Intelligent Experimentation In Artificial Intelligence

Tobias Andreasen - Staff Engineer

intel | SIGOPT

Enable Modelers Everywhere To Accelerate And Amplify Their Impact With An Intelligent Experimentation Platform

Motivation:

- What Is Intelligent Experimentation
- How To Solve An Optimization Problem
- How To Design An Optimization Problem
- Putting It All Together Using SigOpt
- <u>Pop Quiz With Prices</u>

What Is Intelligent Experimentation

Experimentation

"…provides insight into cause-and-effect by demonstrating what outcome occurs when a particular component is manipulated…"

Experimentation

"…provides insight into cause-and-effect by demonstrating what outcome occurs when a particular component is manipulated…"

Intelligent Experimentation

"…makes recommendations, based on design decisions, on how to explore particular components in order to find the optimal solution(s)…"

Design Explore Optimize

Design

Explore

Optimize

Design

<u>Explore</u>

Optimize

Design Explore <u>Optimize</u>

How To Solve An Optimization Problem

Optimize

Black Box Optimization

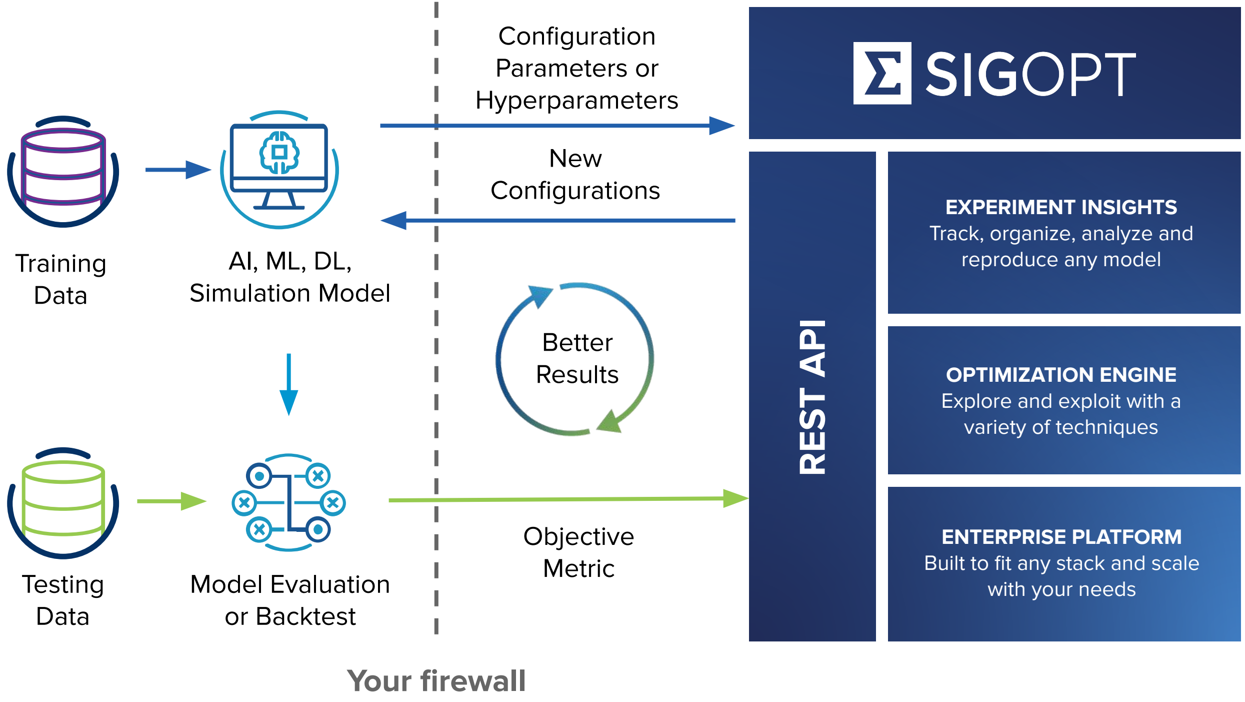
Why Black Box Optimization?

SigOpt is designed to empower you, the expert, to re-define machine learning problems as black box optimization problems with the benefit of:

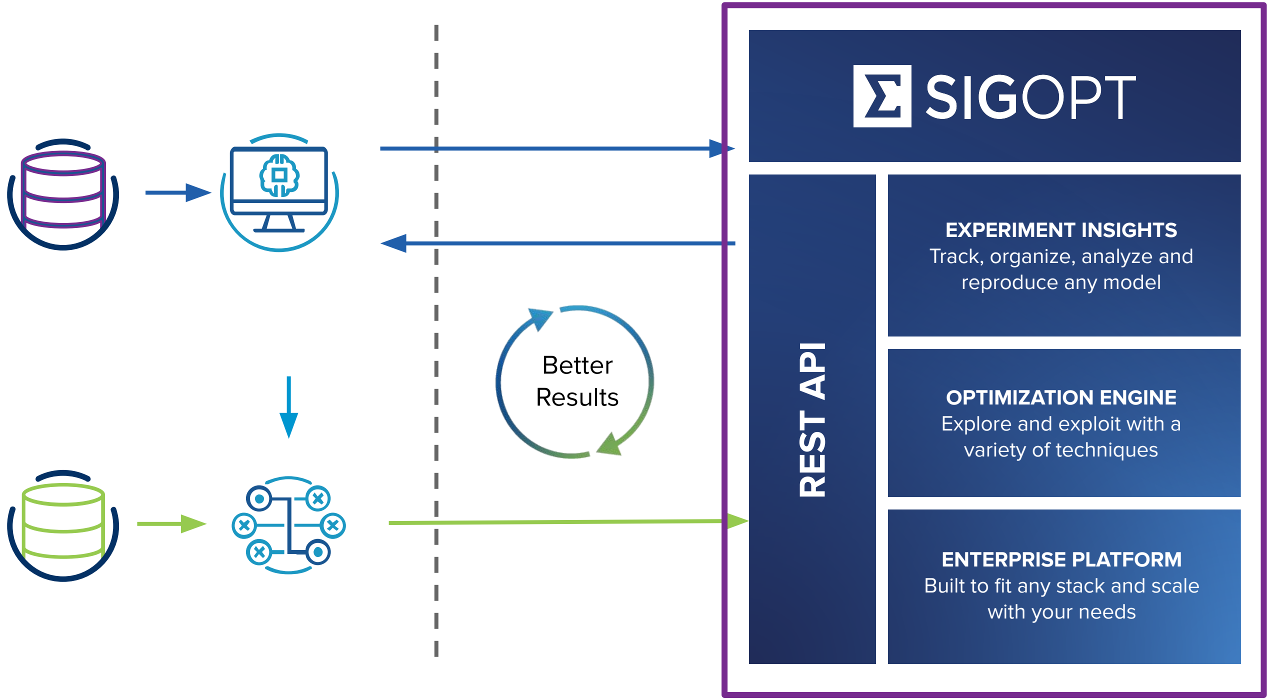
- **Amplified Performance** — incremental gains across success metrics
- **Productivity Gains** — a standard modeling platform for optimizing and other tasks
- **Accelerated Modeling** — early elimination of non-scalable tasks
- **Compute Efficiency** — continuous, full utilization of infrastructure

SigOpt uses an ensemble of Bayesian and Global Optimization methods to solve these black box optimization problems.

