Multi-Fidelity Bayesian Optimization with Unreliable Information Sources

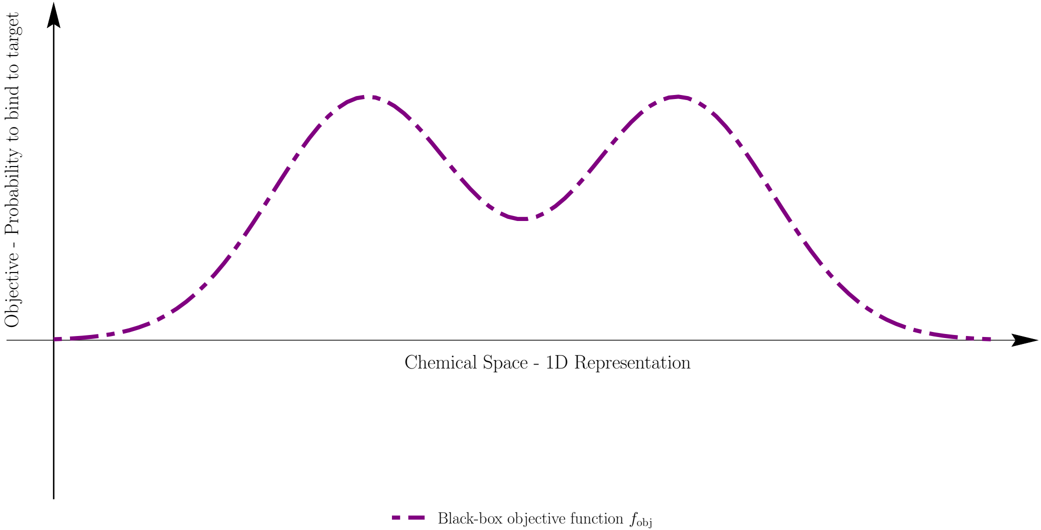
Petrus Mikkola, **Julien Martinelli**, Louis Filstroff, Samuel Kaski

