

## Research Article

## Open Access

Yinjiang Li, Song Xiao, Paolo Di Barba, Mihai Rotaru, and Jan K. Sykulski\*

# Localized probability of improvement for kriging based multi-objective optimization

<https://doi.org/10.1515/phys-2017-0117>

Received October 30, 2017; accepted November 12, 2017

**Abstract:** The paper introduces a new approach to kriging based multi-objective optimization by utilizing a local probability of improvement as the infill sampling criterion and the nearest neighbor check to ensure diversification and uniform distribution of Pareto fronts. The proposed method is computationally fast and linearly scalable to higher dimensions.

**Keywords:** kriging; multi-objective optimization; Pareto front; surrogate-based optimization

**PACS:** 89.20.Kk, 85.85.+j, 85.70.-w, 87.10.Mn

## 1 Introduction

Research on multiple objective optimization (MO) has been attracting significant attention of the engineering community since 1980s; with the aid of fast computers solutions to many complex optimization problems have been made possible. The Vector Evaluated Genetic Algorithm (VEGA) [1] is one of the earliest examples of Multi-Objective Evolutionary Algorithms (MOEAs). The more recent developments include NSGA-II [2] and its modified versions as well as Particle Swarm based methods [3]. A comprehensive review of problem definitions and non-EA based solution methods may be found in [4].

There is an increasing number of indicator-based MOEAs that have been proposed in recent years; the indicator is used as a fitness measure for a set of Pareto

points, and – by optimizing the indicator function – the MO problem essentially becomes a single objective optimization problem as the solver only needs to locate the optimal value of the indicator value and update the generation based on it. One of the best-known indicators is the hypervolume [5]; it has been successfully applied to both EAs and surrogate-based algorithms. Despite its unique feature of being strictly monotonic to Pareto improvements [6], it suffers from high computational cost for higher dimensions.

The general opinion favors EAs as advantageous in solving MO problems by often being population based, thus multiple solutions can be obtained in a single run. However, solutions to practical problems may be expensive in terms of computational time and effort. In the context of electromagnetic devices, the finite element method is a common design tool; it often takes hours or even days to obtain a single solution, therefore surrogate model based algorithms are often preferred.

In this study we propose a new indicator focused Localized Probability of Improvement (LPOI) approach for MO problems; its implementation requires the predicted mean and mean square errors to be available, hence it is not applicable to other EAs, but for Gaussian based surrogate models (including those relying on kriging) it has the advantage of being linearly scalable to problems with higher number of objectives.

## 2 Kriging theory

Modern engineering design often involves implementation of deterministic computer simulation; in electromagnetic design, time consuming finite element models (FEM) are often built to represent the actual devices. Designs are analyzed and optimized before being put into production. In these types of problems, the optimization can be a very time consuming process due to a large number of FEM calls needed. Therefore surrogate modeling techniques are often used to reduce the number of expensive FEM simulations.

**Yinjiang Li, Song Xiao, Mihai Rotaru:** Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom of Great Britain and Northern Ireland, E-mail: yl10e09@ecs.soton.ac.uk, jks@soton.ac.uk

**Paolo Di Barba:** Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy

\***Corresponding Author: Jan K. Sykulski:** Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom of Great Britain and Northern Ireland, E-mail: jks@soton.ac.uk

Kriging is one of the most commonly used surrogate techniques amongst many others. An ordinary kriging model  $Y$  consists of a global mean  $f$  and a local departure  $Z$ :

$$Y(x) = f(x) + Z(x) \tag{1}$$

where  $x$  is the location of any design site.

The local departure follows the Gaussian distribution with a mean of zero, variance  $\sigma^2$  and non-zero covariance. A general exponential correlation function is one of the most commonly used correlation functions, due to its continuous characteristic and flexibility

$$corr[x_i, x_j] = \prod_n^k \exp(-\theta_n |x_{in} - x_{jn}|^{p_n}) \tag{2}$$

where  $x_i$  and  $x_j$  are a pair of observations,  $k$  is the problem dimension, while  $\theta_n$  and  $p_n$  are hyperparameters controlling the shape of the correlation function.

Kriging parameters  $u$ ,  $\sigma^2$  and the hyperparameters  $\theta$  and  $p$  are obtained via the Maximum Likelihood Estimation (MLE), with the maximum likelihood function given by

$$\frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}} |R|^{\frac{1}{2}}} \exp\left[-\frac{y - 1u'R^{-1}(y - 1u)}{2\sigma^2}\right] \tag{3}$$

where  $y$  denotes all observations and  $R$  is the correlation matrix.

