Doctoral Dissertation
Doctoral Program in Aerospace Engineering (36th cycle)

Advanced Optimization Methods for Design and Diagnostic Problems in Aerospace

By

Francesco Di Fiore
******

Supervisor(s):
Prof. Dr. Ing. Laura Mainini, Supervisor
Prof. Paolo Maggiore, Supervisor
Dr. Ing. Matteo D.L. Dalla Vedova, Co-Supervisor

Doctoral Examination Committee:
Prof. Ali Elham, University of Southampton
Prof. Vassili Toropov, Queen Mary University of London
Dr. Nathalie Bartoli, ONERA
Dr. Ana Paula Cuco, EMBRAER
Dr. Vladmir Balabanov, University of Colorado

Politecnico di Torino
2024

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Francesco Di Fiore
2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Acknowledgements

Completing this PhD thesis has been an incredible journey, one that I could not have undertaken without the support and encouragement of numerous individuals who have been fundamental in shaping this endeavor.

First and foremost, I extend my deepest gratitude to my supervisor, Prof. Dr. Ing. Laura Mainini. Your guidance, wisdom, and unwavering support have been the cornerstone of this research. Your dedication to my academic and personal growth, your insightful feedback, and your encouragement have inspired me to reach new heights. I am truly fortunate to have had the opportunity to work under your mentorship.

I am also grateful to my supervisors Prof. Paolo Maggiore and Dr. Ing. Matteo D. L. Dalla Vedova for their valuable insights and constructive support, which have enriched this work and helped me to refine my ideas. I deeply appreciate all you have done to foster my growth as a researcher and as a person.

To my friends, whose camaraderie, laughter, and unwavering support have brought joy and companionship to my journey. Thank you for being a constant source of encouragement, for your willingness to lend an ear, and for sharing in both the triumphs and challenges of this process. Your friendship has enriched my life immeasurably.

To my family, who has been my pillars of strength and support throughout this journey. Your love, understanding, and encouragement have sustained me through the highs and lows of academia. I am endlessly grateful for your unwavering belief in me and for the sacrifices you have made to help me pursue my dreams.

And to my dear fiancée, Martina, I am overwhelmed with gratitude for your unwavering love, support, and understanding throughout this journey. Your presence in my life has been a constant source of strength and inspiration, sustaining me

through the challenges and triumphs of academia. You have stood by me with boundless patience and encouragement, cheering me on during moments of doubt and celebrating every milestone achieved. Your belief in me, even when I struggled to believe in myself, has been a guiding light, reminding me of my capabilities and fueling my determination to succeed. As I embark on the next chapter of my career, I carry with me the knowledge that I am not alone, for I have you by my side. Thank you, my love, for being my rock, my confidante, and my greatest source of joy. I am truly blessed to have you in my life.

Abstract

In science and engineering, the development of complex systems involves many-query optimization problems to identify both optimal design configurations to satisfy competing requirements, and states of the system to monitor their health status. The adoption of high-fidelity models for those optimization problems would be beneficial to identify superior optimization solutions, but would prohibitively rise the demand for computational resources and time required for every query. Multifidelity Bayesian Optimization (MFBO) combines information from models at different levels of fidelity to accelerate the optimization procedure: fast low-fidelity models are massively queried to explore different combinations of optimization variables while high-fidelity models are wisely evaluated to sparingly refine the accuracy of the optimization solution. Most existing MFBO algorithms adopt a suboptimal greedy approach which measures the utility of evaluating the objective function only at the immediate next iteration, and are sequential in nature precluding the parallel computation of high-fidelity models. Another limitation of MFBO lies in the purely data-driven search of optimal solutions which does not include explicit information about the physical domain of the system of interest.

In the first part of this thesis, we propose a Non-Myopic Multifidelity Bayesian Optimization framework (NM2-BO) to grasp the long-term reward from future steps of the optimization. Our computational strategy comes with a two-step lookahead policy that maximizes the cumulative reward obtained measuring the improvement in the solution over two steps ahead. We demonstrate NM2-BO for a large set of analytical benchmark problems and an aerodynamic design optimization problem. Moreover, we devise a Non-Myopic Multipoint Multifidelity Bayesian Optimization (NM3-BO) which relies on a two-step lookahead policy and a local penalization strategy to measure the future utility achieved evaluating multiple design configurations simultaneously. We demonstrate NM3-BO for the multidisciplinary design optimization of a space vehicle.

In the second part, we propose a generalized formulation for physics-aware MFBO (PA-MFBO) to embed forms of domain awareness during the optimization procedure. We formalize a bias in the search that captures the physical structure of the domain. This permits to partially alleviate the data-driven search from learning the domain properties on-the-fly, and sensitively enhances the management of multiple sources of information. PA-MFBO is demonstrated for an aerodynamic design optimization and a structural health monitoring problem. In addition, we develop a non-myopic formulation of the PA-MFBO algorithm (PA-NM2BO) which combines a lookahead policy with the physics-aware search characterizing the PA-MFBO algorithm. PA-NM2BO is validated against wind-tunnel data for an aerodynamic design optimization problem.

In the third part of this thesis, we propose a computational framework to accelerate diagnostics optimization problems to identify onboard incipient damages affecting complex systems. This procedure typically requires an expensive large amount of high-dimensional signals acquired through numerical models of the system. We devise the FREEDOM – Fast REliability Estimate and incipient fault Detection Of Multiphysics aerospace systems – algorithm to address such limitations in diagnostics. FREEDOM combines an original two-stage compression to compute an optimally reduced representation of the diagnostics signals for the minimum demand of onboard resources, and a single-fidelity Bayesian optimization scheme to infer multiple fault modes affecting the equipment. In addition, we extend the FREEDOM methodology to incorporate high-fidelity models directly in the diagnostics procedure, and devise the multifidelity FREEDOM algorithm (MF-FREEDOM). MF-FREEDOM relies on a multifidelity Bayesian scheme to identify fault parameters from the compressed signals: variable cost and fidelity models are optimally queried for a major reduction of the overall computational expense. The FREEDOM and MF-FREEDOM frameworks are demonstrated and validated for aerospace electromechanical actuators for flight controls affected by incipient multimodal faults.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Optimization is becoming essential in science and engineering to enable the next generation of advanced systems toward stringent requirements of safety, sustainability and operational performances. Indeed, optimization procedures play a key role to identify optimal design configurations to maximize the performance and minimize the environmental impact of novel engineering solutions [3], and the detection and identification of damages or faults to monitor the health condition of complex systems to maximize their useful life and minimize waste of resources [4]. Depending on the specific application, the identification of optimal solutions requires the minimization of an objective function that measures the goodness of design configurations with respect to the requirements, or the accuracy of the estimated health status of the system as to measurements.

The identification of optimal designs and health status of aerospace systems is notoriously difficult. Modern aerospace systems are characterized by a significant level of complexity associated with the demand for high-performance while operating in extreme environments. Those systems are characterized by technological architectures consisting of a variety of heterogeneous subsystems to meet the ever-increasing requirements of superior capabilities. The complexity of aerospace systems further increases with the demand for technological advances to improve the efficiency and environmental sustainability of the next generation aircraft [5, 6]. Indeed, new lower emission solutions for green aviation consist of innovative technologies and equipment that might complicate the overall architecture of systems currently implemented onboard. This growth of technical complexity and increasing number of components and physical interactions pose significant challenges both in the identification of optimal design configurations capable to satisfy competing requirements of performance, and identification of damages to anticipate severe faults and support high-regret decision making processes with potentially catastrophic consequences [7].

In recent years, the advances of computing technologies and platforms allowed the availability of advanced analysis and modeling approaches to accurately represent the behaviour of complex systems. This is the case of high-fidelity physics-based models for the numerical solution of governing partial differential equations as computational fluid dynamic solvers to represent viscous fluids, and finite element methods for the numerical analysis of mechanical structures, heath transfer and electromagnetic phenomena. Accordingly, simulation-based optimization methodologies [8–10] emerged as powerful computational frameworks where the objective function is computed through those physics-based numerical analysis that represent the behaviour of the system of interest under a specific combination of optimization variables – e.g. design or damage configurations.

The ambition of using high-fidelity models for optimization is shared by design and diagnostics communities. Indeed, simulation based optimization benefits from expensive high-fidelity models capable to accurately represent the behaviour of complex and multiphysics systems, as the aerospace systems, characterized by physical domains and phenomena difficult to be precisely captured. However, the extensive adoption of these high-fidelity models during the optimization procedure is hampered by the significant computational cost and time required for their evaluation, potentially in the order of months for a single evaluation on high performance computing platforms. This issue becomes more challenging for large-scale optimization problems – which are typically addressed in science and engineering – where the demand for model evaluations grows exponentially with the number of optimization variables to optimize. This makes the inclusion of high-fidelity models the major bottleneck of the simulation-based optimization procedure. In addition, the high level of complexity of the disciplinary analysis that characterizes multiphysics systems precludes the adoption of popular gradient-based optimization algorithms: the computation of the objective function derivatives requires a large amount of high-fidelity simulations and significantly increases the overall computational expense. Therefore, simulation-based optimization usually relies on a black-box approach where the disciplinary analyses are regarded as a pure input/output relationship, whose information about the mathematical properties and derivatives are not available [11]. These limitations preclude the scalability of high-fidelity simulation-based optimization for complex multiphysics systems, and hinder the identification of optimal design configurations and the reliable identification of complex faults affecting innovative components.

To contain the overall computational expense, simulation-based optimization strategies commonly adopt low-fidelity models of the system to reduce the general complexity of the disciplinary analysis and save computational resources with respect to the adoption of costly high-fidelity simulations. As reviewed and proposed by Peherstorfer et al. [12], low-fidelity models range from simplified models directly derived from the high-fidelity counterpart using expert knowledge about the specific disciplinary analysis [13], to projection-based models identifying a low-fidelity subspace that retains the essential features of the subsystem [14], and to surrogate-based representations where the input-output relationships of disciplines are derived from observations of the high-fidelity model [15]. Even if the computational cost of these low-fidelity models could be sensitively reduced if compared with high-fidelity representations, the former simplified models might not be adequate to depict complex non-linear phenomena that frequently characterize the disciplinary domain, while the latter projection-based and surrogate-based models might require a huge amount of high-fidelity data during their construction leading to significant numerical cost.

To address these limitations, this thesis explores surrogate modeling techniques to efficiently approximate the responses of high-fidelity models at untried combinations of optimization variables [16–20]. In particular, we investigate the opportunities offered by surrogate-based optimization strategies [21–23] as computational frameworks that seeks to alleviate the computational burden associated with high-fidelity simulation-based optimization adopting surrogate modeling to approximate the objective function. Surrogate models are computed starting from a set of observations of the objective function for certain combinations of optimization variables. Accordingly, the accuracy and efficiency of the resulting surrogate model are strongly related to the sampling technique adopted to determine this set of objective function evaluations. Among the possible sampling processes, goal-driven adaptive sampling schemes provide resource-efficient policies that aim at efficiently select combinations of optimization variables that maximize the utility of evaluating the objective function with respect to a given goal [19, 24, 25]. This goal can be either compute an accurate surrogate model of the objective function over the entire domain of interest, or compute an highly informative surrogate model to assist and inform the optimization procedure toward the minimum of the objective function.

This thesis focuses on Bayesian Optimization (BO) frameworks as goal-driven approaches for the optimization of expensive black-box objective functions [26–29]. BO aims at efficiently elicit valuable evaluations of the objective function from models of the system of interest to achieve a given goal and contain the computational expense of the optimization procedure. The Bayesian routine iteratively computes a surrogate model of the objective function, and defines a goal-driven sampling process through an acquisition function computed on the surrogate information. This acquisition function measures the merit of samples according to certain infill criteria, and permits to select the next sample that maximizes the query utility with respect to the given optimization goal. BO requires evaluating the objective function several times to asses the solution of the optimization problem. This may result in a prohibitively computational expense for engineering applications requiring the use of time consuming high-fidelity computer-based models, and where a limited number of evaluations of the objective are allowed to identify optimal solutions.

In many applications, scientists and engineers might rely on different models of the objective function with different degrees of accuracy and associated demand for computational resources. Multifidelity methods acknowledge the opportunity offered by low-fidelity representations and offer approaches to address the research gap of including expensive high-fidelity disciplinary analysis into the optimization process [30, 31]. Multifidelity methods combines data extracted from a library of disciplinary models that can be hierarchically ordered according to accuracy and computational cost [32, 33]. The availability of multiple levels of fidelity can be exploited to support the search procedure through a principled elicitation of information: fast low-fidelity models are used to systematically reduce the number of high-fidelity evaluations through a continuous trade-off between cost and accuracy [34, 35, 12, 36, 37]. In particular, this thesis focuses on Multifidelity Bayesian Optimization methodologies extending the BO approach to optimization scenario involving multiple representations of the objective function. Multifidelity Bayesian optimization (MFBO) provides a computational framework for black-box optimization and leverages disciplinary solvers at different levels of fidelity to accelerate the identification of promising solutions [38–42]. MFBO combines the function outputs at different levels of fidelity into a single probabilistic model, and uses this information to compute a multifidelity acquisition function that allows to jointly select the new set of optimization variables and the associated level of fidelity to query. In this setting, the surrogate model synthesizes the outcomes computed with multiple models into a unique emulator, and the acquisition function defines a goal-driven adaptive sampling scheme that identifies the optimal configuration and the associated level of fidelity to query at each iteration.

1.1 Research Gaps and Thesis Objectives

The multifidelity acquisition function is a key element of the multifidelity Bayesian optimization as defines a goal driven learning procedure that selects the next combination of optimization variables – e.g. design configuration or health status of the system – and level of fidelity of the objective function model to be evaluated. Most state-of-the-art acquisition functions have been proposed with the commonality that they realize a greedy selection of combinations of optimization variables: the quantification of the usefulness of evaluating the objective function with a certain level of fidelity considers the effects over the immediate next step without considering the future utility with respect to the given goal. The greedy nature of standard MFBO frameworks precludes greater informative gains that can be acquired through the measure of the long-term reward obtained at future steps of the optimization, and might lead to suboptimal solutions if a finite computational budget is allocated as a consequence of limited availability of computational resources. In such context, we formalize a Non-Myopic Multifidelity Bayesian Optimization (NM2-BO) framework that maximizes the reward over two-step ahead through a non-myopic multifidelity policy to increase the convergence rate and accelerate the optimization procedure. This non-myopic multifidelity learning scheme is derived formalizing MFBO as a dynamic system under uncertainty addressed through a dynamic programming technique, and defines an optimal policy as a sequence of decisions to maximize the two steps ahead utility obtained evaluating a combination of optimization variables with a model of a certain level of fidelity.

In the context of complex multidisciplinary design optimization problems, a significant challenge is the intrinsic demand to scale associated with the need for accurate evaluations of the objective function that can dramatically upscale during the search for improved design solutions. Thus, the standard sequential MFBO search might not be computationally efficient for optimization problems where the simulation of complex interdisciplinary relationships demands for a huge amount of high-fidelity data and associated computational expense. Within this scenario, we formulate the Non-Myopic Multipoint Multifidelity Bayesian Optimization (NM3-BO) algorithm that permits to maximize the long-term reward of a batch of design configurations and levels of fidelity to be evaluated simultaneously in parallel. NM3-BO defines a optimal policy as a sequence of decisions that maximizes the two-step ahead utility of a batch of paired designs and levels of fidelity to query in parallel.

A second limitation of multifidelity Bayesian optimization lies in its fully data-driven adaptive sampling informed exclusively through probabilistic data extracted from the surrogate model. Accordingly, the search procedure has to learn entirely from data both the surrogate model and the characterization of the discrepancies – frequently non-linear – between the different models over the entire physical domain. This can still require a large amount of high-fidelity information to depict the modeling correlations for each level of fidelity available, and results in intensive computations associated with the massive evaluations of accurate numerical models. In the engineering context, prior knowledge about these discrepancies is at disposal, either because it is formalized by the governing equations that represent the physics of the system or because it derives from the know-how of experts about the distinguishing physical phenomena characterizing the system behaviour. In these optimization scenarios, the introduction of the physics-awareness during the search procedure could lead to a principled and efficient use of high-fidelity data according to the prior knowledge about the physics of the system. Given this limitation, we propose an original Physics-Aware Multifidelity Bayesian Optimization (PA-MFBO) that incorporates prior domain knowledge to further improve and accelerate the optimization search in multifidelity settings. This permits to partially alleviate the data-driven search from the characterization of the domain structure while trying to identify the optimal combination of parameters that minimizes the objective function. The PA-MFBO relies on a multifidelity acquisition function distinctively shaped to combine data-driven information extracted from the surrogate model and prior/expert knowledge about the structure of the domain encapsulated during the search through a physics-aware utility function. In addition, we extend this physics-aware learning scheme through a non-myopic multifidelity policy to quantify the future gains achieved by a certain pair of optimization variables and levels of fidelity. This form of awareness allows to define an adaptive sampling scheme that efficiently manages different sources of information, targeting the balance between computational cost and accuracy demanded along the optimization search.

A third limitation arises in the context of diagnostics for complex safety critical engineering systems, and relates to the decrease of the computational efficiency of Bayesian algorithms associated with the high dimensionality of the diagnostic signals and the need for a massive amount of data acquired through the evaluation of expensive numerical models in order to assess incipient multimodal damages. This results in a computationally intensive identification of the health status of the system unfeasible for onboard monitoring of safety critical systems that require the fast and reliable assessment of incipient damages. To address these limitations, this thesis proposes an efficient computational framework for fault detection and isolation to accelerate the identification of incipient multimodal damages affecting complex engineering systems. We name and refer to our framework as FREEDOM as the short for Fast REliability Estimate and incipient fault Detection Of Multiphysics complex systems. In particular, we first devise FREEDOM as a single-fidelity computational framework implementing low-fidelity numerical models to obtain a procedure suitable for onboard monitoring. Following the promising results achieved with FREEDOM, we extend our computational framework through a multifidelity methodology and devise MF-FREEDOM to include high-fidelity models directly in the optimization: those high-fidelity data are extremely beneficial for diagnostic procedures at large since provide an accurate and reliable representation of the system of interest. Both FREEDOM and MF-FREEDOM adopt an original two-stage compression strategy that computes an optimal reduced representation of the diagnostics signal that retains only the most informative elements sensitive to faults. Two projection stages are used to compute this encoding map of the signal throught a combination of Dynamic Mode Decomposition (DMD) and Self Organizing Map (SOM). In FREEDOM, the identification of faults from those compressed signals is addressed through a Bayesian optimization scheme implementing low-fidelity numerical models to contain the computational cost of a single evaluation of the objective function. The original combination of these features substantially accelerates the diagnostics procedure: the Bayesian scheme leverages the informative compressed signals to accurately identify damage parameters with contained computational resources. The identification task through the MF-FREEDOM algorithm is executed through a multifidelity Bayesian scheme, which leverages and combines multiple models of the system at different levels of fidelity to effectively identify the health status of the system. The scheme is developed to embed high-fidelity simulations during the inference stage and aims at improving the effectiveness of the diagnostics; at the same time, lower-fidelity models are used to alleviate the computational cost of the procedure and more efficiently explore potential faults affecting the system.

To summarize, the objectives of this thesis are:

- To develop a Non-Myopic MFBO algorithm (NM2-BO) capable to maximize the long-term reward achieved evaluating a combination of optimization variables with a certain level of fidelity, and accelerate the search for optimal solutions while including costly high-fidelity models in the optimization procedure.
- To devise and implement a Non-Myopic Multipoint MFBO algorithm (NM3-BO) which combines the non-myopic search with the evaluation of a batch of combinations of optimization variables and levels of fidelity in parallel to address many query multidisciplinary optimization problems.
- To develop a Physics-Aware MFBO algorithm (PA-MFBO) to incorporate prior scientific/expert knowledge about the physical system of interest directly in the optimization procedure, and alleviate the expensive data-driven learning of specific characteristics of the physical domain and discrepancies between levels of fidelity.
- To combine the physics-aware feature with a non-myopic multifidelity policy and devise a Physics-Aware Non-Myopic MFBO (PA-NM2BO) to quantify the long-term reward of solutions while including awareness about the specific domain structures.
- To devise a computational framework FREEDOM to accelerate the identification of damage configurations of complex systems affected by incipient multiphysics and multimodal faults.
- To extend FREEDOM to a multifidelity setting and develop MF-FREEDOM to accelerate the diagnostics of incipient multiphysics and multimodal faults while efficiently including high-fidelity numerical model of the system of interest.

1.2 Thesis Outline

The reminder of this thesis is organized as follows. In Chapter 2, we discuss the general mathematical formalization of optimization problems typically addressed in science and engineering, such as design optimization, multidisciplinary design optimization, and diagnostics of complex multiphysics systems. Chapter 3 provides a general methodological background on Bayesian optimization and multifidelity Bayesian optimization, and illustrates popular state-of-the-art formulations of the acquisition function for both single-fidelity and multifidelity scenario. In Chapter 4, we devise our Non-Myopic Multifidelity Bayesian Optimization (NM2-BO) methodology through the formalization of an optimal multifidelity policy for MFBO and illustrate how to robustly approximate it through a Monte Carlo technique. In addition, we propose a multiple decision making strategy to enable parallel computations of a batch of designs and associated levels of fidelity, and develop a Non-Myopic Multipoint Multifidelity Bayesian Optimization (NM3-BO) Algorithm. In Chapter 5, we present our Physics-Aware Multifidelity Bayesian Optimization methodology and formalize a physical bias that embeds the available prior scientific and expert knowledge about the structure of the physical domain during the search procedure. In addition, we formalize a non-myopic physics-aware policy and develop a Physics-Aware Non-Myopic Multifidelity Bayesian Optimization (PA-NM2BO) methodology to measure the long-term reward of optimization solutions incorporating the awareness about the physics and the structures of the domain of interest. In Chapter 6, we devise a computational framework FREEDOM for diagnostics optimization problems and illustrate the original combination of a novel two-stage compression strategy to reduce the dimensionality of the diagnostics signals, and a Bayesian optimization scheme to identify damages affecting complex systems, In addition, we extend the FREEDOM capabilities to incorporate high-fidelity models directly in the diagnostics procedure, and devise the MF-FREEDOM algorithm where a multifidelity Bayesian scheme leverages the compressed signals to identify the faults affecting the system. In Chapter 7, we demonstrate the performance of the proposed algorithms for challenging aerospace design optimization problems, while Chapter 8 investigates and validates the capabilities of our computational frameworks for aerospace diagnostics and health monitoring applications. Finally, we summarize the thesis contributions in Chapter 9.

Chapter 2

Problem Statement

In this thesis, we consider a general decision-making problem in which we seek to minimize an objective function $f(\boldsymbol{x}) : \mathscr{X} \to \mathbb{R}$ by deciding iteratively a certain combination of optimization variables $\boldsymbol{x} \in \mathscr{X}$ within their bounds and subject to constraints. Mathematically, this decision-making problem is formalized as follows:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{with respect to} \quad & \boldsymbol{x}_L \leq \boldsymbol{x} \leq \boldsymbol{x}_U \\ \text{subject to} \quad & \boldsymbol{c}(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{h}(\boldsymbol{x}) = 0 \end{aligned} \tag{1.1}$$

where $\boldsymbol{x}_L$ and $\boldsymbol{x}_U$ are the bounds of the optimization variables, and $\boldsymbol{c}(\boldsymbol{x})$ and $\boldsymbol{h}(\boldsymbol{x})$ are the equality and inequality constraints, respectively.

This general formalization is representative of a variety of optimization problems in science and engineering; those include the support to engineering tasks such as the identification of the best design configurations to maximize the performance and minimize the environmental impact of novel engineering solutions, and the detection and identification of damages or faults to monitor the health condition of complex systems to maximize their useful life and minimize waste of resources. *We emphasize that, in spite of the specific application in design or diagnostics, we are always solving optimization problems.*

The values of the objective and constraints functions are typically computed through numerical models or real-world experiments of the system of interest. In this thesis, we focus on black-box representations of the objective and constraints functions depicted according to a pure input/output relationship whose information about the mathematical properties and derivatives are not available. This black-box approach is typical in real-world applications where computer codes or test-benches operate independently and define relationships between inputs and outputs while hiding the procedure associated with the discipline analysis.

In the following section, we detail the formalization of two major families of optimization problems frequently addressed in science and engineering, namely design optimization (Section 2.1) and multidisciplinary design optimization problems (Section 2.2) for the identification of the best combination of design parameters to maximize systems performance, and fault detection and identification problems (Section 2.3) for the non-destructive identification of systems damages or faults.

2.1 Design Optimization

In engineering design optimization, the objective function $f(\boldsymbol{x}) : \mathscr{X} \to \mathbb{R}$ represents a quantity of interest that determines if one design is better than another, and is computed for a given set of design variables $\boldsymbol{x} \in \mathscr{X}$ that describe the engineering system of interest. This design variables are typically bounded through the design space $\mathscr{X}$ to guarantee design configurations within the specifications of the design problem and avoid nonphysical designs. Accordingly, the design optimization problem seeks to identify the best combination of design variables within the design space that minimizes the objective function subject to constraints, and is mathematically formalized as follows:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{with respect to} \quad & \boldsymbol{x}_L \leq \boldsymbol{x} \leq \boldsymbol{x}_U \\ \text{subject to} \quad & \boldsymbol{c}(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{h}(\boldsymbol{x}) = 0 \\ & \mathscr{R}(\boldsymbol{x}) = 0 \end{aligned} \tag{1.2}$$

where is required the solution of governing disciplinary equations of the system of interest in residual form $\mathscr{R}(\boldsymbol{x})$ to compute the values of the objective and constraints functions. This formalization of the design optimization problem involves the identification of optimal designs for a single physical domain/discipline. The resulting design optimization process is conducted through a sequential approach where each component representative of a specific discipline is optimized separately, which might lead to suboptimal designs.

In addition, high-fidelity representations of the objective function can provide a remarkable contribution to enhance the search and identification of optimal design configurations. Unfortunately, the extensive adoption of these high-fidelity models during the optimization procedure is hampered by the significant computational cost and time required for their evaluation, potentially in the order of months for a single evaluation on high performance computing platforms. This issue becomes more challenging for many-query engineering optimization problems where the demand for model evaluations grows exponentially with the number of parameters to optimize.

2.2 Multidisciplinary Design Optimization

In many real-world applications, most engineering systems are multidisciplinary in nature. Multidisciplinary Design Optimization (MDO) relates to the development of computational methodologies for the design and optimization of complex systems taking into account the interactions of multiple disciplines [43, 44]. Those interactions can span different strength and width of the cross disciplinary couplings [45, 46]. The main benefit of this multidisciplinary approach is the optimization of design variables simultaneously: accounting for the various disciplines and minimizing an objective function that reflects a metric of goodness of the complete system permits to identify the design configuration that automatically provides the optimal trade-off between disciplines.

The general formulation of the MDO problem is formalized as follows [46]:

$$
\begin{aligned}
\text{minimize} \quad & f_0(\boldsymbol{x},z) + \sum_{i=1}^{D} f_i(\boldsymbol{x}_0,\boldsymbol{x}_i,z_i) \\
\text{with respect to} \quad & \boldsymbol{x},z,\bar{z} \\
\text{subject to} \quad & \boldsymbol{c}_0(\boldsymbol{x},z) \leq 0 \\
& c_i(\boldsymbol{x}_0,\boldsymbol{x}_i,z_i) \leq 0 \qquad \text{for } i = 1,...,D \\
& \mathscr{R}_i(\boldsymbol{x}_0,\boldsymbol{x}_i,z,\bar{z}_i) = 0 \quad \text{for } i = 1,...,D
\end{aligned}
\tag{1.3}
$$

The goal is to identify a set of design variables $\boldsymbol{x}$ that minimizes an objective function f subject to design constraints. We refer to design variables $\boldsymbol{x} = [\boldsymbol{x}_0,\boldsymbol{x}_1,...,\boldsymbol{x}_D]$ as variables of the MDO problem that are controlled by the optimizer during the optimization procedure, and consists of design variables $\boldsymbol{x}_0$ that affect all the D disciplines and the design variables $\boldsymbol{x}_i$ that concern the single i-th discipline. The optimizer searches for an improved design configuration that must satisfy the design constraints imposed for the single discipline c_i and for all the disciplines at once c_0. Each discipline is characterized by a simulation or model that permits to represent the domain of the specific discipline through a set of governing equations in residual form $\mathscr{R}_i(\boldsymbol{x}_0,\boldsymbol{x}_i,z,\bar{z}_i)$ which depend on the coupling variables z in output from simulations and the state variables $\bar{z}_i$ associated with the specific discipline. In particular, the coupling responses z depict the interactions between the disciplines in the form of output variables that are required to be exchanged between disciplines to solve the governing equations.

The identification of optimal design configurations that satisfy the couplings between disciplines demand for a large amount of data extracted from high-fidelity disciplinary solvers to closely depict the behaviour of the whole system. These high-fidelity representations are generally in the form of large scale computer-based models for the numerical solution of partial differential equations, and require prohibitive computational cost for their evaluation. This makes the inclusion of high-fidelity models the major bottleneck of the MDO procedure.

2.3 Fault Detection and Isolation

Model-based Fault Detection and Identification (FDI) permits to infer the fault condition of a system from measurements of signals sensitive to damages. The health assessment task is addressed solving an inverse problem: the actual damages affecting the system minimize the discrepancy between the diagnostic signal measured from the real system and the same signal simulated with a monitoring model [47–49]. Accordingly, the FDI problem targets the identification of the health status $\boldsymbol{x} = [x_1, ..., x_{n_k}]$ of the system as the combination of n_k faults parameters that minimizes the discrepancy f between the output reference signal of the real system $\boldsymbol{y}$ and the signal computed with a numerical model of the system $\boldsymbol{y}_M$. Accordingly, the model-based FDI procedure is mathematically formalized as follows:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) = ||\boldsymbol{y}(\boldsymbol{x}^*) - \boldsymbol{y}_M(\boldsymbol{x})|| \\ \text{with respect to} \quad & \boldsymbol{x}_L \leq \boldsymbol{x} \leq \boldsymbol{x}_U \\ \text{subject to} \quad & \boldsymbol{c}(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{h}(\boldsymbol{x}) = 0 \\ & \mathscr{R}(\boldsymbol{x}) = 0 \end{aligned} \tag{1.3}$$

where the solution of the optimization problem determines $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \mathscr{X}} f(\boldsymbol{x})$ the actual health status affecting the system.

The effectiveness of the FDI procedure closely relates to the accuracy of the monitoring model and the associated monitoring signal. In principle, the optimal setting is achieved adopting a monitoring model capable to exactly represent the dynamical behaviour of the damaged system. This would guarantee identical output and monitoring signals when the damaged configuration of the system is identified; conversely, modeling discrepancies between reference and monitoring signals might determine errors in the identification process and reduced accuracy of the health status identification. However, this favourable condition is usually unfeasible in practice: the acquisition of accurate monitoring signals requires the evaluation of expensive numerical models that heavily impact the computational burden of the diagnosis. Moreover, the identification of incipient multiphysics and multimodal faults affecting engineering systems – which is essential for safety critical systems

– further amplifies the demand for high-fidelity simulations to accurately capture the almost imperceptible alterations of the system behaviour during operations. In addition, this scenario is complicated by the high-dimensionality of the diagnostic signals, which are measured with an high acquisition frequency to ensure representativeness of the system health status. The ambition of using high-fidelity monitoring is shared by diagnostics and prognostics communities. However, these limitations preclude the scalability of model-based FDI methodologies for complex multiphysics systems, and hinder the reliable identification of complex faults affecting innovative components.

Chapter 3

Methodological Background

In science and engineering, the use of high-fidelity models to solve optimization problems and identify both optimal designs and health status of complex systems is notoriously difficult. Indeed, the adoption of those highly accurate representations leads to a significant bottleneck: the demand for resources to evaluate the objective function for all the combinations of optimization variables is difficult to be adequately satisfied. Indeed, the acquisition of data from these high-fidelity models involves huge non-trivial computational and economical costs that could arise from the computation of the objective function and its derivatives over ideally the entire optimization domain.

This motivates the growth of interest for the capabilities of surrogate models to efficiently approximate the output of expensive computer codes and real-world experiments at untried testing configurations [16–18]. Accordingly, Surrogate-Based Optimization (SBO) leverages surrogate modeling to efficiently approximate the distribution of the objective function over the domain and speed-up the search for the optimum. SBO is of particular interest for applications where the high-fidelity evaluation of the objective function leads to an unfeasible computation burden required for the optimization procedure. Across different application fields, it has been observed that SBO permits to alleviate the cost of the search task through surrogate models of the objective functions: this reduces its evaluation cost for different combinations of optimization variables to efficiently support the identification of optimal solutions. Examples of SBO applications include, but are not limited to, the identification of optimal designs in aerospace [50] and marine [51] engineering, states identification for medical applications [52] and autonomous systems [53], identification of optimal parameters for material design [54] and quantum computing [55]. Queipo et al. [21] discuss the adoption of surrogate models to reduce the computational cost associated with expensive optimization problems motivated by the application to the multidisciplinary design of complex aerospace systems. Wang and Shan [22] overview a variety of surrogate models techniques dedicating particular attention to global optimization and multi-objective optimization problems in engineering design. Haftka et al. [23] detail strategies for global optimization using surrogate models, and focus mainly on criteria for local and global searches from the perspective of parallelization to alleviate the computational burden of design in engineering.

Surrogate models are computed on evaluations of the objective function acquired through computer codes and/or physical experiments of the system: these sources of information are mostly treated as purely input/output black-box relationship whose analytical form is unknown and not directly accessible to the optimizer. Thus, the accuracy and efficiency of the resulting surrogate are highly dependent on the sampling approach adopted to select informative combinations of optimization variables for the acquisition of data. Among the numerous sampling schemes available in literature, it is possible to identify two major families: one-shot, and sequential schemes. The one-shot strategy defines a grid of samples over the domain all at once. Examples include Latin Hypercube [56], factorial and fractional factorials designs [57, 58], Placket-Burmann [59], and D-optimal [60]. However, it is very hard to identify a priori the best design of those experiments to efficiently compute the most informative surrogate. To overcome these limitations, sequential sampling selects samples over the domain through an iterative process [61, 62]. Among these, adaptive sampling [63] provides resource-efficient techniques that seek to reduce as much as possible the evaluations of the objective function, and target the improvement of the fitting quality across the domain and/or the acceleration of the optimization search [19, 24, 25].

This thesis focuses on goal-driven learning schemes as adaptive sampling procedures in which a learner informed by the available evaluations about the objective function directs the acquisition of new evaluations to maximize their utility with respect to a given goal. In particular, we focus on Bayesian optimization and Multi-fidelity Bayesian optimization as goal-driven methodologies to address black-box optimization problems characterized by the expensive evaluation of the objective function. In this chapter, we recognize that the popular paradigms for Bayesian optimization show substantial synergy with active learning schemes which has not been explicitly discussed and formally described in literature to date. Accordingly, we propose the explicit formalization of this synergy through an original perspective of Bayesian optimization and active learning as symbiotic expressions of adaptive sampling schemes. The aim of this unifying viewpoint is to support the use of those methodologies, and point out and discuss the analogies via their mathematical formalization. This unified interpretation is based on the formulation and demonstration of the analogy between the Bayesian infill criteria and the active learning criteria as the elements responsible for the decision on how to learn from samples to reach the given goal. In support of this unified perspective, this chapter first clarifies the concept of goal-driven learning (Section 3.1), and proposes a general classification of adaptive sampling methods that recognizes Bayesian optimization and active learning as methodologies characterized by goal-oriented search schemes (Section 3.2). Then, we provide an overview Bayesian optimization and multifidelity Bayesian optimization (Section 3.3), and elucidate the synergy between Bayesian optimization and active learning mapping the Bayesian learning features on the active learning properties (Section 3.4). The mapping is discussed through the analysis of three popular Bayesian frameworks for both the case of a single information source, and when a spectrum of multiple sources are available to the search. This chapter is based on the article [64] also available as preprint version [65].

3.1 Goal-Driven Learning

Goal-driven learning is a decision-making process in which each decision is made to acquire specific information about the system of interest that contribute the most to achieve a given goal [66–71]. This learning goal can be the increase of the knowledge of the system behaviour over all the domain of application, or acquire specific knowledge to enhance and accelerate the identification of optimization solutions. Accordingly, a goal-driven learner selects what to learn considering both the current knowledge and information needed, and determines how to learn quantifying the relative utility of alternative options in the current circumstances.

This thesis focuses on Bayesian optimization and active learning as goal-driven procedures where a surrogate model is built to accurately represent the behaviour of a system or effectively inform an optimization procedure to minimize given objectives. This goal-driven process is guided by learning principles that determine the "best" location of the domain to acquire information about the system, and refine the surrogate model towards the goal – improve the accuracy of the surrogate or minimize an objective function over the domain. Formally, these surrogate based modeling and optimization problems can be formulated as a minimization problem of the following form:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \mathscr{X}} f(\boldsymbol{R}(\boldsymbol{x})) \tag{3.1}$$

where $f(\boldsymbol{R}(\boldsymbol{x}))$ denotes the objective function evaluated at the location $\boldsymbol{x} \in \mathscr{X}$ of the domain $\mathscr{X}$. The objective function is of the general form $f = f(\boldsymbol{R}(\boldsymbol{x}))$, where $\boldsymbol{R}(\boldsymbol{x})$ represents the response of the system of interest evaluated through a model – e.g. computer-based numerical simulations or real-world experiments. In surrogate based modeling, the objective function can be represented as the error between the approximation of the surrogate model and the response of the system: the goal is to minimize such error to improve the accuracy of the surrogate over all the domain. In surrogate based optimization, the objective function represents a performance indicator dependent on the system response: the goal is to minimize this indicator to improve the capabilities of the system according to given performance requirements. Goal-driven techniques address Equation (3.1) through a decision-making iterative process where learning principles tailor the acquisition of specific knowledge about the objective function – evaluation of f at certain domain location $\boldsymbol{x}$ – currently needed to update the surrogate and inform the learner towards the given goal.

In this context, the goal-driven learner is the agent that makes decisions based on the current knowledge of the system of interest, and acquires new information to accomplish a given goal while augmenting the awareness about the system itself. In practice, the learner queries the sample that maximizes the utility to achieve the desired goal: specific learning principles quantify this utility based on the surrogate estimate and in response to information needs. At the same time, the surrogate model is dynamically updated once new information are acquired, and informs the learner to focus and tailor on the fly the elicitation of samples to further overarching the goal. Thus, the distinguishing element of a goal-driven learning procedure is represented by the mutual exchange of information between the learner and the surrogate model: the learner assimilates the information from the surrogate to make a decision aimed

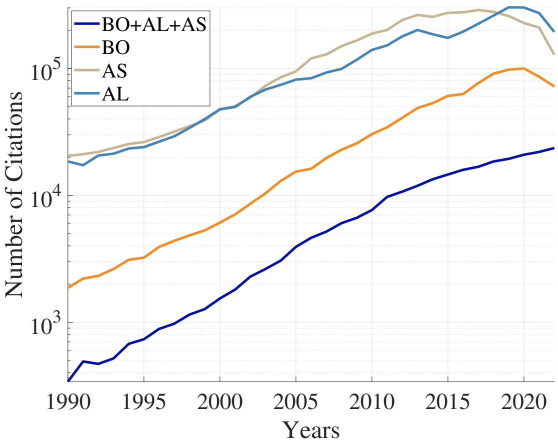


Fig. 3.1 Citations of Bayesian Optimization (BO), Active Learning (AL), Adaptive Sampling (AS) and the three terms combined (BO+AL+AS).

at achieving the goal, and the approximation/prediction of the surrogate is enriched by the result of this decision.

3.2 Adaptive Sampling Classification

Bayesian optimization and active learning realize adaptive sampling schemes to efficiently accomplish a given goal while adapting to the previously collected information. In recent years, there has been a profusion of literature devoted to the general topic of adaptive sampling but arguably a blurring of focus: many contributions from different field provided a deal of interesting advancements, but also led to some degree of confusion around the concepts of adaptive sampling, active learning and Bayesian optimization. Figure 3.1 illustrates the use of the words "adaptive sampling", "active learning", and "Bayesian optimization" from 1990 to 2022. In addition, we report the combined use of all the three words over the same period of time. It can be appreciated both the general increasing trend of use of the three techniques and the associated increase of the use of the three terms combined. Many times the three concepts have been used as complete synonyms, with some growing abuse motivated by the difficulties to map the (shaded) boundaries.

Stemming from these considerations, we recognize that adaptive sampling is not always superimposable with active learning and Bayesian optimization. Figure

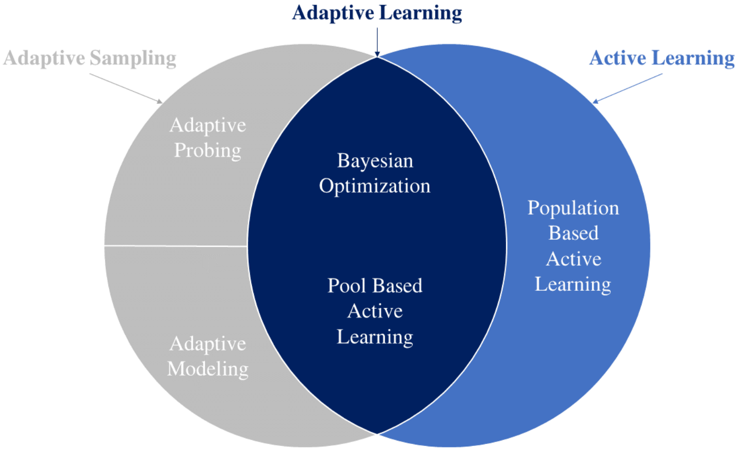


Fig. 3.2 Where adaptive sampling and active learning meet: this thesis focuses on the synergies between Bayesian optimization and active learning as goal-driven learning procedures driven by common learning principles.

3.2 illustrates the relationships between those three methodologies. We propose a classification of adaptive sampling techniques in three main families, namely adaptive probing (Section 3.2.1), adaptive modeling (Section 3.2.2) and adaptive learning (Section 3.2.3). This classification is based on the concept of goal-driven learning as the distinctive element of adaptive learning methodologies: the learner assimilates the information from the surrogate model to make a decision aimed at achieving a goal, and the surrogate is enriched by the result of this decision following a mutual exchange of information. Conversely, adaptive probing and adaptive modeling classes do not realize a goal-driven learning: the former does not rely on a surrogate model to assist the sampling procedure while the latter computes a surrogate model that is not used to inform the search task. This classification permits to clarify the reciprocal positions between adaptive sampling, active learning and Bayesian optimization.

Accordingly, adaptive sampling and active learning do not completely overlap. Active learning strategies are categorized into population-based and pool-based algorithms according to the nature of the search procedure [72, 73]. In population-based active learning, the distribution of the objective function is available: the learner seeks to determine the optimal training input density to generate training points without relying on a surrogate model of the objective function. Conversely, pool-based active learning computes a surrogate model of the unknown objective function that is used to inform the learner toward a given goal, and is updated during the procedure to refine the informative content supporting the learning procedure. Thus, pool-based active learning methods realize goal-driven learning schemes and can be collocated in the adaptive learning class while population-based active learning techniques can not be considered as adaptive samplings. Following the proposed classification, Bayesian optimization represents the logic intersection between active learning and adaptive sampling since (i) BO realizes an adaptive sampling scheme towards a given goal, and (ii) the BO goal-driven learning procedure is guided by learning principles also traceable in active learning schemes. This synergy between Bayesian optimization and active learning is the main focus of this chapter, and the following sections are dedicated to formalize and discuss this dualism. To support this discussion, we provide additional details of the proposed classification for adaptive sampling, and review some popular approaches for each of the three classes. The literature on adaptive sampling is vast, and a complete review goes beyond the purpose of this thesis. Although our discussion will not be comprehensive, the objective is to highlight the distinguishing features of each class and clarify the relative positions of adaptive sampling, active learning and Bayesian optimization.

3.2.1 Adaptive Probing

Adaptive probing schemes exploit the observations of previous samples without computing any surrogate model. These sampling procedures are informed exclusively from the collected data to guide the selection of the next location to query, and exclude the adoption of emulators to support the search. Several adaptive probing frameworks have been developed based on the Monte Carlo method [74, 75]. Among these, adaptive importance samplings [76–78] and adaptive Markov Chain Monte Carlo samplings [79, 80] represent popular methodologies adopted in different practical scenarios, from signal processing [81, 82] to reliability analysis of complex systems [83, 84]. Adaptive importance sampling uses previously observed samples to adapt the proposal densities and locate the regions from which samples should be drawn; this strategy permits to iteratively improve the quality of the samples distribution and enhance the accuracy of the relative inference from these observations. Adaptive Markov Chain Monte Carlo (MCMC) determines the parameters of the MCMC transition probabilities on the fly through already collected information. This adaptively generates new samples from an usually complex and high-dimensional distribution, and enhances the overall computational efficiency and reliability of the procedure. In the next paragraph, we report the mathematical formulation of adaptive importance sampling to illustrate the properties of adaptive probing methodologies and the elements that differentiate them from active learning paradigms.

Adaptive Importance Sampling Adaptive Importance Sampling (AIS) usually considers a generic inference problem characterized by a certain probability density function (pdf) $\tilde{\pi}(\boldsymbol{x})$ of a $d_{\boldsymbol{x}}$-dimensional vector of unknown statistic real parameters $\boldsymbol{x} \in \mathscr{X}$. AIS frameworks aim to provide a numerical approximation of some particular moment of $\boldsymbol{x}$:

$$I(f) = \mathbb{E}_{\tilde{\pi}}\left[f(\boldsymbol{x})\right] = \int f(\boldsymbol{x})\tilde{\pi}(\boldsymbol{x})\, d\boldsymbol{x} \tag{3.2}$$

where $f : \mathscr{X} \rightarrow \mathbb{R}$ can be any function of $\boldsymbol{x}$ integrable with respect to the pdf $\tilde{\pi}(\boldsymbol{x})$

The integral $I(f)$ is representative of different mathematical problems, from Bayesian inference [85] to the estimate of rare events [86]. In many practical scenarios, the integral $I(f)$ can not be computed in closed form. Adaptive importance sampling provides an algorithmic framework to efficiently address this problem.

Let us define a proposal probability density function $q(\boldsymbol{x})$ to simulate samples under the restriction that $q(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$ where $\tilde{\pi}(\boldsymbol{x})f(\boldsymbol{x}) \neq 0$. AIS provides an iterative procedure that improves the quality of one or multiple proposals $q(\boldsymbol{x})$ to approximate a non-normalized non-negative target function $\pi(\boldsymbol{x})$. At the beginning, AIS initializes N proposals $\{q_n(\boldsymbol{x}|\theta_{n,1})\}_{n=1}^{N}$ parameterized through the vector $\theta_{n,1}$. Then, the procedure simulates K samples from each proposal $\boldsymbol{x}_{n,1}^{(k)},\ n = 1,...,N, k = 1,...,K$, and assigns to each sample an associated importance weight formalized as follows:

$$w_n = \frac{\pi(\boldsymbol{x}_n)}{q(\boldsymbol{x}_n)}, \quad n = 1,...,N \tag{3.3}$$

These importance weights measure the representativeness of each sample simulated from the proposal pdf $q(\boldsymbol{x})$ with reference to the distribution of random variables $\tilde{\pi}(\boldsymbol{x})$.

At this point, this set of N weighted samples $\{\boldsymbol{x}_{n,1}^{(k)}, w_{n,1}^{(k)}\}, \quad n = 1,...,N, k = 1,...,K$ are used to define a self-normalized estimator:

$$\hat{I}^N(f) = \sum_{n=1}^{N} \bar{w}_n f(w_n) \tag{3.4}$$

where $\bar{w}_n = w_n / \sum_{j=1}^{N} w_j$ are the normalized weights. This permits to approximate the target function distribution as follows:

$$\tilde{\pi}^N(\boldsymbol{x}) = \sum_{n=1}^{N} \bar{w}_n \delta(\boldsymbol{x} - \boldsymbol{x}_N) \tag{3.5}$$

where δ represents the Dirac measure.

Finally, AIS realizes the adaptation phase and updates the parameters of the n-th proposals from $\theta_{n,1}$ to $\theta_{n,2}$ using the last set of drawn parameters [87] or all the parameters evaluated so far [88]. The whole procedure is repeated until a certain termination criteria is met (e.g. maximum number of iterations).

This adaptive policy permits to gradually evolve the single or multiple proposal densities to accurately approximate the target pdf. The generation of new samples is uniquely driven by the measurement of the importance of previous samples (weighting) that supports the updating of the proposal parameters (adaptation). Thus, AIS adaptively locates promising regions to query without gaining benefits from an overall quantification of the goodness of all the spectrum of samples available in the domain – e.g. through the construction of a surrogate model. On this basis, AIS and the general class of adaptive probing strategies are not considerable a learning procedures since the adaptation phase is not informed by a surrogate model updated on the fly during the procedure, and is not guided by a "learner" that assimilates information from this emulator and adapts the next queries to achieve a given goal.

3.2.2 Adaptive Modeling

Adaptive modeling paradigms sample the domain supported by the information from previous queries, and use the collected data to build a surrogate model. However, the informative content encoded in the emulator is not used to guide the sampling and decide the next point to evaluate. Adaptive modeling approaches have been extensively developed for the reliable propagation and quantification of uncertainties [89, 90], analysis of ordinary or partial differential equations [91, 92], and inverse problems [93, 94]. One common approach is represented by adaptive stochastic collocation methodologies, which use an adaptive sparse grid approximation scheme to construct an interpolant polynomial in a multi-dimensional random space [95, 96]. The adaptive selection of collocation points is driven by an error indicator [97] or estimator [98] that evaluates a certain number of sparse admissible subspaces of the domain: the subspace that exhibits the higher error is included in the grid and the new set of subspaces is identified. Other well-known adaptive modeling approaches are residual-based samplings distribution [99]. This family of techniques is mostly applied to improve the training efficiency of Physics-Informed Neural Networks (PINN) surrogate models. Residual-based approaches enhance the distribution of residual points by placing more samples according to certain properties of the residuals during the training of PINN. This decision can be made on the basis of locations where the residual of the partial differential equation is large [100], according to a probability density function of the residual points [101], and hybrid approaches of the above [99]. This permits to achieve better accuracy of the final PINN surrogate model while containing the computational burden associated with computations. Both stochastic collocation and residual based samplings are intended to build an efficient and accurate surrogate model over the domain of samples. However, the sampling procedure is adapted uniquely to previous evaluated samples without a learning procedure from data: the surrogate model is not used to inform the decision on where to sample, and is not progressively updated with previous information. In the following, we provide general mathematical details about adaptive stochastic collocation to analyze the peculiarities of the adaptive modeling class, and underline the absence of a learning process during the construction of the surrogate model.

Adaptive Stochastic Collocation Adaptive Stochastic Collocation (ASC) builds an interpolation function to approximate the outputs from a model of interest. This emulator is constructed on the evaluations of the model at valuable collocation points of the stochastic inputs to obtain the moments and the probability density function of the outputs.

Consider any point $\boldsymbol{x}$ contained in the random space $\Gamma \subset \mathbb{R}^N$ with probability distribution function $\rho(\boldsymbol{x})$. The goal of ASC is to find an interpolating polynomial $\mathscr{I}(f)$ to approximate a smooth function $f(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}$:

$$\mathscr{I}(f)(\boldsymbol{x}_k) = f(\boldsymbol{x}_k), \quad 1 \leq k \leq q \tag{3.6}$$

for a given set of points $\{\boldsymbol{x}_k\}_{k=1}^{q}$. The selection of the collocation points majorly influences the capability of the interpolating polynomial to be close to the original function f. For multivariate problems, the interpolation function is defined as follows using the tensor product grid:

$$\mathscr{I}(f) = (\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_N})(f) = \sum_{j_1=1}^{n_{i_1}} \cdots \sum_{j_N=1}^{n_{i_N}} f(\boldsymbol{x}_{j_1}^{i_1}, ..., \boldsymbol{x}_{j_N}^{i_N}) \cdot (\mathscr{L}_{j_1}^{i_1} \otimes \cdots \otimes \mathscr{L}_{j_N}^{i_N}) \tag{3.7}$$

where $\mathscr{U}^{i_k}$ is the univariate interpolation function for the level i_k in the k-th coordinate, $\boldsymbol{x}_{j_m}^{i_k}$ is the j_m-th node, and $\mathscr{L}_{j_k}$ are the Lagrange interpolating polynomials.

Equation 3.7 demands for $n_{i_1} \times \cdots \times n_{i_N}$ nodes, which indicate an exponential rate of computational cost growth with the number of dimensions. Adaptive stochastic collocation targets the reduction of this computational effort through an adaptive sparse grid of collocation points: the objective is to wisely place more points of the grid in the important directions to prioritize the collection of highly informative data. This adaptive sparse grid is defined through a subset of the full tensor product grid as follows:

$$\mathscr{A}_{q,N}(f) = \sum_{|\boldsymbol{i}| \leq q} (\Delta\mathscr{U}^{i_1} \otimes \cdots \otimes \Delta\mathscr{U}^{i_N})(f) = \mathscr{A}_{q-1,N}(f) + \sum_{|\boldsymbol{i}|=q} (\Delta\mathscr{U}^{i_1} \otimes \cdots \otimes \Delta\mathscr{U}^{i_N})(f) \tag{3.8}$$

where $\boldsymbol{i} = (i_1, ..., i_N) \in \mathbb{R}^N$, $|\boldsymbol{i}| = i_1 + ... + i_N$, q is the sparseness parameter, and the difference formulas are defined by $\mathscr{U}^0 = 0$ and $\Delta\mathscr{U}^i = \mathscr{U}^i - \mathscr{U}^{i-1}$.

Equation 3.8 leverages the previous results to extend the interpolation from level $q - 1$ to q through the evaluation of the multivariate function on the sparse grid:

$$\mathscr{H}_{q,N} = \bigcup_{|\boldsymbol{i}| \leq q} (\Delta\vartheta^{i_1} \times \cdots \times \Delta\vartheta^{i_N}) = \mathscr{H}_{q-1,N} + \bigcup_{|\boldsymbol{i}|=q} (\Delta\vartheta^{i_1} \times \cdots \times \Delta\vartheta^{i_N}) \tag{3.9}$$

where $\Delta\vartheta^i = \vartheta^i \backslash \vartheta^{i-1}$ are the newly added set of univariate nodes ϑ^{i_k} for level i_k in the k-th coordinate.

This scheme adapts the sampling procedure through the knowledge acquired on the fly, and efficiently leverages data to improve the quality of the interpolation function. In this case, the selection of the collocation points is intended to compute an emulator of the target function, but the adaptive sampling is not driven by the information acquired from this emulator. In addition, the acquisition of data is not used to learn and update the surrogate model. These considerations on ASC can be extended to the general class of adaptive modeling methods: even if the sampling scheme is conceived to construct surrogate models, the selection of promising locations to query is not delegated to a goal-driven learner that leverages a mutual exchange of information with the surrogate.

3.2.3 Adaptive Learning

Adaptive learning methodologies realize goal-driven learning processes characterized by the mutual exchange of information between the surrogate model and the goal-driven learner: the former is updated and refined after new evaluations of samples while the latter decides the next query based on the updated approximation given by the emulator. Bayesian optimization and pool-based active learning belong to this specific class of adaptive sampling techniques. Bayesian frameworks constitute a learning process driven by the mutual informative assimilation between an acquisition function – learner – and a surrogate model [27, 29]. The acquisition function commensurates the benefit of evaluating samples based on the prediction of the surrogate model, and selects the most useful sample to query towards the given goal – either improve the accuracy of the surrogate over the domain or effectively inform the optimization search; at the same time, the emulator is enriched with the data from the new query, and is updated to refine the approximation of the objective function over the domain. Similarly, pool-based active learning methods search the domain through a goal-driven learner informed by a classification model of samples [102, 103]. This process is characterized by the reciprocal flow of information between the learner and the emulator: the classification model is updated through the new evaluations of unsampled locations, and the learner uses this information to select the next query. Mathematical details about pool-based active learning are provided in the following section to better clarify the distinction between this class of adaptive learners, and the other classes which do not realize a goal-driven learning procedure.

Pool-Based Active Learning Pool-based active learning commonly defines an optimal sampling strategy to improve the accuracy of a surrogate model adopted to classify data-points from a target distribution of labels over the domain of samples $\mathscr{X}$. Considering this general classification task, pool-based active learning routine is grounded on a probabilistic estimate of the distribution of features f over the entire domain $\mathscr{X}$ through a surrogate model $\hat{f}$. This emulator is trained on a set of collected data-points, and maps features to labels $f_N(\boldsymbol{x}_n) = \hat{f}_n$ through a predicted probability $p_N(f_n = f|\boldsymbol{x}_n)$ that estimates the distribution of features over the domain. Suppose we have collected from a large pool of unlabelled data $\mathscr{X}$ the – small– dataset $\mathscr{D}_N\{\boldsymbol{x}_n, f(\boldsymbol{x}_n)\}_{n=1}^{N}$ observing the label values $f(\boldsymbol{x}_n)$ in output from an observation model or oracle at some informative locations $\boldsymbol{x}_n$. Based on this dataset, the goal-driven procedure learns a surrogate model $\hat{f}_N$ whose predictive framework emulates the behaviour of samples over the domain based on the previous collected information.

At this point, an utility function acts as the goal-driven learner informed by the surrogate model, and identifies the most promising sample to be labelled by the oracle according to a measure of utility with respect to the given goal – improve the accuracy of the classifier. The next query augments the dataset $\mathscr{D}_{N+1} = \mathscr{D}_N \bigcup \{\boldsymbol{x}_{N+1}, f_{N+1}\}$ and the surrogate model is updated. This utility function defines a learning policy that maps the current predictive distribution to a decision/action on where to sample in the next iteration as follows:

$$\boldsymbol{x}_{N+1} = \arg\max U(p_N(f_n = f|\boldsymbol{x}_n)) \tag{3.10}$$

Equation (3.10) mathematically formalizes the concept of goal-driven learning procedure: the learner leverages the predicted probability of the surrogate $p_N(y_n = y|\boldsymbol{x}_n)$ to make an action $\boldsymbol{x}_{N+1}$; at the same time, the decision is used to enrich the dataset $\mathscr{D}\{\boldsymbol{x}_n, f(\boldsymbol{x}_n)\}_{n=1}^{N+1}$ and update the predicted probability p_{N+1}. This mutual exchange and assimilation between the learner and the surrogate represents the key aspect that defines a goal-driven learning process and the whole class of adaptive learning sampling schemes.

3.3 Bayesian Frameworks

Bayesian optimization constitutes the mid-point between adaptive sampling and active learning. This intersection represents the focal point of this thesis, and motivates the substantial synergy between Bayesian optimization and active learning as adaptive sampling schemes capable to learn from data and accomplish a certain learning goal. The remaining of this section is dedicated to the general overview of Bayesian optimization considering both a single source of information (Section 3.3.1) and when multiple sources are available to the learning procedure (Section 3.3.2). This will guide the reader into the next sections that make explicit the symbiosis between Bayesian frameworks and active learning through our original perspective of Bayesian optimization as a way to actively learn with acquisition functions (Section 3.4).

