TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Maschinenwesen

Lehrstuhl für Leichtbau

Extended Surrogate Modeling Techniques for Large Scale Structural Design Optimization

Qian Xu

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Florian Holzapfel

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. Horst Baier

2. Univ.-Prof. Dr.-Ing. habil. Fabian Duddeck

Die Dissertation wurde am 14 . 11 . 2013 bei der Technische Universität München eingereicht und durch die Fakultät für Maschinenwesen am 03 . 02 . 2014 angenommen.

This dissertation is dedicated to
my parents and my husband
for their endless support,
love and encouragement.

Acknowledgments

First of all, I would like to express my sincere thanks to my supervisor, Prof. Dr.-Ing. Horst Baier. Thank him for giving me the opportunity and resources to conduct research at Technische Universität München. His help, guidance and demands for excellence in these four years has made me a better scientific worker and a better engineer. From his suggestions and advises, I have learned to open my mind to the large industry world, to think and judge more critically and to make more practical engineering designs. Thank him a lot for guiding me and inspiring me through the process of my research in surrogate modeling and engineering design optimization.

I give my special thanks to my colleague Erich Wehrle, who has been helping me since the beginning of my doctor study. The cooperation and discussion with him has both widened and deepened my knowledge in structural optimization and automotive engineering. He has set me an example with his efficient working manner and exploring spirit. Also I want to thank him for being patient in helping me with English writing.

I would also like to give my heartfelt appreciation to all my colleagues. Ögmudur Petersson has given me a lot of patient and careful instructions in aerospace structural design. Jan Both, Markus Schatz and Tanut Ungwattanapanit are always ready to answer my questions related to composite materials. I am also grateful to Jan Both for giving me suggestions for adapting myself to the local living environment of Munich. Working with Markus Schatz has been a very pleasant experience, because he makes me confident by giving me inspiration and encouragement. Many thanks go to Peng He, Bin Wei, Luiz da Rocha-Schmidt, Rainer Wehrle, Alexander Morasch and Peter Krempl for helping me in all kinds of problems. I feel grateful to our secretary Amely Schwörer for being helpful and thoughtful to me. I appreciate the help of André Steinmetz, Manfred Bauer, Bernhard Lerch and Josip Stokic from the workshop of our insitute. Thanks again to all my colleagues for making our working place pleasant and energetic.

Great thanks is given to the China Scholarship Council (CSC) for providing me financial support to live and study in Germany.

Thanks to my friends Xiuna, Xueying, Xinxue and Antonia for going through all the up and downs with me. It is a precious memory and experience, to meet, talk, cook and go out with them. Thanks to all my friends and I have enjoyed a lot of good time during these four and half years. I am very glad and proud to acknowledge here my parents, who worked hard to bring me up. Thank them for giving me a happy family, a healthy body and an aspiring mind. I have saved the very best for the last to my husband. Though we are separated by 8000 km and 7 hours time difference, I can always feel that he is by my side to support me. From our 6000 emails, I can read the belief in love, the courage to overcome all difficulties, and the motivation to seek success. I dedicate this dissertation to him to express my admiration and gratitude.

Garching, September 2013   Qian Xu

Abstract

High-fidelity numerical analysis of structural systems is usually realized through computational structural mechanics, computational fluid dynamics, or coupled discrete methods of fluid-structural interaction. For large-scale structures, the evaluation of system responses can be computationally expensive, thus hindering structural and multidisciplinary design optimization. Surrogate modeling techniques are studied to approximate the system responses and alleviate computational costs in design optimization. This dissertation proposes extended surrogate modeling techniques, which are comprised of four methods. First, engineering knowledge is embedded in surrogate modeling to reduce the dimension of modeling space and improve approximation accuracy. Second, genetic programming is carried out to assist knowledge-based surrogate modeling by searching for the best formulations of regression functions. Third, the quality of surrogate models is improved adaptively by providing additional sample points at critical locations, which are identified according to multiple-infill criteria. Finally, parallelized computing distributes the independent evaluation processes for further accelerating the whole task. The extended surrogate modeling techniques are applied to two typical, large-scale structural problems: the parameter design optimization of an aircraft wingbox under aeroelastic considerations and the front crash system of an automobile under impact loading. It is shown in these applications that the extended surrogate modeling techniques can greatly reduce computational effort while maintaining good accuracy for design optimization.

Kurzfassung

Numerische Analysen von Strukturen werden normalerweise mithilfe numerischer Strukturmechanik, computational fluid dynamics oder gekoppelten diskreten Methoden der Fluid-Struktur-Interaktion durchgeführt. Die Auswertung der Systemantwort kann rechenintensiv werden und steht damit einer strukturellen und multidisziplinären Entwurfsoptimierung entgegen. In dieser Dissertation werden Methoden der parametrischen Ersatzmodellierung untersucht um die Systemantworten zu approximieren und den Rechenaufwand der Entwurfsoptimierung zu verringern. Eine erweiterte Methoden wird vorgeschlagen. Zuerst wird Ingenieurwissen in Ersatzmodellierung eingebettet, um die Dimension des Modellierungsraums zu reduzieren und die Genauigkeit der Approximation zu verbessern. Zweitens wird der genetische Algorithmus durchgeführt um die kenntnisbasierte Ersatzmodelle bei der Suche nach den besten Regressionsfunktionen zu unterstützen. Drittens wird die Qualität der Ersatzmodelle adaptiv durch zusätzliche Abtastpunkte an den kritischen Stellen verbessert, welche gemäß mehrerer Kriterien identifiziert werden. Zuletzt wird die Optimierungsaufgabe unter dem Einsatz von massiver Parallelisierung gelöst, wobei die einzelnen Evaluierungsschritte verteilt und die Lösung weiter beschleunigt wird. Die erweiterten Methoden der parametrischen Ersatzmodellierung werden zur Entwurfsoptimierung von zwei rechnerisch großen Aufgaben der Strukturoptimierung angewendet: Eine Parameter-Konstruktionsoptimierung eines Flugzeug-Flügelkastens unter aerodynamischen Lasten und einen Aufprallabsorber im Vorderbau eines Kraftfahrzeugs. In der Anwendung auf diese Probleme wird gezeigt, dass die erweiterten Methoden der parametrischen Ersatzmodellierung erheblich den Rechenaufwand reduzieren können und gute Genauigkeit für die Entwurfsoptimierung beibehalten.

Contents

List of Figures

List of Tables

Nomenclature

$\boldsymbol{\alpha}$	Vector of fiber angles in a laminate
$\bar{Q}$	Lamina stiffness matrix in the global coordinate system
$\boldsymbol{\beta}$	Coefficient vector of the regression model
$\ddot{\boldsymbol{u}}$	Acceleration vector of a structural system
$\boldsymbol{\omega}$	Weight vector of the radial basis functions
Φ	Cumulative distribution function
$\boldsymbol{\theta}$	Vector of parameters in correlation function of Kriging
$\boldsymbol{\varepsilon}^0$	Mid-plane strains in laminates
$\boldsymbol{A}^*$	In-plane compliance matrix of laminates
$\boldsymbol{A}$	In-plane stiffness matrix of laminates
$\boldsymbol{B}^*$	Bending and in-plane coupling compliance matrix of laminates
$\boldsymbol{B}$	Bending and in-plane coupling stiffness matrix of laminates
$\boldsymbol{D}^*$	Bending compliance matrix of laminates
$\boldsymbol{D}$	Bending stiffness matrix of laminates
$\boldsymbol{F}$	Load matrix of a structural system
$\boldsymbol{g}$	Nonlinear inequality constraints in optimization
$\boldsymbol{g}(\hat{\boldsymbol{y}})$	Surrogate model of nonlinear inequality constraints in optimization
$\boldsymbol{h}$	Nonlinear equality constraints in optimization
$\boldsymbol{h}(\hat{\boldsymbol{y}})$	Surrogate model of nonlinear equality constraints in optimization
$\boldsymbol{K}$	Global stiffness matrix of a structural system
$\boldsymbol{k}$	Plate curvatures in laminates
$\boldsymbol{K}^e$	Element stiffness matrix
$\boldsymbol{M}$	Global mass matrix of a structural system

$\boldsymbol{M}^e$	Element mass matrix
$\boldsymbol{M}_r$	Moment resultants in laminates
$\boldsymbol{N}_r$	Stress resultants in laminates
$\boldsymbol{p}(\boldsymbol{x})$	Vector of design variables in knowledge-based surrogate modeling space
$\boldsymbol{Q}$	Lamina stiffness matrix in the lamina coordinate system
$\boldsymbol{u}$	Displacement vector of a structural system
$\boldsymbol{x}$	Vector of design variables
$\boldsymbol{x}^{(i)}$	The ith vector of sample data sites
$\boldsymbol{x}_c$	Center points of the approximation region in surrogate model
$\boldsymbol{x}_l$	Lower boundary of design variables
$\boldsymbol{x}_u$	Upper boundary of design variables
$\boldsymbol{x}_{\text{fid}}$	Variables indicating particular formats of functions in genetic algorithm
$\boldsymbol{y}$	Vector of system responses
δ	Deflection
$\dot{\boldsymbol{u}}$	Velocity vector of a structural system
$\hat{\boldsymbol{\beta}}$	Estimated coefficients of regression models
$\hat{\boldsymbol{\beta}}^*$	Generalized least-squares estimation of $\boldsymbol{\beta}$ in Kriging
$\hat{\boldsymbol{\theta}}^*$	Maximum likelihood estimation of $\boldsymbol{\theta}$ in Kriging
$\hat{\boldsymbol{y}}$	Vector of surrogate models of system responses
$\hat{z}$	Surrogate model of objective function in optimization
λ_b	The buckling load factor of a laminate
cov	Covariance function of a random process
FI	Failure index
S($\boldsymbol{x}$)	Candidate set in genetic algorithm assisted knowledge-based surrogate modeling
φ	Probability density function
Ψ	Matrix of radial basis functions
ψ	Basis functions of linear superposition for regression

σ_y	Yield stress
σ_1	The in-plane longitudinal stress in a lamina
σ_2	The in-plane transverse stress in a lamina
τ	Shear stress
τ_{12}	The in-plane shear stress in a lamina
ε_{rms}	Normalized root mean square error
A	Cross-section area
A_{sy}	Effective shear cross-sectional area with respect to the y axis
A_{sz}	Effective shear cross-sectional area with respect to the z axis
E	Young's modulus
E_1	Longitudinal elastic modulus of a lamina
E_2	Transverse elastic modulus of a lamina
F	Matrix of regression function values at the sample points in Kriging
F_x	Compression force in a stiffened composite panel
F_{xy}	Shear force in a stiffened composite panel
G	Shear modulus
G_{12}	Shear modulus (longitudinal to transverse) of a lamina
I_y	Area moment of inertia with respect to y axis
I_z	Area moment of inertia with respect to z axis
J_x	Polar area moment of inertia
l	Length
l_e	Number of levels in factorial design
m	The dimension of design space
$m(\boldsymbol{x})$	Mass of structure as a function of design variables
N	Sample size
n_t	The number of test points in one group for cross validation
$n_{\boldsymbol{\beta}}$	Number of coefficients in the regression model

XVIII

n_{samp}	The number of sample points used in surrogate modeling
n_{Ev}	Number of evaluations in the optimization process
n_{It}	Number of iterations in the optimization process
p	Highest order of polynomials
$q(y)$	A knowledge-based surrogate modeling response
R	Matrix of correlation function values between the sample points in Kriging
t	Thickness
v_{12}	Major Poisson ratio (longitudinal to transverse) of a lamina
w_x	Width of a defined approximation region in surrogate model
X	Sample data sites
Y	Sample data responses
Z	Correlation model of Kriging
z	Objective function in optimization
$\mathcal{F}$	Regression model of Kriging
$\mathcal{P}_s$	The profit of using surrogate models in design optimization
$\mathcal{R}$	Correlation function in Kriging
$\mathcal{V}$	Vandermonde matrix
ANN	Artificial neural networks
ASBDO	Adaptive surrogate based design optimization
BLUP	Best linear unbiased prediction
CAD	Computer-aided design
CDF	Cumulative distribution function
CEI	Constrained expected improvement
CFD	Computational fluid dynamics
CFRP	Carbon-fiber-reinforced polymers
CLT	Classical laminate theory
CSM	Computational structural mechanics

DO	Direct optimization
DoE	Design of experiments
DoF	Degree of freedom
EA	Evolutionary algorithm
EI	Expected improvement
FEA	Finite element analysis
FEM	Finite element method
FFD	Full factorial design
FPF	First ply failure
FSI	Fluid-structural interaction
GA	Genetic algorithm
GAKBSM	Genetic algorithm assisted knowledge-based surrogate modeling
GRP	Glass-fiber-reinforced plastic
KBSM	Knowledge-based surrogate modeling
KKT	Karush-Kuhn-Tucker condition
LHS	Latin hypercube sampling
LLB	Lehrstuhl für Leichtbau
MUTE	Electric mobility cooperating project in Technische Universität München
PASBDO	Parallelized adaptive surrogate-based design optimization
PDF	Probability density function
PF	Probability of feasibility
RBF	Radial basis function
RMS	Root mean square error
SBDO	Surrogate-based design optimization
SQP	Sequential quadratic programming
TUM	Technische Universität München

1. Introduction

1.1. Motivation

Efficient design optimization is desired in the industry to support decision-making process for developing new concepts and improving existing products in today's fierce competition environment. It aims to stimulate better designs, reduce production cycle time and cut the cost. Structural design optimization is a subject of maximizing or minimizing the design objective function(s) (e.g. minimizing mass and/or cost), while satisfying all specified constraints (e.g. structural strength, stiffness and stability). Usually a number of evaluations of the objective functions and constraint functions are involved in an optimization process. The design objective and constraint functions in a structural design problem are usually formed by the system equations, which represent the structural responses under all possible working environments. With the development of engineering software packages and high-performance computing, system equations are evaluated by numerical simulations on computers instead of real tests and experiments in the early design stage.

Although the numerical simulations have been developed to be as fast and accurate as
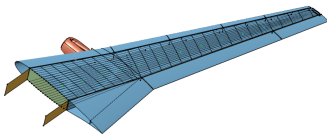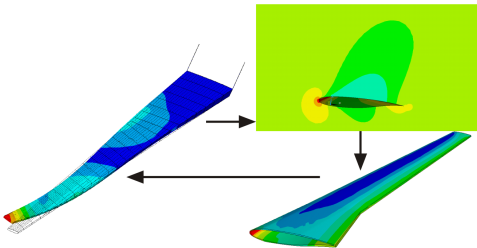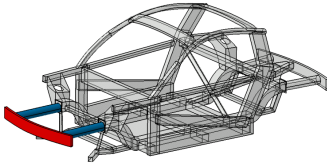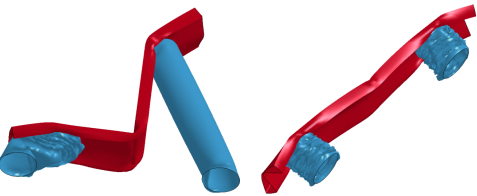
Structure	Numerical simulation	Challenge
Aircraft wing	Structural analysis coupled with aerodynamic analysis	- Computational expensive simulation: coupled computational structural mechanics and fluid dynamics with finite element method - Large number of design variables
Front crash system of MUTE	Crash analysis under different load cases	- Computational expensive simulation: crash simulation with finite element and explicit time integration method - Extremely nonlinear dynamics - Analytical gradients NOT available

Figure 1.1.: Challenges in the structural design optimization of an aircraft wing [Petersson et al. (2010)] and an automobile front crash system [Wehrle (2013)].

possible, the large computational expense per evaluation might still be an obstacle in the

optimization process. For example, problems in the structural design optimization of an aircraft wing and an automobile front crash system are shown in Figure 1.1. The models shown are two typical computationally expensive engineering tasks researched at the Institute of Lightweight Structures (Lehrstuhl für Leichtbau-LLB) of Technische Universität München (TUM). One computer-aided design (CAD) model in Figure 1.1 is a part of LLB-Flieger (aircraft), and the other comes from an electric mobility cooperating project in TUM called MUTE.

The structural design optimization of the aircraft wingbox involves finding the best design values for the geometrical parameters of the skin, spars, ribs and stringers of the wing to reduce weight of the structure. The number of design variables is very large in this problem. The wing has to sustain aerodynamic loads, and its system equations are evaluated by coupled structural and aerodynamic analysis through computational structural mechanics (CSM) and computational fluid dynamics (CFD). The numerical simulation of the aircraft wing is performed with the finite element methods, which is computationally time consuming due to a great number of elements, nodes and large dimensional matrices. Usually the total number of optimization iterations and the number of system evaluations, which are required to search for the best designs increase with the dimension of design space. Since the problem contains a large number of design variables, the computational expense for the optimization process can be extremely high.

For the design optimization of the MUTE front-crash system, computationally expensive simulation is also a barrier. Crash simulation is performed with LS-DYNA. The energy transformation process through plastic deformation is numerically analyzed by explicit time discretization. The structure is spatially discretized and analyzed using finite element method. Usually, many different crash load cases are to be considered in design optimization to ensure the passive safety requirements of the automobile. As a result, the computational cost for evaluating the system equations in the optimization problem goes rather high. The system responses of crash simulations are extremely nonlinear and multimode, which would require a number of iterations for the optimization algorithms to find an adequate feasible design. Besides, the system responses are of non-smooth character, which causes further problem, e.g. convergence problem in the optimization. Last but not least, with the current approach, analytical gradient information of the MUTE system equations is not available yet. Using numerical gradients in optimization require, however, many system evaluations and very high computational expense.

In order to deal with the above mentioned problems, surrogate-based design optimization is studied. Surrogate models are approximation models of the high-fidelity simulations. There are broad sense and narrow sense definitions of surrogate models. The broad definition considers all approximation approaches, e.g. model order reduction, discretization, curve fitting, etc. The narrow definition of surrogate models, which is also the definition in this dissertation, refers to the models constructed by fitting a limited number of sample data of the system responses. This requires that both the original numerical model and the surrogate model are parametric models. The original model contains the whole simulation of physical behavior, while the surrogate model cares only about the input-output relation of the system. Surrogate models are built to fit the input-output relation as closely as pos-

sible, while being computationally inexpensive to evaluate. By using surrogate models in optimization, the time and computational expense can be greatly reduced.

However, surrogate-based design optimization does not necessarily mean more efficient optimization. On the one hand, plenty of sample points are required for the construction of surrogate models and the curse of dimensionality arises when the number of design variables are large. On the other hand, accuracy of the models and the optimization results cannot be assured since the surrogate models are only approximations of the original problem.

In this dissertation, strategies for improving the accuracy of surrogate models and enhancing efficiency of surrogate-based design optimization are studied and implemented.

1.2. State of the Art - Literature Review

In this section, the applications and challenges of surrogate modeling techniques in engineering design optimization are introduced. A number of studies aiming at improving the performance of surrogate modeling and surrogate-based design optimization are reviewed. The advantages and limitations of these methods are discussed.

1.2.1. Applications and Challenges of Surrogate Modeling Techniques

The surrogate modeling techniques are used in a wide range of engineering design optimization problems. For example, Khodaparast et al. (2011) tested a number of surrogate modeling methods in order to approximate the entire flight envelope with a few sampling points in the future fast aeroelastic simulation technologies (FFAST). Blumhardt (2001), Forsberg and Nilsson (2005) tried different surrogate models in the automobile crashworthiness designs. Booker et al. (1997) used surrogate objectives on a Boeing helicopter design optimization. Huang et al. (2011) applied the Kriging surrogate model for shape optimization of an aero engine turbine disc. Leifsson et al. (2013), Ulaganathan and Asproulis (2013) also used different surrogate modeling strategies in aerodynamic shape optimization. Surrogate modeling applications can also be found in the field of electromagnetic-simulation-driven design optimization, simulation-driven antenna design, microwave structure design [Koziel and Leifsson (2013)] and so on . The wide application demonstrates the strong demands of reducing computational and time expense in engineering design for problems containing computationally expensive simulations. It is also shown that the common challenges of surrogate modeling in engineering application are to be addressed, which are concluded as follows:

(i) The number of sample points required in constructing surrogate models can be very large.

(ii) The quality of surrogate models for an engineering problem is not assured by large number of sample points.

(iii) Besides providing sufficient sample points, it is also important to use proper surrogate modeling techniques and to decide the proper design regions to be approximated.

(iv) In case only the input and output data of the engineering system are considered for surrogate modeling, the best distribution of sample points, the necessary density in sampling and the required number of sample points are all unknown before numerous trials.

(v) The proper type or the best formulation of surrogate models for an engineering problem is not known a priori. The properties to be approximated can differ from tasks to tasks and may also differ from regions to regions of the design space.

1.2.2. Advantages and Limitations of Current Strategies

To cope with the challenges in engineering applications of surrogate modeling and surrogate-based optimization, plenty of studies are carried out in both academics and the industries. All studies are oriented to provide robust solutions while reducing computational effort and maintaining promising accuracy [Sacks et al. (1989)]. Strategies are reviewed and concluded into following types.

1.2.2.1. Multiple surrogates

As is explained above, choosing the proper sampling and surrogate modeling methods is challenging. There is no single sampling nor surrogate modeling strategy that is appropriate for all problems. Therefore, the ensemble of multiple surrogate models is performed in approximation of complicate and large-scale problems.
Research and Advantages
Goel (2007) in his dissertation developed an averaging technique for multiple surrogates in order to protect against poor choice of single surrogate. Besides multiple surrogates, multiple error estimation models are also used to give sufficient confidence in the accuracy of surrogates. Hamza and Saitou (2005) showed that different sampling strategies and sample sets were utilized in order to avoid misleading in surrogate modeling by special patterns of sample distribution. Such strategies increase the reliability of surrogate models in engineering application.
Limitations and suggestions
However, proper schemes for weighted average or assembling surrogates and error estimation models should be developed. Even with multiple surrogate models, the refinement of surrogate models in the process of optimization is still needed. Therefore, this strategy may have advantage in approximation accuracy, but it is not necessarily better than conventional surrogate modeling strategies in regarding to the efficiency of constructing surrogates.

1.2.2.2. Trust region method, sequential design space reduction and multi-fidelity

In the whole design space of an engineering optimization problem, usually only a small region or regions are the critical domains that contain the optimal design. If such regions can be identified, then sample points can be distributed mainly there to save the computation effort.

Research and Advantages

Studies are carried out to discover the trust regions with information obtained by validating the optimization results. Therefore, of the whole design space, some regions can be approximated with relatively higher-fidelity surrogates while others with low fidelity. For example, Giunta and Eldred (2000) updated the size of trust region and reduced the design space adaptively, which showed significant reduction of computational expense.

Limitations and suggestions

When the trust region is decided too early, some critical regions could be excluded from the design space. Wang and Li (2012) pointed out that such strategies take the risks of losing characteristics of system responses, especially for highly-nonlinear systems.

1.2.2.3. Gradient information

When gradient information is provided, the approximation accuracy in the nearby region can be greatly increased. Therefore, in some studies, for example, in the work of Laurenceau et al. (2010), derivative information at sample points was fitted in construction of surrogate models.

Research and Advantages

Forrester et al. (2008) showed the ability of more accurate predictions with gradient-enhanced and Hessian-enhanced surrogate modeling.

Limitations and suggestions

For many computational expensive simulations, analytical or semi-analytical gradients are not always available, and numerical gradients are expensive to obtain. Besides, the time required in solving the construction of surrogate models is also longer, since operations of larger matrices have to be handled and the solving process is more complicate.

1.2.2.4. Space mapping and transformation

In high-dimensional design space, the number of sample points required for certain degree of accuracy in surrogate modeling can be extremely large. In this case, a mapped space with reduced dimension is used instead of the original design space.

Research and Advantages

The concept of space mapping was introduced by Bandler et al. (2001) in microwave engineering. The dimension reduction of design space was realized by employing the expert engineering knowledge. The space mapping was then applied and developed in many engineering fields. For example, Baudat and Anouar (2001) transformed a data set into a feature space according to geometric considerations. Physics-based surrogate modeling used by Nair (2002) was an extreme utilization of engineering knowledge in surrogate modeling and optimization, which demonstrated impressive efficiency. The space mapping and parameter transformation strategy has shown great potential in reducing computational effort in engineering problems.

Limitations and suggestions

The only disadvantage is the difficulty to generalize and quantify the expert knowledge, experience or even feelings. The approaches to extract knowledge and to embed it in surrogate

modeling process should be developed for specific engineering fields. This is done in this dissertation for mechanic analysis with finite element method.

