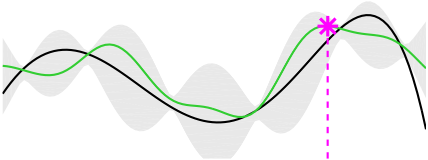
An Introduction to Bayesian Optimisation and (Potential) Applications in Materials Science



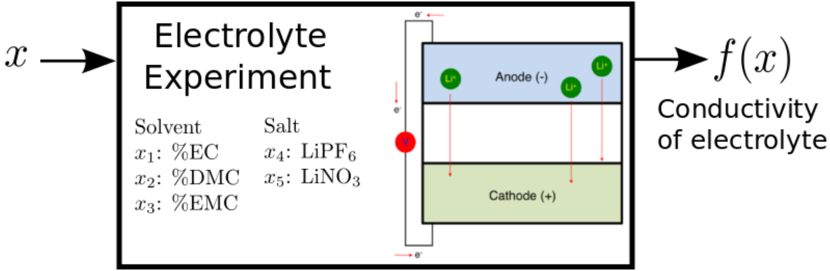
Kirthevasan Kandasamy

Machine Learning Dept, CMU

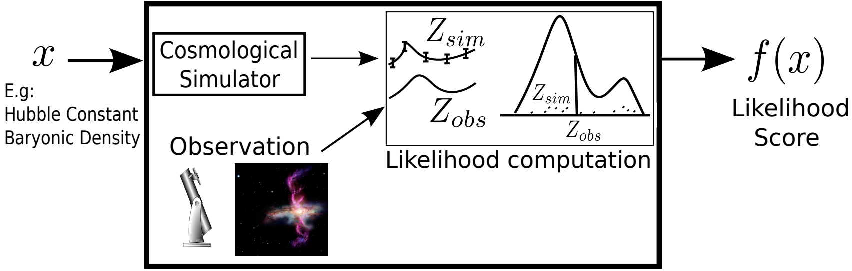
Electrochemical Energy Symposium

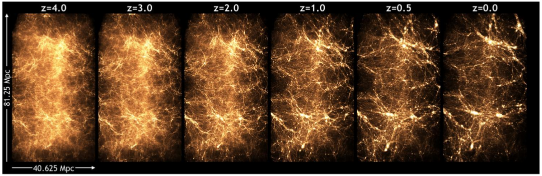
Pittsburgh, PA, November 2017

Designing Electrolytes in Batteries



Black-box Optimisation in Computational Astrophysics

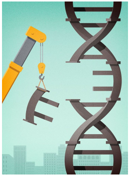




Black-box Optimisation

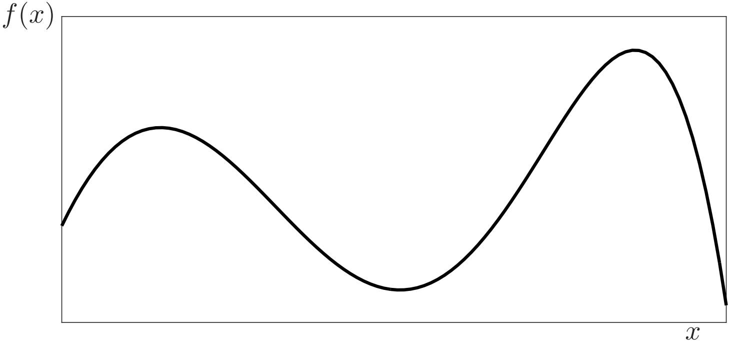
$x \longrightarrow$ **Expensive Blackbox Function** $\longrightarrow f(x)$

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



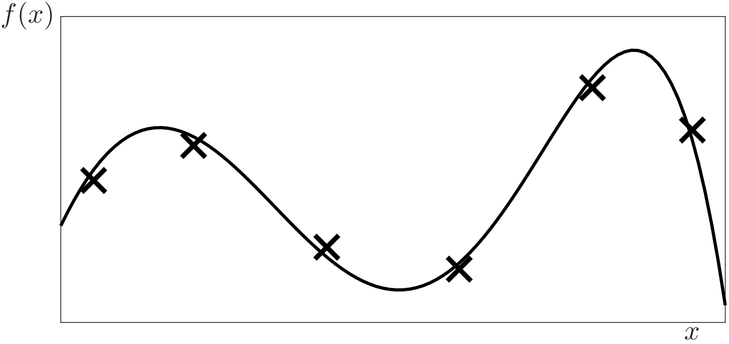
Black-box Optimisation

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



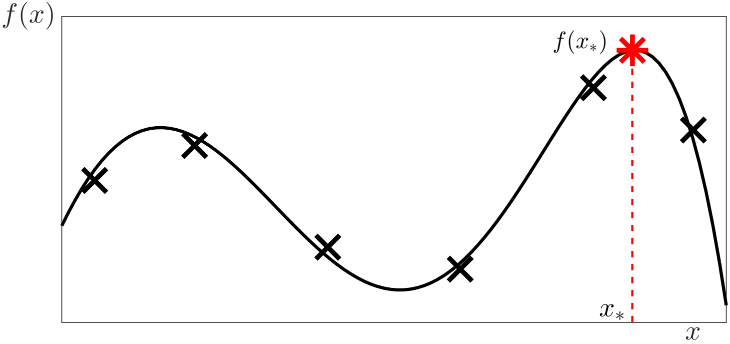
Black-box Optimisation

$f : \mathcal{X} \to \mathbb{R}$ 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 $x_\star = \mathrm{argmax}_x \, f(x)$.



Outline

- Part I: Bayesian Optimisation
  - Bayesian Models for f
  - Two algorithms: upper confidence bounds & Thompson sampling
- Part II: Some Modern Challenges
  - Multi-fidelity Optimisation
  - Parallelisation

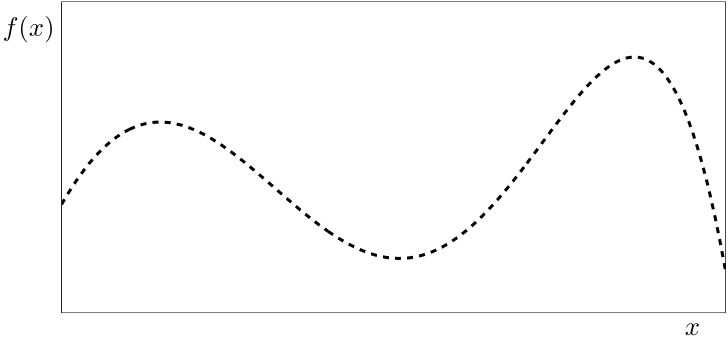
Bayesian Models for f e.g. Gaussian Processes ($\mathcal{GP}$)

$\mathcal{GP}$: 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 $\mathbb{R}$.

