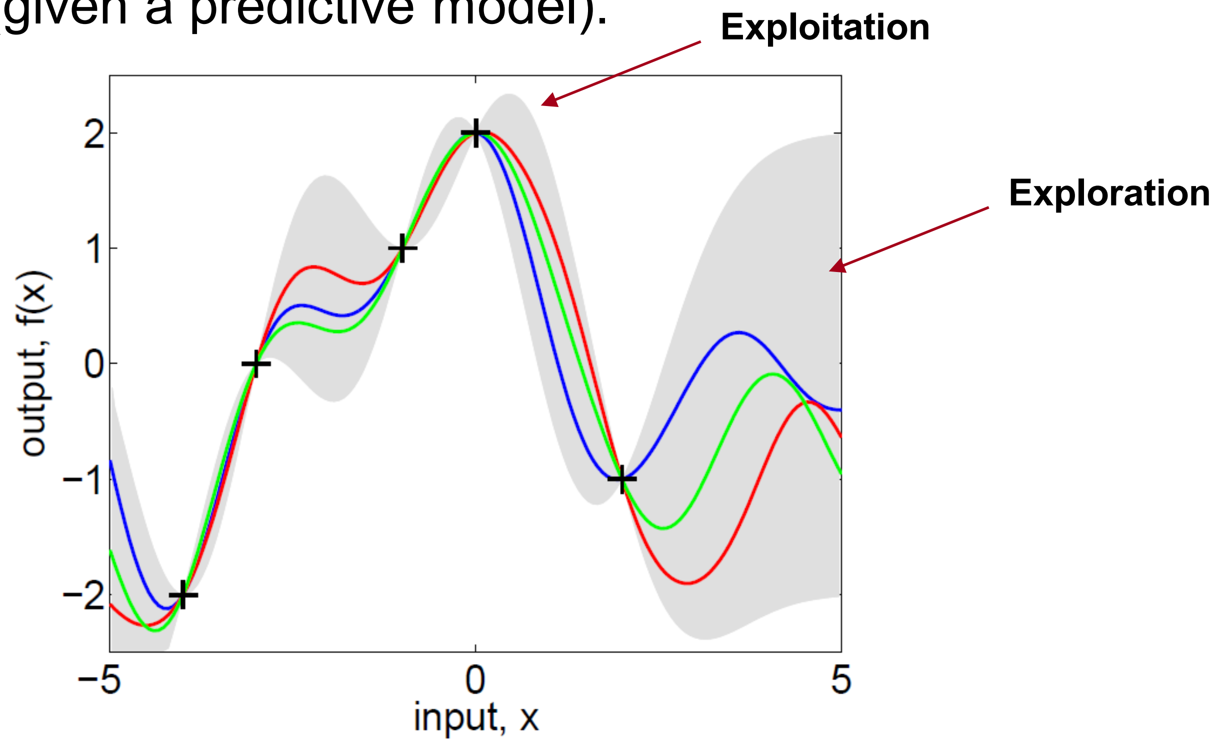
Pousa, A. Ferran, et al. PRAB 26.8 (2023): 084601.

Defining Acquisition Functions



The Acquisition Function

Define a function that characterizes the value of making a potential measurement (given a predictive model).

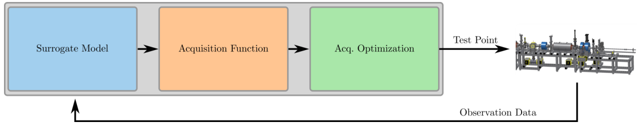


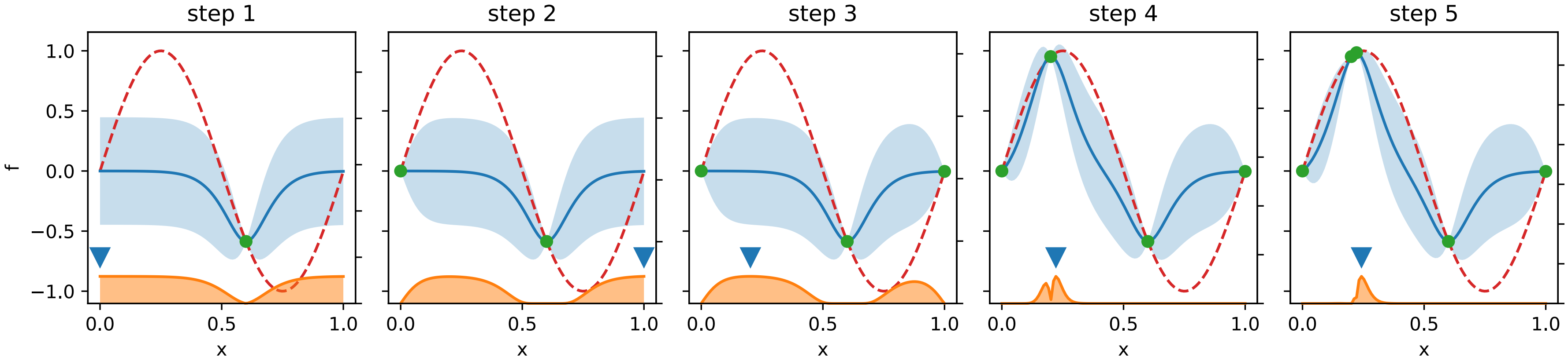
Single Objective Optimization

$$\mathrm{EI}(\boldsymbol{x}) = \mathbb{E}[\max(f(\boldsymbol{x}) - f^*)]$$

$$\boldsymbol{x}_{t+1} = \mathrm{argmax}_{\boldsymbol{x}} \mathrm{EI}(\boldsymbol{x})$$

(Assumes maximization)





Some notes:

