The R2 Indicator: a Study of its Expected Improvement in Case of Two Objectives

André Deutz[1,a)], Kaifeng Yang[1,b)] and Michael Emmerich[1,c)]

[1]*Leiden Institute of Advanced Computer Science, Leiden University, Niels Bohrweg 1, 2333 CA, Leiden, The Netherlands*

[a)]Corresponding author: a.h.deutz@liacs.leidenuniv.nl
[b)]k.yang@liacs.leidenuniv.nl
[c)]m.t.m.emmerich@liacs.leidenuniv.nl

Abstract. By a multi-objective optimization problem (MOP) – aka vector optimization problem – we mean the problem of simultaneously optimizing a finite set of real valued functions with a common domain. The object of interest for multiobjective optimization is the so-called Pareto Front (PF).The indicator based approach in solving multi-objective optimization problems has become very popular. Indicators are used, among others, to compare the quality of approximation sets to PFs produced by an algorithm or different algorithms. Among the indicators used the R2 indicator attracted wide spread interest as it is relatively frugal in using computational resources as compared to other indicators. We will study the expected improvement of this indicator given an approximation set to the PF and given a probability density function of a predictive distribution of objective function vectors. The *improvement* of this indicator is defined as follows: the R2 indicator is evaluated on the given approximation set of the PF to which a point in the image of the feasible set is added and the R2 indicator is evaluated on the the given approximation set of the PF, subsequently from the former the latter is subtracted; the resulting difference is the R2-improvement of the chosen point with respect to the given approximation set. The *expected improvement* is the mean of the improvement over the image of the feasible set with respect to the given pdf. For 2 dimensional MOPs we derive a formula for the expected improvement with respect to a probability density function of a predictive distribution of objective function vectors.

Introduction

By a multi-objective optimization problem (MOP) – aka vector optimization problem – we mean the problem of simultaneously optimizing a finite set of real valued functions with a common domain. In case we are dealing with 4 or more functions to be optimized the term many-objective optimization is used. Without loss of generality we can restrict to simultaneous minimization. Very often on the domain of the functions constraints apply, and the subset of the domain satisfying the constraints is called the feasible space. On the vectors of function values a partial order is introduced: a vector is better than another vector if and only if each component of the first vector is smaller or equal to the corresponding components of the second vector and the vectors are not equal to each other – this partial order goes by the name of Pareto order. By means of the functions and the Pareto order one can introduce a binary relation on the domain as follows: an element of the domain is better than another element of the domain if and only if the vector image of the first element is Pareto better than the vector image of the second. The resulting binary relation is a strict partial order (i.e., irreflexive, antisymmetric, and transitive). Starting from the Pareto order to which the diagonal is added we get a pre-order on the domain by using the given functions; the pre-order so obtained in general is not a partial order. The minimal elements in the image of the feasible set of the functions to be minimized constitute the so called Pareto front (PF). The pre-image of the PF is called the Pareto efficient set (ES). In contrast with single objective optimization the PF and ES usually consist of more than one element and very often of infinitely many elements.

The central object of interest for multiobjective optimization is the so-called Pareto Front (PF). Many algorithms have been developed in order to approximate PFs by finite sets. An approximation set to the PF is a finite, non-dominated set consisting of image points of the feasible set – a set is non-dominated with respect to the Pareto order if for each element of the set one cannot find an element of the set which is Pareto better.

The indicator based approach in solving multi-objective optimization problems including many-objective optimization problems has become very popular. Indicators are used to compare the quality of approximation sets produced by an algorithm or different algorithms, in evolutionary algorithms which produce approximation sets to the PF as a selector, to tune parameters of algorithms, and in creating stopping criteria for algorithms. *Moreover, being able to compute the expected improvement of points in the search space with respect to an approximation set A to the PF and an indicator, is extremely useful in minimizing the number of expensive function evaluations.*

Among the indicators used, the R2 indicator attracted wide spread interest, also in many-objective optimization as it is relatively frugal in using computational resources as compared to other indicators such as the hypervolume indicator. For a definition see the work of M.Hansen and A. Jaszkiewicz [1] and D.Brockhoff, T. Wagner and H. Trautmann [2] and the Section on Preliminaries and Statement of the Problem below.

The concept of expected improvement for single-objective optimization was introduced by A. Törn, A. Žilinskas, and J.Mockus (see [3], [4], and [5]), for multiobjective optimization this was done by M. Emmerich, A.Deutz, and J.Klinkenberg (see [6] and [7]). Since then this notion gave rise to an important subfield of indicator based optimization.

