Exploring the Performance of Stochastic Multiobjective Optimisers with the Second-Order Attainment Function

Carlos M. Fonseca[1], Viviane Grunert da Fonseca[1,2], and Luís Paquete[3,4]

[1] CSI – Centro de Sistemas Inteligentes
Faculdade de Ciências e Tecnologia
Universidade do Algarve
Faro, Portugal
cmfonsec@ualg.pt, vgrunert@csi.fct.ualg.pt
[2] INUAF – Instituto Superior D. Afonso III
Loulé, Portugal
[3] Fachgebiet Intellektik, Fachbereich Informatik
Technische Universität Darmstadt
Darmstadt, Germany
lpaquete@intellektik.informatik.tu-darmstadt.de
[4] Faculdade de Economia
Universidade do Algarve
Faro, Portugal

Abstract. The attainment function has been proposed as a measure of the statistical performance of stochastic multiobjective optimisers which encompasses both the quality of individual non-dominated solutions in objective space and their spread along the trade-off surface. It has also been related to results from random closed-set theory, and cast as a mean-like, first-order moment measure of the outcomes of multiobjective optimisers. In this work, the use of more informative, second-order moment measures for the evaluation and comparison of multiobjective optimiser performance is explored experimentally, with emphasis on the interpretability of the results.

1 Introduction

Stochastic multiobjective optimisers, such as evolutionary algorithms and other metaheuristics, produce Pareto-set approximations which consist of sets of non-dominated points in objective space. Given the stochastic nature of the optimisers, such non-dominated point sets may be seen as realisations of corresponding random non-dominated point sets, the stochastic behaviour of which is tied both to the problem and to the optimiser considered.

The performance of a multiobjective optimiser is intimately related to the quality of the Pareto-set approximations it produces. In the literature, several attempts have been made to quantify the quality of (deterministic) Pareto-set approximations through so-called unary quality indicators, which are functions which assign real values to each Pareto-set approximation. In the face of *random* Pareto-set approximations, unary quality indicators provide a convenient transformation from random sets to random variables. However, this approach suffers from inherent limitations of unary quality indicators. In fact, even a finite number of unary quality indicators which, in combination with each other, would completely describe a deterministic set of non-dominated points in objective space cannot exist in practice [1]. As a result, information is irremediably lost by finite-dimensional unary quality indicators even before any statistical analysis takes place.

To be able to retain all of the information available in the original non-dominated sets, a quality indicator must be infinite-dimensional, such as the binary field [2] derived from the set of goals attained by a Pareto-set approximation (attained set or attained region). Describing a Pareto-set approximation by a (real-valued) function defined over the whole of the objective space may not seem to be very sensible in the deterministic case. However, in the random case, it provides a useful link to existing random set theory, where distributions of random sets are studied directly (possibly, up to a complete distributional description) and not indirectly through distributions of summary measures (indicators) of the sets. The attainment function of a random Pareto-set approximation, for example, has been identified as the first-order moment measure of the binary random field derived from the corresponding random attained set [3], and, as such, is a concept perfectly integrated in random set theory.

By combining fundamentally different features of the quality of Pareto-set approximations, such as the quality of individual solutions and their spread along the trade-off surface, into a real-valued function of the goals, the attainment function can already effectively describe an important aspect of the distribution of random Pareto-set approximations, namely its location. To address the dependence structure of individual solutions within these approximation sets, two additional measures of performance are considered in this work, both of which are of second-order moment type: the second-order attainment function and its centred version, the covariance function.

In section 2, background is given on the attainment function. In section 3, the second-order attainment function and the covariance function are introduced. Empirical estimates of the two functions, obtained from experimental data, are presented graphically, and their interpretation is discussed. Section 4 is devoted to the comparison of optimiser performance using statistical hypothesis tests based on first-order or second-order attainment functions. To illustrate the application of the attainment function approach, experimental results obtained on two different optimisation problems are presented and discussed in section 5. The paper concludes with some remarks and directions for further work in section 6.

2 Background

The outcome of a multiobjective optimiser is considered to be the set of non-dominated objective vectors evaluated during one optimisation run. If the opti-

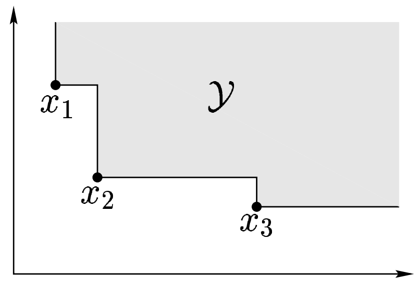


Fig. 1. *RNP-set $\mathcal{X}$ with non-dominated realisations x_1, x_2, and x_3 and the attained set $\mathcal{Y}$ (here as a realisation); compare with Grunert da Fonseca et al. [3].*

miser is stochastic, such Pareto-set approximations are random, and their distribution becomes of interest when assessing the optimiser's performance. Consider the following definitions (assuming minimisation of all objective functions without loss of generality):

Definition 1 (Random non-dominated point set). *A random point-set*

$$\mathcal{X} = \{X_1, \ldots, X_M \in \mathbb{R}^d : \; P(X_i \leq X_j) = 0, \; i \neq j\},$$

where both the number of elements M and the elements X_j themselves are random and $P(0 \leq M < \infty) = 1$, is called a random non-dominated point set (RNP-set).

