Multi-Objective Optimization

Carlos A. Coello Coello

ccoello@cs.cinvestav.mx
CINVESTAV-IPN
Evolutionary Computation Group (EVOCINV)
Computer Science Department
Av. IPN No. 2508, Col. San Pedro Zacatenco
México, D.F. 07360, MEXICO

Patna, India, December, 2016

LECTURE 5

Performance Indicators

Zitzler et al [2000] indicated that, when assessing performance of a MOEA, three are the desirable goals:

1. The distance of the resulting nondominated set to the Pareto optimal front should be minimized.
2. A good distributions of the solutions found is desirable. Normally, the aim is a uniform distribution of solutions.
3. The extent of the obtained Pareto front should be maximized (i.e., for each objective, a wide range of values should be present).

Performance Indicators

It should be clear, however, that these goals are not always desirable or even relevant. For example, if the true Pareto front of a problem is not uniformly distributed, the second goal would not apply.

Also, if the true Pareto front contained a single solution, the third goal would not be relevant.

Because of these issues, normally the first goal is the one with the highest relevance when assessing performance of a MOEA.

A very interesting issue regarding these three goals is that no performance indicator can assess them all. Normally a combination of indicators is required.

Performance Indicators

Before we start describing the most commonly adopted performance indicators, we need to define the preference relations that are normally adopted to compare objective vectors.

Such preference relations are shown in the next slide.

Joshua Knowles, Lothar Thiele and Eckart Zitzler, "**A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers**", Technical Report No. 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, February 2006 (revised version).

Performance Indicators

relation	interpretation in objective space	
strictly dominates	$\mathbf{z}^1 \prec\prec \mathbf{z}^2$	$\mathbf{z}^1$ is better than $\mathbf{z}^2$ in all objectives
dominates	$\mathbf{z}^1 \prec \mathbf{z}^2$	$\mathbf{z}^1$ is not worse than $\mathbf{z}^2$ in all objectives and better in at least one objective
weakly dominates	$\mathbf{z}^1 \preceq \mathbf{z}^2$	$\mathbf{z}^1$ is not worse than $\mathbf{z}^2$ in all objectives
incomparable	$\mathbf{z}^1 \| \mathbf{z}^2$	neither $\mathbf{z}^1 \preceq \mathbf{z}^2$ nor $\mathbf{z}^2 \preceq \mathbf{z}^1$
indifferent	$\mathbf{z}^1 \sim \mathbf{z}^2$	$\mathbf{z}^1$ has the same value as $\mathbf{z}^2$ in each objective

Please note that the corresponding relations on decision vectors are defined on the basis of the associated objective vectors, i.e.,
$\mathbf{x}^1$ *rel* $\mathbf{x}^2 \Leftrightarrow \mathbf{f}(\mathbf{x}^1)$ *rel* $\mathbf{f}(\mathbf{x}^2)$.

Please note that the relations $\succ$, $\succ\succ$ and $\succeq$ are used accordingly with reverse order of the arguments, e.g., $\mathbf{z}^1 \succ \mathbf{z}^2$ is equivalent to $\mathbf{z}^2 \prec \mathbf{z}^1$.

Note that the **indifference** relation only makes sense with regards to decision space; in objective space, it simply means equality.

Performance Indicators

relation	interpretation in objective space	
strictly dominates	$A \prec\prec B$	every $\mathbf{z}^2 \in B$ is strictly dominated by at least one $\mathbf{z}^1 \in A$
dominates	$A \prec B$	every $\mathbf{z}^2 \in B$ is dominated by at least one $\mathbf{z}^1 \in A$
better	$A \lhd B$	every $\mathbf{z}^2 \in B$ is weakly dominated by at least one $\mathbf{z}^1 \in A$ and $A \nsim B$
weakly dominates	$A \preceq B$	every $\mathbf{z}^2 \in B$ is weakly dominated by at least one $\mathbf{z}^1 \in A$
incomparable	$A \| \| B$	neither $A \preceq B$ nor $B \preceq A$
indifferent	$A \sim B$	$A \preceq B$ and $B \preceq A$

Selected preference relations on Pareto front approximations; the corresponding relations on Pareto set approximations are defined by considering the associated Pareto front approximations.

The relations $\succ$, $\succ\succ$, $\succeq$ and $\rhd$ are used accordingly with reversed order of the arguments, e.g., $A \succ B$ is equivalent to $B \prec A$.

Notice that (1) $A \prec\prec B \Rightarrow A \prec B \Rightarrow A \lhd B$ and (ii) two indifferent Pareto front approximations are identical, while this does not need to hold for two indifferent Pareto set approximations.

Performance Indicators

We will now analyze several performance indicators. For this analysis, we will follow the paper from Knowles and Corne [2002].

In that work, they follow Hansen and Jaszkiewicz [1998], who considered the problem of evaluating approximations to the true Pareto front. For that sake, they define a number of **outperformance relations** that express the relationship between two sets of internally nondominated objective vectors, A and B.

In the following, $\mathbf{ND}(S)$ denotes the nondominated points in S:

Weak Outperformance: $AO_WB \iff \mathbf{ND}(A \cup B) = A$ and $A \neq B$. For example, A **weakly outperforms** B if all points in B are 'covered' by those in A (where 'covered' means is equal to or dominates) and there is at least one point in A that is not contained in B.

Performance Indicators

Strong Outperformance: $AO_S B \iff \mathbf{ND}(A \cup B) = A$ and $B \setminus \mathbf{ND}(A \cup B) \neq \emptyset$. In words, A **strongly outperforms** B if all points in B are covered by those in A and some point in B is dominated by a point in A.

Complete Outperformance: $AO_C B \iff \mathbf{ND}(A \cup B) = A$ and $B \cap \mathbf{ND}(A \cup B) = \emptyset$. In words, A **completely outperforms** B if each point in B is dominated by a point in A.

Notice that $AO_C B \Rightarrow AO_S B \Rightarrow AO_W B$. In other words, complete outperformance is the strongest and weak outperformance is the weakest of the relations.

All of these relations describe relationships between approximations to the true Pareto front since they are compatible with, and only depend upon, standard Pareto dominance. They are not metrics of performance, but we can use them to assess the usefulness of nondominated set comparison metrics.

Performance Indicators

Hansen and Jaszkiewicz [1998] formally define **compatibility** and **weak compatibility** with an outperformance relation, as follows:

Weak Compatibility: A comparison metric R is **weakly compatible** with an outperformance relation $\leq$ if for each pair of nondominated sets A, B with $A \leq B$, R will evaluate A as being no worse than B.

Compatibility: A comparison metric R is **compatible** with an outperformance relation $\leq$ if for each pair of nondominated sets A and B, such that $A \leq B$, R will evaluate A as being better than B.

These definitions will be used to compare and contrast different nondominated set comparison metrics.

Performance Indicators

For each of the performance measures that we will discuss next, we will analyze three things:

1. Pareto compatibility
2. Advantages
3. Disadvantages

In the analysis, the purpose of the corresponding performance measure will be provided, as well as the way in which it actually compares two approximations A and B.

Performance Indicators

It is worth noting that there are several possible ways of doing this. A "direct comparative" measure compares A and B directly using a scalar measure $R(A, B)$ to describe how much better A is than B.

If $R(A, B) = c - R(A, B)$ for some constant c for all pairs of nondominated sets A, B, then R is "symmetric".

