A Novel Expected Hypervolume Improvement Algorithm For Lipschitz Multi-Objective Optimisation: Almost Shubert's Algorithm In A Special Case

Heleen J. Otten[1,b)] and Sander C. Hille[1,a)]

[1]*Mathematical Institute, Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands*

[a)]Corresponding author: shille@math.leidenuniv.nl
[b)]hotten@strw.leidenuniv.nl

Abstract. An algorithm is proposed for multi-objective optimisation of Lipschitz objective functions that each satisfy a Lipschitz condition of which a Lipschitz constant is *a priori* known. The number of function evaluations is reduced by determining a good next point of evaluation using an Expected Hypervolume Improvement (EHVI) approach. It is closely related to Shubert's Algorithm for single objective optimisation on one-dimensional decision space, but sampling sequences can be slightly different.

INTRODUCTION

Algorithms for optimising Lipschitz continuous objective functions for which Lipschitz constants are known have attracted some attention over the past decades. Shubert [1] introduced the algorithm (named later after him) for global optimisation of a single Lipschitz continuous objective function on one-dimensional decision space. Žilinskas and Žilinskas [2] introduced an approach to computing the Pareto optimal set for a bi-objective optimisation problem with Lipschitz objective functions on a d-dimensional hyper-rectangular decision space. The Pareto optimal set is approximated by that of a natural Lipschitz lower bound that is iteratively improved. See e.g. [2] for further references.

Here we propose an approach for optimisation of n Lipschitz continuous functions on d-dimensional decision space, motivated by the Expected Hypervolume Improvement (EHVI) method introduced in Emmerich [3] and elaborated upon in Emmerich *et al.* [4]. We show that our EHVI method reduces 'almost' to Shubert's Algorithm in the case $n = 1$, $d = 1$. In multi-objective optimisation of a function $f : D \subset \mathbb{R}^d \to \mathbb{R}^n$ the main objectives are to determine the Pareto optimal solutions (simply called the '*Pareto front*') in $\mathbb{R}^n$ and the corresponding set of decisions in D (cf. Miettinen [5]). In case of minimising, this amounts to determining the points in $f(D)$ that are not dominated by any other point in $f(D)$. We say that an element $\mathbf{y} = (y^1, \ldots, y^d)$ in objective space $\mathbb{R}^n$ is *dominated* by $\mathbf{y}'$, written as $\mathbf{y}' \prec \mathbf{y}$, if $(y')^i \leqslant y^i$ for all $i \in \{1, \ldots, n\}$ and $(y')^i < y^i$ for at least one $i \in \{1, \ldots, n\}$. If $n = 1$ the Pareto front is simply the global minimum.

The objective of the proposed EHVI algorithm is to approximate the Pareto front of a Lipschitz continuous f. Recall that this entails the following:

Definition 1. A function $f : D \subset \mathbb{R}^d \to \mathbb{R}^n$, with $f(\mathbf{x}) = (f^1(\mathbf{x}), \ldots, f^n(\mathbf{x}))$ for any $\mathbf{x} \in D$ is called *Lipschitz continuous on D* or is said *to satisfy a Lipschitz condition on D with constant* $\mathbf{L} = (L^1, \ldots, L^n) \in \mathbb{R}^n_+$ if for all $\mathbf{x}, \mathbf{y} \in D$:

$$|f^k(\mathbf{x}) - f^k(\mathbf{y})| \leqslant L^k \|\mathbf{x} - \mathbf{y}\|, \; k = 1, \ldots, n.$$

Here we take $\|\mathbf{x} - \mathbf{y}\| := \sum_{i=1}^d |x_i - y_i|$, the so-called Manhattan metric. (Note that f^k and L^k are not powers of f and L, but indicate the components of the vector f and $\mathbf{L}$).

