# Experimental optimization

## Lecture 11: Bayesian optimization

David Sweet

# Review

## A/B testing

- Take a measurement
- Randomize: reduce bias
- Replicate: reduce SE
- Result: aggregate measurements w/low SE, easy to compare

# Review

## Multi-armed bandit (MAB)

- Balance **exploration with exploitation**
- Optimize business metric during experiment
  - Instead of limiting false position/negatives rates
- Reduces experimentation cost

# Review

## Response surface methodology (RSM)

- Model response surface with **surrogate function**
- Optimize over surrogate function (fast) instead of experiment (slow)
- Reduces experimentation cost

# Bayesian optimization

## RSM+MAB

- Combines surrogate function with exploration/exploitation
- Also uses modern modeling and optimization methods to
  - Make the whole process automatic
  - Scale to more parameters

# Bayesian optimization

## Procedure

- Steps similar to RSM:

  1. Build surrogate model
  2. Optimize over surrogate model
  3. Take measurement at optimum

Repeat

# Surrogate model

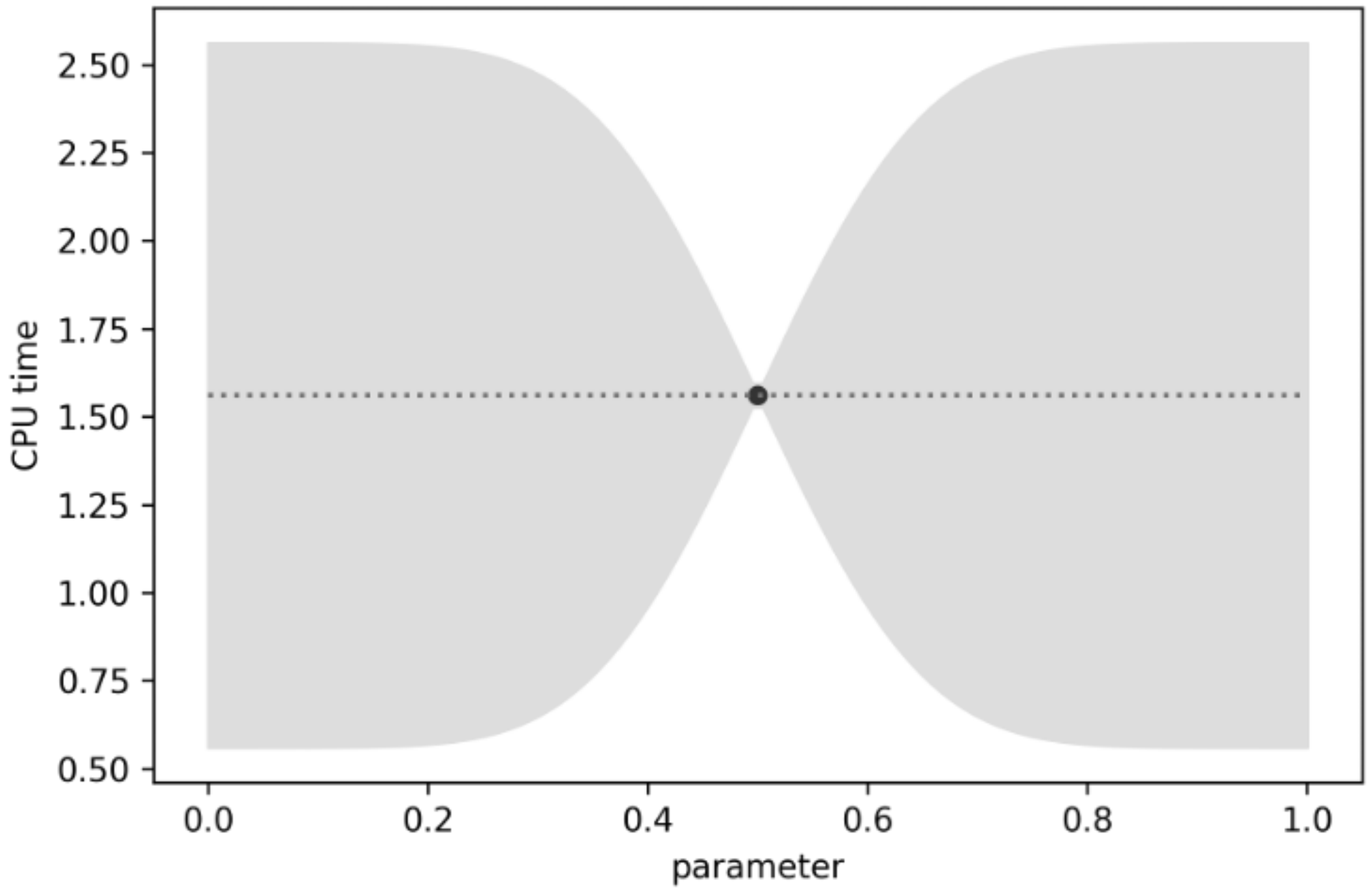
## Gaussian process regression (GPR)

- No need to specify model
  - RSM used linear regression, where engineer decides regressors
- GPR is a non-parametric model
  - Estimates are weighted averages of all measurements
  - No fitting, no betas (as in linear regression, *parametric* model)

