*PC6-19

A Scalarizing One-Stage Algorithm for Efficient Multi-Objective Optimization

Glenn I. Hawe*,† and Jan K. Sykulski†
*Vector Fields, Ltd., 24 Bankside, Kidlington, Oxford, U.K. OX5 1JE
†School of Electronics and Computer Science, University of Southampton, Southampton, U.K. SO17 1BJ
Email: glenn.hawe@vectorfields.co.uk, jks@soton.ac.uk

***Abstract*— A novel kriging-assisted algorithm is proposed for computationally expensive multi-objective optimization problems, such as those which arise in electromagnetic design. The algorithm combines the multiple objectives into a single objective, which it then optimizes using a one-stage method from single-objective optimization. Its efficiency is demonstrated by comparison to a random search on a difficult test function.**

I. Introduction

One popular method of reducing the high computational cost of evaluating objective functions in electromagnetic optimal design problems is the use of surrogate models, such as kriging [1].

Surrogate-model assisted single-objective optimization algorithms may be categorized into 'two-stage' and 'one-stage' varieties [2]. At each iteration of a two-stage algorithm, a surrogate model is constructed from the sampled points (the first stage), and then this model is used to determine where to sample next (the second stage), e.g. [3]. On the other hand, one-stage algorithms choose where to sample next by making hypotheses about the location of the global minimum, and determining the credibility of surrogate models which would pass through each hypothesized optimum and the sampled points, e.g. [4]. The point chosen to be sampled is the hypothesized point which has the most credible surrogate model passing through it. Results on test functions show one-stage methods to be extremely efficient.

One popular technique for solving multi-objective optimization problems (MOOPs) is to combine the multiple objectives into a single objective [5] and then optimize this, e. g. [6]. This paper proposes a novel multi-objective algorithm, which uses a one-stage kriging algorithm to optimize a MOOP, which is scalarized using an augmented Tchbeycheff function [5]. It is believed this is the first time one-stage methodology has been used for multi-objective optimization.

II. One-Stage Kriging Methodology

A brief overview of the one-stage kriging methodology is first given. Suppose that objective function f is a function of an n-dimensional design vector, and suppose N design vectors, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)}$ have been evaluated. Given the objective function values of these N design vectors, a hypothesis is made about the value of the objective function at the global minimum of f. Specifically, the global minimum is hypothesized to have an objective function value f^*. Then defining the Gaussian correlation function R (which expresses how two design vectors $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ are correlated) as

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

(where θ_k determines how rapid the correlation is lost in the k^{th} design variable, and p_k determines the 'smoothness' of the function in the k^{th} design variable), the $N \times 1$ correlation vector $\mathbf{r}$ as

$$\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}^{(1)}), R(\mathbf{x}, \mathbf{x}^{(2)}), \ldots, R(\mathbf{x}, \mathbf{x}^{(N)})]^{\mathrm{T}}, \tag{2}$$

the $N \times N$ correlation matrix $\mathbf{R}$ as the matrix whose $i-j^{\text{th}}$ entry is $R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, the $N \times 1$ vector $\mathbf{y}$ as the vector filled with the objective function values of the sampled points,

$$\mathbf{y} = [f(\mathbf{x}^{(1)}), f(\mathbf{x}^{(2)}), \ldots, f(\mathbf{x}^{(N)})]^{\mathrm{T}} \tag{3}$$

and $\mathbf{1}$ as the $N \times 1$ vector filled with ones, then for any design vector $\mathbf{x}^*$, the likelihood of the N examples conditional upon the surface passing through $(\mathbf{x}^*, f^*)$ is [2]:

$$\frac{1}{(2\pi)^{N/2}(\sigma^2)^{N/2}|\mathbf{C}|^{1/2}} \exp\left(\frac{-(\mathbf{y}-\mathbf{m})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y}-\mathbf{m})}{2\sigma^2}\right) \tag{4}$$

where

$$\mathbf{m} = \mathbf{1}\beta + \mathbf{r}(\mathbf{x}^*)(f^* - \beta) \tag{5}$$

$$\mathbf{C} = \mathbf{R} - \mathbf{r}(\mathbf{x}^*)\mathbf{r}^T(\mathbf{x}^*) \tag{6}$$