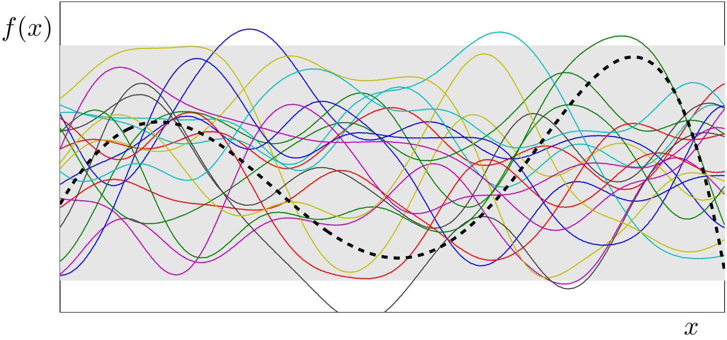
Functions with no observations



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

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

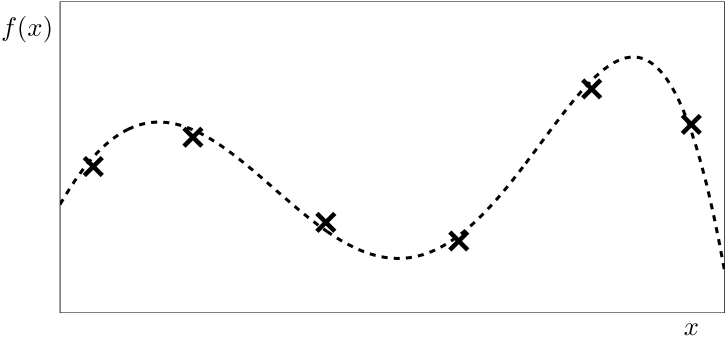
Prior $\mathcal{GP}$



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

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

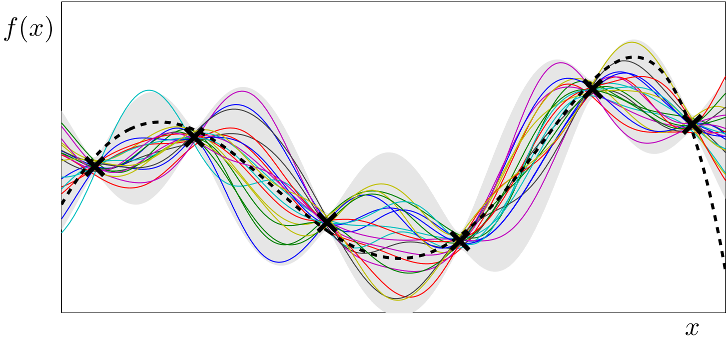
Observations



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

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

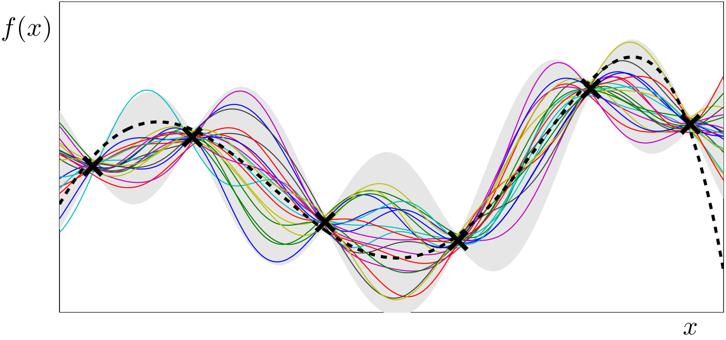
Posterior $\mathcal{GP}$ given observations



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



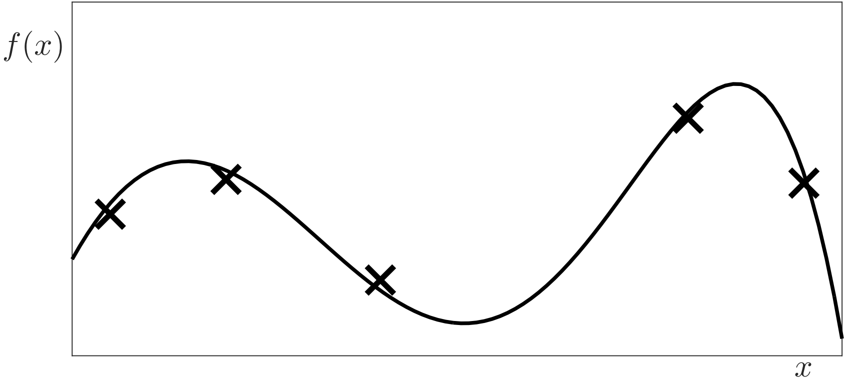
After t observations, $f(x) \sim \mathcal{N}(\,\mu_t(x),\, \sigma_t^2(x)\,)$.

Bayesian Optimisation with Upper Confidence Bounds

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

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)

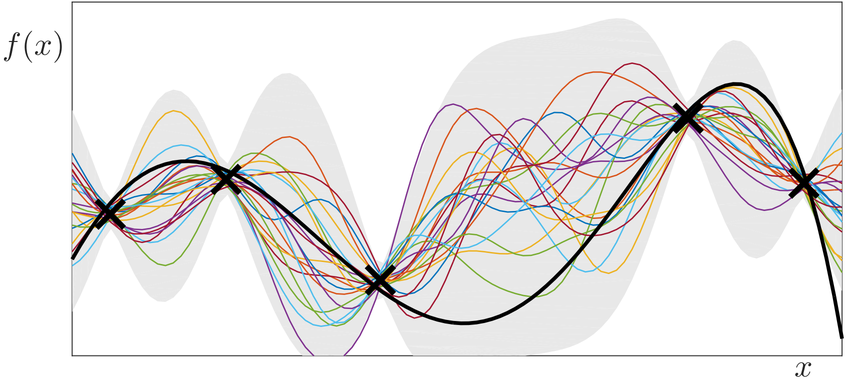


Bayesian Optimisation with Upper Confidence Bounds

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

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



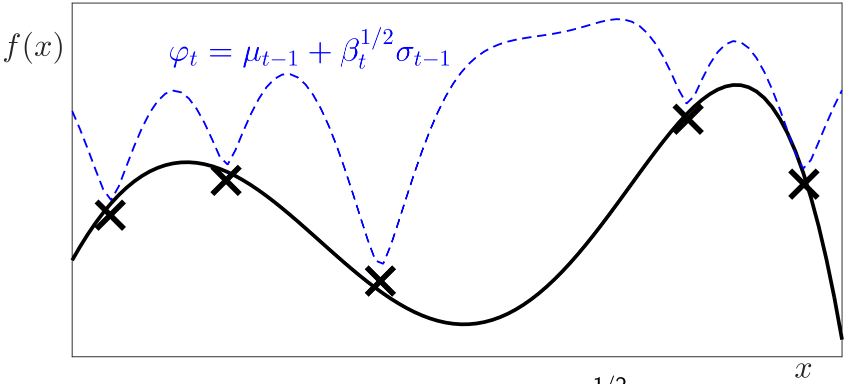
1) Construct posterior $\mathcal{GP}$.

