BAYESIAN OPTIMIZATION IN MULTI-INFORMATION SOURCE AND LARGE-SCALE SYSTEMS

A Dissertation

by

SEYEDE FATEMEH GHOREISHI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Douglas L. Allaire
Committee Members,	Raymundo Arroyave
	Richard J. Malak Jr.
	Ankit Srivastava
Head of Department,	Andreas A. Polycarpou

May 2019

Major Subject: Mechanical Engineering

ABSTRACT

The advancements in science and technology in recent years have extended the scale of engineering problems. Discovery of new materials with desirable properties, drug discovery for treatment of disease, design of complex aerospace systems containing interactive subsystems, conducting experimental design of complex manufacturing processes, designing complex transportation systems all are examples of complex systems. The significant uncertainty and lack of knowledge about the underlying model due to the complexity necessitate the use of data for analyzing these systems. However, a huge time/economical expense involved in data gathering process avoids acquiring large amount of data for analyzing these systems. This dissertation is mainly focused on enabling design and decision making in complex uncertain systems.

Design problems are pervasive in scientific and industrial endeavors: scientists design experiments to gain insights into physical and social phenomena, engineers design machines to execute tasks more efficiently, pharmaceutical researchers design new drugs to fight disease, and environmentalists design sensor networks to monitor ecological systems. All these design problems are fraught with choices, choices that are often complex and high-dimensional, with interactions that make them difficult for individuals to reason about. Bayesian optimization techniques have been successfully employed for experimental design of these complex systems.

In many applications across computational science and engineering, engineers, scientists and decision-makers might have access to a system of interest through several models. These models, often referred to as “information sources", may encompass different resolutions, physics, and modeling assumptions, resulting in different “fidelity" or “skill" with respect to the quantities of interest. Examples of that include different finite-element models in design of complex mechanical structures, and various tools for analyzing DNA and protein sequence data in bioinformatics. Huge computation of the expensive models avoids excessive evaluations across design space. On the other hand, less expensive models fail to represent the objective function accurately. Thus, it is highly desirable to determine which experiment from which model should be conducted at

each time point. We have developed a multi-information source Bayesian optimization framework capable of simultaneous selection of design input and information source, handling constraints, and making the balance between information gain and computational cost. The application of the proposed framework has been demonstrated on two different critical problems in engineering: 1) optimization of dual-phase steel to maximize its strength-normalized strain hardening rate in materials science; 2) optimization of NACA 0012 airfoil in aerospace.

The design problems are often defined over a large input space, demanding large number of experiments for yielding a proper performance. This is not practical in many real-world problems, due to the budget limitation and data expenses. However, the objective function (i.e., experiment's outcome) in many cases might not change with the same rate in various directions. We have introduced an adaptive dimensionality reduction Bayesian optimization framework that exponentially reduces the exploration region of the existing techniques. The proposed framework is capable of identifying a small subset of linear combinations of the design inputs that matter the most relative to the objective function and taking advantage of the objective function representation in this lower dimension, but with richer information. A significant increase in the rate of optimization process has been demonstrated on an important problem in aerospace regarding aerostructural design of an aircraft wing modeled based on the NASA Common Research Model (CRM).

ACKNOWLEDGMENTS

First and foremost, I would like to sincerely thank my academic advisor, Dr. Douglas L. Allaire, for his help, advising, and unconditional encouragement and support. He was more than generous with his expertise and precious time. I also thank my committee members, Dr. Raymundo Arroyave, Dr. Richard Malak and Dr. Ankit Srivastava, for their constructive comments and good-natured support. Special thanks goes to my husband, Mahdi Imani, for being there throughout the entire period of my doctoral studies. All of my success and achievements are undoubtedly because of his endless support. It is my pleasure to also appreciate my family and friends who made this dissertation a beginning of my professional journey.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Dr. Allaire and Dr. Malak of the Department of Mechanical Engineering and Dr. Arroyave and Dr. Srivastava of Department of Materials Science and Engineering. All work for the dissertation was completed by the student, under the advisement of Dr. Allaire of the Department of Mechanical Engineering.

My graduate study was supported by the AFOSR MURI on multi-information sources of multi-physics systems under Award Number FA9550-15-1-0038, and by the National Science Foundation under grant no. CMMI-1663130.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1 Bayesian Optimization in Multi-Information Source Systems

For many applications, there exist multiple models that simulate the same system. These models typically rely on different mathematical formulae or make varying assumptions that simplify the problem. This leads to models with varying degrees of discrepancy from the true output, or fidelity. Generally speaking, a high fidelity model is more accurate than a low fidelity one at the cost of computational expense. Intuitively, using the high fidelity model leads to a reduction in uncertainty due to model inadequacy. However, a low fidelity model can still provide information that should be considered in any decision making process. A method is needed to fuse the information from both models to create a single, fused distribution for the quantities of interest. The goal of this study is to propose a method to utilize the information of all models to generate a fused model with superior predictive capability than any of its constituent models. The other goal of this study is sequential decision making to identify the next design point and information source to evaluate, in order to create a fused distribution using data from models of varying fidelity. To achieve these goals, our methodology estimates the correlation between each model using a model reification approach that eliminates the observational data requirement. The correlation is then used in an updating procedure whereby uncertain outputs from multiple models may be fused together to better estimate some quantity or quantities of interest. These ingredients are used in a decision-theoretic manner to query from multiple information sources sequentially to achieve the maximum expected value-gradient utility per unit cost.

1.2 Bayesian Optimization in Large-Scale Systems

The real practical systems are often complex and defined over large input design space. Finding the design point with the highest quantity of interest is a common goal in these systems. Since the information regarding these systems are often available through data, surrogate models such as Gaussian processes are usually used for modeling these systems. The optimization of these systems

demands querying a number of samples in a sequential fashion, and the knowledge gradient is a popular technique to find the optimal solution with minimum function evaluations. The knowledge gradient method requires a finite set of alternatives to decide the best point in the alternative set which leads to the highest one-step look-ahead expected improvement in the maximum of the objective function. While the method has been successfully applied to many domains, it usually requires a large number of queries to perform well for optimization of systems with large number of input design variables. Indeed, the alternative set exponentially grows by the number of input variables. Furthermore, the set of alternatives has the direct effect on the performance of the optimization problem, since large and naive set increases the exploration process as opposed to the exploitation. This requires a large number of function evaluations to solve the optimization problem. Our aim is to develop a rigorous methodology to provide a rich set of alternatives which is exponentially smaller than the original set. Such a strategy boosts the speed of optimization for systems with large input space. We propose to develop a two-step knowledge gradient method. The method first employs the active subspace technique to adaptively map the original space to a low-dimensional space. Then, the knowledge gradient is performed over this small-dimension to select a design point to query. The inverse mapping of the selected point leads to a hyperplane in the original space, therefore we perform the second knowledge gradient to select a sample in this hyperplane.

2. BAYESIAN OPTIMIZATION IN MULTI-INFORMATION SOURCE SYSTEMS*

2.1 Overview

Design decisions for complex systems often can be made or informed by a variety of information sources. When optimizing such a system, the evaluation of a quantity of interest is typically required at many different input configurations. For systems with expensive to evaluate available information sources, the optimization task can potentially be computationally prohibitive using traditional techniques. This section presents an information-economic approach to the constrained optimization of a system with multiple available information sources. The approach rigorously quantifies the correlation between the discrepancies of different information sources, which enables the overcoming of information source bias. All information is exploited efficiently by fusing newly acquired information with that previously evaluated. Independent decision makings are achieved by developing a two-step look-ahead utility policy and an information gain policy for objective function and constraints respectively. The approach is demonstrated on a one-dimensional example test problem and an aerodynamic design problem, where it is shown to perform well in comparison to traditional multi-information source techniques.

2.2 Introduction

Design decisions for complex engineered systems are increasingly being informed by queries to multiple sources of information. These information sources may consist of variable fidelity models, expert opinions, historical data, or newly conducted physical experiments. Our aim here is to create a systematic procedure for enabling the efficient exploitation of all available information sources. The context we focus on is constrained optimization problems with expensive to evaluate information sources. For such problems, traditional optimization strategies, such as the

*Reprinted with permission from "Multi-Information Source Constrained Bayesian Optimization" by S. F. Ghoreishi, and D. Allaire., 2018. Structural and Multidisciplinary Optimization, Copyright Springer-Verlag GmbH Germany, part of Springer Nature 2018.
*Reprinted with permission from "Multi-information source fusion and optimization to realize icme: Application to dual-phase materials" by S. F. Ghoreishi, A. Molkeri, A. Srivastava, R. Arroyave, and D. Allaire., 2018. Journal of Mechanical Design, Copyright 2018 by ASME.

employment of gradient-based algorithms or heuristic techniques can be inappropriate. This can be, for example, due to sheer expense in the case of heuristics, and the need to calculate finite differences of say an experimental information source in the case of gradient-based algorithms, where resources often would not be spent on queries in such close proximity in input space. Given such considerations, the approach we propose for multi-information source optimization focuses on Bayesian global optimization strategies, where query points are selected based on expected contribution to such quality metrics as information gained regarding a constraint or expected improvement in an objective. Such a strategy enables the natural fusion of information from multiple sources, which we exploit here to ensure that all available information can be used for the optimization task at hand.

Our work builds generally on concepts from multifidelity optimization, where several different computational models are available for a given task. One common approach in multifidelity optimization is to treat the models as a hierarchy and replace or calibrate low-fidelity information with high-fidelity results [2–9]. Trust region methods are used to approximate the high-fidelity model for optimizing an objective function when the derivatives of the objective are available. In Refs. [10] and [11], a trust region based model-management method is employed in which the gradients of the low-fidelity objective function and constraints are scaled or shifted to match those of a high-fidelity model. Ref. [12] presents a convergent multifidelity optimization algorithm using the trust region method that does not require the high-fidelity derivatives, and considers hierarchies between the low and high-fidelity models. In Ref. [13], a multifidelity sequential kriging optimization method is proposed which uses the expected improvement as a measure to determine the next design point and the level of fidelity model to evaluate, and the autoregressive assumption is used to build the multi-fidelity model based on hierarchical relationships among the information sources. In practice, the fidelity of information sources can vary over the design space. Therefore, the assumption of hierarchical relationship among the information sources might not be practical, and all the available data from different information sources need to be fused to take into account all potential information any given source may bring to bear on the optimization problem.

There are several techniques used in practice for achieving the fusion of information from multiple information sources. Among them are the adjustment factors approach [14–16], Bayesian model averaging [17–24] which believe that each model has some probability of being true and the fused estimate is a weighted average of the available models. The other available techniques are fusion under known correlation [25–32], and covariance intersection method [33, 34]. Covariance intersection algorithm has a tendency to discard lower fidelity information sources to achieve the most conservative estimate possible in cases with scalar quantities of interest. It requires that the information source with the highest fidelity be used for the fused estimate. While using the high fidelity model at all points in the design space leads to a reduction in uncertainty due to model inadequacy, this approach leads to a loss of information provided by the low fidelity models that could be used to better estimate the model output if the correlation coefficients were known. Accounting for correlation is critical in information fusion. Aggressive approaches, such as assuming each model is independent from all others, are often incorrect, and can lead to potentially serious misconceptions regarding confidence in quantity of interest estimates.

In this section, we propose a method for the constrained optimization of a quantity of interest where multiple information sources, which are generally expensive to evaluate, are available for approximating the quantity of interest and constraints. To enable informed querying of information sources and to account for each source's discrepancy, we construct Gaussian processes for each source using available evaluations from a given source and that source's discrepancy information. Following standard practice for the fusion of normally distributed data [35], we create fused Gaussian processes for the objective function and constraints, that contain all currently available information from each information source. A novel aspect of this work is the use of fusion techniques that support the rigorous incorporation of correlation both within information sources in terms of information source estimates, and between information sources in terms of information source estimate errors. This latter aspect enables the mitigation of bias among information sources. Specifically, traditional fusion techniques estimate a quantity of interest as a weighted average of information source estimates. This average always resides within the extremes of the information

source estimates. That is, if all information sources are biased in the same direction with respect to a quantity of interest, traditional fusion techniques cannot overcome this bias and will produce a similarly biased estimate. Our work employs a novel model reification technique described in detail in Ref. [36], that overcomes this limitation in the fused Gaussian processes.

By constructing the fused Gaussian processes, the next design to evaluate and the information source to query for the objective function are determined based on our proposed two-step look-ahead policy. This policy is achieved by sequential augmentation process and the knowledge gradient policy as a one-step look-ahead policy to maximize (or minimize) an objective [37–40]. Notice that our proposed two-step look-ahead policy goes beyond the existing Bayesian optimization techniques which only take the next step into account for decision-making process. The next design to evaluate and the information source to query for the constraints are determined based on our proposed information gain policy achieved by sequential augmentation process and the Kullback-Leibler divergence. This is the novel aspect of our methodology that decision-making for querying from constraint information sources is independent of objective function decision-making, and the constraint information sources can be queried at different design locations than the objective information source during the same optimization iteration, and involve entirely separate information sources than those of the objective. The proposed method yields several benefits summarized as follows

- Incorporation of correlations both within and between information sources;
- Direct decision-making on the fused model regarding which information source to query next and where to query it;
- A two-step look-ahead policy and an information gain policy for sequential objective function and constraints decision-making;
- Independent decision-making for querying from constraints and objective function information sources.

Our proposed multi-information source optimization approach is applied to optimize a one-dimensional example test problem and an aerodynamic design example, and the results are com-

pared with a traditional multi-information source optimization method presented in Ref. [1] which is currently the state of the art in constrained Bayesian optimization. It is shown that our proposed method results in higher quality solutions for the demonstrations presented, particularly in the case of a limited budget available for querying.

2.3 Problem Statement

We consider the problem of optimizing a quantity of interest subject to m inequality constraints, where multiple sources of information are available for both the objective and constraints. A mathematical statement of the problem is formulated as

$$\begin{aligned} \mathbf{x}^* &= \underset{\mathbf{x}\in\chi}{\operatorname{argmax}}\, f(\mathbf{x}) \\ \text{s.t.} \quad & g_j(\mathbf{x}) \leq 0, \qquad j = 1, 2, \ldots, m, \end{aligned} \tag{2.1}$$

where f is the real-world objective function, g_j, where $j = 1, 2, \ldots, m$, are the m real-world constraints, and $\mathbf{x}$ is a set of design variables in the vector space χ. In the optimization process, the real-world objective function and the real-world constraints must be estimated at each iteration.

For this estimation task, often several information sources, such as numerical simulation models, experiments, expert opinions, etc., are available. These sources have varying fidelities, or approximation accuracies over the design space, and varying computational costs. The approach we describe here enables the exploitation of all available sources of information, in an information-economic sense, that balances the cost and associated fidelity of each information source when choosing how to estimate the quantity of interest and the constraints. In our formulation, the core issue is how to manage dynamic querying to choose what new information sources to sample and with what input, at each step of the overall decision process. The choice of next design point to query must be based on how much a sample tells us about the design goal at hand and if that sample is expected to satisfy the current constraints. This choice must trade off the costs associated with a particular information source query and the utility or information gained by executing the query.

Here, 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 $\mathbf{x}$. In order to predict the output of each information source at locations that have not yet been queried, an intermediate surrogate is constructed for each information source using Gaussian processes [41]. These surrogates are denoted by $f_{GP,i}(\mathbf{x})$. We consider the prior distributions of the information sources modeled by Gaussian processes as

$$f_{GP,i}(\mathbf{x}) \sim \mathcal{GP}\left(\mathbf{0}, k_i(\mathbf{x}, \mathbf{x})\right), \tag{2.2}$$

where $k_i(\mathbf{x}, \mathbf{x})$ is a real-valued kernel function over the input space. For the kernel function, we consider the commonly used squared exponential covariance function, which is specified as

$$k_i(\mathbf{x}, \mathbf{x}') = \sigma_s^2 \exp\left(-\sum_{h=1}^{d} \frac{(x_h - x_h')^2}{2l_h^2}\right), \tag{2.3}$$

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 via maximum 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 $\mathbf{y}_{N_i} = (f_i(\mathbf{x}_{1,i}), \ldots, f_i(\mathbf{x}_{N_i,i}))$ represents the corresponding outputs from information source i, the posterior distribution of information source i at design point $\mathbf{x}$ is given as

$$f_{GP,i}(\mathbf{x}) \mid \mathbf{X}_{N_i}, \mathbf{y}_{N_i} \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma_{GP,i}^2(\mathbf{x})\right), \tag{2.4}$$

where

$$\begin{aligned} \mu_i(\mathbf{x}) &= K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma_{n,i}^2 I]^{-1} \mathbf{y}_{N_i}, \\ \sigma_{GP,i}^2(\mathbf{x}) &= k_i(\mathbf{x}, \mathbf{x}) - K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma_{n,i}^2 I]^{-1} K_i(\mathbf{X}_{N_i}, \mathbf{x}), \end{aligned} \tag{2.5}$$

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. Note that here we have included the term $\sigma_{n,i}^2$, which can be used to model observation error for information sources based

on experiments and can also be used to guard against numerical ill-conditioning.

To these Gaussian process surrogates, in order to estimate the fused value as will be discussed in Section 3.5, we further quantify the uncertainty of each information source with respect to the true quantity of interest by adding a term associated with the fidelity of the information source. 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

$$\sigma_i^2(\mathbf{x}) = \sigma_{GP,i}^2(\mathbf{x}) + \sigma_{f,i}^2(\mathbf{x}), \tag{2.6}$$

where $\sigma_{f,i}^2(\mathbf{x})$ is the fidelity variance of information source i that has been estimated from, for example, expert opinion or available real-world data. Here, we estimate the fidelity variance of each information source by computing the error values which are 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. Figure 2.1 shows a depiction of total variance for an information source represented by Gaussian process.

Our approach to information sources that can be used to evaluate any of the m constraints in Problem (2.1) follows that of our approach to the objective based information sources. For any given constraint, g_j, where $j \in \{1, 2, \ldots, m\}$, we have available some set of information sources, $g_{i,j}(\mathbf{x})$, where $i \in \{1, 2, ..., G_j\}$, that can be used to estimate the quantity of interest, $g_j(\mathbf{x})$, at the design point $\mathbf{x}$. Following the same approach taken for the objective information sources, we quantify the total variance of the i^{th} information source of the j^{th} constraint as

$$\sigma_{i,j}^2(\mathbf{x}) = \sigma_{GP,i,j}^2(\mathbf{x}) + \sigma_{g_j,i}^2(\mathbf{x}), \tag{2.7}$$

where $\sigma_{GP,i,j}^2(\mathbf{x})$ is the variance associated with the Gaussian process representation of $g_{i,j}$ and $\sigma_{g_j,i}^2(\mathbf{x})$ is the quantified fidelity variance of the information source.

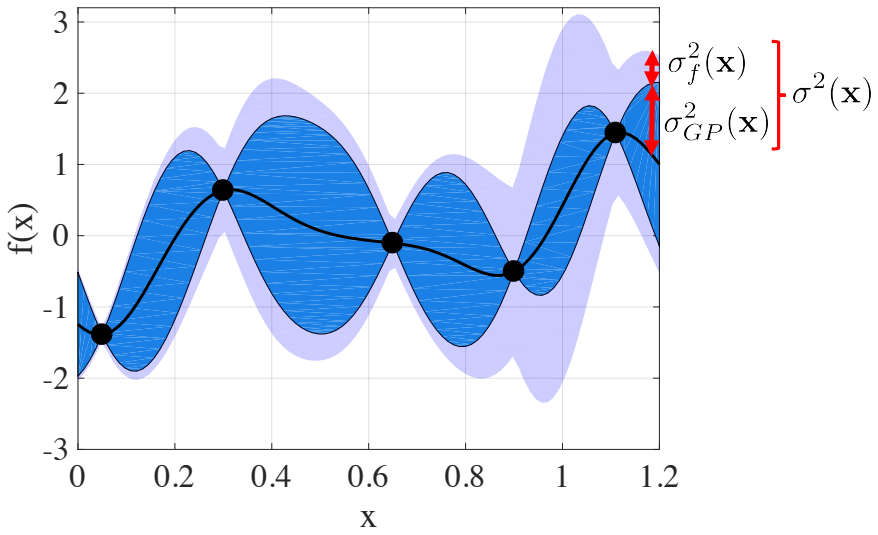


Figure 2.1: A depiction of total variance, which includes both the variance associated with the Gaussian process and variance associated with the fidelity of the information source.

2.4 Approach

This part describes our multi-information source constrained Bayesian optimization approach. We first present how we fuse the knowledge from information sources, and construct fused Gaussian processes for the objective function and the m constraints, required for finding the solution for the maximization problem in Equation (2.1). Then, we present our proposed strategies for choosing the next design points and information sources of objective function and constraints to query. We conclude this section with a complete algorithm for implementing our approach.

2.4.1 Fusion of Information Sources

A fundamental claim in this work is that any information source can provide potentially useful information to a given task at hand. That is, information sources are not assumed to be hierarchical and the goal is not to just optimize whatever is deemed the highest fidelity source. Indeed, according to Equations (2.6) and (2.7), the discrepancy of an information source can vary over its domain. Thus, a strict hierarchy of information source fidelity is an unlikely and restrictive assumption.

To take into account all potential information any given source may bring to bear on the opti-

mization problem (i.e., objective or constraint information at a given input), we fuse our information sources following standard practice for the fusion of normally distributed data [35], since our information sources are represented as Gaussian processes. According to the method of Winkler, the fused mean and variance at point $\mathbf{x}$ can be computed as [35]

$$\mu_{\text{Wink}}(\mathbf{x}) = \frac{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \boldsymbol{\mu}(\mathbf{x})}{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \tag{2.8}$$

$$\sigma^2_{\text{Wink}}(\mathbf{x}) = \frac{1}{\mathbf{e}^T \tilde{\Sigma}(\mathbf{x})^{-1} \mathbf{e}}, \tag{2.9}$$

where $\mathbf{e} = [1, \ldots, 1]^\top$, $\boldsymbol{\mu}(\mathbf{x}) = [\mu_1(\mathbf{x}), \ldots, \mu_S(\mathbf{x})]^\top$ contains the mean values of S models at point $\mathbf{x}$, and $\tilde{\Sigma}(\mathbf{x})$ is the covariance matrix between the models given as

$$\tilde{\Sigma}(\mathbf{x}) = \begin{bmatrix} \sigma_1^2(\mathbf{x}) & \cdots & \rho_{1S}(\mathbf{x})\, \sigma_1(\mathbf{x}) \sigma_S(\mathbf{x}) \\ \rho_{12}(\mathbf{x})\, \sigma_1(\mathbf{x}) \sigma_2(\mathbf{x}) & \cdots & \rho_{2S}(\mathbf{x})\, \sigma_2(\mathbf{x}) \sigma_S(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \rho_{1S}(\mathbf{x})\, \sigma_1(\mathbf{x}) \sigma_S(\mathbf{x}) & \cdots & \sigma_S^2(\mathbf{x}) \end{bmatrix}, \tag{2.10}$$

where $\sigma_i^2(\mathbf{x})$ is the total variance of model i at point $\mathbf{x}$ and $\rho_{ij}(\mathbf{x})$ is the correlation coefficient between the deviations of information sources i and j at point $\mathbf{x}$.