The kriging prediction and the predicted mean square error (MSE) at a given location  $x$  are given as follows

$$\hat{y} = \hat{\mu} + rR^{-1}(y - 1\hat{\mu}) \tag{4}$$

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

with  $\hat{\sigma}^2 = \frac{(1-1\hat{\mu})R^{-1}(y-1\hat{\mu})}{n}$  and  $\hat{\mu} = \frac{1'R^{-1}y}{1'R^{-1}1}$ , where  $\hat{\mu}$  and  $\hat{\sigma}^2$  are the optimal mean and variance, respectively, obtained by solving the maximum likelihood function.

### 3 Localized probability of improvement

Compared to other surrogate modeling methods, kriging has the advantage of providing both the predicted mean and the associated mean square error (MSE) at an unknown location. The probability of improvement  $PoI$  at any location is given by

$$PoI(x) = \Phi\left(\frac{y_t(x) - \hat{y}(x)}{\hat{s}(x)}\right) \tag{6}$$

where  $y_t$  is the target of improvement,  $\hat{y}$  the kriging predicted mean at location  $x$ ,  $\hat{s}$  is a square root of the mean square error at location  $x$  and  $\Phi(\cdot)$  is the cumulative distribution function.

The first improvement target  $y_{ext}$  is associated with the minimum value of each individual objective function; the subscript  $ext$  stands for “extreme value” and  $y_{ext}$  is given by

$$y_{ext}^n = y_{min}^n \cdot (1 - p) \tag{7}$$

where  $y_{min}^n$  is the known minimum value of the  $n^{th}$  objective function and  $p$  is the percentage of improvement to be defined; parameter  $p$  is discussed later in this section. The corresponding  $PoI$  is:

$$PoI_{ext}^n(x) = \Phi\left(\frac{y_{ext}^n - \hat{y}^n(x)}{\hat{s}^n(x)}\right) \tag{8}$$

where  $\hat{y}^n$ ,  $\hat{s}^n$ ,  $y_{ext}^n$  and  $PoI_{ext}^n$  are the corresponding measures of the  $n^{th}$  objective function.

For the first improvement target, we find  $n$  values of  $PoI$ , which equals to the number of objectives, because the  $PoI_{ext}$  is calculated based on the extreme value of each objective function. We consider the maximum potential improvement for all individual objectives, hence

$$PoI_{ext}(x) = \text{maximize} \{PoI_{ext}^n(x)\} \tag{9}$$

The second improvement target  $y_{int}^n(x)$  is associated with the reference point that is defined based on the location of  $x$ . The subscript  $int$  stands for “intermediate” and  $y_{ref}$  is calculated as

$$y_{int}^n = y_{ref}^n \cdot (1 - p) \tag{10}$$

where  $y_{ref}$  is the calculated reference point.

To obtain the reference point  $y_{ref}$ , the algorithm finds the Pareto front for the existing design sites using non-dominated sorting. For each closest set of Pareto points (the number of points is equal to the number of objectives) it calculates the corresponding reference point. The coordinates for the reference point of each dimension is equal to the maximum value of the coordinates for these Pareto points in the same dimension. The coordinates for the corresponding reference point in the  $n^{th}$  dimension  $Ref(x^n)$  is given by:

$$y_{ref}^n = \text{max}\{Y^n\} \tag{11}$$

where  $Y^n$  is the collection of the  $n^{th}$  objective values for all of the points in that Pareto set.

Taking a bi-objective problem as an example, assuming the reference point  $y_{ref}$  is to be determined for Pareto points  $P_1$  and  $P_2$ , the coordinates of  $P_1$  and  $P_2$  are therefore denoted by  $[P_1.x^1, P_1.x^2]$  and  $[P_2.x^1, P_2.x^2]$ , respectively. Note that  $x^n$  is the  $n^{th}$  objective value at the location in the search space associated with  $P$ . The  $x^1$  and  $x^2$

coordinates (in the objective space) of the reference point are thus described as follows

$$y_{ref}^1 = \max \{ P_{1.x^1}, P_{2.x^1} \} \tag{12}$$

$$y_{ref}^2 = \max \{ P_{1.x^2}, P_{2.x^2} \} \tag{13}$$

and the corresponding *PoI* is given as

$$PoI_{ref}^n(x) = \Phi \left( \frac{y_{int}^n(x) - \hat{y}^n(x)}{\hat{s}^n(x)} \right) \tag{14}$$

where  $y_{ref}^n$ ,  $\hat{y}^n$ ,  $\hat{s}^n$ ,  $PoI_{ref}^n$  and  $PoI_{ext}^n$  are the corresponding measures of the  $n^{th}$  objective function.

