# Relaxed Gaussian process interpolation: a goal-oriented approach to Bayesian optimization


Sébastien Petit[1,2]

Joint work with Julien Bect[1] and Emmanuel Vazquez[1]

[1]Université Paris-Saclay, CentraleSupélec, Laboratoire des Signaux et Systèmes

[2] Safran Aircraft Engines


GDR Mascot-Num, juin 2022

# Outline of the presentation

# 1 Introduction

# Goal-oriented modeling

- Consider the task of building a prediction of a function

$$f : \mathbb{X} \subset \mathbb{R}^d \to \mathbb{R}$$

from evaluations at $x_1, x_2, \ldots$ using a Gaussian process model.

# Goal-oriented modeling

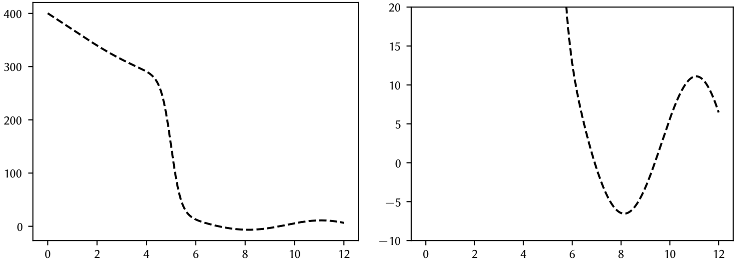
- Consider the task of building a prediction of a function

$$f : \mathbb{X} \subset \mathbb{R}^d \to \mathbb{R}$$

  from evaluations at $x_1, x_2, \ldots$ using a Gaussian process model.

- When such a prediction is used inside a Bayesian optimization algorithm, for example in a minimization problem, it is particularly important to get good predictive distributions on a range of function values corresponding to low values of the function.

# The Steep function



# The Steep function

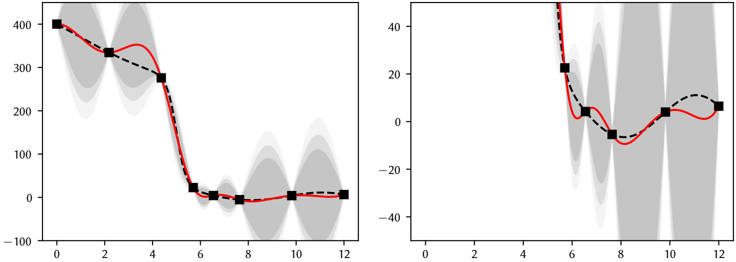



Figure: A stationary GP for building predictive distributions

# Our proposal: relaxed Gaussian process

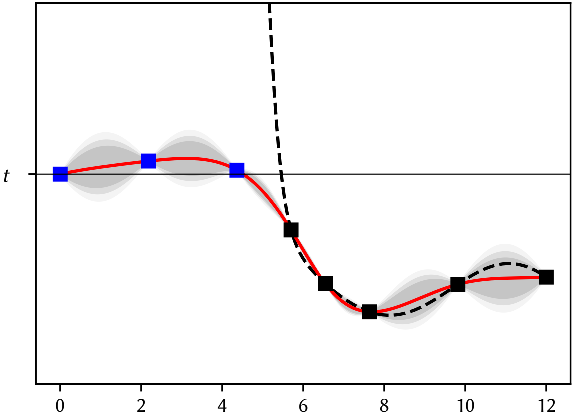



Figure: Relax interpolation constraints above *t*!

## 2 Building predictive distributions with GPs

# Building predictive distributions with GPs

Given data points $(x_i, f(x_i))$, $i = 1, \ldots, n$, a standard practice to get predictive distributions for $f$ is the following procedure:

# Building predictive distributions with GPs

Given data points $(x_i, f(x_i))$, $i = 1, \ldots, n$, a standard practice to get predictive distributions for $f$ is the following procedure:

1. choose a model $\xi \sim \mathrm{GP}(\mu_\theta, k_\theta)$, with $\theta \in \Theta$

# Building predictive distributions with GPs

Given data points $(x_i, f(x_i))$, $i = 1, \dots, n$, a standard practice to get predictive distributions for $f$ is the following procedure:

1. choose a model $\xi \sim \mathrm{GP}(\mu_\theta,\ k_\theta)$, with $\theta \in \Theta$
2. select $\theta$ by maximum likelihood

$$\mathcal{L}(\theta;\ \underline{Z}_n) = -\ln\left(p\left(\underline{Z}_n \,\middle|\, \theta\right)\right)$$

with $\underline{Z}_n = (\xi(x_1),\ \dots,\ \xi(x_n))^\mathsf{T}$

# Building predictive distributions with GPs

Given data points $(x_i, f(x_i))$, $i = 1, \ldots, n$, a standard practice to get predictive distributions for $f$ is the following procedure:

1. choose a model $\xi \sim \mathrm{GP}(\mu_\theta, k_\theta)$, with $\theta \in \Theta$
2. select $\theta$ by maximum likelihood

$$\mathcal{L}(\theta; \underline{Z}_n) = -\ln(p(\underline{Z}_n \mid \theta))$$

   with $\underline{Z}_n = (\xi(x_1), \ldots, \xi(x_n))^\mathsf{T}$

3. compute the posterior distribution $\xi \mid \underline{Z}_n$

## 3 Relaxed Gaussian processes (reGP)

Predictive distributions with interpolation relaxation

Application to Bayesian optimization

Convergence analysis of reGP

# Objective

- Consider $n$ points $\underline{x}_n = (x_1, \ldots, x_n)$ in $\mathbb{X}$ and let $\underline{z}_n = (f(x_1), \ldots, f(x_n))$ be the vector of the corresponding values of $f$.

# Objective

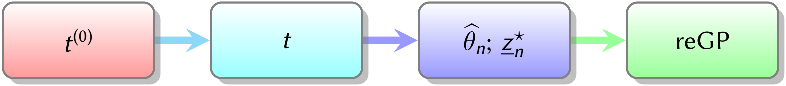
- Consider $n$ points $\underline{x}_n = (x_1, \ldots, x_n)$ in $\mathbb{X}$ and let $\underline{z}_n = (f(x_1), \ldots, f(x_n))$ be the vector of the corresponding values of $f$.
- Consider a family $\xi \sim \mathrm{GP}(\mu_\theta, k_\theta)$, with $\theta \in \Theta$.

# Objective

- Consider $n$ points $\underline{x}_n = (x_1, \ldots, x_n)$ in $\mathbb{X}$ and let $\underline{z}_n = (f(x_1), \ldots, f(x_n))$ be the vector of the corresponding values of $f$.
- Consider a family $\xi \sim \mathrm{GP}(\mu_\theta, k_\theta)$, with $\theta \in \Theta$.
- Our objective is to obtain (good) predictive distributions of $f$ below a threshold $t^{(0)}$, on the range $Q = (-\infty, t^{(0)})$. We accept degraded predictions above $t^{(0)}$.

$\hookrightarrow$ Goal-oriented modeling

# Overview




Given $\underline{x}_n$, $\underline{z}_n$, $t^{(0)}$, and a parametrized GP model $\xi \sim \mathrm{GP}\,(\mu_\theta,\ k_\theta)$:

- Select $t$ automatically above $t^{(0)}$
- Choose $\widehat{\theta}_n \in \Theta$ and modified observations $\underline{z}_n^\star \in \mathbb{R}^n$
- reGP: $\xi \mid \underline{Z}_n = \underline{z}_n^\star$

# reGP predictive distribution given $t$

- Suppose that $\xi \sim \mathrm{GP}(0,\ k)$ with a fixed $k$ for simplicity. Let $\mu_n$ be the posterior mean of $\xi$. We have (Kimeldorf & Wahba 1970)

$$\mu_n = \operatorname{argmin}_{\begin{cases} h \in \mathcal{H}(\mathbb{X}) \\ h(\underline{x}_n) = \underline{z}_n \end{cases}} \|h\|_{\mathcal{H}(\mathbb{X})}.$$

with $\mathcal{H}(\mathbb{X})$ the RKHS attached to $k$.

