

# Dominance-Based Pareto-Surrogate for Multi-Objective Optimization

Ilya Loshchilov, Marc Schoenauer, Michèle Sebag

► **To cite this version:**

Ilya Loshchilov, Marc Schoenauer, Michèle Sebag. Dominance-Based Pareto-Surrogate for Multi-Objective Optimization. Simulated Evolution And Learning (SEAL-2010), Dec 2010, Kanpur, India. 2010. <inria-00522653>

**HAL Id: inria-00522653**

**<https://hal.inria.fr/inria-00522653>**

Submitted on 1 Oct 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Dominance-Based Pareto-Surrogate for Multi-Objective Optimization

Ilya Loshchilov<sup>1,2</sup>, Marc Schoenauer<sup>1,2</sup>, and Michèle Sebag<sup>2,1</sup>

<sup>1</sup> TAO Project-team, INRIA Saclay - Île-de-France\*\*

<sup>2</sup> Laboratoire de Recherche en Informatique (UMR CNRS 8623)

Université Paris-Sud, 91128 Orsay Cedex, France

FirstName.LastName@inria.fr

**Abstract.** Mainstream surrogate approaches for multi-objective problems build one approximation for each objective. Mono-surrogate approaches instead aim at characterizing the Pareto front with a single model. Such an approach has been recently introduced using a mixture of regression Support Vector Machine (SVM) to clamp the current Pareto front to a single value, and one-class SVM to ensure that all dominated points will be mapped on one side of this value. A new mono-surrogate EMO approach is introduced here, relaxing the previous approach and modelling Pareto dominance within the rank-SVM framework. The resulting surrogate model is then used as a filter for offspring generation in standard Evolutionary Multi-Objective Algorithms, and is comparatively validated on a set of benchmark problems.

## 1 Introduction

This paper is concerned with evolutionary Multi-Objective Optimization (EMO) [2], and most specifically focuses on designing and using surrogate models in order to speed up the evolutionary search. Surrogate models, namely computationally light estimates of the objective function, have been extensively used in Evolutionary Algorithms (EAs) since the 1990's [6], as the Achilles's heel of EAs is known to be the high number of times the objective function has to be computed. This high number forbids using mainstream EAs in some application domains, e.g. Optimal Design and Numerical Engineering where the objective functions are computationally demanding. Surrogate-based EAs alleviate this limitation by iteratively estimating the objective function, fueling the EA with the estimate (aka surrogate model), acquiring new examples of the objective function and revising the surrogate model accordingly; the reader will find a comprehensive review of surrogate evolutionary optimization in [6].

Surrogate models are equally useful in evolutionary Multi-Objective Optimization, all the more so when EMO involves several computationally heavy objectives [8]. Current surrogate-based EMO approaches, with the notable exception of [13] and [10], basically extend surrogate-based standard EAs, building

---

\*\* Work partially funded by FUI of System@tic Paris-Region ICT cluster through contract DGT 117 407 *Complex Systems Design Lab (CSDL)*.

one surrogate for each objective function and replacing the objective by its surrogate. The main limitation of such approaches is due to the approximation noise as the number of objectives increases. The estimation cost indeed increases linearly with the number of objectives; but the Pareto dominance test, checking whether one individual is dominated by another one, requires comparing their surrogate values over all objectives; the error thus exponentially increases in the worst case with the number of objectives.

The first mono-surrogate EMO approach was proposed by [13], aimed at characterizing the already visited region of the objective space, although this characterization hardly enables to guide evolution in the decision space. Addressing this limitation, another mono-surrogate approach defined in the decision space aimed at characterizing the current Pareto front and dominated region in the decision space in order to guide further evolution [10] (more in section 2). While this approach, referred to as Aggregated Surrogate Model (ASM), yields significant savings in terms of computational cost on benchmark problems, it relies on a complex adaptation of the Support Vector Machine framework [12], involving regression- and one-class-like constraints. The ASM limitations, related to the diversity of the Pareto front or the care to be exerted when using the surrogate model to filter the offspring, are blamed on the over-constrained formulation of the mono-surrogate model.