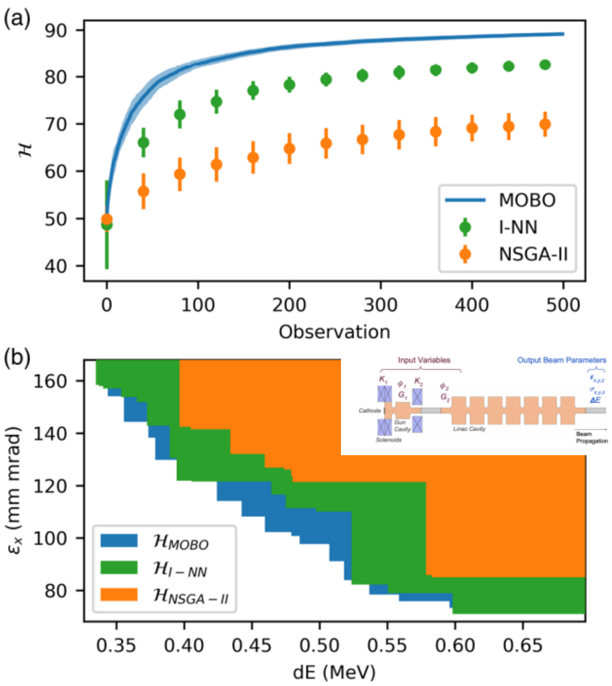
- The model accuracy improves in the region of interest
- Initially the model uncertainty is maximized at the domain boundaries

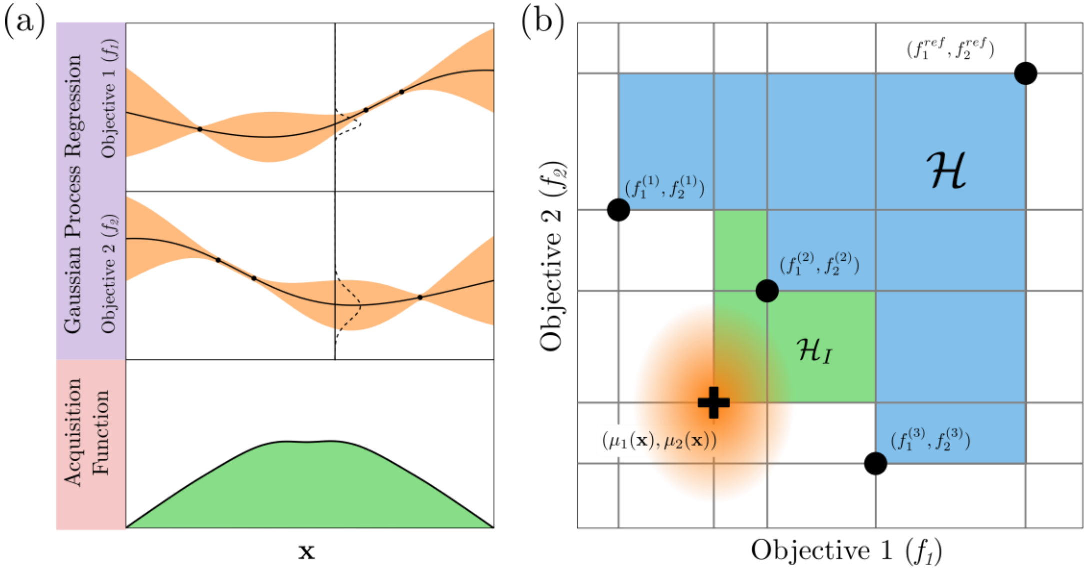
Many examples of this:

- Duris, J. et al. PRL 124.12 (2020): 124801.
- Xu, Chenran, et al. *PRAB* 26.3 (2023): 034601.
- Gao, Y., et al. *PRAB* 25.1 (2022): 014601.
- Miskovich, S. A., et al. *PRAB* 25.4 (2022): 044601.
- and many more…

Multi-Objective Optimization

Determine the optimal trade-off between objectives -> the Pareto front





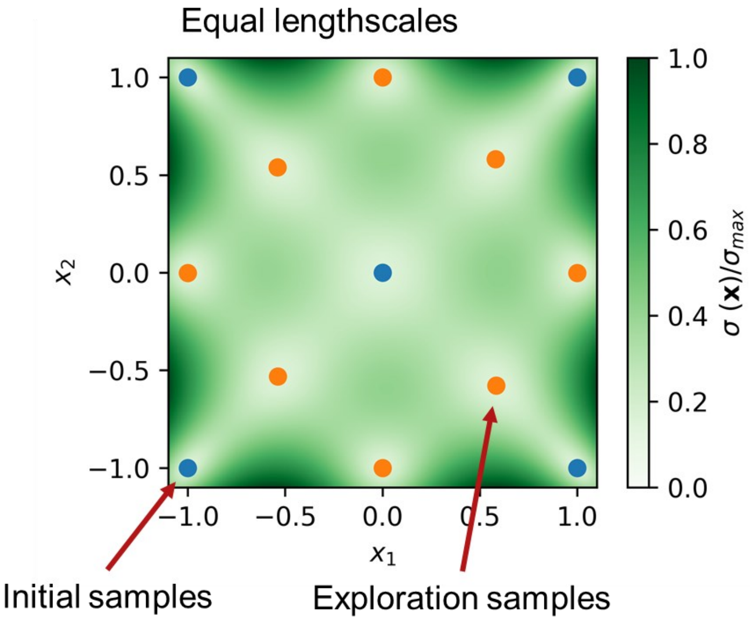
$$\alpha_{EHVI}(\boldsymbol{\mu},\boldsymbol{\sigma},\mathcal{P},\boldsymbol{r}) := \int_{\mathbb{R}^P} \mathrm{HVI}(\mathcal{P},\boldsymbol{y},\boldsymbol{r}) \cdot \xi_{\mu,\boldsymbol{\sigma}}(\boldsymbol{y}) d\boldsymbol{y}$$

