Aerospace Science and Technology ••• (••••) •••–•••

Contents lists available at ScienceDirect

Aerospace Science and Technology

www.elsevier.com/locate/aescte

Development of an optimized trend kriging model using regression analysis and selection process for optimal subset of basis functions

Hakjin Lee [a], Duck-Joo Lee [a,*], Hyungil Kwon [b]

[a] Korea Advanced Institute of Science and Technology (KAIST), Daejeon, 34141, Republic of Korea
[b] Hyundai Motor Group, Hwaseong-si, Gyeonggi-do, 18280, Republic of Korea

ARTICLE INFO

Article history:
Received 28 June 2017
Received in revised form 4 December 2017
Accepted 10 January 2018
Available online xxxx

Keywords:
Kriging surrogate model
Trend function
Regression analysis
Coefficient of determination
Genetic algorithms
Design optimization

ABSTRACT

Surrogate modeling, or metamodeling, is an efficient way of alleviating the high computational cost and complexity for iterative function evaluation in design optimization. Accuracy is significantly important because optimization algorithms rely heavily on the function response calculated by surrogate model and the optimum solution is directly affected by the quality of surrogate model. In this study, an optimized trend kriging model is proposed to improve the accuracy of the existing kriging models. Within the framework of the proposed model, regression analysis is carried out to approximate the unknown trend of the true function and to determine the order of the universal kriging model, which has a fixed form with a mean structure dependent on the order of model. In addition, the selection of an optimal basis function is conducted to separate the useful basis function terms from the full set of the basis function. The optimal subset of the basis function is selected with the global optimization algorithm; which can accurately represent the trend of true response surface. The mean structure of proposed model has been optimized to maximize the accuracy of kriging model depending on the trend of true function. Two- and three-dimensional analytic functions and a practical engineering problem are chosen to validate the proposed model. The results showed that the OTKG model yield the most accurate responses regardless of the number of initial sample points, and can conversed into well-trained model with few additional sample points.

1. Introduction

Industry and academia have continuously been attempting to solve engineering design problems with complex geometry and a highly unsteady flow because computing performance has been continuously growing. To find the optimal solution of engineering system, the objective and constraint functions as a function of the design variables have to be iteratively evaluated. However, high-fidelity analysis of the complex configuration, such as an airplane including the pylon and the intake or the unsteady simulation of rotary systems (e.g., helicopter rotors, wind turbines, and open rotor systems), still requires a computing time of several hours or days to obtain the converged solutions. It is nearly impossible for high-fidelity analysis to be directly applied to the design optimization process because of high computational costs and resources. The surrogate model, which is often called a metamodeling is an efficient way to alleviate this computational burden. It represents a true response surface using a simple mathematical function with evaluated function values of sample points. Then, the iterative and expensive function evaluation can be substituted for modeled response surface instead of actual simulation. Therefore, the accuracy of the surrogate model is significant because the optimization results are significantly affected by the quality of the surrogate model. However, constructing high-fidelity surrogate model for complex problems with numerous variables are challenging because large number of variables has much influence on the efficiency of the optimization process. P. Hao et al. suggested a bi-step surrogate-based optimization framework with adaptive sampling to build high-fidelity surrogate modes with less computational cost for complex engineering designs [1].

Several surrogate models have been developed, such as polynomial response surfaces, Artificial Neural Networks (ANN) [2,3], Genetic Programming (GP) [4], Support Vector Regression (SVR) [5], the Radial Basis Function (RBF) [6], Moving Least Squares (MLS), and the Kriging model [7]. The kriging model is one of the most attractive models because it has a good capability of dealing with nonlinear response. Although the true function is explicitly unknown, the kriging model can provide statistical error information that is modeled using a Gaussian process as well as the predicted function response at an untried point. Therefore, it is widely used

* Corresponding author.
E-mail address: djlee@kaist.ac.kr (D.-J. Lee).

https://doi.org/10.1016/j.ast.2018.01.042
1270-9638/© 2018 Elsevier Masson SAS. All rights reserved.

in various research fields, including spatial analysis, mathematical geology, and engineering.

