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

G. Perrin a,*, C. Soize b, S. Marque-Pucheu a,c, J. Garnier c

a CEA/DAM/DIF, F-91297, Arpajon, France
b Université Paris-Est, MSME UMR 8208 CNRS, Marne-la-Vallée, France
c Laboratoire de Probabilités et Modèles Aléatoires, Laboratoire Jacques-Louis Lions, Université Paris Diderot, 75205 Paris Cedex 13, France

A R T I C L E   I N F O

Article history:
Received 4 April 2016
Received in revised form 24 February 2017
Accepted 31 May 2017
Available online 15 June 2017

Keywords:
Computer experiment
Gaussian process
Nested polynomial trend
Bayesian framework

A B S T R A C T

The role of simulation keeps increasing for the sensitivity analysis and the uncertainty quantification of complex systems. Such numerical procedures are generally based on the processing of a huge amount of code evaluations. When the computational cost associated with one particular evaluation of the code is high, such direct approaches based on the computer code only, are not affordable. Surrogate models have therefore to be introduced to interpolate the information given by a fixed set of code evaluations to the whole input space. When confronted to deterministic mappings, the Gaussian process regression (GPR), or kriging, presents a good compromise between complexity, efficiency and error control. Such a method considers the quantity of interest of the system as a particular realization of a Gaussian stochastic process, whose mean and covariance functions have to be identified from the available code evaluations. In this context, this work proposes an innovative parametrization of this mean function, which is based on the composition of two polynomials. This approach is particularly relevant for the approximation of strongly non linear quantities of interest from very little information. After presenting the theoretical basis of this method, this work compares its efficiency to alternative approaches on a series of examples.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

The numerical cost of many codes to simulate complex physical systems is still very high. To perform sensitivity analyses, uncertainty quantification or reliability studies, these computer models have therefore to be replaced by surrogate models, that is to say by fast and inexpensive mathematical functions. Within the computational science community, when the maximal available information is a finite set of code evaluations, the most widely used surrogate models are the generalized polynomial chaos expansion (PCE) \cite{olgac20061,holmberg2007,orgilsson2008,allaire2010,allaire2013,allaire2014} and the Gaussian process regression (GPR), or kriging (see \cite{rasmussen2006,seeger2009,quillfeldt2014}).

On the one hand, the main idea of PCE is to expand the code output, which is denoted by \( g \) in the following, onto an appropriate basis made of orthonormal multivariate polynomials, which are related to the distribution of the code input variables. As the number of unknown expansion coefficients usually grows exponentially with the number of input parameters, the relevance of these approaches strongly depends on their ability to select the most relevant basis functions. To this end, several penalization techniques, such as the \( 1_1 \)-minimization \cite{allaire2013,allaire2014} and the least Angle Regression (LAR) methods \cite{tibshirani1996,efron2004}, have been introduced to select polynomial basis sets that lead to more accurate PCE than would have been

* Corresponding author.
E-mail address: guillaume.perrin2@cea.fr (G. Perrin).

http://dx.doi.org/10.1016/j.jcp.2017.05.051
0021-9991/© 2017 Elsevier Inc. All rights reserved.
obtained if the basis is a priori fixed. Taking advantage of the tensor-product structure of the multivariate polynomial basis, separated representations, such as low-rank approximations [23,19], have alternatively been proposed to develop surrogate models with polynomial functions in highly-compressed formats.

On the other hand, the GPR is based on the assumption that the code output is a particular realization of a Gaussian stochastic process, \( Y \). This hypothesis, which was first introduced in time series analysis [26] and in optimization [20], is widely used as it allows dealing with the conditional probability and expectation, while leading to very interesting results in terms of computer code prediction. Hence, contrary to the PCE, the GPR is not associated with an a priori projection basis, but requires the introduction of the mean and the covariance functions of \( Y \). In practice, we observe that the role of the mean function of \( Y \) on the prediction decreases when the number of code evaluations increases. This explains that in applications where many code evaluations are available, good GPR-based surrogate models can be obtained using constant or linear trends for the mean function. On the contrary, when the number of code evaluations is low compared to the complexity of \( g \), it can be very useful to optimize it. In that case, searching the mean function of \( Y \) as a well-chosen sum of polynomial functions can indeed strongly improve the relevance of the associated GPR. In particular, the authors refer to [16] and [18] for an illustration of the interest of using variable selection techniques to optimize this polynomial representation of the mean function of \( Y \).