Bayesian Optimisation with Upper Confidence Bounds

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

Gaussian Process Upper Confidence Bound (GP-UCB)

(Srinivas et al. 2010)



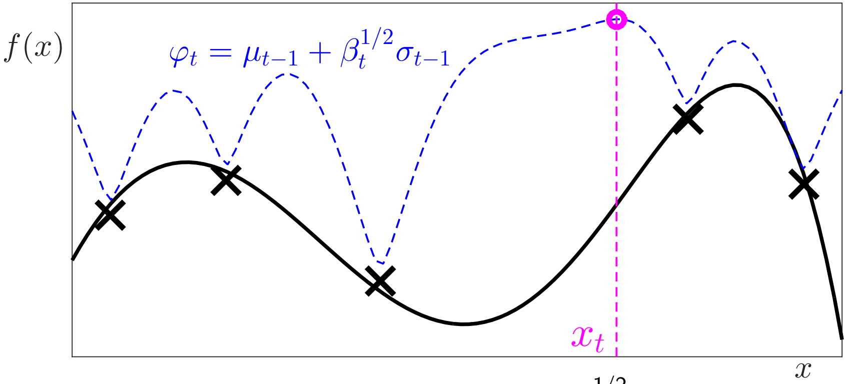
1) Construct posterior $\mathcal{GP}$. 2) $\varphi_t = \mu_{t-1} + \beta_t^{1/2}\sigma_{t-1}$ is a UCB.

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.

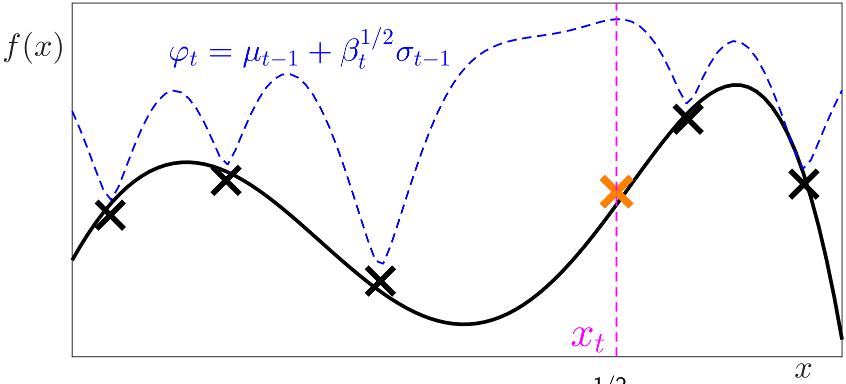
3) Choose $x_t = \mathrm{argmax}_x\, \varphi_t(x)$.

Bayesian Optimisation with Upper Confidence Bounds

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

Gaussian Process Upper Confidence Bound (GP-UCB)

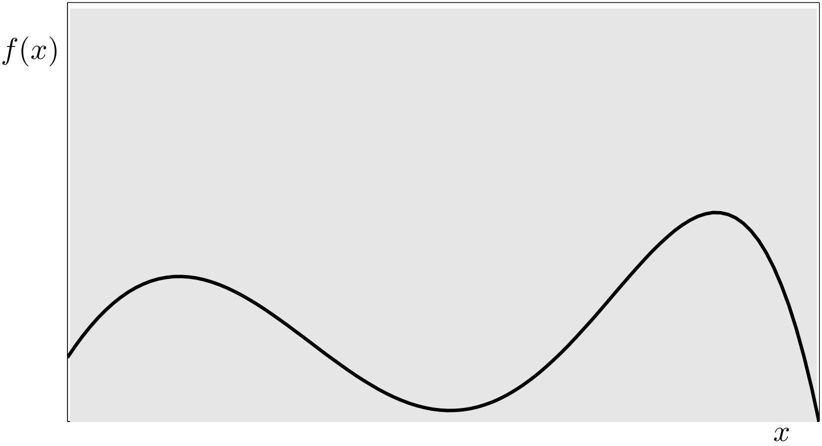
(Srinivas et al. 2010)



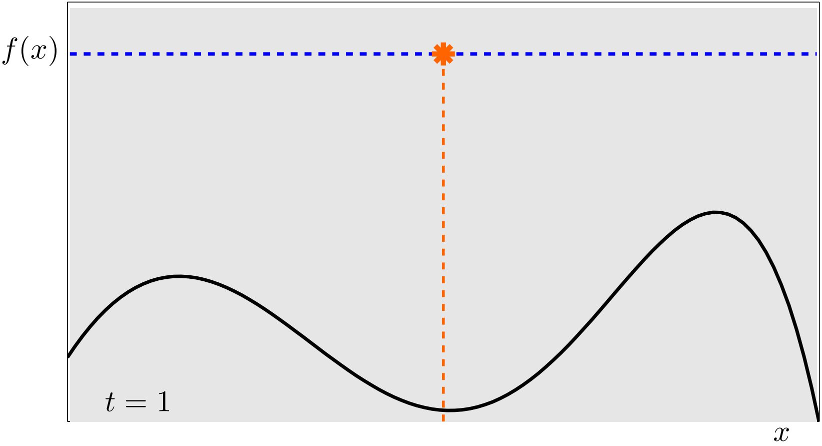
1) Construct posterior $\mathcal{GP}$. 2) $\varphi_t = \mu_{t-1} + \beta_t^{1/2}\sigma_{t-1}$ is a UCB.