Aalto University

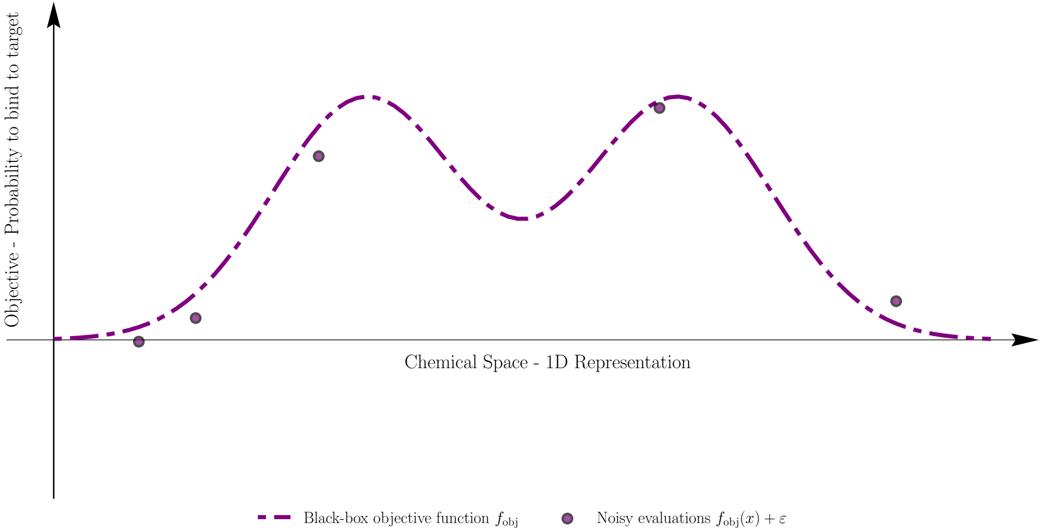
November 16^{th}, 2022

Bayesian Optimization 101

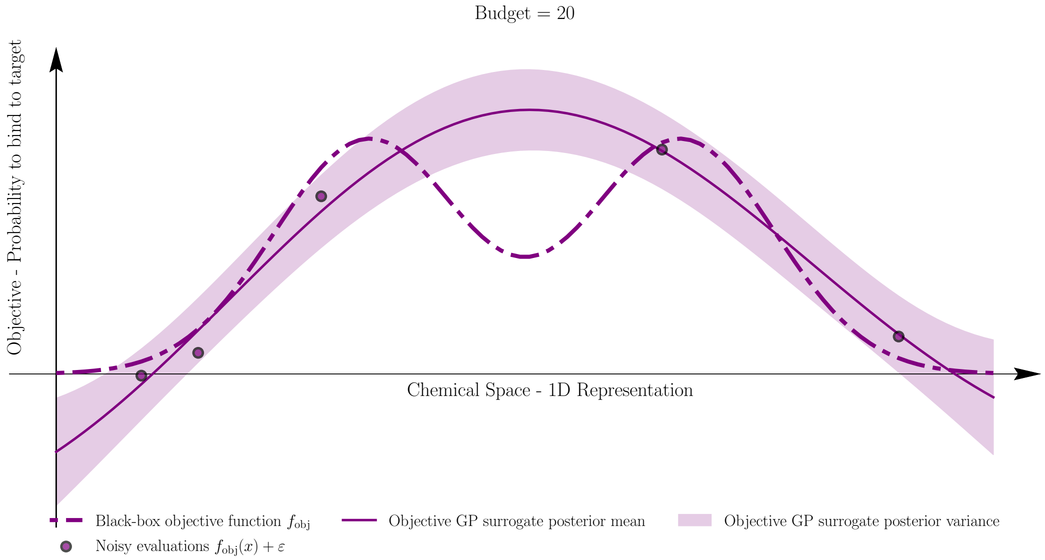
Bayesian Optimization 101



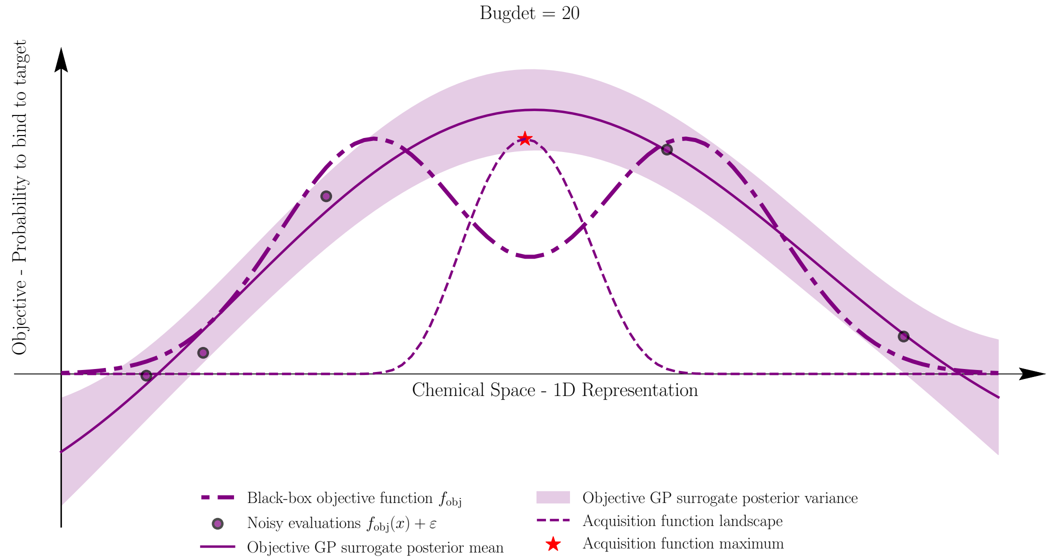
Bayesian Optimization 101



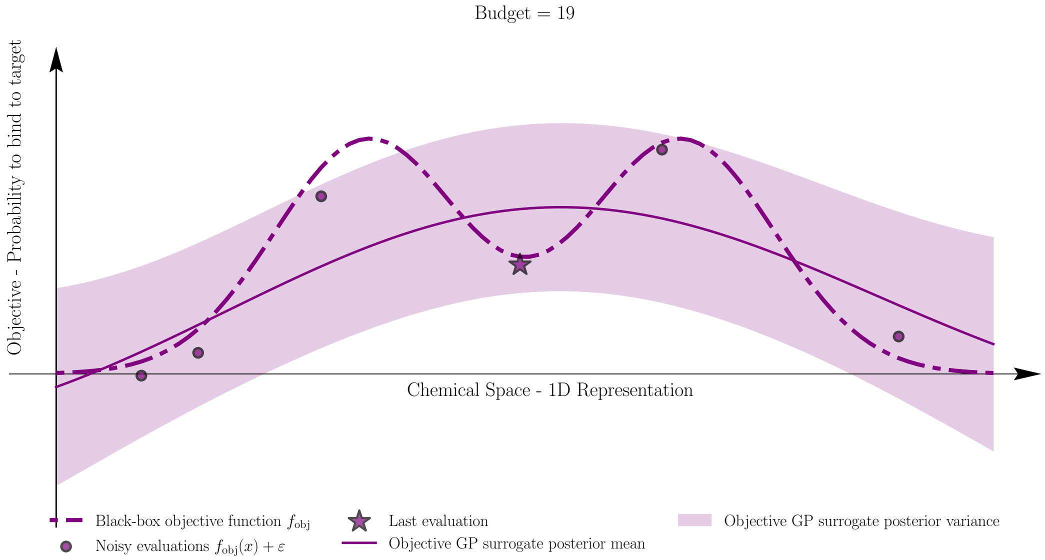
Bayesian Optimization 101



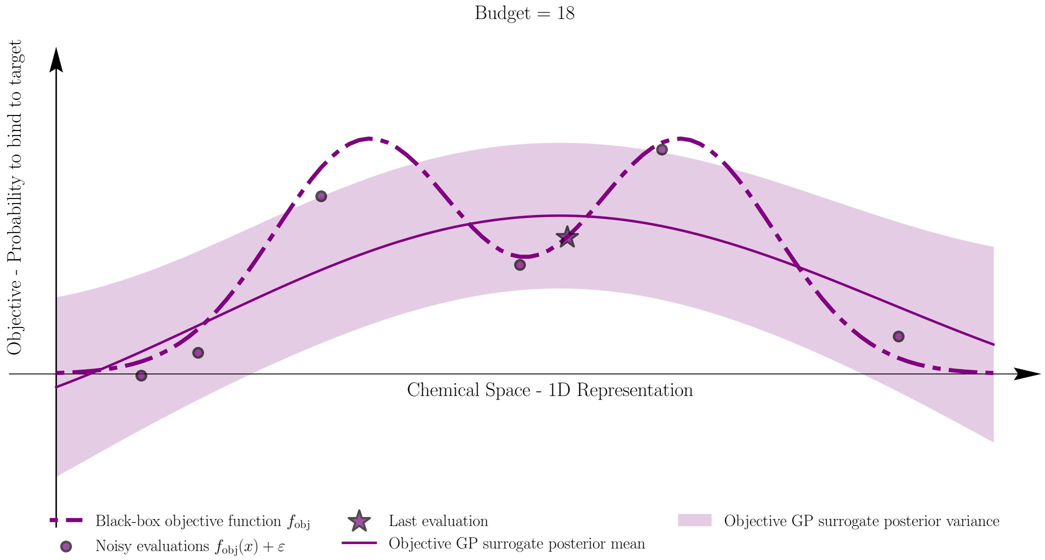
Bayesian Optimization 101



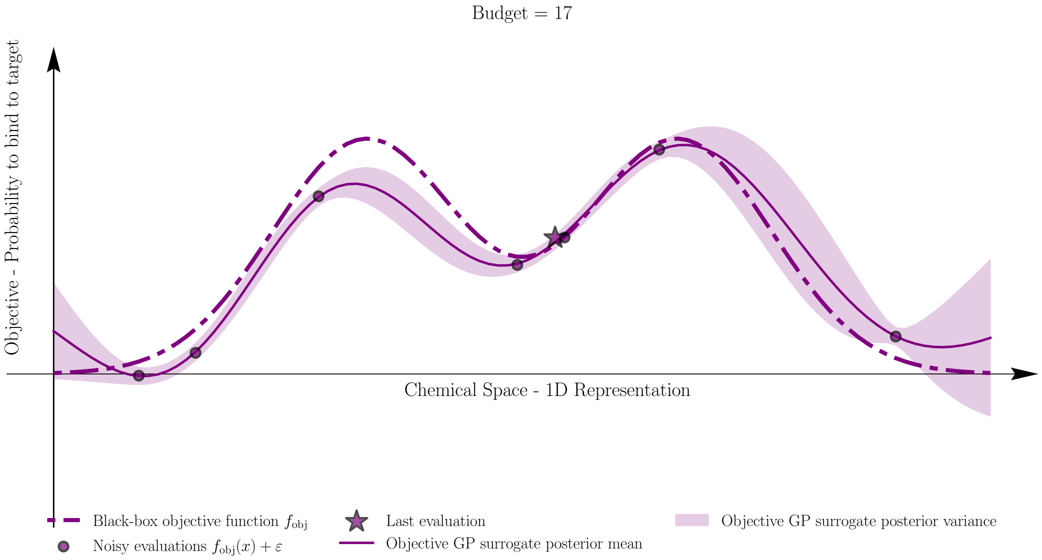
Bayesian Optimization 101



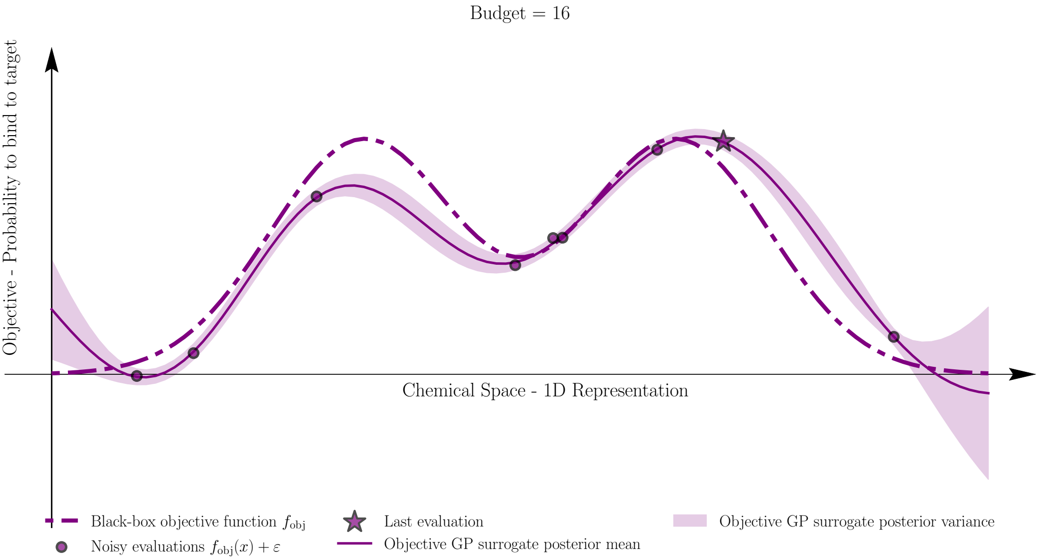
Bayesian Optimization 101



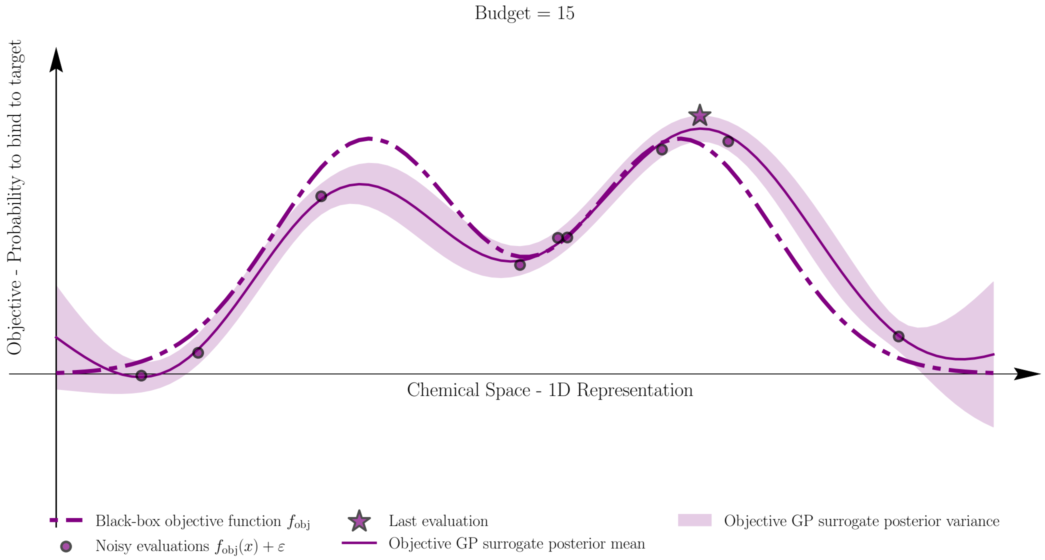