are the conditional mean and conditional correlation matrix respectively. The next design vector to be evaluated is the one which maximizes the conditional likelihood in Eq. (4) (which itself is maximized over $\theta = [\theta_1, \theta_2, \ldots, \theta_n]$, $\mathbf{p} = [p_1, p_2, \ldots, p_n]$, β and σ^2 for each $\mathbf{x}^*$). This design vector is the one which, if it had objective function value f^*, would yield the most credible kriging model interpolating it and the N design vectors already observed. Note that setting the derivatives of Eq. (4) with respect to σ^2 and β equal to zero and rearranging, it is found that:

$$\sigma^2 = \frac{(\mathbf{y}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{y}-\mathbf{m})}{N} \tag{7}$$

$$\beta = \frac{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y} + f^*\mathbf{r}^T\mathbf{C}^{-1}\mathbf{r} - \mathbf{y}\mathbf{C}^{-1}\mathbf{r} - f^*\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{r}}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1} - 2\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{r} + \mathbf{r}^T\mathbf{C}^{-1}\mathbf{r}}. \tag{8}$$

and so Eq. (4) only needs to be maximized over θ and $\mathbf{p}$. The following Section proposes a simple way of extending the one-stage method for use in multi-objective optimization.

III. Scalarizing One-Stage Algorithm for Multi-Objective Optimization

After normalizing the $n_{\rm obj}$ objectives of the MOOP using either known or estimated limits of the objective function space, so that each objective function lies within the range [0,1], they are combined using the augmented Tchebycheff function [5]:

$$f_\lambda(\mathbf{x}) = \max_{j=1}^{n_{\rm obj}}(\lambda_j f_j(\mathbf{x})) + \rho \sum_{j=1}^{n_{\rm obj}} \lambda_j f_j(\mathbf{x}) \qquad (9)$$

where ρ is a small positive value set (arbitrarily) to 0.05, and $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_{n_{\rm obj}}]$ is a normalized weight vector.

A Latin Hypercube experimental design [7] of size $5n$ is initially carried out. This is used to initialize $n_{\rm w} (\geq n_{\rm obj} + 1)$ *independent* optimization searches, where each search:

- uses a different weighting vector λ, so as to converge towards a different region of the Pareto-optimal front,
- is (arbitrarily) $i_{\max} = 5n$ iterations in length, and
- uses the one-stage kriging method described in Section II with a target $f^* = \kappa(i) f_{\min}$ at iteration i, where $f_{\min}$ is the current minimum value of f_λ and $\kappa(i) = 0.95(0.5 + \frac{i}{2i_{\max}})$ is a scaling factor used to make the search less exploratory (more exploitative) as the iterations proceed.

In the initial $n_{\rm obj} + 1$ searches, $n_{\rm obj}$ extreme weighting vectors $(1 - \epsilon, \frac{\epsilon}{n_{\rm obj}-1}, \frac{\epsilon}{n_{\rm obj}-1}, ..., \frac{\epsilon}{n_{\rm obj}-1}), (\frac{\epsilon}{n_{\rm obj}-1}, 1 - \epsilon, \frac{\epsilon}{n_{\rm obj}-1}, ..., \frac{\epsilon}{n_{\rm obj}-1}), ..., (\frac{\epsilon}{n_{\rm obj}-1}, \frac{\epsilon}{n_{\rm obj}-1}, \frac{\epsilon}{n_{\rm obj}-1}, ..., 1 - \epsilon)$ (where $|\epsilon| \ll 1$) which each heavily favor only one objective, are used, as well as the weighting vector which places equal emphasis on each objective, $(\frac{1}{n_{\rm obj}}, \frac{1}{n_{\rm obj}}, ..., \frac{1}{n_{\rm obj}})$. After these first $n_{\rm obj} + 1$ searches, a further $n_{\rm w} - n_{\rm obj} - 1$ searches are then performed, with the weight vector set each time so that the value of its components is the average of the corresponding components of the two weight vectors which yielded points which bound the emptiest region of objective function space. Note that this procedure means the algorithm has a fixed number of iterations, $n_{\rm iter} = 5(n_{\rm w} + 1)n$.

