CATALYSIS CENTER FOR
ENERGY INNOVATION
AN ENERGY FRONTIER RESEARCH CENTER

NEXTorch: A Design and Bayesian Optimization Toolkit for Chemical Sciences and Engineering

Yifan Wang

Dec 2023 Vlab Workshop

U.S. DEPARTMENT OF ENERGY
Office of Science

An Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Basic Sciences

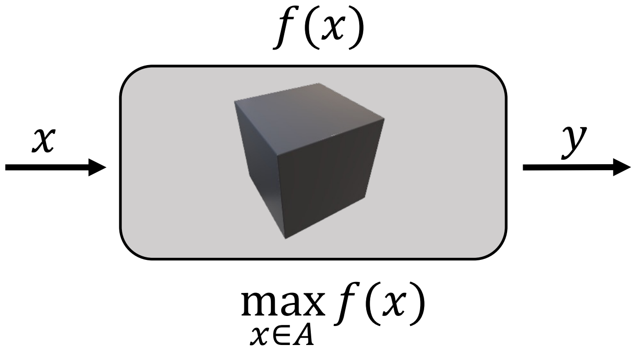
About Me

- Graduated from Vlachos Group Dec 2021
- ML Research Scientist at Meta
- Based in San Francisco
- Research interests: AI data annotation, active learning, generative AI, reinforcement learning from human feedback (RLHF)
- LinkedIn: https://www.linkedin.com/in/wangyifan411/

Global Optimization of Blackbox Functions

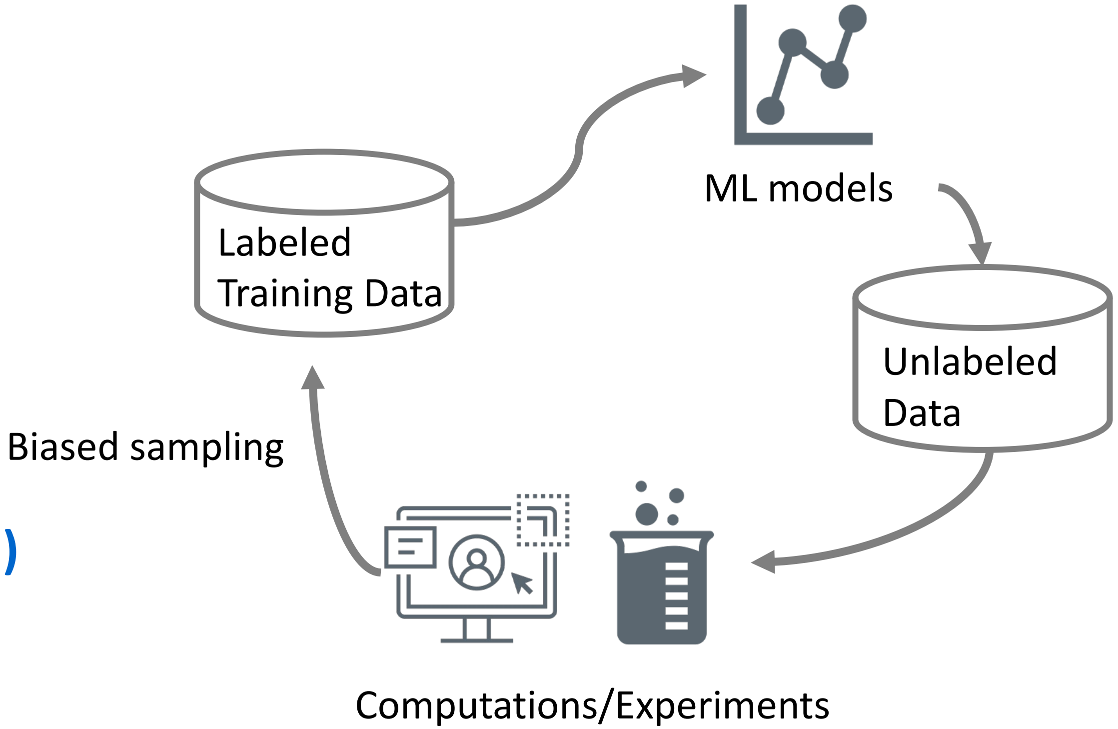
Blackbox functions [1]

- Expensive computer model or laboratory experiments
- Unknown explicit model form
- Multi-dimensional



Active learning [2]

- An algorithm "learning" from data, proposing next experiments, and improving prediction accuracy with fewer training data or lower cost



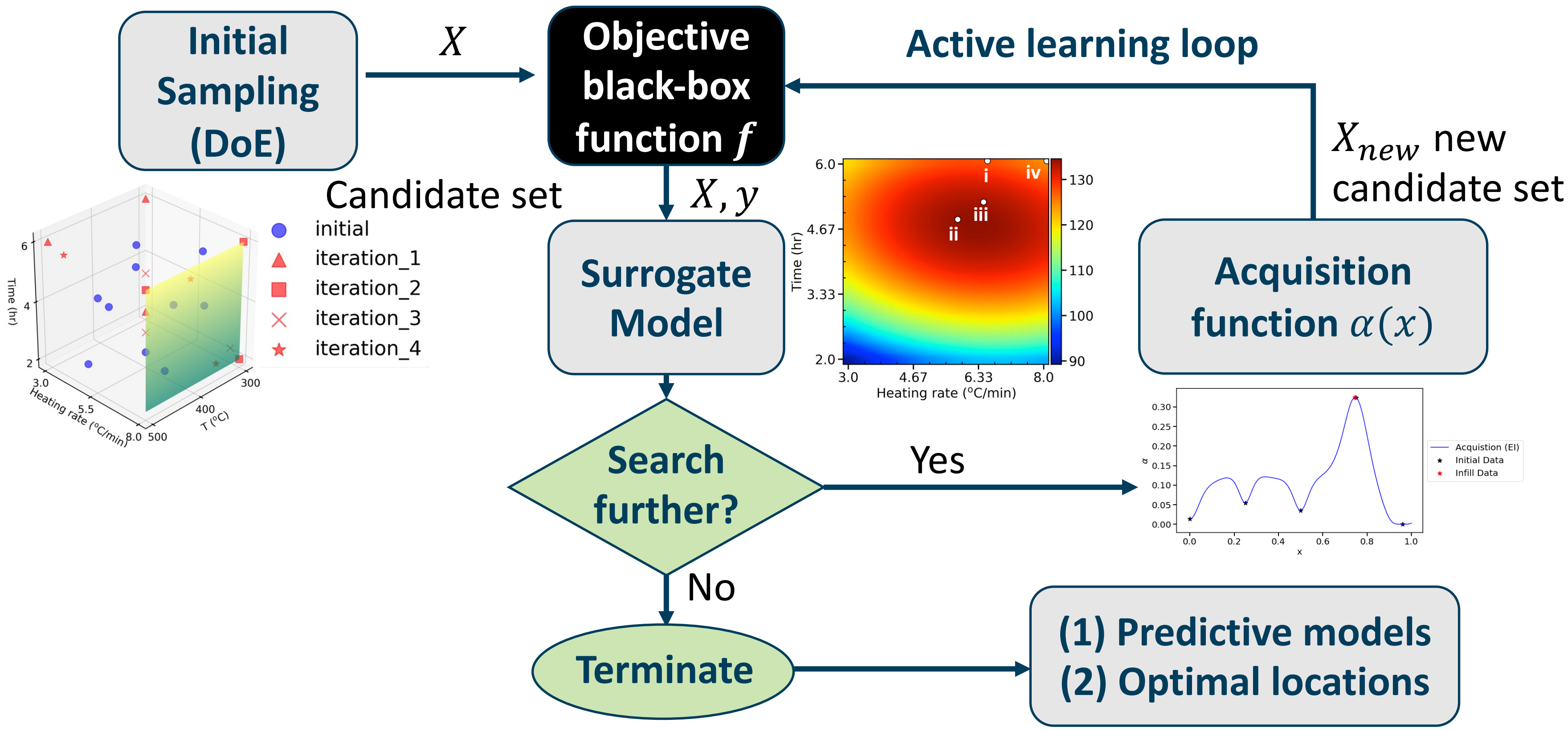
Use active learning to reduce experimental (computational) cost and improve accuracy of the surrogate model

[1] D.R. Jones, M. Schonlau, and W. J. Welch, J. Glob. Optim. **13**, 455 (1998).

[2] Settles, B. *Active Learning Literature Survey*. *Active Learning Literature Survey* (2009).

CATALYSIS CENTER FOR ENERGY INNOVATION

Bayesian Optimization (BO)



Bayesian Statistics

Posterior Data Prior

$$P(f|D) \propto P(D|f)P(f)$$

Thomas Bayes

- Initial sampling can be generated through design of experiments (DoE)
- The surrogate model is typically a Gaussian Process (GP)
- Next experiment points are generated by acquisition functions (exploration vs. exploitation)

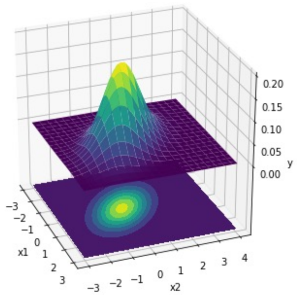
[1] P.I. Frazier, 1 (2018). http://arxiv.org/abs/1807.02811

5

CATALYSIS CENTER FOR ENERGY INNOVATION

BO Building Blocks

Gaussian Process (GP)



= **Mean function $\mu_0(X)$ + Kernel (covariance) function $\sum_0(X, X')$**

Constant $\mu_0(x) = \mu$

Polynomial $\mu_0(x) = \mu + \sum_{i=1}^{p} \beta_i \Psi_i(x)$

Exponential (Gaussian) $\Sigma_0(x, x') = \alpha_0 \exp\left(-||x - x'||^2\right)$