1.2.2.5. Surrogate modeling with genetic programming methodology

Research and Advantages
Alvarez et al. (2000) proposed using genetic programming methodology to generate the best formulation of approximation functions. The genetic optimization of surrogate formations searches for the best surrogates with given sample points, and, therefore, makes the best use of the limited information. Such methods balance the pressure in expert knowledge requirements in space mapping and the blindness in black-box surrogate modeling.
Limitations and suggestions
In the genetic programming of surrogate construction, the candidate basis terms, functions and operations are required, which should be well considered and provided. Otherwise, extra effort is taken for unnecessary search and it can even fail to achieve the required approximation accuracy. A possible solution is the integration of this methodology with knowledge-based surrogate modeling, which is studied and provided in this dissertation.

1.2.2.6. Adaptive sampling and surrogate modeling

Usually more and more information is gathered after performing surrogate-based optimization and validating the feasibility and optimal conditions, e.g. the Karush-Kuhn-Tucker condition (KKT) of the optimization results discovered by Karush (1939), Kuhn and Tucker (1951). It allows us to refine the quality of surrogate models adaptively by properly using the information. Xu et al. (2012) demonstrated that using adaptive sampling and surrogate modeling, higher efficiency and accuracy can be obtained.
Research and Advantages
Infill criterion, such as expected improvement was described by Forrester et al. (2008), which identified positions of additional sample points to maximize the refinement in current surrogate models and the optimization results. Gorissen (2009) introduced in his dissertation a grid-enabled adaptive surrogate modeling method, which was implemented in a surrogate modeling (SUMO) Matlab toolbox. The SUMO toolbox offers large amount of options of sampling and surrogate strategies, error models and infill criteria. Some trials and studies are required to choose the combination of them, which offers the best performance.
Limitations and suggestions
Since only one sample point is added each iteration, the adaptive process can be rather slow if poor initial surrogate model is built. Infill criteria with robust performance, which provide multiple infill points at once is studied and extended in this dissertation to further explore the advantage of adaptive sampling and surrogate modeling.

As can be seen, the current strategies in surrogate modeling and surrogate-based optimization deal with one or more challenges in engineering application. Some of them emphasize the accuracy and reliability of surrogates and some provide possible alleviation to the curse of dimensionality. Some suggests that engineering knowledge to be used while

some try to provide a generalized solution and different options to choose. To summarize the state of the art, the above strategies can be combined, reformed and further developed to gain best accuracy, efficiency and robustness in solving our engineering problems, which is performed in this dissertation.

1.3. Outline of the Dissertation

In this dissertation extended surrogate modeling techniques are studied to achieve high efficiency, accuracy and robustness in surrogate-based optimization of large scale structural optimization problems. The proposed strategy in the optimization process can be demonstrated in Fig. 1.2. As can be seen the surrogate-based design optimization process is mainly
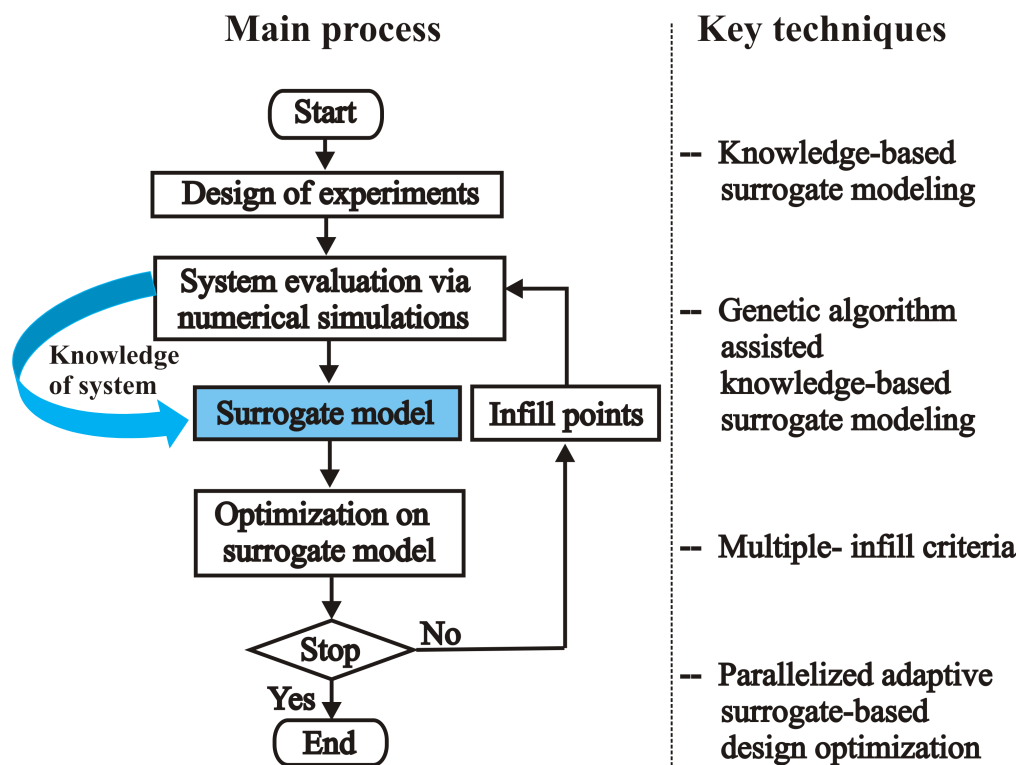
Figure 1.2.: Flowchart and key techniques of the extended surrogate modeling techniques for large scale structural optimization.

composed of the following steps.

(i) Design of experiments (DoE)
    Design of experiments generates a number of sample points in the design space of an optimization problem. Frequently used methods of DoE are described in Section 2.1.

(ii) Numerical simulation on the sample points
    In this dissertation system equations or responses obtained through engineering numerical simulations, such as finite element analysis and computational fluid dynamics, are defined as high-fidelity models. The high-fidelity models are used to provide

reference values for the construction of surrogate models. Introduction to the high-fidelity simulations of the engineering problems in this dissertation can be found in corresponding sections, which are Section 6.2 for the aircraft wingbox and Section 7.1 for the automobile front crash system, respectively.

(iii) Surrogate modeling

Surrogate modeling methods are used to build approximation models based on the information from sample points and particular mathematical hypothesis. In comparison to the engineering numerical simulations, surrogate models are normally not based on the physical behaviors of the engineering system and are defined as low-fidelity models, although they can present very high approximation accuracy. Popular surrogate modeling methods are discussed in Section 2.2.

(iv) Surrogate-based design optimization (SBDO)

Structural design optimization is performed by evaluating the computation inexpensive models instead of the high-fidelity numerical simulations. This technique as well as its formulation and assessment of quality are found in Section 2.3.

(v) Verification of optimization results and refinement of surrogate models

After the surrogate-based optimization, the feasibility of the obtained design is validated on the high-fidelity models and then the optimality conditions KKT are verified for the feasible designs. The verification of optimization results and the refinement of surrogate models are necessary to insure the correctness and goodness of the surrogate-based optimization results. This iterative surrogate-based optimization method stops when the optimal condition is satisfied and the designs converge in the last two loops. This is realized by the multiple-infill criteria developed in this dissertation, which is discussed in Chapter 5.

The key techniques developed in the surrogate modeling and optimization are listed:

(i) A knowledge-based surrogate modeling (KBSM) method is designed, which generates reduced modeling space with basic terms from geometric and mechanical properties of engineering structural system. This is described in Chapter 3.

(ii) Genetic algorithm is used to assist knowledge-based surrogate modeling. The formation of candidate functions and operations, as well as the definition of fitness function are described in Chapter 4.

(iii) Multiple-infill criteria are used to assess the quality of surrogate models and optimization results. The choice of infill criteria and the scheme for assembling the selected criteria are described from Section 5.2 to Section 5.5.

(iv) Parallelized adaptive surrogate-based design optimization (PASBDO) is used to speed up the whole process, which is introduced in Section 5.6.

To validate the performance of the proposed surrogate-based modeling and optimization techniques, test models with ranged model sizes are introduced.

Introduction examples

The introduction examples are of small size and low computational costs. With the introduction examples, it is easier to introduce, modify and validate the methods and algorithms during their developing process. Two examples are used here, one is an I-beam cantilever under vertical load at the tip. This model is used to introduce the knowledge-based surrogate modeling strategy, as can be found in Section 3.5. The other is a one-variable function, which is used to introduce the multiple-infill criteria. It is easier to visualize the sample points, infill points, surrogate function, original function and infill criteria with the one-variable function. This example is used in Section 5.2 to Section 5.6.

Demonstration example

The demonstration example is of medium size and medium computational cost, which is a stiffened composite panel from part of an aircraft wing skin. It is used to demonstrate the procedure and performance of the proposed techniques on engineering problem. With the medium size example, it is easier to compare the performances of different methods.

Application examples

As is mentioned at the beginning of this dissertation, two engineering examples are to be solved, which are of large-scale and high-computational costs. They are the aircraft wing-box and the automobile front crash system, which are used to validate the performance of the proposed method on solving large scale structural engineering problems. They are described in Chapter 6 and Chapter 7.

Finally, conclusion and discussion of the work in this dissertation are given in Chapter 8.

2. Surrogate Model and Surrogate-Based Design Optimization

Surrogate modeling is aimed at relieving high computational effort in engineering design optimization. In the process of design optimization, surrogate models of the system equations are used instead of the high-fidelity numerical simulation. While constructing surrogate models, two major steps are to be taken: design of experiments (DoE) and surrogate modeling. In this chapter, classic methods used in DoE are introduced first, followed by algorithms of surrogate modeling. Finally, the concept and framework of surrogate-based design optimization are introduced.

2.1. Design of Experiments

Design of experiments (DoE), also known as sampling plan, is a technique to generate a limited number of sample points in the design space. The system response at those points will be evaluated with high-fidelity numerical simulations, such as finite element analysis (FEA), which usually requires high computational effort. Those sample points will then be used as reference of the system behavior for the system response approximation. On the one hand, the number of sample points should be sufficient to ensure the quality of the approximation, on the other hand, it should be limited so that the corresponding amount of evaluations is computationally affordable.It is, therefore, important to control the density of sample points and in the meanwhile maintain good distribution in the design space. Different methods of DoE are more or less balancing between the amount and distribution of the sample points. The following sections discuss the properties and performance of some classic DoE methods, which include factorial design method, Monte Carlo method and Latin hypercube sampling.

2.1.1. Factorial Design Methods

Factorial design methods generate sample points that contain combinations of predefined discrete values of the design variables [Fisher (1925)]. When all possible combinations are considered, it is called a full factorial design (FFD). If only a fraction of the combinations are chosen, factional factorial design is resulted.

The design variables that affect system responses are called factors in statistics. The number of factors is, therefore, equal to the dimension of design space, noted as m. The number of the predefined discrete values of each design variable is called the number of levels, noted as l_e. A full factorial design creates l_e^m sample points. For example, full factorial
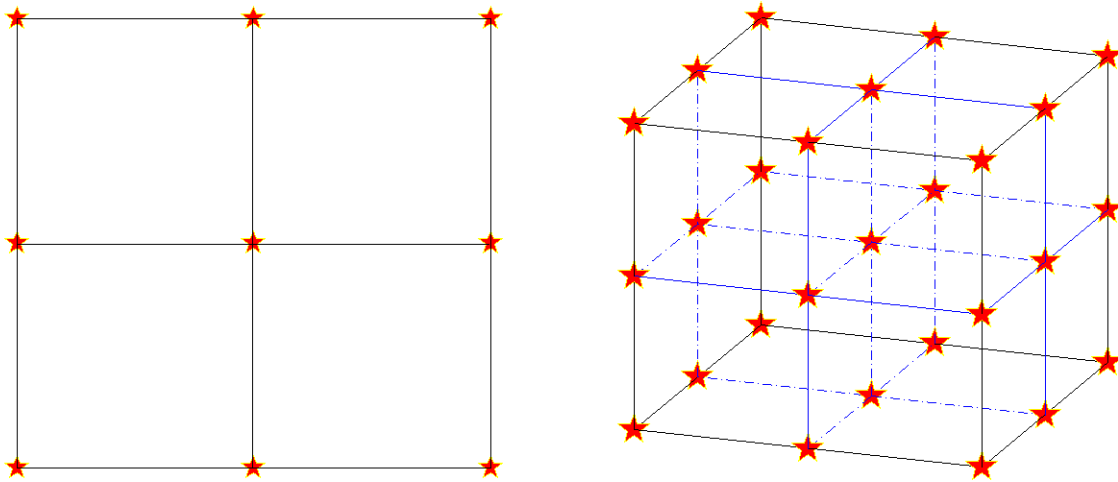
Figure 2.1.: Full factorial design for two (left) and three (right) demensional space with three
levels.

designs for two ($m = 2$) and three ($m = 3$) factors with three levels ($l_e = 3$) contain 9 sample
points and 27 sample points, respectively. They are shown in Figure 2.1.

 The advantages of using full factorial design are explained as follows:

- Sample points are regularly distributed in the design space, which insures good disper-
sity. No samples are gathered together nor spreaded distinctly far away from others.

- The density of sample points in full factorial design can be controled by setting the
factorial level for each dimension.

- The overall and interactive effects of the design variables on the system responses can
be explored. This is because that FFD is composed of all combinations of the design
variables at the values of all levels. In engineering design optimization, this also means
that a primary parameter study can be performed and an overview of the design opti-
mization problem can be obtained with full factorial design.

The disadvantage of full factorial design is the curse of dimensionality. The number of sam-
ple points in full factorial design l_e^m increases exponentially with the dimension of design
space m. Eventually the number of experiments or simulations to be carried out becomes
unaffordable and the surrogate modeling is no more efficient.

 Fractional factorial designs make compromise on full factorial design. They use subset
of the full factorial designs and thus only a fraction of the effort is needed. The trade-offs of
using a fractional factorial design is the aliasing of factor importance. The selection of subset
is, therefore, a problem of balancing all the factors and focusing on the important features of
the system, where knowledge of the studied system is essential.

2.1.2. Monte Carlo Experiments

Besides many other applications, Monte Carlo methods [Ulam (1949)] can also be used to generate sample points. In contrast to factorial designs, Monte Carlo methods generate randomly distributed sample points according to a predefined probability distribution throughout the given domain. A required number of sample points are obtained by blinded and repeated generating of random values. The characteristics of random experiments are discussed as follows:

- The number of sample points required in Monte Carlo experiments is not explicitly determined by the dimension of design space, but by the number of repetitions required to obtain stable statistic conclusions.

- With the probability distribution property, randomized experiments are especially usefully in handling uncertainty in system inputs.

2.1.3. Latin Hypercube Sampling

Latin hypercube sampling (LHS) is a high-dimensional extension of Latin square sampling [Fisher (1935). In two-dimensional space, sample points are generated in squared grids of the design domain with exactly one point in each column and each row. There are a number of solutions to locate sample points that satisfy this rule. Given the required sample size N and the dimension number $m = 2$ of design space, the number of qualified combinations equals to $(N!)^{m-1}$. Choices can be made between all the combinations. Besides, randomness can be assigned to the sample points by allowing the sample points to variate randomly within the located grids. Therefore, Latin hypercube sampling is not a deterministic design. For example, to generate three sample points in a two-dimensional design space, there are basically six combinations, as shown in Figure 2.2. Considering the randomness of sample points distribution within each grids, the actually number of combinations of LHS is infinite. Latin hypercube sampling can generate a design, which is the variant of any of those combinations. For situations where uniformity of sample distribution is needed, optimal design is performed by setting minimum distance or correlation requirements to constrain the variation of sample point.

For higher-dimensional space, $m > 2$, the design domain of each design variable is divided into required number of intervals, which equals to the sample size N. An example of Latin hypercube samples in three dimensional space is shown in Figure 2.3, where 4 sample points are generated and the sample size is $N = 4$. It can be seen that each dimension is divided into 4 intervals. When the sample points are projected to the two-dimensional planes, the distribution of projected points still satisfy the rule of Latin hypercube sampling. In each interval of each design variable, only one sample point is allowed. This means in each dimension, no sample point shares the same value as another sample point.

The feature of LHS distribution is very benefitial in design space mapping strategies. Space mapping is used to reduce the dimension of design space by transforming the original design variables into a smaller amount of quantities, which are still able to represent the system. The mapping usually involves some simple algebraic operations of the design
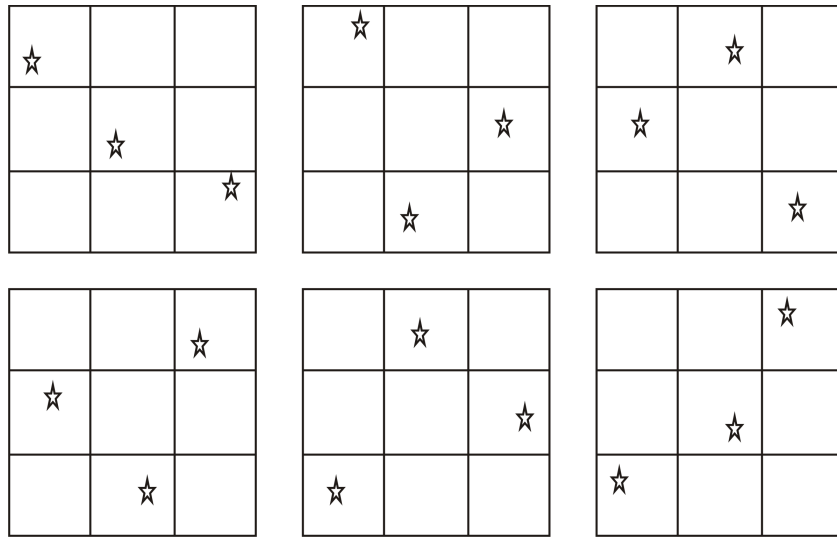
Figure 2.2.: Combinations of the Latin hypercube samples with three points in two-dimensional design space.
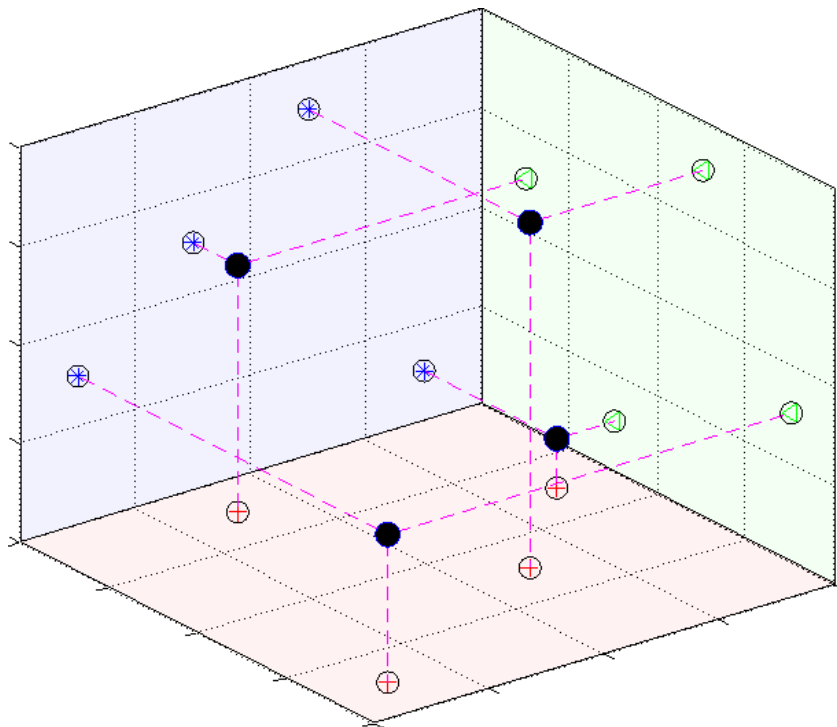
Figure 2.3.: Latin hypercube samples with four points in three-dimensional design space. The points projected at each plane are still Latin hypercube samples.

FFD

LHS
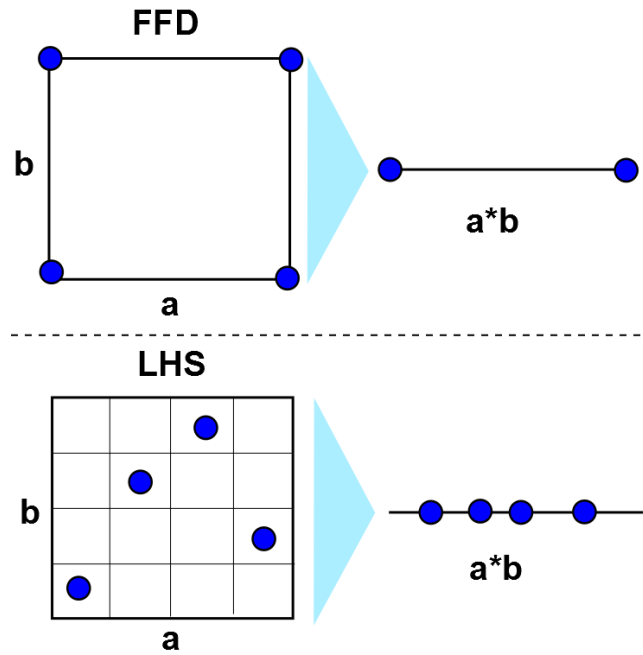
Figure 2.4.: Mapping of sample points from two-dimensional space $[a, b]$ to one-dimensional space $[a \cdot b]$, using four sample points generated with full factorial design (FFD) and Latin hypercube sampling (LHS), respectively.

variables based on their physical meanings, e.g., from the section width a and height b to the cross-section area $a \cdot b$. By doing this, two-dimensional space is reduced to one. Since LHS generates no duplicate values for all dimensions, it has the maximum retainment of sample points' information after space mapping. Therefore, it has absolute advantage over FFD in terms of space mapping. A simple example in Figure 2.4 shows the differnce between full factorial design (FFD) and Latin hypercube sampling (LHS) before and after space mapping. It is shown in Figure 2.4 that after space mapping by multiplication of the two design variables, only 2 sample points are obtained from 4 FFD sample points, while 4 samples are received using LHS.

Further advangtage of LHS over FFD is shown in the case of space projection, another form of space mapping, which is also quite frequently used in model reduction. For example, the differences between space projection of the three-dimensional sample distributions in Figure 2.1 for FFD and in Figure 2.3 for LHS can be seen, which are discussed as follows:

- For the FFD, 27 sample points are generated in 3D space. When projecting them to a coordinate plane, only nine independent sample points are there. When further projecting them to a coordinate axis, only three independent sample points are left.

- For the LHS, when four sample points in 3D space are projected to 2D and 1D, in each case there are still four Latin hypercube sample points.

For surrogate-based design optimization containing computationally expensive evaluations, the number of sample points should be kept small and the remaining amount of data after

space mapping strategies should be kept high. The advantages of Latin hypercube sampling are summarized as follows:

- The minimal number of sample points required in LHS does not have to increase exponentially with the dimension of design space. It is actually decided by the requirement in surrogate modeling, which can be much smaller than the exponential of the dimension. For example, in case of quadratic function, the amount is close to the square of the dimension. Although, the oversampling factor to achieve a certain degree of accuracy in approximation is still increasing largely with the dimension.

- It can easily generate well-scattered samples even in high-dimensional design space. This feature enables Latin hypercube sampling to represent large design spaces with small sample sizes.

- Sample points generated with LHS follow the rule to present certain distribution format, and possess some randomness at the mean time.

- When space mapping and transformation are to be performed, more information is remained with LHS.

Concluding from the above mentioned features, LHS has not only the good sampling density and dispersity as in FFD and some randomness as in Monte Carlo experiments, it also has large data retainment after space mapping. Since space mapping is one important strategy of the extended surrogate modeling techniques, the LHS is, therefore, the selected DoE method throughout this dissertation.

2.1.4. DoE in the Process of Preliminary Analysis and Trial for Design Optimization

Not only the sample points in DoE can be used for parameter study, it works also the other way around. In the process of carrying out an engineering problem, parameter study may be performed already before surrogate modeling is considered. Data from the parameter study can be directly taken as sample points to avoid repetitive computation. This is to say that sampling points can also be obtained during the process of preliminary analysis and trial in design optimization. To perform engineering design optimization, as much information of the problem as possible should be collected before choosing an optimization strategy. It is necessary that a preliminary analysis of the problem is performed in either design optimization or surrogate-based design optimization. There are also trial optimizations, which are carried out in the development process of a project. In such processes, numerical simulations of the structural model are also carried out to gain understanding of the problem, which is important for setting proper design domain, dominating constraints, good initial designs and algorithms for an optimization problem.

For computationally expansive numerical simulations, all responses evaluated with high-fidelity simulations are precious sand should be properly used. Such points can be considered as sample points while performing design of experiments. They can also be used as additional supporting information for refinement of surrogate models, known as infill points.