It should be emphasized that each optimization search is completely independent of the others: the results from one search are not used in another. This has several benefits:

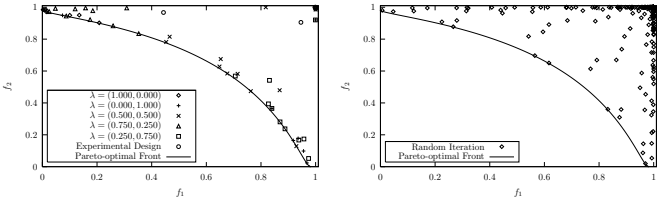
1) The algorithm may be easily parallelized.
2) The conditional correlation matrix $\mathbf{C}$ which is to be inverted to calculate the credibility never exceeds a size of $10n \times 10n$. This is extremely important in keeping the computational cost of the one-stage approach as low as possible.
3) For any given weight vector, the iterations tend to concentrate more and more around one particular region of design variable space. By ignoring the iterations of other searches (using different weight vectors), the degree to which $\mathbf{C}$ becomes ill-conditioned is severely reduced.

IV. Results

The algorithm was tested on a range of test functions; due to lack of space, results are given for one only, VLMOP2 [8]:

$$\begin{aligned} \text{Minimize} \quad f_1(\mathbf{x}) &= 1 - \exp\big(-\sum_{i=1}^{n}(x_i - \frac{1}{\sqrt{n}})^2\big) \\ \text{and} \quad f_2(\mathbf{x}) &= 1 - \exp\big(-\sum_{i=1}^{n}(x_i + \frac{1}{\sqrt{n}})^2\big) \\ \text{with} \quad x_i &\in [-4, 4] \end{aligned}$$

where $n = 2$. $n_{\rm w} = 5$ different weighting vectors were used in the scalarizing one-stage algorithm, giving $n_{\rm iter} = 60$ iterations in total. These iterations are shown in Fig. 1 (a) below (each iteration being identified by the weight vector used). For comparison, a random search of 500 iterations was also performed, and these iterations are shown in Fig. 1 (b). As can be seen, the scalarizing one-stage algorithm is much more efficient at locating solutions close to the Pareto-optimal front. Similar results were observed in other test functions.



(a) 60 iterations of scalarizing one-stage algorithm.

(b) 500 iterations of random search algorithm.

Fig. 1. Results on VLMOP2 test function.

V. Conclusion

A novel algorithm has been proposed which uses, for the first time, a one-stage methodology for multi-objective optimization. It performed efficiently on a difficult test problem; its performance in multi-objective electromagnetic design optimization will be discussed in the full paper.

References

[1] L. Lebensztajn, C. A. R. Marretto, M. C. Costa and J-L. Coulomb, "Kriging: a useful tool for electromagnetic devices optimization", *IEEE Transactions on Magnetics*, vol. 40, no. 2, pp. 1196–1199, 2004.

[2] D. R. Jones, "A Taxonomy of Global Optimization Methods Based on Response Surfaces", *Journal of Global Optimization*, vol. 21, pp. 345–383, 2001.

[3] D. R. Jones, M. Schonlau and W. J. Welch, "Efficient Global Optimization of Expensive Black-Box Functions", *Journal of Global Optimization*, vol. 13, pp. 455–492, 1998.

[4] H. M. Gutmann, "A Radial Basis Function Method for Global Optimization", *Journal of Global Optimization*, vol. 19, pp. 201–227, 2001.

[5] K. M. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, 1999.

[6] J. Knowles, "ParEGO: A Hybrid Algorithm With On-Line Landscape Approximation for Expensive Multiobjective Optimization Problems", *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 50–66, 2006.

[7] M. D. McKay, W. J. Conover and R. J. Beckmantitle, "A comprison of three methods for selecting values of input variables in the analysis of output from a computer code" *Technometrics*, vol. 21, pp. 239-245, 1979.

[8] A. Van. Veldhuizen, "Multiobjective Evolutionary Algorithms: Classifications, Analyzes, and New Innovations", PhD. Dissertation, Air Force Institute of Technology, Wright-Patterson AFB, Jun. 1999.