A new and relaxed version of Aggregated Surrogate Model for EMO is proposed in this paper, inspired from rank-based SVM [4, 7]. Basically, the new surrogate model referred to as RASM (Rank-based ASM) is only required to locally approximate the Pareto dominance relation, enabling to rank neighbor points within the objective space. RASM is still used to filter the offspring, through estimating whether they improve on their parents in terms of approximated Pareto-dominance.

This paper is organized as follows. Section 2 briefly discusses Aggregated Surrogate Models in EMO, detailing the approach proposed in [10] for the sake of self-containedness. Section 3 presents and discusses the RASM approach, which is experimentally validated in Section 4. Section 5 concludes and presents some perspectives for further research.

## 2 Aggregated Surrogate Models

Without pretending to exhaustivity, and referring the reader to [8] for a comprehensive review of surrogate-based EMO approaches, this section focuses on mono-surrogate EMO algorithms. As mentioned in the introduction, the first ever mono-surrogate EMO algorithm proposed by Yu et al. aimed at characterizing the region of the objective space visited so far [13]. The rationale for this approach, based on One-Class SVM [11], is that the envelope of the visited region excludes the Pareto front. In the general case however, the Pareto front in the objective space does not tell much about the Pareto set in decision space (except for specific problems where the Pareto front in the objective space cor-

responds to a set of rectangles in the decision space) and thus hardly allows one to guide the EMO search.

The only mono-surrogate EMO approach in the decision space, to our best knowledge, was referred to as Aggregate Surrogate Model [10]; it combines several variants of the Support Vector Machine framework.

In their initial formulation [12], Support Vector Machines aim at a linear model on the instance space  $X \subset \mathbb{R}^d$ , solution of a quadratic optimization problem:

$$\text{Argmin}_{\{w,\xi\}} \left( \frac{1}{2} \|w\|^2 + C \sum_k \xi_k \right)$$

where the norm of the sought linear solution  $w \in \mathbb{R}^d$  is minimized for the sake of good generalization guarantees, and each  $\xi_k$  stands for the violation of one learning constraint, to be minimized<sup>3</sup>

Typically, mapping a point  $x_i \in \mathbb{R}^d$  onto a desired value  $y_i$  up to some tolerance threshold  $\epsilon$  (regression problem) amounts to four constraints:

$$\begin{aligned} \langle w, x_i \rangle - y_i &\leq \epsilon + \xi_i^{up} & \xi_i^{up} &\geq 0 \\ \langle w, x_i \rangle - y_i &\geq -\epsilon - \xi_i^{low} & \xi_i^{low} &\geq 0 \end{aligned}$$

Likewise, mapping a point  $x_i$  onto a half space  $[a, \infty)$  (one-class problem) involves two constraints:

$$\langle w, x_i \rangle \geq a - \xi_i \quad \xi_i \geq 0$$

The ASM approach presented in [10] hybridizes the above two types of constraints as follows. Let the training set be defined as  $\mathcal{E} = \{x_1, \dots, x_\ell, x_{\ell+1}, \dots, x_m\}$  where the first  $\ell$  points belong to the current Pareto front and the following points  $x_{\ell+1}, \dots, x_m$  are dominated ones (possibly sub-sampling the current population to preserve the diversity in the objective or decision space). The ASM is obtained from the following learning constraints:

- All Pareto points  $x_1 \dots x_\ell$  are mapped on some value  $\rho$  up to tolerance  $\epsilon$  (regression constraints);
- All dominated points are mapped onto  $(-\infty, \rho + \epsilon[$  (one-class constraints).

The intuition behind this formulation is that the *true* Pareto front would then expectedly lie in the ‘half space,  $]\rho + \epsilon, +\infty)$ , thus enabling to guide the exploration of the search space.