Black Box Optimization

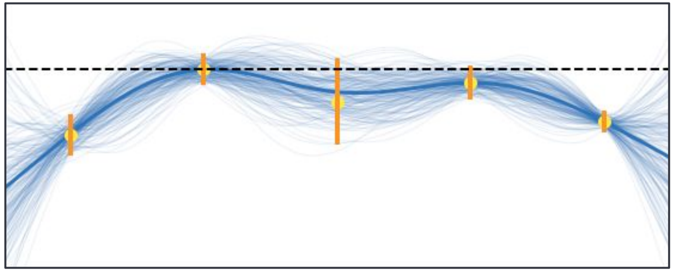


Black Box Optimization



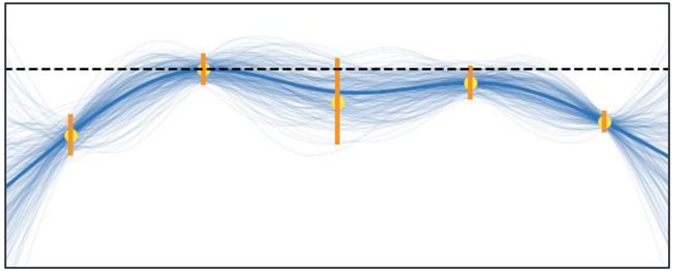
Sequential Model Based Optimization

Build a statistical model

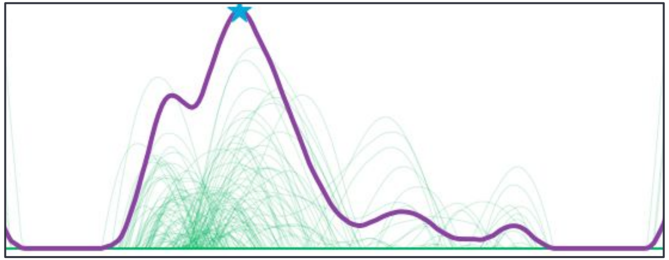


Sequential Model Based Optimization

Build a statistical model

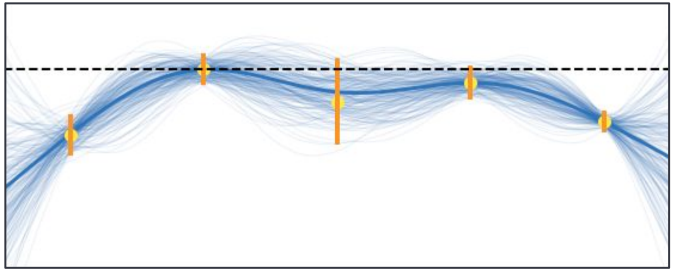


Choose the next point
to maximize the acquisition function

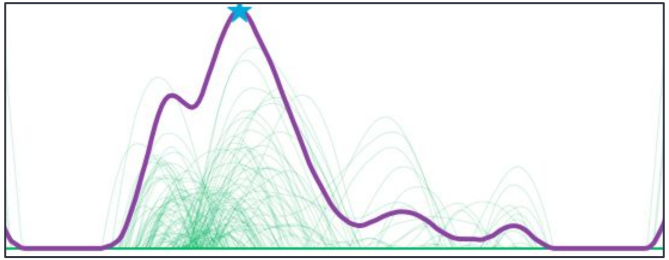


Sequential Model Based Optimization

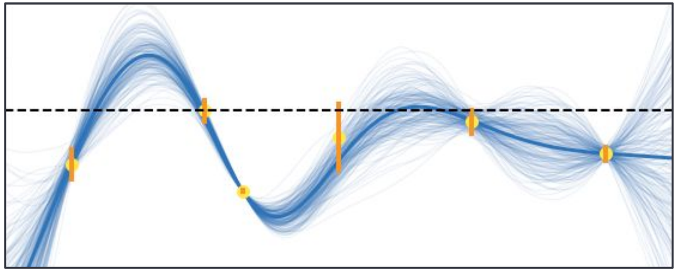
Build a statistical model



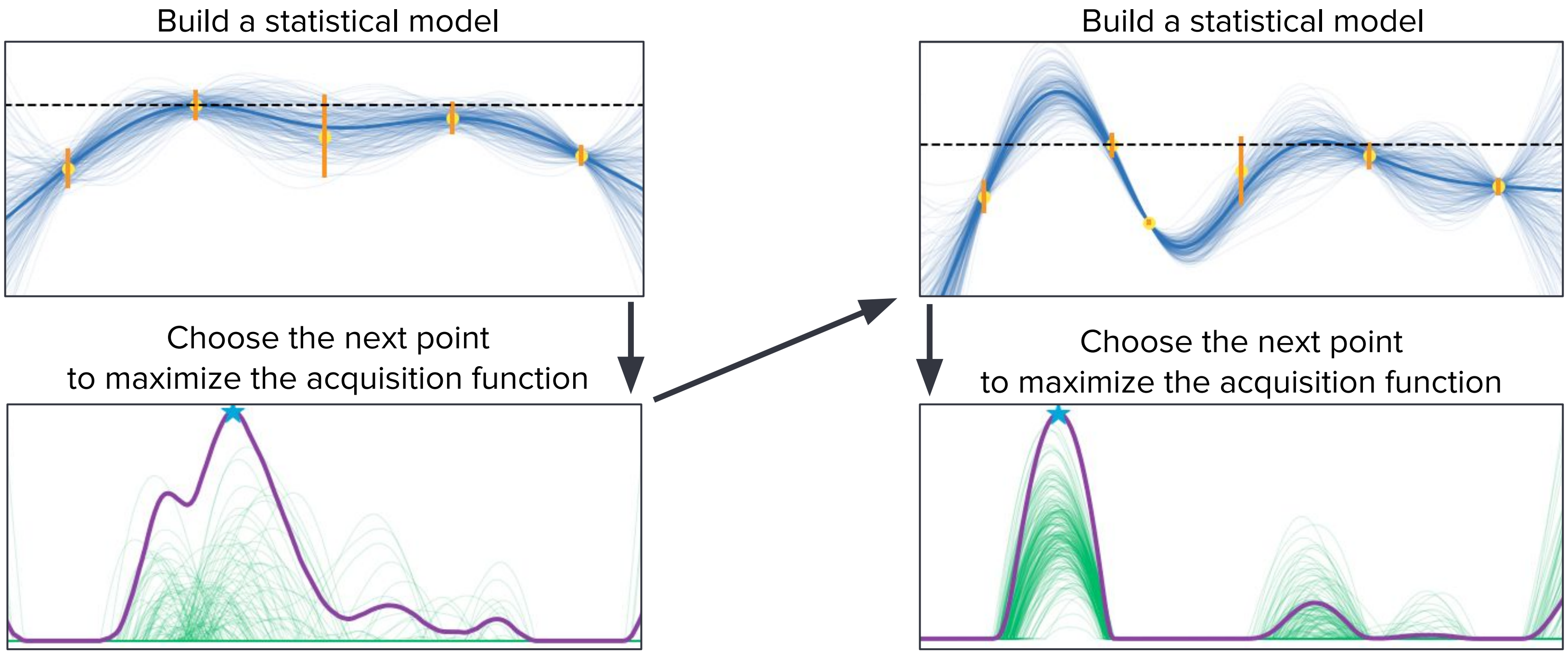
Choose the next point
to maximize the acquisition function



Build a statistical model



Sequential Model Based Optimization



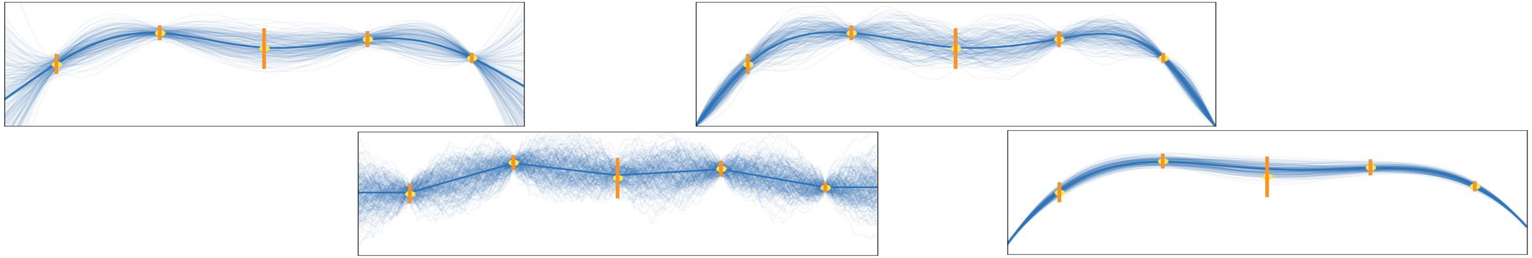
Complexity #1: Find The Right Model

Gaussian processes: a powerful tool for modeling in spatial statistics

A standard tool for building statistical models is the Gaussian process [Fasshauer et al, 2015, Fraizer, 2018].

