DE LA RECHERCHE À L'INDUSTRIE

www.cea.fr

Nested polynomial trends for the improvement of Gaussian process-based predictors

G. Perrin[(1)], C. Soize[(2)], J. Garnier[(3)], S. Marque-Pucheu[(1,3)]

[(1)] CEA/DAM/DIF, F-91297, Arpajon, France,
[(2)] Université Paris-Est, MSME UMR 8208 CNRS, Marne-la-Vallée, France
[(3)] Laboratoire de Probabilités et Modèles Aléatoires, Laboratoire Jacques-Louis Lions, Université Paris Diderot, 75205 Paris Cedex 13, France

Ecole thématique ETICS - Barcelonnette |

June 2016

Outline

Context of uncertainty quantification

- For the last decades, the use of simulation has kept increasing for the analysis of complex and non-linear physical systems.
- Computational models are introduced:
  - to **optimize** the system or its maintenance w.r.t. constraints (geometry, cost, certification criteria...)
  - to **explore** the design space (prototyping, better understanding of the phenomena...)
  - to **evaluate** its robustness and its reliability.

Context of uncertainty quantification

When interested in a complex phenomenon $\boldsymbol{x} \mapsto y(\boldsymbol{x})$, two kinds of problems can be distinguished:

Forward problems

Given information about $\boldsymbol{x}$, we would like to:

1. **infer** the distribution of $\boldsymbol{x}$,
2. **compute** some statistical quantities of y, such as:
   - its mean and its variance,
   - probabilities of exceeding threshold,
   - its full density.

Backward problems

Given information about $\boldsymbol{x}$ and y, we would like to:

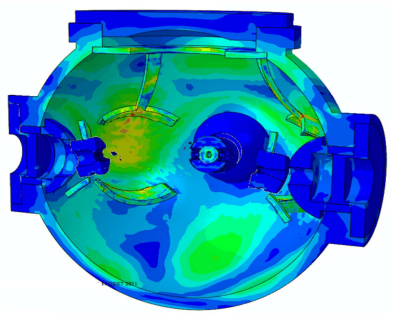
1. **Construct** a parametric model $\mathcal{M}(\cdot; \boldsymbol{\beta})$ for $\boldsymbol{x} \mapsto y(\boldsymbol{x})$,
2. **calibrate** parameters $\boldsymbol{\beta}$ and validate model $\mathcal{M}$,
3. **classify** the influence of each input on the variability of y.

The information about $\boldsymbol{x}$ and $\boldsymbol{y}$ is generally a set of experiments (or code evaluations).

Problematic

Most of the methods associated with these problems are based on the post-processing of large sets of simulations. When the numerical cost associated with one simulation is very high (for instance $\sim$100h on 100 cores in parallel), we need to know how to

- build **cheap mathematical approximations** of the quantities of interest,
- **control the relevance** of these approximations,
- **use these approximations** to solve the problem we are confronted to (by optimized sequential designs for instance).



Outline

Gaussian process-based regression (GPR)

- Let $\mathcal{S}$ be a physical system, which response depends on a d-dimensional input vector $\boldsymbol{x} = (x_1, \ldots, x_d)$, and which performance can be evaluated from the computation of a quantity of interest, $y(\boldsymbol{x})$.
- Function y is a **deterministic** mapping that is assumed to be an element of $L^2(\mathcal{D}^d, \mathbb{R})$, where $\mathcal{D}^d$ is a compact set.
- We suppose that the maximal available information about y is a set of N code evaluations, $\left\{(\boldsymbol{x}^{(1)}, y(\boldsymbol{x}^{(1)})), \ldots, (\boldsymbol{x}^{(N)}, y(\boldsymbol{x}^{(N)}))\right\}$.
- Each computation of $y(\boldsymbol{x})$ is computationally **expensive**, such that N is relatively small compared to the complexity of y.
- Given $\mathcal{F}_N$, the σ-algebra associated with this information, we are interested in the identification of the *best* predictor $y^\star$ of y:

$$\|y - y^\star\|_{L^2}^2 \;\leq\; \|y - \widehat{y}\|_{L^2}^2\,, \quad \widehat{y} \in L^2(\mathcal{D}^d, \mathbb{R}).$$

$$(u, v)_{L^2} := \int_{\mathcal{D}^d} u(\boldsymbol{x}) v(\boldsymbol{x}) d\boldsymbol{x}, \quad \|u\|_{L^2}^2 := (u, u)_{L^2}\,, \quad u, v \in L^2(\mathcal{D}^d, \mathbb{R}).$$