We have therefore obtained  $n$  values of *PoI* for the second improvement target. However, unlike the first improvement target, the second one uses a localized target. Therefore, we consider using the minimum potential improvement for all individual objectives and hence

$$LPoI_{ref}(x) = \text{minimize} \{ PoI_{ref}^n(x) \} \tag{15}$$

Finally, the proposed indicator *LPoI* for any given point is the maximum of these two probability of improvement measures, given by

$$LPoI(x) = \text{maximize} \{ LPoI_{ref}, PoI_{ext} \} \tag{16}$$

where  $PoI_{ext}$ , as described by (9), is due to the fact that the minimum of each individual objective function is always present in the Pareto front, thus the *PoI* at each location  $x$ , over the optimal target of that function, is always considered. This term also contributes to the diversification of the Pareto front.

Furthermore,  $LPoI_{ref}$  – as described by (15) – can be treated as a maximum of the minimum potential improvements to a local target. This term helps to improve the Pareto front both towards the origin and in the direction of the objective value. It contributes to the diversification of the Pareto front, while the max-min method also contributes to the uniformity of the Pareto front.

To obtain the next infill sampling point, the algorithm finds the location  $x$  associated with the maximum *LPoI* measure in the objective space.

The parameter  $p$  – as seen in (7) and (10) – is associated with the magnitude of target improvement; it controls the convergence rate of the algorithm. A smaller amount of improvement will guide the solver towards existing Pareto points, while a larger value will encourage the exploration of the design space. It is crucial to use a proper  $p$ , since too small a value may lead to a false Pareto front, while a large value may result in a slow convergence rate or zero probability of improvement at all unknown sites. Thus it

is advisable to dynamically adjust the value while monitoring the convergence.

We provide a simple self-adjusted method for parameter  $p$  in this paper. First, the initial improvement target percentage  $p_{initial}$  is defined and then the parameter  $p$  is calculated as

$$p = p_{initial} \cdot \max\{LPoI_{prev}\} \tag{17}$$

where  $LPoI_{prev}$  is a complete set of *LPoI* at previous iteration.

The next infill point is taken at the location with a maximum *LPoI*. Therefore, the solver tends to minimise the localised probability of improvement and converges towards the Pareto front. When the design space is well explored, or  $p$  is especially small, the solver will converge towards existing Pareto fronts; at this stage, it is common for the *LPoI* to be equal, or come close, to 1 at multiple unknown sites (extremely likely to improve over the target point). In order to obtain a uniformly distributed Pareto front, the algorithm selects candidates which have the largest Euclidean distance to existing Pareto points compared to the next infill sampling points. For this reason, the maximum value of *LPoI* can be capped between 0.9 and 1 for faster exploitation of the existing Pareto front without degrading the overall performance.

## 4 Test examples

In this section we illustrate the proposed method using a bi-objective example.

The top graph in Figure 1 shows the kriging model (solid line) after 45 iterations, with the red crosses plotted at the true Pareto front, while the bottom plot shows the proposed indicator value for the unknown sites. As