Another option is a "reference measure", which uses a reference set (perhaps the true Pareto front), and then it scores both approximation sets (A and B) against this reference set and compares the results.

Performance Indicators

Any direct comparative measure can also be used as a reference measure by specifying a particular reference set. However, the converse is not true because their definition depends on a particular reference set (normally the true Pareto front).

Finally, we have the "independent measure", which measures some property of each set that is not dependent on any other, or on any reference set.

Another important feature of a performance measure is whether it induces a complete ordering of all possible nondominated sets. This ensures transitivity, so that when A, B and C are compared, if A beats B and B beats C, then it is always true that A beats C.

Performance Indicators

It is often the case that direct comparative measures do not induce a complete ordering, and the relations between different sets may be intransitive.

Using reference sets in such cases would ensure transitivity.

Transitivity is not generally a problem with independent measures, as they all induce a complete ordering.

Finally, it will also be noted if a measure is a cardinal measure (based on counting the number of vectors in some set) or a non-cardinal measure.

Performance Indicators

Pareto compatibility is concerned with compatibility with the outperformance relations O_W, O_S and O_C.

The less compatible the measure is, the more misleading it may be, giving scores for nondominated sets that do not accurately reflect their relative worth in a Pareto sense.

The hardest relation to be (weakly) compatible with is O_W, and the easiest is O_C. Compatibility with O_W is necessary and sufficient to ensure **monotony** and sufficient but not necessary for ensuring **relativity**, which are defined as follows:

- **(weak) monotony:** Given a nondominated set A, adding a nondominated point improves (does not degrade) its evaluation.
- **(weak) relativity:** The evaluation of the true Pareto front is (non)-uniquely optimal (i.e., all other nondominated sets have a strictly inferior (non-superior) evaluation).

Weak compatibility with O_W is sufficient for the weak versions to be exhibited.

Performance Indicators

For the analysis of advantages and disadvantages of the performance measures, several issues will be considered, including: compatibility with the outperformance relations, computational cost, whether or not they are scaling independent (is the ordering of approximations affected if one objective is scaled relative to the others?), and whether it relies on knowledge of the true Pareto front or any other reference set or point, and whether it can differentiate between different levels of complete outperformance.

This means that given three approximation sets A, B and C, with AO_CB, and BO_CC, would the measure give a different evaluation if A and B were compared than if A and C were compared?

The Hypervolume or the $\mathcal{S}$ metric

The hypervolume indicator was proposed by Zitzler [1999] and it is defined as the area of coverage of PF_{known} with respect to the objective space.

This equates to the summation of all the rectangular areas, bounded by some reference point and $(f_1(\vec{x}), f_2(\vec{x}))$ (for the bi-objective case). Mathematically, this is described as follows:

$$\mathbf{HV} \triangleq \left\{ \bigcup_i vol_i | vec_i \in \mathrm{PF}_{known} \right\} \quad (1)$$

This is an independent indicator (although needs a reference point to be chosen), so it induces a complete ordering and is non-cardinal.

Eckart Zitzler, "**Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications**", PhD thesis, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, November 1999.

Performance Indicators

The Hypervolume or the $\mathcal{S}$ metric

Pareto Compatibility: Compatible with O_W provided that the reference point is set so that all feasible nondominated sets are evaluated as positive. Normally, the worst objective function values are used to define this reference point.

Advantages: Compatible with the outperformance relations, independent, differentiates between different degrees of complete outperformance of two sets, scaling independent, and its meaning is intuitive.

Disadvantages: It requires defining some upper boundary of the regions within which all feasible points will lie. This choice does not affect the ordering of nondominated sets, and is relatively arbitrary. It has a large computational overhead $O(n^{k+1})$, which makes it impossible to use it for many-objective optimization. However, in recent years, a lot of research has been done regarding the design of more efficient algorithms to compute it. Another more subtle caveat of the hypervolume is that it multiplies different objectives together, although this is not supposed to matter, since this indicator is scaling independent anyway, and the units are irrelevant.

Suggested Readings on the Hypervolume

Nicola Beume, Carlos M. Fonseca, Manuel Lopez-Ibañez, Luis Paquete and Jan Vahrenhold, "**On the Complexity of Computing the Hypervolume Indicator**", *IEEE Transactions on Evolutionary Computation*, Vol. 13, No. 5, pp. 1075–1082, October 2009.

Lucas Bradstreet, Lyndon While and Luigi Barone, "**A Fast Incremental Hypervolume Algorithm**", *IEEE Transactions on Evolutionary Computation*, Vol. 12, No. 6, pp. 714–723, December 2008.

Karl Bringmann and Tobias Friedrich, "**An Efficient Algorithm for Computing Hypervolume Contributions**", *Evolutionary Computation*, Vol. 18, No. 3, pp. 383–402, Fall 2010.

Hisao Ishibuchi, Noritaka Tsukamoto, Yuji Sakane and Yusuke Nojima, "**Hypervolume Approximation Using Achievement Scalarizing Functions for Evolutionary Many-Objective Optimization**", in *2009 IEEE Congress on Evolutionary Computation (CEC'2009)*, pp. 530–537, IEEE Press, Trondheim, Norway, May 2009.

Error Ratio

This indicator was proposed by Van Veldhuizen [1999] and is defined as:

$$\mathbf{ER} \triangleq \frac{\sum_{i=1}^{n} e_i}{n} \tag{2}$$

where n is the number of vectors in the true Pareto front. In this case, $e_i = 0$ if vector i is in the true Pareto front and 1 otherwise. Lower values or **ER** represent better nondominated sets.

ER is the proportion of non true Pareto points in our approximation. It is a reference indicator using as a reference set to the true Pareto front. It induces a total ordering and is cardinal.

David A. Van Veldhuizen, "**Multiobjective Evolutionary Algorithms: Classifications, Analyses and New Innovations**", PhD thesis, Department of Electrical and Computer Engineering. Graduate School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, May 1999.

Error Ratio

Pareto Compatibility: It is only weakly compatible with O_C. It is not weakly compatible with O_S or O_W. For example, if an algorithm finds two nondominated vectors, one in the true Pareto front, and other far from the true Pareto front, then its error ratio is 0.5.

If it finds one hundred solutions, 99 of which are very close to the true Pareto front (and perhaps uniformly distributed along objective function space), and one (as before) which is in the true Pareto front, then its error ratio will be 0.99.

Clearly, the second set of points is better, but the first has a much better **ER**. It strongly violates monotony; given a nondominated set A with one ore more Pareto optimal points in it, addition of more nondominated but non-Pareto optimal points makes the **ER** score worse.

It violates relativity too, since any non-empty subset of the Pareto optimal set has an optimal error ratio. However, it exhibits weak relativity because the Pareto front itself is evaluated not worse than any other set.

Performance Indicators

Error Ratio

Advantages: It is easy to understand and easy to calculate. It is scaling independent. For test problems, it can be used as a quick and rough means of assessing progress towards the true Pareto front.

Disadvantages: It is required to know the true Pareto front. It is incompatible with the outperformance relations.

