# Expected hypervolume improvement with constraints

Citation of the final article:
Abdolshah, Majid, Shilton, Alistair, Rana, Santu, Gupta, Sunil and Venkatesh, Svetha 2018, Expected hypervolume improvement with constraints, in *ICPR 2018 : Proceedings of the 24th International Conference on Pattern Recognition*, IEEE, Piscataway, N.J., pp. 3238-3243.

Published in its final form at https://doi.org/10.1109/ICPR.2018.8545387.

# Expected Hypervolume Improvement with Constraints

Majid Abdolshah, Alistair Shilton, Santu Rana, Sunil Gupta, Svetha Venkatesh
Centre for Pattern Recognition and Data Analytics,
Deakin University, Australia 3216
Email: {mabdolsh, alistair.shilton, santu.rana, sunil.gupta, svetha.venkatesh}@deakin.edu.au

Abstract—Bayesian optimisation has become a powerful framework for global optimisation of black-box functions that are expensive to evaluate and possibly noisy. In addition to expensive evaluation of objective functions, many real-world optimisation problems deal with similarly expensive black-box constraints. However, there are few studies regarding the role of constraints in multi-objective Bayesian optimisation. In this paper, we extend the Expected Hypervolume Improvement by introducing expectation of constraints satisfaction and merging them into a new acquisition function called Expected Hypervolume Improvement with Constraints (EHVIC). We analyse the performance of our algorithm by estimating the feasible region dominated by Pareto front using 4 benchmark functions. The proposed method is also evaluated on a real-world problem of Alloy Design. We demonstrate that EHVIC is an effective algorithm that provides a promising performance by comparing to a well-known related method.

## I. Introduction

Bayesian optimisation is a well-known tool for solving a variety of optimisation problems. While traditional numerical methods have proved ineffective for solving some optimisation problems, Bayesian optimisation has proved in variety of optimisation problems dealing with black-box objective functions expensive to evaluate [1]. There have been number of studies on the use of Bayesian optimisation on hyperparameter tuning in machine learning and big data [2], expensive multi-objective optimisation for Robotics [3], and experimentation optimisation in product design such as short polymer fiber materials [4].

Most of the real-world problems consist of multiple, conflicting black-box objectives. In such scenarios it is customary to seek for a set of Pareto optimal outcomes often called the Pareto front [5]. There are a number of methods for solving multi-objective problems with Bayesian approaches. For example, in [6] the authors introduced an algorithm for expensive correlated objectives based on Multi-task Gaussian process. Evolutionary algorithms such as NSGAII [7], SPEA2 [8] and surrogate-assisted evolutionary computation [9] as a mixture of surrogate models are also popular, though these studies are not generally designed to work on limited budget of function evaluations as they require massive amount of function evaluations [10].

In addition to expensive evaluation of objective functions, many optimisation problems deal with similarly expensive black-box constraints. Unknown constraints are part of many black-box multi-objective optimisation problems. For example, when tuning SVM hyperparameters we may want to optimise performance subject to a limit on the number of support vectors (and hence the complexity of evaluating the trained classifier) if the trained machine is to be implemented on limited hardware (such as accessible memory). The goal of optimisation in such cases is to minimize the number of black-box function evaluations to find the global optimum of the function with respect to constraints. Bayesian optimisation with inequality constraints [11] and predictive entropy search for Bayesian optimization with unknown constraints [12] are two major studies investigating the role of inequality black-box expensive constraints in single-objective Bayesian optimisation. There are also two recent studies on multi-objective Bayesian optimisation with constraints. For example, in [10] authors proposed a Bayesian multi-objective optimisation (BMOO) approach to solve the single-objective and multi-objective optimisation with non-linear constraints which is in the same context with the proposed problem in this paper. The method handles the constraints using an extended Pareto domination rule that takes both objectives and constraints into account. The authors evaluated their method on the benchmark test functions with respect to hypervolume improvement. They proposed approach is inspired from [13] which relies on highly complex data models. BMOO uses Sequential Monte-Carlo (SMC) in order to compute the integral over the expected improvement formulation. In [14], the authors proposed a method based on predictive entropy search. The authors generated 100 synthetic optimization problems obtained by sampling the objectives and the constraints from their respective GP prior and they did not use benchmark test functions to evaluate their method. This method is generally categorized as information-based methods while our proposed problem is based on hypervolume improvement approaches.

As it was illustrated in literature review there are few studies regarding the role of constraints in multi-objective Bayesian optimisation. This article provides an extension on the well-known expected improvement acquisition function in order to handle the independent constraints and multi-objective functions. We call our method expected hypervolume improvement with constraints (EHVIC) since it is generally based on hypervolume expected improvement. The contributions of our paper are:

- Formulation of the expected hypervolume improvement with constraints based on the simple but effective expected improvement acquisition function.
- Evaluation of the proposed algorithm based on feasible dominated region on 4 benchmark test functions for the first time. We also estimated the volume of the feasible region of the test functions for more accurate evaluation.
- Discussion of the issues involved in the method in terms of the efficiency and size of the problem.

