The Logarithmic Hypervolume Indicator

Tobias Friedrich
Max-Planck-Institut für Informatik
66123 Saarbrücken, Germany

Karl Bringmann
Universität des Saarlandes
66041 Saarbrücken, Germany

Thomas Voß
Institut für Neuroinformatik
Ruhr-Universität Bochum
44780 Bochum, Germany

Christian Igel
Department of Computer Science
University of Copenhagen
2100 Copenhagen Ø, Denmark

ABSTRACT

It was recently proven that sets of points maximizing the hypervolume indicator do *not* give a good multiplicative approximation of the Pareto front. We introduce a new "logarithmic hypervolume indicator" and prove that it achieves a close-to-optimal multiplicative approximation ratio. This is experimentally verified on several benchmark functions by comparing the approximation quality of the multi-objective covariance matrix evolution strategy (MO-CMA-ES) with the classic hypervolume indicator and the MO-CMA-ES with the logarithmic hypervolume indicator.

Categories and Subject Descriptors

F.2 [**Theory of Computation**]:
Analysis of Algorithms and Problem Complexity

General Terms

Measurement, Hypervolume Indicator

Keywords

Multiobjective Optimization, Theory,
Performance Measures, Selection

1. INTRODUCTION

Most real-world optimization problems have to deal with multiple objectives (like time vs. cost) and cannot be easily described by some scalar objective function. The quality of solutions to such multi-criteria optimization problems are measured by vector-valued objective functions. This implies that there is in general no unique optimal value, but a possibly very large set of incomparable optimal values, which form the Pareto front. The corresponding solutions constitute the Pareto set and have the defining property that they cannot be improved in some objective without getting worse in another one.

FOGA'11, January 5–9, 2011, A-6867 Schwarzenberg, Austria.

Many different multi-objective evolutionary algorithms (MOEAs) have been developed to find a Pareto set of (small) size n which gives a *good approximation* of the Pareto front. A popular way to measure the quality of a Pareto set is the hypervolume indicator (HYP), which measures the volume of the dominated space [23]. For small numbers of objectives, MOEAs directly using the hypervolume indicator to guide the search have become the methods of choice. These include for example the generational MO-CMA-ES [10, 20], the SMS-EMOA [3, 8], and variants of IBEA [22, 25]. Despite its popularity, up to recently there was not much rigorously known about the distribution of solution sets that maximize the hypervolume. Such solution sets have been described empirically as "well distributed" in [8, 13, 14]. In contrast to this it was observed that "convex regions may be preferred to concave regions" [17, 23] as well as that HYP is "biased towards the boundary solutions" [7]. For the number of points $n \to \infty$, it is also known that the density of points only depends on the gradient of the function describing the front [2].

At first sight, it is not obvious why maximizing the hypervolume indicator should yield a good approximation of the Pareto front. If we are, for example, interested in a good multiplicative approximation, an "ideal" indicator would directly measure the approximation quality of a solution set P by returning the smallest $\alpha \in \mathbb{R}^+$ such that each element of the Pareto front is dominated by some vector resulting from dividing an element in P by α (we consider minimization and assume positive objective function values). This corresponds to the binary multiplicative ε-indicator [16, 24] applied to the solution set and the (possibly infinite) Pareto front. Unfortunately, such an indicator cannot be used in practice because the Pareto front is usually unknown. This leads to the important question of how close the approximations achieved by realistic indicators such as the hypervolume indicator come to those that could be obtained by an "ideal" indicator. This can be measured by the approximation ratio of a solution set maximizing the hypervolume [4, 5, 9]. Formal definitions of the multiplicative approximation ratio and the alternative additive approximation ratio are given in Definitions 3.1 and 3.4, respectively.

The approximation quality achieved by the hypervolume indicator has been analyzed rigorously for bi-criterion maximization problems by Bringmann and Friedrich [4, 5]. They prove in [5] that for all possible Pareto fronts the multiplicative approximation factor achieved by a set of n solutions maximizing the hypervolume indicator is $1 + \Theta(1/n)$

(cf. Theorem 4.1)[1]. This is *asymptotically equivalent* to the optimal multiplicative approximation factor (cf. Corollary 3.2) [5]. Thus, one can conclude that the hypervolume indicator is guiding the search in the correct direction for sufficiently large n. However, the constant factors hidden by the Θ's are larger for the set maximizing hypervolume compared to the set with best possible approximation factor. Bringmann and Friedrich [4] studied the multiplicative approximation factor relative to the ratio A/a between the largest and smallest coordinate[2]. Using this notation, the precise result of [5] is the computation of the optimal multiplicative approximation ratio as $1 + \log(A/a)/n$ (cf. Corollary 3.2). In [4] it is further shown that the multiplicative approximation ratio for a set maximizing the hypervolume is strictly larger, namely of the order of at least $1+\sqrt{A/a}/n$ (cf. Theorem 4.2). This implies that the multiplicative approximation ratio achieved by a set maximizing the hypervolume can be *exponentially worse* in the order of the ratio A/a. Hence for numerically very wide spread fronts there are Pareto sets which give a much better multiplicative approximation than Pareto sets which maximize the hypervolume.

These results about the multiplicative approximation ratio were surprisingly bad news for the hypervolume indicator. On the other hand, Bringmann and Friedrich [4] also examined the additive approximation ratio and observed that while the multiplicative approximation factor is determined by the ratio A/a, the additive approximation factor is determined by the width of the domain $A - a$. They proved that the optimal additive approximation ratio is $(A - a)/n$ (cf. Theorem 3.3) and upper bound the additive approximation ratio achieved by a set maximizing the hypervolume by $(A-a)/(n-2)$ (cf. Theorem 4.3). This is a very strong statement, because, apart from a small factor of $n/(n - 2)$, the additive approximation ratio achieved when maximizing the hypervolume is optimal. This shows that the hypervolume indicator yields a much better additive than multiplicative approximation for maximization problems.

Our Results

An obvious next question is to verify whether the same results hold for minimization problems. The first part of this paper accordingly confirms this and presents the results of Bringmann and Friedrich [4, 5] in a minimization setting. In the main part, we address the open question whether there are natural indicators which provably achieve a good multiplicative approximation ratio. Since it is known that the hypervolume gives a very good additive approximation, we hypothesize that an indicator achieving a good multiplicative approximation can be constructed by taking the logarithm of all axes before computing the classic hypervolume. We call this new indicator the *logarithmic hypervolume indicator* and analyze its properties in this study. Note that in the setting of weighted hypervolume indicators [25] this corresponds to a reciprocal weight function (cf. Section 4.3). We prove that the logarithmic hypervolume indicator achieves a multiplicative approximation ratio of less than $1 + \log(A/a)/(n-2)$ (cf. Corollary 4.6), which is again optimal apart from the factor $n/(n - 2)$.

