Technische Universität München
TUM School of Engineering and Design

High-dimensional surrogate-based optimization methods for aero-structural aircraft engine blade design

Lisa K. Pretsch

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr.-Ing. habil. Roland Wüchner

Prüfende der Dissertation:

1. Prof. Dr.-Ing. habil. Fabian Duddeck
2. Prof. Dr.-Ing. Fernaß Daoud
3. Prof. Dr. Nathalie Bartoli

Die Dissertation wurde am 09.01.2025 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 01.07.2025 angenommen.

Abstract

Numerical optimization methods show great potential to support the aircraft engine blade design process. They can reduce cost and time and yield better final results, for instance, concerning energy efficiency. A review of the state of the art reveals four open challenges: expensive design evaluations, multiple (physically coupled) disciplines and components, high-dimensional design variables, and high-dimensional coupling variables or constraints. Surrogate-based optimization (SBO) methods can tackle the first challenge. They (partly) replace the expensive high-fidelity evaluations by low-fidelity surrogate models. Bayesian optimization (BO) with Gaussian process (GP) models is particularly efficient for cases with a low number of high-fidelity samples. However, the optimization performance is impaired by the remaining three challenges. The aim of this work is to develop SBO approaches that can solve all four challenges to enhance the computational efficiency of parametric blade shape optimization.

The method developments are motivated by and tested on three industrial use cases with challenging characteristics. The first approach is an inter- instead of multidisciplinary design optimization (MDO), short IDO. It distinguishes main and side discipline, represented by high- and low-fidelity models, respectively. The second advancement extends and combines two adaptive design space reduction methods for high-dimensional constrained BO. More precisely, the hybrid approach reduces the dimensionality by principal component analysis (PCA-BO) and localizes the search inside a trust region (TR-BO). The third method exploits the multi-component structure of high-dimensional multi-stage blade optimization tasks. The novel cooperative components BO (CC-BO) decomposes the large overall problem into smaller interconnected subproblems, either according to the underlying structure or randomly.

In the context of this thesis, the proposed methods enable efficient blade optimizations with up to 223 design variables, 42 constraints, and expensive aerodynamic and structural simulations. In all three use cases, they reduce the optimization wall times for a given design improvement by a factor of four or more. The two BO-based approaches additionally achieve considerably better blade designs after a maximum affordable number of high-fidelity evaluations. Analytical test problems are employed to thoroughly study the working mechanisms of the developed algorithms. In summary, the proposed SBO methods significantly extend the efficiently manageable problem scope and dimensionality for aircraft engine blade design and beyond.

Kurzfassung

Numerische Optimierungsmethoden haben großes Potenzial, den Auslegungsprozess von Triebwerksbeschaufelung zu unterstützen. Sie können Kosten und Zeit einsparen und bessere Endergebnisse liefern, zum Beispiel hinsichtlich der Energieeffizienz. Der Stand der Technik offenbart vier aktuelle Herausforderungen: teure Design-Auswertungen, mehrere (physikalisch gekoppelte) Disziplinen und Komponenten, hochdimensionale Designvariablen und hochdimensionale Kopplungsvariablen oder Nebenbedingungen. Ersatzmodell-basierte Optimierungsmethoden (SBO) können die erste Herausforderung bewältigen. Sie ersetzen die kostspieligen high-fidelity Auswertungen (teilweise) durch low-fidelity Ersatzmodelle. Bayes'sche Optimierung (BO) mit Gauß-Prozess (GP) Modellen ist besonders für Fälle mit einer geringen Anzahl von high-fidelity Auswertungen effizient. Die Optimierungsperformance wird jedoch durch die übrigen drei Herausforderungen beeinträchtigt. Ziel dieser Arbeit ist es, SBO-Ansätze zu entwickeln, die alle vier Herausforderungen adressieren und somit die Recheneffizienz der parametrischen Formoptimierung der Schaufeln verbessern.

Die Methodenentwicklungen sind durch drei industrielle Anwendungsfälle mit herausfordernden Eigenschaften motiviert und werden an diesen getestet. Der erste Ansatz ist eine inter- (IDO) anstatt multidisziplinäre Designoptimierung (MDO). IDO unterscheidet zwischen Haupt- und Nebendisziplin, die durch high-fidelity bzw. low-fidelity Modelle repräsentiert werden. Die zweite Entwicklung erweitert und kombiniert zwei adaptive Methoden zur Verkleinerung des Designraums für hochdimensionale BO mit Nebenbedingungen. Der hybride Ansatz reduziert die Dimension durch Hauptkomponentenanalyse (PCA-BO) und lokalisiert die Suche innerhalb einer Trust Region (TR-BO). Die dritte Methode nutzt die Mehrkomponentenstruktur von hochdimensionalen mehrstufigen Schaufeloptimierungen. Die neuartige kooperative Komponenten BO (CC-BO) zerlegt das große Gesamtproblem in kleinere zusammenhängende Teilprobleme, entweder entsprechend der zugrunde liegenden Struktur oder zufällig.

Die vorgeschlagenen Methoden ermöglichen im Kontext dieser Arbeit effiziente Schaufeloptimierungen mit bis zu 223 Designvariablen, 42 Randbedingungen, sowie teuren aerodynamischen und strukturmechanischen Simulationen. In allen drei Anwendungsfällen reduzieren sie die Optimierungszeiten für eine gegebene Designverbesserung um mindestens Faktor vier. Die beiden BO-basierten Ansätze erzielen zudem deutlich bessere Schaufeldesigns nach einer maximal leistbaren Anzahl von high-fidelity Auswertungen. Anhand von analytischen Testproblemen werden die Wirkmechanismen der entwickelten Algorithmen eingehend untersucht. Zusammenfassend lässt sich sagen, dass die vorgeschlagenen SBO-Methoden den effizient handhabbaren Problemumfang und die Dimensionalität für die Auslegung von Triebwerksbeschaufelung und darüber hinaus erheblich erweitern.

Acknowledgments

First and foremost, I would like to thank Prof. Fabian Duddeck for the opportunity to work as research associate at the Professorship of Computational Solid Mechanics of TUM. I am fortunate to have gotten so much freedom, responsibility, and trust to pursue my research and support where needed. A special thanks goes to Ass.-Prof. Elena Raponi who accompanied the work for the second publication in this thesis during her postdoc time at TUM. I appreciate your mathematical perspective and good advice.

The collaboration with MTU provided much more than just an interesting application. I would like to thank Ilya Arsenyev for his strong commitment to our project. This work profited immensely from your knowledge, experience, blade design test cases, baseline code, and discussions in our weekly meetings. I would also like to thank Andreas Fischersworring-Bunk and Marc Nagel for initiating, accompanying, and supporting our research project from beginning to end. My acknowledgments go to the entire aerodynamic methods team for sharing your office once per week and making my time at MTU enjoyable.

Merci to Prof. Nathalie Bartoli for the exceptionally warm welcome at ONERA in Toulouse. I appreciate your seemingly endless energy and our fruitful discussions, which resulted in the third publication of this thesis. To the entire multidisciplinary methods and integrated concepts team at ONERA, thank you for showing me your working environment and letting me profit from your huge expertise. I am happy to have been so kindly integrated into the PhD student group and had an awesome research stay in Toulouse.

I would also like to thank my (former) colleagues at the Professorship of Computational Solid Mechanics for our nice everyday office life: Catharina Czech, Arne Kaps, Nobert Ludwig, Paolo Ascia, Koushyar Komeilizadeh, Tobias Lehrer, Philippa Weißinger, Elif Gündogan, and Tanja Schnappinger. We shared the ups and downs of a doctorate, long scientific and not-so-scientific discussions, coffee breaks with a view at Vor- and Nachhoelzer, and much more.

The final heartfelt thanks goes to my friends, family, and Lukas for their unwavering love and encouragement. I am lucky to have you!

Appended publications

I L. Pretsch, I. Arsenyev, C. Czech, and F. Duddeck. “Interdisciplinary design optimization of compressor blades combining low- and high-fidelity models”. In: *Structural and Multidisciplinary Optimization* 66.4 (2023). DOI: 10.1007/s00158-023-03516-w

II L. Pretsch, I. Arsenyev, E. Raponi, and F. Duddeck. “Twofold Adaptive Design Space Reduction for Constrained Bayesian Optimization of Transonic Compressor Blades”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2024. DOI: 10.1115/gt2024-121848

III L. Pretsch, I. Arsenyev, N. Bartoli, and F. Duddeck. “Bayesian optimization of cooperative components for multi-stage aero-structural compressor blade design”. Submitted to: *Structural and Multidisciplinary Optimization* on October 14, 2024.

Acronyms

ADP aero design point
AI actual improvement
BBOB black-box optimization benchmarking
BO Bayesian optimization
CC cooperative components
CEI constrained expected improvement
CFD computational fluid dynamics
CO collaborative optimization
CSSO concurrent subspace optimization
DLR Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Center)
DoE design of experiment
DSR design space reduction
EGO efficient global optimization
EI expected improvement
EO engine order
EV expected violation
FE finite element
GP Gaussian process
HPC high-pressure compressor
HPT high-pressure turbine
IDF individual discipline feasible
IDO interdisciplinary design optimization
LE leading edge
LHS Latin hypercube sampling
LPC low-pressure compressor
LPT low-pressure turbine
MDA multidisciplinary analysis
MDF multidisciplinary feasible
MDO multidisciplinary design optimization
MTU MTU Aero Engines AG
ODP off-design point
ONERA office national d'études et de recherches aérospatiales (The French Aerospace Lab)
PCA principal component analysis
PF probability of feasibility
PI probability of improvement
PLS partial least squares
POD proper-orthogonal decomposition
POD+I proper-orthogonal decomposition and interpolation
PS pressure side
RANS Reynolds-averaged Navier-Stokes
ROM reduced order modeling
SBO surrogate-based optimization

SS suction side
SVD singular value decomposition
TE trailing edge
TO take-off
TR trust region
TS Thompson sampling
TUM Technical University of Munich
UCB upper confidence bound

Contents

Abstract ii

Kurzfassung iii

Acknowledgments iv

Appended publications v

Acronyms vi

1 Introduction 1
- 1.1 State of the art . . . 2
- 1.2 Research aim and objectives . . . 4
- 1.3 Thesis outline . . . 5

2 Surrogate-based optimization methods 6
- 2.1 Design of experiments . . . 6
- 2.2 Gaussian process surrogate models . . . 7
  - 2.2.1 Basics . . . 7
  - 2.2.2 High-dimensional input models . . . 9
  - 2.2.3 High-dimensional output models . . . 11
- 2.3 Bayesian optimization . . . 14
  - 2.3.1 Basics . . . 14
  - 2.3.2 High-dimensional optimization . . . 17
  - 2.3.3 Constrained optimization . . . 19
- 2.4 Multidisciplinary design optimization . . . 21
  - 2.4.1 Basics . . . 21
  - 2.4.2 High-dimensional coupling . . . 22

3 Aircraft engine blade design applications 24
- 3.1 Jet engine structure . . . 24
- 3.2 Compressor blade design process . . . 25
- 3.3 Aero-structural design evaluation workflow . . . 28
- 3.4 Application cases . . . 31

4 Summary of appended publications 33
- 4.1 Publication I . . . 33
- 4.2 Publication II . . . 34
- 4.3 Publication III . . . 35

5 Discussion of contributions 37
- 5.1 Publication I . . . 37
- 5.2 Publication II . . . 39
- 5.3 Publication III . . . 40

5.4 Synthesis 42

6 Conclusion **43**

Bibliography **46**

A Publications **57**
A.1 Publication I 57
A.2 Publication II 74
A.3 Publication III 87

1 Introduction

Europe's vision for sustainable aviation [4] comprises social, environmental, and economic aspects. Aviation serves our society by connecting people and cultures. It strives for high safety, customer satisfaction, and a low negative impact on all citizens. The latter also involves a low environmental impact, with the goal of climate-neutral aviation by 2050 [4]. Economic competitiveness requires low development, production, and operation costs, including high energy efficiency.

For a reduced climate impact and energy consumption, the propulsion technology plays a major role. Modern aircraft engines ingest air, compress it, mix it with fuel, combust the mixture, and use the hot exhaust gases to drive a turbine that powers the compressor and, finally, to generate thrust. Axial compressors and turbines consist of a sequence of blade rows, which are alternately rotating and static, as illustrated in Figure 1.1. A rotor plus the associated stator form a stage. Blade design is essential for the engine's overall performance, efficiency, and stability.

For reduced development time and cost, numerical optimization methods can accelerate the design process and eventually yield better designs. They require a clear definition of an objective function, the design parameters, and potential constraints. Optimization algorithms then iteratively adjust design variables and evaluate their effect on objective(s) and constraints via numerical simulations. Thereby, they help engineers to efficiently explore complex design spaces and find superior solutions.

Turbomachinery blade design involves multiple disciplines, notably aerodynamics and structural mechanics. A typical optimization problem is to maximize the aerodynamic performance by varying blade shape parameters while not exceeding a structural stress limit. The design evaluation involves computationally expensive computational fluid dynamics (CFD) and finite element (FE) simulations. Exemplary results are depicted in Figure 1.2. Conflicting disciplinary requirements and interdisciplinary coupling make blade optimization a complex task.

This thesis is based on three first author publications [1, 2, 3], which can be found in Appendix A. All three propose and assess novel optimization approaches aimed at aircraft engine blade design applications and beyond. The underlying research was conducted at the

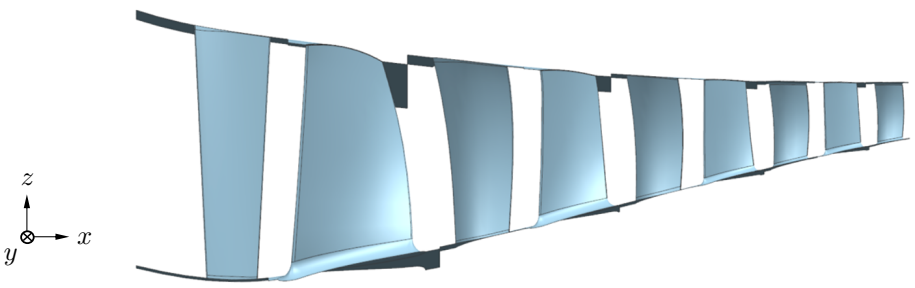


Figure 1.1 Blade rows in a high-pressure compressor frontblock. An inlet guide vane is followed by four stages of rotor and stator blades.

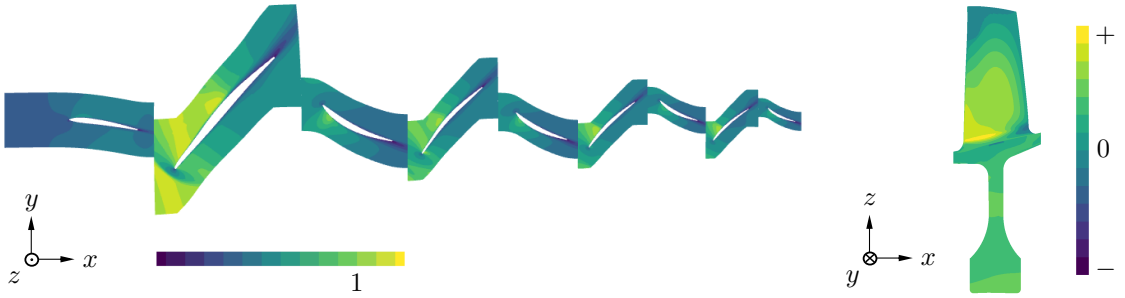


Figure 1.2 Mach number at the mid plane of the geometry in Figure 1.1 as result of a CFD simulation (left) and static principal stress on a rotor blade surface and disc as result of a FE simulation (right).

Technical University of Munich (TUM) in cooperation with the MTU Aero Engines AG and with a research stay at ONERA, the French Aerospace Lab, in Toulouse. It was funded by the *Bayerische Luftfahrtforschungs- und technologieförderung* of the *Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie (StMWi)* and the TUM. The remainder of this chapter outlines the state of the art to identify open challenges. They lead to the research aims and objectives addressed in this thesis.

1.1 State of the art

Reviews on turbomachinery [5, 6] and multidisciplinary optimization [7, 8] all name computational cost as one of the main challenges to be addressed by future research. For turbomachinery optimization in particular, the following sources of high computational cost can be identified [5]: expensive simulations, multidisciplinary and multi-row optimizations, design problems with many design variables and constraints, and optimizations under uncertainty. This section outlines the state of the art of aircraft engine blade optimization with regard to these aspects.

The evaluation of blade designs relies on costly simulations. CFD and FE analyses entail wall times of up to several hours, even in high-performance computing environments. On the one hand, accurate evaluations, including multiple disciplines and blade rows, enable better design choices. This causes the trend of increasing simulation fidelity over the entire design process. On the other hand, optimizations should require as few of these high-fidelity evaluations as possible to save computational cost. Two types of general optimization approaches exist for this purpose: gradient-based and gradient-free, with the latter often being surrogate-based.

Gradient-based approaches enable a fast convergence to a local optimum, while gradient-free ones can find the global optimum in a larger number of iterations. Hottois et al. [9] and Châtel and Verstraete [10] compared a sequential quadratic programming algorithm with adjoint gradient computation and a Gaussian process surrogate model-assisted evolutionary algorithm for aerodynamic turbomachinery blade design. Both algorithms converged to similar results, indicating that the two basic problems might be unimodal anyway. The computational cost of the adjoint method is independent of the number of design variables but grows with the number of objectives and constraints. Gradient-free methods, conversely, scale poorly with the number of design variables but well with many constraints. The two test cases [9, 10] were advantageous for the adjoint method, with 19 and 44 design parameters, respectively, but only 2 constraints each. This led to a difference of factor 2 and 3 in computational cost.

Combinations of gradient-based and -free methods can be employed to get the best of both worlds [11, 12].

For gradient-based methods, the crucial challenge is the local sensitivity computation. Finite difference methods are often prohibitively expensive. Adjoint formulations are much more efficient but intrusive, complex, and hence not always available. Aero-structural optimizations of turbomachinery components require adjoint formulations of both aerodynamic and structural analyses [13, 14, 15]. Another option is to replace the structural simulations with surrogate models [16, 17]. Aero-structural coupling effects are, if at all, imposed via fixed boundary conditions. The author is not aware of adjoint applications with variable aero-structural coupling for aircraft engine component design. Multi-row optimizations require an adjoint formulation of the rotor-stator interface [18, 19], and are often limited by a large memory requirement. For a deeper insight, Lavimi et al. [20] recently reviewed the state of the art of adjoint turbomachinery optimization.

In gradient-free approaches, the complex simulations of multidisciplinary and multi-row problems can be treated as a black box. Surrogate models of the simulations are often employed for convergence with a reduced number of high-fidelity evaluations. They have a negligible evaluation time but lower fidelity. In aero-structural turbomachinery design optimization [21, 22, 23, 24], Gaussian process (GP) models and neural networks are the most commonly used surrogates. They are usually combined with evolutionary or genetic algorithms as optimizers. In all four above references, the surrogate models are updated during the optimization with the enriched high-fidelity database. The ones with GP models [21, 22, 23] all rely on an infill criterion that balances exploration and exploitation to determine the next query point. These approaches are called Bayesian optimization (BO). The above references do not consider variable aero-structural coupling but merely include fixed boundary conditions to represent the static coupling effect or neglect it entirely. Turbomachinery blade optimizations with variable multidisciplinary coupling [25, 26] are the exception because of the prohibitively high computational effort. Multi-row and -stage optimizations usually consider aerodynamic interactions. However, surrogate-based multi-stage optimizations [27, 28, 29] are rare due to the large problem size.

Also besides multi-stage optimizations, there is a trend toward larger problems. The more design parameters and appropriate constraints are considered, the more detailed and hence better the optimization results can get in theory. In practice, these improvements are limited by the surrogate-based optimizers' capabilities. The challenge of a high-dimensional design space is its exponentially growing volume, also known as the curse of dimensionality. The most obvious and popular remedy in the scope of turbomachinery optimization is dimensionality reduction. This comprises variable screening [30], partial least squares methods [31], proper-orthogonal decomposition [32], and active subspaces [33]. An alternative are trust region approaches [34]. The above high-dimensional references deal with original design space dimensions between 30 and 57. With this in mind, the multi-stage blade design problems [27, 28, 29] with 103 to 352 free variables are extremely challenging for surrogate-based optimization (SBO) approaches. The challenge of many constraints from different disciplines is to identify the often very small portion of feasible designs. This can be achieved by penalty approaches [22], special infill criteria [35], or classification techniques [36].

Optimization under uncertainty was addressed in previous dissertations by Antinori [37], Arsenyev [38], and Ludwig [39]. All three were also performed at TUM in cooperation with MTU. Arsenyev implemented, among other things, an efficient global optimization (EGO) algorithm [35]. At MTU, the BO-like optimizer AutoOpti by the German Aerospace Center (DLR) [40] is

Table 1.1 Challenges of the application cases, which are addressed in the appended publications.

challenge	application case / publication		
	I	II	III
expensive design evaluations	✓	✓	✓
multiple disciplines / components	✓		✓
high-dimensional input / design variables		✓	✓
high-dimensional output / constraints	✓		✓

productively used. It was shown to efficiently handle aero-structural [23] and multi-stage [27] turbomachinery design problems with more than 200 design variables [40]. The two turbomachinery optimization algorithms by Arsenyev et al. [35] from TUM and MTU and Siller et al. [40] from the DLR are a special state of the art in this work, as they are used for, compared to, and enhanced by new approaches.

1.2 Research aim and objectives

As outlined in the above state of the art, (coupled) multidisciplinary and multi-stage optimizations are rarely used in an industrial context. Design tasks with many design variables and constraints still pose a challenge. The common limiting factor of these large-scale problems is their high computational effort. Accordingly, the aim of this work is to reduce the computational cost of aircraft engine blade optimization. More efficient methods enable larger-scale optimizations at an industrially acceptable computational effort and runtime. Practicable approaches for larger optimization problems, in turn, yield better blade designs.

The associated challenges for SBO methods lead to the more concrete research objectives of this work. The thesis is based on three publications with a specific application each. Academic examples and analytical benchmarks are useful for method development. Beyond that, the industrial use cases in this work ensure the relevance of the developed approaches for real-world applications. The research objectives are therefore formulated first from a general perspective, and then with regard to the individual application cases. Table 1.1 links the overall challenges to the case-specific aspects.

Overall challenges

Expensive design evaluations can provide high-fidelity results as basis for well-informed design decisions. In this work, the 3D CFD and FE simulations are (partly) replaced by low-fidelity models in SBO algorithms. The first research objective is to generate sufficiently accurate surrogate models based on scarce high-fidelity data. The second is to employ these cheap-to-evaluate models beneficially for efficient optimizations.

Multiple disciplines or components combined in a single optimization problem can yield better designs in a shorter time than sequential single-discipline or -component optimizations. For instance, aerodynamics and structural mechanics often have counteracting objectives that need to be balanced. Aerodynamic interaction effects can be exploited in multi-stage optimizations. Here, the research objective is to use the known problem structure to cope with the increased evaluation cost and problem size, including potential coupling variables.

High-dimensional input or design variables are required for detailed design, like 3D blade shape optimizations, and arise for large control volumes, for example multi-stage configurations. The associated research objective is to overcome the curse of dimensionality in SBO.

High-dimensional output models are needed for coupling variables, like gas load distributions, *or many constraints*, which reflect the complex blade design requirements. The first objective is to reduce the training (and evaluation) cost of surrogate models with high-dimensional output. The second is to efficiently handle a large number of nonlinear and severely restricting constraints in the optimization.

Case-specific aspects

As illustrated in Table 1.1, at least two of the above overall challenges are addressed in each of the three appended publications. Every publication includes a specific application along the design process chain, which serves as both motivation and test case for the developed approaches. All use cases are 3D parametric shape optimizations of turbofan engine high-pressure compressor (HPC) blades under aerodynamic and, in two out of three cases, structural mechanical aspects. All of them rely on 3D CFD (and FE) simulations and hence face the first challenge of expensive design evaluations.

Application I is a single-stage aero-structural blade optimization with focus on structural mechanics. Aero-structural coupling should be considered. Here, one first needs to choose suitable surrogate models for the scalar objective and constraints, as well as the high-dimensional coupling variables. Then, the question is how to combine low- and high-fidelity disciplinary models in the optimization. A suitable approach should both enhance the computational efficiency and maintain the organizational integrability.

Application II is a detailed single-stage aerodynamic blade optimization. Focusing on BO, the high design space dimension makes the GP models inaccurate and the infill criterion optimization inefficient. The specific research objective is to employ adaptive design space reduction approaches to improve the performance of high-dimensional constrained BO.

Application III is a detailed four-stage aero-structural blade optimization with an aerodynamic design objective. For this case, the objective is to develop a coupled problem decomposition method to enable the very high-dimensional and highly constrained BO. The approach should effectively incorporate the known problem structure but also work without this knowledge, to be generally applicable.

1.3 Thesis outline

The remainder of the thesis is structured as follows: Chapter 2 outlines the state of the art of SBO methods. The focus is on GP models and BO for high-dimensional and multidisciplinary problems, in accordance with the above stated challenges. Chapter 3 provides the context for the aircraft engine blade design applications, from the basic working principles to the specific use cases. The two chapters thereby provide the starting point for the method developments in the appended publications. Chapter 4 summarizes the publications' content. In the discussion of individual and joint contributions in Chapter 5, the proposed methods and obtained results are critically reflected with regard to the above state of the art and research objectives. Chapter 6 concludes the thesis by a summary and outlook.

2 Surrogate-based optimization methods

Expensive design evaluations are the connecting challenge of all methods developed in this work, see Table 1.1. In the realm of global optimization, surrogate-based approaches are commonly used to make the most out of scarce high-fidelity evaluation results. Figure 2.1 illustrates the adaptive surrogate-based optimization (SBO) process. To begin with, a design of experiments (DoE) generates initial sample points to be evaluated by the high-fidelity model. Afterward, the following process is iteratively repeated until a termination criterion is met: The surrogate models are fitted to the high-fidelity sample data. New promising points are searched based on the cheap-to-evaluate surrogates. Finally, the resulting infill sample points are evaluated to enrich the high-fidelity database.

A particularly sample-efficient method was proposed by Jones et al. [41, 42] and simply called efficient global optimization (EGO). It is inspired by Bayesian optimization (BO), initially introduced by Mockus [43]. Although EGO refers to a more specific class of algorithms, the two names are often used synonymously. Both profit from the advantage of Gaussian process (GP), also called Kriging [44, 45, 46], surrogate models. GP models predict the model uncertainty in addition to an output value. With this information, the optimizer can not only exploit the model but also efficiently explore the design space to iteratively enhance the surrogate model with good infill points.

Sections 2.1, 2.2, and 2.3 summarize the state of the art of DoE, GP surrogate models and BO, respectively. Section 2.4 outlines strategies for multi-disciplinary (and -component) design optimization (MDO). After the basics, current method variants for high-dimensional problems are reviewed in all sections. Therewith, the methods in this chapter address all challenges from Table 1.1, and provide a starting point for the new developments in the appended publications.

2.1 Design of experiments

DoE techniques [47] should cover the input or design space with as few expensive high-fidelity samples as possible. The quality of the DoE has a high impact on the accuracy of the surrogate model. Popular approaches are Monte Carlo sampling, that is essentially random sampling, and Latin hypercube sampling (LHS). LHS was proposed by McKay et al. [48] and is used in all appended publications.

To generate n samples via LHS, each dimension of the design space is divided into n intervals of equal probability. In case of a uniform probability distribution, this means equally sized intervals. A unique interval is chosen for each of the n samples in each dimension. Within the chosen interval, the sample is placed either centered or randomly. For an improved space filling, the LHS can be optimized with respect to various criteria, like maximin distances [49].

The input or design space volume grows exponentially with its dimension d. Hence, the number of samples required for a constant sample density increases exponentially with d [50, 51].

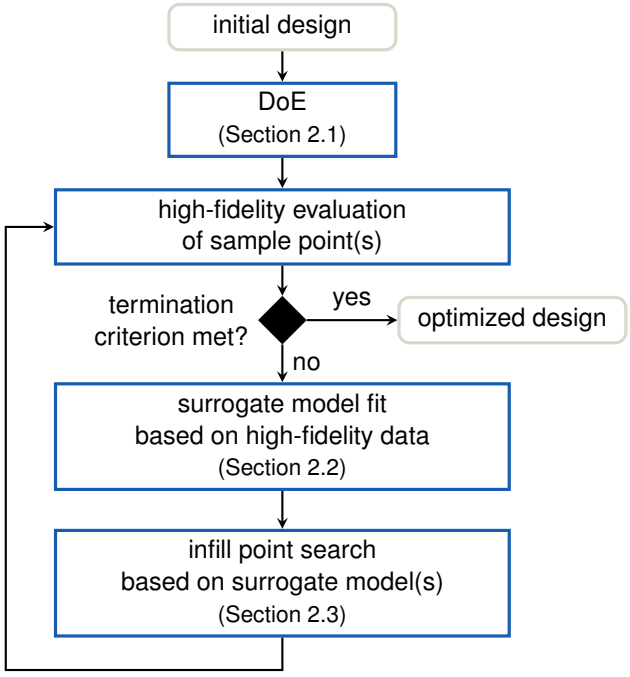


Figure 2.1 Adaptive surrogate-based optimization process.

Defining the number of input or design variables for surrogate models or optimization thus includes a trade-off between computational cost and accuracy. In engineering design, the number of variables is usually chosen based on expert knowledge.

2.2 Gaussian process surrogate models

Surrogate models, or metamodels, approximate the input-output relation $y = f(\boldsymbol{x})$ of expensive evaluations solely based on a finite sample data set $(\boldsymbol{X}, \boldsymbol{y})$. In the context of SBO, inputs are mostly the design variables, and outputs are the objective and constraints. This section focuses on GP models, which are used in all appended publications. Other popular metamodels include polynomials, radial basis functions, support vector regression, and deep neural networks. For comprehensive reviews of surrogate modeling for optimization, the reader is referred to Forrester et al. [52] and Khatouri et al. [53]. The basics of GP modeling are followed by special approaches for high-dimensional model input or output.

2.2.1 Basics

GP models reproduce the fact that there is an infinite number of functions to fit a set of training samples. They use GPs to determine the probability of each of these functions. The resulting probability distribution gives information about the most probable of these functions, as well as the confidence in the prediction. Due to their probabilistic nature, GP models are especially well-suited for scarce data. The basics of GP models for machine learning are summarized in the popular book by Rasmussen and Williams [54]. For a more intuitive understanding of the GP building blocks and the influence of settings and hyperparameters, the interactive publication by Görtler et al. [55] is recommended. The below equations are based on the

books by Keane et al. [56] and Forrester et al. [52], which can be consulted for derivations and more details.

Gaussian process

Gaussian stochastic processes are multivariate normal probability distributions over functions. They are defined by a set of random variables. Any finite number of these random variables has a joint Gaussian distribution. A GP $Z(\boldsymbol{x})$ is defined by a mean $\mu(\boldsymbol{x})$ and covariance $C(\boldsymbol{x}, \boldsymbol{x}')$.

$$\begin{aligned} Z(\boldsymbol{x}) &\sim \mathcal{N}(\mu(\boldsymbol{x}), C(\boldsymbol{x}, \boldsymbol{x}')), \\ \text{with} \quad C(\boldsymbol{x}, \boldsymbol{x}') &= \sigma_z^2 R(\boldsymbol{x}, \boldsymbol{x}'). \end{aligned} \tag{2.1}$$

The covariance is expressed as the product of the process variance σ_z^2 and a correlation function $R(\boldsymbol{x}, \boldsymbol{x}')$, also called kernel. For a d-dimensional input space, the popular squared exponential kernel (also known as radial basis function kernel) is defined as

$$R_{\mathrm{RBF}}(\boldsymbol{x}, \boldsymbol{x}') = \prod_{i=1}^{d} \exp\left(-\theta_i |x_i - x_i'|^2\right). \tag{2.2}$$

The above kernel and its generalization, the Matèrn kernel, belong to the group of stationary kernels. Stationary means that they only depend on the relative position of the two points $\boldsymbol{x}$ and $\boldsymbol{x}'$, here their Euclidean distance, and not their absolute location. The closer the points are, the stronger the corresponding function values are correlated. The lengthscale $\boldsymbol{\theta}$ is a kernel hyperparameter. Here, a vector $\boldsymbol{\theta} \in \mathbb{R}^d$ is considered, but it can also be simplified to a scalar in the isotropic case. Duvenaud [57] provides more information on the kernel choice and formulation.

Gaussian process model

For the purpose of GP surrogate modeling, a response $y = f(\boldsymbol{x})$ is considered as a random variable $Y(\cdot)$. This is contrary to its actual character as the result of deterministic computations. In ordinary Kriging, used herein, it is modeled as the sum of a constant mean β and a stationary GP $Z(\boldsymbol{x})$.

$$Y(\boldsymbol{x}) = \beta + Z(\boldsymbol{x}). \tag{2.3}$$

The mean, or trend, β is an unknown constant and thus needs to be estimated. While a constant trend is sufficient for most cases, it can be replaced by a polynomial in $\boldsymbol{x}$. However, this entails more hyperparameters to be estimated. The stationary GP $Z(\boldsymbol{x})$ has a zero mean $\mu(\boldsymbol{x}) = 0$ and a stationary covariance function $C(\boldsymbol{x}, \boldsymbol{x}')$.

Gaussian process model fit

The GP prior in Equation (2.3) can be fit to a sample data set $(\boldsymbol{X}, \boldsymbol{y})$ using Bayesian inference. The resulting GP posterior can then be used to predict the response and model uncertainty. The model fit involves estimating the mean β and the process variance σ_z^2, as well as optimizing the kernel hyperparameters $\boldsymbol{\theta}$. Maximum likelihood estimation is widely used to find the most probable values of the unknown parameters given the sample data set. For numerical stability reasons, commonly the log-likelihood function is maximized.

$$L(\boldsymbol{\theta}, \beta, \sigma_z^2) = -\frac{1}{2}\left[n \ln(2\pi) + n \ln \sigma_z^2 + \ln |\boldsymbol{R}| + \frac{1}{\sigma_z^2}(\boldsymbol{y} - \mathbf{1}\beta)^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{1}\beta)\right]. \tag{2.4}$$

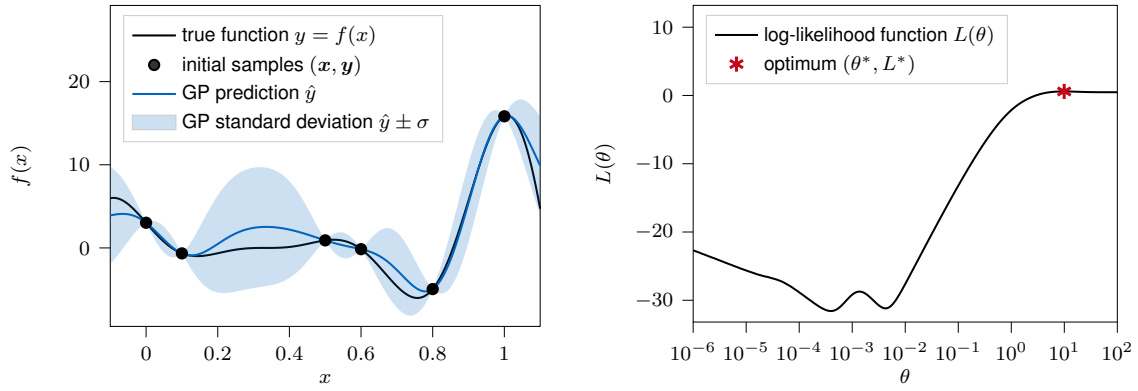


Figure 2.2 GP model (left) of the Forrester [52] function $f(x) = (6x-2)^2 \sin(12x-4)$ and corresponding maximum likelihood estimation (right).

With n samples, $\boldsymbol{R} \in \mathbb{R}^{n \times n}$ is the matrix of correlation among the samples, and $\mathbf{1} \in \mathbb{R}^n$ is a column vector of ones. The correlation matrix $\boldsymbol{R}$ depends on the lengthscale $\boldsymbol{\theta}$. Estimates of β and σ_z^2 are obtained by setting the respective derivatives to zero.

$$\begin{aligned} \hat{\beta} &= (\mathbf{1}^T \boldsymbol{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \boldsymbol{R}^{-1} \boldsymbol{y}, \\ \hat{\sigma}_z^2 &= \frac{1}{n} (\boldsymbol{y} - \mathbf{1}\hat{\beta})^T \boldsymbol{R}^{-1} (\boldsymbol{y} - \mathbf{1}\hat{\beta}). \end{aligned} \tag{2.5}$$

Inserting Equations (2.5) into Equation (2.4) and omitting the constant terms yields the concentrated log-likelihood as a function of the kernel hyperparameters $\boldsymbol{\theta}$ only.

$$L(\boldsymbol{\theta}) = -\frac{1}{2} \left[n \ln \hat{\sigma}_z^2 + \ln |\boldsymbol{R}| \right]. \tag{2.6}$$

It has no analytical solution, so numerical optimization techniques are employed to search for the best $\boldsymbol{\theta}$. The multimodal function landscape makes multi-start gradient-based methods, like BFGS, or global optimizers, like Cobyla, a common choice. The right side of Figure 2.2 shows an example of a concentrated log-likelihood function given the samples on the left.

Gaussian process model prediction

An unknown output $y = f(\boldsymbol{x})$ is assumed to have a joint Gaussian distribution with the samples $\boldsymbol{y}$. Its prediction $\hat{y}$ is the most probable realization of the posterior stochastic process $Y(\boldsymbol{x})$. The predicted response $\hat{y}$ and mean squared error, equal to the variance σ^2, can then be computed as

$$\begin{aligned} \hat{y}(\boldsymbol{x}) &= \hat{\beta} + \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}^{-1} (\boldsymbol{y} - \mathbf{1}\hat{\beta}), \\ \sigma^2(\boldsymbol{x}) &= \hat{\sigma}_z^2 \left[1 - \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}) \right]. \end{aligned} \tag{2.7}$$

Here, $\boldsymbol{r}(\boldsymbol{x}) \in \mathbb{R}^n$ is the vector of correlation between the n samples in $\boldsymbol{X}$ and the new point $\boldsymbol{x}$. Note that $\hat{y}(\boldsymbol{x}) \to \hat{\beta}$ for $\boldsymbol{r}(\boldsymbol{x}) \to 0$. This conservative prediction in case of strong extrapolation makes GP models well-suited for scarce data. The left part of Figure 2.2 shows an example of a posterior GP surrogate model.

2.2.2 High-dimensional input models

With growing input dimension d, the average distance between two points increases for a given number of samples n [50, 51]. Expensive design evaluations limit the affordable number of

samples. The resulting data sparsity makes the surrogate models less accurate. In addition, the difference between the distances of the closest and farthest point pairs decreases [50, 51]. The stationary kernels used in GP, like the one in Equation (2.2), depend on distance measures. As a result, the range of correlations decreases with growing dimension d, making them less meaningful.

Increasing the number of samples, however, involves not only a linearly growing computational effort for the expensive sample evaluation. The computational cost of inverting $\boldsymbol{R} \in \mathbb{R}^{n \times n}$ scales with $\mathcal{O}(n^3)$ [56]. It is required for every concentrated log-likelihood function evaluation in Equation (2.6) (with $\hat{\sigma}_z^2$ from Equation (2.5)) during the iterative GP model fit and for every prediction in Equation (2.7). Moreover, the number of hyperparameters $\boldsymbol{\theta} \in \mathbb{R}^d$ increases. This makes the multimodal concentrated log-likelihood maximization more difficult and, thus, computationally expensive.

Simplified models entail a lower computational effort. If data is scarce, they can even yield more accurate predictions. Therefore, approaches for GP modeling with high-dimensional input generally make some simplifying assumption(s). Binois et al. [58] reviewed approaches to overcome the above challenges. They distinguish three method categories, each relying on a distinct assumption. In the remainder of this section, the main idea of each category is outlined, followed by some examples. The ideas can be directly transferred to high-dimensional BO. Strategies that are tailored to the use within BO are summarized in Section 2.3.2.

Variable selection

Variable selection is based on the assumption that some variables do not affect the output.

$$f(\boldsymbol{x}) \approx g(\boldsymbol{x}_I) \text{ with } I \subset \{1, \dots, d\}, |I| < d. \tag{2.8}$$

Consequently, only a lower-dimensional part $\boldsymbol{x}_I$ of the overall variables $\boldsymbol{x}$ is used as surrogate model input. The influence of the remaining ones is neglected.

Correlation or sensitivity analyses [59] are often employed before the GP modeling. They can provide a decision basis but rely on scarce sample data. The review article by Becker et al. [60] focuses on this difficulty. A correlation analysis is used in publication I, as discussed in Section 5.1. The influence of the variables can also be estimated after the GP model fit. The smaller an optimized kernel lengthscale θ_i, for instance in Equation (2.2), the smaller the correlations and, hence, its relevance. This approach is called automatic relevance determination [54]. It is also implemented in AutoOpti [61], which is employed in publication I and compared in publication III.

A drawback of the above variable selection methods is that the assumption in Equation (2.8) does not hold for most well-posed modeling problems. If all input variables have a non-zero influence, entirely neglecting the less influential ones usually yields less accurate models than the approaches below.

Additive decomposition

Additive GP models assume an additive structure. In the simplest case, the output can be approximated by a sum of d functions of single variables.

$$f(\boldsymbol{x}) \approx \sum_{i=1}^{d} g_i(x_i). \tag{2.9}$$

In contrast to variable selection approaches, not the dimension is reduced, but the interaction between variables is limited.

Generalizing the above assumption, Duvenaud et al. [62] proposed a kernel that can theoretically model additive interactions up to d-th order. The additive GP posterior allows to automatically determine relevant interaction orders. In practice, scarce data usually limits accurately modeled interactions to second order. Another option is to approximate the output as sum of functions of lower-dimensional variable groups. Approaches for both disjoint [63] and overlapping [64] variable groups exist in the literature. Alternatively, the interaction structure can be determined by functional analysis of variance prior to the GP model fit. The structure is then translated into the form of the kernel [65].

The downside of the additive structure is the high number of terms with at least one hyperparameter each. This makes the GP model fit difficult. Various approaches for a more efficient hyperparameter estimation are presented along with the above additive models.

Low-dimensional embeddings

Another structural assumption is that of low intrinsic dimensionality. It means that most of the output variation can be observed in a low-dimensional subspace of the input domain. Considering a linear mapping between original and reduced input space, the output can be approximated as

$$f(\boldsymbol{x}) \approx g(\boldsymbol{A}^T\boldsymbol{x}) = g(\boldsymbol{z}) \text{ with } \boldsymbol{A} \in \mathbb{R}^{d\times r}, r < d. \tag{2.10}$$

The latent variables $\boldsymbol{z} = \boldsymbol{A}^T\boldsymbol{x} \in \mathbb{R}^r$ in the reduced subspace represent the most influential directions in the original design space. Non-linear mappings offer increased flexibility but require more sample data for fitting. The following approaches rely on linear embeddings.

The projection matrix $\boldsymbol{A}$ can be viewed as just another hyperparameter of the GP model. Tripathy et al. [66] followed this concept and use maximum likelihood estimation to first learn an orthogonal $\boldsymbol{A}$ and, second, the remaining hyperparameters of the GP model in the low-dimensional subspace. If derivatives are available as a result of the expensive design evaluations, the active subspace method [67] can be employed to determine $\boldsymbol{A}$. Bouhlel et al. [68] use Kriging with partial least squares (PLS), short KPLS. It finds the directions, called principal components, with maximum correlation between input and output variables. The resulting information is incorporated directly into the correlation function to reduce the number of hyperparameters θ_i from d to r.

Besides estimating a sufficiently accurate linear mapping, finding an appropriate reduced dimension r is not always straightforward. For the above approaches, r is automatically determined via the Bayesian information criterion [66], eigenvalue gaps [69], and Wold's R criterion [68].

2.2.3 High-dimensional output models

The above considerations aim at modeling a scalar output $y = f(\boldsymbol{x})$ as function of an input $\boldsymbol{x} \in \mathbb{R}^d$ based on a set of n samples. However, most engineering applications include more than a single output quantity of interest. This requires surrogate models of a vector-valued output $\boldsymbol{y} \in \mathbb{R}^m$. For instance, constraint values must be satisfied in addition to the objective function. Coupled MDO approaches, see Section 2.4, additionally involve coupling variables.

Vector-valued coupling variables from mesh-based simulations easily exceed $m = 10^3$ outputs and are, hence, considered high-dimensional.

Multi-task GPs [70] exploit potential correlations between multiple outputs for more accurate predictions. This blows up the correlation matrix to $\boldsymbol{R} \in \mathbb{R}^{mn \times mn}$, as well as the computational cost of its inversion to $\mathcal{O}(m^3 n^3)$. Multi-output GPs simply construct a separate model for each output. The associated computational effort only increases linearly with m ($\mathcal{O}(mn^3)$), making them better suited for high-dimensional output. Even though the multi-output model fit and prediction can theoretically be run in parallel, wall times can be prohibitive, as common computational environments cannot efficiently handle that many tasks in parallel. Another issue is the high memory usage.

Reduced order modeling

Reduced order modeling (ROM) approaches can compress a high number of m outputs to $k \ll m$ reduced bases. The underlying idea is that multiple output samples have common features. These features, or reduced bases, can be extracted by modal decomposition methods. The contained information decreases rapidly with the modes' rank. Therefore, the first few modes are sufficient to reconstruct the output without losing much accuracy. In other words, ROM implies a low intrinsic dimensionality. Its assumption is similar to that of the low-dimensional embeddings for high-dimensional input GP models outlined in the previous Section 2.2.2.

In case of high-dimensional output GP models, ROM can drastically reduce the number of outputs to be modeled. As the m-dimensional output can be reconstructed from $k \ll m$ reduced bases, only k GP models are required. Entirely data-driven ROM is called non-intrusive because it does not require access to the high-fidelity evaluation code. That makes it a very flexible tool and well-suited for black-box optimization. Proper-orthogonal decomposition (POD) is one of the most popular non-intrusive ROM approaches and is described below. Alternatives to POD are dynamic mode decomposition or autoencoders. Yu et al. [71] provide a review of non-intrusive ROM with applications in fluid mechanics.

POD is combined with GP interpolation (POD+I) to model aero-structural coupling via the gas load field in publication I. Moreover, the same principle is utilized in the context of high-dimensional BO for design space reduction (DSR) instead of ROM. Table 2.1 in Section 2.3.2 establishes the connection between the two methods. The two publications are summarized and discussed in Chapter 4 and 5.

Proper-orthogonal decomposition

POD decomposes a physical field into orthogonal modes or principal components. The non-intrusive version is called snapshot POD and was initially proposed by Sirovich [72] for fluid mechanics problems. Here, snapshots are time or spatially dependent output samples, for instance a gas load field on the blade surface. The below presented process of snapshot POD is similar to linear principal component analysis (PCA) in the field of statistics. Despite different target applications, the mathematical basics are directly transferable. The tutorial by Weiss [73] is recommended for more detailed explanations of the topic.

Prior to the POD, the output samples are arranged in form of a snapshot matrix $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$. Each row is centered around its origin by subtracting the ensemble mean vector $\bar{\boldsymbol{y}} \in \mathbb{R}^m$.

Singular value decomposition (SVD) is then used to decompose the snapshot matrix into orthogonal basis vectors and singular values.

$$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T. \tag{2.11}$$

Here, $\boldsymbol{U} \in \mathbb{R}^{m \times n}$ is a left-singular matrix with the orthogonal basis vectors or POD modes $\boldsymbol{\phi}_j, j = 1, \ldots, n$ as columns. $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is a diagonal matrix containing the non-negative singular values σ_j in descending order. $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ is a right-singular matrix and holds the same information as $\boldsymbol{U}$. Note that the POD modes are just mathematical objects expressing correlations in multi-dimensional output data. When applied to physical fields, they allow for some physical interpretation but should not be confused with the fields' actual modes.

