This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Multi-Objective Pareto Optimization of Electromagnetic Devices Exploiting Kriging With Lipschitzian Optimized Expected Improvement

S. Xiao[1], G. Q. Liu[1], K. L. Zhang[1], Y. Z. Jing[1], J. H. Duan[1], P. Di Barba[2], and J. K. Sykulski[3], *Fellow, IEEE*

[1]School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China
[2]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, 27100 Pavia, Italy
[3]Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.

This paper focuses on resolving the storage issue of correlation matrices generated by kriging surrogate models in the context of electromagnetic optimization problems with many design variables and multiple objectives. The suggested-improved kriging approach incorporating a direct algorithm is able to maintain memory requirements at a nearly constant level while offering high efficiency of searching for a global optimum. The feasibility and efficiency of this proposed methodology are demonstrated using an example of a classic two-variable analytic function and a new proposed benchmark TEAM multi-objective Pareto optimization problem.

***Index Terms*— Correlation direct algorithm, hybrid kriging, kriging surrogate model, matrices, multi-objective Pareto optimization.**

I. Introduction

KRIGING, as a type of regression model, is able to predict response surface of the objective function through exploiting the spatial correlation of data which based only on limited information [1]–[3]. However, it was found that large-scale tasks—multi-objective and employing many design variables—may lead to a "combinatorial explosion" when all the required correlation matrices are established between the sample points and the design vectors. The partitioning scheme [4] of the correlation matrices, splitting them into manageable sizes, can mitigate to some extent the burden of storing this massive amount of data, but sacrifices may need to be made in terms of computing efficiency at each iteration to achieve more available physical memory. Therefore, a more efficient method capable of removing this bottleneck is sought.

II. Kriging With Lipschitzian Optimized EI

Kriging can exploit the spatial correlation of data to predict the shape of the objective function and search for the global optimum using limited information. Kriging is defined as

$$\hat{y}(x^i) = \sum_{k=1}^{m} \beta_k f_k(x^i) + \varepsilon(x^i) \tag{1}$$

where the sum $\sum_{k=1}^{m} \beta_k f_k(x^i)$, being a linear combination of the values of initial sampled points x^i, may be viewed as a global approximation of the true function. The coefficients β_k are regression parameters and ε are an additive Gaussian noise, representing uncertainty. To be interpolating, the Gaussian distribution $\varepsilon(x^i)$ must be $N(0, \sigma^2)$, with σ^2 to be determined. Two design vectors x^i and x^j, close to each other in the design space, may be expected to have their corresponding objective function values similar. This is modeled statistically by assuming that the errors $\varepsilon(x^i)$ and $\varepsilon(x^j)$ are correlated

$$R(\varepsilon(x^i), \quad \varepsilon(x^j)) = \prod_{k=1}^{n} e^{-\theta_k \left| x_k^i - x_k^j \right|^{p_k}} \tag{2}$$

where θ_k determines how fast the correlation between design vectors drops away in the kth coordinate direction, while p_k determines the smoothness of the function in this direction.

The prediction made by the kriging surrogate model can be viewed as a Gaussian process γ, while a number of updating schemes can be adopted in this process. This allows for the following concept of improvement to be defined: for a single objective to be minimized, the improvement may be measured by comparing the value realized by the objective function with the current minimum of the prediction. This is written as [5]

$$I(x) = \max(f_{\min} - \gamma(x), 0). \tag{3}$$

The expected improvement (EI) (see [5], [6]) may be found by integrating over the likelihood of achieving it, which is given by the normal density function. The maximum of EI indicates the position where the new sampling point should be selected. The EI and the standard error are defined as

$$\mathrm{EIF}[I(x)] = \begin{cases} (f_{\min} - \hat{y}(x))\Phi\left(\frac{f_{\min}-\hat{y}(x)}{s(x)}\right) & \\ \qquad + s(x)\phi\left(\frac{f_{\min}-\hat{y}(x)}{s(x)}\right), & s(x) > 0 \\ 0, & s(x) = 0 \end{cases} \tag{4}$$

$$s(x) = \sqrt{\hat{\sigma}^2\left[1 - r^T R^{-1} r + \frac{(1 - r^T R^{-1} r)}{1^T R^{-1} 1}\right]} \tag{5}$$

where $\hat{y}(x)$ is the value of the objective function predicted by kriging; $f_{\min}$—the minimum of y for the existing samples; $s(x)$ is the root-mean-square error produced by kriging, which contains the correlation R (the correlation matrices between x and the existing sampling points) and r (the correlation between x and the other unknown points); and ϕ are the

