Multi-Fidelity Bayesian Optimization with Unreliable Information Sources

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

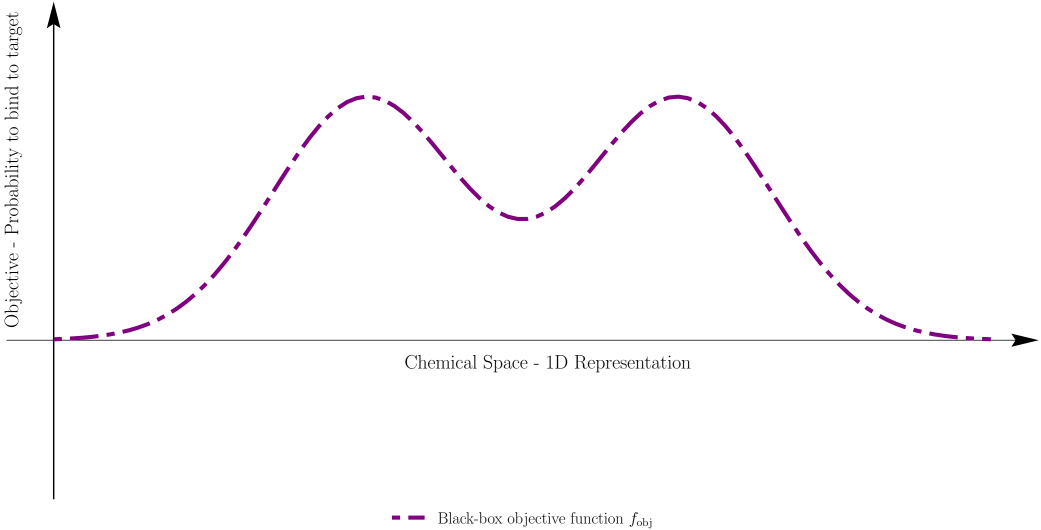
Aalto University

AISTATS 2023

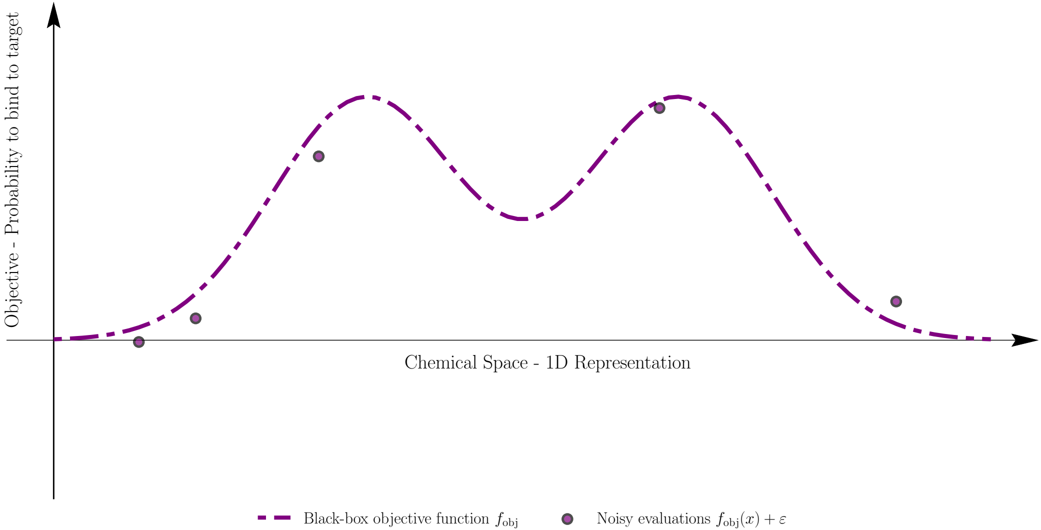
March 25th, 2025

Bayesian Optimization 101

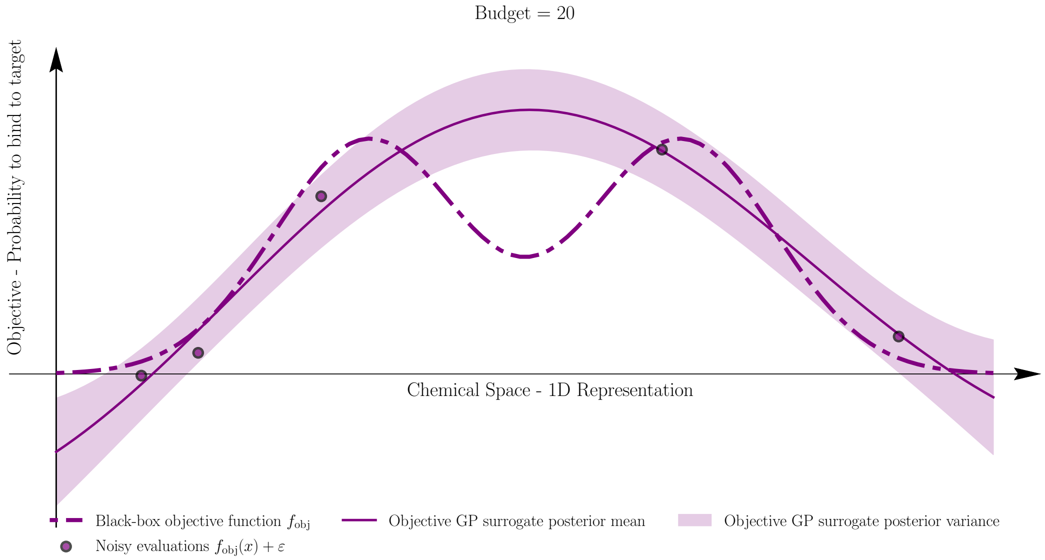
Bayesian Optimization 101



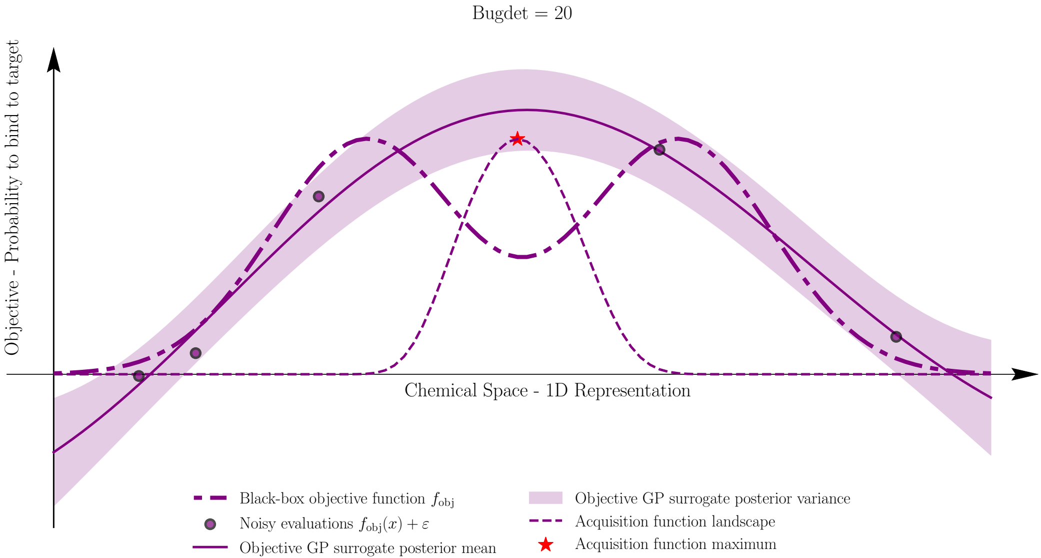
Bayesian Optimization 101