Proper-orthogonal decomposition for reduced order modeling

The higher the singular value σ_j, the higher the influence of the corresponding POD mode $\boldsymbol{\phi}_j$ on the total variance. This sorting enables an approximation of the snapshots in $\boldsymbol{Q}$ by truncating the three matrices after the first k basis vectors and singular values.

$$\boldsymbol{Q} \approx \boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^T. \tag{2.12}$$

The truncated matrices are $\boldsymbol{U}_k \in \mathbb{R}^{m \times k}$, $\boldsymbol{\Sigma}_k \in \mathbb{R}^{k \times k}$, and $\boldsymbol{V}_k \in \mathbb{R}^{n \times k}$. POD is optimal in the sense that it loses as little information as possible for any reduced rank. The reduced rank k can be user-defined or automatically determined such that

$$\frac{\sum_{j=1}^{k} \sigma_j^2}{\sum_{j=1}^{n} \sigma_j^2} \geq \kappa. \tag{2.13}$$

The above criterion means that the variance, or proportion of energy, covered by the first k modes should be larger or equal to a user-defined threshold, often $\kappa = 95\%$. The captured variance usually drops rapidly with the rank. Hence, the number of degrees of freedom can be reduced by several orders of magnitude and $k \ll n, m$.

The lower-dimensional representation of $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$ is obtained by projecting the snapshots onto the reduced subspace spanned by the dominant POD modes via $\boldsymbol{U}_k^T\boldsymbol{Q} \in \mathbb{R}^{k \times n}$. This linear mapping can be applied to any output vector $\boldsymbol{y} \in \mathbb{R}^m$. The result is the POD coefficient vector $\boldsymbol{\alpha} \in \mathbb{R}^k$, containing k coefficients α_j.

$$\begin{aligned} \boldsymbol{\alpha} &= \boldsymbol{U}_k^T(\boldsymbol{y} - \bar{\boldsymbol{y}}), \\ \alpha_j &= \boldsymbol{\phi}_j^T(\boldsymbol{y} - \bar{\boldsymbol{y}}), j = 1, \ldots, k. \end{aligned} \tag{2.14}$$

The back projection is an approximation, as $k \ll n, m$. With $\boldsymbol{U}^{-1} = \boldsymbol{U}^T$, the high-dimensional output is computed by a linear combination of POD modes and coefficients.

$$\boldsymbol{y} \approx \bar{\boldsymbol{y}} + \boldsymbol{U}_k\boldsymbol{\alpha} = \bar{\boldsymbol{y}} + \sum_{j=1}^{k} \boldsymbol{\phi}_j\alpha_j. \tag{2.15}$$

Proper-orthogonal decomposition + interpolation

ROM via POD is, among other things, useful for high-dimensional output surrogates. Modeling only $k \ll m$ coefficients instead of m outputs saves computational cost, run time, and memory.

The approach is known as POD and interpolation (POD+I) [74]. In this work, GP models are used for the interpolation step. A corresponding PCA-GP [75] was also proposed in the statistics community.

First, the output samples are projected onto the low-dimensional subspace according to Equation (2.14). This yields the POD coefficient samples $\boldsymbol{\alpha}_j \in \mathbb{R}^n, j = 1, \ldots, k$. Second, a separate model is fit for each POD coefficient, based on the sample set $(\boldsymbol{X}, \boldsymbol{\alpha}_j)$. The GP model fit and prediction follow the basic procedure from Section 2.2.1. For an unknown input $\boldsymbol{x}$, the predicted POD coefficients $\hat{\alpha}_j(\boldsymbol{x})$ are mapped back to the original space according to Equation (2.15).

2.3 Bayesian optimization

The DoE and surrogate models described above are essential building blocks of every SBO algorithm. The optimization task is to minimize an objective function $f(\boldsymbol{x})$ by varying a design variable vector $\boldsymbol{x}$ within its lower and upper bounds.

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{by varying} && \boldsymbol{x}_{\text{lb}} \leq \boldsymbol{x} \leq \boldsymbol{x}_{\text{ub}} \in \mathbb{R}^d. \end{aligned} \tag{2.16}$$

If the surrogate model is sufficiently accurate, the problem can be efficiently solved by replacing the high-fidelity evaluation $f(\boldsymbol{x})$ with a surrogate model $\hat{f}(\boldsymbol{x})$. This approach is employed in publication I, see Section 4.1 and 5.1. However, global surrogate accuracy is often limited by sparse sample data due to expensive high-fidelity evaluations and a high-dimensional design space. A solution is adaptive SBO methods, which locally refine the surrogates in regions of interest. They were outlined in the beginning of this chapter and illustrated in Figure 2.1. For a deep-dive into SBO, the reader is referred to Forrester et al. [52] and Gramacy [76]. This section focuses on BO as an adaptive SBO technique relying on GP models. In the following, the basic method is described, and approaches for high-dimensional design spaces and constrained problems are reviewed. High-dimensional constrained BO approaches are developed in publication II and III, summarized and discussed in Chapter 4 and 5.

2.3.1 Basics

BO is a class of algorithms for global optimization of black-box functions. It can make the most out of a small high-fidelity evaluation budget. Therefore, BO is particularly suitable for expensive-to-evaluate functions, for instance costly numerical simulations. The adaptive SBO approach leverages probabilistic surrogates, mostly GP models, to find new promising points. These infill points are then evaluated and added to the high-fidelity sample dataset. The review by Shahriari et al. [77] and the book by Garnett [78] provide comprehensive summaries of BO, from basics to more specialized approaches. The review by Lam et al. [79] focuses on BO methods for aerospace design applications.

Infill criterion

GP models form the basis of BO and were already covered in the previous Section 2.2. They predict an objective function value $\hat{f}(\boldsymbol{x})$ for an unknown design variable vector $\boldsymbol{x}$. Additionally, they estimate the model uncertainty in form of the variance $\sigma(\boldsymbol{x})$, see Equation (2.7). In

every BO iteration, the question is where to place the next query point for the high-fidelity evaluation. The most straightforward idea is to exploit the GP model by minimizing $\hat{f}(\boldsymbol{x})$. However, this might mislead the search because the GP model only approximates the high-fidelity evaluation. A countermeasure is to increase the model accuracy by maximizing the model uncertainty $\sigma(\boldsymbol{x})$. The infill point search is, hence, always a trade-off between model exploitation and design space exploration.

Various infill point search strategies exist in the literature. Commonly, an infill criterion, or acquisition function, is optimized based on posterior GP model evaluations. A basic criterion is the probability of improvement (PI) [80] over the previous best design. Going one step further, the expected improvement (EI) [41] quantifies the magnitude of improvement. The upper confidence bound of the GP model (GP-UCB) [81] is a more optimistic criterion when proposed in a maximization setting (in contrast to our problem definition in Equation (2.16)). Finally, the idea of Thompson sampling (TS) [82] is to use a random realization of the GP posterior as acquisition function.

EI is employed as infill criterion throughout this work. The original idea was formulated by Mockus [43]. Later, EI was reintroduced by Jones et al. [41, 42] for the EGO algorithm.

$$\mathrm{EI}(\boldsymbol{x}) = E[I(\boldsymbol{x})] = (f_{\min} - \hat{f}(\boldsymbol{x}))\Phi\left(\frac{f_{\min} - \hat{f}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right) + \sigma(\boldsymbol{x})\phi\left(\frac{f_{\min} - \hat{f}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right). \tag{2.17}$$

The predicted improvement over the current best objective function value $f_{\min}$ for a given point $\boldsymbol{x}$ is $f_{\min} - \hat{f}(\boldsymbol{x})$. Accordingly, the first term favors exploitation. The second term includes the model uncertainty $\sigma(\boldsymbol{x})$ and, hence, promotes exploration. Φ and ϕ are the normal cumulative distribution function and probability density function, respectively (unrelated to and not to be confused with the POD modes in the previous section).

Infill criterion optimization

Figure 2.3 illustrates two BO iterations featuring the GP model and the EI maximization. The start is the initial GP model of the Forrester [52] function from Figure 2.2, replicated on the top left. The corresponding EI maximum on the bottom left is near the current best sample. This means that the first BO iteration is an exploitative step. The infill point is evaluated and added as the red sample to the enhanced GP model on the top right. The corresponding EI on the bottom right is now maximized in the region of the highest model variance. Nevertheless, it still shows a local maximum near the true objective minimum.

The acquisition functions on the bottom of Figure 2.3 are not only multimodal but also show large plateaus with zero EI. These plateaus make the infill criterion optimization extra challenging. Classical population methods like genetic algorithms and differential evolution are common choices for simple problems. Covariance matrix adaptation evolution strategies (CMA-ES) [83] are usually more computationally efficient. Direct methods, such as the dividing rectangles (DIRECT) or Nelder-Mead simplex (NMS) algorithm, were already considered at the beginning of EGO [42] and are still a valid choice.

In this work, the multi-start NMS infill criterion optimization implementation of Arsenyev [38] is used. The NMS algorithm [84, 85] optimizes the acquisition function by reflecting, expanding, contracting, and shrinking a simplex according to a series of rules. It is a local search method and thus requires multiple restarts to cope with the highly multimodal acquisition function landscape. Good starting points for an EI maximization generally lie between the existing samples. Schonlau et al. [42] proposed to generate one starting simplex per sample point.

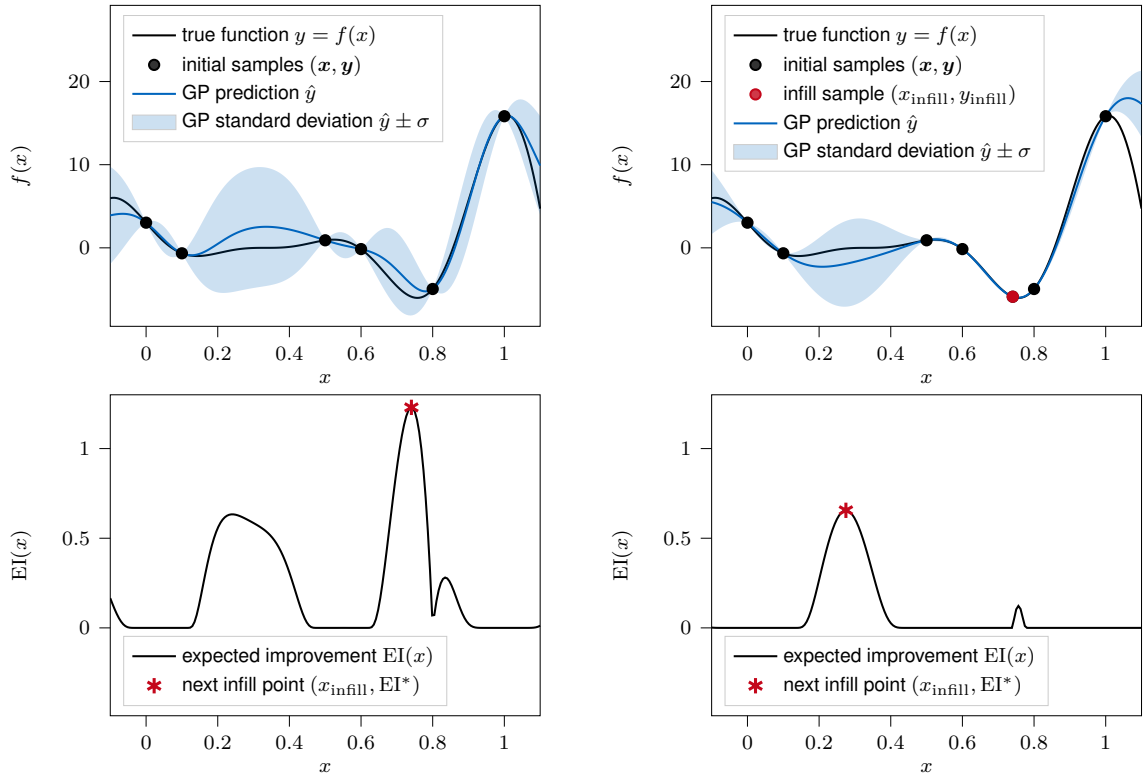


Figure 2.3 Two BO iterations with GP model (top) and infill point search (bottom) based on the respective model. The first iteration (left) yields a model exploitation near the current best sample. The second iteration (right) yields a design space exploration in the region of the highest model variance.

Each simplex has $d + 1$ vertices, which are generated by changing one coordinate of the sample point at a time. It is set to the mean value of the closest adjacent samples in the respective coordinate direction. Arsenyev [38] add some starting simplexes around random points to enhance the exploration capabilities. Moreover, he first maximizes the GP model variance to get the starting simplexes for the EI maximization out of the plateau region.

Batch infill criterion optimization

Multi-start local optimization approaches have the additional advantage that they can easily be run in parallel and yield several local optima. Therefore, they are also well-suited for batch infill point search. Evaluating $q > 1$ infill points in parallel in every BO iteration is a very effective way to save wall time, which is key in most industrial settings. Arsenyev [38] cluster the local optima and then choose the centers of the q best clusters as next infill points, to avoid very close points [86]. The clustering approach is also used in this work.

Another approach for parallel infill is the so-called q-EI [87]. It aims at finding an optimal batch of q infill points at once. An analytical expression exists only for $q = 2$. For $q > 2$, the q-EI is approximated by Monte Carlo simulations. Nevertheless, the q-EI evaluation becomes prohibitively expensive for large q. The q-EI has a qd-dimensional search space for optimization, which also quickly becomes difficult. A more efficient alternative are q sequential EI optimizations [87]. The chosen but not yet evaluated infill points are assigned a value and provisionally added to the sample data set, such that $\sigma(\boldsymbol{x}_{\text{infill}}) = 0$. The assigned value can be a constant or the GP prediction, called constant liar or Kriging believer strategy, respectively.

For arbitrarily large batches, TS is a suitable acquisition function [82]. One can simply draw q independent samples from the GP posterior in parallel.

2.3.2 High-dimensional optimization

BO exploits the advantages of GP models but also inherits their challenges for high design space dimensions d. The causes of sparse samples, inaccurate models, and expensive model fits were already explained in Section 2.2.2. Moreover, acquisition functions, like EI in Equation (2.17), are extremely difficult to optimize in high dimensions. As illustrated at the bottom of Figure 2.3 for $d = 1$, the EI function has flat regions with zero gradient and narrow local optima. With growing dimension d, the flat regions account for an increasing part of the design space [78], where optimizers easily get stuck. Besides, an increasing proportion of the design space is near the domain boundaries. The large model variance near the boundaries causes over-exploration [58]. Finally, a truly global search with a high d and small sample budget is unrealistic.

Most high-dimensional BO approaches rely on design space reduction methods. Dimension and interaction reduction strategies fall into the same three categories as for the high-dimensional GP models in Section 2.2.2. Localization is an additional option, mostly achieved via trust region (TR) approaches. Binois et al. [58] summarized problems and review solutions for high-dimensional BO. Santoni et al. [88] compared standard BO to algorithms of the above categories on the black-box optimization benchmarking (BBOB) functions. While BO is very sample efficient for $d = 10$, its performance deteriorates for $d \geq 20$. For higher d, the benefit of the approaches below depends on the validity of the respective underlying assumptions.

Variable selection

The variable selection approaches for GP models in Section 2.2.2 can also be employed prior to BO. They determine the relevant variables based on usually scarce initial sample data. Fixing the remaining variables before the optimization, one risks to prematurely neglect actually influential variables. Spagnol [89] presented a global sensitivity index tailored to high-dimensional constrained optimization instead of modeling problems, along with an adaptive procedure to select active variables in each BO iteration.

An alternative is a random variable selection in every iteration. In contrast to classical variable selection strategies, all variables are treated as equally relevant and are hence considered at some point in the optimization. The dropout method [90] follows this idea by running BO with a reduced number $r < d$ of randomly selected design variables in every iteration. For the high-fidelity function evaluation, the left-out dimensions are filled in by random values, the best so far values, or a mixture of both.

The cooperative approach CoEGO [91] relies on a similar idea. Instead of an entirely random selection in each iteration, Zhan et al. [91] randomly decompose the design space and then activate a different part in every iteration. Once all parts were considered, a new random decomposition is performed. Moreover, the inactive variables are not entirely neglected in BO but fixed to so-called context vectors. Only the active dimensions are varied in both the GP model fit and the infill criterion optimization. The context vectors are adaptively set to the best so far GP model hyperparameters $\theta_{\mathrm{best},i}$ and design points $x_{\mathrm{best},i}$.

CoEGO is an inspiration for the high-dimensional multi-component BO approach presented in publication III. The proposed cooperative components BO (CC-BO) is summarized and

discussed in Section 4.3 and 5.3. The detailed CC-BO algorithm description can be found in Appendix A.3.

Additive decomposition

Under the assumption of additivity, the objective function can be approximated by a sum of functions, see Section 2.2.2. Each term of the additive GP model then depends only on a lower-dimensional subset of the design variables. Add-GP-UCB [63] applies the same concept to formulate and optimize additive GP-UCB acquisition functions. Each term of the acquisition function then depends only on one term of the above GP models. It can be optimized separately by varying a lower-dimensional subset of the design variables. This means that the infill search space volume is only exponentially proportional to the lower subset dimension $r < d$. Consequently, the infill point search becomes much simpler and more efficient.

Initially proposed for disjoint variable groups [63], Add-GP-UCB was also generalized to overlapping subsets [64]. The overlap requires a message passing algorithm for the infill criterion optimization to incorporate dependency information. Generally unknown additive structures and dependencies limit the practical applicability of the above methods. Several approaches have been proposed to discover a suitable design space decomposition into disjoint [92] and overlapping [93] variable groups. However, learning a decomposition based on sparse sample data can be misleading. Random decompositions [94] are a good alternative, independent of the available data.

Low-dimensional embeddings

Random projections can be used to realize the assumption of low intrinsic dimensionality, see Section 2.2.2. REMBO [95] randomly samples the entries of the projection matrix $\boldsymbol{A} \in \mathbb{R}^{d \times r}$ from $\mathcal{N}(0, 1)$. As defined in Equation (2.10), $\boldsymbol{z} = \boldsymbol{A}^T \boldsymbol{x}$ maps a high-dimensional design variable onto a lower-dimensional embedding. Wang et al. [95] proved that, provided a lower intrinsic dimensionality $r < d$, for any $\boldsymbol{x} \in \mathbb{R}^d$ there is a point $\boldsymbol{z} \in \mathbb{R}^r$, such that

$$f(\boldsymbol{x}) = f((\boldsymbol{A}^T)^{-1}\boldsymbol{z}) = g(\boldsymbol{z}) \text{ with } \boldsymbol{A} \in \mathbb{R}^{d \times r}, r < d. \tag{2.18}$$

Consequently, one can optimize $g(\boldsymbol{z})$ instead of $f(\boldsymbol{x})$ without any loss of accuracy. The GP model fit and the infill point search can be performed in the lower-dimensional subspace. While the projection matrix $\boldsymbol{A}$ is random, the generally unknown intrinsic lower dimension r is still difficult to estimate. Numerous approaches for BO in lower-dimensional subspaces followed REMBO. The low-dimensional embedding is chosen randomly [96, 97, 98], inferred from sample data [99, 100], or determined as a combination of both [101].

Raponi et al. [100] use PCA to learn the linear projection and the reduced dimension from sample data. As stated in Section 2.2.3, linear PCA and snapshot POD have the same mathematical basis. Table 2.1 lists the equivalent variables to POD in Section 2.2.3. In contrast to the high-dimensional output for POD, the high-dimensional design variables are generally not correlated. PCA-BO [100] adds a weighting according to the points' merit in between the centering and SVD to enforce the correlation. The result are the centered and weighted design variable samples $\boldsymbol{X}''$. They are the input for the PCA, equivalent to the snapshot matrix for POD. In each BO iteration, $\boldsymbol{A}$ is computed by a new PCA based on all centered and weighted samples. The reduced dimension r is determined via Equation (2.13). As input to the GP model fit, the unweighted samples $\boldsymbol{X}$ are projected onto the lower-dimensional embedding

Table 2.1 Equivalent variables of PCA for design space reduction (DSR) and POD for reduced order modeling (ROM), see Section 2.2.3. They are used in publication II and I, respectively.

variable	PCA for DSR	POD for ROM
high dimension	d	m
reduced dimension	r	k
high-dimensional vector	$\boldsymbol{x} \in \mathbb{R}^d$	$\boldsymbol{y} \in \mathbb{R}^m$
reduced-dimensional vector	$\boldsymbol{z} \in \mathbb{R}^r$	$\boldsymbol{\alpha} \in \mathbb{R}^k$
snapshot matrix	$\boldsymbol{X}'' \in \mathbb{R}^{n \times d}$	$\boldsymbol{Q} \in \mathbb{R}^{n \times m}$
(back) projection matrix	$\boldsymbol{A} \in \mathbb{R}^{d \times r}$	$\boldsymbol{U}_k \in \mathbb{R}^{m \times k}$

according to Equation (2.14). A penalization of the acquisition function ensures that all infill points belong to the original space $\boldsymbol{x}_{\mathrm{lb}} \leq \boldsymbol{x} \leq \boldsymbol{x}_{\mathrm{ub}}$ after the back projection with Equation (2.15).

PCA-BO is modified to enable batch infill and constraint handling in publication II. This makes it applicable to aircraft engine blade optimization. The novel approach is summarized and discussed in Section 4.2 and 5.2. A detailed algorithm description is provided in Appendix A.2.

Trust regions

Trust regions localize the search by reducing the design variable ranges instead of dimension or interaction. The assumption is that the optimum lies close to the best so far design. Therefore, the trust region is adaptively centered around the current best point. Its commonly box-shaped bounds are contracted or expanded depending on the optimization progress (or model uncertainty). The trust region includes a drastically reduced design space volume. Moreover, the localization counteracts over-exploration.

Regis et al. [102] proposed a trust region implementation, called TRIKE, with the ratio of actual improvement to EI as criterion for contraction or expansion. While the GP model is still fit in the entire space, the EI is then maximized inside the trust region. The trust region BO (TuRBO) by Eriksson et al. [103] uses TS as acquisition function to enable efficient batch infill. The trust region is tightened or extended after a predefined number of iterations without or with improvement, respectively. GP models are fit on the samples in the TR only, to allow for heterogeneous function landscapes.

The drawback of localization can be overcome by restarts once the EI falls below a certain threshold [102]. For batch high-fidelity evaluations, multiple parallel trust regions can be managed by a bandit strategy [103]. Diouane et al. [104] alter between local iterations inside the TR and global ones in the full design space.

We utilize the localization by using a trust region BO (TR-BO) after the more global PCA-BO in publication II. TRIKE is combined with ideas from TuRBO to enable batch infill. The reader is referred to the summary and discussion of contributions in Section 4.2 and 5.2, and the detailed algorithm description in Appendix A.2.

2.3.3 Constrained optimization

The above-referenced TuRBO [103] is one of the very few high-dimensional BO algorithms with an extension to constrained problems [105]. Most engineering design tasks are sub-

ject to constraints, though. Equality constraints can be reformulated as two close inequality constraints. In the problem definition below, inequality constraints are denoted by the vector $\boldsymbol{h}(\boldsymbol{x})$.

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && \boldsymbol{h}(\boldsymbol{x}) \leq \mathbf{0} \in \mathbb{R}^{m-1} \\ &\text{by varying} && \boldsymbol{x}_{\text{lb}} \leq \boldsymbol{x} \leq \boldsymbol{x}_{\text{ub}} \in \mathbb{R}^{d}. \end{aligned} \tag{2.19}$$

For BO, constraint GP models are required in addition to the objective function $f(\boldsymbol{x})$. Approaches for modeling a vector-valued output $\boldsymbol{y} = [f, h_2, \ldots, h_m]^T \in \mathbb{R}^m$ were summarized in Section 2.2.3. New infill points are then searched by exploiting and enhancing these surrogates to find a feasible optimized design.

Infill criterion

Parr et al. [106] compared infill criteria for constrained problems (and batch infill). The unconstrained acquisition functions from Section 2.3.1 can be optimized subject to the constraints [107]. GP model predictions can replace expensive constraint evaluations. Most other approaches transform the constrained problem into an unconstrained one. A classical option is to add a penalty term to the unconstrained infill criterion if the surrogate predicts a constraint violation. The penalty can be a simple constant [108] or computed by an augmented Lagrangian formulation [109]. The above methods do not take the constraint GP model uncertainty into account. Consequently, inaccurate initial constraint surrogates can mislead the search because they are not specifically enhanced.

In contrast, the expected violation (EV) [110] and probability of feasibility (PF) [42] criteria include the predicted variance. They are formulated in analogy to EI and PI, respectively. EV quantifies the magnitude of the constraint violation. If all samples are infeasible, EV can be used as standalone criterion to find feasible regions. In this case, the PF would be mostly zero and therefore difficult to optimize. PF is mostly combined with EI to improve the current best feasible design.

$$\text{PF}_j(\boldsymbol{x}) = P[F_j(\boldsymbol{x})] = \Phi\left(\frac{-\hat{h}_j(\boldsymbol{x})}{\sigma_{h_j}(\boldsymbol{x})}\right). \tag{2.20}$$

All constraints $h_j, j = 2, \ldots, m$ are formulated such that $h_j(\boldsymbol{x}) \leq 0$. The constraint GP model predicts the mean value $\hat{h}_j(\boldsymbol{x})$ and the standard deviation $\sigma_{h_j}(\boldsymbol{x})$ for a given point $\boldsymbol{x}$. Alternatively, PF can be computed based on a probabilistic support vector machine (SVM) [111]. The classifier predicts whether a point is feasible with a single SVM for all constraints. This can be advantageous in case of many constraints. Another approach for BO with high-dimensional constraints is to reduce the number of constraints via PCA [112], similar to POD+I in Section 2.2.3. Apart from these two methods, there is very little work on BO with high-dimensional output constraints.

Finally, the constrained EI (CEI) is the product of the EI of the objective function in Equation (2.17) and the individual PFs for all constraints in Equation (2.20) [42]. CEI is used as constrained acquisition function throughout this work. It can be maximized with the (batch) infill criterion optimization methods from Section 2.3.1.

$$\text{CEI}(\boldsymbol{x}) = E[I(\boldsymbol{x}) \cap \boldsymbol{F}(\boldsymbol{x})] = E[I(\boldsymbol{x})] \cdot \prod_{j=2}^{m} P[F_j(\boldsymbol{x})]. \tag{2.21}$$

2.4 Multidisciplinary design optimization

In engineering design, constraints often come from various disciplines. The basic idea of MDO [113] is to consider multidisciplinary interactions directly in the design optimization. On the design optimization level, these interactions entail trade-offs due to conflicting disciplinary design requirements. On the design evaluation level, they involve coupled multidisciplinary analyses (MDA). For both aspects, MDO can yield better results in less time compared to sequential or iterative disciplinary optimizations. Papageorgiou et al. [7] reviewed recent advancements in aerospace MDO. Martins and Ning [114] summarized engineering design optimization approaches and also cover MDO.

They state that the terms *discipline* and *component* can be used interchangeably in the MDO context. This implies that the same methods can be used for any kind of coupled system. A classical example with conflicting disciplinary requirements is aero-structural design. Applications are shape optimizations of aircraft wings, wind turbine rotors, or the aircraft engine blades herein. Systems with multiple physical components include the multi-stage axial compressors described in the next Chapter 3.

2.4.1 Basics

Design variables $\boldsymbol{x}$ and constraints $\boldsymbol{h}(\boldsymbol{x})$ in MDO problems can be local or global. The first only apply to a single component, while the latter affect more than one component or discipline. In addition, a coupled MDA introduces coupling variables. For a static aero-structural blade analysis, the disciplines are coupled via aerodynamic pressure forces on the blade surface and structural displacements of the blade which affect each other. A coupled MDA is a nonlinear system of equations. It is usually solved by fixed-point iteration methods, such as nonlinear block Jacobi or Gauss-Seidel. The result is a set of coupling variables which satisfies the multidisciplinary equilibrium. The MDA is often itself a computationally expensive task that requires several disciplinary evaluations.

If the MDA is treated as a common evaluation, common optimization techniques, like BO, can be used. However, MDO methods can combine the tasks of satisfying the multidisciplinary equilibrium and finding a feasible optimized design in a more efficient way. Various MDO architectures were proposed for this purpose. The architectures organize the disciplinary analyses and optimization methods (and potential surrogate models). Tailored problem formulations and solution strategies exploit the underlying problem structure. Martins and Lambe [115] provide a comprehensive summary of MDO architectures. They distinguish monolithic and distributed architectures.

The various architectures served as inspiration for the *inter*disciplinary design optimization (IDO) in publication I and the high-dimensional multi-component CC-BO in publication III. Both combine ideas from monolithic and distributed approaches to their advantage. They are summarized and discussed in Chapter 4 and 5.

Monolithic optimization architectures

Monolithic architectures define a single optimization problem. Here, the problem formulations mainly differ in how much of the multidisciplinary interaction is moved from the analysis to the optimization problem. This introduces additional design variables and constraints, increasing the optimization problem size. Monolithic architectures are most frequently used in practice

because they generally show a good convergence behavior [114]. Multidisciplinary feasible (MDF) and individual discipline feasible (IDF) are two of the earliest and most popular monolithic formulations [116].

MDF treats the coupled MDA like any other high-fidelity evaluation. Hence, the optimization problem can be solved with the common methods described above. The full MDA in every iteration guarantees physical compatibility but can be computationally expensive. IDF uncouples the disciplinary analyses by letting the optimizer also solve for coupling variables that satisfy interdisciplinary feasibility. It introduces the coupling variables as additional design variables and additional consistency constraints. IDF accelerates the analysis but increases the number of design variables and constraints. Dubreuil et al. [117, 118] proposed a BO approach for monolithic MDO formulations. It is based on coupled disciplinary GP models and was shown to work with both the MDF and IDF formulation.

Distributed optimization architectures

Distributed architectures decompose the MDO problem into discipline level subproblems, which are coordinated by a system level problem. The idea is to exploit the multidisciplinary problem structure for a more efficient optimization. Distributed strategies mimic the organization of disciplinary development teams and, thus, can fit well into the industrial context. In practice, the high coordination effort on the system level often causes a slow convergence [114]. Nevertheless, they can bring an advantage for special problem structures. One distinguishes distributed MDF and IDF architectures, depending on whether they enforce multidisciplinary consistency via a coupled analysis or in the optimization. A main representative of each class is outlined in the following.

Concurrent subspace optimization (CSSO) [119] is a distributed MDF architecture. In the surrogate-based version by Sellar et al. [120], it solves the system level problem based on metamodels. The discipline subproblems use the high-fidelity analysis for the corresponding discipline and surrogates for all others. The surrogate models are iteratively updated. Collaborative optimization (CO) [121] is a distributed IDF architecture. The system level problem minimizes the objective function, while the discipline subproblems minimize system incompatibility subject to local constraints. With its fully independent disciplinary subproblems, CO is well-suited for a small number of coupling variables. The author is not aware of a BO approach for distributed MDO.

2.4.2 High-dimensional coupling

Monolithic MDF has the smallest single problem formulation. The distributed approaches partition the task into even smaller problems, each including only a subset of the design variables and constraints. On the downside, the subproblems must be coordinated on the system level. The challenges and remedies of a high number of design variables and constraints in SBO were discussed in the previous Sections 2.2 and 2.3. They can be transferred to MDO.

Independent of the architecture, MDO tasks can additionally contain interdisciplinary coupling variables. Evaluations are often performed by high-fidelity numerical methods, such as CFD or FE simulations. They rely on discretized computational meshes. This entails high-dimensional fields of coupling variables, such as pressure or displacement fields with up to several thou-

sand nodal values each. Let us consider two disciplines with vector-valued coupling variables denoted by $\boldsymbol{y}_i \in \mathbb{R}^{m_i}, i = 1, 2$.

$$\boldsymbol{y}_1 = f_1(\boldsymbol{x}, \boldsymbol{y}_2),\ \boldsymbol{y}_2 = f_2(\boldsymbol{x}, \boldsymbol{y}_1). \tag{2.22}$$

In case of expensive disciplinary solvers $f_i, i = 1, 2$, surrogate models of the coupling variables can accelerate the MDA and MDO. The high dimensionality requires some kind of ROM. POD, see Section 2.2.3, is also a popular method in the MDO context. POD+I is employed to model the unidirectional coupling via gas loads in publication I, see the summary and discussion in Section 4.1 and 5.1. Modeling a mutually coupled system entails a much higher computational effort. Two approaches are outlined in the following.

Proper-orthogonal decomposition + interpolation

Coelho et al. [122] proposed to use POD to reduce the high-dimensional coupling variables to a few POD coefficients $\boldsymbol{\alpha}_i \in \mathbb{R}^{k_i}, k_i < m_i, i = 1, 2$.

$$\boldsymbol{\alpha}_1 = g_1(\boldsymbol{x}),\ \boldsymbol{\alpha}_2 = g_2(\boldsymbol{x}). \tag{2.23}$$

This reduces the amount of data to be exchanged and resolves geometry and mesh compatibility issues at the interface. POD+I surrogate models can partly replace the expensive disciplinary analyses. The samples are evaluated by a fully coupled MDA. Hence, several disciplinary analyses are required for a single output sample, which satisfies the multidisciplinary equilibrium.

Berthelin et al. [123] proposed to further increase the computational efficiency by uncoupling the MDA, such that every disciplinary analysis can be used as a sample. The result are disciplinary POD (DPOD) coefficients of the mutually dependent coupling variables.

$$\boldsymbol{\alpha}_1 = g_1(\boldsymbol{x}, \boldsymbol{\alpha}_2),\ \boldsymbol{\alpha}_2 = g_2(\boldsymbol{x}, \boldsymbol{\alpha}_1). \tag{2.24}$$

Implementing the above equation as coupled DPOD+I models is a very complex task. First, it requires a specific sample generation strategy for well-distributed coupling variables. Second, the reduced disciplinary GP models must be locally enriched for an accurate random MDA at a given point $\boldsymbol{x}$. Both POD+I models are shown to perform well for static aero-structural problems with low numbers of 2 [122] to 8 [123] design variables.

3 Aircraft engine blade design applications

The SBO approaches developed in this work are motivated by and tested on aircraft engine blade design applications. The MTU provided three challenging industrial use cases. All of them are aero(-structural) 3D high-pressure compressor (HPC) blade design optimizations.

This chapter first conveys the basic working principle and structure of jet engines, focusing on the HPC blades. Second, it gives the overall blade design process context, including aerodynamic and structural 3D blade design requirements. Third, the modular aero-structural simulation workflow is described, used in all the blade optimizations. Finally, the three application cases are briefly outlined. Their characteristics pose the challenges treated in the appended publications, which are summarized in the next chapter.

3.1 Jet engine structure

Today's aircraft engines are mostly jet engines [125, 126]. They convert the fuel's energy into a fast air jet that propels the aircraft. Their working principle follows the thermodynamic Brayton cycle. It consists of ingestion, compression, combustion, expansion, and exhaust of the air passing through the engine. The working cycle is reflected in the basic jet engine structure, illustrated in Figure 3.1. It is composed of an air intake (here with a fan), a compressor, a combustion chamber, a turbine, and a nozzle. Compressor and turbine are connected via a shaft. Due to the high turbine entry temperature, the turbine can drive the compressor with only a small part of its expansion ratio. The remaining expansion produces the thrust.

Figure 3.1 shows a modern geared turbofan engine for civil aviation. A large fan at the intake is driven by a low-pressure turbine that extracts power from the exhaust gases. The fan accelerates an additional big mass of air to a comparatively low jet velocity in a bypass duct outside the core engine. The specific fuel consumption is proportional to the square of the jet velocity, while thrust increases only linearly with it. Therefore, turbofans require less fuel (and produce lower emissions and noise) for a given thrust. Turbojet engines, in contrast, do not have a fan and are mainly employed for military applications, where a high flight speed and large thrust have top priority. Turboprop and -shaft engines power unducted fans, propellers, or helicopter rotors.

Axial compressors and turbines, like the ones in Figure 3.1, consist of an axial sequence of blade rows that are alternately rotating and static. Both rotor blades and stator vanes have an airfoil cross-section. Rotor blades are mounted on discs that connect them to the shaft that is in turn supported by the casing via bearings. Especially compressor rotors are often manufactured as bladed disks (blisks), that is blades and disk in one piece. Blisks are smaller, lighter, and enable better aerodynamics than conventional designs with separate blades. Stator vanes are directly attached to the engine casing.

For a high compression efficiency, the rotational speed of the fan and compressor rotors should increase in axial direction. This is achieved by a split in two (or sometimes even

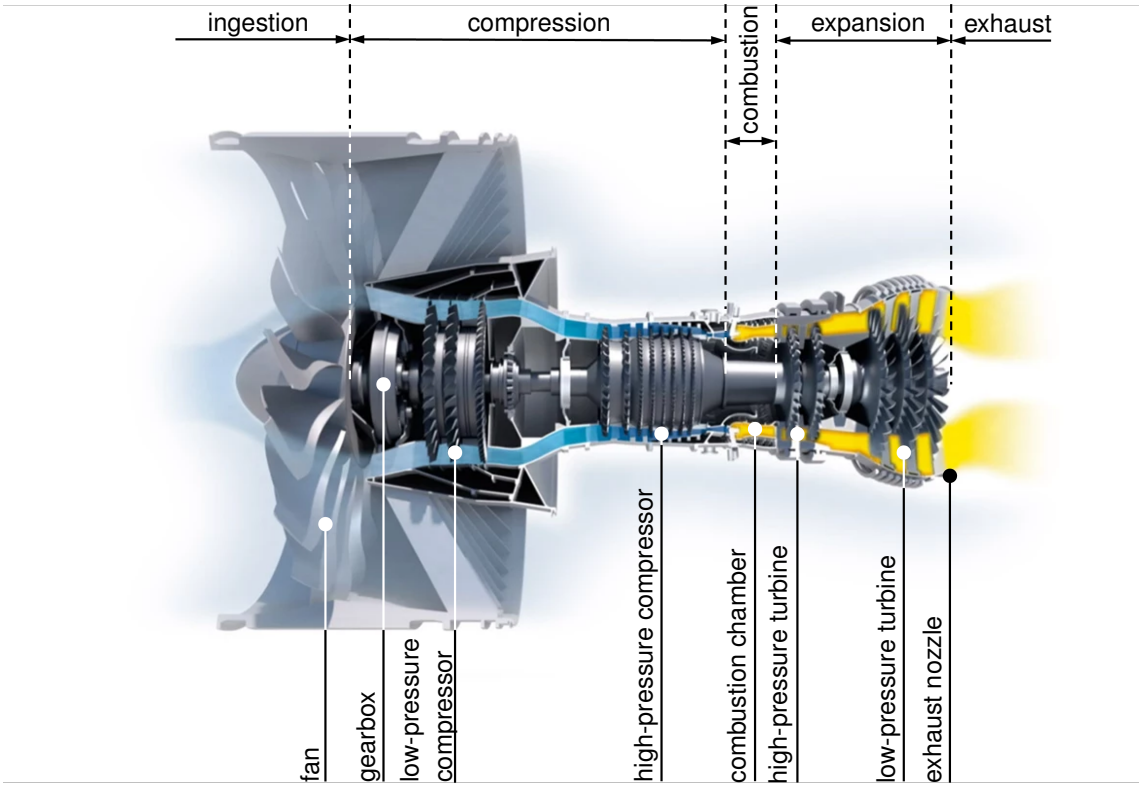


Figure 3.1 Geared turbofan engine with description of working cycle (top) and components (bottom). Modified original illustration by MTU [124].

three) shafts, one with a high-pressure turbine (HPT) and compressor (HPC), and one with a low-pressure turbine (LPT) and compressor (LPC) and a fan. In a geared turbofan engine, the fan can rotate even slower due to a gear to the low-pressure shaft.

All applications in this work focus on the HPC blades. As the air passes the rotor, it absorbs energy in form of a pressure and temperature rise. The stator then removes the swirl and generates a static pressure rise. The combination of a rotor and a downstream stator is called a stage. The stage pressure ratio is proportional to the change in tangential (or whirl) velocity that is limited to avoid flow separation. While pressure and temperature rise from stage to stage, the axial velocity is approximately maintained by a decreasing annulus area and, thus, blade height. Figure 1 in publication III in Appendix A.3 shows a four-stage HPC frontblock annulus with blades.

3.2 Compressor blade design process

Compressor blade design [127, 128] is the art of defining the best possible blade geometry. The compressor's task is to produce a required pressure ratio at a given mass flow. This must be achieved as efficiently as possible at the aero design point (ADP) and stable and safe over the entire operating range. Compressor blade design is a complex iterative process. It includes multiple disciplines with conflicting requirements, most importantly aerodynamics, structural mechanics, and aeroelastics. Nigro et al. [129] presented an example of an entire axial compressor blade design process. It consists of several iteratively adapted optimization problems based on aerodynamic simulations of various fidelities and structural analyses.

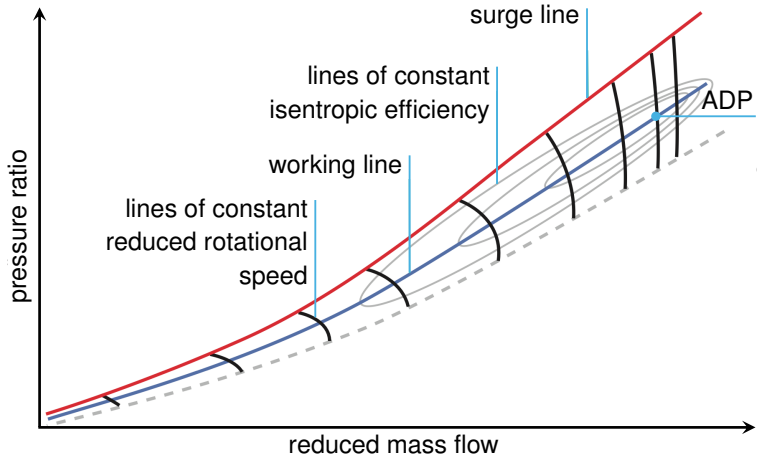


Figure 3.2 Axial compressor overall characteristic map, modified original illustration by Rolls Royce [125]. A sufficient margin between working and surge line ensures stability over a wide operating range.

Aerodynamic design

Aerodynamics is the key discipline and, thus, the starting point of the blade design process. First, the compressor concept is defined based on 1D and 2D methods. An example are throughflow computations by the streamline geometry method. This yields the flow path, the work distribution between the stages, and the number of blades. Even if it does not yet provide blade shapes, the resulting velocity triangles can be used to create preliminary blade geometries. Second, the individual 3D blade shapes are designed based on steady-state 3D CFD simulations. The preliminary blade geometry is refined for increased efficiency and a sufficient surge margin while maintaining the specifications of the previous design steps. Finally, the design of casing treatments and endwall contouring can further enhance efficiency and stability but often requires expensive, unsteady CFD simulations.

For modern transonic HPC profiles, an increased efficiency at the ADP can be attained by reducing secondary flow losses. One distinguishes between profile losses, shock losses, and endwall losses. The isentropic efficiency is the ratio of isentropic and polytropic work, that is loss-free and actual work, for an equivalent compression ratio. It is as a function of the compressor pressure ratio π, temperature ratio τ, and isentropic exponent κ.

$$\eta_{\text{is}} = \frac{\pi^{\frac{\kappa-1}{\kappa}} - 1}{\tau - 1}. \tag{3.1}$$

A sufficient surge margin ensures stability over a wide operating range. It can be inferred from the compressor characteristic map in Figure 3.2. It is the pressure ratio difference between working and surge line. Surge can occur during transient operations like rapid acceleration, when the operating point moves above the working line. It is an oscillation of the mass flow, resulting in a loss in pressure rise and, in the worst case, a loss of thrust.

Structural design

Aerodynamically good compressor blades are usually relatively thin but must also endure high mechanical loads. Static and modal structural analyses are performed by 3D FE simulations. Static forces comprise centrifugal loads, gas loads, and thermal loads. The first dominate,

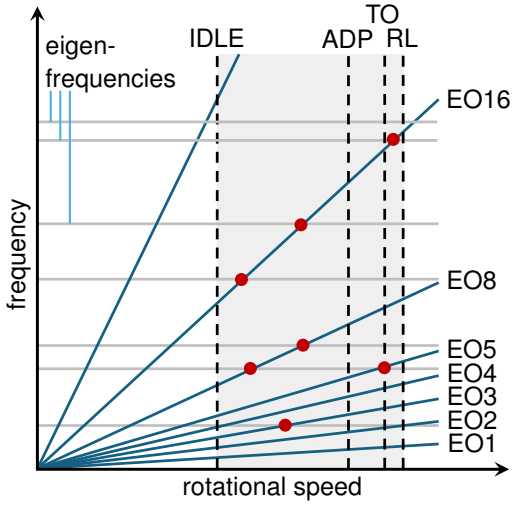


Figure 3.3 Campbell diagram of a compressor rotor. The operating range is from idle, over cruise (ADP), and take-off (TO), to red-line (RL). Engine orders (EO) and eigenfrequencies should not cross within this range to avoid resonances.

while the latter are more important at the higher turbine temperatures. With the FE result of a 3D stress state, the design goal is usually to reduce or limit the von Mises stress or the maximum principal stress. To avoid high dynamic loads, potential resonances should be avoided within the operating range, especially close to ADP speed. They can be identified in the Campbell diagram in Figure 3.3 as crossings of the engine orders (EO) and blade eigenfrequencies. Usually, some resonances remain after frequency tuning. The endurance strength margin indicates to what extent the blades can withstand the remaining dynamic loads.

Aerodynamic and structural behavior are coupled via the gas loads acting on the blade surface and the blade deformations from the so-called *cold* to the *hot* geometry. Moreover, unsteady aeroelastic phenomena such as flutter and forced response can damage the blade. Special cases like bird strike should also be considered in the design process. The blade geometry design must be coordinated with other disciplines and components, also outside the core gas flow path. The designer additionally ensures manufacturing, assembly, maintenance, repair, overhaul, and certification requirements. This might, in turn, require modifications of the blade design.

Design optimization

Numerical optimization methods can support the compressor blade design process. Given the current state of technology, the goal is not a single optimization that yields the *optimal* blade geometry for manufacturing. Instead, parts of the manual design process are automated to quickly get *optimized* blade geometries as input for the next steps in the development process. The engineer's task is to define the optimization problem and analyze the results, which is again an iterative process and requires much expertise. The optimization problem definition consists of blade design parameters and quantities of interest that reflect the above outlined design objectives and constraints. Moreover, an automated evaluation workflow is needed to compute these quantities as a function of the parameters.

The optimization applications in this work focus on the 3D blade design under aerodynamics and structural mechanics aspects. Even if the design problems in the three appended publications differ, they have common design parameters and quantities of interest. The 3D blade shapes are defined by 2D airfoil profiles and their radial stacking. The design parameterization is depicted in all three publications in the Appendix A. It is also used in the productive manual design process. The parameterization allows for large variations of the blade shape and a targeted modification of local flow phenomena. For more details, the reader is referred to the work of Nagel [130, 131].

By varying these design parameters, the isentropic efficiency at the ADP should be increased or maintained. The compatibility to the rest of the system is preserved by proper interface boundary conditions and additional constraints, such as a constant mass flow and downstream flow angle at the ADP. An accurate surge margin computation would require evaluations at many more operating points, which is prohibitively expensive. An additional operating point at the same rotational speed but closer to the surge line is evaluated to implicitly ensure a sufficient surge margin. Structural objectives and constraints include reduced or limited maximum principal stresses and critical eigenfrequencies from the Campbell diagram. Additional geometric aspects, like profile thickness and constriction, ensure manufacturability and generally reasonable designs.

3.3 Aero-structural design evaluation workflow

An automated 3D aero-structural design evaluation workflow provides the above quantities of interest. Parts of this workflow are replaced by surrogate models to enhance the optimizations in this work. For the surrogate modeling, it is helpful to know the characteristics of the input-output relationships, like non-linearity, multi-modality, or continuity. For the optimization method development, an understanding of the multidisciplinary problem structure is beneficial.

The design evaluation workflow comprises blade geometry generations and subsequent parallel CFD simulations, and static and dynamic FE analyses. The blades and vanes of different rows are generated separately. They are then all set into the annulus geometry for the aerodynamic computations or individually attached to the disc geometry for the structural analyses. The two expensive simulations are described in the following. Figure 2 in publication III in Appendix A.3 shows an example of our multi-stage aero-structural evaluation process.

Aerodynamic analysis

The CFD [132] simulations in this work are steady computations of the viscous and compressible 3D flow around rotor blades and stator vanes in the HPC annulus. Denton and Dawes [133] and Adamczyk [134] summarized the basics of CFD for turbomachinery design. Pinto et al. [135] and Sandberg and Michelassi [136] outlined the current state of the art.

