Secondmind

Active learning in the real world

Machine learning for electric motor calibration

Henry Moss

henry.moss@secondmind.ai

Secondmind

Helping engineers design better cars faster

- Tech startup

Secondmind

Helping engineers design better cars faster

- Tech startup
- 50 people

Secondmind

Helping engineers design better cars faster

- Tech startup
- 50 people
- Maintain open-source toolboxes (Gpflow, Trieste)

Secondmind

Helping engineers design better cars faster

- Tech startup
- 50 people
- Maintain open-source toolboxes (Gpflow, Trieste)
- Develop active learning and Bayesian optimisation software products

Secondmind

Helping engineers design better cars faster

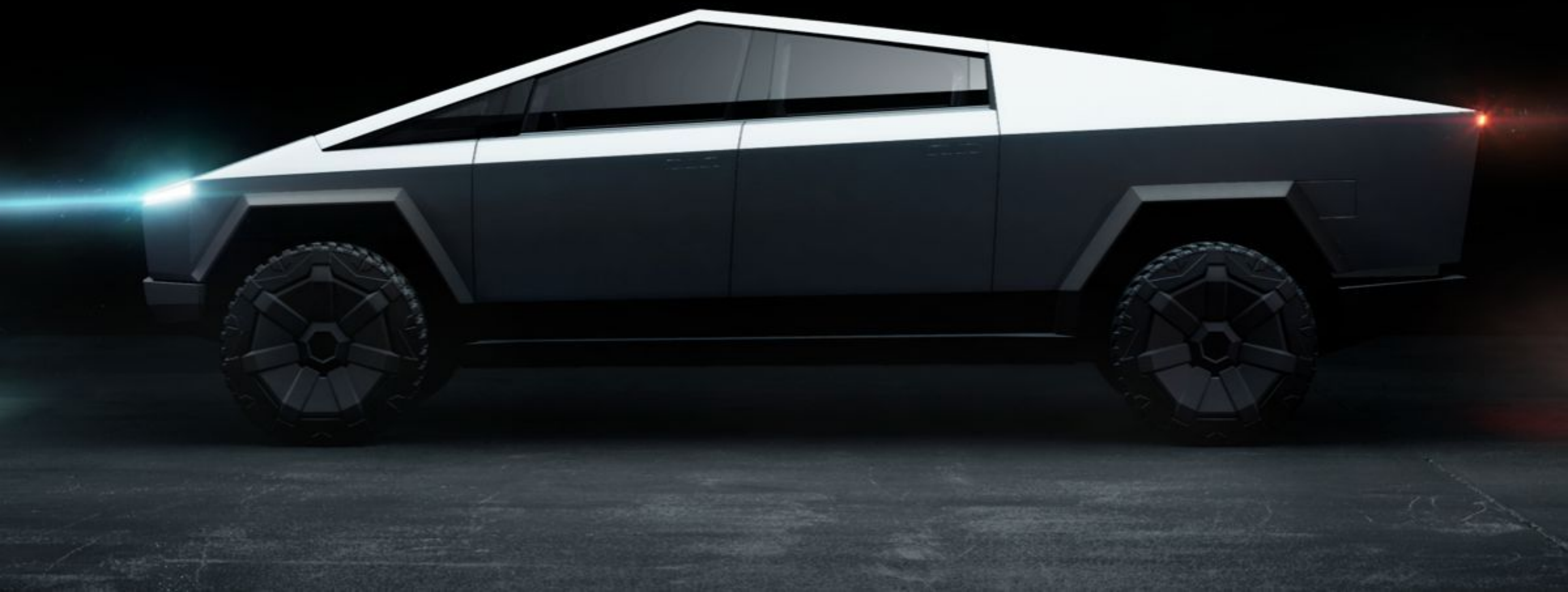
- Tech startup
- 50 people
- Maintain open-source toolboxes (Gpflow, Trieste)
- Develop active learning and Bayesian optimisation software products

How we help

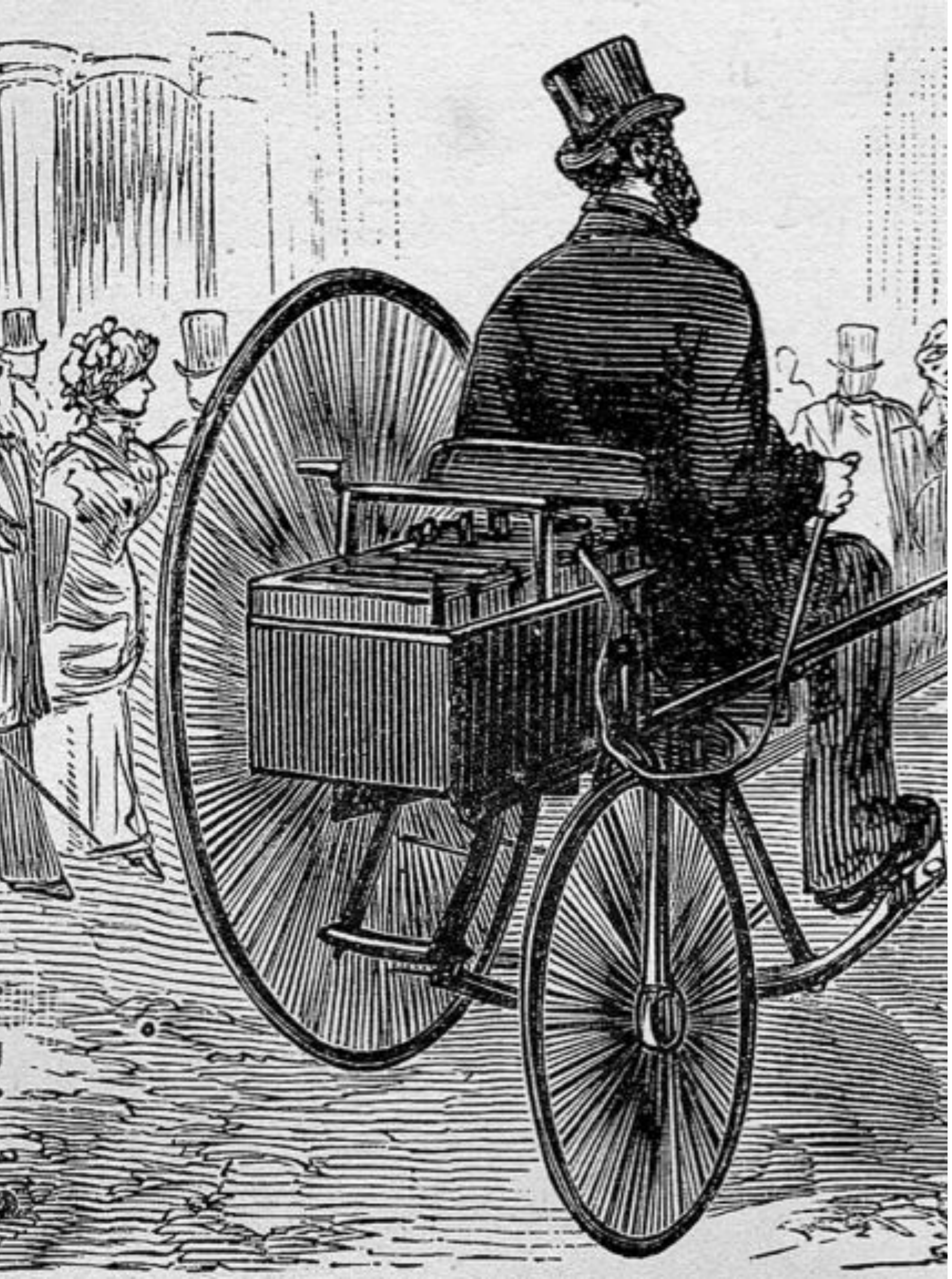
Secondmind Active Learning helps optimize powertrain design and development processes more efficiently and in less time than current techniques, with up to...

80%	50%	40%
less data	less time	less material

Everything you wanted to know about Electric Motors

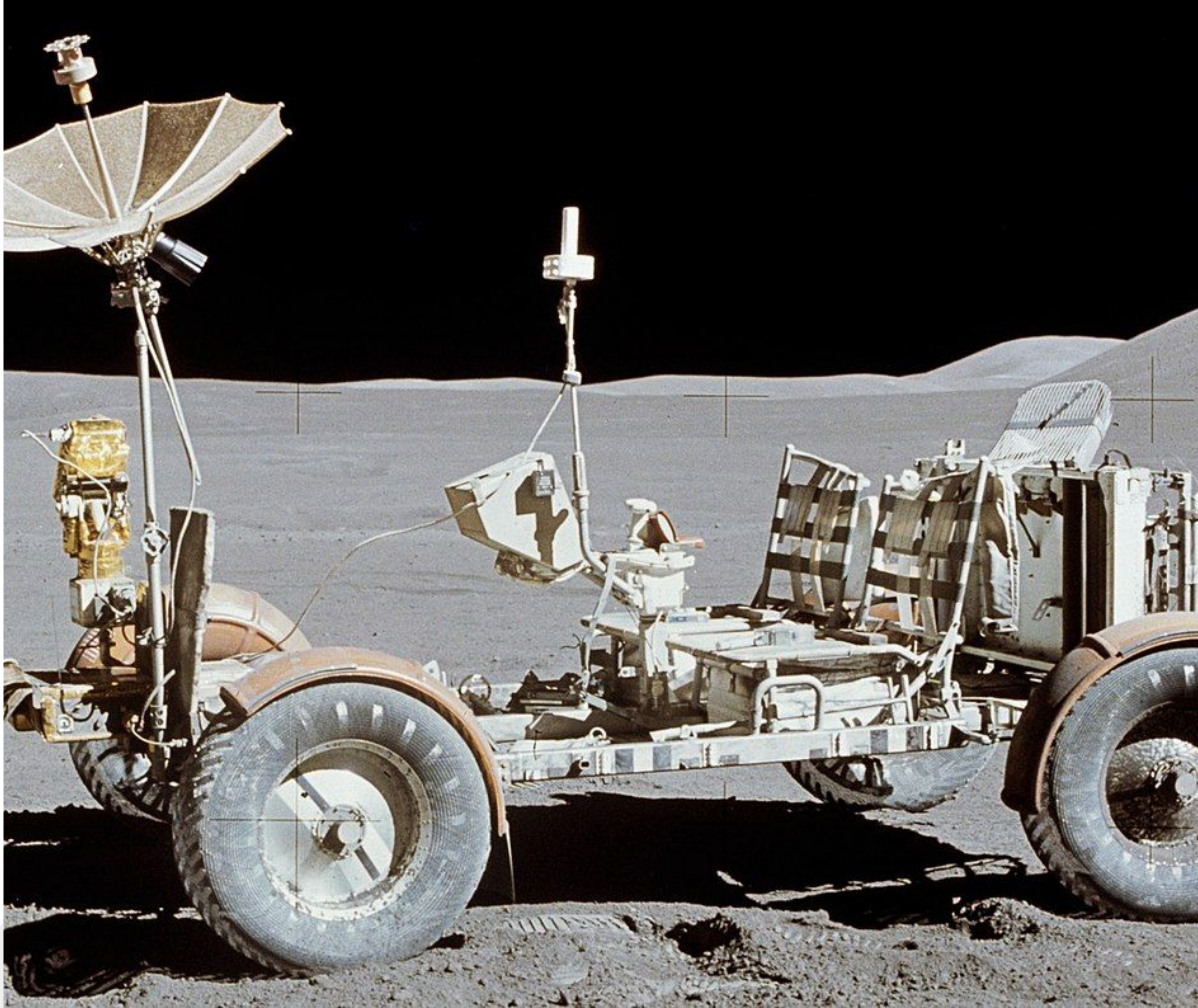


What year was the first all-electric car?