## II. Preliminaries

In Multi-Objective Optimisation (MOO) potential solutions are assessed by their performance in more than one objective [5]. In MOO, based on definition of Pareto optimality, we wish to return a Pareto front that represents the best trade-off possible considering all criteria [15]. More generally, MOO with constraints (MOOC) includes $M$ objective functions and a set of $K$ constraints. We assume that $f_1, f_2, \ldots, f_M$ and $c_1, c_2, \ldots, c_K$ are drawn independently from Gaussian Process (GP) [16] as $f_i \sim \mathrm{GP}(\mu_i^f, \sigma_i^f)$ and $c_i \sim \mathrm{GP}(\mu_i^c, \sigma_i^c)$. Formally:

$$\begin{aligned} \underset{\mathrm{x}}{\text{minimize}} \quad & \mathrm{y} = \mathrm{Z}(\mathrm{x}) = \{f_1(\mathrm{x}), f_2(\mathrm{x}), \ldots, f_M(\mathrm{x})\} \\ \text{subject to} \quad & \mathrm{C}(\mathrm{x}) = \{c_1(\mathrm{x}), c_2(\mathrm{x}), \ldots, c_K(\mathrm{x})\} \leq 0, \\ \text{where:} \quad & \mathrm{x} \in \mathcal{X} \subseteq \mathbb{R}^d \\ & \mathrm{y} \in \mathcal{Y} \subseteq \mathbb{R}^M \end{aligned}$$

We assume that both the objectives and constraints are expensive to evaluate and that observations of same are noisy. We accumulate the observations of objective functions to $D_{z_{\{1:t\}}} = \{\mathrm{x}_{1:t}, Z(\mathrm{x})_{1:t}\}$ and as for constraint function to $D_{c_{\{1:t\}}} = \{\mathrm{x}_{1:t}, C(\mathrm{x})_{1:t}\}$. In the single-objective case, Expected Improvement with Constraints (EIC) or constrained acquisition function is a popular acquisition function that estimates the expected change in the objective function while satisfying constraints [17]. There are a few studies in single-objective Bayesian optimisation, using EIC or related extensions to handle inequality constraints [11]. There are others frameworks such as information-based search [18], and predictive entropy search for Bayesian optimization with unknown constraints [14]. EHVIC is the analogue of the EIC-based methods in multi-objective case with constraints.

### A. Expected Improvement with Constraints

In the context of single-objective Bayesian optimisation, EIC can handle the inequality constraints via EI acquisition function [11]. The EIC relies on Improvement-based acquisition function which favours points that are likely to improve upon $\mathrm{x}^+$, the feasible point with the lowest function value observed. So in single-objective case, the improvement is defined to be $\mathrm{I}(\mathrm{x}) = max\{0, f(\mathrm{x}^+) - f(\mathrm{x})\}$ and $\alpha_{EI}(\mathrm{x})$ is the expected value of $\mathrm{I}(\mathrm{x})$. Assuming independence, we can separate the constraint satisfaction probabilities, so EIC is defined by:

$$\alpha_{EIC}(\mathrm{x}) = \alpha_{EI}(\mathrm{x}) \prod_{k=1}^{K} Pr(c_k(\mathrm{x}) \leq 0). \tag{1}$$

Assuming $Pr(c_k(\mathrm{x}) \leq 0)$ is a simple univariate Gaussian cumulative distribution [17], EIC has a closed form expression:

$$\begin{aligned} \alpha_{EIC}(\mathrm{x}) = \sigma_f(\mathrm{x})(z_f(\mathrm{x}) \times \Phi(z_f(\mathrm{x})) + \phi(z_f(\mathrm{x})))\times \\ \prod_{k=1}^{K}(1 - \Phi(\frac{\mu_k(\mathrm{x})}{\sigma_k(\mathrm{x})})), \end{aligned} \tag{2}$$

where $z_f(\mathrm{x}) = \frac{f(\mathrm{x}^+) - \mu_f(\mathrm{x})}{\sigma_f(\mathrm{x})}$. It should be noted that EIC is used for single-objective problems. However in multi-objective optimisation problems, $\mathrm{I}(\mathrm{x})$ needs to be considered as hypervolume improvement function [5]. Below we focus on the Expected hypervolume Improvement (EHVI) and then we introduce the proposed EHVIC acquisition function.