The ASM problem finally reads:

$$\text{Argmin}_{\{w,\xi\}} \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} (\xi_i^{up} + \xi_i^{low}) + C \sum_{i=\ell+1}^m \xi_i^{up} \right)$$

<sup>3</sup> The extension of the SVM approach to non-linear search spaces relies on the so-called kernel trick, implicitly mapping the instance space  $X$  onto a feature space [12]. See section 3 and [12] for more details.

subject to

$$\begin{aligned}
\langle w, x_i \rangle &\leq \rho + \epsilon + \xi_i^{up} & \xi_i^{up} &\geq 0 & i = 1 \dots \ell \\
\langle w, x_i \rangle &\geq \rho - \epsilon - \xi_i^{low} & \xi_i^{low} &\geq 0 & i = 1 \dots \ell \\
\langle w, x_i \rangle &\leq \rho - \epsilon + \xi_i^{up} & \xi_i^{up} &\geq 0 & i = \ell + 1 \dots m
\end{aligned}$$

As mentioned in the introduction, the ASM problem is overconstrained as all Pareto points must be mapped on a narrow interval  $]\rho - \epsilon, \rho + \epsilon[$ . Another issue is that it should make no difference whether the dominated points are mapped onto  $(-\infty, \rho - \epsilon[$  or  $]\rho + \epsilon, +\infty)$ . Still, the experimental validation on problems ZDT1:3-6 [14] and their rotated variants IHR1:3-6 [5] shows that the most effective variant depends on the underlying benchmark problem. Whereas one can proceed by trying both variants and retaining the most effective one, the approach is clearly unsatisfactory. Some attempts at a symmetrical formulation of the ASM problem failed to address this issue.

### 3 Rank-based Aggregate Surrogate Model

This section gives an overview of the Rank-based Aggregate Surrogate Model (RASM), meant to address the ASM limitations. After stating the RASM formulation and sketching its resolution, it details its use within the EMO framework.

#### 3.1 A surrogate modelling Pareto dominance

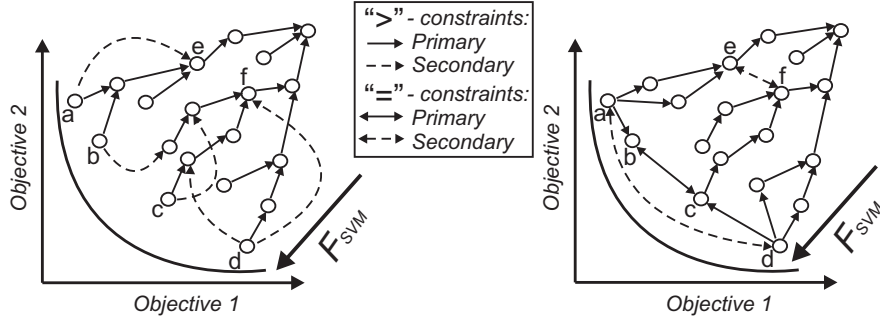
A new learning setting aimed at preference learning, a.k.a. learning to rank, has been addressed within the SVM framework [7]. While preference learning can be cast as a classification problem on  $X \times X$  (the class of  $(x, x')$  is positive iff  $x$  is to be preferred to  $x'$ ), it offers better generalization guarantees to formalize preference learning as an underconstrained regression problem, where the hypothesis  $h$  mapping  $X$  onto the real-valued space  $\mathbb{R}$  is only required to satisfy  $h(x) > h(x')$  whenever  $x$  is preferred to  $x'$ .

Let  $\mathcal{E} = \{x_1, \dots, x_m\}$  and let  $\mathcal{P}$  denote the set of pairs  $(i, j)$  such that  $x_i$  is preferred to  $x_j$ ; the original formulation of rank-based SVM, involving all preference constraints, is as follows:

$$\text{Argmin}_{\{w, \xi\}} \left( \frac{1}{2} \|w\|^2 + C \sum_{(i,j) \in \mathcal{P}} \xi_{i,j} \right) \quad (1)$$

