

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

Glenn I. Hawe<sup>1,2</sup> and Jan K. Sykulski<sup>2</sup>

<sup>1</sup>Vector Fields Ltd., Oxford OX5 1JE, U.K.

<sup>2</sup>University of Southampton, Southampton SO17 1BJ, U.K.

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. Inequality and equality constraint handling techniques are included, as is a method for dealing with failed iterations. Its efficiency is demonstrated on the constrained optimization of a pair of Helmholtz coils.

**Index Terms**—Kriging, optimization methods.

## 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 and robust.

One popular technique for solving multi-objective optimization problems (MOOPs) is to combine the multiple objectives into a single objective [5] and then use a method from single-objective optimization to optimize this, e.g., [6]. This paper proposes a scalarizing multi-objective optimization algorithm, which uses a one-stage methodology for the selection of points; it is believed this is the first time one-stage methodology has been used for multi-objective optimization.

## II. ONE-STAGE KRIGING METHODOLOGY FOR SINGLE-OBJECTIVE OPTIMIZATION

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,  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$  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

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

(where  $\theta_k$  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

$$\mathbf{r}(\mathbf{x}) = \left[ R(\mathbf{x}, \mathbf{x}^{(1)}), R(\mathbf{x}, \mathbf{x}^{(2)}), \dots, R(\mathbf{x}, \mathbf{x}^{(N)}) \right]^T \quad (2)$$

the  $N \times N$  correlation matrix  $\mathbf{R}$  as the matrix whose  $i - j$ 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,

$$\mathbf{y} = \left[ f(\mathbf{x}^{(1)}), f(\mathbf{x}^{(2)}), \dots, f(\mathbf{x}^{(N)}) \right]^T \quad (3)$$

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]

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

where

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

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

are the conditional mean and conditional correlation matrix, respectively. The next design vector to be evaluated is that which maximizes the conditional likelihood in (4) (which itself is maximized over  $\theta$ ,  $\mathbf{p}$ ,  $\beta$  and  $\sigma^2$  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 (4) with

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

$$\beta = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{y} + f^* \mathbf{r}^T \mathbf{C}^{-1} \mathbf{r} - \mathbf{y} \mathbf{C}^{-1} \mathbf{r} - f^* \mathbf{1}^T \mathbf{C}^{-1} \mathbf{r}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} - 2\mathbf{1}^T \mathbf{C}^{-1} \mathbf{r} + \mathbf{r}^T \mathbf{C}^{-1} \mathbf{r}} \quad (8)$$

and so (4) only needs to be maximized over  $\theta$  and  $\mathbf{p}$ .

The concept of the one-stage methodology is illustrated in Fig. 1, on a test function which has been sampled at five points.

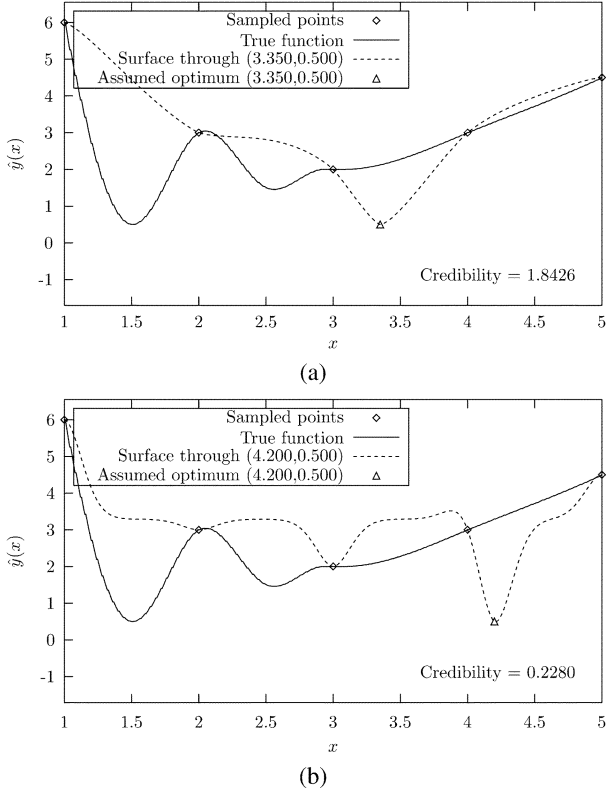


Fig. 1. Illustration of the one-stage methodology on a test function. (a) Hypothesis 1:  $x^* = 3.35$ . (b) Hypothesis 2:  $x^* = 4.20$ .