### B. Hypervolume and EHVI

In this paper we are particularly interested in the EHVI with a closed form expression introduced in [19]. Our aim is to jointly minimize $m > 1$ objective functions $f_m : \mathcal{X} \to \mathbb{R}$ for $m = 1, ..., M$. A point $\mathrm{y} \in \mathbb{R}^M$, is said to dominate a point $\mathrm{y}' \in \mathbb{R}^M$, iff $\forall i \in \{1, ..., M\} : y_i \leq y_i'$ and $\mathrm{y} \neq \mathrm{y}'$; written as $\mathrm{y} \preceq \mathrm{y}'$. For a set of points $\mathcal{Y} = \{\mathrm{y}_1, \mathrm{y}_2, ..., \mathrm{y}_N\}$, Pareto efficient points $P(\mathcal{Y})$, $P(\mathcal{Y}) \subset \mathcal{Y}$ is defined as $P(\mathcal{Y}) = \{\mathrm{y}_i \in \mathcal{Y} : \mathrm{y}_j \npreceq \mathrm{y}_i, \forall \mathrm{y}_j \in \mathcal{Y} \setminus \{\mathrm{y}_i\}\}$ [6]. A point is Pareto efficient if there is no other point that can improve any one objective without making at least one other objective worse.

One of the most popular indicators for multi-objective optimisation is hypervolume, otherwise known as The S-metric or Lebesgue measure [20]. We denote the Lebesgue measure of a set $U$ by $\mathrm{Vol}(U)$. The S-metric or Lebesgue measure for $P(\mathcal{Y}) = \{\mathrm{y}_1, \mathrm{y}_2, ..., \mathrm{y}_P\}$ is defined to be:

$$S(P(\mathcal{Y}), \mathrm{y}^{ref}) = Vol(\{\mathrm{y} \in \mathbb{R}^M | P(\mathcal{Y}) \preceq \mathrm{y} \preceq \mathrm{y}^{ref}), \tag{3}$$

where $\mathrm{y}^{ref}$ is a reference point in $\mathbb{R}^M$ which is always dominated by all elements of $P(\mathcal{Y})$. By considering equation 3, the improvement in hypervolume is defined to be $I(\mathrm{y}, P(\mathcal{Y})) = S(P(\mathcal{Y}) \cup \{\mathrm{y}\}) - S(P(\mathcal{Y}))$, Therefore the expected hypervolume improvement EHVI(x) is:

$$\mathrm{EHVI}(\mathrm{x}) = \int_{\mathrm{y} \in \mathbb{R}^M} I(\mathrm{y}, P(\mathcal{Y})) \times P_{f_{\mathrm{x}}}(\mathrm{y}) \mathrm{d}(\mathrm{y}), \tag{4}$$

where the $P_{f_{\mathrm{x}}}(\mathrm{y})$ is a probability density function of a predictive distribution of objective function vectors.

The EHVI can be computed using piecewise integration over a set of cells defined by $P(\mathcal{Y})$ as shown in Fig. 1 (see [19] for details). We denote this set of cells $S$. After

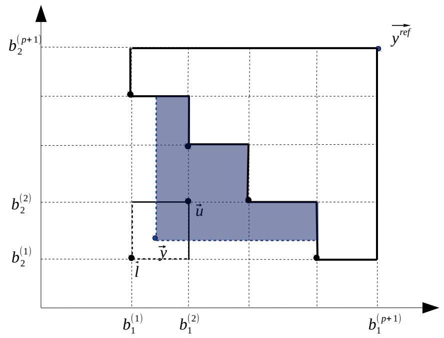



Fig. 1: Sample bi-objective case with partitioned cells. The black points are in set of $P(\mathcal{Y})$ and dominated the $\mathrm{y}^{ref}$. Also the upper (u) bound and lower bound (l) for a sample point y in a cell is illustrated. The $b_1^{(i)}$ and $b_2^{(i)}$, $\{i = 1, ..., p+1\}$ are representative of first and second coordinate.

partitioning the integration space into grid cells, it can be inferred that the expected improvement integral is the sum of improvement integral over the set of non-dominated cells. Non-dominated cells are the one whose points are not dominated by any member of Pareto efficient set [6]: $S^+ = \{s \in S : \forall \mathrm{y} \in s, \mathrm{y}' \in P(\mathcal{Y}), \mathrm{y}' \not\preceq \mathrm{y}\}$. So the expected hypervolume improvement over the cells is defined to be:

$$\mathrm{EHVI}(\mathrm{x}) = \sum_{s \in S^+} \int_{\mathrm{y} \in s} I(\mathrm{y}, P(\mathcal{Y})) P_{f_{\mathrm{x}}}(\mathrm{y}) \mathrm{dy} \tag{5}$$

as $f_1, \ldots, f_M$ are assumed to be generated from independent GPs, the EHVI can be computed analytically [19].

