An Enhanced Probability of Improvement Utility Function for Locating Pareto Optimal Solutions

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— **This paper describes a novel utility function for choosing design vectors to evaluate in multi-objective optimization problems which are statistically most probable to be Pareto-optimal, given the points already evaluated. The method is tunable to the number of existing Pareto-optimal solutions that an unevaluated design vector is sought to dominate, is naturally parallelized, and removes any need for combining the multiple objectives into a single objective with a scalarizing function.**

I. Introduction

In [1], the probability of improvement criteria was identified as an effective utility function to use with kriging surrogate models [2] for locating design vectors to evaluate in single-objective optimization. By modeling the prediction of a kriging model as the realization of a Gaussian distribution, with mean $\hat{f}(\mathbf{x})$ and standard error $s(\mathbf{x})$ as given by the kriging model, the probability of an unevaluated design vector $\mathbf{x}$ having an objective function value less than T is

$$P(f(\mathbf{x}) < T) = \Phi\Big(\frac{T - \hat{f}(\mathbf{x})}{s(\mathbf{x})}\Big) \quad (1)$$

where Φ is the normal cumulative distribution function. This is illustrated below in Fig. 1; the probability of the unevaluated design vector x^* having an objective function value less than T is represented by the shaded region.

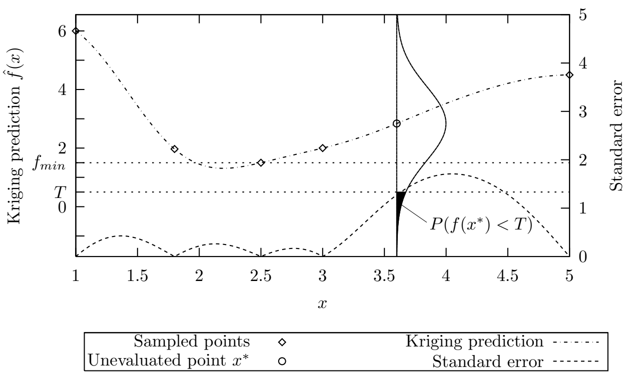

Fig. 1. Uncertainty in objective function value for an unevaluated design vector as predicted by a kriging model.

This method of selecting design vectors to evaluate suffered one drawback however, in that it was sensitive to the level of improvement sought (i.e. the value of T). Two methods were proposed in [1] for overcoming this sensitivity:

1) Evaluate the design vector which maximizes the expectation of the improvement, or
2) Evaluate several design vectors per iteration, each corresponding to a different level of improvement (i.e. a different value of T).

The first method led to the development of the EGO algorithm [3], which has subsequently received significant attention in the literature. However, the second method, referred to as 'enhanced probability of improvement' is the approach recommended in [1]. It has the advantage that it is very robust and easily parallelized, although it suffers in that the different levels of improvement sought in each iteration (i.e. the different values of T used) are arbitrary. This paper extends this enhanced method to the multi-objective case, where it shall be seen that natural levels of improvement exist.

II. Probability of Improvement and Expected Improvement in Multi-Objective Optimization

The concept of improvement in multi-objective optimization has only appeared recently in the literature. First, recall that a solution S_a (with design vector $\mathbf{x}_a$) is said to dominate another solution S_b (with design vector $\mathbf{x}_b$) if and only if S_a is strictly better than S_b in at least one objective, and no worse in all other objectives. The set of Pareto-optimal solutions are then those solutions which are not dominated by any other existing solution. Suppose then, that a set S of N_{par} Pareto-optimal solutions exists, $S = \{S_1, S_2, \ldots, S_{N_{\text{par}}}\}$, after performing an experimental design. Then a design vector $\mathbf{x}$ is said to yield an improvement if it is non-dominated by the solutions in S [4]. This may happen in one of two ways:

1) $\mathbf{x}$ dominates at least one of the solutions in S, or
2) $\mathbf{x}$ does not dominate any solution in S, nor does any solution in S dominate $\mathbf{x}$.

This is shown in Fig. 2 for the two-objective case, when $N_{\text{par}} = 5$ Pareto solutions exist. Design vectors which yield an improvement map to either of the shaded regions; design vectors which dominate at least one solution in S map to the region labeled 'Dominating Designs', whilst design vectors which do not dominate any solution in S (but which still constitute an improvement) map to the region labeled 'Equivalent Designs'. Equations for the probability of an unevaluated design vector yielding an improvement are given in [4] and [5], and the equation for the probability of an unevaluated design vector dominating at least one solution in S for the two-objective case is given in [5]. Both [4] and [5] report problems

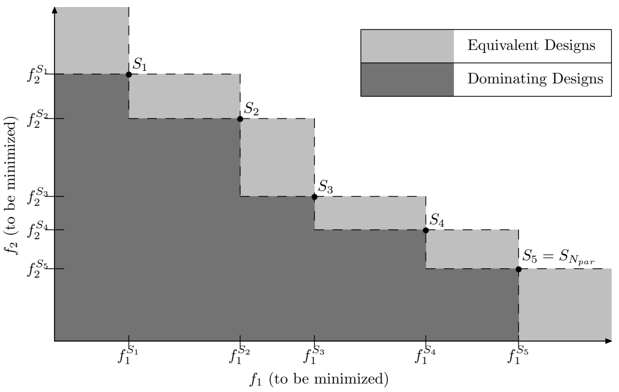

Fig. 2. Regions of dominance and equivalence for $N_{\text{par}} = 5$ Pareto solutions.

with the probability of improvement criteria yielding searches which are not very global, and so only small improvements are made. Both overcome this by instead maximizing the first moment of the probability of improvement around the Pareto-optimal front, which is the multi-objective equivalent of the expected improvement utility function, i.e. they use the first method suggested in Section 1 for overcoming the equivalent sensitivity problem in the single-objective case. The following Section shows how to overcome this sensitivity problem in the multi-objective case using the second method proposed in Section 1, that is, it proposes an enhanced probability of improvement criteria for multi-objective optimization.