These theoretical results indicate that one should get a much better multiplicative approximation of the Pareto-front if one uses the logarithmic instead of the classic hypervolume indicator as a subroutine of an indicator-based evolutionary algorithm. However, the results do not directly apply to solutions returned by such an algorithm: First, these algorithms might fail to return a solution maximizing the (logarithmic) hypervolume indicator, because they did not run long enough or got stuck in a local optimum. Second, in the theoretical part we measure the approximation quality in the worst-case over all possible fronts, which gives only upper, but no lower bounds for "typical" fronts. And third, the factor $n/(n - 2)$ the logarithmic hypervolume is worse in the worst-case, goes to 1 for large n, but the number of non-dominated solutions n can be very small in a solution returned by an evolutionary algorithm. To examine whether the logarithmic hypervolume indicator yields indeed a better multiplicative approximation than the classic hypervolume indicator for a typical indicator-based evolutionary algorithm, we compare the $(\mu+1)$-MO-CMA-ES with the classic hypervolume indicator and the $(\mu+1)$-MO-CMA-ES with the logarithmic hypervolume indicator. This study is performed on the DTLZ benchmark functions [6]. We observe that the results for the theoretical worst-case bounds match well with the empirically measured approximation ratios for these benchmark functions. On all benchmark functions, the approximation achieved by the logarithmic hypervolume indicator compared to the classic hypervolume indicator is better by up to 31% (cf. Table 1). This implies that for multiplicative problems the logarithmic hypervolume indicator should be preferred over the classic hypervolume indicator.

The remainder of this paper is structured as follows. In Section 2 we define the used notation. The following two Sections 3.1 and 3.2 define the concepts of multiplicative and additive approximation ratios. Section 4 introduces the weighted, normal and logarithmic hypervolume indicator. The following Section 4.4 proves the bounds on the multiplicative approximation ratio of the logarithmic hypervolume indicator. Finally, Section 5 presents our experimental framework and results.

2. PRELIMINARIES

All three previous papers on the approximation ratio of the hypervolume indicator [4, 5, 9] only consider maximization problems. As benchmark functions such as DTLZ [6] usually consider minimization problems, the first part of this paper (till Section 3.2) deals with translating the results of Bringmann and Friedrich [4, 5] for the approximation ratios of maximization problems to minimization problems. This is straight-forward for the definitions, but not obvious for all of the theorems. All results still hold analogously if the handling of the boundary solutions is adapted. As these changes are easy to verify, all proofs of theorems which are translated from the maximization setting of [4, 5] are omitted.

As in [4, 5], we consider the case of two objectives where there is a mapping from an arbitrary search space to an objective space which is a subset of $\mathbb{R}^2$. Throughout this paper, we will only work on the objective space. For points from the objective space we define the following dominance

[1]The precise statements of this and the following results is slightly more technical. For details see the respective theorems.

[2]The approximation ratio depends on the ratios in both dimensions. To simplify the presentation in the introduction, we assume that the ratio A/a in the first dimension is equal to the ratio B/b in the second dimension.

relation:

$$(x_1, y_1) \preceq (x_2, y_2) \text{ iff } x_1 \leqslant x_2 \text{ and } y_1 \leqslant y_2,$$
$$(x_1, y_1) \prec (x_2, y_2) \text{ iff } (x_1, y_1) \preceq (x_2, y_2) \text{ and } (x_1, y_1) \neq (x_2, y_2).$$

We restrict ourselves to Pareto fronts that can be written as $\{(x, f(x)) \mid x \in [a, A]\}$ where $f\colon [a, A] \to [b, B]$ is a (not necessarily strictly) monotonically decreasing, lower semi-continuous[3] function with $f(a) = B$, $f(A) = b$ for some $a < A$, $b < B$ with $a, A, b, B \in \mathbb{R}$. We write $\mathcal{F} = \mathcal{F}_{[a,A]\to[b,B]}$ for the set of all such functions f. We will use the term *front* for both, the set of points $\{(x, f(x)) \mid x \in [a, A]\}$, and the function f.

The condition of f being lower semi-continuous cannot be relaxed further as without it the front lacks a certain symmetry in the two objectives: This condition is necessary and sufficient for the existence of the inverse function $f^{-1}\colon [b, B] \to [a, A]$ defined by setting $f^{-1}(y) := \min\{x \in [a, A] \mid f(x) \leqslant y\}$. Without lower semi-continuity this minimum does not necessarily exist. Furthermore, this condition implies that there is a set maximizing the hypervolume indicator.

Note that the set $\mathcal{F}$ of fronts we consider is a very general one. Many papers that theoretically examine the hypervolume indicator assume that the front is continuous and differentiable (e.g. [1, 2, 9]), and are thus not able to give results about discrete fronts.

Let $n \in \mathbb{N}$. For fixed $[a, A], [b, B] \subset \mathbb{R}$ we call a set $P = \{p_1, \ldots, p_n\} \subset [a, A] \times [b, B]$ a *solution set* (of size n) and write $\mathcal{P} = \mathcal{P}_n$ for the set of all such solution sets. A solution set P is said to be *feasible* for a front $f \in \mathcal{F}$, if $y \geqslant f(x)$ for all $p = (x, y) \in P$. We write $\mathcal{P}^f \subseteq \mathcal{P}$ for the set of all solution sets that are feasible for f.

A common approach to measure the quality of a solution set is to use unary indicator functions [26]. They assign to each solution set a real number that somehow reflects its quality, i.e., we have a function $\mathrm{ind}\colon \bigcup_{n=1}^{\infty} \mathcal{P}_n \to \mathbb{R}$. As throughout the paper $n \in \mathbb{N}$ is fixed, it is sufficient to define an indicator $\mathrm{ind}\colon \mathcal{P}_n \to \mathbb{R}$. Note that as we are only working on the objective space, we here slightly deviate from the usual definition of an indicator function where the domain is the search space, not the objective space.

A final remark regarding our notation: We will mark every variable with a $+$ or $*$ depending on whether it belongs to the additive or multiplicative approximation.

3. APPROXIMATING THE PARETO FRONT

When attempting to minimize an indicator function, we actually try to find a solution set $P \in \mathcal{P}$ that constitutes a good approximation of the front f. In the following, we introduce notions of multiplicative and additive approximtion quality.

3.1 Multiplicative Approximation

According to the custom of approximation algorithms, we measure the quality of a solution by its multiplicative approximation ratio. This can be transferred to the world of multi-objective optimization. For this we use the following definition of Papadimitriou and Yannakakis [18], which was also used in [4, 5, 9, 15, 16]. Note that it is crucial to require $a, b > 0$ here, as it is unclear what multiplicatively approximating a negative number should mean. We will always assume this when talking about multiplicative approximation throughout the paper.

[3]Semi-continuity is a weaker property than normal continuity. A function f is said to be lower semi-continuous if for all points x of its domain, $\liminf_{y\to x} f(y) \geqslant f(x)$. Intuitively speaking this means that for all points x the function values for arguments near x are either close to $f(x)$ or greater than $f(x)$. For more details see e.g. Rudin [19].

DEFINITION 3.1. *Let $f \in \mathcal{F}$ and $P \in \mathcal{P}^f$. The solution set P is a* multiplicative α-approximation *of f if for each $\hat{x} \in [a, A]$ there is a $p = (x, y) \in P$ with*

$$\hat{x} \geqslant x/\alpha \quad \text{and} \quad f(\hat{x}) \geqslant y/\alpha$$

where $\alpha \in \mathbb{R}$, $\alpha \geqslant 1$. The multiplicative approximation ratio *of P with respect to f is defined as*

$$\alpha^*(f, P) := \inf\{\alpha \in \mathbb{R} \mid P \text{ is a mult. } \alpha\text{-approximation of } f\}.$$

The quality of an algorithm which calculates a solution set of size n for each Pareto front in $\mathcal{F}$ has to be compared with the respective optimal approximation ratio defined as follows.