- Assume that function values are jointly normally distributed.
- Apply prior beliefs about mean behavior and covariance between observations.
- Posterior beliefs about unobserved locations can be computed rather easily.

Different prior assumptions produce different statistical models:



Complexity #2: What To Do Next?

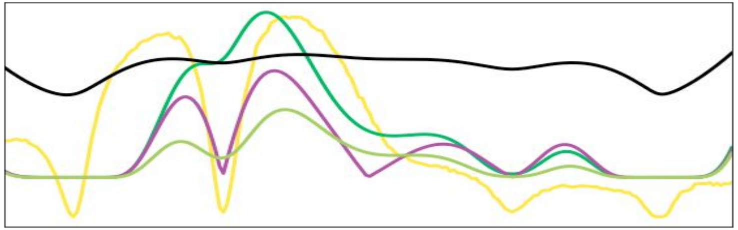
Acquisition function: given a model, how should we choose the next point?

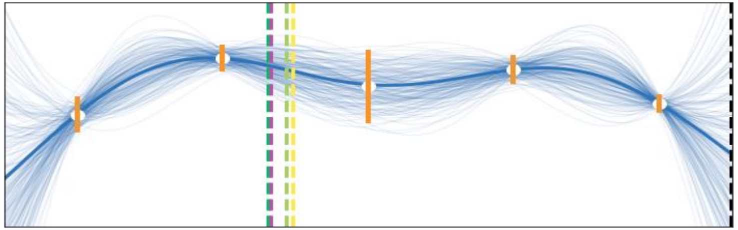
An acquisition function is a strategy for defining the utility of a future sample, given the current samples, while balancing exploration and exploitation [Shahriari et al, 2016].

Exploration:
Learning about the whole function *f*

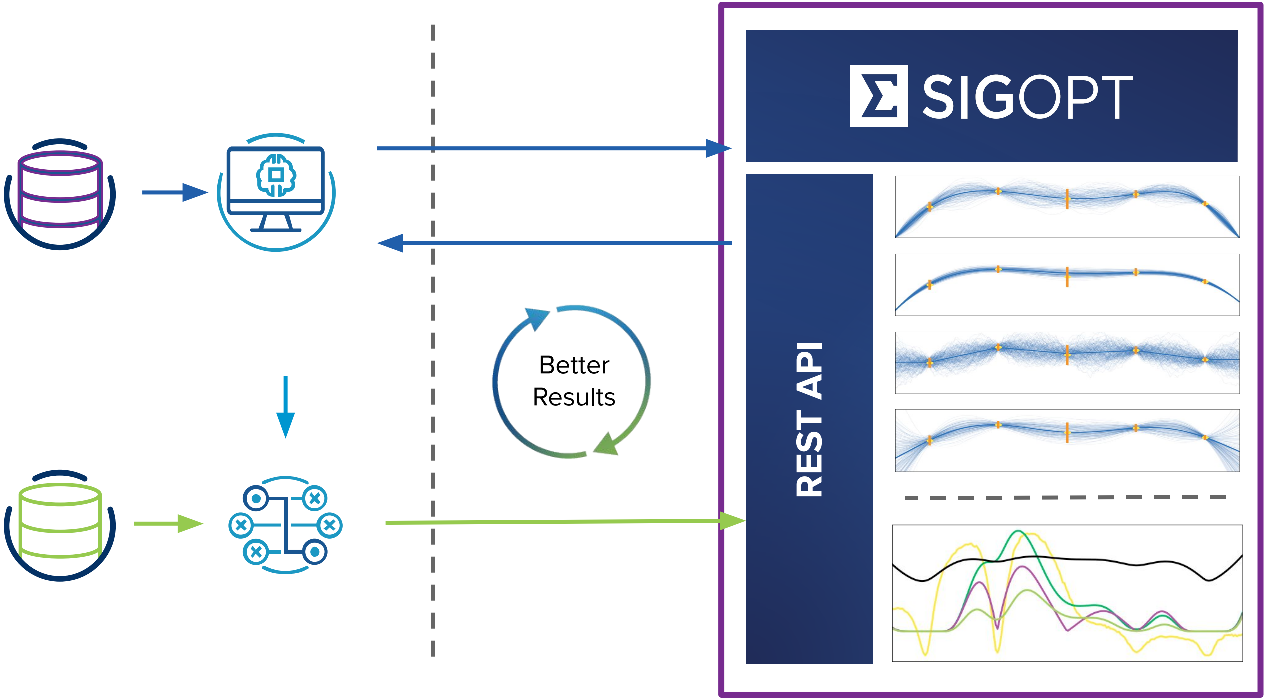
Exploitation:
Further resolving regions where good *f* values have already been observed

Different acquisition functions choose different points (EI, PI, KG, etc.).