Surfaces through two different hypothesized optima are plotted (the value of  $f^*$  is taken to be 0.5 in each case). Intuitively, the surface arising from the first hypothesis looks more plausible than the surface arising from the second hypothesis; this indeed is supported by their respective credibility (conditional likelihood) values. It should be noted that while in this case the value of the objective function at the true minimum (0.5) was used as a target, a value based on the existing objective function values may be used if the value of  $f^*$  is not known. For example, suitable target values may be constructed using

$$f^* = s_{\min} - \alpha(f_{\max} - f_{\min}) \quad (9)$$

with different values of  $\alpha$ , where  $f_{\max}$  and  $f_{\min}$  are the current maximum and minimum objective function values, respectively, and  $s_{\min}$  is the minimum of a kriging surface fit through the sampled points alone.

The use of one-stage methods in the literature has so far been rather limited [2], [4], however, there has been some revived interest [7], and their application has also recently spread to optimal electromagnetic design [8].

### III. SCALARIZING ONE-STAGE ALGORITHM FOR CONSTRAINED MULTI-OBJECTIVE OPTIMIZATION

The general problem under consideration is the constrained multi-objective optimization problem:

$$\begin{aligned} & \text{Minimize} && f_i(\mathbf{x}) \quad i = 1, 2, \dots, M; \\ & \text{subject to} && g_i(\mathbf{x}) \geq 0 \quad i = 1, 2, \dots, J; \\ & && h_i(\mathbf{x}) = 0 \quad i = 1, 2, \dots, K; \\ & && x_i^{(L)} \leq x_i \leq x_i^{(U)} \quad i = 1, 2, \dots, n. \end{aligned} \quad (10)$$

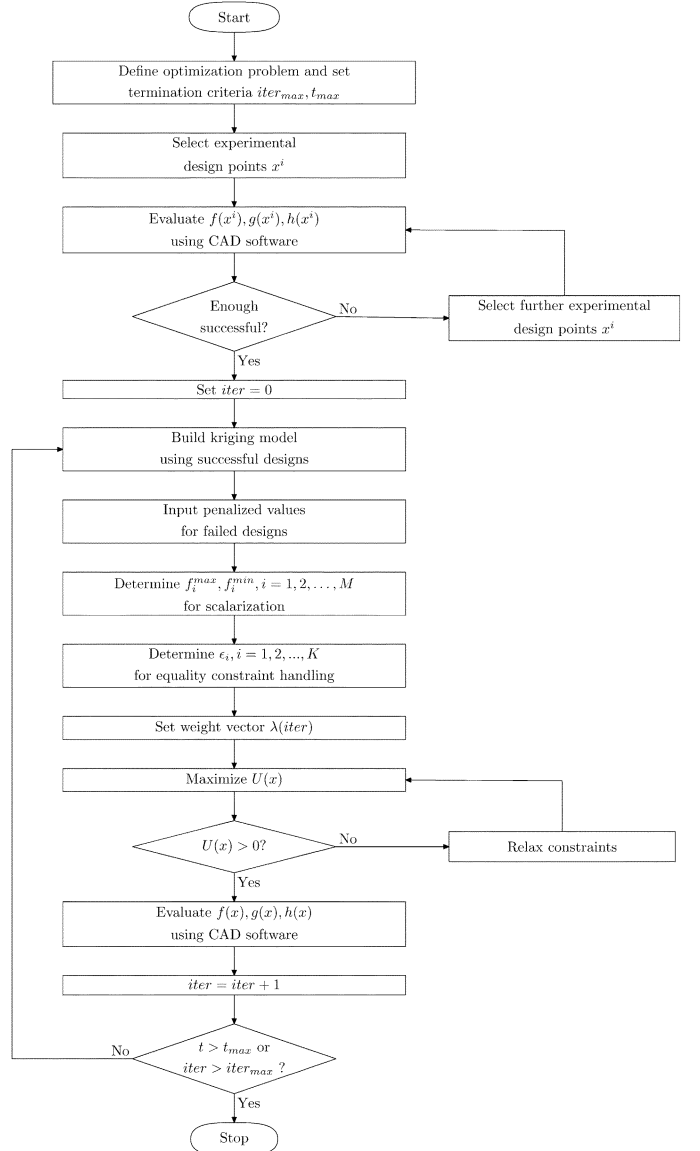


Fig. 2. Overview of the scalarizing one-stage algorithm.

An overview of the algorithm proposed for solving (10) is shown in Fig. 2. In brief, the algorithm proceeds as follows.

