

Multi-objective Bayesian optimisation using an exploitative attainment front acquisition function

Finley J. Gibson
Department of Computer Science,
University of Exeter,
Exeter, UK
F.J.Gibson@exeter.ac.uk

Richard M. Everson
Department of Computer Science,
University of Exeter,
Exeter, UK
R.M.Everson@exeter.ac.uk

Jonathan E. Fieldsend
Department of Computer Science,
University of Exeter,
Exeter, UK
J.E.Fieldsend@exeter.ac.uk

Abstract—Efficient methods for optimising expensive black-box problems with multiple objectives can often themselves become prohibitively expensive as the number of objectives is increased. We propose an infill criterion based on the distance to the summary attainment front which does not rely on the expensive hypervolume or expected improvement computations, which are the principal causes of poor dimensional scaling in current state-of-the-art approaches. By evaluating performance on the well-known Walking Fish Group problem set, we show that our method delivers similar performance to the current state-of-the-art. We further show that methods based on surrogate mean predictions are more often than not superior to the widely used expected improvement, suggesting that the additional exploration produced by accounting for the uncertainty in the surrogate’s prediction of the optimisation landscape is often unnecessary and does not aid convergence towards the Pareto front.

Index Terms—Expensive optimisation, Bayesian optimisation, infill criteria, acquisition functions.

I. INTRODUCTION

The process of optimising black-box functions relies exclusively on querying the underlying objective function in order to advance the understanding of the mapping between parameter space and objective space — so as to aid convergence toward an optimal solution. During the optimisation process a balance must be struck between the exploitation of the function as currently understood in order to find good solutions, and further exploration of the unknown regions of its landscape in order to advance this understanding. Extensive work has gone into theorising how this balance should be managed, particularly in cases where there is a high cost associated with querying the underlying objective. Careful management of this balance is considered essential if a suitable solution is to be found as efficiently as possible. Bayesian optimisation [1], which constructs a probabilistic surrogate model of the objective function, has emerged as an effective means of solving such problems, and has been widely applied to many problems with success.

In the field of multi-objective optimisation (MO), accurately modelling the multi-dimensional objective function is more challenging, as many techniques used in single-objective optimisation require (hyper-) volume measurements and integra-

tion, which are complex in higher dimensions. One widely used technique is to calculate and the expected improvement (EI) [1] in the dominated hypervolume to manage the exploitation/exploration trade-off. However, recent work has shown that in single objective problems with high-dimensional parameter spaces, the low fidelity between the commonly used surrogate models and the underlying function reduces the need to actively explore the function. Instead purely exploitative methods have been demonstrated to outperform methods which balance the explore/exploit trade-off [2]. With this in mind we investigate the use of exploitative approaches in the multi-objective domain, anticipating the increased complexity of modelling problems with multi-dimensional outputs will similarly benefit from exploitative methods.

We are concerned with multi-objective problems (MOPs) for which the cost associated with the evaluation of the objectives for a set of parameters (also known as a *design vector*) is significant, which heavily restricts the evaluation budget during an optimisation. This can be because either each objective is independently expensive and requires a separate evaluation or, more commonly, all objectives are jointly evaluated by a single expensive process. In such MOPs the goal of the optimisation process is to find a suitable solution in as few evaluations of the objective function(s) as possible. Large-scale simulation and embodied optimisation, requiring physical intervention for objective evaluation, has emphasised the importance of these expensive problems [3]–[7] as more traditional multi-objective evolutionary algorithms require prohibitively large numbers of evaluations.

In this paper we explore the use of a cheap to compute minimax infill criterion to avoid the expensive dominated hypervolume calculation. We also investigate whether incorporating the surrogate model uncertainty is effective, finding that in fact an exploitative approach is at least as effective and computationally cheaper. The principal contributions of this work are:

- We propose a novel, cheap to compute, infill criterion for evaluating the suitability of candidate parameters at which to next evaluate the true objective function, based on the distance to the summary attainment front (SAF).
- We give an extensive empirical evaluation on the Walking Fish Group problem set [7] and show that the SAF infill cri-

terion yields state-of-the-art performance in Bayesian multi-objective optimisation, equalling or surpassing the current leading methods, with lower computational complexity.

- We show empirically that exploiting the mean prediction of the surrogate model is usually superior to deliberately exploring regions of uncertainty in the surrogate posterior prediction.

In the following section we summarise the Bayesian optimisation approach, particularly in the multi-objective context and discuss works related to this topic. Then in section III we describe the novel SAF infill criterion, and describe the optimisation process used. The experimental evaluation process is described in section III, and findings discussed in section IV. The work concludes with section V.

II. BACKGROUND

A. Bayesian Optimisation

Bayesian optimisation (BO) is a form of efficient global optimisation (EGO), which searches for the global optimal solution to an objective function $f(\mathbf{x})$ for $f: \mathcal{X} \subset \mathbb{R}^d \mapsto \mathbb{R}$, while making as few evaluations of the objective function as possible. In BO, in order to limit the number of evaluations required of the true objective function, a small number of N initial evaluations are made, and a probabilistic surrogate model is then generated from these evaluations. This surrogate model, which should be cheap to evaluate, can then be used to predict the quality of subsequent candidate solutions without the need to frequently evaluate f . Unlike other surrogate modelling methods, BO constructs a probabilistic surrogate model which allows the model's uncertainty as well as its mean prediction to be taken into account when choosing the next design location for evaluation.

The process of BO involves first selecting a set of N initial candidate solutions, usually via Latin hypercube sampling (LHS) [8], and evaluating the objective function for each of these to produce a set $\mathcal{D} = \{(\mathbf{x}_t, f_t \triangleq f(\mathbf{x}_t))\}_{t=1}^N$. A probabilistic, surrogate model is then fitted to these observations, for which Gaussian processes (GPs) are commonly used; see [9] for a comprehensive introduction. The GP describes the current belief about the objective function, modelling it as a set of random variables with a joint Gaussian distribution. The predictive probability distribution $\hat{f}(\mathbf{x})$ at a location \mathbf{x} is a normal distribution:

$$P(\hat{f}(\mathbf{x}) | \mathbf{x}, \mathcal{D}, \theta) = \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \sigma^2(\mathbf{x})\mathbf{I}) \quad (1)$$

where $\boldsymbol{\mu}$ and σ are mean and standard deviation predictions given by:

$$\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\kappa}(\mathbf{x}, \mathbf{X}) - K^{-1}\boldsymbol{\phi}, \quad (2)$$

$$\sigma^2(\mathbf{x}) = \boldsymbol{\kappa}(\mathbf{x}, \mathbf{x}) - \boldsymbol{\kappa}(\mathbf{x}, \mathbf{X})^\top K^{-1}\boldsymbol{\kappa}(\mathbf{X}, \mathbf{x}). \quad (3)$$