Gaussian process-based regression (GPR)

Hypothesis [Sacks et al., 1989],[Santner et al., 2003]

$\boldsymbol{x} \mapsto y(\boldsymbol{x})$ is supposed to be a realization of a Gaussian process (GP),

$$Y \sim \mathsf{GP}(\mu, C).$$

If the mean and the covariance functions μ and C are known, then :

$$Y \mid \mathcal{F}_N \sim \mathsf{GP}(\mu_N, C_N),$$

- $\mu_N(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})^T [C]^{-1} (\mathbb{Y} - \boldsymbol{\mu})$,
- $C_N(\boldsymbol{x}, \boldsymbol{x}') = C(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{r}(\boldsymbol{x})^T [C]^{-1} \boldsymbol{r}(\boldsymbol{x}')$,

and $\mathbb{E}[Y \mid \mathcal{F}_N] = \mu_N$ is the **best predictor** of y in the L^2 sense.

$\mathbb{Y} = \left(y(\boldsymbol{x}^{(1)}), \ldots, y(\boldsymbol{x}^{(N)})\right)$, $\boldsymbol{\mu} = \left(\mu(\boldsymbol{x}^{(1)}), \ldots, \mu(\boldsymbol{x}^{(N)})\right)$,
$[C]_{n,m} = C(\boldsymbol{x}^{(n)}, \boldsymbol{x}^{(m)})$, $\boldsymbol{r}(\boldsymbol{x}) = \left(C(\boldsymbol{x}, \boldsymbol{x}^{(1)}), \ldots, C(\boldsymbol{x}, \boldsymbol{x}^{(N)})\right)$.

Gaussian process-based regression (GPR)

Parametric formulation

Assuming that μ and C belong to classes that are parameterized by $\boldsymbol{\beta}$ and $\boldsymbol{\Theta}$, such that:

$$Y \mid \boldsymbol{\beta}, \boldsymbol{\Theta} \ \sim \ \mathrm{GP}(\mu(\boldsymbol{\beta}), C(\boldsymbol{\Theta})),$$

for all $\boldsymbol{x}$ in $\mathcal{D}^d$, the best predictor of $y(\boldsymbol{x})$ is given by:

$$\mathbb{E}[Y(\boldsymbol{x}) \mid \mathcal{F}_N] \ = \ \int_{\boldsymbol{\beta},\boldsymbol{\Theta},y} y \times \pi[Y(\boldsymbol{x}) = y \mid \boldsymbol{\beta}, \boldsymbol{\Theta}, \mathcal{F}_N] \, \pi[\boldsymbol{\beta}, \boldsymbol{\Theta} \mid \mathcal{F}_N] \, d\boldsymbol{\beta} d\boldsymbol{\Theta} dy,$$

where:

- $\pi[Y(\boldsymbol{x}) = y \mid \boldsymbol{\beta}, \boldsymbol{\Theta}, \mathcal{F}_N]$ is Gaussian,
- $\pi[\boldsymbol{\beta}, \boldsymbol{\Theta} \mid \mathcal{F}_N]$ is the **posterior** joint distribution of $(\boldsymbol{\beta}, \boldsymbol{\Theta})$.

Gaussian process-based regression (GPR)

Linearized plug-in approach [Bect et al., 2012], [Bichon et al., 2008]

- As a good compromise between complexity, efficiency, and errors control, the "linearized plug-in" approach (or "Universal Kriging") is generally preferred to this "full-Bayesian" approach.
- This method consists in :
  1. computing the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Theta}$, which are denoted by $\boldsymbol{\beta}^\star$ and $\boldsymbol{\Theta}^\star$,
  2. linearizing the mean function around $\boldsymbol{\beta}^\star$, such that:

$$\mu(\boldsymbol{x};\boldsymbol{\beta}) \approx c_0(\boldsymbol{x}) + \langle \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{\beta} \rangle \,,$$
$$c_0(\boldsymbol{x}) := \mu(\boldsymbol{x};\boldsymbol{\beta}^\star) - \langle \boldsymbol{f}(\boldsymbol{x}), \boldsymbol{\beta}^\star \rangle \,, \quad \boldsymbol{f}(\boldsymbol{x}) := \frac{\partial \mu}{\partial \boldsymbol{\beta}}(\boldsymbol{x};\boldsymbol{\beta}^\star),$$

  3. reparameterizing: $\boldsymbol{f}(\boldsymbol{x}) \leftarrow (c_0(\boldsymbol{x}), \boldsymbol{f}(\boldsymbol{x}))$, $\boldsymbol{\beta} \leftarrow (1, \boldsymbol{\beta})$,
  4. assuming that $\boldsymbol{\beta} \sim \mathcal{U}_{\mathbb{R}^{|\boldsymbol{\beta}|_0}}$, and conditioning all the results by $\boldsymbol{\Theta}^\star$.