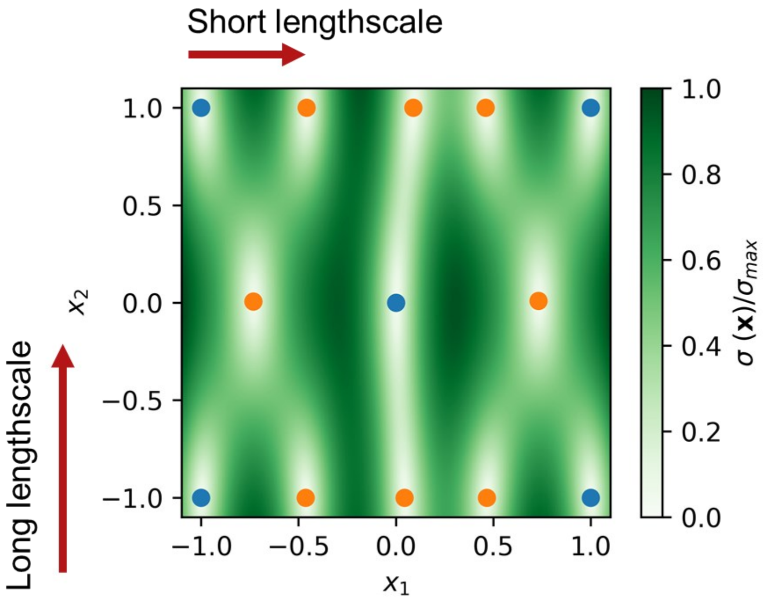
Roussel et. al. PRAB **2021**

Autonomous Characterization – Bayesian Exploration

If the function changes more rapidly along one axis, sample more points along that axis!

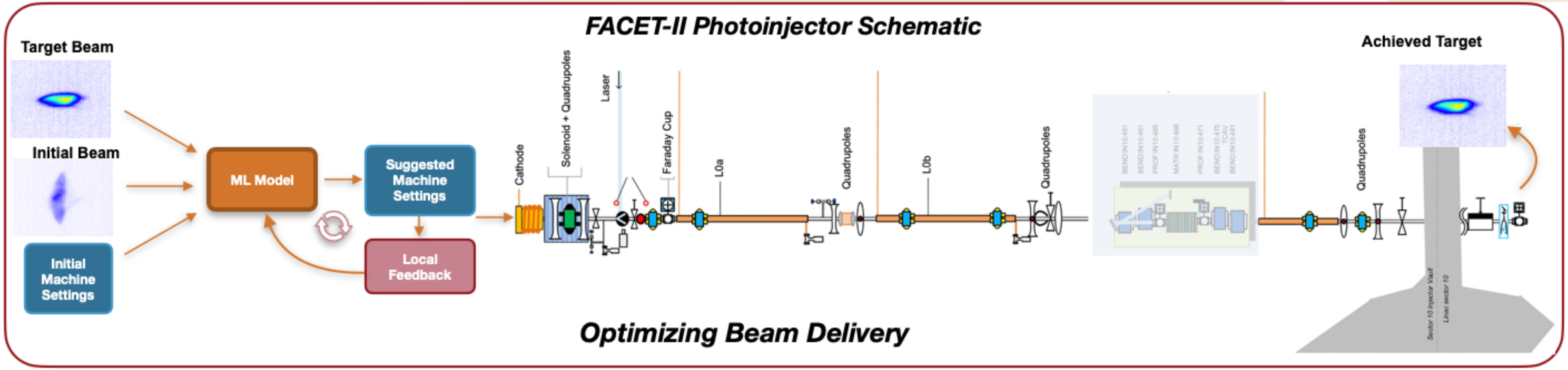
$$\alpha(\boldsymbol{x}) = \sigma(\boldsymbol{x})$$





Roussel et. al. *Nat. Comm.* **2021**

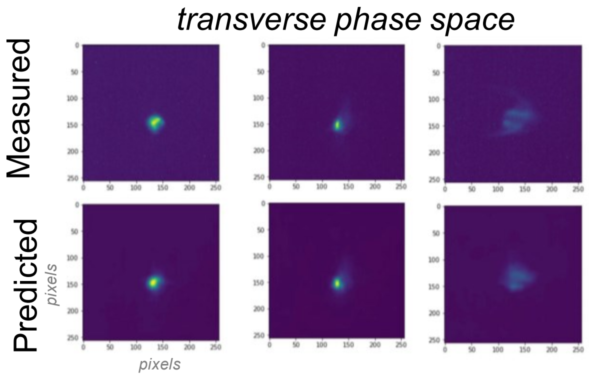
Example: FACET-II Emittance Characterization



Used Bayesian Exploration for efficient high-dimensional characterization (10 variables) at 700pC: **2 hrs for 10 variables compared to 5 hrs for 4 variables with N-D parameter scan**

Data was used to train ML models to predict + optimize beam emittance and injector match

Example of integrated cycle between characterization, modeling, and optimization → now extending to larger system sections and new setups (e.g. two-bunch)