Generational Distance

The Generational Distance (GD) reports how far, on average, PF_{known} (our approximation of the Pareto front) is from PF_{true} (the true Pareto front). It is mathematically defined as [Veldhuizen, 1999]:

$$\mathbf{GD} \triangleq \frac{(\sum_{i=1}^{n} d_i^p)^{1/p}}{n} \tag{3}$$

where n is the number of vectors in PF_{known}, $p = 2$, and d_i is the Euclidean distance in objective space between each member, i, of PF_{known} and the closest member in PF_{true} to that member, i. When $GD = 0$, $\mathrm{PF}_{known} = \mathrm{PF}_{true}$ and, in general, lower values of GD represent better approximations to the true Pareto front.

It is a reference measure which uses the true Pareto front as a reference. It induces a total ordering and is non-cardinal. It can be quite misleading, since it is very easy to produce an example in which it provides an inappropriate result.

Performance Indicators

Generational Distance

Pareto compatibility: It is not weakly compatible with O_W, but is compatible with O_S. It violates weak monotony. For example, the GD score favors one vector close to the true Pareto front over a set containing that vector plus others, as long as the others are not closer on average to the true Pareto front than the first one. It does exhibit weak relativity, since any subset of the true Pareto front has an optimal GD.

Advantages: For a constant size of the nondominated set, GD is compatible with O_S. It is relatively cheap to calculate.

Disadvantages: Since it is not compatible with O_W it cannot be used confidently for nondominated sets that are changing in cardinality (which is something typical of the nondominated portion of a MOEA population over time). It cannot reliably differentiate between different levels of complete outperformance. It is required to know the true Pareto front of the problem. The distance metric will either add or multiply different objectives together, introducing scaling and normalization issues that cannot be properly solved without reference to additional preference information.

Inverted Generational Distance

The Inverted Generational Distance (GD) reports how far, on average, is PF_{true} (the true Pareto front) from PF_{known} (our approximation of the Pareto front). It is an apparently trivial variant of GD, which is mathematically defined as [Coello, 2005]:

$$\mathbf{IGD} \triangleq \frac{(\sum_{i=1}^{n} d_i^p)^{1/p}}{n} \tag{4}$$

where n is the number of vectors in PF_{true}, $p = 2$, and d_i is the Euclidean distance in objective space between each member, i, of PF_{true} and the closest member in PF_{known} to that member, i. When $IGD = 0$, $\mathrm{PF}_{known} = \mathrm{PF}_{true}$ and, in general, lower values of IGD represent better approximations to the true Pareto front.

It is worth noting, however, that by measuring the average distance from the true Pareto front to our Pareto front approximation, many of the problems of the original GD are solved, although this indicator is not free of problems. In fact, it is not Pareto-compliant [Zitzler, 2003].

Performance Indicators

Inverted Generational Distance

Something interesting is that IGD can assess not only convergence, but also spread of solutions along the Pareto front. Due to its simplicity, several authors have used it in many-objective optimization during the last few years, in spite of the criticism that it has received. This is mainly due to the high computational cost of the hypervolume in many-objective optimization, which makes it unaffordable.

Carlos A. Coello Coello and Nareli Cruz Cortés, "**Solving Multiobjective Optimization Problems using an Artificial Immune System**", *Genetic Programming and Evolvable Machines*, Vol. 6, No. 2, pp. 163–190, June 2005.

Performance Indicators

Inverted Generational Distance+

Recently, Ishibuchi [2015] proposed a variant of IGD which is weakly Pareto compliant. His idea is very simple: he proposed modifying the distance calculation between a solution and a reference point in the GD and IGD indicators by taking into account the Pareto dominance relation between them. If a solution is dominated by a reference point, then he uses the Euclidean distance with no modification. However, if they are non-dominated with respect to each other, he calculates the minimum distance from the reference point to the dominated region by the solution.

Inverted Generational Distance+

This distance can be viewed as an amount of the inferiority of the solution (i.e., the insufficiency of its objective values) in comparison with the reference point. Only inferior objective values of the solution to the reference point are used in their distance calculation.

This indicator, which is called $\mathbf{IGD}^+$ requires the definition of an inferiority (i.e., insufficiency) vector $\mathbf{d}^+ = (d_1^+, d_2^+, \ldots d_m^+)$ as follows:

$$d_i^+ = \max\{a_i - z_i, 0\}, \quad i = 1, 2, \ldots, m \tag{5}$$

When $\mathbf{z} \succ \mathbf{a}$ holds, $\mathbf{d}^+$ is the same as $\mathbf{d} = \mathbf{a} - \mathbf{z}$, where $\mathbf{z}$ is our reference point set (the true Pareto front, in this case). However, when $\mathbf{z} \succ \mathbf{a}$ does not hold, $\mathbf{d}^+$ is different from $\mathbf{d} = \mathbf{a} - \mathbf{z}$, since only the positive elements of $\mathbf{d}$ remain in $\mathbf{d}^+$. Thus, the indicator is defined as (assuming minimization):

$$\mathbf{IGD}^+(\mathbf{z}, \mathbf{a}) = \sqrt{{d_1^+}^2 + \ldots + {d_m^+}^2} =$$

$$\sqrt{(\max\{a_1 - z_i, 0\})^2 + \ldots + (\max\{a_m - z_m, 0\})^2} \tag{6}$$

Performance Indicators

Inverted Generational Distance+

This is, with no doubt, a very interesting indicator that, however, has some resemblance with the additive version of the ϵ indicator [Zitzler, 2003], which is used for all elements of all reference points. In other words, the maximum distance over all reference points (and over all objective of each reference point) is calculated in the ϵ indicator, instead of the average distance as done in **IGD**$^+$.

Hisao Ishibuchi, Hiroyuki Masuda, Yuki Tanigaki and Yusuke Nojima, "**Modified Distance Calculation in Generational Distance and Inverted Generational Distance**", in António Gaspar-Cunha et al. (Eds), *Evolutionary Multi-Criterion Optimization, 8th International Conference, EMO 2015*, pp. 110–125, Springer. Lecture Notes in Computer Science Vol. 9019, Guimarães, Portugal, March 29 - April 1, 2015.

Performance Indicators

Δ_p

Another interesting variant of **GD** and **IGD** is the Δ_p indicator proposed by Schütze et al. [2012].

Oliver Schütze, Xavier Esquivel, Adriana Lara and Carlos A. Coello Coello, "**Using the Averaged Hausdorff Distance as a Performance Measure in Evolutionary Multiobjective Optimization**", *IEEE Transactions on Evolutionary Computation*, Vol. 16, No. 4, pp. 504–522, August 2012.

Δ_p

The Δ_p indicator, which can be viewed as an averaged Hausdorff distance between an approximation set and the Pareto front of a MOP, is composed of two (slightly modified) quality indicators: Generational Distance (GD) and Inverted Generational Distance (IGD).

Given an approximation set A and a discretized Pareto front $\text{PF} = (p_1, p_2, \ldots, p_{|\text{PF}|})$ of a MOP, the (slightly modified) GD indicator is defined as:

$$I_{\text{GD}_p} = \left(\frac{1}{|A|} \sum_{i=1}^{|A|} d_i^p \right)^{\frac{1}{p}} \tag{7}$$

where d_i is the Euclidean distance from a_i to its nearest member of PF_{true}.

Δ_p

Given an approximation set A and a discretized Pareto front $PF = (p_1, p_2, \ldots, p_{|PF|})$ of a MOP, the (slightly modified) IGD indicator is defined as:

$$I_{IGD_p} = \left(\frac{1}{|PF|} \sum_{i=1}^{|PF|} \tilde{d}_i^p \right)^{\frac{1}{p}} \tag{8}$$

where d_i is the Euclidean distance from pf_i to its nearest member of A.