1881 - First all-electric car

1971 - First lunar car

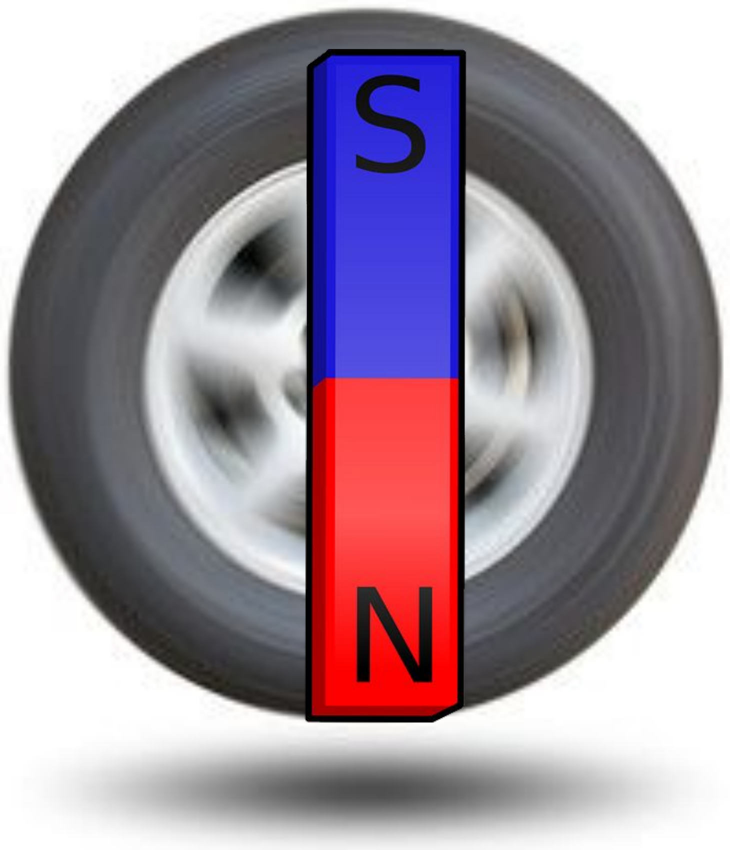


1985 - The C5

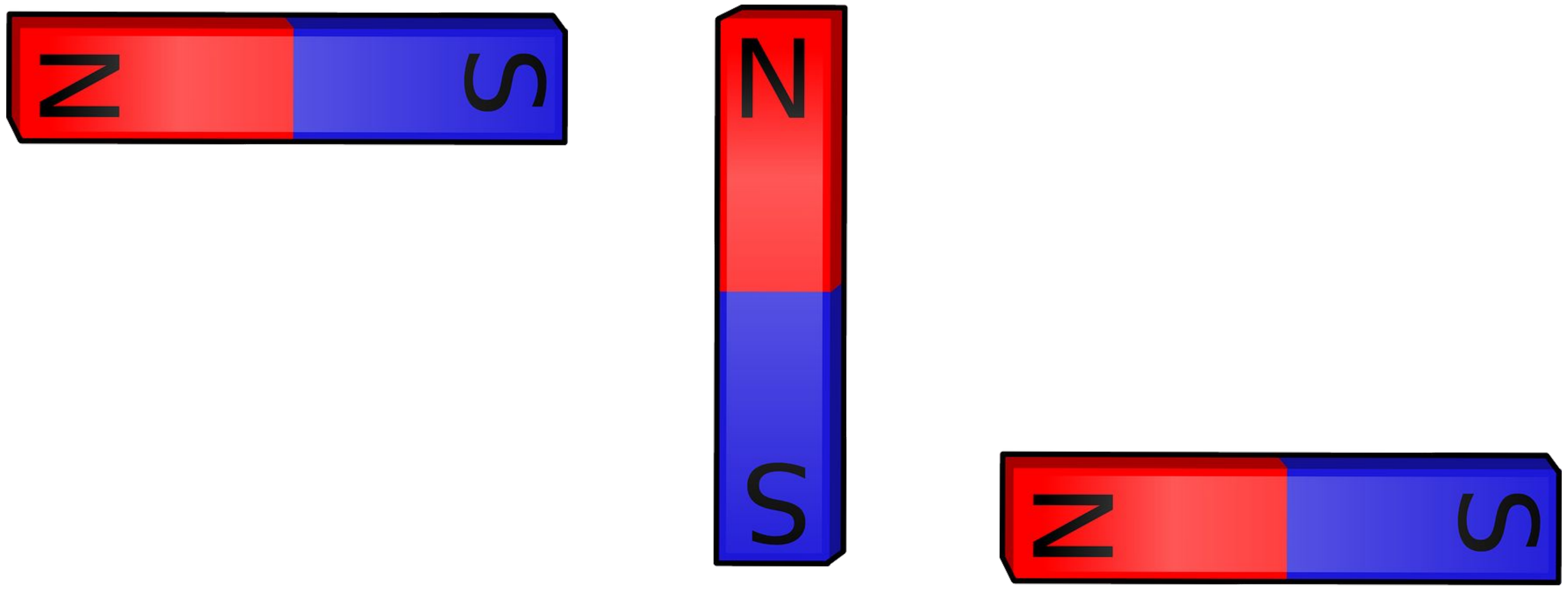
2010 - First electric hatchback

Electric motors

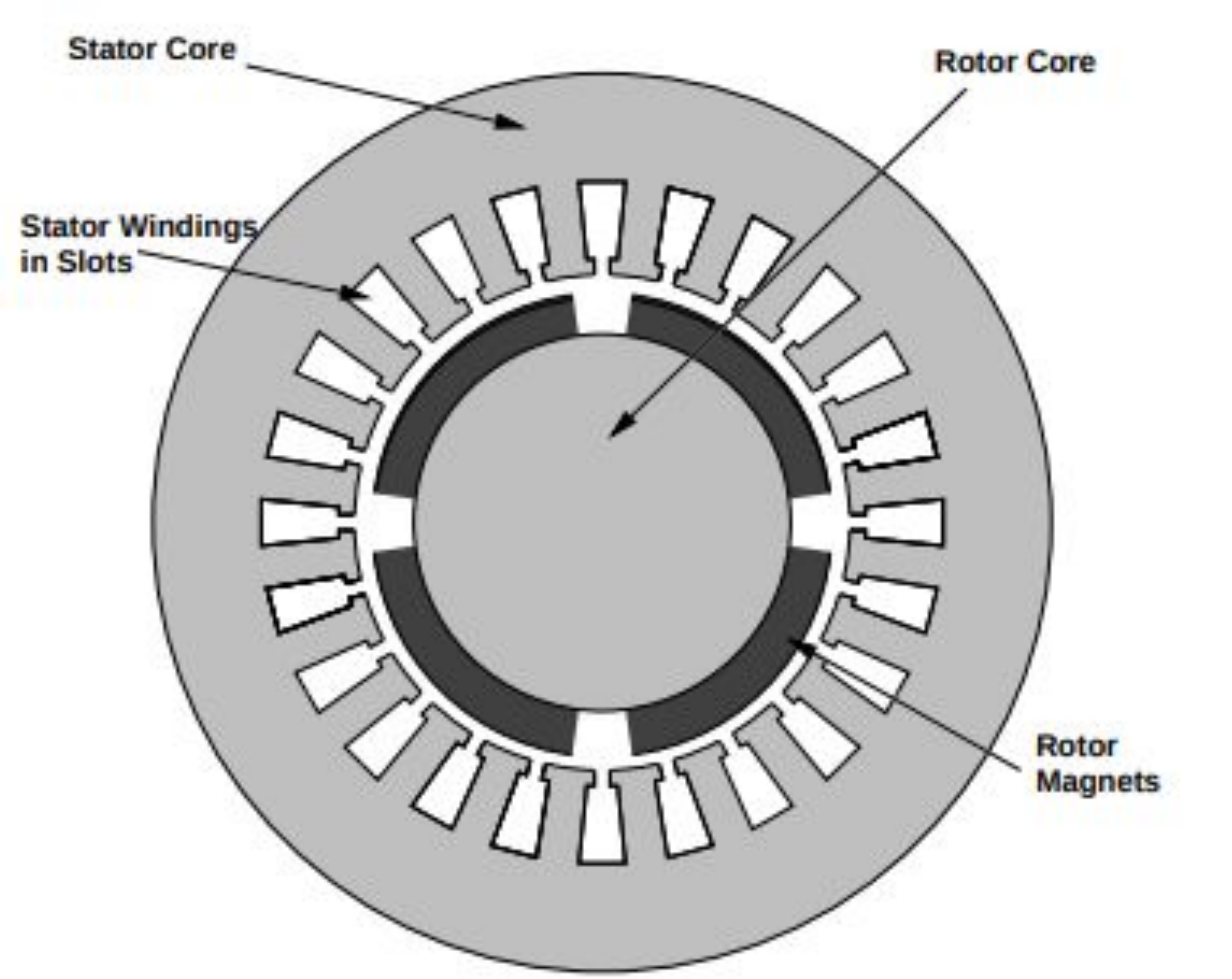
Electric motors



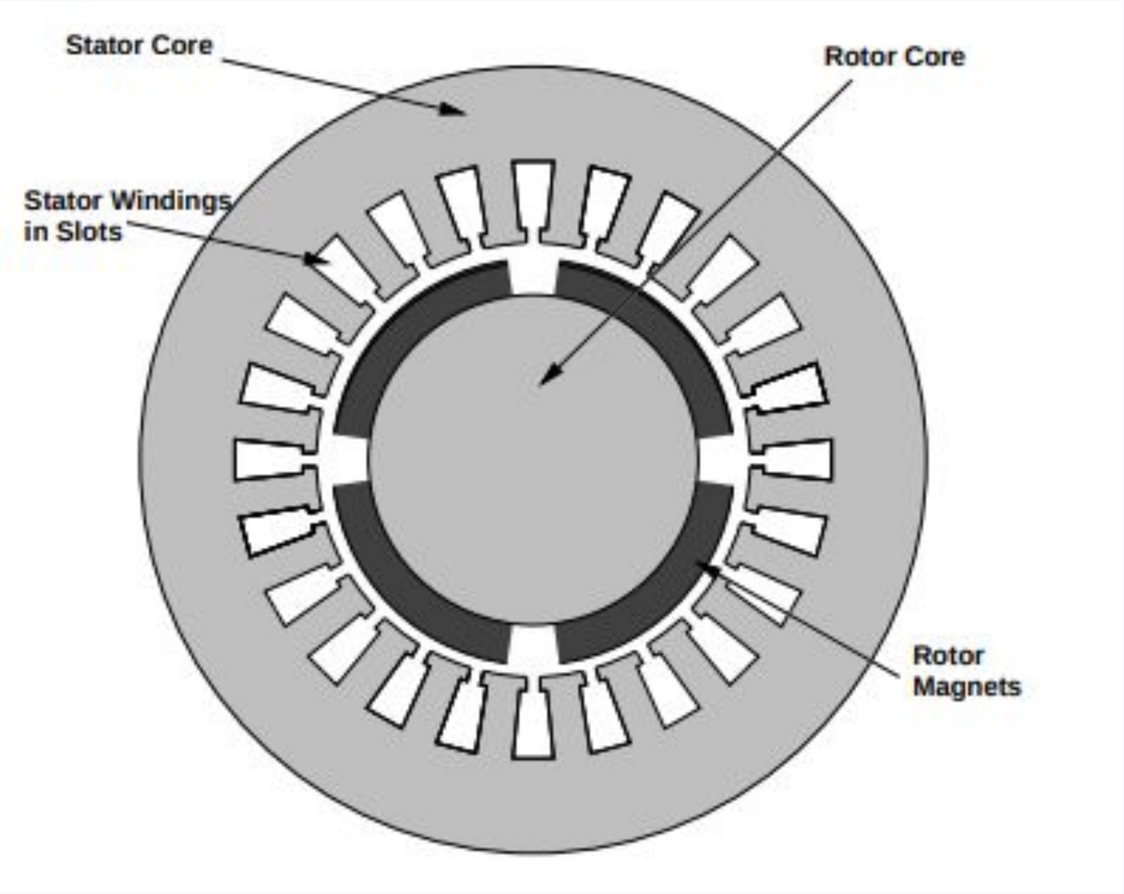
Electric motors

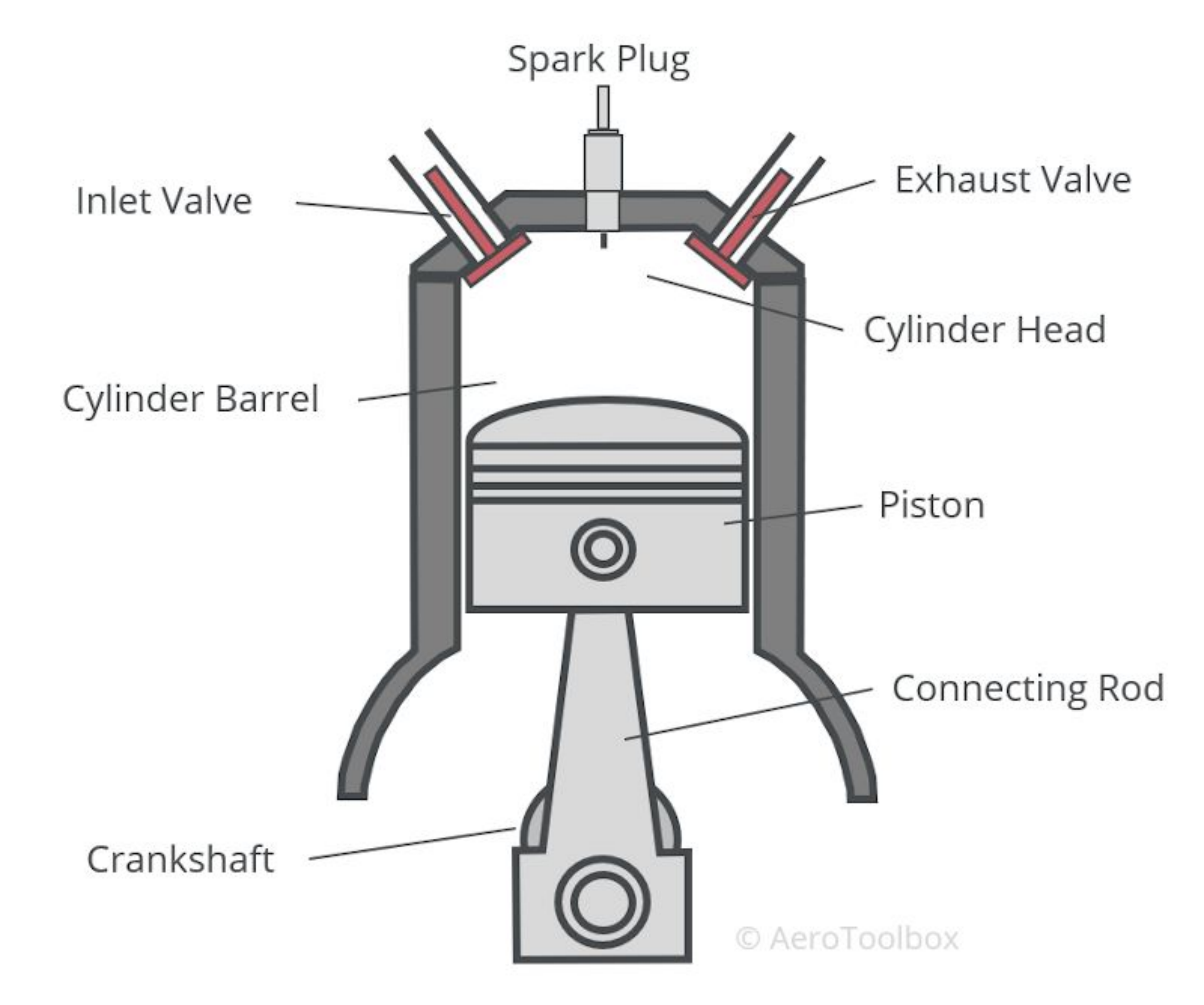


Background: Cross Section view of an EV Motor

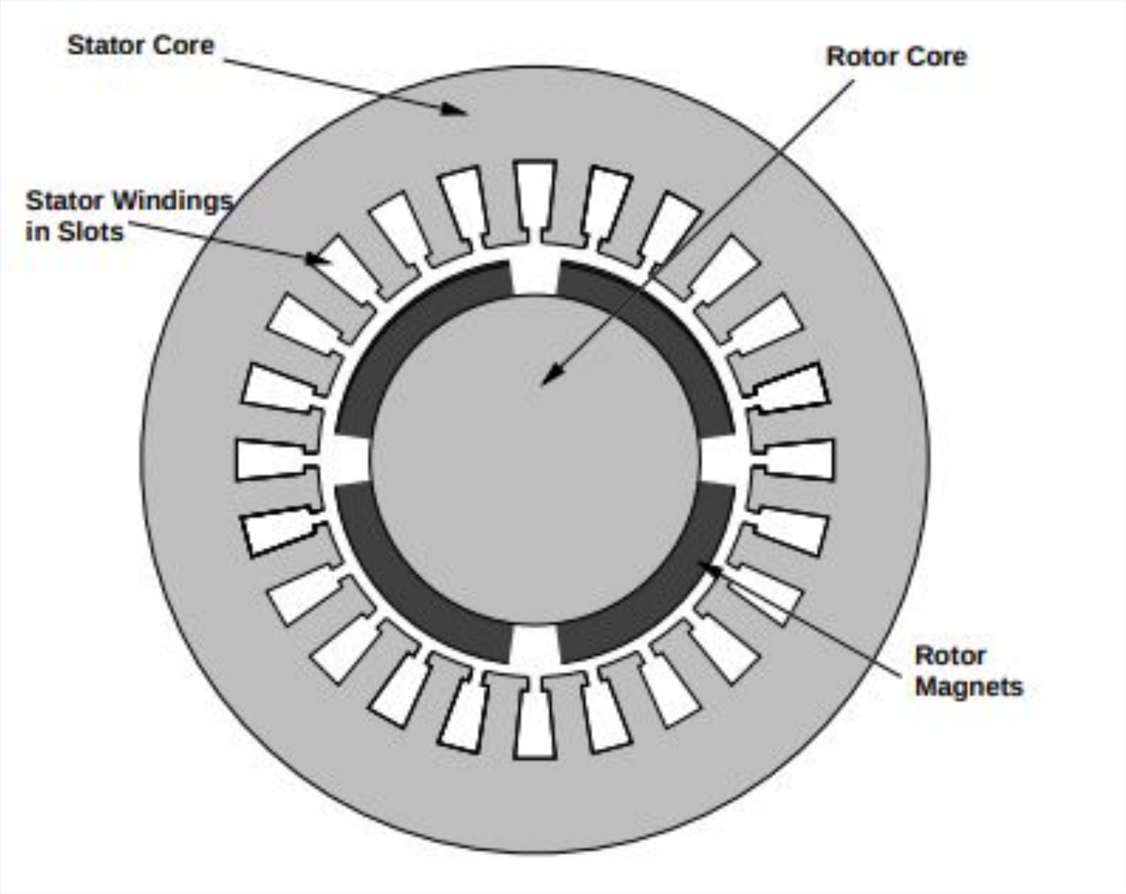


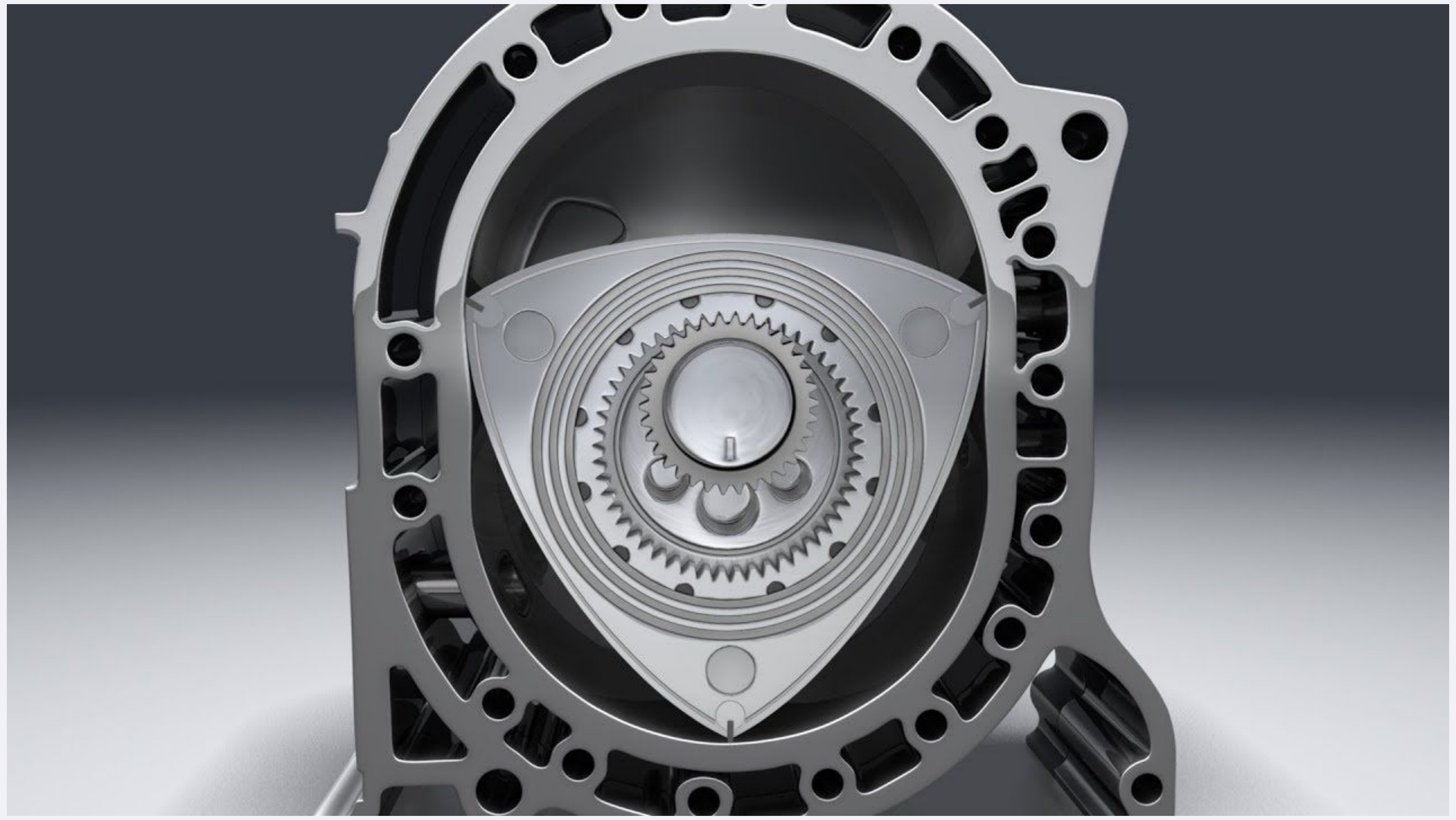
Comparison vs ICE





Comparison vs ICE





Electric Motors

Torque depends on:

The *strength* of the magnet: *Ia* - Current

The *location* of the magnet: *β* - Phase angle of the current

Secondmind

Electric Motors

Torque also depends on:

The rotation speed of the motor

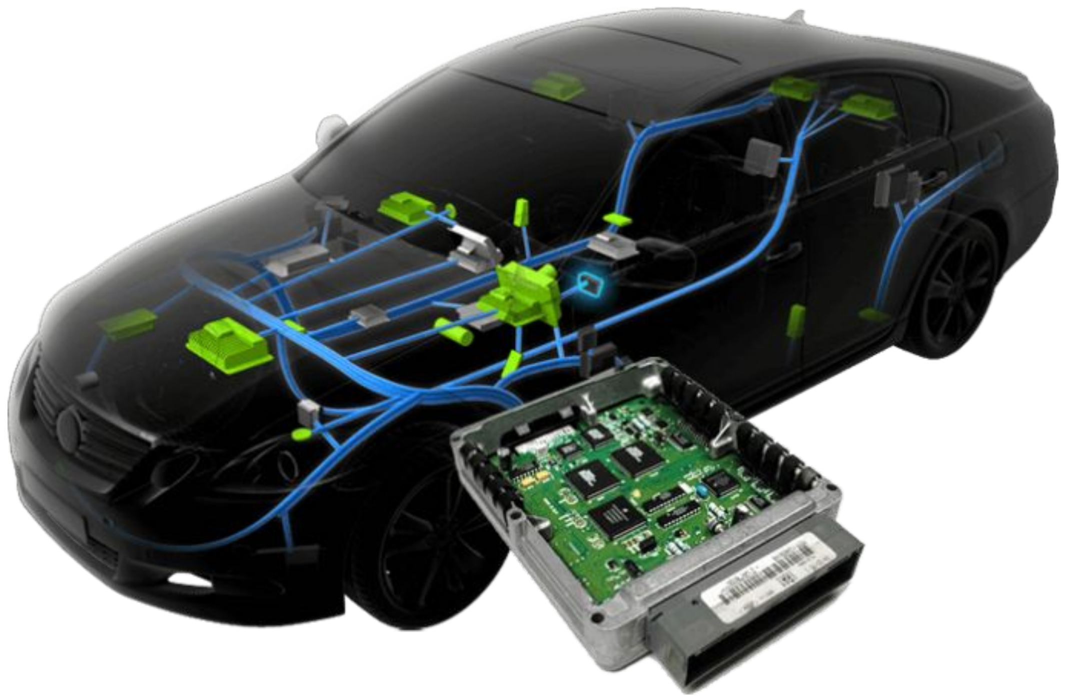
Voltage supplied by the battery

Temperature

Electric Motor Calibration

Engine control unit calibration

Goal: come up with a *look-up table*, a set of optimal engine configurations, given environmental conditions

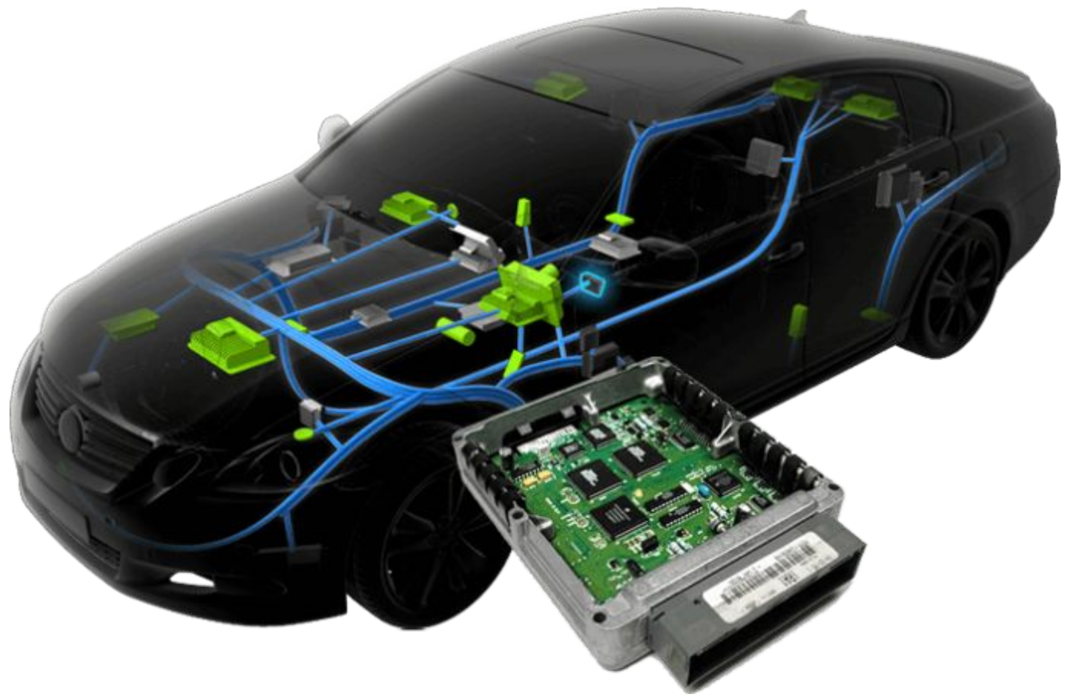


ECU

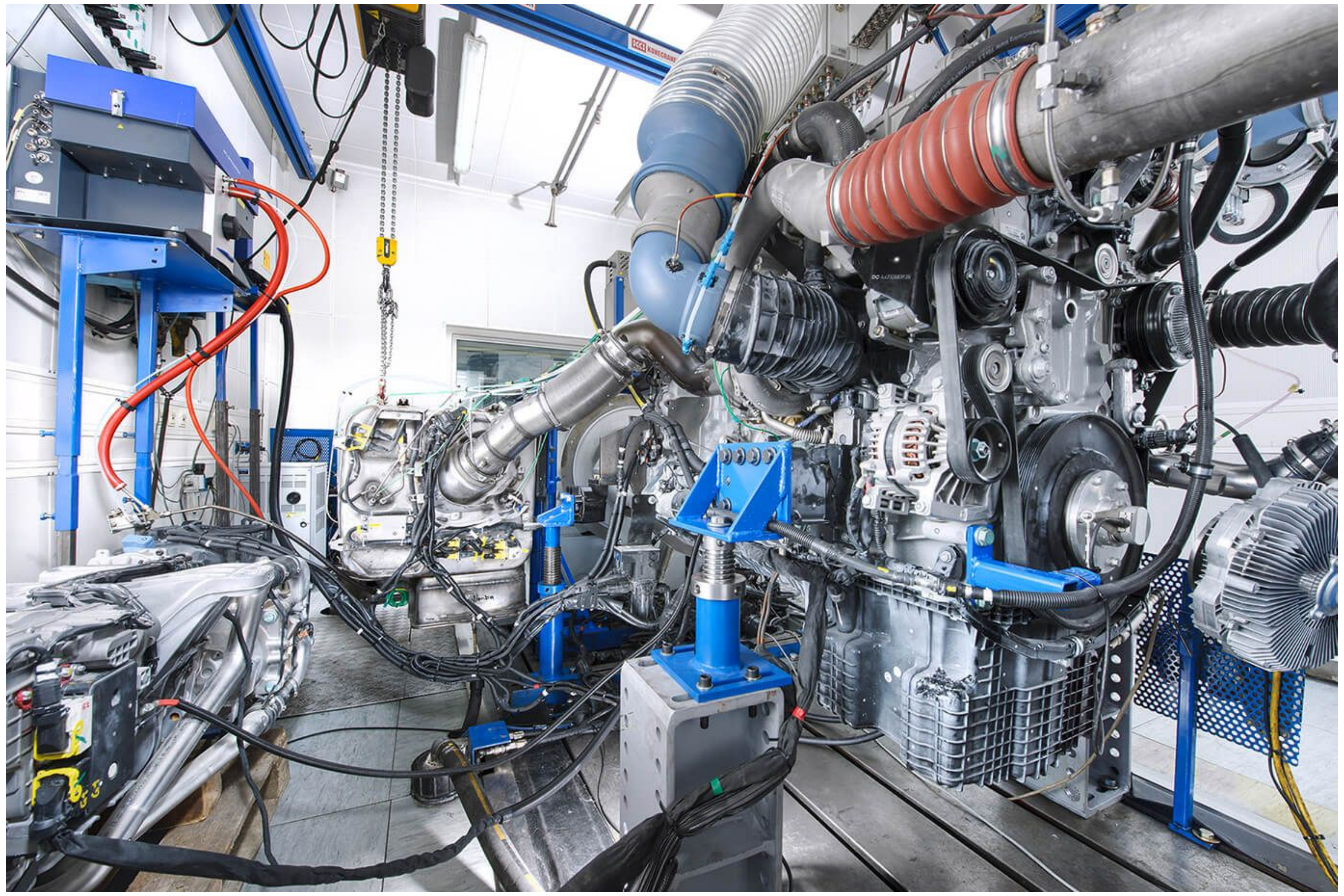
Electric Motor Calibration

Engine control unit calibration

Goal: come up with a *look-up table*, a set of optimal engine configurations, given environmental conditions



ECU



Test bench

Motor Calibration: Some Numbers

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs
- 2 objectives
- 1-3 constraints
- Need to find a look-up table = “profile optimum”
- Noise is heteroscedastic and overall budget = millions of observations

Motor Calibration: Some Numbers

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs ✅
- 2 objectives ✅
- 1-3 constraints ✅
- Need to find a look-up table = “profile optimum” ❓
- Noise is heteroscedastic and overall budget = millions of observations ❓

Motor Calibration: Some Numbers

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs ✅
- 2 objectives ✅
- 1-3 constraints ✅
- Need to find a look-up table = “profile optimum” ❓
- Noise is heteroscedastic and overall budget = millions of observations ❓

- 1 experiment takes < 1 minute
- 1 experiment delivers 100-1000 observations at a time
- Risk adversity
- Large/variable cost of preparing the motor for an experiment

Motor Calibration: Some Numbers

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs ✅
- 2 objectives ✅
- 1-3 constraints ✅
- Need to find a look-up table = “profile optimum” ❓
- Noise is heteroscedastic and overall budget = millions of observations ❓

- 1 experiment takes < 1 minute ❌
- 1 experiment delivers 100-1000 observations at a time ❌
- Risk adversity ❌
- Large/variable cost of preparing the motor for an experiment ❌

Motor Calibration: Some Numbers

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs
- 2 objectives
- 1-3 constraints
- Need to find a look-up table = "profile optimum"
- Noise is heteroscedastic and overall budget = millions of observations

- 1 experiment takes < 1 minute
- 1 experiment delivers 100-1000 observations at a time
- Risk adversity
- Large/variable cost of preparing the motor for an experiment

We need methods that

- scale really well with data,
- are quick,
- are robust (i.e. work all the time, not just **once** for our paper)

Secondmind

Motor Calibration

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs ✅
- 2 objectives ✅
- 1-3 constraints ✅
- Need to find a look-up table = "profile optimum" ❓
- Noise is heteroscedastic and overall budget = millions of observations ❓

- 1 experiment takes < 1 minute ❌
- 1 experiment delivers 100-1000 observations at a time ❌
- Risk adversity ❌
- Large/variable cost of preparing the motor for an experiment ❌

We need methods that

- scale really well with data,
- are quick,
- are robust (i.e. work all the time, not just **once** for our paper)

Motor Calibration

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs
- 2 objectives
- 1-3 constraints
- Need to find a look-up table = “profile optimum”
- Noise is heteroscedastic and overall budget = millions of observations

Bayesian Adaptive Reconstruction of Profile Optima (Ginsbourger et al. 2013)

- 1 experiment takes < 1 minute
- 1 experiment delivers 100-1000 observations at a time
- Risk adversity
- Large/variable cost of preparing the motor for an experiment

We need methods that

- scale really well with data,
- are quick,
- are robust (i.e. work all the time, not just **once** for our paper)

Motor Calibration

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs ✓
- 2 objectives ✓
- 1-3 constraints ✓
- Need to find a look-up table = “profile optimum” ?
- Noise is heteroscedastic and overall budget = millions of observations ?

Bayesian Adaptive Reconstruction of Profile Optima (Ginsbourger et al. 2013)

Chained Gaussian Processes (Saul, Hensman, Vehtari, Lawrence et al 2016.)

- 1 experiment takes < 1 minute ✗
- 1 experiment delivers 100-1000 observations at a time ✗
- Risk adversity ✗
- Large/variable cost of preparing the motor for an experiment ✗

We need methods that

- scale really well with data,
- are quick,
- are robust (i.e. work all the time, not just **once** for our paper)

Motor Calibration

How can we use existing BO stuff and what new innovations do we need?

- 6-10 inputs ✓
- 2 objectives ✓
- 1-3 constraints ✓
- Need to find a look-up table = “profile optimum” ?
- Noise is heteroscedastic and overall budget = millions of observations ?

Bayesian Adaptive Reconstruction of Profile Optima (Ginsbourger et al. 2013)

Chained Gaussian Processes (Saul, Hensman, Vehtari, Lawrence et al 2016.)

- 1 experiment takes < 1 minute
- 1 experiment delivers 100-1000 observations at a time
- Risk adversity
- Large/variable cost of preparing the motor for an experiment

???

We need methods that

- scale really well with data,
- are quick,
- are robust (i.e. work all the time, not just **once** for our paper)

Remainder of the Talk

A story of ML development in industry

Remainder of the Talk

A story of ML development in industry

1. BO recap

Remainder of the Talk

A story of ML development in industry

1. BO recap
2. First steps: Motor calibration Proof Of Concept (POC)
    a. Profile optimisation
    b. Scalable heteroscedastic Gaussian processes

Remainder of the Talk

A story of ML development in industry

1. BO recap
2. First steps: Motor calibration Proof Of Concept (POC)
    a. Profile optimisation
    b. Scalable heteroscedastic Gaussian processes
3. Next steps: Research fun
    i. Smooth BO
    ii. Custom sparse models for BO
    iii. Risk averse BO

Secondmind

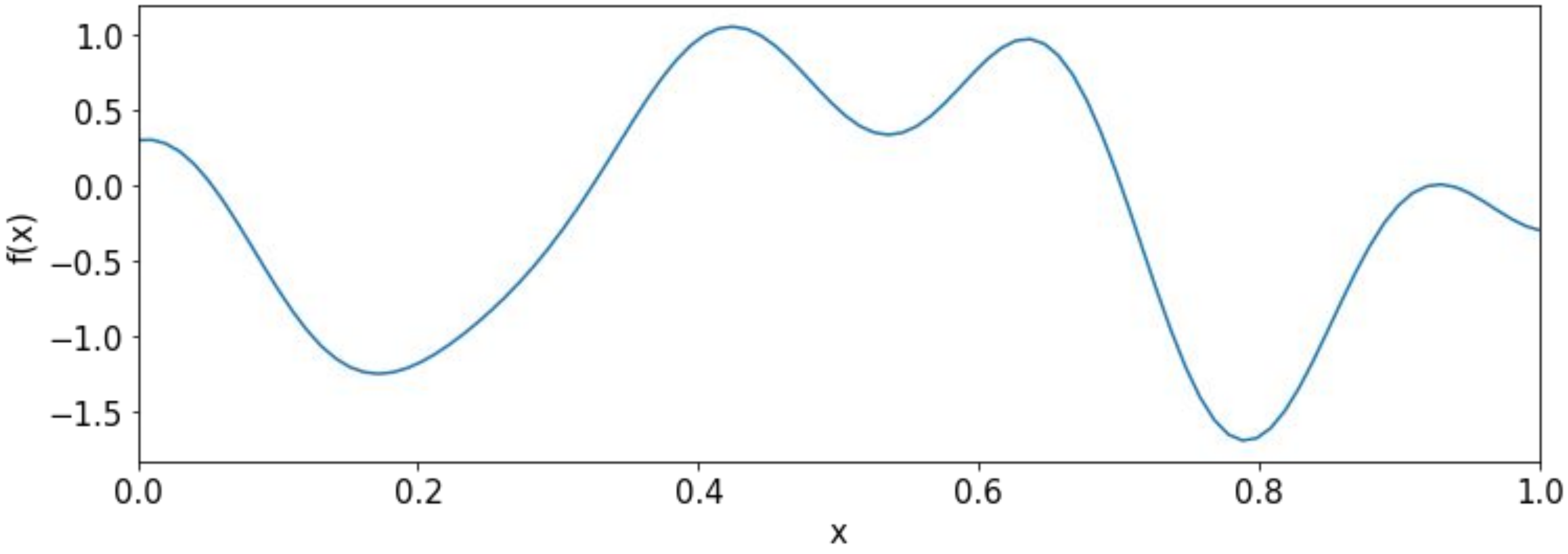
Bayesian Optimization Recap

Model-based global optimization

BO Demo

Let's find the maximum of a 1D function:

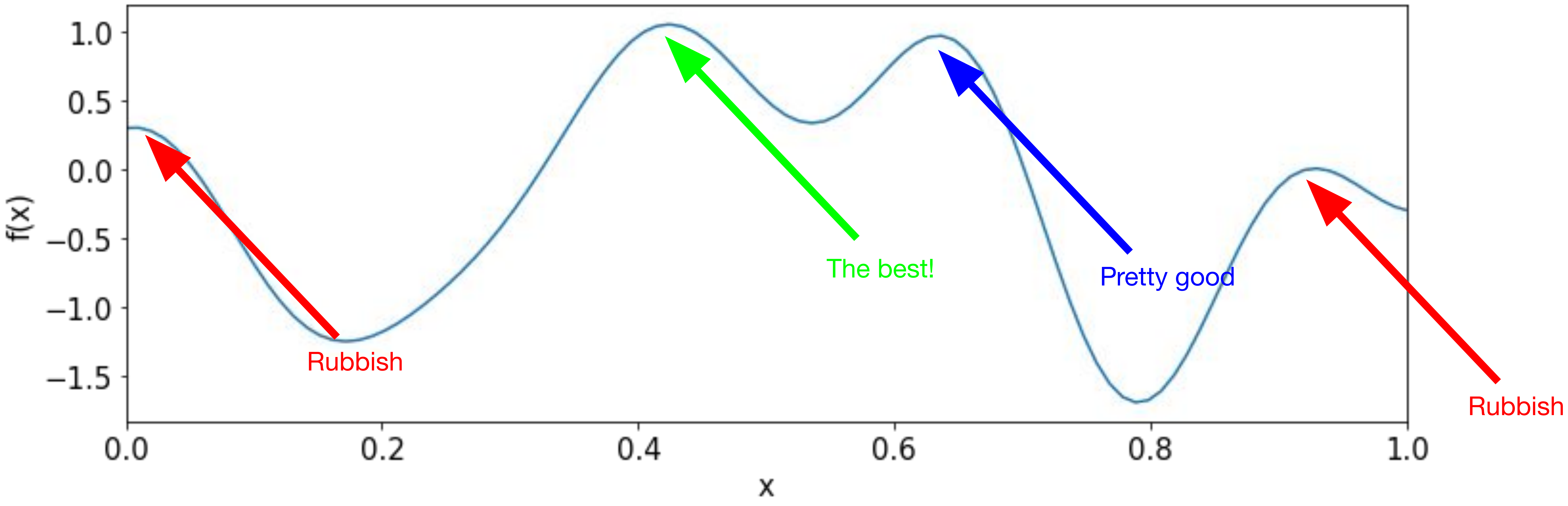
Using as **few** function evaluations as possible!



BO Demo

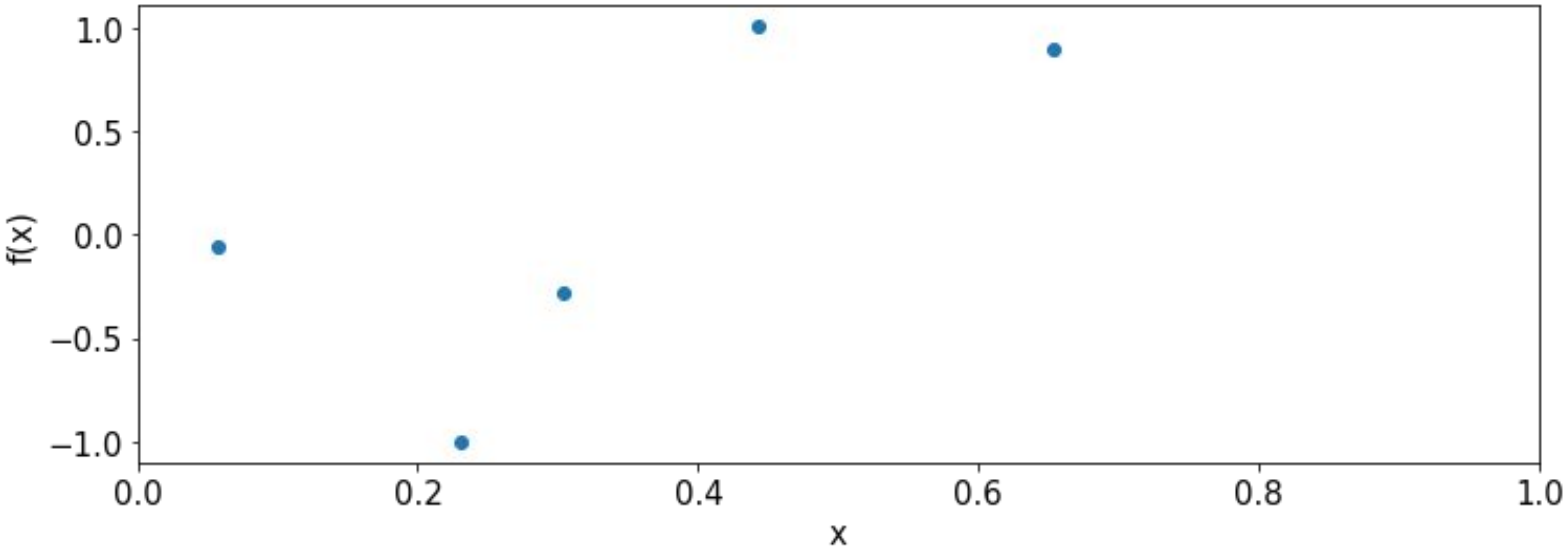
Let's find the maximum of a 1D function:

Using as **few** function evaluations as possible!