2.2. Surrogate Modeling Methods

After system responses are evaluated with high-fidelity numerical simulations on the sample points, surrogate models can be constructed. There are a variety of surrogate modeling methods available. The choice of method to be used should be made according to the features of system responses.

Linear and nonlinear system responses should be treated differently. Linear behavior can be approximated with first-order polynomial regressions, while nonlinear behavior should be approximated with higher-order polynomial regressions or other surrogate models, such as radial basis function (RBF) or Kriging. For example, in engineering design optimization, responses like displacements, stresses and strains are nonlinear, which should be approximated with carefully chosen surrogate modeling methods.

Smooth and nonsmooth system responses should also be distinguished by using regression or interpolation methods. An interpolation model will exactly go through all the sample points, while a regression method predict the global trend of the sample points. In the following sections, brief introductions of different surrogate modeling methods, including polynomial regression, RBF and Kriging will be given.

2.2.1. Polynomial Regression

Polynomial regression methods use polynomial functions to fit the sample points, the parameters of which are usually determined according to the least squares. The least-squares, first introduced by Legendre (1805), is a method to approximate the solution of overdetermined equations, and is a standard approach used in response surface approximation (RSA). The term response surface approximation referred particularly to a second-order polynomial regression when it was introduced by Box and Wilson (1951). Nowadays it is mostly understood as a general term for all approximation models.

A multivariate polynomial regression model is generally formulated as follows.

$$\hat{y}(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{m}\beta_i x_i + \sum_{i=1}^{m}\sum_{j>i}^{m}\beta_{ij}x_i x_j + \sum_{i=1}^{m}\beta_{ii}x_i^2 + \sum_{i=1}^{m}\sum_{j>i}^{m}\sum_{k>j}^{m}\beta_{ijk}x_i x_j x_k + \cdots + \sum_{i=1}^{m}\beta_{ii\cdots i}x_i^p, \quad (2.1)$$

where,

$\boldsymbol{x}$: vector of design variables,

m: dimension of design space,

p: highest order of polynomials,

$\hat{y}$: polynomial regesseion model for a system response y,

$\boldsymbol{\beta}$: coefficients of the regression model.

The coefficients $\boldsymbol{\beta}$ should be determined in order to construct a polynomial regression model. The number of coefficients n_β with respect to the order of the polynomial regression model

is shown as follows:

$$
\begin{aligned}
&\text{when} && p = 0, \ \ n_\beta = 1, && \text{regression with constant value} \\
&\text{when} && p = 1, \ \ n_\beta = 1 + m, && \text{regression with linear function} \\
&\text{when} && p = 2, \ \ n_\beta = \frac{(1+m)(m+2)}{2}, && \text{second-order polynomial regression} \quad (2.2) \\
&\ \vdots && \ \vdots \quad \ \ \vdots && \ \vdots \\
&\text{generally} && \quad\ n_\beta = C_p^{(m+p)} = \frac{(m+p)!}{p!m!}, && \text{higher-order polynomial regression,}
\end{aligned}
$$

where $C_p^{(m+p)}$ is the number of p-combinations. According to the rule of least squares, at least n_β sample points are required to determine the coefficients $\boldsymbol{\beta}$. Suppose that N sample points are obtained from the design of experiments. The observed data sites are $X = \left\{ \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}, \cdots, \boldsymbol{x}^{(N)} \right\}^T$. The corresponding responses are $Y = \left\{ y^{(1)}, y^{(2)}, y^{(3)}, \cdots, y^{(N)} \right\}^T$. A system of linear equations can be written as

$$
\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix}
=
\begin{bmatrix}
1 & x_1^{(1)} & \cdots & x_i^{(1)} & \cdots & (x_i x_j)^{(1)} & \cdots & (x_i x_j x_k)^{(1)} & \cdots & (x_i^p)^{(1)} & \cdots \\
1 & x_1^{(2)} & \cdots & x_i^{(2)} & \cdots & (x_i x_j)^{(2)} & \cdots & (x_i x_j x_k)^{(2)} & \cdots & (x_i^p)^{(2)} & \cdots \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
1 & x_1^{(N)} & \cdots & x_i^{(N)} & \cdots & (x_i x_j)^{(N)} & \cdots & (x_i x_j x_k)^{(N)} & \cdots & (x_i^p)^{(N)} & \cdots
\end{bmatrix}
\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_{ij} \\ \vdots \\ \beta_{ijk} \\ \vdots \\ \beta_{ii\cdots i} \end{bmatrix} .
$$

$$(2.3)$$

Equation 2.3 can also be written in symbolic format as $Y = \mathcal{V}\boldsymbol{\beta}$, where Y is the vector of system responses at all sample points, $\boldsymbol{\beta}$ is the vector of coefficients and $\mathcal{V}$ is the matrix of the sampled design sites evaluated in all required polynomial combinations. $\mathcal{V}$ is a Vandermonde matrix [Luther and Rost (2004)] in case of a one-dimensional polynomial regression. The estimated coefficients $\hat{\boldsymbol{\beta}}$ can be obtained through the least square solution of $Y = \mathcal{V}\boldsymbol{\beta}$, which is

$$
\hat{\boldsymbol{\beta}} = (\mathcal{V}^T \mathcal{V})^{-1} \mathcal{V}^T Y. \tag{2.4}
$$

This theory of polynomial regression is well established and has been used since the early 18th century [Stigler (1974)]. The advantages of polynomial regression model are given as follows:

- It is simply formulated and its properties are well understood. Such features make it easy to be programmed and applied.

- It can be efficiently calculated and simply visualized and, thus, the system exploration can be greatly facilitated.

- It smooths the system responses, which is helpful in handling noise in the system. The convergence behavior can be greatly improved with smoothed system responses in the optimization process. Besides, after smoothing, gradient calculation can be more accurate and efficient in case that gradient-based searching is used in engineering design optimization.

However, the drawbacks should also be noted, which includes:

- Polynomial regression generally has poor interpolatory and extrapolatory properties. Assume no noise is included in the high-fidelity system simulations, then all sample points are considered accurate and are to be respected while building surrogate models. When the sample points do not behave as the given order of polynomials, using polynomial regression in this case will result in the loss of large amount of information. Therefore, the quality of the regression models are not reliable while approximating rather complicated system responses.

- It can be seen from Equation 2.2 that high-order polynomial regression requires large number of coefficients to be determined, especially when the number of design variables are also large. In this case, the cost in constructing regression models is very high.

- The Runge's phenomenon was found by Runge (1901) when high-order polynomial regression models are used to approximate nonlinear responses, which results in large oscillation in some regions, as is demonstrated in Figure 2.5. It is seen that although the higher-order polynomial function fits better with the sample points, the quality of the approximation at positions between sample points, and especially at the boundaries could be very poor.

To conclude, although the polynomial regression is well established and easy to perform, the system behavior, such as nonlinearity and noise, and the dimension of design space should be carefully considered to decide whether to use it or not. In most engineering problems, the second-order polynomial regression shows good performance and is still widely used.

2.2.2. Radial Basis Function

The radial basis function is a weighted linear summation of basis functions with different centers. It is also understood as a single-layer linear artificial neural networks (ANN) [Ewan (2010)]. The term basis function means the basic function format of the elements, which are combined to approximate complicated functions. The basis functions are usually simple and have well understood behavior. The same symbols for sample points as polynomial regression method in section 2.2.1 are used. They are $X = \left\{ \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}, \cdots, \boldsymbol{x}^{(N)} \right\}^{T}$ for the observed data sites and $Y = \left\{ y^{(1)}, y^{(2)}, y^{(3)}, \cdots, y^{(N)} \right\}^{T}$ for the corresponding responses. The surrogate model constructed with radial basis function method is formulated as follows:

$$\hat{y}(\boldsymbol{x}) = \sum_{i=1}^{N} \omega_i \psi(\left\| \boldsymbol{x} - \boldsymbol{x}^{(i)} \right\|), \tag{2.5}$$
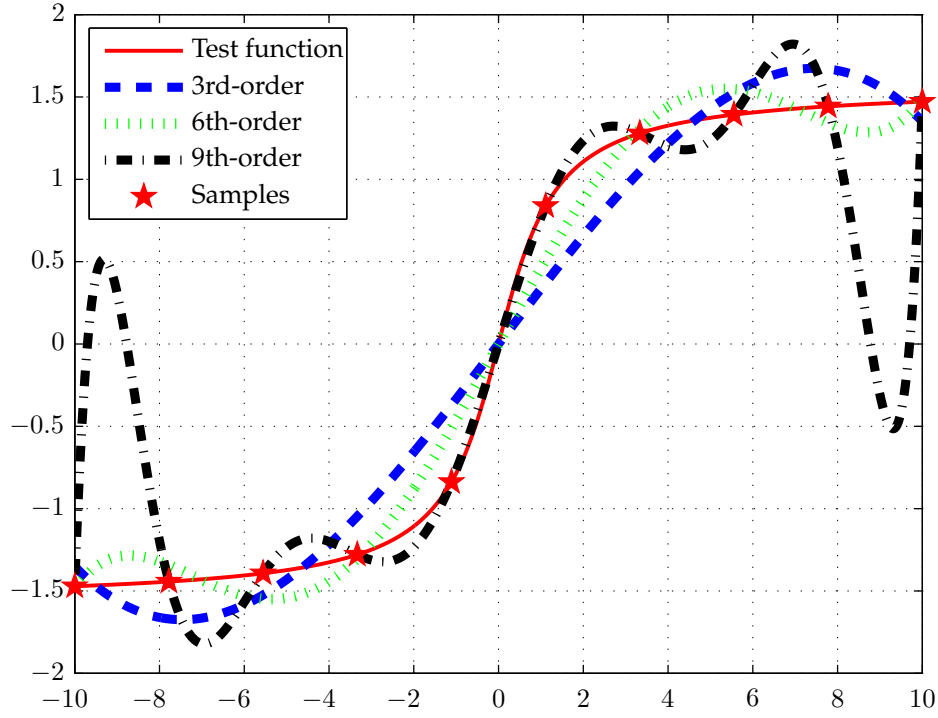
Figure 2.5.: Demonstration of the Runge's phenomenon. Red solid line: test function; red star: sample points; blue dashed line: 3rd-order polynomial regression function; dotted line: 6th-order polynomial regression function; dash-dotted line: 9th-order polynomial regression function.

where,

$\boldsymbol{x}$: vector of design variables,

m: dimension of design space,

$\hat{y}$: radial basis function for y,

$\boldsymbol{\omega}$: weight vector of the basis functions ,

ψ: the basis function, which is variable with $\left\| \boldsymbol{x} - \boldsymbol{x}^{(i)} \right\|$,

$\left\| \boldsymbol{x} - \boldsymbol{x}^{(i)} \right\|$: the Euclidean distance between the unobserved site $\boldsymbol{x}$ and the sampled site $\boldsymbol{x}^{(i)}$,

$\boldsymbol{x}^{(i)}$: the ith observed sample site, $i = 1, 2, \cdots, N$, which serves as the center of the ith basis function.

The popularly employed basis functions are given as follows:

- linear function: $\psi(r) = r$

- cubic function: $\psi(r) = r^3$

- thin plate spline function: $\psi(r) = r^2 \ln r$

- Gaussian function: $\psi(r) = e^{-r^2}$

where r stands for the Euclidean distance of two sites. Using the sample points, a system of linear equations is obtained:

$$
\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix} = \begin{bmatrix} \psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,N} \\ \psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{N,1} & \psi_{N,1} & \cdots & \psi_{N,N} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \end{bmatrix}, \tag{2.6}
$$

where $\psi_{i,j} = \psi(\left\| \boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)} \right\|), (i, j = 1, 2, \cdots, N)$. The equations above can also be written in matrix form as:

$$
Y = \Psi \boldsymbol{\omega}, \tag{2.7}
$$

where the solution for the weighting vector $\boldsymbol{\omega}$ can be obtained by solving the linear equation

$$
\boldsymbol{\omega} = \Psi^{-1} Y. \tag{2.8}
$$

The characteristics of the radial basis function method are summarized as follows:

- Based on the selection of basis functions, a wide range of linear and nonlinear system responses can be approximated with radial basis functions.

- It has both regression and interpolation abilities, which makes it useful in approximating both deterministic and stochastic behaviors. However, it tends to overfit the model and results in large local error.

- It is easy to formulate, but solving the Equation 2.8 can be a tough task because of Ψ^{-1}. The choice of sample sites and basis function plays an important role to assure a positive definite Ψ matrix.

2.2.3. Kriging

Kriging was first developed by the mining engineer Danie G. Krige in 1951 as a technique for prediction of mineral resources [Krige (1951)]. Besides its application in geostatistics, it has also made its way into many engineering field. It became widely used as a surrogate modeling method for design and analysis of computer experiments following the work of Sacks et al. (1989).

Kriging is known as an interpolation method and a model performing best linear unbiased prediction (BLUP) of random effects. The Kriging model is composed of two parts, a deterministic part and a stochastic part. The deterministic part is actually a regression model that fits the sample points according to the rule of generalized least squares. The stochastic

part is a Gaussian process that interpolates the regression residual by using a correlation function. Suppose the sample points are given, where $X = \left\{ \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}, \cdots, \boldsymbol{x}^{(N)} \right\}^T$ are the observed data sites, and $Y = \left\{ y^{(1)}, y^{(2)}, y^{(3)}, \cdots, y^{(N)} \right\}^T$ are the corresponding responses. Kriging is formulated as

$$\hat{y}(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}) + Z(\boldsymbol{x}) = \overbrace{\sum_{i=1}^{p} \beta_i f_i(\boldsymbol{x})}^{\text{regression}} + \underbrace{Z(\boldsymbol{x})}_{\text{correlation}}, \tag{2.9}$$

where $\mathcal{F}(\boldsymbol{x}) = \sum_{i=1}^{p} \beta_i f_i(\boldsymbol{x})$ is the regression model, which is a linear combination of selected functions $f_i(\boldsymbol{x}), i = 1, 2, \cdots, p$. Usually different orders of polynomial functions are chosen as f_i, but the regression functions are not restricted to polynomials. Any continuous function form can be considered and the choice should be made to suit with different applications. $Z(\boldsymbol{x})$ is the Gaussian process with mean value 0 and covariance given by

$$\text{cov}(Z(\boldsymbol{x}^{(i)}), Z(\boldsymbol{x}^{(j)})) = \sigma^2 \mathcal{R} \left(\boldsymbol{\theta}, \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \right), \tag{2.10}$$

where $\mathcal{R}$ is the correlation function with a m-dimensional parameter vector $\boldsymbol{\theta}$. This means for the m-dimensional design variables $\boldsymbol{x}$, each dimension can be assigned with different $\theta_k, j = 1, 2, \cdots, m$, leading to different correlation properties. For continuously differentiable responses, usually Gaussian function is chosen as the correlation function, written as

$$\mathcal{R} \left(\boldsymbol{\theta}, \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \right) = \prod_{k=1}^{m} e^{-\theta_k \left\| x_k^{(i)} - x_k^{(j)} \right\|^2}. \tag{2.11}$$

After chosen proper regression functions f_i and correlation function $\mathcal{R}$, there are only the regression coefficients $\boldsymbol{\beta}$ and the correlation parameter $\boldsymbol{\theta}$ to be determined to fit the sample data. Similarly as in the polynomial regression in section 2.2.1, the regression coefficients $\boldsymbol{\beta}$ can be determined according to the rule of least squares. The difference is that the rule of generalized least squares is used and the estimated regression coefficients $\hat{\boldsymbol{\beta}}$ is written as

$$\hat{\boldsymbol{\beta}} = (F^T R^{-1} F)^{-1} F^T R^{-1} Y. \tag{2.12}$$

In the generalized least squares solution shown in Equation 2.12, F is a matrix containing the value of the regression functions at the observed sites, given as

$$F = \begin{bmatrix} f_1(\boldsymbol{x}^{(1)}) & f_2(\boldsymbol{x}^{(1)}) & \cdots & f_p(\boldsymbol{x}^{(1)}) \\ f_1(\boldsymbol{x}^{(2)}) & f_2(\boldsymbol{x}^{(2)}) & \cdots & f_p(\boldsymbol{x}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(\boldsymbol{x}^{(N)}) & f_2(\boldsymbol{x}^{(N)}) & \cdots & f_p(\boldsymbol{x}^{(N)}) \end{bmatrix}. \tag{2.13}$$

In Equation 2.12, R is a symmetric matrix containing the correlation function values between the observed design sites. R_{ij} is given by

$$R_{ij} = \mathcal{R} \left(\boldsymbol{\theta}, \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \right), \ i, j = 1, 2, \cdots, N, \tag{2.14}$$

where it can be seen that R_{ij} is variable with $\boldsymbol{\theta}$. The optimal correlation parameter $\hat{\boldsymbol{\theta}}$ is obtained by solving the maximum likelihood estimation of the Gaussian process under the observed data. This means in contrast to the least squares estimation of β in deterministic regression, the generalized least squares estimation of β has to be determined iteratively with the correlation parameter $\boldsymbol{\theta}$. With the optimal $\hat{\beta}^*$ and $\hat{\boldsymbol{\theta}}^*$, the Kriging model is able to interpolation the sample points. It is, therefore, a good combination of deterministic model and stochastic process. The advantages of Kriging model are significant, which are summarized as follows:

- It contains a regression model, which uses linear combination of different functions to estimate the trend of the system responses. The selection of the regression functions is flexible, which allows us to implant engineering knowledge in the process of surrogate modeling.

- With the correlation function and maximum likelihood estimation, the stochastic process makes full use of the sample points. It not only gives the Kriging model the capability of interpolating regression residuals but also enables better prediction of nearby locations.

- It provides reasonable balance between approximation accuracy and construction efficiency.

Especially, it should be noted that Kriging contains two parts, a regression model and a correlation model, of which the regression model dominate the trend of the surrogate model. It is, therefore, important to choose proper formulations of the regression functions which are able to represent the general trend of system behavior as precisely as possible. Higher-order polynomial regression functions have the capability of approximating more complex responses, but require more sample points to determine the polynomial coefficients and might result in irregular behavior. By properly generating mixed terms of design variables and selecting the function forms, the required number of sample points can be reduced and the performance of Kriging can be greatly improved. The techniques for generating mixed terms and regression function forms are specifically described in chapter 3 and chapter 4, where knowledge-based surrogate modeling and genetic algorithm assisted surrogate modeling are introduced, respectively.

2.3. Surrogate-Based Design Optimization

In engineering design optimization problems, system equations mostly exist in implicit forms by using numerical simulation methods [Baier et al. (1994)]. For example, in aerospace and automotive designs numerical simulations with the finite element method (FEM) are used to establish a relationship between design variables and system responses. An engi-

neering design optimization problem is mathematically formulated as follows:

$$
\begin{aligned}
\text{minimize} \quad & z(\boldsymbol{x}) \\
\text{such that} \quad & g_j(\boldsymbol{y}(\boldsymbol{x})) \leq 0 , \quad j = 1, \cdots, p \\
\text{and} \quad & h_k(\boldsymbol{y}(\boldsymbol{x})) = 0 , \quad k = 1, \cdots, q \\
& \boldsymbol{x}_l \leq \boldsymbol{x} \leq \boldsymbol{x}_u
\end{aligned}
\tag{2.15}
$$

where the symbols are expressed as follows:

$\boldsymbol{x}$: vector of design variables,

z: objective function,

$\boldsymbol{y}$: vector of system responses, which are to be determined from the system equations,

g_j: a nonlinear inequality constraint,

h_k: a nonlinear equality constraint.

To alleviate the cost in the optimization process, surrogate models are used to replace the computational expensive numerical simulations. Surrogate models The surrogate-based design optimization has a formulation shown as:

$$
\begin{aligned}
\text{minimize} \quad & \hat{z}(\boldsymbol{x}) \\
\text{such that} \quad & g_j(\hat{\boldsymbol{y}}(\boldsymbol{x})) \leq 0 , \quad j = 1, \cdots, p \\
\text{and} \quad & h_k(\hat{\boldsymbol{y}}(\boldsymbol{x})) = 0 , \quad k = 1, \cdots, q \\
& \|\boldsymbol{x} - \boldsymbol{x}_c\| \leq w_x
\end{aligned}
\tag{2.16}
$$

Symbols in Equation 2.16 are expressed as follows:

$\hat{z}$: surrogate objective function,

$\hat{\boldsymbol{y}}$: vector of approximation of system responses, which are evaluated on surrogate models,

$g_j(\hat{\boldsymbol{y}})$: a surrogate nonlinear inequality constraint,

$h_k(\hat{\boldsymbol{y}})$: a surrogate nonlinear equality constraint,

w_x: width of a defined approximation region,

$\boldsymbol{x}_c$: center points of the approximation region.

The implementation difference of the standard engineering design optimization and the surrogate-based design optimization can be demonstrated with Figure 2.6.

In standard engineering design optimization, evaluations of the engineering system has to be done iteratively, while in surrogate-based design optimization, surrogate models are constructed first, which are then evaluated in instead. To get the surrogate models, design of experiments and evaluation of the high-fidelity system equations on the sample points have to be performed first. Usually the computational cost and time expense for a surrogate
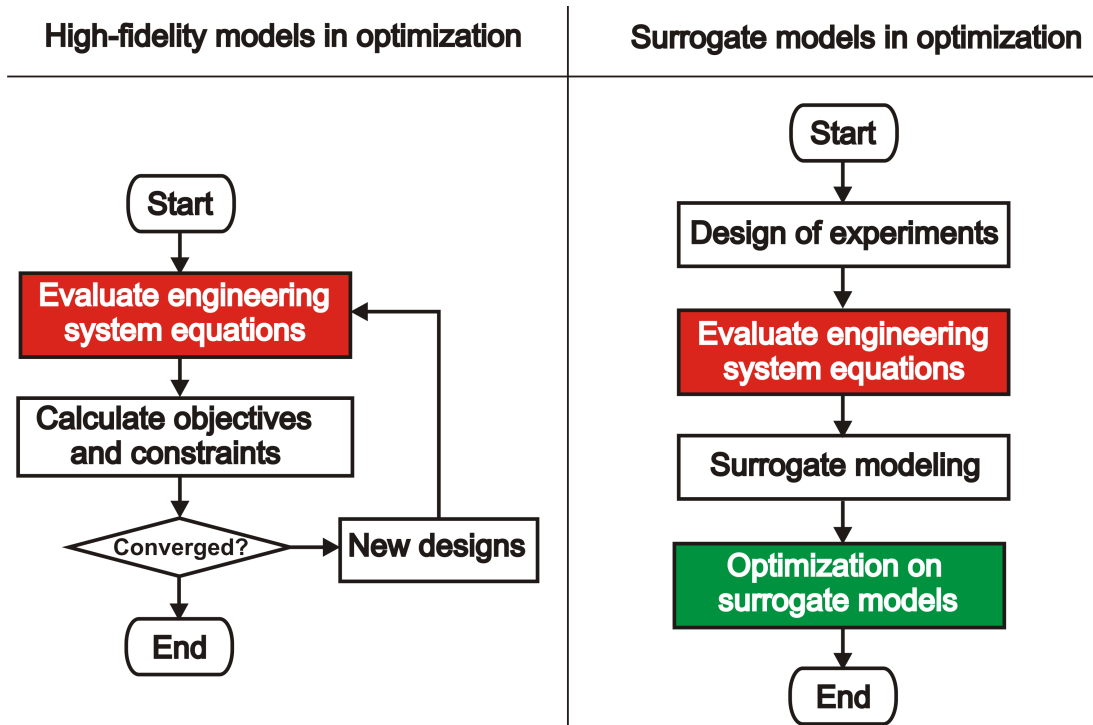
Figure 2.6.: The flowchart of the standard engineering design optimization and the surrogate-based design optimization.

model can be ignored in comparison with a high-fidelity numerical simulationA through, e.g., FEM and CFD. The cost for surrogate-based design optimization thus mainly depends on the construction of surrogate models. The process of surrogate modeling can be graphically represented in Figure 2.7.