To estimate the correlations between the model deviations, we use the reification process defined in [36, 42], which refers to the process of treating each model, in turn, as ground truth. Following this process, the correlation between the errors is computed as follows

$$\rho_{ij}(\mathbf{x}) = \frac{\sigma_j^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})} \tilde{\rho}_{ij}(\mathbf{x}) + \frac{\sigma_i^2(\mathbf{x})}{\sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x})} \tilde{\rho}_{ji}(\mathbf{x}), \tag{2.11}$$

where $\sigma_i^2(\mathbf{x})$ and $\sigma_j^2(\mathbf{x})$ are the total variances of information sources i and j at point $\mathbf{x}$, and $\tilde{\rho}_{ij}(\mathbf{x})$

and $\tilde{\rho}_{ji}(\mathbf{x})$ are computed by reifying models i and j respectively as

$$\begin{aligned}\tilde{\rho}_{ij}(\mathbf{x}) &= \frac{\sigma_i(\mathbf{x})}{\sqrt{\left(\mu_i(\mathbf{x}) - \mu_j(\mathbf{x})\right)^2 + \sigma_i^2(\mathbf{x})}},\\ \tilde{\rho}_{ji}(\mathbf{x}) &= \frac{\sigma_j(\mathbf{x})}{\sqrt{\left(\mu_j(\mathbf{x}) - \mu_i(\mathbf{x})\right)^2 + \sigma_j^2(\mathbf{x})}},\end{aligned} \tag{2.12}$$

where $\mu_i(\mathbf{x})$ and $\mu_j(\mathbf{x})$ are the mean values of models i and j respectively at design point $\mathbf{x}$. Note that the correlation between the errors in (2.11) is the variance weighted average of the two correlation coefficients in (2.12). The detailed process of estimating the correlation between the errors of two models can be found in [36, 42].

To imbue the fused model with the complete learned information, we construct a fused Gaussian process (which also accounts for correlation between design points). To accomplish this, we construct the Gaussian process from a finite set of samples $\mathbf{x}_{1:N_f} \subset \chi$. Let $\boldsymbol{\mu}_{\text{Wink}}(\mathbf{x}_{1:N_f})$ and $\Sigma(\mathbf{x}_{1:N_f}) = diag\left(\sigma^2_{\text{Wink}}(\mathbf{x_1}), \ldots, \sigma^2_{\text{Wink}}(\mathbf{x}_{N_f})\right)$ be the vector of fused means and a diagonal matrix containing the fused variances at $\mathbf{x}_{1:N_f}$ obtained according to Equations (2.8) and (2.9) respectively. Considering the squared exponential covariance function (other options could be employed) and obtaining the hyperparameters via maximum likelihood, we construct the fused Gaussian process. The posterior predictive distribution of fused estimate at any set of samples $\mathbf{X}$ in the design space can be obtained as

$$\hat{f}^{\text{fused}}(\mathbf{X}) \sim \mathcal{N}(\boldsymbol{\mu}^{\text{fused}}(\mathbf{X}), \Sigma^{\text{fused}}(\mathbf{X})), \tag{2.13}$$

where

$$\begin{aligned}\boldsymbol{\mu}^{\text{fused}}(\mathbf{X}) &= K(\mathbf{x}_{1:N_f}, \mathbf{X})^T [K(\mathbf{x}_{1:N_f}, \mathbf{x}_{1:N_f}) + \Sigma(\mathbf{x}_{1:N_f})]^{-1} \boldsymbol{\mu}_{\text{Wink}}(\mathbf{x}_{1:N_f}),\\ \Sigma^{\text{fused}}(\mathbf{X}) &= K(\mathbf{X}, \mathbf{X}) - K(\mathbf{x}_{1:N_f}, \mathbf{X})^T [K(\mathbf{x}_{1:N_f}, \mathbf{x}_{1:N_f}) + \Sigma(\mathbf{x}_{1:N_f})]^{-1} K(\mathbf{x}_{1:N_f}, \mathbf{X}).\end{aligned} \tag{2.14}$$

Fused Gaussian processes are also constructed for each of the constraints to predict the feasibility of the design to evaluate. This approach follows that of the approach taken above for the fused Gaussian process of the objective function.

2.4.2 Two-Step Look-Ahead Utility Function for Objective Function Query

Equipped with fused Gaussian processes for the objective and each constraint, the next step of our methodology for solving Equation (2.1) is identification of the next design point location to query and with which information source to execute the query. The design point to evaluate is based on the fused Gaussian processes of the objective function and constraints. This choice must trade off the goals of learning or exploring the fused model of objective function and maximizing or exploiting this function. Further, this choice must eventually be feasible with respect to the fused constraint models. We denote the feasible design space by χ_{feas} where $\chi_{\text{feas}} \subset \chi$.

To determine the next design point and information source to query, we generate Latin Hypercube samples in the input design space as alternatives denoted as $\mathbf{X}_{\text{alt}}$. Among these alternatives, we seek to identify a set of feasible design points, as determined by dynamically updated constraints described below. From this feasible set denoted by $\mathbf{X}_{\text{feas}}$, we seek to determine which design point and which information source to query that design point will lead to the maximum utility in the fused Gaussian process of the objective function with minimum cost of query. This utility is also described below.

We define a two-step look-ahead policy as the utility function which is able to consider both our current best design given a potential query result and the ability we would have, armed with this potential query, to find an even better design on the next query. In our proposed method, unlike most Bayesian optimization techniques which are developed based on a one-step look-ahead procedure, a two-step look-ahead optimization is achieved.

Let $(\mathbf{x}_{1:N}, y_{1:N})$ be the design points and the corresponding objective values, and $i_{1:N}$ be the indices of the queried information sources of the objective function up to time step N. We define the utility obtained by querying the design point $\mathbf{x} \in \mathbf{X}_{\text{feas}}$ from information source i as

$$U_{\mathbf{x},i} = \mathbb{E}\left[\max_{\mathbf{x}' \in \chi_{\text{feas}}} \mu^{\text{fused}}(\mathbf{x}') + \max_{\mathbf{x}'' \in \chi_{\text{feas}}} EI_{\mathbf{x},i}(\mathbf{x}'') \;\middle|\; \mathbf{x}_{1:N}, y_{1:N}, i_{1:N}, \mathbf{x}_{N+1} = \mathbf{x}, i_{N+1} = i \right], \quad (2.15)$$

where $EI_{\mathbf{x},i}(\mathbf{x}'')$ is the one-step look-ahead expected increase in the maximum of the fused Gaus-

sian process given $\mathbf{x}_{N+1} = \mathbf{x}$ and $i_{N+1} = i$ as

$$EI_{\mathbf{x},i}(\mathbf{x}'') = \mathbb{E}\big[\max_{\mathbf{x}' \in \chi_{\text{feas}}} \mu^{\text{fused}}(\mathbf{x}') \mid \mathbf{x}_{N+2} = \mathbf{x}''\big] - \max_{\mathbf{x}' \in \chi_{\text{feas}}} \mu^{\text{fused}}(\mathbf{x}'). \tag{2.16}$$

The exact computation of Equation (2.15) is not possible. Thus in this paper, we provide a Monte Carlo technique for its approximation. For a given point $\mathbf{x}$ at information source i, the objective function values are normally distributed based on the mean and variance of the corresponding Gaussian process. Drawing N_q independent samples from this normal distribution as $f_i^q(\mathbf{x}) \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma^2_{GP,i}(\mathbf{x})\right)$, $q = 1, \ldots, N_q$, results in the following approximation of Equation (2.15)

$$U_{\mathbf{x},i} \approx \frac{1}{N_q} \sum_{q=1}^{N_q} \big(\max_{\mathbf{x}' \in \chi_{\text{feas}}} \mu^{\text{fused},q}_{\mathbf{x},i}(\mathbf{x}') + \max_{\mathbf{x}'' \in \chi_{\text{feas}}} EI^q_{\mathbf{x},i}(\mathbf{x}'')\big), \tag{2.17}$$

where $\mu^{\text{fused},q}_{\mathbf{x},i}(\mathbf{x}')$ denotes the mean of the fused Gaussian process upon temporarily augmenting $(\mathbf{x}\,,\, f_i^q(\mathbf{x}))$, one at a time, to the available samples of information source i, and $EI^q_{\mathbf{x},i}(\mathbf{x}'')$ is the one-step look-ahead expected increase in the maximum of the fused Gaussian process upon augmentation of query $(\mathbf{x}\,,\, f_i^q(\mathbf{x}))$ to information source i. Note that for each generated sample, $q = 1, \ldots, N_q$, the previously added sample is removed and the next new sample is augmented, then the Gaussian process of information source i and as a result, the fused Gaussian process of the objective function are updated temporarily. We compute the second term in Equation (2.17) by the Knowledge Gradient (KG) approach over the temporary fused Gaussian process in a closed form solution [38–40]. Letting $S^{N+1}_{\mathbf{x}''} = \mathbb{E}[\hat{f}^{\text{fused}}(\mathbf{x}'') \mid \mathbf{x}_{1:N}, y_{1:N}, i_{1:N}, \mathbf{x}_{N+1} = \mathbf{x}, y_{N+1} = f_i^q(\mathbf{x}), i_{N+1} = i]$ be the knowledge state, the value of being at state S^{N+1} is defined as $V^N(S^{N+1}) = \max_{\mathbf{x}'' \in \chi} S^{N+1}_{\mathbf{x}''}$. The knowledge gradient which is a measure of expected increase in the maximum of the objective function, if the design point $\mathbf{x}''$ would be queried at the next time step, can be defined as

$$\nu^{KG,N+1}_{\mathbf{x}''} = \mathbb{E}[V^{N+2}\left(S^{N+2}(\mathbf{x}'')\right) - V^{N+1}(S^{N+1}) \mid S^{N+1}],$$

where the expectation is taken over the stochasticity in the posterior distribution of the fused model

representing the objective function at design point $\mathbf{x}''$.

The utility is evaluated for all feasible, as feasibility will be discussed in the following subsections, alternatives and all information sources. By denoting $C_{\mathbf{x},i}$ as the cost of querying information source i of the objective function at design point $\mathbf{x}$, we find the query $(i_{N+1}, \mathbf{x}_{N+1})$ that maximizes the utility per unit cost, given by

$$(i_{N+1}, \mathbf{x}_{N+1}) = \underset{i\in[1,\ldots,S],\mathbf{x}\in\mathbf{X}_{feas}}{\operatorname{argmax}} \frac{U_{\mathbf{x},i}}{C_{\mathbf{x},i}}. \tag{2.18}$$

Figure 2.2 shows a depiction of computation of our proposed two-step look-ahead utility function for choosing the next query for an objective function with two information sources. The left column shows the Gaussian processes of the two information sources and the fused Gaussian process constructed from these information sources. The right column shows the temporary Gaussian processes of information source 1 and the fused model after augmenting the query $(\mathbf{x}, q)$ to information source 1 $(i = 1)$. The terms of the two-step look-ahead utility function are shown on the temporary fused Gaussian process.

2.4.3 Information Gain Policy for Constraints Query

We propose to use the information gain for selection of the next design point and information source of the constraints to query. In order to measure the information gain, we use the Kullback-Leibler (KL) divergence [43–46] which is a criterion for evaluating the difference between two probability distributions P and Q with densities $p(x)$ and $q(x)$, defined as

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)}\, dx. \tag{2.19}$$

Latin Hypercube sampling is again used for generating alternatives denoted as $\mathbf{X}_g$ in the design space. For a given point $\mathbf{x}$ at information source i of constraint g_j, the constraint values are normally distributed based on the mean and variance of the corresponding Gaussian process. By drawing N_r independent samples from this normal distribution denoted by $g^r_{i,j}(\mathbf{x})$, $r = 1, \ldots, N_r$,

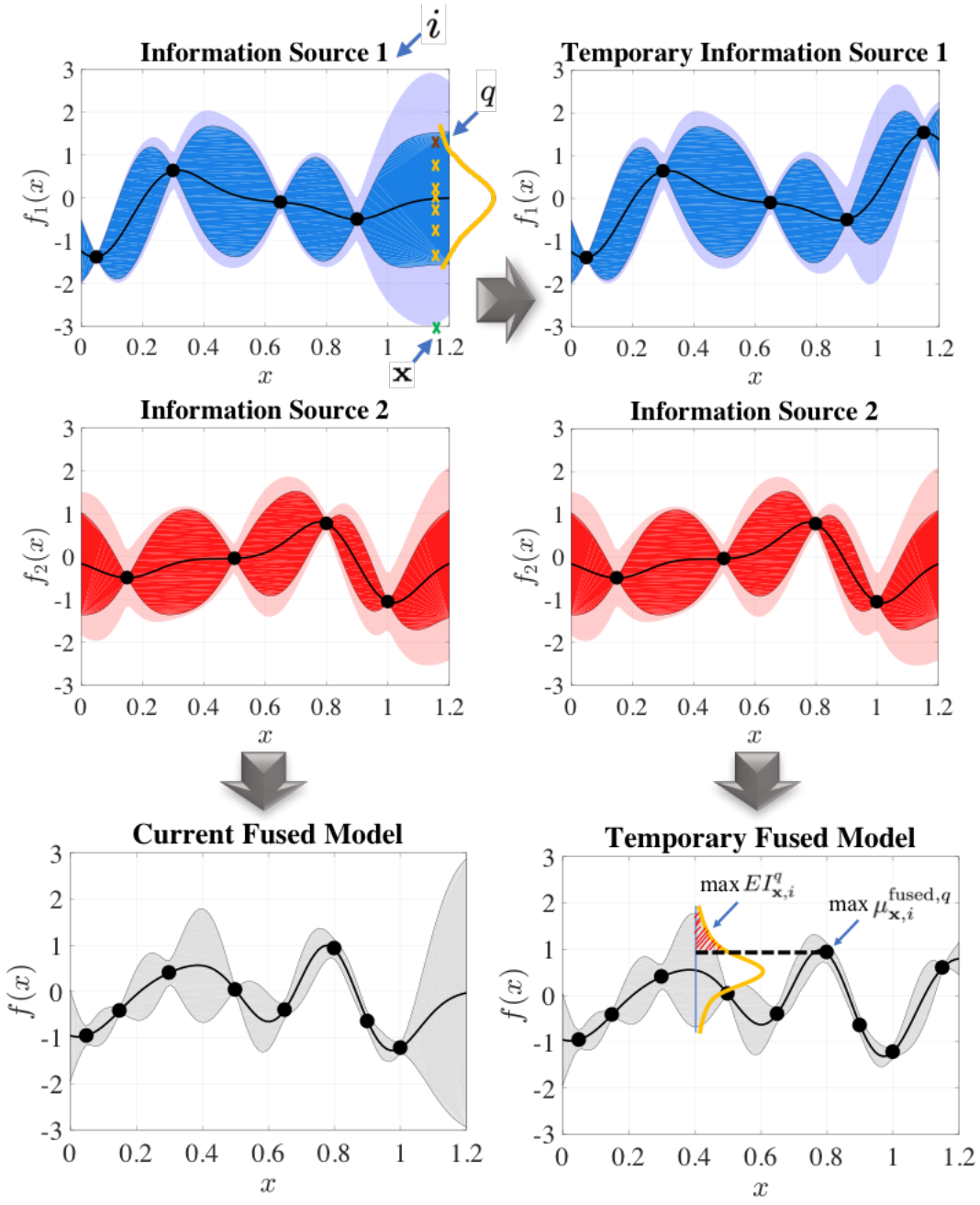


Figure 2.2: A depiction of computing our proposed two-step look-ahead utility function for choosing the next design and information source of objective function to query.

and temporarily augmenting each sample, one at a time, to the available samples of the corresponding information source, the Gaussian process of information source i and consequently, the fused Gaussian process of constraint g_j are temporarily updated. We denote the temporary fused

Gaussian process of constraint g_j after augmenting sample $(\mathbf{x}\,,\, g^r_{i,j}(\mathbf{x}))$ to information source i by $\hat{g}^{\text{fused},r}_{\mathbf{x},i,j}$, and the current actual fused Gaussian process before augmenting any sample by $\hat{g}^{\text{fused}}_j$. We repeat this process for all N_r generated samples. Then, we compute the Information Gain (IG) for the potential query of design point $\mathbf{x}$ at information source i of constraint g_j which is the average Kullback-Leibler divergence between the current and the temporary Gaussian processes over the design space as

$$IG^j_{\mathbf{x},i} = \frac{1}{N_r}\sum_{r=1}^{N_r}\sum_{s=1}^{N_s} D_{KL}\left(\hat{g}^{\text{fused},r}_{\mathbf{x},i,j}(\mathbf{x}'_s) \parallel \hat{g}^{\text{fused}}_j(\mathbf{x}'_s)\right), \tag{2.20}$$

where the Kullback-Leibler divergence over the design space of two Gaussian processes is approximated by Monte Carlo technique at N_s samples $\mathbf{x}'_s \in \chi$. The information gain is then evaluated for all the alternatives and all the information sources of constraint g_j. By denoting $C^j_{\mathbf{x},i}$ as the cost of querying information source i of constraint g_j at design $\mathbf{x}$, we find the query $(i^j_{N+1}, \mathbf{x}^j_{N+1})$ that maximizes the information gain per unit cost, given by

$$(i^j_{N+1}, \mathbf{x}^j_{N+1}) = \underset{i\in[1,\ldots,G_j],\mathbf{x}\in\mathbf{X}_g}{\operatorname{argmax}} \frac{IG^j_{\mathbf{x},i}}{C^j_{\mathbf{x},i}}. \tag{2.21}$$

2.4.4 Feasibility of a Design Point

The specifics of our constraint handling methodology are as follows. The value of fused constraint g_j at a design point $\mathbf{x}$ is estimated to be distributed normally with mean $\mu^{\text{fused}}_j(\mathbf{x})$ and standard deviation $\sigma^{\text{fused}}_j(\mathbf{x})$, according to the fused Gaussian process representation of that constraint. In the first iterations of the approach, since the constraints are not learned, we aim to avoid over-restricting the search space and to involve the infeasible samples in the search process to explore the information they might carry. We achieve this by considering the fused standard deviation of constraints at each design point in addition to the fused mean value of the constraints at that point. Therefore, the design point $\mathbf{x}$ must satisfy

$$\mu^{\text{fused}}_j(\mathbf{x}) - 3\sigma^{\text{fused}}_j(\mathbf{x}) \leq 0, \qquad j = 1, \ldots, m. \tag{2.22}$$

In this strategy, variances are large in the first iterations, which helps the methodology to explore the design space, and as the process goes on, the fused Gaussian processes of the constraints are updated, which results in decreases in the value of their respective variances. This reduces the possibility of an infeasible sample to be selected. Here, a practitioner can control the level of acceptability in terms of constraint violation by modifying the coefficient of $\sigma_j^{\text{fused}}(\mathbf{x})$ in Equation (2.22) as deemed appropriate. It should be noted that in the final step for returning the optimal solution, only the fused means of the constraints are considered for determining the feasibility of the design points, as will be discussed in the next subsection. Figure 2.3 shows this strategy that the feasible region shrinks and gets closer to the true feasible region after one query.

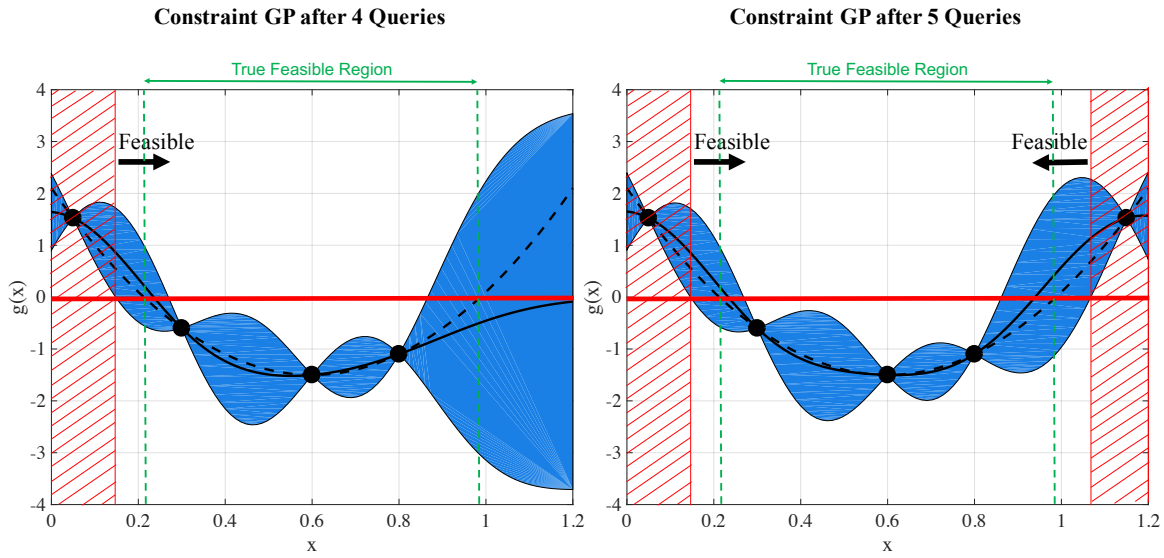


Figure 2.3: A depiction of our constraint handling strategy that the feasible region shrinks and gets closer to the true feasible region after one query.

2.4.5 Multi-Information Source Constrained Bayesian Optimization Approach

Our complete proposed multi-information source constrained Bayesian optimization approach is presented in Figure 2.4. In our proposed approach, the first step is constructing a Gaussian process for each information source of the objective function and all the constraints based on their

available data. These information sources are combined together and the fused Gaussian processes of the objective function and constraints are constructed as discussed in subsection 2.4.1. The next design points and information sources to query are selected based on our proposed two-step look-ahead utility function for the objective function and the information gain measurement for the constraints. After querying the selected design points from the selected information sources, the corresponding Gaussian processes are updated. Therefore, the mean, variance and correlation coefficient values are also updated, resulting in new fused Gaussian processes of the objective function and constraints based on the Gaussian processes of the information sources learned so far in the process and their learned correlations. This process repeats until a termination criterion, such as exhaustion of the querying budget, is met.

Finally, the feasible design point with maximum fused mean of the objective function is selected as the optimum solution of Equation (2.1), as

$$\begin{aligned} \mathbf{x}^* &= \underset{\mathbf{x}\in\chi_{\text{feas}}}{\operatorname{argmax}}\, \mu^{\text{fused}}(\mathbf{x}), \\ \chi_{\text{feas}} &:= \{\mathbf{x} \mid \mu_j^{\text{fused}}(\mathbf{x}) \leq 0, \qquad j = 1, \ldots, m\}. \end{aligned} \tag{2.23}$$

In order to find the best approximation to the optimum solution, we discretize the design space. This discretization must be sufficiently fine to ensure a good approximation of the continuous design space solution. Then, the values of the predicted mean of the objective function and the constraints are computed at each point according to Equation (2.8). Among all the feasible points, the point that has the maximum function value is selected as the optimum solution. We note here, that alternatively, we could also optimize the fused objective function subject to the fused constraint functions using standard gradient-based techniques on the mean functions of the processes.

2.5 Application and Results

In this section, we present the key features of our multi-information source constrained Bayesian optimization approach on three demonstrations. The first case is an analytical problem with one-dimensional input and output, the second demonstration is the minimization of the drag coefficient

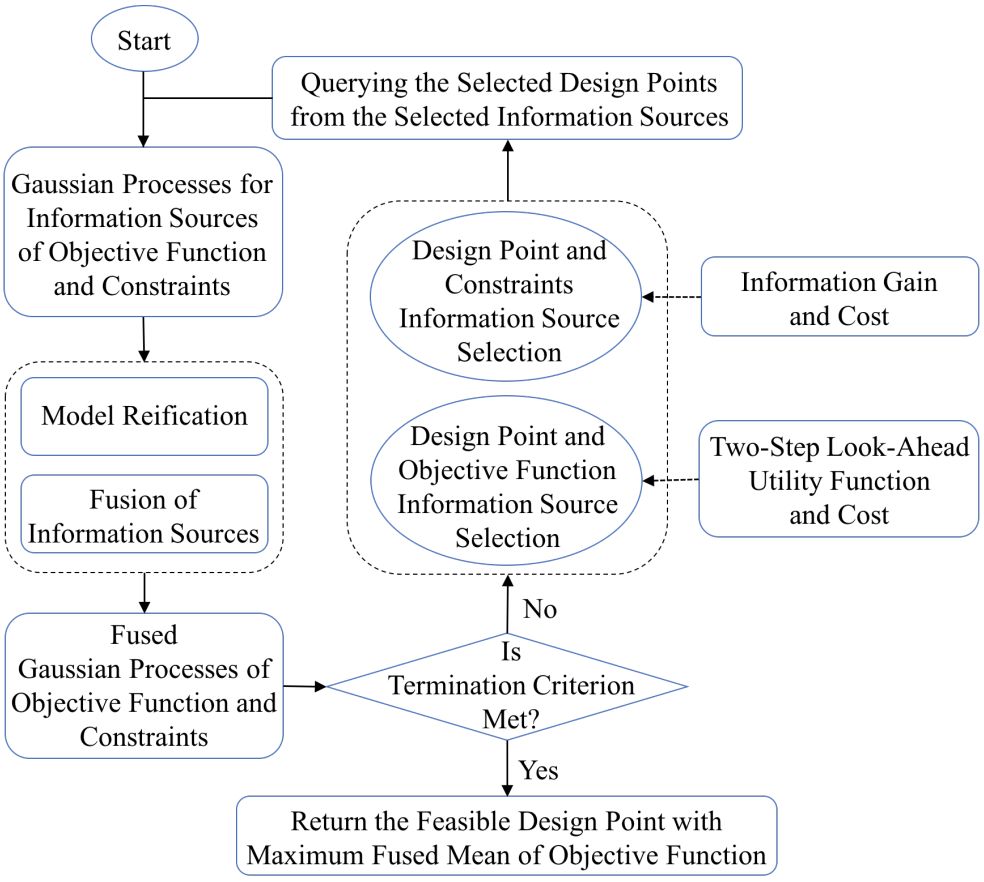


Figure 2.4: A depiction of our proposed multi-information source constrained Bayesian optimization approach.

of a NACA 0012 airfoil subject to a constraint on the lift coefficient, and the third case is optimization of dual-phase steel to maximize its strength-normalized strain hardening rate in materials science.

2.5.1 One-Dimensional Problem

The first example is maximization of a one-dimensional constrained function shown in Figure 2.5 and defined as

$$\begin{aligned} x^* = \underset{\mathbf{x}\in[0\,1.2]}{\operatorname{argmax}} \; & -(1.4-3x)\sin(18x) \\ \text{s.t.} \quad & x^2 - 1.2 \leq 0. \end{aligned} \tag{2.24}$$

The feasible optimal solution for this problem is $(x^*, f(x^*)) = (1.0954, 1.4388)$. We consider two

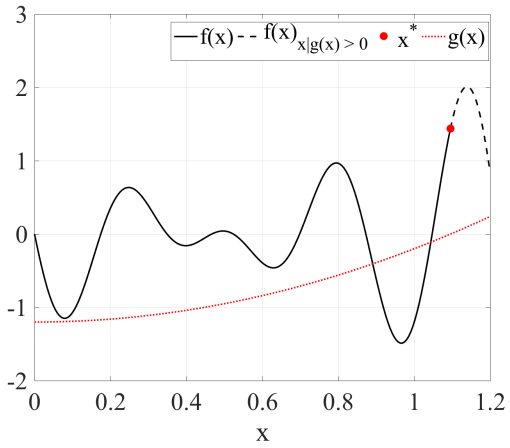


Figure 2.5: One-dimensional optimization problem of Equation (2.24).

information sources of $f_1(x)$ and $f_2(x)$ for the objective function, with constant discrepancies of $\sigma^2_{f,1} = 0.04$ and $\sigma^2_{f,2} = 0.2$, and evaluation costs of $C_1 = 25$ and $C_2 = 20$, given as

$$\begin{aligned} f_1(x) &= -(1.6 - 3x)\sin(18x), \\ f_2(x) &= -(1.8 - 3x)\sin(18x + 0.1). \end{aligned} \tag{2.25}$$

We note that in our methodology, two Gaussian processes are constructed for these two information sources based on their available data and $\sigma^2_{f,1}$ and $\sigma^2_{f,2}$ are added to the uncertainty associated with the Gaussian processes to obtain the total uncertainties.

Two information sources, denoted as $g_{1,1}(x)$ and $g_{2,1}(x)$, are considered for the constraint. These sources have constant discrepancies of $\sigma^2_{g_1,1} = 5e-6$ and $\sigma^2_{g_1,2} = 0.002$, and evaluation costs of $C^1_1 = 10$ and $C^1_2 = 5$, given as

$$\begin{aligned} g_{1,1}(x) &= (x - 0.001)^2 - 1.2, \\ g_{2,1}(x) &= (x + 0.02)^2 - 1.2. \end{aligned} \tag{2.26}$$

We note that the constant discrepancies used in this example are not restrictions of the work. In the next application problem the discrepancies are permitted to vary over the design space.

Figure 2.6 shows the Gaussian processes of the two objective information sources, the fused objective function, and the fused constraint function. We note that owing to the constant discrepancies, we can refer specifically to the information sources as low-fidelity and high-fidelity, where high fidelity refers to the source with lower discrepancy. The fused models shown in the figure represent what has been learned after applying our methodology with a computational budget limited to 450 and 20 Latin Hypercube samples as alternatives. The red solid lines show the true function $f(x)$ and the true constraint $g(x)$, and the black solid lines show the learned mean functions of the Gaussian processes. It is shown that after querying 6 samples from the high-fidelity model and 9 samples from the low-fidelity model of objective function, the fused model provides a reasonable estimate of the true objective function and the feasible optimal solution.

In order to asses the performance of our approach, we compare our methodology with a standard multi-information source optimization method presented in Ref. [1]. We refer to this method as the MF algorithm. This algorithm is the state of art in constrained Bayesian optimization and has been used as a benchmark for many Bayesian optimization techniques [**?**, 47–51]. In Ref. [1], each information source has a separate Gaussian process, and predictions are obtained by fusing the information via the method presented in Ref. [35] without considering the correlation between the information sources. In this algorithm, the acquisition function selects the point to sample and the information source to query separately. The next design to evaluate is selected by applying the expected improvement function on the multi-fidelity surrogate, and the next information source of objective function to query is chosen based on a heuristic that aims to balance information gain and cost of query. It should be noted that there is no decision making with regards to the information sources used for constraints in the MF algorithm. In the MF algorithm, it is assumed that the constraint information source is identical to the objective function information source and the query is based on that of the objective.

Table 2.1 represents the expectation and variance of the design points obtained by our proposed approach and the MF algorithm, as well as the expectation and variance of the objective function at these points over 100 independent replications of the simulations for two different cost values

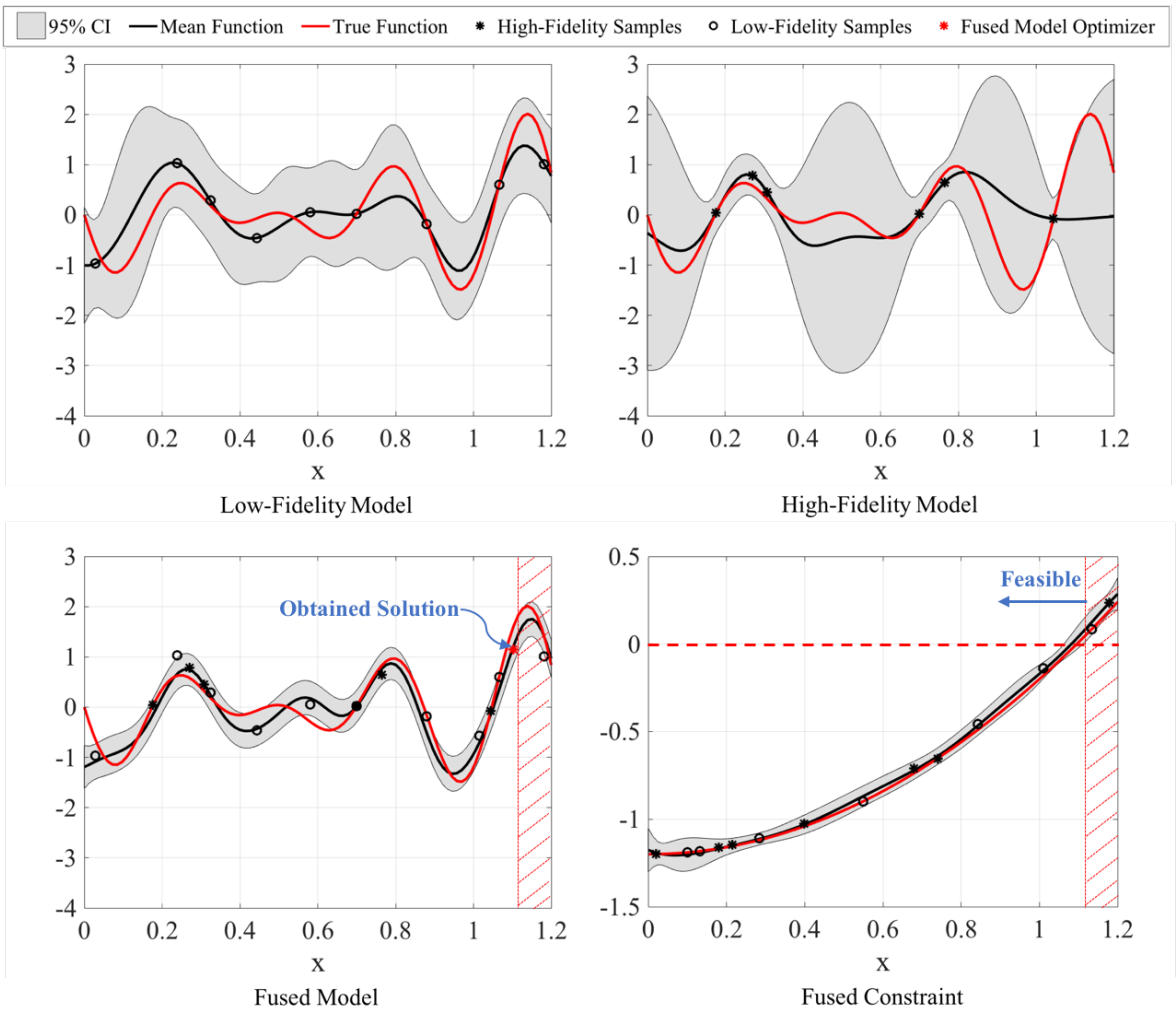


Figure 2.6: The learned low-fidelity, high-fidelity and fused objective functions and fused constraint after applying our approach on optimization problem of Equations (2.24-2.26).

for information sources and budget of 1000 as the stopping criterion. As can be seen, in the case of $C_1 = 20, C_2 = 20$ in which the cost of the first information source is lower, better results are obtained in the fixed budget as more samples can be queried from the information sources in comparison to $C_1 = 25, C_2 = 20$. Figure 2.7 shows the average maximum function values obtained by our approach and MF algorithm for different budgets.

As can be seen, our approach improves on the work of Ref. [1], as the expectations of the ob-

Table 2.1: Expectation and variance of the results obtained by our proposed approach and the MF Algorithm [1] over 100 replications of the simulations of optimization problem in Equations (2.24-2.26) for two different cost values for information sources and budget of 1000 as the stopping criterion. The feasible optimal solution is $(x^*, f(x^*)) = (1.0954, 1.4388)$.

	$C_1 = 25, C_2 = 20$				$C_1 = 20, C_2 = 20$			
	$\mathbb{E}[x^*]$	$var(x^*)$	$\mathbb{E}[f(x^*)]$	$var(f(x^*))$	$\mathbb{E}[x^*]$	$var(x^*)$	$\mathbb{E}[f(x^*)]$	$var(f(x^*))$
Proposed Approach	1.0819	0.0070	1.4139	0.0215	1.0886	0.0068	1.4279	0.0173
MF Algorithm [1]	1.0415	0.0389	1.2216	0.0471	1.0591	0.0308	1.3083	0.0406

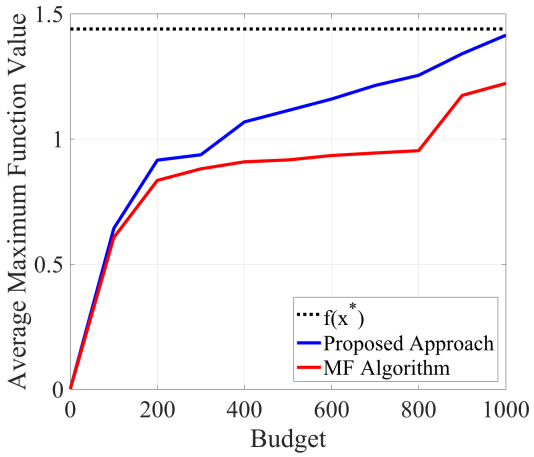


Figure 2.7: The average maximum function value obtained by our proposed approach and the MF Algorithm [1] over 100 independent replications of the simulations of the optimization problem in Equations (2.24-2.26) for different budgets as the stopping criterion.

tained results are closer to the true optimal values and the variances of the estimations are lower. This is due to the fact that our methodology allows for the rigorous exploitation of correlations between information sources and across the design space. This results in reduced uncertainty by querying one new design sample, even if it is queried from an information source with lower fidelity. Thus, we obtain more accurate estimates of the true objective function and also constraint from each sample. Furthermore, allowing for separate decision making with regards to the con-

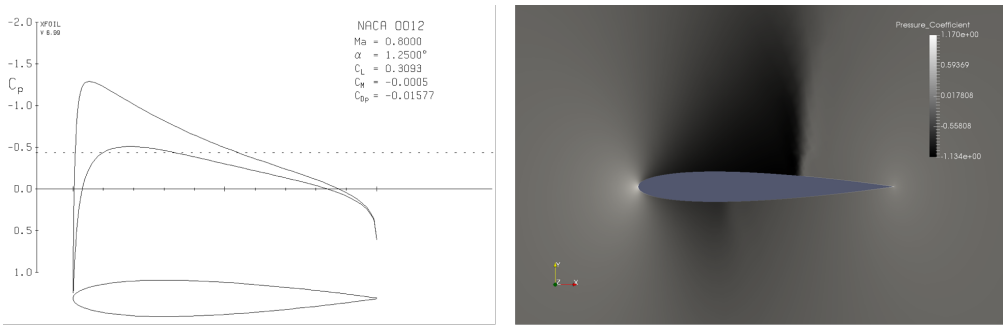
straints adds flexibility to our approach that can also be exploited for efficiency gains. With regards to the exploitation of correlation, we note here that the value of the low fidelity information source at the true optimum is 1.2240 and the value of the high fidelity information source is 1.2862 at the same location. The true objective function value at that point is 1.4388. As can be seen from the table, the estimated objective function value is greater than either information source using our fusion-based approach built with model reification. This is not the case for the MF algorithm, which has no means for overcoming bias. Finally, it should be mentioned that the higher performance of the proposed method comes with higher computational cost. Indeed, our proposed method constructs $|\mathbf{X}_{\text{feas}}| \times N_q \times S$ more Gaussian processes for each query in comparison to the MF algorithm, while these computations are negligible in comparison to the actual experiments' costs. Notice that the computation involved for each query due to the need for multiple Gaussian processes by the proposed method can be done in parallel.

2.5.2 NACA 0012 Drag Coefficient with Lift Coefficient Constraints

The second demonstration of our multi-information source constrained Bayesian optimization approach is a two-dimensional constrained aerodynamic design example. The airfoil of interest is the NACA 0012, a common validation airfoil [52, 53]. In this demonstration, we consider two information sources for the objective and constraint. The sources are the computational fluid dynamics programs XFOIL [54] and SU2 [55]. Figure 2.8 shows an example output of the two simulators that illustrates the difference in fidelity levels.

XFOIL is a solver for the design and analysis of airfoils in the subsonic regime. It combines a panel method with the Karman-Tsien compressibility correction for the potential flow with a two-equation boundary layer model. This causes XFOIL to overestimate lift and underestimate drag [56]. SU2 uses a finite volume scheme and Reynolds-averaged Navier-Stokes (RANS) method with the Spalart-Allmaras turbulence model, which allows SU2 to be significantly more accurate than XFOIL in the more turbulent flow regimes at higher values of Mach number and angle of attack.

In this problem, we are particularly concerned with finding the Mach number M and angle of



(a) XFOIL (b) SU2

Figure 2.8: Example outputs of NACA 0012 airfoil from XFOIL and SU2.

attack α that minimize the coefficient of drag C_D of a NACA 0012 airfoil subject to maintaining a minimum coefficient of lift C_L

$$\begin{aligned} \mathbf{x}^* = \underset{\mathbf{x}\in\chi}{\operatorname{argmax}} \; & -C_D(\mathbf{x}) \\ \text{s.t.} \quad & 0.4 - C_L(\mathbf{x}) \leq 0, \end{aligned} \tag{2.27}$$

where $\mathbf{x}^* = (M^*, \alpha^*)$. The design space is $\chi = I_M \times I_\alpha$ with $I_M = [0.15 \; 0.75]$ and $I_\alpha = [-2.2 \; 13.3]$. To validate our approach, wind tunnel data from NASA and AGARD are used to construct true models using spline interpolation to determine values between the given data points. The true models of the objective and the constraint constructed from the wind tunnel data, as well as a contour plot of objective and constraint are shown in Figure 2.9. These data are also used to estimate the model discrepancy at all points in the design space by taking the difference between the true model and the simulator, XFOIL or SU2. The differences at all available samples are used as training data for a Gaussian process regression. We set the evaluation costs of the objective information sources to be $C_1 = 100$ and $C_2 = 70$, and for information sources of the constraint, we set the evaluation costs to be $C_{g_{11}} = 50$ and $C_{g_{21}} = 20$. The number of Latin Hypercube samples as alternatives is set to 100.

Figures 2.10 and 2.11 show the Gaussian processes and contour plots of XFOIL and SU2 coefficient of drag obtained by our proposed approach and the MF Algorithm respectively. As

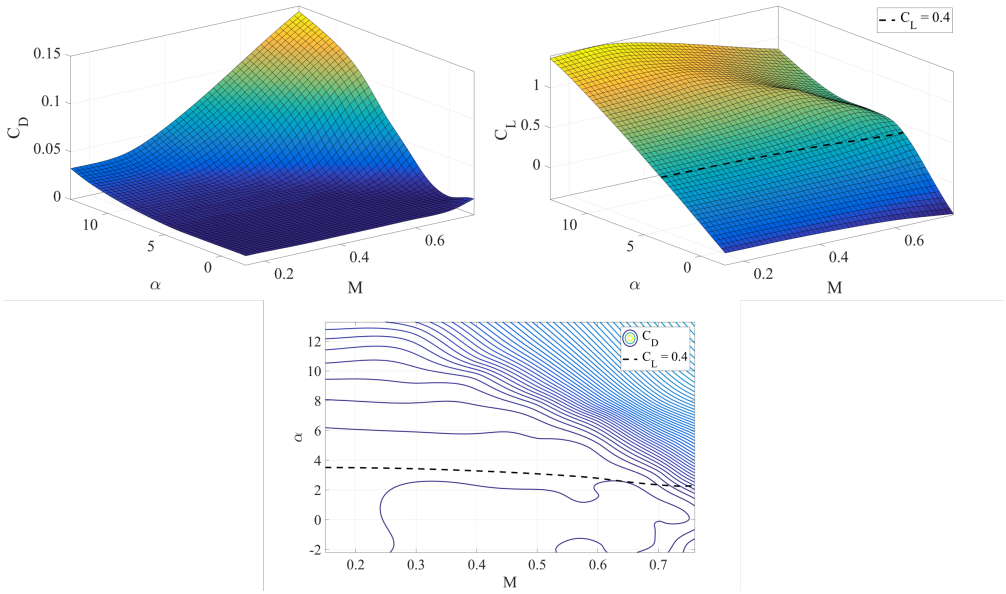


Figure 2.9: Optimization problem of Equation (2.27).

can be seen, in our proposed approach by querying 61 samples from SU2 and 56 samples from XFOIL, the regions around the optimal solution of both information sources are well explored, and the optimal solutions of the learned surrogates of the information sources are close to the optimal design points of the real information sources. However, in the MF algorithm, 8 and 131 samples are queried from the SU2 and XFOIL information sources respectively and the Gaussian process of SU2 is not properly explored, therefore the optimal solution is far from the true value.

