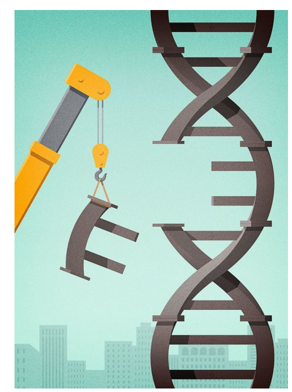
Brief Overview of Bayesian Optimization and Gaussian Process

Black-box Optimisation

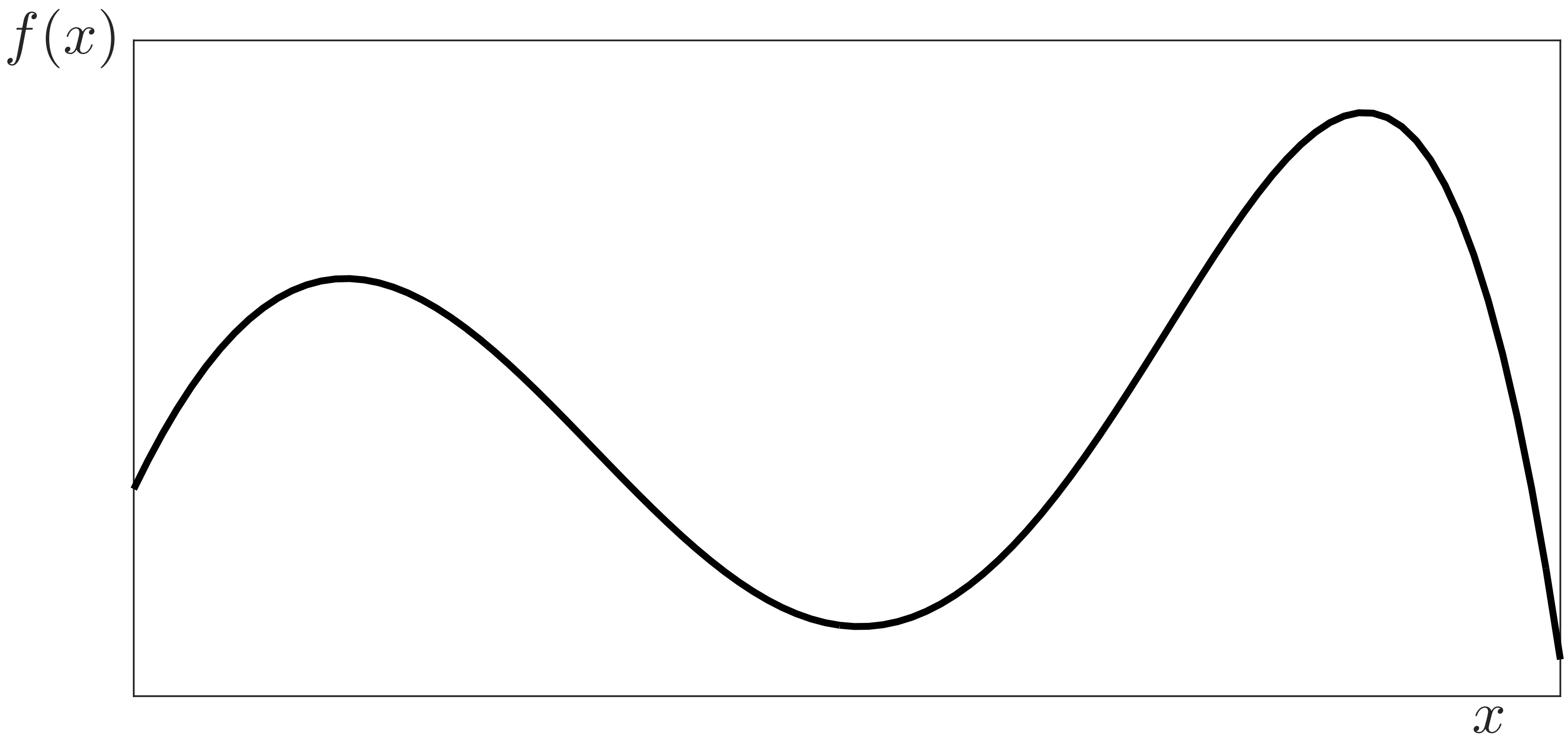
x → **Expensive Blackbox Function** → $f(x)$

Other Examples:
- Pre-clinical Drug Discovery
- Optimal policy in Autonomous Driving
- Synthetic gene design



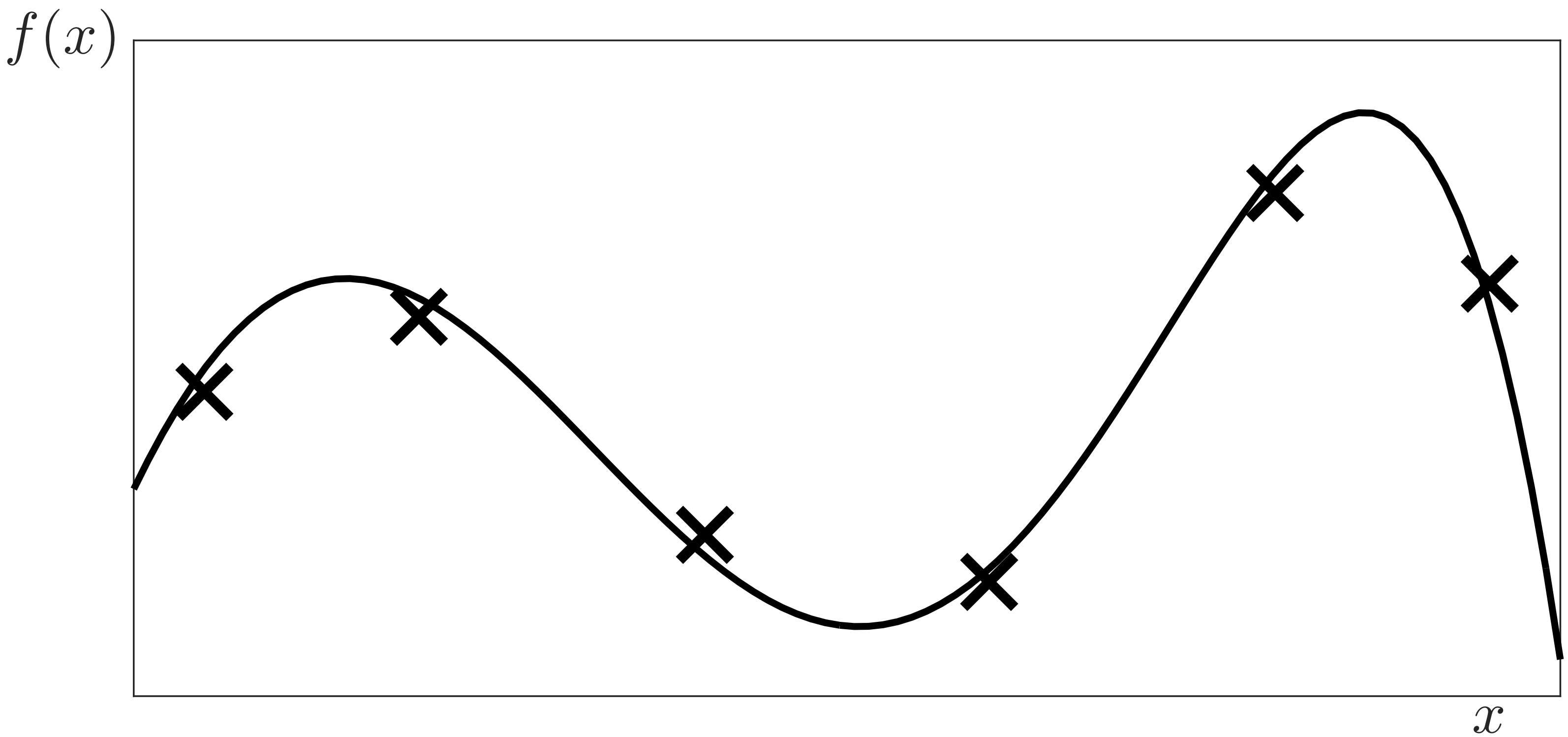
Black-box Optimisation

$f : \mathcal{X} \to \mathbb{R}$ is an expensive, black-box function, accessible only via noisy evaluations.



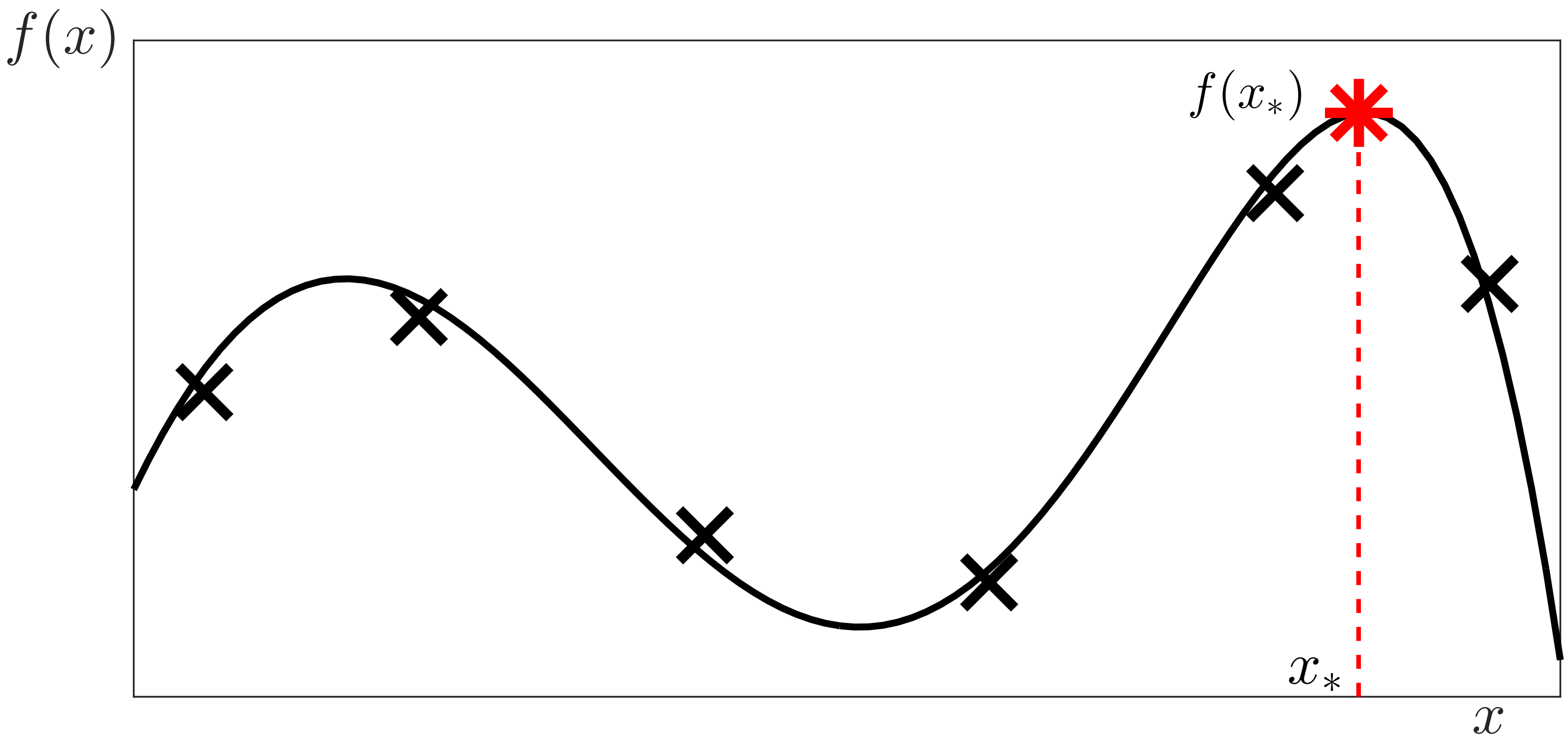
Black-box Optimisation

$f : \mathcal{X} \to \mathbb{R}$ is an expensive, black-box function, accessible only via noisy evaluations.



Black-box Optimisation

$f : \mathcal{X} \to \mathbb{R}$ is an expensive, black-box function, accessible only via noisy evaluations.
Let $x_\star = \mathrm{argmax}_x f(x)$.



