The Attainment-Function Approach to Stochastic Multiobjective Optimiser Assessment and Comparison

C. M. Fonseca,[1] V Grunert da Fonseca[1,2]

[1] Centre for Intelligent Systems, University of Algarve [2] INUAF

CSI
CENTRO DE SISTEMAS INTELIGENTES

Outline

1. Background
2. The attainment function approach
3. Second-order moment measures
4. Comparing optimiser performance
5. Experimental results
6. Quality indicators revisited
7. Modelling performance
8. An integrated view of performance
9. Concluding remarks

CENTRO DE SISTEMAS INTELIGENTES

1. Background

- Optimiser quality is intimately related to
  - the quality of the solutions produced
  - the time taken to produce them
  - the difficulty of the problem considered
- Problem difficulty depends on problem size and/or configuration (hard to quantify)
- For stochastic optimisers, both run time and solution quality are random, and are associated with some probability distribution
- An optimisation run samples from the corresponding distributions, much like an estimator

CENTRO DE SISTEMAS INTELIGENTES

1. Background

- Performance criteria (for estimators *and* optimisers):

  Location Typically, how close to the true/theoretical value?

  Spread Typical variability, best and worst-case behaviour

  Tractability Can this behaviour be modelled?

- Current practice...
  - Solution quality and run time usually studied independently from each other, other factors kept fixed
  - Emphasis on one-off and typical behaviour (especially location), tractability not always a concern

CSI
CENTRO DE SISTEMAS INTELIGENTES

1.1. Run time (Hoos and Stützle, 1998)

1.1.1. Experimental setup

- Execute algorithm n times on a given problem until a valid solution is found or cutoff time $t_{\max}$ is reached
- Record number of successful runs, k, and the corresponding run time of each one, t_i, $i = 1, \ldots, k$

1.1.2. Data analysis (univariate)

- Empirical distribution function, sample statistics
- Estimate mean run time from experimental data, accounting for unsuccessful runs:

$$\hat{E}(T) = \frac{1}{k}\sum_{i=1}^{k} t_i + \frac{n-k}{k} t_{\max}$$

CENTRO DE SISTEMAS INTELIGENTES

1.2. Solution quality – Single objective

1.2.1. Experimental setup

- Execute algorithm n times on a given problem until a given stopping criterion is met (maximum runtime, convergence, etc.)
- Record best objective value found in each run, x_i, for $i = 1, \ldots, n$

1.2.2. Data analysis (univariate)

- Empirical distribution function, sample statistics
- Normality can seldom be assumed
- Hypothesis tests (applies to run time, too)

CENTRO DE SISTEMAS INTELIGENTES

1.3. Solution quality – Multiple objectives

1.3.1. Experimental setup

- Execute algorithm n times on a given problem until a given stopping criterion is met (maximum runtime, convergence, etc.)
- Record all non-dominated objective vectors found in each run, $\{x_{1i}, x_{2i}, \ldots, x_{m_i i}\}$, for $i = 1, \ldots, n$

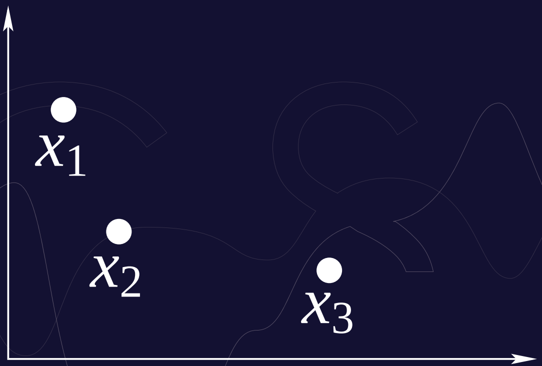
1.3.2. Data analysis

- Each $\{x_{1i}, x_{2i}, \ldots, x_{m_i i}\}$ is a set of non-dominated points in objective space
- These non-dominated point (NDP) sets are random
- How can their stochastic behaviour be described?

CENTRO DE SISTEMAS INTELIGENTES

1.3. Solution quality – Multiple objectives

1.3.3. Example



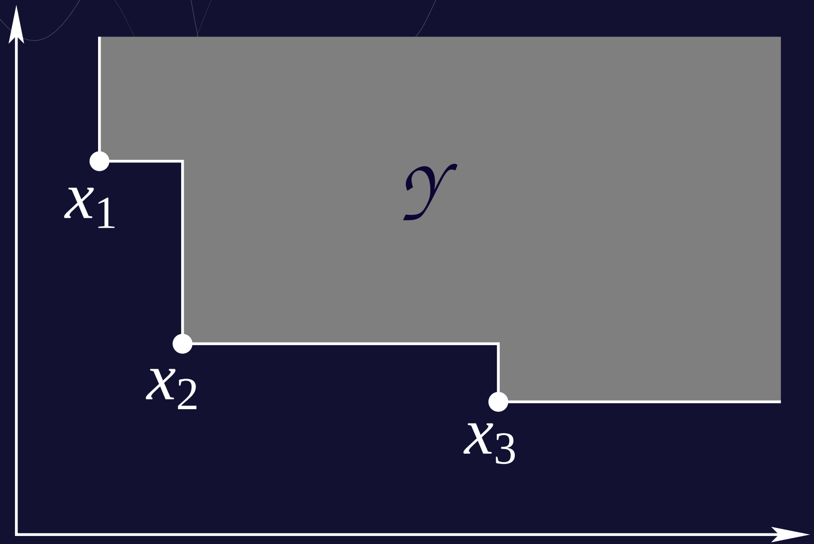
1.4. Quality indicators

- Transform NDP sets into real values or real vectors
- More conventional statistical analysis
- Lose some information in the process (how much?)