Random Pareto-set approximations produced by stochastic multiobjective optimisers on d-objective problems are RNP-sets in $\mathbb{R}^d$.

Definition 2 (Attained set). *The random set*

$$\begin{aligned}\mathcal{Y} &= \{y \in \mathbb{R}^d \mid X_1 \leq y \;\vee\; X_2 \leq y \;\vee \ldots \vee\; X_M \leq y\} \\ &= \{y \in \mathbb{R}^d \mid \mathcal{X} \trianglelefteq y\}\end{aligned}$$

is the set of all goals $y \in \mathbb{R}^d$ attained by the RNP-set $\mathcal{X}$ (see Figure 1).

The distributions of both random sets, $\mathcal{X}$ and $\mathcal{Y}$, are equivalent, i.e. a characterisation of the distribution of $\mathcal{X}$ automatically provides a characterisation of the distribution of $\mathcal{Y}$, and vice versa.

Definition 3 (Attainment indicator). *Let $\mathbf{I}\{\cdot\} = \mathbf{I}_{\{\cdot\}}(z)$ denote the indicator function. Then, the random variable $b_{\mathcal{X}}(z) = \mathbf{I}\{\mathcal{X} \trianglelefteq z\}$ is called the attainment indicator of $\mathcal{X}$ at goal $z \in \mathbb{R}^d$.*

The set of all attainment indicators indexed by $z \in \mathbb{R}^d$ is the binary random field $\{b_{\mathcal{X}}(z), z \in \mathbb{R}^d\}$. For the deterministic case, this binary field fully characterises a single Pareto-set approximation, as one can always be obtained from the other. As an infinite-dimensional quality indicator, it can be used to construct a comparison method which is complete and compatible with respect to weak-dominance [1]. Although this may not seem to be very useful in the deterministic case, such a quality indicator provides an interesting assessment tool in the random case:

Definition 4 (Attainment function). *The function* $\alpha_{\mathcal{X}} : \mathbb{R}^d \longmapsto [0,1]$ *with*

$$\alpha_{\mathcal{X}}(z) = P\big(b_{\mathcal{X}}(z) = 1\big)$$

is called the attainment function of $\mathcal{X}$.

As identified in Grunert da Fonseca *et al.* [3], the attainment function is the first-order moment measure of the binary random field $\{b_{\mathcal{X}}(z), z \in \mathbb{R}^d\}$ derived from $\mathcal{Y}$ (the set attained by $\mathcal{X}$) and, as such, it offers a useful description of the *location* of the distribution of $\mathcal{Y}$ (and also of $\mathcal{X}$). Note that for $M = 1$, the optimiser produces a single random objective vector X per run, and the attainment function reduces to the usual multivariate distribution function $F_X(z) = P(X \leq z)$. A natural empirical counterpart of the (theoretical) attainment function $\alpha_{\mathcal{X}}(\cdot)$ may be defined as follows:

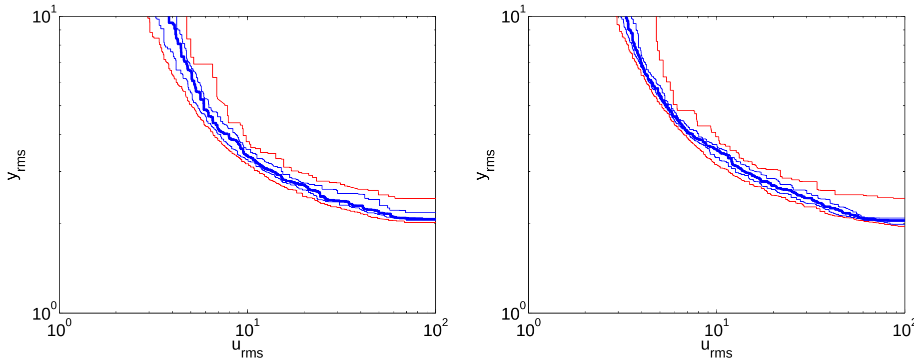
Definition 5 (Empirical attainment function). *Let* $b_1(z), \ldots, b_n(z)$ *be* n *realizations of the attainment indicator* $b_{\mathcal{X}}(z)$, $z \in \mathbb{R}^d$. *Then, the function defined as* $\alpha_n : \mathbb{R}^d \longmapsto [0,1]$ *with*

$$\alpha_n(z) = \frac{1}{n} \cdot \sum_{i=1}^{n} b_i(z)$$

is called the empirical attainment function of $\mathcal{X}$ *(EAF).*

The realizations $b_1(z), \ldots, b_n(z)$ correspond to n runs of the optimiser under study. In Figure 2, contour plots of the EAFs obtained from 21 independent runs each of two simple multiobjective genetic algorithm variants (MOGA-A and MOGA-B) on a bi-objective optimal control problem are presented (see section 5 for additional information). While the theoretical attainment function is continuous, the EAF is a discontinuous function which exhibits transitions not only at the data points but also at other points, the coordinates of which are combinations of the coordinates of the data points, much like in the case of the multivariate empirical distribution function [4]. Contours are drawn (from left to right) at the ϵ-, 0.25-, 0.5-, 0.75- and $(1-\epsilon)$-levels, for arbitrarily small positive ϵ. The function value $\alpha_n(z)$ at a given goal $z \in \mathbb{R}^d$ indicates the proportion of optimisation runs (Pareto-set approximations) which produced at least one solution in objective space attaining that goal. Hence, all goals $z \in \mathbb{R}^d$ which are located on the 0.5-level of the EAF, for example, were attained in 50% of the optimisation runs carried out. This line may also be called the 50%-attainment surface [5].

