Efficient Multi-Information Source Multi-Objective Bayesian Optimization

Danial Khatamsaz*, Lalith Peddareddygari†, Samuel Friedman‡, and Douglas Allaire§
J. Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, TX 77843

Multi-objective optimization is often a difficult task owing to the need to balance competing objectives. A typical approach to handling this is to estimate a Pareto frontier in objective space by identifying non-dominated points. This task is typically computationally demanding owing to the need to incorporate information of high enough fidelity to be trusted in design and decision-making processes. In this work, we present a multi-information source framework for enabling efficient multi-objective optimization. The framework allows for the exploitation of all available information and considers both potential improvement and cost. The framework includes ingredients of model fusion, expected hypervolume improvement and intermediate Gaussian process surrogates. The approach is demonstrated on two test problems and aerostructural wing design problem.

I. Introduction

When estimating a ground truth quantity of interest, for example, fuel burn for an aircraft or a particular material property, we can often consider different mathematical formulations of the analysis or prediction. This, in addition to experimental data and expert opinion, can give rise to the ability to use multiple different sources of information for the estimation task at hand. The different assumptions made lead to differing levels of fidelity among the sources, as well as different costs, both in terms of time and monetarily. In the presence of multiple sources of information, we seek analysis and design processes that exploit the extra information that would not be present if only a single source were available. The opportunity is efficient selection of which information source to query and where to query it on the basis of cost and potential for improvement in the estimation of a quantity of interest. The challenge is ensuring proper fusion of information as it becomes available and a need for a rapid capability for moving from prior predictive information to posterior predictive information without necessarily executing a true information source. While a multi-information source capability can be applicable to a wide variety of contexts, our focus here is on multi-objective optimization. We present a novel framework for exploiting available information sources to identify non-dominated points in objective space to estimate the true Pareto front of a given problem. To handle fusion, we incorporate model reification. To enable rapid assessment of posterior predictive information, we use Gaussian processes as intermediate surrogate models that may be temporarily updated with candidate query points. To drive candidate query points toward the Pareto front, we use the expected hypervolume improvement metric. Our approach is demonstrated on two test problems, which include a two-dimensional input space and a 30-dimensional input space. We then demonstrate our approach on an aerostructural wing design problem involving a 17-dimensional input space. These input space dimensions stress the limits of typical Gaussian process regression modeling and our approach is still shown to perform well. In each demonstration, we consider two objectives, but this is for presentation purposes only and is not a limitation of the work.

The rest of the paper is organized as follows. In Section II, a background on surrogate modeling, fusion, and multi-objective problems is provided. Then, in Section III, the approach to build the optimization algorithm is addressed, supported by pseudocode and a flowchart. Section IV then presents the results on test functions and the aerostructural design optimization problem. Conclusions are then drawn in Section V.

II. Background

Our multi-information source optimization approach for multiple objectives employs Gaussian processes as intermediate surrogate models and fuses information using the process of model reification. We describe each of

*Ph.D. Candidate, AIAA Student Member.
†Graduate Student, AIAA Student Member.
‡Ph.D. Candidate, AIAA Student Member.
§Assistant Professor, AIAA Senior Member, dallaire@tamu.edu.

these ingredients in turn in this section. We then conclude this section with background on a general multi-objective optimization formulation based on the Pareto frontier, which is how we approach such problems here. Once we have established the necessary ingredients of our approach, we move to a description of our formal hypervolume indicator based framework for multi-information source multi-objective optimization in Section III.

A. Gaussian Process Regression Surrogates

For the estimation of a quantity of interest (e.g., an objective or constraint), there are often available several information sources that may be employed. These encompass computational models, experiments, expert opinions, etc., and have varying fidelity (that also varies over the input space) and varying costs. Our approach here seeks to exploit optimally all available information sources by balancing the cost and fidelity of each information source when determining which source and where in the input space to query. Following Refs. [1, 2], we assume we have available some set of information sources, $f_i(\mathbf{x})$, where $i \in \{1, 2, \ldots, S\}$, that can be used to estimate the quantity of interest, $f(\mathbf{x})$, at design point $\mathbf{x}$. We note here that in the multi-objective context we address in this work, the output of the information sources can be vector-valued. To predict the output of each information source at input configurations that have not yet been executed, a surrogate model is constructed for each information source using Gaussian process regression [3]. These surrogates are denoted by $f_{GP,i}(\mathbf{x})$. We consider the prior distributions of the information sources modeled by Gaussian processes as

$$f_{GP,i}(\mathbf{x}) \sim \mathcal{GP}\left(\mathbf{0}, k_i(\mathbf{x}, \mathbf{x}')\right), \tag{1}$$

where $k_i(\mathbf{x}, \mathbf{x}')$ is a real-valued kernel function over the input space. For the kernel function, without loss of generality, we employ the squared exponential covariance function. This is specified as

$$k_i(\mathbf{x}, \mathbf{x}') = \sigma_s^2 \exp\left(-\sum_{h=1}^{d} \frac{(x_h - x_h')^2}{2l_h^2}\right), \tag{2}$$

where d is the dimension of the input space, σ_s^2 is the signal variance, and l_h, where $h = 1, 2, \ldots, d$, is the characteristic length-scale that indicates the correlation between the points within dimension h. The parameters σ_s^2 and l_h associated with each information source can be estimated by maximizing the log marginal likelihood. Assuming we have available N_i evaluations of information source i denoted by $\{\mathbf{X}_{N_i}, \mathbf{y}_{N_i}\}$, where $\mathbf{X}_{N_i} = (\mathbf{x}_{1,i}, \ldots, \mathbf{x}_{N_i,i})$ represents the N_i input samples to information source i and $\mathbf{y}_{N_i} = \left(f_i(\mathbf{x}_{1,i}), \ldots, f_i(\mathbf{x}_{N_i,i})\right)$ represents the corresponding outputs from information source i, the posterior distribution of information source i at design point $\mathbf{x}$ is given as

$$f_{GP,i}(\mathbf{x}) \mid \mathbf{X}_{N_i}, \mathbf{y}_{N_i} \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma_{GP,i}^2(\mathbf{x})\right), \tag{3}$$

where

$$\begin{aligned} \mu_i(\mathbf{x}) &= K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma_{n,i}^2 I]^{-1} \mathbf{y}_{N_i}, \\ \sigma_{GP,i}^2(\mathbf{x}) &= k_i(\mathbf{x}, \mathbf{x}) - K_i(\mathbf{X}_{N_i}, \mathbf{x})^T \\ &\qquad [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma_{n,i}^2 I]^{-1} K_i(\mathbf{X}_{N_i}, \mathbf{x}), \end{aligned} \tag{4}$$

where $K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i})$ is the $N_i \times N_i$ matrix whose m, n entry is $k_i(\mathbf{x}_{m,i}, \mathbf{x}_{n,i})$, and $K_i(\mathbf{X}_{N_i}, \mathbf{x})$ is the $N_i \times 1$ vector whose m^{th} entry is $k_i(\mathbf{x}_{m,i}, \mathbf{x})$ for information source i. Here we have included the term $\sigma_{n,i}^2$, which can be used to model observation error for information sources based on experiments.

For each information source surrogate we further quantify the uncertainty with respect to the ground truth quantity of interest by adding a term associated with the fidelity of the information source at a given location in input space. Specifically, we quantify the total variance, which captures both the variance associated with the Gaussian process representation and the quantified variance associated with the fidelity of the information source over the input space, as

$$\sigma_i^2(\mathbf{x}) = \sigma_{GP,i}^2(\mathbf{x}) + \sigma_{f,i}^2(\mathbf{x}), \tag{5}$$

where $\sigma_{f,i}^2(\mathbf{x})$ is the variance related to the fidelity of information source i that has been estimated from, for example, expert opinion or available real-world data. Here, we estimate this variance by computing the absolute difference between the available data from the true quantity of interest and the information source. A Gaussian process is then performed using the square of these error values as training points to estimate the fidelity variance over the input space as mean of the Gaussian process. This is described in more detail in Ref. [4].

