

The computation of the expected improvement in dominated hypervolume of Pareto front approximations

Michael Emmerich, André Deutz, Jan-Willem Klinkenberg

Leiden Institute for Advanced Computer Science,
Niels Bohrweg 1, Leiden University, The Netherlands

LIACS-TR 9-2008, September 2008

Abstract

The hypervolume measure is used frequently in the design and performance assessment of multiobjective optimization algorithms. Especially in the context of metamodel (or surrogate) assisted optimization [1, 2] it is interesting to look at the following problem. Given an approximation set for the Pareto front and a new candidate solution x that has not yet been evaluated precisely but for which a prediction with uncertainty measure in the form of an independent multivariate Gaussian distribution with mean vector $\vec{\mu}$ and standard deviation vector $\vec{\sigma}$ exists (Figure 1): What is the expected improvement of the hypervolume when x is being added to the population? For multiple objectives up til now only a Monte Carlo method has been reported. This contribution provides a direct computation procedure for the integral expression. This will be useful to enhance both accuracy and speed of computation for this important measure.

1 Introduction

A point $\vec{y} \in \mathbb{R}^m$ is said to dominate a point $\vec{y}' \in \mathbb{R}^m$, iff $\forall i \in \{1, \dots, m\} : y_i \leq y'_i$ and $\vec{y} \neq \vec{y}'$. In shorthand we write $\vec{y} \prec \vec{y}'$. We write $\vec{y} \preceq \vec{y}'$, iff $\forall i \in \{1, \dots, m\} : y_i \leq y'_i$. For a (finite) set of points $P \subset \mathbb{R}^m$ and a vector $\vec{y} \in \mathbb{R}^m$ define: $P \preceq \vec{y}$, iff $\exists \vec{y}' \in P : \vec{y}' \preceq \vec{y}$.

Let $\text{Vol}(U)$ denote the Lebesgue measure of a set U . Then we can define the S-metric as :

$$S(P, \vec{y}^{ref}) = \text{Vol}(\{\vec{y} \in \mathbb{R}^m | P \preceq \vec{y} \preceq \vec{y}^{ref}\}) \quad (1)$$

Next, let us assume a multiobjective optimization problem

$$f_1(x) \rightarrow \min, \dots, f_m(x) \rightarrow \min, x \in X \quad (2)$$

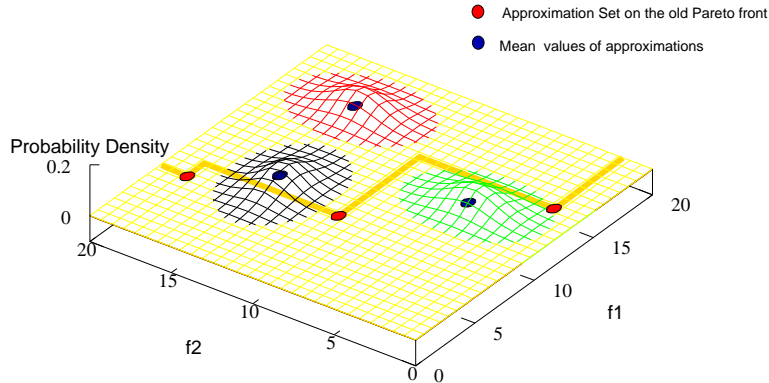


Figure 1: Multivariate distributions of three uncertain predictions and confidence boxes around them. The red line indicates the current approximation to the Pareto front.

for some input space X . We may think of X as \mathbb{R}^n for some n but this is not essential for the following discussion. In addition, let us assume that a prediction tool (e.g. Kriging) provides a prediction for the function values $f_i(x), i = 1, \dots, m$ with an uncertainty measure. The prediction and its uncertainty are defined by a multivariate normal distribution. Whenever we deal with independent models of the m objectives, the multivariate distribution is defined by a vector of mean values $\vec{\mu}$ of the marginal distributions and the vector of their corresponding standard deviations $\vec{\sigma}$.