The EAF serves as an estimator for the theoretical attainment function $\alpha_{\mathcal{X}}(z)$, in the same way as the multivariate empirical distribution function estimates the (theoretical) multivariate distribution function $F_X(z)$, for all $z \in \mathbb{R}^d$.



(a) MOGA-A

(b) MOGA-B

Fig. 2. EAF contour plots (first example).

Thus, the performance of an optimiser on a given optimisation problem, understood in terms of *location* of the corresponding RNP-set distribution, may be assessed via EAF estimates. The farther down and to the left the weight of the attainment function, the greater is the probability of attaining tighter goals (*independently* from one another), and the better is the performance of the algorithm.

In the above sense, performance of two (or more) optimisers operating on the same optimisation problem can be compared by comparing the corresponding attainment functions. A suitable, Smirnov-like, statistical testing procedure based on (two) EAFs has been applied earlier by Shaw *et al.* [6]. Rejecting the null-hypothesis of equal attainment functions in a statistically significant way supports the conclusion that the optimisers under study exhibit different performance. However, if such a null hypothesis cannot be rejected, optimisers may still exhibit different performance, not only because of the statistical error involved, but also because the RNP-set distribution, and thus the performance, of a stochastic multiobjective optimiser is *not completely characterised* by the attainment function.

Whereas the attainment function, as a measure of first-order moment type, describes the distribution of the RNP-set $\mathcal{X}$ in terms of location, it does not address the dependence structure within the non-dominated elements of $\mathcal{X}$. Measures of second-order moment type will now be defined for that purpose.

3 Second-Order Moment Measures

Measures of second-order moment type allow the pairwise relationship between the elements of a random Pareto-set approximation $\mathcal{X}$ to be studied. Consider the following definitions (again assuming a minimisation problem, without loss of generality):

Definition 6 (Second-order attainment function). *The function defined as* $\alpha_{\mathcal{X}}^{(2)} : \mathbb{R}^d \times \mathbb{R}^d \longmapsto [0,1]$, *with*

$$\alpha_{\mathcal{X}}^{(2)}(z_1, z_2) = P\Big(b_{\mathcal{X}}(z_1) = 1 \ \wedge \ b_{\mathcal{X}}(z_2) = 1\Big)$$

is called the second-order attainment function of $\mathcal{X}$.

The second-order attainment function is the second, non-centred, moment measure of the binary random field $\{b_{\mathcal{X}}(z), z \in \mathbb{R}^d\}$ derived from the attained set $\mathcal{Y}$ [3]. In random set theory terminology, the second-order attainment function would be called the covariance of the attained set $\mathcal{Y}$ (see, for example, Stoyan *et al.* [7]). It expresses the probability of (the elements of) the same Pareto-set approximation $\mathcal{X}$ simultaneously attaining two different goals, $z_1, z_2 \in \mathbb{R}^d$. Obviously, the second-order attainment function is symmetric in its arguments, and includes all the information of the (first-order) attainment function, as

$$\alpha_{\mathcal{X}}^{(2)}(z, z) = \alpha_{\mathcal{X}}(z) \quad \text{for all } z \in \mathbb{R}^d,$$

and

$$\alpha_{\mathcal{X}}^{(2)}(z_1, z_2) = \alpha_{\mathcal{X}}^{(2)}(z_2, z_1) = \alpha_{\mathcal{X}}(z_1) \quad \text{for all } z_1 \leq z_2 \in \mathbb{R}^d.$$

A natural empirical counterpart of the (theoretical) second-order attainment function may be defined as follows:

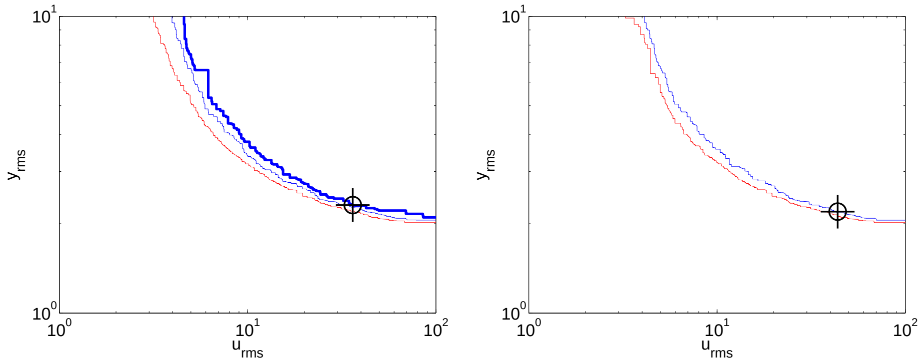
Definition 7 (Second-order empirical attainment function). *Let* $b_1(z)$, $\ldots, b_n(z)$ *be* n *realizations of the attainment indicator* $b_{\mathcal{X}}(z)$, $z \in \mathbb{R}^d$. *Then, the function* $\alpha_n^{(2)} : \mathbb{R}^d \times \mathbb{R}^d \longmapsto [0,1]$ *with*

$$\alpha_n^{(2)}(z_1, z_2) = \frac{1}{n} \cdot \sum_{i=1}^{n} b_i(z_1) \cdot b_i(z_2)$$

is called the second-order empirical attainment function of $\mathcal{X}$ *(second-order EAF).*

The second-order EAF is a discontinuous function, with the values $\alpha_n^{(2)}(z_1, z_2)$ representing the proportion of optimisation runs (Pareto-set approximations) which attained goals z_1 and z_2 simultaneously.

