



# Local optimization of black-box functions with high or infinite-dimensional inputs: application to nuclear safety

Angelina Roche

## ► To cite this version:

Angelina Roche. Local optimization of black-box functions with high or infinite-dimensional inputs: application to nuclear safety. Computational Statistics, 2018, 33 (1), pp.467-485. 10.1007/s00180-017-0751-1 . hal-01144628v3

**HAL Id: hal-01144628**

**<https://hal.science/hal-01144628v3>**

Submitted on 17 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# LOCAL OPTIMIZATION OF BLACK-BOX FUNCTION WITH HIGH OR INFINITE-DIMENSIONAL INPUTS. APPLICATION TO NUCLEAR SAFETY.

ANGELINA ROCHE

**ABSTRACT.** An adaptation of Response Surface Methodology (RSM) when the covariate is of high or infinite dimensional is proposed, providing a tool for black-box optimization in this context. We combine dimension reduction techniques with classical multivariate Design of Experiments (DoE). We propose a method to generate experimental designs and extend usual properties (orthogonality, rotatability,...) of multivariate designs to general high or infinite dimensional contexts. Different dimension reduction basis are considered (including data-driven basis). The methodology is illustrated on simulated functional data and we discuss the choice of the different parameters, in particular the dimension of the approximation space. The method is finally applied to a problem of nuclear safety.

**Keywords:** Functional data analysis. Response surface methodology. Design of experiments.

**AMS Subject Classification 2010:** 62K20, 62K15.

## 1. INTRODUCTION

Black-box optimization problems arises in many applications, for instance when one wants to optimise an output of a computer code or in real-life experiments such as crash test, chemical reactions, medical experiments... In more and more applications, the input is high-dimensional, or even infinite-dimensional (time or space dependent). In this paper, our aim is to minimise an unknown function  $m : \mathcal{R} \rightarrow \mathbb{R}$  where  $\mathcal{R}$  is a subset of a separable Hilbert space  $(\mathbb{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$ , which can be e.g.  $\mathbb{R}^d$  or a function space. The function  $m$  is unknown, but noisy evaluations of  $m$  are available. We suppose that each evaluation of  $m$  is costly, thus the aim is to be as close as possible to an optimum with a given (low) number of evaluations of  $m$ .

In this context, surrogate-based approaches, such as those based on response-surface methodology, kriging, radial basis functions, splines or neural networks are commonly used (Queipo et al., 2005; Simpson et al., 2001). In Response Surface Methodology (RSM), the function  $m$  is locally approximated by a polynomial regression model, typically with order 1 or 2 (see e.g. the review of Khuri and Mukhopadhyay, 2010). With the information given by the evaluation of the function  $m$  on design points chosen by the user, least-squares estimates of the model parameters provide a local approximation of the surface  $y = m(x)$ . This local approximation can be used, for instance, to approximate the gradient or to locate a critical point of the surface. The efficiency of the method rests mainly on the choice of the design of experiments. Hence, a lot of research has been done in order to find sets of points giving the best possible precision of the fitted surface with few evaluations

of  $m$  on the design points. We refer to Khuri and Mukhopadhyay 2010, pp. 131–133 for the description of the most common response surface designs. This subject is still an active research field (see e.g. Georgiou et al. 2014). However, these design generating methods are not tractable when the inputs are high-dimensional (for instance,  $\mathbb{H} = \mathbb{R}^d$  with  $d \geq 100$ ) and can not be directly defined with infinite-dimensional inputs.

Projection-based dimension reduction techniques has become a powerful tool in high-dimensional statistics and are the main tool of most of the methods used to treat functional data. Usually, the data are projected into a subset  $S = \text{Vect}\{\varphi_1, \dots, \varphi_d\} \subset \mathbb{H}$  of reduced dimension. In functional data analysis,  $\mathbb{H}$  is a function space and fixed basis such as Fourier basis, spline or wavelet basis are commonly used, exploiting the regularity properties (smoothness for instance) of the functions in the sample. Another interesting approach consists in using the information given in a learning sample of pairs  $\{(X_i, Y_i), i = 1, \dots, n\}$  to generate the directions  $\varphi_1, \dots, \varphi_d$ . Among them the approaches based on the principal components are the most common: PCA (Hall, 2011), sparse PCA (Zou et al., 2006; Qi and Luo, 2015), regularized PCA (Rice and Silverman, 1991; Lee et al., 2002; Ramsay and Silverman, 2005). Another approach to obtain interesting data-driven basis is Partial Least Squares regression (PLS, Wold, 1975; Preda and Saporta, 2005). The main advantage of PLS is that the directions  $\varphi_1, \dots, \varphi_d$  are chosen in  $\mathbb{H}$  so as to maximize the explained variance of  $Y$ . Hence, PLS uses the information of the whole learning sample  $\{(X_i, Y_i), i = 1, \dots, n\}$  whereas the PCA basis is generated only with the  $X_i$ 's.

The present paper proposes an adaptation of Response Surface Methodology when the input is in a general Hilbert space  $\mathbb{H}$ , via dimension reduction techniques. The specificities of the framework are explained in Section 2. In Section 3, we provide a way to generate RSM design of experiments based on dimension reduction tools. A simulation study is presented in Section 4. The method is finally applied in Section 5 to a nuclear safety problem.

## 2. HIGH-DIMENSIONAL AND FUNCTIONAL CONTEXT

We suppose here that our real response  $y$  depends on a variable  $x$  in an infinite or high-dimensional space  $\mathbb{H}$  equipped with a scalar product  $\langle \cdot, \cdot \rangle$ . For instance,  $\mathbb{H}$  may be the space  $\mathbb{R}^d$  equipped with its usual scalar product  $\langle x, y \rangle = \sum_{j=1}^d x_j y_j$  or a function space, such that  $\mathbb{L}^2(I)$  with  $I$  a measurable subset of  $\mathbb{R}^d$  and  $\langle f, g \rangle = \int_I f(t)g(t)dt$ . We denote by  $\|\cdot\|$  the associated norm ( $\|x\|^2 = \langle x, x \rangle$  for all  $x \in \mathbb{H}$ ).

We suppose here that the function  $m$  is sufficiently smooth, so that the surface  $y = m(x)$  can be approximated reasonably by a first or second-order surface. We consider here generalizations of the classical multivariate first and second-order models. Higher-order polynomial models, or even generalized linear models (Müller and Stadtmüller, 2005; Khuri, 2001), while less standard, could also be considered similarly.

**2.1. First-order model.** We define first-order models in the following form

$$y := \alpha + \langle \beta, x \rangle + \varepsilon,$$

with  $\alpha \in \mathbb{R}$ ,  $\beta \in \mathbb{H}$  and  $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ . If  $\mathbb{H}$  is a function space, this model is known as *functional linear model* (Ramsay and Dalzell, 1991; Cardot et al., 1999) and has been widely studied (see Cardot and Sarda 2011 for a recent overview or Brunel et al. 2016 for a recent work on this subject). Moreover, it is known that, if  $\mathbb{H}$  is of high or infinite dimension, least-squares estimators of the slope parameter  $\beta$  are, in general, unstable. Hence, precautions must be taken, either in the choice of the design (see Section 3), or in the estimation method for the parameter  $\beta$  which is an ill-posed inverse problem (Engl

et al., 1996). The interest of this model is that, if  $m$  is differentiable, for all  $x_0 \in \mathbb{H}$ ,  $m(x) = m(x_0) + \langle m'(x_0), x - x_0 \rangle + o(\|x - x_0\|)$ , where  $m'(x_0)$  is the gradient of  $m$  at the point  $x_0$ , which implies that, if  $x$  is sufficiently close to  $x_0$ , an estimator of  $\beta$  will estimate the gradient of the surface  $y = m(x)$  around the point  $x_0$ .

