

Optimizing Time-consuming Objective Functions: Case Studies in Drug Development and Simulation Calibration

Peter I. Frazier

Department of Operations Research and Information Engineering
Cornell University

June 30, 2011
Joint Genome Institute
Department of Energy
Walnut Creek, CA

Outline

- 1 Overview
- 2 Illustrative Example and Algorithm Design
- 3 Simulation Model Calibration
- 4 Drug Development

Optimization

Optimization means **searching for, and finding, the best option.**

- In drug development, searching through the space of possible treatments to find the one that best treats the disease.
- Tuning the parameters of an assembler (e.g., Velvet) to produce the best assembly from data.
 - To do this, we need a comprehensive way to measure the quality of an assembly. This is ongoing work with Zhong Wang, Rob Egan, and Scott Clark (Cornell).
- Designing an experiment to achieve the best results.
 - e.g., choosing a set of solvents to minimize experimental noise.
- In high throughput screening, finding proteins with a given function, or small molecules with a given biological activity. (active learning)

Optimization

An optimization problem can be written as

$$\max_x f(x)$$

where,

- x is one option.
- $f(x)$ is the value of this option
 - (it might be only observable with noise.)
- f is the objective function.

In the assembler tuning problem, x is a set of parameters inputted to the assembler, and $f(x)$ is the quality of the assembly produced.

Time-Consuming Objective Functions

- In some problems, evaluating the objective function f takes a lot of time.
 - In the drug development problem discussed later, evaluating $f(x)$ takes 12 hours (although we can evaluate $f(x)$ for several x simultaneously).
 - In the simulation calibration problem discussed later, evaluating $f(x)$ takes 3 days.
- In these problems, standard optimization methods that evaluate f many times work poorly.
- Our focus in this talk is on methods that solve the optimization problem (approximately) using few function evaluations.
- In place of “time-consuming”, researchers also use the word “expensive.”

Machine Learning and Statistics

The methods to be discussed will combine **machine learning** with **value of information** calculations.

- ① Use machine learning and statistics to **predict** $f(x)$.
- ② Use value of information calculations to **decide** which options x to evaluate next.
- ③ Use the results from the evaluation to improve our predictions, and repeat.

Derivative-Free Global Optimization

Other aspects of the optimization problems considered:

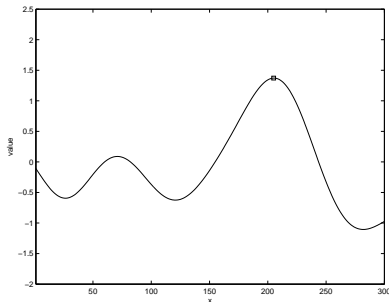
- **Derivative-free:** Many optimization methods assume we can observe the gradient of $f(x)$. We do not assume this.
- **Lack of Convexity:** Many optimization methods assume f is convex. We do not assume this.
- **Global Optimization:** Many optimization methods search only for a local optimum. We search for a global optimum (a full solution to $\max_x f(x)$.)

Outline

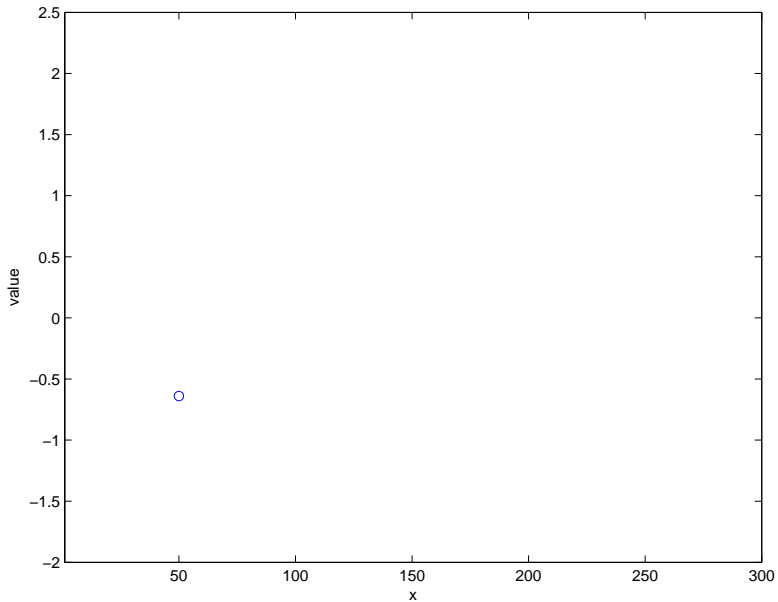
- 1 Overview
- 2 Illustrative Example and Algorithm Design**
- 3 Simulation Model Calibration
- 4 Drug Development

Illustrative 1D Example

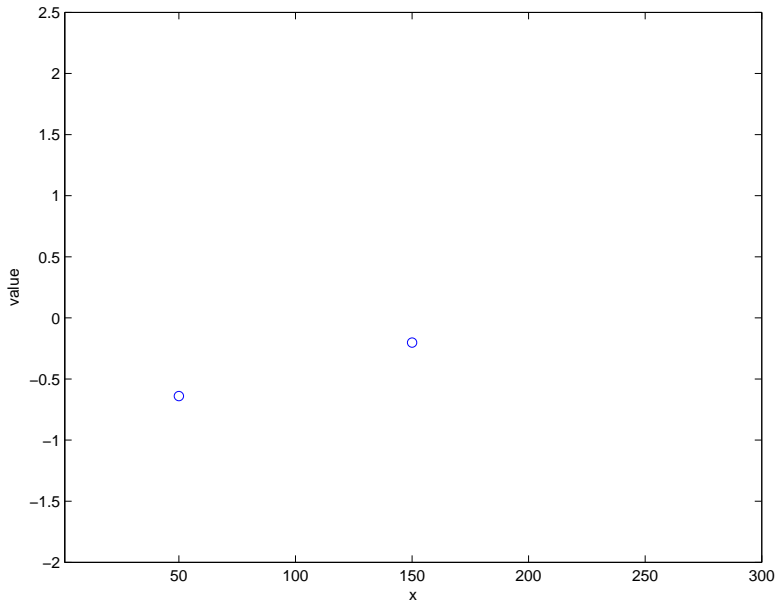
- For illustration, suppose we are tuning the concentration of one reagent in an assay.
- When we run the assay with the reagent at the concentration x , we observe the resulting quality of the assay $f(x)$ with noise.
- If we could run the assay many times, we could calculate the curve $f(x)$ below.
 - The concentration is on the horizontal axis, and the quality $f(x)$ is on the vertical axis.