Manuscript received June 27, 2017; revised October 20, 2017; accepted October 29, 2017. Corresponding author: S. Xiao (e-mail: xiaosong@swjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMAG.2017.2771561

normal density and normal distribution functions, respectively. Along with the increasing number of sampling points selected by kriging throughout the iterative process, the amount of data produced by the correlation matrices in (2) accumulates constantly, which may become problematic especially when dealing with large-scale multi-variable problems. The known sampling points chosen during the iterative process only take account of very limited parts of the full design vectors, thus the correlation matrices between only the existing sampling points are unlikely to cause memory problems. However, the correlation matrices between the known points and all the design vectors, in the whole design space, may become incredibly large and keep increasing. This paper focuses on a methodology to mitigate the burden of data storage in kriging.

The "DIRECT optimization" approach [7], motivated by a revised Lipschitzian algorithm, is able to address difficult global optimization problems with constraints. It only requires a decision based on available information where to search next. The Lipschitzian optimization process can be defined as

$$|f(x) - f(x')| \le \alpha |x - x'| \quad \forall x,\ x' \in M. \tag{6}$$

If the function f is Lipschitz continuous, with a constant α, then this information can be used to seek the minimum of f iteratively. DIRECT begins the optimization by transforming the domain of the design space into the unit hyper-cube. The center of the design space is c_1, the optimization is initialized by finding the $f(c_1)$. The next step is to divide this hyper-cube by evaluating the function at the points $c_1 \pm \delta_i e_i$, $i = 1, \ldots, N$

$$\omega_j = \min\{f(c_1 + \delta_i e_i),\quad f(c_1 - \delta_i e_i)\},\ \ 1 < i < N \tag{7}$$

where δ is one third the side-length of the hyper-cube and e_i is the ith unit vector. The algorithm triggers a loop of identifying potentially optimal hyper-rectangles, dividing them suitably, and sampling at their centers. DIRECT determines which rectangles are potentially optimal and should thus be divided, by searching locally and globally and applying the criteria

$$f(c_j) - \widehat{K} d_j \le f(c_i) - \widehat{K} d_i \quad \forall i \tag{8}$$

$$f(c_j) - \widehat{K} d_j \le f_{\min} - \epsilon |f_{\min}| \tag{9}$$

where $\epsilon > 0$ and $f_{\min}$ is the current best function value; a hyper-rectangle j is said to be potentially optimal, if there exists a certain constant $\widehat{K}$; c_j is the center of the hyper-rectangle j, and d_j represents the measurement for this rectangle. The criteria are used to search optimal hyper-rectangles to converge to the optimum based on the initialization of DIRECT.

The DIRECT algorithm is utilized here to assist kriging in finding the next sampling point with an optimal value of EI [6], rather than constructing a complete EI over the whole design space using very large correlation matrices. This combination of kriging and the direct algorithm constitutes the main novelty of the proposed approach. As a number of sampling points selected by kriging increases, the amount of data produced by the improved algorithm remains nearly constant. The optimizing procedures for kriging and kriging with Lipschitzian optimized EI are shown in Fig. 1, the latter referred to as "hybrid kriging."

III. Numerical Experiments

To verify the advantages of the proposed hybrid kriging methodology, a two-variable ($n = 2$) analytical test function

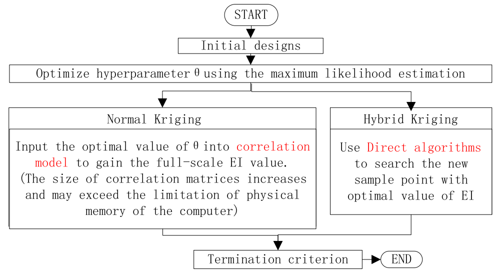

Fig. 1. Decision-making chart for normal kriging and modified kriging.

