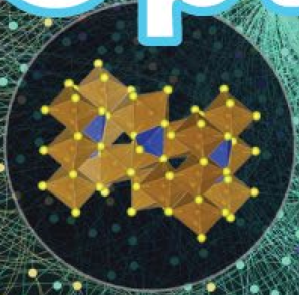
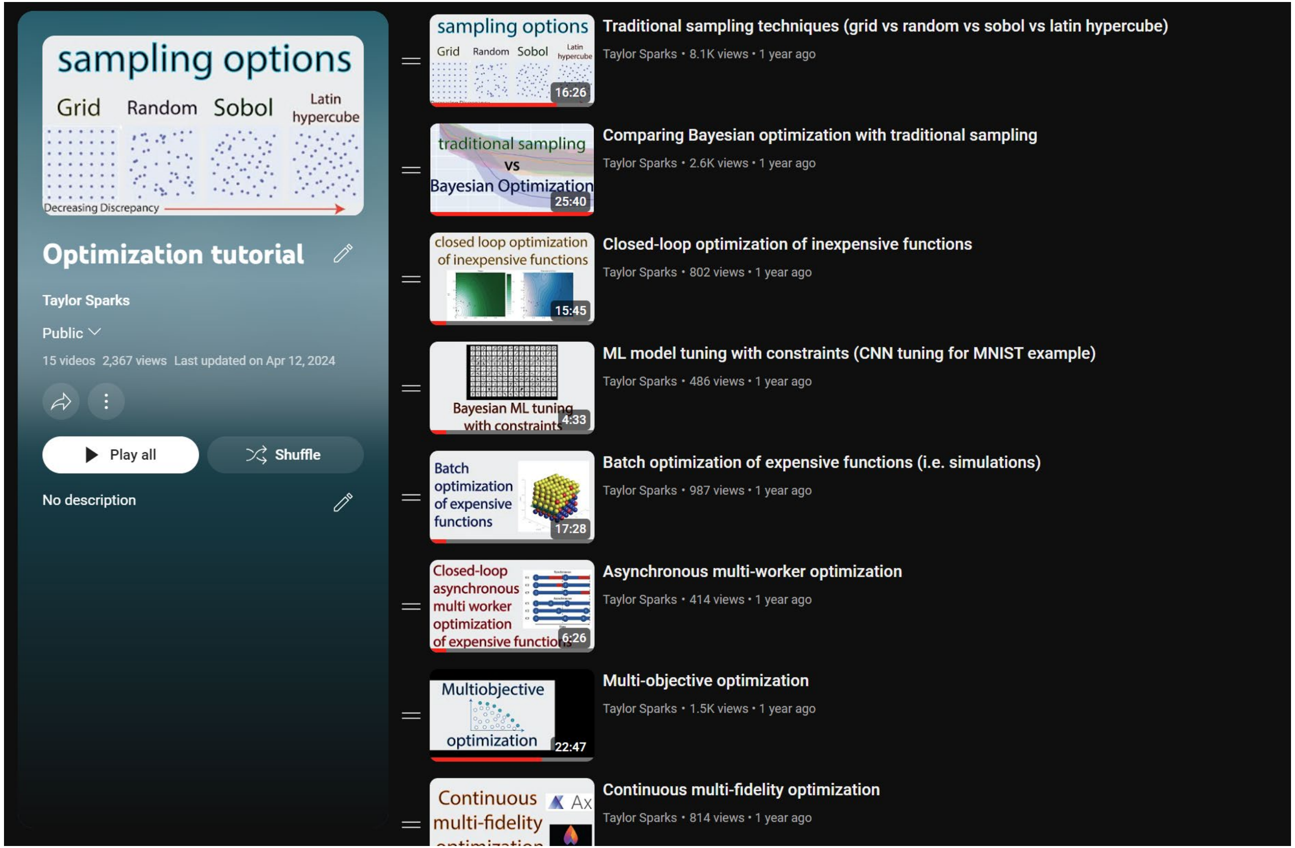
Bayesian Optimization



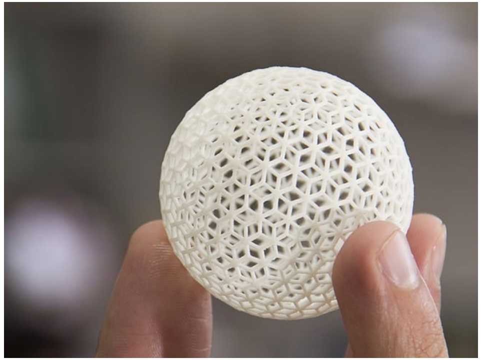
There is A LOT to say about Bayesian Optimization… check out the full playlist!



Materials design is filled with difficult property tradeoff decisions

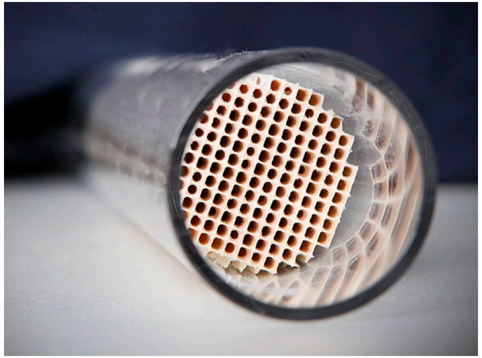
Strength vs ductility

3d print parameters



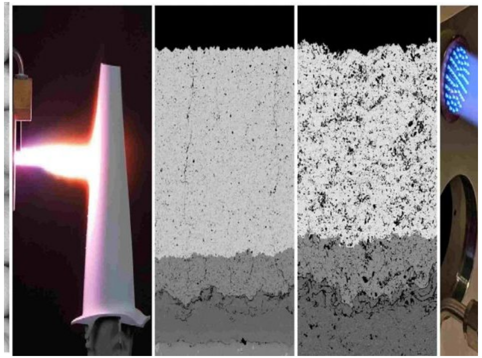
Performance vs biodegradability

Catalytic activity vs selectivity



Thermal vs electrical

Thermal vs mechanical



A motivating example… let’s get that 3d printer up and running!

Your company purchased a brand new 3D printer and wants to use it to fulfill a client request for a custom part. Your manager gives you a week to get the machine dialed in before they print the client’s part.

What is the most efficient way to dial in these params?

- X-offset
- Y-offset
- Prime Delay
- Print Speed

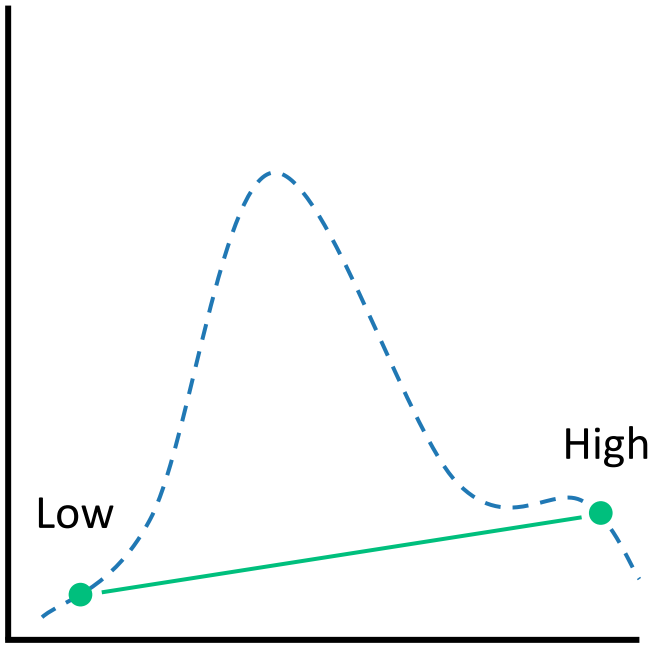
A traditional Design of Experiment (DOE) approach is too slow