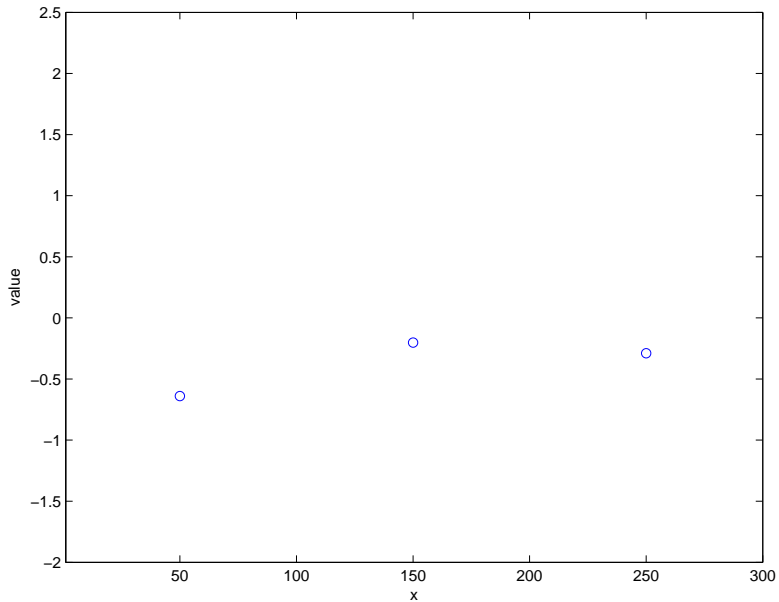
Illustrative 1D Example



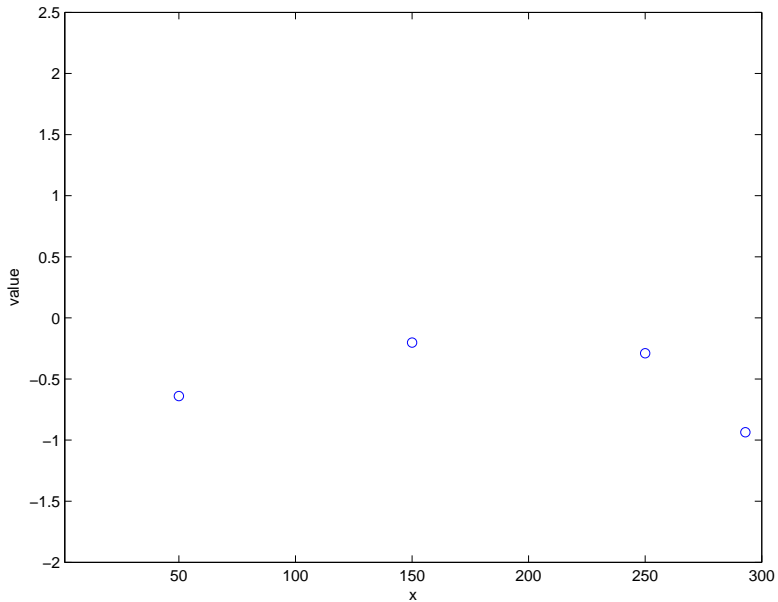
Illustrative 1D Example



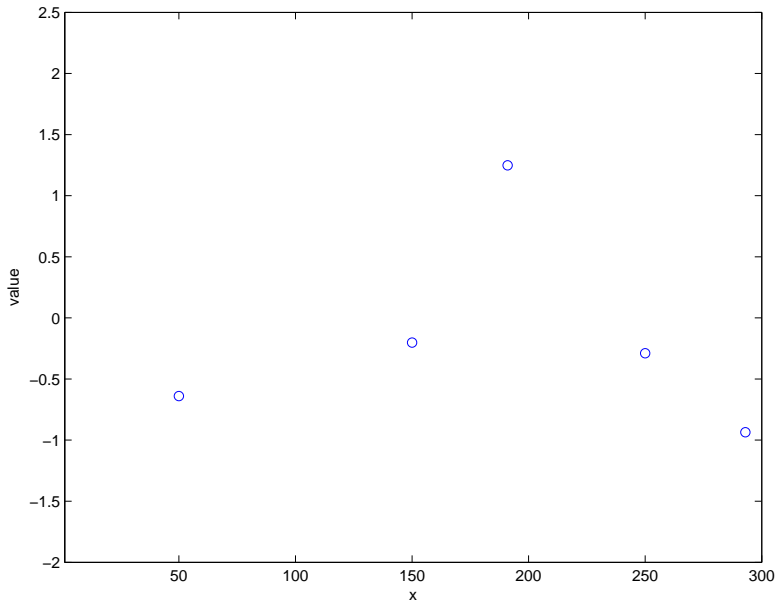
Illustrative 1D Example



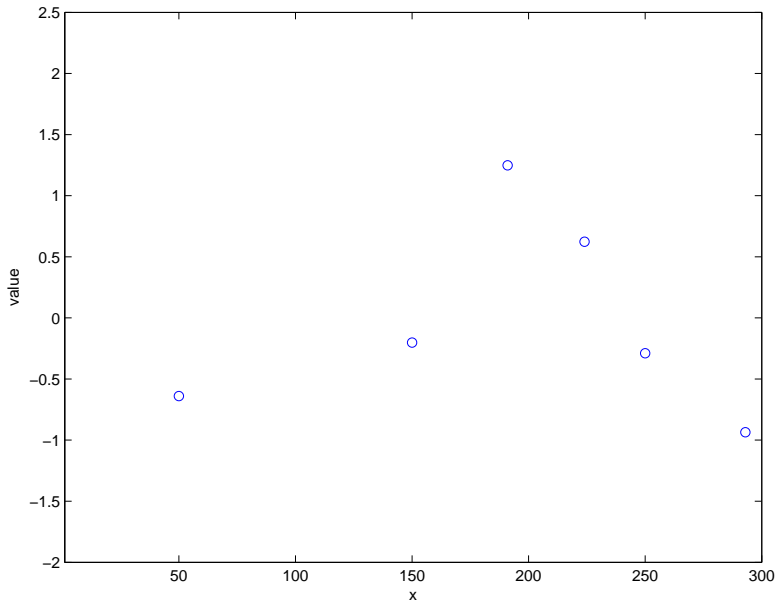
Illustrative 1D Example



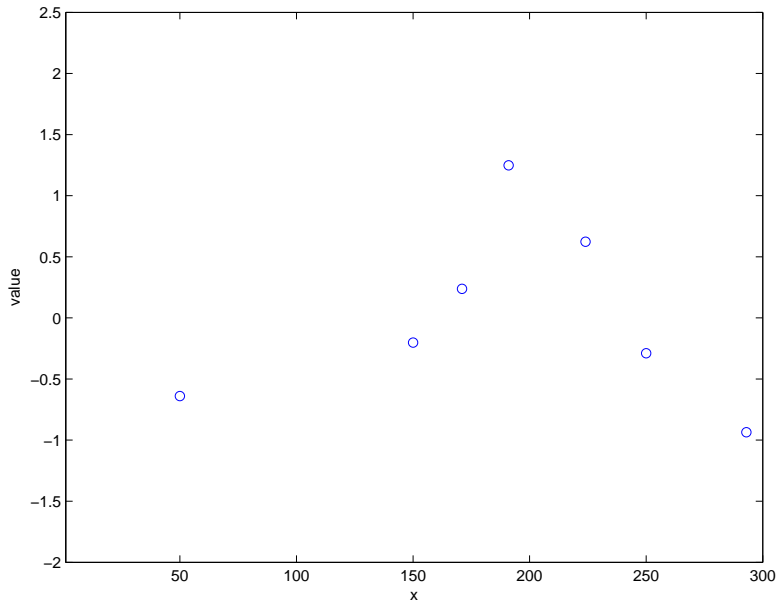
Illustrative 1D Example



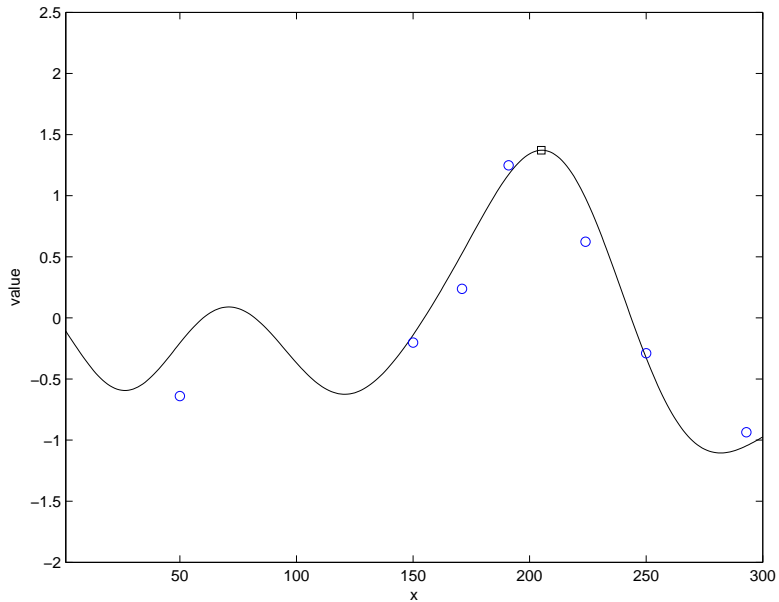
Illustrative 1D Example



Illustrative 1D Example

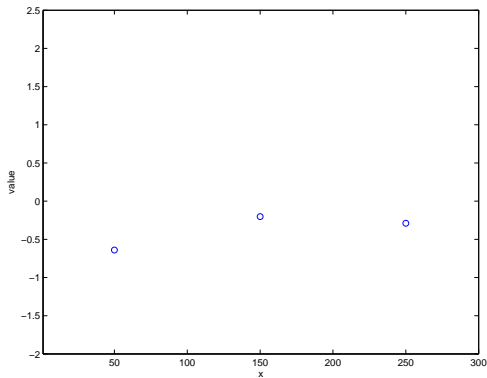


Illustrative 1D Example



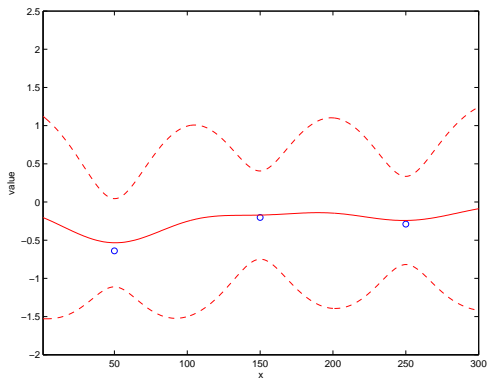
