Surrogate models for optimization in high dimension using a mixed Kriging/PLS method

M. Bouhlel, N. Bartoli, J. Morlier, A. Otsmane

Surrogate models for optimization in high dimension using a mixed Kriging/PLS method

Bouhlel M. A.[a*] and Bartoli N.[a] and Morlier J.[b] and Otsmane A.[c]

[a]ONERA, Toulouse, France
[b]ISAE, Toulouse, France
[c]SNECMA, Paris, France

Introduction

Kriging [1] is a popular metamodel used to replace a computationally expensive simulation model. However, with the development of modern scientific codes in the field of engineering, the number of variables increases in order to capture the physical behaviour and it is well known that Kriging performances are degraded in this condition. In order to overcome this problem, several methods for reducing the dimension are available, particularly the Partial Least Squares method (PLS) [2]. PLS creates new variables (latent variables or principal components) by modelling the relationship between input and output variables while maintaining most of information in the input variables.

In this work, a new method is developed by combining Kriging and PLS. This method consists in using information from the PLS in order to build an efficient Kriging metamodel adapted to high dimension. Numerical experiments on both academic and industrial test cases show the efficiency of such method in terms of accuracy and CPU time up to dimension 60.

1   Overview of Kriging and PLS

From now, d is the dimension of the problem, n is the number of observations (design of experiments), h is the number of principal components to retain, $\mathbf{x}^{(i)}$ $(i = 1, \cdots, n)$ is a point from the design of experiment and $\mathbf{y}$ is a vector containing true responses associated.

1.1   Kriging

A Kriging model is a generalized linear regression model since it accounts for the correlation in the residuals between the regression model and the observations. Correlation $\mathbf{r_{u,v}}$ between two points $\mathbf{u}$ and $\mathbf{v}$ is given by

$$\mathbf{r_{u,v}} = \prod_{j=1}^{d} \exp\left(-\theta_j |\mathbf{u}_j - \mathbf{v}_j|^2\right), \tag{1}$$

where θ_j $(j = 1, \cdots, d)$ is estimated using maximum likelihood. The number of parameters θ to determine equals the problem dimension d. Thus, this estimation may be a difficult task if the number of variables is large.

The mathematical form of Kriging prediction at a point $\mathbf{x}$ is given by Equations (2) and (3)

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r_x}^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{2}$$

---

*Corresponding author. E-mail: ONERA, France. mohamed.bouhlel@onera.fr.

with

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}}, \ \mathbf{R}_{ij} = \mathbf{r}_{\mathbf{x}^{(i)} \mathbf{x}^{(j)}}, \ i, j = 1, \cdots, n, \ \mathbf{r}_{\mathbf{x}} = (\mathbf{r}_{\mathbf{x}\mathbf{x}^{(1)}}, \cdots, \mathbf{r}_{\mathbf{x}\mathbf{x}^{(n)}})^T \text{ and } \mathbf{1} = (1, \cdots, 1)^T. \quad (3)$$

1.2 Partial Least Squares

PLS is a statistical method that finds a linear relation between input variables and output variable (one consider only one output in this study) by projecting input variables to a new space, formed by new variables called principal components $\mathbf{T}$. PLS will try to find the best multidimensional direction in the $\mathbf{X} = (\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(n)})$ space that explains the behaviour of the output $\mathbf{y}$. Thus, principal components are given by $\mathbf{T} = \mathbf{XW}$ with $\mathbf{W}$ the matrix containing in each column the loading vector for the construction of the principal component. PLS is suited when $\mathbf{X}$ has many variables (high dimension) to explain.

2 Kriging combined with PLS

This paper presents a new method -Kriging combined PLS- for computer simulation. This method builds an efficient and a fast Kriging. It uses informations (contained in $\mathbf{W}$) extracted from PLS. The key point consists in substituting the spatial correlation function from the Eq. (1) to the new expression

$$\mathbf{r}_{\mathbf{u},\mathbf{v}} = \prod_{j=1}^{h} \exp\left(-\theta_j \sum_{i=1}^{d} |\mathbf{W}_{i,j}| |\mathbf{u}_i - \mathbf{v}_i|^2\right). \quad (4)$$

The main advantage of this kernel is that the number of θ to estimate will be reduced significantly from d to h. Indeed the number of latent variables to retain is often less than 4. Also, this approach allows us to build an anisotropic spatial correlation matrix with few θ_j to estimate.

3 Results

Academic problem (Griewank function) and two engineering problems are carried out. Relative error ($\frac{||\hat{\mathbf{Y}} - \mathbf{Y}||}{||\mathbf{Y}||} 100$, with $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ are the vectors containing true responses and predicted responses of validation points, respectively) and CPU time (h refers to hour, mn refers to minutes and s to seconds) are reported in Table 1.

Table 1. Relative error and CPU time for Griewank function and two engineering problems.

Surrogate	Griewank function 60D (800 sample points)		Engineering problem 10D (1295 sample points)		Engineering problem 24D (99 sample points)	
	error (%)	CPU time	error (%)	CPU time	error (%)	CPU time
Kriging	11.47	4 mn 53 s	5.37	1 h 30 mn	0.082	8 s
Kriging/PLS (1 component)	7.4	6.88 s	5.07	11.69 s	0.079	0.12 s
Kriging/PLS (2 components)	6.04	12.57 s	5.02	1 mn 22 s	0.079	0.43 s
Kriging/PLS (3 components)	5.23	16.82 s	5.34	7 mn 34 s	0.079	0.82 s

The results of Kriging combined with PLS show a better accuracy and a faster CPU time than ordinary Kriging existing in literature. The main perspective is to couple this method with an efficient global optimizer (EGO algorithm) to carry out an optimization in high dimension.

References

[1] J. Sacks, W. Welch, W. Mitchell, and H. Wynn, "Design and analysis of computer experiments.," *Statistical Science*, 1989.

[2] S. Wold, M. Sjstrm, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems* **58**(2), pp. 109 – 130, 2001.