The visualisation of the second-order EAF is more difficult than that of the first-order EAF, as it is defined in $\mathbb{R}^{2d}$. Even with only two objectives, this results in four dimensions, and direct visualisation is impossible. A useful work-around consists of fixing one goal $z^* \in \mathbb{R}^2$ and depicting the contours of the marginal function $\alpha_n^{(2)}(z, z^*) = \alpha_n^{(2)}(z^*, z)$, defined over all $z \in \mathbb{R}^2$, at given levels. Figure 3(a) shows the contours of the marginal second-order EAF at levels ϵ, 0.25, and 0.5 (from left to right), considering a fixed goal $z^* \in \mathbb{R}^2$. Higher-level contours are not shown, because $\alpha_n(z^*) < 0.75$ in this case, and so must $\alpha_n^{(2)}(z, z^*)$ be, for all $z \in \mathbb{R}^2$.



(a) With a fixed goal $z^* \in \mathbb{R}^2$ (b) With a different fixed goal $z^* \in \mathbb{R}^2$

Fig. 3. Contour plot of the marginal second-order EAF (MOGA-A)

A better impression of the second-order EAF would be offered by an interactive session. By changing the position of the goal z^* in the plot of the marginal function $\alpha_n^{(2)}(z, z^*)$, it is possible to explore how the marginal second-order EAF changes as a function of the goal z^*. By pulling the goal z^* further downwards, Figure 3(b) is obtained. Note how another contour disappears, and how the ones that remain tend to move away from the best Pareto-set approximation known. Continuing to pull the goal z^* downwards until it cannot be attained by any of the runs, all contours would disappear as the marginal second-order EAF became equal to zero over the whole objective space. On the other hand, moving the goal z^* upwards until it can be attained by all optimisation runs will lead to the contours of the (first-order) EAF.

An alternative way of studying the optimiser's second-order behaviour is given by the second, *centred*, moment measure of the binary random field $\{b_{\mathcal{X}}(z), z \in \mathbb{R}^d\}$. In line with the random set theory literature [7], this will be referred to as the covariance function.

Definition 8 (Covariance function). *The function* $\mathrm{cov}_{\mathcal{X}} : \mathbb{R}^d \times \mathbb{R}^d \longmapsto [-0.25, 0.25]$ *with*

$$\mathrm{cov}_{\mathcal{X}}(z_1, z_2) = \alpha_{\mathcal{X}}^{(2)}(z_1, z_2) - \alpha_{\mathcal{X}}(z_1) \cdot \alpha_{\mathcal{X}}(z_2)$$

is called the covariance function of $\mathcal{X}$.

For each pair of goals $(z_1, z_2) \in \mathbb{R}^d \times \mathbb{R}^d$, the value of the covariance function indicates the direction and strength of the relationship between the two (random) attainment indicators $b_{\mathcal{X}}(z_1)$ and $b_{\mathcal{X}}(z_2)$. If the value $\mathrm{cov}_{\mathcal{X}}(z_1, z_2)$ equals zero the two random variables are uncorrelated, i.e. there exists no linear relationship between them. On the other hand, if all elements of $\mathcal{X}$ were independent

$\text{cov}_\mathcal{X}(z_1, z_2)$ would equal zero for all $z_1, z_2 \in \mathbb{R}^d$, $z_1 \neq z_2$. Moreover, if the value $\text{cov}_\mathcal{X}(z_1, z_2)$ is positive, there is a positive correlation between $b_\mathcal{X}(z_1)$ and $b_\mathcal{X}(z_2)$, in the sense that the differences $b_\mathcal{X}(z_1) - \alpha_\mathcal{X}(z_1)$ and $b_\mathcal{X}(z_2) - \alpha_\mathcal{X}(z_2)$ will be likely to show the same sign. In other words, the attainment of goal z_1 tends to coincide with the attainment of goal z_2. The tendency *not* to attain two particular goals simultaneously is reflected by a negative covariance function value. The maximum of $\text{cov}_\mathcal{X}(z_1, z_2)$ equals 0.25 and is reached by the variance function

$$\text{var}_\mathcal{X}(z) = \text{cov}_\mathcal{X}(z, z)$$

at all $z \in \mathbb{R}^d$ where $\alpha_\mathcal{X}(z) = 0.5$. The minimum possible value of $\text{cov}_\mathcal{X}(z_1, z_2)$ is -0.25 and will be reached at any pairs of goals which *cannot* be attained together, but where each goal has probability of being attained individually equal to 0.5.

An empirical counterpart of the covariance function may be defined as follows:

Definition 9 (Empirical covariance function). *The function* $\text{cov}_n : \mathbb{R}^d \times \mathbb{R}^d \longmapsto [-0.25, 0.25]$ *with*

$$\text{cov}_n(z_1, z_2) = \alpha_n^{(2)}(z_1, z_2) - \alpha_n(z_1) \cdot \alpha_n(z_2)$$

is called the empirical covariance function of $\mathcal{X}$ (ECF).