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

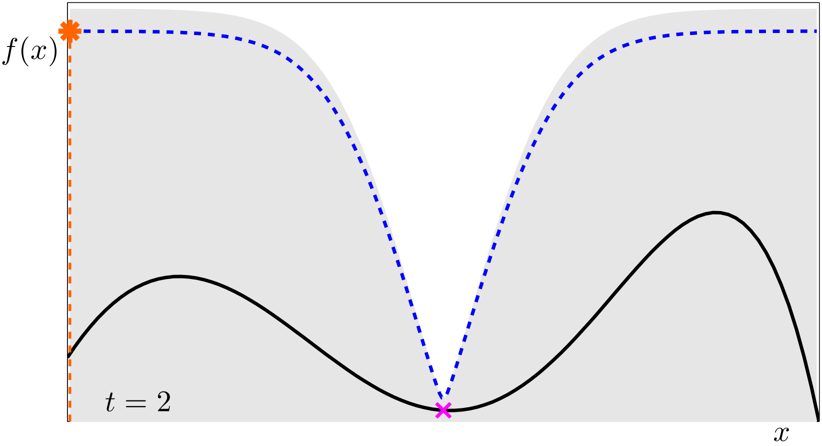
GP-UCB (Srinivas et al. 2010)



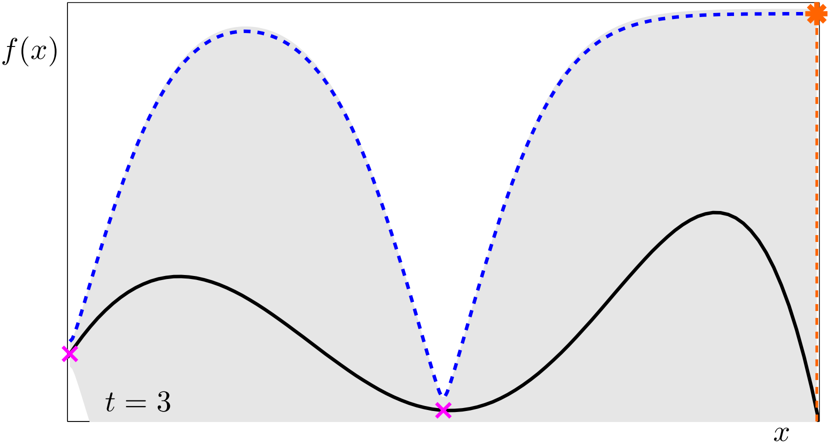
GP-UCB (Srinivas et al. 2010)



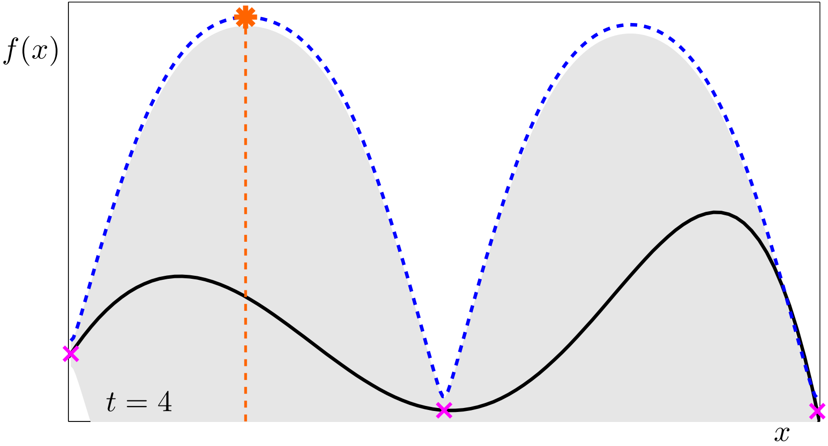
GP-UCB (Srinivas et al. 2010)



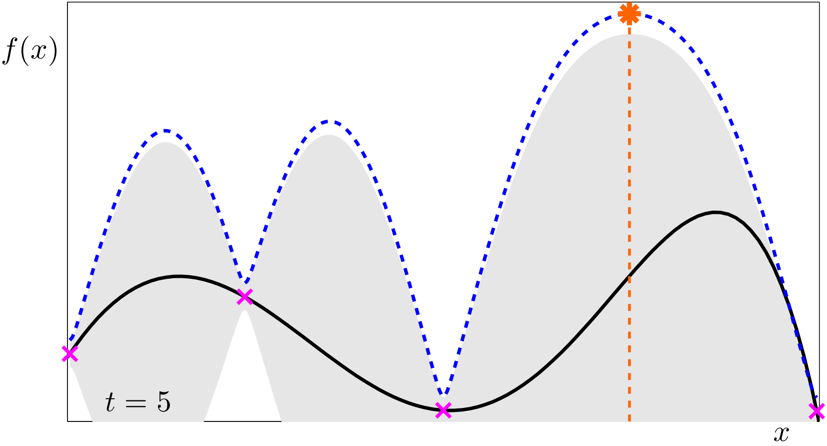
GP-UCB (Srinivas et al. 2010)



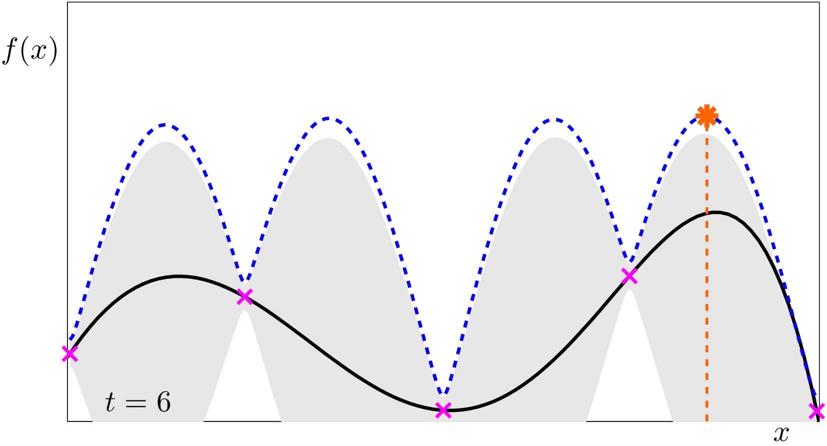
GP-UCB (Srinivas et al. 2010)



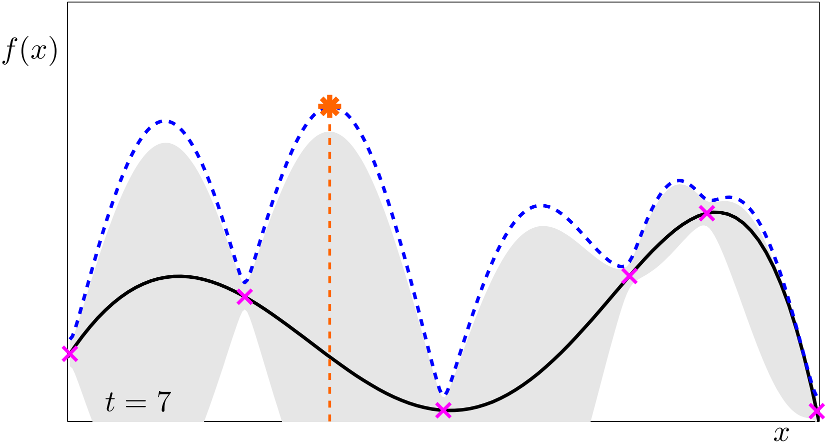
GP-UCB (Srinivas et al. 2010)