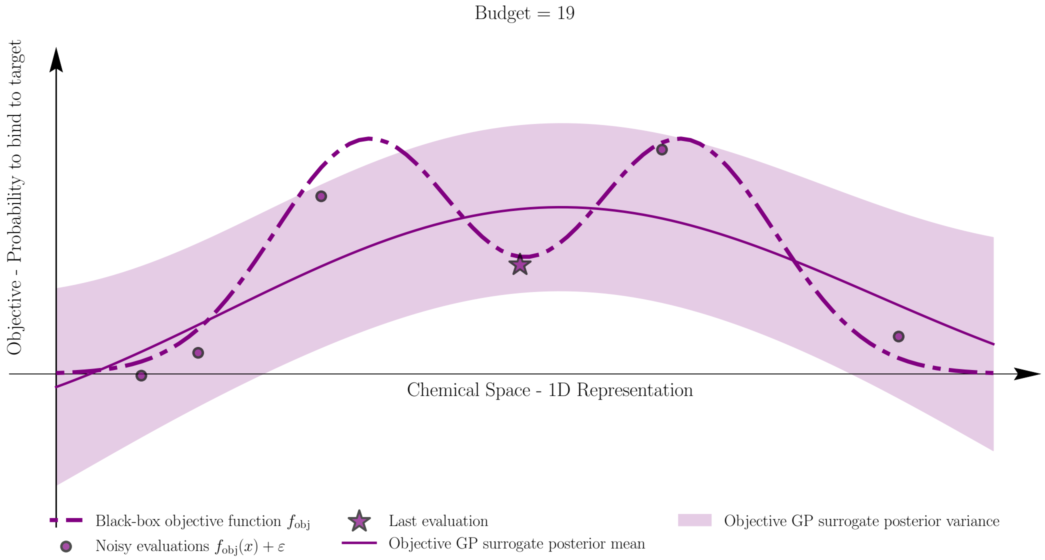
Bayesian Optimization 101



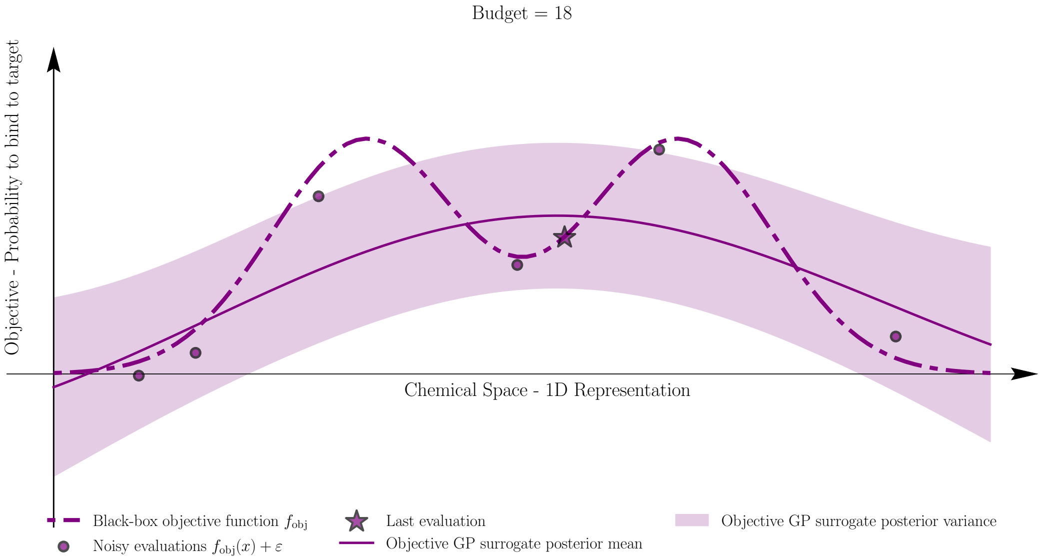
Bayesian Optimization 101



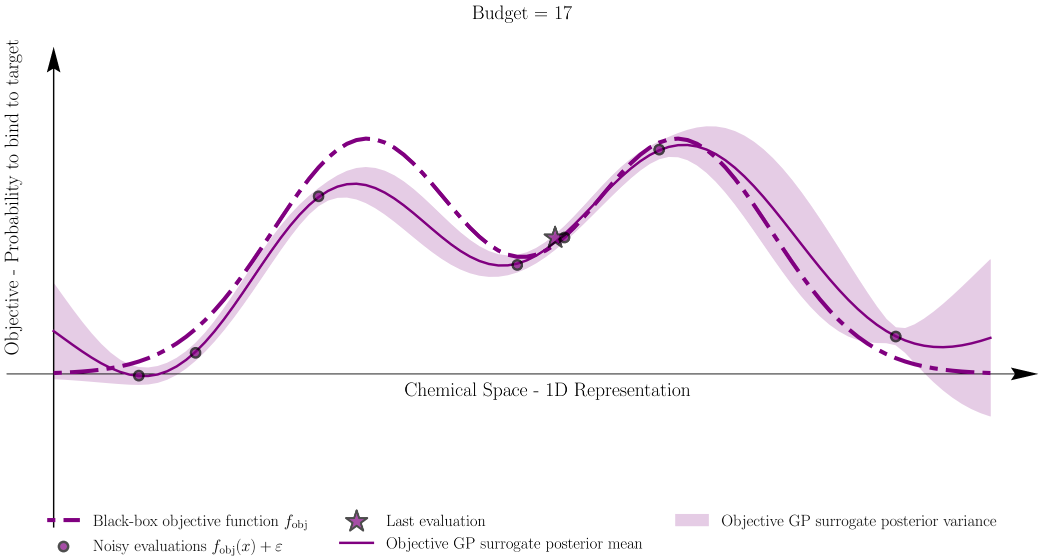
Bayesian Optimization 101



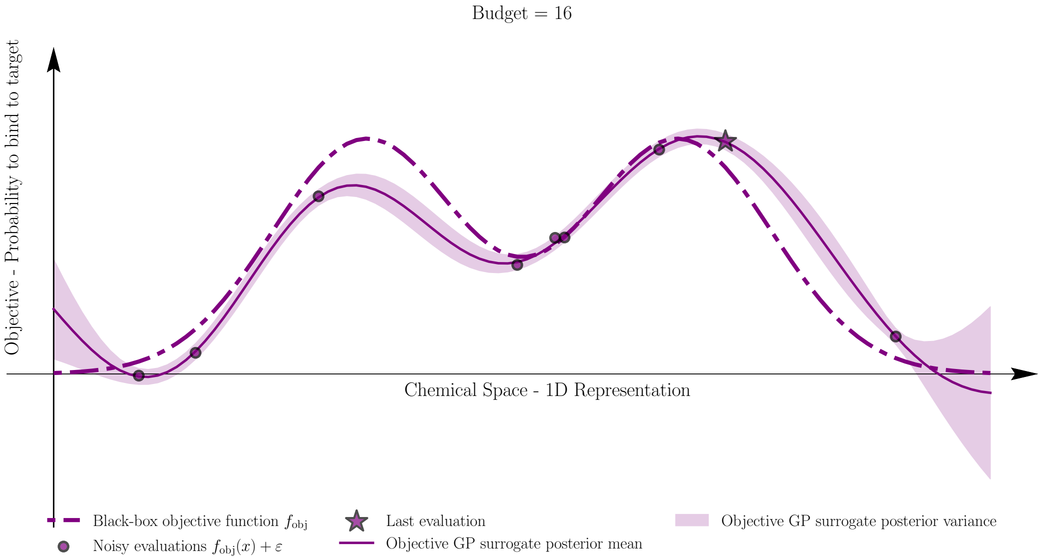
Bayesian Optimization 101



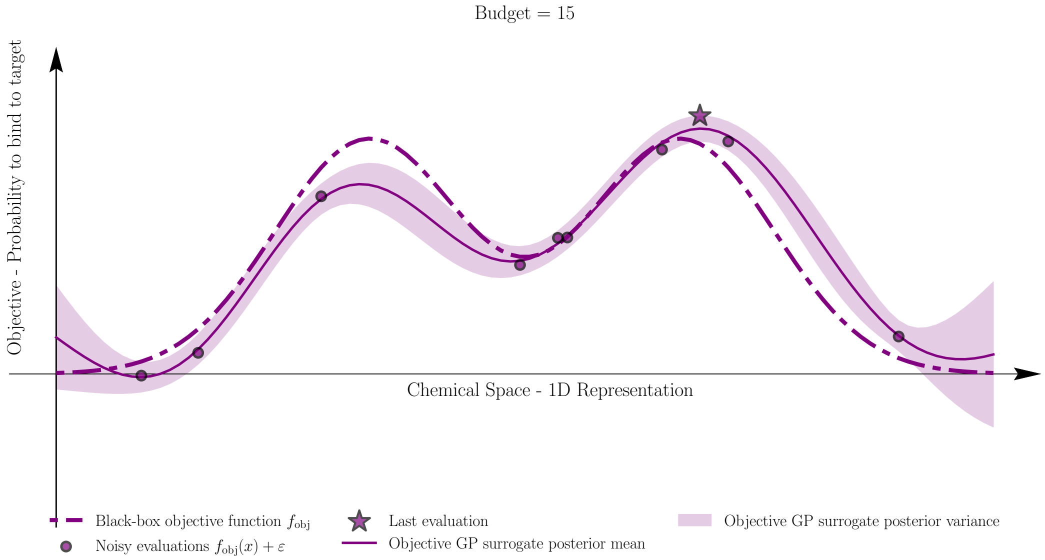
Bayesian Optimization 101



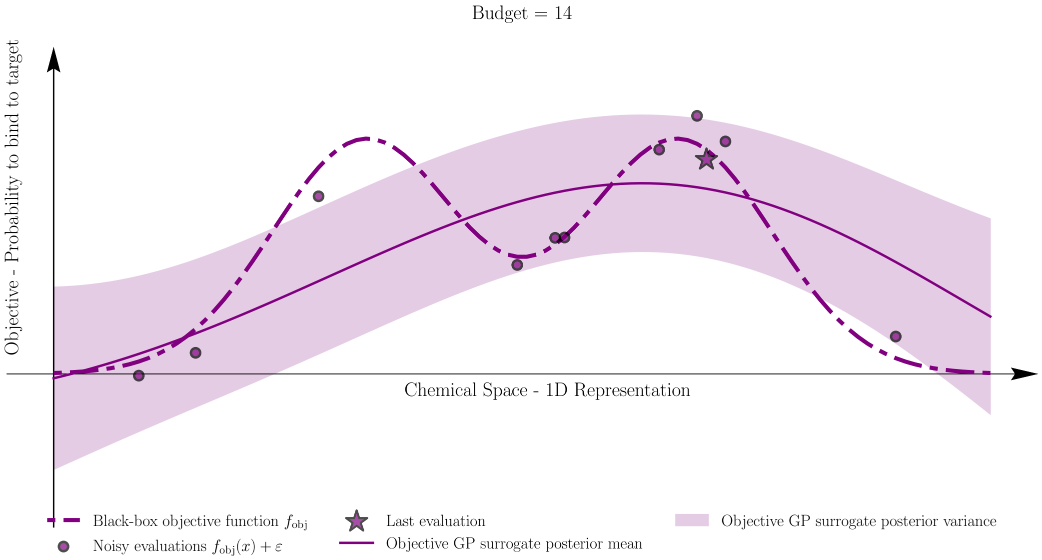
Bayesian Optimization 101