$$\text{subject to } \left. \begin{aligned} \langle w, x_i \rangle - \langle w, x_j \rangle &\geq 1 - \xi_{i,j} \\ \xi_{i,j} &\geq 0 \end{aligned} \right\} \forall (i, j) \in \mathcal{P} \quad (2)$$

where  $\xi_{i,j}$  stands for the slack variable associated to the violation of the preference constraint associated to  $(x_i, x_j)$  along the same lines as in Section 2. However, for the sake of tractability, the resolution of Eq. (1-2) proceeds iteratively, considering a set  $\Omega_{active}$  of active constraints which is initially empty. Eq. (1)-(2) then only imply  $\Omega_{active}$  instead of  $\mathcal{P}$ . At each iteration, the most violated constraint in  $\mathcal{P}$  is added to  $\Omega_{active}$ , and optimization proceeds [7].



**Fig. 1.** Constraints involved in Rank-based Aggregated Surrogate Models. Left: The current RASM. Right: Further extensions; see section 5.

Rank-SVM is adapted to the EMO framework as follows. Let  $\mathcal{E}$  define the current training set (population or archive), and consider the preference order defined by non-dominated sorting. Several requirements on the rank-based surrogate model are defined. First of all, the number of constraints should be linear or sub-linear in the population size for the sake of tractability. Secondly, the constraints should enforce an accurate model in terms of generality w.r.t. Pareto dominance. Lastly, the model should support the diversity of the population along the Pareto front. To comply with these requirements, only dominance constraints have been considered so far (Fig. 1.left)<sup>4</sup>. Specifically, let **primary dominance constraints** be associated to pairs  $(x_i, x_j)$  such that  $x_j$  is the nearest neighbor of  $x_i$  conditionally to the fact that  $x_i$  dominates  $x_j$  (continuous arrow, Fig. 1.left), and let **secondary dominance constraints** be associated to pairs  $(x_i, x_j)$  such that  $x_i$  belongs to the current Pareto front and  $x_j$  belongs to another front from non-dominated sorting (dotted arrow, Fig. 1.left).

Formally, the RASM is built by solving the dual problem associated to Eqs.(1-2); due to space limitations the reader is referred to [10] for more details. Let  $K_{a,b}$  denote the scalar product  $\langle \Phi(x_a), \Phi(x_b) \rangle$  where  $\Phi$  denotes the mapping from the initial to the feature space (kernel trick), and let  $\alpha_{a,b}$  denote the Lagrange multiplier associated to the constraint relating  $x_a$  and  $x_b$ . Then it comes:

**Dual form:**

$$\begin{aligned} & \text{Argmax}_{\{\alpha\}} \sum_{i,j=1}^m \alpha_{ij} - \frac{1}{2} \sum_{i,j,u,v=1}^m \alpha_{ij} \alpha_{uv} (K_{iu} - K_{iv} - K_{ju} + K_{jv}) \\ & \text{subject to } 0 \leq \alpha_{ij} \leq C \end{aligned} \quad (3)$$

The Lagrangian (Eq. 3) is maximized iteratively by optimizing a single  $\alpha_{ij}$  multiplier (uniformly selected in  $\Omega_{active}$ ) in each iteration. Taking inspiration from [7], RASM maintains a set  $\Omega_{active}$  of active constraints which is initialized to the set of primary dominance constraints. After the first  $1000 \times |\Omega_{active}|$

<sup>4</sup> Other possibilities, illustrated in Fig. 1.right, will be discussed in section 5.

iterations, every  $10 \times |\Omega_{active}|$  iterations the most violated secondary constraint is thereafter added to  $\Omega_{active}$ , until a given number of secondary constraints have been added (typically 10% the number of primary constraints in the presented experiments, see Section 4).

### 3.2 RASM-based EMO

The RASM is integrated within existing EMO algorithms along the same lines as the ASM [10]. Two different EMO algorithms have been considered, respectively NSGA-II and MO-CMA-ES. At each generation, the training set is built from the current population and an archive of points that have been visited by the algorithm (more in Section 4.1). The standard variation operators of the underlying EMO are used to generate many offspring; these offspring are thereafter filtered according to the RASM as follows. Formally, the quality of an offspring  $z$  is estimated as  $RASM(z) - RASM(x_z)$  where  $x_z$  is the individual from population nearest to  $z$  in decision space.