In Figure 2.7 it is shown that the core expense of surrogate modeling lies on the evaluation of the engineering system at sample points. First, the design sites for a number of sample points are generated with DoE, then a corresponding amount of high-fidelity system evaluations are performed. Surrogate modeling method is then used to fit the sample points. Supposing the computational and time expense in data transferring and surrogate model evaluation can be neglected, the speed-up ratio of surrogate-based design optimization using sequential computing can be quantified by

$$\mathcal{P}_s = \frac{N_r}{N_s},\tag{2.17}$$

where N_r refers to the number of high-fidelity evaluations of the engineering system in standard engineering design optimization, and N_s is the total number of sample points used for construction of surrogate models. $\mathcal{P}_s$ is the speed-up ratio of surrogate-based design optimization, which means the profit of using surrogate models in design optimization. Further, if parallel computing is used, the surrogate-based design optimization can be even more efficient, which will be introduced in Section 5.6. Although parallelized system evaluation is also possible in the standard high-fidelity model based optimization, the number of parallelized processes is usually limited. This is because that parallelization can only be
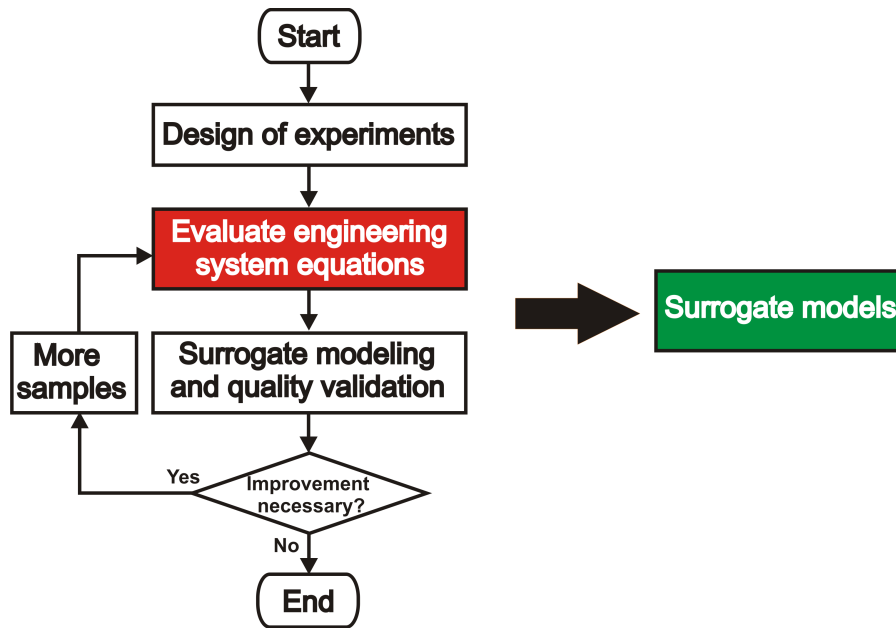
Figure 2.7.: The process of constructing surroagte models.

implemented within each iteration of optimization, which contains a certain number of independent evaluations for individual points or a limited number of independent gradient calculations. In surrogate-based design optimization, parallelization can be performed on a massive number of sample points all in one iteration as long as the hardware resource is sufficient. Therefore, surrogate-based design optimization has great potential of speeding up the engineering optimitation process.

Surrogate models are inexpensive to calculate but of low fidelity. Therefore, the quality of surrogate models should be validated and if necessary, refined before it is used in design optimization. Cross-validation [Hastie et al. (2008)] are usually used to test the quality of surrogate models. The accuracy of surrogate models is quantified by the normalized root mean square error (RMS) of the approximation at a number of test points. This is given by

$$\varepsilon_{\text{rms}} = \frac{1}{y_{\text{max}} - y_{\text{min}}} \sqrt{\frac{\sum_{i=1}^{n} \left[\hat{y}^{(i)} - y^{(i)}\right]^2}{n}}. \tag{2.18}$$

Good understanding of the optimization problem, as well as carefully chosen DoE method and surrogate modeling method should work in close collaboration to improve the efficiency and accuarcy of surrogate-based design optimization.

2.4. Summary of Chapter

In this chapter, methods for design of experiments are introduced first. The charicteristics of each method are discussed, which can be used as a guide for the selection of DoE for different problems. Then, the formulations of three typical surrogate models are given. The construction and effort for solving each model, as well as the applicability of different methods are described. Finally, the structure of surrogate-based design optimization is

demonstrated. The quantitative definition for the efficiency and accuracy of surrogate modeling are given. It is pointed out that the understanding of the optimization problem and knowledge of the engineering system should be considered the most important guide in surrogate-based design optimization.

3. Knowledge-Based Surrogate Modeling

The use of engineering knowledge can drastically improve efficiency and accuracy in surrogate-based design optimization, especially when a large number of design variables are involved. In this chapter, knowledge-based surrogate modeling will be introduced, which considers physical properties of the engineering system while constructing surrogate models.

3.1. Motivation and Overview of the Chapter

The engineering systems are generally understood by engineers and, therefore, are not of black-box nature. All knowledge of the engineering problem can be considered to assist the construction of surrogate models, which is studied and named as knowledge-based surrogate modeling (KBSM) in this dissertation. In able to use the engineering knowledge in surrogate models, focus in put on reducing the design space and selecting the quantities to be approximated. In other words, engineering knowledge is used in the study of this dissertation to choose modeling space and responses. As a result, in KBSM there is a mapping from the original design space and system responses to the surrogate modeling space and responses. The difference between the standard surrogate modeling and the knowledge-based surrogate modeling is demonstrated in Figure 3.1.

The sample points, which contain design sites x from design space and the evaluated system responses y are the information used in surrogate modeling methods. The standard surrogate modeling method only takes the information from sample points regardless of the physical properties of an engineering system. While, in knowledge-based surrogate modeling, the design space and system responses are mapped into modeling space p and modeling responses q based on the understanding of the corresponding engineering system. p and q are functions of x and y, respectively. By proper knowledge-based mapping, it is expected that dimension of the modeling space is reduced, which is

$$\mathbb{R}^{n_x} \longrightarrow \mathbb{R}^{n_p} \tag{3.1}$$

where $n_x \ll n_p$. n_x and n_p are the number of original design variables and mapped modeling variables, respectively. The $q(y)$ is then determined according to the knowledge of the system to be more directly related to modeling variables and it can be more accurately approxiamted. After knowledge-based mapping, surrogate modeling is no more a black-box process.

In this chapter, focus will be given to choosing function formulations of design variables
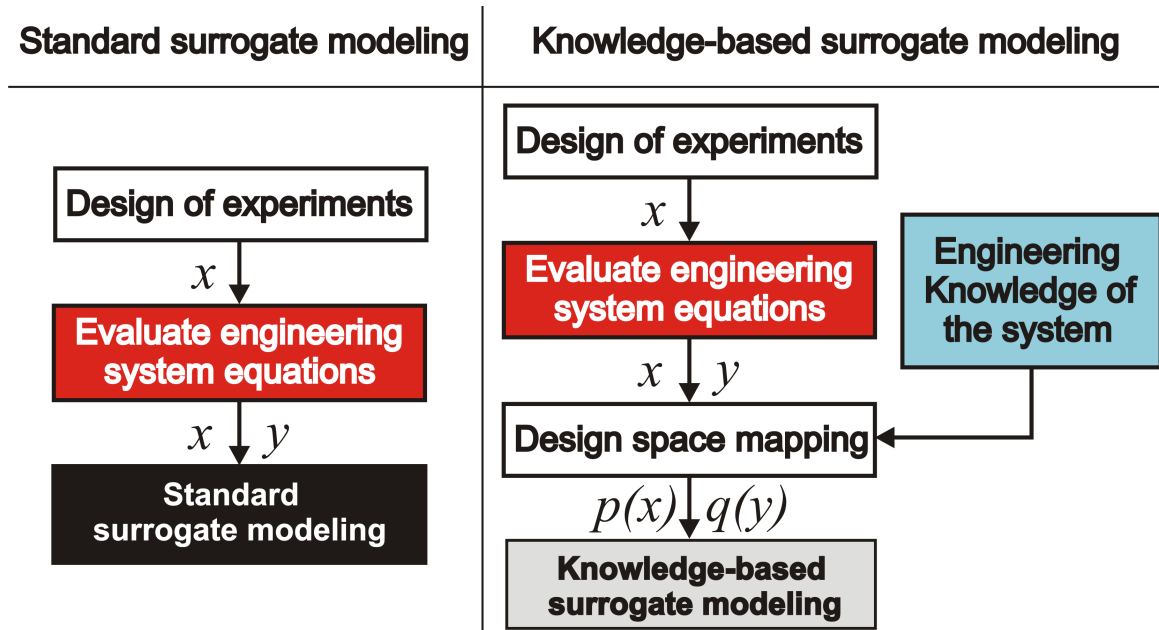
Figure 3.1.: The difference between the standard surrogate modeling and knowledge-based surrogate modeling methods.

and system responses for the mapping. The implementation of knowledge-based surrogate modeling is realized by embedding knowledge-based terms in the surrogate functions, which are mapped from design space and system responses to represent the system properties. With those terms forming modeling space and modeling responses, fewer surrogate parameters are to be determined and the relationship to be approximated becomes more directly related or even linearly related. As a result, efficiency and accuracy in surrogate-based design optimization can be improved. To show the implementation and the performance of knowledge-base surrogate modeling method, a small engineering problem is used as an introduction example. The efficiency and accuracy of the standard surrogate modeling and KBSM methods are compared.

It is to be noted that the emphasis of this chapter is to illustrate the approximation efficiency and quality of standard vs. knowledge-based surrogates, the comparison of design optimization behaviors of them are presented later in the following chapters.

3.2. Implementation of Knowledge-Based Surrogate Modeling with Kriging

As is introduced in section 2.2.3, the Kriging model is composed of a deterministic part and a stochastic part, represented by a regression model and a correlation model, respectively. The format of the regression and the correlation functions are rather flexible, which can be adjusted to different engineering problems. The kriging model constructed considering the mechanical system properties is called knowledge-based Kriging in this dissertation. It is

formulated as

$$\hat{q}(y(\boldsymbol{x})) = \mathcal{F}(\boldsymbol{p}(\boldsymbol{x}))) + Z(\boldsymbol{p}(\boldsymbol{x}))) = \overbrace{\sum_{i=1}^{k} \beta_i f_i(\boldsymbol{p}(\boldsymbol{x}))}^{\text{regression}} + \underbrace{Z(\boldsymbol{p}(\boldsymbol{x}))}_{\text{correlation}}, \tag{3.2}$$

where instead of $\boldsymbol{x}$, $\boldsymbol{p}(\boldsymbol{x})$ is used as the basic element in the regression model and the correlation model. $\boldsymbol{p}(\boldsymbol{x})$ is a vector of algebraical combinations of the design variables, which are formed based on engineering knowledge to represent, for example, the structural mechanical properties of a structural system. Since only physically sensible mixed-terms of design variables are considered to form the regression model, the total number of regression functions can be reduced. For example, unnecessary mixed-terms in a full quadratic polynomial regression model can be removed. $q(y)$ is a chosen function of system response y according to engineering knowledge, which forms a relationship that can be more accurately represented by Kriging. For example, it is a common knowledge in mechanical engineering that axial stress σ_a in a rod is inversely related to its cross-section area A_c, and therefore, the relation between $q(y) = \dfrac{1}{\sigma_a}$ and A_c can be more accurately approximated than the relation between $y = \sigma_a$ and A_c. The goal of implementing knowledge-based Kriging is to obtain following advantages:

- lower-dimensional surrogate modeling space,

- fewer coefficients β_i to be determined,

- smaller computational effort,

- higher approximation accuracy.

3.3. Knowledge-Based Terms for Beam and Shell Elements

In this section, the construction of knowledge-based modeling space for structural system is introduced. Beam and shell elements are frequently employed elements in aerospace and automobile industries. The formulations of element matrices for them will be extracted, and the knowledge-based mixed-terms for surrogate modeling will be generated from those formulations. To identify the mixed-terms of design variables, which are capable of representing the mechanical properties of a structural system, the force equilibrium is considered here

$$\boldsymbol{M}\ddot{\boldsymbol{u}} + \boldsymbol{C}\dot{\boldsymbol{u}} + \boldsymbol{K}\boldsymbol{u} = \boldsymbol{F}, \tag{3.3}$$

where $\boldsymbol{M}$, $\boldsymbol{C}$ and $\boldsymbol{K}$ are the global mass, damping and stiffness matrices, respectively. In elastic analysis, the load matrix $\boldsymbol{F}$ is often fixed and is not a variable of the geometric parameters, such as length, thickness or cross-section area of the structural components. The displacement vector $\boldsymbol{u}$ and the related variables such as velocity $\dot{\boldsymbol{u}}$, acceleration $\ddot{\boldsymbol{u}}$, as well

as the resulted stress and strain are dependent variables, which are the system responses to be analyzed. The mechanical behavior of the structural system depends on $\boldsymbol{K}$ and $\boldsymbol{M}$, which could be the desired combinations of design variables that directly represent the relation between geometric parameters and system responses. However, the global matrices are usually implicitly related to the design variables and the mixed-terms cannot be found directly, especially for large and complicate structural systems. The global matrices $\boldsymbol{K}$ and $\boldsymbol{M}$ are assembled from the element matrices $\boldsymbol{K}^e$ and $\boldsymbol{M}^e$, which can be explicitly derived with respect to the design variables. The detailed formulations of the mass and stiffness matrices for beam and shell elements, as well as their knowledge-based terms for surrogate modeling are derived in the following two sections, respectively. A demonstration example of surrogate modeling with the knowledge-based terms of beam and shell elements is given in section 3.5 of this chapter.

3.3.1. Knowledge-based terms for beam elements

For beam elements, the geometric parameters of the cross sections are usually considered as design variables in engineering design optimization. The axial, bending, shear and torsional stiffness as well as the translational and rotational inertia represent the mechanical properties of the beam element. To derive these terms, the complete stiffness matrix for a 12 degree-of-freedom beam element is given in Equation 3.4, as one can also refer to any basic mechanical literature, such as Gupta and J.L.Meek (2003).

$$\boldsymbol{K}^e = \begin{bmatrix} \boldsymbol{K}_{11} & \boldsymbol{K}_{12} \\ sym & \boldsymbol{K}_{22} \end{bmatrix} \tag{3.4}$$

in which

$$\boldsymbol{K}_{11} = \begin{bmatrix} \frac{EA}{l} & 0 & 0 & 0 & 0 & 0 \\ & \frac{12EI_z}{l^3\Psi_y} & 0 & 0 & 0 & \frac{6EI_z}{l^2\Psi_y} \\ & & \frac{12EI_y}{l^3\Psi_z} & 0 & -\frac{6EI_y}{l^2\Psi_z} & 0 \\ & sym & & \frac{GJ}{l} & 0 & 0 \\ & & & & \frac{\Psi'_z EI_y}{l} & 0 \\ & & & & & \frac{\Psi'_y EI_z}{l} \end{bmatrix}, \tag{3.5}$$

where E, G, A and l are the Young's modulus, the shear modulus, the cross-section area and the length of the beam element, respectively. I_y and I_z are the area moment of inertia with respect to y and z axes. Further, the following substitutions are used

$$\Psi_y = 1 + \frac{12EI_z}{GA_{sy}l^2}, \quad \Psi'_y = \frac{3+\Psi_y}{\Psi_y}$$

$$\Psi_z = 1 + \frac{12EI_y}{GA_{sz}l^2}, \quad \Psi'_z = \frac{3+\Psi_z}{\Psi_z}. \tag{3.6}$$

A_{sy} and A_{sz} in the substitutions stand for effective shear cross-sectional areas. $\boldsymbol{K}_{12}$ and $\boldsymbol{K}_{22}$ can also be derived analogously. In these equations, the terms EA, $\frac{EI}{\Psi}$, $EI\Psi'$, and GJ are combinations of design variables that represent mechanical properties of axial, bending,

shear and torsional stiffness, respectively. Similarly, the element mass matrix $\boldsymbol{M}^e$ is also given

$$\boldsymbol{M}^e = \begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ sym & \boldsymbol{M}_{22} \end{bmatrix}, \tag{3.7}$$

in which

$$\boldsymbol{M}_{11} = \begin{bmatrix} \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ & \frac{13}{53} + \frac{6I_z}{5Al^2} & 0 & 0 & 0 & \frac{11l}{210} + \frac{I_z}{10Al} \\ & & \frac{13}{53} + \frac{6I_y}{5Al^2} & 0 & -\frac{11l}{210} - \frac{I_y}{10Al} & 0 \\ & sym & & \frac{J_x}{3A} & 0 & 0 \\ & & & & \frac{l^2}{105} + \frac{2I_y}{15A} & 0 \\ & & & & & \frac{l^2}{105} + \frac{2I_z}{15A} \end{bmatrix}. \tag{3.8}$$

$\boldsymbol{M}_{12}$ and $\boldsymbol{M}_{22}$ are derived analogously.. Terms such as $\frac{I}{A}$ and $\frac{J}{A}$ are mixed-terms of design variables representing translational and rotational inertia, respectively. As is discussed in 3.2, these mixed-terms of design variables derived from Equation 3.4 to Equation 3.8 can be used to form the modeling space for knowledge-based surrogate modeling. More specifically, these terms are used to replace $\boldsymbol{x}$ as the basic elements $\boldsymbol{p}(\boldsymbol{x})$ in the regression functions of Kriging. For approximation of different responses under particular load cases, different combinations of those mechanical properties can be selected to represent the system behaviors. Since the number of the terms for the mechanical properties of beam elements are small, the dimension of the modeling space can be controlled and the required number of sample points can be reduced, even when very large number of design variables are involved.

3.3.2. Knowledge-based terms for shell elements

For shell elements, the thicknesses are often considered as design variables in engineering sizing design optimization. To prevent local buckling of the shell structures, they are usually stiffened at certain distances, which may also be considered as design variables. Suppose that t, b and l represents the thickness, the horizontal and the vertical sizes, respectively. The membrane, bending and shear stiffness of a shell element can be expressed as follows. The membrane stiffness matrix is

$$\boldsymbol{K}^e_M = \frac{t}{4A} \begin{bmatrix} \boldsymbol{P} & \boldsymbol{R} \\ sym & \boldsymbol{Q} \end{bmatrix}, \tag{3.9}$$

and the bending stiffness is

$$\boldsymbol{K}^e_B = \frac{t^3}{48A} \begin{bmatrix} \boldsymbol{Q} & -\boldsymbol{R}^T \\ sym & \boldsymbol{P} \end{bmatrix}, \tag{3.10}$$

and the shear stiffness is

$$\boldsymbol{K}^e_S = \frac{t}{4A} \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} & \boldsymbol{S}_{13} \\ & \boldsymbol{S}_{22} & \boldsymbol{S}_{23} \\ sym & & \boldsymbol{S}_{33} \end{bmatrix}, \tag{3.11}$$

where $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{R}$ and $\boldsymbol{S}$ are position dependent matrices, which are composed of quadratic terms of the horizontal and the vertical sizes b and l, one can refer to Gupta and J.L.Meek (2003) for details. $\boldsymbol{A}$ represents the surface area, which is directly proportional to the production bh. According to the Equations 3.9 to 3.11, t, t^3, $\frac{b}{l}$ and $\frac{l}{b}$, as well as the inverse of them should be considered as knowledge-based terms for shell elements in a structure. The selected combinations of those terms are then used to form the modeling space to perform knowledge-based surrogate modeling.

3.4. Knowledge-Based Terms for Laminated Fiber-Composite Plate

Owing to the high strength-to-weight ratio and many other properties, composite materials have got widespread applications in the latest products of the aerospace and automobile industries. For example, the rate of composite structure to structural weight is about 25% in Airbus 380, 53% in Airbus 350XWB [Hellard (2008)] and 50% in Boeing 787. Fiber-reinforced composite is a type of composites, which includes carbon-fiber-reinforced polymers (CFRP), glass-fiber-reinforced plastic (GRP), metal-fiber-reinforced metals and etc. Mechanical properties of fiber-reinforced composites can be tailored to suit to specific applications very flexibly by, e.g. selecting materials of fibers and matrices, orientating fibers, adjusting thicknesses of plies, stacking plies with different sequences or combinations of the above. Analysis, design and optimization of fiber-reinforced composite structures usually involves large number of design variables (e.g., orientations and thicknesses of all plies of laminates), highly-nonlinear responses and nonconvex design space. Such problems indicate a requirement for large number of system evaluations in design optimization. Due to computationally expensive high-fidelity simulations of many problems, surrogate models are desired. However, the construction of standard surrogate models becomes inefficient owing to the need of vast number of sample points. Therefore, the adoption of engineering knowledge in surrogate modeling for structures containing composite materials becomes imperative. Laminated fiber-composite plates is used as an example in this dissertation to demonstrate the necessary theory of composites, which is employed to assist knowledge-based surrogate modeling.

Similar with using stiffness of isotropic elastic materials as knowledge-based terms, stiffness terms of a structure with composite materials is also explored for knowledge-based surrogate modeling. Considering that laminated fiber-composite plates commonly used in aircraft and automobile structures are usually relatively thin and flat, the classical lamination theory (CLT), as one can refer to B.D.Agarwal and Broutman (1980), is adopted to formulate knowledge-based terms for surrogate modeling. The typical constitutive equation according to CLT describes the relation between loads and structural displacement, which is

$$
\left\{ \begin{array}{c} N_r \\ M_r \end{array} \right\} = \left[\begin{array}{cc} A & B \\ B & D \end{array} \right] \left\{ \begin{array}{c} \varepsilon^0 \\ k \end{array} \right\}, \tag{3.12}
$$

where $\boldsymbol{N}_r$ and $\boldsymbol{M}_r$ are vectors representing the forces and moments in a laminate cross section. $\boldsymbol{\varepsilon}^0$ and $\boldsymbol{k}$ are the mid-plane strains and the plate curvatures, respectively. $\boldsymbol{A}$ is the in-plane stiffness matrix, $\boldsymbol{B}$ is the coupling stiffness matrix, and $\boldsymbol{D}$ is the bending stiffness matrix. For thin laminates, which are usually simplified as two-dimensional plates, $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ are reduced from the full 6×6 matrices to 3×3 matrices. All the three matrices are symmetric, and as a result, there are six independent matrix elements for each of the $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ matrices, which are derived from the lamina stiffness matrix $\bar{\boldsymbol{Q}}$ in the global coordinate system. This is given by

$$\boldsymbol{A}_{ij} = \sum_{k=0}^{n} (\bar{Q}_{ij})_k (h_k - h_{k-1})$$

$$\boldsymbol{B}_{ij} = \tfrac{1}{2} \sum_{k=0}^{n} (\bar{Q}_{ij})_k (h_k^2 - h_{k-1}^2) \tag{3.13}$$

$$\boldsymbol{D}_{ij} = \tfrac{1}{3} \sum_{k=0}^{n} (\bar{Q}_{ij})_k (h_k^3 - h_{k-1}^3).$$

The two in-plane axes of the local/lamina coordinate are so defined that one principle axis is parallel to the fiber orientation and the other principle axis is orthogonal to it. The transformation from local stiffness matrix $\boldsymbol{Q}$ to the global stiffness matrix $\bar{\boldsymbol{Q}}$ is expressed as:

$$\left[\bar{\boldsymbol{Q}} \right] = \left[\boldsymbol{T} \right]^{-1} \left[\boldsymbol{Q} \right] \left[\boldsymbol{R} \right] \left[\boldsymbol{T} \right] \left[\boldsymbol{R} \right]^{-1}, \tag{3.14}$$

where,

$$[\boldsymbol{T}] = \begin{bmatrix} \cos^2 \alpha & \sin^2 \alpha & 2 \sin \alpha \cos \alpha \\ \sin^2 \alpha & \cos^2 \alpha & -2 \sin \alpha \cos \alpha \\ -\sin \alpha \cos \alpha & \sin \alpha \cos \alpha & \cos^2 \alpha - \sin^2 \alpha \end{bmatrix}, \tag{3.15}$$

and

$$[\boldsymbol{R}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \tag{3.16}$$

$\boldsymbol{\alpha}$ in Equation 3.15 is a vector of fiber angles in a laminate. The global stiffness matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{D}$ and $\bar{\boldsymbol{Q}}$ are combination of design variables $\boldsymbol{\alpha}$. They are capable of directly represent the system properties of laminated fiber composite plates. Usually, rather than the relationship in Equation 3.12, the stresses and strains in a laminate are to be determined for known loads. As discussed in section 3.2, it improves the accuracy of approximation by performing inverse operations in mapped modeling space and modeling responses instead of approximating inverse relationship in surrogate models. To achieve this, the system equations are rewritten as

$$\left\{ \begin{array}{c} \varepsilon^0 \\ k \end{array} \right\} = \left[\begin{array}{cc} \boldsymbol{A}^* & \boldsymbol{B}^* \\ \boldsymbol{B}^* & \boldsymbol{D}^* \end{array} \right] \left\{ \begin{array}{c} \boldsymbol{N} \\ \boldsymbol{M} \end{array} \right\} \tag{3.17}$$

where,

$$\begin{aligned} [\boldsymbol{A}^*] &= [\boldsymbol{A}^{-1}] - [\boldsymbol{A}^{-1}] [\boldsymbol{B}] [[\boldsymbol{D}] - [\boldsymbol{B}] [\boldsymbol{A}^{-1}] [\boldsymbol{B}]]^{-1} [[\boldsymbol{A}^{-1}] [\boldsymbol{B}]]^T \\ [\boldsymbol{B}^*] &= -[\boldsymbol{A}^{-1}] [\boldsymbol{B}] [[\boldsymbol{D}] - [\boldsymbol{B}] [\boldsymbol{A}^{-1}] [\boldsymbol{B}]]^{-1} \\ [\boldsymbol{C}^*] &= [\boldsymbol{B}^*] \\ [\boldsymbol{D}^*] &= [[\boldsymbol{D}] - [\boldsymbol{B}] [\boldsymbol{A}^{-1}] [\boldsymbol{B}]]^{-1} \end{aligned} \tag{3.18}$$

The stresses on a lamina are then calculated as:

$$\sigma_k = \left[\bar{Q}\right]_k \varepsilon^0 + z \left[\bar{Q}\right]_k k \tag{3.19}$$

Therefore, the laminate stresses and strains are directly related to the terms in matrices A^*, B^*, D^* and $\bar{Q}$, namely the compliance matrices of the laminate. Using elements of such matrices instead of the original design variables α to form a modeling space allows more accurate approximation of the system responses. The system responses such as stresses, strains, and deflection are approximated with linear or quadratic relations of these matrix elements. Selected combinations of these matrix elements are then used in Kriging as knowledge-based terms $p(x)$ to perform surrogate modeling for different system responses. With engineering knowledge the required number of sample points is greatly reduced when the number of fiber angles is very large. Application examples of knowledge-based surrogate modeling for laminated composite panel can be found in section 4.6 of this dissertation, and design optimization of the composite plate is given in section 5.7.