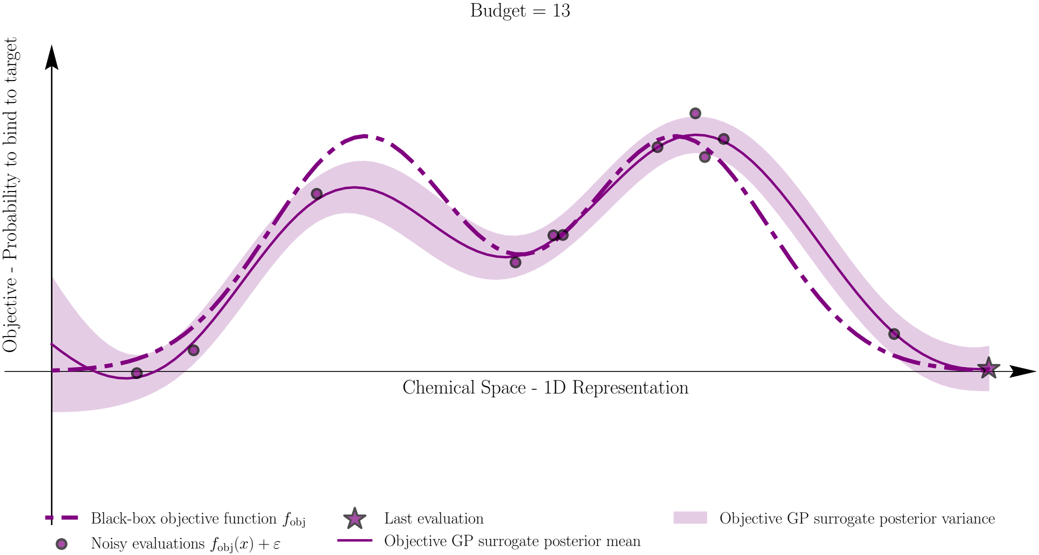
Bayesian Optimization 101



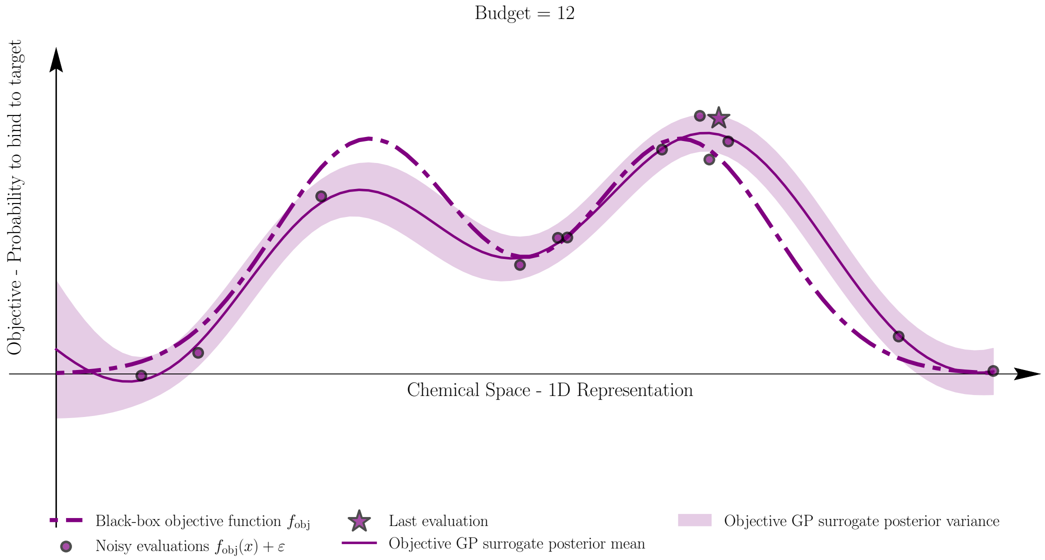
Bayesian Optimization 101



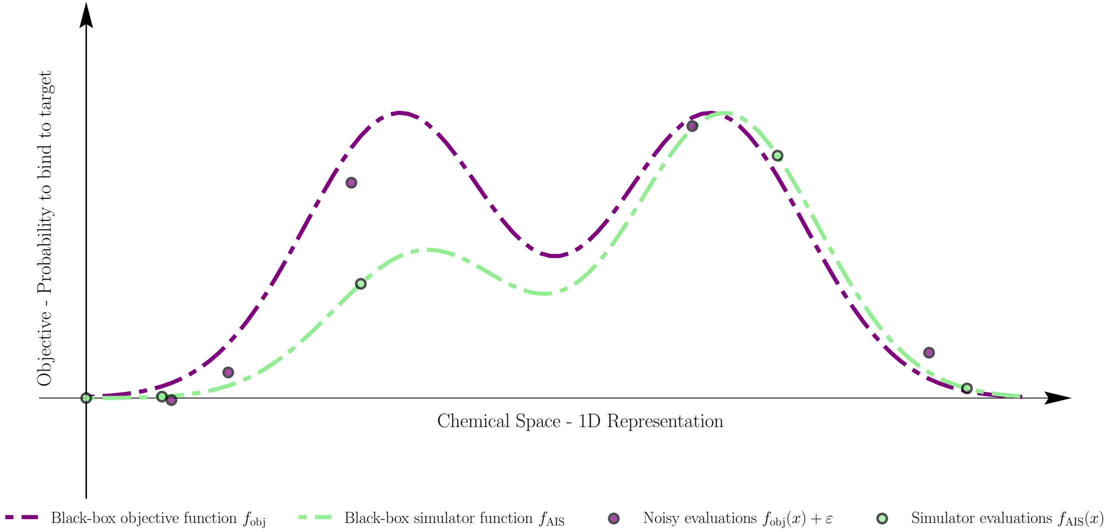
Bayesian Optimization 101



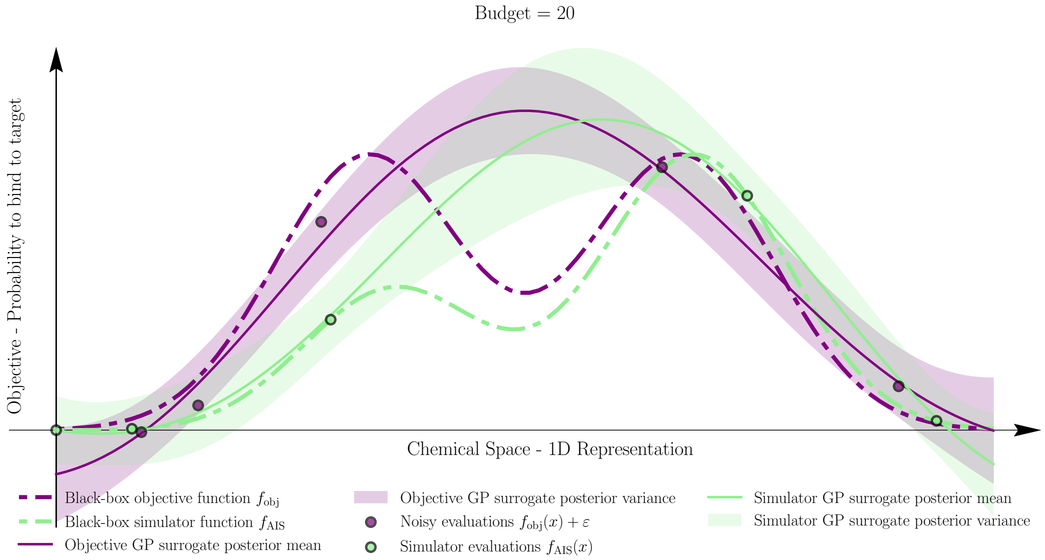
Bayesian Optimization 101



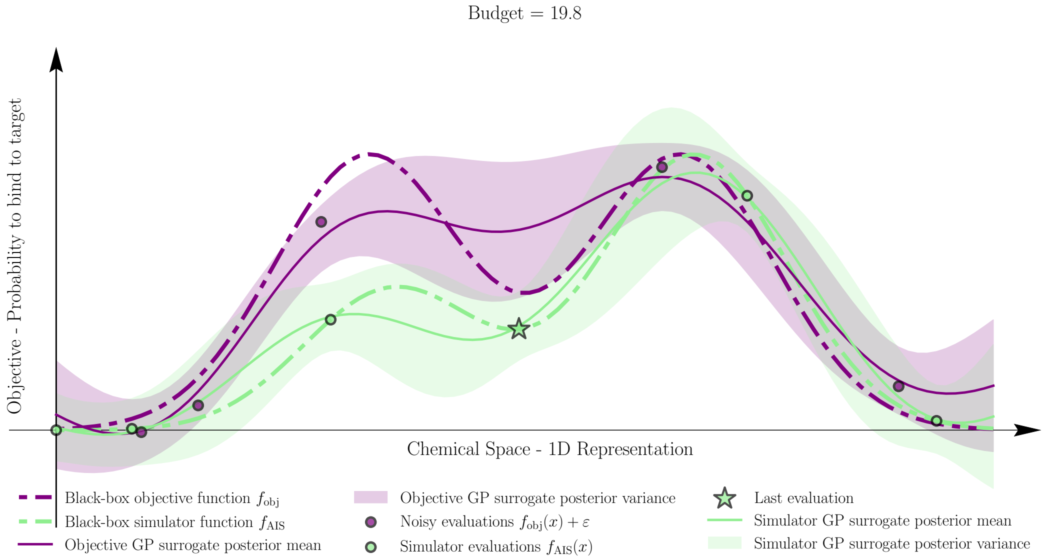
Multi Fidelity Bayesian Optimization 101



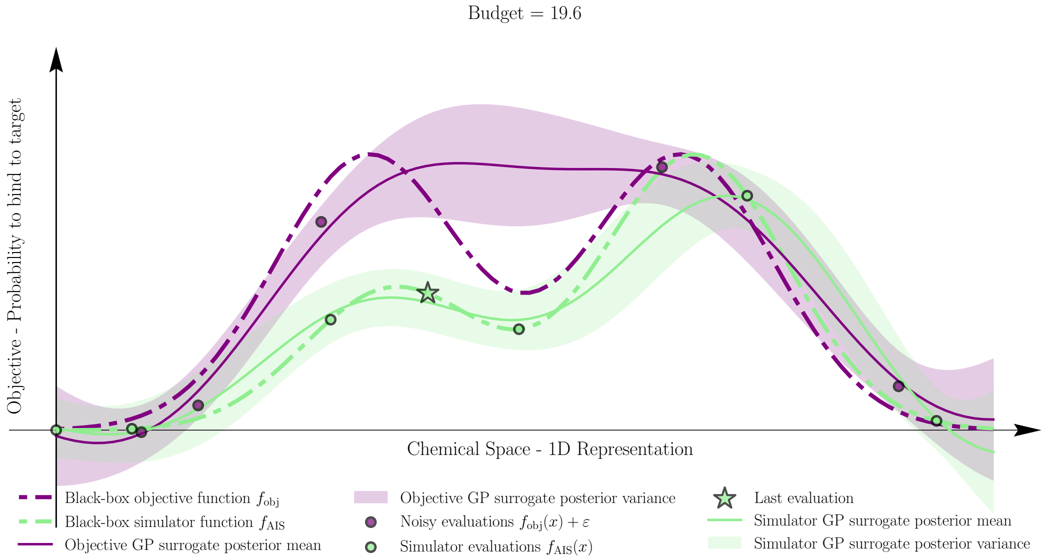
Multi Fidelity Bayesian Optimization 101