In the context of efficient global optimization and metamodel-assisted local search it is desirable to know what the expected improvement of a point is. By maximizing the *expected improvement* over a selection of points from X we can detect promising solutions using the prediction tool which subsequently may be evaluated by means of a real experiment or a cost expensive analysis tool.

In this paper we are particularly interested in the expected improvement in hypervolume. The *improvement in hypervolume* is measured relative to a population of solution vectors $P = \{\vec{y}^{(1)}, \dots, \vec{y}^{(k)}\}$ in the objective space \mathbb{R}^m by:

$$I(\vec{y}, P) = S(P \cup \{\vec{y}\}) - S(P). \quad (3)$$

The **expected improvement in hypervolume** is now defined as

$$\text{ExI}(x) = \int_{\vec{y} \in \mathbb{R}^m} I(\vec{y}, P) \cdot \text{PDF}_x(\vec{y}) d\vec{y} \quad (4)$$

In [1] this integral is proposed for the purpose of metamodel-assisted multi-objective optimization. A Monte-Carlo integration method was described for its computation. However, this method has limited accuracy and a direct computation would be desirable. This paper proposes for the first time such a *direct*

method for computing the expected improvement integral as a closed form expression for $m > 1$, thereby generalizing the direct computation of Jones et al. [2] that is only defined for the single-objective case, i.e. $m = 1$.

2 Computation Procedure

As with the probability of non-domination, i.e. $1 - \int_{\vec{y} \in \mathbb{R}^m} \mathbb{T}(P \preceq \vec{y}) d\vec{y}$, described in [1], the integration region \mathbb{R}^m can be partitioned into a set of interval boxes, and then piecewise integration can solve the problem of computing the integral directly. To provide an intuition on the grid-variables and areas introduced in the following, Fig. 3 may serve the reader.

Let $b_i^{(1)}, b_i^{(2)}, \dots, b_i^{(k)}$ denote the sorted list of the all the i -th coordinates of vectors $\vec{y}^{(1)}, \dots, \vec{y}^{(k)}$. For technical reasons we define $b_i^{(0)} = -\infty$ and $b_i^{(k+1)} = y_i^{ref}, b_i^{(k+2)} = \infty$.

The grid-coordinates $b_j^{(i)}$ give rise to a partitioning into grid cells. We can enumerate these grid cells as follows. For each (i_1, \dots, i_m) , where $i_s \in \{0, \dots, k+1\}$, the grid cell named $C(i_1, \dots, i_m)$ is determined by $(b_1^{(i_1)}, \dots, b_m^{(i_m)})^T$ and $(b_1^{(i_1+1)}, \dots, b_m^{(i_m+1)})^T$ as the half open (from below) interval box (i.e., $(b_1^{(i_1)}, b_1^{(i_1+1)}] \times (b_2^{(i_2)}, b_2^{(i_2+1)}] \times \dots \times (b_m^{(i_m)}, b_m^{(i_m+1)}]$). We call $(b_1^{(i_1)}, \dots, b_m^{(i_m)})^T$ the lower bound of $C(i_1, \dots, i_m)$ denoted by $\vec{l}(i_1, \dots, i_m)$, likewise we call $(b_1^{(i_1+1)}, \dots, b_m^{(i_m+1)})^T$ the upper bound of $C(i_1, \dots, i_m)$ denoted by $\vec{u}(i_1, \dots, i_m)$. With this notation we will also denote $C(i_1, \dots, i_m)$ by $(\vec{l}(i_1, \dots, i_m), \vec{u}(i_1, \dots, i_m)]$. See Figure 2 for an example.

It can be directly observed that there are many cells the integration over which adds a contribution of zero to the integral. These are cells that

1. have lower bounds $\vec{l}(i_1, \dots, i_m)$ that are dominated or equal to points in P , i.e. $P \preceq \vec{l}(i_1, \dots, i_m)$, or
2. that have upper bounds $\vec{u}(i_1, \dots, i_m)$ that do not dominate the reference point, i.e. $\vec{u}(i_1, \dots, i_m) \not\prec \vec{y}^{ref}$.