As with the second-order EAF, visualisation of the ECF requires a workaround. In Figure 4, the pairs of goals $(z_1, z_2) \in \mathbb{R}^2 \times \mathbb{R}^2$ which exhibit a covariance above or below a certain threshold are indicated in objective space by a solid bracket beginning at one goal and ending at the other, while the contours of first-order EAF are plotted as a reference in the background. Figure 4(a) shows that goals which are likely to be attained simultaneously by MOGA-B are generally located close to each other in objective space, whereas goals which are likely to be attained in alternative to each other are located farther apart, as Figure 4(b) indicates.

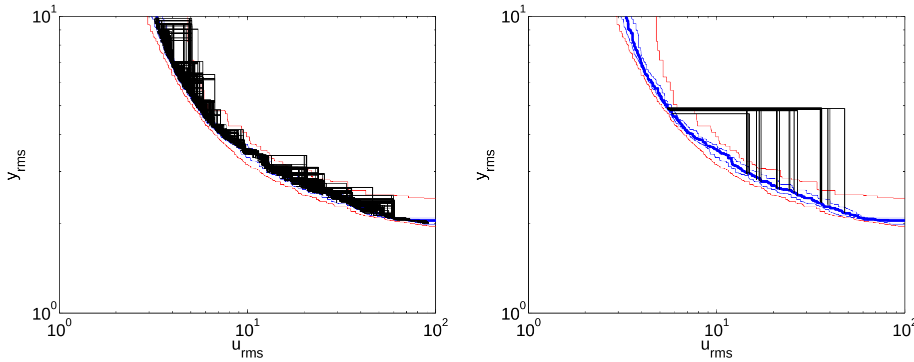
4 Comparison of optimiser performance

The performance of two multiobjective stochastic optimisers may be compared through statistical hypothesis tests of Smirnov-type based on either the (first-order) attainment function or the second-order attainment function.

4.1 First-order attainment function comparison

Given two optimisers A and B applied to the same optimisation problem independently from each other, the following two-sample test problem can be formulated for first-order comparison:

$$\begin{aligned} H_0 &: \alpha_{\mathcal{X}_A}(z) = \alpha_{\mathcal{X}_B}(z) \qquad \text{for all } z \in \mathbb{R}^d \\ &\quad \text{v.s.} \\ H_1 &: \alpha_{\mathcal{X}_A}(z) \neq \alpha_{\mathcal{X}_B}(z) \qquad \text{for at least one } z \in \mathbb{R}^d, \end{aligned}$$



(a) Covariance value greater than 0.21

(b) Covariance value less than -0.21

Fig. 4. Pairs (z_1, z_2) showing a covariance below or above a threshold (MOGA-B)

where $\alpha_{\mathcal{X}_A}(\cdot)$ and $\alpha_{\mathcal{X}_B}(\cdot)$ represent the (first-order) attainment function for optimiser A and optimiser B, respectively. Note that the equality of the two attainment functions stated in the null hypothesis H_0 is *not* equivalent to the equality of the performance of the two optimisers, since attainment functions solely address the *location* of the RNP-set distributions in objective space.

Let $\alpha_n^A(\cdot)$ and $\alpha_m^B(\cdot)$ be the EAF of optimiser A and B, respectively, after n and m optimisation runs. The above null hypothesis H_0 is rejected when the observed value of the test statistic

$$D_{n,m} = \sup_{z \in \mathbb{R}^d} |\alpha_n^A(z) - \alpha_m^B(z)|$$

is large. The test is a generalisation of the multivariate two-sided Kolmogorov-Smirnov test for two independent samples, and like the latter it is not distribution-free under H_0. However, critical values can be obtained by using the permutation argument [8], i.e. reject H_0 if $D_{n,m}$ is greater than the $(1-\alpha)$-quantile of the resulting permutation distribution of the test statistic under H_0. Note that, the permutation approach has also been used to formulate a distribution-free version for the multivariate two-sample Kolmogorov-Smirnov test [9].

4.2 Second-order attainment function comparison

A similar test problem can be formulated based on the second-order attainment function:

$$H_0 \ : \ \alpha^{(2)}_{\mathcal{X}_A}(z_1, z_2) = \alpha^{(2)}_{\mathcal{X}_B}(z_1, z_2) \qquad \text{for all } z_1, z_2 \in \mathbb{R}^d$$

v.s.

$$H_1 \ : \ \alpha^{(2)}_{\mathcal{X}_A}(z_1, z_2) \neq \alpha^{(2)}_{\mathcal{X}_B}(z_1, z_2) \qquad \text{for at least one pair } (z_1, z_2) \in \mathbb{R}^d \times \mathbb{R}^d,$$

where $\alpha^{(2)}_{\mathcal{X}_A}(\cdot,\cdot)$ and $\alpha^{(2)}_{\mathcal{X}_B}(\cdot,\cdot)$ represent the second-order attainment function for optimiser A and optimiser B, respectively. Again, the equality of the two second-order attainment functions stated in H_0 is *not* equivalent with the equality of the performance of the two optimisers. However, the Smirnov-like test formulated in the following should already inherit more statistical power than the respective test based on the (first-order) attainment function as it uses more information from the data. In other words, not rejecting H_0 is generally less likely to be a wrong decision.