Bayesian Optimization 101



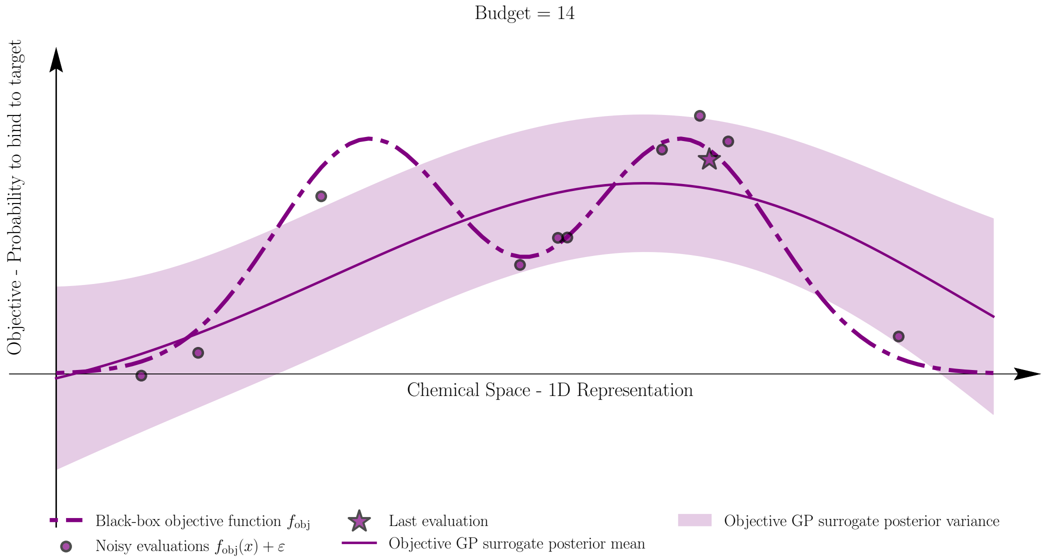
Bayesian Optimization 101



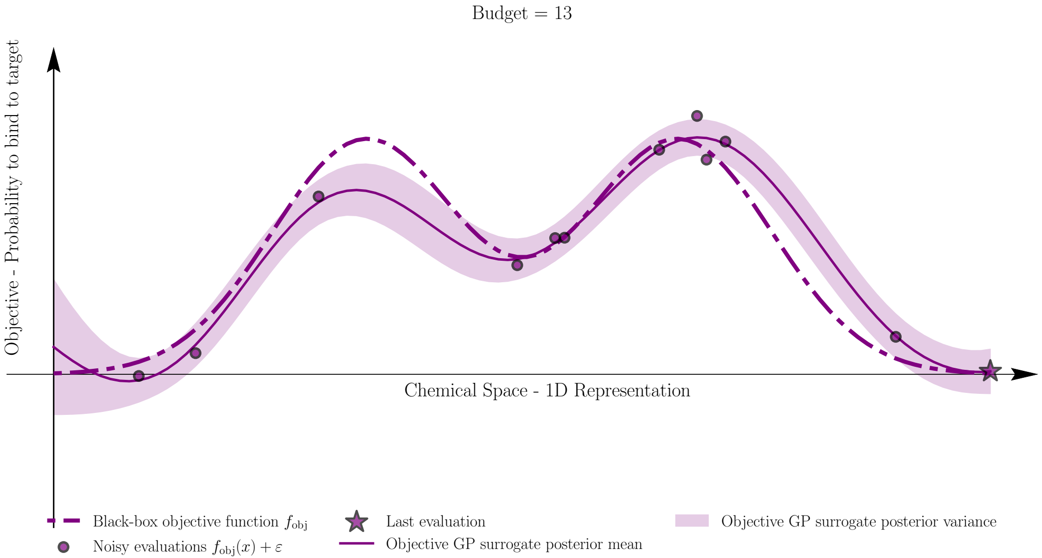
Bayesian Optimization 101



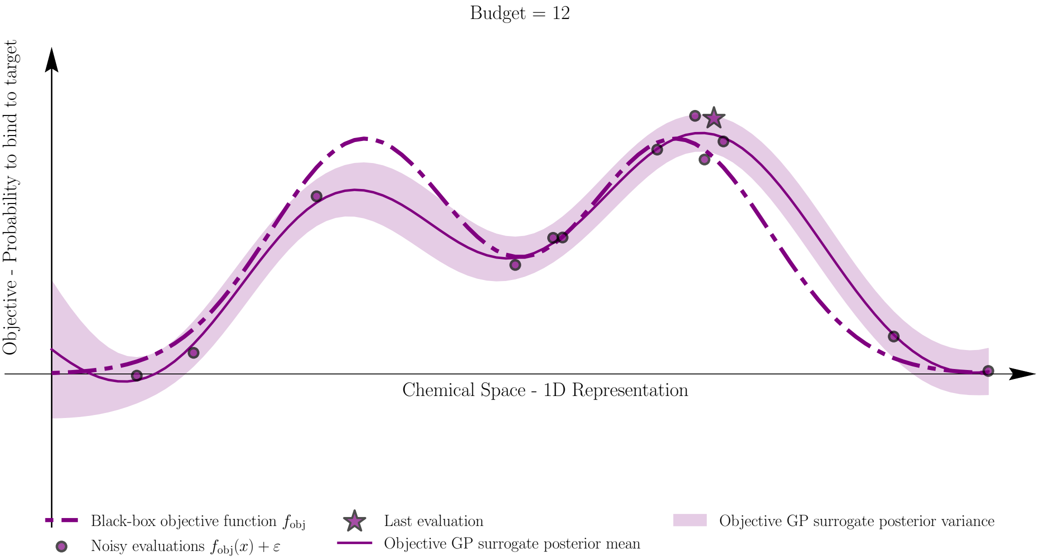
Bayesian Optimization 101



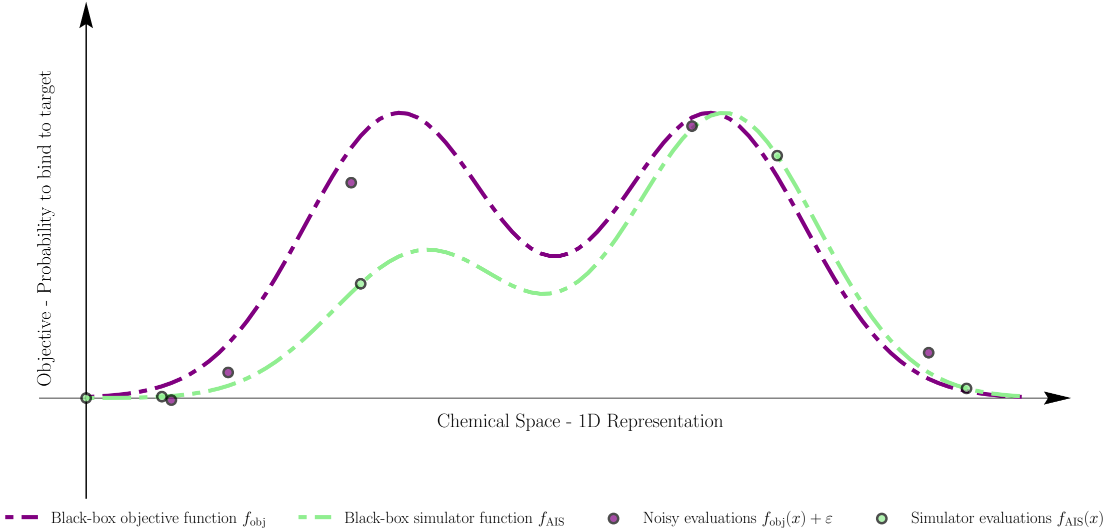
Bayesian Optimization 101



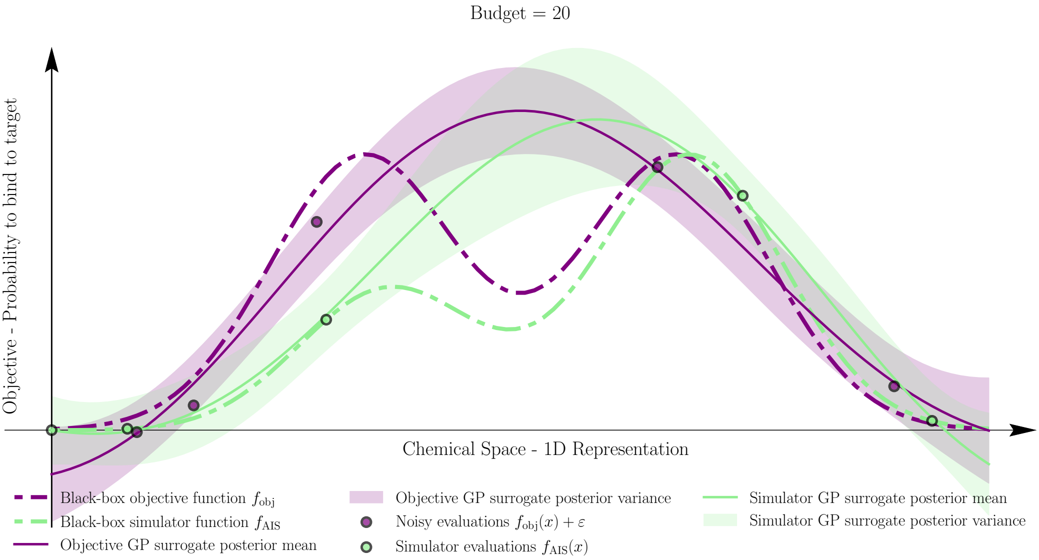
Bayesian Optimization 101



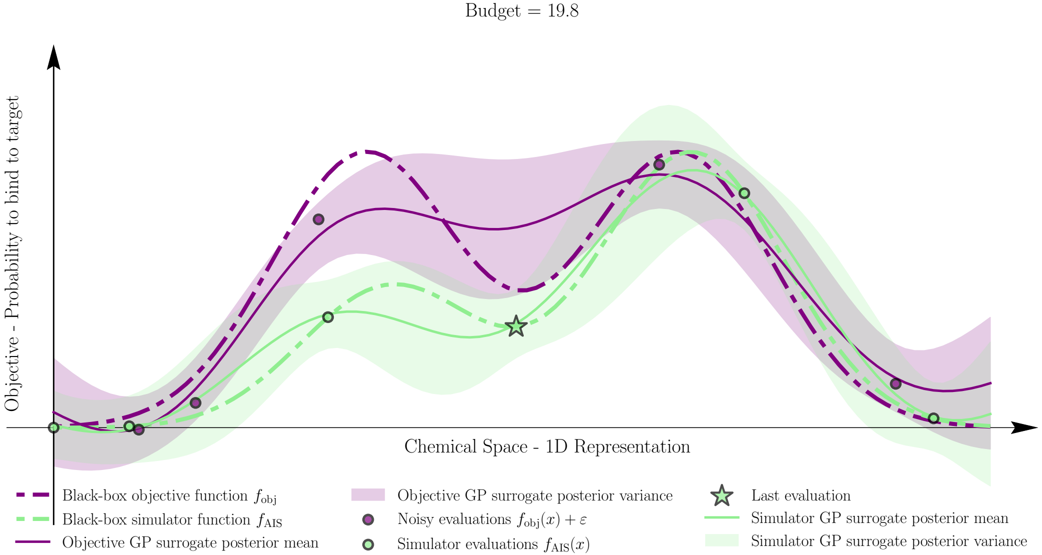
Multi Fidelity Bayesian Optimization 101



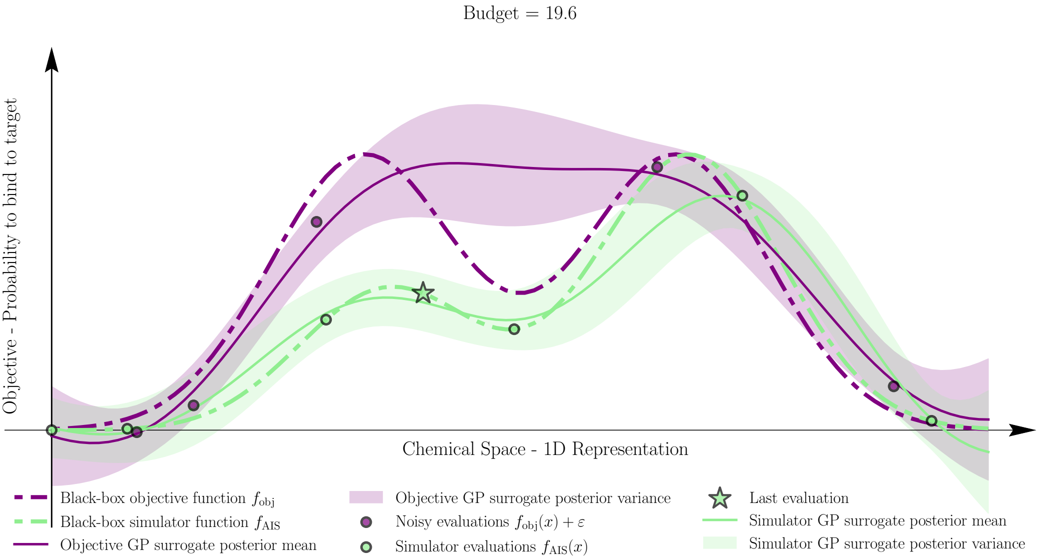
Multi Fidelity Bayesian Optimization 101