The second criterion is fulfilled by grid cells $C(i_1, \dots, i_m)$ with $i_1 = k+1$ or $i_2 = k+1 \dots$ or $i_m = k+1$, i.e. at least one of their coordinates has index $k+1$.

All cells that fulfill criterion (1) or (2) will be called inactive cells, while the other cells will be termed active cells. *Active cells are cells the inner points of which are dominating the reference point and dominate at least one point in P .*

Obviously, the expected improvement integral is the sum of all contributions of integration of the improvement integral over the set of active cells \mathcal{C}^+ , i.e.

$$\text{ExI}(x) = \sum_{C(i_1, \dots, i_m) \in \mathcal{C}^+} \delta(i_1, \dots, i_m) \quad (5)$$

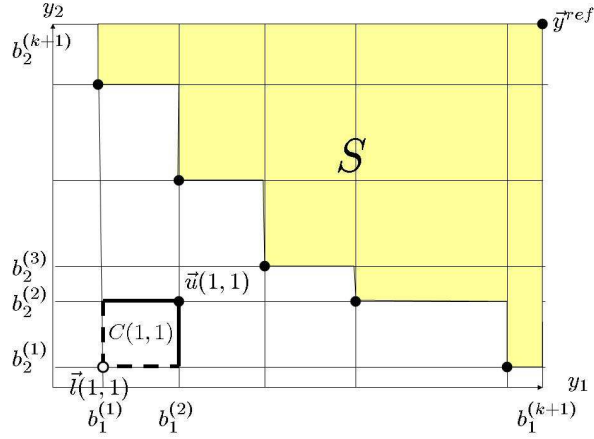


Figure 2: Schematic drawing of a population, its hypervolume, and grid in the bi-objective case. The black points are the points of the population, except the point in the upper right corner that marks the position of the reference point for the hypervolume. The yellow region defines the measured hypervolume S . The grid coordinates are indicated by $b_1^{(i)}$ and $b_2^{(i)}$ for the first and second coordinate, respectively. Grid-cell $C(1,1)$ is highlighted by a thick black boundary.

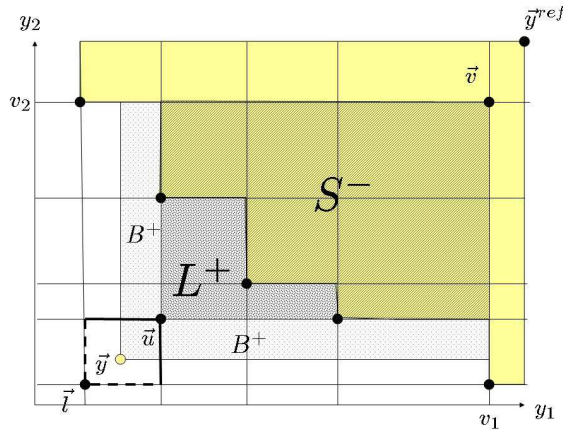


Figure 3: Schematic drawing of the integration area and grid in the bi-objective case. ($S^- = \{z \in \mathbb{R}^m | P(\vec{u}) \preceq z \preceq \vec{u}\}$, where $P(\vec{u}) = \{\vec{p} \in P | \vec{u} \prec \vec{p}\}$)

, where

$$\delta(i_1, \dots, i_m) = \int_{\vec{y} \in (\vec{l}(i_1, \dots, i_m), \vec{u}(i_1, \dots, i_m)]} \mathbf{I}(\vec{y}, P) \cdot \text{PDF}_x(\vec{y}) d\vec{y} \quad (6)$$

. Note, that it is important to choose the half-open interval boxes here, as the lower bounds can reach out to $-\infty$ and we have to avoid overlapping of boundaries, which might cause double integration of non-zero contributions in case of zero standard deviations. Moreover, because of the possibility of zero standard deviations, Lebesgue integration is assumed.