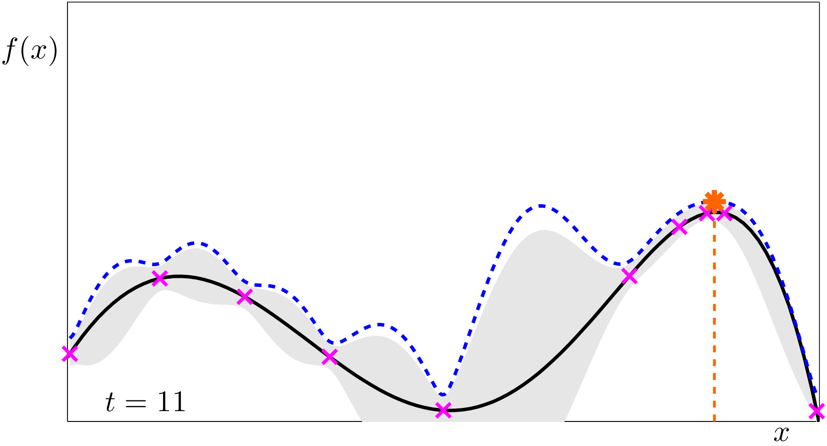
GP-UCB (Srinivas et al. 2010)



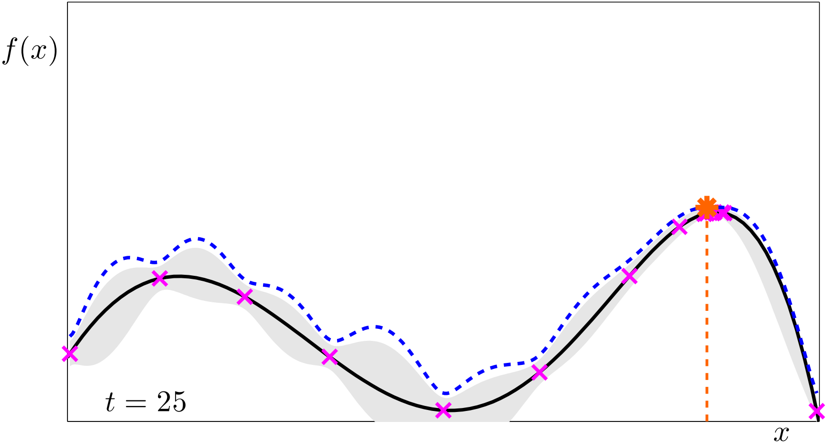
GP-UCB (Srinivas et al. 2010)



GP-UCB (Srinivas et al. 2010)



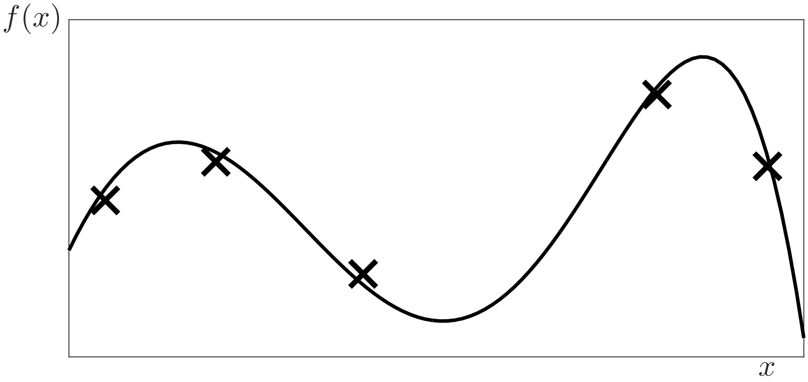
GP-UCB (Srinivas et al. 2010)



Bayesian Optimisation with Thompson Sampling

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

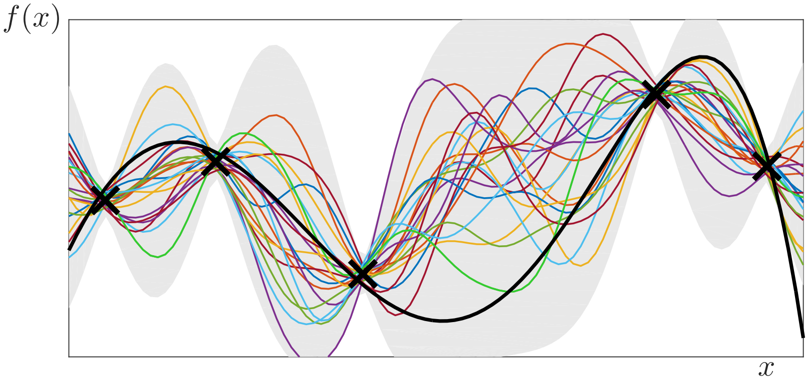
Thompson Sampling (TS) (Thompson, 1933).



Bayesian Optimisation with Thompson Sampling

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

Thompson Sampling (TS) (Thompson, 1933).

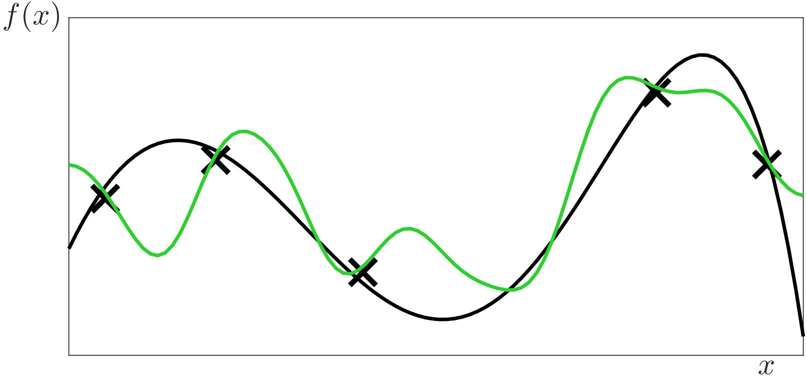


1) Construct posterior $\mathcal{GP}$.

Bayesian Optimisation with Thompson Sampling

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

Thompson Sampling (TS) (Thompson, 1933).

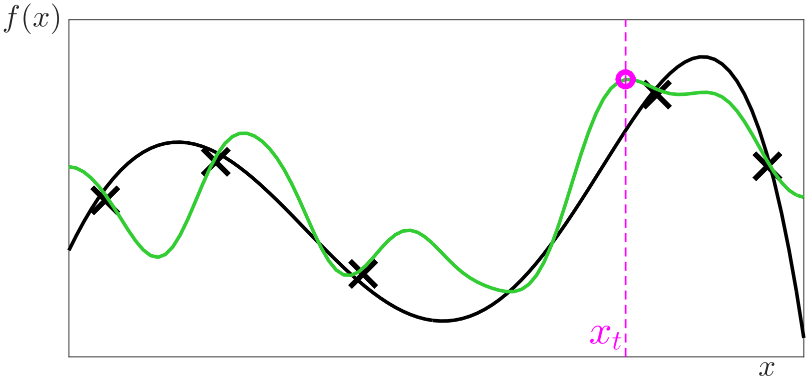


1) Construct posterior $\mathcal{GP}$. 2) Draw sample g from posterior.