We will study the expected improvement of this indicator given an approximation set to the PF and given the probability density function of a predictive distribution of objective function vectors. The *improvement* of this indicator is defined as follows: the R2 indicator is evaluated on the given approximation set of the PF to which a point in the image of the feasible set is added, the R2 indicator is evaluated on the given approximation set of the PF, subsequently from the former the latter is subtracted; the resulting difference is the R2-improvement of the chosen point with respect to the given approximation set. The *expected improvement* is the mean of the improvement over the image of the feasible set with respect to the given pdf.

Preliminaries and Statement of the Problem.

Consider $\mathbf{f} = (f_1, \cdots f_n)^\top : \mathcal{X} \subseteq \mathbb{R}^m \to \mathbb{R}^n$. The multiobjective optimization problem is stated as follows:

$$\text{minimize "simultaneously" } f_i : \mathcal{X} \to \mathbb{R}^m, i = 1, \cdots, n.$$

Without loss of generalization we concentrate on minimizing each of the component functions as maximization can easily be reduced to minimization by considering the opposite of the component functions. A point in $\mathbb{R}^n$ is called *utopian* if and only if it is not dominated by an element of $\mathbf{f}(\mathcal{X})$. Technically the R2 indicator for finite approximation sets $A \subseteq \mathbb{R}^n$ to the Pareto front with respect to some utopian point $\mathbf{z}^* \in \mathbb{R}^n$ and $U = \{u_1, \cdots, u_s\}$ – $s \in \mathbb{N}$ – a finite set of utility functions each provided with a probability $p_i, p_1 + \cdots p_s = 1$ is defined as follows:

$$R2(A, U, \mathbf{z}^*) := \sum_{i=1}^{s} p_i(u_i(\mathbf{z}^*) - \max_{a \in A}\{u_i(a)\}).$$

For a uniform distribution - the one we will use exclusively for the *definition of R2* – this boils down to:

$$R2(A, U, \mathbf{z}^*) := \frac{1}{s} \sum_{i=1}^{s}(u_i(\mathbf{z}^*) - \max_{a \in A}\{u_i(a)\}).$$

This reduces to by keeping the utopian point fixed

$$R2(A, U) = \frac{1}{s} \sum_{i=1}^{s} \min_{a \in A}\{u_i(a)\}, \text{as } \frac{1}{s} \sum_{i=1}^{s} u_i(\mathbf{z}^*) \text{ is a constant.}$$

We can specialize U to a set of utility functions derived from a set of distance functions as follows: for each $u \in U$ there is a distance funtion d on $\mathbb{R}^m$ such that $u(a) = d(\mathbf{z}^*, a), a \in \mathbb{R}^n$. Note also if d is a distance function, then λd, where λ is a positive, real number, is a distance function as well.

From now on we will consider solely utility functions derived from distance functions and more specifically distance functions on $\mathbb{R}^n$ which are weighted Chebyshev distance funtions, i.e., functions of the form $d_{wc}(a, b) := \max_{i \in \{1, \cdots, n\}} \lambda_i |a_i - b_i|$ for some $\lambda_i \in \mathbb{R}^{\ltimes}{}_+, i = 1, \cdots, n$, i.e., λ_i is positive. Let $a, b, \lambda \in \mathbb{R}^n$ such that the components of λ are non-negative. Then the *weighted Chebyshev distance*, notation d_{wc}, is defined as follows.

$$d_{wc}(a, b, \lambda) := \max_{i \in \{1, \cdots, n\}} \lambda_i |a_i - b_i|.$$

In case it is clear which λ is used from the context, we will write $d_{wc}(a, b)$ instead of $d_{wc}(a, b, \lambda)$. Let d be a distance function, $d : W \times W \to \mathbb{R}$ and let $V \subset W$ and $w \in W$. Then $d(w, V) := \min_{v \in V}\{d(w, v)\}$.

Expected Improvement of the R2 Indicator for the Biobjective Case

Consider the situation where the utility function is derived from a weighted Chebyshev distance:

$$u : \mathbb{R}^n \to \mathbb{R} \text{ is such that } a \mapsto d_{wc}(0, a, \lambda),$$

where 0 is the utopian point (which we assume to be the origin) and $a, \lambda \in \mathbb{R}^n$ – for λ each component is non-negative.