Let us now discuss, how the contribution $\delta(i_1, \dots, i_m)$ can be computed. First the computation will be described. Later on, we will detail the derivation of the expression. Readers, who are only interested in the implementation of the integral can omit the latter part.

The approach is discussed in a top-down way, we will provide a general expression for computing $\delta(i_1, \dots, i_m)$ and then its components are defined.

The contribution $\delta(i_1, \dots, i_m)$ of an active grid cell can be computed via

$$\delta(i_1, \dots, i_m) = \left(\prod_{j=1}^m \delta_j(i_1, \dots, i_m) \right) - \text{Vol}(S^-) \prod_{i=1}^m \left(\Phi\left(\frac{u_i - \mu_i}{\sigma_i}\right) - \Phi\left(\frac{l_i - \mu_i}{\sigma_i}\right) \right). \quad (7)$$

$$\delta_j(i_1, \dots, i_m) = \quad (8)$$

$$\Psi(v_j(i_1, \dots, i_m), u_j(i_1, \dots, i_m), \mu_j, \sigma_j) - \Psi(v_j(i_1, \dots, i_m), l_j(i_1, \dots, i_m), \mu_j, \sigma_j) \quad (9)$$

with $j = 1, \dots, m$. (Ψ is defined in Equation 10. For S^- see Figure 3.)

We define a vector $\vec{v}(i_1, \dots, i_m) \in \mathbb{R}^m$ whose coordinates are defined as follows. The j -th coordinate of $\vec{v}(i_1, \dots, i_m)$ is the j -th coordinate of the intersection point, called $\vec{h}^{(j)}$, of the attainment surface of P and the ray $\{\vec{y} \in \mathbb{R}^m \mid \exists t \geq 0 : \vec{y} = \vec{l}(i_1, \dots, i_m) + t\vec{e}^{(j)}\}$ (where $j \in \{1, \dots, m\}$). We can construct the j -th coordinate of \vec{v} as follows by finding the first vector in the sequence $\vec{l}(i_1, \dots, i_j, \dots, i_m), \vec{l}(i_1, \dots, i_j+1, \dots, i_m), \vec{l}(i_1, \dots, i_j+2, \dots, i_m)$ etc. which does not strictly dominate some point of P . The j -th coordinate of this point is announced as the j -th coordinate of \vec{v} (cf. Fig 3).

The computation of this boundary coordinate can be done by looping only at the discrete steps given by the grid nodes in direction of the unit vector $\vec{e}^{(j)}$. In case $m = 2$ the monotony of the staircase structure of the polyline describing the attainment curve can be exploited to determine this value efficiently .

The integration of the marginal normal distributions is captured in the expression:

$$\Psi(a, b, \mu, \sigma) = \sigma \cdot \phi\left(\frac{b - \mu}{\sigma}\right) + (a - \mu)\Phi\left(\frac{b - \mu}{\sigma}\right) \quad (10)$$

. In this expression ϕ denotes the probability density function of the standard normal distribution, and Φ denotes the cumulative probability density function of the normal distribution, i.e.:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2), \quad \Phi(x) = \frac{1}{2}(1 + \text{erf}(x/\sqrt{2})) \quad (11)$$

Finally, $\text{Vol}(S^-)$ is a correction term. It denotes the hypervolume measure for the subset of points in P dominated or equal to $\vec{u}(i_1, \dots, i_m)$ and reference point $\vec{v}(i_1, \dots, i_m)$ (see also Figure 3). This can be computed by means of a standard procedure for hypervolume computation. For the 2-D case differences of the cumulated contributions of to the hypervolume, when integrating it along f_1 along the staircase, can be exploited for its efficient computation. \square

3 Derivation

The direct computation of the expected improvement contribution $\delta(i_1, \dots, i_m)$ can best be motivated by means of decomposing the general integral expression. For the sake of clarity in the discussion of this section we assume that the grid indices i_1, \dots, i_m and omit them as parameters, i.e. we set $\vec{l} = \vec{l}(i_1, \dots, i_m)$, $\vec{u} = \vec{u}(i_1, \dots, i_m)$, $C = C(i_1, \dots, i_m)$, and so forth.