**2.2. Second-order model.** A second-order model can also be written as follows

$$(1) \quad y := \alpha + \langle \beta, x \rangle + \frac{1}{2} \langle Hx, x \rangle + \varepsilon,$$

where  $\alpha \in \mathbb{R}$ ,  $\beta \in \mathbb{H}$  and  $H : \mathbb{H} \rightarrow \mathbb{H}$  is a linear self-adjoint operator. Up to our knowledge, this model (at least in this particular form) has not been studied yet in the literature of functional data analysis. As for the first-order linear model, classical least-squares estimation is not a good choice and we have to be careful either in the choice of the design, or in the estimation method. If  $\mathbb{H} = \mathbb{R}^d$ , the operator  $H$  is a symmetric and positive semi-definite matrix  $H = (h_{i,j})_{1 \leq i,j \leq d}$  and the model (1) is the classical second-order multivariate linear model which can be written

$$y = \alpha + \sum_{j=1}^d \beta_j x_j + \sum_{\substack{j,k=1 \\ j < k}}^d h_{j,k} x_j x_k + \sum_{j=1}^d h_{j,j} x_j^2 + \varepsilon.$$

As for the second-order model, if  $m$  is twice differentiable and if the design points are sufficiently close to  $x_0$ , an estimator of  $\beta$  is an estimator of the gradient of  $m$  and an estimator of  $H$  gives an estimator of the Hessian matrix of  $m$ . In particular, if  $x_0$  is close to a critical point, then the estimator of  $\beta$  is close to 0 and the estimator of  $H$  may help to precise the exact location and the nature of the critical point.

### 3. GENERATION OF DESIGN OF EXPERIMENTS

**3.1. General principle.** The method is based on dimension reduction coupled with classical multivariate designs. The main idea is the following: suppose that we want to generate a design around  $x_0 \in \mathbb{H}$ , we choose an orthonormal basis  $(\varphi_j)_{j \geq 1}$  of  $\mathbb{H}$ , a dimension  $d$  and a  $d$ -dimensional design  $\{\mathbf{x}_i, i = 1, \dots, n\} = \{(x_{i,1}, \dots, x_{i,d}), i = 1, \dots, n\}$  around  $0 \in \mathbb{R}^d$  and we define a functional design  $\{x_i, i = 1, \dots, n\}$  verifying

$$(2) \quad x_i := x_0 + \sum_{j=1}^d x_{i,j} \varphi_j.$$

The advantage of such a method is its flexibility: all multivariate designs and all basis of  $\mathbb{H}$  can be used. Then, by choosing an appropriate design and an appropriate basis, we can generate designs satisfying some constraints defined by the context.

Remark that  $\{\varphi_1, \dots, \varphi_d\}$  can be seen as predefined optimisation directions in the sense that the input  $x$  will vary exclusively in the directions of the space  $\text{span}\{\varphi_1, \dots, \varphi_d\}$ . Therefore their choice have a great influence on the precision of the results and has to be made carefully.

**3.2. Data-driven directions of optimisation.** If a training sample  $\{(X_i, Y_i), i = 1, \dots, n\}$  is available, with  $X_i \in \mathbb{H}$ , for all  $i$  and  $\mathbb{E}[Y_i] = m(X_i)$ , it may be relevant to use the information of this sample to find a suitable basis. We consider here two method to generate data-driven basis:

- Principal Components which is the basis of  $\mathbb{H}$  verifying

$$\frac{1}{n} \sum_{i=1}^n \|X_i - \hat{\Pi}_d X_i\|^2 = \min_{\Pi_d} \left\{ \frac{1}{n} \sum_{i=1}^n \|X_i - \Pi_d X_i\|^2 \right\},$$

where  $\hat{\Pi}_d$  is the orthogonal projector on  $\text{span}\{\varphi_1, \dots, \varphi_d\}$ ,  $\|\cdot\|$  is a norm on the space  $\mathbb{H}$  and the minimum on the right-hand side is taken over all orthogonal projectors  $\Pi_d$  on  $d$ -dimensional subspaces of  $\mathbb{H}$ .

- Partial Least Squares (Wold, 1975; Preda and Saporta, 2005) which permits to take into account the interaction between  $X$  and  $Y$ . It is computed iteratively by the procedure described in Delaigle and Hall (2012). For theoretical results on the PLS basis see Delaigle and Hall (2012); Blazère et al. (2014). For practical implementation, see Algorithm 1.

**Data:** Training sample  $\{(X_i, Y_i), i = 1, \dots, n\}$

Initialization :

$$X_i^{[0]} = X_i - \frac{1}{n} \sum_{j=1}^n X_j, \quad Y_i^{[0]} = Y_i - \frac{1}{n} \sum_{j=1}^n Y_j$$

**for**  $j = 1, \dots, d$  **do**

Estimate  $\varphi_j$  by the empirical covariance of  $X_i^{[j-1]}$  and  $Y_i^{[j-1]}$ :

$$\varphi_j = \sum_{i=1}^n Y_i^{[j-1]} X_i^{[j-1]} / \left\| \sum_{i=1}^n Y_i^{[j-1]} X_i^{[j-1]} \right\|$$

Fit the models  $Y_i^{[j-1]} = \beta_j \langle X_i^{[j-1]}, \varphi_j \rangle + \varepsilon_i^{[j]}$  and  $X_i^{[j-1]} = \delta_j \langle X_i^{[j-1]}, \varphi_j \rangle + W_i^{[j]}$  by least-squares that is

$$\hat{\beta}_j := \sum_{i=1}^n Y_i^{[j-1]} \langle X_i^{[j-1]}, \varphi_j \rangle / \sum_{i=1}^n \langle X_i^{[j-1]}, \varphi_j \rangle^2$$

and

$$\hat{\delta}_j := \sum_{i=1}^n \langle X_i^{[j-1]}, \varphi_j \rangle X_i^{[j-1]} / \sum_{i=1}^n \langle X_i^{[j-1]}, \varphi_j \rangle^2$$

Define  $X_i^{[j]} := X_i^{[j-1]} - \langle X_i^{[j-1]}, \varphi_j \rangle \hat{\delta}_j$  and  $Y_i^{[j]} := Y_i^{[j-1]} - \hat{\beta}_j \langle X_i^{[j-1]}, \varphi_j \rangle$  the residuals of the two fitted models;

**end**

**Algorithm 1:** practical implementation of PLS basis (Delaigle and Hall, 2012, Section A.2)

**3.3. Multivariate designs.** In this article, we focus on the most classical designs. However, the method we propose is flexible and can be used with any multivariate design, for instance Latin Hypercube Sampling (see Liu et al., 2015, and references therein), small composite designs (Draper and Lin, 1990), augmented-pair designs (Morris, 2000)... We also refer to Georgiou et al. (2014) and references therein for the recent advances on the subject.

The  $2^d$  *factorial design* is one of the simplest. It is a *first-order design* in the sense that it is frequently used to fit a first-order linear model. For each explanatory variable  $x_1, \dots, x_d$ , we choose two levels (coded by  $+1$  and  $-1$ ) and we take all the  $2^d$  combinations of these two levels. When  $d$  is large, it may be impossible to achieve the  $2^d$  factorial experiments, hence *fractional factorial design* keeps only a certain proportion (e.g. a half,

a quarter,...) of points of a  $2^d$  factorial design. Typically, when a fraction  $1/(2^p)$  is kept from the original  $2^d$  design, this design is called  $2^{d-p}$  factorial design. The points removed are carefully chosen, we refer e.g. to Gunst and Mason (2009) for more details. In our context, since we have the freedom to choose the dimension  $d$ , the interest of considering fractional factorial designs relies on its flexibility. For a given number  $2^k$  of design points, all pairs  $(d, p)$  of positive integers such that  $d - p = k$  gives a different design, the choice  $d = k$  and  $p = 0$  leads to the full factorial design while larger values of  $p$  allow to explore a higher-dimensional space keeping the number of experiments low.