Both I_{GD_p} and I_{IGD_p} have (weak) metric properties:

- I_{GD_p} and I_{IGD_p} are non-negative
- I_{GD_p} and I_{IGD_p} are non-symmetric
- I_{GD_p} and I_{IGD_p} don't satisfy the (relaxed) triangle inequality

Performance Indicators

Δ_p

Given an approximation set A and a discretized Pareto front $\text{PF} = (p_1, p_2, \ldots, p_{|\text{PF}|})$ of a MOP, the Δ_p indicator is defined as:

$$I_{\Delta_p} = \max\left(I_{\text{GD}_p}, I_{\text{IGD}_p}\right) \tag{9}$$

The Δ_p indicator has better metric properties than either the GD or the IGD indicators:

- It is positive and symmetric: I_{Δ_p} is a semi-metric.
- If the magnitudes of the sets are bounded, the relaxed triangle inequality is satisfied and I_{Δ_p} is a pseudo-metric.
- If $p = \infty$ then I_{Δ_p} is a metric (the Hausdorff distance).

The Δ_p indicator is not Pareto compliant but its properties can be exploited by a MOEA to guide its search.

Cynthia A. Rodríguez Villalobos and Carlos A. Coello Coello, "**A New Multi-Objective Evolutionary Algorithm Based on a Performance Assessment Indicator**", in *2012 Genetic and Evolutionary Computation Conference (GECCO'2012)*, pp. 505–512, ACM Press, Philadelphia, USA, July 2012, ISBN: 978-1-4503-1177-9.

Maximum Pareto Front Error

It was proposed by Van Veldhuizen [1999] and it measures the largest distance between any vector in our Pareto front approximation and the corresponding closest vector in the true Pareto front.

It is defined as follows:

$$\textbf{MPFE} = \max_j(\min_i|f_1^i(\vec{x}) - f_1^j(\vec{x})|^p + |f_2^i(\vec{x}) - f_2^j(\vec{x})|^p)^{\frac{1}{p}} \quad (10)$$

where $p = 2$, $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$ are index vectors in our Pareto front approximation (PF_{known}) and the true Pareto front (PF_{true}), respectively.

A value of 0 indicates $PF_{known} \subseteq \mathrm{PF}_{true}$. Any other value indicates that at least one vector of PF_{known} is not in PF_{true}.

It is a reference measure that uses the true Pareto front as a reference. It induces a complete ordering and is non-cardinal.

Performance Indicators

Maximum Pareto Front Error

Pareto Compatibility: It is not weakly compatible with any outperformance relation. It violates weak monotony. It is better, according to MPFE, to find one solution close to the true Pareto front than to find ten solutions, nine of which are in the true Pareto front, and one which is some distance away. This does not sit well with typical intuitions about the quality of a nondominated set. It exhibits weak relativity because any subset of the true Pareto front is optimal.

Advantages: It is cheap to compute. It provides information about whether any points found are far from the true Pareto front.

Disadvantages: Even if a nondominated set has a very low value of MPFE, it does not make it a good front, and doesn't necessarily make it better than another one with a much worse value of MPFE. As with other distance measures, different objectives must be combined to get a single figure of merit, bringing in scaling and normalization issues. Also, it is required to know the true Pareto front of the problem.

Performance Indicators

Average Pareto Front Error

This is a variant of MPFE that was also proposed by Van Veldhuizen [1999]. It also attempts to measure the convergence property of a MOEA by using distance to the true Pareto front.

From each solution in our Pareto front approximation, its perpendicular distance to the true Pareto front is determined by approximating the true Pareto front as a combination of piece-wise linear segments with the average of these distances defining the metric value (e.g., Deb [2002] used 500 segments).

It has the same problems as MPFE.

Overall Nondominated Vector Generation

The Overall Nondominated Vector Generation (**ONVG**) measures the total number of nondominated vectors found during MOEA execution [Veldhuizen, 1999].

It is defined as:

$$\textbf{ONVG} \triangleq |\mathrm{PF}_{known}| \tag{11}$$

Measuring the number of distinct nondominated points produced makes of this an independent measure.

Additionally, ONVG induces a complete ordering on the set of approximations, and it is a cardinal measure.

Overall Nondominated Vector Generation

Pareto Compatibility: It is not weakly compatible with any outperformance relation. It does not exhibit either weak monotony or weak relativity.

Advantages: It is easy to compute. It is scaling independent. There are a few pathological cases where this measure can be used to gauge the quality of a nondominated set, for example, if the entire search space contains only nondominated points.

Disadvantages: Its lack of Pareto compatibility. In general, it is straightforward to come up with scenarios in which A outperforms B on this measure, but in which B is clearly 'better' than A. For example, let's assume that A has one million nondominated points and B contains just 1, but this point dominates all of those in A.

Overall Nondominated Vector Generation Ratio

The Overall Nondominated Vector Generation Ratio (**ONVGR**) was also proposed by Van Veldhuizen [1999]. It measures the ratio of the total number of nondominated vectors found PF_{known} during MOEA execution to the number of vectors found in PF_{true}.

It is defined as:

$$\textbf{ONVGR} \triangleq \frac{|\mathrm{PF}_{known}|}{|\mathrm{PF}_{true}|} \tag{12}$$

When $ONVGR = 1$, this states only that the same number of points have been found in both PF_{true} and PF_{known}. It does not infer that $\mathrm{PF}_{true} = \mathrm{PF}_{known}$.

It requires that we know PF_{true}. Schott [1995], used this performance indicator (although defined over the Pareto optimal set, i.e., $| P_{known} |$). Genotypically or phenotypically defining this indicator is probably a matter of preference, but again note multiple solutions may map to an identical vector, or put another way, $| P_{known } | \geq | PF_{known} |$.

Performance Indicators

Overall Nondominated Vector Generation Ratio

Although counting the number of nondominated solutions gives some feeling for how effective the MOEA is in generating desired solutions, it does not reflect on how "far" from the true Pareto front the vectors in our Pareto front approximation are. Additionally, too few vectors and the representation of our Pareto front approximation may be poor; too many vectors may overwhelm the distance measure.

It is difficult to determine what good values for *ONVG* might be. The cardinality of our Pareto front approximation may change at various computational resolutions as well as differing (perhaps radically) between MOPs. Reporting the ratio of the Pareto front approximation cardinality to the discretized true Pareto front gives some feeling for the number of nondominated vectors found versus how many exist to be found.

It is a reference measure that uses the true Pareto front as its reference set. It induces a complete ordering on the set of approximations and it is a cardinal measure.

Performance Indicators

Overall Nondominated Vector Generation Ratio

Pareto Compatibility: It is not weakly compatible with any outperformance relation. It does not exhibit weak monotony or weak relativity.

Advantages: It is easy to compute. It is scaling independent.

Disadvantages: The true (discretized) Pareto front is required. It is not compatible with any outperformance relation. This indicator is useful in the pathological case when the whole search space is the Pareto front. However, other than this pathological case, it is difficult to thing of a possible situation in which this indicator can be useful.

Joshua D. Knowles, "**Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization**", PhD Thesis, The University of Reading, Reading, UK, January 2002.