3.5. Introductory Example: Knowledge-Based Surrogate Modeling for an I-Beam Cantilever

To demonstrate how engineering knowledge of a structural system is used to assist surrogate modeling, the design problem of an I-beam cantilever is given as an introductory example.

3.5.1. Introduction of the I-Beam Cantilever

The cantilever model is loaded with vertical force at the tip and is clamped at the opposite end, shown in Figure 3.2. The cross section and the geometric design variables of the I-beam
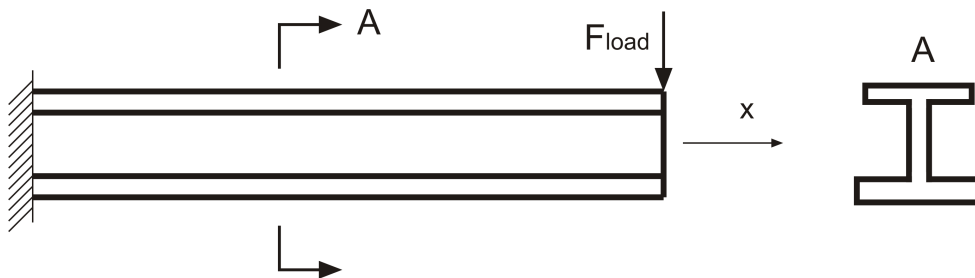
Figure 3.2.: The cantilever model under vertical load at the tip.

cantilever are shown in Figure 3.3.

The engineering design optimization for this cantilever model is so defined, that the geometric parameters are to be determined to minimize the structural mass subject to a number of structural constraints. The design of the I-beam cantilever involves six design variables and six design constraints. The structural constraints include strength, stiffness and stability requirements, which are described as follows:
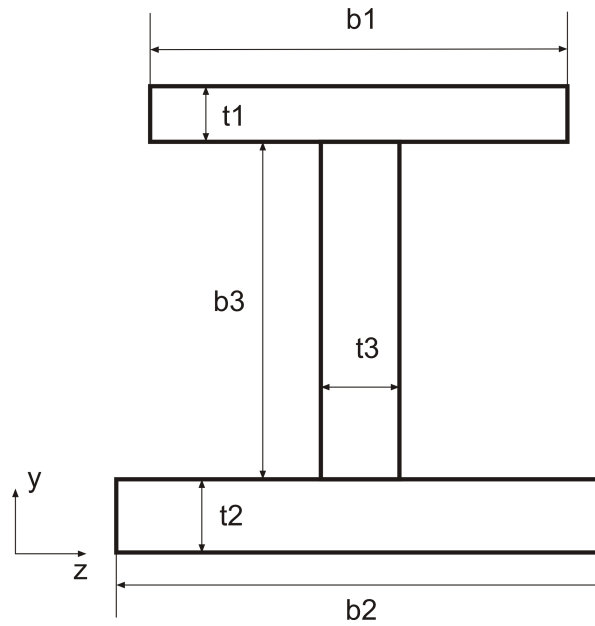
Figure 3.3.: Cross section of a cantilever with geometrical design variables for optimization.

- The maximum bending stress σ_b resulting in the cantilever should not be larger than the yield stress of the material σ_y. That is $\sigma_b < \sigma_y$, which is a strength requirements so that the structure would not deform plastically.

- The maximum shear stress τ in the cantilever should not exceed half of the yield stress of the material. That is $\tau < \frac{\sigma_y}{2}$, a strength constrain to prevent the ductile material used for the cantilever from failure in shear.

- The tip deflection δ of the cantilever should not exceed a critical value δ_{crit}. That is $\delta < \delta_{\mathrm{crit}}$, so that the cantilever is stiff enough to prevent overlarge deformation.

- The critical force F_c which actives the twist of the structure around the x-axis must be larger than the loaded force F_{load}. That is $F_c > F_{\mathrm{load}}$, which is a stability constraint so that the coupled bending-torsion deformation does not happen under the given load.

- The first resonance frequency of bending f_b must be greater than a given frequency value f_0. That is $f_b > f_0$, so that structural resonance is avoided when the given force is simple harmonic with frequency up to f_0.

- The critical buckling stress σ_c should be larger than the yield stress of the material. That is $\sigma_c > \sigma_y$, which is a local stability constraint to avoid inelastic buckling of web. The web of the cross-section is actually a thin-walled plate that has the potential to buckle locally under the given load.

Therefore, the design optimization problem is formulated as follows:

$$
\begin{aligned}
\text{minimize} \quad & m(\boldsymbol{x}) \\
\text{such that} \quad & \frac{\sigma_b(\boldsymbol{x})}{\sigma_y} - 1 \leq 0 \\
& \frac{2\tau(\boldsymbol{x})}{\sigma_y} - 1 \leq 0 \\
& \frac{\delta(\boldsymbol{x})}{\delta_{\text{crit}}} - 1 \leq 0 \\
& 1 - \frac{F_{\text{c}}(\boldsymbol{x})}{F_{\text{load}}} \leq 0 \\
& 1 - \frac{f_b(\boldsymbol{x})}{f_0} \leq 0 \\
& 1 - \frac{\sigma_{\text{c}}(\boldsymbol{x})}{\sigma_y} \leq 0 \\
\text{and} \quad & \boldsymbol{x}_l \leq \boldsymbol{x} \leq \boldsymbol{x}_u,
\end{aligned}
\tag{3.20}
$$

where $\boldsymbol{x} = \{b_1, b_2, b_3, t_1, t_2, t_3\}^T$ (see Fig. 3.3) is a vector of the design variables. $\boldsymbol{x}_l$ and $\boldsymbol{x}_u$ refer to the lower and upper boundaries of the design variables. m is the mass of the structure and is the objective function of the optimization problem.

3.5.2. Task Definition for the I-beam Cantilever

From the definition of the cantilever optimization problem, it is known that system responses to be analyzed are mass m, the maximum bending stress σ_b, the maximum shear stress τ, the tip deflection δ, the critical force F_{c} for twisting, the first resonance frequency of bending f_b and the critical buckling stress σ_{c}. As is know that the mass of a structure can be usually explicitly derived with respect to the design variables very efficiently and is not needed to be approximated with a surrogate model. The task in this introduction example is, therefore, to build surrogate models for all system responses in the constraint functions. It is to be noted that the computational effort to calculate the system responses are actually extremely low for this cantilever model. The purpose of the example is, however, to demonstrate the effect of utilizing engineering knowledge in surrogate modeling. Surrogate models of the system responses σ_b, τ, δ, F_{c}, f_b and σ_{c} are built with respect to $\boldsymbol{x}$ using both standard surrogate modeling and knowledge-based surrogate modeling. The quality of surrogate models will be judged by both efficiency and accuracy, quantified by the required number of samples and the relative root mean square error, respectively.

3.5.3. Standard Surrogate Modeling for the I-Beam Cantilever

Standard surrogate modeling using Kriging method with a quadratic polynomial regression model is performed. The design optimization problem involves six design variables. For six-dimensional design space, according to Equation 2.2, at least 28 sample points are required to formulate a full quadratic polynomial regression model. Considering an oversampling factor of 1.3 (the typical oversampling factor is around 1.5, and higher for more

complicate problems [Sieberz et al. (2010))], 36 sample points are generated by Latin hyper-cube sampling. For this simple example, we can generate sufficient test points to visualize the quality of different surrogate models. For example, considering a full factorial level of five for the most critical design variable b_3, four for t_1, t_2 and t_3, and three for b_1 and b_2, then the total number of test points is 2880. The approximation results for all the six responses are shown in Figure 3.4.

It shows the normalized values of σ_b, τ, δ, F_c, f_b and σ_c from the left to the right of the
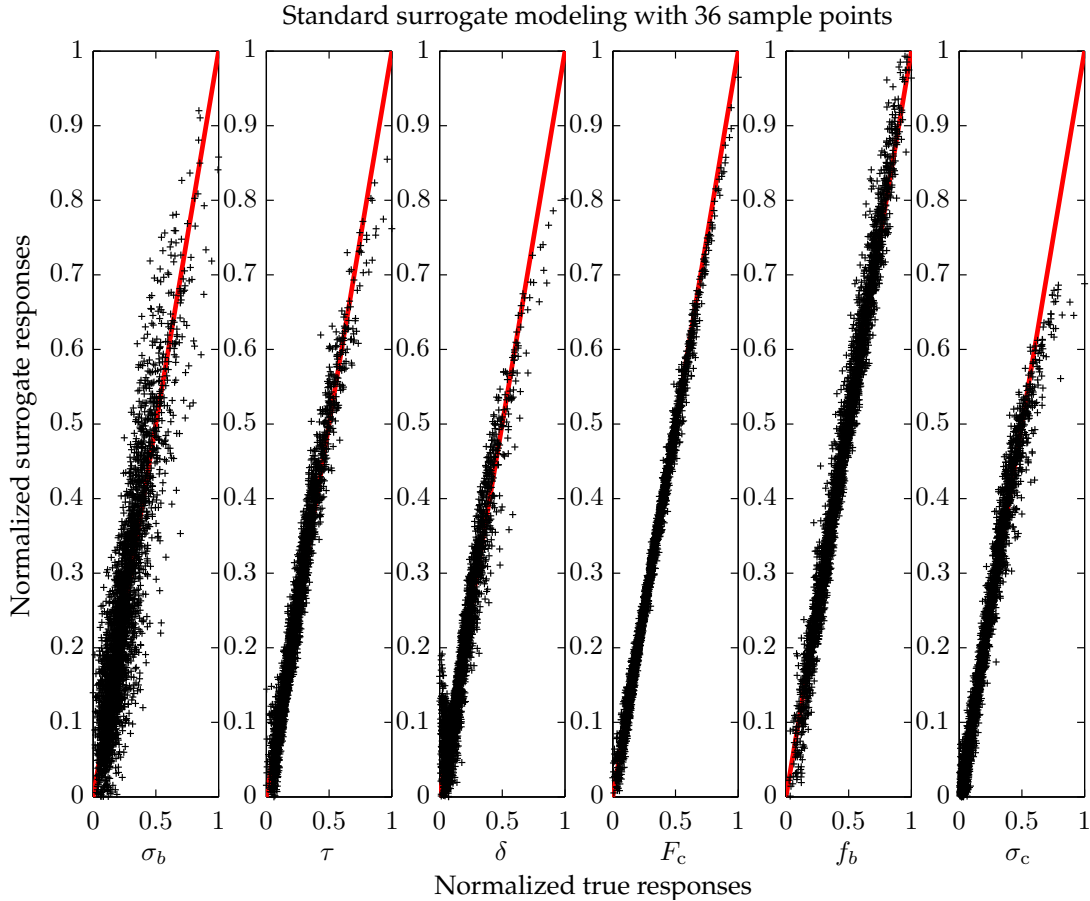
Figure 3.4.: Standard surrogate modeling results of the I-beam cantilever.

graphs in Figure 3.4. More specifically, the consistency of the response values predicted by surrogate models at the testing points with the exact system responses are shown. The purpose is to give an intuitive impression of the approximation quality of the surrogate models. The horizontal axis stands for the true system responses and the vertical axis stands for the surrogate system responses. Therefore, the diagonal line in each graph indicates the exact system model. The closer the predicted values of the surrogate models are to that line, the better the approximation accuracy of the models is. As can be seen, there is a large potential to further improve the quality of the standard surrogate models, especially those for the bending stress σ_b, tip deflection δ and the critical buckling stress σ_c.

3.5.4. Knowledge-Based Surrogate Modeling for the I-Beam Cantilever

To compare with the standard surrogate modeling and illustrate the effect of utilizing engineering knowledge, the knowledge-based surrogate modeling is performed in this section. As is discussed in section 3.3, according to the knowledge of the structural mechanics, the system responses of the I-beam cantilever depend on the geometrical and mechanical properties. Those properties are the cross-sectional area A, the position of the center of the cross section, the area moment of inertia I_z and I_y, etc. For different system responses, different terms should be selected according to basic engineering knowledge and understanding of the structure. For the above described system responses of the I-beam cantilever, the chosen knowledge-based terms of each response are discussed as follows:

- The maximum bending stress σ_b is inversely proportional to the area moment of inertia I_z of the cross section, and is directly proportional to the distance from the geometrical center h_{max}.

- The maximum shear stress τ is inversely proportional to the cross-section area.

- The tip deflection δ is inversely proportional to the area moment of inertia I_z.

- The critical force F_{c} is depend on both the torsional rigidity and the bending rigidity of the cross section of the cantilever. Therefore, the polar area moment of inertia about the x axis, $J_x = I_y + I_z$, and the area moment of inertia about the z axis, I_z are selected.

- The first resonance frequency of bending f_b is decided by the square root of the stiffness to mass ratio.

- The critical local buckling stress σ_{c} is related to be slenderness ratio of the cross section according to the knowledge-based terms for shells. Therefore, the thickness to height ratio of the web $\frac{t_3}{b_3}$ and the ratio of the average width of the flange to the height of the web $\frac{b_1+b_2}{2b_3}$ are chosen.

More intuitively, the knowledge-based terms for each of the structural responses are summarized in Table 3.1. These listed terms are used as knowledge-based terms $\boldsymbol{p}(\boldsymbol{x})$ to form

Table 3.1.: The structural responses and the respective geometrical properties for knowledge-based surrogate modeling.

Structural responses	σ_b	τ	δ	F_{c}	f_b	σ_{c}
Geometrical properties	$h_{\mathrm{max}}, \frac{1}{I_z}$	$\frac{1}{A}$	$\frac{1}{T_z}$	J_x, I_z	$\sqrt{\frac{I_z}{A}}$	$\frac{t_3}{b_3}, \frac{b_1+b_2}{2b_3}$

regression and correlation functions in Kriging. Since the geometrical properties J_x, I_z, A, the slenderness ratio and so on are directly related to the system responses, fewer coefficients in the regression model are to be determined so that fewer sample points are required. Knowledge-based surrogate models are, therefore, constructed with only 12 sample points and the results is shown in Figure 3.5.
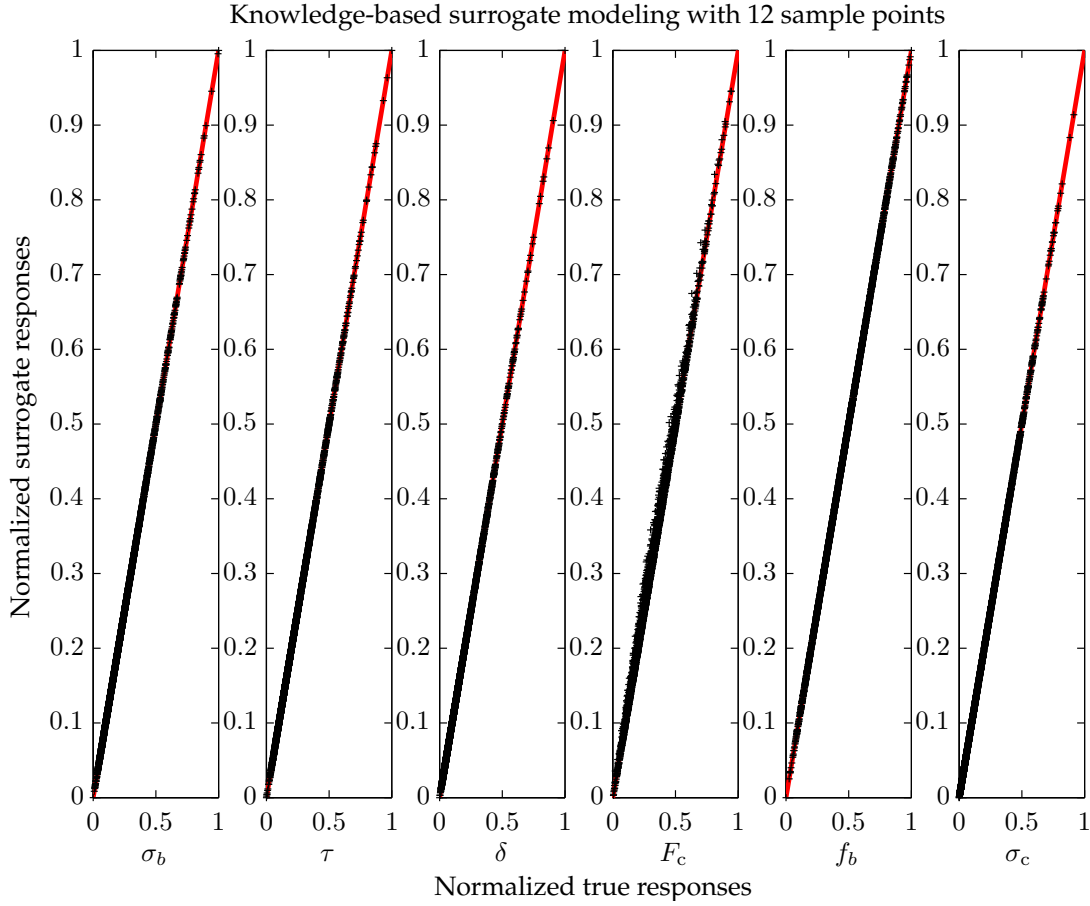
Figure 3.5.: Knowledge-based surrogate modeling results of the I-beam cantilever.

3.5.5. Comparison of Surrogate Models for the I-Beam Cantilever

As is expected, from Figure 3.4 and 3.5 it can be seen that the approximation quality has been greatly improved by knowledge-based surrogate modeling in comparison with the standard surrogate modeling. To quantitatively describe the improvements, the normalized root mean square error (RMS) (see section 2.3) of surrogate models of both methods are given in Table 3.2.

In Table 3.2, n_{samp} represents the number of sample points used in surrogate modeling.

Table 3.2.: Results of the standard surrogate modeling method and the knowledge-based surrogate modeling method for the I-beam cantilever.

Surrogate models	n_{samp}	$\varepsilon_{\mathrm{rms}}(\sigma_b)$	$\varepsilon_{\mathrm{rms}}(\tau)$	$\varepsilon_{\mathrm{rms}}(\delta)$	$\varepsilon_{\mathrm{rms}}(F_{\mathrm{c}})$	$\varepsilon_{\mathrm{rms}}(f_b)$	$\varepsilon_{\mathrm{rms}}(\sigma_{\mathrm{c}})$
Standard	36	7.46%	3.08%	3.75%	1.44%	3.35%	3.19%
Knowledge-based	12	0.00%	0.00%	0.00%	0.96%	0.00%	0.00%

It can be seen that the knowledge-based surrogate modeling requires only $\frac{1}{3}$ of the sample points used in standard surrogate modeling to approximate all the system responses with even much higher accuracy. As a result, it is shown with this simple introductive example

that the utilization of engineering knowledge is essential for improving efficiency and accuracy in surrogate modeling. For large and complicate mechanical structures, the mechanical properties would not be as obvious as the cantilever example shown here. However, according to the derivation and discussion in section 3.3 and sections 3.4, there are clues to be found from engineering knowledge, understanding of the system and engineering experiences as well.

3.6. Summary of Chapter

In this chapter, knowledge-based surrogate modeling method is introduced. The implementation of engineering knowledge in surrogate modeling is achieved by using engineering knowledge-based terms in the Kriging model. By doing this, the dimension of modeling space can be reduced and the relationships to be approximated are simplified. This allows to approximate engineering system responses with much higher efficiency and accuracy. Knowledge-based terms for the frequently employed elements in aerospace and automobile industries are given. A simple cantilever model is used as an example to introduce and demonstrate the knowledge-based surrogate modeling method. Comparison of the standard surrogate modeling and knowledge-based surrogate modeling on this example proves the advantage of using engineering knowledge in surrogate modeling.

4. Genetic Algorithm Assisted Knowledge-Based Surrogate Modeling

A technique is developed in this chapter to assist the formulation of regression functions in knowledge-based surrogate modeling (KBSM). This method is aimed at achieving more accurate approximation of system equations for complicated engineering structures, where knowledge of the system is not complete. Based on the existing sample points, a genetic algorithm (GA) is used to search for the best combination of knowledge-based terms and the format of regression functions in surrogate models. An connection between KBSM and traditional GA programming is created, where candidate functions and operations are defined to formulate the design variables and fitness function for implementation of GA. The method is called genetic algorithm assisted KBSM (GAKBSM).

4.1. Motivation and Overview of the Chapter

As is indicated in Chapter 3, the knowledge-based surrogate modeling method can be expected to have better performance in both approximation accuracy and efficiency than standard surrogate modeling. In KBSM the knowledge-based terms are the candidate terms to be operated on instead of the original design variables of an engineering design problem. The relation between knowledge-based terms and the system responses are usually easier to be represented with simple formats of functions, such as polynomials. However, the knowledge of complicated engineering structures is often not complete. Usually a number of knowledge-based terms are available according to mechanical theories and engineering experience, but the detailed format of those terms in the system equations are not clear. Therefore, it is necessary to develop a general method to search for the best combinations and formats of the knowledge-based terms to assist surrogate modeling, which is in this dissertation the genetic algorithm assisted KBSM (GAKBSM). Since the surrogate modeling techniques are implemented based on the Kriging model, the GA assisted method is introduced by taking Kriging as an example, which can be however also implemented on other kinds of surrogate models. The formulation of the standard Kriging model and the GA assisted Kriging (GA-Kriging) is given as follows:

$$\text{Standard Kriging} \quad : \quad \hat{y} = \mathcal{F}(\boldsymbol{x}) + Z(\boldsymbol{x})$$

$$\text{GA-Kriging} \quad : \quad \hat{y} = \mathcal{F}^*(\boldsymbol{x}) + Z(\boldsymbol{x})$$

(4.1)

```
┌─────────────────────────────────┐
│     Define candidate functions  │
│        and initialize GA        │
└─────────────────────────────────┘
                │ $\mathcal{F}$
                ▼
┌─────────────────────────────────┐
│        Surrogate modeling       │
│       and cross validation      │
└─────────────────────────────────┘
                │ $\varepsilon_{\text{rms}}^{(i)}$
                ▼
┌─────────────────────────────────┐
│        Optimization with GA     │
│    min $\sum \varepsilon_{\text{rms}}^{(i)}$        │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│                                 │
│           $\mathcal{F}^*$               │
│                                 │
└─────────────────────────────────┘
```
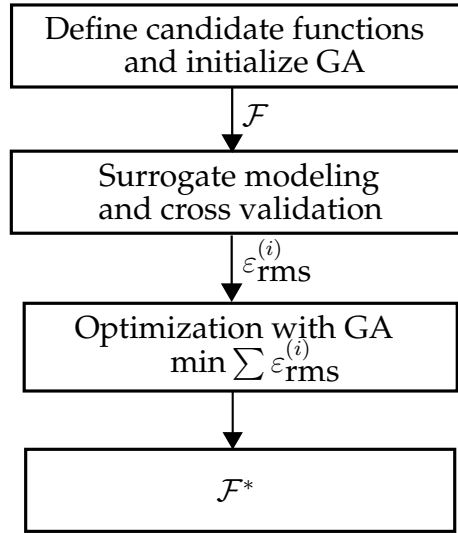
Figure 4.1.: The flowchart of genetic algorithm assisted surrogate modeling.

