

Sequential Optimization Method for the Design of Electromagnetic Device

Gang Lei , K. R. Shao , Youguang Guo , Jianguo Zhu , and J. D. Lavers , *Fellow, IEEE*

¹ College of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China

² Faculty of Engineering, University of Technology, Sydney, N.S.W. 2007, Australia

³ Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4 Canada

Three sequential optimization methods, sequential least square method, sequential Kriging method, and sequential linear Bayesian method, are presented for the optimization design of electromagnetic device. Sequential optimization method (SOM) is composed of coarse optimization process and fine optimization process. The main purpose of the former is to reduce the design space; while the target of the latter is to update the optimal design parameters. To illustrate the performance of the proposed methods, an analytic test function and the TEAM Workshop Problem 22 are investigated. Experimental results of test function demonstrate that SOM can obtain satisfactory solutions; and practical application illustrates that the number of finite element sample points is less than 1/10 compared with that by direct optimization method, while the optimal results are even better than that by direct optimization method.

Index Terms—Kriging, linear Bayesian estimation (LBE), response surface model, sequential optimization method (SOM).

I. INTRODUCTION

OPTIMIZATION is an art of getting the best solution among the feasible solutions. The target of electromagnetic device optimization is to determine a set of physical parameters to satisfy certain design specifications. Engineers have been using optimization techniques for devices, components, and system modeling for decades.

Many electromagnetic optimization problems are solved by means of highly accurate models (e.g., finite element model) with direct optimization algorithms, such as genetic algorithm [1], clonal selection algorithm [2], and differential evolution algorithm (DEA) [3]. However, these approaches are computationally expensive. As an alternative, many statistical approximate models, such as response surface model (RSM), radial basis functions model and Kriging model are also widely employed in practical engineering problems [4]–[7]. They are proved fast, but not very accurate.

In the traditional optimization method, models and algorithms were almost discussed separately, but, indeed, optimization is a simultaneous updating process about models and algorithms. In this paper, we try to discuss both of them simultaneously within the framework of SOM.

II. PARAMETER ESTIMATION THEORY FOR RSM

Give a set of n sample points $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ with $x_i \in \mathbb{R}^m$ and responses $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ with $y_i \in \mathbb{R}$. Now, for an m dimensional input x , the response value $y(x)$ can be expressed as a realization of a low-order polynomial regression model with a random error function. Quadratic polynomials are the most commonly used in electromagnetic opti-

mization problems, which are also used in this work. Then we have the approximate model as

$$y(x) = \beta_0 + \sum_{i=1}^m \beta_i x_i + \sum_{i=1}^m \sum_{j=1}^m \beta_{ij} x_i x_j + w(x) \quad (1)$$

where $w(x)$ is the error function and β s are the regression parameters. To estimate the model parameters, (1) can be written in matrix notation as

$$\mathbf{y} = \mathbf{H}\beta + \mathbf{w} \quad (2)$$

where \mathbf{y} is an $n \times 1$ vector of observations with respect to n sample points \mathbf{x} , \mathbf{H} is a known $n \times l$ observation matrix of rank l with $l = (m+1)(m+2)/2$ and \mathbf{w} is an $n \times 1$ error vector with zero mean and covariance \mathbf{C}_w (the probability density function of \mathbf{w} is otherwise arbitrary).

Least square method is always used to estimate the parameters β . Minimize the weighted least square error

$$J = (\mathbf{y} - \mathbf{H}\beta)^T (\mathbf{y} - \mathbf{H}\beta). \quad (3)$$

Then the least square estimation (LSE) of β is

$$\hat{\beta} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}. \quad (4)$$

It should be noted that LSE is only a coarse estimation. When the dimension of the optimized problem is high or the number of the response points is very large, many difficulties are that we must face, such as matrix singular. So we present two improved parameter estimation methods, which are termed Kriging and linear Bayesian estimation (LBE), respectively.