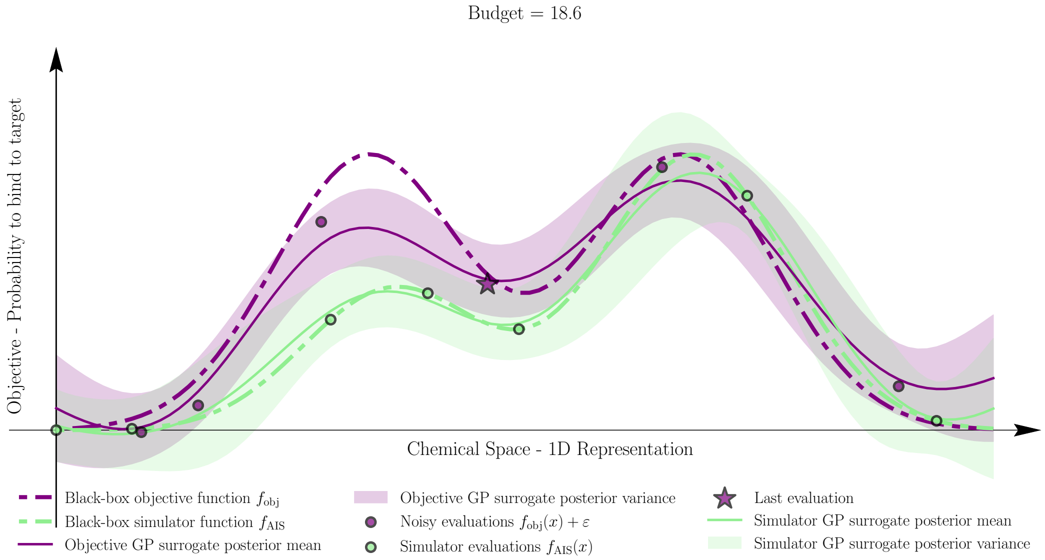
Multi Fidelity Bayesian Optimization 101



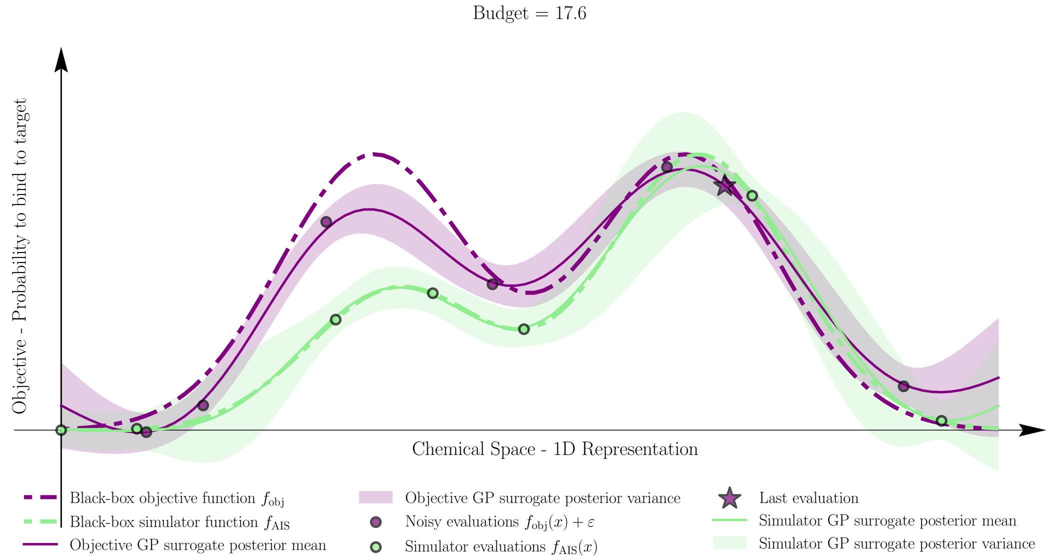
Multi Fidelity Bayesian Optimization 101



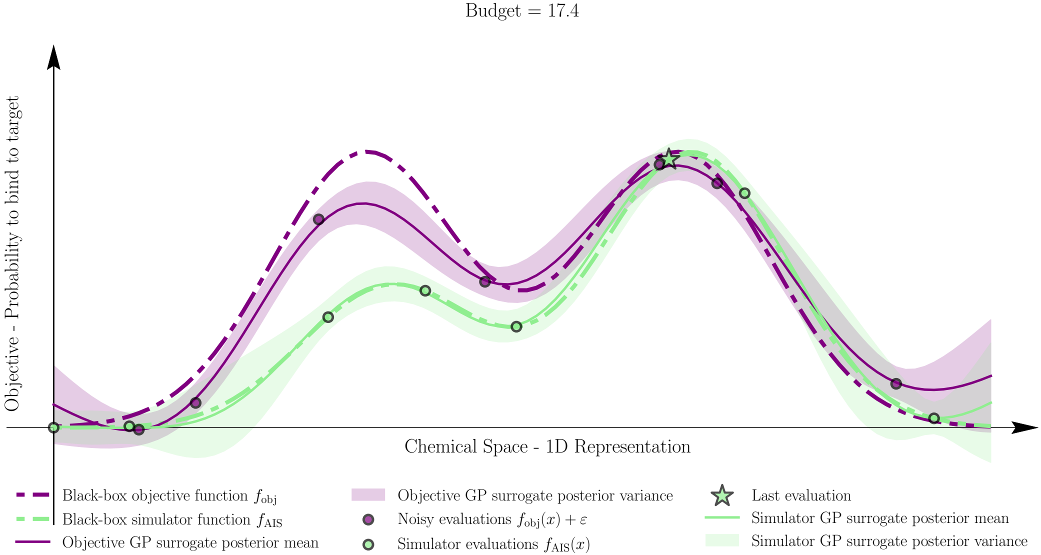
Multi Fidelity Bayesian Optimization 101



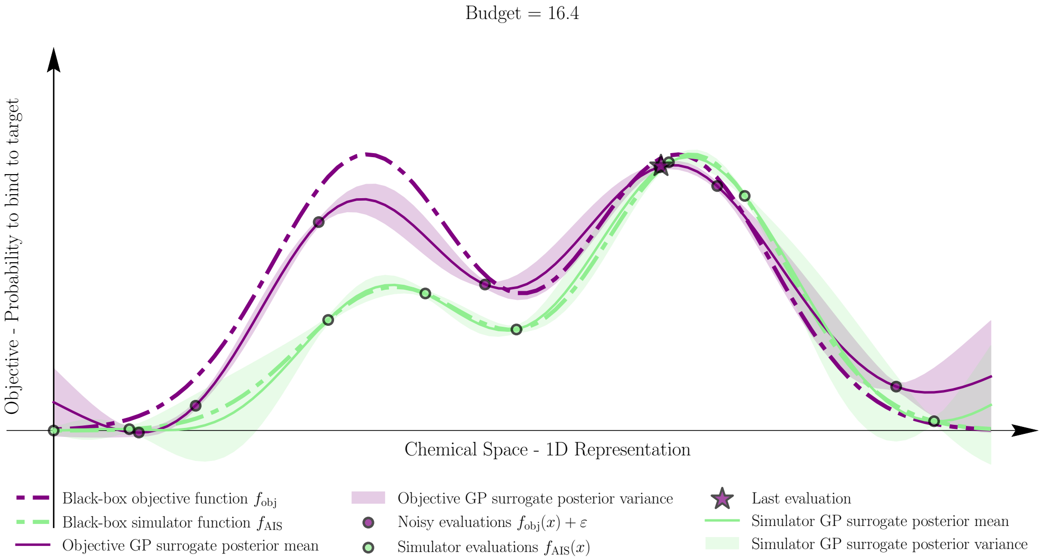
Multi Fidelity Bayesian Optimization 101



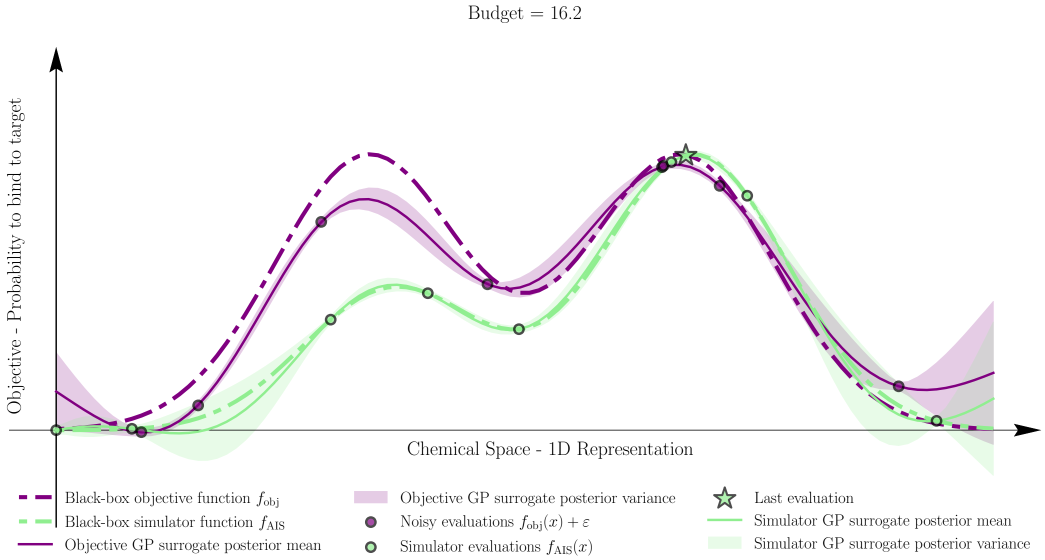
Multi Fidelity Bayesian Optimization 101



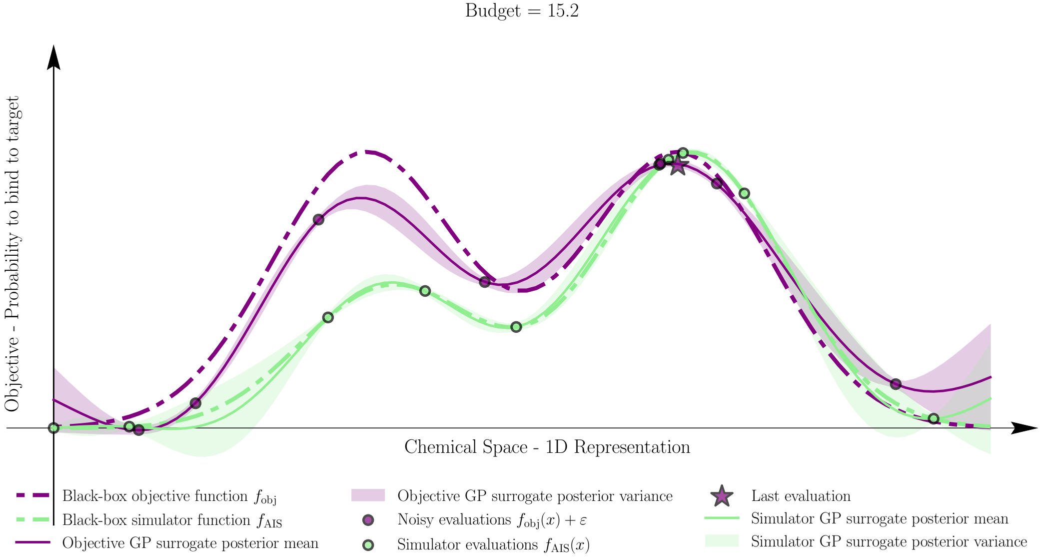
Multi Fidelity Bayesian Optimization 101