3.3.1 Bayesian Optimization

The birth of Bayesian optimization can be retraced in 1964 with the work of Kushner [104] where unconstrained one-dimensional optimization problems are addressed through a predictive framework based on the Wiener process surrogate model, and a sampling scheme guided by the probability of improvement acquisition function. Further contributions have been proposed by Zhilinskas [105] and Mockus [106], and the methodology has been extended to high dimensional optimization problems in the works of Stuckman [107] and Elder [108]. Bayesian optimization achieved resounding success after the introduction of the Efficient Global Optimization (EGO) algorithm by Jones et al. [109]. EGO uses a Kriging surrogate model to predict the distribution of the objective function, and adopts the expected improvement acquisition function to measure the improvement of the optimization procedure obtained evaluating unknown samples.

The EGO methodology paves the way to the application of Bayesian optimization over a wide range of problems in science and engineering. These research fields demand for the efficient management of the information from black-box representations of the objective function – the procedure is only aware of the input and output without a priori knowledge about the function – to guide the optimization search. Engineering has been a pioneer in the adoption of Bayesian optimization: the design optimization of complex systems is frequently characterized by computationally intensive black-box functions which require efficient global optimization methods. Early applications relate to engineering design optimization [110], computer vision [111] and combinatorial problems [112]. Nowadays, the Bayesian framework becomes widely adopted in many fields including and not limited to engineering [113–116], robotics and reinforcement learning [117–119], finance and economics [120, 121], automatic machine learning [122, 123], and preference learning [124, 125].

Given a black-box expensive objective function $f : \mathscr{X} \rightarrow \mathbb{R}$, Bayesian optimization seeks to identify the input $\boldsymbol{x}^* \in \min_{\boldsymbol{x} \in \mathscr{X}} f(\boldsymbol{x})$ that minimizes the objective f over an admissible set of queries $\mathscr{X}$ with a reduced computational cost. To achieve this goal, Bayesian optimization relies on an adaptive learning scheme based on a surrogate model that provides a probabilistic representation of the objective f, and uses this information to compute an acquisition function $U(\boldsymbol{x}) : \mathscr{X} \rightarrow \mathbb{R}^+$ that drives the selection of the most promising sample to query. Let us consider the available information regarding the objective function f stored in the dataset $\mathscr{D}_N = \{(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_n, y_n)\}$ where $y_n \sim \mathscr{N}(f(\boldsymbol{x}_n), \sigma_\varepsilon(\boldsymbol{x}_n))$ are the noisy observations of the objective function and σ_ε is the standard deviation of the normally distributed noise.

At each iteration of the optimization procedure, the surrogate model depicts possible explanations of f as $f \sim p(f|\mathscr{D}_N)$ applying a joint distribution over its behaviour at each sample $\boldsymbol{x} \in \mathscr{X}$. Typically, Gaussian Processes (GPs) have been widely used as the surrogate model for Bayesian optimization [126, 127]. In GP regression, the prior distribution of the objective $p(f)$ is combined with the likelihood function $p(\mathscr{D}_N|f)$ to compute the posterior distribution $p(f|\mathscr{D}_N) \propto p(\mathscr{D}_N|f)p(f)$, representing the updated belief about f. The GP posterior is a joint Gaussian distribution $p(f|\mathscr{D}_N) = \mathscr{N}(\mu(\boldsymbol{x}), \kappa(\boldsymbol{x}, \boldsymbol{x}'))$ completely specified by its mean $\mu(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})]$ and covariance (also referred as kernel) function $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - \mu(\boldsymbol{x}))(f(\boldsymbol{x}') - \mu(\boldsymbol{x}'))]$, where $\mu(\boldsymbol{x})$ represents the prediction of the GP model at $\boldsymbol{x}$ and $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ the associated uncertainty. More details about the Gaussian process regression are provided in the following.

BO uses this statistical belief to make the decision of where to sample assisted by an acquisition function U, which identifies the most informative sample $\boldsymbol{x}_{new} \in \mathscr{X}$ that should be evaluated via maximization $\boldsymbol{x}_{new} \in \max_{\boldsymbol{x} \in \mathscr{X}} U(\boldsymbol{x})$. Then, the objective function is evaluated at $\boldsymbol{x}_{new}$ and this information is used to update the dataset $\mathscr{D}_N = \mathscr{D}_N \cup (\boldsymbol{x}_{new}, y(\boldsymbol{x}_{new}))$. Acquisition functions are designed to guide the search for the optimum solution according to different infill criteria which provide a measure of the improvement that the next query is likely to provide with respect to the current posterior distribution of the objective function. In engineering applications, we could retrieve different implementations proposed for the acquisition function, which differ for the infill schemes adopted to sample pursuing the optimization goal. Examples include the Probability of Improvement (PI) [128], Expected Improvement (EI) [109], Entropy Search (ES) [129] and Max-Value Entropy Search (MES) [130], Knowledge-Gradient (KG) [131], and non-myopic acquisition functions [132, 133].

The Probability of Improvement (PI) acquisition function encourages the selection of samples that are likely to obtain larger improvements over the current minimum predicted by the surrogate model, while the Expected Improvement (EI) considers not only the PI but also the expected gain in the solution of the optimization problem achieved evaluating a certain sample. Other popular schemes are entropy-based acquisition functions such as the Entropy Search (ES) and Max-Value Entropy Search (MES), which rely on estimating the entropy of the location of the optimum and the minimum function value, respectively, to maximize the mutual information between the samples and the location of the global optimum. Knowledge-gradient sampling procedures are conceived for applications where the evaluations of the objective function are affected by noise, recommending the location that maximizes the increment of the expected value that would be acquired by taking a sample from the location. Through the adoption of non-myopic acquisition functions, the learner maximizes the predicted improvement at future iterations of the optimization procedure, overcoming myopic schemes where the improvement of the solution is measured at the immediate step ahead.

The most widely used acquisition functions determine a greedy and sequential adaptive sampling scheme that considers only the immediate effect of evaluating the objective function for a single design, and do not consider the potential gains introduced in future evaluations. To address this type of greedy limitation, BO has been formalized as a partially observable Markov decision process [134], and several works [127, 135–137] provide solutions to this process and formalize non-myopic multifidelity acquisition functions. In addition, multipoint formulations of the BO framework have been proposed to evaluate in parallel multiple designs with a single level of fidelity of the disciplinary model [138, 139, 71]. However, the combination of non-myopic and multipoint formulations in literature are conceived exclusively for a single-fidelity framework only: the optimization process relies on the responses of disciplinary models at one single fixed level of fidelity. In the MDO context, this single fidelity approach could hinder the expensive high-fidelity disciplinary models to be interrogated directly during the search, which otherwise would result in prohibitive computational costs. In addition, the computational cost becomes unmanageable as it scales exponentially when the disciplinary couplings are also considered during the process: the identification of an optimal design that satisfies all the interactions and couplings across the disciplines would require massive evaluations of high-fidelity disciplinary models with the associated growth of the computational demand.

Gaussian Process

Gaussian process (GP) regression is a flexible and efficient framework to approximate the objective function f through a non-parametric kernel-based statistical model [126, 140, 141]. GP allows to predict the values of this objective function across the domain $\mathscr{X}$ based on its observations at previous evaluated points, and quantifies the uncertainty associated with the prediction. In the following, we briefly summarize the theoretical formulation of GP.

Let $\mathscr{D}_N = \{\boldsymbol{x}_n, y(\boldsymbol{x}_n)\}_{n=1}^N$ denote the dataset of N paired combinations of optimization variables $\boldsymbol{x}_n \in \mathscr{X} \subseteq \mathbb{R}^D$ and noisy observations of the objective function $y(\boldsymbol{x}_n) \sim \mathscr{N}(f(\boldsymbol{x}_n), \sigma_\varepsilon)$, where σ_ε is the standard deviation of the normally distributed noise. GP is a non parametric model characterized by its mean function $\mu(\boldsymbol{x}) : \mathscr{X} \to \mathbb{R}$ and the covariance function also defined as kernel function $\kappa(\boldsymbol{x}, \boldsymbol{x}') : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$. We assume that the observations of the objective function $\boldsymbol{\Delta} = \{f_n\}_{n=1}^N$ are jointly Gaussian, and the output $\boldsymbol{F} = \{y(\boldsymbol{x}_n)\}_{n=1}^N$ is normally distributed given f:

$$\boldsymbol{\Delta} \mid \boldsymbol{x} \sim \mathscr{N}(\mu, \boldsymbol{K}) \tag{3.11}$$

$$\boldsymbol{F} \mid \boldsymbol{\Delta}, \sigma_\varepsilon^2 \sim \mathscr{N}(\boldsymbol{\Delta}, \sigma_\varepsilon^2 \boldsymbol{I}) \tag{3.12}$$

where $\mu := \mu(\boldsymbol{x}_n)$, and $K_{i,j} =: \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Using the Bayesian inference principle, the GP regression combines the prior belief about the objective $p(f)$ with the likelihood function $p(\mathscr{D}_N|f)$ to compute the posterior distribution $p(f|\mathscr{D}_N) \propto P(\mathscr{D}_N|f)P(f)$, representing the updated surrogate model of the objective function. Assuming the prior of the objective as a GP: $f \sim GP(0, \kappa(\boldsymbol{x},\boldsymbol{x}'))$ with zero mean function $\mu(\boldsymbol{x}) = 0$, the posterior distribution is a GP completely defined by its mean μ and variance σ^2:

$$\mu(\boldsymbol{x}) = \kappa_i(\boldsymbol{x})^T \left(\boldsymbol{K} + \sigma_\varepsilon \boldsymbol{I}\right)^{-1} \boldsymbol{F} \tag{3.13}$$

$$\sigma^2(\boldsymbol{x}) = \kappa(\boldsymbol{x},\boldsymbol{x}) - \kappa_i(\boldsymbol{x})^T \left(\boldsymbol{K} + \sigma_\varepsilon \boldsymbol{I}\right)^{-1} \kappa_i(\boldsymbol{x}) \tag{3.14}$$

where κ_i is defined as $\kappa_i(\boldsymbol{x}) \doteq (\kappa(\boldsymbol{x},\boldsymbol{x}_0), \cdots, \kappa(\boldsymbol{x},\boldsymbol{x}_i))$. The posterior mean μ represents the maximum a posteriori probability estimate of the objective function f, and the posterior standard deviation σ quantifies the uncertainty of the surrogate model. These information are used to compute the acquisition function that guides the search towards the optimal solution of the optimization problem of interest.

In addition, the estimate of the uncertainty σ associated with the prediction μ of the objective function through the kernel function κ represents a significant property that potentially confers to BO a form of intrinsic reliability. Indeed, kernel functions form a reproducing kernel Hilbert space in which the objective function f has bounded norm [142]. This provides reliable confidence intervals on the objective function that determine a measure of the prediction and therefore reliability of the Gaussian process surrogate model.

3.3.2 Multifidelity Bayesian Optimization

The evaluation of black-box functions in engineering and science frequently requires time-consuming lab experiments or expensive computer-based models, which would dramatically increase the computational burden for the optimization procedure. This is the case of many-query problems such as design exploration and optimization [3], or the parametric analyses of complex system performance for trade off studies, diagnostics, and prognostics [143], where the evaluation of the objective function for enough samples can not be afforded in practice. In many real-world applications, the objective function can be computed using multiple representations at different levels of fidelity $\{f^{(1)}, ..., f^{(L)}\}$, where the lower the level of fidelity the less accurate but also less time-consuming the evaluation procedure. Multifidelity methods recognize that different representative levels of fidelity and associated cost can be used to accelerate the optimization process, and enable a flexible trade-off between computational cost and accuracy of the solution. In particular, multifidelity optimization leverages low-fidelity data to massively query the domain, and uses a reduced number of high-fidelity observations to refine the belief about the objective function toward the optimum [35, 12, 37].

Accordingly, Multifidelity Bayesian Optimization (MFBO) learns a surrogate model that synthesizes through stochastic approximation the multiple levels of fidelity available, and uses an acquisition function as the learner that selects the most promising sample and associated level of fidelity to interrogate. This learning procedure provides potential accelerations of the optimization procedure that is reflected in the likely improvement of the surrogate accuracy. According to Godino et al. [32], the improvement in performance occurs usually if the acquisition of large amount of high-fidelity data is hampered by the computational expense, the correlation between high-fidelity and low-fidelity data is high, low-fidelity models are sufficiently inexpensive; Under different circumstances, multifidelity optimization might not deliver substantial accelerations and quality of the surrogate. In recent years, multifidelity Bayesian optimization has been successfully adopted for optimization problems ranging from engineering design optimization [144, 145, 71, 146, 147], automatic machine learning [148–150], applied physics [151, 152], and medical applications [153, 154].

Multifidelity Bayesian optimization determines a learning procedure informed by a multifidelity surrogate model which defines an approximation of the objective $f^{(l)} \sim p(f^{(l)}|(\boldsymbol{x}, l), \mathscr{D}_N)$ at different level of fidelity, and represents the belief about the distribution of the objective function over the domain $\mathscr{X}$ based on data. A popular practice for MFBO is to adopt an autoregressive scheme to extend the Gaussian process surrogate model to a multifidelity setting and formalize a multifidelity Gaussian process surrogate model [155]. The posterior of the multifidelity Gaussian process is completely specified by the multifidelity mean function $\mu^{(l)}(\boldsymbol{x}, l) = \mathbb{E}\left[f^{(l)}(\boldsymbol{x})\right]$ that represents the approximation of the objective function at different levels of fidelity, and the multifidelity covariance function $\kappa^{(l)}((\boldsymbol{x}, l), (\boldsymbol{x}', l)) = \mathbb{E}\left[(f^{(l)}(\boldsymbol{x}, l) - \mu^{(l)}(\boldsymbol{x}, l))(f^{(l)}(\boldsymbol{x}', l) - \mu^{(l)}(\boldsymbol{x}', l))\right]$ that defines the associated uncertainty for each level of fidelity. The following section provides specific details about the formalization of the multifidelity Gaussian process surrogate model.

The availability of multiple representations of the objective function poses a further decision task that has to be accounted by the learner during the sampling of unknown locations: the selection of the most promising sample is effected with the simultaneous designation of the information source to be evaluated. This is obtained through a learner represented by the multifidelity acquisition function $U(\boldsymbol{x}, l)$ that extends the infill criteria of Bayesian optimization, and selects the pair of sample and the associated level of fidelity to query $(\boldsymbol{x}_{new}, l_{new}) \in \max_{\boldsymbol{x}\in\mathscr{X}, l\in\mathscr{L}} U(\boldsymbol{x}, l)$ that is likely to provide higher gains with a regard for the computational expenditure. Among different formulations, well known multifidelity acquisition functions to address optimization problems are the Multifidelity Probability of Improvement (MFPI) [156], Multifidelity Expected Improvement (MFEI) [38], Multifidelity Predictive Entropy Search (MFPES) [157], Multifidelity Max-Value Entropy Search (MFMES) [42], and non-myopic multifidelity expected improvement [70]. These formulations of the acquisition function define adaptive learning schemes that retain the infill principles characterizing the single-fidelity counterpart, and account for the dual decision task balancing the gains achieved through accurate queries with the associated cost during the optimization procedure.

Multifidelity Gaussian Process

In the multifidelity scenario, the GP regression synthesizes the simulations from the models at different levels of fidelity $[f^{(1)}, f^{(2)}, ..., f^{(L)}]$ into a unique predictive framework. The Multifidelity Gaussian Process (MFGP) is formalized through an autoregressive relation between adjacent levels of fidelity [155]. Let us assume we have collected paired input/output observations in the dataset $\mathscr{D}_N = \{\boldsymbol{x}_n, y^{(l_n)}(\boldsymbol{x}_n), l_n\}_{n=1}^{N}$, where the output $\boldsymbol{y} = \{y^{(l_n)}(\boldsymbol{x}_n)\}_{n=1}^{N}$ is normally distributed given $\boldsymbol{\Delta} = \{f_n^{(l_n)}\}_{n=1}^{N}$:

$$\boldsymbol{y} \mid \boldsymbol{\Delta}, \sigma_\varepsilon^2 \sim \mathscr{N}(\boldsymbol{\Delta}, \sigma_\varepsilon^2 \boldsymbol{I}) \tag{3.15}$$

assuming the same variance of the measurement noise σ_ε^2 for each level of fidelity. The MFGP relies on the linear autoregressive information fusion proposed by Kennedy and O'Hagan to approximate the objective function function [158]. This scheme assigns a GP prior to the lower fidelity model $f^{(1)} \sim GP(0, \kappa_1(\mathbf{x}, \mathbf{x}'))$ with mean function $\mu^{(l)} = 0$ and kernel function $\kappa_1(\mathbf{x}, \mathbf{x}')$, and defines recursively the higher-fidelity levels as follows:

$$f^{(l)} = \rho^{(l-1)}(\boldsymbol{x}) f^{(l-1)}(\mathbf{x}) + \zeta^{(l)}(\mathbf{x}) \quad l = 2, ..., L \tag{3.16}$$

where the scale factor $\rho^{(l-1)}(\boldsymbol{x})$ models the correlation between the outputs of adjacent levels of fidelity, and $\zeta^{(l)} \sim GP(0, \kappa^{(l)}(\mathbf{x}, \mathbf{x}'))$ represents the modeling discrepancy between two adjacent levels of fidelity as a Gaussian process with mean function $\mu^{(l)} = 0$ and kernel function $\kappa^{(l)}(\mathbf{x}, \mathbf{x}')$.

Following the Bayesian inference principle, the MFGP regression combines the prior belief about the objective $P(f^{(l)})$ with the likelihood function $P(\mathscr{D}_N | f^{(l)})$, and computes the updated posterior distribution of the objective function function $P(f^{(l)} | \mathscr{D}_N) \propto P(\mathscr{D}_N | f^{(l)}) P(f^{(l)})$. This multifidelity posterior is a Gaussian process completely specified by its mean function $\mu^{(l)}$ and variance function $\sigma^{2(l)}$:

$$\mu^{(l)}(\mathbf{x}) = \kappa_N^{(l)}(\mathbf{x})^T (\boldsymbol{K} + \sigma_\varepsilon \boldsymbol{I})^{-1} \boldsymbol{y} \tag{3.17}$$

$$\sigma^{2(l)}(\mathbf{x}) = \kappa((\mathbf{x}, l), (\mathbf{x}, l)) - \kappa_N^{(l)}(\mathbf{x})^T (\boldsymbol{K} + \sigma_\varepsilon \boldsymbol{I})^{-1} \kappa_N^{(l)}(\mathbf{x}) \tag{3.18}$$

where $\kappa_N^{(l)}$ is defined as $\kappa_N(\mathbf{x}) \doteq (\kappa((\mathbf{x}, l), (\mathbf{x}_1, l_1)), \cdots, \kappa((\mathbf{x}, l), (\mathbf{x}_N, l_N)))$, and $\boldsymbol{K}$ is the kernel matrix defined as follows:

$$\boldsymbol{K} = \begin{pmatrix} \kappa^{(l-1)}(\boldsymbol{x}, \boldsymbol{x}') \boldsymbol{K}^{(l-1)} & \rho \kappa^{(l-1)}(\boldsymbol{x}, \boldsymbol{x}') \boldsymbol{K}^{(l-1)} \\ \rho \kappa^{(l-1)}(\boldsymbol{x}, \boldsymbol{x}') \boldsymbol{K}^{(l-1)} & \rho^2 \kappa^{(l-1)}(\boldsymbol{x}, \boldsymbol{x}') \boldsymbol{K}^{(l-1)} + \kappa^{(l)}(\boldsymbol{x}, \boldsymbol{x}') \boldsymbol{K}^{(l)} \end{pmatrix} \tag{3.19}$$

where $\boldsymbol{K}^{(l-1)}(i, j) = \kappa\left((\boldsymbol{x}_i, l-1), (\boldsymbol{x}_j, l-1)\right)$ and $\boldsymbol{K}^{(l)}(i, j) = \kappa\left((\boldsymbol{x}_i, l), (\boldsymbol{x}_j, l)\right)$.

The mean function $\mu^{(l)}(\mathbf{x})$ represents the prediction of the objective function $f^{(l)}(\boldsymbol{x})$ over the possible combinations of optimization variables in the domain $\mathscr{X}$, and $\sigma^{2(l)}(\mathbf{x})$ quantifies the associated uncertainty.

3.4 An Active Learning Perspective

Bayesian frameworks and Active learning schemes exhibit a strong synergy: in both cases the learner seeks to design an efficient sampling policy to accomplish the learning goal, and is guided by a surrogate model that informs the learner and is continuously updated during the learning procedure. Active learning literature is vast an include a multitude of approaches [159–165, 73]. According to the well accepted classification proposed by Sugiyama and Nakajima [72], active learning strategies can be categorized in population-based and pool-based active learning frameworks according to the nature of the sampling scheme defined by the learner. Population-based active learning targets the identification of the best optimal density of the samples for training known the target distribution. Conversely, pool-based active learning defines an efficient sampling scheme to improve the efficiency of a surrogate model of the unknown target distribution over the domain of samples.

This section explicitly formalizes and discusses Bayesian frameworks as an active learning procedure realized through acquisition functions. In particular, pool-based active learning shows in essence a strong dualism with Bayesian frameworks. We emphasize this synergy through the dissertation on the correspondence between learning criteria and infill criteria; the former drive the sampling procedure in pool-based active learning, while the latter guide the search in Bayesian schemes through the acquisition function. This symbiosis is evidenced for the case of a single source of information adopted to query samples, and when multiple sources are at disposal of the learner to interrogate new input. Accordingly, we review and discuss popular sampling policies commonly adopted in pool-based active learning, and discern the learning criteria to accomplish a specific learning goal (Section 3.4.1). Then, the attention is dedicated to the identification of the infill criteria realized through popular acquisition functions in Bayesian optimization (Section 3.4.2). The objective is to explicitly formalize the synergy between Bayesian frameworks and Active learning as adaptive sampling schemes guided by common principles. The same avenue is followed to formalize this dualism for the case of multiple sources of information available during the learning procedure. In particular, we identify the learning criteria adopted in pool-based active learning with multiple oracles (Section 3.4.3), and compare them with the infill criteria specified by well-established multifidelity acquisition functions in multifidelity Bayesian optimization (Section 3.4.4). The objective is to clarify the shared principles and the mutual relationship that characterize the two adaptive learning schemes when the decision of the sample to query requires also the selection of the appropriate source of information to be evaluated.

3.4.1 Learning Criteria

Pool-based active learning determines a tailored sampling policy to ensure the maximum computational efficiency of the adaptive sampling procedure – limited and well selected amount of samples to query. This adaptive learning demands for principled guidelines to decide whether or not evaluate a certain sample based on a measure of its goodness. Learning criteria permit to establish a metric for quantifying the gains of all the possible learner decisions, and prescribe an optimal decision based in information acquired from the surrogate model. The vast majority of the literature concerning pool-based active learning identifies three essential learning criteria: informativeness, representativeness and diversity [166–168, 73, 169, 103]:

1. **Informativeness** measures the amount of information encoded by a certain sample. This means that the sampling policy is driven by the maximum likely contribution of queries that would significantly benefit the objective of the learning procedure.
2. **Representativeness** quantifies the similarity of a sample or a group of samples with respect to a target sample representative of the target distribution. Thus, the sampling policy exploits the structure underlying the domain to direct the queries in locations where a sample can represent a large amount of neighbouring samples.
3. **Diversity** estimates how well the queries are disseminated over the domain of samples. This is reflected in a sampling policy that selects samples scattering across the full domain, and prevents the concentration of queries in small local regions.

Figure 3.3 illustrates a watering optimization problem that attempts to clarify the peculiarities of each learning criteria. This simple toy problem requires identifying the areas of a wheat field where the crop is ripe and where it is still unripe for irrigation purposes. The learning goal is formalized as the identification of the area where the wheat is lower, which means an unripe cultivation and maximum requirements for irrigation. We assume that the learner can explore a maximum of five sites on the field during the procedure. A learner driven by the pure informativeness criterion (Figure 3.3(a)) would place observations in regions where the height of the wheat is minimum. This maximizes information on where it is strictly necessary to irrigate,

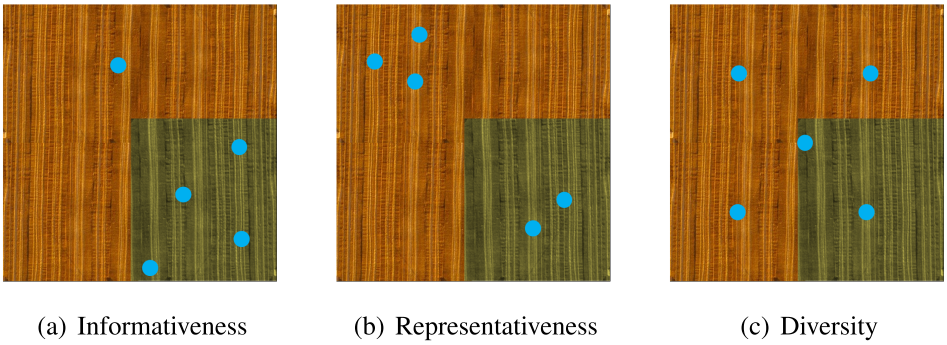

(a) Informativeness (b) Representativeness (c) Diversity

Fig. 3.3 Learning criteria: watering optimization problem.

but nothing is known about the regions where the grain is higher. Conversely, a purely representative sampling (Figure 3.3(b)) would probe the field by agglomerating observations to ensure the representativeness of the samples. This allows to partially know even areas where copious irrigation is not necessary, but increases the overall uncertainty given the small amount of samples for each agglomeration. If the learner pursues only the diversity of queries (Figure 3.3(c)), samples would scatter the field minimizing the maximum distance between measurements. Although this allows the queries to be distributed across the entire domain, the uncertainty is high as only one sample covers a respective area of the field.

The remaining of this section is dedicated to the revision and discussion of popular pool-based active learning schemes. We aim to provide a broad spectrum of approaches that exemplify the implementation of different learning criteria both individually and in combination. This permits to highlight the driving principles of learning procedures, and will help to better clarify the existing synergy between active learning and Bayesian optimization accounted in the following sections. Figure 3.4 summarizes the relationship between the methodologies reviewed in the following and the three learning criteria.

Informativeness-Based Learning procedures characterized by a pure informative criterion can be traced in uncertainty-based sampling policies. These approaches make the query decision based on the predictive uncertainty of the surrogate model, and seek to improve the density of samples in regions that exhibit the largest uncertainty with respect to a specific learning goal. Popular uncertainty-based active learning algorithms are uncertainty sampling and query-by-committee methods. Un-

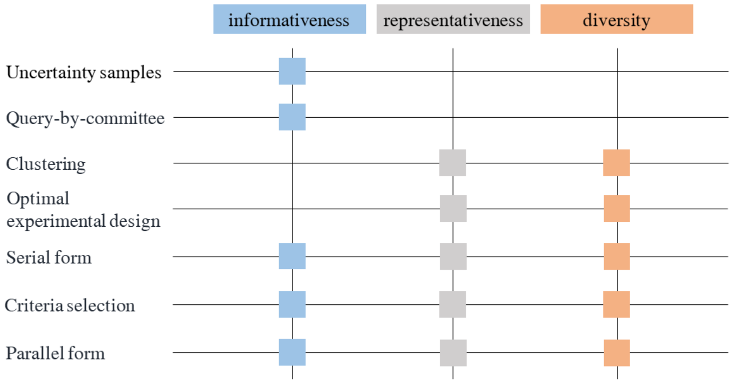


Fig. 3.4 Mapping methodologies to learning criteria.

certainty sampling algorithms probe the domain to improve the overall accuracy of the surrogate model according to a measure of the predictive uncertainty. Examples include the quantification of the uncertainty associated with samples [170], and its alternatives as margin-based [171], least confident [172] and entropy-based [173] approaches. Other strategies define sampling policies which promote the minimization of the surrogate model predicted variance [174] to maximize, respectively, the decrease of loss augmenting the training set [102], and the gradient descend [175]. Other uncertainty-based strategies are query-by-committee sampling schemes [160, 176], where the most informative sample to query is selected through the maximization of the disagreement between the predictions of a committee of surrogate models computed on subsets of the locations.

Representativeness/Diversity-Based Other pool-based active learning algorithms relies exclusively on representativeness and diversity learning frames: usually these learning criteria are implemented at the once in the learning procedure to drive the domain probing. This blend is justified by the mutual complementary relationship between representativeness and diversity: pure representativeness might concentrate the sampling in congregated representative domain regions without a proper dispersion of queries, while pure diversity might lead to the over-query of the domain and divert the learning procedure from the actual goals. The combination of both the learning criteria permits on one hand to leverage the representativeness of samples to accomplish a certain learning goal, on the other hand prevents the selection of redundant samples and high densities of queries only in circumstanced regions of the domain. Representative/diversity-based algorithms include a multitude of approaches that are commonly classified in two main schemes: clustering methodology and optimal experimental design. The former clustering algorithms identifify the most representative locations exploiting the underlying structures of the domain: the utility of samples is obtained as a function of their distance from the cluster centers. Popular examples include hierarchical clustering and k-center clustering. The former identifies a hierarchy of clusters based on the encoded information, and selects samples closer to the cluster centers [177]; the latter determines a subset of k congruent clusters that together cover the sampling space and whose radius is minimized, and the best sample minimizes the maximum distance of any point to a center [178]. Optimal experimental design defines a sampling policy based on a transductive approach: the learning procedure conducts the queries through a data reconstruction framework that measure the samples representativeness based on the capacity to reconstruct the training dataset. The selection of the most representative sample comes from an optimization process that maximizes the local acquisition of information about the parameters of the surrogate model [179–181].

Hybrid Recent avenues explore the combination of both informativeness and representativeness/diversity learning criteria to combine the goal oriented query of the first, and the use of underlying structures preventing over-density of the second. Accordingly, combined-based algorithms integrates multiple learning criteria to improve the overall sampling performance. Those approaches are commonly classified into three main classes [182, 103]: serial-form, criteria selection, and parallel-form approaches. Serial-form algorithms use a switching approach to take advantages from all the three learning criteria: informativeness-based techniques are used to select a subset of highly informative samples, and then representativeness/diversity techniques identifies the centers of the clusters on this subset as the querying locations [167]. Criteria selection algorithms rely on a selection parameter informed by a measure of the learning improvement that suggests the appropriate learning criteria to be used during the procedure [183]. Both serial-form and criteria selection strategies combine the three learning criteria through a sequential approach where each criteria is used consecutively during the learning procedure. Parallel-form methods combine simultaneously multiple learning criteria: the utility of each sample is judged by weighting informativeness and representativeness/diversity at the same time; then, valuable samples are selected through a multi-objective optimization of the weights to maximize at the same time the improvement in terms of learning goals and the exploitation of potentially useful structures of the domain [184–186].

3.4.2 Acquisition Functions and Infill Criteria

The synergy between active learning and Bayesian optimization relies on the substantial analogy between the learning criteria driving the active learning procedure and the infill criteria that characterize the Bayesian learning scheme. Infill criteria provide a measure of the information gain in terms of utility acquired evaluating a certain location of the domain. In Bayesian optimization, the acquisition function is formalized according to a certain infill criterion: this permits to quantify the merit of each sample with respect to a specific learning goal. Accordingly, the sample that maximizes the querying utility is observed to enrich the learning procedure towards this goal.

In particular, Bayesian learning schemes rely on two main infill criteria: global exploration ad local exploitation toward the optimum. The former exploration criterion concentrates the samples in regions of the domain where the uncertainty predicted by the surrogate is higher; this enhances the global awareness about the distribution of the objective function over the domain, but the resources might not be directed toward the goal of the procedure – e.g. minimum of the objective function. The latter exploitation criterion condensates the samples on regions where the surrogate model indicates that the objective is likely to be located – e.g. minimum of the Gaussian process mean function; exploitation realizes a goal-oriented sampling procedure that privileges the search for the objective without a potentially accurate knowledge of the overall distribution of interest.

The dilemma between exploration and exploitation represents a key challenge to be carefully addressed. On one hand, a learning procedure based on pure exploration might use a large amount of samples to improve the overall accuracy of the surrogate model without searching toward the learning goal. On the other hand, an exploitation-based learner might anchor a high density of samples to a suboptimal local solution as a consequence of information from an unreliable surrogate model. These extreme behaviours demonstrate the need to find a compromise between exploration and exploitation criteria.

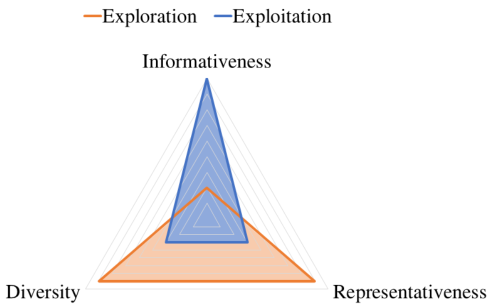


Fig. 3.5 Mapping of the learning criteria in active learning and infill criteria in Bayesian optimization.

In principle, infill criteria in Bayesian optimization are strongly related to the learning criteria commonly adopted in active learning. In particular:

- The concept of **exploration** is close to the **representativeness/diversity** criterion: both this learning schemes leverage underlying structures of the target distribution predicted by an accurate surrogate model to improve the awareness about the objective over the domain.
- The concept of **exploitation** is close to the **informativeness** criterion: the learner directs the selection of samples toward the believed objective without considering the global behaviour of the objective over the domain.

Figure 3.5 summarizes the mapping between infill criteria and learning criteria. The following sections discuss the formalization of (infill) active learning criteria for three most popular formulations of Bayesian acquisition functions, namely the expected improvement, probability of improvement, and max-value entropy search.

Expected Improvement

The Expected Improvement (EI) acquisition function quantifies the expected value of the improvement in the solution of the optimization problem achieved evaluating a certain location of the domain [109]. EI at the generic location $\boldsymbol{x}$ relies on the predicted improvement over the best solution of the optimization problem observed so far. Considering the Gaussian process as the surrogate model for Bayesian optimization, EI can be expressed as follows:

$$U_{EI}(\boldsymbol{x}) = \sigma(\boldsymbol{x})(I(\boldsymbol{x})\Phi(I(\boldsymbol{x}))) + \mathcal{N}(I(\boldsymbol{x});0,1) \tag{3.20}$$

where $I(\boldsymbol{x}) = (f(\hat{\boldsymbol{x}}^*) - \mu(\boldsymbol{x}))/\sigma(\boldsymbol{x})$ is the predicted improvement, $\hat{\boldsymbol{x}}^*$ is the current location of the best value of the objective sampled so far, $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution, μ is the mean function and σ is the standard deviation of the GP. The computation of $U_{EI}(\boldsymbol{x})$ requires limited computational resources and the first-order derivatives are easy to calculate:

$$\frac{\partial U_{EI}(\boldsymbol{x})}{\mu(\boldsymbol{x})} = -\Phi(I(\boldsymbol{x})) \tag{3.21}$$

$$\frac{\partial U_{EI}(\boldsymbol{x})}{\sigma(\boldsymbol{x})} = \phi(I(\boldsymbol{x})). \tag{3.22}$$

Both Equation (3.21) and Equation (3.22) demonstrate that $U_{EI}(\boldsymbol{x})$ is monotonic with respect to the increase of both the mean and the uncertainty of the GP surrogate model. This highlights a form of trade-off between exploration and exploitation: the formulation of the EI permits to balance the sampling in locations of the domain where is likely to have a significant improvement of the solution with respect to the current best solution, and the observations of regions where the improvement might be contained but the prediction is highly uncertain. In principle, it is possible to state that EI is driven by a combination of informativeness and representativeness/diversity criteria adopted in active learning. On one hand, the learner seeks to direct the computational resources towards the maximization of the learning contribution and achievement of the goal – informativeness; on the other hand, the learner pursues the awareness of the objective distribution over the domain to improve the quality of the prediction and better drive the search – representativeness/diversity. The predictive framework of the surrogate model regulates the learning thrusts privileging the one over the other on the basis of the information about the objective function acquired over the iterations.

Probability of Improvement

The Probability of Improvement (PI) acquisition function targets the locations characterized by the highest probability of achieving the goal, based on the information from the current surrogate model [128, 187]. PI measures the probability that the prediction of the surrogate model at the generic location is lower than the best observation of the objective function so far. Under the Gaussian process surrogate model, the PI acquisition function is computed in closed form as follows:

$$U_{PI}(\boldsymbol{x}) = \Phi\left(I(\boldsymbol{x})\right) \tag{3.23}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution and $\boldsymbol{x}^*$ is the current location of the best value of the objective. Similarly to EI, also $U_{PI}(\boldsymbol{x})$ is inexpensive to compute and the evaluation of the first-order derivatives requires simple calculations:

$$\frac{\partial U_{PI}(\boldsymbol{x})}{\partial \mu(\boldsymbol{x})} = -\frac{1}{\sigma(\boldsymbol{x})}\phi\left(I(\boldsymbol{x})\right) \tag{3.24}$$

$$\frac{\partial U_{PI}(\boldsymbol{x})}{\partial \sigma(\boldsymbol{x})} = -\frac{I(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\phi\left(I(\boldsymbol{x})\right) \tag{3.25}$$

where ϕ is the standard Gaussian probability density function. As demonstrated by Equation (3.24), regions of the input space characterized by lower values of the posterior mean of the GP are preferred for sampling, at fixed uncertainty of the surrogate. Moreover, Equation (3.25) shows that if $\mu(\boldsymbol{x}) < f(\boldsymbol{x}^*)$ the regions characterized by lower uncertainty are preferred and, conversely, PI increases with uncertainty. Overall, the PI acquisition function can be considered as an exploitative scheme that determines the most informative location as the one that potentially produces a larger reduction of the minimum value of the objective function observed so far. This is achieved sampling regions where the surrogate model is reliable and characterized by lower levels of uncertainty. In principle, this sampling scheme makes PI in accordance with the informativeness criterion: the search toward the optimum is uniquely directed in regions of the domain that exhibit the higher probability of achieving the goal according to the emulator prediction.

Entropy Search and Max-value Entropy Search

The Entropy Search (ES) acquisition function measures the differential entropy of the believed global minimum location of the objective function, and targets the reduction of uncertainty selecting the sample that maximizes the decrease of differential entropy [129]. The ES acquisition function is formulated as follows:

$$U_{ES}(\boldsymbol{x}) = H(p(\boldsymbol{x}^*|\mathscr{D})) - \mathbb{E}_{f(\boldsymbol{x})|\mathscr{D}}[H(p(\boldsymbol{x}^*|f(\boldsymbol{x}),\mathscr{D}))] \tag{3.26}$$

where $H(p(\boldsymbol{x}^*))$ is the entropy of the posterior distribution at the current iteration on the location of the minimum of the objective function $\boldsymbol{x}^*$, and $\mathbb{E}_{f(\boldsymbol{x})}[\cdot]$ is the expectation over $f(\boldsymbol{x})$ of the entropy of the posterior distribution at the next iteration on $\boldsymbol{x}^*$. Typically, the exact calculation of the second term of Equation (3.26) is not possible and requires complex and expensive computational techniques to provide an approximation of $U_{ES}(\boldsymbol{x})$.

The Max-value entropy search (MES) [130] acquisition function is derived from the ES acquisition function and allows to reduce the computational effort required to estimate Equation (3.26) measuring the differential entropy of the minimum-value of the objective function:

$$U_{MES}(\boldsymbol{x}) = H(p(f|\mathscr{D})) - \mathbb{E}_{f(\boldsymbol{x})|\mathscr{D}}[H(p(f|f^*,\mathscr{D}))] \tag{3.27}$$

where the first and the second term are now computed on the minimum value of the objective function f^*. This permits to simplify the computations and to approximate the second term through a Monte Carlo strategy [130]. The analysis of the derivatives is not possible for the MES acquisition function since the formulation of the second term of Equation (3.27) is intractable.

As reported by Wang et al. [130] in their experimental analysis, MES targets the balance between the exploration of locations characterized by higher uncertainty of the surrogate model, and the exploitation toward the believed optimum of the objective function. However, Nguyen et al. [188] demonstrate that MES might suffer from an imbalanced exploration/exploitation trade-off due to noisy observations of the objective function, and to the discrepancy in the computation of the mutual information in the second term of Equation (3.27). As a result, MES might over-exploit the domain in presence of noise in measurements, and over-explore when the discrepancy in the evaluation issue determines a pronounced sensitivity to the uncertainty of the surrogate model. Overall, the adaptive sampling scheme determined by the MES acquisition function follows both the informativeness and the representativeness/diversity learning criteria: the most promising sample is ideally selected targeting the balance between the search toward the believed minimum predicted by the emulator, and the decrease of uncertainty about the objective function distribution.

3.4.3 Learning Criteria with Multiple Oracles

Most of the active learning paradigms rely on a unique and supposed omniscient source of information about the target distribution. This oracle is iteratively queried by the learner to evaluate the value of the distribution at certain locations, and is assumed that its estimate is exact. In many other scenarios, the learner can elicit information from multiple imperfect oracles at different levels of reliability, accuracy and cost. Accordingly, the active learning community introduces a multitude of annotator-aware algorithms which are capable to efficiently learn from multiple sources of information. This require to make an additional decision during the learning procedure: the learner has to select at each iteration the most useful sample and the associated information source to query. In this context, the original learning criteria of informativeness and representativeness/diversity (Section 3.4.1) evolve and extend to quantify the utility of querying the domain with a certain level of accuracy and associated cost:

1. **Informativeness** seeks to maximize the amount of information from deciding the sample and information source to query. Thus, the learner might privilege the evaluations from accurate and yet costly oracles to capitalize from high-quality information and potentially reach the objective.

2. **Representativeness** attempts to identify underlying structures of the domain to better inform the search procedure. In this case, the decision making process might prefer to interrogate less expensive sources of information to contain the required effort, especially if cheap predictions of the target distribution exhibit good correlation with the estimate of the accurate oracle.

3. **Diversity** scatters the sampling effort over the domain to pursue a proper distribution of evaluations and augment the awareness about the target distribution. This might be favored by a major use of less accurate predictions of the target distribution, which are more likely to well address the cost/effectiveness trade-off during the diversity sampling.

The remaining of this section provides an overview of different multiple oracles active learning methodologies to present and further clarify popular extensions of the learning criteria to a multi-oracle setting.

Typically, active learning paradigms are extended to the multiple-oracle setting through relabeling, repeating-labeling, probabilistic and transfer knowledge, and cost-aware algorithms. Relabeling approaches query samples multiple times using the library of sources of information available, and the final query is obtained via majority voting [189]. Popular methodologies following this scheme pursue the identification of a subset of oracles according to the proximity of their upper confidence bound to the maximum upper confidence bound, and apply the majority voting technique only considering the queries of this informative subset [190]. Other multi-oracle active learning methods use a repeating-labeling procedure: the learner integrates the repeated – often noisy – prediction of the oracles to improve the quality of the evaluation process and the accuracy of the surrogate model learned from data [191]. Both relabeling and repeating-labeling approaches share a common drawback: the same unknown sample is evaluated multiple times with different oracles, which results in a sub-optimal usage of the available sources of information. Probabilistic and transfer learning methodologies attempt to overcome this limitation. Probabilistic frameworks rely on surrogate models specifically conceived for the multi-source scenario that provide a predictive framework to estimate the accuracy of each oracle in the evaluation of samples over the domain [192, 193]. Transfer knowledge approaches enhance the simultaneous selection of the most informative location to sample and the associated most profitable source to query; this is achieved through the transfer of knowledge from samples not evaluated in auxiliary domains to support the estimate of the oracle reliability [194]. Recent advancements in multiple oracles active learning are cost-effective algorithms, where the cost of an oracle is evaluated considering both the overall reliability of the prediction and the quality of samples in specific locations [195–197]. The cost-effectiveness property enhances the use of computational resources for the evaluation of samples, and targets the search toward the learning objectives while guarantees an optimal trade-off between evaluation accuracy and computational cost.

From the examined literature, the three learning criteria appear frequently coupled together during the learning procedure with multiple sources to query. This appears as a natural evolution of what has already been observed in the literature for active learning with single information source: the overall learning procedure usually benefits from a balanced learning scheme driven by informativeness and representativeness/diversity. In particular, informativeness permits to direct the search toward the learning goal, while representativeness/diversity augments the learner awareness about the target distribution over the domain; the combination of these learning criteria – in different measures – contributes to improve the performance of the active learning algorithms by using efficiently the computational resources and the information from multiple oracles.

3.4.4 Multifidelity Acquisition Functions and Infill Criteria

This section further investigates and highlights the synergy between active learning and Bayesian optimization for the specific case of multiple source of information used to accomplish the learning goal. Similarly to the single source setting, this symbiotic relationship is revealed through common principles characterizing the infill criteria in multifidelity Bayesian optimization and the learning criteria in active learning with multiple oracles. The multifidelity scenario imposes an additional decision to be made: the learner has to identify the appropriate information source to query according to an accuracy/cost trade-off. This is reflected in the formalization of infill criteria capable to define an efficient and balanced sampling policy, targeting either the wise selection of the samples and the level of fidelity which ensure the maximum benefits with the minimum cost. Accordingly, the multifidelity acquisition function formalizes an adaptive sampling scheme based on one or multiple infill criteria to quantify the utility of querying a location of the domain with a specific level of fidelity.

Based on this considerations, the exploration and exploitation infill strategies are extended according to the peculiarities of the multifidelity setting:

- **Exploration** is close to the **representativeness/diversity** criterion and defines a sampling policy that incentives the overall reduction of the surrogate uncertainty. Accordingly, the selection of the appropriate level of fidelity is driven by a trade-off between accuracy and evaluation cost. This might be accomplished through less-expensive low-fidelity information to contain the demand for computational resources during exploration.

- **Exploitation** is close to the **informativeness** criterion: concentrates the sampling process in the regions of the domain where optimal solutions are likely to be located. For this purpose, the learner might emphasize the use of accurate evaluations of the target function to refine the solution of the learning procedure toward the specific goal.

Similarly to the acquisition functions in Bayesian optimization (Section 3.4.2), the symmetry between informativeness and exploitation criterion, and between representativeness/diversity and exploration criterion is preserved in the multifidelity setting. The following sections are dedicated to the revision and discussion of popular multifidelity acquisition function, namely the multifidelity expected improvement (Section 3.4.4), multifidelity probability of improvement (Section 3.4.4) and multifidelity max-value entropy search (Section 3.4.4). The goal is to highlight the equivalent principles driving both the learning schemes, and further clarify the elements that encode the symbiotic relationship that exists between multifidelity Bayesian optimization and multi-oracle active learning.

Multifidelity Expected Improvement

The Multifidelity Expected Improvement (MFEI) extends the expected improvement acquisition function to define a learning scheme in the multifidelity setting as follows [38]:

$$U_{MFEI}(\boldsymbol{x},l) = U_{EI}(\boldsymbol{x},L)\alpha_1(\boldsymbol{x},l)\alpha_2(\boldsymbol{x},l)\alpha_3(\boldsymbol{x},l) \tag{3.28}$$

where $U_{EI}(\boldsymbol{x},L)$ is the expected improvement depicted in Equation (3.20) evaluated at the highest level of fidelity L, and the utility functions α_1, α_2 and α_3 are defined as follows:

$$\alpha_1(\boldsymbol{x},l) = corr\left[f^{(l)}, f^{(L)}\right] \tag{3.29}$$

$$\alpha_2(\boldsymbol{x}, l) = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma^{2(l)}(\boldsymbol{x}) + \sigma_\varepsilon^2}} \tag{3.30}$$

$$\alpha_3(l) = \frac{\lambda^{(L)}}{\lambda^{(l)}}. \tag{3.31}$$

The first element α_1 is the posterior correlation coefficient between the level of fidelity l and the high-fidelity level L, and accounts for reduction of the expected improvement when a sample is evaluated with a low fidelity model. This term reflects a measure of the informativeness of the l-th source of information at the location $\boldsymbol{x}$, and balances the amount of improvement achievable evaluating the high-fidelity level L with the reliability of the prediction associated with the level of fidelity l. Accordingly, α_1 modifies the learning scheme by adding a penalty in the formulation that reduces the U_{MFEI} when $1 \leq l < L$: this includes awareness about the increase of uncertainty associated with a low-fidelity prediction. The second element α_2 is conceived to adjust the expected improvement when the output at the l-th level of fidelity contains random errors. This is equivalent to consider the reduction of the uncertainty on the Gaussian process prediction after a new evaluation of the objective function is added to the dataset $\mathscr{D}$. This function allows to improve the robustness of U_{MFEI} when the representation of $f^{(l)}$ at different levels of fidelity is affected by noise in the measurements. The third element α_3 is formulated as the ratio between the computational cost of the high-fidelity level L and the l-th level of fidelity. This permits to balance the informative contributions of high- and a lower-fidelity observation and the related computational resources required for the evaluation. The effect of this term is to encourage the use of low-fidelity representations if almost the same expected improvement can be achieved with a high-fidelity evaluation. This directs wisely the use of computational resources to achieve the representativeness/diversity of samples, and prevents a massive use of expensive accurate queries during exploration phases.

Multifidelity Probability of Improvement

The Multifidelity Probability of Improvement (MFPI) acquisition function provides an extended formulation of the probability of improvement suitable for the multifidelity scenario as follows [156]:

$$U_{MFPI}(\boldsymbol{x}, l) = U_{PI}(\boldsymbol{x}, L)\eta_1(\boldsymbol{x}, l)\eta_2(l)\eta_3(\boldsymbol{x}, l) \tag{3.32}$$

where the PI acquisition function (Equation (3.23)) is computed considering the highest-fidelity level L available, and the utility function η_1, η_2 and η_3 are defined as follows:

$$\eta_1(\boldsymbol{x}, l) = corr\left[f^{(l)}, f^{(L)}\right] \tag{3.33}$$

$$\eta_2(l) = \frac{\lambda^{(L)}}{\lambda^{(l)}} \tag{3.34}$$

$$\eta_3(\boldsymbol{x}, l) = \prod_{i=1}^{n_l}\left[1 - R\left(\boldsymbol{x}, \boldsymbol{x}_i^{(l)}\right)\right]. \tag{3.35}$$

The first term η_1 shares the same formalization of the utility function α_1 in Equation (3.29), and accounts for the increase of uncertainty associated with low-fidelity representations $1 \leq l < L$ if compared with the high-fidelity output L. This reduces the probability of improvement if a low-fidelity representation is queried at a specific location of the input space $\boldsymbol{x}$. As already highlighted for the multifidelity expected improvement, η_1 incentives a form of informativeness learning where the information source is selected according to its capability to accurately represent the objective function. Similarly, the second utility function η_2 is also included in the multifidelity expected improvement in Equation (3.31) as the α_3 term. This element balances the computational costs and the informative contributions achieved through the l-th level of fidelity. This prevents the rise of computational demand produced by the over-exploitative nature of the probability of improvement (Section 3.4.2): η_2 encourages the use of fast low-fidelity data if the discrepancy between the l-th level of fidelity and the high-fidelity L – quantified by η_1 – is not significant. The third element η_3 is the sample density function and is computed as the product of the complement to unity of the spatial correlation function $R(\cdot)$ [198] evaluated for the n_l samples considering the l-th level of fidelity. This term reduces the probability of improvement in locations with an high sampling density – over exploitation of the domain – to prevent the clustering of data. Accordingly, η_3 promotes a form of representativeness/diversity learning scheme and encourages the exploration to augment the awareness about the domain structure.

Multifidelity Entropy Search and Multifidelity Max-Value Entropy Search

The Multifidelity Entropy Search (MFES) acquisition function is formulated extending the entropy search acquisition function to query multiple sources of information [157]

$$U_{MFES}(\boldsymbol{x}) = H(p(\boldsymbol{x}^*|\mathscr{D})) - \mathbb{E}_{f^{(l)}(\boldsymbol{x})|\mathscr{D}}[H(p(\boldsymbol{x}^*|f^{(l)}(\boldsymbol{x}),\mathscr{D}))] \tag{3.36}$$

where the expectation term $\mathbb{E}_{f^{(l)}(\boldsymbol{x})}[\cdot]$ considers multiple levels of fidelity $l = 1,...,L$. Similarly to the entropy search acquisition function, the computation of the expectation in Equation (3.36) is not possible in closed-form and requires an intensive procedure to provide a reliable approximation.

The Multifidelity Max-Value Entropy Search (MFMES) acquisition function can be formulated extending the max-value entropy search to a multifidelity setting as follows [42]:

$$U_{MFMES}(\boldsymbol{x}) = \left[H(p(f^{(l)}|\mathscr{D})) - \mathbb{E}_{f^{(l)}(\boldsymbol{x})|\mathscr{D}}[H(p(f^{(l)}|f^{*(L)},\mathscr{D}))]\right]/\lambda^{(l)} \tag{3.37}$$

where the differential entropy is measured on the minimum value of the objective function $f^{*(L)}$ considering the high-fidelity representation L. In this case, the approximation of the expectation term in Equation (3.37) relies on a Monte Carlo strategy that allows to contain the computational cost if compared with the procedure used for the MFES acquisition function [42].

In the multifidelity scenario, the MFMES acquisition function measures the information gain obtained evaluating the objective function $f^{(l)}(\boldsymbol{x})$ at a certain location $\boldsymbol{x}$ and associated level of fidelity l with respect to the global minimum of the objective function. This can be interpreted as an informativeness-driven learning based on the reduction of the uncertainty associated with the minimum value of the objective f^{*L} through the observation $f^{(l)}(\boldsymbol{x})$, where this uncertainty is measured as the differential entropy associated with the l-th level of fidelity. At the same time, the information gain is also sensitive to the accuracy of the surrogate predictive framework, and realizes a form of representativeness/diversity balancing to improve the awareness about the distribution of the objective function over the domain. The sensitivity to the computational cost $\lambda^{(l)}$ of the l-th level of fidelity is introduced in Equation (3.37) to balance the quality of the source – quantified by the information gain – and the demand for computational resources.

Chapter 4

Non-Myopic Multifidelity Bayesian Optimization

In this section, we present and formalize an original non-myopic multifidelity Bayesian optimization framework to sensitively accelerate the optimization search, and efficiently include high-fidelity responses to enhance the identification of superior optimization solutions. The proposed Non-Myopic Multifidelity Bayesian Optimization (NM2-BO) algorithm grasps the long-term reward from future steps of the optimization obtained evaluating the objective function with a certain level of fidelity. This is realized through the formalization of a two-step lookahead multifidelity acquisition function that maximizes the cumulative reward obtained measuring the improvement in the solution of the optimization problem over two steps ahead.

In addition, we propose the extension of the NM2-BO sequential adaptive sampling to enable multiple queries that can be issued in parallel. This is particularly attractive to address large scale multidisciplinary design optimization problems in science and engineering, where the demand for high-fidelity evaluations of the objective function leads to an intractable demand for computational resources. Accordingly, we formalize the Non-Myopic Multipoint Multifidelity Bayesian Optimization (NM3-BO) framework to enable the parallel evaluation of multiple information sources with different fidelities. NM3-BO selects a batch of promising combinations of optimization variables to be evaluated in parallel, and quantifies the expected long-term improvement at future steps of the optimization. The NM3-BO learning scheme leverages an original acquisition function based on the combination of a two-step lookahead policy and a local penalization strategy to measure the future utility achieved evaluating multiple configurations simultaneously.

In this context, the non-myopic optimal solution strategy for multifidelity Bayesian optimization is derived from the formulation of MFBO as a dynamic programming instance. We provide a general overview of the dynamic programming approach (Section 4.1) and develop the connection between multifidelity Bayesian optimization and dynamic programming. This permits to formalize the two-step lookahead multifidelity acquisition function (Section 4.2) solving the specific MFBO dynamic programming problem. The solution of our acquisition function requires an approximation of its intractable formulation for which we adopt the Monte Carlo approach (Section 4.3). The lookahead strategy is then combined with a Bayesian approach for multifidelity optimization into our non-myopic MFBO scheme (Section 4.4). In addition, we extend the optimal non-myopic policy for MFBO to a multiple decision making strategy and enable parallel computations of a batch of combinations of optimization variables and associated levels of fidelity (Section 4.5), and present and discuss the NM3-BO algorithmic framework (Section 4.6). Finally, we numerically investigate the proposed NM2-BO algorithm over a comprehensive set of benchmark problems specifically defined to stress test an validate goal-driven methods (Section 4.7). This chapter is based on the articles [199, 200, 70].

4.1 Dynamic Programming Problem

This section is dedicated to the general overview of Markov decision processes and dynamic programming as strategies to address the optimal decision making process for dynamic systems under uncertainty.

Markov Decision Processes (MDPs) are discrete-time stochastic control processes that permit to model the sequential decision making process under uncertainty [201]. The objective of MDP frameworks is to determine an optimal strategy as a set of decisions to achieve given objectives over the time. To address this setting, MDPs rely on three main ingredients: (i) a Markov chain model represents the transitions of the system to future states given an initial state, (ii) a decision-making model implements an action at each sequential state transition, and (iii) an objective formalized as an utility function measures the utility from the current state of the system, and alters the opportunities to gain utility in the future.

The solution of MDPs involves the statistical inference on the behaviour of the system at future states, and requires a dynamic model to represent the transitions of the dynamic system from one state to the other after exploring every possible decision. Dynamic Programming (DP) offers an optimization strategy to address the solution of MDPs partitioning the problem in simpler sub-problems defined through a recursive approach across several transitions [202]. Accordingly, DP addresses the MDP decision making process formalizing an optimal policy as a sequence of rules that, regardless of the initial state and the decision, makes the remaining decisions optimal over time with respect to the state resulting from the first decision. Thus, the current decision produces an effect on the dynamic of the system that contributes to both the current utility and to future utility through its effects on the evolution of the system at future states.

Let us consider a generic dynamic system fully characterized at each stage z by a state $s_z \in \mathscr{S}_z$, where $\mathscr{S}_z$ denotes a set of states that represent all the possible configurations of the system at each time step. The objective of the DP strategy is the maximization of an utility function over a time horizon T which quantifies the gains obtained applying an action $c_z \in \mathscr{C}_z$ to a certain state. The dynamic system evolves according to a transition model depicted as a Markov process in which the future state s_{z+1} depends on the current state s_z and the action adopted c_z:

$$s_{z+1} = \mathscr{F}_z(s_z, c_z, d_z) \tag{4.1}$$

where $\mathscr{F}_z : \mathscr{S}_z \times \mathscr{C}_z \times \mathscr{D}_z$ is the dynamic of the system at the stage z, and $d_z \in \mathscr{D}_z(s_z, c_z)$ is a random disturbance modelled as a random variable with probability distribution $P(\cdot|s_z, c_z)$.

DP seeks to construct an optimal policy that maximizes the chance of achieving the objective over a certain time horizon. A policy $\boldsymbol{\pi} = \{\pi_1, ..., \pi_T\}$ is defined as a sequence of rules $\pi_z : \mathscr{S}_z \rightarrow \mathscr{C}_z$ that maps each state s_z to the action c_z applied at that state. At each stage z, it is possible to define the value function $J_z^{\boldsymbol{\pi}}(s_z)$ as the expected reward obtained using the generic policy $\boldsymbol{\pi}$ at the state s_z:

$$J_z^{\boldsymbol{\pi}}(s_z) = \mathbb{E}\left[\beta^{T+1} r_{T+1}(s_{T+1}) + \sum_{z=1}^{T} \beta^{z-1} r_z(s_z, \pi_z(s_z), d_z)\right] \tag{4.2}$$

where $r_z : \mathscr{S}_z \times \mathscr{C}_z \times \mathscr{D}_z$ is the stage reward function that quantifies the immediate reward of applying an action c_z to a state z_z under the disturbances d_z, and β is the discount factor representing the relative reward gained in the next period with reference to the current time. Equation (4.2) includes two elements: the discounted terminal reward r_{T+1} which is a function of the state s_{T+1} acquired by the system after the last decision, and the sum of the discounted utilities from stage z to the horizon T.

