We Still Don't Understand High-Dimensional Bayesian Optimization

Colin Doumont [et al.]
Donney Fan, Natalie Maus, Jacob R. Gardner, Henry Moss, Geoff Pleiss

Tübingen AI Center
c.doumont@uni-tuebingen.de

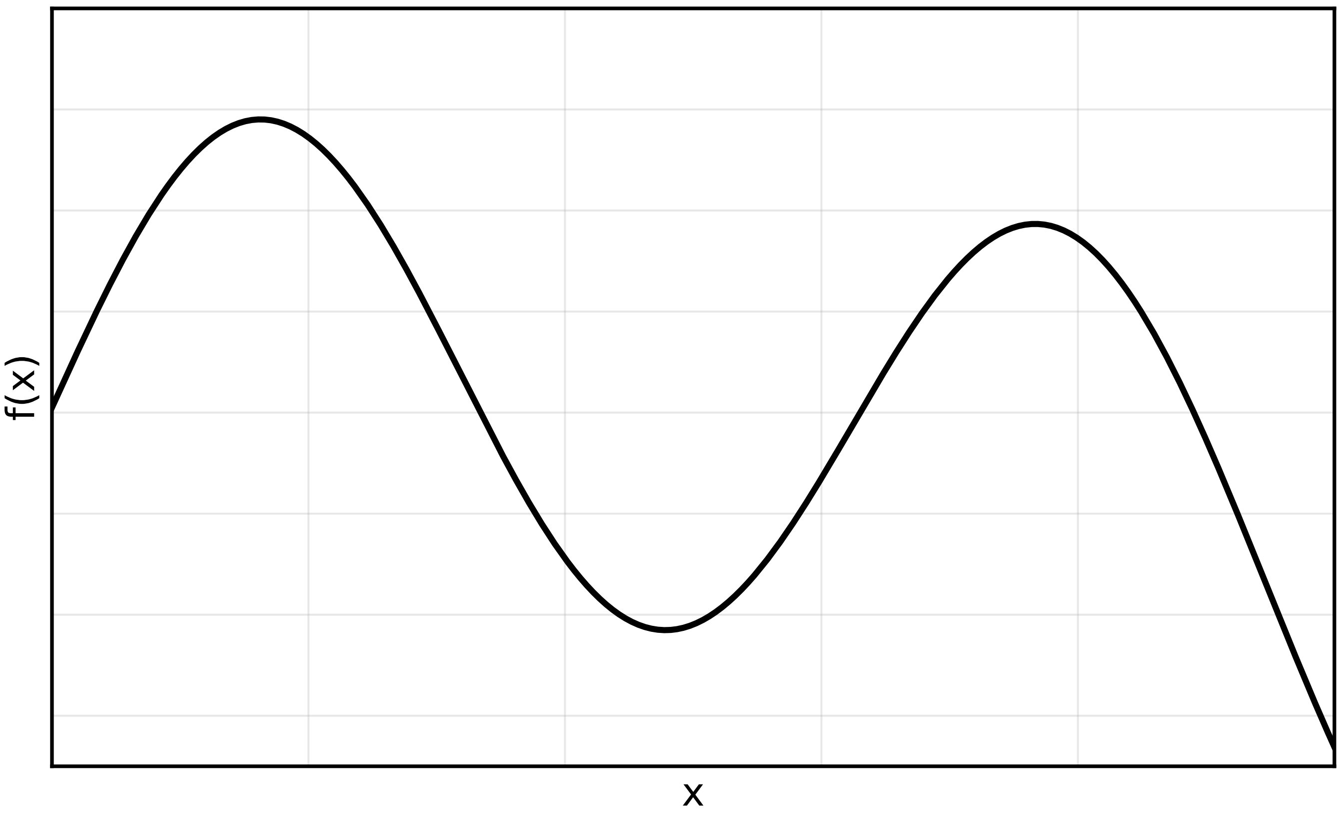
3 May 2026

VECTOR INSTITUTE

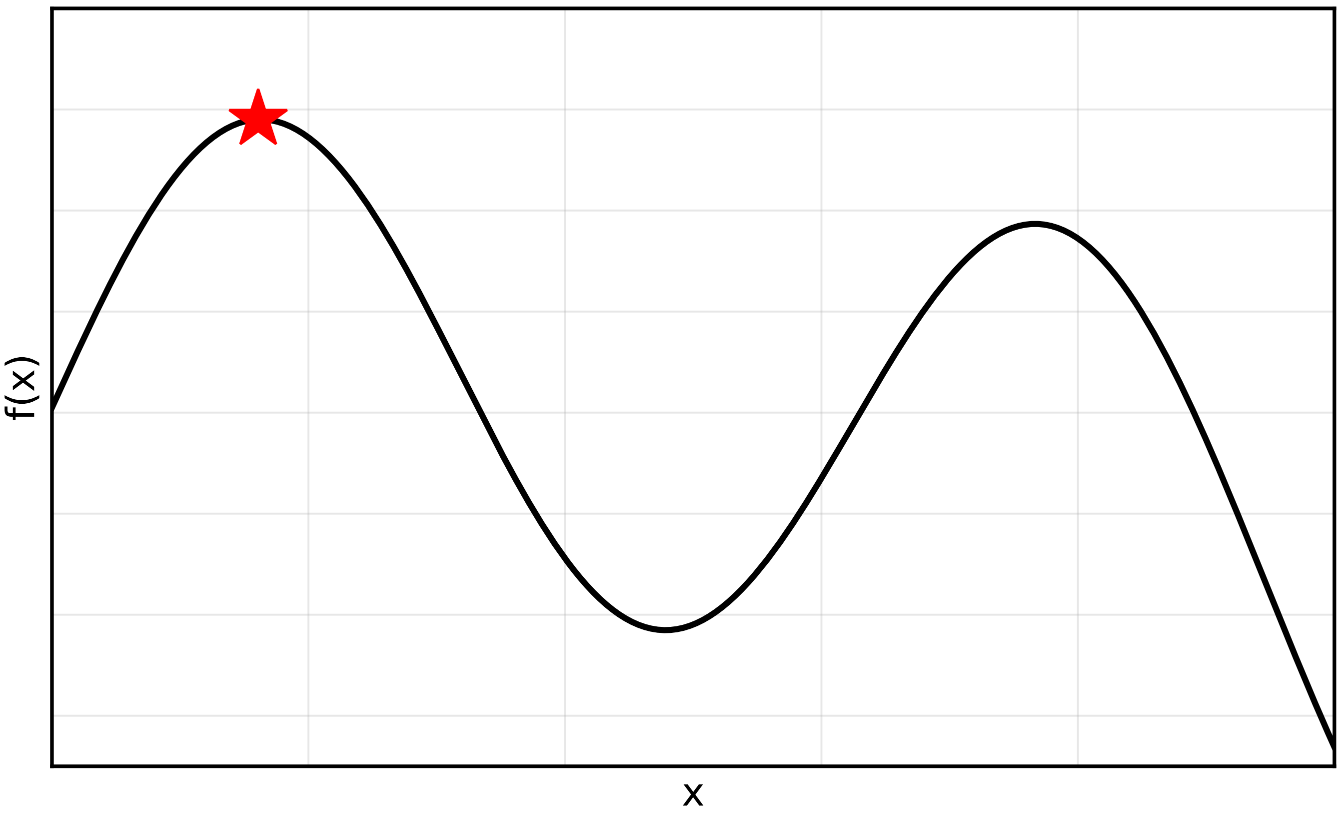
Content

1. What is Bayesian optimization (BO)?

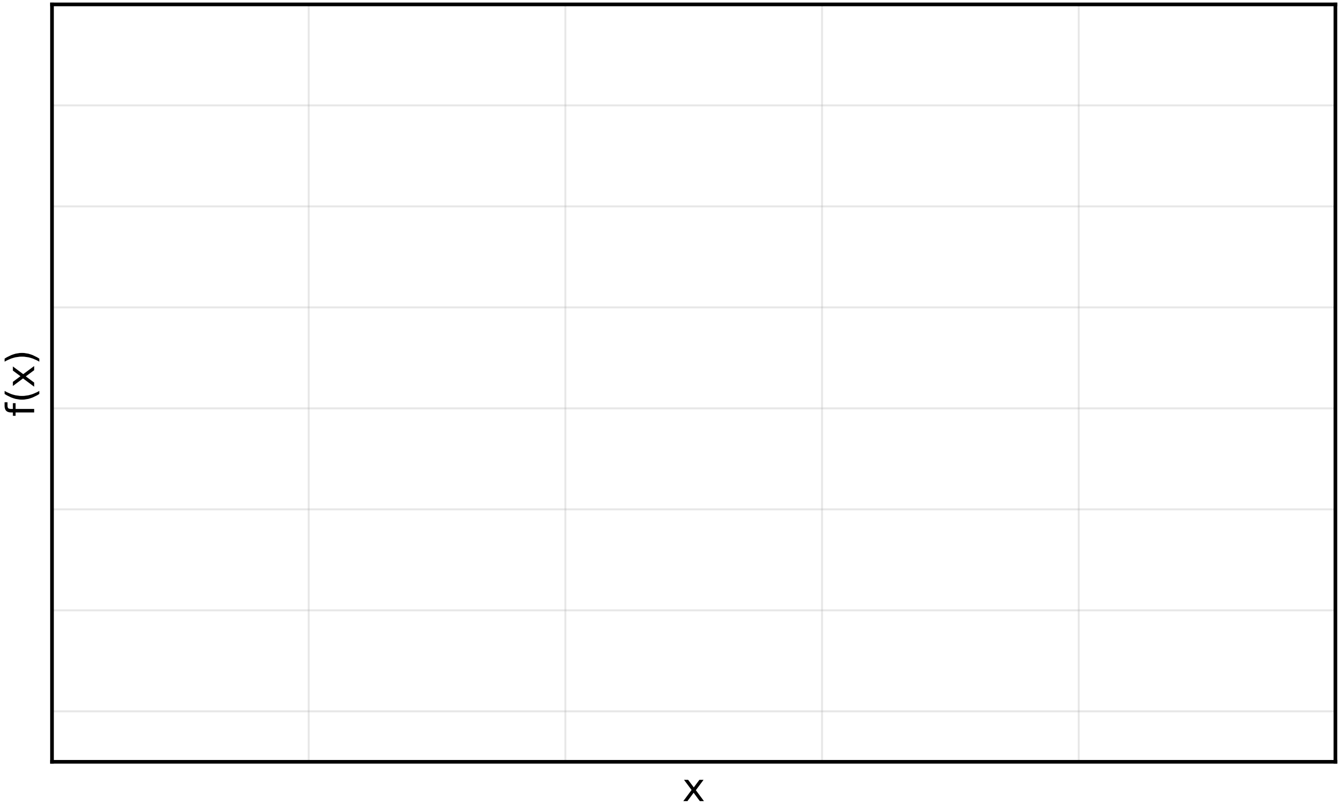