Gaussian process-based regression (GPR)

Linearized plug-in approach [Bect et al., 2012], [Bichon et al., 2008]

It comes:

$$Y \mid \boldsymbol{\Theta}^\star, \mathcal{F}_N \;\sim\; \mathsf{GP}(\widehat{\mu}_N, \widehat{C}_N),$$

$$\widehat{\mu}_N(\boldsymbol{x}) = \left\langle \boldsymbol{f}(\boldsymbol{x}), \widehat{\boldsymbol{\beta}} \right\rangle + \boldsymbol{r}(\boldsymbol{x})^T [C]^{-1} (\mathbb{Y} - [F]\widehat{\boldsymbol{\beta}}),$$

$$\widehat{C}_N(\boldsymbol{x}, \boldsymbol{x}') = C(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{r}(\boldsymbol{x})^T [C]^{-1} \boldsymbol{r}(\boldsymbol{x}') + \boldsymbol{u}(\boldsymbol{x})^T ([F]^T [C]^{-1} [F])^{-1} \boldsymbol{u}(\boldsymbol{x}').$$

$$[F] = [\boldsymbol{f}(\boldsymbol{x}^{(1)}) \;\cdots\; \boldsymbol{f}(\boldsymbol{x}^{(N)})]^T, \; \widehat{\boldsymbol{\beta}} = ([F]^T [C]^{-1} [F])^{-1} [F]^T [C]^{-1} \mathbb{Y},$$
$$\boldsymbol{u}(\boldsymbol{x}) = [F]^T [C]^{-1} \boldsymbol{r}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}).$$

An analytic expression is found back for the best predictor of $y(\boldsymbol{x})$ in any non-computed point $\boldsymbol{x} \in \mathcal{D}^d$:

$$\mathbb{E}[Y(\boldsymbol{x}) \mid \mathcal{F}_N] \approx y^{\text{lin-plug}}(\boldsymbol{x}) := \widehat{\mu}_N(\boldsymbol{x}).$$

Gaussian process-based regression (GPR)

One of the main advantage of this approximated approach comes from the fact that the relevance of the "linearized plug-in" predictor of y can "easily" be computed **without any additional evaluation** of y from **cross-validation** procedures
[Miller, 1974, Dubrule, 1983, Blatman and Sudret, 2011, Bachoc, 2013]:

$$\left\| y - y^{\text{lin-plug}} \right\|_{L_2}^2 \approx \epsilon_{\text{LOO}}^2 := \frac{1}{N} \sum_{n=1}^{N} \left(y(\boldsymbol{x}^{(n)}) - y_{(-n)}^{\text{lin-plug}}(\boldsymbol{x}^{(n)}) \right)^2 ,$$

$$y(\boldsymbol{x}^{(n)}) - y_{(-n)}^{\text{lin-plug}}(\boldsymbol{x}^{(n)}) = \frac{([R]\mathbb{Y})_n}{[R]_{nn}},$$

$$[R] = [C]^{-1} \left\{ [I] - [F]([F]^T[C]^{-1}[F])^{-1}[F]^T[C]^{-1} \right\}.$$

Outline

Optimization of the GPR

To optimize the results associated with such a plug-in approach, two directions can be explored:

- work on the **parameterization of the covariance function** (stationary or not, regularity...),
- work on the **parameterization of the mean function** (linear or not with respect to β, sparse representations...).

Optimization of the GPR - covariance function

- By definition, for all $\boldsymbol{x}, \boldsymbol{x}'$ in $\mathcal{D}^d$:

$$C(\boldsymbol{x}, \boldsymbol{x}') := \mathbb{E}\left[\left(Y(\boldsymbol{x}) - \mathbb{E}\left[Y(\boldsymbol{x})\right]\right) \times \left(Y(\boldsymbol{x}') - \mathbb{E}\left[Y(\boldsymbol{x}')\right]\right)\right],$$

  such that C is *a priori* any symmetric and non-negative definite kernel that is defined on $\mathcal{D}^d \times \mathcal{D}^d$.
- Remembering that the actual goal is to predict function y, which is defined on $\mathcal{D}^d$ only, it is clear that only "simple" parametric expressions of C can be considered (from a limited number of evaluations of y, we do not pretend to be able to precisely identify C).