Table 2.2 represents averaged results and their variances obtained by our approach and the MF algorithm for a computational budget limited to 100,000. The expected results are averaged over 1000 replications of the simulations. The true models used to validate the method and estimate the model discrepancies are the full set of real-world NACA 0012 wind tunnel data, which include 68 data points for objective (C_D) and 64 data points for constraint (C_L) throughout the design space. The feasible optimal design point and the corresponding coefficient of drag for the true model are $\mathbf{x}^* = (0.6574, 3.0132)$ and $C_D{}^* = 0.0064$ respectively. It is clear that our proposed methodology outperforms the MF Algorithm by obtaining the lower expected objective value, which is closer to

the optimal solution obtained from the true model. Furthermore, our approach obtained expected design points closer to the true feasible design points with lower variances. We emphasize here that our approach is developed to best approximate the true optimum design and objective and not to identify the optimal point on the highest fidelity model. In this case, as in the previous demonstration, the fusion of our information sources, with the incorporation of model reification, enabled a more accurate estimate of the true optimal point, which could not have been achieved with either model in isolation, or with traditional fusion techniques.

Table 2.2: Expectation and variance of the results obtained by our proposed approach and MF Algorithm [1] over 1000 replications of the simulations of optimization problem in Equation (2.27). True optimal values are $\mathbf{x}^* = (0.6574, 3.0132)$ and ${C_D}^* = 0.0064$.

	$\mathbb{E}[M^*]$	$\mathrm{v}ar(M^*)$	$\mathbb{E}[\alpha^*]$	$\mathrm{v}ar(\alpha^*)$	$\mathbb{E}[{C_D}^*]$	$\mathrm{v}ar({C_D}^*)$
Proposed Approach	0.6169	0.0135	3.0240	0.0696	0.0071	0.0018
MF Algorithm [1]	0.4258	0.0231	4.3744	2.3556	0.0088	0.0018

2.5.3 Strength-Normalized Strain Hardening Rate in Dual-Phase Materials

In this part, we demonstrate the application of our framework to the optimization of the ground truth strength normalized strain hardening rate for the dual-phase steel application. We stress here that the purpose of our framework is the optimization of ground truth. Thus, we seek to identify the best candidate for a ground truth experiment with whatever resources we have available. Once those resources are exhausted, a ground truth experiment takes place based on the recommendation of our framework. The result of that experiment can then be fed back into the framework. If more resources are then allocated, perhaps on the basis of promising results, then the framework can be employed again.

The specific demonstration consists of the use of the three reduced-order models (isostrain,

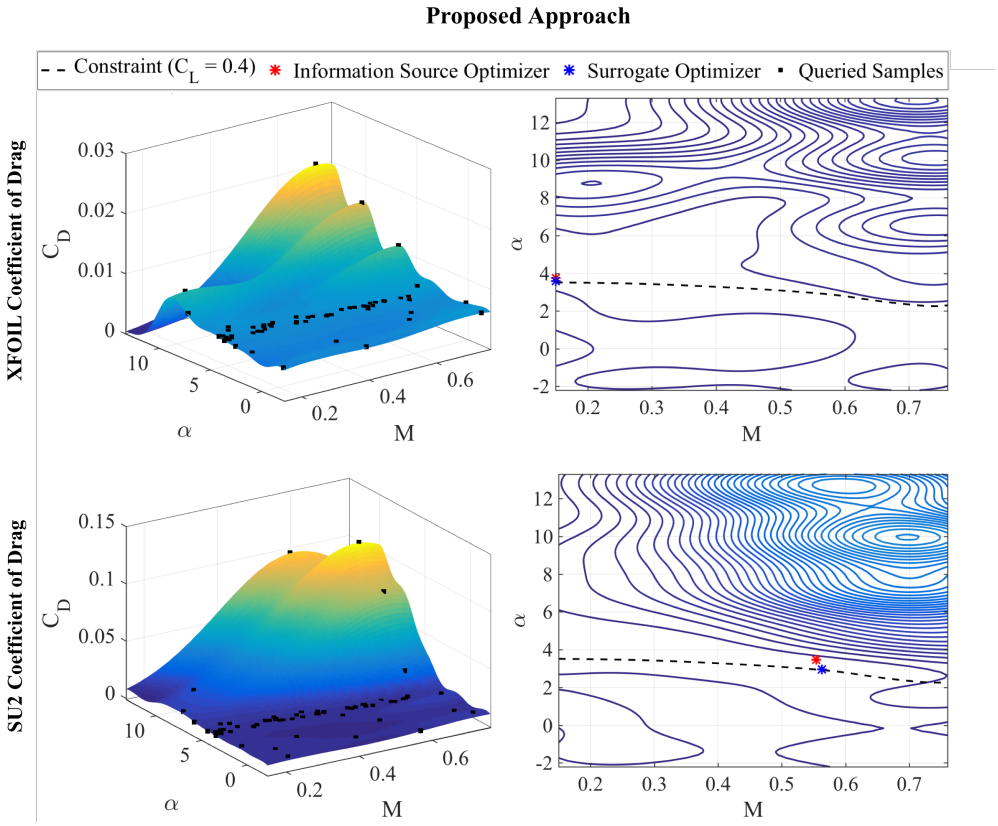


Figure 2.10: Gaussian processes and contour plots of XFOIL and SU2 coefficient of drag obtained by our proposed approach.

isostress, and isowork) to query the impact of quantifiable microstructural attributes on the mechanical response of a composite microstructure—in this case a dual-phase steel. The ground truth in this case is the finite element model of the dual-phase material. The objective is the maximization of the (ground truth) normalized strain hardening rate at $\epsilon_{pl} = 1.5\%$. The design variable is the percentage of the hard phase, f_{hard}, in the dual-phase material. The variation of the strength normalized strain-hardening rate, $(1/\tau)(\mathrm{d}\tau/\mathrm{d}\epsilon_{pl})$, with the volume fraction of the hard phase, f_{hard}, estimated at a strain level, $\epsilon_{pl} = 1.5\%$, using the microstructure-based finite element calculations, is shown in Figure 2.12.

We assume that our resources limit us to five total queries to (any of) the information sources before we must make a recommendation for a ground truth experiment. Given promising ground

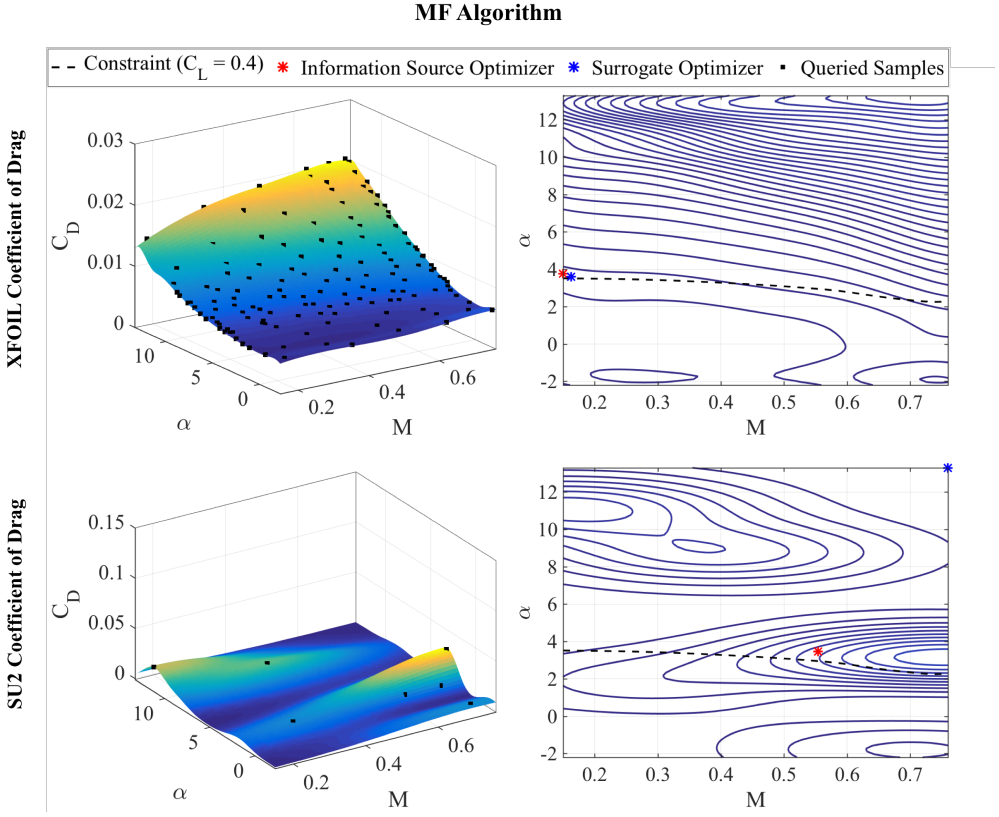


Figure 2.11: Gaussian processes and contour plots of XFOIL and SU2 coefficient of drag obtained by MF Algorithm.

truth results, five more queries can be allocated to the information sources. The framework is initialized with one query from each information source and one query from the ground truth. This information is used to construct the initial intermediate Gaussian process surrogates.

The value-gradient policy per unit cost of our framework is used to select the next information source and the location of the query in the input space for each iteration of the process. For comparison purposes, the KG policy operating directly on the ground truth is also used to reveal the gains that can be had by considering all available information sources. For this, a Gaussian process representation is created and updated after each query to ground truth. The convergence results of our proposed approach using all information sources and the KG policy on the ground truth are shown in Figure 2.13. On the figure, the dashed line represents the optimal value of

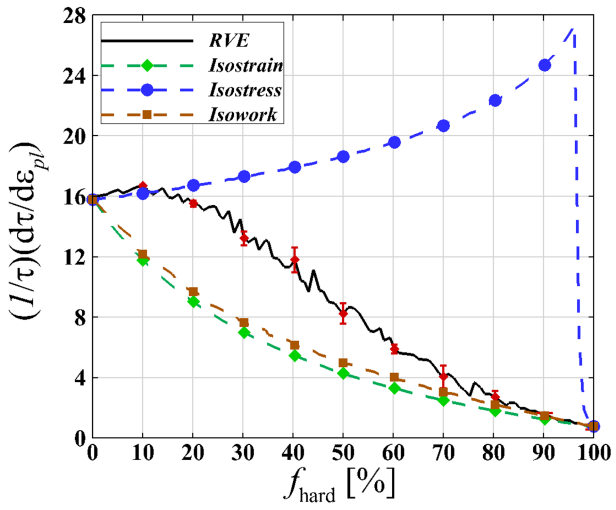


Figure 2.12: Comparison of the variation of the strength normalized strain-hardening rate, $(1/\tau)(\mathrm{d}\tau/\mathrm{d}\epsilon_{pl})$ at $\epsilon_{pl} = 1.5\%$, with the volume fraction of the hard phase, f_{hard}, as predicted by the three reduced-order models and the microstructure-based finite element model.

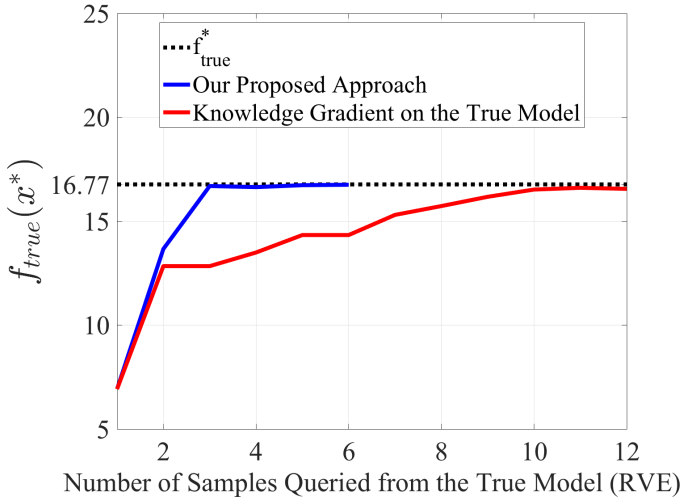


Figure 2.13: The optimal solution obtained by our proposed approach and by applying the knowledge gradient on a GP of only the true data (RVE) for different number of samples queried from the true model.

the ground truth quantity of interest. It is clear from this figure that our approach outperforms the knowledge gradient applied to directly to ground truth, and in doing so, saves considerable expense by reducing the number of needed ground truth experiments. The superior performance

of our approach can be attributed to its ability to efficiently utilize the information available from the three low fidelity information sources to better direct the querying at ground truth. We note that the original sample from ground truth used for initialization was taken at $f_{\text{hard}} = 95\%$, which is far away from the true optimal. This can be seen below in Figure 2.14 in the left column. Thus, the framework, by leveraging the three inexpensive available information sources, is able to quickly direct the ground truth experiment to a higher quality region of the design space.

Table 2.3 presents the results of each ground truth experiment conducted according to the recommendation of our framework. From the table we see that the third recommendation for a ground truth experiment produces a nearly optimal design. The final three experiments show that little more is gained in terms of ground truth objective and that the fused model has learned more about the ground truth in that region. At this point, it is likely that more resources would not be allocated to this design problem and the framework is able to successfully find the best design.

Table 2.3: The optimal solution obtained by the fused model, and the true value at the obtained optimum design point. The true optimal solution by the RVE model is $(\mathbf{x}^*\ ,\ f^*) = (8.54\ ,\ 16.77)$.

Experiment	$\mathbf{x}^*_{fused}$	f^*_{fused}	$f_{true}(\mathbf{x}^*_{fused})$
2	28.64	7.50	13.66
3	10.05	11.98	16.69
4	10.55	13.92	16.64
5	9.55	15.47	16.73
6	8.80	16.71	16.75

Updates to each information source Gaussian process surrogate model and the fused model representing our knowledge of ground truth are also shown in Figure 2.14 for iterations 1, 15, and 30 of the information source querying process. Here, an iteration occurs when an information source is queried. This is distinct from any queries to ground truth. As can be seen from the left column, the first experiment from ground truth and the first query from each information source told us little about the location of the true objective. However, on iteration 15, the fused model,

shown by the smooth red curve, has identified the best region of the design space, although it under-predicts the ground truth at this point. We note that at this point, only 3 expensive ground truth experiments have been conducted. By iteration 30, the fused model is very accurate in the region surrounding the optimal design for ground truth. At this point, six ground truth experiments have been conducted. From Figure 2.12 and this figure, it is clear that none of the information sources share the ground truth optimum. The ability of the framework to find this optimum rested upon the use of correlation exploiting fusion, and would not have been possible using traditional methods.

To conclude this demonstration we present the history of the queries to each information source and the ground truth. This information is provided in Fig. 2.15. Note, that the iteration now counts queries to each information source as well as ground truth experiments. From the figure, it is clear that all three information sources are exploited to find the ground truth optimal design, implying that, however imperfect, all sources available to the designer must be used, in an optimal manner, in order to identify the optimal ground truth.

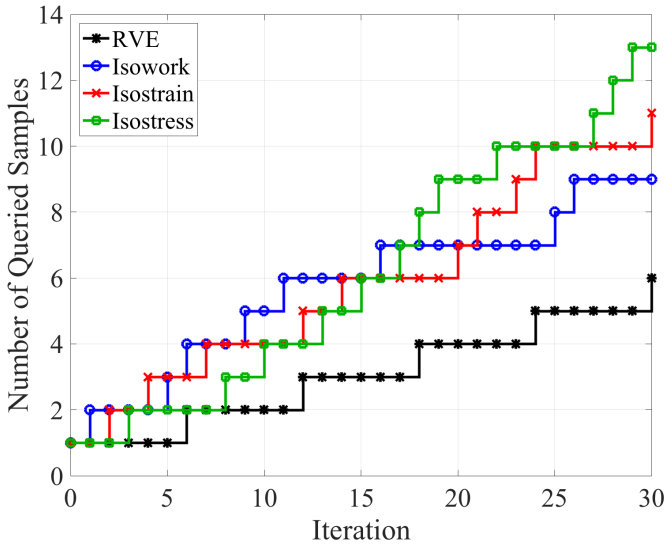


Figure 2.15: Number of samples queried from the true model (RVE) and the information sources in each iteration.

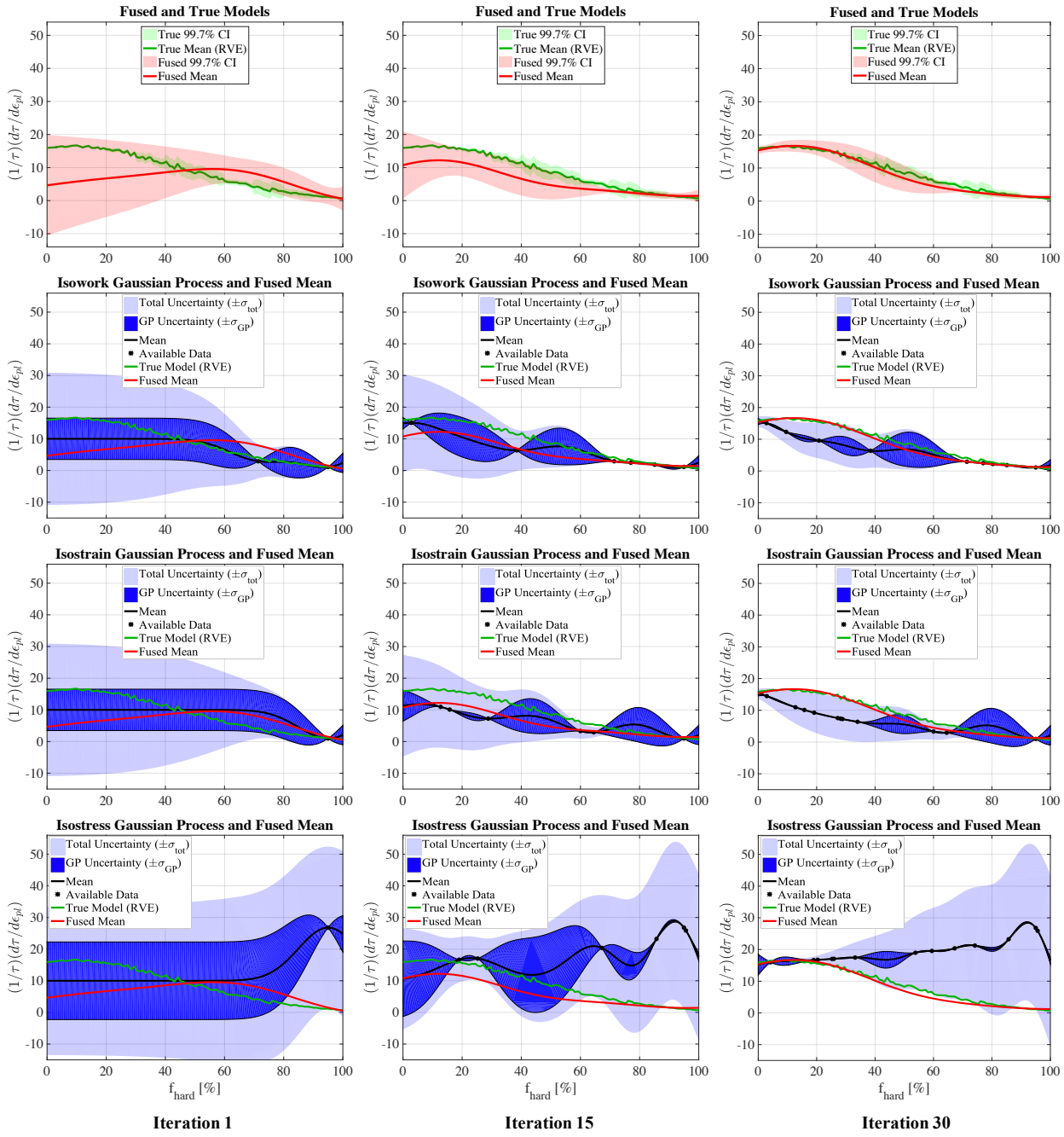


Figure 2.14: The fused model and Gaussian processes of the isowork, isostrain and isostress models in comparison with the true (RVE) model in iterations 1, 15 and 30.

3. BAYESIAN OPTIMIZATION IN LARGE-SCALE SYSTEMS*

3.1 Overview

Available computational models for many engineering design applications are both expensive and black-box. This renders traditional optimization techniques difficult to apply, including gradient-based optimization and expensive heuristic approaches. For such situations, Bayesian global optimization approaches, that both explore and exploit a true function while building a metamodel of it, are applied. These methods often rely on a set of alternative candidate designs over which a querying policy is designed to search. For even modestly high dimensional problems, such an alternative set approach can be computationally intractable, due to the reliance on excessive exploration of the design space. To overcome this, we have developed a framework for the optimization of expensive black-box models, which is based on active subspace exploitation and a two-step knowledge gradient policy. We demonstrate our approach on three benchmark problems and a practical aerostructural wing design problem, where our method performs well against traditional direct application of Bayesian global optimization techniques.

3.2 Introduction

For many engineering design applications, available computational models are expensive and of a black-box nature. In such situations, Bayesian global optimization techniques, such as efficient global optimization [57–59], sequential Kriging optimization [60–62], value-based global optimization [63–66], and the knowledge gradient [38, 67, 68] can be very effective. These techniques generally rely on the simultaneous learning and optimization of a metamodel, or surrogate, of the true model. Often, a querying policy is implemented over a set of alternatives to explore and exploit information from the true function. In many cases, the size of this alternatives set increases exponentially with the input dimension of the true function. This renders even modestly

*Reprinted with permission from "Adaptive Dimensionality Reduction for Fast Sequential Optimization with Gaussian Processes" by S. F. Ghoreishi, S. Friedman, and D. Allaire., 2019. Journal of Mechanical Design, Copyright 2019 by ASME.

high dimensional problems intractable, and requires recourse to gradient-based Bayesian global optimization approaches, which can be susceptible to local minima. To better understand the challenges in high-dimensional problems, consider a d-dimensional design space where a simple discretization of each dimension to n points, results in n^d different alternative samples for the exploration process. This number increases to n^{d+m} in a $(d+m)$-dimensional problem. This exponential growth in the alternatives set makes the optimization process slow or intractable, limiting the applicability of alternative-based Bayesian optimization techniques to problems with less than 4 or 5 dimensions.