DEFINITION 3.2. *For fixed $[a, A]$, $[b, B]$, and n, let*

$$\alpha^*_{OPT} := \sup_{f\in\mathcal{F}} \inf_{P\in\mathcal{P}^f} \alpha^*(f, P).$$

The value α^*_{OPT} is chosen such that every front in $\mathcal{F}$ can be approximated by n points to a ratio of α^*_{OPT}, and there is a front which cannot be approximated better. In [5] the following results was shown.

THEOREM 3.1. $\alpha^*_{OPT} = \min\{A/a, B/b\}^{1/n}$.

As shown in [5], this implies the following corollary.

COROLLARY 3.2. *For all $n \geqslant \log(\min\{A/a, B/b\})/\varepsilon$ and $\varepsilon \in (0, 1)$,*

$$\alpha^*_{OPT} \geqslant 1 + \frac{\log(\min\{A/a, B/b\})}{n},$$
$$\alpha^*_{OPT} \leqslant 1 + (1 + \varepsilon)\,\frac{\log(\min\{A/a, B/b\})}{n}.$$

We further want to measure the approximation of the solution set of size n maximizing an indicator ind. As there might be several solution sets maximizing ind, we consider the worst case and use the following definition.

DEFINITION 3.3. *For a unary indicator* ind *and fixed $[a, A]$, $[b, B], n$, and $f \in \mathcal{F}$ let*

$$\mathcal{P}^f_{\mathrm{ind}} := \left\{P \in \mathcal{P}^f \;\middle|\; \mathrm{ind}(P) = \max_{Q\in\mathcal{P}^f} \mathrm{ind}(Q)\right\} \text{ and}$$
$$\alpha^*_{\mathrm{ind}} := \sup_{f\in\mathcal{F}} \sup_{P\in\mathcal{P}^f_{\mathrm{ind}}} \alpha^*(f, P).$$

The set $\mathcal{P}^f_{\mathrm{ind}}$ is the set of all feasible solution sets that maximize ind on f. The value α^*_{ind} is chosen such that for every front f in $\mathcal{F}$ every solution set maximizing ind approximates f by a ratio of at most α^*_{ind}.

3.2 Additive Approximation

Depending on the problem at hand, one can also consider an additive approximation ratio. Analogous to Definition 3.1 we use the following definition.

Definition 3.4. *Let $f \in \mathcal{F}$ and $P \in \mathcal{P}^f$. The solution set P is an* additive α-approximation *of f if for each $\hat{x} \in [a, A]$ there is a $p = (x, y) \in P$ with*

$$\hat{x} \geqslant x - \alpha \quad \text{and} \quad f(\hat{x}) \geqslant y - \alpha$$

where $\alpha \in \mathbb{R}$, $\alpha \geqslant 0$. The additive approximation ratio *of P with respect to f is defined as*

$$\alpha^{+}(f, P) := \inf\{\alpha \in \mathbb{R} \mid P \text{ is an add. } \alpha\text{-approximation of } f\}.$$

Again, we are interested in the optimal approximation ratio for Pareto fronts in $\mathcal{F}$. Analogous to Definition 3.2 we use the following definition.

Definition 3.5. *For fixed $[a, A]$, $[b, B]$, and n, let*

$$\alpha^{+}_{OPT} := \sup_{f \in \mathcal{F}} \inf_{P \in \mathcal{P}^f} \alpha^{+}(f, P).$$

Bringmann and Friedrich [4] showed the following result which identifies α^{+}_{OPT} equivalently to Theorem 3.1 for α^{*}_{OPT}. It will be reproven in Section 4.4 to illustrate the relationship between additive and multiplicative approximation ratios.

Theorem 3.3. $\alpha^{+}_{OPT} = \min\{A - a, B - b\}/n$.

Moreover, the analog for α^{*}_{ind} is defined similarly to Definition 3.3.

Definition 3.6. *For a unary indicator* ind *and fixed $[a, A]$, $[b, B]$, n, and $f \in \mathcal{F}$ let*

$$\alpha^{+}_{\text{ind}} := \sup_{f \in \mathcal{F}} \sup_{P \in \mathcal{P}^f_{\text{ind}}} \alpha^{+}(f, P).$$

4. HYPERVOLUME INDICATORS

In this section we come to concrete indicators for which upper bounds for α^{*}_{ind} or α^{+}_{ind} are known. First, we recap the general framework of the weighted hypervolume indicator. Then we review the results for the classic hypervolume indicator. After that, a new indicator designed for multiplicative approximation—the logarithmic hypervolume indicator—is proposed. We then show how to carry over additive approximation results to multiplicative approximation. Further, we discuss the combination of classic and logarithmic indicator.

4.1 The Weighted Hypervolume

The classic definition of the hypervolume indicator is the volume of the dominated portion of the objective space relative to a fixed footpoint called the reference point $R = (R_x, R_y) \succeq (A, B)$. As a general framework for our two indicators we use the more general weighted hypervolume indicator of [25]. It weights points with a weight distribution $w\colon \mathbb{R}^2 \to \mathbb{R}$. The *hypervolume* $\text{HYP}_w(P, R)$ (or $\text{HYP}_w(P)$ for short) of a solution set $P \in \mathcal{P}$ is then defined as

$$\text{HYP}_w(P) := \text{HYP}_w(P, R) := \iint_{\mathbb{R}^2} A(x, y)\, w(x, y)\, dy\, dx$$

where the attainment function $A\colon \mathbb{R}^2 \to \mathbb{R}$ is an indicator function on the objective space which describes the space below the reference point which weakly dominates P. Formally, $A(x, y) = 1$ if $(R_x, R_y) \succeq (x, y)$ and there is a $p = (p_x, p_y) \in P$ such that $(x, y) \succeq (p_x, p_y)$, and $A(x, y) = 0$ otherwise.

The original purpose of the weighted hypervolume indicator was to allow the decision maker to stress certain regions of the objective space. In this paper we unleash one of its hidden powers by showing that one gets a better multiplicative approximation choosing the right weight distribution.

4.2 The Classic Hypervolume

If w is the all-ones functions $\mathbb{1}$ with $\mathbb{1}(x, y) = 1$ for all $x, y \in \mathbb{R}$, above definition matches to the classic definition of the hypervolume indicator. In this case we write $\text{HYP} = \text{HYP}_{\mathbb{1}}$ for short. Bounds for this indicator are of particular interest. Bringmann and Friedrich [5] examined α^{*}_{HYP} and showed the following upper bound that has the same asymptotic behavior as α^{*}_{OPT}, but a much larger constant factor

Theorem 4.1. *Let $f \in \mathcal{F}$ and $n > 4$. If we have*

- $R_x \geqslant A + \frac{1}{n-2} \min\{\sqrt{Aa}\, B/b, A\sqrt{B/b}\}$ *and*
- $R_y \geqslant B + \frac{1}{n-2} \min\{\sqrt{Bb}\, A/a, B\sqrt{A/a}\}$

for the reference point $R = (R_x, R_y)$, then

$$\alpha^{*}_{HYP} \leqslant 1 + \frac{\sqrt{A/a} + \sqrt{B/b}}{n - 4}.$$

This shows that for sufficiently large n the hypervolume yields a good multiplicative approximation. However, this does not hold for small n as shown by the following lower bound of [5] for the case $A/a = B/b$.

Theorem 4.2. *Let $n \geqslant 7$ and $\frac{A}{a} = \frac{B}{b} \geqslant 13$. Then*

$$\alpha^{*}_{HYP} \geqslant 1 + \frac{2\sqrt{A/a - 1}}{3(n + 1)}.$$

Hence the multiplicative approximation ratio of HYP is exponentially worse in the ratio A/a. On the other hand, the following theorem of [4] shows that HYP has a close to optimal additive approximation ratio.

