Surrogate Modeling Based on Statistical Techniques for Multi-fidelity Optimization

by

Rémi Lam

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 16 2014
LIBRARIES

© Massachusetts Institute of Technology 2014. All rights reserved.

Signature redacted

Author .....
Department of Aeronautics and Astronautics
May 22, 2014

Signature redacted

Certified by.....
Karen Willcox
Professor of Aeronautics and Astronautics
Thesis Supervisor

Signature redacted

Accepted by .....
Paulo C. Lozano
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

Surrogate Modeling Based on Statistical Techniques for Multi-fidelity Optimization

by

Rémi Lam

Submitted to the Department of Aeronautics and Astronautics
on May 22, 2014, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics and Astronautics

Abstract

Designing and optimizing complex systems generally requires the use of numerical models. However, it is often too expensive to evaluate these models at each step of an optimization problem. Instead surrogate models can be used to explore the design space, as they are much cheaper to evaluate.

Constructing a surrogate becomes challenging when different numerical models are used to compute the same quantity, but with different levels of fidelity (i.e., different levels of uncertainty in the models). In this work, we propose a method based on statistical techniques to build such a multi-fidelity surrogate. We introduce a new definition of fidelity in the form of a variance metric. This variance is characterized by expert opinion and can vary across the design space. Gaussian processes are used to create an intermediate surrogate for each model. The uncertainty of each intermediate surrogate is then characterized by a total variance, combining the posterior variance of the Gaussian process and the fidelity variance. Finally, a single multi-fidelity surrogate is constructed by fusing all the intermediate surrogates. One of the advantages of the approach is the multi-fidelity surrogate capability of integrating models whose fidelity changes over the design space, thus relaxing the common assumption of hierarchical relationships among models.

The proposed approach is applied to two aerodynamic examples: the computation of the lift coefficient of a NACA 0012 airfoil in the subsonic regime and of a biconvex airfoil in both the subsonic and the supersonic regimes. In these examples, the multi-fidelity surrogate mimics the behavior of the higher fidelity samples where available, and uses the lower fidelity points elsewhere. The proposed method is also able to quantify the uncertainty of the multi-fidelity surrogate and identify whether the fidelity or the sampling is the principal source of this uncertainty.

Thesis Supervisor: Karen Willcox
Title: Professor of Aeronautics and Astronautics

Acknowledgements

The research presented in this thesis would not have been possible without the help of many people. I would like to thank my advisor Professor Karen Willcox for her patience and constant guidance. She has always encouraged me to explore my ideas, provided me with valuable advices and made time available for our discussions. Her passion for outdoor activities plays an important role in making our group a friendly environment.

I would like to thank Professor Douglas Allaire for his help. He was always available to answer my questions and facilitated my transition when I joined the ACDL.

I would also like to thank all the members of the ACDL. Everyday they stimulate my curiosity, push me to deepen my understanding of sciences, and fill my life with laughter. I am especially thankful to Alessio Spantini for supporting me when I prepared for the quals and as I write this thesis. I would also like to thank Patrick Blonigan (for fostering enthusiasm), Eric Dow, Giulia Pantalone, Chi Feng and Chaitanya Talnikar for making the ACDL a place where I enjoy working.

I would like to thank my friends at MIT, especially Irene Dedoussi who was always present to help me drinking from the fire hose and my two roommates Louis Boulanger and Alexandre Constantin for the many memories in the "Maison du bonheur": you have become my overseas family.

Finally, I would like to thank my parents and my brother for always encouraging me to pursue my dreams, even if it means living thousands miles away from them.

This work was supported by the AFOSR MURI on Uncertainty Quantification, grant FA9550-09-0613, program manager F. Fahroo.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Designing complex systems is one of the challenges of engineering: how to exploit complicated physical phenomena and take advantages of the many interactions of those systems in order to improve their design? Advances in computational power are now allowing the engineer to partially answer this question by increasing the capabilities of numerical simulations. High-fidelity simulations are now used in analysis to evaluate performance, often replacing experiments.

Ideally, these high-fidelity simulations would be used every time one needs to evaluate a quantity of interest, but their cost can be prohibitive. For instance, to compute the drag of a wing, one would need to mesh the geometry of that wing and run a computational fluid dynamics (CFD) simulation: this is an expensive procedure both in terms of human resources (to create the mesh) and in terms of computational time. In an optimization setting, one would need to run this high-fidelity simulation for a large number of iterations to converge to an optimal solution, which can be expensive.

This leaves the engineer with the following question: while managing the uncertainty at acceptable levels, how to reduce the cost of optimization? Techniques such as multi-fidelity optimization and surrogate modeling aim at tackling this issue. In the traditional setting, multi-fidelity optimization develops techniques that

use models of different fidelities to drive the optimization process. The goal is to leverage the cheap, but less accurate, simulation to speed up the convergence and use the expensive, but accurate, simulation only when needed. Surrogate modeling is another promising technique to alleviate the cost of optimization. The objective is to construct an approximate model of the high-fidelity simulation based on previous calculations and use it in lieu of the expensive code. To be useful, the surrogate needs to be cheaper to evaluate than the high-fidelity simulation but still somewhat accurate. It can then replace the expensive code in the optimization.

By combining those two techniques, the cost of optimization can be reduced. However, there are some open questions.

- Constructing a single surrogate with data collected from different sources, and with different level of fidelities is challenging. The construction of the surrogate should use all the information available, but treat differently data of different levels of fidelity (relying more on the high-fidelity information than on the low-fidelity information). How do we incorporate several fidelities in a consistent and mathematically rigorous way?

- Surrogate-based optimization often exploits the surrogate itself as well as its uncertainty. Surrogates have inherent uncertainty stemming from the sparsity of the data they are built with. By combining information of various fidelities how can the uncertainty associated with the surrogate be quantified ?

1.2 Challenges in Multi-fidelity Modeling

1.2.1 Beyond the naive approach of Multi-fidelity

Model fidelity is often associated with model uncertainty; however there are multiple forms of uncertainty. The type of uncertainty considered in this work is model inadequacy, a notion defined by Kennedy and O'Hagan [24] as the "difference between the true mean value of the real world process and the code output at the true value of the

inputs". In particular, it represents the model inability to capture the full physical phenomena for a given input. A model is said to be high-fidelity when it has a low model inadequacy and low-fidelity when it has a high model inadequacy (this relation will be made quantitative later on).

A low-fidelity model can be the result of using simplifying assumptions to solve a set of equations. For example, in the fluid mechanics community, there is a clear hierarchy of models including (from high-fidelity to low-fidelity) [32]: Direct Numerical Simulation (DNS), Large Eddy Simulation (LES), Reynolds Averaged Navier-Stokes (RANS), Euler equations. Another example of fidelity can be found in Alexandrov et al.'s work [2]: solving the same equation on both a coarse and a fine mesh also results in a low-fidelity and high-fidelity model.

Even before the formalization of multi-fidelity frameworks, engineers have tried to take advantage of the several levels of fidelity available. Going from low-fidelity to high-fidelity simulations is a natural strategy [32]. First, use a low-fidelity model in the early design phase to evaluate many ambitious designs at a cheap cost. Then, as decisions are made about a configuration, optimize in a narrower design space with a medium-fidelity model. And finally do the fine adjustments using the highest fidelity model available. This strategy of exploring the design space before allocating expensive computations has limitations: if the low-fidelity model does not capture physical phenomena (for instance flow separation, turbulence, stall) that considerably deteriorate the performance of the design, it will only be noticed at a later phase of the design, when some decisions about the configuration have already been frozen [32]. In those cases, the corrections necessary to fix a poor design are often expensive or even impossible. Hence high-fidelity simulations need to be used earlier in the design process and in parallel with low-fidelity simulations. The approach proposed in this thesis intends to take advantage of such situations, when information of different fidelities is available, making it a useful tool for multi-fidelity optimization.

1.2.2 Gaussian Processes as Surrogate Models

Surrogate models can naturally be integrated in a multi-fidelity framework. A surrogate is an approximation model, cheap to evaluate, constructed on a finite set of evaluation points called data set. In this regard, surrogate models can incorporate data computed by simulations of different fidelities and take advantage of a multi-fidelity approach. A popular choice to construct surrogates is Gaussian Processes (GP) [35], also known as Kriging in the geo-sciences community [29]. It has been used extensively in optimization, including in multi-fidelity approaches (see [17] for a good review of multi-fidelity optimization based on surrogates, and Kriging in particular) and in multi-disciplinary optimization [37]. GP is a powerful nonlinear interpolation technique that can be interpreted in different contexts: from supervised machine learning to optimization of regularized networks.

1.2.3 Existing Multi-fidelity Strategies in Design Optimization

Strategies to perform multi-fidelity optimization have been studied extensively. These includes creating response surface surrogates using both low and high-fidelity model evaluations [22], or computing higher fidelity model samples when the difference between two lower fidelity evaluations is larger than a threshold [8]. Multi-fidelity optimization strategies can be divided in global and local approaches. Global approaches try to find the best design in the entire feasible domain: existing strategies include building an interpolation of the high-fidelity model as in Efficient Global Optimization (EGO) [21]. A multi-objective version of EGO developed in [33, 34] combines the use of a surrogate to both minimize the quantity of interest and explore unevaluated regions. In [23, 20], Gaussian processes are used to approximate the difference between a high and a lower fidelity model, this surrogate is then incorporated in the EGO framework . When gradients are available, Co-Kriging methods have been developed to build multi-fidelity surrogates [9, 10, 18].

Local approaches, on the other hand, only search for local optimum. Booker et al.

[7] developed a gradient-free pattern search algorithm for multi-fidelity optimization. When gradients are available, trust region methods use a local approximation of the high-fidelity function to perform the optimization. One significant extension of the trust region method to multi-fidelity optimization is the work by Alexandrov et al. [1, 2]. In particular, it proves that trust region methods converge to the optimum of the high-fidelity model when used with certain surrogate models. These surrogates are created based on the low-fidelity model in a way that satisfies the first-order consistency condition, i.e., the equality of the value and gradient of the surrogate and high-fidelity model at each trust region center. More recently, March and Willcox developed a multi-fidelity trust-region approach in the derivative-free setting [27].

1.2.4 Toward Multi-fidelity Surrogates Handling Non-hierarchical Fidelity Structures

Relaxing the Assumption of Hierarchy Among Models

A current limitation of surrogate modeling in the context of multi-fidelity, is the assumption that there exists a clear hierarchy of models. For instance, one would have a high-fidelity model f^h, and a low-fidelity model f^l, and assume that f^h has a lower model inadequacy than f^l everywhere on the design space. In other words, there is a hierarchical relationship among models. One way of constructing multi-fidelity surrogates is to approximate the difference between a low-fidelity model and a high-fidelity model. This correction, also known as calibration, clearly defines a hierarchy of models [26, 16, 20, 33]. Another way of constructing multi-fidelity surrogates is to use Co-Kriging [11, 18]. Co-Kriging is an extension of Gaussian Processes used to learn several correlated functions when several data sets are available. The evaluations of the low-fidelity model define a data set, and the evaluations of the high-fidelity model defines another one. A common assumption for Co-Kriging in multi-fidelity contexts is the auto-regressive assumption [23]. This imposes that, once a high-fidelity code has been evaluated at a point, nothing can be learned from a lower-fidelity code at this given point. This assumption is the only way fidelity is taken into account in Co-

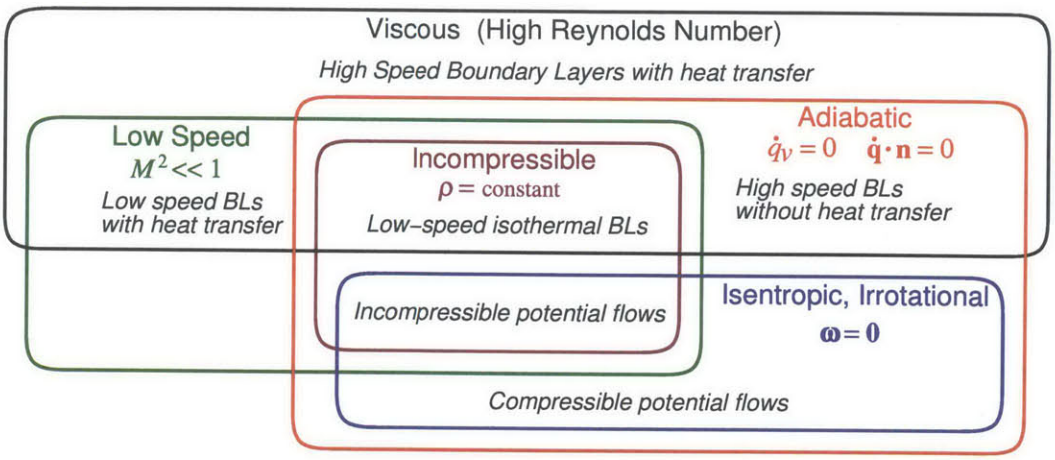


Figure 1-1: Aerodynamic flow categories. Drela [13]

Kriging and also imposes a hierarchy of models. However, in some situations there might not be a clear hierarchy of model to exploit. For example, consider solving an aerodynamic problem as illustrated in Fig.1-1. If one decides to use a solver for incompressible potential flows, one would expect the results to be high fidelity in a domain where viscous effects are negligible and at low Mach number. If one would use the same solver at higher Mach number, compressibility effects would not be taken into account, leading to a low-fidelity model.

Leveraging Information from every Fidelity

This motivates the new definition of the concept of fidelity proposed in our approach. Instead of associating a level of fidelity to a model (a low-fidelity model, or a high-fidelity model), we endow each model with a fidelity that varies across the design space. Thus, the proposed approach extends the notion of fidelity and allows us to handle non-hierarchical relationships among models. We also leverage our new definition of fidelity to reduce uncertainty: all the data available will be combined following Winkler [40]; the importance of each datum will depend on the function that quantifies its fidelity.

1.2.5 Assessing the Uncertainty of a Surrogate

Once quantified by a metric, the notion of fidelity can be integrated in a mathematically rigorous way to construct surrogates, allowing the uncertainty of the surrogate to be assessed.

Quantifying uncertainty is essential to guarantee both the optimality and the robustness of a design. Indeed, it is possible that the discrepancy between the performances predicted by a surrogate with high uncertainty and the performances of a real world process exceeds the confidence specifications given to the engineer. To avoid this situation, it is thus crucial to certify the results predicted by the surrogate by assessing its uncertainty.

Uncertainty quantification (UQ) is also a valuable tool for fidelity management. Indeed, when confidence requirements are not met, it is essential to identify the different sources of uncertainty. In the case of multi-fidelity surrogates, those sources of uncertainty include: sparse sampling, and model uncertainty. If a lack of samples (used to build a surrogate) is identified as the leading contributor to uncertainty, it is possible to determine which sample should be evaluated next in order to update the surrogate and reduce its variance [33]. This is the focus of the field of experimental design. On the other hand, in the case where several sub-models (computing different physical values) interact to evaluate a quantity of interest, if the uncertainty is driven by the fidelity of one sub-model, techniques such as global sensitivity analysis (GSA) [38, 36, 3] can determine which sub-model is the largest contributor to uncertainty. Once identified, that sub-model can be replaced by another sub-model with higher fidelity in the region of interest.

Finally, getting access to surrogate uncertainty quantification paves the way to devising optimization algorithms that rely on trust regions [28], or surrogate uncertainty (e.g., Efficient Global Optimization [21]) to explore the design space. Providing a practical tool for multi-fidelity optimization is one of the end goals of multi-fidelity surrogate modeling.

1.3 Objectives

The primary objective of this research is to provide techniques to construct surrogates that can be exploited in a multi-fidelity optimization. Specifically, we will propose an extension of the notion of fidelity that covers a broader range of situations than the traditional definition. The notion of fidelity will be made quantitative via a metric given by expert opinion and treated as an input in this work. A new approach to construct multi-fidelity surrogates will be proposed, combining several intermediate surrogates built with Gaussian processes and integrating several level of fidelity information. And finally we will quantify the uncertainty associated with this new surrogate.

To summarize, the research objectives are:

- to propose a richer definition of fidelity that admits non-hierarchical fidelity relationships among models.
- to construct multi-fidelity surrogates that leverage all the information available and can be used in multi-fidelity optimization.
- to quantify the uncertainty of that multi-fidelity surrogate with a metric that accounts for the fidelity and the sampling quality.

1.4 Thesis Outline

This thesis is organized as follows. In Chapter 2, we start with a review of the mathematical background on Gaussian process modeling. Then, we present the methodology and the algorithm developed to construct multi-fidelity surrogates: we define a new metric to quantify the uncertainty associated with fidelity and finally we explain how several GP models can be combined into a single multi-fidelity surrogate. In Chapter 3, the proposed method is applied to two examples. The first one is the characterization of a NACA 0012 airfoil lift coefficient as a function of Mach number and angle of attack using a panel code with boundary layer coupling (XFOIL) and an

Euler solver (SU2). The second example focuses on the lift coefficient of a biconvex airfoil with a supersonic panel code, a shock expansion theory model, and the inviscid solver of XFOIL. Chapter 4 summarizes the results of this research and offers leads for future work.

Chapter 2

Multi-fidelity Surrogate Modeling Methodology

This chapter presents the multi-fidelity surrogate modeling methodology proposed in this thesis. Section 2.1 presents an overview of the proposed method. Section 2.2 reviews the mathematical background of Gaussian Processes. Finally, Section 2.3 presents the new approach that we propose.

2.1 Overview of the Method

The proposed approach uses Gaussian Processes (GP), often used for regression, to construct multi-fidelity surrogates. The idea is to construct an intermediate surrogate for each model available using GP, but assign them uncertainty that takes into account the fidelity of the model. To be able to handle non-hierarchical relationships between models, the uncertainty of each model is quantified by a function that varies across the design space. Based on this metric for uncertainty, the intermediate surrogates are combined into a single multi-fidelity surrogate that gives more weight to intermediate surrogates with lower uncertainty.

