

Fast Multilevel Optimization using a Multiparametric Strategy and a Cokriging Metamodel

L. Laurent, P.A. Boucard and B. Soulier

LMT-Cachan

(ENS Cachan/CNRS/Université Paris 6/PRES UniverSud Paris)

Cachan, France

Abstract

This paper deals with a multilevel model optimization strategy for structural assemblies. Two levels are introduced: the full mechanical model and a metamodel. The general objective is to reduce computation costs; here, we focus on the costs which are associated with the generation of the metamodel. Our goal is achieved through the introduction of two main elements: what we call a “multiparametric strategy” based on the LATIN method, which reduces the computation costs when the parameters vary, and the use of a cokriging metamodel taking gradients into account. Several examples illustrate the efficiency of these two elements.

Keywords: multilevel optimization, metamodel, cokriging, multiparametric strategy, LATIN method, assemblies.

1 Introduction

Structural optimization involves two very different domains of expertise: numerical simulation and optimization. Of the two, simulation is the only one which generates high computation costs. The main difficulty in the context of optimization is that the optimizer often requires a large number of simulations to locate the optimum of an objective function. (In contact and friction problems, these simulations are nonlinear and, thus, even more costly.) Therefore, new methods for the resolution of complex optimization problems are needed. This paper presents an optimization strategy based on multilevel model optimization. In order to reduce computation costs, we introduce two elements: a cokriging metamodel and a multiparametric strategy.

In the first part of this paper, the context of multilevel model optimization is reviewed. In the second part, a specific mechanical resolution process is introduced and its performance is discussed in detail. The third part concerns the kriging class of

metamodels. Finally, the last part presents the application of the coupled multiparametric-metamodel approach to a mechanical test case.

2 Multilevel model optimization

The proposed multilevel model strategy is based on the approach developed in [1]. In our case, we consider the two-level strategy illustrated in Figure 1. The first level consists in a metamodel defined using a limited amount of data. The zones where the optimum can be located are determined using, for example, a genetic algorithm. This information is transferred to the second level, where a precise search of the minimum is carried out. Therefore, the second level consists in an optimization process based on the full mechanical model. Two elements contribute to the computation cost: the construction of the metamodel from the responses of the full mechanical model, and the direct optimization based on the full mechanical model. Both elements are involved both levels of the multilevel model strategy. Therefore, the strategy proposed here leads to a reduction in the computation cost of both phases.

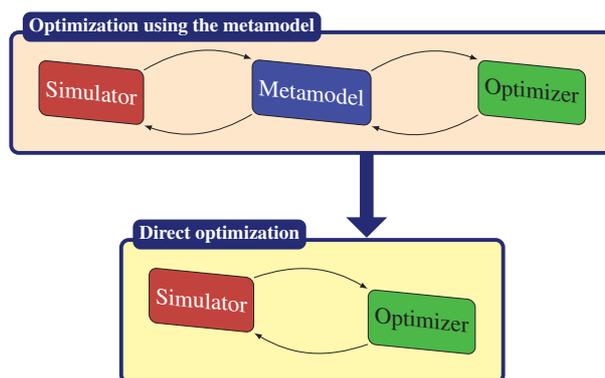


Figure 1: The multilevel optimization

The scope of this paper is limited to a study of the cost associated with the generation of the metamodel. Many types of metamodels can be used to find an approximate solution: polynomial regression [2], methods based on neural networks [3], radial basis functions [4], etc... We chose to use a particular category of metamodels called kriging approximations and, more precisely, a cokriging metamodel using derivatives. These approximations are presented in Section 4.

3 The multiparametric strategy

3.1 The context

We are considering assemblies of linear elastic structures under the assumption of small perturbations. The only nonlinearities are considered to be due to contact or

friction between parts. The strategy we use to solve this type of problem, based on twenty years of development at LMT-Cachan, relies on an iterative algorithm introduced by P. Ladevèze [5, 6].

In the context of assemblies, this strategy consists of 3 steps:

- the structure being studied is divided into *substructures* and *interfaces*;
- a dedicated iterative algorithm is used to solve the mechanical problem;
- the operators of the method remain constant and depend neither on the loading nor on the parameters (friction coefficient, gap) of the interface.

3.2 The resolution process

The resolution strategy, known as the LATIN (LArge Time INcrement) approach, consists of alternative resolutions of two groups of equations. The first group contains the *local* equations Γ (which may be nonlinear) related to the interfaces, and the second group contains the *linear* equations \mathbf{A}_d related to the substructures. The iterations between the two groups are handled through the use of search directions which are parameters of the method. This resolution process, shown in Figure 2, leads to the solution defined as the intersection of spaces Γ and \mathbf{A}_d .

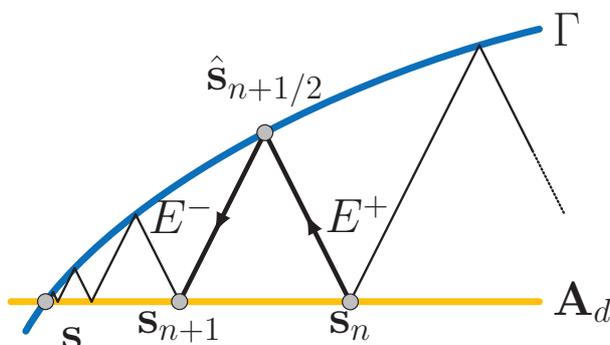


Figure 2: Schematic representation of the LATIN process

Friction and contact problems along with their specific approaches and laws [7] are part of equation group Γ . The schematic *linear* representation of \mathbf{A}_d is due to the *linear* behavior of the substructures. In the context of elastic assemblies, the boundary conditions (forces \mathbf{F}_i and displacements \mathbf{W}_i) at each interface i between two substructures are sufficient to define the solution: the internal behavior is the result of a classical elastic problem. Thus, an approximate solution is described entirely by

$$s = \bigcup_{\text{all the interfaces}} (\mathbf{W}_i, \mathbf{F}_i).$$

3.3 A method for reducing computation costs

A parametric optimization involves many calculations, each carried out with a different set of the parameters of the problem (design variables such as friction coefficients, preloads, gaps, etc...). This leads to the resolution of many similar problems in the sense that only the parameters of the problems vary. Therefore, it is essential to find a method to accelerate these calculations. The method we use, called the multiparametric strategy, was introduced in [8], then studied and applied to various sample problems in [9, 10].