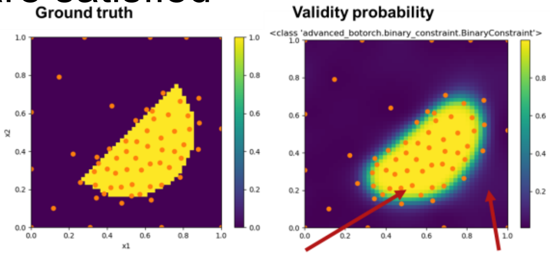


Incorporating Constraints

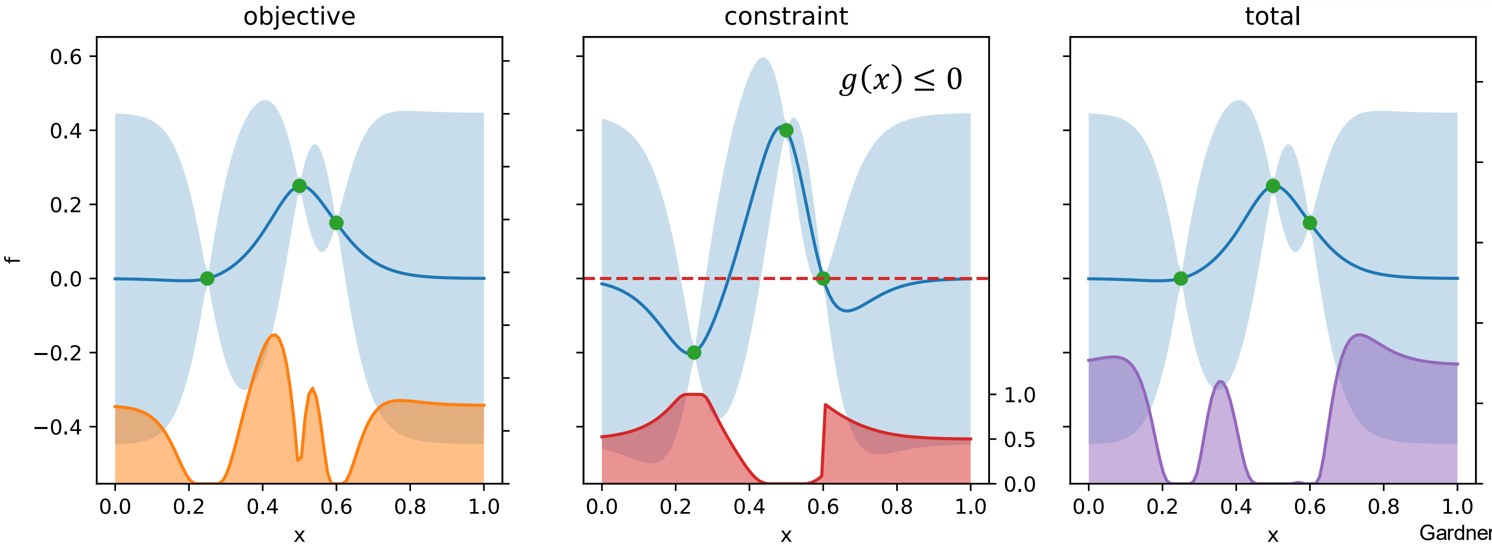
Weight the acquisition function by the probability that constraints are satisfied

$$\hat{\alpha}(x) \rightarrow \alpha(x) \prod_i p[g_i(x) \leq h_i]$$

Warning: Requires $\alpha(x) \geq 0$



2D Example



Gardner et. al. ICML 2014

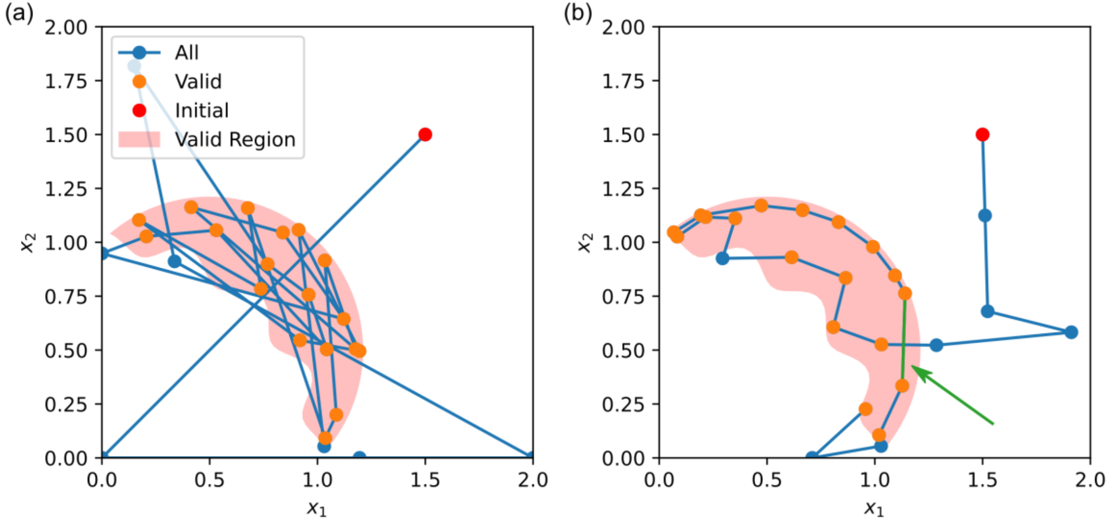
Proximal Biasing

Prevents instabilities during optimization experimental beamlines

Weight the acquisition function by travel distance → better than hard limits

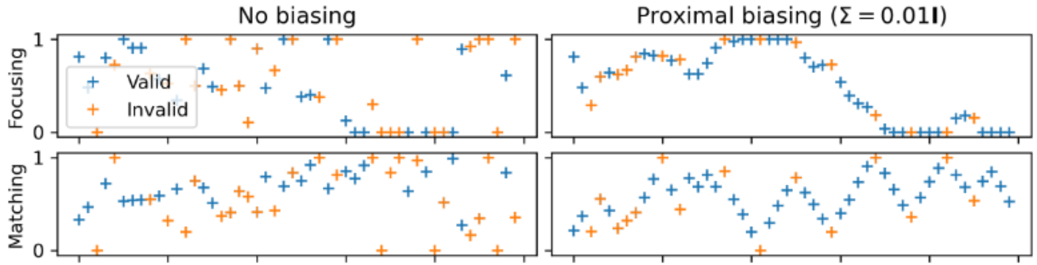
$$\hat{\alpha}(x) = \alpha(x) \exp\left(-\frac{(x - x_0)^2}{2\sigma^2}\right)$$