Theorem 4.3. *If $n > 2$ and*

$$(n - 2) \min\{R_x - A, R_y - B\} \geqslant \sqrt{(A - a)\,(B - b)}$$

we have

$$\alpha^{+}_{HYP} \leqslant \frac{\sqrt{(A - a)\,(B - b)}}{n - 2}.$$

Note that the precondition is fulfilled if n is large enough *or* if the reference point is sufficiently far away from (a, b). Compared to Theorem 3.3 it means that for $A - a \approx B - b$ and moderately sized n, α^{+}_{HYP} is very close to α^{+}_{OPT}. For $A - a \ll B - b$ (or the other way around) the constant is the geometric mean of $A - a$ and $B - b$ instead of the minimum of both.

4.3 The Logarithmic Hypervolume

Up to now we have mainly reviewed the results for the approximation ratios of the hypervolume indicator for maximization problems and have confirmed that the classic hypervolume indicator yields a good additive approximation also in the minimization setting. For getting a good multiplicative approximation HYP turned out to be inapplicable. We propose the *logarithmic hypervolume indicator* to address this problem. For a solution set $P \in \mathcal{P}$ and reference point $R = (R_x, R_y)$ with $(R_x, R_y) \succeq (A, B)$ we define

$$\text{logHYP}(P, R) := \text{HYP}_{\mathbb{1}}(\log P, \log R),$$

where $\log P := \{(\log x, \log y) \mid (x, y) \in P\}$ and $\log R := (\log R_x, \log R_y)$. Here, as in the classic case, the reference point is a parameter to be chosen by the user. Note, that we do not really change the axes of the problem to logarithmic scale: We only change the calculation of the hypervolume, not the problem itself.

Above definition is nice in that it allows to compute logHYP using existent implementations of algorithms for HYP, only wiring the input differently.

It is very illustrative, though, to observe that the logarithmic hypervolume indicator fits very well in the weighted hypervolume framework: An equivalent definition of logHYP is

$$\text{logHYP}(P, R) := \text{HYP}_{\hat{w}}(P, R),$$

where $\hat{w}(x, y) = 1/(xy)$ is the appropriate weight distribution.

Lemma 4.4. $\text{HYP}_{\mathbb{1}}(\log P, \log R) = \text{HYP}_{\hat{w}}(P, R)$.

Proof. Let $\{(x_1, y_1), \ldots, (x_k, y_k)\} \subseteq P$ be the set points in P not dominated by any other point in P with $x_1 < \ldots < x_k$, $y_1 > \ldots > y_k$. With $x_{k+1} := R_x$ we can then compute HYP as

$$\begin{aligned} \text{HYP}_{\mathbb{1}}(\log P, \log R) &= \sum_{i=1}^{k} \int_{\log x_i}^{\log x_{i+1}} \int_{\log y_i}^{\log R_y} 1 \, dy \, dx \\ &= \sum_{i=1}^{k} \int_{x_i}^{x_{i+1}} \int_{y_i}^{R_y} \frac{1}{xy} \, dy \, dx \\ &= \text{HYP}_{\hat{w}}(P, R). \end{aligned}$$ □

The first main result of this paper is now that the logarithmic hypervolume indicator yields a good multiplicative approximation, just like the classic hypervolume indicator yields a good additive approximation. The following result will be shown in Section 4.4.

Theorem 4.5. *If $n > 2$ and*

$$(n-2) \log \min\{R_x/A, R_y/B\} \geqslant \sqrt{\log(A/a)\log(B/b)}$$

we have

$$\alpha^*_{logHYP} \leqslant \exp\left(\frac{\sqrt{\log(A/a)\log(B/b)}}{n-2}\right).$$

Note that the precondition is fulfilled if n is large enough *or* we choose the reference point far enough away from (A, B).

This is a very good upper bound compared to $\alpha^*_{OPT} = \exp(\min\{\log(A/a), \log(B/b)\}/n)$. Also compare the next corollary to Corollary 3.2. Its proof is analogous to the one of Corollary 3.2 in [5].

Corollary 4.6. *For $\varepsilon \in (0, 1)$ and all*

$$n \geqslant 2 + \sqrt{\log(A/a)\log(B/b)} / \min\{\varepsilon, \log(R_x/A), \log(R_y/B)\}$$

we have

$$\alpha^*_{logHYP} \leqslant 1 + (1+\varepsilon)\, \frac{\sqrt{\log(A/a)\log(B/b)}}{n-2}.$$

Hence we get a much better constant factor than in the bound of α^*_{HYP}.

4.4 Relationship Between Additive and Multiplicative Approximation

Now we describe a relationship that allows to transfer results on multiplicative approximation into results on additive approximation and the other way around. This proves Theorems 3.3 and 4.5 and gives the intuition behind the logarithmic hypervolume indicator, as it is the classic hypervolume indicator transferred into the world of multiplicative approximation.

Consider a front $f^* \in \mathcal{F}_{[a^*,A^*]\to[b^*,B^*]}$ and a solution set $P^* \in \mathcal{P}^{f^*}$ that is a multiplicative α^*-approximation of f^*. This means that we have for any $\hat{x}^* \in [a^*, A^*]$ a point $(x^*, y^*) \in P^*$ with

$$\hat{x}^* \geqslant x^*/\alpha^* \quad \text{and} \quad f^*(\hat{x}^*) \geqslant y^*/\alpha^*.$$

Logarithmizing both inequalities gives

$$\log \hat{x}^* \geqslant \log x^* - \log \alpha^* \quad \text{and} \quad \log f^*(\hat{x}^*) \geqslant \log y^* - \log \alpha^*.$$

This corresponds to an additive approximation. We set $x^+ := \log x^*$, $y^+ := \log y^*$, $\hat{x}^+ := \log \hat{x}^*$, $\alpha^+ := \log \alpha^*$ and $f^+ := \log \circ f^* \circ \exp$ and get

$$\hat{x}^+ \geqslant x^+ - \alpha^+ \quad \text{and} \quad f^+(\hat{x}^+) \geqslant y^+ - \alpha^+.$$

This means that $P^+ := \{(\log x, \log y) \mid (x, y) \in P^*\}$ is an additive α^+-approximation of the front $f^+ \in \mathcal{F}_{[a^+,A^+]\to[b^+,B^+]}$ with $a^+ = \log a^*, A^+ = \log A^*, b^+ = \log b^*, B^+ = \log B^*$. Observe that this corresponds to logarithmizing both axes.

All operations we used above are invertible, so that we can do the same thing the other way round: Having a solution set P^+ on a front f^+ achieving an additive α^+-approximation, we get a solution set $P^* = \{(\exp x, \exp y) \mid (x, y) \in P^+\}$ on a front $f^* = \exp \circ f^+ \circ \log$ achieving a multiplicative α^*-approximation, with $\alpha^* = \exp \alpha^+$. Thereby the interval bounds like a^+ are also exponentiated and we get $a^* = \exp a^+$.