Bayesian optimization



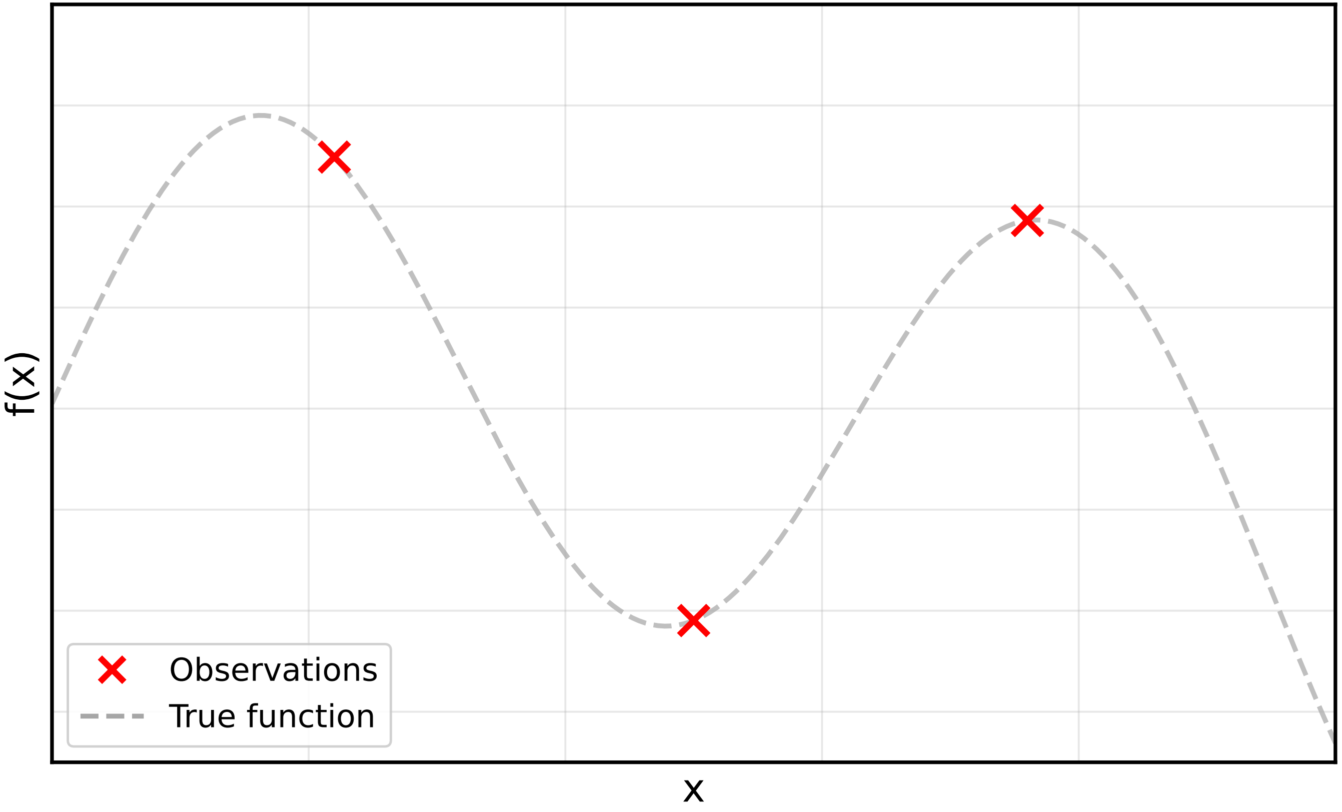
Bayesian optimization



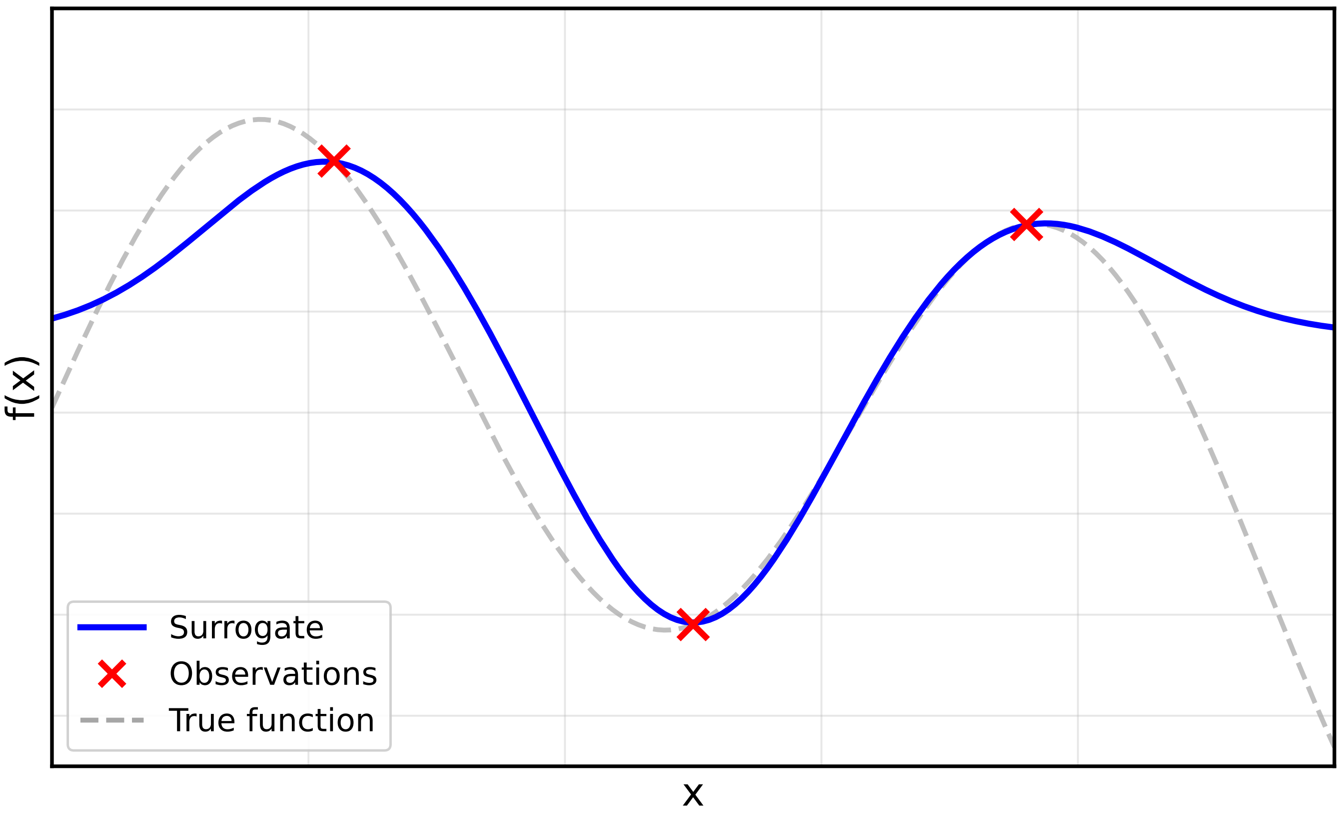
Bayesian optimization



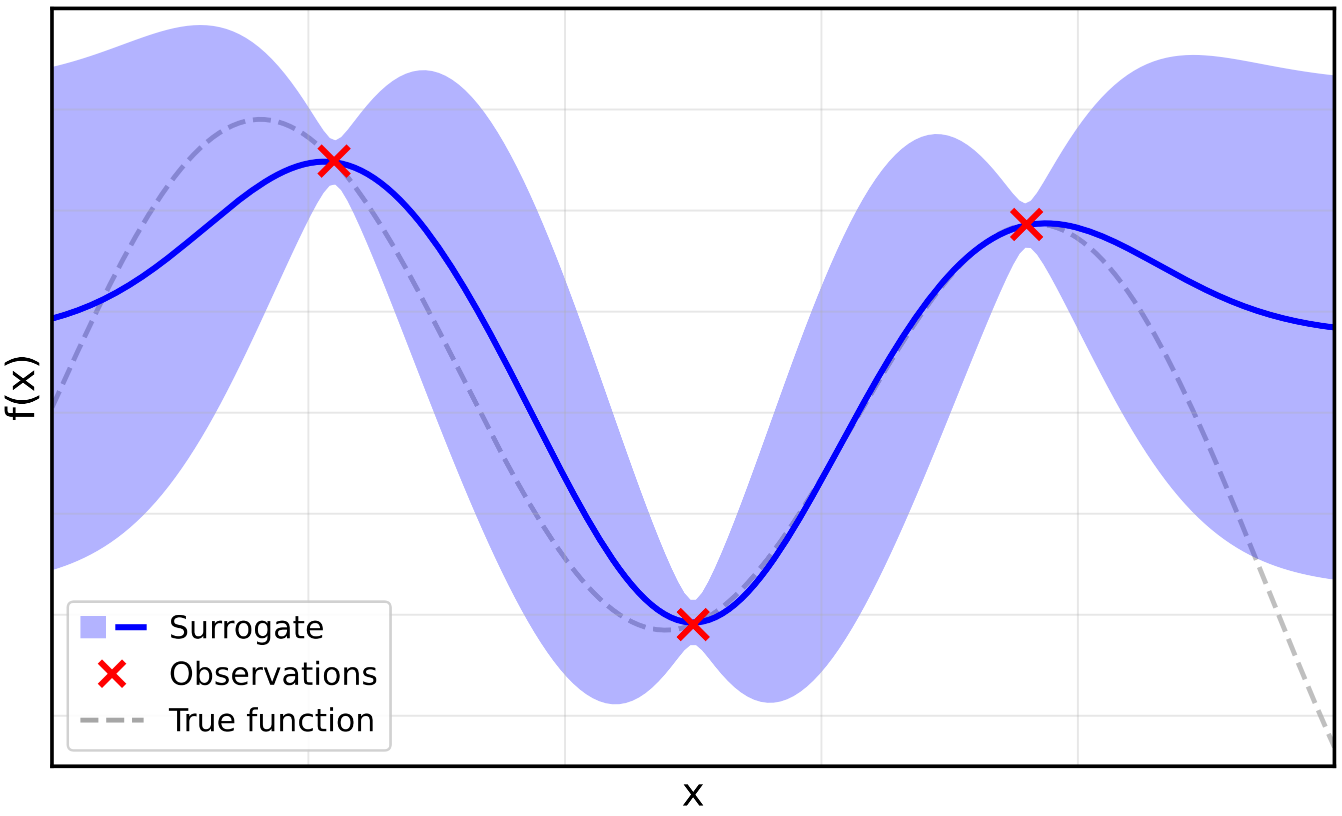
Bayesian optimization



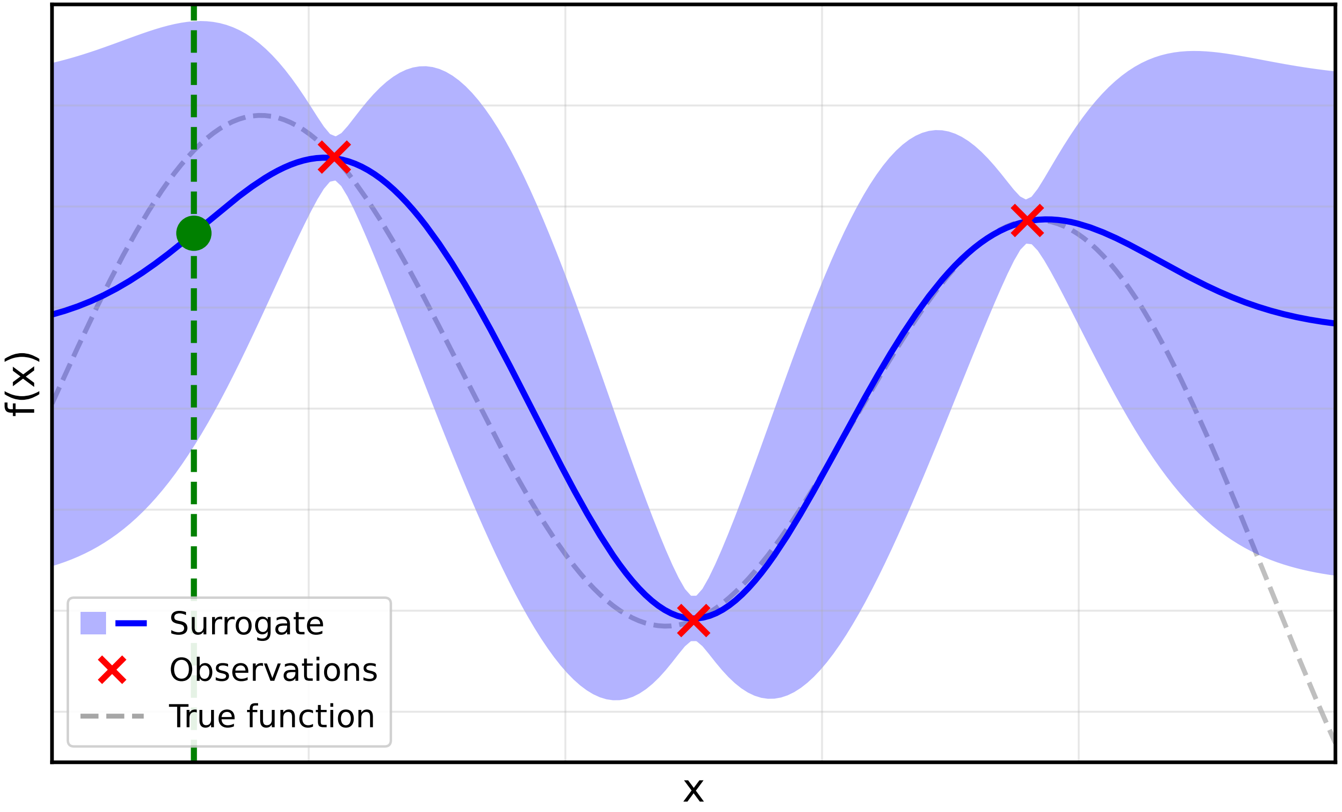
Bayesian optimization



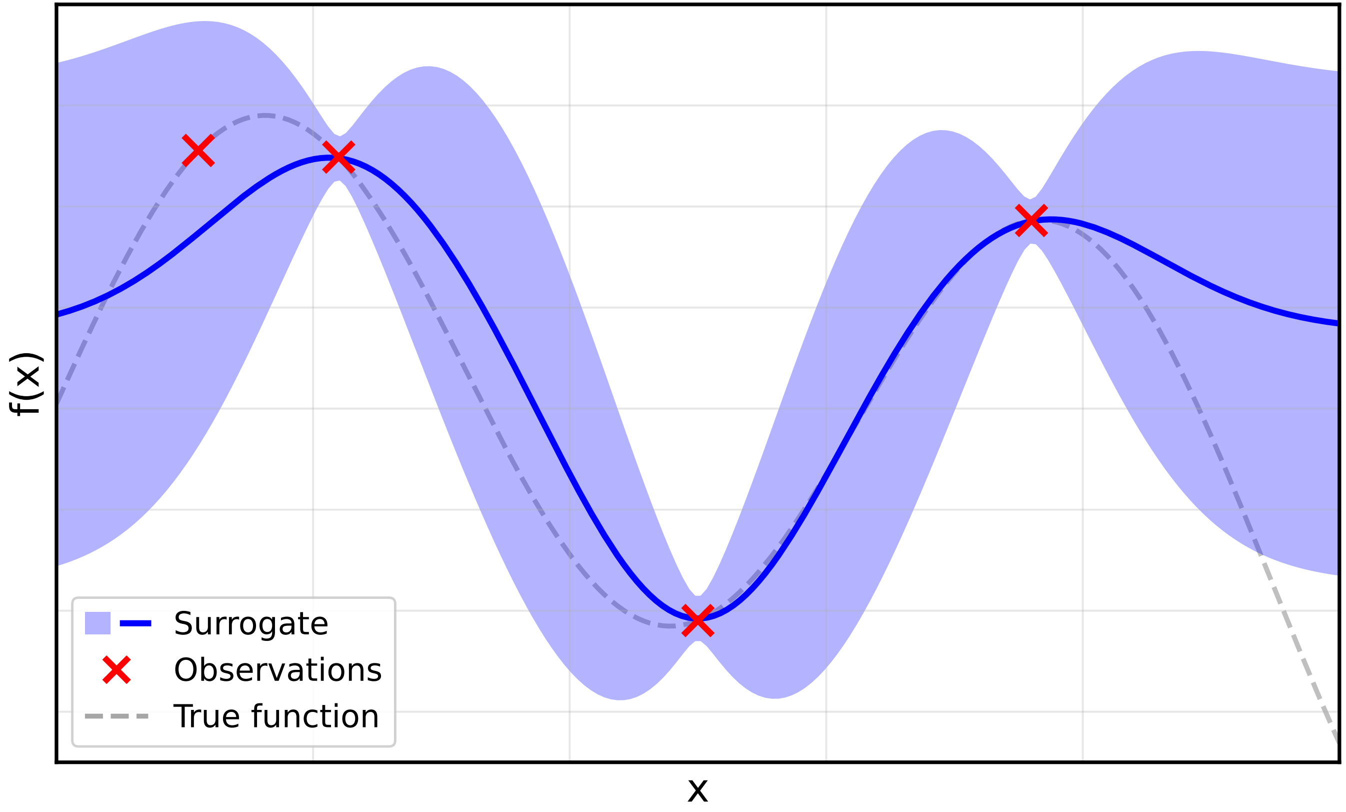
Bayesian optimization



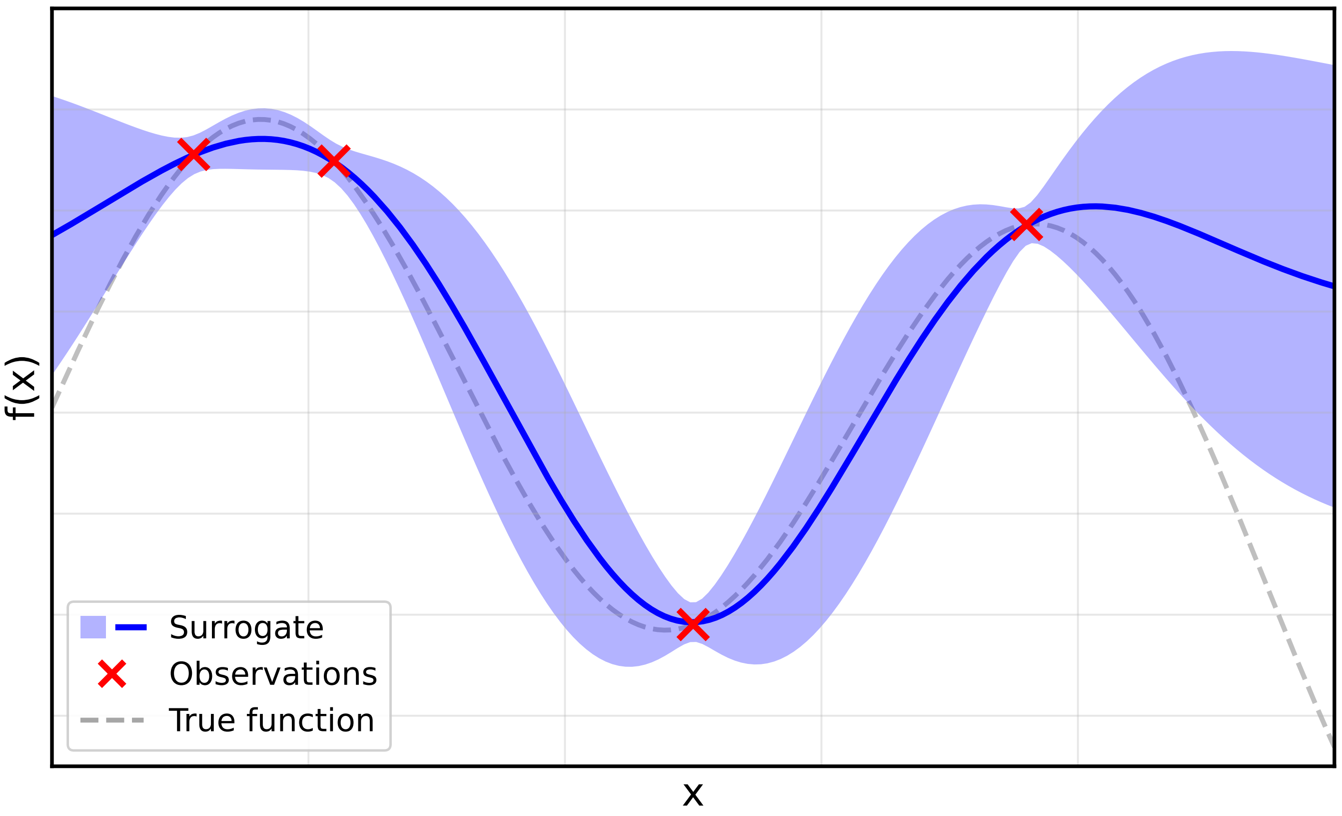