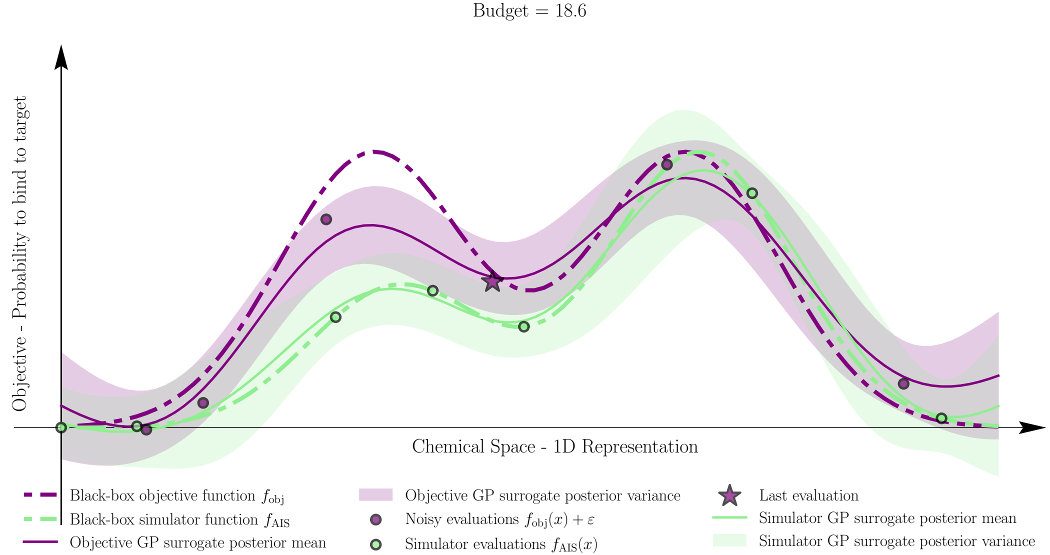
Multi Fidelity Bayesian Optimization 101



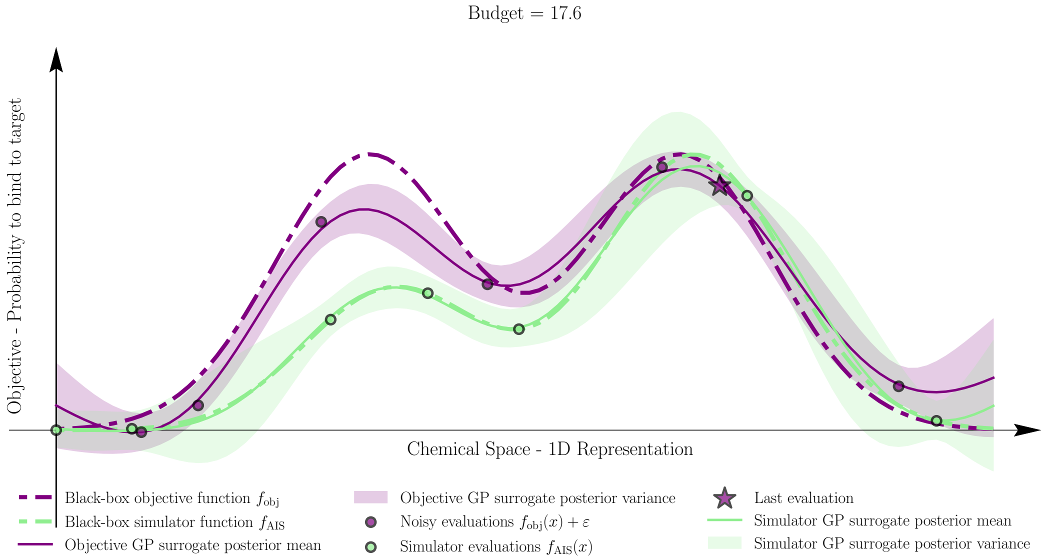
Multi Fidelity Bayesian Optimization 101



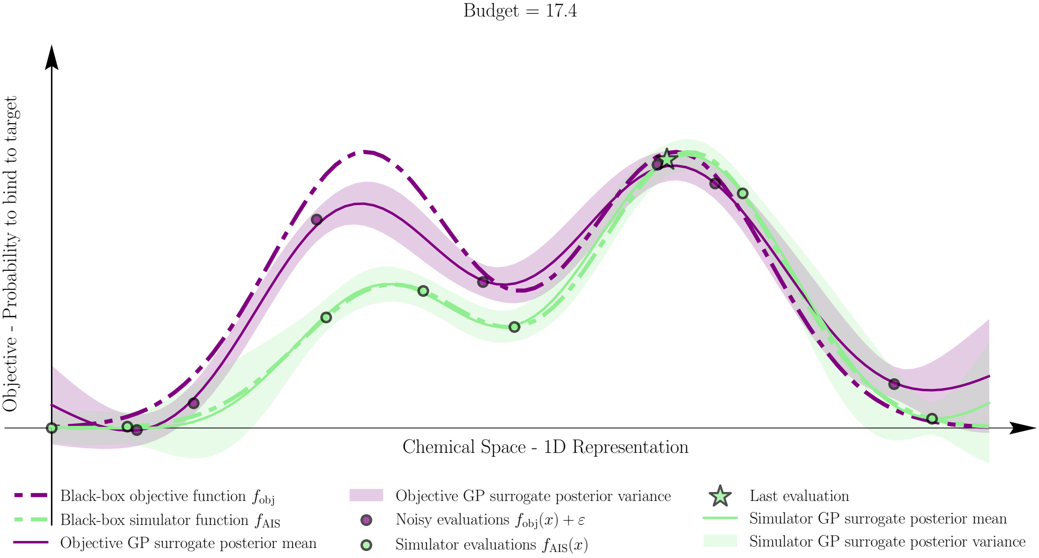
Multi Fidelity Bayesian Optimization 101



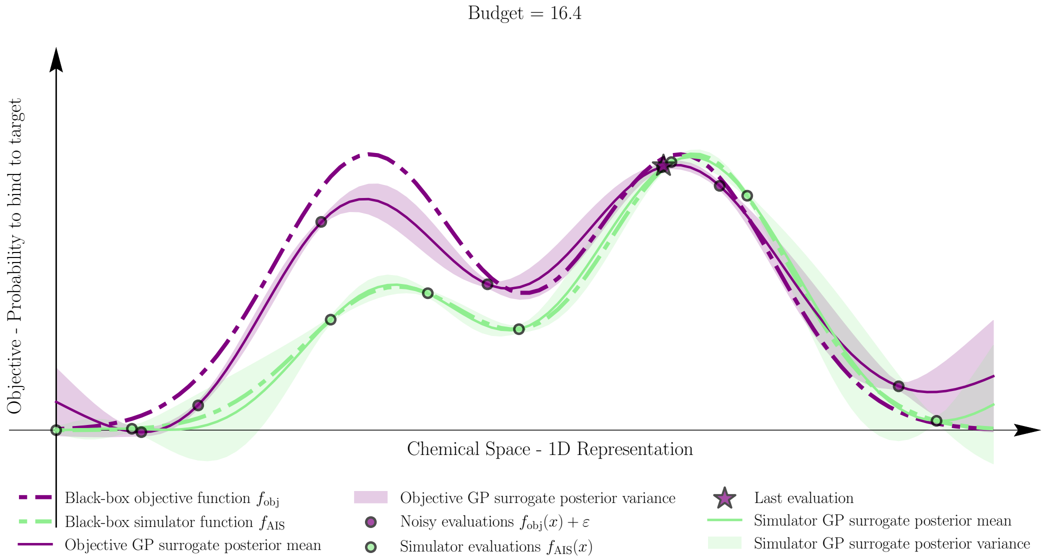
Multi Fidelity Bayesian Optimization 101



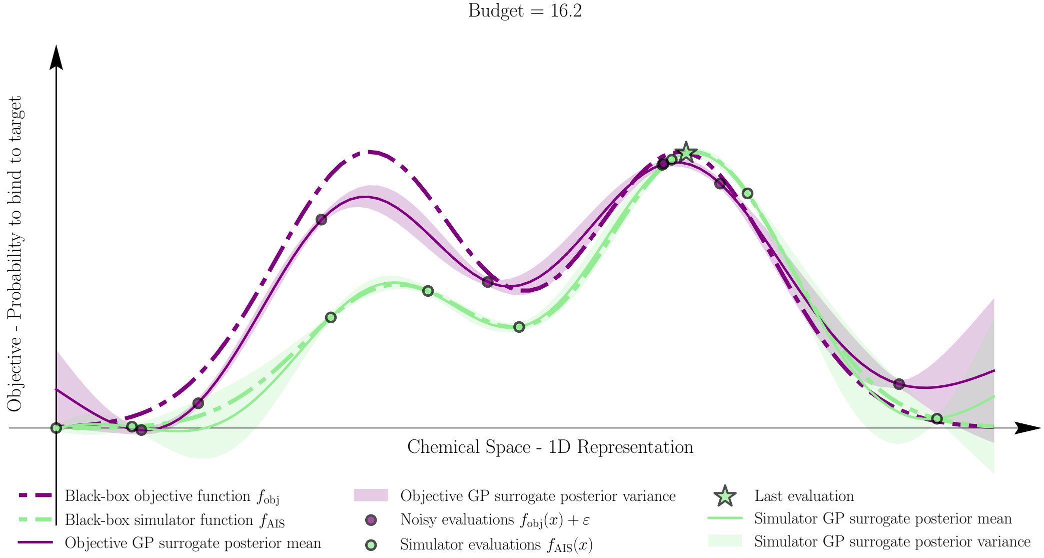
Multi Fidelity Bayesian Optimization 101



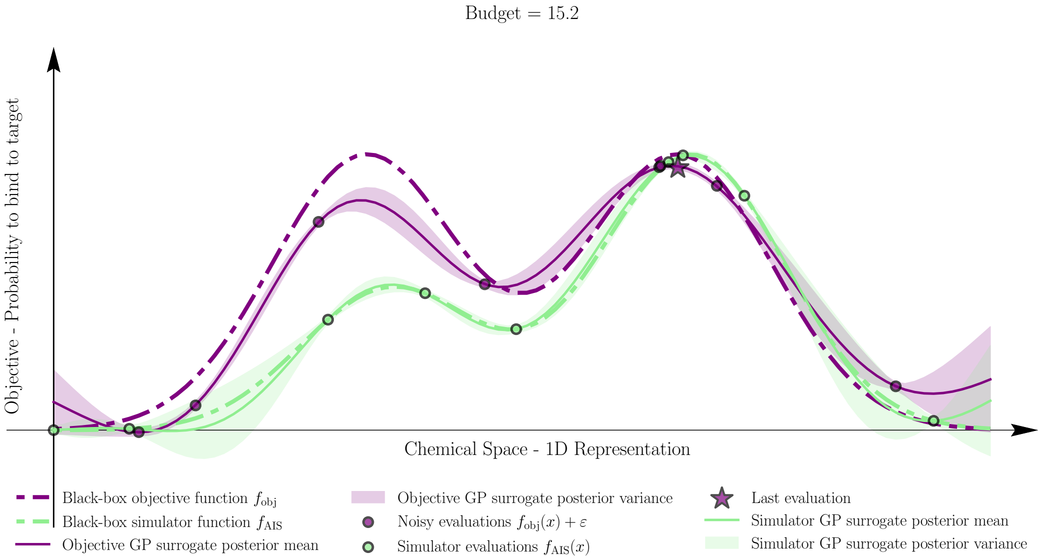
Multi Fidelity Bayesian Optimization 101