Bayesian Optimisation with Thompson Sampling

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

Thompson Sampling (TS) (Thompson, 1933).



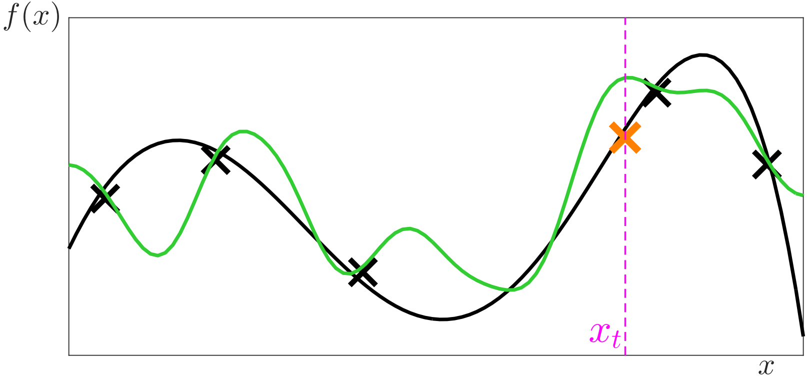
1) Construct posterior $\mathcal{GP}$. 2) Draw sample g from posterior.

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

More on Bayesian Optimisation

Theoretical results: Both UCB and TS will eventually find the optimum under certain smoothness assumptions of f.

More on Bayesian Optimisation

Theoretical results: Both UCB and TS will eventually find the optimum under certain smoothness assumptions of f.

Other criteria for selecting x_t:

- Expected improvement (Jones et al. 1998)
- Probability of improvement (Kushner et al. 1964)
- Predictive entropy search (Hernández-Lobato et al. 2014)
- Information directed sampling (Russo & Van Roy 2014)

More on Bayesian Optimisation

Theoretical results: Both UCB and TS will eventually find the optimum under certain smoothness assumptions of f.

Other criteria for selecting x_t:

- Expected improvement (Jones et al. 1998)
- Probability of improvement (Kushner et al. 1964)
- Predictive entropy search (Hernández-Lobato et al. 2014)
- Information directed sampling (Russo & Van Roy 2014)

Other Bayesian models for f:

- Neural networks (Snoek et al. 2015)
- Random Forests (Hutter 2009)

Some Modern Challenges/Opportunities

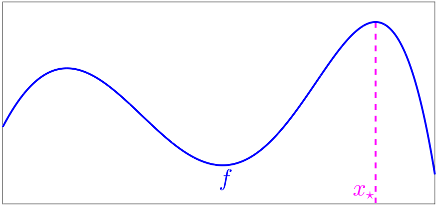
1. Multi-fidelity Optimisation (Kandasamy et al. NIPS 2016 a&b, Kandasamy et al. ICML 2017)

2. Parallelisation (Kandasamy et al. Arxiv 2017)

1. Multi-fidelity Optimisation

(Kandasamy et al. NIPS 2016 a&b, Kandasamy et al. ICML 2017)

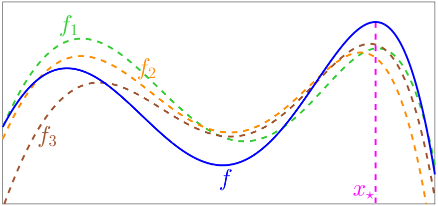
Desired function f is very expensive, but . . .
we have access to cheap approximations.



1. Multi-fidelity Optimisation

(Kandasamy et al. NIPS 2016 a&b, Kandasamy et al. ICML 2017)

Desired function f is very expensive, but . . .
we have access to cheap approximations.

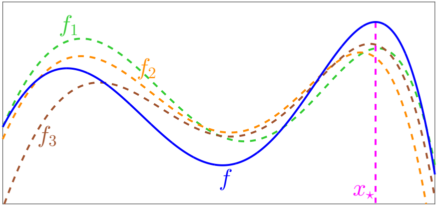


$f_1, f_2, f_3 \approx f$ which are cheaper to evaluate.

1. Multi-fidelity Optimisation

(Kandasamy et al. NIPS 2016 a&b, Kandasamy et al. ICML 2017)

Desired function f is very expensive, but ...
we have access to cheap approximations.



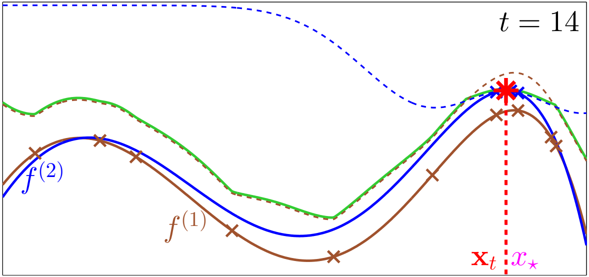
$f_1, f_2, f_3 \approx f$ which are cheaper to evaluate.

E.g. f: a real world battery experiment
f_2: lab experiment
f_1: computer simulation

MF-GP-UCB (Kandasamy et al. NIPS 2016b)

Multi-fidelity Gaussian Process Upper Confidence Bound

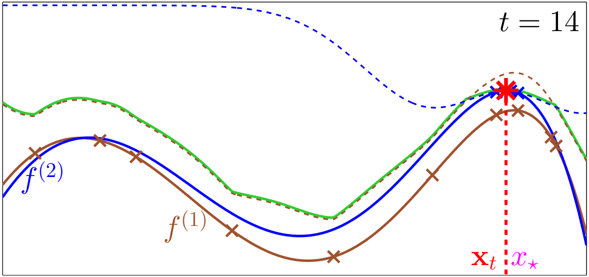
With 2 fidelities (1 Approximation),



MF-GP-UCB (Kandasamy et al. NIPS 2016b)

Multi-fidelity Gaussian Process Upper Confidence Bound

With 2 fidelities (1 Approximation),

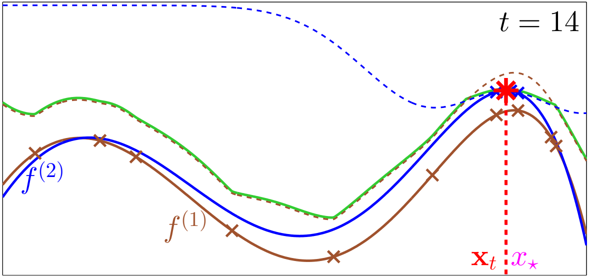


Theorem: MF-GP-UCB finds the optimum $x_\star$ with less resources than GP-UCB on $f^{(2)}$.