BO Demo

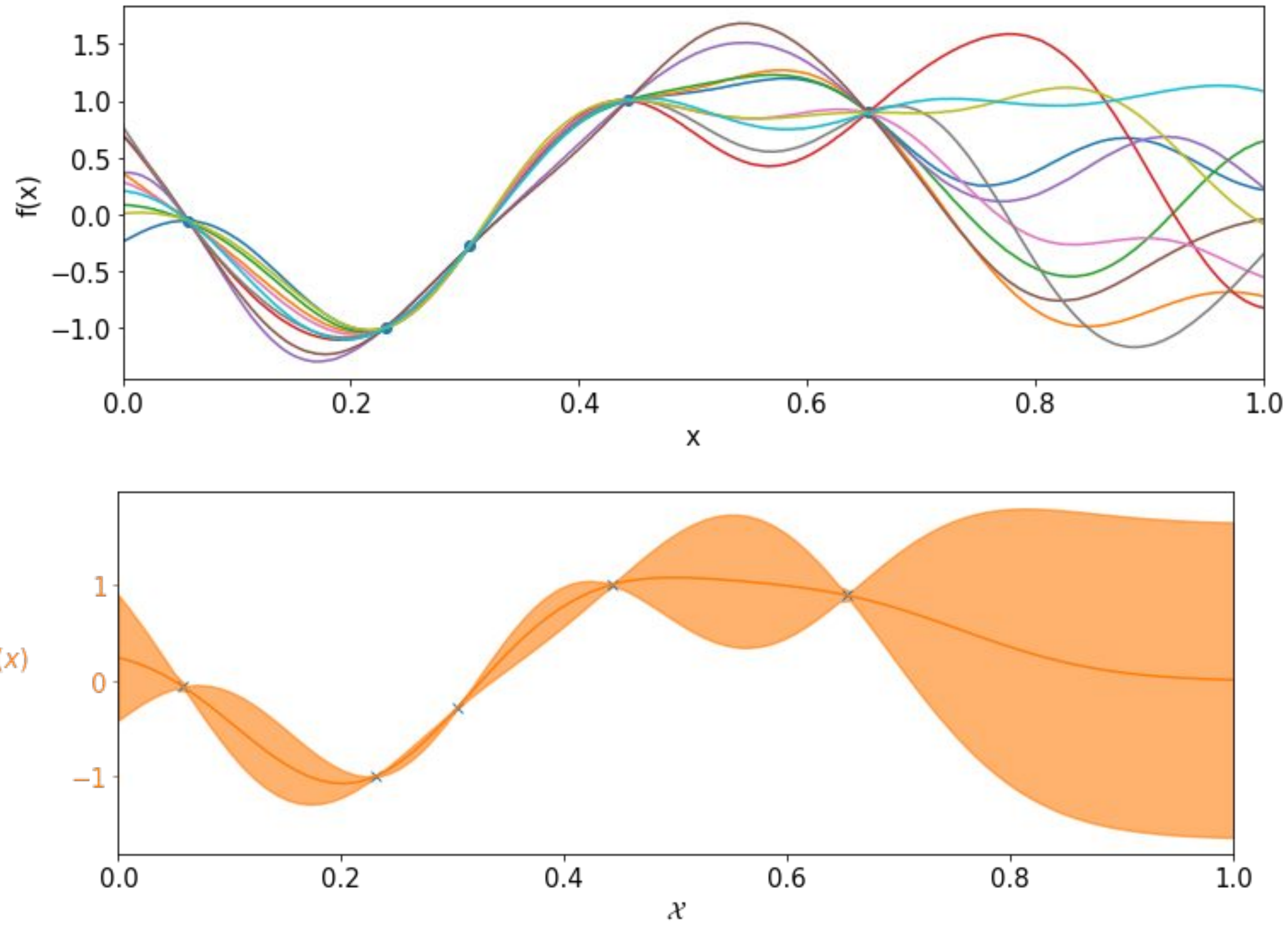
Suppose we make 5 evaluations



Where should we next evaluate? Explore/Exploit?

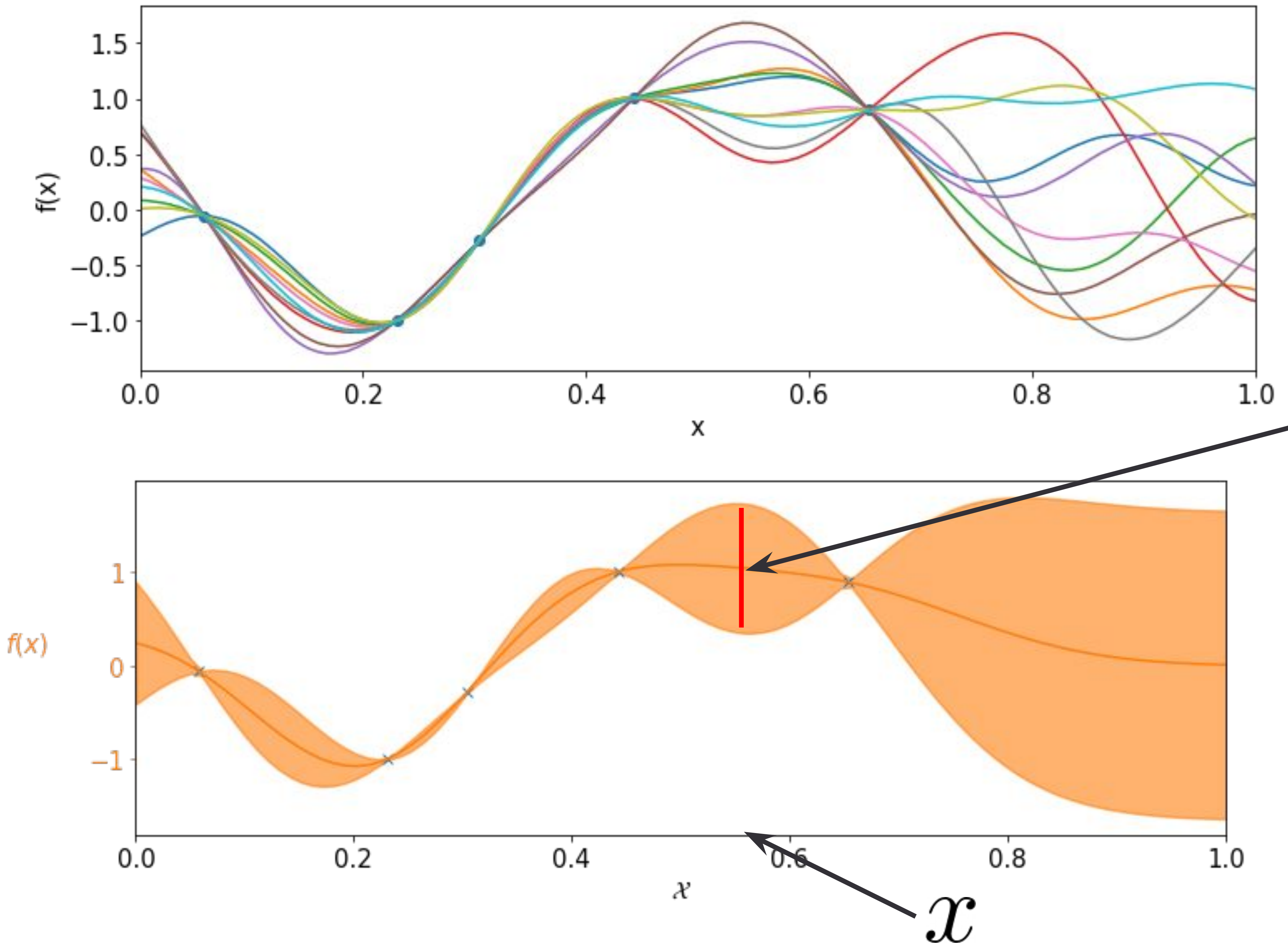
How to automate BO: step 1

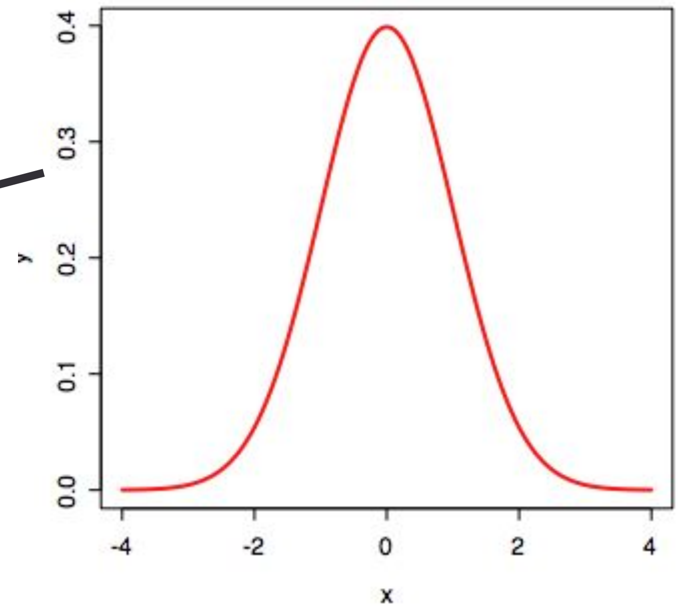
Use a statistical model like a Gaussian process



How to automate BO: step 1

Use a statistical model like a Gaussian process

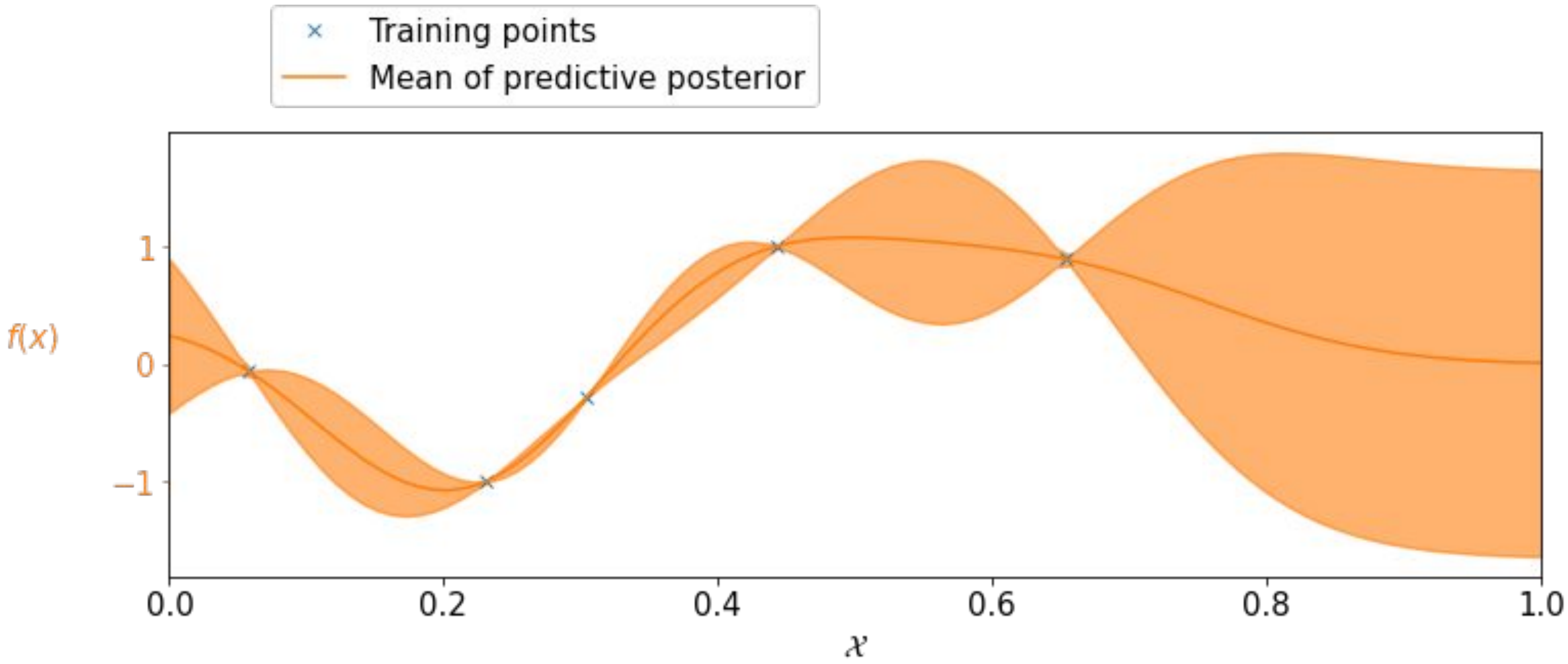




$$f(x) \sim N(\mu(x), \sigma^2(x))$$

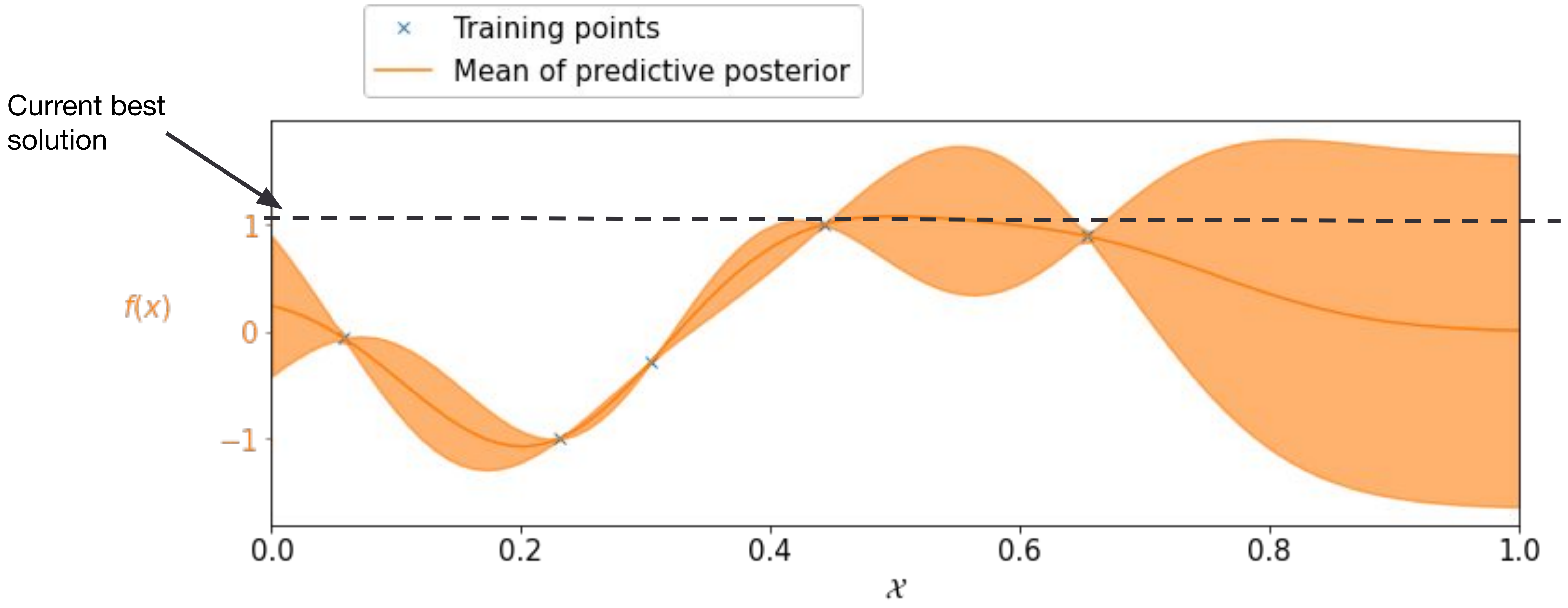
How to automate BO: step 2

Automated decision making via an acquisition function like expected improvement



How to automate BO: step 2

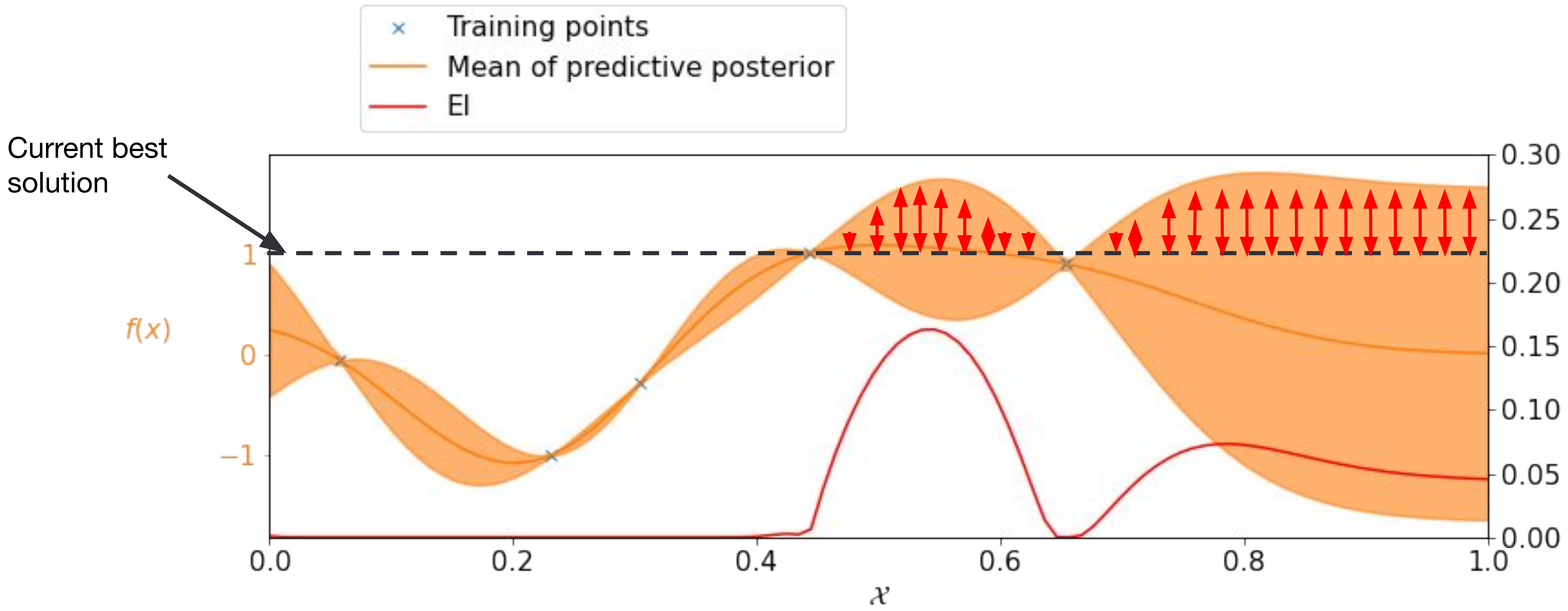
Automated decision making via an acquisition function like expected improvement



Secondmind

How to automate BO: step 2

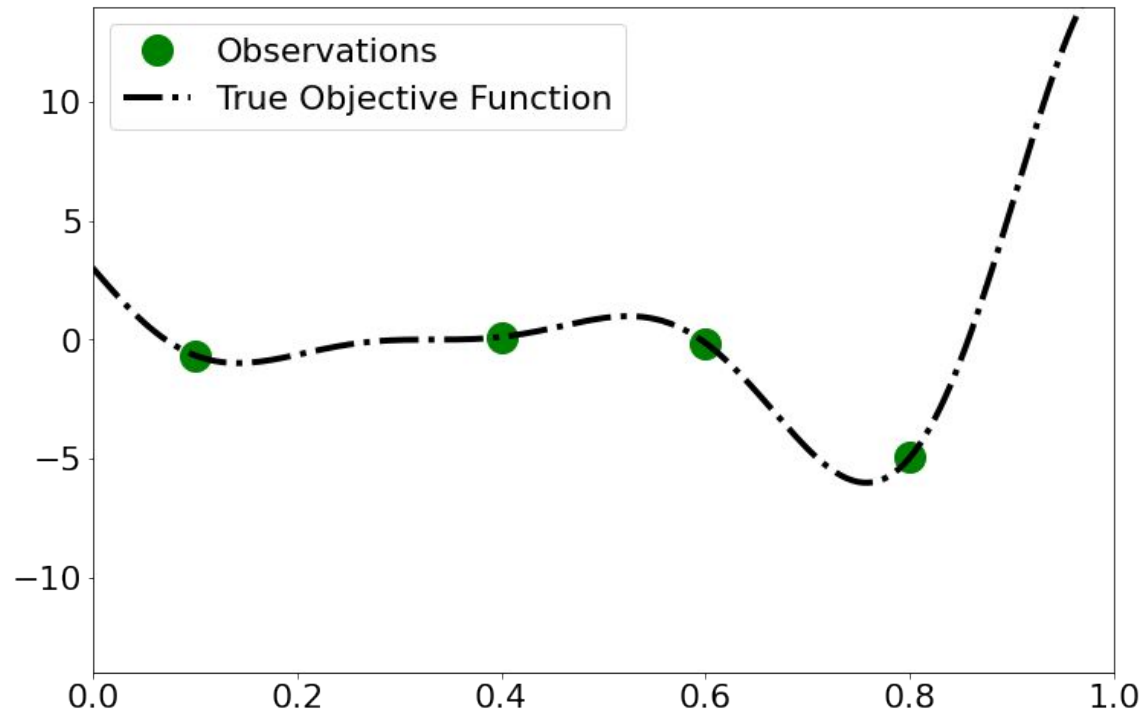
Automated decision making via an acquisition function like expected improvement