Performance Indicators

Generational Nondominated Vector Generation (GNVG)

It tracks how many nondominated vectors are produced at each MOEA generation and is defined as [Veldhuizen, 1999]:

$$GNVG \triangleq \mid PF_{current}(t) \mid \ . \tag{13}$$

Nondominated Vector Addition (NVA)

As *globally* nondominated vectors are sought, one hopes to add new nondominated vectors (that may or may not dominate existing vectors) to the Pareto front approximation at each generation. It is then defined as [Veldhuizen, 1999]:

$$NVA \triangleq \mid PF_{known}(t) \mid - \mid PF_{known}(t-1) \mid \ . \tag{14}$$

Performance Indicators

However, **NVA** may be misleading. A single vector added to $PF_{known}(t)$ may dominate and thus remove several others. $PF_{known}(t)$'s size may also remain constant for several successive generations even if $GNVG \neq 0$.

These two indicators (**GNVG** and **NVA**) are not Pareto compatible and do not exhibit either weak relativity or weak monotony.

Schott's Spacing

Schott [1995] proposed this indicator (which he called **efficient set spacing**) that measures the distance variance of neighboring vectors in PF_{known}. It is defined as:

$$\mathbf{SS} \triangleq \sqrt{\frac{1}{|\mathrm{PF}_{known}| - 1} \sum_{i=1}^{|\mathrm{PF}_{known}|} (\bar{\mathbf{d}} - \mathbf{d_i})^2} \tag{15}$$

and

$$\mathbf{d_i} = min_j(|f_1^i(\vec{x}) - f_1^j(\vec{x})| + |f_2^i(\vec{x}) - f_2^j(\vec{x})|) \tag{16}$$

where $d_i = \min_j(|\ f_1^i(\vec{x}) - f_1^j(\vec{x})\ | + |\ f_2^i(\vec{x}) - f_2^j(\vec{x})\ |)$, $i, j = 1, \ldots, n$, $\overline{d}$ is the mean of all d_i, and n is the number of vectors in PF_{known}. When $S = 0$, all members are spaced evenly apart. This indicator is based on Holder's metric of degree one discussed by Horn [1993].

Jason R. Schott, "**Fault Tolerant Design Using Single and Multicriteria Genetic Algorithm Optimization**", Master's thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Massachusetts, May 1995.

Performance Indicators

Schott's Spacing

This indicator does not require to know PF_{true}, although it is normally assumed that a MOEA has already converged prior to applying this indicator. It is an independent measure, it induces a complete ordering and it is cardinal.

Pareto Compatibility: It is not even weakly compatible with O_W. It exhibits neither monotony nor relativily, since the true Pareto front may be non-uniform.

Advantages: Used in conjunction with other indicators (as it is designed to be), it provides information about the distribution of vectors obtained. It has low computational overhead. It can be generalized to more than two objectives by extending the definition of d_i.

Disadvantages: Schott's definition of d_i does not specify the use of normalized distances, which may be problematic. Its incompatibility with the outperformance relations and the fact that it violates both monotony and relativity make it unreliable. It has to be properly adapted for special cases (e.g., disjoint Pareto fronts).

Deb's Spacing

It was introduced in the original paper of the NSGA-II [Deb, 2000] and it's very similar to Schott's spacing. It is defined as:

$$\Delta = \sum_{i=1}^{|PF_{known}|} \frac{|d_i - \bar{d}|}{|PF_{known}|} \tag{17}$$

where d_i is the Euclidean distance between two consecutive vectors in the Pareto front approximation and $\bar{d}$ is the average of these distances.

The purpose of this indicator is to gauge how evenly the points in the approximation set are distributed in the objective space.

Kalyanmoy Deb, Samir Agrawal, Amrit Pratap and T. Meyarivan, "**A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II**", in Marc Schoenauer et al. (Eds), *Proceedings of the Parallel Problem Solving from Nature VI Conference*, pp. 849–858. Springer, 2000.

Performance Indicators

Deb's Spacing

It is an independent measure, it induces a complete ordering on the set of approximations, and it is a non-cardinal measure.

Pareto Compatibility: It is not weakly compatible with any outperformance relation. It does not exhibit weak monotony or weak relativity. It is quite possible that the true Pareto front has a non-uniform distribution of points.

Advantages: If used in conjunction with other indicators (as it is designed to be), it may provide information about the distribution of vectors obtained. It has low computational overhead.

Disadvantages: It is only suitable for two-dimensional objective spaces because it is not clear how "consecutive" would be defined in the case of more than two objectives. It suffers from normalization and scaling issues, as with other indicators that combine objectives. The incompatibility of this indicator with the outperformance relations and the properties of monotony and relativity make it an unreliable means of making judgements about the overall quality of a nondominated set.

Relative Coverage Comparison of Two Sets

Zitzler et al. [2000] proposed a binary indicator termed *relative coverage comparison of two sets*. Consider $A, B \subseteq X$ as two sets of vectors. $\mathcal{C}$ is defined as the mapping of the order pair (A, B) to the interval $[0, 1]$ as follows:

$$\mathcal{C}(A, B) = \frac{|\{b \in B | \exists a \in A : a \preceq b\}}{|B|} \tag{18}$$

The value $\mathcal{C}(A, B) = 1$ means that all decision vectors in B are weakly dominated by A. The opposite $\mathcal{C}(A, B) = 0$, represents the situation when none of the points in B is weakly dominated by A. Note that always both orderings have to be considered, since $\mathcal{C}(A, B)$ is not necessarily equal to $1 - \mathcal{C}(B, A)$.

Eckart Zitzler, Kalyanmoy Deb and Lothar Thiele, "**Comparison of Multiobjective Evolutionary Algorithms: Empirical Results**, *Evolutionary Computation*, **8**(2):173-195, Summer 2000.

Performance Indicators

Relative Coverage Comparison of Two Sets

The purpose of this indicator is to compare two nondominated sets for overall quality. It is a direct comparative approach giving a single figure of merit that not symmetric. It is a cardinal measure. It is difficult to establish whether this indicator induces a complete ordering on the set of approximations because it is not clear how the pair of $\mathcal{C}$ values should be interpreted together.

Pareto Compatibility: Since this indicator gives two values when comparing sets A and B, $\mathcal{C}(A, B)$ and $\mathcal{C}(B, A)$, it is more difficult to analyze whether it is compatible with the outperformance relations. Its compatibility will depend on how we interpret or combine the two outputs of the indicator.

Given two stes A and A' such that $A \subset A'$ and $\mathbf{ND}(A') = A'$ (i.e., A' is a mutually nondominated set), then $\mathcal{C}(A, A') < 1$ and $\mathcal{C}(A', A) = 1$ so the indicator evaluates A' better than A if we take it that in general, a set C is evaluated better than a set D according to this indicator if $\mathcal{C}(C, D) = 1$ and $\mathcal{C}(D, C) < 1$. Accepting this convention, this indicator is compatible with the weak outperformance relation.

Relative Coverage Comparison of Two Sets

Clearly, this indicator can also be used with reference sets. Consider now the sets A and A' as previously defined, and a set R, where R is a reference set, with $\mathbf{ND}(R) = R$ and $A' \subseteq R$. Then, $\mathcal{C}(R, A) = \mathcal{C}(R, A') = 1$ and $\mathcal{C}(A, R) < \mathcal{C}(A', R)$. Now, for compatibility with O_W we wish A to be evaluated worse than A', so we may simply make the convention for general sets C, D, R with $D \subseteq R$, that if $\mathcal{C}(R, C) = \mathcal{C}(R, D) = 1$ and $\mathcal{C}(C, R) < \mathcal{C}(D, R)$, then we say C is evaluated 'worse' than D. Then *all* sets C, D where D O_W C will be correctly evaluated via a reference set R *provided* $D \subseteq R$.