2. The attainment function approach

(Grunert da Fonseca *et al*, 2001)

- Considers the region attained by each non-dominated point set
- Studies the set distributions directly through their moments
- Higher-order moments provide additional information



2. The attainment function approach

Definition 1 (Random non-dominated point set)

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

Definition 2 (Attained set)

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

- The distributions of random sets $\mathcal{X}$ and $\mathcal{Y}$ are equivalent

Definition 3 (Attainment indicator)

$$b_{\mathcal{X}}(z) = \mathbf{I}\{\mathcal{X} \trianglelefteq z\}$$

- The binary random field $\{b_{\mathcal{X}}(z), z \in \mathbb{R}^d\}$ provides yet another way to look at the distribution of $\mathcal{X}$

CENTRO DE SISTEMAS INTELIGENTES

2.1. First-order attainment function

Definition 4 (Attainment function)

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

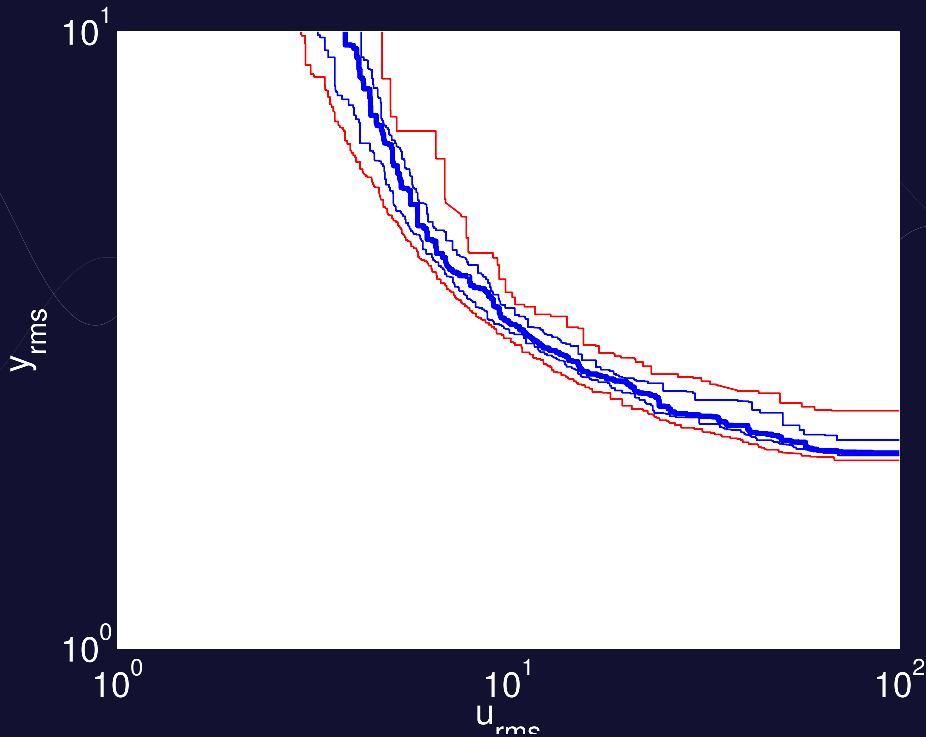
- Probability of attaining a given goal z
- First-order moment measure of the binary random field $\{b_{\mathcal{X}}(z), z \in \mathbb{R}^d\}$
- Describes the location of the Pareto-set approximations
- Reduces to the multivariate distribution function when $M = 1$
- Can be estimated from experimental data

Definition 5 (Empirical attainment function (EAF))