3.3.1 Principle of the method

The main idea of the multiparametric strategy is to take advantage of a feature of the LATIN algorithm: at each iteration, the solver yields an approximate solution over the whole loading path and at all points in the structure. Then, if a new calculation associated with other values of the parameters is requested, the algorithm can be reinitialized using a previously converged solution. Figure 3 illustrates the use of the multiparametric strategy to obtain an approximate solution starting from a previous calculation. Here, we assume that the only parameters which are different are the interface parameters, so space Γ alone is affected by the change.

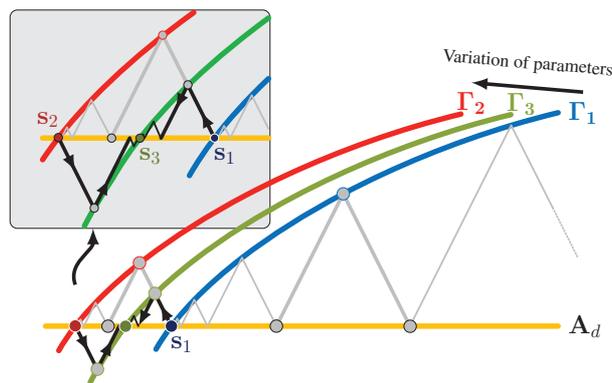


Figure 3: The multiparametric strategy using the LATIN algorithm

With this approach, the solution converges in fewer iterations of the algorithm and, therefore, in less time.

3.3.2 Illustration of the performance of the method

The example considered here is a quasi-static academic problem which was presented in [10]. Figure 4 shows the geometry of the problem, which consists of three square parts ($h = 50mm$, Young's modulus $E = 2 \cdot 10^5 MPa$ and Poisson's coefficient $\nu = 0.3$) which are in contact with friction. Each part was represented by a single substructure discretized into 24×24 bilinear quadrangles. The parametric study consisted in varying the friction coefficients μ_1 and μ_2 of the two contact interfaces. The

loading consisted of two stages: first, a progressive vertical pressure P_1 up to a maximum of 50MPa applied at the top of substructure Ω_3 (the preloading stage); then, a progressive horizontal load from 0 to 30MPa applied to substructure Ω_2 .

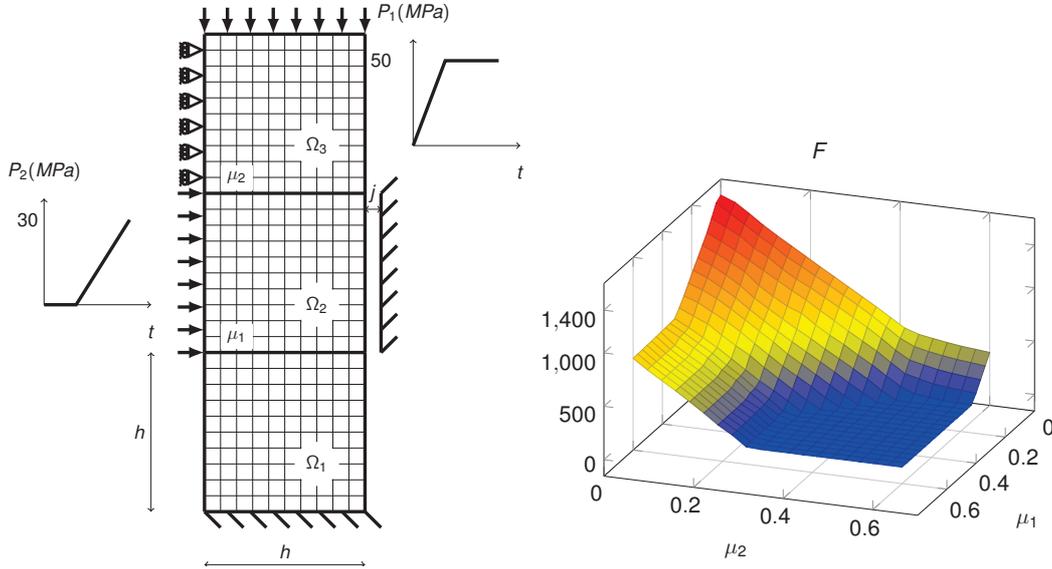


Figure 4: The geometry of the problem and the response surface

In this test case, variations of the friction coefficients μ_1 and μ_2 between 0 and 0.6 were considered, and the function studied was the reaction force on the rigid wall. Figure 4 shows the response surface of this function obtained with 18×18 values of the friction coefficients.

In order to illustrate the performance of the multiparametric strategy, the two-variable design space was sampled with a 4×4 regular grid. For each sample, the force and gradients were calculated (using a classical finite difference method for the gradients). Table 1 summarizes the characteristics of these calculations:

The gain obtained with our method (compared to a classical calculation without the multiparametric strategy) was estimated using the following expression:

$$\text{Gain} = \frac{\text{Number of calculations} \times \text{CPU time of the first calculation}}{\text{CPU time using the multiparametric strategy}} \quad (1)$$

We considered that the time of each calculation without the multiparametric strategy was almost constant and equal to the time of the first calculation ($8.2s$).

The results presented in Table 1 show that the solver can reuse a previously converged solution to accelerate the resolution of similar problems. The most remarkable point is the gain obtained in the evaluation of the gradients. These were calculated using a finite difference method: each gradient required the value of the response at 3 points (a sample point plus 2 points obtained by very small variations of the parameters). The results show that for the same number of evaluations the cost was less when the gradients were calculated along with the responses than when the responses alone

	Eval.	Eval. gradients	Total
Number of evaluations	16	2×16	48
Total CPU time	42.3s	58.1s	100.4s
Avg CPU time per calc.	2.7s	1.8s	2.1s
Min CPU time	0.2s	0.2	0.2
Max CPU time	8.2s	6s	8.2s
Nb. of iterations	1,896	2,580	4,476
Avg nb. of iterations (per calc.)	118.5	80.6	93.3
Min nb. of iterations	5	5	5
Max nb. of iterations	381	260	381
Gain	3.11	4.54	3.94

Table 1: The mechanical calculations using the multiparametric strategy

were calculated: in our example, 16 evaluations took 42.3s whereas 16 gradients (32 evaluations) took 58s. Thus, the strategy is particularly efficient in calculating the gradients, thanks to the ability of the method to reduce the computation time for sample points which are close together. This led to our decision which was to use a dedicated gradients metamodel.

4 The cokriging metamodel

This type of metamodel is similar to the kriging metamodel proposed by [11] and developed by [12, 13]. The cokriging strategy stems from multivariate geostatistics [14]. In our case, the cokriging metamodel is built using not only the responses, but also their gradients.