Considering a set of admissible policies Π, the optimal policy $\boldsymbol{\pi}^*$ is determined assuming that the last action c_T taken at the horizon time T is optimal, and conditions the actions at the previous stages. Thus, the optimal value function $J_z^{\boldsymbol{\pi}^*}$ is defined maximizing the long-term expected reward:

$$J_z^{\boldsymbol{\pi}^*}(s_z) = \max_{\boldsymbol{\pi} \in \Pi} J_z^{\boldsymbol{\pi}}(s_z) \tag{4.3}$$

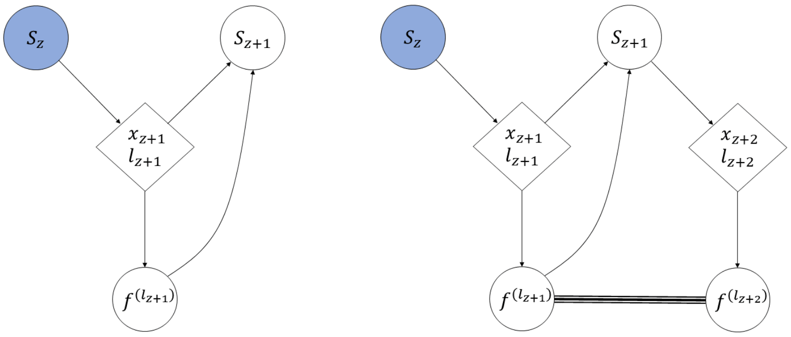
Following the DP backward iteration algorithm, it is possible to define a recursive relationship that identifies for each state a value function $J_z(s_z)$ for the stage z, given that stage $J_{z+1}(s_{z+1})$ has already been solved:

$$J_z^{\boldsymbol{\pi}^*}(s_z) = \max_{c_z \in \mathscr{C}_z} \mathbb{E}\left[r_z(s_z, c_z, d_z) + \beta J_{z+1}(\mathscr{F}_z(s_z, c_z, d_z))\right] \tag{4.4}$$

Then, the procedure is initialized by setting $J_{T+1}(s_{T+1}) = r_{T+1}(s_{T+1})$ and the solution is determined working backward from $z = T$ to $z = 1$.

4.2 Optimal Decision Making Process over Two-Step Ahead

Multifidelity Bayesian optimization can be regarded and formalized as a dynamic programming (DP) problem. DP permits to address the MFBO problem representing the decision making process under uncertainty as a dynamic system characterized by a dynamics consisting of a sequence of decisions to achieve a specific given goal [202, 203]. In particular, we consider MFBO as a dynamic system governed by a discrete stage dynamics. Accordingly, the associated DP problem is formalized through three main elements: (i) the probabilistic surrogate model represents the dynamic system operating under uncertainty, and synthesizes the multifidelity representations of the



(a) Myopic MFBO (b) Non-Myopic MFBO

Fig. 4.1 Bayesian networks describing the myopic (Figure 4.1(a)) and the non-myopic (Figure 4.1(b)) multifidelity Bayesian scheme. The shaded nodes ($\mathscr{D}_z$) are known, and the diamond nodes are the results of a decision. In the non-myopic scheme, the two f nodes are correlated under the multifidelity Gaussian process model (Section 3.3.2).

objective function over the entire domain, (ii) a system dynamics that describes the update of the surrogate model after a new observation of the objective function with a certain level of fidelity is collected, and (iii) a goal that measures the information gains achieved over two-step ahead as the long-term improvement of the quality of the optimization solution achieved evaluating a specific combination of optimization variables with a certain level of fidelity.

DP addresses the MFBO decision making process through the formalization of an optimization policy that determines the optimal sequence of decisions given the current information about the state of the MFBO dynamic system. In the remaining of this section, we formalize an optimal policy for MFBO that allows to identify the combination of optimization variables and the associated level of fidelity to query that maximizes the benefits two steps ahead with respect to the optimization goal. Figure 4.1 summarizes the main differences between the decision making process realized by the standard myopic MFBO (Figure 4.1(a)) and the proposed non-myopic MFBO paradigm (Figure 4.1(b)). The myopic procedure makes a sub-optimal decision with respect to the current knowledge of the system without considering possible future scenarios. Conversely, the non-myopic multifidelity scheme do not ignores the future decision that will be made by the algorithm in the future step, and capitalizes from these information to maximize the utility of the decision.

Let us consider the generic stage z of the MFBO dynamic system, and define the combination of optimization variables $\boldsymbol{x}$ and the associated evaluation of the objective function $f^{(l)}$ at the l-th level of fidelity, where the levels of fidelity can be in principle extended to any finite number. The MFBO system is fully characterized by a state s_z corresponding to the training dataset $\mathscr{D}_z = \{\boldsymbol{x}_n, f^{(l_n)}, l_n\}_{n=1}^{N}$ of N observations, and an action $c_z = \{\boldsymbol{x}_{z+1}, l_{z+1}\}$ that activates the dynamics of the system. Thus, a policy π_z is defined as a function that maps the state s_z to an action $c_z = \pi_z(s_z)$.

Given the state and the action of the system, we introduce the disturbances to represent a simulated value of the objective function at $\{\boldsymbol{x}_{z+1}, l_{z+1}\}$, defined as a random variable $d_z^{(l)} \sim \mathscr{N}(\mu_z^{(l)}(\boldsymbol{x}_{z+1}), \sigma_z^{2(l)}(\boldsymbol{x}_{z+1}))$ characterized with the mean and variance of the multifidelity Gaussian process (Section 3.3.2). At the new stage $z+1$, the system evolves to a new state s_{z+1} following its dynamics, corresponding to the augmented dataset $\mathscr{D}_{z+1} = \mathscr{D}_z \cup \{\boldsymbol{x}_{z+1}, f^{(l_{z+1})}, l_{z+1}\}$:

$$\mathscr{D}_{z+1} = \mathscr{F}(\boldsymbol{x}_{z+1}, f^{(l_{z+1})}, l_{z+1}, \mathscr{D}_z) \tag{4.5}$$

The disturbances d_{z+1} are then characterized using the multifidelity Gaussian process conditioned on $\mathscr{D}_{z+1}$. We define a stage reward function to measure the benefits of applying the action c_z to a state s_z subject to the disturbances d_z. For the MFBO dynamic system, we formulate the stage reward function as the reduction of the objective function achieved at the stage $z+1$ with respect to z:

$$r_z(\boldsymbol{x}_{z+1}, f^{(l_{z+1})}, l_{z+1}, \mathscr{D}_z) = (f_z^{*(L)} - f_{z+1}^{(L)})^+ \tag{4.6}$$

where $f_z^{*(L)}$ is the minimum value of the objective function at z evaluated at the highest level of fidelity. Thus, we can formulate the two-step lookahead multifidelity acquisition function at a generic stage z as the expected reward:

$$\begin{aligned} U_z^{\boldsymbol{\pi}}(\boldsymbol{x}_{z+1}, l_{z+1}, \mathscr{D}_z) = & \\ & \mathbb{E}[r_z(\boldsymbol{x}_{z+1}, f^{(l_{z+1})}, l_{z+1}, \mathscr{D}_z) + \\ & + J_{z+1}(\mathscr{F}(\boldsymbol{x}_{z+1}, f^{(l_{z+1})}, l_{z+1}, \mathscr{D}_z))] \end{aligned} \tag{4.7}$$

where the expectation is taken with respect to the disturbances, $J_{z+1}(\mathscr{F}(\cdot))$ is the long-term expected reward, and $\mathbb{E}[r_z(\cdot)] = U_{MFEI}(\boldsymbol{x}_{z+1}, l_{z+1})$ is the multifidelity expected improvement (Equation (5.1)).

Following the DP principle, the objective is to maximize the long term reward defining an optimal policy π_z^*. For the case of our two-step lookahead multifidelity acquisition function, the optimal policy is the one that identifies the optimal combination of optimization variables and the associated level of fidelity to query at the second future step to minimize the cumulative expected loss. Thus, we define the long term reward J_{z+1} as the maximum of the multifidelity expected improvement conditioned on the training set $\mathscr{D}_{z+1}$:

$$J_{z+1} = \max(U_{MFEI}(\boldsymbol{x}_{z+2}, l_{z+2})) \tag{4.8}$$

Combining Equation (4.7) with Equation (4.8), we formalize the two-step lookahead multifidelity acquisition function:

$$\begin{aligned} U_z^{\boldsymbol{\pi}^*}(\boldsymbol{x}_{z+2}, l_{z+2}, \mathscr{D}_{z+1}) = & \\ & U_{MFEI}(\boldsymbol{x}_{z+1}, l_{z+1}) + \\ & \mathbb{E}\left[\max(U_{MFEI}(\boldsymbol{x}_{z+2}, l_{z+2}))\right] \end{aligned} \tag{4.9}$$

Figure 4.2 shows the results for a synthetic optimization problem comparing three iterations of our non-myopic multifidelity search with the standard myopic counterpart based on the multifidelity expected improvement. The myopic acquisition function concentrates the high-fidelity samples near a local optimum of the objective function, without exploring the domain with accurate evaluations. This provides a relatively small improvement of the global solution of the optimization problem: the algorithm might use several queries of the objective function to depart from the suboptimal valley and move towards the global solution. Conversely, the non-myopic methodology uses high-fidelity queries to significantly explore the domain during the first iterations. Those information are beneficial for future evaluations: the exploration provides awareness about the distribution of the objective function and its local optimum that is used to maximize the future gains and place the last sample near the actual global optimum – third iteration.

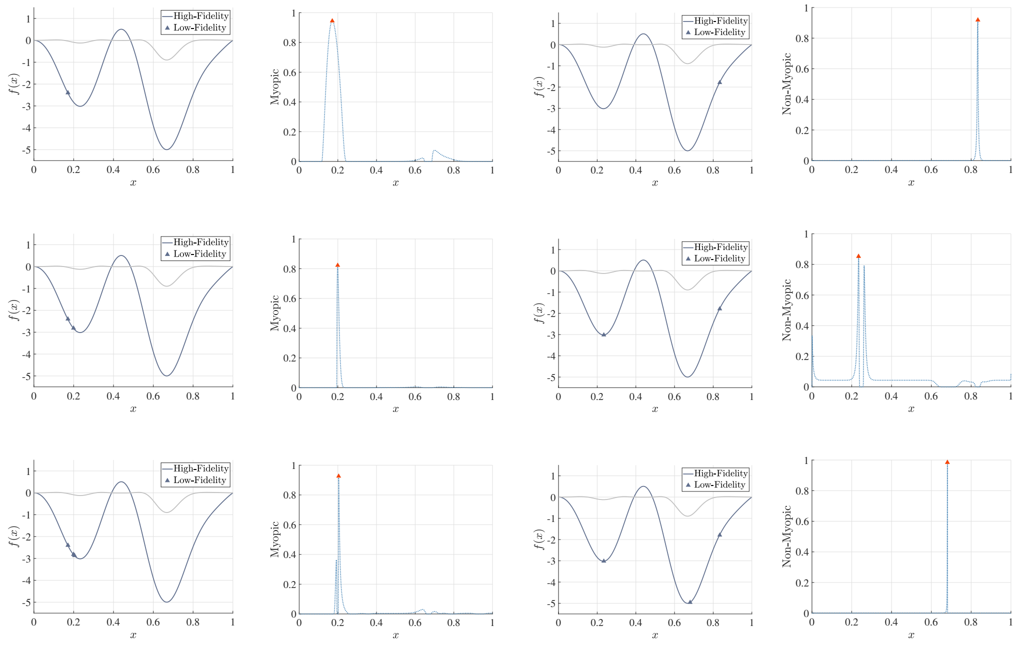


Fig. 4.2 Illustration of the searches conducted through a non-myopic MFBO and myopic MFBO for the toy example of the minimization of a 1-dimensional multifidelity synthetic function. Each row represents an iteration of the optimization procedure, and reports the high-fidelity queries together with the corresponding acquisition function for myopic MFBO (first two columns) and non-myopic MFBO (last two columns).

4.3 Robust Approximation of the Optimal Decision Making Process

The evaluation of the non-myopic multifidelity acquisition function (Equation (4.9)) demands for the solution of nested expectations and maximizations computationally intractable. We adopt the Monte Carlo approach to estimate $U_z^{\pi^*}$ as illustrated in Figure 4.3, using the reparameterization strategy proposed by Wilson et al. [204] to formulate the value of the objective function at the first step ahead with l-th level of fidelity:

$$f^{(l)}(\boldsymbol{x}_{z+1}) = \mu_z^{(l)} + \boldsymbol{C}_z^{(l)}(\boldsymbol{x}_{z+1})Z \tag{4.10}$$

where $\boldsymbol{C}_z^{(l)}$ is the Cholesky decomposition of the covariance matrix $\mathbf{K}_z$, and Z is an independent standard normal random variable. Then, we use Equation (4.10) to compute the mean and variance of the multifidelity Gaussian process at the step

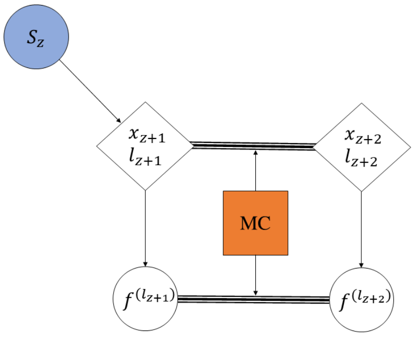


Fig. 4.3 Monte Carlo approximation of the relationship between the future stage $z+2$ and the stage $z+1$. The shaded nodes ($\mathscr{D}_z$) are known, and the diamond nodes are the results of a decision. Compare with Figure 4.1(b), the nested maximization problems required for each pair $\boldsymbol{x}_{z+1}, l_{z+1}$ and $\boldsymbol{x}_{z+2}, l_{z+2}$ are now address approximating our acquisition function $U_z^{\pi^*}$ through the Monte Carlo strategy.

$z+1$ for the generic input $\boldsymbol{x}$:

$$\mu_{z+1}^{(l)}(\boldsymbol{x}) = \mu_z^{(l)}(\boldsymbol{x}) + \boldsymbol{H}_z^{(l)}(\boldsymbol{x}) Z \tag{4.11}$$

$$\sigma_{z+1}^{(l)}(\boldsymbol{x}) = \sigma_z^{(l)}(\boldsymbol{x}) - \boldsymbol{H}_z^{(l)}(\boldsymbol{x}) \boldsymbol{H}_z^{(l)}(\boldsymbol{x})^T \tag{4.12}$$

where $\boldsymbol{H}_z^{(l)}(\boldsymbol{x}) = \kappa_z^{(l)}(\boldsymbol{x}) \boldsymbol{C}_z^{(l)-1}(\boldsymbol{x})$.

Equation (4.11) and Equation (4.12) are used to estimate the multifidelity expected improvement at the step $z+1$:

$$U_{MFEI}(\boldsymbol{x}_{z+2}, l_{z+2}) \sim \widehat{U_{MFEI}}(\boldsymbol{x}_{z+2}, l_{z+2}, Z) \tag{4.13}$$

To approximate the expectation term of Equation (4.9), we sample the random variable Z and compute $\mathbb{E}[\widehat{U}_z^{\pi^*}(\boldsymbol{x}_{z+2}, l_{z+2}, \mathscr{D}_{z+1}, Z)]$ averaging over many realizations of the two-step lookahead multifidelity acquisition $\widehat{U}_z^{\pi^*}$ evaluated using Equation (4.13).

This Monte Carlo approach requires evaluations of $\widehat{U}_z^{\pi^*}$ of the order of thousands to provide a reliable approximation of the non-myopic acquisition function. However, we emphasize that the proposed Monte Carlo technique is based on the evaluation of Equation (4.11) and Equation (4.12) which are inexpensive to compute.

Algorithm 1 NM2-BO: Non-Myopic Multifidelity Bayesian Optimization

```
Input: Feasible set $\mathscr{X} \in \mathbb{R}^D$, multifidelity objective function $f^{(l)}(\boldsymbol{x})$ and the multifidelity Gaussian process prior $GP(0, \kappa^{(l)}(\boldsymbol{x}, \boldsymbol{x}'))$
Output: $f^* = \min f(\boldsymbol{x})$
$\mathscr{D}_0 \leftarrow \{\boldsymbol{x}_n, f^{(l_n)}, l_n\}_{n=1}^{N_0}$ collect initial observations
$\mu_0^{(l)}, \sigma_0^{2(l)} \leftarrow$ learn the initial multifidelity GP
$i \leftarrow 1$
repeat
    Load $\boldsymbol{x}_i$ and associated $l_i$
    Evaluate $f^{(l_i)}(\boldsymbol{x}_i)$
    $\mathscr{D}_i \leftarrow \mathscr{D}_{i-1} \cup \{\boldsymbol{x}_i, f^{(l_i)}, l_i\}$
    $\mu_i^{(l)}, \sigma_i^{2(l)} \leftarrow$ update the multifidelity GP
    $z \leftarrow i$
    Compute $U_{MFEI}(\boldsymbol{x}_{z+1}, l_{z+1})$
    for $j \leftarrow 1, N_{MC}$ do
        $Z_j \leftarrow \mathscr{N}(0,1)$
        $\mu_{z+1}^{(l)}, \sigma_{z+1}^{2(l)} \leftarrow$ estimate the multifidelity GP
        Compute $\widehat{U_{MFEI}}(\boldsymbol{x}_{z+2}, l_{z+2}, Z_j)$
    end for
    return $\widehat{\boldsymbol{U}}_z^{\pi^*} = \{\widehat{U}_{zj}^{\pi^*}\}_{j=1}^{N_{MC}}$
    $U_i^{\pi^*} = \mathbb{E}\left[\widehat{\boldsymbol{U}}_z^{\pi^*}\right]$
    $[\boldsymbol{x}_{i+1}, l_{i+1}] \leftarrow \max(U_i^{\pi^*})$
    $i + 1 \leftarrow i$
until $B_i \leq B_{max}$
return $\boldsymbol{x}^*$ that minimize $f(\boldsymbol{x})$ over $\mathscr{D}_i$
```

Thus, the overall computational cost associated with the approximation of the two-step lookahead multifidelity acquisition function is negligible if compared with the computational cost of real-world optimization problems in science and engineering applications, where usually the evaluation of the objective function requires the query of expensive high-fidelity physics-based models.

4.4 NM2-BO Algorithm

The Algorithm 3 and the Figure 4.4 provide details about the Non-Myopic Multifidelity Bayesian Optimization (NM2-BO) routine, and illustrate the integration of

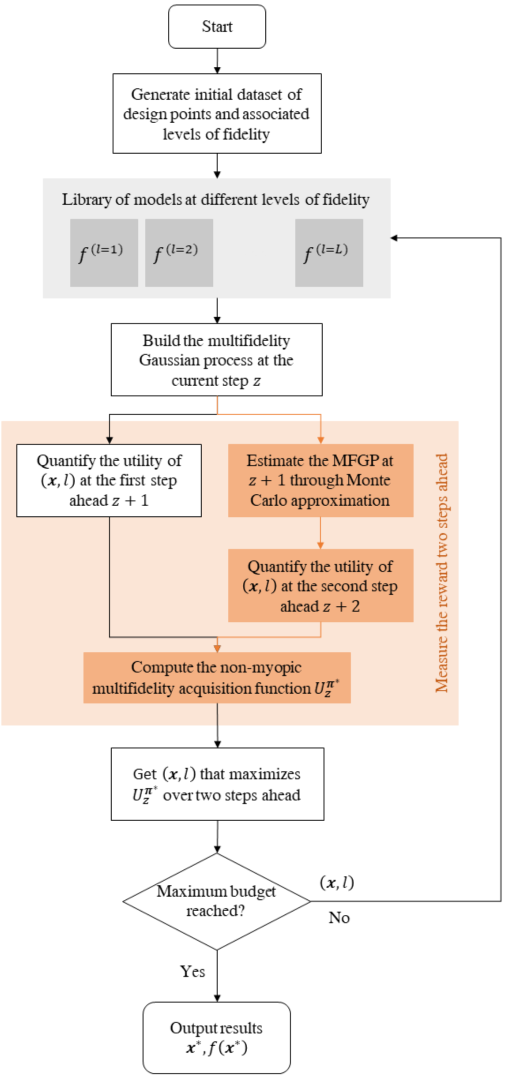


Fig. 4.4 Schematic representation of the NM2-BO framework. In orange are reported the distinguishing processes of NM2-BO with respect to a standard multifidelity Bayesian optimization framework.

our non-myopic multifidelity acquisition function (Equation (4.9)) into a multifidelity Bayesian framework. The optimization procedure begins with the definition of the D-dimensional domain $\mathscr{X} \in \mathbb{R}^D$ assembled through a Latin Hypercube design of experiments [205], together with the library of the objective function representations at different levels of fidelity $f^{(l)}$, and the multifidelity Gaussian process prior $GP(0, \kappa^{(l)}(\boldsymbol{x},\boldsymbol{x}'))$. The algorithm computes an initial subset of feasible combinations of optimization variables $\{\boldsymbol{x}_n\}_{n=1}^{N_0}$ and associated levels of fidelity $\{l_n\}_{n=1}^{N_0}$, where

N_0 is the number of points sampled at the beginning of the optimization. Contextually, observations of the objective function $\{f^{(l_n)}\}_{n=1}^{N_0}$ are collected and adopted to determine the initial state of the MFBO dynamic system $\mathscr{D}_0 = \{\boldsymbol{x}_n, f^{(l_n)}, l_n\}_{n=1}^{N_0}$. The multifidelity GP prior together with the initial state $\mathscr{D}_0$ induce the multifidelity Gaussian process posterior with mean $\mu^{(l)}$ and variance $\sigma^{2(l)}$: this represents the first approximation of the objective function based on the initial observations.

Let us now consider the generic iteration i of the optimization search. The NM2-BO algorithm selects the domain location $\boldsymbol{x}_i$ along with the associated level of fidelity l_i at the previous step $i-1$, and evaluates the objective function $f^{(l_i)}(\boldsymbol{x}_i)$. This information is used to augment the state $\mathscr{D}_i = \mathscr{D}_{i-1} \cup \{\boldsymbol{x}_i, f^{(l_i)}, l_i\}$ and update the multifidelity Gaussian process prediction conditioned on $\mathscr{D}_i$.

At the same iteration i, we compute the two-step lookahead multifidelity acquisition function to determine the next combination of optimization variables $\boldsymbol{x}_{i+1}$ to evaluate and the associated representation of the objective function l_{i+1} to query. We indicate with $z = i$ the current step of the optimization and with $z+1$ and $z+2$ the first and second step ahead, respectively. The first term of the acquisition function (Equation (4.9)) is computed on the current state $\mathscr{D}_z = \mathscr{D}_i$, while the second term is estimated using the Monte-Carlo approach discussed in Section 4.3. Considering the j-th Monte Carlo simulation, the NM2-BO procedure samples independently the normally distributed random variable Z_j, and estimates the mean $\mu_{s+1}^{(l)}$ and variance $\sigma_{s+1}^{2(l)}$ of the multifidelity GP computing Equation (4.11) and Equation (4.12). Then, we evaluate the second term of the acquisition function through Equation (4.13) and compute $\widehat{U}_{z_j}^{\pi^*}$. The Monte Carlo algorithm iterates till a maximum number of simulations N_{MC} is reached and returns the realizations of the approximated acquisition function $\{\widehat{U}_{z_j}^{\pi^*}\}_{j=1}^{N_{MC}}$. This allows to estimate the two-step multifidelity acquisition function U_i as the expectation taken over the Monte Carlo simulations.

At this point, we maximize $U_i^{\pi^*}$ and determine the next combination of optimization variables to evaluate $\boldsymbol{x}_{i+1}$ and the associated level of fidelity l_{i+1}. The NM2-BO algorithm is iterated until a maximum computational budget $B_i = B_{max}$ is reached, where B_i is the cumulative computational cost expended until iteration i.

4.5 Enabling Multiple Decisions

The non-myopic multifidelity sampling scheme defined through Equation (4.9) sequentially identifies the most promising combination of optimization variables and the respective level of fidelity that maximizes the cumulative informative gains over two steps ahead. However, complex multidisciplinary design optimization problems open major challenges for the intrinsic demand to scale: the required accurate evaluations of the objective function can dramatically upscale during the search for improved design solutions.

To address this complex multiphysics optimization scenario, we extend the optimal policy $\boldsymbol{\pi}^*$ to enable multiple decisions in a single iteration of the optimization procedure. Accordingly, this policy defines a decision making process where a batch of informative combinations of optimization variables and associated levels of fidelity $\mathscr{B}_i^{n_b} = [(\boldsymbol{x}_{i,1}, l_{i,1}), ..., (\boldsymbol{x}_{i,n_b}, l_{i,n_b})]$ are iteratively selected while improving the optimization solutions over future iterations. The potential of multipoint formulations has been illustrated by [206] for a greedy single-fidelity Bayesian framework, and motivates our proposal of a multipoint sampling strategy for the non-myopic MFBO in multidisciplinary settings.

Accordingly, the formulation of our non-myopic multifidelity acquisition function $U_z^{\boldsymbol{\pi}^*}$ is extended through a local penalization maximization strategy as follows:

$$\boldsymbol{x}_{i,k}, l_{i,k} = \max \left[U_z^{\boldsymbol{\pi}^*}(\boldsymbol{x}_{z+2}, l_{z+2}, \mathscr{D}_{z+1}) \prod_{j=1}^{k-1} \psi(\boldsymbol{x}, \boldsymbol{x}_j) \right] \tag{4.14}$$

where $(\boldsymbol{x}_{i,k}, l_{i,k}) \in \mathscr{B}_i^{n_b}$, and ψ is the local penalty function which quantifies the probability that a point in the optimization domain $\boldsymbol{x}$ is a potential minimum not belonging to the hypersphere $\{\boldsymbol{x} \in \mathscr{X} : ||\boldsymbol{x}_j - \boldsymbol{x}|| \leq (\hat{f}^* - f^L(\boldsymbol{x}_j))/L\}$:

$$\psi(\boldsymbol{x}, \boldsymbol{x}_j) = \frac{1}{2} erfc \left[\frac{1}{\sqrt{2\sigma^{2(L)}(\boldsymbol{x}_j)}} \left(L||\boldsymbol{x}_j - \boldsymbol{x}|| - \hat{f}^* + \mu^{(L)}(\boldsymbol{x}_j) \right) \right] \tag{4.15}$$

where $erfc$ is the complementary error function, $\hat{f}^* = \min_{\boldsymbol{x} \in \mathscr{X}} \mu^{(L)}(\boldsymbol{x})$ is the minimum predicted by the surrogate model, and $L = \max_{\boldsymbol{x} \in \mathscr{X}} ||\mu_{\nabla}^{(L)}(\boldsymbol{x})||$ is the

Gaussian process Lipschitz constant [206] defined as the maximum of the surrogate gradient.

The formalization of the penalty function ψ allows to locally penalize the non-myopic acquisition function and creates exclusion zones whose amplitude is determined by the Lipschitz constant L. Thus, larger exclusion zones are created in regions of the search domain where the mean function $\mu^{(L)}$ is far from the predicted minimum $\hat{f}^*$. Conversely, small sized exclusion zones are determined by values of the mean function $\mu^{(L)}$ closer to the predicted minimum $\hat{f}^*$. The objective of this penalization strategy is to mimic an adaptive sampling scheme over multiple iterations that would have been achieved by a sequential scheme considering available the previous evaluations of the combinations of optimization variables in the batch.

In addition, we provide an adaptive batch size formulation that identifies the number of domain points in a batch $n_b(i) = 1 + \varpi/(\sqrt{2}i)$ as a function of the optimization iterations i and the initial batch size ϖ. This strategy targets the efficient use of computational resources during the optimization procedure: the number of evaluations of the objective function increases at the beginning of the optimization to improve the awareness about the distribution of the unknown objective function over the entire domain, and is progressively reduced to catalyze the resources toward the analyses of believed optimal solutions.

4.6 NM3-BO Algorithm

Algorithm 2 illustrates the numerical implementation of the Non-Myopic Multipoint Multifidelity Bayesian Optimization (NM3-BO) scheme. The computations are initialized through the same approach described in Section 4.4 for the NM2-BO algorithm.

For a generic iteration i of the NM3-BO algorithmic flow, the surrogate model is updated through the collected observations of the objective function $f^{(l_{i,k})}(\boldsymbol{x}_{i,k})$ at each $n_b(i)$ pair of combination of optimization variables $\boldsymbol{x}_{i,k}$ and levels of fidelity $l_{i,k}$ that constitutes the batch $\mathscr{B}_i^{n_b(i)}$ selected at the previous iteration $i-1$. This represents the updated state of the MFBO dynamic system $\mathscr{D}_i = \mathscr{D}_{i-1} \cup \{\boldsymbol{x}_{i,k}, f^{(l_{i,k})}(\boldsymbol{x}_{i,k}), l_{i,k}\}_{k=1}^{n_b(i)}$. At this stage, the NM3-BO algorithm computes and maximizes the non-myopic multipoint acquisition function $U_z^{\boldsymbol{\pi}^*}$ to select next batch of

Algorithm 2 NM3-BO: Non-Myopic Multipoint Multifidelity Bayesian Optimization

```
Input: Design space $\mathscr{X} \in \mathbb{R}^D$, multifidelity models of the objective function $f^{(l)}(\boldsymbol{x})$
    and the multifidelity Gaussian process prior $GP(0, \kappa^{(l)}(\boldsymbol{x}, \boldsymbol{x}'))$
Output: $f^* = \min f(\boldsymbol{x})$
$\mathscr{D}_0 \leftarrow \{\boldsymbol{x}_n, f^{(l_n)}, l_n\}_{n=1}^{N_0}$ collect initial observations
$\mu_0^{(l)}, \sigma_0^{2(l)} \leftarrow$ learn the initial multifidelity GP
$i \leftarrow 1$
repeat
    Load the $k$ pairs of $\boldsymbol{x}_{i,k}$ and associated $l_{i,k}$ from the batch $\mathscr{B}_{i-1}^{n_b(i-1)}$
    Evaluate $f^{(l_{i,k})}(\boldsymbol{x}_{i,k})\ \forall n_b(i-1)$
    $\mathscr{D}_i = \mathscr{D}_{i-1} \cup \{\boldsymbol{x}_{i,k}, f^{(l_{i,k})}(\boldsymbol{x}_{i,k}), l_{i,k}\}_{k=1}^{n_b(i)}$
    $\mu_i^{(l)}, \sigma_i^{2(l)} \leftarrow$ update the multifidelity Gaussian process
    $z \leftarrow i$
    Compute $U_{MFEI}(\boldsymbol{x}_{z+1}, l_{z+1})$
    for $j \leftarrow 1, N_{MC}$ do
        $Z_j \leftarrow \mathscr{N}(0,1)$
        $\mu_{z+1}^{(l)}, \sigma_{z+1}^{2(l)} \leftarrow$ estimate the multifidelity GP
        Compute $\widehat{U}_{MFEI}(\boldsymbol{x}_{z+2}, l_{z+2}, Z_j)$
    end for
    return $\widehat{\boldsymbol{U}}_z^{\boldsymbol{\pi}^*} = \{\widehat{U}_{z_j}^{\boldsymbol{\pi}^*}\}_{j=1}^{N_{MC}}$
    $U_i^{\boldsymbol{\pi}^*} = \mathbb{E}\left[\widehat{\boldsymbol{U}}_z^{\boldsymbol{\pi}^*}\right]$
    for $k \leftarrow 1, n_b(i)$ do
        $\boldsymbol{x}_{i,k}, l_{i,k} = \max\left[U_z^{\boldsymbol{\pi}^*}(\boldsymbol{x}_{z+2}, l_{z+2}, \mathscr{D}_{z+1}) \prod_{j=1}^{k-1} \psi(\boldsymbol{x}, \boldsymbol{x}_j)\right]$
    end for
    return $\mathscr{B}_i^{n_b(i)} = [(\boldsymbol{x}_{i,1}, l_{i,1}), ..., (\boldsymbol{x}_{i,n_b(i)}, l_{i,n_b(i)})]$
    $i + 1 \leftarrow i$
until $B_i \leq B_{max}$
return $\boldsymbol{x}^*$ that minimize $f(\boldsymbol{x})$ over $\mathscr{D}_i$
```

domain points and model of the objective function $\mathscr{B}_{i+1}^{n_b(i+1)}$ to evaluate. Let now indicate with $z = i$ the current step of the optimization and with $z+1$ and $z+2$ the first and the second step ahead, respectively. The first element of $U_z^{\boldsymbol{\pi}^*}$ is determined using the information extracted from the surrogate model updated at the current state of the MFBO system $\mathscr{D}_z = \mathscr{D}_i$. The second element requires our Monte Carlo technique to approximate the intractable nested expectation and maximization, and measure the informative reward at future iterations. Similarly to the NM2-BO algorithm, NM3-BO samples independently a random variable Z_j normally distributed for the j-th Monte Carlo realization, and simulates the future optimization scenario through the estimate of the mean $\mu_{z+1}^{(l)}$ and variance $\sigma_{z+1}^{2(l)}$ of the surrogate model by the computation of Equation (4.11) and Equation (4.12). This provides an approximation of the multifidelity acquisition function $\widehat{U}_{z_j}^{\boldsymbol{\pi}^*}$ as the expectation taken over the realizations $\{\widehat{U}_{z_j}^{\boldsymbol{\pi}^*}\}_{j=1}^{N_{MC}}$. Then, the penalized maximization of the acquisition function determines the next batch $\mathscr{B}_{i+1}^{n_b(i+1)}$ of design configurations and the levels of fidelity to be evaluated in parallel at the next iteration. This optimization procedure iterates until a maximum computational budget $B_i = B_{max}$ is reached, where B_i is the cumulative computational cost adopted until iteration i.

4.7 Numerical Experiments

In this section, we illustrate and discuss the proposed NM2-BO algorithm in comparison with popular MFBO approaches over a set of well accepted benchmark problems specifically conceived to stress-test multifidelity methods [207, 208]. This permits to investigate the performance of the NM2-BO framework over challenging mathematical properties of the objective function frequently encountered in real-word applications, such as global and local non-linearity, discontinuity, multimodality, oscillatory behaviours, and noise in the measurements. As competing algorithms, we adopt and implement three standard MFBO frameworks based on multifidelity expected improvement (MFEI) [38], multifidelity probability of improvement (MFPI) [41], and multifidelity max-value entropy search (MFMES) [42]. All the competing algorithms are initialized adopting the same initial settings: we define the number of points $n_0^{(l)}$ for each l-th fidelity level that characterize the initial dataset $\mathscr{D}_0$ determined through a Latin Hypercube Sampling [209]. In addition, we set the maximum computational budget B_{max} for each benchmark problem computed as the sum of the computational costs associated with observations of the objective function at the l-th level of fidelity. The initial evaluations of the objective function at different levels of fidelity are used to compute the first surrogate model as depicted in Section 4.4. We use the square exponential kernels for all the multifidelity GP covariances, and optimize the hyperparameters of the kernel and mean functions of the multifidelity GP via Maximum Likelihood Estimation [210].

Function	$n_0^{(4)}$	$n_0^{(3)}$	$n_0^{(2)}$	$n_0^{(1)}$	B_{max}	$\lambda^{(4)}$	$\lambda^{(3)}$	$\lambda^{(2)}$	$\lambda^{(1)}$
Forrester	1	2	3	5	100	1	0.5	0.1	0.05
Jump Forrester	-	-	2	5	100	-	-	1	0.2
Rastrigin	-	10	20	30	200	-	1	0.065	0.0039
ALOS 1D	-	-	2	5	100	-	-	1	0.2
ALOS 2D	-	-	5	10	30	-	-	1	0.2
ALOS 3D	-	-	7	14	300	-	-	1	0.2
Rosenbrock 2D	-	-	5	10	200	-	-	1	0.5
Rosenbrock 5D	-	-	15	30	500	-	-	1	0.5
Rosenbrock 10D	-	-	50	250	1000	-	-	1	0.5
Paciorek	-	-	5	10	200	-	-	1	0.2
Mass Spring	-	-	4	10	400	-	-	1	1/60
Borehole	-	-	100	500	800	-	-	1	0.5

Table 4.1 Summary of the experiments setup

4.7.1 Analytical Benchmark Problems

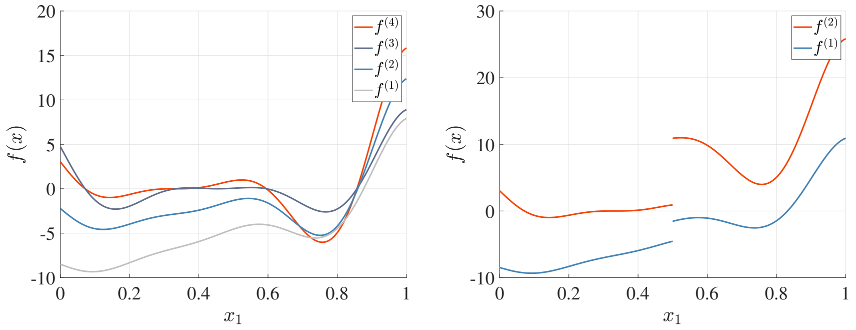
The mathematical formalization of the benchmark problems is illustrated in the following and includes the Forrester function standard and discontinuous, the Rastrigin function shifted and rotated, the ALOS functions, the Rosenbrock function, the Paciorek function, a mass spring system optimization problem, and the Borehole function. This set of functions is specifically selected to emulate mathematical characteristics that are frequent in physics-based optimization problems, keeping contained the computational expense associated with their evaluation [207, 208]. Table 4.1 summarizes the experimental setup for each test function.

Forrester Function

The Forrester function [210] is a one-dimensional multimodal function well accepted as a benchmark test for multifidelity algorithms:

$$f^{(4)}(\boldsymbol{x}) = (6\boldsymbol{x}-2)^2\sin(12\boldsymbol{x}-4) \tag{4.16}$$

The optimization problem is defined over the domain $\mathscr{X} = [0,1]$ with the minimum of the objective located at $\boldsymbol{x}^* = 0.7572$ and given by $f^* = -6.0207$. We define a total of three lower-fidelity levels as follows:



(a) Forrester (b) Jump Forrester

Fig. 4.5 Forrester function benchmark problems

$$f^{(3)}(\boldsymbol{x}) = (5.5\boldsymbol{x} - 2.5)^2 \sin(12\boldsymbol{x} - 4) \tag{4.17}$$

$$f^{(2)}(\boldsymbol{x}) = 0.75 f_1(\boldsymbol{x}) + 5(\boldsymbol{x} - 0.5) - 2 \tag{4.18}$$

$$f^{(1)}(\boldsymbol{x}) = 0.5 f_1(\boldsymbol{x}) + 10(\boldsymbol{x} - 0.5) - 5. \tag{4.19}$$

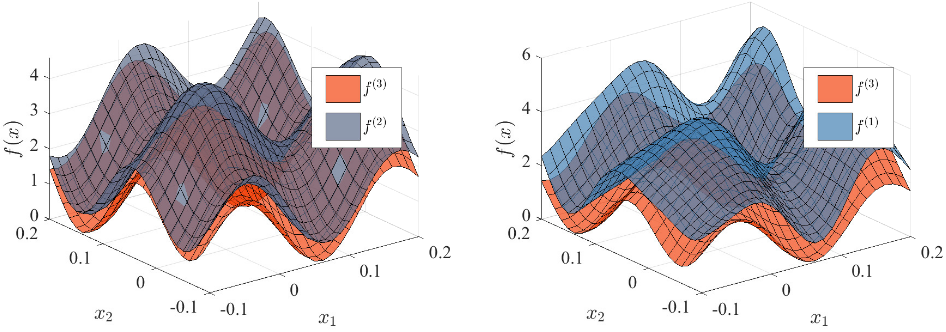
Figure 4.5(a) reports the four levels of fidelity for the Forrester function over the search domain.

Jump Forrester Function

The jump Forrester function [211] problem introduces a discontinuous behaviour in the formulation of the Forrester function in order to assess the performance of the algorithms in presence of objective functions that are not continuous in all the input space. The high-fidelity jump Forrester is defined as follows:

$$f^{(2)}(\boldsymbol{x}) = \begin{cases} (6\boldsymbol{x} - 2)^2 \sin(12\boldsymbol{x} - 4) & 0 \leq \boldsymbol{x} \leq 0.5 \\ (6\boldsymbol{x} - 2)^2 \sin(12\boldsymbol{x} - 4) + 10 & 0.5 < \boldsymbol{x} \leq 1 \end{cases} \tag{4.20}$$

The input ranges is defined as $\mathscr{X} = [0,1]$ and the optimum equal to $f^* = -0.9863$ at $\boldsymbol{x}^* = 0.1426$. The low-fidelity function is formulated as:



(a) Rastrigin $f^{(3)}$, $f^{(2)}$ (b) Rastrigin $f^{(3)}$, $f^{(1)}$

Fig. 4.6 Rastrigin function shifted and rotated benchmark problem

$$f^{(1)}(\boldsymbol{x}) = \begin{cases} 0.5f^{(2)}(\boldsymbol{x}) + 10(\boldsymbol{x}-0.5)-5 & 0 \leq \boldsymbol{x} \leq 0.5 \\ 0.5f^{(2)}(\boldsymbol{x}) + 10(\boldsymbol{x}-0.5)-2 & 0.5 < \boldsymbol{x} \leq 1 \end{cases} \tag{4.21}$$

Figure 4.5(b) reports the two levels of fidelity available during the search procedure.

Rastrigin Function Shifted and Rotated

The Rastrigin function is commonly used as test function to represent real-world applications where the objective function might present an high multimodal behaviour. We adopt a benchmark problem based on the original formulation of the Rastrigin function shifted and rotated as follows (Figure 4.6):

$$f^{(l)}(\boldsymbol{x}) = v_2^2 + v_1^2 - \cos(10\pi v_2) - \cos(10\pi v_1) + 2 \tag{4.22}$$

where

$$\boldsymbol{v} = R(\vartheta)(\boldsymbol{x} - \boldsymbol{x}^*) \tag{4.23}$$

with

$$R(\vartheta) = \begin{bmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{bmatrix} \tag{4.24}$$

where R is the rotation matrix, $\vartheta = 0.2$ is the rotation angle. The optimization problem is defined in $\mathscr{X} = [-0.1, 0.2]^2$ and the minimum is located at $\boldsymbol{x} = [0.1, 0.1]^2$ given by $f^* = 0$ To obtain the levels of fidelity, we define a resolution error [212]:

$$e(v, \varphi) = \sum_{i=1}^{2} \theta(\varphi) \cos^2(10\pi\Theta(\varphi)v_i + 0.5\pi\Theta(\varphi) + \pi) \tag{4.25}$$

where $\Theta(\varphi) = 1 - 0.0001\varphi$. We assign the value $\varphi = 10000$ for the high-fidelity model ($l = 3$), $\varphi = 5000$ for the intermediate level of fidelity ($l = 2$), and $\varphi = 2500$ for the low fidelity model ($l = 1$).

ALOS functions

The Agglomeration of Locally Optimized Surrogate (ALOS) models are a set of functions conceived to evaluate the capability of multifidelity optimization in presence of objective function characterized by oscillatory phenomena at different frequency along the domain. The high-fidelity ALOS functions are defined for one and two dimensions [213], and extended to three dimensions as follows [207]:

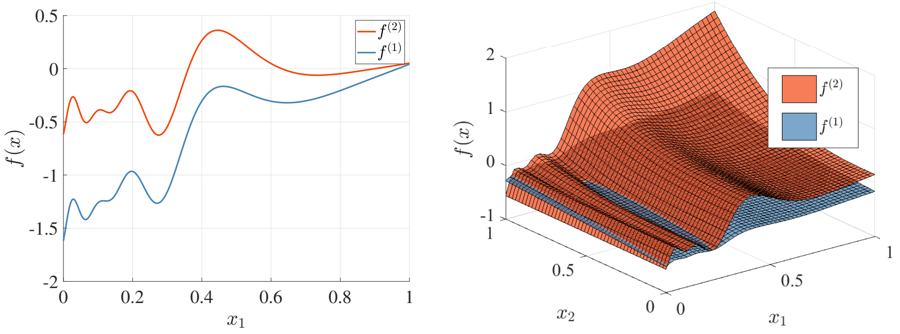
$$f^{(2)}(x_1) = \sin[30(x_1 - 0.9)^4] \cos[2(x_1 - 0.9)] + (x_1 - 0.9)/2 \tag{4.26}$$

$$f^{(2)}(x_1, x_2) = f^{(2)}(x_1) + 2x_2^2 \sin(x_1 x_2) \tag{4.27}$$

$$f^{(2)}(x_1, x_2, x_3) = f^{(2)}(x_1, x_2) + 3x_3^3 \sin(x_1 x_2 x_3) \tag{4.28}$$

The domain of the input is defined as the interval $[0, 1]^D$ where $D = 1, 2, 3$ is the dimensionality of the problem. The optimum is $f^* = -0.62489$ located at $\boldsymbol{x}^* = 0.2746$ for $D = 1$, $f^* = -0.56271$ corresponding to the input $\boldsymbol{x}^* = [0, ..., 0]^D$ for $D = 2, 3$. The low-fidelity ALOS functions are derived through bridge functions:

$$f^{(1)}(x_1) = (f^{(1)}(x_1) - 1 + x_1)/(1 + 0.25x_1) \tag{4.29}$$



(a) ALOS $D = 1$ (b) ALOS $D = 2$

Fig. 4.7 ALOS function benchmark problems over the $D = 1$ and $D = 2$ dimensional domain

$$f^{(1)}(x_1,x_2) = (f^{(1)}(x_1,x_2) - 2 + x_1 + x_2)/(5 + 0.25x_1 + 0.5x_2) \tag{4.30}$$

$$f^{(1)}(x_1,x_2,x_3) = (f^{(1)}(x_1,x_2,x_3) - 2 + x_1 + x_2 + x_3)/(5 + 0.25x_1 + 0.5x_2 - 0.75x_3) \tag{4.31}$$

Figure 4.7 illustrates the high and low-fidelity ALOS function for $D = 1$ (Figure 4.7(a)) and $D = 2$ (Figure 4.7(b)).

Rosenbrock Function

The Rosenbrock function [214] is a popular non-convex D-dimensional optimization benchmark problem defined in the interval $\mathscr{X} = [-2, 2]^D$ as follows:

$$f^{(2)}(\boldsymbol{x}) = \sum_{i=1}^{D-1} 100(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i^2)^2 + (1 - \boldsymbol{x}_i)^2 \tag{4.32}$$

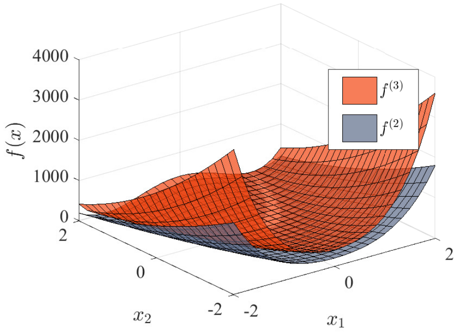


Fig. 4.8 Rosenbrock function benchmark problem over the $D = 2$ dimensional domain

The minimum of Equation (4.32) is located at $\boldsymbol{x}^* = [1,...,1]^D$ corresponding to $f^* = 0$. We consider as the low fidelity level the linear mapping defined as follows:

$$f^{(1)}(\boldsymbol{x}) = \sum_{i=1}^{D-1} 50(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i^2)^2 + (-2 - \boldsymbol{x}_i)^2 - \sum_{i=1}^{D} 0.5\boldsymbol{x}_i \tag{4.33}$$

In the experiments, we consider $D = 2, 5, 10$ to investigate the performances of the algorithms as the dimension of the design space increases. Figure 4.8 illustrates the library of two levels of fidelity for the $D = 2$ Rosenbrock function.

Paciorek Function with Noise

This multifidelity problem considers the Paciorek function [215] with a noise term as the high-fidelity model, and uses a noisy low-fidelity representation that permits to simulate real-world applications where the measurements of the objective function are affected by noise [207]. The noisy high-fidelity Paciorek function is defined as (Figure 4.9):

$$f^{(2)}(\boldsymbol{x}) = \sin\left(\frac{1}{\boldsymbol{x}_1\boldsymbol{x}_2}\right) + rand.normal(0, 0.0125) \tag{4.34}$$

where a uniformly distributed random noise parameter is included to simulate the noise. Across the input domain $\mathscr{X} = [0.3, 1]^2$ the optimum is equal to $f^* = -1$ and located in a belt of equal-minimum input points. The noisy low-fidelity function is formulated as follows:

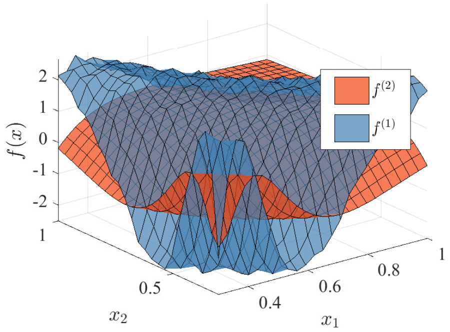


Fig. 4.9 Paciorek function benchmark problem

$$f^{(1)}(\boldsymbol{x}) = f^{(2)}(\boldsymbol{x}) - 2.25\cos\left(\frac{1}{\boldsymbol{x}_1\boldsymbol{x}_2}\right) + rand.normal(0, 0.075) \tag{4.35}$$

Mass Spring System

This benchmark problem consists of a coupled mass spring system, where two masses are connected by two springs. We consider a friction-less dynamic along an horizontal plane, where the masses are treated as point masses concentrated at their center of gravity and the elastic force of the springs is modeled through the Hooke's law. Assuming that the system is connected to an ideal wall through the first spring, its equations of motion are given by:

$$m_1\ddot{p}_1(t) = -k_1 p_1(t) + k_2[p_2(t) - p_1(t)] \tag{4.36a}$$
$$m_2\ddot{p}_2(t) = -k_2[p_2(t) - p_1(t)] - k_1 p_2(t) \tag{4.36b}$$

where m_1, m_2 are the masses, k_1, k_2 are the Hooke's constants of the springs and $p_1(t), p_2(t)$ are the positions of the masses as a function of time t. We use the fourth-order Runge-Kutta method to solve the system of first-order ODEs (4.36), setting the time-step size $dt = 0.01$ for the high-fidelity and $dt = 0.6$ for the low-fidelity analysis. The optimization task consists of determining the input variable

$\boldsymbol{x} = [m_1, m_2, k_1, k_2]$ that minimizes $p_1(t = 6)$ over the design space $\mathscr{X} = [1,4]^4$, with initial conditions of motion $p_1 = 1, p_2 = 0, \dot{p}_1 = \dot{p}_2 = 0$.

Borehole Function

The Borehole function models the water flows through a borehole. We adopt the following multifidelity setting to compute the water flow rate [216]:

$$f^{(2)}(\boldsymbol{x}) = \frac{2\pi x_3(x_4 - x_6)}{\log(x_2/x_1)\left(1 + \frac{2x_7x_3}{\log(x_2/x_1)x_1^2x_8} + \frac{x_3}{x_5}\right)} \tag{4.37}$$

$$f^{(1)}(\boldsymbol{x}) = \frac{5x_3(x_4 - x_6)}{\log(x_2/x_1)\left(1.5 + \frac{2x_7x_3}{\log(x_2/x_1)x_1^2x_8} + \frac{x_3}{x_5}\right)} \tag{4.38}$$

where $x_1 \in [0.05, 0.15]$ is the radius of the borehole, $x_2 \in [100, 50000]$ is the radius of influence, $x_3 \in [63070, 115600]$ is the trasmissivity of upper aquifer, $x_4 \in [990, 1110]$ is the potentiometric head of lower aquifer, $x_5 \in [63.1, 116]$ is the trasmissivity of lower aquifer, $x_6 \in [700, 820]$ is the potentiometric head of lower aquifer, $x_7 \in [1120, 1680]$ is the lenght of the borehole, and $x_8 \in [9855, 12045]$ is the hydraulic conductivity of the borehole. The minimum of the Equation (4.37) is $f^* = 9.1884\, m^3/yr$ corresponding to the design $\boldsymbol{x}^* = [0.0504, 3584, 87130, 1007, 112.3, 813.2, 1.658, 9.858]$.

4.7.2 Results and Discussion

In the following, the results are observed and discussed in terms of difference error normalized in the domain of the objective function:

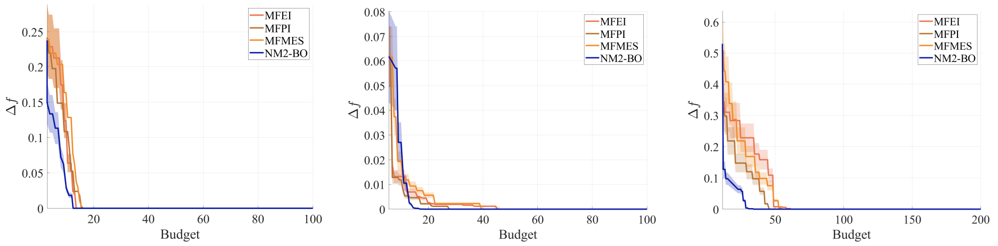
$$\Delta f = \frac{f(\boldsymbol{x}_i^*) - f^*}{f_{max} - f^*} \tag{4.39}$$

where $f(\boldsymbol{x}_i^*)$ is the minimum value of the objective function identified by the algorithm at the i-th iteration, f^* is the analytical solution of the optimization problem and $f_{max} = \max_{\boldsymbol{x} \in \mathscr{X}} f(\boldsymbol{x})$. The error Δf is represented as a function of the computational Budget B to compare the optimization results achieved by the competing algorithms. To quantify and compensate the influence of the random initial design of experiments, we run the optimization tasks 25 times considering a random initialization for each benchmark problem, and reported the result in terms of median values of the error together with the values falling in the interval between the 25-th and 75-th percentiles.

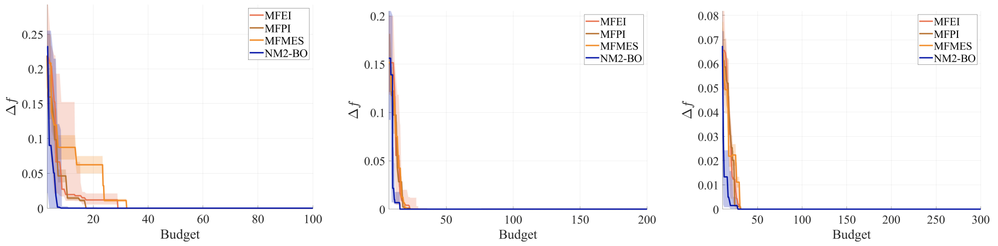
Figure 4.10 illustrates the results achieved with the competing multifidelity Bayesian optimization algorithms for the proposed set of analytical benchmark problems (Section 4.7.1). The outcomes reveal that the proposed Non-Myopic Multifidelity Bayesian Optimization (NM2-BO) achieves superior optimization performance both in terms of reduction of the error and resource efficiency, and outperforms the baseline competing multifidelity Bayesian frameworks (MFEI-MFPI-MFMES) for all the benchmark problems.

The results achieved for the Forrester function (Figure 4.10(a)) considering a library of four levels of fidelity suggest the acceleration of the convergence to the optimum solution achieved by NM2-BO ($\Delta f = 0$) with a significant reduction of the computational budget if compared with the baseline MFBO algorithms. In addition, the capabilities of the proposed algorithm are confirmed for the case of the Jump Forrester problem (Figure 4.10(b)). Indeed, the non-myopic property of the NM2-BO algorithm permits to identify the optimum with a fraction of the allocated computational budget, while the competing algorithms demand for significantly higher computational resources: MFPI requires almost double the resource expenditure and both MFMES and MFEI demand for more than triple the cost of NM2-BO to converge.

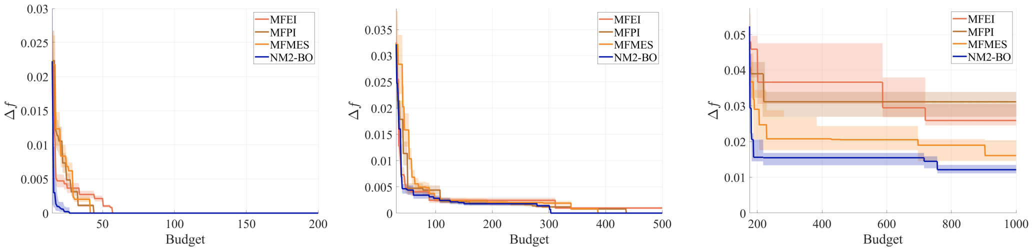
Another remarkable outcome is the performance of our non-myopic multifidelity framework for the optimization of the highly multimodal Rastrigin function (Figure 4.10(c)) considering a total of three levels of fidelity. We note a step-like reduction of the error with the computational budget: this behaviour occurs when the optimizer identifies a local minimum of the multimodal objective function that leads to a constant error until the search procedure identifies a new minimum. It is possible to observe that the myopic baseline MFBO methods are characterized by an emphasized step-like convergence of the error, while is almost not appreciable for the NM2-BO convergence history. This suggests that the non-myopic property improves the convergence performance in presence of marked multimodality of the objective function. A possible explanation for these outcomes might be attributed to an efficient form of exploration depicted in Section 4.2, where the observation of



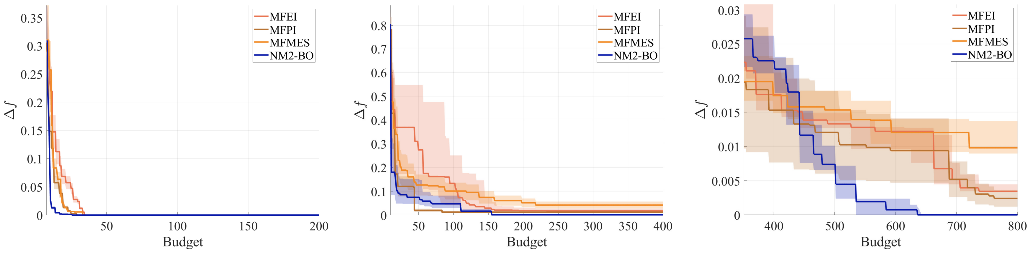
(a) Forrester (b) Jump Forrester (c) Rastrigin



(d) ALOS 1D (e) ALOS 2D (f) ALOS 3D



(g) Rosenbrock 2D (h) Rosenbrock 5D (i) Rosenbrock 10D



(j) Paciorek (k) Mass spring system (l) Borehole

Fig. 4.10 Error of the minimum of the objective function obtained with the NM2-BO algorithm and compared to the baseline MFBO algorithms over the analytical benchmark problems.

high-uncertainty regions of the search domain improves future evaluations and leads to an effective exploitation toward the optimum. This is in agreement with what observed by Ginsbourger and LeRiche [134] for their non-myopic single-fidelity optimization algorithm.