# reGP predictive distribution given $t$

- Suppose that $\xi \sim \mathrm{GP}(0,\ k)$ with a fixed $k$ for simplicity. Let $\mu_n$ be the posterior mean of $\xi$. We have (Kimeldorf & Wahba 1970)

$$\mu_n = \operatorname{argmin}_{\begin{cases} h \in \mathcal{H}(\mathbb{X}) \\ h(\underline{x}_n) = \underline{z}_n \end{cases}} \|h\|_{\mathcal{H}(\mathbb{X})}.$$

  with $\mathcal{H}(\mathbb{X})$ the RKHS attached to $k$.

- The core idea is to build a predictive distribution with a mean given by the relaxed interpolator:

$$\tilde{\mu}_n = \operatorname{argmin}_{\begin{cases} h \in \mathcal{H}(\mathbb{X}) \\ h(\underline{x}_{n,0}) = \underline{z}_{n,0} \\ h(\underline{x}_{n,1}) \geq t \end{cases}} \|h\|_{\mathcal{H}(\mathbb{X})}.$$

  with $\underline{x}_n = \left(\underline{x}_{n,\,0},\ \underline{x}_{n,\,1}\right)$ and $\underline{z}_n = \left(\underline{z}_{n,\,0},\ \underline{z}_{n,\,1}\right)$, such that $\underline{z}_{n,\,0} < t$ and $\underline{z}_{n,\,1} \geq t$ wlog.

# reGP predictive distribution given $t$

Recall that $\underline{z}_n = \left(\underline{z}_{n,0},\, \underline{z}_{n,1}\right)$, where $\underline{z}_{n,0} < t$ and $\underline{z}_{n,1} \geq t$.

## Definition

The predictive distribution reGP is defined as the conditional distribution $\mathrm{P}_n^t$ of $\xi$ given

$$\begin{cases} \xi(\underline{x}_{n,0}) & = & \underline{z}_{n,0} \\ \xi(\underline{x}_{n,1}) & = & \underline{z}^{\star}_{n,1} \end{cases} \qquad (1)$$

where $\underline{z}^{\star}_{n,1}$ is the solution of the extended negative log likelihood

$$\left(\widehat{\theta}_n,\, \underline{z}^{\star}_{n,1}\right) = \mathrm{argmin}_{\theta\in\Theta,\ z_{n,1}\geq t}\, \mathcal{L}\left(\theta;\, \underline{z}_{n,0},\, z_{n,1}\right). \qquad (2)$$

# reGP predictive distribution given $t$

Recall that $\underline{z}_n = \left(\underline{z}_{n,0},\ \underline{z}_{n,1}\right)$, where $\underline{z}_{n,0} < t$ and $\underline{z}_{n,1} \geq t$.

## Definition

The predictive distribution reGP is defined as the conditional distribution $\mathrm{P}_n^t$ of $\xi$ given

$$\begin{cases} \xi(\underline{x}_{n,0}) & = & \underline{z}_{n,0} \\ \xi(\underline{x}_{n,1}) & = & \underline{z}^{\star}_{n,1} \end{cases} \tag{1}$$

where $\underline{z}^{\star}_{n,1}$ is the solution of the extended negative log likelihood

$$\left(\widehat{\theta}_n,\ \underline{z}^{\star}_{n,1}\right) = \operatorname{argmin}_{\theta\in\Theta,\ z_{n,1}\geq t} \mathcal{L}\left(\theta;\ \underline{z}_{n,0},\ z_{n,1}\right). \tag{2}$$

For a given $\theta$, the mean of the reGP predictive distribution is given by the relaxed interpolator $\tilde{\mu}_n$.

- In more details, let $\underline{\mu}_\theta = (\mu_\theta(x_1), \ldots, \mu_\theta(x_n))^\mathsf{T}$, $K_\theta = \left(k_\theta\left(x_i, x_j\right)\right)_{i,j}$, and write $z_n = \left(\underline{z}_{n,0}^\mathsf{T}, z_{n,1}^\mathsf{T}\right)^\mathsf{T}$, then

$$\mathcal{L}\left(\theta; \underline{z}_{n,0}, z_{n,1}\right) \propto \log\left(\det\left(K_\theta\right)\right) + \underbrace{(z_n - \underline{\mu}_\theta)^\mathsf{T} K_\theta^{-1}(z_n - \underline{\mu}_\theta)}_{\text{quadratic form w.r.t. } (z_{n,1})} + \text{constant},$$

- The likelihood can be optimized jointly w.r.t. $\theta$ and $z_{n,1}$ with an L-BFGS-B algorithm (Byrd et al., 1995), for instance.

