NEURIPS 2026

# Bayesian Optimization with Fisher Information Geometry

## Gradient Bounds and Trust-Region Methods

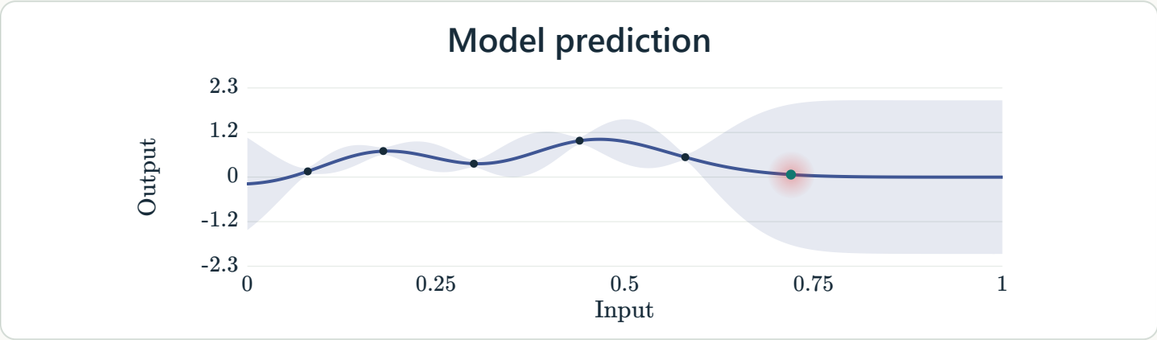
**Saksham Kiroriwal**

Julius Pfrommer · Jürgen Beyerer

Fraunhofer IOSB · Karlsruhe Institute of Technology

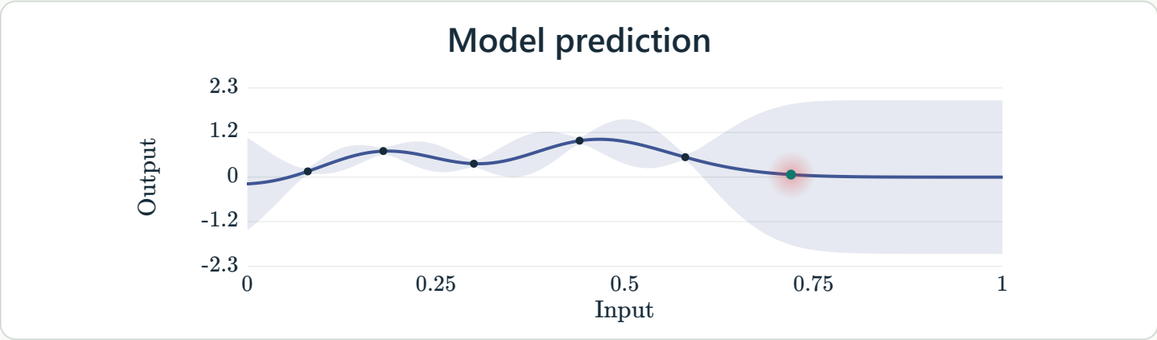
01 · MANIFOLD VIEW

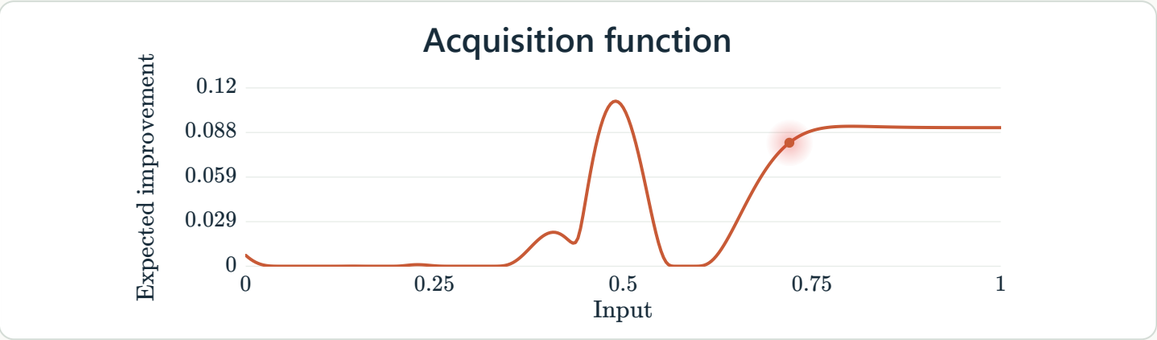
# The surrogate maps inputs to predictive distributions