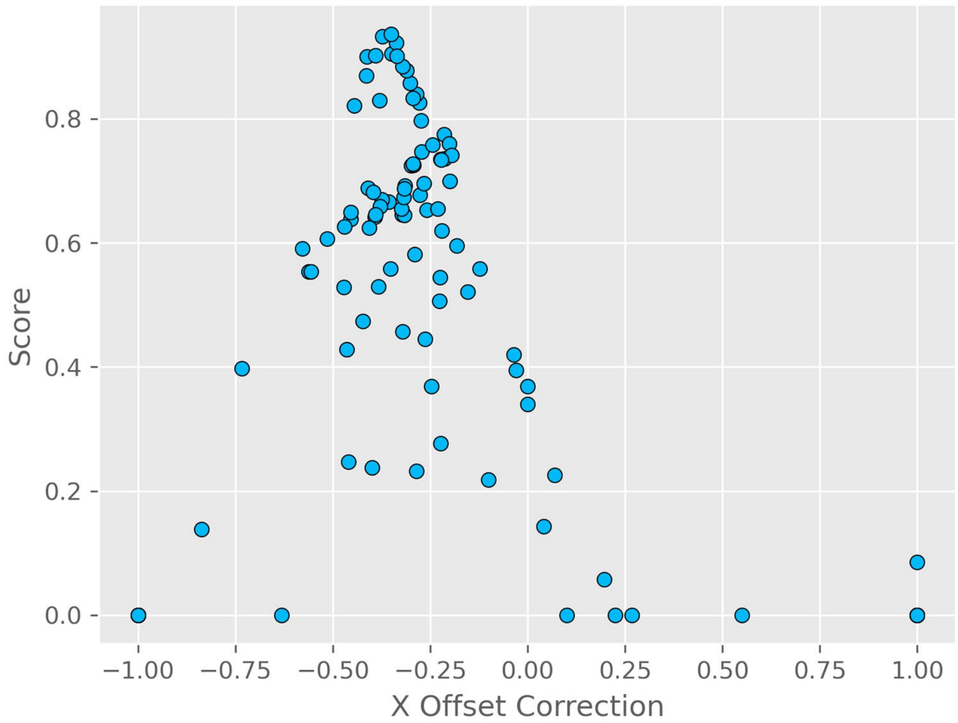
Create statistically orthogonal testing points to learn the impact of each parameter on the printed part properties.

For 2 level DOE approach with 3 replicates:

Full Factorial $3 * 2^4 = 48$ Experiments ← And that's just the beginning!

DOE isn't just slow though, it can miss optimal solutions!

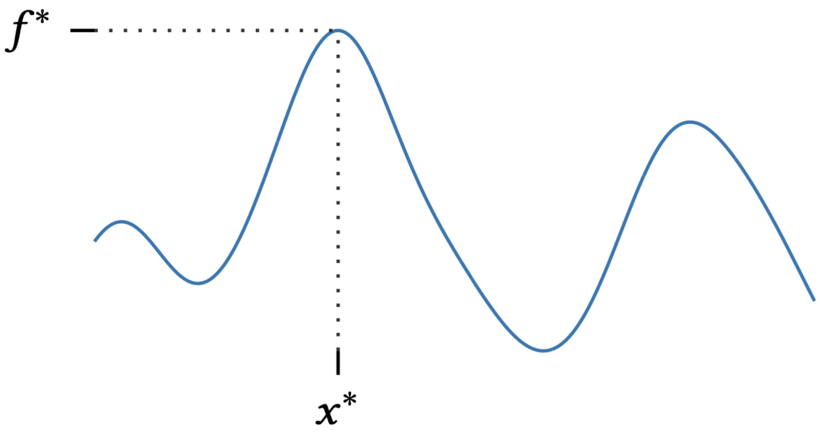




Optimization formulation

objective function, f

domain of objective function, $\mathcal{X}$



$$x^* \in \arg\max_{x \in \mathcal{X}} f(x); \qquad f^* = \max_{x \in \mathcal{X}} f(x) = f(x^*)$$

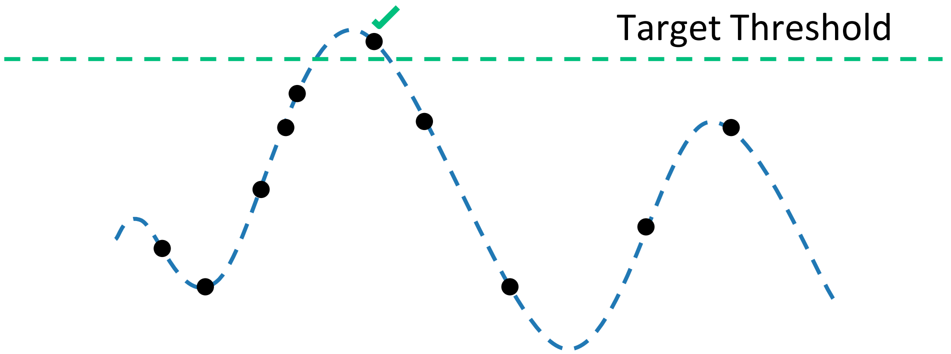
Goal: Systematically search the domain for a point x^* attaining the globally maximal value f^*

Figure Referenced From

Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press

We don't actually need to know f(x) in order to find the optima…

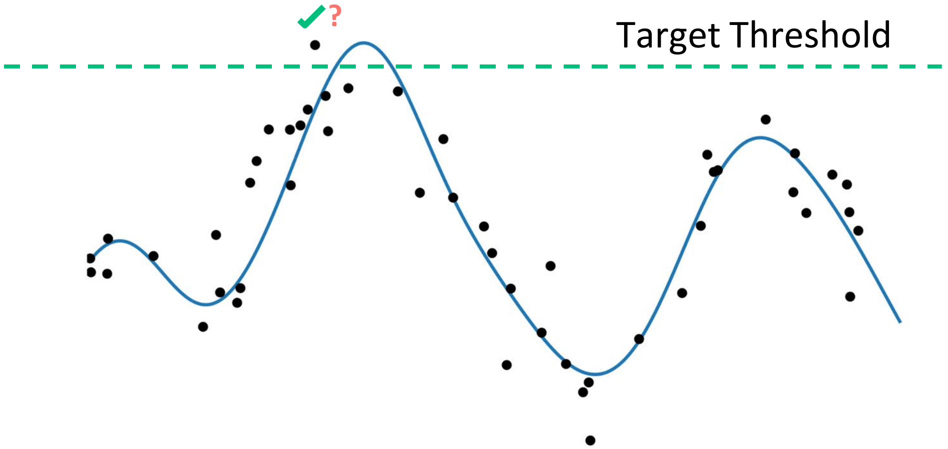
input: initial dataset $\mathcal{D}$ ▸ can be empty
repeat
  $x \leftarrow \text{POLICY}(\mathcal{D})$ ▸ select the next observation location
  $y \leftarrow \text{OBSERVE}(x)$ ▸ observe at the chosen location
  $\mathcal{D} \leftarrow \mathcal{D} \cup \{(x, y)\}$ ▸ update dataset
until termination condition reached ▸ e.g., budget exhausted
return $\mathcal{D}$



You **could** keep trying experiments until you get something above your target threshold.

Table Referenced From
Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press

Measurement noise can derail a simple direct observation approach

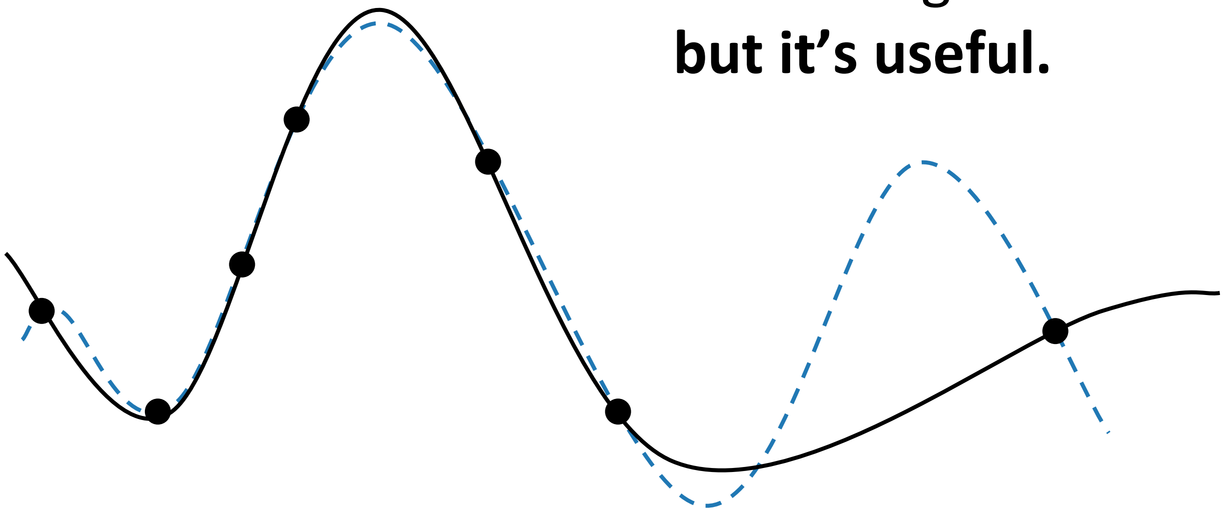


We don’t want to rely on a simple observational model.