The outcomes achieved for the ALOS benchmark problem confirm the substantial acceleration of the optimization convergence adopting our NM2-BO algorithm, which outperforms the standard MFBO frameworks for the one- (Figure 4.10(d)), two- (Figure 4.10(e)) and three- (Figure 4.10(f)) dimensional ALOS problem. The performance of the non-myopic algorithm might be explained with the computation of an efficient multifidelity Gaussian process surrogate model that effectively assists and informs the non-myopic search toward the optimization goal even in presence of objective functions with oscillatory nature at different frequencies.

The experiments for the Rosenbrock function are conducted increasing the number of the optimization variables ($D = 2, 5, 10$) to investigate the capabilities of the competing algorithms for various dimensionality of the search domain. Considering the case of $D = 2$ optimization variables (Figure 4.10(g)), NM2-BO permits to identify the optimum of the objective function with a fraction of the computational cost demanded by the standard MFBO algorithms. This confirms the substantial accelerations of the optimization procedure through non-myopic multifidelity policies. Increasing the number of optimization variables $D = 5$ (Figure 4.10(h)), MFEI identifies sub-optimal solutions with an error $\Delta f > 0$ while the other multifidelity methods are capable to minimize the error and identify the optimum solution. However, our NM2-BO algorithm exhibits the lower demand for computational resources to accurately identify the optimum. For an higher dimensionality of the domain $D = 10$ (Figure 4.10(i)), all the algorithms fail to identify the optimum of the Rosenbrock function, and converge to sub-optimal solutions. This undesirable performance is due to the difficulties of the multifidelity Gaussian process to accurately represent high dimensional objective functions. After the initial stages, the error decreases only with large computational expense for both the non-myopic and the standard MFBO schemes. We observed that the algorithms here mostly query the low-fidelity model to contain the computational expense during a secondary exploration phase.

The results achieved for the Paciorek benchmark function (Figure 4.10(j)) show that the NM2-BO algorithm provides superior accelerations of the optimization procedure even in presence of noise in measurements if compared with the competing algorithms. In addition, the superior performance of NM2-BO is confirmed also for the optimization problem of the mass spring system (Figure 4.10(k)) and of the Borehole function (Figure 4.10(l)), where the non-myopic property permits to obtain results consistent with the one obtained for the other benchmark functions.

The experiments on the analytical set of benchmark problems reveal the capabilities of the proposed non-myopic multifidelity strategy: NM2-BO leads to overall better performance and accelerations with respect to the standard multifidelity settings over a variety of challenging mathematical properties of the objective function. In particular, those results demonstrate the remarkable impact on the search performance obtained maximizing the reward over two steps ahead of the optimization procedure: the non-myopic optimizer effectively explores the search domain with contained computational resources and leverages the acquired resources to direct resources for the local exploitation toward the optimum.

Chapter 5

Physics-Aware Multifidelity Bayesian Optimization

In this section, we propose and formalize an original Physics-Aware Multifidelity Bayesian Optimization (PA-MFBO) framework that incorporates prior scientific and expert knowledge about the domain of the objective function to improve and accelerate the optimization search in multifidelity settings. The goal is to partially alleviate the purely data-driven search of standard Bayesian frameworks from the characterization of the domain structure on the fly while searching for the optimal combination of optimization variables that minimizes the objective function.

The proposed PA-MFBO framework relies on a physics-aware multifidelity acquisition function conceived to wisely combine (i) data-driven information extracted from the multifidelity surrogate model of the objective function and (ii) prior knowledge about the structure of the domain encapsulated during the search through a physics-aware utility function. This formalizes a sort of physics-aware search that introduces a bias during the learning procedure and influences the selection of specific numerical models – from the library of models at different levels of fidelity – targeting the balance between the computational cost required to evaluate the objective function and the accuracy demanded for specific combinations of optimization variables along the optimization search.

In addition, we extend the proposed physics-aware multifidelity acquisition function through the non-myopic formulation proposed in Chapter 3: this permits to formalize an acquisition function that is capable to (i) unlock the improvement of the solution quality obtainable through a coordinated action across future evaluations through the lookahead multifidelity property, and (ii) select the appropriate level of fidelity of the objective function model combining data-driven and physics-based information through the physics-aware formulation.

The next section is dedicated to the formalization of the physics-aware multifidelity acquisition function (Section 5.1). Then, building on the non-myopic formulation developed in Chapter 3, we formalize the non-myopic physics-aware multifidelity acquisition function in Section 5.2. This chapter is based on the articles [217, 147, 146, 218, 219].

5.1 Physics-Aware Multifidelity Acquisition Function

The physics-aware multifidelity Bayesian optimization framework incorporates the prior scientific and expert knowledge about the physical phenomena and the specific structure of the optimization domain through the original physics-aware multifidelity acquisition function $U_{PA}(\boldsymbol{x},l)$ based on the multifidelity expected improvement [38]:

$$U_{PA}(\boldsymbol{x},l) = U_{EI}(\boldsymbol{x})\alpha_1(\boldsymbol{x},l)\alpha_2(\boldsymbol{x},l)\alpha_3(l)\alpha_4(\boldsymbol{\psi},l) \tag{5.1}$$

where $U_{EI}(\boldsymbol{x})$ is the expected improvement acquisition function evaluated at the highest level of fidelity [220]:

$$U_{EI}(\boldsymbol{x}) = \sigma(\boldsymbol{x})(I(\boldsymbol{x})\Phi(I(\boldsymbol{x}))) + \mathcal{N}(I(\boldsymbol{x});0,1) \tag{5.2}$$

where $I(\boldsymbol{x}) = (f^{(L)}(\hat{\boldsymbol{x}}^*) - \mu(\boldsymbol{x}))/\sigma(\boldsymbol{x})$ is the predicted improvement, $\hat{\boldsymbol{x}}^*$ is the current location of the best value of the objective sampled so far, $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution, while the terms α_1, α_2 and α_3 are formalized as follows:

$$\alpha_1(\boldsymbol{x},l) = corr\left[f^{(l)}, f^{(L)}\right] = \frac{\kappa((\boldsymbol{x},l),(\boldsymbol{x},L))}{\sqrt{\sigma^{2(l)}\sigma^{2(L)}}} \tag{5.3}$$

$$\alpha_2(\boldsymbol{x},l) = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma^{2(l)}(\boldsymbol{x}) + \sigma_\varepsilon^2}} \tag{5.4}$$

$$\alpha_3(l) = \frac{\lambda^{(L)}}{\lambda^{(l)}}. \tag{5.5}$$

α_1 is defined as the posterior correlation coefficient between the l-th level of fidelity and the highest-fidelity available at the same location of the domain. This utility function reflects the reduction of the acquisition function when samples are evaluated with lower-fidelity models, and accounts for the decrease of the accuracy associated with a low-fidelity representation of the objective function. Accordingly, the use of a high-fidelity model is solicited when a low-fidelity estimate might produce unreliable observations of the objective function. α_2 considers the reduction of the uncertainty associated with the prediction of the multifidelity Gaussian process after a new observations of the objective function with a certain level of fidelity is added to the dataset $\mathscr{D}_N$. This permits to consider the stochastic nature of the objective function during the optimization, and prevents the systematic sampling in already explored regions of the domain characterized by lower uncertainty. α_3 is formulated as the ratio between the computational cost $\lambda^{(L)}$ associated with the evaluation of the high-fidelity model and the computational cost $\lambda^{(l)}$ required to compute the l-th fidelity model. This utility function is conceived to include awareness about the computational resources required for the evaluation of the objective function adopting the l-th level of fidelity. The purpose of this term is to privilege the selection of lower-fidelity queries when similar improvements of the solution are obtained from higher-fidelity observations, and balance the computational cost and the informative contribution of different fidelity levels.

$\alpha_4(\boldsymbol{\psi}, l)$ is the physics-aware utility function that embeds a source of prior knowledge represented by a set of physical variables $\boldsymbol{\psi}$ relevant for the specific optimization problem of interest. The goal of this element of to introduce a learning bias that captures the scientific knowledge and expertise underlying the physical domain of the system. The formulation of the physics-aware utility function is conceived to incorporate any form of prior knowledge about the optimization problem. Without limiting the informative content that can be embedded during the search procedure, we identify two major sources of prior knowledge in the form of scientific and expert knowledge. Scientific knowledge refers to the general body of rules formalized and validated adopting the scientific method such as conservation laws, physical principles or phenomenological behaviors that depict the physics of interest. Examples include the Navier-Stokes partial differential equations in fluid dynamics to model the motion of viscous fluids, and numerical methodologies to approximate the solution of the governing equations as the finite element method in structural mechanics. Expert knowledge represents the information that is held by a community of experienced specialists and validated implicitly over several years of experience in a specific field. Examples include the common knowledge within the engineering or physics community resulting from training, research and personal experience.

5.2 Toward Physics-Aware Non-Myopic Multifidelity Bayesian Optimization

In this section, we define an optimal optimization policy which permits to include the awareness about the physical domain of the system of interest while maximizing the long-term reward over two steps ahead of the optimization procedure. The proposed approach naturally combines the non-myopic formalization of the multifidelity acquisition function presented in Section 4.2 with the physics-aware utility function illustrated in Section 5.1. We refer to this multifidelity framework as Physics-Aware Non-Myopic Multifidelity Bayesian Optimization (PA-NM2BO).

Following the dynamic programming recursive strategy formalized in Section 4.2, we define the physics-aware expected reward at the generic step z:

$$\begin{aligned} J_z^{\boldsymbol{\pi}}(\boldsymbol{x}_{z+1}, l_{z+1}, \mathscr{D}_z) = & \\ & \mathbb{E}[r_z(\boldsymbol{x}_{z+1}, y^{(l_{z+1})}, l_{z+1}, \mathscr{D}_z) + \\ & J_{z+1}(\mathscr{F}(\boldsymbol{x}_{z+1}, y^{(l_{z+1})}, l_{z+1}, \mathscr{D}_z))] \end{aligned} \tag{5.6}$$

where $\mathbb{E}[r_z(\cdot)] = U_{PA}(\boldsymbol{x}_{z+1}, l_{z+1}, \boldsymbol{\psi}_{z+1})$ is now the physics-aware multifidelity acquisition function, and $J_{z+1}(\mathscr{F}(\cdot))$ is the long-term expected reward. Accordingly, we formalize the two-step lookahead physics-aware multifidelity acquisition function through an optimal policy $\boldsymbol{\pi}^*$ that maximizes the cumulative expected reward over two-step ahead of a pair of combination of optimization variables $\boldsymbol{x}_{z+2}$ and level of fidelity l_{z+2}:

$$\begin{aligned} U_z^{\boldsymbol{\pi}^*}(\boldsymbol{x}_{z+2}, l_{z+2}, \mathscr{D}_{z+1}) = & \\ & U_{PA}(\boldsymbol{x}_{z+1}, l_{z+1}, \boldsymbol{\psi}_{z+1}) + \\ & \mathbb{E}\left[\max(U_{PA}(\boldsymbol{x}_{z+2}, l_{z+2}, \boldsymbol{\psi}_{z+2}))\right] \end{aligned} \tag{5.7}$$

where we define the long term reward $J_{z+1} = \max(U_{PA}(\boldsymbol{x}_{z+2}, l_{z+2}))$ as the maximum of the physics-aware multifidelity acquisition function conditioned on the dataset $\mathscr{D}_{z+1}$.

Chapter 6

FREEDOM: Fast Reliability Estimate and Incipient Fault Detection of Multiphysics Systems

In this section, we propose and formalize a novel computational framework for Fault Detection and Isolation (FDI) to accelerate the identification of incipient damages of complex multiphysics systems. We aim to obtain an efficient procedure capable to monitor the health status of complex systems and innovative technologies onboard to potentially ease the transition towards the adoption of greener technologies. We name and refer to our framework as FREEDOM as the short for Fast REliability Estimate and incipient fault Detection Of Multiphysics aerospace systems. The main features of the FREEDOM computational framework are: i) a Bayesian scheme for the inference of damage reduces the demand for costly evaluations of numerical models of the system of interest, ii) a novel two-stage highly informative compression strategy reduces the dimensionality of the diagnostics signals while retaining a high-quality informative content, and iii) the original combination of both Bayesian inference and two-stage compression accelerates the identification of incipient damages affecting the system: the highly informative compressed diagnostic signals are leveraged during the Bayesian inference procedure to reduce the dimensionality of the fault detection and isolation problem to accurately identify damage parameters with contained computational resources.

FREEDOM is conceived to adopt an unique monitoring model of the system of interest with a fixed level of fidelity. The testing and validation reported in Section 8.2.5 reveal that FREEDOM achieves satisfactory performances in terms of the accuracy of the incipient damages inference adopting low-fidelity monitoring models. For the case of an high-fidelity monitoring model, FREEDOM identifies the exact health status of the system with a total computational cost not suitable for on-board monitoring of innovative technologies. Indeed, the effectiveness of the FDI procedure closely relates to the accuracy of this monitoring model and the associated monitoring signal. In principle, the optimal setting is achieved adopting a monitoring model capable to exactly represent the dynamical behaviour of the damaged system. This would guarantee identical output and monitoring signals when the damaged configuration of the system is identified; conversely, modeling discrepancies between reference and monitoring signals might determine errors in the identification process and reduced accuracy of the health status identification, as observed in [221]. However, this favourable condition is usually unfeasible in practice: the acquisition of accurate monitoring signals requires the evaluation of expensive high-fidelity numerical models that heavily impact the computational burden of the diagnosis. Moreover, the identification of incipient damages – which is essential for safety critical systems – further amplifies the demand for high-fidelity simulations to accurately capture the almost imperceptible alterations of the system behaviour during operations.

To address this scenario, we extend the FREEDOM algorithm and formalize a multifidelity FREEDOM computational framework that embeds a library of monitoring models at different levels of fidelity. This permits to efficiently incorporate high-fidelity monitoring models in model-based FDI procedures for complex multiphysics systems, where the reliable identification of complex faults affecting innovative components might be hindered by a single-fidelity Bayesian inference scheme. Multifidelity FREEDOM is based on the original combination of i) the two-stage highly informative compression also adopted in the FREEDOM algorithm to contain the dimensionality of the FDI procedure, and ii) a multifidelity Bayesian scheme for inversion leverages and combines multiple models of the system at different levels of fidelity to effectively identify the health status of the system. Thus, multifidelity FREEDOM includes high-fidelity simulations during the inference stage and aims at improving the effectiveness of the diagnostic procedure; at the same time, lower-fidelity monitoring models permits to alleviate the overall computational cost of the procedure and more efficiently explore potential faults affecting the system.

In the remaining of this chapter, Section 6.1 introduces the proposed framework FREEDOM including the two-stage compression strategy (Section 6.1.1) and the Bayesian scheme for inference (Section 6.1.2). Then, we formalize the multifidelity FREEDOM computational framework in Section 6.2 and detail the multifidelity Bayesian inference procedure for damage identification (Section 6.2.1). This chapter is based on the articles [222–224, 221].

6.1 FREEDOM: Methodology and Computational Framework

This section illustrates FREEDOM as a methodology and computational framework for fault detection and isolation to accelerate the identification of incipient damages affecting complex multiphysics systems. The main methodological procedures consist of: i) an offline two-stage compression strategy to compute a synthesized representation of the diagnostics signals – encoding the dynamics of the system of interest – and reduce the amount of data to be stored and processed during the inference of damages, and ii) an online inference process based on a Bayesian optimization scheme to accurately infer the health status of the system containing the evaluations of numerical models and computational expense.

The FREEDOM computational framework is schematically illustrated through the diagram in Figure 6.1. Let us consider a generic complex system characterized by a certain output signal $\boldsymbol{y}(\boldsymbol{x},\boldsymbol{m})$ of n_e elements that encodes its dynamics and is sensitive to its health status $\boldsymbol{x} = [x_1,...,x_{n_k}]$ defined as a combination of n_k faults modes, and measured in certain locations $\boldsymbol{m}$. The output signal is typically measured with an high acquisition frequency to guarantee the sensitivity to the damages affecting the dynamics of the system.

In the offline phase, we compute a reference dataset using a scaled Latin Hypercube sampling process [225] to determine a set of n_s health conditions $\boldsymbol{X} = [\boldsymbol{x}_1,...,\boldsymbol{x}_{n_s}]$ that are used to compute the output signals $\boldsymbol{Y} = \left[\boldsymbol{y}_0(\boldsymbol{x}_0,\boldsymbol{m}),...,\boldsymbol{y}_{n_s}(\boldsymbol{x}_{n_s},\boldsymbol{m})\right]$. This scaled design of experiments methodology augments the density of the sampled health conditions close to the nominal condition to enhance the informative content

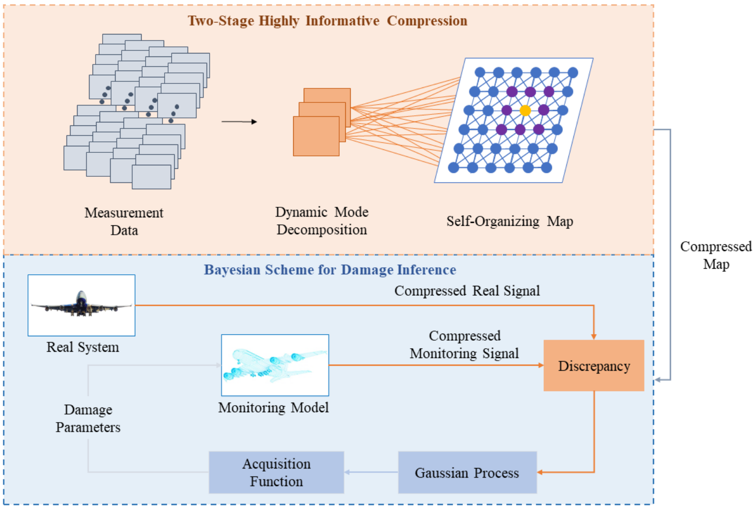


Fig. 6.1 Schematic representation of the proposed FREEDOM framework.

about the behaviour of the system under incipient damages. The dataset $\boldsymbol{Y}$ is obtained by evaluating a high-fidelity accurate model or through a real-world test bench of the system, and is considered the ground truth reference dataset of the system. Those information are usually high-dimensional and demand for intensive computations to be stored and processed during FDI. Therefore, we propose an original two-step signal compression based on two projection methods to determine a synthesized representation of the signal $\hat{\boldsymbol{y}}(\boldsymbol{x},\hat{\boldsymbol{m}})$ and retain only the highly informative $n_w << n_e$ data.

Online, the fault detection and isolation procedure is based on a Bayesian scheme for damage inference to identify the health status of the system. This is formalized as an inverse problem where the damage condition affecting the system minimizes the discrepancy between the output signal of the real system and the signal computed with a fast monitoring model. In this stage, we use the compressed informative map $(\hat{\boldsymbol{m}},\hat{\boldsymbol{y}})$ to place measurement points and reduce the dimension of the output signal. Specifically, the Bayesian routine uses the Gaussian process surrogate model (Section 3.3.1) to approximate the discrepancy between the real-world measurements and numerical responses. This predictive framework informs an adaptive sampling scheme realized by the acquisition function that quantifies the utility of querying the fault domain, and directs the computational resources towards the health assessment of the system.

The following Sections describe in detail the two main elements that characterize FREEDOM. Specifically, Section 6.1.1 deals with the original two steps signal compression strategy, and Section 6.1.2 illustrates the FDI procedure adopting the Bayesian scheme for inference.

6.1.1 Two-Stage Highly Informative Compression Map

A two-stage highly informative compression technique is proposed to be conducted offline and consists of two main stages: i) Dynamic Mode Decomposition (DMD) is used to identify a set of dominant coherent structures explaining the dynamic behaviour of the system, and ii) a Self Organizing Map (SOM) is adopted to project those coherent structures into a lower-dimensional space and compute a map that encodes the overall dynamic of the system. The procedure is inspired by the two-stage compression proposed by Mainini [226] and Mainini and Willcox [227] where Proper Orthogonal Decomposition and Self Organizing Maps jointly identify the optimal sensor placement in structural health monitoring applications, and applied by Berri et al. [225] for the FDI of aircraft electromechanical actuators.

Dynamic Mode Decomposition

The first stage of our compression strategy adopts the Dynamic Mode Decomposition (DMD) technique to extract the dynamical features of the system in the form of dynamic modes [228, 229]. From a general perspective, a dynamical system affected by damages could be considered as a non-linear system ψ whose output signal $\boldsymbol{y}(\boldsymbol{x},\boldsymbol{m}) \in \mathbb{R}^{n_e}$ is sensitive to the health status $\boldsymbol{x}$ of the system and to the measurement locations $\boldsymbol{m}$. DMD seeks to identify the dynamical properties of this system as the dominant eigenvalues and eigenvectors of the informative matrix $\tilde{\boldsymbol{A}} \in \mathbb{R}^{n_e \times n_e}$ such that:

$$\tilde{\boldsymbol{A}}\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{Y}}' \tag{6.1}$$

where

$$\tilde{\boldsymbol{Y}} = [\tilde{\boldsymbol{y}}(\boldsymbol{x}_0,\boldsymbol{m}) \quad \tilde{\boldsymbol{y}}(\boldsymbol{x}_1,\boldsymbol{m}) \quad \dots \quad \tilde{\boldsymbol{y}}(\boldsymbol{x}_{n_s-1},\boldsymbol{m})] \in \mathbb{R}^{n_e \times n_s}$$

$$\tilde{\boldsymbol{Y}}' = [\tilde{\boldsymbol{y}}(\boldsymbol{x}_1,\boldsymbol{m}) \quad \tilde{\boldsymbol{y}}(\boldsymbol{x}_2,\boldsymbol{m}) \quad \dots \quad \tilde{\boldsymbol{y}}(\boldsymbol{x}_{n_s},\boldsymbol{m})] \in \mathbb{R}^{n_e \times n_s}$$

are the snapshot matrices. Those snapshots are computed offline through an high-fidelity representation of the system ψ, and collect a series of n_s paired high-fidelity output signals $\tilde{\boldsymbol{y}}(\boldsymbol{x}_{j-1},\boldsymbol{m})$ and $\tilde{\boldsymbol{y}}(\boldsymbol{x}_j,\boldsymbol{m})$ computed for incipient fault conditions $\{\boldsymbol{x}_j\}_{j=1}^{j=n_s}$ assembled via scaled Latin hypercube sampling scheme [225].

The computation of the informative matrix $\tilde{\boldsymbol{A}}$ is usually unfeasible: the direct solution of Equation (6.1) is hampered by the significant computational cost required for large snapshot matrices – frequently encountered in many real-world applications in science and engineering. A popular approach to overcome this issue adopts the Singular Value Decomposition (SVD) technique [230] to identify a set of $n_w << n_s$ informative modes $\boldsymbol{B} \in \mathbb{C}^{n_e \times n_s}$. This permits to retain a large fraction of information embedded in the snapshot matrices, and quantifies this informative content as the cumulative sum $\sum_{i=1}^{n_w} \lambda_i / \sum_{i=1}^{n_s} \lambda_i$ of the eigenvalues λ associated to the considered modes. As a result, the snapshot matrices $\tilde{\boldsymbol{Y}}$ and $\tilde{\boldsymbol{Y}}'$ can be projected onto the first n_w modes, and compute the matrices $\boldsymbol{Y}$ and $\boldsymbol{Y}'$:

$$\boldsymbol{Y} = \boldsymbol{B}^* \tilde{\boldsymbol{Y}} \in \mathbb{R}^{n_w \times n_s} \,, \quad \boldsymbol{Y}' = \boldsymbol{B}^* \tilde{\boldsymbol{Y}}' \in \mathbb{R}^{n_w \times n_s} \tag{6.2}$$

where $\boldsymbol{B}^*$ is the Hermitian transpose of $\boldsymbol{B}$. Accordingly, the DMD problem (Equation (6.1)) becomes manageable and the information matrix can be computed as follows:

$$\boldsymbol{A} = \boldsymbol{Y}' \boldsymbol{Y}^+ \tag{6.3}$$

where $\boldsymbol{Y}^+$ is the Moore-Penrose pseudoinverse of $\boldsymbol{Y}$. The eigendecomposition of the informative matrix $\boldsymbol{A}$ yields the dominant eigenvalues and eigenvectors that encode the fundamental properties of the underlying dynamical system in the form of dynamic modes $\boldsymbol{\Upsilon} = [\boldsymbol{\upsilon}_1, \dots, \boldsymbol{\upsilon}_{n_w}]$.

Self Organizing Map

In the second stage of compression, a Self Organizing Map (SOM) projects the DMD modes $\boldsymbol{\Upsilon}$ into a lower-dimensional space, and computes an efficient compression map of the output signal $\hat{\boldsymbol{y}}(\boldsymbol{x},\hat{\boldsymbol{m}})$ that preserves the topological properties of those modes. This representation retains only the highly informative elements $n_w << n_e$ of the output signal, and is used in the online phase to alleviate the computational burden of the FDI procedure.

Self-organizing map (SOM) is a competitive artificial neural network trained with unsupervised learning paradigms to produce a low-dimensional representation of the input space of the training samples [231, 232]. Considering a non-linear damaged system ψ, the training set $\boldsymbol{T}$ consists of the dynamic modes of the system $\boldsymbol{\Upsilon}$ and the measurement locations $\boldsymbol{m}$ of the output signal:

$$\boldsymbol{T} = [\boldsymbol{m}, \boldsymbol{\upsilon}_1, ..., \boldsymbol{\upsilon}_{n_w}] \tag{6.4}$$

During the SOM training process, each row $\boldsymbol{\tau}_i$ of $\boldsymbol{T}$ is presented to the network and the node that shows greater similarity to this input is selected to be the winner. The weight of this winning node is updated towards the input vector, and the weights of the other nodes are influenced according to their topological distances from the winner. In particular, the winning neuron is selected as the one that minimizes the distance between the associated weight vector $\boldsymbol{w}_l$ and the current training point $\boldsymbol{\tau}_i$:

$$l = \arg\min_j \left(||\boldsymbol{\tau}_i - \boldsymbol{w}_j|| \right) \tag{6.5}$$

where $||\cdot||$ denotes the L2 norm.

In the final layout of a trained SOM, adjacent nodes have an high similarity to each other and are dissimilar to nodes that are located far apart. The resulting net represents a non-linear projection of the training set $\boldsymbol{T}$ to the lower dimensional space of the neurons, and the weight vectors in the input space encode representative vectors for clusters of self-similar points [233, 226, 227]. This permits to extract the latent structure of the training input and compute the efficient compression map $\hat{\boldsymbol{y}}(\boldsymbol{x},\hat{\boldsymbol{m}})$.

6.1.2 Bayesian Scheme for Damage Inference

The goal of the online FDI stage is to infer the combination of fault parameters $\boldsymbol{x}^*$ that minimizes the discrepancy f between the output signal $\boldsymbol{y}(\boldsymbol{x}^*,\boldsymbol{m})$ measured in the

Algorithm 3 Bayesian scheme for the inference of damage parameters

```
Input: Definition of the faults domain $\mathscr{X} \in \mathbb{R}^D$, discrepancy function $f(\boldsymbol{x},\hat{\boldsymbol{m}})$ and
    the Gaussian process surrogate model prior $GP(0,\kappa(\boldsymbol{x},\boldsymbol{x}'))$
Output: Actual damage configuration affecting the system $\boldsymbol{x}^* = \arg\min f(\boldsymbol{x},\hat{\boldsymbol{m}})$
$\mathscr{D}_0 \leftarrow \{\boldsymbol{x}_n, f(\boldsymbol{x}_n,\hat{\boldsymbol{m}})\}_{n=1}^{N_0}$ collect initial $N_0$ noisy observations of the discrepancy
    function $f(\boldsymbol{x}_n) \sim \mathscr{N}(f(\boldsymbol{x}_n,\hat{\boldsymbol{m}}),\sigma_\varepsilon)$
$\mu_0,\sigma_0 \leftarrow$ compute the initial mean and standard deviation of the Gaussian
    process surrogate model
$i \leftarrow 1$
repeat
    Load the new combination of damage parameters $\boldsymbol{x}_{N_i}$
    Compute the observation of the discrepancy function $f(\boldsymbol{x}_{N_i},\hat{\boldsymbol{m}})$
    $\mathscr{D}_i \leftarrow \mathscr{D}_{i-1} \cup \{\boldsymbol{x}_{N_i}, f(\boldsymbol{x}_{N_i},\hat{\boldsymbol{m}})\}$ update the dataset of observations
    $\mu_i,\sigma_i \leftarrow$ update the mean and standard deviation of the Gaussian process
        surrogate model
    Compute the acquisition function $U(\boldsymbol{x} \mid \mathscr{D}_i)$ on the updated dataset $\mathscr{D}_i$
    Maximize the acquisition function to select the next damage configuration to
        query $\boldsymbol{x}_{N_{i+1}} = \arg\max_{\boldsymbol{x}\in\mathscr{X}} U(\boldsymbol{x} \mid \mathscr{D}_i)$
    $i+1 \leftarrow i$
until Convergence criteria is met
return Combination of damage parameters $\boldsymbol{x}^*$ that minimizes the discrepancy
    function $f(\boldsymbol{x}_n,\hat{\boldsymbol{m}})$ over the faults domain $\mathscr{X}$
```

locations $\boldsymbol{m}$ from the physical system and the same signal $\boldsymbol{y}_M(\boldsymbol{x},\boldsymbol{m})$ computed with a numerical model. This problem may be challenging to be solved with a contained computational cost since the output signals $\boldsymbol{y}$ and $\boldsymbol{y}_M$ encode a significant number of measurements. To reduce the dimensionality of the problem, we leverage the compression mask $\hat{\boldsymbol{y}}(\boldsymbol{x},\hat{\boldsymbol{m}})$ computed with our two-step compression strategy and formalize the FDI task as follows:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}\in\mathscr{X}}{\arg\min} f(\boldsymbol{x},\hat{\boldsymbol{m}}) \tag{6.6}$$

where $f(\boldsymbol{x},\hat{\boldsymbol{m}}) = ||\hat{\boldsymbol{y}}(\boldsymbol{x}^*,\hat{\boldsymbol{m}}) - \hat{\boldsymbol{y}_M}(\boldsymbol{x},\hat{\boldsymbol{m}})||$.

To solve this identification problem, we use a Bayesian optimization scheme based on the Gaussian process surrogate model to approximate the discrepancy function f, and on an acquisition function to iteratively select the combination of fault parameters that is likely to reduce – ideally minimize – the objective/discrepancy function. Algorithm 3 illustrates the main steps of the inverse Bayesian FDI framework. BO sequentially collects noisy observations of f and learns the probabilistic Gaussian process surrogate model to quantify the current belief about the discrepancy between the response of the real system and the outcome of the monitoring model. At this point, the acquisition function uses the surrogate information to measure the utility of making any given evaluation of f. The process iterates until a maximum number of evaluations of the monitoring model are reached to contain the total computational expense.

6.2 Multifidelity FREEDOM: Methodology and Computational Framework

This section extends the FREEDOM methodology to a multifidelity setting where a library of monitoring models of the system is available during the online FDI procedure. Accordingly, we propose and formalize the multifidelity FREEDOM computational strategy to efficiently combine data from multiple representations of the monitoring signal: fast low-fidelity models are adopted to improve the awareness about possible damage configurations, and contained high-fidelity evaluations are used to improve the accuracy of the identified health status of the system. The objective is to efficiently include high-fidelity simulations during the FDI without negatively affect the overall computational cost of the process.

In particular, the multifidelity FREEDOM methodology combines two constitutive phases: the first one aims to optimize the informative content of the diagnostic signals minimizing the computational burden associated with their management through the two-stage highly informative compression strategy presented in Section 6.1.1); the second one seeks to efficiently include high-fidelity models to support the accurate inference of incipient faults through a multifidelity Bayesian scheme for inference (Section 6.2.1).

Figure 6.2 illustrates the main features of our multifidelity fault detection and isolation framework. Similarly to the single-fidelity FREEDOM algorithm, the procedure starts collecting offline a reference dataset of measurements obtained computing the output signal $\boldsymbol{y}(\boldsymbol{x},\boldsymbol{m})$ for different incipient fault conditions $\{\boldsymbol{x}_j\}_{j=1}^{j=n_s}$ determined through a scaled Latin hypercube sampling scheme [225]. This permits

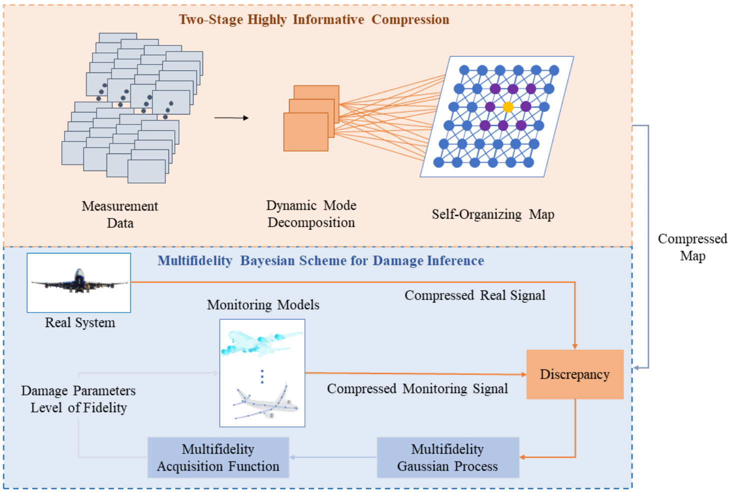


Fig. 6.2 Schematic representation of the proposed multifidelity FDI framework.

to increase the probability distribution of samples near the nominal condition and augment the informative content about the behaviour of the system under incipient damages. The dataset is collected evaluating an high-fidelity accurate representation of the system – e.g. accurate numerical model or physical test-bench of the system – and is considered as the ground truth reference data.

The output signal $\boldsymbol{y}$ is typically high-dimensional and unpractical to store and process, which results in a computationally intensive FDI task unfeasible for onboard applications. The first phase of the proposed strategy adopts the two-stage compression strategy (Section 6.1.1) applied offline on the reference dataset to determine a compressed map $\hat{\boldsymbol{y}}(\boldsymbol{x},\hat{\boldsymbol{m}})$ that retains only the highly informative elements of the original signal.

The second phase is the online damage inference; this step attempts to identify the specific health status of the system $\boldsymbol{x}^*$ with contained computational resources, and aims at including expensive high-fidelity simulations to improve the overall accuracy of the inference procedure. During the online phase, the informative map computed offline $(\hat{\boldsymbol{m}},\hat{\boldsymbol{y}})$ is used to reduce the dimensionality of the FDI problem and improve the efficiency of the inference step. We consider a spectrum of representations of the output monitoring signal at different levels of fidelity $\{\boldsymbol{y}_M^{(1)},\dots,\boldsymbol{y}_M^{(L)}\}$, where the higher the level of fidelity l the more accurate and costly the acquisition of the signal $\boldsymbol{y}_M^{(l)}$. A multifidelity Bayesian scheme for damage inference is adopted to solve the fault detection and isolation task. This multifidelity technique approximates the discrepancy f between the real and monitoring signals using a multifidelity Gaussian process surrogate model (Section 3.3.2); then, the multifidelity acquisition function is computed using the prediction of the surrogate, and iteratively selects the combination of fault parameters and level of fidelity that is likely to minimize f and identify the health status of the system.

6.2.1 Multifidelity Bayesian Scheme for Damage Inference

The informative compressed map $\hat{\boldsymbol{y}}(\boldsymbol{x},\hat{\boldsymbol{m}})$ computed with the two-stage compression strategy (Section 6.1.1) is used to reduce the dimensionality of the FDI problem:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}\in\mathscr{X}}{\arg\min} f^{(l)}(\boldsymbol{x},\hat{\boldsymbol{m}}) \tag{6.7}$$

where $f^{(l)}(\boldsymbol{x},\hat{\boldsymbol{m}}) = ||\hat{\boldsymbol{y}}(\boldsymbol{x}^*,\hat{\boldsymbol{m}}) - \hat{\boldsymbol{y}_M}^{(l)}(\boldsymbol{x},\hat{\boldsymbol{m}})||$ is now evaluated only for the n_w informative elements of the reference compressed signal $\hat{\boldsymbol{y}}$ and monitoring compressed output $\hat{\boldsymbol{y}_M}$. To address this inverse problem, we adopt a multifidelity Bayesian scheme to leverage queries of the discrepancy function $[f^{(1)}, f^{(2)}, ..., f^{(L)}]$ at different levels of fidelity and accelerate the inference of the damages affecting the system [33, 234, 71, 37].

Algorithm 4 describes the main methodological stages of the multifidelity FDI. The damage inference procedure is dynamically conducted online through an iterative process and relies on two key elements: the multifidelity Gaussian process surrogate model and the multifidelity acquisition function. The surrogate approximates the discrepancy function based on collected information, and provides a predictive framework that synthesizes the data from multifidelity models into a unique emulator. At each iteration, the acquisition function informed by this predictive distribution is maximized to select the damage configuration that is likely to actually affect the system together with the associated fidelity of the representation to query. Meanwhile, the new observation is used to update the posterior distribution of the emulator, and the above process is repeated until a convergence metric is met.

Algorithm 4 Multifidelity Bayesian scheme for the inference of damage parameters

```
Input: Definition of the faults domain $\mathscr{X} \in \mathbb{R}^D$, discrepancy function $f^{(l)}(\boldsymbol{x},\hat{\boldsymbol{m}})$, the multifidelity Gaussian process surrogate model prior $GP(0,\kappa^{(l)}(\boldsymbol{x},\boldsymbol{x}'))$, and the library of monitoring models of the discrepancy function at different levels of fidelity $[f^{(1)},f^{(2)},...,f^{(L)}]$
Output: Actual damage configuration affecting the system $\boldsymbol{x}^* = \arg\min f^{(L)}(\boldsymbol{x},\hat{\boldsymbol{m}})$
$\mathscr{D}_0 \leftarrow \{\boldsymbol{x}_n, f^{(l_n)}(\boldsymbol{x}_n,\hat{\boldsymbol{m}})\}_{n=1}^{N_0}$ collect initial $N_0$ noisy observations of the discrepancy function $f^{(l_n)}(\boldsymbol{x}_n,\hat{\boldsymbol{m}}) \sim \mathscr{N}(f^{(l_n)}(\boldsymbol{x}_n,\hat{\boldsymbol{m}}),\sigma_\varepsilon)$ at different levels of fidelity
$\mu_0^{(l)},\sigma_0^{(l)} \leftarrow$ compute the initial mean and standard deviation of the multifidelity Gaussian process surrogate model
$i \leftarrow 1$
repeat
    Load the new combination of damage parameters $\boldsymbol{x}_{N_i}$ and associated level of fidelity $l_{N_i}$
    Compute the observation of the discrepancy function $f^{(l_{N_i})}(\boldsymbol{x}_{N_i},\hat{\boldsymbol{m}})$
    $\mathscr{D}_i \leftarrow \mathscr{D}_{i-1} \cup \{\boldsymbol{x}_{N_i}, f^{(l_{N_i})}(\boldsymbol{x}_{N_i},\hat{\boldsymbol{m}})\}$ update the dataset of observations
    $\mu_i^{(l)},\sigma_i^{(l)} \leftarrow$ update the mean and standard deviation of the multifidelity Gaussian process surrogate model
    Compute the multifidelity acquisition function $U(\boldsymbol{x} \mid \mathscr{D}_i)$ on the updated dataset $\mathscr{D}_i$
    Maximize the multifidelity acquisition function to select the next damage configuration and associated level of fidelity to query $\boldsymbol{x}_{N_{i+1}}, l_{N_{i+1}} = \arg\max_{\boldsymbol{x}\in\mathscr{X}} U(\boldsymbol{x} \mid \mathscr{D}_i)$
    $i+1 \leftarrow i$
until Convergence criteria is met
return Combination of damage parameters $\boldsymbol{x}^*$ that minimizes the discrepancy function $f(\boldsymbol{x}_n,\hat{\boldsymbol{m}})$ over the faults domain $\mathscr{X}$
```

Chapter 7

Experiments and Discussion: Design Applications

In this chapter, we investigate the performance of the proposed algorithms for aerospace engineering design optimization problems. In particular:

- Section 7.1 investigates the capabilities of the NM2-BO methodology for an aerodynamic shape optimization problems.
- Section 7.2 demonstrates the optimization performance of the PA-MFBO algorithm for a cross-regime aerodynamic shape optimization problem.
- Section 7.3 illustrates the results achieved with the PA-NM2BO, PA-MFBO, and NM2-BO for a cross-regime aerodynamic design optimization problem.
- Section 7.4 demonstrates the capabilities of the NM3-BO and PA-MFBO algorithms for the multidisciplinary design optimization problem of a re-entry space vehicle.

7.1 Aerodynamic Shape Optimization Problem

This section illustrates and discusses the results achieved with the Non-Myopic Multifidelity Bayesian Optimization (NM2-BO) framework (Chapter 4) for an aerodynamic shape design optimization problem. This design optimization problem has been specifically selected and conceived by an international community of experts in multifidelity methods to stress-test multifidelity algorithms over challenging engineering test-cases [235].

Specifically, this aerodynamic shape optimization problems aims at identify the optimized airfoil geometrical shape that minimizes the drag coefficient C_d of a RAE 2822 airfoil, subject to maintaining a constant lift coefficient C_l. Additional constraints include the pitching momentum C_m, the flight conditions, and the geometry of the airfoil. The geometry of the RAE 2822 airfoil is modified through the code WG2AER developed by [236], where the initial geometry of the airfoil is linearly combined with shape modification functions through the weights $\boldsymbol{w}$. Specifically, we use six shape modification polynomial functions and assign the weights $\boldsymbol{w} = [w_1, ..., w_6]$ at each function to modify the shape of the RAE 2822 airfoil. The design optimization problem is formulated as follows:

$$
\begin{aligned}
\min_{\boldsymbol{x}\in\mathscr{X}} \quad & C_d(\boldsymbol{x}) \\
& \boldsymbol{x} = [w_1, ..., w_6] && \text{(7.1a)} \\
s.t. \quad & C_l = 0.824 && \text{(7.1b)} \\
& -0.1 \leq C_m \leq -0.01 && \text{(7.1c)} \\
& M = 0.734 && \text{(7.1d)} \\
& h = 10000m && \text{(7.1e)} \\
& t/c = 0.1211 && \text{(7.1f)} \\
& r \geq 0.007c && \text{(7.1g)} \\
& \tau \geq 5^\circ && \text{(7.1h)} \\
& t_{85}/c \geq 0.02 && \text{(7.1i)} \\
& \mathscr{X} = [-1,1]^6 && \text{(7.1j)}
\end{aligned}
$$

where M is the flight Mach number and h is the flight altitude, t represents the thickness of the airfoil, c is the chord, τ is the trailing edge angle, and t_{85} is the thickness of the airfoil at the 85% of the chord. The search for the optimal airfoil shape aims at identifying the combination of design variables, namely the weights of the shape modification functions $\boldsymbol{w}$, that minimizes the drag coefficient $C_d(\boldsymbol{x})$ over the design space $\mathscr{X}$ defined by the move limits for the weights $I_w = [-1,1]^6$. The constraint on the lift coefficient is satisfied considering the angle of attack α as a free parameter in the CFD simulations and the constraint on the maximum thickness is automatically satisfied scaling the airfoil after the geometry modifications. The remaining constraints are considered formalizing the corresponding unconstrained optimization problem through the quadratic penalty functions as follows:

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathscr{X}} C_d + c_1 g^+(C_m,-0.1) + c_1 g^-(C_m,-0.01) + \\ c_2 g^+(t/c, 0.007) + c_3 g^+(\tau, 5^\circ) + c_4 g^+(t_{85}/c, 0.02)\end{aligned} \tag{7.2}$$

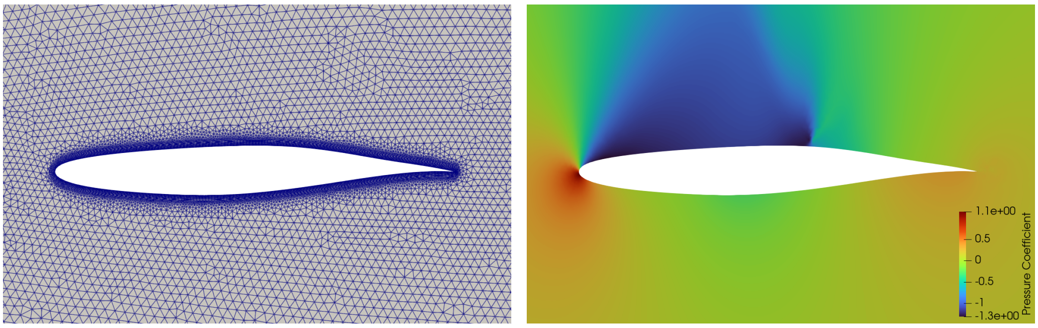
where g^+ and g^- are the quadratic penalty functions [236], and $c_1 = 1000$, $c_2 = 5000$, $c_3 = 10$, and $c_4 = 30$.

7.1.1 Aerodynamic Models

The flow-field modeling consists of two levels of fidelity that are assembled using an open-source computational tool for aerodynamic analysis [236]. Both the aerodynamic models take in input the airfoil geometry modified through the assignment of the weights $\boldsymbol{w}$ together with the flight altitude h, the Reynolds number Re and the free-stream Mach number M, and return the drag coefficient C_d and the pitching moment coefficient C_m. We assume the flow-field in a fully turbulent condition and model the aerodynamic domain through the Reynolds Averaged Navier-Stokes (RANS) equations. The software numerically solves the RANS equations using a finite volume method, with an automatic procedure to generate a specific mesh for each modified shape of the RAE 2822. The differential formulation of RANS is mathematically expressed as follows:

$$\boldsymbol{R}(\boldsymbol{U}) = \frac{\partial(\boldsymbol{U})}{\partial t} + \nabla\cdot\boldsymbol{F}^c - \nabla\cdot\boldsymbol{F}^v - \boldsymbol{Q} = 0 \quad in\,\Omega, \quad t>0 \tag{7.3}$$

where Ω is the computational domain, $\boldsymbol{R}$ are the numerical residuals, $\boldsymbol{Q}$ is the source term, $\boldsymbol{U} = (\rho, \rho v, \rho E)$ are the conservative variables, namely the air density



(a) (b)

Fig. 7.1 (a) High-fidelity discretization of the computational domain with 90000 elements, and (b) high-fidelity pressure coefficient contours for the unmodified RAE-2822 airfoil.

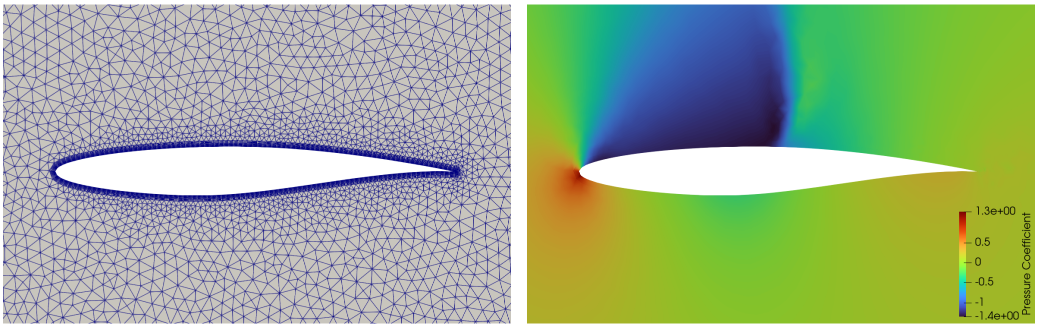
$\rho = \rho(h)$, the free-stream velocity v and the total energy E, and $\boldsymbol{F}^c$ and $\boldsymbol{F}^v$ are the convective and viscous fluxes, respectively:

$$\boldsymbol{F}^c = \begin{pmatrix} \rho v \\ \rho v \otimes v + \boldsymbol{I} p \\ \rho E v + p v \end{pmatrix} \tag{7.4}$$

$$\boldsymbol{F}^v = \begin{pmatrix} \cdot \\ \bar{\boldsymbol{\tau}} \\ \bar{\boldsymbol{\tau}} v + k \nabla T \end{pmatrix} \tag{7.5}$$

where $T = T(h)$ is the free-stream temperature, $p = p(h)$ is the free-stream static pressure, $k = k(h)$ is the thermal conductivity and $\bar{\boldsymbol{\tau}}$ is the tensor of viscous stresses.

We determine the level of fidelity of the aerodynamic simulation by modifying the element scale factor ES that defines the density of the computational mesh, where the lower the element scale the more refined is the discretization of the flow-field. For the optimization problem in Equation (7.1), two levels of fidelity are considered for the aerodynamic modeling: we set $ES = 2.5$ for the high-fidelity model corresponding to a grid of about 90000 cells, and $ES = 12$ for the low-fidelity model with about 30000 cells. Figure 7.1 illustrates the high-fidelity computational grid (Figure 7.1(a)) and the corresponding distribution of the pressure coefficient (Figure 7.1(b)) computed with the high-fidelity aerodynamic model, while Figure 7.2 provides details about the low-fidelity mesh (Figure 7.2(a)) and correlated distributions of the pressure



(a)

(b)

Fig. 7.2 (a) Low-fidelity discretization of the computational domain with 30000 elements, and (b) low-fidelity pressure coefficient contours for the unmodified RAE-2822 airfoil.

coefficient (Figure 7.2(b)) in output from the low-fidelity aerodynamic model; all those outcomes are obtained considering the unmodified RAE-2822 airfoil. The flow solver leverages for all the levels of fidelity the finite-volume code SU2 v6.2.0 to compute numerical solutions of the RANS equation using the Spalart-Allmaras turbolence model, a II order monotone upstream-centered scheme and an adaptive CFL number [237]. We require a computational residuals minor than 10^{-6} as the convergence criteria, with a fixed maximum number of iterations of 20000. The computational grid is generated through the GMSH software v4 where an automated procedure allows to determine an hybrid grid of triangles and quadrangles elements, balancing the accuracy and efficiency of the CFD computation with the robustness of the grid generation when the airfoil geometry is modified [238].

7.1.2 NM2-BO Results and Discussion

We present the results achieved with our NM2-BO framework for the aerodynamic constrained shape optimization of the RAE 2822 airfoil. The NM2-BO capabilities are compared against standard multifidelity Bayesian optimization methodologies implementing different acquisition function, namely the Multifidelity Expected Improvement (MFEI) [38], Multifidelity Max-Value Entropy Search (MFMES) [42], and Multifidelity Probability of Improvement (MFPI) [41]. The outcomes are discussed in terms of:

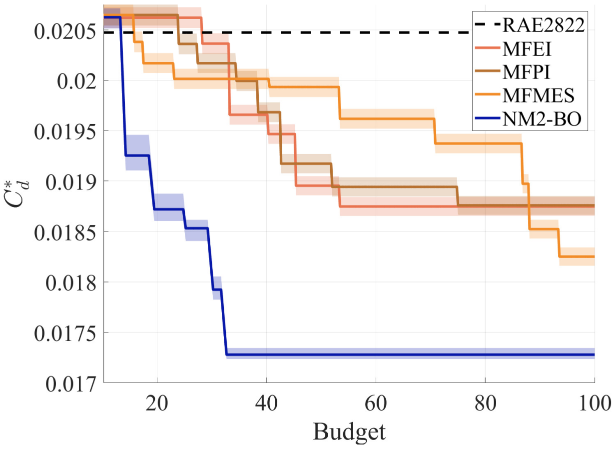


Fig. 7.3 Minimum drag coefficient C_d^* obtained with the NM2-BO algorithm and the baseline MFBO algorithms.

$$C_d^* = \min(C_d(\boldsymbol{x})) \tag{7.6}$$

The minimum drag coefficient C_d^* provides a measure of the improvement of the design solution and allows to assess the accuracy of the optimal shape configuration identified by the competing methods. We consider a statistics over 25 experiments for each methodology starting from different initial samples determined through a Latin hypercube design of experiments [209]. The objective of this experimental setting is to measure and compensate the influence of the random initialization during the assessment of the performance of the competing algorithms. In particular, the multifidelity searches are initialized with 30 initial design configurations, of which $n_0^{(1)} = 25$ samples are observed with the low-fidelity aerodynamic model and $n_0^{(2)} = 5$ are computed with the high-fidelity aerodynamic model.

Figure 7.3 illustrates the values of the minimum drag coefficient C_d^* as a function of the computational budget B to compare the optimization outcomes achieved by the competing algorithms. The results are reported in terms of median values of the minimum drag coefficient together with the associated statistics in between the 25-th and 75-th percentiles. It is possible to notice that the optimization procedure for all the algorithm has been initialized with design configurations that score worse than the baseline solution (C_d^* >0.020475): the algorithms progressively identify improved design solution through the evaluation of the aerodynamic models and reduce the

B	C_d^* MFEI	C_d^* MFPI	C_d^* MFMES	C_d^* NM2-BO
15	0.02062	0.02064	0.02065	0.01925
25	0.02062	0.02036	0.02001	0.01872
50	0.01895	0.01917	0.01993	0.01727
100	0.01875	0.01876	0.01825	0.01727

Table 7.1 Comparison between NM2-BO and standard MFBO algorithms median values of the minimum drag coefficient C_d^*.

Method	w_1	w_2	w_3	w_4	w_5	w_6
NM2-BO	−0.9875	0.9755	−0.1987	0.3927	0.8424	−0.9451
MFEI	−0.3785	−0.1015	−0.4479	0.3031	0.2993	−0.4426
MFPI	−0.4580	0.4574	0.2391	0.4266	0.4096	−0.2365
MFMES	−0.4752	0.4857	0.2124	−0.1824	0.3324	−0.1692

Table 7.2 Comparison between the best design solutions evaluated with the NM2-BO and the standard MFBO algorithms.

minimum drag coefficient C_d^* below the baseline design. We can observe that our non-myopic multifidelity framework provides larger reduction of the drag coefficient and remarkable accelerations of the optimization process: NM2-BO converges to an optimal combination of optimization variables with a fraction of the computational cost required by the baseline MFBO algorithms to identify sub-optimal design solutions.

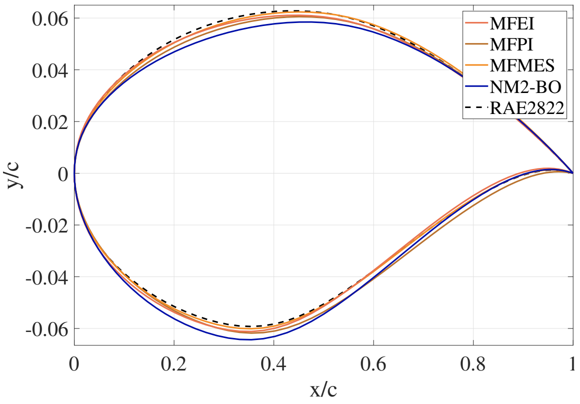
Table 7.1 reports the convergence history for the competing algorithms in terms of specific median values that characterize the distribution of the multifidelity experiments at $B = 15$, 25, 50 and 100. In particular, the NM2-BO algorithm achieves a median value of the drag coefficient equal to $C_d^* = 0.01872$ for $B = 25$ which corresponds to a design improvement of the 8.6% if compared with the baseline solution, and converges to an optimal median value of $C_d^* = 0.01727$ with a design upgrade of the 15.7% consuming a budget of just $B = 32.7$. The second best performing algorithm is the MFMES, which scores $C_d^* = 0.01825$ and provides a design improvement of the 10.9% with respect to the baseline airfoil adopting a computational budget $B = 93.6$.

Table 7.2 reports the optimal design solutions identified by the competing algorithms. We observe that our NM2-BO identifies an aerodynamic design configuration capable to deliver a drag coefficient of $C_d^* = 0.01723$, which corresponds to an upgrade of the 15.9% with respect to the baseline solution. These outcomes demonstrates the remarkable accelerations and superior design solutions achieved by our NM2-BO algorithm: the capability to capitalize from the two-step ahead informative gains enhances the wise selection of optimization variables and aerodynamic models to query while balances resource expenditure and accuracy of the design solution.

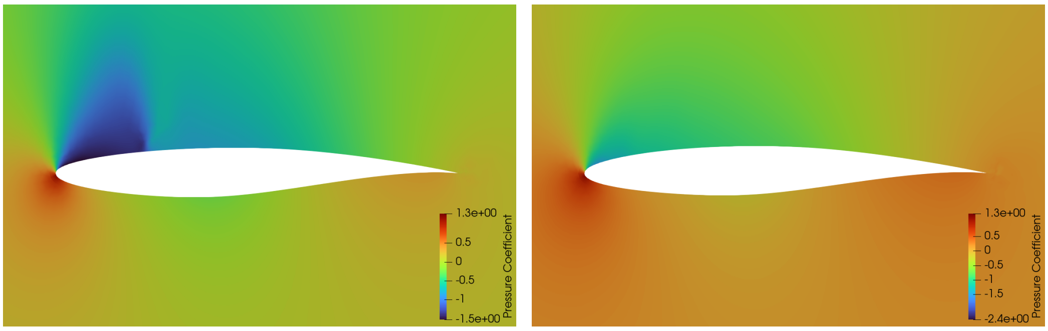
Figure 7.4 illustrates the best optimal airfoil geometries obtained with the competing algorithms compared with the baseline RAE2822 airfoil, and the related distributions of the pressure coefficient. In particular, the optimal airfoil computed with our non-myopic algorithm allows to efficiently expand the flow at the upper surface leading edge while reducing the intensity of the shock waves that occur at transonic speed; indeed, the corresponding distribution of the pressure coefficient (Figure 7.4(e)) shows that the shock wave on the upper surface is reduced in intensity, and the supersonic bubble evolves with an almost constant pressure plateau. These outcomes are achieved through the reduction of the adverse pressure gradient which results in the substantial decrease of the drag coefficient.

7.2 Cross-Regime Aerodynamic Shape Optimization Problem

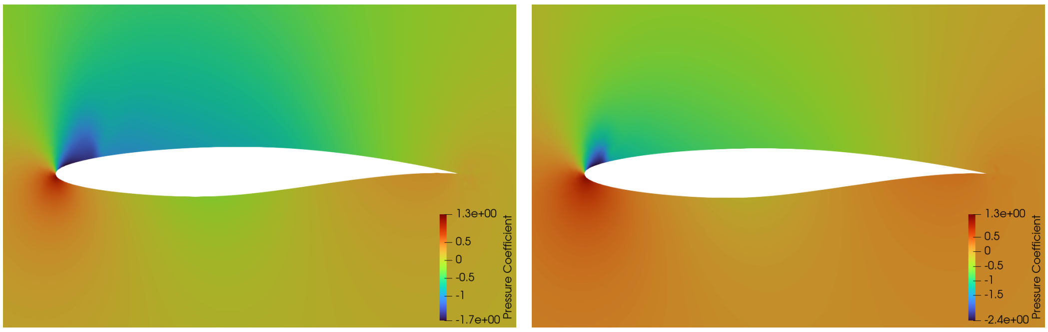
In this section, we consider an aerodynamic shape optimization problem within a cross-regime scenario: the fluid regime and the associated physical phenomena evolve during the optimization process according to the combination of design variables considered. This defines a robust optimization procedure that potentially ensures to obtain optimal airfoil shapes for different operational conditions, without limiting the improvement of performance to a single application context [239, 240]. We address this specific aerodynamic problem through our physics-aware multi-fidelity Bayesian optimization framework (Chapter 5) given the capability to incorporate the prior scientific knowledge about the structure of the domain related to the transition of the fluid dynamic regimes during the optimization. Thus, the physics-aware utility function is formalized to bias the search procedure according to the evolution of the physical domain (Section 7.2.2).