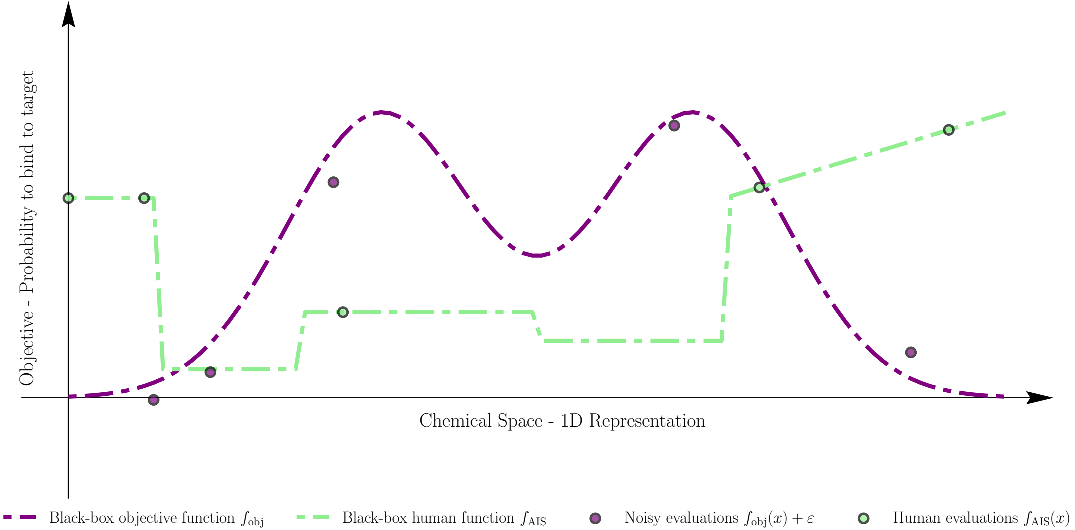
Multi Fidelity Bayesian Optimization 101



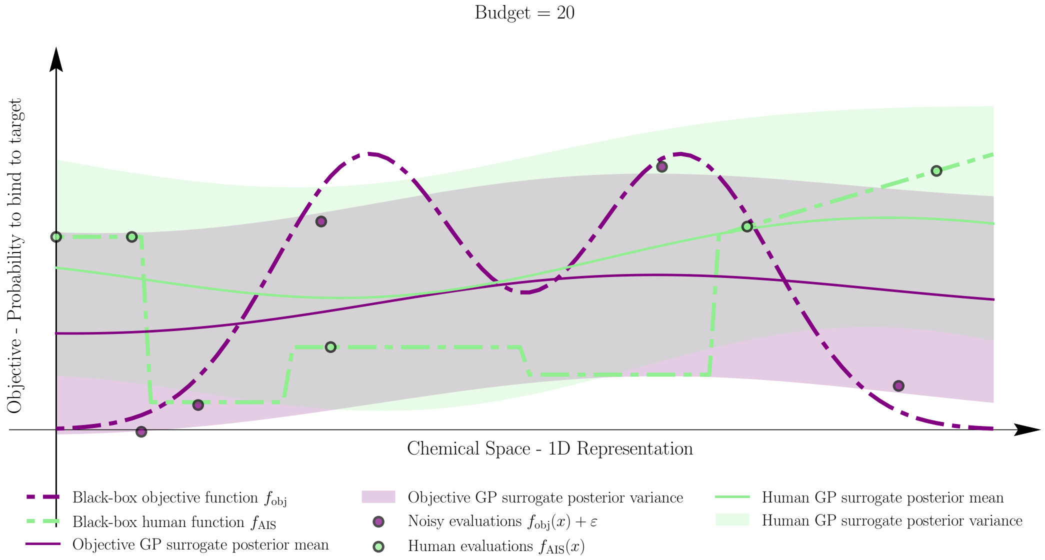
Multi Fidelity Bayesian Optimization 101



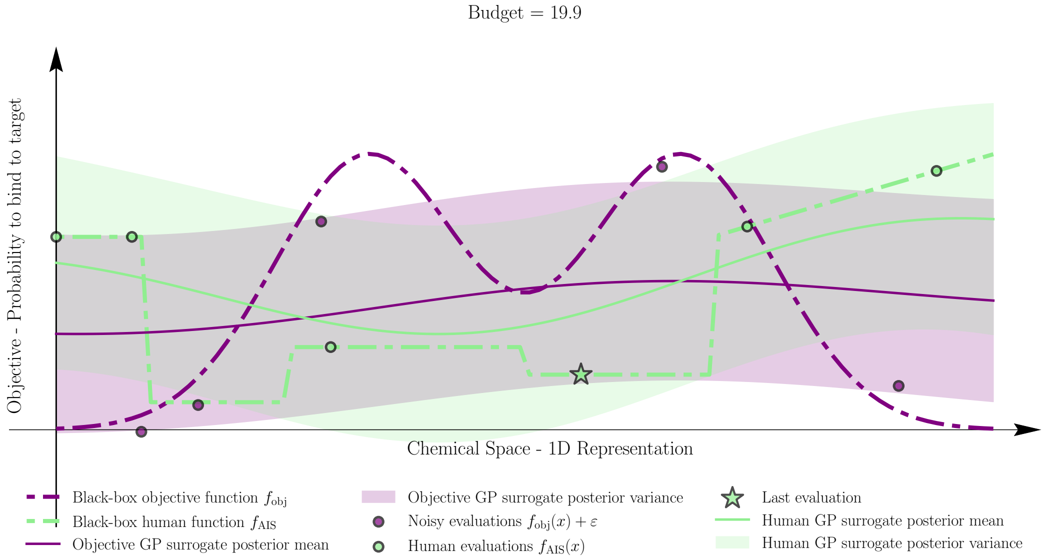
Multi Fidelity Bayesian Optimization with Unreliable Sources



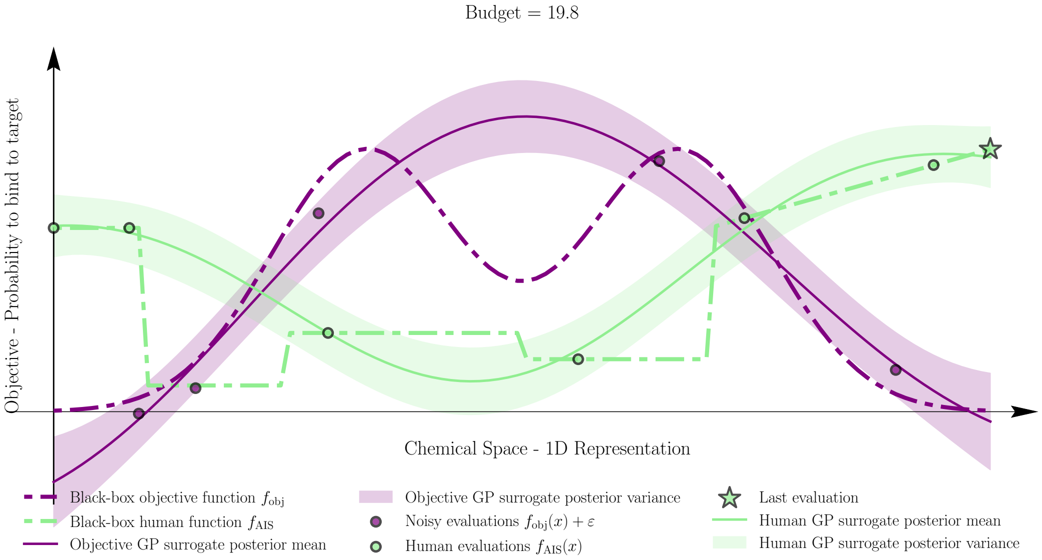
Multi Fidelity Bayesian Optimization with Unreliable Sources



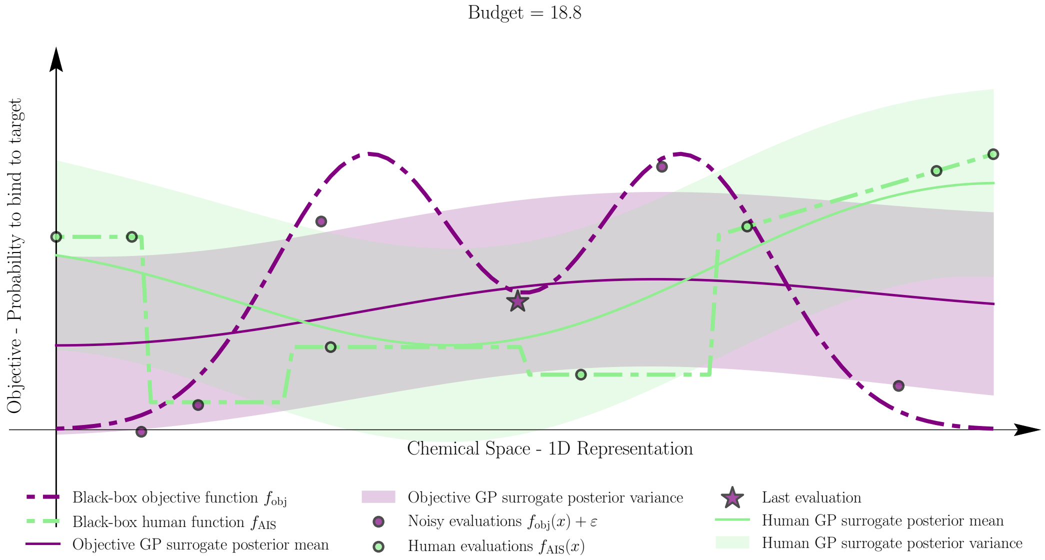
Multi Fidelity Bayesian Optimization with Unreliable Sources