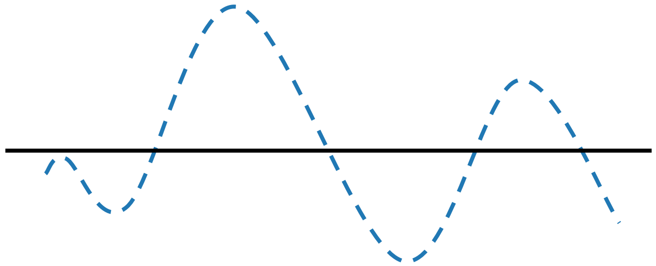
Figure Referenced From
Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press

A surrogate model can help us understand the objective function

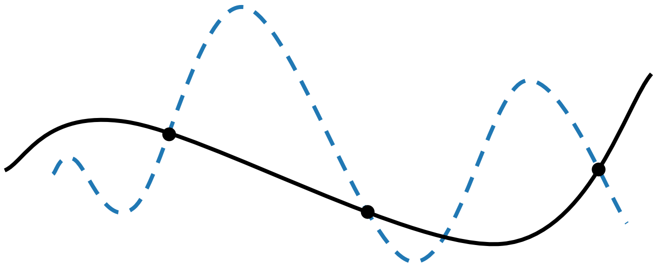
The surrogate model is wrong,
but it's useful.



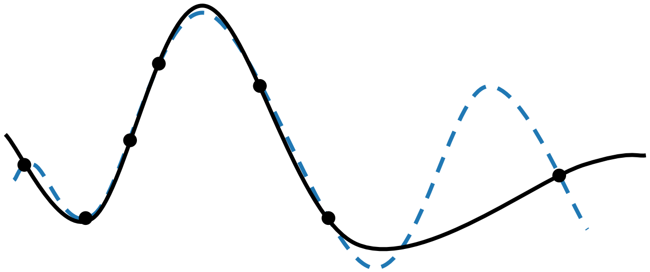
We can improve our understanding with each new data point we collect



Our initial assumption about the objective function



Observing some points gives the function some form.



Our belief about that form changes with more observations.

A Bayesian framework is ideal for an approach where we are constantly updating our model

Consider a situation where we want to model a value y with a distribution ϕ with parameters x

$\phi(x)$

Prior: What do we believe about ϕ before observing data.

Likelihood: Where is ϕ most compatible with the observation y

Posterior: What is the distribution of ϕ having observed y.

$$p(\phi \mid x, y) = \frac{p(\phi \mid x)\, p(y \mid x, \phi)}{p(y \mid x)}.$$

Bayes approach makes more sense visually

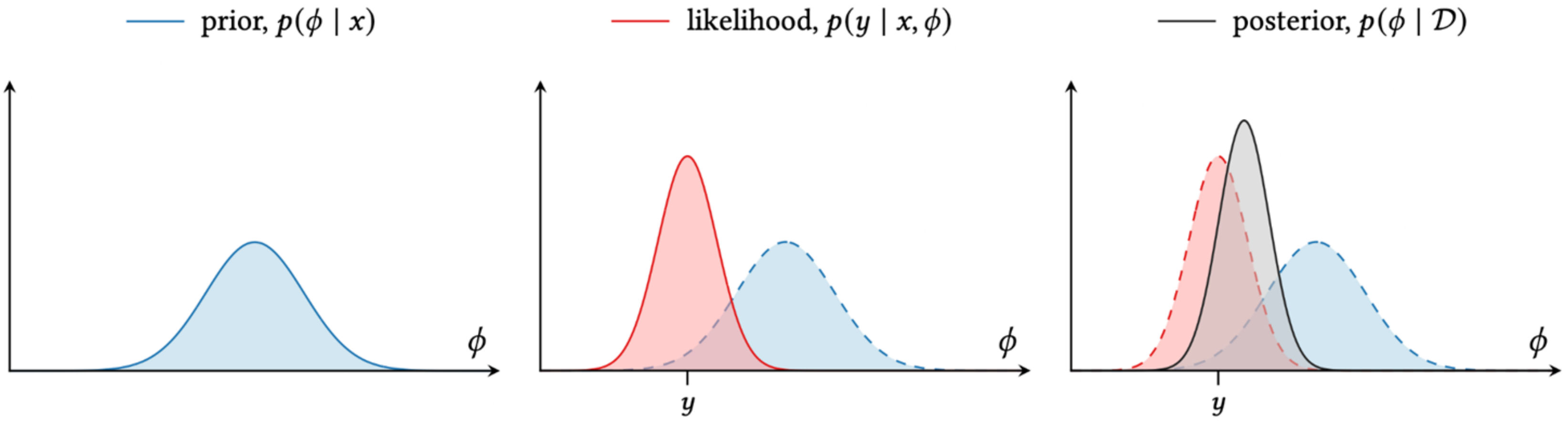
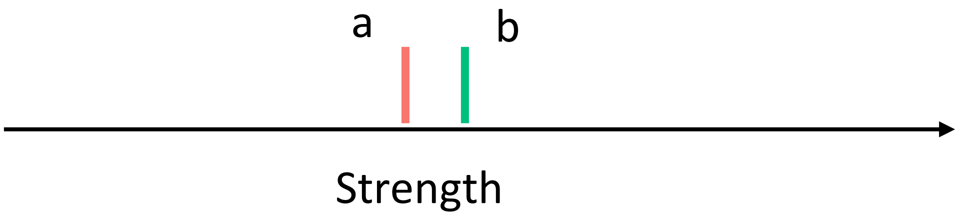


Figure Referenced From

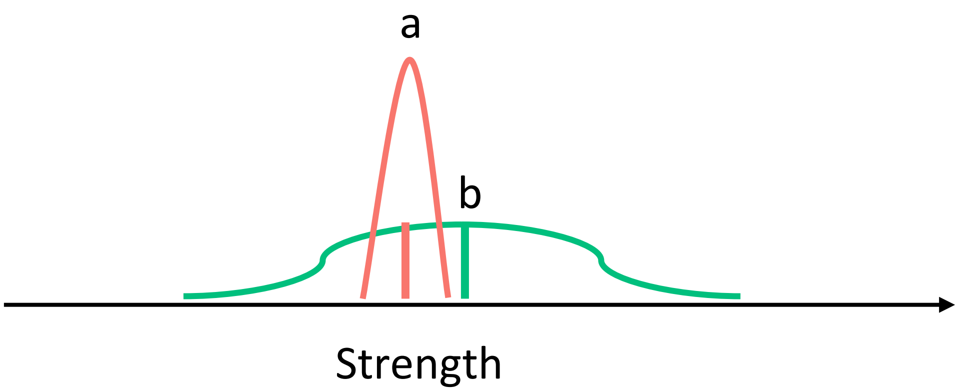
Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press

Uncertainty changes the decision-making process



Which material would you choose for a bridge support?

Uncertainty changes the decision-making process



Which material would you
choose for a bridge support?