Here \mathbf{X} is the d by t matrix of locations at which $f(\mathbf{x})$ has been previously evaluated, $\boldsymbol{\phi} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t)$. $\boldsymbol{\kappa}(\mathbf{x}, \mathbf{x}')$ is a predefined covariance function, or *kernel*, between \mathbf{x} and \mathbf{x}' , and K is the covariance matrix comprised of all covariances $\boldsymbol{\kappa}(\mathbf{x}, \mathbf{x}') \forall \mathbf{x}, \mathbf{x}' \in \mathbf{X}$. A vector of the covariances between \mathbf{x}

and each of the t locations in \mathbf{X} is denoted $\boldsymbol{\kappa}(\mathbf{x}, \mathbf{X}) \in \mathbb{R}^t$. The vector of covariances between each of the t locations in \mathbf{X} and \mathbf{x} is denoted $\boldsymbol{\kappa}(\mathbf{X}, \mathbf{x}) \in \mathbb{R}^t$. The parameters of the covariance function (and any noise model) are denoted by θ : these are learned on receipt of each new $(\mathbf{x}, f(\mathbf{x}))$ pair by maximising the likelihood of the data [9].

Determination of how desirable a new evaluation of the objective function would be at a new location is achieved via an acquisition function $\alpha(\mathbf{x}; \theta)$, which balances the exploitation of locations predicted by the surrogate (with parameters θ) to be good with high confidence, with the exploration of regions that have high uncertainty and might therefore contain the optimum. Maximisation of the acquisition function yields the next location \mathbf{x}' at which to evaluate the real objective function:

$$\mathbf{x}' = \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmax}} \alpha(\mathbf{x}; \theta). \quad (4)$$

A widely used acquisition function is the Expected Improvement [1]. Here the improvement of a value f , over the best solution evaluated so far, $f^* = \min\{f(\mathbf{x}_i)\}_{i=1}^t$ is $I(\mathbf{x}, f^*) = \max(f^* - f, 0)$. The expected improvement at \mathbf{x} on the basis of the model is therefore

$$\alpha(\mathbf{x}, f^*) = \int_{-\infty}^{\infty} I(\mathbf{x}, f^*) p(f | \mathbf{x}, \mathcal{D}) df \quad (5)$$

$$= \sigma(\mathbf{x}) (s\Phi(s) - \phi(s)) \quad (6)$$

where $s = (f^* - \boldsymbol{\mu}(\mathbf{x}))/\sigma(\mathbf{x})$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the Gaussian probability density function and cumulative density functions. This acquisition function is essentially the improvement weighted by the part of the posterior predictive distribution that lies below the evaluated minimum f^* and thus balances the exploitation of solutions which are very likely to be a little better than f^* with the exploration of others which may, with lower probability, turn out to be much better. Other acquisition functions, which achieve the exploration-exploitation balance in different ways include the probability of improvement (PI) [10], optimistic strategies such as UCB [11], expected improvement [1], ϵ -greedy strategies [2], [12], and information-theoretic approaches, e.g. [13], [14]. Under certain conditions EI and UCB have been shown to converge [11], [12].

Since evaluation of the acquisition function entails evaluating the model rather than f itself, it can be cheaply optimised using methods such as an evolutionary algorithm. This process of fitting a surrogate, and optimisation of the acquisition function and evaluation of f is repeated, expanding \mathcal{D} , until a good solution is found and some stopping criteria is satisfied or the computational budget is exhausted.

B. Multi-objective Bayesian Optimisation

In MOPs, rather than a single criterion by which the success of a solution is assessed, a series of conflicting objectives is employed, between which some compromise must be reached. We denote the M conflicting objective functions by $f_m(\mathbf{x})$, $m = 1, \dots, M$, so that the MOP may be expressed as

$$\underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} \mathbf{f}(\mathbf{x}), \quad (7)$$

where $\mathcal{X} \subset \mathbb{R}^d$ is the feasible space and $\mathbf{f} : \mathcal{X} \mapsto \mathbb{R}^M$.

A solution \mathbf{x} is said to dominate another \mathbf{x}' (denoted $\mathbf{x} \prec \mathbf{x}'$) if $f_m(\mathbf{x}) \leq f_m(\mathbf{x}')$ for $m = 1, \dots, M$ and $f_m(\mathbf{x}) < f_m(\mathbf{x}')$ for at least one m . Since in most cases with conflicting objectives there is no single dominating solution, the goal of most multi-objective optimisation algorithms is to produce a number of solutions which well represent the Pareto set, that is the maximal set of solutions that are not dominated by any other solutions in the feasible space:

$$\mathcal{P} = \{\mathbf{x} \in \mathcal{X} : \mathbf{x}' \not\prec \mathbf{x} \forall \mathbf{x}' \in \mathcal{X}\}. \quad (8)$$

The image of \mathcal{P} under \mathbf{f} is the Pareto front, \mathcal{F} . Since \mathcal{P} may contain infinitely many elements, usually in multi-objective optimisation the best we can aspire to is an *approximation* $\tilde{\mathcal{P}}$, as represented by a set of mutually non-dominating solutions. Given a set of solutions \mathcal{D} , we define the function $\text{nondom}(\mathcal{D})$ as the function that returns the non-dominated solutions in \mathcal{D} ; thus $\tilde{\mathcal{P}} = \text{nondom}(\mathcal{D})$.

Methods of Bayesian optimisation for MOPs generally fall into one of two categories: single surrogate approaches and multi-surrogate approaches. The mono-surrogate approach aggregates the M objective functions using a scalarising function $g : \mathbb{R}^M \mapsto \mathbb{R}$. Then using $f(\mathbf{x}) \triangleq g(\mathbf{f}(\mathbf{x}))$ the standard BO algorithm is used to optimise f and thus \mathbf{f} . For example, the hypervolume improvement of a modelled Pareto front over the current $\tilde{\mathcal{P}}$ might be used as an scalarising function.