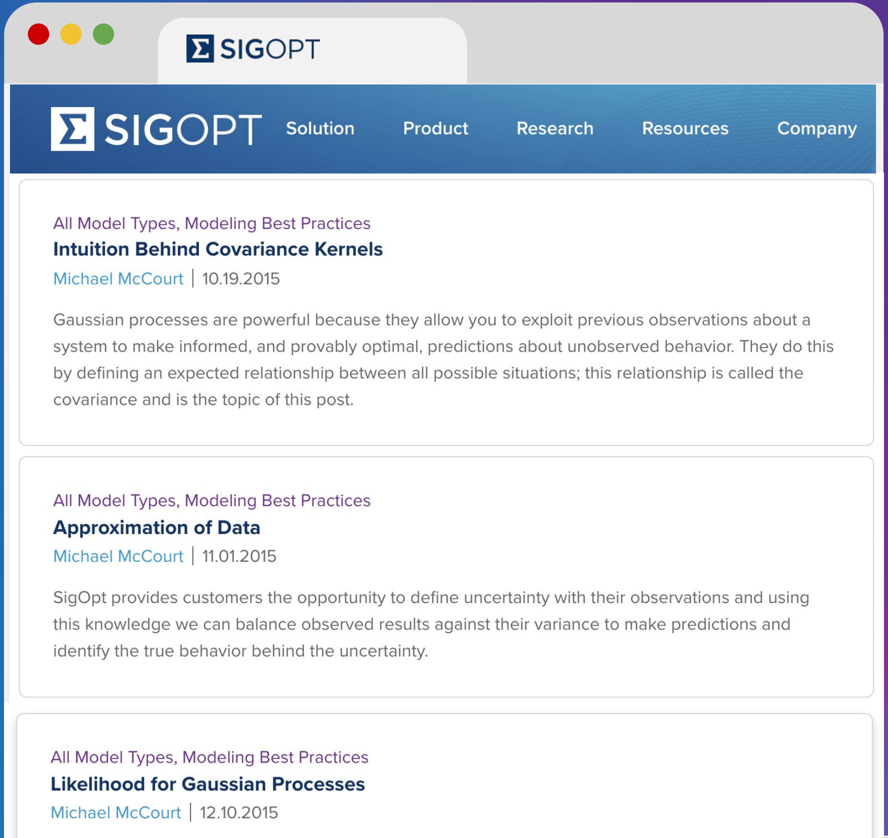




The Benefit Of Using SigOpt



SigOpt Blog Posts: Intuition Behind Bayesian Optimization



Some Relevant Blog Posts

- Intuition Behind Covariance Kernels
- Approximation of Data
- Likelihood for Gaussian Processes
- Profile Likelihood vs. Kriging Variance
- Intuition behind Gaussian Processes
- Dealing with Troublesome Metrics

Find more blog posts visit:
https://sigopt.com/blog/

How To Design An Optimization Problem

Design

Optimizing Multiple Competing Metrics

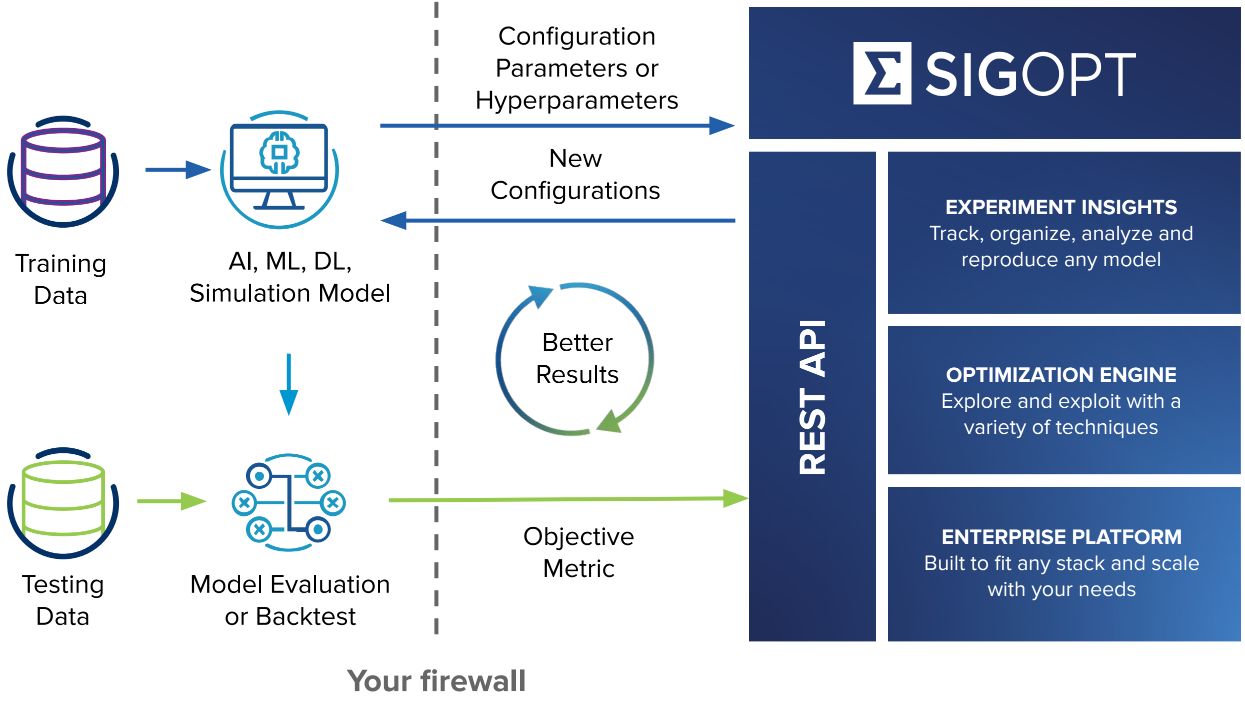
Why optimize against multiple competing metrics?