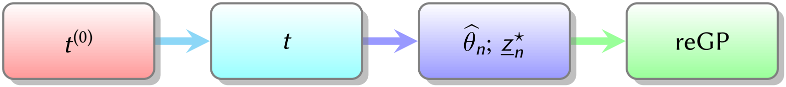
- In more details, let $\underline{\mu}_\theta = (\mu_\theta(x_1), \ldots, \mu_\theta(x_n))^\mathsf{T}$, $K_\theta = \left(k_\theta\left(x_i, x_j\right)\right)_{i,j}$, and write $z_n = \left(\underline{z}_{n,0}^\mathsf{T}, z_{n,1}^\mathsf{T}\right)^\mathsf{T}$, then

$$\mathcal{L}\left(\theta; \underline{z}_{n,0}, z_{n,1}\right) \propto \log\left(\det\left(K_\theta\right)\right) + \underbrace{(z_n - \underline{\mu}_\theta)^\mathsf{T} K_\theta^{-1} (z_n - \underline{\mu}_\theta)}_{\text{quadratic form w.r.t. } (z_{n,1})} + \text{constant},$$

- The likelihood can be optimized jointly w.r.t. $\theta$ and $z_{n,1}$ with an L-BFGS-B algorithm (Byrd et al., 1995), for instance.
- Moreover, note that,

$$\mathcal{L}\left(\theta; \underline{z}_{n,0}, z_{n,1}\right) = -\underbrace{\ln\left(p\left(\underline{z}_{n,0} \mid \theta\right)\right)}_{\text{likelihood under } t} - \underbrace{\ln\left(p\left(z_{n,1} \mid \theta, \underline{z}_{n,0}\right)\right)}_{\text{imputation term}},$$

# Overview




Given $\underline{x}_n$, $\underline{z}_n$, $t^{(0)}$, and a parametrized GP model $\xi \sim \mathrm{GP}\,(\mu_\theta,\ k_\theta)$:

- Select $t$ automatically above $t^{(0)}$
- Choose $\widehat{\theta}_n \in \Theta$ and modified observations $\underline{z}_n^\star \in \mathbb{R}^n$
- reGP: $\xi \mid \underline{Z}_n = \underline{z}_n^\star$

# Choosing $t \geq t^{(0)}$

Which data are useful for prediction below $t^{(0)}$?

# Choosing $t \geq t^{(0)}$

Which data are useful for prediction below $t^{(0)}$?

$\longrightarrow$ select a good value of $t$ between:

- $t = +\infty$ yielding a standard GP,
- $t = t^{(0)}$, when only the data within $Q = (-\infty, t^{(0)})$ help for prediction

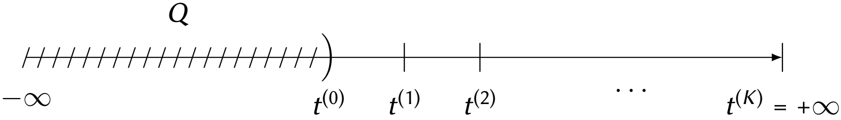
# Choosing $t \geq t^{(0)}$

Which data are useful for prediction below $t^{(0)}$?

$\longrightarrow$ select a good value of $t$ between:

- $t = +\infty$ yielding a standard GP,
- $t = t^{(0)}$, when only the data within $Q = (-\infty, t^{(0)})$ help for prediction

We consider a finite set of candidates $t^{(0)} < t^{(1)} < \cdots < t^{(K)} = +\infty$



# A goal-oriented goodness-of-fit criterion?

- How to select $t$ among $t^{(0)} < t^{(1)} < \cdots < t^{(K)} = +\infty$?