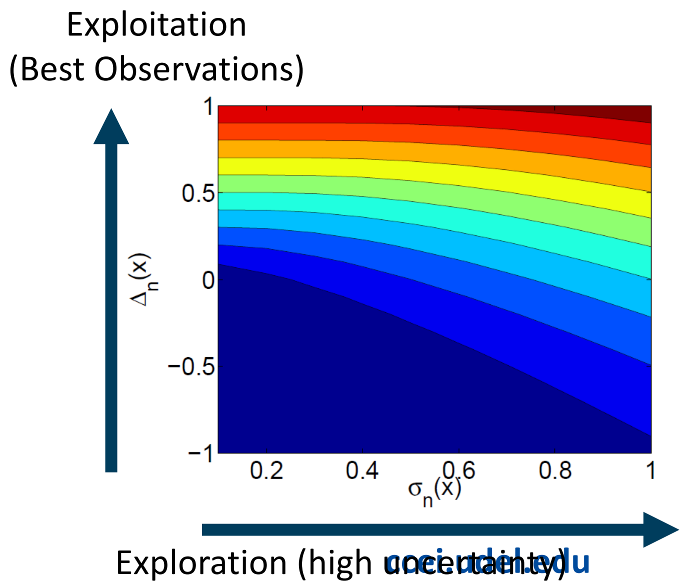
Matern $\Sigma_0(x, x') = \alpha_0 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}||x - x'||\right)^{\nu} K_\nu(\sqrt{2\nu}||x - x'||)$

The hyperparameters are determine by maximizing the cost function – maximum likelihood estimate (MLE)

Acquisition Function

= Controls the trade-off between **exploration** and **exploitation**

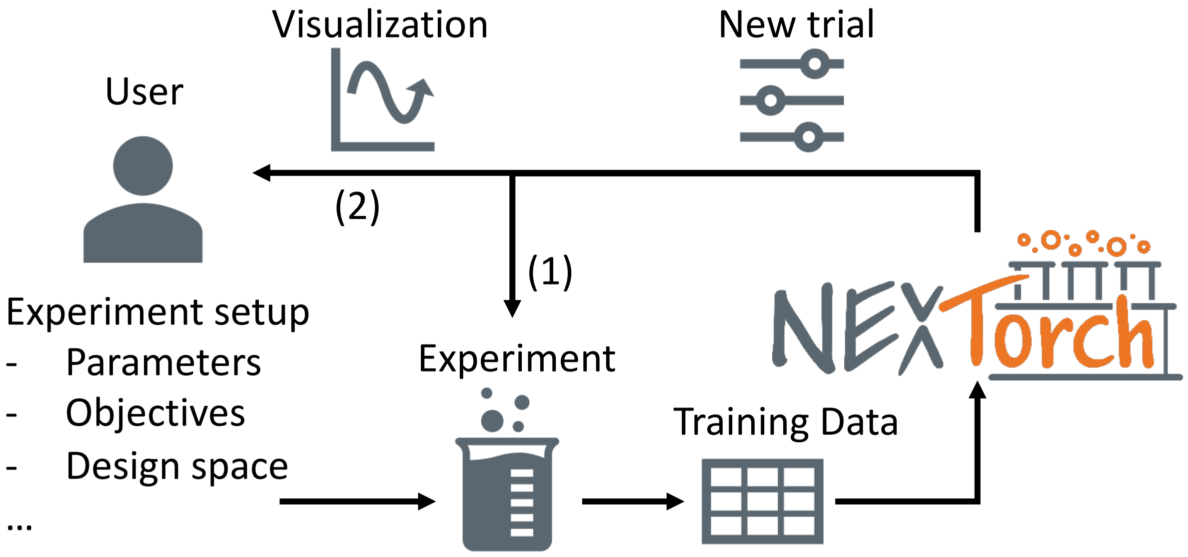
Expected improvement (EI)
Probability of improvement (PI)
Upper confidence bound (UCB)



1 P.I. Frazier, 1 (2018). http://arxiv.org/abs/1807.02811
2 C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning* (2000).

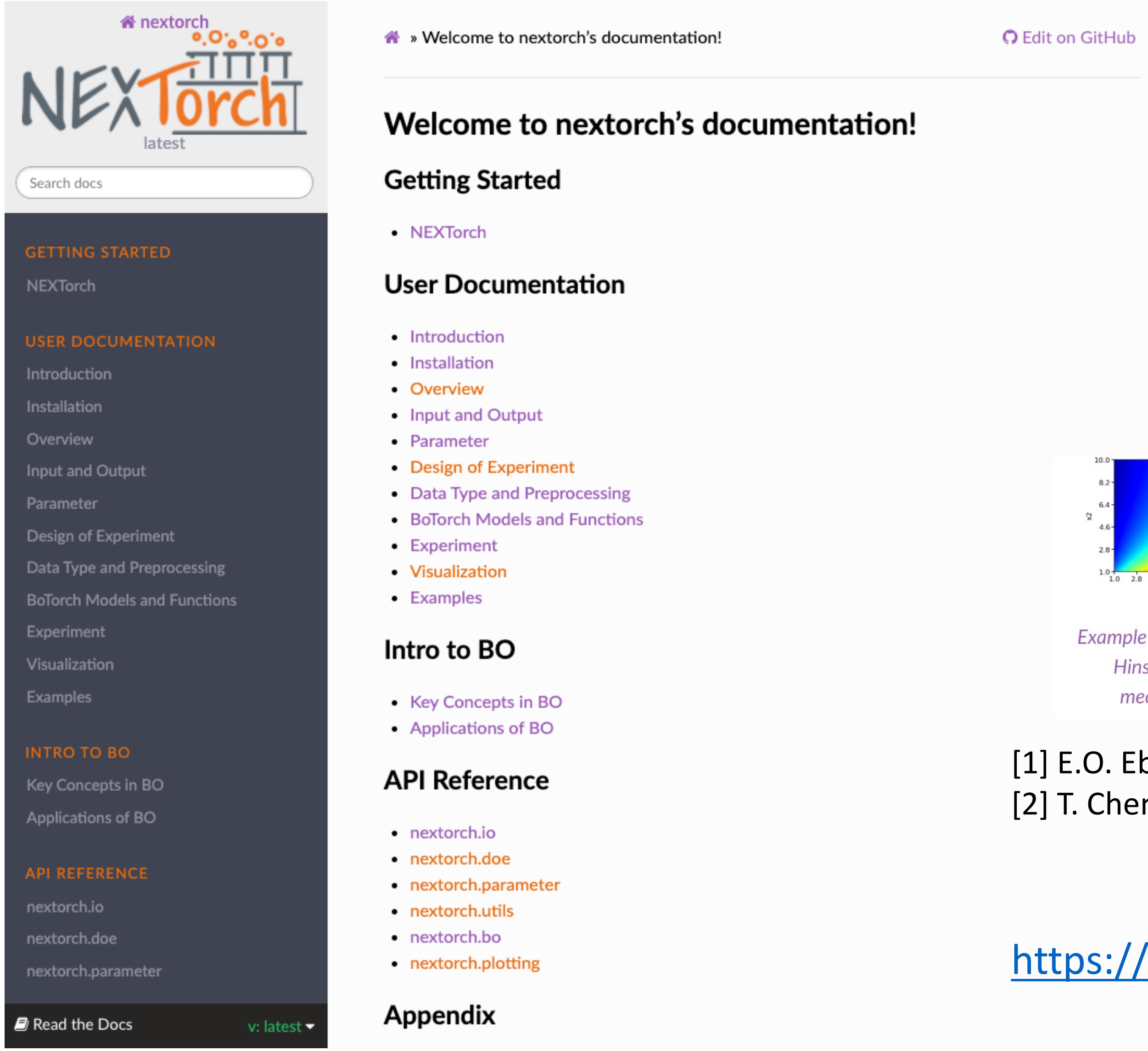
DOE and BO Toolkit - NEXTorch

- Key dependencies: python, pyDOE (pythonhosted.org/pyDOE/), PyTorch (pytorch.org), BoTorch (botorch.org)
- GPU acceleration, modern BO algorithms, visualization
- Connect BO implement to chemistry or engineering problems
  - (1) Automated optimization (good for computations);
  - (2) Human-in-the-loop optimization (good for laboratory experiments)



[1] **Y. Wang**, T. Chen, and D.G. Vlachos, J. Chem. Inf. Model. 61, 5312–5319 (2021).
GitHub: https://github.com/VlachosGroup/nextorch
Documentation: https://nextorch.readthedocs.io/en/latest/index.html

Online Documentation Page


nextorch
NEXTorch
latest
Search docs

GETTING STARTED
NEXTorch

USER DOCUMENTATION
Introduction
Installation
Overview
Input and Output
Parameter
Design of Experiment
Data Type and Preprocessing
BoTorch Models and Functions
Experiment
Visualization
Examples

INTRO TO BO
Key Concepts in BO
Applications of BO

API REFERENCE
nextorch.io
nextorch.doe
nextorch.parameter

Read the Docs v: latest

» Welcome to nextorch's documentation! Edit on GitHub

Welcome to nextorch's documentation!