- 1) Initialization by experimental design.
- 2) Penalization of failed designs.
- 3) Transformation of the MOOP to a single-objective optimization problem (SOOP).
- 4) Selection of a design vector for evaluation using a one-stage methodology, with constraint handling methods.

The algorithm iterates over stages 2–4, until some pre-defined termination criteria is met. The remainder of this section deals with each of these stages in turn.

#### A. Experimental Design

Several modern space-filling experimental designs now exist, and the choice of which to use is inevitably arbitrary. However, empirical studies (e.g., [9]) suggest that uniform designs tend to give better results. For this reason, a Hammersley sequence [10] was used for the initial selection of design vectors.

The number of examples in the initial design,  $N_{\text{exp}}$ , is also arbitrary, and most studies scale it according to the number of

design variables (e.g., 10  $n$  is used in [3]). In constrained optimization however, it turns out to be quite beneficial to have at least one example in the experimental design which is feasible (i.e., satisfies the constraints), and for this reason the experimental design size is also scaled according to the number of constraints

$$N_{\text{exp}} = 5(n + J + 2K). \quad (11)$$

Note, that in the rare case of insufficient successful designs being found from the experimental design to allow construction of an initial kriging model, further experimental design points are selected using higher order Hammersley sequences.

### B. Dealing With Failed Designs

In optimal electromagnetic design, the process of evaluating objective functions and constraints typically involves building a model, generating a finite element mesh, creating a database, running a solver and then performing some post-processing using CAD software. Each of these stages is prone to failure, meaning that it is not always the case that information will be obtainable from a design selected for evaluation.

Following advice in [11], for each failed design  $\mathbf{x}^*$ , each objective function value  $f_i$  is penalized using a value of

$$f_i(\mathbf{x}^*) = \begin{cases} \hat{f}_i(\mathbf{x}^*) + s(\mathbf{x}^*) & \text{if } f_i \text{ is to be minimized} \\ \hat{f}_i(\mathbf{x}^*) - s(\mathbf{x}^*) & \text{if } f_i \text{ is to be maximized} \end{cases} \quad (12)$$

where  $\hat{f}_i(\mathbf{x}^*)$  is the kriging prediction of  $f_i(\mathbf{x}^*)$  based on the successful designs, and  $s(\mathbf{x}^*)$  is the standard error in this prediction (see e.g., [3] for derivations of these expressions). The kriging predictions  $\hat{g}_i(\mathbf{x}^*)$ ,  $\hat{h}_i(\mathbf{x}^*)$  (again based on successful designs) are input for the constraint vectors of each failed design; penalization of the objective function values alone is deemed sufficient to guide the algorithm away from the regions with high probability of failure.

### C. Scalarization

Once the experimental design stage is finished, and each of the failed designs penalized, the  $M$  objectives of the MOOP are normalized so that each objective function lies within the range [0,1], using

$$\bar{f}_i(\mathbf{x}) = \frac{f_i(\mathbf{x}) - f_i^{\min}}{f_i^{\max} - f_i^{\min}} \quad (13)$$

where  $f_i^{\min}$ ,  $f_i^{\max}$  are the true lower and upper limits of objective  $f_i$ , respectively. If the true limits  $f_i^{\min}$ ,  $f_i^{\max}$  are unknown (as will often be the case), then the maximum and minimum values obtained so far are used. This leads to an ideal objective function value of  $\mathbf{z}^* = (0, 0, \dots, 0)^T$ .

Once normalized, the objectives are combined using the augmented weighted Tchebycheff function [5]

$$f_\lambda(\mathbf{x}) = \max_{j=1}^M (\lambda_j |\bar{f}_j(\mathbf{x}) - z_j^*|) + \rho \sum_{j=1}^M \lambda_j (|\bar{f}_j(\mathbf{x}) - z_j^*|) \quad (14)$$

where  $\rho$  is a small positive value set (arbitrarily) to 0.05, and  $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_M]$  is a normalized weight vector (in this implementation, random weighting vectors are used at each iteration). This metric has the advantage that solutions situated on

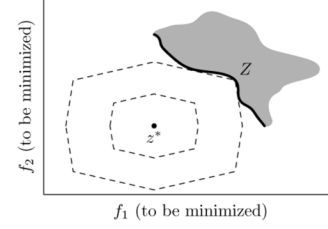


Fig. 3. Contour lines for the weighted augmented Tchebycheff metric.

concave parts of Pareto-optimal fronts may be located, as can be seen from the contours of the Tchebycheff function illustrated schematically for a two-objective problem in Fig. 3. This ability to locate concave parts of Pareto-fronts is something which is not true of scalarizing functions in general.