(a) Optimal airfoils



(b) MFEI (c) MFPI



(d) MFMES (e) NM2-BO

Fig. 7.4 (a) Optimal airfoils geometry and associated pressure coefficient contours obtained with (b) MFEI, (c) MFPI, (d) MFMES, and (e) our NM2-BO algorithm.

The Mach number M is the main physical variable that captures the evolution of the fluid domain, and constitutes a measure of the compressibility effects that modify the fluid structure. According to the fluid mechanics theory [241, 242], the flow-field around streamlined bodies is defined subsonic for values of the Mach number lower than 0.8, and represents a condition characterized by the absence of discontinuities and the fluid properties vary continuously. As the Mach number approaches the sonic condition, discontinuities in the form of local shock waves start to appear in the fluid domain with the consequent separation of the viscous boundary layer. This mixed subsonic-supersonic flow field emerges for values of the Mach number between 0.8 and 1.2, and is commonly referred as the transonic regime. The interactions between shock waves and boundary layer determine an increase of the drag force, and unsteady effects generated by a shift of the center of pressure of the aerodynamic body. Therefore, the cross-regime scenario poses significant challenges associated with the modeling of complex physics. On one hand, the subsonic regime can be represented adopting simplifications in the aerodynamic modeling due to the smooth evolution of the flow field; on the other hand, the transonic regime requires the implementation of accurate and robust modeling techniques to capture non-linear phenomena in the unsteady mixed subsonic-supersonic flow.

The cross-regime aerodynamic design optimization problem consists in the identification of the optimal combination of design parameters that minimizes the drag coefficient C_d of a transonic airfoil, subject to a variety of aerodynamic and geometric constraints. For this demonstrative test-case, we adopt the RAE 2822 transonic airfoil that is modified through the code WG2AER developed by [236]. The aerodynamic design optimization problem is formulated as follows:

$$
\begin{aligned}
\min_{\boldsymbol{x} \in \mathscr{X}} \quad & C_d(\boldsymbol{x}) \\
& \boldsymbol{x} = [w_1, ..., w_6, M] && \text{(7.7a)} \\
s.t. \quad & C_l = 0.824 && \text{(7.7b)} \\
& -0.1 \leq C_m \leq -0.01 && \text{(7.7c)} \\
& t/c = 0.1211 && \text{(7.7d)} \\
& r \geq 0.007c && \text{(7.7e)} \\
& \tau \geq 5^\circ && \text{(7.7f)} \\
& t_{85}/c \geq 0.02 && \text{(7.7g)} \\
& \mathscr{X} = I_w \times I_M && \text{(7.7h)}
\end{aligned}
$$

where the design parameters $\boldsymbol{x} = [w_1, ..., w_6, M]$ consist of six weights w_i assigned to the shape modification polynomial functions and the Mach number M. The feasibility of the design configuration is subject to obtain certain aerodynamic performances in terms of lift coefficient C_l and pitching momentum coefficient C_m, and the modified geometry must accomplish the constraints on the airfoil thickness t, chord c, trailing edge angle τ, and thickness of the airfoil at the 85% of the chord t_{85}. The search for optimal design configuration is limited to the domain $\mathscr{X}$ bounded by the move limits imposed for the weights $I_w = [-1, 1]^6$ and for the Mach number $I_M = [0.6, 0.99]$. This allows for the exploration of different aerodynamic configurations, and improves the robustness of the optimization procedure in presence of an evolution of the fluid domain from the low subsonic to the transonic regime.

The constraints of this cross-regime optimization problem (Equation 7.7) are handled formalizing the corresponding unconstrained optimization problem through the quadratic penalty functions as follows:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathscr{X}} C_d + c_1 g^+(C_m, -0.1) + c_1 g^-(C_m, -0.01) + \\ c_2 g^+(t/c, 0.007) + c_3 g^+(\tau, 5^\circ) + c_4 g^+(t_{85}/c, 0.02) \end{aligned} \tag{7.8}$$

where g^+ and g^- are the quadratic penalty functions [236], and $c_1 = 1000$, $c_2 = 5000$, $c_3 = 10$, and $c_4 = 30$, while the constraint on the maximum thickness is satisfied through an automatic procedure that scales the modified geometry, and the constraint on the lift coefficient is satisfied considering the angle of attack α as a free parameter in the CFD simulations.

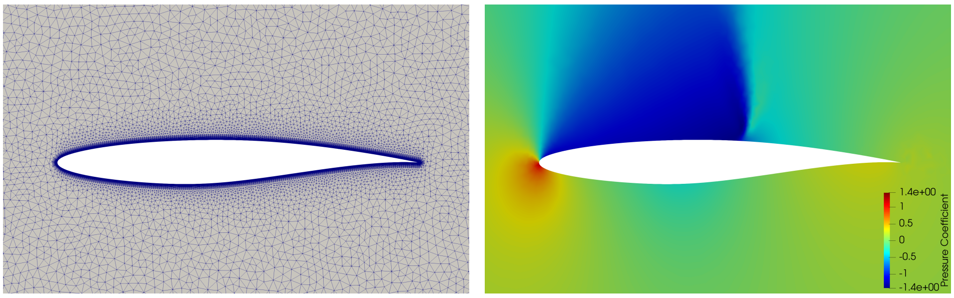
7.2.1 Aerodynamic Models

The distribution of the pressure field around the RAE 2822 is represented through a library of three numerical models at different accuracy and related CPU time associated with their evaluation. The aerodynamic models return the drag coefficient C_d and the aerodynamic constraints on lift C_l and pitching momentum coefficients C_m given the selected configuration of optimization variables in terms of weights $\boldsymbol{w}$ and Mach number M.

We model the aerodynamic domain through the numerical solution of the RANS equations (Equation (7.3)) adopting a Computational Fluid Dynamic (CFD) solver. This permits to obtain a finite-dimensional approximation of the pressure coefficient over the fluid domain. We use the SU2 v6.2.0 CFD code based on the finite-volume method to discretize the RANS partial differential equations considering a fully turbulent flow-field [237]. To ensure the robustness of the aerodynamic outcomes, the convergence criteria is set for a computational residuals minor than 10^{-6} with a fixed maximum number of 20000 iterations. The fluid domain is discretized through a computational mesh generated using the GMSH software v4 [238] where an automated procedure embedded within the adopted computational tool adapts the hybrid grid of triangles and quadrangles elements with the modified geometry of the airfoil. This permits to target the balance between accuracy and efficiency of the CFD computations.

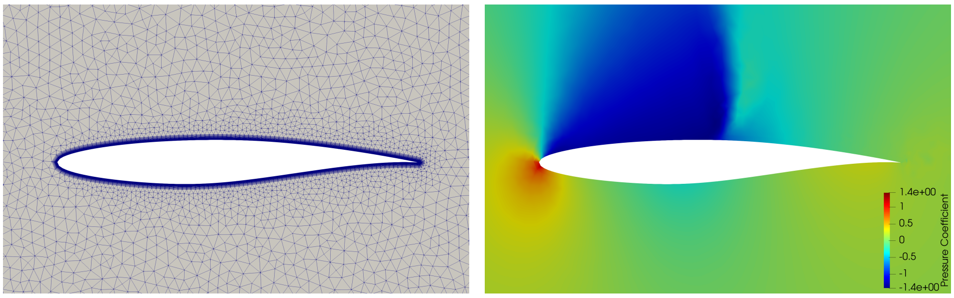
The fidelity of the aerodynamic simulations is determined controlling the granularity of the computational mesh through the associated element scale factor ES where the higher the value of ES the coarser the discretization of the fluid domain. Specifically, three levels of fidelity are considered for the aerodynamic modeling: we set $ES = 2.5$ for the high-fidelity model corresponding to a grid of about 90000 cells, $ES = 12$ for the mid-fidelity model with about 30000 cells, and $ES = 20$ for the low-fidelity model consisting of a mesh with 15000 cells.

The high-fidelity model achieves an accurate representation of complex aerodynamic phenomena that occurs at higher regimes of speed including discontinuities, shock-waves and unsteadiness of the flow-field. This provides a close prediction of the mixed subsonic-supersonic fluid domain that characterizes the transonic regime. The mid-fidelity model reduces the demand for CPU if compared with the high-fidelity model by decreasing the number of cells that discretize the fluid domain. This produces a reliable estimate of the aerodynamic coefficients for Mach number regimes far from the sonic condition where the unsteady phenomena have marginal effects, and a reduced accuracy for discontinuous flows that occur at the transonic regime. The low-fidelity model further decreases the number of elements adopted to discretize the fluid domain, and leads to an inaccurate representation of the compressibility effects that characterize the more turbulent flows at higher Mach values ($M > 0.65$). However, the coarser discretization reduces the computational cost required for its evaluation of the 65% with respect to the cost associated with the high-fidelity model.



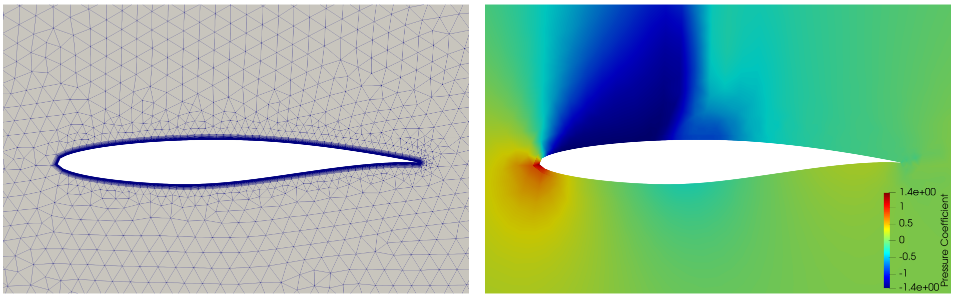
(a) (b)

Fig. 7.5 (a) high-fidelity discretization of the computational domain, and (b) high-fidelity pressure coefficient contours for the RAE-2822 airfoil.



(a) (b)

Fig. 7.6 (a) mid-fidelity discretization of the computational domain, and (b) mid-fidelity pressure coefficient contours for the RAE-2822 airfoil.



(a) (b)

Fig. 7.7 (a) low-fidelity discretization of the computational domain, and (b) low-fidelity pressure coefficient contours for the RAE-2822 airfoil.

The high-fidelity (Figure 7.5), mid-fidelity (Figure 7.6), and low-fidelity (Figure 7.7) meshes and distributions of the pressure coefficient highlight the increasing discrepancy of the flow field representations as the level of fidelity decreases.

7.2.2 Physics-Aware Utility Function for Aerodynamic Design

We formulate our physics-aware utility function α_4 (Equation (5.1)) to include a learning bias that wisely measures the utility of evaluating an aerodynamic model – at a certain level of fidelity – according to the characteristics of the fluid dynamic regime. This is achieved by formalizing the learning bias dependent on the Mach number $\boldsymbol{\psi} = M$ as the variable representative of the physical phenomena involved in the aerodynamic domain. Accordingly, $\alpha_4(M,l)$ is depicted as follows:

$$\alpha_4(M,l) = \begin{cases} 1 & \text{if} \quad l = 1,...,L-1 \\ \frac{M_s}{M_s - M} & \text{if} \quad l = L \quad M_s = 1 \end{cases} \tag{7.9}$$

This physics-aware utility function encourages the query of the high-fidelity model for values of the Mach number close to the sonic condition ($M = 1$). Indeed, α_4 increases the value of the multifidelity acquisition function (Equation (5.1)) when an aerodynamic configuration is evaluated with the high-fidelity model in the transonic regime ($M > 0.8$). The goal is to capture large-scale separation of the fluid vein and unsteady effects that deeply influence the overall performance of the aerodynamic system. This permits to better support and improve the search for optimal designs through the a priori scientific knowledge about the aerodynamic domain structure derived directly from the governing equations.

7.2.3 PA-MFBO Results and Discussion

This section illustrates and discusses the results achieved with the physics-aware multifidelity Bayesian optimization (PA-MFBO) framework for the aerodynamic design optimization problem of the RAE 2822 transonic airfoil. The effectiveness of the PA-MFBO algorithm is compared with other existing methods commonly adopted to address black-box optimization problems, namely the single-fidelity efficient global optimization (EGO) algorithm [109] adopting the high-fidelity aerodynamic model, and the multifidelity Bayesian optimization based on the multifidelity expected improvement acquisition function (MFEI) [243]. We select the minimum drag coefficient as the assessment metric to evaluate the capabilities of the com-

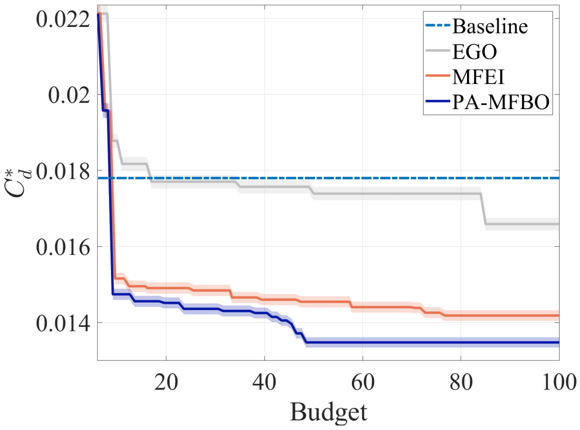


Fig. 7.8 Statistics over 25 runs of the minimum drag coefficient C_d^* obtained with the competing algorithms.

peting algorithms, and provide a measure of the improvement of the aerodynamic performance achieved by the identified design configurations:

$$C_d^* = \min(C_d(\boldsymbol{x})) \tag{7.10}$$

Figure 7.8 reports the convergence history of the optimization procedure in terms of values of the minimum drag coefficient C_d^* as a function of the computational budget B. We compute 25 independent replications of the experiment for each methodology to measure and compensate the influence of the random initial sampling procedure and ensure a fair comparison of the algorithms. The outcomes of the statistics are represented through the median values of the assessment metric C_d^* together with the the associated values in between the 25-th and 75-th percentiles. We consider as the baseline design solution the drag coefficient $C_d^* = 0.017796$ obtained for the unmodified RAE 2822 airfoil corresponding to the design configuration $\boldsymbol{x} = [0,0,0,0,0,0,0.65]$. The convergence histories show that all the competing methodologies are capable to identify improved design solutions if compared with the baseline RAE 2822 design. In particular, our PA-MFBO delivers better aerodynamic designs in terms of reduction of the drag coefficient if compared with the baseline EGO and MFEI algorithms. As can be seen, PA-MFBO reduces the drag coefficient at the beginning of the optimization procedure and identifies optimal design configurations consuming a fraction of the available computational budget. In addition, we note from the convergence of the PA-MFBO experiments that the algorithm starts the search allocating budget for the exploration of different

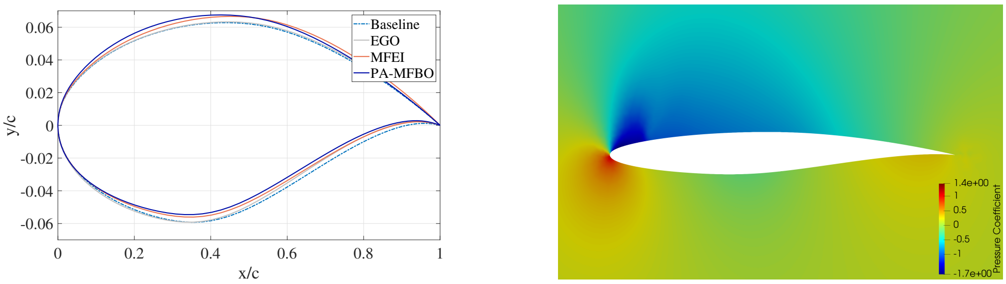
B	C_d^* EGO	C_d^* MFBO	C_d^* PA-MFBO
6	0.02212 (-24.30 %)	0.02212 (-24.30 %)	0.02212 (-24.30 %)
10	0.01887 (-6.055 %)	0.01515 (14.87 %)	0.01455 (18.24 %)
25	0.01770 (0.5394 %)	0.01484 (16.61 %)	0.01435 (19.36 %)
50	0.01738 (2.337 %)	0.01454 (18.30 %)	0.01347 (24.31 %)
100	0.01658 (6.833 %)	0.01418 (20.32 %)	0.01348 (24.31 %)

Table 7.3 Median values of the minimum of the drag coefficient C_d^* and corresponding design improvement $(\cdot)$ obtained with the competing algorithms.

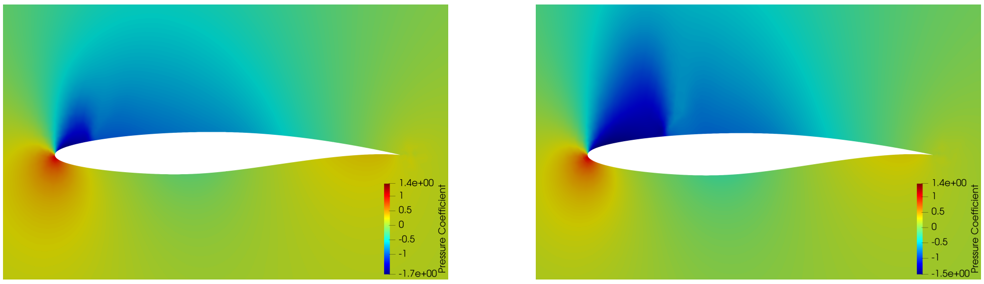
design configurations over the domain, which corresponds to a moderate reduction of the drag coefficient. Then, the computational resources are directed towards the exploitation phase reducing the values of the design objective.

Table 7.3 reports the median values of the minimum drag coefficient for incremental computational expense $B = 6$, 10, 25, 50, and 100 to better measure the performance of the proposed PA-MFBO in comparison with the baseline algorithms. After the expenditure of the initial resources for the initialization phase ($B = 6$), all the algorithms identify design solutions characterized by higher values of the drag coefficient with respect to the baseline design. At $B = 10$, the multifidelity algorithms capitalize from the collected evaluations of the objective function and identify improved designs with respect to the baseline solution, while the single-fidelity EGO still achieves worst designs if compared with the unmodified RAE 2822 airfoil. The PA-MFBO methodology realizes the larger design improvement of the 24.31% before consuming a Budget of $B = 50$, which is superior to the MFEI design upgrade of the 20.32% obtained adopting much more computational resources. Moreover, the EGO methodology is capable to deliver a design improvement of only the 6.83% using all the available computational budget.

To clarify and interpret the results obtained, Figure 7.9 illustrates the aerodynamic performance of the optimal designs determined by all the algorithm. In particular, we report the optimal airfoil shapes corresponding to the best aerodynamic design (Figure 7.9(a)), and the related pressure coefficient distribution for the PA-MFBO (Figure 7.9(b)), MFEI (Figure 7.9(c)), and EGO (Figure 7.9(d)) design solutions. It can be noticed that the superior performance of the design configuration identified with the PA-MFBO can be explained with the efficient expansion of the fluid at the upper surface leading edge that induces low-intensity shock waves if compared



(a) (b) PA-MFBO



(c) MFEI (d) EGO

Fig. 7.9 (a) Optimal airfoils geometry and associated pressure coefficient contours obtained with (b) the PA-MFBO, (c) MFBO, and (d) EGO algorithm.

with the other modified airfoils. This results from the increase of the leading edge radius and aft camber that produces a reduction of the adverse pressure gradient, and permits a smooth evolution of the pressure coefficient in the supersonic bubble. This features of the aerodynamic domain determine the substantial decrease of the drag coefficient and enhance the overall efficiency of the modified airfoil. In this design test case, the remarkable performance of the PA-MFBO framework is related to the physical bias introduced in the sampling scheme. This enables the capitalization from the prior scientific knowledge about the fluid dynamic regime, and permits to accelerate and improve the optimization search through the wise selection of the aerodynamic model to query with a continuous balance between computational cost and accuracy of the solution.

7.3 Cross-Regime Aerodynamic Design Optimization Problem

In this section, we address an aerodynamic design optimization problem considering a cross-regime optimization setting – the cross-regime optimization has been previously discussed in Section 7.2. Specifically, the design optimization problem aims at identify the optimal combination of Mach number M and angle of attack that minimizes the drag coefficient C_d of a NACA 0012 airfoil subject to maintaining a minimum coefficient of lift C_l at a certain altitude h. This specific aerodynamic design optimization problem is used to demonstrate and validate the performance of the methodologies proposed in this thesis including the non-myopic NM2-BO (Section 4.4), the physics-aware PA-MFBO (Section 5.1), and the physics-aware non-myopic PA-NM2BO (Section 5.2) frameworks. Indeed, this NACA 0012 optimization problem is frequently adopted as validation test case [244–247] due to the availability of the real-world solution of the optimization problem and experimental data of the lift and drag coefficients [248, 249]. This permits to assess the performance of the algorithms with respect to the experimental optimum and real-world aerodynamic data, which usually is not the case for real-world experiments where the optimum solution is not computable.

The cross-regime aerodynamic design optimization problem is formalized as follows:

$$
\begin{aligned}
\min_{\boldsymbol{x}\in\mathscr{X}} \quad & C_d(\boldsymbol{x}) \\
s.t. \quad & 0.45 - C_l(\boldsymbol{x}) \leq 0 && \text{(7.11a)} \\
& h - h_0 = 0 && \text{(7.11b)} \\
& \mathscr{X} = I_M \times I_\alpha && \text{(7.11c)}
\end{aligned}
$$

where $\boldsymbol{x} = [M, \alpha]$ are the design variables, h is the flight altitude, $C_l = 0.45$ is the constraint on the lift coefficient, $h_0 = 10000m$ is the constraint on the flight altitude, and $\mathscr{X} = I_M \times I_\alpha$ is the design space with $I_M = [0.6, 0.99]$ and $I_\alpha = [0°, 4.5°]$. The constraint on the lift coefficient (Equation 7.11a), on the flight altitude (Equation 7.11b), and on the domain boundaries (Equation 7.11c) are imposed to replicate the flight scenario of a civil aircraft during the cruise phase. In particular, the Mach number limits are defined to explore design configurations in both the subsonic and transonic regime with the appropriate level of fidelity to grasp design solutions that may be otherwise discarded. To handle the aerodynamic constraint during the optimization process, we approximate the lift coefficient over the entire design space through a multifidelity Gaussian process $C_l^{(l)} = GP(\mu_C^{(l)}, \sigma_C^{(l)})$ and adopt a conditional selection strategy that defines an additional surrogate model $GP(\tilde{\mu}^{(l)}, \tilde{\sigma}^{(l)})$ to represent the feasible regions of the design space from which the new design to be evaluated will be chosen, where $\tilde{\mu}^{(l)} = \mu^{(l)}(\mu_C^{(l)} > 0.45)$ and $\tilde{\sigma}^{(l)} = \sigma^{(l)}(\mu_C^{(l)} > 0.45)$.

7.3.1 Aerodynamic Modeling

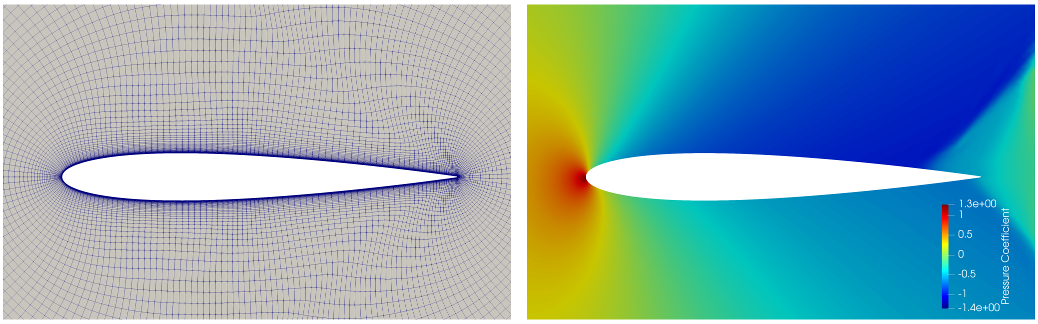
We consider a library of three aerodynamic models at different levels of fidelity. All the aerodynamic models compute the drag coefficient C_d and lift coefficient C_l, given the geometry of the airfoil, the flight altitude h, the Reynolds number Re, the free-stream Mach number M and the flight angle of attack α. The high-fidelity aerodynamic models are based on a CFD solver for the numerical solution of the Reynolds Averaged Navier-Stokes equations; the mid-fidelity model is conceptually equivalent to the high-fidelity model, except for a coarser discretization of the aerodynamic domain; the low-fidelity model solves numerically the potential flow equation through the panel method.

Specifically, the high-fidelity and mid-fidelity aerodynamic models represent the fluid domain through the Reynolds Averaged Navier-Stokes (RANS) equations (Equation (7.3)) to capture the effects of turbulence that occur at higher regimes of Mach number and angle of attack. We use the finite volume method to discretize Equation (7.3) in space with a standard edge-based data structure. The numerical solution of the RANS equations is computed through the computational fluid dynamic code SU2 version 7.0.3 in RANS mode with the Spalart-Allmaras turbolence model [250], using the Jameson-Schmidt-Turkel model (JST) and the Scalar upwind model as the flow and turbolence numerical method, respectively, and Euler implicit for the time discretization. The convergence criterion is set as satisfied for values of the computational residuals minor than 10^{-6}.

The high-fidelity computational domain consists of a grid of 27125 quadrilateral elements with the farfield boundary extended to $500c$ away from the airfoil surface, where $c = 1m$ is the Reynolds length. Figure 7.10(a) provides details about the discretization of the computational domain for the high-fidelity model. We consider this model as the high-fidelity representation of the aerodynamic phenomena for the capability to provide an accurate estimate of aerodynamic coefficients C_d and C_l even in the presence of discontinuity and separation phenomena typical of the mixed flow condition that characterizes the transonic regime. Figure 7.10(b) illustrates the distribution of the pressure coefficient around the NACA 0012 airfoil computed with the high-fidelity aerodynamic model, for a Mach number of 0.9, angle of attack of 3° and an altitude of $10000m$.

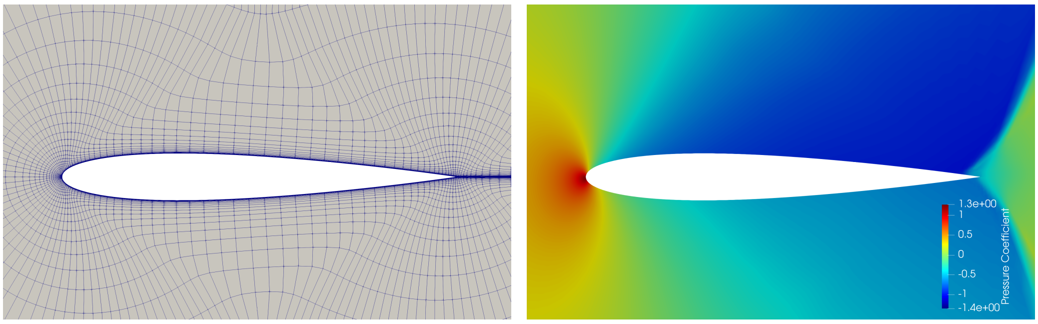
The mid-fidelity aerodynamic model adopts the same governing RANS equations and the same solution method of the high-fidelity model, with the grid density being far coarser than that of the high-fidelity one. In particular, we discretize the computational domain with 14336 quadrilateral elements, which leads to a lower computational time required to estimate the aerodynamic coefficients C_d and C_l in comparison to the high-fidelity model. Figure 7.11(a) illustrates the discretization of the computational domain for the mid-fidelity aerodynamic model. As a result of the coarser mesh, this model is not suitable for the computation of the aerodynamic field for Mach numbers close to the sonic condition, as it may not capture the discontinuities of the flow that occur at transonic speed regimes. However, the mid-fidelity representation can still provides close approximations of the aerodynamic coefficients for Mach number regimes far from the sonic condition. Figure 7.11(b) presents the distribution of the pressure coefficient around a NACA 0012 airfoil in output from the mid-fidelity aerodynamic model, for a Mach number of 0.9, angle of attack of 3° and an altitude of $10000m$.

The low-fidelity aerodynamic model approximates the flow-field around the airfoil through the potential flow panel method with an integral boundary layer formulation and an approximate e^N envelope method to calculate the aerodynamic transition. We use XFOIL [251] to solve the potential flow equation through the panel method with the Karman-Tsien compressibility correction to rapidly predict the airfoil performances. The code is developed to evaluate the aerodynamic coefficients at low Reynolds numbers and is capable to calculate the viscous pressure distribution, capturing the influence of limited trailing edge separation and laminar separation bubbles. We use 200 panels to discretize the airfoil contour and set 200



(a) (b)

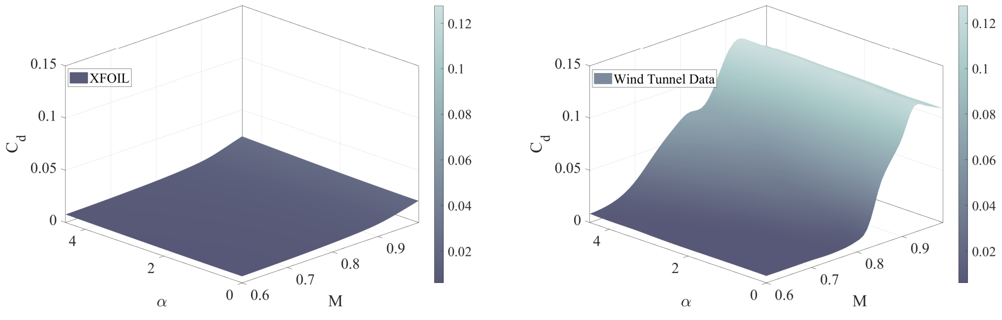
Fig. 7.10 (a) High-fidelity discretization of the computational domain with 27125 quadrilateral elements, and (b) pressure coefficient contours for a Mach number of 0.9, angle of attack of 3° and an altitude of 10000 m.



(a) (b)

Fig. 7.11 (a) Mid-fidelity discretization of the computational domain with 14336 quadrilateral elements, and (b) pressure coefficient contours for a Mach number of 0.9, angle of attack of 3° and an altitude of 10000 m.

iterations as the convergence criteria to contain the computation time. We consider the aerodynamic coefficients computed with this model being the low-fidelity approximation of the real values, as the potential flow panel method and the compressibility correction are not capable to achieve an accurate representation of the flow-field in the more turbulent flow at higher values of Mach number ($M > 0.65$) and angle of attack ($\alpha > 4°$). Moreover, the combination of the panel method with the Karman-Tsien correction causes XFOIL to overestimate lift and underestimate drag [252]. Figure 7.12 illustrates the comparison between the drag coefficient of a NACA 0012 airfoil computed with the low-fidelity model over the design space, and the wind tunnel prediction of the drag coefficient obtained interpolating with a spline function the experimental data [1].



(a) (b)

Fig. 7.12 Prediction of the drag coefficient C_d as a function of Mach number M and angle of attack α obtained by interpolating data computed by the low-fidelity model (7.12(a)) and experimental wind tunnel data (7.12(b)) for the NACA0012 airfoil [1]

7.3.2 Results and Discussion PA-NM2BO

This section presents and discusses the results achieved with the physics-aware non-myopic MFBO (PA-NM2BO) for the constrained aerodynamic optimization problem of a NACA 0012 airfoil subject to a multi-regime flow-field. The PA-NM2BO is compared with state of the art Bayesian frameworks such as the efficient global optimization (EGO) [109], a single-fidelity non-myopic framework (lookEGO) [127, 134, 136, 137], MFBO based on the multifidelity expected improvement acquisition function (MFEI) [38], the proposed physics-aware multifidelity Bayesian optimization (PA-MFBO), and our non-myopic multifidelity Bayesian optimization (NM2-BO).

All the multifidelity algorithms consider the complete spectrum of three levels of fidelity for the aerodynamic modeling, while the single fidelity frameworks elicitate evaluations of the high-fidelity CFD model only. For the numerical experiments discussed in this section, the initial set of design configurations to be evaluated is determined through a Latin hypercube strategy, and are used to compute the first surrogate model of the objective function at the first iteration. Specifically, the multifidelity searches are initialized with 13 initial design configurations among which 10 observations of the objective function are evaluated with the low-fidelity model, 2 are computed with the mid-fidelity model, and the last one is evaluated with the high-fidelity model, while for the single fidelity frameworks we consider 2 initial design configurations at which we compute high-fidelity observations. The

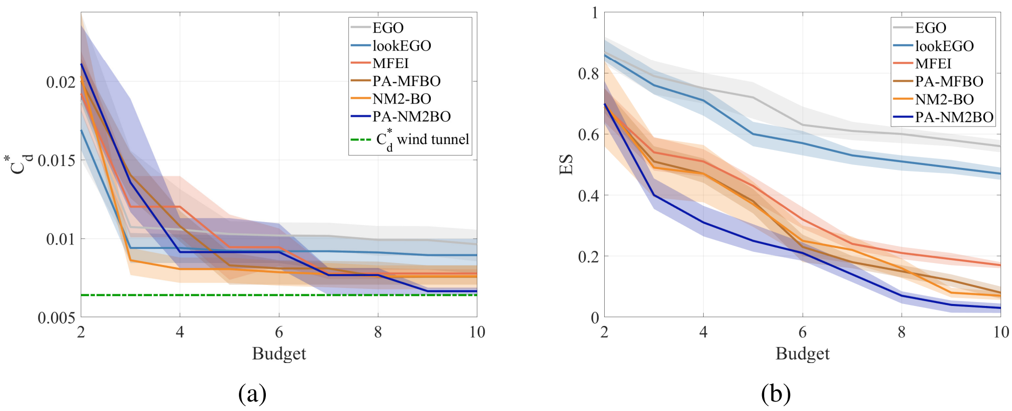


Fig. 7.13 Statistics over 20 runs of the minimum drag coefficient C_d^* (7.13(a)) and of the error of the surrogate model *ES* (7.13(b)) computed as the Gaussian process of the objective function obtained with the competing algorithms

computational cost B is evaluated as the sum of the computational costs required to estimate the aerodynamic domain with a certain level of fidelity. To total computational budget assigned is fixed at $B_{max} = 10$, that is equivalent to 10 evaluations of the high fidelity model. The overall budget includes also the cost for the initial samples. The single-fidelity competing methods adopt the Gaussian process as the surrogate model of the objective function [126], and the multifidelity algorithms implement the multifidelity Gaussian process surrogate model (Section 3.3.2) to synthesize the responses from the spectrum of aerodynamic model into a unique representation of the objective function. We adopt the square exponential kernel for the covariance function of the GP models and provide an estimate of the hyperparameters using the maximum likelihood estimation technique [253].

Figure 7.13(a) presents the results achieved in terms of reduction of the minimum drag coefficient $C_d^* = min(C_d(\boldsymbol{x}))$ while Figure 7.13(b) illustrates the error of the surrogate model obtained from a randomized statistic of 20 experiments starting from different initial samples for all the algorithms. In particular, the evaluation of the minimum drag coefficient provides a metric reflecting the goodness of the design solution identified by the competing algorithms and provides a measure of the improvement in the solution of the optimization problem. The measure of the surrogate error reflects the capability of the competing methods to combine the responses from the library of aerodynamic model to improve the accuracy of the surrogate model: this permits to better inform the acquisition function with a reliable estimate of the objective function and potentially lead to superior optimization performance. The outcomes are reported in terms of the median values (solid line) together with the observations falling in the interval between the 25-th and 75-th percentiles (shaded area). The error of the surrogate model of the objective function ES is defined as the relative error between the values of the drag coefficient $\tilde{C}_d$ predicted by the surrogate model of the objective function and the values of $C_{d,wt}$ determined through real world NACA 0012 wind tunnel data [1], since the high fidelity model has been validated with this experiments:

$$ES = \frac{1}{n_{wt}} \sum_{i=1}^{n_{wt}} \frac{\tilde{C}_d(\boldsymbol{x}_{i,wt}) - C_{d,wt}(\boldsymbol{x}_{i,wt})}{C_{d,wt}(\boldsymbol{x}_{i,wt})} \tag{7.12}$$

where $n_{wt} = 120$ is the number of experimental data available and $\boldsymbol{x}_{i,wt}$ is the i-th design configuration evaluated in the wind tunnel experiments. The experimental optimum is $C^*_{d,wt} = 0.0064$ (dashed line in Figure 7.13) corresponding to the optimum design $\boldsymbol{x}^*_{wt} = [M^*_{wt}, \alpha^*_{wt}] = [0.6574, 3.013]$, and is also estimated through wind tunnel experiments. The results suggest that the PA-NM2BO algorithm provides superior optimization performance than the competing algorithms, and achieves larger design improvements reducing both the drag coefficient and the surrogate error earlier on in the first stages of the optimization process. After the exploitation of all the computational budget available, it can be observed that the PA-NM2BO identifies superior design configurations of the NACA0012 if compared with the reference methodologies: the joint contributions of the two-step lookahead multifidelity policy and the physics-aware formulation allow to efficiently learn the aerodynamic surrogate and minimize the drag coefficient. In addition, we notice that the inclusion of the awareness about the fluid-dynamic physics enhances the optimization procedure, and obtains sensitive improvements of the accuracy of the multifidelity surrogate thanks to the wise use high-fidelity models at higher mach number regimes. Table 7.4 depicts the minimum median values of the drag coefficient achieved by the algorithms, and reports the drag coefficient along with the design configurations and the queries of the aerodynamic models corresponding to the experiments that give the best results. Our framework allows to achieve a design solution almost identical to the best design from wind tunnel data using 993 observations of the low-fidelity model, 10 evaluations of the mid-fidelity model and only 4 queries of the high-fidelity model. Table 7.5 indicates the percentage errors computed on the mean and on the best values of the drag coefficient determined by each competing methods

Method	mean C_d^*	best C_d^*	best M^*	best α^*	HF	MF	LF
EGO	0.00963	0.00858	0.621	2.622	10	-	-
lookEGO	0.00894	0.00759	0.658	3.192	10	-	-
MFEI	0.00777	0.00742	0.649	3.230	4	11	497
PA-MFBO	0.00756	0.00706	0.636	3.078	5	8	988
NM2-BO	0.00759	0.00676	0.633	2.959	4	10	997
PA-NM2BO	0.00664	0.00642	0.656	3.000	4	10	993
Wind Tunnel	-	0.0064	0.657	3.013	-	-	-

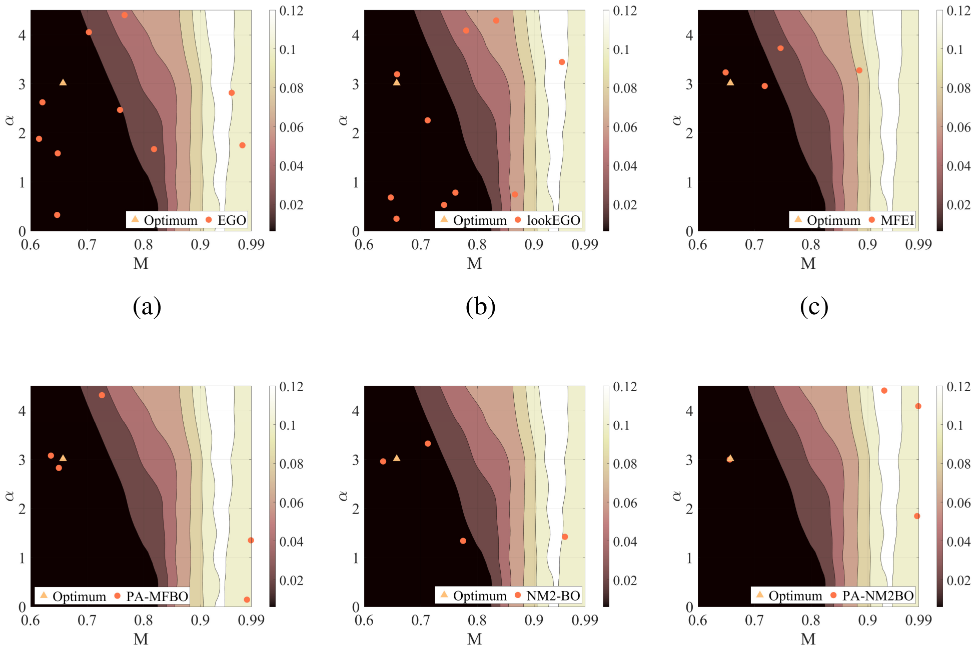
Table 7.4 Comparison across median and best values of the minimum drag coefficient C_d^* over 20 experiments for the competing algorithms, including the related optimum design variables M^* and α^*, and the queries of the high-fidelity (HF), mid-fidelity (MF) and low-fidelity (LF) model.

Error C_d^*	EGO	lookEGO	MFEI	PA-MFBO	NM2-BO	PA-NM2BO
Mean	50.5%	39.7%	21.5%	18.2%	18.7 %	3.88%
Best	34.2%	18.6%	15.9%	10.4%	5.75 %	0.422%

Table 7.5 Comparison of the percentage error of the estimate of the minimum drag coefficient C_d^* with respect to the wind tunnel optimum $C_{d,wt}^*$: mean and best values over the 20 trials.

with reference to the wind tunnel optimum. On average, the PA-NM2BO strategy permits to achieve an error of about 3.88%, whereas the second-best NM2-BO algorithm achieves an error of the 5.75%. Moreover, our framework scores the overall maximum reduction of the error with a value of about 0.422%. This demonstrates the capability of the PA-NM2BO to leverage multiple aerodynamic models that jointly with the lookahead property and the physics-awareness enhance the optimization performances.

Figure 7.14 illustrates the design space obtained using a spline interpolation of wind tunnel data and the high-fidelity sampling process corresponding to the experiments that provide the best results in terms of reduction of the drag coefficient for all the algorithms considered. It can be observed that the multifidelity strategies allow to reduce the total number of high-fidelity observations respect to the single-fidelity algorithms, leveraging the information from the mid-fidelity and low-fidelity



(a) (b) (c)

(d) (e) (f)

Fig. 7.14 High-fidelity evaluations of C_d called by the different algorithms: EGO (7.14(a)), lookEGO (7.14(b)), MFEI (7.14(c)), PA-MFBO (7.14(d)), NM2-BO (7.14(e)), and PA-NM2BO (7.14(f)). The contour map indicates the C_d value obtained interpolating the wind tunnel data as a function of Mach number M and angle of attack α.

models. The proposed PA-NM2BO framework uses the high-fidelity model to explore the regions of the design space characterized by higher Mach numbers and angles of attack thanks to the physics-awareness property, and effectively exploit the design space sampling closely to the optimum configuration through the lookahead feature (Figure 7.14(f)). These aspects permit to accurately predict the aerodynamic domain at transonic regimes taking advantage of lower-fidelity models to explore design configurations at lower speeds and angle of attack. Also the PA-MFBO algorithm samples the design space with a similar approach thanks to the domain-aware formulation (Figure 7.14(d)). However, the exploitation is not effective as PA-NM2BO demonstrating the significant role of the lookahead strategy to improve the optimization performance.

The outcomes for this cross-regime aerodynamic design problem reveal the potential impact of combining non-myopic multifidelity policies with physics-awareness features for a broader application to aerodynamic design and optimization problems. PA-NM2BO could be particularly attractive for optimization problems where the use of high-fidelity models is computationally unfeasible but essential to represent the full order complexity of the aerodynamic domain in certain fluid regimes. Indeed, a multifidelity framework that is conscious about the evolution of the physics domain can improve the wise inclusion of high-fidelity data during the optimization, and a non-myopic formulation can accelerate the identification of superior design configurations and contain the total computational cost.

7.4 Space Vehicle Multidisciplinary Design Optimization

The design of a space re-entry vehicle is a multidisciplinary optimization problem that well carries the computational challenges associated with the design of complex engineering systems [37, 207, 254]. This section uses this demanding MDO application to demonstrate both the physics-aware multifidelity algorithm PA-MFBO and the non-myopic multipoint multifidelity algorithm NM3-BO, and discuss them in comparison with popular standard BO and MFBO algorithms.

Re-entry vehicles are spacecrafts conceived to enter the planetary atmosphere and safely land on the planet surface. The main objective of the re-entry mission is achieve a proper balance between the deceleration across the re-entry orbit and the thermal loads on the structural frame. Indeed, these vehicles operate in an extreme environment characterized by intense aerothermodynamic phenomena that cause significant stress on the vehicle structural frame. This space vehicle MDO problem captures the multi-physics nature of the atmospheric re-entry and involves several disciplinary analyses, namely the contributions of the propulsion system, the re-entry descend trajectory, the aerothermodynamic effects that occurs during the descend path, and the thermo-structural interaction between the re-entry flow-field and the thermal protection system. Figure 7.15 illustrates the concept of operations of the re-entry mission. This involves several phases, namely a maneuver sequence to introduce a thrust component that shapes the re-entry trajectory, the heat peak along the descent caused by the hypersonic aerothermodynamic phenomena, and the deployment of the parachutes during the landing phase. The entry phase is

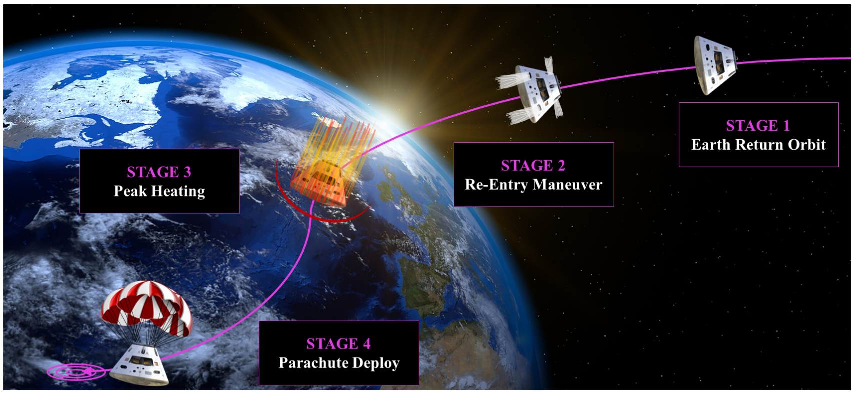


Fig. 7.15 Re-Entry Mission Concept of Operations.

characterized by a maneuver sequence to shape the descend trajectory, introducing a thrust component opposite to the direction of motion to reduce the approaching velocity, and a small normal component to calibrate the trajectory. During the re-entry flight, the vehicle is subject to significant thermal stresses determined by the largely hypersonic flow-field. Surface heating and temperatures are key design aspects that influence the design of the spacecraft thermal protection system (TPS), and influence the shape of the descend trajectory and the entry maneuver. The TPS is designed to withstand the severe re-entry heat fluxes, keeping acceptable the internal temperatures of the vehicle. At the end of descend phase, the parachutes are deployed and the vehicle lands on the planetary surface.

Figure 7.16 illustrates the design structure matrix [255] of the re-entry vehicle optimization problem. We adopt the multidisciplinary feasible architecture [46] to address the MDO problem through a single optimization procedure where the design variables and constraints are under the direct control of the optimizer. The disciplinary analyses flow follows the diagonal of the DSM, while the feed forward flows are represented on the upper triangle and the couplings between disciplines are reported on the lower side. The propulsion system is modeled according to the chemical rocket theory, and comprises primary and secondary chemical thrusters fueled by an hypergolic propellant. The trajectory solver models the descend trajectory as a bi-dimensional orbit propagated through the numerical integration of the non-linear re-entry planetary ordinary differential equations. The aerothermodynamic analysis consists of two disciplinary solvers at different levels of fidelity. The high-fidelity model simulates the full order aerothermodynamic physics through the numerical solution of the Reynolds-Averaged Navier-Stokes equations. The

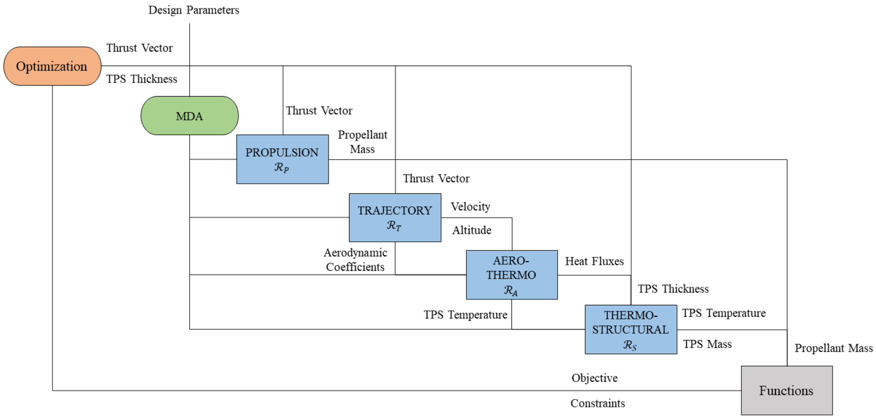


Fig. 7.16 Design structure matrix of the space vehicle MDO problem.

low-fidelity model uses the Oswatitsch Mach number independence principle jointly with the Tauber-Sutton and Sutton-Grave formulations to provide an approximated representation of the aerothermodynamic domain. The high-fidelity model requires hours of computation on an high performance computing cluster, while the low-fidelity analysis is three orders of magnitude faster on a standard computing platform. The thermo-structural analysis models the interaction between the flow-field and the structure of the Thermal Protection System (TPS) through the thermo-elastic equations. Section 7.4.1 provides more details about the disciplinary models adopted in the space vehicle MDO problem.

The design optimization problem targets the best design configuration $\boldsymbol{x} = [F_V, F_N, s_{TPS}]$ of the re-entry vehicle in terms of thrust capabilities $F = [F_V, F_N]$ and TPS structural thickness s_{TPS} that jointly minimizes the temperature T_{TPS} reached by the TPS frame, the overall structural mass m_{TPS} of the TPS, and the mass of propellant m_P burned during the re-entry maneuver. The multidisciplinary design optimization problem is formulated as follows:

$$\begin{aligned}
\text{minimize} \quad & f(\boldsymbol{x}) = 0.4\frac{m_{TPS}(\boldsymbol{x})}{m_{TPS0}} + 0.4\frac{T_{TPS}(\boldsymbol{x})}{T_{TPS0}} + 0.2\frac{m_P(\boldsymbol{x})}{m_{P0}} \\
\text{with respect to} \quad & \boldsymbol{x} = [F_V, F_N, s_{TPS}] \\
\text{subject to} \quad & 100km \leq h^*(\boldsymbol{x}) \leq 125km \\
& \mathscr{R}_P(\boldsymbol{x}) = 0 \\
& \mathscr{R}_T(\boldsymbol{x}) = 0 \\
& \mathscr{R}_A^{(l=1)}(\boldsymbol{x}) = 0 \\
& \mathscr{R}_A^{(l=2)}(\boldsymbol{x}) \leq 10^{-6} \\
& \mathscr{R}_S(\boldsymbol{x}) = 0
\end{aligned} \tag{4.11}$$

where the objectives are evaluated with reference to the baseline values for the TPS mass $m_{TPS0} = 700\ kg$ and temperature $T_{TPS0} = 1000\ K$, and for the mass of propellant $m_{P0} = 150\ kg$ derived from similar re-entry capsules [256]. The search is bounded by the move limits of the design space $\mathscr{X} = \mathscr{X}_{F_V} \times \mathscr{X}_{F_N} \times \mathscr{X}_{s_{TPS}}$, where the thrust capabilities tangential $\mathscr{X}_{F_V} = [29.2\ kN, 146\ kN]$ and normal $\mathscr{X}_{F_N} = [0.48\ kN, 2.4\ kN]$ to the trajectory are defined according to the propulsion system specifications, and the limits on the TPS thickness $\mathscr{X}_{s_{TPS}} = [0.03\ m, 0.1\ m]$ are imposed from expert knowledge. The MDO problem requires a specific range of altitudes h^* for the re-entry maneuver to simulate a real-world mission. Additional constraints include the feasibility of the physics-based models at each iteration of the optimization procedure, namely the complete resolution of the propulsion system model $\mathscr{R}_P(\boldsymbol{x}) = 0$, the trajectory model $\mathscr{R}_T = 0$, the low-fidelity aerothermodynamic model $\mathscr{R}_A^{(l=1)}(\boldsymbol{x}) = 0$, the high-fidelity aerothermodynamic model ensured reducing the computational residuals below $\mathscr{R}_A^{(l=2)}(\boldsymbol{x}) \leq 10^{-6}$, and the thermo-structural model $\mathscr{R}_S(\boldsymbol{x}) = 0$.

7.4.1 Disciplinary Models

The MDO problem of the re-entry space vehicle accounts for the interactions and couplings between the aerothermodynamics, the atmospheric flight trajectory, the

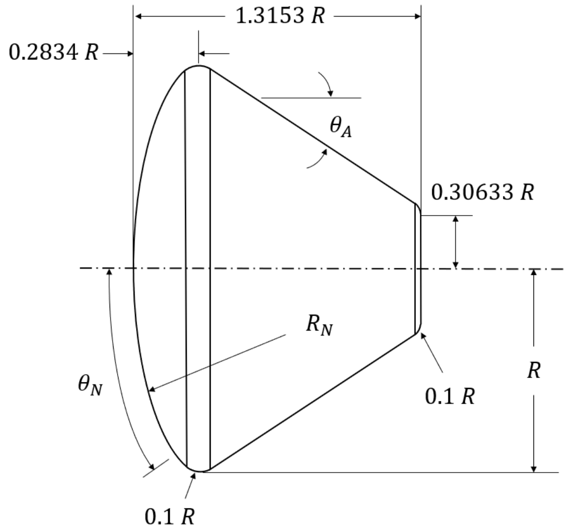


Fig. 7.17 Geometry of the Orion-like re-entry capsule.

capabilities of the propulsion system and the thermo-structural phenomena. In particular, we consider the case of an Orion-like capsule re-entering the Earth atmosphere. Figure 7.17 illustrates the geometry of the capsule ([257]) and Table 7.6 indicates the values of the geometric parameters which are considered given and not varying for our problem. The capsule is equipped with two sets of thrusters to guide the entry maneuver: the first set consists of two primary thrusters and the second set includes six secondary thrusters. Both the primary and secondary thrusters are chemical rocket engines that use an hypergolic combination of monomethyl hydrazine (MMH) as propellant and dinitrogen tetra oxide (NTO) as oxidant. Table 7.7 reports the number of thrusters, the maximum thrust in vacuum, the effective exhaust velocity and the burning time for both the primary and secondary engines, summarizing the details of the overall propulsion system. The propulsion system is modeled according to the chemical rocket theory, that accounts for an impulsive orbital maneuver considering the short burning period. The model of the trajectory considers a bi-dimensional orbit of re-entry that is propagated starting from a fixed entry point in the atmosphere. Table 7.8 summarizes the trajectory parameters of the entry point, the mass and reference area of the capsule, the model of the Earth atmosphere ([258]), and the model of the Earth gravitational field. Two representations are included for the aerothermodynamic phenomena: the first models the full order physics and the second one provides a simplified physics-based representation. The TPS is modeled approximating the vehicle to a sphere with a radius equal to the

Geometry Parameters	Description	Unit
Frontal section diameter	$R = 5.0$	m
Nose radius	$R_N = 2.4R$	m
Upperside ablator inclination	$\theta_A = 32.5$	deg
TPS aperture	$\theta_D = 23.04$	deg

Table 7.6 Design parameters of the Orion-like geometry.

Thrusters Parameters	Description	Unit
Number of primary thruster	2	-
Number of secondary thruster	6	-
Maximum thrust primary thruster (Vacuum)	$F_{max1} = 73$	kN
Maximum thrust secondary thruster (Vacuum)	$F_{max2} = 4.87$	kN
Effective exhaust velocity primary thruster	$c_1 = 2305$	m/s
Effective exhaust velocity secondary thruster	$c_2 = 2943$	m/s
Burning time	$\Delta t = 5$	s

Table 7.7 Design parameters of the primary and the secondary thrusters.

radius of the nose of the capsule; the sphere is then discretized into finite elements for the numerical solution of the thermo-elastic equations. The material chosen for the TPS structural frame is the composite zirconium diboride ultra-high temperature monolithic (UHTC ZrB2), whose properties are detailed in Table 7.9. This material has been demonstrated to be very popular for this kind of application, given the capability to withstand high temperatures ([259, 260]).

Propulsion System Model

The propulsion system model estimates the total mass of propellant demanded to complete the re-entry maneuver, and accounts for the specifications of the engines and the propulsive thrust (Table 7.7). We model the thrust vector $\boldsymbol{F} = [F_V, F_N]$ considering two major components: F_V tangential to the trajectory and F_N normal to the trajectory. Accordingly, the magnitude of the thrust vector $F = |\boldsymbol{F}|$ is defined as per the chemical rocket propulsion theory ([261]):

$$F = \dot{m}_P c \tag{7.18}$$

where $\dot{m}_P$ is the propellant mass flow rate and c is the effective exhaust velocity. Accordingly, we compute the mass of the propellant m_P as the propellant mass flow integrated over the entire burning time $\Delta = t_{off} - t_{on}$ from the beginning t_{on} to the end t_{off} of the maneuver :

$$m_P = \int_{t_{on}}^{t_{off}} \frac{F}{c} dt = \frac{F}{c} \Delta t \tag{7.19}$$

Trajectory Model

The re-entry trajectory model estimates the parameters characterizing the re-entry profile such as the descend velocity V, the flight path angle γ, the altitude during the re-entry h and the longitude angle β. The re-entry orbit is completely computed given the thrust vector $\boldsymbol{F}$, the aerodynamic force coefficients provided by either the low or high-fidelity aerothermodynamic representation, the design parameters of the trajectory (Table 7.8) including the point of the atmosphere where the entry maneuver begins, the mass M and reference area of the vehicle A_{ref}, the spherical gravitational model $g(h)$, and the model of the atmosphere ([258]).