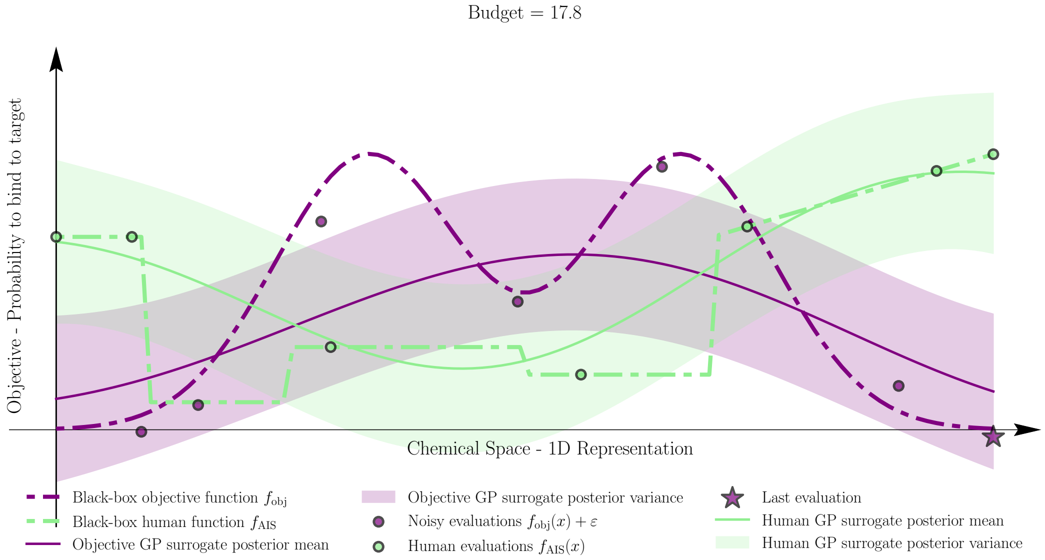
Multi Fidelity Bayesian Optimization with Unreliable Sources



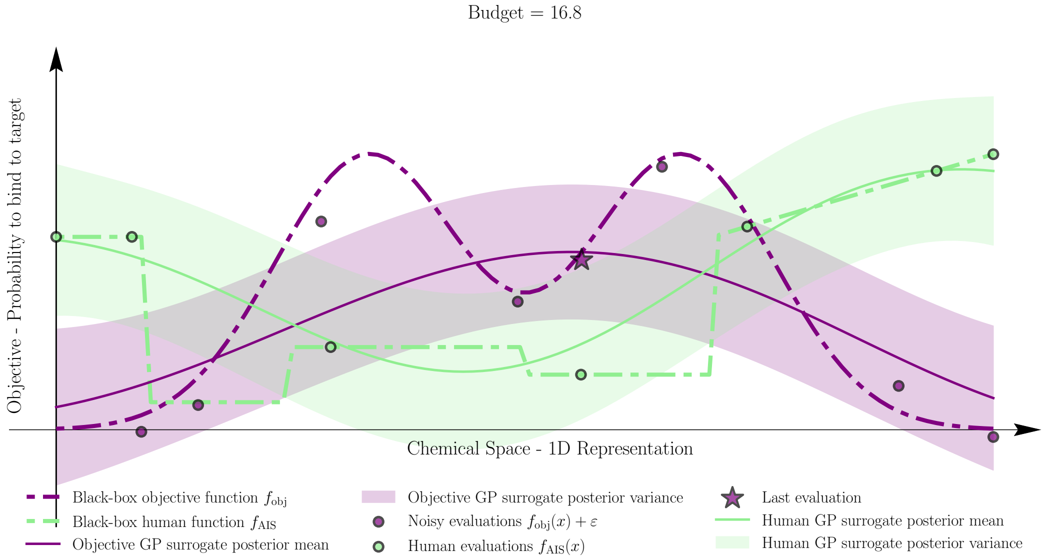
Multi Fidelity Bayesian Optimization with Unreliable Sources



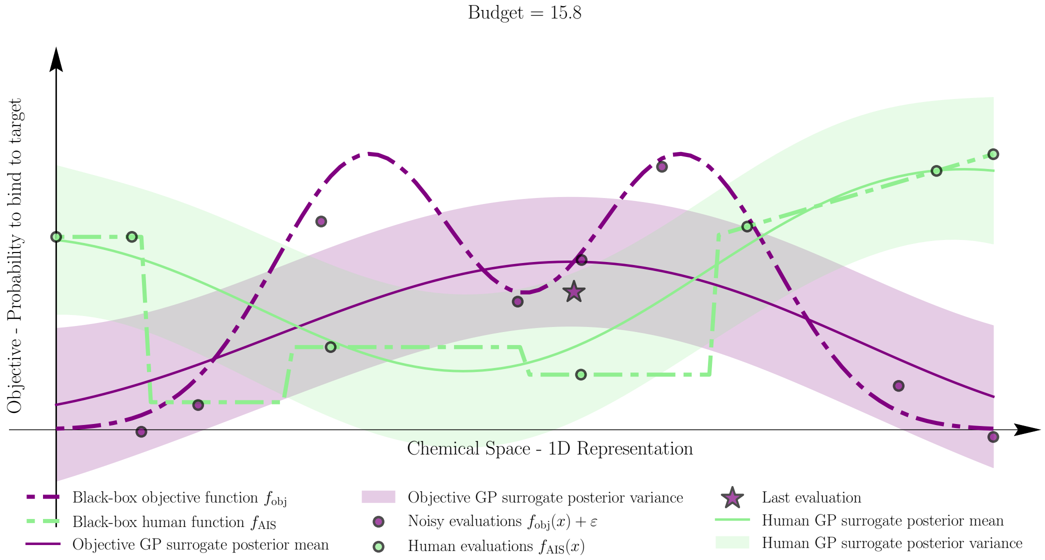
Multi Fidelity Bayesian Optimization with Unreliable Sources



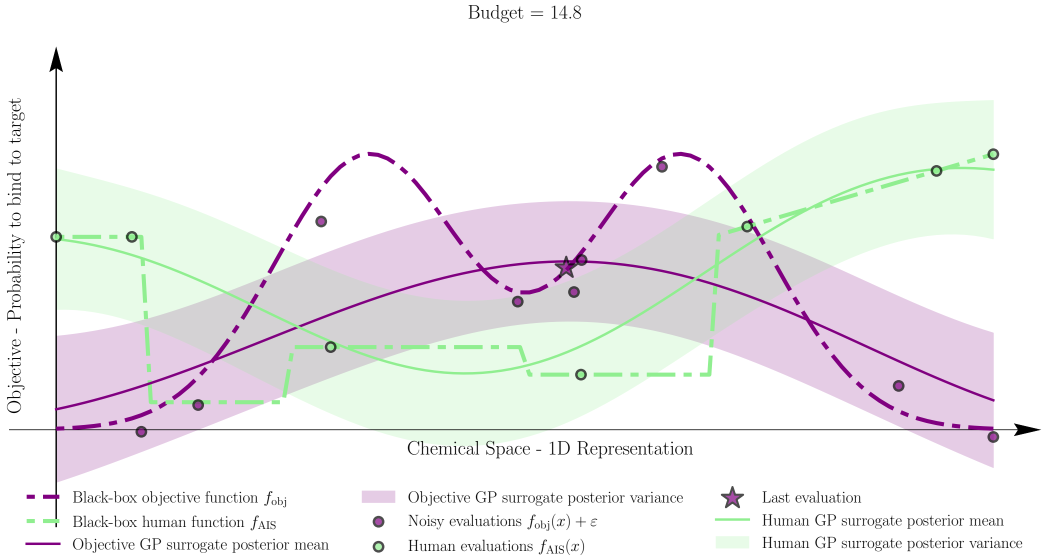
Multi Fidelity Bayesian Optimization with Unreliable Sources



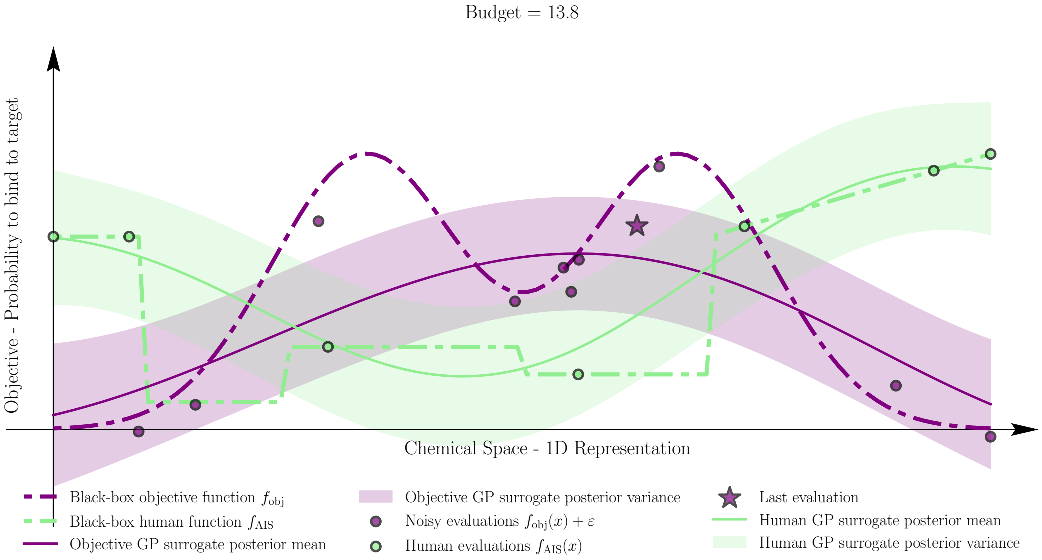
Multi Fidelity Bayesian Optimization with Unreliable Sources