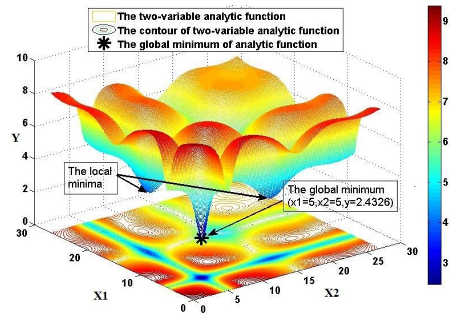

Fig. 2. Two-variable analytical test function.

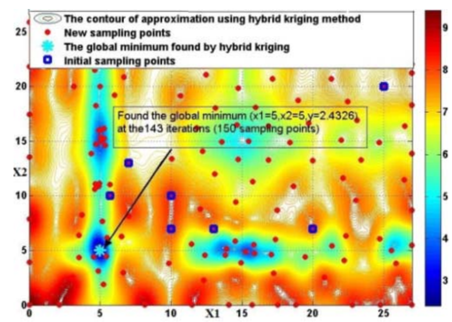

Fig. 3. Highly efficient approximation of hybrid kriging (fewer iterations).

of Fig. 2, with one global minimum and several local minima, has been attempted. The analytical function [8] is defined as

$$f(x) = 10 - \sum_{i=1}^{n} \left[\frac{3.5}{1 + (x_i - 5)^2} + \frac{2.2}{1 + (x_i - 15)^2/10} + \frac{1.2}{1 + (x_i - 25)^2/30} \right] \tag{10}$$

for $0 \le x_i \le 27$. The hybrid kriging located the global minimum after 143 iterations (Fig. 3) and is more than twice as efficient as the kriging assisted EI requiring 324 iterations (Fig. 4) [6].

More significantly, however, the peak memory occupied by the hybrid kriging at each iteration is maintained at a nearly constant level, as presented in Fig. 5, whereas the memory required by the kriging with EI increases linearly throughout the optimization process. On the other hand, the computing times, simultaneously monitored, show similarity for both the hybrid kriging and the normal kriging (see Fig. 6).

IV. Proposed Benchmark TEAM Optimization Problem

The normal kriging, with different sampling strategies for balancing exploration and exploitation, was previously applied to the benchmark TEAM 22 and 25 problems [9]. The results were good but the memory issue clearly visible. The hybrid

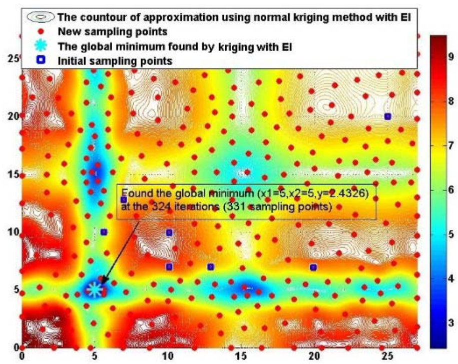

Fig. 4. Kriging assisted EI [6] with more iterations.

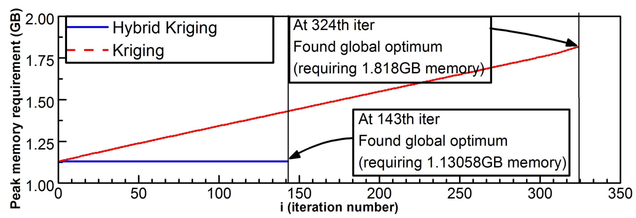

Fig. 5. Monitoring of peak memory requirements of each iteration.