The descend trajectory is approximated as a planar orbit assuming the planet as non-rotating and a constant flight path azimuth angle. Accordingly, the re-entry equations are formalized as follows:

$$\frac{dV}{dt} = -\frac{(D + F_V)}{M} - g \sin\gamma \tag{7.20a}$$

$$V\frac{d\gamma}{dt} = \frac{(L + F_N)}{M} L - g\cos\gamma + \frac{V^2}{(h + R_E)} \cos\gamma \tag{7.20b}$$

$$\frac{dh}{dt} = V \sin\gamma \tag{7.20c}$$

$$\frac{d\beta}{dt} = \frac{V \cos\gamma}{(h + R_E)} \tag{7.20d}$$

where t is the re-entry time, M is the mass of the vehicle, D is the aerodynamic drag, g is the acceleration of gravity, L is the aerodynamic lift and $R_E = 6.378 \cdot 10^6 m$ is the Earth radius. We use the spherical gravitational model to provide an estimate of the gravitational acceleration as a function of altitude:

Trajectory Parameters	Description	Unit
Initial entry altitiude	$h_0 = 125000$	m
Initial entry velocity	$V_0 = 7900$	m/s
Initial entry longitude	$\beta_0 = 0$	deg
Initial entry time	$t_0 = 0$	s
Capsule mass	$M = 7500$	kg
Capsule reference area	$A_{ref} = 78.54$	m^2
Atmosphere model: free stream density	$\rho_\infty = \rho_\infty(h)$	kg/m^3
Atmosphere model: free stream temperature	$T_\infty = T_\infty(h)$	K
Atmosphere model: free stream pressure	$p_\infty = p_\infty(h)$	Pa
Gravitational model	$g = g(h)$	m/s^2
Altitude for parachute deployment	$h = 5000$	m

Table 7.8 Design parameters of the re-entry trajectory.

$$g(h) = g_0 \left(\frac{R_E}{R_E + h} \right)^2 \tag{7.21}$$

where $g_0 = 9.81 m/s^2$ is gravitational acceleration at sea level. The aerodynamic forces of lift L and drag D are defined as follows :

$$L = \frac{1}{2} \rho_\infty V^2 A_{ref} C_L \tag{7.22}$$

$$D = \frac{1}{2} \rho_\infty V^2 A_{ref} C_D \tag{7.23}$$

where the free stream density $\rho_\infty(h)$ is computed as a function of the altitude through the atmosphere model, A_{ref} is the area of the mid-ship section of the vehicle; C_L and C_D are the lift and the drag coefficients, respectively, given by the aerothermodynamic models.

The trajectory parameters V, γ, h and β are computed by solving the system of non-linear ODEs (7.20) using Runge-Kutta method. The equations are integrated over the re-entry time from the time when the space vehicle enters the atmosphere t_1 until the parachute is deployed t_f – when Equation (7.20) is no longer valid since the additional resistance due to the parachute is not considered. Given the short burning time of the chemical engines, the entry maneuver is considered impulsive and the thrust components F_V and F_N are considered exclusively for the first integration step,

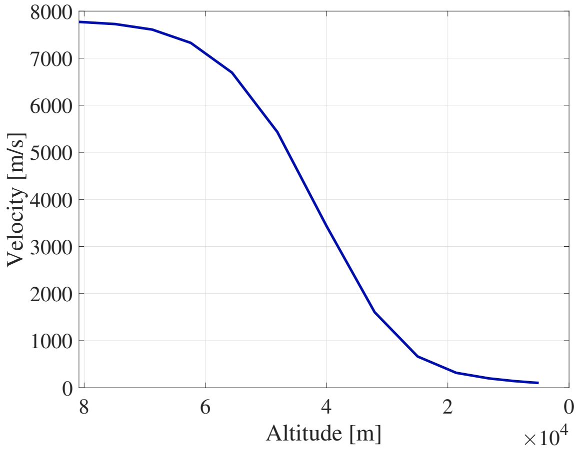


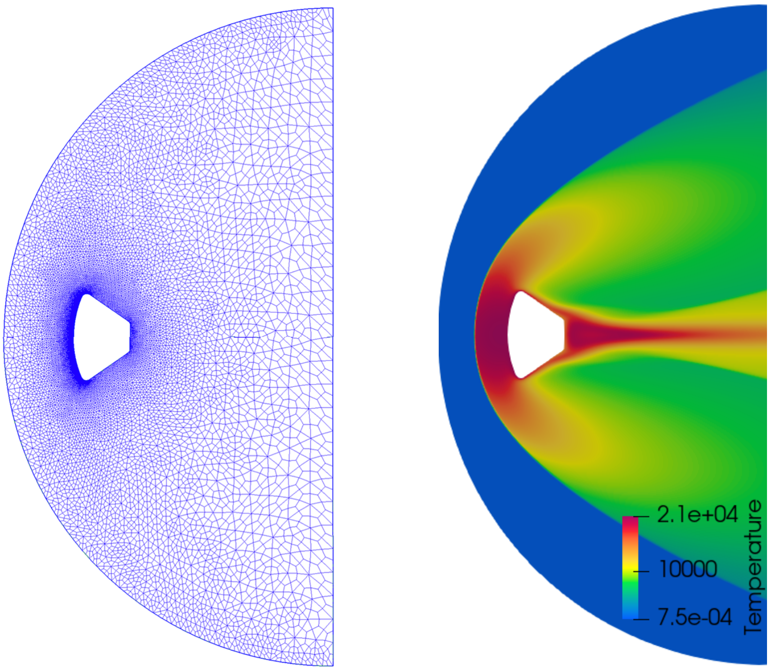
Fig. 7.18 Profile of the re-entry velocity evaluated with the re-entry trajectory model, for the case of an unpowered re-entry of the Orion-like capsule.

while are set to zero for the following time-steps. Figure 7.18 illustrates the re-entry velocity profile for the different altitudes crossed along the descend path, for the case of the unpowered re-entry of the Orion-like capsule.

High-fidelity Aerothermodynamic model

The high-fidelity aerothermodynamic model computes the stagnation point heat flux $\dot{q}$ and the aerodynamic coefficients of lift C_L and drag C_D, given the geometry parameters of the vehicle (Table 7.6), the velocity V and the altitude profile h computed with the trajectory model, the atmosphere model (Table 7.8), and the temperature of the TPS structure given by the thermo-structural model of the TPS.

We consider the full set of Reynolds Averaged Navier-Stokes equations for viscous fluid as the governing equations of the fluid domain during the atmospheric re-entry. The high-fidelity aerothermodynamic model uses the finite volume method to discretize the equations in space, with a standard edge-based data structure where the convective and viscous fluxes are evaluated at the midpoint of the edges. The computational domain is defined as a semicircle of radius equal to $6.3R_N$, where the capsule is placed on the axis of symmetry with the nose at $2.5R_N$ from the center, and is discretized with a mesh of $9.2 \cdot 10^4$ quads elements. The density of the computational grid is variable in the domain and more refined in the proximity of the nose to better capture the discontinuities of the flow field and the temperature



(a) (b)

Fig. 7.19 (a) Discretization of the computational domain with approximately $9.2 \cdot 10^4$ quad elements, and (b) temperature contours around the Orion-like capsule for an altitude of $60km$ and for a Mach number of 20.

gradients that are critical for the structure of the TPS. The boundary conditions of the computational domain include the body of the re-entry capsule and the inlet flow; the outline of the capsule is defined as a marker wall on which the temperature of the TPS structure T_{TPS} is imposed, while the inlet marker is defined in terms of the re-entry velocity V, the free stream density $\rho_\infty(h)$, the free stream temperature $T_\infty(h)$, and the free stream pressure $p_\infty(h)$. Figure 7.19(a) provides details about the discretization of the computational domain considered in the model. We use gmsh [262] to generate the computational grid of the fluid domain and SU2 [263] version 7.0.3 in RANS steady mode to solve the Reynolds-averaged Navier-Stokes equations and predict the effects of the turbulence; The RANS equations are integrated through Euler implicit and the convergence criteria are set minor than 10^{-6}. Figure 7.19(b) illustrates the temperature distribution around the Orion-like capsule, at an altitude of $h = 60\ km$ and for a Mach number of 20.

We consider the temperature distribution and the flow-field computed with this model as the high-fidelity representation of the stagnation point heat load $\dot{q}$ acting on the vehicle and of the aerodynamic coefficients C_L and C_D.

Low-fidelity Aerothermodynamic model

The low-fidelity aerothermodynamic model computes the aerodynamic coefficients of lift C_L and drag C_D, and the stagnation point heat flux acting on the TPS structure $\dot{q}$, given the re-entry flight parameters from the trajectory model, the model of the atmosphere (Table 7.8) and the geometry of the capsule (Table 7.6). This low-fidelity representation is characterized by surrogate models based on the Oswatitsch Mach number Independence principle to approximate the aerodynamic coefficients, and on Sutton-Grave and Tauber-Sutton formulations to evaluate the convective heat flux and the radiative heat flux. The Oswatitsch Mach number independence principle ([264]) is based on the inviscid representation of the re-entry flow modeled through the Euler equations. Accordingly, at large values of Mach number the inviscid flow field behind the bow shock tends to a limit condition, where the lift coefficient $C_L = 0$ and drag coefficient $C_D = 6.14$ are constant with altitude.

The total heat flux at the stagnation point transfers to the structure of the TPS through convective and radiative energy exchanges. The convective heat load is estimated according to the Sutton-Grave formulation ([265]):

$$\dot{q}_{conv} = k_s \sqrt{\frac{\rho_\infty}{R_N}} \left(\frac{V}{1000} \right)^{3.15} \tag{7.24}$$

where $k_s = 5.1564 \cdot 10^{-5}$ is a constant for the Earth atmosphere, R_N is the radius of the nose of the capsule and V is the re-entry flight velocity.

The radiative heat load is computed using the Tauber-Sutton formulation ([266]):

$$\dot{q}_{rad} = C R_N^a \rho_\infty^b f(V) \tag{7.25}$$

where $C = 4.736 \cdot 10^4$ and $b = 1.22$ are constants adopted for the Earth atmosphere, $a = 1.072 \cdot 10^6 V^{-1.88} \rho_\infty^{-0.325}$ is given in function of the descend velocity V and the density of the atmosphere $\rho_\infty(h)$, and $f(V)$ is a tabulated function of velocity.

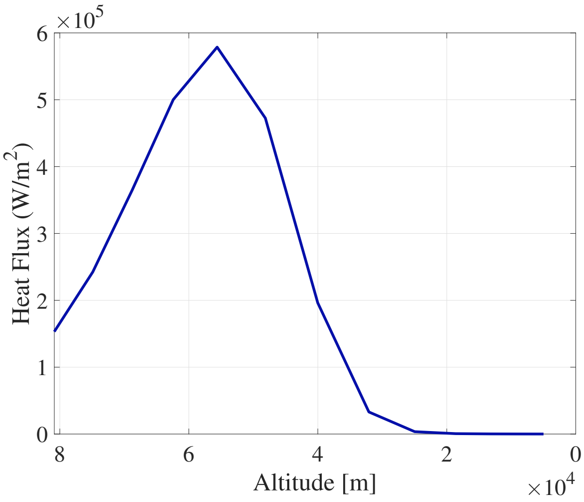


Fig. 7.20 Heat flux evaluated with the low-fidelity aerothermodynamic model, for the case of an unpowered re-entry of the Orion-like capsule.

The Sutton-Grave and Tauber-Sutton formulations give the total heat load at the stagnation point $\dot{q}$:

$$\dot{q} = \dot{q}_{conv} + \dot{q}_{rad} \tag{7.26}$$

This model constitutes the low-fidelity aerothermodynamic representation within our framework of disciplinary models.

Figure 7.20 illustrates the values of the stagnation point heat flux computed along the altitude profile of the re-entry with this low-fidelity aerothermodynamic model, for the case of an unpowered re-entry of the Orion-like re-entry capsule.

Thermo-structural Model of the Thermal Protection System

The Thermo-structural model estimates the temperature of the TPS frame T_{TPS} and the mass of the TPS structure m_{TPS}, given the total heat load $\dot{q}$ provided by either the low or the high-fidelity aerothermodynamic model, the thickness of the TPS structure s_{TPS}, the material property of the TPS (Table 7.9) and the geometry of the capsule (Table 7.6).

We model the structure of the TPS as an arc of circumference discretized into $n_e = 1000$ linear elements, approximating the re-entry capsule with a sphere of radius equal to the radius of the nose R_N. Figure 7.21(a) shows an example of the discretization of the TPS with $n_e = 4$ finite elements for illustration purposes. Figure 7.21(b) represents a generic e-th finite element of the discretized TPS where η is the local references axis, $\hat{\boldsymbol{n}}_e$ is the versor normal to the element, 1^e and 2^e are the nodes of the element, l_e is the length of the element, and $\hat{\boldsymbol{w}}$ is the versor representing the direction of the heat flux vector $\dot{\mathbf{q}} = \dot{q}\hat{\boldsymbol{w}}$. Accordingly, the heat equation is specialized for the generic e-th finite element and linearized considering uniform the thermal conductivity κ_{TPS} and the thickness of the TPS structure s_{TPS}:

$$\rho_{TPS} c_P s_{TPS} \frac{dT}{dt} - \kappa_{TPS} s_{TPS} \frac{\partial^2 T}{\partial \eta^2} + 4\sigma\varepsilon_{TPS} T_\infty^3 T - 4\sigma\varepsilon_{TPS} T_\infty^4 - \dot{q}_s = 0 \tag{7.27}$$

where ρ_{TPS}, c_P and ε_{TPS} are respectively the density, the specific heat at constant pressure and the emissivity coefficient of the TPS material, σ is the Stephan-Boltzmann constant, $\dot{q}_s$ is the heat source term and $T_\infty(h)$ is the temperature of the atmosphere. We use the Galerkin method over the discretized structural domain of the TPS to numerical solve Equation (7.27). Thus, the associated weak formulation is defined as follows:

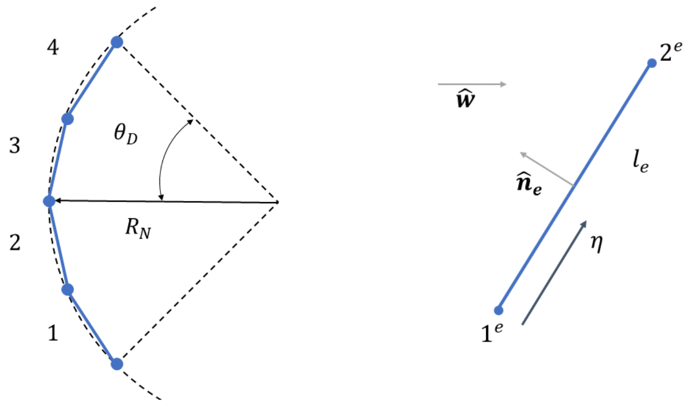
$$f(T(\eta,t)) = 0 \tag{7.28}$$

The temperature along each element is defined using the technique of the separation of variables:

$$T(\eta,t) = \mathbf{N}(\eta)\boldsymbol{\Omega}_e \tag{7.29}$$

where $\mathbf{N}(\eta)$ are the linear shape functions and $\boldsymbol{\Omega}_e = [\Omega_{1^e}, \Omega_{2^e}]$ are the nodal temperatures of the e-th element. The problem is formulated with the Galerkin method as follows:

$$\int_0^{l_e} \mathbf{N}^T(\eta) f(T(\eta,t))\, d\eta = 0 \tag{7.30}$$



(a) (b)

Fig. 7.21 (a) Example of the discretization of the TPS with 4 elements, and (b) details of the *e*-th finite element

TPS Parameter	Description	Unit
TPS mass density	$\rho_{TPS} = 6000$	kg/m^3
TPS specific heat (constant pressure)	$c_P = 628$	J/kgK
TPS emissivity	$\varepsilon_{TPS} = 0.9$	$-$
TPS thermal conductivity ($T = 300K$)	$\kappa_{TPS} = 58$	W/mK
TPS thermal conductivity ($T = 1300K$)	$\kappa_{TPS} = 64$	W/mK
TPS thermal conductivity ($T = 2300K$)	$\kappa_{TPS} = 134$	W/mK
TPS maximum temperature	$T_{max} = 2273.15$	K

Table 7.9 Design parameters of the thermal protection system.

where $\dot{q}_s = \dot{\mathbf{q}} \cdot \hat{\boldsymbol{n}}_e$ is the initial condition. Problem (7.30) is a system of ODEs where the nodal temperatures $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, ..., \boldsymbol{\Omega}_{n_e}]$ are the unknowns computed via Crank-Nicolson method.

Among them, the nodal temperature at the stagnation point $\Omega_{stag} = \max(\boldsymbol{\Omega})$ is the most stressfull for the frame and we consider it as the temperature of the TPS structure T_{TPS}. Figure 7.22 illustrates the profile of the TPS temperature T_{TPS} as a function of the re-entry altitudes for the case of unpowered re-entry of the Orion-like capsule.

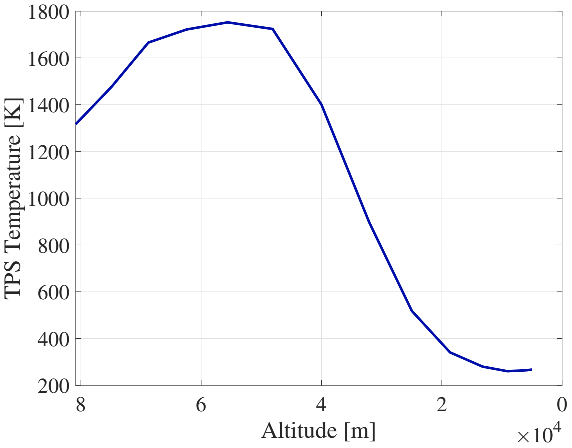


Fig. 7.22 Temperature profile of the TPS structure evaluated with the thermo-structural model, for the case of an unpowered re-entry of the Orion-like capsule

The model of the thermal protection system computes the mass of the structural frame of the TPS m_{TPS}:

$$m_{TPS} = \rho_{TPS} S_{TPS} s_{TPS} \tag{7.31}$$

Where S_{TPS} is the frontal surface of the spherical shell that approximates the structure of the TPS, given by the area of the circle with radius equal to the radius of the nose of the capsule R_N.

7.4.2 NM3-BO Results and Discussion

This section investigates the performance of the proposed Non-Myopic Multipoint Multifidelity algorithm NM3-BO for the MDO problem of the re-entry space vehicle. We compare our methodology against standard Multifidelity Bayesian Optimization frameworks: all those MFBO algorithms rely on the multifidelity Gaussian process surrogate model (Section 3.3.2) and implement different formulations of the acquisition function, including the Multifidelity Expected Improvement [38] (MFEI), Multifidelity Max-value Entropy Search [42] (MFMES), and Multifidelity Probability of Improvement [41] (MFPI). In addition, we report the outcomes achieved with the Efficient Global Optimization [109] (EGO) algorithm using only high-fidelity

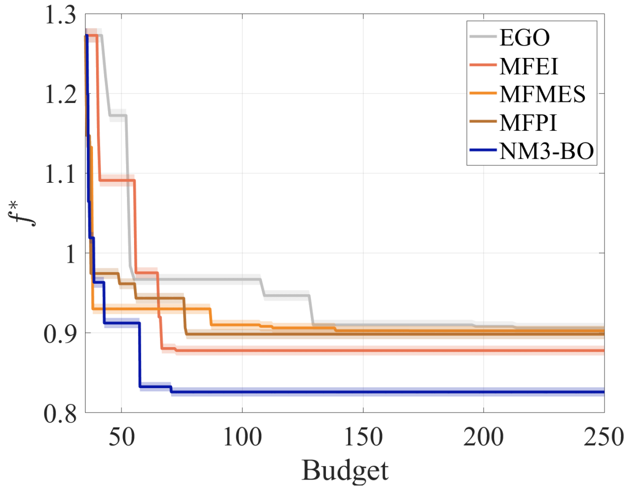


Fig. 7.23 Statistics over 25 experiments of the minimum of the objective function f^* obtained with the competing algorithms.

queries to provide a comparison with a popular single fidelity Bayesian optimization methodology.

The performance of the competing algorithms are evaluated in terms of the minimum of the objective function $f^*(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\mathscr{X}} f(\boldsymbol{x})$ as a function of the computational budget $B = \sum \lambda_i^{(l)}$ at each iteration i of the optimization procedure. The computational costs for the aerothermodynamic analyses are imposed at $\lambda^{(2)} = 1$ for the high-fidelity model and $\lambda^{(1)} = 0.001$ for the low-fidelity model; these specific values reflect the time required to complete the aerothermodynamic simulation adopting either the CFD solver or the low-fidelity formulations. We consider a statistics over 25 experiments for each algorithm, and initialize the searches with random initial samples collected through a Latin hypercube sampling scheme. This experimental methodology permits to quantify the influence of different initialization on the algorithms performance. In particular, the multifidelity algorithms are initialized with 1000 design configurations evaluated with the low-fidelity model, and 34 design points computed with the high-fidelity analysis. The single-fidelity algorithm starts the search with an initial set of 35 designs evaluated with the high-fidelity model.

Figure 7.23 illustrates the median values (solid line) of the minimum of the objective function f^* along with the observations falling between the 25-th and 75-th percentiles (shaded area). At the beginning of the optimization, all the competing algorithms relies on a set of initial samples that corresponds to design configurations

B	f^*_{EGO}	f^*_{MFEI}	f^*_{MFMES}	f^*_{MFPI}	f^*_{NM3-BO}
50	1.172 (-17)	1.091 (-9.1)	0.930 (6.9)	0.961 (3.8)	0.912 (8.7)
75	0.967 (3.3)	0.877 (12.2)	0.930 (6.9)	0.943 (5.6)	0.826 (17.2)
100	0.967 (3.3)	0.877 (12.2)	0.910 (8.9)	0.898 (10.1)	0.826 (17.2)
150	0.910 (8.9)	0.877 (12.2)	0.902 (9.7)	0.898 (10.1)	0.826 (17.2)
200	0.908 (9.1)	0.877 (12.2)	0.902 (9.7)	0.898 (10.1)	0.826 (17.2)
250	0.906 (9.3)	0.877 (12.2)	0.902 (9.7)	0.898 (10.1)	0.826 (17.2)

Table 7.10 Median values of the minimum of the objective function f^* and corresponding percentage design improvement $(\cdot)\%$ obtained with the competing algorithms.

worse than the baseline solution ($f > 1$), and all the methodologies are capable to identify design configurations superior to the baseline design solution within the maximum computational budget available $B = 250$. It is possible to observe that the multifidelity algorithms outperform the single-fidelity EGO both in terms of computational efficiency and design improvement. This indicates that the combination of disciplinary responses from multiple models allows to efficiently explore the design space and contain the computational cost. However, we notice that the NM3-BO algorithms achieves remarkable accelerations of the MDO procedure, and obtains superior design solutions – larger reductions of the objective function – with a fraction of the computational cost required by the competing algorithms to identify suboptimal designs. This outcome suggests that the combination of the non-myopic scheme and the multiple decision making process capitalizes from the design evaluations adopting different sources of information, and effectively accelerates the search toward optimal design solutions.

To further highlights the advantages of the proposed methodology, Table 7.10 summarizes the median values of f^* for discrete values of the computational budget B. We can observe that the NM3-BO achieves the higher design improvement (8.77%) after the consumption of a budget $B = 50$, whereas the other strategies score worst in terms of design upgrades. A remarkable outcome is the overall acceleration of the MDO procedure provided by NM3-BO: our framework converges for a computational budget below $B = 75$ and leads to a design upgrade of the 17.4%. This result is outstanding if compared with the design improvement of about the 10% obtained by the EGO, MFMES, and MFPI algorithms, and the design upgrade around the 12% achieved by the MFEI at convergence.

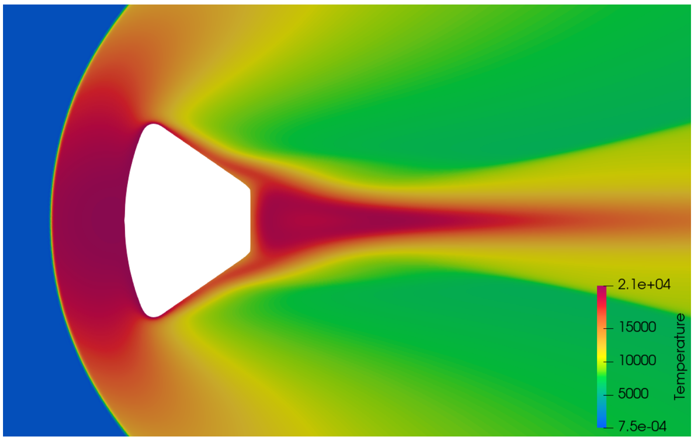


Fig. 7.24 Temperature contours at the heat peak condition evaluated with the high-fidelity aerothermodynamic model considering the best design solution achieved with our NM3-BO algorithm.

Method	$f^*(\boldsymbol{x}^*)$	$\boldsymbol{x}^* = [F_V^*, F_N^*, s_{TPS}^*]$	m_{TPS}^*	T_{TPS}^*	m_P^*
EGO	10.0 %	$[33.63\,kN, 0.969\,kN, 0.0396\,m]$	$476\,kg$	$1320\,K$	$74.6\,kg$
MFEI	12.8 %	$[35.67\,kN, 1.561\,kN, 0.0341\,m]$	$410\,kg$	$1326\,K$	$80.0\,kg$
MFMES	10.3 %	$[35.97\,kN, 2.046\,kN, 0.0373\,m]$	$448\,kg$	$1329\,K$	$81.5\,kg$
MFPI	10.7 %	$[35.40\,kN, 0.691\,kN, 0.0377\,m]$	$453\,kg$	$1322\,K$	$77.9\,kg$
NM3-BO	17.9 %	$[29.53\,kN, 0.807\,kN, 0.0304\,m]$	$365\,kg$	$1310\,K$	$65.4\,kg$

Table 7.11 Comparison between the best design solutions identified with the competing algorithms.

Table 7.11 compares the best design solutions obtained with the competing algorithms over the collected experiments. NM3-BO identifies an optimal design configuration of the re-entry vehicle that delivers an upgrade of the 17.98%, and privileges lower thrust capabilities and contained thickness of the thermal protection system. This determines a lower storage of propellant $m_P^* = 65.45\,kg$ on-board and permits to navigate a safe re-entry trajectory that contains the heat loads affecting the frame. As a result, the temperature of the TPS structure is kept below $T_{TPS}^* = 1310\,K$ with a total TPS structural mass of $m_{TPS}^* = 365.17\,kg$. Figure 7.24 provides details about the temperature distribution achieved adopting the best re-entry capsule computed with the NM3-BO at the heat peak condition. It should be noticed that all the design solutions identified by the algorithms prioritize the reduction of both the TPS and propellant mass, and penalize the temperature reached by the heat shield. On one hand, this permits to contain the overall mass of the vehicle with consequent savings in terms of launch costs; on the other hand, the temperature peaks experienced by the structural frame are far below the thermal properties of the TPS material: this guarantees the survival of the vehicle during the atmospheric descent.

7.4.3 Physics-Aware Utility Function for the Space Vehicle MDO

The physics-aware utility function α_4 (Equation (5.1)) is specifically conceived to incorporate a learning bias that guides the selection of the aerothermodynamic model to query – at a certain level of fidelity – according to the severity of the aerothermodynamic phenomena. This is achieved introducing a learning bias dependent on the re-entry flight altitude h as the trajectory parameter sensitive to the heat loads affecting the TPS structure. Accordingly, $\alpha_4(h,l)$ is formalized as follows:

$$\alpha_4(h,l) = \begin{cases} 1 & \text{if } \; l < L \\ 200\frac{h}{h_0} & \text{if } \; l = L \wedge h \in H \\ 1 & \text{if } \; l = L \wedge h \notin H \end{cases} \tag{7.32}$$

where $H = [35km, 65km]$ and $h_0 = 50km$. The range of altitudes H represents specific conditions of the re-entry trajectory where the thermal loads acting on the heat shield achieve their maximum values, as per real-world data measured during the atmospheric re-entry of capsules and probes [267–269]. The goal is to encourage the use of the high-fidelity aerothermodynamic model $l = L$ in this range of altitudes: this permits to accurately estimate the thermal effects and the heat loads since the temperatures are more likely to be critical for the survivability of the TPS material. This motivates the choice to set α_4 to achieve values much more larger than 1 for $l = L$, when the capsule navigates the risky range of altitudes, encouraging the selection of the high-fidelity model to compute the heat fluxes affecting the capsule.

7.4.4 PA-MFBO Results and Discussion

This section discusses the results obtained with our physics-aware multifidelity algorithm (PA-MFBO) for the space vehicle MDO problem. The PA-MFBO methodology is compared to the single fidelity-Bayesian framework based on the efficient

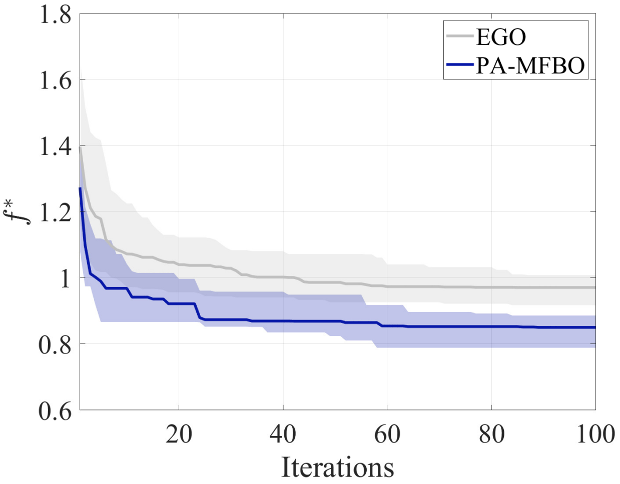


Fig. 7.25 PA-MFBO and EGO statistics.

global optimization formulation ([220]) eliciting evaluations of the low fidelity aerothermodynamic model only.

Figure 7.25 illustrates the convergence of the algorithms in terms of minimum of the objective function $f^* = \min(f(\mathbf{x}))$ obtained for 100 iterations of both the EGO and the multifidelity PA-MFBO optimization strategies. The EGO experiments consist of 50 tests initialized with different $n = 200$ initial design configurations; the PA-MFBO results are obtained for 20 searches each starting with different $n = 200$ combinations of design variables, where $n = 198$ designs are evaluated with the low-fidelity aerothermodynamic model and $n = 2$ designs are observed with high-fidelity evaluations. The results of the statistics are reported in terms of the median values (solid line) of f^* together with the observations falling in the interval between the 25-th and 75-th percentiles (shaded area), and Table 7.12 reports the median that characterize the distribution of the multifidelity and single-fidelity tests at i=1, 5, 25, 50 and 100.

At the beginning of the MDO procedure, all the initial design configurations score worse than the baseline re-entry vehicle $(f^*(\mathbf{x}) > 1)$, and both the EGO and PA-MFBO experiments progressively learn from the physics-based models to search for improved design solutions. Our PA-MFBO algorithm identifies superior design configurations and achieves larger improvement with contained resource expenditure. Specifically, it is possible to notice that the PA-MFBO permits to identify better design solutions with respect to the baseline design after only $i = 5$ iterations for

i	PA-MFBO Median	EGO Median
$i = 1$	1.2730	1.396
$i = 5$	0.9894	1.178
$i = 25$	0.8732	1.037
$i = 50$	0.8687	0.9859
$i = 100$	0.8499	0.9705

Table 7.12 Comparison between PA-MFBO and EGO median values of the minimum of the objective function.

Method	$\mathbf{x} = \{F_V, F_N, s_{TPS}\}$	m_P	T_{TPS}	m_{TPS}
EGO	$\mathbf{x} = \{70.52kN, 2.098kN, 0.0320m\}$	$156.6kg$	$1850K$	$384.3kg$
PA-MFBO	$\mathbf{x} = \{41.79kN, 1.621kN, 0.03057m\}$	$93.41kg$	$1476K$	$367.4kg$

Table 7.13 Comparison between the best design solutions evaluated with the PA-MFBO and the EGO algorithms.

a fraction of experiments, and after $i = 25$ iterations for all the experiments. In contrast, only a fraction of the EGO experiments achieves improved designs even at $i = 100$. On average, after 100 iterations, our PA-MFBO strategy achieves a design improvement of about 15% with respect to the baseline, whereas the EGO optimization obtains design upgrades of about 3%. The best design configuration identified with the EGO algorithm scores $f^* = 0.9169$, which corresponds to a design improvement of 8%. The best result obtained with the PA-MFBO algorithm is $f^* = 0.7905$, corresponding to a design improvement of 21% with respect to the baseline. Table 7.13 compares the two best design solutions identified with the EGO and PA-MFBO optimization strategies.

Figure 7.26 illustrates the space of the objectives and its projections to illustrate the search sequence corresponding to the PA-MFBO optimization experiment that identifies the best design configuration, which corresponds to a space vehicle characterized by a TPS thickness of $0.03057m$ whose propulsion system can generate $41.79kN$ of tangential thrust and $1.621kN$ of normal thrust. This design solution is characterized by an overall TPS structural mass of $367.4kg$ that permits to keep the temperature of the TPS structure below $1476K$ and requires $93.41kg$ of propellant mass to complete the re-entry maneuver. It is possible to observe that the initial exploration phase corresponds to high values of the TPS temperature: this corresponds to the evaluations of the design configurations with the low-fidelity

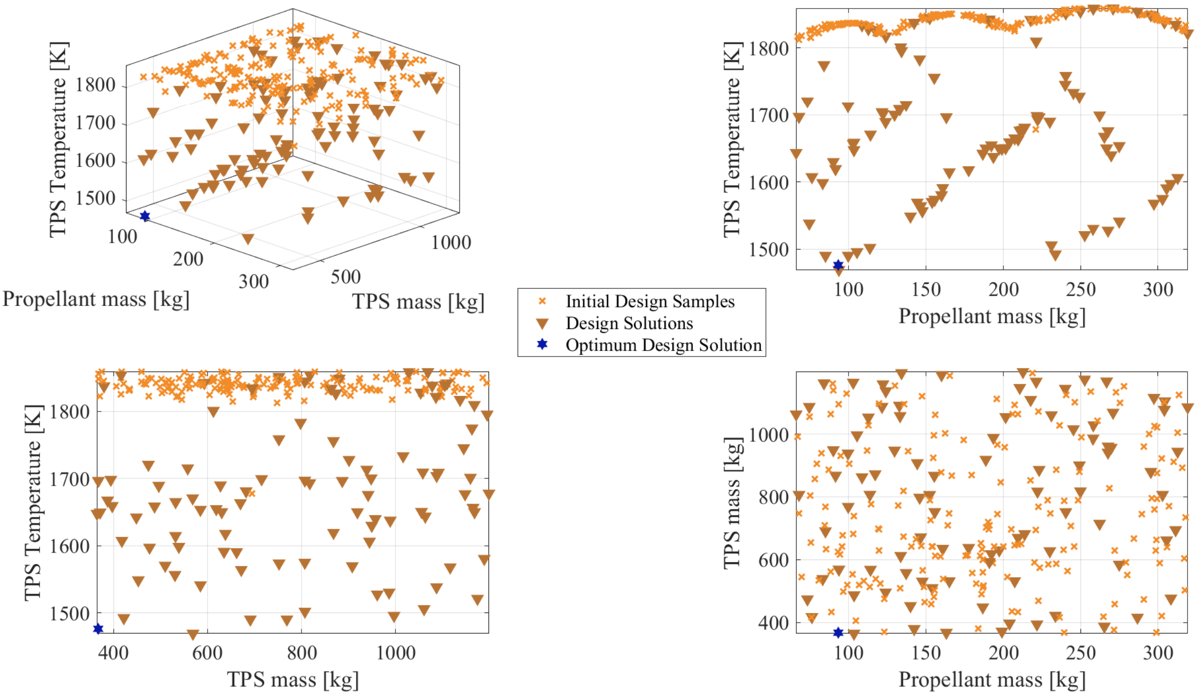


Fig. 7.26 Space of the objectives related to the best PA-MFBO test.

aerothermodynamic model that overestimates the thermodynamic phenomena. The search progressively capitalizes the knowledge gained through a limited number of high-fidelity aerothermodynamics data towards improved (lower) values of all the components of the objective functions, with particular benefits in terms of reduction of the temperature of the TPS structure. It is interesting to note that the search structures can be observed in the $T_{TPS} - m_P$ plane: at first, the higher m_P values are explored in favor of lower T_{TPS}; then, three main paths are explored which jointly reduce m_P and T_{TPS}.

Figure 7.27 illustrates the design space and its projections to highlight the sampling sequence of the corresponding PA-MFBO experiment in Figure 7.26. The initial set of design configurations consists of 200 samples obtained through a Latin Hypercube sampling of the design space. Each design point requires the evaluation of all the time-steps of the re-entry trajectory. The subsequent search targets the identification of an optimal design configuration through a continuous trade-off between the exploration of the design space and exploitation of information to identify better (eventually optimal) design solutions

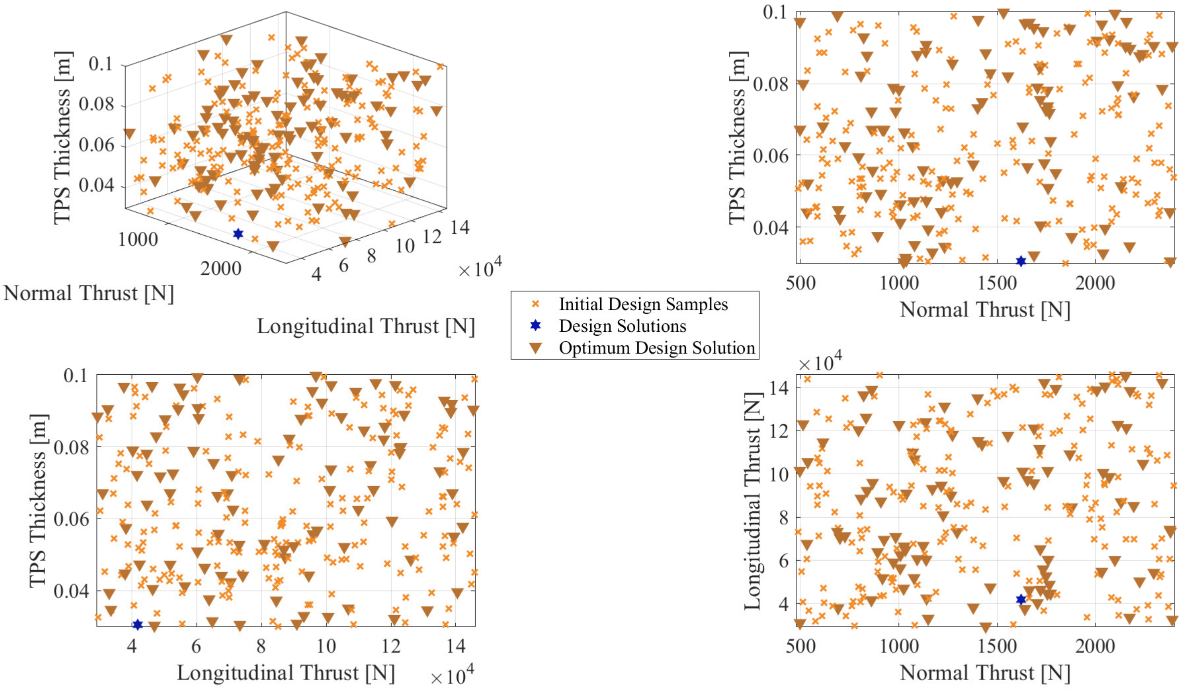


Fig. 7.27 Space of the design variables related to the best PA-MFBO test.

Figure 7.28 shows the evolution of the heat flux at the stagnation point of the TPS structure over the re-entry altitude for the best design solution identified by the PA-MFBO algorithm. The diagram focuses on the results obtained for an altitude below $85km$ where the hypothesis of continuous flow holds and RANS provide an accurate estimate of the aerothermodynamic phenomena. The maximum heat load of $\dot{q} = 55610W/m^2$ is achieved at $h = 57.87km$ and is computed with the high-fidelity aerothermodynamic model. The PA-MFBO method incentives the evaluation of the aerothermodynamic phenomena through the high-fidelity model for three points of the trajectory which correspond to altitudes where the re-entry conditions are critical for the survivability of the capsule. This illustrates the role of the physics-aware utility function α_4 that characterize the particular formulation of our multifidelity acquisition function: it allows to enrich the low-fidelity aerothermodynamics information with expensive simulations capturing the need for higher fidelity information when the heat load becomes critical for the survivabilty of the vehicle.

The results discussed in this section are obtained running groups of 4 experiments in parallel. We run each test on a single core of a desktop PC with Intel

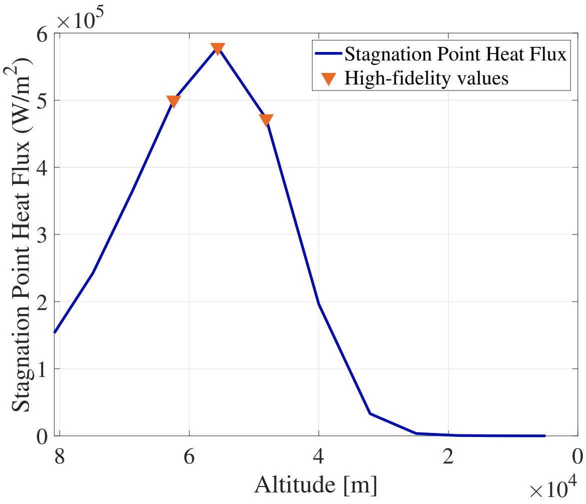


Fig. 7.28 Stagnation point heat flux with altitude outcoming from the best PA-MFBO analysis.

Core i7-8700 (3.2 GHz) and 32 GB of RAM. A single iteration of the PA-MFBO optimization takes approximately 3.6×10^2 *min*, while each iteration of the EGO algorithm based on the low-fidelity aerothermodynamics takes about 2*min*. To assess the computational time possibly associated with the single fidelity Bayesian optimization based on high-fidelity evaluations only, consider that a single evaluation of the aerothermodynamic model in Section 7.4.1 takes about 150*min* on the same computing platform. This would amount to about 150×22 *min* for a single HF only iteration with a re-entry trajectory discretized into 22 stages. This considerations remark the need for computational strategies to accelerate the MDO procedure while efficiently including high-fidelity simulations. The proposed PA-MFBO leverages a physics-aware formulation of the multifidelity acquisition function that captures the expert knowledge about the range of altitudes where the heat loads are expected to be critical. Accordingly, the algorithm prioritizes the interrogation of the costly aerothermodynamic model when higher fidelity estimations are essential contributing to both accelerate the identification of superior design and reduce the demand for computational resources.

Chapter 8

Experiments and Discussion: Diagnostics Applications

In this chapter, we apply the proposed methodologies to aerospace engineering diagnostics optimization problems and demonstrate their performance. Specifically;

- Section 8.1 investigates the capabilities of the PA-MFBO methodology for a structural health monitoring problem of a composite plate for aerospace structures.
- Section 8.2 demonstrates and validates the capabilities of the FREEDOM and MF-FREEDOM computational frameworks for the diagnostics of aerospace electromechanical actuators for flight controls systems.

8.1 Structural Health Monitoring of a Composite Plate

This structural health monitoring problem requires the assessment of the health status of a composite skin plate of an aircraft wing. Particular attention is dedicated to the incipient fracture of the carbon fiber: this represents one of the most critical failure for laminates since involves the degradation of the mechanical properties of the material and cannot be easily detected by standard non-destructive health monitoring techniques [270]. For this application, the expert knowledge about the physics relates to specific structures of the domain characterized by damage conditions that might be misinterpreted by simplified modeling approaches. We adopt our physics-aware multifidelity Bayesian optimization framework (Chapter 5) to address this structural health monitoring problem, and include the expert knowledge about the problem in the search procedure through a physics-aware utility function (Section 5.1) that biases the query of numerical models to accurately distinguish the actual fault condition affecting the plate.

The structural health monitoring problem demands for the identification of the damage parameters affecting a composite plate subject to a cut in the fibers. The composite plate is constituted of four layers of plain weave fabric of carbon prepreg (IM7/8552 AS4) laminated with a stacking sequence $[45^\circ/0^\circ/0^\circ/45^\circ]$, and with dimension of 102 mm transversal length, 456 mm longitudinal length, and 0.76 mm thickness of each ply. The material properties for the IM7/8552 AS4 considered in this application are reported in the data sheet published from the national center for advanced materials [271]. To reproduce an operational condition, we consider a load applied along the major dimension of the plate which represents a simplified load condition of a wing panel during the flight. The damage consists in a cut of the fibers along the transversal direction in the third layer, and is selected to simulate a critical condition where the fault involves the layer with 0° orientation that mainly contributes to support the load.

The health status of the system is represented through different damage parameters $\boldsymbol{x} = [x_1, x_2, x_3, x_4]$, including the transversal x_1 and longitudinal x_2 position of the cut, the extension of the cut x_3 along the transversal direction, and the load x_4 acting on the structure. Accordingly, the health monitoring task aims at identify the health status of the composite plate minimizing the discrepancy f between a real-world signal measured from the real system and the same signal computed evaluating a structural numerical model. For this procedure, we adopt the strain field $\boldsymbol{y}$ as the output signal to determine the health status of the structure: this signal is sensitive to failures in the fibers and can be easily measured in real-world applications and in laboratory. Formally, this health monitoring task is formulated as an inverse optimization problem:

$$\boldsymbol{x}^* = \min_{\boldsymbol{x} \in \mathscr{X}} f(\boldsymbol{x}) \tag{8.1}$$

where the discrepancy function $f(\boldsymbol{x}) = RMSE\left(\boldsymbol{y}_{ref}(\boldsymbol{x}), \boldsymbol{y}_{mon}^{(l)}(\boldsymbol{x})\right)$ is the root mean square error between the reference strain field $\boldsymbol{y}_{ref}(\boldsymbol{x})$ measured from the real system and the strain field $\boldsymbol{y}_{mon}^{(l)}(\boldsymbol{x})$ computed with the l-th level of fidelity numerical model. The domain of the fault parameters $\mathscr{X} = I_{x_1} \times I_{x_2} \times I_{x_3} \times I_{x_4}$ bounds the transversal $I_{x_1} = [0, 102]$ mm and longitudinal $I_{x_2} = [0, 456]$ mm position of the cut according to the maximum dimensions of the plate, while the intervals for the length of the cut $I_{x_3} = [0, 30]$ mm and the load $I_{x_4} = [0, 20]$ N are imposed from the expert knowledge about the specific structural health monitoring problem.

8.1.1 Structural Models

The strain field $\boldsymbol{y}$ of the composite plate is modeled through the Reissner-Mindlin plate equations [272] and numerically solved adopting the Finite Element Method (FEM). The structural modeling approach depicts the composite material of the undamaged structure as an orthotropic material assigning the properties of the carbon prepreg IM7/8552 AS4, and the cut in the fiber is modeled as an homogeneous material with the mechanical properties of the matrix. The boundary conditions impose a clamp in the lower section and a displacement in the upper portion of the plate with a region extended for the 10% of the total longitudinal length. This represents a simplification of the aerodynamic load acting on a composite panel adopted for the skin of an aircraft wing.

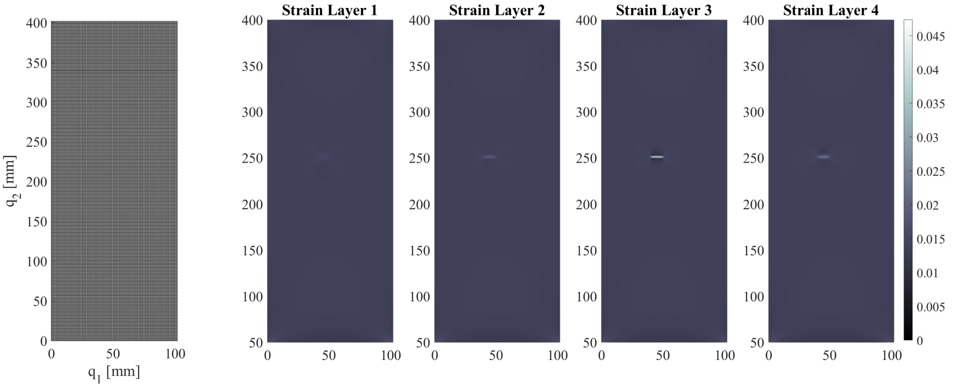
We use the software MSC Patran and MSC Nastran to develop two FEM models and compute the strain field of the damaged composite plate at different levels of fidelity. The high-fidelity model consists of a three-dimensional representation of the structure discretized through an adaptive grid of HEXA8 3D elements characterized by a dimension of 1 mm in both the longitudinal and transversal direction near the cut region thought the border, and an increasingly coarse discretization away from the cut. This permits to capture the variation of the strain field that occurs in a small region near the cut with an high level of accuracy, while containing the overall computational cost reducing the number of elements far from the damaged location. The thickness of the plate is modeled inserting three HEXA8 elements for each of the four layers along the thickness direction to further enhance the accurate representation of the strain field. The cut is represented as a rectangular parallelepiped in the third layer characterized by a transversal extension discretized with HEXA8 elements, longitudinal extension equal to one element and thickness of three elements.

The low-fidelity model approximates the composite plate through a two-dimensional representation discretized using an adaptive mesh of QUAD4 elements with transversal dimension of 2 mm and longitudinal dimension of 4 mm around the cut, and progressively increases the coarseness towards the boundaries of the plate. The cut is modeled through the same methodology of the accurate numerical model.

The high-fidelity model provides a reliable representation of the strain field as a result of the refined computational grid near the damage. This guarantees an high sensitivity to small incipient faults for which the variation of the strain field occurs in a contained region around the cut. In addition, this model allows to distinguish variations in the strain field caused by the application of intense loads in presence of a small cut in fiber – which leads to a significant variation of the strain field even in regions far from the damage – from an extended cut of the fiber – which produces large strains in an extended region due to the size of the damage. We consider the high-fidelity structural model as an emulator of the real-world composite plate that is adopted to compute the reference strain field $\boldsymbol{y}_{ref}(\boldsymbol{x})$, and is used as the highest level of fidelity available to evaluate the monitoring signal $\boldsymbol{y}_{mon}^{(L=2)}(\boldsymbol{x})$.

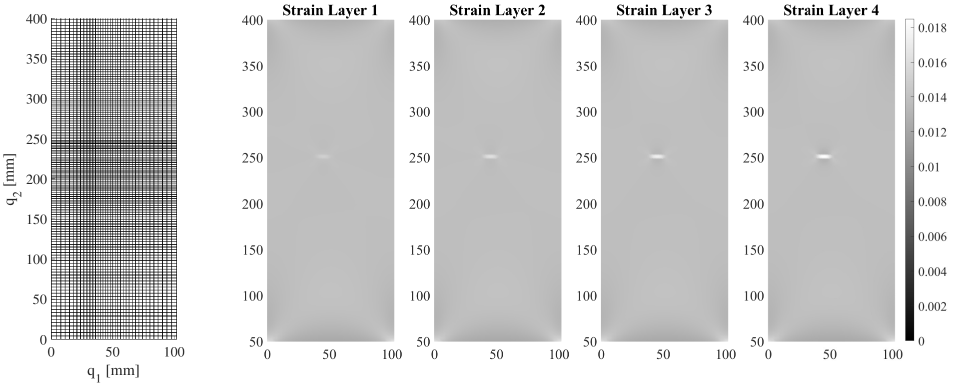
The low-fidelity representation reduces the computational burden if compared with the high-fidelity model, and achieves a satisfactory accuracy of the strain field prediction for damages characterized by an extended cut in the fiber. However, the coarse discretization entails an inaccurate evaluation of the strain field for small incipient damages of the composite plate, and fails in differentiating the increase of the strain associated with small cuts in presence of significant loads from extended damages in the fiber. This results in an approximated representation of the monitoring strain signal $\boldsymbol{y}_{mon}^{(l=1)}(\boldsymbol{x})$.

Figure 8.1 and Figure 8.2 illustrate the computational mesh and strain field over the four layers of the plate computed with the high-fidelity and low-fidelity structural model, respectively. These results are achieved for a cut in the fiber of the third layer located horizontally at 40 mm and vertically at 250 mm considering a cut length of 10 mm and load equal to 5 N.



(a) (b)

Fig. 8.1 (a) high-fidelity discretization of the computational domain, and (b) high-fidelity strain distribution for the four layers of the damaged composite plate.



(a) (b)

Fig. 8.2 (a) low-fidelity discretization of the computational domain, and (b) low-fidelity strain distribution for the four layers of the damaged composite plate.

8.1.2 Physics-Aware Utility Function for Structural Health Monitoring

The physics-aware utility function is conceived to incorporate expert knowledge about the appropriate structural model to be evaluated in presence of a small incipient cut concurrently with a significant load condition, or an extended damage in the fiber of the composite plate. This is realized through a bias in the search procedure $\boldsymbol{\psi} = [x_3, x_4]$ that encodes the specific structure of the domain, and is induced by the length of the fiber cut x_3 and the load applied on the plate x_4. Thus, we formalize $\alpha_4(x_3, x_4, l)$ for the health monitoring problem as follows:

$$\alpha_4(x_3, x_4, l) = \begin{cases} 1 & \text{if} \quad l = 1, ..., L-1 \\ 0.5\frac{x_{3max}}{x_3} + 0.5\frac{1}{x_{4max} - x_4} & \text{if} \quad l = L \end{cases} \tag{8.2}$$

where x_{3max} is the maximum length of the cut in the fiber and x_{4max} is the maximum load applied on the plate. This physics-aware utility function realizes a sort of expert reasoning and privileges the evaluation of the high-fidelity structural model for small incipient damages and high load values. Indeed, $\alpha_4(x_3, x_4, l)$ increases the value of the multifidelity acquisition function (Equation (5.1)) when the health status of the plate affected by an incipient cut and high load condition is evaluated with the high-fidelity numerical model. This ensures an accurate estimate of the strain field and permits to distinguish a narrow cut that generates large variations of the strain field amplified by high loading conditions form the magnification of strains generated by an extended cut in the fiber.

8.1.3 PA-MFBO Results and Discussion

This section reports and discusses the results achieved with the PA-MFBO framework for the structural health monitoring example to evaluate the capabilities of the algorithm against a damage identification problem. The outcomes of the PA-MFBO are compared against the efficient global optimization (EGO) algorithm [109] and the multifidelity Bayesian optimization implementing the multifidelity expected improvement (MFEI) [38]. To assess the performance of the optimization algorithms on this test case, we compute the following assessment metrics:

$$e(x_i) = \frac{|x_i^* - x_i|}{x_i^*} \cdot 100 \tag{8.3}$$

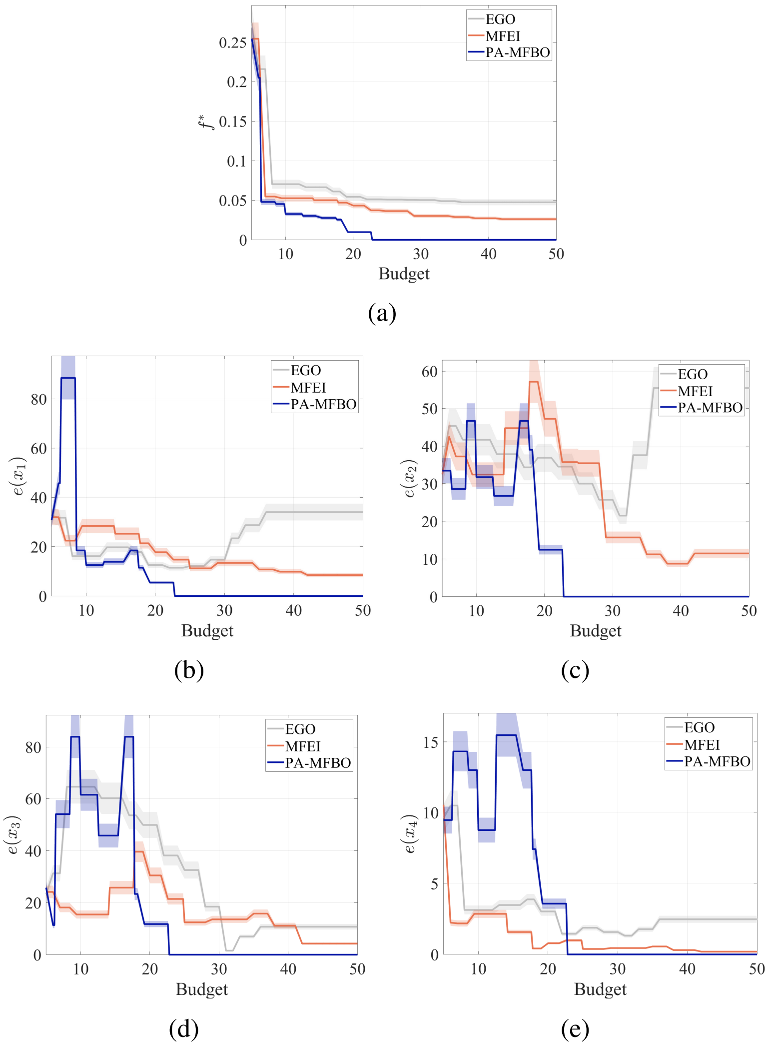
$$f^* = \min(f(\boldsymbol{x})) \tag{8.4}$$

where x_i^* is the actual level of damage that affects the composite plate, x_i is the level of damage inferred by the algorithm considering the i-th fault parameter, and

$f(\boldsymbol{x})$ is the value of the discrepancy between the reference strain signal and the strain field computed with the high-fidelity model. The percentage relative error $e(x_i)$ quantifies the accuracy related to the identification of the faults parameters, and f^* represents the minimum value of the discrepancy computed by the algorithms and provides a measure of the improvement in the solution of the optimization procedure.

We consider a statistics over 25 different combinations of fault parameters determined through the scaled Latin hypercube sampling process proposed by Berri et al. [225]. This design of experiments permits to increase the distribution of the fiber cut length located in proximity of the undamaged condition, and improves the amount of incipient damages evaluated during the experiments.

Figure 8.3 reports the outcomes in terms of median and interval between the 25-th and 75-th percentiles for both the assessment metrics, and Table 8.1 summarizes the convergence results. Overall, the multifidelity algorithms – PA-MFBO and MFEI – achieve lower values of the identification error rather than the single fidelity strategy implementing the high-fidelity structural model – EGO. However, it can be noticed that the proposed PA-MFBO is the only optimization method capable to infer the exact health status of the composite plate ($e(x_i) = 0\%$) with a computational budget of just $B = 22.8$, which corresponds to less than half of the budget consumed by EGO and MFBO ($B = 50$) to converge to suboptimal values of the identification error $e(x_i) > 0\%$. These results suggest that the introduction of prior expert knowledge about the health monitoring problem enhances the accuracy of the damage identification procedure. A remarkable outcome is that the PA-MFBO algorithm is the only optimization framework capable to accurately identify the health status of the composite plate within the allocated budget. This outcome suggests that the inclusion of the expert knowledge about the structures of the domain and the behaviour of the numerical models over those structures allow to obtain a fast and robust inference performance.



(a)

(b) (c)

(d) (e)

Fig. 8.3 Statistics over 25 runs of the minimum discrepancy f^* and percentage relative error of the inference of the damage parameters $e(x_i)$ obtained with the competing algorithms.

Method	$e(x_1)$	$e(x_2)$	$e(x_3)$	$e(x_4)$	f^*
PA-MFBO	0.00%	0.00%	0.00%	0.00%	0.00
MFBO	8.46%	11.5%	4.26%	0.19%	0.0261
EGO	34.1%	55.5%	10.7%	2.47%	0.0474

Table 8.1 Convergence results of the percentage relative error of the inference of the damage parameters $e(x_i)$, and minimum discrepancy value f^* obtained with the competing algorithms.