As before, construct a permutation test which rejects H_0 if

$$D^{(2)}_{n,m} = \sup_{z_1,z_2\in\mathbb{R}^d} |\alpha^{A(2)}_n(z_1,z_2) - \alpha^{B(2)}_m(z_1,z_2)|$$

is greater than the $(1-\alpha)$-quantile of the resulting permutation distribution of the test statistic under H_0. Here, $\alpha^{A(2)}_n(\cdot,\cdot)$ and $\alpha^{B(2)}_n(\cdot,\cdot)$ are the second-order EAF of optimiser A and optimiser B, respectively, after n and m optimisation runs.

5 Experimental results

To illustrate how the first-order and the second-order attainment functions may be used to study the performance of stochastic multiobjective optimisers, two application examples are considered here.

The first example consists of the optimisation of a multiobjective Linear-Quadratic-Gaussian controller design problem proposed by Barratt and Boyd [10] under controller complexity constraints, as formulated in [11], with multiobjective genetic algorithms (MOGA). Two MOGAs were applied to the problem, one without sharing or mating restriction (MOGA-A) and another one with sharing and mating restriction in the decision variable domain (MOGA-B), as described in [12]. Each algorithm was run 21 times for 100 generations, and the cumulative set of non-dominated objective vectors found in each run was taken as the outcome of that run.

The second example consists of the optimisation of a multiobjective Travelling Salesman Problem (TSP) by stochastic local search techniques. The problem itself is a benchmark instance based on a pair of 100-city TSP instances, kroA100.tsp and kroB100.tsp, available at TSPLIB[5]. The optimiser used was the Pareto local search (PLS) algorithm proposed in Paquete *et al.* [13], which uses an archive of non-dominated solutions and an acceptance criterion which takes into account the concept of Pareto optimality. Two different neighbourhoods were considered, the standard 2-opt neighbourhood and a 2-opt extension proposed by Bentley [14] and known as 2H-opt, leading to two variants of the algorithm, respectively PLS-A and PLS-B. Each variant was run 25 times until all solutions in the archive were locally Pareto-optimal, and the corresponding non-dominated objective vectors were taken as the outcome of each run. Table 1

[5] http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/

Table 1. Pareto-set approximation statistics

Optimiser	No. of runs	No. of elements min	average	max
MOGA-A	21	48	120.38	191
MOGA-B	21	87	170.95	259
PLS-A	25	1973	2386.1	2891
PLS-B	25	2052	2541.5	3032

Table 2. Hypothesis test results ($\alpha = .05$)

Optimiser	Hypothesis test	Test statistic	Critical value	p-value	decision
MOGA	1st-order EAF	0.571	0.571	0.091	do not reject H_0
MOGA	2nd-order EAF	0.762	0.714	0.016	reject H_0
PLS	1st-order EAF	0.680	0.560	0.004	reject H_0
PLS	2nd-order EAF	0.840	0.720	0.002	reject H_0

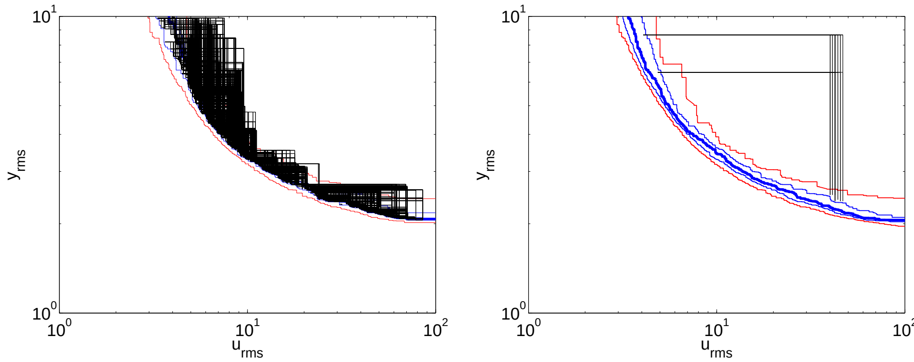
gives some additional information about the Pareto-set approximations obtained in the two examples.

For each set of runs, the (first-order) EAF was computed at all relevant goals in objective space (where the EAF exhibits transitions) and a record was kept of which runs attained which goals. This record was used to compute the second-order EAF and the corresponding empirical covariance function values. The observed values of the test statistic of each of the two Smirnov-like tests were determined by pooling the outcomes of the two optimisers involved in the test, and computing the first-order EAF of the resulting pooled sample. Then, the maximum absolute difference between the two individual EAFs to be compared (whether first or second-order) was determined over the the set of goals (or pairs of goals) at which the corresponding EAF of the pooled sample exhibited transitions. Finally, the permutation distribution of the two test statistics was simulated by considering 10000 random permutations of the runs in the corresponding pooled samples, assigning half of the runs to each algorithm variant, and recomputing the test statistic for each permutation. The results of the tests are summarised in Table 2, and will be discussed separately for each example.