Getting Started

- NEXTorch

User Documentation

- Introduction
- Installation
- Overview
- Input and Output
- Parameter
- Design of Experiment
- Data Type and Preprocessing
- BoTorch Models and Functions
- Experiment
- Visualization
- Examples

Intro to BO

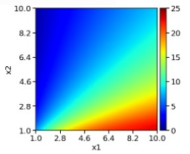
- Key Concepts in BO
- Applications of BO

API Reference

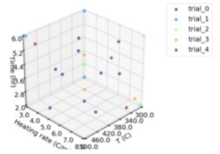
- nextorch.io
- nextorch.doe
- nextorch.parameter
- nextorch.utils
- nextorch.bo
- nextorch.plotting

Appendix

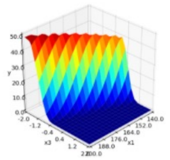
- NEXTorch modules and functions
- Tutorials with code examples
- Introduction to BO theory
- BO applications in literature



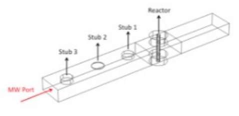
Example 3 - Langmuir Hinshelwood mechanism



Example 4 - Nitrogen-doped carbon catalysts



Example 5 - Plug flow reactor yield



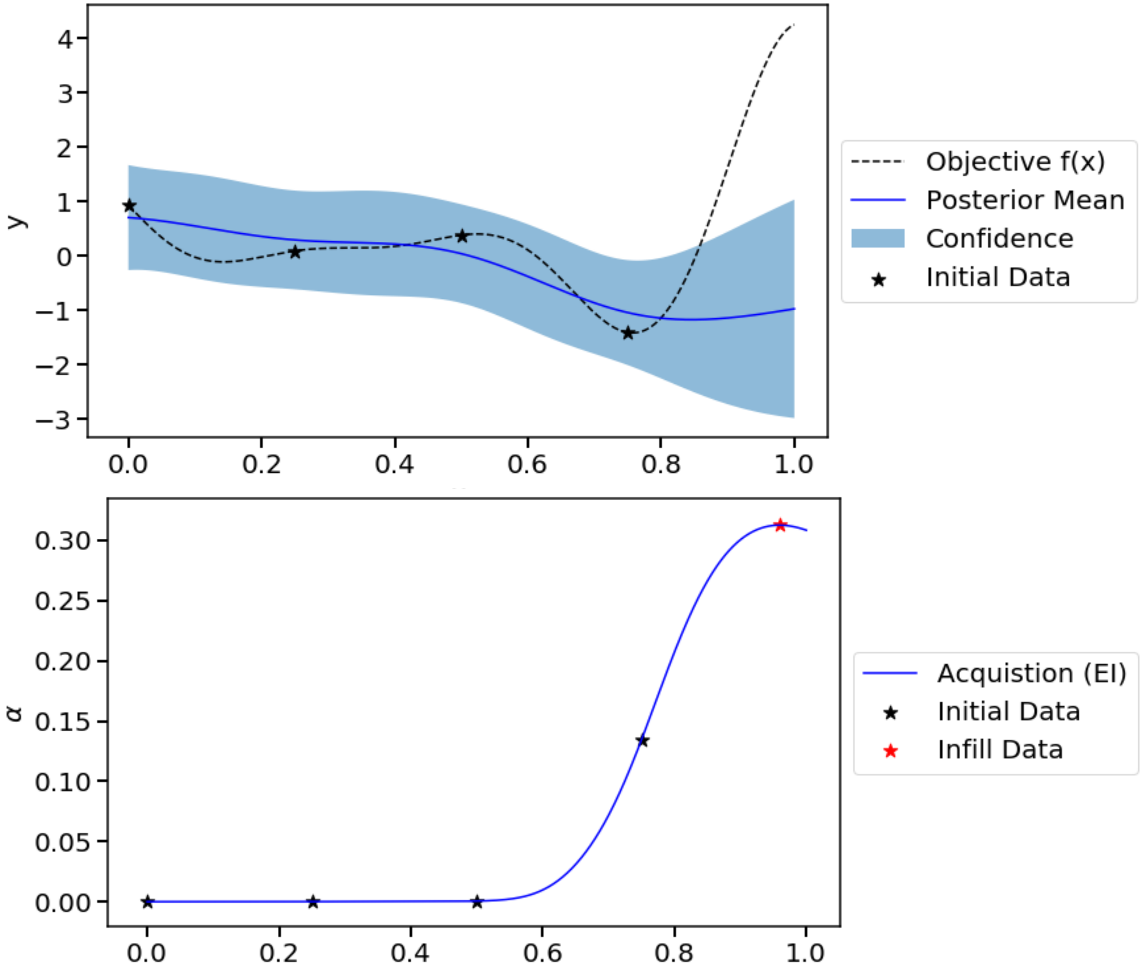
Example 8 - Stub tuner of the microwave cavity

[1] E.O. Ebikade, et al, React. Chem. Eng. (2020).
[2] T. Chen, et al, Ind. Eng. Chem. Res. 59, 10418 (2020).

https://nextorch.readthedocs.io/en/latest/index.html

BO Examples in 1D

- Find the minima of $f(x) = (6x-2)^2 sin(12x-4); x \in [0,1]$
- Starting from $x = 0, 0.25, 0.5, 0.75$

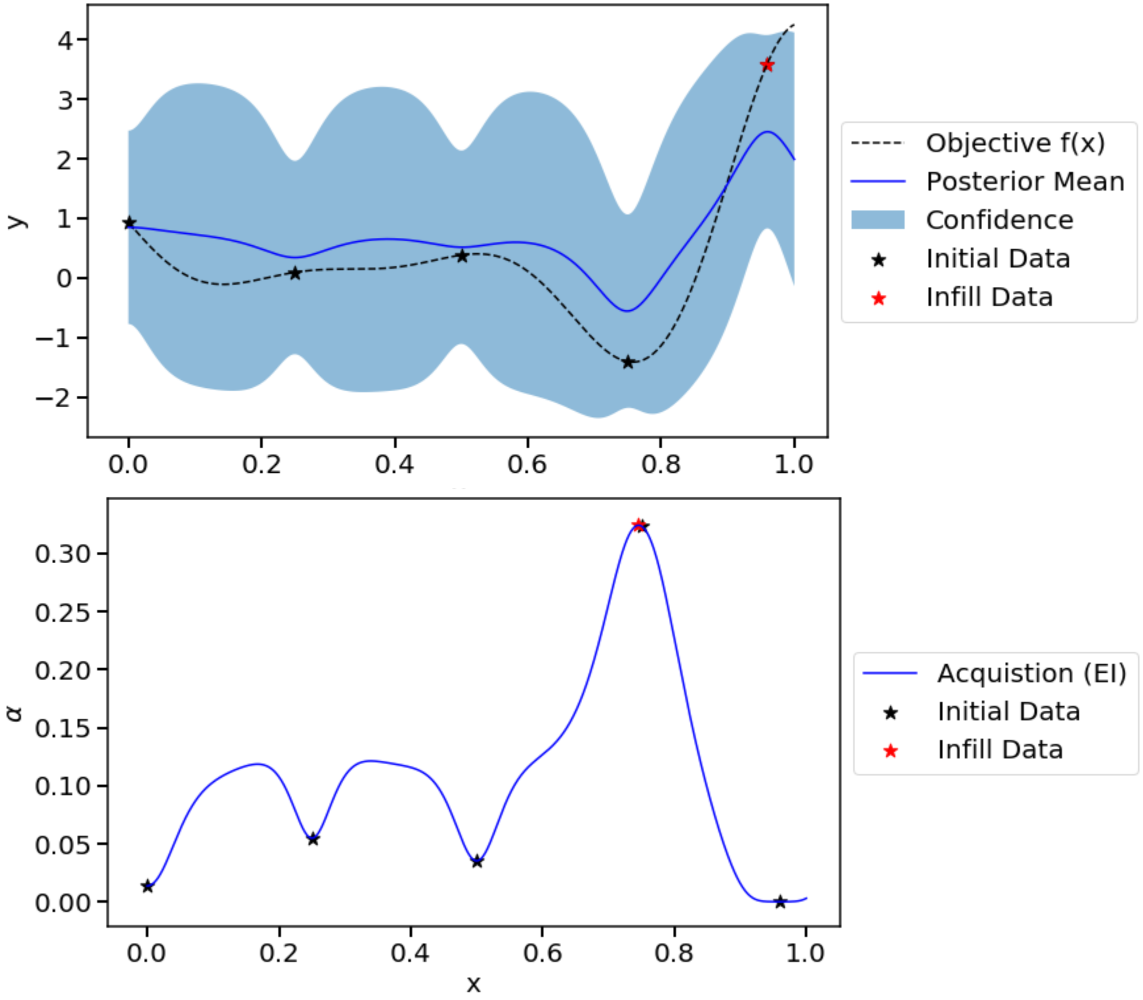


Demo instructions: https://github.com/VlachosGroup/vlab_workshop_2023/tree/main/NEXTorch
Notebook location: https://github.com/VlachosGroup/nextorch/blob/main/examples/notebooks/01_simple_1d.ipynb