# A goal-oriented goodness-of-fit criterion?

- How to select $t$ among $t^{(0)} < t^{(1)} < \cdots < t^{(K)} = +\infty$?
- Recall the standard leave-one-out cross-validation (Currin et al. 1988)

$$\frac{1}{n}\sum_{i=1}^{n} S\left(\mathrm{P}_{n,\,-i};\ z_i\right)$$

where

→ $\mathrm{P}_{n,\,-i}$ is the loo predictive distribution at $x_i$

→ and $S$ is a scoring rule (see, e.g., Gneiting & Raftery 2007): $S(P; z)$ represents a loss when using the distribution $P$ to predict $z$

# A goal-oriented goodness-of-fit criterion?

- How to select $t$ among $t^{(0)} < t^{(1)} < \cdots < t^{(K)} = +\infty$?
- Recall the standard leave-one-out cross-validation (Currin et al. 1988)

$$\frac{1}{n}\sum_{i=1}^{n} S\left(\mathrm{P}_{n,\,-i};\ z_i\right)$$

where

→ $\mathrm{P}_{n,\,-i}$ is the loo predictive distribution at $x_i$

→ and $S$ is a scoring rule (see, e.g., Gneiting & Raftery 2007): $S(P; z)$ represents a loss when using the distribution $P$ to predict $z$

- A scoring rule focusing on $Q = \left(-\infty,\ t^{(0)}\right)$?

# Truncated continuous ranked probability score

Recall the continuous ranked probability score (Matheson and Winkler 1976; Hersbach 2000; Gneitting 2004):

$$S^{\mathrm{CRPS}}(P,\ z) = \int_{-\infty}^{+\infty} \left(F_P(u) - \mathbb{1}_{z \le u}\right)^2 \mathrm{d}u\,,$$

# Truncated continuous ranked probability score

Recall the continuous ranked probability score (Matheson and Winkler 1976; Hersbach 2000; Gneitting 2004):

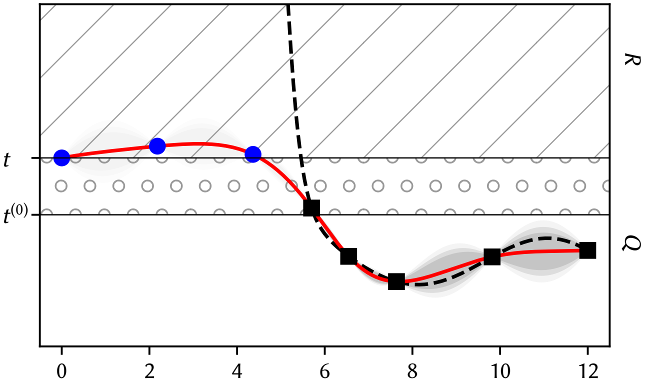
$$S^{\mathrm{CRPS}}(P,\ z) = \int_{-\infty}^{+\infty} \left(F_P(u) - \mathbb{1}_{z \leq u}\right)^2 \mathrm{d}u\,,$$

We propose the truncated continuous ranked probability score:

$$S^{\mathrm{tCRPS}}_{t^{(0)}}(P;\ z) = \int_{-\infty}^{t^{(0)}} \left(F_P(u) - \mathbb{1}_{z \leq u}\right)^2 \mathrm{d}u\,.$$

- If $z < t^{(0)} \rightarrow$, the tCRPS asks that $P \simeq \delta_z$
- If $z \geq t^{(0)} \rightarrow$ the value of $S^{\mathrm{tCRPS}}_{t^{(0)}}(P;\ z)$ does not depend on the specific value $z$, and decreases when $P$ is concentrated above $t^{(0)}$
- We give closed-form expressions for $S^{\mathrm{tCRPS}}_{t^{(0)}}$ when $P$ is Gaussian

The cross-validation criterion to select $t$ using the tCRPS scoring rule is called the LOO-tCRPS



# Application to Bayesian optimization

# Efficient Global Optimization (EGO, Jones et al. 1998)

- Given $f : \mathbb{X} \to \mathbb{R}$, consider the minimization problem

$$m = \min_{x \in \mathbb{X}} f$$

$\hookrightarrow$ construct a sequence of evaluations points $x_1,\ x_2,\ \ldots$, such that for $n > 0$, $m_n = \min f(x_1), \ldots, f(x_n)$ is close to $m$