Bayesian optimization



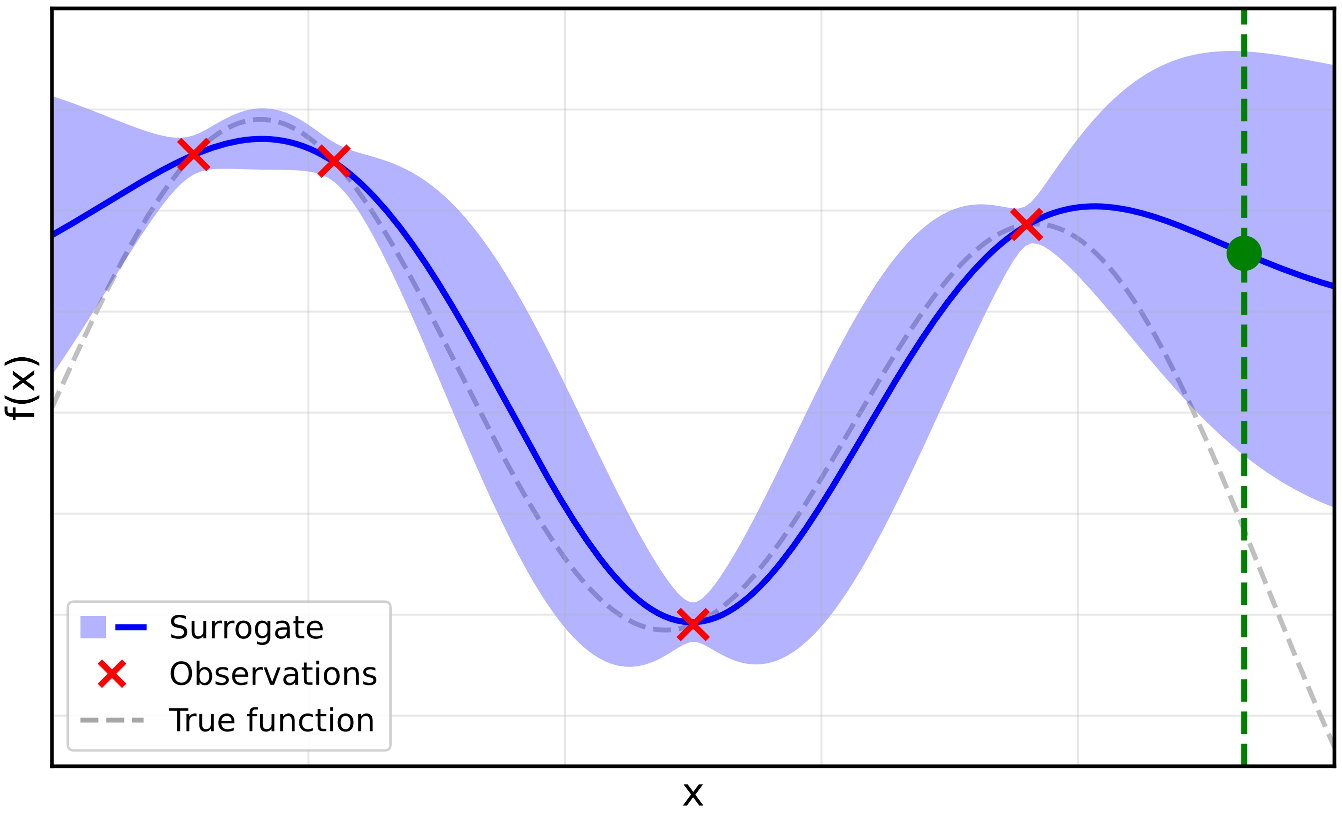
Bayesian optimization



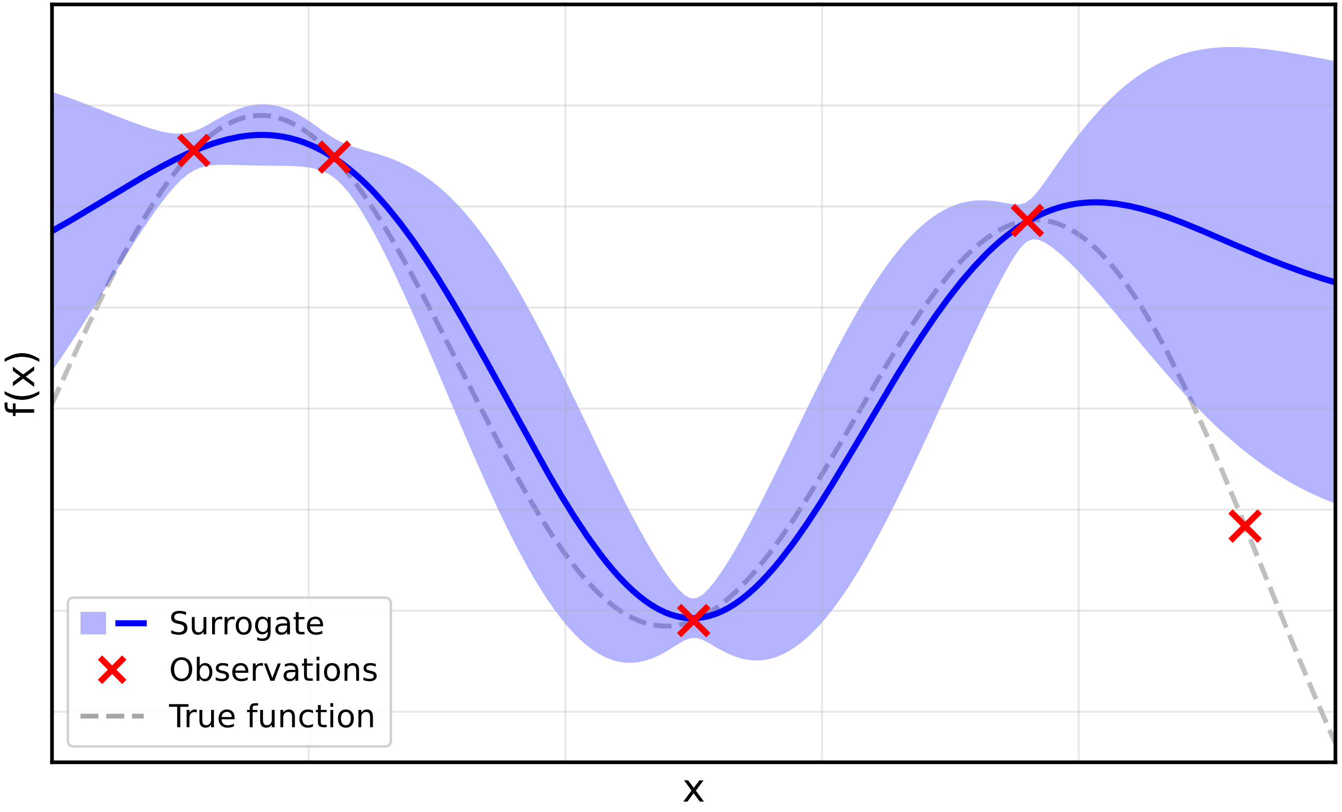
Bayesian optimization



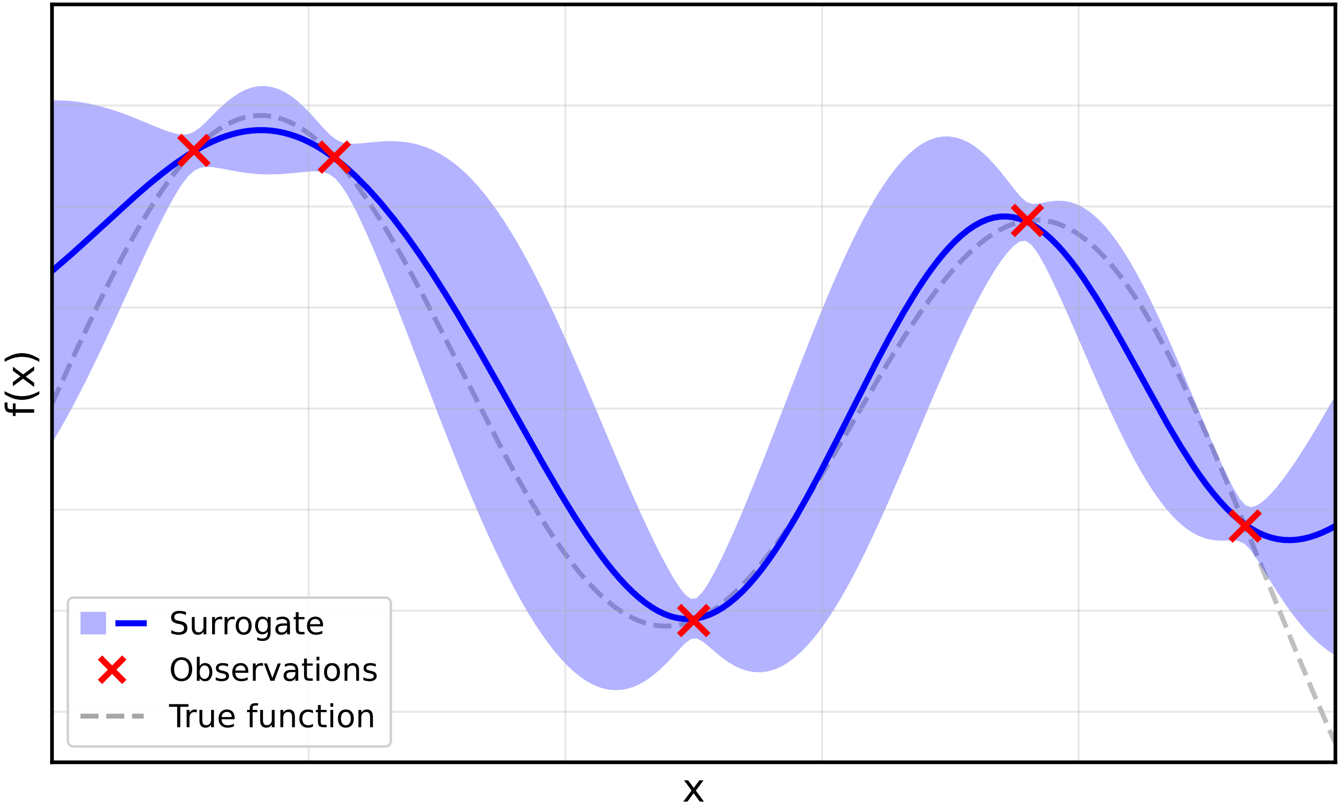
Bayesian optimization



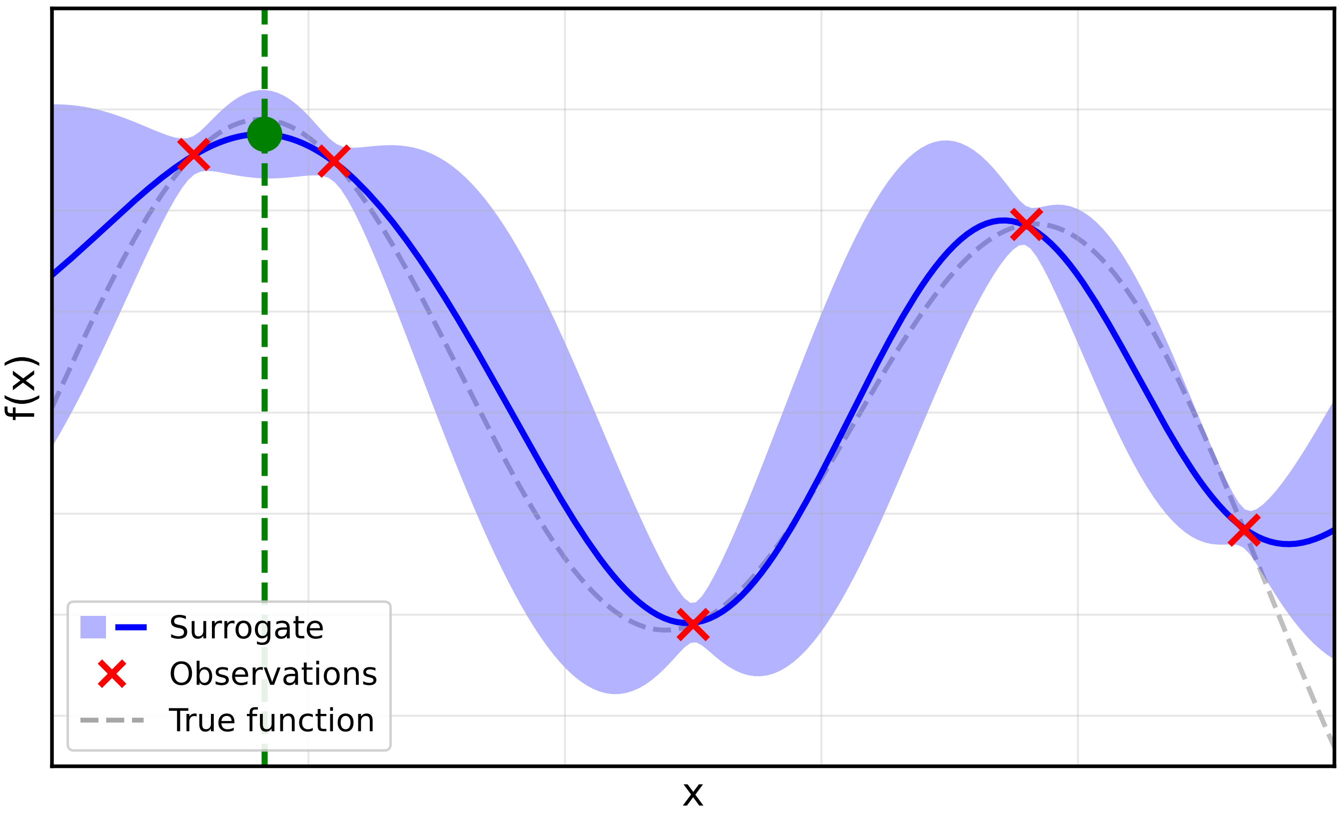
Bayesian optimization



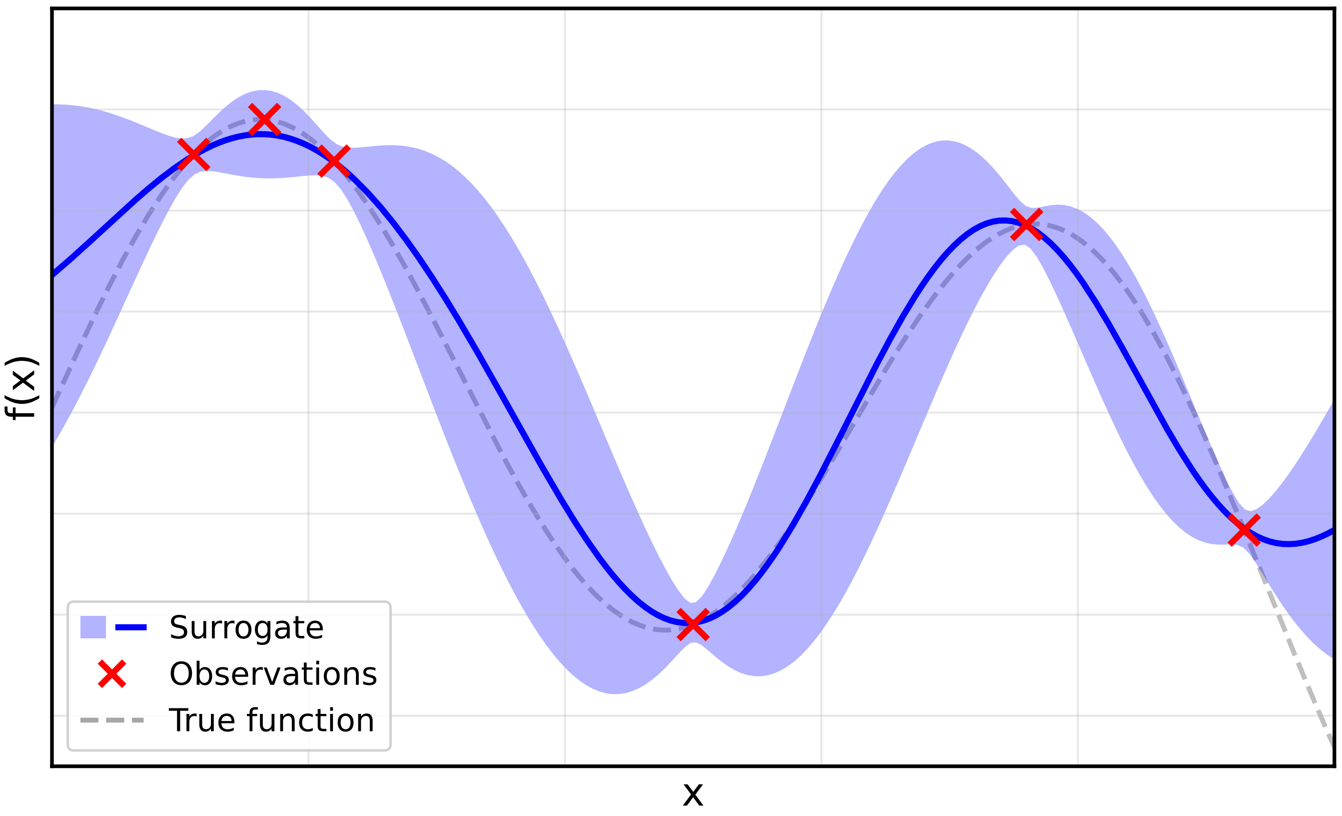
Bayesian optimization