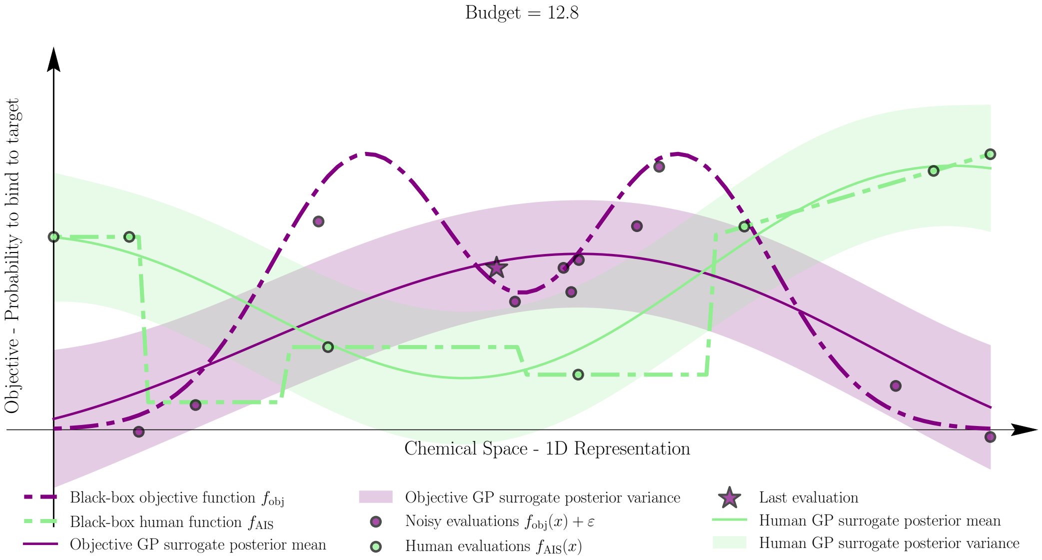
Multi Fidelity Bayesian Optimization with Unreliable Sources



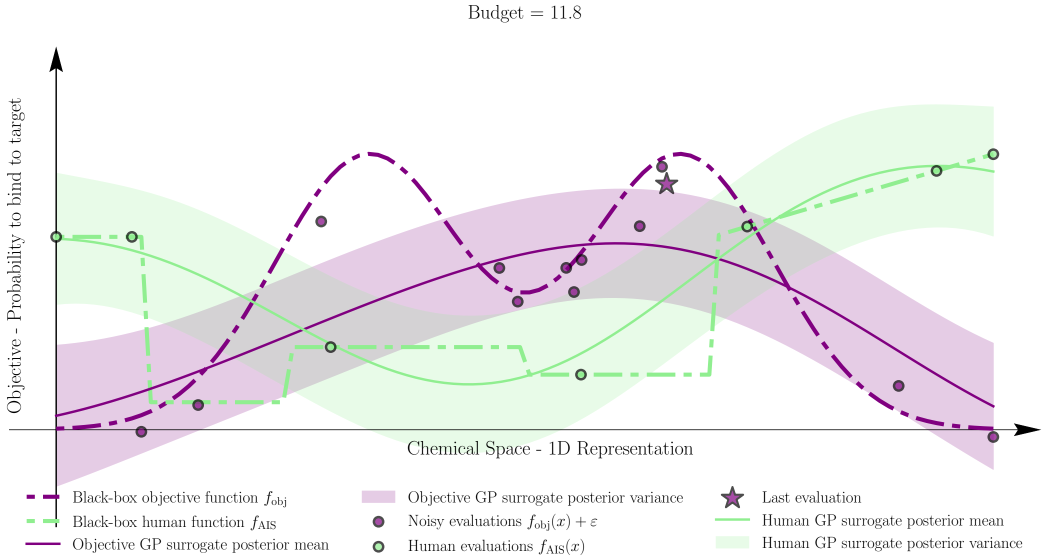
Multi Fidelity Bayesian Optimization with Unreliable Sources



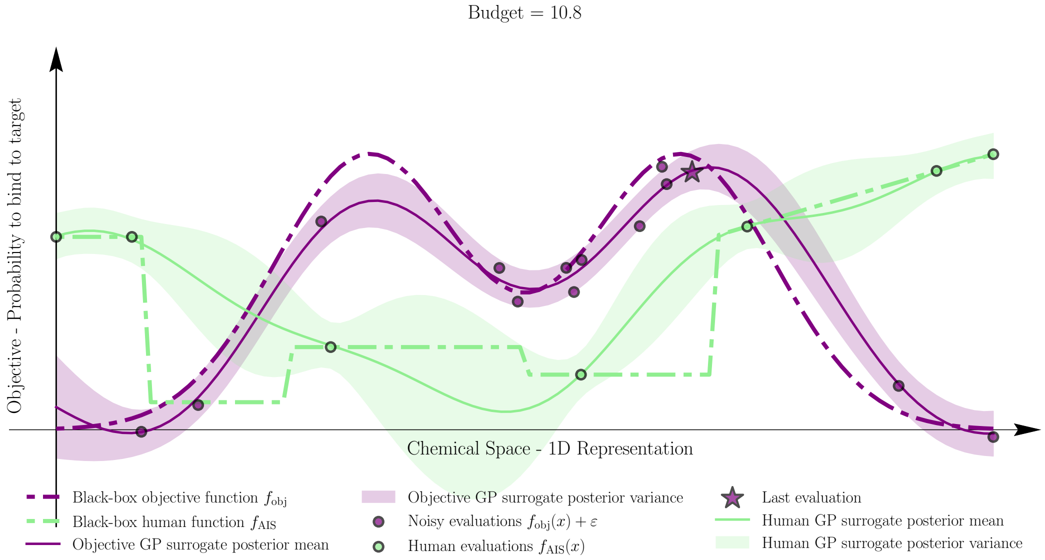
Multi Fidelity Bayesian Optimization with Unreliable Sources



Multi Fidelity Bayesian Optimization with Unreliable Sources

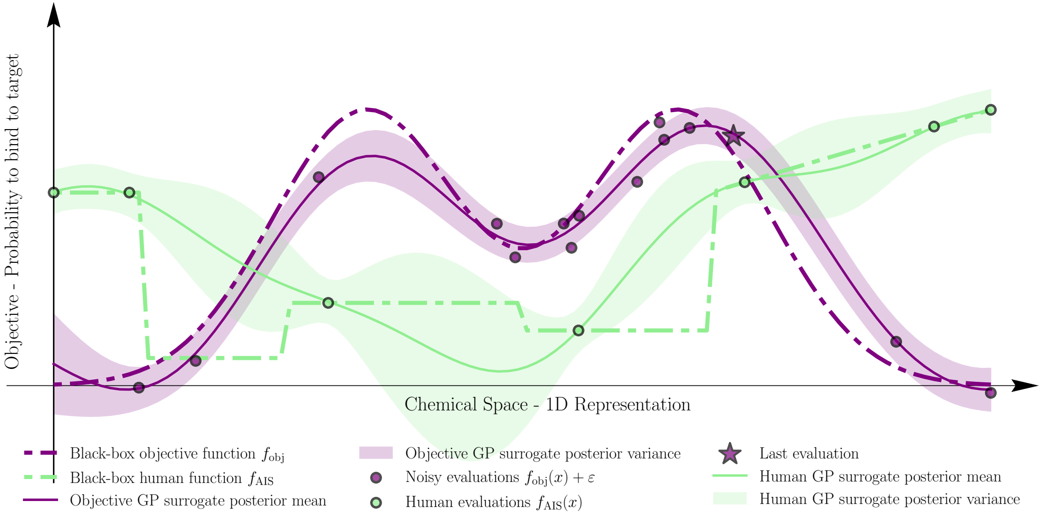


Multi Fidelity Bayesian Optimization with Unreliable Sources

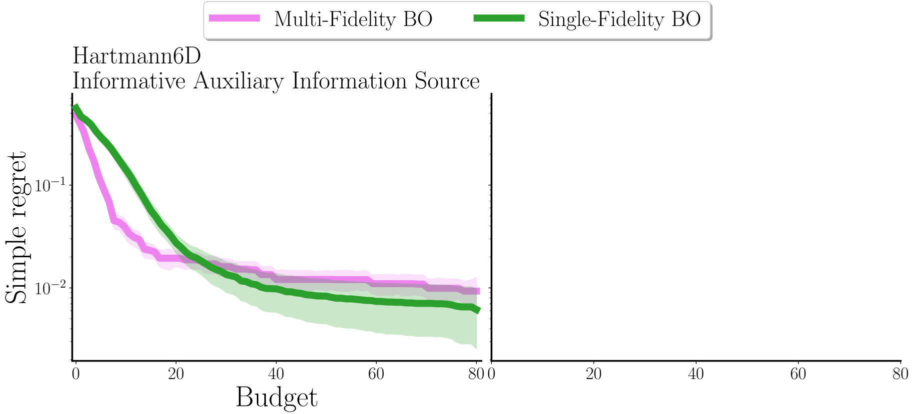


Multi Fidelity Bayesian Optimization with Unreliable Sources

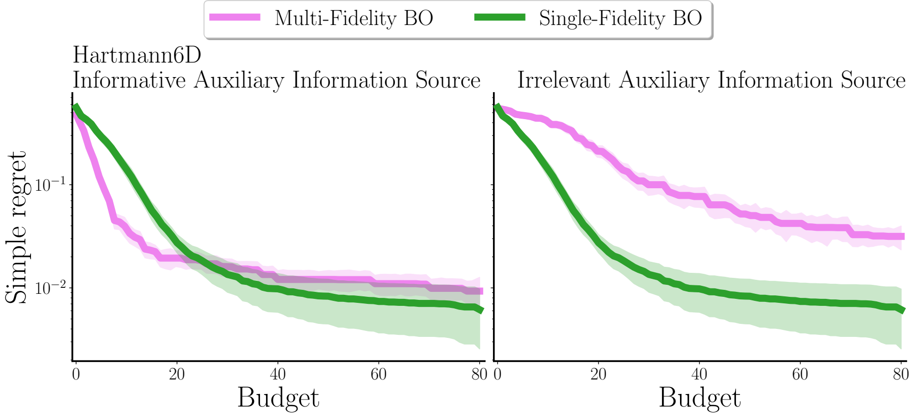
Budget = 9.8



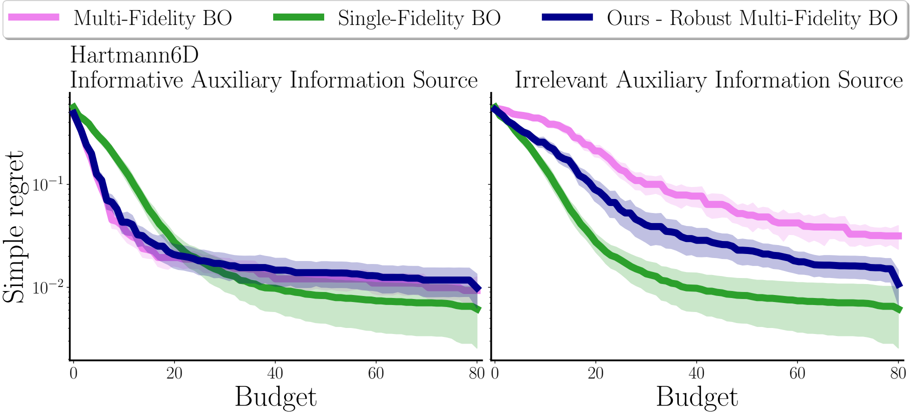
Multi-Fidelity BO is not robust to unreliable Information Sources