## III. Proposed Method

In EHVIC, the improvement is defined only when all the constraints are satisfied. Since we are not certain about the values of the constraints, we need to extend our model to include this uncertainty. Because $f_{1..M}(\mathrm{x})$ and $c_{1..K}(\mathrm{x})$ are both black-box function and expensive to evaluate, we need to use the Bayesian formalism to model each of $f_{1..M}(\mathrm{x})$ and $c_{1..K}(\mathrm{x})$, given x. We consider $D_{z_{\{1:t\}}}$ and $D_{c_{\{1:t\}}}$ as set of observations for objective functions and constraint functions. During the Bayesian optimisation, by picking a candidate point like x, we can evaluate $Z(\mathrm{x})$ and $C(\mathrm{x})$ and add the point to $D_{z_{\{1:t\}}}$ and $D_{c_{\{1:t\}}}$ respectively. To consider the most general case, we assume that the constraints and the objective functions are independent. So the Gaussian Process posterior is used to independently update $c_i(\mathrm{x}) \sim \mathcal{N}(\mu_i^c(\mathrm{x}), \sigma_i^c(\mathrm{x})); \quad i = 1, ..., K$ for constraint functions and $f_i(\mathrm{x}) \sim \mathcal{N}(\mu_i^f(\mathrm{x}), \sigma_i^f(\mathrm{x})); \quad i = 1, ..., M$ for objective functions.

The proposed acquisition function is made of two main parts, namely the hypervolume improvement with respect to the objective functions as explained in section II-B, and the expectation of constraints satisfaction $\Delta(s)$. We define $\Delta(s)$ for each given cell $s$ as:

$$\begin{aligned}\Delta(s) &= \int_{\mathrm{x}|\mathrm{y} \in s} Pr(c(\mathrm{x}) \leq 0) \mathrm{dx} \\ &= \int_{\mathrm{x}|\mathrm{y} \in s} \int_{-\infty}^{0} p(c(\mathrm{x})|\mathrm{x}, D_{c_{\{1:t\}}}) \mathrm{d}c(\mathrm{x}) \mathrm{d}(\mathrm{x})\end{aligned} \tag{6}$$

Since the expectation of satisfaction for constraints is defined on each cell, the integral bound is on the inputs x whose corresponding objective values satisfies $\mathrm{y} \in s$ that is $\mathrm{x} = \{\mathrm{x} \in \mathcal{X} : \mathrm{y} \in s\}$. Due to marginal Gaussianity of $c(\mathrm{x})$ and the independence of constraints, the value of $\int_{-\infty}^{0} p(c(\mathrm{x})|\mathrm{x}, D_{c_{\{1:t\}}}) \mathrm{d}c(\mathrm{x})$ is equal to product of $K$ Gaussian univariate cumulative distribution functions on constraint functions. The required integral is intractable without making further approximations. To solve this integral, we approximate it using a set of samples from the posterior distribution of the objective functions. The input points of each sample are selected if the posterior sample satisfies the interval criteria $(\mathrm{x}|\mathrm{y} \in s)$. Then, the well-known Monte-Carlo sampling method in [21] is used to estimate the value of $\Delta(s)$. Thus the EHVIC acquisition function results as following equation:

$$\mathrm{EHVIC}(\mathrm{x}) = \sum_{s \in S^+} \left\{ \Delta(s) \times \underbrace{\int_{\mathrm{y} \in s} I(\mathrm{y}, P(\mathcal{Y})) \times P_{f_{\mathrm{x}}}(\mathrm{y}) dy}_{\text{EHVI calculation for given cell s}} \right\} \tag{7}$$

It is noteworthy to mention that, while infeasible points may be selected in our experiment, they are never considered for calculation of Pareto set, though they help to shape the Gaussian processes posteriors. By allowing the Gaussian Process to discern the regions which are more likely to be feasible, the number of evaluations that is required to find the first feasible point, will be reduced. Algorithm 1 illustrates the EHVIC method.

It is worth noting that the stopping criteria in Algorithm 1 is usually related to resources, such as the number of maximum black-box function evaluation, the value of similarity to the best estimated Pareto front, etc.. Moreover, volume of the region dominated by the Pareto front can be used as stopping point and also as an evaluation factor [10].

## IV. Experiments

In this section, we evaluate the performance of EHVIC on a suite of benchmark test functions for multi-objective optimisation with constraints. All test problems are minimization problems. To the best of our knowledge, there is only one study, BMOO [10] based on hypervolume improvement and Bayesian approaches that investigates the role of black-box constraints on multi-objective optimisation. The other studies are based on genetic or evolutionary algorithms such as NSGAII [7], SPEA2 [8] and Surrogate-assisted evolutionary computation [9] which are not designed to work on limited budget of