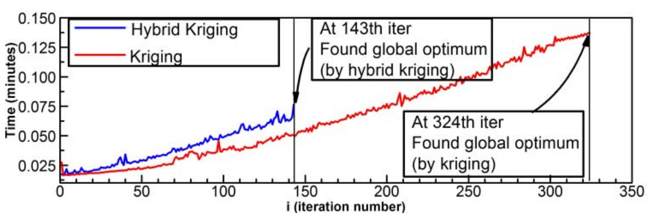

Fig. 6. Monitoring of computing times of each iteration.

kriging, with the burden of memory accumulation removed, has now been applied to the proposed new TEAM problem, with ten variables, details of which may be found in [11]. To summarize, an air-cored single-layer solenoid made up of 20 coils carries a certain current. The target is to find the optimal distribution of the 20 radii that yields the prescribed flux density in a specified region along the solenoid axis. The two goals are to minimize the discrepancy between the prescribed $((B_0(z_q))$ and the actual $(B(z_q, r(\xi_l)))$ field along the solenoid axis and to minimize the field sensitivity with respect to perturbations in the solenoid radii. Hence, the objectives are

$$f_1(r) = \max_{q=1,np} |B(z_q, r(\xi_l)) - B_0(z_q)|, \quad l = 1,\ n_t \tag{11}$$

$$f_2(r) = \max_{q=1,np} [\|B^+ - B(r(\xi_l))\| + \|B(r(\xi_l)) - B^-\|] \tag{12}$$

where B^+ and B^- are the flux density values computed after an expansion, or a contraction, of all radii with respect to the unperturbed configuration. A scalarizing method [10] has been applied to assist the hybrid kriging to combine the multiple objectives using a weighted sum (the weights ω_i are set to 1)

$$\text{Minimize } f(x) = \sum_{i=1}^{M} \omega_i \hat{f}_i(x) \ (M = 2). \tag{13}$$

For an assumed current of 6 A, Fig. 7 shows the objective function trajectory of the sampling points, obtained by hybrid kriging when equal weights of $w_1 = 0.5$ and $w_2 = 0.5$ are placed on both $f_1(r)$ and $f_2(r)$. For a randomly chosen initial point the hybrid kriging required 201 sampling points. The best results a_1 and b_1, for the minima of f_1 and f_2, respectively, are depicted in Fig. 8, while point c_1 denotes the minimum of the objective function among all sampling points. Throughout the iterative search, the minimum of the objective function is traced at the 190th iteration, shown in Fig. 8. The test was terminated manually after 200 iterations for better clarity; it is recommended that ultimately the termination criterion may be formulated so that when the EI of the sampling points declines at a specific value, the hybrid kriging predictor will be stopped.

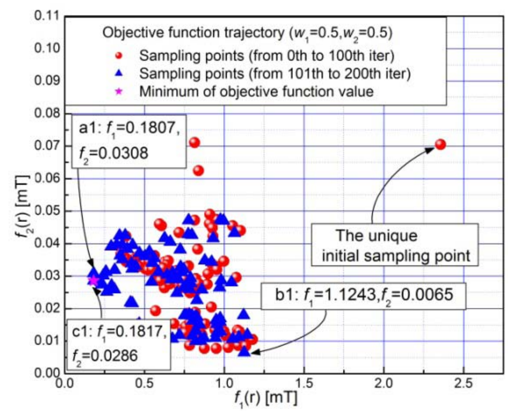

Fig. 7. Objective function trajectory for $(f_1(r)$ and $f_2(r)$, $(w_1 = 0.5$ and $w_2 = 0.5)$.

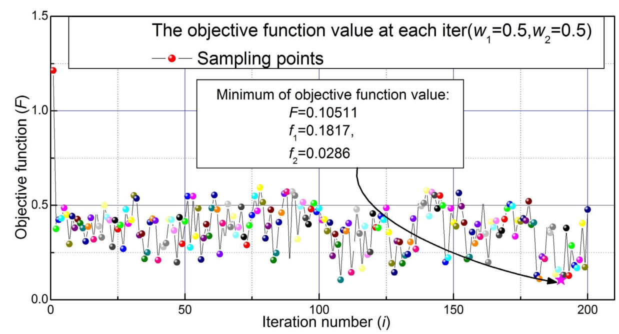

Fig. 8. History of the iterative optimization process ($w_1 = 0.5$ and $w_2 = 0.5$).