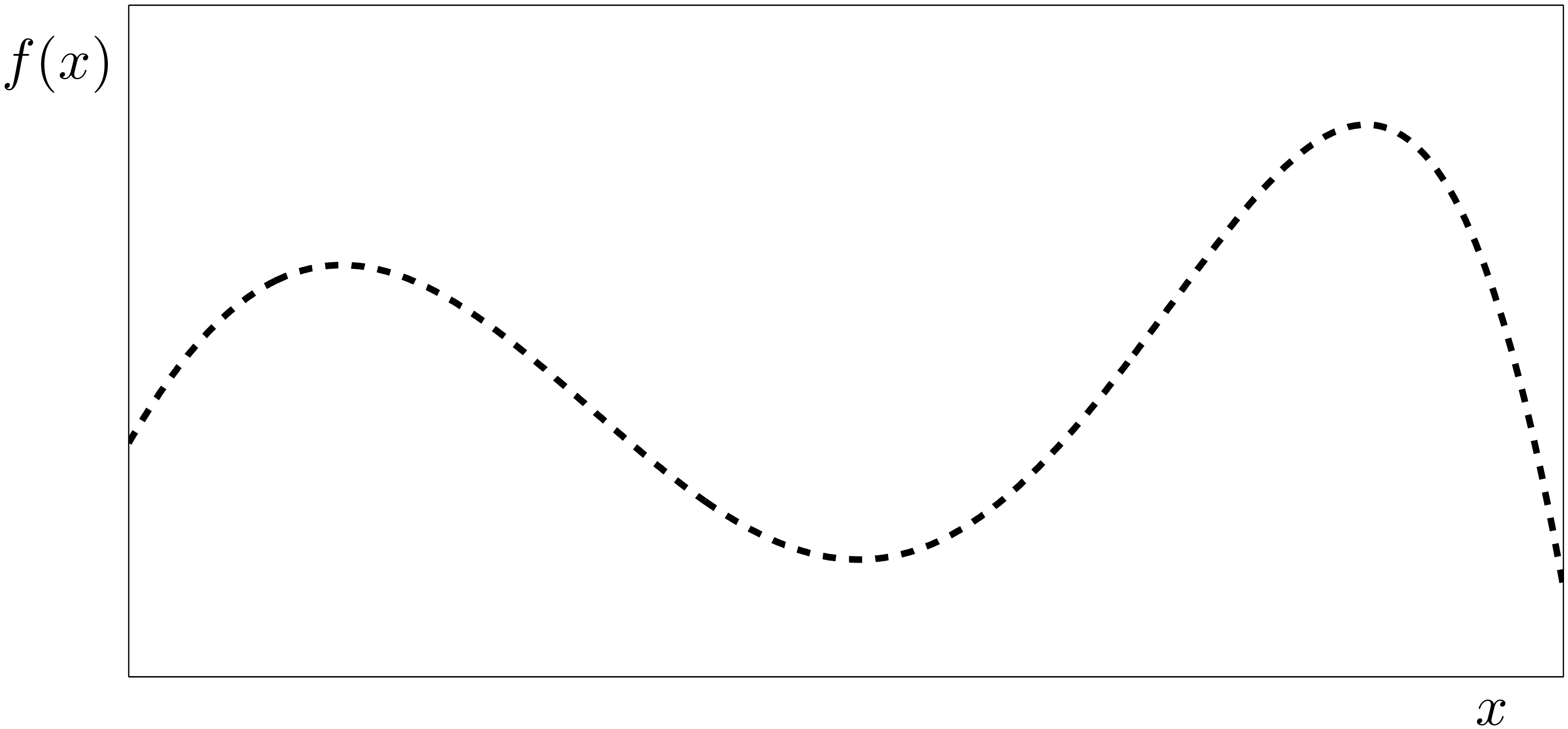
Bayesian Models for f e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