B. Fusion of Information from Multiple Sources

Following Refs. [2, 4–6], we assume that every information source contains potentially useful information regarding a given quantity of interest. Given that this information may be correlated across information sources, we aim to accurately fuse available new information from each source query so as to ensure we use our resources as efficiently as possible. Our approach, unlike most traditional multifidelity approaches [7–14], does not assume a hierarchy of information sources and our goal is not optimization with the highest fidelity source but optimization with respect to ground truth.

Several approaches exist for fusing multiple sources of information. Among these are approaches such as Bayesian modeling averaging [15–20], the use of adjustment factors [21–24], covariance intersection methods [25, 26], and fusion under known correlation [27–29]. As noted previously, we assume that every information source contains useful information regarding the ground truth quantity of interest. Thus, as more information sources are incorporated into a fusion process, we expect the variance of quantity of interest estimates to decrease. This is not necessarily the case for all of the aforementioned fusion techniques with the exception of fusion under known correlation. Thus, there is significant value in determining correlations prior to fusion.

Following the work of Refs. [2, 4–6], we note that since our information sources are represented by intermediate Gaussian processes, their fusion follows that of normally distributed information. Under the case of known correlations between the discrepancies of information sources, the fused mean and variance are shown to be [29]

$$\mathbb{E}[\hat{f}(\mathbf{x})] = \frac{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \boldsymbol{\mu}(\mathbf{x})}{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \tag{6}$$

$$\mathrm{Var}\left(\hat{f}(\mathbf{x})\right) = \frac{1}{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \tag{7}$$

where $\mathbf{e} = [1, \ldots, 1]^T$, $\boldsymbol{\mu}(\mathbf{x}) = [\mu_1(\mathbf{x}), \ldots, \mu_S(\mathbf{x})]^T$ given S models, and $\tilde{\Sigma}(\mathbf{x})^{-1}$ is the inverse of the covariance matrix between the information sources.

To estimate the correlation coefficients between information sources over the domain we use the reification process defined in [5, 6]. In this process, to estimate the correlation coefficients between the deviations of information sources i and j, each of the information sources i and j, one at a time, is reified, or, treated as ground truth. Assuming that information source i is reified, the correlation coefficients between the information sources i and j, for $j = 1, \ldots, i-1, i+1, \ldots, S$, are given as

$$\rho_{ij}(\mathbf{x}) = \frac{\sigma_i^2(\mathbf{x})}{\sigma_i(\mathbf{x})\sigma_j(\mathbf{x})} = \frac{\sigma_i(\mathbf{x})}{\sqrt{\left(\mu_i(\mathbf{x}) - \mu_j(\mathbf{x})\right)^2 + \sigma_i^2(\mathbf{x})}}, \tag{8}$$

where $\mu_i(\mathbf{x})$ and $\mu_j(\mathbf{x})$ are the mean values of the Gaussian processes of information sources i and j respectively, at $\mathbf{x}$, and $\sigma_i^2(\mathbf{x})$ and $\sigma_j^2(\mathbf{x})$ are the total variances at $\mathbf{x}$. Next, information source j is reified to estimate $\rho_{ji}(\mathbf{x})$. The variance weighted average of the two estimated correlation coefficients can then be used as the estimate of the correlation between the errors as

$$\bar{\rho}_{ij}(\mathbf{x}) = \frac{\sigma_j^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})}\rho_{ij}(\mathbf{x}) + \frac{\sigma_i^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})}\rho_{ji}(\mathbf{x}). \tag{9}$$

These average correlations are then used to estimate the fused mean and variance in Eqs. 6 and 7.

C. Multi-Objective Optimization

A multi-objective optimization problem can be defined as

$$\text{minimize} \;\; \{f_1(x), ..., f_n(x)\}, x \in \mathcal{X} \tag{10}$$

where $f_1(x), \ldots, f_n(x)$ are the objectives and $\mathcal{X}$ is the feasible design space. Throughout this work, we develop unconstrained approaches, however, the inclusion of penalty terms could be considered for constraint handling. For problems such as (10), it is usually the case that no single point optimizes each individual objective simultaneously. To deal with this, approaches based on the creation of a scalar objective using utility theory are common, as well as approaches based on finding non-dominated solutions approaching the Pareto frontier. We focus on the latter here. For this case, optimal solutions, $\mathbf{y}$, to a multi-objective problem with n objectives are denoted as $\mathbf{y} \prec \mathbf{y}'$, and are defined as

$$\{\mathbf{y} : \mathbf{y} = (y_1, y_2, \ldots, y_n),\ y_i \leq y_i' \;\; \forall\, i \in \{1, 2, \ldots, n\} \;\text{ and }\; \exists\, j \in \{1, 2, \ldots, n\} : y_j < y_j'\}, \tag{11}$$

where $\mathbf{y}' = (y_1', y_2', \ldots, y_n')$ denotes any possible objective output. The set of $\mathbf{y} \in \mathcal{Y}$, where $\mathcal{Y}$ is the objective space, is the Pareto front of the problem. This is shown conceptually for a biobjective problem in Figure 1. All points on the Pareto front are non-dominated. Our approach, which is common in the literature, is thus to find the Pareto front as efficiently as possible for a given multi-objective optimization problem.

There are many techniques in use for approximating Pareto frontiers for multi-objective optimization problems. Among these are the weighted sum approach [30], the adaptive weighted sum approach [31], normal boundary intersection methods [32], hypervolume indicator methods [33–39], and others. The hypervolume indicator approach is well-suited to expected improvement based algorithms, which have been shown to work well in a multiple information source setting (see, e.g., Refs. [1, 2, 40, 41]). Thus, our approach proposes the incorporation of hypervolume indicator improvement within a multi-information source querying framework. Hypervolume indicator approaches are based on the concept of a hypervolume in objective space. These hypervolumes are measured relative to a fixed reference point and the enclosed volume between the approximated set of Pareto points and the reference point is computed. The concept is shown notionally in Figure 2. Here, the shaded area is the hypervolume to be computed. In general, if a given set of points has a higher hypervolume than another set, then the given set is a better estimator of the Pareto front. Hypervolume indicator algorithms seek to maximize the hypervolume in objective space so as to best approximate the Pareto front. Thus, the value of new query points can be estimated (using prior predictive distributions if using Gaussian processes) by measuring the expected improvement in the hypervolume that would occur given the query takes place. In Figure 2, the shaded area shows the amount of increase in hypervolume when a new non-dominated point is found and added to the solution set.

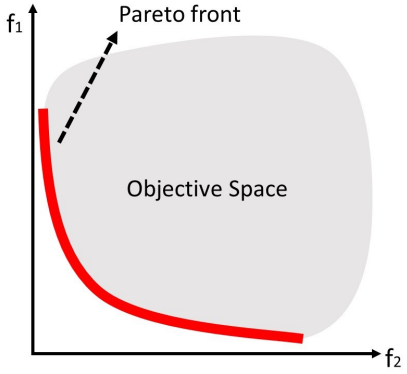


Figure 1. The Pareto front (red curve) of a two objective minimization problem. All points on the red line are non-dominated and constitute the solution set.

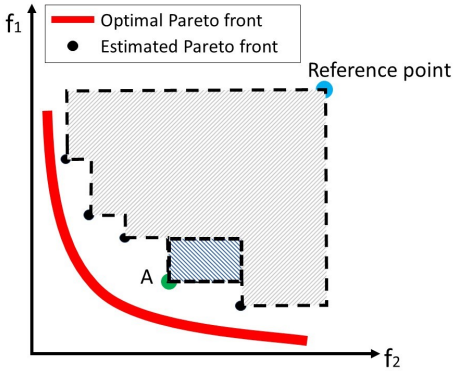


Figure 2. Hypervolume of the estimated Pareto front. When point A is added to the solution set, the blue shaded region corresponds to the resulting hypervolume improvement.

III. Approach

Generally, our approach is based on determining, with available prior information, where to query and what source to query to maximize the hypervolume indicator while being budget aware. To achieve this, we make use of the updateable Gaussian process surrogates described previously for each information source. These surrogates can be used as prior predictive distributions that can be temporarily updated with potential query locations that result in potential changes to the hypervolume indicator. By searching over the space of potential query locations and potential information sources with these prior predictive surrogates, we are able to efficiently identify the next best query to execute. Once this query is executed, all surrogates (including correlation information) may be updated, and then can serve as prior predictive distributions for the next iteration. In this section, we describe in detail our approach to achieving this. We begin with necessary preliminaries regarding the fast calculation of the expected hypervolume improvement (EHVI) [42] within a multi-information source framework. This discussion follows largely from Ref. [43] where more details can be found if desired. We then describe our algorithm for multi-information source multi-objective optimization. In Section IV we demonstrate the use of this framework on test problems and an aircraft wing design problem.