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

2.1. First-order attainment function

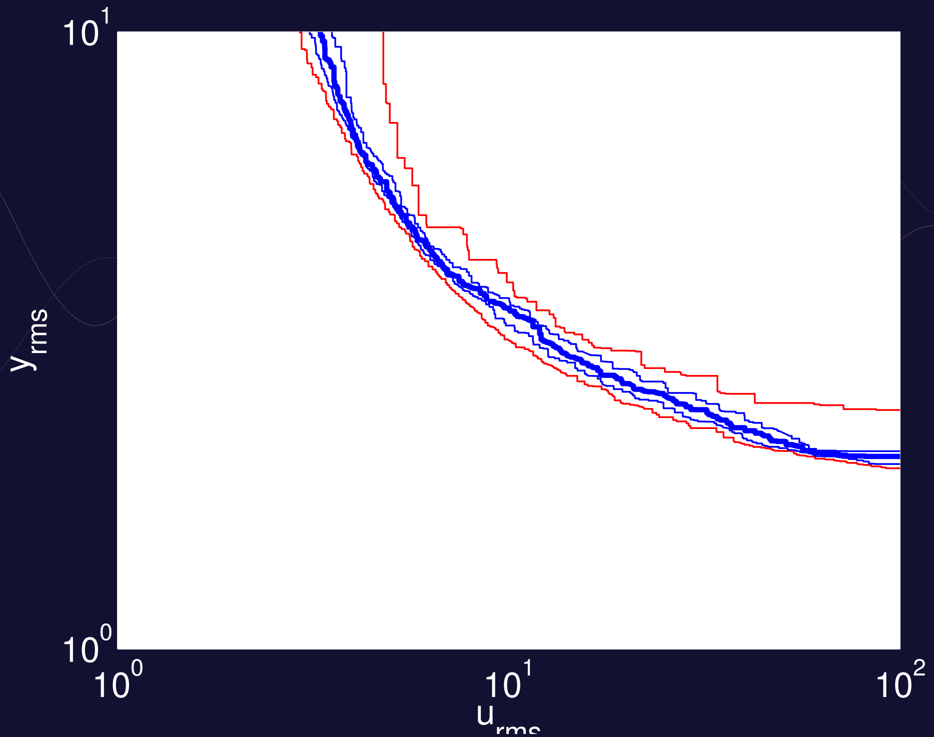
2.1.1. EAF example



MOGA-A

2.1. First-order attainment function

2.1.2. Another EAF example



MOGA-B

3. Second-order moment measures

(Fonseca *et al*, 2005)

3.1. Second-order attainment function

Definition 6 (Second-order attainment function)

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

- Probability of attaining two goals simultaneously
- Second, non-centred, moment of $\{b_{\mathcal{X}}(z), z \in \mathbb{R}^d\}$
- Can be estimated from experimental data

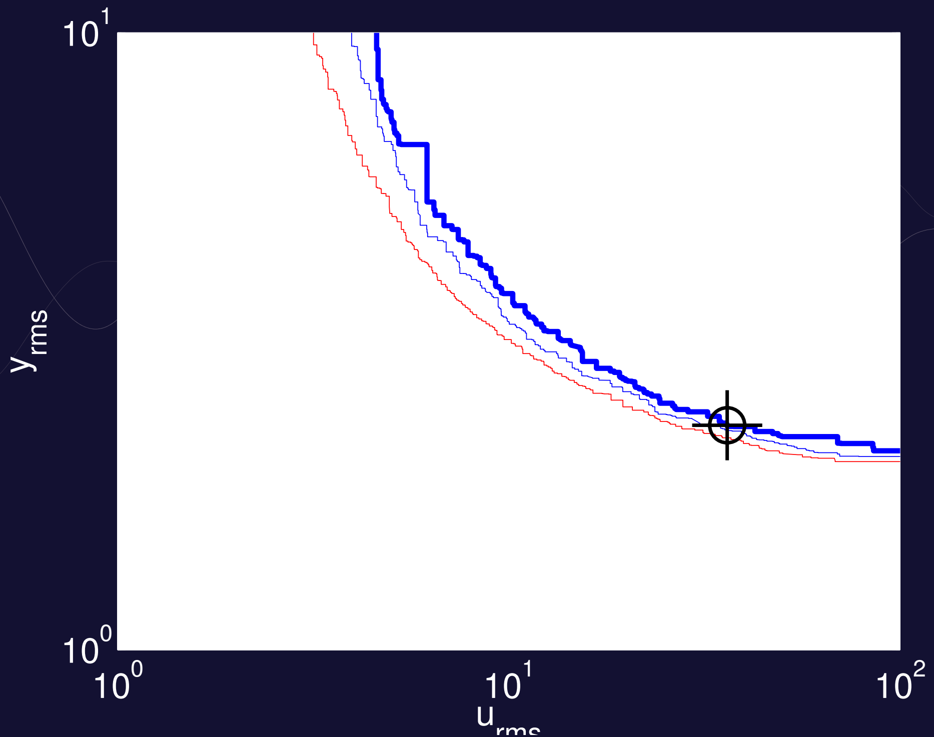
Definition 7 (Second-order empirical attainment function)