4.1 Notations

In this section, we will use the following notations: $\mathbf{x}^{(i)}$ denotes a point in the design space \mathcal{D} ($\mathbf{x}^{(i)}$ is one of the n_s sample points, while $\mathbf{x}^{(0)}$ is a non-sample point); x_i denotes the i^{th} coordinate of a point \mathbf{x} ; $Y(\mathbf{x})$ and $\tilde{Y}(\mathbf{x})$ denote respectively the response of the analytical function (or the response of the mechanical model) and the approximate response given by the metamodel; finally, $R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ is a correlation function.

4.2 Principle

The model can be viewed as the sum of two components: a linear model (which represents the trend of the data) and a departure from the linear model (which represents the fluctuations around the trend).

$$\tilde{Y}(\mathbf{x}^{(0)}) = \mu(\mathbf{x}^{(0)}) + Z(\mathbf{x}^{(0)}), \quad \forall \mathbf{x}^{(0)} \in \mathcal{D} \quad (2)$$

$$\mathbb{E}(Z) = 0, \text{Cov}[Z(\mathbf{x}^{(i)}), Z(\mathbf{x}^{(j)})] \neq 0, \forall (\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \in \mathcal{D}^2 \quad (3)$$

where \mathbb{E} and Cov denote the classical statistical expected value and covariance.

The main idea is to consider a covariance relation between the discrete responses of the deterministic function. This covariance depends only on the distances among the samples. Thus the departure is represented by a zero-mean, second-order stationary process. (The mean and the variance are constant, with a correlation depending on the distance.) Depending on the definition of the function μ , one can build different types of kriging or cokriging metamodels (simple kriging, where μ is the average of the response at sample design points; ordinary kriging, where μ is an unknown constant; or universal kriging, where μ is a polynomial function).

Function Z has a zero expected value and its covariance structure is a function of a generalized distance among the sample responses. The covariance structure can be written as:

$$\forall (\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \in \mathcal{D}^2, \text{cov}[Z(\mathbf{x}^{(i)}), Z(\mathbf{x}^{(j)})] = \sigma^2 R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sigma^2 c_{ij} \quad (4)$$

In our case, the correlation function chosen is a Gaussian function or a Matérn function[15, 16].

In the case of a cokriging metamodel, additional covariance relations involving the different variables are introduced [17]. In our case, the primary variables are the evaluations of the function being studied, and the secondary variables are the gradients:

$$\text{Cov}\left[\frac{\partial Z}{\partial x_k}(\mathbf{x}^{(i)}), Z(\mathbf{x}^{(j)})\right] = -\sigma^2 \frac{\partial R}{\partial x_k}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sigma^2 c_{kij} \quad (5)$$

$$\text{Cov}\left[Z(\mathbf{x}^{(i)}), \frac{\partial Z}{\partial x_k}(\mathbf{x}^{(j)})\right] = -\sigma^2 \frac{\partial R}{\partial x_k}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sigma^2 c_{kij} \quad (6)$$

$$\text{Cov}\left[\frac{\partial Z}{\partial x_k}(\mathbf{x}^{(i)}), \frac{\partial Z}{\partial x_l}(\mathbf{x}^{(j)})\right] = -\sigma^2 \frac{\partial^2 R}{\partial x_k \partial x_l}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sigma^2 c_{kilj} \quad (7)$$

With this relation, one can also build the kriging or cokriging metamodel by considering the following linear predictors of the non-sample point $\mathbf{x}^{(0)}$:

Kriging	$\tilde{Y}(\mathbf{x}^{(0)}) = \sum_{i=1}^{n_s} \lambda_i(\mathbf{x}^{(0)}) Y(\mathbf{x}^{(i)})$
Cokriging	$\tilde{Y}(\mathbf{x}^{(0)}) = \sum_{i=1}^{n_s} \lambda_{0i}(\mathbf{x}^{(0)}) Y(\mathbf{x}^{(i)}) + \sum_{j=1}^{n_d} \sum_{i=1}^{n_s} \lambda_{ij}(\mathbf{x}^{(0)}) \frac{\partial Y}{\partial x_j}(\mathbf{x}^{(i)})$

Thus, the determination of the Best Linear Unbiased Predictor (BLUP) leads to the λ 's which, for a kriging metamodel, minimize the *Mean Square Error*:

$$\text{MSE}[\tilde{Y}(\mathbf{x}^{(0)})] = \mathbb{E}\left[\left(\sum_{i=1}^{n_s} \lambda_i(\mathbf{x}^{(0)}) Y_i - Y(\mathbf{x}^{(0)})\right)^2\right] \quad (8)$$

subject to the unbiasedness condition:

$$\mathbb{E} \left[\sum_{i=1}^{n_s} \lambda_i(\mathbf{x}^{(0)}) Y(\mathbf{x}^{(i)}) \right] = \mathbb{E} [Y(\mathbf{x}^{(0)})] \quad (9)$$

The construction of a cokriging metamodel follows the same process.

In the universal kriging case, the best linear predictor of $Y(\mathbf{x}^{(0)})$ can be written in matrix form [18] as:

$$\tilde{Y}(\mathbf{x}^{(0)}) = \mathbf{x}_0^T \hat{\boldsymbol{\beta}} + \mathbf{c}_0^T \mathbf{C}^{-1} (\mathbf{Y}_s - \mathbf{X} \hat{\boldsymbol{\beta}}) \quad (10)$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{C}^{-1} \mathbf{Y}_s \\ \mathbf{x}_0 &= [f_1(\mathbf{x}^{(0)}) \quad f_2(\mathbf{x}^{(0)}) \quad \cdots \quad f_k(\mathbf{x}^{(0)})] \\ \mathbf{c}_0 &= [c_{01} \quad c_{02} \quad \cdots \quad c_{0n_s}] \\ (\mathbf{X})_{ij} &= f_j(\mathbf{x}^{(j)}) \\ (\mathbf{C})_{ij} &= c_{ij} \end{aligned}$$

The regressor vector $\hat{\boldsymbol{\beta}}$ is the generalized least-squares approximation of $\boldsymbol{\beta}$. Thus, the first part of Equation 10 is the generalized least-squares prediction at point $\mathbf{x}^{(0)}$. The second part can be viewed as a correction of the generalized least-squares response surface to obtain the interpolating kriging model.