Consider sets A, A' and R again, with $A \subset A'$ as before, but this time $A' \nsubseteq R$. Consider the case where $\mathcal{C}(R, A) = \mathcal{C}(R, A') = 1$. Now, in this case, it is possible that $\mathcal{C}(A, R) = \mathcal{C}(A', R)$ even though A' O_W A. But we cannot have $\mathcal{C}(A, R) > \mathcal{C}(A', R)$ since $A \subset A'$ so it is not possible that A can cover more of R than A'. Thus, for a reference set R and two sets A, A', with A' $O_W A$ such that $\mathcal{C}(R, A) = \mathcal{C}(R, A') = 1$ then $\mathcal{C}(A, R) \leq \mathcal{C}(A', R)$ and so for this case $\mathcal{C}$ is weakly compatible with O_W.

Relative Coverage Comparison of Two Sets

But the $\mathcal{C}$ indicator cannot detect that O' O_W A using a general reference set R, that is if it is not the case that $\mathcal{C}(R, A) = \mathcal{C}(R, A') = 1$. $\mathcal{C}$ is not even weakly compatible with O_W for such general R.

Any pair of $\mathcal{C}$ indicator scores for a pair of sets A and B in which neither $\mathcal{C}(A, B) = 1$ nor $\mathcal{C}(B, A) = 1$, indicates that the two sets are incompatible according to the weak outperformance relation.

Drawing any further conclusions from the output of the $\mathcal{C}$ indicator in this case is inadvisable.

Also, it is worth noticing that the $\mathcal{C}$ indicator does not give an output which is even representative of our intuitions about the relative quality of two sets *unless* the two sets contain very evenly distributed points, and are of very similar cardinality.

Performance Indicators

Relative Coverage Comparison of Two Sets

Advantages: It has low computational overhead compared to the $\mathcal{S}$ indicator (hypervolume). It is compatible with O_S. It is scale and reference point independent. It does not require any knowledge of the efficient set or ranges of the feasible set. For two evenly-distributed sets, of the same cardinality, the $\mathcal{C}$ indicator gives results compatible with intuitive notions of quality, to some extent.

Disadvantages: Its incompatibility with O_W. If two sets are of different cardinality and/or the distribution of the sets are non-uniform, the the $\mathcal{C}$ indicator gives unreliable results. It cannot determine the degree of outperformance if one set completely outperforms the other. The purpose of the $\mathcal{C}$ indicator could be better served by simply using the outperformance relations themselves.

$D1_R$

This indicator was proposed by Czyzak and Jaszkiewicz [1998] and is defined as follows:

$$D1_R(A, A) = \frac{1}{|R|} \sum_{\mathbf{r} \in R} \min_{\mathbf{z} \in A} \{d(\mathbf{r}, \mathbf{z})\} \tag{19}$$

where A is the approximation set, R is a reference set, $d(\mathbf{r}, \mathbf{z}) = \max_k \{\lambda_k (r_k - z_k)\}$ and $\Lambda = [\lambda_1, \lambda_2, \ldots, \lambda_k]$, $\lambda_k = 1/R_k$, $k = 1, \ldots, K$ with R_k being the range of objective k in set R.

The purpose of $D1_R$ is to measure the mean distance, over the points in a reference set, of the nearest point in an approximation set. This is a reference indicator which induces a complete ordering on the set of approximations. It is a non-cardinal measure.

P. Czyzak and A. Jaszkiewicz, "**Pareto simulated annealing–a metaheuristic technique for multiple-objective combinatorial optimization**", *Journal of Multi-Criteria Decision Analysis*, **7**:34-47, 1998.

Performance Indicators

$D1_R$

Pareto Compatibility: $D1_R$ is weakly compatible with O_W. However, it is not compatible even with O_C.

Advantages: It is cheap to compute. Its weak compatibility with the outperformance relations. It *can* differentiate between different levels of complete outperformance but this will depend upon an appropriate choice of reference set.

Disadvantages: This indicator effectively calculates a weighted average where the reference points have equal weight. This means that the score is strongly dependent upon the distribution of points in the reference set.

$R1$ and $R1_R$

These indicators were proposed by Hansen and Jaszkiewicz [1998]. $R1$ is defined as follows:

$$R1(A, B, U, p) = \int_{u \in U} C(A, B, u) p(u) du, \text{ where}$$

$$C(A, B, u) = \begin{cases} 1 & \text{if } u^*(A) > u^*(B) \\ 1/2 & \text{if } u^*(A) = u^*(B) \\ 0 & \text{if } u^*(A) < u^*(B) \end{cases}$$

where A and B are two approximation sets, U is some set of utility functions, $u : \mathbb{R}^K \rightarrow \mathbb{R}$ which maps each point in the objective space into a measure of utility, $p(u)$ is an intensity function expressing the probability density of the utility $u \in U$, and $u^*(A) = \max_{\mathbf{z} \in A} \{u(\mathbf{z})\}$ and similarly for $u^*(B)$.

Michael Pilegaard Hansen and Andrzej Jaszkiewicz, "**Evaluating the quality of approximations to the non-dominated set**", Technical Report IMM-REP-1998-7, Technical University of Denmark, March 1998.

Performance Indicators

$R1$ and $R1_R$

$R1$ is based on calculating the probability that approximation A is better than B over an entire set of utility functions. It is a direct comparative indicator. It does not induce a total ordering on the set of approximations. It is a non-cardinal measure.

$R1_R$ is $R1$ when it is used with a reference set (i.e., as a reference indicator). This indicator does not induce a total ordering on the set of approximations.

Pareto Compatibility: Making the convention that we are maximizing all objectives, a utility function u is strictly compatible with the dominance relation **iff** $\forall \mathbf{z}^1, \mathbf{z}^2, \mathbf{z}^1 > \mathbf{z}^2 \Rightarrow u(\mathbf{z}^1) > u(\mathbf{z}^2)$. The set of all utility functions that are strictly compatible with the dominance relation is U_{SC}.

Let $U(A > B) = \{u \in U | u^*(A) > u^*(B)\}$. If the probability density function $p(u)$ is such that the probability of selecting a utility function $u \in U(A > B)$ is positive whenever $U(A > B) \neq \emptyset$ and $U \subseteq U_{SC}$ then $R1$ is compatible with O_W.

Performance Indicators

$R1$ and $R1_R$

Under the same conditions, $R1_R$ is only weakly compatible with O_W and is not compatible even with O_C.

Advantages: These indicators are scaling independent. They have a lower computational overhead than the $\mathcal{S}$ indicator. They are compatible with the outperformance relations. The $R1_R$ indicator can differentiate between different levels of complete outperformance provided that an appropriate reference set is chosen.

Disadvantages: The $R1$ indicator cannot differentiate between different levels of complete outperformance. It is cycle-inducing. These indicators depend upon being able to define a set of utility functions. In general, this can be achieved without any knowledge of the Pareto front or the search space, however.