A. Preliminaries

Following Ref. [43] for the development of the fast computation of EHVI, we present here our implementation within a multi-information source setting. We begin by considering a current solution set, $\mathcal{S}$, of non-dominated points in objective space at some point during a multi-objective optimization process. The dominated hypervolume, denoted as $\mathcal{H}(\mathcal{S})$,can then be computed given $\mathcal{S}$ and a reference point. Improvement to the hypervolume due to adding a new solution vector $\mathbf{y}$ is then defined as

$$\mathcal{H}_I(\mathbf{y}, \mathcal{S}) = \mathcal{H}(\mathcal{S} \cup \mathbf{y}) - \mathcal{H}(\mathcal{S}). \tag{12}$$

If $\mathcal{H}_I(\mathbf{y}, \mathcal{S}) > 0$, then $\mathbf{y}$ is in the non-dominated region of $\mathcal{S}$ and can be used to update the solution set. Otherwise, there is no improvement over $\mathcal{H}(\mathcal{S})$ by adding $\mathbf{y}$ and the query adds no value. In the context of Bayesian optimization, $\mathbf{y}$ is a random output of a probabilistic model related to a potential solution in the design space. Hence, $\mathcal{H}_I(\mathbf{y}, \mathcal{S})$ is also a random variable. Therefore, it is possible to calculate its expected value, which is the EHVI. Comparing EHVI values for different potential solutions in the design space and finding the maximum EHVI leads to an information-economic querying policy that ensures maximum gains are achieved in each successive query. In a multi-information source context, however, the different cost of querying each source should also be taken into account.

The formula for calculating EHVI as outlined in [44] is given as

$$\mathbb{E}[\mathcal{H}_I(\mathbf{y})] = \int_U \mathbb{P}(\{\mathbf{y} \prec \mathbf{y}'\}) d\mathbf{y}' \tag{13}$$

where $\mathbb{P}(\mathbf{y} \prec \mathbf{y}')$ is the probability that $\mathbf{y}'$ is dominating y and U is the dominated hypervolume. In our context, this can be computed in closed-form as will be shown below. Given that we have independent Gaussian process models for every objective for each information source, the posterior predictive output of each model given the data is a random variable identified as $y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ where $i \leq m$ and μ_i, σ_i^2 are the mean and covariance of the i^{th} objective accordingly (note, we have not included information source specific indices here for notational clarity). For a new potential solution in the design space, we have the following equation:

$$\mathbb{P}(\mathbf{y} \prec \mathbf{y}') = \prod_{i=1}^{n} \Phi\left(\frac{y_i' - \mu_i}{\sigma_i}\right) \tag{14}$$

where Φ is the cumulative distribution function of the standard normal random variable. The next step in the computation process is the decomposition of the dominated space, U, into high-dimensional boxes where we can then integrate $\mathbb{P}(\{\mathbf{y} \prec \mathbf{y}'\})$ in each box. Indicating the lower and upper bounds of each box as $\mathbf{l} = (l_1, l_2, ...l_m)$ and $\mathbf{u} = (u_1, u_2, ..., u_m)$ respectively and the high-dimensional box as $\mathcal{B}(\mathbf{l}, \mathbf{u}) = (l_1, u_1] \times (l_2, u_2] \times ... \times (l_m, u_m]$, the integral of $\mathbb{P}(\{\mathbf{y} \prec \mathbf{y}'\})$ inside $\mathcal{B}(\mathbf{l}, \mathbf{u})$ is:

$$\delta(\mathcal{B}(\mathbf{l}, \mathbf{u})) = \int_{\mathcal{B}(\mathbf{l}, \mathbf{u})} \prod_{i=1}^{m} \Phi\left(\frac{y_i' - \mu_i}{\sigma_i}\right) dy' = \prod_{i=1}^{m} \int_{l_i}^{u_i} \Phi\left(\frac{y_i' - \mu_i}{\sigma_i}\right) dy_i'. \tag{15}$$

Using the fact that the integral of $\Phi(a + bx)$ can be written as $\int_{-\infty}^{t} \Phi(a + bx)dx = \frac{1}{b}[(a + bt)\Phi(a + bt) + \phi(a + bt)]$ [45], Equation 15 can be written:

$$\delta(\mathcal{B}(\mathbf{l}, \mathbf{u})) = \prod_{i=1}^{m} [\psi(u_i, \mu_i, \sigma_i) - \psi(l_i, \mu_i, \sigma_i)], \tag{16}$$

where

$$\psi(a, \mu, \sigma) = \int_{-\infty}^{a} \Phi\left(\frac{y - \mu}{\sigma}\right) dy = (a - \mu)\Phi\left(\frac{a - \mu}{\sigma}\right) + \sigma\phi\left(\frac{a - \mu}{\sigma}\right). \tag{17}$$

In Figure 3, a box with the upper and lower bounds is illustrated in the non-dominated region of a solution set consisting of four vectors. Two important points here are first, any decomposition methods can be used in the calculation of Equation 16 as the boxes do not need to have a certain structure and second, the integral in each box is independent from the integral in the other boxes. There are of course different ways to decompose a region into hyperrectangles. Here we choose the Walking Fish Group (WFG) approach [46] since it is known to be a fast algorithm for computing hypervolume [47]. To compute the hypervolume of a non-dominated set, $\mathcal{S}$, WFG uses the following equation:

$$\mathcal{H}(\mathcal{S}) = \sum_{i=1}^{n} \mathcal{H}_I(y_i, \{y_{i+1}, ..., y_n\}) \tag{18}$$

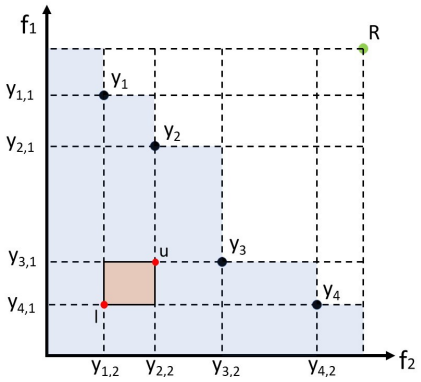


Figure 3. A 2D $\mathcal{B}(\mathbf{l}, \mathbf{u})$ in the non-dominated region is indicated [adapted from [43]].

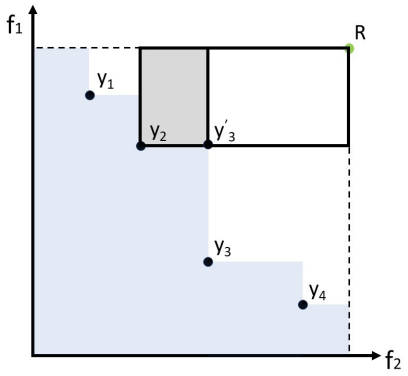


Figure 4. The second step of WFG decomposition method adapted from [43].

where n is the number of vectors in the solution set, $\mathcal{S}$, and $\mathcal{H}_I$ is the hypervolume improvement defined in Equation 12. To avoid calculating $\mathcal{H}(\mathcal{S})$ considering that $\mathcal{H}_I(y_1, \{y_2, ...y_n\}) = \mathcal{H}(\mathcal{S}) - \mathcal{H}(\{y_2, ...y_n\})$, the WFG method decomposes the hypervolume into a series of boxes using $\mathcal{H}_I(y_i, \{y'_{i+1}, ..., y'_n\})$. In Figure 4, the second step of the calculation by the WFG method is shown. The point y'_3 is determined by

$$y'_3 = \text{limit}(y_2, y_3) = (\max(y_{2_1}, y_{3_1}), ..., \max(y_{2_m}, y_{3_m})) \tag{19}$$

As a result, the hypervolume is calculated by summation and differences of a series of boxes constructed with the WFG method. The final step is to compute the EHVI. In summary, we decompose the dominated hypervolume into a series of boxes and calculate the Equation (15) for all boxes. This is then subtracted from the integral of the whole region represented by the box $\mathcal{B}(-\infty, R)$.

B. Multi-information Source Multi-Objective Optimization Framework

Using the Gaussian process as the surrogate model for each objective and EHVI as the acquisition function, we can perform Bayesian optimization to approximate a solution set for a multi-objective optimization problem. It is necessary to notice that the model discrepancies are changing whenever new information is found about the ground truth by querying the information sources. The model discrepancy is defined as the difference between the predicted value of the model built with data from an information source and the model built with the available data from ground truth for a specific design space point. Therefore, model discrepancies should be updated regularly. However, querying the ground truth to update its model is costly. Thus, we need define a condition for when to query the ground truth. Such a condition can be, for example, when a certain number of updates have been made to available information sources, or also when a specific amount of the total allotted budget is spent. The later, is the criterion used in our algorithm and the ground truth is queried each time 10 percent of the budget is spent. This method allows the decision maker to query cheap information source more between ground truth queries if it finds the cheap information source is still providing useful knowledge about the ground truth. This is inline with expected intuition regarding the exploitation of cheap information sources given their nearly negligible cost in comparison to expensive sources and ground truth itself.

Algorithm 1 presents our overall framework. Our procedure to optimize a multi-objective function is established assuming the function has m objectives, and there are n information sources of differing fidelity available to provide information about the ground truth. The first step involves the construction of Gaussian processes (GPs) and to create the initial Pareto front. The can be established by finding non-dominated design points of initial data available from the ground truth. Next, the fusion step takes place, which involves the previously described model reification process. This is followed by the generation of candidate query points which are tested for EHVI potential given the current set of GPs. The best candidate (query point and information source) is selected and executed. This is followed by another fusion step given the new information. The budget condition is then checked, which would lead to a ground truth query or a check on whether the budget is exhausted. If the budget is exhausted, the process terminates with a final analysis of the estimated Pareto front from the fused GP, which leads to subsequent evaluations of best points from the ground truth. If the budget is not exhausted, the process resamples candidate points and repeats. If the budget condition is met, leading to a ground truth query, then ground truth is queried, its GP is updated, and then all other GPs are updated (owing to a

Algorithm 1 : Multi-Objective Bayesian Optimization

construct GP_1 to GP_m given available data from the ground truth
for i from 1 to n **do**
  for j from 1 to m **do**
    construct $GP_{j,i}$ for objective (j) of the information source (i) given the data
  end for
end for
fuse models and construct the initial Pareto front
while available budget > 0 **do**
  X-sample set ⟵ Latin Hypercube Sampling
  for k from 1 to n **do**
    for s from 1 to size(X-sample set) **do**
      Y-sample ⟵ query X-samples from GP_{1k} to GP_{mk}
      construct temporary GPs by updating GP_{1k} to GP_{mk} using sample s
      updated_fused_values ⟵ fuse other models with the updated one
      generate test_samples using fused_means and fused_variances
      improvement(s,k) ⟵ **EHVI**(test_samples,updated_fused_values,Pareto front)
    end for
  end for
  X = sample to be queried , V = Information Source(IS) to query from ⟵ Max(improvement)
  $Y = (y_1, ...y_m) = \mathrm{IS}_{(V)}(X)$
  update $GP_{1,V}$ to $GP_{m,V}$ using X and Y
  fuse models
  U ⟵ query a randomly generated set of design points from fused model
  find non-dominated vectors in U to update Pareto front
  if requirements to query ground truth is met **then**
    G ⟵ a set of design points with arbitrary size distributed along the Pareto front
    Y_G ⟵ query set G from the ground truth
    update GP_1 to GP_m
    update model discrepancies
    fuse models and update the Pareto front
  end if
end while
fuse models to construct fusedGP_1 to fusedGP_m
$\mathcal{U}$ ⟵ query a randomly generated set of design points from fused models
$\mathcal{S}$ ⟵ find non-dominated vectors in $\mathcal{U}$ to update the Pareto front
$\mathcal{X}$ ⟵ the design space corresponding to non-dominated set $\mathcal{S}$

change in the discrepancies and correlations given new ground truth information). The budget exhaustion condition is then checked and the process proceeds as previously described from this point. A complete flow chart of this process is provided in Figure 5.

When the decision to query ground truth is made based on the budget condition, a certain number of points, N, are considered. Although choosing higher values of N results in more information gain and higher accuracy to estimate the model discrepancies, it is costly. Hence, a trade-off should be considered in assigning a value to N. For the purposes of the demonstration cases that follow in Section IV, we have set $N = 10$, however, the study of this parameter is a topic of future work. Algorithm 2 presents our ground truth querying strategy.

IV. Application and Results

To evaluate the performance of our proposed algorithm, we have applied it on two different test functions from Ref. [48]. The first one is Poloni's two objective test function that maps points from a two dimensional design space to the two dimensional objective space. The second test function is a two objective optimization problem referred to as DTLZ2 that can be scaled to any number of dimensions in the design space. A comparison between the optimal Pareto

Algorithm 2 : Querying the Ground Truth

1: divide the most updated Pareto front into N slices
2: construct a smooth Gaussian process for each slide given data points in the slice
3: $P \longleftarrow$ choose the closest point to the Gaussian process mean in each slice
4: $P_Y \longleftarrow$ query the design space point corresponding to P from the ground truth

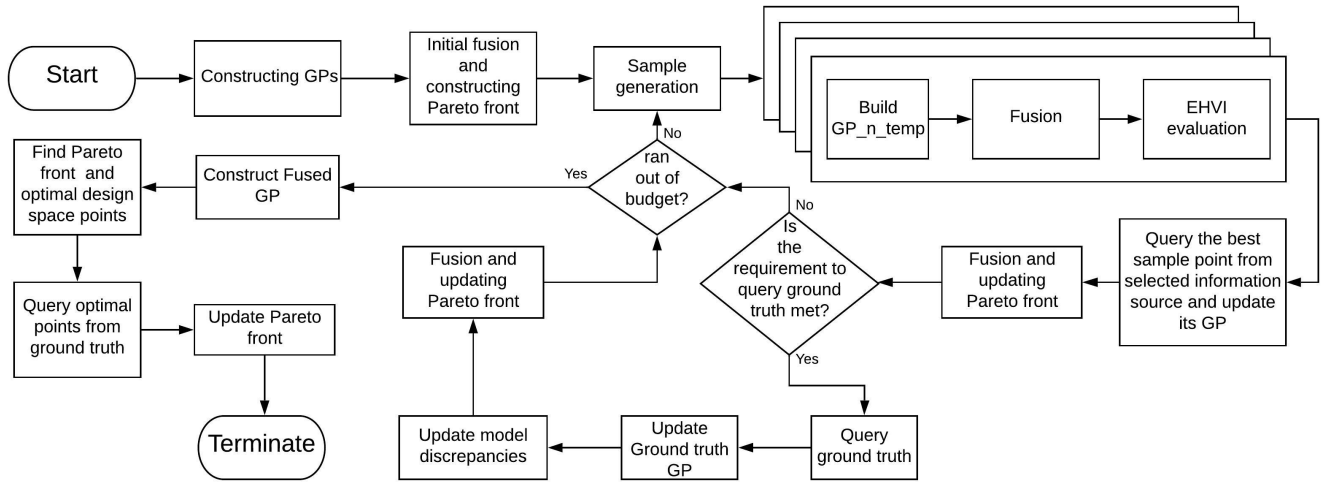


Figure 5. Our proposed multi-information source Bayesian optimization approach for multi-objective optimization using Gaussian processes as the surrogate models and EHVI as the criterion to guide the search toward promising design space regions. The requirement to query the ground truth is set by user. After spending available resources, the most updated Pareto front is taken as the final estimation of the true Pareto front.

front associated to each problem and approximated Pareto front is made to show the effectiveness of the algorithm. To apply the concept of multi-fidelity approach in optimizing the test functions, we have constructed 2 other functions close to each test function by changing the coefficients and constants. The test functions themselves are considered ground truth. We follow the demonstration of our framework on the two test functions with its application to an aircraft wing design problem using OpenAeroStruct [49]. We describe the software and the problem for this application in Section IVC.

A. Poloni's test function

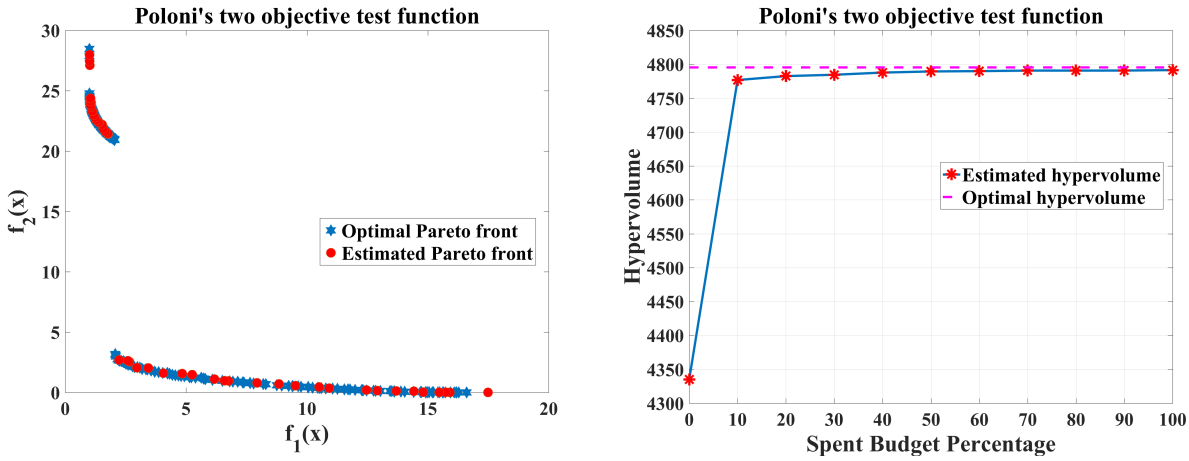
Poloni's two objective test function is defined as:

$$\begin{aligned} &\text{minimize}: \quad f_1(\mathbf{x}) = 1 + (A_1 - B_1(\mathbf{x}))^2 + (A_2 - B_2(\mathbf{x}))^2 \\ &\text{minimize}: \quad f_2(\mathbf{x}) = (x_1 + 3)^2 + (x_2 + 1)^2 \end{aligned}$$

where

$$\begin{aligned} &-\pi \leq x_1, x_2 \leq \pi \\ &A_1 = 0.5\sin(1) - 2\cos(1) + \sin(2) - 1.5\cos(2) \\ &A_2 = 1.5\sin(1) - \cos(1) + 2\sin(2) - 0.5\cos(2) \\ &B_1(\mathbf{x}) = 0.5\sin(x_1) - 2\cos(x_1) + \sin(x_2) - 1.5\cos(x_2) \\ &B_2(\mathbf{x}) = 1.5\sin(x_1) - \cos(x_1) + 2\sin(x_2) - 0.5\cos(x_2). \end{aligned}$$

Figure 6a shows the optimal Pareto front versus estimation of the Pareto front at the end of our optimization algorithm. The estimated Pareto front is found using the knowledge of lower fidelity models about the ground truth. Figure 6b is showing the hypervolume is increasing as a result of improved estimation of the Pareto front admitting that the budget is spent effectively. Hypervolume is initially computed using available data from the ground truth. In Figure 6b, a



(a) The true and estimated Pareto front.

(b) Computed hypervolume during the optimization process.

Figure 6. (a) True and estimated Pareto front. Our approach has found the optimal solutions with a good accuracy for Poloni's two objective test function. (b) The hypervolume during the optimization process as a function of spent budget. The dashed line shows the optimal hypervolume with respect to the reference point (70,70).

significant improvement is achieved before spending the first 10% of the budget. This was achieved largely through the exploitation of the cheaper lower fidelity model. We present a similar result and the information source query history for the OpenAeroStruct demonstration in Section IVC. The diminishing returns in hypervolume are expected given that finding new non-dominated points becomes more difficult as more points are found.

B. DTLZ2 test function

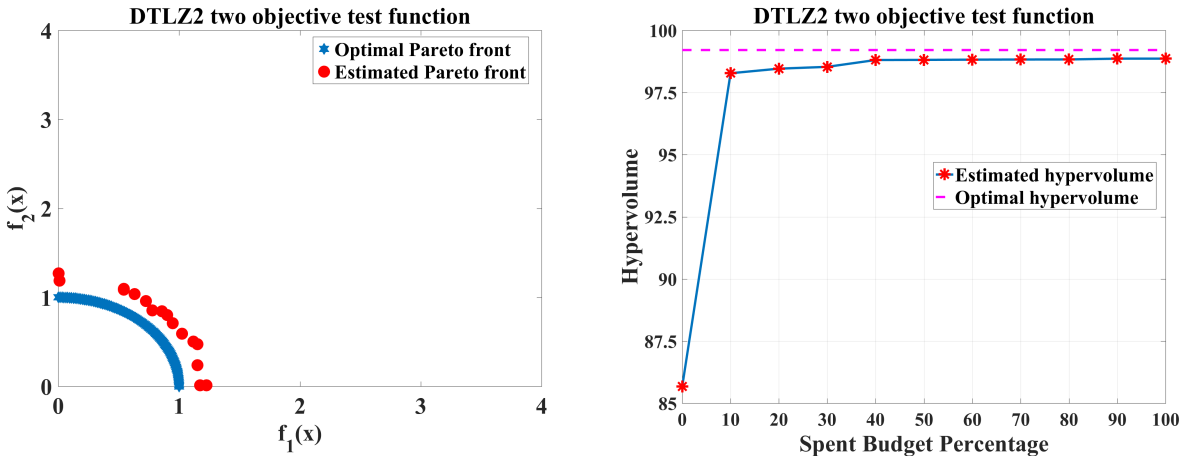
The DTLZ test function suite [48] can be extended to any number of objectives and any number of inputs. To make use of DTLZ test functions to evaluate our optimization algorithm, a two objective DTLZ2 test function is created with a 30-dimensional design space. This test function is harder to optimize than previous test function since the Gaussian processes made for the information sources needs to deal with a 30-dimensional input space and there is not enough resources to evaluate many points. Consequently, it tries to get as close as possible to the optimal Pareto front with a very limited budget. The DTLZ2 problem with two objectives used here is defined as

$$\text{minimize}: \quad f_1(\mathbf{x}) = (1 + g(\mathbf{xm})) \cos\left(x_1 \frac{\pi}{2}\right)$$
$$\text{minimize}: \quad f_2(\mathbf{x}) = (1 + g(\mathbf{xm})) \sin\left(x_1 \frac{\pi}{2}\right)$$

where

$$0 \leq x_1, x_2 \leq 1, \quad \text{and}$$
$$g(\mathbf{xm}) = \sum_{i=1}^{k} (xm_i - 0.5)^2$$

The variable k is found by the formula $k = n + m - 1$ where n is the size of its design space, m is the number of objectives, and, $\mathbf{xm}$ is a vector created by the last k elements of $\mathbf{x}$. The optimal Pareto front for the DTLZ2 problem is found when $\{xm_i = 0.5,\ 1 \leq i \leq k\}$. For this particular test function, optimization is not easy due to special characteristics of DTLZ2 problem. For example, Latin hypercube sampling is not able to produce useful samples since it divides the range from 0 to 1 into 30 segments and pick a value out of each segment. Thus, it can never find a point in the design space close to the optimal solutions. Therefore, for this problem, all samples are generated from uniform random variables. Figure 7a shows how close the estimated Pareto front is to the true Pareto front. In Figure 7b, it is illustrated that the hypervolume is not increasing significantly anymore and that the algorithm has again achieved significant early gains in hypervolume improvement. We note again that this problem is extremely challenging owing to its high dimensionality in the input space. Given this, our approach performed well. We will compare these results to algorithms such as ParEgo and NSGA-II in future work, but we point out here that these algorithms are not set up to take advantage of



(a) The true and estimated Pareto front.

(b) Computed hypervolume over optimization process.

Figure 7. (a) The true and estimated Pareto front are illustrated for the DTLZ2 problem. (b) he hypervolume during the optimization process as a function of spent budget. The dashed line shows the optimal hypervolume with respect to the reference point (10,10).

multiple information sources and hence, would either never find the Pareto front (owing to the use of non-ground truth) or would take a prohibitively long time if ground truth is directly queried. In either case, it is expected that the approach proposed here would significantly outperform these other options.

C. OpenAeroStruct Demonstration

OpenAeroStruct is an open-source software developed in NASA's OpenMDAO framework [50], which can be used for fast tightly coupled aerostructural design optimization. The framework implements the coupled adjoint method to compute the aerostructural derivatives used for efficient gradient-based optimization. As noted in Ref. [49], OpenAeroStruct combines a vortex lattice method (VLM) and 1-D finite element analysis using six degrees of freedom 3-D spatial beam elements to model lifting surfaces [49, 51]. A common aerostructural single objective optimization problem is the fuel burn minimization problem using the Breguet range equation. Structural mass minimization of the wing is also frequently considered and thus is used as a second objective in demonstrating our proposed multi-information source multi-objective optimization framework.

The OpenAeroStruct application, as described in Ref. [49], uses the Breguet range equation to compute the fuel burn as a function of structural weight and aerodynamic performance. Design variables consist of twist distributions, spar thickness distributions, and planform variables such as skin thickness, thickness over cord ratio, and angle of attack. The first four variables are 4-dimensional as 4 control surfaces are considered for the wing. Hence, the problem has a 17-dimensional design space. Constraints in the standard problem ensure lift equals weight, and that structural failure does not occur.

The mesh in the OpenAeroStruct is defined by the number of the spanwise and chordwise points. The fidelity of each model depends on the number of points used to define the lifting surface. A model with a finer mesh is considered to have higher fidelity compared to a model with a coarser mesh. We use three different mesh resolutions in this demonstration to serve as three different multifidelity information sources. These meshes are shown in Table 1, where

Fidelity level	Num_y	Num_x	Cost
Low	15	3	1
High	35	11	23.69
Ground truth	55	19	149.4

Table 1. Mesh sizes and costs for different fidelity models

Num_y is the number of span-wise points and Num_x is the number of chord-wise points. In Figure 9, the three different

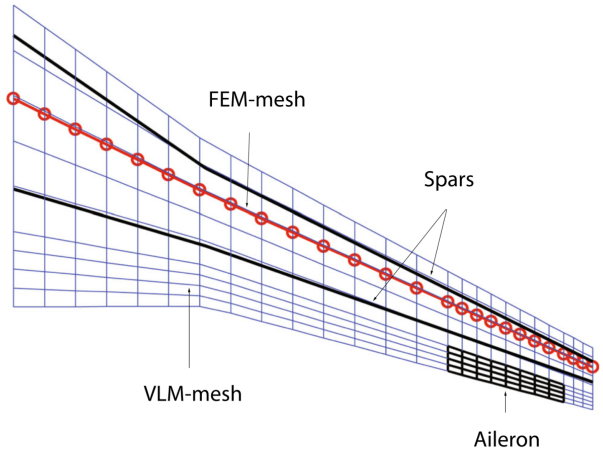
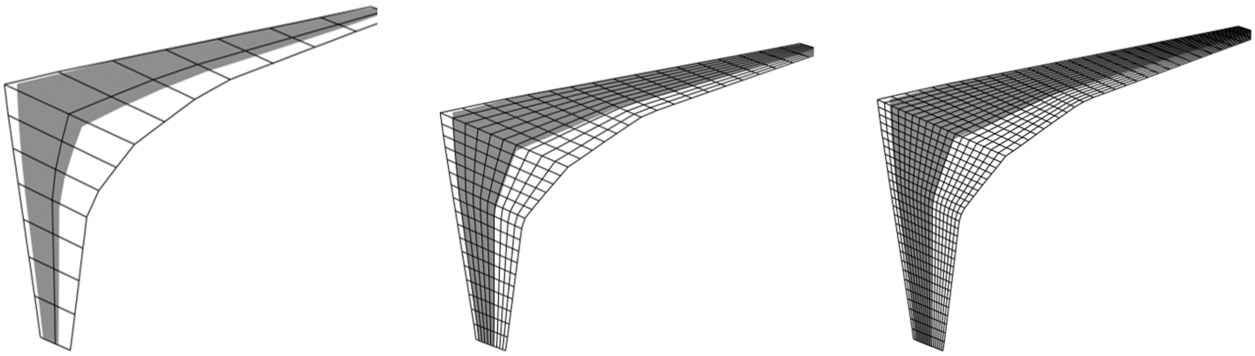


Figure 8. A wing with the aerodynamic and structural Meshes [52].

meshes are shown. The cost of querying each model is computed by comparing the running times a single query. The low fidelity model's cost is considered as the unit cost and other model's costs are found relatively.



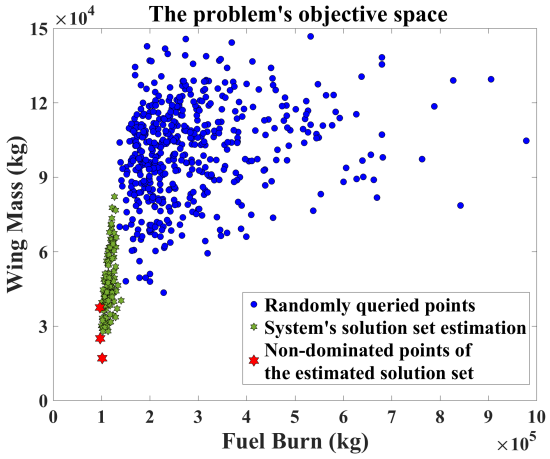
(a) Low fidelity model. **(b)** High fidelity model. **(c)** The ground truth.

Figure 9. Illustration of the meshed wing with different fidelity models. The number of meshes in each models is presented in table 1

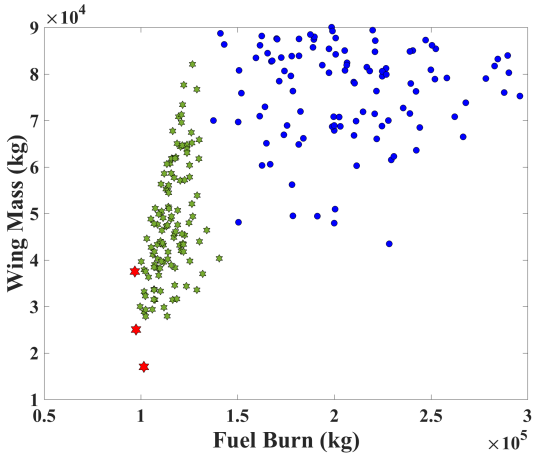
We applied our approach to multi-information source multi-objective optimization to this two objective OpenAeroStruct problem with a ground truth and two other information sources. We assumed a budget of 1000 units for this demonstration. The results are shown in Figure 10, where random points are shown in blue to show the objective space (these are not part of the algorithm and are for visualization only), the green points are those points selected by our approach with the fused GP, and the red points are the non-dominated green points that have been evaluated with the ground truth (that is, the final step of our algorithm). The figure reveals that our approach has done well in identifying the non-dominated region in the objective space for this 17 dimensional problem.

In Figure 11a, the hypervolume is computed during the optimization process each time 10% of the budget is spent. The reference point is fixed as $(10^6, 10^7)$ to make sure it is dominated by all points in the objective space. The most significant change in the hypervolume quantity is made when spending the first 10% of the budget (as was seen in the previous test cases as well). Beyond that, there are improvements in the hypervolume but the returns are diminishing as expected.

Figure 11b shows the hypervolume as function of iteration, where iteration is a query to any information source. The points on the plot denote queries to ground truth. The larger distance between these points indicates more low fidelity queries, since they are cheaper. We see that early in the process then, our approach is querying the cheaper information source. This matches the intuition that the cheap source, if cheap enough, should be queried nearly exhaustively in promising regions while there is still value to do so. In this demonstration, when the value of the lower fidelity source is decreased enough the high fidelity information source is queried more. This results in spending budget faster and hence,

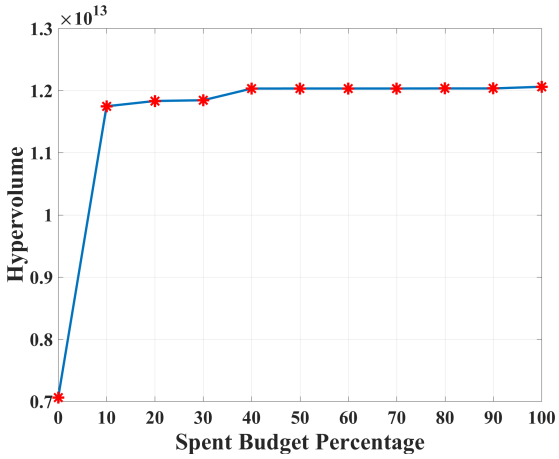


(a) OpenAeroStruct objective space with random points, fused GP points, and the estimated Pareto front.

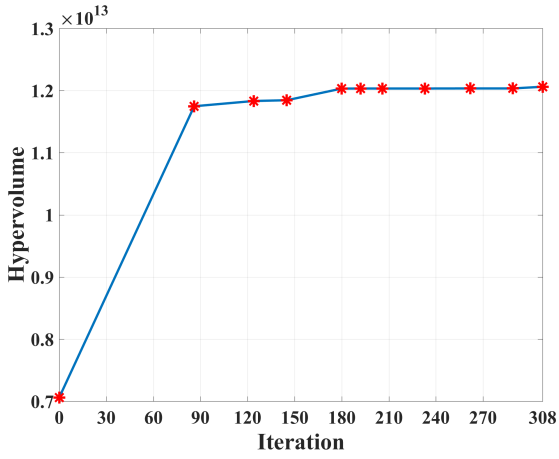


(b) Zooming in on the suggested region by the fused model.

Figure 10. Illustration of the estimated Pareto front (red stars) and the non-dominated points (green stars) found by the fused model. Blue circles are found by randomly querying the ground truth. The right plot is a zoomed in version of the left plot.



(a)



(b)

Figure 11. (a) Hypervolume is updated when the ground truth is queried. After spending 40% of the budget, the hypervolume is not increasing significantly anymore and the system has found the best estimation of the optimal solution set. (b) The updated hypervolume is presented with respect to the number of iterations. A larger number of iteration between the two steps of querying the ground truth shows that the cheaper information source is queried more. When the high fidelity information source is involved, there will be smaller number of iterations between two steps of querying the ground truth.

closer points on the plot in Figure 11b.

Figure 12 reveals the cumulative sum of queries from any information source plotted against overall iteration, where iteration is defined as it was previously. As depicted in the figure, initially, the low fidelity information source has been queried much more than the high fidelity information source. After some number of iterations (about 120 in this case), the value of the low fidelity source has diminished enough that some queries to the higher fidelity source are now necessary. This continues until the budget is spent. We see in this case that the low fidelity source continues to be queried as well. This is due to the fact that as higher fidelity information is obtained, the correlations between the low and high fidelity sources are updated, which results in possible renewed value in lower fidelity information. This was the case here.

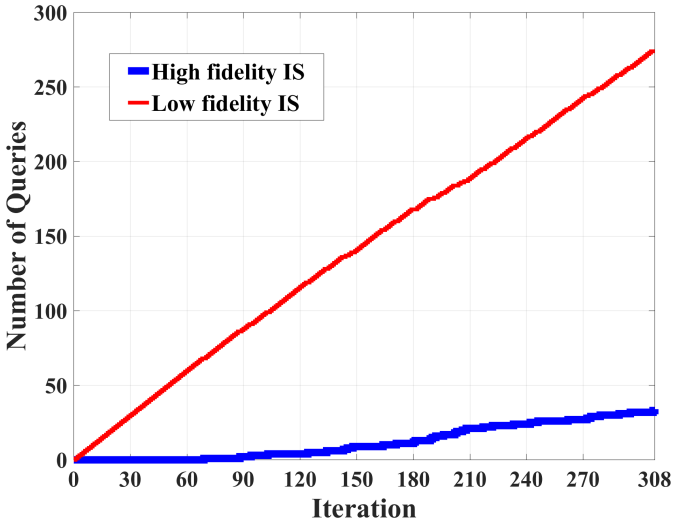


Figure 12. The step plot shows the cumulative sum of queries from each information source. When one of the information sources is queried, the corresponding marker is raised one step in the figure while the other one does not change. Consequently, the number of iteration is the sum of queries from all information sources as we query one information source at each iteration.

V. Conclusion

In this paper, a methodology to optimize expensive multi-objective functions is presented. The methodology seeks to exploit all available information sources for efficiently identifying non-dominated points in the objective space as a means of estimating the true Pareto front. The approach was based on the fast evaluation of the expected hypervolume improvement through the use of temporarily updated Gaussian process surrogate models of each information source. The process also incorporates model reification to fuse new information rigorously as it becomes available through proper accounting for correlation between the sources. The conclusion of this study is that multi-information source Bayesian optimization approaches to directing efficient querying when budget constrained can be an effective way of estimating the Pareto front of a multi-objective problem. In particular, the ability to rapidly query lower fidelity sources while accounting for their correlation with higher fidelity sources and ground truth has enabled efficient (less than 10% of budget for the problems studied here) identification of promising regions for non-dominated point searching. In future work, we will compare our approach to state-of-the-art multi-objective approaches, though in general, these are designed only with a single information source (often ground truth or the highest possible fidelity model) in mind.

Acknowledgments

This work was supported by the AFOSR MURI on multi-information sources of multi-physics systems under Award Number FA9550-15-1-0038, program manager, Dr. Fariba Fahroo and by the National Science Foundation under grant no. CMMI-1663130. Opinions expressed in this paper are of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Ghoreishi, S. F., and Allaire, D., "Multi-information source constrained Bayesian optimization," *Structural and Multidisciplinary Optimization*, Vol. 59, No. 3, 2019, pp. 977–991.

[2] Ghoreishi, S. F., Molkeri, A., Srivastava, A., Arroyave, R., and Allaire, D., "Multi-information source fusion and optimization to realize icme: Application to dual-phase materials," *Journal of Mechanical Design*, Vol. 140, No. 11, 2018, p. 111409.

[3] Rasmussen, C. E., and Williams, C. K. I., *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.

[4] Ghoreishi, S. F., Thomison, W. D., and Allaire, D., "Sequential Information-Theoretic and Reification-Based Approach for Querying Multi-Information Sources," *Journal of Aerospace Information Systems*, Vol. 16, No. 12, 2019, pp. 575–587.

[5] Allaire, D., and Willcox, K., "Fusing information from multifidelity computer models of physical systems," *2012 15th International Conference on Information Fusion*, IEEE, 2012, pp. 2458–2465.

[6] Thomison, W. D., and Allaire, D. L., "A model reification approach to fusing information from multifidelity information sources," *19th AIAA non-deterministic approaches conference*, 2017, p. 1949.

[7] Alexandrov, N., Lewis, R., Gumbert, C., Green, L., and Newman, P., "Approximation and Model Management in Aerodynamic Optimization with Variable-Fidelity Models," *AIAA Journal*, Vol. 38, No. 6, 2001, pp. 1093–1101.

[8] Balabanov, V., Haftka, R., Grossman, B., Mason, W., and Watson, L., "Multifidelity Response Surface Model for HSCT Wing Bending Material Weight," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, St. Louis, MO, 1998. AIAA 1998-4804.

[9] Balabanov, V., and Venter, G., "Multi-Fidelity Optimization with High-Fidelity Analysis and Low-Fidelity Gradients," *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany, New York, 2004. AIAA 2004-4459.

[10] Choi, S., Alonso, J. J., and Kroo, I. M., "Two-level multifidelity design optimization studies for supersonic jets," *Journal of Aircraft*, Vol. 46, No. 3, 2009, pp. 776–790.

[11] Eldred, M., Giunta, A., and Collis, S., "Second-order corrections for surrogate-based optimization with model hierarchies," 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2004. AIAA 2004-4457.