Traditional second-order designs are factorial designs, central composite designs and Box-Behnken designs.

- $3^d$  or  $3^{d-p}$  factorial designs are similar to  $2^d$  and  $2^{d-p}$  factorial designs but with three levels (+1, -1 and 0).
- *Central Composite Designs (CCD)* are obtained by adding to the two-level factorial design (fractional or not) two points on each axis of the control variables on both sides of the origin and at distance  $\alpha > 0$  from the origin.
- *Box-Behnken Designs (BBD)* are widely used in the industry. It is a well-chosen subset of the  $3^d$  factorial design. Box-Behnken designs are not used when  $d = 2$ . For  $d \geq 4$ , we refer to Myers et al. (2009, 7.4.7).

**3.4. Design properties.** One of the interests of the design generation method(2) is that all design properties (orthogonality, rotatability and alphabetic optimality) verified by the multivariate design  $\{\mathbf{x}_i, i = 1, \dots, n\}$  are also verified for the design  $\{x_i, i = 1, \dots, n\} \subset \mathbb{H}$ .

To explain that we focus on first-order and second-order designs but the same reasoning may apply to other models and other kind of optimality properties related to the model considered. Let us first rewrite these models.

*First-order model.* Recall that, for all  $i = 1, \dots, n$ ,  $x_i = x_0 + \sum_{j=1}^d x_{i,j} \varphi_j$ , then the first-order model can be rewritten

$$(3) \quad y_i := \alpha + \langle \beta, x_0 \rangle + \sum_{j=1}^d x_{i,j} \langle \beta, \varphi_j \rangle + \varepsilon_i, \text{ for } i = 1, \dots, n.$$

With our choice of design points, this model is a first-order multivariate model and can be written

$$(4) \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,d} \\ 1 & x_{2,1} & & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,d} \end{pmatrix}$$

and coefficients  $\boldsymbol{\beta} = (\alpha + \langle \beta, x_0 \rangle, \langle \beta, \varphi_1 \rangle, \dots, \langle \beta, \varphi_d \rangle)^t$ . Then, the first-order linear model in  $\mathbb{H}$  is in fact a first-order multivariate linear model, with inputs  $\{\mathbf{x}_i, i = 1, \dots, n\}$ .

*Second-order model.* Now we can see that a similar conclusion holds for the second-order model, which can also be written  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  with

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,d} & x_{1,1}^2 & x_{1,1}x_{1,2} & \dots & x_{1,d}^2 \\ 1 & x_{2,1} & & x_{2,d} & x_{2,1}^2 & x_{2,1}x_{2,2} & \dots & x_{2,d}^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,d} & x_{n,1}^2 & x_{n,1}x_{n,2} & \dots & x_{n,d}^2 \end{pmatrix}$$

3.4.1. *Design properties.* An important property is the *orthogonality*. An orthogonal design is a design for which the matrix  $\mathbf{X}^t\mathbf{X}$  is diagonal. This implies that the vector  $\hat{\boldsymbol{\beta}}$  is also a Gaussian random vector with independent components and makes it easier to test the significance of the components of  $\boldsymbol{\beta}$  in the model.  $2^d$  factorial designs are orthogonal first-order designs. However, fractional designs have to be constructed carefully in order to keep the orthogonality property. For second-order designs, we refer to Box and Hunter (1957) for general criteria applied to factorial and fractional factorial designs. Central Composite Designs are orthogonal if  $\alpha = \sqrt{\left(\sqrt{F(F+2d+n_0)} - F\right)/2}$ , where  $F$  is the number of points of the initial factorial design (see Myers et al. 2009).

A design is said to be *rotatable* if  $\text{Var}(\hat{y}(\mathbf{x}))$  depends only on the distance between  $\mathbf{x}$  and the origin. This implies that the prediction variance is unchanged under any rotation of the coordinate axes. We refer to Box and Hunter (1957) for conditions of rotatability. All first-order orthogonal designs are also rotatable. This is not the case for second-order designs, for instance a CCD design is rotatable if  $\alpha = F^{1/4}$  which means that a CCD design can be rotatable and orthogonal only for some specific values of  $n_0$  and  $F$ . Box-Behnken designs are rotatable for  $d = 4$  and  $d = 7$ . Some measures of rotatability have been introduced (Khuri, 1988; Draper and Guttman, 1988; Draper and Pukelsheim, 1990; Park et al., 1993) in order to measure how close a design is to the rotatability property.

The important point is that all the design properties cited above only depends on the design matrix  $\mathbf{X}$ . Hence, all properties of the multivariate design  $\{\mathbf{x}_i, i = 1, \dots, n\}$  are automatically verified for the design  $\{x_i, i = 1, \dots, n\}$ .

Since all alphabetic optimality criteria (Pázman, 1986) are also exclusively based on properties of the design matrix  $\mathbf{X}$ , it is possible to define e.g.  $D$ -optimal designs for data in  $\mathbb{H}$  with Equation (2) by taking a  $D$ -optimal multivariate design.

#### 4. NUMERICAL EXPERIMENTS

In this section,  $\mathbb{H} = \mathbb{L}^2([0, 1])$ .

4.1. **Functional designs.** We use here the functions *cube*, *ccd* and *bbd* of the R-package *rsm* (Lenth, 2009) to generate respectively  $2^d$  factorial designs, Central Composite Designs (CCD) and Box-Behnken Designs (BBD).

*Functional designs with Fourier basis.* In this section, we set  $\varphi_1 \equiv 1$  and for all  $j \geq 1$ , for all  $t \in [0, 1]$ ,

$$\varphi_{2j}(t) = \sqrt{2} \cos(2\pi jt) \text{ and } \varphi_{2j+1}(t) = \sqrt{2} \sin(2\pi jt).$$

The curves of the generated designs are given in Figure 1.

*Functional design with data-driven bases.* We simulate a sample  $\{X_1, \dots, X_n\}$  comprising of  $n = 500$  realizations of the random variable

$$X(t) = \sum_{j=1}^J \sqrt{\lambda_j} \xi_j \psi_j(t),$$

with  $J = 50$ ,  $\lambda_j = e^{-j}$ ,  $(\xi_j)_{j=1, \dots, J}$  an i.i.d. sequence of standard normal random variables and  $\psi_j(t) := \sqrt{2} \sin(\pi(j - 0.5)t)$ .

The PCA basis only depends on  $\{X_i, i = 1, \dots, n\}$ . In order to see the influence of the law of  $Y$  on the PLS basis we define two training samples  $\{(X_i, Y_i^{(j)}), i = 1, \dots, n\}$  for

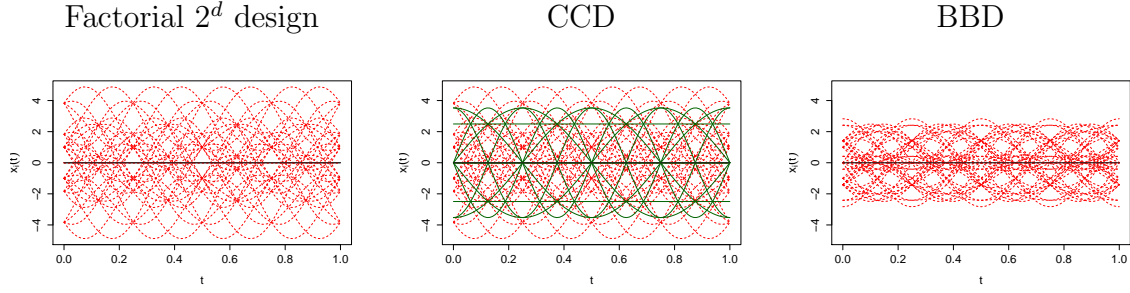


FIGURE 1. Functional designs with the Fourier basis ( $d = 5$ ). Gray thick line:  $x_0 \equiv 0$ , red lines: points of the original  $2^d$  or  $3^d$  (for BBD) factorial design, green dotted lines: points added to the factorial design (for CCD).