The fundamental formulation of the kriging model is consisted of the two parts: the drift function and the deviation function. The former represents the global trend of the kriging model, while the latter is a localized variation between the true and the drift functions. The accuracy of the kriging model relies greatly on how to formulate them. Many studies have been conducted to improve the accuracy of the kriging model. H.S. Chung and J.J. Alonso used secondary information, such as the values of the gradient, in addition to primary function values at sample points for constructing a covariance matrix of the deviation terms [8]. Z.H. Han et al. suggested a new Cokriging model that utilized both the function values at sample points obtained by the variable fidelity analysis and gradient values computed by the adjoint method to generate the kriging model [9,10]. Their results show that the accuracy of the kriging model can be enhanced by using the gradient information and the function values computed by variable fidelity analysis. They have focused on the modification of deviation terms to improve the quality of the kriging model. In contrast, V.R. Joseph et al. proposed the blind kriging model that uses the optimally selected basis functions to model the trend function. The optimal subset of basis functions can be selected by the Bayesian forward selection process [11]. However, the Bayesian forward selection process could easily get stuck in the local optimum solution rather than finding the global optimum. This converging problem was overcome in dynamic kriging model which was suggested by L. Zhao et al. In dynamic kriging model, the optimization problem of selecting basis functions from the candidates of basis functions was solved by using genetic algorithm which is one of the most popular global optimization algorithm. The kriging process variance was used as the objective function of the optimization problem for finding the optimal subset of basis function. It was found that the quality of the kriging model can be enhanced by excluding unnecessary polynomial terms in the full set of basis function [12]. However, H. Liang and M. Zhu pointed out that the kriging process variance cannot be set to be the objective function of the optimization problem for searching optimal basis functions and genetic algorithm cannot converge to the global optimum. It is analytically proved [13]. A revised dynamic kriging model has been proposed to design the trend function using cross-validation method, and the cross-validation root mean-square error and cross-validation error correlation coefficients were used to be the objective function in the optimization problem of designing the trend function [14]. To find the optimal subset of basis function in the optimization problem, the highest-order of trend function needs to be determined first. In dynamic and revised dynamic kriging model, it is determined to satisfy a constraint associated with the number of samples and the total number of possible candidates of basis functions. However, this constraint depends strongly on the number of sample points and does not consider the trend of the true response. H.I. Kwon and S.I. Choi has developed the R^2 indicator based on regression analysis. The coefficient of determination, denoted R^2 indicates that how well the regression model can approximate the trend of sample points. The unknown trend of the true response could be approximately predicted, and the well-matched order of the universal kriging (UKG) model can be determined depending on the coefficients of determination. It is called the trended kriging (TKG) model because its mean structure is constructed to fit the trend of the true response more accurately by considering the trend of the true function. The results showed that the TKG can improve the accuracy of the model by adjusting its drift function to the identified trend of the true function [15]. However, the form of the drift function in the mean structure is fixed as a p-th order polynomial function. Although the order of the drift function is properly determined from the regression analysis, the unnecessary terms in the fixed form of the drift function could deteriorate the quality of the kriging model.

In this study, an optimized trend kriging (OTKG) model is suggested to improve the accuracy of the TKG model by excluding the unnecessary terms from full set of basis function in mean structure. Therefore, we adopted the global optimization algorithm to separate the useful terms from the fixed form of the basis function of the TKG. In order to validate the OTKG model and compare its accuracy with the ordinary kriging (OKG) model and the UKG models, two- and three-dimensional analytic functions were applied. The validation results verified that the proposed OTKG model can be applied to any trend of response and provide a more accurate response surface than existing kriging models. The proposed OTKG model was also applied to a practical engineering problem. The numerical example shows that the OTKG model can more accurately represent the true response, despite a lack in the number of sample points.

The outline of this paper is as follows. The methods for the optimized OTKG model, including the basic background of the kriging model, trend identification and optimal basis selection process, are introduced in the following section. The detailed validation procedure and the results of using two- and three dimensional analytic functions are described in section 3. Section 4 explains a practical engineering problem and shows the results of model comparison, depending on the dimension of the problem, and the accuracy of the proposed model and the existing model are compared. Our conclusions are discussed in Section 5.

2. Background and methods

2.1. Kriging model

The kriging model was initially suggested to find locations for a borehole by D.G. Krige [7] and mathematically formulated by G. Matheron [16]. It is an interpolation-based surrogate model and perfectly passes through all sample points which are extracted by the Design of Experiment (DoE) approach. The function values of selected sample points must be evaluated by numerical simulation or experimentation. In the kriging model, the deterministic form of the true function is assumed to be the stochastic form of the function. As mentioned above, the kriging model is modeled as the sum of the drift function and the deviation function, as shown by Eq. (1). The first term on the right-hand side of Eq. (1) is the mean structure of the model that globally presents and emulates a mean trend of the true response, while the second term is a deviation between the true and drift functions.

$$\mathbf{y} = \mathbf{F}\boldsymbol{\beta} + \mathbf{Z} \tag{1}$$

The drift function in the kriging model can be formulated using the p-th order polynomial function which is called as p-th order universal kriging (UKG) model. Its drift function can be written as shown by Eqs. (2)–(4), where $\mathbf{y}$ is the vector of the response values at the sample points, $\mathbf{x}$ is the vector of the sample points ($\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]^T$ with $\mathbf{x}_i \in \mathbf{R}^n$), n is the number of design variables (the dimension of the design space), and m is the number of sample points. In this study, the Latin Hypercube Sampling (LHS) method is used to randomly select the sample points in the design space. It is known that the LHS method is well-fitted to the kriging model [17]. $\mathbf{F}$ is the $m \times k$ model matrix that is composed of the p-th order polynomial form of the basis function, where k is the number of elements in the full basis function, $\mathbf{f}(\mathbf{x})$. $\boldsymbol{\beta}$ is the vector of the regression coefficients for the polynomial function that is determined with the Generalized Least Square (GLS) method [18].