A. Kriging

Kriging is a general term used for the family methods of minimum mean square error estimation. Under the assumption of (2), the best linear unbiased estimation of β has the form as

$$\hat{\beta} = (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{y}. \quad (5)$$

As we know, the above estimation is the minimum variance linear unbiased estimation, which is the one whose variance for each component is minimum among all unbiased estimations. It is an optimal estimation in the statistical sense.

The covariance matrix $C_w = [c_{ij}]$ can be defined as

$$c_{ij} = \sigma^2 \mathbf{R} [R(x_i, x_j)], \quad i, j = 1, 2, \dots, n \quad (6)$$

where \mathbf{R} is the correlation matrix, and R is the user-specified correlation function. Gaussian correlation functions are most commonly used [6], [7]. They have the form as

$$R(x_i, x_j) = \exp \left\{ - \sum_{k=1}^n \alpha_k |x_i - x_j|^2 \right\}. \quad (7)$$

Implementing a Kriging method consists of estimating the parameters β in (2), σ^2 in (6) and α_k in (7). All parameters can be easily estimated by the software package Design and Analysis of Computer Experiments (DACE) [7].

B. Linear Bayesian Estimation

LSE and Kriging were gained without any prior information of β . If some prior information is available, LBE can be a better choice. LBE is implemented and quantified within the formalism of Bayesian statistics and linear unbiased estimation theory. In the data model of (2), and in the assumption of β is a $l \times 1$ random vector of parameters whose realization is to be estimated and it has prior mean μ_β and covariance matrix C_β (the probability density function of β is otherwise arbitrary). For the estimation to be linear, we require

$$\hat{\beta}_i = \sum_{k=1}^n a_{ik} x_k + a_{i0}, \quad i = 1, 2, \dots, l \quad (8)$$

where a_{ik} are constants yet to be determined for $k = 0, 1, \dots, n$. In order for $\hat{\beta}_i$ to be unbiased we require

$$E(\hat{\beta}_i) = a_{i0} + \sum_{k=1}^n a_{ik} E(x_k) = \beta_i, \quad i = 1, 2, \dots, l \quad (9)$$

where E is the expectation, and the Bayesian mean square error (BMSE) of $\hat{\beta}_i$ is [8]

$$\text{BMSE}(\hat{\beta}_i) = E \left[(\beta_i - \hat{\beta}_i)^2 \right], \quad i = 1, 2, \dots, m. \quad (10)$$

Then the LBE is found by minimizing (10) subject to the constraints of (9), which lead to

$$\hat{\beta} = \mu_\beta + \left(C_\beta^{-1} + H^T C_w^{-1} H \right)^{-1} H^T C_w^{-1} (y - H \mu_\beta). \quad (11)$$

The main work in the LBE is the selection of the prior information for β . Fortunately, LSE is a coarse estimation which can be simply gained. So in our latter implementation of LBE, LSE is used for its prior information. Obviously, Kriging and LBE are more accurate and efficient than LSE, especially for the high dimension optimization problems. However, they are also more complex than the LSE.

III. SEQUENTIAL OPTIMIZATION METHOD

SOM can be summarized in six main steps and its flowchart is shown in Fig. 1.