Gaussian Processes are flexible, probabilistic surrogate models

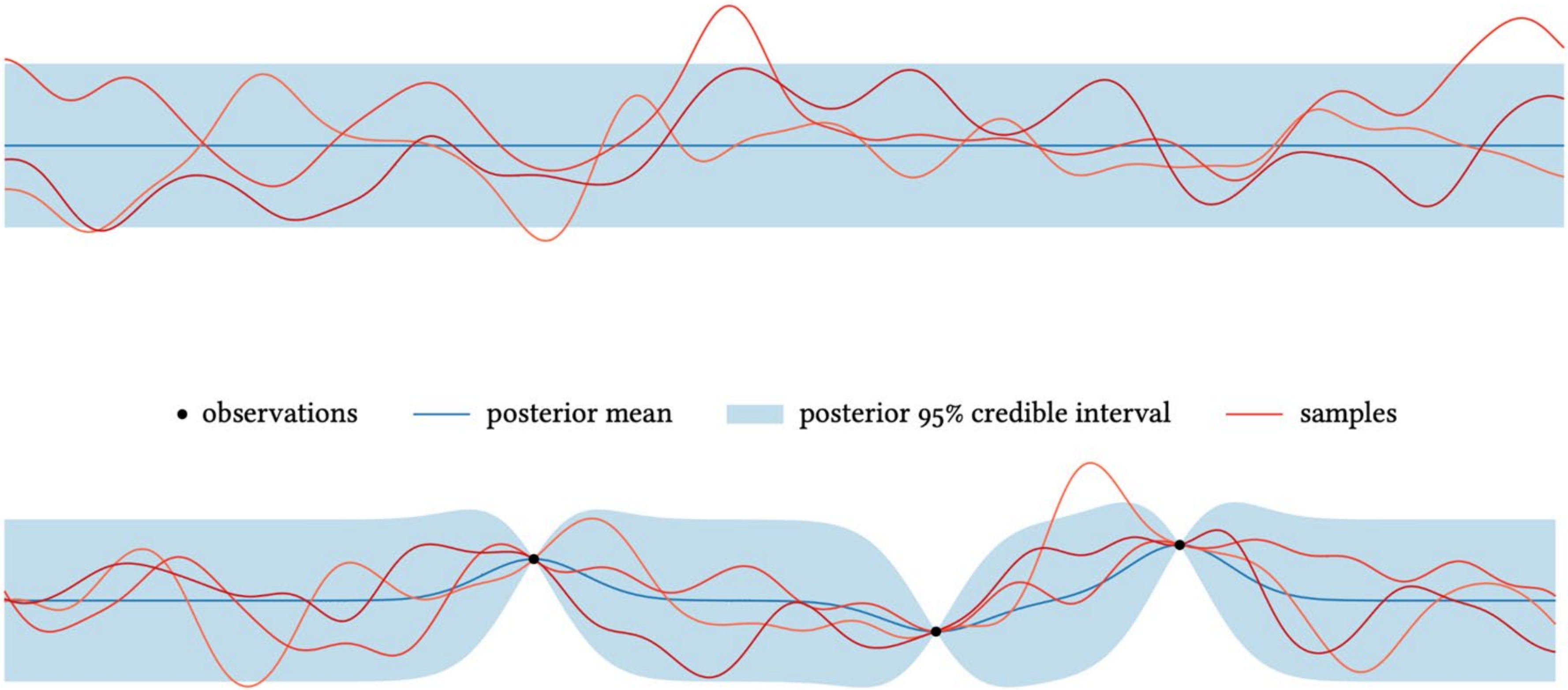
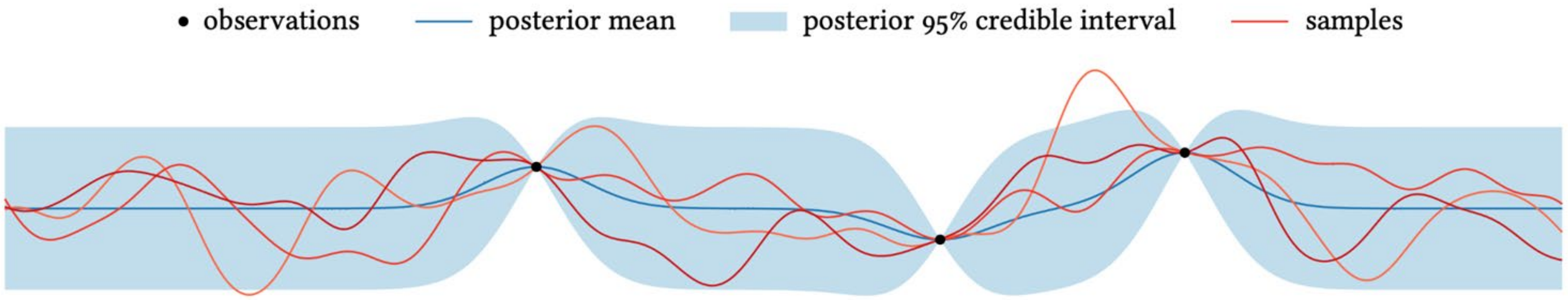


Figure Referenced From
Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press

How does uncertainty change our selection of the next data point?



Where would you look next? why?

Figure Referenced From
Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press

Optimization is a balance of exploration and exploitation

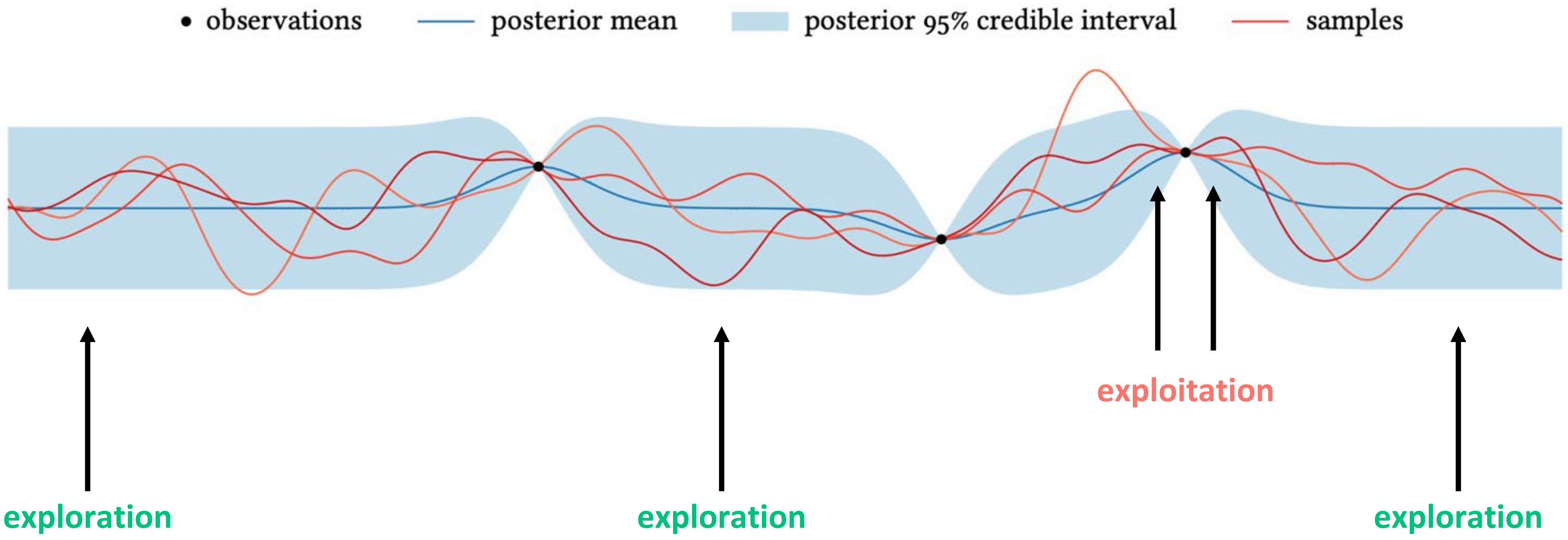
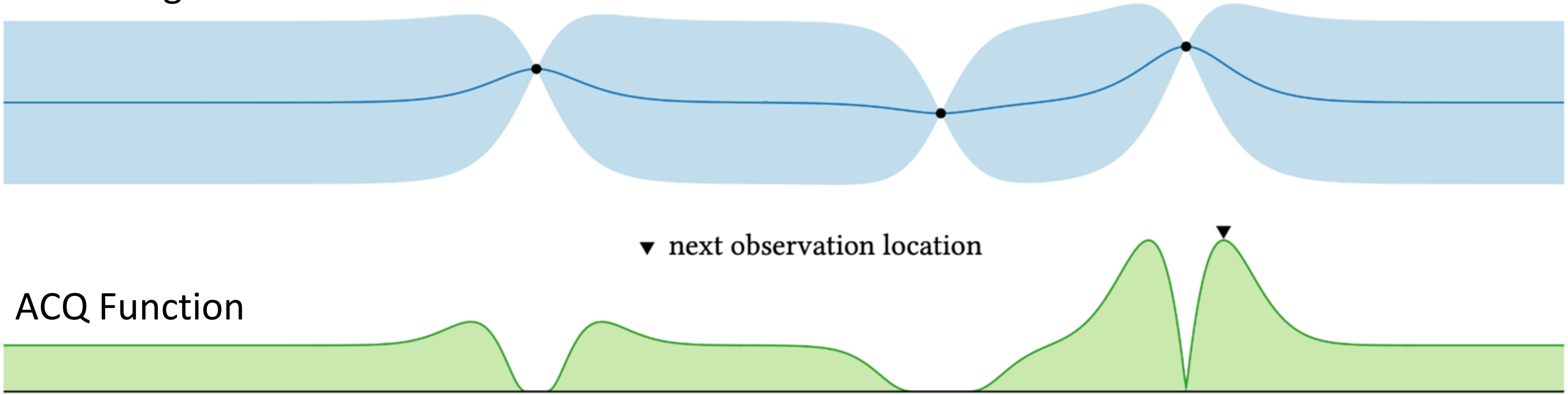