Optimization of the GPR - covariance function

A very commonly used representation for C is the following tensorized Matern expression:

$$C(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \prod_{i=1}^{d} \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(2\sqrt{\nu} h_i\right)^{\nu} \mathcal{B}_{\nu}^{III}\left(2\sqrt{\nu} h_i\right),$$

where:

- $h_i = g_i\left(\frac{|x_i - x_i'|}{\ell_i}\right)$,
- $\nu \leftrightarrow$ regularity of y,
- $\sigma^2 \leftrightarrow$ *a priori* uncertainty,
- $\ell_i \leftrightarrow$ correlation lengths,
- $g_i \leftrightarrow$ scaling function (to take into account the fact that the model can be non-stationary with respect to x_i).

Optimization of the GPR - covariance function

- All the parameters associated with covariance function C have to be identified from the available evaluations of y, or be *a priori* fixed (from expert judgment for instance...).
- With only very few information about y and its regularity is available, it is generally accepted that the **Matern-5/2** class is a good *a priori* choice for C, which corresponds to the case when the realizations of Y are twice mean-square differentiable:

$$C(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2(1 + \sqrt{h} + (5/3)h^2) \times \exp(-\sqrt{5}\ h),$$

$$h = \sum_{i=1}^{d} |x_i - x_i'|/\ell_i.$$

  - Fixed parameters: $\nu = 5/2$, $\gamma = \sqrt{5}$, $g(h) = h$,
  - Free parameters: σ^2, $\{\ell_1, \ldots, \ell_d\}$.

Optimization of the GPR - mean function

- By definition, for all $\boldsymbol{x}$ in $\mathcal{D}^d$:

$$\mu(\boldsymbol{x}; \boldsymbol{\beta}) := \mathbb{E}\left[Y(\boldsymbol{x})\right],$$

  such that μ is *a priori* any function that is defined on $\mathcal{D}^d$.
- Any parametric expression can therefore be used to characterize μ.
- For instance, one can think to linear or non-linear functions such as polynomials (which can be orthogonal or not in $L^2(\mathcal{D}^d, \mathbb{R})$), cos and sin, neural networks, Low Rank representations...

Optimization of the GPR - mean function

- Once again, without information about y, polynomials are generally chosen for $\boldsymbol{f}$. Indeed, the set $\left\{m_{\boldsymbol{\alpha}},\ \boldsymbol{\alpha} \in \mathbb{N}^d\right\}$, with

$$m_{\boldsymbol{\alpha}}(\boldsymbol{x}) := x_1^{\alpha_1} \times \cdots \times x_d^{\alpha_d}, \quad \boldsymbol{x} \in L^2(\mathcal{D}^d, \mathbb{R}),$$

  defines a basis of $L^2(\mathcal{D}^d, \mathbb{R})$.
- $\boldsymbol{\beta} \mapsto \mu(\boldsymbol{\beta})$ is therefore chosen **linear**, and for any given value of M, the number of polynomials we want to consider for the representation of μ, the idea is to identify the best M-dimensional subset of $\left\{m_{\boldsymbol{\alpha}},\ \boldsymbol{\alpha} \in \mathbb{N}^d\right\}$ to minimize $\|y - \mathbb{E}\left[Y \mid \mathcal{F}_N\right]\|_{L^2}$.
- At last, convergence analyses based on the formerly introduced LOO error can be carried out to identify the most accurate value of M.

Optimization of the GPR - mean function

- In practice, this very complex optimization problem is replaced by an optimization over a **finite** dimensional subset of $\left\{m_{\boldsymbol{\alpha}},\ \boldsymbol{\alpha} \in \mathbb{N}^d\right\}$.
- Different truncation schemes have thus been proposed to choose such a relevant subset, which are mostly based on the assumption that the most influential elements of $\left\{m_{\boldsymbol{\alpha}},\ \boldsymbol{\alpha} \in \mathbb{N}^d\right\}$ correspond to the elements of **lowest total polynomial order**:

$$\mathcal{P}(r,d) := \left\{ m_{\boldsymbol{\alpha}} \mid \boldsymbol{\alpha} \in \mathbb{N}^d,\ \sum_{i=1}^{d} |\alpha_i| \leq r \right\}.$$

Optimization of the GPR - mean function

- A **penalization technique**, such as the $\ell - 1$ or the $\ell - 2$ penalizations or the Least Angle Regression (LAR) method [Hastie et al., 2002, Efron et al., 2004, Blatman and Sudret, 2011] are generally used to find these M most significant terms in $\mathcal{P}(r, d)$.
- Cross validation procedures are once again used to avoid overfitting.
- Such an approach will be referred as "LAR+UK" approach in the following [Kersaudy et al., 2015].