$R2$ and $R2_R$

These indicators were also defined by Hansen and Jaszkiewicz [1998]. $R2$ is defined as follows:

$$\begin{aligned} R2(A,B,U,p) &= E(u^*(A)) - E(u^*(B)) \\ &= \int_{u\in U} u^*(A)p(u)d(u) - \int_{u\in U} u^*(B)p(u)du \\ &= \int_{u\in U} (u^*(A) - u^*(B))p(u)du \end{aligned}$$

where A and B are two approximation sets, U is some set of utility functions, $u : \mathbb{R}^K \rightarrow \mathbb{R}$ which maps each point in the objective space into a measure of utility, $p(u)$ is an intensity function expressing the probability density of the utility $u \in U$, and $u^*(A) = \max_{\mathbf{z}\in A}\{u(\mathbf{z})\}$ and similarly for $u^*(B)$.

Where $R1$ just uses the function $C(A,B,u)$ to decide which of two approximations is better on utility function u, without measuring by *how much*, $R2$ takes into account the expected values of the utility. $R2$ calculates the expected *difference* in the utility of an approximation A with another one B. It is a direct comparative indicator.

Performance Indicators

$R2$ and $R2_R$

$R2$ induces a complete ranking in the set of all approximations. It is a non-cardinal measure. $R2_R$ is $R2$ when used as a reference indicator. It also induces a complete ranking in the set of all approximations.

$R2$ is compatible with O_W subject to the same set of conditions on the set of utility functions used as outlined for $R1$. $R2_R$ is also compatible with O_W given this set of conditions.

Advantages: The advantages of $R2$ arise from its compatibility with all of the outperformance relations and the fact that it can differentiate between different levels of complete outperformance.

Disadvantages: The application of $R2$ depends upon the assumption that it is meaningful to add the values of different utility functions from the set U. This simply means that each utility function in U must be appropriately scaled with respect to the others and its relative importance.

$R3$ and $R3_R$

Hansen and Jaszkiewicz [1998] also proposed a similar indicator to $R2$ whereby the *ratio* of the best utility values is calculated instead of the differences. These indicators are called $R3$ and $R3_R$. The latter is similar to the approach used in single objective optimization, where an approximate solution is evaluated by the ratio of its value to that of a fixed bound.

Performance Indicators

Other Performance Indicators

Many other performance indicators exist. For example:

- ϵ indicator
- Distributed Spacing
- Progress Measure
- Attainment Functions
- Size of the Space Covered

ϵ indicator

Given two approximate sets, A and B, the ϵ-indicator measures the smallest amount, ϵ, that must be used to translate the set, A, so that every point in B is covered. This is a weakly Pareto compliant quality indicator.

Let $A, B \subseteq X$. Then, the ϵ-indicator $I_\epsilon(A, B)$ is defined as the minimum $\epsilon \in \mathbb{R}$ such that any solution $b \in B$ is ϵ-dominated by at least one solution $a \in A$:

$$I_\epsilon(A, B) = \min\{\epsilon \in \mathbb{R} | \forall b \in B \exists a \in A : a \succ_\epsilon b\} \tag{20}$$

So, when $I_\epsilon(A, B) < 1$, all solutions in B are dominated by a solution in A. If $I_\epsilon(A, B) = 1$ and $I_\epsilon(B, A) = 1$, then A and B represent the same Pareto front approximation. If $I_\epsilon(A, B) > 1$ and $I_\epsilon(B, A) > 1$, then A and B are incomparable (i.e., they both contain solutions not dominated by the other set).

Distributed Spacing (ι)

Srinivas and Deb [1994], defined a measure expressing how well a MOEA has distributed Pareto optimal solutions over a nondominated region (the Pareto optimal set).

This indicator is defined as:

$$\iota \triangleq (\sum_{i=1}^{q+1}(\frac{n_i - \overline{n}_i}{\sigma_i})^p)^{1/p} , \tag{21}$$

where q is the number of desired optimal points and the $(q+1)$-th subregion is the dominated region, n_i is the actual number of individuals in the ith subregion (niche) of the nondominated region, $\overline{n}_i$ is the expected number of individuals in the ith subregion of the nondominated region, $p = 2$, and σ_i^2 is the variance of individuals serving the ith subregion of the nondominated region.

N. Srinivas and Kalyanmoy Deb, "**Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms**", *Evolutionary Computation*, **2**(3):221-248, Fall 1994.

Performance Indicators

Distributed Spacing (ι)

They show that if the distribution of points is ideal with $\overline{n}_i$ number of points in the ith subregion, the performance measure $\iota = 0$.

Thus, a low performance measure characterizes an algorithm with a good distribution capacity. This indicator may be modified to measure the distribution of vectors within the Pareto front.

Progress Measure (P, RP)

Bäck defined a parameter used in assessing single-objective EA convergence velocity called a **Progress Measure** [Bäck, 1996], which quantifies *relative* rather than *absolute* convergence improvement by:

$$P \triangleq \ln \sqrt{\frac{f_{max}(0)}{f_{max}(T)}}, \tag{22}$$

where $f_{max}(i)$ is the best objective function value in the parent population at generation i.

To account for the (possible) multiple solutions in our Pareto front approximation, this definition is modified as follows:

$$RP \triangleq \ln \sqrt{\frac{G_1}{G_T}}, \tag{23}$$

where G_1 is the generational distance at generation 1, and G_T the distance at generation T.

Performance Indicators

Attainment Functions

A set of nondominated points in objective space define a region that is dominated by them. The boundary of this region is a "surface" called the **attainment surface** [Fonseca, 1996].

By measuring the location and extent of this attainment surface, one can judge how good the approximation to the true Pareto front is, in a way that is consistent with our intuitive notions of quality.

However, its use in generating statistical distributions is reflected in a graphical evaluation of performance based upon empirical data from the true Pareto front.

Carlos M. Fonseca and Peter J. Fleming, "**On the Performance Assessment and Comparison of Stochastic Multiobjective Optimizers**", in Hans-Michael Voigt et al. (Eds), *Parallel Problem Solving from Nature–PPSN IV*, Springer-Verlag, Lecture Notes in Computer Science, pages 584-593, Berlin, Germany, September 1996.

Performance Indicators

Attainment Functions

Fonseca and Fleming [1996] note that when several runs of an optimizer are performed it is possible to overlay the attainment surfaces from each independent run. This overlaying of surfaces gives a far clearer picture of the different runs than overlaying the points found themselves. Importantly, the combination of the surfaces define a sample worst boundary and a sample best boundary that can easily be identified. In fact, the individual surfaces could be erased, leaving only the upper and lower boundary attainment surfaces. This gives a very clear indication of the range of quality of the approximation of an algorithm.

However, although giving the best and worst that an optimizer attains over some sample number of runs is indeed useful, and is certainly far more informative than plotting the vectors from just one run for each optimizer, it would be desirable if one could calculate a "typical" attainment surface, or a representative range in which the attainment surface is expected to lie in some proportion of the runs. Even better would be the ability to make use of the whole distribution of runs to make some statistical inferences about them.

Attainment Functions

Fortunately, these goals can be achieved, to a certain resolution, by sampling the surfaces using lines angled in the direction of increasing value of all objectives.

Fonseca and Fleming [1996] indicate several weaknesses of this approach. One weakness is that the calculated percentile attainment surfaces do not properly represent the probability of attaining the whole surface in the given percentage of runs. Each point on the surface just independently represents the level achieved in the given percentage of runs. This may be a serious weakness if an algorithm were to generate alternatively one extreme or other of the Pareto front in different runs, but never all extremes at once. This may not be distinguishable from an algorithm that on some runs does find the WHOLE extent of the Pareto front and on others, it does not.