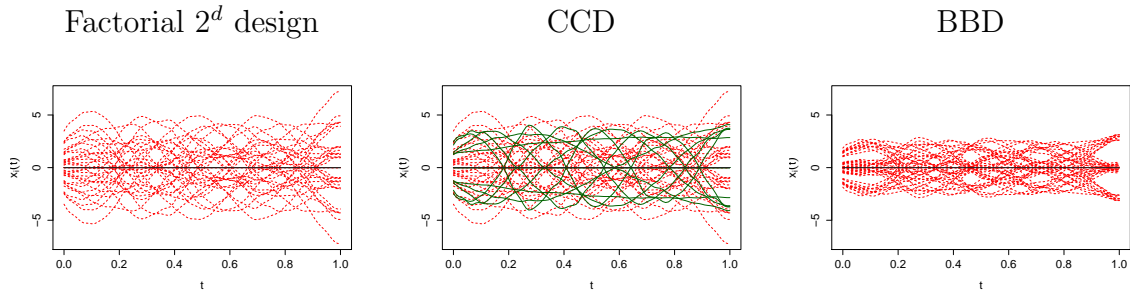


FIGURE 2. Functional designs with the PCA basis associated to  $\{X_i, i = 1, \dots, n\}$  ( $d = 5$ ). The legend is the same as the one of Figure 1.

$j = 1, 2$  with

$$Y_i^{(j)} := m_j(X_i) + \varepsilon_i,$$

$m_j(x) := \|x - f_j\|^2$ , where

$$\begin{aligned} f_1(t) &:= \cos(4\pi t) + 3\sin(\pi t) + 10, \\ f_2(t) &:= \cos(8.5\pi t) \ln(4t^2 + 10) \end{aligned}$$

and  $\varepsilon_1, \dots, \varepsilon_n$ , i.i.d.  $\sim \mathcal{N}(0, 0.01)$ .

The curves of the design generated by the PLS basis (Figure 3) are much more irregular than those generated by the PCA basis (Figure 2). However, remark that the designs generated by the PLS basis (Figure 3) of the two samples show significant differences, which illustrates that the PLS basis effectively adapts to the law of  $Y$ .

**4.2. Estimation of the response surface.** We use here the PLS basis calculated from the training sample  $\{(X_i, Y_i^{(j)}), i = 1, \dots, n\}$  with  $j = 1$  or  $j = 2$ . The aim is to approach the minimum  $f_j$  of  $m_j : x \in \mathbb{H} \rightarrow \|x - f_j\|^2$ .

We use the training sample a second time to determine the starting point of the algorithm. We take

$$x_0^{(0)} := X_{i_{\min}} \text{ where } i_{\min} := \arg \min_{i=1, \dots, n} \{Y_i\}.$$

The dimension is set to  $d = 8$ .

*Approximation of  $f_1 = \cos(4\pi t) + 3\sin(\pi t) + 10$ .*

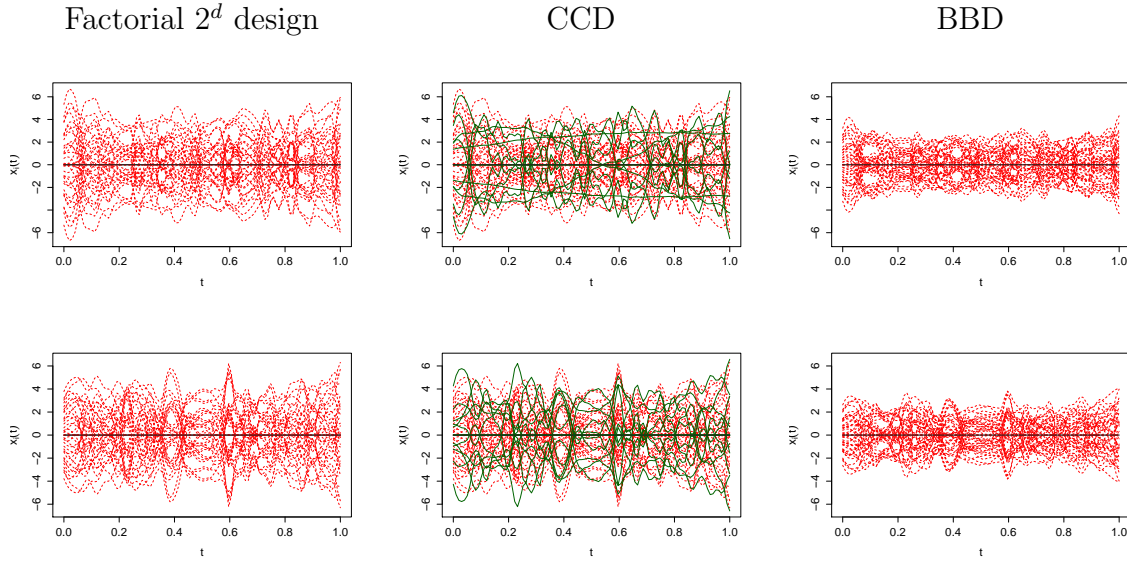


FIGURE 3. Functional designs with the PLS basis of the training sample  $\{(X_i, Y_i^{(j)}, i = 1, \dots, n\}$ ,  $j = 1$  (first line) and  $j = 2$  (second line),  $d = 5$ . The legend is the same as the one of Figure 1.

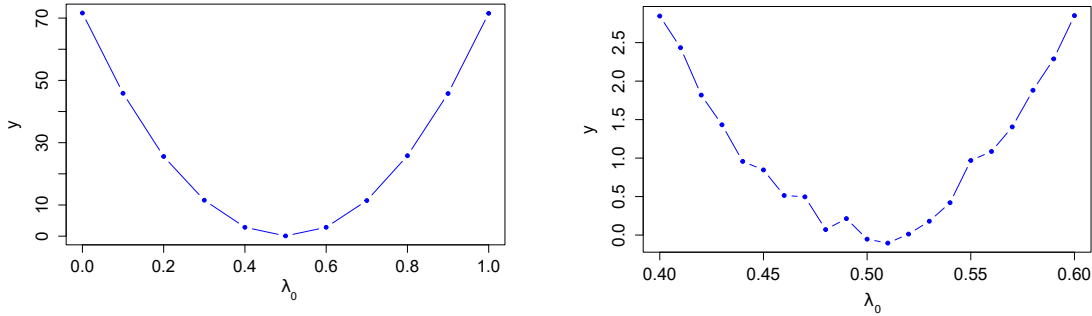


FIGURE 4. Results of experiments on the direction of steepest descent for the estimation of  $f_1$ .  $x$ -axis:  $\lambda_0$ ,  $y$ -axis: response  $Y = m_1(x_0^{(0)} - \lambda_0 \hat{\beta}^{(0)}) + \varepsilon$ .

**Descent step:** We generate a factorial  $2^d$  design (Figure 3 – left)  $(x_1^{(0)}, \dots, x_{n_0}^{(0)})$  (here  $n_0 = 2^d$ ) and we fit a first-order model

$$Y_i^{(0)} = \alpha^{(0)} + \sum_{j=1}^d \beta_j^{(0)} x_{i,j}^{(0)} + \varepsilon_i^{(0)},$$

to estimate the gradient. We realize two series of experiments along the line of steepest descent  $x_0^{(0)} - \lambda_0 \hat{\beta}^{(0)}$  ( $\lambda_0 > 0$ ). The first one (Figure 4– top left) suggests that the optimal value of  $\lambda_0$  is between 0.4 and 0.6 and with the results of the second one we fix  $\lambda_0 = 0.50$ . We set  $x_0^{(1)} := x_0^{(0)} - \lambda_0 \hat{\beta}^{(0)}$ .

The value of  $m_1$  at the starting point was  $m(x_0^{(0)}) = 70.4 \pm 0.1$ . At this step, we have  $m_1(x_0^{(1)}) = 6.33 \times 10^{-3} \pm 10^{-5}$  and we have done only  $2^d + 24 = 280$  experiments to reach this result.