Optimization of the GPR - conclusions

Finally, given $\mathcal{F}_N$, the "LAR+UK" approach follows the following steps:

1. identify the M most interesting polynomials to represent y, which are gathered in the vector-valued function $\boldsymbol{f} = (f_1, \ldots, f_M)$, such that:

$$\mu(\boldsymbol{\beta}) := \langle \boldsymbol{f}, \boldsymbol{\beta} \rangle ,$$

2. choose an *a priori* adapted parametric expression for $C(\boldsymbol{\Theta})$,
3. assume that y is a particular realization of $Y \sim \mathrm{GP}\left(\mu(\boldsymbol{\beta}), C(\boldsymbol{\Theta})\right)$,
4. identify $\boldsymbol{\Theta}^\star$ the maximum likelihood estimation of $\boldsymbol{\Theta}$,
5. compute the posterior distribution of Y that is conditioned by $\mathcal{F}_N$, $\boldsymbol{\Theta}^\star$ and $\boldsymbol{f}$.

The optimized GPR is thus given by the **mean function** of this conditioned Gaussian process, $\mathbb{E}\left[Y(\boldsymbol{x}) \mid \mathcal{F}_N, \boldsymbol{\Theta}^\star, \boldsymbol{f}\right]$.

Optimization of the GPR - conclusions

$$\mathbb{E}\left[Y(\boldsymbol{x}) \mid \mathcal{F}_N, \boldsymbol{\Theta}^\star, \boldsymbol{f}\right] = y^{\text{trend}}(\boldsymbol{x}) + y^{\text{cond}}(\boldsymbol{x}),$$

$$y^{\text{trend}}(\boldsymbol{x}) := \left\langle \boldsymbol{f}(\boldsymbol{x}), \widehat{\boldsymbol{\beta}} \right\rangle$$
$$y^{\text{cond}}(\boldsymbol{x}) := \boldsymbol{r}(\boldsymbol{x})^T [C(\boldsymbol{\Theta}^\star)]^{-1} (\mathbb{Y} - [F]\widehat{\boldsymbol{\beta}}).$$

Comments

- when N tends to infinity, the role of y^{trend} becomes negligible,
- on the contrary, when N is relatively small compared to the complexity of y, the role of y^{trend} is crucial.

Outline

Nested polynomial trends

- When N, the number of code evaluations, is low compared to the complexity of y, such approaches are limited by the fact that **only low values** of M, the dimension of the projection family, can be considered to avoid extra-fitting.
- In order to be able to deal with higher values of M, without increasing the number of unknown parameters to be identified, we can propose a new parameterization of the polynomial trend, which is based on a **nested structure**.
- It will then be shown to what extent such an approach shows very promising results when only very limited information is available ($\leftrightarrow$ N is small).

Nested polynomial trends

Main ideas

- a **composition of two polynomials** for the mean function, μ, of the GP associated with $\boldsymbol{x} \mapsto y(\boldsymbol{x})$ is proposed,
- the composition of polynomials being polynomials, it is thus possible to span a **large subset** of $L^2(\mathcal{D}^d, \mathbb{R})$ from a **small number** of independent parameters.

$$\mu(\boldsymbol{x}; \boldsymbol{b}_1, \boldsymbol{b}_2) := \left\langle \boldsymbol{m}^{(p_2)} \left(\left\langle \boldsymbol{m}^{(p_1)}(x_1), \boldsymbol{b}_1 \right\rangle \right), \boldsymbol{b}_2 \right\rangle,$$

where $\boldsymbol{m}^{(p)}$ gathers all the polynomials, which total degree is less than $p \in \mathbb{N}^*$.

- $\boldsymbol{b}_1 \leftrightarrow$ parameterization of the inner polynomials,
- $\boldsymbol{b}_2 \leftrightarrow$ parameterization of the outer polynomials.

Nested polynomial trends

Main difficulties

1. the fonction

$$\boldsymbol{b}_1 \mapsto \left\langle \boldsymbol{m}^{(p_2)} \left(\left\langle \boldsymbol{m}^{(p_1)}(x_1), \boldsymbol{b}_1 \right\rangle \right), \boldsymbol{b}_2 \right\rangle,$$

   is strongly non-linear.

2. different values of $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ could lead to the same nested-representation.

In that prospect, it has been proposed in [Perrin et al., 2016] to work on:

- a **minimal parametrization** of this nested structure to avoid redundancies,
- a **linearization** of the mean around the maximimun likelihood estimates of $(\boldsymbol{\Theta}, \boldsymbol{b}_1, \boldsymbol{b}_2)$,
- **iterative algorithms** to solve the maximum likelihood maximization.

Nested polynomial trends

Comments on the proposed nested representation

- contrary to the "LAR+UK" approach that looks for **sparse** representations of μ, the nested representation looks for **"full"** polynomial representations, which are however characterized by a very **limited number of independent coefficients**. This is particularly interesting for the modeling of complex phenomena with very limited information.
- For $d > 1$, it allows us to model separately the **dependency** structure between the different input parameters, and the **individual** actions of each input parameter. Hence, analyzing the final structure of the mean function can give us information about the structure of y (is the model additive up to a transformation of its input parameters or are the dependencies more complex?).

Nested polynomial trends

For $d = 1$ and $\mathcal{D}^d = [-1, 1]$, three analytic examples are proposed:

- case 1: $y(x) = P_2^{(4)} \circ P_1^{(4)}(x)$,
- case 2: $y(x) = \sin((x+1)^3)$,
- case 3: $y(x) = \sin(20x)\cos(2x)$,

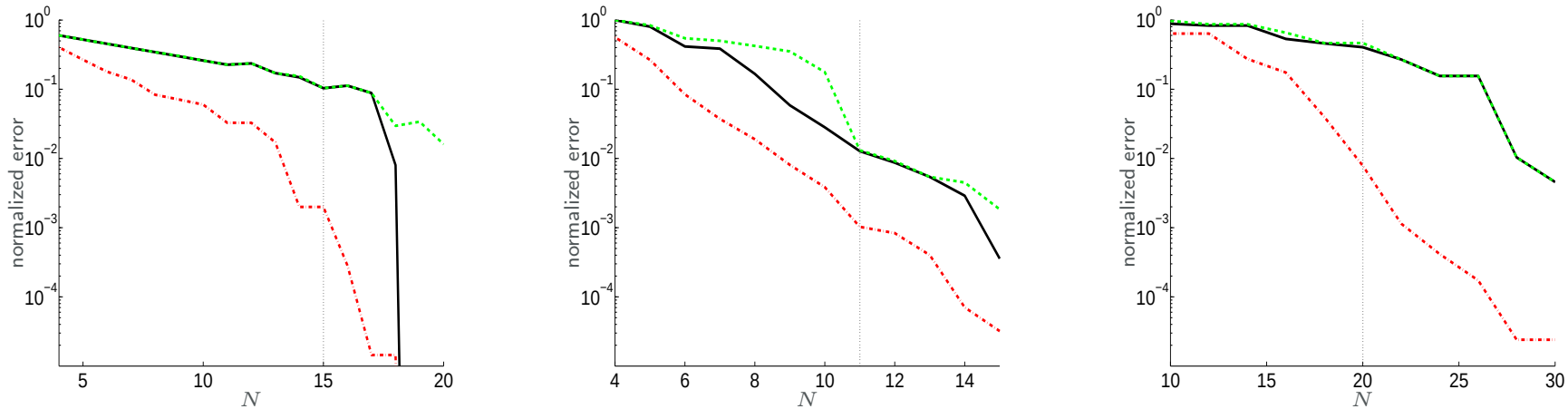


Figure: Mean value of the errors associated with these 10 repetitions. Solid black line: error associated with the LAR+UK approach. Red dotted line: error associated with the nested approach. Green dotted line: UK with linear trend