Bayesian Models for f e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

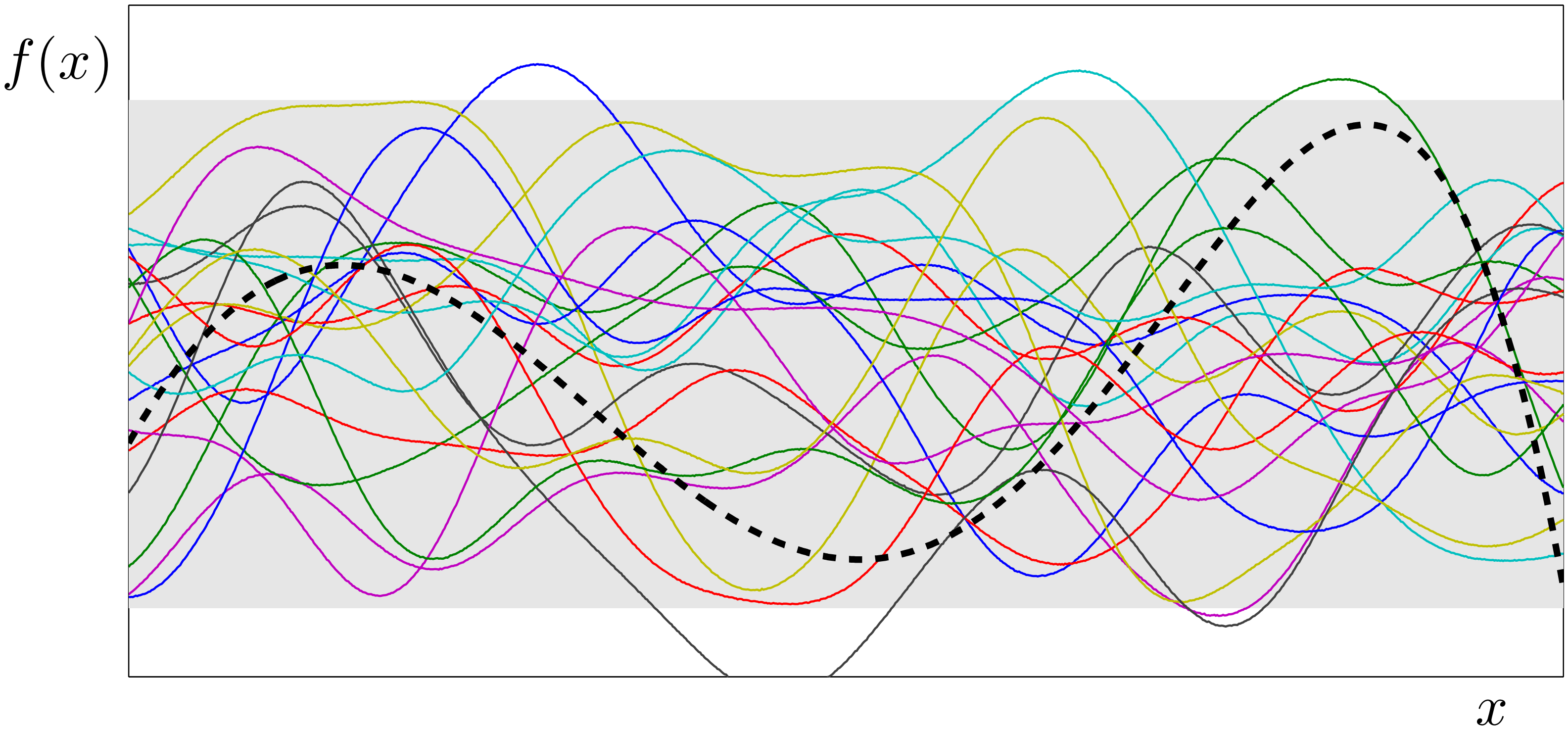
Functions with no observations



Bayesian Models for f e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

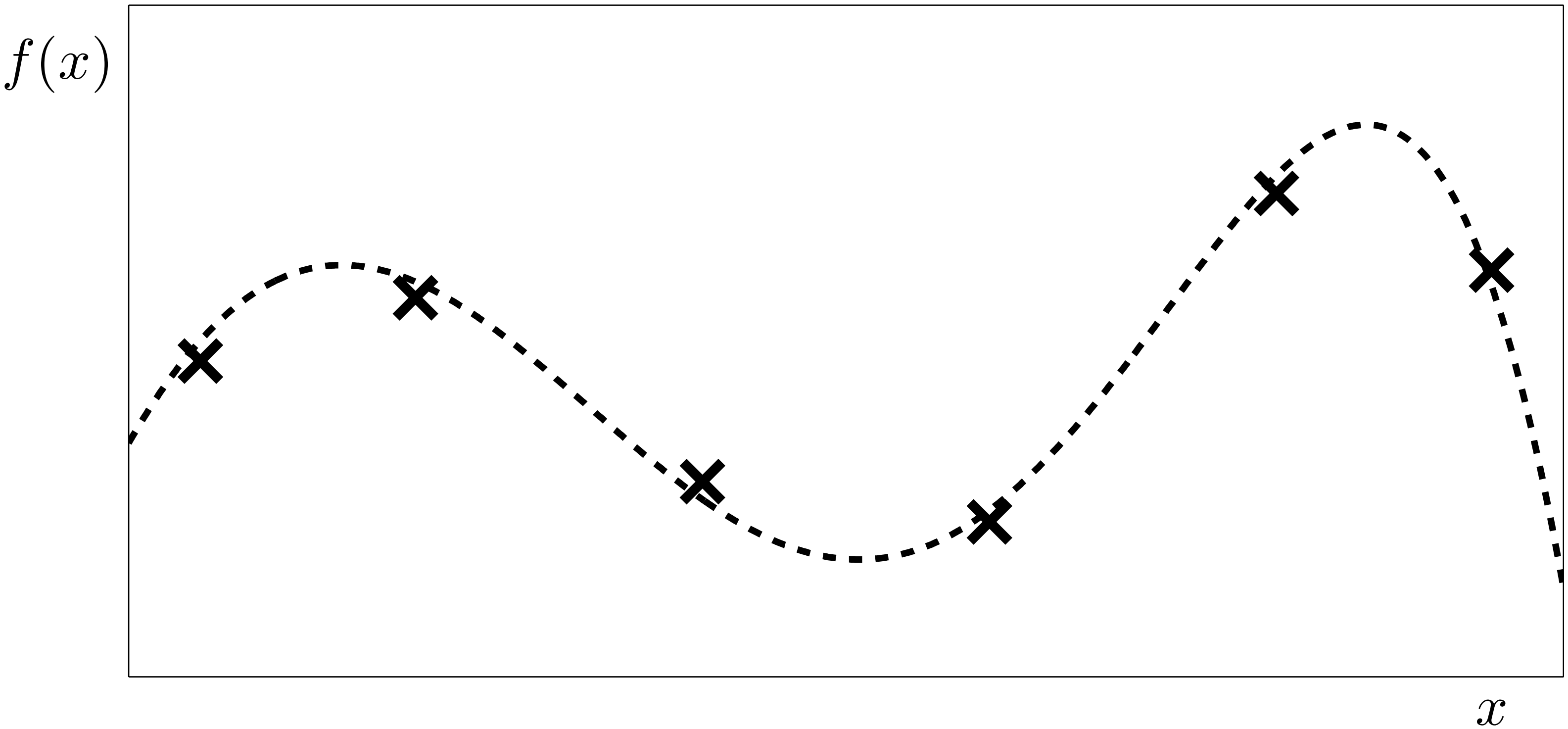
Prior $\mathcal{GP}$



Bayesian Models for f e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

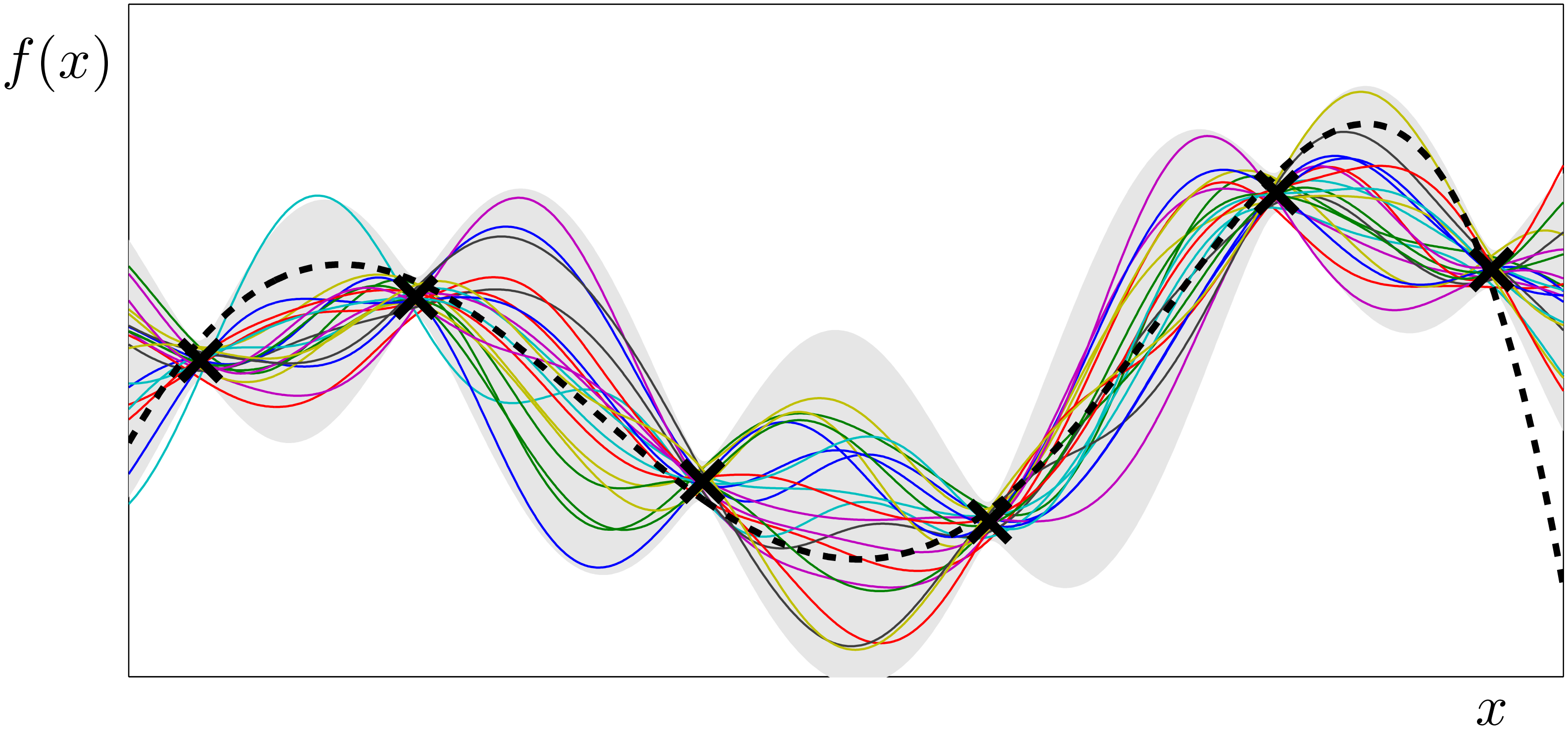
Observations



Bayesian Models for f e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

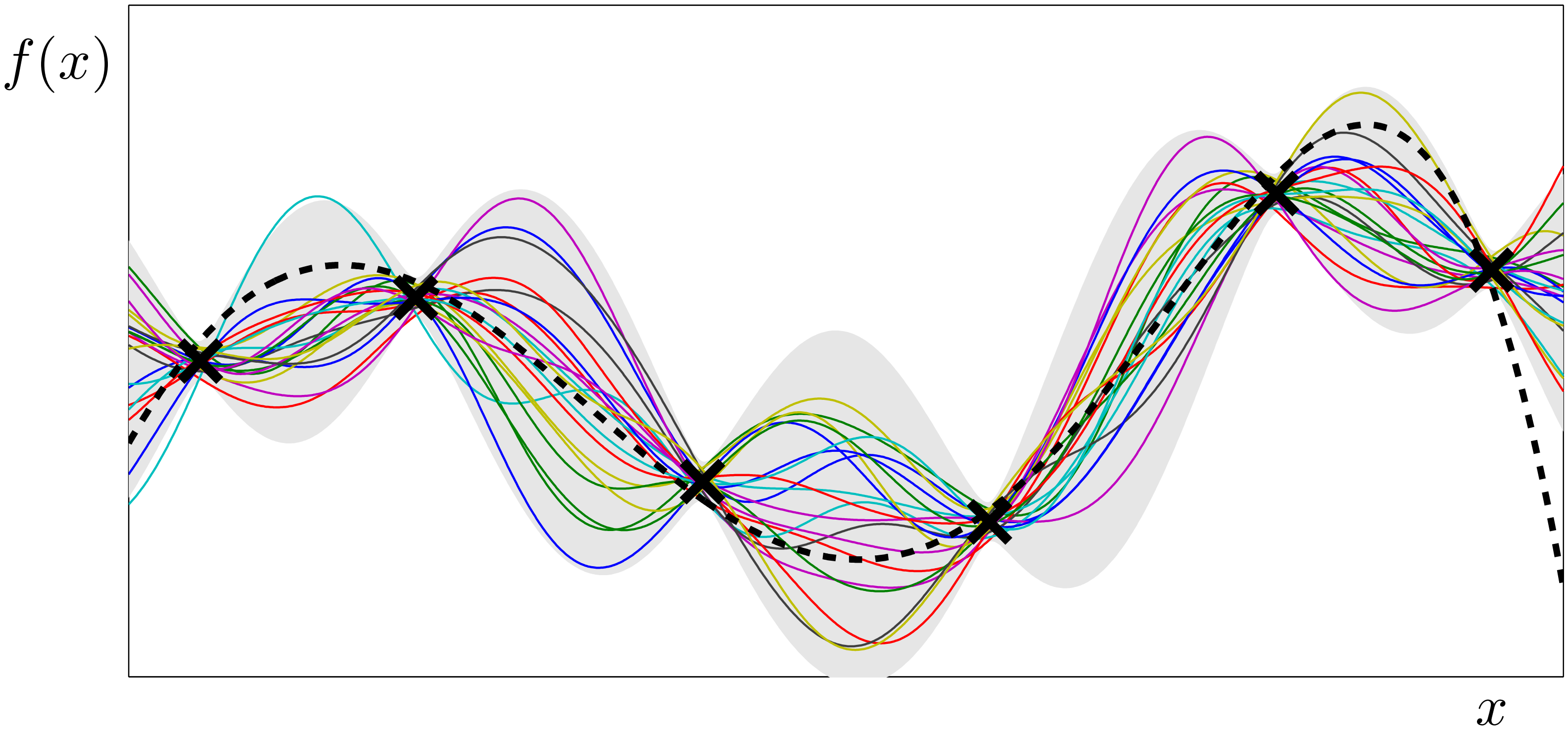
Posterior $\mathcal{GP}$ given observations



Bayesian Models for f e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: A distribution over functions from $\mathcal{X}$ to $\mathbb{R}$.

Posterior $\mathcal{GP}$ given observations



After t observations, $f(x) \sim \mathcal{N}(\, \mu_t(x), \, \sigma_t^2(x)\,)$.

Mathematical Foundations: Definition

Gaussian process = generalization of multivariate Gaussian distribution to infinitely many variables.

Definition: a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.

A Gaussian distribution is fully specified by a mean vector, $\boldsymbol{\mu}$, and covariance matrix Σ:

$$\mathbf{f} = (\mathrm{f}_1, \ldots, \mathrm{f}_n) \sim \mathcal{N}(\mu, \Sigma), \text{ indices } i = 1, \ldots, n$$

A Gaussian process is fully specified by a mean function $m(\mathbf{x})$ and covariance function $\mathrm{K}(\mathbf{x}, \mathbf{x}')$:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), \mathrm{K}(\mathbf{x}, \mathbf{x}')\right), \text{ indices } \mathbf{x}$$

Gaussian Processes – Noise free observations

- Model
  - (x, f) are the observed locations and values (training data)
  - (x*, f*) are the test or prediction data locations and values.

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

- After observing some noise free data (x, f),

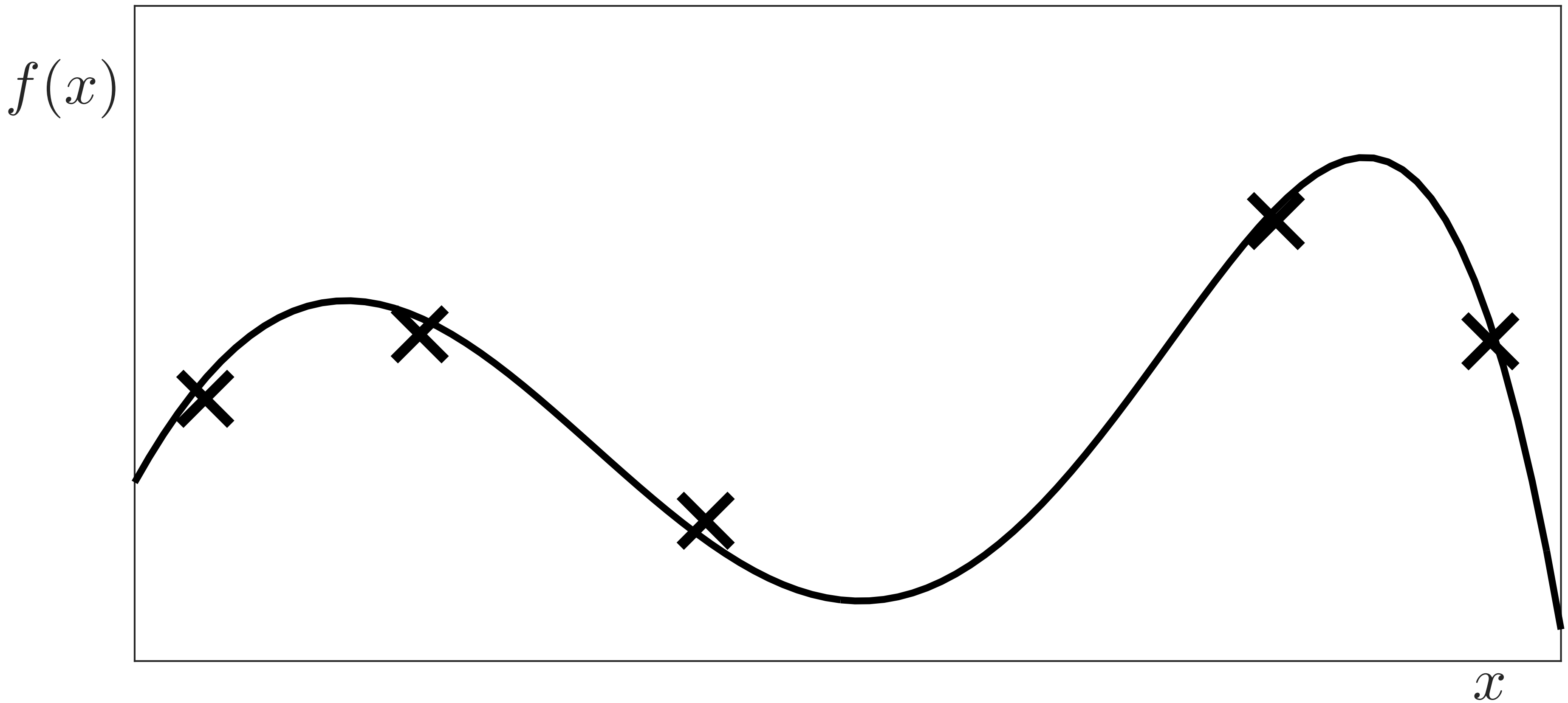
$$\mathbf{f}_* | X_*, X, \mathbf{f} \sim \mathcal{N}\big(K(X_*,X)K(X,X)^{-1}\mathbf{f}, \\ K(X_*,X_*) - K(X_*,X)K(X,X)^{-1}K(X,X_*)\big)$$

Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)

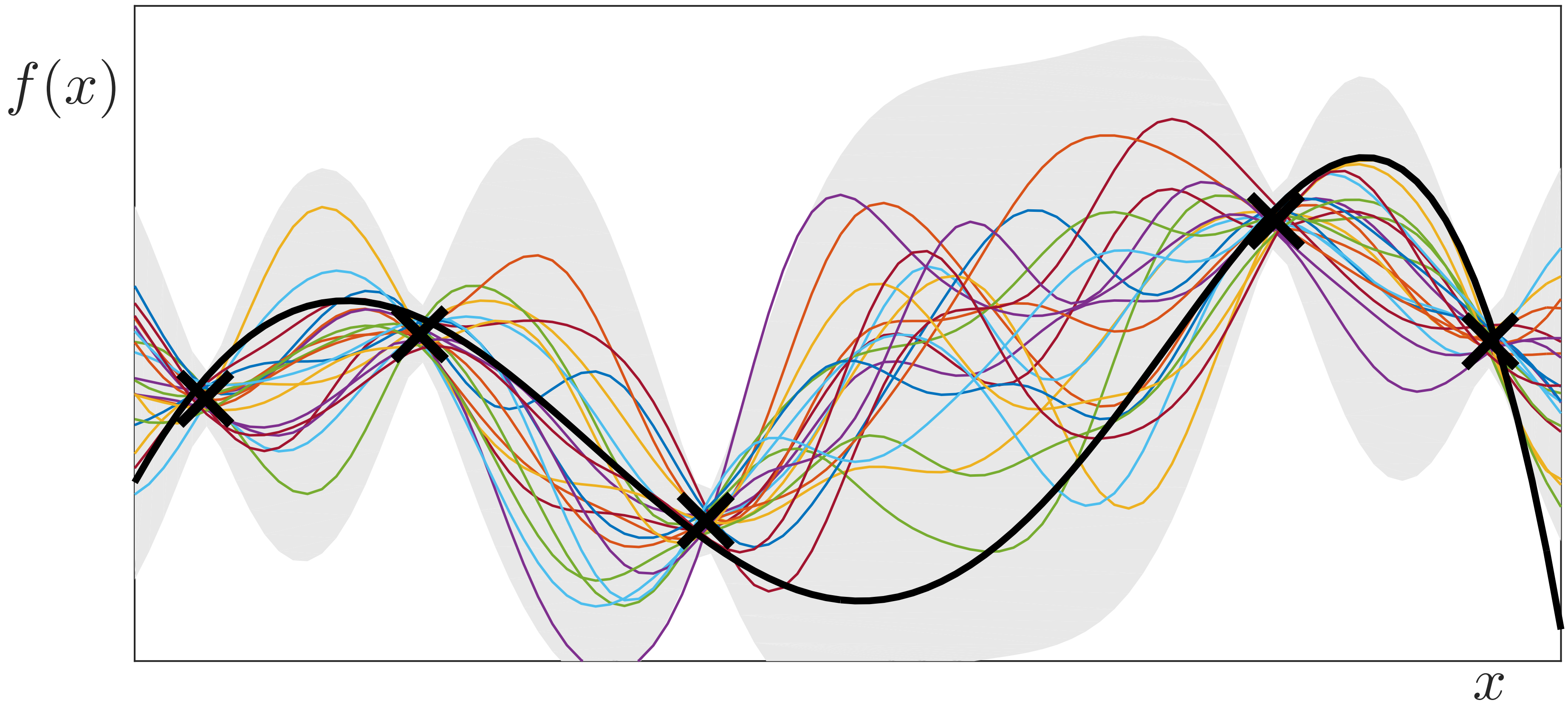


Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



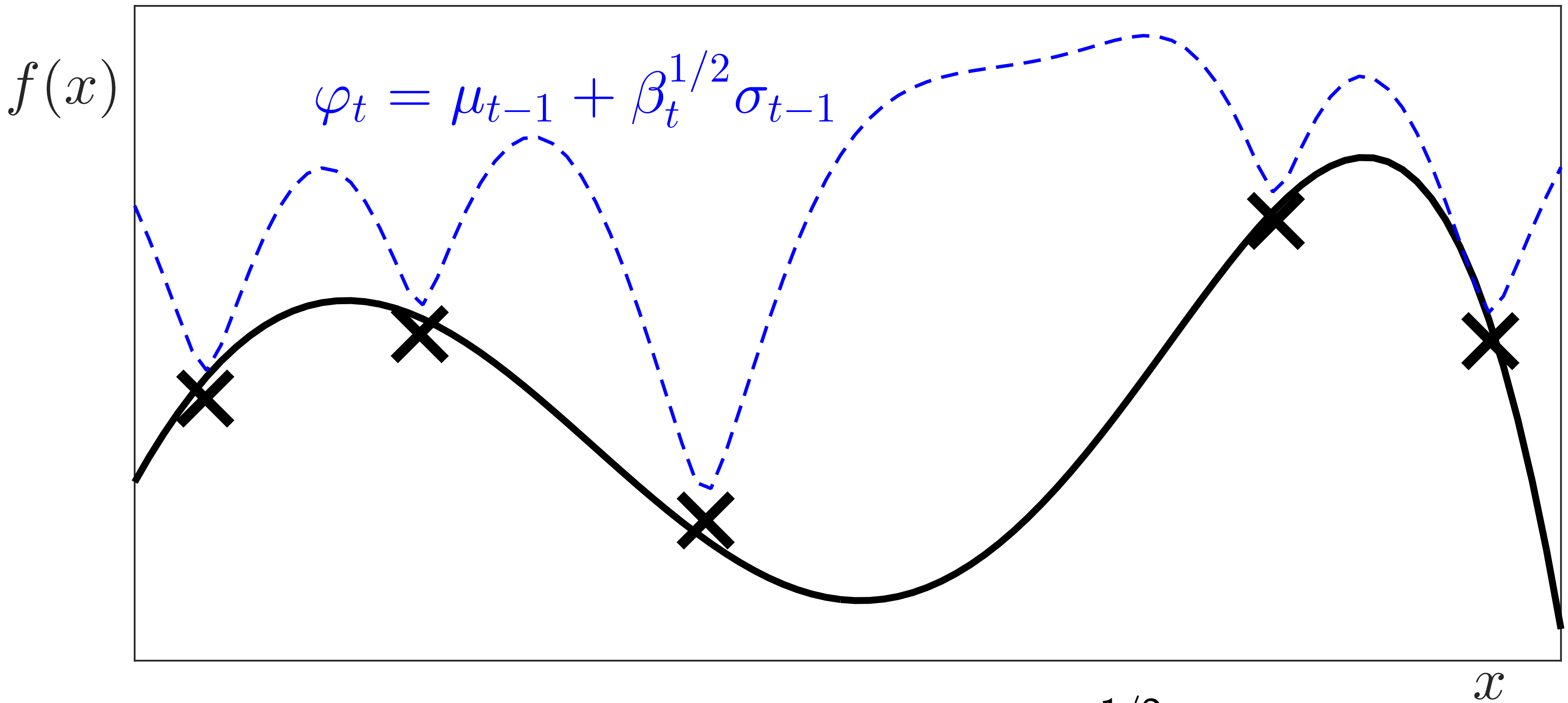
1) Construct posterior $\mathcal{GP}$.

Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



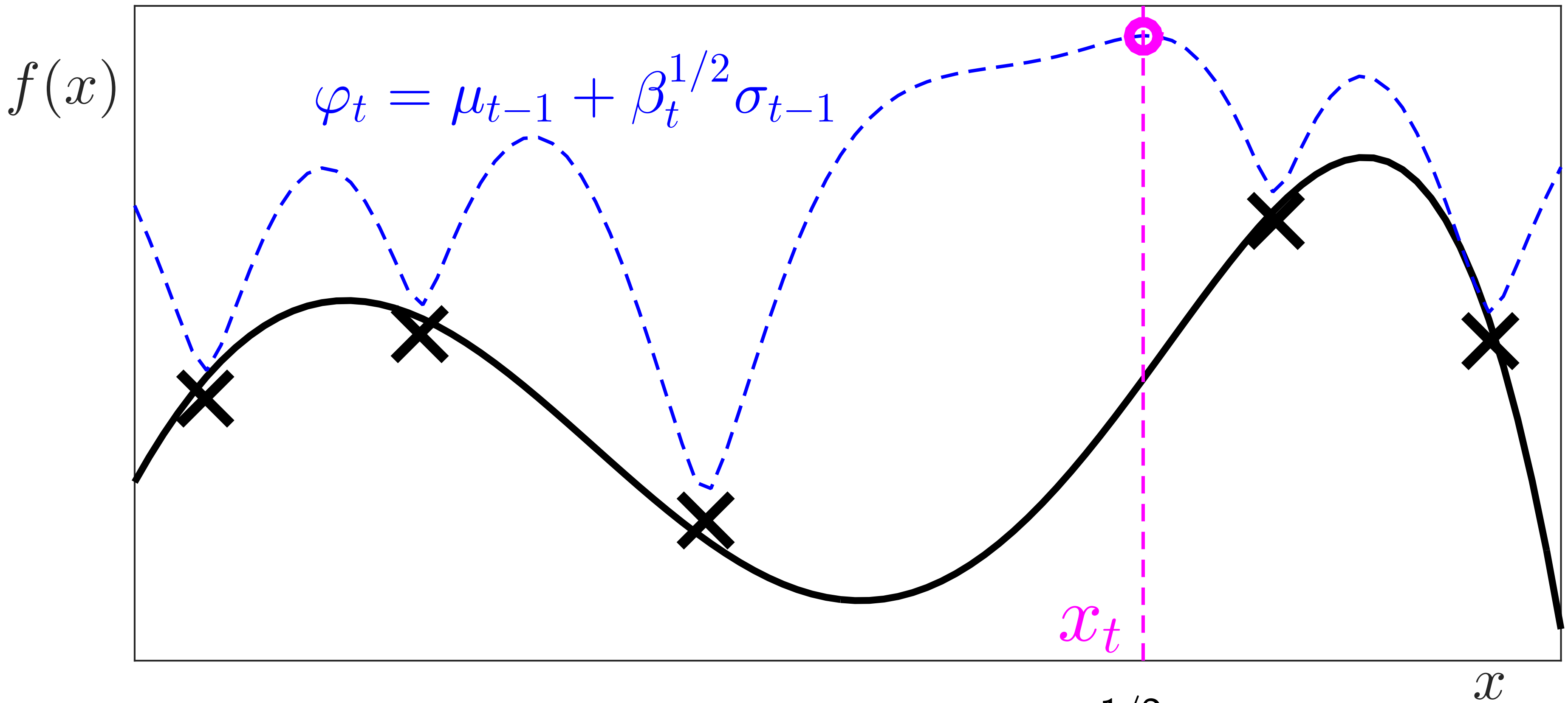
1) Construct posterior $\mathcal{GP}$. 2) $\varphi_t = \mu_{t-1} + \beta_t^{1/2}\sigma_{t-1}$ is a UCB.

Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



1) Construct posterior $\mathcal{GP}$. 2) $\varphi_t = \mu_{t-1} + \beta_t^{1/2}\sigma_{t-1}$ is a UCB.

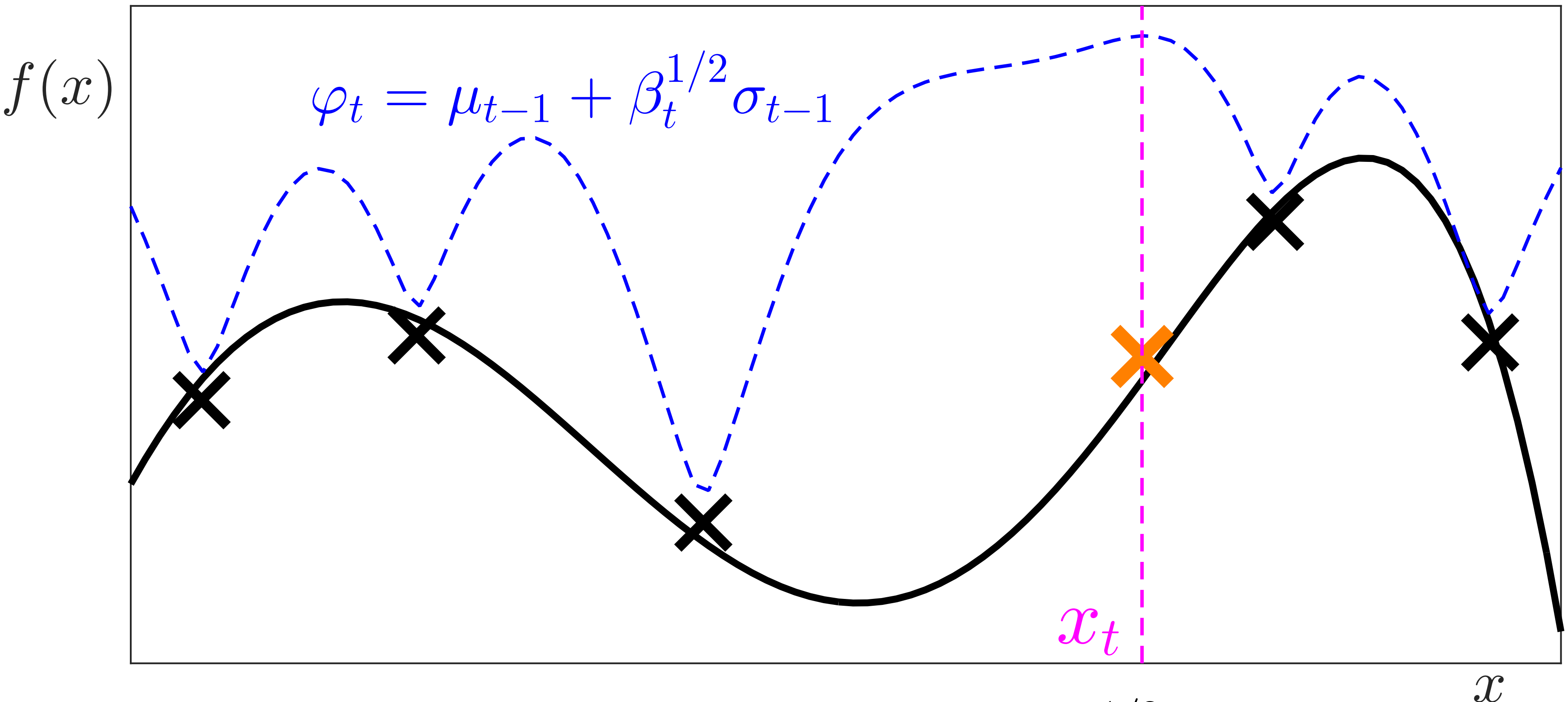
3) Choose $x_t = \operatorname{argmax}_x \varphi_t(x)$.

Bayesian Optimisation with Upper Confidence Bounds

Model $f \sim \mathcal{GP}$.

Gaussian Process Upper Confidence Bound (GP-UCB)

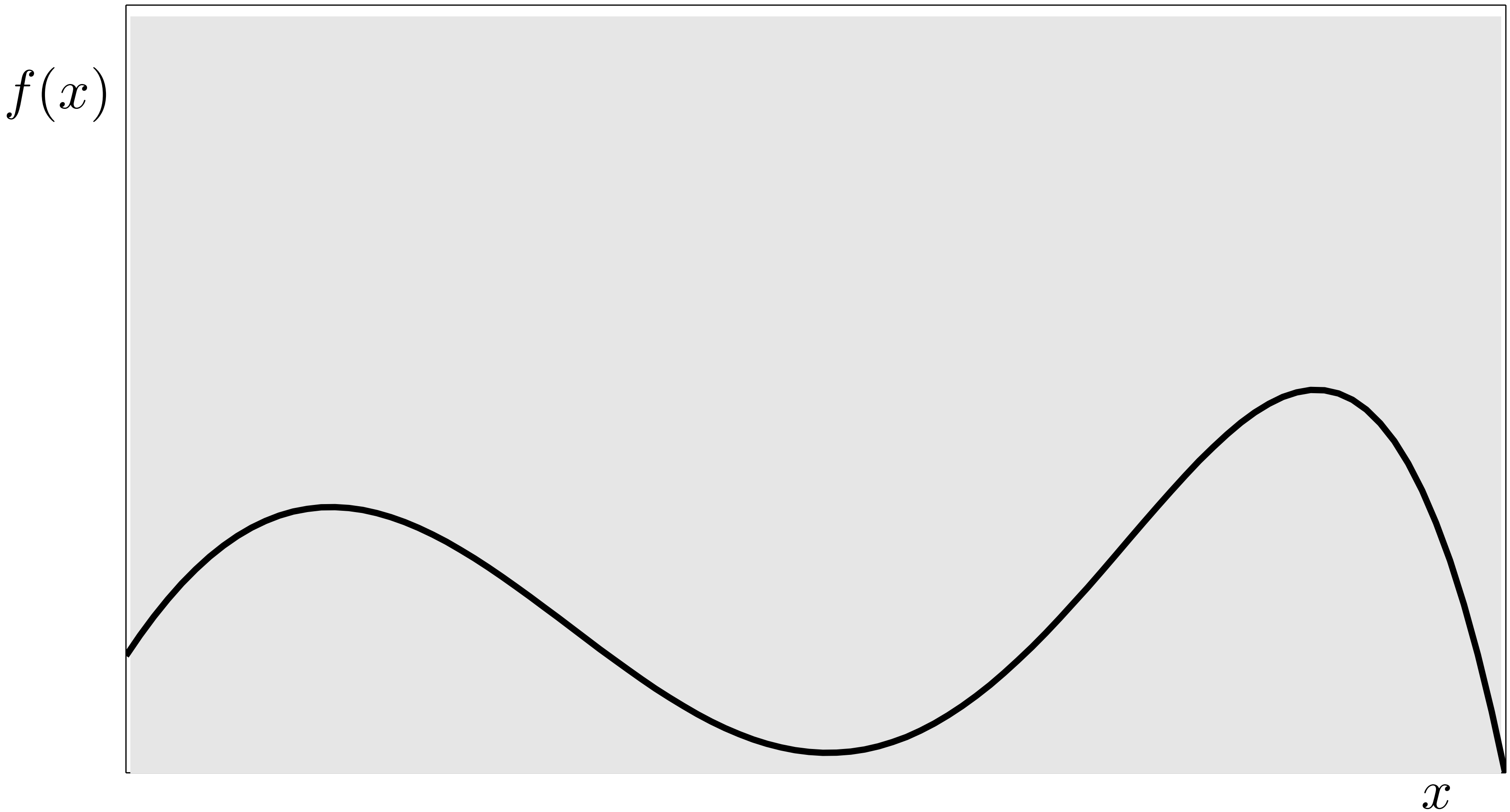
(Srinivas et al. 2010)