Warning: Requires $\alpha(x) \geq 0$



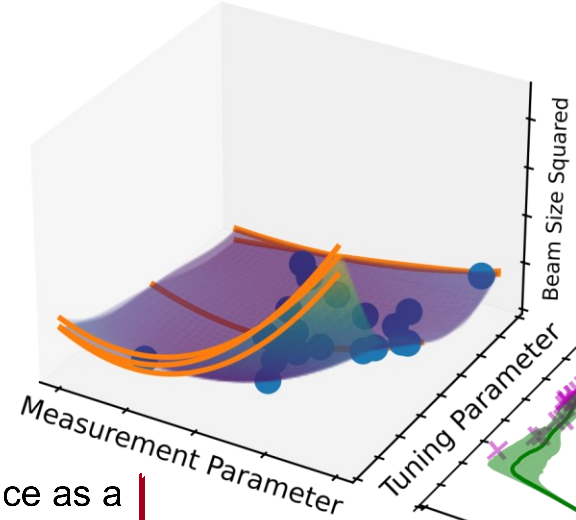
Roussel et. al. *Nat. Comm.* **2021**

Reduces travel distances during exploration

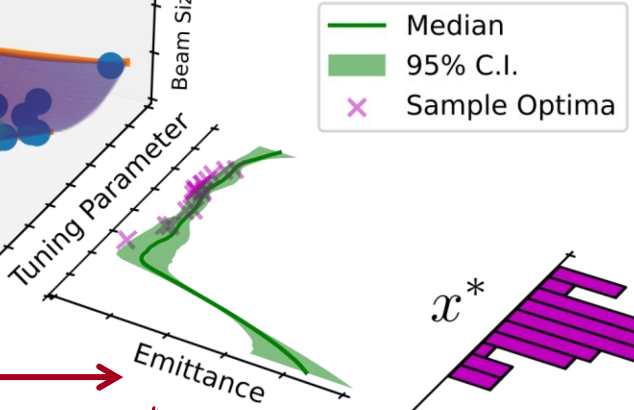


Our goal:
Determine the **ideal tuning configuration x^*** which minimizes emittance

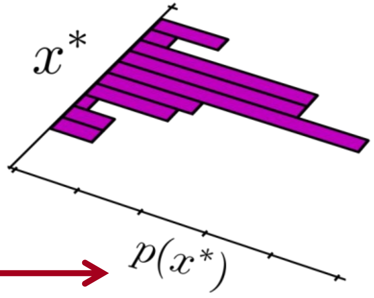
Beam size model



Predict beam emittance as a function of tuning config. (with uncertainty)



Predict optimal tuning config. (with uncertainty)

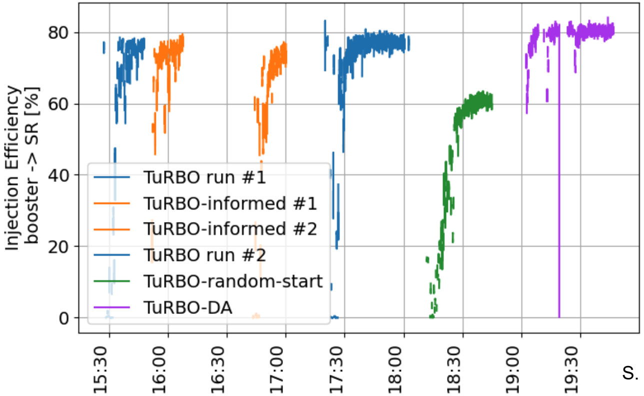


The BAX algorithm chooses beam size measurements that reduce uncertainty in x^* **without measuring emittance directly → 20x speed up**

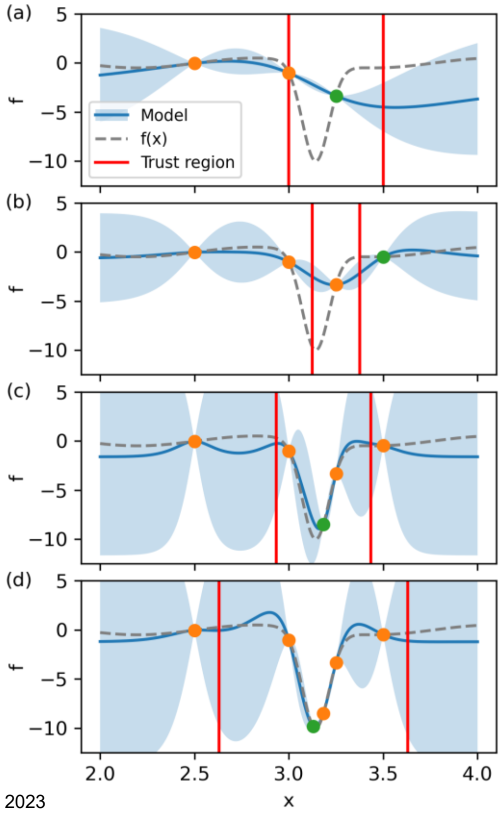
Miskovich, et. al., arXiv:2209.04587

Trust Region Bayesian Optimization (TuRBO)

- Bayesian optimization tends to **over-prioritize exploration** to find global optima
- **Restrict search region** to local area around best point
- Expand / contract **“trust” region** based on algorithm successes / failures on-the-fly
- Helps find local extrema in high dimensional problems
- Optimization success at ESRF led to the **highest ever observed lifetime** using 8 knobs in under 30 mins



S. M. Liuzzo, MO3AO01, ICALEPS 2023



Xopt: Flexible Optimization of Arbitrary Problems

SLAC

Easy to control

```python
# create Xopt object.
X = Xopt(YAML)

# take 10 steps and view data
for _ in range(10):
    X.step()

X.data
```

Python interface

```yaml
xopt:
    max_evaluations: 6400

generator:
    name: cnsga
    population_size: 64
    population_file: test.csv
    output_path: .

evaluator:
    function: xopt.resources.test_functions.tnk.evaluate_TNK
    function_kwargs:
      raise_probability: 0.1

vocs:
    variables:
        x1: [0, 3.14159]
        x2: [0, 3.14159]
    objectives: {y1: MINIMIZE, y2: MINIMIZE}
    constraints:
        c1: [GREATER_THAN, 0]
        c2: [LESS_THAN, 0.5]
    linked_variables: {x9: x1}
    constants: {a: dummy_constant}
```

Xopt

Text file interface

Badger GUI interface

Simple to connect with simulations / machine
(single python function!)

```python
evaluate(inputs: dict) -> dict
```

Many optimization algorithms
- Genetic algorithms (NSGA-II, etc.)
- Nelder-Mead Simplex
- Bayesian optimization, almost everything shown in this presentation

Currently used at many facilities: SLAC, DESY, Argonne, ESRF, BNL, LBNL, etc.

