11th World Congress on Structural and Multidisciplinary Optimization
7th - 12th, June 2015, Sydney Australia

A Parallel Optimization Method Based on Kriging Model

Zhaojun Li[1,2], Xicheng Wang[2]

[1] School of Computer Science and Technology, Dalian University of Technology, Dalian 116023, China, lizhaojun@mail.dlut.edu.cn
[2] State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116023, China, guixum@dlut.edu.cn

1. Abstract
In many practical engineering designs, the forms of objective functions are not easily available in terms of the design variables. So as to obtain relatively accurate optimal solutions, a lot of computing iterations must be taken into account, which results in huge computational consumption. A parallel optimization method based on Kriging model is proposed in this paper: by developing entropy-based expected improvement algorithm on the Kriging model, this approach progressively provides a designer a rich and evenly distributed set of Pareto optimal points; meanwhile the implementation of the method proposed on a super-server is discussed, with speedup and efficiency described by statistical data. Two mathematical optimization problems and two engineering optimization problems are presented as testing cases, with the results showing that the proposed method can be effectively used for engineering optimization designs. The method developed in this paper could help the designers to search and compare near optimal design alternatives intelligently in the early design stages by utilizing high performance computing resources.
2. Keywords: expected improvement, engineering optimization, Kriging model

3. Introduction

Optimization refers to finding the values of decision variables, which correspond to and provide the maximum or minimum of one or more desired objectives. Problems may arise in engineering optimization at present: 1) The physical and mathematical models for multidiscipline field are complicated to be optimized directly, so that some general analysis programs, such as Ansys, Abaqus, Fluent, Moldflow, etc, have been used as black-box programs, how to develop the efficient optimization methods based on these black-box programs; 2)High performance computing that began to appear in the late 1970s and continue to undergo rapid development, how to solve a complex optimization problem by using advanced computing infrastructure efficiently [1-3]. Physical and mathematical models are hard to describe explicitly under normal engineering optimization situations, which makes "black-box optimization" a very effective way in solving these problems. This "only-use-function-values" optimization method is hailed as "the most useful algorithm" [4].
In this paper, a parallel optimization method based on Kriging model is developed. By weighted coefficient method of multi-objective optimization, an evenly and rich distributed set of Pareto solutions can be provided, in which a Pareto point obtained by optimizing weighted coefficient–with entropy-based expected improvement (EEI) algorithm. To make the method well performed within parallel computing environment, the parallel strategy is given. Test cases are developed, with numerical data showing that the proposed method not only effectively gives more precise optimal solutions through EEI algorithm but accelerates the whole optimization process using parallel strategy at the same time.

4. Related Work

4.1. "Black-box Optimization"
A typical "black-box optimization" process performed in this article is be defined as: 1) Use sampling method (Latin hypercube sampling (**LHS**) [5], for example) to get a set of fairly well-distributed samples which are consistent with the limits of the design variables; 2) Compute the responses of all the samples using a black-box analysis program; 3) Try to set up a proximate model between the design variables and the responses so as to analyze the system features; 4) Calculate the optimal solution of the objective function under the approximation model established before with an appropriate sampling guidance function such as EI method; 5) Convergence analysis: if the optimal solution has met the given accuracy then stop; Otherwise, put the optimal solution into the sampling set as a new sample and go to step 2) [6].

4.2. Kriging Model

Samples mentioned above can be used as input data for “black-box optimization” and their responses would be acquired after necessary calculations performed by black-box analysis software. With samples $X = [X_1, X_2, \ldots, X_n]$ and their responses y, Kriging model can be established as [7,8]:

$$\hat{y}(X) = f^{\mathrm{T}}(X)\beta + z(X) \tag{1}$$

coefficient β is the regression ratio. Deterministic drift $f(X)$, provided the global approximation of simulation in the design domain, is often described as a polynomial of X. $z(X)$, known as fluctuation, offers the local approximation of simulation.