01 · MANIFOLD VIEW

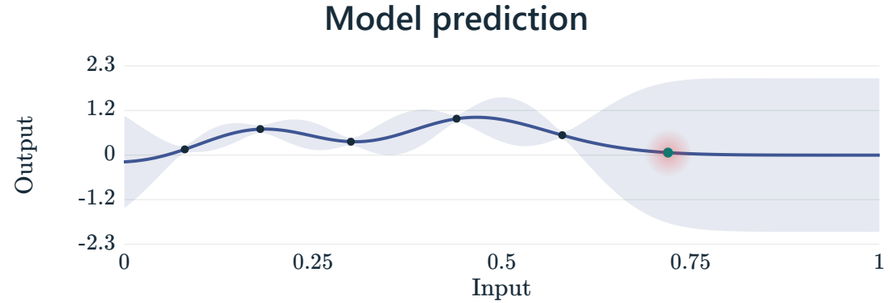
# The surrogate maps inputs to predictive distributions.

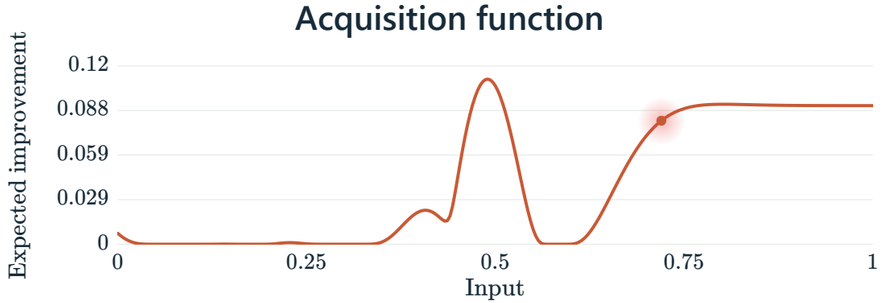


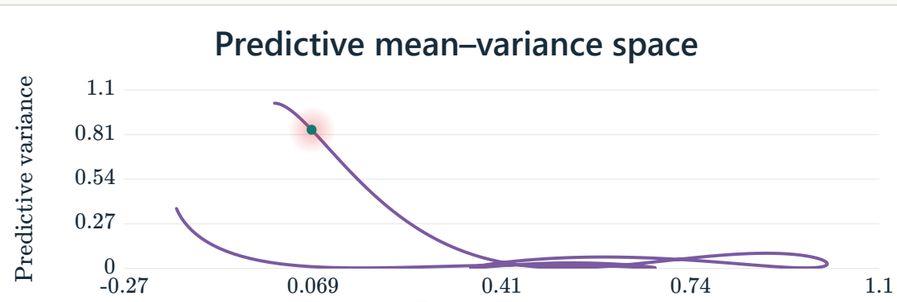



# The surrogate maps inputs to predictive distributions.

$$\boldsymbol{x} \xrightarrow{\varphi_t} \boldsymbol{\theta}_{t,\boldsymbol{x}} = (\mu, \Sigma) \longmapsto \bar{\alpha}(\mu, \Sigma)$$