Secondmind

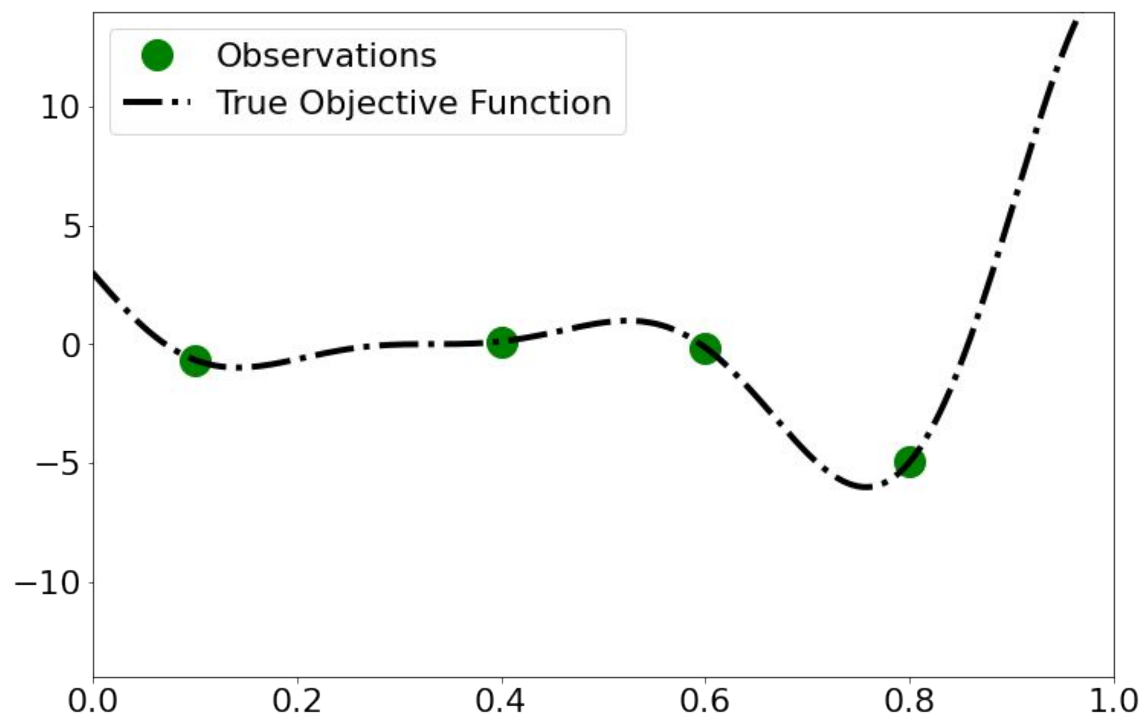
Expected Improvement

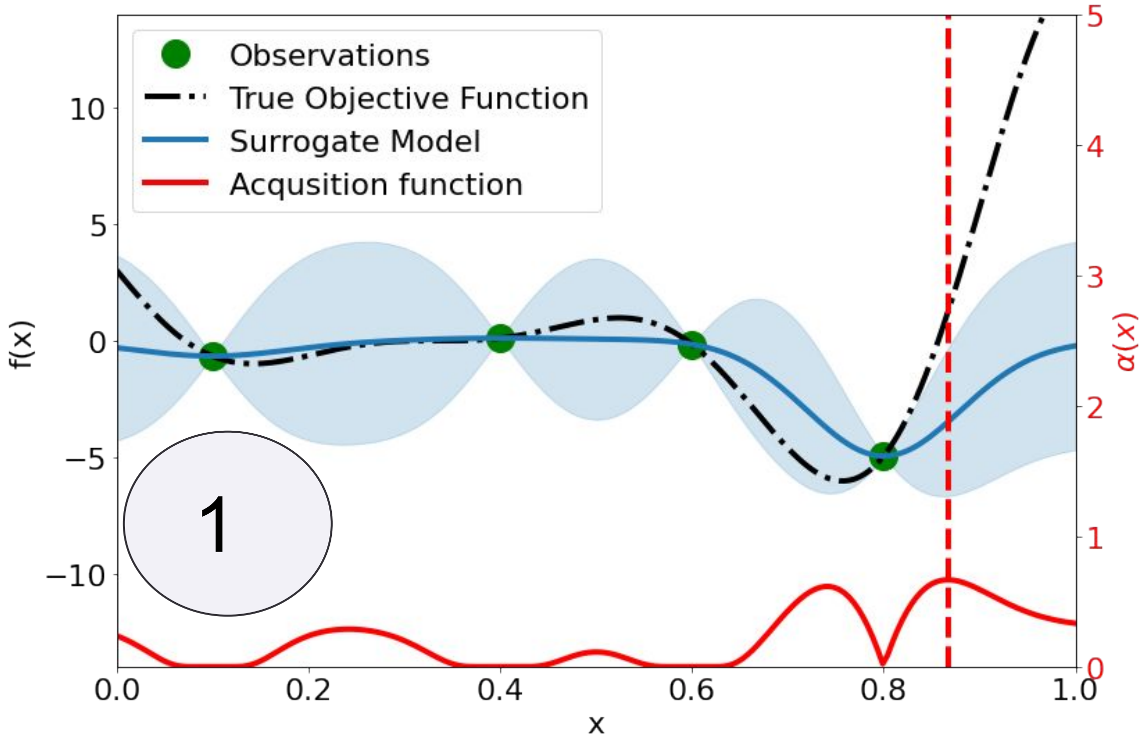
Demo BO loop



Expected Improvement

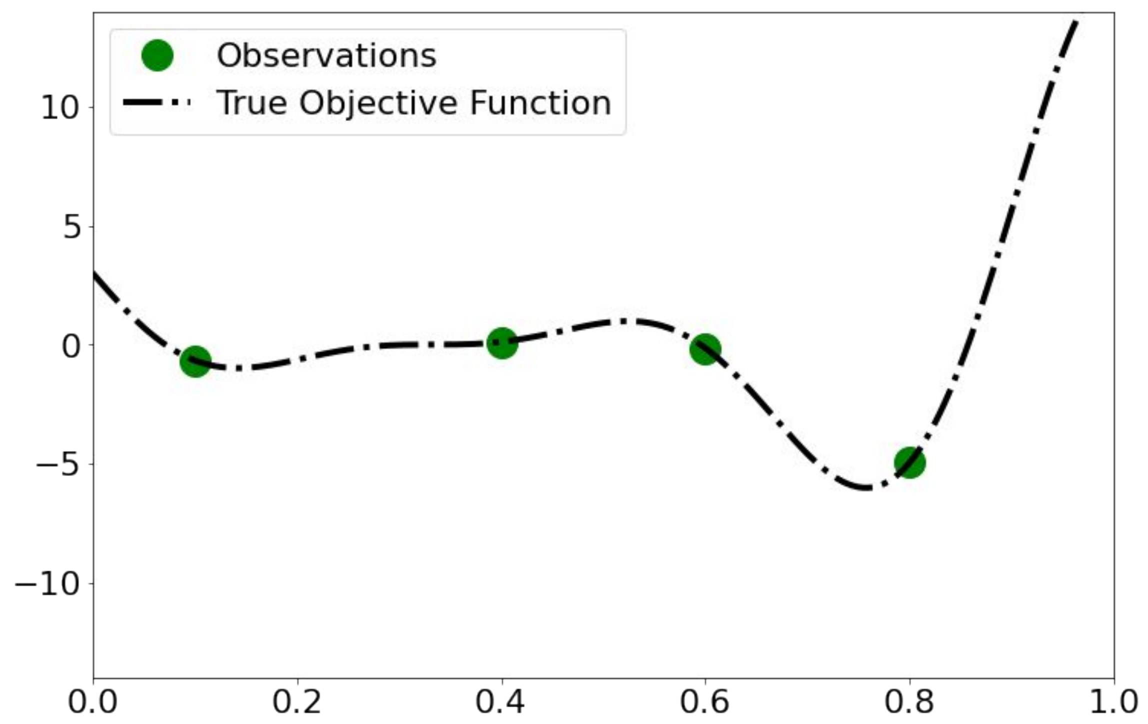
Demo BO loop

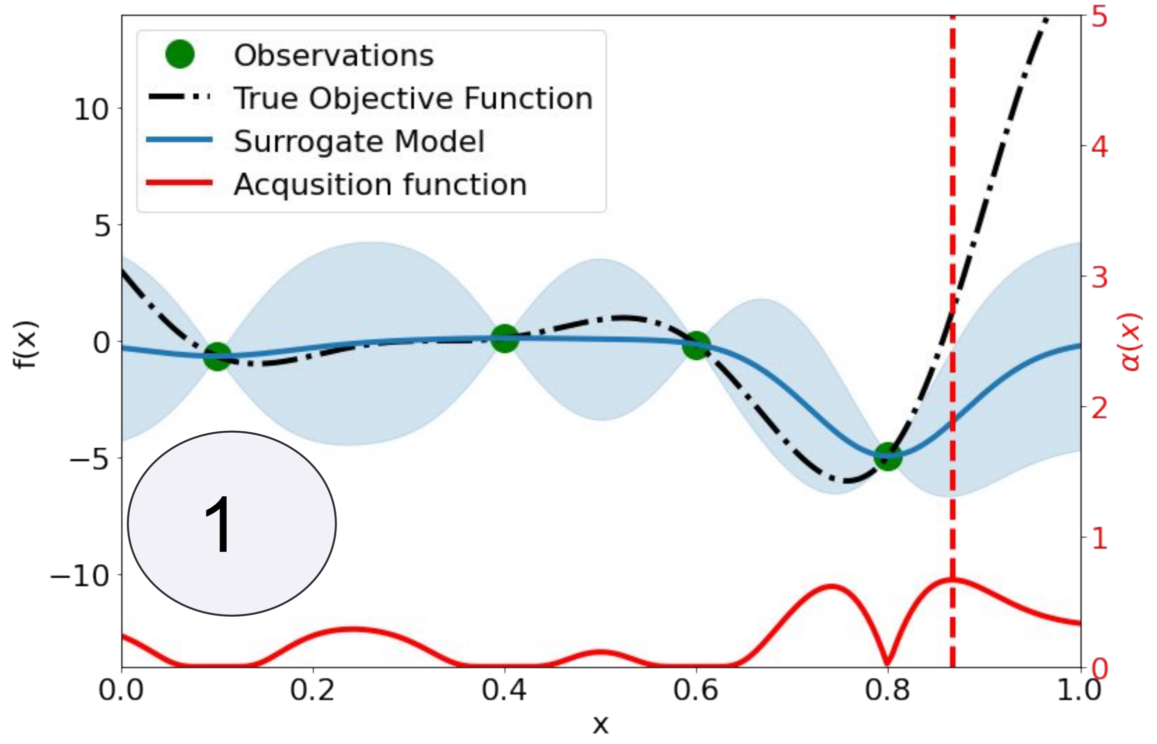


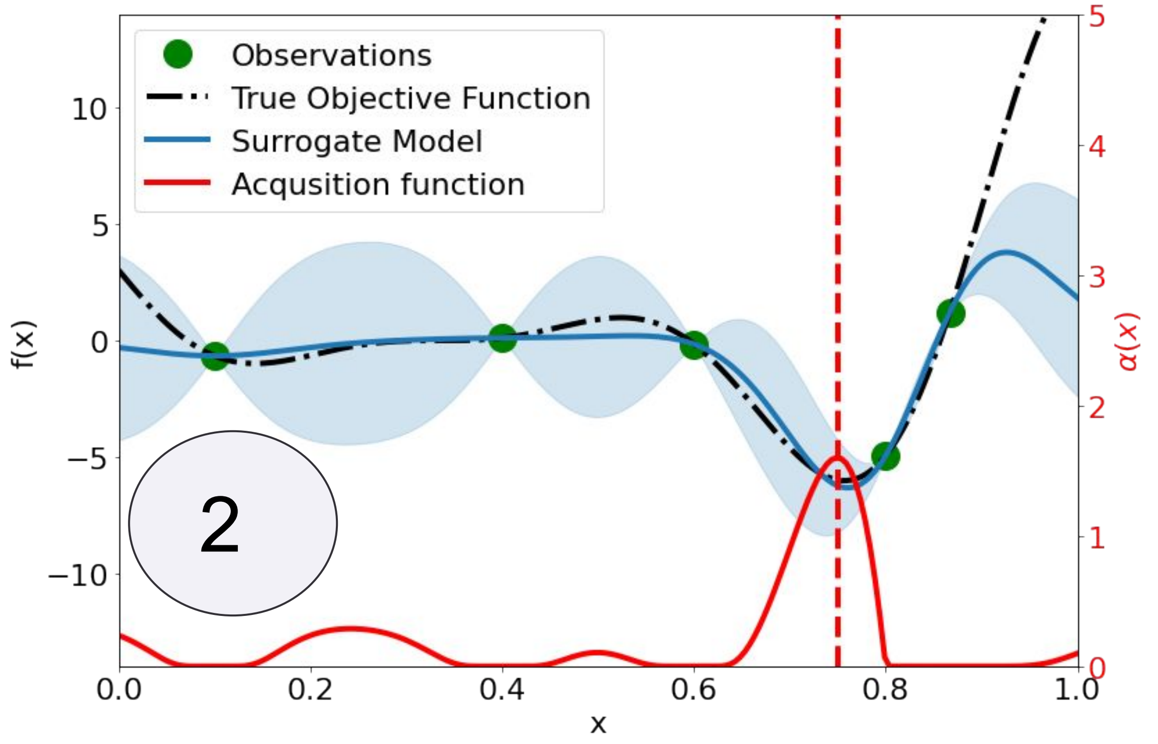


Expected Improvement

Demo BO loop



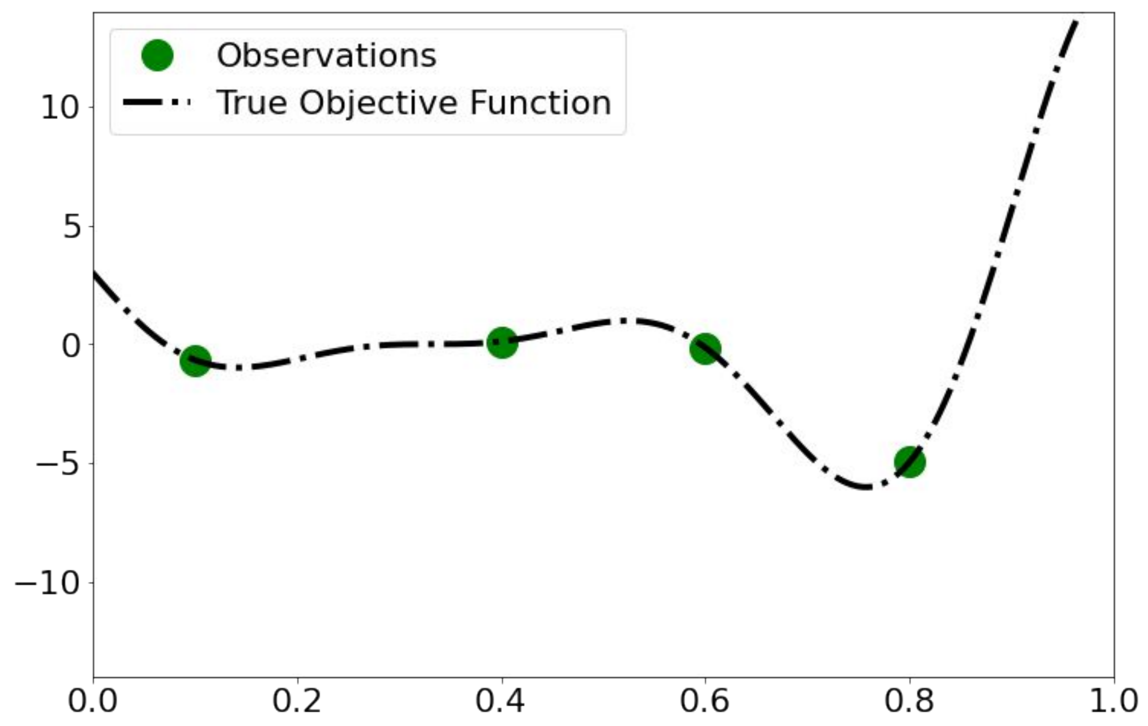


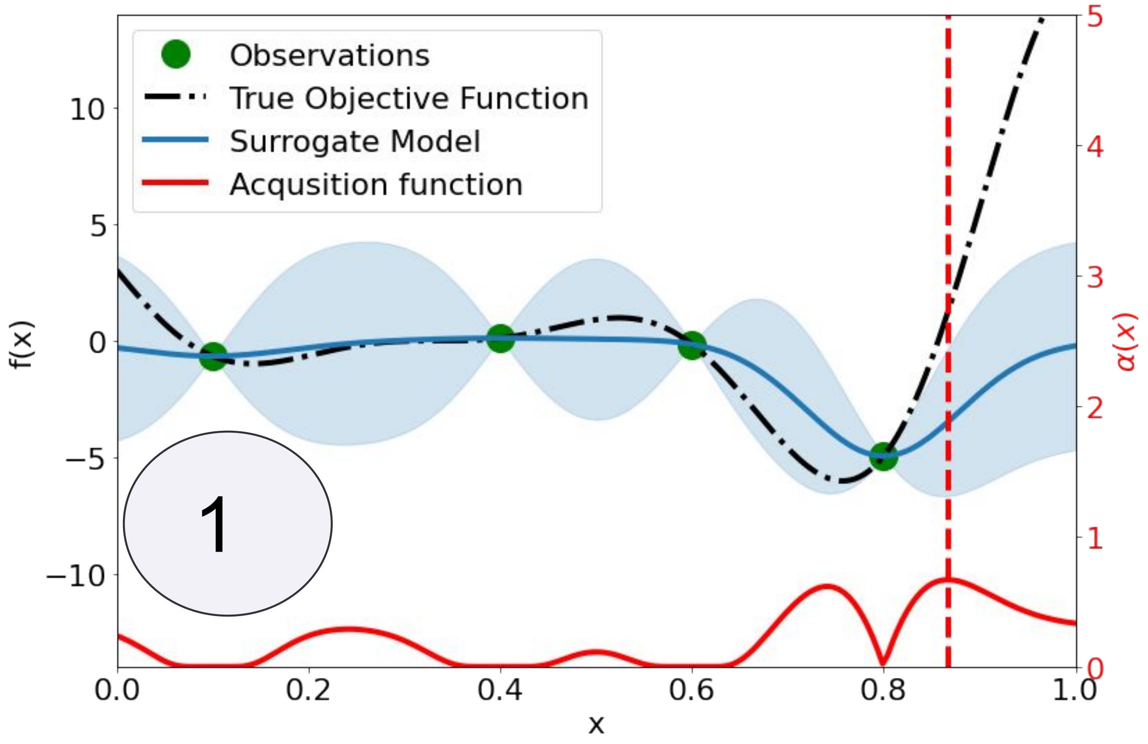


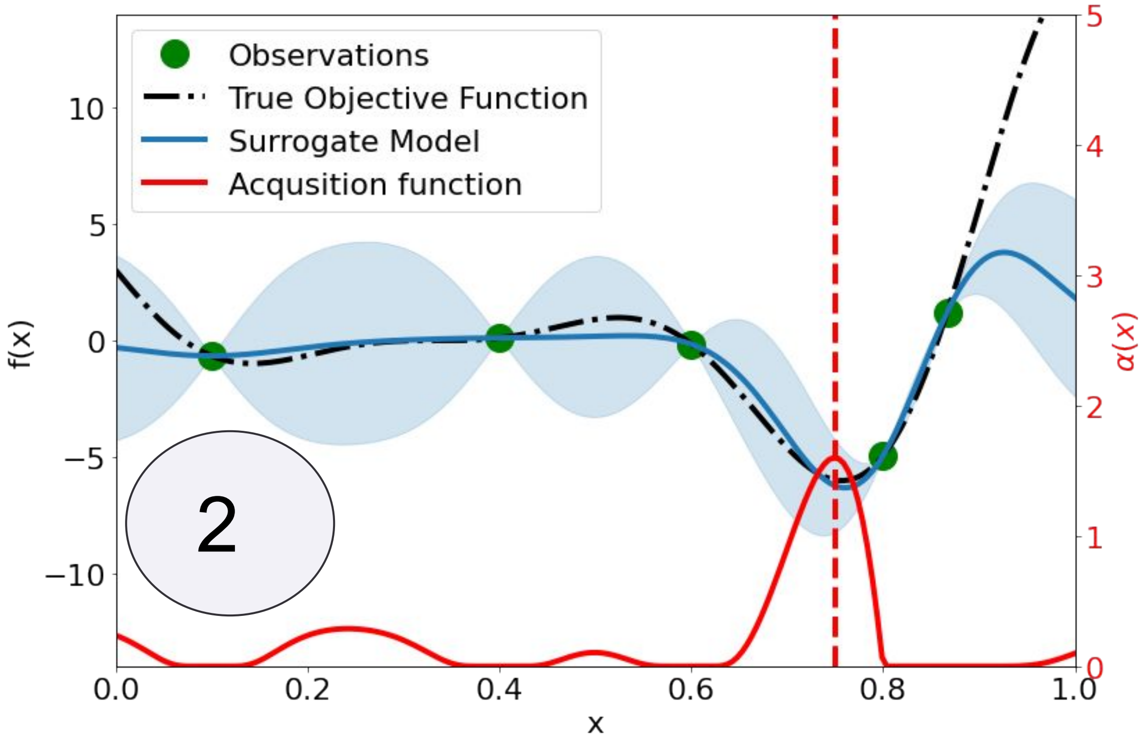
Secondmind

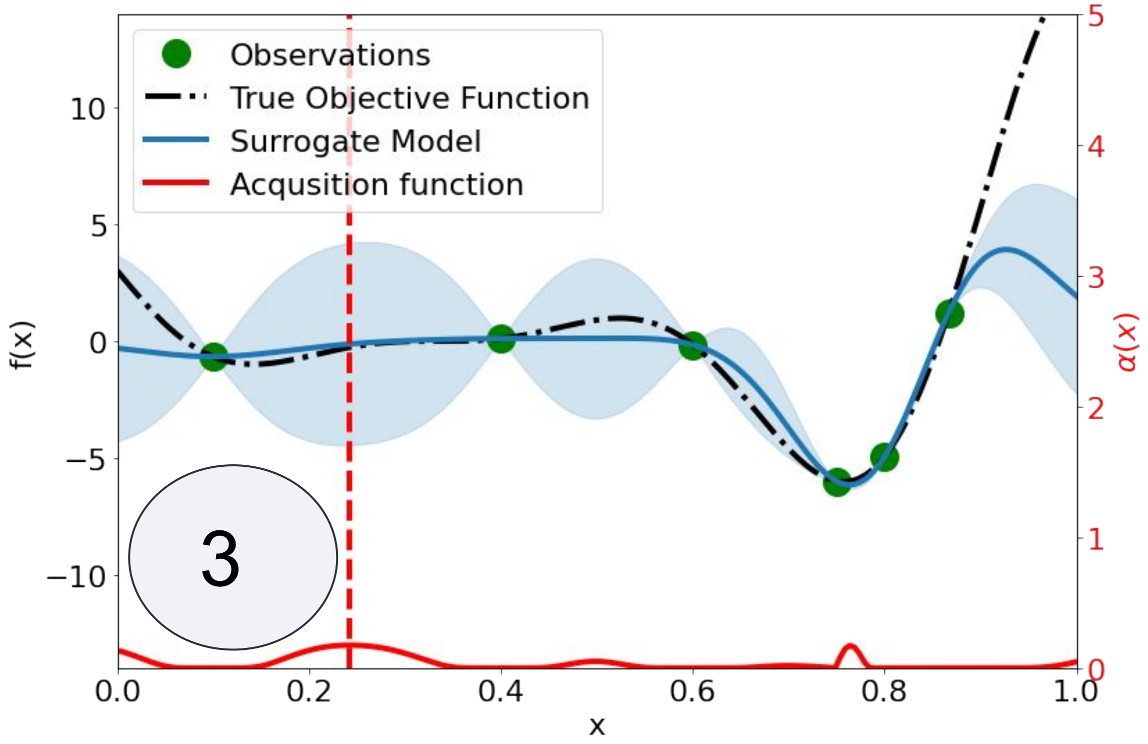
Expected Improvement

Demo BO loop





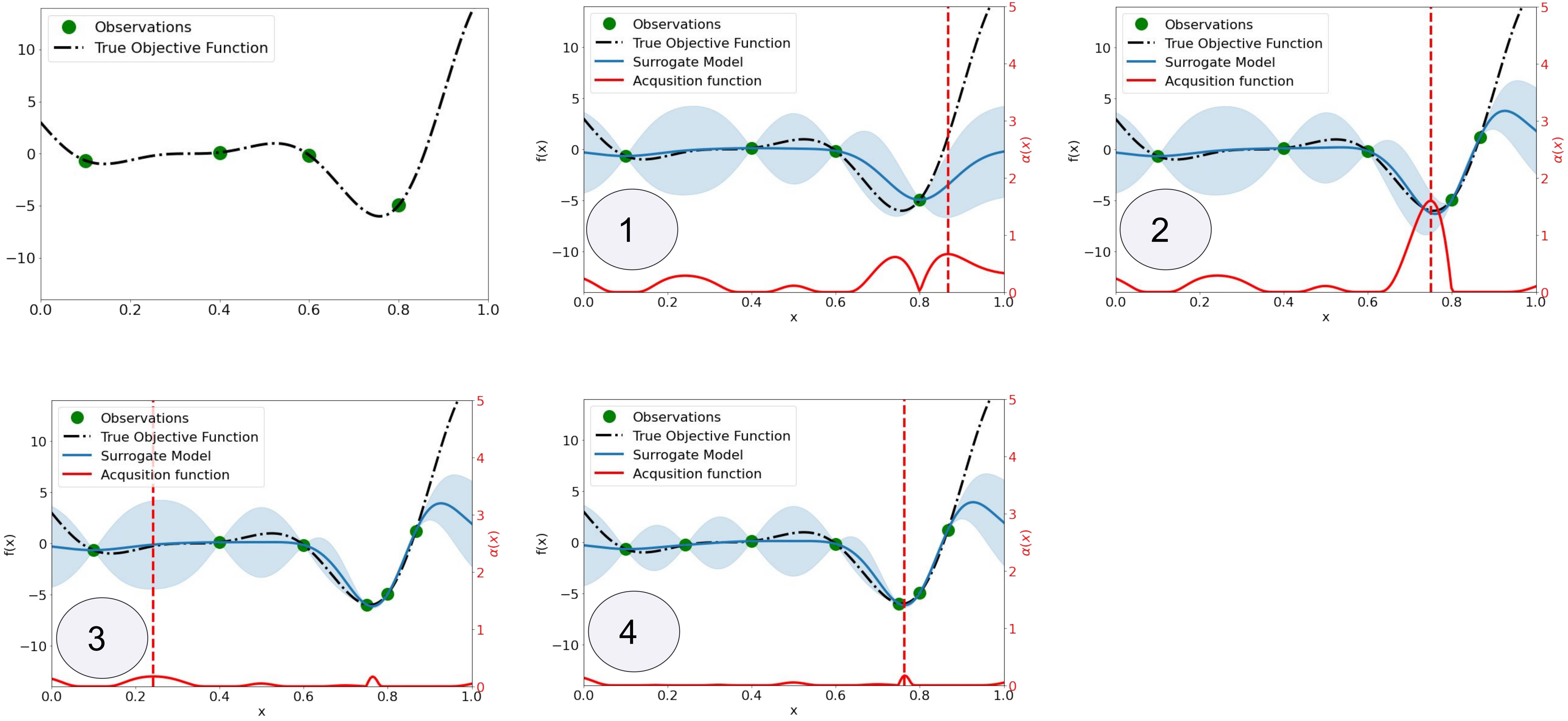




Secondmind

Expected Improvement

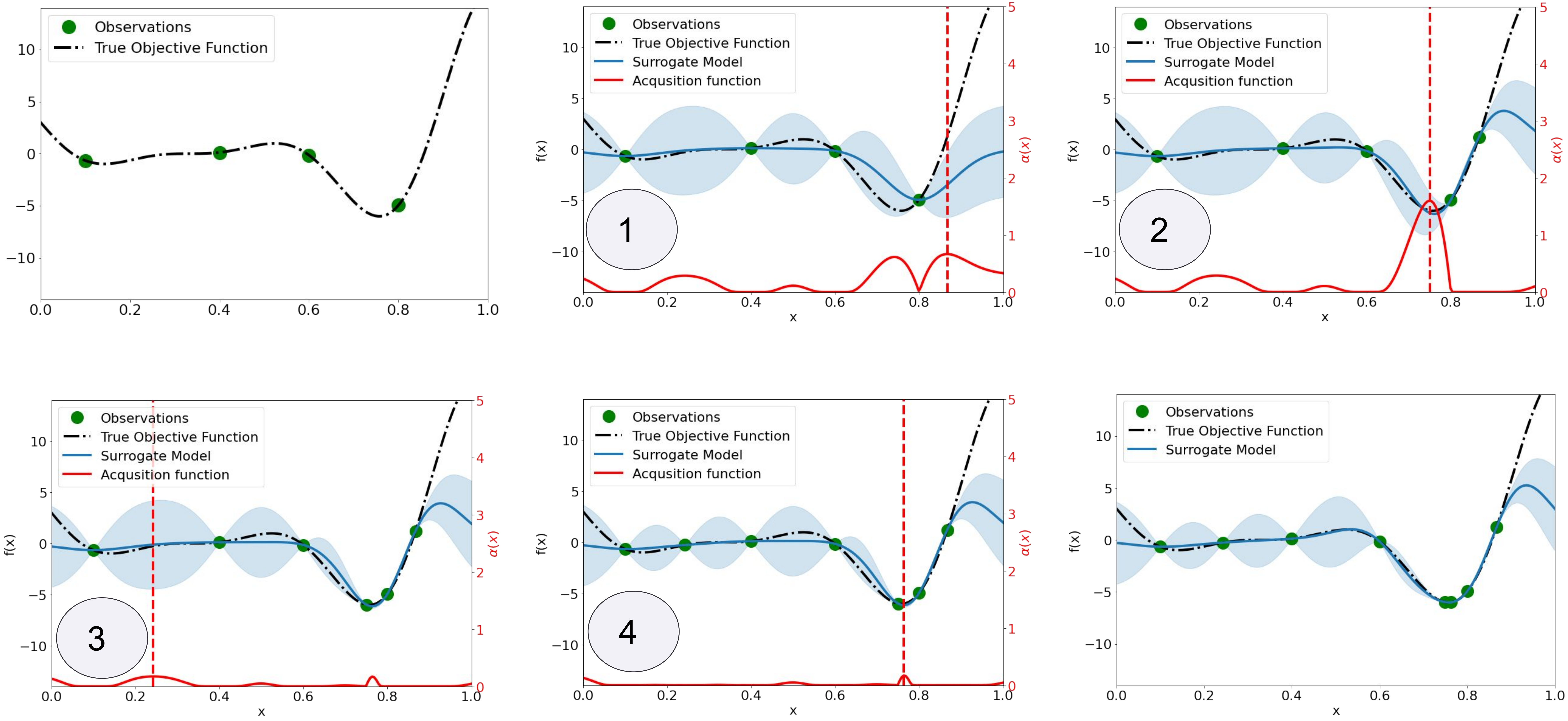
Demo BO loop



Secondmind

Expected Improvement

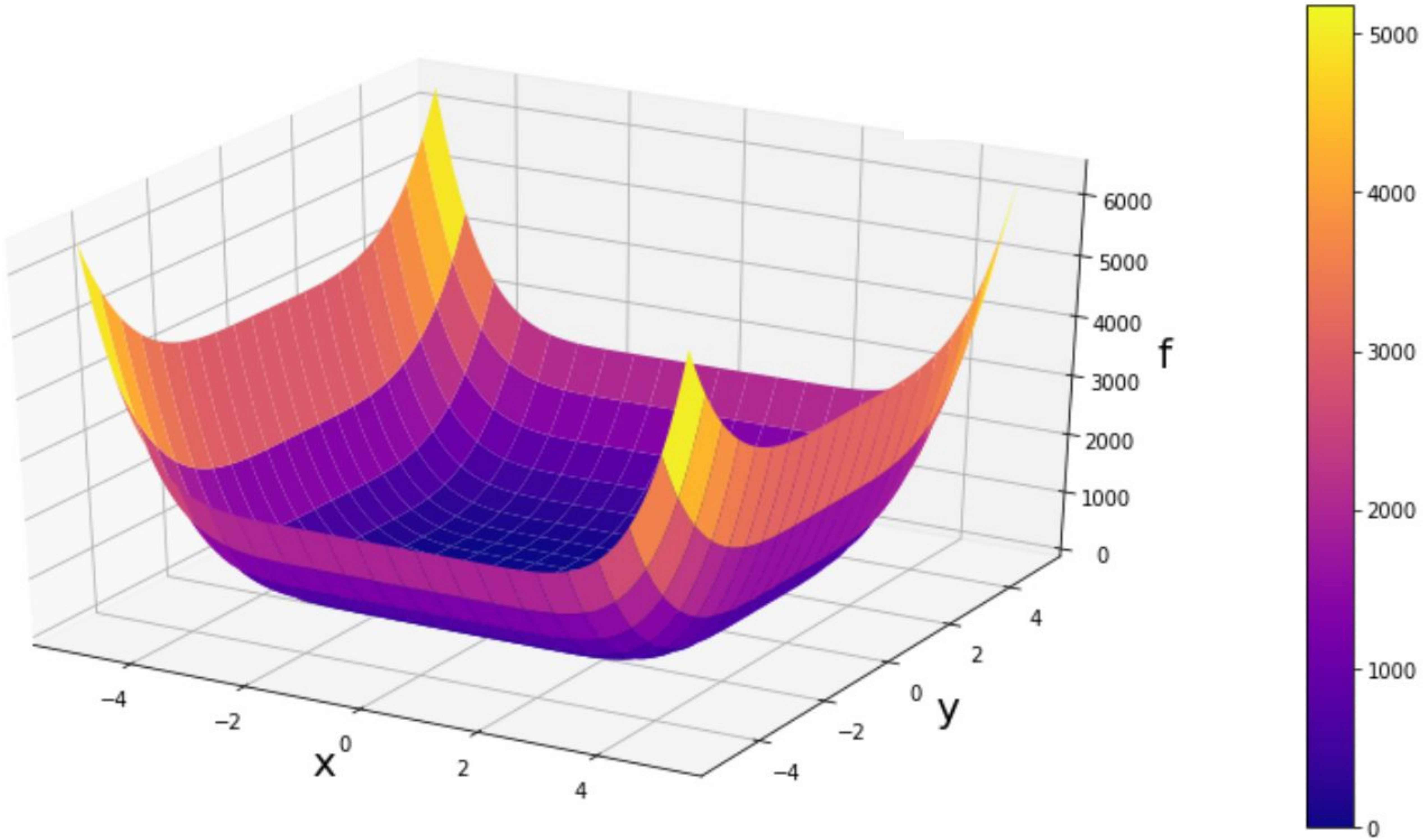
Demo BO loop



Secondmind

BO Demo 2

Let minimize the 6 Hump Camel function

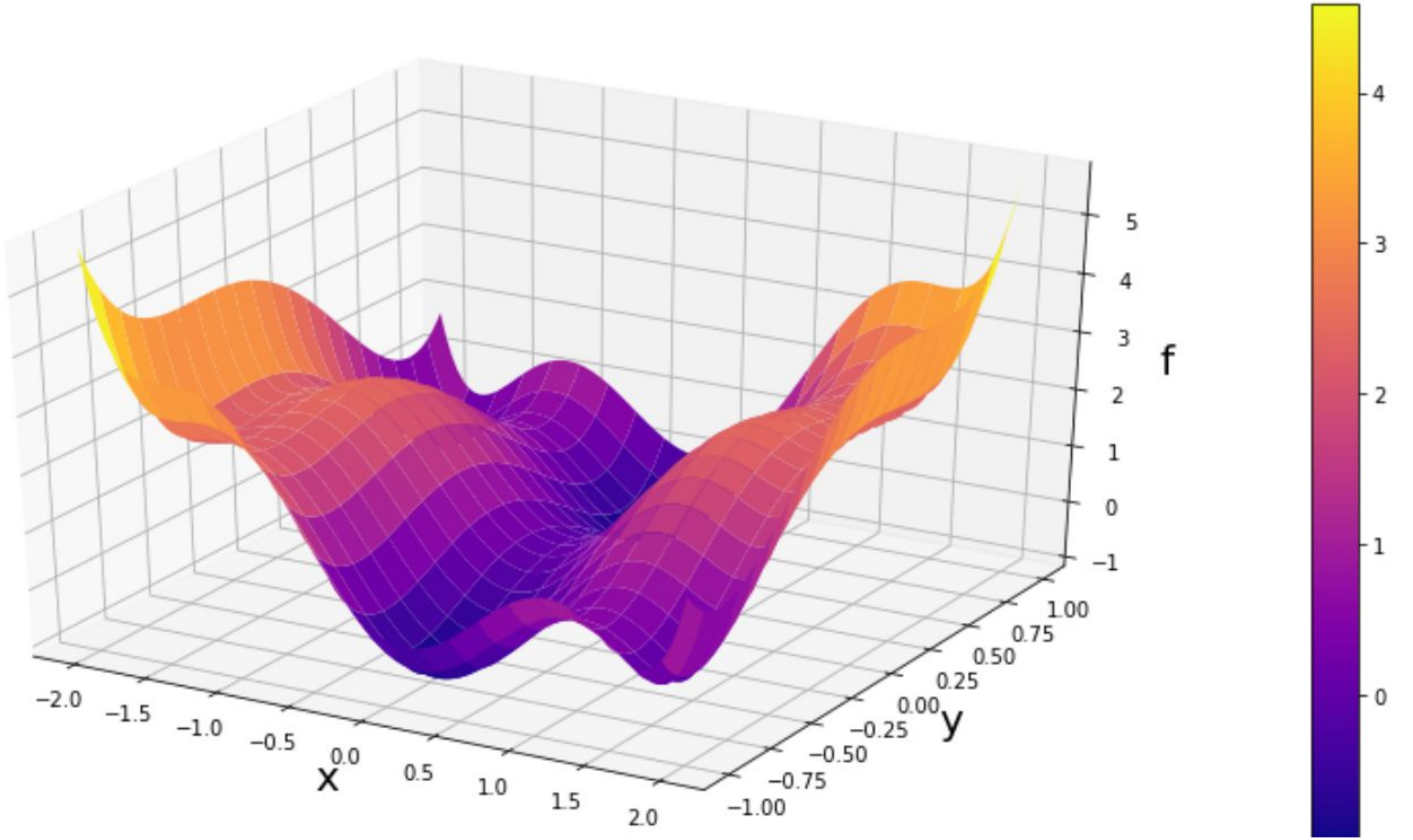


Looks like we **can** use a local optimizer!

Secondmind

BO Demo 2

Zoom in: Perhaps not quite as easy?

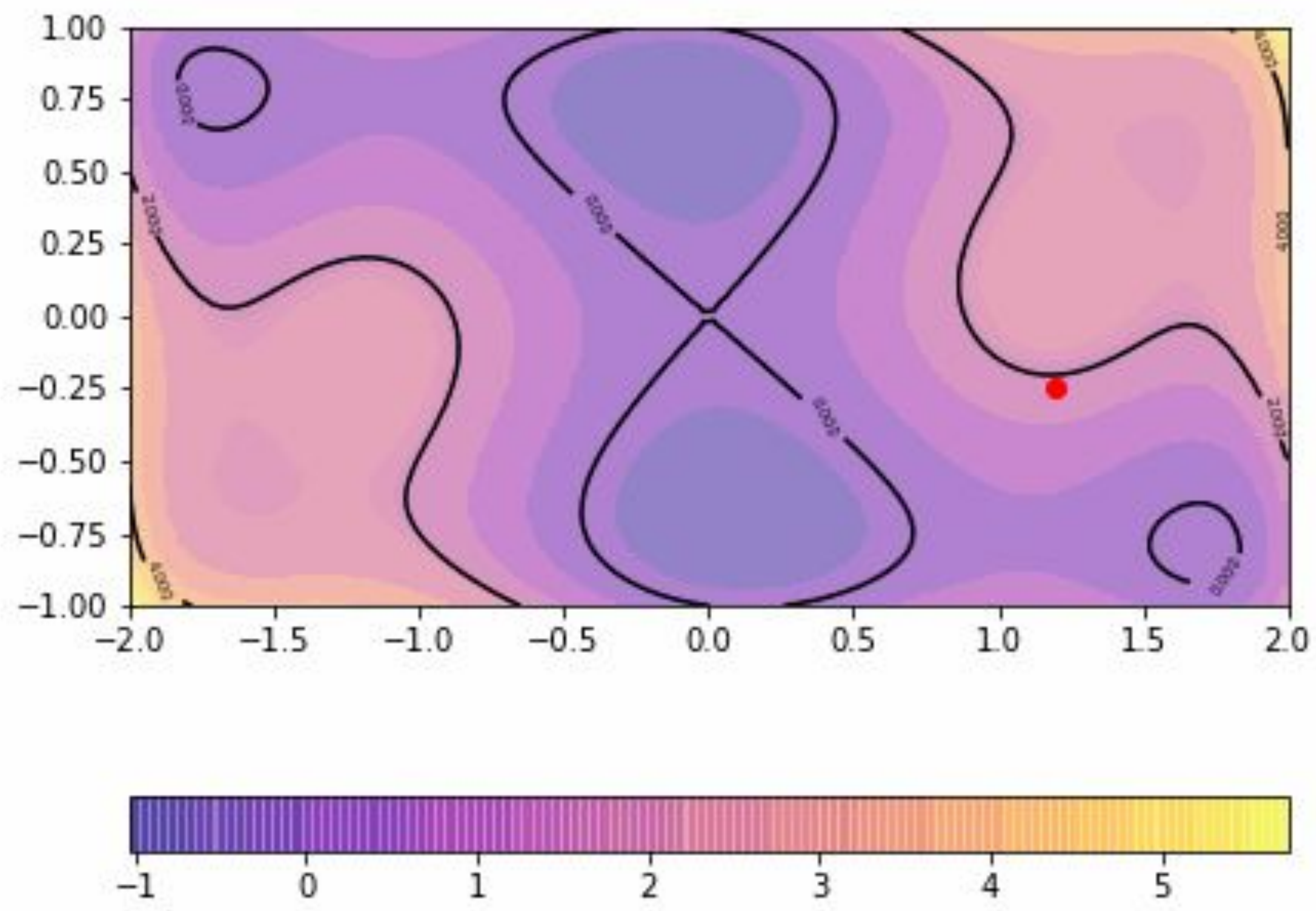


Looks like we **cannot** use a local optimizer!

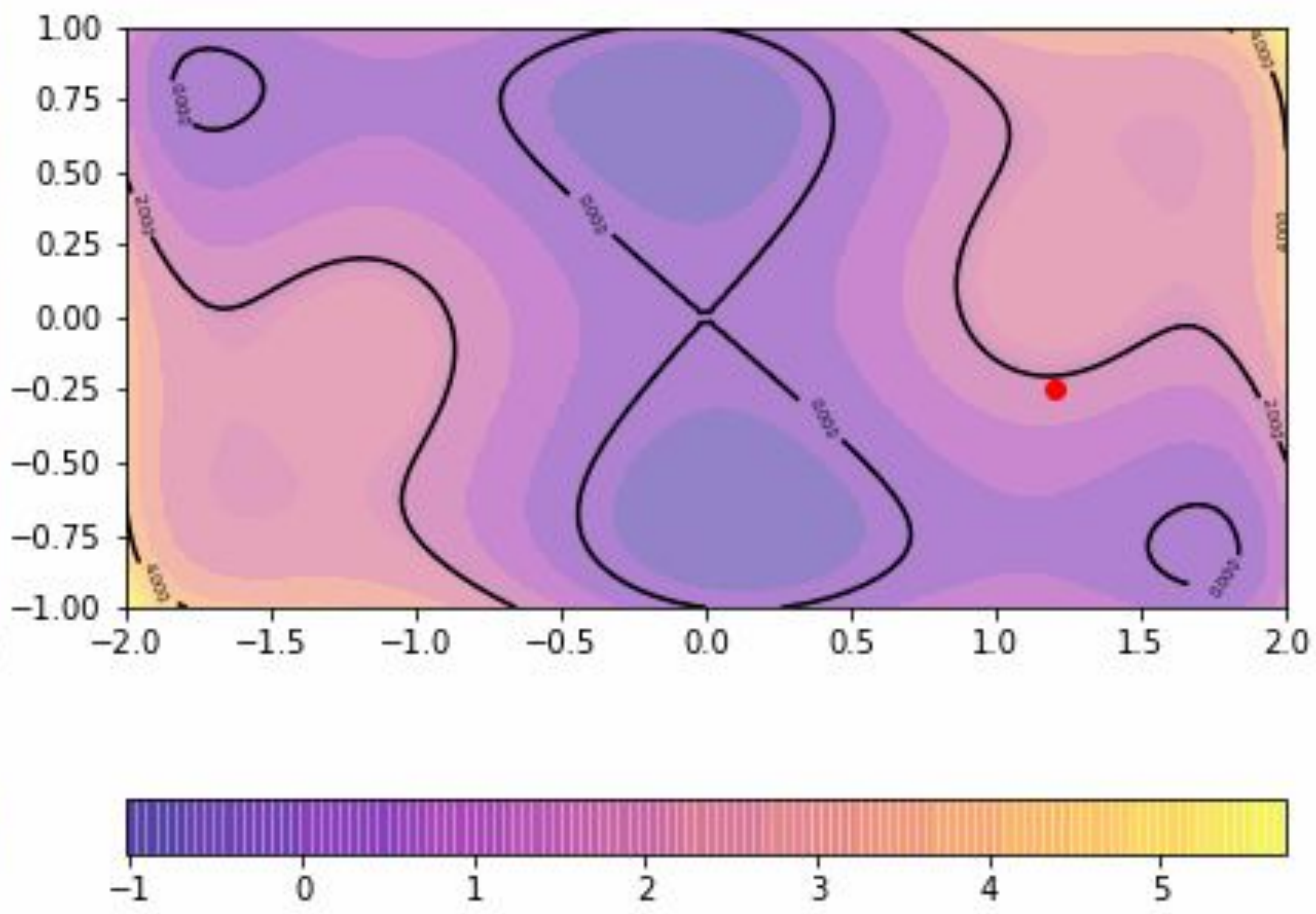
BO Demo 2

Bayesian optimization is a global optimizer

Bayesian optimization (global)



Gradient descent (local)



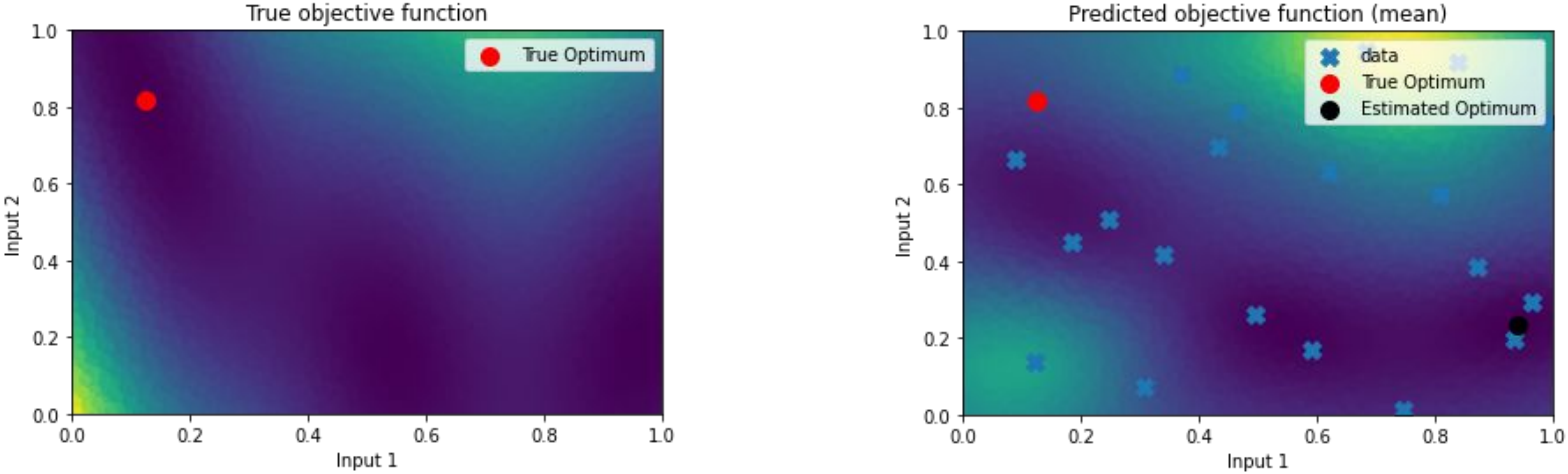