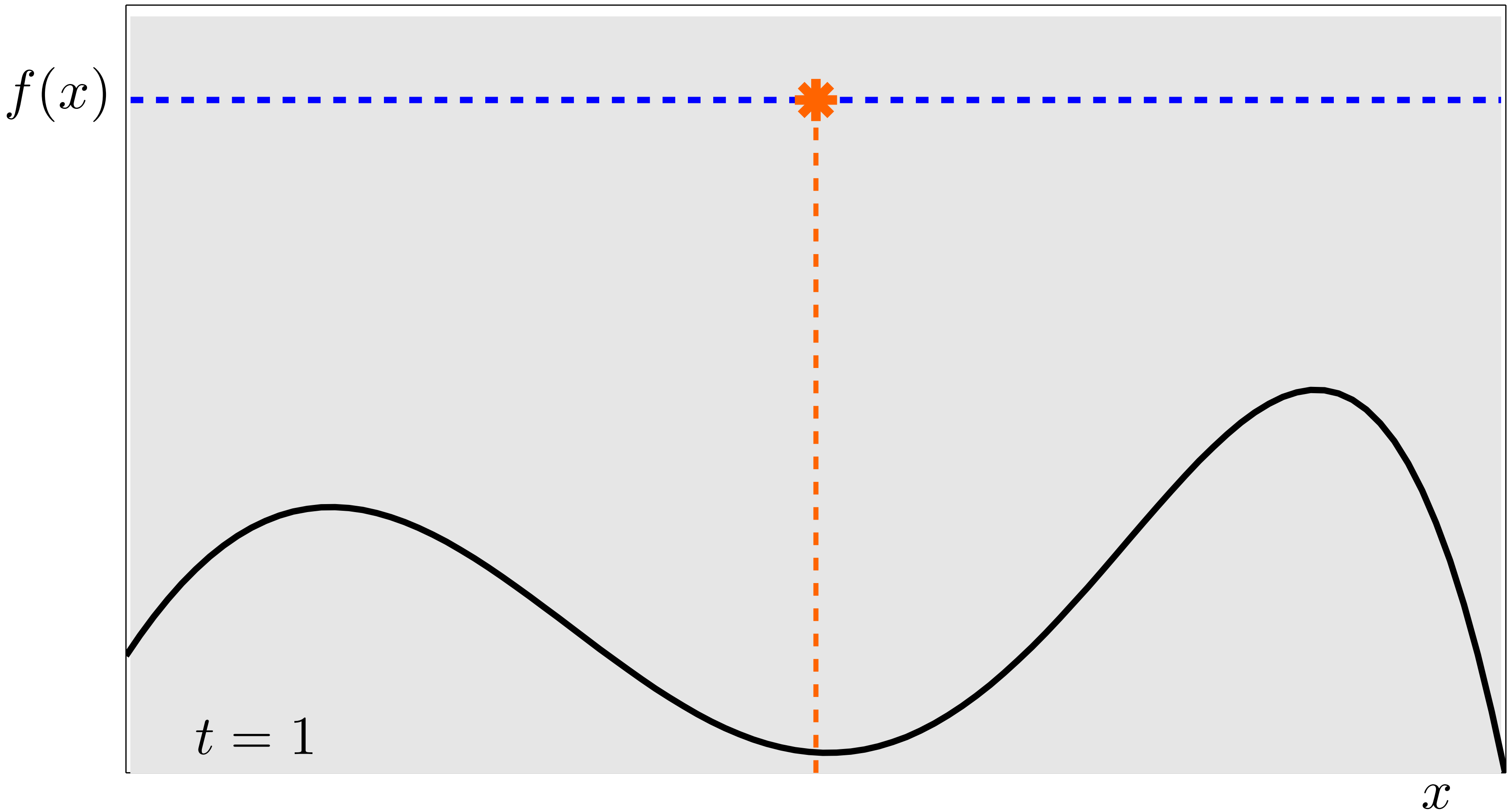
1) Construct posterior $\mathcal{GP}$. 2) $\varphi_t = \mu_{t-1} + \beta_t^{1/2}\sigma_{t-1}$ is a UCB.

3) Choose $x_t = \operatorname{argmax}_x \varphi_t(x)$. 4) Evaluate f at x_t.

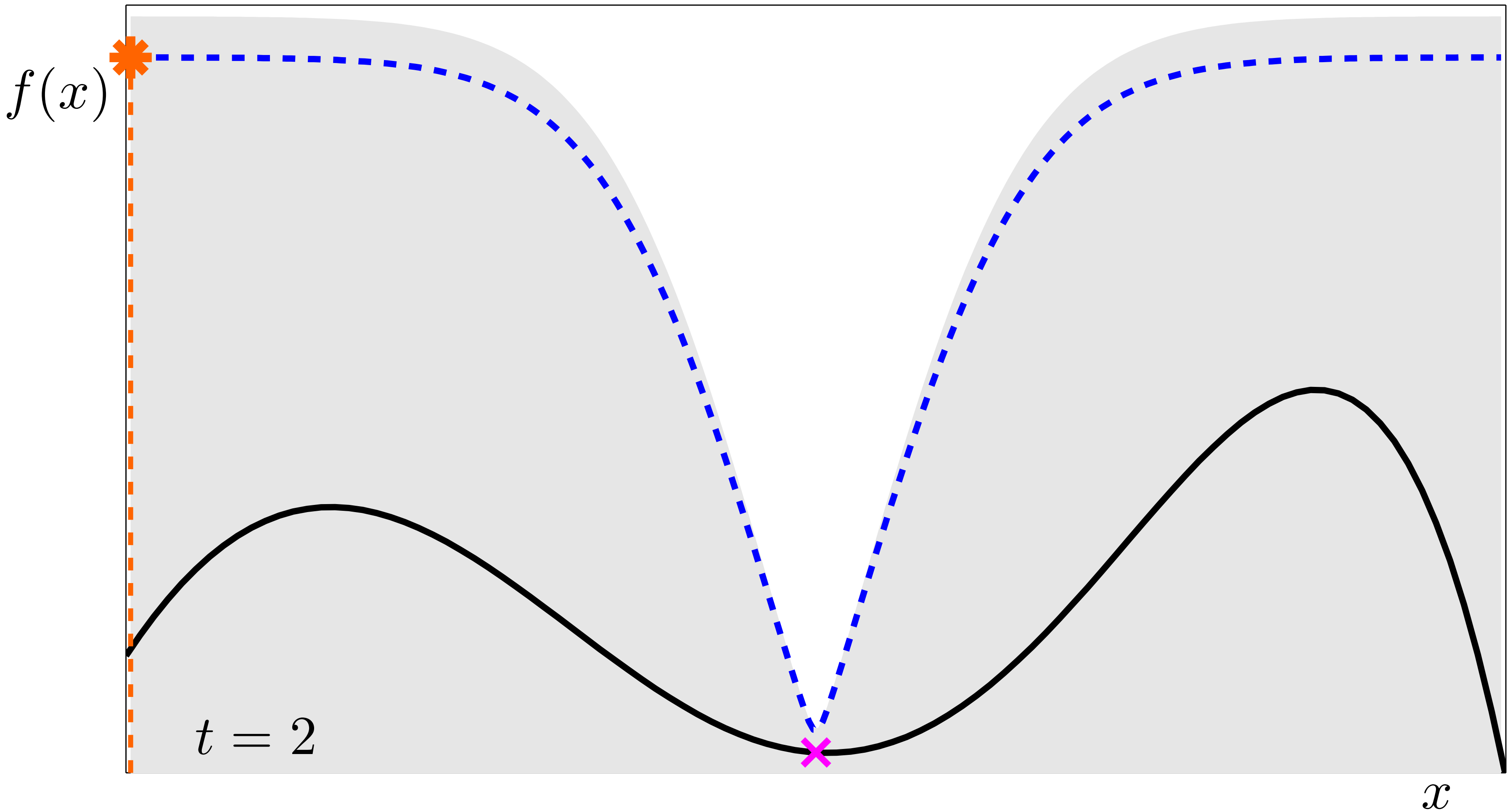
GP-UCB (Srinivas et al. 2010)



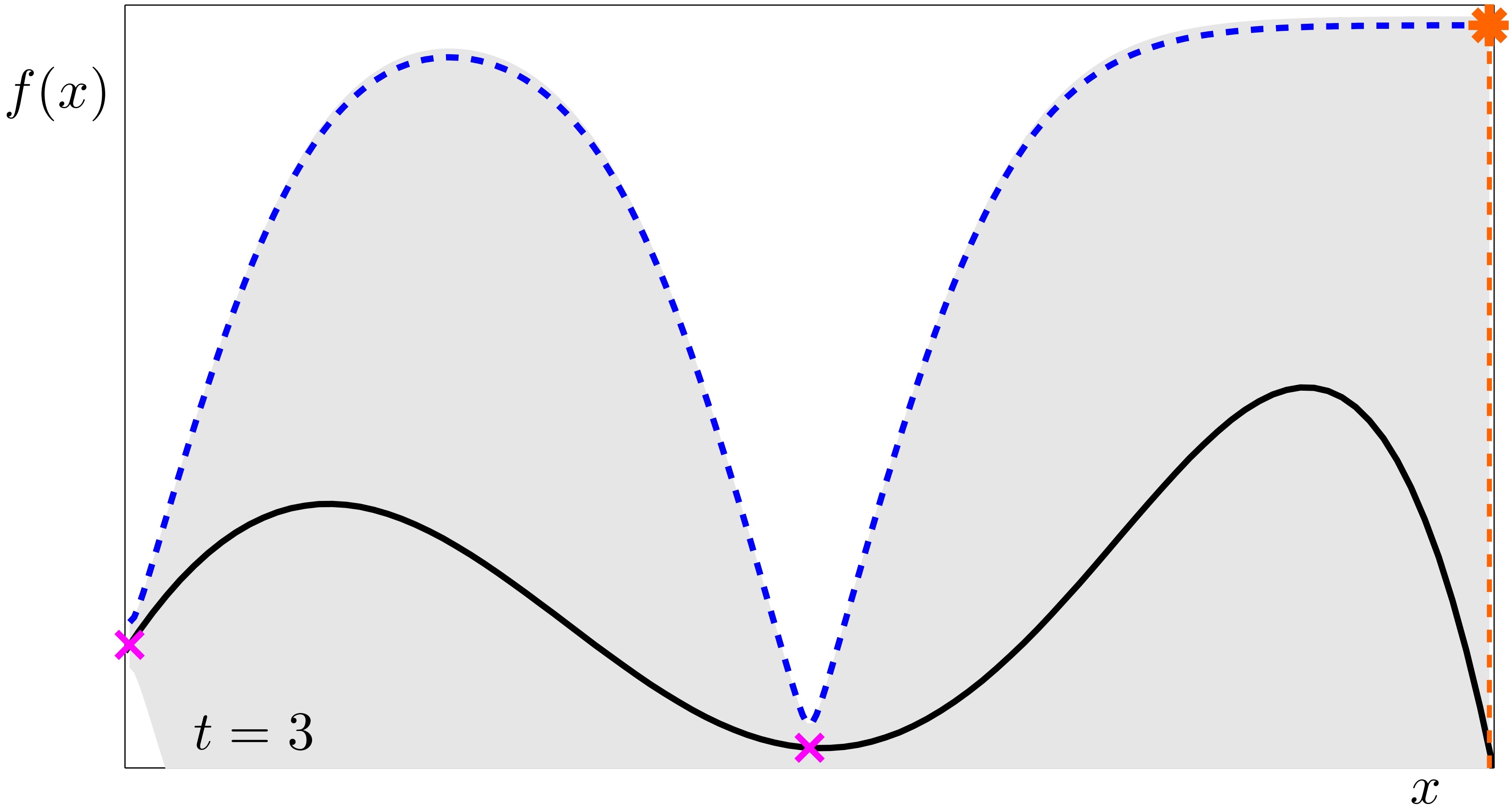
GP-UCB (Srinivas et al. 2010)



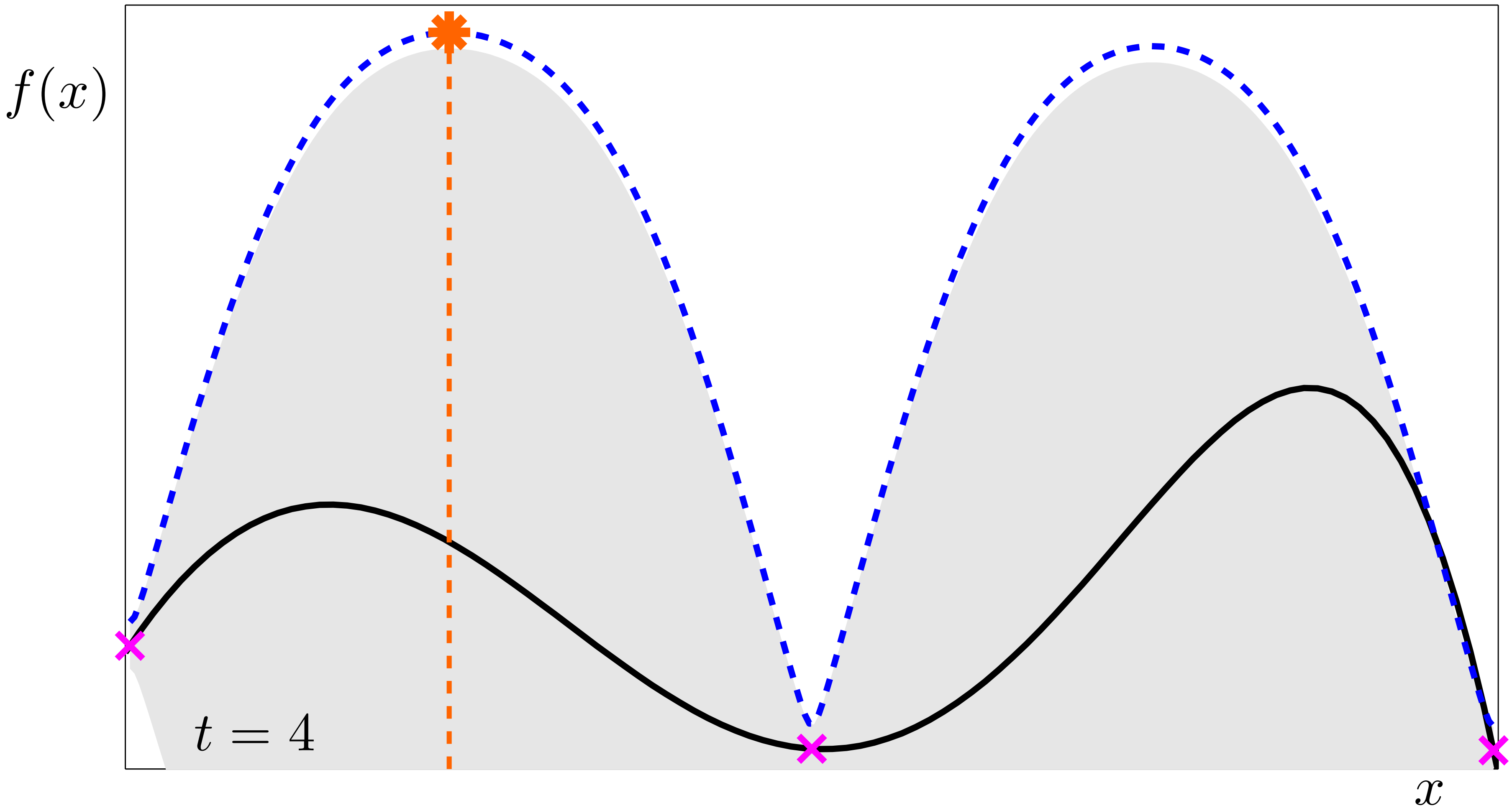
GP-UCB (Srinivas et al. 2010)