In a multi-surrogate approach, each objective f_m is modelled individually by its own separate surrogate which models $P(f_m(\mathbf{x} | \mathcal{D}))$. These models are considered to be independent, thus ignoring cross-correlations between the models and the $P(f_m | \mathbf{x}, \mathcal{D})$ are multiplied together to form the joint density $P(\mathbf{f} | \mathbf{x}, \mathcal{D}) = \prod_{m=1}^M P(f_m | \mathbf{x}, \mathcal{D})$. A scalarising function or infill criterion $g : \mathbb{R}^M \mapsto \mathbb{R}$ is used to evaluate the quality of any modelled solution (for notational convenience and w.l.o.g. we assume that smaller g means better solutions). Thus an expected improvement in $g(\cdot)$ may be found as:

$$\alpha_{EI}(\mathbf{x}, g^*) = \int_{-\infty}^{\infty} \max(g^* - g(\mathbf{f}(\mathbf{x})), 0) P(\mathbf{f} | \mathbf{x}, \mathcal{D}) d\mathbf{f} \quad (9)$$

where $g^* = \min\{g(\mathbf{f}(\mathbf{x}_i))\}_{i=1}^t$. Note that this integral is multi-dimensional and, since it does not admit any closed form solution as in the single objective case (5), must be evaluated numerically, usually by Monte Carlo sampling from the distribution $P(\mathbf{f} | \mathcal{D})$.

Practical experience shows that multi-surrogate methods tend to be superior to mono-surrogate approaches because a single GP surrogate is unable to effectively model the highly complex function formed by the aggregated objectives, whereas each of the GPs in a multi-surrogate approach has the simpler task of modelling a single objective function [15].

C. Related work

Bayesian approaches to solving MOPs tend to use Gaussian Process surrogates and largely differ in the acquisition functions by which parameters for the next objective function evaluations are selected. The acquisition functions which have

proven most effective leverage measurement of the expected improvement to the dominated hypervolume, also known as the \mathcal{S} -metric [16], first suggested by Emmerich *et al.* [17]. The \mathcal{S} -metric improvement to the dominated hypervolume is the Heaviside difference between the dominated hypervolume made over the estimated Pareto optimal solutions $\tilde{\mathcal{P}}$ and the dominated hypervolume measurement over these solutions and the GP model posterior prediction. For those locations which are not predicted by the model to improve the dominated hypervolume by some margin ϵ , this is zero, and to these cases a penalty function which roughly depends on the distance from $\tilde{\mathcal{F}}$ is used to overcome the resulting plateau in objective space [16], [18]. The resulting infill criterion for an $M = 2$ objective space is illustrated in Fig. 1b. This approach naturally converges towards a set of solutions which well represent the Pareto front, as maximising the dominated hypervolume is equivalent to finding the true Pareto set [19].

The computation required to calculate the dominated hypervolume, and therefore the \mathcal{S} -metric, scales poorly with the number of objectives and the number of solutions forming $\tilde{\mathcal{P}}$. As observed by Yang *et al.* [20], the expected hypervolume improvement (EHVI) calculation is an NP hard problem in M , but polynomial in $|\tilde{\mathcal{P}}|$ for any fixed value of M . Their recent method for decomposition into hyperboxes for EHVI having a complexity of $\mathcal{O}(2^{M-1} \cdot |\tilde{\mathcal{P}}|^{\lfloor M/2 \rfloor})$. Furthermore increasing M results in larger $\tilde{\mathcal{P}}$ because of the increasing proportion of objective space that is mutually non-dominating. Nevertheless acquisition functions which compute the hypervolume improvement currently comprise the state-of-the art for expensive MOPs. In addition, optimisation using dominated hypervolume depends on the setting of a well-placed reference point for the \mathcal{S} -metric calculation. Positioning of this reference vector requires some knowledge of the scales of the objectives, and poor positioning can bias the optimisation.

Less expensive methods exist and ParEGO [21] is a well regarded alternative without the requirement for hypervolume computation. Throughout the optimisation ParEGO periodically generates a random weight vector which is used to weight the objectives and the aggregated objective is optimised using a single surrogate. Although computationally efficient ParEGO tends to converge to a poorer approximation of the Pareto front than \mathcal{S} -Metric Selection EMO (SMS-EGO) [22] in most cases.

Alternatively, other simple methods which model the desirability of posterior predictions by interpolating the unexplored regions in objective-space have been implemented with success. Keane [23] uses the signed objective-space Euclidean distance to the nearest point in $\tilde{\mathcal{P}}$ to interpolate between observations and then the EI is computed over this to form the acquisition function. However, this fails to preserve the dominance relation [18], and is outperformed by SMS-EGO. Svenson and Santner retain the dominance relation by using the maximin distance between the GP posterior prediction and the approximate Pareto front $\tilde{\mathcal{F}}$ [24]. The EI is then approximated using Monte-Carlo integration. In experiments on some synthetic benchmark problems they achieve similar performance using this infill criteria compared with directly

calculating the dominated hypervolume. Although neither of these interpolation methods require setting of a reference point as in SMS-EGO, they are susceptible to bias depending on the scales of the objectives. Rahat *et al.* [15] compare a range of acquisition functions arguing that, although performance is problem dependent, their proposed inexpensive minimum probability of improvement (MPoI) acquisition function performs comparably to using the hypervolume improvement and similarly to ParEGO. All such inexpensive acquisition functions however are surpassed consistently by the current state of the art, SMS-EGO.

In single-objective Bayesian optimisation, it has recently been suggested by De Ath *et al.* [2] that exploitative optimisation methods are preferable to those which balance exploitation with exploration in problems where the complexity of the problem causes low fidelity between the surrogate models and the objective function. They argue that exploitative methods are sufficiently *fortuitously* exploitative in their nature without the need for active exploration. Their work showed that deploying an exploitative, ‘ ϵ -greedy’ optimisation strategy was often preferable for high-dimensional, complex problems.

III. EXPLOITATIVE SAF

We propose using a signed maximin distance, in the objective space, from the summary attainment front (SAF), similar to that used by Svenson and Santer [24] as the infill criterion for a multi-surrogate assisted EMO optimisation process with GP surrogates as proposed by Emmerich *et al.* [17]. In addition to using this SAF infill criterion with the expected improvement, we explore the use of using only the surrogate posterior mean prediction. This disregards the model uncertainty and therefore explores less, embodying a more exploitative approach as suggested by De Ath *et al.* [2].