First Steps in Automatic Motor Calibration

Hacking a few things together to get a working algorithm

Standard Bayesian Optimization

Finding **global** minimum of a function

- We care about the **Estimated optimum**: the minimizer of our surrogate model's mean function
- Require model to be accurate around the all local optima (not the case below!)

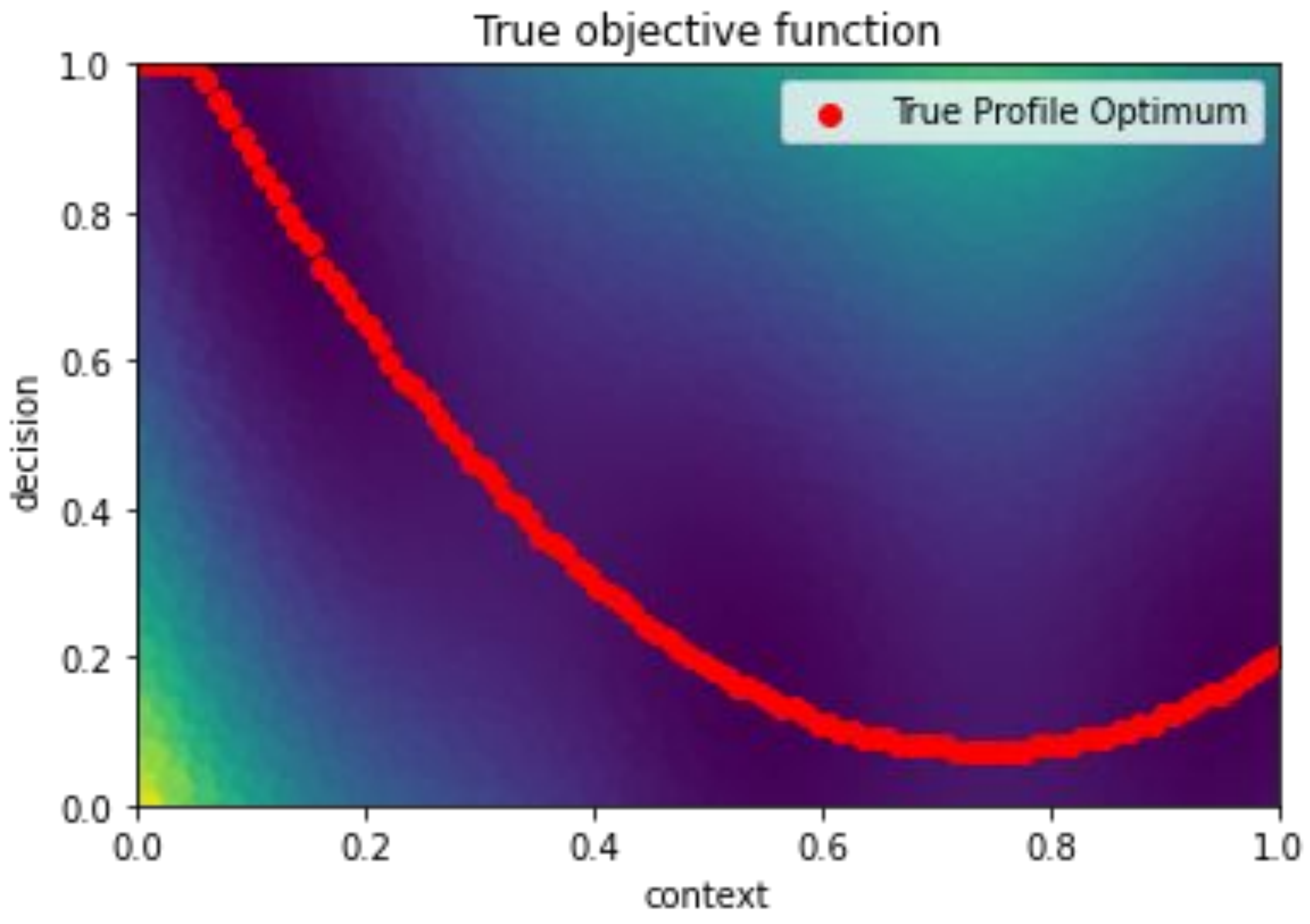


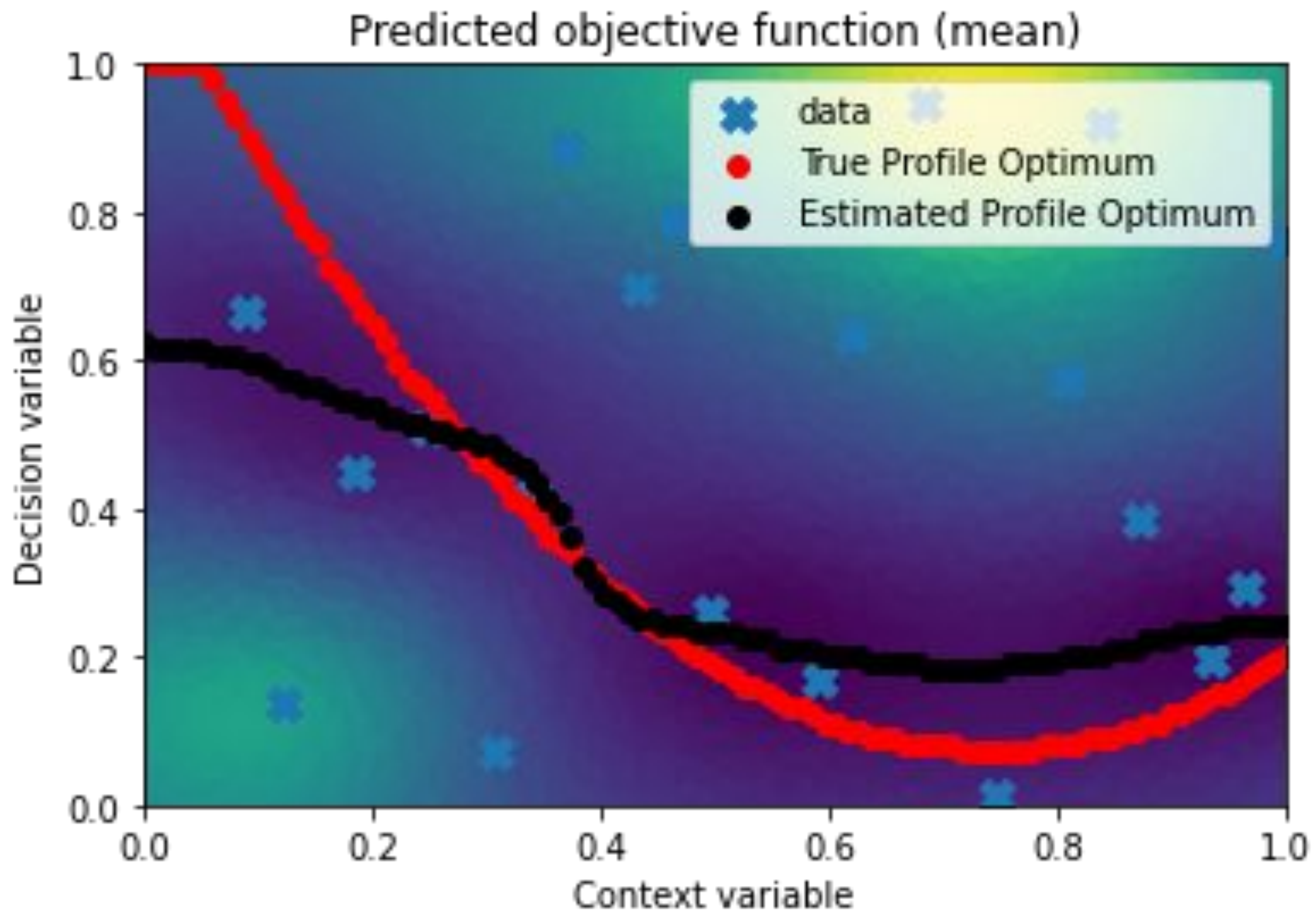
Secondmind

1) Profile Bayesian Optimization for learning lookup tables

Finding optimal **decision** for all **contexts**

- We care about the **Profile Optimum**: the set of minimizers for each possible context
- i.e. a trajectory with elements accessed by optimizing a slice of the search space
- Require our model to be accurate across **more** than just the local minima





1) Profile Bayesian Optimization

Finding optimal **decision** for all **contexts**

- We care about the **Profile Optimum**: the set of minimizers for each possible context
- i.e. a trajectory with elements accessed by optimizing a slice of the search space
- Require our model to be accurate across **more** than just the local minima

Decisions

The *strength* of the magnet: *Ia* - Current

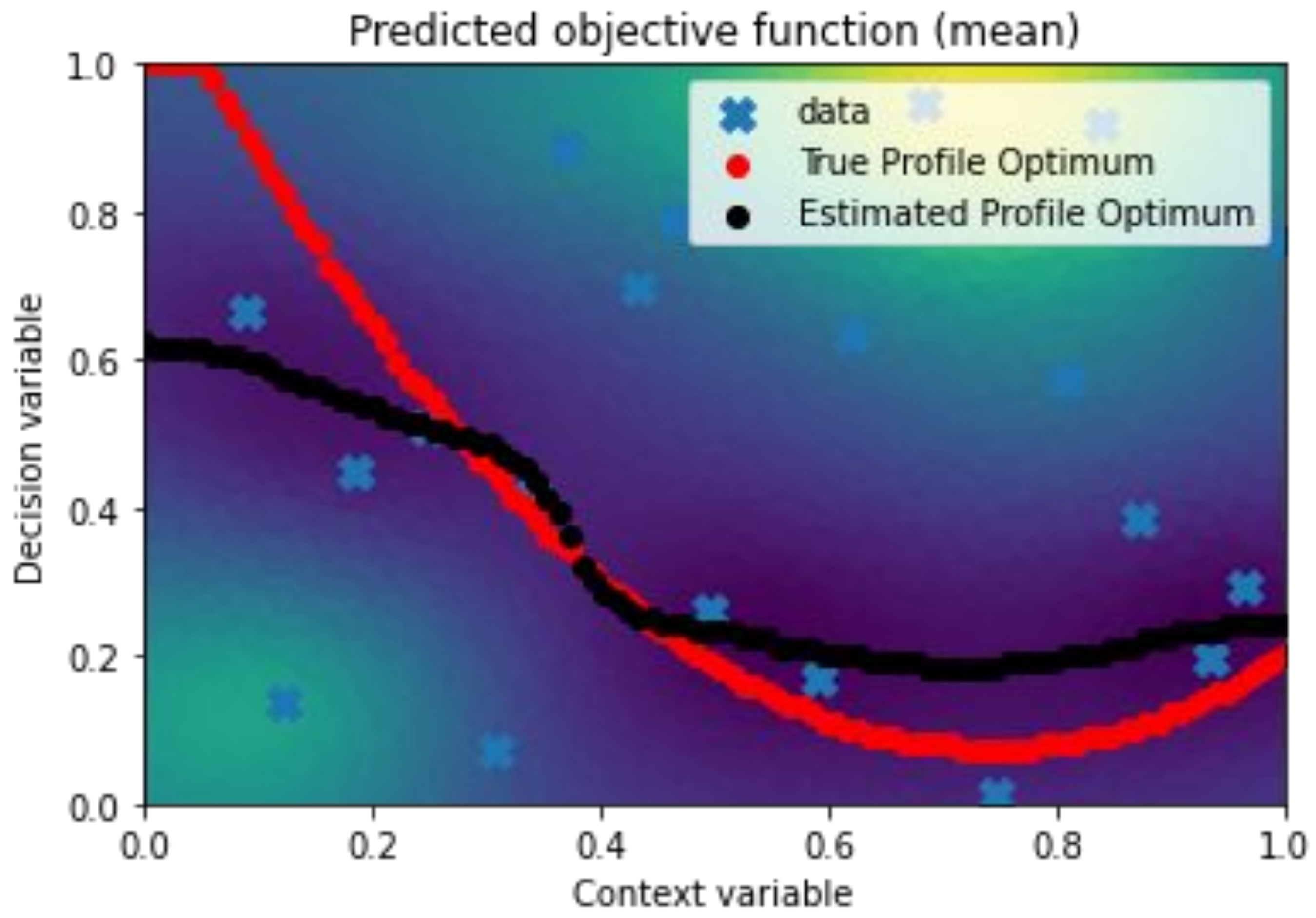
The *location* of the magnet: *β* - Phase angle of the current

Contexts

The rotation speed of the motor

Voltage supplied by the battery

Temperature

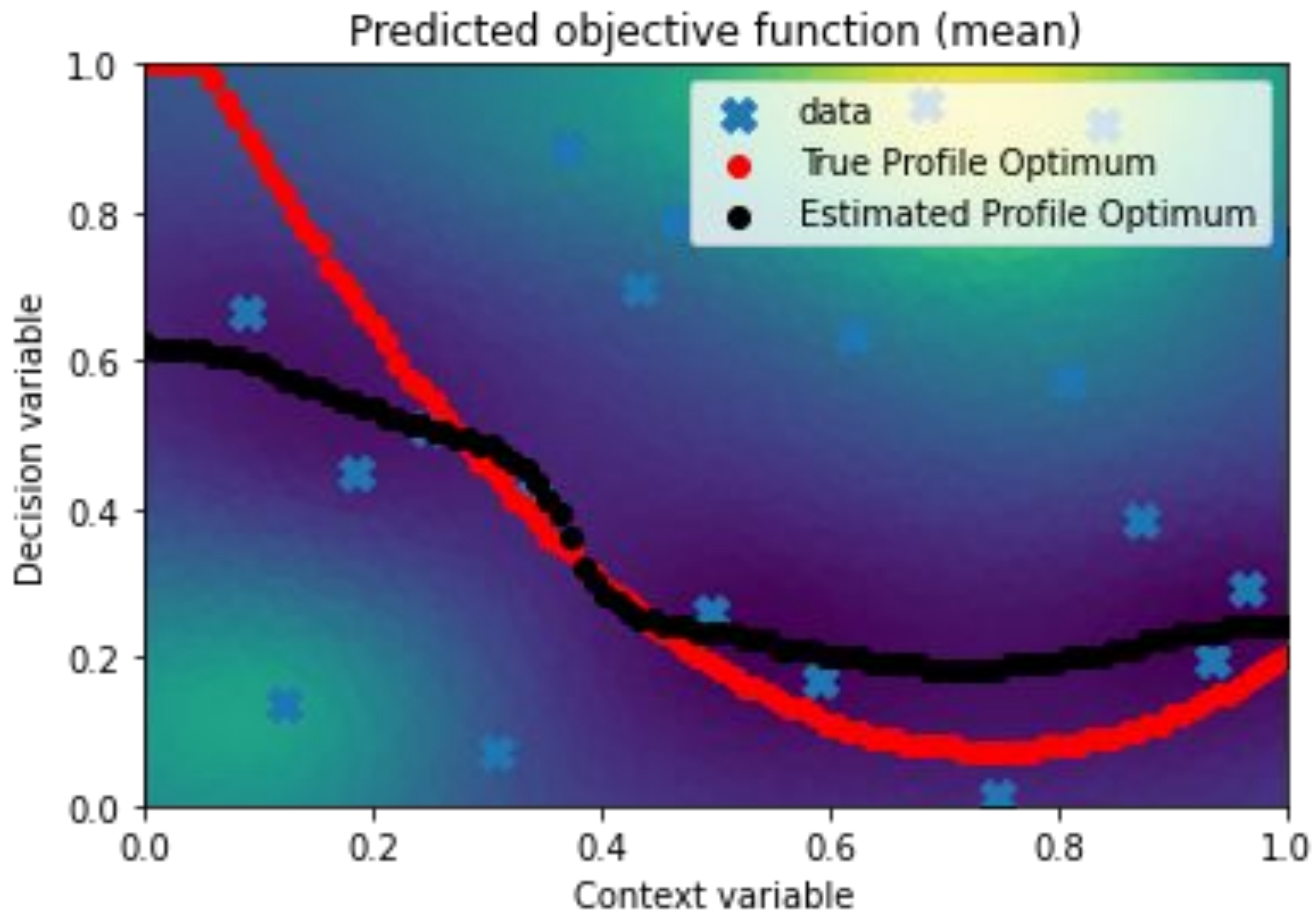


1) Profile Bayesian Optimization

Finding optimal **decision** for all **contexts**

- We care about the **Profile Optimum**: the set of minimizers for each possible context
- i.e. a trajectory with elements accessed by optimizing a slice of the search space
- Require our model to be accurate across **more** than just the local minima

- We can use the acquisition function of Ginsbourger et al. 2013
- But this does not support constraints



2) Scalable surrogate models

A surrogate model suitable for large data

- Standard GP incurs $O(N^3)$

2) Scalable surrogate models

A surrogate model suitable for large data

- Standard GP incurs $O(N^3)$
- For us N>>1,000,000

2) Scalable surrogate models

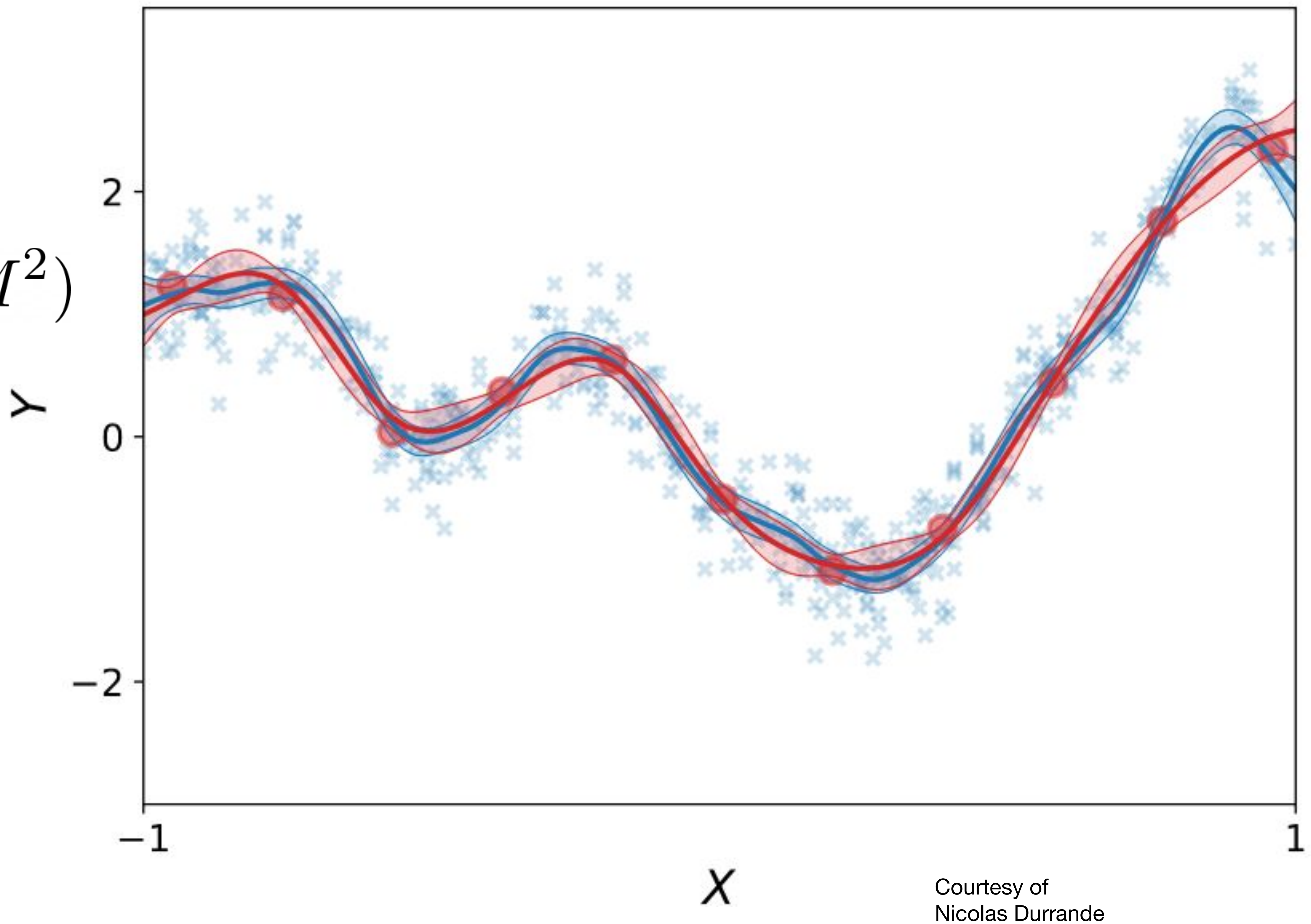
A surrogate model suitable for large data

- Standard GP incurs $O(N^3)$
- For us N>>1,000,000
- Use SVGP (Hensman et al. 2013)

2) Scalable surrogate models

A surrogate model suitable for large data

- Standard GP incurs $O(N^3)$
- For us N>>1,000,000
- Use SVGP (Hensman et al. 2013)
- Replace with M representative points $O(NM^2)$



Courtesy of
Nicolas Durrande

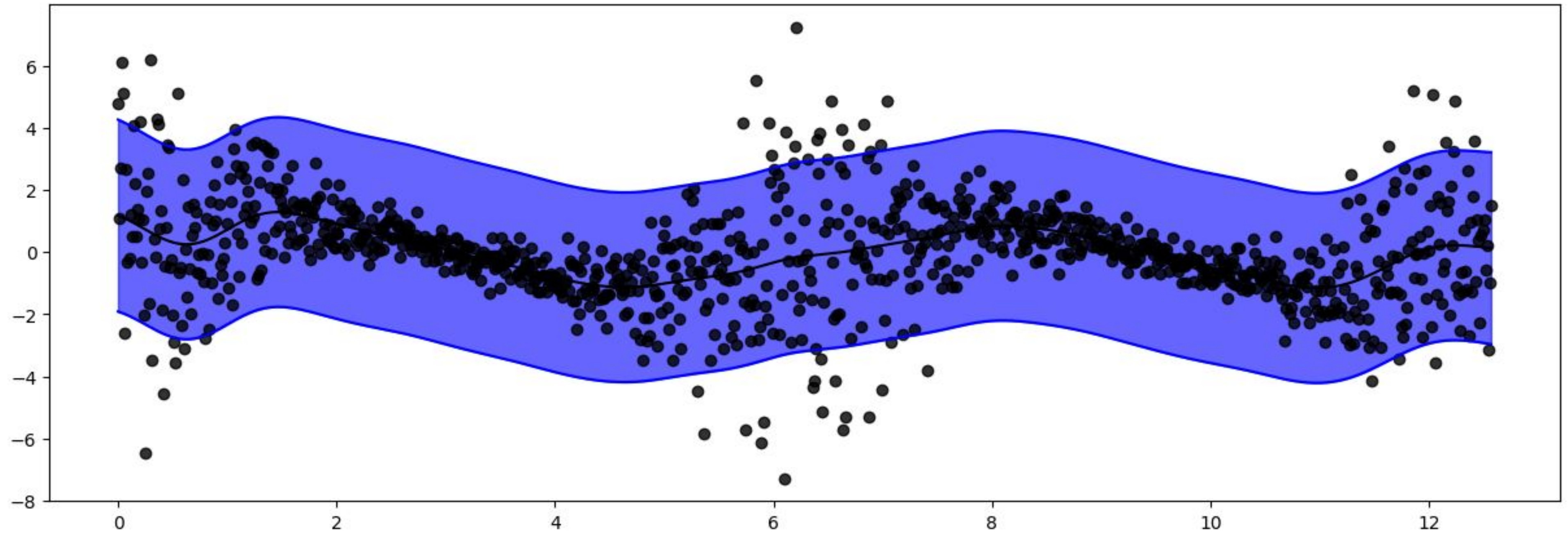
2) Scalable surrogate models

A **heteroscedastic** surrogate model suitable for large data

2) Scalable surrogate models

A **heteroscedastic** surrogate model suitable for large data

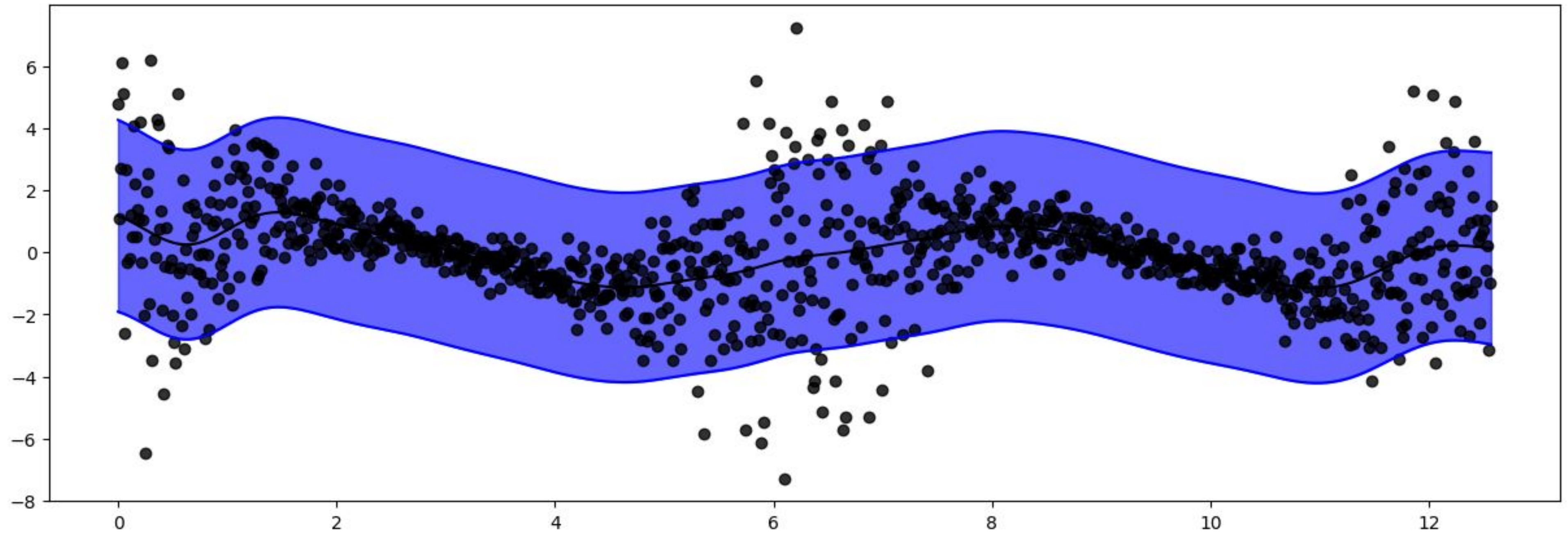
- Standard GP (SVGP) assume fixed noise levels $y_i \sim \mathcal{N}(f(\mathbf{x}_i), \sigma^2)$



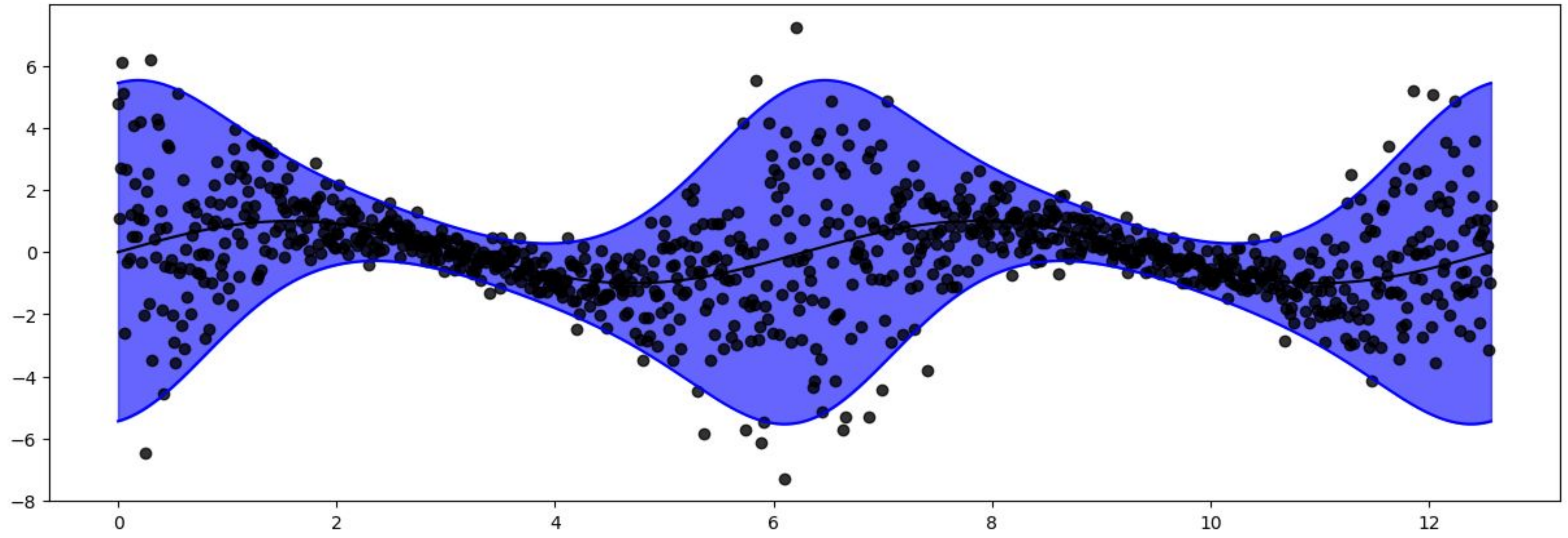
2) Scalable surrogate models

A **heteroscedastic** surrogate model suitable for large data

- Standard GP (SVGP) assume fixed noise levels $y_i \sim \mathcal{N}(f(\mathbf{x}_i), \sigma^2)$



- We use the chained GP of Saul et al. (2016) $y_i \sim \mathcal{N}(f(\mathbf{x}_i), e^{g(\mathbf{x}_i)})$



Secondmind

2) Scalable surrogate models

Unfortunately unsuitable for BO : Small tweaks required

- A balancing act to fit this model (two key failure modes)

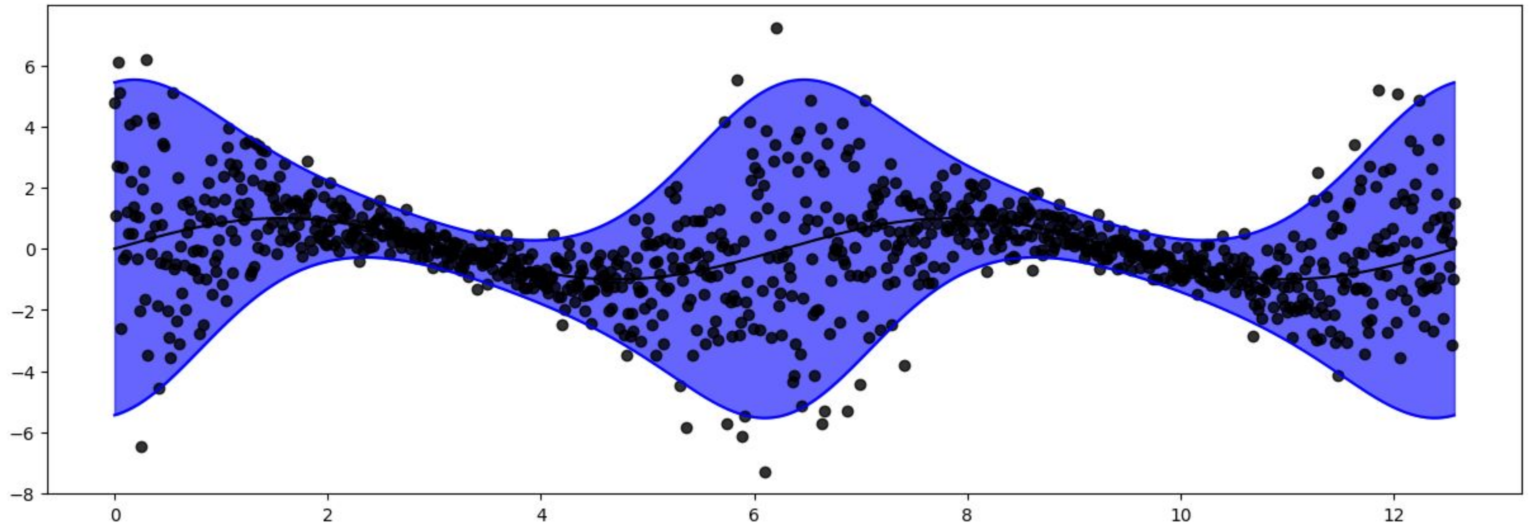
$$y_i \sim \mathcal{N}(f(\mathbf{x}_i), e^{g(\mathbf{x}_i)})$$

2) Scalable surrogate models

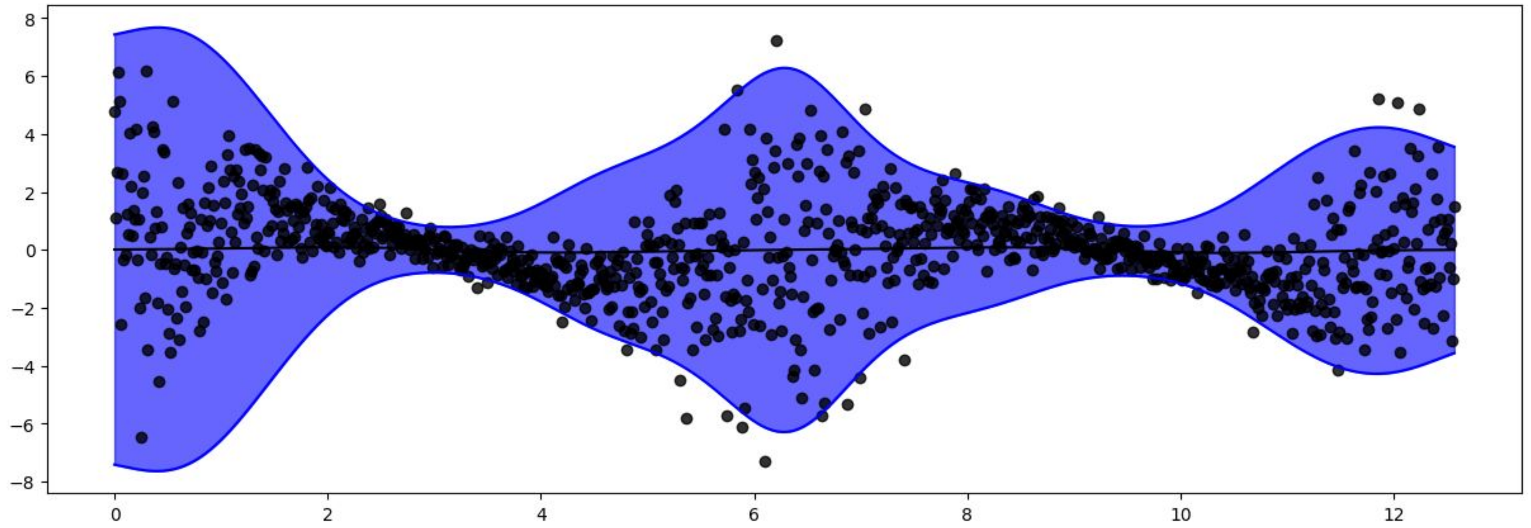
Unfortunately unsuitable for BO : Small tweaks required

- A balancing act to fit this model (two key failure modes)

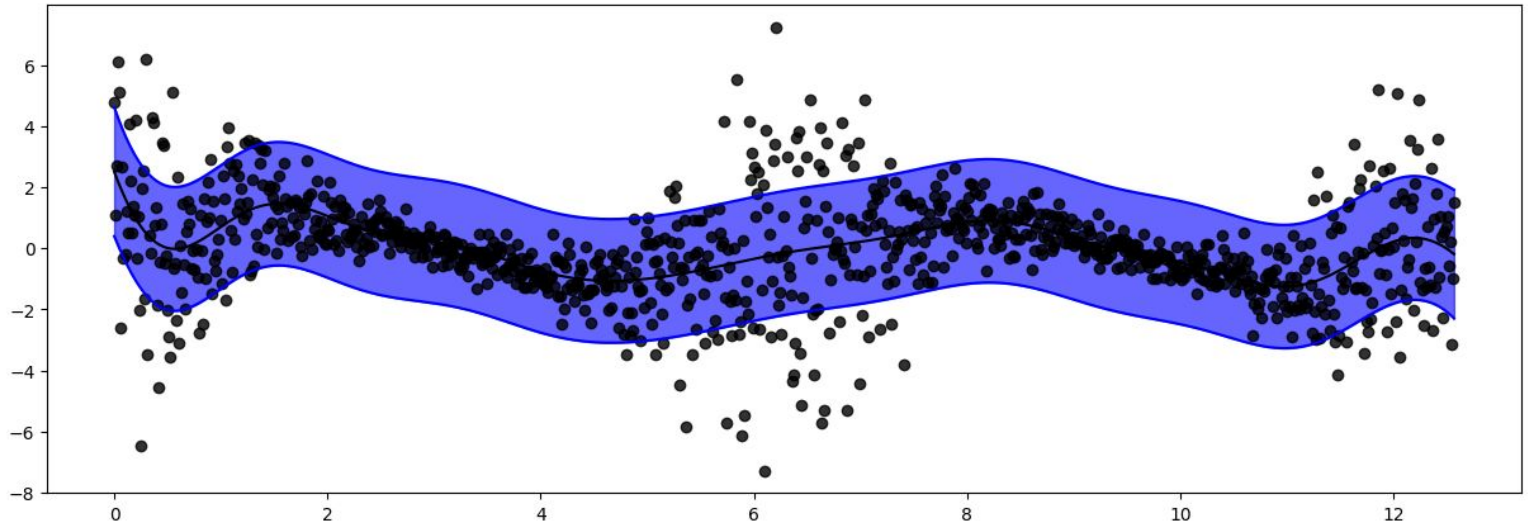
$$y_i \sim \mathcal{N}(f(\mathbf{x}_i), e^{g(\mathbf{x}_i)})$$