Algorithm Design

The central question when designing an optimization algorithm is: **Given our observations so far, which option x should we try next?**

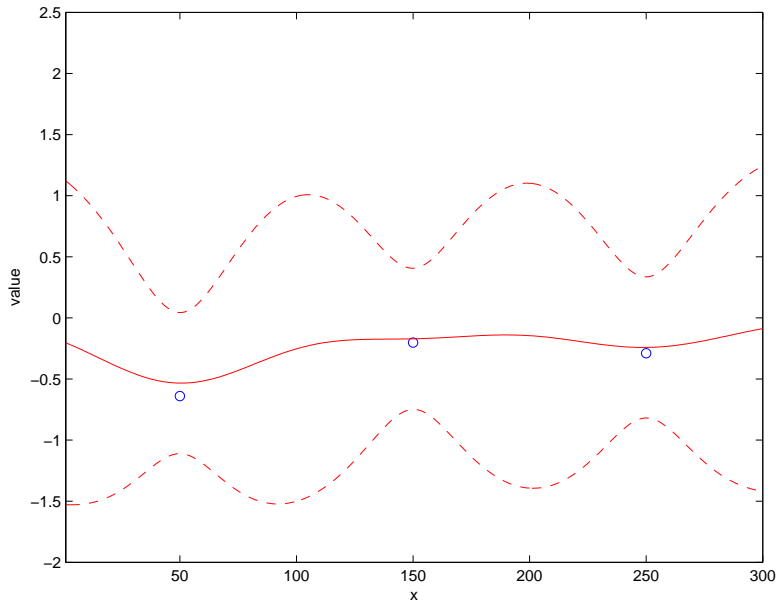


Prediction and Uncertainty

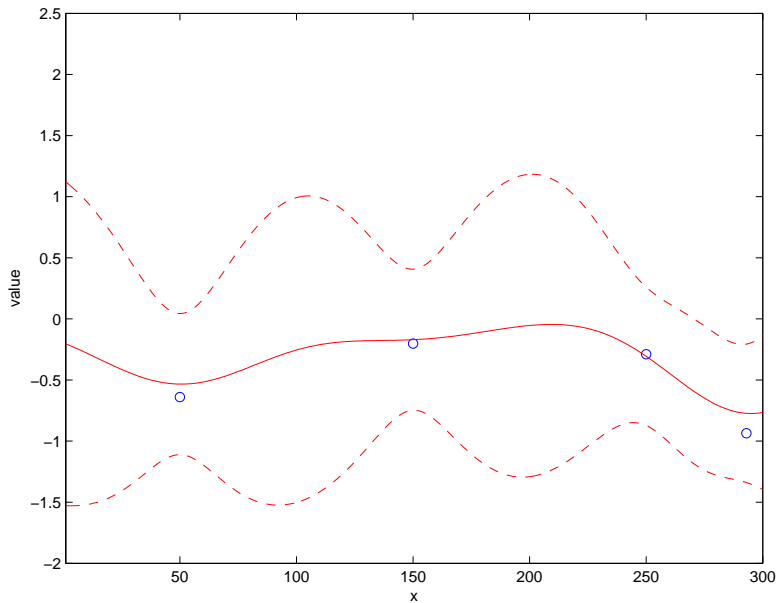
- To help choose which option to try next, we use statistics.
- In this problem, we use a Bayesian Gaussian process prior [Rasmussen & Williams 2006] on the objective function f , and calculate a posterior distribution from the observations so far:



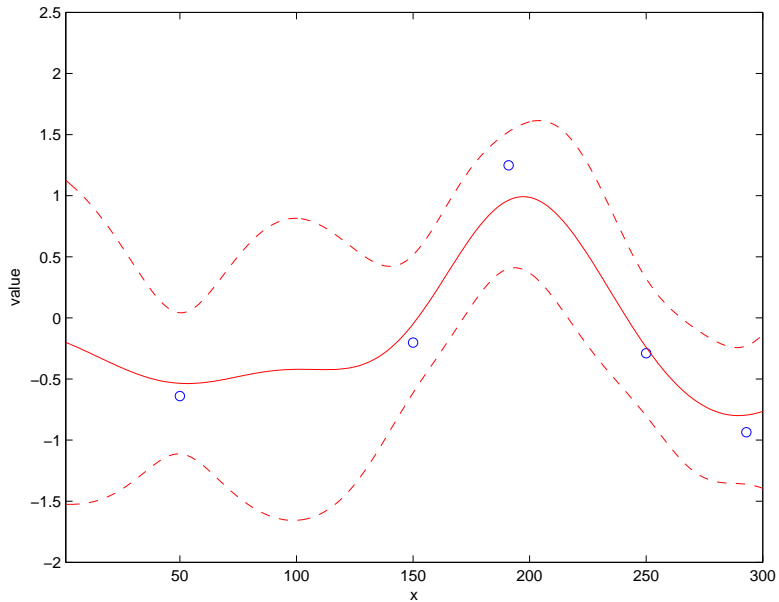
Illustrative 1D Example



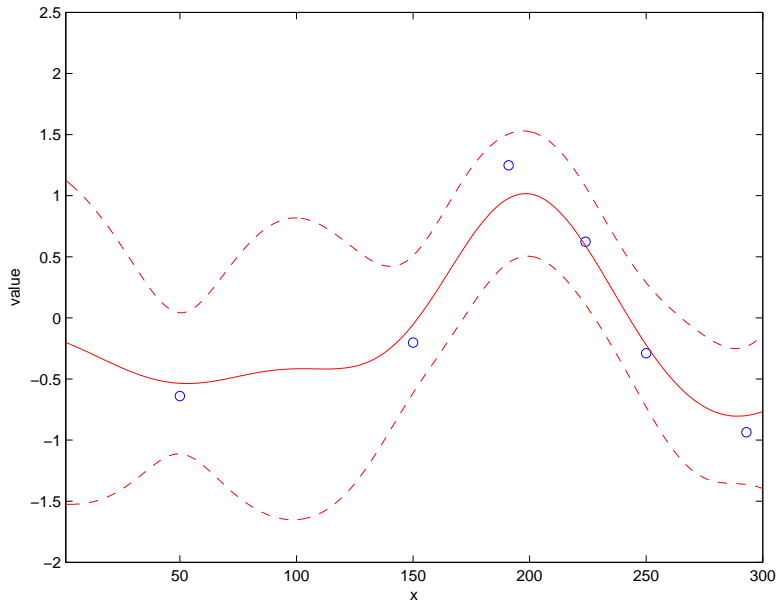
Illustrative 1D Example



Illustrative 1D Example

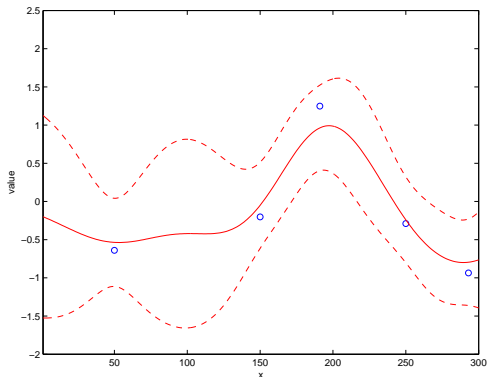


Illustrative 1D Example



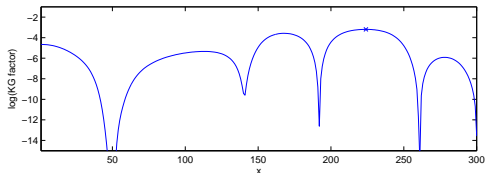
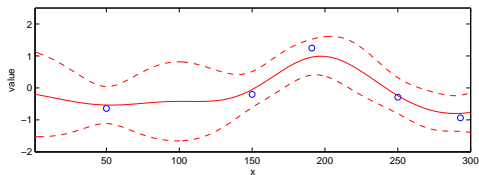
Algorithm Design

Given our observations so far, which option x should we try next?



Algorithm Design

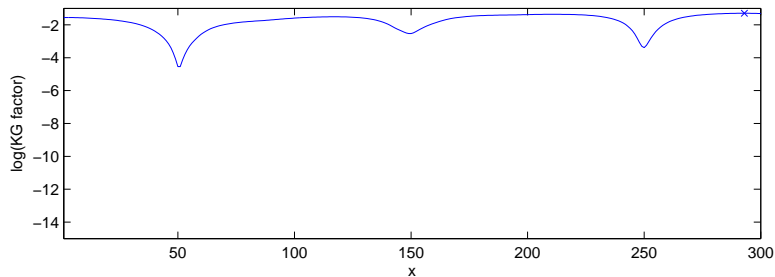
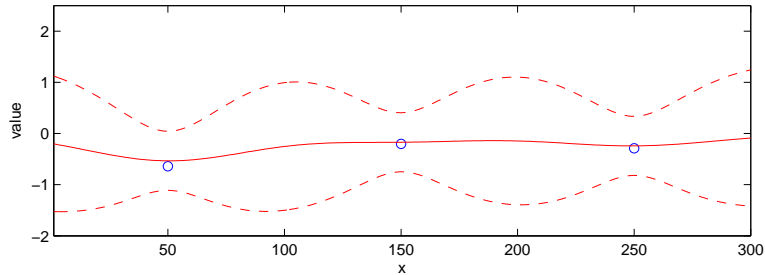
- We can calculate the **value of the information** that we can obtain, as a function of where we sample.
- A natural optimization algorithm is then to measure at the x whose value of information is largest.