A. Infill criterion

Similarly to Svenson and Santer [24] we measure the quality of a solution \mathbf{f}' in relation to the current approximation to the Pareto front $\tilde{\mathcal{P}} = \text{nondom}(\mathcal{D})$ as the maximin distance of \mathbf{f}' to the summary attainment front. The attainment front $\mathcal{A}_{\mathcal{F}}$ is a conservative interpolation of the elements of the non-dominated set \mathcal{F} , so that every element of $\mathcal{A}_{\mathcal{F}}$ is weakly dominated by an element of \mathcal{F} . More formally, the attainment front is the boundary of the region in objective space which is dominated by the elements of \mathcal{F} . If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^M$, we say that \mathbf{u} *properly dominates* \mathbf{v} (denoted $\mathbf{u} \triangleleft \mathbf{v}$) iff $u_m < v_m \forall m = 1, \dots, M$. Then if

$$\mathcal{H} = \{\mathbf{y} \mid \mathbf{u} \prec \mathbf{y} \text{ for some } \mathbf{u} \in \mathcal{F}\} \quad (10)$$

$$\mathcal{U} = \{\mathbf{y} \mid \mathbf{u} \triangleleft \mathbf{y} \text{ for some } \mathbf{u} \in \mathcal{F}\} \quad (11)$$

the attainment front is $\mathcal{A}_{\mathcal{F}} = \mathcal{H} \setminus \mathcal{U} = \partial \mathcal{U}$ [25]. Let $\tilde{\mathcal{F}} = \{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in \tilde{\mathcal{P}}\}$ be the approximate Pareto front corresponding to the evaluations of \mathbf{f} . Then the summary attainment front infill criterion is defined as:

$$\text{SAF}(\mathbf{y}, \tilde{\mathcal{F}}) = \max_{m=1, \dots, M} \min_{\mathbf{y}' \in \tilde{\mathcal{F}}} (y_m - y'_m). \quad (12)$$

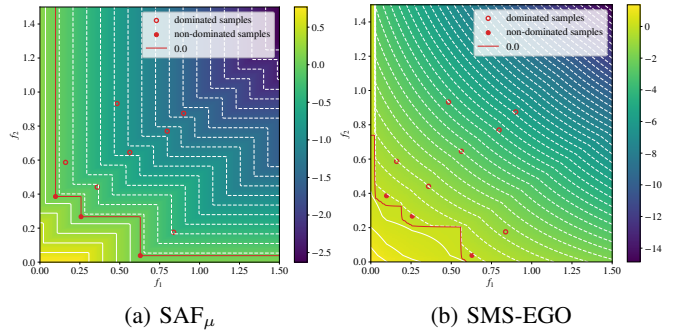


Fig. 1: Visualisation of infill criteria of a two objective problem, $M = 2$. The color scale shows the value of the SAF_{μ} infill criterion (left) and the SMS-EGO infill criterion (right) based on the set of points shown, of which 3 are not dominated and form $\tilde{\mathcal{F}}$. The red curve shows contour where the infill criterion is zero and for SAF_{μ} corresponds to the attainment front $\mathcal{A}_{\tilde{\mathcal{F}}}$. A uniform uncertainty of $\sigma = 0.05$ has been applied to the SMS-EGO predictions, with $\epsilon = [0.058, 0.058]$ and a reference point for the hypervolume computation of $\mathbf{R} = (2, 2)$.

An image of the resultant SAF distance generated by 3 non-dominated locations in a two-dimensional objective space is shown in Fig. 1a. The attainment front $\mathcal{A}_{\tilde{\mathcal{F}}}$ is shown in red and corresponds to the $\text{SAF}(\mathbf{y}, \tilde{\mathcal{F}}) = 0$ contour. The SAF infill criterion is positive behind $\tilde{\mathcal{F}}$ (i.e. for locations dominated by $\tilde{\mathcal{F}}$) and negative in front of $\tilde{\mathcal{F}}$. The figure also shows the infill criterion corresponding to SMS-EGO, which depends on the dominated hypervolume. As can be seen, the two share broadly similar characteristics, although the infill criterion derived from the hypervolume is smoother. We note however, that the hypervolume is much more expensive to calculate and scales poorly with the number of objectives.

We compare two multi-objective Bayesian optimisation algorithms based on the SAF infill criterion. In the first (denoted by SAF_{EI}), we use the SAF infill criterion in conjunction with the expected improvement acquisition function. The expected improvement was calculated using 3000M samples from the posterior distribution $p(\mathbf{f}(\mathbf{x}) \mid \mathcal{D})$ for each proposed \mathbf{x} . In single-objective Bayesian optimisation it has recently been suggested that purely exploitative methods (or ones that devote very little computational resource to deliberate exploration) are preferable over those which deliberately explore more such as the EI [2]. The reason advanced for this is that exploitative methods are sufficiently *fortuitously* exploitative in their nature without the need for active exploration. Their work showed that deploying an exploitative, ϵ -greedy optimisation strategy was often preferable in problems where the parameter space was high-dimensional and thus the objective function complex. For this reason, we also use the SAF infill criterion, but only use the mean posterior prediction $\boldsymbol{\mu}(\mathbf{x}) = (\mu_1(\mathbf{x}), \mu_2(\mathbf{x}), \dots, \mu_M(\mathbf{x}))^{\top}$ where the $\mu_m(\mathbf{x})$ are

Algorithm 1: Multi-objective Bayesian optimisation with SAF

Result: Approximation to the Pareto set for $\mathbf{f}(\mathbf{x})$

Input : N - number of initial evaluations

Input : B - evaluation budget

```

1  $X \leftarrow \text{LHS}(\mathcal{X}, N)$            Generate initial samples
2 for  $t = 1 \rightarrow N$  do
3   |  $\mathbf{f}_t \leftarrow \mathbf{f}(\mathbf{x}_t)$        Expensively evaluate initial samples
4 end
5  $\mathcal{D} \leftarrow \{(\mathbf{x}_t, \mathbf{f}_t)\}_{t=1}^N$ 
6 for  $t = N + 1 \rightarrow B$  do
7   | for  $j = 1 \rightarrow M$  do
8     |  $\theta_j \leftarrow \text{train } \mathcal{GP}(\mathcal{D})$  Train GP for each objective
9   | end
10  |  $\tilde{\mathcal{F}} \leftarrow \text{nondom}(\mathcal{D})$        Approximate Pareto front
11  | if exploitative then
12    |  $\mathbf{x}_t \leftarrow \underset{\mathbf{x} \in \mathcal{X}}{\text{argmax}} \text{SAF}(\boldsymbol{\mu}(\mathbf{x}), \tilde{\mathcal{F}}, \theta_{1:M})$     $\text{SAF}_\mu$ 
13  | else
14    |  $\mathbf{x}_t \leftarrow \underset{\mathbf{x} \in \mathcal{X}}{\text{argmax}} \text{EI}(\mathbf{x}, \tilde{\mathcal{F}}, \theta_{1:M})$         $\text{SAF}_{EI}$ 
15  | end
16  |  $\mathbf{f}_t \leftarrow \mathbf{f}(\mathbf{x}_t)$            Expensively evaluate  $\mathbf{x}_t$ 
17  |  $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\mathbf{x}_t, \mathbf{f}_t)\}$    Augment data
18 end
19  $\tilde{\mathcal{P}} \leftarrow \text{nondom}(\mathcal{D})$ 
output:  $\tilde{\mathcal{P}}$ 