The algorithm consists of three steps:

- construct an intermediate surrogate for each model f_m using GP. This allows us to compute a posterior mean μ_m, used as the intermediate surrogate, and a

posterior variance $\sigma^2_{GP,m}$ associated with the GP.

- use expert opinion to quantify the fidelity of each model across the design space in the form of variance function $\sigma^2_{f,m}$. A total variance $\sigma^2_{t,m}$ can then be defined, taking into account both the uncertainty associated with the GP and the fidelity uncertainty.

- combine the intermediate surrogates of each model into a single surrogate $\overline{\mu}$, with variance $\overline{\sigma}^2$, using fusion of information.

2.2 Gaussian Process - Background Review

This section gives a review of the mathematical background of GP. The first subsection is dedicated to the definition of GP and to fundamental results of inference with GP. The second subsection reviews results of regularized approaches and offers a different interpretation of the intermediate surrogate as the solution of a minimization problem. The third subsection gives a brief overview of hyper-parameter and kernel selection. Finally, a practical algorithm to compute the GP posterior statistics is reviewed.

2.2.1 Gaussian Process Regression

Problem statement

We illustrate the problem we intend to solve using GP. As described in the overview of the method, the first step of the algorithm is to construct a surrogate for each model available. First, consider only one of those models, f_m (the m^{th} model). This model f_m maps the design space to a real-value quantity of interest or performance metric. Formally:

$$\begin{array}{rccc} f_m: & \mathcal{X} \subset \mathbb{R}^d & \rightarrow & \mathbb{R} \\ & \mathbf{x} & \mapsto & y \end{array} \tag{2.1}$$

where d is the dimension of the design space $\mathcal{X}$.

The deterministic model f_m is endowed with a Gaussian Process that defines a random variable for every design of $\mathcal{X}$. Those random variables are used to build

a surrogate and quantify its uncertainty. The function f_m is only known at a finite number N of designs (also referred to as samples or data points): the i^{th} evaluated design $\mathbf{x}_i$ is associated with a performance y_i and the set $S_N = \{\mathbf{x}_i, y_i\}_{i=1}^N$ is called the data set or training data. The data set is then used to train the GP. The objective is to compute the posterior mean $\mu_m(\mathbf{x}^*)$ of the GP and use it as a surrogate for $f_m(\mathbf{x}^*)$, for any unevaluated design $\mathbf{x}^*$ of the design space. One of the features of Gaussian Processes is the quantification of their uncertainty with a posterior variance $\sigma^2_{GP,m}(\mathbf{x}^*)$. Formal definitions of Gaussian Processes, posterior mean and variance of a GP are reviewed in the following section. The problem to solve is illustrated in Fig. 2-1 in the case of a one dimensional design space.

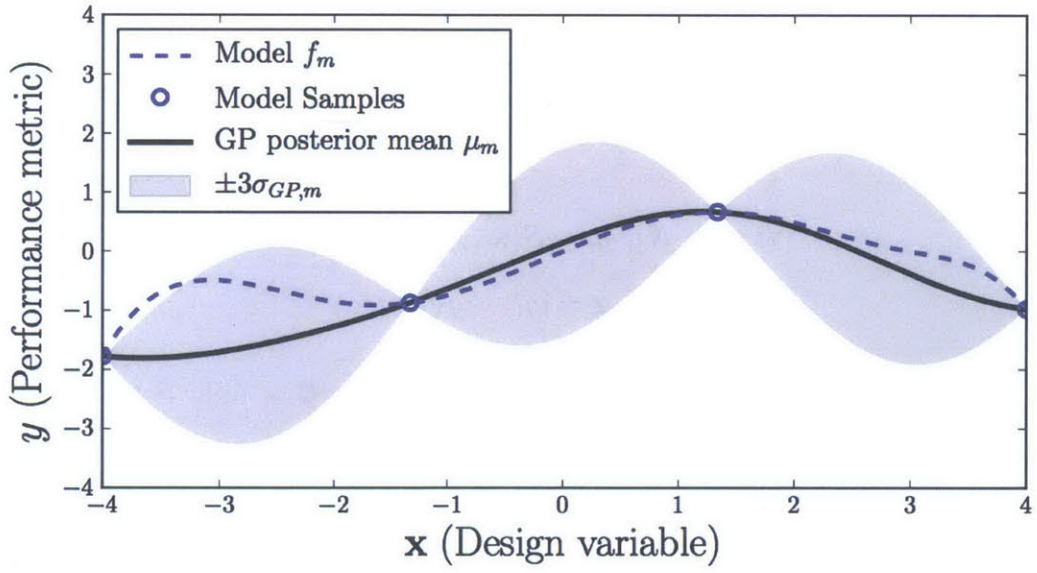


Figure 2-1: The blue dashed line is the model f_m to be recovered, the blue circles represent the data set, the solid black line is the posterior mean μ_m of the GP and the shading represents plus or minus three times the standard deviation associated with the GP.

Definition of Gaussian Processes - Function-space view

Several approaches can be used to describe and explain Gaussian Processes. Rasmussen and Williams [35] offer a complete review of GP in the context of machine learning. In this section, we present the key elements necessary to understand GP using the function-space view. This consists of defining a prior over a functional

space, and doing inference on that same functional space. An equivalent approach (not covered in this thesis), the weight-space view, can be found in [35].

A Gaussian Process is a special case of a stochastic process. A stochastic process can be defined as follows [25]. Let $\mathcal{X}$ be the design space and $(\Omega, \Sigma, \mathbb{P})$ a probability space (with Ω a sample space, Σ a σ-algebra over Ω, and $\mathbb{P}$ a probability measure). The function

$$\begin{aligned} \mathcal{S}: \quad \mathcal{X} \times \Omega &\rightarrow \quad \mathbb{R} \\ (\mathbf{x}, \omega) &\mapsto \quad \mathcal{S}(\mathbf{x}, \omega) \end{aligned} \tag{2.2}$$

is a stochastic process if for every fixed $\mathbf{x} \in \mathcal{X}$, the function $\mathcal{S}(\mathbf{x}, .) : \omega \in \Omega \rightarrow \mathbb{R}$ is a random vector on $(\Omega, \Sigma, \mathbb{P})$.

A stochastic process $\mathcal{S}$ is a Gaussian Process if, for any $N \in \mathbb{N}$, and any N designs $\mathbf{x}_1, \cdots \mathbf{x}_N$, the random vector $\mathbf{g}$ defined by

$$\forall i \in [1..N], \; g_i = \mathcal{S}(\mathbf{x}_i, .) : \omega \in \Omega \rightarrow \mathbb{R} \tag{2.3}$$

$$\mathbf{g} = [g_1 \cdots g_N]^\top \tag{2.4}$$

has a joint Gaussian distribution. A Gaussian Process $\mathcal{G}$ is fully determined by a mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ defined as:

$$m(\mathbf{x}) = \mathbb{E}_\omega[\mathcal{G}(\mathbf{x}, \omega)] \tag{2.5}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}_\omega[(\mathcal{G}(\mathbf{x}, \omega) - m(\mathbf{x}))(\mathcal{G}(\mathbf{x}', \omega) - m(\mathbf{x}'))]. \tag{2.6}$$

A realization of a GP is a function f_{GP} (hence the function-space view) and can be written:

$$f_{\text{GP}}(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{2.7}$$

Inference on Gaussian Processes

Gaussian Processes are powerful tools to construct surrogates: by updating the prior with information gained with the training points, the posterior mean of the GP can

be used as a surrogate of the original model. Three elements need to be specified before inference can occur:

- A prior mean for the GP: this represents the a priori mean of the function to be recovered. This can be any function, but in the rest of the section this will be set to zero for ease of notation, without loss of generality.

- A covariance function: this determines how strongly correlated are two values $f_{\text{GP}}(\mathbf{x})$ and $f_{\text{GP}}(\mathbf{x}')$ of a realization f_{GP} of the GP.

- A data set: to train the GP, or equivalently to condition the posterior of the GP on the data.

Consider a training set $S_N = \{\mathbf{x}_i, y_i\}_{i=1}^N$, composed of N evaluated designs. The performances $\{y_i\}_{i=1}^N$ can be written in a more compact way using a vector notation: $\mathbf{y}$ a $N \times 1$ vector. The designs $\{\mathbf{x}_i\}_{i=1}^N$ can also be written $X \in \mathbb{R}^{d \times N}$ where $X = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$. Define $\mathbf{g} = [g_1, \cdots, g_N]^\top$ where g_i is the random variable used to represent $f(\mathbf{x}_i)$ at the evaluated design $\mathbf{x}_i$. Now, define g_* as the random variable associated with the unevaluated design $\mathbf{x}^*$ (test point). Before seeing the performances of the evaluated designs, under the Gaussian Process assumption, these random variables $\mathbf{g}$ and g_* have a joint Gaussian distribution and can be written:

$$\begin{bmatrix} \mathbf{g} \\ g_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) & K(X,\mathbf{x}^*) \\ K(\mathbf{x}^*,X) & K(\mathbf{x}^*,\mathbf{x}^*) \end{bmatrix}\right) \tag{2.8}$$

where $K(X,X)$ is the $N \times N$ matrix of the covariance evaluated at all pairs of training points X. Similarly, $K(\mathbf{x}^*,X)$ is the $1 \times N$ matrix of covariance evaluated at all pairs of test point $(\mathbf{x}^*)$ and training points (X). This is the prior distribution of the GP that reflects the state of knowledge before seeing data.

Using the data set, the prior knowledge can be updated to take into account all the information available. Mathematically, this consists of conditioning the prior on the training points:

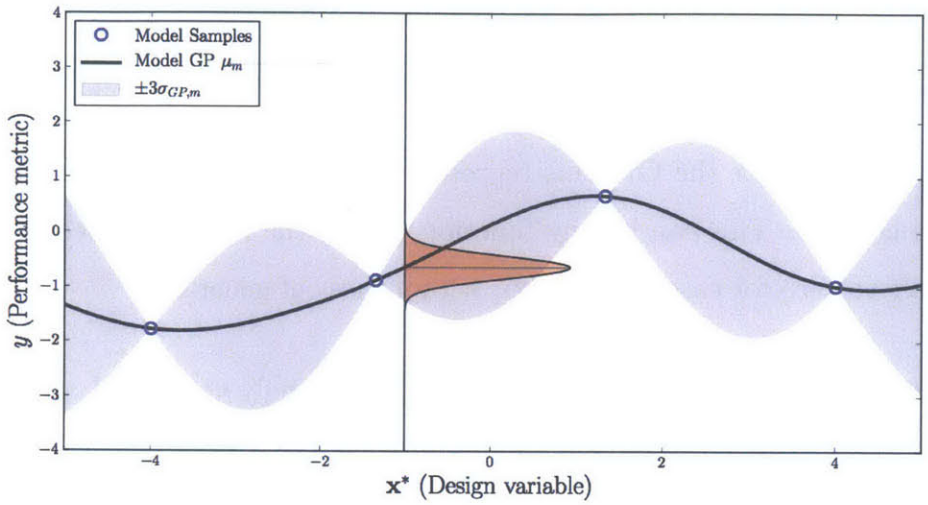


Figure 2-2: The blue dots are the training set, the solid black line is the posterior mean μ_m of the GP , and the grey shading represents plus or minus three times $\sigma_{GP,m}$. The red shading represents the distribution of the random variable at an unevaluated design $\mathbf{x}^*$ conditioned on the training set: it is a Gaussian distribution with mean $\mu_m(\mathbf{x}^*)$ and variance $\sigma^2_{GP,m}(\mathbf{x}^*)$.

$$g_*|\mathbf{y}, X \sim \mathcal{N}(\mu_m(\mathbf{x}^*), \sigma^2_{GP,m}(\mathbf{x}^*)) \tag{2.9}$$

The posterior is still a random variable with a Gaussian distribution. However, the posterior mean $\mu_m(\mathbf{x}^*)$ and variance $\sigma^2_{GP,m}(\mathbf{x}^*)$ are different from the prior. They have a closed form defined as:

$$\mu_m(\mathbf{x}^*) = K(X, \mathbf{x}^*)^\top K(X, X)^{-1}\mathbf{y} \tag{2.10}$$

$$\sigma^2_{GP,m}(\mathbf{x}^*) = K(\mathbf{x}^*, \mathbf{x}^*) - K(X, \mathbf{x}^*)^\top K(X, X)^{-1}K(X, \mathbf{x}^*). \tag{2.11}$$

There are a few interesting things to note about the posterior statistics. For instance, the mean $\mu_m(\mathbf{x}^*)$ is the scalar product of the vector $K(\mathbf{x}^*, X)$ with the vector $\alpha = K(X, X)^{-1}\mathbf{y}$. Once α has been computed, evaluating μ_m at a new design $\mathbf{x}^*$ only requires $\mathcal{O}(N)$ work. This is an important statement as μ_m will be used to construct a surrogate that needs to be cheap to evaluate. It can also be noted that

the posterior mean can be written as a linear combination of the covariance function:

$$\mu_m(\mathbf{x}^*) = \sum_{i=1}^{N} \alpha_i k(\mathbf{x}^*, \mathbf{x}_i). \tag{2.12}$$

where $\alpha_i \in \mathbb{R}$ is the i^{th} component of α. Hence, the choice of the covariance function (or kernel) decides which classes of function are admissible for the posterior mean. Moreover, the posterior variance can be decomposed into two terms: the prior variance minus a second term that encompasses the knowledge learned by conditioning on the training set and reduces the posterior variance.

Even if the model f_m used to evaluate training points is deterministic, it is common to consider that there exists a discrepancy between a model evaluation $f_m(\mathbf{x})$ and the true output $y(\mathbf{x})$ [17]. The discrepancy can be modeled by additive Gaussian noise:

$$y(\mathbf{x}) = f_m(\mathbf{x}) + \epsilon(\mathbf{x}) \tag{2.13}$$

$$\epsilon(\mathbf{x}) \sim \mathcal{N}(0, \lambda), \tag{2.14}$$

with λ the variance of the noise. This additive noise leads to a smoother posterior mean μ_m, with fewer oscillations. It also reduces the risk of over-fitting the data (the situation when the posterior mean μ_m agrees well with the model f_m on the training set but not on unevaluated points). This additive noise also ensures the positive definiteness of the covariance matrix which is necessary to compute its inverse (or its Cholesky decomposition).

Assuming that the noise at each training point is independent, the covariance function is now:

$$k(\mathbf{x}, \mathbf{x}') + \delta(\mathbf{x}, \mathbf{x}')\epsilon(\mathbf{x}), \tag{2.15}$$

leading to the prior

$$\begin{bmatrix} \mathbf{g} \\ g_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \lambda I & K(X, \mathbf{x}^*) \\ K(\mathbf{x}^*, X) & K(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix}\right). \tag{2.16}$$

The previous results for the posterior statistics still hold replacing the matrix $K(X,X)$ by $K(X,X)+\lambda I$:

$$\mu_m(\mathbf{x}^*) = K(X,\mathbf{x}^*)^\top [K(X,X)+\lambda I]^{-1}\mathbf{y} \tag{2.17}$$

$$\sigma^2_{GP,m}(\mathbf{x}^*) = K(\mathbf{x}^*,\mathbf{x}^*) - K(X,\mathbf{x}^*)^\top [K(X,X)+\lambda I]^{-1} K(X,\mathbf{x}^*). \tag{2.18}$$

The posterior mean μ_m of the GP can then be used as an intermediate surrogate for f_m. Even if it is not the case in general, μ_m often interpolates the training set S_N in situations when the data is sparse in the design space $\mathcal{X}$ (which typically occurs in an optimization setting).

2.2.2 Parallel with Regularized Approach

Interpreting μ_m as the solution of a minimization problem

The posterior mean μ_m computed using properties of Gaussian Processes can be derived from a different equation, leading to an additional interpretation of what the posterior mean represents. It can be shown [19] that μ_m is actually the solution of a minimization problem, in a particular space $\mathcal{H}$:

$$\mu_m = \underset{f\in\mathcal{H}}{\operatorname{argmin}}\, I_s[f] + \lambda \|f\|^2_{\mathcal{H}}, \tag{2.19}$$

where $\mathcal{H}$ is a Reproducing Kernel Hilbert Space (RKHS), I_s is the empirical risk, λ a regularizing term and $\|f\|_{\mathcal{H}}$ is the norm of f induced by the inner product defined on $\mathcal{H}$ (also called penalizing term).

The empirical risk is defined with respect to a training set $S_N = \{\mathbf{x}_i, y_i\}_{i=1}^N$ generated by f_m. It is the sum over every training point of a loss function. In this particular case, the loss function is the square error. Hence, the empirical risk can be written as :

$$I_s[f] = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(\mathbf{x}_i))^2. \tag{2.20}$$

The solution of the minimization problem can thus be interpreted as the element of

$\mathcal{H}$ that has a low error on the training set, but also has a low norm in $\mathcal{H}$. For specific spaces $\mathcal{H}$ (e.g., RKHS induced by Gaussian reproducing kernel), this norm can be related to the Fourier decomposition of the solution, and thus to its smoothness. Imposing a condition on the norm is necessary to ensure well-posedness of the problem. Without this regularizer term, multiple functions could have a zero empirical risk. The relative importance between the empirical risk and keeping a low norm is balanced by λ.