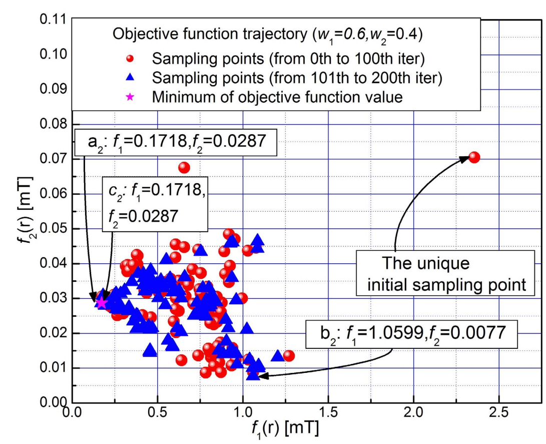

Fig. 9. Objective function trajectory for $f_1(r)$ and $f_2(r)$, $(w_1 = 0.6$ and $w_2 = 0.4)$.

To understand better the impact on the optimal results of applying different weights to the objectives f_1 and f_2, more emphasis was placed on the minimization of the discrepancy

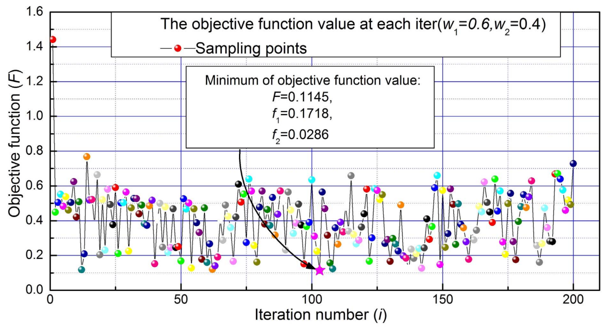

Fig. 10. History of the iterative optimization process ($w_1 = 0.6$ and $w_2 = 0.4$).

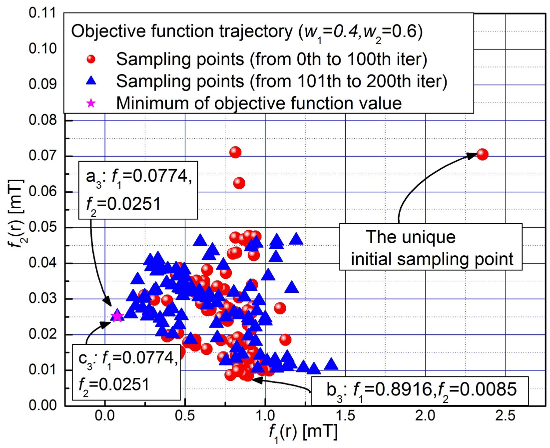

Fig. 11. Objective function trajectory for $f_1(r)$and$f_2(r)$, ($w_1 = 0.4$ and $w_2 = 0.6$).

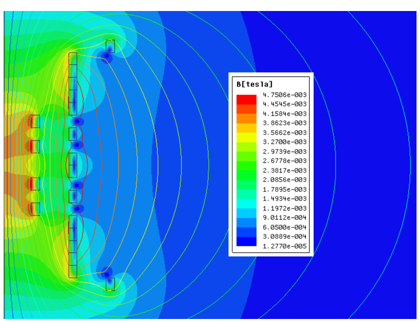

Fig. 12. Optimal solution simulated by finite element method ($w_1 = 0.5$ and $w_2 = 0.5$).

between the actual field along the solenoid axis and the prescribed value, rather than the reduction of the sensitivity with respect to variations of the coil radii. The combination of weights on f_1 and f_2 was set to $w_1 = 0.6$ and $w_2 = 0.4$, respectively. As shown in Fig. 9, the point c_2 found after 103 iterations (Fig. 10) dominates the minimum value of the objective function, as well as the minimum of f_2. Compared with the results obtained using equal weights, a better optimal value of f_1 is achieved while maintaining the optimizing effect in terms of f_2.