### D. Selection of Design Vectors and Constraint Handling

With the original MOOP now transformed to a SOOP, the one-stage methodology described in Section II may now be used to select design vectors for evaluation. In this case, as each objective has been normalized, an ideal target value of  $f^* = 0$  is used at each iteration.

Most practical engineering design problems involve a set of constraints, and so to make the algorithm as versatile as possible, constraint handling techniques are included. Each equality constraint  $h_i(\mathbf{x}) = 0$  is transformed to two separate inequality constraints,  $|h_i(\mathbf{x})| \leq \epsilon_i$ , where  $\epsilon_i$  is taken to be a small percentage ( $x\%$ ) of the range of the values of  $h_i$  of the designs so far sampled

$$\epsilon_i = \frac{x}{100} (h_i^{\max} - h_i^{\min}). \quad (15)$$

In this implementation,  $x$  is set to be 5. This results in a MOOP with  $J + 2K$  inequality constraints. These are dealt with by penalizing the value of the utility function  $U$  (which is based on the conditional likelihood in (4)) used to select the design vector to be evaluated, as follows:

$$U(\mathbf{x}) = \begin{cases} \frac{1}{(2\pi)^{N/2} (\sigma^2)^{N/2} |\mathbf{C}|^{1/2}} \exp\left(\frac{-(\mathbf{y}-\mathbf{m})^T \mathbf{C}^{-1} (\mathbf{y}-\mathbf{m})}{2\sigma^2}\right) & \text{if } \hat{g}_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, J \text{ and} \\ & |\hat{h}_i(\mathbf{x})| \leq \epsilon_i, i = 1, 2, \dots, K \\ 0 & \text{if } \exists i \text{ such that } \hat{g}_i(\mathbf{x}) > 0 \text{ or } |\hat{h}_i(\mathbf{x})| > \epsilon_i. \end{cases} \quad (16)$$

The design vector which maximizes  $U$  is then selected for evaluation. This method of dealing with constraints suffers in that it is possible for  $U(\mathbf{x}) = 0 \forall \mathbf{x}$  (as a consequence of at least one constraint being predicted to be violated for every  $\mathbf{x}$ ); when this happens, the constraints are gradually relaxed, until a design vector which yields a non-zero value of  $U$  is found, and this is then selected for evaluation.

### E. Termination Criteria

After evaluation of the design vector which maximizes the utility function  $U$ , it is added to the set of examples. The algorithm then repenalizes the failed designs, using a kriging model

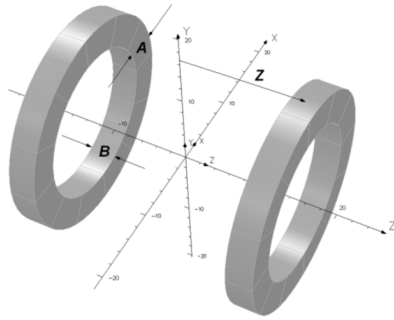


Fig. 4. Pair of Helmholtz coils with design variables, and  $Z$  labeled.

based on the updated set of examples; objectives are then renormalized and rescaled (using a different weighting vector  $\lambda$ ), and another design vector is selected using the updated utility function  $U$ . This procedure continues until one of two possible termination criteria is met: either a maximum time limit  $t_{\max}$  has been exceeded, or a maximum number of iterations  $iter_{\max}$  have been carried out.

#### IV. OPTIMIZATION OF HELMHOLTZ COILS

The algorithm was applied to a pair of Helmholtz coils (with current density  $10^4 \text{ cm}^{-2}$ ), to achieve a uniform central field of 7000 Gauss. The width  $A$ , coaxial length  $B$ , and half  $z$  separation  $Z$  of the coils, as shown in Fig. 4, were allowed to vary within specific ranges. To achieve the goal of a uniform central field of 7000 Gauss, the Legendre polynomial coefficients  $A_{00}, A_{20}, A_{40}, A_{60}$  of the  $z$ -component of the magnetic flux density were calculated over a sphere (of radius 3) centered around the origin. Then, to ensure field uniformity, the magnitude of the error harmonics were minimized, while to achieve the desired central field value, the magnitude of the  $A_{00}$  coefficient was constrained to be equal to 7000

$$\begin{aligned} &\text{Minimize } |A_{20}|, |A_{40}|, |A_{60}| (\text{Error harmonics}) \\ &\text{subject to } |A_{00}| = 7000 \\ &\quad \text{with } 0.5 \leq A \leq 5 \text{ cm} \\ &\quad \quad 0.5 \leq B \leq 8 \text{ cm} \\ &\quad \quad 0.5 \leq Z \leq 15 \text{ cm}. \end{aligned}$$