SigOpt allows the user to specify multiple competing metrics for either optimization or tracking to better align success of the experimentation with business value:

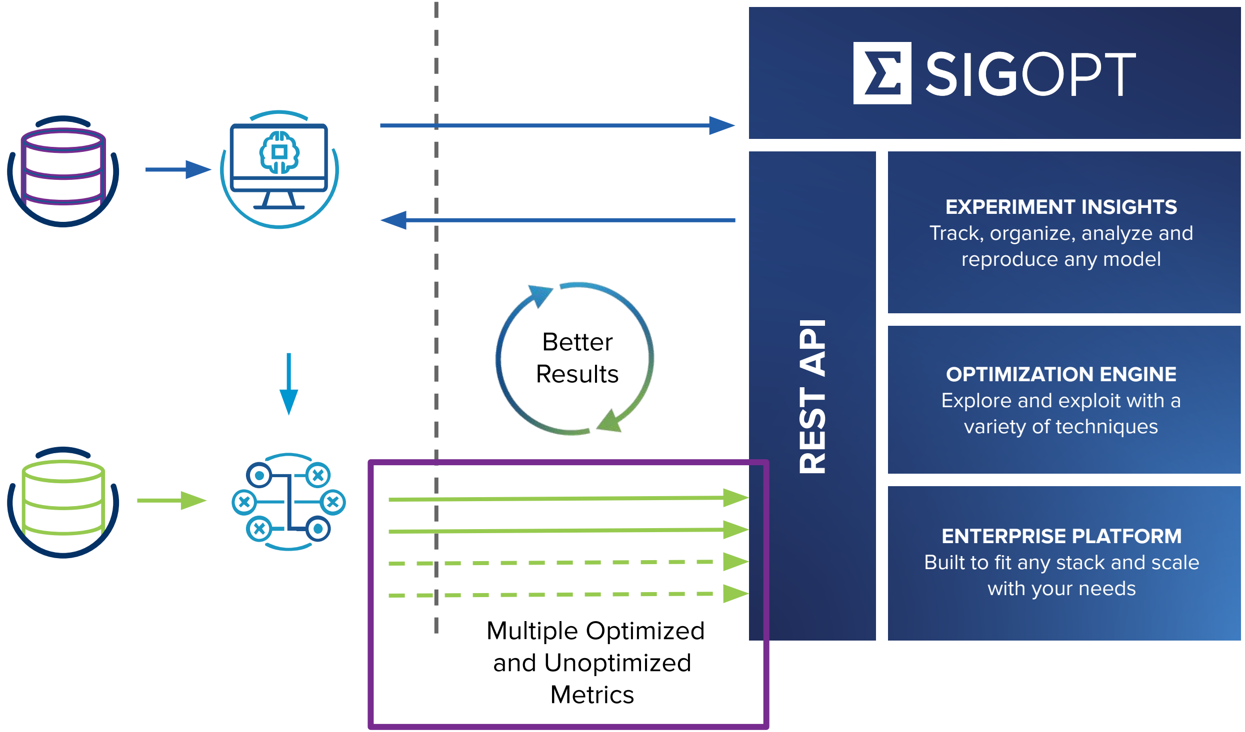
- **Multiple Metrics** — The option to defining multiple metrics, which can yield new and interesting results
- **Insights, Metric Storage** — Insights through tracking of optimized and unoptimized metrics
- **Thresholds** — The ability to define thresholds for success to better guide the optimizer

This process gives models that deliver more reliable business outcomes by helping optimally make the tradeoffs inherent in the modeling process and real world applications.

Optimizing Multiple Competing Metrics



Optimizing Multiple Competing Metrics



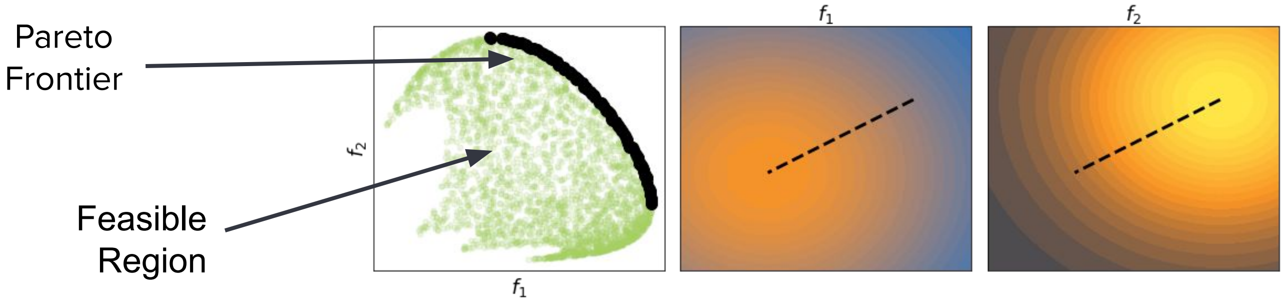
Looking for the right Balance

Balancing competing metrics to find the Pareto frontier

Most problems of practical relevance involve 2 or more *competing metrics*.