```

given by (2). Thus

$$\alpha(\mathbf{x}, \tilde{\mathcal{F}}) \triangleq \text{SAF}(\boldsymbol{\mu}(\mathbf{x}), \tilde{\mathcal{F}}, \theta_{1:M}). \quad (13)$$

We denote this algorithm by SAF_μ . Use of only the mean prediction has the computational advantage of not requiring a Monte Carlo integration to approximate the EI. In addition we have noticed that the posterior uncertainty is often so great that model predictions are made well beyond the actual Pareto front. These predictions can carry significant weight in the EI estimation leading to the optimisation focusing on regions, which although apparently promising, are in fact infeasible. This effect is reduced if only the mean prediction is used. We also note that since the posterior uncertainty is not required, a wider range of surrogates (e.g. random forests, support vector machines) could be employed.

Algorithm 1 summarises the multi-objective optimisation procedure using SAF. Initially a series of N evaluations of the objective function are made via Latin hypercube sampling. These initial samples are then expensively evaluated by the objective function to produce a set of observed function values $\mathcal{D} = \{(\mathbf{x}_t, \mathbf{f}_t)\}_{t=1}^N$ (line 5). Separate GP surrogates are then fitted, one for each of the M objectives (line 8). Here we use the Matérn 5/2 kernel (as recommended for modelling realistic functions [26]) and the parameters of these GPs are learned by maximising the marginal log likelihood. The location \mathbf{x}_t for the next expensive evaluation is found by

maximising either the SAF infill criterion using the mean predictions from the surrogate model or the EI calculated via Monte Carlo sampling from the posterior prediction using the well-known known CMA-ES optimiser [27]; lines 12 and 14. The objective function is then evaluated at \mathbf{x}_t and the process repeated until the evaluation budget B is exhausted or satisfactory convergence has been obtained. The best estimate of the Pareto set is, at each stage of the algorithm, the maximal non-dominated set of locations that have been expensively evaluated.

IV. EXPERIMENTAL EVALUATION

A. Process

In this section an analysis is conducted of the performance of SAF_μ over 150 evaluations of a set of challenging, synthetic MOPs. For comparison we benchmark the state-of-the-art SMS-EGO [22] and the competitive, yet computationally efficient ParEGO [21]. We also compare the, cheap-to-evaluate minimum probability of improvement (MPOI) infill criterion proposed by Rahat *et al.* [15]. For SMS-EGO we include the later improvements to the treatment of the objective space ϵ -dominated by the attainment surface described by Wagner *et al.* [18]. The prior knowledge of the objective space required to set the reference point for the hypervolume calculation is assumed to be unknown. Instead the reference point \mathbf{R} is updated at each evaluation to the maximum value of the function observed in each objective, plus an offset of 1, as in the original work $R_m = \max_{\mathbf{f} \in \tilde{\mathcal{F}}} (f_m) + 1$. In order to test whether more a exploitative search will be beneficial, we also include the EI based maximin model proposed by Svenson and Santner [24], (hereby denoted as SAF_{EI}), and also a version of SMS-EGO, where the uncertainty from the GP is ignored, instead using mean prediction only, with no ϵ in the dominance relation (denoted SMS-EGO_μ). Finally, we also compared with Latin hypercube sampling (LHS) as a pseudo random search of the parameter space.

Each optimisation was started from 10 initial LHS samples, over 31 repeat optimisations of each test function. To ensure comparability the repeats are structured so that the same random seed and LHS samples are used for all optimisation methods within each repeat. The dominated hypervolume is calculated using the Fonseca *et al.* method [28].

The chosen objective functions for the benchmark tests are a subset the Walking Fish Group problems [7] WFG1-WFG6, which include deceptive functions and functions in which the parameters determining distance from, and position along the Pareto front are not separable. The dimension of the parameter-space d for the WFG problems is scalable, as is the number of objectives M . We tested each function over three configurations of each problem in WFG1-WFG6, with 2, 3 and 4 objectives, and d ranging from 3-12 dimensions. More details of the test functions can be seen in Table I.

Two metrics were used to quantify the performance of each optimiser. The first was the dominated hypervolume or \mathcal{S} -

¹Multi-modal in only the first objective f_1 , uni-modal in $f_2 \cdots f_M$.

Function	(M, d)	\mathcal{X}	Comment
WFG1	(2, 3), (3, 4), (4, 5)	$[0, 2d]^d$	separable, uni-modal
WFG2	(2, 6), (3, 6), (4, 10)	$[0, 2d]^d$	non-separable, multi-modal ¹ , \mathcal{F} is discontinuous.
WFG3	(2, 6), (3, 6), (4, 10)	$[0, 2d]^d$	non-separable, uni-modal
WFG4	(2, 6), (3, 8), (4, 8)	$[0, 2d]^d$	separable, multi-modal
WFG5	(2, 6), (3, 8), (4, 10)	$[0, 2d]^d$	separable, deceptive
WFG6	(2, 10), (3, 6), (4, 12)	$[0, 2d]^d$	non-separable, uni-modal

TABLE I: Walking Fish Group functions, showing the 18 (M, d) combinations tested.