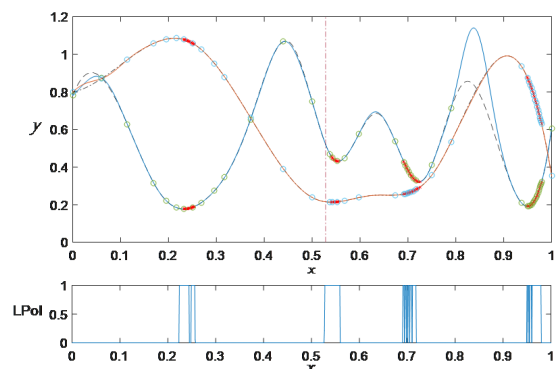


Figure 1: The kriging model and the LPoI criterion in the search space

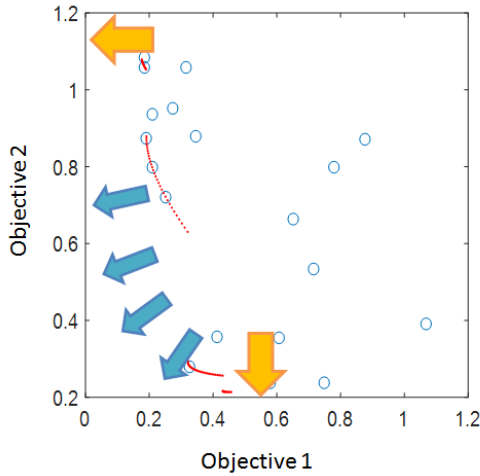


Figure 2: Existing design sites in the criterion space after 20 iterations

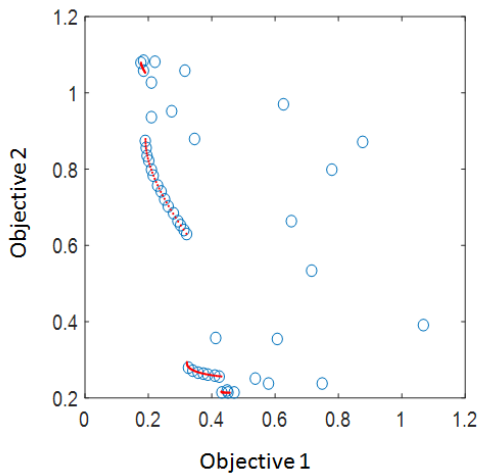


Figure 3: Existing design sites in the criterion space after 45 iterations

can be seen, the algorithm has correctly converged to all four Pareto point clusters in the search space and thus further sampling will lead to more exploitation on the Pareto front. The sampled design sites in the objective space are plotted in Figure 2, where the red dots indicate the location of the true Pareto front. The improvement direction imposed by the two improvement targets are illustrated in Figure 2, where the yellow arrows show the improvement direction for the first improvement target, and the blue arrows indicate the improvement direction for the second improvement target.

## 5 Solving the new TEAM problem

A new TEAM problem was proposed at the Compumag conference in Korea, June 2017 [7]. This is devoted specifically to multi-objective optimization. In its extended version, an additional objective has been added and there are therefore in total three objectives. The model consists of an air-cored multi-turn winding. By arranging the current-carrying coils, a desired magnetic field distribution is to be obtained. The flux density at point  $z$  is given by

$$B(z) = \frac{\mu_0}{2} \int_{-d}^d \int_{r_i}^{r_s} \frac{Jr^2(\xi) dr d\xi}{\sqrt{[r^2(\xi) + (z - \xi)^2]^3}} \quad (18)$$

The problem is specified as follows: given the current density  $J$  within the coil, and prescribed flux density, find the optimal  $r$  distribution of radii  $r(z)$ ,  $-d \leq z \leq d$  that yields the prescribed flux density  $B_0(z)$ .

An initial arrangement of turns was given in the extended paper of [7], the width of each turn  $w$  and the height  $h$  are 1 mm and 1.5m m, respectively.

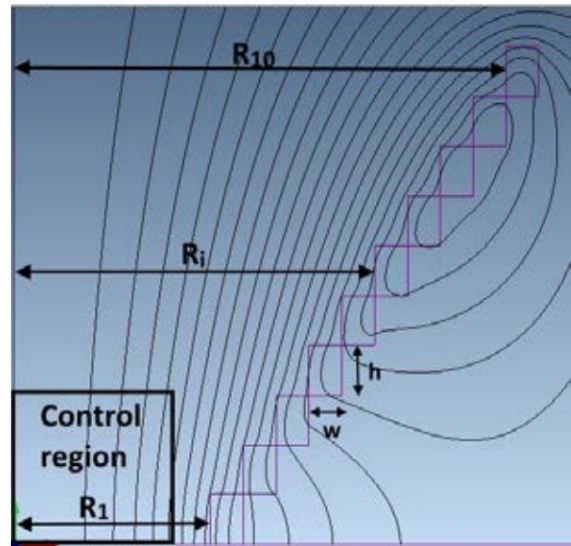


Figure 4: Example of radii distribution; geometry and magnetic flux lines

The model consists of 20 turns connected in series, symmetrically distributed, hence there are 10 radii which need to be optimized (the main objective  $f_1$ ). Two additional objectives were proposed to complement the first objective  $f_1$ . The three objectives  $f_1, f_2$  and  $f_3$  may be described as follows:

$f_1$ : find the optimal distribution of  $r$ , so that the discrepancy between the prescribed flux density  $B_0$  and the actual