The improvement $I(\vec{y})$ in the grid cell C always can be decomposed into the contribution of L^+ that is the hypervolume measure of the non-dominated part in $[\vec{u}, \vec{v}]$ and a 'boundary' contribution V that depends on the integration variable \vec{y} and forms a boundary layer around the lower parts of $[\vec{u}, \vec{v}]$. The set V is determined by the set difference

$$V = [\vec{y}, \vec{v}] - [\vec{u}, \vec{v}] \quad (12)$$

(in the following, we assume always the grid indices i_1, \dots, i_m). In summary for points \vec{y} within the cell $C(i_1, \dots, i_m)$ it holds:

$$I(\vec{y}) = \text{Vol}([\vec{y}, \vec{v}] - [\vec{u}, \vec{v}] + L^+) \quad (13)$$

. An important observation is that only the first term, namely $[\vec{y}, \vec{v}]$ depends on \vec{y} .

After these preliminaries, let us recall the expected improvement integral (expression 4). The contribution δ of cell $C(i_1, \dots, i_m)$ is given by:

$$\delta = \int_{l_1}^{u_1} \dots \int_{l_m}^{u_m} I(y_1, \dots, y_m, P) \cdot \text{PDF}_x(y_1, \dots, y_m) dy_1 \dots dy_m \quad (14)$$

, where $\vec{l}(i_1, \dots, i_m) = (l_1, \dots, l_m)$ and $\vec{u}(i_1, \dots, i_m) = (u_1, \dots, u_m)$. In order to compute δ explicitly by reducing, among others, to an S-metric computation of S^- (i.e., $-\text{Vol}(S^-)$), we define three quantities A_1, A_2 , and A_3 as follows.

$$A_1 = \int_{l_1}^{u_1} \dots \int_{l_m}^{u_m} \prod_{i=1}^m (v_i - y_i) \dots \text{PDF}_x(y_1, \dots, y_m) dy_1 \dots dy_m \quad (15)$$

,

$$A_2 = \int_{y_1=l_1}^{u_1} \dots \int_{l_m}^{u_m} \prod_{i=1}^m (v_i - u_i) \cdot \text{PDF}_x(y_1, \dots, y_m) dy_1 \dots dy_m \quad (16)$$

, and

$$A_3 = \int_{l_1}^{u_1} \dots \int_{l_m}^{u_m} \text{Vol}(L^+) \cdot \text{PDF}_x(y_1, \dots, y_m) dy_1 \dots dy_m \quad (17)$$

. Clearly $\delta = A_1 - A_2 + A_3$. The expression $-A_2 + A_3$ is equal to:

$$(\text{Vol}(L^+) - \prod_{i=1}^m (v_i - u_i)) \int_{l_1}^{u_1} \cdots \int_{l_m}^{u_m} \text{PDF}_x(y_1, \dots, y_m) dy_1 \dots dy_m \quad (18)$$

Moreover the expression $\text{Vol}(L^+) - \prod_{i=1}^m (v_i - u_i)$ is the negative hypervolume measure for the set of points in P dominated or equal to \vec{u} with reference point \vec{v} (i.e., $\text{Vol}(L^+) - \prod_{i=1}^m (v_i - u_i) = -\text{Vol}(S^-)$, where $S^- = \{\vec{z} \in \mathbb{R}^m | P(\vec{u}) \preceq \vec{z} \preceq \vec{u}\}$ with $P(\vec{u}) = \{\vec{p} \in P | \vec{u} \prec \vec{p}\}$, see also Figure 3).