Nested polynomial trends

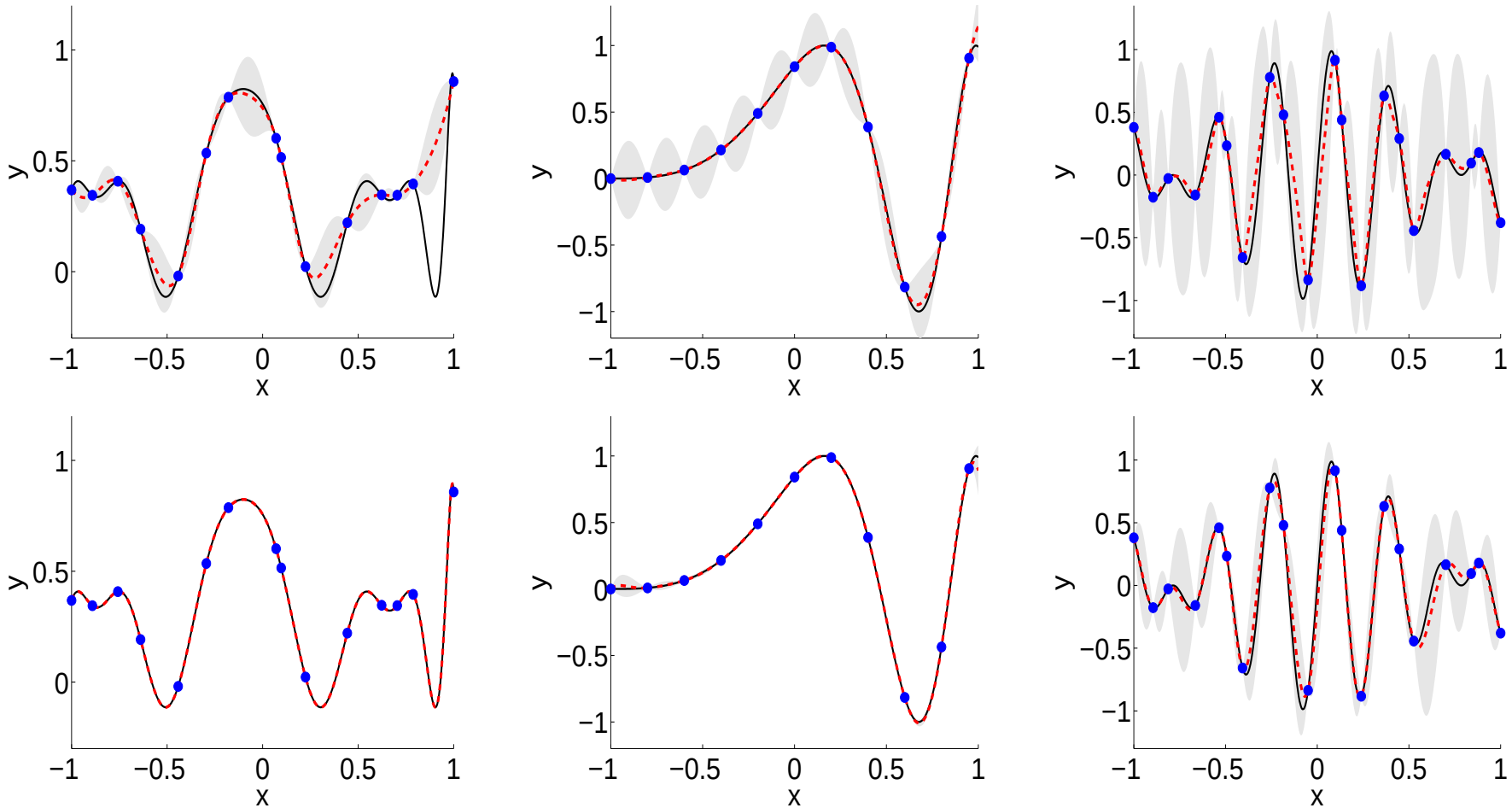


Figure: black $\leftrightarrow y_2$, blue points $\leftrightarrow$ observations, red $\leftrightarrow \widehat{\mu}_N$. top : the LAR+UK approach, bottom : the proposed approach, grey areas $\leftrightarrow 95\%$ confidence region for the prediction. Left: case 1, middle: case 2, right: case 3.

Nested polynomial trends

For $d > 1$, the same kinds of results are found:

- case 1: $y(\boldsymbol{x}) = (1 - x_1^2)\cos(7x_1) \times (1 - x_2^2)\sin(5x_2)$,
- case 2: $y(\boldsymbol{x}) = \sin(x_1) + 7\sin(x_2)^2 + 0.1\sin(x_1)x_3^4$ (**Ishigami**),
- case 3: $y(\boldsymbol{x}) = g^{(1)} \circ g^{(2)}(\boldsymbol{x})$, $g^{(1)}(\boldsymbol{z}) = 0.1\cos\left(\sum_{i=1}^{6} z_i\right) + \sum_{i=1}^{6} z_i^2$, $\boldsymbol{g}^{(2)}(\boldsymbol{x}) = (\cos(\pi x_1 + 1), \cos(\pi x_2 + 2), \ldots, \cos(\pi x_6 + 6))$.