It is also worth mentioning that Fonseca and Fleming [1996] do not indicate that this approach could be used with reference surfaces to give quantitative, statistical measures of performance for an individual algorithm. This approach was proposed by Knowles [2002].

Performance Indicators

Size of the Space Covered

Zitzler & Thiele [1999] proposed an indicator called **Size of the Space Covered** (SSC).

This indicator estimates the size of the global dominated set in objective function space. The core idea of this indicator is to compute the area of objective function space covered by the nondominated vectors generated by a MOEA.

Eckart Zitzler and Lothar Thiele, "**Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach**", *IEEE Transactions on Evolutionary Computation*, **3**(4):257-271, November 1999.

Performance Indicators

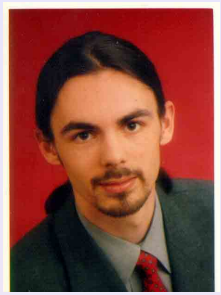
Size of the Space Covered

For problems with two objective functions, each nondominated vector represents a rectangle defined by the points (0,0) and $(f_1(x_i), f_2(x_i))$, where $f_1(x_i)$ and $f_2(x_i)$ are nondominated solutions.

Therefore, SSC is calculated as the union of the areas of all the rectangles that correspond to the nondominated vectors that have been generated.

It is worth noting, however, that this indicator can produce erroneous results when the Pareto front is non-convex.

Performance Indicators



Size of the Space Covered

Laumanns et al. [1999], used the concept of "space covered" to compare problems with more than two objectives.

For this sake, they adopted an m-dimensional cuboid as the reference set, from which a MOEA must cover as much as possible of the dominated space.

Marco Laumanns, Günter Rudolph and Hans-Paul Schwefel, "**Approximating the Pareto Set: Concepts, Diversity Issues and Performance Assessment**", Technical Report CI-72/99, Dortmund: Department of Computer Science/LS11, University of Dortmund, Germany, March 1999. ISSN 1433-3325.

Performance Indicators

Size of the Space Covered

Each nondominated solution provides a cone of dominated solutions. The intersection of this cone with the reference cuboid (which is also a cuboid) is aggregated to the dominated volume. When computing the dominated volume, we avoid counting multiple times the overlapped parts of the different solutions available.

With this method, the reference cuboid is developed using the optimum solutions considering each objective separately. This means that such solutions must be known a priori, or must be relatively easy to generate. It is worth noting, however, that in real-world problems, the cost associated with generating the optima considering each objective separately, may be prohibitively high.

The value of the space covered varies with the number of nondominated solutions and their distribution along the Pareto front. This indicator attempts to combine convergence, spread and number of Pareto optimal elements into a single measure. Therefore, it doesn't work properly when two algorithms differ in more than one of these 3 elements.

Other Indicators

There are many other interesting proposals. For example, Farhang-Menhr [2003] proposed to use entropy as an index that can quantitatively assess the distribution quality of a Pareto set. This index can be used to capture and compare the capability of different population-based MOEAs in generating well-distributed solution sets.

Lizárraga et al. [2008] proposed the G-metric, which takes *m* nondominated sets as argument and assigns a real number to each of them based on its convergence and dispersion-extension. This indicator is compatible with all outperformance relations.

Ali Farhang-Mehr, "**Entropy Approach to Meta-Modeling, Multi-Objective Genetic Algorithm and Quality Assessment of Solution Sets for Design Optimization**", PhD thesis, Department of Mechanical Engineering, University of Maryland, College Park, Maryland, USA, 2003.

Giovanni Lizárraga Lizárraga, Arturo Hernández Aguirre and Salvador Botello Rionda, "**G-Metric: an M-ary Quality Indicator for the Evaluation of Non-dominated Sets**", in *2008 Genetic and Evolutionary Computation Conference (GECCO'2008)*, pp. 665–672, ACM Press, Atlanta, USA, July 2008, ISBN 978-1-60558-131-6.

Performance Indicators

A Word of Caution

Zitzler et al. [2003] provided a careful analysis of performance indicators in order to determine if they are appropriate or not. One of the key results of this study is that the hypervolume is the only unary indicator (from the many unary indicators adopted in the study) that fulfills the property of Pareto compliance.

The most astonishing result from this paper is, however, that they are able to prove that even the combination of unary indicators does not allow to derive solid conclusions about our results (e.g., algorithm A is better than algorithm B). It is worth indicating, however, that this last result has been refuted by other researchers (see [Lizárraga et al., 2008a]).

Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca and Viviane Grunert da Fonseca, "**Performance Assessment of Multiobjective Optimizers: An Analysis and Review**", *IEEE Transactions on Evolutionary Computation*, Vol. 7, No. 2, pp. 117–132, April 2003.

Giovanni Lizárraga, Arturo Hernández and Salvador Botello, "**Some Demonstrations about the Cardinality of Important Sets of Non-dominated Sets**", in Alexander F. Gelbukh and Eduardo F. Morales (editors), *MICAI 2008: Advances in Artificial Intelligence, 7th Mexican International Conference on Artificial Intelligence*, pp. 440–450, Springer. Lecture Notes in Computer Science Vol. 5317, Atizapán de Zaragoza, Mexico, October 27-31, 2008. ISBN 978-3-540-88635-8.

A Nice Tutorial

Another interesting document is a technical report generated from a tutorial delivered at EMO'2005 and updated in 2006, in which the appropriate use of performance measures is properly exemplified. Same as in the paper from Zitzler et al. [2003], this tutorial favors the use of binary indicators.

If we can systematically apply binary performance measures (by pairs of output files, corresponding to the executions of the two MOEAs being compared), it is feasible to use nonparametric analysis afterwards, in order to determine the confidence intervals of the results. Today, it is very common to adopt Wilcoxon's signed-rank test in order to report statistical significance of the results obtained by a MOEA.

Joshua Knowles, Lothar Thiele and Eckart Zitzler, "**A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers**", Technical Report No. 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, February 2006 (revised version).

Performance Indicators

Further Readings

A. Farhang-Mehr and S. Azarm, "**Minimal Sets of Quality Metrics**", in Carlos M. Fonseca et al. (Eds), *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pp. 405–417, Springer. Lecture Notes in Computer Science. Volume 2632, Faro, Portugal, April 2003.

Arnaud Liefooghe and Bilel Derbel, "**A Correlation Analysis of Set Quality Indicator Values in Multiobjective Optimization**", in *2016 Genetic and Evolutionary Computation Conference (GECCO'2016)*, pp. 581–588, ACM Press, Denver, Colorado, USA, 20-24 July, 2016, ISBN 978-1-4503-4206-3.

Giovanni Lizárraga, Marco Jimenez Gomez, Mauricio Garza Castañon, Jorge Acevedo-Davila and Salvador Botello Rionda, "**Why Unary Quality Indicators Are Not Inferior to Binary Quality Indicators**", in Arturo Hernández Aguirre, Raúl Monroy Borja and Carlos Alberto Reyes García (editors), *MICAI 2009: Advances in Artificial Intelligence. 8th Mexican International Conference on Artificial Intelligence*, pp. 646–657, Springer, Lecture Notes in Artificial Intelligence Vol. 5845, Guanajuato, México, November 2009.