The formulation of the cokriging metamodel is very similar to that of the kriging metamodel. For simplicity's sake, following the idea in [19], we propose to use only an ordinary cokriging metamodel:

$$\tilde{Y}(\mathbf{x}^{(0)}) = \hat{\boldsymbol{\beta}}_c + \mathbf{c}_{c0}^T \mathbf{C}_c^{-1} (\mathbf{Y}_{sc} - \mathbf{X}_c \hat{\boldsymbol{\beta}}_c) \quad (11)$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}}_c &= (\mathbf{X}_c^T \mathbf{C}_c^{-1} \mathbf{X}_c)^{-1} \mathbf{X}_c^T \mathbf{C}_c^{-1} \mathbf{Y}_{sc} \\ \mathbf{c}_{c0} &= [c_{01} \quad c_{02} \quad \cdots \quad c_{0n_s} \quad c_{110} \quad c_{210} \quad \cdots \quad c_{n_d n_s 0}] \\ \mathbf{X}_c &= [1, 1, \cdots, 1, 0, 0, \cdots, 0]^T \\ \mathbf{C}_c &= \begin{bmatrix} \mathbf{C} & \mathbf{C}_d \\ \mathbf{C}_d^T & \mathbf{C}_{dd} \end{bmatrix} \end{aligned}$$

and where

$$\begin{aligned}
(\mathbf{C})_{ij} &= c_{ij} \\
(\mathbf{C}_d)_{ij} &= \begin{bmatrix} c_{111} & c_{211} & \cdots & c_{n_d 11} & c_{121} & \cdots & c_{n_d n_s} \\ c_{112} & c_{212} & \cdots & c_{n_d 12} & c_{122} & \cdots & c_{n_d n_s 2} \\ c_{113} & c_{213} & \cdots & & & & \\ \vdots & & \ddots & & & & \vdots \\ c_{11 n_s} & \cdots & & & & & c_{n_d n_s n_s} \end{bmatrix} \\
(\mathbf{C}_{dd})_{ij} &= \begin{bmatrix} c_{1111} & c_{1121} & \cdots & c_{11 n_d 1} & c_{1112} & \cdots & c_{11 n_d n_s} \\ c_{2111} & c_{2121} & \cdots & & & & \vdots \\ \vdots & & \ddots & & & & \\ & & & c_{n_d 1 n_d 1} & c_{n_d 112} & \cdots & \\ & & & & \ddots & & \\ c_{n_d n_s 11} & \cdots & & & & & c_{n_d n_s n_d n_s} \end{bmatrix}
\end{aligned}$$

In the kriging case, vector \mathbf{Y}_s contains only the responses of the function at the sample point, whereas in the cokriging case vector \mathbf{Y}_{sc} contains both the responses and the gradients. With these formulations, the models can supply approximate responses of the actual function at all the points in the design space. In our case, we consider the response of the function to be deterministic and we obtain both of the interpolating models. This type of metamodel has other advantages: for example, it provides statistical information (the expected value and the variance of the process). Due to the use of the unbiasedness condition, the expected value of \tilde{Y} is given by the trend model μ and the mean square error of \tilde{Y} [18]:

$$\text{MSE} \left[\tilde{Y}(\mathbf{x}^{(0)}) \right] = \sigma^2 \left[1 - [\mathbf{x}_0 \quad \mathbf{c}_0] \begin{bmatrix} \mathbf{0} & \mathbf{X}^T \\ \mathbf{X} & \mathbf{C} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{c}_0 \end{bmatrix} \right] \quad (12)$$

This relation also holds in the cases of simple or ordinary kriging or cokriging models, provided the appropriate forms of vectors and matrices \mathbf{x}_0 , \mathbf{c}_0 , \mathbf{X} and \mathbf{C} are used.

4.3 Estimation of the parameters

The model's parameters (such as the characteristic correlation length scale l , the variance σ of the random process Z or the regression coefficients) can be determined by maximizing the likelihood [20]. In our case, we use this technique to determine l and σ :

$$(l, \sigma^2) = \arg \min_{l, \sigma^2} \left[-\frac{n_s}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log(|\mathbf{C}(l)|) - \frac{1}{2} (\mathbf{Y}_s - \mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{C}(l)^{-1} (\mathbf{Y}_s - \mathbf{X}\hat{\boldsymbol{\beta}}) \right] \quad (13)$$

The variance σ^2 can be determined analytically through the derivation of the likelihood function:

$$\hat{\sigma}^2 = \frac{1}{n_s} (\mathbf{Y}_s - \mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{C}(l)^{-1} (\mathbf{Y}_s - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (14)$$

One can also use an optimizer to determine the value of the correlation length scale. This method has some drawbacks [21, 22, 23]: in many cases with very few points, the log-likelihood is monotonous; often, the correlation matrix is affected by conditioning problems which make finding the minimum difficult. When such problems arise, one sets the parameters (in particular the correlation length scale) to fixed values.

4.4 Analytical applications

In this section, we apply kriging and cokriging to one- and two-dimensional analytical test functions. We will use the abbreviations OK and OCK to designate respectively **Ordinary Kriging** and **Ordinary CoKriging**.

4.4.1 The one-dimensional test function

First, we applied the two types of metamodels to an analytical function, chosen to be $y(x) = \exp(-x/10) \cos(x) + x/10$. We used 5 sample responses of the analytical function to build the OK metamodel, and an additional 5 sample derivatives to build the OCK metamodel. The correlation function was the Matérn function. The sample points were obtained using *Latin Hypercube Sampling*.

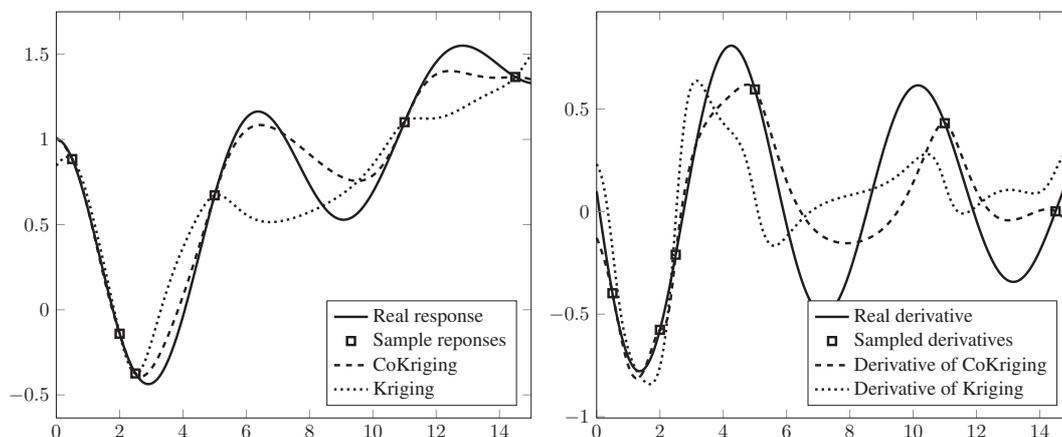


Figure 5: The ordinary kriging and cokriging metamodels and their derivatives

Figure 5 illustrates the capability of the cokriging metamodel to interpolate not only the values of the responses, but also the sample derivatives. In this example, the quality of the cokriging metamodel was better. This statement remained true as long as we worked with only a few points. We can also observe that for a relatively smooth function the kriging metamodel converges quickly toward a good approximation of the real function when the number of points becomes large enough.