Figure Referenced From
Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press

Enter the acquisition function: a guide for what to do next

GP Surrogate

ACQ Function



The acquisition function applies preference to the value of our surrogate model at any given point.

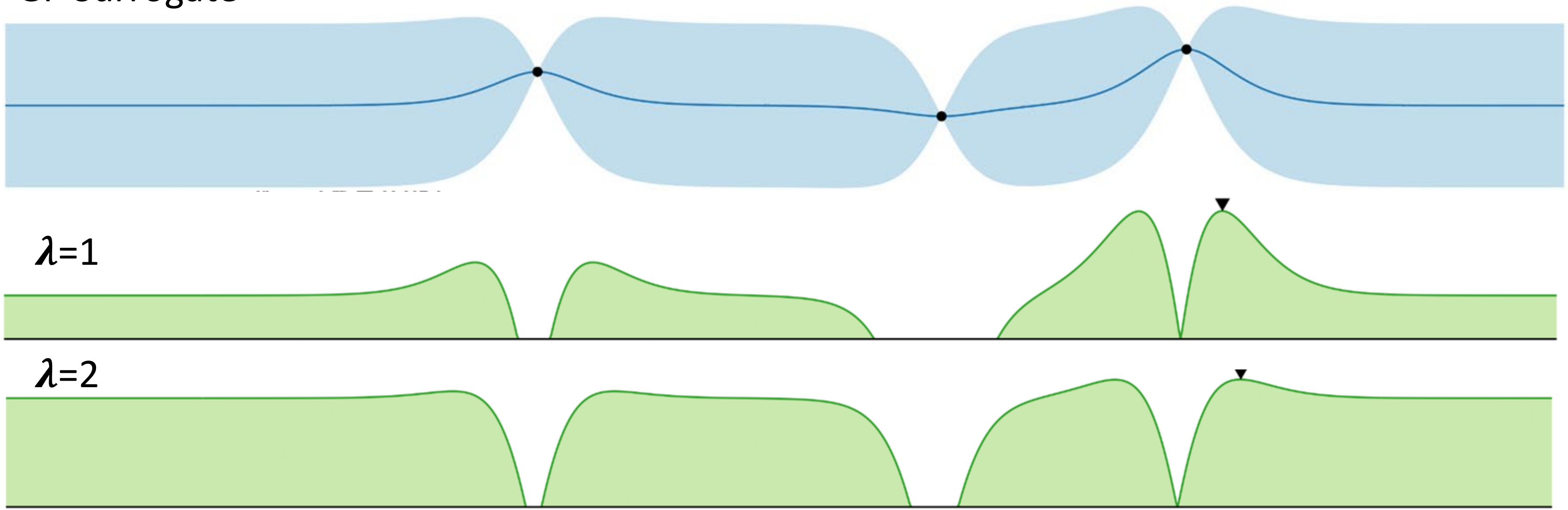
Figure Referenced From
Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press

Upper confidence bound is one popular acquisition function

$$a(x;\lambda) = \mu(x) + \lambda\sigma(x)$$

Mean + Uncertainty

GP Surrogate



λ=1

λ=2

Figure Referenced From: *Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press*

Expected Improvement is another very popular acquisition function

$$\text{EI}(x) = \left(\mu - f(x^\star)\right)\Phi\left(\frac{\mu - f(x^\star)}{\sigma}\right) + \sigma\varphi\left(\frac{\mu - f(x^\star)}{\sigma}\right)$$

Mean + Uncertainty

GP Surrogate

ACQ Function

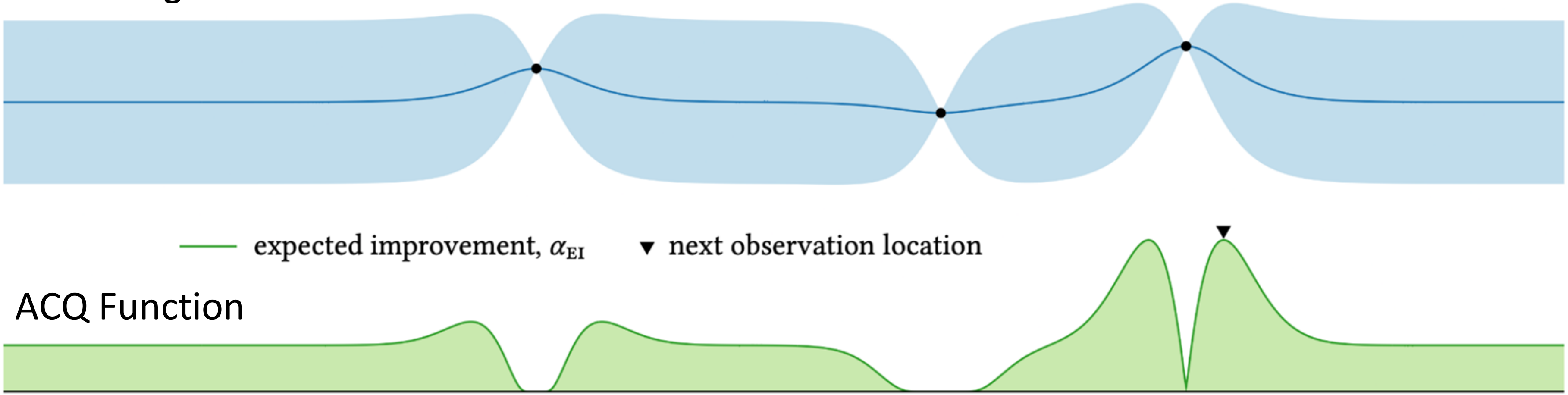


Figure Referenced From: *Garnett, Roman. Bayesian Optimization. 2023. Cambridge University Press*

The acquisition function is just an opinion… and there are many!