To enable the use of alternatives-based Bayesian global optimization for the optimization of expensive black-box functions, we develop here a novel two-step knowledge gradient policy that exploits potential active subspaces of a given true function. In our approach, with previously queried data, the active subspace method [69] is used to map the problem to one of smaller dimension based on directions of greatest variability of the true function. Exponentially fewer alternatives are required on this smaller dimensional problem and the knowledge gradient can readily be applied. The solution to the problem on the active subspace is then mapped to the original space where a second knowledge gradient is applied to a hyperplane orthogonal to the active subspace. The result is a method that enables the application of the alternatives-based knowledge gradient policy to moderately high dimensional problems. Further, due to the fact that the initial knowledge gradient step is focused in directions defined by the largest variability of the true function, our method is significantly more efficient than direct application of the knowledge gradient to the true function. That is, exploiting the active subspace, when one exists, can lead to significant gains in terms of iterations required for convergence.

3.3 Background

Optimization of expensive to evaluate black-box models is often made tractable by the incorporation of lower fidelity metamodels, or surrogate models. These methods have seen significant application in engineering design optimization. Most of these approaches build corrections to low-fidelity information from higher fidelity sources, such as adding global response surface corrections

to low-fidelity models [70–72], using low-fidelity information for coarse-grained search while using high-fidelity function values for fine grained decisions [73–76], creating a response surface using both high- and low-fidelity results [77, 78], and running higher-fidelity models when two or more lower-fidelity models disagree [79, 80]. More formal multifidelity optimization frameworks use either a local approach, such as trust region model management and other surrogate management techniques [81–88], or a global approach constructed via interpolation of the high-fidelity objective function. For example, efficient global optimization (EGO), sequential Kriging optimization (SKO), and knowledge gradient (KG) optimization use a Gaussian process model to estimate the location of high-fidelity optima and guide multifidelity sampling [38, 57, 60, 89–92].

Another common means of optimizing expensive to evaluate problems is the use of either local or global sensitivity analysis. In local sensitivity analysis, gradient information is obtained (e.g., using finite differences for black-box functions). This information can then be used to inform the search direction and plays a significant role in several of the surrogate model management frameworks noted above. For certain optimization methodologies, the dimension of the design vector is a key contributor to computational expense. For example, in many Bayesian global optimization techniques, such as EGO and KG, an optimization problem should be solved over the alternative set according to an acquisition function over the surrogate model. This can be computationally expensive or intractable for large alternative sets. To handle this situation, methods for reducing the dimension of the design vector in the alternative set are desired. Global sensitivity analysis provides a well-known technique for dimension reduction that maintains the original design variables. By considering total effect sensitivity indices [93–95], design variables that do not play a significant role can be fixed to a nominal value. However, in situations where effective lower dimensional approximations do not align well with the design variable directions, alternative subspace approximation techniques can be employed. It should be noted that while global or local sensitivity information can be used for dimensionality reduction, the process of obtaining this global/local information is still affected by the curse of dimensionality, since it is based on sampling of the black box model or of a surrogate constructed from samples of it. Hence, the problems that can be

solved by sensitivity analysis techniques are still limited by the size of the original design space.

Subspace approximation approaches are widely used in optimization [96], model reduction [97], optimal control [98], and other tasks. The proper orthogonal decomposition technique, also known as principal component analysis [99], has been developed to reduce the dimension of the state space of high-order systems. In this method, data are projected to a new coordinate system defined by principal components. Singular value decomposition is a common approach for selection of dimensions using principal component analysis [100]. These techniques are used in modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions and are surveyed in Refs. [101–103]. These techniques are typically used to either reduce the dimension of the output space, for example the objectives of a multidimensional optimization; or the dimension of an input space that has been conditioned by some process, for example on the optimized samples on a pareto-front [104].

A recent approach to subspace approximation is the Active Subspace method [69, 105, 106]. In this method, the directions in which a function has the largest variability are detected to construct an approximate model in a low-dimensional subspace of the function's input design space. This is achieved using first order derivatives. These derivatives can be computed by several techniques, such as adjoint methods [107, 108] and algorithmic differentiation [109]. Upon computation of the function's gradients at a set of input design points, the active subspace technique detects the important directions, followed by rotating the input design space in the directions in which the function has the highest variability. Then, the input design space is projected to the low-dimensional subspace where the function is then approximated. Following Ref. [69], this low-dimensional subspace is called the active subspace [105]. This approach has been applied to design optimization [110, 111], inverse analysis [112], spatial sensitivity analysis [113], aerospace shape optimization [104], and uncertainty quantification for multiphysics scramjet models [114].

In the following section, we state the problem we seek to solve. Our approach to solving this problem then follows. This approach brings to bear the tools and techniques of Bayesian global optimization with the active subspace method in a novel manner that enables more efficient

optimization of expensive black-box models when an active subspace is present.

3.4 Problem Statement

We consider the problem of constrained optimization of an expensive to evaluate and black-box objective function $f(\mathbf{x})$ with design space, $\chi \subseteq \mathbb{R}^m$. The specific problem is to find a design according to

$$\begin{aligned} \mathbf{x}^* &= \arg\max_{\mathbf{x}\in\chi} f(\mathbf{x}) \\ \text{s.t.} \quad & c_j(\mathbf{x}) \leq 0, \qquad j = 1, 2, \ldots, s, \end{aligned} \tag{3.1}$$

where $\mathbf{x}$ is a set of design variables in the input design space χ, and $c_j(\mathbf{x})$, where $j = 1, 2, \ldots, s$, are a set of constraints that must be satisfied. These constraints are also assumed to be expensive to evaluate black-box functions.

For problems such as that given by Equation 3.1, it is common to employ sequential querying policies built off of learned surrogate models as part of a Bayesian global optimization process designed to both explore and exploit based on learned function values. For this, there are two traditional techniques for choosing what to query next [37]. These are efficient global optimization [57] and its extensions, such as sequential Kriging optimization [115] and value-based global optimization [63], and the knowledge gradient [27, 32, 38, 116, 117]. EGO uses a Gaussian process but assumes no noise [118, 119]. SKO also uses Gaussian processes, but includes a tunable weighting factor to lean towards decisions with higher uncertainty [37]. KG uses Gaussian processes and differs from EGO in that the overall best objective value, as determined by a current learned surrogate is used, rather than a best queried value. In Ref. [40] the superiority of the KG policy was demonstrated on several benchmark functions.

The KG policy, as with other Bayesian optimization techniques, performs poorly in systems with large input design spaces. This is generally due to the scaling of the alternatives set with input dimension. In other words, Bayesian optimization techniques rely on proper exploration of the design space at the beginning of the optimization process, which gets exponentially more expensive as the dimension of the input space increases. Thus, in this paper, we develop an approach based

on the active subspace method and a two-step knowledge gradient technique to adaptively reduce the dimension of the input design space and boost the speed of the optimization process. In the following subsections, the key ingredients of our approach which include the active subspace, Gaussian process regression, and the knowledge gradient policy are discussed.

3.4.1 Active Subspace Method

The active subspace method is a technique for discovering the directions of largest variability of a function. Once discovered, an approximation of the function can then be constructed on a lower dimensional subspace defined by these directions. The result is the potential for learning a subspace with significantly lower dimension than that of the original problem [69]. This potential, when it exists, can be exploited to create significant efficiency gains in the application of knowledge gradient policies to expensive black-box optimization problems.

Following Ref. [105], let f be a scalar objective function of m input variables $\mathbf{x}$ in the input design space χ, and $\nabla_{\mathbf{x}} f$ be the column vector of gradient of f at design point $\mathbf{x}$. That is

$$f = f(\mathbf{x}), \quad \nabla_{\mathbf{x}} f = \nabla_{\mathbf{x}} f(\mathbf{x}), \quad \mathbf{x} \in \chi. \tag{3.2}$$

The goal is to find an n-dimensional, $n < m$, active subspace of the input variables that contains most of the variability of the objective function, and a function $g : \mathbb{R}^n \to \mathbb{R}$ that is the approximate representation of f in the active subspace.

The first step is estimating the covariance of the gradient by computing the expectation over the probability density function of $\mathbf{x}$ on χ. We note here that the use of concepts from probability is entirely for convenience. There are no requirements of stochasticity in the underlying variables or functions. Thus, the expectation is usually taken considering uniformly distributed design variables. The $m \times m$ covariance matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = \mathbb{E}[\nabla_{\mathbf{x}} f(\mathbf{x}) \nabla_{\mathbf{x}} f(\mathbf{x})^{\top}]. \tag{3.3}$$

The exact computation of the covariance matrix in Equation 3.3 is not possible in most problems, especially in the case of black-box functions. However, it can be approximated using Monte Carlo methods. This can be achieved by first drawing M samples in the design space and computing the gradient values at those points (e.g., using finite differences). Then the covariance matrix can be estimated as

$$\mathbf{C} \approx \frac{1}{M} \sum_{i=1}^{M} \nabla_{\mathbf{x}} f(\mathbf{x}_i) \nabla_{\mathbf{x}} f(\mathbf{x}_i)^{\top}. \tag{3.4}$$

In order to identify the most effective directions of the input design space, the eigenvectors of the covariance matrix, which is a symmetric and positive semidefinite matrix, are computed. The covariance matrix can then be written based on the eigenvalue decomposition as

$$\mathbf{C} = \mathbf{W} \boldsymbol{\Lambda} \mathbf{W}^{\top}, \tag{3.5}$$

where $\mathbf{W}$ is a $m \times m$ column matrix of eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues. Note that here, the eigenvalues and eigenvectors are placed in descending order. The first n eigenvectors are then selected to form a reduced-order basis. This partitions the eigenvectors and eigenvalues as

$$\mathbf{W} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & \\ & \boldsymbol{\Lambda}_2 \end{bmatrix}, \tag{3.6}$$

where $\mathbf{U}$ contains the first n columns of $\mathbf{W}$, and defines the active subspace of the input design space. Now, the original full space can be transferred to the active subspace as

$$\mathbf{z} = \mathbf{U}^{\top} \mathbf{x}, \tag{3.7}$$

and the function f can be approximated in this active subspace as

$$f(\mathbf{x}) \approx g(\mathbf{U}^{\top} \mathbf{x}) = g(\mathbf{z}), \tag{3.8}$$

where the domain of g is

$$\mathcal{Z} = \{\mathbf{z} = \mathbf{U}^\top \mathbf{x},\ \ \mathbf{x} \in \chi\} \subset \mathbb{R}^n. \tag{3.9}$$

3.4.2 Gaussian Process Regression

Gaussian processes are powerful statistical tools for probabilistic modeling purposes. A Gaussian process can be thought of as a generalized version of the Gaussian distribution applied over a continuous input space. In other words, it is an infinite-dimensional normal distribution where each sample in the input space has a corresponding normal distribution that is characterized by mean and covariance functions [41]. This class of models is widely used in engineering applications due to its flexibility, the ability to incorporate prior knowledge, and the ability to work with small sample sizes.

Gaussian process regression is a non-parametric Bayesian approach that conditions a probabilistic function to training data. Following Ref. [41], Gaussian process regression is approached by conditioning a multivariate normal distribution as

$$f \sim \mathcal{N}\left(\mu, \Sigma\right), \tag{3.10}$$

where f is a normally distributed function with mean μ and covariance matrix Σ. Assuming that N training data are available, represented by $\mathbf{X}_N = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ and $\mathbf{y}_N = (y_1, \ldots, y_N)$ as input and output samples respectively, the posterior distribution of f at any design point $\mathbf{x}$ in the input design space is given as

$$f(\mathbf{x}) \mid \mathbf{X}_N, \mathbf{y}_N \sim \mathcal{N}\left(\mu(\mathbf{x}), \sigma^2(\mathbf{x})\right), \tag{3.11}$$

where

$$\mu(\mathbf{x}) = K(\mathbf{X}_N, \mathbf{x})^T [K(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 I]^{-1} \mathbf{y}_N, \tag{3.12}$$

$$\sigma^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - K(\mathbf{X}_N, \mathbf{x})^T [K(\mathbf{X}_N, \mathbf{X}_N) + \sigma_n^2 I]^{-1} K(\mathbf{X}_N, \mathbf{x}), \tag{3.13}$$

where k is a real-valued kernel function over the input space, $K(\mathbf{X}_N, \mathbf{X}_N)$ is the $N \times N$ matrix

whose m, n entry is $k(\mathbf{x}_m, \mathbf{x}_n)$, and $K(\mathbf{X}_N, \mathbf{x})$ is the $N \times 1$ vector whose m^{th} entry is $k(\mathbf{x}_m, \mathbf{x})$. Note that the term $\sigma_{n,i}^2$ can be used to model observation error and can also be used to guard against numerical ill-conditioning. In this paper, the following exponential kernel function has been employed

$$k(\mathbf{x}, \mathbf{x}') \; = \; \sigma_f^2 \exp\left(\frac{-||\mathbf{x} - \mathbf{x}'||_2^2}{2l^2}\right), \tag{3.14}$$

where $||.||_2^2$ is square of the L_2-norm, σ_f^2 determines the prior variance and l denotes the characteristic length-scale. The parameters of the Gaussian process, i.e., σ_f^2, l and σ_n^2, can be updated at each time new information is obtained by using a maximum likelihood approach or Bayesian techniques [41].

3.4.3 Knowledge Gradient

Given a Gaussian process representation of the underlying function to be optimized, the next requirement is the decision on where to query next. For this, we incorporate the knowledge gradient policy. Let $\{\mathbf{x}_{1:N}, y_{1:N}\}$ be the set of design points and the corresponding objective values that have been used to construct the Gaussian process of the objective function. Let $f(\mathbf{x})$ represent this posterior distribution of the function given this available information. The best expected objective value can be computed as

$$f_N^* = \max_{\mathbf{x} \in \chi} \mathbb{E}[f(\mathbf{x}) \mid \mathbf{x}_{1:N}, y_{1:N}]. \tag{3.15}$$

Similarly, if one additional design point can be queried to update the posterior distribution of the model, the best expected objective value would be

$$f_{N+1}^* = \max_{\mathbf{x} \in \chi} \mathbb{E}[f(\mathbf{x}) \mid \mathbf{x}_{1:N+1}, y_{1:N+1}]. \tag{3.16}$$

The difference $f_{N+1}^* - f_N^*$ specifies the improvement in value of the function resulting from the additional query. The idea is to select a design point to query that maximizes this improvement. Since the Gaussian process is the probabilistic representation of the objective function, when choosing $\mathbf{x}_{N+1}$, there is stochasticity in the value of the objective function, i.e., y_{N+1}, upon querying $\mathbf{x}_{N+1}$.

Thus, one needs to compute the expected value of improvement using the posterior predictive distribution of the objective function.

The knowledge gradient technique [31, 37–39, 120] is a method for selecting the design point that maximizes the expected increase in the objective value. Letting $S^N = \mathbb{E}[f(\mathbf{x}) \mid \mathbf{x}_{1:N}, y_{1:N}]$ be the knowledge state, the value of being at state S^N is defined as $V^N(S^N) = \max_{\mathbf{x}\in\chi} S^N$. The knowledge gradient which is a measure of expected improvement, if the design point $\mathbf{x}$ would be queried at the next time step, can be defined as

$$\nu_{\mathbf{x}}^{KG,N} = \mathbb{E}[V^{N+1}\left(S^{N+1}(\mathbf{x})\right) - V^N(S^N) \mid S^N], \tag{3.17}$$

where the expectation is taken over the stochasticity in the posterior distribution of the objective at design point $\mathbf{x}$, i.e., $f(\mathbf{x})$.

Now, let χ_{alt} be the alternative set, which denotes a finite set of design samples. The knowledge gradient policy for sequentially choosing the next query is then given as

$$\mathbf{x}^{KG,N} = \arg\max_{\mathbf{x}\in\chi_{\text{alt}}} \nu_{\mathbf{x}}^{KG,N}. \tag{3.18}$$

Calculation of the knowledge gradient, which is based on a piecewise linear function, is discussed in detail in two algorithms presented in Ref. [40].

3.5 Approach

Our approach to solving Equation 3.1 is based on bringing together the ideas of Bayesian global optimization, via the knowledge gradient, and subspace approximation, via the active subspace. From a general standpoint, the approach alternates between applying the knowledge gradient on a learned active subspace and applying the knowledge gradient on a hyperplane of the original design space (and orthogonal to the active subspace). The result is an efficient method for the solution of Equation 3.1 when an active subspace exists. We stress here that the learned active subspace itself is adaptive in our methodology, allowing for rapid initial progress followed by better and better

approximation of the true underlying subspace. The details of our approach are provided in the following paragraphs. The section concludes with an algorithm for implementing our approach and a flowchart depicting the main steps.

The approach starts by constructing Gaussian processes for the objective function, $f(\mathbf{x})$, and the constraints, $c_j(\mathbf{x})$, $j = 1, \ldots, s$, in the original m-dimensional input design space based on their available data. As discussed in Section 3.4.2, these Gaussian processes are given as

$$f(\mathbf{x}) \mid \mathbf{X}_N, \mathbf{y}_N \sim \mathcal{N}\left(\mu_h(\mathbf{x}), \sigma_h^2(\mathbf{x})\right), \tag{3.19}$$

$$c_j(\mathbf{x}) \mid \mathbf{X}_{N_j}, \mathbf{c}_{N_j} \sim \mathcal{N}\left(\mu_{c_j}(\mathbf{x}), \sigma_{c_j}^2(\mathbf{x})\right), \tag{3.20}$$

where μ_h and σ_h^2 are the mean and variance of the objective function at design point $\mathbf{x}$ in the original high-dimensional space, and μ_{c_j} and $\sigma_{c_j}^2$ are those of constraint j given the available data $\mathbf{X}_{N_j}$ and $\mathbf{c}_{N_j}$.

After constructing the Gaussian processes of the objective function and constraints, the next step is to choose the next design point to query. Here, this is achieved via a two-step knowledge gradient process. Rather than directly applying the knowledge gradient on the potentially high dimensional original space, we first map the original space to an active subspace. In this lower dimensional space we apply the knowledge gradient policy. Since this lower dimensional space is an active subspace, useful objective function improvement is expected at significantly less computational expense. The active subspace itself is found by evaluating the gradient of the function at the N currently available samples. This can be done through finite differences with the current Gaussian process or other means that may be available for a given function. Then, the covariance matrix is estimated according to Equation 3.4 and the normalized eigenvalues are computed for this matrix. The eigenvectors associated to the normalized eigenvalues larger than a user defined threshold specify the transformation matrix $\mathbf{U}$ to find the active subspace. Notice that this threshold should take value between 0 and 1. Setting larger values for this threshold results in more dimensionality reduction, whereas smaller values are associated with active subspaces with a number of

dimensions closer to the original space. In practice, this value could be chosen adaptively during the optimization process. This could be done in such a way that larger values are selected at the beginning of the process for fast exploration, and as more queries are made and the covariance matrix of gradients is better approximated, smaller values of the threshold are selected to avoid unnecessary reduction and error in the optimization process.

After identifying the active subspace, the N available input vectors are projected onto the subspace as

$$\mathbf{Z}_N = \mathbf{U}_N^\top \mathbf{X}_N, \tag{3.21}$$

where $\mathbf{Z}_N$ are the input vectors in the low-dimensional space, $\mathbf{U}_N$ is the active subspace, and $\mathbf{X}_N$ are the available input vectors in the original space. We note again that this active subspace is adaptive. In the initial stages of the approach, the covariance matrix might be estimated inaccurately due to its sensitivity to the initial quality/accuracy of the surrogate model and its correlation parameters, which can result in inaccurate detection of the active subspace in early stages. As the Gaussian process of the objective function is updated, a new active subspace is computed. Thus, as the more points are queried, the approximation of the active subspace becomes closer to the true active subspace as a result of the Gaussian process representation of the function becoming closer to the true underlying function.

After the original space has been mapped to the active subspace, we have the lower dimensional representation of the previously queried points according to Equation 3.21. From these data, we construct a Gaussian process in the active subspace as

$$g(\mathbf{z}) \mid \mathbf{Z}_N, \mathbf{y}_N \sim \mathcal{N}\left(\mu_l(\mathbf{z}), \sigma_l^2(\mathbf{z})\right), \tag{3.22}$$

where g is the posterior distribution of the objective function in the low-dimensional space, and μ_l and σ_l^2 are the mean and variance of the objective function at the design point $\mathbf{z}$ in this space respectively. The next step of the approach is then to use the knowledge gradient to find the next best point to query in the active subspace. For this, we generate Latin Hypercube samples in the

current active subspace, which we denote as the alternative set $\mathbf{Z}_f$. Among these alternatives, we select the one that leads to the maximum knowledge gradient in the current active subspace, as

$$\mathbf{z}_{N+1} = \arg\max_{\mathbf{z}\in\mathbf{Z}_f} \nu_{\mathbf{z}}^{KG,N}. \tag{3.23}$$

Upon selection of the best point in the current active subspace, this sample, $\mathbf{z}_{N+1}$, needs to be mapped back to the original input space. For this, we use the following transformation

$$\underbrace{\begin{bmatrix} z_{1N+1} \\ \vdots \\ z_{nN+1} \end{bmatrix}}_{\mathbf{z}_{N+1}} = \underbrace{\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} & \cdots & u_{1m} \\ & & & \vdots & & \\ u_{n1} & u_{n2} & \cdots & u_{nn} & \cdots & u_{nm} \end{bmatrix}}_{\mathbf{U}_N^T} \underbrace{\begin{bmatrix} x_{1N+1} \\ \vdots \\ x_{nN+1} \\ \vdots \\ x_{mN+1} \end{bmatrix}}_{\mathbf{x}_{N+1}}, \tag{3.24}$$

$$\text{s.t.} \quad \mathbf{x}_{N+1} \in \chi,$$

where $\mathbf{x}_{N+1}$ is the solution to the above linear equations. It is easy to verify that the solution to the inverse mapping in Equation 3.24 is not unique and the transformation will thus lead to an infinite set of design points in the original space. To keep the computation tractable, we propose a strategy to choose a single point $\mathbf{x}_{N+1} \in \chi$ given each $\mathbf{z}_{N+1} \in \mathcal{Z}$.