As can be seen in Equation 4.1, the standard Kriging is the basic one, which contains a regression model $\mathcal{F}(x)$ and a correlation model $Z(x)$. The GA-Kriging uses an optimized regression model, which is obtained by searching for the best formulations with GA. The flowchart of GA assisted surrogate modeling is shown in Figure 4.1.

First, a number of candidate functions are provided to generate the basic terms for forming regression functions, which can be obtained with the knowledge of an engineering system. The GA is initialized, which gives a generation of individuals with different regression models. The surrogate models are then constructed and cross validation is used to measure the quality of the surrogates in the form of root mean square error $\varepsilon_{\text{rms}}^{(i)}$. Cross validation is a kind of validatation method, which divides the sample points into several groups, while each group plays the role of test points in turn. Details of cross validation will be given in the following of this chapter. By minimizing the summation of root mean square errors $\sum \varepsilon_{\text{rms}}^{(i)}$, the optimal regression model is finally found.

Refering to section 3.2, the knowledge-based changes the modeling inputs and outputs from x and y to $p(x)$ and $q(y)$, which are the mapped modeling space and responses, according to engineering understanding of the system. The knowledge-based Kriging and the GA assisted Kriging can be combined, which is GAKB-Kriging and is formulated as:

$$\text{GAKB-Kriging} \ : \ q(\hat{y}) = \mathcal{F}^*(p(x)) + Z(p(x)) \tag{4.2}$$

The process is described in details in the rest of this chapter. Reasons for developing the genetic algorithm assisted knowledge-based surrogate modeling technique are listed as follows:

(i) Although a number of knowledge-based terms can be identified to support surrogate modeling, the function formulation with those terms for best representation of the engineering physical behavior needs to be decided.

(ii) GA is a well developed heuristic search and optimization method. The use of GA to

search for the best formulation of the regression functions in surrogate modeling is a more general solution for engineering approximation problems.

(iii) GA is available as a toolbox in numerical computing environment, such as MATLAB. Therefore, no effort is taken on programming the codes and focus can be put on implementing the connection between GA and KBSM.

(iv) The best combination of knowledge-based terms can be selected from a wide variety of function forms with GA. Therefore, the formulation of the regression functions is more flexible, which is possible to gain better approximation quality than fitting with the fixed formats.

(v) Besides, for engineering problems containing large number of design variables, the sample size required by genetic algorithm assisted knowledge-based surrogate modeling is no more as large as that required by full quadratic or higher-order polynomial functions.

(vi) Last but not least, the searching process only involves the evaluation of surrogate models, no additional high-fidelity evaluation is necessary. Though many iterations might be needed for the search, it is computationally inexpensive. Therefore, it is possible to make best use of the sample points with humble computational effort.

In this chapter, a brief introduction of GA is given first. The procedure of GA is shown and its interface with knowledge-based surrogate modeling is pointed out. Then the optimization problem of GAKBSM for searching the best function formats of knowledge-based terms is defined. Details of GAKBSM are explained, which include the definition, the creation and the evaluation of the fitness function, followed by the introduction of the complete process. Finally, the demonstration example is given to show the procedure and performance of GAKBSM in approximation of engineering problems.

4.2. Genetic Algorithm

Genetic algorithm is a heuristic search and optimization method, which imitates the inheritance and evolution process of the nature [Barricelli (1957)]. The individuals are selected according to their fitness along the evolutionary process and new generations are produced by inheriting and mutating properties of their parents [Arora (2004)]. Such process goes on for a number of generations or until the fitness of the generation is stabilized. A general process of GA can be shown with the flowchart in Figure 4.2. It has to be noted that usually the GA is used to find the optimized solution of a design problem. However, in the GAKBSM, GA is used to find the best function format of the knowledge-based terms to fit with a number of sample points. To introduce the interface of the GA with knowledge-based surrogate modeling, the terms in the flowchart are interpreted accordingly.

- Design variables:
  The design variables here are not the design variables of the engineering design problem, but the values indicating particular formats of functions. To differ them in this

```
                         ( Start )
                            │
                            ▼
              ┌─────────────────────────────┐
              │      Initial population      │
              │      (design variables)      │
              └─────────────────────────────┘
                            │
                            ▼
              ┌─────────────────────────────┐
              │      Fitness evaluation      │
              │      (system equations)      │
              └─────────────────────────────┘
                            │
                            ▼
              ┌─────────────────────────────┐
              │      Initial generation      │
              └─────────────────────────────┘
                            │
                            ▼
              ┌─────────────────────────────┐
              │          Selection           │◄──────────┐
              │    recombination, mutation   │           │
              └─────────────────────────────┘           │
                            │                            │
                            ▼                            │
              ┌─────────────────────────────┐            │
              │        New population        │           │
              └─────────────────────────────┘           │
                            │                            │
                            ▼                            │
              ┌─────────────────────────────┐            │
              │      Fitness evaluation      │           │
              │      (system equations)      │           │
              └─────────────────────────────┘           │
                            │                            │
                            ▼                            │
              ┌─────────────────────────────┐            │
              │        New generation        │           │
              └─────────────────────────────┘           │
                            │         No                 │
                            ▼  ◄──────────────────────────┘
                        ◇ Stop ◇
                       Yes │
                            ▼
                         ( End )
```
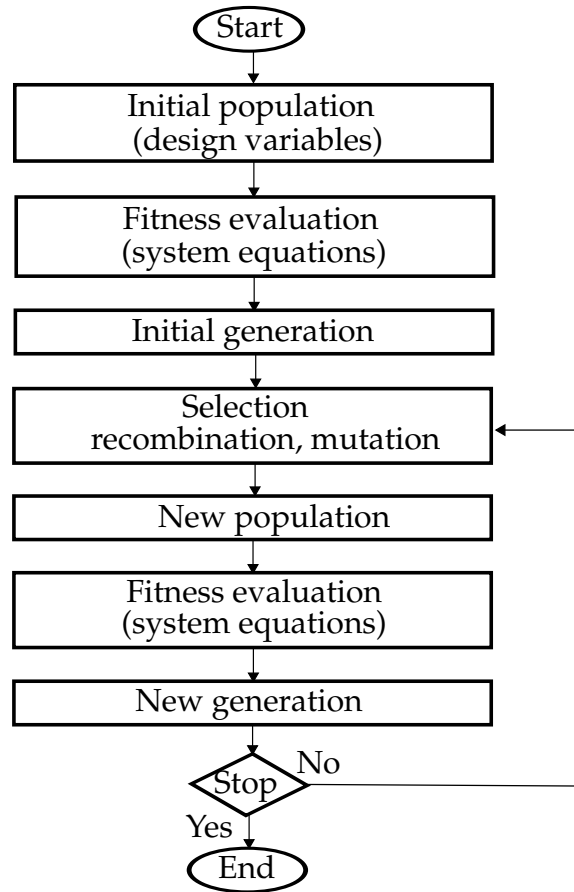
Figure 4.2.: The flowchart of a genetic algorithm.

dissertation, the variables indicating function formats in GAKBSM is noted as x_{fid}, while design variables of the engineering design problem is noted as x.

- Initial population:
  This is a randomly generated group of individuals. Each individual is a specifical assignment of x_{fid}, which represents one format of regression functions.

- Fitness evaluation:
  This evaluates the objective functions and constraint functions (if exist) of the individuals to assess the goodness of the regression functions, as well as the specific assignment of x_{fid}. In GAKBSM, only one objective function is involved, and there is no constraint function. The objective function in this case is the approximation error, noted as $\mathrm{Loss}(x_{\mathrm{fid}})$. $\mathrm{Loss}(x_{\mathrm{fid}})$ means the summation of approximation error $\sum \varepsilon_{\mathrm{rms}}^{(i)}$ of the surrogate models constructed by adopting the regression functions according to x_{fid}.

- Selection, recombination and mutation:
  These are the GA operations for producing better individuals during the evolutionary process, so that the regression function formats for surrogate models are also improved progressively for better approximation accuracy.

- Stopping criterion:

  In GAKBSM, GA is stopped when the formulation of regression functions is converged or the maximum number of generation is reached.

4.3. Definition of the Optimization Problem in Genetic Algorithm Assisted Knowledge-Based Surrogate Modeling

GAKBSM is aimed at searching for the best formats of regression functions, so that the surrogate models can approximate the engineering system equations with least error. The search of different function formats is a discrete problem. Continuous variables are divided into uniform segments, each of which is used as an indicator of a function format. The approximation error is assessed by the loss in a cross-validation method. In Stone (1974), the loss is defined as the average root mean square error of an estimation cross-validated at all test points. In this dissertation, the summation of root mean square error is used, which differs from the definition of Stone (1974) only by a scaling factor and is basically the same for optimization. The optimization task of GAKBSM is, therefore, mathematically formulated as follows:

$$
\begin{aligned}
\text{minimize} \quad & \text{Loss}(\boldsymbol{x}_{\text{fid}}) = \sum_{i=1}^{5} \varepsilon_{\text{rms}}^{(i)} \\
\text{such that} \quad & \mathbf{0} \leq \boldsymbol{x}_{\text{fid}} < \mathbf{1},
\end{aligned}
\tag{4.3}
$$

where $\boldsymbol{x}_{\text{fid}}$ is a vector of continuous values in the normalized domain $[0, 1)$, which indicates particular formats of regression functions. The mapping of the $\boldsymbol{x}_{\text{fid}}$ to formats of regression functions is introduced in Section 4.4.1. $\text{Loss}(\boldsymbol{x}_{\text{fid}})$ measures the quality of the regression model by summation of root mean square error $\varepsilon_{\text{rms}}^{(i)}$ from cross validation. According to a general rule of cross validation described in Forrester et al. (2008), the sample points are divided into 5 groups, while each group plays the role of test points in turn. For every regression model there are in total five different surrogate models constructed according to different groups of sample points. Each surrogate model is validated by a group of test points and a $\varepsilon_{\text{rms}}^{(i)}$ is obtained. The loss function is, therefore, the summation of error evaluated at all five groups of test points. The implementation of the cross validation is described with more details in Section 4.4.2.

4.4. Fitness Evaluation in Genetic Algorithm Assisted Knowledge-Based Surrogate Modeling

The fitness evaluation is a process of computing the approximation error $\text{Loss}(\boldsymbol{x}_{\text{fid}})$ for given function indicator $\boldsymbol{x}_{\text{fid}}$. The process is demonstrated in Figure 4.3. First, $\boldsymbol{x}_{\text{fid}}$ is transformed into function formats:

$$
f_l(\boldsymbol{p}(\boldsymbol{x}), \boldsymbol{x}_{\text{fid}}), \quad l = 1, 2, \cdots, m,
\tag{4.4}
$$
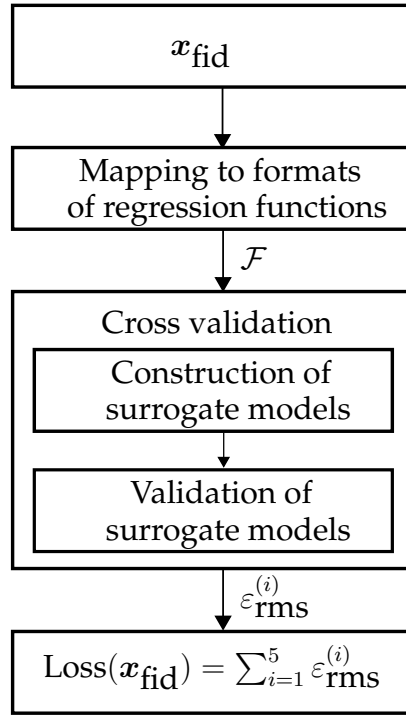
Figure 4.3.: The flowchart of the fitness evaluation in genetic algorithm assisted surrogate modeling.

and

$$\mathcal{F} = \sum_{l=1}^{m} \beta_l f_l(\boldsymbol{p}(\boldsymbol{x}), \boldsymbol{x}_{\text{fid}}), \tag{4.5}$$

where β_l are the regression coefficients, which are to be determined by Kriging modeling. Then those function formats are used as regression functions of Kriging model in knowledge-based surrogate modeling. This surrogate model is formulated as follows:

$$\hat{q}(y) = \mathcal{F}(\boldsymbol{p}(\boldsymbol{x})) + Z(\boldsymbol{p}(\boldsymbol{x})) = \overbrace{\sum_{l=1}^{m} \beta_l f_l(\boldsymbol{p}(\boldsymbol{x}), \boldsymbol{x}_{\text{fid}})}^{\text{regression}} + \underbrace{Z(\boldsymbol{p}(\boldsymbol{x}))}_{\text{correlation}}, \tag{4.6}$$

where $\boldsymbol{p}(\boldsymbol{x})$ is the vector of knowledge-based terms, which are already introduced in Chapter 3. The approximation error of the surrogate models are evaluated with a cross validation method to get the fitness of each individual in a generation. The best formats of regression functions $f_l(\boldsymbol{p}(\boldsymbol{x}), \boldsymbol{x}_{\text{fid}})$ are to be determined with GA, while the coefficients β_l are determined with Kriging model with sufficient number ($\geq m$: the number of coefficients) of sample points. With f_l and β_l determined, the optimal regression model $\mathcal{F}^*$ is obtained.

4.4.1. Mapping of Regression Functions

The mapping of x_{fid} to the formats of regression functions in Kriging is expressed as:

$$x_{\text{fid}} \quad \longrightarrow \quad f_l(p(x), x_{\text{fid}}), \quad l = 1, 2, \cdots, m, \tag{4.7}$$

where m is the number of basic functions in the regression model of Kriging. Each f_l is generated by mapping a subset of x_{fid} following the predefined rules. The number of coefficients β_l to be determined equals m, which means at least m sample points are required. The dimension of x_{fid} depends on the predefined rules of mapping. The more complicate the mapping rule is, the larger the dimension of x_{fid} will be. When the mapping rule is simple and the number of basic functions involved in the regression model is small, the sample size required for model construction can also be very small. In that case, although larger sample size is not necessary for the surrogate modeling, it is still helpful for GA to converge correctly, otherwise, GA might converge to a design that only fits with the given sample points, but has poor performance at other positions. As a result, the following two contradictory aspects are very important in defining the mapping rule:

- The mapping should be comprehensive enough to generated as many different formats of functions as possible. Therefore, large variety of function formats are available as candidates to be searched with GA, thus increasing the chance of getting the best surrogate models.

- The mapping should be rather compact that it does not involve too large dimension of x_{fid}. Therefore, the number of sample points can also be kept relatively small.

As introduced in Chapter 3, the knowledge-based terms are so generated that the their relationship with the system responses can be represented with polynomial terms and the inverse of polynomials. The orders of the polynomials are usually low, for example, linear, quadratic and cubic. The mapping rule is composed of three steps, which are listed below and described in the following three sections.

 (i) Generating subsets from the candidate set of knowledge-based terms,

 (ii) Unary operation on the terms of subsets,

(iii) Binary operation on the terms generated by unary operation.

4.4.1.1. Step 1: generating subsets from the candidate set

The first step is to generate a map between the function indicators x_{fid} and the knowledge-based terms $p(x)$. As is demonstrated in Figure 4.4, the candidate set $S(x)$ has $n+1$ elements, which include a constant value 1, and n knowledge-based terms (from $p_1(x)$ to $p_n(x)$). Three elements from the candidate set $S(x)$ will be selected based on the value of the first three function indicators $(x_{\text{fid}}(j), j = 1, 2, 3)$. The selected terms will be operated in the next two steps. The constant value 1 in the candidate set makes sure that lower-order polynomial functions are also possible to be generated, because in multiplication operation the constant

$$S = [1 \quad p_1(\boldsymbol{x}) \quad p_2(\boldsymbol{x}) \quad \cdots \quad p_n(\boldsymbol{x})]$$
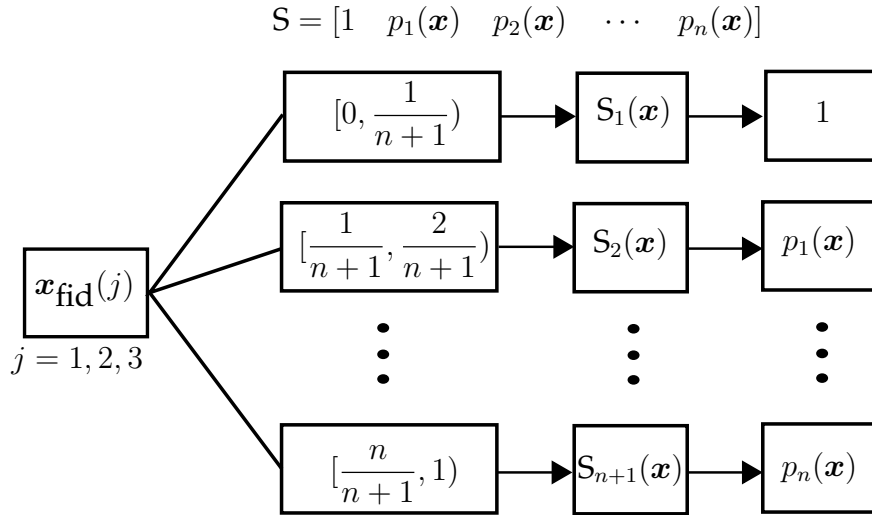
Figure 4.4.: Mapping between the function indicators $\boldsymbol{x}_{\text{fid}}$ and the knowledge-based terms $\boldsymbol{p}(\boldsymbol{x})$ for generating subsets from the candidate set of knowledge-based terms.

value does not increase the order of polynomials. Assume that each element in $S(\boldsymbol{x})$ should have equal chance to be selected, the design domain $[0,1)$ of $\boldsymbol{x}_{\text{fid}}$ is equally divided into $n+1$ segments. Each segment is corresponding to one element of the candidate set. According to the segment, which the value of $\boldsymbol{x}_{\text{fid}}(j)$ belongs to, a corresponding term in the set $S(\boldsymbol{x})$ will be selected. In this step, three elements from the vector $\boldsymbol{x}_{\text{fid}}$ are used. Each of them is responsible for selecting one element from the candidate set, thus three terms are selected as a subset to be operated later on. There is no restriction for repeated selection, which means it is possible that durings these three selections some elements are selected more than once, which is the case to generate quadratic and cubic terms in the next steps.

4.4.1.2. Step 2: unary operation
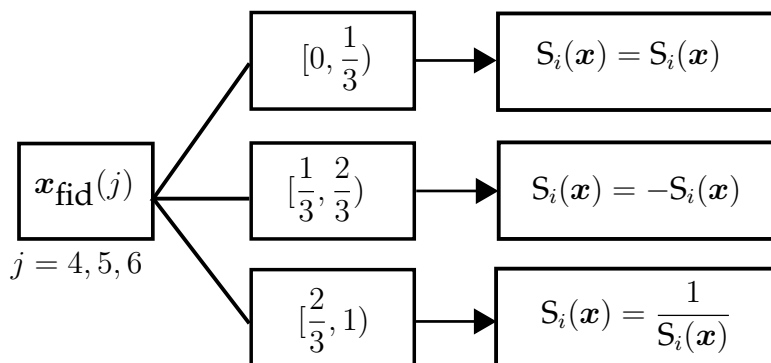
Figure 4.5.: Mapping between the function indicators $\boldsymbol{x}_{\text{fid}}$ and the unary operations, which are acted on the subsets generated from the previous step.

In this step, unary operation will be executed on each of the elements generated from

step 1. As can be seen in Figure 4.5, three different unary operations are possible. They are:

   (i) the equal assignment, $S_i(\boldsymbol{x}) = S_i(\boldsymbol{x})$,

   (ii) the opposite assignment, $S_i(\boldsymbol{x}) = -S_i(\boldsymbol{x})$,

   (iii) the inverse assignment, $S_i(\boldsymbol{x}) = \dfrac{1}{S_i(\boldsymbol{x})}$.

It is to be noticed that the inverse assignment is responsible for generating the inverse and the division operation of polynomial terms in the next step. In this step, the design domain $[0, 1)$ is equally divided into three segments. Each segment is mapped to one of the three unary operations. According to the value of $\boldsymbol{x}_{\mathrm{fid}}(j)$, a specific operation is selected to be executed on one term from step 1. Three $\boldsymbol{x}_{\mathrm{fid}}(j)$ are used in this step, so that all three terms from step 1 are unary-operated.

4.4.1.3. Step 3: binary operation

Two different binary operations are considered in this step, as is shown in Figure 4.6.They are:

   (i) weighted summation, $f_1 + c \cdot f_2$,

   (ii) production, $f_1 \cdot f_2$.
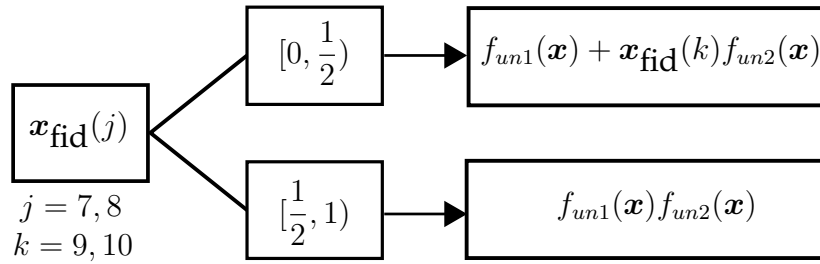
Figure 4.6.: Mapping between the function indicators $\boldsymbol{x}_{\mathrm{fid}}$ and the binary operations on the terms generated by unary operations.

The design domain $[0, 1)$ is equally divided into two segments, each represents one of the above listed operation. Based on the value of $\boldsymbol{x}_{\mathrm{fid}}(j)$, a specific operation is selected to be executed on two terms generated from step 2. In Figure 4.6, $f_{un1}(\boldsymbol{x})$ and $f_{un2}(\boldsymbol{x})$ represent any two terms obtained from the unary operation. These two terms are selected randomly and they can also be the same term from step 2. Quadratic and cubic polynomials, as well as the inverse of them are generated when the same terms are selected from step 1 and the multiplication operation is active in this step. Since there are three terms generated from step 2, two $\boldsymbol{x}_{\mathrm{fid}}(j)$ are used to generate the operations between all three terms. One $\boldsymbol{x}_{\mathrm{fid}}(j)$ is responsible for the binary operation of the first two terms, and the other $\boldsymbol{x}_{\mathrm{fid}}(j)$ is responsible for the binary operation of the resulted term of these two with the third term. In case the weighted summation operation is active, the weighting factor is the value of $\boldsymbol{x}_{\mathrm{fid}}(k)$. Two $\boldsymbol{x}_{\mathrm{fid}}(k)$ are involved for the operation between three terms from step 2. The function generated by the binary operation in step 3 is noted as $f_{bn}(\boldsymbol{x})$.