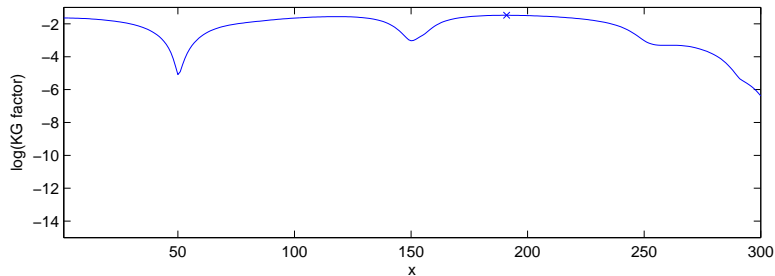
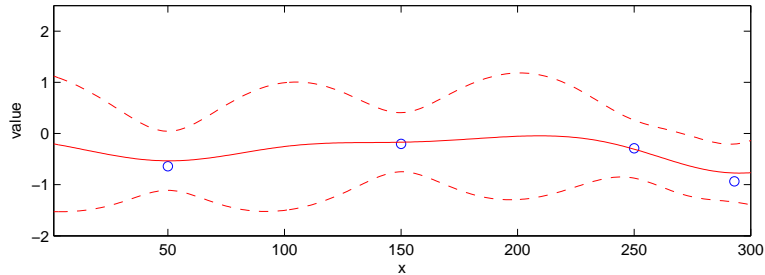
Value of Information

- The value of the information obtained from measuring x can be calculated as follows:
 - ① If we measure at x and observe y , we quantify the improvement in our ability to solve the optimization problem.
 - ② Since we haven't observed y yet, we take the expectation over all possible values of y that we might observe.
- The way in which we calculate step 1 determines the resulting value of information computed.
- A number of algorithms use this type of approach, including [Jones et al. 1998, Huang et al. 2006, Frazier et al. 2009, Scott et al. 2011].
- One such algorithm is the **Knowledge-Gradient (KG)** Algorithm [Frazier et al. 2009].

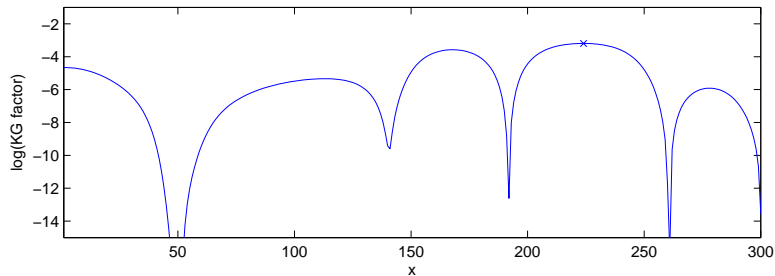
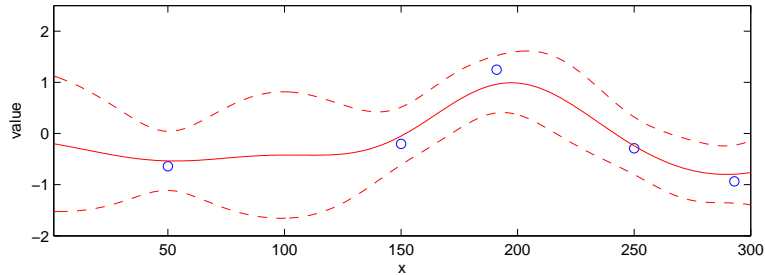
Illustrative 1D Example



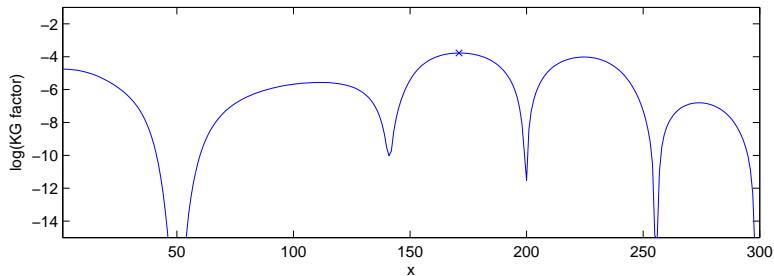
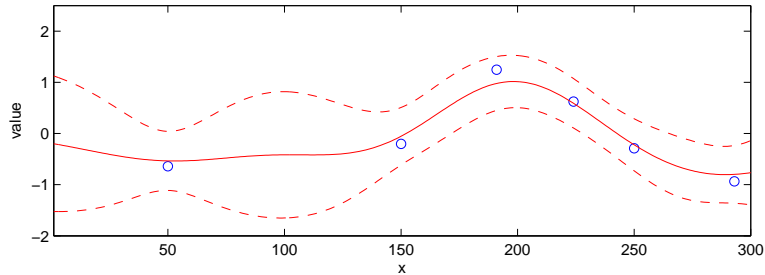
Illustrative 1D Example



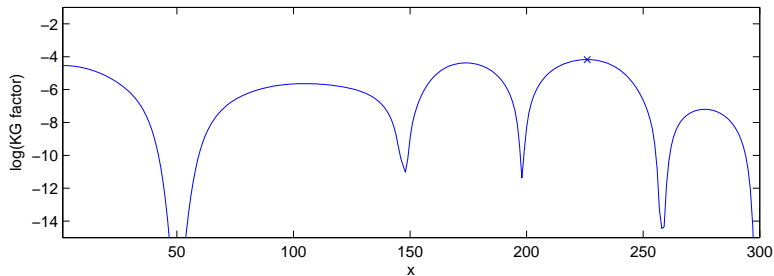
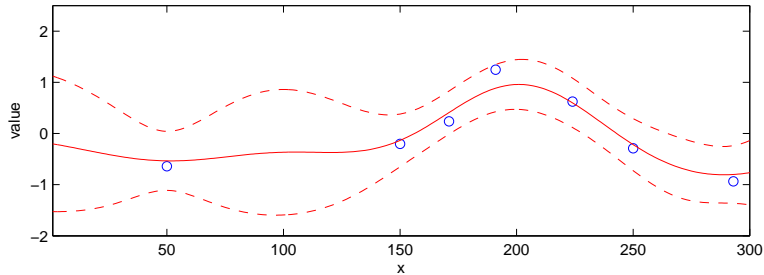
Illustrative 1D Example



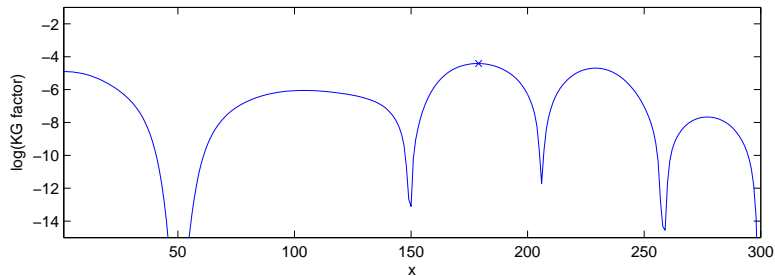
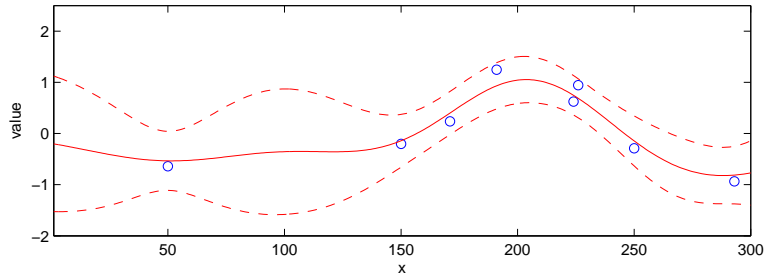
Illustrative 1D Example