Following on these works, the idea of this paper is to propose an alternative parametrization of the mean function of \( Y \), which is particularly adapted to the case when the number of code evaluations is small compared to the complexity of \( g \). Instead of searching sparse polynomial approximations, we look for high dimensional polynomial approximations that are characterized by a small number of parameters. In other words, if we want to model a complex code response with a very limited number of code evaluations, we believe that it can be more efficient to use complex but approximated models. We thus propose to consider the composition of two polynomials for the mean function of \( Y \). Indeed, the composition of two polynomial functions is still a polynomial function, but of much higher order. In particular, such a formalism can be used to model separately a transformation of each code input and the dependence structure between them.

The main difficulty concerning this specific representation is the identification of the parameters of the two combined polynomials. Indeed, by composing two polynomial functions that are linear with respect to their parameters, we get a strongly non-linear representation, which is likely to be very sensitive to small changes in the parameters values. In addition, distinct values for these parameters can lead to the same nested representation, which does not help for the identification. To avoid such redundancies, minimal nested parametrizations are introduced, and we show to what extent integrating this nested structure in the Gaussian process formalism can increase the robustness of the results, make easier the error control, and limit as much as possible over-fitting.

The outline of this work is as follows. First, Section 2 presents the theoretical framework for the definition of a Gaussian-process regression with a linear polynomial trend. Then, the nested polynomial trends we propose are detailed in Section 3. At last, the efficiency of the method is illustrated on a series of analytic examples in Section 4.

## 2. Gaussian process predictors

### 2.1. General framework

For \( d \geq 1 \), let \( L^2(D_d, \mathbb{R}) \) be the space of square integrable functions on any compact subset \( D_d \) of \( \mathbb{R}^d \), with values in \( \mathbb{R} \), equipped with the inner product \((.,.)_{L^2} \), and the associated norm \( \|\cdot\|_{L^2} \), such that for all \( u \) and \( v \) in \( L^2(D_d, \mathbb{R}) \),

\[
(u,v)_{L^2} := \int_{D_d} u(x)v(x)\text{d}x, \quad \|u\|_{L^2}^2 := (u,u)_{L^2}.
\]  

(1)

Let \( S \) be a physical system, whose response depends on a \( d \)-dimensional input vector \( x = (x_1, \ldots, x_d) \), and whose performance can be evaluated from the computation of a quantity of interest, \( g(x) \). Function \( g \) is a deterministic mapping that is assumed to be an element of \( L^2(D_d, \mathbb{R}) \). In this work, it is supposed that the maximal available information about \( g \) is a set of \( N \) code evaluations at the points \( \{x^{(1)}, \ldots, x^{(N)}\} \) in \( D_d \). Given this information, we are interested in the identification of the best predictor \( g^* \) of \( g \), in the sense that:

\[
\forall \hat{g} \in L^2(D_d, \mathbb{R}), \quad \|g - g^*\|_{L^2}^2 \leq \|g - \hat{g}\|_{L^2}^2.
\]  

(2)

In that context, the Gaussian process regression (GPR), or kriging, plays a major role [30,24,31,34]. It is indeed able to provide a prediction of \( g(x) \), which is optimal in the class of the linear predictors of \( g \), and whose precision can be a posteriori quantified. Such a method considers function \( g \) as a sample path of a real-valued Gaussian stochastic process \( Y \), which is defined on the probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Let \( \mu \) and \( C \) be respectively the mean and the covariance functions of \( Y \):

\[
Y \sim \text{GP}(\mu, C).
\]  

(3)

We can introduce \( \mathcal{F}_N \) the \( \sigma \)-algebra generated by the available information about \( g \),
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات