The 9th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2024
3–7 June 2024, Lisboa, Portugal

MULTI-FIDELITY VARIABLE COST OPTIMISATION

MARK J. D. PELLOWE[1], FELIX PRUTTON[2], GARY PAGE[1], ADRIAN SPENCER [1]

[1] National Centre for Combustion and Aerothermal Technology
Loughborough University
Loughborough, United Kingdom
e-mail: m.pellowe@lboro.ac.uk

[2] Whittle Laboratory
University of Cambridge
Cambridge, United Kingdom

Key words: Bayesian Optimisation, Gaussian Processes, Multi-Fidelity Analysis

Summary. In this work, a novel cost function is introduced within the domain of multi-fidelity Bayesian Optimisation, addressing a significant gap in the consideration of costs associated with experimental design in real-world applications. This newly proposed cost function encapsulates the property of most studies - some parameters are cheaper to change than others. This new cost function is applied to a synthetic test function and an engineering dataset with relevant properties using a number of strategies, including single and multi-fidelity treatments. It is found that, in both single-fidelity and multi-fidelity settings, adaptive cost treatment generally improves the performance of optimisers. However, in certain cases, care needs to be taken in determining the relative costs due to 'scanning' behaviour - especially when fidelities are not well correlated.

1 Introduction

Bayesian Optimisation (BO) describes the problem of constructing a strategy for finding optima of unknown functions through repeated sampling. A BO strategy works by taking samples from the unknown function, building a surrogate model of the unknown function based on that sample data, and sequentially selecting further points to sample based on the outputs of that surrogate model. A typical assumption of BO is that sampling is costly, and that a useful strategy is one that converges on a global optimum while minimising the costs associated with sampling.

Multi-fidelity Bayesian Optimisation (MF-BO) is a variation of the BO problem but where the experimentalist has access to a set of lower-fidelity functions, correlated to the original function from which they can take samples. These correlated functions are less costly to sample from than the original, and so successful MF-BO strategies must balance trade-offs between costs and accuracy when evaluating points to sample.

Many problems in engineering can be framed as instances of BO and MF-BO problems. MF-BO problems often arise when an experimentalist has the choice between either taking physical measurements of a system, or running cheaper (but less accurate) computer simulations of that same system [10], such as when determining the aerodynamic performance of a wing: a physical

experiment, or a Reynolds Averaged Navier Stokes (RANS) simulation can be performed, where the more trusted results would come from the physical experiment - but at a much higher cost.

Multi-fidelity systems are ubiquitous in fluid dynamics. From the initial design space coverage using low-order or simplified models to determine desirable attributes, to more detailed analysis of particular candidate designs with more robust methods such as Large Eddy Simulations (LES), and even experiments, there exists different fidelity data in different quantities of any system - and each fidelity exhibits better agreement with the true system, which typically correlates with increased cost of observation.

Many different types of MF-BO problems have been explored previously [6, 13, 9, 15]. In its most basic form, the cost of sampling is solely dependent on the fidelity of the sampled function. Further work has seen cost functions that take experiment parameters into account [15], so that not all experiments at the same fidelity are equally as expensive; a practical consideration central to much practical design.

The availability of these information sources of variable fidelities presents a number of practical problems to an experimentalist:

1. What is the best way to combine the results of different types of experiments (with different biases and uncertainties across the design space)?

2. Which type of experiments should be performed (and in what regions of the design space) to gain the most useful information in the most cost-effective manner?

In this paper, we introduce a new, dynamical cost function to MF-BO. That is, a cost is associated with the sampling of a point in the design space, but this function also depends on the previously sampled point. The fact that some experiments are more expensive to perform than others has been incorporated by others in previous research but, to the authors' knowledge, there has been no attempt to incorporate dynamic cost functions of the type described in the multi-fidelity context.

The paper discusses the previous related work on the topic and relevant theoretical background. Then, the additional dynamic cost function is explained in detail, as well as how it fits into a multi-fidelity context. After this, an algorithm to efficiently optimise an expensive 'black-box' function with multiple information sources, with a dynamical cost function is proposed and is then applied to one synthetic test case and one real engineering case.

2 Methodology

2.1 Gaussian Processes

Gaussian Processes (GPs) are stochastic processes - that is, a set of indexed random variables - with the defining property that any finite subset of the indexed variables is jointly normally distributed [12]. The covariance between two random variables in the normally distributed subset is defined by some specified function called the kernel $K(x, x'; \theta)$, with θ being a vector of hyperparameters.

The covariance kernel can take many forms and can be selected based upon the expected form of the function. The most common is the smooth and continuous Radial Basis Function kernel which is used in this work. For an excellent discussion of different kernels and their properties, one can refer to [5].

The prediction of the mean and the variance of the GPs are performed with the equations [12]:

$$y(x^*) = \mu\mathbf{1} + \mathbf{K}(x^*, x)\left(\mathbf{K}(x, x) + \sigma^2 I\right)^{-1}(y - \mathbf{1}\mu) \tag{1}$$

$$\mathbf{\Sigma}(x^*) = \mathbf{K}(x^*, x^*) - \mathbf{K}(x^*, x)\left(\mathbf{K}(x, x) + \sigma^2 I\right)^{-1}\mathbf{K}(x, x) \tag{2}$$

where x^* are the prediction points, x are the training points, $\mathbf{K}$ are the kernels conditioned on hyperparameters according to the prior covariance function, I is an identity matrix, σ^2 is the noise, and $\mathbf{1}$ is a vector of ones.

2.2 Multi-fidelity Gaussian Processes

Multi-fidelity Gaussian Processes, in a variety of forms, have found use in a broad range of engineering problems for BO. There are two models which are generally used when optimising problems with access to cheaper information sources. One is the Auto-Regressive model of Kennedy and O'Hagan [9] and has been shown to be useful for hierarchical information structures, where the highest fidelity is the 'truth' and all other observations are corruptions of this 'truth'.

The other commonly used model, which will be the model used in this paper, is the Linear Model of Coregionalisation (LMC). The LMC model is referred to as a 'multi-output' model; since it correlates different outputs together whilst not assuming a strict hierarchy [4].

The model works by transforming the covariance matrix into:

$$\mathbf{K}(x, x') = \sum_{i=1}^{t} \mathbf{B}_i k_i(x, x')$$

where $\mathbf{B}$ is a coregionalisation matrix [2] with components $\sum_{j=1}^{r_i} a_{s,i}^j a_{s',i}^j$ with r referring to the rank of matrix $\mathbf{B}$, which controls the number of GPs that share the same covariance structure, and a is a scaling parameter between each GP. Inference with this model follows the same strategy as in eqs. (1) and (2). In this work, the model is implemented using GPyTorch [8] to take advantage of the automatic differentiation capabilities.

2.3 Acquisition Functions and Bayesian Optimisation

In the simplest terms, BO can be summarised as [7]: $\max_{x\in\mathcal{A}} f(x)$ where $\mathcal{A}$ is the domain of the problem. This maximisation can be translated to minimisation by simply negating the function at all points. Generally, since the function $f(x)$ is expensive to evaluate, it's desirable for the search to try to balance exploration and exploitation. Exploration utilises the confidence intervals from the model to determine where the model is uncertain of the form of the function. For BO to be efficient, these uncertainties should be well-defined and derived from noise in the data rather than uncertainties in the model's parameters. The exploitation aspect of BO concerns utilising the model where optimal values are likely to already exist. This allows the optimiser to resolve the area of the desired optimum to the finest extent, whilst avoiding 'wasteful' observations where an optimum is unlikely to exist.

As the most probable value of the function is μ_* then, since some uncertainty, σ_* exists in the model, the expectation of improvement becomes:

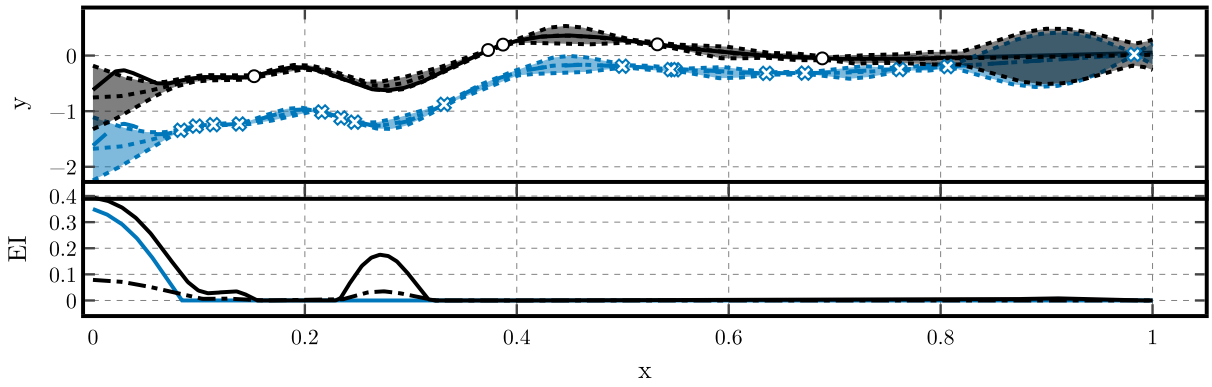


Figure 1: Multi-Fidelity Bayesian Optimisation. The top figure demonstrates the true high, , the true low, , the samples from the high and low-fidelity functions as circles and crosses; respectively, and the GP model fits, . The bottom figure shows the acquisition functions before, , and after scaling, .