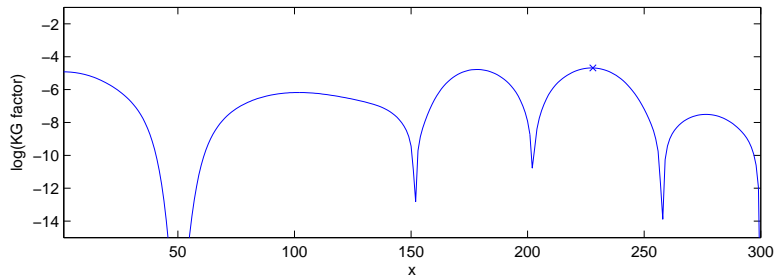
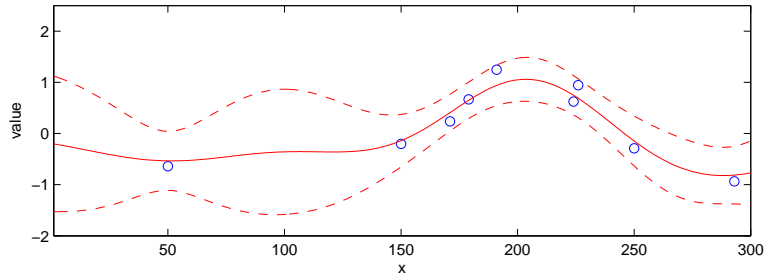
Illustrative 1D Example



Illustrative 1D Example



Illustrative 1D Example



Keep in Mind

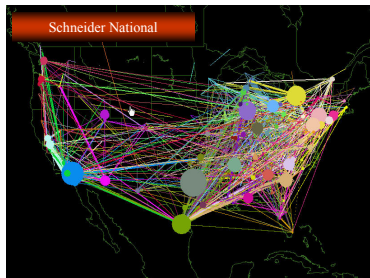
- This illustrative example is in one dimension with a continuous input, but...
- The methods discussed can be used in problems with continuous inputs in **higher dimensions** (e.g., 10 dimensions).
- They can also be used in problems with discrete input spaces (e.g., x is a **chain of amino acids**, or x is a **small molecule**).
- In these more difficult problems, solving the optimization problem by hand is much more difficult, and so it is for these problems that mathematical methods often provide the most value.

Outline

- 1 Overview
- 2 Illustrative Example and Algorithm Design
- 3 Simulation Model Calibration**
- 4 Drug Development

Simulation Model Calibration at Schneider National

- The logistics company Schneider National uses a large simulation-based optimization model to try “what if” scenarios.
- The model has several input parameters that must be tuned to make its behavior match reality before it can be used.
- The model is tuned by hand once per year on the most recent data. Each tuning effort requires between 1 and 2 weeks.



(Joint work with Warren B. Powell and Hugo Simão, Princeton University)

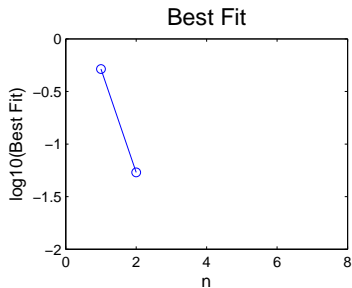
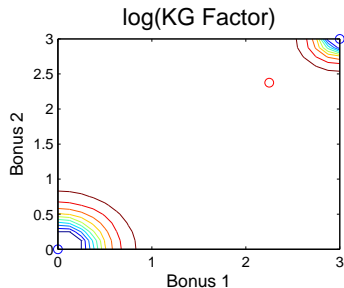
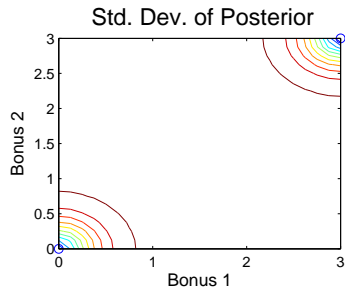
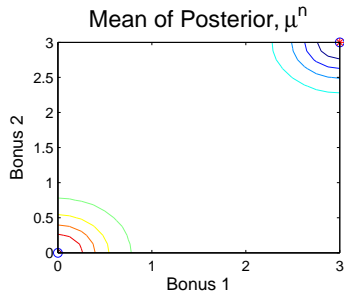
Model Parameters

- Input parameters to the model include:
 - time-at-home bonuses.
 - “pacing” parameters describing how fast and far drivers drive per day.
 - gas prices
 - ...
- Output parameters from the model include:
 - billed miles
 - driver utilization
 - average number of trips home per driver per 4 weeks.
 - proportion of drivers without time at home over 4 weeks.
 - ...
- Some of these inputs are known (e.g., gas prices), but some are unknown (e.g. time-at-home bonuses).
- Goal: adjust the inputs to make the optimal solution found by the model match current practice.

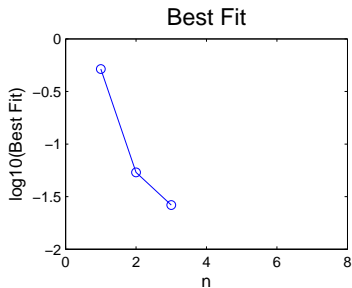
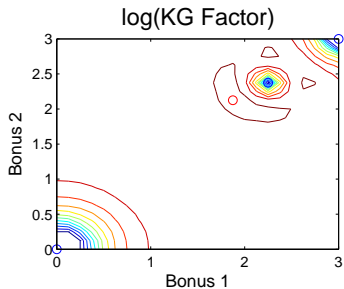
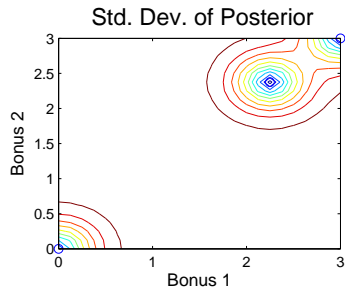
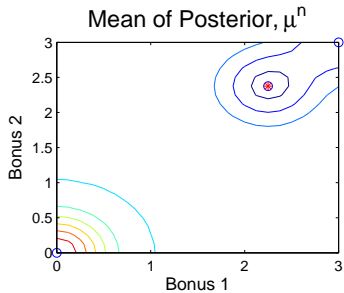
Simulation Model Calibration

- Goal: adjust the inputs to make the optimal solution found by the ADP model match current practice.
 - x is a set of inputs to the simulator.
 - $f(x)$ is how closely the simulator output matches history.
- Running the simulator for one set of bonuses takes 3 days, making calibration difficult.
- The model may be run for shorter periods of time, e.g. 12 hours, to obtain noisy output estimates.