The algorithm was allowed to run for 65 iterations in total (25 experimental design points, followed by 40 iterations). The variation of the central field value over these iterations is shown in Fig. 5. In all, 34 feasible solutions were located, 12 of which were non-dominated; the positions of these 12 Pareto solutions in objective function space are shown in Fig. 6. By comparison, a random search of 250 iterations produced only 17 feasible solutions (10 of which were non-dominated).

#### V. CONCLUSION

A scalarizing algorithm has been proposed which uses, for the first time, a one-stage methodology for multi-objective optimization, and which includes methods for handling non-linear constraints and dealing with failed designs. Areas of future

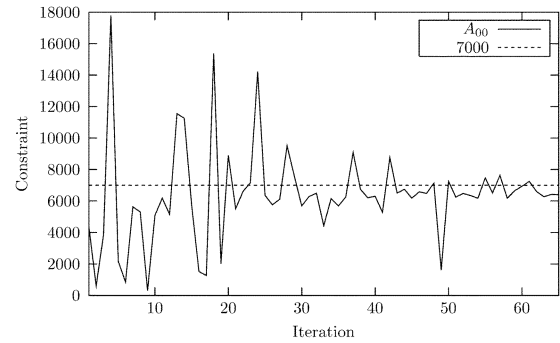


Fig. 5. Variation of  $|A_{00}|$  (central field value in Gauss) over 65 iterations.

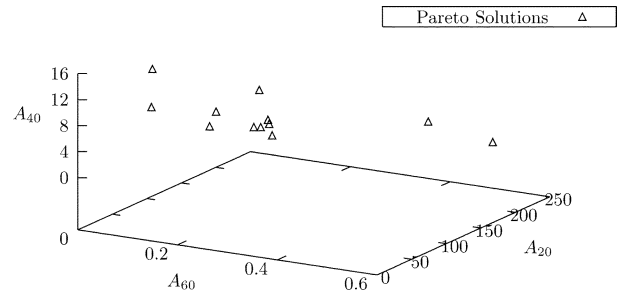


Fig. 6. Positions of Pareto solutions in objective function space.

potential development include the selection of multiple design vectors during each iteration, (using a selection of multiple weighting vectors), and extension to deal with discrete design variables.

#### REFERENCES

- [1] L. Leblensztajn, C. A. R. Marretto, M. C. Costa, and J.-L. Coulomb, "Kriging: A useful tool for electromagnetic devices optimization," *IEEE Trans. Magn.*, vol. 40, no. 2, pt. 2, pp. 1196–1199, Mar. 2004.
- [2] D. R. Jones, "A taxonomy of global optimization methods based on response surfaces," *J. Global Optim.*, vol. 21, pp. 345–383, 2001.
- [3] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, pp. 455–492, 1998.
- [4] H. M. Gutmann, "A radial basis function method for global optimization," *J. Global Optim.*, vol. 19, pp. 201–227, 2001.
- [5] K. M. Miettinen, *Nonlinear Multiobjective Optimization*. Boston, MA: Kluwer Academic, 1999.
- [6] J. Knowles, "ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Trans. Evol. Comput.*, vol. 10, no. 1, pp. 50–66, Feb. 2006.
- [7] R. G. Regis and C. A. Shoemaker, "Improved strategies for radial basis function methods for global optimization," *J. Global Optim.*, vol. 37, no. 1, pp. 113–135, 2007.
- [8] G. I. Hawe and J. K. Sykulski, "A hybrid one-then-two stage algorithm for computationally expensive electromagnetic design optimization," *COMPEL*, vol. 26, no. 2, pp. 236–246, 2007.
- [9] T. W. Simpson, D. K. J. Lin, and W. Chen, "Sampling strategies for computer experiments: Design and analysis," *Int. J. Rel. Appl.*, vol. 2, no. 3, pp. 209–240, 2001.
- [10] J. R. Kalagnanam and U. M. Diwekar, "An efficient sampling technique for off-line quality control," *Technometrics*, vol. 39, no. 2, pp. 131–148, 1997.
- [11] A. I. J. Forrester, A. Sobester, and A. J. Keane, "Optimization with missing data," *Proc. Roy. Soc. A*, vol. 462, pp. 935–945, 2006.