

# Multifidelity Optimization using Statistical Surrogate Modeling for Non-Hierarchical Information Sources

Remi Lam\*

*Massachusetts Institute of Technology, Cambridge, MA, 02139, USA*

Douglas Allaire<sup>†</sup>

*Texas A&M University, College Station, TX, 77843, USA*

Karen Willcox<sup>‡</sup>

*Massachusetts Institute of Technology, Cambridge, MA, 02139, USA*

Designing and optimizing complex systems often requires numerous evaluations of a quantity of interest. This is typically achieved by querying potentially expensive numerical models in an optimization process. To alleviate the cost of optimization, surrogate models can be used in lieu of the original model, as they are cheaper to evaluate. In addition, different information sources with varying fidelity, such as numerical models, experimental results or historical data may be available to estimate the quantity of interest. This work proposes a strategy to adaptively construct and exploit a multifidelity surrogate when multiple information sources of varying fidelity are available. One of the distinguishing features of the proposed approach is the relaxation of the common assumption of hierarchical relationships among information sources. This is achieved by endowing the surrogate representation with uncertainty functions that can vary across the design space; this uncertainty quantifies the fidelity of the underlying information source. The resulting multifidelity surrogate is used in an optimization setting to identify the next design to evaluate, as well as to select the information sources with which to perform the evaluation, based on information source evaluation cost and fidelity. For an aerodynamic design example, the proposed strategy leverages multifidelity information to reduce the number of evaluations of the expensive information source needed during the optimization.

## I. Introduction

Designing and optimizing complex systems generally requires the evaluation for different designs of quantities of interest describing designer preferences and design requirements. Numerical optimization achieves this by evaluating a model: a mapping from the design space to estimates of the quantities of interest. It is often the case that the quantities of interest can be evaluated by different means. For instance, they could be computable with numerical models, measurable by experiments, or informed by historical data. We refer to these as different *information sources* (IS), where each IS has an associated fidelity that represents its ability to estimate the true value of a quantity of interest. To speed up the optimization process, it is desirable to leverage these multiple IS. One way to do this is to construct cheap-to-evaluate surrogates that embody the fidelity of the underlying IS.

Strategies to perform multifidelity optimization have been studied extensively in the case where a hierarchy of fidelity exists among models. These include creating response surface surrogates using both low and high-fidelity model evaluations,<sup>1</sup> or computing higher fidelity model samples when the difference between two lower fidelity evaluations is larger than a threshold.<sup>2</sup> Multifidelity optimization strategies can be divided into global and local approaches. Global approaches try to find the best design in the entire

---

\*Graduate student, Department of Aeronautics and Astronautics, rlam@mit.edu.

<sup>†</sup>Assistant Professor, Department of Mechanical Engineering, dallaire@tamu.edu, Member AIAA.

<sup>‡</sup>Professor, Department of Aeronautics and Astronautics, kwillcox@mit.edu, Associate Fellow AIAA.

feasible domain: existing strategies include building an interpolation of the high-fidelity model as in Efficient Global Optimization<sup>3</sup> (EGO). A multi-objective version of EGO developed in Rajnarayan<sup>4</sup> and Rajnarayan et al.<sup>5</sup> combines the use of a surrogate to both minimize the quantity of interest and explore unevaluated regions. In Kennedy and O’Hagan,<sup>6</sup> Huang et al.<sup>7</sup> and Xiong and Chen,<sup>8</sup> Gaussian processes are used to approximate the difference between a high-fidelity and a lower fidelity model. Co-Kriging methods have been developed to build multifidelity surrogates.<sup>9,10,11,12</sup> Moore et al.<sup>13</sup> sequentially construct a Kriging surrogate using several multifidelity models, where the fidelity is quantified by a metric accounting for the correlation between each model and the real world process. In that work, the next design to evaluate and the model to query are determined by maximizing a value of information metric. Correlations between models have also been exploited to create a multifidelity approach for optimization under uncertainty.<sup>14</sup>

Local multifidelity optimization approaches, on the other hand, focus the search on finding a local optimum. Booker et al.<sup>15</sup> developed a gradient-free pattern search algorithm for multifidelity optimization. When gradients are available, trust region methods use a local approximation of the high-fidelity model to perform the optimization. One significant extension of the trust region method to multifidelity optimization is presented by Alexandrov et al.<sup>16,17</sup> In that work, 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. This enables a guarantee that the multifidelity trust region method is provably convergent with respect to an optimum of the high-fidelity model. March and Willcox<sup>18</sup> extended a multifidelity trust region approach to the derivative-free setting.

In this work, we propose a method based on statistical techniques to adaptively build a multifidelity surrogate for multifidelity optimization. We propose 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 IS. 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 multifidelity surrogate is constructed by fusing all the intermediate surrogates. One of the advantages of the approach is the multifidelity surrogate capability of integrating IS whose fidelity changes over the design space, thus relaxing the common assumption of hierarchical relationships among IS. We also introduce an optimization algorithm that leverages our multifidelity surrogate. The exploration of the design space is based on an extension of the expected improvement criteria to balance the need to improve the surrogate and its exploitation. A heuristic selects the next IS to query, taking into account evaluation cost and fidelity.

The proposed multifidelity surrogate approach is applied to the aerodynamic example of computing the lift coefficient of a NACA 0012 airfoil in the subsonic regime. In this example, the multifidelity 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 multifidelity surrogate and identify whether the fidelity or the sampling is the principal source of this uncertainty. The proposed optimization algorithm is applied to minimize the drag coefficient of a NACA 0012 airfoil with a constraint on lift coefficient. It is shown to reduce the overall evaluation cost of the optimization in the case where one IS is more expensive than the others.

This paper is organized as follows. Section II presents the construction of the multifidelity surrogate. Section III details the proposed optimization algorithm. In Section IV, the multifidelity optimization algorithm is applied to an aerodynamic example. Finally, Section V concludes the paper.

## II. Constructing a Multifidelity Surrogate

This section describes the proposed approach for constructing multifidelity surrogates. This consists of the following two steps. First, for each IS, construct an intermediate surrogate using a Gaussian process to define the mean and associated variance quantifying the uncertainty. Second, build the multifidelity surrogate by fusing information from the intermediate surrogates.

### II.A. Building an Intermediate Surrogate with Gaussian Processes

#### II.A.1. Mean of the Intermediate Surrogate

We consider the case where we have  $M$  IS available  $f_1, \dots, f_M$ , all mapping from the design space  $\mathcal{X} \subset \mathbb{R}^d$  to  $\mathbb{R}$ . The  $m^{\text{th}}$  IS is evaluated at a finite number of designs, specified by the training set  $\mathcal{D}_m$ , which is

composed of  $N_m$  evaluated design points written in matrix form as  $X_m = [\mathbf{x}_{1,m}, \dots, \mathbf{x}_{N_m,m}] \in \mathbb{R}^{d \times N_m}$  and  $N_m$  performances estimates written in vector form as  $\mathbf{y}_m = [y_{1,m}, \dots, y_{N_m,m}]^\top \in \mathbb{R}^{N_m}$  with  $\mathbf{x}_{i,m}$  the  $i^{\text{th}}$  evaluated design point of the  $m^{\text{th}}$  IS and  $y_{i,m} = f_m(\mathbf{x}_{i,m})$  its associated performance.

The first step in our approach is to construct an intermediate surrogate for each IS  $f_m$  using Gaussian processes (see Rasmussen and Williams<sup>19</sup> for an overview of Gaussian processes). We choose a square exponential covariance function with additive Gaussian noise and select the hyper-parameters of the covariance function as well as the noise variance with the maximum marginal likelihood method. The hyper-parameters and the noise variance depend on the data set  $\mathcal{D}_m$ , leading to a different covariance function  $k_m$  and noise variance  $\lambda_m$  for each IS  $f_m$ . The posterior mean  $\mu_m$  of each Gaussian process can be computed in closed form for any unevaluated design  $\mathbf{x} \in \mathcal{X}$  and used as a surrogate of  $f_m$ :

$$\mu_m(\mathbf{x}) = K_m(X_m, \mathbf{x})^\top [K_m(X_m, X_m) + \lambda_m I]^{-1} \mathbf{y}_m, \quad (1)$$

with  $K_m(X_m, X_m)$  the  $N_m \times N_m$  matrix whose  $ij^{\text{th}}$  entry is  $k_m(\mathbf{x}_{i,m}, \mathbf{x}_{j,m})$  and  $K_m(X_m, \mathbf{x})$  the  $N_m \times 1$  vector whose  $i^{\text{th}}$  entry is  $k_m(\mathbf{x}_{i,m}, \mathbf{x})$ .

### II.A.2. Sources of Uncertainty in Building an Intermediate Surrogate

There are different sources of uncertainty that arise in the construction of the proposed multifidelity surrogate. In this section, we briefly describe these uncertainties and their models.

#### Variance Associated with Gaussian Process

For each IS  $f_m$ , after conditioning the Gaussian process with the training data  $\mathcal{D}_m$ , the prior statistics are updated and a posterior variance  $\sigma_{GP,m}^2$  can be computed everywhere on the design space

$$\sigma_{GP,m}^2(\mathbf{x}) = k_m(\mathbf{x}, \mathbf{x}) - K_m(X_m, \mathbf{x})^\top [K_m(X_m, X_m) + \lambda_m I]^{-1} K_m(X_m, \mathbf{x}). \quad (2)$$

The prior variance is reduced where training data are available, forming uncertainty bubbles (Fig. 1 left panel). When the training points are sparse in the design space,  $\sigma_{GP,m}^2$  is an indicator of the quality of the sampling.

#### Variance Associated with Fidelity

We consider a second source of uncertainty to be that inherent to the IS used to produce the data set. This uncertainty is the model inadequacy defined by Kennedy and O’Hagan<sup>20</sup> 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 knowing the true mean value of the real world process. Instead, we associate model inadequacy with fidelity and quantify it for IS  $f_m$  with a fidelity variance  $\sigma_{f,m}^2$ . This fidelity variance provides a lower bound on the uncertainty of the intermediate surrogate (i.e., the surrogate cannot be more accurate than the underlying IS used to create it). Unlike the traditional view of multifidelity modeling, which imposes a hierarchy among the IS, we consider the fidelity to be free to change across the design space. We achieve this by letting  $\sigma_{f,m}^2(\mathbf{x})$  be a function of the design variable  $\mathbf{x} \in \mathcal{X}$ . This function is specific to each IS  $f_m$ . In this work, we assume it to be provided as an input by an expert. This approach to modeling fidelity is similar in spirit but differs in the mathematical representation from that in Moore et al.,<sup>13</sup> which models the correlation between an IS and the real world process.

#### Total Variance of an Intermediate Surrogate

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 IS, we propose the definition of a total variance  $\sigma_{t,m}^2$  for the IS  $f_m$ :

$$\forall \mathbf{x} \in \mathcal{X}, \sigma_{t,m}^2(\mathbf{x}) = \sigma_{GP,m}^2(\mathbf{x}) + \sigma_{f,m}^2(\mathbf{x}), \quad (3)$$

where  $\sigma_{GP,m}^2(\mathbf{x})$  and  $\sigma_{f,m}^2(\mathbf{x})$  are respectively the variance associated with the Gaussian process and the fidelity variance of IS  $f_m$ .

The total variance  $\sigma_{t,m}^2$  is bounded below by the fidelity variance  $\sigma_{f,m}^2$ , 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 IS itself. Fig. 1 illustrates how the intermediate surrogate is built for a given IS  $f_m$ .

For any design  $\mathbf{x} \in \mathcal{X}$ , and each IS  $f_m$ , we define the  $m^{\text{th}}$  intermediate surrogate at  $\mathbf{x}$  to be the random variable  $S_{\mathbf{x},m}$  with the following distribution

$$S_{\mathbf{x},m} \sim \mathcal{N}(\mu_m(\mathbf{x}), \sigma_{t,m}^2(\mathbf{x})), \quad (4)$$

where  $\mathcal{N}$  denotes a Gaussian distribution, and the mean  $\mu_m(\mathbf{x})$  and variance  $\sigma_{t,m}^2(\mathbf{x})$  are defined by Eqs. (1) and (3).

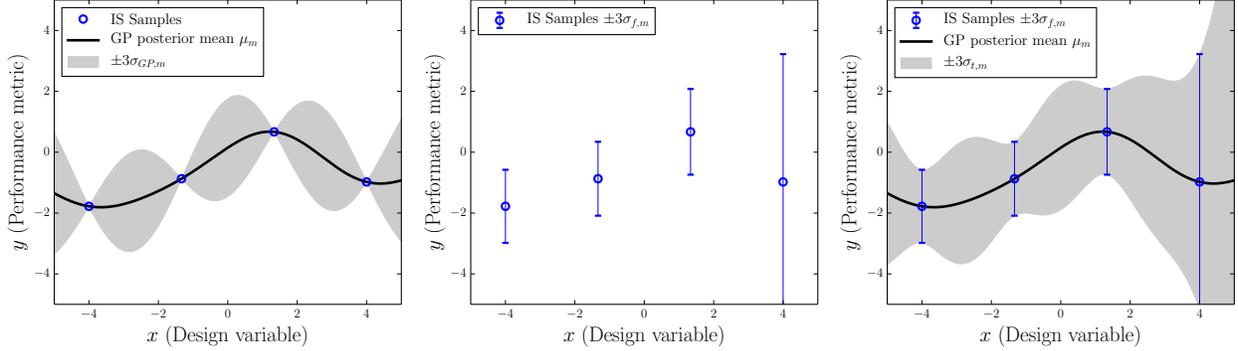


Figure 1: Left panel: Posterior mean  $\mu_m$  and standard deviation  $\sigma_{GP,m}$  computed for a given IS. Middle panel: Samples with associated fidelity standard deviation  $\sigma_{f,m}$ . Right panel: Intermediate surrogate with mean  $\mu_m$  and total standard deviation  $\sigma_{t,m}$ .

## II.B. Building a Multifidelity Surrogate: Fusion of Information

Once the intermediate surrogates have been computed for each of the  $M$  IS, a single multifidelity surrogate can be built combining all the information available. For any  $\mathbf{x}$  in  $\mathcal{X}$ , following Winkler,<sup>21</sup> we fuse the random variables  $S_{\mathbf{x},1}, \dots, S_{\mathbf{x},M}$ , representing the intermediate surrogates, into a single random variable  $S_{\mathbf{x}}$  to obtain our multifidelity surrogate:

$$S_{\mathbf{x}} \sim \mathcal{N}(\bar{\mu}(\mathbf{x}), \bar{\sigma}^2(\mathbf{x})). \quad (5)$$

Fig. 2 illustrates this notion for  $M = 2$  IS. We make the simplifying assumption that the  $S_{\mathbf{x},m}$  are uncorrelated random variables. Characterizing the correlation between the IS is a challenging but important topic of future work. Given this simplifying assumption, the mean  $\bar{\mu}(\mathbf{x})$  and variance  $\bar{\sigma}^2(\mathbf{x})$  of the multifidelity surrogate can be written

$$\bar{\sigma}^2(\mathbf{x}) = \left( \sum_{m=1}^M \frac{1}{\sigma_{t,m}^2(\mathbf{x})} \right)^{-1} \quad (6)$$

$$\bar{\mu}(\mathbf{x}) = \bar{\sigma}^2(\mathbf{x}) \sum_{m=1}^M \frac{\mu_m(\mathbf{x})}{\sigma_{t,m}^2(\mathbf{x})}. \quad (7)$$

The mean  $\bar{\mu}(\mathbf{x})$  of the multifidelity surrogate  $S_{\mathbf{x}}$  then provides an approximate estimation of the quantity of interest  $\mu(\mathbf{x})$ .

## II.C. Application to NACA 0012 Airfoil Lift Coefficient Modeling

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 performance. Since the fluid dynamics are governed by the Navier-Stokes equations, which are expensive to solve, simplified equations have been developed to solve this problem (e.g., Reynolds Averaged Navier-Stokes equations, Euler equations). These simplified equations are typically derived using additional assumptions and are valid in given regimes. This extended set of equations

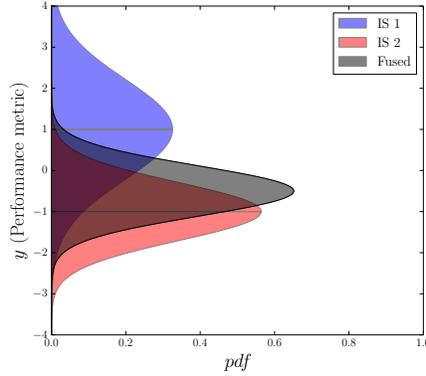


Figure 2: At a given design  $\mathbf{x}$ , the random variable  $S_{\mathbf{x},1}$  and  $S_{\mathbf{x},2}$  associated with the Gaussian process of IS 1 and 2 have the distributions represented by the blue and red shadings. The black shading is the distribution of the random variable  $S_{\mathbf{x}}$  after fusion: a normally distributed random variable of mean  $\bar{\mu}(\mathbf{x})$  and variance  $\bar{\sigma}^2(\mathbf{x})$ .

available to solve the same problem leads to information of variable 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 some regimes but not in others. This motivates the new definition of fidelity proposed in this paper, which no longer associates a unique fidelity to an IS, as in the traditional setting, but extends the notion allowing the fidelity of an IS 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 multifidelity surrogates, especially in an optimization context that requires multiple evaluations. This section demonstrates the multifidelity surrogate approach developed in this paper on an aerodynamics example: characterizing the lift coefficient  $C_L$  of a NACA 0012 airfoil in the subsonic regime.

### II.C.1. Problem Formulation

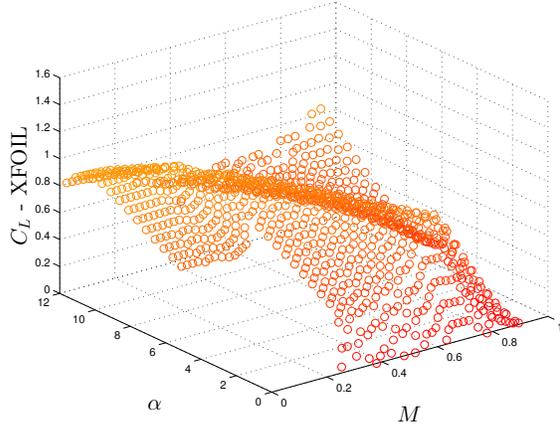
In this example, the quantity of interest is the lift coefficient  $C_L$  of the airfoil, the design space  $\mathcal{X}$  is two dimensional:  $\mathcal{X} = I_M \times I_\alpha$  where  $I_M = [0, 1]$  is the range of the Mach number  $M$  and  $I_\alpha = [0, 12^\circ]$  is the range of the angle of attack  $\alpha$ . There are two IS available:  $f_1$  is the lift coefficient computed by XFOIL<sup>22</sup> with viscous terms, and  $f_2$  is the lift coefficient computed by SU2<sup>23</sup> using Euler equations. These models are described in the following subsections.

### II.C.2. XFOIL

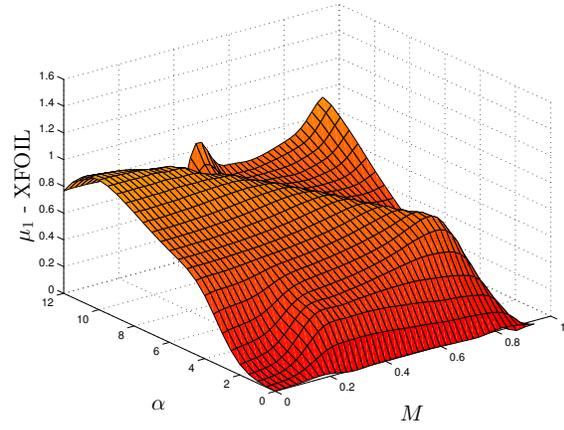
XFOIL is a solver for the design and analysis of airfoils in the subsonic regime. It couples a panel method with the Karman-Tsien compressibility correction for the potential flow with a two-equation boundary layer model. The laminar-turbulent transition is governed by an  $e^N$  envelope criterion. More about the XFOIL framework can be found in Drela,<sup>22</sup> the boundary layer treatment is described in Drela and Giles,<sup>24</sup> and a general overview of the panel method used can be found in Drela.<sup>25</sup>

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  $5^\circ$  (Fig. 3a) for a Reynolds number  $R_e = 10^5$ . Based on this data set  $\mathcal{D}_1$ , the posterior mean  $\mu_1$  of the intermediate surrogate for IS 1 is computed (Fig. 3b) as well as the posterior variance  $\sigma_{GP,1}^2$ . Fig. 3c shows the posterior standard deviation of the Gaussian process for XFOIL. It can be seen that the uncertainty of the Gaussian process 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 number, low angle of attack region.

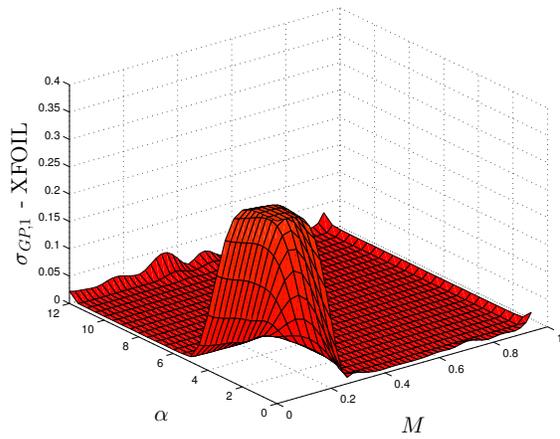
XFOIL is a solver for subsonic airfoils, and the compressibility correction used is only valid for relatively low Mach number, hence the lift coefficient computed at high Mach number ( $M \geq 0.65$ ) has a low fidelity. Similarly, for high angle of attack, physical phenomena, such as separation, are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



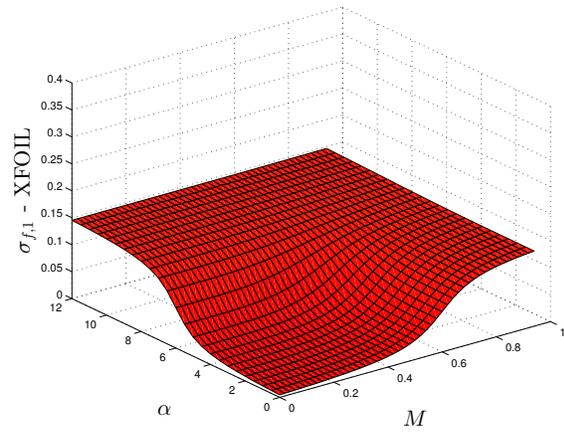
(a)  $C_L$  samples  $\mathcal{D}_1$  computed with XFOIL



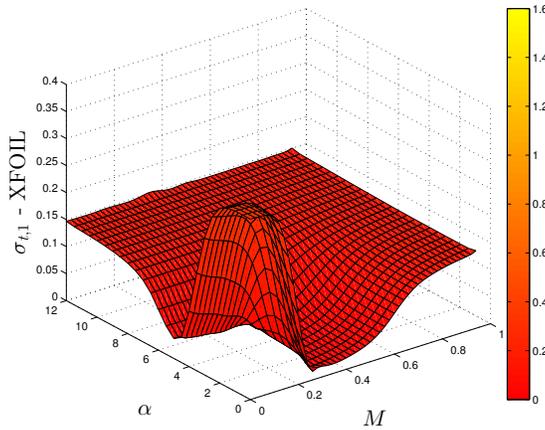
(b) Intermediate surrogate mean using XFOIL



(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: XFOIL samples  $\mathcal{D}_1$ , posterior mean  $\mu_1$  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.

The total variance  $\sigma_{t,1}^2$  of the intermediate surrogate of XFOIL is computed using Eq. (3) and is shown on Fig. 3e. The uncertainty is dominated by the fidelity uncertainty at high Mach number and high angle of attack, but dominated by the uncertainty of the Gaussian process in the low Mach number and low angle of attack area.

### II.C.3. SU2 Euler Code

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 Palacios et al.<sup>23</sup> Because SU2 is more expensive to evaluate than XFOIL, the design space is sampled uniformly on  $\mathcal{X}$  but more sparsely (Fig. 4a). Those evaluations define a second data set  $\mathcal{D}_2$  used to construct the intermediate surrogate for  $f_2$  (Fig. 4b).

The posterior variance  $\sigma_{GP,2}^2$  exhibits an oscillatory behavior but only in one direction of the design space Fig. 4c. 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 Mach number  $M$  appears sparser than the sampling in angle of attack  $\alpha$ , leading to uncertainty bubbles along the Mach number axis.

Given that the Euler equations do not account for viscous terms, the fidelity variance of SU2 is low at low Mach number and low angle of attack (Fig. 4d). In this region,  $\sigma_{f,2} \geq \sigma_{f,1}$  because XFOIL has a better representation of the physics than the Euler equations solved by SU2.

The total variance  $\sigma_{t,2}^2$  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 Gaussian process in the low Mach number, low angle of attack region (Fig. 4e).

### II.C.4. Combining Intermediate Surrogates

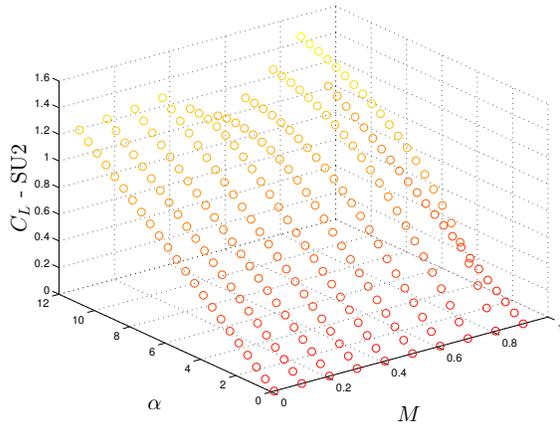
Fig. 5 shows the results after combining the two intermediate surrogate means  $\mu_1$  and  $\mu_2$ : the multifidelity surrogate mean  $\bar{\mu}$  exhibits the behavior of  $\mu_1$  (respectively  $\mu_2$ ) in regions where the uncertainty of  $\mu_1$  (respectively  $\mu_2$ ) is low (Fig. 5e). The uncertainty in the multifidelity surrogate is high in the region where neither model has a high fidelity (high Mach number and high angle of attack). However, the uncertainty has been reduced in the low Mach number, low angle of attack region, where IS 1 (XFOIL) is lacking samples (Fig. 5f) but IS 2 (SU2) has been sampled uniformly.

## III. Optimizing Using a Multifidelity Surrogate

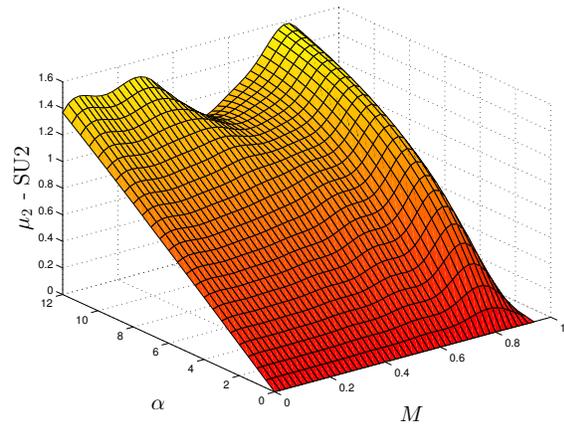
We now introduce an algorithm that adaptively builds the multifidelity surrogate introduced in the previous section and leverages it to solve an optimization problem. We first formulate the optimization problem in the case where multiple IS are available. Then, we propose a metric to choose the next design to evaluate and introduce a heuristic to choose which IS to query. This new training point is used to adaptively update the multifidelity surrogate. Finally, we illustrate how constraints can be added.

### III.A. Problem Formulation

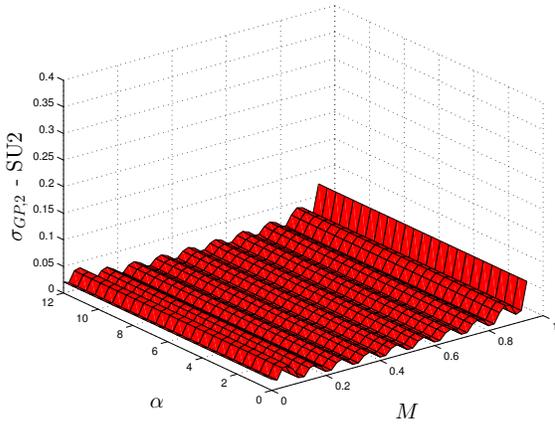
When a single IS  $f_1$  is available, writing the (unconstrained) optimization problem is straightforward:  $\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} f_1(\mathbf{x})$ . However, the optimization problem needs a careful definition in the case where several IS with varying fidelity are available. In this paper, we consider each IS to be uncertain but containing relevant information about the real world process. In particular, none of the IS is considered to be the real world process or “truth.” As stated in Section II, each IS  $f_m$  is endowed with a fidelity variance  $\sigma_{f,m}^2$  that quantifies its fidelity. For any design  $\mathbf{x} \in \mathcal{X}$ , the IS  $f_m$  is represented by a Gaussian random variable  $P_{\mathbf{x},m} \sim \mathcal{N}(f_m(\mathbf{x}), \sigma_{f,m}^2(\mathbf{x}))$ . Following Winkler,<sup>21</sup> we define the estimation of the real world process, given  $M$  different IS all evaluated at design  $\mathbf{x}$ , to be  $P_{\mathbf{x}}$ : the fusion of the  $M$  random variables  $P_{\mathbf{x},m}$ . In the simplifying case where the  $P_{\mathbf{x},m}$  are independent Gaussian random variables, the estimation of the real world process is the random variable  $P_{\mathbf{x}} \sim \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$  with:



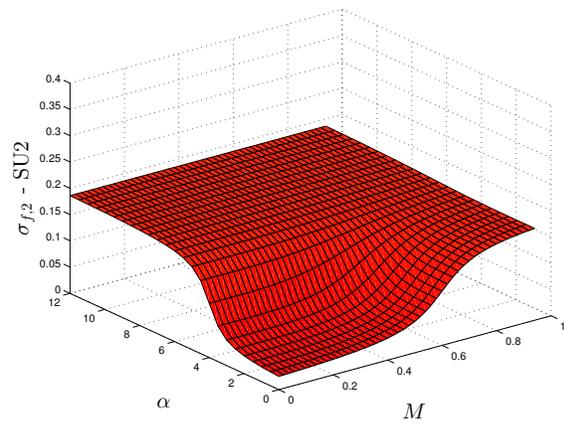
(a)  $C_L$  samples  $\mathcal{D}_2$  computed with SU2



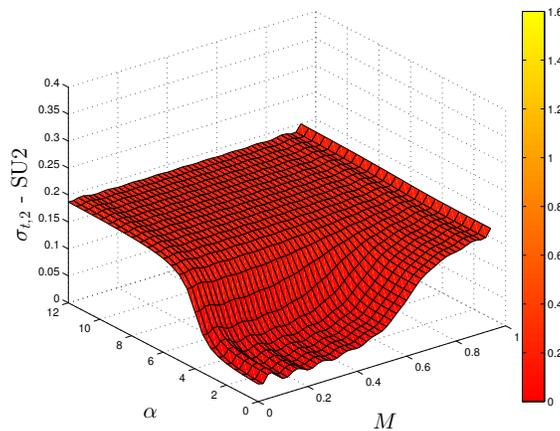
(b) Intermediate surrogate mean using SU2



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

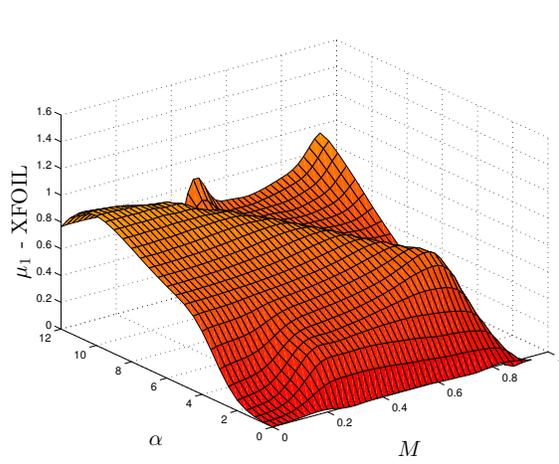


(d) Fidelity standard deviation  $\sigma_{f,2}$  of SU2

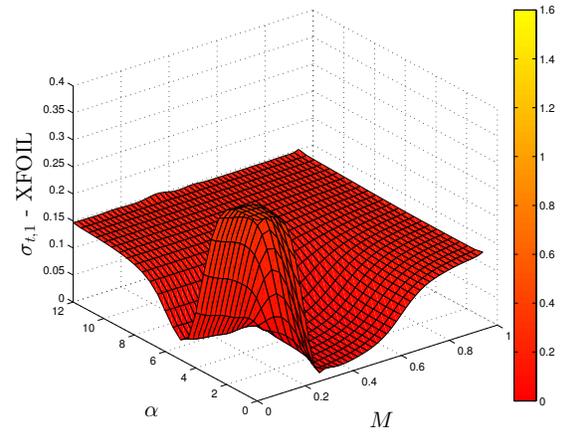


(e) Total standard deviation  $\sigma_{t,2}$  of SU2 surrogate

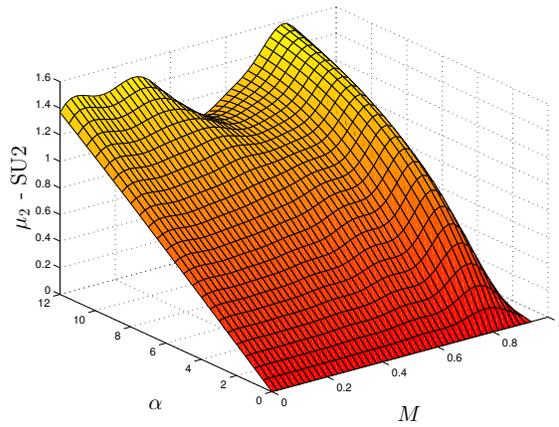
Figure 4: SU2 samples  $\mathcal{D}_2$ , posterior mean  $\mu_2$  and standard deviation  $\sigma_{GP,2}$ , fidelity standard deviation  $\sigma_{f,2}$  and total standard deviation  $\sigma_{t,2}$ .



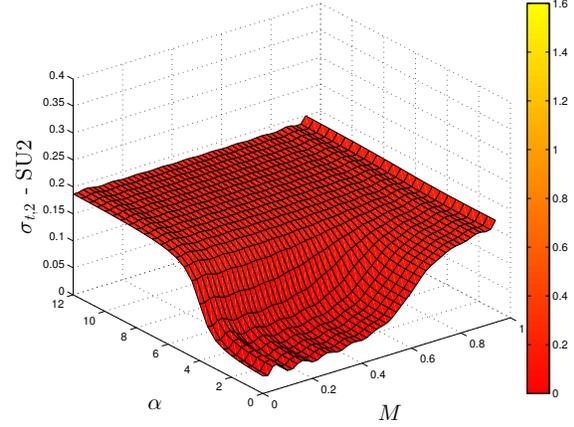
(a) Intermediate surrogate mean  $\mu_1$  (XFOIL)



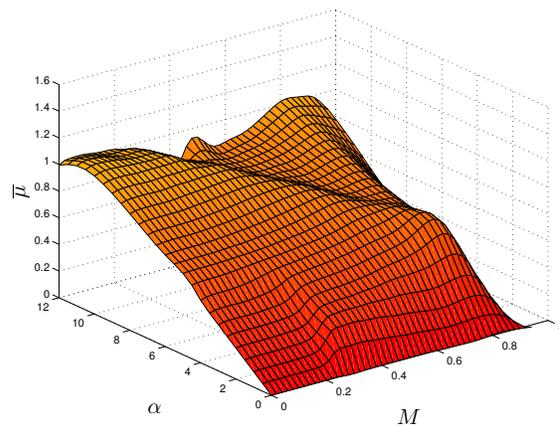
(b) Total standard deviation of XFOIL surrogate,  $\sigma_{t,1}$



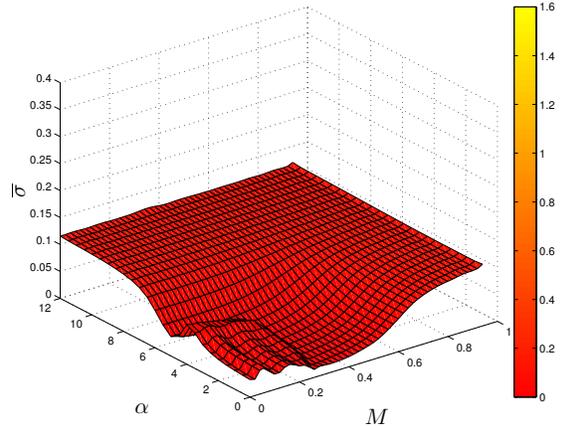
(c) Intermediate surrogate mean  $\mu_2$  (SU2)



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



(e) Multifidelity surrogate mean  $\bar{\mu}$



(f)  $\bar{\sigma}$  for the multifidelity surrogate

Figure 5: Fusing intermediate surrogates for XFOIL and SU2 into a single multifidelity surrogate.

$$\sigma^2(\mathbf{x}) = \left( \sum_{m=1}^M \frac{1}{\sigma_{f,m}^2(\mathbf{x})} \right)^{-1} \quad (8)$$

$$\mu(\mathbf{x}) = \sigma^2(\mathbf{x}) \sum_{m=1}^M \frac{f_m(\mathbf{x})}{\sigma_{f,m}^2(\mathbf{x})}. \quad (9)$$

Note that the mean  $\mu(\mathbf{x})$  of the estimation of the real world process is a weighted sum of the  $M$  IS values  $f_m(\mathbf{x})$ : the higher the fidelity, the higher the weight. We now formulate the multiple IS optimization problem as:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} \mu(\mathbf{x}). \quad (10)$$

The estimation of the real world process and the multifidelity surrogate have similar formulas for their variance and mean. The multifidelity surrogate replaces the fidelity variance  $\sigma_{f,m}^2$  with the total variance  $\sigma_{t,m}^2$ , and replaces the IS value  $f_m$  with the intermediate surrogate mean  $\mu_m$  (Table 1). In regions that have been well sampled, the intermediate surrogate mean  $\mu_m$  becomes a good approximation of the IS value  $f_m$  and the posterior variance tends to  $\sigma_{GP,m}^2 \approx 0$  (hence  $\sigma_{t,m}^2 \approx \sigma_{f,m}^2$ ). Thus, the mean  $\bar{\mu}(\mathbf{x})$  (respectively the variance  $\bar{\sigma}^2(\mathbf{x})$ ) of the multifidelity surrogate is also expected to be a good approximation of the mean  $\mu(\mathbf{x})$  (respectively the variance  $\sigma^2(\mathbf{x})$ ) of the estimation of the real world process. Using an optimization algorithm that through adaptivity samples the design space near the minimizer of  $\mu$  and updates the corresponding surrogate, we expect the multifidelity surrogate mean  $\bar{\mu}$  to become a good approximation of  $\mu$  near the minimizer. In this way, by adaptively constructing the multifidelity surrogate and performing optimization on its mean  $\bar{\mu}$ , we expect to obtain solutions close to those of the original problem defined by Eq. (10).

	RV	Mean	Variance
Estimation of real world process	$P_{\mathbf{x}}$	$\mu(\mathbf{x}) = \sigma^2(\mathbf{x}) \sum_{m=1}^M \frac{f_m(\mathbf{x})}{\sigma_{f,m}^2(\mathbf{x})}$	$\sigma^2(\mathbf{x}) = \left( \sum_{m=1}^M \frac{1}{\sigma_{f,m}^2(\mathbf{x})} \right)^{-1}$
Multifidelity surrogate	$S_{\mathbf{x}}$	$\bar{\mu}(\mathbf{x}) = \bar{\sigma}^2(\mathbf{x}) \sum_{m=1}^M \frac{\mu_m(\mathbf{x})}{\sigma_{t,m}^2(\mathbf{x})}$	$\bar{\sigma}^2(\mathbf{x}) = \left( \sum_{m=1}^M \frac{1}{\sigma_{t,m}^2(\mathbf{x})} \right)^{-1}$

Table 1: Comparison of means and variances of the random variables representing the best estimation of the real world process and the multifidelity surrogate.

### III.B. Choosing the Next Design to Evaluate

Starting with an initial training set for each of the  $M$  IS, we construct an initial multifidelity surrogate (the designs in the training sets can be different for each IS). At each iteration  $n$  of the optimization algorithm, a new design  $\mathbf{x}_n$  is evaluated with IS number  $m_n \in \{1, \dots, M\}$ . The new training point ( $\mathbf{x}_n, y_n = f_{m_n}(\mathbf{x}_n)$ ) is added to the training set  $\mathcal{D}_{m_n}$ . The  $m_n^{\text{th}}$  intermediate surrogate and the multifidelity surrogate are then updated. The design to evaluate at iteration  $n + 1$  is chosen based on this updated multifidelity surrogate. This decision balances two different objectives. We are interested in (1) learning the function  $\bar{\mu}$  that approximates  $\mu$  and (2) minimizing  $\mu$  by minimizing  $\bar{\mu}$ . The first objective corresponds to the exploration of the design space while the second objective corresponds to the exploitation of the multifidelity surrogate learned so far.

#### *Expected Improvement with Single Information Source*

In the case of a single IS  $f_1$ , algorithms such as Efficient Global Optimization<sup>3</sup>(EGO) evaluate the design that maximizes the expected improvement (EI). Given a training set  $\mathcal{D}_1$  composed of  $L$  training points,

the Gaussian process posterior distribution is computed. This defines, for any  $\mathbf{x}$  in  $\mathcal{X}$ , a random variable  $Y_{\mathbf{x}} \sim \mathcal{N}(\mu_1(\mathbf{x}), \sigma_{GP,1}^2(\mathbf{x}))$ . The EI at design  $\mathbf{x}$  is then defined by:

$$EI_Y(\mathbf{x}) = \mathbb{E}[\max(y_{min} - Y_{\mathbf{x}}, 0)] \quad (11)$$

$$y_{min} = \min_{l \in \{1, \dots, L\}} f_1(\mathbf{x}_{l,1}). \quad (12)$$

The EI is known in closed form for Gaussian random variables:

$$EI_Y(\mathbf{x}) = (y_{min} - \mu_1(\mathbf{x}))\Phi\left(\frac{y_{min} - \mu_1(\mathbf{x})}{\sigma_{GP,1}(\mathbf{x})}\right) + \sigma_{GP,1}(\mathbf{x})\phi\left(\frac{y_{min} - \mu_1(\mathbf{x})}{\sigma_{GP,1}(\mathbf{x})}\right), \quad (13)$$

with  $\Phi$  the cumulative density function and  $\phi$  the probability density function of the standard Gaussian random variable.

Designs with large EI are likely to improve the objective function because (1) the posterior variance is large (exploration) or because (2) the posterior mean is close to the minimum performance  $y_{min}$  evaluated at iteration  $n$  (exploitation). This notion is also illustrated in Fig. 6.

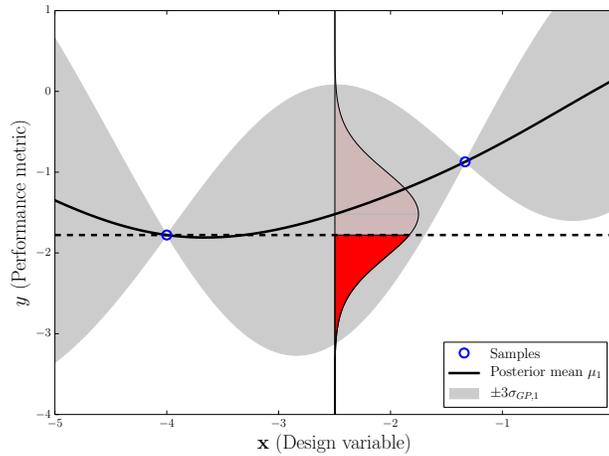


Figure 6: The blue circles are samples used to compute the posterior distribution of a Gaussian process. The Gaussian process posterior mean  $\mu_1$  is shown by the black solid line, and three times the posterior standard deviation  $\sigma_{GP,1}$  is shown by the grey shading. The minimum performance  $y_{min}$  is the left blue circle, also shown by the black dashed line. At a given design  $\mathbf{x}$ , the distribution of  $Y_{\mathbf{x}}$  is shown by the light red shading. The solid red shading is the probability  $\mathbb{P}(Y_{\mathbf{x}} \leq y_{min})$ .

#### Extension of Expected Improvement to Multiple Information Sources

For the EGO algorithm to work, it is necessary that  $EI_Y(\mathbf{x})$  reduces to a value close to zero once  $\mathbf{x}$  is evaluated to avoid evaluating the same design at a future iteration. Thus, in our case it is not possible to use the EI of  $S_{\mathbf{x}}$  to pick  $\mathbf{x}_{n+1}$ , because the variance of  $S_{\mathbf{x}}$  is lower bounded by  $\sigma^2(\mathbf{x})$ , leading to a strictly positive lower bound for the EI. This motivates the definition of a new random variable  $Q_{\mathbf{x}} \sim \mathcal{N}(\bar{\mu}(\mathbf{x}), \hat{\sigma}^2(\mathbf{x}))$ , where  $\bar{\mu}$  is the mean of the multifidelity surrogate and

$$\hat{\sigma}^2(\mathbf{x}) = \left( \sum_{m=1}^M \frac{1}{\sigma_{GP,m}^2(\mathbf{x})} \right)^{-1}. \quad (14)$$

The random variable  $Q_{\mathbf{x}}$  has expected improvement  $EI_Q(\mathbf{x})$ :

$$EI_Q(\mathbf{x}) = (y_{min} - \bar{\mu}(\mathbf{x}))\Phi\left(\frac{y_{min} - \bar{\mu}(\mathbf{x})}{\hat{\sigma}(\mathbf{x})}\right) + \hat{\sigma}(\mathbf{x})\phi\left(\frac{y_{min} - \bar{\mu}(\mathbf{x})}{\hat{\sigma}(\mathbf{x})}\right), \quad (15)$$

where  $y_{min}$  is the minimum of  $\bar{\mu}$  over the designs already evaluated at iteration  $n$ . Since  $\sigma_{GP,m}^2(\mathbf{x})$  reduces to close to zero if  $\mathbf{x}$  is evaluated with the  $m^{th}$  IS,  $EI_Q(\mathbf{x})$  will also decrease to a small value. This permits

the algorithm to avoid picking designs that have already been evaluated with at least one IS. Thus the next design to evaluate is:

$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} EI_Q(\mathbf{x}). \quad (16)$$

Selecting  $\mathbf{x}_{n+1}$  requires solving the auxiliary optimization problem Eq. (16). The EI is only a function of  $y_{min}$ ,  $\bar{\mu}$  and  $\hat{\sigma}$ , which are cheap to evaluate compared to evaluating the IS. However, EI is a multimodal function of the design  $\mathbf{x}$  (see Fig. 7 bottom right panel) and finding its maximizer can be a challenging task. In particular, after a large number of iterations  $n$  of the multifidelity algorithm, using an optimization algorithm to solve this auxiliary problem often requires finding an isolated peak in an almost flat domain (see Fig. 8 bottom right panel). The design selected by a numerical optimization algorithm used to solve Eq. (16) is likely to be in a flat region of the EI, providing a poor choice for exploration and exploitation. This typically occurs when the EI is low across the design space, indicating that the multifidelity surrogate has low uncertainty in regions that are believed to contain the minimizer. For this reason, when the maximum EI found is lower than a threshold  $\epsilon_{ei}$ , we instead choose  $\mathbf{x}_{n+1}$  to be the design minimizing the multifidelity surrogate:

$$\mathbf{x}_{n+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \bar{\mu}(\mathbf{x}). \quad (17)$$

This corresponds to exploiting the multifidelity surrogate.

### III.C. Choosing the Next Information Source to Query

Once the algorithm chooses a design  $\mathbf{x}_{n+1}$  to evaluate at the next iteration, we use a heuristic to choose  $m_{n+1}$ , i.e., which IS to next query. We seek large information gain for a low evaluation cost and propose to use the following heuristic:

$$m_{n+1} = \operatorname{argmin}_{m \in \{1, \dots, M\}} \frac{C_m(\mathbf{x}_{n+1})}{\bar{\sigma}^2(\mathbf{x}_{n+1}) - \tilde{\sigma}_m^2(\mathbf{x}_{n+1})}, \quad (18)$$

$$\tilde{\sigma}_m^2(\mathbf{x}_{n+1}) = \left( \frac{1}{\sigma_{f,m}^2(\mathbf{x}_{n+1})} + \sum_{\substack{i=1 \\ i \neq m}}^M \frac{1}{\sigma_{t,i}^2(\mathbf{x}_{n+1})} \right)^{-1}, \quad (19)$$

where  $C_m(\mathbf{x}_{n+1})$  is the evaluation cost of the  $m^{th}$  IS at design  $\mathbf{x}_{n+1}$ ,  $\bar{\sigma}^2(\mathbf{x}_{n+1})$  the multifidelity surrogate variance and  $\tilde{\sigma}_m^2(\mathbf{x}_{n+1})$  the multifidelity surrogate variance if the  $m^{th}$  IS is known up to its fidelity at design  $\mathbf{x}_{n+1}$ . The denominator quantifies how much information could be learned by evaluating the  $m^{th}$  IS. Thus, this metric selects the IS with large information gain and low evaluation cost. This is a two-step approach, which first chooses the design and then the IS to query. A better approach might be that in Moore et al.,<sup>13</sup> who choose the design and the IS at the same time by maximizing a value of information metric. In that work, the authors note that the value of information tends to be maximized in the same design space region for each IS. This suggests that the two-step approach of choosing the design and then the IS would lead to similar results, although this should be investigated in future work.

#### *Unconstrained 1D example*

We now illustrate the key features of the algorithm on a one-dimensional unconstrained example. We consider the design space  $\mathcal{X} = [0, 1]$  and  $M = 2$  information sources  $f_1$  and  $f_2$  (shown as red and green dashed lines on Figs. 7–8). The best estimation of the real world process  $\mu$  is shown as the black dashed line. The evaluation cost of IS 1 and IS 2 are set to  $C_1 = 100$  and  $C_2 = 1$  (independent of the evaluated design) and their respective fidelity variances change across the design space ( $\pm 3\sigma_{f,m}$  is shown by the dark grey shading on the left panels of Figs. 7–8). IS 1 is high fidelity near  $\mathbf{x} = 0$  but a low fidelity near  $\mathbf{x} = 1$ . IS 2 is medium fidelity across the design space. Fig. 7 shows the first iteration of the algorithm. The left panels show the intermediate surrogates, their associated total and fidelity variances, and the samples used to construct the surrogates. The top right panel shows the multifidelity surrogate and its variances. The bottom right panel

is the expected improvement for the multifidelity surrogate. It is multimodal and a multistart optimization method is able to find the maximum. Fig. 8 shows the iteration  $n = 10$  of the algorithm. The multifidelity surrogate minimum is close to the minimum of  $\mu$  and is also designated by the location of the maximum EI. At this iteration, finding this maximum is challenging.

The algorithm has evaluated the inexpensive IS 2 extensively, while the expensive IS 1 has only been evaluated in the region close to the minimum or where its fidelity is high. It should be noted that there is a large uncertainty in the intermediate surrogate of IS 1 in  $[0.4, 0.7]$  because IS 1 has not been queried there. However, the intermediate surrogate of IS 2 has low uncertainty and is confident that this region does not contain the minimum, so IS 1 is not queried. This leads to a reduction in the total evaluation cost of the algorithm. This example illustrates how IS with varying fidelity and different costs can be exploited to reduce the number of queries of the expensive IS.

### III.D. Adding Constraints

We now consider the case in which the optimization is subject to inequality constraints. For ease of notation, we consider the case in which there is only one quantity that must satisfy an inequality constraint. Each IS provides a different estimation  $c_m(\mathbf{x})$  of that quantity with a fidelity variance  $\sigma_{f_c,m}^2(\mathbf{x})$ . For any  $\mathbf{x}$  in  $\mathcal{X}$ , we define the estimation of the real world constraint to be the random variable  $P_{\mathbf{x}}^{(c)} \sim \mathcal{N}(\mu_c(\mathbf{x}), \sigma_c^2(\mathbf{x}))$  such that:

$$\sigma_c^2(\mathbf{x}) = \left( \sum_{m=1}^M \frac{1}{\sigma_{f_c,m}^2(\mathbf{x})} \right)^{-1} \quad (20)$$

$$\mu_c(\mathbf{x}) = \sigma_c^2(\mathbf{x}) \sum_{m=1}^M \frac{c_m(\mathbf{x})}{\sigma_{f_c,m}^2(\mathbf{x})}. \quad (21)$$

We can now define the optimization problem as:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} \mu(\mathbf{x}) \quad (22)$$

$$\text{s.t. } \mu_c(\mathbf{x}^*) \leq 0. \quad (23)$$

Since the constraints  $c_m$  are typically expensive to evaluate, we construct a multifidelity surrogate for the estimation of the real world constraint  $\mu_c$ . We define an intermediate surrogate for each IS, and characterize it with a random variable  $S_{\mathbf{x},m}^{(c)} \sim \mathcal{N}(\mu_{c,m}(\mathbf{x}), \sigma_{t_c,m}^2(\mathbf{x}))$ . Here,  $\mu_{c,m}$  is the posterior mean of the Gaussian process associated with  $c_m$ , and  $\sigma_{t_c,m}^2$  is the sum of the posterior variance of the Gaussian process  $\sigma_{G_{P_c,m}}^2$  and the fidelity variance  $\sigma_{f_c,m}^2$ . The multifidelity surrogate  $S_{\mathbf{x}}^{(c)} \sim \mathcal{N}(\bar{\mu}_c(\mathbf{x}), \bar{\sigma}_c^2(\mathbf{x}))$  of the constraint is the fusion of the  $M$  random variables  $S_{\mathbf{x},m}^{(c)}$ , where  $\bar{\mu}_c$  and  $\bar{\sigma}_c^2$  are defined in Table 2. The multifidelity surrogate for the constraint is also constructed adaptively.

	RV	Mean	Variance
Best estimation of real world process	$P_{\mathbf{x}}^{(c)}$	$\mu_c(\mathbf{x}) = \sigma_c^2(\mathbf{x}) \sum_{m=1}^M \frac{c_m(\mathbf{x})}{\sigma_{f_c,m}^2(\mathbf{x})}$	$\sigma_c^2(\mathbf{x}) = \left( \sum_{m=1}^M \frac{1}{\sigma_{f_c,m}^2(\mathbf{x})} \right)^{-1}$
Multifidelity surrogate	$S_{\mathbf{x}}^{(c)}$	$\bar{\mu}_c(\mathbf{x}) = \bar{\sigma}_c^2(\mathbf{x}) \sum_{m=1}^M \frac{\mu_{c,m}(\mathbf{x})}{\sigma_{t_c,m}^2(\mathbf{x})}$	$\bar{\sigma}_c^2(\mathbf{x}) = \left( \sum_{m=1}^M \frac{1}{\sigma_{t_c,m}^2(\mathbf{x})} \right)^{-1}$

Table 2: Comparison of means and variances of the random variables representing the best estimation of the real world process and the multifidelity surrogate for the constraints. The intermediate surrogate of the  $m^{\text{th}}$  IS has mean  $\mu_{c,m}$  and total variance  $\sigma_{t_c,m}^2$ .

Following Sec. III.B, the next design  $\mathbf{x}_{n+1}$  is picked by maximizing the EI. This optimization is now subject to constraints. Since there is uncertainty in the constraint surrogate, we use a loose constraint

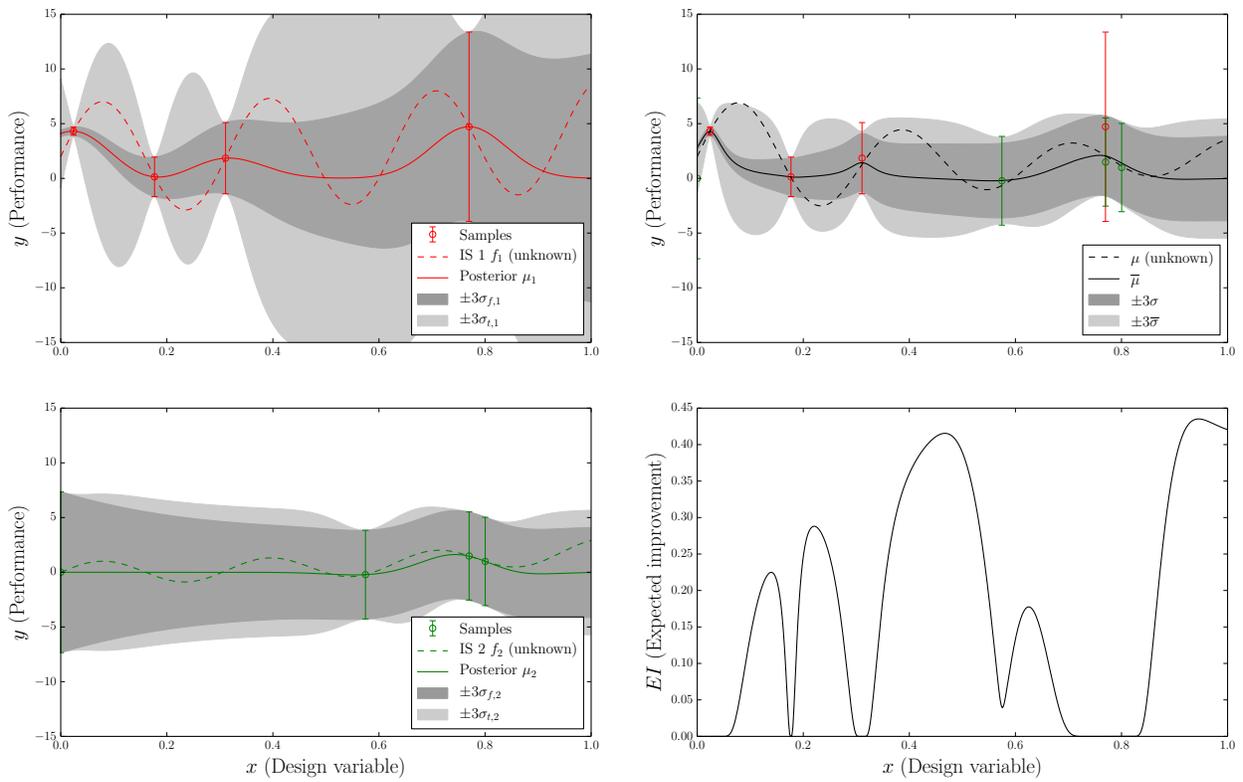


Figure 7: 1D unconstrained example: Iteration 1. Top left panel: intermediate surrogate of IS 1. Bottom left panel: intermediate surrogate of IS 2. Top right panel: multifidelity surrogate. Bottom right panel: Expected improvement of the multifidelity surrogate.

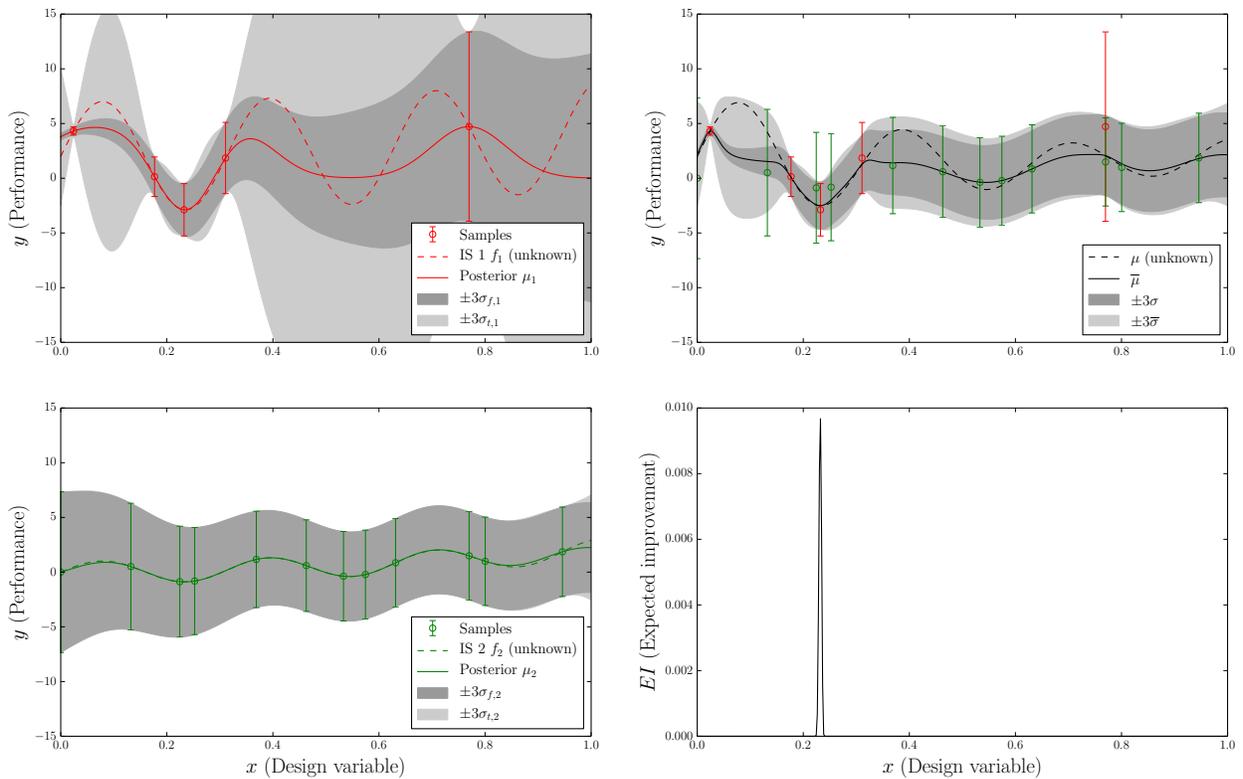


Figure 8: 1D unconstrained example: Iteration 10. Top left panel: intermediate surrogate of IS 1. Bottom left panel: intermediate surrogate of IS 2. Top right panel: multifidelity surrogate. Bottom right panel: Expected improvement of the multifidelity surrogate.

$\bar{\mu}_c - 3\bar{\sigma}_c \leq 0$  instead of the strict constraint  $\bar{\mu}_c \leq 0$ ; the search space defined by the loose constraint is a superset of the one defined by the strict constraint. This avoids over-restricting the search space when exploration is needed and the constraint surrogate is uncertain. As the constraint surrogate is updated,  $\bar{\sigma}_c$  reduces and the loose constraint becomes close to the strict constraint in regions that have been sampled (see Fig. 10). The next design is found by solving:

$$\mathbf{x}_{n+1} = \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} EI_Q(\mathbf{x}) \quad (24)$$

$$\text{s.t. } \bar{\mu}_c(\mathbf{x}_{n+1}) - 3\bar{\sigma}_c(\mathbf{x}_{n+1}) \leq 0 \quad (25)$$

where  $EI_Q(\mathbf{x})$  is defined by Eq. (15) with  $y_{min}$  the minimum of  $\bar{\mu}$  over the designs already evaluated at iteration  $n$  and satisfying the loose constraint. When the maximum EI found for this auxiliary problem is smaller than the threshold  $\epsilon_{ei}$ , the next design  $\mathbf{x}_{n+1}$  is chosen by minimizing the multifidelity surrogate  $\bar{\mu}$ . Since this corresponds to exploiting the multifidelity surrogate, we use a strict constraint in this case:

$$\mathbf{x}_{n+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \bar{\mu}(\mathbf{x}) \quad (26)$$

$$\text{s.t. } \bar{\mu}_c(\mathbf{x}_{n+1}) \leq 0. \quad (27)$$

The choice of the IS to query is chosen in the same way described in Sec. III.C.

## IV. Results

This section demonstrates the multifidelity optimization algorithm developed in Section II and Section III on two different cases. The first case is a modified 2D Rosenbrock problem with a nonlinear constraint. The second case is the minimization of the drag coefficient of a NACA 0012 airfoil subject to a constraint on the lift coefficient.

### IV.A. Rosenbrock with Constraints

#### IV.A.1. Problem Formulation

We demonstrate the key features of the algorithm on a two-dimensional constrained example. We consider the design space  $\mathcal{X} = [-2, 2] \times [-2, 2]$  and  $M = 2$  IS  $f_1$  and  $f_2$ . IS 1 is the Rosenbrock function, while IS 2 is the Rosenbrock function with an added oscillatory component:

$$f_1(\mathbf{x}) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2 \quad (28)$$

$$f_2(\mathbf{x}) = f_1(\mathbf{x}) + 0.1 \sin(10x_1 + 5x_2). \quad (29)$$

The two IS are similar except in the boomerang shaped valley, where the oscillatory term becomes more significant as  $f_1$  is close to zero. We set the evaluation costs to be  $C_1 = 1000$  and  $C_2 = 1$ . The fidelity variances are set to  $\sigma_{f,1}^2 = 10^{-3}$  and  $\sigma_{f,2}^2 = 10^{-2}$ .

The constraints are:

$$c_1(\mathbf{x}) = -x_1^2 - \frac{(x_2 - 1)^2}{2}, \quad (30)$$

$$c_2(\mathbf{x}) = c_1(\mathbf{x}) + 0.1 \sin(10x_1 + 5x_2). \quad (31)$$

The fidelity variances for the constraints are set to  $\sigma_{f_c,1}^2 = 10^{-3}$  and  $\sigma_{f_c,2}^2 = 10^{-2}$ .

The two IS being analytical functions, we can afford to compute the minimizer  $\mathbf{x}^*$  of  $\mu$  subject to the constraint  $\mu_c$  (i.e., the original problem defined by Eq. (23)) and use it to assess the performance of the multifidelity algorithm. The multifidelity surrogate is initialized with 5 training points for IS 1 and 5 training points for IS 2, chosen at random.

#### IV.A.2. Results

The results of the multifidelity optimization are shown in Fig. 10. Most of the evaluations are computed in the valley with the cheap IS 2 (Fig. 9). The expensive IS 1 is only evaluated in the valley, close to the

minimizer  $\mathbf{x}^*$  of  $\mu$  where an IS with high fidelity is required. The minimizer of the multifidelity surrogate at the final iteration  $\bar{\mathbf{x}}^*$  is shown by the yellow star and is close to  $\mathbf{x}^* = [0.58, 0.34]$  (not plotted). The feasible space is approximated well in the regions densely sampled and in particular near the multifidelity surrogate minimizer. The feasible space based on the loose constraint  $\bar{\mu}_c - 3\bar{\sigma}_c$ , shown as the light blue shading, collapses to the feasible space based on the strict constraint  $\bar{\mu}_c$  near the minimizer, because  $\bar{\sigma}_c$  is close to  $\sigma_c = 0.03$  (i.e., the contribution to the variance from the GP model of the constraint is close to 0). Recall from Section III.D that once the maximum EI found is lower than the threshold  $\epsilon_{ei} = 10^{-5}$ , the next design is chosen to satisfy the strict constraint by solving the surrogate-based optimization problem defined in Eqs. (26) and (27). Hence, the final solution  $\bar{\mathbf{x}}^*$  also satisfies the strict constraint.

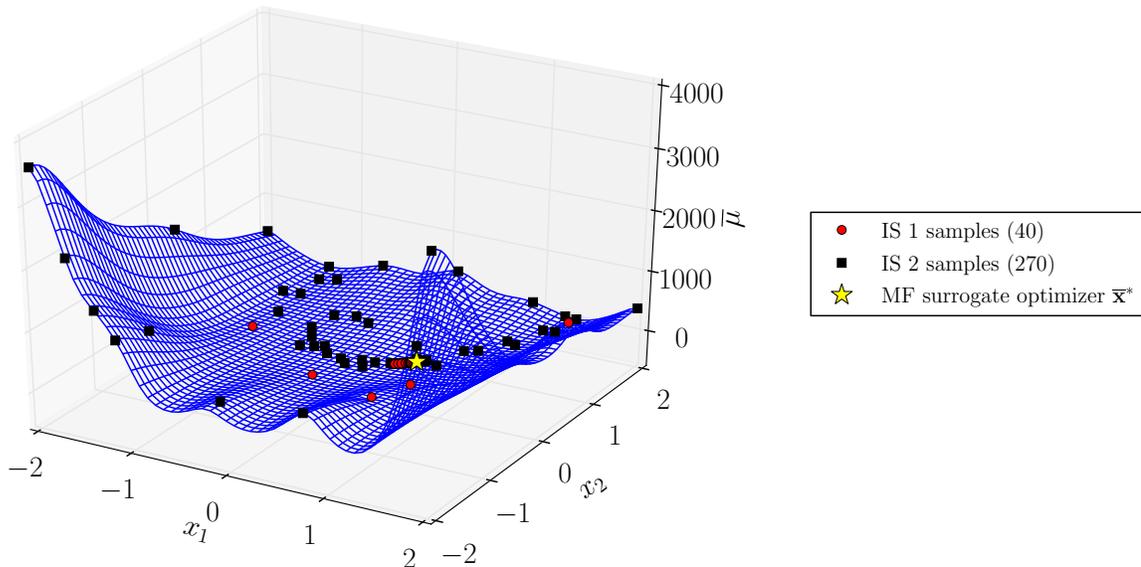


Figure 9: 2D constrained Rosenbrock example. 3D visualization of the multifidelity surrogate.

Fig. 11 (top left panel) shows the distance between the minimizer  $\mathbf{x}^*$  of the estimation of the real world process  $\mu$  and the minimizer  $\mathbf{x}_n^*$  at iteration  $n$  of the multifidelity surrogate mean  $\bar{\mu}$  as a function of the iteration  $n$ . The expensive IS 1 is queried for the first time after more than 60 iterations. Most of the evaluations are made with the cheap IS 2. The distance  $\|\mathbf{x}_k^* - \mathbf{x}^*\|$ , where  $\mathbf{x}_k^*$  is the minimizer of  $\bar{\mu}$  subject to the strict constraint at the  $k^{\text{th}}$  query of the expensive IS 1, drops to approximately  $10^{-3}$  in about 25 evaluations of IS 1 (Fig. 11 top right panel). The multifidelity surrogate mean value  $\bar{\mu}(\mathbf{x}_k^*)$  is a good approximation of  $\mu(\mathbf{x}_k^*)$ , which lies in the  $\pm 3\bar{\sigma}(\mathbf{x}_k^*)$  bars (Fig. 11 bottom right panel). The bias between  $\bar{\mu}(\mathbf{x}_k^*)$  and  $\mu(\mathbf{x}_k^*)$  seen in Fig. 11 bottom right panel is due to training points far from  $\mathbf{x}_k^*$  that influence the multifidelity surrogate mean through a large characteristic length scale. This suggests that, once the minimizer is trusted to be in a small region, a new multifidelity surrogate should be constructed using only the subset of training points included in the neighborhood of the minimizer. The optimization would be restricted to that small region and the multifidelity surrogate would be updated as new training points are computed. This could potentially be achieved through use of a trust region framework.

#### IV.A.3. Comparison with the Efficient Global Optimization Algorithm

The proposed multifidelity optimization algorithm is compared to the EGO algorithm<sup>3</sup> on the problem defined by Eqs. (22) and (23). We allow EGO to query  $\mu$  (which involves evaluating  $f_1$  and  $f_2$ ) and  $\mu_c$  (which involves evaluating  $c_1$  and  $c_2$ ) at each iteration. We also initialize EGO with a larger training set than the multifidelity algorithm, composed of the designs of both the initial training sets of IS 1 and IS 2 evaluated at  $\mu$ . Given those two advantages: access to evaluation of  $\mu$  (instead of  $\bar{\mu}$ ) and a larger initial training set, the EGO algorithm performance is assessed. Fig. 11 (bottom left panel) shows the distance  $\|\mathbf{x}^* - \mathbf{x}_k^*\|$  as a function of the number of IS 1 evaluations  $k$ , where  $\mathbf{x}_k^*$  is the minimizer of the EGO surrogate subject to the strict constraint  $\bar{\mu}_c \leq 0$  after  $k$  evaluations of the expensive IS 1 (necessary to evaluate  $\mu$ ).

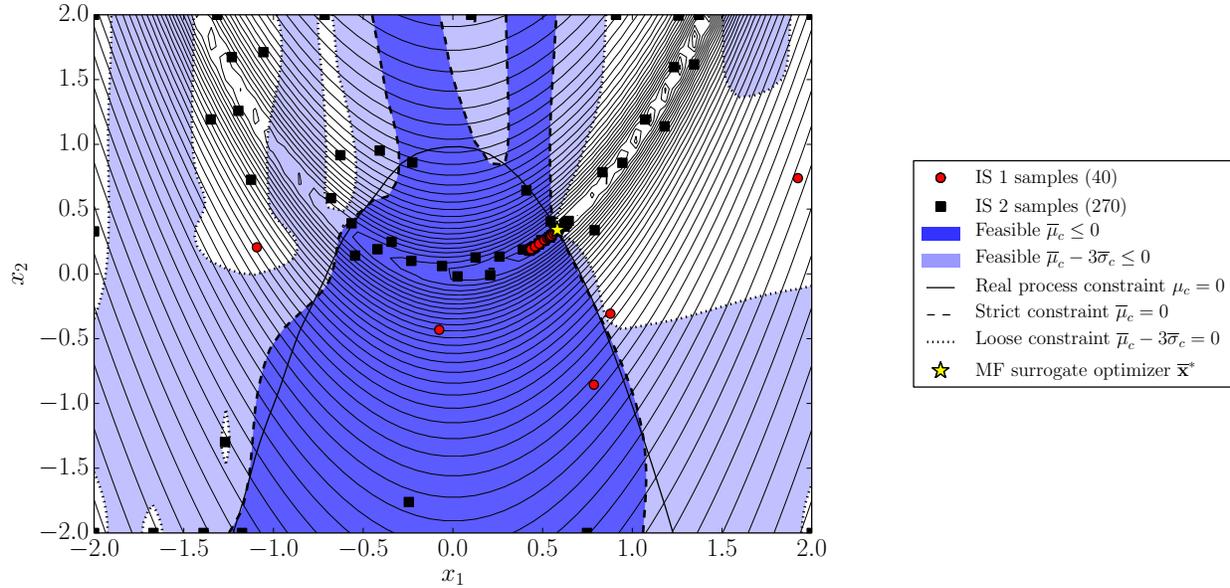


Figure 10: 2D constrained Rosenbrock example. 2D visualization of the loose and strict feasible space and evaluated design locations.

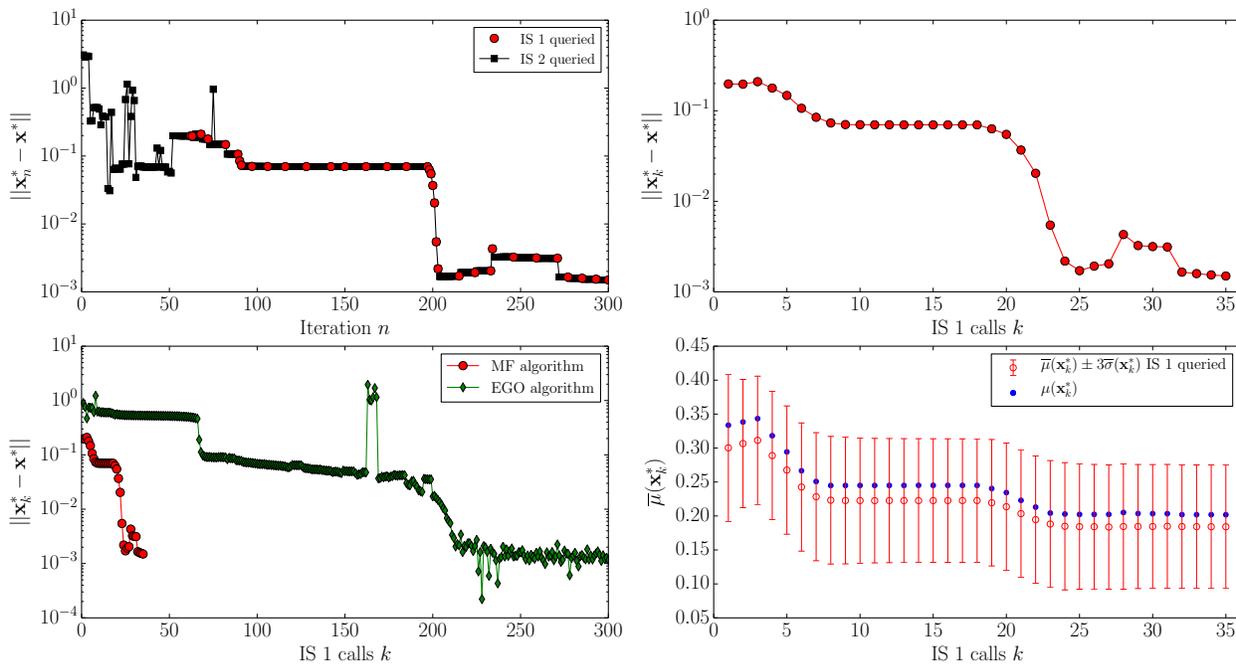


Figure 11: Convergence plots of the multifidelity optimization algorithm for the 2D constrained Rosenbrock problem.

The multifidelity algorithm and EGO both converge to the solution within a tolerance of  $10^{-3}$ . In this particular example, the multifidelity surrogate only requires 25 expensive evaluations of IS 1 to reach that tolerance, compared with 210 evaluations of IS 1 for EGO. This convergence acceleration is the consequence of leveraging the information provided at cheap cost by IS 2. A more complete study of the impact of the fidelity variances on the convergence is necessary before drawing general conclusions about the performances of the multifidelity algorithm compared to EGO.

## IV.B. NACA 0012 Drag Coefficient with Lift Coefficient Constraints

### IV.B.1. Problem Formulation

We apply the proposed multifidelity algorithm to a two-dimensional constrained aerodynamic design example. We seek the Mach number  $M$  and angle of attack  $\alpha$  that minimize the drag coefficient  $C_D$  of a NACA 0012 airfoil subject to a lift coefficient  $C_L$  larger than 0.4. We consider the design space  $\mathcal{X} = I_M \times I_\alpha$  with  $I_M = [0.1, 0.5]$  and  $I_\alpha = [0.01^\circ, 8^\circ]$ . Two information sources are available: SU2 with viscous terms (IS 1) and XFOIL (IS 2) can each compute both  $C_D$  and  $C_L$ . The evaluation of SU2 requires about one hour of computation while XFOIL can be evaluated in less than a second. Thus we set the evaluation costs to be  $C_1 = 3600$  and  $C_2 = 1$ . The fidelity variances are set to  $\sigma_{f,1}^2 = 1$  and  $\sigma_{f,2}^2 = 10$ . The fidelity variances for the constraints are also set to  $\sigma_{f_c,1}^2 = 1$  and  $\sigma_{f_c,2}^2 = 10$ . The multifidelity surrogate is initialized with 5 designs chosen at random for each IS. The number of iterations of the algorithm is limited to 300 with a maximum of 10 evaluations of SU2. This corresponds to a scenario where the expensive IS is given a fixed computational budget.

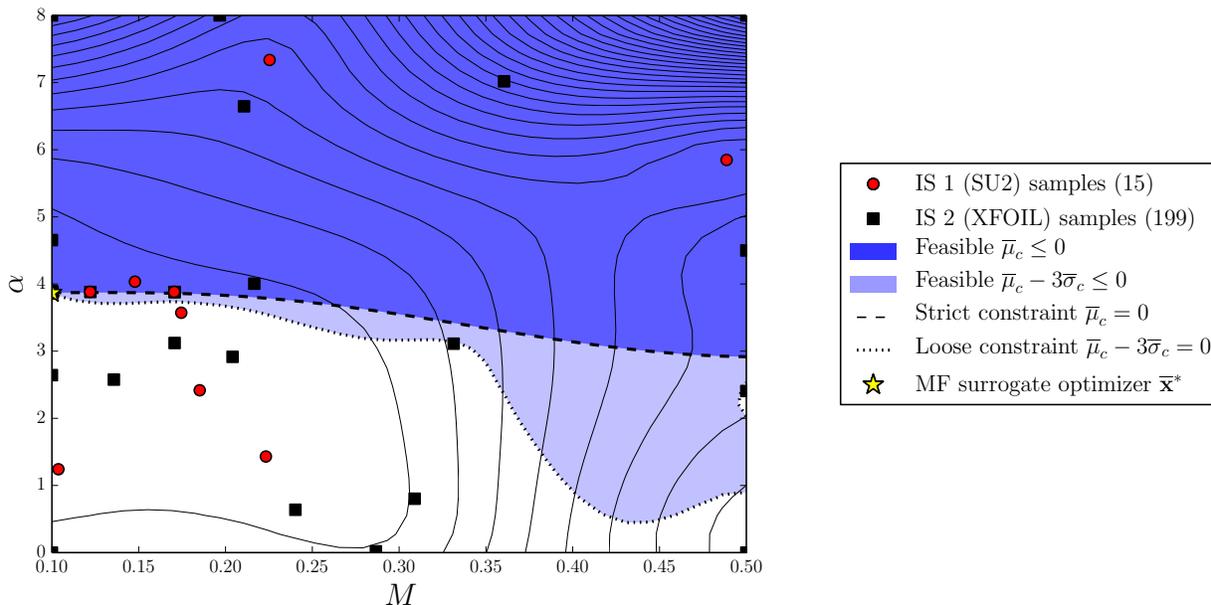
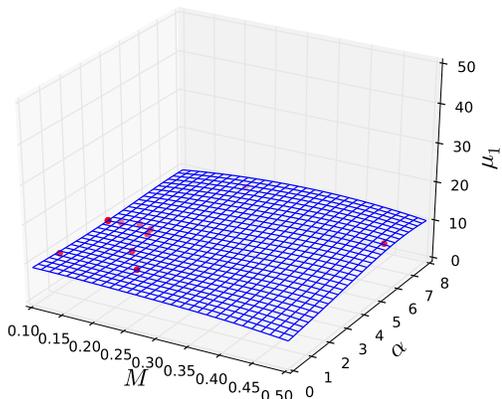


Figure 12: Visualization of the feasible domain and the location of the evaluated designs.