# Efficient Global Optimization (EGO, Jones et al. 1998)

- Given $f : \mathbb{X} \to \mathbb{R}$, consider the minimization problem

$$m = \min_{x \in \mathbb{X}} f$$

$\hookrightarrow$ construct a sequence of evaluations points $x_1, x_2, \ldots$, such that for $n > 0$, $m_n = \min f(x_1), \ldots, f(x_n)$ is close to $m$

- Standard Bayesian approach $\to$ build predictive distributions for $f$ using a GP $\xi$, and use the strategy

$$x_{n+1} = \operatorname{argmax}_{x \in \mathbb{X}} \rho_n(x)$$

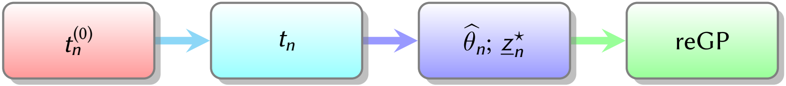
where $\rho_n$ is the expected improvement (Mockus $\sim$ 1970$s$)

$$\rho_n(x) = \mathbb{E}\left((m_n - m_{n+1})_+ \mid \underline{Z}_n = \underline{z}_n\right)$$

with $\underline{Z}_n = (\xi(x_1), \ldots, \xi(x_n))$

# Efficient global optimization with relaxation (EGO-R)

At each step $n$, construct a reGP model



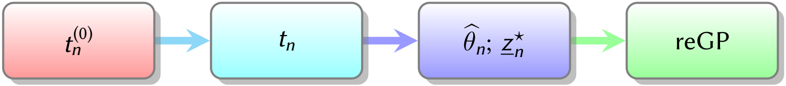

and choose $x_{n+1}$ using the expected improvement

$$\rho_n(x) = \mathbb{E}\left((m_n - m_{n+1})_+ \;\middle|\; \underline{Z}_n = \underline{z}_n^\star\right),$$

sequentially under the reGP distribution.

# Efficient global optimization with relaxation (EGO-R)

At each step $n$, construct a reGP model




and choose $x_{n+1}$ using the expected improvement

$$\rho_n(x) = \mathbb{E}\left((m_n - m_{n+1})_+ \;\middle|\; \underline{Z}_n = \underline{z}_n^\star\right),$$

sequentially under the reGP distribution.

Several strategies for determining a validation threshold $t_n^{(0)}$ may be considered (e.g., set a fixed threshold $t_n^{(0)} = t^{(0)}$ from an initial DoE)

# EGO-R on a “difficult” function: Goldstein-Price

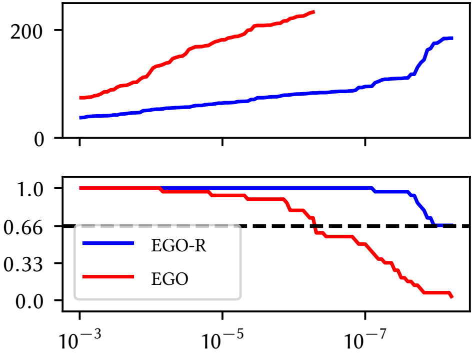



Figure: average number of iterations to reach a target on the horizontal axis (top) and fraction of repetitions reaching the target

# Relaxation on Goldstein-Price

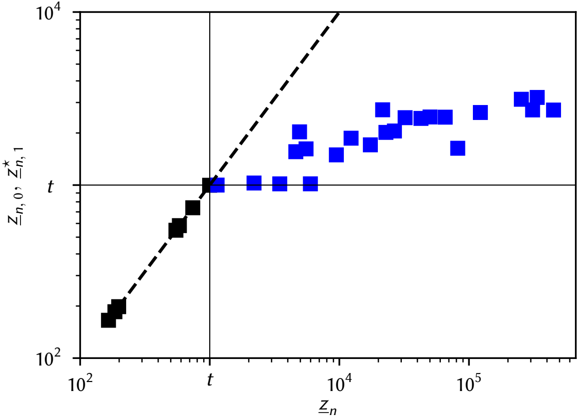



Figure: relaxed observations $\left(\underline{z}_{n,0}, \underline{z}^\star_{n,1}\right)$ versus $\underline{z}_n$.