CATALYSIS CENTER FOR ENERGY INNOVATION

BO Examples in 1D

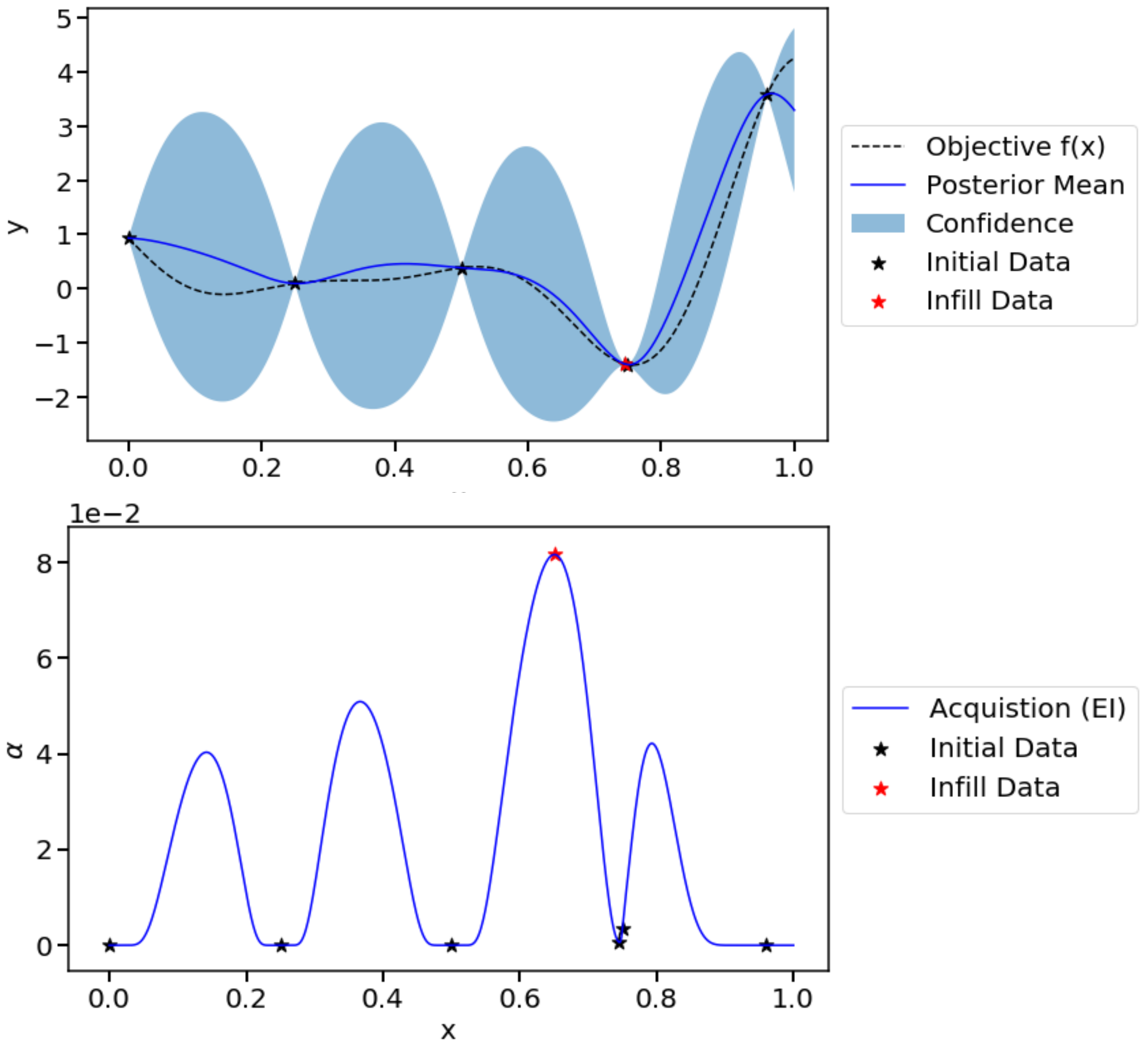
- Find the minima of $f(x) = (6x-2)^2 sin(12x-4); x \in [0,1]$
- Iteration 1, acquisition function – expected improvement (EI)



CATALYSIS CENTER FOR
ENERGY INNOVATION

BO Examples in 1D

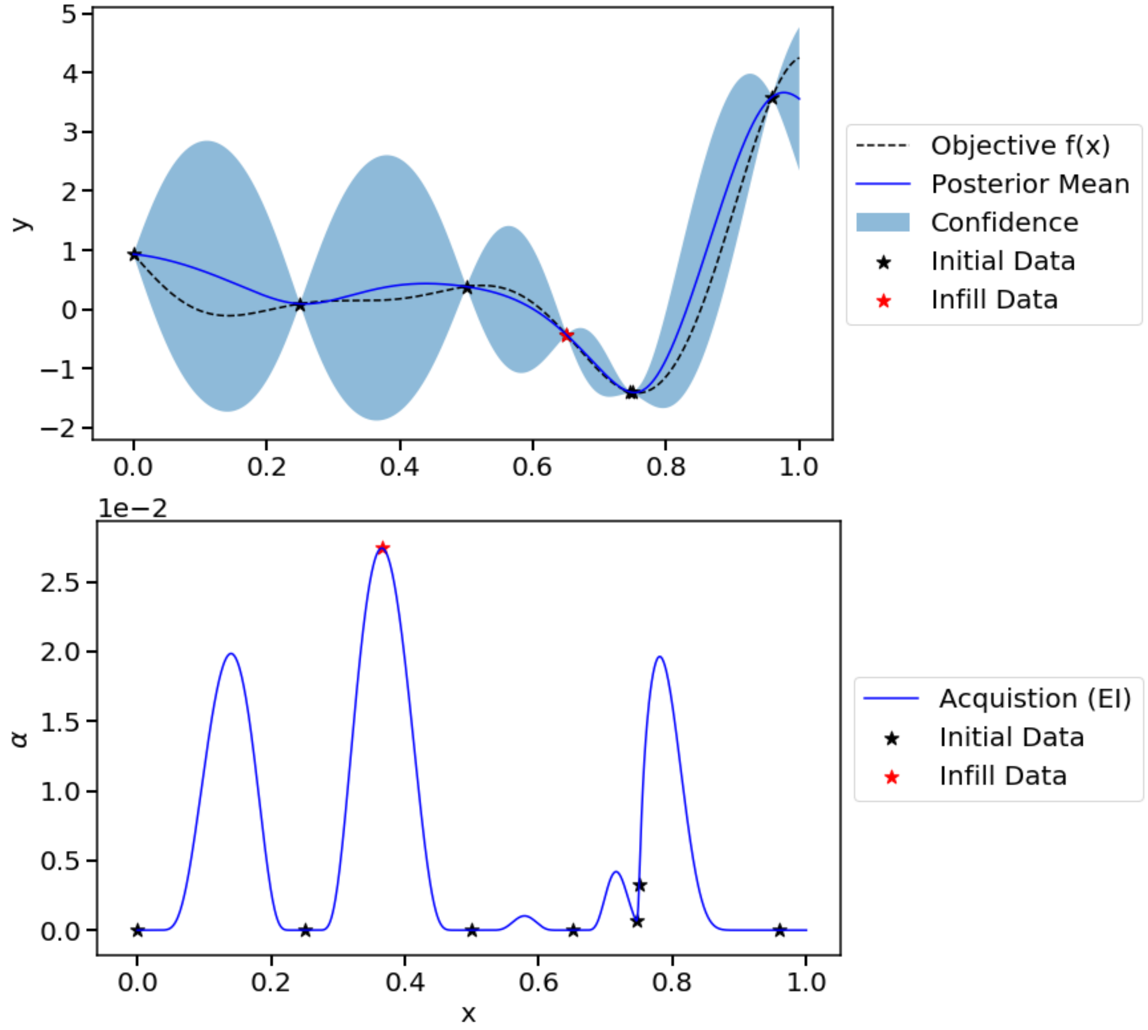
- Find the minima of $f(x) = (6x - 2)^2 sin(12x - 4); x \in [0,1]$
- Iteration 2



CATALYSIS CENTER FOR ENERGY INNOVATION

BO Examples in 1D

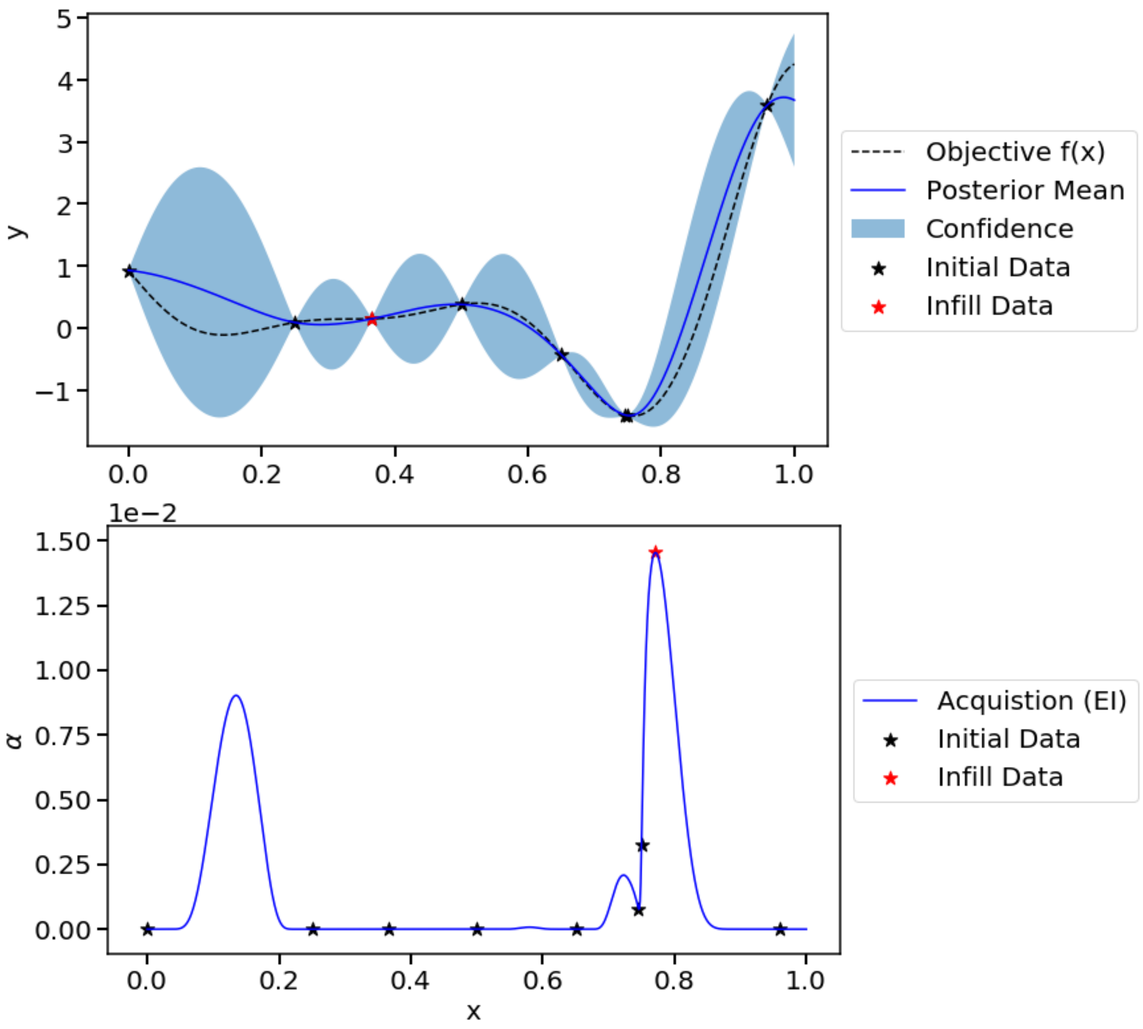
- Find the minima of $f(x) = (6x - 2)^2 sin(12x - 4); x \in [0,1]$
- Iteration 3