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

CENTRO DE SISTEMAS INTELIGENTES

3.1.1. Second-order EAF visualization

With a fixed goal $z^* \in \mathbb{R}^2$

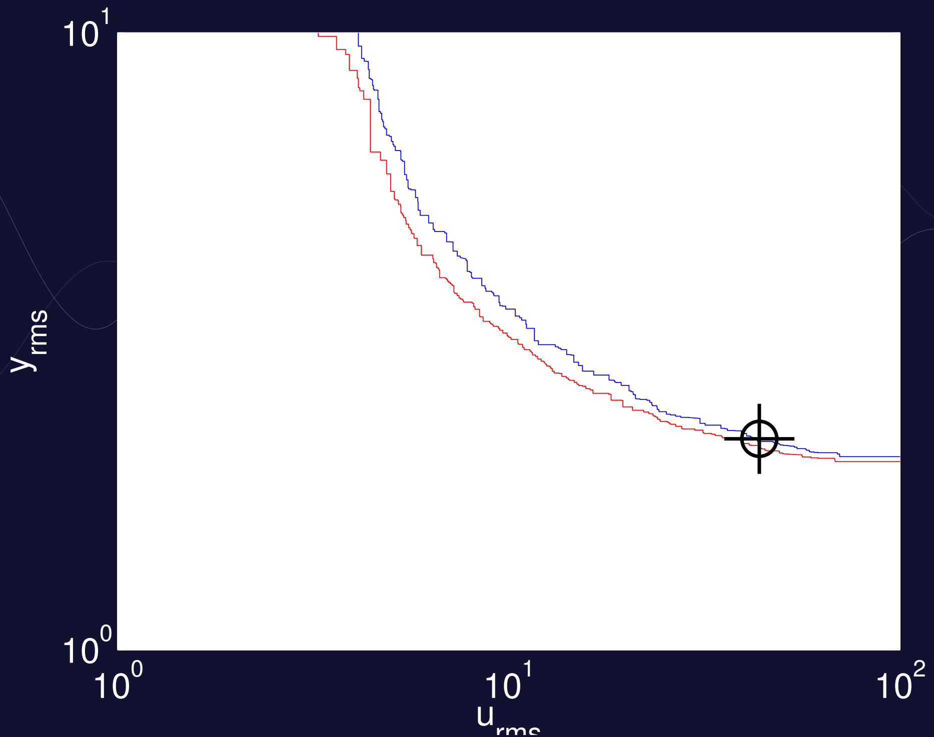


MOGA-A

CSI
CENTRO DE SISTEMAS INTELIGENTES

3.1.1. Second-order EAF visualization

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



MOGA-A

CENTRO DE SISTEMAS INTELIGENTES

3.2. Covariance function

Definition 8 (Covariance function)

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

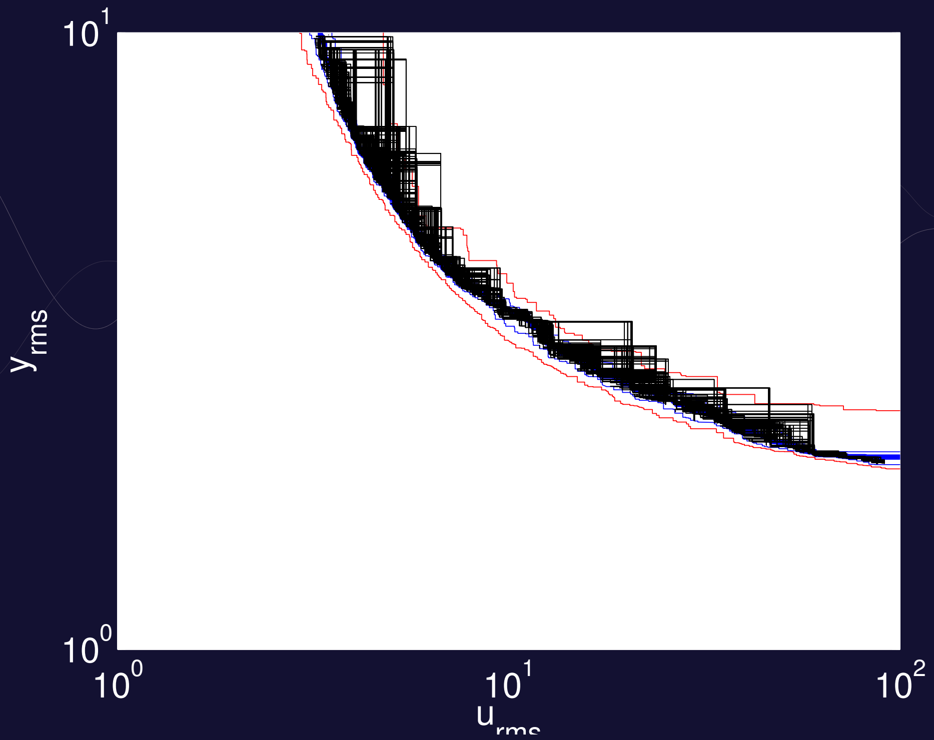
- Second, centred, moment of $\{b_{\mathcal{X}}(z), z \in \mathbb{R}^d\}$
- Indicates how likely two different goals are to be attained together in the same run in comparison to being attained independently in different runs
- Can be estimated from experimental data

Definition 9 (Empirical covariance function (ECF))