The objective is to use as few functions evaluations $f(\mathbf{x})$ as possible, because in applications the evaluation $f(\mathbf{x})$ can be computationally quite expensive. The EHVI algorithm exploits the *a priori* knowledge of a Lipschitz constant $\mathbf{L}$ to determine a position $\mathbf{x} \in D$ for the next evaluation, given the previous evaluated points and corresponding computed

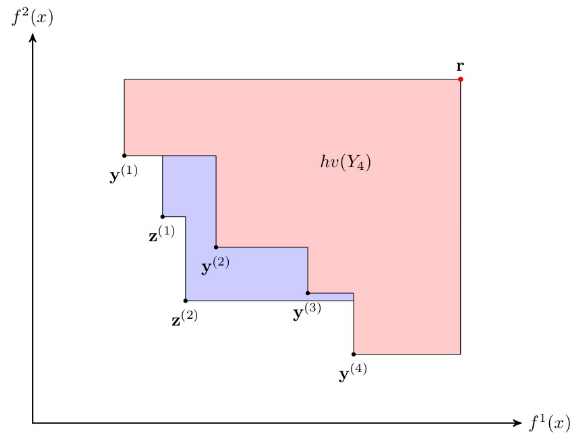

FIGURE 1. *The set of points that are dominated by a set $Y_4 = \{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(4)}\} \subset \mathbb{R}^2$ relative to reference point $\mathbf{r} \in \mathbb{R}^2$ (red) and its hypervolume indicator (the area). The area of the blue region is the hypervolume improvement of $Z = \{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}\}$ relative to Y_4.*

values, that maximises the expected improvement – in a suitable sense – of the approximation of the Pareto front. This 'educated guess' of the new position $\mathbf{x}$ is based on the hypervolume improvement measure, that we discuss next.

Expected hypervolume improvement

Fix a reference point $\mathbf{r} \in \mathbb{R}^n$. For $Y \subset \mathbb{R}^n$, the *set of points dominated by Y* (relative to $\mathbf{r}$) is the set

$$\mathrm{Dom}_{\mathbf{r}}(Y) := \{\mathbf{u} \in \mathbb{R}^n \mid \mathbf{u} \prec \mathbf{r} \text{ and there exists } \mathbf{y} \in Y : \ \mathbf{y} \prec \mathbf{u}\}. \tag{1}$$

Definition 2. The *hypervolume improvement of Z over Y* is the increase of size of the set of dominated points relative to Z compared to that relative to Y, as measured by n-dimensional Lebesgue measure λ_n:

$$\mathrm{HVI}(Z \mid Y) := \lambda_n\big(\mathrm{Dom}_{\mathbf{r}}(Z) \setminus \mathrm{Dom}_{\mathbf{r}}(Y)\big). \tag{2}$$

Figure 1 illustrates the concepts discussed so far. If $Z = \{z\}$, a single point, we shall write $\mathrm{HVI}(z \mid Y)$.

Emmerich *et al.* [4] showed that the *expected* hypervolume improvement is a useful tool for global optimisation. Suppose one has evaluated the Lipschitz objective function f (with constant $\mathbf{L}$) at the points $\mathbf{x} \in X_k := \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}\}$. Let $Y_k := f(X_k)$ and write $\mathbf{y}^{(j)} := f(\mathbf{x}^{(j)})$. Because f is Lipschitz continuous, we know that if we evaluate f in $\mathbf{x} \in \mathbb{R}^d$, the corresponding value $\mathbf{y} := f(\mathbf{x}) \in \mathbb{R}^n$ satisfies for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, k\}$:

$$f^i\big(\mathbf{x}^{(j)}\big) - L^i\|\mathbf{x} - \mathbf{x}^{(j)}\| \leqslant y_i \leqslant f^i\big(\mathbf{x}^{(j)}\big) + L^i\|\mathbf{x} - \mathbf{x}^{(j)}\|. \tag{3}$$

That is, $\mathbf{y}$ has to be in the hyper-rectangle $E_{\mathbf{x}}(X_k)$ that is an n-fold Cartesian product of intervals in $\mathbb{R}$:

$$E_{\mathbf{x}}(X_k) := \bigtimes_{i=1}^{n} \left[\max_j \left\{f^i\left(\mathbf{x}^{(j)}\right) - L^i\|\mathbf{x} - \mathbf{x}^{(j)}\|\right\}, \min_j \left\{f^i\left(\mathbf{x}^{(j)}\right) + L^i\|\mathbf{x} - \mathbf{x}^{(j)}\|\right\}\right]. \tag{4}$$

Since one has no further information on the location of $\mathbf{y}$ within $E_{\mathbf{x}}(X_k)$, we assume that its location is a random variable Y that is homogeneously distributed over $E_{\mathbf{x}}(X_k)$. Write $E_{\mathbf{x}} = E_{\mathbf{x}}(X_k)$ and – motivated by [4] – define

Definition 3. The *expected hypervolume improvement (EHVI)* of a point $\mathbf{x} \in D$ relative to the set X_k of previously evaluated points and corresponding values $Y_k = f(X_k)$ is $\mathrm{EI}(\mathbf{x} \mid X_k) := \mathbb{E}\left[\mathrm{HVI}(Y \mid Y_k)\right]$.