4.3. Expected Improvement

Expected Improvement (**EI**) criterion is a sampling guidance function which considers both the predicted value and the predicted variance while infilling a sample into the optimization models. To a new sample X, the Kriging model can predict its average value $\overline{y}(X)$ and its average variance σ^2. Let y_{min} be the current minimum response value, therefore I (lets assume that it is a minimization problem and $y(X) = y_{min} - I$) will be the improvement of the response value at the given point. I is normally distributed with mean $\overline{y}(X)$ and variance σ^2. The likelihood of this improvement is given by the normal density function [9]:

$$\frac{1}{\sqrt{2\pi}\sigma(X)} \exp\left[-\frac{\left(y_{min} - I - \overline{y}(X)\right)^2}{2\sigma^2(X)}\right] \tag{2}$$

The expected improvement is simply the expected value of the improvement found by integrating over this density

$$\mathrm{E}\left[I(X)\right] = \int_{I=0}^{I=\infty} I\left\{\frac{1}{\sqrt{2\pi}\sigma(X)} \exp\left[-\frac{\left(y_{min} - I - \overline{y}(X)\right)^2}{2\sigma^2(X)}\right]\right\} dI \tag{3}$$

using integration by parts, one can show that

$$\mathrm{E}\left[I(X)\right] = \sigma(X)\left[v\Phi(v) + \varphi(v)\right], v = \frac{y_{min} - \overline{y}(X)}{\sigma(X)} \tag{4}$$

where Φ and φ are the normal cumulative distribution function and density function, respectively.

5. Developed Methodology

5.1. Entropy-based Expected Improvement

In order to facilitate parallel computing, entropy-based expected improvement (**EEI**) is developed to optimize the infilling samples in the design space. The optimization of the weighted expected improvement (**WEI**) can be written as

$$\max\ \mathrm{E}\left(I(X)\right) = \sum_{j=1}^{2} \lambda_j \mathrm{E}_j\left(I(X)\right) \tag{5}$$

and

$$\begin{cases} \mathrm{E}_1\left(I(X)\right) = \Phi(v)\left(y_{min} - \overline{y}(X)\right), \\ \mathrm{E}_2\left(I(X)\right) = \sigma(X)\phi(v), \\ \sum\limits_{j=1}^{2} \lambda_j = 1, \lambda_j \in [0,1] \end{cases} \tag{6}$$

Shannon entropy can be introduced to measure the uncertainty about the searching range. If $\lambda_j \in [0,1]$ are here defined as a probability that the optimal solution occurs in the local space and other space, respectively, then the Shannon entropy will be decreased during optimization process of Eq.5 and an entropy-based optimization model for the optimizing weighting coefficient can be constructed as

$$\begin{cases} \min\ -\sum\limits_{j=1}^{2} \lambda_j \mathrm{E}(I) \\ \min\ H = -\sum\limits_{j=1}^{2} \lambda_j \ln(\lambda_j), \\ \sum\limits_{j=1}^{2} \lambda_j = 1, \lambda_j \in [0,1] \end{cases} \tag{7}$$

where H is the information entropy. It can easily be proved that Eq.5 and Eq.7 having the same optimal solutions. By Lagrange multiplier method, an augmented function can be obtained as

$$\begin{cases} L(I,\eta,\mu) = -(1-\eta)\sum\limits_{j=1}^{2}\lambda_j \mathrm{E}_j - \eta\sum\limits_{j=1}^{2}\lambda_j \ln(\lambda_j) + \mu(\sum\limits_{j=1}^{2}\lambda_j - 1), \\ \eta \in (0,1) \end{cases} \tag{8}$$

where η and μ are the weighting coefficient of multi-object optimization and Lagrange multiplier, respectively. Solving Kuhn-Tucker condition Eq.9 as follows

$$\frac{\partial L}{\partial \lambda_j} = 0, \frac{\partial L}{\partial \mu} = 0 \tag{9}$$

gives

$$\lambda_j^* = \frac{\exp\left[r\mathrm{E}_j(I)\right]}{\sum\limits_{j=1}^{2}\exp\left[r\mathrm{E}_j(I)\right]} \tag{10}$$