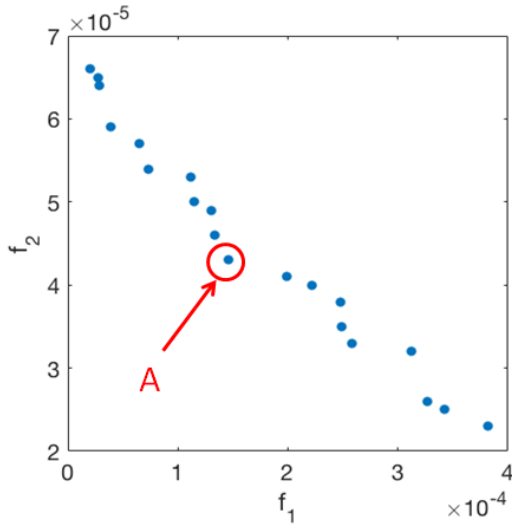


Figure 5: Pareto front of  $f_1$  and  $f_2$  in the objective space

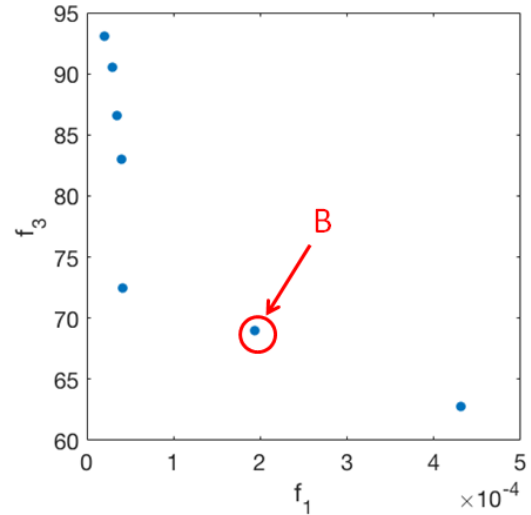


Figure 6: Pareto front of  $f_1$  and  $f_3$  in the objective space

induction field  $B$  is minimized;

$f_2$ : minimize the sensitivity function;

$f_3$ : minimize the power loss related function.

Mathematically the three objective functions are expressed as

$$f_1(r) = \sup_{q=1, n_p} |B(r(\xi_l), z_q) - B_0(r_q, z_q)|, \quad l = 1, n_t \quad (19)$$

$$f_2(r) = \sup_{l=1, n_t} \left[ \|B^+ - B(r(\xi_l), z_q)\| + \|B(r(\xi_l), z_q) - B^-\| \right], \quad q = 1, n_p \quad (20)$$

$$f_3(r) = \sum_{l=1}^{n_t} r(\xi_l) \quad (21)$$

where  $B^+ = B(r(\xi_l + \Delta\xi), z_q)$ ,  $B^- = B(r(\xi_l - \Delta\xi), z_q)$ ,  $l = 1, n_t$  and  $q = 1, n_p$ .  $\Delta\xi = 0.5$  mm. At this stage it was suggested to consider only two objectives at the time,  $f_1$  and either  $f_2$  or  $f_3$ .

The optimization results are illustrated by Figs. 5 and 6, where objectives 2 and 3 are plotted against objective 1, respectively. The globally optimal points A and B (defined by the closeness to the respective utopia points) are defined by the radii distributions [11.4, 8.6, 9.1, 12.1, 8.9, 8.3, 7.0, 6.4, 6.8, 5.9] and [7.2, 10.6, 7.2, 6.6, 9.0, 5.2, 9.2, 5.0, 5.4, 6.9], respectively.

## 6 Conclusion

A novel approach to kriging-based multi objective optimization is put forward relying on the Localized Prob-

ability of Improvement. For illustration purposes a bi-objective test problem is provided, as well as the recently introduced TEAM benchmark problem. It is shown that the proposed method addresses efficiently both the diversification and uniformity of the Pareto solution, is computationally efficient and is linearly scalable to higher number of objectives.

## References

- [1] Schaffer J.D., Multiple objective optimization with vector evaluated genetic algorithms, Proceedings of the 1st International Conference on Genetic Algorithms, 1985, 93-100
- [2] Deb K., Agrawal S., Pratap A., Meyarivan T., A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, in Schoenauer M. et al. (eds), Parallel Problem Solving from Nature PPSN VI, 2000
- [3] Parsopoulos K.E., Vrahatis M.N., Particle swarm optimization method in multiobjective problems, SAC'02 Proceedings of the 2002 ACM Symposium on Applied Computing, 2002, 603-607
- [4] Marler R.T., Arora J. S., Survey of multi-objective optimization methods for engineering, Structural and Multidisciplinary Optimization, 2004, 26, 6, 369
- [5] Zitzler E., Thiele L., Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach, IEEE Transactions on Evolutionary Computation, 1999, 3, 4, 257-271
- [6] Knowles J., Corne D., On metrics for comparing nondominated sets, Evolutionary Computation, Proceedings of CEC'02, Honolulu, 2002, 711-716
- [7] Barba P.D., Mognaschi M.E., Song X., Lowther D.A., Sykulski J.K., A benchmark TEAM problem for multiobjective Pareto optimization of electromagnetic devices, IEEE Transactions on Magnetics, 2017, PP, 99