As could have been expected, greedily selecting the offspring with maximal quality leads to premature convergence due to the approximation noise. The offspring are therefore ordered according to their quality estimate; a probabilistic selection attached to the offspring indices is achieved, using a normal distribution with variance  $\sigma_{sel}^2$  (Fig. 2.right). Granted that the number of offspring is sufficiently large, parameter  $\sigma_{sel}^2$  thus controls the selection pressure and the exploration vs exploitation trade-off. However we use at the moment a small value  $\sigma_{sel}^2 = 0.001$  for the normal distribution for the ranked point to be chosen (Fig. 2.right) to simulate the situation when new offspring always has the first rank, and allow comparison with results from [10].

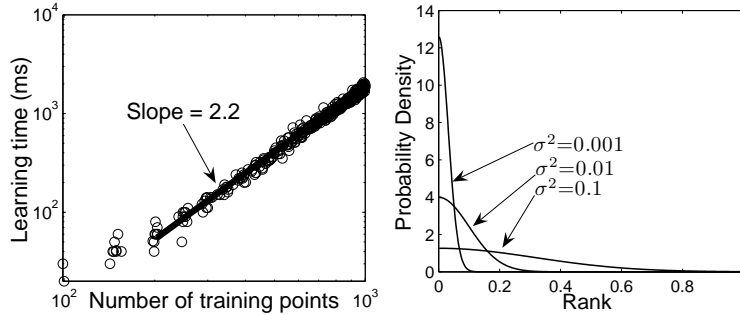
## 4 Experimental Validation

### 4.1 Experimental Setting

For the sake of a fair comparison between the presented RASM and the ASM first presented in [10], the same experimental validation procedure is used: two state-of-the-art EMO algorithms are considered as baselines, the (100+100)-S-NSGA-II [3], NSGA-II with hypervolume indicator as second-sorting criterion, and the  $100 \times (1+1)$ -MO-CMA-ES, the multi-objective version of CMA-ES [5]. Both ASM and RASM are integrated within these algorithms.

Both ASM and RASM are based on the Radial Basis Functions kernel:  $K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2}$ , where the bandwidth  $\sigma$  is set as the average distance between all pairs of training points. The SVM penalization constant  $C$  is set to 1000.

The training set  $\mathcal{E}$  that is used at each generation to build the RASM model is an archive that contains at most  $N_{archive} = 1000$  points. The current population is added to the archive at each generation. When this archive gets larger than  $N_{archive}$ , it is pruned by removing the worst individuals after non-dominated



**Fig. 2.** Left: Learning time of the proposed dominance-based RASM on ZDT1 function. Right: Mapping the ranks of pre-children to a normal distribution.

sorting. Furthermore, in order to improve the diversity of the training set (many points too close together can lead to poor surrogate model), an additional filtering procedure is applied to the archive. The 2-objective space has been divided into  $100 \times 100$  boxes, and at most one point among the archived non-dominated points of each box is retained in the archive.

As detailed in section 3, RASM maintains the set  $\Omega_{active}$  of active constraints, initialized to the set of primary dominance constraints. After an initial round of  $1000 |\Omega_{active}|$  iterations,  $\Omega_{active}$  is incrementally enriched every  $10 |\Omega_{active}|$  iterations with the most violated constraint among the secondary dominance constraints, until a total of  $0.1 |\Omega_{active}|$  secondary constraints have been added.

**Performance Measures** Many ways of measuring the performance of EMO algorithms have been proposed. After [9], this study uses Pareto-compliant quality indicators, more particularly the widely used hypervolume indicator  $I_H$ . Let  $P$  be a  $\mu$ -size approximation of Pareto front and let  $P^*$  be the approximate  $\mu$ -optimal distribution of optimal Pareto points [1]. The approximation error of the Pareto front is defined by  $\Delta H(P^*, P) = I_H(P^*) - I_H(P)$ . All reported results are averaged over 10 independent runs with at most 100,000 fitness evaluations.