4.4.1.4. Resulted formats of functions

The possible formats of f_l generated by the whole procedure are listed but not limited to those in Equation 4.8.

$$
\begin{array}{ccccccc}
c_0 & p_i & p_i^2 & p_i^3 & \dfrac{1}{p_i} & \dfrac{1}{p_i^2} & \dfrac{1}{p_i^3} \\[3ex]
p_i p_j & p_i p_j^2 & p_i p_j p_k & p_i \pm c_1 p_j & (p_i \pm c_1 p_j)p_k \quad p_i \pm c_1 p_j^2 & & p_i \pm c_1 p_j p_k \\[3ex]
\dfrac{p_i}{p_j} & \dfrac{p_i}{p_j^2} & \dfrac{1}{p_i p_j} & \dfrac{1}{p_i p_j^2} & \dfrac{p_i p_j}{p_k} & \dfrac{p_i}{p_j p_k} & \dfrac{1}{p_i p_j p_k} \\[3ex]
p_i \pm \dfrac{c_1}{p_j} \quad p_i \pm \dfrac{c_1}{p_j^2} \quad \dfrac{1}{p_i} \pm \dfrac{c_1}{p_j} \quad \dfrac{1}{p_i} \pm \dfrac{c_1}{p_j p_k} & & \dfrac{p_i \pm c_1 p_j}{p_k} & & p_i \pm c_1 p_j \pm c_2 p_k & & \dfrac{1}{p_i} \pm c_1 \dfrac{1}{p_j} \pm c_2 \dfrac{1}{p_k}
\end{array}
$$

$$(4.8)$$

The complete procedure of generating regression functions for individuals of each generation in a genetic algorithm is demonstrated in Figure 4.7. As can be seen in Equation 4.8,
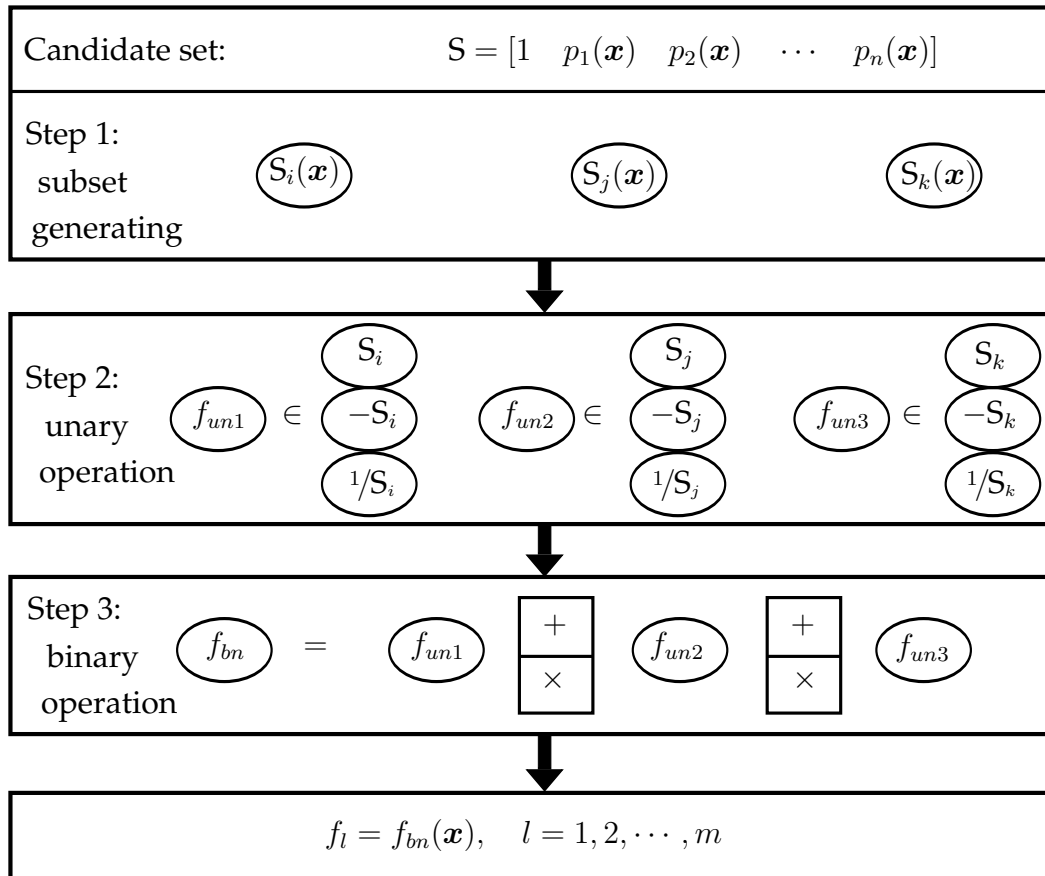
Figure 4.7.: Procedure of generating regression functions with individuals of each generation in a genetic algorithm.

according to the mapping rule the possible formats of functions have large variety. These in-

clude polynomial terms from zeroth order to the third order, as well as the inverse of those polynomials. Also covered in these formats are the mixed terms of polynomial functions and the combination of summation and production operations. Those formats are the typical elements to represent the relation between knowledge-based terms and the system responses of an engineering problem. Compared with traditional polynomial functions, they have larger variety and more flexibility, which allows GA to search for the best formulations of the regression models. In total, x_{fid} should have a dimension of ten to generate one f_l, which is a reasonably small dimension. It has to be pointed out that the dimension of x_{fid} increases as more basic functions f_l are required in the regression model for higher approximation quality. Therefore, it is suggested to start with small number of functions (small value of m), and then increase m gradually until the approximation quality is satisfactory.

4.4.2. Cross Validation

Once formulations of regression functions are generated, knowledge-based surrogate modeling can be performed. After that, it is necessary to validate the approximate performance with a few test points to check the quality of the surrogate models, so as to evaluate the goodness of the formats of regression functions. In order to avoid further computational expensive simulations, the quality of surrogate models are evaluated with the existing sample points. This is realized by a cross validation method, as shown in Figure 4.8. This validation method is based on prediction sum of squares (PRESS), pioneered by Allen (1971) and extended to regression diagnostics by Quan (1988). In this disseration it is implemented by adopting the K-fold cross-validation [Picard and Cook (1984)], where a typical setting of $K = 5$ is made.

First, the existing high-fidelity points from DoE are quasi-equally divided into five

Figure 4.8.: The procedure of cross validation.

groups. Then one group is taken out as test points, while the other four groups are used

as sample points for construction of surrogate models. The root mean square error of the approximation on the test points is noted as $\varepsilon_{\text{rms}}^{(i)}$, which is calculated as follows:
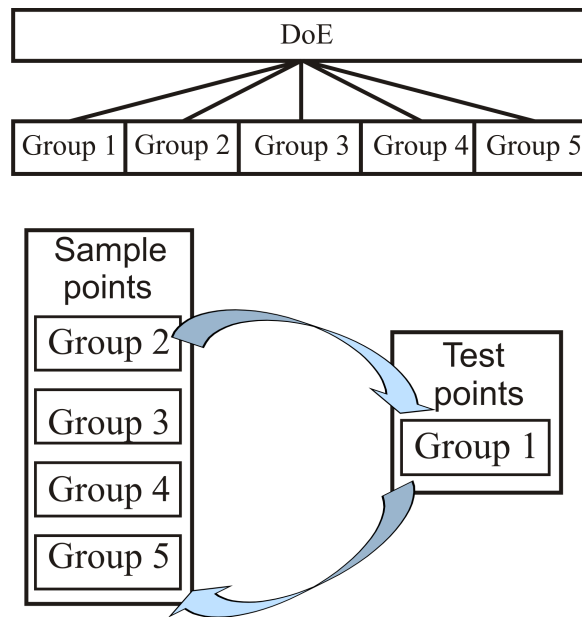
$$\varepsilon_{\text{rms}}^{(i)} = \sqrt{\frac{\sum_{j=1}^{n_t}(y_j - \hat{y}_j)^2}{n_t}}, \quad j = 1, 2, \cdots, 5 \tag{4.9}$$

where n_t is the number of test points in group 1. y and $\hat{y}$ are the high-fidelity value and the surrogate model value of the system responses, respectively. In order not to lose generality and that the approximation error is not affected by the choosing of test points from the groups, the root mean square error is calculated five times. Each time a different group is chosen as test points, the resulted approximation errors are noted as $\varepsilon_{\text{rms}}^{(i)}$, $i = 1, 2, 3, 4, 5$. The loss function is the summation of those approximation errors, formulated as:

$$\text{Loss}(\boldsymbol{x}_{\text{fid}}) = \sum_{i=1}^{5} \varepsilon_{\text{rms}}^{(i)}. \tag{4.10}$$

4.5. Genetic Algorithm Assisted Knowledge-Based Surrogate Modeling

The GAKBSM is a comprehensive searching and approximation technique. By integrating the genetic algorithm and knowledge-based surrogate modeling method, the best formulations for the regression models in Kriging is found. The mathematical tool of GA and physical knowledge are combined to improve approximation quality. For engineering problems, which are physically well understood, less effort is required in the searching process of GA. On the other hand, for engineering systems which are not (yet) well understood, the GA plays a great role in assisting the construction of surrogate models. The key procedures involved in GAKBSM are listed as follows.

(i) Studying, analyzing and decomposing the engineering problem to generate knowledge-based terms. Those terms are used to replace the original design space and form the knowedge-based surrogate modeling space.

(ii) Define the inputs for GA according to the mapping rule, which is designed to generate formats of regression functions.

(iii) Generate population for GA.

(iv) Perform fitness evaluation based on cross validation, which involves construction of knowledge-based kriging models and validation of approximation quality on test points.

(v) Search for the best formats of regression functions along the evolutionary process.

4.6. Demonstrative Example: Genetic Algorithm Assisted Knowledge-Based Surrogate Modeling for Fiber Angle Design of Stiffened Composite Panel

The implementation of GA assisted knowledge-based surrogate modeling is demonstrated with a fiber angle design problem. The structure is a stiffened composite panel from an aircraft wing. Large number of design variables are involved in this example, which can lead to the curse of dimensionality and great computational effort for surrogate modeling and design optimization. To cope with these problems, engineering knowledge in laminated composites is utilized in constructing the surrogate models. Genetic algorithm is used to search for the best regression models and optimize the quality of the approximation. The surrogate system responses of the composite panel can be further used to ease computational effort in engineering designs, such as fiber angle design optimization. Besides, in the application of stacking sequence optimization of composites, surrogate models have also shown to be very effective, as one can found in the work of Todoroki and Sekishiro (2008).

4.6.1. Introduction of the Stiffened Composite Panel

A piece of skin taken from the aircraft wing model developed in LLB is studied in this section. The geometric model of the wing is based on the commercial aircraft A320 of Airbus. The piece of skin from the wing is called a stiffened panel in this section. The stiffeners are stringers with T-section, as is shown in Figure 4.9. The whole wing has 26 ribs and is stiffened with 16 long stringers starting from the root of the wing and ending with eight short stringers at the tip. The panel is spanwisely taken between two adjacent ribs of the wing, and chordwisely it contains three stringers.

Composite materials are used to achieve large strength to weight ratio for the stiffened
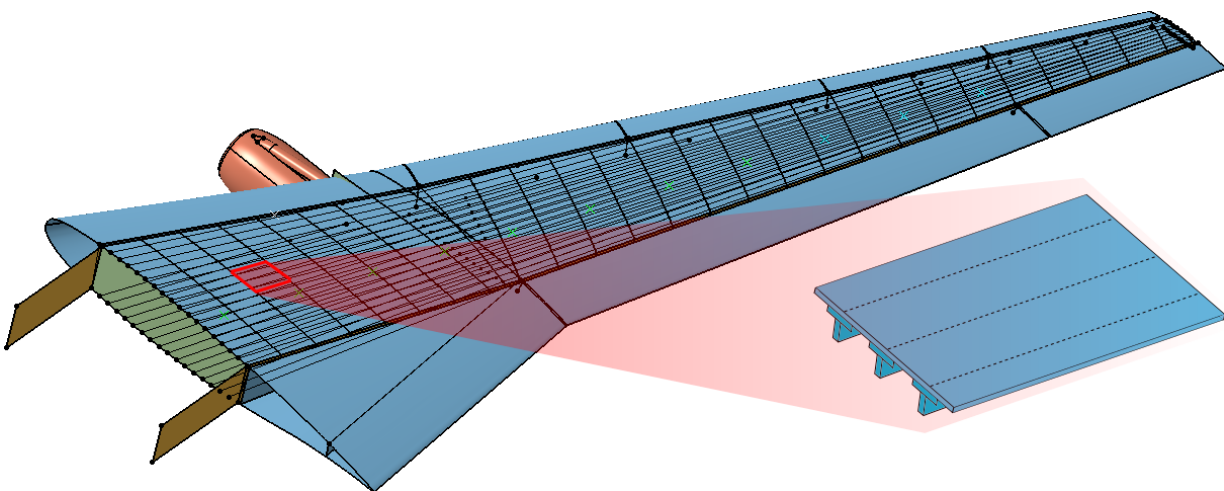
Figure 4.9.: A piece of aerocraft wing skin from the aircraft wing model developed in LLB.

panel. The skin, webs and flanges of the panel are all composed of laminated carbon fiber reinforced polymer (CFRP) as demonstrated in Figure 4.10.

Table 4.1.: Geometric parameters of the stiffened composite panel

Parameter name	Symbol	Property	Value	Unit
Lamina Orientation	α	Variable	$[-90, 90]$	degree
Lamina thickness	t_l	Constant	0.0931	mm
Skin thickness	t_s	Constant	$40 \times t_l$	mm
Flange thickness	t_f	Constant	$40 \times t_l$	mm
Web thickness	t_w	Constant	$60 \times t_l$	mm
Skin length	l	Constant	600	mm
Skin width	b	Constant	480	mm
Flange width	w	Constant	60	mm
Web height	h	Constant	40	mm

The skin and the flanges are composed of 40 plies, while the webs contain 60 plies. All
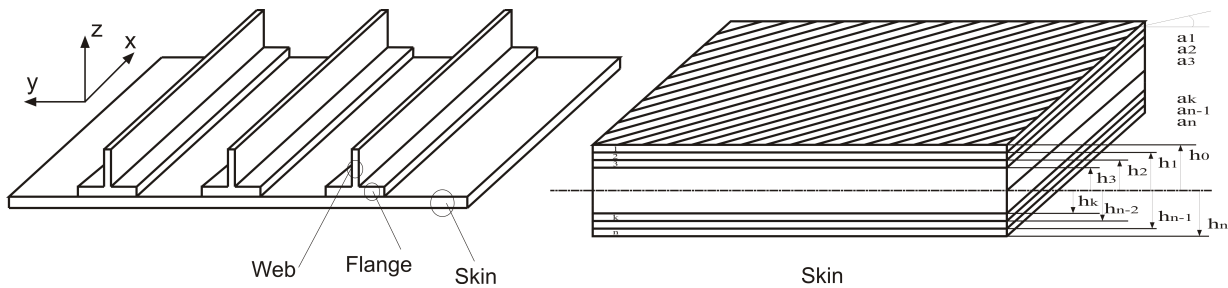
Figure 4.10.: Stiffened aircraft wing skin constructed with fiber reinforced laminates.

plies have the same thickness and the same material properties, only the fiber orientation of each ply is variable. All laminates are symmetrically laid up with respect to their middle planes, which means fiber orientations for only half of the plies are to be determined. The flanges are designed to have the same stacking sequences as the skin. As a result, in total 50 fiber angles (20 for the skin and 30 for the webs) are to be designed. The geometric parameters of the stiffened panel model is shown in Table 4.1.

4.6.2. Loads and Boundary Conditions

The main function of the aircraft skin is to encapsulate the components of wing and to hold the airfoil shape so that lift can be continuously provided. With the force of lift, the upper skin endures compression and the lower skin is under extension. The lift generated on the skin is transformed to the ribs and spars through their connections. Shear forces are produced along the edges where the skin and the other components connects. Therefore, the stiffened panel model, which is taken from the upper skin of the wing sustains compression forces and shear forces, shown in Figure 4.11.

The four edges of the panel are named counter clockwisely Edge1, Edge2, Edge3 and Edge4, respectively (see Figure 4.11). Edge3 is perpendicular to the wing span, and are acted
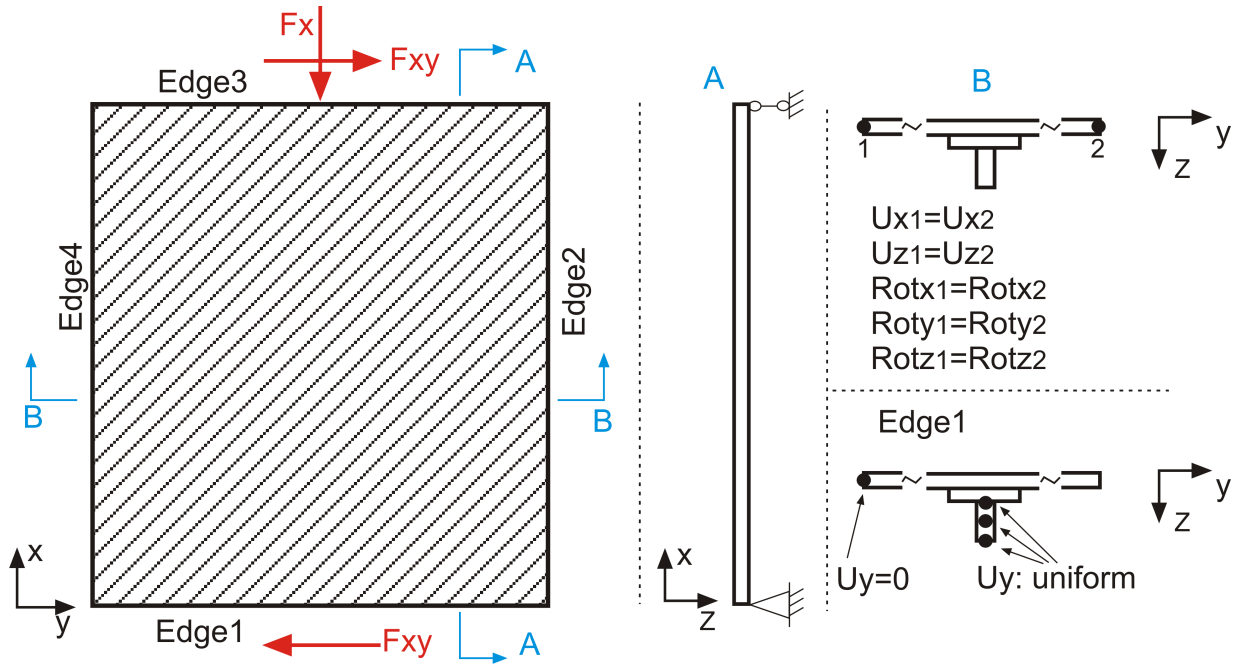
Figure 4.11.: Boundary conditions of the stiffened aircraft wing skin.

with the compression force F_x. Edge1 and Edge3 are loaded with the shear force F_{xy}. The boundary conditions are defined as follows:

- Edge1 and Edge3 are simply supported. As is demonstrated in Figure 4.11, section A is cut at any position of the model parallel to the xz plane. The displacements in x and z direction on Edge1 is not allowed, that is $U_x = 0$ and $U_z = 0$. On Edge3 it has uniform U_x and $U_z = 0$. Rotation about both edges are free.

- Edge2 and Edge4 are considered as symmetric edges with respect to the xz plane and are given the same constraints in all degree of freedom (DoF) except the displacement in y direction. This boundary condition is shown in section B of Figure 4.11, which is cut from any position on Edge2 and Edge4 parallel to the yz plane.

- There are three cross sections of the stringers at both Edge1 and Edge3. The nodes on the web of each section have uniform displacement in y direction. The U_y DoF of the leftmost node on Edge1 is fixed.

4.6.3. Failure Criterion

The panel should satisfy all strength, stiffness and stability constraints under the loads and boundary conditions defined in Section 4.6.2. The assumption of perfect bond between adjacent layers is made and failures such as laminate delamination [Reifsnider (1982)] are not considered. Interlaminar effects, such as in-situ stresses and the process of crack propagation [Camanho et al. (2006)] in the laminate are also not within the design consideration of this example. For the composite panel example used in this chapter and through the whole

Table 4.2.: Material properties and strength constants of the material for the fiber reforced composite panel.

Parameter name	Symbol	Value	Unit
Longitudinal elastic modulus	E_1	126	GPa
Transverse elastic modulus	E_2	11.1	GPa
Shear modulus (longitudinal to transverse)	G_{12}	6.6	GPa
Major Poisson ratio (longitudinal to transverse)	v_{12}	0.28	–
Allowable tensile stress (longitudinal)	$\bar{\sigma}_1^+$	1950	MPa
Allowable tensile stress (transverse)	$\bar{\sigma}_2^+$	48	MPa
Allowable compressive stress (longitudinal)	$\bar{\sigma}_1^-$	−1480	MPa
Allowable compressive stress (transverse)	$\bar{\sigma}_2^-$	−200	MPa
Allowable shear stress (longitudinal to transverse)	$\bar{\tau}_{12}$	79	MPa

dissertation, the first ply failure (FPF) criterion [Reddy and Pandey (1987) and Hinton et al. (2004)] is used as the strength constraint of the fiber angle design problem. According to FPF, the values of the failure criterion index are calculated for all the plies in all laminates of the structure, and the maximum one is used to decide whether the structure is failed or not. In this design problem, the Tsai-Hill failure criterion index [Tsai and Wu (1971) and Hill (1993)] is used to measure the stresses resulted in each ply with respect to the strength of the material. It is supposed that all plies are in plane-stress state, so that the Tsai-Hill failure criterion index FI is formulated as follows:

$$\text{FI} = \frac{\sigma_1^2}{\bar{\sigma}_1^2} + \frac{\sigma_2^2}{\bar{\sigma}_2^2} + \frac{\tau_{12}^2}{\bar{\tau}_{12}^2} - \frac{\sigma_1 \sigma_2}{\bar{\sigma}_1^2}, \tag{4.11}$$

where σ_1, σ_2 and τ_{12} refer to the in-plane longitudinal, transverse and shear stresses, respectively. Accordingly, $\bar{\sigma}_1$, $\bar{\sigma}_2$ and $\bar{\tau}_{12}$ refer to the in-plane longitudinal, transverse and shear strength constants of the lamina, respectively. Of the three strength constants, $\bar{\sigma}_1$ and $\bar{\sigma}_2$ is either the tensile strength or the compressive strength of the lamina depending on the direction of σ_1 and σ_2. That is when $\sigma_1 > 0$, $\bar{\sigma}_1$ takes the value of longitudinal tensile strength constant $\bar{\sigma}_1^+$ and when $\sigma_1 < 0$, $\bar{\sigma}_1 = \bar{\sigma}_1^-$. The same rule is defined for σ_2. The strength constraint is, therefore, given as:

$$\text{FI} < 1. \tag{4.12}$$

The lamina material used for the panel model is AS4 3501-6, the material properties and strength constants of which are given in Table 4.2 according to the databank of ESAComp [Palanterä and Mönicke (2013)].

Besides the strength requirement, stiffness and stability constraints are defined by the buckling load factor λ_b. λ_b is obtained after a buckling analysis on the finite element model. Physical meaning of λ_b is the ratio of the load that actives the first buckling mode of the structure to the applied loads [Jones (1998)]. To prevent the structure from buckling instability, the buckling load factor λ_b has to satisfy the following equation:

$$\lambda_b > 1. \tag{4.13}$$

4.6.4. Task Definition for Fiber Angle Design of Stiffened Composite Panel

Since large number of design variables are involved in the fiber angle design problem of the stiffened composite panel, computational efficient and sufficiently accurate surrogate models are expected to assist the design process. The task here is, therefore, to constructed surrogate models with as fewer sample points as possible to approximate the Tsai-Hill failure criterion index FI and the buckling load factor λ_b.

4.6.5. Parameter Study

Parameter studies of the stiffened panel model are performed. There are 50 design variables in this problem. The purpose of performing parameter studies in this section is to have a primary understanding of the system behavior, which helps us to decide what kind of surrogate modeling method to use. To achieve this, the variation of the system responses with respect to one and two design variables is taken as an example, namely one-dimensional and two-dimensional parameter studies, respectively. The one and two-dimensional parameter studies are easily visualized, and give us a clear view of the system behavior.

First, one-dimensional parameter study is shown. Here only one fiber angle is chang-

Figure 4.12.: One-dimensional parameter study of the laminate fiber angle design problem. The graphs show the variation of the system responses FI and λ_b with the first ply fiber angle of the skin α_{s1}.

ing, while all the other fiber angles are kept constant. The fiber angle of the first ply of the skin α_{s1} and the fiber angle of the first ply of the web α_{w1} are selected as the single variable parameter separately. In Figure 4.12 it is shown how the system responses FI and λ_b are varying with the α_{s1}, and in Figure 4.13 the system responses are varying with α_{w1}.

It can be seen the two system responses FI and λ_b have distinct behaviors and they vary in completely different forms with respect to α_{s1} and α_{w1}. From the trend of the curves it

Figure 4.13.: One-dimensional parameter study of the laminate fiber angle design problem. The graphs show the variation of the system responses FI and λ_b with the first ply fiber angle of the web α_{w1}.

is seen that the relationships between system responses and the design variables are neither linear or quadratic. This means the standard surrogate modeling methods will be impossible to approximate the system responses accurately.

Second, a two-dimensional parameter study is performed, in which both α_{s1} and α_{w1} are variable, while all the other fiber angles are set constant. The relationship of the system responses FI and λ_b with the two variable parameters are demonstrated in Figure 4.14. The two-dimensional parameter study shows the interactive effects of different design variables on the system responses.