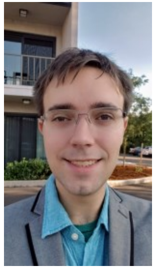
https://christophermayes.github.io/Xopt/

Conclusion

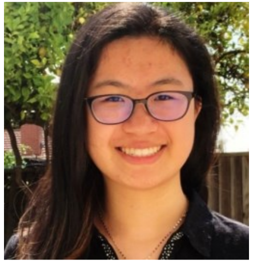
- We have identified algorithms available for automating many online and offline optimization tasks in accelerator physics
  - Some work is needed to make faster decisions, characterize safety-performance trade-offs, integrate into control systems
- These algorithms can be used out of the box to improve EIC operations and accelerator design in simulation

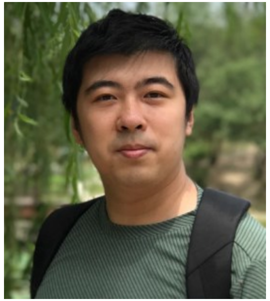
Questions?

Thanks to the team!







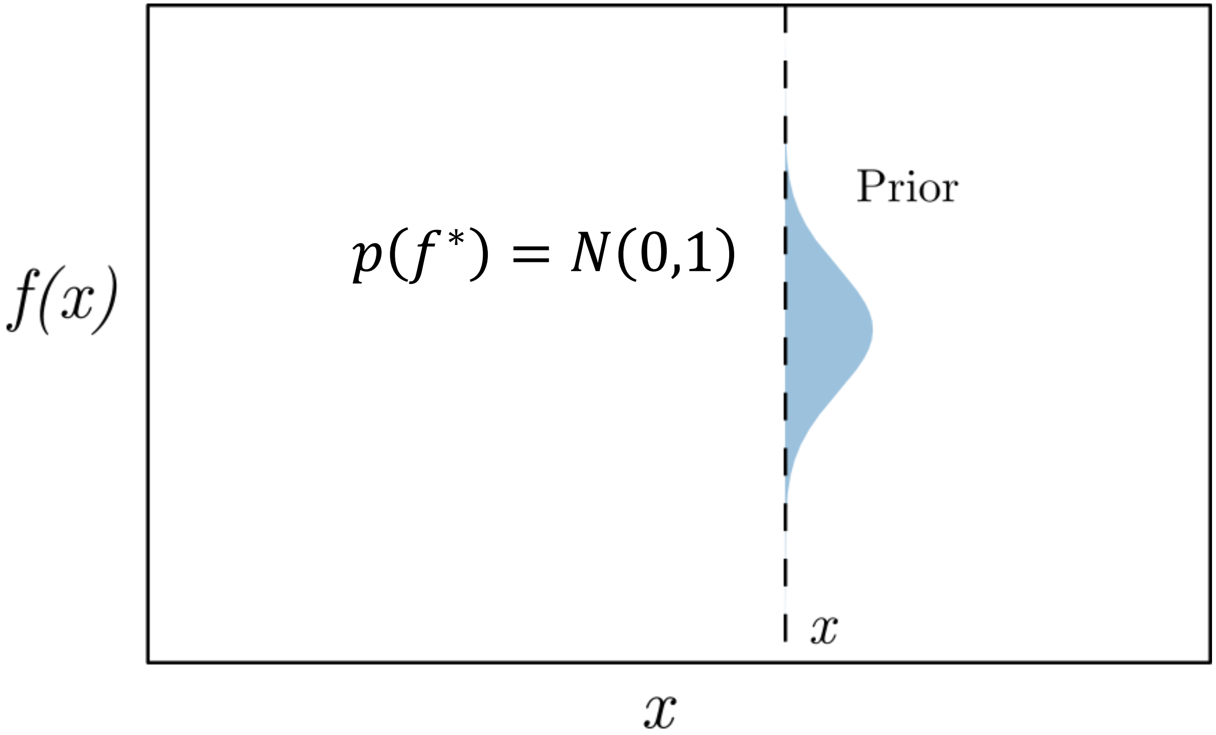




Gaussian Process Math

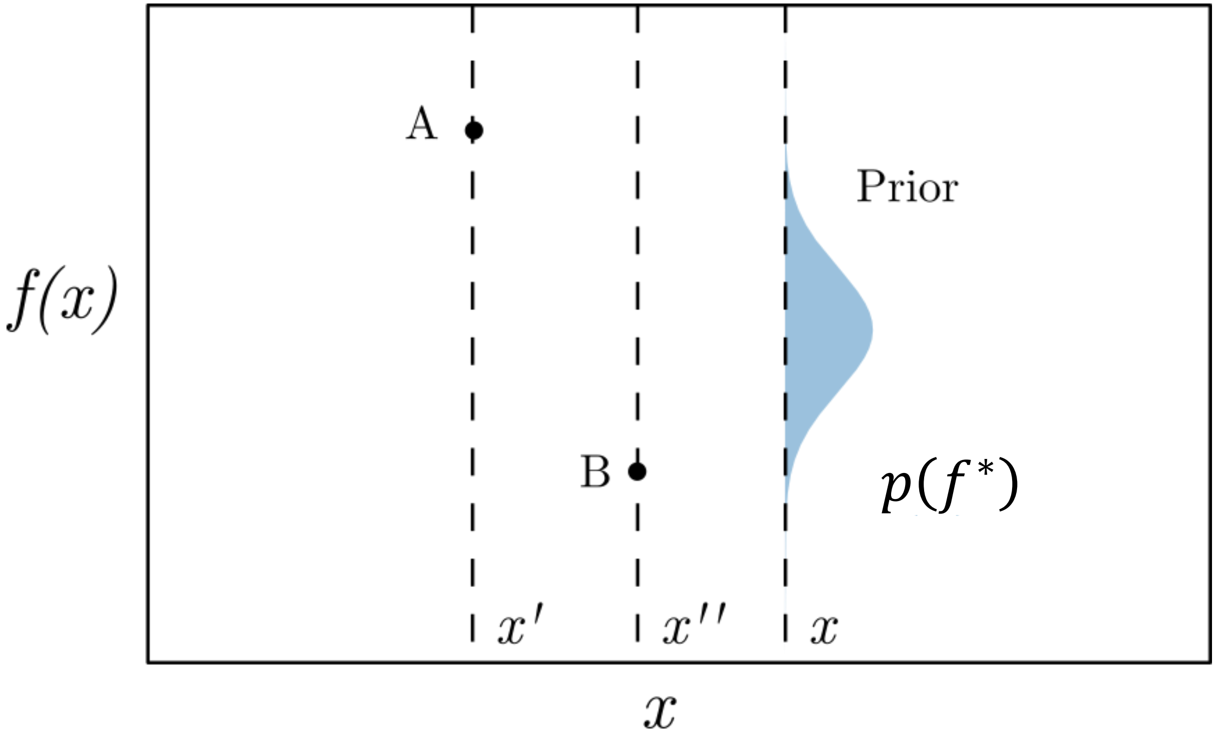
Some intuition…

Let's predict the function value f^* at the point x



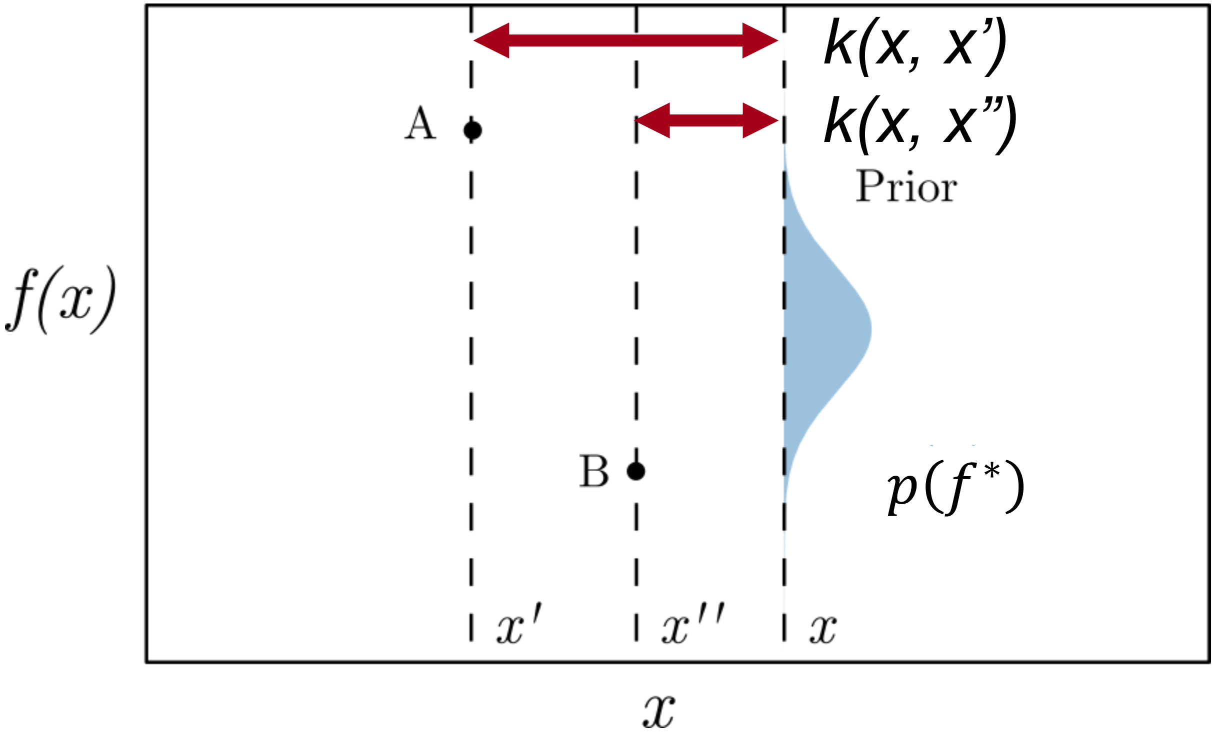
Some intuition…

Which observation will have a larger impact on changing *p(f)*?



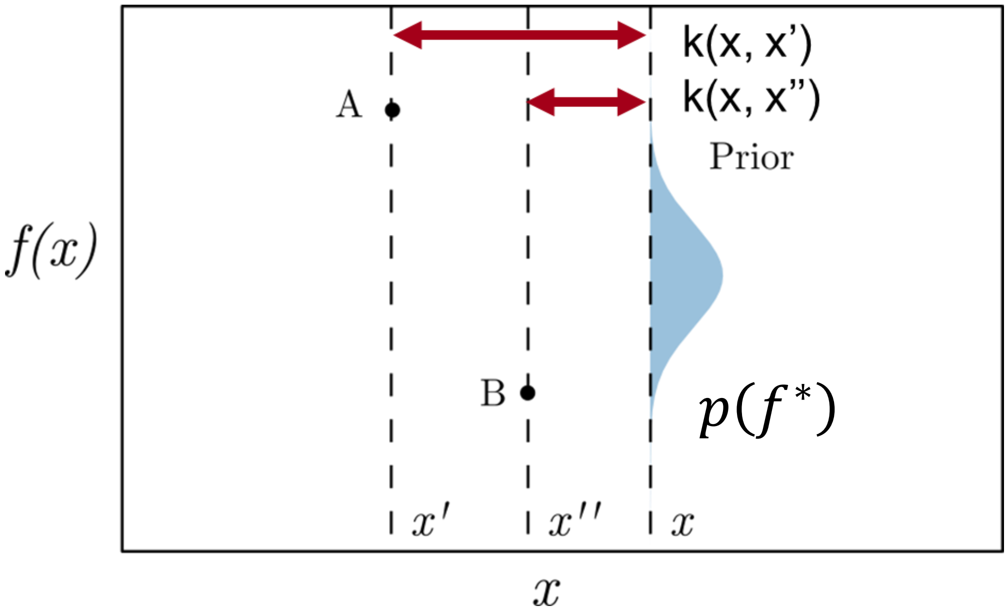
Adding some math

Which observation will have a larger impact on changing *p(f)*?