[12] Allaire, D., and Willcox, K., "Surrogate modeling for uncertainty assessment with application to aviation environmental system models," *AIAA journal*, Vol. 48, No. 8, 2010, pp. 1791–1803.

[13] March, A., and Willcox, K., "Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives," *AIAA Journal*, Vol. 50, No. 5, 2012, pp. 1079–1089.

[14] March, A., and Willcox, K., "Convergent multifidelity optimization using Bayesian model calibration," *Structural and Multidisciplinary Optimization*, Vol. 46, No. 1, 2012, pp. 93–109.

[15] Leamer, E., *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, John Wiley & Sons, New York, NY, 1978.

[16] Madigan, D., and Raftery, A., "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window," *American Statistical Association*, Vol. 89, No. 428, 1994, pp. 1535–1546.

[17] Draper, D., "Assessment and Propagation of Model Uncertainty," *Journal of the Royal Statistical Society Series B*, Vol. 57, No. 1, 1995, pp. 45–97.

[18] Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C., "Bayesian Model Averaging: A Tutorial," *Statistical Science*, Vol. 14, No. 4, 1999, pp. 382–417.

[19] Clyde, M., "Model Averaging," In Subjective and Objective Bayesian Statistics, 2nd ed., Chapter 13, Wiley-Interscience, 2003.

[20] Clyde, M., and George, E., "Model Uncertainty," *Statistical Science*, Vol. 19, 2004, pp. 81–94.

[21] Mosleh, A., and Apostolakis, G., "The Assessment of Probability Distributions from Expert Opinions with an Application to Seismic Fragility Curves," *Risk Analysis*, Vol. 6, No. 4, 1986, pp. 447–461.

[22] Zio, E., and Apostolakis, G., "Two Methods for the Structured Assessment of Model Uncertainty by Experts in Performance Assessments of Radioactive Waste Repositories," *Reliability Engineering & System Safety*, Vol. 54, No. 2-3, 1996, pp. 225–241.

[23] Reinert, J., and Apostolakis, G., "Including Model Uncertainty in Risk-informed Decision Making," *Annals of Nuclear Energy*, Vol. 33, No. 4, 2006, pp. 354–369.

[24] Riley, M., and Grandhi, R., "Quantification of Modeling Uncertainty in Aeroelastic Analyses," *Journal of Aircraft*, Vol. 48, No. 3, 2011, pp. 866–873.

[25] Julier, S., and Uhlmann, J., "A Non-Divergent Estimation Algorithm in the Presence of Unknown Correlations," In proceedings of the American Control Conference, pp. 2369-2373, 1997.

[26] Julier, S., and Uhlmann, J., "General Decentralized Data Fusion with Covariance Intersection," In D. Hall and J. Llinas, editor, Handbook of Data Fusion. CRC Press, Boca Raton FL, USA, 2001, ????

[27] Geisser, S., "A Bayes Approach for Combining Correlated Estimates," *Journal of the American Statistical Association*, Vol. 60, 1965, pp. 602–607.

[28] Morris, P., "Combining Expert Judgments: A Bayesian Approach," *Management Science*, Vol. 23, 1977, pp. 679–693.

[29] Winkler, R., "Combining Probability Distributions from Dependent Information Sources," *Management Science*, Vol. 27, No. 4, 1981, pp. 479–488.

[30] Marler, R. T., and Arora, J. S., "The weighted sum method for multi-objective optimization: new insights," *Structural and multidisciplinary optimization*, Vol. 41, No. 6, 2010, pp. 853–862.

[31] Kim, I. Y., and de Weck, O. L., "Adaptive weighted-sum method for bi-objective optimization: Pareto front generation," *Structural and multidisciplinary optimization*, Vol. 29, No. 2, 2005, pp. 149–158.

[32] Das, I., and Dennis, J. E., "Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems," *SIAM journal on optimization*, Vol. 8, No. 3, 1998, pp. 631–657.

[33] Zitzler, E., and Thiele, L., "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," *IEEE transactions on Evolutionary Computation*, Vol. 3, No. 4, 1999, pp. 257–271.

[34] Emmerich, M. T., Deutz, A. H., and Klinkenberg, J. W., "Hypervolume-based expected improvement: Monotonicity properties and exact computation," *2011 IEEE Congress of Evolutionary Computation (CEC)*, IEEE, 2011, pp. 2147–2154.

[35] Beume, N., "S-metric calculation by considering dominated hypervolume as klee's measure problem," *Evolutionary Computation*, Vol. 17, No. 4, 2009, pp. 477–492.

[36] Bradstreet, L., While, L., and Barone, L., "A fast many-objective hypervolume algorithm using iterated incremental calculations," *IEEE Congress on Evolutionary Computation*, IEEE, 2010, pp. 1–8.

[37] Fonseca, C. M., Paquete, L., and López-Ibánez, M., "An improved dimension-sweep algorithm for the hypervolume indicator," *2006 IEEE international conference on evolutionary computation*, IEEE, 2006, pp. 1157–1163.

[38] Yang, Q., and Ding, S., "Novel algorithm to calculate hypervolume indicator of Pareto approximation set," *International Conference on Intelligent Computing*, Springer, 2007, pp. 235–244.

[39] Russo, L. M., and Francisco, A. P., "Quick hypervolume," *IEEE Transactions on Evolutionary Computation*, Vol. 18, No. 4, 2013, pp. 481–502.

[40] Ghoreishi, S. F., and Allaire, D. L., "Gaussian process regression for Bayesian fusion of multi-fidelity information sources," *2018 Multidisciplinary Analysis and Optimization Conference*, 2018, p. 4176.

[41] Lam, R., Allaire, D. L., and Willcox, K. E., "Multifidelity optimization using statistical surrogate modeling for non-hierarchical information sources," *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2015, p. 0143.

[42] Couckuyt, I., Deschrijver, D., and Dhaene, T., "Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization," *Journal of Global Optimization*, Vol. 60, No. 3, 2014, pp. 575–594.

[43] Zhao, G., Arroyave, R., and Qian, X., "Fast Exact Computation of Expected HyperVolume Improvement," *arXiv preprint arXiv:1812.07692*, 2018.

[44] Feliot, P., Bect, J., and Vazquez, E., "A Bayesian approach to constrained single-and multi-objective optimization," *Journal of Global Optimization*, Vol. 67, No. 1-2, 2017, pp. 97–133.

[45] Patel, J. K., and Read, C. B., *Handbook of the normal distribution*, Vol. 150, CRC Press, 1996.

[46] While, L., Bradstreet, L., and Barone, L., "A fast way of calculating exact hypervolumes," *IEEE Transactions on Evolutionary Computation*, Vol. 16, No. 1, 2011, pp. 86–95.

[47] Jaszkiewicz, A., "Improved quick hypervolume algorithm," *Computers & Operations Research*, Vol. 90, 2018, pp. 72–83.

[48] Deb, K., Thiele, L., Laumanns, M., and Zitzler, E., "Scalable multi-objective optimization test problems," *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*, Vol. 1, IEEE, 2002, pp. 825–830.

[49] Jasa, J. P., Hwang, J. T., and Martins, J. R., "Open-source coupled aerostructural optimization using Python," *Structural and Multidisciplinary Optimization*, Vol. 57, No. 4, 2018, pp. 1815–1827.

[50] Gray, J. S., Hwang, J. T., Martins, J. R. R. A., Moore, K. T., and Naylor, B. A., "OpenMDAO: an open-source framework for multidisciplinary design, analysis, and optimization," *Structural and Multidisciplinary Optimization*, Vol. 59, No. 4, 2019, pp. 1075–1104. https://doi.org/10.1007/s00158-019-02211-z, URL https://doi.org/10.1007/s00158-019-02211-z.

[51] Chauhan, S. S., and Martins, J. R., "Low-fidelity aerostructural optimization of aircraft wings with a simplified wingbox model using OpenAeroStruct," *International Conference on Engineering Optimization*, Springer, 2018, pp. 418–431.

[52] Elham, A., and van Tooren, M. J., "Coupled adjoint aerostructural wing optimization using quasi-three-dimensional aerodynamic analysis," *Structural and Multidisciplinary Optimization*, Vol. 54, No. 4, 2016, pp. 889–906.