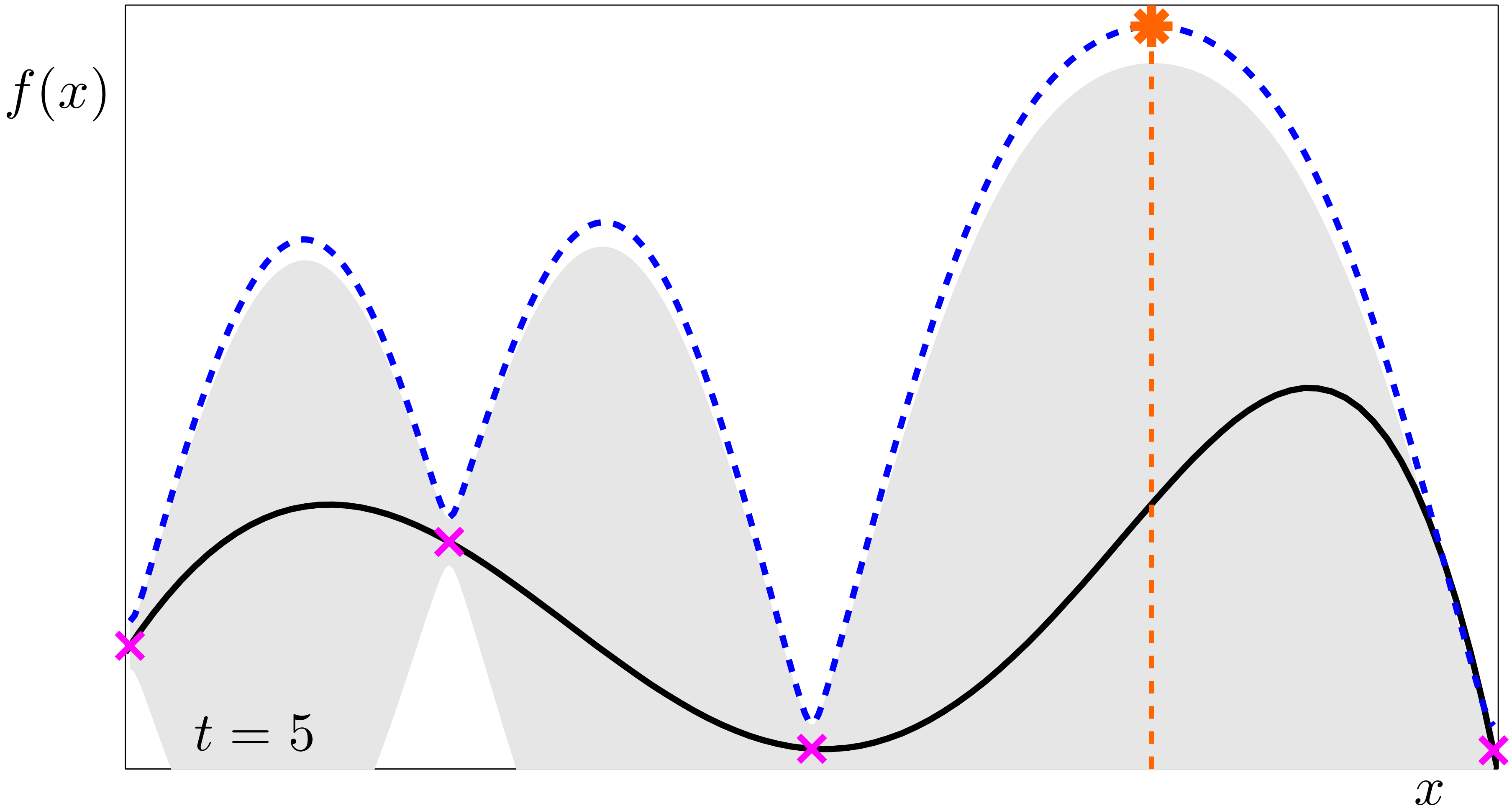
GP-UCB (Srinivas et al. 2010)



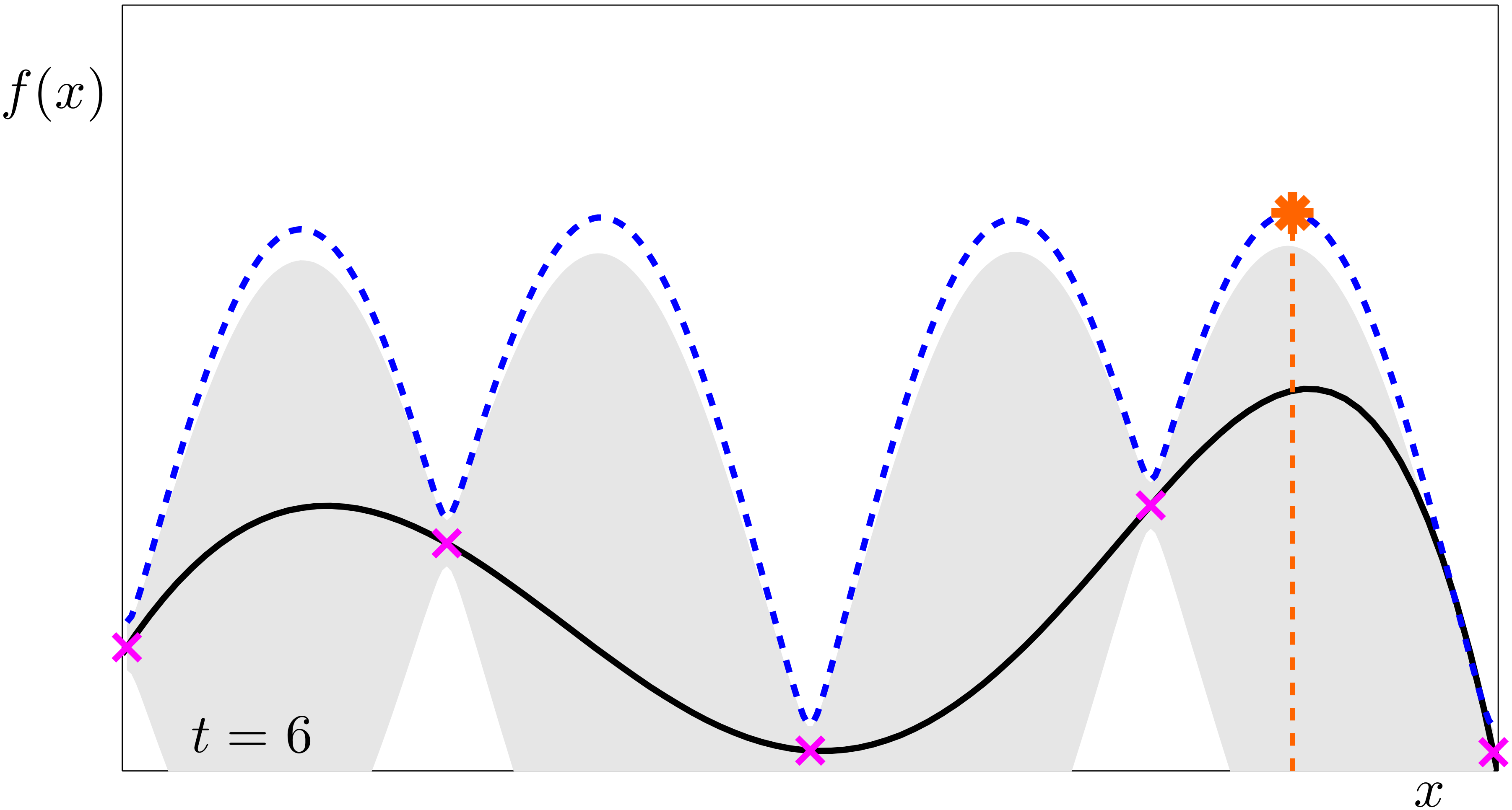
GP-UCB (Srinivas et al. 2010)



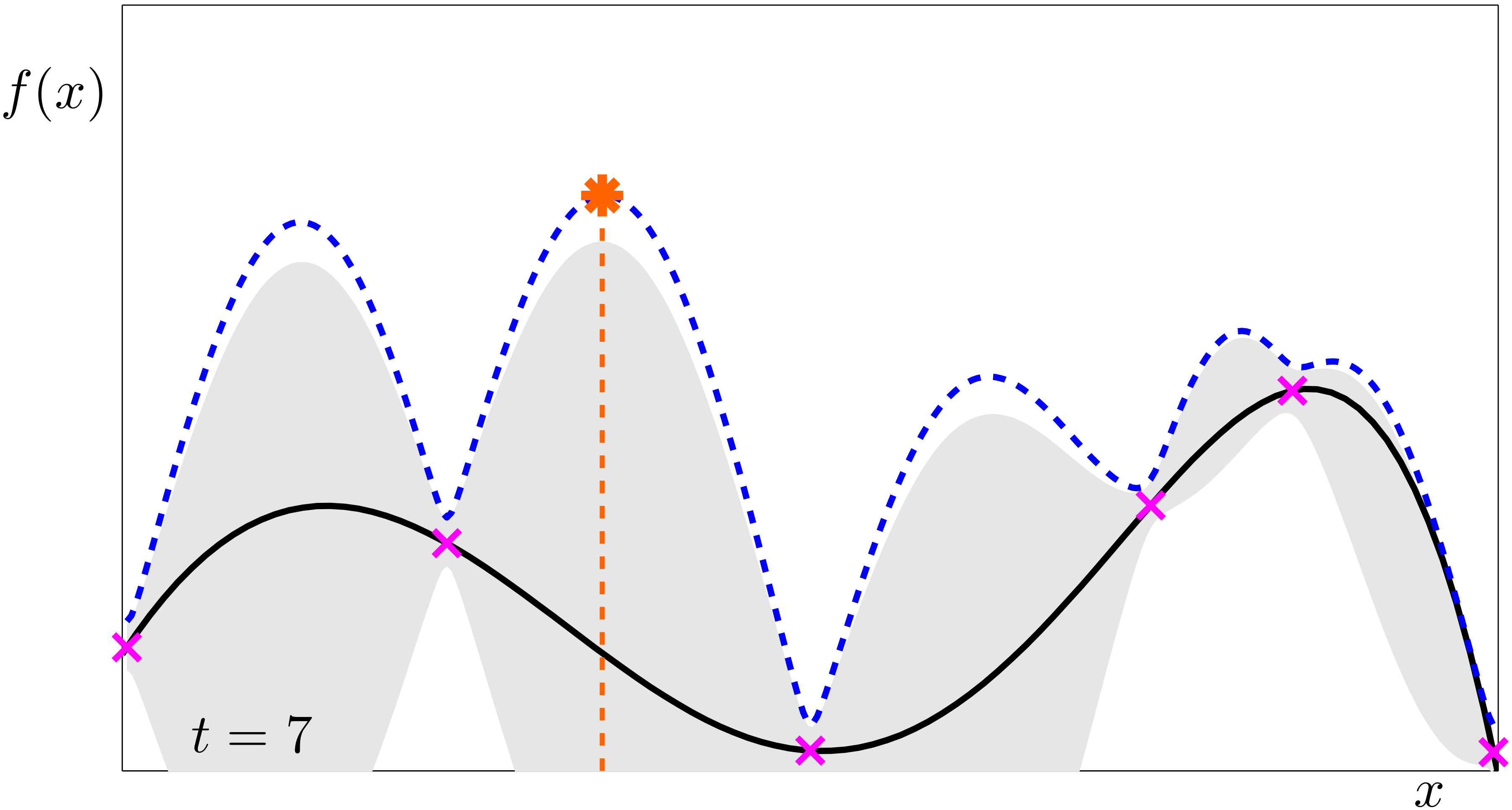
GP-UCB (Srinivas et al. 2010)