in which $r = (\eta-1)/\eta$ is called as the quasi-weighting coefficient (here $\eta = 0.5$). λ_j^* is here called as the optimal weighting coefficient. Then a Pareto optimal sample can be obtained by solving Eq.5 with the optimal weight coefficient λ_j^* of Eq.10. For a specific weight coefficient $\lambda_j \in [0,1]$, the optimization scheme stops when

$$\frac{\mathrm{E}\left[I(X)\right]}{y_{max} - y_{min}} \le \varepsilon_1 \tag{11}$$

where ε_1 is the stopping tolerance, y_{max} and y_{min} are the maximal and minimal function value in samples, respectively. Then considering the accuracies of both Kriging model and optimization simultaneously, besides Eq.11 the convergence condition of Eq.12

$$\frac{|f(X_n) - \hat{y}_n|}{f(X_n)} \le \varepsilon_2 \tag{12}$$

should be satisfied too. Where, ε_1 and ε_2 are given errors (here take $\varepsilon_1 = \varepsilon_2 = 0.01$), $\hat{y}_n$ is the approximate response value of the Kriging surrogate model.

5.2 A Parallel Strategy of EEI on Super Servers

It is assumed here that p cores can be used for the computation on super servers, and the parallel algorithm is described as follows

Step.1. Generate a set of N_s samples (each point corresponding to a group design variables) using LHS. Divide the N_s samples into p parts, which used as input data for black-box analysis programs, the first $p-1$ cores get $\lceil N_s/p \rceil$ samples, the last core gets the remaining samples;

Step.2. Compute the responses of the samples using black-box program on each core respectively;

Step.3. Build a Kriging surrogate model between the sample sets and their corresponding output response values;

Step.4. Solve the maximizing problems of Eq.5 by the p cores, in which the weighted parameter of λ_1 is obtained by Eq.10 and $\lambda_i = (i-1)/p, i = 2,3,\ldots,p$ for the rest of the computing cores;

Step.5. Compute the responses of the samples after step 4 is accomplished by using same black-box program;

Step.6. Check convergence. If the convergence criteria Eq.11 and Eq.12 are satisfied then stop, the optimization solution is obtained; Put new samples and their responses into the sampling set and the responding set respectively, redirect to step 3.

6. Test Cases

6.1 Math problems

In this section, the results for some classic mathematical problems are presented. For all of the problems, LHS is used and the number N_s of initial samples is 10, available computing cores p is 4.

6.1.1 Ackley Function

Ackley function is widely accepted by form of:

$$\begin{aligned} &F(x_1,x_2) = -2\exp(-0.2\sqrt{x_1^2 + x_2^2}) - \exp((cos(2\pi x_1) + \\ &cos(2\pi x_2)) + 2 + \exp(2), (x_1,x_2) \in (-1.5,1.5) \end{aligned} \tag{13}$$

theoretically, the minimum value of Ackley function is 0 at (0,0). The iteration histories of Ackleys problem are shown in Fig. 1. The optimal value 0(0,0) is obtained after 33 iterations of WEI method and 10 iterations of EEI

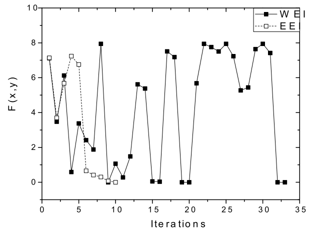

Figure 1: Iteration histories of Ackley problem

method, respectively.

6.1.2 Branin Function

Branin function is defined as:

$$F(x_1,x_2) = (x_2 - 5.1x_1^2/(4\pi^2) + 5x_1/\pi - 6)^2 + 10(1-0.125/\pi)cos(x_1) + 10, \\ x_1 \in [-5,10.0], x_2 \in [0,15.0] \tag{14}$$

its optimal solution is 0.3979 at point (-3.1416, 12.28). Iteration histories of this problem is shown in Fig. 2. The optimization value 0.784(9.342, 3) is obtained by WEI method after 16 iterations and 0.5031(-3.271, 12.75)

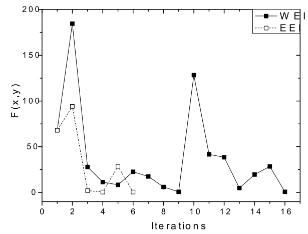

Figure 2: Iteration histories of Branin problem

obtained by EEI method after 6 iterations. Method of EEI has given a more precise solution which is closer to the optimal solution should be in theory.