## 4.2 Result Analysis

Some experiments are first conducted to estimate the complexity of the surrogate training on 30-dimensional ZDT1 problem. The empirical complexity with respect to the number of training points is circa 2.2 (slope on Fig. 2-Left in log scale). The fact that the complexity is super-quadratic is not surprising since the SVM procedure relies on the Gram matrix  $K(x_i, x_j)$  for all points  $x_i$  and  $x_j$  in the training set. The complexity however remains bounded as the size of the training set (extracted from the archive) is less than 1,000, limiting *de facto* the computational cost of the RASM learning.



Table 1 shows the comparative results of all baseline, ASM and RASM-based EMOs; in the latter cases, both  $p = 2$  and  $p = 10$  pre-offspring are considered. These results first confirm that *S*-NSGA-II performs best on separable functions ZDTx and *MO*-CMA-ES on non-separable functions IHRx<sup>5</sup>. They also show that both *RASM*-NSGA and *RASM*-*MO*-CMA work nearly 1.5 times faster with  $p = 2$  and more than 2 times faster with  $p = 10$  than the baseline versions with regards to the  $\Delta H$  value and the number of function evaluations.

*ASM*-NSGA and *RASM*-NSGA yield comparable performances. A more thorough analysis shows that *RASM*-*MO*-CMA is usually faster at the beginning (up to 10000-15000 function evaluations) though it might suffer from a premature convergence thereafter: experiments on concave IHR2 (and to some extent, also on ZDT2) show that *RASM*-*MO*-CMA converges to the value  $\Delta H = 0.1$  nearly 1.5 times faster than with *ASM* model, and fails to go further. This failure is blamed on the fact that the diversity of the population is hardly preserved; a small part of the optimal Pareto front is sampled. Indeed, *RASM* learning and the *RASM*-based offspring selection only aim at speeding up the convergence; further work will be required to extend the approach and approximate the  $\mu$ -optimal distribution of nearly-optimal Pareto points.

## 5 Discussion and Perspectives

The main contribution of the present paper is to show how to train a single surrogate model to reflect Pareto dominance in an EMO framework, using a *Learning to Rank* framework. *RASM*, the resulting surrogate model, does not require that all Pareto points are mapped onto the same value. It is thus both more constrained in the dominated region, and less constrained on the Pareto front, than *ASM*, previous work of the authors along similar lines [10].

Furthermore, this approach opens new and interesting perspectives for real world multi-objective problems, enabling for instance to account for the user’s preferences in a flexible way by simply adding user-defined constraints to the order-based SVM formulation. Most importantly, the rank constraint formalization enables to accommodate conflicting preferences: to the best of our knowledge, this corresponds to a significant advance on the state of the art. This property is likely to be important for simulated evolution, since the ability to predict an environment is a prerequisite for intelligent behavior.

The experimental validation of the proposed approach shows that *RASM*-EMO usually converges faster than *ASM*-EMO, with the caveat that it sometimes leads to premature convergence (e.g., on ZDT2 and IHR2 problems). This premature convergence was blamed on the selection pressure and the adjustment of parameter  $\sigma_{sel}^2$ . A further work will explore the adaptation of the famed 1/5-th rule to adjust  $\sigma_{sel}^2$ , using the hypervolume indicator  $\Delta H$  as measure of success.

The number of constraints that are added to the primary constraints could also be made adaptive by considering the stability of the surrogate model. How-

<sup>5</sup> *MO*-CMA-ES penalization parameter  $\alpha$  is 1.0 for all problems in order to prevent evolution from being biased toward exploring the boundaries of the decision space.

ever, such potential improvement would require the computation of all  $\xi_{ij}$  and  $F_{svm}$  values for all points of the archive at each generation, and would hence be computationally costly.