The governing equations of viscous fluid flow are the Navier-Stokes equations. They are derived from the conservation equations of mass, momentum, and energy. The nonlinear second-order partial differential equations have no analytical solution but can be solved by CFD simulations. In practice, mostly the Reynolds-averaged form of the Navier-Stokes equations (RANS) is used. The temporal averaging of turbulent effects yields additional unknown source terms. They are approximated by turbulence models, here the k-ω eddy viscosity model [137].

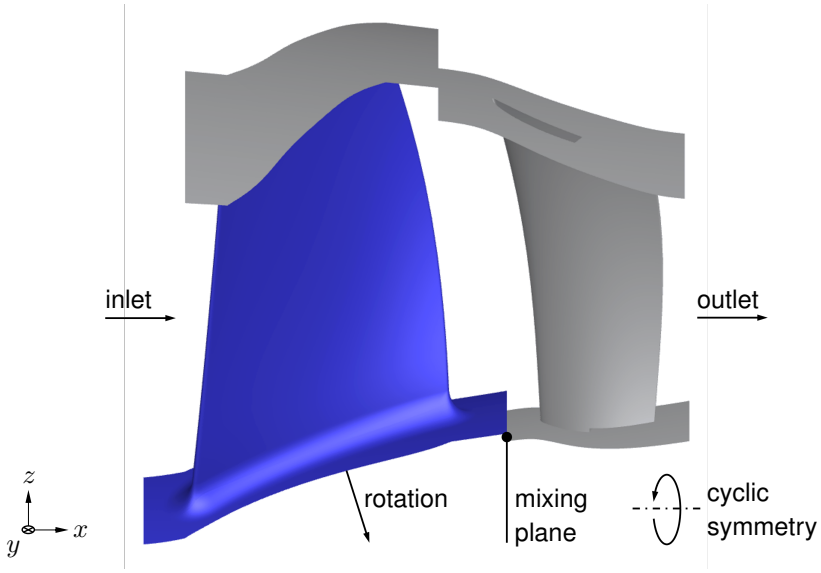


Figure 3.4 CFD geometry of an HPC stage with boundary conditions. The rotating parts are marked in blue, the stationary ones in gray.

The flow field is discretized by a numerical mesh. It consists of multiple structured mesh blocks. The very dense mesh perpendicular to the surfaces is used for a fine resolution of the boundary layer flow. Here, a good mesh quality is especially challenging to achieve. An unstructured meshing would be used for details like cavities. The aerodynamic mesh is generated by *AutoGrid* [138].

Cyclic symmetry can be used to model only one blade or vane per row for computational efficiency. However, the number of blades or vanes usually differs from row to row. This impedes the use of cyclic symmetry in multi-row simulations. One solution is the mixing plane approach [139]. Here, the flow at the rotor-stator interface is averaged in circumferential direction. Once the simulations are converged, the averaged values on both sides of the interface match.

Figure 3.4 shows an example of a single stage geometry with boundary conditions. The flow domain boundaries are very close to the blades and, hence, highly influential. Blade and annulus surfaces are modeled as non-slip walls. Inlet and outlet conditions are usually set based on 2D simulations of a bigger control volume that provide radial distributions of the flow field quantities. At the inlet, the upstream flow state is prescribed. At the outlet, the downstream potential effect is imposed via the static pressure. Both are set as non-reflecting boundary conditions [140].

The 3D RANS equations are solved by the finite volume method using the flow solver *TRACE* [141], developed by the DLR in cooperation with MTU. The result is a 3D flow field, fully characterized by a velocity vector, density, pressure, and temperature. For example, Figure 2 in publication I in Appendix A.1 shows the pressure field on the blade surface and annulus walls. Figure 7 in publication II shows the Mach number field on the mid plane, with a shock on the rotor blade suction side. The flow solution is post-processed to get the scalar quantities of interest specified in the previous section. Depending on the quantity, this is achieved by mass-, flux-, or area-averaging.

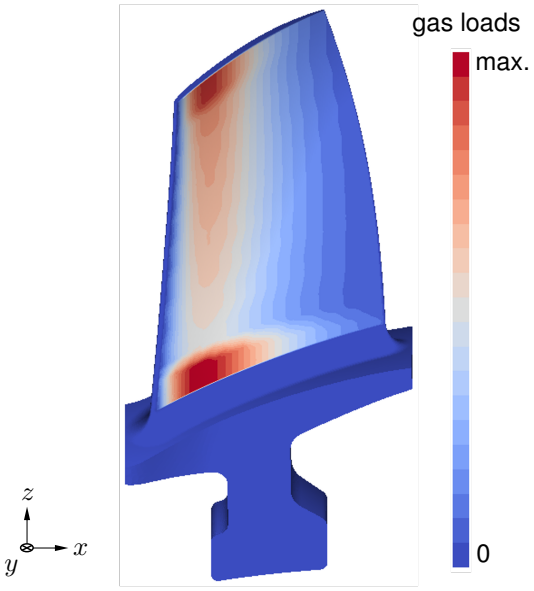


Figure 3.5 FE geometry of the HPC rotor in Figure 3.4 with gas load field on the pressure side.

Structural analysis

The 3D FE [142] simulations in the application cases comprise nonlinear static and modal analyses of the rotor blades with discs. Dhondt [143] covers the topic of FE methods for structural (and thermo-) mechanics. The book contains the theory behind *CalculiX*, an open-source FE software including pre- and post-processor. It is used for all structural analyses in this work.

The governing equations in structural mechanics are kinematic, constitutive, and balance (or conservation) equations. The weak form of the latter are the governing equations for the FE method. Blades and disc are discretized by a numerical mesh of solid elements that can be structured or unstructured. The titanium alloy of the blade is modeled as isotropic linear elastic material. The same hot blade geometry is used for both aerodynamic and structural analyses. There is no hot-to-cold transformation prior to the FE simulations.

Figure 3.5 shows an example of a rotor geometry for the structural analysis. Like for the aerodynamic analysis, modeling one rotor blade is sufficient due to cyclic symmetry. The discs are fixed in the casing by prescribing zero displacement in circumferential and axial direction. Centrifugal loads depend on the rotational speed and the changing blade geometry. Gas loads are given as radial distributions, computed by 2D simulations and thus independent of the 3D blade shape. They are then interpolated onto the blade pressure side surface and remain fixed throughout the optimization. Figure 3.5 shows the resulting distribution. The peaks at hub and tip compensate for the stresses below and above the pressure side surface, respectively. Thermal loads are also imposed as fixed distributions independent of the 3D blade shape. They only affect the material properties.

The result of the static analyses are displacement, strain, and stress distributions. An example of the maximum principal stress field on the rotor blade surface are shown in Figure 3 and 14 of publication I in Appendix A.1, with a stress concentration in the fillet. The maximum principal stress is a nodal value that can change its position and is therefore particularly difficult to predict. Modal analyses yield eigenmodes and -frequencies as basis for the Campbell diagram.

All evaluations are performed on a high-performance computing cluster. Single simulations are distributed over several CPUs, and several simulations can be run in parallel to save wall time. For the set-ups in this work [1, 2, 3], geometry generation and structural simulations take up to 3 and 6 minutes, respectively. The major part of the evaluation time comes from the aerodynamic analyses, with 18 to 119 minutes for our use cases. It must be noted that the CFD simulations in this work are set up to be as fast as possible while still being representative for the method development. In industrial practice, wall times can be much longer. The high computational cost makes aero-structural coupling inside optimizations often prohibitively expensive. Especially if CFD simulations are used, sample-efficient optimization methods are essential for industrial applicability.

3.4 Application cases

In the context of the above described jet engine structure, design, and analysis process, all three applications in this work are 3D HPC blade design optimizations based on steady CFD and static and dynamic FE analyses. The characteristics of the three application cases are outlined in Table 3.1.

Case I is a fillet stress minimization of an HPC rotor while maintaining the aerodynamic performance. This is, for instance, useful after the purely aerodynamic preliminary 1D and 2D design phase. 3D blade shape-dependent gas loads can be considered for more accurate structural simulations.

Case II is a detailed design to maximize the efficiency of an HPC stage. Structural aspects can here be implicitly taken into account by geometric constraints, like minimum blade thicknesses.

Case III is an efficiency maximization of the first four HPC stages under structural constraints. Considering inter-stage effects and allowing for changes in the stage loading can enable an even better performance. Multi-stage 3D blade design is not yet part of the above outlined design process because the problem is too large to be treated by customary optimization algorithms.

The general notion of large-scale can be specified according to its source(s). In this work, that is long simulation times, many design variables or constraints, or the inclusion of multiple disciplines or components. These characteristics are also quantified in Table 3.1. The connection to the corresponding challenges addressed in this work is established in Table 1.1.

Table 3.1 Characteristics of the 3D HPC blade design optimization application cases, which are treated in the appended publications.

characteristic	**application case / publication**		
	I	**II**	**III**
design variables	blade shape parameters HPC stage 3	blade shape parameters HPC stage 2	blade shape parameters HPC stages 1-4
objective function	min. worst principal stresses in fillet	max. isentropic efficiency	max. isentropic efficiency
constraint disciplines	aerodynamics, structural mechanics, geometry	aerodynamics, geometry	aerodynamics, structural mechanics, geometry
aero-structural coupling	variable low-fidelity gas loads	-	fixed
wall time per design evaluation	33 min	20 min	122 min
number of design variables	18	55	223
number of objectives	1	1	1
number of constraints	25	20	42

4 Summary of appended publications

Each application case serves as motivation and test for one of the three publications in Appendix A. This chapter contains a one-page summary of each publication. It is followed by a list of individual author contributions. Methods and results are further discussed in Chapter 5.

4.1 Publication I

Existing MDO architectures, outlined in Section 2.4, consider all disciplines as equally important. However, industrial optimizations are seldom truly *multi*disciplinary, as most companies are organized in disciplinary design teams. Their primary objective is disciplinary improvement, with the secondary constraint not to interfere with the other discipline(s).

In publication I, we proposed a distinction between primary and secondary, or main and side discipline, reflected in the model fidelity. Our novel *inter*disciplinary design optimization (IDO) approach facilitates the organizational integration and improves the computational efficiency. The main discipline is represented as accurately as possible by a high-fidelity simulation model. The side discipline and its coupling to the main discipline are computed as accurately as necessary by low-fidelity surrogate models. Scalar side discipline constraints are modeled by GPs and high-dimensional coupling variables by a combination of POD and GPs.

The proposed approach is tested on the application case I, summarized in Section 3.4 and Table 3.1. An HPC rotor blade should be optimized for minimum worst principal stresses in the fillet while satisfying geometric, aerodynamic, and structural constraints. The main discipline structural mechanics is computed by 3D FE simulations. The training samples for the side discipline aerodynamics are evaluated with 3D CFD. A high-fidelity aerodynamic simulation takes about 5 times more wall time than a structural one.

The surrogate models are trained on 300 samples. This is half of the number of evaluations in the optimization. The accuracy estimate is based on 10-fold cross-validation. The posterior GP models of the aerodynamic constraints achieve the predefined goals of maximum 5% root mean square error (RMSE) and minimum 90% coefficient of determination R^2. The POD of the gas load field is performed separately for the four blade sides. Altogether, the over 4000 nodal values can be reduced to 33 POD coefficients. The POD+GP models also reach the accuracy goals, with a slight deviation at the leading edge (LE) due to high gradients. The surrogate models are straightforward to hand over from the aerodynamic to the structural mechanics team. The shared data is small, and the mathematical surrogate model evaluation requires neither expert knowledge nor costly software licenses.

The multi-fidelity IDO is compared to a high-fidelity MDO with a fixed coupling, as described in Section 3.3. A variable high-fidelity coupling would be prohibitively expensive. The experiments comprise 5 runs per method with 30 optimization iterations with $q = 20$ parallel evaluations each. AutoOpti [40] is used as optimizer. All 10 runs show a similar convergence behavior and achieve more than 40% stress reduction. The design evaluations dominate the

wall time, which IDO reduces by 80%. The optimized designs obtained with IDO are validated by high-fidelity aerodynamic simulations. The results show that the constraints are largely satisfied, and the gas loads are visibly accurate, especially on the highly loaded pressure side (PS). The predicted small error on the LE is confirmed. Finally, the optimized designs of MDO and IDO are similar. This means that for IDO, both the increase in gas load accuracy and the decrease in aerodynamic model accuracy are not visible in the optimization result.

Individual contribution of authors

- Lisa Pretsch: Conceptualization, Methodology, Software, Formal analysis and investigation, Writing
- Ilya Arsenyev: Formal analysis and investigation, Writing - review and editing, Supervision
- Catharina Czech: Writing - review and editing
- Fabian Duddeck: Writing - review and editing, Supervision

4.2 Publication II

Various approaches to overcome the curse of dimensionality for BO are reviewed in Section 2.3.2. Nevertheless, only a few of them are applied to industrial engineering optimizations, let alone aircraft engine blade design. Most existing methods are not readily applicable because they cannot handle constraints or exploit simulation parallelism.

In publication II, we propose a twofold adaptive design space reduction strategy for BO, which fulfills the above requirements. We extend two existing unconstrained single-infill approaches, PCA-BO [100], see Section 2.3.2, and a trust region BO (TR-BO) [102, 103], see Section 2.3.2. Their sequential combination exploits the respective advantages of the two methods. They complement each other, as PCA-BO usually exhibits a fast initial convergence but then stagnates. TR-BO localizes the search and is, hence, better suited for a detailed search close to the optimum.

First, PCA-BO acts on a lower-dimensional embedding inferred from a PCA of weighted samples. We modified the weighting so that feasible samples have more influence on the PCA than infeasible ones. The resulting lower-dimensional parameterization is unsuitable for the GP models. Therefore, we fit them in the original design space. The batch infill would make the reduced dimension drop quickly over the number of iterations. The reduced dimension is, therefore, kept constant after the first iteration. Second, TR-BO tightens the design space bounds. The basic method follows TRIKE [102]. The trust region update criterion must be fulfilled for several consecutive iterations to work more robustly with batch infill, like in TuRBO [103]. The switch from PCA-BO with reduced dimension and wide bounds to TR-BO with full dimension and tightened bounds is with the first TR contraction.

The Rastrigin function was used to test the two base methods [100, 102] and is therefore also employed in publication II. The analytical test problem has 40 design variables and 4 constraints. The blade design application case II was characterized in Section 3.4 and Table 3.1. It is an aerodynamic blade optimization with 55 design variables and 20 constraints. The objective is to maximize the isentropic efficiency. The proposed hybrid approach and its parts are compared to the reference BO on both problems.

For the 40D analytical problem, the experiments comprise 30 runs per method with 90 initial samples and 36 optimization iterations with $q = 10$ evaluations each. PCA-BO reduces the design space dimension from $d = 40$ to $r = 24$ on average. The switch and first TR reduction is after about 8 iterations. The convergence results qualitatively match their unconstrained single-infill counterparts [100, 102]. Compared to BO, PCA-BO exhibits a steeper initial convergence, and TR-BO enables further improvements in the later iterations. As expected, experiments with GP model fit in the reduced space show a worse performance than the proposed PCA-BO and TR-BO with surrogates in the original space. The algorithm runtime is more than halved using one or both design space reduction strategies.

For the 55D blade design task, 5 runs per method were performed with 245 initial samples, followed by 96 optimization iterations with $q = 10$ parallel high-fidelity evaluations each. The dimensionality reduction for PCA-BO is comparatively small, from $d = 55$ design variables to $r = 40$ to 42 principal components. Nevertheless, the initial convergence slope of PCA-BO is about twice as steep as that of BO and TR-BO. Unlike for the analytical problem, PCA-BO exhibits continuous improvements until the final iteration. TR-BO achieves only minor additional improvements. As a result, PCA-BO and the hybrid approach reach the final feasible objective function value of BO and TR-BO in only 30 iterations instead of the full 96. They finally achieve about 25% higher isentropic efficiencies. The especially good performance of PCA-BO can be explained by its better ability to propose infill points that yield successful evaluations and feasible designs. This underlines its suitability for complex real-world applications.

Individual contribution of authors

- Lisa Pretsch: Conceptualization, Methodology, Software, Formal analysis and investigation, Writing
- Ilya Arsenyev: Formal analysis and investigation, Writing - review and editing, Supervision
- Elena Raponi: Writing - review and editing, Supervision
- Fabian Duddeck: Writing - review and editing, Supervision

4.3 Publication III

Distributed MDO architectures, summarized in Section 2.4.1, seem promising for our very high-dimensional multi-stage blade design case III. They decompose the overall optimization into smaller discipline or component subproblems. However, the coordination in an additional system subproblem hampers the convergence.

In publication III, we developed an interconnected problem decomposition approach for BO without the need for a system level coordination. Our novel cooperative components BO (CC-BO) approach is inspired by distributed MDF architectures and the cooperative EGO (CoEGO) algorithm [91]. We decompose the optimization problem according to the underlying multi-component problem structure. In every BO iteration, the GP model fit and infill criterion optimization are performed sequentially for each subproblem. The component subproblems follow the overall objective but are solved by varying only the associated component design variables. The remaining design variables are fixed to so-called context vectors. This ensures the inter-component cooperation. The evaluations are treated as multi-component system analyses and do not require a decomposition.

The lower-dimensional component design spaces simplify the subproblem optimizations. In this way, CC-BO enables a sample-efficient convergence for high-dimensional constrained problems. The same concept is also applicable without a multi-component problem structure. In contrast to the informed CC-BO described above, we proposed a more general random CC-BO. Here, the design space is decomposed randomly into a user-defined number of components in every BO iteration.

An analytical test case was developed to represent the relevant problem characteristics: a multi-component structure, high design space dimension, and constraints. The popular constrained Branin function [106] is transformed into a 100D 2-component problem. The industrial application case III was outlined in Section 3.4 and Table 3.1. It is a 4-stage aero-structural HPC blade design with 223 design variables, 42 constraints, and 2h wall time per high-fidelity evaluation. The individual stages are aerodynamically coupled. Conversely, the stages' blade geometry and structural behavior are uncoupled. For our informed CC-BO approach, this causes two additional advantages: The surrogate models of geometric and structural constraints have lower-dimensional input variables. Furthermore, the component subproblems have fewer active constraints.

For the analytical case, we compare our informed and random CC-BO to BO as baseline and PCA-BO [100, 2], SCBO [105], and AutoOpti [40] as state of the art. The experiments comprise 10 repetitions per algorithm with 101 initial samples and 40 optimization iterations with batches of $q = 10$ parallel evaluations each. CC-BO and AutoOpti largely outperform the other algorithms. Informed CC-BO with $c = 2$ components yields the fastest convergence. Additional numerical experiments on the working mechanisms reveal that the cooperative components approach is beneficial for high-dimensional GP model fit and infill criterion optimization. Moreover, the informed CC-BO is an effective way to incorporate information on the multi-component problem structure.

Based on the analytical results, informed and random CC-BO, BO, and AutoOpti [40] are run on the multi-stage blade design case. We perform 5 runs per method with 241 initial samples, followed by 100 optimization iterations with $q = 10$ parallel evaluations each. While the baseline BO performs poorly, the two CC-BO versions yield the best results. Unlike the analytical case, random CC-BO with $c = 20$ components converges even faster than the informed version with $c = 4$ physical components. Informed CC-BO exhibits a better constraint fulfillment. The results show the particular advantages of the two CC-BO versions: random CC-BO allows a flexible choice of a low subproblem dimension, while informed CC-BO exploits prior knowledge. Consequently, both informed and random CC-BO can be the best choice for the problem at hand.

Individual contribution of authors

- Lisa Pretsch: Conceptualization, Methodology, Software, Formal analysis and investigation, Writing
- Ilya Arsenyev: Formal analysis and investigation, Writing - review and editing, Supervision
- Nathalie Bartoli: Writing - review and editing, Supervision
- Fabian Duddeck: Writing - review and editing, Supervision

5 Discussion of contributions

This chapter discusses first the individual and then the joint contribution of the three publications in Appendix A. They are compared to the state of the art of aircraft engine blade optimization from Section 1.1. Moreover, the contributions are embedded in the SBO methods from Chapter 2. Finally, the research objectives formulated in the introduction Section 1.2 provide the basis for a critical reflection of the presented work. The discussion reveals limitations of the proposed methods and identifies possibilities for future work.

5.1 Publication I

Comparison to state of the art

The contribution of publication I is an IDO approach tailored to aero-structural aircraft engine blade optimization. The first novelty is the interdisciplinary perspective, which is reflected by the distinction of main and side discipline with different model fidelities. Of the aero-structural blade optimizations referenced in Section 1.1, Cuciumita et al. [16, 17] come closest to this idea. They combine high-fidelity aerodynamic simulations with surrogate models of the structural constraints in an adjoint optimization. All others keep the fully multidisciplinary view with similar model fidelity for all disciplines.

The second novelty is the variable low-fidelity coupling via POD+I for aero-structural blade design. The surrogate models reuse the sample evaluations of the constraint models and predict the gas load field in milliseconds. Most references in Section 1.1 neglect coupling due to its high computational cost. Note that the coupling effects are not always entirely ignored but imposed via precomputed values, which are independent of the 3D blade shape. This is also the case for our reference MDO, and was described in Section 3.3. The few coupled blade optimizations [25, 26] use a CO architecture, which is only suitable for a small number of coupling variables [115]. None of them employ POD+I models to approximate the high-dimensional coupling variables.

Embedding in existing methods

The GP models of the aerodynamic constraints have a moderate input dimension $d = 18$. Nevertheless, a correlation analysis is used to identify variables with little effect on the output, reducing the input dimension to $r = |I| = 12$. The remaining variables are selected as GP model input, as outlined in Section 2.2.2. The next applications II and III have a considerably higher dimension but comparable initial sampling budget. It becomes clear that the models can no longer be globally accurate. Therefore, BO (as adaptive SBO) is combined with more elaborate design space reduction techniques in the next publications.

The interdisciplinary coupling via POD+I models is assumed to be unidirectional. The assumption resolves issues of high sampling cost for a coupled MDA [122] or a very complicated procedure to uncouple the disciplinary solvers [123], discussed in Section 2.4.2. The

assumption is valid if the mutual coupling does not dominate effect of 3D blade shape modifications on the aerodynamic or structural behavior. It is, for instance, no longer true for the often large deflections and aerodynamic loads of aircraft wings. In this case, one of the two POD+I approaches in Section 2.4.2 can be used.

In our case, the difference between the variable low-fidelity coupling in IDO and the fixed coupling in MDO is invisible in the optimization results. The main reason are the much higher centrifugal loads, which dominate the structural behavior. The enhanced coupling effect is smaller than the dispersion of the optimized designs of different runs. The dispersion is due to the probabilistic nature of the SBO in combination with a limited number of iterations. Therefore, the coupling remains fixed in publication III. Nonetheless, the trade-off between disciplinary requirements is important [14, 16]. It can be effectively taken into account in an IDO or MDO.

The proposed IDO incorporates some MDO concepts but overall differs from the classical MDO architectures in Section 2.4. The IDO problem formulation is similar to a single CSSO subproblem. As such, it is closer to a monolithic MDF, with a single optimization problem and coupling via the analysis. Besides, IDO adopts some ideas of distributed MDO. It leans on industrial design team organization and aims for disciplinary independence. The simple structure with a single optimization problem and unidirectional coupling enables a straightforward extension to more than one side discipline.

Reflection of research objectives

The first case-specific objective in Section 1.2 was to choose suitable surrogate models. GP models are well-established in aero-structural blade optimization [21, 22, 23]. Comparisons against polynomial response surfaces and radial basis function models confirmed their high accuracy. The comparisons were omitted in publication I for the sake of brevity. POD+I models are rarely used in turbomachinery design [144]. To the best knowledge of the author, publication I marks the first instance of POD+I for blade aero-structural coupling. The choice is backed up by its good performance in aero-structural wing optimization [122, 123].

The second objective was to combine these models for computational efficiency and organizational integrability. IDO tackles both challenges at once. The efficiency is affected by the computational cost of main and side discipline. The 80% wall time gains of IDO in publication I are caused entirely by the 80% shorter wall time of the FE simulations compared to the CFD simulations. This advantage is lost if the main discipline is more expensive to evaluate than the side discipline. Independent of the evaluation times, the overall CPU hours are always reduced if less samples are evaluated in the DoE than in the overall optimization. In publication I, 300 sample evaluations are opposed to 600 evaluations per optimization.

The computational savings increase even further if the evaluation results or surrogate models are used in more than one optimization. It is common industrial practice to repeat optimizations with modified objectives or constraints, see Section 3.2. The evaluation results provide training data for models with the same input but possibly different output. The surrogate models from publication I can be reused provided the same input and output. This restriction no longer applies when learning the entire numerical solution instead of the input-output relationship. For example, graph neural networks are well-suited for mesh-based simulations [145] and can learn entire flow fields [146]. Various outputs can then be extracted in a postprocessing step.

5.2 Publication II

Comparison to state of the art

The contribution of publication II is a twofold adaptive design space reduction approach for BO. It enables efficient and effective high-dimensional constrained turbomachinery blade design. The state of the art in Section 1.1 includes literature references for both PCA (or POD) [32] and trust regions [34]. They were applied for blade optimizations with 30 and 48 design variables, respectively. The first novelty of publication II is employing these design space reduction methods with BO. BO is arguably one of the most popular SBO methods in engineering design and, thus, highly relevant.

Compared to Zhang et al. [32], another novelty is the adaptive use of PCA in every iteration, instead of only prior to the optimization. The original BO profits from its adaptive nature, and so does PCA-BO [100]. The third novelty is the combination of the two methods, in a way that compensates for their weaknesses and profits from their strengths.

Embedding in existing methods

The combination is not only a novelty for blade design applications but for high-dimensional BO methods in general, see Section 2.3.2. Santoni et al. [88] suggested combining PCA-BO [100] and TuRBO [103] in future work but do not present a corresponding approach. The BAxUS [147] algorithm comes closest to this idea, with a combination of random embeddings [95, 96] and trust regions [103]. For our learned embeddings in PCA-BO, a parallel instead of sequential combination with trust regions, like in BAxUS, might also be a promising approach.

The other methodological novelty is the extension of PCA-BO and TR-BO to handle constraints and enable batch infill. The extension makes the design space reduction methods applicable to industrial blade optimization in the first place. A major modification is the GP model fit in the original design space. The results for the analytical test case show a clear advantage of this choice for both PCA-BO and TR-BO. For PCA-BO, the principal components that are suitable to find a feasible objective function improvement are unsuitable to represent the individual response variations. For TR-BO, the part of the samples inside the trust region provide a worse training base than all available samples, especially if the function landscape is rather homogeneous.

Other options for high-dimensional GP modeling are summarized above in Section 2.2.2. In particular, KPLS [68] was tested for the two test cases in publication II. The results were not published to avoid confusions due to the similarities between PCA and PLS. Note that KPLS significantly accelerates the fitting but does generally not enhance the model accuracy [68]. This can only be achieved by additional information. The informed CC-BO in publication III uses prior knowledge on input-output dependencies for variable selection.

The main working mechanism of PCA-BO is dimension reduction. However, the reduced dimension $r = 40$ can still be considered high-dimensional. For such a high intrinsic dimensionality, random embedding methods like REMBO [95] are expected to perform poorly. In addition to the mere dimensionality reduction, PCA filters out bad designs. The filtering ability was already used in previous applications of PCA or POD prior to wing optimization [148, 149]. For our blade design case, bad designs are ones with a high constraint violation. PCA-BO doubles the ratio of feasible infill points compared to BO and TR-BO. These findings reveal the big potential of PCA-BO for constrained problems.

SCBO [105] is one, if not the only, high-dimensional BO algorithm with dedicated constraint handling and batch infill. The trust region approach relies on TS as infill criterion. We therefore chose to modify TRIKE [102] instead. It uses EI and is as such consistent with our original BO algorithm [38], PCA-BO [100], as well as many other BO methods for engineering design.

Reflection of research objective

The research objective for application II in Section 1.2 was to employ adaptive design space reduction approaches to improve the performance of high-dimensional constrained BO for a detailed industrial blade design. We chose PCA-BO [100] and TR-BO [102, 103] because they showed promising results in benchmarking studies [88] and complement each other well. The original methods are modified to meet the needs of our application. They fulfill the aim of a faster convergence to eventually better designs. Beyond that, PCA-BO exhibits an exceptionally high ratio of successfully evaluated and feasible points. This ability is very valuable for a robust performance in real-world problems.

With regard to the multi-stage test case III, the question is how well our design space reduction approach scales to even higher dimensions. The good performance for the single-stage application II relies heavily on PCA-BO. It learns the lower-dimensional parameterization based on sparse sample data. This becomes increasingly inaccurate, if the evaluation budget does not grow proportionally with the dimension. Moreover, the possibly inaccurate initial PCA is highly influential. All subsequent PCAs are based on the initial plus all infill sample points. The latter lie on the previous principal components and thereby reinforce their influence. Despite these limitations, PCA-BO outperforms the reference BO for the 100D analytical problem in publication III.

5.3 Publication III

Comparison to state of the art

The contribution of publication III is CC-BO, a coupled problem decomposition strategy for high-dimensional constrained BO. It overcomes the limitations of the design space reduction methods used in publication II. Therewith, CC-BO clearly outperforms them and other state-of-the-art BO algorithms on a 100D and a 223D test case.

The development is motivated by the multi-stage aero-structural blade design application III. Even if a detailed multi-stage optimization is not a complete novelty, it is still rare because of the large problem size. References of previous multi-stage SBO approaches [27, 28, 29] were given in the state of the art in Section 1.1. The main novelty is that we exploit the underlying multi-component structure to solve this problem with informed CC-BO. Random CC-BO relies on the same concept, but with the novelty of exploiting the partial separability for a sample-efficient convergence.

Embedding in existing methods

CoEGO [91], summarized in Section 2.3.2, provides the methodological basis to implement our idea. The original algorithm aims for a lower algorithm runtime. In contrast, we strive for convergence with a lower number of iterations, that is with a lower number of expensive

high-fidelity evaluations. The required modifications include sequential solutions of all subproblems in every iteration, and, thus, a novel context vector update. Moreover, we adapt the algorithm to handle constraints and find batches of infill points, like in publication II. The physical component decomposition in informed CC-BO is another difference to CoEGO.

Distributed MDO approaches, outlined in Section 2.4.1, provide the conceptual basis for the proposed informed CC-BO algorithm. On the evaluation level, the multi-component coupling is resolved in the MDA. CC-BO can, hence, be classified as distributed MDF architecture. On the optimization level, the multi-component coupling is represented by the context vectors. The sequential component subproblem optimizations make a system subproblem unnecessary. This is the main difference to and advantage over common distributed MDO strategies. Thereby, the problem decomposition enhances the convergence for high-dimensional problems.

The open question is in how far informed CC-BO can be used for other high-dimensional MDO problems. First, our multi-component problem has only local design variables, which affect a single component. Additional annulus parameters would be global design variables. They could be grouped in one or more additional system components, which are treated just like the other components. Nevertheless, like most distributed architectures, CC-BO makes most sense with mainly local variables. Second, we have both global aerodynamic objective and constraints and local structural and geometric constraints. Third, the coupled multi-stage CFD simulations are treated as a black box. Coupling variables do hence not appear in the optimization. This is typical for MDF architectures. Besides the coupled components, our multi-stage problem features the uncoupled disciplines aerodynamics and structural mechanics. Coupled disciplines would add another layer of complexity, which could be handled by EGMDO [117, 118] or simplified by IDO from publication I.

In contrast to the informed version, random CC-BO is applicable to any high-dimensional constrained optimization task. It does not require a particular problem structure. The high-dimensional GP modeling and BO methods reviewed in Section 2.2.2 and 2.3.2 also exploit structural assumptions, like additivity or low intrinsic dimensionality. For instance, the additive structure is either based on prior knowledge [63] or sample data [92], or chosen randomly [94]. Informed CC-BO exploits prior knowledge. For general applicability, we initially tried to infer this knowledge from sample data via a sensitivity analysis with sparse PCE [150]. However, the sparse samples allowed only for very inaccurate results. Finally, the random decomposition was shown to be successful.

Reflection of research objective

The case-specific research objective in Section 1.2 is to employ interconnected problem decomposition for efficient very high-dimensional and constrained BO. Informed CC-BO is tailored to the multi-stage application III. Here, the decomposition into stage subproblems yields an additional advantage for the constrained problem. The local geometric and structural constraints depend only on single stage design variables. In the GP modeling, this allows for an informed variable selection and, hence, more accurate models. In the infill criterion optimization, only a subset of all constraints are active. This simplifies the constraint satisfaction, which is an important factor for highly constrained problems.

The generalization to random CC-BO is not only generally applicable. It additionally allows for a flexible choice of the number of components c, and, thereby, an arbitrarily low subproblem dimension d_i. In publication III, we recommended $d_i \approx 10$. This is still well manageable by BO

and not too small. The blade design application showed that the lower component dimension can even outweigh the advantage of additional information. The good performance with the chosen settings provides the starting point for further studies on the effect of the number of components.

5.4 Synthesis

The three appended publications have similar applications and associated challenges. Moreover, they share the base methods GP and BO. The proposed solutions rely on the same principles of dimension reduction for POD+I and PCA-BO, as well as MDO strategies for IDO and CC-BO. Yet the appended publications each present a new independent approach for a specific application case. The proposed methods were not combined because good scientific publications present a clearly defined contribution to the state of the art. Combining many aspects makes the method description confusing and the individual performances difficult to assess. Nevertheless, we took care that the three approaches can be combined without any problems.

The performance gains in all three publications are caused by multiple factors. We identified these factors in a thorough result analysis. The success of IDO depends on the surrogate accuracy, which we estimated prior to the optimization and validated afterward. In addition to our hybrid PCA-BO and TR-BO algorithm, we also tested its parts to see their individual contribution and looked at the design space reduction progress. We ran random CC-BO not only with a supposedly beneficial number of components but also with the same number as informed CC-BO for a meaningful comparison. The CC-BO working mechanisms were broken down into the separate and combined effects of its distinctive building blocks.

The final question is whether a combination of the proposed approaches could be advantageous. IDO can be used with the proposed BO algorithms, just like it was with AutoOpti [40]. It would reduce the computational effort but not yield better optimized designs for a given number of iterations. The component subproblems in CC-BO can be solved with PCA-BO and TR-BO. PCA makes most sense when applied to each subproblem separately. The subsequent TR application is straightforward. Preliminary results for the multi-stage test case III with the combination of CC-BO, PCA-BO, and TR-BO showed slight improvements over CC-BO alone. A well-founded statement requires further tests.

6 Conclusion

The state of the art in Chapter 1 reveals four open challenges of aircraft engine blade optimization: expensive design evaluations, multiple disciplines and components, high-dimensional input or design variables, and high-dimensional output or constraints. They impede the search for the best blade designs at an industrially acceptable time and cost. The aim of this work was to reduce the computational effort of high-dimensional aero-structural blade optimizations. It contributes to the overarching goal of the fast development of better aircraft engines.

The design evaluations are computationally expensive and do not all provide gradients. This makes SBO methods, summarized in Chapter 2, an evident choice. GP surrogate models used within BO are especially sample-efficient. GP models and BO are frequently used for various engineering design tasks. Their advancement for multidisciplinary and high-dimensional applications is, thus, highly relevant.

Furthermore, the method developments in this work are motivated by specific industrial use cases. Their demanding characteristics, described in Chapter 3 and Table 3.1, reveal the research gaps. The proposed methods were integrated into a productive industrial optimization framework, where they enable significant improvements in real applications. The presented solutions are, hence, not only academically interesting but also industrially relevant.

Proposed methods

Each of the three appended publications presented a novel approach for one of the application cases and beyond. The publications were summarized in Chapter 4 and are attached in Appendix A. As a conclusion, the individual methods are reconnected to the overall challenges formulated in the introduction Section 1.2. Table 6.1 provides an overview of the solutions as an answer to Table 1.1.

Expensive design evaluations dominate the optimization wall time in blade design applications. The computational efficiency can, hence, be increased by reducing the evaluation time, as in publication I. Another option is to reduce the number of high-fidelity evaluations required for a good optimized design, as in publications II and III. Throughout this work, the expensive evaluations are partly replaced by GP surrogate model predictions. Given an adequate sample density, the models can be sufficiently accurate for blade design applications. This was confirmed by the high-fidelity validation in publication I. In publications II and III, the initially sparse samples, and thus inaccurate models, are adaptively enhanced by BO.

Multiple disciplines are weakly coupled in our novel IDO approach. With only global design variables that affect both disciplines, IDO is a monolithic strategy. It mimics the interdisciplinary perspective of disciplinary design teams. The distinction between high-fidelity main discipline and low-fidelity side discipline considerably reduces the overall design evaluation effort. *Multiple components* are fully coupled in our proposed CC-BO approach. The purely local design variables make the distributed architecture an appropriate choice. We decompose

Table 6.1 Solutions to challenges of the application cases, which are proposed in the appended publications.

challenge	**solution for application case / in publication**		
	I	**II**	**III**
expensive design evaluations	partial replacement by GP models	BO	BO
multiple disciplines / components	interdisciplinary optimization (IDO)	-	cooperative components optimization (CC-BO)
high-dimensional input / design variables	-	dimensionality reduction (PCA-BO) + trust region (TR-BO)	problem decomposition
high-dimensional output / constraints	dimensionality reduction (POD+I)	-	problem decomposition

the optimization problem into its physical components, the stages. The smaller interconnected subproblems enable a fast convergence for high-dimensional problems.

High-dimensional input or design variables were shown to slow down the convergence of BO for dimensions $d = 40$ and 55 in publication II and prevent notable improvements for $d = 100$ and 223 in publication III. In case II, the adaptive design space reduction approaches considerably speed up the convergence. The reduction in our modified and combined PCA-BO and TR-BO is based on sample data and the optimization progress, respectively. In case III, the novel problem decomposition method enables significant improvements in the first place. The decomposition in our CC-BO is based on prior information or generated randomly.

High-dimensional output models entail a high training effort. High-dimensional coupling variables were reduced by POD and then interpolated by GP models (POD+I) in publication I. *Many constraints* hinder the application of most existing BO approaches. However, they severely restrict the feasible design space for the blade design applications. The superior constraint handling of both our modified PCA-BO and novel informed CC-BO contributes to their good performance. PCA-BO filters out infeasible designs, while informed CC-BO reduces the number of active constraints per subproblem.

Innovative contributions

The originality of the proposed methods was discussed in Chapter 5. The innovative contributions of the individual publications are summarized below.

Publication I presents the inter- instead of multidisciplinary optimization approach IDO. Its novelty is the distinction between main and side discipline, which is reflected by different model fidelities in the optimization. The low-fidelity coupling is the first application of POD+I to approximate high-dimensional coupling variables in aircraft engine blade design.

Publication II features a novel sequential combination of PCA-BO and TR-BO to overcome the course of dimensionality. The existing methods are extended to handle constraints and batch infill, both of which are rare in high-dimensional BO. This enables the application to

industrial turbomachinery blade design, where it marks the first instance of adaptive design space reduction for BO.

Publication III introduces an innovative problem decomposition approach for high-dimensional BO. Both informed and random CC-BO are the first methods of their kind to exploit partial separability for a sample-efficient convergence. The informed version was initially developed for multi-stage aircraft engine blade design. In contrast to previous multi-stage approaches, it effectively exploits the underlying multi-component structure.

Perspectives

The discussion in Chapter 5 revealed the limitations of the proposed individual approaches. Moreover, it indicated interesting directions for further studies and developments. This thesis concludes with perspectives for future work, which contribute to the overall aim of efficient high-dimensional aero-structural blade optimization.

This work was focused on reducing the computational effort or number of high-fidelity evaluations because they dominate the wall time. However, the time spent on modeling and optimization also grows with the problem size. High-dimensional GP modeling could be accelerated by KPLS [68], see Section 2.2.2. High-dimensional BO with TS [82] replaces the expensive infill criterion optimization by a simple sampling, compare Section 2.3.1. TS and all methods in this work are well-suited for batch evaluations, which reduce the wall time per high-fidelity simulation. Just like the evaluations, the optimization algorithm could be parallelized, as started in publication III.

In all appended publications, GP surrogate models were employed to approximate the objective, constraints, and coupling variables as a function of the design variables. Models of the resulting flow or displacement field [145, 146] as a function of the design variables or even mesh and boundary conditions would be more versatile. Such low-fidelity models could be reused for various problems, for example within our IDO. Besides, surrogate models could predict and, hence, help to avoid simulation failures. Especially if the failures occur at the physical stability limits [151], classifiers can provide a benefit for BO [152]. Furthermore, surrogate models could enhance or accelerate the high-fidelity evaluations. For instance, one could approximate the surge margin [153]. The important but expensive-to-evaluate stability indicator is only considered implicitly in our applications, as explained in Section 3.2.

Future use cases include full aero-structural coupling and even larger control volumes. The coupling can either be considered by a high-fidelity MDA, via ROMs [123], or within EGMDO [117, 118]. The two latter options could be combined with our PCA-BO, TR-BO, or CC-BO to maintain their good performance for high-dimensional design spaces with limited samples. A larger control volume could include the entire, in our case 8-stage, HPC. Moreover, one could add annulus design parameters for greater design flexibility. The resulting increase in design space dimension might already be manageable due to the proposed methods.

Bibliography

[1] L. Pretsch, I. Arsenyev, C. Czech, and F. Duddeck. “Interdisciplinary design optimization of compressor blades combining low- and high-fidelity models”. In: *Structural and Multidisciplinary Optimization* 66.4 (2023). DOI: 10.1007/s00158-023-03516-w.

[2] L. Pretsch, I. Arsenyev, E. Raponi, and F. Duddeck. “Twofold Adaptive Design Space Reduction for Constrained Bayesian Optimization of Transonic Compressor Blades”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2024. DOI: 10.1115/gt2024-121848.

[3] L. Pretsch, I. Arsenyev, N. Bartoli, and F. Duddeck. “Bayesian optimization of cooperative components for multi-stage aero-structural compressor blade design”. Submitted to: *Structural and Multidisciplinary Optimization* on October 14, 2024.

[4] European Commission and Directorate-General for Research and Innovation. *Fly the Green Deal : Europe’s vision for sustainable aviation*. Publications Office of the European Union, 2022. DOI: doi/10.2777/732726.

[5] Z. Li and X. Zheng. “Review of design optimization methods for turbomachinery aerodynamics”. In: *Progress in Aerospace Sciences* 93 (2017), pp. 1–23. DOI: 10.1016/j.paerosci.2017.05.003.

[6] J. Hammond, N. Pepper, F. Montomoli, and V. Michelassi. “Machine Learning Methods in CFD for Turbomachinery: A Review”. In: *International Journal of Turbomachinery, Propulsion and Power* 7.2 (2022), p. 16. DOI: 10.3390/ijtpp7020016.

[7] A. Papageorgiou, M. Tarkian, K. Amadori, and J. Ölvander. “Multidisciplinary Design Optimization of Aerial Vehicles: A Review of Recent Advancements”. In: *International Journal of Aerospace Engineering* (2018), pp. 1–21. DOI: 10.1155/2018/4258020.

[8] P. Ramu, P. Thananjayan, E. Acar, G. Bayrak, J. W. Park, and I. Lee. “A survey of machine learning techniques in structural and multidisciplinary optimization”. In: *Structural and Multidisciplinary Optimization* 65.9 (2022). DOI: 10.1007/s00158-022-03369-9.

[9] R. Hottois, A. Châtel, G. Coussement, T. Debruyn, and T. Verstraete. “Comparing Gradient-Free and Gradient-Based Multi-Objective Optimization Methodologies on the VKI-LS89 Turbine Vane Test Case”. In: *Journal of Turbomachinery* 145.3 (2022). DOI: 10.1115/1.4055577.

[10] A. Châtel and T. Verstraete. “Comparison Between Gradient-Free and Gradient-Based Optimizations of the SRV2 Radial Compressor”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2023. DOI: 10.1115/gt2023-102387.

[11] N. Lachenmaier, D. Baumgärtner, H.-P. Schiffer, and J. Kech. “Gradient-Free and Gradient-Based Optimization of a Radial Turbine”. In: *International Journal of Turbomachinery, Propulsion and Power* 5.3 (2020), p. 14. DOI: 10.3390/ijtpp5030014.

[12] A. Karimian, R. Schmidt, and C. Janke. “Robust Optimization of a Compressor Blade Through Combination of an Adjoint-Based Multistart Approach and Global Gaussian Process Regression”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2024. DOI: 10.1115/gt2024-121351.

[13] T. Verstraete, L. Mueller, and J.-D. Mueller. “Multidisciplinary Adjoint Optimization of Trubomachinery Components Including Aerodynamic and Stress Performance”. In: *35th AIAA Applied Aerodynamics Conference*. American Institute of Aeronautics and Astronautics, 2017. DOI: 10.2514/6.2017-4083.

[14] L. Mueller, T. Verstraete, and M. Schwalbach. “Adjoint-Based Multidisciplinary, Multipoint Optimization of a Radial Turbine Considering Aerodynamic and Structural Performances”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2019. DOI: 10.1115/gt2019-91823.

[15] P. He, C. A. Mader, J. R. R. A. Martins, and K. J. Maki. “DAFoam: An Open-Source Adjoint Framework for Multidisciplinary Design Optimization with OpenFOAM”. In: *AIAA Journal* 58.3 (2020), pp. 1304–1319. DOI: 10.2514/1.j058853.

[16] C. Cuciumita, A. John, N. Qin, and S. Shahpar. “Structurally Constrained Aerodynamic Adjoint Optimisation of Highly Loaded Compressor Blades”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2021. DOI: 10.1115/gt2021-59717.

[17] C. Cuciumita, N. Qin, and S. Shahpar. “Adjoint based aero-structural design optimisation of a transonic fan blade”. In: *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy* 237.6 (2023), pp. 1141–1157. DOI: 10.1177/09576509231162169.

[18] C. Ma, X. Su, and X. Yuan. “An Efficient Unsteady Adjoint Optimization System for Multistage Turbomachinery”. In: *Journal of Turbomachinery* 139.1 (2016). DOI: 10.1115/1.4034185.

[19] S. Vitale, M. Pini, and P. Colonna. “Multistage Turbomachinery Design Using the Discrete Adjoint Method Within the Open-Source Software SU2”. In: *Journal of Propulsion and Power* 36.3 (2020), pp. 465–478. DOI: 10.2514/1.b37685.

[20] R. Lavimi, A. E. Benchikh Le Hocine, S. Poncet, B. Marcos, and R. Panneton. “A review on aerodynamic optimization of turbomachinery using adjoint method”. In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 238.13 (2024), pp. 6405–6441. DOI: 10.1177/09544062231221625.

[21] M. H. Aissa and T. Verstraete. “Metamodel-Assisted Multidisciplinary Design Optimization of a Radial Compressor”. In: *International Journal of Turbomachinery, Propulsion and Power* 4.4 (2019), p. 35. DOI: 10.3390/ijtpp4040035.

[22] Z. Song, X. Zheng, B. Wang, K. Zhou, and R. Amankwa Adjei. “Aerodynamic and structural multidisciplinary optimization design method of fan rotors based on blade curvature constraints”. In: *Aerospace Science and Technology* 136 (2023), p. 108187. DOI: 10.1016/j.ast.2023.108187.

[23] R. Schaffrath, E. Nicke, N. Forsthofer, O. Kunc, and C. Voß. “Gradient-Free Aerodynamic Optimization With Structural Constraints and Surge Line Control for Radial Compressor Stage”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2023. DOI: 10.1115/gt2023-101593.

[24] M. Ghalandari, A. Ziamolki, A. Mosavi, S. Shamshirband, K.-W. Chau, and S. Bornassi. “Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments”. In: *Engineering Applications of Computational Fluid Mechanics* 13.1 (2019), pp. 892–904. DOI: 10.1080/19942060.2019.1649196.

[25] D. Hu, J. Mao, R. Wang, Z. Jia, and J. Song. “Optimization Strategy for a Shrouded Turbine Blade Using Variable-Complexity Modeling Methodology”. In: *AIAA Journal* 54.9 (2016), pp. 2808–2818. DOI: 10.2514/1.j054742.

[26] D. Meng, S. Yang, Y. Zhang, and S. Zhu. “Structural reliability analysis and uncertainties-based collaborative design and optimization of turbine blades using surrogate model”. In: *Fatigue and Fracture of Engineering Materials and Structures* 42.6 (2018), pp. 1219–1227. DOI: 10.1111/ffe.12906.

[27] M. Schnoes, A. Schmitz, G. Goinis, C. Voß, and E. Nicke. “Strategies for Multi-Fidelity Optimization of Multi-Stage Compressors with Throughflow and 3D CFD”. In: *Conference of the International Society for Air Breathing Engines (ISABE)*. 2019. URL: https://elib.dlr.de/129457/.