Algorithm 1 Expected Hypervolume Improvement with Constraints (EHVIC) Algorithm

```
Require:
Z(x) = {f_1(x), f_2(x), ..., f_M(x)}
C(x) = {c_1(x), c_2(x), ..., c_K(x)}
Generate the initial design 𝒳, 𝒴, D_{z{1:t}} and D_{c{1:t}}
Calculate the initial values of Pareto set P_t = P(𝒴)
procedure EHVIC(x)
    while !(Stopping Criteria) do
        Fit the independent GP models based on D_{z{1:t}}
        and D_{c{1:t}}
        Compute Δ(s) and EHVI(x) for ∀s ∈ S^+
        Find the x_{t+1} = argmax_{x∈𝒳} EHVIC(x)
        Evaluate the Z(x_{t+1}) and C(x_{t+1})
        Add {x_{t+1}, Z(x_{t+1})} and {x_{t+1}, C(x_{t+1})}
        to D_{z{1:t}} and D_{c{1:t}}
    end while
    Update P_t = P(𝒴 ∪ y_{t+1})
end procedure
```

function evaluations [10]. However, we still compared the performance of the proposed method with NSGAII to demonstrate the efficiency of our method for expensive functions. The BMOO method introduced by [10] used the volume of the region dominated by the Pareto front relative to a reference point as the evaluation factor expressing the quality and diversity of Pareto front. According to our experiments the volume of the feasible region should be considered to better quantify the quality of the obtained Pareto set. The reason to use the volume of the feasible region is illustrated in Fig. 2 based on BNH function introduced in Table II. The shaded region in Fig. 2 is the dominated area based on the true Pareto set (Black dots). The darker shaded area is the feasible dominated region. Note that the sample Pareto set in Fig. 2 dominated almost 80% of the dominated region. Though the sample Pareto set just dominated 60% of the feasible dominated region. Considering this fact, we believe constraints should not be ignored while evaluating the quality of the obtained Pareto front. So feasible dominated region is a more accurate criterion to be used while handling both objective and constraint functions.

All simulations and obtained results were written in Python. The objectives and constraint functions are modeled by independent GPs. We applied the standard approach to set all GP hyper-parameters with maximum likelihood estimation [16]. The maximum number of function evaluation is set to be 200 for $M = 2$ and the algorithm is initialized with $5d = 10$ ($d = 2$) function evaluations. Note that, if no feasible solution is found in this initial random set then the acquisition function keeps exploring so that it can discern the feasible region.

For hypervolume based comparison, fast method to calculate multi-objective probability of improvement and

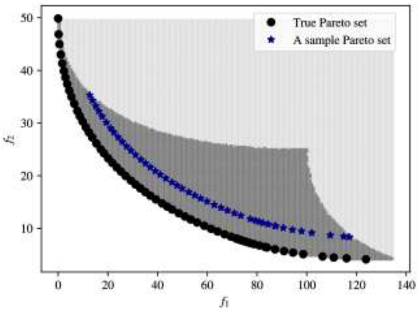



Fig. 2: Comparison of feasible dominated region with the whole dominated region for evaluation of a sample Pareto set. Although the sample Pareto set covered more than 80% of the dominated region, but almost 40% of the feasible dominated region has not been dominated by the sample Pareto set.

expected improvement is implemented [5]. The amount of volume dominated in feasible region by Pareto front for each function is estimated by approximating the true Pareto front and estimating the volume of the feasible region by extracting the feasible cells.

TABLE I: Experimental results on benchmark test functions. The average number of function evaluations over 10 independent runs in order to dominate 90%, 95% and 99% of the target volume ($V$). The corresponding standard deviation for each of the iteration numbers are written in parentheses. We use "≥" sign when the Pareto front approximation is beyond the function evaluation limits.

| Function | Target 90% | Target 95% | Target 99% |
|---|---|---|---|
| BNH | 11.3 (0.82) | 37.5 (1.74) | 64.7 (13.69) |
| SRN | 25.7 (6.70) | 48.31 (9.27) | 82.04 (6.88) |
| TNK | 45.57 (9.21) | 71.40 (5.10) | 105.7 (23.52) |
| OSY | 61.8 (11.35) | 97.1 (13.89) | ≥ 200 (−) |

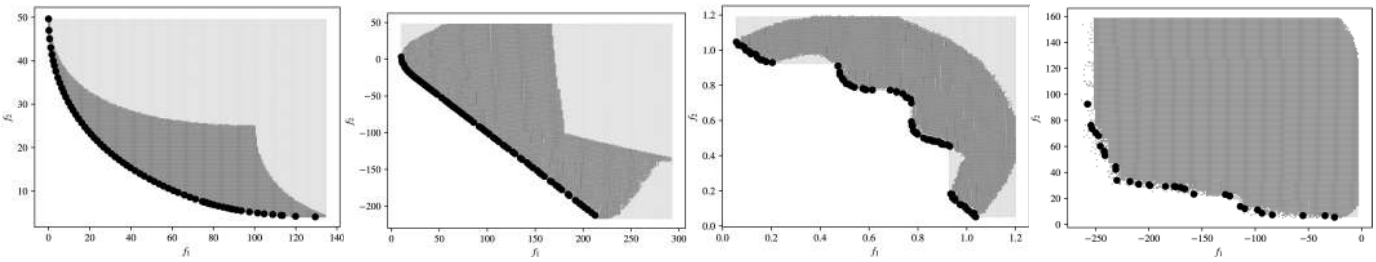
For finding feasible cells, Monte-Carlo method is used with same configuration as explained in section III. All objective functions and constraint functions are scaled down to $[0, 1]$ space while performing EHVIC. The experimental results are obtained by 10 independent runs of the method. We have used four benchmark multi-objective functions with constraints as the test problems. A summary of the selected benchmark functions can be found in Table II.