Desired fit



g dominates



f dominates

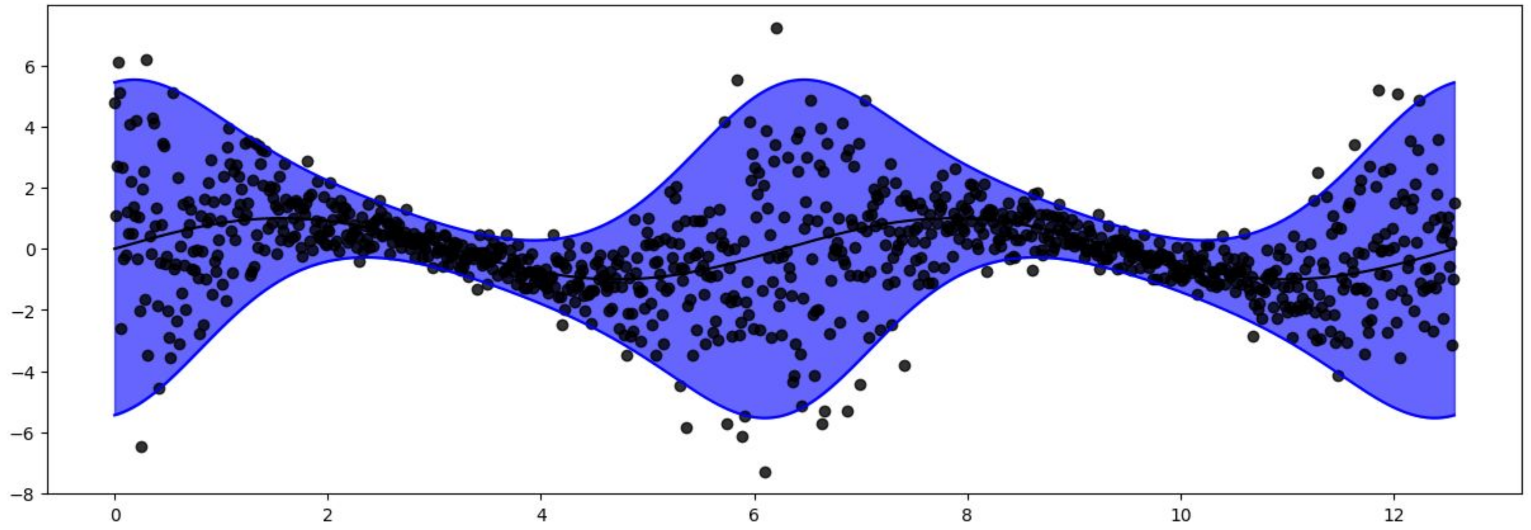
2) Scalable surrogate models

Unfortunately unsuitable for BO : Small tweaks required

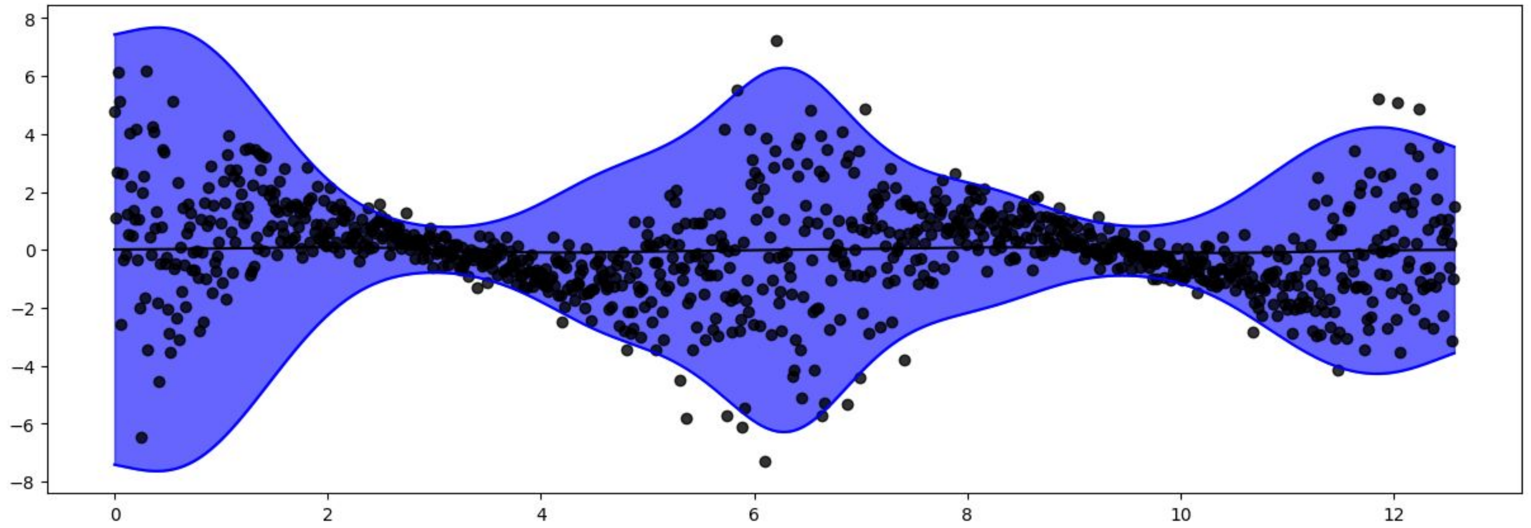
- A balancing act to fit this model (two key failure modes)
- Also relatively expensive to fit (for each BO step)

$$y_i \sim \mathcal{N}\left(f(\mathbf{x}_i), e^{g(\mathbf{x}_i)}\right)$$

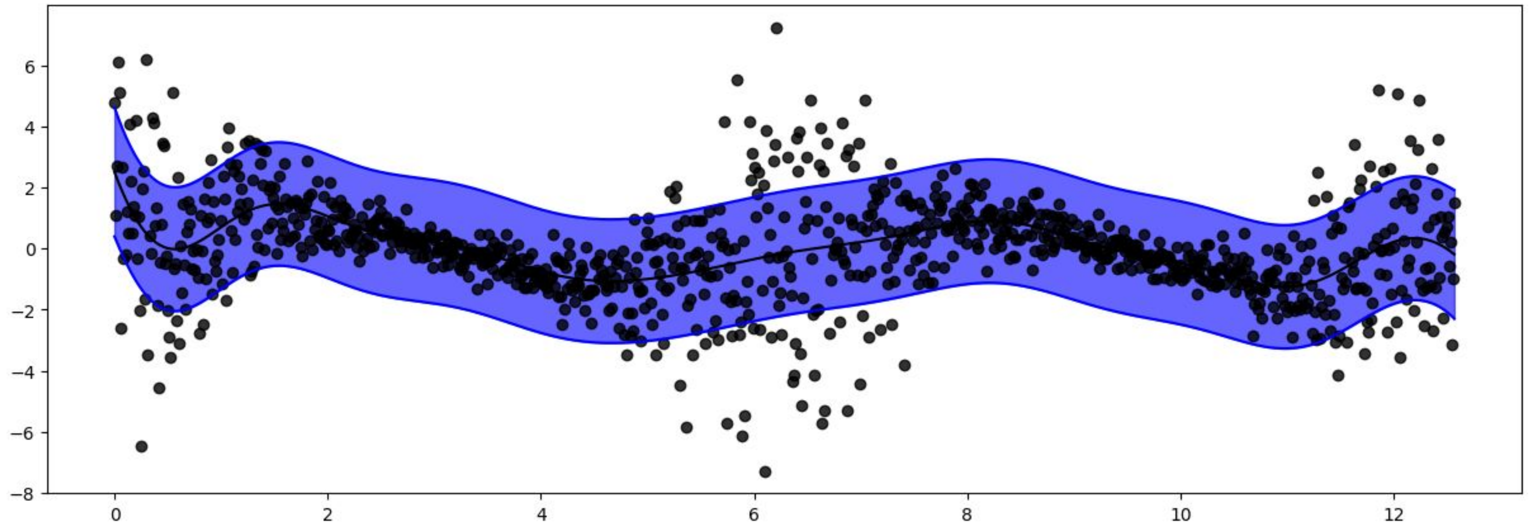
Desired fit



g dominates



f dominates



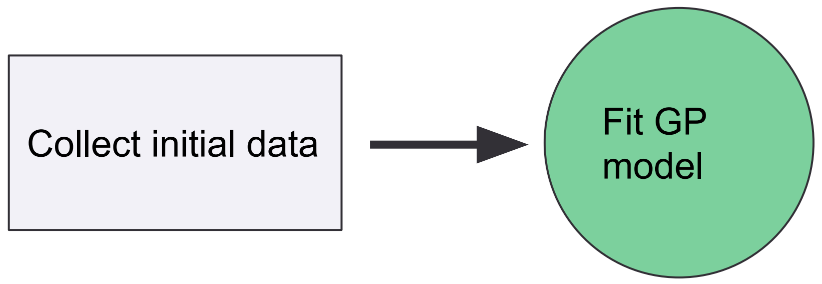
2) Scalable surrogate models

Tweaks required to use HetGP in practice

Collect initial data

2) Scalable surrogate models

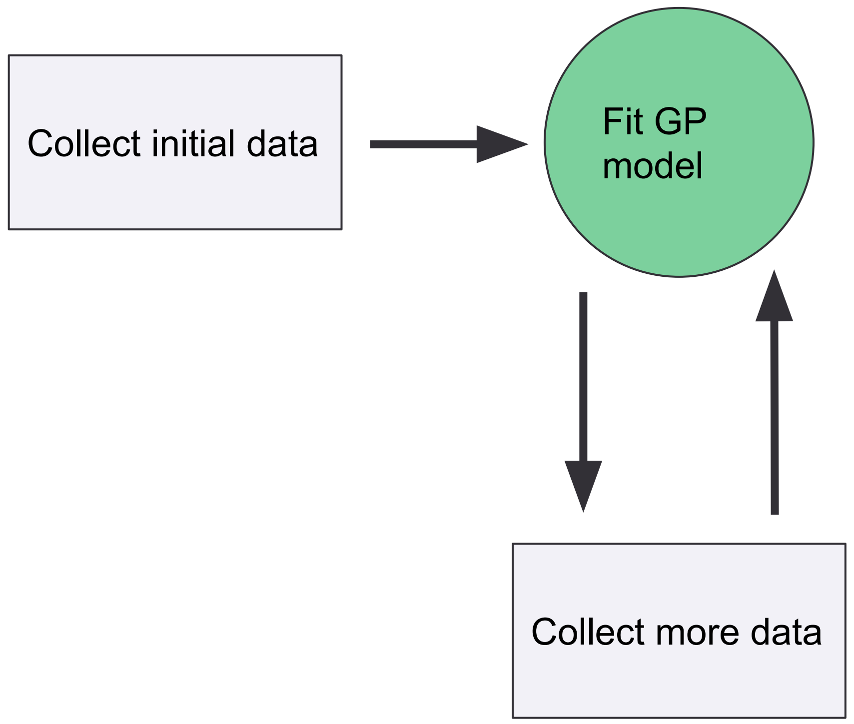
Tweaks required to use HetGP in practice



Secondmind

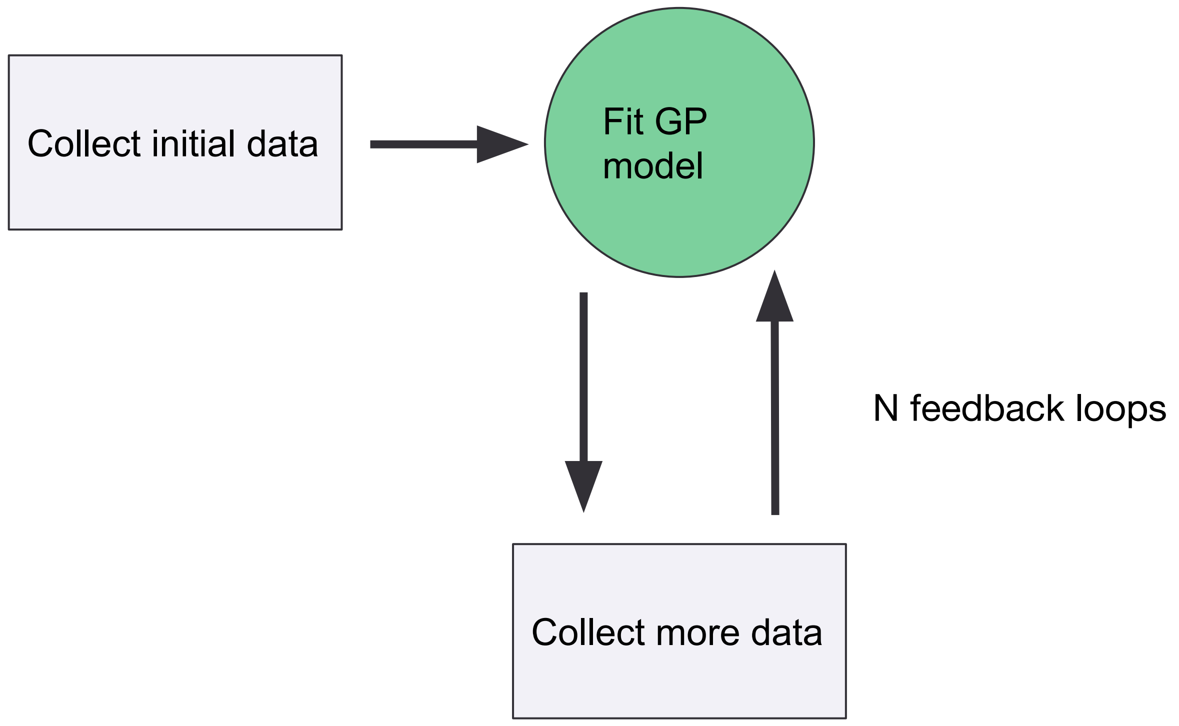
2) Scalable surrogate models

Tweaks required to use HetGP in practice



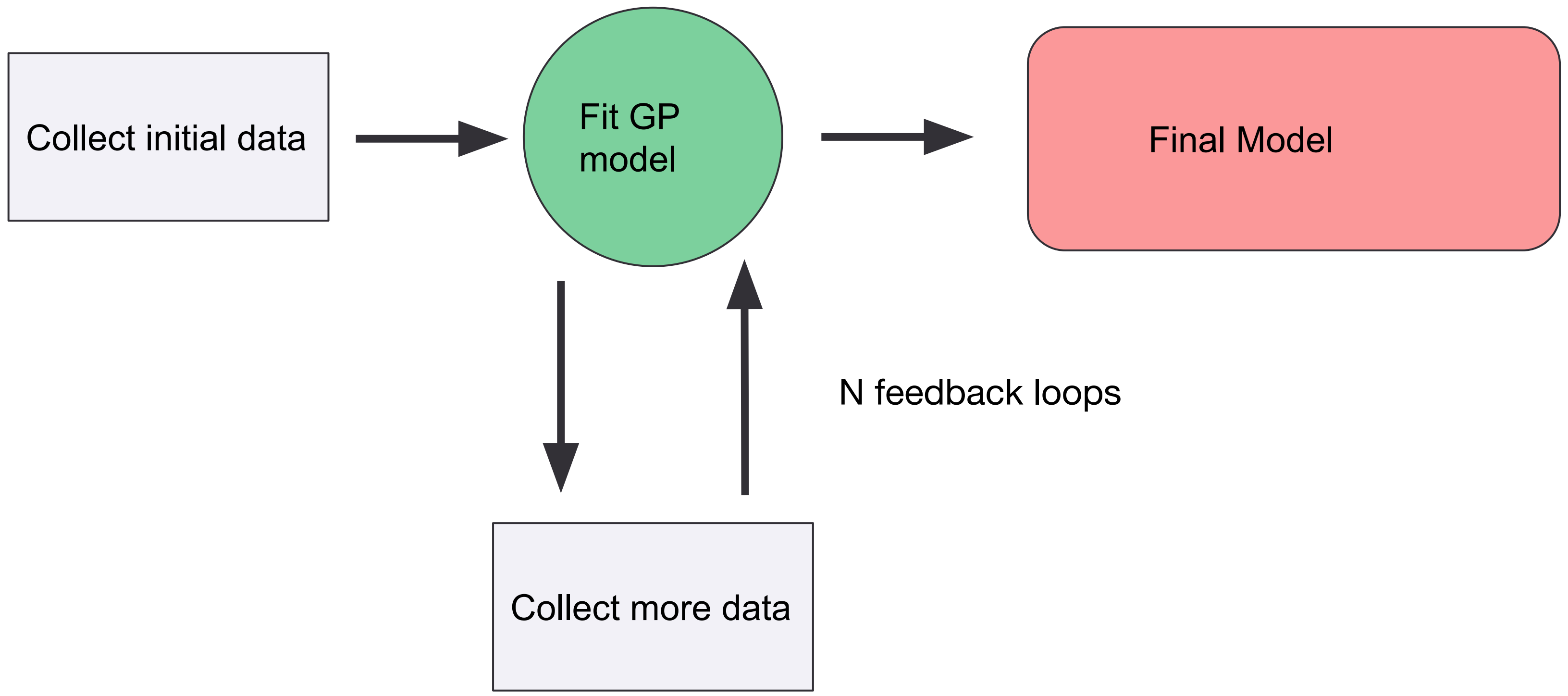
2) Scalable surrogate models

Tweaks required to use HetGP in practice



2) Scalable surrogate models

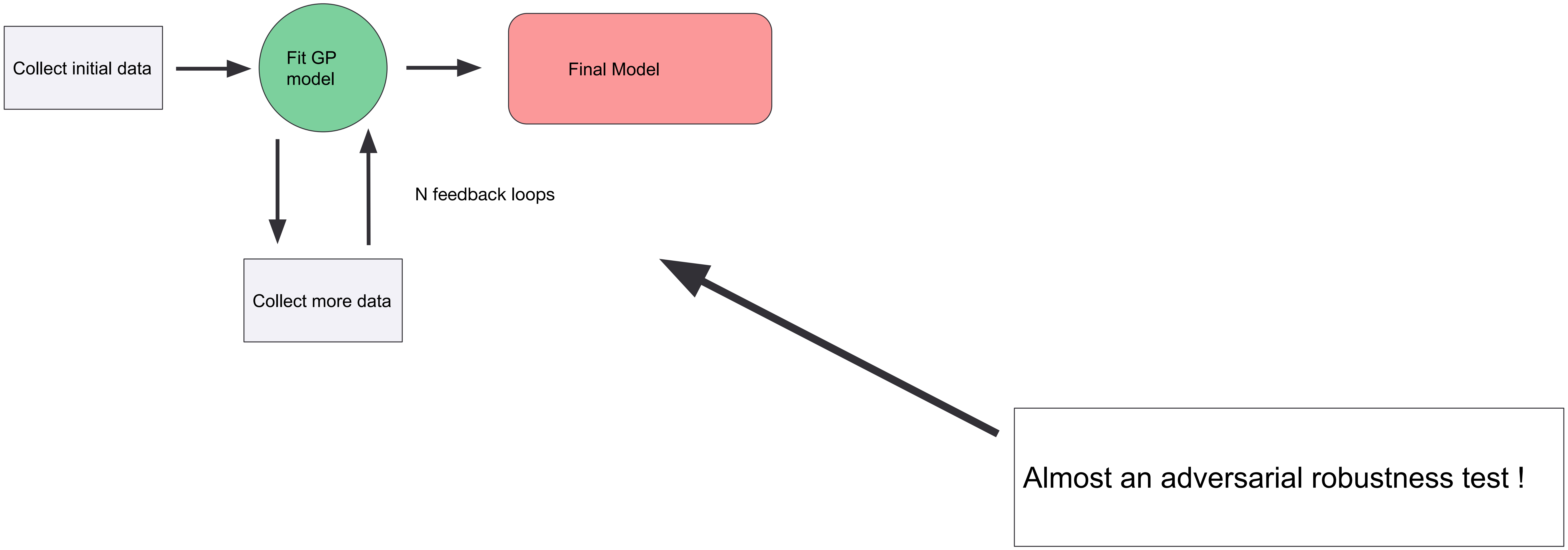
Tweaks required to use HetGP in practice



Secondmind

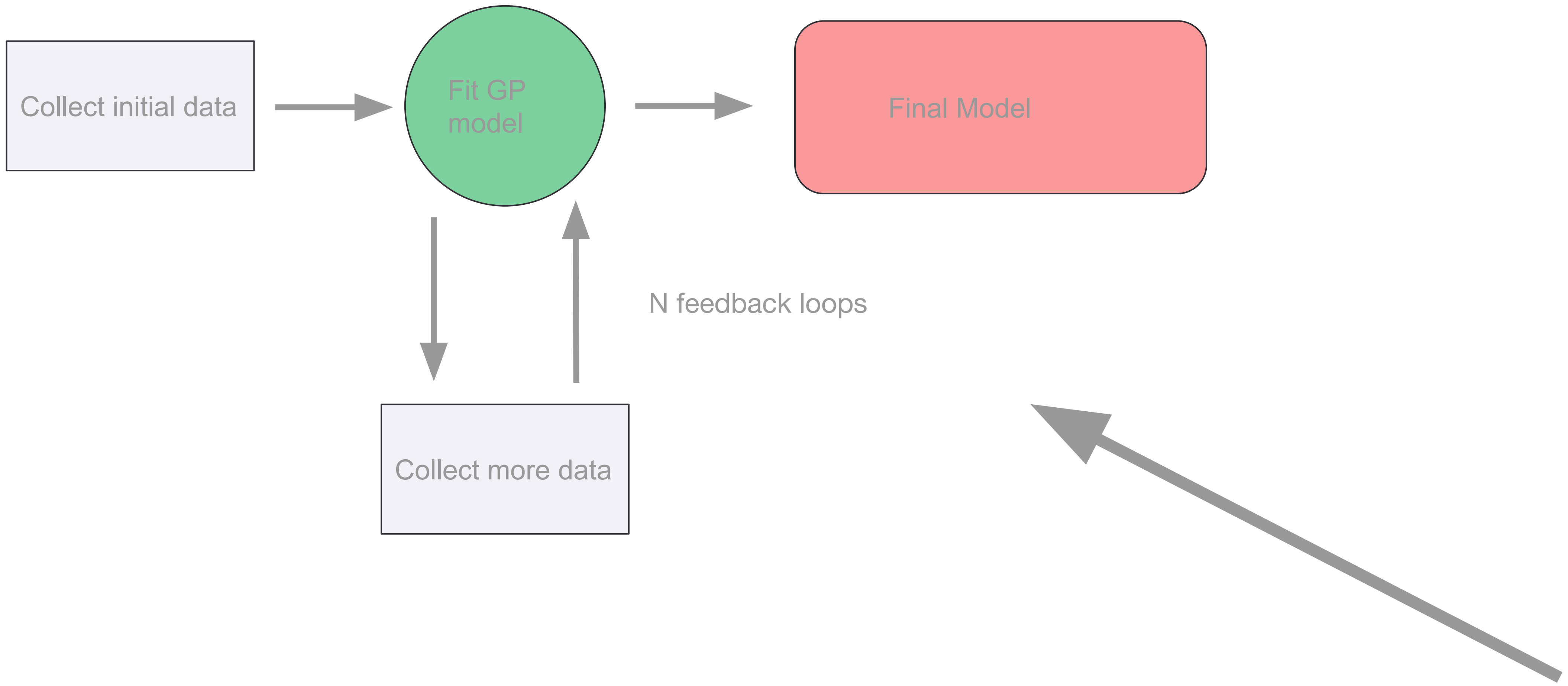
2) Scalable surrogate models

Tweaks required to use HetGP in practice



2) Scalable surrogate models

Tweaks required to use HetGP in practice



Almost an adversarial robustness test !

- Incremental model updates
- Clever initialisation
- Carefully defined optimisers
- Strong priors

A first ML solution

A handful of existing algorithms + Some clever hacks → Proof of concept/value solution

Next Steps

Research time!

Motor Calibration

What else do we need?

- 6-10 inputs
- 2 objectives
- 1-3 constraints
- Need to find a look-up table = “profile optimum”
- Noise is heteroscedastic and overall budget = millions of observations

Bayesian Adaptive Reconstruction of Profile Optima (Ginsbourger et al. 2013)

Chained Gaussian Processes (Saul, Hensman, Vehtari, Lawrence et al 2016.)

- Large/variable cost of preparing the motor for an experiment
- 1 experiment delivers 100-1000 observations at a time
- Risk adversity

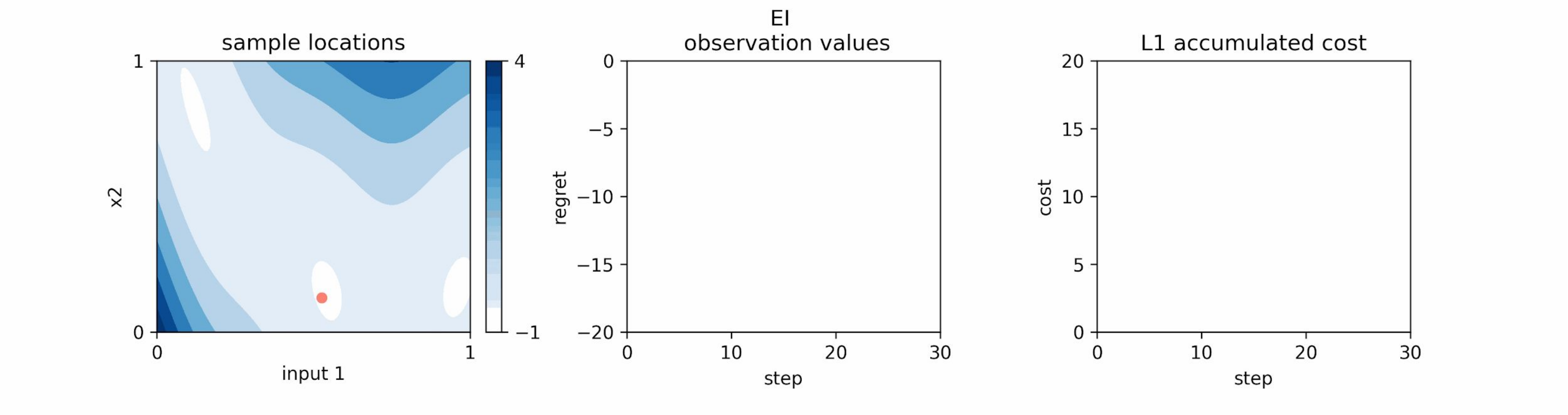
1) Smooth Bayesian Optimisation

Avoid large costs for changing engine settings

1) Smooth Bayesian Optimisation

Avoid large costs for changing engine settings

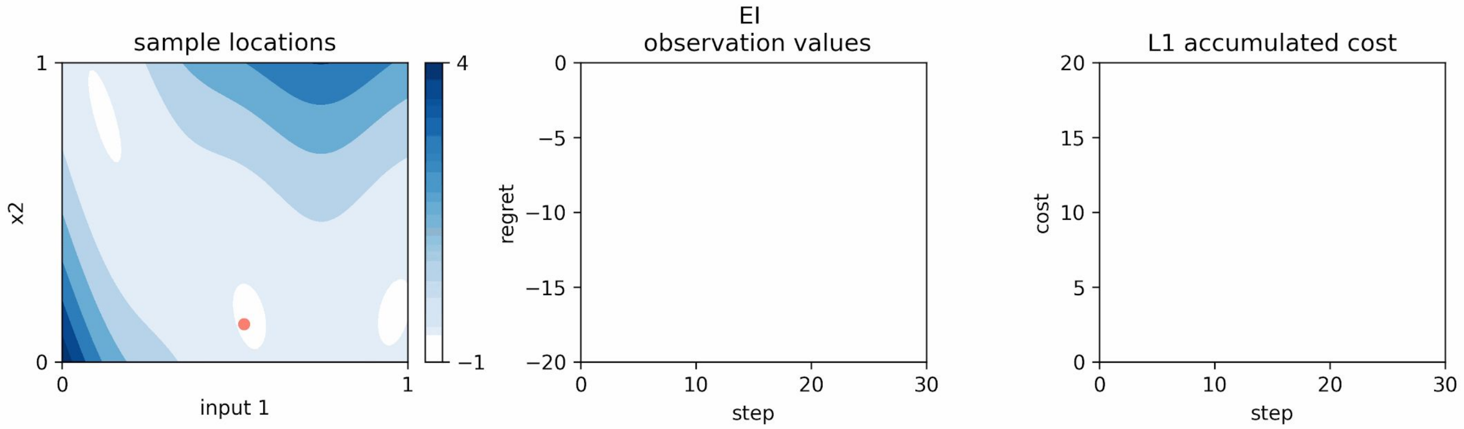
- Need to minimise movement costs but still achieve global optimisation



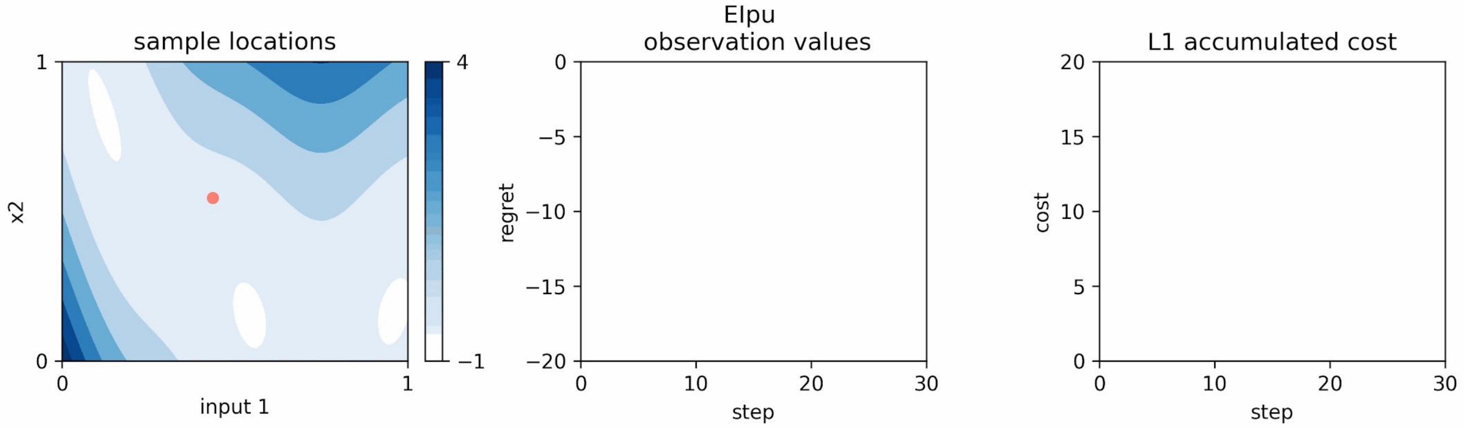
1) Smooth Bayesian Optimisation

Avoid large costs for changing engine settings

- Need to minimise movement costs but still achieve global optimisation



- Constraining maximum movement is not sufficient (needs to be non-myopic)