[28] L. Baert, E. Chérière, C. Sainvitu, I. Lepot, A. Nouvellon, and V. Leonardon. “Aerodynamic Optimization of the Low-Pressure Turbine Module: Exploiting Surrogate Models in a High-Dimensional Design Space”. In: *Journal of Turbomachinery* 142.3 (2020). DOI: 10.1115/1.4046232.

[29] X. Zhang, Y. Ju, Z. Li, F. Liu, and C. Zhang. “Optimization of Three-Dimensional Blade and Variable Stators for Efficiency and Stability Enhancement of Multistage Axial Flow Compressor at Variable Speeds”. In: *Journal of Turbomachinery* 146.4 (2023). DOI: 10.1115/1.4064080.

[30] W. Song and A. J. Keane. “Surrogate-Based Aerodynamic Shape Optimization of a Civil Aircraft Engine Nacelle”. In: *AIAA Journal* 45.10 (2007), pp. 2565–2574. DOI: 10.2514/1.30015.

[31] L. Hartwig and D. Bestle. “Compressor blade design for stationary gas turbines using dimension reduced surrogate modeling”. In: *IEEE Congress on Evolutionary Computation (CEC)*. 2017. DOI: 10.1109/cec.2017.7969493.

[32] L. Zhang, D. Mi, C. Yan, and F. Tang. “Multidisciplinary Design Optimization for a Centrifugal Compressor Based on Proper Orthogonal Decomposition and an Adaptive Sampling Method”. In: *Applied Sciences* 8.12 (2018), p. 2608. DOI: 10.3390/app8122608.

[33] D. I. Lopez, T. Ghisu, and S. Shahpar. “Global Optimisation of a Transonic Fan Blade Through AI-Enabled Active Subspaces”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2021. DOI: 10.1115/gt2021-59166.

[34] R. C. Schlaps, S. Shahpar, and V. Gümmer. “Automatic Three-Dimensional Optimisation of a Modern Tandem Compressor Vane”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2014. DOI: 10.1115/gt2014-26762.

[35] I. Arsenyev, F. Duddeck, and A. Fischersworring-Bunk. “Adaptive Surrogate-Based Multi-Disciplinary Optimization for Vane Clusters”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2015. DOI: 10.1115/gt2015-42164.

[36] L. Baert, P. Beaucaire, M. Leborgne, C. Sainvitu, and I. Lepot. “Tackling Highly Constrained Design Problems: Efficient Optimisation of a Highly Loaded Transonic Compressor”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2017. DOI: 10.1115/gt2017-64610.

[37] G. Antinori. “Uncertainty analysis and robust optimization for low pressure turbine rotors”. PhD thesis. Technical University of Munich, 2017. ISBN: 9783844051551.

[38] I. Arsenyev. “Efficient surrogate-based robust design optimization method”. PhD thesis. Technical University of Munich, 2018. ISBN: 9783844060904.

[39] N. Ludwig. “Direct and inverse methods to account for epistemic and lack-of-knowledge uncertainties in turbine development”. PhD thesis. Technical University of Munich, 2024. URL: https://mediatum.ub.tum.de/1724423.

[40] U. Siller, C. Voß, and E. Nicke. “Automated Multidisciplinary Optimization of a Transonic Axial Compressor”. In: *47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition*. American Institute of Aeronautics and Astronautics, 2009. DOI: 10.2514/6.2009-863.

[41] D. R. Jones, M. Schonlau, and W. J. Welch. “Efficient Global Optimization of Expensive Black-Box Functions”. In: *Journal of Global Optimization* 13.4 (1998), pp. 455–492. DOI: 10.1023/a:1008306431147.

[42] M. Schonlau, W. J. Welch, and D. R. Jones. “Global versus local search in constrained optimization of computer models”. In: *Institute of Mathematical Statistics Lecture Notes - Monograph Series*. Institute of Mathematical Statistics, 1998, pp. 11–25. DOI: 10.1214/lnms/1215456182.

[43] J. Mockus. *Bayesian Approach to Global Optimization*. Springer Netherlands, 1989. DOI: 10.1007/978-94-009-0909-0.

[44] D. G. Krige. “A statistical approach to some basic mine valuation problems on the Witwatersrand”. In: *Journal of The South African Institute of Mining and Metallurgy* 52 (1951), pp. 201–203.

[45] G. Matheron. “Principles of geostatistics”. In: *Economic Geology* 58 (1963), pp. 1246–1266.

[46] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. “Design and Analysis of Computer Experiments”. In: *Statistical Science* 4.4 (1989). DOI: 10.1214/ss/1177012413.

[47] A. Giunta, S. Wojtkiewicz, and M. Eldred. “Overview of Modern Design of Experiments Methods for Computational Simulations (Invited)”. In: *41st Aerospace Sciences Meeting and Exhibit*. American Institute of Aeronautics and Astronautics, 2003. DOI: 10.2514/6.2003-649.

[48] M. D. McKay, R. J. Beckman, and W. J. Conover. “Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code”. In: *Technometrics* 21.2 (1979), pp. 239–245. DOI: 10.1080/00401706.1979.10489755.

[49] M. Johnson, L. Moore, and D. Ylvisaker. “Minimax and maximin distance designs”. In: *Journal of Statistical Planning and Inference* 26.2 (1990), pp. 131–148. DOI: 10.1016/0378-3758(90)90122-b.

[50] R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961. ISBN: 1400874661.

[51] N. Altman and M. Krzywinski. “The curse(s) of dimensionality”. In: *Nature Methods* 15.6 (2018), pp. 399–400. DOI: 10.1038/s41592-018-0019-x.

[52] A. I. J. Forrester, A. Sóbester, and A. J. Keane. *Engineering Design via Surrogate Modelling*. Wiley, 2008. DOI: 10.1002/9780470770801.

[53] H. Khatouri, T. Benamara, P. Breitkopf, and J. Demange. “Metamodeling techniques for CPU-intensive simulation-based design optimization: a survey”. In: *Advanced Modeling and Simulation in Engineering Sciences* 9.1 (2022). DOI: 10.1186/s40323-022-00214-y.

[54] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2006. ISBN: 026218253X. URL: https://gaussianprocess.org/gpml/.

[55] J. Görtler, R. Kehlbeck, and O. Deussen. “A Visual Exploration of Gaussian Processes”. In: *Distill* 4.4 (2019). DOI: 10.23915/distill.00017.

[56] A. Keane and P. Nair. *Computational Approaches for Aerospace Design: The Pursuit of Excellence*. Wiley, 2005. 582 pp. DOI: 10.1002/0470855487.

[57] D. K. Duvenaud. “Automatic Model Construction with Gaussian Processes”. PhD thesis. University of Cambridge, Cambridge, UK, 2014.

[58] M. Binois and N. Wycoff. “A Survey on High-dimensional Gaussian Process Modeling with Application to Bayesian Optimization”. In: *ACM Transactions on Evolutionary Learning and Optimization* 2.2 (2022), pp. 1–26. DOI: 10.1145/3545611.

[59] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd, 2007. DOI: 10.1002/9780470725184.

[60] W. E. Becker, S. Tarantola, and G. Deman. “Sensitivity analysis approaches to high-dimensional screening problems at low sample size”. In: *Journal of Statistical Computation and Simulation* 88.11 (2018), pp. 2089–2110. DOI: 10.1080/00949655.2018.1450876.

[61] M. Aulich and U. Siller. “High-Dimensional Constrained Multiobjective Optimization of a Fan Stage”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2011. DOI: 10.1115/gt2011-45618.

[62] D. Duvenaud, H. Nickisch, and C. E. Rasmussen. “Additive Gaussian Processes”. In: *Advances in Neural Information Processing Systems* (2011). DOI: 10.48550/arXiv.1112.4394.

[63] K. Kandasamy, J. Schneider, and B. Poczos. “High Dimensional Bayesian Optimisation and Bandits via Additive Models”. In: *Proceedings of the 32nd International Conference on Machine Learning*. 2015. DOI: 10.48550/arXiv.1503.01673.

[64] P. Rolland, J. Scarlett, I. Bogunovic, and V. Cevher. “High-Dimensional Bayesian Optimization via Additive Models with Overlapping Groups”. In: *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*. Vol. 84. Proceedings of Machine Learning Research. 2018, pp. 298–307. DOI: 10.48550/arXiv.1802.07028.

[65] T. Muehlenstaedt, O. Roustant, L. Carraro, and S. Kuhnt. “Data-driven Kriging models based on FANOVA-decomposition”. In: *Statistics and Computing* 22.3 (2011), pp. 723–738. DOI: 10.1007/s11222-011-9259-7.

[66] R. Tripathy, I. Bilionis, and M. Gonzalez. “Gaussian processes with built-in dimensionality reduction: Applications to high-dimensional uncertainty propagation”. In: *Journal of Computational Physics* 321 (2016), pp. 191–223. DOI: 10.1016/j.jcp.2016.05.039.

[67] P. G. Constantine, E. Dow, and Q. Wang. “Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces”. In: *SIAM Journal on Scientific Computing* 36.4 (2014), A1500–A1524. DOI: 10.1137/130916138.

[68] M. A. Bouhlel, N. Bartoli, A. Otsmane, and J. Morlier. “Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction”. In: *Structural and Multidisciplinary Optimization* 53.5 (2016), pp. 935–952. DOI: 10.1007/s00158-015-1395-9.

[69] P. G. Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. Society for Industrial and Applied Mathematics, 2015. DOI: 10.1137/1.9781611973860.

[70] E. V. Bonilla, K. Chai, and C. Williams. “Multi-task Gaussian Process Prediction”. In: *Advances in Neural Information Processing Systems*. Vol. 20. 2007.

[71] J. Yu, C. Yan, and M. Guo. “Non-intrusive reduced-order modeling for fluid problems: A brief review”. In: *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 233.16 (2019), pp. 5896–5912. DOI: 10.1177/0954410019890721.

[72] L. Sirovich. “Turbulence and the dynamics of coherent structures. II. Symmetries and transformations”. In: *Quarterly of Applied Mathematics* 45.3 (1987), pp. 573–582. DOI: 10.1090/qam/910463.

[73] J. Weiss. “A Tutorial on the Proper Orthogonal Decomposition”. In: *AIAA Aviation Forum*. American Institute of Aeronautics and Astronautics, 2019. DOI: 10.2514/6.2019-3333.

[74] H. V. Ly and H. T. Tran. “Modeling and control of physical processes using proper orthogonal decomposition”. In: *Mathematical and Computer Modelling* 33.1–3 (2001), pp. 223–236. DOI: 10.1016/s0895-7177(00)00240-5.

[75] D. Higdon, J. Gattiker, B. Williams, and M. Rightley. “Computer Model Calibration Using High-Dimensional Output”. In: *Journal of the American Statistical Association* 103.482 (2008), pp. 570–583. DOI: 10.1198/016214507000000888.

[76] R. B. Gramacy. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. Chapman and Hall/CRC, 2020. DOI: 10.1201/9780367815493.

[77] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. “Taking the Human Out of the Loop: A Review of Bayesian Optimization”. In: *Proceedings of the IEEE* 104.1 (2016), pp. 148–175. DOI: 10.1109/jproc.2015.2494218.

[78] R. Garnett. *Bayesian Optimization*. Cambridge University Press, 2023. DOI: 10.1017/9781108348973.

[79] R. Lam, M. Poloczek, P. Frazier, and K. E. Willcox. “Advances in Bayesian Optimization with Applications in Aerospace Engineering”. In: *2018 AIAA Non-Deterministic Approaches Conference*. American Institute of Aeronautics and Astronautics, 2018. DOI: 10.2514/6.2018-1656.

[80] H. J. Kushner. “A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise”. In: *Journal of Basic Engineering* 86.1 (1964), pp. 97–106. DOI: 10.1115/1.3653121.

[81] N. Srinivas, A. Krause, S. Kakade, and M. Seeger. “Gaussian process optimization in the bandit setting: no regret and experimental design”. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. 2010, pp. 1015–1022. ISBN: 9781605589077.

[82] K. Kandasamy, A. Krishnamurthy, J. Schneider, and B. Poczos. “Parallelised Bayesian Optimisation via Thompson Sampling”. In: *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. Vol. 84. Proceedings of Machine Learning Research. 2018, pp. 133–142.

[83] N. Hansen. “The CMA Evolution Strategy: A Tutorial”. 2016. DOI: 10.48550/arXiv.1604.00772.

[84] J. A. Nelder and R. Mead. “A Simplex Method for Function Minimization”. In: *The Computer Journal* 7.4 (1965), pp. 308–313. DOI: 10.1093/comjnl/7.4.308.

[85] M. A. Luersen and R. Le Riche. “Globalized Nelder-Mead method for engineering optimization”. In: *Computers & Structures* 82.23–26 (2004), pp. 2251–2260. DOI: 10.1016/j.compstruc.2004.03.072.

[86] W. Ponweiser, T. Wagner, and M. Vincze. “Clustered multiple generalized expected improvement: A novel infill sampling criterion for surrogate models”. In: *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. 2008. DOI: 10.1109/cec.2008.4631273.

[87] D. Ginsbourger, R. Le Riche, and L. Carraro. “Kriging Is Well-Suited to Parallelize Optimization”. In: *Adaptation Learning and Optimization*. Springer Berlin Heidelberg, 2010, pp. 131–162. DOI: 10.1007/978-3-642-10701-6_6.

[88] M. L. Santoni, E. Raponi, R. De Leone, and C. Doerr. “Comparison of High-Dimensional Bayesian Optimization Algorithms on BBOB”. In: *ACM Transactions on Evolutionary Learning and Optimization* (2024). DOI: 10.1145/3670683.

[89] A. Spagnol. “Kernel-based sensitivity indices for high-dimensional optimization problems”. PhD thesis. Université de Lyon, 2020. URL: https://theses.hal.science/tel-03173192.

[90] C. Li, S. Gupta, S. Rana, V. Nguyen, S. Venkatesh, and A. Shilton. “High dimensional Bayesian optimization using dropout”. In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017, pp. 2096–2102. DOI: 10.48550/arXiv.1802.05400.

[91] D. Zhan, J. Wu, H. Xing, and T. Li. “A cooperative approach to efficient global optimization”. In: *Journal of Global Optimization* (2023). DOI: 10.1007/s10898-023-01316-6.

[92] J. Gardner, C. Guo, K. Weinberger, R. Garnett, and R. Grosse. “Discovering and Exploiting Additive Structure for Bayesian Optimization”. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. Vol. 54. Proceedings of Machine Learning Research. 2017, pp. 1311–1319. URL: https://proceedings.mlr.press/v54/gardner17a.html.

[93] E. Han, I. Arora, and J. Scarlett. “High-Dimensional Bayesian Optimization via Tree-Structured Additive Models”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. Association for the Advancement of Artificial Intelligence. 2021, pp. 7630–7638. DOI: 10.48550/arXiv.2012.13088.

[94] J. K. Ziomek and H. Bou Ammar. “Are Random Decompositions all we need in High Dimensional Bayesian Optimisation?” In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. Proceedings of Machine Learning Research. 2023, pp. 43347–43368. URL: https://proceedings.mlr.press/v202/ziomek23a.html.

[95] Z. Wang, F. Hutter, M. Zoghi, D. Matheson, and N. D. Feitas. “Bayesian Optimization in a Billion Dimensions via Random Embeddings”. In: *Journal of Artificial Intelligence Research* 55 (2016), pp. 361–387. DOI: 10.1613/jair.4806.

[96] A. Nayebi, A. Munteanu, and M. Poloczek. “A framework for Bayesian optimization in embedded subspaces”. In: *Proceedings of the 36th International Conference on Machine Learning*. Vol. 97. Proceedings of Machine Learning Research. 2019, pp. 4752–4761.

[97] M. Binois, D. Ginsbourger, and O. Roustant. “On the choice of the low-dimensional domain for global optimization via random embeddings”. In: *Journal of Global Optimization* 76.1 (2019), pp. 69–90. DOI: 10.1007/s10898-019-00839-1.

[98] B. Letham, R. Calandra, A. Rai, and E. Bakshy. “Re-Examining Linear Embeddings for High-Dimensional Bayesian Optimization”. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 1546–1558. DOI: 10.48550/arXiv.2001.11659.

[99] T. W. Lukaczyk, P. Constantine, F. Palacios, and J. J. Alonso. “Active Subspaces for Shape Optimization”. In: *10th AIAA Multidisciplinary Design Optimization Conference*. American Institute of Aeronautics and Astronautics, 2014. DOI: 10.2514/6.2014-1171.

[100] E. Raponi, H. Wang, M. Bujny, S. Boria, and C. Doerr. “High Dimensional Bayesian Optimization Assisted by Principal Component Analysis”. In: *Parallel Problem Solving from Nature – PPSN XVI*. Springer International Publishing, 2020, pp. 169–183. DOI: 10.1007/978-3-030-58112-1_12.

[101] R. Priem, Y. Diouane, N. Bartoli, S. Dubreuil, and P. Saves. “High-Dimensional Bayesian Optimization Using Both Random and Supervised Embeddings”. In: *AIAA Journal* (2024), pp. 1–13. DOI: 10.2514/1.j063488.

[102] R. G. Regis. “Trust regions in Kriging-based optimization with expected improvement”. In: *Engineering Optimization* 48.6 (2015), pp. 1037–1059. DOI: 10.1080/0305215x.2015.1082350.

[103] D. Eriksson, M. Pearce, J. R. Gardner, R. Turner, and M. Poloczek. “Scalable global optimization via local Bayesian optimization”. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019.

[104] Y. Diouane, V. Picheny, R. L. Riche, and A. S. D. Perrotolo. “TREGO: a trust-region framework for efficient global optimization”. In: *Journal of Global Optimization* (2022). DOI: 10.1007/s10898-022-01245-w.

[105] D. Eriksson and M. Poloczek. “Scalable Constrained Bayesian Optimization”. In: *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*. Vol. 130. Proceedings of Machine Learning Research. 2021, pp. 730–738.

[106] J. M. Parr, A. J. Keane, A. I. Forrester, and C. M. Holden. “Infill sampling criteria for surrogate-based optimization with constraint handling”. In: *Engineering Optimization* 44.10 (2012), pp. 1147–1166. DOI: 10.1080/0305215x.2011.637556.

[107] M. Sasena, P. Papalambros, and P. Goovaerts. “Global Optimization of Problems with Disconnected Feasible Regions via Surrogate Modeling”. In: *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*. American Institute of Aeronautics and Astronautics, 2002. DOI: 10.2514/6.2002-5573.

[108] M. J. Sasena, P. Papalambros, and P. Goovaerts. “Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization”. In: *Engineering Optimization* 34.3 (2002), pp. 263–278. DOI: 10.1080/03052150211751.

[109] R. B. Gramacy, G. A. Gray, S. Le Digabel, H. K. H. Lee, P. Ranjan, G. Wells, and S. M. Wild. “Modeling an Augmented Lagrangian for Blackbox Constrained Optimization”. In: *Technometrics* 58.1 (2016), pp. 1–11. DOI: 10.1080/00401706.2015.1014065.

[110] C. Audet, J. Denni, D. Moore, A. Booker, and P. Frank. “A surrogate-model-based method for constrained optimization”. In: *8th Symposium on Multidisciplinary Analysis and Optimization*. American Institute of Aeronautics and Astronautics, 2000. DOI: 10.2514/6.2000-4891.

[111] A. Basudhar, C. Dribusch, S. Lacaze, and S. Missoum. “Constrained efficient global optimization with support vector machines”. In: *Structural and Multidisciplinary Optimization* 46.2 (2012), pp. 201–221. DOI: 10.1007/s00158-011-0745-5.

[112] H. Maathuis, R. De Breuker, and S. G. P. Castro. “High-Dimensional Bayesian Optimisation with Large-Scale Constraints – An Application to Aeroelastic Tailoring”. In: *AIAA SCITECH Forum*. American Institute of Aeronautics and Astronautics, 14, 2023. DOI: 10.2514/6.2024-2012.

[113] J. Sobieszczanski-Sobieski and R. T. Haftka. “Multidisciplinary aerospace design optimization: survey of recent developments”. In: *Structural Optimization* 14.1 (1997), pp. 1–23. DOI: 10.1007/bf01197554.

[114] J. R. R. A. Martins and A. Ning. *Engineering Design Optimization*. Cambridge University Press, 2021. 475 pp. DOI: 10.1017/9781108980647.

[115] J. R. R. A. Martins and A. B. Lambe. “Multidisciplinary Design Optimization: A Survey of Architectures”. In: *AIAA Journal* 51.9 (2013), pp. 2049–2075. DOI: 10.2514/1.j051895.

[116] E. J. Cramer, J. E. Dennis Jr., P. D. Frank, R. M. Lewis, and G. R. Shubin. “Problem Formulation for Multidisciplinary Optimization”. In: *SIAM Journal on Optimization* 4.4 (1994), pp. 754–776. DOI: 10.1137/0804044.

[117] S. Dubreuil, N. Bartoli, C. Gogu, and T. Lefebvre. “Towards an efficient global multidisciplinary design optimization algorithm”. In: *Structural and Multidisciplinary Optimization* 62.4 (2020), pp. 1739–1765. DOI: 10.1007/s00158-020-02514-6.

[118] I. Cardoso, S. Dubreuil, N. Bartoli, C. Gogu, and M. Salaün. “Constrained efficient global multidisciplinary design optimization using adaptive disciplinary surrogate enrichment”. In: *Structural and Multidisciplinary Optimization* 67.2 (2024). DOI: 10.1007/s00158-023-03736-0.

[119] J. Sobieszczanski-Sobieski. *Optimization by decomposition: A step from hierarchic to non-hierarchic systems*. Tech. rep. NASA Langley Research Center, 1989.

[120] R. Sellar, S. Batill, and J. Renaud. “Response surface based, concurrent subspace optimization for multidisciplinary system design”. In: *34th Aerospace Sciences Meeting and Exhibit*. American Institute of Aeronautics and Astronautics, 1996. DOI: 10.2514/6.1996-714.

[121] R. Braun, P. Gage, I. Kroo, and I. Sobieski. “Implementation and performance issues in collaborative optimization”. In: *6th Symposium on Multidisciplinary Analysis and Optimization*. American Institute of Aeronautics and Astronautics, 1996. DOI: 10.2514/6.1996-4017.

[122] R. F. Coelho, P. Breitkopf, C. Knopf-Lenoir, and P. Villon. “Bi-level model reduction for coupled problems”. In: *Structural and Multidisciplinary Optimization* 39.4 (2009), pp. 401–418. DOI: 10.1007/s00158-008-0335-3.

[123] G. Berthelin, S. Dubreuil, M. Salaün, N. Bartoli, and C. Gogu. “Disciplinary proper orthogonal decomposition and interpolation for the resolution of parameterized multidisciplinary analysis”. In: *International Journal for Numerical Methods in Engineering* (2022). DOI: 10.1002/nme.6981.

[124] MTU Aero Engines AG. “How does a turbofan engine work?” In: *AEROREPORT* (2024). Accessed: 2024-11-29. URL: https://aeroreport.de/en/good-to-know/how-does-a-turbofan-engine-work.

[125] Rolls-Royce. *The jet engine*. Wiley, 2015. 1288 pp. ISBN: 1523111143.

[126] W. J. Bräunling. *Flugzeugtriebwerke*. Springer Berlin Heidelberg, 2015. DOI: 10.1007/978-3-642-34539-5.

[127] N. Cumpsty and A. Heyes. *Jet Propulsion: A Simple Guide to the Aerodynamics and Thermodynamic Design and Performance of Jet Engines*. Cambridge University Press, 2015. DOI: 10.1017/cbo9781316223116.

[128] H. Rick. *Gasturbinen und Flugantriebe*. Springer Berlin Heidelberg, 2013. DOI: 10.1007/978-3-540-79446-2.

[129] R. Nigro, L. Baert, F. Nyssen, J. de Cazenove, J. Dominique, I. Lepot, M. Veglio, and R. Princivalle. “Multi-Fidelity Aeromechanical Design Framework for High Flow Speed Multistage Axial Compressors”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2024. DOI: 10.1115/gt2024-122244.

[130] M. G. Nagel. “Numerische Optimierung dreidimensional parametrisierter Turbinenschaufeln mit umfangsunsymmetrischen Plattformen: Entwicklung, Anwendung und Validierung.” PhD thesis. Universität der Bundeswehr München, 2004. URL: https://athene-forschung.rz.unibw-muenchen.de/doc/85317/85317.pdf.

[131] M. G. Nagel and R.-D. Baier. “Experimentally Verified Numerical Optimization of a Three-Dimensional Parametrized Turbine Vane With Nonaxisymmetric End Walls”. In: *Journal of Turbomachinery* 127.2 (2005), pp. 380–387. DOI: 10.1115/1.1773848.

[132] J. H. Ferziger, M. Perić, and R. L. Street. *Computational Methods for Fluid Dynamics*. Springer International Publishing, 2020. DOI: 10.1007/978-3-319-99693-6.

[133] J. D. Denton and W. N. Dawes. "Computational fluid dynamics for turbomachinery design". In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 213.2 (1998), pp. 107–124. DOI: 10.1243/0954406991522211.

[134] J. J. Adamczyk. "Aerodynamic Analysis of Multistage Turbomachinery Flows in Support of Aerodynamic Design". In: *Journal of Turbomachinery* 122.2 (1999), pp. 189–217. DOI: 10.1115/1.555439.

[135] R. N. Pinto, A. Afzal, L. V. D'Souza, Z. Ansari, and A. D. Mohammed Samee. "Computational Fluid Dynamics in Turbomachinery: A Review of State of the Art". In: *Archives of Computational Methods in Engineering* 24.3 (2016), pp. 467–479. DOI: 10.1007/s11831-016-9175-2.

[136] R. D. Sandberg and V. Michelassi. "The Current State of High-Fidelity Simulations for Main Gas Path Turbomachinery Components and Their Industrial Impact". In: *Flow, Turbulence and Combustion* 102.4 (2019), pp. 797–848. DOI: 10.1007/s10494-019-00013-3.

[137] D. C. Wilcox. "Reassessment of the scale-determining equation for advanced turbulence models". In: *AIAA Journal* 26.11 (1988), pp. 1299–1310. DOI: 10.2514/3.10041.

[138] NUMECA Ingenieurbüro. *AutoGrid*. https://www.numeca.de/en/products-meshing-solutions/. Accessed: 2024-11-29.

[139] J. D. Denton. "The Calculation of Three-Dimensional Viscous Flow Through Multistage Turbomachines". In: *Journal of Turbomachinery* 114.1 (1992), pp. 18–26. DOI: 10.1115/1.2927983.

[140] M. B. Giles. "Nonreflecting boundary conditions for Euler equation calculations". In: *AIAA Journal* 28.12 (1990), pp. 2050–2058. DOI: 10.2514/3.10521.

[141] German Aerospace Center (DLR). *TRACE*. http://www.trace-portal.de/userguide/trace/index.html. Accessed: 2024-11-29. 2024.

[142] T. Belytschko. *Nonlinear Finite Elements for Continua and Structures*. 1st ed. New York Academy of Sciences Series. Description based on publisher supplied metadata and other sources. Newark: John Wiley & Sons, Incorporated, 2014. 1834 pp. ISBN: 9781118700051.

[143] G. Dhondt. *The Finite Element Method for Three-Dimensional Thermomechanical Applications*. Wiley, 2004. 362 pp. DOI: 10.1002/0470021217.

[144] T. Benamara, P. Breitkopf, I. Lepot, C. Sainvitu, and P. Villon. "Multi-fidelity POD surrogate-assisted optimization: Concept and aero-design study". In: *Structural and Multidisciplinary Optimization* 56.6 (2017), pp. 1387–1412. DOI: 10.1007/s00158-017-1730-4.

[145] T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia. "Learning Mesh-Based Simulation with Graph Networks". In: *International Conference on Learning Representations*. 2021. DOI: 10.48550/arXiv.2010.03409.

[146] S. Strönisch, M. Sander, M. Meyer, and A. Knüpfer. "Prediction of Steady and Unsteady Flow Quantities Using Multiscale Graph Neural Networks". In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2024. DOI: 10.1115/gt2024-121697.

[147] L. Papenmeier, L. Nardi, and M. Poloczek. "Increasing the Scope as You Learn: Adaptive Bayesian Optimization in Nested Subspaces". In: *Advances in Neural Information Processing Systems* (2022). DOI: 10.48550/arXiv.2304.11468.

[148] D. J. J. Toal, N. W. Bressloff, A. J. Keane, and C. M. E. Holden. “Geometric Filtration Using Proper Orthogonal Decomposition for Aerodynamic Design Optimization”. In: *AIAA Journal* 48.5 (2010), pp. 916–928. DOI: 10.2514/1.41420.

[149] D. Cinquegrana and E. Iuliano. “Efficient Global Optimization of a Transonic Wing with Geometric Data Reduction”. In: *35th AIAA Applied Aerodynamics Conference*. American Institute of Aeronautics and Astronautics, 2017. DOI: 10.2514/6.2017-3057.

[150] G. Blatman and B. Sudret. “Efficient computation of global sensitivity indices using sparse polynomial chaos expansions”. In: *Reliability Engineering & System Safety* 95.11 (2010), pp. 1216–1229. DOI: 10.1016/j.ress.2010.06.015.

[151] L. Baert, C. Dumeunier, M. Leborgne, C. Sainvitu, and I. Lepot. “Agile SBO Framework Exploiting Multisimulation Data: Optimising Efficiency and Stall Margin of a Transonic Compressor”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2018. DOI: 10.1115/gt2018-76639.

[152] A. Tfaily, M. Kokkolaras, N. Bartoli, and Y. Diouane. “Efficient Acquisition Functions for Bayesian Optimization in the Presence of Hidden Constraints”. In: *AIAA AVIATION Forum*. American Institute of Aeronautics and Astronautics, 2023. DOI: 10.2514/6.2023-4261.

[153] Z. Guo, W. Chu, H. Zhang, C. Liang, and D. Meng. “Statistical evaluation of stability margin of a multi-stage compressor with geometric variability using adaptive polynomial chaos-Kriging model”. In: *Physics of Fluids* 35.7 (2023). DOI: 10.1063/5.0158821.

A Publications

A.1 Publication I

L. Pretsch, I. Arsenyev, C. Czech, and F. Duddeck. “Interdisciplinary design optimization of compressor blades combining low- and high-fidelity models”. In: *Structural and Multidisciplinary Optimization* 66.4 (2023). DOI: 10.1007/s00158-023-03516-w

Structural and Multidisciplinary Optimization (2023) 66:70
https://doi.org/10.1007/s00158-023-03516-w

RESEARCH PAPER

Interdisciplinary design optimization of compressor blades combining low- and high-fidelity models

Lisa Pretsch[1] · **Ilya Arsenyev**[2] · **Catharina Czech**[1] · **Fabian Duddeck**[1]

Received: 19 August 2022 / Revised: 23 January 2023 / Accepted: 30 January 2023 / Published online: 16 March 2023

Abstract
Multidisciplinary design optimization has great potential to support the turbomachinery development process by improving designs at reduced time and cost. As part of the industrial compressor design process, we seek for a rotor blade geometry that minimizes stresses without impairing the aerodynamic performance. However, the presence of structural mechanics, aerodynamics, and their interdisciplinary coupling poses challenges concerning computational effort and organizational integration. In order to reduce both computation times and the required exchange between disciplinary design teams, we propose an *inter*- instead of *multi*disciplinary design optimization approach tailored to the studied optimization problem. This involves a distinction between main and side discipline. The main discipline, structural mechanics, is computed by accurate high-fidelity finite element models. The side discipline, aerodynamics, is represented by efficient low-fidelity models, using Kriging and proper-orthogonal decomposition to approximate constraints and the gas load field as coupling variable. The proposed approach is shown to yield a valid blade design with reasonable computational effort for training the aerodynamic low-fidelity models and significantly reduced optimization times compared to a high-fidelity multidisciplinary design optimization. Especially for expensive side disciplines like aerodynamics, the multi-fidelity interdisciplinary design optimization has the potential to consider the effects of all involved disciplines at little additional cost and organizational complexity, while keeping the focus on the main discipline.

Keywords Multidisciplinary design optimization · Multi-fidelity methods · Kriging · Proper-orthogonal decomposition · Turbomachinery

1 Introduction

Reduced development costs and decreased emissions are two of the goals set by the aerospace industry and the European Commission (2011) in the Flightpath 2050. Reduced development costs imply shorter design cycles with less iterations between the disciplines, motivating the use of multidisciplinary approaches. Decreased emissions can be achieved by either innovative concepts or by optimizing existing aircraft components.

One of the main aircraft engine components is the compressor, responsible for raising the pressure level of the intake air, while it passes through several rows of rotor and stator blades, see Fig. 1. The compressor development involves several engineering disciplines, notably aerodynamics and structural mechanics.

1.1 Aero-structural optimization problem statement

As part of the industrial aero-structural compressor design process, we seek for the best possible blade geometry. After the initial design optimization by the aerodynamic department, the second step is a structural optimization of individual blades. However, design changes that are beneficial from the structural mechanics point of view may counteract the aerodynamic performance and disturb the aerodynamic integration with the neighboring stages. Structural blade design

Responsible Editor: Lei Wang

✉ Lisa Pretsch
lisa.pretsch@tum.de

[1] TUM School of Engineering and Design, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

[2] MTU Aero Engines, Dachauer Str. 665, 80995 Munich, Germany

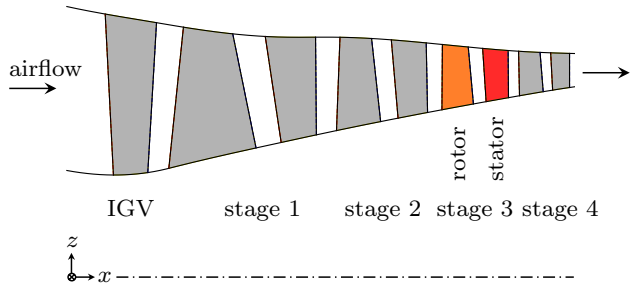


Fig. 1 High-pressure compressor frontblock annulus geometry illustration with inlet guide vane (IGV) and four stages of rotor and stator blades. The third-stage rotor and stator blades, highlighted in orange and red, are the subject of the treated optimization problem. The arrows indicate the airflow in x-direction

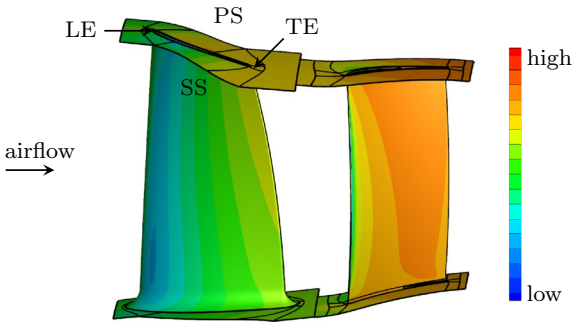


Fig. 2 Pressure field on the third-stage rotor and stator blade surface and annulus walls as a result of the CFD simulations. The four profile sides pressure side (PS), suction side (SS), leading edge (LE), and trailing edge (TE) are indicated on the rotor blade

problems should thus include aerodynamic constraints and aero-structural coupling.

A frequently arising optimization problem is the minimization of stresses in the compressor blades without impairing the aerodynamic performance. It is herein studied for the third-stage rotor blade, highlighted in orange in Fig. 1. Figure 2 shows the qualitative pressure field on the third-stage surfaces, as a result of CFD simulations with the initial geometry. It induces aerodynamic loads, the gas loads, which in turn affect the stress field, depicted in Fig. 3 for the rotor suction side.

The blade design problem can be treated either by an iterative process between disciplinary design teams or in a multidisciplinary design optimization (MDO) (Sobieszczanski-Sobieski and Haftka 1997). Papageorgiou et al. (2018) reviewed recent advancements and challenges in MDO of aerial vehicles and provided a roadmap including nine fundamental elements, among others computational efficiency and organizational integration.

For the above-mentioned aero-structural compressor blade design task, these two elements are particularly challenging and inhibit the application of MDO as a standard tool in the industrial development process. First, computing times can be prohibitively long if expensive aerodynamic simulations are performed in addition to the structural analyses in every optimization iteration. Interdisciplinary coupling additionally increases computational cost and complexity. It is therefore often neglected or strongly simplified at the expense of model accuracy. Second, the integration into company structures is usually difficult, as disciplinary department boundaries impede multidisciplinary developments. The aim for a practical optimization approach thus is as little disciplinary interdependence as possible and as much as needed for a useful optimization result. In other words, a multidisciplinary feasible solution should be obtained while maintaining the division of work between responsible discipline-focused company units.

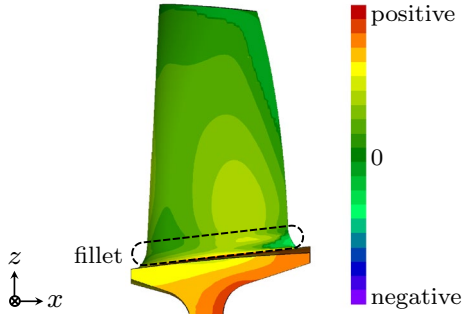


Fig. 3 Maximum principal stress field on the rotor blade suction side as a result of the FE simulations. The stresses depend on the rotor blade geometry and are induced by centrifugal forces, thermal loads, and gas loads. The latter result from the aerodynamic pressures in Fig. 2. The dashed line indicates the fillet region

1.2 Multi-fidelity approaches

The first challenge, the computational efficiency, is commonly tackled by parallelized computations and multi-fidelity methods. An additional option is adjoint methods for efficient gradient-based optimization, not discussed herein, as used among others by He et al. (2020) in an uncoupled aero-structural compressor blade optimization. Multi-fidelity methods (Viana et al. 2014) speed up computations by complementing a comparatively small number of high-fidelity simulations, here finite element (FE) and computational fluid dynamics (CFD) simulations, with a large amount of low-fidelity model evaluations. The latter are cheaper to evaluate, while providing useful information within the design domain. They are also called surrogate models or metamodels. Before outlining our ideas, we review recent literature on the application of low-fidelity models in aero-structural optimization of aircraft engine blades.

The most widely used low-fidelity models are data-fit models. Cuciumita et al. (2021) approximated a maximum stress constraint by a radial basis function (RBF) model. Aissa and Verstraete (2019) proposed a bounded Kriging model as robust method for surrogate-assisted MDO of compressor blades. Neural networks were recently used by Ghalandari et al. (2019) to model aerodynamic performance and stress levels, or by Vanti et al. (2018) in an uncoupled aeroelastic optimization.

Another type of low-fidelity models are reduced-order models, including proper-orthogonal decomposition (POD). Benamara et al. (2017) proposed multi-fidelity non-intrusive POD to predict isentropic efficiency and pressure ratio. Instead of approximating the response variables, Zhang et al. (2018) replaced the 30 original design parameters in the MDO by only four POD coefficients to decrease computation time.

In the above turbomachinery references, static aero-structural coupling is mostly neglected. In aircraft wing optimization, Coelho et al. (2009) used reduced-order models to approximate pressure loading and displacement field. The low-fidelity models enabled coupling at little additional computational cost, enhancing the accuracy of the disciplinary high-fidelity results.

The second challenge, the organizational integration of MDO, is a much less active research area (Papageorgiou et al. 2018). Lian and Liou (2006) are one of the few references to address both challenges together for a statically coupled aero-structural compressor optimization. They used genetic algorithms solely based on quadratic response surface models. These low-fidelity models were in turn based on high-fidelity samples with unidirectional coupling via the aerodynamic pressure field, realized by sequential CFD and FE simulations. Hu et al. (2016) employed a collaborative optimization (CO) strategy for a bidirectionally coupled aero-structural turbine blade optimization. The distributed CO strategy guarantees disciplinary autonomy for better organizational integration. Its commonly prohibitive computational expense was reduced by quadratic response surface models, based on data from medium-fidelity aerodynamic and structural analyses and high-fidelity fluid–structure interaction simulations.

The above literature references have in common that they treat both disciplines as equally important. They do not exploit the fact that industrial aero-structural optimizations are often performed with focus on monodisciplinary objectives, and include multidisciplinary coupling and constraints merely not to interfere with the respective other discipline.

1.3 Proposed approach

This work addresses the question of how low-fidelity models can be employed to improve the aero-structural optimization of compressor blades with regard to both computational efficiency and organizational integration.

For the compressor blade optimization problem stated in the beginning, we propose a multi-fidelity optimization process which is *inter*- instead of *multi*disciplinary. This involves a distinction between main and side discipline. The main discipline is structural mechanics, as it is the objective and focus of the optimization problem. The side discipline is aerodynamics. The main discipline is represented by a high-fidelity model. Additionally, the side discipline and the coupling from side to main discipline is taken into account by low-fidelity models only. Thereby, the main discipline is evaluated as accurately as possible, while the side discipline is computed only as accurately as necessary for a useful optimization result.

The benefit of the proposed approach is shown by comparison to a common uncoupled and purely high-fidelity MDO. Despite a significant computation speed-up, the approach provides a valid blade design that is not only structurally optimized, but also respects aerodynamic effects and constraints.

The paper is structured as follows: In Sect. 2, multi-fidelity and multidisciplinary optimization approaches are outlined, followed by the employed low-fidelity modeling methods in Sect. 3. The proposed interdisciplinary optimization process is presented in Sect. 4. Afterward, in Sect. 5, the computational set-up and concrete optimization problem of the compressor blade application are defined. The results are presented in Sect. 6 and the paper is concluded by a summary and outlook in Sect. 7.

2 Multi-fidelity and multidisciplinary optimization methods

In the present context, optimization (Martins and Ning 2021) uses numerical methods to seek the overall best design of an engineering system. It can yield an improved system performance at reduced time and cost compared to a conventional development process. However, the correct problem formulation is crucial and requires expertise in both the involved engineering discipline(s) and numerical optimization.

A standard optimization problem is to minimize an objective function by varying the design variables within their prescribed bounds subject to equality and inequality constraints. For compressor blade design, common objectives are minimum stresses or maximum efficiency, the design parameters define the blade geometry, and constraints

concern for example mass flow, pressure ratio, surge margin, stresses, and eigenfrequencies.

In every optimization iteration, objectives and constraints need to be evaluated. With aerodynamics and structural mechanics as the most influential disciplines in compressor blade design, this usually entails thousands of CFD and FE simulations.

2.1 Multi-fidelity optimization processes

Expensive high-fidelity simulations can be partly or fully replaced by efficient low-fidelity model evaluations within the optimization. Models of various fidelities are combined in multi-fidelity optimization (MFO) processes, either by adaptation, fusion, or filtering (Peherstorfer et al. 2018; Khatouri et al. 2022).

In adaptive MFO, the low-fidelity model is improved by high-fidelity results at new sampling points during the optimization. Fusion strategies combine the outputs of models of different fidelities. Filtering means that the high-fidelity model is only evaluated if the low-fidelity output meets a predefined criterion. Note that all of these approaches combine models of multiple fidelities with the same output parameters and thereby differ from the approach we propose in Sect. 4.

2.2 Multidisciplinary optimization processes

Just like models of various fidelities are embedded in a multi-fidelity optimization process, the presence of several disciplines allows different approaches for the efficient organization of the disciplinary analyses and optimization methods. MDO processes, or architectures, were extensively reviewed by Martins and Lambe (2013) and compared by Tedford and Martins (2009) and Gray et al. (2013) based on benchmark problems. They classify MDO architectures as monolithic or distributed.

The former treat the MDO problem in a single optimization and are therefore simple to implement and well-suited for small problems. However, they are inefficient for problems with a large number of disciplines. Here, distributed architectures may perform better. They decompose the MDO problem into disciplinary optimization subproblems which are coordinated by a system-level optimization subproblem. Thereby, they try to mimic the structure of industrial development teams and strive for independence between the disciplines. Although the ideas of distributed architectures seem promising, they exhibit a slow convergence for many problems (Martins and Ning 2021).

The basic monolithic MDO architecture is the multidisciplinary feasible (MDF) approach. It treats the MDO as a common disciplinary optimization problem, with the difference that objective and constraints are computed by multidisciplinary analyses in every optimization iteration. MDF has the advantage that it makes use of established optimization and multidisciplinary analysis methods. The optimization problem remains as small as possible and is thus also suited for gradient-free methods. Moreover, the multidisciplinary analysis ensures physical compatibility after every optimization iteration. This is especially useful for engineering and industry applications, where the aim usually is finding a better design, rather than a mathematical optimum, and optimizations are often terminated prematurely. The main disadvantage is the computational effort and often slow convergence of the multidisciplinary analyses in case of strong interdisciplinary coupling.

With regard to the use of low-fidelity models, concurrent subspace optimization (CSSO) is an interesting distributed approach. It was introduced by Sobieszczanski-Sobieski (1989) and extended to the version considered herein by Sellar et al. (1996). In CSSO, the coordinating system-level subproblem is optimized based on low-fidelity models. Each disciplinary subproblem is optimized based on high-fidelity models for the corresponding discipline and low-fidelity models for all other disciplines. The disciplinary subproblem results are used to update the low-fidelity models. Despite the fast low-fidelity model evaluations, the slow convergence hinders its efficient application.

3 Low-fidelity models

Before starting the interdisciplinary optimization process, low-fidelity models of the aerodynamic quantities of interest must be generated. According to Peherstorfer et al. (2018), low-fidelity models can be divided into three categories: data-fit models, projection-based models, and simplified models. While data-fit models are purely mathematical response surfaces, projection-based models, hereinafter referred to as reduced-order models, represent a system by its most important eigenmodes and thereby provide a certain physical interpretability. Simplified models, for example with a coarser mesh, are based on the original problem physics. However, their evaluation often takes a lot longer than for the other two categories and they may be difficult to combine with the high-fidelity models; that is why they are omitted in this work.

The process of data-fit and reduced-order model generation followed for the optimization problem herein is illustrated as a flowchart in Fig. 4. First, a set of input points in the design space is selected, also called design of experiments (DoE) (Giunta et al. 2003). Popular approaches are Monte Carlo sampling, Latin hypercube sampling (LHS), and low discrepancy sequences. Then, the corresponding outputs are computed by high-fidelity simulations, here CFD for aerodynamics.

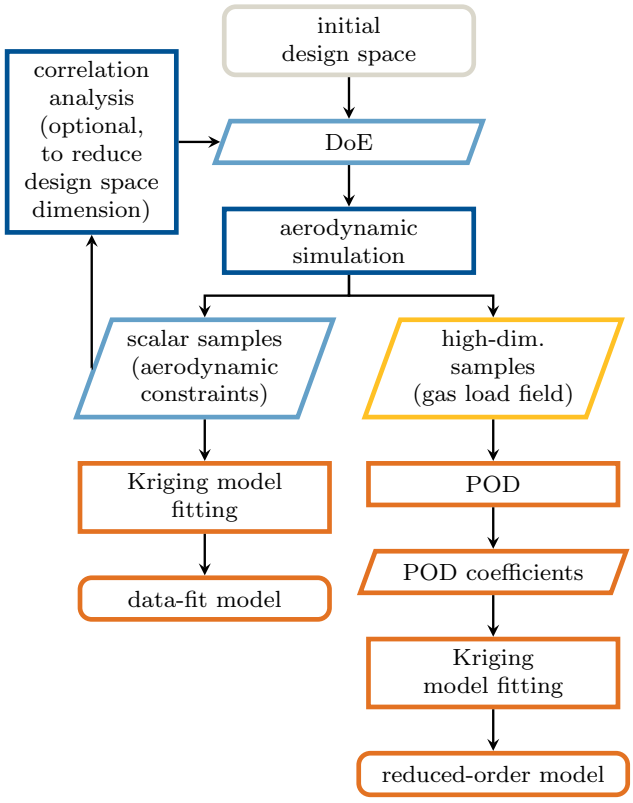


Fig. 4 Flowchart of the low-fidelity model generation process. The high-fidelity aerodynamic model, CFD, is approximated by a data-fit model, Kriging, and a reduced-order model, consisting of POD and Kriging

Based on the resulting data set, a correlation or sensitivity analysis (Saltelli et al. 2007), for instance in form of Spearman correlation coefficients (Spearman 1904), can provide a better understanding of the problem at hand and help to identify variables with negligible influence on the quantities of interest. Omitting these variables during the low-fidelity model generation can allow for more accurate models with respect to the remaining variables, especially in case of limited sampling budget.