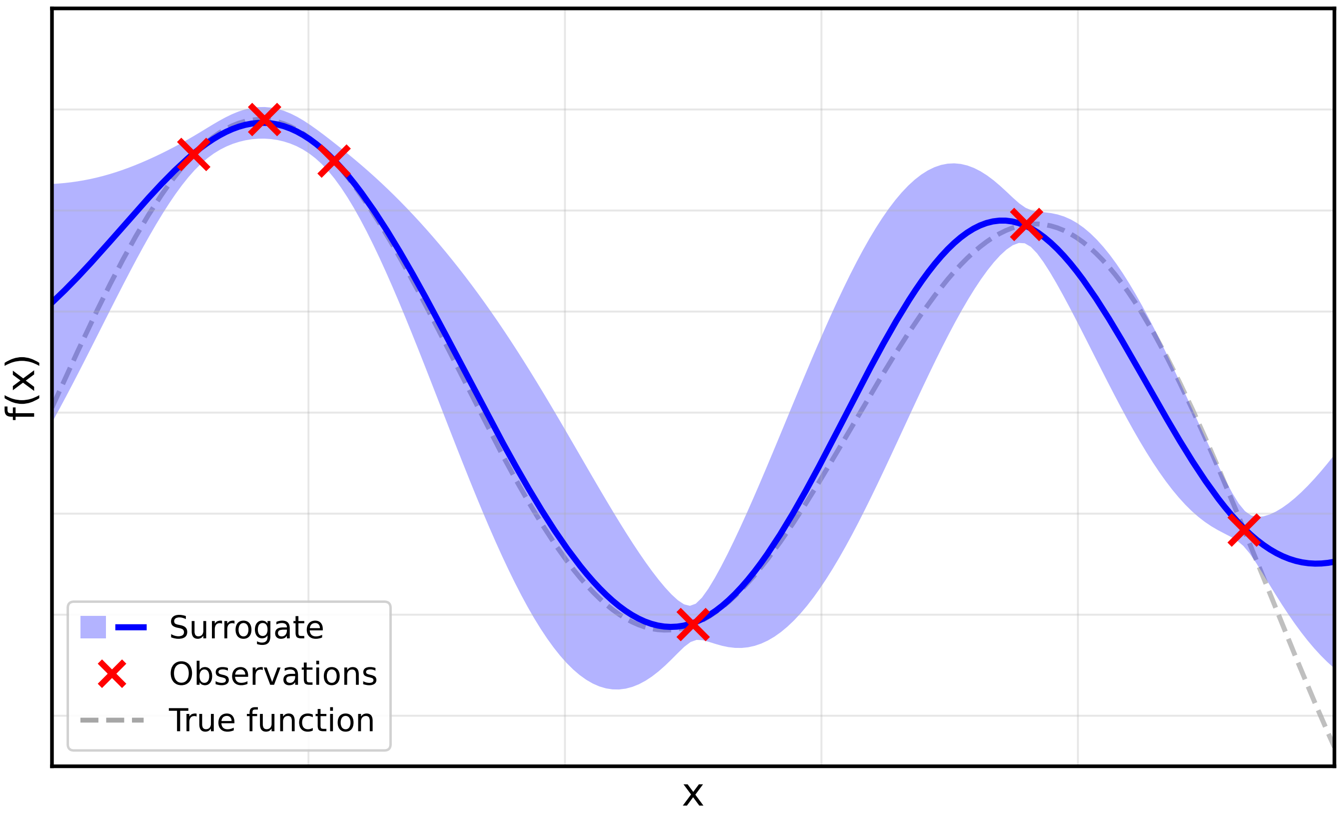
Bayesian optimization



Bayesian optimization



Bayesian optimization



Content

1. What is Bayesian optimization (BO)?

Content

1. What is Bayesian optimization (BO)?
2. What changes in high-dimensional spaces (HDBO)?

Curse of dimensionality – problem

Curse of dimensionality — problem

GP-UCB simple regret (Srinivas et al., 2010):