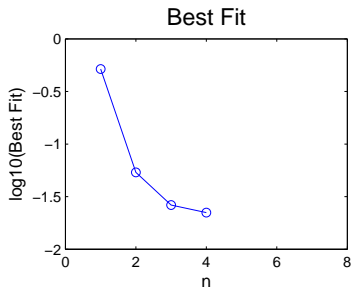
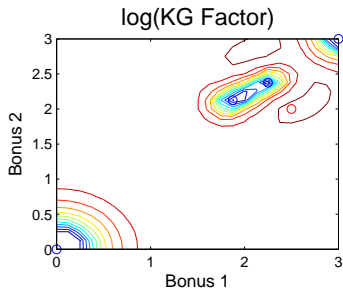
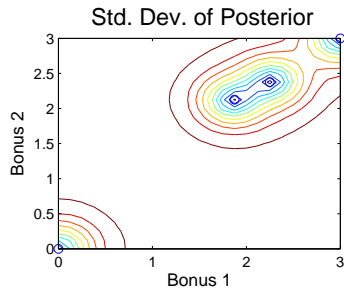
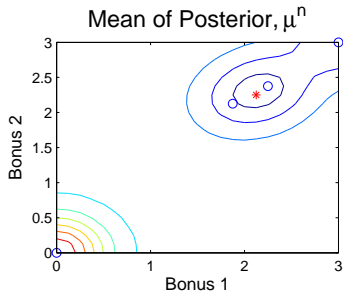
Simulation Model Calibration Results



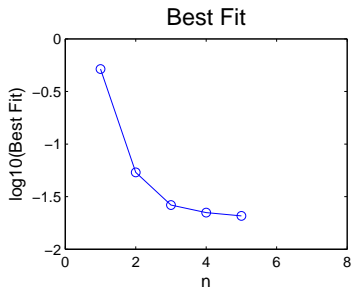
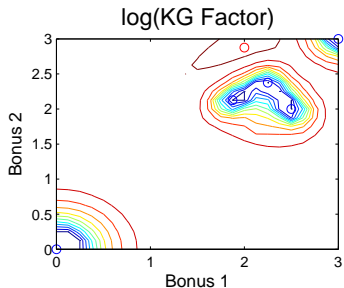
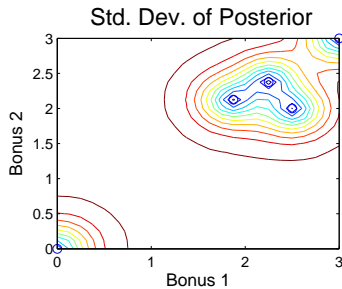
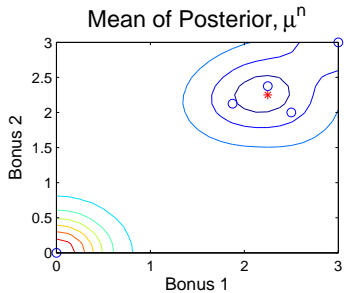
Simulation Model Calibration Results



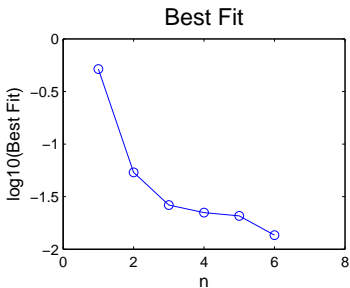
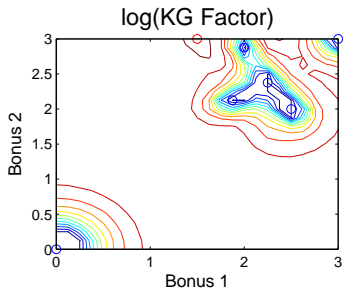
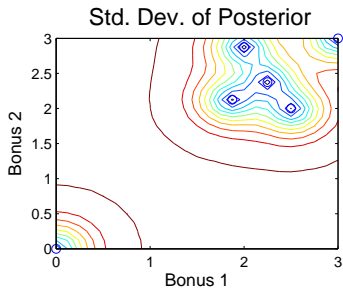
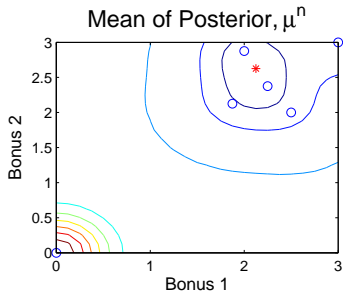
Simulation Model Calibration Results



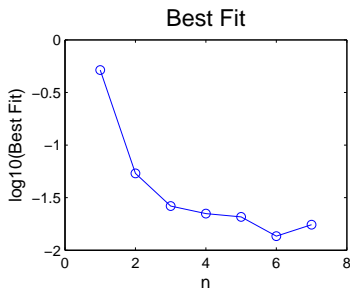
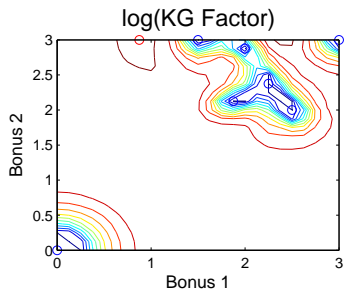
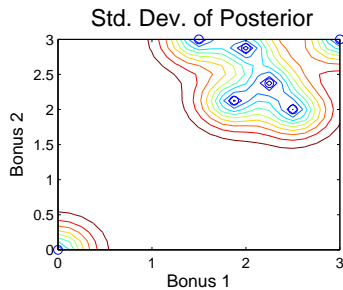
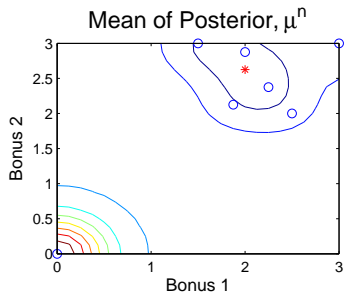
Simulation Model Calibration Results



Simulation Model Calibration Results



Simulation Model Calibration Results



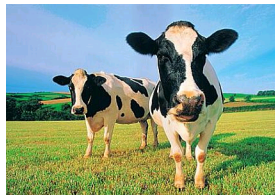
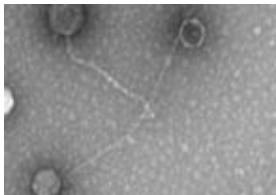
Simulation Model Calibration Results

- The KG method calibrates the model in approximately 3 days, compared to 7 – 14 days when tuned by hand.
- The calibration is automatic, freeing the human calibrator to do other work.
- The KG method calibrates as accurately or better than does by-hand calibration.
- Current practice uses the year's calibrated bonuses for each new “what if” scenario, but to enforce the constraint on driver at-home time it would be better to recalibrate the model for each scenario. Automatic calibration with the KG method makes this feasible.