6.2 Optimization Problem of Turbine Foundation

A turbine foundation is normally a concrete base for turbine generators. The turbine foundation optimization (TFO) problem can be described as:

$$\begin{aligned} &\min F(X) = \min F(X_s, X_g)^T = \min f_l(X_s, X_g)^T \\ &s.t. \begin{cases} u_q(X_s,X_g) \le \bar{u}_q, q = 1,2,\dots,Q \\ M\ddot{u}(t) + C\dot{u}(t) + Ku(t) = P(t) \\ \underline{x}_s^i \le x_s^i \le \vec{x}_s^i, \underline{x}_g^i \le x_g^i \le \vec{x}_g^i, \\ l = 1,2,\dots,L, i = 1,2,\dots,I, j = 1,2,\dots,J \end{cases} \end{aligned} \tag{15}$$

where, $X = (X_s, X_g)$ stands for design variables: X_s is a size design vector with I beam section areas, X_g is geometry design vector with J columns node coordinates, $\underline{x}_s^i$, $\vec{x}_s^i$ and $\underline{x}_g^i$, $\vec{x}_g^i$ represent section and geometry limits, i and j are defined as variable numbers; here $L = 2$, f_1 and f_2 are separately defined as amplitude computed by black-box analysis program (here Ansys software is used) dividing current average amplitude and construction weight (a linear function of the section areas) dividing current average weight of the foundation; $u_q(X)$ is the amplitude of the points concerned; Q is the number of points concerned; $\bar{u}_q$ is the maximum amplitude of the foundation; $P(t)$ is disturbing force of the generator. TFO is a multi-objective engineering optimization problem. By means of the weighted coefficient method for solving multi-objective optimization, a weighted parameter ω to unify amplitude and weight to one objective function in order to reduce the complexity of modeling using Kriging method in

Table 1: Speedup and efficiency of test cases

Test Case	WEI		EEI	
	Speedup	Efficiency(%)	Speedup	Efficiency(%)
6.1.1	3.243	81.075	3.333	83.325
6.1.2	3.332	83.300	3.320	83.000
6.2.1	3.259	81.475	3.350	83.750
6.2.2	3.351	83.775	3.299	82.475

engineering practice. In this article, ω is defined as 0.5 for considering both structural weight and the amplitude at the same level. Therefore should $F(X)$ become $F(X) = 0.5f_1(X) + 0.5f_2(X)$ by now.
6.2.1 A 600MW Turbine Foundation Example
There are 15 size design variables and a geometry design variable for this example. The iteration histories are available in Fig. 3.

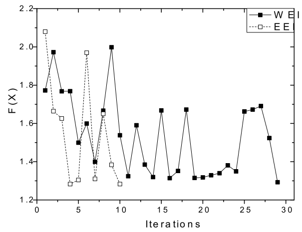

Figure 3: Iteration histories of the 600MW turbine foundation

Fig. 3 gives that iterations of EEI (10 optimization steps at optimal solution of $F(X) = 1.27$) is less than WEI (29 optimization steps at optimal solution of $F(X) = 1.28$).

6.2.2 A 300MW Turbine Foundation Example
There are 17 size design variables and two geometry design variables for this example. Fig. 4 and has described the iteration processes.

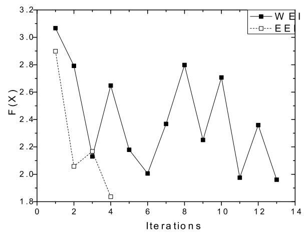

Figure 4: Iteration histories of the 300MW turbine foundation

According to Fig. 4, optimization iterations needed by EEI method (4 steps) is less than WEI method (13 steps); meanwhile WEI method has also got a not so good optimization result ($F(X) = 1.92$) than EEI ($F(X) = 1.83$), which is a very huge effort for this problem as a 0.1 more less of the objective function would gain nearly about $10^5 kg$ saving of the foundations construction of weight consume.