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

3.2.1. Empirical covariance function visualization

Covariance function values greater than 0.21

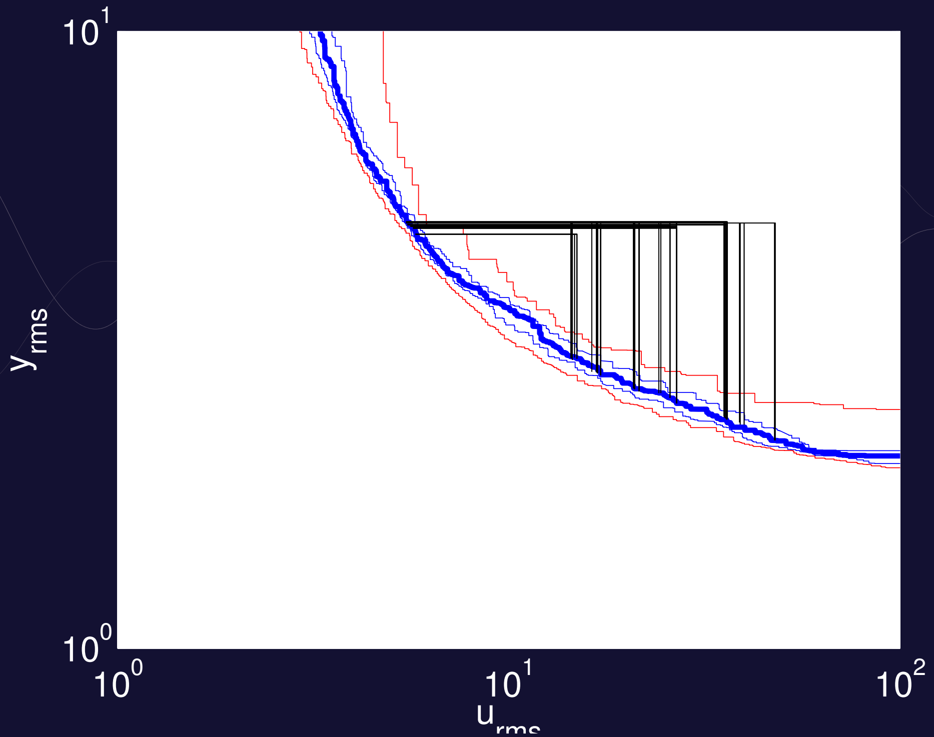


MOGA-B

CENTRO DE SISTEMAS INTELIGENTES

3.3. Empirical covariance function visualization

Covariance function values less than -0.21



MOGA-B

CENTRO DE SISTEMAS INTELIGENTES

4. Comparing optimiser performance

Performance may be compared through EAF-based hypothesis tests

4.1. First-order attainment function comparison

$$H_0 \; : \; \alpha_{\mathcal{X}_A}(z) = \alpha_{\mathcal{X}_B}(z) \qquad \text{for all } z \in \mathbb{R}^d$$

vs.

$$H_1 \; : \; \alpha_{\mathcal{X}_A}(z) \neq \alpha_{\mathcal{X}_B}(z) \qquad \text{for at least one } z \in \mathbb{R}^d,$$

- Reject if the test statistic $D_{n,m} = \sup_{z \in \mathbb{R}^d} |\alpha_n^A(z) - \alpha_m^B(z)|$ is large
- Permutation argument allows critical values to be obtained

CENTRO DE SISTEMAS INTELIGENTES

4.2. Second-order attainment function comparison

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

vs.

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

- Reject if

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

  exceeds the $(1-\alpha)$-quantile of the permutation distribution of the test statistic under H_0

- One-sided tests could be formulated in a similar way

5. Experimental results

- First example based on a multiobjective LQG controller design problem, under complexity constraints
  - MOGA-A (no niching) vs. MOGA-B (sharing and mating restriction)
  - Two sets of 21 runs for 100 generations
- Second example based on a multiobjective TSP instance
  - PLS-A (2-opt neighbourhood) vs. PLS-B (2H-opt)
  - Two sets of 25 runs until archive contained only local optima
- 10000 permutations used to estimate critical values

CENTRO DE SISTEMAS INTELIGENTES

5.1. Pareto-set approximation statistics

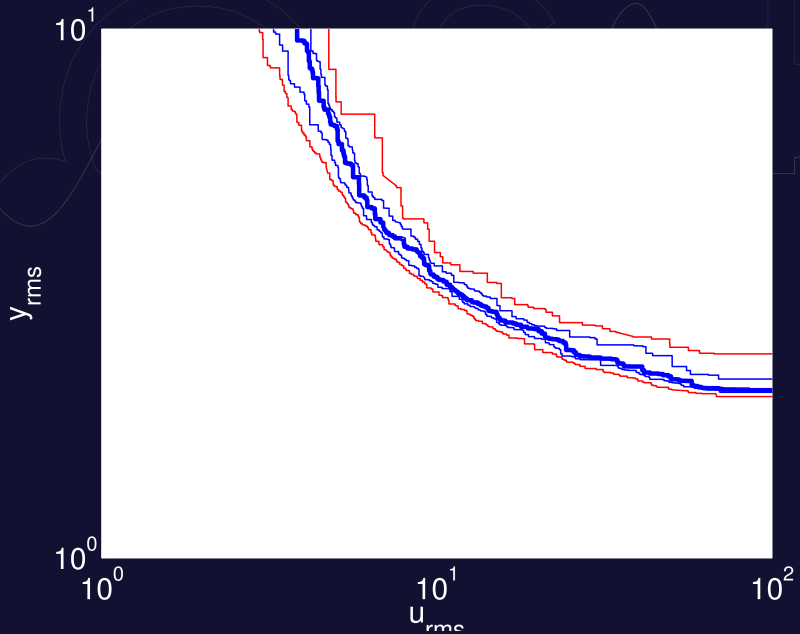
Optimiser	No. of runs	No. of elements		
		min	average	max
MOGA-A	21	48	120.38	191
MOGA-B	21	87	170.95	259
PLS-A	25	1973	2386.1	2891
PLS-B	25	2052	2541.5	3032