MF-GP-UCB (Kandasamy et al. NIPS 2016b)

Multi-fidelity Gaussian Process Upper Confidence Bound

With 2 fidelities (1 Approximation),



Theorem: MF-GP-UCB finds the optimum $x_\star$ with less resources than GP-UCB on $f^{(2)}$.

Can be extended to multiple approximations and continuous approximations.

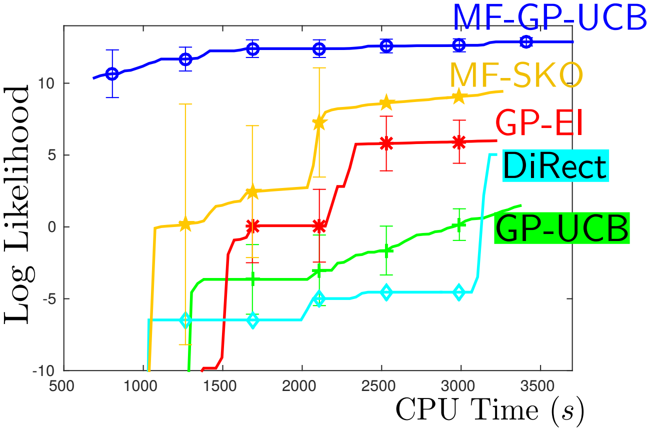
Experiment: Cosmological Maximum Likelihood Inference

- Type Ia Supernovae Data
- Maximum likelihood inference for 3 cosmological parameters:
  - Hubble Constant H_0
  - Dark Energy Fraction Ω_Λ
  - Dark Matter Fraction Ω_M
- Likelihood: Robertson Walker metric (Robertson 1936)
  Requires numerical integration for each point in the dataset.

Experiment: Cosmological Maximum Likelihood Inference

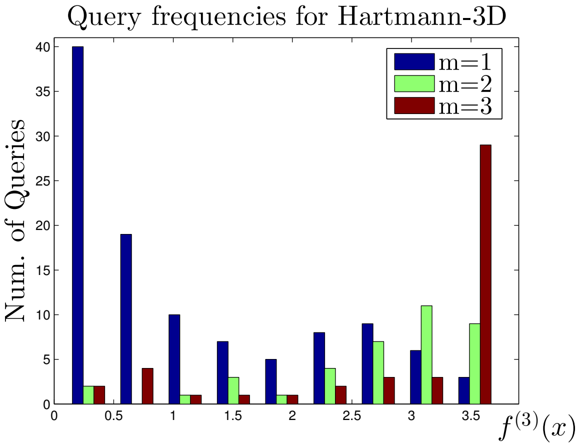
3 cosmological parameters. $(d = 3)$

Fidelities: integration on grids of size $(10^2, 10^4, 10^6)$. $(M = 3)$



Experiment: Hartmann-3D

2 Approximations (3 fidelities).
We want to optimise the $m = 3^{\text{rd}}$ fidelity, which is the most expensive. $m = 1^{\text{st}}$ fidelity is cheapest.



2. Parallelising function evaluations

Parallelisation with M workers: can evaluate f at M different points at the same time.

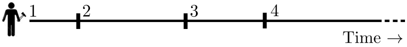
E.g.: Test M different battery solvents at the same time.

2. Parallelising function evaluations

Parallelisation with M workers: can evaluate f at M different points at the same time.

E.g.: Test M different battery solvents at the same time.

Sequential evaluations with one worker

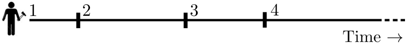


2. Parallelising function evaluations

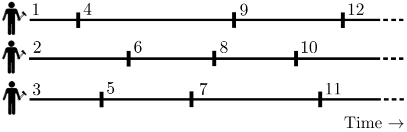
Parallelisation with M workers: can evaluate f at M different points at the same time.

E.g.: Test M different battery solvents at the same time.

Sequential evaluations with one worker



Parallel evaluations with M workers (Asynchronous)

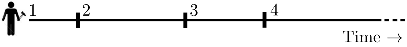


2. Parallelising function evaluations

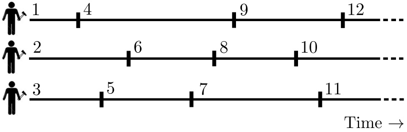
Parallelisation with M workers: can evaluate f at M different points at the same time.

E.g.: Test M different battery solvents at the same time.

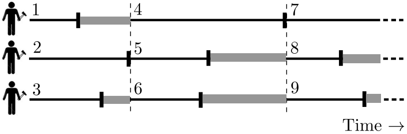
Sequential evaluations with one worker



Parallel evaluations with M workers (Asynchronous)



Parallel evaluations with M workers (Synchronous)

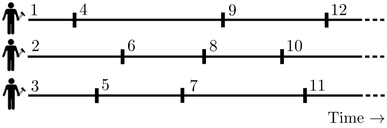


Parallel Thompson Sampling (Kandasamy et al. Arxiv 2017)

Asynchronous: asyTS

At any given time,

1. $(x', y') \leftarrow$ Wait for a worker to finish.
2. Compute posterior $\mathcal{GP}$.
3. Draw a sample $g \sim \mathcal{GP}$.
4. Re-deploy worker at $\operatorname{argmax} g$.



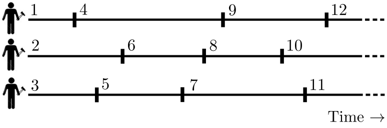
Parallel Thompson Sampling

(Kandasamy et al. Arxiv 2017)

Asynchronous: asyTS

At any given time,

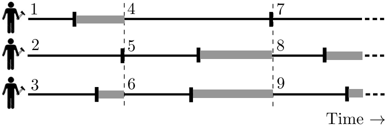
1. $(x', y') \leftarrow$ Wait for a worker to finish.
2. Compute posterior $\mathcal{GP}$.
3. Draw a sample $g \sim \mathcal{GP}$.
4. Re-deploy worker at $\mathop{\mathrm{argmax}} g$.