The characteristics of the two metamodels are summarized in the following table:

	R^2	RAAE	RMAE	Q_1	Q_2	Q_3
OK	0.4442	0.5036	1.726	0.4192	23.46	$7.796 \cdot 10^{-2}$
Ock	0.9719	0.1153	0.4087	$2.351 \cdot 10^{-2}$	1.186	$3.940 \cdot 10^{-3}$

Table 2: Characteristics of the two previous metamodels

Criteria RAAE and RMAE stand for *Relative Average Absolute Error* and *Relative Maximum Absolute Error*. Criteria Q_i , which compare the actual response and the responses of the metamodels at n_c points with $n_c \gg n_s$, were calculated as follows:

$$\begin{aligned}
 Q_1 &= \sup_{i \in \{1, 2, \dots, n_c\}} e_i \\
 Q_2 &= \sum_{i=1}^{n_c} e_i \\
 Q_3 &= \frac{Q_2}{n_c}
 \end{aligned}$$

where $\forall i \in \{1, 2, \dots, n_c\}$, $e_i = \frac{(Y(\mathbf{x}^{(i)}) - \tilde{Y}(\mathbf{x}^{(i)}))^2}{\sup_{j \in \{1, 2, \dots, n_c\}} Y(\mathbf{x}^{(j)})^2}$

Based on the statistical information from the kriging and cokriging metamodels, one can derive confidence intervals. The two diagrams of Figure 6 show the 95% Confidence Intervals (CIs) obtained with Expression 15. In these types of metamodels, the size of the confidence envelopes is determined mainly by the distance between each pair of neighboring points.

$$CI^\pm(\mathbf{x}) = \tilde{Y}(\mathbf{x}) \pm 2\sqrt{MSE[\tilde{Y}(\mathbf{x})]} \quad (15)$$

4.4.2 The two-dimensional test function

The same two types of metamodels were used to approximate an analytical function of two variables. In order to illustrate the performance of the cokriging metamodel, we chose a very bumpy function: the *six-hump camel back function* ($\forall (x_1, x_2) \in [-3, 3] \times [-2, 2]$, $f(x_1, x_2) = (4 - 2.1x_1^2 + x_1^4)x_1^2 + x_1x_2 + 4(x_2^2 - 1)x_2^2$). The two metamodels were constructed using 16 evaluations of the function for the kriging metamodel, and an additional 16 evaluations of the gradients for the cokriging metamodel. In both cases the correlation function was the Matérn function.

The characteristics of these two metamodels are given in Table 3.

For this 2D test function, the cokriging metamodel led to a relatively accurate approximation of the actual function using only a few sample points. Taking into account

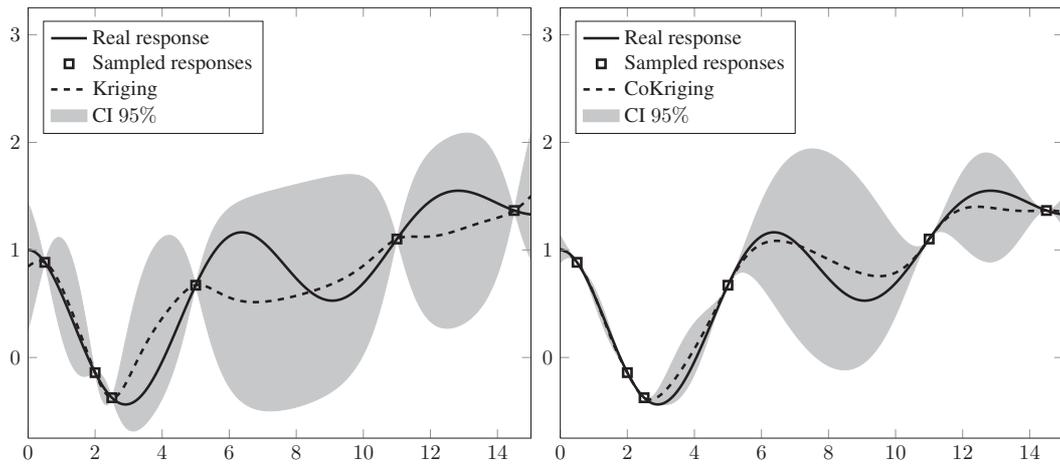


Figure 6: The confidence envelopes for the kriging and cokriging metamodels

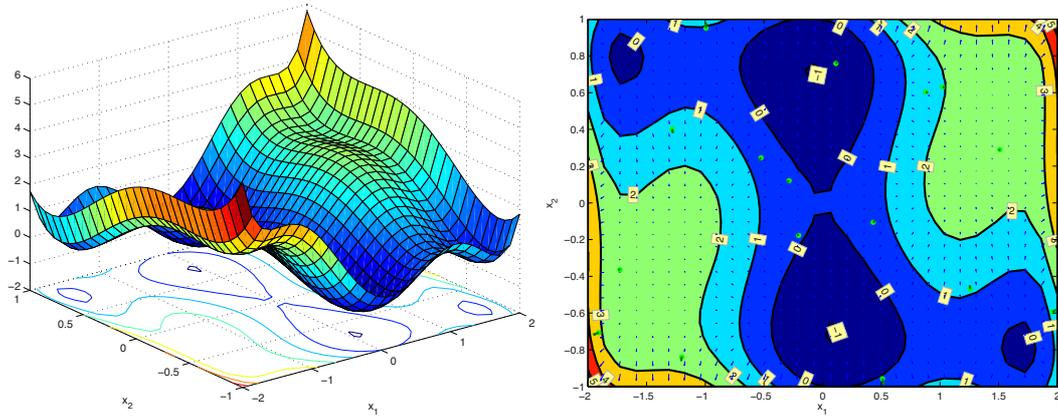


Figure 7: The actual function and its gradients (blue arrows)