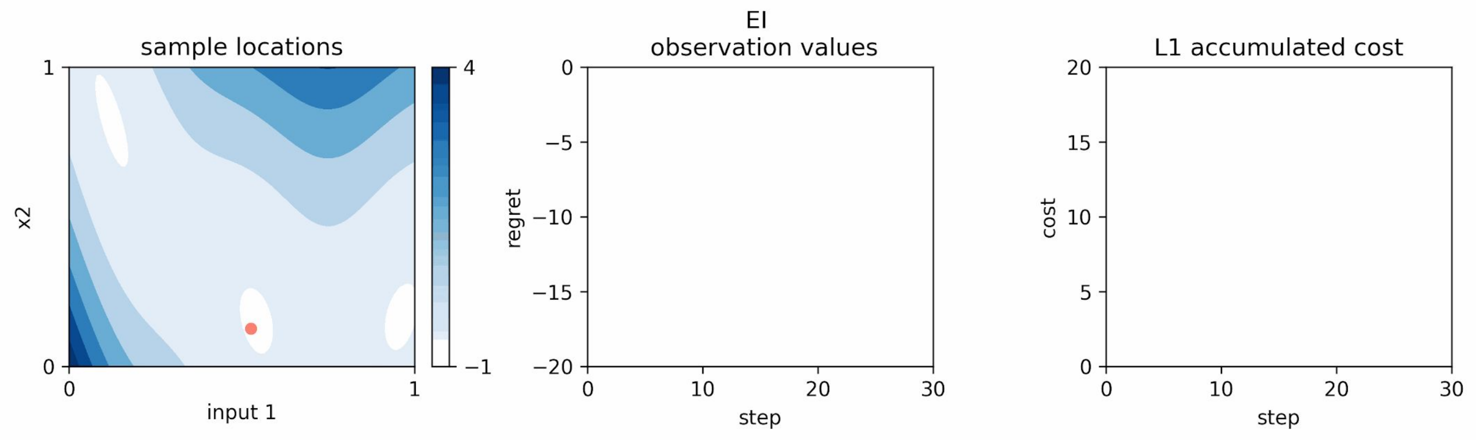


Secondmind

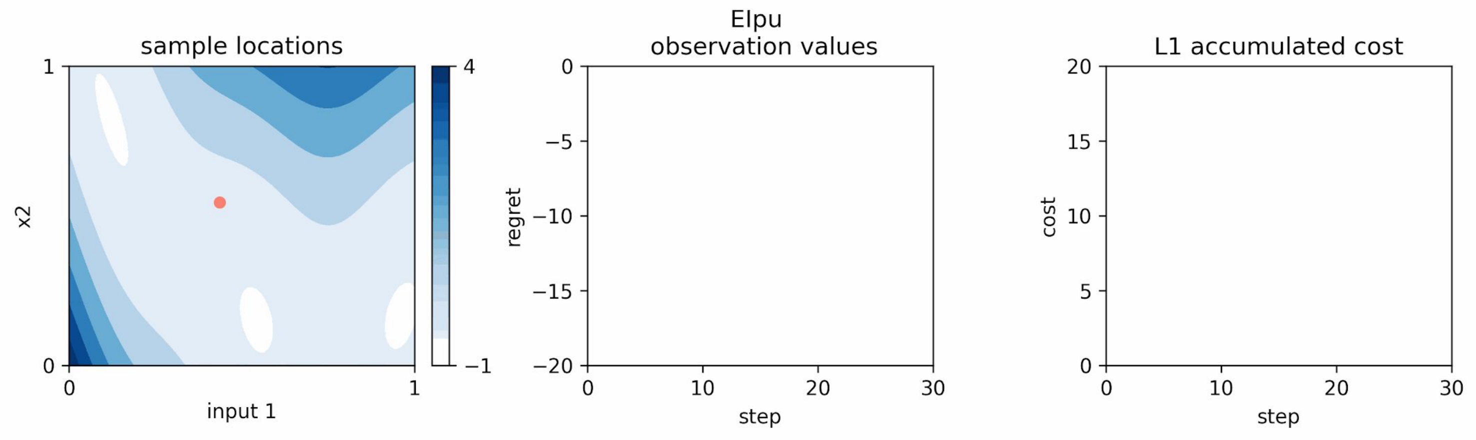
1) Smooth Bayesian Optimisation

Avoid large costs for changing engine settings

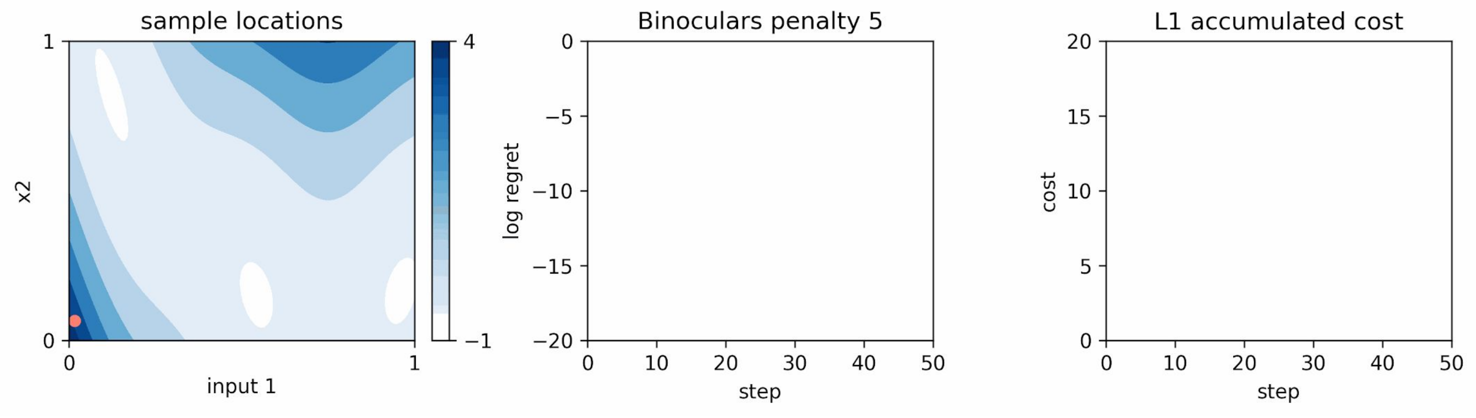
- Need to minimise movement costs but still achieve global optimisation



- Constraining maximum movement is not sufficient (needs to be non-myopic)



- We need a non-myopic strategy



Secondmind

1) Smooth Bayesian Optimisation

Learn a non-myopic cost-aware strategy using an LSTM

1) Smooth Bayesian Optimisation

Learn a non-myopic cost-aware strategy using an LSTM

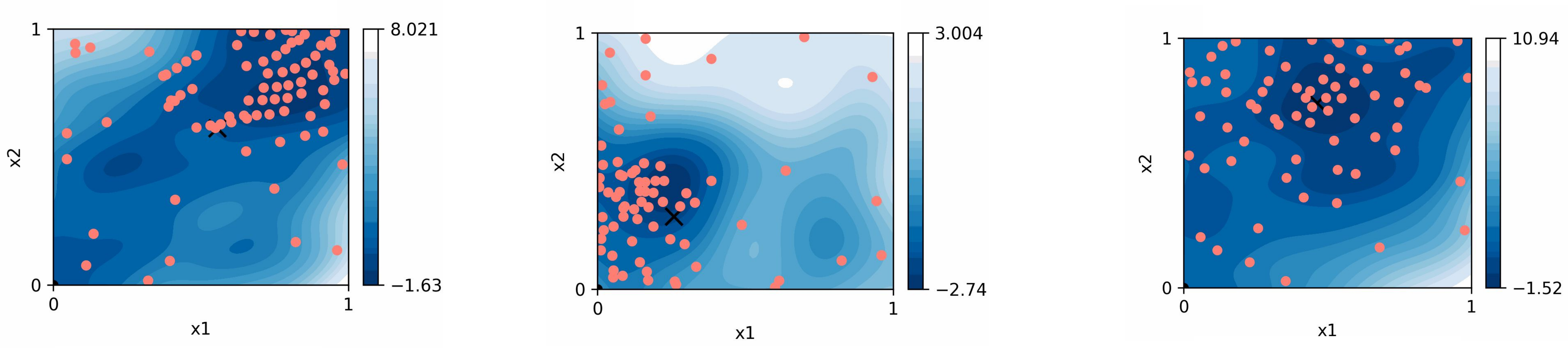
- Trained on GP samples (allow specification of prior knowledge)
- Instant inference provides scalability

1) Smooth Bayesian Optimisation

Learn a non-myopic cost-aware strategy using an LSTM

- Trained on GP samples (allow specification of prior knowledge)
- Instant inference provides scalability

Standard BO

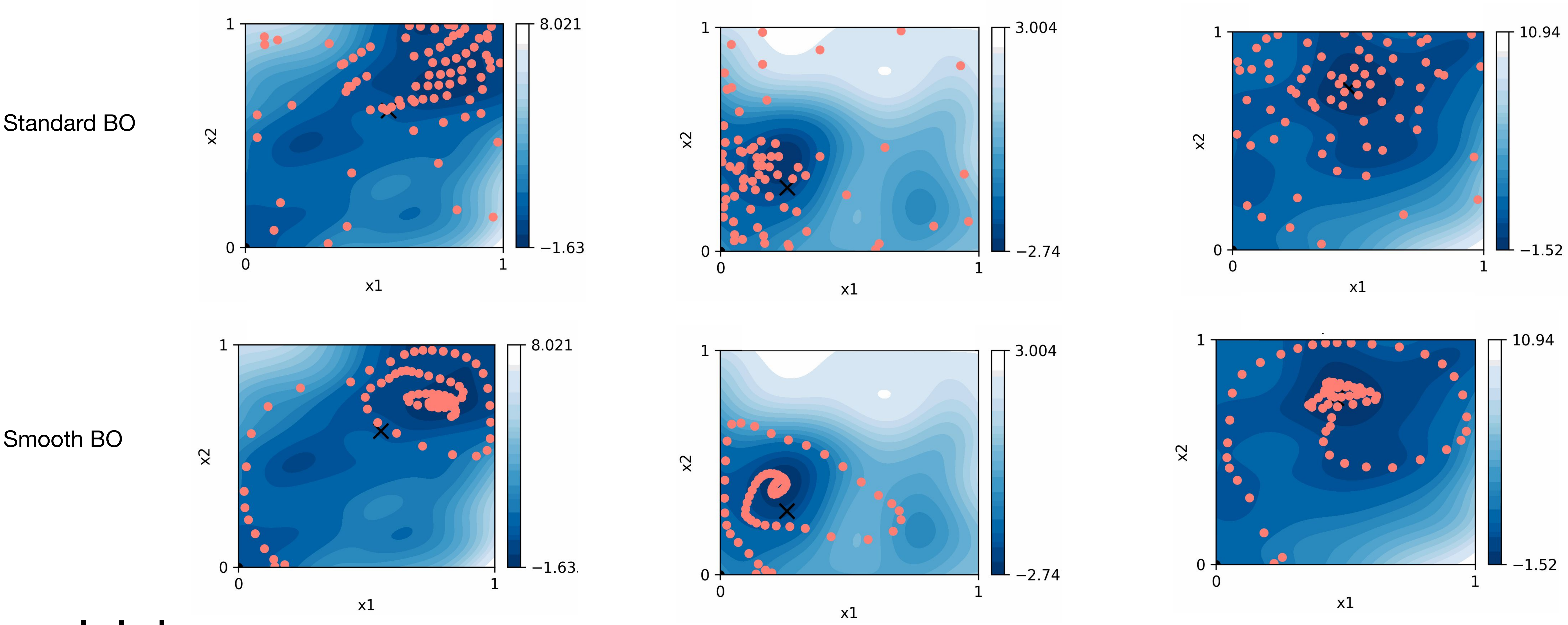


Secondmind

1) Smooth Bayesian Optimisation

Learn a non-myopic cost-aware strategy using an LSTM

- Trained on GP samples (allow specification of prior knowledge)
- Instant inference provides scalability

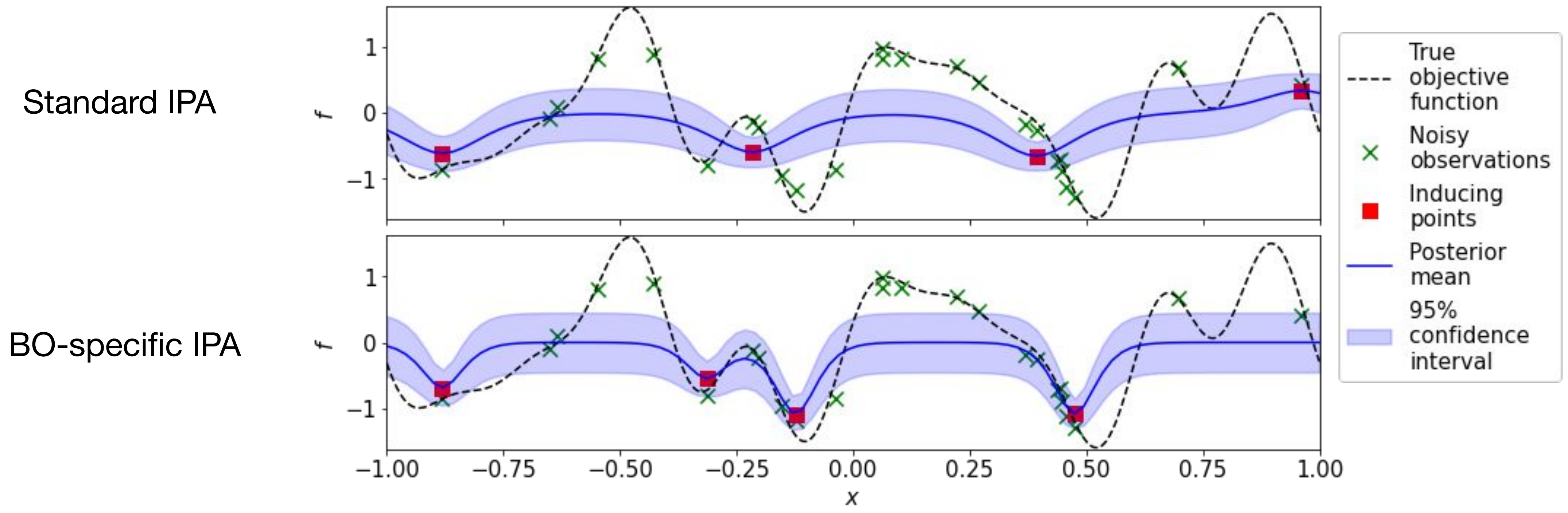


Secondmind

2) Inducing point allocation (IPA) in BO loops

Sparse Gaussian process should be customised to the task at hand

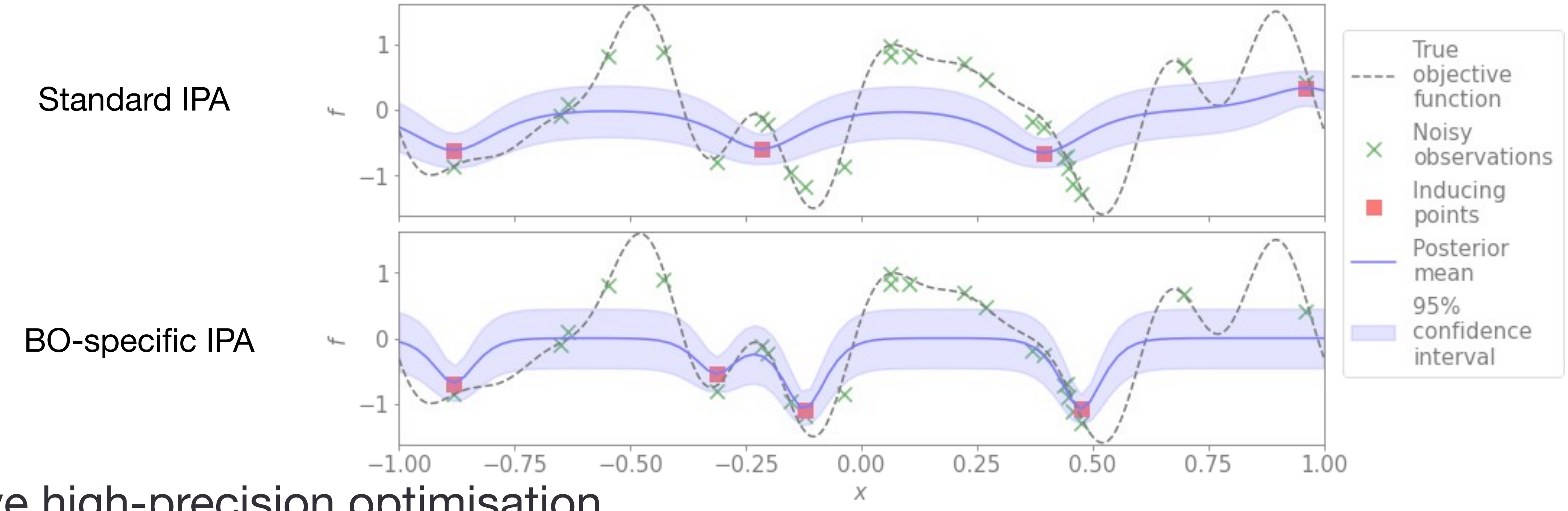
- Standard approaches for inducing point allocation are not suitable for BO



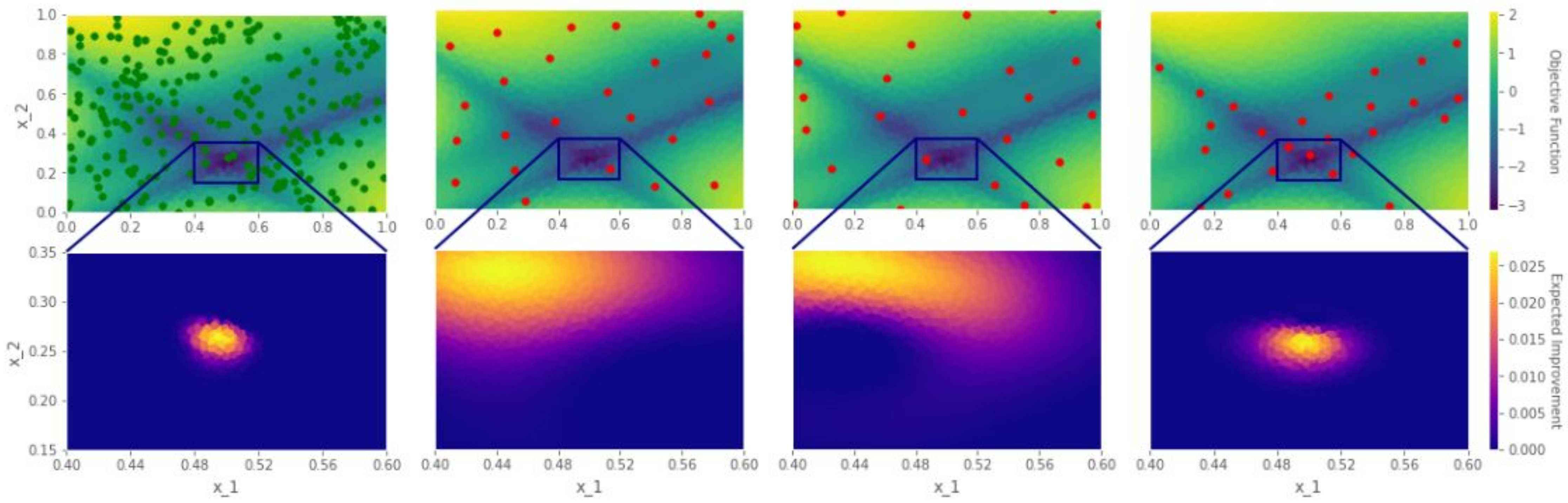
2) Inducing point allocation (IPA) in BO loops

Sparse Gaussian process should be customised to the task at hand

- Standard approaches for inducing point allocation are not suitable for BO



- Cannot achieve high-precision optimisation



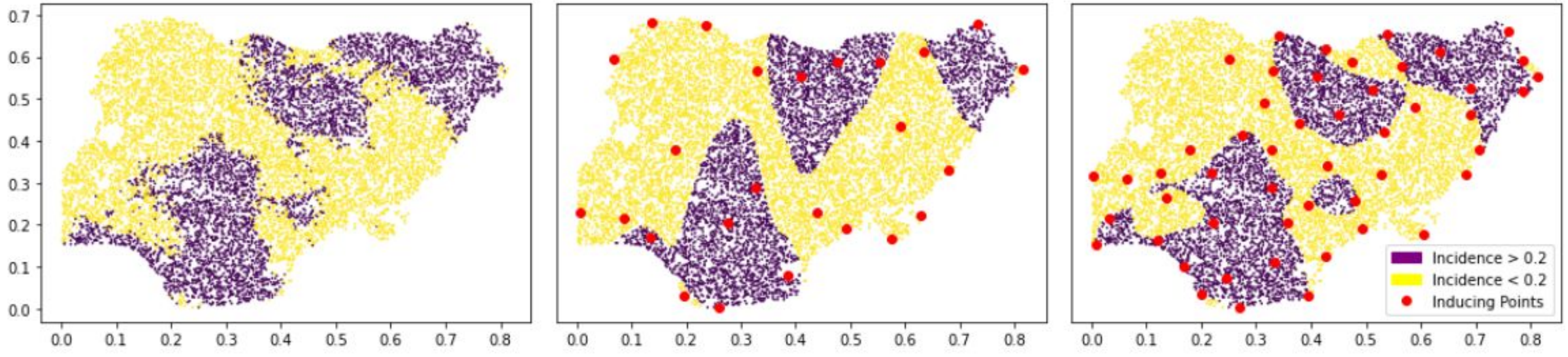
(a) Exact GP (b) SVGP with K-means (c) SVGP with CVR (d) SVGP with IMP-DPP

Secondmind

2) Inducing point allocation in BO loops

Sparse Gaussian process should be customised to the task at hand

- Existing IPA strategies are also unsuitable for active learning

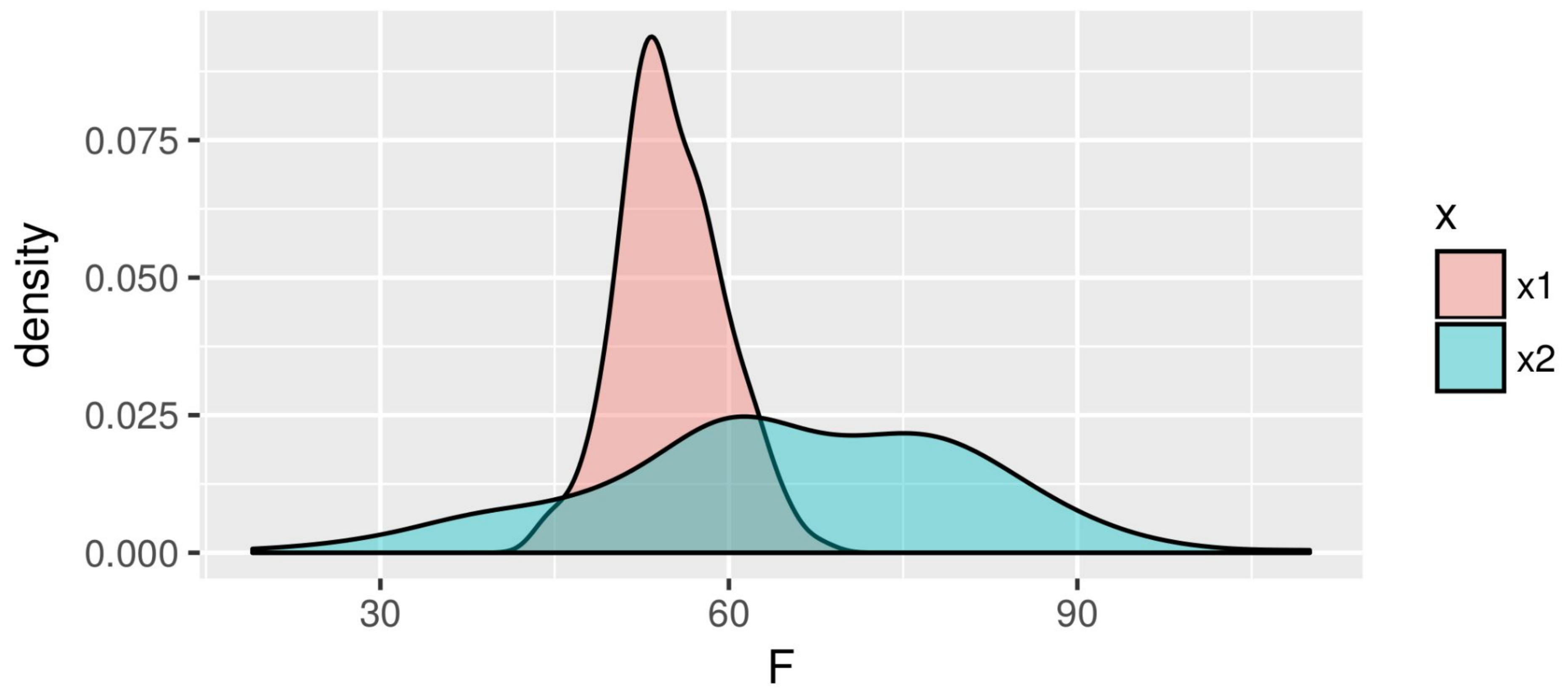


(a) Ground truth of test set (b) SVGP using CVR (c) SVGP using our active learning IPA

3) Risk averse Bayesian Optimisation

Decision makers need to protect themselves from extreme events

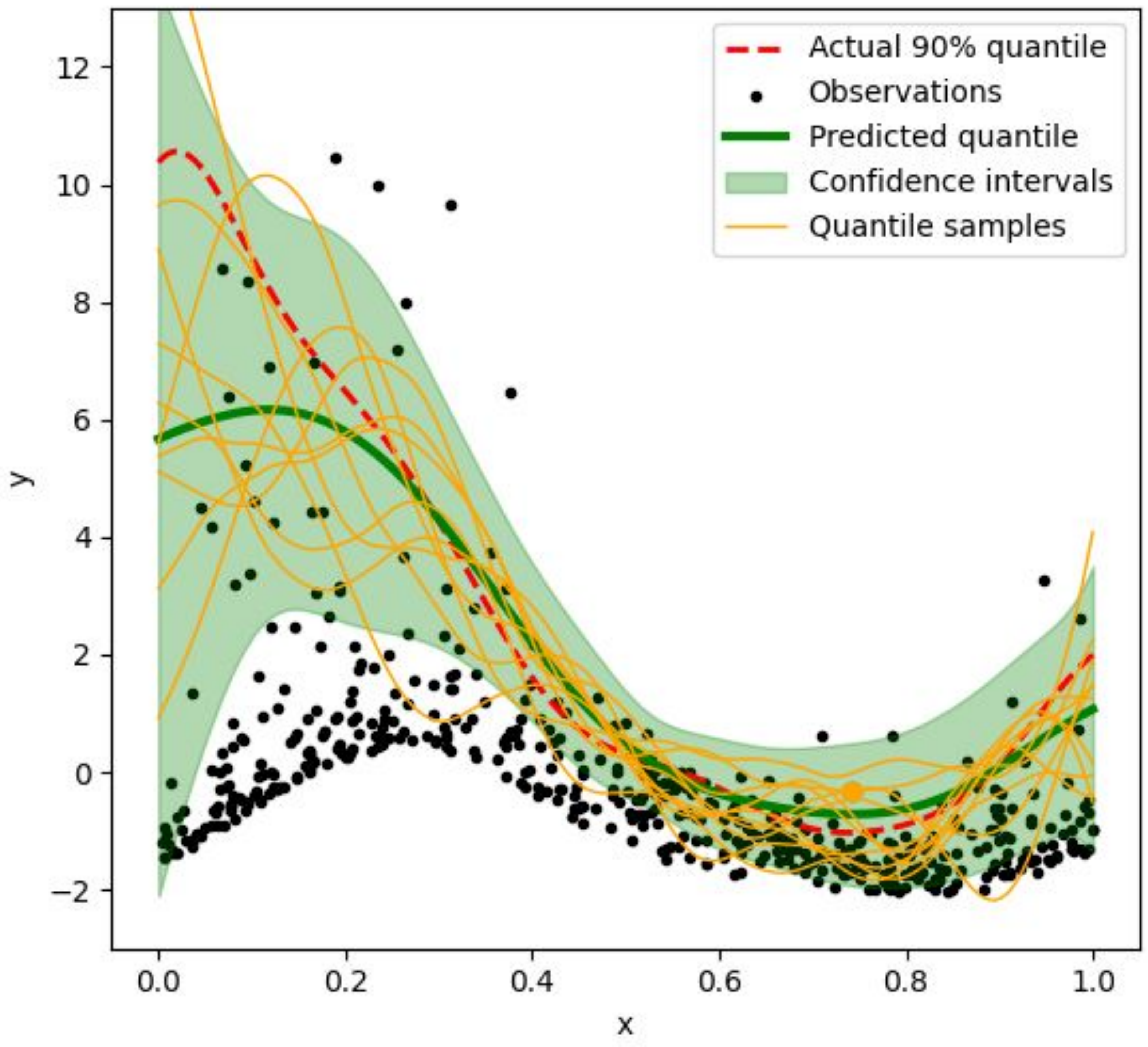
- Objective function is a conditional quantile rather than conditional expectation



3) Risk averse Bayesian Optimisation

Decision makers need to protect themselves from extreme events

- We model observations as quantile + noise



Secondmind