Outline

- 1 Overview
- 2 Illustrative Example and Algorithm Design
- 3 Simulation Model Calibration
- 4 Drug Development**

Bacteriophage-based Treatments for Metritis



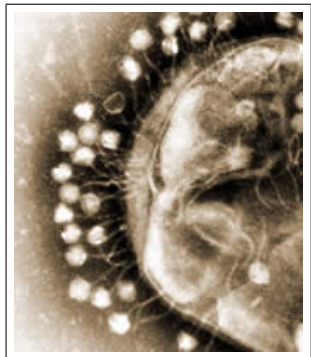
- Metritis is a bacterial infection of the uterus.
 - A leading cause of loss of milk production and fertility in dairy cows.
 - Caused primarily by *E. coli* and *A. pyogenes* bacteria.
- We are developing new treatments for metritis that use bacteriophages (viruses that kill bacteria) instead of antibiotics.

Joint work with Rodrigo C. Bicalho, Thiago M.A. Santos, and André G.V. Teixeira, (Cornell College of Veterinary Medicine), Zachary Owen, Rolf Waeber (Cornell ORIE).

Current treatment: Broad-spectrum Antibiotics

- Antibiotics are used to treat metritis in sick cows.
- Antibiotics are also given to well cows as a preventative measure.
- Causes to be concerned about the indiscriminate use of broad-spectrum antibiotics:
 - ① Bacteria develop **resistance to antibiotics** if they are used too widely.
 - ② Releasing large quantities of antibiotics into the environment via farm effluent may have negative **environmental effects**.

Advantages of Bacteriophage-based Treatments



- Reduced risk of **bacterial resistance**
 - Increasing the number of available treatments mitigates the problem of bacterial resistance.
 - Each phage is specific for a few strains of a single bacterium, limiting its use.
- Reduced **environmental impact**:
 - Phages already exist at dairy farms.
 - Each phage interacts with a very limited number of species.

The Challenge of Specificity

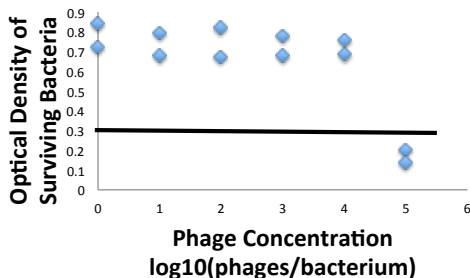
- Specificity:
 - Each phage kills only a few specific strains of a single species of bacterium.
 - An infection could be caused by any one of a number of strains of bacteria.
 - An effective treatment must then be a cocktail of phages that will be effective against each common metritis-causing strains of bacteria.
- A **cocktail** is a collection of phages.
 - We also specify a concentration (# phages/mL) for each phage in the cocktail.
- Challenge: Find a cocktail that works well and can be produced economically.

Production Costs

- A cocktail is a collection of phages, with a concentration specified for each phage.
- Production costs
 - A cocktail is cheaper to produce if it contains few phages.
 - A cocktail is cheaper to produce if the concentrations are small.
- **What is the cheapest cocktail that kills all targeted bacteria?**

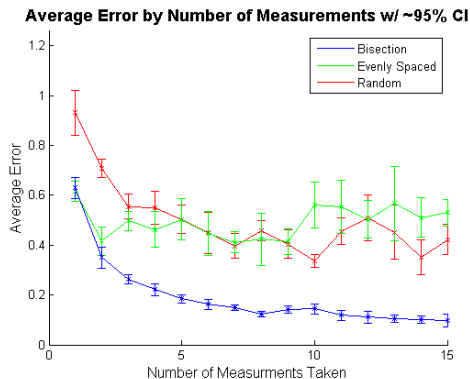
Optimization Problem 1: Minimal Effective Concentration

- Optimization Problem: Given a phage (or collection of phages) and a bacterium, find the **minimal** concentration that kills the bacterium.



- Each experiment is time-consuming (1 day).
- x is the phage concentration and $f(x)$ measures the distance of the optical density to the critical threshold 0.3.
- We act **sequentially**, basing each new experiment on previous results. (Experiments are done in batches of 96 per day. To allow sequential decision-making, we consider 96 phage-bacteria pairs simultaneously.)

Optimization Problem 1: Experimental Results

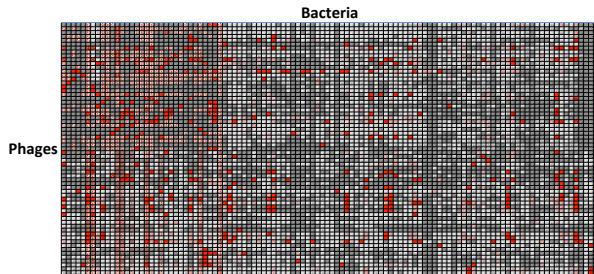


- These results suggest we can use 3 experiments per phage-bacteria pair instead of 6 (currently used), and get better accuracy.
- This reduces time required to **300 days** down from **600 days** to compute the root for every pair of one phage and one bacterium.

Optimization Problem 2: Global Optimization

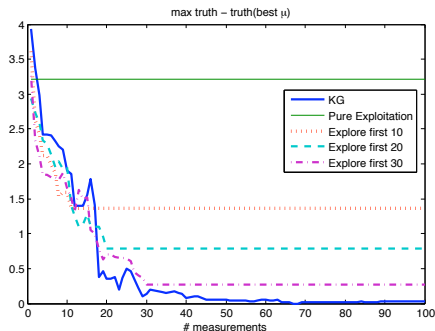
- Problem 1 searched for the minimal effective concentration of a single phage against a single bacterium.
- Problem 2: Find the phage cocktail with minimal production costs that is effective against each bacterium individually.
- x is a 50-dimensional vector (a concentration for each phage, with 0 for phages not in the cocktail) and $f(x)$ combines production costs with effectiveness.

Optimization Problem 2: Global Optimization



- This matrix is obtained from a faster spot assay experiment.
 - The color in each cell gives performance of a single phage against a single bacterium in the spot assay.
 - Better spot assay performance indicates a lower concentration of the phage may be sufficient to kill the bacterium.
- In our prior, we use that the effectiveness of a cocktail against a bacterium is approximately the effectiveness of the best phage in the cocktail against that bacterium.
- Performing this optimization is ongoing work.

Results (for a related problem)



- Results for Optimization Problem 2 are not ready yet.
- The numerical results shown are from on a simulation study of a related problem (design of an analgesic based on benzomorphan)
- Results show effectiveness of the best drug compound discovered so far as a function of the number of measurements for a collection of 1000 benzomorphan compounds. (Negoescu, Frazier, Powell 2010)

Conclusion

- Problems of interest:
 - Optimization or classification.
 - Time-consuming function evaluations.
 - Noise is ok (actually desirable).
 - Lack of typical optimization structure (lack of derivatives, lack of convexity).
- Methods combine prediction using machine learning and statistics with value of information calculations.
- I'm glad to discuss possible collaborations!

Thank You!