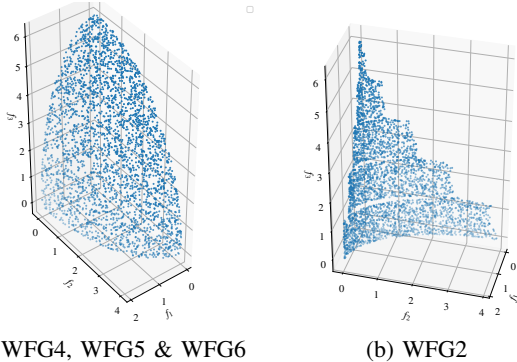


Fig. 2: IGD⁺ reference points, for 3 objective WFG problems.

metric measure. This is a commonly used convergence measure, but as it is also the principal quantity driving SMS-EGO. We also compared algorithms using the Inverted Generational Distance plus (IGD⁺) [29] in order to identify any undue bias in favour of the SMS-EGO and SMS-EGO_μ methods. Given a set of reference points \mathcal{Z} on the true Pareto front, IGD⁺ is

$$\text{IGD}^+(\tilde{\mathcal{F}}, \mathcal{Z}) = \frac{1}{|\mathcal{Z}|} \sum_{\mathbf{z} \in \mathcal{Z}} \min_{\mathbf{f} \in \tilde{\mathcal{F}}} d^+(\mathbf{f}, \mathbf{z}) \quad (14)$$

where d^+ is the modified Euclidean distance given by:

$$d^+(\mathbf{z}, \mathbf{f}) = \sqrt{\sum_{m=1}^M \max(f_m - z_m, 0)^2} \quad (15)$$

Measurements using IGD⁺ rely on a set of reference points \mathcal{Z} on the problem's Pareto front. In order to ensure no bias toward solutions found in localised regions of this surface it is important that the points be distributed evenly across the front. The non-linearity of the objective functions means that the images of uniformly distributed samples from the Pareto set in parameter space are not uniformly distributed on \mathcal{F} . In order to achieve an approximately uniform distribution, we first sampled a non-uniform, but dense population of points on this Pareto front. We then uniformly sampled the attainment surface using the methods described by Smith *et al.* [25]. Attainment surface points which did not lie on the Pareto front were then discarded if their distance to the nearest neighbour in the set of Pareto optimal points, was greater than a certain threshold. This sampling method was used for all functions with the exception of WFG3, for which the Pareto surface is straightforward to compute and sample uniformly. Example reference points are shown in Fig. 2a.

B. Results

The results of the optimisations, as measured by dominated hypervolume and IGD⁺, are shown in table II. These tables show the median result and interquartile range (IQR) for each optimiser over the 31 repeated runs on each test problem. The best median performance is shaded in dark grey, with any statistically similar results, as assessed by paired Wilcoxon signed-rank test with Holm-Bonferroni correction ($p > 0.05$) shaded in light grey. Space constraints preclude the inclusion of detailed convergence plots for individual algorithms and each problem, but code implementing these ideas and detailed plots may be found in the Supplementary Material², and an exemplar convergence plot is shown in figure 4.

After the 150 budgeted evaluations of the objective functions, SAF_μ produced either the best median dominated hypervolume performance, or was statistically equivalent to the best median performance in 12 out of the 18 tested problem configurations. This was the most out of any of the considered alternatives, with SMS-EGO giving best or equivalent performance on 8 objective functions. The results as measured by IGD⁺ were similar, with SAF_μ again offering the best or equivalent performance in 12 out of 18 functions, and SMS-EGO in 10. The sets of non-dominated solutions produced by each method were not characteristically different, and both generally produced good coverage of the Pareto front. For examples of this please see the supplementary materials.

As may be expected, performance of the optimisers varies between problems. The purely exploratory LHS method was not the best strategy for any problem, and in fact was the worst performer on all problems aside from the difficult WFG2 problem. While ParEGO and LHS excelled here relative to their performances on other objective functions (beating SMS-EGO, SMS-EGO_μ and SAF_{EI} in the two-objective configuration), few solutions were found in much of the attainable objective space.

The measurements in Table II only consider a slice of the optimisation process, when the algorithms have largely converged. Figure 3 shows average ranked performance at each step of the optimisation over all 31 repeats and all 18 test function and (M, d) combinations. Where the difference between optimisations of similar rank were not statistically significant (according to Wilcoxon signed-rank test) the two were assigned the mid-point between their respective ranks. From this it is clearer that over the full range of test problems SAF_μ shows the best overall performance in the final stages whether measuring by IGD⁺ or dominated hypervolume, but also shows competitive performance throughout the process. SMS-EGO and SMS-EGO_μ rank similarly and competitively with SAF_μ, but SAF_{EI} falls well behind after a few initial, favourable evaluations. This is surprising as one might expect SAF_{EI} to find better solutions in the later stages, when a more complete representation of the optimised function has been formed by the surrogate [2]. However, we speculate that even

²https://github.com/FinleyGibson/SAF_EMO.git

	WFG1 2M3d	WFG1 3M4d	WFG1 4M5d	WFG2 2M6d	WFG2 3M6d	WFG2 4M10d
	Median	IQR	Median	IQR	Median	IQR
LHS	0.134	0.015	0.107	0.015	0.097	0.013
MPoI	0.261	0.074	0.216	0.052	0.172	0.047
ParEGO	0.326	0.043	0.281	0.053	0.228	0.03
SMS-EGO	0.295	0.026	0.284	0.045	0.234	0.067
SMS-EGO _μ	0.293	0.038	0.281	0.035	0.239	0.072
SAF _{EI}	0.23	0.028	0.2	0.035	0.186	0.021
SAF _μ	0.319	0.038	0.276	0.064	0.241	0.054

	WFG3 2M6d	WFG3 3M10d	WFG3 4M10d	WFG4 2M6d	WFG4 3M8d	WFG4 4M8d
	Median	IQR	Median	IQR	Median	IQR
LHS	0.725	0.024	0.7	0.021	0.727	0.022
MPoI	0.807	0.077	0.815	0.102	0.882	0.049
ParEGO	0.852	0.029	0.864	0.05	0.873	0.035
SMS-EGO	0.976	0.005	0.942	0.025	0.906	0.056
SMS-EGO _μ	0.973	0.006	0.947	0.014	0.899	0.053
SAF _{EI}	0.927	0.014	0.912	0.018	0.938	0.019
SAF _μ	0.968	0.005	0.93	0.02	0.932	0.028

	WFG5 2M6d	WFG5 3M8d	WFG5 4M10d	WFG6 2M10d	WFG6 3M6d	WFG6 4M12d
	Median	IQR	Median	IQR	Median	IQR
LHS	0.617	0.024	0.478	0.027	0.376	0.022
MPoI	0.75	0.066	0.624	0.098	0.462	0.105
ParEGO	0.846	0.043	0.636	0.086	0.492	0.081
SMS-EGO	0.834	0.078	0.69	0.087	0.496	0.09
SMS-EGO _μ	0.774	0.07	0.678	0.056	0.495	0.099
SAF _{EI}	0.682	0.047	0.573	0.069	0.46	0.059
SAF _μ	0.793	0.05	0.673	0.058	0.572	0.065