Let $\mathcal{F}^* := \mathcal{F}_{[a^*,A^*]\to[b^*,B^*]}$ and $\mathcal{F}^+ := \mathcal{F}_{[a^+,A^+]\to[b^+,B^+]}$. Then we have a bijection $\mathcal{F}^* \to \mathcal{F}^+$, $f^* \mapsto f^+$ and for any $f^* \in \mathcal{F}^*$ a bijection $\mathcal{P}^{f^*} \to \mathcal{P}^{f^+}$, $P^* \mapsto P^+$ that satisfies $\alpha^+(f^+, P^+) = \log \alpha^*(f^*, P^*)$. Though Theorem 3.3 was already proven in [4] (for maximization problems), it is interesting to reprove it to illustrate above technique as follows.

Proof of Theorem 3.3. We want to prove $\alpha^+_{OPT} = \min\{A^+ - a^+, B^+ - b^+\}/n$. By definition and the above bijection $(*)$ we know that

$$\begin{aligned} \alpha^+_{OPT} &= \sup_{f^+ \in \mathcal{F}^+} \inf_{P^+ \in \mathcal{P}^{f^+}} \alpha^+(f^+, P^+) \\ &\overset{(*)}{=} \sup_{f^+ \in \mathcal{F}^+} \inf_{P^+ \in \mathcal{P}^{f^+}} \log \alpha^*(f^*, P^*) \\ &\overset{(*)}{=} \sup_{f^* \in \mathcal{F}^*} \inf_{P^* \in \mathcal{P}^{f^*}} \log \alpha^*(f^*, P^*) \\ &= \log \sup_{f^* \in \mathcal{F}^*} \inf_{P^* \in \mathcal{P}^{f^*}} \alpha^*(f^*, P^*). \end{aligned}$$

The last expression matches the definition of α^*_{OPT}. We replace α^*_{OPT} using Theorem 3.1 and a^* by $\exp a^+$ etc. and get

$$\begin{aligned}\alpha^{+}_{OPT} &= \log \alpha^{*}_{OPT} \\ &= \log\left(\min\{A^*/a^*, B^*/b^*\}^{1/n}\right) \\ &= \min\{\log A^* - \log a^*, \log B^* - \log b^*\}/n \\ &= \min\{A^+ - a^+, B^+ - b^+\}/n. \qquad \square\end{aligned}$$

With similar reasoning we can now also prove Theorem 4.5.

Proof of Theorem 4.5. We want to show that

$$\alpha^{*}_{logHYP} \leqslant \exp\left(\frac{\sqrt{\log(A^*/a^*)\log(B^*/b^*)}}{n-2}\right).$$

For a solution set $P^* \in \mathcal{P}^*$ and a reference point $R^* = (R^*_x, R^*_y)$ we defined LOGHYP by setting $\text{LOGHYP}(P^*, R^*) = \text{HYP}_{\mathbb{1}}(\log P^*, \log R^*)$ with $\log P^* = \{(\log x, \log y) \mid (x,y) \in P^*\}$ and $\log R^* = (\log R^*_x, \log R^*_y)$. This $\log P^*$ is exactly P^+ as defined above. Writing $R^+ := \log R^*$ we thus have $\text{LOGHYP}(P^*, R^*) = \text{HYP}(P^+, R^+)$. Now, consider a solution set P^* maximizing $\text{LOGHYP}(P^*, R^*)$, thus, also maximizing $\text{HYP}(P^+, R^+)$. We know that P^+ is an α^{+}_{HYP}-approximation of the front f^+, so using Theorem 4.3 and above bijections we get

$$\begin{aligned}\alpha^*(f^*, P^*) &= \exp \alpha^+(f^+, P^+) \\ &\leqslant \exp\left(\sqrt{(A^+ - a^+)(B^+ - b^+)}/(n-2)\right) \\ &= \exp\left(\sqrt{\log(A^*/a^*)\log(B^*/b^*)}/(n-2)\right).\end{aligned}$$

The observation that the precondition of Theorem 4.3 transforms directly into the precondition of Theorem 4.5 concludes the proof. □

Note that we could also have proceeded the other way round: Proving a bound for α^{*}_{logHYP} and transforming it into a result for α^{+}_{HYP}. Above proof also makes clear why we defined LOGHYP as it is: Maximizing $\text{HYP}(P^+, R^+)$ gives a very good additive approximation which transforms into a very good multiplicative approximation going back to P^*.

4.5 A Hybrid Indicator

The results of Bringmann and Friedrich [4, 5] imply that guiding the search with the hypervolume indicator is an appropriate choice if we want an additive approximation. On the other hand, guiding the search with the proposed logarithmic hypervolume indicator is preferable if we want a multiplicative approximation.

Of course, it may happen that one wants an additive approximation of some objectives and a multiplicative of others. We propose a simple rule of thumb for this case: Logarithmize all objectives of the second type, i.e., that should get multiplicatively approximated (and leave the objectives of the first type as they are) and then compute the hypervolume indicator. This hybrid indicator should work as intended, i.e., maximizing it should give a good additive approximation of the objectives of the first type and a good multiplicative approximation of the objectives of the second type.

For details, assume we have two objectives, x and y, and want to approximate x additively and y multiplicatively. Then we use the hybrid indicator $\text{ind}(P, R) := \text{HYP}(P', (R_x, \log R_y))$, where $P' = \{(x_i, \log y_i) \mid (x_i, y_i) \in P\}$ and R is again a reference point. This indicator logarithmizes the y-axis and applies HYP afterwards. Along the lines of the proofs in this paper one can show that maximizing ind on a front f yields a solution set P with the following property: For any $\hat{x} \in [a, A]$ there is a $p = (x, y) \in P$ with

$$\hat{x} \geqslant x - \alpha^+ \quad \text{and} \quad f(\hat{x}) \geqslant y/\alpha^*,$$

where $\alpha^* = \exp \alpha^+$ and $\alpha^+ \leqslant \frac{\sqrt{(A-a)(\log(B)-\log(b))}}{n-2}$. This means that we get an additive approximation of x and a multiplicative approximation of y, as desired.

5. EXPERIMENTS

Above theoretical results indicate that for getting good multiplicative approximations one should maximize LOGHYP instead of HYP. This section substantiates this claim experimentally by taking a particular indicator-based selection scheme and running it with the indicator LOGHYP instead of HYP on typical test problems.

5.1 Experimental Setup

For the empirical evaluation, we implemented both the classic as well as the logarithmic hypervolume indicator in the indicator-based selection strategy of the $(\mu + 1)$-MO-CMA-ES (see [11, 21]) using the Shark software library [12]. We evaluate both algorithms on a set of benchmark functions taken from the literature, namely DTLZ1-7 (see [6]). Note that for all considered fitness functions, we lower bound the individual objectives to 10^{-6} as otherwise any multiplicative approximation ratio would be ∞ (if the solution set does not include exactly the leftmost point of the front). For each algorithm and each fitness function, we conducted 25 independent trials with 50,000 fitness function evaluations each. The parent population size μ has been chosen as 50. For both variants of the $(\mu + 1)$-MO-CMA-ES, we rely on the parameter setup presented in [21].

We evaluated the final fronts obtained by both variants of the $(\mu + 1)$-MO-CMA-ES with respect to the absolute hypervolume indicator and with respect to the multiplicative approximation ratio. In the latter case, we rely on a logarithmic sample of 10000 points of the true Pareto-optimal front as reference. For the statistical testing procedure, we refer again to [21].