Since the design point needs to be inverse mapped from n dimensions to m dimensions ($n < m$), one needs to discretize $m - n$ arbitrary dimensions of the original space and plug in these discretized values in the equations and solve the n equations to find the remaining n values of $\mathbf{x}_{N+1}$. For simplicity and without loss of generality, N_f samples from the last $m - n$ dimensions are generated using techniques such as Latin Hypercube sampling. Then, the other dimensions of these N samples are obtained by solving Equation 3.24. We note here that constraints are also checked at this point. Specifically, only those samples that satisfy $\mu_{c_j}(\mathbf{x}) - 3\sigma_{c_j}(\mathbf{x}) \leq 0$ for $j = 1, \ldots, s$ are kept. This set is denoted by $\mathbf{X}_f$, and called the inverse-mapped set. The constraint

handling strategy follows that of Ref. [31], where it was found to have good performance. Since Gaussian process variances are large in early iterations, it is easy to satisfy the probabilistic constraints, which enables good exploration potential. As the process moves forward, the variances are reduced via learning, and the probabilistic constraints converge to the true constraints given by Equation 3.1. In general, a practitioner can control the level of acceptability in terms of constraint violation by modifying the coefficient of $\sigma_{c_j}(\mathbf{x})$ as deemed appropriate.

A depiction of the proposed inverse mapping process is shown in Fig. 3.1. In this figure, the

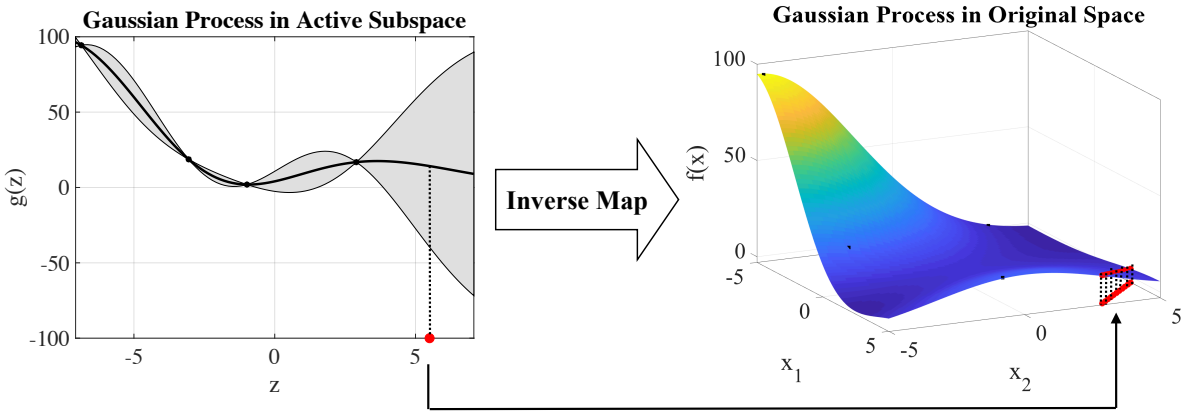


Figure 3.1: A depiction of the inverse mapping process.

left plot shows the Gaussian process of an objective function in its active subspace and the right plot shows the Gaussian process in the original space. In the left plot, the red circle shows the selected point, i.e. $\mathbf{z}_{N+1}$, by the first-step knowledge gradient, and the red samples in the right plot show the inverse-mapped set. Now, among all the samples in the inverse-mapped set, the best design point to query needs to be selected (i.e., $\mathbf{x}_{N+1} \in \mathbf{X}_f$). It is desired to select a sample that yields the highest expected improvement in the original space. We propose to select this sample by applying the knowledge gradient policy to the inverse-mapped set in the original space.

Applying the knowledge gradient in the whole original space leads to poor performance in selection of the best design point to query. This is due to the fact that the alternative set should

be enumerated in the high-dimensional original space. This number of alternatives in this set will suffer from the curse of dimensionality. Further, if there is an active subspace, then exploring certain directions of the high-dimensional original space may be inefficient due to a lack of variability along those directions. We overcome this issue in our approach by applying two knowledge gradient steps. The first on the active subspace and the second orthogonal to the active subspace in the original space. Thus, the alternative set in the original, high-dimensional space can be enumerated on a hyperplane of the original space, as shown by the red samples in the right plot of Fig. 3.1. Now, performing the knowledge gradient leads to selection of a sample with the highest expected improvement as

$$\mathbf{x}_{N+1} = \arg\max_{\mathbf{x}\in\mathbf{X}_f} \nu_{\mathbf{x}}^{KG,N}. \tag{3.25}$$

where $\nu_{\mathbf{x}}^{KG,N}$ is defined in Equation 3.17.

After observing y_{N+1} corresponding to the selected design point $\mathbf{x}_{N+1}$, the Gaussian processes of the high-dimensional objective function and all the constraints are updated based on $\mathbf{X}_{N+1}$, $\mathbf{y}_{N+1}$ and $\mathbf{c}_{N+1_j}$. Afterwards, the active subspace method is applied to find the new low-dimensional subspace according to the current updated knowledge about the objective function, and the entire process is repeated. Thus, our approach is adaptive in finding the active subspace in each iteration. The process continues by finding the next design point to query, and repeats until a termination criterion, such as exhaustion of the querying budget, is met. The final solution to Equation 3.1 is then found from the current Gaussian process in the original space. Our proposed approach for adaptive dimensionality reduction for fast sequential optimization with Gaussian processes is presented in Algorithm 1, and Fig. 3.2 presents a schematic diagram of our approach.

3.6 Benchmark Applications

In this section, we present the key features of our proposed sequential adaptive dimensionality reduction approach for fast optimization of expensive black-box functions. We focus here on benchmark problems designed to highlight the strengths and limitations of the approach. We begin with an analytic two-dimensional constrained optimization problem, which we use to demonstrate

Algorithm 1 Adaptive Dimensionality Reduction for Fast Sequential Optimization with Gaussian Processes

1: Construct Gaussian processes for the objective function and constraints in the original space χ.

repeat

- 2: Find the active subspace corresponding to the normalized eigenvalues of the covariance matrix in Equation 3.4 larger than a user defined threshold.
- 3: Transform the data available in the original space to the active subspace.
- 4: Construct the Gaussian process of the objective function in the active subspace $\mathcal{Z}$.
- 5: Generate Latin Hypercube samples in the active subspace $\mathcal{Z}$.
- 6: Apply the first-step knowledge gradient to select a design point in the active subspace.
- 7: Inverse map the selected design point to the original space according to Equation 3.24.
- 8: Apply the second-step knowledge gradient to select the best design point in the original space according to Equation 3.25.
- 9: Update the Gaussian processes of the constraints and the objective function in the original space based on the observations obtained at the selected design sample.

until termination

10: Return the feasible design point with the largest estimated objective value according to the Gaussian processes of the constraints and the objective function in the original space.

the effectiveness of our approach when an active subspace exists. We follow that demonstration with the application of our methodology to standard benchmark problems from the literature to reveal the behavior of our approach when an active subspace does not exist.

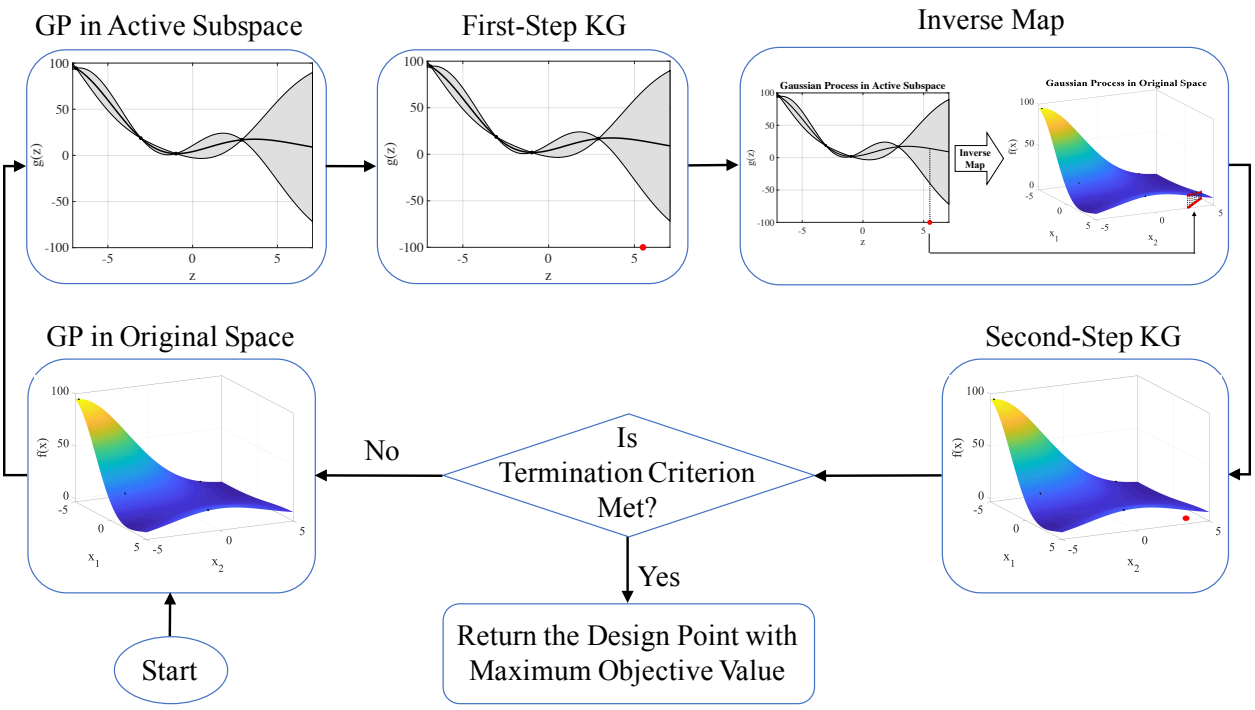


Figure 3.2: A depiction of our proposed approach for fast sequential optimization. “GP" stands for Gaussian process and “KG" stands for knowledge gradient.

3.6.1 Two-Dimensional Function

We first consider the constrained maximization of an analytic two-dimensional function. The problem is given as

$$\begin{aligned} \mathbf{x}^* = \underset{\mathbf{x}\in[-5,\,5]^2}{\operatorname{argmax}}\ & x_1^2 + x_2^2 + 2x_1x_2 \\ \text{s.t.}\quad & x_1 + x_2 < 6. \end{aligned} \tag{3.26}$$

The feasible optimal solution for this problem is $\mathbf{x}^* = (-5, -5)$ with the optimal objective value $f(\mathbf{x}^*) = 100$. A depiction of this example in the original two-dimensional space is shown in Figure 3.3, where the plane separating the feasible and infeasible regions is shown in red. Figure 3.4 represents the transformed objective function to its active subspace, which is a one-dimensional space for this function. We note that these two figures are shown for the sake of visualization, and these are assumed to be unknown prior to applying our approach.

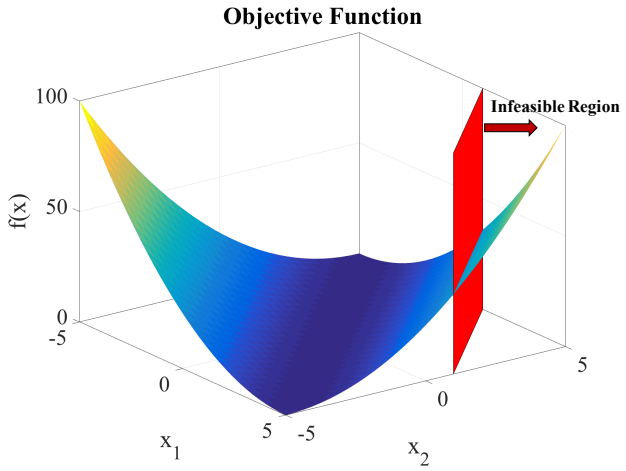


Figure 3.3: A depiction of the two-dimensional function in Equation 3.26

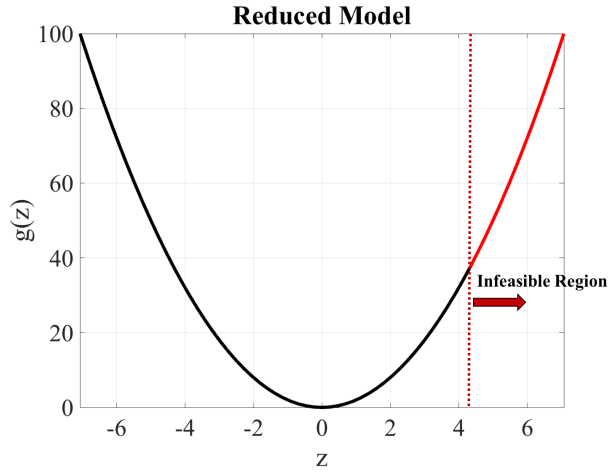


Figure 3.4: A depiction of the function of Equation 3.26 in its one-dimensional active subspace

Figure 3.5 demonstrates the learned Gaussian processes for this problem both with and without use of the active subspace. Each column of the figure represents an iteration (here the third, sixth, and ninth iterations of the optimization process). The top row provides the Gaussian process representation in the active subspace. The middle row shows the Gaussian process in the original space when our approach (exploiting the active subspace) is utilized. The bottom row shows the

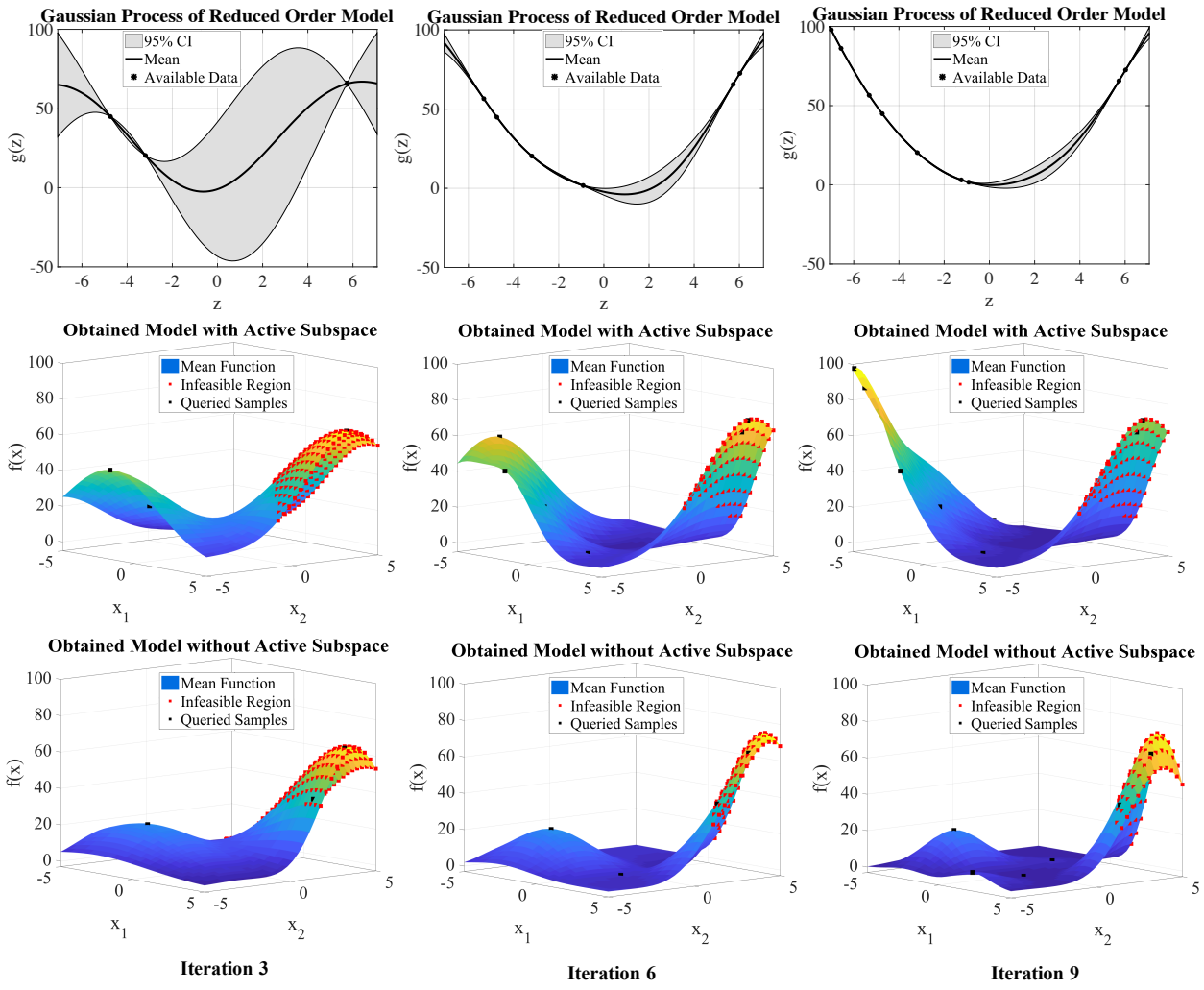


Figure 3.5: Gaussian processes of the objective function in the original two-dimensional space obtained by our active subspace exploiting approach and via direct knowledge gradient application to the full model.

Gaussian process in the original space when the knowledge gradient is applied directly to the full model. From the figure it is clear that the use of the active subspace significantly improves the Gaussian process representation in the original space. This is the result of more efficient use of queries to the full model that were made possible through identification of the active subspace for this problem.

3.6.2 Benchmarks with and without Active Subspaces

In this subsection, we focus on the situation where a function we seek to optimize does not have an active subspace. In most situations, given the black-box nature of the functions we seek to optimize, whether or not a given function has an active subspace will be unknown *a priori*. Therefore, it is necessary to understand the behavior of our approach when an active subspace does not exist for a given function. For this, we test our methodology on three unconstrained analytic problems, two of which have a negligible active subspace. The first of the three functions studied

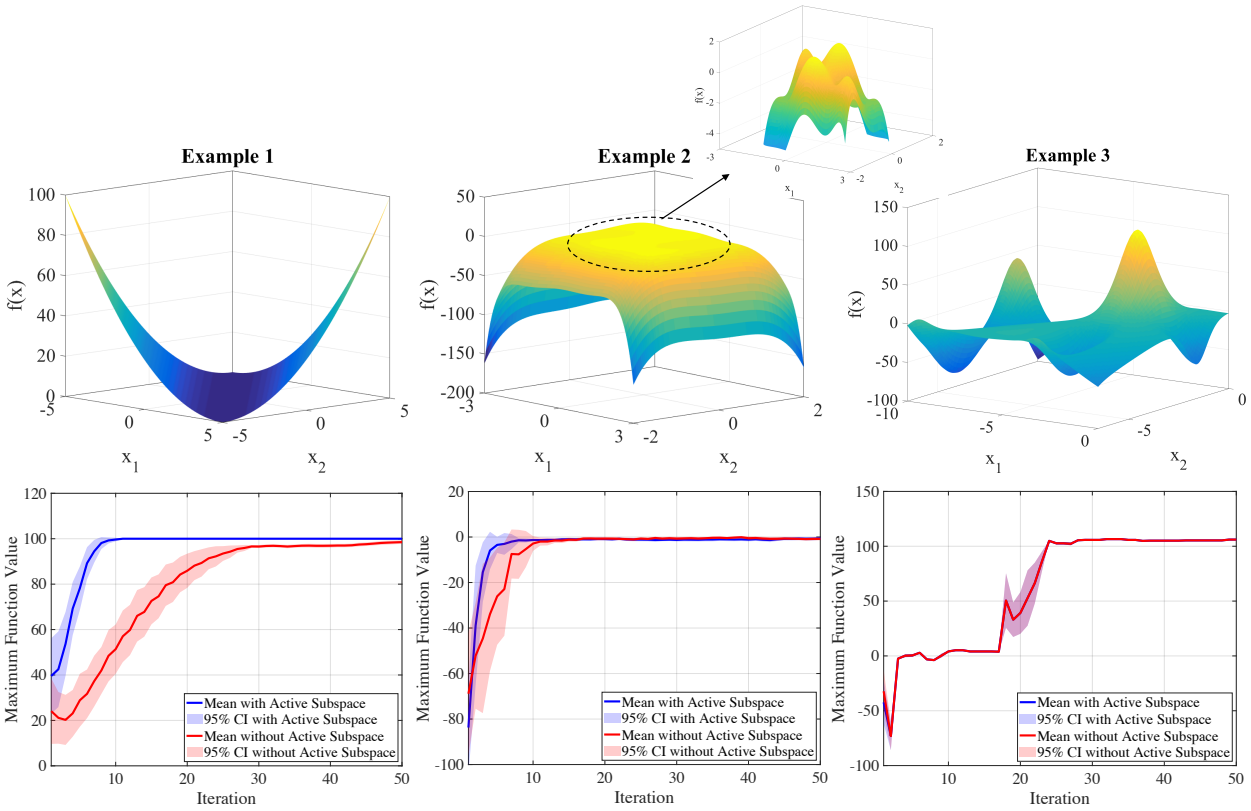


Figure 3.6: The two-dimensional function examples and the mean and 95% confidence interval of the maximum function values obtained by our approach with active subspace exploitation and with direct knowledge gradient application in each iteration over 100 independent runs.

here is the objective function given in the previous subsection and will be denoted as example 1 in what follows. The second (example 2) is the negated six-hump camel back function of Ref. [121].

This function is given as

$$f(\mathbf{x}) = -\Big((4 - 2.1x_1^2 + \frac{x_1^4}{3})x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2\Big), \tag{3.27}$$

where $x_1 \in [-3,\ 3]$ and $x_2 \in [-2,\ 2]$, and the optimal solutions are $\mathbf{x}^* = (0.0898\,,\, -0.7126)$ and $(-0.0898\,,\, 0.7126)$ with the objective value $f(\mathbf{x}^*) = 1.0316$. The third benchmark problem (example 3) is the negated Mishra's Bird function, which can be found in Ref. [122]. This function is given as

$$\begin{aligned} f(\mathbf{x}) = -\Big(&\sin(x_2)\exp[(1-\cos x_1)^2] + \cdots \\ &\cos(x_1)\exp[(1-\sin x_2)^2] + (x_1 - x_2)^2\Big), \end{aligned} \tag{3.28}$$

where $x_1 \in [-10,\ 0]$ and $x_2 \in [-6.5,\ 0]$, and the optimal solution is $\mathbf{x}^* = (-3.1302\,,\, -1.5821)$ with the objective value $f(\mathbf{x}^*) = 106.7645$.

Figure 3.6 shows the objective functions of examples 1, 2 and 3, as well as the mean and 95% confidence interval of the maximum objective function value obtained in each iteration of applying our approach with active subspace exploitation and performing the knowledge gradient method in the original space without the active subspace. The results are obtained over 100 independent runs. As was seen in the previous subsection, the first example has a one-dimensional active subspace. Our approach is able to exploit this and obtain a larger average maximum objective value in each iteration as compared to direct application of the knowledge gradient policy to the original function. In example 2, the function gently varies in all directions and there is no useful active subspace. As shown in the center plots of Fig. 3.6, our approach thus performs similarly to the direct application of the knowledge gradient. We note here, however, that in early iterations some efficiency gains are still had through the use of our method. In the extreme case of example 3, the function varies significantly in all directions. Therefore, there is no possible reduction to a meaningful active subspace. Our approach, with the exception of the first iteration, performs exactly the same as

when the knowledge gradient is applied to the function in the original space. The key takeaway here is that the worst case scenario of applying our method, that is, the situation where there is no active subspace to take advantage of, simply reverts our approach to that of direct knowledge gradient application to the original function. However, as can be seen from the left plots of Fig. 3.6, if there is an active subspace, our approach can exploit it for significant gains in efficiency.