$$\mathbb{E}[I(x)] = \left((\mu_*(x) - y_{best})\Phi\left(\frac{\mu_*(x) - y_{best}}{\sigma_*(x)}\right) + \varphi\left(\frac{\mu_*(x) - y_{best}}{\sigma_*(x)}\right)\right) \tag{3}$$

where Φ is the cumulative distribution function, and φ is the probability distribution function, which are both Gaussian. Equation (3) dictates that when $\sigma_* = 0$, then $\mathbb{E}[I(x)] = 0$ as there is nothing more to learn about the function in this area; this prevents the suggestion of resampling.

Figure 1 shows a multi-fidelity function where the lighter colour denotes high-fidelity and darker denotes low-fidelity. The aim is to find the minimum of the high-fidelity function by using the EI acquisition function from eq. (3), but dividing it by a cost. The bottom figure shows how a sample is to be taken from the high-fidelity source, but after scaling by cost (assuming the high-fidelity is five times the cost of low-fidelity), the optimiser elects to add the low-fidelity point. This occurs over and over until a suitable convergence criterion is met. After this, since different fidelities are generally of different costs and offer different levels of information, the 'value' of querying a sample can differ; hence, authors began to implement strategies of deciding when low-fidelity and high-fidelity queries are appropriate based on normalising the amount of information gain of a query by a suitable cost function.

2.3.1 Multi-Fidelity Specific Acquisition Functions

The method used in this paper is a Knowledge gradient method, which are a relatively inexpensive way of normalising information gain per unit cost when compared to the first introduced entropy-based methods [10].

The optimisation occurs in rounds of exploration where the next sample is chosen by evaluating each of the fidelities and their expected information gain upon the next iteration. To calculate this, the authors in [11] developed the Multi-Information Source Optimisation using the Knowledge Gradient (misoKG) algorithm. Having modelled n data points at various fidelities and points in the design space, the 'misoKG factor' is given by [11]:

$$\mathrm{MKG}^n(t,x) = \mathbb{E}\left[\frac{\max_{x'\in\mathcal{D}}\mu^{(n+1)}(0,x') - \max_{x'\in\mathcal{D}}\mu^{(n)}(0,x')}{C_t(x)}\right] \tag{4}$$

where $\mu^{(n)}$ is the posterior mean of the model at iteration n, $C_t(x)$ is the cost of the information source at fidelity t at point x in the domain, and t^{n+1} and x^{n+1} are the next fidelity and sample locations, respectively, $\mu^i(0,x)$ represents the expected value of the true objective function at x after sampling i data points. Monte-Carlo sampling is used to estimate eq. (4) using the software package BOtorch from [3].

2.4 Cost-Adaptive Methodology

In a situation where it is possible to choose from a wide variety of experiments, it is inevitable that the cost of performing different experiments will vary. Not only will some experiments be intrinsically more expensive to perform (perhaps they use more power, or take more time than other experiments), but in practice, there are costs associated with changing the parameters of the experiments or switching from one experiment type to another. It is clear that the cost of performing the next experiment in a series is therefore not only a function of the experiment parameters but also of the parameters of the past experiments too. Under these assumptions, we can construct cost functions that better reflect the practical nature of experiments.

As well as being of variable cost, different experiments are likely to result in variable utility. The assessment of the utility of a potential experiment is a key aspect of (BO): a surrogate model in combination with an acquisition function can be used to assess the potential utility. Dividing the utility of a potential experiment with the cost it would take to perform it offers a useful metric - utility per unit cost. It is this metric of an experiment (or series of experiments) that should be maximised.

2.4.1 A New Cost Function

As an example, we consider a system with a QOI and two continuous parameters that can be altered from one experiment to another experiment. We imagine that the cost of a potential experiment at a location in the parameter space (x_1, x_2) is a function of the new parameters, as well as a function of the old ones $(x_{1.prev}, x_{2.prev},)$. We assume some facts about this cost function, that should generalise to a large number of experimental systems:

1. There is some intrinsic, non-zero cost to every experiment: $C_{\text{intrinsic}} = C_{\text{in}}(x_1, x_2) > 0$.
2. There is an additional cost incurred if a parameter is changed from its previous value: $C_{\text{change}} = \lambda_1\delta(x_1 - x_{1,\text{prev}}) + \lambda_2\delta(x_2 - x_{2,\text{prev}})$ where $\delta(x)$ is defined by: $\delta(0) = 0$ and $\delta(x \neq 0) = 1$ and λ_1, λ_2 are suitably chosen coefficients.
3. There is an associated cost with switching between experiments: C_{switch}.

The coefficients chosen, $\lambda_x = \lambda_1 \ldots \lambda_n$ should be normalised by the most expensive parameter. For example, the expensive parameter is 1, whereas a parameter that is only 20% of the cost of this will therefore have a value of 0.2.

The total cost of an experiment is then given by the sum of these terms:

$$C_{\text{total}} = C_{\text{in}}(x_1, x_2) + C_{\text{change}}(x_1, x_2, x_{1,\text{prev}}, x_{2,\text{prev}}) + C_{\text{switch}} \tag{5}$$

It is trivial to expand to higher dimensions, and in a general n dimensional case becomes:

$$C_{\text{total}} = C_{\text{in}}(\mathbf{x}, \mathbf{x}_{\text{prev}}) + C_{\text{change}}(\mathbf{x}, \mathbf{x}_{\text{prev}}) + C_{\text{switch}} \tag{6}$$

where we can define $\mathbf{x}$ as the collection of inputs at the observation point, $(x_1, x_2, \ldots, x_n)$; and $\mathbf{x}_{\text{prev}}$ as the inputs from the previous observation, $(x_{1,\text{prev}}, x_{2,\text{prev}}, \ldots, x_{n,\text{prev}})$.

As the cost can depend on several factors, such as time and money, defining exactly what the cost of an information source is can be difficult. Where cost-aware optimisation strategies have been introduced, such as in [10, 13, 15] they have assumed that the cost of each information source is constant across every dimension of the domain. This, however, becomes problematic when optimising problems that contain real multi-fidelity data.

2.4.2 Using the Adaptive Cost-Function

In this work, the new cost function is applied to eq. (4) as $C(x)$ which works as a subclass of the BOtorch [3] `AffineFidelityCostModel` class. The implementation saves the previous observation, and uses the λ values, stored as a tensor, to reduce costs down the appropriate dimensions - it also does this in the t-batches of data, a feature of BOtorch, to allow the acquisition function to be aware when dispatching across multiple cores. The cost function changes how expensive it is to take a sample given the previous observation in each fidelity. It applies to both a single-fidelity setting, where an experiment to determine the optimal configuration needs to be determined, but it works more intuitively in a multi-fidelity setting. This is because, in a single-fidelity setting, the optimisation doesn't have an indication from lower-information sources to suggest where the optimal configuration might be. Hence, the exploratory phase in a single-fidelity setting will be highly robust and, especially in experiments, already be enough resolved to remove the need for further exploratory analysis.

The difference in this workflow is that it's an assumption that more exploration in the highest-fidelity search space can be reasonably performed. If the optimiser can suggest which fidelity of samples to take at subsequent searches such as in [11, 6], then the relationship between each of the fidelities must be robust enough to ensure the high-fidelity mapping is appropriate.

There are two distinct ways of interpreting the reduced costs, and each can be incorporated in slightly different ways. The first is accessing cheaper observations based on an observation that immediately opens up cheaper exploration paths - such as in an experimental setup where changing a set of certain parameters is now relatively much cheaper than going from another observation. The second way is considering in an experiment that a certain expensive geometry may exist, of which certain parameters can be changed, but the fact that the geometry already exists allows for a slightly cheaper observation along one of the dimensions this geometry can travel through.

3 Results

The results that follow demonstrate the effects of the variable-dimension-aware cost function. Two particular attributes of the systems were investigated in detail:

Table 1: The optimisation strategies employed.

Strategy Name	Surrogate Model	Cost Function	Acquisition Function	Samples from all fidelities?
SFOC	GP	Original	EI	No
SFAC	GP	Adaptive	EI	No
MFOC	LMC GP	Original	misoKG	Yes
MFAC	LMC GP	Adaptive	misoKG	Yes
MFSFSOC	LMC GP	Original	EI	No
MFSFSAC	LMC GP	Adaptive	EI	No

1. The strength of the Pearson's correlation coefficient, ρ, between the low-fidelity and high-fidelity functions.

2. The values of the λ parameters.

These aspects of the system were varied and the resulting effects of these variations were observed and are discussed in this section.

Table 1 specifies the different optimisation search strategies employed in this paper. The six strategies explore different combinations of surrogate models, cost functions, acquisition functions and sampling techniques. For example, the MFSFSAC strategy uses the same multi-fidelity model as the MFAC strategy but a different acquisition function, intended to only add to the high-fidelity data. Such a strategy might be suitable in circumstances where the cost of switching between fidelities is prohibitively high, for instance.

These strategies can be adapted to work on problems of any dimension. However, in the interest of improving the interpretability of the results, the datasets in this paper are restricted to two-dimensions.

3.1 Six-Hump Camel Back Functions

The Six-Hump Camel Back is a two-dimensional system, comprising two functions (high-fidelity and low-fidelity) of the form $Q(x_1, x_2) : [-2, 2] \times [-1, 1] \subset \mathbb{R}^2 \to \mathbb{R}$. The two functions, $\{Q_H, Q_L\}$, are defined as follows:

$$Q_H(x_1, x_2) = x_1^2 \cdot \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right) + x_1 \cdot x_2 + x_2^2 \cdot (-4 + 4x_2^2) \tag{7}$$

$$Q_L(x_1, x_2) = Q_H\left(\mathbb{A} \cdot (x_1, x_2)\right) + x_1 \cdot x_2 - 15 \tag{8}$$

where $0 \leq \mathbb{A} \leq 1$ acts as a parameter that determines the correlation between the low and high-fidelity sources.

The high-fidelity function consists of 6 local minima, of which two are global minima. The global minima occur at $(-0.0899, 0.713)$ and the point corresponding to this minimum's reflection in the line $y = x$. At these points, the function takes a value of -1.032.

A number of multi-fidelity optimisation strategies were tested on the 6HCB system. The search strategies were initialised with 50 random samples from the low-fidelity function and 4 random samples from the high-fidelity function. For this set of experiments, the cost of sampling from the high-fidelity function - the 'Cost Ratio' = $\frac{C_h}{C_l}$ - was varied, as was the correlation between the low- and high-fidelity functions. The correlation between the functions varies between

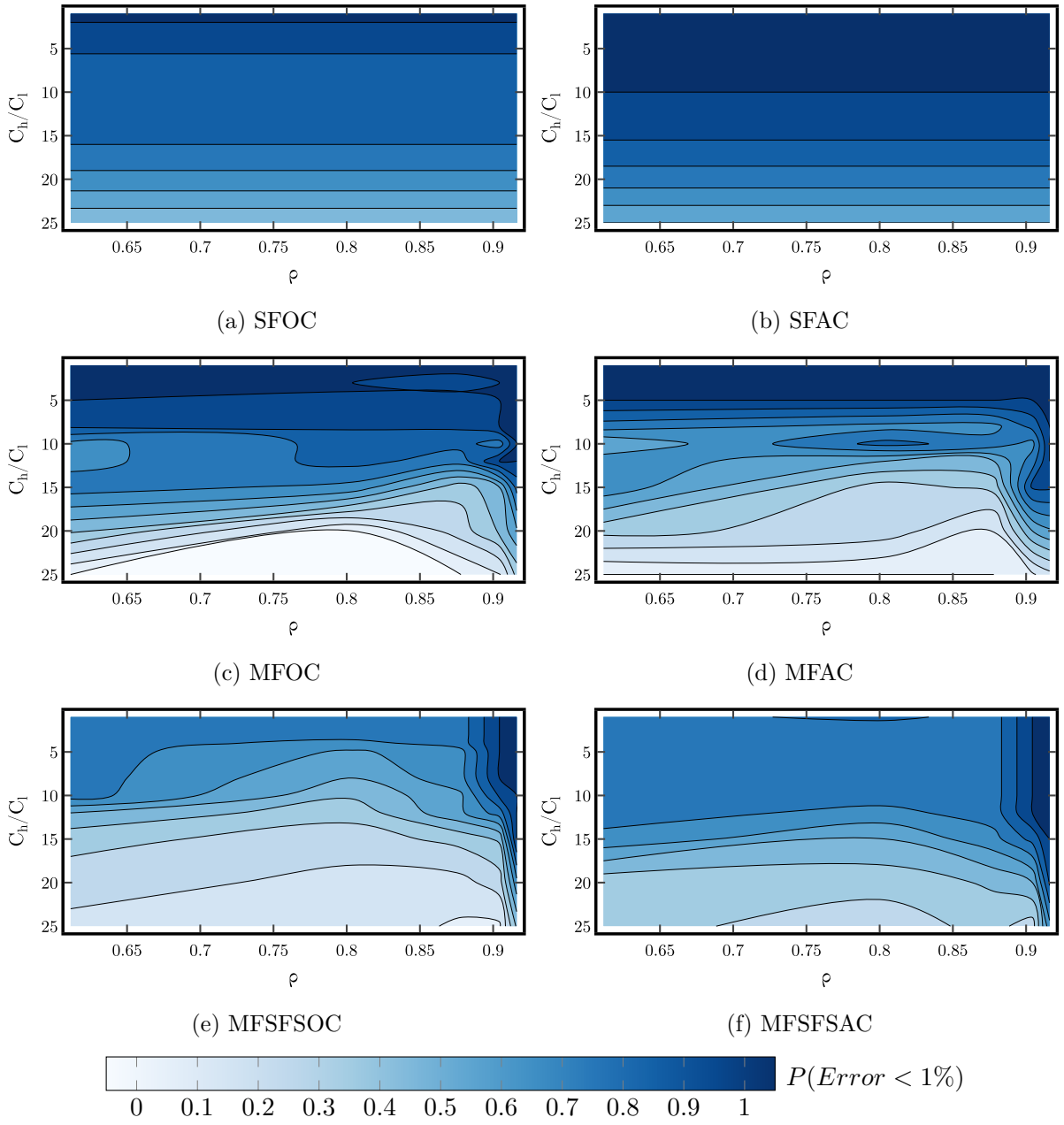


Figure 2: Probability of a strategy producing a minimum within 1% of the true minimum for the Six-Hump Camel Back Function.

0.6 and 1, and the Cost Ratio varies between 2 and 25. A low-fidelity sample is defined to cost 1 unit, and the budget for each strategy is set at 334 units.

Figures 2c and 2d depict the performance of the MFOC and MFAC strategies with this system when configured with different cost ratios and correlations. For the adaptive cost case, the parameters associated with the cost vectors were $\lambda_1 = 1.0$, $\lambda_2 = 0.1$. The plots demonstrate how higher correlations between the low- and high-fidelity functions leads to lower error rates

when using these particular multi-fidelity strategies. If correlations are too low, the optimiser can more easily become misguided by the poor relationship, resulting in strategies getting stuck in local minima and missing the true global minimum. However, the addition of the cheaper dimension in the MFAC model allows for more exploration in the high-fidelity function, which allows for more consistent optimisation at lower correlations, the low correlations still have a tendency to misguide the optimiser.

When the correlation is very strong - around 0.9 or higher - both multi-fidelity strategies are able to outperform the single-fidelity strategies up to a cost ratio of 12. When the MFAC model is used, a more favourable cost-vector is able to guarantee convergence of optimisers up to a cost ratio of 14, after this, too much of the budget is expended on the low-fidelity, such that even with favourable cost-vectors, the strategy can not properly learn the inter-fidelity mappings.

The cost vector effects on the multi-fidelity models were consistently better than the original cost counterparts, as opening a cheaper path up allowed the optimiser to search the high-fidelity space more effectively. This was particularly visible for the MFSFSOC and MFSFSAC models in figs. 2e and 2f, respectively, where the cheaper dimensions led to consistently better performance over the non-cost-adaptive variant.

The results of the single-fidelity strategies are shown in figs. 2a and 2b. Varying the correlation between the functions does not affect the single-fidelity results - in these strategies, low-fidelity samples are ignored entirely and have no impact on future sampling.

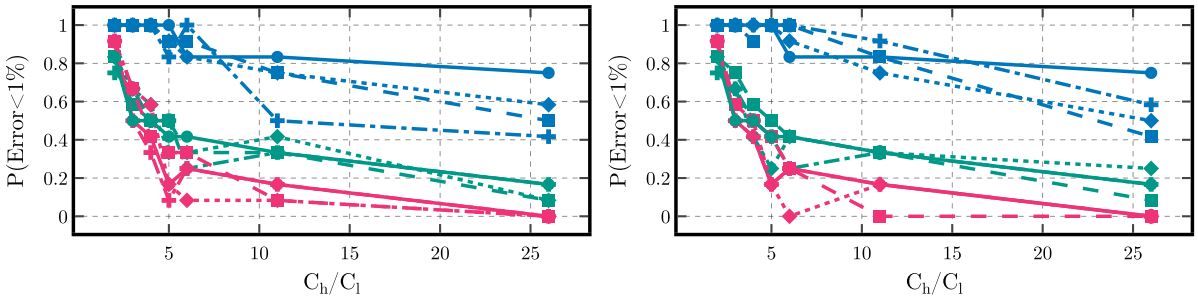
However, the same cannot be said for differences in the λ parameters. The SFAC strategy assumes the existence of a cheaper-to-explore dimension: in this case, $\lambda_1 = 1.0, \lambda_2 = 0.1$. SFOC, with no preferential dimension, has $\lambda_1 = 1, \lambda_2 = 1$. The SFAC strategy obtains minima within 1% of the true global minimum close to 100% of the time, for C_h/C_l values of up to 11. The SFAC performance above cost ratios of 15 outperforming the multi-fidelity variants is due to having to waste any budget on low-fidelity approximations in the initialisation, so the only learned values are those which are from the high-fidelity.

Overall, the best performer was the MFSFSAC model at the maximum correlation value ($\mathbb{A} = 1$), with the optimiser guaranteeing the optimum is found within 1% up to a cost ratio of 15. This is due to initialisation with the low-fidelity information providing useful information about the form of the high-fidelity function. The

3.2 Optimising the Lift-to-Drag Ratio of an Aerofoil

This case concerns the optimisation of the lift-to-drag ratio of a NACA4415 aerofoil as a function of angles of attack; $\alpha \in [-5, 20]$ and Reynolds number; $Re \in [9\times10^5, 1\times10^7]$. The system is composed of data from three sources with different fidelities and associated costs. These sources (and their costs) are as follows, in order of decreasing fidelity: measurements from a wind-tunnel test [1] (variable cost - $C \in [2, 26]$ per sample); SU2 RANS simulations with the $k-\omega$ SST turbulence model (costs 0.5 $\times$ wind-tunnel sample); and X-foil simulations (fixed cost = 1 per sample). The correlation between these sources was 0.95 for the SU2 data, and 0.91 for the X-foil data - where the slight discrepancies are explained by X-foils under predictions of drag due to viscous effects not being included.

The wind-tunnel data has an additional relevant property, in that one of its variables is preferential to change over the other - Re is much more practical to alter than α as it only requires changing the wind tunnel speed - thus making this system an interesting test case of



(a) λ_α changes

(b) λ_{Re} changes

Figure 3: The effects of changing the λ values for the angle of attack, fig. 3a and for the Reynolds number, fig. 3b. The most favourable λ are at values of 0.1, (-■-) , the second most favourable are 0.5, (···◆···) , and the least favourable are 0.9, (·-✚-·) . The original cost models, without changes in λ, equivalent to 1.0, are denoted by (—●—) . All 6 strategies are shown where the AC models correspond to the changed lambda values. Multi-Fidelity (MF) (■) , Multi-Fidelity Single-Fidelity Search (MFSFS) (■) , and Single-Fidelity (SF) (■) .

adaptive cost functions. C_h/C_l is defined as the ratio between the wind-tunnel cost and the X-foil cost.

When assessing the sensitivity of this case to different initial conditions, a method of simulating different initialisations must be produced. One method of initialisation could be the random selection of points from the dataset, but this constrains the intialisations somewhat, and could produce results specific to this particular instance of sampling, as opposed to offering a more general insight into the behaviour under different strategies. Instead, for each fidelity, the entire dataset is used to train a single-fidelity GP, and the mean of the resulting posterior is sampled at random points to produce an initialisation. For the analysis that follows, the data is initialised with 16 X-foil samples, 4 SU2 samples, and 2 wind-tunnel samples with a budget of 282.

Figure 3 demonstrates the impact of changing the values of λ_1 and λ_2. The plots show how, on 12 random initialisations, the probability of the search strategies getting within 1% of the true global optimum are influenced by these parameters, as well as the cost ratio of wind tunnel to X-foil data.

Figure 3a represents the behaviour of the system under the assumption that Re is the more expensive variable to adjust - a realistic assumption. In these cases, MF strategies were superior - most of the instances obtained minima within 1% of the true optimum at $\frac{C_h}{C_l} = 26$. The effects of changing λ_α were varied - when $C_h/C_l < 5$, each change in λ_α performed similarly.

However, at higher cost ratios, the effects of an increasingly preferential dimension are unfavourable. When λ_α was larger, the strategy outperformed strategies with the smallest λ_α by almost 20% in the high-cost-ratio regime. This can be understood by considering that high costs will encourage the optimiser to take observations along the cheapest direction, which results in slightly more coverage in scans of the high-fidelity function. In the angle of attack case, since the relationship is relatively simple - increasing the angle of attack increases lift significantly up to a certain point (the stall angle), while drag also increases but at a slower rate. However, as the angle of attack approaches and exceeds the critical stall angle, drag increases dramatically, and lift decreases sharply due to flow separation over the aerofoil. This relationship can be

smoothed over, and an underprediction of the lift to drag ratio in this area can lead to the optimiser missing the global optimum, and being encouraged to scan down other paths. This should be partially mitigated with the inclusion of multi-fidelity data, but overconfidence from poorer local correlations (the prediction of the stall angle is poor in the lower fidelities) hinder performance for this particular case.