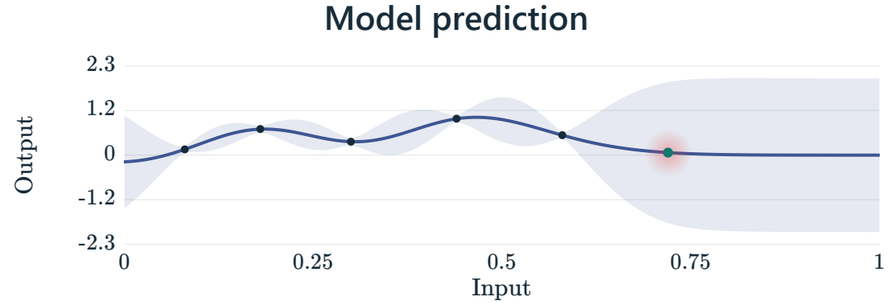


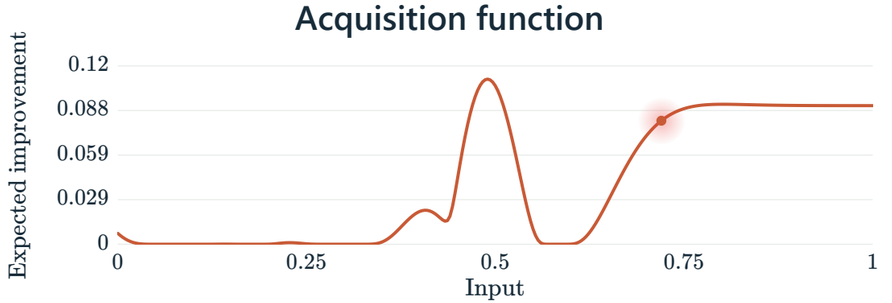



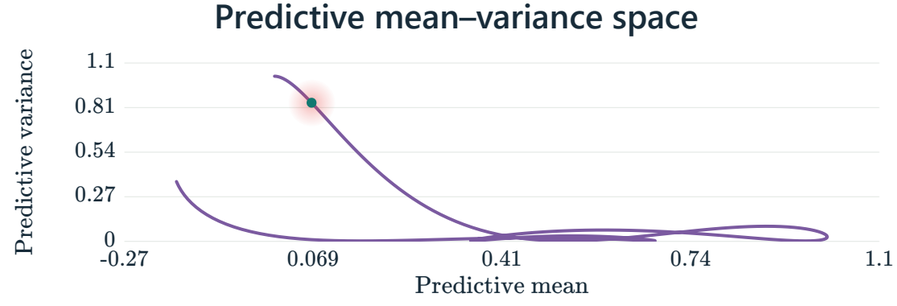
# The surrogate maps inputs to predictive distributions.

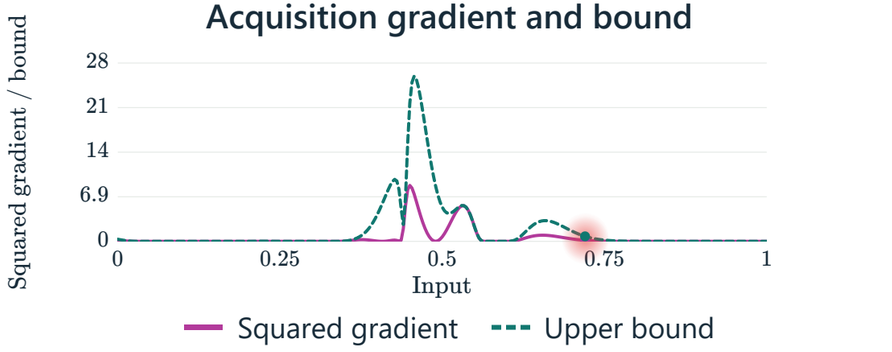
$$\boldsymbol{x} \xrightarrow{\varphi_t} \boldsymbol{\theta}_{t,\boldsymbol{x}} = (\mu, \Sigma) \longmapsto \bar{\alpha}(\mu, \Sigma)$$




$$\|\nabla_{\boldsymbol{x}} \alpha(\boldsymbol{x})\|_2^2 \leq \underbrace{\operatorname{tr}(\mathbf{G}_t(\boldsymbol{x}))}_{\text{predictive sensitivity}} \quad \underbrace{C_\alpha(\boldsymbol{\theta}_{t,\boldsymbol{x}})}_{\text{acquisition sensitivity}}$$