$k(x, x') < k(x, x'')$

Adding some math



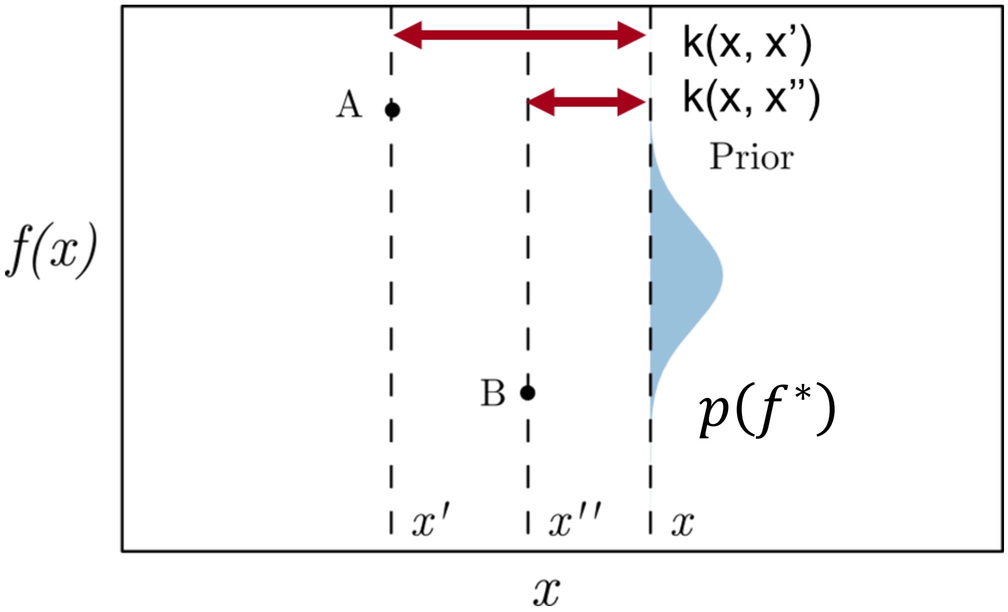
$$p(f_A, f_B, f^*) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} k(x',x') & k(x',x'') & k(x',x) \\ k(x',x'') & k(x'',x'') & k(x'',x) \\ k(x',x) & k(x'',x) & k(x,x) \end{pmatrix}$$

$$p(f^* | f_A, f_B) = \frac{p(f_A, f_B | f^*) p(f^*)}{p(f_A, f_B)} = \frac{p(f_A, f_B, f^*)}{p(f_A, f_B)}$$

Bayes rule

Adding some math



$$p(f^* | f_A, f_B) = N(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*)$$

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} + K^* K^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$
$$\boldsymbol{\sigma}^* = K^{**} - K^{*T} K^{-1} K^*$$

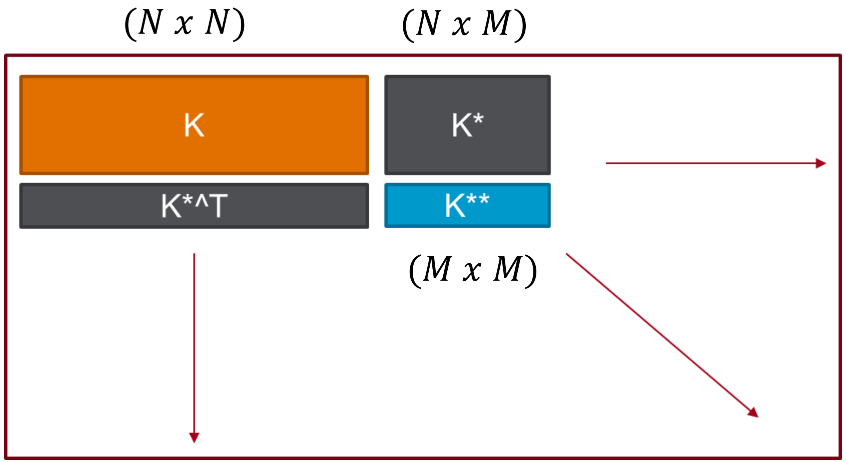
$$p(f^* | f_A, f_B) = \frac{p(f_A, f_B | f^*) p(f^*)}{p(f_A, f_B)} = \frac{p(f_A, f_B, f^*)}{p(f_A, f_B)}$$

$K^{-1} \sim \mathcal{O}(N^3)$!

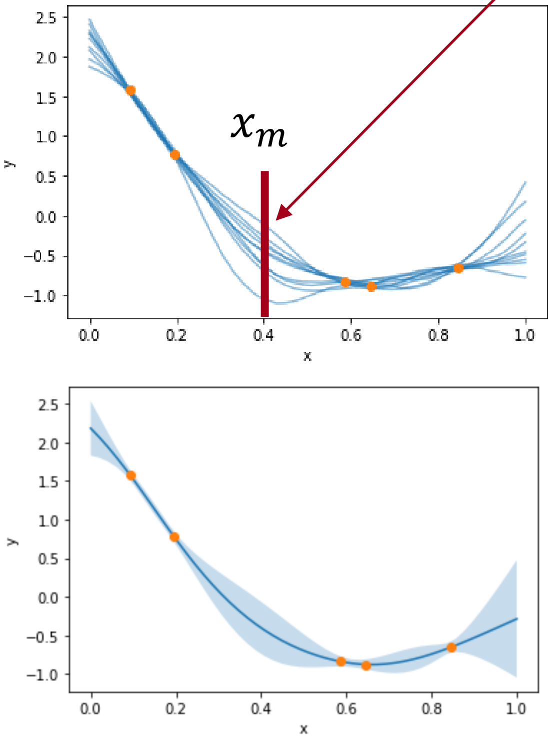
Making predictions with GP's

What about multiple predictions?

$$p(f_0^*, f_1^*, \ldots, f_M^* | f_0, f_1, \ldots, f_N) = N(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*)$$



$p(f_m^* | f_0, f_1, \ldots, f_N)$



Samples

Distribution

Draw function samples? Sample from the joint posterior distribution at requested points