3.6.3 Rosenbrock Function

Here, in order to assess the effectiveness of the proposed method in higher dimensions, we consider the negated 10- and 20-dimensional Rosenbrock function, which can be found in Ref. [122] and is given as

$$f(\mathbf{x}) = -\sum_{i=1}^{N-1}[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2], \tag{3.29}$$

where $x_i \in [-2.048,\ 2.048]$, and the optimal solution is $\mathbf{x}^* = (1\,,\,\ldots\,,\,1)$ with the objective value $f(\mathbf{x}^*) = 0$. Figures 3.7 and 3.8 show the mean and 95% confidence interval of the maximum objective function value obtained in each iteration of applying our approach with active subspace exploitation and performing the knowledge gradient method in the original space without the active subspace. The results are obtained over 100 independent runs. As can be seen, our approach, by exploiting the active subspace, shows significant improvements in performance versus the case without active subspace exploitation. Figures 3.9 and 3.10 show the average dimension of the active subspace as a function of the iteration number. As can be seen, in the initial stages of the approach, the dimension is decreased substantially and as the process goes on, the dimension of the problem remains less than 5 and 9 dimensions on average for 10- and 20-dimensional Rosenbrock function respectively.

3.7 Aerostructural Demonstration Problem

To demonstrate the effectiveness of our approach on a realistic problem, we consider the aerostructural design of an aircraft wing for fuel burn minimization. The aerostructural model of the wing used for this demonstration is based on the NASA Common Research Model (CRM) [123], which is a commonly used representation of a long-range commercial airliner wing operat-

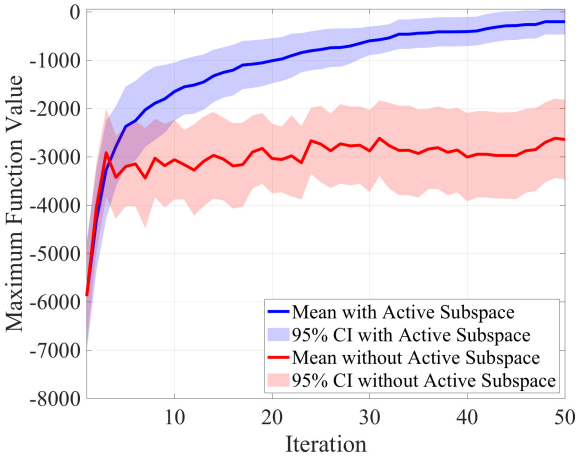


Figure 3.7: The mean and 95% confidence interval of the maximum function values obtained by our approach with active subspace exploitation and without active subspace exploitation in each iteration obtained over 100 independent simulations of the 10-dimensional Rosenbrock function.

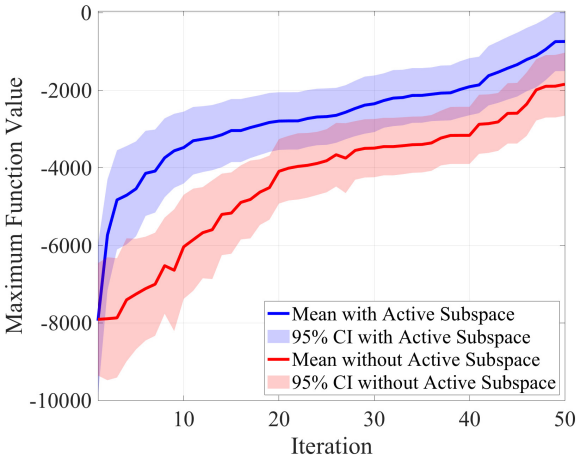


Figure 3.8: The mean and 95% confidence interval of the maximum function values obtained by our approach with active subspace exploitation and without active subspace exploitation in each iteration obtained over 100 independent simulations of the 20-dimensional Rosenbrock function.

ing in transonic flight [124]. The aerostructural model consists of a three-dimensional aerodynamic and structural deformation coupled system. The aerodynamic lift is calculated using a vortex lattice method and the structural deformation of the wing is calculated using a six degree of freedom

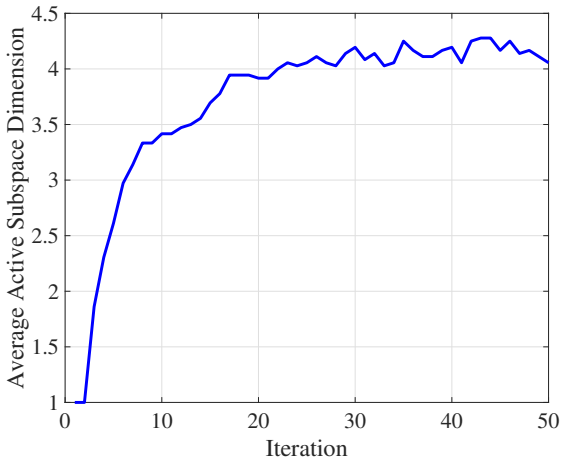


Figure 3.9: Active subspace dimension in each iteration averaged over 100 independent simulations of the 10-dimensional Rosenbrock function.

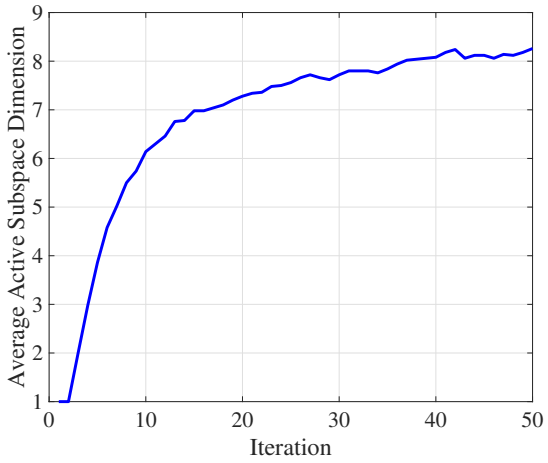


Figure 3.10: Active subspace dimension in each iteration averaged over 100 independent simulations of the 20-dimensional Rosenbrock function.

spatial beam finite element method. The coupled nature of the two subsystems is a consequence of the aerodynamic lifting force deforming the wing geometry, which in turn has a feedback effect on the aerodynamics. This is shown schematically in Fig. 3.11, which is adapted from Ref. [125] and uses the XDSM methodology of Ref. [126]. In the figure, y represents the fuel burn and c_i

represents the constraints. A more detailed schematic of the aerostructural system is presented in Ref. [124]. The model's design variables are used to determine the undeformed wing geometry and

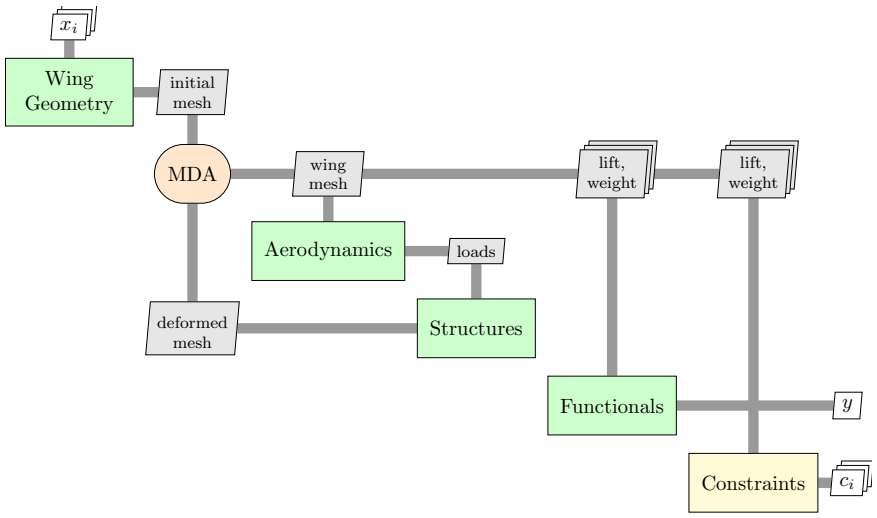


Figure 3.11: A simplified depiction of the aerostructural system

internal spar properties. We start with the CRM wing shape and then vary its twist distribution, taper ratio, and chord ratio, and vary the spar thickness distribution. The aircraft is assumed to be in steady flight at cruising altitude and the spar elements are assumed to be made from aluminum.

The twist and thickness distribution design variables are control points used in a B-spline interpolation to vary the design along the wing's span. We used two control points each for the twist and thickness distributions. The wing's twist effectively changes the wing's angle of attack along its span. The spar thickness affects the wing's weight as well as it's strength. The taper and chord ratios affect the chord length of the wing, which varies linearly along the span of the wing.

The coupled system analysis is conducted using a Gauss-Seidel iterator [127] until the aerodynamic loads and structural deformations converge to a fixed point. After the variables in the coupled system converge, the fuel range is computed along with other performance-based values and problem constraints. The model employs three constraints in the optimization: 1) it is ensured

that stress on structural spars due to aerodynamic loads remain below the yield stress for the spar material properties; 2) it is ensured that the spar thickness does not exceed the wing thickness; and 3) it is assumed that the aircraft is in steady flight by enforcing the total aerodynamic lifting force be equal to the aircraft weight. In this problem, the goal is to minimize the fuel burn subject to these three constraints as

$$\begin{aligned} \mathbf{x}^* = \arg\min_{\mathbf{x}\in\chi} \ & \text{fuel burn}(\mathbf{x}) \\ \text{s.t.} \quad & \text{lift-weight constraint } (L = W) \\ & \text{structural failure constraint} \\ & \text{structural intersection constraint} \end{aligned} \tag{3.30}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \text{Twist Distribution 1} \\ \text{Twist Distribution 2} \\ \text{Thickness Distribution 1} \\ \text{Thickness Distribution 2} \\ \text{Taper Ratio} \\ \text{Chord Ratio} \end{bmatrix}. \tag{3.31}$$

The boundaries of the design variables are listed in Table 3.1.

To identify the true optimum for this problem, we use a gradient-based nonlinear optimizer using sequential least squares programming (SLSQP) [128]. The optimal solution obtained by SLSQP is $\mathbf{x}^* = [12.8037\ ,\ 14.7378\ ,\ 0.0377\ ,\ 0.0718\ ,\ 0.2\ ,\ 0.9]$ with minimum fuel burn equal to 1.0190×10^5. Figure 3.12 shows the mean and 95% confidence interval of the minimum fuel burn obtained by our approach with active subspace exploitation and without active subspace exploitation in each iteration obtained over 100 independent simulations. Our approach, on average, obtains lower values of fuel burn in each iteration compared to the case that does not apply the

Table 3.1: Design variable bounds for the aerostructural problem. LB is the lower bound, UB is the upper bound, and # is the number of variables of a given type.

Design Variable	#	LB	UB
Twist Distribution	2	-15	15
Thickness Distribution	2	0.001	0.25
Taper Ratio	1	0.2	1.5
Chord Ratio	1	0.9	1.1

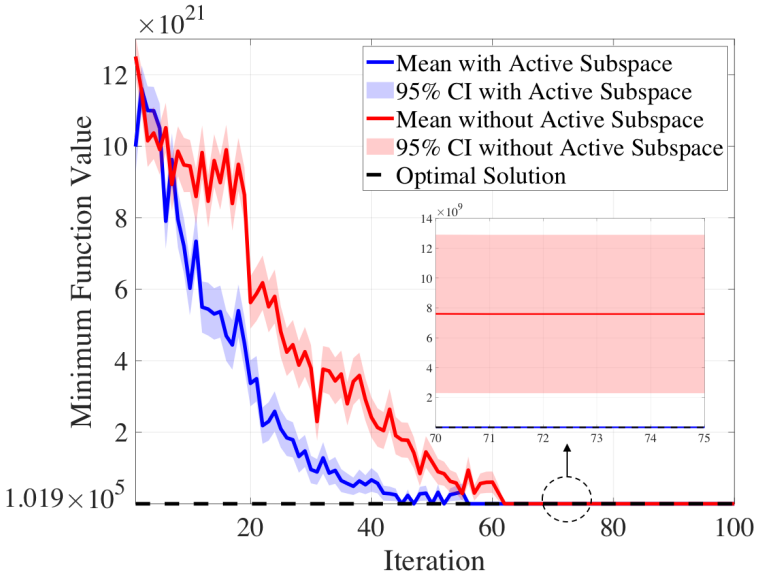


Figure 3.12: The mean and 95% confidence interval of the minimum function values obtained by our approach with active subspace exploitation and without active subspace exploitation in each iteration obtained over 100 independent simulations of the aerostructural problem.

active subspace method. Figure 3.13 shows the mean and 95% confidence interval of the absolute difference between the optimal design variables and the design variables obtained by our approach with active subspace exploitation and without active subspace exploitation in each iteration. These results are obtained over 100 independent runs. From the figure, it is clear that the design variables obtained by our approach are closer to the optimal design variables than the case without applying the active subspace method. Figure 3.14 shows the average dimension of the active subspace as a

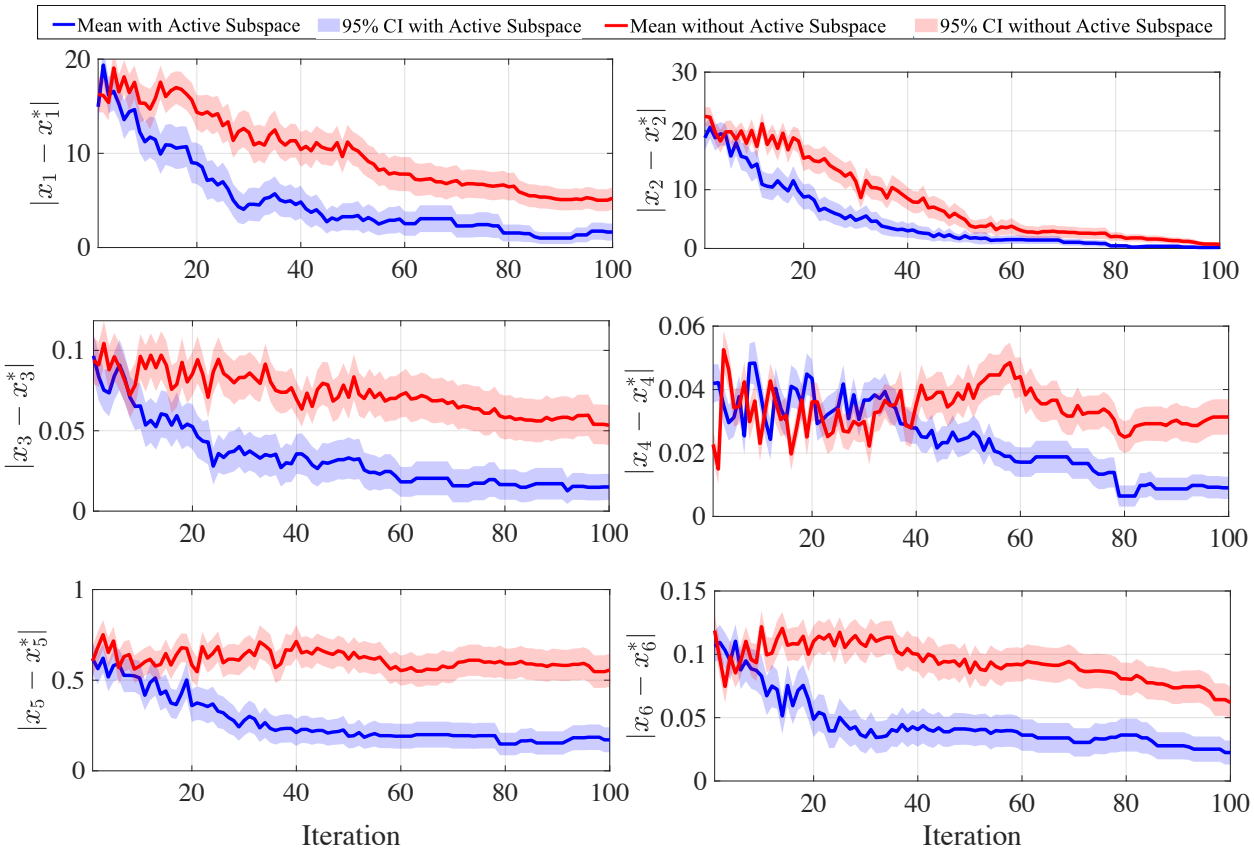


Figure 3.13: The mean and 95% confidence interval of the absolute difference between the optimal design variables and the design variables obtained by our approach with active subspace exploitation and without active subspace exploitation in each iteration obtained over 100 independent simulations of the aerostructural problem.

function of the iteration number. As can be seen, our approach reduces the dimension of the problem from 6 to less than 2 dimensions on average. This reduction enables far more efficient search over alternatives when using a Bayesian global optimization search strategy, as we have done here.

It should be noted that the performance improvement by our approach with active subspace exploitation is achieved by an additional cost for surrogate model fitting for the objective function in the active subspace. In other words, the selection process by our proposed method has approximately two times more complexity than the original Knowledge Gradient method, due to the double selections in the active subspace and the original space. However, the most important part of the computational expense for real-world applications is often the number of black-box function

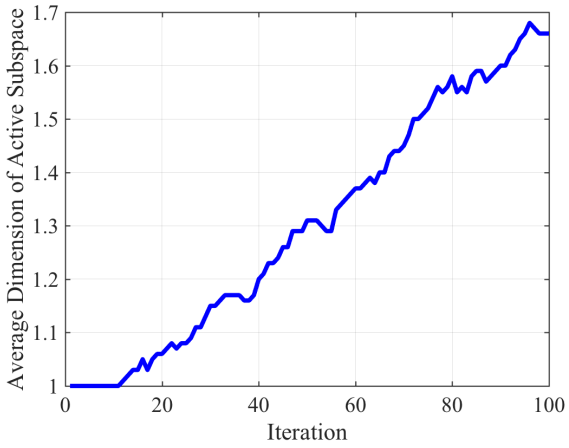


Figure 3.14: Active subspace dimension in each iteration averaged over 100 independent simulations of the aerostructural problem.

evaluations, which is far less in our approach with active subspace exploitation in comparison to the case without active subspace exploitation.

4. CONCLUSIONS AND FUTURE WORK

This dissertation has focused on development of methodologies for Bayesian optimization in multi-information source and large-scale systems.

In section 2, an approach is presented to perform constrained optimization of expensive to evaluate functions when different information sources with varying fidelities and evaluation costs are available for the objective function and constraints. This is achieved by estimating the correlation between the information sources and fusing the information obtained from each information source to construct the fused Gaussian processes. Then, the cost of querying and a two-step look-ahead utility function obtained from the fusion process and the knowledge gradient policy, are incorporated to identify the next design and objective information source to query. The approach also considers the trade off between cost and information gain as quantified by the Kullback-Leibler divergence for identifying design points and information sources for constraint handling. The proposed strategy samples the design space by balancing exploration and exploitation tasks both between and within the available information sources to efficiently move sequentially toward the true, real world optimum of a constrained objective function. We demonstrated our approach on the constrained optimization of a one-dimensional example test problem, an aerodynamic design problem, and the ground truth strength normalized strain hardening rate for a dual-phase material. With these demonstrations, it has been shown that the proposed approach estimates well feasible optimal values of the objective in an efficient decision-theoretic manner which reveals the promise of this framework as a suitable methodology for accelerating the design problems.

In future, the information fusion framework developed here can be validated against larger sets of ground truth data and be demonstrated in higher dimensions. The framework can also be extended to handle information sources with misaligned input-output interfaces, which is a key challenge facing many design problems. Moreover, the optimization framework can be extended to handle multiple objectives and studied for scalability to high dimensional input spaces. Additionally, the possibility of carrying out optimal sequential queries in which the sources of information

are not input/output aligned can be explored. A specific scenario, for example, would be combining sources that establish relationships between processing parameters/conditions and microstructure with sources that connect microstructures to properties/performance. Much remains to be done, but this work presents a plausible research program toward tool (information source) integration.

In section 3, a sequential decision-theoretic approach is presented for fast constrained optimization of problems with several input variables. The general use case is for expensive to evaluate black-box functions. Using the fact that certain directions in the input design space have less impact on the objective function than others, we have developed an adaptive methodology to map the high-dimensional problem to a lower dimension using the active subspace method. The Gaussian process model constructed for the objective function in the low-dimensional space is used in combination with the knowledge gradient approach to select the best sample to query in the reduced domain. Upon selection of the best design point in the low-dimensional space, an inverse mapping process to the original design space is performed. Since the inverse mapping is not one-to-one, the method selects the point which has the maximum knowledge gradient over the inverse-mapped points in the original space. We demonstrated our approach on the optimization of three two-dimensional example test problems, 10- and 20-dimensional Rosenbrock function, and an aerostructural wing design problem. It has been shown that the proposed approach can take advantage of existing active subspaces, leading to more efficient querying. It has also been shown that when no active subspace exists, the approach proposed here performs as well as direct application of Bayesian global optimization approaches to the full high-dimensional problem of interest. It should be noted that this framework can be generalized to any Bayesian optimization technique by only changing the acquisition function.

Estimating the active subspace requires access to gradients. If gradients are not available and finite differences are not feasible, then one may seek to estimate active subspaces without gradients. Some initial work has been done along these lines [129, 130]. How such gradient-free approaches can be used in the framework can be explored in the future research. Furthermore, some methods are of particular interest for using fewer evaluations of the gradient to compute the

directions defining the active subspace. Some strategies can also be pursued that make better use of the function evaluations acquired during the gradient evaluation, and compute the gradients of the objective function more accurately by proper selection of the kernel function. The future application of this framework can be in multi-disciplinary optimization problems, where surrogates can enable distributed frameworks to more efficiently couple different design disciplines.

REFERENCES

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

[2] N. Alexandrov, R. Lewis, C. Gumbert, L. Green, and P. Newman, "Optimization with variable-fidelity models applied to wing design," in *38th Aerospace Sciences Meeting and Exhibit*, p. 841, 2000.