# Surrogate model

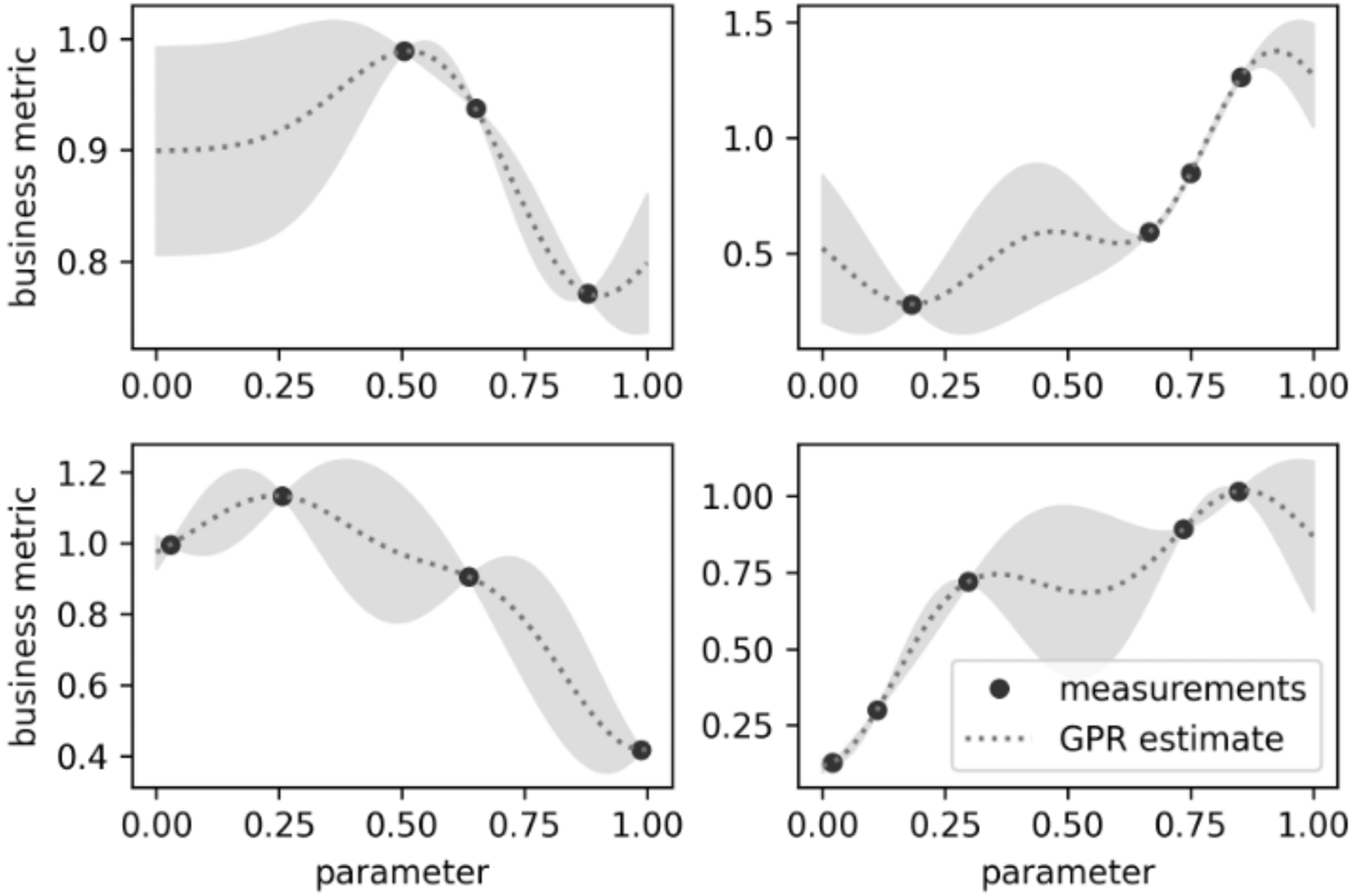
## Gaussian process regression (GPR)

- Also estimates uncertainty in its own estimates
  - Due to having few aggregate measurements
  - Not the same as standard error of a single agg. meas.
  - Larger uncertainty when farther from measurement