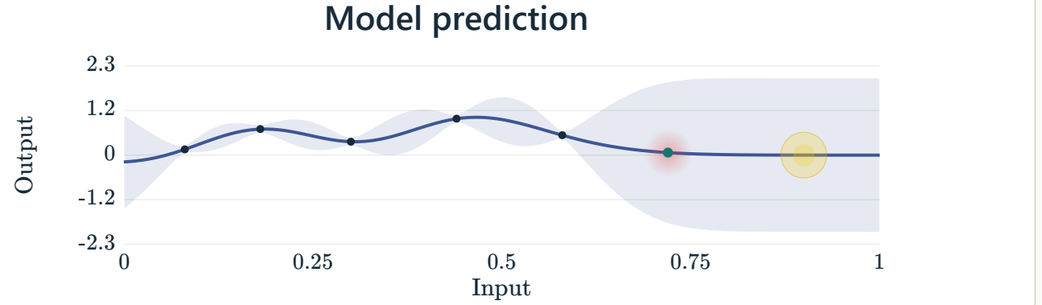



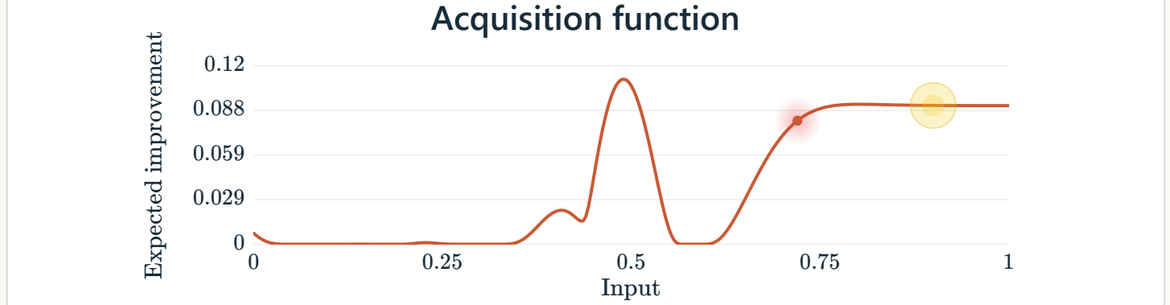


Fraunhofer IOSB

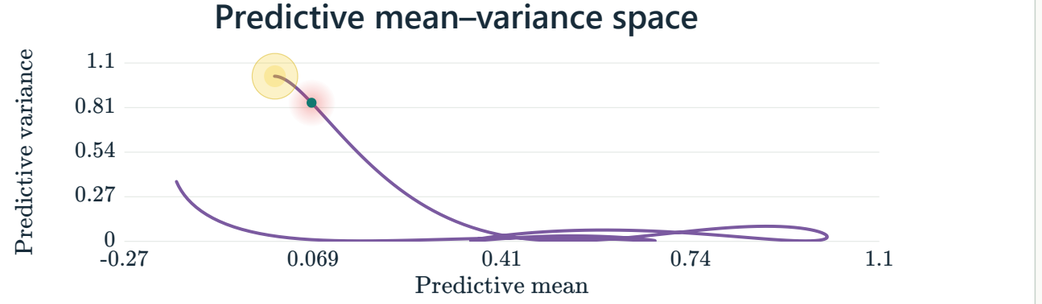
# The surrogate maps inputs to predictive distributions.

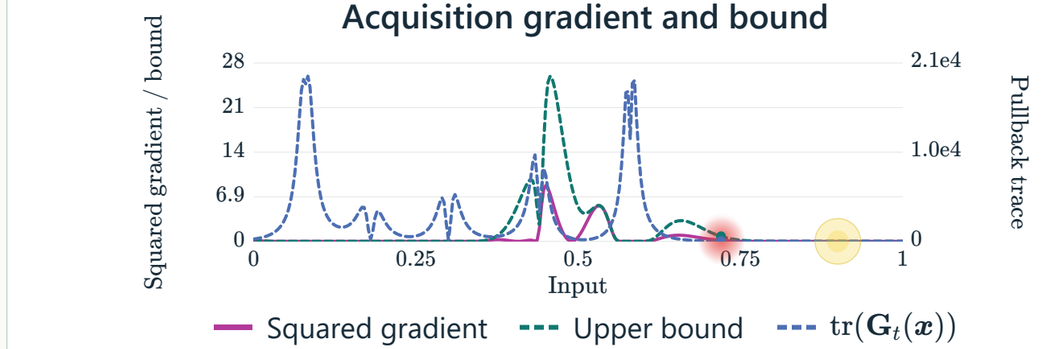
$$\boldsymbol{x} \xrightarrow{\varphi_t} \boldsymbol{\theta}_{t,\boldsymbol{x}} = (\mu, \Sigma) \longmapsto \bar{\alpha}(\mu, \Sigma)$$




$$\|\nabla_{\boldsymbol{x}} \alpha(\boldsymbol{x})\|_2^2 \leq \underbrace{\operatorname{tr}(\mathbf{G}_t(\boldsymbol{x}))}_{\text{predictive sensitivity}} \quad \underbrace{C_\alpha(\boldsymbol{\theta}_{t,\boldsymbol{x}})}_{\text{acquisition sensitivity}}$$




**With bounded acquisition sensitivity, trace collapse forces vanishing gradients.**

# Characteristics of stationary points for acquisition function optimization

Scalar Gaussian predictive with positive variance we can express acquisition function gradients as

$$\nabla_{\boldsymbol{x}}\alpha(\boldsymbol{x}) = \mathbf{A}^\top \boldsymbol{u}$$

$$\mathbf{A} = \begin{bmatrix} (\nabla_{\boldsymbol{x}}\mu(\boldsymbol{x}))^\top / \sqrt{\Sigma} \\ (\nabla_{\boldsymbol{x}}\Sigma(\boldsymbol{x}))^\top / (\sqrt{2}\Sigma) \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} \sqrt{\Sigma}\,\partial_\mu \bar{\alpha}(\mu, \Sigma) \\ \sqrt{2}\Sigma\,\partial_\Sigma \bar{\alpha}(\mu, \Sigma) \end{bmatrix}, \quad \mathbf{M} = \mathbf{A}\mathbf{A}^\top$$

$\nu_{\min}, \nu_{\max}$ are the eigenvalues of $\mathbf{M}$.