[3] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[4] M. Imani and U. Braga-Neto, "Optimal finite-horizon sensor selection for Boolean Kalman filter," in *2017 51th Asilomar Conference on Signals, Systems and Computers*, IEEE, 2017.

[5] A. J. Booker, J. Dennis Jr, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset, "A rigorous framework for optimization of expensive functions by surrogates," *Structural optimization*, vol. 17, no. 1, pp. 1–13, 1999.

[6] D. R. Jones, "A taxonomy of global optimization methods based on response surfaces," *Journal of global optimization*, vol. 21, no. 4, pp. 345–383, 2001.

[7] M. Imani and U. Braga-Neto, "Optimal control of gene regulatory networks with unknown cost function," in *Proceedings of the 2018 American Control Conference (ACC 2018)*, IEEE, 2018.

[8] M. J. Sasena, P. Papalambros, and P. Goovaerts, "Exploration of metamodeling sampling criteria for constrained global optimization," *Engineering optimization*, vol. 34, no. 3, pp. 263–278, 2002.

[9] S. F. Ghoreishi and D. Allaire, "Gaussian process regression for Bayesian fusion of multi-fidelity information sources," in *19th AIAA/ISSMO Multidisciplinary Analysis and Opti-*

mization Conference, 2018.

[10] N. M. Alexandrov, J. E. Dennis Jr, R. M. Lewis, and V. Torczon, "A trust-region framework for managing the use of approximation models in optimization," *Structural optimization*, vol. 15, no. 1, pp. 16–23, 1998.

[11] N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman, "Approximation and model management in aerodynamic optimization with variable-fidelity models," *Journal of Aircraft*, vol. 38, no. 6, pp. 1093–1101, 2001.

[12] A. March and K. Willcox, "Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives," *AIAA journal*, vol. 50, no. 5, pp. 1079–1089, 2012.

[13] D. Huang, T. Allen, W. Notz, and R. Miller, "Sequential kriging optimization using multiple-fidelity evaluations," *Structural and Multidisciplinary Optimization*, vol. 32, no. 5, pp. 369–382, 2006.

[14] A. Mosleh and G. Apostolakis, "The assessment of probability distributions from expert opinions with an application to seismic fragility curves," *Risk Analysis*, vol. 6, no. 4, pp. 447–461, 1986.

[15] J. M. Reinert and G. E. Apostolakis, "Including model uncertainty in risk-informed decision making," *Annals of nuclear energy*, vol. 33, no. 4, pp. 354–369, 2006.

[16] M. E. Riley, R. V. Grandhi, and R. Kolonay, "Quantification of modeling uncertainty in aeroelastic analyses," *Journal of Aircraft*, vol. 48, no. 3, pp. 866–873, 2011.

[17] E. E. Leamer, *Specification searches: Ad hoc inference with nonexperimental data*, vol. 53. John Wiley & Sons Incorporated, 1978.

[18] D. Madigan and A. E. Raftery, "Model selection and accounting for model uncertainty in graphical models using occam's window," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1535–1546, 1994.

[19] S. Xie, M. Imani, E. R. Dougherty, and U. M. Braga-Neto, "Nonstationary linear discriminant analysis," in *Signals, Systems, and Computers, 2017 51st Asilomar Conference on*, pp. 161–165, IEEE, 2017.

[20] E. Hajiramezanali, M. Imani, U. Braga-Neto, X. Qian, and E. R. Dougherty, "Scalable optimal Bayesian classification of single-cell trajectories under regulatory model uncertainty," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 596–597, ACM, 2018.

[21] P. Honarmandi, T. C. Duong, S. F. Ghoreishi, D. Allaire, and R. Arroyave, "Bayesian uncertainty quantification and information fusion in calphad-based thermodynamic modeling," *Acta Materialia*, vol. 164, pp. 636–647, 2019.

[22] E. Hajiramezanali, M. Imani, U. Braga-Neto, X. Qian, and E. R. Dougherty, "Scalable optimal Bayesian classification of single-cell trajectories under regulatory model uncertainty," *BMC genomics*, 2019.

[23] D. Draper, "Assessment and propagation of model uncertainty," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 45–97, 1995.

[24] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: a tutorial," *Statistical science*, pp. 382–401, 1999.

[25] S. Geisser, "A bayes approach for combining correlated estimates," *Journal of the American Statistical Association*, vol. 60, pp. 602–607, 1965.

[26] M. Imani and U. Braga-Neto, "Optimal state estimation for Boolean dynamical systems using a Boolean Kalman smoother," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 972–976, IEEE, 2015.

[27] S. F. Ghoreishi and D. Allaire, "Multi-information source constrained bayesian optimization," *Structural and Multidisciplinary Optimization*, pp. 1–15.

[28] M. Imani and U. Braga-Neto, "Optimal gene regulatory network inference using the Boolean Kalman filter and multiple model adaptive estimation," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, pp. 423–427, IEEE, 2015.

[29] P. Morris, "Combining expert judgments: A bayesian approach," *Management Science*, vol. 23, pp. 679–693, 1977.

[30] R. Winkler, "Combining probability distributions from dependent information sources," *Management Science*, vol. 27, no. 4, pp. 479–488, 1981.

[31] S. F. Ghoreishi and D. L. Allaire, "A fusion-based multi-information source optimization approach using knowledge gradient policies," in *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, p. 1159, 2018.

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

[33] S. Julier and J. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations." In proceedings of the American Control Conference, pp. 2369-2373, 1997.

[34] S. Julier and J. Uhlmann, "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.

[35] R. L. Winkler, "Combining probability distributions from dependent information sources," *Management Science*, vol. 27, no. 4, pp. 479–488, 1981.

[36] W. D. Thomison and D. L. Allaire, "A model reification approach to fusing information from multifidelity information sources," in *19th AIAA Non-Deterministic Approaches Conference*, p. 1949, 2017.

[37] W. Scott, P. Frazier, and W. Powell, "The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression," *SIAM Journal on Optimization*, vol. 21, no. 3, pp. 996–1026, 2011.

[38] P. I. Frazier, W. B. Powell, and S. Dayanik, "A knowledge-gradient policy for sequential information collection," *SIAM Journal on Control and Optimization*, vol. 47, no. 5, pp. 2410–2439, 2008.

[39] W. B. Powell and I. O. Ryzhov, *Optimal learning*, vol. 841. John Wiley & Sons, 2012.

[40] P. Frazier, W. Powell, and S. Dayanik, "The knowledge-gradient policy for correlated normal beliefs," *INFORMS journal on Computing*, vol. 21, no. 4, pp. 599–613, 2009.

[41] C. E. Rasmussen and C. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

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

[43] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[44] S. Ghoreishi and D. Allaire, "Adaptive uncertainty propagation for coupled multidisciplinary systems," *AIAA Journal*, pp. 3940–3950, 2017.

[45] S. F. Ghoreishi and D. L. Allaire, "Compositional uncertainty analysis via importance weighted gibbs sampling for coupled multidisciplinary systems," in *18th AIAA Non-Deterministic Approaches Conference*, p. 1443, 2016.

[46] S. F. Ghoreishi, "Uncertainty analysis for coupled multidisciplinary systems using sequential importance resampling," Master's thesis, Texas A&M University, 2016.

[47] J. Wang, *Bayesian optimization with parallel function evaluations and multiple information sources: Methodology with applications in biochemistry, aerospace engineering, and machine learning*. Cornell University, 2017.

[48] M. Poloczek, J. Wang, and P. Frazier, "Multi-information source optimization," in *Advances in Neural Information Processing Systems*, pp. 4291–4301, 2017.

[49] B. Peherstorfer, K. Willcox, and M. Gunzburger, "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *Preprint*, pp. 1–57, 2016.

[50] A. Chaudhuri, J. Jasa, J. Martins, and K. E. Willcox, "Multifidelity optimization under uncertainty for a tailless aircraft," in *2018 AIAA Non-Deterministic Approaches Conference*, p. 1658, 2018.

[51] K. Kandasamy, G. Dasarathy, J. Schneider, and B. Poczos, "Multi-fidelity bayesian optimisation with continuous approximations," *arXiv preprint arXiv:1703.06240*, 2017.

[52] J. Thibert, M. Grandjacques, L. Ohman, *et al.*, "Naca 0012 airfoil," *AGARD Advisory Report*, vol. 138, 1979.

[53] C. Rumsey, "2d naca 0012 airfoil validation case," *Turbulence Modeling Resource, NASA Langley Research Center*, p. 33, 2014.

[54] M. Drela, "Xfoil: An analysis and design system for low reynolds number airfoils," in *Low Reynolds number aerodynamics*, pp. 1–12, Springer, 1989.

[55] F. Palacios, J. Alonso, K. Duraisamy, M. Colonno, J. Hicken, A. Aranake, A. Campos, S. Copeland, T. Economon, A. Lonkar, *et al.*, "Stanford university unstructured (su 2): an open-source integrated computational environment for multi-physics simulation and design," in *51st AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, p. 287, 2013.

[56] R. Barrett and A. Ning, "Comparison of airfoil precomputational analysis methods for optimization of wind turbine blades," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1081–1088, 2016.

[57] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[58] M. Imani and U. M. Braga-Neto, "Finite-horizon LQR controller for partially-observed Boolean dynamical systems," *Automatica*, vol. 95, pp. 172–179, 2018.

[59] A. Bahadorinejad, M. Imani, and U. Braga-Neto, "Adaptive particle filtering for fault detection in partially-observed Boolean dynamical systems," *IEEE/ACM transactions on computational biology and bioinformatics*, 2018.

[60] D. Huang, T. T. Allen, W. I. Notz, and R. A. Miller, "Sequential kriging optimization using multiple-fidelity evaluations," *Structural and Multidisciplinary Optimization*, vol. 32, no. 5, pp. 369–382, 2006.

[61] M. Imani, R. Dehghannasiri, U. M. Braga-Neto, and E. R. Dougherty, "Sequential experimental design for optimal structural intervention in gene regulatory networks based on the mean objective cost of uncertainty," *Cancer informatics*, vol. 17, p. 1176935118790247, 2018.

[62] M. Imani and U. M. Braga-Neto, "Point-based methodology to monitor and control gene regulatory networks via noisy measurements," *IEEE Transactions on Control Systems Technology*, 2018.

[63] R. A. Moore, D. A. Romero, and C. J. Paredis, "Value-based global optimization," *Journal of Mechanical Design*, vol. 136, no. 4, p. 041003, 2014.

[64] M. Imani and U. Braga-Neto, "Gene regulatory network state estimation from arbitrary correlated measurements," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 1, p. 22, 2018.

[65] M. Imani and U. Braga-Neto, "Control of gene regulatory networks using Bayesian inverse reinforcement learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018.

[66] M. Imani and U. Braga-Neto, "Particle filters for partially-observed Boolean dynamical systems," *Automatica*, vol. 87, pp. 238–250, 2018.

[67] M. Imani, S. F. Ghoreishi, D. Allaire, and U. Braga-Neto, "MFBO-SSM: Multi-fidelity Bayesian optimization for fast inference in state-space models," in *AAAI*, 2019.

[68] M. Imani, *Estimation, Inference and Learning in Nonlinear State-Space Models*. PhD thesis, Texas A&M University, 2019.

[69] T. M. Russi, *Uncertainty quantification with experimental data and complex system models*. University of California, Berkeley, 2010.

[70] V. Balabanov, R. Haftka, B. Grossman, W. Mason, and L. Watson, "Multifidelity response surface model for HSCT wing bending material weight," in *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, (St. Louis, MO), September 2-4 1998. AIAA 1998-4804.

[71] M. Imani and U. Braga-Neto, "State-feedback control of partially-observed Boolean dynamical systems using RNA- seq time series data," in *American Control Conference (ACC), 2016*, pp. 227–232, IEEE, 2016.

[72] T. W. Simpson, T. M. Mauery, J. J. Korte, and F. Mistree, "Kriging models for global approximation in simulation-based multidisciplinary design optimization," *AIAA journal*, vol. 39, no. 12, pp. 2233–2241, 2001.

[73] V. Balabanov and G. Venter, "Multi-fidelity optimization with high-fidelity analysis and low-fidelity gradients," in *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, (Albany, New York), August 30-September 1, 2004. AIAA 2004-4459.

[74] M. Imani and U. Braga-Neto, "Point-based value iteration for partially-observed Boolean dynamical systems with finite observation space," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pp. 4208–4213, IEEE, 2016.

[75] R. A. Moore and C. J. Paredis, "Variable fidelity modeling as applied to trajectory optimization for a hydraulic backhoe," in *ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 79–90, American Society of Mechanical Engineers, 2009.

[76] M. Imani and U. M. Braga-Neto, "Maximum-likelihood adaptive filter for partially observed Boolean dynamical systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 359–371, 2017.

[77] A. J. Keane, "Wing optimization using design of experiment, response surface, and data fusion methods," *Journal of Aircraft*, vol. 40, pp. 741–750, July-August 2003.

[78] S. Chen, Z. Jiang, S. Yang, and W. Chen, "Multimodel fusion based sequential optimization," *AIAA Journal*, vol. 55, no. 1, pp. 241–254, 2016.

[79] S. Choi, J. Alonso, and I. Kroo, "Multi-fidelity design optimization of low-boom supersonic business jets.," in *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, (Albany, New York), August 30–September 1 2004. AIAA 2004-4371.

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

[81] N. Alexandrov, J. Dennis, R. Lewis, and V. Torczon, "A trust region framework for managing the use of approximation models in optimization," Tech. Rep. CR-201745, NASA, October 1997.

[82] M. Imani and U. M. Braga-Neto, "Control of gene regulatory networks with noisy measurements and uncertain inputs," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 2, pp. 760–769, 2018.

[83] N. Alexandrov, R. Lewis, C. Gumbert, L. Green, and P. Newman, "Optimization with variable-fidelity models applied to wing design," Tech. Rep. CR-209826, NASA, December 1999.

[84] N. Alexandrov, R. Lewis, C. Gumbert, L. Green, and P. Newman, "Approximation and model management in aerodynamic optimization with variable-fidelity models," *AIAA Journal*, vol. 38, pp. 1093–1101, November-December 2001.

[85] M. Imani and U. Braga-Neto, "Multiple model adaptive controller for partially-observed Boolean dynamical systems," in *Proceedings of the 2017 American Control Conference (ACC 2017), Seattle, WA*, pp. 1103–1108, IEEE, 2017.

[86] B. Talgorn, S. Le Digabel, and M. Kokkolaras, "Statistical surrogate formulations for simulation-based design optimization," *Journal of Mechanical Design*, vol. 137, no. 2, p. 021405, 2015.

[87] C. Audet, M. Kokkolaras, S. Le Digabel, and B. Talgorn, "Order-based error for managing ensembles of surrogates in mesh adaptive direct search," *Journal of Global Optimization*, vol. 70, no. 3, pp. 645–675, 2018.

[88] B. Talgorn, C. Audet, S. Le Digabel, and M. Kokkolaras, "Locally weighted regression models for surrogate-assisted design optimization," *Optimization and Engineering*, vol. 19, no. 1, pp. 213–238, 2018.

[89] M. Imani, S. F. Ghoreishi, and U. M. Braga-Neto, "Bayesian control of large MDPs with unknown dynamics in data-poor environments," in *Advances in neural information processing systems*, pp. 8156–8166, 2018.

[90] L. D. McClenny, M. Imani, and U. M. Braga-Neto, "Boolean Kalman filter with correlated observation noise," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pp. 866–870, IEEE, 2017.

[91] L. D. McClenny, M. Imani, and U. Braga-Neto, "BoolFilter package vignette," *The Comprehensive R Archive Network (CRAN)*, 2017.

[92] L. D. McClenny, M. Imani, and U. M. Braga-Neto, "BoolFilter: an R package for estimation and identification of partially-observed Boolean dynamical systems," *BMC bioinformatics*, 2017.

[93] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[94] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and computers in simulation*, vol. 55, no. 1, pp. 271–280, 2001.

[95] S. Kucherenko, M. Rodriguez-Fernandez, C. Pantelides, and N. Shah, "Monte carlo evaluation of derivative-based global sensitivity measures," *Reliability Engineering & System Safety*, vol. 94, no. 7, pp. 1135–1148, 2009.

[96] P. E. Gill, W. Murray, and M. H. Wright, "Practical optimization," 1981.

[97] A. C. Antoulas, *Approximation of large-scale dynamical systems*. SIAM, 2005.

[98] K. Zhou, J. C. Doyle, K. Glover, *et al.*, *Robust and optimal control*, vol. 40. Prentice hall New Jersey, 1996.

[99] G. H. Dunteman, *Principal components analysis*. No. 69, Sage, 1989.

[100] C. Ding, X. He, H. Zha, and H. D. Simon, "Adaptive dimension reduction for clustering high dimensional data," in *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pp. 147–154, IEEE, 2002.

[101] S. Shan and G. G. Wang, "Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions," *Structural and Multidisciplinary Optimization*, vol. 41, no. 2, pp. 219–241, 2010.

[102] F. Chinesta, A. Huerta, G. Rozza, and K. Willcox, "Model order reduction: a survey," Wiley, 2016.

[103] P. Benner, S. Gugercin, and K. Willcox, "A survey of projection-based model reduction methods for parametric dynamical systems," *SIAM review*, vol. 57, no. 4, pp. 483–531, 2015.

[104] T. Lukaczyk, F. Palacios, J. J. Alonso, and P. Constantine, "Active subspaces for shape optimization," in *Proceedings of the 10th AIAA Multidisciplinary Design Optimization Conference*, pp. 1–18, 2014.

[105] P. G. Constantine, E. Dow, and Q. Wang, "Active subspace methods in theory and practice: applications to kriging surfaces," *SIAM Journal on Scientific Computing*, vol. 36, no. 4, pp. A1500–A1524, 2014.

[106] S. F. Ghoreishi, S. Friedman, and D. Allaire, "Adaptive dimensionality reduction for fast sequential optimization with gaussian processes," *Journal of Mechanical Design*, 2019.

[107] A. E. Bryson, *Applied optimal control: optimization, estimation and control*. CRC Press, 1975.

[108] A. Jameson, "Aerodynamic design via control theory," *Journal of scientific computing*, vol. 3, no. 3, pp. 233–260, 1988.

[109] A. Griewank and A. Walther, *Evaluating derivatives: principles and techniques of algorithmic differentiation*. SIAM, 2008.

[110] H. Chen, Q. Wang, R. Hu, and P. Constantine, "Conditional sampling and experiment design for quantifying manufacturing error of transonic airfoil," in *Proceedings of the 49th Aerospace Sciences Meeting*, 2011.

[111] E. Dow and Q. Wang, "Output based dimensionality reduction of geometric variability in compressor blades," in *Proceedings of the 51st Aerospace Sciences Meeting*, 2013.

[112] P. G. Constantine, Q. Wang, A. Doostan, and G. Iaccarino, "A surrogate accelerated bayesian inverse analysis of the hyshot ii flight data," in *Proceedings of the 49th Aerospace Sciences Meeting*, 2011.

[113] P. G. Constantine, Q. Wang, and G. Iaccarino, "A method for spatial sensitivity analysis," *Center for Turbulence Research, Stanford, CA*, 2012.

[114] P. G. Constantine, M. Emory, J. Larsson, and G. Iaccarino, "Exploiting active subspaces to quantify uncertainty in the numerical simulation of the hyshot ii scramjet," *Journal of Computational Physics*, vol. 302, pp. 1–20, 2015.

[115] D. Huang, T. T. Allen, W. I. Notz, and N. Zeng, "Global optimization of stochastic black-box systems via sequential kriging meta-models," *Journal of global optimization*, vol. 34, no. 3, pp. 441–466, 2006.

[116] S. S. Gupta and K. J. Miescke, "Bayesian look ahead one stage sampling allocations for selecting the largest normal mean," *Statistical Papers*, vol. 35, no. 1, pp. 169–177, 1994.

[117] S. S. Gupta and K. J. Miescke, "Bayesian look ahead one-stage sampling allocations for selection of the best population," *Journal of statistical planning and inference*, vol. 54, no. 2, pp. 229–244, 1996.

[118] M. Schonlau and D. Welch, William J and, "Global optimization with nonparametric function fitting," *Proceedings of the ASA, Section on Physical and Engineering Sciences*, pp. 183–186, 1996.

[119] M. Schonlau, W. J. Welch, and D. R. Jones, "Global versus local search in constrained optimization of computer models," *Lecture Notes-Monograph Series*, pp. 11–25, 1998.

[120] J. Wu and P. Frazier, "The parallel knowledge gradient method for batch bayesian optimization," in *Advances in Neural Information Processing Systems*, pp. 3126–3134, 2016.

[121] F. H. Branin, "Widely convergent method for finding multiple solutions of simultaneous nonlinear equations," *IBM Journal of Research and Development*, vol. 16, pp. 504–522, Sept 1972.

[122] M. Jamil and X.-S. Yang, "A literature survey of benchmark functions for global optimisation problems," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 4, no. 2, pp. 150–194, 2013.

[123] J. Vassberg, M. Dehaan, M. Rivers, and R. Wahls, "Development of a common research model for applied cfd validation studies," in *26th AIAA Applied Aerodynamics Conference*, p. 6919, 2008.

[124] J. P. Jasa, J. T. Hwang, and J. R. Martins, "Open-source coupled aerostructural optimization using python," *Structural and Multidisciplinary Optimization*, pp. 1–13, 2018.

[125] S. Friedman, S. F. Ghoreishi, and D. L. Allaire, "Quantifying the Impact of Different Model Discrepancy Formulations in Coupled Multidisciplinary Systems," in *19th AIAA Non-Deterministic Approaches Conference*, AIAA SciTech Forum, American Institute of Aeronautics and Astronautics, jan 2017.

[126] A. B. Lambe and J. R. Martins, "Extensions to the design structure matrix for the description of multidisciplinary design, analysis, and optimization processes," *Structural and Multidisciplinary Optimization*, vol. 46, no. 2, pp. 273–284, 2012.

[127] S. Yoon and A. Jameson, "Lower-upper symmetric-gauss-seidel method for the euler and navier-stokes equations," *AIAA journal*, vol. 26, no. 9, pp. 1025–1026, 1988.

[128] D. Kraft, "A software package for sequential quadratic programming," *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*, 1988.

[129] P. G. Constantine, A. Eftekhari, and M. B. Wakin, "Computing active subspaces efficiently with gradient sketching," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 353–356, IEEE, 2015.

[130] W. Li, G. Lin, and B. Li, "Inverse regression-based uncertainty quantification algorithms for high-dimensional models: Theory and practice," *Journal of Computational Physics*, vol. 321, pp. 259–278, 2016.