The CFD results serve as a common training base for both kinds of low-fidelity models. We use a data-fit model, Kriging, to approximate the scalar aerodynamic constraints. A reduced-order model, consisting of POD and Kriging, approximates the gas load field for aero-structural coupling. Both are explained in what follows.

3.1 Data-fit model: Kriging

Popular data-fit models are polynomial response surface models, moving least-squares, RBF interpolation, Kriging, or support vector regression, summarized by Forrester et al. (2008). We use Kriging (Krige 1951; Matheron 1963; Sacks et al. 1989), because it does not assume a certain problem structure and can thus yield accurate predictions for various function forms.

Kriging is also known as Gaussian process (GP) regression (Rasmussen and Williams 2006). It predicts an unknown output by interpolation between the values in its vicinity. The response is considered as a random variable, contrary to its actual properties as observed response obtained by deterministic computations. This takes into account the actual uncertainty of the prediction.

Since the method is very popular, we only briefly recapitulate the main steps of Kriging model fitting and prediction. To start with, the a-priori GP is defined as sum of mean and covariance function. In ordinary Kriging, used herein, the mean is an unknown constant. The covariance is specified by kernel functions. For modeling physical quantities, the Matérn kernels are in general preferred over the squared exponential kernel, because the latter are unrealistically smooth (Rasmussen and Williams 2006). Then, the model is fitted to the training data by optimizing the hyperparameters, that is the unknown mean and the kernel parameters, commonly using maximum likelihood estimation (MLE). The resulting *a-posteriori* GP can be used to compute the most probable output values and the prediction uncertainty at new input points.

3.2 Reduced-order model: POD + Kriging

While data-fit models approximate scalar quantities, like the efficiency, reduced-order models approximate multi-dimensional outputs, like the gas load field. The underlying idea is to extract the most important features, or modes, of the physical field and then represent the field as a linear combination of these modes. The resulting lower-dimensional system representation is much easier to handle.

The most popular method for obtaining the lower-dimensional representation is POD. It is similar to principal component analysis (PCA) from the field of statistics. Non-intrusive snapshot POD, developed by Sirovich (1987), is employed in this work and will be outlined in what follows. Case studies of snapshot POD in both aerodynamics and structural mechanics were presented by Swischuk et al. (2019). An overview of non-intrusive reduced-order models was provided by Yu et al. (2019).

Following the right branch of Fig. 4, the lower-dimensional outputs, that is the POD coefficients, are approximated by Kriging models, like in the previous subsection. The procedure of reduced-order model generation and prediction is summarized hereinafter. For more details and underlying equations of snapshot POD, we refer to the above references.

The starting point of POD is the snapshot matrix containing n samples of m-dimensional simulation results. This matrix is projected onto the subspace spanned by the dominant modes. Mathematically, this is achieved using thin

singular value decomposition (SVD). The resulting matrices can then be truncated after the first k modes, loosing as little information as possible for a given reduced rank k. The reduced rank k can either be specified in advance or chosen such that the proportion of energy, or variance, captured by the first k modes should be above or equal to a given threshold κ, typically $\kappa \geq 90\%$. In case of computational meshes with thousands of nodes, the energy usually decreases rapidly with increasing rank and $k \ll m$. As lower-dimensional representation of the high-dimensional samples, the POD coefficients are computed by projecting the snapshot vectors onto the first k POD modes.

Afterward, a Kriging model can be trained to approximate the POD coefficients as a function of the original input. Note that the number of degrees of freedom to be modeled decreases from m nodal values to $k \ll m$ coefficients due to the model order reduction.

For the low-fidelity model prediction, the POD coefficients are predicted by the Kriging models. The full-order physical field can then be approximated as a linear combination of POD coefficients and modes.

4 Proposed interdisciplinary optimization process

The goal of this work is a fast and industrially integrable aero-structural optimization approach for the compressor blade design problem stated in Sect. 1.1. For this purpose, we combine the methods explained in the two previous sections. We propose a combination of a simple MDF architecture and a multi-fidelity approach inspired by the CSSO disciplinary subproblems. The resulting interdisciplinary optimization (IDO) process can be categorized as monolithic architecture and is illustrated as a flowchart in Fig. 5.

The underlying idea is the distinction between main and side discipline, leading to an *inter*- rather than *multi*disciplinary optimization problem. It follows the concept of a disciplinary design team, that maintains an interdisciplinary exchange with its neighboring teams, but does not have fully multidisciplinary competences.

The main discipline is the focus of the aforementioned design team and the optimization objective, here structural mechanics to minimize stresses. It is evaluated by a high-fidelity FE model. While the side discipline, aerodynamics, is neither the optimization objective, nor the expertise of the design team, it needs to be considered for a useful optimization result. Therefore, the side discipline is treated as a low-fidelity model, which completely replaces the high-fidelity CFD simulations. It thus generates very little additional computational effort in each optimization iteration. Instead of exchanging large amounts of data, complex CFD models, and simulation tools plus the corresponding licenses, only simple low-fidelity models need to be provided by the aerodynamic design team.

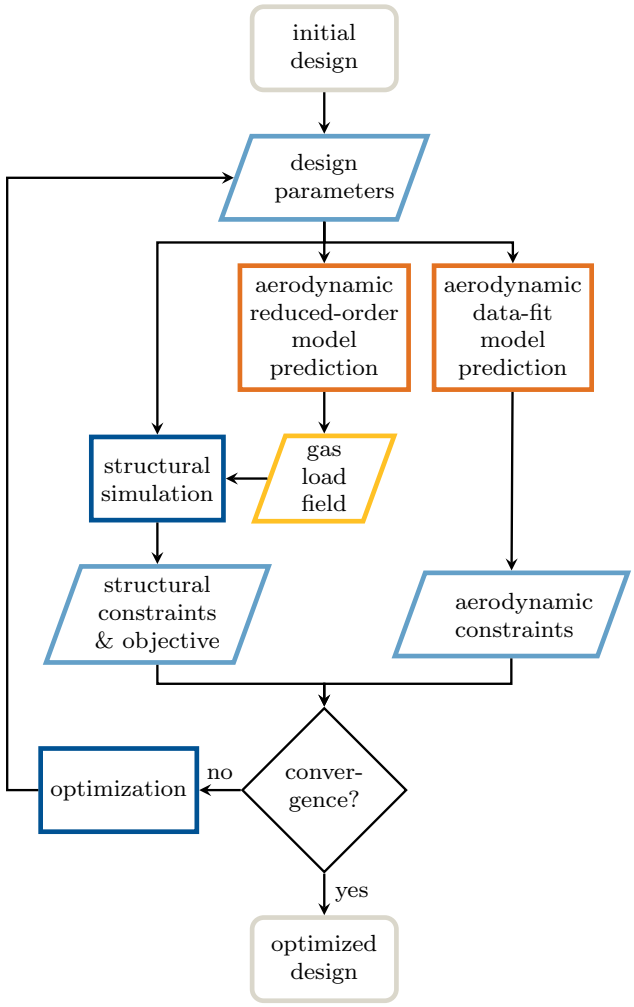


Fig. 5 Flowchart of the proposed multi-fidelity IDO process. The structural mechanics simulations are based on high-fidelity FE models, the aerodynamic simulations are replaced by low-fidelity models

Nevertheless, the aerodynamic influence is considered via both constraints and coupling variables. The constraints are represented by data-fit models, here Kriging, the high-dimensional coupling variables by reduced-order models combining POD and Kriging, see Sect. 3. The effect of the aerodynamics on the structure is represented by unidirectional coupling, transferring the aerodynamic load field as a boundary condition for the static structural analysis. The rapid approximation enables a concurrent coupling, meaning that it is computed during the structural simulation set-up and requires no additional time, contrary to common sequential computations. Cross-coupling, in our case a bidirectional exchange of aerodynamic loads and structural displacements, would make the sampling much more complex and expensive. Approaches for POD and interpolation

(POD + I) model generation for cross-coupled multidisciplinary analyses were presented by Coelho et al. (2009) and Berthelin et al. (2022). Herein, unidirectional coupling via the aerodynamic loads is assumed to yield sufficiently accurate results, because the low-fidelity model is not fully accurate by definition. This also resolves any multidisciplinary analysis convergence issues.

The most widely used approach for aero-structural turbomachinery blade optimization is a MDF architecture without coupling, as employed in most references in Sect. 1.2. The proposed approach differs in that the aerodynamic simulation is replaced by the low-fidelity models (orange building blocks in Fig. 5), enabling a speed up and a coupling via the gas load field (yellow block in Fig. 5). Contrary to common MFO strategies, as presented by Peherstorfer et al. (2018), in the proposed approach the multiple fidelities are not combined within one discipline, but separated between the disciplines. In particular, adaptation is omitted, because it would hinder the interdisciplinary applicability.

To summarize, the main discipline structural mechanics is evaluated as accurately as possible, including updated coupling via gas loads, while the side discipline aerodynamics is computed only as accurately as necessary for a useful optimization result. The IDO is much faster than a complete MDO and better integrable into company structures, because complex high-fidelity CFD models are no longer required in the optimization. Provided sufficiently accurate surrogate models, it fulfills the aim of an industrial MDO, which is an improved design, rather than a mathematical optimum.

5 Compressor problem set-up

To illustrate the proposed approach, it is applied to a concrete compressor rotor blade design problem. More precisely, we seek to optimize the rotor blade geometry of a compressor stage so as to minimize the stresses in the fillet, while fulfilling a number of structural, aerodynamic, and geometrical constraints. The simulation set-up and optimization problem are specified in what follows.

5.1 Computational set-up

Our model represents the third stage of a high-pressure compressor in a next-generation turbofan aircraft engine. It is based on a model from the *Clean Sky 2* (Clean Aviation 2021) project by the European Union.

Geometry parameterization and generation are carried out using *AirFoil Designer pdesk* (atech GmbH 2022) and are explained in Sect. 5.2.

The FE simulations including meshing are performed in the open-source program *CalculiX* (Dhondt 2004). They consist of a static analysis at cruise conditions, and a dynamic analysis at red line conditions to prevent resonance. The structural analyses are carried out for the rotor blades only, which are a critical component in the considered problem. A structural analysis takes on average 6 min on 4 CPUs.

For the CFD simulations, the mesh is generated using *AutoGrid* (Cadence 2022). Then, the unsteady RANS flow solver *TRACE* (German Aerospace Center 2022) is employed, which is specialized on compressor and turbine components. The aerodynamics are computed for cruise conditions, that is at the aero design point (ADP). Here, the entire stage is of interest, to investigate the interaction between rotor and stator. One CFD simulation takes on average 33 min on 8 CPUs.

A fully coupled multidisciplinary analysis would require several sequential iterations of FE and CFD simulations to ensure compatibility of structural displacements and aerodynamic pressures on the blade surface. This would take a prohibitively long time in each optimization iteration. Therefore, the structural displacements are neglected in the CFD model, also because the aerodynamics are only treated as side discipline. The pressure field is obtained from a 2D flow solver, which is only capable of estimating a 1D radial distribution of the total pressure loads. It represents an approximation of the difference between PS and SS pressures and does not account for 3D blade geometry changes. The resulting loads are then projected only onto the blades' PS surface using a heuristic tool to produce the 2D gas load distribution. The gas load field is thus considered constant and only applied to the PS surface unless replaced by the reduced-order model, which is based on 3D CFD samples. Thermal loads are applied in a similar manner.

The computational effort of the CFD is considerably higher than for the FE simulations. This further motivates the use of aerodynamic low-fidelity models in an optimization focusing on structural mechanics. For each design evaluation herein, FE and CFD simulations are run in parallel. Additionally, 20 designs are evaluated in parallel to speed up computations in both sampling and optimization.

The optimizations are run using *AutoOpti* (Siller et al. 2009), a program tailored to the multidisciplinary optimization of turbomachinery components, which can handle high-dimensional non-linearly constrained problems. It is based on an efficient global optimization (Jones et al. 1998) approach, combining evolutionary algorithms, adaptive Kriging surrogate models, and an expected improvement infill criterion. However, our data-fit models for the side discipline are not directly integrated into the optimizer for the sake of modularity. The optimizer is not the focus of this work and can theoretically be exchanged with other gradient-free methods.

5.2 Optimization problem definition

Our optimization problem represents the second step in the industrial blade design process, after the initial aerodynamic design step. We focus on structural mechanics as the main discipline, but also consider aerodynamic aspects, in order to reduce the number of subsequent interdisciplinary iterations. The optimization problem is defined in Table 1.

We seek for a rotor blade geometry, which minimizes the maximum static principal stress $\sigma_{\mathrm{I,\,max}}$ in the fillet, that is the transition between blade hub and disk in a blisk (blade integrated disk), indicated in Fig. 3. Low maximum stresses in the fillet region are crucial for robust compressor designs. Structural constraints are the maximum static principal stresses on the four blade sides PS, SS, LE, and TE, indicated in Fig. 2. Moreover, the first two eigenfrequencies are constrained to prevent resonance. Aerodynamic constraints refer to the mass flow, the isentropic efficiency, and an incidence criterion at five radial positions. The equality constraint for the mass flow, in order to maintain the initial operating point, is implemented as narrow two-sided inequality constraint. In addition, the maximum blade thickness for the upper part of the blade is constrained to avoid extremely thin profiles. This is particularly important to prevent significant changes in the higher frequency torsion modes of the blade, which are not explicitly set as structural constraints. The constriction at eleven radial positions is limited for good manufacturability.

The design variables are rotor blade geometry parameters, as illustrated in Fig. 6.

The free design parameters are the blade angles α, the stagger β_{S}, the wedge angles γ, and the distance wedge l. These parameters define the 2D blade profiles for each of the three radial airfoil sections at hub, mid, and tip by setting the four base points. The profiles are then constructed by spline

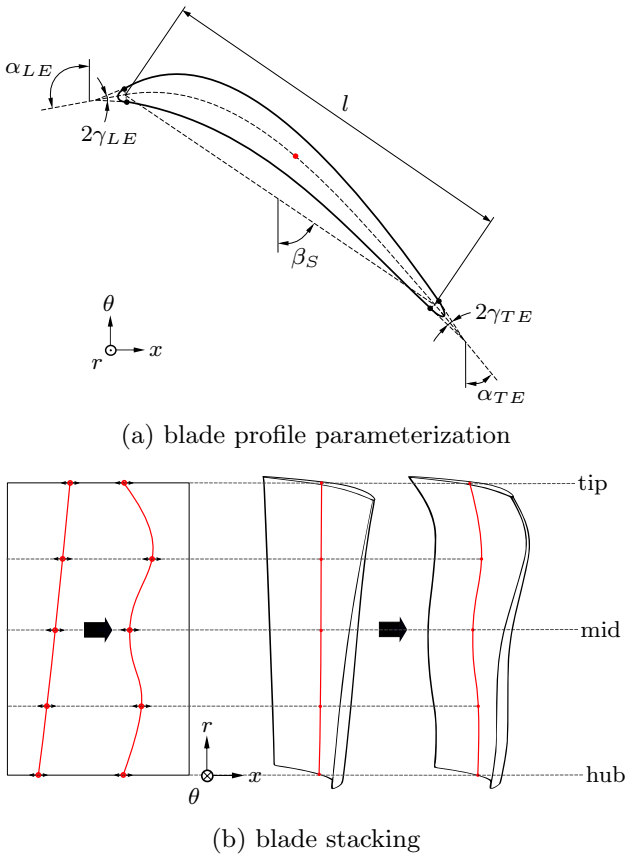


Fig. 6 Blade geometry parameterization. Modified from Arsenyev (2018). The parameters in **a** at blade hub, mid, and tip (see **b**) are the design variables of the optimization problem

Table 1 Interdisciplinary compressor rotor blade optimization problem with main discipline structural mechanics

Objective	Type	Discipline
Max. static principal stress $\sigma_{\mathrm{I,\,max}}$ in the fillet	Minimization	Structural mechanics
Constraints	**Type**	**Discipline**
Max. static principal stress $\sigma_{\mathrm{I,\,max}}$ PS, SS, LE, TE	Upper bound	Structural mechanics
Eigenfrequency f 1, 2	Lower & Upper bound	Structural mechanics
Mass flow $\dot{m}$	Equality constraint (= baseline)	Aerodynamics
Isentropic efficiency η_{is}	Lower bound (= baseline)	Aerodynamics
Incidence criterion slice 1–5	Upper bound (= baseline)	Aerodynamics
Max. blade thickness near tip	Lower bound	Geometry
Constriction slice 1–11	Lower bound	Geometry
Design variables (each for rotor hub, mid, tip)	**Range (relative to baseline)**	**Discipline**
Blade angle α LE, TE	±5%	Geometry
Stagger β_{S}	±4%	Geometry
Wedge angle γ LE, TE	±10%	Geometry
Distance wedge l	±5%	Geometry

interpolation between these base points, which also separate the four sides PS, SS, LE, and TE. The resulting 2D profiles are stacked in radial direction to obtain the 3D blade geometry. The profiles' position in space is defined by the axial and circumferential location of their centers (highlighted in red), which are fixed within the optimization. The blade shape between the predefined profiles is interpolated, for instance by fourth-order splines defined by five radial control points as shown in Fig. 6b. In this work, the three control points at hub, mid, and tip allow for second-order splines to interpolate the design variables. Finally, the transition between blade and disk is smoothed by a fillet with fixed radius. The stator blade geometry is generated analogously and remains constant throughout the optimization, which is common for a structural optimization.

In total, the optimization problem sums up to 26 response variables (one objective plus 25 constrained variables) and 18 design variables.

6 Results

In the following, data-fit models of the aerodynamic constrained variables and reduced-order models of the aerodynamic loads are generated to replace the expensive CFD simulations in the optimization. Multi-fidelity IDO results are validated and compared to purely high-fidelity MDO results as a reference.

6.1 Data-fit model generation

The proposed approach involves data-fit models that predict the aerodynamic constrained variables: the mass flow, the efficiency, and the five incidence criteria. For the Kriging model generation, we follow the flowchart in Fig. 4, taking the left branch. The prediction accuracy is estimated by 10-fold cross-validation. The aim is a NRMSE < 5% (normalized by the respective variable range) and a correlation coefficient $R^2 > 0.9$.

The initial design space is the one used later in the optimization, see Table 1. The sampling points are obtained by LHS. For a reasonable sampling effort, the number of CFD simulations is limited to half of what would be required for a comparable high-fidelity optimization. 300 sampling points are evaluated, of which 11 fail, leaving 289 samples available for the training.

Correlations between design and response variables are analyzed based on Spearman's correlation coefficients and scatter plots. The former are shown in Fig. 15 in the appendix. The results indicate that blade angle and stagger variables correlate strongly with the quantities of interest. Furthermore, wedge angles and distance wedge at hub and tip are only weakly correlated to the aerodynamic constraints. These six design variables are therefore kept constant in the sampling.

The Kriging model implementation relies on the Python package *scikit-learn* (Pedregosa et al. 2011) that in turn builds upon the formulations by Rasmussen and Williams (2006). An isotropic Matérn 5/2 kernel is chosen, as it yields the best fit among various considered kernels.

The distribution of the errors over the response variables is shown by the gray bars in Fig. 7. Although the mean error measures are below the self-imposed limits, the prediction of the isentropic efficiency still poses a challenge.

Based on the response variable distribution in the training data, three significant outliers can be recognized in the set of 289 samples. They can be associated with very unrealistic geometries generated by the automated sampling. Removing these outliers, the errors clearly improve, especially for the isentropic efficiency and the mass flow, see the orange bars in Fig. 7. Consequently, a Kriging model with Matérn 5/2 kernel, based on the training set without outliers, is chosen as low-fidelity model for all aerodynamic constraints.

With 33 min per design evaluation and 20 evaluations in parallel, the sampling runs 8.25 h. Afterward, the aerodynamic data-fit model generation takes only about 1.5 s. Its evaluation takes less than one millisecond, instead of 33 min for a CFD simulation in each optimization iteration.

6.2 Reduced-order model generation

For a concurrent aero-structural coupling in the IDO, we train reduced-order models of the gas load field on the

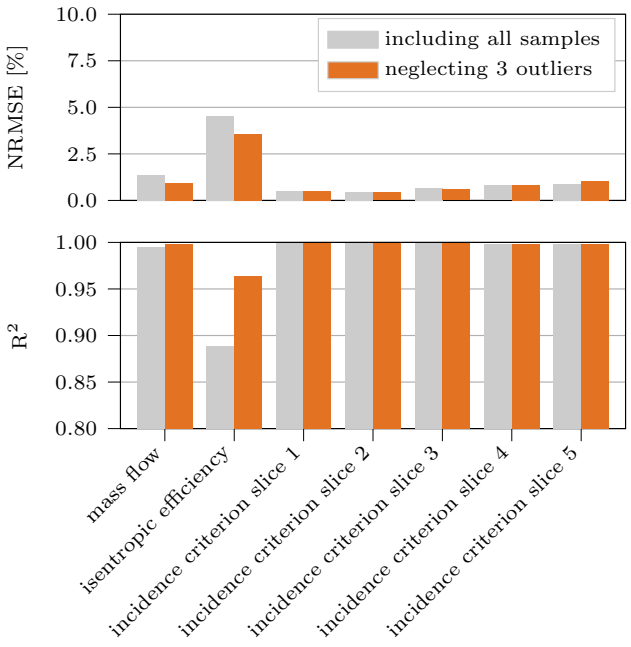


Fig. 7 Error measures for the Kriging data-fit models with and without inclusion of three outliers in the set of 289 samples, based on a 10-fold cross-validation

blade surface, following the right branch of the flowchart in Fig. 4. First, the high-dimensional field is reduced to a low-dimensional representation using POD. Then, the POD coefficients are predicted by a Kriging model. The NRMSE (normalized by the range at the respective node) and R^2 are estimated using a 10-fold cross-validation based on the sampling data.

The training data originate from the same sampling as for the data-fit models and thus requires no additional run time. For each sample, the pressure field on the blade surface is interpolated onto the mesh of the structural simulations. Note that the structural mesh topology remains constant throughout all simulations. This is a prerequisite for a direct construction of the snapshot matrix from the nodal gas load values.

The implementation of the POD uses the SVD algorithm from *scipy* (Virtanen et al. 2020). It is embedded in a self-implemented reduced-order model generation, closely following the theory in Sect. 3.2.

The reduced rank k is computed such that $\kappa = 99\%$ of the energy is conserved. Table 2 holds the resulting reduced ranks, compared to the original full rank, that is the number of degrees of freedom m, equal to the number of nodes. The POD is conducted separately for each of the four blade sides. For pressure and suction side, the rank is reduced by more than two orders of magnitude, for leading and trailing edge by a factor of more than ten. Over the whole blade surface, 4048 nodal values are reduced to 33 POD coefficients to be approximated by the data-fit model.

After the model order reduction, the POD coefficients are computed as output variables for the data-fit model generation. Here, a Kriging model with Matérn 5/2 kernel is used, as in Sect. 6.1, and all 289 samples are considered. The overall low-fidelity model error estimates, due to the reduction plus the data-fit approximation, are shown in the lower part of Table 2. The combination of high loads and a small area make the prediction on the leading edge challenging. At the other three blade sides, the self-imposed accuracy targets (NRMSE $< 5\%$, $R^2 > 0.9$) are achieved, notably also for the highly loaded pressure side.

Based on the above accuracy estimates, the reduced-order models of the gas loads on the blade surface can be considered a large improvement in accuracy compared to the previously used constant gas loads. First, the loads predicted by the reduced-order model are applied to the entire blade surface, instead of the pressure side only. Moreover, they are based on 3D CFD, instead of 2D flow solutions, and hence take into account the blade geometries. As geometry-dependent boundary conditions in the FE analyses, the reduced-order model gas loads enable a weak aero-structural coupling.

Since the reduced-order model is generated based on the same samples as the data-fit models in the previous section, no additional sampling time is required. The reduced-order model generation takes less than 10 s. The gas load field is predicted within milliseconds and the respective input files for the FE models are written in about 0.2 s during the structural simulation set-up. Consequently, the structural analyses in the optimization can start without delay despite the unidirectional coupling.

Table 2 Ranks and error measures for the reduced-order models of the gas load field on the four blade sides

	PS	SS	LE	TE
Full rank m	1932	1932	92	92
Reduced rank k	9	11	7	6
NRMSE (%)	3.10	2.73	6.53	3.62
R^2	0.965	0.977	0.869	0.966

The full rank is equal to the number of degrees of freedom m on the respective side. The reduced rank k is obtained by snapshot POD with $\kappa = 99\%$ energy conservation, based on 289 samples. The error measures of the final model (POD + Kriging) are averaged over the associated nodes and are based on 10-fold cross-validation

6.3 Optimization studies

The proposed multi-fidelity IDO process is compared to a purely high-fidelity MDO approach as a reference, to assess the potential to solve the aero-structural compressor blade optimization problem defined in Sect. 5.2. The two approaches are specified in Table 3.

Both involve high-fidelity structural analysis via 3D FEM, as the structural behavior is the focus of the optimization and thus the main discipline. The aerodynamic constraints are computed either by high-fidelity 3D CFD or the low-fidelity data-fit model. The aero-structural coupling is realized either by a constant gas load field from 2D computations, or by the low-fidelity reduced-order model predictions based on 3D CFD. A coupling directly via high-fidelity 3D CFD gas loads would be prohibitively expensive.

Since the employed evolutionary optimization algorithm is non-deterministic, both approaches are repeated five times with the same low- and high-fidelity models. 20 individuals are evaluated in parallel for 30 optimization iterations,

Table 3 Model fidelities in the two compared optimization approaches

	Structural mechanics	Aerodynamics	Coupling via gas loads
MDO	High-fidelity	High-fidelity	Constant
IDO	High-fidelity	Low-fidelity	Low-fidelity

MDO is the high-fidelity reference multidisciplinary optimization, IDO is the proposed multi-fidelity interdisciplinary approach

which makes 600 design evaluations in total. Figure 8 shows the development of the objective function values over the number of iterations.

Both optimization approaches converge quickly and most runs achieve a stress reduction by about 40% after 20 iterations, compared to the initial blade from the predesign step. Afterward, there is only little improvement. The differences in aerodynamic model fidelity and type of gas loads do not seem to considerably affect the convergence behavior.

The optimizer itself requires a similar amount of time for both high- and multi-fidelity optimizations. Consequently, the aero-structural analysis time, illustrated in Fig. 9, determines the difference in overall computation time.

Considering 30 iterations with 33 min each, the analysis time can be reduced by more than 80% from 16.5 h to only 3 h due to the low-fidelity models in the IDO. Additionally, the 8 CPUs for the CFD simulations are no longer necessary. For a holistic consideration, the 3 h of simulation time in the multi-fidelity IDO must be offset by the sampling time, adding up to 11.25 h. Note that multiple optimizations with varying objective functions and constraints can be run with a single sampling, reducing the sampling time per optimization. Moreover, the sampling can be arbitrarily parallelized, depending on the computational resources.

Figure 10 shows the aerodynamic responses for the IDO results.

The low-fidelity data-fit model results are indicated in orange. They satisfy all optimization constraints, that is they lie inside the light gray ranges, as expected for a feasible optimization result. For validation, the low-fidelity results are recomputed by high-fidelity CFD simulations, indicated in black. They show that all aerodynamic inequality constraints are also satisfied for the high-fidelity model. The

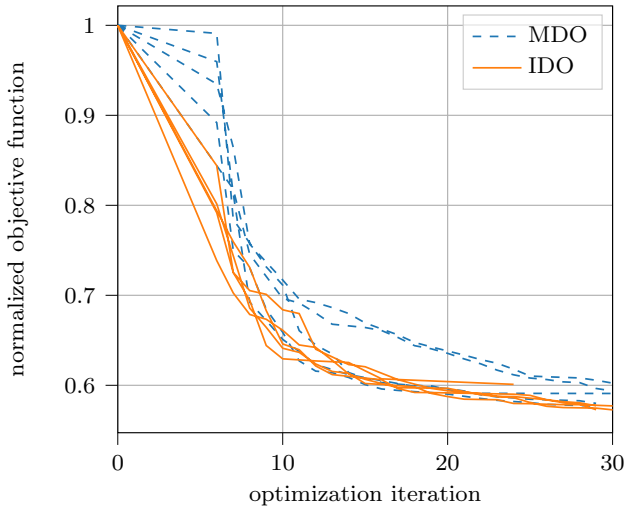


Fig. 8 Convergence of the objective function, normalized with its initial value, as a function of the optimization iteration. Only feasible results with improved objective values are plotted

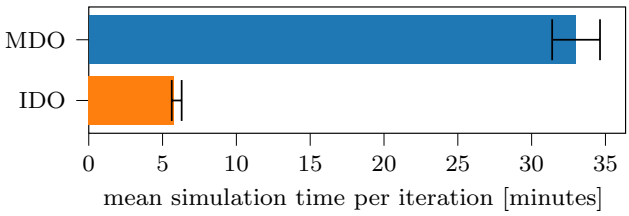


Fig. 9 Mean simulation time per optimization iteration for the two different optimization approaches. The whiskers show minimum and maximum values of the five repetitions

relaxed equality constraint for the mass flow with very narrow bounds is slightly violated. The clear trends in the deviations between low- and high-fidelity model predictions can be explained by all runs converging to a similar region in the design space, see Fig. 12. The mean, minimum, and maximum errors are shown in Table 4 in the appendix. In summary, all deviations are within reasonable limits and do not undermine the validity of the overall optimization results.

In addition to the aerodynamic result variables, the gas loads on the rotor blade surface are predicted in the IDO approach. As an example of the prediction quality, Fig. 11 compares the CFD result and the reduced-order model prediction of the gas load field on the pressure side for an optimized blade design.

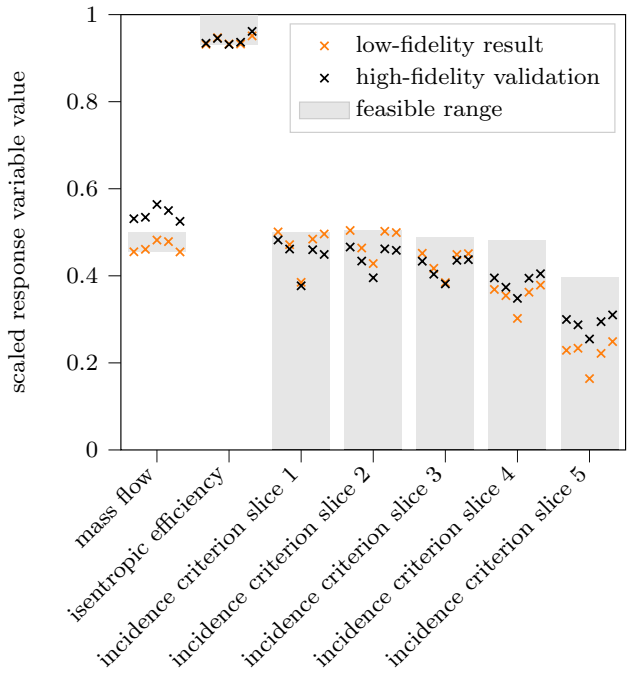


Fig. 10 Aerodynamic response variable values for the multi-fidelity IDO results, scaled with the respective variable ranges from the sampling. The low-fidelity data-fit model results are indicated in orange, the high-fidelity CFD validations in black. The feasible ranges, that is where the optimization constraints are satisfied, are marked in light gray

Comparing the high- and low-fidelity results in Fig. 11a and b, distributions are in good agreement. The normalized error in Fig. 11c indicates that the prediction is very accurate on the majority of the blade's pressure side. The error only increases in the area of high gradients near the leading edge. The mean, minimum, and maximum errors are shown in Table 5 in the appendix. Overall, the gas load prediction can be considered a large improvement in accuracy compared to the constant loads from the 2D flow simulations. The computational effort for model generation and evaluation is negligible, as it exploits existing samples from the scalar aerodynamic response predictions.

The optimized design variables of the two approaches are shown in the parallel coordinates plot in Fig. 12.

Both yield the same trends compared to the initial design. The variation among the results of each approach is larger than the variations between the approaches. Moreover, the gas load accuracy improvement in the IDO does also have no visible effect on the optimization results. Its impact may be weakened by dominating centrifugal loads in the present case. One can conclude that the multi-fidelity approach is accurate enough to provide design proposals for industrial applications, where optimizations are stopped when improvements become minor, like between 20 and 30 iterations in the convergence plot in Fig. 8.

Figure 13 shows an optimized blade design from the proposed IDO approach in orange, in comparison to the initial geometry in gray. Starting at the blade hub, the profile becomes considerably thicker, for higher stiffness in the fillet region, which in turn reduces the stresses and thus the objective function. This is realized by an increase in the wedge angles at both LE and TE hub, that also show a strong correlation to the fillet stresses, see Figs. 12 and 15. Continuing toward the blade tip, curvature and chord length were significantly changed, which is likely to affect higher order eigenmodes and eigenfrequencies. They should thus be included for an improved problem formulation in future optimizations.

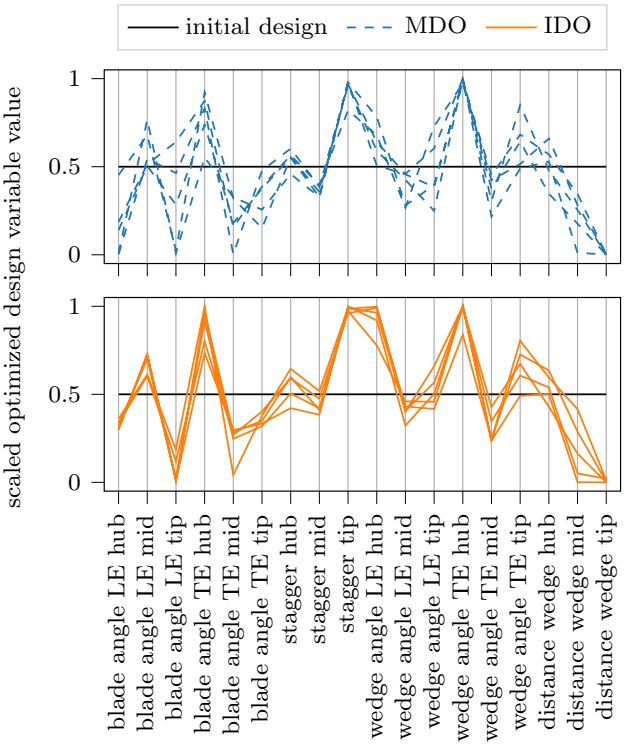
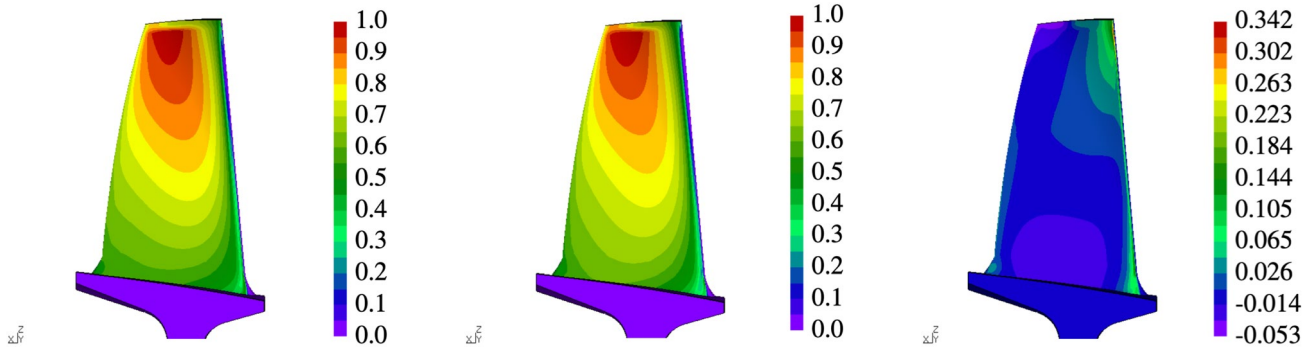


Fig. 12 Parallel coordinates plot comparing the results of the two optimization approaches to the initial design. The design variable values are scaled with their respective range

The effect of the design changes can be observed in the stress field, notably the maximum static principal stresses. The maxima occur on the pressure side, which is shown in Fig. 14. We compare again the initial and optimized blade design from the proposed IDO approach. The maximum is shifted from the fillet to a second peak in the middle, to relieve the fillet region without exceeding the upper bound set as constraint in the optimization problem.



(a) high-fidelity CFD result (b) reduced-order model prediction (c) error = (a) - (b)

Fig. 11 Gas load field on the pressure side of an optimized rotor blade from the proposed IDO approach, scaled (a,b), and normalized (c) with the range from the high-fidelity CFD result

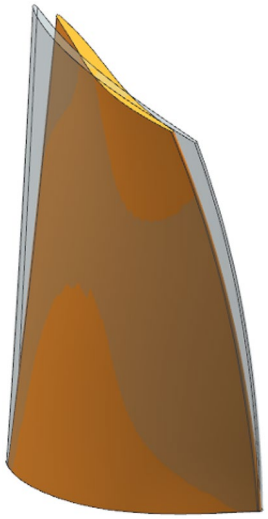


Fig. 13 Optimized rotor blade design from the proposed IDO approach in orange. For comparison, the initial design is depicted in semitransparent gray

7 Conclusions

An interdisciplinary approach for the optimization of compressor blade geometries so as to minimize structural stresses without impairing the aerodynamic performance is proposed in this work. We consider structural mechanics as main discipline, aerodynamics as side discipline, and static aero-structural coupling. Combining high- and low-fidelity models, we imitate the interdisciplinary workflow of design teams in industry. The main discipline is the objective and focus of the optimization and therefore evaluated by high-fidelity FE models. The side discipline computations inside the optimization are entirely replaced by low-fidelity model evaluations with almost negligible computation time and model complexity compared to high-fidelity CFD. The scalar aerodynamic constraints are approximated by Kriging models. The gas load field is predicted using POD combined with Kriging. The latter enables a unidirectional concurrent aero-structural coupling at no additional time in each optimization iteration and thereby enhances the high-fidelity structural analysis accuracy.

Consequently, aerodynamic effects of design changes can be considered at very low cost inside the optimization. Moreover, the required exchange of data and expertise between disciplinary design teams is greatly reduced, facilitating the organizational integration compared to common multidisciplinary optimization approaches.

The proposed approach is illustrated by a structural compressor blade optimization with 25 structural, aerodynamic, and geometric constraints and 18 design variables. The results show that sufficiently accurate low-fidelity models (NRMSE $< 5\%$, $R^2 > 0.95$) can be obtained at reasonable cost (300 sampling points, as opposed to 600 design evaluations in each optimization). The multi-fidelity optimizations yield valid designs. Design evaluation times in the optimization are reduced by more than 80% compared to the high-fidelity reference. This shows the great potential to reduce computation time, especially for expensive side disciplines like aerodynamics.

Based on the above results, a number of interesting directions for future research and developments can be discussed. For different problems, for example if the aerodynamic design team sets the isentropic efficiency as objective function, main and side disciplines could be reversed. Since high-fidelity aerodynamic evaluations are more expensive than structural ones, low-fidelity structural constraints would not reduce the overall simulation times for parallel disciplinary evaluations. Here, additional adaptive low-fidelity models for the main discipline may be useful. For unidirectional coupling variables, however, the low-fidelity models would still enable a concurrent coupling. The benefit in terms of organizational integration remains, independent of

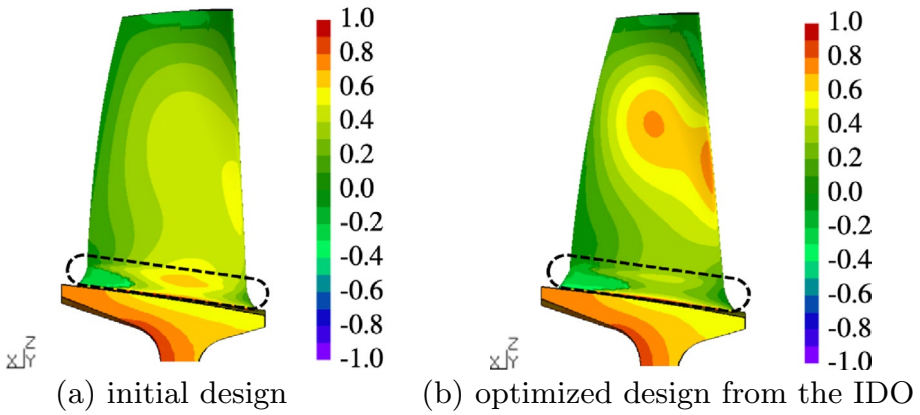


Fig. 14 Maximum principal stress field on the rotor blade pressure side, normalized with the maximum allowed value defined in the optimization problem. In the fillet region, indicated by the dashed line, the stress peak is minimized. On the rest of the blade, the stresses stay below the maximum allowed value

the computational effort. Besides, the proposed IDO approach is readily extendable to additional side disciplines, for example thermodynamics in the compressor blade design. With regard to the organizational structure of most companies, more than one main discipline does usually not make sense. Finally, the potential of the proposed approach should be confirmed for even larger problems, for example by a multi-stage compressor blade optimization. The problem size may then require not only to decouple disciplines, but also split the design domain, for instance into single stages. Here, reduced-order models have great potential to predict physical interfaces.

Appendix

See Fig. 15 and Tables 4 and 5.

Table 5 Mean, minimum, and maximum errors of the reduced-order model gas load predictions for the optimized rotor blades from the multi-fidelity IDO, normalized by the respective nodal range from the sampling

Error (%)	PS	SS	LE	TE
Mean	7.88	5.27	7.93	12.11
Min	7.26	4.78	4.88	10.75
Max	8.84	5.65	17.53	13.97

Acknowledgements The authors would like to thank the *MTU Aero Engines* and *Technical University of Munich* for their support in this work.

Author Contributions Conceptualization: LP; Methodology: LP; Software: LP; Formal analysis and investigation: LP, IA; Writing—original draft preparation: LP; Writing—review and editing: LP, IA, CC, FD; Supervision: IA, FD.

Funding Open Access funding enabled and organized by Projekt DEAL. This project is funded by the *Bayerische*

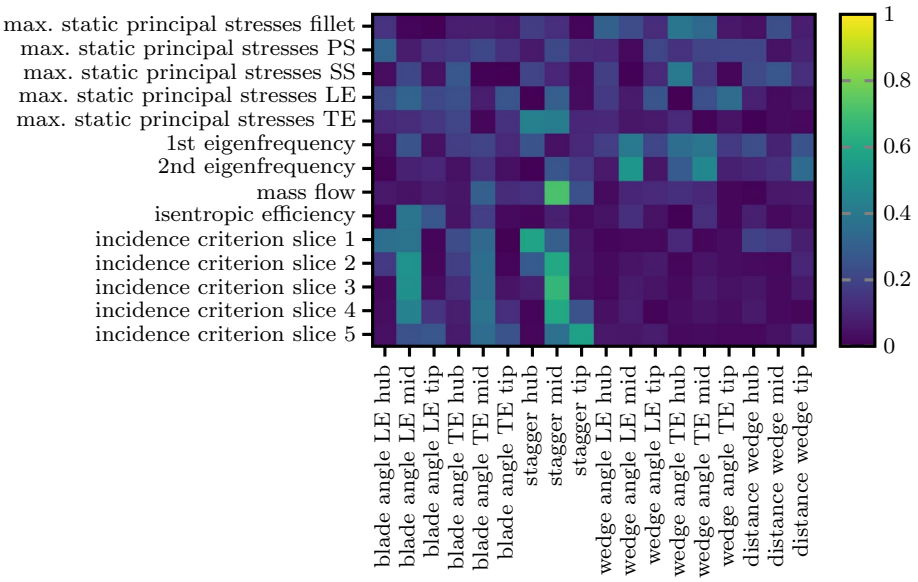


Fig. 15 Spearman's correlation coefficients for the design and response variables of the optimization problem defined in Table 1. The coefficients indicate a relevant monotonic relationship of the structural responses to all design variables. The aerodynamic responses show a strong correlation with blade angles and stagger in the middle of the blade, but only weakly relate to wedge angles and distance wedge

Table 4 Mean, minimum, and maximum errors of the data-fit model predictions for the optimized rotor blades from the multi-fidelity IDO, normalized by the respective variable range from the sampling

Error (%)	Mass flow	Isentropic efficiency	Incidence criterion slice 1	Incidence criterion slice 2	Incidence criterion slice 3	Incidence criterion slice 4	Incidence criterion slice 5
Mean	7.46	0.40	2.16	3.64	1.24	3.00	7.00
Min	7.03	0.12	0.80	3.00	0.31	1.92	5.38
Max	8.17	1.03	4.70	4.08	1.83	4.60	9.10

Luftfahrtforschung- und technologieförderung of the *Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie (StMWi)*.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Replication of results The implementation of the low-fidelity modeling methods is described in detail and the associated Python code can be obtained from the corresponding author on reasonable request. For the purpose of a realistic industrial application, the compressor blade optimization was performed in the design environment of *MTU Aero Engines* and is therefore unfortunately subject to strict confidentiality. Different simulation models and software may produce different results, but should yield similar trends.