5.2 Results

The results of the performance evaluation are presented in Figures 1, 2, and 3 and Table 1. For all fitness functions the $(\mu{+}1)$-MO-CMA-ES with the logarithmic hypervolume indicator outperformed its counterpart maximizing the original hypervolume indicator regarding the multiplicative approximation ratio (at a significance level of $p < 0.001$). For the functions DTLZ2, DTLZ4, DTLZ5, DTLZ6, and DTLZ7, both variants find solution sets very close to the optimal approximation ratio 1. However, the solutions found by guiding the search with the logarithmic hypervolume indicator still give a slightly better multiplicative approximation ratio. On DTLZ1 and DTLZ3, both variants are far from the optimal approximation ratio 1. Here the logarithmic hypervolume indicator has the largest improvement of more than 17% and 31%, respectively (cf. Table 1). Additionally, we also examined how both algorithms perform with respect to the classic hypervolume indicator. As expected, guiding the $(\mu + 1)$-MO-CMA-ES with the classic hypervolume indica-

	Empirical multiplicative approximation ratio for		
	$(\mu+1)$-MO-CMA-ES with HYP	$(\mu+1)$-MO-CMA-ES with logHYP	
DTLZ1	3.1581	**2.6864**	17.56% smaller
DTLZ2	1.0285	**1.0245**	0.4% smaller
DTLZ3	3.0734	**2.3398**	31.35% smaller
DTLZ4	1.0618	**1.0461**	1.51% smaller
DTLZ5	1.0285	**1.0245**	0.039% smaller
DTLZ6	1.0120	**1.0117**	0.029% smaller
DTLZ7	1.0133	**1.0131**	0.019% smaller

Table 1: Experimental results for the $(\mu+1)$-MO-CMA-ES with the classic and the logarithmic hypervolume indicator. Best values in each row are marked in bold if they are statistically significant at a significance level of $p < 0.001$.

	Theoretical worst-case bound for multiplicative approximation ratio	
OPT	$1 + \frac{\log(\min\{A/a, B/b\})}{n}$	(cf. Corollary 3.2)
HYP	$1 + \frac{\sqrt{A/a} + \sqrt{B/b}}{n-4}$	(cf. Theorem 4.1)
logHYP	$1 + \frac{\sqrt{\log(A/a)\,\log(B/b)}}{n-2}$	(cf. Corollary 4.6)

Table 2: Theoretical results for the optimal approximation ratio and upper bounds for the approximation ratios of HYP and LOGHYP. See the cited theorems for the precise statements. The results for OPT and HYP are proven in [4, 5]. The result for LOGHYP is shown in Section 4.4 of this paper.

tor gives a larger (classic) hypervolume for DTLZ1, DTLZ2, DTLZ5, and DTLZ6 (see Figures 1, 2, and 3). In case of DTLZ4, both variants are on par w.r.t. the classic hypervolume. Surprisingly, this is not the case for DTLZ3 and DTLZ7. For these two fitness functions, the variant with the logarithmic hypervolume indicator achieves a large classic hypervolume after 50,000 fitness function evaluations. We expect that this results from the limited number of fitness function evaluations and that also for DTLZ3 and DTLZ7 the classic hypervolume indicator achieves a larger classic hypervolume for a larger number of fitness function evaluations.

On the set of DTLZ fitness functions our empirical evaluation matches with the theoretical results obtained beforehand. They further suggest to rely on the logarithmic hypervolume indicator if an optimal multiplicative approximation ratio is desired.

6. CONCLUSION

After it was shown in [4, 5] that the classic hypervolume indicator does not give an optimal multiplicative approximation factor, it was natural to ask what other indicator might have this desirable property. We defined a new indicator LOGHYP and proved that it yields a close-to-optimal multiplicative approximation ratio. This was confirmed empirically on a set of typical benchmark functions.

References

[1] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Investigating and exploiting the bias of the weighted hypervolume to articulate user preferences. In *Proc. 11th annual Conference on Genetic and Evolutionary Computation (GECCO '09)*, pp. 563–570. ACM Press, 2009.

[2] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the hypervolume indicator: optimal μ-distributions and the choice of the reference point. In *Proc. 10th International Workshop on Foundations of Genetic Algorithms (FOGA '09)*, pp. 87–102. ACM Press, 2009.

[3] N. Beume, B. Naujoks, and M. T. M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181:1653–1669, 2007.

[4] K. Bringmann and T. Friedrich. Tight bounds for the approximation ratio of the hypervolume indicator. In *Proc. 11th International Conference Parallel Problem Solving from Nature (PPSN XI)*, Vol. 6238 of *LNCS*, pp. 607–616. Springer, 2010.

[5] K. Bringmann and T. Friedrich. The maximum hypervolume set yields near-optimal approximation. In *Proc. 12th annual Conference on Genetic and Evolutionary*

Computation (GECCO '10), pp. 511–518. ACM Press, 2010.

[6] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable multi-objective optimization test problems. In *Proc. Congress on Evolutionary Computation (CEC '02)*, pp. 825–830. IEEE Press, 2002.

[7] K. Deb, M. Mohan, and S. Mishra. Evaluating the ε-domination based multi-objective evolutionary algorithm for a quick computation of Pareto-optimal solutions. *Evolutionary Computation*, 13:501–525, 2005.

[8] M. T. M. Emmerich, N. Beume, and B. Naujoks. An EMO algorithm using the hypervolume measure as selection criterion. In *Proc. Third International Conference on Evolutionary Multi-Criterion Optimization (EMO '05)*, pp. 62–76. Springer, 2005.

[9] T. Friedrich, C. Horoba, and F. Neumann. Multiplicative approximations and the hypervolume indicator. In *Proc. 11th annual Conference on Genetic and Evolutionary Computation (GECCO '09)*, pp. 571–578. ACM Press, 2009.

[10] C. Igel, N. Hansen, and S. Roth. Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15:1–28, 2007.

[11] C. Igel, T. Suttorp, and N. Hansen. Steady-state selection and efficient covariance matrix update in the multi-objective CMA-ES. In *Proc. 4th International Conference on Evolutionary Multi-Criterion Optimization (EMO '07)*, Vol. 4403 of *LNCS*, pp. 171–185. Springer, 2007.

[12] C. Igel, T. Glasmachers, and V. Heidrich-Meisner. Shark. *Journal of Machine Learning Research*, 9:993–996, 2008.

[13] J. D. Knowles and D. Corne. Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Trans. Evolutionary Computation*, 7:100–116, 2003.

[14] J. D. Knowles, D. W. Corne, and M. Fleischer. Bounded archiving using the Lebesgue measure. In *Proc. IEEE Congress on Evolutionary Computation (CEC '03)*, Vol. 4, pp. 2490–2497. IEEE Press, 2003.

[15] R. Kumar and N. Banerjee. Running time analysis of a multiobjective evolutionary algorithm on simple and hard problems. In *Proc. 8th International Workshop on Foundations of Genetic Algorithms (FOGA '05)*, Vol. 3469 of *LNCS*, pp. 112–131. Springer, 2005.

[16] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10(3):263–282, 2002.

[17] G. Lizarraga-Lizarraga, A. Hernandez-Aguirre, and S. Botello-Rionda. G-metric: an m-ary quality indicator for the evaluation of non-dominated sets. In *Proc. 10th annual Conference on Genetic and Evolutionary Computation (GECCO '08)*, pp. 665–672. ACM Press, 2008.

[18] C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. In *Proc. 41st annual Symposium on Foundations of Computer Science (FOCS '00)*, pp. 86–92. IEEE Press, 2000.

[19] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, 3rd edition, 1987.

[20] T. Suttorp, N. Hansen, and C. Igel. Efficient covariance matrix update for variable metric evolution strategies. *Machine Learning*, 75:167–197, 2009.