Another shortcoming of Aggregate Surrogate Models is how to resist the loss of diversity. It is emphasized that RASM might incorporate additional specific constraints in each generation. Some possible constraints are described in Figure 1-Right: such **non-dominance** constraints involved points on the current Pareto front, and include inequality constraints from the extremal points over their neighbors (continuous arrows), and equality constraints for all neighbor pairs on the Pareto front (continuous double arrow), as well as between extremal points (dotted double arrows). Such equality constraints can be rewritten as two symmetrical inequality constraints in order to preserve the particular form of the formulation (Eq. 2). Along the same lines, constraints could be weighted, e.g. the weight of constraints related to points with the largest hypervolume contributions can be increased online. This is however a topic for further work.

## References

1. A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the hypervolume indicator: Optimal  $\mu$ -distributions and the choice of the reference point. In *FOGA*, pages 87–102. ACM, 2009.
2. K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley, 2001.
3. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A Fast Elitist Multi-Objective Genetic Algorithm: NSGA-II. *IEEE TEC*, 6:182–197, 2000.
4. R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. In A. Smola, P. Bartlett, and B. Schölkopf, editors, *Advances in Large Margin Classifiers*, pages 115–132, Cambridge, MA, 2000. MIT Press.
5. C. Igel, N. Hansen, and S. Roth. Covariance Matrix Adaptation for Multi-objective Optimization. *Evolutionary Computation*, 15(1):1–28, 2007.
6. Y. Jin. A Comprehensive Survey of Fitness Approximation in Evolutionary Computation. *Soft Computing*, 9(1):3–12, 2005.
7. T. Joachims. A support vector method for multivariate performance measures. In L. De Raedt and S. Wrobel, editor, *Proc. ICML*, pages 377–384. ACM, 2005.
8. J. Knowles and H. Nakayama. Meta-modeling in multiobjective optimization. In J. Branke et al., editor, *Multiobjective Optimization*, number 5252 in LNCS, pages 245–284. Springer Verlag, 2008.
9. J. Knowles, L. Thiele, and E. Zitzler. A tutorial on the performance assessment of stochastic multiobjective optimizers. Technical report, 2006.
10. I. Loshchilov, M. Schoenauer, and M. Sebag. A Mono Surrogate for Multiobjective optimization. In J. Branke et al., editor, *GECCO'2010*, pages 471–478. ACM, 2010.
11. B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, 13:1443–1471, 2001.
12. V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
13. Y. Yun, H. Nakayama, and M. Arakava. Generation of pareto frontiers using support vector machine. In *MCDM'04*, 2004.
14. E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8:173–195, 2000.

**Table 1.** Comparative results of two baseline EMOAs, namely S-NSGA-II and MO-CMA-ES and their ASM and RASM variants. Median number of function evaluations (out of 10 independent runs) to reach  $\Delta H_{\text{target}}$  values, normalized by Best: a value of 1 indicates the best result, a value  $X > 1$  indicates that the corresponding algorithm needed  $X$  times more evaluations than the best to reach the same precision.