- **Neural networks** — Balancing accuracy and inference time
- **Materials design** — Balancing performance and maintenance cost
- **Algo trading** — Balancing Sharpe Ratio and book size

In a situation with Competing Metrics, the set of all **efficient** points (the Pareto frontier) is the solution.



Finding the Best Trade-offs

Balancing competing metrics to find the Pareto frontier

As shown before, the goal in multi objective or multi criteria optimization the goal is to find the optimal set of solution across a set of function [Knowles, 2006].

- This is formulated as finding the maximum of the set functions f_1 to f_n over the same domain x
- No single point exist as the solution, but we are actively trying to maximize the size of the efficient frontier, which represent the set of solutions
- The solution is found through scalarization methods such as convex combination and epsilon-constraint

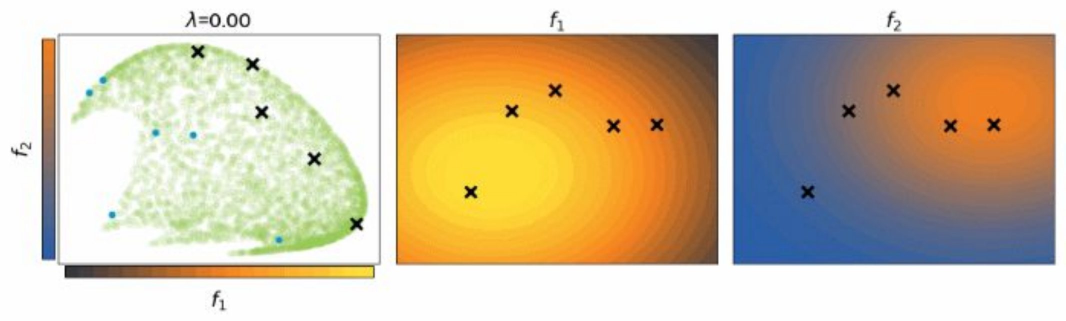
Optimizing Multiple Competing Metrics

First Approach: Convex Combination Scalarization

Idea: *If we can convert the multimetric problem into a scalar problem, we can solve this problem using Bayesian optimization.*

One possible scalarization is through a convex combination of the objectives.

$$f_\lambda(\mathbf{x}) = \lambda f_1(\mathbf{x}) + (1-\lambda) f_2(\mathbf{x}), \qquad 0 \le \lambda \le 1$$



intel. | SIGOPT

Optimizing Multiple Competing Metrics

Second Approach: Balancing competing metrics to find the Pareto frontier with threshold

As shown before, the goal in multi objective or multi criteria optimization the goal is to find the optimal set of solution across a set of function [Knowles, 2006].

- This is formulated as finding the maximum of the set functions f_1 to f_n over the same domain x
- No single point exist as the solution, but we are actively trying to maximize the size of the efficient frontier, which represent the set of solutions
- The solution is found through constrained scalarization methods such as convex combination and epsilon-constraint
- Allow users to change constraints as the search progresses [Letham et al, 2019]

Rephrase the Entire Problem

Constrained Scalarization

1. **Model all metrics independently**.
    - Requires no prior beliefs of how metrics interact.
    - Missing data removed on a per metric basis if unrecorded.
2. **Expose the efficient frontier through <u>constrained</u> scalar optimization**.
    - Enforce user constraints when given.
    - Iterate through sub constraints to better resolve efficient frontier, if desired.
    - Consider different regions of the frontier when parallelism is possible.
3. **Allow users to change constraints as the search progresses**.
    - Allow the problems/goals to evolve as the user's understanding changes.