5.1 First example

First-order EAF plots for the outcomes of MOGA-A and MOGA-B have already been presented in Figure 2. Visual inspection suggests that MOGA-B may be more likely to attain certain goals than MOGA-A, as the weight of the EAF of MOGA-B seems to be more concentrated downwards and to the left than in the case of MOGA-A. However, the hypothesis test based on the first-order EAF does not lead to the rejection of the null hypothesis at the 0.05-significance level, indicating that any differences between the first-order attainment functions of the two algorithms are not statistically significant.

With respect to second-order behaviour, MOGA-B was shown in Figure 4 to exhibit strong negative covariance at certain pairs of goals. In comparison,



(a) Covariance value greater than 0.21 (MOGA-A)

(b) Difference between second-order EAFs exceeding the critical value

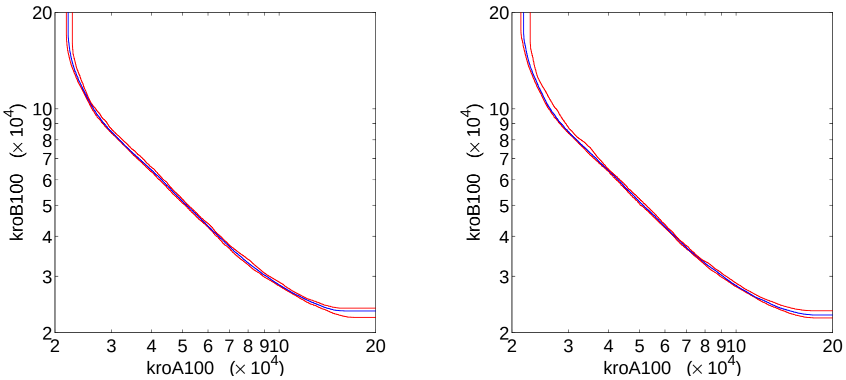
Fig. 5. Additional results (first example)

MOGA-A does not exhibit as strong a negative covariance anywhere in objective space. Thus, the effect of the niche induction techniques used in MOGA-B seems to be the imposition of stronger dependencies between the elements of the Pareto-set approximations. As for strong positive covariance, the analogue to Figure 4(a) for MOGA-A is presented in Figure 5(a). MOGA-A seems to exhibit stronger positive covariance between more distant goals in objective space, which may simply be due to the fact that the corresponding Pareto-set approximations generally contain fewer points than those produced by MOGA-B.

Finally, the hypothesis test based on the second-order EAF does lead to rejection of the null hypothesis at the 0.05-significance level. Hence, there is evidence for a statistically significant difference between the second-order attainment functions and, consequently, in the performance of the two algorithms. Figure 5(b) represents the pairs of goals where the absolute difference between second-order EAFs exceeded the critical value, with the contours of the first-order EAF of the pooled sample plotted in the background. Closer inspection of the individual second-order EAFs indicates that MOGA-B has a greater probability of attaining these pairs of goals than MOGA-A, which supports the idea that niche induction techniques had a positive effect on the performance of MOGA-B.

5.2 Second example

First-order EAF contour plots for the outcomes of PLS-A and PLS-B on the multiobjective TSP example are presented in Figure 6. Due to the characteristics of both EAFs, only the ϵ-, 0.5- and $(1-\epsilon)$-levels are displayed. The hypothesis test based on the first-order EAF leads to a clear rejection of the null hypothesis,



(a) PLS-A (b) PLS-B

Fig. 6. EAF contour plots (second example)

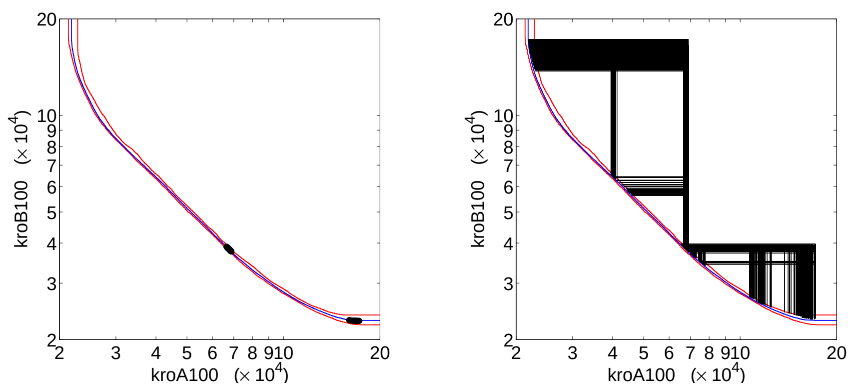
even at significance level 0.01. In Figure 7(a), the goals at which the critical value of the test statistic is exceeded are marked with black dots. Inspection of the individual EAFs shows that PLS-B is more likely to attain those goals than PLS-A, which is hardly surprising since the 2H-opt neighbourhood contains the 2-opt neighbourhood and both algorithms were run until the archive contained exclusively local optima.

Also not surprisingly, the hypothesis test based on the second-order EAF leads to an even clearer rejection of the null hypothesis, with absolute differences in second-order EAF values greater than the critical value being observed at many pairs of goals (see Figure 7(b)). Again, the probability of PLS-B attaining these pairs of goals is greater than that of PLS-A, confirming the benefits of a larger neighbourhood under the experimental conditions considered.

6 Concluding remarks

In this paper, it was shown how the performance of multiobjective optimisers may be studied using the attainment function approach. Whereas the first-order attainment function expresses the probability of given goals being attained independently from each other in one optimisation run, the second-order attainment function and the covariance function consider the probability of pairs of goals being attained simultaneously and, thus, take the dependence between the nondominated elements of individual Pareto-set approximations into account.