References

Aissa MH, Verstraete T (2019) Metamodel-assisted multidisciplinary design optimization of a radial compressor. Int J Turbomach Propuls Power 4(4):35. https://doi.org/10.3390/ijtpp4040035

Arsenyev I (2018) Efficient surrogate-based robust design optimization method. PhD thesis, Technische Universität München, Munich, Germany

atech GmbH (2022) Airfoil designer pdesk. http://www.atech.de/produkte/airfoil-designer-pdesk/. Accessed 21 Mar 2022

Benamara T, Breitkopf P, Lepot I, Sainvitu C (2017) LPC blade and non-axisymmetric hub profiling optimization using multi-fidelity non-intrusive POD surrogates. In: Proceedings of the ASME Turbo Expo, vol 2C-2017. American Society of Mechanical Engineers. https://doi.org/10.1115/gt2017-65106

Berthelin G, Dubreuil S, Salaün M, Bartoli N, Gogu C (2022) Disciplinary proper orthogonal decomposition and interpolation for the resolution of parameterized multidisciplinary analysis. Int J Numer Methods Eng. https://doi.org/10.1002/nme.6981

Cadence (2022) Autogrid. https://www.numeca.com/product/omnis-autogrid. Accessed 21 Mar 2022

Clean Aviation JU (2021) Clean sky 2. https://www.clean-aviation.eu/clean-sky-2. Accessed 21 Mar 2022

Coelho RF, Breitkopf P, Knopf-Lenoir C, Villon P (2009) Bi-level model reduction for coupled problems. Struct Multidisc Optim 39(4):401–418. https://doi.org/10.1007/s00158-008-0335-3

Cuciumita C, John A, Qin N, Shahpar S (2021) Structurally constrained aerodynamic adjoint optimisation of highly loaded compressor blades. In: Proceedings of the ASME Turbo Expo, vol 2D-2021. American Society of Mechanical Engineers. https://doi.org/10.1115/gt2021-59717

Dhondt G (2004) The Finite Element Method for Three-Dimensional Thermomechanical Applications. Wiley. https://doi.org/10.1002/0470021217

European Commission, Directorate-General for Mobility and Transport, Directorate-General for Research and Innovation (2011) Flightpath 2050: Europe's vision for aviation: maintaining global leadership and serving society's needs. Publications Office of the European Union. https://doi.org/10.2777/50266

Forrester AIJ, Sóbester A, Keane AJ (2008) Engineering design via surrogate modelling. Wiley. https://doi.org/10.1002/9780470770801

German Aerospace Center (DLR) (2022) Trace. http://www.trace-portal.de/userguide/trace/index.html. Accessed 21 Mar 2022

Ghalandari M, Ziamolki A, Mosavi A, Shamshirband S, Chau KW, Bornassi S (2019) Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments. Eng Appl Comput Fluid Mech 13(1):892–904. https://doi.org/10.1080/19942060.2019.1649196

Giunta A, Wojtkiewicz S, Eldred M (2003) Overview of modern design of experiments methods for computational simulations (invited). In: 41st aerospace sciences meeting and exhibit. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2003-649

Gray JS, Moore KT, Hearn TA, Naylor BA (2013) Standard platform for benchmarking multidisciplinary design analysis and optimization architectures. AIAA J 51(10):2380–2394. https://doi.org/10.2514/1.J052160

He P, Mader CA, Martins JRRA, Maki KJ (2020) DAFoam: an open-source adjoint framework for multidisciplinary design optimization with OpenFOAM. AIAA J 58(3):1304–1319. https://doi.org/10.2514/1.j058853

Hu D, Mao J, Wang R, Jia Z, Song J (2016) Optimization strategy for a shrouded turbine blade using variable-complexity modeling methodology. AIAA J 54(9):2808–2818. https://doi.org/10.2514/1.j054742

Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. J Global Optim 13(4):455–492. https://doi.org/10.1023/a:1008306431147

Khatouri H, Benamara T, Breitkopf P, Demange J (2022) Metamodeling techniques for CPU-intensive simulation-based design optimization: a survey. Adv Model Simul Eng Sci. https://doi.org/10.1186/s40323-022-00214-y

Krige DG (1951) A statistical approach to some basic mine valuation problems on the witwatersrand. J S Afr Inst Min Metall 52:201–203

Lian Y, Liou MS (2006) Aerostructural optimization of a transonic compressor rotor. J Propul Power 22(4):880–888. https://doi.org/10.2514/1.15397

Martins JRRA, Lambe AB (2013) Multidisciplinary design optimization: a survey of architectures. AIAA J 51(9):2049–2075. https://doi.org/10.2514/1.j051895

Martins JRRA, Ning A (2021) Engineering design optimization. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781108980647

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266

Papageorgiou A, Tarkian M, Amadori K, Ölvander J (2018) Multidisciplinary design optimization of aerial vehicles: a review of recent advancements. Int J Aerosp Eng 2018:1–21. https://doi.org/10.1155/2018/4258020

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Peherstorfer B, Willcox K, Gunzburger M (2018) Survey of multifidelity methods in uncertainty propagation, inference, and optimization. SIAM Rev 60(3):550–591. https://doi.org/10.1137/16m1082469

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT Press, Cambridge. https://gaussianprocess.org/gpml/

Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. Stat Sci 4(4). https://doi.org/10.1214/ss/1177012413

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2007) Global sensitivity analysis. Wiley, The Primer. https://doi.org/10.1002/9780470725184

Sellar R, Batill S, Renaud J (1996) Response surface based, concurrent subspace optimization for multidisciplinary system design. In: 34th Aerospace Sciences Meeting and Exhibit. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.1996-714

Siller U, Voß C, Nicke E (2009) Automated multidisciplinary optimization of a transonic axial compressor. In: 47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2009-863

Sirovich L (1987) Turbulence and the dynamics of coherent structures. II. Symmetries and transformations. Q Appl Math 45(3):573–582. https://doi.org/10.1090/qam/910463

Sobieszczanski-Sobieski J (1989) Optimization by decomposition: a step from hierarchic to non-hierarchic systems. Tech. rep, NASA Langley Research Center

Sobieszczanski-Sobieski J, Haftka RT (1997) Multidisciplinary aerospace design optimization: survey of recent developments. Struct Optim 14(1):1–23. https://doi.org/10.1007/bf01197554

Spearman C (1904) The proof and measurement of association between two things. Am J Psychol 15(1):72–101. https://doi.org/10.2307/1412159

Swischuk R, Mainini L, Peherstorfer B, Willcox K (2019) Projection-based model reduction: formulations for physics-based machine learning. Comput Fluids 179:704–717. https://doi.org/10.1016/j.compfluid.2018.07.021

Tedford NP, Martins JRRA (2009) Benchmarking multidisciplinary design optimization algorithms. Optim Eng 11(1):159–183. https://doi.org/10.1007/s11081-009-9082-6

Vanti F, Pinelli L, Arnone A, Schneider A, Astrua P, Puppo E (2018) Aeroelastic optimization of an industrial compressor rotor blade geometry. In: Volume 2D: Turbomachinery. American Society of Mechanical Engineers, https://doi.org/10.1115/gt2018-76474

Viana FAC, Simpson TW, Balabanov V, Toropov V (2014) Special section on multidisciplinary design optimization: metamodeling in multidisciplinary design optimization: how far have we really come? AIAA J 52(4):670–690. https://doi.org/10.2514/1.j052375

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy 1.0 Contributors (2020) SciPy 1.0: fundamental algorithms for scientific computing in python. Nat Methods 17(3):261–272. https://doi.org/10.1038/s41592-019-0686-2

Yu J, Yan C, Guo M (2019) Non-intrusive reduced-order modeling for fluid problems: a brief review. Proc Inst Mech Eng G 233(16):5896–5912. https://doi.org/10.1177/0954410019890721

Zhang L, Mi D, Yan C, Tang F (2018) Multidisciplinary design optimization for a centrifugal compressor based on proper orthogonal decomposition and an adaptive sampling method. Appl Sci 8(12):2608. https://doi.org/10.3390/app8122608

A.2 Publication II

L. Pretsch, I. Arsenyev, E. Raponi, and F. Duddeck. “Twofold Adaptive Design Space Reduction for Constrained Bayesian Optimization of Transonic Compressor Blades”. In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2024. DOI: 10.1115/gt2024-121848

Proceedings of ASME Turbo Expo 2024
Turbomachinery Technical Conference and Exposition
GT2024
June 24-28, 2024, London, United Kingdom

GT2024-121848

TWOFOLD ADAPTIVE DESIGN SPACE REDUCTION FOR CONSTRAINED BAYESIAN OPTIMIZATION OF TRANSONIC COMPRESSOR BLADES

Lisa Pretsch[1,*], Ilya Arsenyev[2], Elena Raponi[3], Fabian Duddeck[1]

[1]Technical University of Munich, Munich, Germany
[2]MTU Aero Engines AG, Munich, Germany
[3]Leiden University, Leiden, Netherlands

ABSTRACT

As turbomachinery designs become more complex, shape optimization tasks involve computationally expensive simulations and many constraints and design variables. Bayesian optimization (BO) is a class of adaptive surrogate-based methods for global optimization. It can efficiently utilize a small budget of high-fidelity evaluations and handle a large number of constraints. However, it suffers from a hampered convergence rate for problems with a large number of design variables. Adaptive design space reductions via principal component analysis (PCA) and trust region (TR) approaches have been shown to improve the scalability in different phases of the optimization. Extending existing methods, we implement the ability to benefit from parallel high-fidelity evaluations and to handle constraints. We then sequentially combine our PCA-BO and TR-BO in a hybrid method to profit from their respective complexity reduction strategies. We assess the performance of our hybrid algorithm by comparing it to vanilla BO on two problems: a 40D analytical test function and a 55D aerodynamic compressor blade design. The empirical results show that PCA-BO enhances the convergence rate in the initial optimization phase, while TR-BO allows for further improvements in the later iterations. Moreover, the algorithm computation time is more than halved. For the compressor blade case, our approach yields an equally good design as vanilla BO after only 20 instead of 96 iterations. The proposed approach has the potential to extend the good performance of BO to even higher-dimensional constrained problems, including multi-stage turbomachinery optimizations.

Keywords: Bayesian optimization, Efficient global optimization, Design space reduction, Dimensionality reduction, Principal component analysis, Trust region, Shape optimization, High-pressure compressor

*Corresponding author: lisa.pretsch@tum.de

NOMENCLATURE

f	objective function
h	constraint
m	number of constraints
d	number of design variables
r	reduced number of design variables
w	weight for rescaling in PCA-BO
$\boldsymbol{x}$	vector of design variables
$\boldsymbol{z}$	vector of reduced design variables
n	number of sample points
q	number of parallel infill points
l	trust region box side length
τ	consecutive iteration threshold
σ	standard deviation
Φ	cumulative distribution function
ϕ	probability density function

1. INTRODUCTION

Motivated by the vision of climate neutral aviation by 2050 [1, 2], novel aircraft engine concepts are needed in a short time. Whether it is entirely new (hybrid-)electric systems or the next generation of gas turbines, the vision's realization depends on fast development in every detail. This can be supported by design optimization methods.

As simulation capabilities evolve, turbomachinery designs become more complex and shape optimizations contain more design variables and constraints. If robust adjoint solvers are available, gradient-based local optimization methods can be employed to tackle this challenge. Their computational cost is almost independent of the number of design variables. However, they scale poorly with a high number of constraints (and objectives) to be evaluated separately. Gradient-free global methods generally perform well for many constraints [3].

This work focuses on gradient-free optimization by means of Bayesian optimization (BO) [4]. BO became popular in engineering under the name efficient global optimization (EGO) [5, 6]. It

is a class of adaptive surrogate-based methods, which can efficiently compute high-quality solutions for optimization problems with a small budget of high-fidelity evaluations. This is beneficial when dealing with the generally non-convex problems and expensive simulations in turbomachinery design, where BO is frequently applied [7–9]. However, high-dimensional design spaces pose a challenge for standard BO methods. Problem complexity increases exponentially with the input dimension, leading to the so-called curse of dimensionality.

One of the most commonly employed strategies is to reduce the problem dimensionality. Song and Keane [10] used parameter screening in a Kriging-based aerodynamic engine nacelle optimization. Even if the simple approach of fixing less important design variables to their initial value may reduce the computational cost of the optimization, the cost of the preceding screening also increases exponentially with the input dimension. Moreover, the approach is less flexible than most more recently applied methods.

Lopez et al. [11] combined active subspaces [12] with a global optimization algorithm, which relies on artificial neural networks. The approach was compared to an adjoint method, yielding similar results at reduced computational cost for a 35-dimensional jet engine fan blade optimization.

Hartwig and Bestle [13] used partial least squares (PLS) in a Kriging-based 57-dimensional aero-structural optimization of a gas turbine compressor blade. First, PLS was used to reduce the internal dimensionality of the Kriging model, as in [14]. Second, PLS was used to transform but not reduce the design space for the optimization.

Zhang et al. [15] employed proper-orthogonal decomposition (POD) for dimensionality reduction prior to an aero engine compressor optimization. The 30 geometric design variables could be replaced by 4 POD coefficients in the optimization. This reduced problem was simpler and faster to be optimized, however at the cost of accuracy as the POD was not adapted during the optimization. POD is also known as principal component analysis (PCA), as it will be called hereinafter.

PCA was also applied for geometric filtration and dimensionality reduction in airfoil shape optimization in various ways [16–19]. The underlying idea is that a set of geometries with good aerodynamic performance has common geometric features, which can be extracted by PCA. The benefit of a PCA-based parameterization is threefold: it filters out "bad" designs, reduces the number of optimization parameters, and yields a well-posed optimization problem with orthogonal design variables.

Trust region methods are another approach to reduce the design space in large-scale surrogate-based optimizations. The above techniques aim at reducing the dimensionality of the problem, while keeping the original ranges for the optimization. Here, the dimensionality of the problem is kept constant and the parameter variation ranges are iteratively reduced. Trust region methods rely on the assumption that the current best point gets closer and closer to the optimum as the optimization evolves. Thus, their nature is more exploitative than explorative. Schlaps et al. [20] showed that their combination of surrogate-based optimization with a trust region method works well for a highly constrained compressor vane optimization with 48 design parameters.

In combination with BO, the above methods were rarely applied in turbomachinery (and airfoil) optimization. None of the above outlined PCA-based approaches adapt the PCA during the optimization. In contrast, the good performance of BO originates especially from the adaptive enhancement of the Kriging surrogate models.

In this paper, we encapsulate both the concepts of dimensionality reduction and trust regions under the term design space reduction. We present a twofold adaptive design space reduction approach for constrained BO. The resulting approach should improve the performance of BO for high-dimensional optimization problems with multiple constraints, particularly in industrial turbomachinery applications.

Our paper contains three methodological novelties: First, we sequentially combine dimensionality reduction and trust region approaches, namely PCA-BO [21] and a trust region-based BO [22, 23], to complement their individual strengths. Second, we adapt the two existing methods to handle multiple, potentially severely restrictive, constraints. The third novelty is a method extension to generate and evaluate several designs in parallel, considering the availability of high-performance computing environments and the importance of wall time for the industrial design process. These modifications make the adaptive design space reduction approaches usable for large-scale industrial engineering design problems in the first place.

We analyze the performance of the proposed approach on an analytical test function and a realistic turbomachinery design problem. The first is the multimodal Rastrigin function minimization subject to 4 constraints by varying 40 design variables, taking up previous studies for the original methods. The latter is a detailed transonic blade design of the second stage of the high-pressure compressor (HPC) in a modern high-bypass geared turbofan engine. We seek for a blade design that maximizes the efficiency subject to 20 aerodynamic and geometric constraints by varying 55 blade geometry parameters. This high-dimensional and severely constrained optimization confirms the proposed method's adaptability to complex scenarios.

The remainder of the article is structured as follows: Section 2 summarizes the basics of constrained and high-dimensional BO as reference. The subsequent Section 3 presents our proposed method, split into the two parts PCA-BO and TR-BO. Its performance is tested on the analytical Rastrigin function in Section 4. The aerodynamic blade design problem is defined in Section 5, followed by the corresponding optimization results in Section 6. The paper is concluded by a summary and outlook in Section 7.

2. CONSTRAINED HIGH-DIMENSIONAL BAYESIAN OPTIMIZATION

The optimization problem is posed in the form

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) && \\ &\text{subject to} && h_j(\boldsymbol{x}) \leq 0 && j = 1, \ldots, m \\ &\text{by varying} && x_{i,\text{lb}} \leq x_i \leq x_{i,\text{ub}} && i = 1, \ldots, d. \end{aligned} \tag{1}$$

The goal is to minimize a single objective function $f(\boldsymbol{x})$, subject to m inequality constraints $h_j(\boldsymbol{x})$, by varying a vector $\boldsymbol{x}$ of d design variables within their respective lower and upper bounds.

2.1 Constrained BO

The basic BO procedure can be understood by the flowchart in Figure 1, omitting the parts highlighted in color. All steps with their employed methods and settings are specified below. Our constrained BO approach is built upon the implementation described by Arsenyev et al. [24], which can be consulted for more details. For a broader understanding of surrogate-based optimization in general, we additionally point the reader to the reviews by Forrester and Keane [25] and, specifically for BO, Garnett [26].

DoE. First, a design of experiments (DoE) provides points to be evaluated by the high-fidelity model, forming an initial sample set. We choose an optimized Latin hypercube sampling (LHS) [27] to obtain a more uniform point distribution over the design space.

High-fidelity evaluation. Second, the high-fidelity evaluations are performed. In our blade design case, these are geometric and aerodynamic simulations as described in Section 5. Points with failed simulations get no special treatment and are omitted in the training data set for lack of evaluation results. We do not employ a dedicated failure handling [28], because the constraints are chosen so as to guide the optimization towards successful design space regions.

Best design update. If feasible points are available, the current best design is the feasible one with the best objective function value. Otherwise, the current best design is the one with the smallest sum of absolute constraint violations. For a well-balanced impact, all constraint values are scaled to zero mean and unit variance beforehand.

GP model fit. The samples are then used to fit a Gaussian process (GP), or Kriging, low-fidelity model. The resulting posterior distribution returns the mean, that is the predicted output values, and the prediction uncertainty for arbitrary points in the design space. The GP prediction cost is negligible compared to the one of the high-fidelity evaluations. Herein, the GP surrogates are generated using the surrogate modeling toolbox [29]. The default settings are kept except for a Matérn $3/2$ kernel instead of a squared exponential kernel. The latter is unrealistically smooth for modeling physical quantities [30], and the GP models with the Matérn $3/2$ kernel showed the highest accuracy among various tested kernels for our problem. The anisotropic kernel in combination with the high-dimensional input and output variables and an increasingly high number of training points makes the GP model's hyperparameter fit time consuming. Therefore, the hyperparameters are optimized for all output variables at once, maximizing the sum of likelihood function values. This can be considered sufficiently accurate in view of the time savings for the large number of constraints.

Infill criterion optimization. Based on this probabilistic surrogate model, an infill criterion, or acquisition function, is optimized. The aim is to find new infill points which either exploit the model to converge to the optimal design, or explore the design space to reduce the prediction uncertainty. A popular infill criterion for unconstrained problems is the expected improvement EI. With constraints, a common criterion, which we also use herein, is the product of EI and the individual probabilities of feasibility

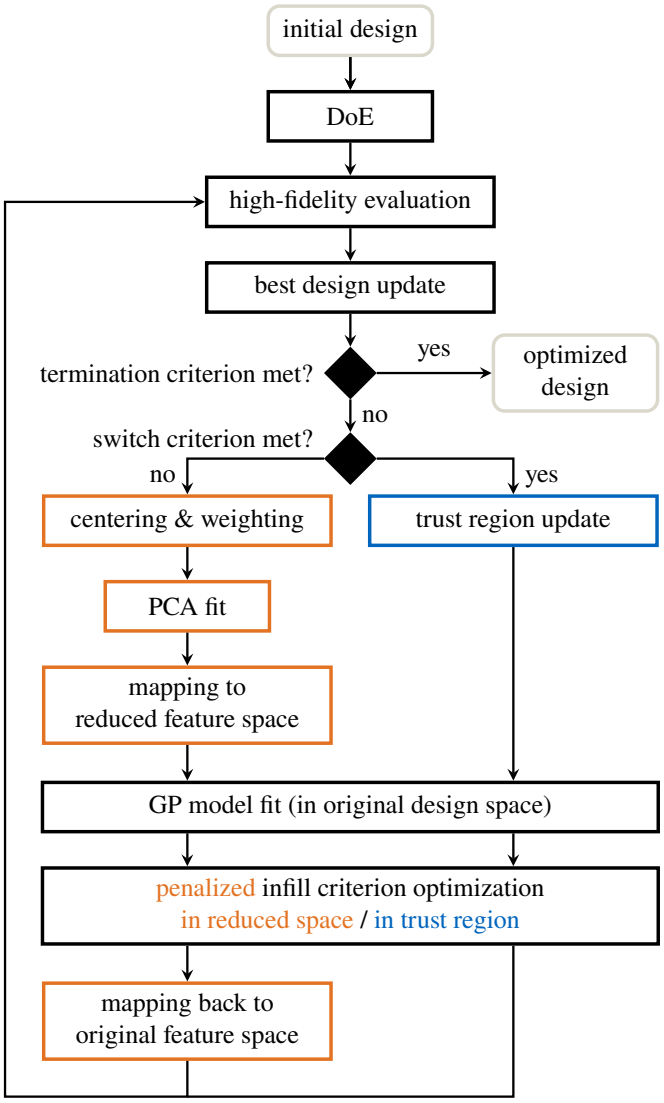


FIGURE 1: Flowchart of the proposed constrained PCA-BO + TR-BO approach. The parts which differ from the reference BO are highlighted in orange for PCA-BO and blue for TR-BO.

PF_j [6]. The resulting constrained infill criterion is defined as

$$\mathrm{CEI}(\boldsymbol{x}) = \mathrm{EI}(\boldsymbol{x}) \cdot \prod_{j=1}^{m} \mathrm{PF}_j(\boldsymbol{x}). \tag{2}$$

The CEI maximization is performed by a multi-start Nelder-Mead simplex method. It yields multiple local optima, which are then clustered to choose several infill point candidates to be evaluated in parallel, as explained by Arsenyev et al. [24].

Repeat until termination criterion is met. Finally, high-fidelity evaluations at these infill points provide new adaptive samples to enhance the GP model. This process of GP model fit, infill criterion optimization, high-fidelity evaluation, and best design update is repeated until a termination criterion is met. We chose the termination criterion as a maximum number of iterations or minimum distance between infill points.

2.2 High-dimensional BO

High dimensionality and the associated exponentially increasing problem complexity concerns both major components of the BO process. The GP model becomes inaccurate for limited training data. The infill criterion optimization becomes difficult and inefficient within affordable budgets. As remedy, three method classes for dimensionality reduction can be distinguished: variable screening [31], additive models [32], and (non-) linear embeddings. Examples for the latter are combinations of BO with random embeddings [33], active subspaces [34], PLS [14], or PCA, as used herein. A comprehensive summary of challenges and recent approaches of high-dimensional BO is provided by Binois and Wycoff [35].

They conclude with the idea that a combination of low-dimensional modeling and trust region approaches [22, 23, 36] might be the key to make BO practical for high-dimensional complex problems. In accordance, the comparison of various high-dimensional BO approaches by Santoni et al. [37] closed with the perspective of combining a PCA-assisted BO for fast convergence towards optimal regions of the design space with a trust region-based BO for efficient solution refinement.

3. PROPOSED APPROACH: CONSTRAINED PCA-BO + TR-BO

We take up these suggestions and propose to sequentially apply PCA-BO [21] and a trust region-based BO (TR-BO) [22, 23]. PCA-BO reduces the design space dimensionality and is thus better suited for global search in the initial phase of the optimization run. TR-BO localizes the search by reducing the design space size and is therefore more efficient in the later iterations. Starting with PCA-BO, we switch to TR-BO once there is no improvement for a predefined number of iterations.

Both of these adaptive design space reduction methods for BO [21, 22] were originally proposed for unconstrained problems and tested mainly on analytical functions. As mentioned in the introduction, they need modifications to enable industrial aero engine design applications. They must handle constrained high-dimensional problems and simultaneously evaluate several design candidates per iteration to exploit the available parallelized computing environment. The resulting two individual BO approaches are directly explained along with the respective original methods in the following Subsections 3.1 and 3.2.

3.1 Constrained PCA-BO

Raponi et al. [21] proposed a PCA-BO algorithm for efficient global optimization of unconstrained high-dimensional problems. Here, a PCA procedure applied to weighted samples restricts the search over an adaptive lower-dimensional manifold. This linear manifold is defined as the space spanned by the orthogonal directions in the design space along which the objective function value changes rapidly. Consequently, the infill criterion optimization is performed in a target-oriented reduced subspace, mitigating the curse of dimensionality.

A version with nonlinear manifolds based on kernel PCA was also published [38]. The much higher complexity does, however, usually not pay off by a considerably better performance [37]. Its major drawback is that kernel PCA does not provide an explicit map back. To overcome this limitation, kernel PCA-BO reconstructs the counterimages of solutions found in the low-dimensional space through sub-optimization procedures, inevitably introducing computational overhead.

The procedure proposed in this work for constrained problems is illustrated in the left branch of the flowchart in Figure 1. The differences to vanilla BO are highlighted in orange and are explained below. Please refer to Raponi et al. [21] for more mathematical details on the parts adopted from the unconstrained version.

Centering & weighting. As preparation for the PCA fit, the successfully evaluated sample points $\boldsymbol{x}^{(k)}$, $k = 1, ..., n$, must be rescaled by centering and weighting $\boldsymbol{x}'^{(k)} = \bar{\boldsymbol{x}}^{(k)} w^{(k)}$. This allows PCA to find directions in the design space going through the most promising regions. To compute the weights, the points are first attributed a rank $r^{(k)}$. In the original unconstrained formulation [21], the rank is determined according to their objective function value.

If the method must also handle constraints, this is no longer appropriate. We therefore rank the points so that among feasible points, the rank is determined by the objective function value. Infeasible points are ranked according to the total constraint violation. The infeasible points' ranks are then increased by adding the number of feasible points, so that feasible points are always ranked superior to all infeasible ones. The weights are computed as $\tilde{w}^{(k)} = \ln n - \ln r^{(k)}$, and normalized by $w^{(k)} = \tilde{w}^{(k)} / \sum \tilde{w}^{(k)}$. The better the point, the lower its rank and the larger its weight.

PCA fit. The PCA is trained on the weighted points $\boldsymbol{x}'^{(k)}$. Therefore, the d-dimensional sample points are centered again and assembled in a matrix $\boldsymbol{X}'' \in \mathbb{R}^{n \times d}$. The dominant features can then be determined by a singular value decomposition (SVD) of $\boldsymbol{X}''$. The resulting matrices contain the orthonormal basis vectors, the so-called principal components, and the singular values in descending order. This sorting enables an approximation of $\boldsymbol{X}''$ by truncating the matrices after the first r basis vectors and singular values, losing as little information as possible for any reduced rank $r \leq d$.

The reduced rank can either be specified in advance or chosen such that the first r variances, that is the squared singular values, sum up to a given percentage threshold of the total variance of the data. In this work, we determine r in the initial iteration by considering a 95% explained variance ratio. We then keep this value constant throughout the optimization procedure to prevent premature stagnation. The PCA implementation relies on the Python package *scikit-learn* [39] with a full SVD.

Mapping to reduced feature space & back. The PCA yields a linear map between the original and the reduced feature space. In the unconstrained PCA-BO [21], all sample points are mapped to the reduced feature space as input for a faster GP model fit. With the additional constraints, the reduced feature space is influenced by various response variables and is thus not suitable to describe the variation of each individual response. Surrogates in the reduced space may exhibit high multimodality and be prone to inaccurate prediction of feasible regions. For instance, sample points deemed infeasible might be projected relatively close to feasible ones in the reduced space, or vice versa, potentially compromising constraint-driven optimization efforts.

Nevertheless, the infill criterion is optimized based on the adaptive lower-dimensional parameterization $z_i, i = 1, \ldots, r$ from the PCA. Operating in the new reduced feature space entails the need for new variable bounds $z_{i,\mathrm{lb}} \leq z_i \leq z_{i,\mathrm{ub}}$. Neither should good regions be neglected, nor should the infill points lie outside the original design space when mapped back. Therefore, as in [21], we define a larger than necessary hyper-cubic reduced design space. Its center is the original design space center mapped to the reduced space. Its side length is the diameter of the hyper-sphere tangent to the original space. Finally, every surrogate model evaluation in the infill point search and every high-fidelity evaluation of new infill points requires a map back through $L : \mathbb{R}^r \to \mathbb{R}^d$.

Penalized infill criterion optimization in reduced space. A penalized constrained infill criterion is used to ensure that mapped back infill points lie within the original design space:

$$\mathrm{PCEI}(\boldsymbol{z}) = \begin{cases} \mathrm{CEI}(L(\boldsymbol{z})) & \text{if } x_{i,\mathrm{lb}} \leq L(\boldsymbol{z})_i \leq x_{i,\mathrm{ub}}, i = 1, \ldots, d, \\ -\mathrm{P}(L(\boldsymbol{z})) & \text{otherwise.} \end{cases} \tag{3}$$

If the considered infill point $\boldsymbol{z}$ lies outside the original design space when mapped back via $L(\boldsymbol{z})$, the CEI from Equation (2) is replaced by a negative penalty value P. The penalty is determined by summing the distances from the original domain bounds, calculated along each direction. Hence, points outside the original space are automatically discarded in the infill criterion optimization.

3.2 Constrained TR-BO

Regis [22] proposed a trust region implementation in Kriging-based optimization with expected improvement. In other words, he developed a trust region approach for BO of unconstrained high-dimensional problems. We adjust the approach to our problem by combining it with elements from Eriksson et al. [23], and denote the proposed method as TR-BO.

Like in PCA-BO, the infill criterion optimization is performed in an iteratively adapted reduced design space. However, here not the design space dimension, but the size is reduced by narrowing its bounds. Thereby, the infill point search becomes more local and thus simplified for large-scale problems. It is performed in a smaller region of the design space, in which we expect our optimum. The differences between TR-BO and BO are highlighted in blue in the right branch of Figure 1 and explained in the following. Note that the steps have partly been rearranged with respect to the original method [22], but the working principles are the same.

Trust region update. The trust region is a hypercube centered around the current best design $\boldsymbol{x}_{\mathrm{best},t}$ in every iteration t. Its size is also set iteratively and depends on the optimization progress. We consider all design variables to be scaled to $[0, 1]$ in the original design space, and set the trust region input parameters as defined in [22]: Initially, the trust region covers the entire design space, even if the center lies in a corner. The initial side length is hence $l_0 = l_{\max} = 2$, and we allow a reduction down to $l_{\min} = 1/64$. At iteration t, the side length l_t is halved if there is no actual improvement $\mathrm{AI} = f(\boldsymbol{x}_{\mathrm{best},t-1}) - f(\boldsymbol{x}_{\mathrm{best},t})$. It is doubled if the AI is greater than or equal to the (non-constrained) $\mathrm{EI} = \mathrm{EI}(\boldsymbol{x}_{\mathrm{best},t})$. The criterion AI/EI does not only consider sufficient surrogate accuracy, but emphasizes the optimization progress.

Compared to Regis' [22] original approach, we modify the update criteria to get a more robust algorithm for parallel infill points. Inspired by Eriksson et al. [23], we require the original update criteria to be fulfilled for τ consecutive iterations, instead of only one. Eriksson et al. set different values of τ for trust region expansion and shrinking. For the sake of simplicity, we set $\tau = d/q/2$ for a d-dimensional design space and q parallel infill points per iteration. At iteration t, the trust region size is thus adapted as

$$l_t = \begin{cases} \max(l_{t-1}/2, l_{\min}) & \text{if AI} = 0 \text{ for } \tau \text{ consecutive} \\ & \text{iterations since last resizing,} \\ \min(2l_{t-1}, l_{\max}) & \text{if AI/EI} \geq 1 \text{ for } \tau \text{ consecutive} \\ & \text{iterations since last resizing,} \\ l_{t-1} & \text{otherwise.} \end{cases} \tag{4}$$

Finally, the reduced range is updated to

$$\begin{aligned} \max(x_{i,\mathrm{best},t} - l_t/2, x_{i,\mathrm{lb}}) \leq x_i \\ \leq \min(x_{i,\mathrm{best},t} + l_t/2, x_{i,\mathrm{ub}}), i = 1, \ldots, d. \end{aligned} \tag{5}$$

The presence of constraints can be considered to have little impact on the trust region update. Constraint satisfaction is already implied by the update criterion, because full constraint satisfaction is a prerequisite for every new best point.

As part of the proposed hybrid approach, the criterion for switching from PCA-BO to TR-BO is the first TR contraction. We switch to full dimension and reduced side length $l_t = 1$ if $\mathrm{AI} = 0$ for the last τ iterations. The subsequent TR-BO steps are performed as specified above.

Infill criterion optimization in trust region. The infill criterion optimization is localized inside the trust region. As the included hyper-volume is exponentially reduced, also fewer sample points with $\mathrm{CEI} = 0$ are included. This makes the acquisition function less multi-modal and hence easier and faster to optimize. In contrast, we still fit the surrogate model in the full domain including all training points for maximum accuracy, especially near the trust region bounds. A GP fit within the trust region would allow for region-specific hyperparameter tuning, a feature beneficial for handling inhomogeneous functions [23], which are not expected in this context.

The downside of the localization is the risk of getting stuck in a local optimum. Remedies are restarts once the infill criterion falls below a threshold [22], or simultaneous optimization in q local trust regions [23]. Both come at the cost of an increased number of expensive high-fidelity evaluations. We do not incorporate any of these globalization strategies, as our hybrid approach relies on localization through trust regions as the operational mechanism in the second phase.

4. RASTRIGIN TEST FUNCTION OPTIMIZATION RESULTS

We first assess the performance of our hybrid algorithm and its two individual parts by comparing them to a vanilla BO on an analytical test problem, the Rastrigin function. Additionally,

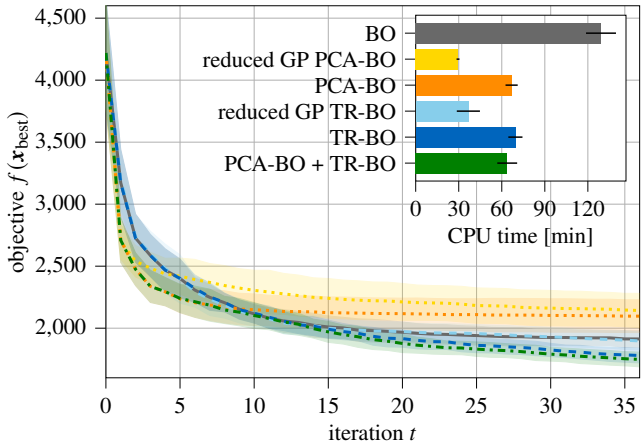


FIGURE 2: Objective function (Rastrigin function minimization) over optimization iteration. The mean and standard deviation over 30 runs per method are plotted based on the current best feasible objective values of each run and method. Additionally, the mean and standard deviation of the total CPU time per run is plotted. The prefix *reduced GP* denotes versions with GP model fit in the respective reduced space.

versions of PCA-BO and TR-BO with the GP model fit in the respective reduced space are compared to justify our choice of a GP model fit in the full design space. The high-fidelity function evaluations are computationally cheap and thus enable many repetitions for statistically meaningful comparisons. The unconstrained Rastrigin function was amongst others used by Raponi et al. [21] with $d = 10, 20, 40$, and Regis [22] with $d = 30$ to test their respective methods we build upon. It is part of the black-box optimization benchmarking (BBOB) test suite [40], which also provides a constrained version. We examine our methods based on an instance of the 40-dimensional multi-modal function with 4 nonlinear inequality constraints. They are generated according to the constrained BBOB formulation by Dufossé et al. [41].

Following the experimental set-up of Raponi et al. [21], we fix our high-fidelity evaluation budget to $10d + 50$, of which 20% are used for the DoE. Given our test problem dimension $d = 40$, this yields 90 initial samples and 36 optimization iterations with $q = 10$ parallel high-fidelity evaluations per iteration. BO is non-deterministic, with the DoE and the initial points of the acquisition function optimization as main sources of randomness. Therefore, we perform 30 runs of each method. Every run has an individual seed to initialize the random state for a fair comparison between the methods within a single run. All computations are performed on a single Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz.

Figure 2 shows the convergence of the Rastrigin function minimization over the optimization iterations. Only fully feasible points are taken into account. PCA-BO reduces the design space dimension from $d = 40$ to $r = 24$ on average. TR-BO performs the first trust region reduction after 9 iterations on average, our hybrid algorithm (PCA-BO + TR-BO) already after 8 iterations. We therefore roughly divide our analysis in the initial phase up until 10 iterations and the remainder.

In the initial phase, the TR-BO curve in blue overlays the BO in gray, and the PCA-BO + TR-BO in green overlays the PCA-BO curve in orange, because the trust region does not yet actively reduce the design space. The PCA-BO methods significantly speed up the initial convergence, just like for the unconstrained single-infill results by Raponi et al. [21].

PCA-BO stagnates after 10 iterations, as the reduced set of adaptive design variables represents a limit for further convergence. BO also mostly stagnates after about 25 iterations. In both cases, TR-BO enables a further convergence, because it adaptively localizes and hence simplifies the infill point search. This qualitatively matches the unconstrained single-infill results by Regis [22].

Since we test on an analytical function, the CPU time per run is mainly spent on the BO algorithm. The proposed hybrid algorithm and its individual parts can approximately halve the runtime. This is because they perform the computationally expensive infill criterion optimization in reduced design spaces. The versions with also the GP model fit in the respective reduced space halve the runtime once more. However, these savings come at the cost of decreased model accuracy, causing an earlier stagnation for both PCA-BO and TR-BO.

The first tests on the constrained 40D Rastrigin function show that our constrained and parallel variants of the PCA-BO and TR-BO algorithms can effectively reduce the design space in different phases of the optimization. Applied consecutively in our hybrid method, they can yield a better feasible design at reduced algorithm runtime for both small and larger high-fidelity evaluation budgets.

5. HPC AERODYNAMIC BLADE DESIGN

As realistic industrial application, we deal with the second stage of a generic HPC with eight stages in total. It is inspired by a modern high-bypass geared turbofan engine.

The annulus geometry and initial blade profiles are predesigned based on a previous optimization using 2D streamline curvature methods. The initial 3D airfoils are generated from the 2D solution with an in-house software and minor manual adjustments. We seek for a detailed rotor blade design with maximum isentropic efficiency, while fulfilling various geometric and aerodynamic constraints. The detailed blade design employs 3D blade geometry generation and computational fluid dynamics (CFD).

5.1 Blade geometry parameterization

The design problem is characterized by 46 rotor plus 9 stator blade geometry parameters, which is considered a high dimensionality to be effectively addressed by BO [35]. They are required to enable large variations of complex blade geometries. The parameters are as mutually independent as possible, and are chosen for a target-oriented modification of local flow phenomena.

The blade geometry parameters are illustrated in Figure 3. 2D profiles are defined along the streamlines at hub, mid, and tip, and then stacked to generate the 3D geometry. For the 2D profiles, four base points are set as joint points between the four profile segments pressure side (PS), suction side (SS), leading edge (LE), and trailing edge (TE). Their location with respect to the profile center is defined by the blade angles α, the stagger β_S, the wedge angles γ, and the distance wedge l. The shapes

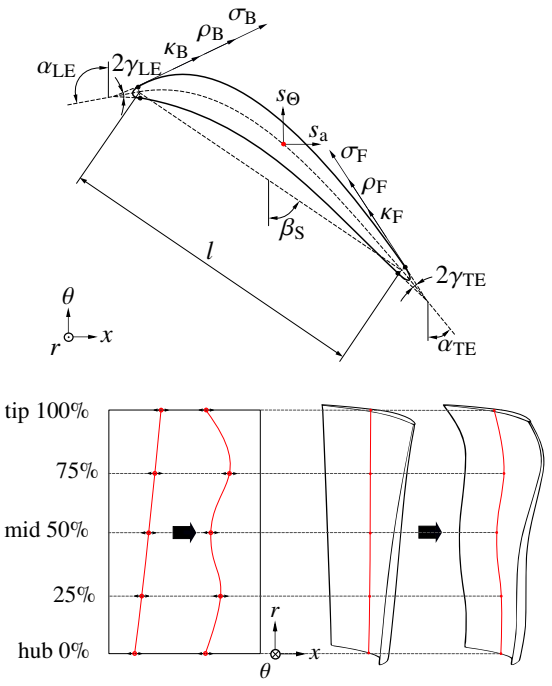


FIGURE 3: Blade geometry parameterization. Modified from [42]. The profile parameters (top) at blade hub, mid, and tip (bottom) are the design variables of the optimization problem

of the PS and SS segment splines are described by the curvature κ, and the aspect ratio parameters of slope ρ and curvature σ. These parameters are symmetric both at LE (backward), and TE (forward).

For a smooth 3D blade, the 2D profiles are radially stacked and interpolated with polynomials of specified orders. The location of the blade centers are controlled by the shift s in axial and circumferential direction. All of these parameters are free to modify the rotor blades, while only blade angles α and stagger β_S are considered to alter the stator. The focus is on the rotor blade design, while the stator parameters only adjust to the changing boundary conditions due to the large rotor modifications.

5.2 Aerodynamic design evaluation

The above geometry parameters are the input for the high-fidelity evaluation of new designs, which is outlined in what follows. The geometry generation is performed by in-house tools and takes about 2 minutes per design. To enable large design variations, we set wide variable ranges in the optimization problem. With the high number of variables, the resulting design space contains regions where the geometry generation fails. The complex blade geometries make it difficult to identify these regions in advance. Therefore, the optimizer must handle the simulation failures.

The aerodynamic analysis is performed by 3D CFD simulations. The flow solution is based on the RANS equations and computed by *TRACE* [43], a program specialized on turbomachinery components developed by the *Institute of Propulsion Technology* of the *German Aerospace Center (DLR)* and the *MTU Aero Engines AG*. We consider two operating points: the aero design point (ADP) at cruise conditions, and an off-design point (ODP). The ODP is a throttled state with the same rotational speed as the ADP but higher back pressure, which is used to ensure acceptable off-design performance. The flow around the two blades is transonic. The aerodynamic analysis has by far the highest simulation expense, with about 18 minutes on 16 CPUs per design point, thus 32 CPUs in total.

TABLE 1: Compressor blades optimization problem definition.

1 objective $f(\boldsymbol{x})$	**type**
isentropic efficiency η_{is} at ADP	maximization
20 constraints $h_j(\boldsymbol{x})$	**type of bound**
isentropic efficiency η_{is} at ODP	lower (=baseline)
mass flow $\dot{m}$ at ADP	lower & upper
mass flow $\dot{m}$ at ODP	lower (=baseline)
downstream flow angle α_Z at ADP	lower & upper
constriction at 11 radial slices of rotor	lower
max. blade thickness delta at 4 radial slices near rotor tip	lower
max. blade thickness at rotor tip	lower & upper
55 design variables x_i	**type of bound**
see Figure 3	lower & upper

5.3 Design optimization problem

Our detailed blade design optimization problem for the second stage of a HPC is defined in Table 1. We aim at maximizing the isentropic efficiency at cruise conditions by modifying the rotor and stator blade geometry parameters described above. The optimized design must fulfill a number of inequality constraints. They are computed along with the objective function by the 3D simulation process outlined before.

When optimizing a single stage, we must ensure that it fits back into the system. Therefore, boundary conditions are extracted from the full HPC CFD simulations and additional aerodynamic constraints are introduced. In this case, the variation of mass flow and downstream flow angle is limited by narrow lower and upper bounds. Additionally, the performance at the throttled ODP should not degrade during the optimization. This is realized by a lower bound for the isentropic efficiency and a minimal requirement for the absolute mass flow.

Concerning the rotor blade geometry, the constriction at 11 radial positions should not be too small to guarantee the manufacturability. For the same reason, the tip region should not become too thin.

The described optimization problem represents a realistic industrial use case. It requires an optimization algorithm that can deal with 55 design parameters and 31 constrained response variables.

6. BLADE OPTIMIZATION RESULTS

We test the proposed approach on the above described HPC aerodynamic blade design problem by comparing the performance of the constrained PCA-BO, TR-BO, their combination,

and a vanilla BO algorithm as reference. Besides the initial sampling and the optimization convergence, also the progress of design space reduction, the ability to handle both simulation failures and a highly constrained design space, and the final designs are analyzed.

Due to the increased complexity of the blade design problem, we double the high-fidelity evaluation budget of the Rastrigin function optimization to $20d+100$. With design space dimension $d = 55$, we aim to evaluate 240 initial points and then perform 96 optimization iterations with $q = 10$ parallel high-fidelity evaluations each. The high CFD simulation times limit the number of runs per method to 5. Each of the 5 runs has an individual initial seed for a fair method comparison.

As mentioned in Section 5.2 and specified in Table 2, the geometry generation fails in about 77% of the design space. This is mostly because pressure and suction side surfaces interpenetrate or become very close. In this case, also the constriction constraints are not satisfied, because the profile thickness is reduced in between the LE and TE base points (see Figure 3). The geometrical constraints thereby drive the optimization towards regions with successful evaluations and make a dedicated failure handling in the optimizer unnecessary.

Since the geometry generation is the first and least computationally expensive step of the high-fidelity evaluation, these failed simulations do not add much time to the initial sample generation. We initially evaluate $20d$ design points to get on average 245 successful samples. The initial sampling phase takes about 6 hours in total, as 20 designs are evaluated in parallel. Apart from the initial point in the center of the design space, denoted by $\boldsymbol{x}_0$, all DoE points violate at least one constraint, indicating a small feasible design space.

PCA-BO is highly dependent on the initial sampling, because the initial principal components bias the further optimization. Figure 4 shows the variance ratio captured by an increasing number r of principal components which were computed based on the initial sampling. With only 10 out of 55 dimensions capturing half the variance, and 25 dimensions capturing about 80% of the variance, it becomes clear that a design space reduction is possible. The 95% threshold is reached with $r = 40$ to 42 components for the different runs. The resulting dimension reduction by a fourth shows that the intrinsic problem dimensionality is not much lower than the original one. In other words, our design space dimension is not artificially high. The negligible standard deviation over the 5 different initial sample sets indicates an almost sample-independent choice of the reduced dimension r.

The optimization progress of the four methods is illustrated in Figures 5 and 6. Figure 5 shows the convergence of the objective, isentropic efficiency maximization, over the number of iterations on the top and over the wall time on the bottom. Figure 6 shows the associated design space reduction over the number of iterations, with the reduced dimension r by PCA-BO on the top and the reduced side length l by TR-BO on the bottom. Note that designs are only considered as current best $\boldsymbol{x}_{\text{best}}$ if they satisfy all constraints.

The switch from PCA-BO to TR-BO in our hybrid algorithm can be seen in the upper part of Figure 6. It is indicated by the jumps up to the original dimension $d = 55$ between 16 and 35

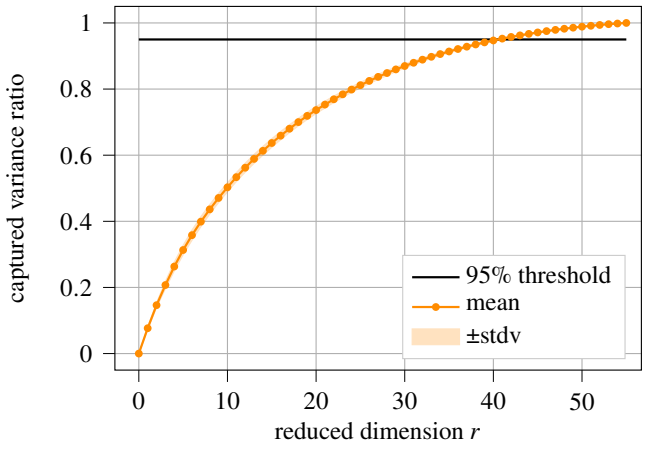


FIGURE 4: Variance ratio captured by the first r principal components based on the initial sampling. For PCA-BO, we choose r so as to explain at least 95% of the variance. The mean and the (very small) standard deviation over 5 runs are plotted, as well as the 95% threshold.

iterations. Therefore, we roughly split our discussion in the first 20 iterations and the remainder.

In the first iterations, the BO and TR-BO curves and the PCA-BO and PCA-BO + TR-BO curves are similar but not identical, unlike in Figure 2 for the Rastrigin test function. The small deviations are caused by numerical errors in the high-fidelity evaluations. The initial convergence slope of the two PCA-BO approaches is about twice as steep as of the other two methods. For the improvement achieved with adaptive dimension reduction via PCA in 20 iterations, the methods with full dimension need about 80 iterations. This behavior is especially useful for industrial engineering applications, where we seek for a quickly improved design proposal rather than the mathematical optimum.

Also unlike the analytical tests, PCA-BO does not stagnate in the later iterations but continuously finds better designs. However, the standard deviation of the objective is significantly higher for PCA-BO compared to the results obtained with full dimension. This might be due to the strong influence of the adaptive PCA-based reduced design variables, which in turn depend on different training points gathered during each run.

After the first 20 iterations, all three proposed methods show a similar convergence rate, but a superior performance to vanilla BO. TR-BO yields an additional improvement in both cases. The adaptive design space contraction is, however, less effective than the dimension reduction via PCA-BO and cannot make up for the head start in the first iterations. After the full 96 optimization iterations, BO reaches on average only 80% of the objective function improvement of the proposed hybrid approach. This is a big difference in a detailed design optimization like the one addressed in this study.

In addition, the vanilla BO algorithm computations take more than double the time of the three proposed approaches. This can be deduced from the lower part of Figure 5, which shows the objective function over the wall time. The simulations take about 96·20 minutes = 32 hours for all four methods and runs. PCA-BO

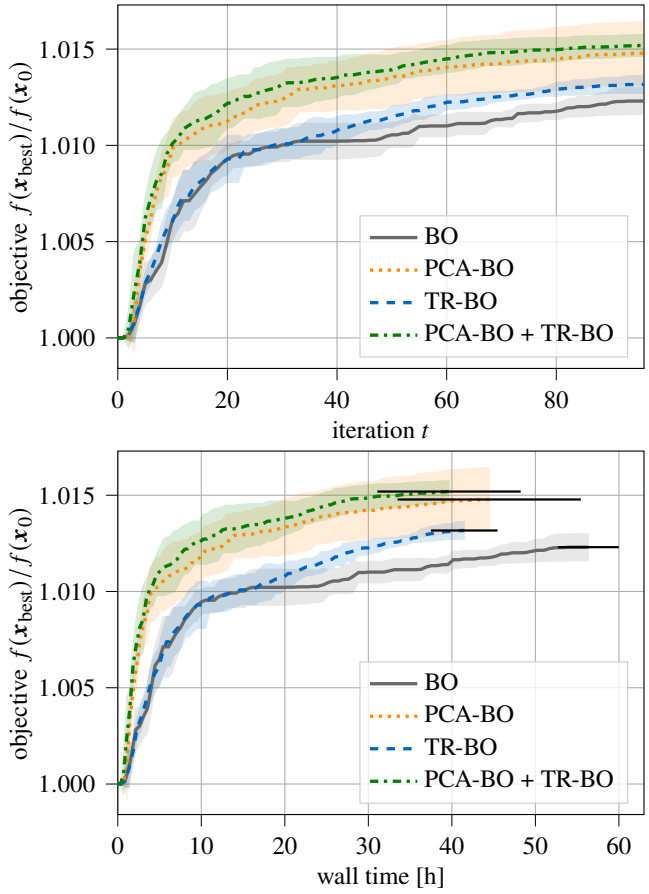


FIGURE 5: Objective function (isentropic efficiency maximization) over optimization iteration (top) and wall time (bottom). The mean and standard deviation over 5 runs per method are plotted based on the current best feasible normalized objective values of each run and method. On the bottom plot, additionally, the standard deviation of the total wall time per run is indicated with black horizontal lines.