CENTRO DE SISTEMAS INTELIGENTES

5.2. First-example

5.2.1. Empirical covariance function

Covariance function values less than -0.21



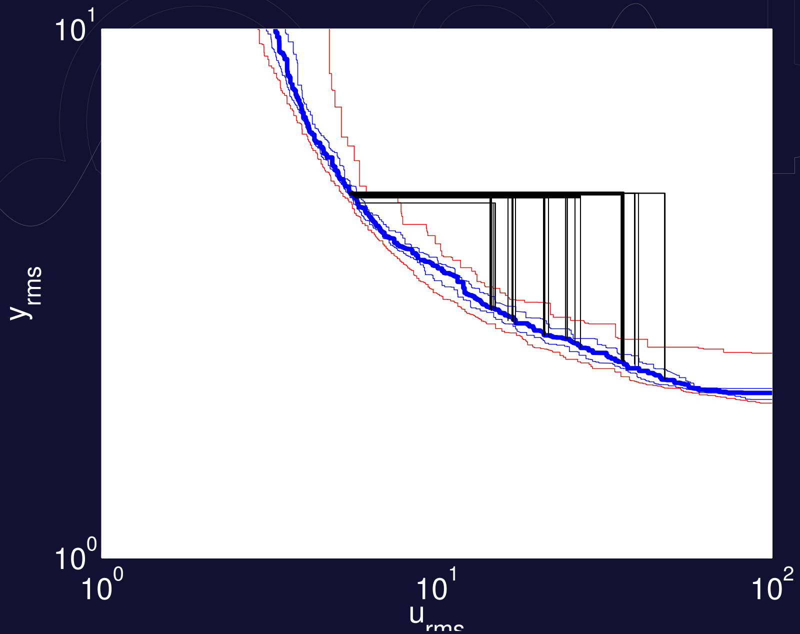
MOGA-A

CSI
CENTRO DE SISTEMAS INTELIGENTES

5.2. First-example

5.2.1. Empirical covariance function

Covariance function values less than -0.21



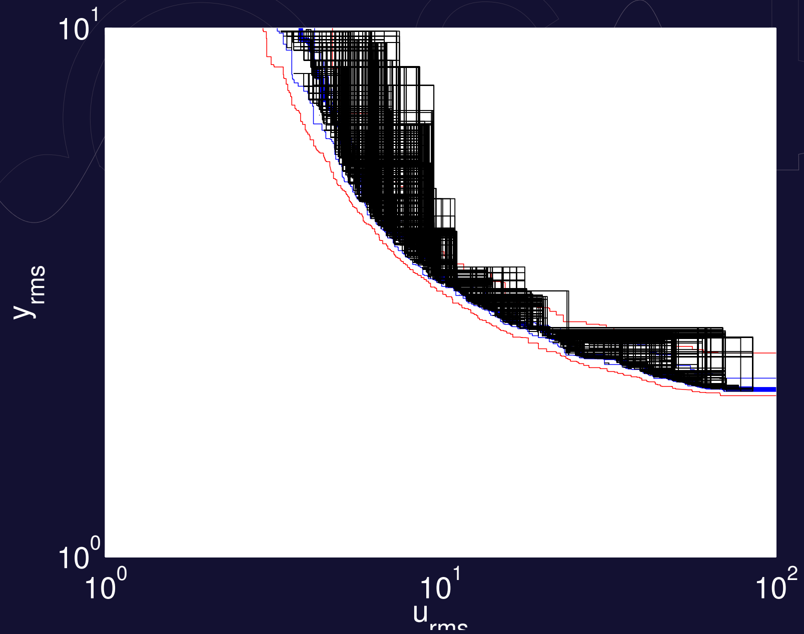
MOGA-B

CSI
CENTRO DE SISTEMAS INTELIGENTES

5.2. First-example

5.2.1. Empirical covariance function

Covariance function values greater than 0.21



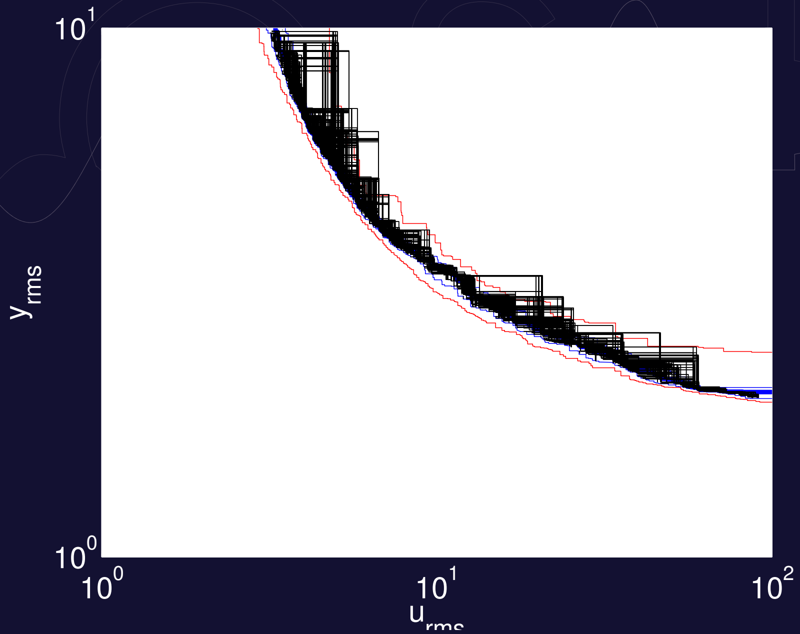
MOGA-A

CSI
CENTRO DE SISTEMAS INTELIGENTES

5.2. First-example

5.2.1. Empirical covariance function

Covariance function values greater than 0.21

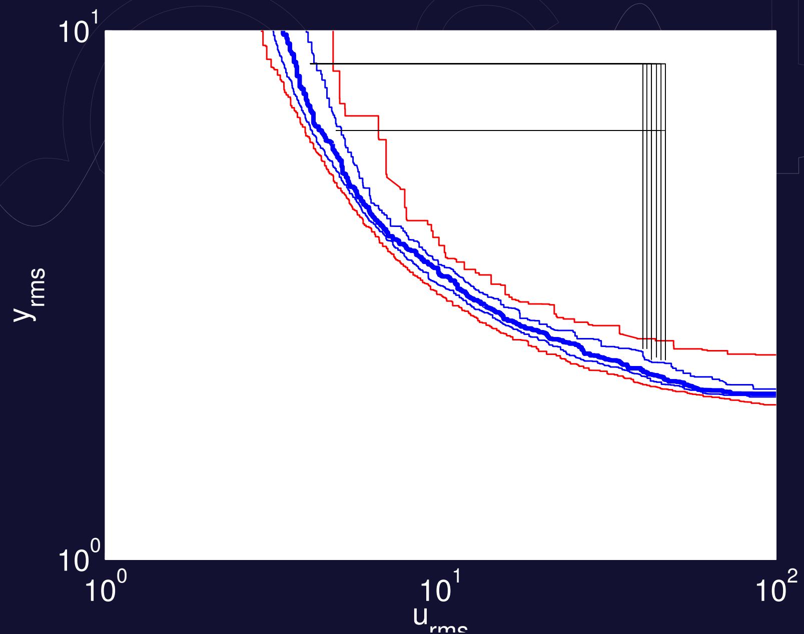


MOGA-B

CSI
CENTRO DE SISTEMAS INTELIGENTES

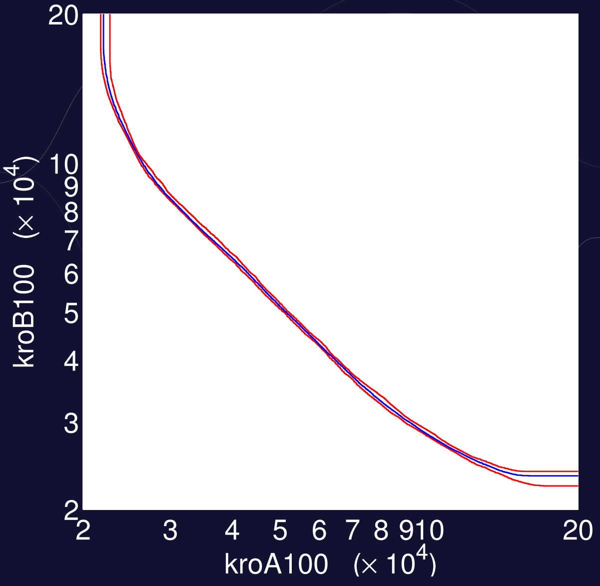
5.2. First-example

5.2.2. Second-order EAF test ($\alpha = .05$)

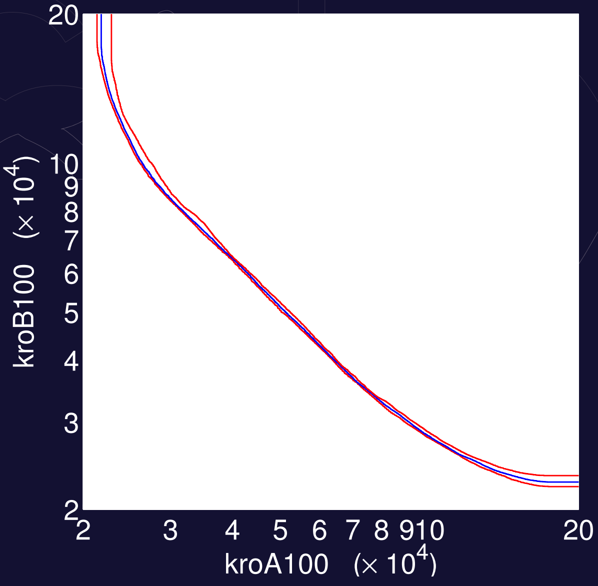


5.3. Second example

5.3.1. EAF contour plots



PLS-A

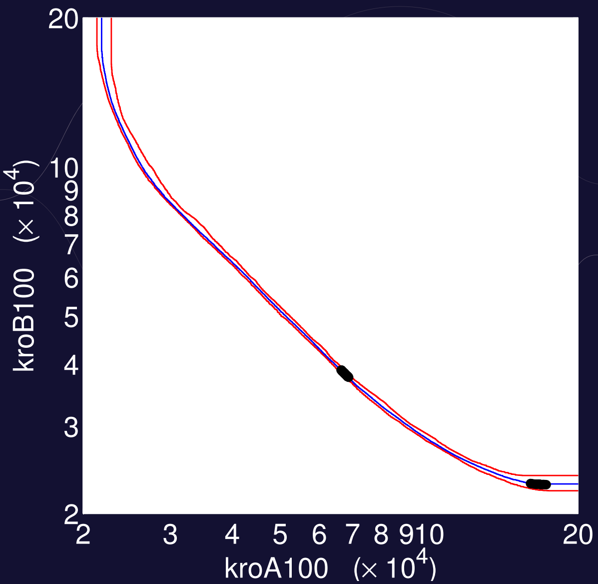


PLS-B

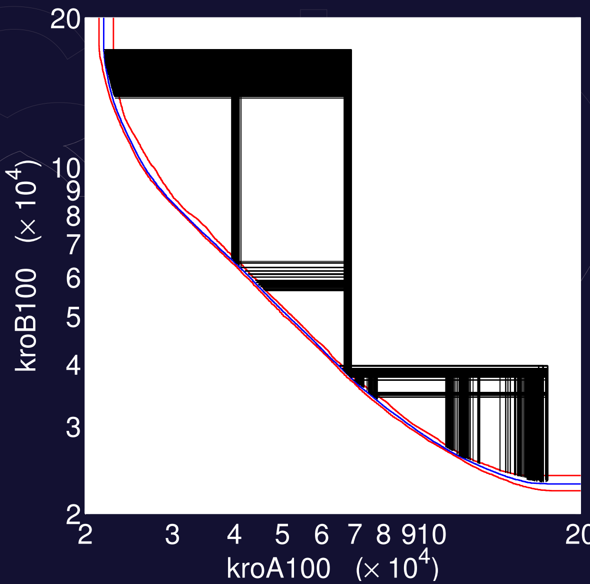
CENTRO DE SISTEMAS INTELIGENTES

5.3. Second example

5.3.2. Hypothesis test results ($\alpha = .05$)



1st-order



2nd-order

CENTRO DE SISTEMAS INTELIGENTES

5.4. Result summary

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

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

6. Quality indicators revisited

- Quality indicators transform NDP sets into real values
- If the multiobjective optimiser is stochastic, both outcome NDP sets and quality indicator values will be random
- The quality indicator distribution depends only on the underlying random NDP set distribution