$$r_T = \mathcal{O}\left(\sqrt{\frac{\log(T)^{D+1}}{T}}\right)$$

Achieving ε-accuracy requires T growing **exponentially** in D!

Curse of dimensionality — solutions

Curse of dimensionality — solutions

1. Structural assumptions:

Curse of dimensionality — solutions

1. Structural assumptions:
   - trust regions
   - random embeddings
   - variable selection
   - ...

Curse of dimensionality — solutions

1. Structural assumptions:
    - trust regions
    - random embeddings
    - variable selection
    - …

2. Non-linear surrogate

Curse of dimensionality — solutions

1. ~~Structural assumptions:~~
    - ~~trust regions~~
    - ~~random embeddings~~
    - ~~variable selection~~
    - ~~…~~

2. Non-linear surrogate (smoother)

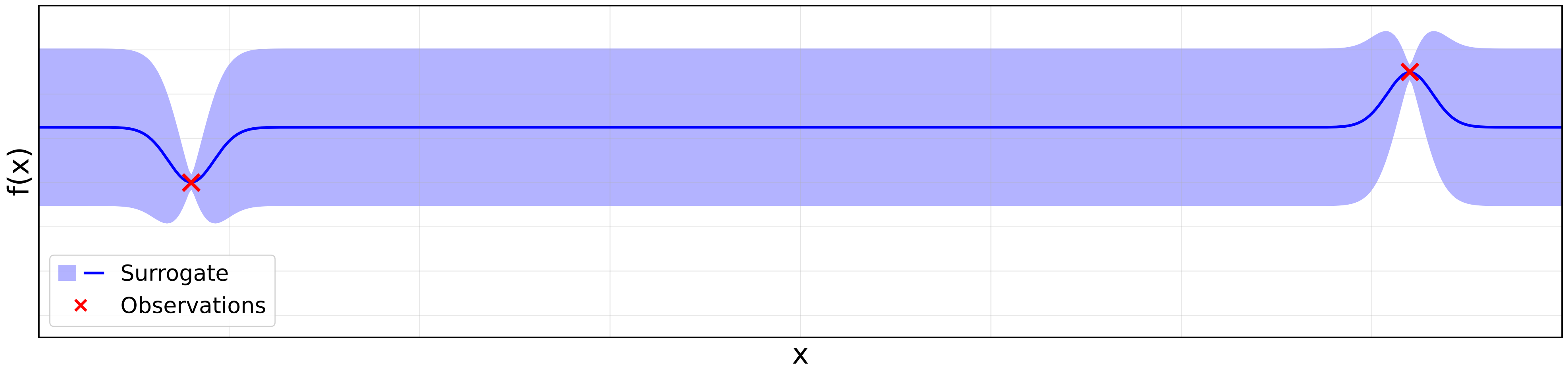
Smoother surrogates ("Vanilla BO")