Reproducing Kernel Hilbert Space properties

Solving Eq. (2.19) is difficult, as $\mathcal{H}$ can be infinite dimensional (e.g., with space of real function on $\mathbb{R}$). However, if $\mathcal{H}$ is a RKHS, some properties allow simplification of the problem. A Hilbert space $\mathcal{H}$ of real-valued functions on $\mathcal{X}$ is said to be a Reproducing Kernel Hilbert Space if for all $\mathbf{x} \in \mathcal{X}$ the linear map $L_{\mathbf{x}} : f \mapsto f(\mathbf{x})$ from $\mathcal{H}$ to $\mathbb{R}$ is a bounded operator. Then, the Riesz representation theorem allows unique definition of an element $K_{\mathbf{x}} \in \mathcal{H}$ such that:

$$\forall \mathbf{x} \in \mathcal{X}, \exists! K_{\mathbf{x}} \in \mathcal{H}, \forall f \in \mathcal{H}, f(\mathbf{x}) = L_{\mathbf{x}}(f) = \langle K_{\mathbf{x}}, f \rangle_{\mathcal{H}}. \tag{2.21}$$

Since $K_{\mathbf{x}}$ is also an element of $\mathcal{H}$, it can also be written as:

$$\forall \mathbf{y} \in \mathcal{X}, \exists! K_{\mathbf{y}} \in \mathcal{H}, K_{\mathbf{x}}(\mathbf{y}) = \langle K_{\mathbf{x}}, K_{\mathbf{y}} \rangle_{\mathcal{H}}. \tag{2.22}$$

The function $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$,

$$\begin{array}{rccc} k: & \mathcal{X} \times \mathcal{X} & \rightarrow & \mathbb{R} \\ & (\mathbf{x}, \mathbf{y}) & \mapsto & \langle K_{\mathbf{x}}, K_{\mathbf{y}} \rangle_{\mathcal{H}}, \end{array} \tag{2.23}$$

is the reproducing kernel of $\mathcal{H}$. It can be shown that a RKHS uniquely defines a reproducing kernel, and any symmetric positive definite kernel uniquely defines a RKHS (Moore-Aronszajn Theorem) [4].

Simplification to a finite dimensional problem

The fundamental mathematical result that allows solving the minimization problem (2.19) is known as the Representer theorem [39]. It states that, for any RKHS $\mathcal{H}$, the minimizer of Eq.(2.19) satisfies:

$$\exists \alpha \in \mathbb{R}^N \text{ s.t. } \forall \mathbf{x}^* \in \mathcal{X} \ , \mu_m(\mathbf{x}^*) = \sum_{i=1}^{N} \alpha_i k(\mathbf{x}^*, \mathbf{x}_i), \tag{2.24}$$

where $\alpha_i \in \mathbb{R}$ the i^{th} component of α and $\{\mathbf{x}_i\}_{i=1}^N$ the training samples. This result is key to solving the minimization problem as it reduces the solution space from an infinite dimensional space to a space of dimension N. The original problem is now simplified to [19]:

$$\mu_m = \underset{\alpha \in \mathbb{R}^N}{\operatorname{argmin}} \|\mathbf{y} - K(X,X)\alpha\|_2^2 + \lambda \alpha^\top K(X,X)\alpha \tag{2.25}$$

where $\mathbf{y} = [y_1, \cdots, y_N]^\top$, and $K(X,X)$ is the $N \times N$ matrix whose ij^{th} entry is $k(\mathbf{x}_i, \mathbf{x}_j)$.

The solution of (2.25) is derived from the KKT conditions:

$$\mu_m(\mathbf{x}^*) = K(X,\mathbf{x}^*)^\top [K(X,X) + \lambda I]^{-1}\mathbf{y}, \tag{2.26}$$

with $K(X,\mathbf{x}^*)$ the $N \times 1$ vector whose i^{th} entry is $k(\mathbf{x}_i, \mathbf{x}^*)$. This result is similar to Eq.(2.17), replacing the covariance function by the reproducing kernel.

Insight gained from the regularized approch

As stated in the beginning of this section, the regularized approach allows us to interpret the posterior mean of a Gaussian Process μ_m as the solution of a minimization problem. One interesting feature is that it is not just the minimizer of the empirical risk: in particular, μ_m does not necessarily interpolate the training points. Indeed, μ_m minimizes a weighted sum of the empirical risk and its norm in $\mathcal{H}$. In the special

case where the kernel is a Gaussian with variance σ^2, the norm can be written [15]:

$$\|f\|_{\mathcal{H}} \propto \int |F(\omega)| \exp\left(\frac{\sigma^2\omega^2}{2}\right) d\omega, \tag{2.27}$$

which directly relates the norm $\|f\|_{\mathcal{H}}$ to the Fourier transform of f and penalizes high frequencies. This can be viewed as minimizing the error on the training set while keeping μ_m smooth. The relative importance of those two terms is controlled by the regularizer term λ, which plays the same role as the variance of the noise in the GP framework.

Again, the choice of the reproducing kernel, the equivalent of the covariance function in the GP framework, defines on which space $\mathcal{H}$ the minimizer should live, and thus imposes the form of the solution to satisfy Eq.(2.24). The important role that the kernel, or covariance function, plays in the quality of the solution μ_m motivates the following discussion about its selection.

2.2.3 Kernel and Selection of Hyper-parameters

The Square Exponential Kernel

In the previous section, the framework of Gaussian Processes has been presented, yielding a closed form of the posterior mean μ_m that can be computed based on a training set. The form of this posterior mean is strongly dependent on the covariance matrix $K(X, X)$ defined by the matrix of the covariance function (also refered to as the kernel) evaluated at the training points.

A popular choice for the covariance function is the square exponential kernel. It is defined as follows:

$$\begin{array}{rccc} k: & \mathcal{X} \times \mathcal{X} & \rightarrow & \mathbb{R} \\ & (\mathbf{x}, \mathbf{x}') & \mapsto & C \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2L^2}\right), \end{array} \tag{2.28}$$

where C is the maximum covariance, and L is a characteristic length scale. This choice of kernel implies an assumption of continuity for the solution μ_m and any realization

f_{GP} of the GP. Recalling that the covariance function is also used to evaluate

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f_{GP}(\mathbf{x}) - m(\mathbf{x}))(f_{GP}(\mathbf{x}') - m(\mathbf{x}'))] = \mathbb{C}ov[f_{GP}(\mathbf{x}), f_{GP}(\mathbf{x}')], \quad (2.29)$$

this means that the covariance of $f_{GP}(\mathbf{x})$ and $f_{GP}(\mathbf{x}')$ is only a function of the distance between $\mathbf{x}$ and $\mathbf{x}'$, with a maximum covariance reached when $\mathbf{x}$ and $\mathbf{x}'$ are equal. The square exponential kernel imposes more than continuity, it imposes smoothness. Indeed, the kernel is infinitely differentiable, leading to an infinitely mean-square differentiable process [35]. As previously mentioned, the square exponential kernel also leads to a norm in the $\mathcal{H}$ space that penalizes high frequencies.

In the case where a different characteristic length scale exists for each dimension of the design space $\mathcal{X} \subset \mathbb{R}^d$, it is possible to use a different kernel:

$$\begin{array}{rccc} k: & \mathcal{X} \times \mathcal{X} & \rightarrow & \mathbb{R} \\ & (\mathbf{x}, \mathbf{x}') & \mapsto & C \exp\left(-\sum_{i=1}^{d} \frac{\|x^{(i)} - x'^{(i)}\|^2}{2L_i^2}\right) \end{array} \quad (2.30)$$

where $x^{(i)}$ is the i^{th} component of the vector $\mathbf{x}$, and L_i is the characteristic length scale associated with dimension i of the design space.

Hyper-parameters

Choosing the Values for Hyper-parameters

Since the choice of the hyper-parameter values is important, one might want to have systematic way of selecting them. There are several standard methods for selecting the hyper-parameters [35] including cross validation and maximum marginal likelihood.

Cross validation consists of dividing the available data into two sets. One set is used to train the Gaussian Process while the other set is used as a reference to quantify the quality of the inference. The hyper-parameters that lead to the best average performance on the test set are kept. One popular technique of cross validation is the Leave One Out Cross Validation (LOOCV) which only uses one data point as the test set.

The maximum marginal likelihood method consists of finding the hyper-parameters that maximize the probability of observing the training set. The marginal likelihood of $\mathbf{y}$ conditioned on X (marginalizing over the functions f),

$$p(\mathbf{y}|X) = \int p(f|X)p(\mathbf{y}|f, X)df, \tag{2.31}$$

is the integral of the prior of f conditioned on X times the likelihood of $\mathbf{y}$ given f, conditioned on X. Under the Gaussian Process assumption, the prior and the likelihood are known:

$$f|X \sim \mathcal{N}(0, K(X, X)) \tag{2.32}$$

$$\mathbf{y}|f \sim \mathcal{N}(f, \lambda I). \tag{2.33}$$

This yields a closed form of the marginal likelihood, often expressed as the log marginal likelihood:

$$\log p(\mathbf{y}|X) = -\frac{1}{2}\mathbf{y}^\top \left(K(X, X) + \lambda I\right)^{-1} \mathbf{y} - \frac{1}{2}\log |K(X, X) + \lambda I| - \frac{N}{2}\log 2\pi. \tag{2.34}$$

The derivative of the log marginal likelihood with respect to the hyper-parameters is also known in closed form, allowing for an efficient optimization. In the rest of this work, the hyper-parameters are selected using maximum marginal likelihood.

2.2.4 Algorithm to Compute the Gaussian Process Posterior Mean and Variance

Inputs:

- Training set : $X = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$, $\mathbf{y} = [y_1, \cdots, y_N]^\top$
- Test point : $\mathbf{x}^* \in \mathcal{X}$
- Covariance function : $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

1. Compute the hyper-parameters $\theta = \{C, \lambda, L_i\}$ using the maximum marginal likelihood method

2. Compute

   - $K(X, X)$ the $N \times N$ matrix whose ij^{th} entry is $k(\mathbf{x}_i, \mathbf{x}_j)$
   - $K(X, \mathbf{x}^*)$ the $N \times 1$ vector whose i^{th} entry is $k(\mathbf{x}_i, \mathbf{x}^*)$
   - $K(\mathbf{x}^*, \mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*)$

3. Compute the Cholesky decomposition L of $K(X, X) + \lambda I$
   $L = \mathbf{Chol}[K(X, X) + \lambda I]$

4. Compute the vector $\alpha = \left(L^\top\right)^{-1} (L^{-1}\mathbf{y})$

5. Compute the posterior mean at $\mathbf{x}^*$:
   $\mu(\mathbf{x}^*) = K(X, \mathbf{x}^*)^\top \alpha$

6. Compute the posterior variance:
   $v = L^{-1} K(X, \mathbf{x}^*)$
   $\sigma^2_{GP} = K(\mathbf{x}^*, \mathbf{x}^*) - v^\top v$

2.3 Multi-fidelity Surrogate Modeling - Proposed Approach

We now introduce the proposed approach for constructing multi-fidelity surrogates. This consists of the three steps outlined in the overview of the method (Section 2.1). Those steps are explained in detail in the following sections.

2.3.1 Building an Intermediate Surrogate with Gaussian Processes

We consider the case where we have M models available $f_1, \cdots, f_M$, all mapping from $\mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}$. The m^{th} model has a training set $S_{N,m}$ with N_m training points written in matrix form as $X_m \in \mathbb{R}^{d \times N_m}$ and N_m performances written in vector form as $\mathbf{y}_m \in \mathbb{R}^{N_m}$. The first step in our approach is to construct an intermediate surrogate for each model f_m using Gaussian Processes (GP). As explained in Section 2.2, the posterior mean μ_m and variance $\sigma^2_{GP,m}$ of each GP can be computed in closed form and used as a surrogate of f_m. For each model, a covariance function k_m and its matrix form K_m are defined.

2.3.2 Sources of Uncertainty in Building a Multi-fidelity Surrogate

There are different sources of uncertainty that arise in the construction of the proposed multi-fidelity surrogate. In this section, we briefly describe those uncertainties and provide some insight into their meaning.

Variance Associated with Gaussian Process

For each model f_m, after conditioning the Gaussian Process with the training data $S_{N,m}$, the prior statistics are updated and a posterior variance $\sigma^2_{GP,m}$ can be computed everywhere on the design space. The prior variance is reduced where training data is available, forming uncertainty bubbles (Fig. 2-3). Note that the posterior variance is not necessarily 0 at a training point $\mathbf{x}_i \in S_{N,m}$. Recalling the formula to compute the posterior variance:

$$\sigma^2_{GP,m}(\mathbf{x}^*) = K_m(\mathbf{x}^*, \mathbf{x}^*) - K_m(X_m, \mathbf{x}^*)^\top [K_m(X, X) + \lambda I]^{-1} K_m(X_m, \mathbf{x}^*), \quad (2.35)$$

one should notice that, for a given covariance function k_m, $\sigma^2_{GP,m}$ does not depend on the model output $\mathbf{y}$, but only on the designs X. In some sense, when the training

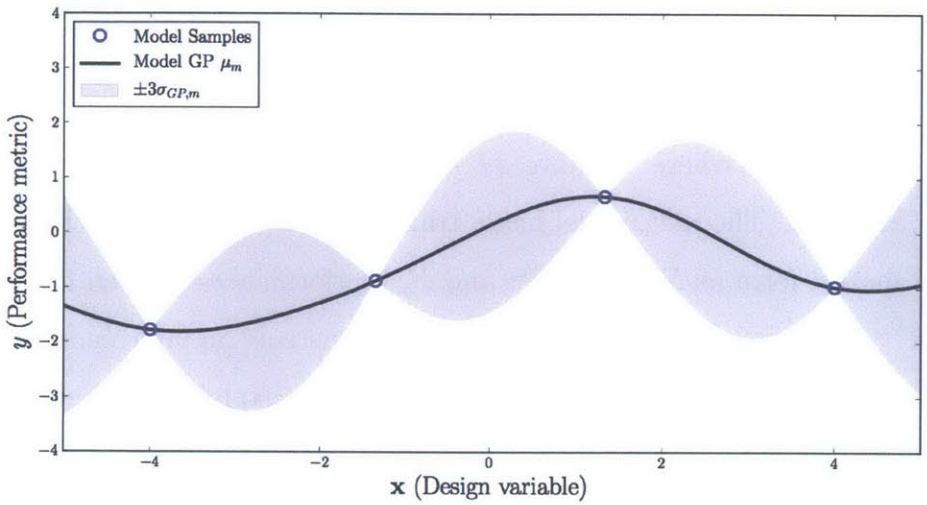


Figure 2-3: The blue dots are the training set, the solid black line is the posterior mean μ_m of the GP of f_m , and the grey shading represents plus or minus three times $\sigma_{GP,m}$.

points are sparse in the design space, it is an indicator of the quality of the sampling.

Variance Associated with Fidelity

Recalling the definition of fidelity (Section 1.2), we consider a second source of uncertainty to be that inherent to the model used to produce the data set. This uncertainty is the model inadequacy defined by Kennedy and O'Hagan [24] as the "difference between the true mean value of the real world process and the code output at the true value of the input". The model inadequacy cannot be computed as it requires to know the true mean value of the real process. Instead, we associate model inadequacy with fidelity and quantify it for model f_m with a variance $\sigma^2_{f,m}$. Unlike the traditional view of multi-fidelity, we consider the fidelity to be free to change across the design space, by letting $\sigma^2_{f,m}(\mathbf{x})$ be a function of the design variable $\mathbf{x} \in \mathcal{X}$. This function is specific to each model f_m and is assumed to be provided as an input by an expert.

Total Variance Associated with a Model

To quantify the uncertainty of each intermediate surrogate in a way that takes account of the uncertainty in the Gaussian Process and the uncertainty stemming from the fidelity of the model, we propose the definition of a total variance $\sigma^2_{t,m}$ for the model

f_m:

$$\forall \mathbf{x} \in \mathcal{X}, \sigma^2_{t,m}(\mathbf{x}) = \sigma^2_{GP,m}(\mathbf{x}) + \sigma^2_{f,m}(\mathbf{x}), \tag{2.36}$$

where $\sigma^2_{GP,m}(\mathbf{x})$ and $\sigma^2_{f,m}(\mathbf{x})$ are respectively the variance associated with the GP and the fidelity of model f_m.

The total variance $\sigma^2_{t,m}$ is bounded below by the fidelity variance $\sigma^2_{f,m}$ which means that, no matter how many samples are used to compute the intermediate surrogate, the uncertainty cannot be reduced below the uncertainty of the model itself. Fig. 2-4 illustrates how the intermediate surrogate is built for a given model f_m.

2.3.3 Building a Multi-fidelity Surrogate: Fusion of Information

Once the intermediate surrogates and the total variance have been computed for each of the M models, a single multi-fidelity surrogate can be built combining all the information available.