[21] T. Voß, N. Hansen, and C. Igel. Improved step size adaptation for the MO-CMA-ES. In *Proc. 12th Annual Conference on Genetic and Evolutionary Computation Conference (GECCO '10)*, pp. 487–494. ACM Press, 2010.

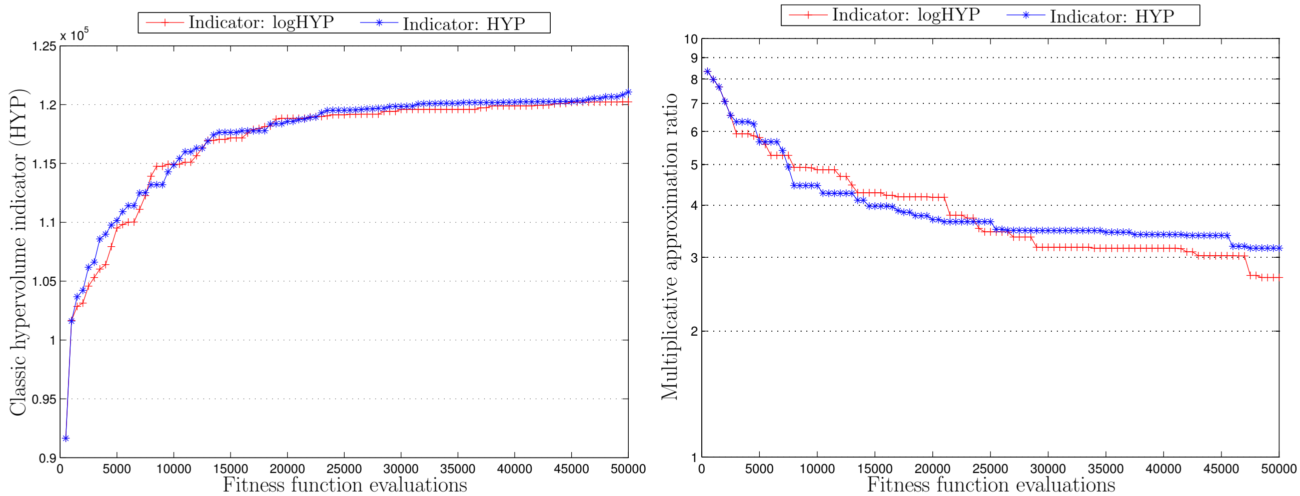
[22] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *Proc. 8th International Conference Parallel Problem Solving from Nature (PPSN VIII)*, Vol. 3242 of *LNCS*, pp. 832–842. Springer, 2004.

[23] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Trans. Evolutionary Computation*, 3:257–271, 1999.

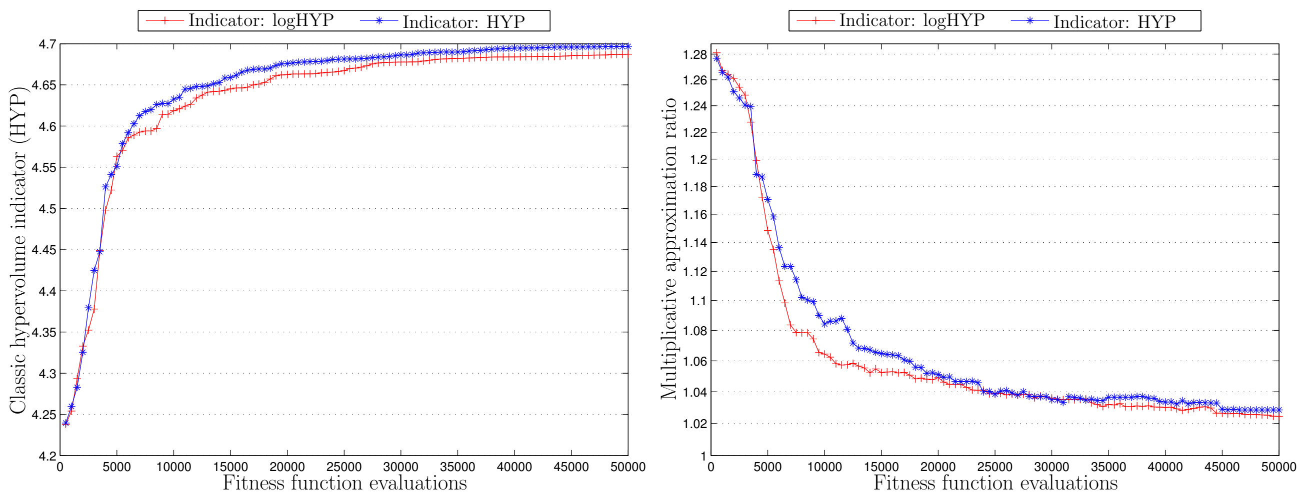
[24] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Trans. Evolutionary Computation*, 7:117–132, 2003.

[25] E. Zitzler, D. Brockhoff, and L. Thiele. The hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted integration. In *Proc. Fourth International Conference on Evolutionary Multi-Criterion Optimization (EMO '07)*, Vol. 4403 of *LNCS*, pp. 862–876. Springer, 2007.

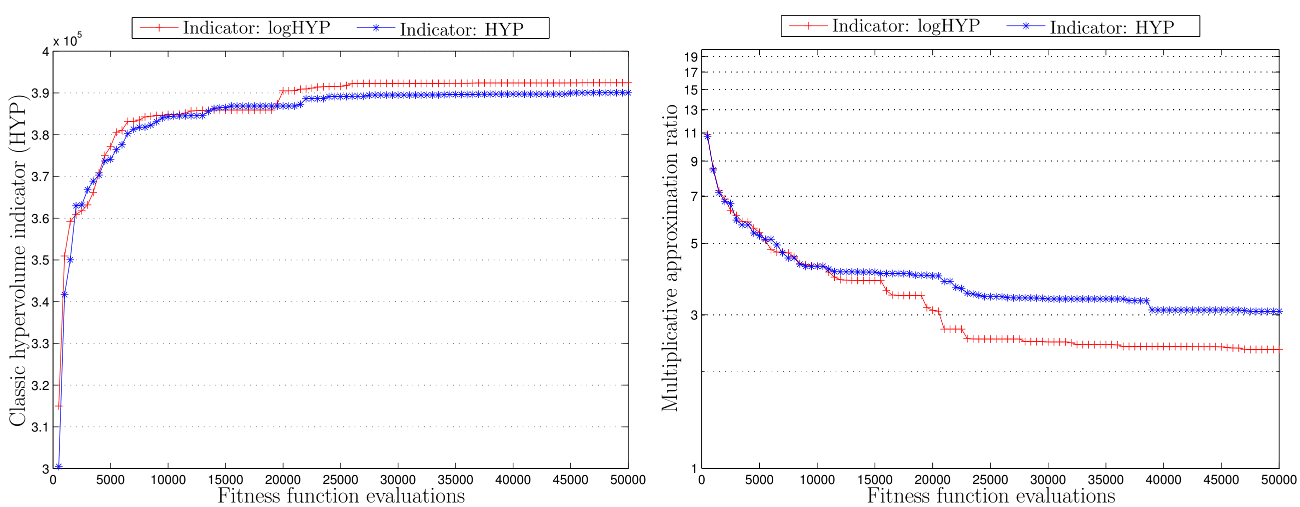
[26] E. Zitzler, L. Thiele, and J. Bader. On set-based multiobjective optimization. *IEEE Trans. Evolutionary Computation*, 14:58–79, 2010.