III. Enhanced Probability of Improvement in Multi-Objective Optimization

In the multi-objective case, $k = N_{\text{par}}$ natural levels of improvement may be defined, where the k^{th} level of improvement yields a solution which dominates *exactly* k of the existing Pareto-optimal solutions. In addition, a level of equivalence may also be defined ($k = 0$), which yields an additional Pareto-optimal solution which does not dominate any of the existing Pareto-optimal solutions (i.e a design vector which maps to the region labeled 'Equivalent Designs' in Fig. 2). These levels of improvement are shown in Fig. 3 for the $N_{\text{par}} = 5$ Pareto solutions considered earlier.

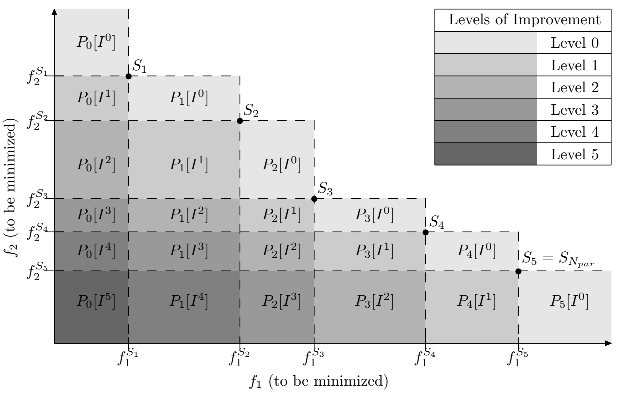

Fig. 3. Probability of improvement levels for $N_{\text{par}} = 5$ Pareto solutions.

As can be seen, in the two-objective case, for an improvement level k, a design vector may map to $N_{\text{par}} - k + 1$ regions of objective function space. Denoting $P(I^k(\mathbf{x}))$ as the probability that an unknown design vector $\mathbf{x}$ yields a level of improvement k (i.e. it dominates exactly k existing Pareto-optimal solutions), $P_i(I^k(\mathbf{x}))$ as the probability that design vector $\mathbf{x}$ will dominate the k Pareto solutions $S_{i+1}, S_{i+2}, \ldots, S_{i+k}$ (these sub-regions are labeled in Fig. 3), and defining

$$\Phi_1^i(\mathbf{x}) := \Phi\left(\frac{f_1^{S_i} - \hat{f}_1(\mathbf{x})}{s_1(\mathbf{x})}\right) \quad (2)$$

$$\Phi_2^i(\mathbf{x}) := \Phi\left(\frac{f_2^{S_i} - \hat{f}_2(\mathbf{x})}{s_2(\mathbf{x})}\right) \quad (3)$$

with

$$\Phi_1^0(\mathbf{x}) := 0 \quad (4)$$

$$\Phi_1^{N_{\text{par}}+1}(\mathbf{x}) := 1 \quad (5)$$

$$\Phi_2^0(\mathbf{x}) := 1 \quad (6)$$

$$\Phi_2^{N_{\text{par}}+1}(\mathbf{x}) := 0 \quad (7)$$

where $\hat{f}_1(\cdot)$ and $s_1(\cdot)$ are the kriging predictions and standard errors for the first objective function respectively (similarly for the second objective function), and $f_1^{S_i}$ is the first objective function value of the i^{th} Pareto solution (similarly for the second objective function), then:

$$P(I^k(\mathbf{x})) = \sum_{i=0}^{N_{\text{par}}-k} P_i(I^k(\mathbf{x})) \quad (8)$$

$$= \sum_{i=0}^{N_{\text{par}}-k} (\Phi_1^{i+1}(\mathbf{x}) - \Phi_1^i(\mathbf{x}))(\Phi_2^{k+i}(\mathbf{x}) - \Phi_2^{k+i+1}(\mathbf{x})). \quad (9)$$

Furthermore, denoting by $P^*(I^k(\mathbf{x}))$ the probability that $\mathbf{x}$ will dominate *at least* k existing Pareto-optimal solutions, then

$$P^*(I^k(\mathbf{x})) = \sum_{j=k}^{N_{\text{par}}} \sum_{i=0}^{N_{\text{par}}-j} (\Phi_1^{i+1}(\mathbf{x}) - \Phi_1^i(\mathbf{x}))(\Phi_2^{j+i}(\mathbf{x}) - \Phi_2^{j+i+1}(\mathbf{x})). \quad (10)$$

Equations (9) and (10) are two multi-objective equivalents of the 'enhanced probability of improvement' in single-objective optimization, for the case of two objectives. The method is extensible to higher numbers of objectives.

IV. Conclusions

A novel utility function, which is easily parallelized and which does not require normalization of objective functions, has been proposed for use in computationally expensive multi-objective optimization. Its performance in optimal electromagnetic design will be discussed in the full paper.

References

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

[2] 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.

[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] M. T. M. Emmerich, K. C. Giannakoglou and B. Naujoks, "Single- and Multiobjective Evolutionary Optimization Assisted by Gaussian Random Field Metamodels", *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 4, pp. 421-439, 2006.

[5] A. J. Keane, "Statistical Improvement Criteria for Use in Multiobjective Design Optimization", *AIAA Journal*, vol. 44, no. 4, pp. 879–891, 2006.