By relating the probability of simultaneously attaining given pairs of goals with the probability of attaining them independently from each other, the covariance function provides information about the behaviour of the optimiser. In



(a) Difference between first-order EAFs exceeding the critical value

(b) Difference between second-order EAFs exceeding the critical value

Fig. 7. Hypothesis test results (second example)

the first example, the use of niche induction techniques in a MOGA lead to a different covariance structure, suggesting stronger dependencies between the points in each random Pareto-set approximation when those techniques were used.

The second-order attainment function, on the other hand, may be said to extend the first-order attainment function in the task of assessing the ability of an optimiser to consistently produce good solutions, not only in isolation, but also in combination with each other. This is particularly relevant when *comparing* the performance of two optimisers, as differences in performance may not be related only to the quality of the non-dominated solutions produced, but also to whether or not they are likely to occur together. In the first example, only the hypothesis test based on the second-order attainment function was able to detect significant performance differences between the two algorithms. In the second example, the first-order attainment function could already distinguish the performance of the two optimisers in a statistically significant way. In that case, it is reasonable to expect even larger differences between second-order attainment functions to arise, as was observed in this example.

The two examples presented here show that the attainment function approach to optimiser assessment is currently applicable to bi-objective problems, using realistic sample sizes (numbers of runs) and large non-dominated point sets. Determining critical values for the second-order EAF hypothesis tests is by far the most computationally demanding aspect of the approach (6.5 hours and 40 days on a single Athlon MP 1900+ processor for the first and second examples, respectively, using 10000 permutations). Clearly, there is still much scope for work on the computational aspects of the approach, especially as the number of objectives and/or the size of the data sets grow.

References

1. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Grunert da Fonseca, V.: Performance assessment of multiobjective optimizers: An analysis and review. IEEE Transactions on Evolutionary Computation **7** (2003) 117–132
2. Goutsias, J.: Modeling random shapes: An introduction to random closed set theory. In Haralick, R.M., ed.: Mathematical Morphology: Theory and Hardware. Oxford Series in Optical & Imaging Sciences. Oxford University Press, New York (1996)
3. Grunert da Fonseca, V., Fonseca, C.M., Hall, A.O.: Inferential performance assessment of stochastic optimisers and the attainment function. In Zitzler, E., Deb, K., Thiele, L., Coello Coello, C.A., Corne, D., eds.: Evolutionary Multi-Criterion Optimization. First International Conference, EMO 2001, Zürich, Switzerland, March 2001, Proceedings. Number 1993 in Lecture Notes in Computer Science. Springer Verlag, Berlin (2001)
4. Justel, A., Peña, D., Zamar, R.: A multivariate Kolmogorov-Smirnov test of goodness of fit. Statistics and Probability Letters **35** (1997) 251–259
5. Fonseca, C.M., Fleming, P.J.: On the performance assessment and comparison of stochastic multiobjective optimizers. In Voigt, H.M., Ebeling, W., Rechenberg, I., Schwefel, H.P., eds.: Parallel Problem Solving from Nature – PPSN IV. Number 1141 in Lecture Notes in Computer Science. Springer Verlag, Berlin, Germany (1996) 584–593
6. Shaw, K.J., Nortcliffe, A.L., Thompson, M., Love, J., Fleming, P.: Assessing the performance of multiobjective genetic algorithms for optimization of a batch process scheduling problem. In: Proceedings of the Congress on Evolutionary Computation (CEC99). Volume 1., Washington DC (1999) 37–45
7. Stoyan, D., Kendall, W.S., Mecke, J.: Stochastic Geometry and its Applications. 2nd edn. Wiley Series in Probability and Statistics. Wiley & Sons, Chichester (1995)
8. Good, P.I.: Permutation Tests: A practical Guide to Resampling Methods for Testing Hypotheses. 2nd edn. Springer Series in Statistics. Springer Verlag, New York (2000)
9. Bickel, P.J.: A distribution free version of the Smirnov two sample test in the p-variate case. The Annals of Mathematical Statistics **40** (1969) 1–23
10. Barratt, C., Boyd, S.: Example of exact trade-offs in linear control design. IEEE Control Systems Magazine **9** (1989) 46–52
11. Fonseca, C.M., Fleming, P.J.: Multiobjective optimal controller design with genetic algorithms. In: Proc. IEE Control'94 International Conference. Volume 1., Warwick, U.K. (1994) 745–749
12. Fonseca, C.M., Fleming, P.J.: Multiobjective genetic algorithms made easy: selection, sharing and mating restriction. In: First IEE/IEEE International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications, Sheffield, U.K. (1995) 45–52
13. Paquete, L., Chiarandini, M., Stützle, T.: Pareto local optimum sets in the biobjective traveling salesman problem: An experimental study. In Gandibleux, X., Sevaux, M., Sörensen, K., T'kindt, V., eds.: Metaheuristics for Multiobjective Optimisation. Volume 535 of Lecture Notes in Economics and Mathematical Systems. Springer Verlag (2004) 177–200
14. Bentley, J.: Fast algorithms for geometric traveling salesman problems. ORSA Journal on Computing **4** (1992) 387–411