(a) Fitness function DTLZ1.

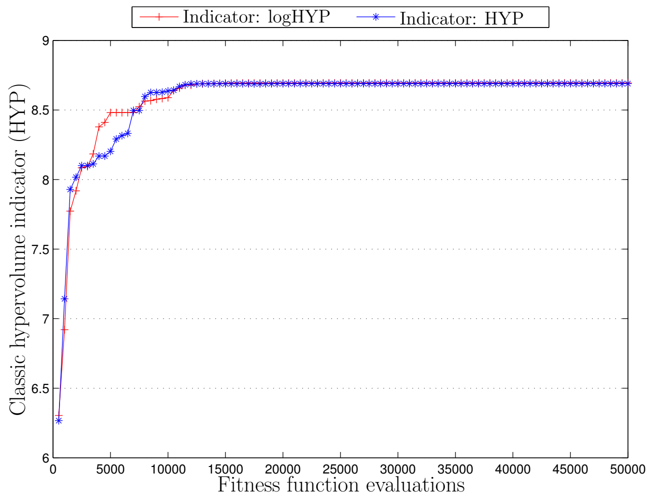


(b) Fitness function DTLZ2.

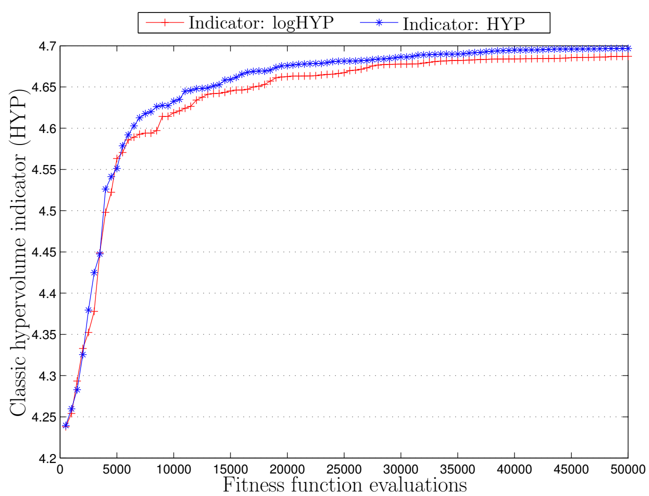


(c) Fitness function DTLZ3.

Figure 1: Performance results of the $(\mu + 1)$-MO-CMA-ES with the two different indicators. The left column shows the results for the classic hypervolume indicator. The right column shows the results for the multiplicative approximation ratio (scaled logarithmically).

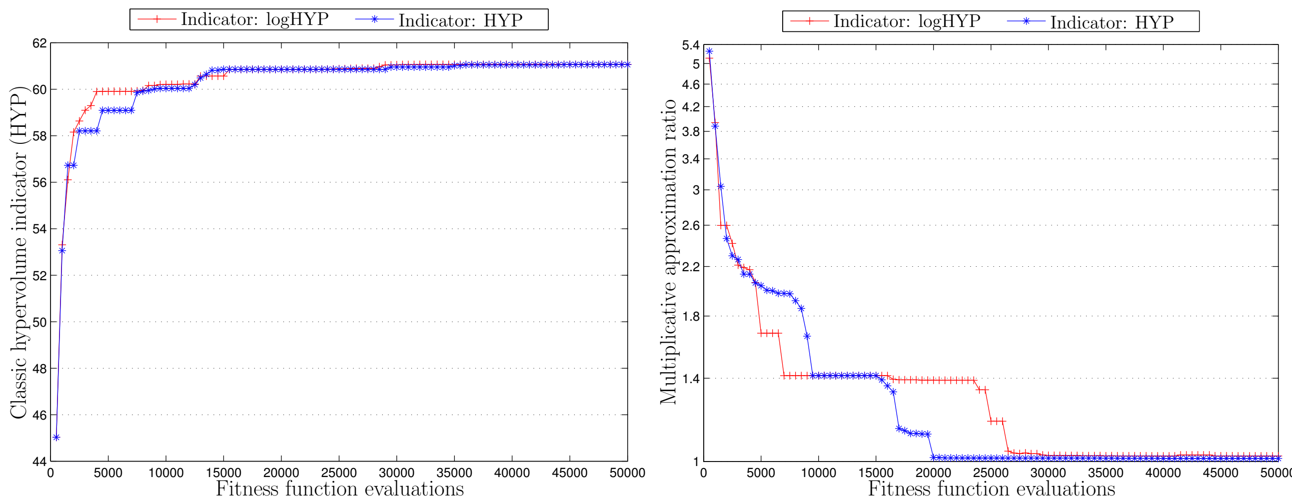


(a) Fitness function DTLZ4.

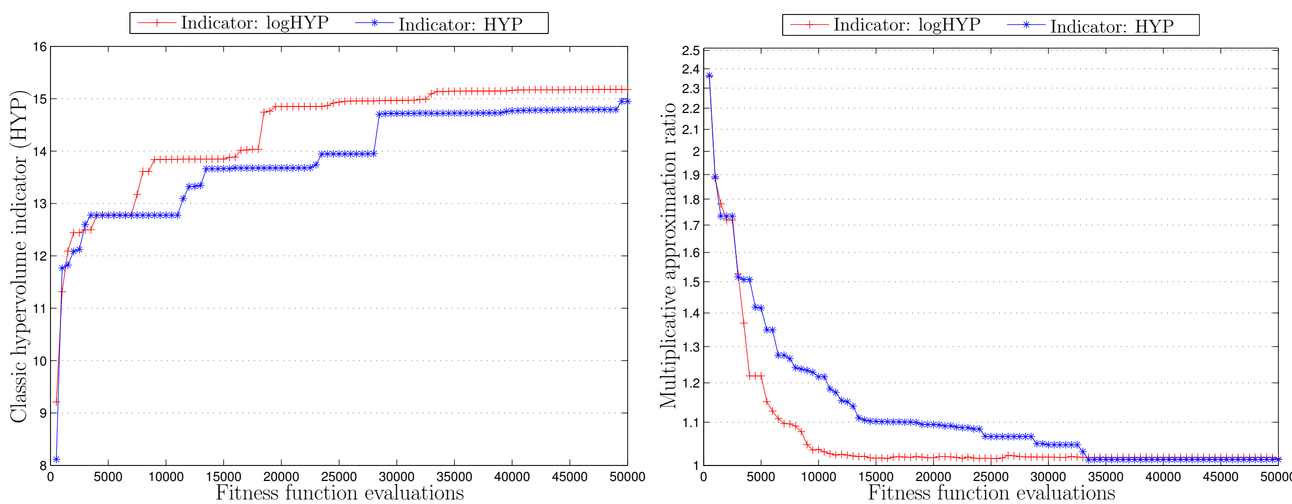


(b) Fitness function DTLZ5.

Figure 2: Performance results of the $(\mu + 1)$-MO-CMA-ES with the two different indicators. The left column shows the results for the classic hypervolume indicator. The right column shows the results for the multiplicative approximation ratio (scaled logarithmically).



(a) Fitness function DTLZ6.



(b) Fitness function DTLZ7.

Figure 3: Performance results of the $(\mu + 1)$-MO-CMA-ES with the two different indicators. The left column shows the results for the classic hypervolume indicator. The right column shows the results for the multiplicative approximation ratio (scaled logarithmically).