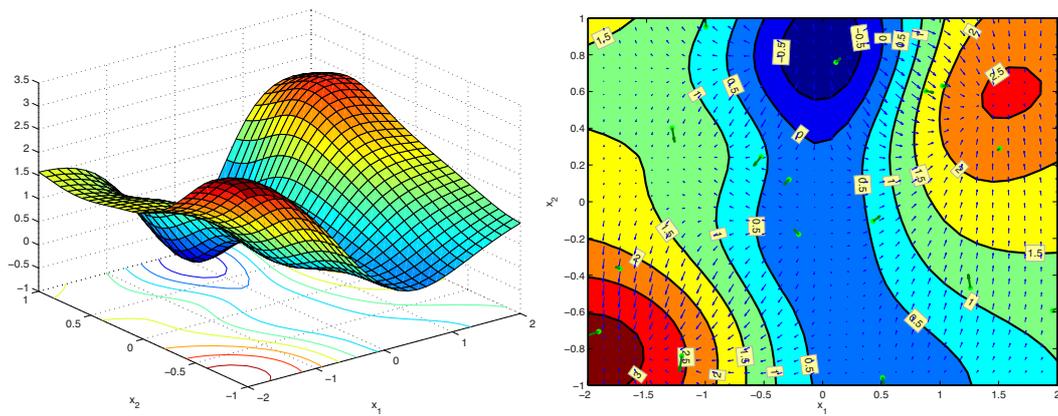


Figure 8: The kriging function and its gradients (blue arrows)

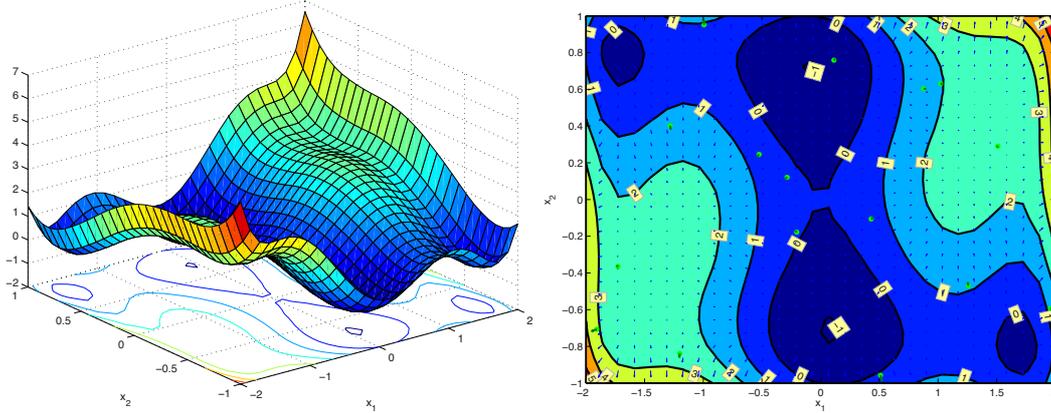


Figure 9: The cokriging function and its gradients (blue arrows)

	R^2	RAAE	RMAE	Q_1	Q_2	Q_3
OK	0.6991	0.4042	2.845	0.3946	13.192	$1.466 \cdot 10^{-2}$
OCK	0.9958	$3.027 \cdot 10^{-2}$	$5.330 \cdot 10^{-1}$	$1.391 \cdot 10^{-2}$	0.1824	$2.027 \cdot 10^{-4}$

Table 3: Characteristics of the two previous metamodels (Figure 8-9)

the gradients, we were able to develop more efficient approximate models. But for a problem involving the evaluation of mechanical responses the computation cost for the determination of n_s responses and n_s gradients is, of course, higher than that required to obtain n_s responses alone. (In the former case, due to the use of finite differences to obtain gradients, the construction of the metamodel requires $3n_s$ evaluations of the mechanical model.)

Now let us take another approach: the idea is to build metamodels using the same number of evaluations. In the following case, we used the same six-hump function, but the kriging metamodel was constructed based on the responses at 27 sample points (Figure 10), while the cokriging metamodel was still constructed using the responses and the gradients at 9 sample points (Figure 11).

	R^2	RAAE	RMAE	Q_1	Q_2	Q_3
OK	0.8175	0.2566	2.933	0.4195	8.001	$8.890 \cdot 10^{-3}$
OCK	0.8160	0.2580	2.541	0.3150	8.066	$8.9627 \cdot 10^{-3}$

Table 4: Characteristics of the two previous metamodels (Figure 10-11)

Table 4 shows that the quality of the prediction achieved with the two metamodels was similar, but a very important advantage of the cokriging metamodel was that it led to more zones where a minimum could be found than the kriging metamodel. If the problem involved the calculation of mechanical responses, the computation cost associated with the 9 points and their gradients, thanks to the multiparametric strategy and as shown in Table 1, would be much lower than that associated with the 27 points.

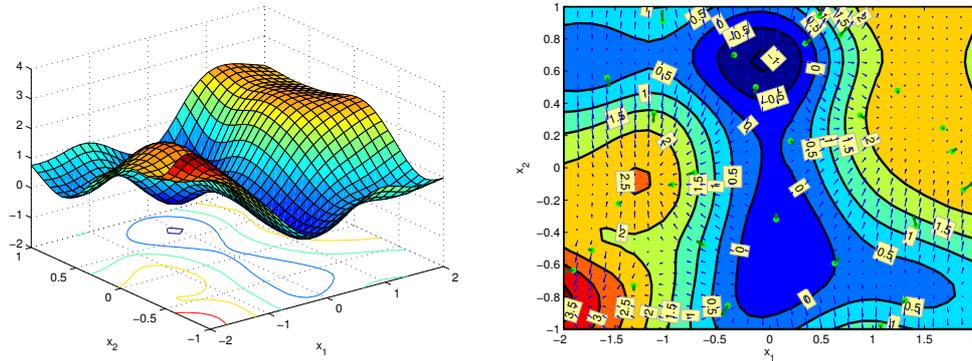


Figure 10: The kriging surface obtained with 27 responses (27 evaluations of the function)