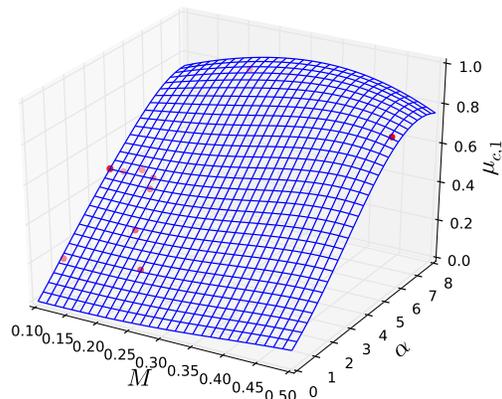
### IV.B.2. Results

The first iterations are queried in the infeasible space, and as the constraint is learned, the search space is reduced: the standard deviation  $\bar{\sigma}_c$  is close to zero where  $M < 0.2$  and  $\alpha \approx 0.4$  (see Fig. 12). The constraint is not learned (the dashed and dotted lines are not close in space Fig. 12) where the domain is not sampled (i.e.  $M > 0.3$ ), this corresponds to regions unlikely to contain the minimum. Finally the evaluations are restricted to the feasible domain, near  $M = 0.1$  and  $\alpha \approx 4^\circ$ , where a minimum is likely to be present. The optimization stopped after reaching the maximum number of evaluations of the expensive IS 1 (SU2). It completed 10 queries to IS 1 (SU2) and 194 queries to IS 2 (XFOIL). The intermediate surrogates at the final iteration (Figs. 13a,13c) for the drag coefficient are computed. SU2 predicts a lower drag coefficient than XFOIL. The multifidelity surrogate (Fig. 13e) gives more weight to the SU2 intermediate surrogate based on the fidelity variances. The intermediate surrogates (Fig. 13b,13d) for the lift coefficient are computed and the two information sources provide similar predictions. The border between the feasible and infeasible domains as predicted by the mean  $\bar{\mu}_c(\mathbf{x})$  of the multifidelity surrogate  $S_x^{(c)}$  is shown by the solid black line in Fig. 13f. The solution of the problem (see Fig. 12) is the lowest Mach number allowed by box constraints and the lowest angle of attack satisfying the lift coefficient requirement:  $\bar{\mathbf{x}}^* = (0.1, 3.870^\circ)$ . This result is obtained with 15 evaluations of the CFD code SU2 (5 initial points and 10 calls during the optimization). In the multifidelity optimization, the cheap-to-evaluate IS 2 (XFOIL) is used to explore the design space,

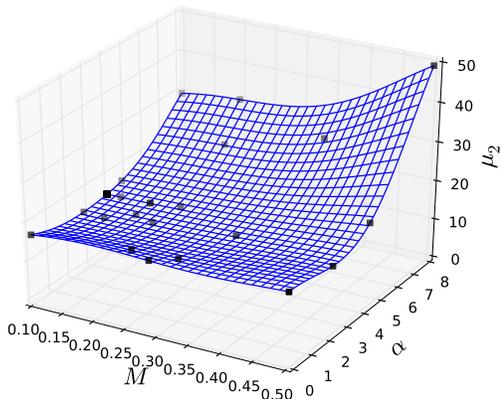
while the expensive IS 1 (SU2) is only used to refine the results. Significant parts of the design space are not evaluated with SU2, because they are unlikely to contain the minimum. Thus, the multifidelity optimization algorithm finds a good approximation  $\bar{\mathbf{x}}^*$  of  $\mathbf{x}^*$  with a reduced evaluation cost.



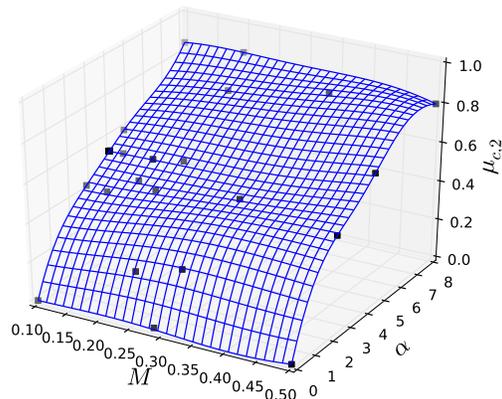
(a)  $\mu_1 (10^3 \times C_D)$  computed with SU2



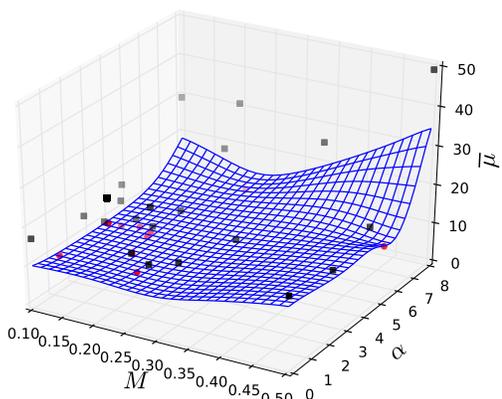
(b)  $\mu_{c,1} (C_L)$  computed with SU2



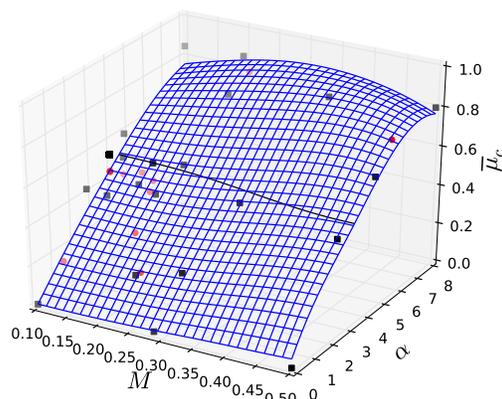
(c)  $\mu_2 (10^3 \times C_D)$  computed with XFOIL



(d)  $\mu_{c,2} (C_L)$  computed with XFOIL



(e) Multifidelity surrogate mean  $\bar{\mu} (10^3 \times C_D)$



(f) Multifidelity surrogate mean  $\bar{\mu}_c (C_L)$

Figure 13: Intermediate and multifidelity surrogate mean for the objective function  $C_D$  and the constraints  $C_L$  using SU2 and XFOIL.

## V. Conclusions

This work has proposed a method to perform optimization when several information sources, all mapping from the same design space to the same quantity of interest, are available. This is achieved by constructing a multifidelity surrogate that can leverage past information, and incorporate evaluations of different IS into a single, cheap-to-evaluate surrogate. This multifidelity surrogate is built adaptively without using the common assumption of hierarchical IS. Instead, each IS is endowed with a fidelity that can vary across the design space. We leverage this multifidelity surrogate in an optimization algorithm. The proposed strategy samples the design space balancing two competing objectives: exploitation of the surrogate, and exploration to improve the surrogate. The proposed algorithm does not require evaluating each IS at every iteration, yielding evaluation cost reduction. The IS to evaluate is chosen based on a heuristic that balances potential information gain and evaluation cost. The evaluation cost reduction of the proposed algorithm has been demonstrated on an analytical problem and an aerodynamic problem.

## Acknowledgments

This work was supported in part by the AFOSR MURI on Uncertainty Quantification, grant FA9550-09-0613, program manager F. Fahroo, and by the Singapore-MIT Alliance Computational Engineering Programme.

## References

- <sup>1</sup>Keane, A. J., “Wing optimization using design of experiment, response surface, and data fusion methods,” *Journal of Aircraft*, Vol. 40, No. 4, 2003, pp. 741–750.
- <sup>2</sup>Choi, S., Alonso, J. J., Kroo, I. M., and Wintzer, M., “Multifidelity design optimization of low-boom supersonic jets,” *Journal of Aircraft*, Vol. 45, No. 1, 2008, pp. 106–118.
- <sup>3</sup>Jones, D. R., Schonlau, M., and Welch, W. J., “Efficient global optimization of expensive black-box functions,” *Journal of Global Optimization*, Vol. 13, No. 4, 1998, pp. 455–492.
- <sup>4</sup>Rajnarayan, D., *Trading risk and performance for engineering design optimization using multifidelity analyses*, Ph.D. thesis, Stanford University, CA, June 2009.
- <sup>5</sup>Rajnarayan, D., Haas, A., and Kroo, I., “A multifidelity gradient-free optimization method and application to aerodynamic design,” *12th AIAA/ISSMO multidisciplinary analysis and optimization conference, Victoria, British Columbia, AIAA*, 2008.
- <sup>6</sup>Kennedy, M. C. and O’Hagan, A., “Predicting the output from a complex computer code when fast approximations are available,” *Biometrika*, Vol. 87, No. 1, 2000, pp. 1–13.
- <sup>7</sup>Huang, D., Allen, T. T., Notz, W. I., and Miller, R. A., “Sequential Kriging optimization using multiple-fidelity evaluations,” *Structural and Multidisciplinary Optimization*, Vol. 32, No. 5, 2006, pp. 369–382.
- <sup>8</sup>Xiong, Y., Chen, W., and Tsui, K., “A new variable-fidelity optimization framework based on model fusion and objective-oriented sequential sampling,” *Journal of Mechanical Design*, Vol. 130, No. 11, 2008, pp. 111401.
- <sup>9</sup>Chung, H. and Alonso, J. J., “Design of a low-boom supersonic business jet using cokriging approximation models,” *AIAA Paper*, Vol. 5598, 2002, pp. 2002.
- <sup>10</sup>Chung, H. and Alonso, J. J., “Using gradients to construct cokriging approximation models for high-dimensional design optimization problems,” *AIAA Paper*, Vol. 317, 2002, pp. 14–17.
- <sup>11</sup>Forrester, A. I. J., Sóbester, A., and Keane, A. J., “Multi-fidelity optimization via surrogate modelling,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, Vol. 463, No. 2088, 2007, pp. 3251–3269.
- <sup>12</sup>Romero, D. A., Amon, C. H., and Finger, S., “Multiresponse Metamodeling in Simulation-Based Design Applications,” *Journal of Mechanical Design*, Vol. 134, No. 9, 2012, pp. 091001.
- <sup>13</sup>Moore, R. A., Romero, D. A., and Paredis, C. J. J., “A rational design approach to gaussian process modeling for variable fidelity models,” *ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, 2011, pp. 727–740.
- <sup>14</sup>Ng, L. W. T. and Willcox, K. E., “Multifidelity approaches for optimization under uncertainty,” *International Journal for Numerical Methods in Engineering*, Vol. 100, No. 10, 2014, pp. 746–772.
- <sup>15</sup>Booker, A. J., Dennis Jr., J. E., Frank, P. D., Serafini, D. B., Torczon, V., and Trosset, M. W., “A rigorous framework for optimization of expensive functions by surrogates,” *Structural optimization*, Vol. 17, No. 1, 1999, pp. 1–13.
- <sup>16</sup>Alexandrov, N. M., Dennis Jr., J. E., Lewis, R. M., and Torczon, V., “A trust-region framework for managing the use of approximation models in optimization,” *Structural Optimization*, Vol. 15, No. 1, 1998, pp. 16–23.
- <sup>17</sup>Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., and Newman, P. A., “Approximation and model management in aerodynamic optimization with variable-fidelity models,” *Journal of Aircraft*, Vol. 38, No. 6, 2001, pp. 1093–1101.
- <sup>18</sup>March, A. and Willcox, K. E., “Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives,” *AIAA Journal*, Vol. 50, No. 5, 2012, pp. 1079–1089.
- <sup>19</sup>Rasmussen, C. E. and Williams, C. K. I., *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.

<sup>20</sup>Kennedy, M. C. and O'Hagan, A., "Bayesian Calibration of Computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 3, 2001, pp. 425–464.

<sup>21</sup>Winkler, R. L., "Combining probability distributions from dependent information sources," *Management Science*, Vol. 27, No. 4, 1981, pp. 479–488.

<sup>22</sup>Drela, M., "XFOIL: An Analysis and Design System for Low Reynolds Number Airfoils," *Low Reynolds Number Aerodynamics*, Springer, Dordrecht, Germany, 1989.

<sup>23</sup>Palacios, F., Alonso, J. J., Duraisamy, K., Colonno, M., Hicken, J., Aranake, A., Campos, A., Copeland, S., Economon, T. D., Lonkar, A., et al., "Stanford University Unstructured (SU2): An open-source Integrated Computational Environment for Multi-physics Simulation and Design," *51st AIAA Aerospace Sciences Meeting and Exhibit*, 2013.

<sup>24</sup>Drela, M. and Giles, M. B., "Viscous-inviscid Analysis of Transonic and Low Reynolds Number Airfoils," *AIAA Journal*, Vol. 25, No. 10, 1987, pp. 1347–1355.

<sup>25</sup>Drela, M., *Flight Vehicle Aerodynamics*, MIT Press, Cambridge, MA, 2014.