Observe that the hypervolume improvement of Y relative to Y_k will be 0 if $Y \in \mathrm{Dom}_r(Y_k) \cap E_{\mathbf{x}}$. Otherwise it will be $\mathrm{HVI}(Y \mid Y_k)$. Therefore,

$$\mathrm{EI}(\mathbf{x} \mid X_k) = \frac{1}{\mathrm{Vol}(E_{\mathbf{x}})} \int_{E_{\mathbf{x}} \setminus \mathrm{Dom}_r(Y_k)} \mathrm{HVI}(\mathbf{y} \mid Y_k)\, \mathrm{d}\mathbf{y}, \tag{5}$$

where $\mathrm{Vol}(E_{\mathbf{x}})$ is readily obtained from equation (4).

THE EXPECTED HYPERVOLUME IMPROVEMENT ALGORITHM

The proposed EHVI algorithm for approximating the Pareto front consists of the following steps:

1. Select $\mathbf{x}^{(1)} \in D$ and put $X_1 := \{\mathbf{x}^{(1)}\}$.
2. Compute $\mathbf{y}^{(1)} := f(\mathbf{x}^{(1)})$ and put $Y_1 := \{\mathbf{y}^{(1)}\}$.
3. Select $\mathbf{x}^{(k+1)} \in \arg\max_{\mathbf{x}\in D} \mathrm{EI}(\mathbf{x} \mid X_k)$ and put $X_{k+1} := X_k \cup \{\mathbf{x}^{(k+1)}\}$.
4. Compute $\mathbf{y}^{(k+1)} := f(\mathbf{x}^{(k+1)})$ and put $Y_{k+1} := Y_k \cup \{\mathbf{y}^{(k+1)}\}$.
5. Stop if $\mathrm{EI}(\mathbf{x}^{(k+1)} \mid X_k) \leqslant \varepsilon$, otherwise increase k and return to Step 3.

After stopping, the subset of Y_{k+1} consisting of those points that are not dominated by any other point in Y_{k+1} provide an approximation of the part of the Pareto front of f in $\{\mathbf{u} \in \mathbb{R}^n \mid \mathbf{u} \prec \mathbf{r}\}$, to an accuracy that is controlled by $\varepsilon > 0$. This algorithm is interesting to consider – roughly speaking – when computing a global maximum of the functions $D \to \mathbb{R} : \mathbf{x} \mapsto \mathrm{EI}(\mathbf{x} \mid X_k)$ $(k = 1, 2, 3, \ldots)$, required in Step 3, is computationally more efficient than evaluating f.

RELATION TO SHUBERT'S ALGORITHM

Now we will take a closer look at the case for $n = 1$ and $d = 1$, i.e. single objective optimisation in one dimensional decision space. We take $D = [a, b] \subset \mathbb{R}$ and the single objective function $f : [a, b] \to \mathbf{R}$ is assumed to satisfy a Lipschitz condition with constant L. Bruno O. Shubert introduced in 1972 an algorithm to approximate the global maximum of f on $[a, b]$ in [1]. Our main conclusion concerning the relationship to Shubert's Algorithm, which will be made precise below, is:

The sampling sequence of the Expected Hypervolume Improvement Algorithm applied to single objective optimisation $(n = 1)$ *of a Lipschitz continuous objective function on* $[a, b] \subset \mathbb{R}$ $(d = 1)$ *will generally follow that of Shubert's Algorithm, but may deviate at steps, occasionally.*

