STOR-i

BOSH

Bayesian Optimisation Sampled Hierarchically

Henry Moss, David Leslie and Paul Rayson

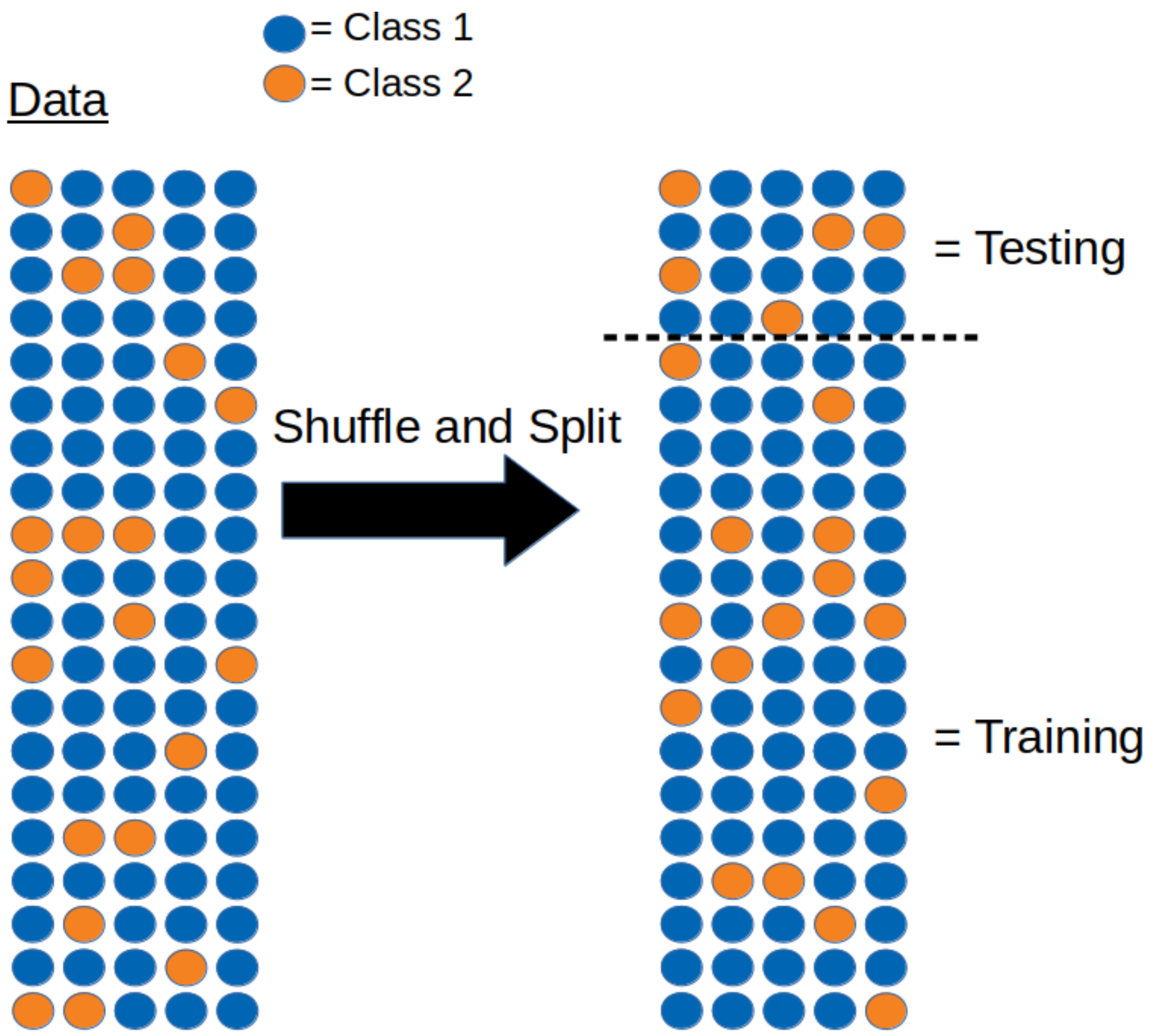
Lancaster University, UK

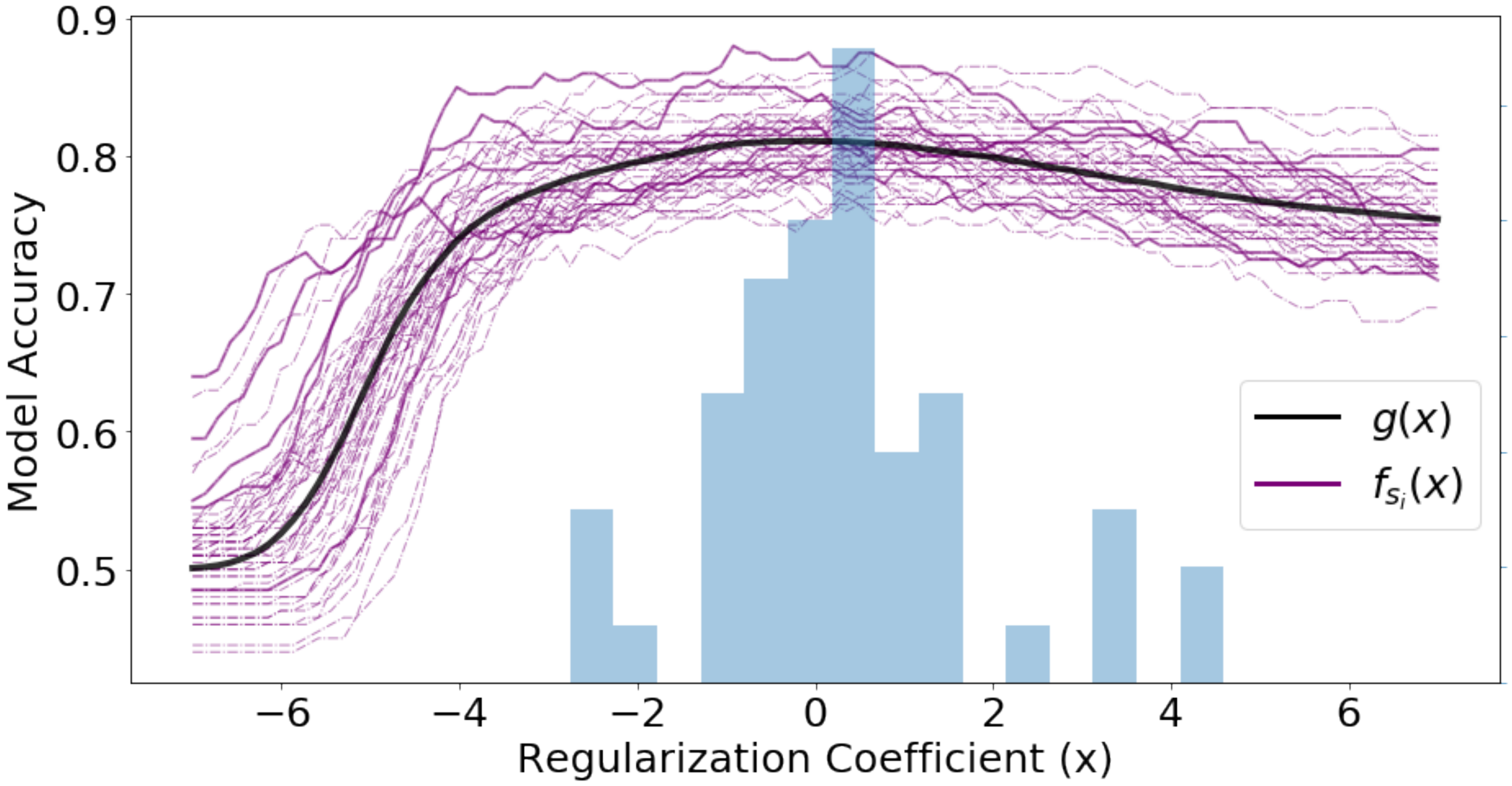
h.moss@lancaster.ac.uk

@Henrymossmoss

Stochastic Optimisation

Realisations of the objective function depend on choice of seed





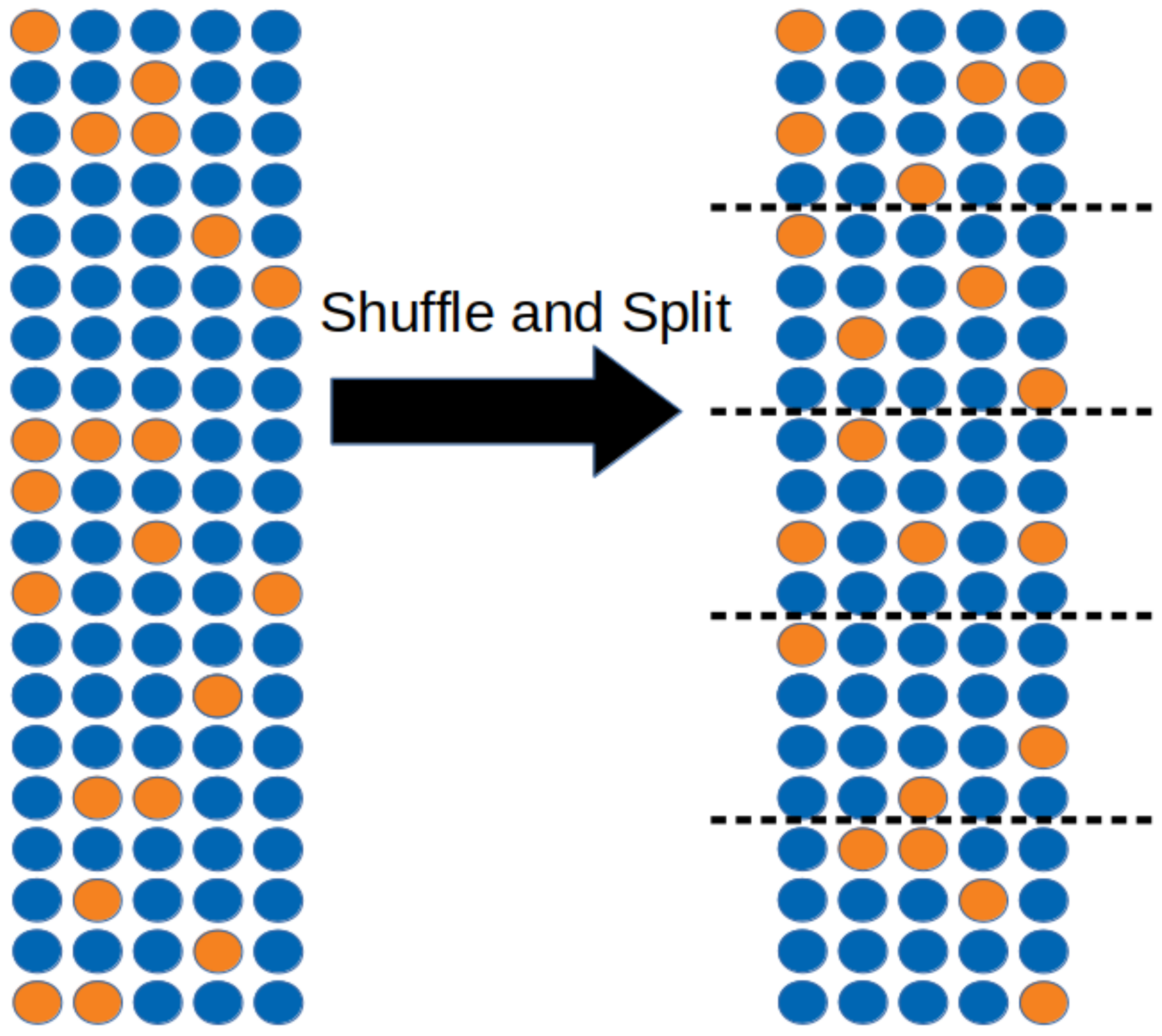
Stochastic Optimisation

Common approach: use an fixed set of seeds

1. Define an evaluation strategy of K realisations $S = \{s_1, \ldots, s_K\}$
2. Find $\mathrm{x}_S^* = \mathrm{argmax}\ \tilde{g}_S(\mathrm{x})$