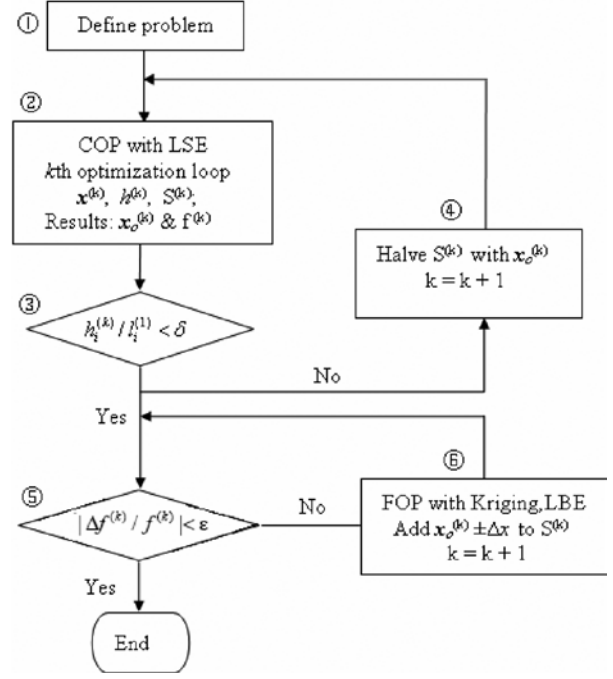


Fig. 1. Flowchart of SOM.

- 1) Define problem. This is the pretreatment process of the optimized problem, Such as using experiment design approach to select sample points and using analysis of variance to determine the distinct factors.
- 2) Coarse optimization process (COP). The target of COP is to reduce the design space. LSE is employed to reconstruct the the design space of the k th optimization process is $[x_{li}^{(k)}, x_{ui}^{(k)}]$ and $l_i^{(k)}$ is the interval of the i th variable. $h_i^{(k)}$ is the step size and $h_i^{(k)} = h_i^{(k-1)}/2$; and $N_i^{(k)}$ is the number of sample points. Note that the value of the step size must ensure that the minimum of the number of sample points is no smaller than 3, otherwise singularity matrix may appear in the matrix inversion process. In our implementation, if the step size can not ensure this request, we use $N_i^{(k)} = 3$ to evaluate the step size. Then we can reconstruct the response surface with the sample set $S^{(k)}$. Through the optimization of the response surface, we can obtain the current optimal design parameters and corresponding response value. They are noted as $\bar{x}_{oi}^{(k)}, \dots, \bar{x}_{oi}^{(k)}$ and $f^{(k)}$, respectively.
- 3) Step size termination. If $h_i^{(k)} / l_i^{(1)} < \delta$ (δ is a positive constant), stop COP and go to step 5, otherwise go to step 4.
- 4) COP loop. Halve the sample space with the optimal value. Under the boundary condition of the design space, the design space of next step is updated as follows:

$$\begin{cases} x_{li}^{(k+1)} = \max \left\{ x_{li}^{(k)}, \text{round}(2a) \cdot h_i^{(k)} / 2 \right\} \\ x_{ui}^{(k+1)} = \min \left\{ x_{ui}^{(k)}, \text{round}(2b) \cdot h_i^{(k)} / 2 \right\} \end{cases} \quad (12)$$

where $a = (x_{oi}^{(k)} - l_i^{(k)} / 4) / h_i^{(k)}$, $b = (x_{oi}^{(k)} + l_i^{(k)} / 4) / h_i^{(k)}$, function $\text{round}(x)$ is round to the nearest integer of x . The

number of sampling points can be remarkably reduced by this updating method.

For example, supposing the initialization of design space is $[0, 1]$ and step size is 0.2; then 6 sample points constitute the first sample data $S^{(1)} = \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. And supposing the current optimal value is 0.35, from (12) we can get the next sample space is $[0.1, 0.6]$. If we also use uniform sampling method with half of the step size, the sample data are $S^{(2)} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$. Obviously, three sample points have been sampled in the last step, in other words, 50% computation cost is reduced. Similarly, supposing the current optimal value is 0.12, the new sample space is $[0, 0.4]$, and new sample data are $S^{(2)} = \{0, 0.1, 0.2, 0.3, 0.4\}$, in this case, 60% computation cost is reduced.