For design $\mathbf{x}^* \in \mathcal{X}$, and each model f_m, we define a random variable $h_{*,m}$, representing the intermediate surrogate $\mu_m(\mathbf{x}^*)$, with the following distribution

$$h_{*,m} \sim \mathcal{N}(\mu_m(\mathbf{x}^*), \sigma^2_{t,m}(\mathbf{x}^*)) \tag{2.37}$$

where $\mu_m(\mathbf{x}^*)$ and $\sigma^2_{t,m}(\mathbf{x}^*)$ are defined as in the previous section. For $\mathbf{x}^* \in \mathcal{X}$, it is possible to fuse the random variables $h_{*,1}$, ... ,$h_{*,M}$ into a single random variable H_*, assuming that they are independent (this hypothesis will be discussed in Chapter 4) [40] to obtain:

$$H_* \sim \mathcal{N}(\overline{\mu}(\mathbf{x}^*), \overline{\sigma}^2(\mathbf{x}^*)), \tag{2.38}$$

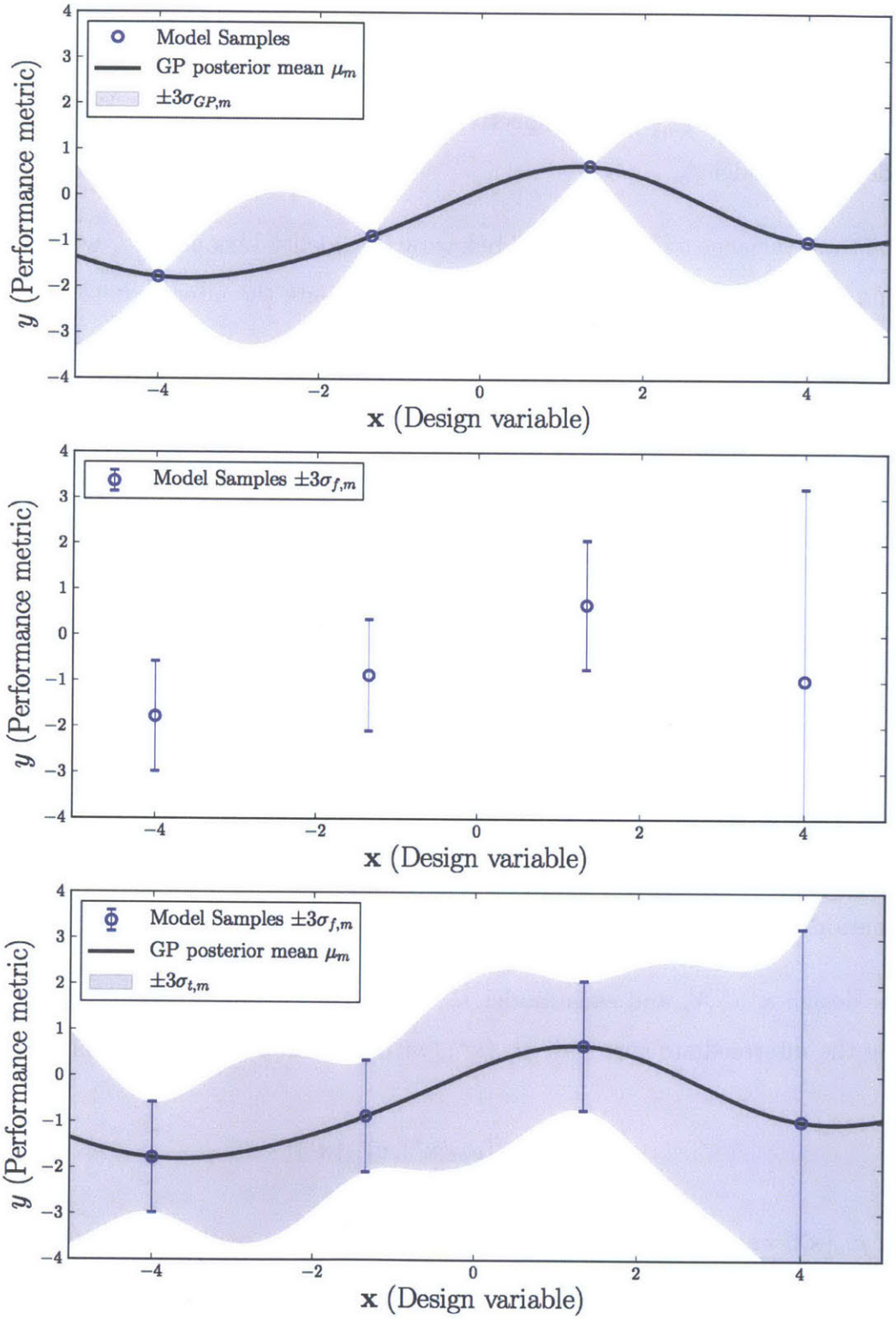


Figure 2-4: Top panel: Posterior mean μ_m and variance $\sigma^2_{GP,m}$ computed for a given model. Middle panel: Samples with associated fidelity variance $\sigma^2_{f,m}$. Bottom panel: Intermediate surrogate with mean μ_m and variance $\sigma^2_{t,m} = \sigma^2_{GP,m} + \sigma^2_{f,m}$.

with $\overline{\mu}(\mathbf{x}^*)$ and $\overline{\sigma}^2(\mathbf{x}^*)$ given by:

$$\overline{\sigma}^2(\mathbf{x}^*) = \left(\sum_{m=1}^{M} \frac{1}{\sigma_{t,m}^2(\mathbf{x}^*)} \right)^{-1} \tag{2.39}$$

$$\overline{\mu}(\mathbf{x}^*) = \overline{\sigma}^2(\mathbf{x}^*) \sum_{m=1}^{M} \frac{\mu_m(\mathbf{x}^*)}{\sigma_{t,m}^2(\mathbf{x}^*)}. \tag{2.40}$$

The fused model $\overline{\mu}$ can then be used as a multi-fidelity surrogate of the original quantity of interest. Fig. 2-5 illustrates this notion.

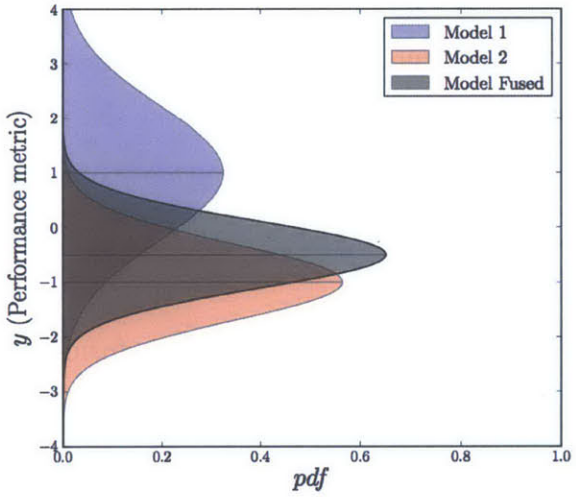


Figure 2-5: At a given design $\mathbf{x}^*$, the random variable $h_{*,1}$ and $h_{*,2}$ associated with the GP of model 1 and 2 have the distributions represented by the blue and red shadings. The black shading is the distribution of the random variable H_* after fusion: a normally distributed random variable of mean $\overline{\mu}(\mathbf{x}^*)$ and variance $\overline{\sigma}^2(\mathbf{x}^*)$.

Example 1

We demonstrate the approach through a simple illustrative example. In this example, we are interested in constructing a multi-fidelity surrogate for a performance metric y. The design space $\mathcal{X}$ is one-dimensional, and there are $M = 2$ models available to compute y: f_1 and f_2. Expert opinion characterizes the fidelity of each model through the variances $\sigma_{f,1}^2$ and $\sigma_{f,2}^2$.

For the first example, the model f_1 has been sampled evenly on $\mathcal{X} = [-4, 4]$ with 15 samples. The posterior statistics of the GP are computed, providing μ_1 and $\sigma_{GP,1}^2$.

Since the sampling covers the entire design space of interest, the variance of the GP is negligible compared to the fidelity variance $\sigma^2_{f,1}$ provided by the expert opinion. The total variance is then computed and $\sigma^2_{t,1} \approx \sigma^2_{f,1}$. Fig. 2-6 (top panel) shows the intermediate surrogate μ_1 for model f_1 and $\pm 3\sigma_{t,1}$. As mentioned, the total variance (and thus the standard deviation) is dominated by the fidelity variance: additional sampling of f_1 would not lead to uncertainty reduction for the intermediate surrogate of model 1.

Model 2 has only been evaluated twice. Thus the training data is sparser than for model 1; however, each of the two evaluations of f_2 has a higher fidelity. As a result, the variance of the GP collapses to zero at the training data (Fig. 2-7). The total variance for model 2 increases where there is no data point and reduces to the fidelity variance at the training points. The uncertainty of the intermediate surrogate for model 2 , illustrated by the grey shading on Fig. 2-6 (middle panel), could be reduced with more samples.

Fig. 2-6 (bottom panel) shows the surrogate after fusion of the two intermediate surrogates. The multi-fidelity surrogate $\overline{\mu}$ provides estimates that are close to the samples of model 2 (red dots), because the uncertainty of model 2 is low at those points. However, the multi-fidelity surrogate follows the trend of model 1 where the uncertainty of model 2 increases. The overall uncertainty $\overline{\sigma}^2$ of the multi-fidelity surrogate is reduced everywhere on the design space $\mathcal{X}$ with respect to the uncertainty ($\sigma^2_{t,1}$ and $\sigma^2_{t,2}$) of each intermediate surrogate, due to the information fusion step in our approach.

Example 2

In this second example, model 1 is identical to the first example (Fig. 2-8 top panel). Model 2, however, has been evaluated at different designs compared to the first example. The designs evaluated with f_2 are concentrated in $\mathcal{X} = [-3.3, -3]$. Therefore, the uncertainty of the intermediate surrogate for model 2 is low in $\mathcal{X} = [-4, -2]$ but increases steeply away from the samples (Fig. 2-8 middle panel). This leads to an intermediate surrogate with low uncertainty in $\mathcal{X} = [-4, -2]$, but very little

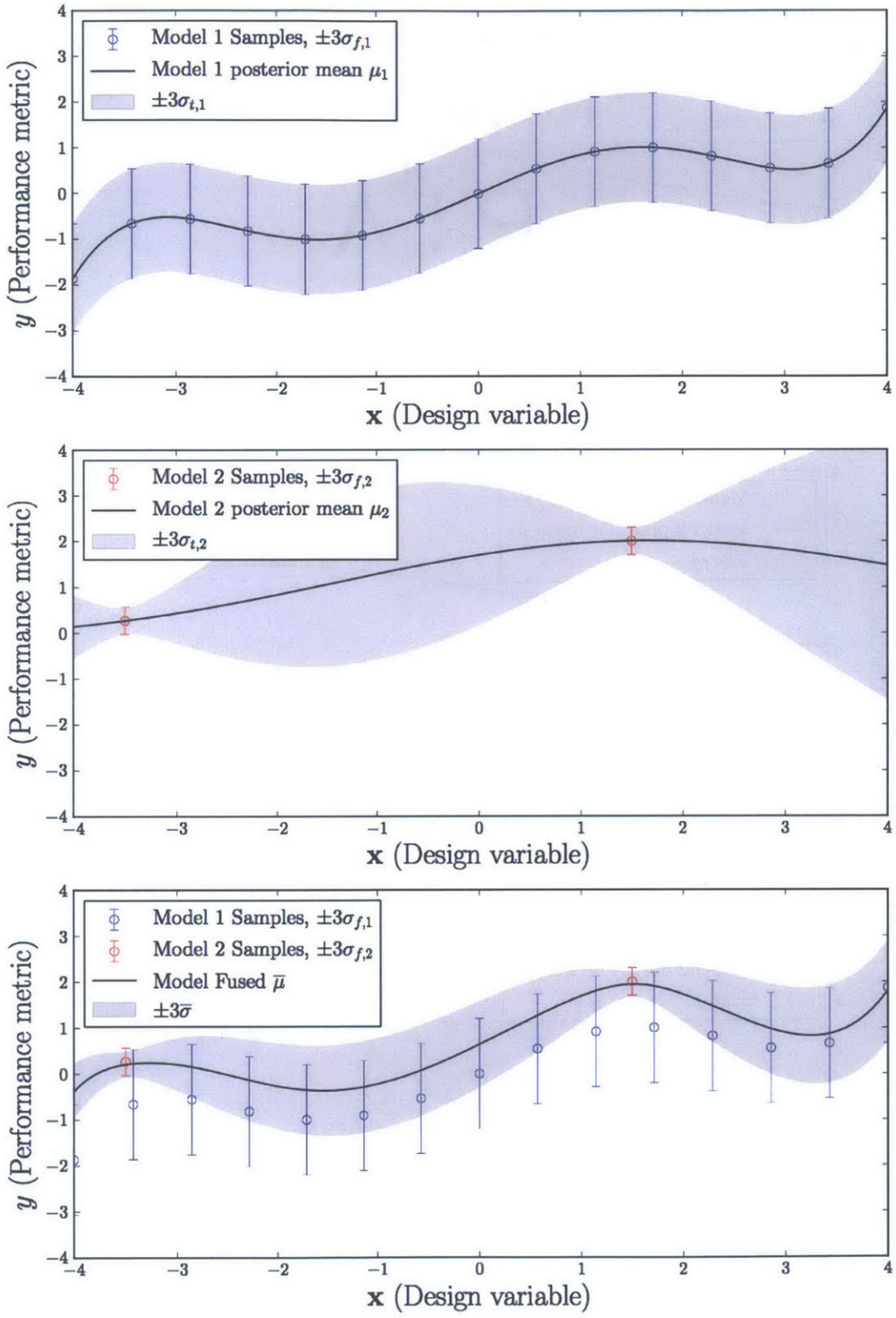


Figure 2-6: Example 1. Top panel: Intermediate surrogate for model 1, using the total variance $\sigma_{t,1}^2$. Middle panel: Intermediate surrogate for model 2, using the total variance $\sigma_{t,2}^2$. Bottom panel: Multi-fidelity surrogate, using the variance $\overline{\sigma}^2$.

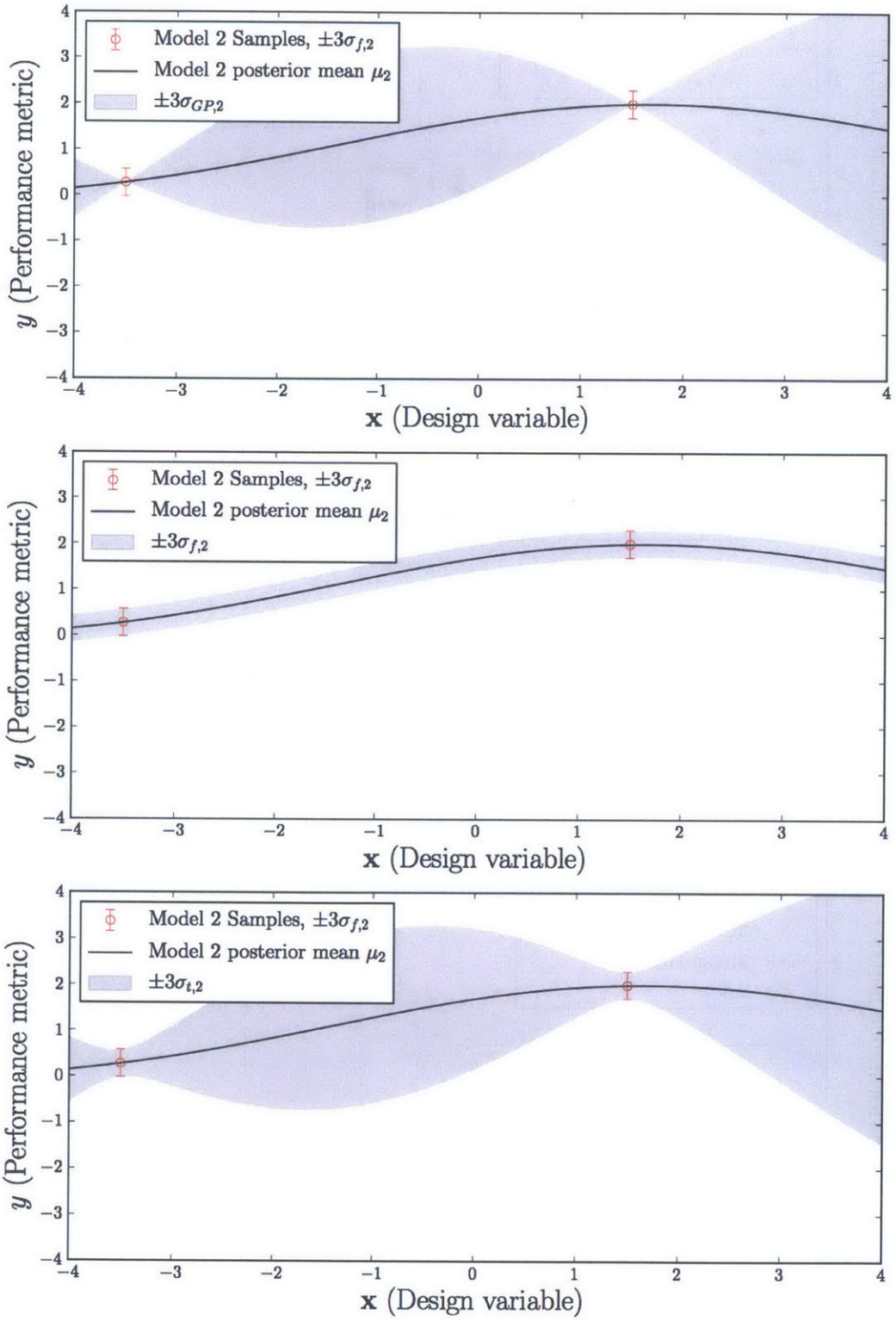


Figure 2-7: Example 1. Top panel: Posterior variance $\sigma^2_{GP,2}$ of the GP of model 2. Middle panel: Fidelity variance $\sigma^2_{f,2}$. Bottom panel: Total variance $\sigma^2_{t,2}$ of model 2.

information in $\mathcal{X} = [-2, 4]$.

Fig. 2-8 (bottom panel) shows the multi-fidelity surrogate after fusion of the two intermediate surrogates. The result is a smooth transition from the intermediate surrogate μ_2 of model 2 in $\mathcal{X} = [-4, -2]$ to the intermediate surrogate μ_1 of model 1 in $\mathcal{X} = [-2, 4]$. The uncertainty in the multi-fidelity surrogate takes advantage of the best part of each intermediate surrogate: the uncertainty of $\overline{\mu}$ collapses to the uncertainty of μ_2 in $\mathcal{X} = [-4, -2]$, and to the uncertainty of μ_1 in $\mathcal{X} = [-2, 4]$.

In the next chapter, we apply this approach to a more realistic aerodynamic analysis problem.

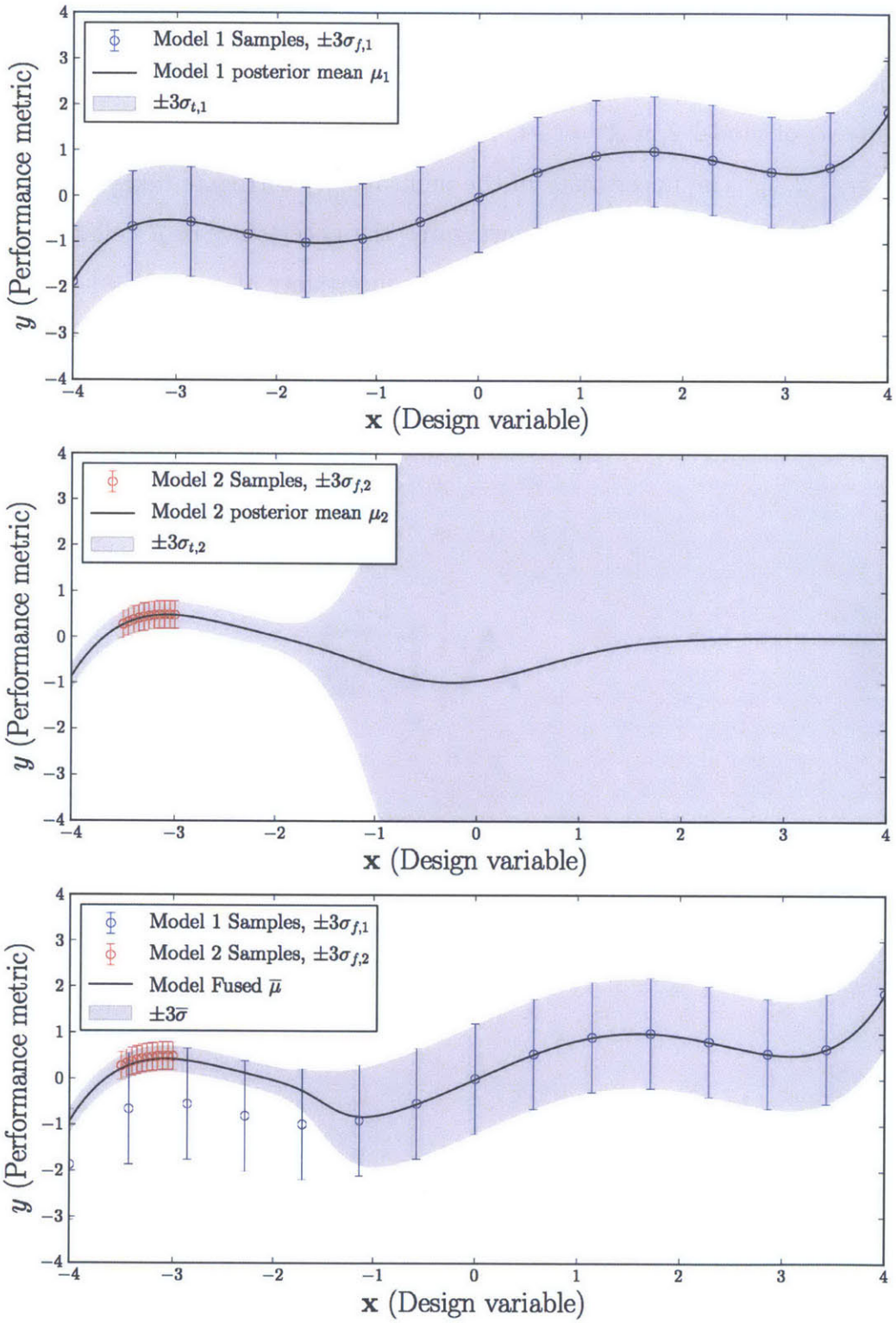


Figure 2-8: Example 2. Top panel: Intermediate surrogate for model 1, using the total variance $\sigma_{t,1}^2$. Middle panel: Intermediate surrogate for model 2, using the total variance $\sigma_{t,2}^2$. Bottom panel: Multi-fidelity surrogate, using the variance $\overline{\sigma}^2$.

Chapter 3

Application of Multi-fidelity Surrogate Modeling to Aerodynamic Examples

Estimating the lift and drag coefficient of an airfoil, a wing, or an entire aircraft is necessary to assess the quality of an airplane design and evaluate its performances. Since the dynamics of the fluid are governed by the Navier-Stokes equations, which are difficult to solve, simplified equations have been developed to solve this problem (e.g., Reynolds Averaged Navier-Stokes equation, Euler equations). These simplified equations are typically derived using additional assumptions and are valid in given regimes. This extended set of equations available to solve the same problem leads to information of multiple fidelity. The fidelity depends not only on the equations used, but also on the design tested, since the assumptions used to derive the equations could be valid in certain regimes but not in others. This motivates the new definition of fidelity proposed in Chapter 2, which no longer associates a unique fidelity to a model, as in the traditional setting, but extends the notion allowing the fidelity of a model to vary across the design space. In addition, the computational cost of solving the models can vary from a fraction of second to days, which motivates the use of surrogates, in particular multi-fidelity surrogates, especially in an optimization context.

This chapter demonstrates the multi-fidelity surrogate approach developed in Chapter 2 on three aerodynamic cases. Those three problems consist of characterizing the lift coefficient C_L of an airfoil. The first example is a NACA 0012 airfoil in the subsonic regime, the second example a is a biconvex airfoil in both subsonic and supersonic regimes, and the third example extends the second one by adding extra expert knowledge to improve the multi-fidelity surrogate.

3.1 Subsonic Airfoil - NACA 0012 - Case 1

The NACA 0012 airfoil (Fig. 3-1) is part of the four-digit airfoils series designed by the National Advisory Committee for Aeronautics (NACA), and is often used as a test case for Computational Fluid Dynamics (CFD) codes. The NACA 0012 is a symmetrical airfoil (no camber) and can be described by its half thickness, for which the equation is given by the 4-digit series:

$$\eta = \frac{t}{0.2}c\left[0.2969\sqrt{\frac{\xi}{c}} - 0.1260\frac{\xi}{c} - 0.3516\left(\frac{\xi}{c}\right)^2 + 0.2843\left(\frac{\xi}{c}\right)^3 - 0.1015\left(\frac{\xi}{c}\right)^4\right], \tag{3.1}$$

with c the chord, ξ the position along the chord, η the half thickness and t the maximum thickness to chord ratio (for the NACA 0012, $t = 0.12$).

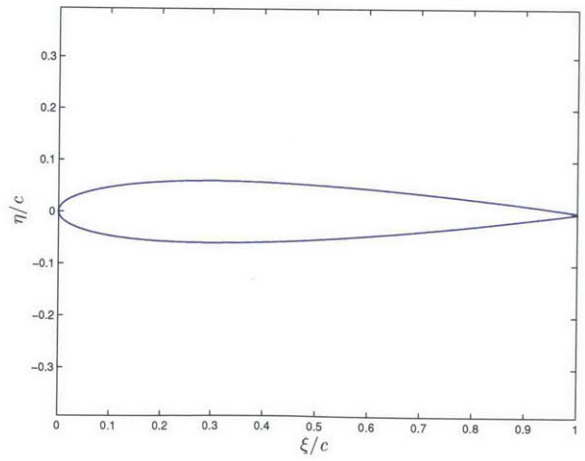


Figure 3-1: NACA 0012 airfoil profile.

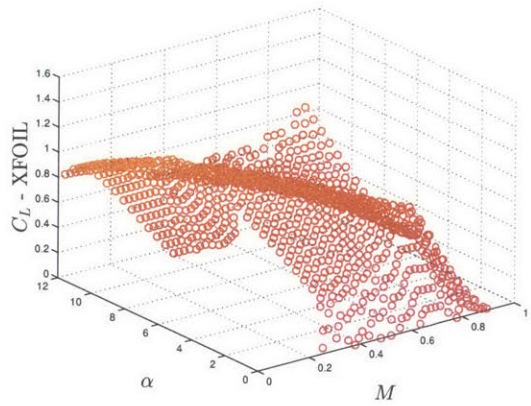
In this first example, the quantity of interest is the lift coefficient C_L of the airfoil, the design space $\mathcal{X}$ is two dimensional: $\mathcal{X} = I_M \times I_\alpha$ where $I_M = [0, 1]$ is the range of the Mach number M and $I_\alpha = [0, 12°]$ is the range of the angle of attack α. There are two models available: f_1 is the lift coefficient computed by XFOIL [12] with viscous terms, and f_2 is the lift coefficient computed by SU2 [30] using Euler equations. These models are described in the following subsections.

3.1.1 XFOIL

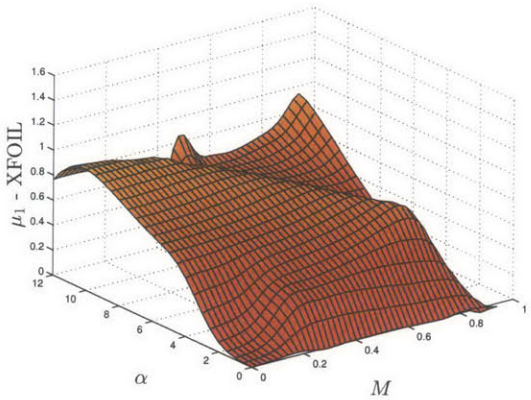
XFOIL is a solver for the design and analysis of airfoils in subsonic regime. It couples a panel method with the Karman-Tsien compressibility correction for the potential flow with a two equations boundary layer code. The laminar-turbulent transition is governed by an e^N envelope criterion. More about the XFOIL framework can be found in [12], the boundary layer treatment is described in [14], and a general overview of the panel method used can be found in [13].

The design space $\mathcal{X}$ is sampled and evaluated densely except in the region of Mach number lower than 0.3 and angle of attack lower than 5° (Fig. 3-2a) for a Reynolds number $R_e = 10^5$. Based on this data set $S_{N,1}$, the posterior mean μ_1 of the intermediate surrogate for model 1 is computed (Fig. 3-2b) as well as the posterior variance $\sigma^2_{GP,1}$. Fig. 3-2c shows the posterior standard deviation of the GP for XFOIL. One can notice that the uncertainty of the GP is low ($\sigma_{GP,1} \approx 0.001$) where samples are available, but increases elsewhere, i.e. at Mach number larger than 0.9 and especially in the low Mach, low angle of attack region. This means that the intermediate surrogate μ_1 has a low uncertainty except in the aforementioned regions.

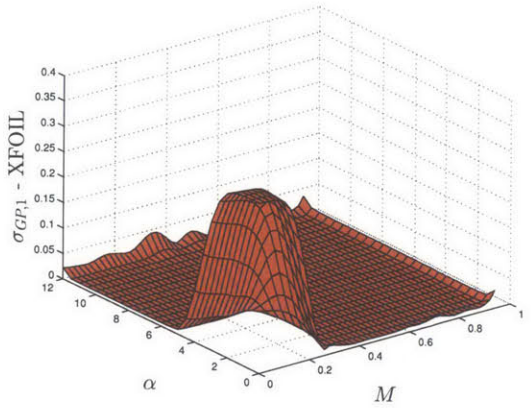
XFOIL is a solver for subsonic airfoils, and the compressibility correction used is only valid for relatively low Mach number, hence the lift coefficient computed at high Mach number ($M \geq 0.65$) has a low fidelity. Similarly, for high angle of attack, physical phenomena such as separation are expected, and cannot be handled by XFOIL, leading to a low fidelity calculation of C_L for $\alpha \geq 5$. This is illustrated in Fig. 3-2d, where the standard deviation of the fidelity $\sigma_{f,1}$ is low at low Mach



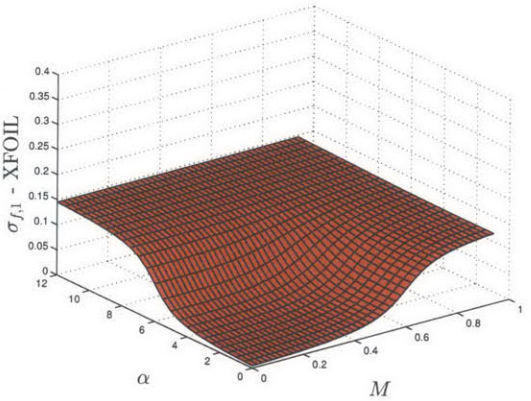
(a) C_L samples $S_{N,1}$ computed with XFOIL



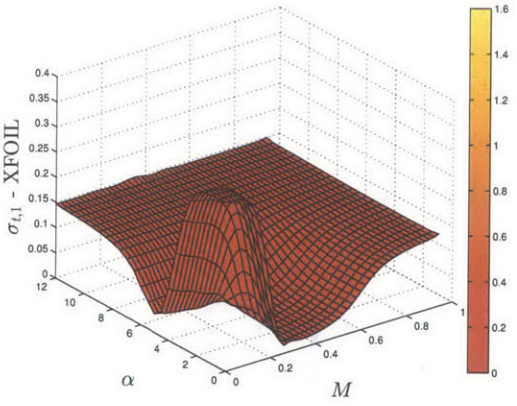
(b) Intermediate surrogate using XFOIL



(c) Posterior standard deviation $\sigma_{GP,1}$ of XFOIL



(d) Fidelity standard deviation $\sigma_{f,1}$ of XFOIL



(e) Total standard deviation $\sigma_{t,1}$ of XFOIL surrogate

Figure 3-2: Case 1. XFOIL samples $S_{N,1}$, posterior mean μ_1 and standard deviation $\sigma_{GP,1}$, fidelity standard deviation $\sigma_{f,1}$ and total standard deviation $\sigma_{t,1}$. In this and all the following figures, the color bar scale applies to all subplots.

number and low angle of attack, but increases in regions where XFOIL is not trusted to capture the physics.

The total variance $\sigma^2_{t,1}$ of the intermediate surrogate of XFOIL can be computed and is shown on Fig. 3-2e. The uncertainty is dominated by the fidelity uncertainty at high Mach number and high angle of attack, but dominated by the uncertainty of the GP in the low Mach number and low angle of attack area.

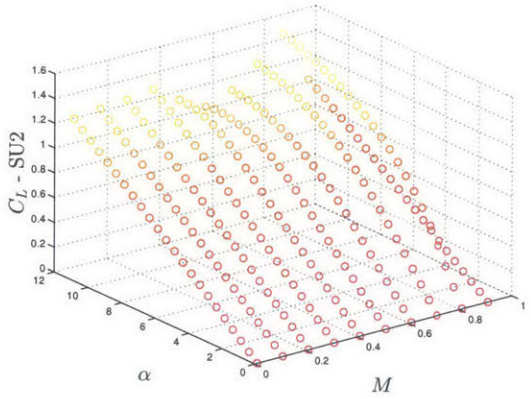
3.1.2 SU2 Euler

Stanford University Unstructured (SU2) suite is a collection of software for the analysis of partial differential equations (PDEs) and the optimization of PDE-constrained problems. We use SU2 to solve the Euler equations on a NACA 0012 airfoil using a finite volume scheme. Details about the equations solved and their implementation can be found in [30]. Because, SU2 is more expensive to evaluate than XFOIL, the design space is sampled uniformly on $\mathcal{X}$ but more sparsely (Fig. 3-3a). Those evaluations define a second data set $S_{N,2}$ used to construct the intermediate surrogate for f_2 (Fig. 3-3b).

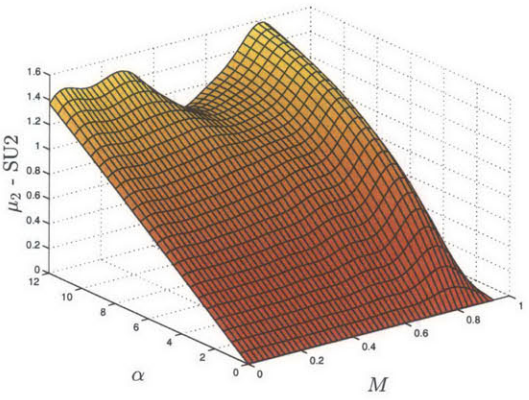
The posterior variance $\sigma^2_{GP,2}$ exhibits the bubble behavior shown in Section 2.3.2, but only in one direction of the design space 3-3c. This can be explained by the different characteristic length-scales of this problem: the variations of the lift coefficient with respect to the angle of attack have a longer length scale than the variations of C_L with respect to the Mach number. Hence, the sampling in the axis of the Mach number appears sparser than the sampling in the axis of the angle of attack α, leading to uncertainty bubbles in the Mach number axis.

Because the Euler equations do not account for viscous terms, the fidelity variance of SU2 is low only at low Mach number and low angle of attack (Fig. 3-3d). In this region, $\sigma_{f,2} \geq \sigma_{f,1}$ because XFOIL has a better representation of the physics than the Euler equation solved by SU2.

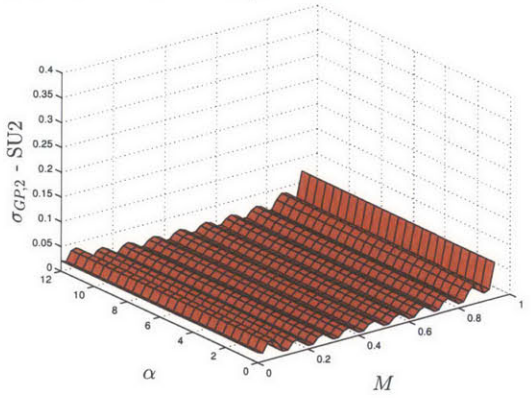
The total variance $\sigma^2_{t,2}$ of the intermediate surrogate for f_2 is dominated by the fidelity variance at high Mach number and high angle of attack, and is dominated by the variance of the GP in the low Mach number, low angle of attack region (Fig.



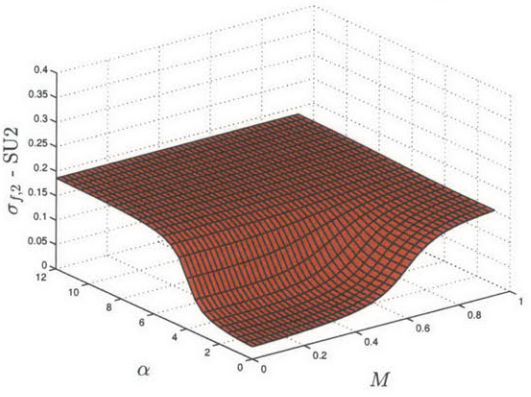
(a) C_L samples $S_{N,2}$ computed with SU2



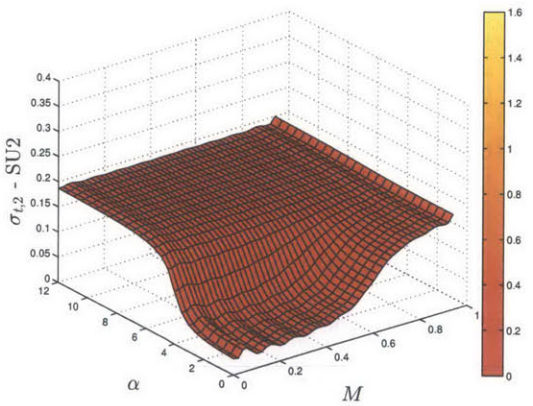
(b) Intermediate surrogate using SU2



(c) Posterior standard deviation $\sigma_{GP,2}$ of SU2

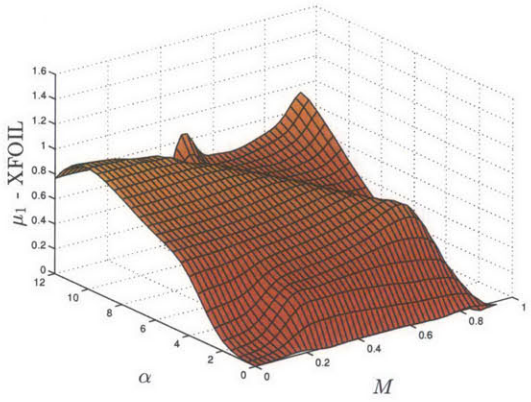


(d) Fidelity standard deviation $\sigma_{f,2}$ of SU2

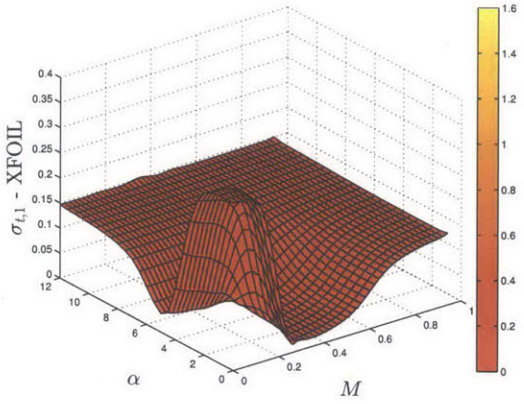


(e) Total standard deviation $\sigma_{t,2}$ of SU2 surrogate

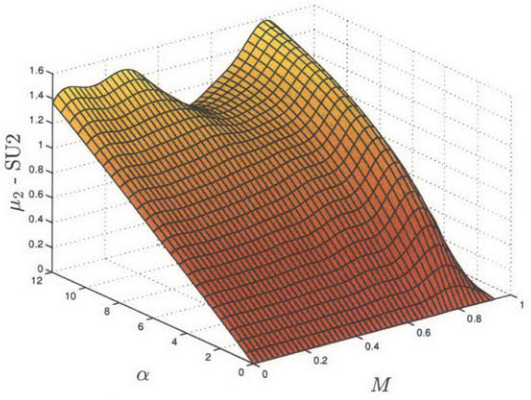
Figure 3-3: Case 1. SU2 samples $S_{N,2}$, posterior mean μ_2 and standard deviation $\sigma_{GP,2}$, fidelity standard deviation $\sigma_{f,2}$ and total standard deviation $\sigma_{t,2}$.



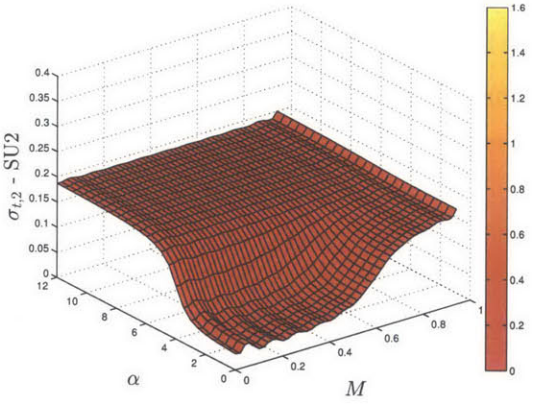
(a) Intermediate surrogate μ_1 (XFOIL)



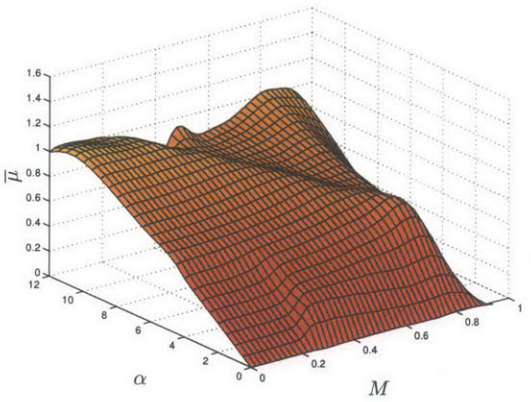
(b) Total standard deviation of XFOIL surrogate $\sigma_{t,1}$



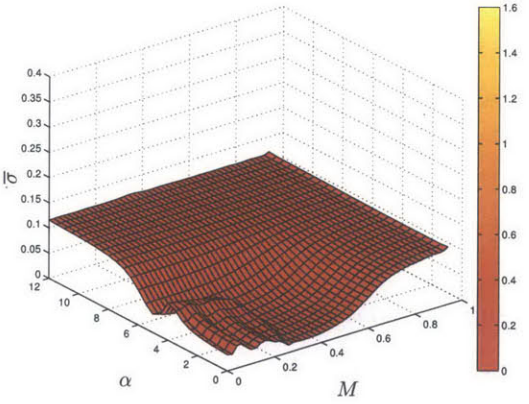
(c) Intermediate surrogate μ_2 (SU2)



(d) Total standard deviation of SU2 surrogate $\sigma_{t,2}$



(e) Multi-fidelity surrogate $\overline{\mu}$



(f) $\overline{\sigma}$ for the multi-fidelity surrogate

Figure 3-4: Case 1. Fusing intermediate surrogates for XFOIL and SU2 into a single multi-fidelity surrogate.

3-3e).

3.1.3 Combining Intermediate Surrogates

Fig. 3-4 shows the results after combining the two intermediate surrogates μ_1 and μ_2: the multi-fidelity surrogate $\overline{\mu}$ exhibits the behavior of μ_1 (respectively μ_2) in region where the uncertainty of μ_1 (respectively μ_2) is low (Fig. 3-4e). In some sense, the multi-fidelity surrogate uses XFOIL where data is available, and pads with data from SU2 elsewhere. The uncertainty in the multi-fidelity surrogate is high in the region where neither of the model has a high fidelity (high Mach number and high angle of attack), but has been reduced in the low Mach number, low angle of attack region where model 1 (XFOIL) is lacking samples (Fig. 3-4f) but model 2 has been sampled uniformly.

3.2 Supersonic Biconvex Airfoil - Case 2

In this second example, we consider a supersonic airfoil: a biconvex airfoil with a 5% thickness Fig. 3-5. The quantity of interest is the lift coefficient C_L and the design variables are the Mach number $M \in I_M = [0, 2]$ and the angle of attack $\alpha \in I_\alpha = [0, 6°]$. There are three models available: the two first models are valid in the supersonic regime [28], f_1 is a shock expansion code, f_2 is a supersonic linear panel code. The third model f_3 is XFOIL in inviscid mode for the subsonic regime.

3.2.1 Shock Expansion Theory

The shock expansion (SE) method solves the pressure at the surface of the airfoil based on geometry changes using non-linear equations. The lift coefficient C_L is then computed by integrating the local pressure coefficient on the surface of the airfoil.

The design space is sampled densely in the supersonic regime Fig. 3-6a. However, the shock expansion code does not return any result for some designs, which are removed from the data set $S_{N,1}$ shown on Fig. 3-6b. The posterior mean μ_1 and

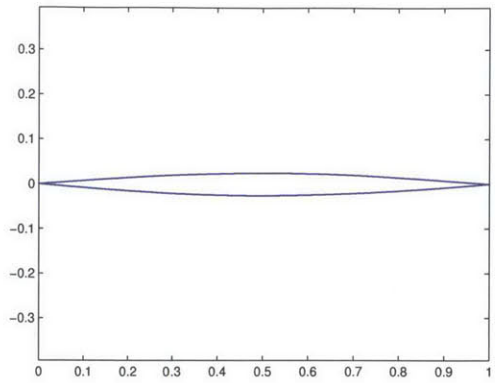


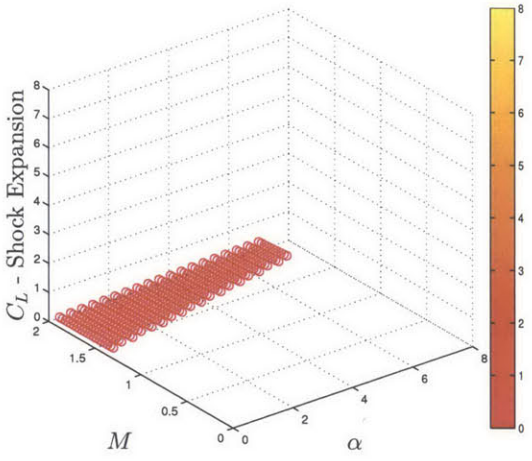
Figure 3-5: Biconvex airfoil profile.

posterior variance $\sigma^2_{GP,1}$ are computed using the proposed approach (Fig. 3-6c and Fig. 3-6d). Wherever the shock expansion method cannot provide results, the fidelity variance $\sigma^2_{f,1}$ is set to a high value (Fig. 3-6e). The total variance $\sigma^2_{t,1}$ can be computed and leads to a high uncertainty in the surrogate model where there is no data available. The uncertainty where samples are available is approximately the uncertainty due to fidelity (Fig. 3-6f).

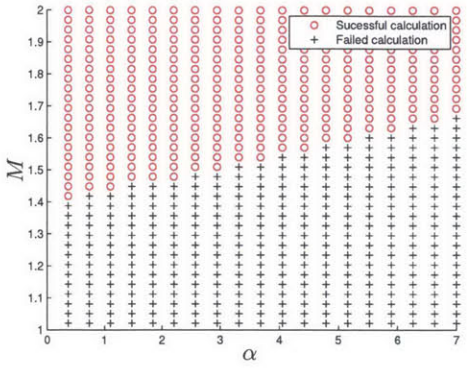
3.2.2 Panel code

The panel code (PC) also computes the pressure on the surface of the airfoil but assumes a linear relationship between the changes in the geometry and the changes in the local pressure. The lift coefficient is then computed by integrating the pressure coefficient.

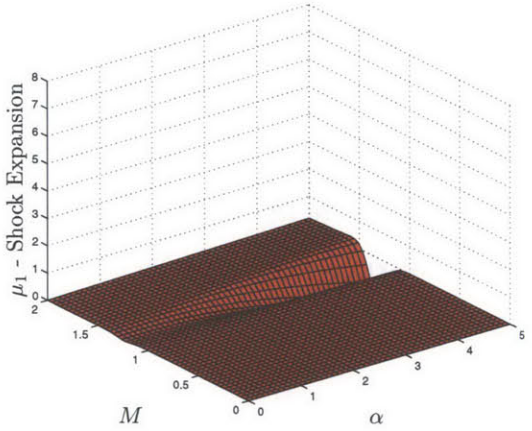
Since the panel code is only valid in supersonic regime, the design space is only sampled for Mach numbers $M \geq 1$. The data set $S_{N,2}$ is once again used to compute the intermediate surrogate μ_2 and the posterior variance $\sigma^2_{GP,2}$. It can be noticed that $\sigma_{GP,2}$ steeply increases in the subsonic region, because no sample is available (Fig. 3-7c). The fidelity variance $\sigma^2_{f,2}$ is set to be low in the supersonic region and high in the subsonic region (Fig. 3-7d). The results are shown in Fig. 3-7.



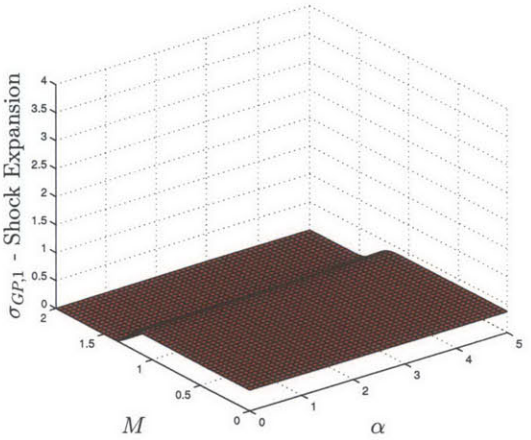
(a) Data set $S_{N,1}$ of SE



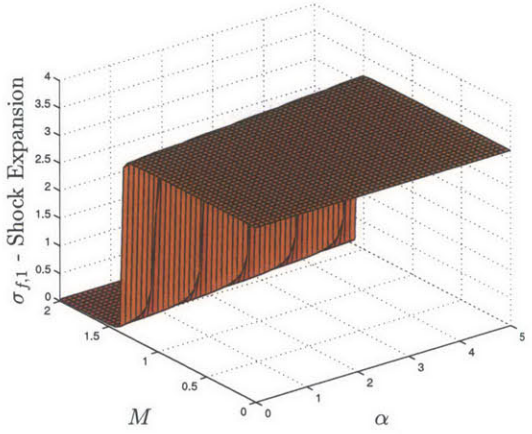
(b) Successful and failed calculations of SE



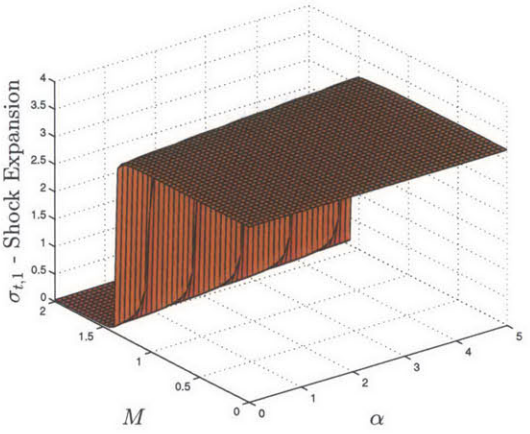
(c) Intermediate surrogate of SE



(d) Posterior standard deviation $\sigma_{GP,1}$ of SE

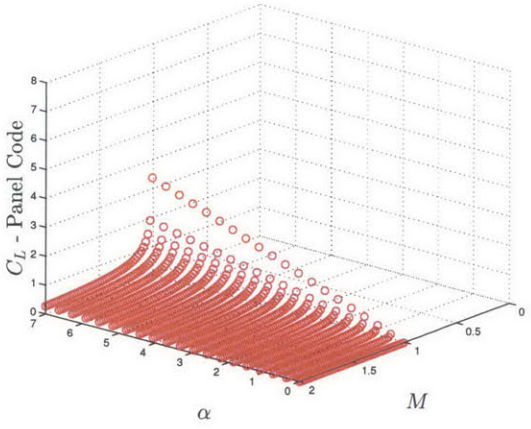


(e) Fidelity standard deviation $\sigma_{f,1}$ of SE

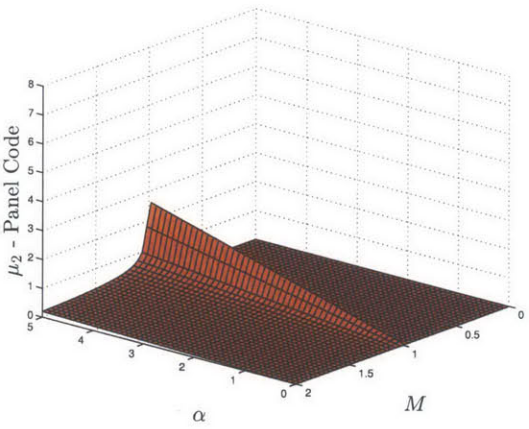


(f) Total standard deviation $\sigma_{t,1}$ of SE

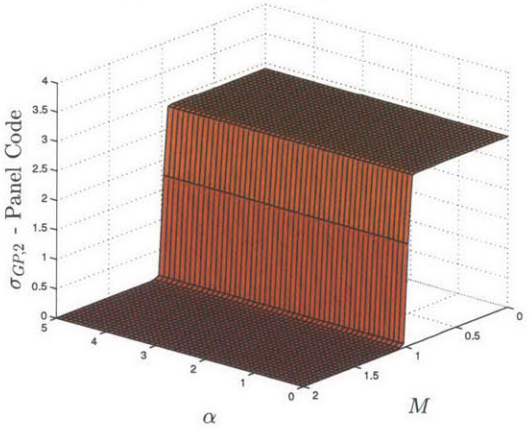
Figure 3-6: Case 2. Shock Expansion data set $S_{N,1}$ and posterior mean μ_1 and standard deviation $\sigma_{GP,1}$, fidelity standard deviation $\sigma_{f,1}$ and total standard deviation $\sigma_{t,1}$.



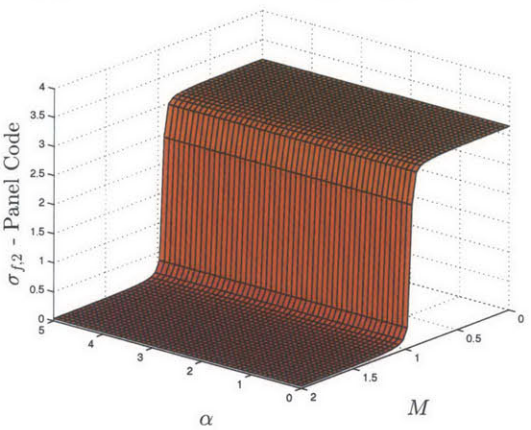
(a) Data set $S_{N,2}$ of PC



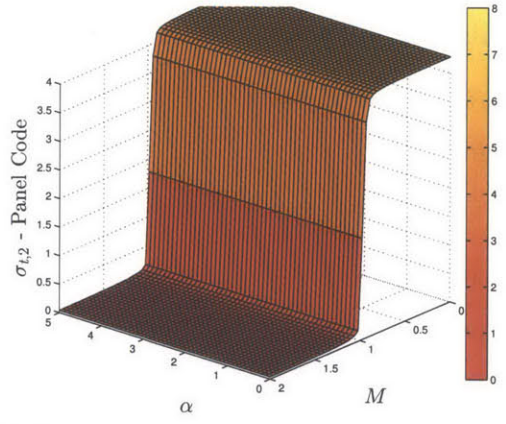
(b) Intermediate surrogate μ_2 of PC



(c) Posterior standard deviation $\sigma_{GP,2}$ of PC



(d) Fidelity standard deviation $\sigma_{f,2}$ of PC



(e) Total standard deviation $\sigma_{t,2}$ of PC

Figure 3-7: Case 2. Panel Code data set $S_{N,2}$ and posterior mean μ_2 and standard deviation $\sigma_{GP,2}$, fidelity standard deviation $\sigma_{f,2}$ and total standard deviation $\sigma_{t,2}$ (Note that the orientation of the axis has changed).

3.2.3 XFOIL

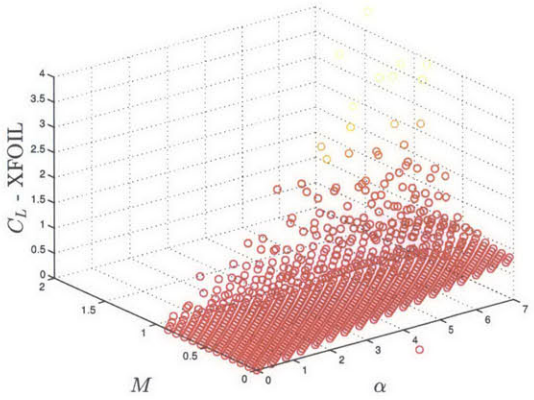
XFOIL is used in inviscid mode. Since XFOIL is only valid in subsonic regime, the design space $\mathcal{X}$ is sampled for Mach number $M \in [0, 1]$. The data set is shown on Fig. 3-8a: because of the sharp leading edge of the biconvex airfoil, we expect that the computation will lead to flows with unrealistic high velocities around that leading edge, and invalid compressibility corrections, which explain the noisy evaluations at high Mach number. This corresponds to a common situation where models are used outside of their range of validity; the proposed approach accounts for such situations by allowing to increase the fidelity variance in regions where the model is invalid.

With this data set $S_{N,3}$ (Fig. 3-8a), the intermediate surrogate μ_3 is computed (Fig. 3-8b) as well as the posterior variance $\sigma^2_{GP,3}$ (Fig. 3-8c). The noisy measurements lead to a posterior mean μ_3 that poorly approximates the lift coefficient and the posterior variance $\sigma^2_{GP,3}$ is relatively high. Recalling that, for given hyper-parameters, the posterior variance only depends on the location of the samples (not on the values of the samples), this results can seem surprising because the sampling is quite uniform. However, the hyper-parameters are computed to maximize the marginal likelihood, which depends on the observations. Hence, the noisy measurements lead to a regularizer hyper-parameter λ (or noise hyper-parameter) larger than if the data were not noisy. The regularizer hyper-parameter enters directly in the calculation of the variance through the covariance function, which explain how noisy measurements yield high posterior variance $\sigma^2_{GP,3}$.

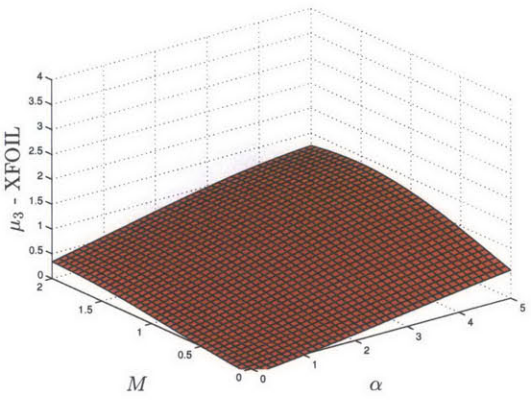
As previously mentioned, the results are not expected to capture reality for high Mach number, hence the fidelity variance $\sigma^2_{f,3}$ is set to steeply increase when the compressibility correction is not valid (Fig. 3-8d). Finally, the total variance is computed (Fig. 3-8e).

3.2.4 Combining Intermediate surrogates

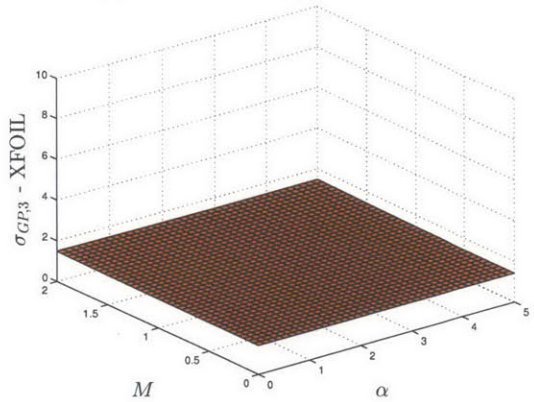
The intermediate surrogates and the multi-fidelity surrogate are shown in Fig. 3-9 and their uncertainty in Fig. 3-10. After fusion of the three intermediate surrogates,



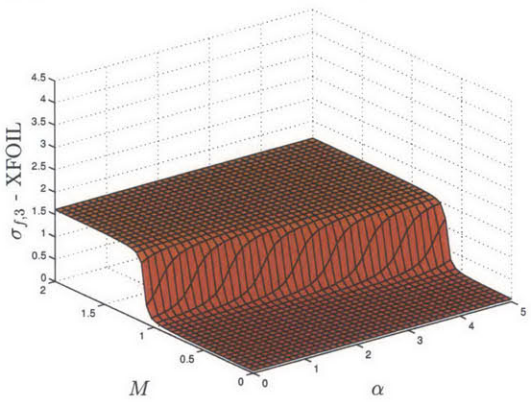
(a) Data set $S_{N,3}$ of XFOIL



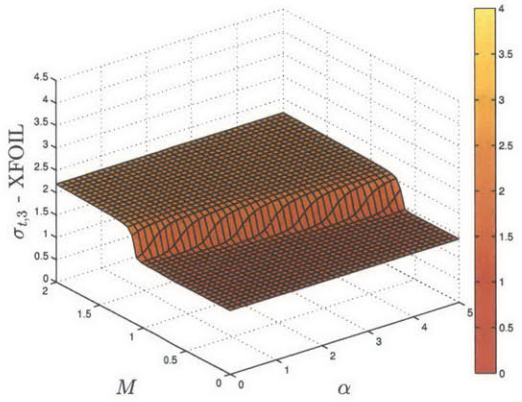
(b) Intermediate surrogate μ_3 of XFOIL



(c) Posterior standard deviation $\sigma_{GP,3}$ of XFOIL

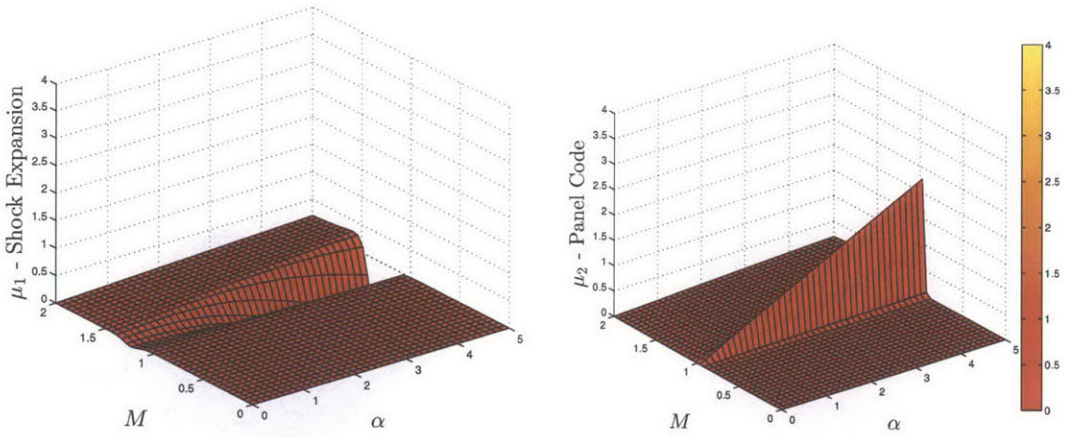


(d) Fidelity standard deviation $\sigma_{f,3}$ of XFOIL



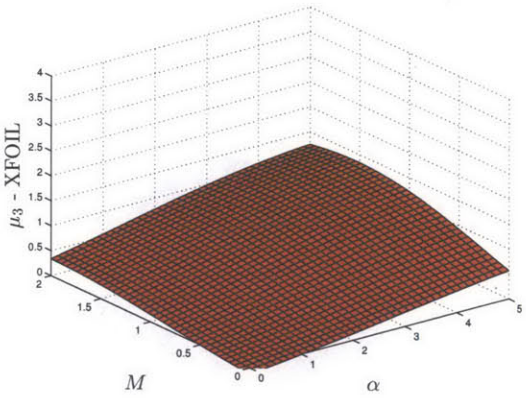
(e) Total standard deviation $\sigma_{t,3}$ of XFOIL

Figure 3-8: Case 2. XFOIL dataset $S_{N,3}$ and posterior mean μ_3 and standard deviation $\sigma_{GP,3}$, fidelity standard deviation $\sigma_{f,3}$ and total standard deviation $\sigma_{t,3}$.

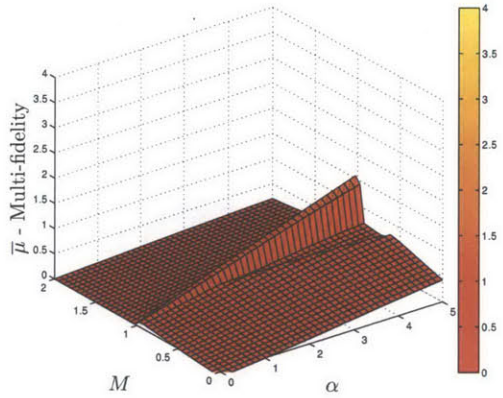


(a) Intermediate surrogate μ_1 of SE

(b) Intermediate surrogate μ_2 of PC



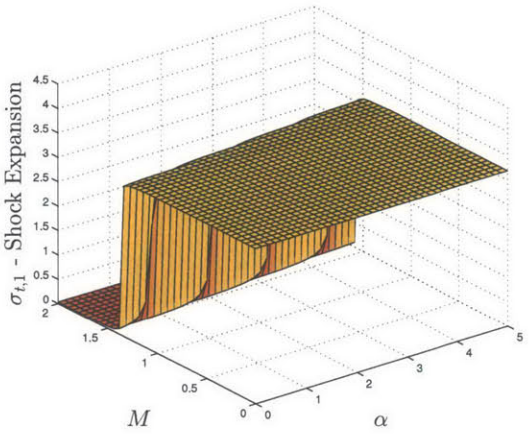
(c) Intermediate surrogate μ_3 of XFOIL



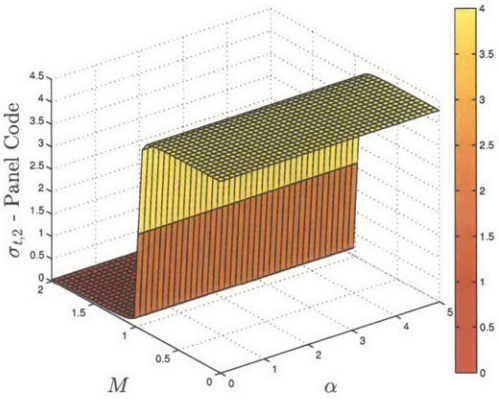
(d) Multi-fidelity surrogate $\overline{\mu}$

Figure 3-9: Case 2. Intermediate surrogate of the three models (SE, PC, XFOIL) and multi-fidelity surrogate after fusion.

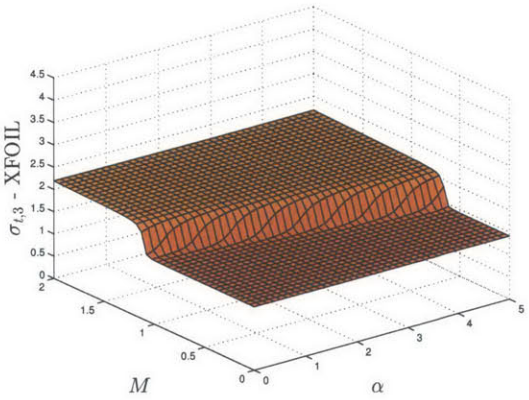
the multi-fidelity surrogate is in good agreement with the data of $S_{N,1}$ and $S_{N,2}$, but poorly approximates the subsonic behavior. In the supersonic region, the two intermediate surrogates μ_1 and μ_2 have a low uncertainty, thus μ_3 has little influence on $\overline{\mu}$ in that region. In the subsonic region, both μ_1 and μ_2 have high uncertainty. The multi-fidelity surrogate is dominated by μ_3 and its uncertainty is driven by the uncertainty of the GP of model 3 at low Mach number and by the uncertainty of the GP and the fidelity of model 3 in the transonic region. Since the uncertainty is mostly explained by model 3, the solution to reduce the variance in the subsonic regime would be to sample the transonic region with a model valid in transonic regime and improve the GP of model 3 in the low Mach number region (for instance by restricting its evaluation to the domain where the compressibility correction is valid).



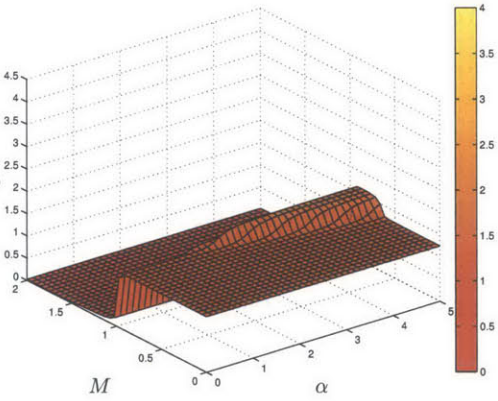
(a) Total standard deviation $\sigma_{t,1}$ of SE intermediate surrogate



(b) Total standard deviation $\sigma_{t,2}$ of PC intermediate surrogate



(c) Total standard deviation $\sigma_{t,3}$ of XFOIL intermediate surrogate



(d) Standard deviation $\overline{\sigma}$ of the multi-fidelity surrogate

Figure 3-10: Case 2. Total standard deviation of the three models (SE, PC, XFOIL) and standard deviation of the multi-fidelity surrogate after fusion.

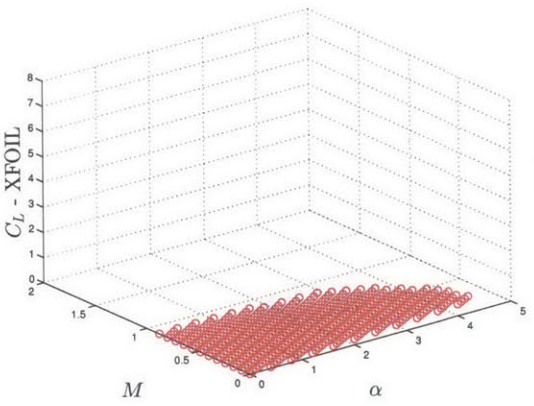
3.3 Supersonic Biconvex Airfoil: Incorporating Additional Expert Information - Case 3

We now consider the same example as Section 3.2, but bring additional expert knowledge to the data set produced by XFOIL. As previously mentioned, because of the sharp leading edge of the biconvex airfoil, we expect that the computation will lead to flows with unrealistic high velocities around that leading edge, and invalid compressibility corrections. This motivates the choice to remove from the data set $S_{N,3}$ noisy points (caused by numerical effects). We go further to ensure a "clean" data set: we remove from $S_{N,3}$ any point with a lift coefficient $C_L \geq 0.5$. This could be considered as information given by an expert on what realistic values should be.

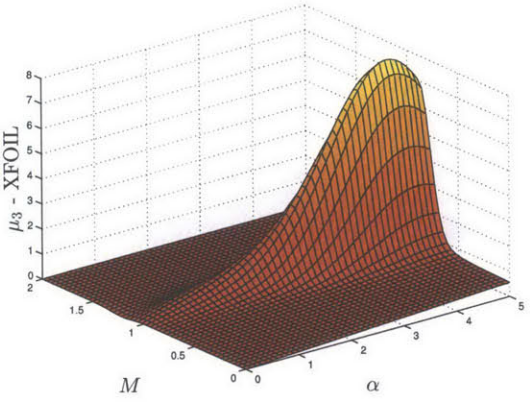
With this data set $S_{N,3}$ (Fig. 3-11a), the intermediate surrogate μ_3 is computed (Fig. 3-11b) as well as the posterior variance $\sigma^2_{GP,3}$ (Fig. 3-11c). The results are not expected to capture reality for high Mach number, hence the variance of the fidelity is set to steeply increase when the compressibility correction is not valid (Fig. 3-11d). Finally, the total variance is computed (Fig. 3-11e).

Combining Intermediate Surrogates

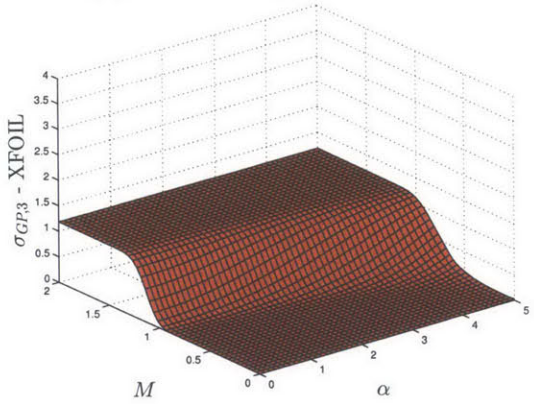
The intermediate surrogates and the multi-fidelity surrogate are shown in Fig. 3-12 and their uncertainty in Fig. 3-13. After fusion of the three surrogates, the multi-fidelity surrogate is in good agreement with the data of each data set ($S_{N,1}$, $S_{N,2}$ and $S_{N,3}$). This improvement is explained by the additional expert knowledge incorporated that significantly improved the intermediate surrogate μ_3 and reduced its total variance $\sigma^2_{t,3}$ to the fidelity variance $\sigma^2_{f,3}$. In the transonic region (where no samples are available), $\overline{\mu}$ is dominated by the intermediate surrogate μ_3 (Fig. 3-12d). The uncertainty of the multi-fidelity surrogate is high in the transonic region, indicating that the values of $\overline{\mu}$ should be used with caution in that area (Fig. 3-13d). Since the uncertainty is mostly explained by the lack of samples rather than evaluations of low fidelity, the solution to reduce the variance in the transonic regime would be to



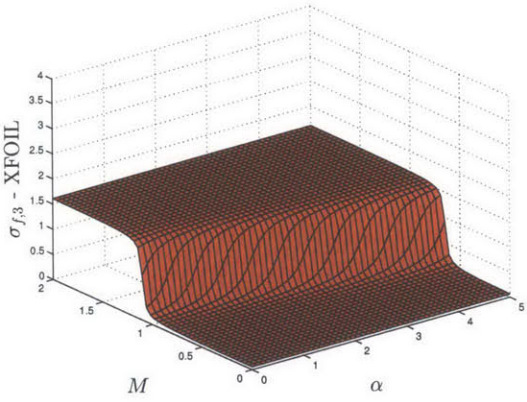
(a) Data set $S_{N,3}$ of XFOIL



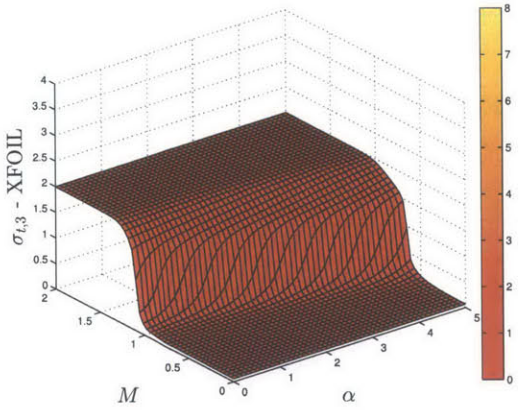
(b) Intermediate surrogate μ_3 of XFOIL



(c) Posterior standard deviation $\sigma_{GP,3}$ of XFOIL

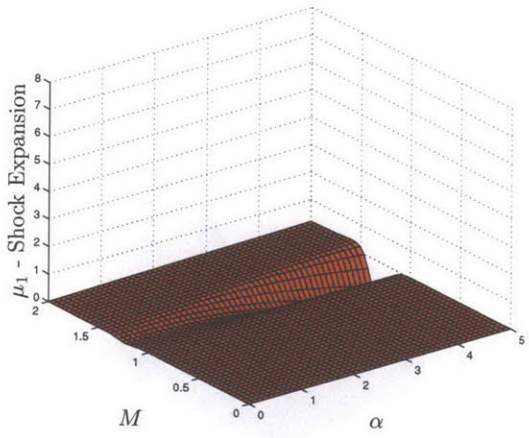


(d) Fidelity standard deviation $\sigma_{f,3}$ of XFOIL

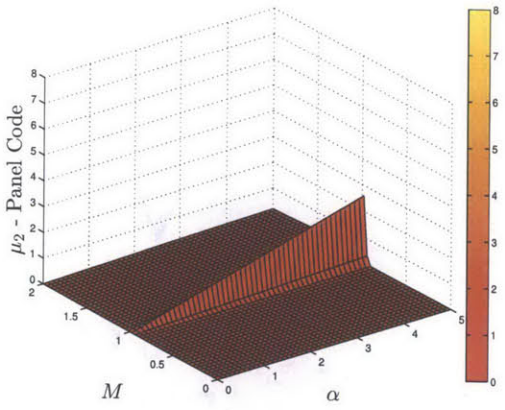


(e) Total standard deviation $\sigma_{t,3}$ of XFOIL

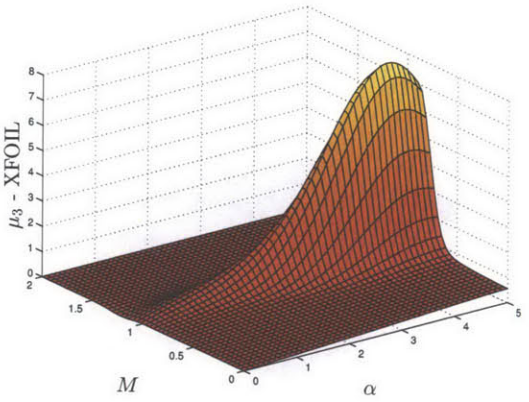
Figure 3-11: Case 3. XFOIL data set $S_{N,3}$ and posterior mean μ_3 and standard deviation $\sigma_{GP,3}$, fidelity standard deviation $\sigma_{f,3}$ and total standard deviation $\sigma_{t,3}$.



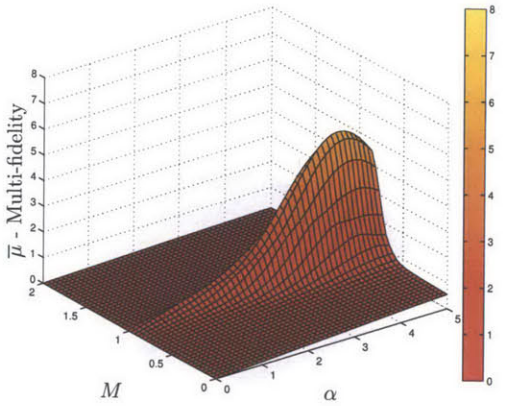
(a) Intermediate surrogate μ_1 of SE



(b) Intermediate surrogate μ_2 of PC



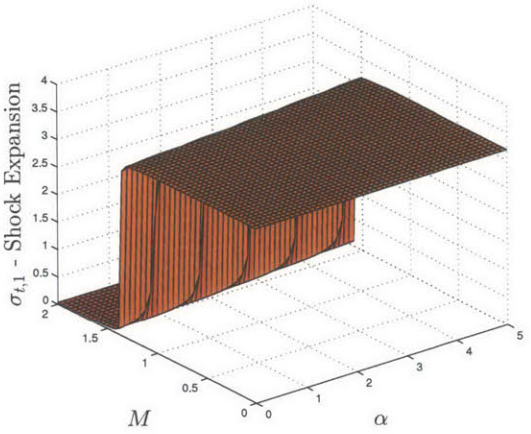
(c) Intermediate surrogate μ_3 of XFOIL



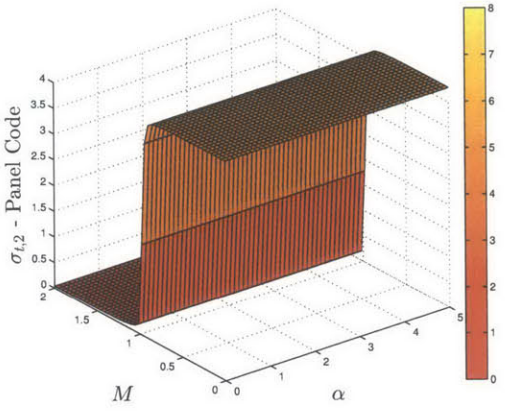
(d) Multi-fidelity surrogate $\overline{\mu}$

Figure 3-12: Case 3. Intermediate surrogate of the three models (SE, PC, XFOIL) and multi-fidelity surrogate after fusion.

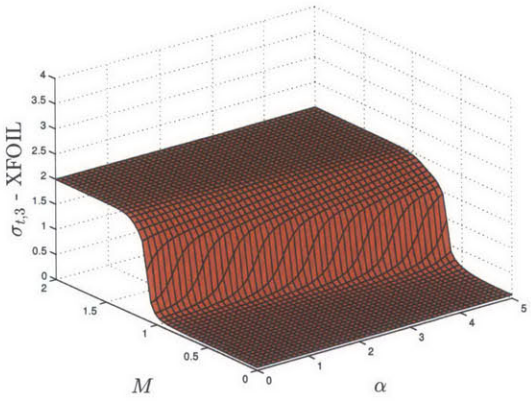
sample the transonic region with a model valid in the transonic regime.



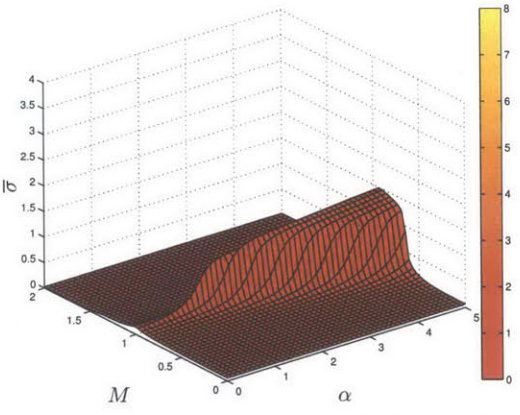
(a) Total standard deviation $\sigma_{t,1}$ of SE intermediate surrogate



(b) Total standard deviation $\sigma_{t,2}$ of PC intermediate surrogate



(c) Total standard deviation $\sigma_{t,3}$ of XFOIL intermediate surrogate



(d) Standard deviation $\overline{\sigma}$ of the multi-fidelity surrogate

Figure 3-13: Case 3. Total standard deviation of the three models (SE, PC, XFOIL) and standard deviation of the multi-fidelity surrogate after fusion.

Chapter 4

Conclusions

4.1 Summary

One of the objectives of this work is to construct a multi-fidelity surrogate able to handle non-hierarchical relationships among models. This is achieved by proposing a new quantitative definition of fidelity based on variance. This definition allows us to extend the notion of fidelity beyond the simple classification of low and high fidelity. By doing so, we are able to quantify by how much the fidelity of two models differs for a given design. This point is crucial in creating a **multi-fidelity surrogate that incorporates all the data available**, giving more weight to high fidelity information compared to lower fidelity information. The variance that quantifies the fidelity of a model is characterized by expert opinion and can vary over the design space. This allows us to handle **non-hierarchical relationships among models**. Finally, a variance-based definition of the fidelity yields a **quantification of the uncertainty of the multi-fidelity surrogate**.

The proposed approach is applied to two aerodynamic cases. In the NACA 0012 case, the multi-fidelity surrogate is shown to improve the surrogate prediction and to reduce uncertainty in regions where one model is lacking samples but the other is densely evaluated. This is an interesting feature especially in a multi-fidelity optimization setting. Indeed, it is common that an expensive model is sparsely evaluated, while a cheaper model is used to sample the design space more densely. In that situ-

ation the multi-fidelity surrogate mimics the behavior of the available higher fidelity samples , and uses the lower fidelity points elsewhere.

The biconvex airfoil case highlights how the multi-fidelity surrogate is able to identify regions where the predictions have a high uncertainty. With our approach, it is possible to pinpoint the source of uncertainty: either the fidelity of the model is too low, or the uncertainty due to sparse sampling is too high. This diagnostic is essential to devising strategies regarding which model should be evaluated next, and where. This second case also illustrated how expert opinion can be added to improve the characterization of each model's uncertainty and thus improve the quality of the multi-fidelity surrogate.

4.2 Future Work

Exploiting Model Relatedness

The proposed approach aimed at leveraging information of all available fidelities. This was achieved by constructing an intermediate surrogate for each model and fusing the information into a single multi-fidelity surrogate. This effort toward taking advantage of all data available could be developed further by exploiting the existing structure between models. Given that every model represents the same quantity of interest, these models are likely to be related in some way. In the proposed approach, this structure is not explicitly exploited. We plan on addressing this point by using machine learning techniques such multi-task learning (e.g., [5, 6]) that learn every model at the same time by jointly using their data sets . It has been shown that this can improve the learning process when several models are well correlated [31].

Relaxing the Independence Assumption of the Fusion Step

If the relationship between two models is learned in the form of a covariance [6], additional improvement can be made to the proposed approach. Specifically, in the fusion step of the algorithm, the assumption of independence of the random variable

$h_{*,m}$ representing the intermediate surrogate of model f_m can be relaxed to fully integrate the covariance terms [40]. This would result in a more realistic estimation of the variance of the multi-fidelity surrogate, avoiding over-confident predictions.

Optimization

Future work also includes developing optimization strategies leveraging multi-fidelity surrogates. These strategies needs to address two possibly competitive goals:

- exploring the design space to find the optimum of the quantity of interest
- sampling designs that will improve the multi-fidelity surrogate in a sequential manner

The challenge is to devise algorithms that not only select a design to evaluate, but also suggest which model should be used for the evaluation based on its fidelity.

Bibliography

[1] 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*, 15(1):16–23, 1998.

[2] 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*, 38(6):1093–1101, 2001.

[3] D. L. Allaire and K. E. Willcox. A variance-based sensitivity index function for factor prioritization. *Reliability Engineering & System Safety*, 107:107–114, 2012.

[4] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[5] E. V. Bonilla, F. V. Agakov, and C. Williams. Kernel multi-task learning using task-specific features. In *International Conference on Artificial Intelligence and Statistics*, volume 20, pages 43–50, 2007.

[6] E. V. Bonilla, K. M. A. Chai, and C. Williams. Multi-task gaussian process prediction. In *Neural Information Processing Systems*, 2007.

[7] A. J. Booker, J. E. 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*, 17(1):1–13, 1999.

[8] S. Choi, J. J. Alonso, I. M. Kroo, and M. Wintzer. Multifidelity design optimization of low-boom supersonic jets. *Journal of Aircraft*, 45(1):106–118, 2008.

[9] H. Chung and J. J. Alonso. Design of a low-boom supersonic business jet using cokriging approximation models. *AIAA Paper*, 5598:2002, 2002.

[10] H. Chung and J. J. Alonso. Using gradients to construct cokriging approximation models for high-dimensional design optimization problems. *AIAA Paper*, 317:14–17, 2002.

[11] N. Cressie and C. K. Wikle. *Statistics for Spatio-temporal Data*. John Wiley & Sons, Hoboken, NJ, 2011.

[12] M. Drela. Xfoil: An analysis and design system for low reynolds number airfoils. In *Low Reynolds Number Aerodynamics*. Springer, Dordrecht, Germany, 1989.

[13] M. Drela. *Flight Vehicle Aerodynamics*. MIT Press, Cambridge, MA, 2014.

[14] M. Drela and M. B. Giles. Viscous-inviscid analysis of transonic and low reynolds number airfoils. *AIAA Journal*, 25(10):1347–1355, 1987.

[15] T. Evgeniou, M. Pontil, and T. Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13(1):1–50, 2000.

[16] A. I. J. Forrester, N. W. Bressloff, and A. J. Keane. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 462(2071):2177–2204, 2006.

[17] A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50–79, January 2009.

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

[19] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani. *The Elements of Statistical Learning*, volume 2. Springer, Dordrecht, Germany, 2009.

[20] D. Huang, T. T. Allen, W. I. Notz, and R. A. Miller. Sequential Kriging optimization using multiple-fidelity evaluations. *Structural and Multidisciplinary Optimization*, 32(5):369–382, 2006.

[21] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

[22] A. J. Keane. Wing optimization using design of experiment, response surface, and data fusion methods. *Journal of Aircraft*, 40(4):741–750, 2003.

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

[24] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[25] O. P. Le Maître and O. M. Knio. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer, Dordrecht, Germany, 2010.

[26] S. J. Leary, A. Bhaskar, and A. J. Keane. A knowledge-based approach to response surface modelling in multifidelity optimization. *Journal of Global Optimization*, 26(3):297–319, 2003.

[27] A. March and K. E. Willcox. Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives. *AIAA Journal*, 50(5):1079–1089, 2012.

[28] A. I. March. *Multidelity methods for multidisciplinary system design.* PhD thesis, Massachusetts Institute of Technology, MA, June 2012.

[29] G. Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963.

[30] F. Palacios, J. J. Alonso, K. Duraisamy, M. Colonno, J. Hicken, A. Aranake, A. Campos, S. Copeland, T. D. Economon, A. Lonkar, et al. Stanford university unstructured (su2): An open-source integrated computational environment for multi-physics simulation and design. In *51st AIAA Aerospace Sciences Meeting and Exhibit*, 2013.

[31] S. J. Pan and Q. Yang. A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on*, 22(10):1345–1359, 2010.

[32] P. Piperni, A. DeBlois, and R. Henderson. Development of a multilevel multidisciplinary-optimization capability for an industrial environment. *AIAA Journal*, 51(10):2335–2352, October 2013.

[33] D. Rajnarayan. *Trading risk and performance for engineering design optimization using multifidelity analyses.* PhD thesis, Stanford University, CA, June 2009.

[34] D. Rajnarayan, A. Haas, and I. Kroo. A multifidelity gradient-free optimization method and application to aerodynamic design. In *12th AIAA/ISSMO multidisciplinary analysis and optimization conference, Victoria, British Columbia, AIAA*, 2008.

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

[36] A. Saltelli, K. Chan, E. M. Scott, et al. *Sensitivity analysis.* Wiley, New York, 2000.

[37] 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*, 39(12):2233–2241, 2001.

[38] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.

[39] G. Wahba. *Spline Models for Observational Data.* SIAM, Philadelphia, PA, 1990.

[40] R. L. Winkler. Combining probability distributions from dependent information sources. *Management Science*, 27(4):479–488, 1981.