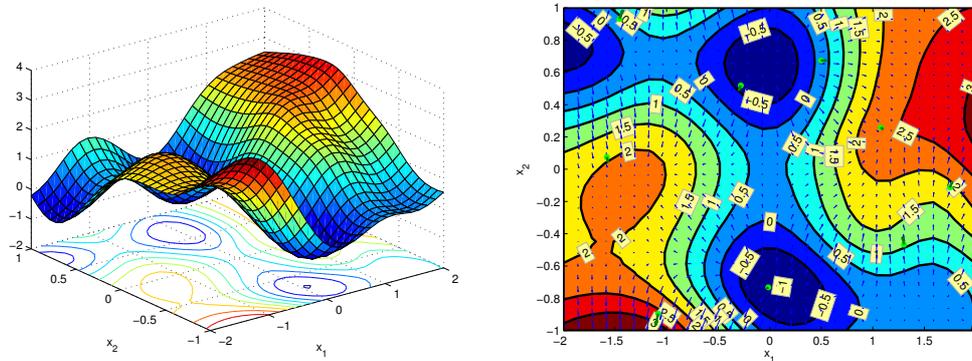


Figure 11: The cokriging surface obtained with 9 responses and 9 gradients (27 evaluations of the function)

5 Application to a mechanical problem

In this section, we present the construction of the two previous metamodels (kriging and cokriging) using the multiparametric strategy presented in the first part of this paper and the actual responses of a mechanical model. The test case being considered is the *three-squares example* presented in Page 4. We carried out two studies: one with a fixed number of mechanical calls, and the other with a fixed quality of the metamodels.

5.1 Case of a fixed number of mechanical calls

Both metamodels were constructed using 15 mechanical evaluations. The kriging metamodel was determined using 15 values of the force, and the cokriging metamodel was determined with 5 values of the force and 5 gradients. The sample points were obtained through *Latin Hypercube Sampling*. Our calculations led to the following

two response surfaces (Figures 12 and 13) to be compared with the actual response surface of Figure 4:

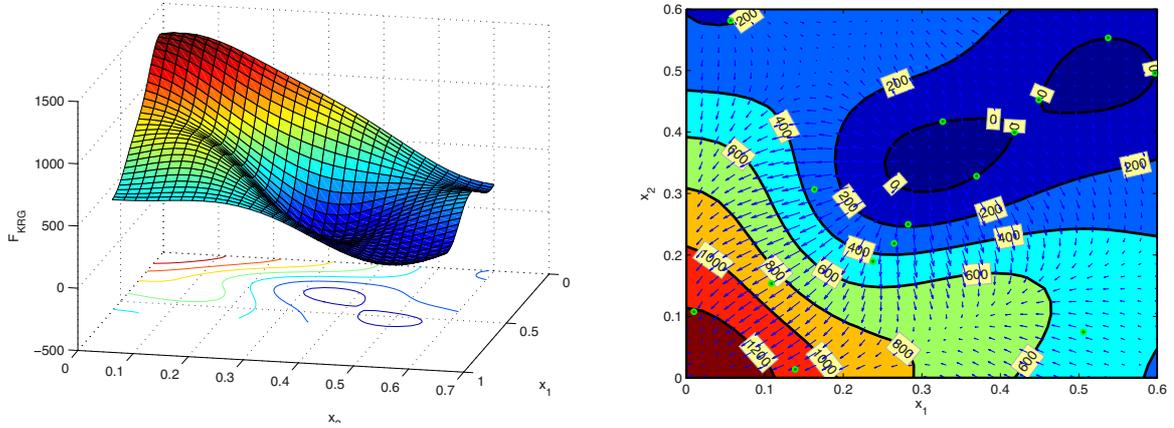


Figure 12: The kriging response surface of the *three-squares test case* (using 15 responses)

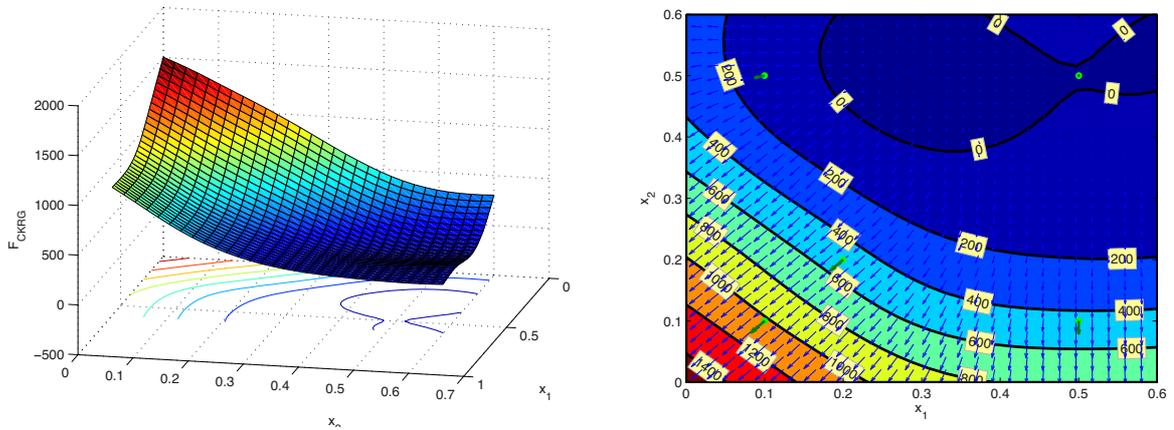


Figure 13: The cokriging response surface of the *three-squares test case* (using 5 responses and 5 gradients)

Tables 5 and 6 show the characteristics of the calculations and of the metamodels.

In this case, the cokriging metamodel led to a much better approximation of the mechanical model than the kriging metamodel. Moreover, this result was obtained at a lower computation cost.