Fig. 3 shows the obtained set of Pareto front after 128 iteration of the proposed algorithm. The figure shows that EHVIC is able to uniformly maintain solutions in feasible regions for multi-objective functions.

Considering $\Gamma$ as the estimate of the percentage of the feasible volume to that of the whole search space, it is important to note that smaller values of $\Gamma$ result in more computations to shape the feasible region based on independent GPs on black-box constraints. Also in this case, the number of expected particles in the feasible cells such $s$ for calculation of $\Delta(s)$ is typically small. As a

TABLE II: Benchmark functions used for constrained multi-objective optimisation problems. Note that $D$ is the dimension of input space, objective-functions, and constraint functions written as $d_{input}, d_{objectives}, d_{constraints}$. $\Gamma$ is the estimate of the percentage of the feasible volume to that of the whole search space. $V$ is the representative for the volume of the region dominated by Pareto Front. $\mathrm{y}^{ref}$ is defined same as the BMOO method in [10] for volume calculation of the dominated region.

| Function | $D$ | Bounds | Objectives | Constraints | $\Gamma$ | $V$ | $\mathrm{y}^{ref}$ |
|---|---|---|---|---|---|---|---|
| BNH | $2,2,2$ | $x_1 \in [0,5]$ <br> $x_2 \in [0,3]$ | $\begin{cases} f_1(x_1,x_2) = 4x_1^2 + 4x_2^2 \\ f_2(x_1,x_2) = (x_1-5)^2 + (x_2-5)^2 \end{cases}$ | $\begin{cases} c_1(x_1,x_2) = (x_1-5)^2 + x_2^2 \leq 25 \\ c_2(x_1,x_2) = (x_1-8) + (x_2+3)^2 \geq 7.7 \end{cases}$ | 93.6% | 1716 | $[140,50]$ |
| SRN | $2,2,2$ | $x_1 \in [-20,20]$ <br> $x_2 \in [-20,20]$ | $\begin{cases} f_1(x_1,x_2) = (x_1-2)^2 + (x_2-1)^2 + 2 \\ f_2(x_1,x_2) = 9x_1 - (x_2-1)^2 \end{cases}$ | $\begin{cases} c_1(x_1,x_2) = x_1^2 + x_2^2 \leq 225 \\ c_2(x_1,x_2) = x_1 - 3x_2 \leq -10 \end{cases}$ | 16.1% | 23861 | $[200,50]$ |
| TNK | $2,2,2$ | $x_1 \in [0,\pi]$ <br> $x_2 \in [0,\pi]$ | $\begin{cases} f_1(x_1,x_2) = x_1 \\ f_2(x_1,x_2) = x_2 \end{cases}$ | $\begin{cases} c_1(x_1,x_2) = -x_1^2 - x_2^2 + 1 + 0.1cos(16arctan(\frac{x_1}{x_2})) \leq 0 \\ c_2(x_1,x_2) = (x_1-0.5)^2 + (x_2-0.5))^2 \leq 0.5 \end{cases}$ | 5.1% | 0.48 | $[1.2,1.2]$ |
| OSY | $6,2,6$ | $x_1 \in [0,10]$ <br> $x_2 \in [0,10]$ <br> $x_3 \in [1,5]$ <br> $x_4 \in [0,6]$ <br> $x_5 \in [1,5]$ <br> $x_6 \in [0,10]$ | $\begin{cases} f_1(x_1,x_2) = -[25(x_1-2)^2 + (x_2-2)^2 + (x_3-1)^2 + (x_4-4)^4 + (x_5-1)^2] \\ f_2(x_1,x_2) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 \end{cases}$ | $\begin{cases} c_1(x_1,x_2) = x_1 + x_2 - 2 \geq 0 \\ c_2(x_1,x_2) = 6 - x_1 - x_2 \geq 0 \\ c_3(x_1,x_2) = 2 - x_2 + x_1 \geq 0 \\ c_4(x_1,x_2) = 2 - x_1 + 3x_2 \geq 0 \\ c_5(x_1,x_2) = 4 - (x_3-3)^2 - x_4 \geq 0 \\ c_6(x_1,x_2) = (x_5-3)^2 + x_6 - 4 \geq 0 \end{cases}$ | 3.2% | 16079 | $[0,80]$ |



(a) BNH (b) SRN (c) TNK (d) OSY

Fig. 3: Result of EHVIC algorithm on test problems explained in Table II for 128 evaluations of test functions. Black dots represent non-dominated solutions while the area in gray is the feasible region for each test function.

result, the Pareto set on such test functions might not be accurate compared to other forms of test scenarios (see Fig. 3(d) as an example).