We proceed with computing the expected improvement of the R2 indicator (in case of *two* objectives, the definitions we will keep for any dimension of the objective space). Let A be a finite non-dominated subset of $\mathbb{R}^2$ which is an approximation set to the/a Pareto front. Assume that the utopian point $\mathbf{z}^*$ is the origin. We also assume for now that the set of utility functions is a singleton. We will, for this special case, compute the expected improvement of the R2 indicator. The singleton in question contains a utility function derived from a weighted Chebyshev metric. The expected improvement for sets of utility functions which are derived from distance functions can be computed as the sum of expected improvement of the individual utility functions in the finite set of utility functions.

Definition 1 *Individual Contribution. Let A be a finite, non-dominated subset of $\mathbb{R}^n$ (i.e., A is a finite approximation to the Pareto Front) and let $u : \mathbb{R}^n \to \mathbb{R}$ be a utility function and $a \in A$, then the individual contribution – notation $IC(A, a)$ – to the R2 indicator is defined as follows. Then*

$$IC(A, a) = \min_{y \in A}\{u(y)\} - \min_{y \in A \setminus \{a\}}\{u(y)\}$$

We will specialize the utility function in the above definition (Definition 1) to the case where u is derived from a weighted Chebyshev function: $u(a) := d_{wc}(\mathbf{z}^*, a)$ for an appropriate utopian point $\mathbf{z}^*$, we start by assuming it is the origin. The expected improvement of a single weighted Chebyshev 2D distance function can be computed as follows – we assume that the utopian point is the origin and we assume that the weighted Chebyshev distance – with weights λ_1 and λ_2 – of the origin to the approximation set A is m. Then the rectangle specified by $[(0, 0), (\frac{m}{\lambda_1}, \frac{m}{\lambda_2})]$ (which is the weighted Chebyshev ball with radius m and center $(0, 0)$ intersected with the positive quadrant) is the set of points for which the distance to the utopian point is less or equal to the distance of A to the utopian point. In general, $(\frac{m}{\lambda_1}, \frac{m}{\lambda_2})$ is not a member of A, but it is the point which determines the integration region in computing the expected improvement. (The point of A which realizes the minimum distance to the origin lies either on the upper horizontal line or the right vertical line of the aforementioned rectangle.) We also assume that that predictive distribution $\mathrm{pdf}(y_1, y_2) = \mathrm{pdf}(y_1) \cdot \mathrm{pdf}(y_2)$ at a search point can be written as product and moreover we assume that each of the pdfs is a normal distribution.

$$\text{expected improvement at } x \text{ for } A = \int_0^\infty \int_0^\infty (\text{ IC } (A \cup \{(y_1, y_2)^\top\}, (y_1, y_2)^\top)\mathrm{pdf}(y_1, y_2) dy_1 dy_2$$

This reduces to – as we need to integrate over the weighted Chebyshev ball with radius $d_{wc}(\mathbf{z}^*, A)$ and center the origin:

$$\int_0^{\frac{m}{\lambda_2}} \int_0^{\frac{m}{\lambda_1}} (d_{wc}(\mathbf{z}*, (y_1, y_2)^\top) - d_{wc}(\mathbf{z}^*, (m_1, m_2)^\top))\mathrm{pdf}(y_1) \cdot \mathrm{pdf}(y_2) dy_1 dy_2,$$

where m is the minimum distance of $\mathbf{z}^*$ to A, and $\frac{m}{\lambda_1}$ and $\frac{m}{\lambda_2}$ are the integration upper bounds - as discussed before (see also Figure 1).

Unraveling the integration we get an explicit formula for the expected improvement in terms of the error function (erf) as follows. We split up the calculation in two parts, the first part pertains to the lower triangle in Figure 1. Subsequently we integrate over the upper triangle in Figure 1.

$$\int_0^{\frac{m}{\lambda_1}} \left(\int_0^{\frac{\lambda_1 x}{\lambda_2}} \frac{1}{2\pi} \lambda_1 x e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}y^2} dy \right) dx = \tag{1}$$