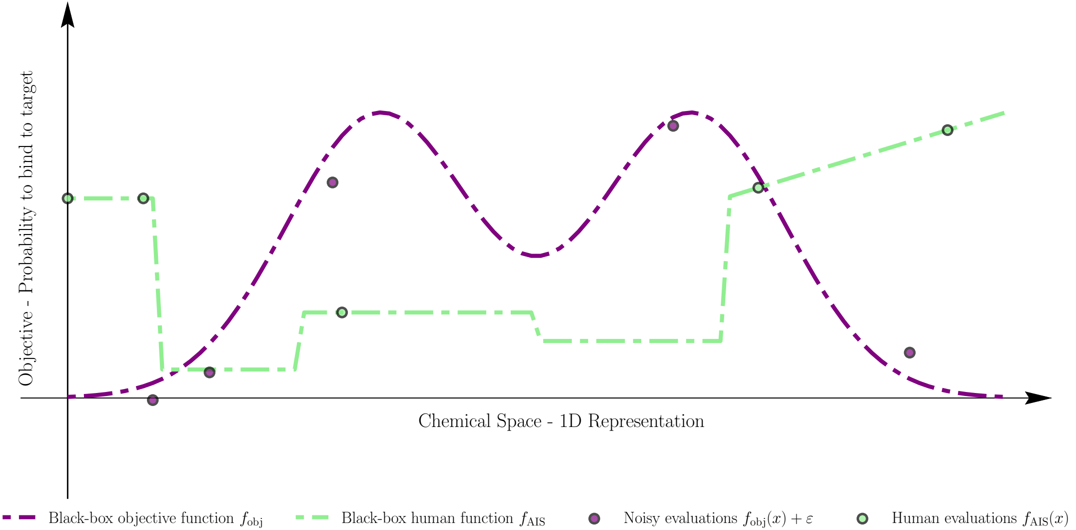
Multi Fidelity Bayesian Optimization 101



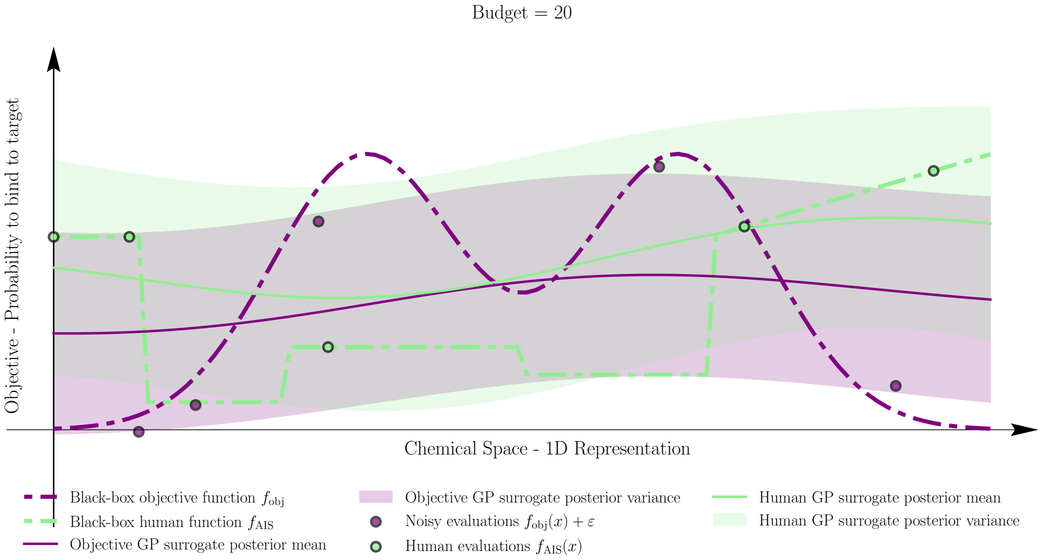
Multi Fidelity Bayesian Optimization 101



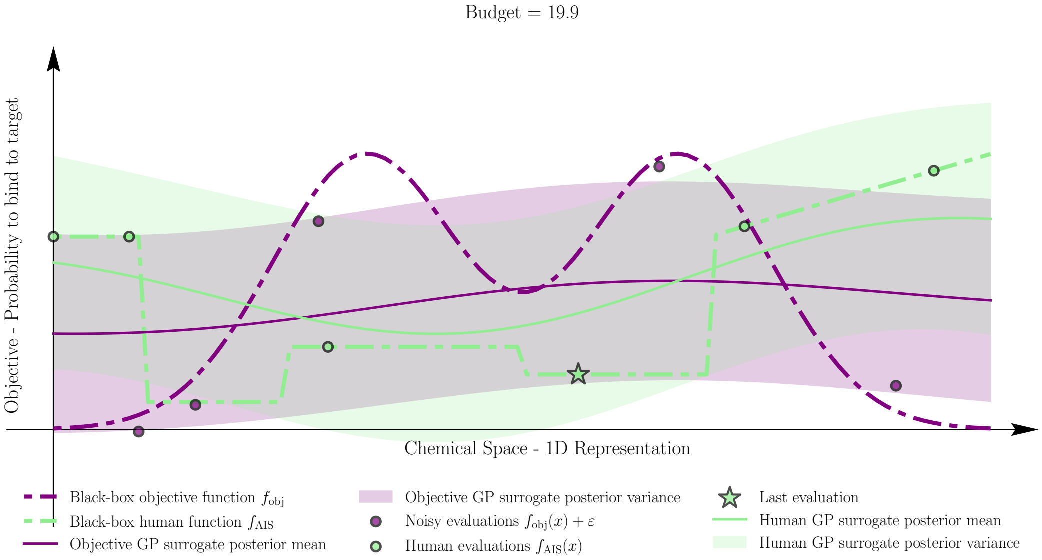
Multi Fidelity Bayesian Optimization with Unreliable Sources



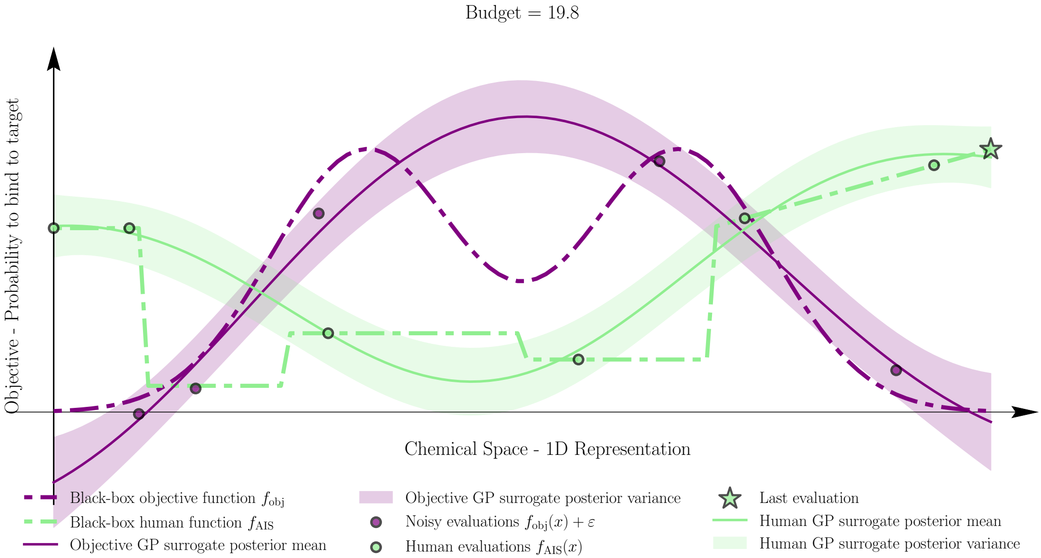
Multi Fidelity Bayesian Optimization with Unreliable Sources



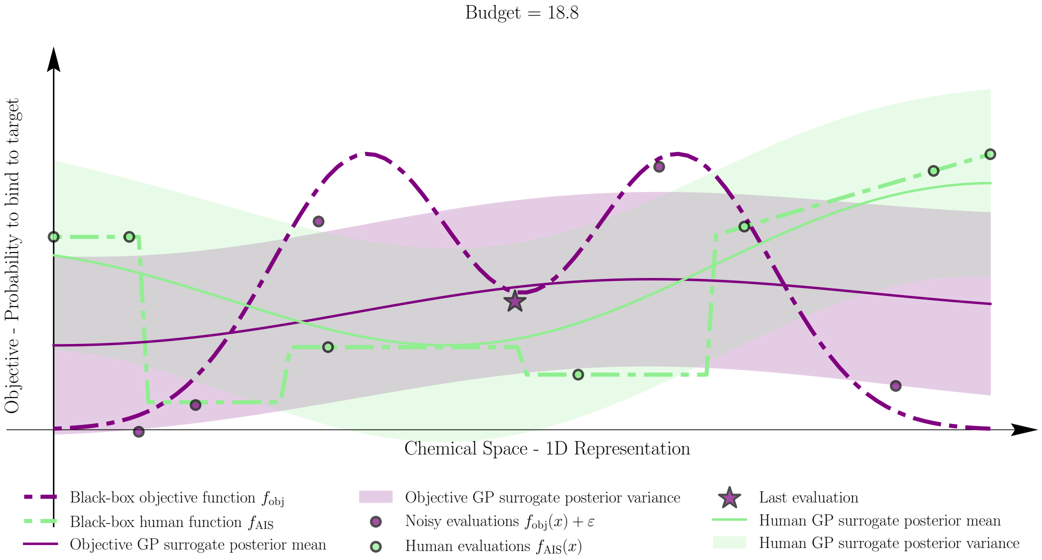
Multi Fidelity Bayesian Optimization with Unreliable Sources