- 5) Optimal results termination. If $|\Delta f^{(k)}/f^{(k)}| < \epsilon$ (ϵ is a small positive constant), stop optimization process and output the optimal results, otherwise go to next step.
- 6) Fine optimization process (FOP): The target of this process is to find the optimal results. Kriging and LBE are employed to reconstruct the response surface of RSM in this process. Moreover, LSE may be also used in FOP. It should be noted that there is no need to update the design space in the FOP. To ensure the accuracy and robustness of optimization process, local multipoint samples updating method is involved here. Given current optimal value, the next sample set $S^{(k+1)}$ is updated by (13), which is constructed by the N_p perturbations around the current optimal value

$$S^{(k+1)} = S^{(k)} \cup \left\{ \mathbf{x}_o^{(k)} \pm \Delta \mathbf{x}_o^{(p)} \mid p = 1, \dots, N_p \right\}. \quad (13)$$

In this paper, $N_p = 2^m$. Two points in each variable is selected, one is the optimal value, and the other is the mean value of the two sample points near the optimal result.

IV. EXPERIMENTS AND RESULTS

Considering the different parameter estimation methods of RSM, we have three SOMs, which are termed as sequential least square method (SLSM), sequential Kriging method (SKM), and sequential linear Bayesian method (SLBM). RSM is constructed with LSE for all of them in COP, while LSE, Kriging, and LBE are used in FOP, respectively. In this paper, DEA is used as the optimization algorithm [3]. The algorithm parameters are: mutation scaling factor is 0.8, crossover factor is 0.8, the maximum number of iteration is 3000, and the maximum stall generation is 100, which is selected for the stop criterion [9].

Table I shows the unconstrained optimization results of an analytic function [5], [6]

$$f = 0.01 \sum_{i=1}^2 [(x_i + 0.5)^4 - 30x_i^2 - 20x_i]. \quad (14)$$

The initial sample space is $[-5.12, 5.12] \times [-5.12, 5.12]$. The initialization of parameters are: $\delta = 5\%$, $\epsilon = 5\%$ and $N^{(1)} = [9, 9]$. It should be noted that the optimal results given in Table I are the best one among 100 runs in each optimization process. Actually, no less than 90 runs can get the best solution (also for

TABLE I
OPTIMIZATION RESULTS OF AN ANALYTIC FUNCTION

Method	χ_1	χ_2	f
Exact values	-4.4538	-4.4538	-5.2328
SLSM	-4.4119	-4.4119	-5.2305
SKM SLBM	-4.4564	-4.4562	-5.2327

TABLE II OPTIMIZATION RESULTS
OF SMES

Var Unit	R_2 m	$h_2/2$ m	d_2 m	B_{stray} mT	B_m T	E_s MJ	FESP
TEAM	3.08	0.239	0.394	0.909	4.73	179.86	---
DEA	3.18	0.428	0.211	1.032	3.83	180.00	2310
SLSM	3.09	0.246	0.376	0.916	4.65	179.63	154
SKM SLBM	3.08	0.239	0.394	0.909	4.73	179.86	162

the Table II). Compared with the exact values, we can obtain satisfactory solutions from our presented methods.

These methods were also applied to the TEAM Workshop problem 22 which is a well known benchmark problem for optimization algorithms [1], [2], [4], [10]. Fig. 2 shows the three-variable case of Superconducting magnetic energy storage (SMES). And SMES consists of two solenoids. The dimensions of the inner solenoid ($R_1, h_1/2, d_1$) stay fixed, while the dimensions of the outer solenoid ($R_2, h_2/2, d_2$) should be optimized to reduce stray fields while keeping the stored energy close to an expectation value. One objective function and two constraints are included in this problem [2]. To be detailed, the objective function is given by

$$f(\vec{x}) = B_{\text{stray}}/B_{\text{norm}} \quad (15)$$

where $B_{\text{norm}} = 3$ mT; $B_{\text{stray}} = \sqrt{\sum_{i=1}^{21} |B_{\text{stray}}^i|^2}/21$, which is the magnetic flux density evaluated along 21 equidistant points on lines $h(\vec{x})$ and $g(\vec{x})$ are two constraints, the former is an equality constraint, concerning the amount of energy stored at the SMES device; and the latter is an inequality constraint, concerning the quench condition that guarantees superconductivity