	WFG1 2M3d	WFG1 3M4d	WFG1 4M5d	WFG2 2M6d	WFG2 3M6d	WFG2 4M10d
	Median	IQR	Median	IQR	Median	IQR
LHS	1.563	0.034	1.981	0.058	2.26	0.078
MPoI	1.063	0.269	1.483	0.231	1.883	0.187
ParEGO	0.969	0.036	1.318	0.113	1.655	0.146
SMS-EGO	0.993	0.035	1.314	0.114	1.634	0.248
SMS-EGO _μ	0.992	0.035	1.297	0.092	1.614	0.286
SAF _{EI}	1.324	0.067	1.636	0.1	1.814	0.128
SAF _μ	1.014	0.035	1.343	0.184	1.621	0.187

	WFG3 2M6d	WFG3 3M10d	WFG3 4M10d	WFG4 2M6d	WFG4 3M8d	WFG4 4M8d
	Median	IQR	Median	IQR	Median	IQR
LHS	0.418	0.042	0.61	0.065	0.635	0.07
MPoI	0.28	0.13	0.402	0.221	0.293	0.139
ParEGO	0.242	0.052	0.336	0.213	0.408	0.108
SMS-EGO	0.038	0.006	0.136	0.053	0.298	0.086
SMS-EGO _μ	0.043	0.012	0.131	0.04	0.303	0.109
SAF _{EI}	0.112	0.031	0.222	0.065	0.274	0.053
SAF _μ	0.047	0.01	0.164	0.042	0.242	0.072

	WFG5 2M6d	WFG5 3M8d	WFG5 4M10d	WFG6 2M10d	WFG6 3M6d	WFG6 4M12d
	Median	IQR	Median	IQR	Median	IQR
LHS	0.436	0.043	0.768	0.064	1.262	0.144
MPoI	0.229	0.099	0.584	0.204	1.182	0.369
ParEGO	0.141	0.044	0.539	0.128	1.09	0.295
SMS-EGO	0.131	0.073	0.434	0.15	1.2	0.446
SMS-EGO _μ	0.19	0.092	0.443	0.115	1.132	0.452
SAF _{EI}	0.325	0.069	0.69	0.212	1.252	0.341
SAF _μ	0.178	0.064	0.458	0.122	0.875	0.189

TABLE II: The median relative dominated hypervolume (left) and IGD⁺ (right) measured after 150 function evaluations and the associated interquartile ranges (IQR), over 31 repeated optimisations of the WFG test functions. The best median performance is shaded in dark grey, while performances which are statistically equivalent are shaded light grey. The theoretical optimum for each is 0, however the limited cardinality of \mathcal{Z} and $\tilde{\mathcal{F}}$ prohibit attaining this in practice.

after 150 evaluations the surrogate is insufficiently faithful that deliberate exploratory evaluations are required.

Neither MPoI or SAF_{EI}, which account for the surrogate’s uncertainty, are competitive over the broad range of functions tested, with both falling well short of the exploitative SAF_μ approach in the majority of functions tested. In problems where SAF_{EI} failed to match the results obtained by SAF_μ the solutions found were similar, but there were fewer non-dominated solutions produced after 150 evaluations; for example, SAF_{EI} finds an average of 98 non-dominated in WFG5 ($M = 3$ and $d = 8$) compared to 128 found by SAF_μ.

As measured by dominated hypervolume SAF_μ is never surpassed in a statistically significant manner by any of the computationally inexpensive alternatives (LHS, MPoI, ParEGO & SAF_{EI}), with the exception of two of the WFG2 benchmarks, which present a particularly difficult challenge and were poorly optimised by all methods. SAF_μ is also better when measured by IGD⁺ in the majority of cases, and its average ranked performance is better than any of these methods beyond the initial 10 steps of the optimisation, whichever measurement method is used (Figure 3).

The two methods which utilise hypervolume calculations in their infill criteria, SMS-EGO and SMS-EGO_μ, did not demonstrate better overall performance than the cheap SAF_μ method, nor did the posterior uncertainty information leveraged by SMS-EGO benefit it significantly over SMS-EGO_μ. SMS-EGO had slightly more *winning* results, being the best or equivalent in 10 out of 18 experiments compared to 8 by SMS-EGO_μ when measured by dominated hypervolume (both with 10 by IGD⁺), but the average ranked performances were very similar. SMS-EGO represents the state of the art, and was inferior in the majority functions tested in this set of

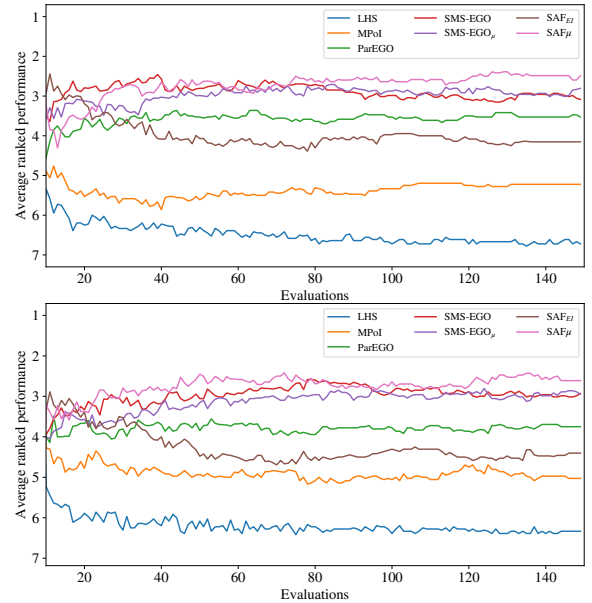


Fig. 3: Average ranked performance of optimisers over all 18 test function and (M, d) combinations shown in Table II; dominated hypervolume (above) and IGD⁺ (below).

experiments to the proposed SAF_μ method.

V. CONCLUSION

When optimising expensive, multi-objective problems efficient use of objective evaluations is critical to convergence within a constrained budget. Current state-of-the-art methods for optimising such problems in few evaluations rely on dominated hypervolume based infill criteria, which can themselves

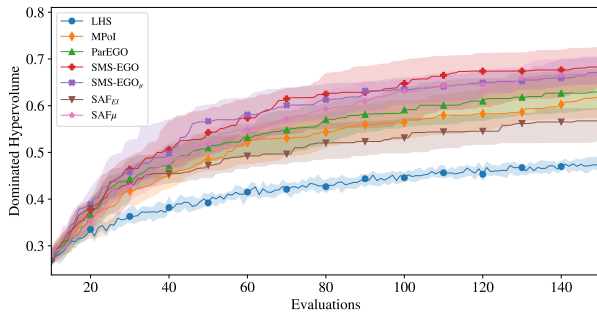


Fig. 4: Convergence on WFG5 ($M = 3$, $d = 8$) over 150 evaluations, starting from 10 LHS samples. Solid lines show median values over 31 repeats, shaded regions show the IQR.

become prohibitively costly as the number of objectives grows and the number of solutions increases.