# Surrogate model

## Gaussian process regression (GPR)



# Optimization

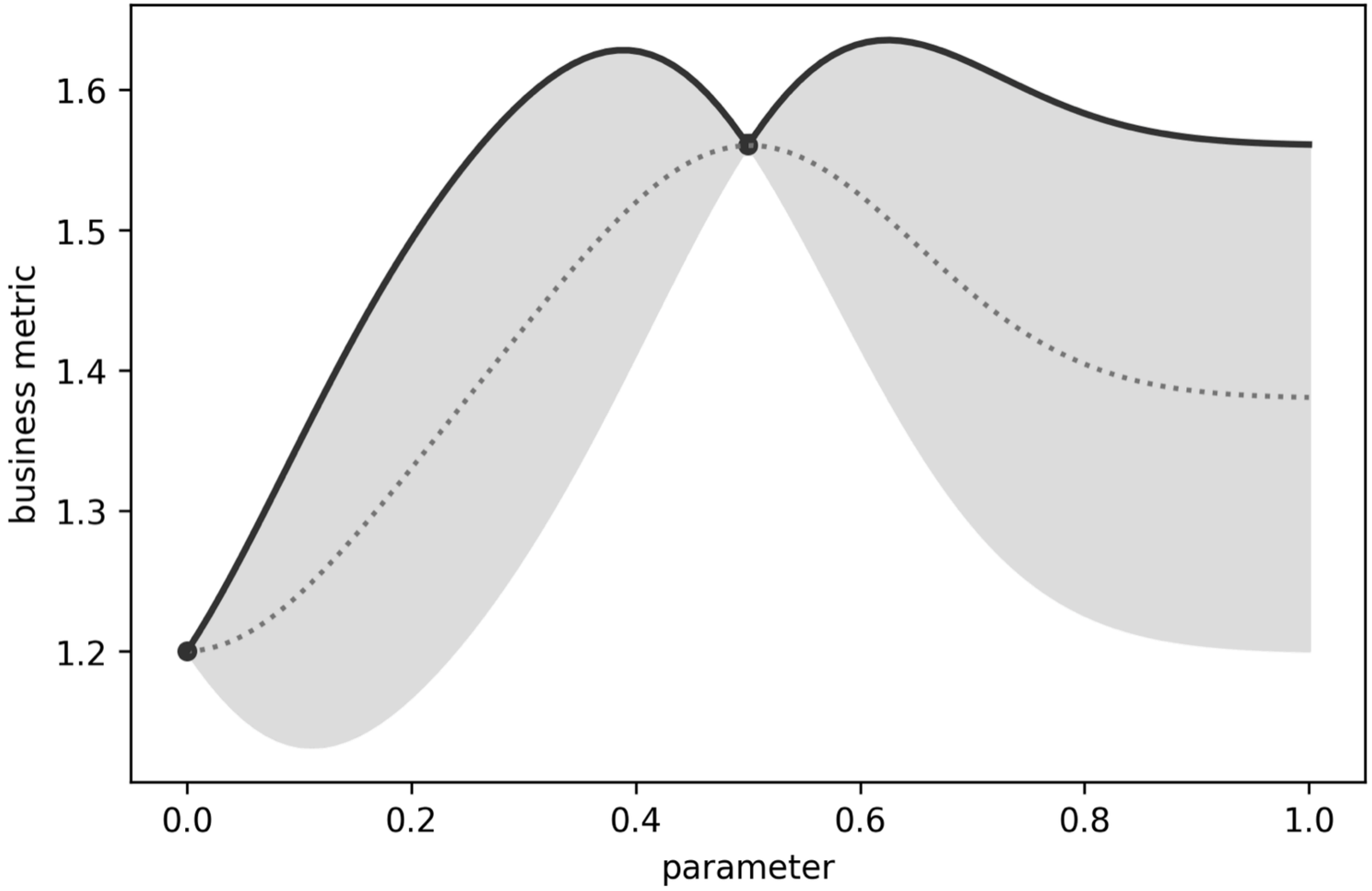
## Acquisition function

- Make use of surrogate’s uncertainty estimates
- Optimize over some function of BM estimates **and** uncertainty
  - called *acquisition function*
- Optimization chooses next parameter value to measure
- Exploitation: measure where BM is high to get BM
- Exploration: measure where uncertainty is high to improve (next) surrogate

# Optimization

## Acquisition function

- BM estimate is gray, dotted line
- Black line is BM + uncertainty
- Black line is acquisition function



# Optimization

## Acquisition function

- Acquisition functions are heuristics; no one is “optimal”
- Examples:
  - Upper confidence bound: previous slide, BM + uncertainty
  - Expected improvement: looks for parameter with most grey above the best measurement so far
  - Thompson sampling: looks for **best BM estimate** from a single function randomly drawn from the grey area (that still fits the data)

# Optimization

## Numerical optimization methods

- Grid search, for 1-2 parameters
- Random search, for 2-3 parameters
- CMA-ES (or NES) for 3+ parameters
  - CMA-ES is a SOTA “black box optimizer” (BBO)
- BBO: only requires function evaluations, not derivatives
- Bayesian optimization is also a BBO

| | + Pursue BM<br>Exploration/exploitation | |
|---|---|---|
| | A/B testing<br>(measurement) | Multi-armed bandits |
| + Surrogate function | Response surface modeling | Bayesian optimization |

# Bayesian optimization

## Miscellany

- Has been extended to categorical and discrete parameters
- Active research to extend beyond 20 parameters
- More complex code and math than previous methods
  - SigOpt — high quality software, just made free by Intel
  - Ax — also high quality, open sourced by Facebook

# Summary

## Bayesian optimization (BO)

- BO combines two ideas to reduce experimentation costs
  - surrogate function (from RSM)
  - balancing exploration with exploitation (from MAB)
- Surrogate model is Gaussian process regression (GPR)
- Acquisition function biases measurements toward uncertainty in surrogate
- Scales to 20 parameters
- Automated (compare to RSM)