Variation on Expected Improvement

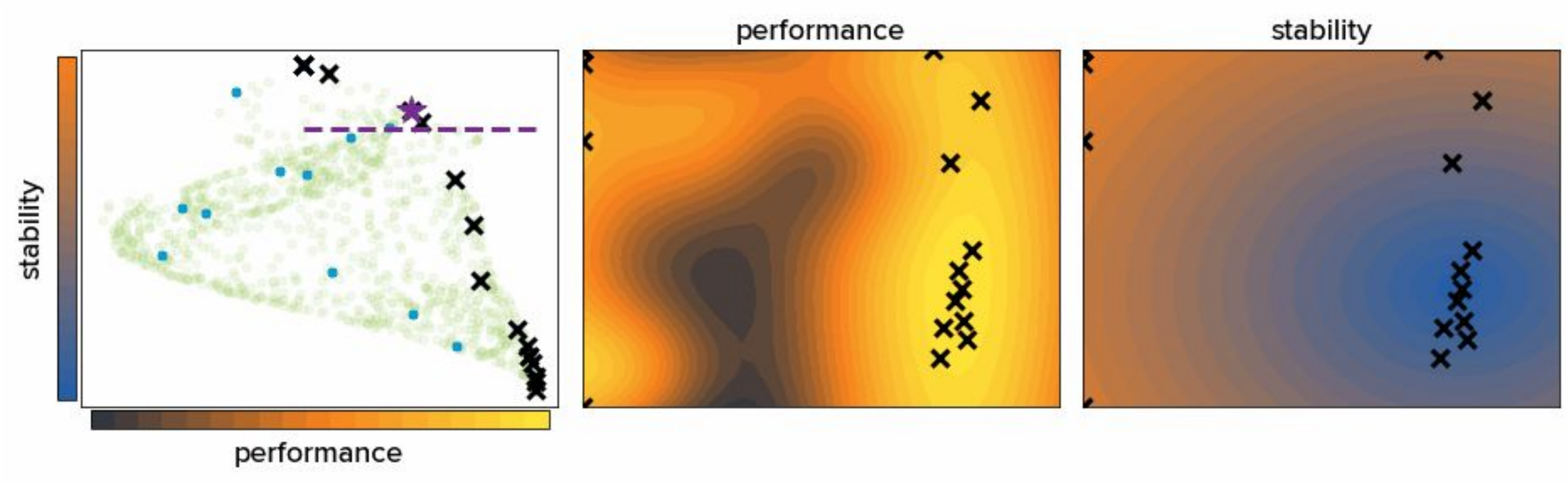
[Letham et al, 2019]

Constraints give customers more control over the circumstances and more ability to understand our actions.

$$\alpha_{\mathrm{NEI}}(\mathbf{x}|\mathcal{D}) = \int_{\mathbf{f}^n} \int_{\mathbf{c}^n} \alpha_{\mathrm{EIx}}(\mathbf{x}|\mathbf{f}^n, \mathbf{c}^n) p(\mathbf{f}^n|\mathcal{D}_f) \prod_{j=1}^{J} p(\mathbf{c}_j^n|\mathcal{D}_{c_j}) d\mathbf{c}^n d\mathbf{f}^n.$$

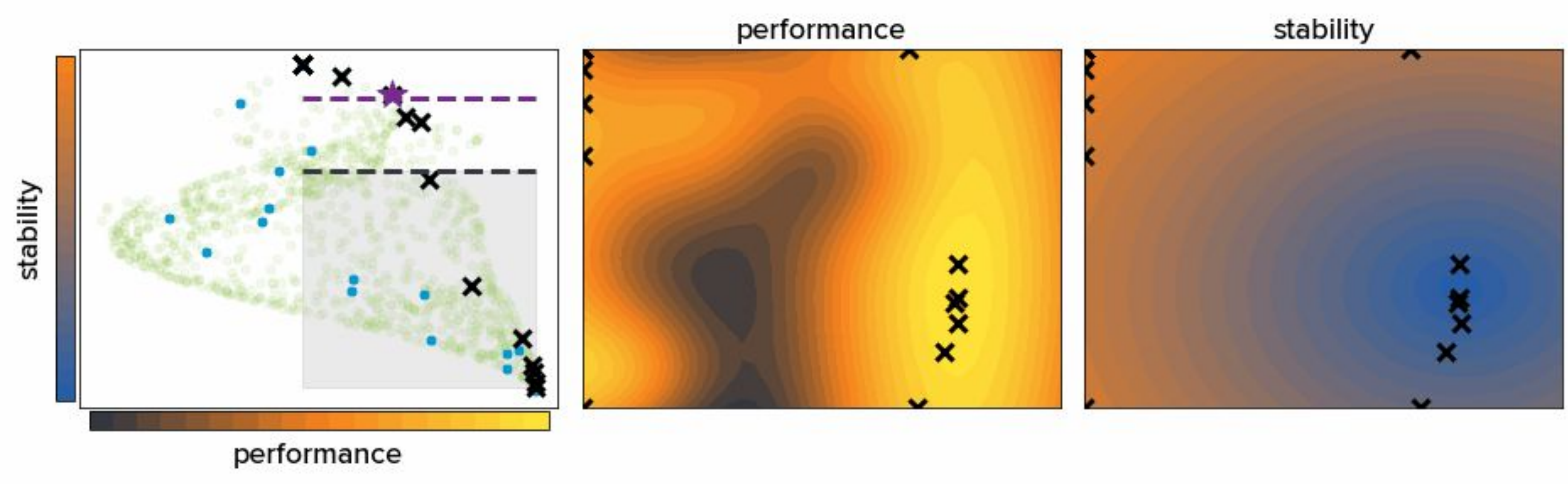
Use Epsilon Constraints to Focus Search

Intuition: Scalarization and Epsilon Constraints



Add Thresholds to Limit Frontier

Intuition: Constrained Scalarization and Epsilon Constraints



Putting It All Together Using SigOpt

Follow Along On Your Own:

1) Sign Up For A SigOpt Account:

sigopt.com/signup

2) Navigate To The Course Material:

tinyurl.com/lrzsigopt

Pop Quiz With Prices

Thank You

Please Give
Feedback

Tobias Andreasen - Staff Engineer, tobias@sigopt.com

intel | SIGOPT

SIGOPT AI & HPC SUMMIT

Registration: sigopt.com/summit

November 16th

Free & Virtual

accenture intel Mindtree

anastasia Stanford University Massachusetts Institute of Technology Novelis

Numenta PayPal The University of Arizona University of Pittsburgh