Vanilla Bayesian Optimization Performs Great in High Dimensions

Carl Hvarfner [1] **Erik O. Hellsten** [1 2] **Luigi Nardi** [1 2]

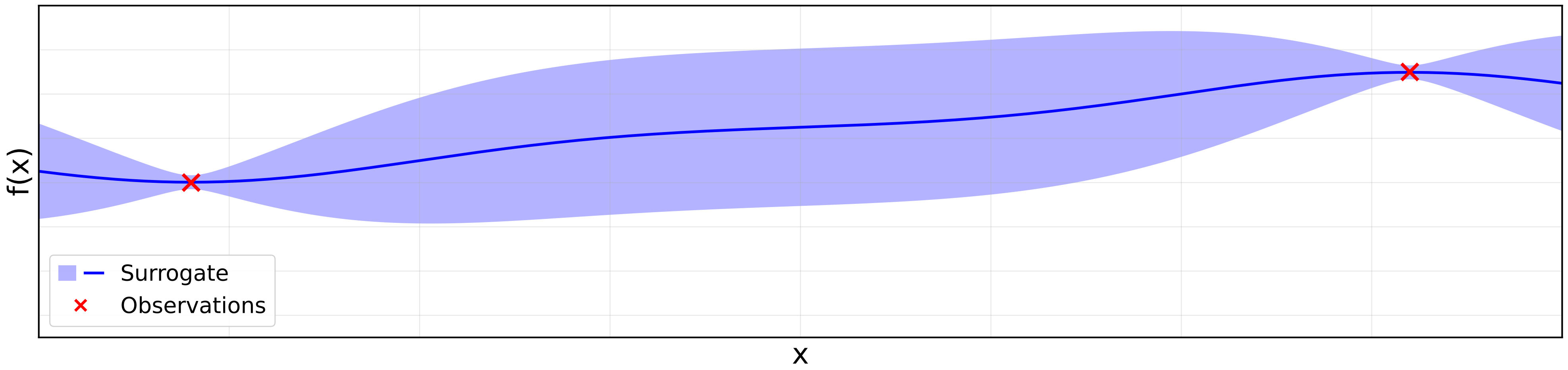
Smoother surrogates ("Vanilla BO")

$$\mathrm{Cov}[f(\mathbf{x}), f(\mathbf{x}')] \propto \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2}\right)$$



Smoother surrogates ("Vanilla BO")

$$\mathrm{Cov}[f(\mathbf{x}), f(\mathbf{x}')] \propto \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2D}\right)$$



Content

1. What is Bayesian optimization (BO)?
2. What changes in high-dimensional spaces (HDBO)?

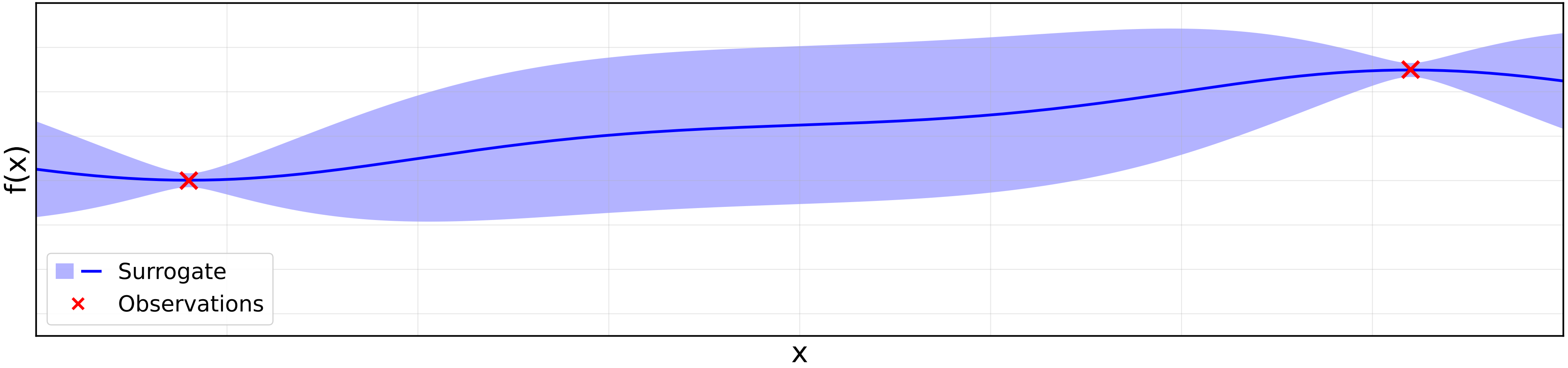
Content

1. What is Bayesian optimization (BO)?
2. What changes in high-dimensional spaces (HDBO)?
3. Why do we still not understand HDBO?

Pushing smoothness to the extreme

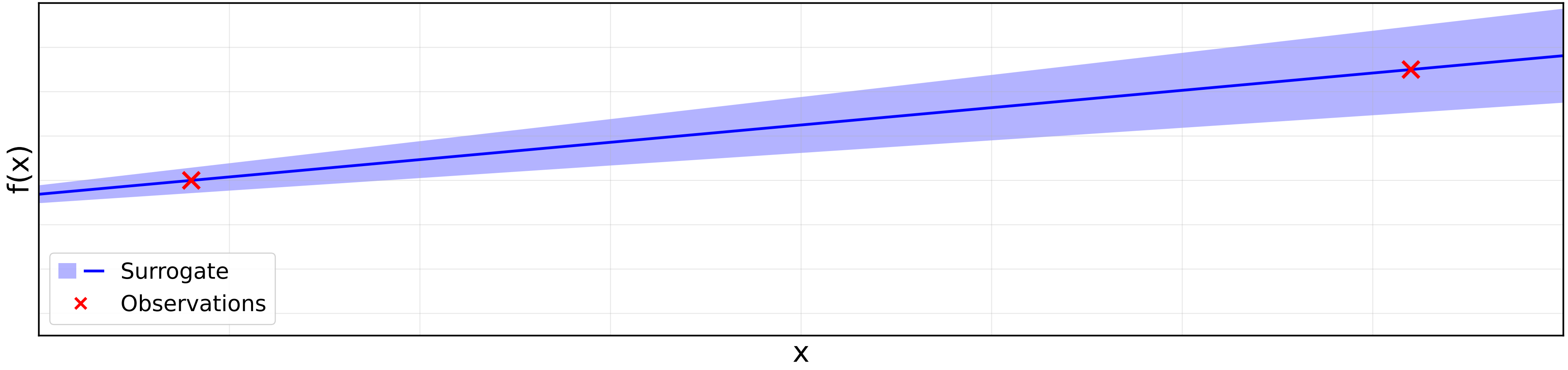
Pushing smoothness to the extreme

What if we replace the non-linear surrogate...



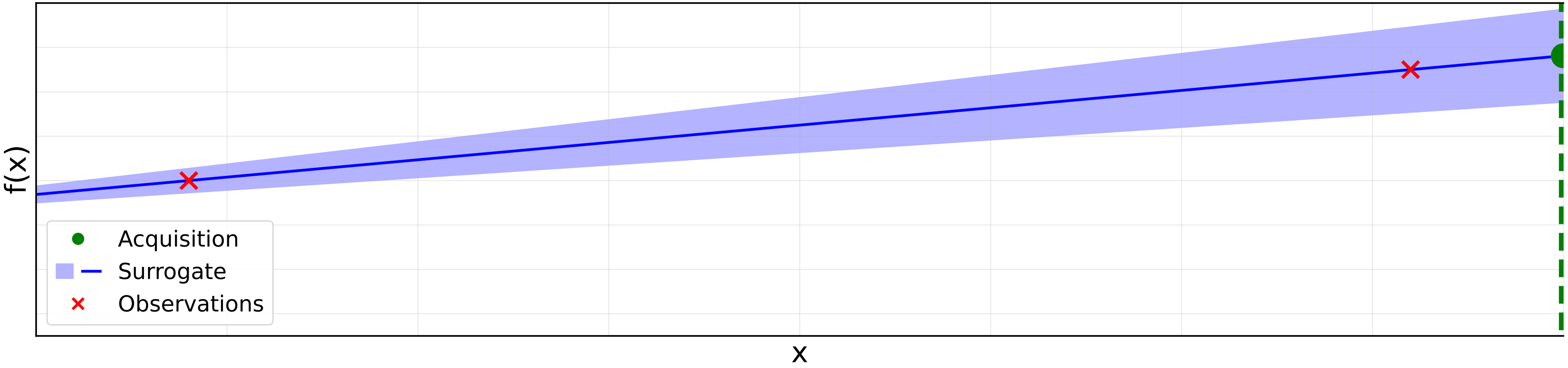
Pushing smoothness to the extreme

...by a (Bayesian) linear regression?



Pushing smoothness to the extreme

We always obtain acquisitions on the boundary!