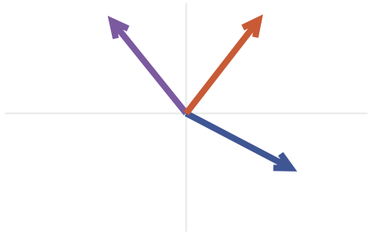
$$\nu_{\min} \leq \frac{\boldsymbol{u}^\top \mathbf{M} \boldsymbol{u}}{\|\boldsymbol{u}\|_2^2} = \frac{\|\nabla_{\boldsymbol{x}}\alpha(\boldsymbol{x})\|_2^2}{\|\boldsymbol{u}\|_2^2} \leq \nu_{\max}$$

$$\nabla_{\boldsymbol{x}}\alpha(\boldsymbol{x}) = \mathbf{0} \;\Rightarrow\; \nu_{\min} = 0 \;\Rightarrow\; \operatorname{rank}(\mathbf{A}) = 1$$

$$\Rightarrow\; \nabla_{\boldsymbol{x}}\mu(\boldsymbol{x}) \parallel \nabla_{\boldsymbol{x}}\Sigma(\boldsymbol{x})$$

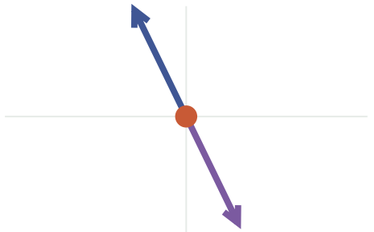
A necessary condition for a location $\boldsymbol{x}$ to be stationary.

Non-stationary point



$\nabla_{\boldsymbol{x}}\alpha(\boldsymbol{x}) \neq \mathbf{0}$

Interior stationary point

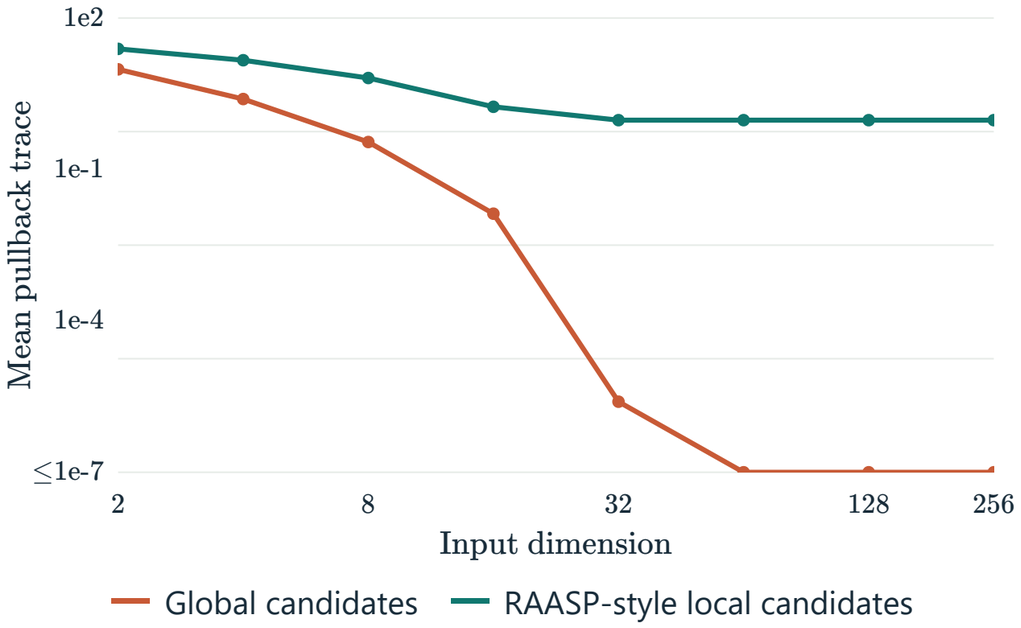


$\nabla_{\boldsymbol{x}}\alpha(\boldsymbol{x}) \approx \mathbf{0}$

— $\nabla_{\boldsymbol{x}}\mu(\boldsymbol{x})$ — $\nabla_{\boldsymbol{x}}\Sigma(\boldsymbol{x})$ — $\nabla_{\boldsymbol{x}}\alpha(\boldsymbol{x})$