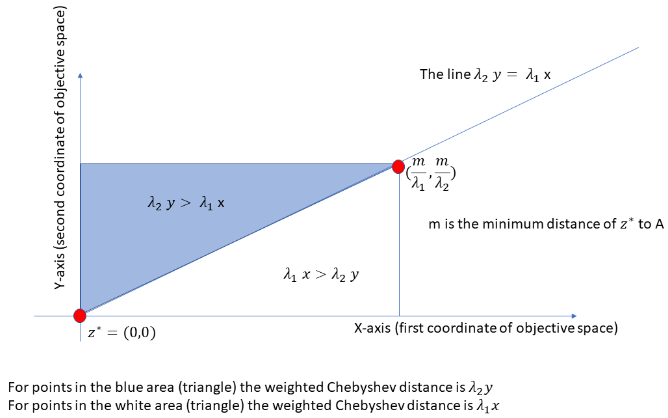

FIGURE 1. The area over which integration is needed for the expected improvement

$$\int_0^{\frac{m}{\lambda_1}} \lambda_1 \frac{1}{\sqrt{2\pi}} x e^{-\frac{1}{2}x^2} \frac{1}{2} \mathrm{erf}(\frac{\lambda_1 x}{\lambda_2 \sqrt{2}}) dx - \int_0^{\frac{m}{\lambda_1}} \lambda_1 \frac{1}{\sqrt{2\pi}} x e^{-\frac{1}{2}x^2} \frac{1}{2} \mathrm{erf}(0) dx \tag{2}$$

In the above equation, Equation 2, the first term can be expressed in terms of the exponential and erf by using integration by parts and the second term can be expressed by using substitution in terms of the exponential and the constant erf(0) as follows.

The first term of Equation 2 is equal to:

$$\frac{1}{\sqrt{2\pi}} \lambda_1 \frac{1}{2} \left[-e^{-(\frac{m}{\lambda_1})^2 \frac{1}{2}} \mathrm{erf}(\frac{m}{\lambda_2 \sqrt{2}}) + \frac{\lambda_1}{\lambda_2} \frac{1}{\sqrt{2c}} \mathrm{erf}(\sqrt{c} \frac{m}{\lambda_1}) + e^0 \mathrm{erf}(0) - \frac{\lambda_1}{\lambda_2} \frac{1}{\sqrt{2c}} \mathrm{erf}(0) \right] \tag{3}$$

The second term of Equation 2 obtained via substitution is as follows:

$$\int_0^{\frac{m}{\lambda_1}} \frac{1}{\sqrt{2\pi}} \lambda_1 \frac{1}{2} \mathrm{erf}(0) x e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \lambda_1 \frac{1}{2} \left[-e^{-\frac{1}{2}x^2} \right]_0^{\frac{m}{\lambda_1}} = \frac{1}{\sqrt{2\pi}} \lambda_1 \frac{1}{2} \left[1 - e^{-(\frac{m}{\lambda_1})^2 \frac{1}{2}} \right] \tag{4}$$

We now compute over the upper triangle in Figure 1.

$$\int_0^{\frac{m}{\lambda_1}} \left(\int_{\frac{\lambda_1 x}{\lambda_2}}^{\frac{m}{\lambda_2}} \frac{1}{2\pi} \lambda_2 y e^{-\frac{1}{2}y^2} e^{-\frac{1}{2}x^2} dy \right) dx = \frac{\lambda_2}{2\pi} e^{-\frac{1}{2}(\frac{m}{\lambda_2})^2} \left(\sqrt{\frac{\pi}{2}} \mathrm{erf}(\frac{m}{\lambda_1 \sqrt{2}}) - \sqrt{\frac{\pi}{2}} \mathrm{erf}(0) \right) - \frac{\lambda_2}{2\pi} \left(\frac{\sqrt{\pi}}{2\sqrt{c}} \mathrm{erf}(\sqrt{c} \frac{m}{\lambda_1}) - \frac{\sqrt{\pi}}{2\sqrt{c}} \mathrm{erf}(0) \right). \tag{5}$$

The sum of the result in the lower triangle and the result in the upper triangle is the expected improvement.

REFERENCES

[1] M. Hansen and A. Jaszkiewicz, "Evaluating the quality of approximations to the non-dominated set," Technical Report IMM-REP-1998-7, Technical University of Denmark 1998.

[2] D. Brockhoff, T. Wagner, and H. Trautmann, "On the Properties of the R2 Indicator," in *Proceedings of the 2012 Genetic and Evolutionary Computation Conference* (ACM, 2012), pp. 465–472.

[3] A. Törn and A. Žilinskas, *Global Optimization* (LNCS 350, Springer Verlag, New York, 1989).

[4] A. Žilinskas and J. Mockus, Avtomatica i Vicheslitel'naya Teknika **4**, 42–44 (1972).

[5] J. Mockus, *Bayesian approach to global optimization: theory and applications*, Vol. 37 (2012).

[6] M. Emmerich, "Single- and multi-objective evolutionary design optimization assisted by gaussian random field metamodels," Ph.D. thesis, University of Dortmund 2005.

[7] M. Emmerich, A. Deutz, and J. Klinkenberg, "Hypervolume-based expected improvement: Monotonicity properties and exact computation," in *IEEE Congress on Evolutionary Computation* (IEEE, 2011), pp. 2147–2154.