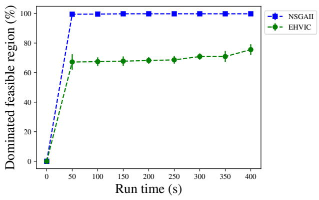
Table I shows the average number of iterations with the corresponding standard deviation required to dominate 90%, 95% and 99% of the target volume ($V$) for each benchmark function. The BNH test function with the highest value of $\Gamma$ requires less number of iterations to dominate $V$ and OSY with the lowest value of $\Gamma$ needs significantly more iterations to dominate 95% of $V$. However $\Gamma$ is not the only determining factor. Another important factor is the ability of the GP to accurately model the function.

Fig. 4 compares the EHVIC with NSGAII when there exists a limited time budget for run time. Also we assume different values of Function Evaluation Time (FET) both for objective functions and constraint functions to simulate expensive evaluation cost of the functions. Fig. 4(a),(b) show that NSGAII outperforms EHVIC in regard to dominated feasible region when the FET $\simeq 0$ $ms$ or FET $= 10$ $ms$. As the value of FET increases, NSGAII fails to evaluate same number of points at a certain time budget and as a result this affects the performance of NSGAII and EHVIC outperforms NSGAII (see Fig. 4(c)). Finally, significant improvement are made by using EHVIC comparing to NSGAII when FET $= 1$ $s$ (see Fig. 4(d)). Although higher values of FET significantly affected the performance of NSGAII, but EHVIC has almost the same performance since it is able to reduce the number of redundant evaluations of the objective functions, and making it more efficient.

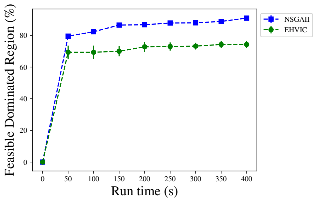
In a practical application we used our method jointly with our metallurgist colleague for alloy optimisation. We seek for the compositions that obtain a set of maximized $\eta$ hardened $Ni$ alloy with as minimum $\sigma$ phase as possible at a temperature of 850°C with two hard constraints on the full dissolution of all structural phases at temperature 1100°C and 1200°C. We have a 9-dimensional optimization problem (9 elements as input) and ThermoCalc is used as a thermodynamic phase simulator for this design. Our result in Fig. 5 represent that we successfully find a set of Pareto frontiers using EHVIC. The results show the effectiveness of our methods for the real world application of alloy design. The obtained results achieved with 64 experiment evaluations. In future we hope to evaluate these alloys by actually casting them. We have released our implementation as an open-source package available on Github.

## V. Conclusion

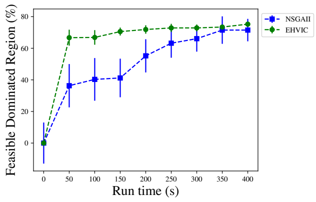
In this paper we have tackled the problem of multi-objective optimisation with constraints by proposing a new but simple acquisition function (EHVIC). EHVIC



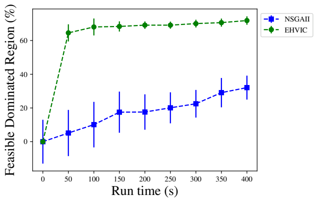

(a) FET $\simeq 0\ ms$




(b) FET $= 10\ ms$




(c) FET $= 100\ ms$




(d) FET $= 1\ s$

Fig. 4: Comparing NSGAII method with EHVIC on the TNK function (see Table II). Optimising the TNK function subject to constraints. (a) FET $\simeq 0\ ms$, (b) FET$= 10\ ms$, (c) FET$= 100\ ms$, and (d) FET$= 1\ s$. Experiments for each algorithm are repeated 10 times, the mean performance and standard deviation are plotted.

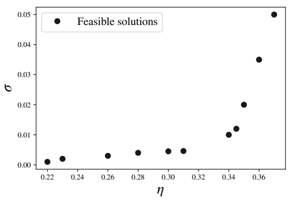



Fig. 5: Set of solutions obtained by EHVIC for maximizing $\eta$ and minimizing $\sigma$ in compositions. Each point represents a feasible solution.

calculates the expected hypervolume improvement and the expectation of constraints satisfaction for all the cells in each iteration and returns the point with the highest value of expected improvement. In contrast to other method proposed for solving this problem, we evaluated our method in regard to the feasible dominated region of the Pareto front. We have also compared the proposed method with NSGAII with different function evaluation time and variety of time budget limitation. Our method proved to outperform the NSGAII when the objective and constraint functions are black-box and expensive to evaluate.