CATALYSIS CENTER FOR ENERGY INNOVATION

BO Examples in 1D

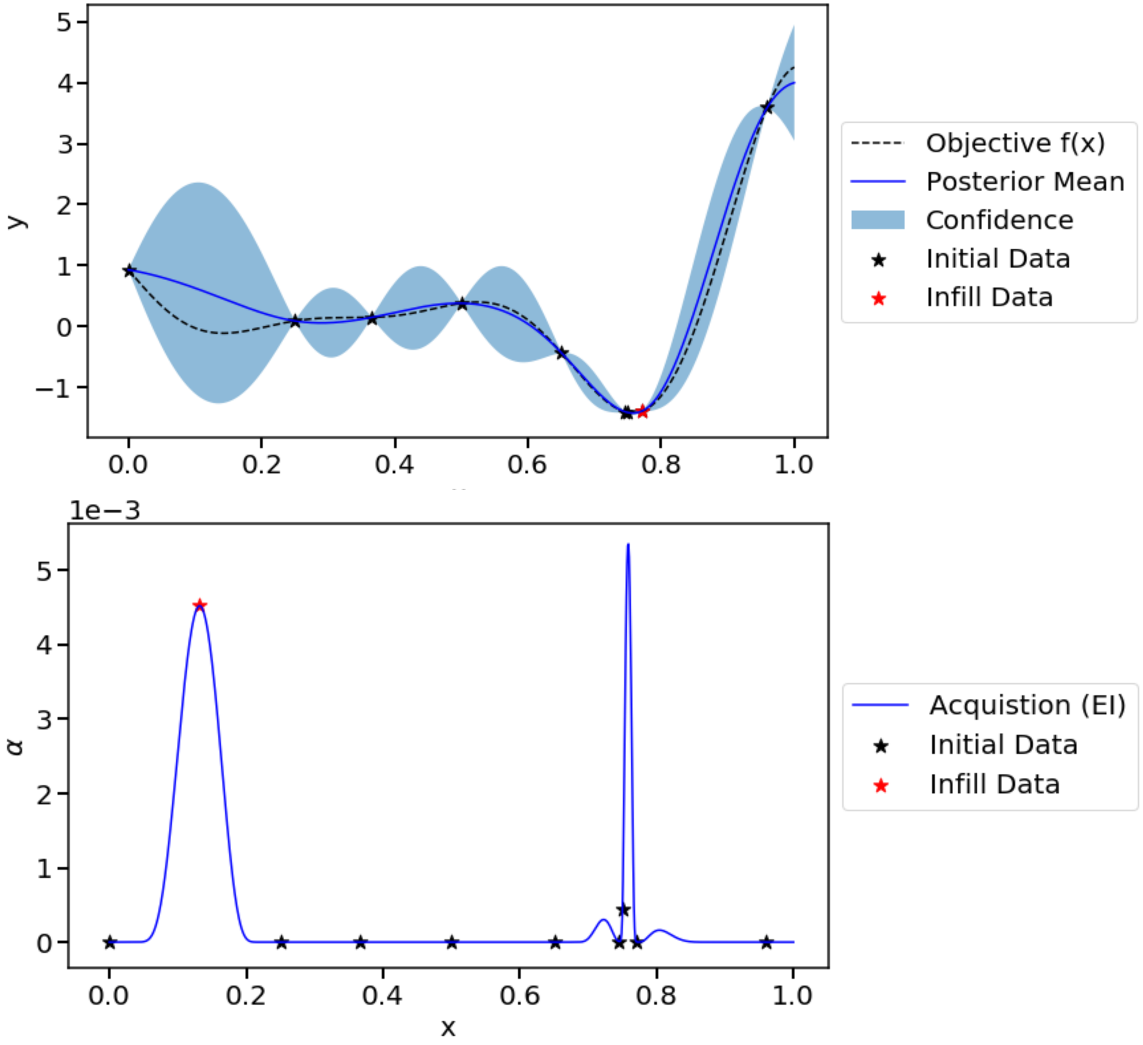
- Find the minima of $f(x) = (6x-2)^2 sin(12x-4); x \in [0,1]$
- Iteration 4



CATALYSIS CENTER FOR ENERGY INNOVATION

BO Examples in 1D

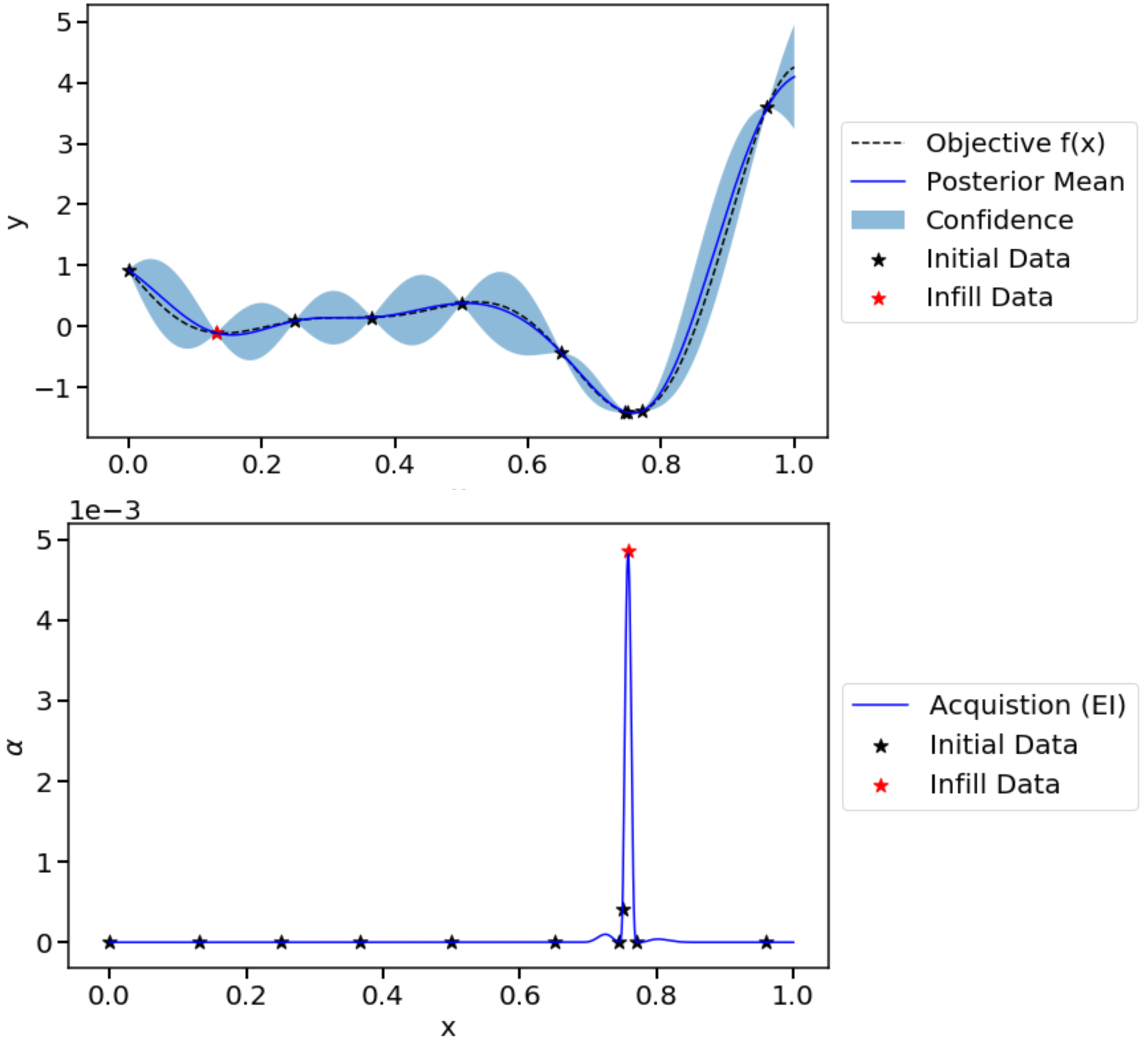
- Find the minima of $f(x) = (6x-2)^2 sin(12x-4); x \in [0,1]$
- Iteration 5