$$h(\vec{x}) = |E_s/E_{\text{ref}} - 1| = 0 \quad (16)$$

$$g(\vec{x}) = B_m - 4.92 \leq 0 \quad (17)$$

where E_s is the energy stored by the SMES, E_{ref} is the reference energy, and B_m is the highest magnetic field value. In this paper, the constraints were treated using a penalty function as

$$F(\vec{x}) = f(\vec{x}) + 1000 \left[h(\vec{x})^2 + \max(g(\vec{x}), 0)^2 \right]. \quad (18)$$

Table II shows the optimization results given by our proposed methods. $N^{(1)} = [5, 4, 3]$, δ and ϵ are 5%. To illustrate the efficiency of the proposed methods, the results given by direct optimization method are also given in Table II, which is implemented by the finite element model and DEA. Furthermore, as a reference, the best known results (TEAM [10]) are also given in Table II. In order to make these results to be comparable, all response values, B_{stray} , B_m , and E_s , are calculated from the ANSYS with the same preprocessing process.

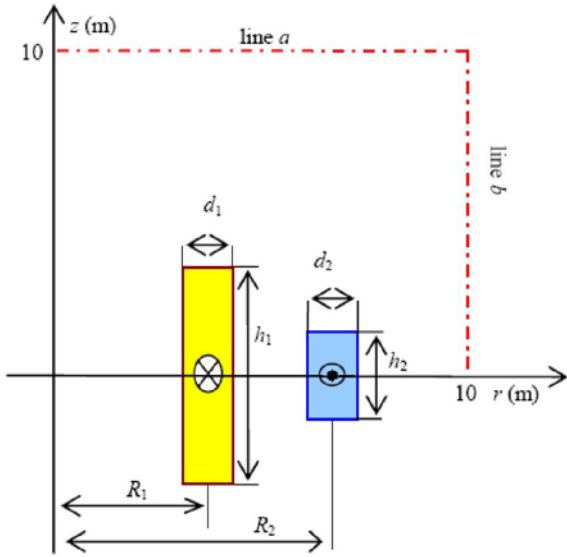


Fig. 2. Design parameters d_1 , d_2 and h_2 of SMES.

Meanwhile, the average convergence speed of the SOM is ranged from 900 and 1320 with DEA. The average convergence time is half of the time of a finite element sampling process. So, the time for optimization algorithm can be neglected compared with finite element sampling process. Three main conclusions are drawn from the Table II.

- 1) For the direct optimization method, the optimal design parameters given by DEA are [3.18 0.428 0.211]. The mean stray field B_{stray} is 1.032 mT, energy is 180.00 MJ, and 2310 finite element sample points (FESP) are needed.
- 2) For the optimization with SLSM, only 154 FESP, less than 7% compared with that of DEA, are needed to get the optimal results, which are [3.09 0.246 0.376]. The energy is 179.63 MJ, and the B_{stray} is 0.916 mT, which is smaller than that given by DEA.
- 3) For the optimization with SKM and SLBM, the optimal results are the same. Only 162 FESP, about 7% compared with that of DEA, are needed to get the optimal results. The optimal results are [3.08 0.239 0.394], which are the same as the results of TEAM. B_{stray} is 0.909 mT, which is smaller than that given by SLSM and DEA. The error of E_s is 0.14 MJ, which is smaller than that of SLSM.

Obviously, SKM and SLBM are better than SLSM and direct optimization method. Finally, as an example to illustrate all the sample points sampled in optimization process, the entire sample data of SKM are displayed in Fig. 3. Actually, five sampling processes (3 COP and 2 FOP) are needed to implement the SKM and SLBM. And the fourth sampling processes in Fig. 3 are the entire sample data of SLSM.

V. CONCLUSION

In this paper, three parameter estimation methods, LSE, Kriging and LBE, are firstly presented for the reconstruction of RSM. Then SOM of models and algorithms are addressed to electromagnetic device optimization problems. SOM only

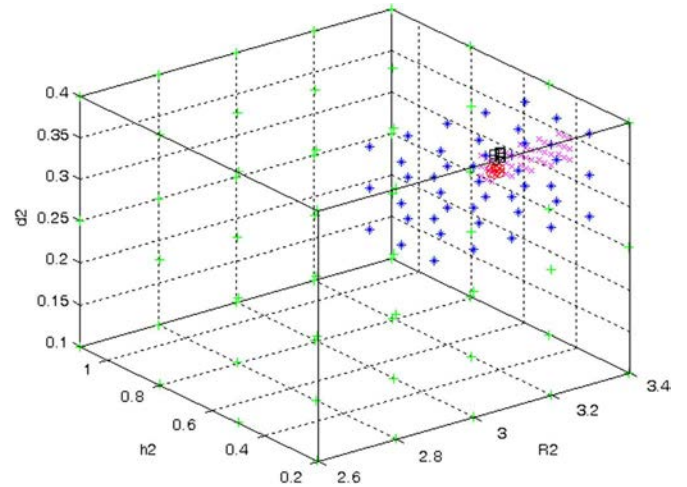


Fig. 3. Entire sample data of SKM (\times for the first COP; \times for the second COP; \times for the third COP; \circ for the first FOP; \square for the second FOP).

need a small sample data, and the overall computational effort needed is much less than that by direct optimization method.

Obviously, SLSM, SKM and SLBM are very successful in optimizing the analytic function and the SMES device. Especially the SKM and SLBM, of which the efficiency is better than that by DEA, while its finite element computation cost is obviously less than 1/10 compared with that by DEA.

It should be noted that SOM can be extended to other approximation models, such as radial basis functions model and other forms of Kriging model. Furthermore, other sampling methods, such as Latin hypercube sampling method, can also be used in the SOM. All these methods can be widely used in engineering electromagnetic optimization problems.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 50877029.

REFERENCES

- [1] R. H. C. Takahashi, J. A. Vasconcelos, J. A. Ramírez, and L. Krahenbuhl, *IEEE Trans. Magn.*, vol. 39, no. 3, pp. 1321–1324, Mar. 2003.
- [2] F. Campelo, F. G. Guimaraes, H. Igarashi, and J. A. Ramirez, *IEEE Trans. Magn.*, vol. 41, no. 5, pp. 1736–1739, May 2005.
- [3] R. Storn and K. Price, *J. Global Optim.*, vol. 11, pp. 341–359, 1997.
- [4] L. D. Wang and D. A. Lowther, *IEEE Trans. Magn.*, vol. 42, no. 2, pp. 1227–1230, Feb. 2006.
- [5] J.-L. Coulomb, A. Kobetski, M. C. Costa, Y. Marechal, and U. Jonsson, *COMPEL*, vol. 22, pp. 616–629, 2003.
- [6] L. Lebensztajn, C. A. R. Marretto, M. C. Costa, and J. L. Coulomb, *IEEE Trans. Magn.*, vol. 40, no. 2, pp. 1196–1199, Feb. 2004.
- [7] S. N. Lophaven, H. B. Nielsen, and J. Sondergaard, DACE: A MATLAB Kriging Toolbox Version 2.0. Technical Univ. Denmark, Copenhagen, Denmark, Tech. Rep. IMM-TR-2002-12, 2002.
- [8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993, pp. 242–250, and 392–400.
- [9] Differential Evolution Homepage [Online]. Available: <http://www.icsi.berkeley.edu/~storn/code.html>
- [10] B. Brandstaetter, SMES Optimization Benchmark [Online]. Available: <http://www.igte.tugraz.at/archive/team/team3dis.htm>