Multi-Fidelity BO is not robust to unreliable Information Sources



Multi-Fidelity BO is not robust to unreliable Information Sources



Not far from a well-known problem in transfer learning: **negative transfer**

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

Introducing robust MFBO (rMFBO)

- In a nutshell, carry MFBO ... **BUT** ...
- ...Keep track of what would have looked like the acquisition trajectory without AIS using a single-output GP $\Longrightarrow$ **HOW?**

At iteration t, choose between two queries:

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

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

If $\sigma_{\mathsf{MF}}(\mathrm{x}_t^{\mathsf{pSF}}, \mathsf{obj}) \leq c_1 \Longrightarrow$ joint model reliable at $\mathrm{x}_t^{\mathsf{pSF}}$, i.e. $\mu_{\mathsf{MF}}^{\mathsf{obj}}\left(\mathrm{x}_t^{\mathsf{pSF}}\right) \approx f^{\mathsf{obj}}\left(\mathrm{x}_t^{\mathsf{pSF}}\right)$

Therefore $\mathscr{D}^{\mathsf{pSF}} \leftarrow \left(\mathrm{x}_t^{\mathsf{pSF}}, \mu_{\mathsf{MF}}^{\mathsf{obj}}\left(\mathrm{x}_t^{\mathsf{pSF}}\right)\right)$: creating a *pseudo* single fidelity track

Summary

$$\sigma_{\mathsf{MF}}(\mathsf{x}_t^{\mathsf{pSF}}, \mathsf{obj}) \leq c_1 \implies \begin{cases} \text{pick } (\mathsf{x}_t^{\mathsf{MF}}, \ell_t) \\ \mathscr{D}^{\mathsf{MF}} \leftarrow (\mathsf{x}_t^{\mathsf{MF}}, f^{(\ell_t)}(\mathsf{x}_t^{\mathsf{MF}})) \\ \mathscr{D}^{\mathsf{pSF}} \leftarrow (\mathsf{x}_t^{\mathsf{pSF}}, \mu_{\mathsf{MF}}^{\mathsf{obj}}(\mathsf{x}_t^{\mathsf{pSF}})) \end{cases}$$

Summary

$$\sigma_{\mathsf{MF}}(\mathrm{x}_t^{\mathsf{pSF}}, \mathsf{obj}) \leq c_1 \implies \begin{cases} \text{pick } (\mathrm{x}_t^{\mathsf{MF}}, \ell_t) \\ \mathscr{D}^{\mathsf{MF}} \leftarrow (\mathrm{x}_t^{\mathsf{MF}}, f^{(\ell_t)}(\mathrm{x}_t^{\mathsf{MF}})) \\ \mathscr{D}^{\mathsf{pSF}} \leftarrow (\mathrm{x}_t^{\mathsf{pSF}}, \mu_{\mathsf{MF}}^{\mathsf{obj}}(\mathrm{x}_t^{\mathsf{pSF}})) \end{cases}$$

because $\mathrm{x}_t^{\mathsf{pSF}}$ only brings little information

Summary

$$\sigma_{\mathrm{MF}}(\mathrm{x}_t^{\mathrm{pSF}}, \mathrm{obj}) \leq c_1 \Longrightarrow \begin{cases} \text{pick } (\mathrm{x}_t^{\mathrm{MF}}, \ell_t) \\ \mathscr{D}^{\mathrm{MF}} \leftarrow (\mathrm{x}_t^{\mathrm{MF}}, f^{(\ell_t)}(\mathrm{x}_t^{\mathrm{MF}})) \\ \mathscr{D}^{\mathrm{pSF}} \leftarrow (\mathrm{x}_t^{\mathrm{pSF}}, \mu_{\mathrm{MF}}^{\mathrm{obj}}(\mathrm{x}_t^{\mathrm{pSF}})) \end{cases}$$

Ensures pseudo-queries added to SFBO are trustworthy: **unreliable case**

because $\mathrm{x}_t^{\mathrm{pSF}}$ only brings little information

Summary

$\sigma_{\mathrm{MF}}(\mathrm{x}_t^{\mathrm{pSF}}, \mathrm{obj}) \leq c_1 \implies \begin{cases} \text{pick } (\mathrm{x}_t^{\mathrm{MF}}, \ell_t) \\ \mathscr{D}^{\mathrm{MF}} \leftarrow (\mathrm{x}_t^{\mathrm{MF}}, f^{(\ell_t)}(\mathrm{x}_t^{\mathrm{MF}})) \\ \mathscr{D}^{\mathrm{pSF}} \leftarrow (\mathrm{x}_t^{\mathrm{pSF}}, \mu_{\mathrm{MF}}^{\mathrm{obj}}(\mathrm{x}_t^{\mathrm{pSF}})) \end{cases}$

Ensures pseudo-queries added to SFBO are trustworthy: **unreliable case**

because $\mathrm{x}_t^{\mathrm{pSF}}$ only brings little information

If not satisfied:

1. Pick $(\mathrm{x}_t^{\mathrm{pSF}}, \mathrm{obj})$
2. $\mathscr{D}^{\mathrm{MF}} \leftarrow \left(\mathrm{x}_t^{\mathrm{pSF}}, f^{\mathrm{obj}}\left(\mathrm{x}_t^{\mathrm{pSF}}\right)\right)$
3. $\mathscr{D}^{\mathrm{pSF}} \leftarrow \left(\mathrm{x}_t^{\mathrm{pSF}}, f^{\mathrm{obj}}\left(\mathrm{x}_t^{\mathrm{pSF}}\right)\right)$

$\mathscr{D}^{\mathrm{pSF}}$ and $\mathscr{D}^{\mathrm{SF}}$ only differ at the points where we inputted $\mu_{\mathrm{MF}}^{\mathrm{obj}}(\mathrm{x}_t^{\mathrm{pSF}})$!

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 twice differentiable
- $\mathbb{P}\left(\sup_{\mathrm{x} \in \mathcal{X}} \left| \frac{\partial f^{\text{obj}}}{\partial x_j} \right| > L \right) \leq a e^{-(L/b_j)^2} \quad \forall\ j \in \{1, \dots, d\}$, for $a, b_j > 0$

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 twice differentiable
- $\mathbb{P}\left(\sup_{\mathrm{x} \in \mathscr{X}} \left|\frac{\partial f^{\text{obj}}}{\partial x_j}\right| > L\right) \leq a e^{-(L/b_j)^2} \quad \forall\ j \in \{1, \ldots, d\}$, for $a, b_j > 0$
- The mapping $(\mathrm{x}, \mathscr{D}) \mapsto \alpha(\mathrm{x}|p(f|\mathscr{D}))$ is twice differentiable

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 twice differentiable
- $\mathbb{P}\left(\sup_{\mathrm{x}\in\mathscr{X}} \left|\frac{\partial f^{\text{obj}}}{\partial x_j}\right| > L\right) \leq a e^{-(L/b_j)^2} \quad \forall\ j \in \{1, \dots, d\}$, for $a, b_j > 0$
- The mapping $(\mathrm{x}, \mathscr{D}) \mapsto \alpha(\mathrm{x}|p(f|\mathscr{D}))$ is twice differentiable

Theorem:

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

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 twice differentiable
- $\mathbb{P}\left(\sup_{\mathrm{x}\in\mathcal{X}} \left|\frac{\partial f^{\text{obj}}}{\partial x_j}\right| > L\right) \leq a e^{-(L/b_j)^2} \quad \forall\, j \in \{1, \ldots, d\}$, for $a, b_j > 0$
- The mapping $(\mathrm{x}, \mathcal{D}) \mapsto \alpha(\mathrm{x}|p(f|\mathcal{D}))$ is twice differentiable

Theorem:

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

No assumption on the amount of information that f^{AIS} can provide about f^{obj}!

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 twice differentiable
- $\mathbb{P}\left(\sup_{\mathrm{x}\in\mathscr{X}} \left|\frac{\partial f^{\text{obj}}}{\partial x_j}\right| > L\right) \leq ae^{-(L/b_j)^2} \quad \forall\ j \in \{1, \dots, d\}$, for $a, b_j > 0$
- The mapping $(\mathrm{x}, \mathscr{D}) \mapsto \alpha(\mathrm{x}|p(f|\mathscr{D}))$ is twice differentiable

Theorem:

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

No assumption on the amount of information that f^{AIS} can provide about f^{obj}!

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

$c_1(\varepsilon, q) = \frac{\varepsilon}{\sqrt{-2\log(1-q)}}.$

Interested? Have a look at the paper ¯_(ツ)_/¯

Petrus Mikkola, **Julien Martinelli**, Louis Filstroff, Samuel Kaski
Multi-Fidelity Bayesian Optimization with Unreliable Sources. AISTATS 2023.

Some stuff (continued) ¯_(ツ)_/¯

$$k_{\mathsf{MISO}}((\mathrm{x},\ell),(\mathrm{x}',\ell')) = \begin{cases} k_{\mathsf{input}}(\mathrm{x},\mathrm{x}') + k_{\ell}(\mathrm{x},\mathrm{x}') & \ell = \ell' \neq 1 \\ k_{\mathsf{input}}(\mathrm{x},\mathrm{x}') & \text{otherwise} \end{cases}$$

$$k_{\mathsf{LT}}((\mathrm{x},\ell),(\mathrm{x}',\ell')) = \begin{cases} k_{\mathsf{input}}(\mathrm{x},\mathrm{x}') + (1-\ell)(1-\ell')k_{\mathsf{IS}}(\mathrm{x},\mathrm{x}') & \ell \neq 1,\ \ell' \neq 1 \\ k_{\mathsf{input}}(\mathrm{x},\mathrm{x}') & \text{otherwise} \end{cases}$$

$$k_{\mathsf{DS}}((\mathrm{x},\ell),(\mathrm{x}',\ell')) = \begin{cases} ck_{\mathsf{input}}(\mathrm{x},\mathrm{x}') + (1-\ell)(1-\ell')k_{\mathsf{input}}(\mathrm{x},\mathrm{x}') & \ell \neq 1,\ \ell' \neq 1 \\ ck_{\mathsf{input}}(\mathrm{x},\mathrm{x}') & \text{otherwise} \end{cases}$$