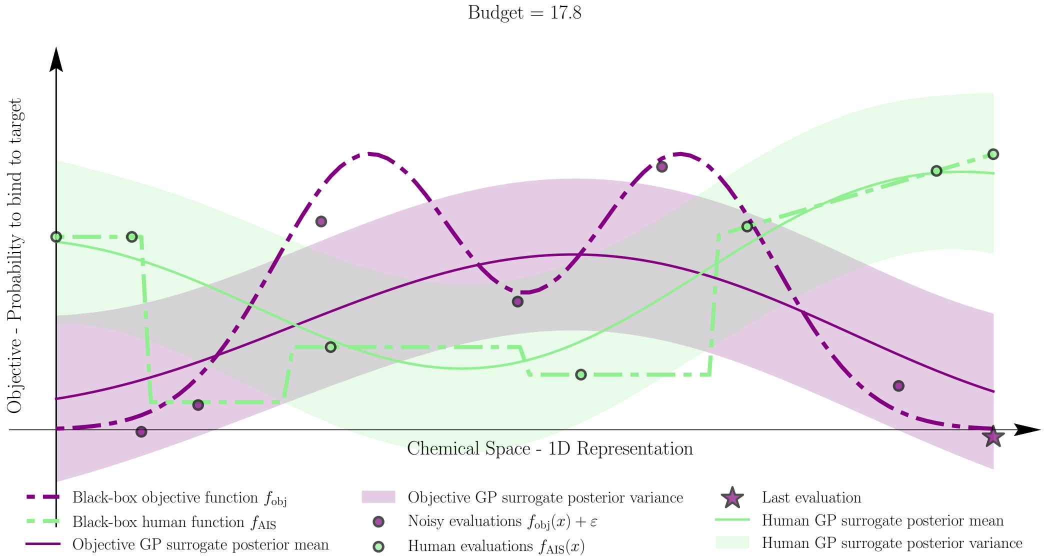
Multi Fidelity Bayesian Optimization with Unreliable Sources



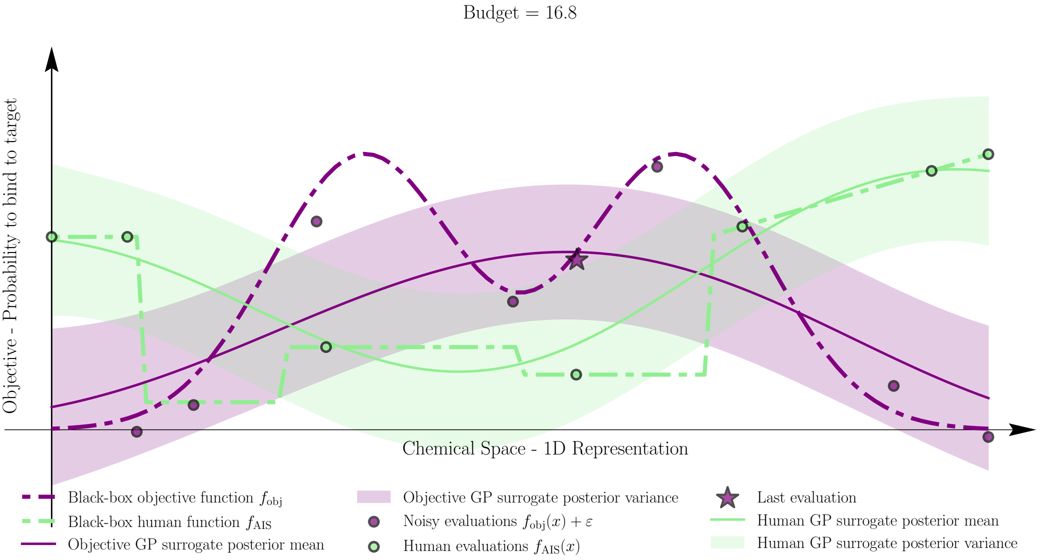
Multi Fidelity Bayesian Optimization with Unreliable Sources



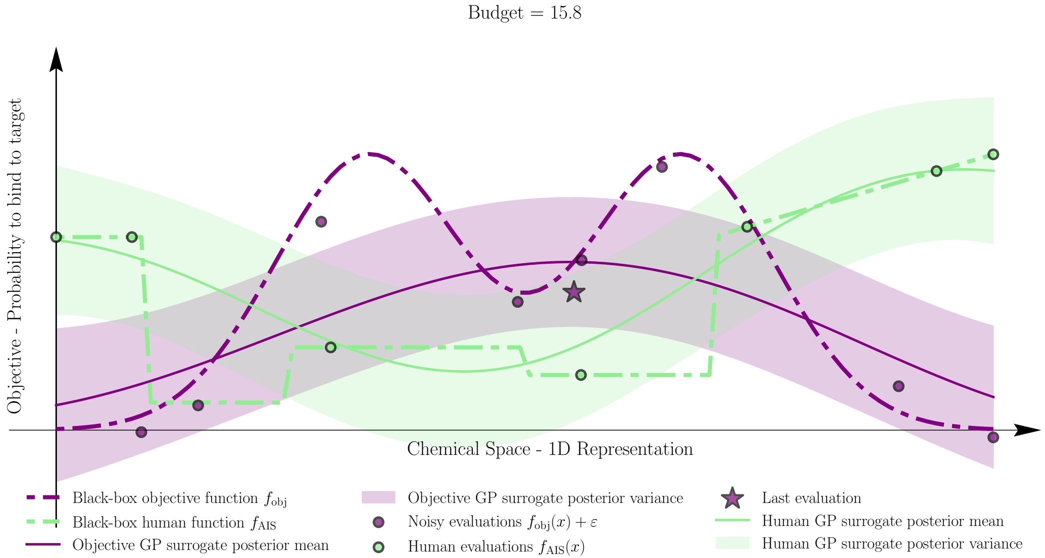
Multi Fidelity Bayesian Optimization with Unreliable Sources



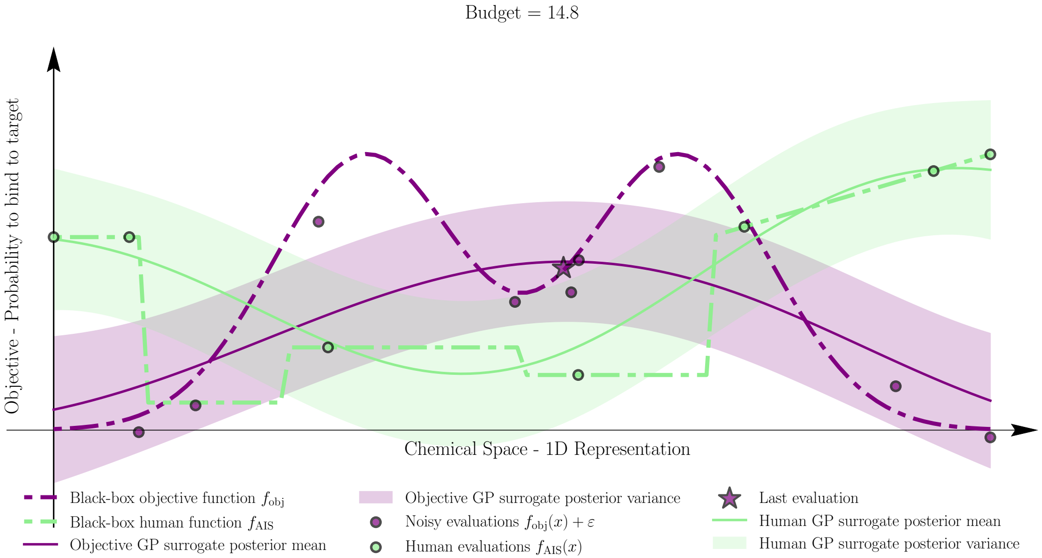
Multi Fidelity Bayesian Optimization with Unreliable Sources



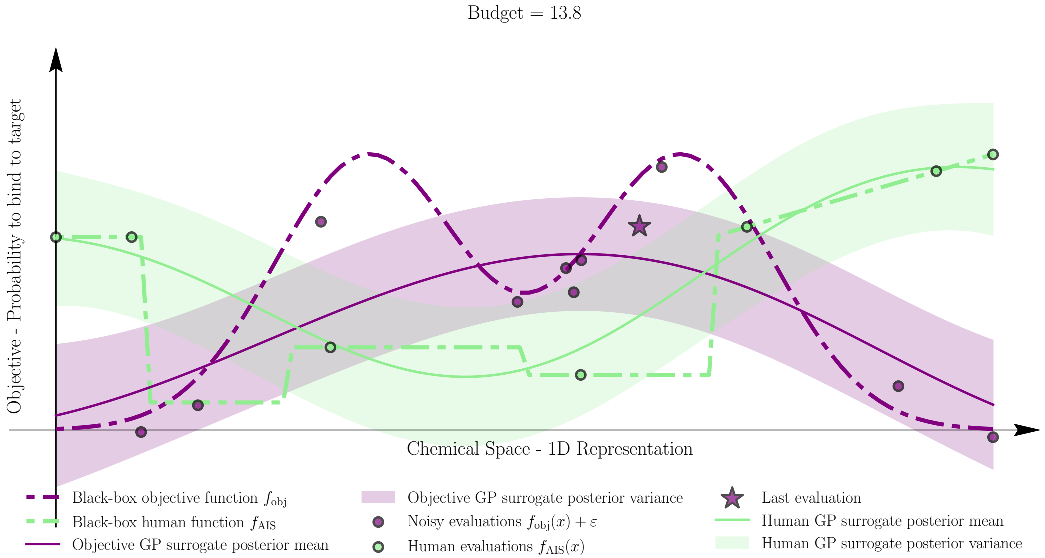
Multi Fidelity Bayesian Optimization with Unreliable Sources



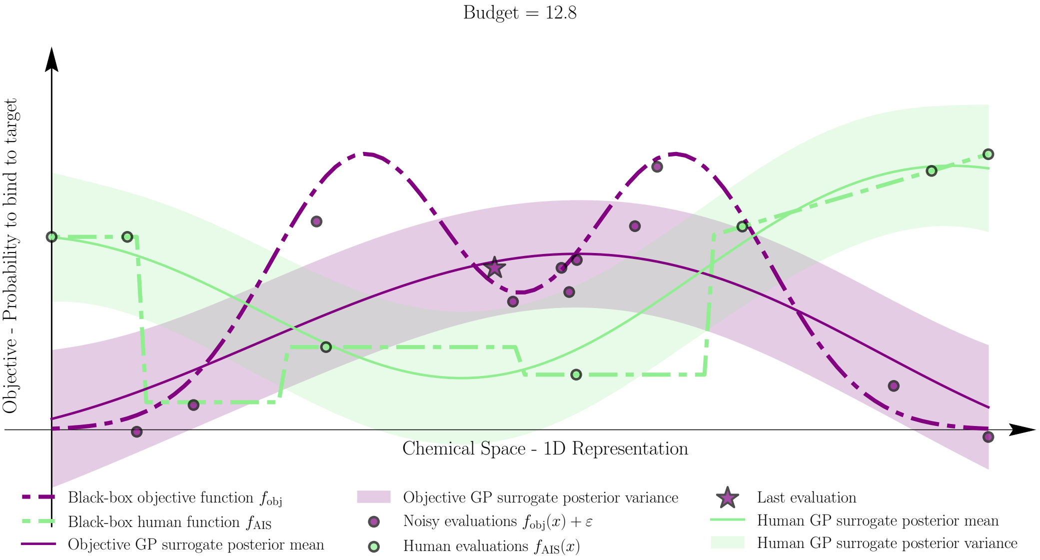
Multi Fidelity Bayesian Optimization with Unreliable Sources