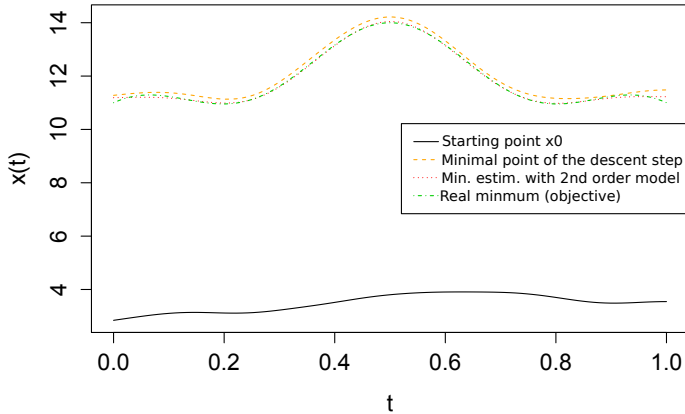


FIGURE 5. Result of optimization algorithm.

We fit a first-order model once again with a  $2^d$  factorial design and find that the norm of  $\hat{\beta}^{(1)}$  is very small ( $\|\hat{\beta}^{(1)}\| < 0.02$ ) compared to  $\|\hat{\beta}^{(0)}\| = 16.8 \pm 0.1$  which suggests that we are very close to a stationary point. We also note that the  $p$ -value of the Fisher's test  $H_0 : \beta_1^{(1)} = \dots = \beta_d^{(1)} = 0$  against  $H_1 : \exists j \in \{1, \dots, d\}, \beta_j^{(1)} \neq 0$  is very close to 1 which tends to confirm this assertion.

**Final step:** To improve the approximation, we fit a second-order model on the design points given by a Central Composite Design (Figure 3 – center). The matrix  $\hat{H}$  at this step is an estimation of the matrix of the restriction to the space  $\text{span}\{\varphi_1, \dots, \varphi_d\}$  of the Hessian operator of  $m$  at the point  $x_0^{(1)}$ . All the eigenvalues of  $\hat{H}$  are greater than  $1.96 > 0$ , this suggests that we are close to a minimum. We set  $x_0^{(2)} := -\hat{H}^{-1}\hat{\beta}^{(1)}$  and we have  $m_1(x_0^{(2)}) := 5.45 \times 10^{-3} \pm 10^{-5}$ . The CCD with  $d = 8$  counts 280 elements then we have realized 280 experiments for the descent step plus 280 for the final step, this rises to 557 the total number of experiments performed. Figure 5 represents the different results.

*Approximation of  $f_2(t) = \cos(8.5\pi t) \ln(4t^2 + 10)$ .* We have here  $m_2(x_0^{(0)}) = 2.88 \pm 0.01$ .

We follow the same steps as in the previous paragraph. Figure 6–left represents the evolution of the response along the direction of steepest descent. Here, since the response is noisy, refining the result without doing a too large number of experiments seems to be difficult. Then, we fix  $\lambda_0 = 0.5$  and  $x_0^{(1)} = x_0^{(0)} - \lambda_0 \hat{\beta}^{(0)}$ . We have  $m_2(x_0^{(1)}) = 1.99 \pm 0.01$ . At this step, we have improved the response of about 31%. This is not as important as the improvement of the first step of estimation of  $f_1$  but that is significant.

This time, the  $p$ -value of the Fisher's test  $H_0 : \beta_1^{(1)} = \dots = \beta_d^{(1)} = 0$  against  $H_1 : \exists j \in \{1, \dots, d\}, \beta_j^{(1)} \neq 0$  is very small ( $< 2 \times 10^{-4}$ ) which indicates that we are not close to a stationary point. Then, we try to improve the response doing a second descent step.

**4.3. Choice of basis.** In this section, we compare the three bases proposed in Section 4.1. We generate  $n_s = 50$  training samples of size  $n = 500$  and compare the results of the first descent step when the design is generated by the Fourier basis, the PCA basis and the PLS basis. The starting point is the same:  $x_0^{(0)} = X_{i_{\min}}$  for  $i_{\min} = \arg \min_{i=1, \dots, n} \{Y_i\}$  (then for the Fourier basis the training sample is only used to set the starting point). The

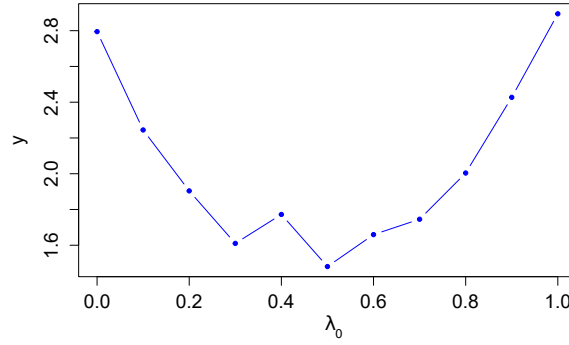


FIGURE 6. Results of experiments on the direction of steepest descent for the estimation of  $f_2$ .

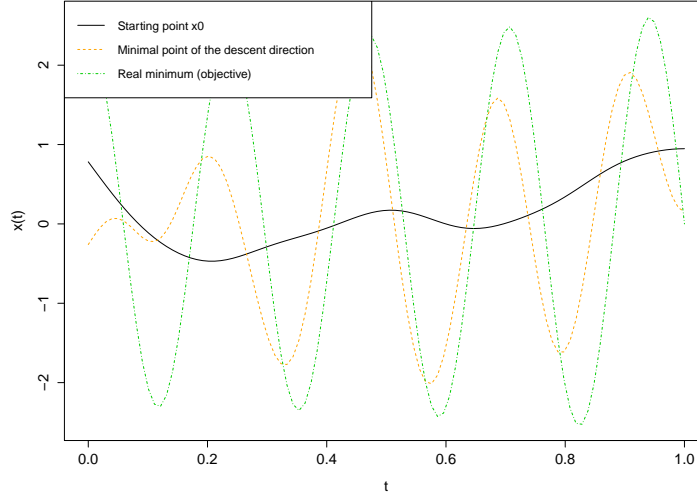


FIGURE 7. Result of optimization algorithm. Black curve:  $x_0^{(0)}$ , orange curve:  $x_0^{(1)}$ , red curve:  $x_0^{(2)}$ , green curve:  $f_2$ .

results are given in Figure 8. We see immediately that the PLS basis seems to be a better choice than the PCA one. However, surprisingly, the choice between the PLS basis and the Fourier basis is less clear.

**4.4. Choice of dimension  $d$ .** Figures 9 and 10 show that, except when the design is generated by the PCA basis for the approximation of  $f_1$ , the percentage of improvement increases when the dimension increases, which is coherent with the fact that the number of experiments grows exponentially with the dimension.

We then decide to study the properties of the method when the number of design points is fixed and the dimension  $d$  varies. For this purpose, we consider  $2^{d-p}$  fractional factorial designs. We can see on Figure 11 the results of the Monte-Carlo study. It seems that there is a significant improvement of the method when the dimension  $d$  increases. This suggests that it is always better to explore new dimensions, even if we perform less experiments along each direction.