where $\tilde{g}_S(\mathrm{x})$ is the average realisation

e.g. K-fold cross validation



Stochastic Optimisation

However

$$E_S\left[g(\mathrm{x}^*) - g(\mathrm{x}_S^*)\right] = O\left(\frac{1}{K}\right)$$

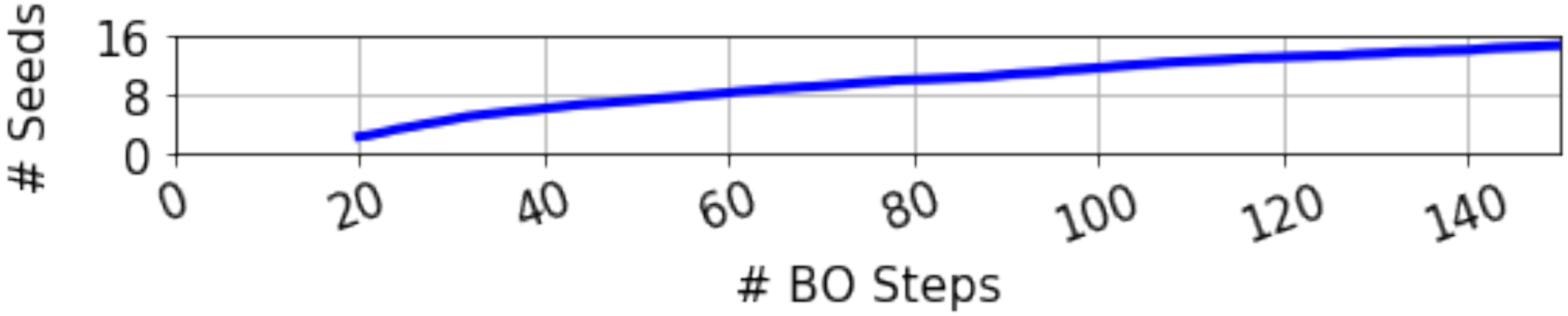
fixed evaluation strategies can be either

- **efficient**
- **precise**

but not both!