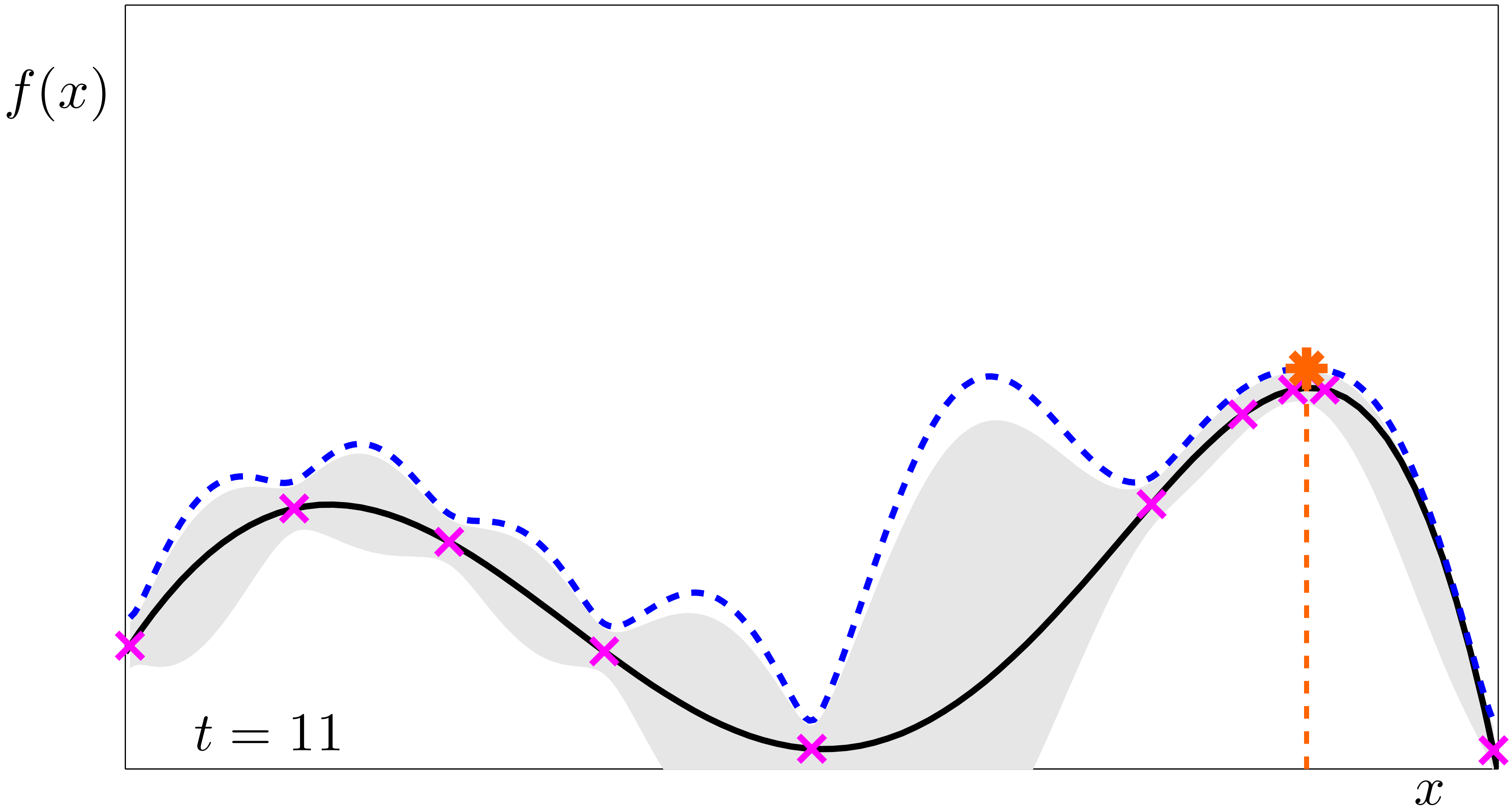
GP-UCB (Srinivas et al. 2010)



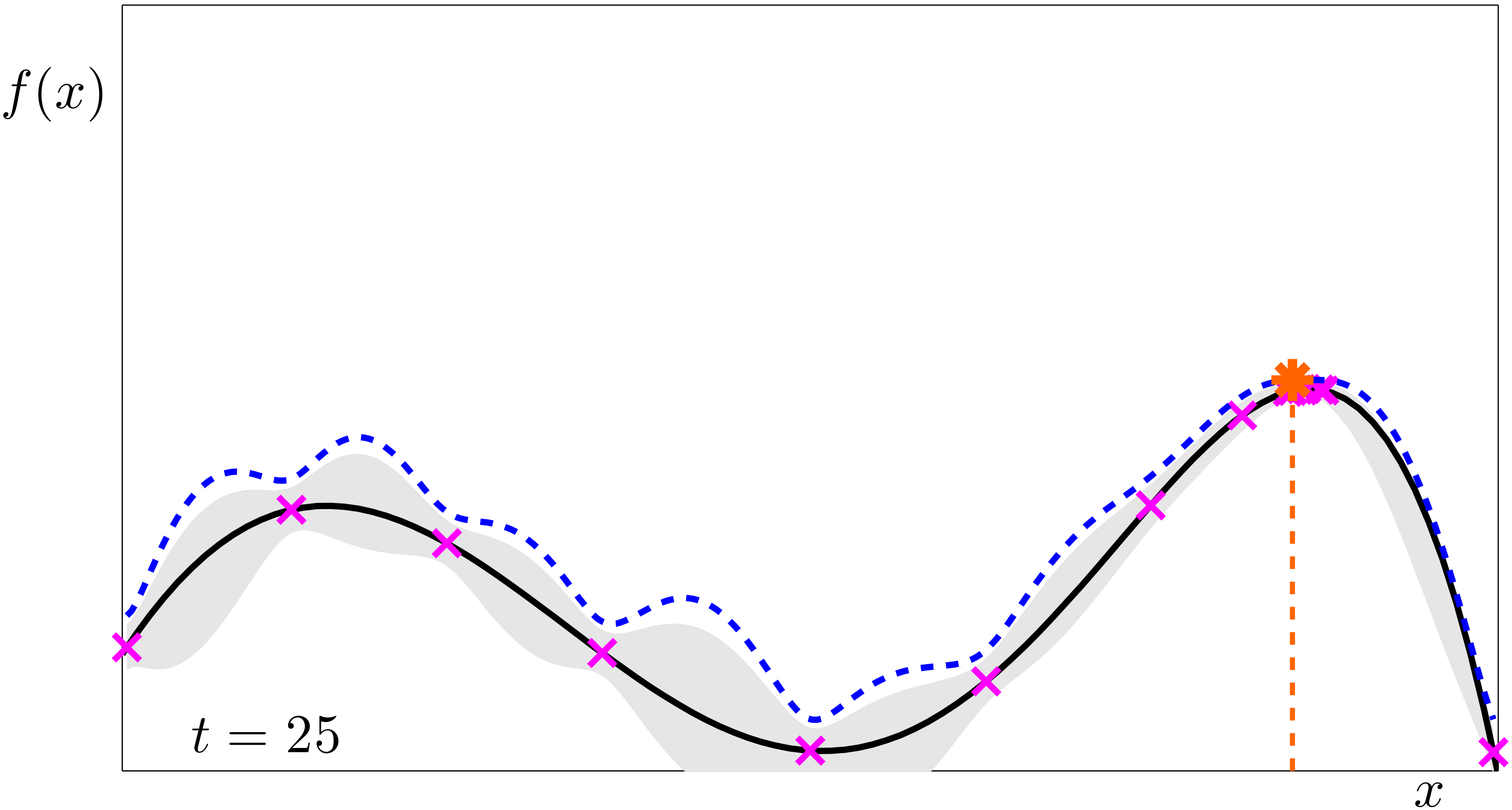
GP-UCB (Srinivas et al. 2010)



GP-UCB (Srinivas et al. 2010)



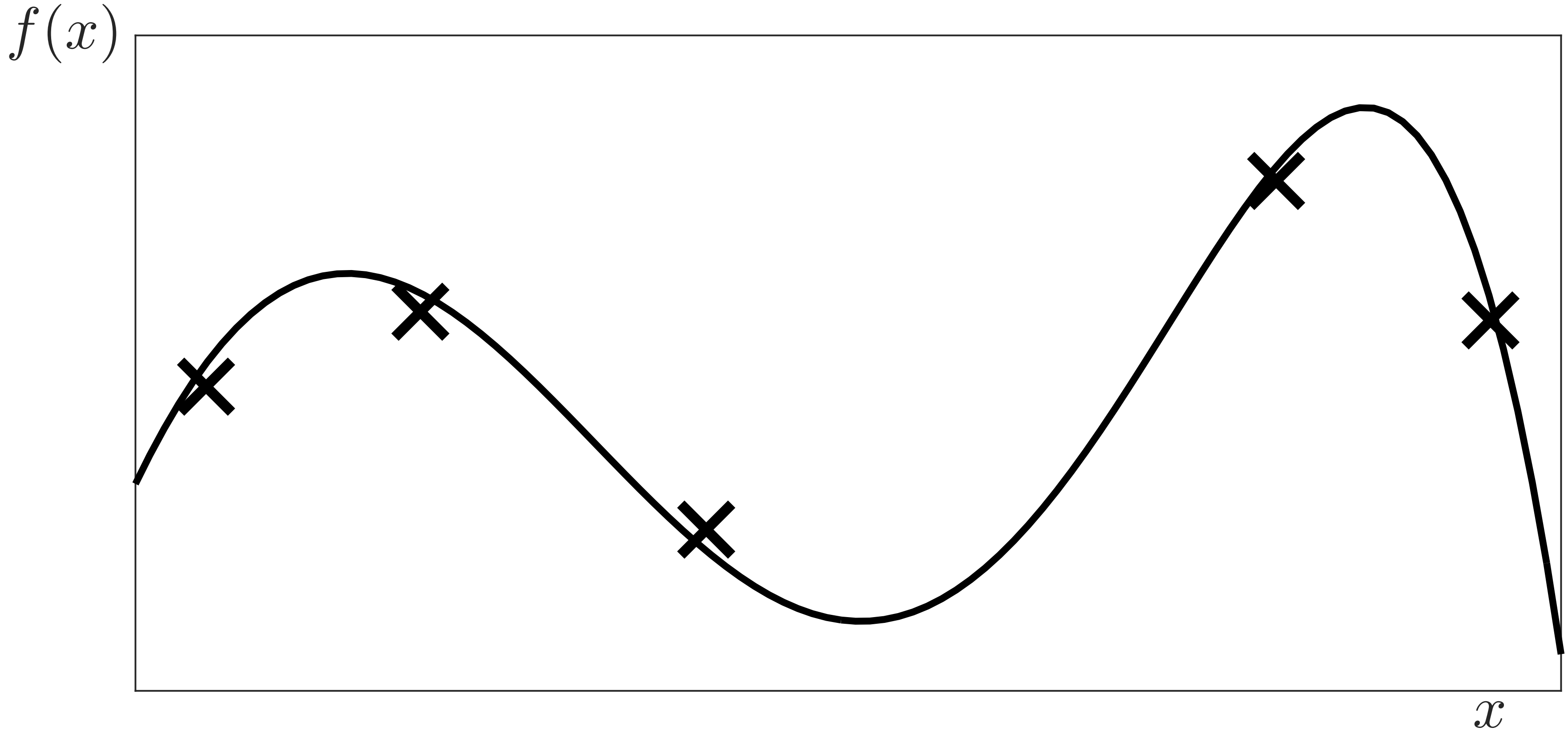
GP-UCB (Srinivas et al. 2010)



Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

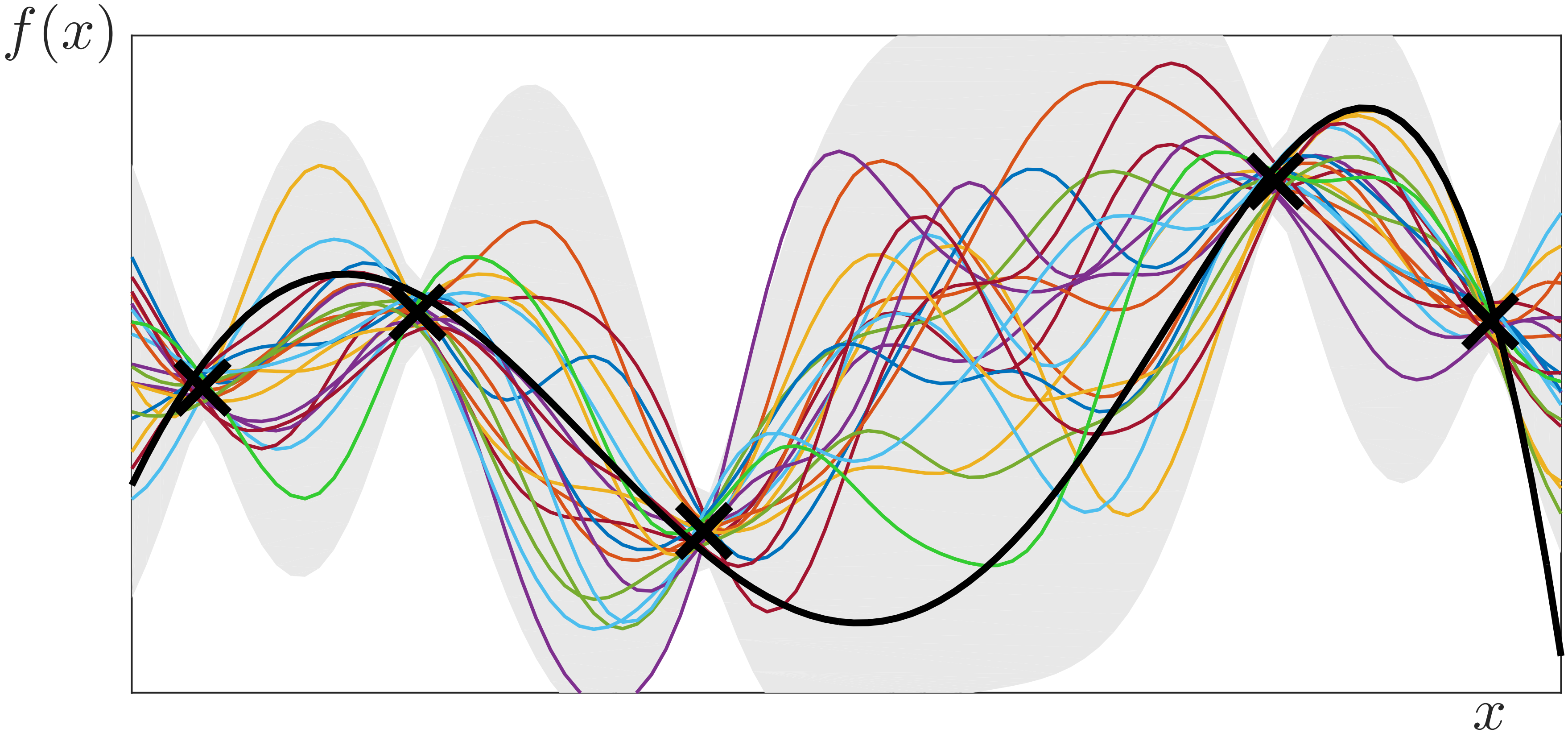
Thompson Sampling (TS) (Thompson, 1933).



Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

Thompson Sampling (TS) (Thompson, 1933).

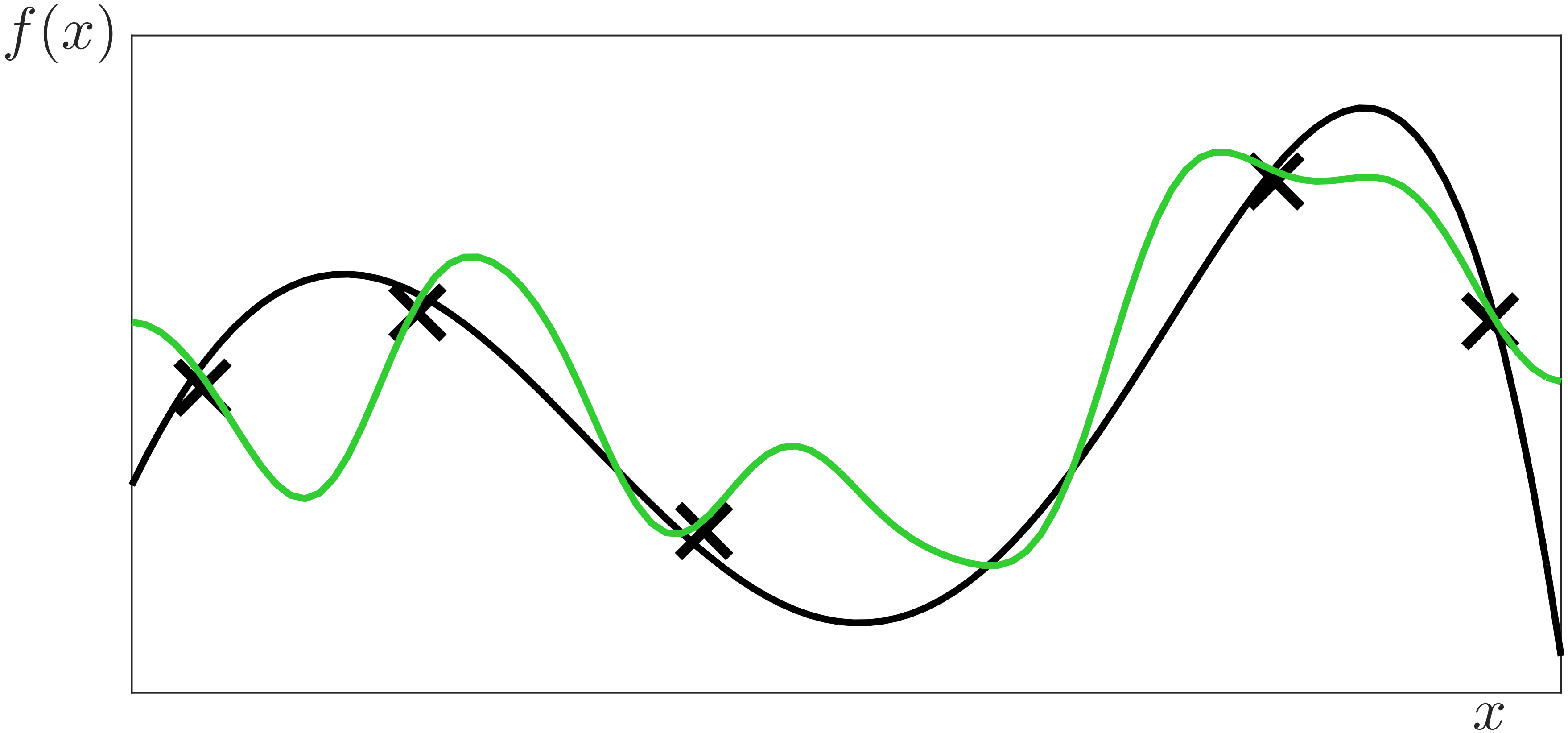


1) Construct posterior $\mathcal{GP}$.

Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

Thompson Sampling (TS) (Thompson, 1933).

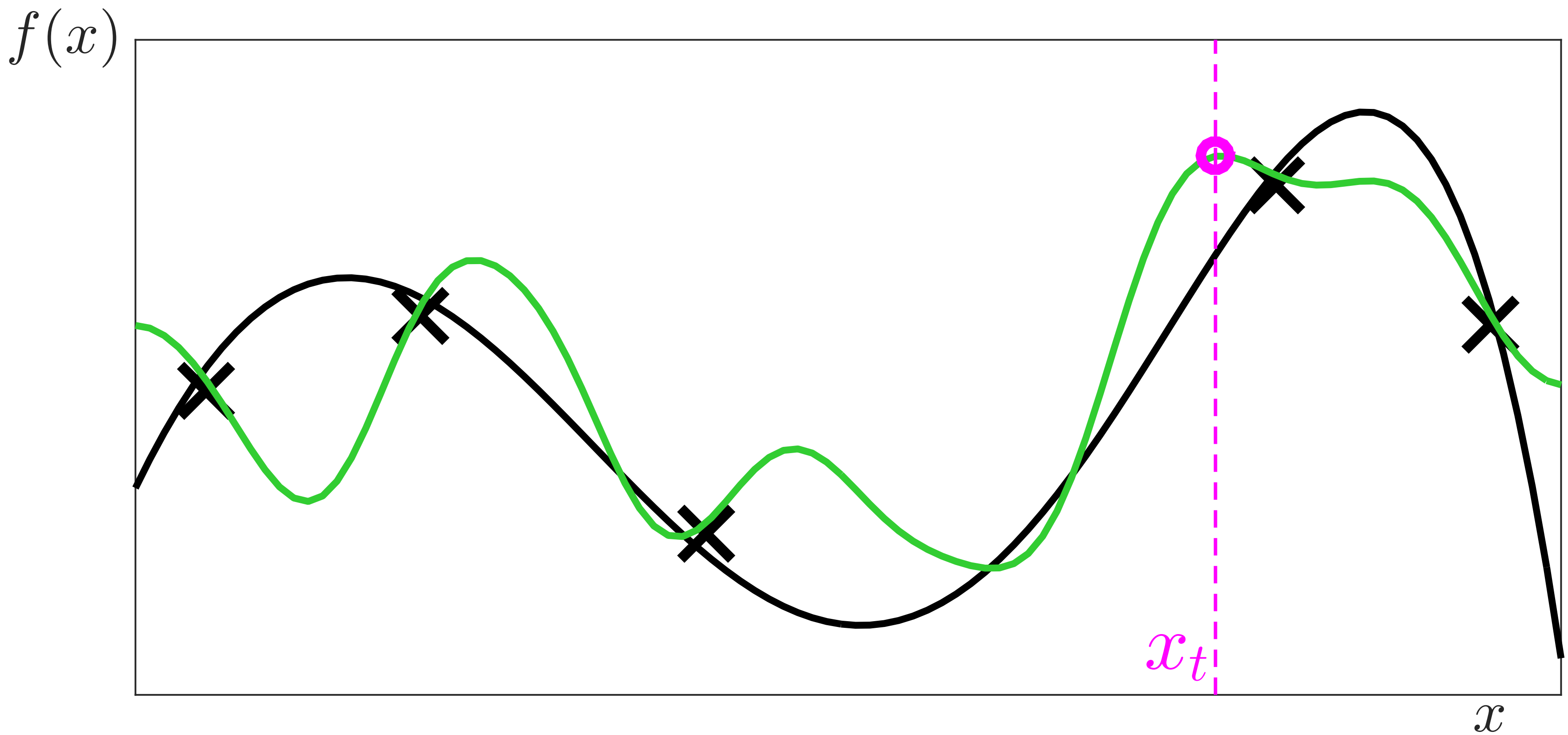


1) Construct posterior $\mathcal{GP}$. 2) Draw sample g from posterior.

Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

Thompson Sampling (TS) (Thompson, 1933).



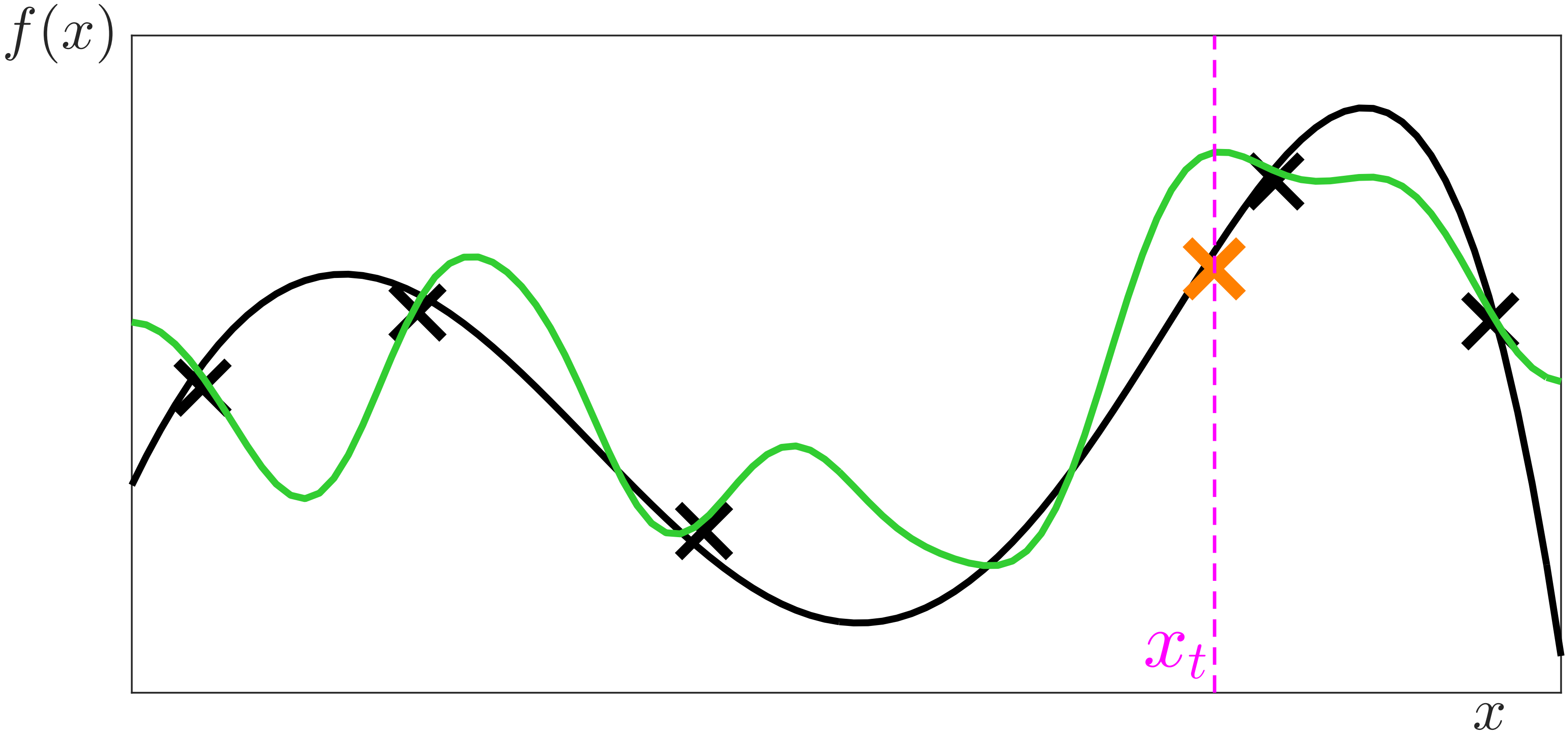
1) Construct posterior $\mathcal{GP}$. 2) Draw sample g from posterior.

3) Choose $x_t = \mathrm{argmax}_x\, g(x)$.

Bayesian Optimisation with Thompson Sampling

Model $f \sim \mathcal{GP}(\mathbf{0}, \kappa)$.

Thompson Sampling (TS) (Thompson, 1933).



1) Construct posterior $\mathcal{GP}$. 2) Draw sample g from posterior.

3) Choose $x_t = \text{argmax}_x\, g(x)$. 4) Evaluate f at x_t.

Reference for more detailed tutorial

- "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning" (https://arxiv.org/pdf/1012.2599.pdf)

Slides from

- Kirthevasan Kandasamy’s talk on “An Introduction to Bayesian Optimisation and (Potential) Applications in Materials Science” (https://people.eecs.berkeley.edu/~kandasamy/talks/electrochem-bo-slides.pdf)

- University of Washington course stat527 recitation Slides 2 (https://sites.stat.washington.edu/courses/stat527/s14/recitation/Slides2.pptx)

- CMU S19 10-403 slides on “Bayesian Optimization - Gaussian Processes” (https://www.andrew.cmu.edu/course/10-403/slides/S19GaussianProcesses.pdf)