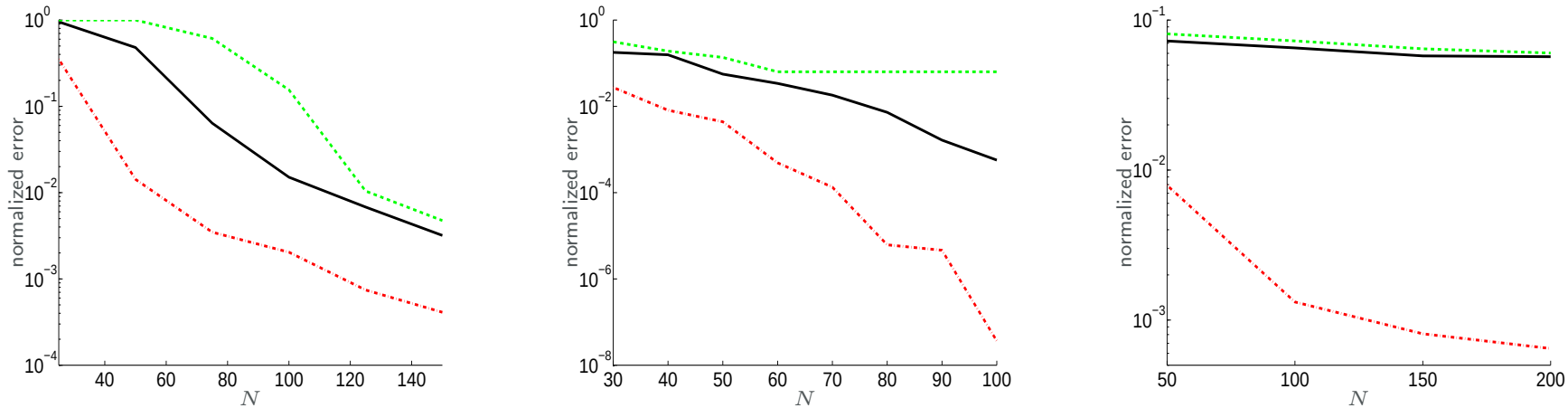


Figure: Mean value of the errors associated with these 10 repetitions. Solid black line: error associated with the LAR+UK approach. Red dashed line: error associated with the nested approach. Green dotted line: UK with linear trend.

Outline

Conclusions and prospects

- A lot of methods in uncertainty quantification and computer experiments require **many code evaluations**.
- When the numerical cost associated with each evaluation is high, **surrogate models** are generally introduced.
- When interested in deterministic mappings (when the dimension of the input space is not too large), one of the most used method is the **GPR**.
- **Working on the polynomial trend** can strongly improve the relevance of the GPR.
- When little information about the code is available, considering **nested polynomial trends** shows promising results.
- Enabling these techniques to deal with **high dimensional problem** ($d \gg 1$), even if a lot of code evaluations ($N \gg 1$) are available is still an open questions.

Conclusion

Thank you for your attention.

References


Bachoc, F. (2013).
Estimation paramétrique de la fonction de covariance dans le modèle de Krigeage par processus Gaussiens. Application à la quantification des incertitudes en simulation numérique.
PhD thesis, University Paris Diderot, France.


Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vasquez, E. (2012).
Sequential design of computer experiments for the estimation of a probability of failure.
Statistics and Computing, 22.


Bichon, B., Eldred, M., Swiler, L., Mahadevan, S., and McFarland, J. (2008).
Efficient global reliability analysis for non linear implicit performance functions.
AIAA Journal, 46(10).

References II

Blatman, G. and Sudret, B. (2011).
Adaptative sparse polynomial chaos expansion based on least angle regression.
Journal of Computational Physics, 230.

Dubrule, O. (1983).
Cross validation of kriging in a unique neighborhood.
Mathematical Geology, 15(6):687–699.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004).
Least angle regression.
Ann. Stat., 32:407–499.

Hastie, T., Tibshirani, R., and Friedman (2002).
Elements of Statistical Learning.
Springer, New York.

References III

Kersaudy, P., Sudret, B., Varsier, N., and Picon, O. (2015).
A new surrogate modeling technique combining kriging and polynomial chaos expansions - application to uncertainty analysis in computational dosimetry.
Journal of Computational Physics, 286:103–117.

Miller, R. G. (1974).
The jackknife - a review.
Biometrika, 61:1–15.

Perrin, G., Soize, C., Garnier, J., and Marque-Pucheu, S. (2016).
Nested polynomial trends for the improvement of gaussian process-based predictors.
Journal of Computational Physics, in review.

References IV

- Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4:409–435.
- Santner, T. J., Williams, B., and Notz, W. (2003). *The design and analysis of computer experiments*. Springer, New York.