6.3 Speedup & efficiency statistic data for test cases
Table 1 has given the characteristics belonging to all the test cases including speedup and efficiency using the parallel strategy discussed in 5.2.

6.4 Results analysis & discussion

As can be seen from the above figures, the developed EEI method can normally calculate better optimization solutions and WEI method (and also EI method, obviously). This is a behavior caused by a reasonable weighted parameter is given while doing the optimization, which stands for a correct and direct way for searching final optimal solutions generally. A better solution always means better objective function values, which indicates great construction cost savings and better other related engineering technical indicators are obtained.

An ordinary way to save time cost while doing optimization must be making it into parallel pattern. Reading from Table 1, both WEI and EEI methods have expressed high speedup and efficiency for all test cases. This is very simple to comprehend that black-box optimization has established based on Kriging models: black-box programs are scheduled without too much correspondence to each other henceforth a much more considerable speedup could be made. Using the parallel strategy proposed above, engineering optimization problems can be estimated to be performed in a quicker mode.

7. Conclusion

This paper has developed an improved expected improvement method called entropy-based expected improvement (EEI) method, adapting to address computing-intensive optimization problems. It is based on black-box optimization method and global Kriging models. An optimal weighted parameter is calculated at each iteration, which is used for sampling guidance next iteration: it has similar process steps to normal WEI method but gives more accurate solutions under normal circumstances. A parallel strategy described as responses computed by different computing cores with black-box programs has been successfully implemented on a super server to deal with the problem encountered when EEI method needs to consider several samples and points at the same iteration, which could get large speedup and efficiency for engineering problems due to black-box optimizations manifest excellent character to be paralleled.

In order to verify the effectiveness of the proposed method, several examples consist of both math and engineering problems are developed. Numerical results have demonstrated that the raised method could get good speedup and better optimal solutions simultaneously, which surely gives an advance in parallel engineering optimization. We hope the method discussed may further be used in other related engineering optimization problems due to the common behaviors in black-box optimization as to ease the cost made by all of them.

8. Acknowledgements

The work proposed in this paper has sponsored by the following funds: the National Basic Research Program of China (No. 2012CB025905), the Fundamental Research Funds for the Central Universities of China (No. DUT14RC(3)060), the National Natural Science Funds of China (No. 11202049) and the 111 Project (No. B14013).

9. References

[1] Rao S S and Rao S S, *Engineering optimization: theory and practice*, John Wiley & Sons, 2009.

[2] Wang L, Ng A H C, Deb K, et al, *Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction*, Springer London, 2011.

[3] Prof Geng J, Prof Yan W, Dr. Xu W, et al, *Application of Commercial FEA Software*, Springer Berlin Heidelberg, 2008.

[4] Powell M, Direct search algorithms for optimization calculations, *Acta Numerica*, 7, 287-336, 1998.

[5] Iman R L, *Latin hypercube sampling*, Wiley Online Library, 2008.

[6] Regis R G, Shoemaker C A, Constrained global optimization of expensive black box functions using radial basis functions, *Journal of Global Optimization*, 31(1), 153-171, 2005.

[7] Oliver M A R, Kriging: A method of interpolation for geographical information systems, *Int. J. Geograph. Inf. Syst.*, 4, 313-332, 1990.

[8] Shi H, Gao Y, Wang X, Optimization of injection molding process parameters using integrated artificial neural network model and expected improvement function method, *The International Journal of Advanced Manufacturing Technology*, 48(9-12), 955-962, 2010.

[9] Vazquez E, Bect J, Convergence properties of the expected improvement algorithm with fixed mean and covariance functions, *Journal of Statistical Planning and inference*, 140(11), 3088-3095, 2010.