In addition, the integral

$$\int_{l_1}^{u_1} \cdots \int_{l_m}^{u_m} \text{PDF}_x(y_1, \dots, y_m) dy_1 \dots dy_m$$

is, in case of independent normal distributions, equal to the term

$$\prod_{i=1}^m \left(\Phi\left(\frac{u_i - \mu_i}{\sigma_i}\right) - \Phi\left(\frac{l_i - \mu_i}{\sigma_i}\right) \right)$$

. An expression for A_1 is more difficult to derive, due to the nonlinearity. First of all we decompose the integral into the product of one-dimensional integrals:

$$\begin{aligned} A_1 &= \prod_{i=1}^m \int_{l_i}^{u_i} (v_i - z) \phi\left(\frac{y_i - \mu_i}{\sigma_i}\right) dz \quad (19) \\ &= \prod_{i=1}^m \left(\int_{-\infty}^{u_i} (v_i - z) \phi\left(\frac{z - \mu_i}{\sigma_i}\right) dz - \int_{-\infty}^{l_i} (v_i - z) \phi\left(\frac{z - \mu_i}{\sigma_i}\right) dz \right) \\ &= \prod_{i=1}^m (\Psi(v_i, u_i, \mu_i, \sigma_i) - \Psi(v_i, l_i, \mu_i, \sigma_i)) \end{aligned}$$

. The last step is motivated by the following equality and the definition in Equation 10 :

$$\int_{-\infty}^b (a - z) \phi\left(\frac{y - \mu}{\sigma}\right) dz = \sigma \phi\left(\frac{b - \mu}{\sigma}\right) + (a - \mu) \Phi\left(\frac{b - \mu}{\sigma}\right). \quad (20)$$

. The computation of Ψ reduces to the integration of a one-dimensional integration the details of which are as follows: Compute $\int_{-\infty}^b (a - y) \text{pdf}(y) dy$, where $\text{pdf}(y) = \frac{2}{\sigma\sqrt{\pi}} \exp(-\frac{1}{2}(\frac{y-\mu}{\sigma})^2)$. Answer: $\int_{-\infty}^b (a - y) \frac{2}{\sigma\sqrt{\pi}} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} dy = \frac{2}{\sigma\sqrt{\pi}} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} + \sqrt{2}(a - \mu) \text{erf}\left(\frac{\sqrt{2}}{2\sigma}(b - \mu)\right) + \sqrt{2}(a - \mu)$. Simplified: $\sigma \phi\left(\frac{b - \mu}{\sigma}\right) + (a - \mu) \Phi\left(\frac{b - \mu}{\sigma}\right)$ with $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ being the probability density function of the standard normal distribution and $\Phi(y) = \frac{1}{2}(1 + \text{erf}(\frac{y}{\sqrt{2}}))$ being the cumulative density function of the standard normal distribution.

Now, all components used in the computation of δ have been derived and we get the computation procedure summarized in the previous section. The

computation is relatively simple to implement. It needs no numerical integration (except, of course, for the computation of the function *erf*). Implementations of the *erf* function can be found, however, in most programming languages and statistical packages.

4 Conclusion

The integral of the expected improvement in hypervolume can be computed directly in case of more than one objectives. The proposed procedure for computation is highly accurate and it can be computed for arbitrary dimensions of the objective space. In the 2-dimensional case the running time is at most quadratic in the number of points in the Pareto front approximation set. An open problem is whether it is possible to find a more *efficient* computation of the expected improvement for higher dimensional objective spaces, for example by merging neighboring interval boxes. However, the given procedure can be considered as a very accurate and often faster computation alternative to the Monte carlo integration method. It is notable, that despite the nonlinearities caused by the normal distribution function, the described procedure does solely use expressions that can be directly computed and no numerical integration is needed (except for the computation of *erf*).

References

- [1] M. Emmerich, Single- and Multiobjective Evolutionary Design Optimization Using Gaussian Random Field Metamodels, PhD Thesis, FB Informatik, University of Dortmund, 2005
- [2] Jones, D., Schonlau, M., Welch, W., (1998) Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization, Vol. 13, 455-492.
- [3] J.W. Klinkenberg, M. Emmerich, A. Deutz: Expected Improvement of the S-Metric for finite Pareto front Approximations, Proc. of MCDM 2008, Auckland, NZ