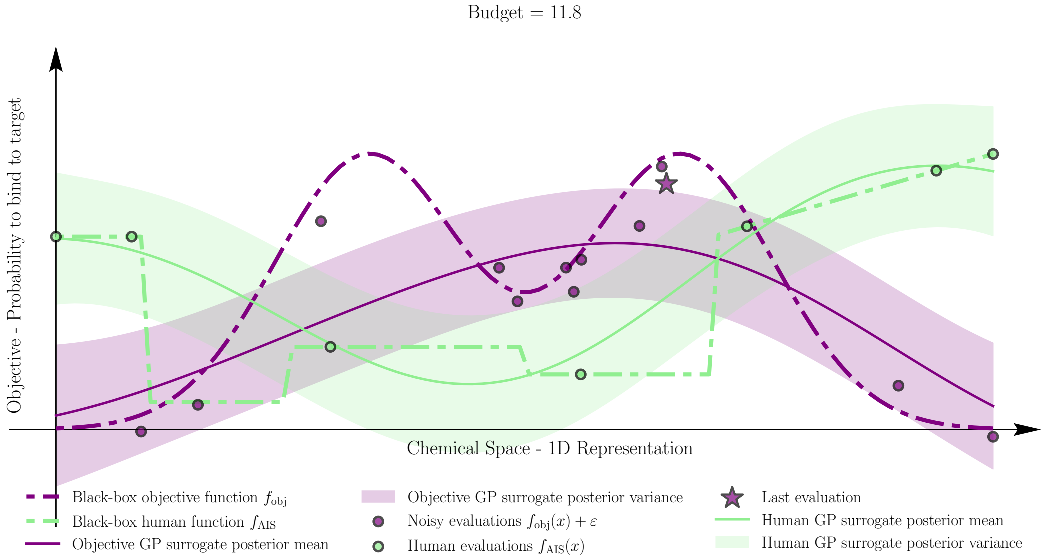
Multi Fidelity Bayesian Optimization with Unreliable Sources



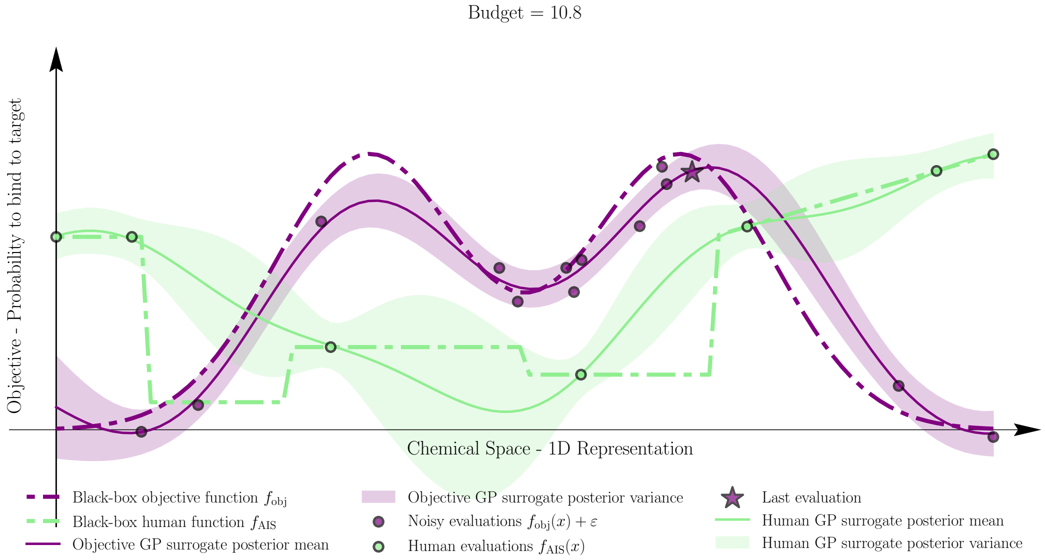
Multi Fidelity Bayesian Optimization with Unreliable Sources



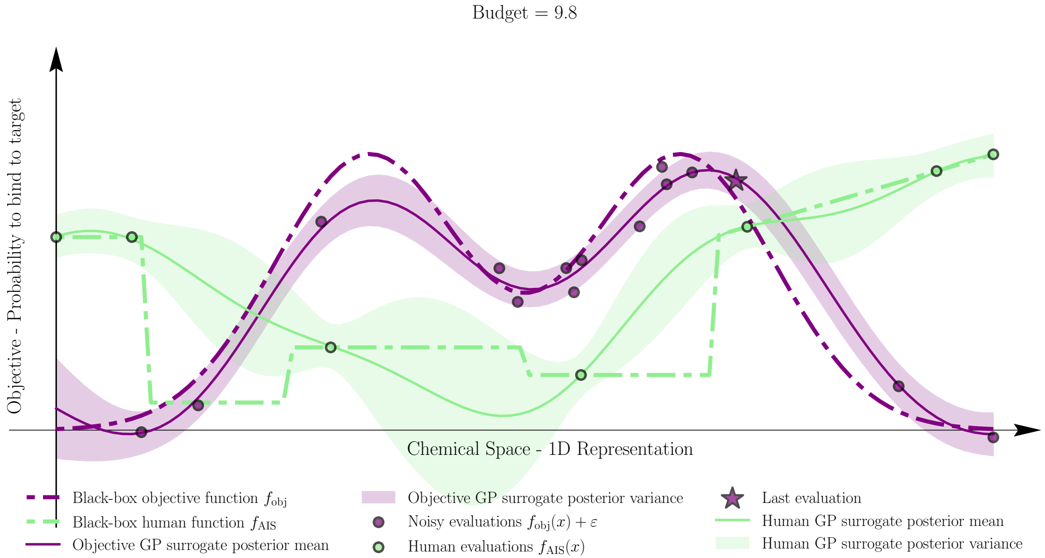
Multi Fidelity Bayesian Optimization with Unreliable Sources



Multi Fidelity Bayesian Optimization with Unreliable Sources



Multi Fidelity Bayesian Optimization with Unreliable Sources



A 6D case: the Hartmann problem

For $A, P \in \mathscr{M}_{4,6}(\mathbb{R})$ two matrices, $\mathrm{x} \in [0,1]^6$ $\ell \in [0,1]$, we define

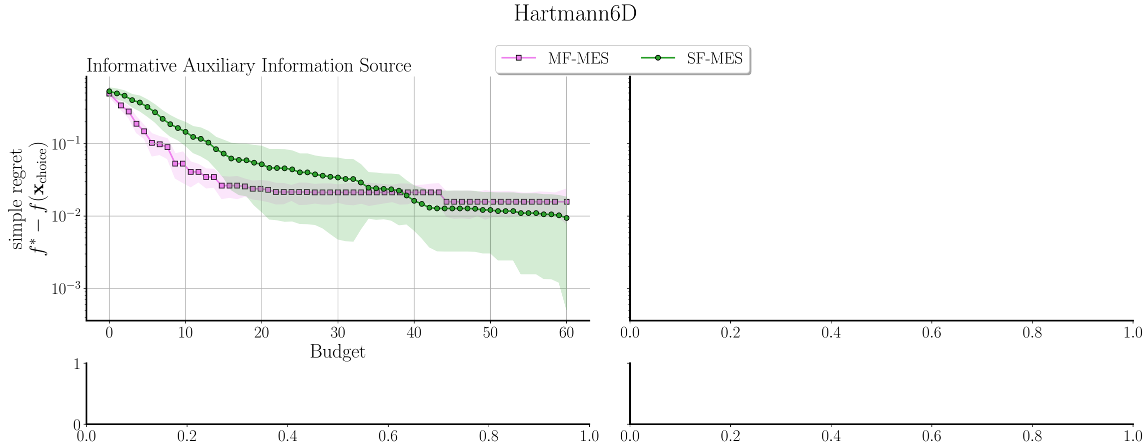
$$f^{(\ell)}(\mathrm{x}) = -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}(x_j - P_{ij})\right)$$

$$\alpha = (1.0 - 0.1(1-\ell), 1.2, 3.0, 3.2)^T$$

The objective is $f^{(1)}$, with query cost $\lambda_{\text{obj}} = 1$. BO is performed in 3 scenarios, using:

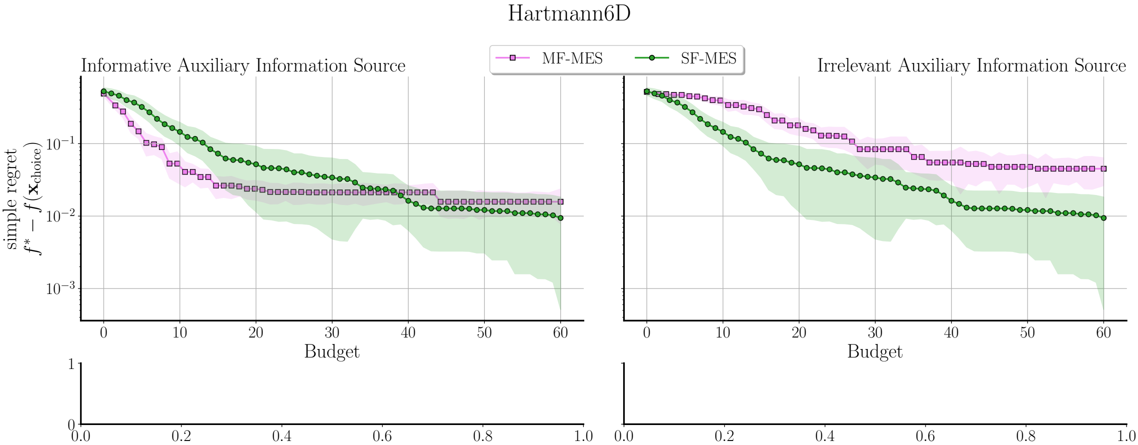
- Only $f^{(1)}$ (Single-Fidelity BO)
- $f^{(1)}$ and an **informative** AIS: $f^{(0.2)}$, $\lambda_{\text{AIS}} = 0.2$ (Multi-Fidelity BO)
- $f^{(1)}$ and an **irrelevant** AIS: $f(\mathrm{x}) = \sum_{i=1}^{5}\left(100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right)$, $\lambda_{\text{AIS}} = 0.2$