As future work, we plan for eliminating the calculation of the cells in each iteration since it is relatively slow in high dimensions. Also we will focus on modeling the correlation among the objective and constraint functions.

## References


[1] M. A. Gelbart, J. Snoek, and R. P. Adams, "Bayesian optimization with unknown constraints," arXiv preprint arXiv:1403.5607, 2014.

[2] T. T. Joy, S. Rana, S. Gupta, and S. Venkatesh, "Hyperparameter tuning for big data using bayesian optimisation," in Pattern Recognition (ICPR), 2016 23rd International Conference on. IEEE, 2016, pp. 2574–2579.

[3] M. Tesch, J. Schneider, and H. Choset, "Expensive multiobjective optimization for robotics," in Robotics and Automation (ICRA), 2013 IEEE International Conference on. IEEE, 2013, pp. 973–980.

[4] C. Li, D. R. de Celis Leal, S. Rana, S. Gupta, A. Sutti, S. Greenhill, T. Slezak, M. Height, and S. Venkatesh, "Rapid bayesian optimisation for synthesis of short polymer fiber materials," Scientific reports, vol. 7, no. 1, p. 5683, 2017.

[5] I. Couckuyt, D. Deschrijver, and T. Dhaene, "Fast calculation of multiobjective probability of improvement and expected improvement criteria for pareto optimization," Journal of Global Optimization, vol. 60, no. 3, pp. 575–594, 2014.

[6] A. Shah and Z. Ghahramani, "Pareto frontier learning with expensive correlated objectives," in International Conference on Machine Learning, 2016, pp. 1919–1927.

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," IEEE transactions on evolutionary computation, vol. 6, no. 2, pp. 182–197, 2002.

[8] E. Zitzler, M. Laumanns, and L. Thiele, "Spea2: Improving the strength pareto evolutionary algorithm," 2001.

[9] Y. Jin, "Surrogate-assisted evolutionary computation: Recent advances and future challenges," Swarm and Evolutionary Computation, vol. 1, no. 2, pp. 61–70, 2011.

[10] P. Feliot, J. Bect, and E. Vazquez, "A bayesian approach to constrained single-and multi-objective optimization," Journal of Global Optimization, vol. 67, no. 1-2, pp. 97–133, 2017.

[11] J. R. Gardner, M. J. Kusner, Z. E. Xu, K. Q. Weinberger, and J. P. Cunningham, "Bayesian optimization with inequality constraints." in ICML, 2014, pp. 937–945.

[12] E. C. Garrido-Merchán and D. Hernández-Lobato, "Predictive entropy search for multi-objective bayesian optimization with constraints," arXiv preprint arXiv:1609.01051, 2016.

[13] A. Oyama, K. Shimoyama, and K. Fujii, "New constraint-handling method for multi-objective and multi-constraint evolutionary optimization," Transactions of the Japan Society for Aeronautical and Space Sciences, vol. 50, no. 167, pp. 56–62, 2007.

[14] J. M. Hernández-Lobato, M. Gelbart, M. Hoffman, R. Adams, and Z. Ghahramani, "Predictive entropy search for bayesian optimization with unknown constraints," in International Conference on Machine Learning, 2015, pp. 1699–1707.

[15] R. Calandra, J. Peters, and M. Deisenrothy, "Pareto front modeling for sensitivity analysis in multi-objective bayesian optimization," in NIPS Workshop on Bayesian Optimization, vol. 5, 2014.

[16] C. E. Rasmussen and C. K. Williams, Gaussian processes for machine learning. MIT press Cambridge, 2006, vol. 1.

[17] M. A. Gelbart, "Constrained bayesian optimization and applications," Ph.D. dissertation, 2015.

[18] J. M. Hernández-Lobato, M. A. Gelbart, R. P. Adams, M. W. Hoffman, and Z. Ghahramani, "A general framework for constrained bayesian optimization using information-based search," 2016.

[19] M. Emmerich and J.-w. Klinkenberg, "The computation of the expected improvement in dominated hypervolume of pareto front approximations," Rapport technique, Leiden University, vol. 34, 2008.

[20] M. Laumanns, E. Zitzler, and L. Thiele, "A unified model for multi-objective evolutionary algorithms with elitism," in Evolutionary Computation, 2000. Proceedings of the 2000 Congress on, vol. 1. IEEE, 2000, pp. 46–53.

[21] M. T. Emmerich, K. C. Giannakoglou, and B. Naujoks, "Single- and multiobjective evolutionary optimization assisted by gaussian random field metamodels," IEEE Transactions on Evolutionary Computation, vol. 10, no. 4, pp. 421–439, 2006.