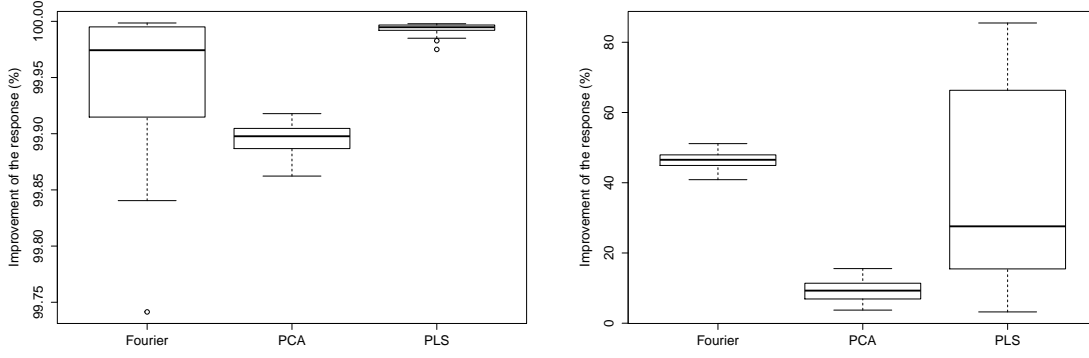


FIGURE 8. Monte-Carlo study of response improvement  $\frac{m(x_0^{(0)}) - m(x_0^{(1)})}{m(x_0^{(0)})}$  after the first descent step. Left-hand side: estimation of  $f_1$ , right-hand side: estimation of  $f_2$ .

## 5. APPLICATION TO NUCLEAR SAFETY

**5.1. Data and objectives.** An hypothetical cause of nuclear accident is the loss of coolant accident (LOCA). This is caused by a breach on the primary circuit. In order to avoid reactor meltdown, the safety procedure consists in incorporating cold water in the primary circuit. This can cause a pressurised thermal shock on the nuclear vessel inner wall which increases the risk of failure of the vessel.

The parameters influencing the probability of failure are the evolution over time of temperature, pressure and heat transfer in the vessel. Obviously, the behavior of the reactor vessel during the accident can be hardly explored by physical experimentation and numerical codes have been developed, for instance by the CEA<sup>1</sup>, reproducing the mechanical behavior of the vessel given the three mentioned parameters (temperature, pressure, heat transfer). Figure 12 represents different evolution of each parameter during the procedure depending on the value of several input parameters, which can be used as a learning sample.

The aim is to find the temperature transient which minimizes the risk of failure. We have access here to the margin factor (MF) which decreases when the risk of failure increases. Hence, the aim is to maximise the MF.

**5.2. Generation of design.** Considering that the inputs are the three temperature, pressure and heat penetration curves, we set  $\mathbb{H} = (\mathbb{L}^2([0, T]))^3$  (with  $T = 5000s$ ) equipped with the natural scalar product

$$\langle (x_1^{(T)}, x_1^{(P)}, x_1^{(H)}), (x_2^{(T)}, x_2^{(P)}, x_2^{(H)}) \rangle = \int_0^T x_1^{(T)}(t)x_2^{(T)}(t)dt + \int_0^T x_1^{(P)}(t)x_2^{(P)}(t)dt + \int_0^T x_1^{(H)}(t)x_2^{(H)}(t)dt.$$

We define the starting point of the algorithm as the triplet  $(X_i^{(T)}, X_i^{(P)}, X_i^{(H)})$  of the learning sample maximizing the response.

In view of the simulation results of Section 4 and the presence of a learning sample, we focus on the PLS basis and generate a functional design based on a minimum aberration

<sup>1</sup>French Alternative Energies and Atomic Energy Commission (Commissariat à l'énergie atomique et aux énergies alternatives), government-funded technological research organisation. <http://www.cea.fr/>

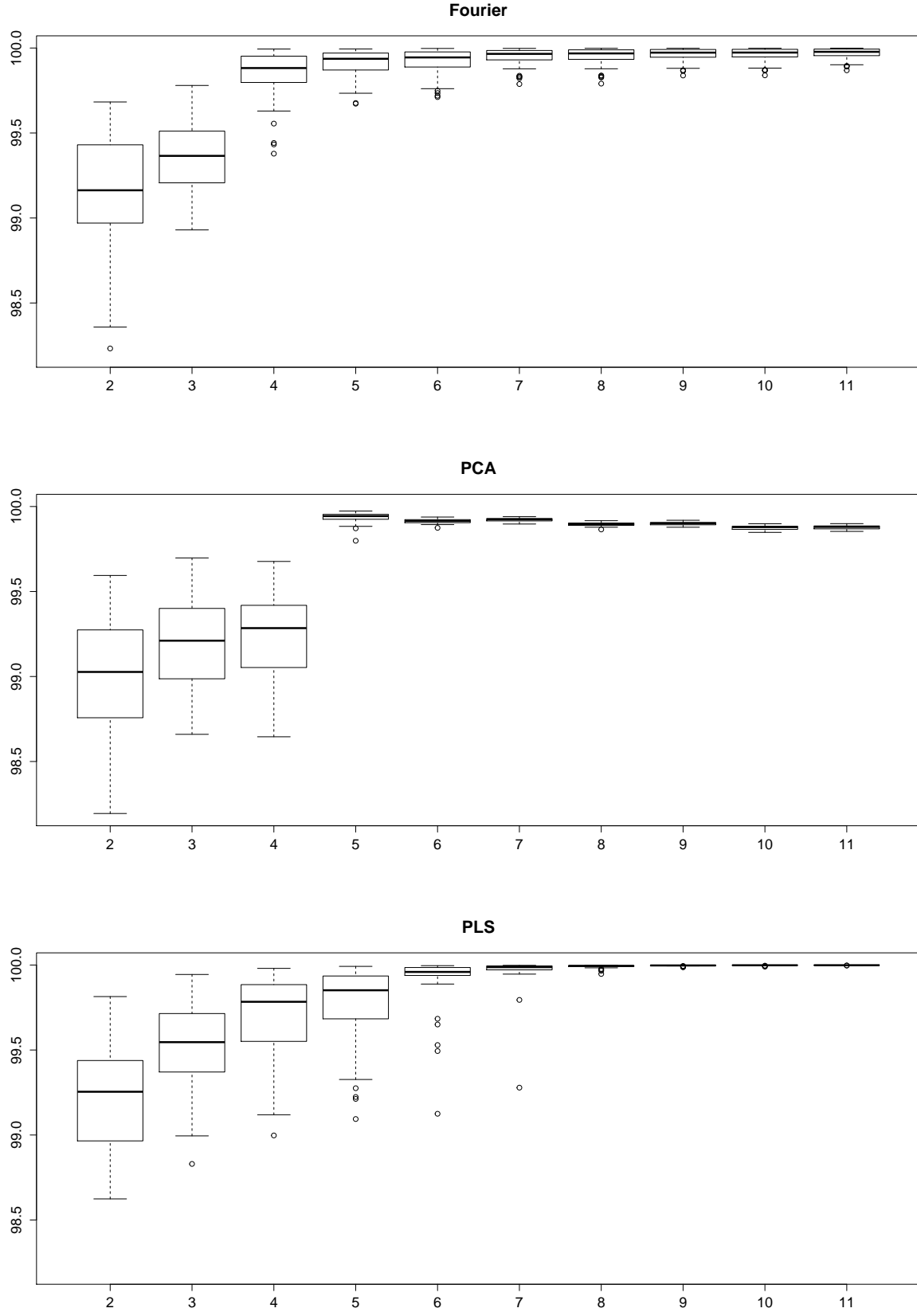


FIGURE 9. Monte-Carlo study of response improvement for the approximation of  $f_1$  as a function of the dimension  $d$ .