When $w_1 = 0.4$ and $w_2 = 0.6$, the sampling point c_3 arrived at the 195th iteration dominates the minimum of the objective function f_1 as well as the minimum of f_2 (Fig. 11). When more weight is put on f_2 the hybrid kriging makes a better attempt at searching for the minimum; consequently, the tests have demonstrated the importance of appropriate choice of the weights in terms of the ability of the optimizer to achieve the ideal solution (global minimum) while doing it efficiently (fewer sampling points). Unfortunately, for

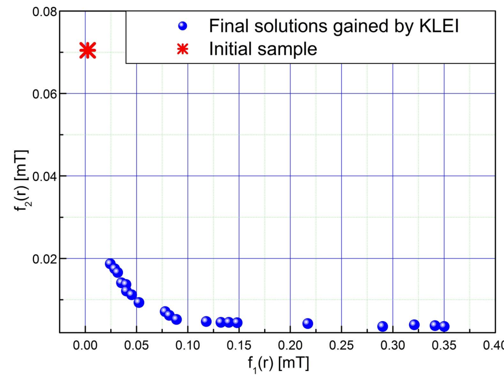

Fig. 13. Pareto front for the bi-objective optimization.

multi-objective problems, the choice of weights is usually problem dependent.

A typical optimal field solution for one of the cases (with equal weights) is shown in Fig. 12 and a Pareto front is shown in Fig. 13.

In the aforementioned tests the following critical parameters were assumed: in (2), θ_k varies from 0.1 to 30 (initially 2); the number of variables k is set as 10; the parameter P_k is set as 2; the number of hyper-cubes i in (7) is set to 100. The 20 radii of the coils in (11) and (12) vary between 0.001 and 0.0145 m. The starting point was chosen randomly.

V. Conclusion

A novel Lipschitzian Optimized EI ("hybrid") kriging model has been proposed to resolve the storage issue of accumulating data during normal kriging. The new algorithm outperforms previous models in terms of efficiency. A proposed benchmark TEAM problem with ten design variables for the multi-objective Pareto optimization of electromagnetic devices has been utilized to verify the feasibility and efficiency of the proposed kriging. The importance of selecting appropriate weights for the auxiliary objectives has been emphasized.

References

[1] B. Xia, S. Hong, K. Choi, and C. S. Koh, "Optimal design of winding transposition of power transformer using adaptive co-Kriging surrogate model," *IEEE Trans. Magn.*, vol. 53, no. 6, Jun. 2017, Art. no. 7203904.

[2] B. Xia, Z. Ren, and C. Koh, "A novel reliability-based optimal design of electromagnetic devices based on adaptive dynamic Taylor Kriging," *IEEE Trans. Magn.*, vol. 53, no. 6, Jun. 2017, Art. no. 7201504.

[3] T. V. Tran, F. Moussouni, S. Brisset, and P. Brochet, "Adapted output space-mapping technique for a bi-objective optimization," *IEEE Trans. Magn.*, vol. 46, no. 8, pp. 2990–2993, Aug. 2010.

[4] S. Xiao, M. Rotaru, and J. K. Sykulski, "Correlation matrices in Kriging assisted optimisation of electromagnetic devices," *IET Sci., Meas. Technol.*, vol. 9, no. 2, pp. 189–196, 2015.

[5] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.

[6] S. Xiao, M. Rotaru, and J. K. Sykulski, "Adaptive weighted expected improvement with rewards approach in Kriging assisted electromagnetic design," *IEEE Trans. Magn.*, vol. 49, no. 5, pp. 2057–2060, May 2013.

[7] D. R. Jones, C. D. Pertunnen, and B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *J. Optim. Theory Appl.*, vol. 79, no. 1, pp. 157–181, 1993.

[8] Z. Ren, M.-T. Pham, M. Song, D.-H. Kim, and C. S. Koh, "A robust global optimization algorithm of electromagnetic devices utilizing gradient index and multi-objective optimization method," *IEEE Trans. Magn.*, vol. 47, no. 5, pp. 1254–1257, May 2011.

[9] *International Compumag Society*. Accessed: Sep. 21, 2017. [Online]. Available: http://www.compumag.org/jsite/team.html

[10] P. Di Barba, *Multiobjective Shape Design in Electricity and Magnetism*. Berlin, Germany: Springer, 2010.

[11] P. Di Barba, M. E. Mognaschi, D. A. Lowther, and J. K. Sykulski, "A benchmark TEAM problem for multi-objective Pareto optimization of electromagnetic devices," *IEEE Trans. Magn.*, to be published.