- It must be possible to describe it as a function of an attainment function of sufficiently high order
- Results should be both of theoretical and of practical value

6.1. Example: The unary ε-indicator

- May be written as (Z is the reference set)

$$
\begin{aligned}
I_{\varepsilon,Z}(X) &= \inf\left\{\varepsilon \in \mathbb{R}^+ : X \trianglelefteq \varepsilon \cdot z,\ \forall z \in Z\right\} \\
&= \inf\left\{\varepsilon \in \mathbb{R}^+ : \prod_{z \in Z} b_X(\varepsilon \cdot z) = 1\right\}
\end{aligned}
$$

- Has distribution function

$$P\left[I_{\varepsilon,Z}(\mathcal{X}) \leq c\right] = \alpha_{\mathcal{X}}^{(k)}(c \cdot z_1, c \cdot z_2, \ldots, c \cdot z_k)$$

given a reference set $Z = \{z_1, z_2, \ldots, z_k\}$

- In particular, when $Z = \{z\}$,

$$P\left[I_{\varepsilon,Z}(\mathcal{X}) \leq c\right] = \alpha_{\mathcal{X}}(c \cdot z)$$

7. Modelling performance

7.1. Run time (Hoos and Stützle, 1998)

- Run-time distributions may be related to the exponential distribution
- Doing so may help decide, e.g., when an algorithm should be restarted
- Other survival-time distributions (e.g., Weibull)

7.2. Solution quality (Hüssler et al., 2003)

- Solution-quality distributions may be related to the Weibull extreme value distribution, at least in certain ideal cases
- The parameters of the Weibull distribution generally depend on the function being optimised, and may give information about, e.g., whether the optimum is likely to be close or far away

8. An integrated view of performance

- The best-so-far trace of a single-objective optimisation run represents an observed run-time/solution-quality tradeoff