CATALYSIS CENTER FOR
ENERGY INNOVATION

BO Examples in 1D

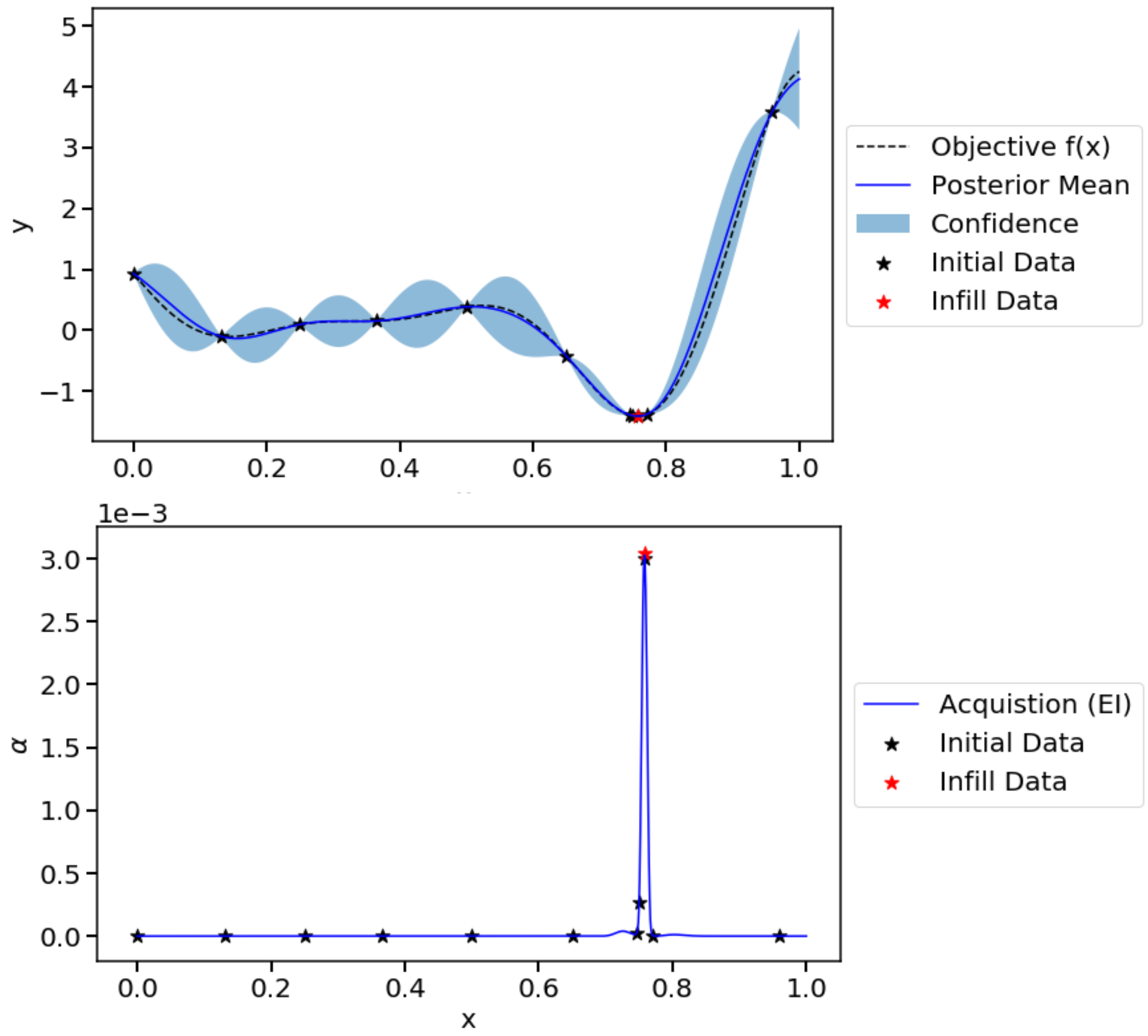
- Find the minima of $f(x) = (6x-2)^2 sin(12x-4); x \in [0,1]$
- Iteration 6



CATALYSIS CENTER FOR ENERGY INNOVATION

BO Examples in 1D

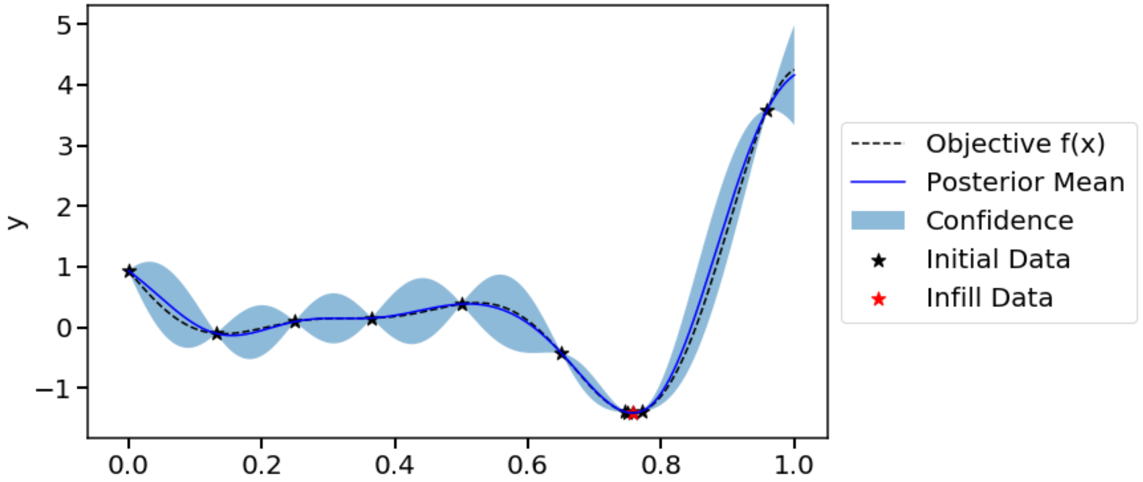
- Find the minima of $f(x) = (6x-2)^2 sin(12x-4); x \in [0,1]$
- Iteration 7

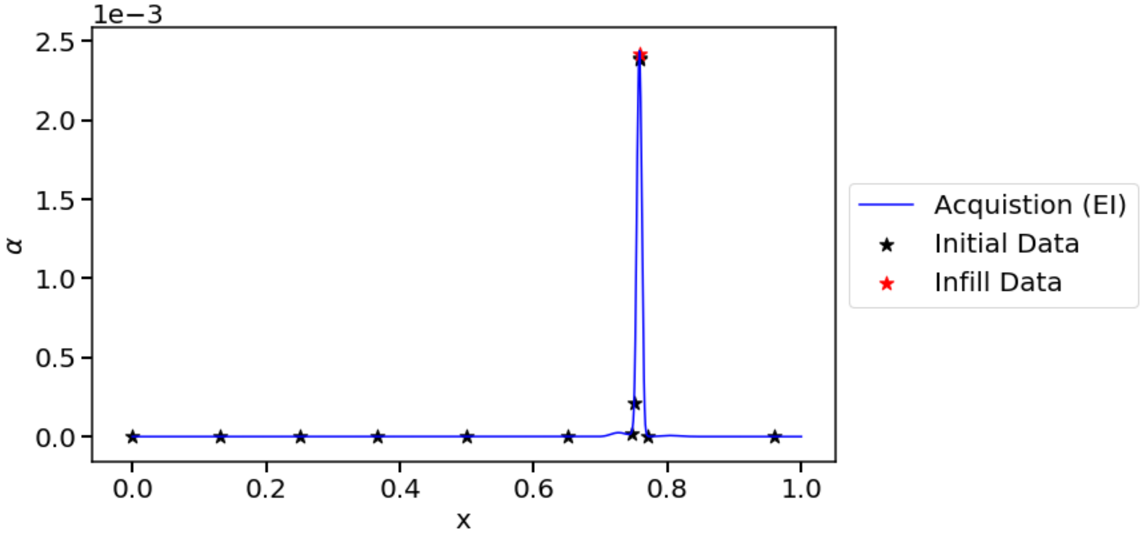


CATALYSIS CENTER FOR
ENERGY INNOVATION

BO Examples in 1D

- Find the minima of $f(x) = (6x-2)^2 sin(12x-4); x \in [0,1]$
- Iteration 8





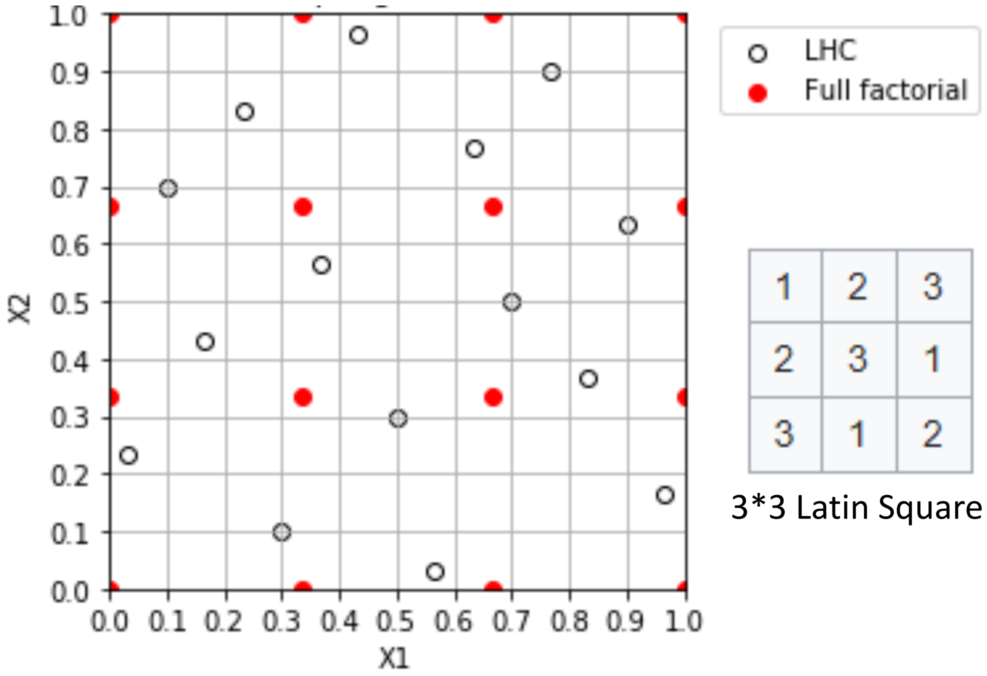
CCEI CATALYSIS CENTER FOR ENERGY INNOVATION

Optimized Latin Hypercube design (LHS)

Initial Sampling plan for high dimensional space:

- Monte Carlo sampling
- Maximize the distance between points
- Maximize the information gain in the sampling space

2D example (level = 4, 16 samples)



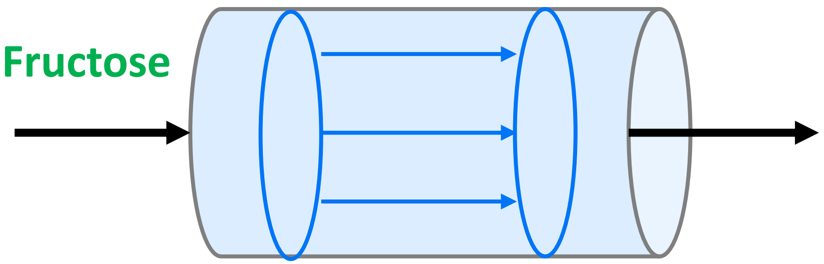
1	2	3
2	3	1
3	1	2

3*3 Latin Square

CATALYSIS CENTER FOR ENERGY INNOVATION

Case Study – HMF Yield Optimization

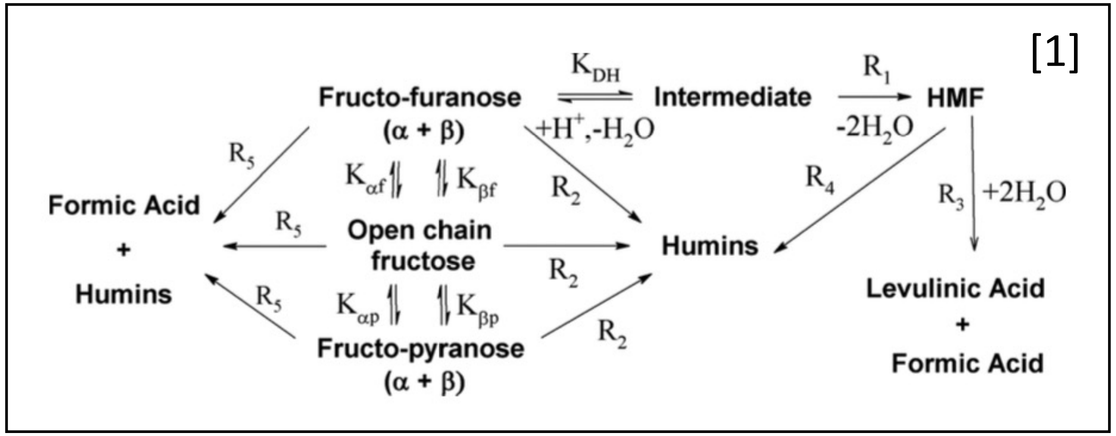
- Fructose dehydrated to produce valuable platform chemical, HMF.
- It's important to maximize the HMF yield (Y) to improve productivity and reduce downstream costs.
- Three key input parameters ($\boldsymbol{X}$) include the reaction temperature, pH, and residence time.



$$u\frac{dC_i}{dx} = R_i$$

$R_{FRU} = -r_1-r_2-r_3$
$R_{HMF} = r_1-r_3-r_4$
$R_{LA} = r_3$
$R_{FA} = r_3+r_5$

$$r_1 = k_1(T)C_{Fru}\left(\frac{K_{DH}(T)C_{H+}}{C_{H_2O}}\right)$$
$$r_2 = k_2(T)C_{Fru}C_{H+}$$
$$r_3 = k_3(T)C_{HMF}C_{H+}$$
$$r_4 = k_4(T)C_{HMF}C_{H+}$$
$$r_5 = k_5(T)C_{Fru}C_{H+}$$



[1] TD Swift et al., *ACS Catal.* 4, **2014**, 259.

CATALYSIS CENTER FOR ENERGY INNOVATION

Step 1 - Define Parameter Space and Sampling Plan

Code example 1. Import, Parameter Space and DOE

```
from nextorch import bo, doe
import PFR_yield

X_name_list = ['T', 'pH', 'log10(tf)']
Y_name = 'Yield %'
var_names = X_name_list + [Y_name]

# Objective function
objective_func = PFR_yield

# Set the operating range for each parameter
X_ranges = [[140, 200],
            [0, 1],
            [-2, 2]]

# Get the information of the design space
n_dim = len(X_name_list) # the dimension of inputs
n_objective = 1 # the dimension of outputs

# Latin hypercube design with 10 initial points
n_init_lhs = 10
X_init_lhs = doe.latin_hypercube(n_dim=n_dim, n_points=n_init_lhs, seed=1)
# Get the initial responses
Y_init_lhs = bo.eval_objective_func(X_init_lhs, X_ranges, objective_func)
```

- The ranges of temperature, pH, and residence time are defined.
  - Temperature: 140 – 200 [°C]
  - pH: 0 – 1 [-]
  - Residence time: 0.01 – 100 [min]
- Three different sampling plans are tested.
  - Full factorial
  - Latin Hypercube
  - Completely random

CATALYSIS CENTER FOR ENERGY INNOVATION

Step 2 – Initialize the Object and Do the Optimization

Code example 2. *Experiment* Object

```
# Set its name, the files will be saved under the folder with the same name
Exp_lhs = bo.Experiment('PFR_yield_lhs')
# Import the initial data
Exp_lhs.input_data(X_init_lhs, Y_init_lhs, X_ranges=X_ranges, unit_flag=True)
# Set the optimization specifications
Exp_lhc.set_optim_specs(objective_func=objective_func,
                        maximize=True)
```

Code example 3. Optimization Loops

```
# Set the number of iterations
n_total = 64
n_trials_lhs = n_total - n_init_lhs
for i in range(n_trials_lhc):
    # Generate the next experiment point
    X_new, X_new_real, acq_func = Exp_lhs.generate_next_point()
    # Get the response at this point
    Y_new_real = objective_func(X_new_real)
    # Retrain the model by input the next point into Exp object
    Exp_lhs.run_trial(X_new, X_new_real, Y_new_real)

# lhc optimum
y_opt_lhs, X_opt_lhs, index_opt_lhs = Exp_lhs.get_optim()
```

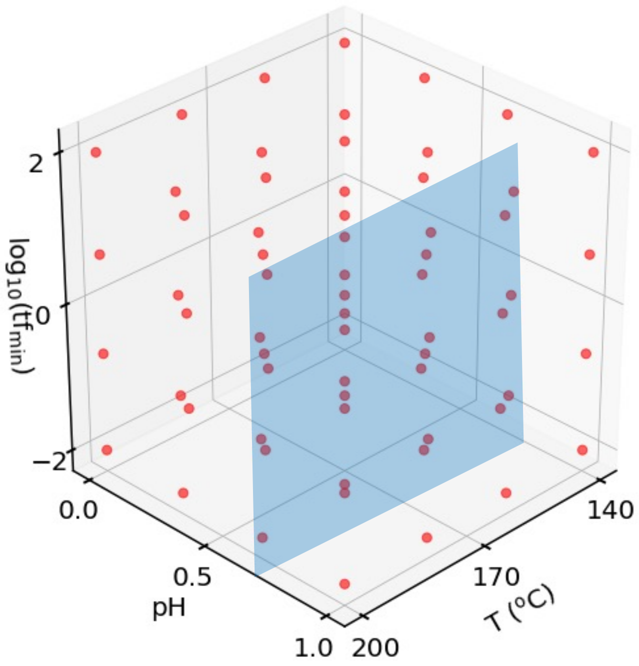
- Set the objective function as PFR model and the goal as maximization
- Perform the optimization using an explicit (human-in-the-loop) format
  - Users can access the values of parameters and responses in real units.

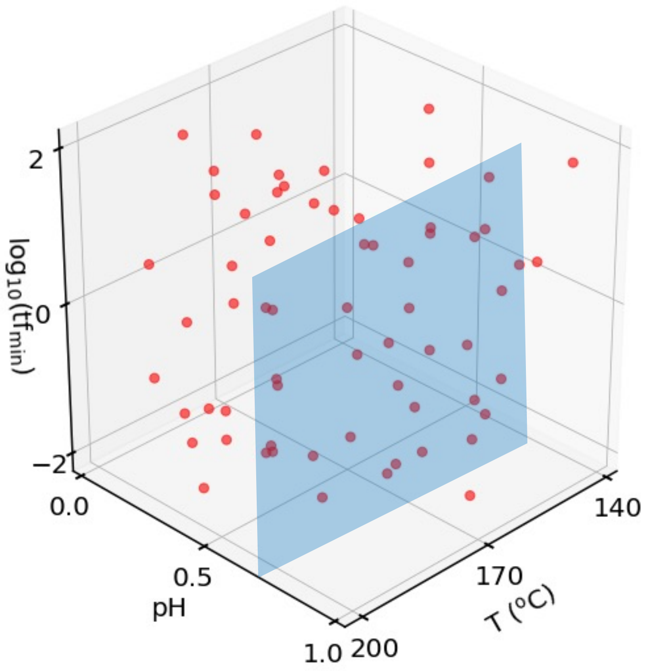
Sampling Plans

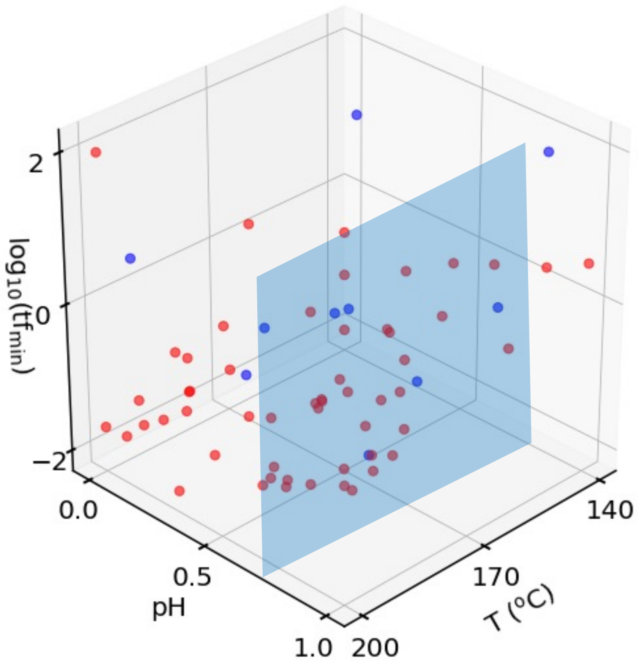
Design 1
Full factorial (level = 4)
64 samples

Design 2
64 random samples

Design 3
10 samples from Latin hypercube (LHS)
54 samples from 54 BO loops,
Acquisition function – EI