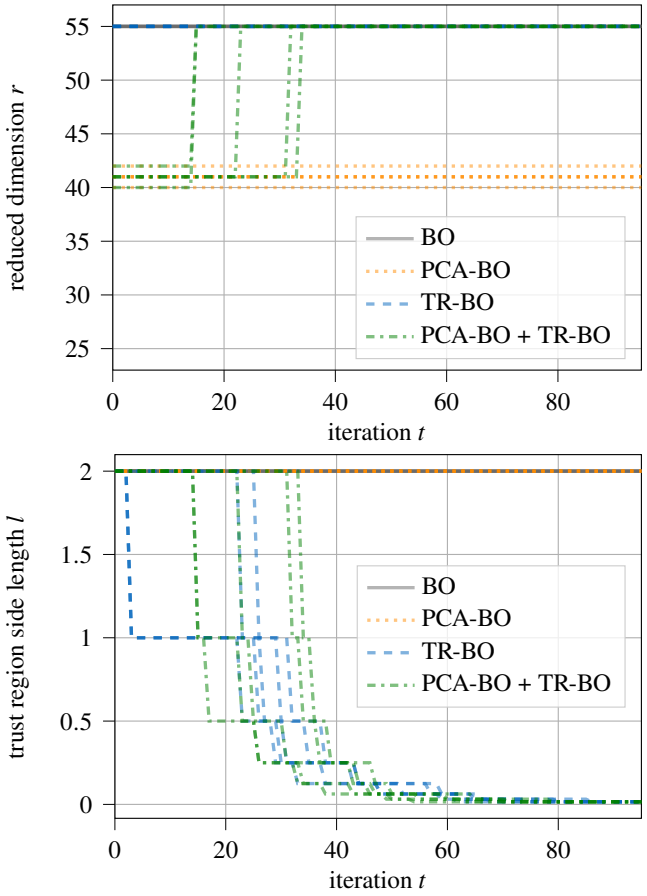


FIGURE 6: Design space reduction over number of optimization iterations: The top figure shows the (reduced) design space dimension r (= d if there is no reduction). The bottom figure shows the trust region side length l. All 5 runs per method are plotted semi-transparent, so that less transparent lines indicate several runs with the same course.

+ TR-BO takes on average 40 hours to complete the 96 iterations, so 8 hours are spent on the optimization algorithm. PCA-BO and TR-BO individually take only a little longer. Vanilla BO, in contrast, takes on average 16 hours longer. This indicates again that design space reductions mitigate problem complexity. Thereby, they save much more computation time than the additional reduction steps in the algorithm take.

Coming back to the optimization objective, we compare the average isentropic stage efficiency improvement $f(\boldsymbol{x}_{\text{best}})/f(\boldsymbol{x}_0)$ for the proposed hybrid algorithm and BO at two different wall times: 8 hours, corresponding to 20 iterations of our approach, and 40 hours, corresponding to the final iteration. In the first 8 hours, the average increase in isentropic stage efficiency is 0.85% without design space reductions, compared to 1.22% with the implemented reductions. Despite the considerably better start of our proposed approach, the improvement between 8 and 40 hours is similar, yielding on average in total 1.16% versus 1.52% efficiency improvement with respect to the initial design. These are significant isentropic stage efficiency improvements given the preoptimized initial design.

The blade optimization also tests our approaches' ability to handle simulation failures and complex constraints. Table 2 lists the mean ratios of successful high-fidelity evaluations and feasible designs in the initial DoE and during the optimizations with 5 runs per method. Among the compared optimization approaches, PCA-BO's ability to find successful and feasible points is clearly superior. With 37% feasible designs it achieves nearly twice the ratio of BO and TR-BO. This matches our expectations, as the principal components are computed based on only successful and preferably feasible designs. Moreover, it explains the even better performance of PCA-BO on the blade design problem compared to the analytical test case without failures.

For the final best designs, all 20 constraints are satisfied, in accordance with our definition of $\boldsymbol{x}_{\text{best}}$. The isentropic efficiency increases along with the objective function. All other aerodynamic constraints are active, meaning that the response value lies on the constraint bounds. The geometric profile constriction constraints are only partly active for the tip sections. The remaining geometric constraints are mainly needed to drive the search to-

TABLE 2: Mean ratios of successful simulations and feasible designs during the initial DoE and the iterative phase of the compared optimization algorithms.

	DoE	BO	PCA-BO	TR-BO	PCA-BO + TR-BO
successful	23%	47%	57%	36%	43%
feasible	0%	19%	37%	19%	25%

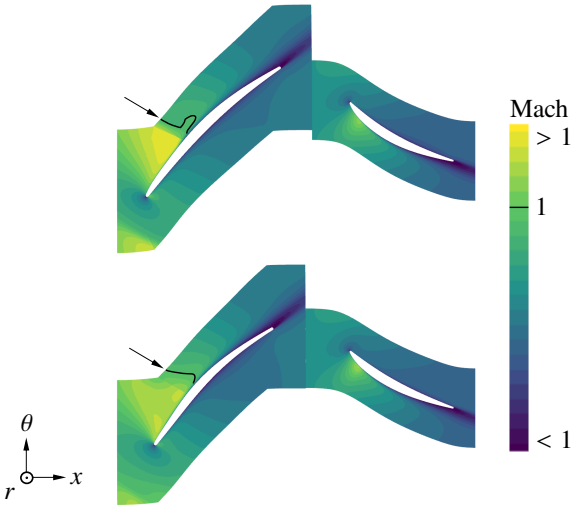


FIGURE 7: Mach number distribution at the mid plane for initial (top) and optimized (bottom) design obtained by the proposed BO approach in the run closest to the mean. The switch from super- to subsonic flow at Mach = 1 on the rotor suction side is marked by the black lines.

wards reasonable geometries and are therefore easily fulfilled for the final best designs.

To conclude the result analysis, we take a look at the main source of the efficiency improvements. Since the blades are transonic, there is a shock where the flow over the blade's suction side is decelerated to subsonic speeds. The shock-induced boundary layer separation increases the drag and thus reduces the efficiency. Figure 7 shows the Mach number distribution at the mid plane for the initial and optimized design. The latter is from the run of our hybrid approach with the result closest to the mean. At the rotor blade suction side, the maximum Mach number is lowered and the deceleration to Mach < 1 is spread over a longer blade span, softening the shock for an increase in efficiency.

7. CONCLUSIONS

The outcome of this work is a BO approach for high-dimensional constrained problems, which is shown to significantly accelerate the optimization and further improve the isentropic efficiency of an industrial aero engine blade design. We propose to facilitate the surrogate-based infill point search in the high-dimensional design space by a twofold adaptive reduction. In the first phase, the design space dimension is adaptively reduced via PCA. In the second phase, the design space ranges are adaptively reduced by a trust region approach. Both individual parts, PCA-BO and TR-BO, are extensions of existing approaches to address parallel-constrained scenarios, providing simultaneously candidate solutions fulfilling constraint feasibility.

We test our approach on two optimization problems: the analytical 40D Rastrigin function and an industrial 55D HPC aerodynamic blade design. The results for the constrained 40D Rastrigin function show that PCA-BO improves the convergence rate in the first, more explorative, optimization iterations, and then stagnates. The algorithm then switches to TR-BO, which localizes the search for a further objective improvement compared to vanilla BO. An even larger improvement is achieved for the 55D HPC aerodynamic blade design. Our hybrid approach not only achieves an equally good design as vanilla BO within just 20 iterations of the full 96, but also demonstrates further capability of improvement beyond this point. This can be explained by the improved failure handling of PCA-BO. Moreover, the algorithm computation times are more than halved. The quick initial convergence makes PCA-BO a useful tool for industrial engineering applications, while TR-BO can contribute to better optimize detailed designs, where minor improvements might matter.

Potential future work includes a speed up of the GP model fit, for instance by a GP-internal dimension reduction via PLS [14] or other design space reduction approaches [35]. Besides, PCA-BO and TR-BO can be employed simultaneously for a smoother transition. Alternatively, the final local optimization phase can be done using adjoint methods. This study contributes to the long-term goal to make the benefits of BO available for even higher-dimensional constrained design problems. One example is a multi-stage turbomachinery configuration, where a joint detailed design can yield better results as aerodynamic inter-stage coupling is taken into account.

ACKNOWLEDGMENTS

The authors would like to thank the *MTU Aero Engines AG* and *Technical University of Munich* for their support in this work. This project is funded by the *Bayerische Luftfahrtforschung- und technologieförderung* of the *Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie (StMWi)*. Elena Raponi acknowledges funding by the PRIME programme of the German Academic Exchange Service (DAAD) with funds from the German Federal Ministry of Education and Research (BMBF).

REFERENCES

[1] European Commission and Directorate-General for Research and Innovation. *Fly the Green Deal : Europe's vision for sustainable aviation*. Publications Office of the European Union (2022). DOI doi/10.2777/732726.

[2] U.S. Department of Transportation – Federal Aviation Administration. *United States Aviation Climate Action Plan* (2021). URL https://www.faa.gov/sites/faa.gov/files/2021-11/Aviation_Climate_Action_Plan.pdf.

[3] Martins, Joaquim R. R. A. and Ning, Andrew. *Engineering Design Optimization*. Cambridge University Press (2021). DOI 10.1017/9781108980647.

[4] Mockus, Jonas. *Bayesian Approach to Global Optimization*.

Springer Netherlands (1989). DOI 10.1007/978-94-009-0909-0.

[5] Jones, Donald R., Schonlau, Matthias and Welch, William J. "Efficient Global Optimization of Expensive Black-Box Functions." *Journal of Global Optimization* Vol. 13 No. 4 (1998): pp. 455–492. DOI 10.1023/a:1008306431147.

[6] Schonlau, Matthias, Welch, William J. and Jones, Donald R. "Global versus local search in constrained optimization of computer models." *Institute of Mathematical Statistics Lecture Notes - Monograph Series*. Institute of Mathematical Statistics (1998): pp. 11–25. DOI 10.1214/lnms/1215456182.

[7] Dharmadhikari, Susheel, Berdanier, Reid A., Thole, Karen A. and Basak, Amrita. "An Automated Gaussian Process Integrated With Bayesian Optimization Approach to Designing Spline-Based Pin-Fin Arrays." *Proceedings of the ASME Turbo Expo*. 2023. American Society of Mechanical Engineers. DOI 10.1115/gt2023-102018.

[8] Schaffrath, Robert, Nicke, Eberhard, Forsthofer, Nicolai, Kunc, Oliver and Voß, Christian. "Gradient-Free Aerodynamic Optimization With Structural Constraints and Surge Line Control for Radial Compressor Stage." *Proceedings of the ASME Turbo Expo*. 2023. American Society of Mechanical Engineers. DOI 10.1115/gt2023-101593.

[9] Zhang, Yiming, Ghosh, Sayan, Vandeputte, Thomas and Wang, Liping. "Bayesian Optimization for Multi-Objective High-Dimensional Turbine Aero Design." *Proceedings of the ASME Turbo Expo*. 2021. American Society of Mechanical Engineers. DOI 10.1115/gt2021-59745.

[10] Song, Wenbin and Keane, Andy J. "Surrogate-Based Aerodynamic Shape Optimization of a Civil Aircraft Engine Nacelle." *AIAA Journal* Vol. 45 No. 10 (2007): pp. 2565–2574. DOI 10.2514/1.30015.

[11] Lopez, Diego I., Ghisu, Tiziano and Shahpar, Shahrokh. "Global Optimisation of a Transonic Fan Blade Through AI-Enabled Active Subspaces." *Volume 2A: Turbomachinery - Axial Flow Fan and Compressor Aerodynamics*. 2021. American Society of Mechanical Engineers. DOI 10.1115/gt2021-59166.

[12] Constantine, Paul G. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. Society for Industrial and Applied Mathematics (2015). DOI 10.1137/1.9781611973860.

[13] Hartwig, Lennard and Bestle, Dieter. "Compressor blade design for stationary gas turbines using dimension reduced surrogate modeling." *2017 IEEE Congress on Evolutionary Computation (CEC)*. 2017. IEEE. DOI 10.1109/cec.2017.7969493.

[14] Bouhlel, Mohamed Amine, Bartoli, Nathalie, Regis, Rommel G., Otsmane, Abdelkader and Morlier, Joseph. "Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method." *Engineering Optimization* Vol. 50 No. 12 (2018): pp. 2038–2053. DOI 10.1080/0305215x.2017.1419344.

[15] Zhang, Lizhang, Mi, Dong, Yan, Cheng and Tang, Fangming. "Multidisciplinary Design Optimization for a Centrifugal Compressor Based on Proper Orthogonal Decomposition and an Adaptive Sampling Method." *Applied Sciences* Vol. 8 No. 12 (2018): p. 2608. DOI 10.3390/app8122608.

[16] Toal, David J. J., Bressloff, Neil W., Keane, Andy J. and Holden, Carren M. E. "Geometric Filtration Using Proper Orthogonal Decomposition for Aerodynamic Design Optimization." *AIAA Journal* Vol. 48 No. 5 (2010): pp. 916–928. DOI 10.2514/1.41420.

[17] Berguin, Steven H. and Mavris, Dimitri N. "Dimensionality Reduction In Aerodynamic Design Using Principal Component Analysis With Gradient Information." *10th AIAA Multidisciplinary Design Optimization Conference*. 2014. American Institute of Aeronautics and Astronautics. DOI 10.2514/6.2014-0112.

[18] Cinquegrana, Davide and Iuliano, Emiliano. "Efficient Global Optimization of a Transonic Wing with Geometric Data Reduction." *35th AIAA Applied Aerodynamics Conference*. 2017. American Institute of Aeronautics and Astronautics. DOI 10.2514/6.2017-3057.

[19] Poole, Daniel J., Allen, Christian B. and Rendall, Thomas C. S. "Efficient aeroelastic wing optimization through a compact aerofoil decomposition approach." *Structural and Multidisciplinary Optimization* Vol. 65 No. 3 (2022). DOI 10.1007/s00158-022-03174-4.

[20] Schlaps, R. C., Shahpar, S. and Gümmer, V. "Automatic Three-Dimensional Optimisation of a Modern Tandem Compressor Vane." *Volume 2B: Turbomachinery*. 2014. American Society of Mechanical Engineers. DOI 10.1115/gt2014-26762.

[21] Raponi, Elena, Wang, Hao, Bujny, Mariusz, Boria, Simonetta and Doerr, Carola. "High Dimensional Bayesian Optimization Assisted by Principal Component Analysis." *Parallel Problem Solving from Nature – PPSN XVI*. Springer International Publishing (2020): pp. 169–183. DOI 10.1007/978-3-030-58112-1_12.

[22] Regis, Rommel G. "Trust regions in Kriging-based optimization with expected improvement." *Engineering Optimization* Vol. 48 No. 6 (2015): pp. 1037–1059. DOI 10.1080/0305215x.2015.1082350.

[23] Eriksson, David, Pearce, Michael, Gardner, Jacob R., Turner, Ryan and Poloczek, Matthias. "Scalable global optimization via local Bayesian optimization." *Advances in Neural Information Processing Systems*, Vol. 32. Conference paper. 2019.

[24] Arsenyev, Ilya, Duddeck, Fabian and Fischersworring-Bunk, Andreas. "Adaptive Surrogate-Based Multi-Disciplinary Optimization for Vane Clusters." *Volume 2C: Turbomachinery*. 2015. American Society of Mechanical Engineers. DOI 10.1115/gt2015-42164.

[25] Forrester, Alexander I.J. and Keane, Andy J. "Recent advances in surrogate-based optimization." *Progress in Aerospace Sciences* Vol. 45 No. 1-3 (2009): pp. 50–79. DOI 10.1016/j.paerosci.2008.11.001.

[26] Garnett, Roman. *Bayesian Optimization*. Cambridge University Press (2023). DOI 10.1017/9781108348973.

[27] Fang, Kai-Tai, Li, Runze and Sudjianto, Agus. *Design and Modeling for Computer Experiments*. Chapman and Hall/CRC (2005). DOI 10.1201/9781420034899.

[28] Baert, Lieven, Dumeunier, Christophe, Leborgne, Michaël, Sainvitu, Caroline and Lepot, Ingrid. "Agile SBO Framework Exploiting Multisimulation Data: Optimising Efficiency and Stall Margin of a Transonic Compressor." *Volume 2D: Turbomachinery*. 2018. American Society of Mechanical Engineers. DOI 10.1115/gt2018-76639.

[29] Bouhlel, Mohamed Amine, Hwang, John T., Bartoli, Nathalie, Lafage, Rémi, Morlier, Joseph and Martins, Joaquim R.R.A. "A Python surrogate modeling framework with derivatives." *Advances in Engineering Software* Vol. 135 (2019): p. 102662. DOI 10.1016/j.advengsoft.2019.03.005.

[30] Rasmussen, Carl E. and Williams, Christopher K. I. *Gaussian processes for machine learning*. MIT Press, Cambridge, Massachusetts, USA (2006). URL https://gaussianprocess.org/gpml/.

[31] Salem, Malek Ben, Bachoc, Francois, Roustant, Olivier, Gamboa, Fabrice and Tomaso, Lionel. "Gaussian Process-Based Dimension Reduction for Goal-Oriented Sequential Design." *SIAM/ASA Journal on Uncertainty Quantification* Vol. 7 No. 4 (2019): pp. 1369–1397. DOI 10.1137/18m1167930.

[32] Rolland, Paul, Scarlett, Jonathan, Bogunovic, Ilija and Cevher, Volkan. "High-Dimensional Bayesian Optimization via Additive Models with Overlapping Groups." *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, Vol. 84: pp. 298–307. 2018. PMLR. DOI 10.48550/ARXIV.1802.07028.

[33] Wang, Ziyu, Hutter, Frank, Zoghi, Masrour, Matheson, David and Feitas, Nando De. "Bayesian Optimization in a Billion Dimensions via Random Embeddings." *Journal of Artificial Intelligence Research* Vol. 55 (2016): pp. 361–387. DOI 10.1613/jair.4806.

[34] Nayebi, Amin, Munteanu, Alexander and Poloczek, Matthias. "A framework for Bayesian optimization in embedded subspaces." *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97: pp. 4752–4761. 2019. PMLR.

[35] Binois, Mickaël and Wycoff, Nathan. "A Survey on High-dimensional Gaussian Process Modeling with Application to Bayesian Optimization." *ACM Transactions on Evolutionary Learning and Optimization* Vol. 2 No. 2 (2022): pp. 1–26. DOI 10.1145/3545611.

[36] Diouane, Youssef, Picheny, Victor, Riche, Rodolophe Le and Perrotolo, Alexandre Scotto Di. "TREGO: a trust-region framework for efficient global optimization." *Journal of Global Optimization* (2022)DOI 10.1007/s10898-022-01245-w.

[37] Santoni, Maria Laura, Raponi, Elena, Leone, Renato De and Doerr, Carola. "Comparison of High-Dimensional Bayesian Optimization Algorithms on BBOB." (2023)URL 2303.00890.

[38] Antonov, Kirill, Raponi, Elena, Wang, Hao and Doerr, Carola. "High Dimensional Bayesian Optimization with Kernel Principal Component Analysis." (2022). DOI 10.48550/ARXIV.2204.13753.

[39] Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* Vol. 12 (2011): pp. 2825–2830. URL http://jmlr.org/papers/v12/pedregosa11a.html.

[40] Hansen, Nikolaus, Auger, Anne, Ros, Raymond, Mersmann, Olaf, Tušar, Tea and Brockhoff, Dimo. "COCO: a platform for comparing continuous optimizers in a black-box setting." *Optimization Methods and Software* Vol. 36 No. 1 (2020): pp. 114–144. DOI 10.1080/10556788.2020.1808977.

[41] Dufossé, Paul, Hansen, Nikolaus, Brockhoff, Dimo, Sampaio, Phillipe R., Atamna, Asma and Auger, Anne. "Building scalable test problems for benchmarking constrained optimizers." Technical report no. 2022. URL http://numbbo.github.io/coco-doc/bbob-constrained/.

[42] Arsenyev, Ilya. "Efficient surrogate-based robust design optimization method." Ph.D. Thesis, Technical University of Munich. 2018.

[43] German Aerospace Center (DLR). "TRACE." http://www.trace-portal.de/userguide/trace/index.html (2023). Accessed: 2023-10-04.

A.3 Publication III

L. Pretsch, I. Arsenyev, N. Bartoli, and F. Duddeck. “Bayesian optimization of cooperative components for multi-stage aero-structural compressor blade design”. Submitted to: *Structural and Multidisciplinary Optimization* on October 14, 2024.

This article is currently under review.

Associate Professorship of Computational Solid Mechanics
TUM School of Engineering and Design
Technical University of Munich

Bayesian optimization of cooperative components for multi-stage aero-structural compressor blade design

Preprint

Lisa Pretsch[1] ✉, **Ilya Arsenyev**[2], **Nathalie Bartoli**[3], **and Fabian Duddeck**[1]

[1]TUM School of Engineering and Design, Technical University of Munich
[2]MTU Aero Engines AG
[3]ONERA, University of Toulouse

✉ lisa.pretsch@tum.de

October 14, 2024

Abstract In turbomachinery axial compressor development, detailed multi-stage 3D blade optimizations enable better designs than their common single-stage counterparts. However, this design problem entails a prohibitive computational effort for industrial application. It is very high-dimensional, highly constrained, and involves expensive aerodynamic and structural design evaluations. Bayesian optimization (BO) is well suited for the two latter characteristics but suffers from the curse of dimensionality. In contrast to previous approaches, we exploit the problem's multi-component structure to overcome this challenge. We propose to decompose the overall optimization task into lower-dimensional component subproblems. Component interactions are fully taken into account by a sequential cooperative procedure, named cooperative components BO (CC-BO). This enormously facilitates the BO without the need to modify the overall system evaluations. Additionally, we propose a variant with random subproblem decomposition. We analyze the working mechanisms of informed and random CC-BO and compare their performance to standard BO and state-of-the-art algorithms on two problems: a 100D multi-component Branin function and a 223D 4-stage aero-structural compressor blade design. Both proposed approaches significantly enhance the originally poorly performing BO and largely outperform the compared algorithms. The informed version shows an additional advantage for the well-known multi-component structure of the analytical problem. The increased flexibility of random CC-BO makes it the best choice for an efficient global optimization of the multi-stage blade design. It can be readily applied to other high-dimensional constrained optimization tasks.

Keywords Bayesian optimization, Efficient global optimization, Problem decomposition, Multi-disciplinary design optimization, Turbomachinery, Blade design

Nomenclature

$\boldsymbol{\theta}$	hyperparameters vector
$\boldsymbol{h}$	constraints vector
$\boldsymbol{x}$	design variables vector
ψ	reduced likelihood function
c	number of components
d	number of design variables
f	objective function
m	number of response variables
n	number of sample points
q	number of parallel infill points

1 Introduction

Sustainable aviation requires more efficient aircraft engines. Their development can be supported by numerical optimization [1]. In classical axial turbomachinery design, a preliminary multi-stage flowpath and blading optimization is followed by detailed single-stage airfoil shape optimizations. The first are commonly based on 1D mean-line and 2D throughflow analyses, while the latter rely on 3D computational fluid dynamics (CFD) simulations.

However, stage interface boundary conditions that are fixed after the low-fidelity predesign restrain improvements in the subsequent high-fidelity detailed design step. Detailed multi-stage optimizations based on 3D CFD would allow to improve not only the individual stage aerodynamics but also inter-stage effects. Besides, additional structural mechanics constraints must be imposed to guarantee safe and reliable operation. The resulting multi-stage aero-structural design problem is challenging to optimize, because it contains a high number of design variables, many constraints from multiple disciplines, and computationally expensive high-fidelity simulations.

The high-dimensional design space can be well handled by gradient-based optimization algorithms. Wang

et al. [2, 3] enabled multi-stage steady aerodynamic optimizations by proposing an adjoint treatment of the mixing planes, a boundary condition formulation for the blade row interfaces. Ma et al. [4] introduced an unsteady adjoint formulation for multi-stage turbomachinery optimization for a more accurate simulation of the blade row interactions. More recent adjoint multi-stage turbomachinery optimization publications mostly focus on 1.5-stage applications [5, 6, 7]. Drawbacks of gradient-based optimization are its high computational cost in case of many constraints, and its huge memory requirement for large control volumes. Besides, robust adjoint solvers may not be available for all disciplines and finite difference computations are prohibitively expensive.

This makes gradient-free optimization methods a popular choice. Schnoes et al. [8] combined low-fidelity 2D streamline curvature computations with high-fidelity 3D RANS CFD simulations in a Co-Kriging-based optimization approach. For a 4-stage aero-structural compressor design with 119 design parameters, the 2D methods were not accurate enough for a benefit compared to a purely high-fidelity optimization. Nevertheless, their high-fidelity bi-objective optimization via AutoOpti [9] converged in less than 1200 high-fidelity evaluations. Baert et al. [10] performed an aerodynamic optimization of a 4-stage low-pressure turbine by a radial basis function network-based evolutionary algorithm. Despite the high number of 352 design parameters, they obtained an almost 50% higher efficiency gain compared to sequential stage-wise optimizations. Zhang et al. [11] optimized the blade geometries in a 3.5-stage transonic axial compressor with variable inlet guide vanes (IGVs) and stators at two operating points. They used a metamodel-interpreted data mining method to reduce the 103 design parameters to 15 key variables and optimized these by a surrogate-based multi-objective approach. Efficiency and stability improvements were shown to be achieved by inter- and intra-stage loading redistribution. This shows once more the need for efficient multi-stage optimization procedures.

The target application in this work is a detailed parametric blade geometry optimization of the 4-stage transonic high-pressure compressor (HPC) frontblock shown in Figure 1. The frontblock's efficiency should be maximized while fulfilling constraints from geometry, static structural mechanics, modal analysis, and aerodynamics. To obtain these multi-disciplinary response values, new designs are automatically evaluated using the aero-structural simulation workflow

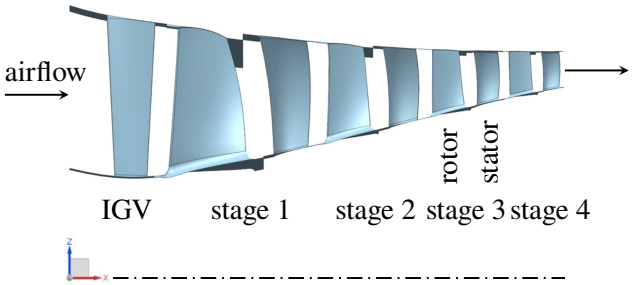


Figure 1 High-pressure compressor frontblock annulus geometry composed of aerodynamically coupled stages. Each stage consists of one row of rotor blades followed by one row of stator vanes. The first stage additionally includes an inlet guide vane (IGV).

shown in Figure 2. As the workflow indicates, we focus on the overall aerodynamic stage interactions and the structural behavior of the first two rotors. Gas loads are kept constant during the optimization, because the aero-structural coupling effects are usually much less influential than the centrifugal loads. Our generic industrial optimization problem amounts to 223 design variables, 43 nonlinear constraints, and about 2 hours per high-fidelity evaluation.

Multi-stage turbomachinery optimization can hence be classified as a high-dimensional, constrained, and expensive-to-evaluate problem. We focus on surrogate-based optimization methods, more precisely Bayesian optimization (BO), because they are well suited for the two latter characteristics. BO [12] is a global optimization approach, that iteratively identifies new infill points to be evaluated by the high-fidelity model based on an adaptive probabilistic surrogate model. Like most gradient-free methods, it suffers from the curse of dimensionality, that is a poor scalability to a large number of design variables [13].

Shifting from the purely mathematical to the engineering point of view, our design problem can also be classified as multi-component and multi-disciplinary. The design space can be split into coupled components, while the response space can be divided into uncoupled disciplines. The simulation workflow in Figure 2 additionally indicates which component design variables affect which discipline response variables. This information is illustrated in Table 1. Even if the multi-stage aero-structural optimization may not be a typical multi-disciplinary optimization (MDO) [14] problem, similar approaches can be applied. Distributed MDO architectures [15] exploit the problem structure to decompose the overall optimization task into several smaller and simpler interconnected sub-problems.

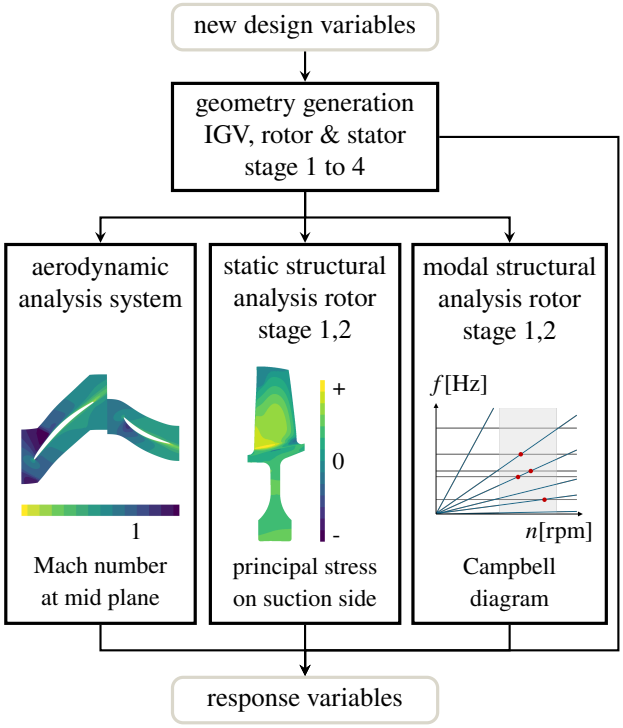


Figure 2 Aero-structural high-fidelity evaluation workflow composed of uncoupled disciplinary simulations. The blade geometries are generated row by row. The aerodynamic simulations are performed for the overall 4-stage system. The structural behavior is computed separately for the most highly loaded rotors in the first two stages.

The innovative idea in this work is to use the concept of distributed MDO to tackle the challenge of sample-efficient BO with a large number of design variables and constraints. The resulting methodological goal is to develop a novel interconnected problem decomposition approach for BO. It has the major benefit of smaller subproblems at optimization level, without requiring a system decomposition at evaluation level. Thereby, we enable efficient global multi-stage aero-structural turbomachinery blade optimizations.

We propose a cooperative components BO (CC-BO), inspired by distributed MDO strategies and the cooperative efficient global optimization (CoEGO) approach by Zhan et al. [16]. We decompose the overall optimization task into component subproblems according to the known underlying structure. Component interactions are fully taken into account by a sequential cooperative procedure. The subproblems are simpler to solve, because of lower-dimensional component design spaces, partly lower-dimensional surrogate models, and a reduced number of component constraints. This makes CC-BO an effective way to incorporate the problem structure information for a

Table 1 Component-to-discipline dependency information for multi-stage aero-structural blade design. Stage 1 includes the IGV.

discipline (group of response variables y)	stage 1	stage 2	stage 3	stage 4
aero	■	■	■	■
static struct 2		■		
static struct 1	■			
modal struct 2		■		
modal struct 1	■			
geo 4				■
geo 3			■	
geo 2		■		
geo 1	■			

component (group of design variables $\boldsymbol{x}$)

significantly faster convergence over the number of iterations.

The proposed approach is motivated by a particular problem structure and requires Boolean information on the relation of design variables to components and response variables. For the case that no information is available, we additionally propose a generalized approach, with a random design space decomposition into a predefined number of components in every BO iteration. It outweighs the loss of information by the advantage of increased flexibility. Depending on the problem at hand, both informed and random CC-BO can hence be the better method for high-dimensional, constrained, and expensive-to-evaluate optimization.

The rest of this paper is structured as follows: Section 2 outlines various optimization approaches, as context for the proposed CC-BO method presentation in Section 3. Afterwards, in Section 4, we develop a multi-component analytical problem and perform comparisons to state-of-the art algorithms, as well as an analysis of the CC-BO working mechanisms. Section 5 provides an optimization problem description and in-depth result analyses for the multi-stage aero-structural compressor blade design. The article is concluded by a summary and outlook in Section 6.

2 Optimization approaches

We assume an optimization problem of the form

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && \boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0} \\ &\text{by varying} && \boldsymbol{x}_{\text{lb}} \leq \boldsymbol{x} \leq \boldsymbol{x}_{\text{ub}}, \quad \boldsymbol{x} \in \mathbb{R}^d. \end{aligned} \tag{1}$$

Our m quantities of interest are a single objective function $f(\boldsymbol{x})$ and a set of $m-1$ inequality constraints $\boldsymbol{h}(\boldsymbol{x})$. We seek for a feasible minimum by varying a d-dimensional design variable vector $\boldsymbol{x}$.

2.1 Bayesian optimization

BO [12] is a popular method in engineering design, where optimization problems are often constrained and contain expensive-to-evaluate functions. In this context, it is also known as efficient global optimization (EGO) [17]. The basic process is illustrated by the flowchart in Figure 3, omitting all colored parts. In this section, we specify the employed methods and settings step by step. For a deeper understanding, we refer the reader to the review on surrogate-based optimization in general by Forrester and Keane [18] and BO in particular by Garnett [19].

DoE. The initial step is a design of experiments (DoE) that distributes points in the design space. For the analytical test problem, we use an optimized Latin Hypercube Sampling (LHS) [15] with uniform distribution in each direction. For the turbomachinery problem, we opt for a LHS with normal distribution centered around the initial design to reduce the number of simulation failures due to extreme designs.

High-fidelity evaluation. The responses for the sample points are then computed by high-fidelity evaluations, in the compressor case according to the workflow in Figure 2. Points with failed simulations are filtered out. The resulting sample set of design points and corresponding response values acts as training set for the surrogate model.

Best design update. The current best design is the sample point with the best objective function value, provided that it satisfies all constraints. It is updated at the end of every BO iteration.

GP model fit. In every iteration, the Gaussian process (GP) surrogate models, also referred to as Kriging models, are fitted based on all available sample points. Our a-priori GP is the sum of a constant mean and a stationary GP, defined by its variance and squared exponential correlation function, also known as kernel. The model fit is a log-marginal-likelihood function ψ maximization to optimize the d-dimensional kernel hyperparameter vector $\boldsymbol{\theta}$. It is performed by a multi-start Cobyla algorithm [20]. The resulting a-posteriori GP models not only provide a response prediction, but also a model uncertainty estimate for every point in the design space. We use the implementation from the surrogate modeling toolbox (SMT) [21].

Infill criterion optimization. Based on the cheap-to-evaluate GP models, we then search for new promising points to be evaluated by the high-fidelity model. Therefore, an infill criterion or acquisition function is optimized, here the product of expected improvement (EI) and the individual probabilities of feasibility (PF), yielding the constrained expected improvement (CEI) [17]. A multi-start Nelder-Mead simplex algorithm is used as optimizer. It yields multiple local optima, that are clustered [22] to choose a batch of q infill points to be evaluated in parallel with the expensive high-fidelity model.

Repeat until termination criterion is met. BO is an iterative process, which is repeated until a termination criterion, here a maximum number of high-fidelity evaluations, is met.

2.2 High-dimensional Bayesian optimization

With higher design space dimension, the volume increases exponentially and the distance between a constant number of sample points grows. This effect is known as the curse of dimensionality. It impedes the good performance of BO for high-dimensional problems, affecting the two main building blocks of BO. First, the GP models become inaccurate and the fit becomes slow for a fixed number of training points. Second, the infill criterion optimization becomes complicated and inefficient. The infill criterion is generally difficult to optimize, because of its many local maxima in promising unexplored regions and large plateaus in the other parts, and even more so for a high design space dimension.

These challenges and recent approaches for high-dimensional BO were summarized by Binois and Wycoff [13]. They distinguish four groups of remedies: additive models, (non-)linear embeddings, trust region approaches, and variable selection. The assumption of an additive objective function [23] does not hold for the compressor efficiency [24]. The assumption of a low intrinsic problem dimensionality [25] is also not reasonable for detailed compressor blade design [26]. Besides, literature on high-dimensional BO methods for constrained problems is comparatively sparse. Transformations into unconstrained problems via penalty approaches are not suitable for the complex constraints of many engineering design problems [27]. Therefore, some originally unconstrained methods were extended to handle constrained problems.

An example for linear embeddings is PCA-BO [28], where a lower-dimensional design space is determined

by a principal component analysis (PCA) of weighted samples. In its extension to constrained problems by the authors [26], the infill criterion optimization is performed in this lower-dimensional subspace. The approach was shown to significantly speed up the convergence in the first BO iterations on both a 40D constrained analytical problem and a 55D compressor blade design.

The arguably most popular trust region BO approach, TuRBO [29], was extended to scalable constrained BO (SCBO) [30] by Eriksson et al. It combines three building blocks. First, it counteracts overexploration by confining the design space inside a trust region. Second, it treats heterogeneous functions by transforming objective and constraints so as to emphasize optimum and feasible regions, respectively. Third, it enables a fast generation of large infill point batches in each iteration by Thompson sampling instead of CEI maximization. SCBO was shown to outperform state-of-the-art methods on highly constrained test problems up to 124D. We use the version provided by BoTorch [31] with the settings from the SCBO paper [30].

Tailored to turbomachinery design, the BO-like optimizer AutoOpti [9] by the German Aerospace Center (DLR) is productively used in industry. It employs GP models to approximate high-fidelity responses. A multi-objective evolutionary strategy is used to optimize a combination of infill criteria, including EI. Its performance is enhanced by various state-of-the-art techniques, such as automatic selection of relevant input dimensions, robust GP training, trust regions, and asynchronous parallel infill points generation. AutoOpti was shown to perform well on blade design problems with up to 230 design parameters [9]. It was also successfully employed for a 4-stage blade optimization [8].

With regard to our idea of an informed problem decomposition, the cooperative approach to EGO (CoEGO) by Zhan et al. [16] provides a good basis. CoEGO does not include any problem information and therefore randomly decomposes the high-dimensional design space into a number of lower-dimensional subproblems. The GP model fit and the infill criterion optimization are then performed sequentially for each subproblem. Context vectors of GP hyperparameters and design variables maintain the cooperative connection between the subproblems. After each subproblem solution, a high-fidelity evaluation is run and the context vectors are updated accordingly. Once all subproblems are done, the process is repeated, starting with a new random design space decomposition.

CoEGO was tested on a set of analytical 100D functions [16]. It was shown to yield a nearly linear reduction in optimization time as the number of subproblems increases. However, the convergence over the number of high-fidelity evaluations slowed down with a growing number of subproblems. Both of these effects were enhanced by parallel instead of sequential subproblem optimizations. The final optimization results of CoEGO were competitive with various high-dimensional BO approaches.

2.3 Multi-disciplinary and multi-component optimization

Many large-scale problems in engineering design relate to systems with multiple disciplines or components. Disregarding their structure, they can be treated as a common optimization problem with a multi-disciplinary model and solved by the above methods. This approach is called multi-disciplinary feasible (MDF) and belongs to the group of monolithic MDO architectures [32].

The term MDO encompasses all kinds of interactions, no matter if between components or disciplines. The underlying idea is that optimizing a coupled system, like the multi-stage compressor, can yield better designs in a shorter time than sequential optimizations of the uncoupled parts, for instance the single compressor stages. Martins and Lambe [32] provide an overview of MDO methods for engineering design.

For large problems, the group of distributed MDO architectures seems promising. They decompose the optimization problem into multiple smaller subproblems, which only consider a part of the design variables and constraints. The idea is to speed up the optimization by exploiting the problem structure. The decomposition mimics the traditional industrial development process, where complex tasks are split into manageable sub-tasks to be performed by disciplinary design teams. Despite the promising idea, distributed architectures are rarely used because of their slow convergence [15]. This is due to the high coordination effort, that is required to guarantee the multi-disciplinary consistency and usually enforced by an additional system-level optimization.

3 Proposed approach: cooperative components Bayesian optimization

In contrast to most common MDO problems, our aim is not to decouple the entire system, for example into

single-stage analyses and optimizations. Instead, we only want to split the optimization, for instance into single-stage subproblems. The component subproblems have fewer design variables and constraints and can thus be more efficiently solved by BO. Multi-component consistency is thereby always fulfilled for the analyses and only needs to be maintained inside the optimizations, resolving the main disadvantage of distributed MDO approaches.

We propose a cooperative components BO (CC-BO), which combines the above ideas of distributed MDO and CoEGO. Similar to distributed MDO, we split the coupled problem into component subproblems according to the known underlying structure. Similar to CoEGO, we perform sequential subproblem optimizations linked by context vectors.

CoEGO provides a methodological basis but requires significant changes to fit our purpose. First, we shift the focus from an accelerated optimization algorithm to a fast convergence over the number of iterations, which is decisive in case of expensive high-fidelity evaluations. Second, we make sure that every high-fidelity evaluation provides entirely new results despite the non-random decomposition. Therefore, we only perform a high-fidelity evaluation once all component subproblems were solved. The modification also entails a different context vector update. Moreover, we add constraint handling and batch evaluation capabilities. They are required in most engineering problems and to exploit modern computational environments, respectively.

The problem structure information is provided in the form of two Boolean matrices $\{0,1\}^{d\times c}$ and $\{0,1\}^{d\times m}$. They describe which design variables belong to which components and which design variables affect which responses. Accordingly, d is the number of design variables, c is the number of components, and m is the number of responses, that is the objective and constraints. We introduce the index $i = 1,\dots,c$ of the currently active component, that has a reduced number of design variables $d_i < d$ and responses $m_i < m$. Moreover, $j = 1,\dots,m$ indexes the GP models and $k = 1,\dots,q$ the infill point batches. Figure 3 illustrates the proposed CC-BO process. The steps that differ to standard BO are highlighted in green and will be explained in what follows.

Shuffle component order. In each CC-BO iteration, we randomly shuffle the component order before performing the sequential component optimizations. This guarantees a well-balanced influence of all components and can hence enable better overall results.

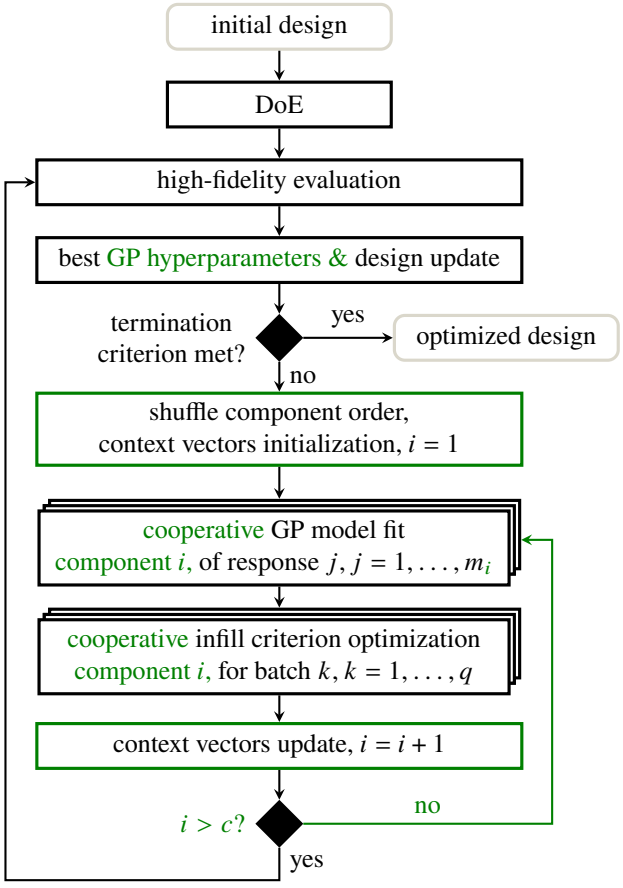


Figure 3 Flowchart of the proposed CC-BO approach, with the difference to the reference BO highlighted in green.

Context vectors initialization. The sequential subproblem optimizations consist of a GP model fit and an infill criterion optimization for each component $i = 1,\dots,c$. They are BO-internal maximization problems of the marginal likelihood $\psi(\boldsymbol{\theta})$ and the CEI($\boldsymbol{x}$), respectively. Both the GP model hyperparameters $\boldsymbol{\theta} \in \mathbb{R}^d$ and the design variables $\boldsymbol{x} \in \mathbb{R}^d$ are high-dimensional vectors. In CC-BO, these internal optimizations are performed by varying only the respective component parameters and fixing all others to so-called cooperative context vectors. The context vectors thereby establish the connection to the other subproblems. They are initialized with the current best parameters. We set the GP hyperparameter context vectors to $\boldsymbol{\theta}_{\text{coop}} = \boldsymbol{\theta}_{\text{best}}$ for each response $j = 1,\dots,m$ and initialize the infill point context vector as $\boldsymbol{x}_{\text{coop}} = \boldsymbol{x}_{\text{best}}$.

Cooperative components GP model fit. The m_i GP models are then fitted in parallel. The component i training set is $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, with the set of n samples of component design parameters $\boldsymbol{X}_i \in \mathbb{R}^{n\times d_i}$ and responses $\boldsymbol{Y}_i \in \mathbb{R}^{n\times m_i}$. The cooperative GP model hy-

perparameter fit of component i can then be formulated as

$$\begin{aligned} &\text{maximize} && \psi(\boldsymbol{\theta}) \\ &\text{by varying} && \boldsymbol{\theta}_{\mathrm{lb},i} \leq \boldsymbol{\theta}_i \leq \boldsymbol{\theta}_{\mathrm{ub},i}, \quad \boldsymbol{\theta}_i \in \mathbb{R}^{d_i} \\ &\text{while fixing} && \boldsymbol{\theta}_l = \boldsymbol{\theta}_{\mathrm{coop},l}, \quad l \neq i, l = 1, \ldots, c. \end{aligned} \tag{2}$$

The hyperparameter vector $\boldsymbol{\theta}$ is composed of the variable component hyperparameter vector $\boldsymbol{\theta}_i$ and the cooperative context vector $\boldsymbol{\theta}_{\mathrm{coop},l}$ for all other components $l \neq i$. The cooperative components GP model fit was added as an application in the version 2.7.0 of the open-source SMT [21].

Besides, there may be responses that exclusively depend on the component i design variables. Including this problem information, they can be modeled by ordinary GPs. Their lower dimensionality yields more accurate models. As an example, we look at stage 1 from our turbomachinery problem. With the information from Table 1, only models of four instead of all nine disciplines' responses need to be fitted. Only the aerodynamic responses are modeled by a cooperative GP. Geometry, modal, and structural responses can be approximated by lower-dimensional ordinary GP models.

Cooperative components infill criterion optimization. Based on the surrogate models, we search for a batch of q component infill points to be evaluated in parallel. For the first component $i = 1$, there is a single context vector $\boldsymbol{x}_{\mathrm{coop}} = \boldsymbol{x}_{\mathrm{best}}$. Therefore, the first component $i = 1$ search is a single optimization. It yields q component infill points and hence q design variable context vectors to continue with. For the subsequent components $i > 1$, a separate infill point search is run in parallel for each of the q context vectors. Each of the parallel searches returns a single infill point, for a total of q infill points to be evaluated in parallel after the last component $i = c$. Similar to the cooperative GP fit in Equation (2), the cooperative constrained infill criterion optimization of component i is formulated as

$$\begin{aligned} &\text{maximize} && \mathrm{CEI}(\boldsymbol{x}) = \mathrm{EI}(\boldsymbol{x}) \cdot \prod_{j=2}^{m} \mathrm{PF}_j(\boldsymbol{x}) \\ &\text{by varying} && \boldsymbol{x}_{\mathrm{lb},i} \leq \boldsymbol{x}_i \leq \boldsymbol{x}_{\mathrm{ub},i}, \quad \boldsymbol{x}_i \in \mathbb{R}^{d_i} \\ &\text{while fixing} && \boldsymbol{x}_l = \boldsymbol{x}_{\mathrm{coop},l}, \quad l \neq i, l = 1, \ldots, c. \end{aligned} \tag{3}$$

Moreover, there may be constraints that do not depend on the varied component i design parameters. Throughout the component i infill search, they remain constant and so does their PF. This additionally speeds up the infill search by less surrogate model evaluations and simplifies the CEI function to be maximized. Considering again stage 1 as an example, according to Table 1, only the responses of four of the nine disciplines affect the CEI for the single-stage infill point search.