$\Delta H_{\text{target}}$	1	0.1	0.01	1e-3	1e-4	1	0.1	0.01	1e-3	1e-4
	ZDT1					ZDT2				
Best	1100	3000	5300	7800	38800	1400	4200	6600	8500	32700
S-NSGA-II	1.6	2	2	2.3	1.1	1.8	1.7	1.8	2.3	1.2
ASM-NSGA p=2	1.2	1.5	1.4	1.5	1.5	1.2	1.2	1.2	1.4	<b>1</b>
ASM-NSGA p=10	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	.	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	.
RASM-NSGA p=2	1.2	1.4	1.4	1.6	<b>1</b>	1.3	1.2	1.2	1.5	<b>1</b>
RASM-NSGA p=10	<b>1</b>	1.1	1.1	1.5	.	1.1	<b>1</b>	<b>1</b>	1.2	.
MO-CMA-ES	16.5	14.4	12.3	11.3	.	14.7	10.7	10	10.1	.
ASM-MO-CMA p=2	6.8	8.5	8.3	8	.	5.9	8.2	7.7	7.5	.
ASM-MO-CMA p=10	6.9	10.1	10.4	12.1	.	5	.	.	.	.
RASM-MO-CMA p=2	5.1	7.7	7.6	7.4	.	5.2	.	.	.	.
RASM-MO-CMA p=10	3.6	4.3	4.9	7.2	.	3.2	.	.	.	.
	ZDT3					ZDT6				
Best	1300	3500	7100	10100	15200	2500	3600	5200	12300	.
S-NSGA-II	1.4	1.9	1.6	1.9	2.2	2.1	3.4	3.8	2.7	.
ASM-NSGA p=2	1.1	1.3	1.1	1.2	1.3	1.4	2.4	2.6	2	.
ASM-NSGA p=10	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	1.1	1.8	2.3	2.3	.
RASM-NSGA p=2	1.1	1.3	1.2	1.4	1.6	1.5	2.4	2.8	2.1	.
RASM-NSGA p=10	<b>1</b>	1.1	1.1	2	.	1.4	2	2.3	1.8	.
MO-CMA-ES	15.4	17.8	.	.	.	2.5	2.6	2.5	2	.
ASM-MO-CMA p=2	9	.	.	.	.	1.1	1.2	1.1	<b>1</b>	.
ASM-MO-CMA p=10	8	25.6	.	.	.	<b>1</b>	1.1	1.3	2.5	.
RASM-MO-CMA p=2	8.5	.	.	.	.	1.5	1.2	1.2	<b>1</b>	.
RASM-MO-CMA p=10	8.1	.	.	.	.	<b>1</b>	<b>1</b>	<b>1</b>	1.6	.
	IHR1					IHR2				
Best	500	2000	35300	41200	50300	1700	7000	12900	52900	.
S-NSGA-II	1.6	1.5	.	.	.	1.1	3.2	6.2	.	.
ASM-NSGA p=2	1.2	1.3	.	.	.	<b>1</b>	3.9	4.9	.	.
ASM-NSGA p=10	<b>1</b>	1.5	.	.	.	1.4	6.4	4.6	.	.
RASM-NSGA p=2	1.2	1.2	.	.	.	1.5	.	.	.	.
RASM-NSGA p=10	<b>1</b>	<b>1</b>	.	.	.	1.2	5.1	4.8	.	.
MO-CMA-ES	8.2	6.5	1.1	1.2	1.2	5.8	2.7	2.1	<b>1</b>	.
ASM-MO-CMA p=2	4.6	2.9	<b>1</b>	<b>1</b>	<b>1</b>	3.1	1.6	1.4	1.1	.
ASM-MO-CMA p=10	9.2	6.1	1.3	1.2	.	5.9	2.6	2.4	.	.
RASM-MO-CMA p=2	2.6	2.3	2.4	2.1	.	2.2	<b>1</b>	<b>1</b>	.	.
RASM-MO-CMA p=10	1.8	1.9	.	.	.	.	.	.	.	.
	IHR3					IHR6				
Best	800	.	.	.	.	16500	.	.	.	.
S-NSGA-II	1.5	.	.	.	.	5.4	.	.	.	.
ASM-NSGA p=2	1.1	.	.	.	.	3.8	.	.	.	.
ASM-NSGA p=10	<b>1</b>	.	.	.	.	.	.	.	.	.
RASM-NSGA p=2	1.3	.	.	.	.	2.2	.	.	.	.
RASM-NSGA p=10	1.1	.	.	.	.	2.6	.	.	.	.
MO-CMA-ES	9.6	.	.	.	.	2	.	.	.	.
ASM-MO-CMA p=2	7.2	.	.	.	.	2	.	.	.	.
ASM-MO-CMA p=10	12.1	.	.	.	.	.	.	.	.	.
RASM-MO-CMA p=2	3.3	.	.	.	.	<b>1</b>	.	.	.	.
RASM-MO-CMA p=10	2.6	.	.	.	.	<b>1</b>	.	.	.	.