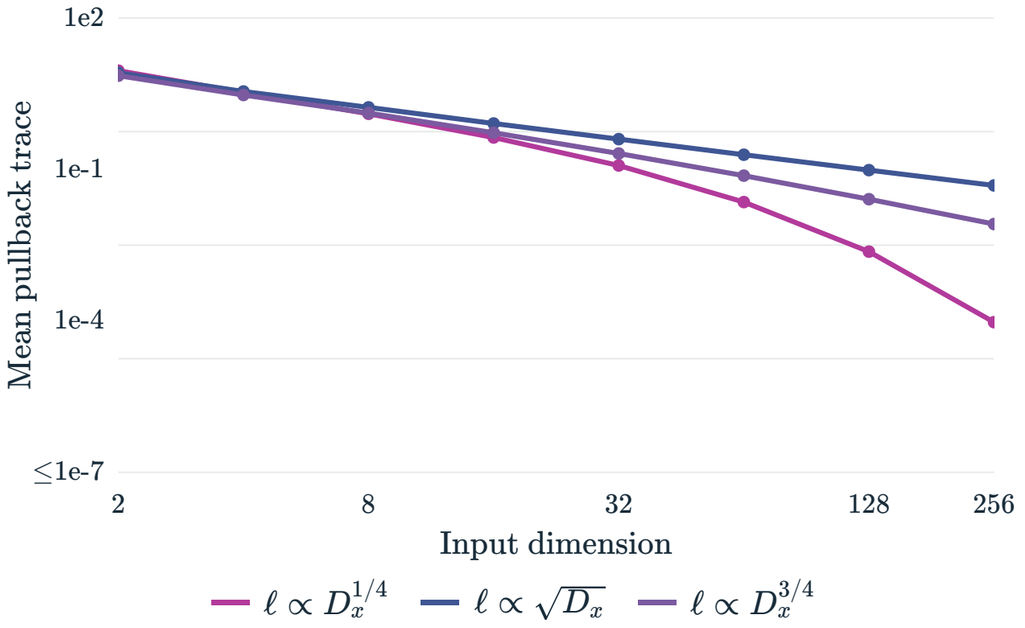
# Gradient bound as unifying lens for RAASP and dimension-scaled priors

## RAASP: stay near observations




Local perturbations retain predictive sensitivity.