Shubert's algorithm

We reformulate the algorithm in Shubert [1] for minimisation. Put $\phi := \min_{x\in[a,b]} f(x)$ and $\Phi := \arg\min_{x\in[a,b]} f(x)$. Shubert's Algorithm defines a sampling sequence $x_0, x_1, x_2, \ldots$ of points from $[a, b]$ recursively, by selecting (arbitrarily) $x_0 \in [a, b]$. Once $x_0, \ldots, x_n$ have been selected, x_{n+1} is selected according to

$$F_n(x) := \max_{k=0,\ldots,n} \left(f(x_k) - L|x - x_k|\right), \qquad x_{n+1} \in \arg\min_{x\in[a,b]} F_n(x). \tag{6}$$

It is shown in [1] that the sequence (x_n) converges to a point in Φ and that the minimal values $M_n := \min_{x\in[a,b]} F_n(x)$ converges to ϕ. In practice one usually starts with $x_0 = a$ after which one can take $x_1 = b$. This version of the algorithm one may call the *Canonical Shubert Algorithm (CSA)*. An example is visualised in Figure 2 (left).

Computation of the expected hypervolume improvement

Select a reference point $r \in \mathbb{R}$ sufficiently large, such that $r \geqslant \max_{x\in[a,b]} f(x)$. Suppose that evaluations have been made at points $x_0, \ldots, x_{k-1}$, with $k \geqslant 1$. Put $X_k := \{x_0, \ldots, x_{k-1}\}$ and $Y_k := f(X_k)$. Assume for simplicity of exposition that $a, b \in X_k$. Fix $x \in [a, b]\backslash X_k$ and define x^- as the point in X_k closest to x such that $x^- < x$. Similarly, x^+ is the point closest to x with $x^+ > x$, see Figure 2 (right). Put $y_{\min} := \min(Y_k)$ and define

$$M_x := \min\left\{f(x^-) + L(x - x^-), f(x^+) - L(x - x^+)\right\}, \quad m_x := \max\left\{f(x^-) - L(x - x^-), f(x^+) + L(x - x^+)\right\}. \tag{7}$$

The computation of an expression for $\mathrm{EI}(x \mid X_k)$ and its maximisation are established in the following lemmas.

Lemma 4. $E_x(X_k) \subset \mathbb{R}$ *is determined by the evaluations at* x^- *and* x^+ *only:* $E_x(X_k) = [m_x, M_x]$.

Lemma 5. $\mathrm{HVI}(y \mid Y_k) = y_{\min} - y$ *for* $y \in E_x(X_k)\backslash\mathrm{Dom}_r(Y_k) = \left[\min(m_x, y_{\min}),\ y_{\min}\right]$.

Lemma 6. $\mathrm{EI}(x \mid X_k) = \frac{(y_{\min} - m_x)^2}{2(M_x - m_x)}$ *if* $m_x < y_{\min}$, *and* $\mathrm{EI}(x \mid X_k) = 0$ *otherwise.*

Lemma 7. *Define* $F_{x^-,x^+}(\xi) := \min\{f(x^-) - L(\xi - x^-), f(x^+) + L(\xi - x^+)\}$. *Then* $\arg\max_{x\in[x^-,x^+]} \mathrm{EI}(x \mid X_k) = \{x_L\}$, *where* x_L *is the location of the unique minimum of* F_{x^-,x^+}*:*

$$x_L = \tfrac{1}{2}\left(x^- + x^+ + \tfrac{1}{L}[f(x^-) - f(x^+)]\right). \tag{8}$$