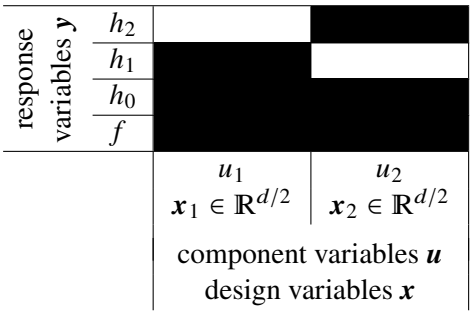
Context vectors update. After each component i optimization, the context vectors are updated with the optimized quantities. For the cooperative GP models, $\boldsymbol{\theta}_{\mathrm{coop},i} = \boldsymbol{\theta}_i^*$ for each component response $j = 1, \ldots, m_i$. The infill point context vectors are updated as $\boldsymbol{x}_{\mathrm{coop},i} = \boldsymbol{x}_{\mathrm{infill},i}$ for each batch $k = 1, \ldots, q$.

In summary, the informed problem decomposition has several advantages for BO in a high-dimensional design space, with a high number of constraints, and expensive high-fidelity evaluations. Concerning the high-dimensional design space, the series of lower-dimensional component subproblems simplify the GP model fit and infill criterion optimization. Moreover, information on design-to-response variable dependencies allows for partly lower-dimensional and thus more accurate GP models. With the same information, also the large number of constraints can be partly decomposed, reducing the number of non-constant constraints per subproblem. This additionally accelerates and simplifies the component infill criterion optimizations. Finally, the combination of the above effects in sequential cooperative component optimizations allows for a fast convergence over the limited number of iterations, which is the main goal of CC-BO.

Although the structure is problem-specific, the proposed decomposition approach is generally applicable, even without information about the problem structure. If the allocation of design variables to components is unclear, CC-BO can be applied with a new random split in a user-defined number of c components in every BO iteration. We call this generalization random CC-BO, as opposed to the informed CC-BO described above. A small c can simplify the cooperation, whereas a small component dimension d_i can improve the component optimizations. The subproblem dimension should therefore be chosen as high as still well manageable by BO, for instance $d_i = 10$.

If the relation of design to response variables is unknown, they can be assumed to be fully dependent. Consequently, the advantages of partly lower-dimensional GP models and fewer non-constant component constraints are lost, but the main working mechanism of simplified lower-dimensional component BOs is maintained.

Table 2 Design-to-response variable dependency information for multi-component Branin function.



4 Multi-component Branin function optimization

We first assess our approach on an analytical test function. It is easily reproducible and enables in-depth analyses of our method's performance and working principles with much less computational effort compared to a simulation-based problem.

4.1 Branin optimization problem

The test problem should have similar characteristics to our turbomachinery problem. This means it must be subject to constraints, exhibit a multi-component structure, and have a high-dimensional design space. We choose the popular two-dimensional modified Branin test function with Gomez#3 constraint, defined by Parr et al. [27], as a basis. For the multi-component structure, we decompose the Gomez#3 function into a set of 3 constraints, where $h_{0,2D}$ depends on both u_1 and u_2, $h_{1,2D}$ depends only on u_1, and $h_{2,2D}$ is only a function of u_2. This problem structure is illustrated in Table 2, analogous to our turbomachinery problem structure in Table 1. The resulting objective and constraint functions are formulated in Equation (4).

$$
\begin{aligned}
f_{2D}(\boldsymbol{u}) &= \left(u_2 - \frac{5.1}{4\pi^2} + \frac{5}{\pi}u_1 - 6\right)^2 \\
&\quad + 10\left[\left(1 - \frac{1}{8\pi}\right)\cos u_1 + 1\right] + \frac{5u_1 + 25}{15} \\
h_{0,2D}(\boldsymbol{u}) &= u_1 u_2 + 0.2 \\
h_{1,2D}(\boldsymbol{u}) &= \left(4 - 2.1u_1^2 + \frac{1}{3}u_1^4\right)u_1^2 + 3\sin[6(1-u_1)] + 3 \\
h_{2,2D}(\boldsymbol{u}) &= \left(-4 + 4u_2^2\right)u_2^2 + 3\sin[6(1-u_2)] + 1
\end{aligned}
\tag{4}
$$

The objective function f_{2D} is defined on the two-dimensional design space $u_1 \in [-5, 10]$ and $u_2 \in [0, 15]$, the multi-component constraint $h_{0,2D}$ is defined on $u_1, u_2 \in [0, 1]$, and the single-component constraints $h_{1,2D}, h_{2,2D}$ are defined on $u_1, u_2 \in [-1, 1]$. All functions are normalized to $u_1, u_2 \in [-1, 1]$. The constraints in Equation (4) are already standardized for $h \leq 0$ to be consistent with the optimization problem formulation in Equation (1).

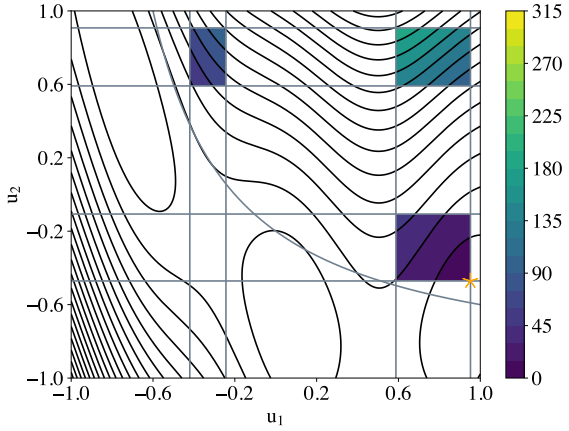


Figure 4 Branin function with multi-component constraints. The black contours represent the objective function, the gray lines show the constraint bounds, and the colored parts indicate the feasible regions. The global optimum is at the yellow star.

Figure 4 illustrates the resulting two-dimensional problem. The starting point at $u_1 = u_2 = 0.8$ is feasible, like for the turbomachinery problem. Similar to the original problem by Parr et al. [27], we have two local and one global minimum, each of which is associated to a feasible region. The three feasible regions are disconnected and include different ranges of function values. The local optimum in the initial feasible region is at $f = 91.48480$, the second local minimum is at $f = 42.56271$, and the global optimum is at $f^* = 7.20185$ with $u_1^* = 0.95151$ and $u_2^* = -0.47102$.

The high-dimensional design space $\boldsymbol{x} \in [-1, 1]^d$ is obtained by an artificial scaling, similar to the high-dimensional unconstrained modified Branin function by Priem et al. [33]. In order to keep our two-component problem structure, the random scaling matrix $\boldsymbol{A} \in \mathbb{R}^{2\times d}$ of Priem et al. [33] is replaced by a separate vector $\boldsymbol{a}_i \in \mathbb{R}^{d/2}$ for each of the two components $i = 1, 2$. This vector is generated randomly and then normalized such that $u_i = \boldsymbol{a}_i^T \boldsymbol{x}_i \in [-1, 1]$ for all $\boldsymbol{x}_i \in [-1, 1]^{d/2}$. The objective function is thereby evaluated as $f(\boldsymbol{x}) = f_{2D}(\boldsymbol{a}_1^T \boldsymbol{x}_1, \boldsymbol{a}_2^T \boldsymbol{x}_2)$ and the three constraints are computed analogously.

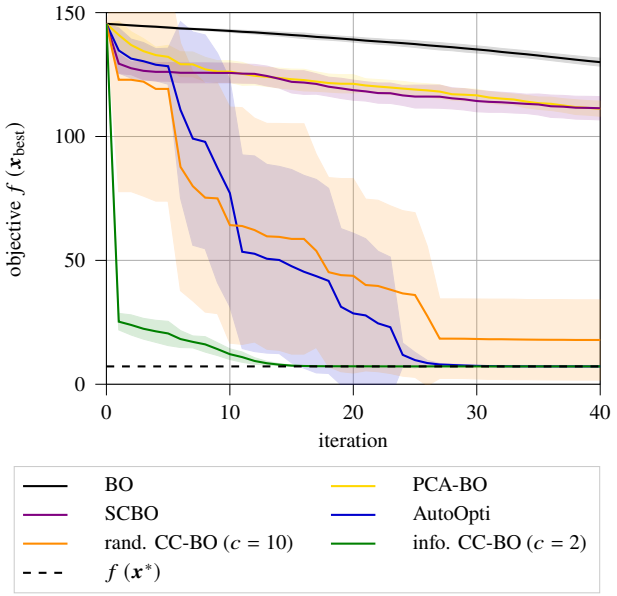


Figure 5 Convergence of various BO algorithms, including the proposed CC-BO, for the 100D multi-component constrained Branin test function minimization. The mean and standard deviation of the best feasible objective function values of 10 runs per algorithm are shown.

4.2 Branin optimization results

We compare our reference BO and the proposed CC-BO to various state-of-the art high-dimensional constrained BO algorithms. Moreover, we analyze the individual and joint effects of the distinguishing elements of our CC-BO algorithm on the overall convergence. The high dimension $d = 100$ is chosen for a similar component design space dimension $d_i = 50$ as for our turbomachinery problem. Each algorithm is run 10 times with a different seed, which yields 10 different initial DoEs. We start the BOs with $d+1 = 101$ initial LHS points and set the maximum number of high-fidelity evaluations to $5d+1 = 501$. With a batch of $q = 10$ infill points per iteration for all algorithms, this is equivalent to 40 optimization iterations.

Besides BO and informed and random CC-BO, we compare constrained PCA-BO, SCBO, and AutoOpti, which were outlined in Section 2.2. Figure 5 shows the convergence mean and standard deviation of the 10 runs each. All six methods find feasible improvements of the objective function. However, BO, PCA-BO, and SCBO do not even find the highest local minimum in the limited number of iterations. The poor convergence of the reference BO was to be expected, see Section 2.2.

PCA-BO [26] shows a faster initial convergence than BO but quickly slows down. The design space dimension is reduced to 53 or 54 principal components, despite a captured variance ratio of 99%. This indicates that $d+1$ initial samples are not sufficient for a good PCA with original design space dimension $d = 100$.

SCBO [30], in contrast, does not actually operate in a reduced design space, as the trust region side length never falls below the initial value. The reason might be the large failure tolerance of $\tau_f = d/q = 10$, meaning that the side length could be reduced no sooner than every tenth iteration. With a smaller tolerance of $\tau_f = 3$, an improved objective function value of around 100 could be reached. Besides, the Thompson sampling makes it by far the fastest among the tested algorithms.

AutoOpti [9] reliably reaches the global minimum within 40 iterations. This confirms its good performance on high-dimensional and highly constrained problems, which are typical for turbomachinery design. The big standard deviations in the runs with AutoOpti can be explained by the different objective function values of the three feasible regions, see Figure 4. This results in a jump of the objective function value when a better feasible region is found.

Random CC-BO performs a new random split in $c = 10$ components with $d_i = 10$ in every iteration for. It converges to the global minimum in 7 of 10 runs. This can be understood as a proof of concept, that sequential component optimizations connected by cooperative context vectors can counteract the curse of dimensionality and thereby enable high-dimensional BO. Therefore, the cooperative components approach could also be applied to problems without a particular problem structure.

Informed CC-BO runs with $c = 2$ components, according to the problem structure in Table 2. It already finds the feasible region that contains the global optimum in the first iteration and then only takes about 15 iterations to reliably find the global optimum. It is important to state that the other high-dimensional BO algorithms are not generally inferior, as informed CC-BO has the advantage of knowing the problem structure. The results rather show that our CC-BO approach is a very effective way of incorporating this information. Despite being based on our indeed inferior reference BO, it can outperform state-of-the art algorithms like SCBO or AutoOpti.

In addition to the overall CC-BO performance, we study the effect of the two cooperative building blocks on the convergence and wall time. We compare all four combinations of ordinary and cooperative GP model fits and infill searches to analyze their individual and joint influence. The term ordinary refers to the building blocks used in BO, as described in Section 2.1.

Additionally, we distinguish whether problem information is available or not. The upper part of Figure 6 shows the resulting convergence of the feasible best point over the number of iterations.

In the versions without information in Figure 6a, the number of components is set to $c = 2$ for a meaningful comparison. Even without any information, all runs with one or both cooperative building blocks do at least reach the highest local minimum at $f \approx 91.5$ in the first 20 iterations. Afterwards, at least one run of each combination also finds the global optimum.

As shown in Figure 6b, problem information alone has a negligible effect on the convergence, even if it might make the surrogate models more accurate by providing design-to-response variable dependencies. However, it greatly enhances the positive impact of cooperative GP model fits and infill searches, where it additionally provides information on how to decompose the design space. All of these informed runs with one or both cooperative building blocks show a very steep initial convergence. The version with only cooperative model fits gets stuck at the second local minimum at $f \approx 42.6$ in 6 out of 10 runs. The two others, with the green curve overlapping the blue one, converge to the global optimum at $f \approx 7.2$ in at least 20 iterations.

The lower part of Figure 6 shows the mean wall times per optimization with standard deviations. All computations are performed on ten Intel(R) Xeon(R) Gold 6142 CPUs @ 2.60GHz. Qualitatively, the information mostly accelerates the cooperative procedures. Cooperative component model fits additionally speed up the algorithms, whereas the cooperative infill searches take longer. The numbers can be interpreted as algorithm run times, because the high-fidelity analytical function evaluations are negligibly fast. Compared to the reference BO, informed CC-BO takes 27% longer and random CC-BO adds 73% wall time. All tested algorithms stay below 2 minutes per high-fidelity evaluation. This is not critical when simulation times of several hours are expected, like for our engineering test case.

The main performance indicator is thus the convergence over the number of iterations. The above results without information show that the cooperative components approach is advantageous for high-dimensional design spaces. If problem information is available, CC-BO is a highly effective way to incorporate it, even for the simplest possible structure of our test problem with only two components and three constraints.

5 Multi-stage aero-structural HPC blade design

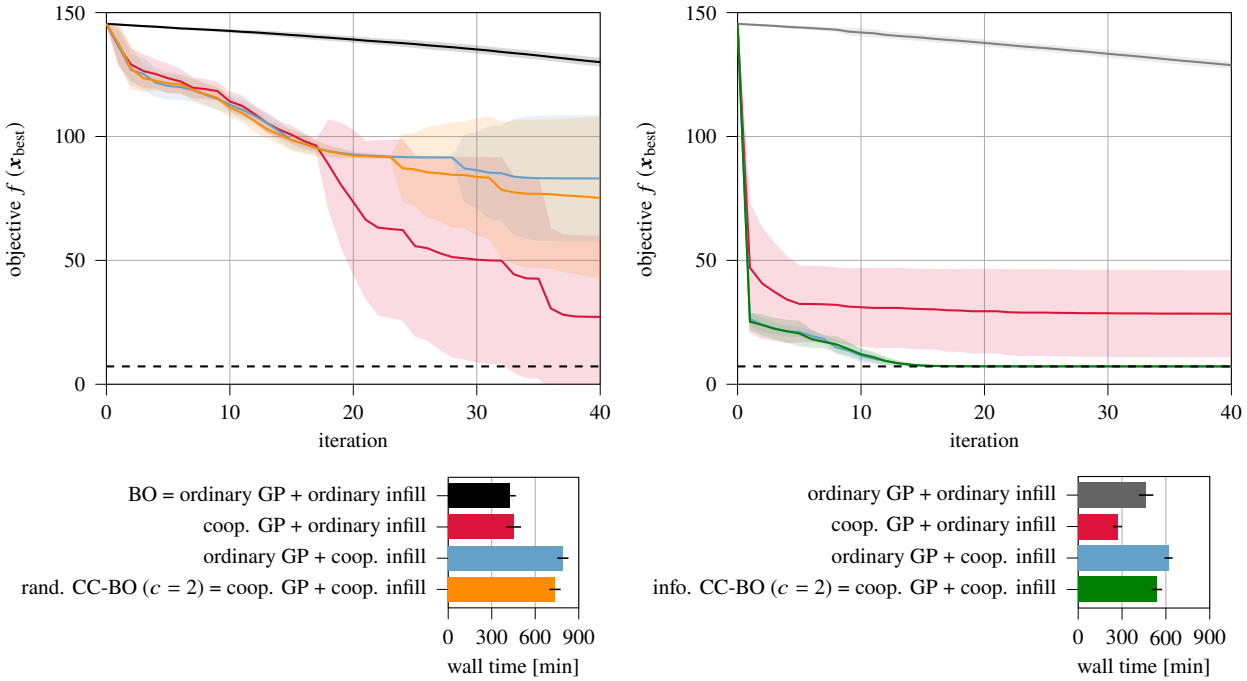
The motivation for CC-BO comes from the industrial engineering problem of multi-stage aero-structural turbomachinery blade design. Therefore, we assess our algorithms on the detailed design of the first four stages, see Figure 1, of a generic 8-stage HPC that could be part of a modern high-bypass geared turbofan engine.

5.1 Aero-structural design evaluation

Our optimizations are based on 3D blade geometry generations, finite element (FE) analyses, and CFD simulations. The high-fidelity evaluation workflow is illustrated in Figure 2. All building blocks are specified in the following.

Geometry generation. The blade shapes are defined by a set of parameters that also serve as design variables in the optimization. They are illustrated in Figure 7. First, 2D profile parameters are specified at various radial positions, mostly at blade hub, mid, and tip. The locations of the four base points at the intersection of leading edge (LE), pressure side (PS), trailing edge (TE), and suction side (SS), are set with respect to the blade center. They are defined by the blade angles α, the stagger β_S, the wedge angles γ, and the distance wedge l. Second, the splines of the PS and SS segments are refined by symmetric values of the curvature κ and the aspect ratio parameters of slope ρ and curvature σ. The 2D profile centers, highlighted in red in Figure 7, are shifted in axial and circumferential direction. The smooth 3D blade is then obtained by spline interpolation, with the order depending on the number of radial positions. This process of blade geometry generation is performed with in-house tools. It takes about 3 minutes for the nine blade rows in the 4-stage frontblock. The generated hot blade geometries are an input for both structural and aerodynamic simulations. For the structural models, the disk sector geometry is added to the rotor blades. For the aerodynamic model, the blades are assembled to form the multi-stage frontblock.

Static & modal structural analysis. The 3D structural simulations are carried out in the open-source FE program *CalculiX* [35]. The rotor blades operate under centrifugal, gas, and thermal loads. Centrifugal loads are determined in the FE analysis based on rotational speed. Gas and thermal loads are computed in advance, prescribed as radial pressure and temperature distributions, and finally mapped onto the respective



(a) without information

(b) with information

Figure 6 Influence of the CC-BO building blocks (cooperative components GP model fit, cooperative components infill criterion optimization) without (a) and with (b) problem structure information for the 100D multi-component constrained Branin test function minimization. The mean and standard deviation of the best feasible objective function values (top) and the wall time (bottom) of 10 runs per algorithm are shown.

blade. The structural analyses are performed at cruise operating conditions and run about 5 minutes on 6 CPUs per row.

Aerodynamic analysis. The 3D CFD simulations are performed with the steady-state RANS flow solver *TRACE* [36] by the *German Aerospace Center (DLR)* and *MTU Aero Engines AG*. For multi-row aerodynamic analyses, single blade row passages are modeled by the mixing-plane approach. This means that the flow state at one side of the row interface is averaged in peripheral direction and prescribed as a boundary condition to the other side. The converged flow solution has consistent flow states on both sides of the mixing plane. In addition to the aerodynamic design point (ADP) at cruise conditions, an off-design point (ODP) is considered. It is a throttled state with the same rotational speed and a higher back pressure. The CFD simulations are by far the most expensive part of the high-fidelity evaluation process and take on average 119 minutes on 32 CPUs per design point.

5.2 Blade optimization problem

Based on the 3D simulation workflow specified above, we seek for a detailed blade design with maximized efficiency subject to constraints on aerodynamics, static stresses, eigenfrequencies, and geometry. The initial annulus geometry and blade profiles of all four stages were designed via 2D throughflow computations. Moreover, the blades in the first two stages, that are subject to structural constraints, were preoptimized separately for a feasible initial design. Our multi-stage aero-structural compressor blade optimization problem is defined in Table 3. The problem has a single objective, 42 constraints, and 223 design parameters.

The objective is to maximize the isentropic efficiency at cruise conditions. The optimized 4-stage frontblock must fit back into the overall engine. This requires to keep the mass flow and downstream flow angle close to the initial state, which is imposed by narrow double-sided inequality constraints. In order to maintain a good off-design performance, we set a lower limit for the isentropic efficiency and the mass flow at the throttled ODP.

Structural mechanics constraints are imposed on the rotor blades of the first two stages. Being relatively thin and highly loaded, they are most critical for the structural integrity and life of the blisks. We constrain the maximum static principal stresses within the rotor blade and in the fillet region between blade and disk. Moreover, the stress difference between PS and SS is

Table 3 Multi-stage aero-structural compressor blades optimization problem definition.

1 objective $f(\boldsymbol{x})$	**type**	**number**	**discipline**
isentropic efficiency η_{is} at ADP	maximization	1	aero
42 constraints $\boldsymbol{h}(\boldsymbol{x})$	**type of bound**	**number**	**discipline**
mass flow $\dot{m}$ at ADP	lower & upper	1	aero
downstream flow angle α_{Z} at ADP	lower & upper	1	aero
isentropic efficiency η_{is} at ODP	lower (=baseline)	1	aero
mass flow $\dot{m}$ at ODP	lower (=baseline)	1	aero
max. static principal stress $\sigma_{I,\text{max}}$ on rotor surface, fillet	upper	2+2	struct 1+2
max. static principal stress difference $\Delta\sigma_{I,\text{max}}$ between PS and SS on rotor surface, fillet	lower & upper	2+2	struct 1+2
eigenfrequencies	lower &/ upper	9+7	modal 1+2
max. blade thickness at rotor tip	lower & upper	1+1	geo 1+2
max. blade thickness delta at radial slices near rotor tip	lower	4+4	geo 1+2
blade profile constriction	lower	1+1+1+1	geo 1+2+3+4

223 design variables $\boldsymbol{x}$	**number**								
	IGV	R1	S1	R2	S2	R3	S3	R4	S4
see Figure 7	6	51	9	46	9	42	9	42	9

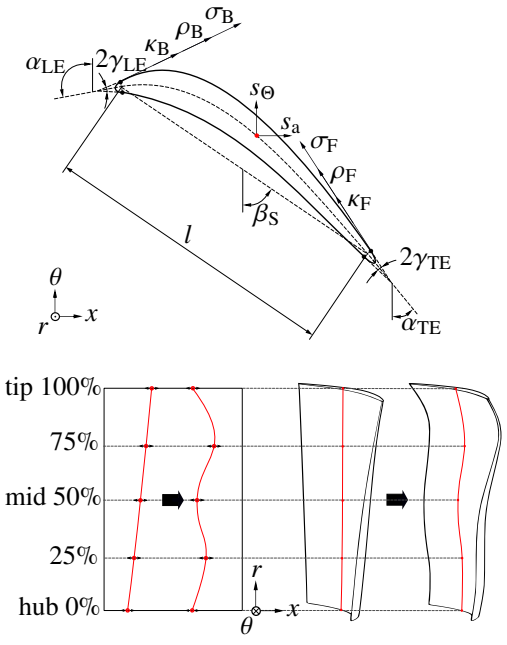


Figure 7 Blade geometry parameterization. Modified from [34].

bounded to guide the optimization to a better balanced blade design. In addition, several rotor eigenfrequencies are constrained to avoid resonances within the engine's operating range. The critical modes and associated frequency ranges are identified based on the Campbell diagram, taking into account fundamental engine orders and excitation from the neighboring stators.

For the first and second stage rotor, the structural feasibility could be enforced by large modifications of the blade thickness distribution in the tip region. This is unintended and is counteracted by a narrow lower and upper bound for the maximum tip thickness. Moreover, we prescribe a decreasing blade thickness in radial direction near the tip. The combination of a high number and wide ranges of our rotor blade design variables can yield unreasonable blade profiles, like ones with negative constriction. A corresponding aggregated constraint for each of the four stages prevents such blade shapes.

The design variables are illustrated in Figure 7 and were described in the previous section. The distribution of design variables in the IGV, rotor blades (R), and stator vanes (S) of the four stages is also listed in Table 3. The rotor blades require a more detailed design than the stator vanes, because they are aerodynamically and structurally more complex. The stators are only modified to adjust to the changing flow conditions due to the rotor blade optimizations. Moreover, the initial geometry of the first rotor is so complex that it requires partly higher order polynomials compared to the second rotor, and has therefore some more design variables. For the rotors of stage 3 and 4, where no structural analysis is performed, the chord length distribution, the ratio of blade thickness to chord length,

and the stacking in radial direction are kept unchanged to avoid large effects on their structural behavior.

The total of 223 design variables poses a challenge for BO, but is required for a detailed design of the 4-stage HPC frontblock. For a single-stage optimization with a similar problem and parameterization, a PCA showed that the design space dimension is not artificially high [26]. The about four times higher dimensionality and limited number of samples do not allow for reasonable PCA results for our 4-stage problem. This also makes variable screening and sensitivity analyses inaccurate, risking to neglect actually important parameters.

5.3 Blade optimization results

Due to the high simulation cost, we focus on four optimization approaches. The main reference is our BO, that is the base for the proposed informed and random cooperative components algorithms. We additionally run AutoOpti, because it was successfully employed for a similar multi-stage problem [8] and showed a good performance for the analytical test problem in Section 4. Our analysis includes the DoE, initial model accuracy, convergence, simulation success, and constraint fulfillment. Moreover, we compare the stage-wise quantities of interest and flow field of the initial and optimized design.

We perform 5 runs per algorithm with different seeds. Hence, we initialize the optimizations with 5 different DoEs. Like for the analytical problem, we aim for $d + 1$ initial samples and choose a batch size of $q = 10$ high-fidelity evaluations per optimization iteration. The maximum number of iterations is set to 100.

The large design space allows parameter combinations which lead to impossible blade geometries. In the optimization, the geometric constraints usually drive the search away from these failure regions. In the DoE, however, this causes many high-fidelity evaluation failures. We use a LHS with a d-dimensional normal distribution with center at the initial design $\boldsymbol{x}_0$ and standard deviation $0.2 \cdot (\boldsymbol{x}_{\mathrm{ub}} - \boldsymbol{x}_{\mathrm{lb}})$, truncated at the upper and lower bounds. This yields less extreme designs than a uniform distribution but still only 22% successfully evaluated sample points. Simulation failures occur in 75% of the cases in the geometry generation, 15% in the structural analyses, and the remaining 10% in the aerodynamic simulations. Therefore, the computational overhead of simulation failures in the DoE is small. With this in mind, we evaluate $5d + 1 = 1116$ designs and get initial training sets with 241 samples on average. Except for the initial design, none of the samples fulfills all constraints. This indicates a small feasible part of the design space.

The high number of constraints would make the GP model fits a very time consuming task, despite the parallelization. Moreover, even for very fast predictions, the infill criterion optimization times scale nearly linearly with the number of surrogate model evaluations. For a linear speed-up, we group the responses according to the 13 physical quantities that are expected to behave similarly: constriction stage 1-4, thickness stage 1+2, stresses stage 1+2, frequencies stage 1+2, mass flows, downstream velocity angle, and efficiencies. We then fit only one hyperparameter vector $\boldsymbol{\theta}$ per group by maximizing the sum of the individual likelihood functions $\psi(\boldsymbol{\theta})$ [21]. As expected, this accelerates the prediction by 70% and even increases the initial surrogate accuracy.

The initial model error is estimated by a 5-fold cross-validation, with one of the initial samplings as training set and the remaining four as test set. All errors are normalized with the respective test set variable range. Without design-to-response variable dependency information (for BO and random CC-BO), the mean error is 5.45% for one model per response and 5.29% for one model per group. With information (for informed CC-BO), the mean error is 4.67% for one model per response and 4.60% for one model per group. Especially the thickness, stress, and frequency models profit from the information. All structural models additionally improve with the grouping, presumably because the averaging counteracts overfitting. These initial model accuracies are acceptable, given the very small DoE and the adaptive enhancement in every optimization iteration.

The convergence over the optimization iterations is shown in Figure 8, in form of the maximum feasible objective function improvement over the initial value. In addition to the individual 5 runs per algorithm, the respective mean is indicated by the thicker lines. Our reference BO only achieves minor feasible improvements for the 223D blade design problem. This was to be expected, given the theoretical background in Section 2.2 and the 100D test problem results in Figure 5.

AutoOpti [9] can handle the large problem much better. It improves the objective function by 0.49% compared to the initial design in the 100 iterations. After the first few iterations, the slope is nearly steady with no sign of convergence. The long simulation

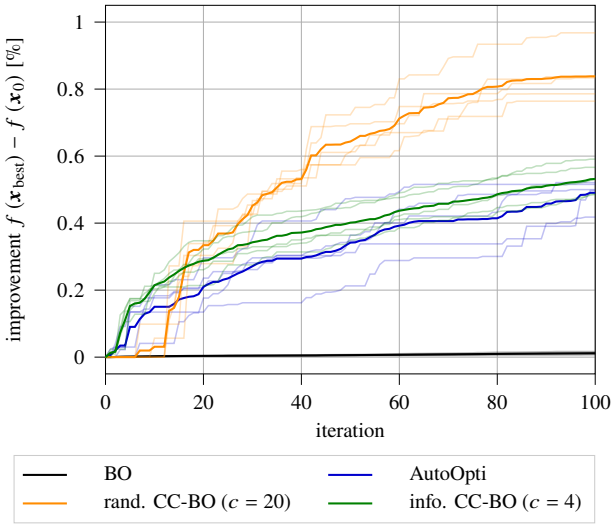


Figure 8 Convergence of various BO algorithms, including the proposed CC-BO, for the 223D multi-stage isentropic efficiency maximization. The mean (thick) and individual (thin) best feasible objective function improvements of 5 runs per algorithm are shown.

times prevent further iterations beyond the 1000 high-fidelity evaluations per run.

For random CC-BO, $c = 20$ components are randomly set in each iteration. The resulting subproblem dimensions of $d_i = 11$ to 12 should be well manageable by BO. The algorithm struggles to find a feasible improved design in the first 10 iterations. Afterwards, random CC-BO exhibits by far the best convergence of the four compared approaches. It reaches the 0.49% improvement, that AutoOpti yields in 100 iterations, after only 35 iterations. It finally converges to on average 0.84% isentropic efficiency improvement. The best run even achieves 0.97%. These improvements are considerable for a detailed optimization, especially as the first two of four stages were already preoptimized separately.

Informed CC-BO shows the best convergence in the first iterations. Afterwards, it exhibits a similar convergence slope to AutoOpti. The final designs after 100 iterations have on average a 0.04% higher isentropic efficiency than the ones obtained with AutoOpti. Conversely to the Branin case in Figure 5, the lower design space dimension of the random components outperforms the advantage of problem structure information. The $c = 20$ random components with $d_i = 11$ to 12 enable much higher efficiency improvements than the $c = 4$ stage components with $d_i = 51$ to 66. Moreover, a decomposition into different (and varying) variable groups might be preferable from the aerodynamic point of view.

Table 4 Mean percentage of successful evaluations and feasible designs during the initial DoE and the optimization iterations of the compared algorithms.

	DoE	**BO**	**Auto-Opti**	**rand. CC-BO**	**info. CC-BO**
successful	22%	51%	67%	62%	80%
feasible	0%	44%	30%	14%	61%

The optimizers must handle simulation crashes and 42 influential constraints. Accordingly, the ratios of successful evaluations and feasible designs in Table 4 provide further insights. On average at least 5 of the $q = 10$ parallel simulations per iteration are successful for all algorithms. The failures are thus no problem. Random CC-BO yields a remarkably low ratio of feasible designs. This explains the convergence difficulties in the first iterations in Figure 8. In contrast, informed CC-BO shows by far the highest ratios of both successful and feasible designs, which also reflects the steep initial convergence. Here, the advantage of the problem structure information takes effect. Each stage subproblem has fewer geometric and structural constraints, and their GP models are lower-dimensional and thus more accurate.

In other words, the physical components are advantageous for the stage-wise geometric and structural constraints. However, good constraint handling is less decisive than the authors expected. Overall aerodynamic aspects are more important for an efficient multi-stage blade optimization. The main working mechanism is hence the reduced subproblem design space dimension. The success of the problem decomposition approach shows that the problem is more separable, though not necessarily along stage interfaces, than the objective and constraint formulations indicate.

To conclude the result analysis, we compare the initial and best optimized designs obtained with the tested BO algorithms. The stage characteristics in Figure 9 provide insights on how the overall isentropic efficiency improvements are achieved. Quantities of interest are the stage-wise efficiency and intra- and inter-stage loading. The initial design curves are overlapped by the BO results. This means that BO attains its minor improvements by tiny design variations.

The other three approaches show significant design changes. The isentropic efficiency is increased in all four stages, with the biggest gains in the third and fourth stage. This was to be expected, as the two first stages were preoptimized. The reaction coefficient is

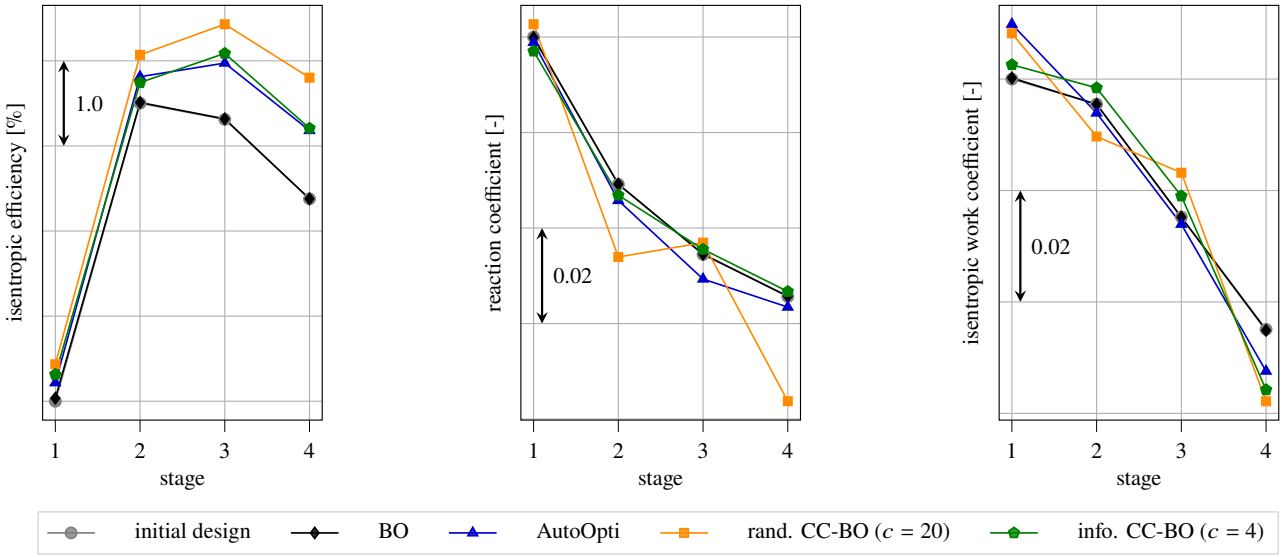


Figure 9 Stage-wise quantities of interest of the initial design and the best optimized designs of various BO algorithms. From left to right, the isentropic efficiency, reaction coefficient, and isentropic work coefficient of all four optimized stages are plotted.

the ratio of static enthalpy rise in the rotor and overall stage. It is only significantly changed by random CC-BO, increasing the static pressure rise for the stators of stage 2 and 4. This flexibility in the intra-stage loading might be due to the fact that random CC-BO is the only approach that does not optimize all variables of a stage at once. The isentropic work coefficient quantifies the inter-stage loading. All three approaches shift the work from the last to the first stage(s). Random CC-BO additionally relieves stage 2 and increases the load of stage 3. In summary, random CC-BO makes the biggest multi-stage design changes, which allow for the largest feasible efficiency improvements.

The Mach number distributions in Figure 10 show how the design optimization affects the flow field and, thus, the overall efficiency. We compare the initial to the best optimized design. The latter corresponds to the best random CC-BO run. In the transonic HPC frontblock, the main loss mechanism is shock-induced boundary layer separation. Especially in the last two stages, the shocks on the blades' suction sides are visibly mitigated. The peak Mach numbers are decreased and the deceleration to subsonic flow speeds behind the shocks is less abrupt, explaining the substantial efficiency gains.

6 Conclusions

The outcome of this work is a novel interconnected problem decomposition approach for high-dimensional constrained BO. The proposed CC-BO enables an efficient multi-stage aero-structural turbomachinery blade optimization. Following the ideas of distributed MDO [32], we exploit problem structure information for a decomposition into multiple interconnected lower-dimensional component subproblems. CC-BO differs in that the decomposition is only on the optimization level, while the overall system evaluations remain unaffected. An additional generalized version without information assumes a new random structure in every optimization iteration. The methodological implementation is inspired by CoEGO [16], with sequential subproblem optimizations connected by context vectors. We adapt the approach to our aim of an accelerated convergence over the number of iterations.

We examine the proposed method on a 100D multi-component Branin function as analytical test, and a 223D 4-stage aero-structural HPC blade design as generic industrial case. For both problems, the cooperative component approach shows a remarkable impact on the originally poorly performing BO. Informed and random CC-BO can efficiently handle the two high-dimensional optimization problems and largely outperform state-of-the-art algorithms. In the analytical case, the multi-component structure is well known and can be effectively used to the advantage of informed CC-BO. In the compressor case, the informed decomposition into the four stages is beneficial for constraint fulfillment. However, random CC-BO enables a lower subproblem dimensionality and an increased flexibility. This proves to be decisive for an

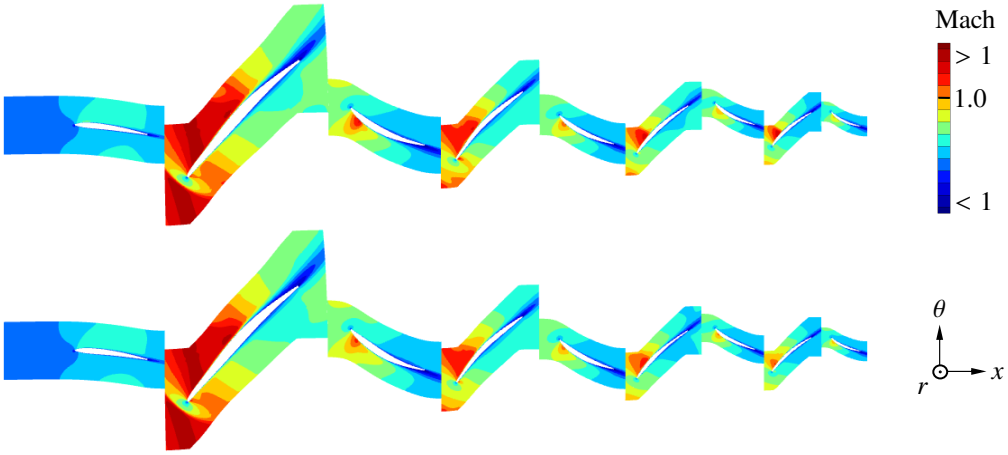


Figure 10 Mach number field at the mid plane for the initial (top) and overall best optimized (bottom) multi-stage design.

efficient multi-stage aerodynamic design. While BO only achieves marginal improvements, informed CC-BO gains around 0.5% isentropic efficiency, and random CC-BO even converges to a substantial average improvement of 0.84%.

Potential future research includes extensions and studies of both the multi-stage turbomachinery design application and the proposed method. Concerning the application, integrating cavities in the CFD simulations could yield more accurate results. Additional free annulus geometry parameters would enable even more flexibility in the multi-stage design. Methodologically, the effect of the number of components c for random CC-BO, and alternative (non-physical) components for informed CC-BO could be investigated. The cooperative components procedure should be readily transferable to other surrogate-based optimization approaches. The proposed CC-BO can efficiently handle high-dimensional constrained optimization problems with and without multi-component structure.

Statements and Declarations

Acknowledgments The authors would like to thank the *Technical University of Munich*, *MTU Aero Engines AG*, and *ONERA* for their support in this work.

Conflicts of interest The authors declare that they have no conflict of interest.

Funding This project is partly funded by the *Bayerische Luftfahrtforschung- und technologieförderung* of the *Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie (StMWi)*.

Replication of results An open-source implementation of the proposed cooperative components GP model fit is provided in version 2.7.0 of SMT [21]. The remaining parts of the proposed optimization algorithm are described in detail. The analytical Branin problem can be reproduced based on the provided equations. For the purpose of a realistic industrial application, the HPC blade design problem was set up in the design environment of *MTU Aero Engines AG* and is therefore subject to strict confidentiality.

References

[1] Z. Li and X. Zheng. "Review of design optimization methods for turbomachinery aerodynamics". In: *Progress in Aerospace Sciences* 93 (2017), pp. 1–23. DOI: 10.1016/j.paerosci.2017.05.003.

[2] D. X. Wang and L. He. "Adjoint Aerodynamic Design Optimization for Blades in Multistage Turbomachines—Part I: Methodology and Verification". In: *Journal of Turbomachinery* 132.2 (2010). DOI: 10.1115/1.3072498.

[3] D. X. Wang et al. "Adjoint Aerodynamic Design Optimization for Blades in Multistage Turbomachines—Part II: Validation and Application". In: *Journal of Turbomachinery* 132.2 (2010). DOI: 10.1115/1.3103928.

[4] C. Ma, X. Su, and X. Yuan. "An Efficient Unsteady Adjoint Optimization System for Multistage Turbomachinery". In: *Journal of Turbomachinery* 139.1 (2016). DOI: 10.1115/1.4034185.

[5] G. Ntanakas, M. Meyer, and K. C. Giannakoglou. "Employing the Time-Domain Unsteady Discrete Adjoint Method for Shape Optimization of Three-Dimensional Multirow Turbomachinery Configurations". In: *Journal of Turbomachinery* 140.8 (2018). DOI: 10.1115/1.4040564.

[6] S. Vitale, M. Pini, and P. Colonna. "Multistage Turbomachinery Design Using the Discrete Adjoint Method Within the Open-Source Software SU2". In: *Journal of Propulsion and Power* 36.3 (2020), pp. 465–478. DOI: 10.2514/1.b37685.

[7] H. Wu et al. "Multi-Row Turbomachinery Aerodynamic Design Optimization by an Efficient and Accurate Discrete Adjoint Solver". In: *Aerospace* 10.2 (2023), p. 106. DOI: 10.3390/aerospace10020106.

[8] M. Schnoes et al. "Strategies for Multi-Fidelity Optimization of Multi-Stage Compressors with Throughflow and 3D CFD". In: *Conference of the International Society for Air Breathing Engines (ISABE)*. 2019. URL: https://elib.dlr.de/129457/.

[9] U. Siller, C. Voß, and E. Nicke. "Automated Multidisciplinary Optimization of a Transonic Axial Compressor". In: *47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition*. American Institute of Aeronautics and Astronautics, 2009. DOI: 10.2514/6.2009-863.

[10] L. Baert et al. "Aerodynamic Optimization of the Low-Pressure Turbine Module: Exploiting Surrogate Models in a High-Dimensional Design Space". In: *Journal of Turbomachinery* 142.3 (2020). DOI: 10.1115/1.4046232.

[11] X. Zhang et al. "Optimization of Three-Dimensional Blade and Variable Stators for Efficiency and Stability Enhancement of Multistage Axial Flow Compressor at Variable Speeds". In: *Journal of Turbomachinery* 146.4 (2023). DOI: 10.1115/1.4064080.

[12] J. Mockus. *Bayesian Approach to Global Optimization*. Springer Netherlands, 1989. DOI: 10.1007/978-94-009-0909-0.

[13] M. Binois and N. Wycoff. "A Survey on High-dimensional Gaussian Process Modeling with Application to Bayesian Optimization". In: *ACM Transactions on Evolutionary Learning and Optimization* 2.2 (2022), pp. 1–26. DOI: 10.1145/3545611.

[14] J. Sobieszczanski-Sobieski and R. T. Haftka. "Multidisciplinary aerospace design optimization: survey of recent developments". In: *Structural Optimization* 14.1 (1997), pp. 1–23. DOI: 10.1007/bf01197554.

[15] J. R. R. A. Martins and A. Ning. *Engineering Design Optimization*. Cambridge University Press, 2021. 475 pp. DOI: 10.1017/9781108980647.

[16] D. Zhan et al. "A cooperative approach to efficient global optimization". In: *Journal of Global Optimization* (2023). DOI: 10.1007/s10898-023-01316-6.

[17] M. Schonlau, W. J. Welch, and D. R. Jones. "Global versus local search in constrained optimization of computer models". In: *Institute of Mathematical Statistics Lecture Notes - Monograph Series*. Institute of Mathematical Statistics, 1998, pp. 11–25. DOI: 10.1214/lnms/1215456182.

[18] A. I. Forrester and A. J. Keane. "Recent advances in surrogate-based optimization". In: *Progress in Aerospace Sciences* 45.1-3 (2009), pp. 50–79. DOI: 10.1016/j.paerosci.2008.11.001.

[19] R. Garnett. *Bayesian Optimization*. Cambridge University Press, 2023. DOI: 10.1017/9781108348973.

[20] M. J. D. Powell. "A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation". In: *Advances in Optimization and Numerical Analysis*. Springer Netherlands, 1994, pp. 51–67. ISBN: 9789401583305. DOI: 10.1007/978-94-015-8330-5_4.

[21] M. A. Bouhlel et al. "A Python surrogate modeling framework with derivatives". In: *Advances in Engineering Software* 135 (2019), p. 102662. DOI: 10.1016/j.advengsoft.2019.03.005.

[22] I. Arsenyev, F. Duddeck, and A. Fischersworring-Bunk. "Adaptive Surrogate-Based Multi-Disciplinary Optimization for Vane Clusters". In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2015. DOI: 10.1115/gt2015-42164.

[23] K. Kandasamy, J. Schneider, and B. Poczos. "High Dimensional Bayesian Optimisation and Bandits via Additive Models". In: *Proceedings of the 32nd International Conference on Machine Learning*. 2015. DOI: 10.48550/arXiv.1503.01673.

[24] W. J. Bräunling. *Flugzeugtriebwerke*. Springer Berlin Heidelberg, 2015. DOI: 10.1007/978-3-642-34539-5.

[25] Z. Wang et al. "Bayesian Optimization in a Billion Dimensions via Random Embeddings". In: *Journal of Artificial Intelligence Research* 55 (2016), pp. 361–387. DOI: 10.1613/jair.4806.

[26] L. Pretsch et al. "Twofold Adaptive Design Space Reduction for Constrained Bayesian Optimization of Transonic Compressor Blades". In: *Proceedings of the ASME Turbo Expo*. American Society of Mechanical Engineers, 2024. DOI: 10.1115/gt2024-121848.

[27] J. M. Parr et al. "Infill sampling criteria for surrogate-based optimization with constraint handling". In: *Engineering Optimization* 44.10 (2012), pp. 1147–1166. DOI: 10.1080/0305215x.2011.637556.

[28] E. Raponi et al. "High Dimensional Bayesian Optimization Assisted by Principal Component Analysis". In: *Parallel Problem Solving from Nature – PPSN XVI*. Springer International Publishing, 2020, pp. 169–183. DOI: 10.1007/978-3-030-58112-1_12.

[29] D. Eriksson et al. "Scalable global optimization via local Bayesian optimization". In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019.

[30] D. Eriksson and M. Poloczek. "Scalable Constrained Bayesian Optimization". In: *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*. Vol. 130. Proceedings of Machine Learning Research. 2021, pp. 730–738.

[31] M. Balandat et al. "BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization". In: *Advances in Neural Information Processing Systems* 33 (2019). DOI: 10.48550/arXiv.1910.06403.

[32] J. R. R. A. Martins and A. B. Lambe. "Multidisciplinary Design Optimization: A Survey of Architectures". In: *AIAA Journal* 51.9 (2013), pp. 2049–2075. DOI: 10.2514/1.j051895.

[33] R. Priem et al. "High-Dimensional Bayesian Optimization Using Both Random and Supervised Embeddings". In: *AIAA Journal* (2024), pp. 1–13. DOI: 10.2514/1.j063488.

[34] I. Arsenyev. "Efficient surrogate-based robust design optimization method". PhD thesis. Technical University of Munich, 2018. ISBN: 9783844060904.

[35] G. Dhondt and K. Wittig. *CalculiX*. http://www.calculix.de. Accessed: 2024-07-01. 1998.

[36] German Aerospace Center (DLR). *TRACE*. http://www.trace-portal.de/userguide/trace/index.html. Accessed: 2024-07-02. 2024.