# EGO-R on an “easy” function: log of Golstein-Price

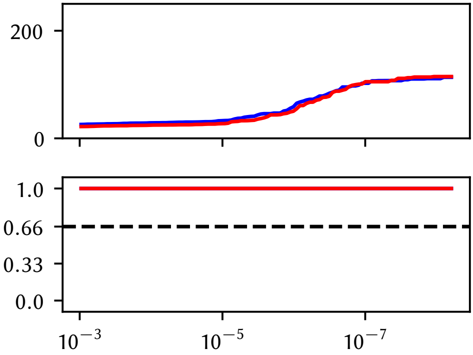



Figure: average number of iterations to reach a target (top) and fraction of repetitions reaching the target (bottom)

# Convergence analysis of reGP

# Setup

- Consider $\xi \sim \mathrm{GP}(0, k)$, where $k$ is a fixed covariance with finite smoothness (e.g, a Matérn covariance function with regularity $0 < \nu < \infty$)
- Denote by $\mathcal{H}(\mathbb{X})$ the reproducing kernel Hilbert space (RKHS) attached to $k$.

# Convergence of EGO-R

Prop.

*Let $f \in \mathcal{H}(\mathbb{X})$ and consider the EGO-R algorithm for building $(x_n)_{n\geq 1} \in \mathbb{X}^{\mathbb{N}}$, with*

$$t_n^{(0)} > \min(f(x_1), \dots, f(x_n)), \quad \forall n \geq 1.$$

*Then, the sequence $(x_n)_{n\geq 1}$ is dense in $\mathbb{X}$.*

# Known results about GPs

- Denote by $\mu_{n,f}$ the posterior distribution of $\xi$ given the data
- If $(x_n)_{n\geq 1}$ is dense, then $\mu_{n,f} \to f$ in $\mathcal{H}(\mathbb{X})$

# Known results about GPs

- Denote by $\mu_{n,f}$ the posterior distribution of $\xi$ given the data
- If $(x_n)_{n\geq 1}$ is dense, then $\mu_{n,f} \to f$ in $\mathcal{H}(\mathbb{X})$
- If $k$ has smoothness $\nu > 0$ and $\mathbb{X}$ is "nice", then

$$\|f - \mu_{n,f}\|_{L^\infty(\mathbb{X})} \lesssim h_n^\nu \|f\|_{\mathcal{H}(\mathbb{X})}, \quad h_n = \sup_{x\in\mathbb{X}} \min_{1\leq i\leq n} \|x - x_i\|$$

(Arcangeli et al. 2007)

# Convergence of reGP

Let $t$ be a fixed relaxation threshold and $\mathcal{H}(t, f)$ be the space of functions $g \in \mathcal{H}(\mathbb{X})$ such that, for all $x \in \mathbb{X}$

$$\begin{cases} g(x) \geq t & \text{if } f(x) \geq t, \\ g(x) = f(x) & \text{otherwise.} \end{cases} \tag{3}$$

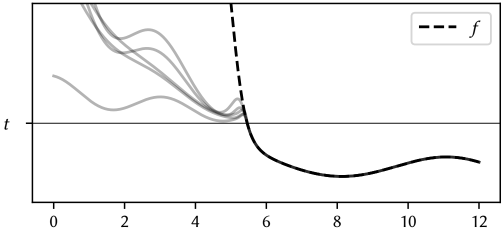



Figure: $\mathcal{H}(t, f)$

# Convergence of reGP

Let $\tilde{\mu}_{n,f} : \mathbb{X} \to \mathbb{R}$ be the mean of the reGP predictive distribution

## Prop.

*Assume $(x_n)_{n\geq 1}$ is dense. Then, the sequence $\left(\tilde{\mu}_{n,f}\right)_{n\geq 1}$ converges to the unique minimum norm element $s_{t,\mathbb{X}}$ of $\mathcal{H}(t, f)$.*