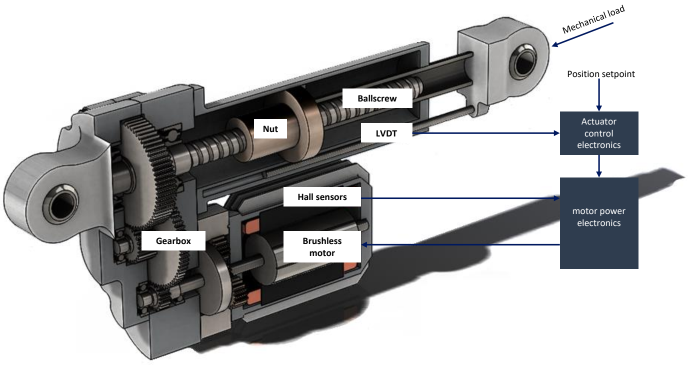


Fig. 8.4 An aerospace electromechanical actuator

8.2 Aerospace Electromechanical Actuator Diagnostics

In this section, we demonstrate and validate our computational frameworks FREEDOM and MF-FREEDOM for the FDI problem of an ElectroMechanical Actuator (EMA) adopted for the flight control system of manned aircraft. EMAs represent a complex multiphysics system that constitute a potential enabling technology for sustainable aviation according to the more electric and all electric aircraft philosophies: these systems permits the elimination of a centralized hydraulic power generation system with benefits in terms of lower emissions and weight reduction [273–278]. The progressive switch to EMA based flight control systems is currently underway in upcoming platforms, and will be facilitated by the adoption of reliable fault detection strategies. EMAs constitute a significant challenge for the FDI procedure due to the presence of different subsystems characterized by heterogeneous and coupled physical domains. The common subsystems of an aerospace EMA (Figure 8.4) includes a brushless motor with its power electronics, and a mechanical transmission with a reduction gearbox and a screw device to convert rotary motion to linear translation of the output. A network of sensors measures positions, speeds, temperatures and electrical parameters to inform the control electronics and close the feedback loops.

FDI for Aerospace EMAs is a particularly challenging task as the monitored system combines multi-physical behaviors involving the interaction between electrical, mechanical and thermal subsystems, which can exhibit highly nonlinear behaviors

fault parameter	fault mode	$x_i = 0$ (no damage)	$x_i = 1$ (full damage)
x_1	friction	nominal	3 times nominal
x_2	backlash	nominal	100 times nominal
x_3	phase A short circuit	absent	complete
x_4	phase B short circuit	absent	complete
x_5	phase C short circuit	absent	complete
x_6	rotor eccentricity	absent	air gap width
x_7	eccentricity phase	$-180°$	$180°$

Table 8.2 Definition of the health status of the EMA system in terms of fault parameters x_i

under different conditions. As a consequence, EMAs may show multiple fault modes that can interact with each other through causal relationships (i.e. an initial fault can propagate to other components) yielding to nonlinear combinations of effects on the actuator's performances. Additional challenges to FDI are posed by the harsh environment in which aerospace actuators are required to operate: EMAs are often exposed to extreme temperature changes, high vibration and acceleration levels, and electromagnetic disturbances, all of which can hamper the acquisition of accurate and reliable sensor data.

For this study, the health condition of the system is encoded in $\boldsymbol{x} \in \mathbb{R}^7$ and includes four different failure modes, namely friction (x_1) and backlash (x_2) increases, partial short circuit of the three stator windings ($x_{3,4,5}$ respectively) and rotor eccentricity ($x_{6,7}$ for the eccentricity amplitude and phase). Table 8.2 summarizes the fault modes affecting the considered EMA system; these failure conditions are selected among the most common EMA faults during the in-flight operations [279, 280]. For each element of $\boldsymbol{x}$, a null value represents a nominal condition without faults, while a unit value is a full failure state. We consider as incipient damages values of fault parameters $x_i < 0.05$ which represent a condition where the system is still capable to meet the operational requirements with a small degradation of the performance.

The variable monitored for the FDI task is the equivalent DC stator current $I = \frac{1}{2}(|i_A|+|i_B|+|i_C|)\text{sign}(T_m)$, where T_m is the motor torque and $i_{A,B,C}$ are the stator phase currents. This variable is sensitive to faults and is already measured to close the torque feedback loop. Thus, the FDI problem of the considered aerospace EMA is formulated as follows:

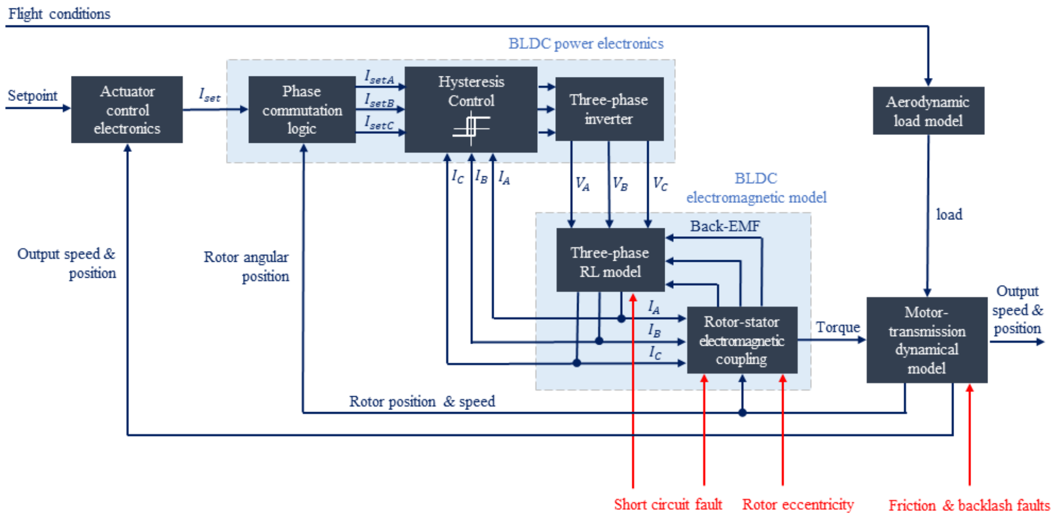


Fig. 8.5 Block diagram of the HF EMA model

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathscr{X}} ||\hat{\boldsymbol{I}_{ref}}(\boldsymbol{x}^*,\hat{\boldsymbol{m}}) - \hat{\boldsymbol{I}_{mon}}(\boldsymbol{x},\hat{\boldsymbol{m}})|| \tag{8.5}$$

where $\boldsymbol{x} = [x_1,...,x_7]$ represents the health status of the considered system, and $\mathscr{X} = [0,1]^7$ is the space of the fault parameters.

We use two physics-based numerical models of the EMA at different levels of fidelity to demonstrate the capabilities of our FDI algorithms. In addition, we employ the data acquired from a real-world EMA test-bench to validate the efficiency of the proposed methodologies. The following sections briefly describe the two numerical models and the physical test-bench of the EMA. Further details about those numerical models and the physical test-bench are discussed in [281, 282].

8.2.1 High Fidelity (HF) model

The High Fidelity (HF) model of the actuator is a detailed, physics-based emulator of the EMA's dynamical behavior. The model has a high accuracy in simulating the response of a physical system, as validated experimentally in [283]. Figure 8.5 shows the block diagram of the high-fidelity models. At the core of the HF model is a three-phase simulation of the stator currents, including the hysteresis closed-loop current control and the resistive-inductive (RL) model of the stator coils.

The electromagnetic coupling between the wound stator and the permanent magnet rotor is mapped as a function of the rotor angular position and magnetic flux that concatenates with the stator coils. The magnetic flux time derivative is leveraged to evaluate both the back-EMF and the torque generated by the motor. The magnetic model is sensitive to asymmetries in the air gap distribution and to unbalances in the windings, therefore faults like rotor eccentricity and partial short circuits of the stator can be simulated directly.

The mechanical transmission is modelled as a second order dyamical system, that includes the simulation of several nonlinear effects, namely: backlash, dry friction, mechanical endstops, and the finite stiffness of load-carrying components. The mechanical load on the actuator is simulated through the linearized longitudinal model of the F-16 jet aircraft proposed by Stevens [284]. Stator currents i_A, i_B and i_C are measured to close the current/torque control loop; the velocity and position loops rely on three Hall effect sensors on the motor shaft and a linear displacement sensor (LVDT) on the actuator output, respectively. The current loop is managed by three individual hysteresis controllers, one for each phase; the position and velocity loops feature Proportional-Integral-Derivative (PID) regulators with full anti-windup, derivative filtering and dead-band functions to inhibit limit cycles.

We utilize this HF model as an emulator of a real-world actuator to evaluate the proposed FDI strategy and gather ground truth data for our two-step compression methodology. However, the high computational cost required to estimate the dynamic of the EMA system is nearly two orders of magnitude higher than the simulated time interval, making the FDI task with the high-fidelity model alone impractical with limited computational resources. As a result, we aim to develop a low-fidelity model of the EMA system that introduces approximations to alleviate the computational burden while maintaining an acceptable level of accuracy in simulating the actuator's dynamics.

8.2.2 Low Fidelity (LF) model

The Low Fidelity model of the actuator introduces simplifications to the physical representation of the EMA in order to reduce the computational cost in evaluation while retaining an acceptable accuracy. The block diagram of the LF model is

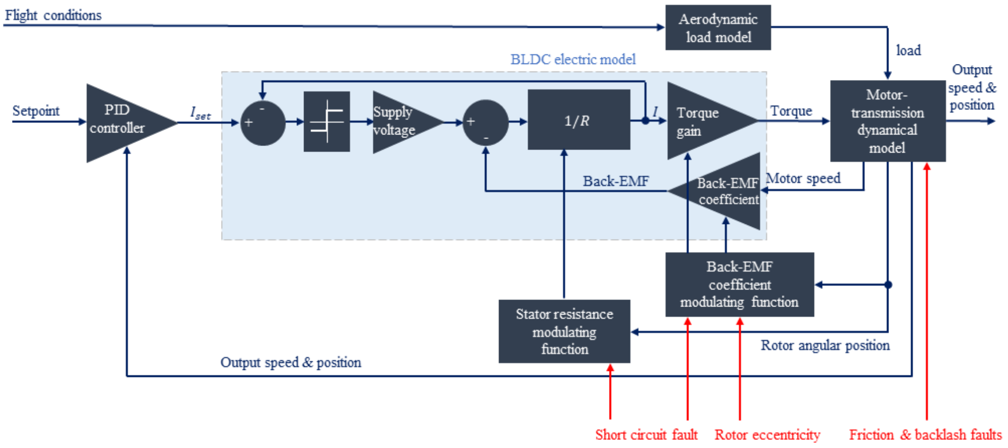


Fig. 8.6 Block diagram of the LF EMA model

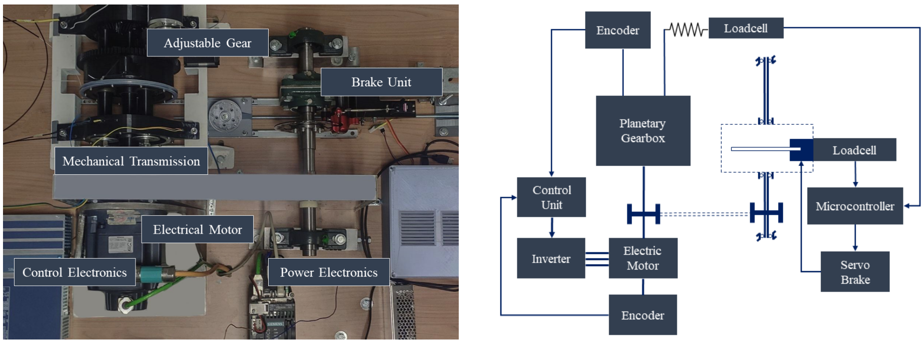
shown in Figure 8.6. The main simplification with respect to the HF model is the replacement of the three-phase RL stator simulation with a single-phase equation:

$$V = RI + \kappa_v \omega \tag{8.6}$$

where κ_v is the motor's back-EMF coefficient and ω is the rotor's angular speed. This approach enables to speed up computations significantly, as the stator circuit is pre-solved and does not require to run an iterative solver at each timestep. This approach requires to introduce the sensitivity to fault modes empirically through modulating functions for the motor parameters, as proposed by Berri et al. [285]. In addition, the control logic only includes a linear PID for the speed and position loops, and a simplified hysteresis controller – consisting in a single sign function – manages the current loop. The aerodynamic load is estimated from the aircraft attitude with a proportional gain, neglecting the longitudinal dynamics of the entire vehicle. In this work, the LF model is used alongside with the HF one as a source of information to compute the monitoring current signal of the EMA system.

8.2.3 Physical test bench

The proposed FDI strategy is validated experimentally with data from a real-world EMA test-bench capable to simulate the presence of mechanical faults in the transmission. The setup, shown in Figure 8.7, includes a permanent magnet brushless



(a) (b)

Fig. 8.7 (a) EMA test-bench and (b) the corresponding block diagram.

motor connected to a planetary reduction gearbox. The motor is driven by a current-controlled, 400 V three-phase inverter. A separate 24 VDC bus drives the control logic. A pair of high resolution encoders measure the position of the motor shaft and the position of the gearbox output shaft: the former is used for phase commutation and speed feedback while the latter closes the position control loop. A repeatable mechanical load is provided by a servo-actuated brake module, whose torque is measured by a loadcell and controlled in closed loop. This permits to simulate different load profiles and include the effect of a friction increase fault. Moreover, the last gear pair driving the output shaft is adjustable, and allows to simulate the backlash fault through a controlled variation of the mechanical play. Acquisitions from the test bench were initially used to validate the HF and LF models in nominal conditions and in presence of mechanical faults [286]. In addition, the measurements of current signals from the EMA test-bench are used as the reference signal to validate the performance of the proposed multifidelity FDI strategy.

8.2.4 Experiments

For all the experiments of this study, we collect a reference dataset of 100 combinations of incipient faults whose corresponding system responses are evaluated with the high-fidelity EMA model. This set of faults is determined through a modified scaled Latin hypercube sampling to increase the density of sampling points near the nominal condition [225]. Offline, we use this dataset to compute an encoding map of $n_w = 30$ informative points using our original two-step compression strategy presented in Section 6.1.1. Figure 8.8 shows the placement of those informative

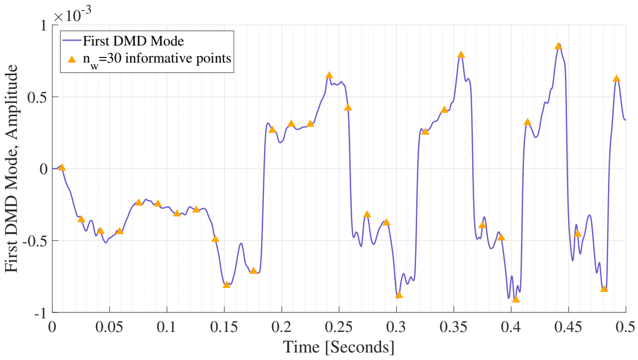


Fig. 8.8 Set of $n_w = 30$ informative points over the first DMD mode

points determined through the self organizing map for the first mode computed using the dynamic mode decomposition technique: our compression strategy places the n_w points in correspondence of the most significant locations of the mode including minima and maxima. This compressed representation is adopted online to identify the health status of the EMA system.

We conducted both numerical and physical experiments considering different experimental configurations to investigate the overall behaviour of the proposed single-fidelity FREEDOM algorithm and multifidelity MF-FREEDOM algorithm. In particular, we first tested FREEDOM implementing the high-fidelity numerical model of the EMA (Section 8.2.5) to compute both the reference output signal and the monitoring signal: the goal is to assess the performance of the proposed methodology in the identification of the health status of the system without introducing a modeling error between the reference and the monitoring signals. This procedure permits to highlight the main difficulties in solving the EMA identification problem caused by the interaction of heterogeneous and multi-domain fault modes that affect simultaneously the system, and excludes the influence of modeling approximations in the computation of the monitoring signal. Then, we investigate our FREEDOM methodology implementing the low-fidelity representation of the EMA (Section 8.2.5) as the monitoring model, to evaluate the actual performance in fault identification considering a real-world scenario where a fast digital twin is required to limit the computational resources for the evaluation of the monitoring signal. This allows to assess both the accuracy and robustness of the proposed methodology in presence of multiple fault conditions and modeling error between reference and monitoring system. Finally, we validate FREEDOM for the health assessment of the real-world

EMA system (Section 8.2.5) using the low-fidelity EMA model as the monitoring model.

In the multifidelity numerical experiments (Section 8.2.6), the proposed MF-FREEDOM is investigated for a setting where the high-fidelity EMA model is adopted to compute the reference current signal of the EMA system, and the multifidelity FDI framework leverages both the high-fidelity and low-fidelity EMA numerical models to compute the monitoring current signal: the multifidelity Bayesian procedure selects at each iteration of the identification process the appropriate EMA numerical model seeking to efficiently direct the computational resources towards the identification of the actual damages. In the real-world multifidelity physical experiments, the EMA test-bench permits to acquire the reference current signal considering the real-world system affected by damages, and both the high-fidelity and low-fidelity EMA numerical models are used during the FDI procedure to compute the monitoring current signal and identify the EMA health status through the multifidelity Bayesian identification procedure.

For the numerical experiments, the output current signal of the EMA is computed considering a linear chirp actuation command characterized by a 0.5 s duration, $5 \cdot 10^{-3}$ rad amplitude, 0 Hz start frequency and 15 Hz end frequency, and an aerodynamic load of 0.5 Nm applied to the actuator. The measurement locations consist of multiple acquisitions of the current signal every 10^{-6} seconds over the actuation time for the high-fidelity numerical model, and every 10^{-4} for the low-fidelity numerical model. This results in a computational cost of the high-fidelity model of 10 seconds on average, while the low-fidelity model requires approximately 1 second. For the real-world EMA test bench, the current signal is acquired considering a sinusoidal actuation command of 24 seconds with 0.12 rad amplitude and 0.2 Hz frequency and without external loads applied to the actuator. This real-world current signal is acquired in measurement locations along the actuation time each $4 \cdot 10^{-3}$ seconds apart. All the algorithms are implemented in the Matlab environment and the experiments are conducted on a laptop PC with Intel Core i7-6700HQ and 32GB memory.

8.2.5 FREEDOM Results and Discussion

In this section, we illustrate and discuss the results achieved with the proposed FREEDOM approach for the EMA diagnostic problem. We compare three different formulations of the acquisition function (Section 6.1.2) for the Bayesian damage inference stage, including the expected improvement (EGO) [109], upper-confidence bound (UCB) [287] and two-step lookahead acquisition function (lookEGO) [137]. To assess the diagnostics performance of FREEDOM, we define the following assessment metrics:

$$e(x_i) = \frac{||x_i^* - \widetilde{x}_i||}{x_i^*} \times 100\% \tag{8.7}$$

$$f^* = \min(f(\boldsymbol{x}^*, \hat{\boldsymbol{m}})) \tag{8.8}$$

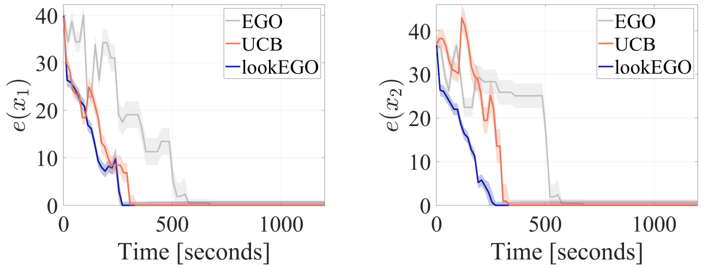
where x_i^* is the actual level of damage that affects the system, $\widetilde{x}_i$ is the level of damage inferred by the algorithm considering the i-th fault mode, and $f(\boldsymbol{x}^*, \hat{\boldsymbol{m}})$ is the value of the discrepancy between the reference and monitoring output signals. In particular, $e(x_i)$ measures the percentage relative error in the identification of the health status of the system, and f^* represents the minimum value of the discrepancy achieved by the algorithm.

We evaluate $e(x_i)$ and f^* as functions of the computational time required by the different implementations to complete the FDI task. For the results obtained against numerical experiments, we consider a statistics over 100 different incipient faults combinations $\boldsymbol{x}^* = (x_1^*, ..., x_7^*)$ including all the spectrum of mechanical and electrical damages. For the results obtained against physical experiments, we consider a statistics over 10 different combinations of incipient mechanical faults $\boldsymbol{x}^* = (x_1^*, x_2^*)$. For both the numerical and physical experiments, the incipient damage conditions are determined through the scaled Latin hypercube sampling proposed by Berri et al. [225]. This choice is motivated by the increased probability distribution of the fault combinations located near the nominal condition, in order to collect a large amount of incipient faults without limiting the identification process bounding the space of faults. In the remaining, the results are compared in term of median values of the inference error $e(x_i)$ and discrepancy f^* together with the associated statistics in between the 25-th and 75-th percentiles.

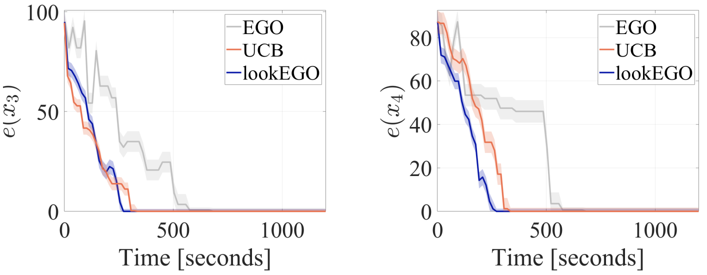
Numerical Experiments: High-Fidelity Monitoring

Figure 8.9 illustrates the outcomes obtained for an experimental setup where the high-fidelity numerical model of the EMA (Section 8.2.1) is used as the source of information for both the reference and monitoring current signal. All the competing formulations are capable to exactly identify the health status of the EMA system reducing the inference error $e(x_i)$ for both the mechanical (x_1, x_2) and electrical faults $(x_3, x_4, x_5, x_6, x_7)$. We notice that the two-step lookahead acquisition function (lookEGO) provides superior convergence performance, and provides a significant reduction of the inference error for all the damages with reduced computational time with respect to UCB and EGO. This suggest the opportunity offered by non-myopic strategies to quantify the improvement in the solution of the FDI problem achievable in future iterations and determine an effective and efficient inference stage.

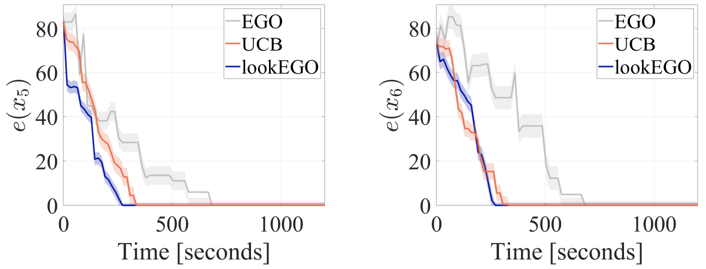
It is possible to observe that the correct inference of the EMA damage configuration, which correspond to the computational time for which the inference error is equal to zero, coincides also with zero discrepancy between the reference and monitoring current signals (Figure 8.9(h)). Indeed, the use of the high-fidelity numerical model to evaluate both the reference and monitoring current signal involves the computation of the same reference and monitoring responses when the damages are correctly identified, which in turn determine a value of the discrepancy function equal to zero. This permits to highlight the challenges associated with the EMA health assessment problem without introducing a modeling error between reference and monitoring signals (which is typically the case of real-world problems). In particular, the convergence diagrams for all the fault modes affecting the EMA indicate that a reduction of the discrepancy between the reference and monitoring current signal is not related with a continuous decrease of the inference error (Figure 8.9). This suggests that the EMA identification problem is ill-posed: a reduction of the discrepancy function is not always related to an improvement in the accuracy of the faults inference. As a consequence, the identified fault parameters are subjected to instability and uncertainty that raise when the reduction of the discrepancy causes an increase in the inference error, further complicating the fault detection and isolation task. These results are justified with the presence of multiple and multiphysics faults affecting simultaneously the EMA system. In particular, a variation of both the friction in the mechanical transmission (x_1) and the partial short circuit in the three phases of the electric motor $(x_3,\ x_4,\ x_5)$ determines opposite effects in the



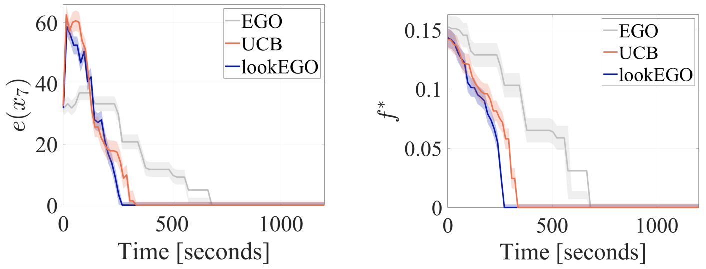
(a) Friction (b) Backlash



(c) Phase A short circuit (d) Phase B short circuit



(e) Phase C short circuit (f) Rotor eccentricity



(g) Eccentricity phase (h) Discrepancy

Fig. 8.9 Statistics over 100 incipient fault conditions of the percentage relative error of the inference of the fault parameters $e(x_i)$ and minimum discrepancy value f^* obtained for the high-fidelity monitoring.

Method	$e(x_1)$	$e(x_2)$	$e(x_3)$	$e(x_4)$	$e(x_5)$	$e(x_6)$	$e(x_7)$	f^*	Time
EGO	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0	583 s
UCB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0	335 s
lookEGO	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0	271 s

Table 8.3 Convergence results of the percentage relative error of the inference of the fault parameters $e(x_i)$, minimum discrepancy value f^*, and computational time obtained for the high-fidelity monitoring.

Method	$e_m(x_1)$	$e_m(x_2)$	$e_m(x_3)$	$e_m(x_4)$	$e_m(x_5)$	$e_m(x_6)$	$e_m(x_7)$
EGO	1.06%	1.05%	1.53%	1.51%	1.07%	1.43%	1.16%
UCB	0.87%	1.30%	1.10%	1.64%	1.35%	1.26%	1.13%
lookEGO	0.88%	0.91%	1.21%	1.25%	0.95%	1.20%	1.05%

Table 8.4 Convergence results of the maximum percentage relative error of the inference of the fault parameters $e_m(x_i)$ obtained for the high-fidelity monitoring.

dynamical response of the system: increasing x_1 reduces the speed of the actuation while increasing $x_{3:5}$ causes an increase in the motor speed. These effects indicate a distinct multimodality of the discrepancy function, which is characterized by the presence of multiple suboptimal local minima that stresses the search towards the actual health status of the system.

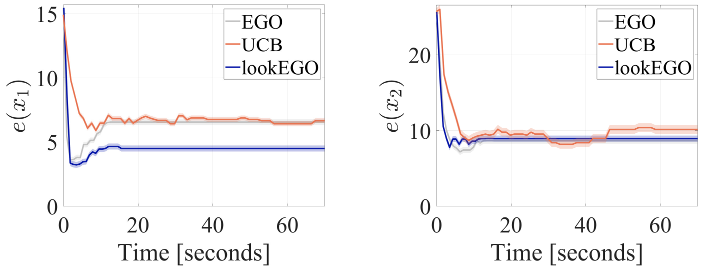
Table 8.3 and Table 8.4 show and summarize the convergence history for the three formulations of the acquisition function in terms of median values of the inference error $e(x_i)$ and minimum discrepancy f^*, and maximum inference error $e_m(x_i)$ achieved at convergence, respectively. It is possible to observe that the demand for computational time required to identify the exact EMA health status is on average 271 seconds for the lookEGO, while the other competing formulations require further time expenditure to infer the damaged condition (Table 8.3). A significant outcome observed at convergence is that all the algorithms achieve a maximum inference error lower than 1.64% over the experimental setup of 100 incipient fault combinations, highlighting the robust identification of the health status of the system (Table 8.4).

The outcome from the high-fidelity monitoring experiments demonstrate both the accuracy and robustness of the FREEDOM algorithm even in presence of multiple and interacting incipient failure modes. In addition, it is shown the capability of our framework to overcome the issues related to the ill-posed identification problem and multimodality of the discrepancy function. However, the demand for computational resources makes the FDI procedure with high-fidelity numerical models unfeasible for the fast identification of the health status of the system. This justifies the adoption of the low-fidelity numerical model to approximate the output signal of the system, decreasing the computational cost required for the FDI procedure.

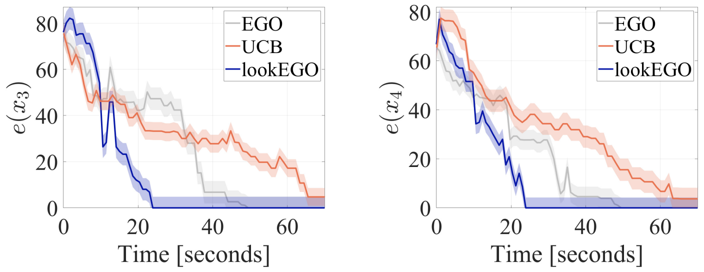
Numerical Experiments: Low-Fidelity Monitoring

Figure 8.10 shows the results achieved for the experiments implementing the low-fidelity numerical model of the EMA system (Section 8.2.2) to compute the monitoring current signal. Similarly to the previous experiments, we observe that the lookEGO acquisition function provides a superior reduction of the inference error with a contained computational time if compared with EGO and UCB. Nevertheless, all the competing formulations identify accurately the level of damage for the electrical faults (x_3, x_4, x_5, x_6, x_7) and allow to reduce the inference error for friction (x_1) and backlash (x_2) fault modes below the 6.66% and 10.1%, respectively. However, the convergence values of $e(x_1)$ and $e(x_2)$ are higher with respect to the values of the error for the electrical faults. Figure 8.10(a) and Figure 8.10(b) report the convergence history of the inference error for the mechanical faults affecting the EMA. Both $e(x_1)$ and $e(x_2)$ show similar trends: the error decreases for lower computational times while tends to reach values higher than the minimum computed so far as the FDI algorithm proceeds in reducing the discrepancy f. This result can be justified with the introduction of a modeling error between the reference – high-fidelity – and monitoring – low-fidelity – signals that is reflected in a computed minimum of the discrepancy f^* different from zero (Figure 8.10(h)). Indeed, the same combination of fault parameters determines discrepant current responses between the high-fidelity and low-fidelity numerical model caused by the physical approximations adopted to reduce the computational cost of the low-fidelity representation. Those effects poses further difficulties to efficiently assess the health status of the system.

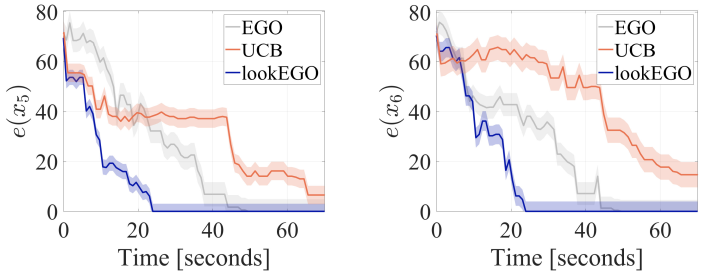
Table 8.5 illustrates the median values of the inference error $e(x_i)$ and minimium discrepancy f^* reached at the convergence of the FDI process, while Table 8.6 shows the corresponding maximum inference error $e_m(x_i)$. In particular, the lookEGO identifies the electrical damages with an error below the 0.12%, while permits the identification of the mechanical faults with an error of the 4.51% and 8.94% for friction and backlash, respectively (Table 8.5). Although the inference of mechanical



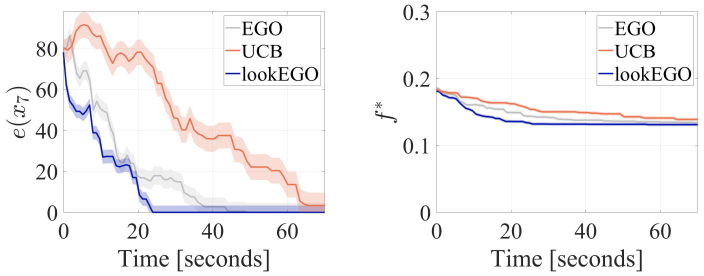
(a) Friction (b) Backlash



(c) Phase A short circuit (d) Phase B short circuit



(e) Phase C short circuit (f) Rotor eccentricity



(g) Eccentricity phase (h) Discrepancy

Fig. 8.10 Statistics over 100 incipient fault conditions of the percentage relative error of the inference of the fault parameters $e(x_i)$ and minimum discrepancy value f^* obtained for the low-fidelity monitoring.

Method	$e(x_1)$	$e(x_2)$	$e(x_3)$	$e(x_4)$	$e(x_5)$	$e(x_6)$	$e(x_7)$	f^*	Time
EGO	6.5%	8.6%	.08%	.09%	.11%	.10%	.05%	0.136	51 s
UCB	6.6%	10%	4.7%	3.6%	6.5%	14%	3.3%	0.139	69 s
lookEGO	4.5%	8.9%	.05%	.09%	.12%	.07%	.06%	0.132	25 s

Table 8.5 Convergence results of the percentage relative error of the inference of the fault parameters $e(x_i)$, minimum discrepancy value f^*, and computational time obtained for the low-fidelity monitoring.

Method	$e_m(x_1)$	$e_m(x_2)$	$e_m(x_3)$	$e_m(x_4)$	$e_m(x_5)$	$e_m(x_6)$	$e_m(x_7)$
EGO	6.77%	9.07%	4.90%	4.05%	4.82%	4.61%	4.53%
UCB	6.86%	10.73%	8.60%	8.21%	10.23%	19.77%	9.64%
lookEGO	4.73%	9.33%	4.81%	4.24%	3.12%	3.98%	3.35%

Table 8.6 Convergence results of the maximum percentage relative error of the inference of the fault parameters $e_m(x_i)$ obtained for the low-fidelity monitoring.

Metric	GA	PSO	DE	GWO	FREEDOM
Average Time	2322 s	1710 s	405 s	709 s	25 s
Average Error $e(\boldsymbol{x})$	2.813%	0.711%	3.001%	4.378%	1.977%

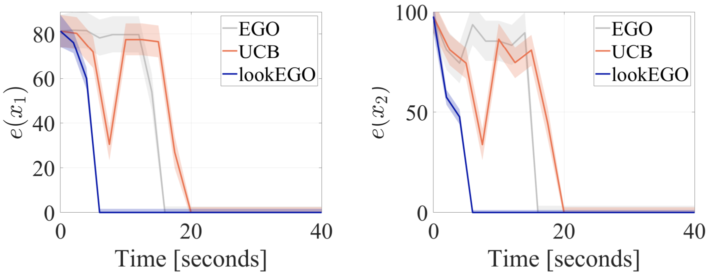
Table 8.7 Comparison of the computational time and average inference error between the FREEDOM algorithm and the meta-heuristic algorithms proposed by [2], namely Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Differential Evolution (DE) and Grey Wolf Optimization (GWO).

failures is not as accurate as the identification of electrical damages, the overall accuracy of our methodology is adequate for the detection of incipient failure modes, and permits the adoption of corrective actions to contain the early stage effects on the degradation of the system performance. In addition, the identification of the health status of the system requires on average 25 seconds for the lookEGO, which correspond to a reduction of the 91% if compared with the time required to assess the EMA health status with the high-fidelity monitoring. This allows to infer barely noticeable failure modes early on before the propagation, and permits to prevent and counteract to excessive loss of performance of the system with potentially severe effects.

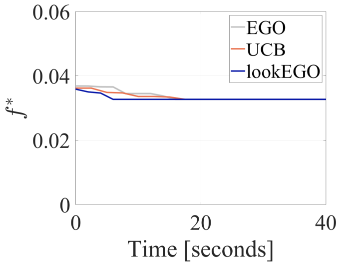
Moreover, it can be noticed that the lookEGO algorithm achieves a maximum inference error at convergence below the 4.81% for the electrical damages, and equal to 4.73% for friction and 9.33% for backlash considering a total of 100 incipient fault combinations (Table 8.6). These outcomes demonstrate that our methodology FREEDOM provides a robust identification of the health status of the system even in presence of modeling errors, ill-posedness, and multimodality of the identification problem. However, the maximum inference error computed at convergence is higher if compared with the outcomes of the high-fidelity monitoring test-case (Table 8.4), suggesting that the introduction of the modeling error also affects the overall robustness of the FDI process.

Table 8.7 compares the results obtained with the FREEDOM algorithm and with popular meta-heuristic approaches as the one proposed by [2]. In particular, we compare the two-step lookahead implementation of FREEDOM with the diagnostics outcomes obtained with Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Differential Evolution (DE) and Grey Wolf Optimization (GWO). FREEDOM is capable to achieve a comparable average reduction of the inference error with a remarkable acceleration up to two orders of magnitude if compared with the standard algorithms. These results remark and emphasize the capability of the FREEDOM framework to leverage the combination of efficiently compressed highly informative signals with the Bayesian inference stage to provide an accurate health status assessment with a fraction of the computational cost required by competing algorithms.

The outcomes computed adopting the low-fidelity numerical model demonstrate a substantial acceleration of the inference of incipient damages with satisfactory accuracy and robustness. In particular, the implementation of a low-fidelity monitoring model significantly reduces the computational burden associated with the FDI procedure if compared with the time expenditure required for the high-fidelity monitoring test-case. Moreover, the computational time required to complete the FDI task with our methodology is two orders of magnitude lower than state of the art model-based FDI algorithms [2, 288–290]. This reveals the capabilities of our algorithm to efficiently exploit data from the low-fidelity monitoring model and contain the demand for computing resources. However, the introduction of a modeling error between the reference and monitoring signal moderately affects the accuracy and robustness of the FDI procedure for the failures in the mechanical transmission.



(a) Friction (b) Backlash



(c) Discrepancy

Fig. 8.11 Statistics over 10 incipient fault conditions of the percentage relative error of the inference of the fault parameters $e(x_i)$ and minimum discrepancy value f^* obtained for the real-world experiments.

Real-World Experiments

Figure 8.11 illustrates the results obtained with the FREEDOM algorithm for the health assessment of the real-world EMA system (Section 8.2.3) implementing the low-fidelity monitoring model. All the three formulations are capable to identify the health status of the system minimizing the inference error with a contained computational time. The overall convergence trends for both the inference error of friction (Figure 8.13(a)) and backlash (Figure 8.13(b)) confirm that the identification problem is ill-posed and multimodal – the decrease of the discrepancy (Figure 8.13(h)) is not always related to a reduction of the inference error – and affected by the modeling error between the current signal measured on the real-world system and the same signal computed with the low-fidelity model – the minimum of the discrepancy is greater than zero when the health status of the system is inferred.

Table 8.8 provides the experimental convergence median values of the inference error $e(x_i)$ and minimum discrepancy f^* for the three acquisition functions considered in this paper. All the implementations accurately identify the health status of

Method	$e(x_1)$	$e(x_2)$	f^*	Time
EGO	0.03%	0.08%	0.0326	16 s
UCB	0.03%	0.08%	0.0326	20 s
lookEGO	0.03%	0.08%	0.0326	6 s

Table 8.8 Convergence results of the percentage relative error of the inference of the fault parameters $e(x_i)$, minimum discrepancy value f^*, and computational time obtained for the real-world experiments.

Method	$e_m(x_1)$	$e_m(x_2)$
EGO	2.69%	3.35%
UCB	2.39%	2.59%
lookEGO	1.62%	1.24%

Table 8.9 Convergence results of the maximum percentage relative error of the inference of the fault parameters $e_m(x_i)$ obtained for the real-world experiments.

the system, and lead to an inference error of the 0.03% for friction and 0.08% for backlash. However, the lookEGO implementation achieves the best experimental results in terms of efficiency, identifying the incipient mechanical faults with a computational time within 6 seconds; this corresponds to a reduction of the 62% and 70% of time expenditure if compared with the outcomes of the EGO and UCB methods, respectively. Table 8.9 illustrates the maximum inference errors $e_m(x_i)$ at convergence. All the formulations keep the maximum error below the 3.35%, validating the robust identification of the incipient mechanical faults; in particular, the two-step lookahead acquisition function shows the higher level of robustness with an inference error at convergence lower than the 1.62% and 1.24% for friction and backlash, respectively.

These outcomes validate the performance of FREEDOM as a computational framework that enables the acceleration of the FDI procedure, and permits the accurate and robust identification of incipient faults affecting a complex and multiphysics aerospace system. A significant achievement of our methodology is the reduction of computational time required to identify the health status of the real-system, which corresponds to an inference procedure two orders of magnitude faster than standard model-based FDI approaches [2, 288–290].

8.2.6 MF-FREEDOM Results and Discussion

In this section, we illustrate and discuss the results achieved with the proposed multifidelity FREEDOM (MF-FREEDOM) approach for the EMA diagnostic problem. We compare our framework with the state-of-the-art Efficient Global Optimization (EGO) algorithm, a popular Bayesian algorithm where the expected improvement works as the acquisition function and only the high-fidelity EMA model is available during the FDI procedure [109].

The FDI performance are evaluated using two assessment metrics: the percentage relative inference error $e(x_i)$ of the x_i fault, and the minimum of the discrepancy function f^* between the reference and monitoring signal:

$$e(x_i) = \frac{||x_i^* - \widetilde{x}_i||}{x_i^*} \times 100\% \tag{8.9}$$

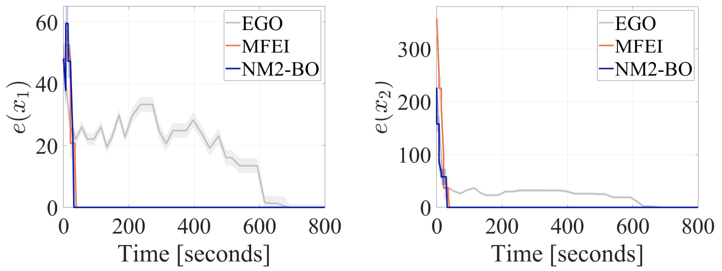
$$f^* = \min(f(\boldsymbol{x}^*, \hat{\boldsymbol{m}})) \tag{8.10}$$

where x_i^* is the actual damage affecting the EMA system and $\widetilde{x}_i$ is the damage inferred by the FDI algorithm. The relative inference error $e(x_i)$ measures the accuracy related to the identification of the damages affecting the EMA system, and f^* provides a measure of the improvement in the solution of the optimization procedure through our multifidelity Bayesian FDI method. For both numerical and physical experiments, these metrics are measured over a statistics of, respectively, 50 and 10 different incipient damage conditions sampled through a scaled Latin hypercube scheme [225]. This sampling procedure increases the probability distribution of damages near the nominal condition, and expands the amount of small incipient faults without completely excluding more serious damages. The goal is to investigate the capabilities of the FDI algorithm to identify the health status of the EMA system before the damages become severe – and easily assessed.

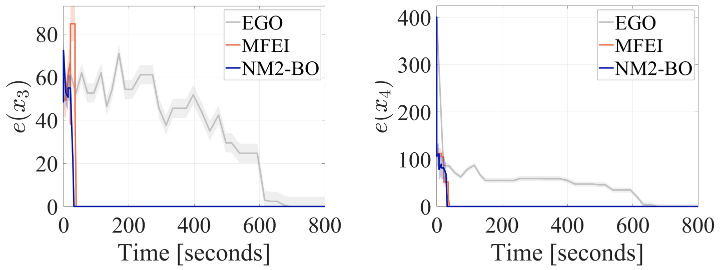
Figure 8.12 and Figure 8.13 illustrate the results of the competing algorithms for the numerical and physical experiments, respectively. The outcomes are reported in terms of median values of the assessment metrics (solid lines) together with the statistics of observations falling in between the 25-th and 75-th percentiles (shaded areas). We show and advance the following empirical observations. First, our MF-FREEDOM algorithm consistently converges much faster than the baseline EGO algorithm with all the formulations of the multifidelity acquisition function – both MFEI and NM2-BO. In particular, the multifidelity paradigm achieves remarkable accelerations and reduces the identification time by more than one order of magnitude. Moreover, the rate of convergence of NM2-BO is moderately higher that the MFEI acquisition function. However, in all the experimental campaign the proposed MF-FREEDOM outperforms the single-fidelity EGO algorithm. In addition, the proposed algorithm provides the exact inference of the incipient faults affecting the EMA system ($e(x_i) = 0$). These results demonstrate and validate the performance of our framework, which provides a fast and robust inference of incipient multiphysics and multimodal faults affecting the EMA. In numerical experiments, the average runtime of NM2-BO and MFEI is within 32 seconds and 39 seconds, respectively, which is notably lower if compared with the 691 seconds runtime of EGO. Similarly, the diagnosis of the real-world EMA takes 26 seconds for NM2-BO and 42 seconds for MFEI on average, while EGO increases the duration of the FDI process to 273 seconds.

The convergence of the minimum discrepancy in both the settings is not strictly related to the continuous decrease of the inference error over the runtime. This can be justified with the ill-posedness of the EMA inverse problem: the FDI procedure is subject to instability caused by the joint decrease of the discrepancy function and increase of the inference error. As highlighted by [221], this behaviour results from the opposite effects of heterogeneous damages on the EMA dynamics: the increase of the mechanical friction x_1 decreases the actuation speed while the increase of the partial short circuit $x_{3:5}$ increases the speed of the motor shaft. As a consequence, the discrepancy function is characterized by a strong multimodality that stresses the FDI framework.

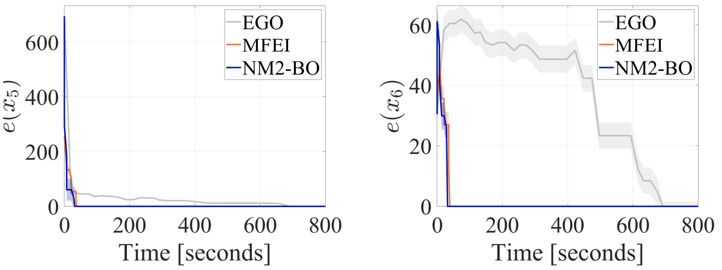
To better illustrate and demonstrate the performance of the proposed multifidelity FDI methodology, Figure 8.14 and Figure 8.15 present the search processes as the sampling in the faults domain selected for a single experiment made by our MF-FREEDOM algorithm and the competing EGO method in the numerical and physical experiments, respectively. As it can be observed, our multifidelity FDI method evaluates different damage configurations adopting both the high-fidelity (circle) and low-fidelity (triangle) EMA numerical model, while the EGO algorithm uses only high-fidelity evaluations of the faults space. All the algorithms seek to balance the exploration of different damage configurations to improve the awareness about the distribution of the discrepancy function over the faults domain, and the exploitation



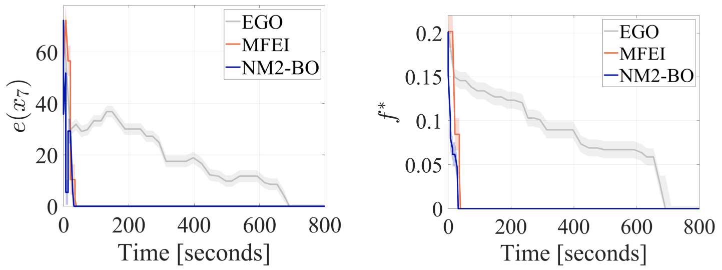
(a) Friction (b) Backlash



(c) Phase A short circuit (d) Phase B short circuit

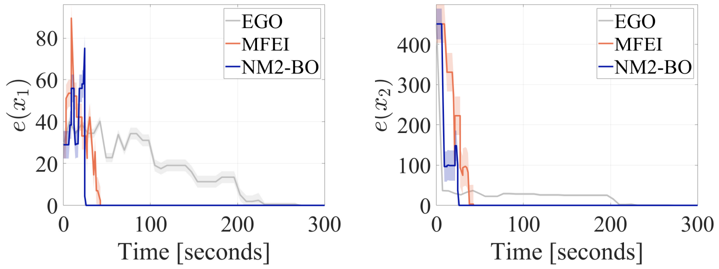


(e) Phase C short circuit (f) Rotor eccentricity

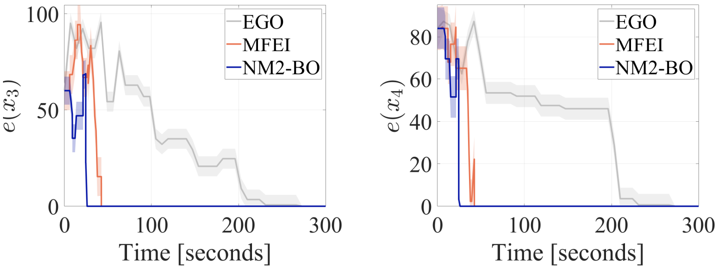


(g) Eccentricity phase (h) Discrepancy

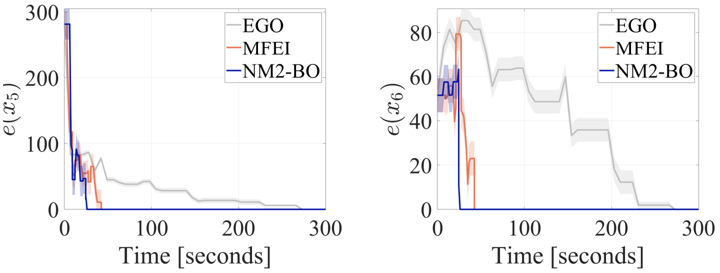
Fig. 8.12 Statistics over 50 incipient fault conditions of the percentage relative error of the inference of the fault parameters $e(x_i)$ and minimum discrepancy value f^* obtained for the MF-FREEDOM algorithm and EGO on numerical experiments.



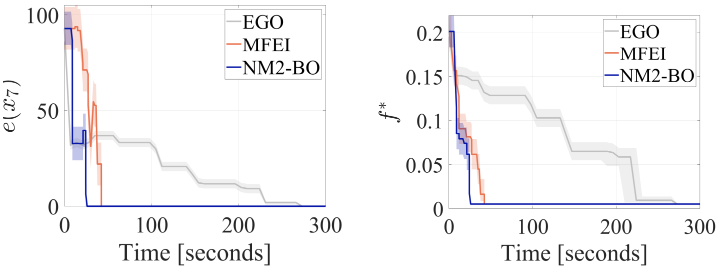
(a) Friction (b) Backlash



(c) Phase A short circuit (d) Phase B short circuit



(e) Phase C short circuit (f) Rotor eccentricity

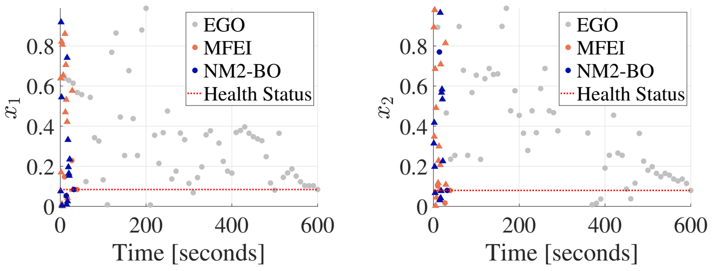


(g) Eccentricity phase (h) Discrepancy

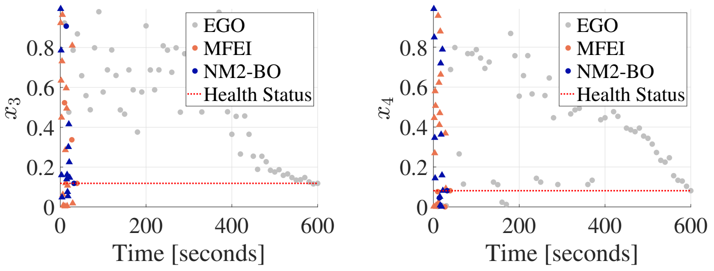
Fig. 8.13 Statistics over 10 incipient fault conditions of the percentage relative error of the inference of the fault parameters $e(x_i)$ and minimum discrepancy value f^* obtained for the MF-FREEDOM algorithm and EGO on physical experiments.

towards the actual damages affecting the system (dashed red line). However, the EGO method requires a large amount of computational time to assess the EMA health status if compared with MF-FREEDOM: high-fidelity data are employed at the beginning of the FDI process to explore damage alternatives, and afterwards directed toward the actual damage level; the unique adoption of the high-fidelity EMA numerical model dramatically penalizes the convergence time. Conversely, the proposed MF-FREEDOM algorithm takes advantage on fast low-fidelity responses: the low-fidelity EMA model is adopted at the beginning of the FDI procedure to explore the faults domain with contained time expenditure while the expensive high-fidelity model is sparingly adopted in the exploitation phase to accurately identify the damages affecting the EMA. In particular, the NM2-BO acquisition function permits to better capitalize the informative content acquired during the low-fidelity exploration phase, and converges faster than the MFEI acquisition function.

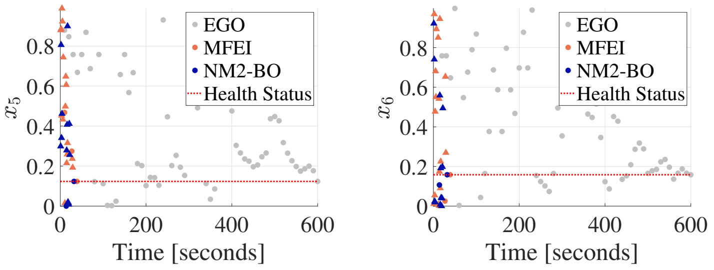
The remarkable accelerations achieved in the experimental campaigns demonstrate and validate the efficiency and robustness of our multifidelity FDI framework even in presence of incipient faults and marked multimodality. These results are due to the combination of two distinguishing features we introduced in our method: i) the optimal two-stage compression reduces the signals into highly informative representations that reduces the dimensionality of the FDI problem, and ii) the multifidelity Bayesian scheme for inference wisely queries high-fidelity models to enhance the identification procedure while contains the overall computational expenditure through low-fidelity evaluations.



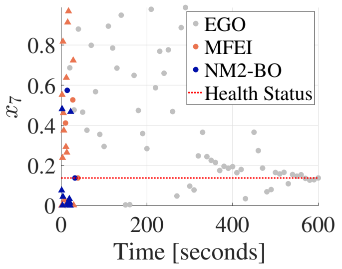
(a) Friction (b) Backlash



(c) Phase A short circuit (d) Phase B short circuit

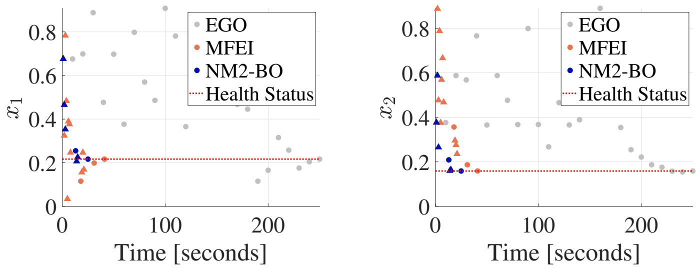


(e) Phase C short circuit (f) Rotor eccentricity

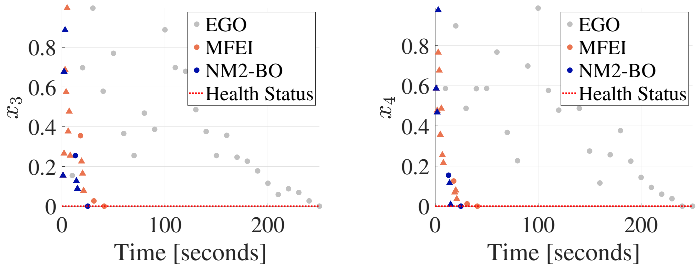


(g) Eccentricity phase

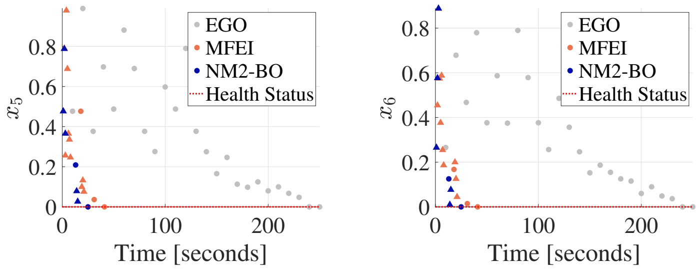
Fig. 8.14 Searches over the faults space of the MF-FREEDOM and EGO algorithms on numerical experiments.



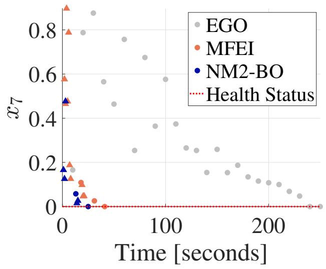
(a) Friction (b) Backlash



(c) Phase A short circuit (d) Phase B short circuit



(e) Phase C short circuit (f) Rotor eccentricity



(g) Eccentricity phase

Fig. 8.15 Searches over the faults space of the MF-FREEDOM and EGO algorithms on physical experiments.

Chapter 9

Conclusions

In chapter 4, we presented a Non-Myopic Multifidelity Bayesian Optimization framework (NM2-BO) to accelerate the optimization of expensive objective functions using multiple representations at different levels of fidelity and evaluating costs. This is achieved by measuring the informative gains acquired from future steps of the optimization procedure through a sampling scheme based on an original two-step lookahead multifidelity acquisition function. We demonstrate the performances of our NM2-BO algorithm against standard multifidelity Bayesian frameworks on a large set of benchmark optimization problems (Section 4.7). These analytical experiments have revealed that the proposed non-myopic multifidelity scheme is capable to outperform the standard algorithms leading to a superior solution of the optimization procedure with a fraction of the computational cost over a variety of mathematical properties that are frequently encountered in real-world applications in science and engineering – such as non linearity, multimodality, oscillatory nature, non-convexity and presence of noise. In addition, in Chapter 7 we demonstrated our approach for the physics-based aerodynamic design optimization problem of a RAE-2822 airfoil subject to aerodynamic and geometric constraints (Section 7.1). The NM2-BO permits to improve the aerodynamic design of the 15.7% requiring a 65% lower computational cost if compared with the best performing standard algorithm, which achieved a design improvement of the 10.9%. Across all the demonstration campaign, the non-myopic multifidelity scheme allows to effectively capitalize from the use of limited high-fidelity information to identify improved optimal solutions, while targeting a trade-off between computational cost and future information gains as prescribed by our two-step lookahead multifidelity acquisition function.

In addition, we propose a Non-Myopic Multipoint Multifidelity Bayesian Optimization (NM3-BO) framework to significantly accelerate expensive multidisciplinary design optimization problems (Section 4.6). NM3-BO combines two distinguishing features: a non-myopic decision making process maximizes the cumulative reward of design solutions over future iterations and a penalization strategy enables multiple decisions as a batch of design configurations and associated level of fidelity to evaluate simultaneously. This search scheme identifies promising batches through the measure of their future utility, and leverages parallel computations to reduce the overall computational cost of the MDO procedure. The NM3-BO algorithm is demonstrated for the MDO problem of a space re-entry vehicle (Section 7.4.2). The method permits substantial accelerations and identifies superior design solutions compared to state-of-the-art multifidelity and single-fidelity algorithms. In particular, NM3-BO delivers on average a space vehicle design improvement of the 17.4% with a fraction of the computational resources adopted by competing algorithms to identify suboptimal solutions.