- In general, the observed performance in an n-objective optimisation run can be described through an augmented, $n+1$-objective NDP set, including the run-time dimension
- The distribution of such NDP sets may be studied through empirical attainment functions
- The Weibull distribution is both an extreme value distribution and a distribution used in survival analysis, and has been shown to be useful in modelling both run-time and solution-quality behaviour of optimisers
- Parametric models of attainment functions, valid under certain ideal conditions, are (still) under development.

CSI
CENTRO DE SISTEMAS INTELIGENTES

9. Concluding remarks

- Optimiser performance involves many criteria
- Many of the questions faced when addressing optimiser performance are similar to those addressed at the optimisation stage
- Other relevant questions pertain to experimental methodology
- First-order attainment function describes the distribution of random NDP sets in terms of location
- Covariance function provides insight into the dependencies within the NDP sets
- Second-order attainment function favours multiple good solutions

CSI
CENTRO DE SISTEMAS INTELIGENTES

9. Concluding remarks

- Hypothesis tests enable comparison of optimisers
- Multiple comparisons still need to be addressed
- The attainment-function methodology is fully usable with two objectives (computational developments still needed for more dimensions)
- Provides a theoretical basis for analysing other quality indicators
- Supports combined time-quality performance evaluation
- Scalability of optimisers is also a performance-related question
- Did you say "sadistics"?

CENTRO DE SISTEMAS INTELIGENTES