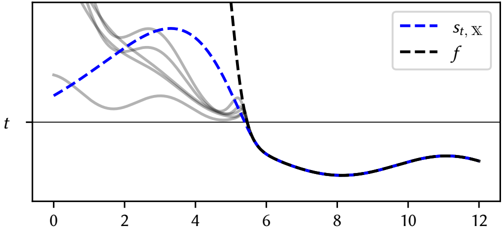



Figure: $\mathcal{H}(t, f)$

# Convergence analysis of reGP

Let $\mathbb{X}_0 = \{x \in \mathbb{X},\, f(x) < t\}$. For every $g \in \mathcal{H}(t, f)$, standard GP interpolation on $g$ yields

$$\|f - \mu_{n,g}\|_{L^\infty(\mathbb{X}_0)} \lesssim h_n^\nu \|g\|_{\mathcal{H}(\mathbb{X})}.$$

# Convergence analysis of reGP

Let $\mathbb{X}_0 = \{x \in \mathbb{X},\, f(x) < t\}$. For every $g \in \mathcal{H}(t, f)$, standard GP interpolation on $g$ yields

$$\|f - \mu_{n,g}\|_{L^\infty(\mathbb{X}_0)} \lesssim h_n^\nu \|g\|_{\mathcal{H}(\mathbb{X})}.$$

Observe that $s_{t,\mathbb{X}}$ optimizes the bound.

# Convergence analysis of reGP

Let $\mathbb{X}_0 = \{x \in \mathbb{X},\ f(x) < t\}$. For every $g \in \mathcal{H}(t, f)$, standard GP interpolation on $g$ yields

$$\|f - \mu_{n,g}\|_{L^\infty(\mathbb{X}_0)} \lesssim h_n^\nu \|g\|_{\mathcal{H}(\mathbb{X})}.$$

Observe that $s_{t,\mathbb{X}}$ optimizes the bound.

**Prop.**

*For $n \geq 1$ and a "nice" $B \subset \mathbb{X}_0$:*

$$\|f - \tilde{\mu}_{n,f}\|_{L^\infty(B)} \lesssim h_n^\nu \|s_{t,\mathbb{X}}\|_{\mathcal{H}(\mathbb{X})}.$$

# Convergence analysis of reGP

Let $\mathbb{X}_0 = \{x \in \mathbb{X},\, f(x) < t\}$. For every $g \in \mathcal{H}(t, f)$, standard GP interpolation on $g$ yields

$$\|f - \mu_{n,g}\|_{L^\infty(\mathbb{X}_0)} \lesssim h_n^\nu \|g\|_{\mathcal{H}(\mathbb{X})}.$$

Observe that $s_{t,\mathbb{X}}$ optimizes the bound.

**Prop.**

*For $n \geq 1$ and a "nice" $B \subset \mathbb{X}_0$:*

$$\|f - \tilde{\mu}_{n,f}\|_{L^\infty(B)} \lesssim h_n^\nu \|s_{t,\mathbb{X}}\|_{\mathcal{H}(\mathbb{X})}.$$

We also have a weaker kind of guarantee outside $\mathbb{X}_0$

# Convergence analysis of reGP

What about $\|f\|_{\mathcal{H}(\mathbb{X})}/\|s_{t,\mathbb{X}}\|_{\mathcal{H}(\mathbb{X})}$ ?

# Convergence analysis of reGP

What about $\|f\|_{\mathcal{H}(\mathbb{X})} / \|s_{t,\mathbb{X}}\|_{\mathcal{H}(\mathbb{X})}$ ?

**Prop.**

*If* $\max f > t$ *and* $k$ *has finite smoothness* $0 < \nu < \infty$, *then*

$$\sup_{g \in \mathcal{H}(t,f)} \|g\|_{\mathcal{H}(\mathbb{X})} = \infty.$$

# Conclusion & perspectives

- reGP $\to$ a goal-oriented GP-based model to good obtain predictive distributions on a range of function values (if we accept degraded predictions outside the range of interest)
- Easy to use
- Low / moderate algorithmic complexity
- EGO-R seems preferable to EGO and can sometimes achieve significant convergence acceleration
- Main future work: extend reGP to the case of regression