## Scale lengthscales with dimension

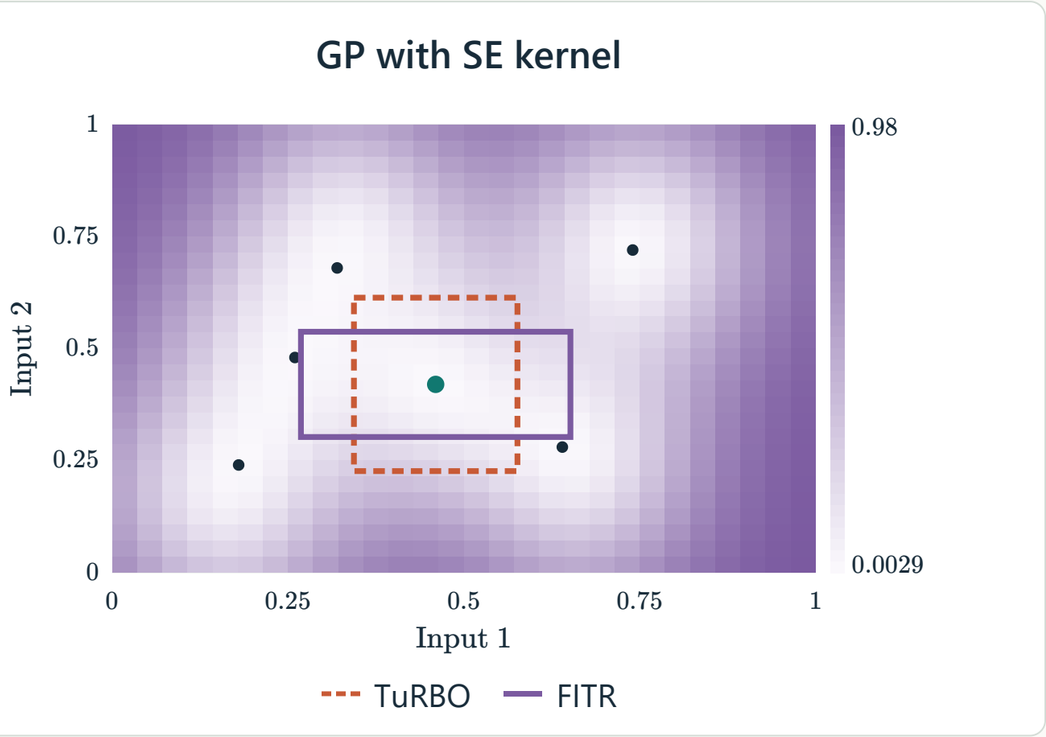


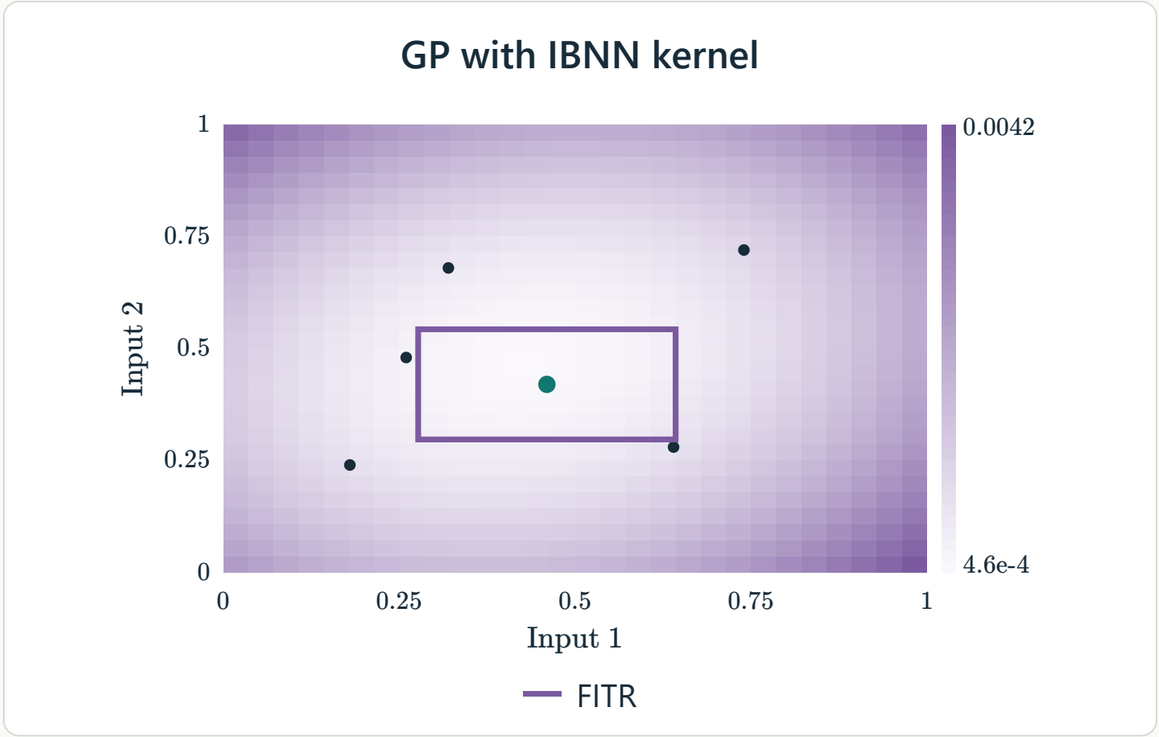

$\ell \propto \sqrt{D_x}$ gives the slowest asymptotic trace decay.

Gaussian process with squared exponential kernel

RAASP: Papenmeier et al., ICML 2025 · Dimension-scaled priors: Hvarfner et al., ICML 2024