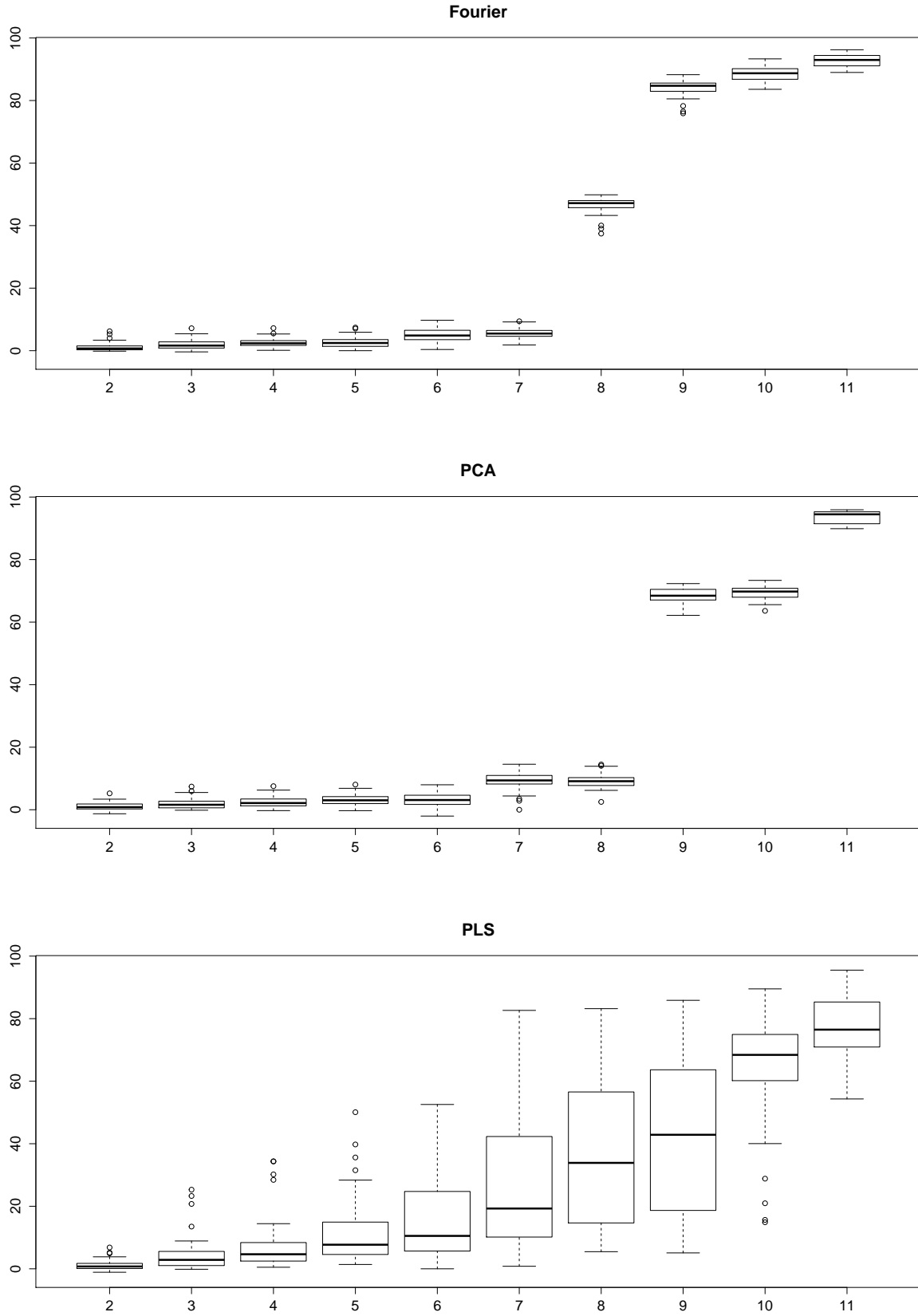


FIGURE 10. Monte-Carlo study of response improvement for the approximation of  $f_2$  as a function of the dimension  $d$ .

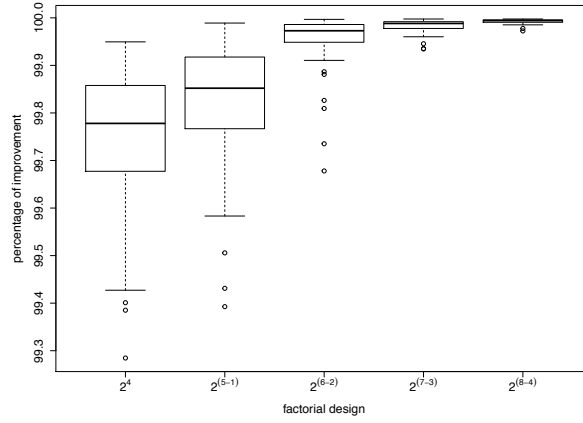


FIGURE 11. Monte-Carlo study of response improvement for the approximation of  $f_1$  with a  $2^{d-p}$  factorial design, for different values of  $d$  and  $p$  such that  $d - p = 4$ .

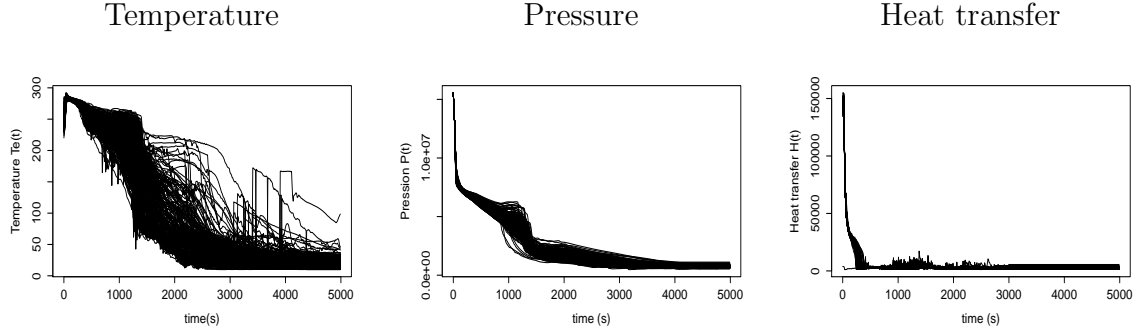


FIGURE 12. Evolution of temperature, pressure and heat transfer (learning sample). Source: CEA.

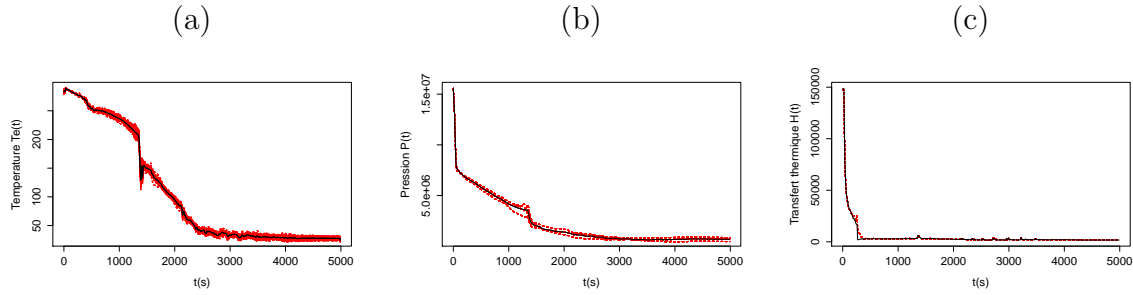


FIGURE 13. Functional experimental design around the initial curves.

$2^{10-5}$  fractional design for the temperature, a  $2^{3-2}$  design for the pressure and the heat transfer. As some design points of the functional design around the initial heat transfer curve took negative values (which can not correspond to the physic since the heat transfer is always positive), we remove it and keep only the design points which are always positive. The design points are plotted in Figure 13. The resulting design, which is a combination of all curves of the three designs obtained (for temperature, pressure and heat penetration) counts 128 design points.