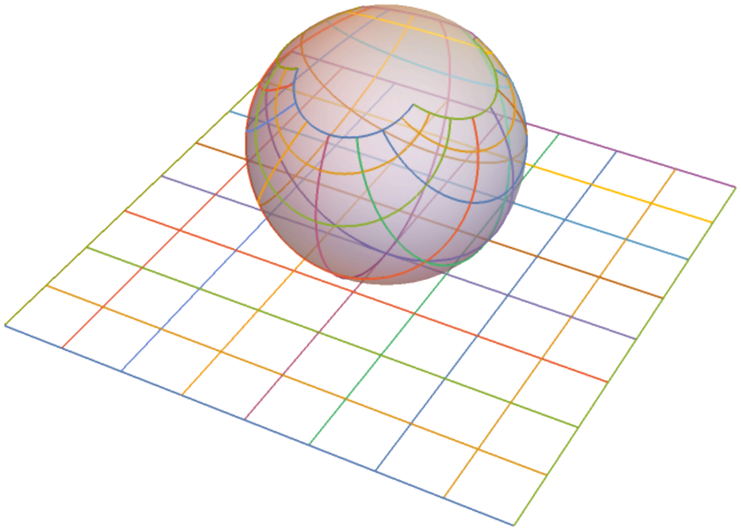


Pushing smoothness to the extreme

Theorem 1. For acquisition functions increasing in posterior mean and variance (e.g. expected improvement), Bayesian linear models will maximize acquisition on the boundary of the search space.

Spherical projection prevents boundary acquisitions

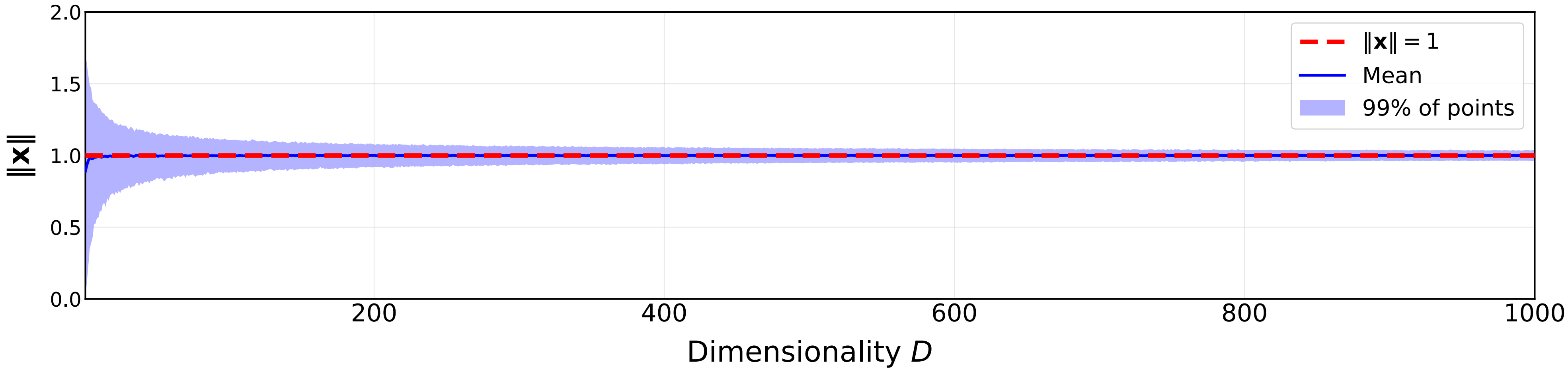
Spherical projection prevents boundary acquisitions



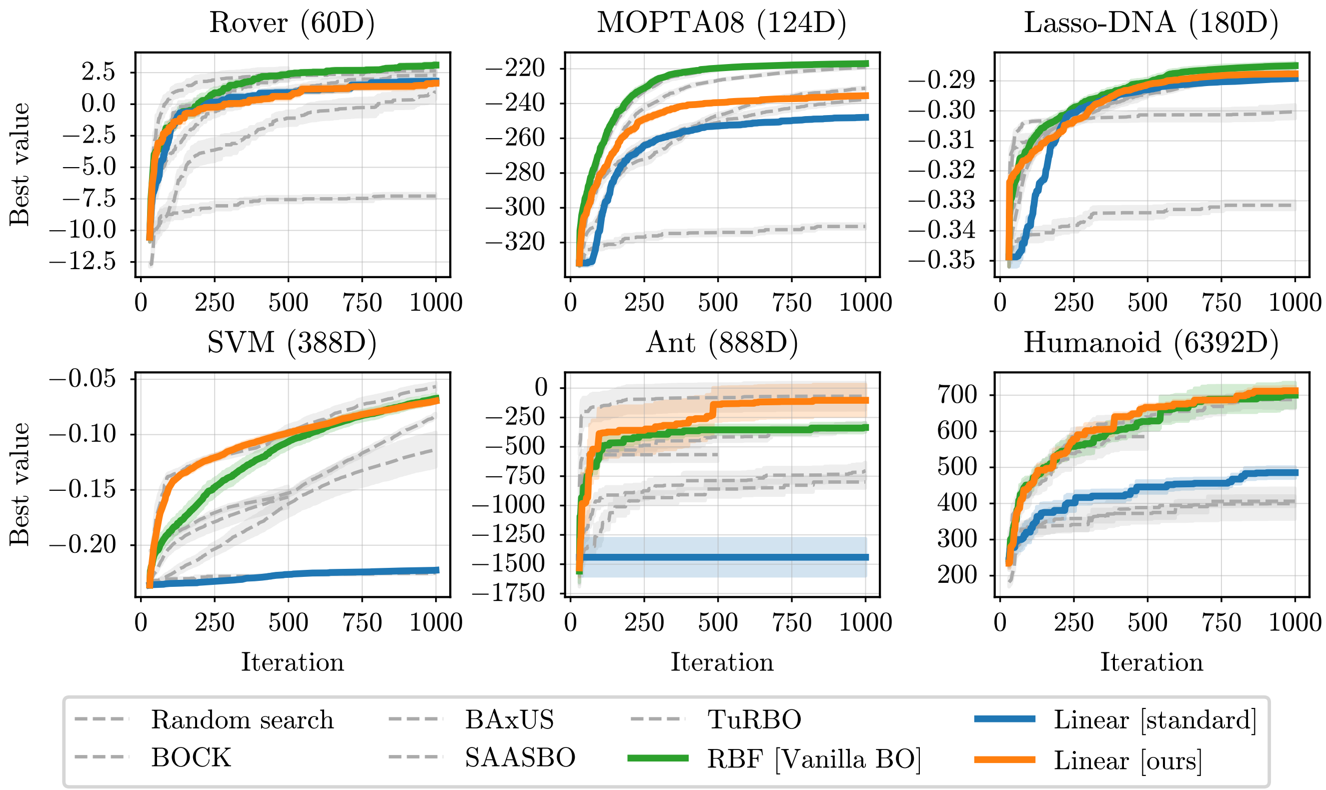
$$\underbrace{\mathbf{x}}_{[-1,1]^D} \rightarrow \underbrace{P(\mathbf{x})}_{\mathbb{S}^D} = \frac{1}{\|\mathbf{x}\|^2+1}\left[2x_1,\ \cdots,\ 2x_D,\ \|\mathbf{x}\|^2-1\right]$$

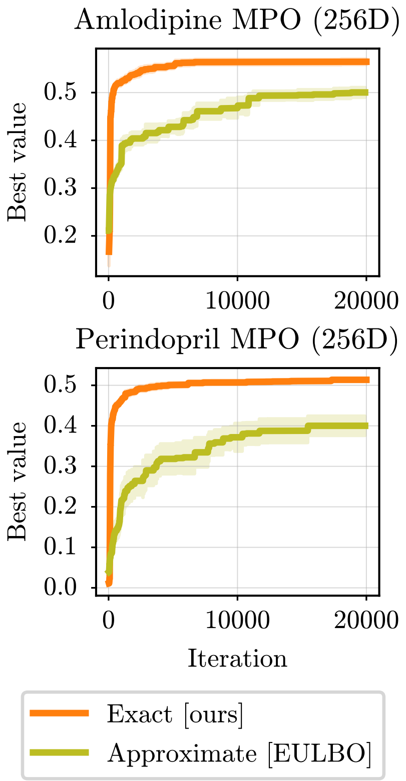
Spherical projection in high dimensions

In high D, $\|\mathbf{x}\| \approx 1 \Longrightarrow P(\mathbf{x}) \approx [\mathbf{x}; 0]$ (projection is nearly a no-op)



Empirical results





Why are these results so surprising?

Why are these results so surprising?

To achieve state-of-the-art in high-dimensional BO, we need:

Why are these results so surprising?

To achieve state-of-the-art in high-dimensional BO, we need:

- structural assumption + non-linear surrogate (past ~10 years)

Why are these results so surprising?

To achieve state-of-the-art in high-dimensional BO, we need:

- ~~structural assumption + non-linear surrogate~~
- non-linear surrogate* (past ~2 years)

Why are these results so surprising?

To achieve state-of-the-art in high-dimensional BO, we need:

- ~~structural assumption + non-linear surrogate~~
- ~~non-linear surrogate*~~
- linear surrogate* (this talk)

What’s next?

What’s next?

Less:

- inventing new (complicated) structural-assumption methods

What’s next?

Less:

- inventing new (complicated) structural-assumption methods

More:

- understanding fundamental mechanisms driving HDBO

Thank you!

Colin Doumont

Donney Fan

Natalie Maus

Jacob R. Gardner

Henry Moss

Geoff Pleiss