Probability of Improvement
Knowledge Gradient
Mutual Information
Thompson Sampling
Entropy Search
Max-Value Entropy Search
…

The data you start with is also critical

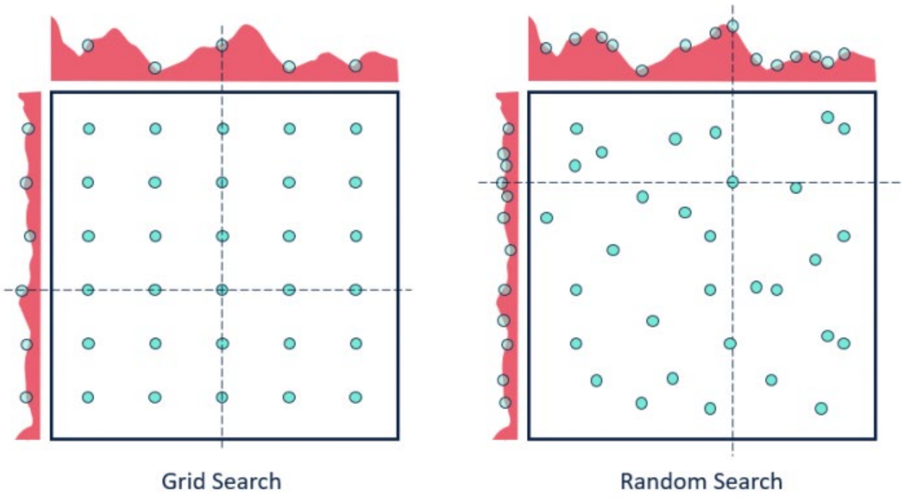
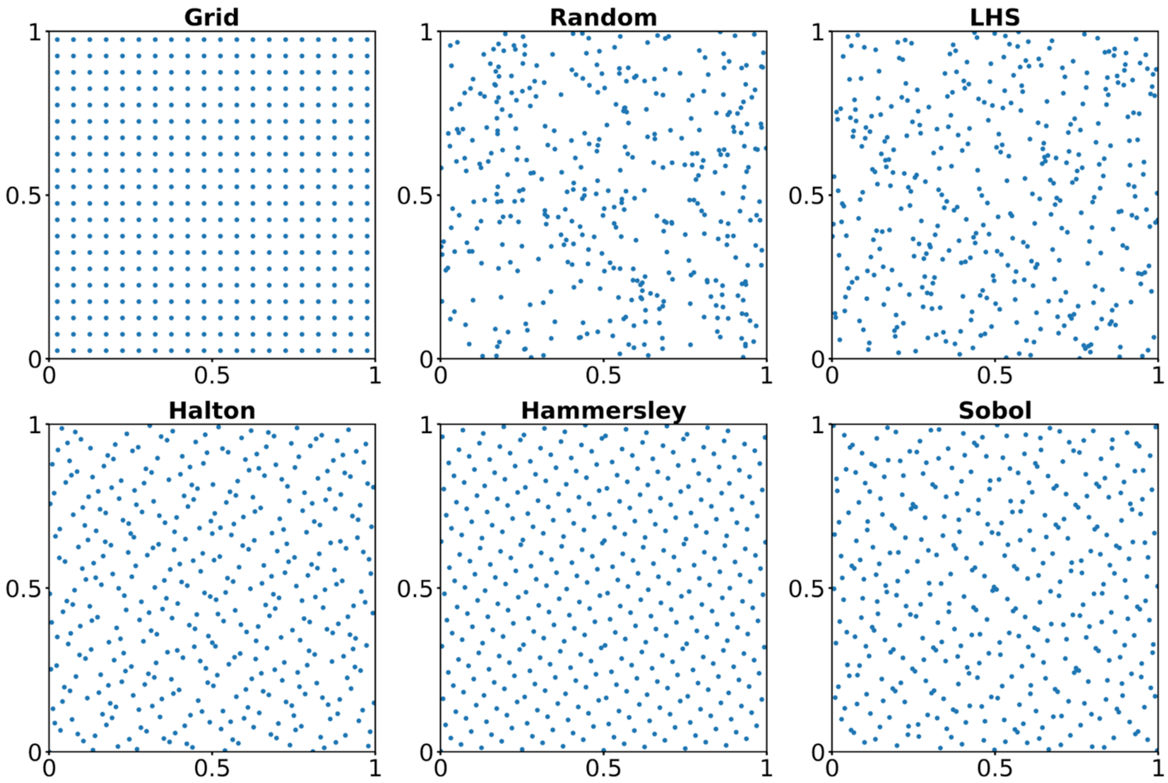


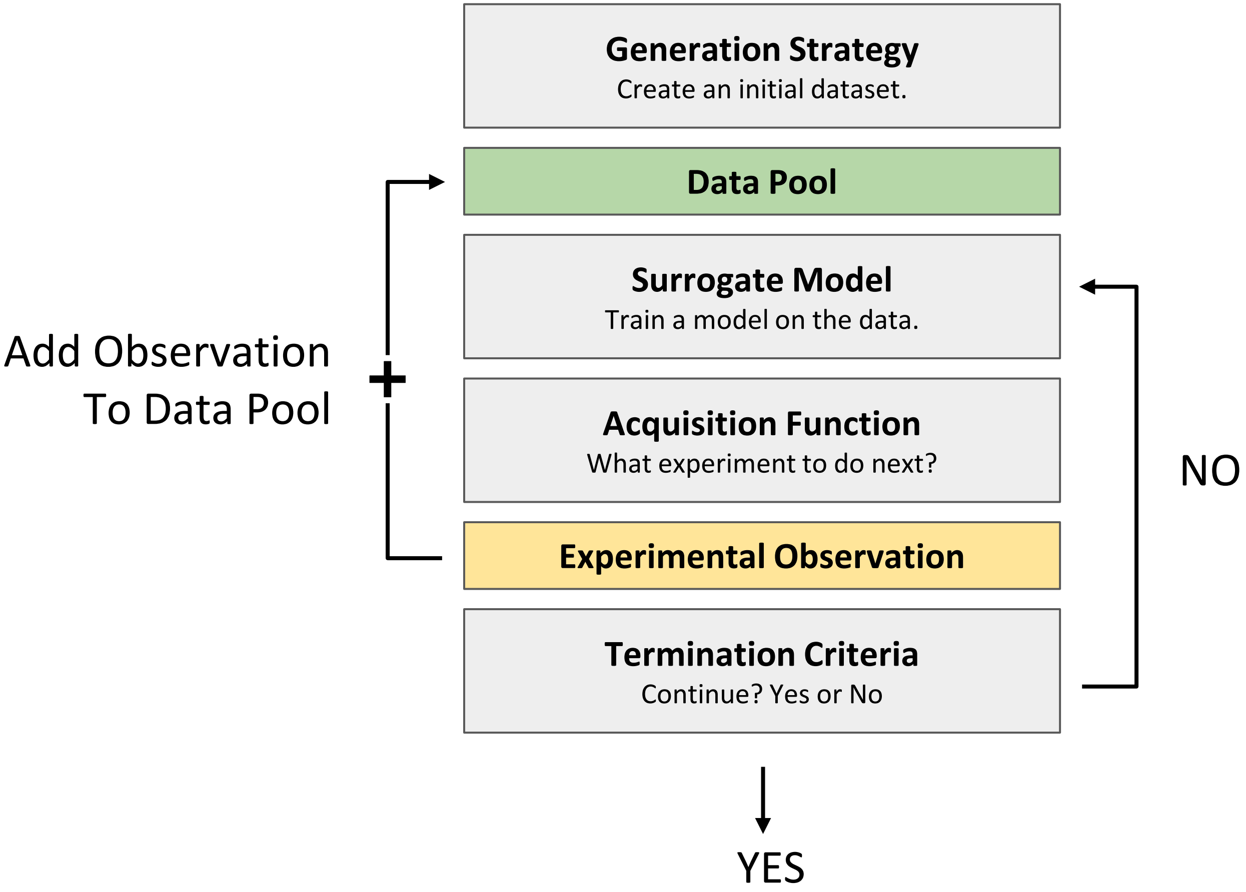
Figure Referenced From:
https://larevueia.fr/3-methodes-pour-optimiser-les-hyperparametres-de-vos-modeles-de-machine-learning/

Quasi-random sequences offer pretty good starting points



Wu C, et al. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. Computer Methods in Applied Mechanics and Engineering. 2023

Let’s pause and make sure we understand the key blocks in Bayesian Optimization



What do we do when we introduce a second optimization criteria?

Maximize Strength
Minimize Cost

The highest strength material is probably really expensive, and the cheapest material is probably weak.

What is optimal?
How much are you willing to spend for an increase in strength?

Scalarization is one approach to optimize more than one objective

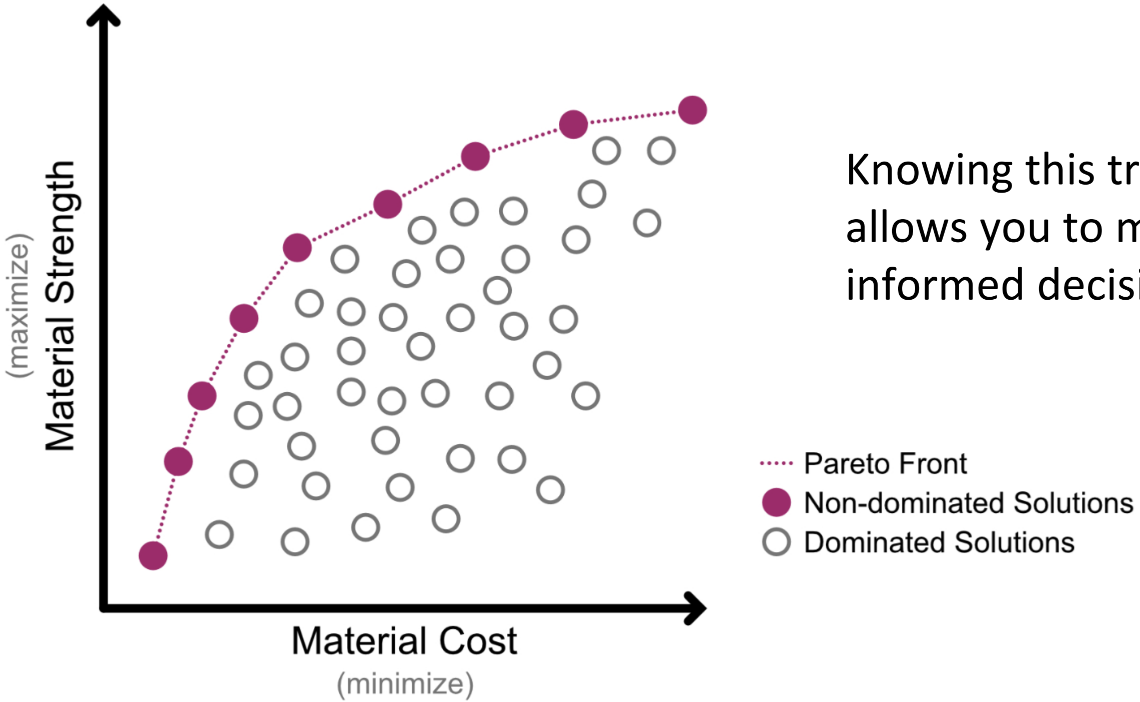
Score = (λ)(Strength Sore) + (1-λ)(Cost Score)

Where λ is the relative weight of each score

This can be as simple or as complex as you want.

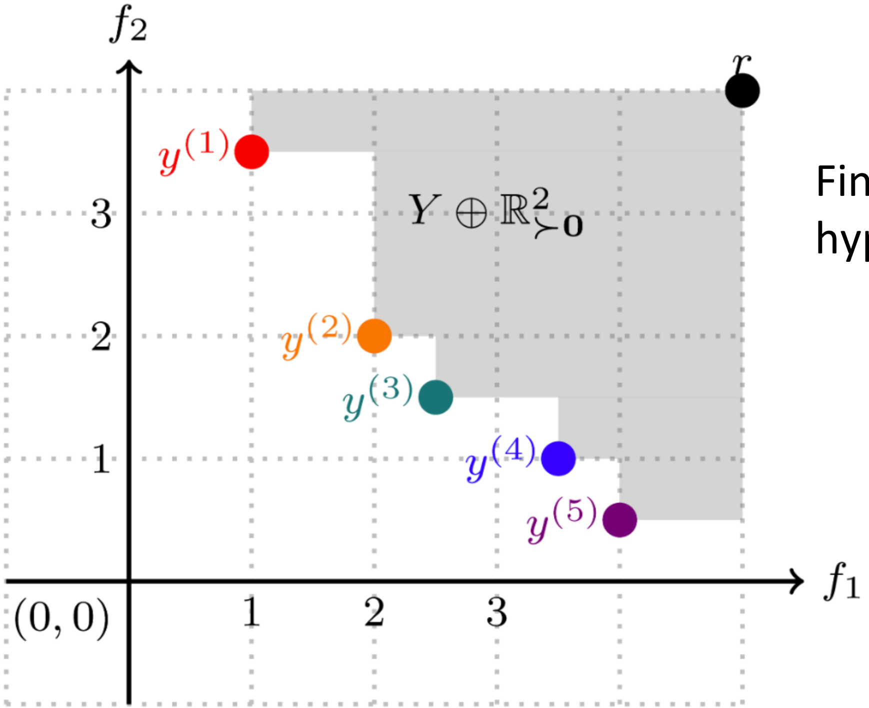
What is the strength and weakness of this approach?

The pareto front gives you the line of maximum tradeoff



Knowing this tradeoff allows you to make informed decisions.

“Hypervolume” allows us to extend the approach to n-dimensions

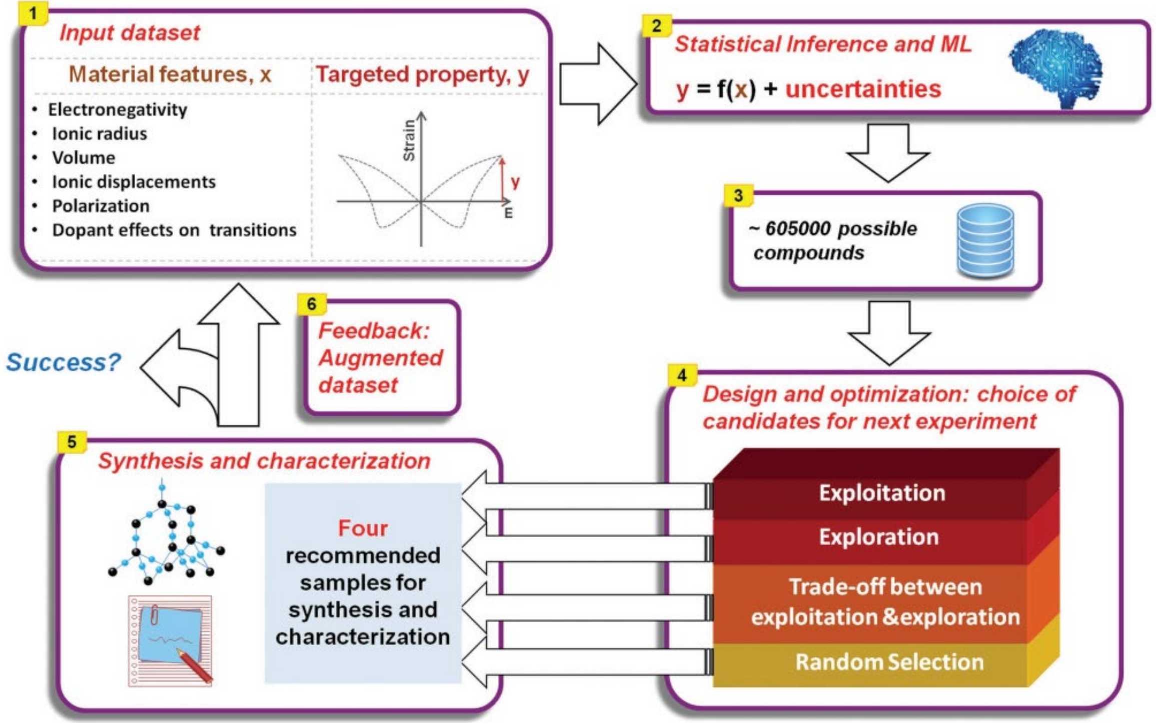


Find points that will maximize the hypervolume.