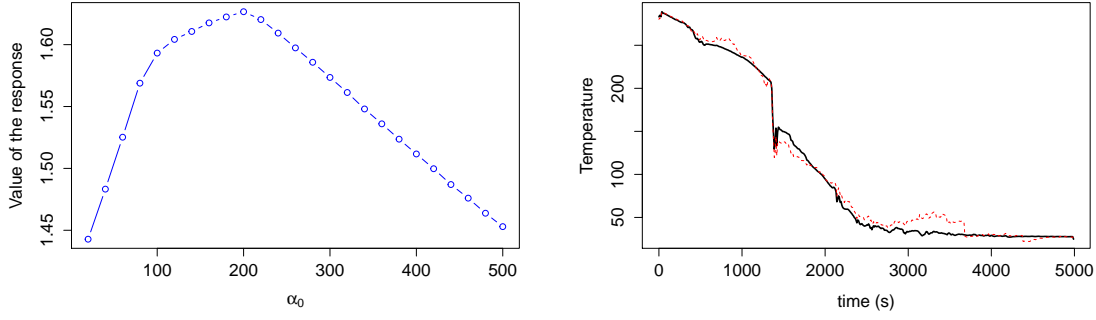


FIGURE 14. Left: value of the response on the estimated steepest ascent direction. Right: solid line initial temperature point, dotted line: optimal temperature transient estimated.

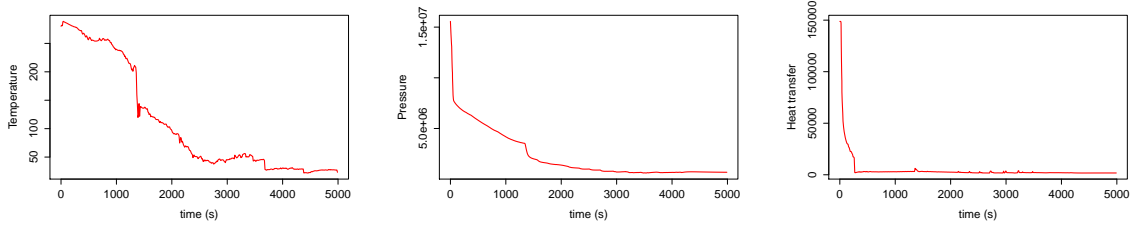


FIGURE 15. Point of the estimated steepest ascent direction maximizing the response.

**5.3. Results.** We compute an estimation of the gradient with the results of the experiments on the design points given in Figure 13. The results are given in Figure 14. We take  $\lambda_0 = 200$ . The final estimates of the optimal curves are given in Figure 15.

The main change between the starting point and the minimal point of the ascent direction lies in the temperature transient (the changes in the pressure and heat temperature transient are minimal), especially in the evolution of temperature between 500 s and 2000 s after the simulation starts and between 3000 s and 4000 s.

#### ACKNOWLEDGEMENT

The data were provided by the French Alternative Energies and Atomic Energy Commission (CEA). The author wants to thank especially Michel Marques for its patience and cooperation, which was essential to obtain the final results. I also want to thank Élodie Brunel and André Mas for their helpful advices and careful reading of this work as well as Hervé Cardot and Peggy Cénac for their precious help on stochastic optimization.

#### REFERENCES

- M. Blazère, F. Gamboa, and J.-M. Loubes. PLS: a new statistical insight through the prism of orthogonal polynomials. 05 2014. URL <http://arxiv.org/abs/1405.5900>.  
 G. E. P. Box and J. S. Hunter. Multi-factor experimental designs for exploring response surfaces. *Ann. Math. Stat.*, 28(1):195–241, 1957.

- É. Brunel, A. Mas, and A. Roche. Non-asymptotic adaptive prediction in functional linear models. *J. Multivariate Anal.*, 143:208 – 232, 2016. doi: <http://dx.doi.org/10.1016/j.jmva.2015.09.008>.
- H. Cardot and P. Sarda. Functional linear regression. In *The Oxford handbook of functional data analysis*, pages 21–46. Oxford Univ. Press, Oxford, 2011.
- H. Cardot, F. Ferraty, and P. Sarda. Functional linear model. *Stat. Probabil. Lett.*, 45(1): 11–22, Oct. 1999.
- A. Delaigle and P. Hall. Methodology and theory for partial least squares applied to functional data. *Ann. Statist.*, 40(1):322–352, 2012.
- N. R. Draper and I. Guttman. An index of rotatability. *Technometrics*, 30(1):105–111, 1988.
- N. R. Draper and D. K. J. Lin. Small response-surface designs. *Technometrics*, 32(2): 187–194, 1990.
- N. R. Draper and F. Pukelsheim. Another look at rotatability. *Technometrics*, 32(2): 195–202, 1990.
- H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- S. D. Georgiou, S. Stylianou, and M. Aggarwal. A class of composite designs for response surface methodology. *Comput. Statist. Data Anal.*, 71:1124–1133, 2014.
- R. F. Gunst and R. L. Mason. Fractional factorial design. *WIREs Comp. Stat.*, 1(2): 234–244, 2009.
- P. Hall. Principal component analysis for functional data: methodology, theory, and discussion. In *The Oxford handbook of functional data analysis*, pages 210–234. Oxford Univ. Press, Oxford, 2011.
- A. I. Khuri. A measure of rotatability for response-surface designs. *Technometrics*, 30(1): 95–104, 1988.
- A. I. Khuri. An overview of the use of generalized linear models in response surface methodology. In *Proceedings of the Third World Congress of Nonlinear Analysts, Part 3 (Catania, 2000)*, volume 47, pages 2023–2034, 2001. no. 3.
- A. I. Khuri and S. Mukhopadhyay. Response surface methodology. *Wiley Interdiscip. Rev. Comput. Stat.*, 2(2):128–149, 2010.
- S.-Y. Lee, W. Zhang, and X.-Y. Song. Estimating the covariance function with functional data. *Br. J. Math. Stat. Psych.*, 55(2):247–261, 2002.
- R. V. Lenth. Response-surface methods in R, using rsm. *J. Statist. Software*, 32(7):1–17, 2009.
- Z.-z. Liu, W. Li, and M. Yang. Two General Extension Algorithms of Latin Hypercube Sampling. *Math. Probl. Eng.*, pages Art. ID 450492, 9, 2015.
- M. D. Morris. A class of three-level experimental designs for response surface modeling. *Technometrics*, 42(2):111–121, 2000.
- H.-G. Müller and U. Stadtmüller. Generalized functional linear models. *Ann. Statist.*, 33(2):774–805, 2005.
- R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook. *Response surface methodology*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2009. Process and product optimization using designed experiments.
- S. H. Park, J. H. Lim, and Y. Baba. A measure of rotatability for second order response surface designs. *Ann. Inst. Statist. Math.*, 45(4):655–664, 1993.

- A. Pázman. *Foundations of optimum experimental design*, volume 14 of *Mathematics and its Applications (East European Series)*. D. Reidel Publishing Co., Dordrecht, 1986. Translated from the Czech.
- C. Preda and G. Saporta. PLS regression on a stochastic process. *Comput. Statist. Data Anal.*, 48(1):149–158, 2005.
- X. Qi and R. Luo. Sparse principal component analysis in Hilbert space. *Scand. J. Stat.*, 42(1):270–289, 2015.
- N. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Tucker. Surrogate-based analysis and optimization. *Prog. Aerosp. Sci.*, 41(1):1–28, 2005.
- J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2 edition, 2005.
- J. O. Ramsay and C. J. Dalzell. Some tools for functional data analysis. *J. Roy. Stat. Soc. B Met.*, pages 539–572, 1991.
- J. A. Rice and B. W. Silverman. Estimating the mean and covariance structure non-parametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, 53(1):233–243, 1991.
- T. Simpson, J. Poplinski, P. Koch, and J. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, 17(2):129–150, 2001.
- H. Wold. Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. In *Perspectives in probability and statistics (papers in honour of M. S. Bartlett on the occasion of his 65th birthday)*, pages 117–142. Applied Probability Trust, Univ. Sheffield, Sheffield, 1975.
- H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *J. Comput. Graph. Statist.*, 15(2):265–286, 2006.

MAP5 UMR CNRS 8145, UNIVERSITÉ PARIS DESCARTES AND CEREMADE UMR CNRS 7534, UNIVERSITÉ PARIS DAUPHINE.

*E-mail address:* `angelina.roche@dauphine.fr`

*URL:* `https://www.ceremade.dauphine.fr/~roche/`