BOSH has an adaptive evaluation strategy

BOSH considers a growing collection of realisations

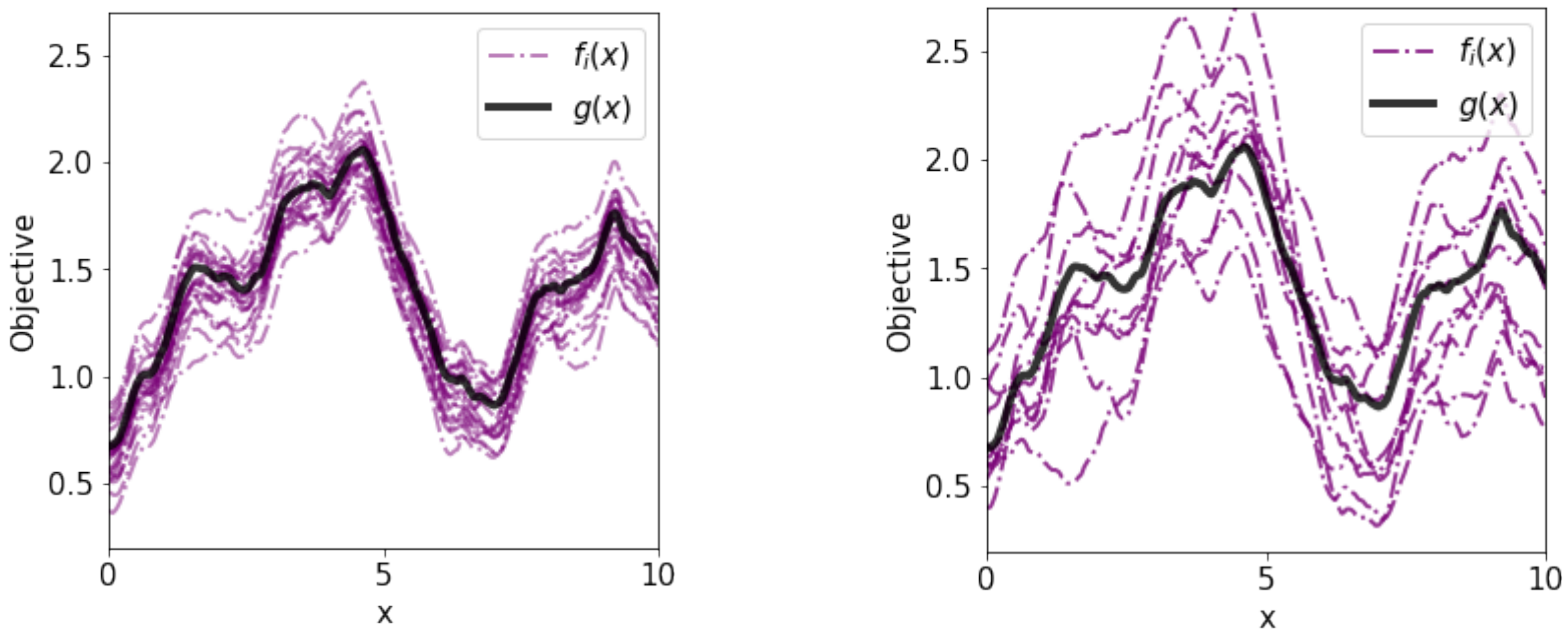


- ☑ Initial small # realisations —-> efficient early optimisation
- ☑ Eventually large # realisations —> precise end optimisation

BOSH has a HGP surrogate model

Instead of using BO to optimise $\tilde{g}_S(x)$, BOSH optimises g(x) directly using a

Hierarchical Gaussian Process (Hensman et al. 2013)



BOSH has an information-theoretic acquisition function

Either evaluate a realisation in current pool or generate a new realisation

$$\alpha_n(\mathsf{x}, s) = MI(g^*; y_s(\mathsf{x}) \,|\, D_n) = H(g^* \,|\, D_n) - E_{y \sim y_s(\mathsf{x})}\left[H(g^* \,|\, D_n, y)\right]$$

TAKE HOME MESSAGE

BOSH is an

- efficient
- high precision

global optimiser for highly stochastic and and high cost functions

BOSH can be used for

- Hyper-parameter tuning of machine learning algorithms
- Reinforcement learning
- Simulation Optimisation
- `Batch design problems