In chapter 5, this thesis recognizes that domain knowledge is commonly available in science and engineering, and can be used to accelerate and improve the multifidelity optimization process. We propose a Physics-Aware Multifidelity Bayesian Optimization – PA-MFBO – framework that incorporates forms of prior scientific and expert knowledge about the physical domain during the search procedure. This is achieved introducing a learning bias formalized as a physics-aware multifidelity acquisition function that leverages the knowledge about the structure of the domain to enhance the accuracy of the solution and alleviate the computational cost for optimization. The results achieved with the PA-MFBO are observed and discussed for a cross-regime aerodynamic design problem (Section 7.2) and a structural health monitoring problem (Section 8.1). In the design test-case, the PA-MFBO introduces a bias to pursue the awareness about the transition of fluid regimes through the Mach number. In the health assessment task, PA-MFBO incorporates a bias based on the expert knowledge about the features of the domain for specific combinations of load and extension of the damage. We note that for the two optimization problems our methodology outperforms standard single-fidelity and multifidelity Bayesian algorithms in terms of accuracy and acceleration of the search. In particular, the PA-MFBO identifies aerodynamic design solutions capable to deliver a performance improvement of the 24.31% in less than half the computational time required by competing algorithms to search suboptimal designs. Moreover, PA-MFBO is the only algorithm that permits the robust identification of damages in the composite plate, which otherwise would have required more computational resources or might have led to an inaccurate health assessment. In Section 7.4.4, we further investigate the capabilities of the PA-MFBO algorithm for the multidisciplinary design optimization problem of a space vehicle. In particular, we embed during the search for optimal designs the physics-awareness about the stressful aerothermodynamic phenomena that occur during the re-entry trajectory. The sensitivity to the evolution of the domain is introduced through a specific range of altitudes where heat loads threaten the survival of the spacecraft, and high-fidelity responses are essential for the accurate estimate of thermal loads. The results for this MDO problem demonstrate that the physics-awareness capability introduced with our PA-MFBO scheme allows to effectively identify superior design improvements of the 21% with respect to the design upgrade of the 8% obtained with the single fidelity BO counterpart informed by the low-fidelity aerothermodynamic model.

In addition, we incorporate the physics-aware learning scheme with the non-myopic multifidelity policy and formalize a Physics-Aware Non-Myopic MFBO (PA-NM2BO) algorithm in Section 5.2. This learning scheme permits to embed the awareness about the scientific/expert knowledge about the optimization problem of interest while selecting the most valuable set of optimization variables and levels of fidelity through a two-step lookahead sampling policy that measures the long term improvement of the solution of the optimization problem. In Section 7.3, we demonstrate our framework and validate our results against wind tunnel data for the cross-regime constrained aerodynamic optimization of an airfoil subject to maintain a minimum lift coefficient at a fixed flight altitude. Our non-myopic physics-aware multifidelity framework is compared against standard multifidelity and single-fidelity Bayesian frameworks. The results show that our strategy outperforms the competing algorithms allowing for a significant reduction of the drag coefficient with reduced computational expense through the construction and use of an efficient aerodynamic surrogate model. Specifically, the proposed framework allows to lower the error on the identification of the minimum drag coefficient down to the 3.88% on average respect to the experimental optimum, which goes down to 0.422% for the best test case over the experiments conducted. This is achieved thanks to the two key features of the proposed multifidelity acquisition function which allow to consider the evolution of the fluid regime through the Mach number by including physics awareness in the formulation, and to account for the anticipated gain through the two-step lookahead scheme that permits the efficient exploration of the design space.

In Chapter 6.1, we introduce an original fault detection and isolation (FDI) methodology FREEDOM to accelerate the identification of the health status of complex aerospace systems. Our algorithm permits to identify the incipient failure condition affecting the system through the original combination of a two-step compression strategy – which reduces the amount of data to be stored and processed – and an inverse Bayesian approach – to infer failure modes through the efficient use of data. The objective is to accelerate the FDI task and provide an accurate and robust inference of the system health status with a major reduction of the demand for computational resources. We demonstrate and validate our computational framework for the fault detection and isolation problem of an electromechanical actuator, considering heterogeneous mechanical and electrical incipient fault modes. Specifically, we conducted both numerical and physical experiments comparing three formulations of the Bayesian acquisition function to investigate the performance of our methodology with different search criteria, and compared the results achieved with popular model-based FDI algorithms (Section 8.2.5). The results demonstrate that FREEDOM provides an accurate and robust inference of the health status of the system in presence of incipient faults with a reduced time expenditure. This is particularly relevant to capture the early stages of damages propagation, and prevents severe effects due to the degradation of the performance of the system. The comparison over the different formulations reveals that the two-step lookahead acquisition function provides a robust identification of the electrical faults with an error below the 0.1% while keeping the error under the 9% for the mechanical damages; this is due to the intrinsic properties of the low-fidelity monitoring model that introduces a modeling error which increases the difficulties in the identification of mechanical faults. Nevertheless, the major outcome is observed in terms of a large reduction of the computational time required to complete the FDI task. Indeed, the fault detection and isolation procedure is completed within 25 seconds when the low-fidelity monitoring model is used implementing the two-step lookahead acquisition function. This corresponds to a reduction of two orders of magnitude with respect to the computational cost required by standard model-based FDI algorithms. The outcomes are validated through physical experiments on the real-world system affected by mechanical failures. In particular, the two-step lookahead acquisition function permits to achieve the exact identification of the health status of the system within 6 seconds with the use of the low-fidelity monitoring model. The major acceleration of the FDI procedure is enabled by the original combination of a highly informative two-step compression mapping that allows to speed up the data processing, and the data efficient Bayesian inference that permits to contain the number of model evaluations. In addition, our experiments illustrate that the adoption of a low-fidelity monitoring model – and therefore the introduction of a modeling error between the reference and monitoring signals – affects the fault detection and isolation process, precluding the possibility of high accuracy in health status inference.

We overcome these issues formalizing the multifidelity FREEDOM (MF-FREEDOM) computational framework (Section 6.2). The proposed MF-FREEDOM framework permits to accelerate the identification of multiphysics and multimodal incipient faults of complex systems while including high-fidelity evaluations of the monitoring signals. The algorithm relies on the original combination of the two-stage highly informative compression strategy to reduce the dimensionality of the diagnostic signals, and a multifidelity Bayesian scheme to infer incipient failure modes through the combination of multiple models with variable fidelities and costs. We demonstrate and validate the proposed methodology for the FDI of an aerospace electromechanical actuator affected by incipient mechanical and electrical faults. We conducted both numerical and physical experiments where our multifidelity scheme is compared with the standard single-fidelity efficient global optimization algorithm (Section 8.2.6). The results show that our multifidelity FDI outperforms standard single-fidelity algorithms in terms of accuracy and acceleration of the inference. In particular, the non-myopic search identifies the exact EMA damage condition 95.4 % faster than the single-fidelity approach. The physical experiments confirm and validate the efficiency of the proposed framework: our strategy identifies with the non-myopic search the exact damage condition of the real-world EMA reducing the computational cost by 90.5 % if compared with the single-fidelity algorithm. The sensitive accelerations and improved accuracy of the FDI procedure are enabled by the combination of the efficient compression of the diagnostic signals through the encoding map, and the effective use of high-fidelity responses to accurately identify multiphysics damages. The results make the ambition of using high-fidelity monitoring more achievable, and even suggest the possibility of using them for nearly real-time diagnostics.

References

[1] Charles D Harris. Two-dimensional aerodynamic characteristics of the naca 0012 airfoil in the langley 8 foot transonic pressure tunnel. 1981.

[2] Matteo Davide Lorenzo Dalla Vedova, Pier Carlo Berri, and Stefano Re. Metaheuristic bio-inspired algorithms for prognostics: Application to on-board electromechanical actuators. In *2018 3rd International Conference on System Reliability and Safety (ICSRS)*, pages 273–279. IEEE, 2018.

[3] Joaquim RRA Martins and Andrew Ning. *Engineering design optimization*. Cambridge University Press, 2021.

[4] Nam-Ho Kim, Dawn An, and Joo-Ho Choi. Prognostics and health management of engineering systems. *Switzerland: Springer International Publishing*, 2017.

[5] Paul Upham, Janet Maughan, David Raper, and Callum Thomas. *Towards sustainable aviation*. Routledge, 2012.

[6] Frederico Afonso, Martin Sohst, Carlos MA Diogo, Simão S Rodrigues, Ana Ferreira, Inês Ribeiro, Ricardo Marques, Francisco FC Rego, Abdolrasoul Sohouli, Joana Portugal-Pereira, et al. Strategies towards a more sustainable aviation: A systematic review. *Progress in Aerospace Sciences*, 137:100878, 2023.

[7] Kavindu Ranasinghe, Roberto Sabatini, Alessandro Gardi, Suraj Bijjahalli, Rohan Kapoor, Thomas Fahey, and Kathiravan Thangavel. Advances in integrated system health management for mission-essential and safety-critical aerospace applications. *Progress in Aerospace Sciences*, 128:100758, 2022.

[8] Averill M Law and Michael G McComas. Simulation-based optimization. In *Proceedings of the Winter Simulation Conference*, volume 1, pages 41–44. IEEE, 2002.

[9] Abhijit Gosavi et al. *Simulation-based optimization*. Springer, 2015.

[10] Wilson Trigueiro de Sousa Junior, José Arnaldo Barra Montevechi, Rafael de Carvalho Miranda, and Afonso Teberga Campos. Discrete simulation-based optimization methods for industrial engineering problems: A systematic literature review. *Computers & Industrial Engineering*, 128:526–540, 2019.

[11] Charles Audet and Warren Hare. Derivative-free and blackbox optimization. 2017.

[12] Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Siam Review*, 60(3):550–591, 2018.

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

[14] MJ Mifsud, David G MacManus, and ST Shaw. A variable-fidelity aerodynamic model using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 82(10):646–663, 2016.

[15] Rhea P Liem, Charles A Mader, and Joaquim RRA Martins. Surrogate models and mixtures of experts in aerodynamic performance prediction for aircraft mission analysis. *Aerospace Science and Technology*, 43:126–151, 2015.

[16] Timothy W Simpson, Jesse D Poplinski, Patrick N Koch, and Janet K Allen. Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers*, 17:129–150, 2001.

[17] Felipe AC Viana, Christian Gogu, and Raphael T Haftka. Making the most out of surrogate models: tricks of the trade. In *International design engineering technical conferences and computers and information in engineering conference*, volume 44090, pages 587–598, 2010.

[18] Timothy Simpson, Vasilli Toropov, Vladimir Balabanov, and Felipe Viana. Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come-or not. In *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*, page 5802, 2008.

[19] Felipe AC Viana, Timothy W Simpson, Vladimir Balabanov, and Vasilli Toropov. Special section on multidisciplinary design optimization: metamodeling in multidisciplinary design optimization: how far have we really come? *AIAA journal*, 52(4):670–690, 2014.

[20] Raul Yondo, Esther Andrés, and Eusebio Valero. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Progress in aerospace sciences*, 96:23–61, 2018.

[21] Nestor V Queipo, Raphael T Haftka, Wei Shyy, Tushar Goel, Rajkumar Vaidyanathan, and P Kevin Tucker. Surrogate-based analysis and optimization. *Progress in aerospace sciences*, 41(1):1–28, 2005.

[22] Gary Wang and S. Shan. Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical Design - J MECH DESIGN*, 129, 04 2007.

[23] Raphael T Haftka, Diane Villanueva, and Anirban Chaudhuri. Parallel surrogate-assisted global optimization with expensive functions–a survey. *Structural and Multidisciplinary Optimization*, 54:3–13, 2016.

[24] Haitao Liu, Yew-Soon Ong, and Jianfei Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Structural and Multidisciplinary Optimization*, 57:393–416, 2018.

[25] Lisia Dias, Atharv Bhosekar, and Mariathi Ierapetritou. Adaptive sampling approaches for surrogate-based optimization. In *Computer Aided Chemical Engineering*, volume 47, pages 377–384. Elsevier, 2019.

[26] Jonas Mockus, Vytautas Tiesis, and Antanas Zilinskas. The application of bayesian methods for seeking the extremum. *Towards global optimization*, 2(117-129):2, 1978.

[27] Jonas Mockus. *Bayesian approach to global optimization: theory and applications*, volume 37. Springer Science & Business Media, 2012.

[28] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

[29] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

[30] Stefan Goertz, Mohammad Abu-Zurayk, Caslav Ilic, Tobias F Wunderlich, Stefan Keye, Matthias Schulze, Christoph Kaiser, Thomas Klimmek, Özge Süelözgen, Thiemo Kier, et al. Overview of collaborative multi-fidelity multidisciplinary design optimization activities in the dlr project victoria. In *AIAA Aviation 2020 Forum*, page 3167, 2020.

[31] Stanislav Karpuk, Valerio Mosca, Chuanzhen Liu, and Ali Elham. Development of a multi-fidelity design, analysis, and optimization environment for future transport aircraft. In *AIAA Scitech 2022 Forum*, page 0686, 2022.

[32] M Giselle Fernández-Godino, Chanyoung Park, Nam H Kim, and Raphael T Haftka. Issues in deciding whether to use multifidelity surrogates. *Aiaa Journal*, 57(5):2039–2054, 2019.

[33] Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Siam Review*, 60(3):550–591, 2018.

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

[35] Alexander IJ Forrester, András Sóbester, and Andy J Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the royal society a: mathematical, physical and engineering sciences*, 463(2088):3251–3269, 2007.

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

[37] Philip S Beran, Dean Bryson, Andrew S Thelen, Matteo Diez, and Andrea Serani. Comparison of multi-fidelity approaches for military vehicle design. In *AIAA AVIATION 2020 FORUM*, page 3158, 2020.

[38] Deng Huang, Theodore T Allen, William I Notz, and R Allen Miller. Sequential kriging optimization using multiple-fidelity evaluations. *Structural and Multidisciplinary Optimization*, 32(5):369–382, 2006.

[39] Kirthevasan Kandasamy, Gautam Dasarathy, Junier Oliva, Jeff Schneider, and Barnabás Póczos. Gaussian process optimisation with multi-fidelity evaluations. In *Proceedings of the 30th/International Conference on Advances in Neural Information Processing Systems (NIPS'30)*, pages 1–9, 2016.

[40] Yehong Zhang, Trong Nghia Hoang, Bryan Kian Hsiang Low, and Mohan Kankanhalli. Information-based multi-fidelity bayesian optimization. In *NIPS Workshop on Bayesian Optimization*, pages 1–5, 2017.

[41] Xiongfeng Ruan, Ping Jiang, Qi Zhou, Jiexiang Hu, and Leshi Shu. Variable-fidelity probability of improvement method for efficient global optimization of expensive black-box problems. *Structural and Multidisciplinary Optimization*, 62:3021–3052, 2020.

[42] Shion Takeno, Hitoshi Fukuoka, Yuhki Tsukada, Toshiyuki Koyama, Motoki Shiga, Ichiro Takeuchi, and Masayuki Karasuyama. Multi-fidelity bayesian optimization with max-value entropy search and its parallelization. In *International Conference on Machine Learning*, pages 9334–9345. PMLR, 2020.

[43] Jaroslaw Sobieszczanski-Sobieski. Multidisciplinary design optimization: an emerging new engineering discipline. In *Advances in structural optimization*, pages 483–496. Springer, 1995.

[44] Natalia M Alexandrov, M Yousuff Hussaini, et al. *Multidisciplinary design optimization: State of the art.* SIAM, 1997.

[45] Jeremy Agte, Olivier De Weck, Jaroslaw Sobieszczanski-Sobieski, Paul Arendsen, Alan Morris, and Martin Spieck. Mdo: assessment and direction for advancement—an opinion of one international group. *Structural and Multidisciplinary Optimization*, 40:17–33, 2010.

[46] Joaquim RRA Martins and Andrew B Lambe. Multidisciplinary design optimization: a survey of architectures. *AIAA journal*, 51(9):2049–2075, 2013.

[47] Rolf Isermann. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in control*, 29(1):71–85, 2005.

[48] Ali Zolghadri. Advanced model-based fdir techniques for aerospace systems: Today challenges and opportunities. *Progress in Aerospace Sciences*, 53:18–29, 2012.

[49] Julien Marzat, Hélène Piet-Lahanier, Frédéric Damongeot, and Eric Walter. Model-based fault diagnosis for aerospace systems: a survey. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of aerospace engineering*, 226(10):1329–1360, 2012.

[50] Ke-Shi Zhang, Zhong-Hua Han, Zhong-Jian Gao, and Yuan Wang. Constraint aggregation for large number of constraints in wing surrogate-based optimization. *Structural and Multidisciplinary Optimization*, 59:421–438, 2019.

[51] Amin Nazemian and Parviz Ghadimi. Multi-objective optimization of trimaran sidehull arrangement via surrogate-based approach for reducing resistance and improving the seakeeping performance. *Proceedings of the institution of mechanical engineers, part M: journal of engineering for the maritime environment*, 235(4):944–956, 2021.

[52] Truong Dang, Anh Vu Luong, Alan Wee Chung Liew, John McCall, and Tien Thanh Nguyen. Ensemble of deep learning models with surrogate-based optimization for medical image segmentation. In *2022 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8. IEEE, 2022.

[53] Halil Beglerovic, Michael Stolz, and Martin Horn. Testing of autonomous vehicles using surrogate models and stochastic optimization. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE, 2017.

[54] César Ramírez-Márquez, Edgar Martín-Hernández, Mariano Martín, and Juan Gabriel Segovia-Hernández. Surrogate based optimization of a process of polycrystalline silicon production. *Computers & Chemical Engineering*, 140:106870, 2020.

[55] Ryan Shaffer, Lucas Kocia, and Mohan Sarovar. Surrogate-based optimization for variational quantum algorithms. *Physical Review A*, 107(3):032415, 2023.

[56] Michael D McKay, Richard J Beckman, and William J Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, 2000.

[57] Richard F Gunst and Robert L Mason. Fractional factorial design. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(2):234–244, 2009.

[58] Douglas C Montgomery. *Design and analysis of experiments*. John wiley & sons, 2017.

[59] Anders Gustafsson, Andreas Herrmann, and Frank Huber. *Conjoint measurement: Methods and applications*. Springer Science & Business Media, 2013.

[60] Raymond H Myers, Douglas C Montgomery, and Christine M Anderson-Cook. *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons, 2016.

[61] Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770, 1959.

[62] Ruichen Jin, Wei Chen, and Agus Sudjianto. On sequential sampling for global metamodeling in engineering design. In *International design engineering technical conferences and computers and information in engineering conference*, volume 36223, pages 539–548, 2002.

[63] Foster Provost, David Jensen, and Tim Oates. Efficient progressive sampling. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 23–32, 1999.

[64] Francesco Di Fiore, Michela Nardelli, and Laura Mainini. Active learning and bayesian optimization: a unified perspective to learn with a goal. *Archives of Computational Methods in Engineering*, 2024.

[65] Francesco Di Fiore, Michela Nardelli, and Laura Mainini. Active learning and bayesian optimization: a unified perspective to learn with a goal. *arXiv preprint arXiv:2303.01560*, 2023.

[66] Ashwin Ram and David B Leake. *Goal-driven learning*. MIT press, 1995.

[67] J Tinsley Oden and Kumar S Vemaganti. Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials: I. error estimates and adaptive algorithms. *Journal of Computational Physics*, 164(1):22–47, 2000.

[68] Tan Bui-Thanh, Karen Willcox, Omar Ghattas, and Bart van Bloemen Waanders. Goal-oriented, model-constrained optimization for reduction of large-scale systems. *Journal of Computational Physics*, 224(2):880–896, 2007.

[69] Chad Lieberman and Karen Willcox. Goal-oriented inference: Approach, linear theory, and application to advection diffusion. *siam REVIEW*, 55(3):493–519, 2013.

[70] Francesco Di Fiore and Laura Mainini. Non-myopic multifidelity bayesian optimization. *arXiv preprint arXiv:2207.06325*, 2022.

[71] Francesco Grassi, Giorgio Manganini, Michele Garraffa, and Laura Mainini. Raal: Resource aware active learning for multifidelity efficient optimization. *AIAA Journal*, 61(6):2744–2753, 2023.

[72] Masashi Sugiyama and Shinichi Nakajima. Pool-based active learning in approximate linear regression. *Machine Learning*, 75(3):249–274, 2009.

[73] Dongrui Wu. Pool-based sequential active learning for regression. *IEEE transactions on neural networks and learning systems*, 30(5):1348–1359, 2018.

[74] Alexander Shapiro. Monte carlo sampling methods. *Handbooks in operations research and management science*, 10:353–425, 2003.

[75] Art B Owen. Quasi-monte carlo sampling. *Monte Carlo Ray Tracing: Siggraph*, 1:69–88, 2003.

[76] A Karamchandani, P Bjerager, and CA Cornell. Adaptive importance sampling. In *Structural Safety and Reliability*, pages 855–862. ASCE, 1989.

[77] Monica F Bugallo, Victor Elvira, Luca Martino, David Luengo, Joaquin Miguez, and Petar M Djuric. Adaptive importance sampling: The past, the present, and the future. *IEEE Signal Processing Magazine*, 34(4):60–79, 2017.

[78] Benjamin Peherstorfer, Tiangang Cui, Youssef Marzouk, and Karen Willcox. Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering*, 300:490–509, 2016.

[79] Yves F Atchadé and Jeffrey S Rosenthal. On adaptive markov chain monte carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

[80] Yves Atchade, Gersende Fort, Eric Moulines, and Pierre Priouret. Adaptive markov chain monte carlo: theory and methods. *Bayesian time series models*, 1, 2011.

[81] Mónica F Bugallo, Luca Martino, and Jukka Corander. Adaptive importance sampling in signal processing. *Digital Signal Processing*, 47:36–49, 2015.

[82] Xiuzhuang Zhou, Yao Lu, Jiwen Lu, and Jie Zhou. Abrupt motion tracking via intensively adaptive markov-chain monte carlo sampling. *IEEE Transactions on Image Processing*, 21(2):789–801, 2011.

[83] Ning-Cong Xiao, Hongyou Zhan, and Kai Yuan. A new reliability method for small failure probability problems by combining the adaptive importance sampling and surrogate models. *Computer Methods in Applied Mechanics and Engineering*, 372:113336, 2020.

[84] HA Jensen, DJ Jerez, and M Valdebenito. An adaptive scheme for reliability-based global design optimization: A markov chain monte carlo approach. *Mechanical Systems and Signal Processing*, 143:106836, 2020.

[85] Christian P Robert, George Casella, and George Casella. *Monte Carlo statistical methods*, volume 2. Springer, 1999.

[86] Tim Hesterberg. Weighted average importance sampling and defensive mixture distributions. *Technometrics*, 37(2):185–194, 1995.

[87] Luca Martino, Victor Elvira, David Luengo, and Jukka Corander. An adaptive population importance sampler: Learning from uncertainty. *IEEE Transactions on Signal Processing*, 63(16):4422–4437, 2015.

[88] Yousef El-Laham, Luca Martino, Víctor Elvira, and Mónica F Bugallo. Efficient adaptive multiple importance sampling. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5. IEEE, 2019.

[89] John D Jakeman and Stephen G Roberts. Local and dimension adaptive stochastic collocation for uncertainty quantification. In *Sparse grids and applications*, pages 181–203. Springer, 2012.

[90] John D Jakeman, Michael S Eldred, Gianluca Geraci, and Alex Gorodetsky. Adaptive multi-index collocation for uncertainty quantification and sensitivity analysis. *International Journal for Numerical Methods in Engineering*, 121(6):1314–1343, 2020.

[91] Max Gunzburger, Clayton G Webster, and Guannan Zhang. An adaptive wavelet stochastic collocation method for irregular solutions of partial differential equations with random input data. In *Sparse Grids and Applications-Munich 2012*, pages 137–170. Springer, 2014.

[92] Martin Eigel, Oliver G Ernst, Bjorn Sprungk, and Lorenzo Tamellini. On the convergence of adaptive stochastic collocation for elliptic partial differential equations with affine diffusion. *SIAM Journal on Numerical Analysis*, 60(2):659–687, 2022.

[93] Youssef Marzouk and Dongbin Xiu. A stochastic collocation approach to bayesian inference in inverse problems. *Communications in Computational Physics*, 6(4):826–847, 2009.

[94] Xiang Ma and Nicholas Zabaras. An efficient bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method. *Inverse Problems*, 25(3):035013, 2009.

[95] Abdul-Lateef Haji-Ali, Fabio Nobile, Lorenzo Tamellini, and Raúl Tempone. Multi-index stochastic collocation for random pdes. *Computer Methods in Applied Mechanics and Engineering*, 306:95–122, 2016.

[96] Jens Lang, Robert Scheichl, and David Silvester. A fully adaptive multilevel stochastic collocation strategy for solving elliptic pdes with random data. *Journal of Computational Physics*, 419:109692, 2020.

[97] Thomas Gerstner and Michael Griebel. Dimension–adaptive tensor–product quadrature. *Computing*, 71:65–87, 2003.

[98] Diane Guignard and Fabio Nobile. A posteriori error estimation for the stochastic collocation finite element method. *SIAM Journal on Numerical Analysis*, 56(5):3121–3143, 2018.

[99] Chenxi Wu, Min Zhu, Qinyang Tan, Yadhu Kartha, and Lu Lu. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 403:115671, 2023.

[100] Lu Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. Deepxde: A deep learning library for solving differential equations. *SIAM review*, 63(1):208–228, 2021.

[101] Mohammad Amin Nabian, Rini Jasmine Gladstone, and Hadi Meidani. Efficient training of physics-informed neural networks via importance sampling. *Computer-Aided Civil and Infrastructure Engineering*, 36(8):962–977, 2021.

[102] Burr Settles. Active learning literature survey. 2009.

[103] Xueying Zhan, Huan Liu, Qing Li, and Antoni B Chan. A comparative survey: Benchmarking for pool-based active learning. In *IJCAI*, pages 4679–4686, 2021.

[104] Harold J Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. 1964.

[105] AG Zhilinskas. Single-step bayesian search method for an extremum of functions of a single variable. *Cybernetics*, 11(1):160–166, 1975.

[106] Jonas Močkus. On bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference: Novosibirsk, July 1–7, 1974*, pages 400–404. Springer, 1975.

[107] Bruce E Stuckman. A global search method for optimizing nonlinear systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(6):965–977, 1988.

[108] John F Elder. Global r/sup d/optimization when probes are expensive: the grope algorithm. In *[Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics*, pages 577–582. IEEE, 1992.

[109] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[110] Thomas A Zang. *Needs and opportunities for uncertainty-based multidisciplinary design methods for aerospace vehicles*. National Aeronautics and Space Administration, Langley Research Center, 2002.

[111] Tai-Pang Wu and Chi-Keung Tang. A bayesian approach for shadow extraction from a single image. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 480–487 Vol. 1, 2005.

[112] Jingpeng Li and Uwe Aickelin. A bayesian optimization algorithm for the nurse scheduling problem. In *The 2003 Congress on Evolutionary Computation, 2003. CEC'03.*, volume 3, pages 2149–2156. IEEE, 2003.

[113] Peter I Frazier and Jialei Wang. Bayesian optimization for materials design. *Information science for materials discovery and design*, pages 45–75, 2016.

[114] Rémi Lam, Matthias Poloczek, Peter Frazier, and Karen E Willcox. Advances in bayesian optimization with applications in aerospace engineering. In *2018 AIAA Non-Deterministic Approaches Conference*, page 1656, 2018.

[115] Huifang Kong, Jiapeng Yan, Hai Wang, and Lei Fan. Energy management strategy for electric vehicles based on deep q-learning using bayesian optimization. *Neural Computing and Applications*, 32:14431–14445, 2020.

[116] Rémy Charayron, Thierry Lefebvre, Nathalie Bartoli, and Joseph Morlier. Multi-fidelity bayesian optimization strategy applied to overall drone design. In *AIAA SCITECH 2023 Forum*, page 2366, 2023.

[117] Felix Berkenkamp, Andreas Krause, and Angela P Schoellig. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. *Machine Learning*, pages 1–35, 2021.

[118] Sreejith Balakrishnan, Quoc Phong Nguyen, Bryan Kian Hsiang Low, and Harold Soh. Efficient exploration of reward functions in inverse reinforcement learning via bayesian optimization. *Advances in Neural Information Processing Systems*, 33:4187–4198, 2020.

[119] M Todd Young, Jacob D Hinkle, Ramakrishnan Kannan, and Arvind Ramanathan. Distributed bayesian optimization of deep reinforcement learning algorithms. *Journal of Parallel and Distributed Computing*, 139:43–52, 2020.

[120] Joan Gonzalvez, Edmond Lezmi, Thierry Roncalli, and Jiali Xu. Financial applications of gaussian processes and bayesian optimization. *arXiv preprint arXiv:1903.04841*, 2019.

[121] Ehsan Sadeghi Pour, Hossein Jafari, Ali Lashgari, Elaheh Rabiee, and Amin Ahmadisharaf. Cryptocurrency price prediction with neural networks of lstm and bayesian optimization. *European Journal of Business and Management Research*, 7(2):20–27, 2022.

[122] A Helen Victoria and Ganesh Maragatham. Automatic tuning of hyperparameters using bayesian optimization. *Evolving Systems*, 12:217–223, 2021.

[123] Ryan Turner, David Eriksson, Michael McCourt, Juha Kiili, Eero Laaksonen, Zhen Xu, and Isabelle Guyon. Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020. In *NeurIPS 2020 Competition and Demonstration Track*, pages 3–26. PMLR, 2021.

[124] John J Dudley, Jason T Jacques, and Per Ola Kristensson. Crowdsourcing interface feature design with bayesian optimization. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–12, 2019.

[125] Yuki Koyama, Issei Sato, and Masataka Goto. Sequential gallery for interactive visual design optimization. *ACM Transactions on Graphics (TOG)*, 39(4):88–1, 2020.

[126] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

[127] Michael A Osborne, Roman Garnett, and Stephen J Roberts. Gaussian processes for global optimization. In *3rd international conference on learning and intelligent optimization (LION3)*, pages 1–15. Citeseer, 2009.

[128] Harold J Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. 1964.

[129] Philipp Hennig and Christian J Schuler. Entropy search for information-efficient global optimization. *Journal of Machine Learning Research*, 13(6), 2012.

[130] Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient bayesian optimization. In *International Conference on Machine Learning*, pages 3627–3635. PMLR, 2017.

[131] Warren Scott, Peter Frazier, and Warren Powell. The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression. *SIAM Journal on Optimization*, 21(3):996–1026, 2011.

[132] Remi Lam and Karen Willcox. Lookahead bayesian optimization with inequality constraints. *Advances in Neural Information Processing Systems*, 30, 2017.

[133] Jian Wu and Peter Frazier. Practical two-step lookahead bayesian optimization. *Advances in neural information processing systems*, 32, 2019.

[134] David Ginsbourger and Rodolphe Le Riche. Towards gaussian process-based optimization with finite time horizon. In *mODa 9–Advances in Model-Oriented Design and Analysis*, pages 89–96. Springer, 2010.

[135] Remi Lam, Karen Willcox, and David H Wolpert. Bayesian optimization with a finite budget: An approximate dynamic programming approach. *Advances in Neural Information Processing Systems*, 29:883–891, 2016.

[136] Javier González, Michael Osborne, and Neil Lawrence. Glasses: Relieving the myopia of bayesian optimisation. In *Artificial Intelligence and Statistics*, pages 790–799. PMLR, 2016.

[137] Jian Wu and Peter Frazier. Practical two-step lookahead bayesian optimization. *Advances in neural information processing systems*, 32:9813–9823, 2019.

[138] Hao Wang, Bas van Stein, Michael Emmerich, and Thomas Back. A new acquisition function for bayesian optimization based on the moment-generating function. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 507–512. IEEE, 2017.

[139] Roy de Winter, Bas van Stein, and Thomas Bäck. Multi-point acquisition function for constraint parallel efficient multi-objective optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 511–519, 2022.

[140] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[141] Eric Schulz, Maarten Speekenbrink, and Andreas Krause. A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85:1–16, 2018.

[142] Felix Berkenkamp, Andreas Krause, and Angela P Schoellig. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. *Machine Learning*, pages 1–35, 2021.

[143] Majdi Mansouri, Mohamed-Faouzi Harkat, Hazem Nounou, and Mohamed N Nounou. *Data-driven and model-based methods for fault detection and diagnosis*. Elsevier, 2020.

[144] L Bonfiglio, P Perdikaris, S Brizzolara, and GE Karniadakis. Multi-fidelity optimization of super-cavitating hydrofoils. *Computer Methods in Applied Mechanics and Engineering*, 332:63–85, 2018.

[145] Mostafa Meliani, Nathalie Bartoli, Thierry Lefebvre, Mohamed-Amine Bouhlel, Joaquim RRA Martins, and Joseph Morlier. Multi-fidelity efficient global optimization: Methodology and application to airfoil shape design. In *AIAA aviation 2019 forum*, page 3236, 2019.

[146] Francesco Di Fiore, Paolo Maggiore, and Laura Mainini. Multifidelity domain-aware learning for the design of re-entry vehicles. *Structural and Multidisciplinary Optimization*, 64(5):3017–3035, 2021.

[147] Francesco Di Fiore and Laura Mainini. Nm-mf: Non-myopic multifidelity framework for constrained multi-regime aerodynamic optimization. *AIAA Journal*, 61(3):1270–1280, 2023.

[148] Jian Wu, Saul Toscano-Palmerin, Peter I Frazier, and Andrew Gordon Wilson. Practical multi-fidelity bayesian optimization for hyperparameter tuning. In *Uncertainty in Artificial Intelligence*, pages 788–798. PMLR, 2020.

[149] Kirthevasan Kandasamy, Gautam Dasarathy, Jeff Schneider, and Barnabás Póczos. Multi-fidelity bayesian optimisation with continuous approximations. In *International Conference on Machine Learning*, pages 1799–1808. PMLR, 2017.

[150] Mahdi Imani and Seyede Fatemeh Ghoreishi. Scalable inverse reinforcement learning through multifidelity bayesian optimization. *IEEE transactions on neural networks and learning systems*, 2021.

[151] Faran Irshad, Stefan Karsch, and Andreas Döpp. Multi-objective and multi-fidelity bayesian optimization of laser-plasma acceleration. *Physical Review Research*, 5(1):013063, 2023.

[152] JM Winter, R Abaidi, JWJ Kaiser, S Adami, and NA Adams. Multi-fidelity bayesian optimization to solve the inverse stefan problem. *Computer Methods in Applied Mechanics and Engineering*, 410:115946, 2023.

[153] Paris Perdikaris and George Em Karniadakis. Model inversion via multi-fidelity bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond. *Journal of The Royal Society Interface*, 13(118):20151107, 2016.

[154] Simone Pezzuto, Paris Perdikaris, and Francisco Sahli Costabal. Learning cardiac activation maps from 12-lead ecg with multi-fidelity bayesian optimization on manifolds. *arXiv preprint arXiv:2203.06222*, 2022.

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

[156] Xiongfeng Ruan, Ping Jiang, Qi Zhou, Jiexiang Hu, and Leshi Shu. Variable-fidelity probability of improvement method for efficient global optimization of expensive black-box problems. *Structural and Multidisciplinary Optimization*, 62(6):3021–3052, 2020.

[157] Yehong Zhang, Trong Nghia Hoang, Bryan Kian Hsiang Low, and Mohan Kankanhalli. Information-based multi-fidelity bayesian optimization. In *NIPS Workshop on Bayesian Optimization*, 2017.

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

[159] Naoki Abe. Query learning strategies using boosting and bagging. *Proc. of 15^< th> Int. Cmf. on Machine Learning (ICML98)*, pages 1–9, 1998.

[160] Robert Burbidge, Jem J Rowland, and Ross D King. Active learning for regression based on query by committee. In *International conference on intelligent data engineering and automated learning*, pages 209–218. Springer, 2007.

[161] Wenbin Cai, Muhan Zhang, and Ya Zhang. Batch mode active learning for regression with expected model change. *IEEE transactions on neural networks and learning systems*, 28(7):1668–1681, 2016.

[162] Beguem Demir and Lorenzo Bruzzone. A multiple criteria active learning method for support vector regression. *Pattern recognition*, 47(7):2558–2567, 2014.

[163] Tirthankar RayChaudhuri and Leonard GC Hamey. Minimisation of data collection by active learning. In *Proceedings of ICNN'95-International Conference on Neural Networks*, volume 3, pages 1338–1341. IEEE, 1995.

[164] Burr Settles and Mark Craven. An analysis of active learning strategies for sequence labeling tasks. In *proceedings of the 2008 conference on empirical methods in natural language processing*, pages 1070–1079, 2008.

[165] H Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 287–294, 1992.

[166] Tianxu He, Shukui Zhang, Jie Xin, Pengpeng Zhao, Jian Wu, Xuefeng Xian, Chunhua Li, and Zhiming Cui. An active learning approach with uncertainty, representativeness, and diversity. *The Scientific World Journal*, 2014, 2014.

[167] Dan Shen, Jie Zhang, Jian Su, Guodong Zhou, and Chew Lim Tan. Multi-criteria-based active learning for named entity recognition. In *Proceedings of the 42nd annual meeting of the Association for Computational Linguistics (ACL-04)*, pages 589–596, 2004.

[168] Dongrui Wu, Vernon J Lawhern, Stephen Gordon, Brent J Lance, and Chin-Teng Lin. Offline eeg-based driver drowsiness estimation using enhanced batch-mode active learning (ebmal) for regression. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 000730–000736. IEEE, 2016.

[169] Robert Munro Monarch. *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI.* Simon and Schuster, 2021.

[170] David D Lewis and Jason Catlett. Heterogeneous uncertainty sampling for supervised learning. In *Machine learning proceedings 1994*, pages 148–156. Elsevier, 1994.

[171] Maria-Florina Balcan, Andrei Broder, and Tong Zhang. Margin based active learning. In *International Conference on Computational Learning Theory*, pages 35–50. Springer, 2007.

[172] Mingkun Li and Ishwar K Sethi. Confidence-based active learning. *IEEE transactions on pattern analysis and machine intelligence*, 28(8):1251–1261, 2006.

[173] Alex Holub, Pietro Perona, and Michael C Burl. Entropy-based active learning for object recognition. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8. IEEE, 2008.

[174] David Cohn. Neural network exploration using optimal experiment design. *Advances in neural information processing systems*, 6, 1993.

[175] Wenbin Cai, Ya Zhang, and Jun Zhou. Maximizing expected model change for active learning in regression. In *2013 IEEE 13th international conference on data mining*, pages 51–60. IEEE, 2013.

[176] Yue Zhao, Ciwen Xu, and Yongcun Cao. Research on query-by-committee method of active learning and application. In *International Conference on Advanced Data Mining and Applications*, pages 985–991. Springer, 2006.

[177] Sanjoy Dasgupta and Daniel Hsu. Hierarchical sampling for active learning. In *Proceedings of the 25th international conference on Machine learning*, pages 208–215, 2008.

[178] Ozan Sener and Silvio Savarese. A geometric approach to active learning for convolutional neural networks. *arXiv preprint arXiv:1708.00489*, 7, 2017.

[179] Rita Chattopadhyay, Zheng Wang, Wei Fan, Ian Davidson, Sethuraman Panchanathan, and Jieping Ye. Batch mode active sampling based on marginal probability distribution matching. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(3):1–25, 2013.

[180] Yifan Fu, Bin Li, Xingquan Zhu, and Chengqi Zhang. Active learning without knowing individual instance labels: a pairwise label homogeneity query approach. *IEEE Transactions on Knowledge and Data Engineering*, 26(4):808–822, 2013.

[181] Tobias Reitmaier, Adrian Calma, and Bernhard Sick. Transductive active learning–a new semi-supervised learning approach based on iteratively refined generative models to capture structure in data. *Information Sciences*, 293:275–298, 2015.

[182] Yu Zhao, Zhenhui Shi, Jingyang Zhang, Dong Chen, and Lixu Gu. A novel active learning framework for classification: Using weighted rank aggregation to achieve multiple query criteria. *Pattern Recognition*, 93:581–602, 2019.

[183] Wei-Ning Hsu and Hsuan-Tien Lin. Active learning by learning. In *Twenty-Ninth AAAI conference on artificial intelligence*, 2015.

[184] Xin Li and Yuhong Guo. Adaptive active learning for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 859–866, 2013.

[185] Zheng Wang and Jieping Ye. Querying discriminative and representative samples for batch mode active learning. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 9(3):1–23, 2015.

[186] Ying-Peng Tang and Sheng-Jun Huang. Self-paced active learning: Query the right thing at the right time. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5117–5124, 2019.

[187] Donald R Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, 21(4):345–383, 2001.

[188] Quoc Phong Nguyen, Bryan Kian Hsiang Low, and Patrick Jaillet. Rectified max-value entropy search for bayesian optimization. *arXiv preprint arXiv:2202.13597*, 2022.

[189] Liyue Zhao, Gita Sukthankar, and Rahul Sukthankar. Incremental relabeling for active learning with noisy crowdsourced annotations. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 728–733. IEEE, 2011.

[190] Pinar Donmez, Jaime G Carbonell, and Jeff Schneider. Efficiently learning the accuracy of labeling sources for selective sampling. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 259–268, 2009.

[191] Panagiotis G Ipeirotis, Foster Provost, Victor S Sheng, and Jing Wang. Repeated labeling using multiple noisy labelers. *Data Mining and Knowledge Discovery*, 28(2):402–441, 2014.

[192] Yan Yan, Romer Rosales, Glenn Fung, and Jennifer G Dy. Active learning from crowds. In *ICML*, 2011.

[193] Yan Yan, Rómer Rosales, Glenn Fung, Faisal Farooq, Bharat Rao, and Jennifer Dy. Active learning from multiple knowledge sources. In *Artificial Intelligence and Statistics*, pages 1350–1357. PMLR, 2012.

[194] Meng Fang, Jie Yin, and Dacheng Tao. Active learning for crowdsourcing using knowledge transfer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

[195] Sheng-Jun Huang, Jia-Lve Chen, Xin Mu, and Zhi-Hua Zhou. Cost-effective active learning from diverse labelers. In *IJCAI*, pages 1879–1885, 2017.

[196] Guoxian Yu, Xia Chen, Carlotta Domeniconi, Jun Wang, Zhao Li, Zili Zhang, and Xiangliang Zhang. Cmal: Cost-effective multi-label active learning by querying subexamples. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[197] Ruijiang Gao and Maytal Saar-Tsechansky. Cost-accuracy aware adaptive labeling for active learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2569–2576, 2020.

[198] Yixin Liu, Shishi Chen, Fenggang Wang, and Fenfen Xiong. Sequential optimization using multi-level cokriging and extended expected improvement criterion. *Structural and Multidisciplinary Optimization*, 58:1155–1173, 2018.

[199] Francesco Di Fiore and Laura Mainini. Non-myopic multipoint multifidelity bayesian framework for multidisciplinary design. *Scientific Reports*, 13(1):22531, 2023.

[200] Francesco Di Fiore and Laura Mainini. Nm2-bo: Non-myopic multifidelity bayesian optimization. *Knowledge-based systems (Under Review)*, 2023.

[201] Martin L Puterman. Markov decision processes. *Handbooks in operations research and management science*, 2:331–434, 1990.

[202] Dimitri Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 1. Athena scientific, 1995.

[203] Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.

[204] James T Wilson, Frank Hutter, and Marc Peter Deisenroth. Maximizing acquisition functions for bayesian optimization. *arXiv preprint arXiv:1805.10196*, 2018.

[205] Michael D McKay. Latin hypercube sampling as a tool in uncertainty analysis of computer models. In *Proceedings of the 24th conference on Winter simulation*, pages 557–564, 1992.

[206] Javier González, Zhenwen Dai, Philipp Hennig, and Neil Lawrence. Batch bayesian optimization via local penalization. In *Artificial intelligence and statistics*, pages 648–657. PMLR, 2016.

[207] Serani A. Rumpfkeil M. P. Minisci E. Quagliarella D. Pehilvan H. Yildiz S. Ficini S. Pellegrini R. Bryson D. Nikbay M. Diez M. Beran P. Mainini L., Di Fiore F. Analytical benchmark problems for multifidelity optimization methods. *arXiv preprint arXiv:2204.07867*, 2022.

[208] Serani A. Rumpfkeil M. P. Minisci E. Quagliarella D. Pehilvan H. Yildiz S. Ficini S. Pellegrini R. Bryson D. Nikbay M. Diez M. Beran P. Mainini L., Di Fiore F. Analytical benchmark problems for multifidelity optimization methods. *NATO STO AVT-354 Research Workshop on Multi-fidelity methods for military vehicle design, Varna, Bulgaria*, 2022.

[209] Felipe AC Viana. A tutorial on latin hypercube design of experiments. *Quality and reliability engineering international*, 32(5):1975–1985, 2016.

[210] Alexander Forrester, Andras Sobester, and Andy Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.

[211] Seungjoon Lee, Felix Dietrich, George E Karniadakis, and Ioannis G Kevrekidis. Linking gaussian process regression with data-driven manifold embeddings for nonlinear data fusion. *Interface focus*, 9(3):20180083, 2019.

[212] Handing Wang, Yaochu Jin, and John Doherty. A generic test suite for evolutionary multifidelity optimization. *IEEE Transactions on Evolutionary Computation*, 22(6):836–850, 2017.

[213] Daniel L Clark Jr, Ha-Rok Bae, Koorosh Gobal, and Ravi Penmetsa. Engineering design exploration using locally optimized covariance kriging. *AIAA Journal*, 54(10):3160–3175, 2016.

[214] HoHo Rosenbrock. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3):175–184, 1960.

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

[216] Shifeng Xiong, Peter ZG Qian, and CF Jeff Wu. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1):37–46, 2013.

[217] Francesco Di Fiore and Laura Mainini. Physics-aware multifidelity bayesian optimization: A generalized formulation. *Computers & Structures*, 296:107302, 2024.

[218] Francesco Di Fiore and Laura Mainini. *Non-Myopic Multifidelity Method for Multi-regime Constrained Aerodynamic Optimization*, page 3716. 2022.

[219] Francesco Di Fiore and Laura Mainini. Multifidelity domain-aware scheme for cross-regime airfoil shape optimization. In *Proceedings of the Eleventh International Conference on Engineering Computational Technology*, volume 300. Elsevier, 2022.

[220] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[221] Francesco Di Fiore, Pier Carlo Berri, and Laura Mainini. Rapid assessment of incipient multimodal faults of complex aerospace systems. In *AIAA SCITECH 2023 Forum*, page 1092, 2023.

[222] Francesco Di Fiore, Pier Carlo Berri, and Laura Mainini. Rapid High-Fidelity Diagnostics of Incipient Faults for a Faster Adoption of More Sustainable Aerospace Technologies. *AIAA Journal (Under Review)*, 2023.

[223] Francesco Di Fiore, Pier Carlo Berri, and Laura Mainini. FREEDOM: Validated Method for Rapid Assessment of Incipient Faults of Aerospace Systems. *AIAA Journal*, 62(2):776–790, 2024.

[224] Francesco Di Fiore, Pier Carlo Berri, and Laura Mainini. Multifidelity framework for the efficient identification of damages in complex aerospace systems. In *AIAA AVIATION 2023 Forum*, page 4449, 2023.

[225] Pier Carlo Berri, Matteo Davide Lorenzo Dalla Vedova, and Laura Mainini. Real-time fault detection and prognostics for aircraft actuation systems. In *AIAA Scitech 2019 Forum*, page 2210, 2019.

[226] Laura Mainini. Structural assessment and sensor placement strategy for self-aware aerospace vehicles. *Structural Health Monitoring 2017*, 2017.

[227] Laura Mainini and Karen E Willcox. Sensor placement strategy to inform decisions. In *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 3820, 2017.

[228] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.

[229] Peter J Schmid, Larry Li, Matthew P Juniper, and O Pust. Applications of the dynamic mode decomposition. *Theoretical and Computational Fluid Dynamics*, 25(1):249–259, 2011.

[230] Michael E Wall, Andreas Rechtsteiner, and Luis M Rocha. Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, pages 91–109. Springer, 2003.

[231] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[232] Teuvo Kohonen. *Self-organizing maps*, volume 30. Springer Science & Business Media, 2012.

[233] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.

[234] Jialin Song, Yuxin Chen, and Yisong Yue. A general framework for multi-fidelity bayesian optimization with gaussian processes. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3158–3167. PMLR, 2019.

[235] Clark D. Bryson D. Beran P. Yildiz S. Nikbay M. Mainini L. Di Fiore F. Minisci E. Leyland P. Serani A. Diez M. Quagliarella D., Thelen A. Reproducible industrial multifidelity optimization benchmark problems for air, space, and sea vehicles. *NATO STO AVT-354 Research Workshop on Multi-fidelity methods for military vehicle design, Varna, Bulgaria*, 2022.

[236] Domenico Quagliarella and Matteo Diez. An open-source aerodynamic framework for benchmarking multi-fidelity methods. In *AIAA AVIATION 2020 FORUM*, page 3179, 2020.

[237] Thomas D Economon, Francisco Palacios, Sean R Copeland, Trent W Lukaczyk, and Juan J Alonso. Su2: An open-source suite for multiphysics simulation and design. *Aiaa Journal*, 54(3):828–846, 2016.

[238] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, 79(11):1309–1331, 2009.

[239] Mark Drela. Pros and cons of airfoil optimization. *Frontiers of computational fluid dynamics*, 1998:363–381, 1998.

[240] Wu Li, Luc Huyse, and Sharon Padula. Robust airfoil optimization to achieve drag reduction over a range of mach numbers. *Structural and Multidisciplinary Optimization*, 24(1):38–50, 2002.

[241] Donald F Elger, Barbara A LeBret, Clayton T Crowe, and John A Roberson. *Engineering fluid mechanics*. John Wiley & Sons, 2020.

[242] Donald F Young, Bruce R Munson, Theodore H Okiishi, and Wade W Huebsch. *A brief introduction to fluid mechanics*. John Wiley & Sons, 2010.

[243] Deng Huang, Theodore T Allen, William I Notz, and R Allen Miller. Sequential kriging optimization using multiple-fidelity evaluations. *Structural and Multidisciplinary Optimization*, 32(5):369–382, 2006.

[244] Rémi Lam, Douglas L Allaire, and Karen E Willcox. Multifidelity optimization using statistical surrogate modeling for non-hierarchical information sources. In *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 0143, 2015.

[245] Seyede Fatemeh Ghoreishi and Douglas L Allaire. A fusion-based multi-information source optimization approach using knowledge gradient policies. In *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 1159, 2018.

[246] Anna Bertram and Ralf Zimmermann. Theoretical investigations of the new cokriging method for variable-fidelity surrogate modeling. *Advances in Computational Mathematics*, 44(6):1693–1716, 2018.

[247] Qi Zhou, Yuda Wu, Zhendong Guo, Jiexiang Hu, and Peng Jin. A generalized hierarchical co-kriging model for multi-fidelity data fusion. *Structural and Multidisciplinary Optimization*, 62(4):1885–1904, 2020.

[248] JJ Thibert, M Grandjacques, LH Ohman, et al. Naca 0012 airfoil. *AGARD Advisory Report*, 138, 1979.

[249] Christopher Rumsey, B Smith, and G Huang. 2d naca 0012 airfoil validation case. *Turbulence Modeling Resource, NASA Langley Research Center*, page 33, 2014.

[250] Thomas D Economon, Francisco Palacios, Sean R Copeland, Trent W Lukaczyk, and Juan J Alonso. Su2: An open-source suite for multiphysics simulation and design. *Aiaa Journal*, 54(3):828–846, 2016.

[251] Mark Drela. Xfoil: An analysis and design system for low reynolds number airfoils. In *Low Reynolds number aerodynamics*, pages 1–12. Springer, 1989.

[252] Ryan Barrett and Andrew Ning. Comparison of airfoil precomputational analysis methods for optimization of wind turbine blades. *IEEE Transactions on Sustainable Energy*, 7(3):1081–1088, 2016.

[253] András Sobester, Alexander Forrester, and Andy Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.

[254] Nicolas P Bons, Joaquim RRA Martins, Felipe IK Odaguil, and Ana Paula C Cuco. Aerostructural wing optimization of a regional jet considering mission fuel burn. *ASME Open Journal of Engineering*, 1, 2022.

[255] D Steward. The design structure matrix: A method for managing the deisgn of complex systems. *IEEE Transactions on Engineering Management*, 28(1981):pp, 1981.

[256] Richard Harris, Michael Stewart, and William Koenig. Thermal protection systems technology transfer from apollo and space shuttle to the orion program. In *2018 AIAA SPACE and Astronautics Forum and Exposition*, page 5134, 2018.

[257] Brian R Hollis and Salvatore Borrelli. Aerothermodynamics of blunt body entry vehicles. *Progress in Aerospace Sciences*, 48:42–56, 2012.

[258] US Standard Atmosphere. National oceanic and atmospheric administration. *National Aeronautics and Space Administration, United States Air Force, Washington, DC*, 1976.

[259] Antonio Viviani and Giuseppe Pezzella. Heat transfer analysis for a winged reentry flight test bed. *International Journal of Engineering*, 3(3):329–345, 2009.

[260] Thomas H Squire and Jochen Marschall. Material property requirements for analysis and design of uhtc components in hypersonic applications. *Journal of the European Ceramic Society*, 30(11):2239–2251, 2010.

[261] George P Sutton and Oscar Biblarz. *Rocket propulsion elements*. John Wiley & Sons, 2016.

[262] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, 79(11):1309–1331, 2009.

[263] Francisco Palacios, Michael R Colonno, Aniket C Aranake, Alejandro Campos, Sean R Copeland, Thomas D Economon, Amrita K Lonkar, Trent W Lukaczyk, Thomas WR Taylor, and Juan J Alonso. Stanford university unstructured (su2): An open-source integrated computational environment for multi-physics simulation and design. *AIAA paper*, 287:2013, 2013.

[264] Klaus Oswatitsch. Ähnlichkeitsgesetze für hyperschallströmung. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, 2(4):249–264, 1951.

[265] Kenneth Sutton and Randokh A Graves Jr. A general stagnation-point convective-heating equation for arbitrary gas mixtures. *Work*, 50:7885, 1971.

[266] Michael E Tauber and Kenneth Sutton. Stagnation-point radiative heating relations for earth and mars entries. *Journal of Spacecraft and Rockets*, 28(1):40–42, 1991.

[267] Michael J Wright, Dinesh K Prabhu, and Edward R Martinez. Analysis of apollo command module afterbody heating part i: As-202. *Journal of Thermophysics and heat Transfer*, 20(1):16–30, 2006.

[268] Kerry A Trumble, Ioana Cozmuta, Steve Sepka, Peter Jenniskens, and Michael Winter. Postflight aerothermal analysis of the stardust sample return capsule. *Journal of Spacecraft and Rockets*, 47(5):765–774, 2010.

[269] Jay Grinstead, Peter Jenniskens, Alan Cassell, Jim Albers, and Michael Winter. Airborne observation of the hayabusa sample return capsule re-entry. In *42nd AIAA Thermophysics Conference*, page 3329, 2011.

[270] Victor Giurgiutiu. Structural health monitoring of aerospace composites. 2015.

[271] Hexcel Corporation. Hexply® 8552 product datasheet. 2016.

[272] RD0044 Mindlin. Influence of rotatory inertia and shear on flexural motions of isotropic, elastic plates. 1951.

[273] Ana Garcia Garriga, Sangeeth S Ponnusamy, and Laura Mainini. A multi-fidelity framework to support the design of more-electric actuation. In *AIAA AVIATION 2018 Multidisciplinary Analysis and Optimization Conference*, page 3741, 2018.

[274] Ana Garcia Garriga, Laura Mainini, and Sangeeth Saagar Ponnusamy. A machine learning enabled multi-fidelity platform for the integrated design of aircraft systems. *Journal of Mechanical Design*, 141(12), 2019.

[275] Andreas W Schäfer, Steven RH Barrett, Khan Doyme, Lynnette M Dray, Albert R Gnadt, Rod Self, Aidan O'Sullivan, Athanasios P Synodinos, and Antonio J Torija. Technological, economic and environmental prospects of all-electric aircraft. *Nature Energy*, 4(2):160–166, 2019.

[276] Pat Wheeler. Technology for the more and all electric aircraft of the future. In *2016 IEEE International Conference on Automatica (ICA-ACCA)*, pages 1–5. IEEE, 2016.

[277] Ashkan Barzkar and Mona Ghassemi. Electric power systems in more and all electric aircraft: A review. *Ieee Access*, 8:169314–169332, 2020.

[278] Bulent Sarlioglu and Casey T Morris. More electric aircraft: Review, challenges, and opportunities for commercial transport aircraft. *IEEE transactions on Transportation Electrification*, 1(1):54–64, 2015.

[279] Edward Balaban, Prasun Bansal, Paul Stoelting, Abhinav Saxena, Kai F Goebel, and Simon Curran. A diagnostic approach for electro-mechanical actuators in aerospace systems. In *2009 IEEE Aerospace conference*, pages 1–13. IEEE, 2009.

[280] Edward Balaban, Abhinav Saxena, Sriram Narasimhan, Indranil Roychoudhury, Kai F Goebel, and Michael T Koopmans. Airborne electro-mechanical actuator test stand for development of prognostic health management systems. Technical report, NATIONAL AERONAUTICS AND SPACE ADMINISTRATION MOFFETT FIELD CA AMES RESEARCH . . . , 2010.

[281] Pier Carlo Berri, Matteo Davide Lorenzo Dalla Vedova, and Paolo Maggiore. A lumped parameter high fidelity ema model for model-based prognostics. *Proceedings of the 29th ESREL, Hannover, Germany*, pages 22–26, 2019.

[282] Pier Carlo Berri, Matteo Davide Lorenzo Dalla Vedova, Paolo Maggiore, and Francesco Viglione. A simplified monitoring model for pmsm servoactuator prognostics. In *MATEC Web of Conferences*, volume 304, page 04013. EDP Sciences, 2019.

[283] Pier Carlo Berri, Matteo Davide Lorenzo Dalla Vedova, and Laura Mainini. Learning for predictions: Real-time reliability assessment of aerospace systems. *AIAA Journal*, 60(2):566–577, 2022.

[284] Brian L Stevens, Frank L Lewis, and Eric N Johnson. *Aircraft control and simulation: dynamics, controls design, and autonomous systems*. John Wiley & Sons, 2015.

[285] Pier Carlo Berri, Matteo Davide Lorenzo Dalla Vedova, and Paolo Maggiore. A smart electromechanical actuator monitor for new model-based prognostic algorithms. *International Journal of Mechanics and Control*, 17(2):19–25, 2016.

[286] Leonardo Baldo, Matteo Bertone, Matteo Davide Lorenzo Dalla Vedova, and Paolo Maggiore. High-fidelity digital-twin validation and creation of an experimental database for electromechanical actuators inclusive of failures. In *2022 6th International Conference on System Reliability and Safety*, pages 19–25, 2022.

[287] Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

[288] Matteo Davide Lorenzo Dalla Vedova, Alfio Germanà, Pier Carlo Berri, and Paolo Maggiore. Model-based fault detection and identification for prognostics of electromechanical actuators using genetic algorithms. *Aerospace*, 6(9):94, 2019.

[289] Yao Cheng, Rixin Wang, and Minqiang Xu. A combined model-based and intelligent method for small fault detection and isolation of actuators. *IEEE Transactions on Industrial Electronics*, 63(4):2403–2413, 2015.

[290] Zhengyang Yin, Niaoqing Hu, Jiageng Chen, Yi Yang, and Guoji Shen. A review of fault diagnosis, prognosis and health management for aircraft electromechanical actuators. *IET Electric Power Applications*, 16(11):1249–1272, 2022.