Synchronous: synTS

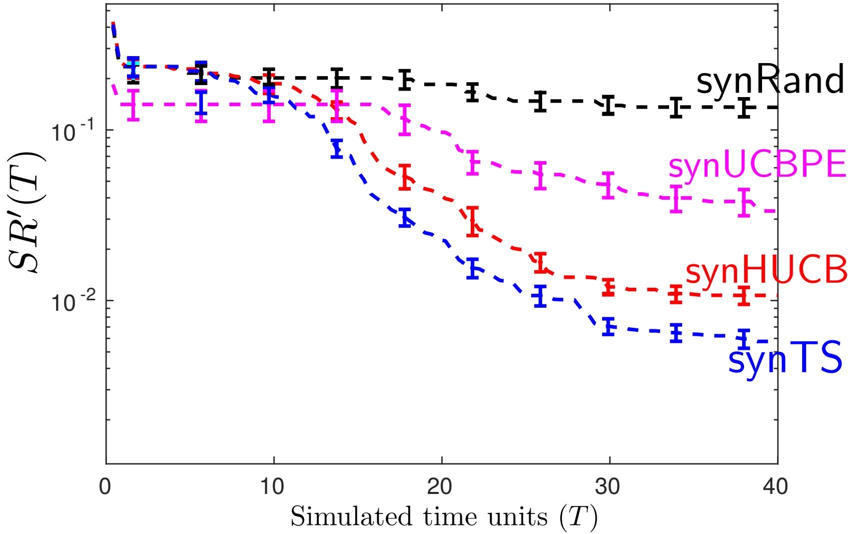
At any given time,

1. $\{(x'_m, y'_m)\}_{m=1}^{M} \leftarrow$ Wait for all workers to finish.
2. Compute posterior $\mathcal{GP}$.
3. Draw M samples $g_m \sim \mathcal{GP}$, $\forall m$.
4. Re-deploy worker m at $\mathop{\mathrm{argmax}} g_m$, $\forall m$.



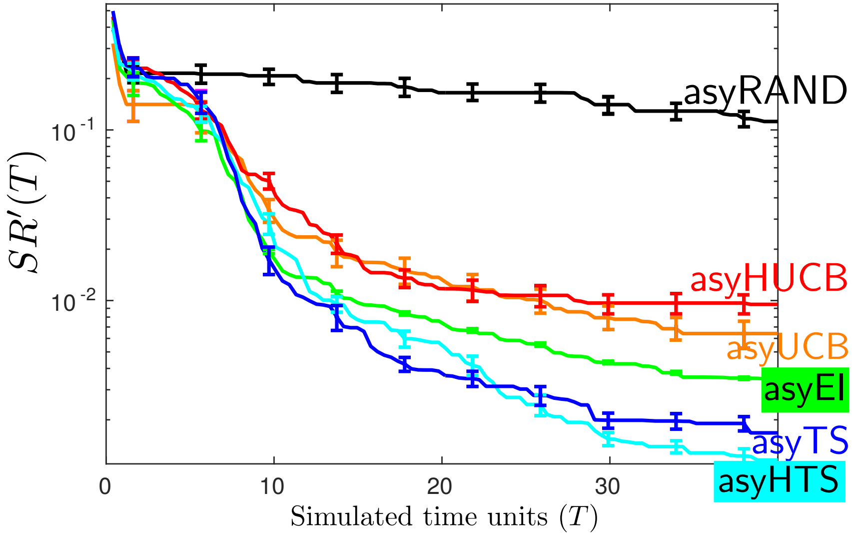
Experiment: Branin-2D $M = 4$

Evaluation time sampled from a uniform distribution



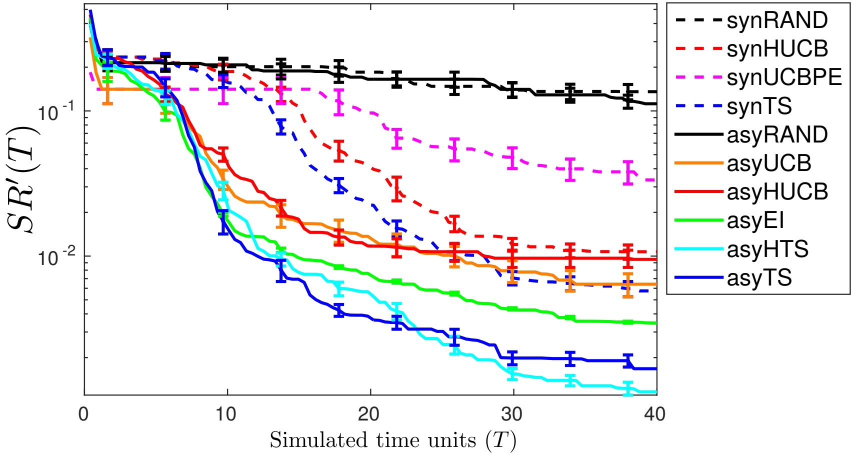
Experiment: Branin-2D $M = 4$

Evaluation time sampled from a uniform distribution



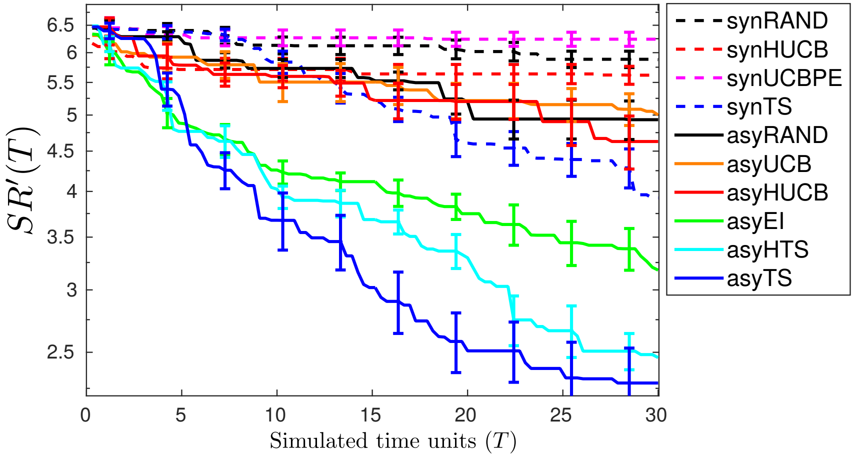
Experiment: Branin-2D $M = 4$

Evaluation time sampled from a uniform distribution



Experiment: Hartmann-18D $M = 25$

Evaluation time sampled from an exponential distribution



Summary

- Black-box Optimisation methods are used in several scientific and engineering applications.
- Bayesian Optimisation: A method for black-box optimisation which uses Bayesian uncertainty estimates for f.
- Some modern challenges
  - Multi-fidelity optimisation
  - Parallel evaluations
  - and several more ...

Summary

- Black-box Optimisation methods are used in several scientific and engineering applications.
- Bayesian Optimisation: A method for black-box optimisation which uses Bayesian uncertainty estimates for f.
- Some modern challenges
  - Multi-fidelity optimisation
  - Parallel evaluations
  - and several more ...

Thank you.

Slides are up on my website: `www.cs.cmu.edu/~kkandasa`