Multi-Fidelity BO is not robust to unreliable Information Sources



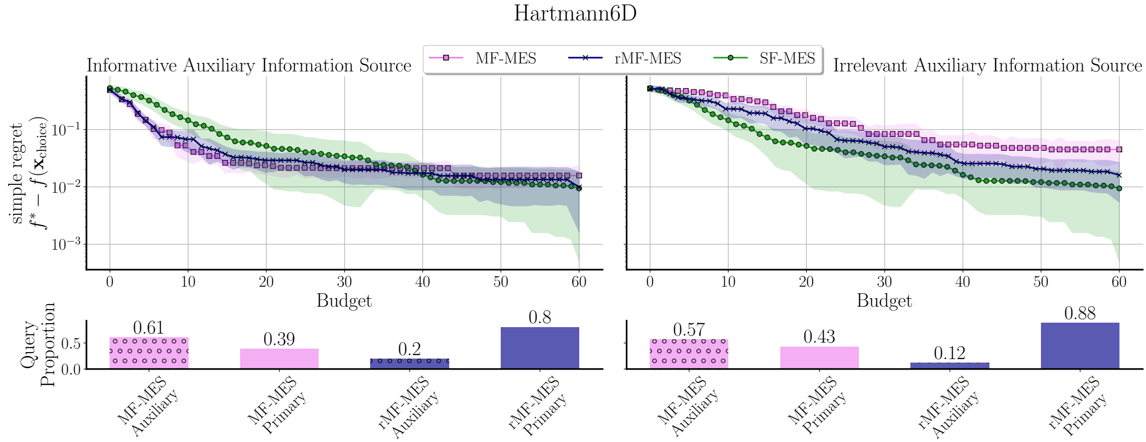
- SF-MES: Single-Fidelity BO w/ Maximum Entropy Search Acquisition Function
- MF-MES: Multi-Fidelity BO w/ Maximum Entropy Search Acquisition Function

Multi-Fidelity BO is not robust to unreliable Information Sources



- SF-MES: Single-Fidelity BO w/ Maximum Entropy Search Acquisition Function
- MF-MES: Multi-Fidelity BO w/ Maximum Entropy Search Acquisition Function

Multi-Fidelity BO is not robust to unreliable Information Sources



- Main aim of our contribution: **robustness** to irrelevant AIS...
- ...While still accelerating convergence for relevant AIS

Introducing robust MFBO (rMFBO)

We modify the BO loop with a **building block added on top of any** MFBO method

- Two separate GPs: $\mu_{\mathsf{SF}}, \sigma_{\mathsf{SF}}$ using objective queries only $\mathscr{D}^{\mathsf{pSF}}$, and a MOGP $\mu_{\mathsf{MF}}, \sigma_{\mathsf{MF}}$ using $\mathscr{D}^{\mathsf{MF}}$

$$(\mathrm{x}_t^{\mathsf{MF}}, \ell_t) = \underset{\mathrm{x}\in\mathscr{X}, \ell\in\{\mathsf{obj}, \mathsf{AIS}\}}{\operatorname{argmax}} \alpha(\mathrm{x}, \ell | \mu_{\mathsf{MF}}, \sigma_{\mathsf{MF}}, \mathscr{D}^{\mathsf{MF}})$$

$$(\mathrm{x}_t^{\mathsf{pSF}}, \mathsf{obj}) = \underset{\mathrm{x}\in\mathscr{X}}{\operatorname{argmax}}\, \alpha(\mathrm{x} | \mu_{\mathsf{SF}}, \sigma_{\mathsf{SF}}, \mathscr{D}^{\mathsf{pSF}})$$

- $\sigma_{\mathsf{MF}}(\mathrm{x}_t^{\mathsf{pSF}}, \mathsf{obj}) \leq c_1$ $\longrightarrow$ do I trust my joint model at the objective?
- $s(\mathrm{x}_t^{\mathsf{MF}}, \ell_t) \geq c_2$ $\longrightarrow$ is my joint model suggestion informative enough?

- Upon satisfaction: query $(\mathrm{x}_t^{\mathsf{MF}}, \ell_t)$ and add *pseudo-observation* of objective: $\mathscr{D}^{\mathsf{pSF}} \leftarrow (\mathrm{x}_t^{\mathsf{pSF}}, \mu_{\mathsf{MF}}(\mathrm{x}_t^{\mathsf{pSF}}, \mathsf{obj}))$ $\longrightarrow$ what if we had queried the objective?
- Otherwise, query $(\mathrm{x}_t^{\mathsf{pSF}}, \mathsf{obj})$

Theorem: rMFBO regret can be tied to that of SFBO

Assumptions:

- f_{obj} is drawn from a GP with zero-mean and covariance function $\kappa(\mathrm{x}, \mathrm{x}')$
- κ is known and at least twice differentiable
- $\mathbb{P}\left(\sup_{\mathrm{x} \in \mathscr{X}} \left|\frac{\partial f_{\text{obj}}}{\partial x_j}\right| > L\right) \leq a e^{-(L/b_j)^2}, \quad \forall\ j \in [\![d]\!]$ for $a, b_j > 0$ (satisfied for RBF kernel)
- The acquisition function is twice differentiable

Theorem: rMFBO regret can be tied to that of SFBO

Assumptions:

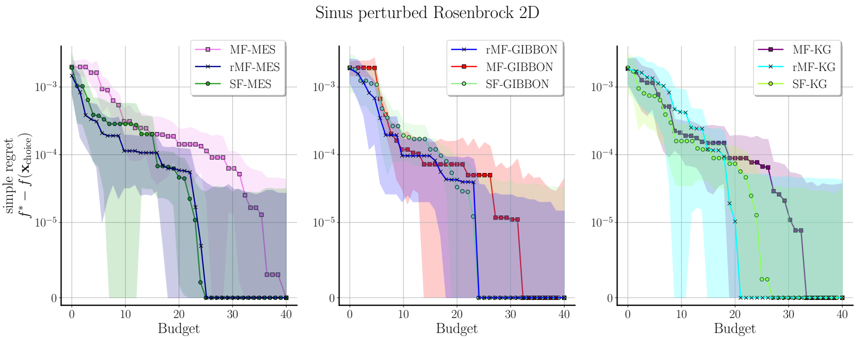
- f_{obj} is drawn from a GP with zero-mean and covariance function $\kappa(\mathrm{x}, \mathrm{x}')$
- κ is known and at least twice differentiable
- $\mathbb{P}\left(\sup_{\mathrm{x}\in\mathcal{X}} \left|\frac{\partial f_{\text{obj}}}{\partial x_j}\right| > L\right) \leq a e^{-(L/b_j)^2}, \quad \forall\, j \in [\![d]\!]$ for $a, b_j > 0$ (satisfied for RBF kernel)
- The acquisition function is twice differentiable

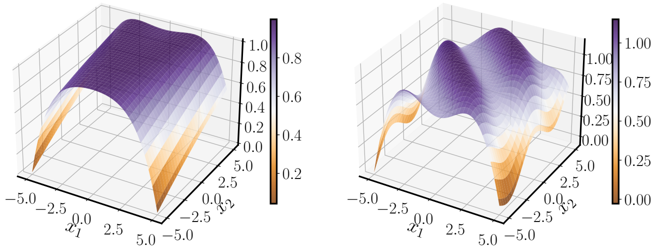
Theorem: $R(\Lambda, \mathrm{x}_T^{\text{rMF}}) \leq R(\Lambda, \mathrm{x}_T^{\text{SF}}) + \varepsilon \max\left\{T \hat{M}_T d^{T+1}, 2\right\}$ w/ prob $\geq q\left(1 - da \exp(-\frac{1}{b^2})\right)$

$c_1(\varepsilon, q) = \frac{\varepsilon}{\sqrt{-2\log(1-q)}}$. Theorem does not depend on c_2.

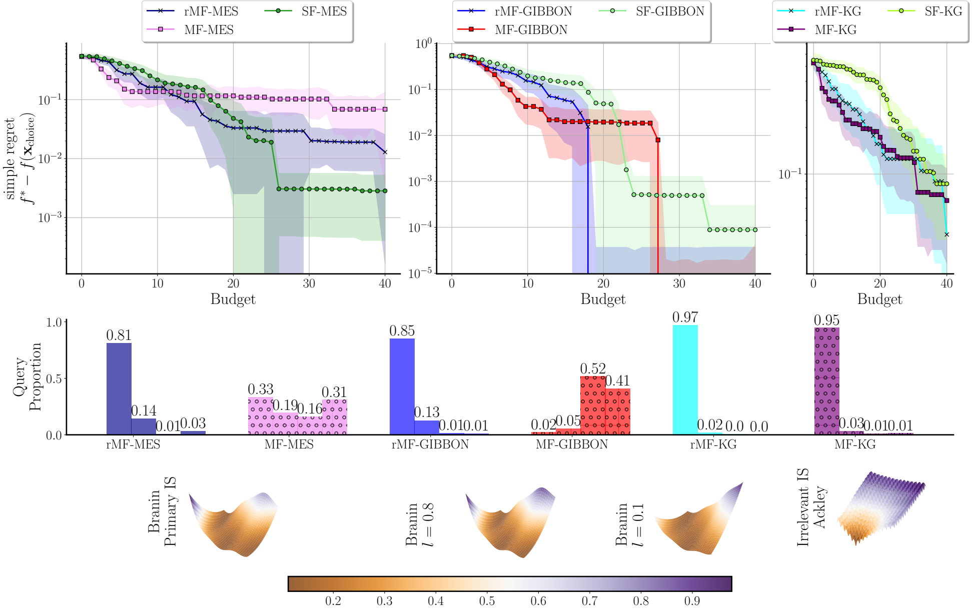
For any AIS, the difference in regrets achieved by SFBO and rMFBO can be bounded.

Results on 2D case





Multiple Information Sources of varying relevance - 2D case



Perspectives

For the method itself:

- Theoretical insights on how to set hyperparameter c_2
- Try against more joint models

From an applied point of view:

- Carry user studies
- Concrete application with human experts