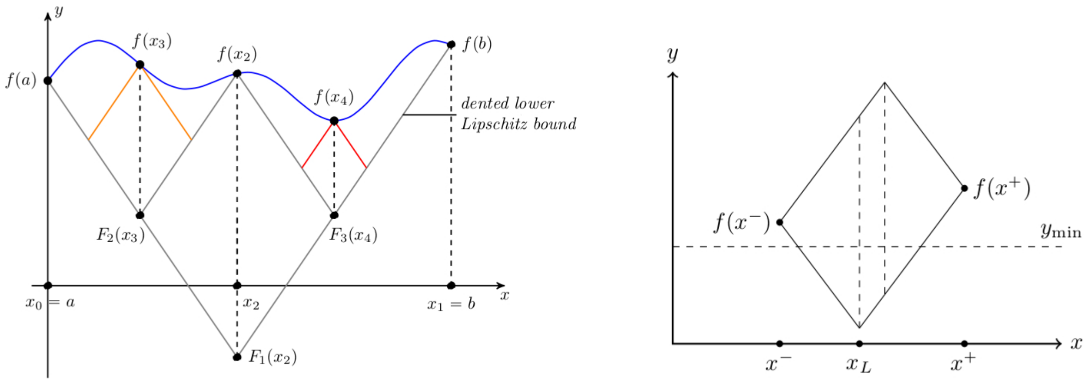

FIGURE 2. *Left: Visualisation of a sampling sequence in the Canonical Shubert's Algorithm, where $x_0 = a$ and $x_1 = b$. Right: The upper and lower bound for the values of $f(x)$ in between two evaluated points x^- and x^+. The set $E_x(X_k)$ of possible values for $f(x)$ is denoted by the vertical dashed lines. x_L is the position of the minimum of the lower bound F_{x^-,x^+}.*

Comparison

Let $x'_0 < x'_1 < \cdots < x'_{k-1}$ be the enumeration of X_k in increasing order and put $y'_i := f(x'_i)$. In Shubert's Algorithm the next point x_k is chosen at a position where $F_{k-1}(x)$ is minimal. F_{k-1} is the minimum of the functions $F_{x'_i,x'_{i+1}}$ defined in Lemma 7, $i \in \{0, 1, \ldots, k-2\}$. Let $x_{L,i}$ be the x_L-location of the interval $[x'_i, x'_{i+1}]$. Put $y_{L,i} := F_{x'_i,x'_{i+1}}(x_{L,i})$. Then $x_k = x_{L,i*}$ for index $i*$ for which $y_{L,i*}$ is minimal. Hence, $z_{i*} := y_{\min} - y_{L,i*}$ is maximal.

In our EHVI algorithm the next point x_k is chosen where $\mathrm{EI}(\mathbf{x} \mid X_k)$ is maximal. According to Lemma 7, x_k is one of the points $x_{L,i}$. A computation shows that $M_{x_{L,i}} - m_{x_{L,i}} = 2[\min(y'_i, y'_{i+1}) - y_{L,i}]$. Thus, Lemma 6 yields

$$E_i := \mathrm{EI}(x_{L,i} | X_k) = \frac{1}{4} \frac{(y_{\min} - y_{L,i})^2}{\min(y'_i, y'_{i+1}) - y_{L,i}} = \frac{1}{4} \frac{z_i^2}{w_i + z_i}, \qquad \text{with } z_i := y_{\min} - y_{L,i},\ w_i := \min(y'_i), y'_{i+1}) - y_{\min}. \quad (9)$$

Then x_k equals $x_{L,i}$ for i for which E_i is maximal. This is *not necessarily* at i with maximal z_i, as in Shubert's Algorithm. Depending on the values w_i, the EHVI algorithm may select a next point x_k different from Shubert's Algorithm. It remains to be investigated how this phenomenon affects convergence rates to global minimum.

ACKNOWLEDGMENTS

This work was part of H. Otten's research project for obtaining her Master's degree in Mathematics at Leiden University under supervision of dr. S.C. Hille (Mathematical Institute) and dr. M.T.M. Emmerich (LIACS).

REFERENCES

[1] B. O. Shubert, SIAM Journal on Numerical Analysis **9**, 379–388 (1972).
[2] A. Žilinskas and J. Žilinskas, Communications in Nonlinear Science and Numerical Simulation **21**, 89 – 98 (2015), numerical Computations: Theory and Algorithms (NUMTA 2013), International Conference and Summer School.
[3] M. Emmerich, "Single- and multi-objective evolutionary design optimization assisted by gaussian random field metamodels," Ph.D. thesis, Fachbereich Informatik, Chair of Systems Analysis, University of Dortmund 2005.
[4] M. Emmerich, K. Yang, A. Deutz, H. Wang, and C. M. Fonseca, "A multicriteria generalization of bayesian global optimization," in *Advances in Stochastic and Deterministic Global Optimization* (Springer International Publishing, 2016), pp. 229–242.
[5] K. Miettinen, *Nonlinear multiobjective optimization* (Springer, 1999).