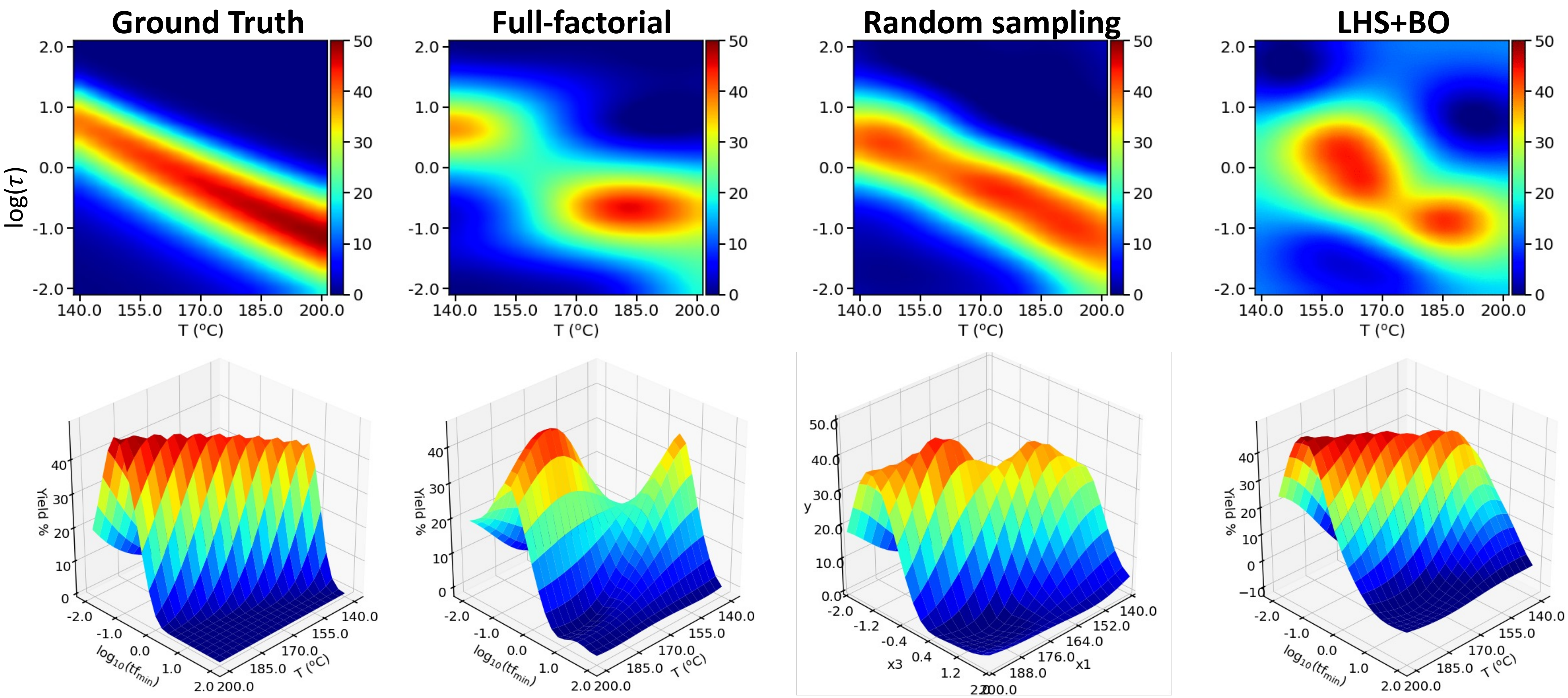




- LHS is an efficient space-filling, Monte Carlo sampling method
- We compare response surfaces of HMF yield at **pH=0.7** with varying temperature and residence time

CATALYSIS CENTER FOR ENERGY INNOVATION

Surrogate Model Performance



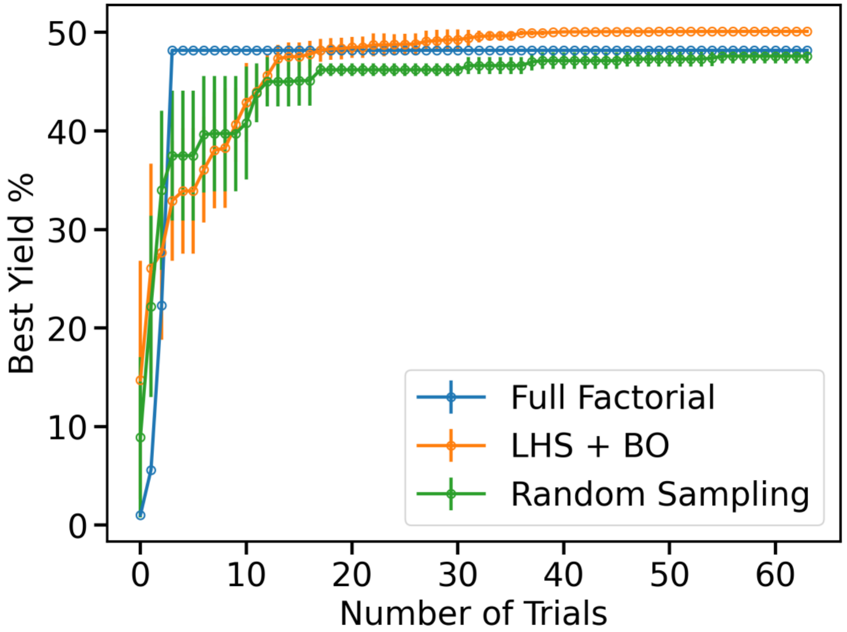
- LHS+BO produces more accurate surrogate models

CATALYSIS CENTER FOR ENERGY INNOVATION

Single-, Multi-Objective Optimization

Maximize HMF Yield

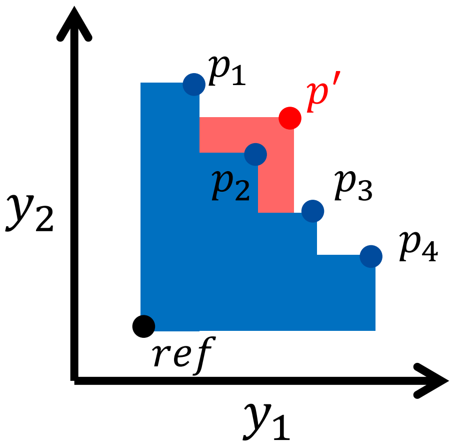
- Optimal condition: Temperature – 200 ° C
  pH – 0.705
  Residence time – 0.076 min (4.56 s)

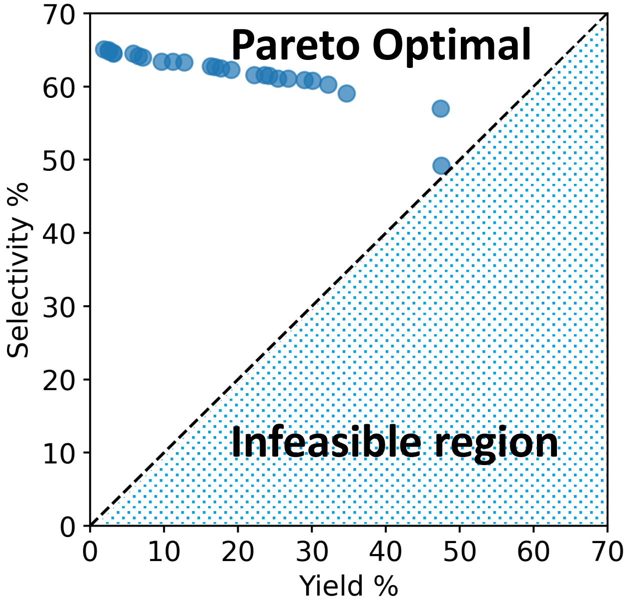


Co-maximize HMF Yield and Selectivity

- HMF Yield = Fructose Conversion × HMF Selectivity
- Fructose Conversion ≤ 100 %

Expected Hypervolume Improvement (EHVI)



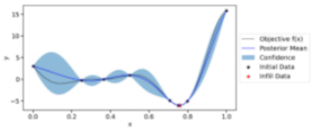


- LHS+BO locates a higher optimal value compared to others
- The runtime of core BO functions completes in seconds per iteration on a laptop CPU
- NEXTorch requires little code and reduces the time or materials for **computations** or **lab experiments**

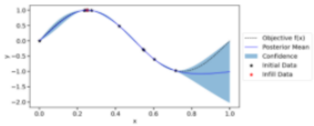
CATALYSIS CENTER FOR
ENERGY INNOVATION

More Examples

Basic API Usage

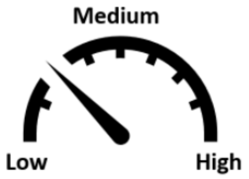


Example 1 - Simple 1d nonlinear function



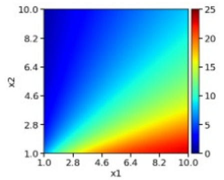
Example 2 - Sin(x) 1d function

Mixed Type Parameters

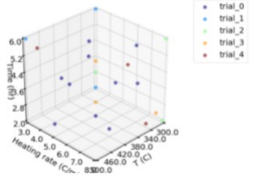


Example 10 - Plug flow reactor yield with mixed type inputs

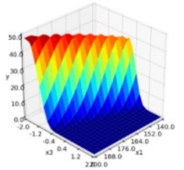
Applications in Reaction Engineering



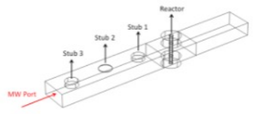
Example 3 - Langmuir Hinshelwood mechanism



Example 4 - Nitrogen-doped carbon catalysts

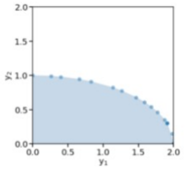


Example 5 - Plug flow reactor yield

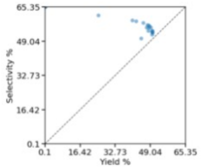


Example 8 - Stub tuner of the microwave cavity

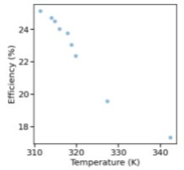
Multi-Objective Optimization(MOO)



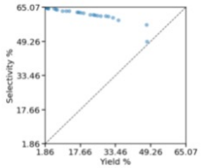
Example 6 - Multi-objective optimization for an ellipse function



Example 7 - Multi-objective optimization for plug flow reactor



Example 9 - Multi-objective optimization for Microwave operating conditions



Example 11 - Multi-objective optimization for plug flow reactor

Documentation: https://nextorch.readthedocs.io/en/latest/index.html