Figure 3b shows the effects of changing λ_{Re} - this represents an implicit assumption that changing Re is more costly than changing α. Similar general trends to the λ_α changes are observed for all MFSFS and SF models. MF models performed better for the most part in this regime: these strategies resulted in minima within 1% of the global optimum for cost-ratios up to 11 in at least 60% of random initialisations. The best results occurred when $\lambda_{Re} = 0.9$. A drop-off in performance occurs at smaller values of λ_{Re} for reasons analogous to those discussed previously - a particularly cheap dimension encourages the model to explore only along 'sweeps', hindering a more thorough exploration of the entire domain.

A drawback of using an adaptive-cost methodology to this dataset is that of sensitivity to the random sampling at initialisation. The GP surrogate model used here reverts to zero-mean when extrapolating beyond the data, and so struggles to cope with the limited initial data provided in the highest fidelity. An unfortunate random initialisation can result in an overconfident, poorly informed surrogate model, leading to poor subsequent selection. Cheaper λ values can further exacerbate these issues, by discouraging 'expensive' exploration, and failing to update and correct the surrogate model.

4 Conclusion

In this paper, a novel adaptive-cost multi-fidelity Bayesian optimisation framework was introduced and a number of strategies of dealing with problems of this type were proposed. These strategies were tested on a synthetic test case - The Six-Hump Camelback Functions - and a real dataset, comprising simulations and experimental measurements of the NACA4415 aerofoil. It was found that multi-fidelity strategies tended to outperform the single-fidelity strategies when correlations were above 0.9, up to cost ratios that do not expend budget too quickly, something that was observed in [14]. The best treatments of this new cost function exist when exploration in the high-fidelity function need to be encouraged - which is common in many systems. Some limitations to the strategies were highlighted in the 'real data' test case. It was observed that, when data is sparse from a particular data source, some strategies can suffer from unfortunate initialisations. Problems can also arise when costs to vary a particular experimental parameter are much lower than the corresponding costs of other parameter changes. Systems of this kind can encourage the strategies to favour exploration only along a single dimension in the parameter space, leading to a lack of exploration of the rest of the domain, and thus, getting stuck in local minima. The overall effects were better performance overall for a single-fidelity system, and an increase in performance within a constrained budget for different cost-ratios for a multi-fidelity system.

REFERENCES

[1] Ira Herbert Abbott and Albert Edward Von Doenhoff. *Theory of Wing Sections, Including a Summary of Airfoil Data*. Dover Publications, January 1959.

[2] Mauricio A. Alvarez, Lorenzo Rosasco, and Neil D. Lawrence. Kernels for Vector-Valued Functions: A Review, April 2012.

[3] Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Advances in Neural Information Processing Systems 33*, 2020.

[4] Loïc Brevault, Mathieu Balesdent, and Ali Hebbal. Overview of Gaussian process based multi-fidelity techniques with variable relationship between fidelities, application to aerospace systems. *Aerospace Science and Technology*, 107:106339, 2020.

[5] David Duvenaud, James Robert Lloyd, Roger Grosse, Joshua B. Tenenbaum, and Zoubin Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. *30th International Conference on Machine Learning, ICML 2013*, 28(PART 3):2203–2211, 2013.

[6] Alexander I.J. Forrester, András Sóbester, and Andy J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3251–3269, 2007.

[7] Peter I Frazier. A tutorial on bayesian optimization. Technical report, US Air Force, 2018.

[8] Jacob R Gardner, Geoff Pleiss, David Bindel, Kilian Q Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, 2018.

[9] M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.

[10] Alonso Marco, Felix Berkenkamp, Philipp Hennig, Angela P. Schoellig, Andreas Krause, Stefan Schaal, and Sebastian Trimpe. Virtual vs. real: Trading off simulations and physical experiments in reinforcement learning with Bayesian optimization. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 1557–1563, 2017.

[11] Matthias Poloczek, Jialei Wang, and Peter I. Frazier. Multi-Information Source Optimization, November 2016.

[12] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Massachusetts, 2006.

[13] Kevin Swersky, Jasper Snoek, and Ryan P. Adams. Multi-task bayesian optimization. *Advances in Neural Information Processing Systems*, pages 1–9, 2013.

[14] David J. J. Toal. Some considerations regarding the use of multi-fidelity Kriging in the construction of surrogate models. *Structural and Multidisciplinary Optimization*, 51(6):1223–1245, June 2015.

[15] Zahra Zanjani Foumani, Mehdi Shishehbor, Amin Yousefpour, and Ramin Bostanabad. Multi-fidelity cost-aware Bayesian optimization. *Computer Methods in Applied Mechanics and Engineering*, 407:115937, March 2023.