The novel summary attainment front infill criterion was empirically shown to produce state-of-the-art results over a range of optimisation problems. Both SMS-EGO and the SAF infill criterion require the Pareto set of solutions to be determined from the evaluated solutions, which can take $\mathcal{O}(t^2)$ time after t iterations. Determining the SAF infill criterion for a single putative solution takes $\mathcal{O}(|\tilde{\mathcal{P}}|)$, that is $\mathcal{O}(t)$ worst case. On the other hand SMS-EGO, using an efficient hypervolume calculation, requires $\mathcal{O}(2^{M-1} \cdot |\tilde{\mathcal{P}}|^{\lfloor M/2 \rfloor})$ [20]. The SAF infill criterion is thus much cheaper to compute than SMS-EGO. This theoretical analysis is borne out by empirically: with $|\tilde{\mathcal{P}}| = 100$ and $M = 2$, SAF is approximately 40 times faster to compute than SMS-EGO; when $M = 9$ the ratio is ≈ 1750 . Detailed timings are given in the Supplementary Materials.

We also demonstrated that there is often no need for deliberately exploratory evaluations of the objective function, and that exploitative evaluation of candidates with the best mean performance predicted by the surrogate leads to better convergence towards an approximation of the Pareto set in the majority of cases. This finding invites the exploration of alternative surrogates which may be better able to model complex objective functions, but do not provide uncertainty estimates for posterior prediction, such as support vector machines or random forests.

REFERENCES

- [1] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.
- [2] G. De Ath, R. M. Everson, A. A. M. Rahat, and J. E. Fieldsend, "Greed is Good: Exploration and Exploitation Trade-offs in Bayesian Optimisation," *ACM Trans. on Evolutionary Learning and Optimization*, 2021 (In Press).
- [3] I. G. Osio and C. H. Amon, "An engineering design methodology with multistage Bayesian surrogates and optimal sampling," *Research in Engineering Design*, vol. 8, no. 4, pp. 189–206, 1996.
- [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, pp. 1–10, 2017.
- [5] S. Jeong, M. Murayama, and K. Yamamoto, "Efficient optimization design method using kriging model," *Journal of Aircraft*, vol. 42, no. 2, pp. 413–420, 2005.

- [6] J. Fang, G. Sun, N. Qiu, N. H. Kim, and Q. Li, "On design optimization for structural crashworthiness and its state of the art," *Structural and Multidisciplinary Optimization*, vol. 55, no. 3, pp. 1091–1119, 2017.
- [7] S. Huband, L. Barone, L. While, and P. Hingston, "A scalable multi-objective test problem toolkit," in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 280–295, Springer, 2005.
- [8] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 42, pp. 55–61, 2000.
- [9] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*, pp. 63–71, Springer, 2003.
- [10] H. J. Kushner, "A new method of locating the maximum point of an arbitrary mul tipeak curve in the presence of noise," *Journal Basic Engineering*, vol. 86, no. 1, pp. 97–106, 1964.
- [11] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," in *International Conference on Machine Learning*, pp. 1015–1022, 2010.
- [12] A. D. Bull, "Convergence rates of efficient global optimization algorithms," *J. Machine Learning Research*, vol. 12, pp. 2879–2990, 2011.
- [13] W. Scott, P. Frazier, and W. Powell, "The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression," *SIAM Journal on Optimization*, vol. 21, no. 3, pp. 996–1026, 2011.
- [14] B. Ru, M. A. Osborne, M. Mcleod, and D. Granzio, "Fast information-theoretic Bayesian optimisation," in *International Conference on Machine Learning*, pp. 4384–4392, PMLR, 2018.
- [15] A. A. Rahat, R. M. Everson, and J. E. Fieldsend, "Alternative infill strategies for expensive multi-objective optimisation," in *Proc. of the Genetic and Evolutionary Computation Conference*, pp. 873–880, 2017.
- [16] 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, pp. 7–3, 2008.
- [17] 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.
- [18] T. Wagner, M. Emmerich, A. Deutz, and W. Ponweiser, "On expected-improvement criteria for model-based multi-objective optimization," in *International Conference on Parallel Problem Solving from Nature*, pp. 718–727, Springer, 2010.
- [19] M. Fleischer, "The measure of Pareto optima applications to multi-objective metaheuristics," in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 519–533, Springer, 2003.
- [20] K. Yang, M. Emmerich, A. Deutz, and T. Bäck, "Efficient computation of expected hypervolume improvement using box decomposition algorithms," *J. Global Optimization*, vol. 75, no. 1, pp. 3–34, 2019.
- [21] J. Knowles, "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Trans. on Evolutionary Computation*, vol. 10, no. 1, pp. 50–66, 2006.
- [22] W. Ponweiser, T. Wagner, D. Biermann, and M. Vincze, "Multiobjective optimization on a limited budget of evaluations using model-assisted S -metric selection," in *International Conference on Parallel Problem Solving from Nature*, pp. 784–794, Springer, 2008.
- [23] A. J. Keane, "Statistical improvement criteria for use in multiobjective design optimization," *AIAA journal*, vol. 44, no. 4, pp. 879–891, 2006.
- [24] J. Svenson and T. Santner, "Multiobjective optimization of expensive-to-evaluate deterministic computer simulator models," *Computational Statistics & Data Analysis*, vol. 94, pp. 250–264, 2016.
- [25] K. I. Smith, R. M. Everson, and J. E. Fieldsend, "Dominance measures for multi-objective simulated annealing," in *Proceedings of the 2004 Congress on Evolutionary Computation*, vol. 1, pp. 23–30, IEEE, 2004.
- [26] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *arXiv:1206.2944*, 2012.
- [27] N. Hansen, S. D. Müller, and P. Koumoutsakos, "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)," *Evol. Comp.*, vol. 11, pp. 1–18, 2003.
- [28] C. M. Fonseca, L. Paquete, and M. López-Ibáñez, "An improved dimension-sweep algorithm for the hypervolume indicator," in *2006 IEEE Conference on Evolutionary Computation*, pp. 1157–1163, 2006.
- [29] H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima, "Modified distance calculation in generational distance and inverted generational distance," in *International conference on evolutionary multi-criterion optimization*, pp. 110–125, Springer, 2015.