Figure Referenced From: *https://ax.dev/tutorials/multiobjective_optimization.html*

Example: SVM/inference predicts $BaTiO_3$-based piezoelectric with largest electrostrain (0.23%)

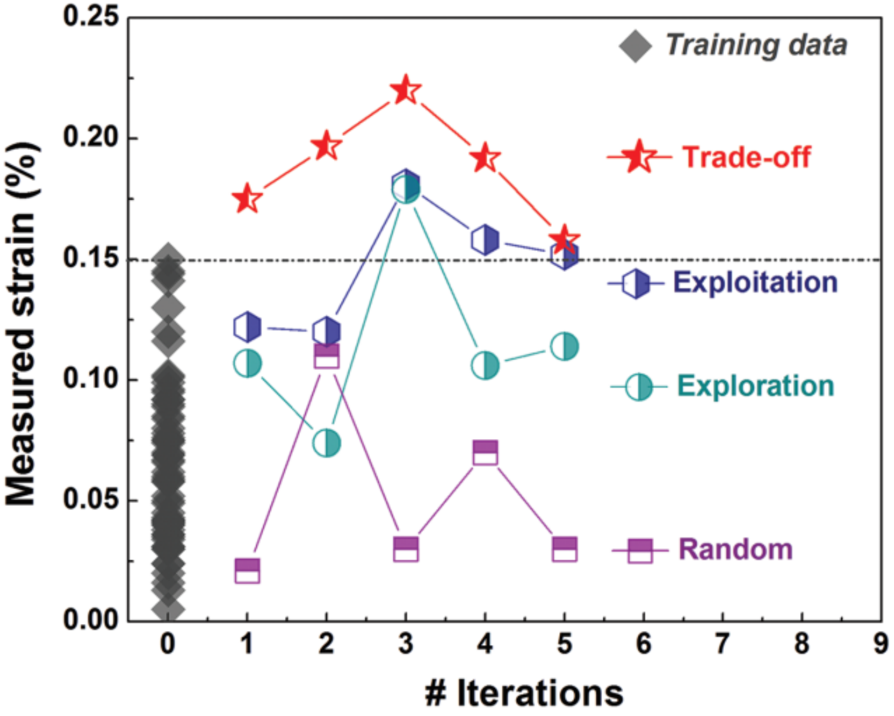
~50 training datapoints & five iterations, adaptive design with trade-off of high uncertainty and best predictions (Efficient Global Optimization)



Yuan et al. *Advanced Materials*, 30 (7) 2018

SVM/inference predicts $BaTiO_3$-based piezoelectric with largest electrostrain (0.23%)

~50 training datapoints & five iterations, adaptive design with trade-off of high uncertainty and best predictions (Efficient Global Optimization)



Yuan et al. *Advanced Materials*, 30 (7) 2018

How to implement these?

ADAPTIVE EXPERIMENTATION PLATFORM

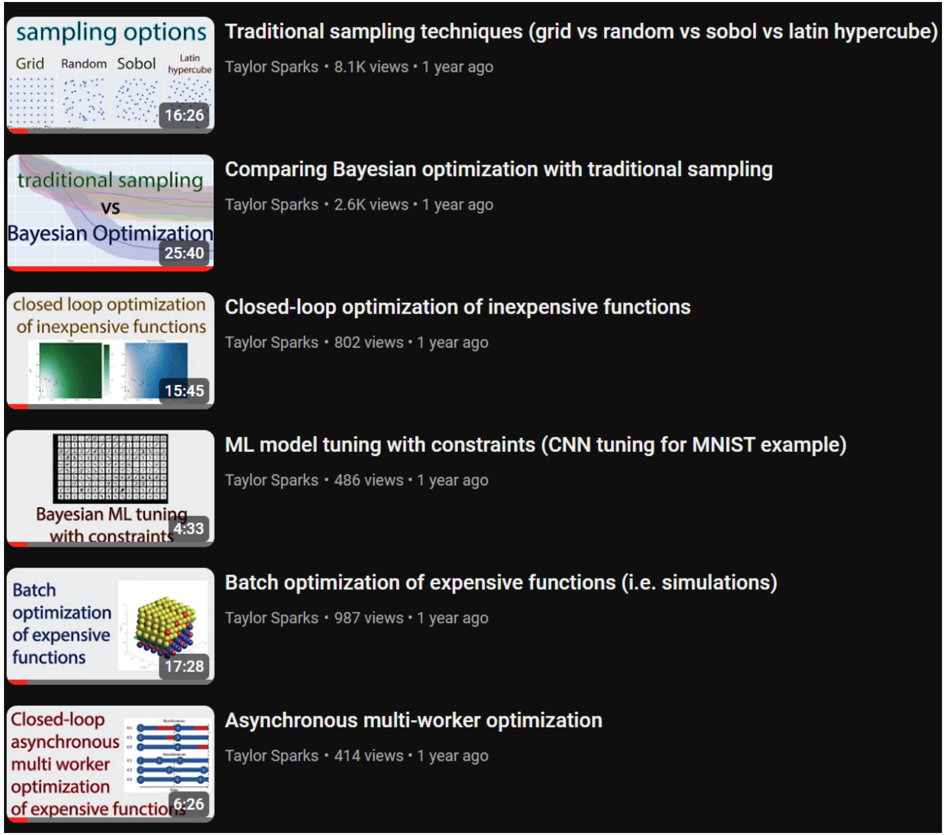
Dragonfly
Scalable Bayesian Optimisation

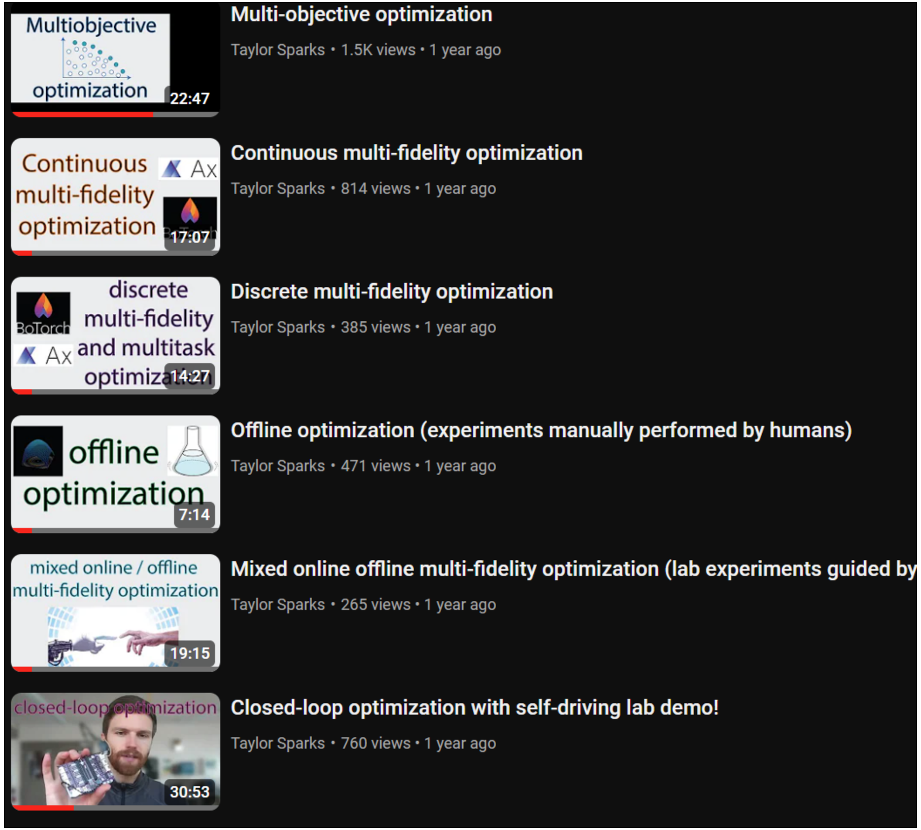
BoTorch
BAYESIAN OPTIMIZATION IN PYTORCH

BayBE
Bayesian Back End

Or you can do it by hand!!

My former student Sterling Baird has an amazing series on BO in much more detail





Large Language Models in Materials