# Trust region Bayesian Optimization using Fisher Information



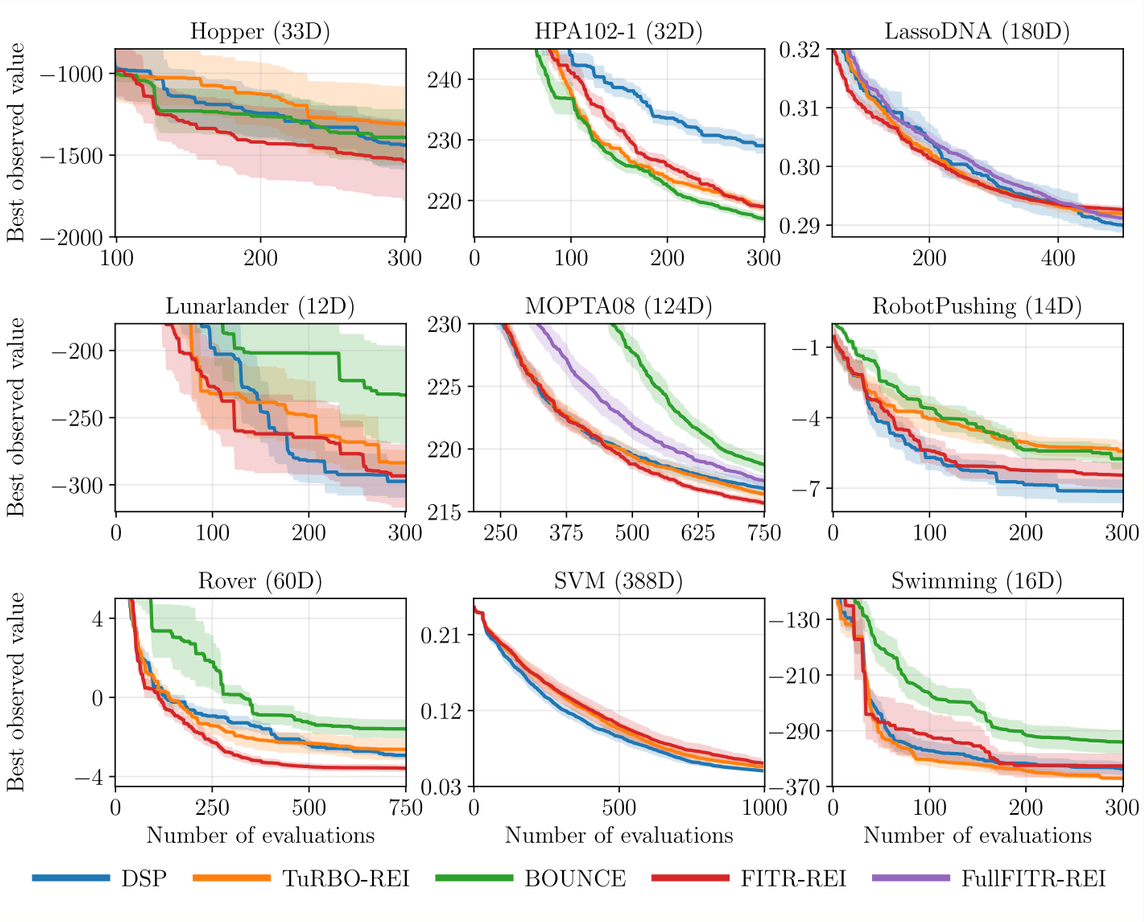

**More sensitive directions → narrower trust region**

Builds on TuRBO: Eriksson et al., NeurIPS 2019

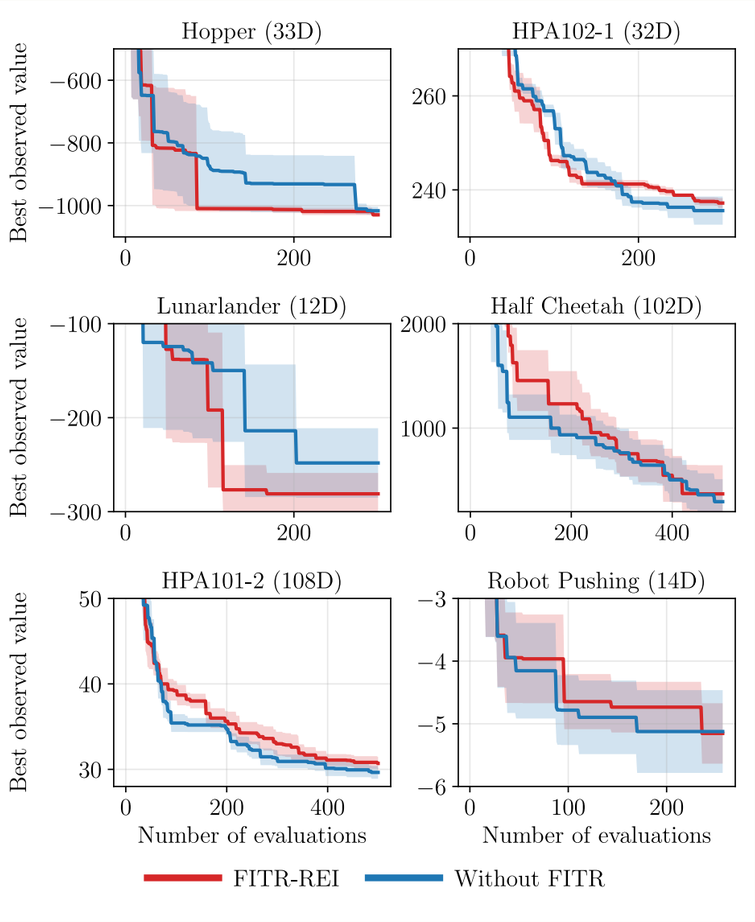
**Extendable to non-stationary kernels · Location dependent trust regions**

# BO performance for GP with SE and IBNN kernels

## GP with SE kernel




## GP with IBNN kernel




Mean BO performance with standard error over eleven trials. Lower is better.