5.2 Case of a fixed metamodel quality

Another way to study the cost of the metamodels consists in constructing kriging and cokriging metamodels with the same quality. For this study, we constructed the

Gradients	without	with
Total CPU Time	44.3s	36.4s
Avg CPU Time (/calc)	2.95s	2.13s
Min CPU Time	0.6s	0.2s
Max CPU Time	8.8s	8.2s
LATIN iterations	2,031	1,611
Avg iterations (/calc)	135	107
Min iterations	25	5
Max iterations	411	391
Gain	2.97	3.40

Table 5: Characteristics of the mechanical calculations

	R^2	RAAE	RMAE	Q_1	Q_2	Q_3
OK	0.9865	$8.751 \cdot 10^{-2}$	0.6624	$2.746 \cdot 10^{-2}$	0.3385	$8.463 \cdot 10^{-4}$
OCK	0.9822	0.1015	0.4024	$1.014 \cdot 10^{-2}$	0.4441	$1.110 \cdot 10^{-3}$

Table 6: Characteristics of the metamodels (OK: ordinary kriging/OCK: ordinary cokriging)

metamodels using sample points obtained through full factorial sampling. We chose to build the kriging metamodel with 9×9 samples and the cokriging metamodel with 3×3 and 5×5 samples. The results are shown in Table 7:

	OK 9×9	OCK 3×3	OCK 5×5
Total CPU Time	154.7s	61.6s	127s
Avg CPU Time (/calc)	1.9s	2.3s	1.7s
Min CPU Time	0.2s	0.2s	0.2s
Max CPU Time	8.9s	8.1s	9.1s
LATIN iterations	6,576	2,731	5,436
Avg iterations (/calc)	81.2	101	72.5
Min iterations	5	5	5
Max iterations	381	381	381
Gain	4.65	3.53	5.39
R^2	0.9878	0.9702	0.9942
RAAE	$2.511 \cdot 10^{-2}$	0.1234	$5.147 \cdot 10^{-2}$
RMAE	0.1442	0.5703	0.3243
Q_1	$1.301 \cdot 10^{-3}$	$2.036 \cdot 10^{-2}$	$6.585 \cdot 10^{-3}$
Q_2	$3.044 \cdot 10^{-2}$	0.7439	0.1450
Q_3	$7.610 \cdot 10^{-5}$	$1.860 \cdot 10^{-3}$	$3.626 \cdot 10^{-4}$

Table 7: The metamodels constructed using full factorial sampling

5×5 full factorial sampling led to a cokriging metamodel with the same quality

as the 9×9 kriging metamodel. As in the previous case with a fixed number of mechanical calls, the cost with the cokriging metamodel was lower. In the case of more complex problems and more parameters, one can expect even better gains.

6 Conclusion

In this paper, we associated a cokriging metamodel with a dedicated multiparametric strategy. These two elements enabled us to achieve a significant reduction in the computation cost of constructing the metamodel.

Our multiparametric strategy takes advantage of a property of the LATIN method which is that the result of a previous calculation constitutes an efficient starting point for a subsequent calculation. This property is particularly advantageous for the evaluation of gradients, which are obtained using finite differences (*i.e.* with very small variations of the parameters). Thus, our strategy lends itself naturally to the use of a gradient-based metamodel such as the cokriging model presented in the paper.

Moreover, since this metamodel is part of a multilevel model optimization strategy, besides reducing the computation cost associated with the construction of the cokriging metamodel, one can also reduce the cost of the second level, in which algorithms based on gradient descent are used. This will be the subject of an upcoming work.

Acknowledgment

This work was supported by the *Agence Nationale de la Recherche* as part as ANR-08-COSI-007-10: Distributed Multi-Disciplinary Optimization.

References

- [1] G. Robinson, A. Keane, “A case for multi-level optimisation in aeronautical design”, *Aeronautical Journal*, 103(1028): 481–485, 1999.
- [2] G. Box, K. Wilson, “On the Experimental Attainment of Optimum Conditions”, *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(1): 1–45, 1951.
- [3] W. McCulloch, W. Pitts, “A logical calculus of the ideas immanent in nervous activity”, *Bulletin of Mathematical Biology*, 5(4): 115–133, Dec. 1943.
- [4] R. Hardy, “Multiquadric Equations of Topography and Other Irregular Surfaces”, *J. Geophys. Res.*, 76: 1905–1915, 1971.
- [5] P. Ladevèze, “New algorithms: mechanical framework and development” (in French), *Compte rendu de l’académie des Sciences*, 300(2): 41–44, 1985.
- [6] P. Ladevèze, “*Nonlinear computational structural mechanics: new approaches and non-incremental methods of calculation*”, Springer Verlag, 1999.

- [7] C. Blanzé, L. Champaney, J. Cognard, P. Ladevèze, “A modular approach to structure assembly computations: application to contact problems”, *Engineering Computations*, 13(1): 15–32, 1995.
- [8] P. Boucard, P. Ladeveze, “A multiple solution method for non-linear structural mechanics”, *Mechanical Engineering*, 50(5): 317–328, 1999.
- [9] O. Allix, P. Vidal, “A new multi-solution approach suitable for structural identification problems”, *Computer Methods in Applied Mechanics and Engineering*, 191(25-26): 2727–2758, 2002.
- [10] P. Boucard, L. Champaney, “A suitable computational strategy for the parametric analysis of problems with multiple contact”, *International Journal for Numerical Methods in Engineering*, 57(9): 1259–1281, 2003.
- [11] D. Krige, “A statistical approach to some mine valuation and allied problems on the Witwatersrand”, Master’s thesis, 1951.
- [12] G. Matheron, “Treaty of applied geostatistic” (in French), *Mémoires du Bureau de Recherches Géologiques et Minières*, 14, 1962.
- [13] G. Matheron, “Principles of geostatistics”, *Economic geology*, 58(8): 1246, 1963.
- [14] H. Wackernagel, “*Multivariate geostatistics: an introduction with applications*”, Springer Verlag, 1995.
- [15] B. Matérn, “Spatial Variation (Lecture Notes Statist. 36)”, Springer, Berlin, 1960.
- [16] M. Stein, “*Interpolation of Spatial Data: some theory for kriging*”, Springer Verlag, 1999.
- [17] M. Morris, T. Mitchell, D. Ylvisaker, “Bayesian design and analysis of computer experiments: use of derivatives in surface prediction”, *Technometrics*, 35(3): 243–255, 1993.
- [18] J. Sacks, W. Welch, T. Mitchell, H. Wynn, “Design and analysis of computer experiments”, *Statistical science*, 4(4): 409–423, 1989.
- [19] H. Chung, J. Alonso, “Using gradients to construct cokriging approximation models for high-dimensional design optimization problems”, in 40th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada. Citeseer, 2002.
- [20] K. Mardia, R. Marshall, “Maximum likelihood estimation of models for residual covariance in spatial regression”, *Biometrika*, 71(1): 135, 1984.
- [21] J. Warnes, B. Ripley, “Problems with likelihood estimation of covariance functions of spatial Gaussian processes”, *Biometrika*, 74(3): 640, 1987.
- [22] K. Mardia, A. Watkins, “On multimodality of the likelihood in the spatial linear model”, *Biometrika*, 76(2): 289, 1989.
- [23] M. Sasena, “Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations”, PhD thesis, Citeseer, 2002.