It is shown in Figure 4.14 that α_{s1} and α_{w1} affect the system responses in different extents. The two system responses vary much more sharply with respect to α_{s1} than to α_{w1}. This is in accordance with the physical roles of the skin and the webs in the structure. The skin is the main component of the panel in sustaining and transferring loads, while the webs belong to the stiffeners, which enhance the structure at local regions. Therefore, it is again indicated that using standard surrogate modeling is not a wise choice in approximating the system responses of fiber composites. For example, in a full quadratic polynomial approximation model, all the design variables are formed in the same manner and have up to the second order. It requires a great number of sample points, $\frac{(m+1)(m+2)}{2} = 1326$ ($m = 50$ is the number of design variables in this case), to determine the coefficients for all terms involving all design variables. The required number of sample points is even larger considering that an oversampling factor is required to adjusting the regression error using least squares. Besides, it can be seen from the parameter studies in Figure 4.12 to Figure 4.14 that the relationships are incapable to be well approximated by quadratic curves. The parameter studies intuitively explain the difficulties of achieving high quality approximation with the standard surrogate modeling methods. Therefore, in order to achieve higher efficiency and accuracy in surrogate modeling for this composite panel problem, it is important to use engineering
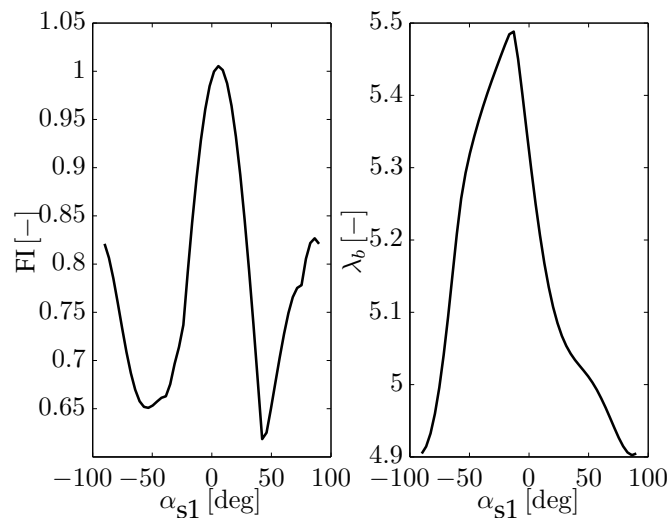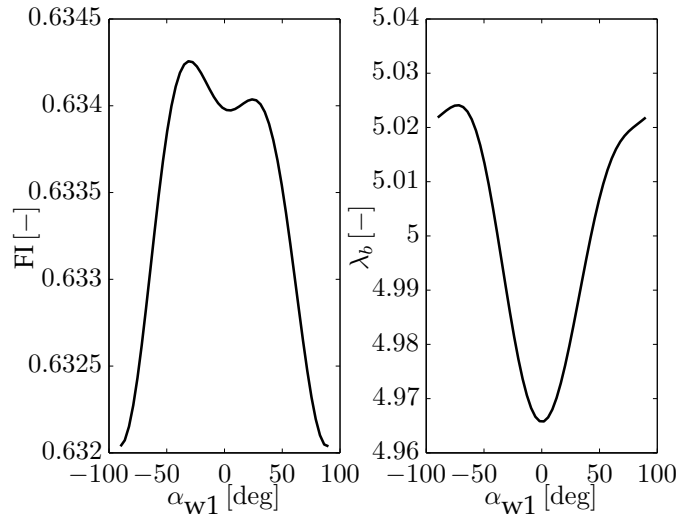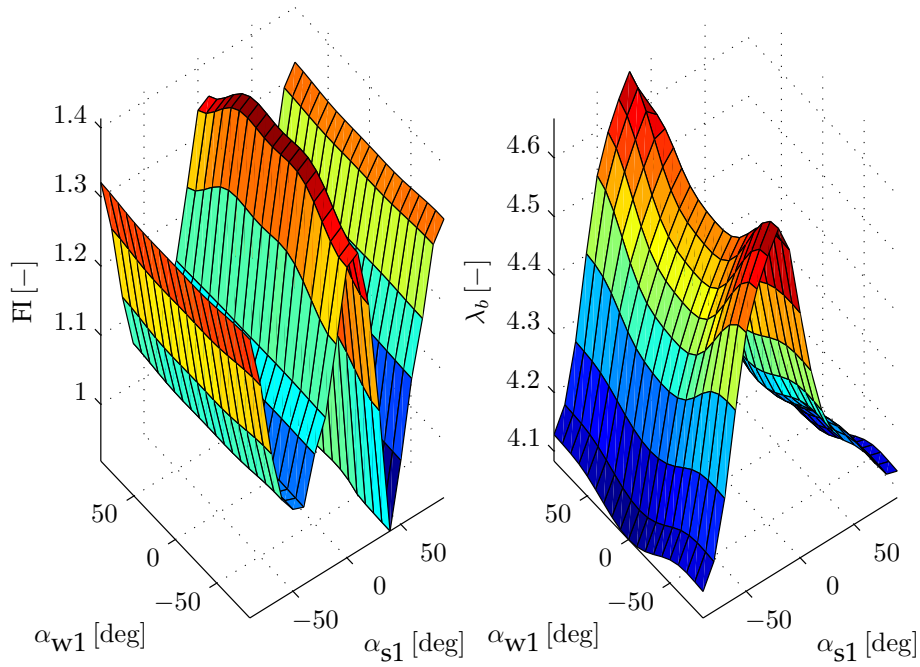
Figure 4.14.: Two-dimensional parameter study of the laminate fiber angle design problem. The graphs show the variation of the system responses FI and λ_b with the first ply fiber angle of the skin α_{s1} and the web α_{w1}.

knowledge of the structural system.

4.6.6. Standard Surrogate Modeling for the Stiffened Composite Panel

The standard Kriging model, which is composed of a second-order polynomial regression model and a Gaussian correlation model is used as an example in the fiber angle design problem. The number of sample points and the normalized root mean square error on a number of test points are used to quantify the efficiency and accuracy of the approximation.

As is calculated in the section above, 1326 sample points are required to fulfill the requirements of least squares. Considering a small oversampling factor, 1500 sample points are tried to construct the standard surrogate models for FI and λ_b with respect to the design variables $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_{50}]$. The results are shown in Figure 4.15, which shows the extent of consistency of the surrogate and the test system responses. The abscissa axis represents the value of system responses on the test points, and the verticle axis stands for both the test and the surrogate system responses. The test points all locate at the diagonal line of each graph, which are the benchmark of the comparison. The closer the surrogate system responses are with the benchmarks, the better the approximation quality is. As is seen, the surrogate responses are not in consistence with the test results. Actually, the disperse of the surrogate responses is quite large from the benchmark. The normalized root mean square error of the approximation for FI, and λ_b are 22.93% and 36.79%, respectively. This proves

Standard surrogate modeling with 1500 sample points
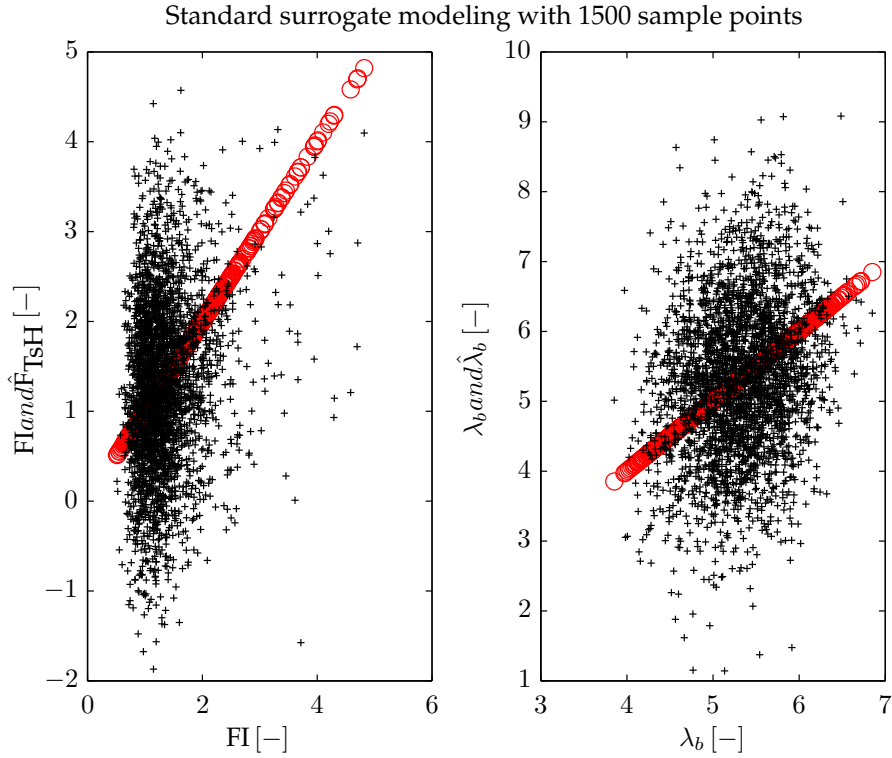
Figure 4.15.: The standard surrogate modeling for the stiffened composite panel: the consistency between the surrogate and the test system responses. Circle (o): test system responses; Cross (+): surrogate system responses.

the conclusion in the parameter studies, that the standard surrogate modeling method have difficulties to fit the system responses of the laminated panel, even with the large sample size 1500.

4.6.7. Genetic Algorithm Assisted Knowledge-Based Surrogate Modeling for the Stiffened Composite Panel

As discussed in Section 3.4, the dimensionality of surrogate modeling space can be reduced, and the relationship to be approximated can be simplified by adopting the structural engineering knowledge. The knowledge-based terms in this case are the elements of the laminate matrices: A^*, B^*, D^* and $\bar{Q}$. A^*, B^* and D^* are the compliance matrices calculated on the cross sections of the laminates in the global coordinate system. $\bar{Q}$ is the stiffness matrix of a lamina calculated in the global coordinate system. The modeling space for surrogate modeling is, therefore, a space of the knowledge-based terms instead of the original design variables $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_{50}]$. In this example, laminates are laid symmetrically, so that the extension and bending of the laminate are uncoupled, and the coupling matrices B^* are zero in all laminates, which is:

$$\boldsymbol{B}^* = 0. \tag{4.14}$$

As a result, the knowledge-based terms are elements of $\boldsymbol{A}^*$, $\boldsymbol{D}^*$ and $\bar{\boldsymbol{Q}}$, which are:

$$\boldsymbol{A}^* = \begin{bmatrix} A_{11}^* & A_{12}^* & A_{16}^* \\ & A_{22}^* & A_{26}^* \\ sym & & A_{66}^* \end{bmatrix}, \boldsymbol{D}^* = \begin{bmatrix} D_{11}^* & D_{12}^* & D_{16}^* \\ & D_{22}^* & D_{26}^* \\ sym & & D_{66}^* \end{bmatrix}, \bar{\boldsymbol{Q}} = \begin{bmatrix} \bar{Q}_{11} & \bar{Q}_{12} & \bar{Q}_{16} \\ & \bar{Q}_{22} & \bar{Q}_{26} \\ sym & & \bar{Q}_{66} \end{bmatrix}. \quad (4.15)$$

In order to formulate the knowledge-based terms for the whole panel, the structure is decomposed into a number of typical simple laminates. As is shown in Figure 4.16, a segment of the panel is geometrically separated into three components.

Cross sections of the three different components of this structure are chosen to represent
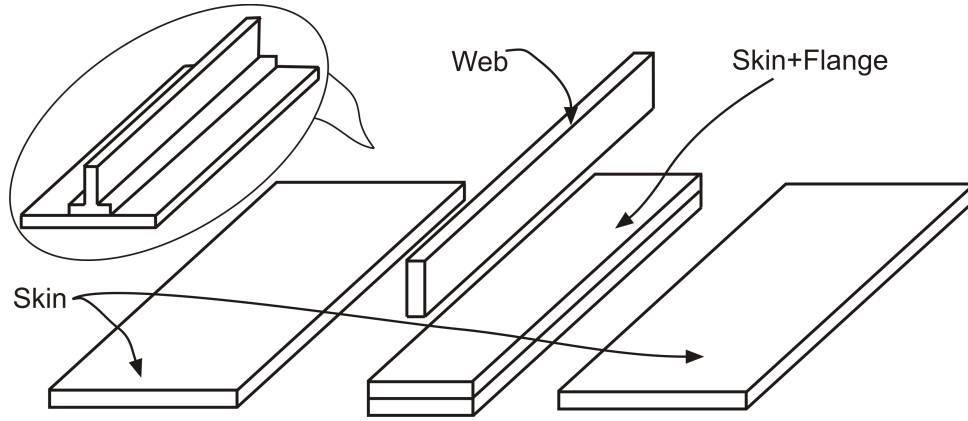
Figure 4.16.: The three components with different cross sections in a segment of the stiffened composite panel.

the system. Laminate matrices are calculated accordingly on those three sections, which are named as $Skin$, Web and $Skin + Flange$ sections, respectively. The components and the knowledge-based terms of these sections are described as follows:

(i) $Skin$: cross sections containing only the skin component. Such sections have 40 symmetrically laid plies, represented by 20 independent fiber angles $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_{20}]$. 12 independent matrix elements in $\boldsymbol{A}_s^*$ and $\boldsymbol{D}_s^*$ (6 in each matrix) and 6 independent matrix elements of $\bar{\boldsymbol{Q}}_s$ for each ply are to be considered as candidates in forming the knowledge-based modeling space.

(ii) Web: cross sections containing only the webs. Such sections have 60 symmetrically laid plies, represented by the rest 30 fiber angles $\boldsymbol{\alpha} = [\alpha_{21}, \alpha_{22}, \cdots, \alpha_{50}]$. The candidate terms in $\boldsymbol{A}_w^*$, $\boldsymbol{D}_w^*$ and $\bar{\boldsymbol{Q}}_w$, which are selected to form the knowledge-based modeling space are obtained the same way as that of the $Skin$ section.

(iii) $Skin + Flange$: cross sections containing both the skin and the flange. Such sections have 80 plies (40 plies in skin and 40 plies in flange), represented by the same angle as in $Skin$. In these sections, the $\bar{\boldsymbol{Q}}_f$ matrices have the same element as in that of $Skin$, and the $\boldsymbol{A}_f^*$ is a fraction of $\boldsymbol{A}_s^*$ in $Skin$. Therefore, only the 6 matrix elements of $\boldsymbol{D}_f^*$ are to be considered.

As is noted, the subscripts s, w and f in $\boldsymbol{A}_s^*$, $\boldsymbol{D}_s^*$, $\bar{\boldsymbol{Q}}_s$, $\boldsymbol{A}_w^*$, $\boldsymbol{D}_w^*$, $\bar{\boldsymbol{Q}}_w$ and $\boldsymbol{D}_f^*$ distinguish the laminate matrices ($\boldsymbol{A}^*$, $\boldsymbol{D}^*$ and $\bar{\boldsymbol{Q}}$, see Equation 4.15) for the three different sections. Among those selected matrices, all the $\boldsymbol{A}^*$ and $\boldsymbol{D}^*$ matrices are responsible for the global behavior of the laminates, which are related to the deflection and curvature in the middle ply of a laminate. The $\bar{\boldsymbol{Q}}$ matrices are related to the strength of each ply, and, therefore, they are only considered when a particular ply are of interest. For a particular system response, not all candidate terms are necessarily included in forming the modeling space. The relevant terms of the modeling space for the two system responses FI and λ_b are explained as follows:

- FI:
  FI is the maximum Tsai-Hill failure criterion index value among all plies in the structure. According to Equation 4.11, the FI is related to stress state of particular plies. Since the stress state is decided by both the global behavior and the strength of a ply $\boldsymbol{A}^*$, $\boldsymbol{D}^*$ and $\bar{\boldsymbol{Q}}$ are all responsible for FI. The number of candidate knowledge-based terms is 50.

- λ_b:
  λ_b is related to the buckling of the structure, which can be expressed by the system behavior at the middle plies of all the laminate components in the panel. Therefore, compliance matric elements in all the three sections are considered, which are elements in $\boldsymbol{A}_s^*$, $\boldsymbol{D}_s^*$, $\boldsymbol{A}_w^*$, $\boldsymbol{D}_w^*$ and $\boldsymbol{D}_f^*$. The total number of the selected terms is 30.

According to the classical laminate theory, the relationship of the system responses, such as stresses and displacements in the laminates with respect to the compliance matrices can be directly proportional to squares or even directly proportional. This means fewer coefficients in the regression model are to be determined in surrogate modeling. Therefore, the required number of sample points to approximate the system responses by embedding the above selected knowledge-based terms is much smaller than that required in a full quadratic model. It is expected that even if all linear and pure quadratic functions of the 50 knolwdge-based terms are used in the regression model, the number of coefficients to be determined is only 100, which is a great reduction in comparison with 1326 in the full quadratic polynomial function of $\boldsymbol{\alpha}$. However, the exact relation between the system responses and those candidate terms of the stiffened composite panel is not clear. As discussed earlier in this chapter, the best formats of the knowledge-based terms in the regression model for both system responses can be searched with GA.

The candidate formats of functions are generated according to the mapping rule introduced in Section 4.4.1, which include all function forms in Equation 4.8. Considering an oversampling factor of 1.5, 150 sample points are generated for the construction of the knowledge-based surrogate models. To perform the GA assisted surrogate modeling, the sample points are divided into five groups. Each group contains 30 high-fidelity points, four groups of which are used for model construction and the other one acts as test points for error calculation. Each group plays the role of test points in turn for cross validation to obtain the loss function (see Equation 4.10), which forms the fitness function in GA. By optimizing with GA, the best formats of the regression models for both system responses are obtained, which are listed as follows:

- FI:

$$
\begin{aligned}
\mathcal{F} = \ & \beta_0 + \beta_1(A^*_{s16} - A^*_{w22}) + \beta_2 A^*_{s22} + \beta_3(A^*_{s26} - 0.50A^*_{s66}) + \beta_4(D^*_{s11} + D^*_{w12} - D^*_{w66}) + \\
& \beta_5 D^*_{s16} + \beta_6(D^*_{s22} - 0.25D^*_{s26} + 0.25A^*_{w66}) + \beta_7 A^*_{w11} + \beta_8 A^*_{w26} + \\
& \beta_9(D^*_{w11} + 0.69D^*_{w16}) + \beta_{10} D^*_{f11} + \beta_{11} D^*_{f16} + \beta_{12}(D^*_{f22} - 0.29D^*_{f66}) + \\
& \beta_{13}[\textstyle\sum_{i=1}^{3}(Q^*_{s2i}A^*_{si1})\sum_{i=1}^{3}(Q^*_{s2i}D^*_{si2})] + \beta_{14}[\sum_{i=1}^{3}(Q^*_{s2i}A^*_{si2})\sum_{i=1}^{3}(Q^*_{s2i}A^*_{si6})],
\end{aligned}
$$

$$(4.16)$$

- λ_b:

$$
\begin{aligned}
\mathcal{F} = \ & \beta_0 + \beta_1(A^*_{s12} - 0.70D^*_{s16} + D^*_{s66}) + \beta_2(A^*_{s16} + 0.10A^*_{s66} + 0.40D^*_{s12}) + \beta_3 A^*_{s22} + \\
& \beta_4(A^*_{s26} + 0.29D^*_{s22} + 0.29D^*_{w66}) + \beta_5(A^*_{s66} - 0.28D^*_{s26}) + \beta_6 D^*_{s11} + \\
& \beta_7(A^*_{w16} + 0.47A^*_{w26}) + \beta_8(A^*_{w22} - 0.76D^*_{w11}) + \beta_9(D^*_{f16} + 0.60D^*_{f22} + 0.50D^*_{f26}).
\end{aligned}
$$

$$(4.17)$$

It is seen in Equation 4.16 and 4.17 that the regression models found by GA contains much fewer coefficients than the full quadratic models. For FI, the regression model has only 15 basic functions, which involves in total only 22 knowledge-based terms of those 50 candidate knowledge-based terms. As is expected, to approximate FI, both linear and quadratic functions of the candidate terms are requires. It is shown that the first 13 functions are linear, while both the last two functions contain the multiplication of two quadratic terms, which are obtained by utilizing the mapping rule in Section 4.4.1 twice. For λ_b, the regression model has only 10 basic functions formulated with the mapping rule. The regression model involves 21 knowledge-based terms of the 30 candidates, all of which are in linear forms. As is expected, it is sufficient to use only linear functions of the knowledge-based terms to approximate the λ_b. Besides, in both Equation 4.16 and 4.17 the laminate parameters of the skin occupy a larger proportion than those of the webs and flanges, which is also in accordance with the mechanical understanding of the stiffened panel.

Results of the GA assisted knowledge-based surrogate modeling are checked on the test points, which are shown in Figure 4.17. In comparison with the performance of the standard surrogate models shown in Figure 4.15, the accuracy of the approximation for both system responses are greatly improved by GAKBSM. It has to be noticed that the GAKBSM mothod requires only 10% of the sample points used in the standard surrogate modeling method. As can be seen in Figure 4.17, the surrogate responses are generally in good consistence with the test responses, except some offset on a few points for FI. Althogh those points have relatively large approximation error, they are far away from the feasible design domain (which requires FI $<$ 1), and so are the corresponding test points. Therefore, these points will be automatically eliminated in the engineering design optimization process, thus will not be a problem to the feasibility of the final design. With only 150 sample points, the normalized root mean square error of the approximation for FI, and λ_b are 4.71% and 2.20%, respectively. In comparison with those data in the standard surrogate models, these results show that the GAKBSM method has much better performance, this will be discussed in further in the following section.
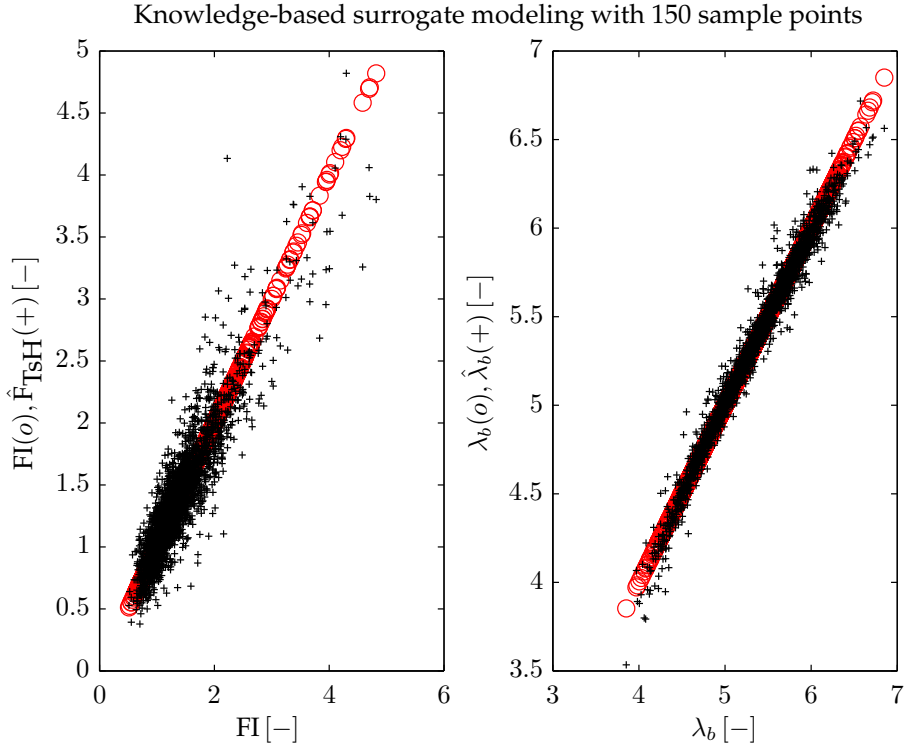
Figure 4.17.: The GA assisted knowledge-based surrogate modeling for the stiffened composite panel: the consistency between the surrogate and the test system responses. Circle (o): test system responses; Cross (+): surrogate system responses.

4.6.8. Comparison of Surrogate Models for the Stiffened Composite Panel

A comparison of the performance of the standard and the GAKBSM methods is shown in Table 4.3, where the number of sample points used in each method and the root mean square error of the approximation models for both system responses (FI and λ_b) are shown. The

Table 4.3.: Results of the standard and the GA assisted knowledge-based surrogate modeling methods for the stiffened composite panel.

Surrogate models	n_{samp}	$\varepsilon_{\text{rms}}(\text{FI})$	$\varepsilon_{\text{rms}}(\lambda_b)$
Standard	1500	22.93%	36.79%
Knowledge-based	150	4.71%	2.20%

standard surrogate models are constructed in 50-dimensional design space using the standard Kriging method, which has a quadratic regression model. The GA assisted knowledge-based surrogate modeling method is capable of reducing the dimension of the surrogate modeling space, and are thus able to constructe surrogate models for FI and λ_b in 22 and 21-dimensional modeling space, respectively. In the regression models of Kriging, the elements

of the compliance and stiffness matrices of the structure are used instead of the original design variables. As can be seen, the approximation quality is much better than the standard surrogate models, even with only 10% of the sample points. According to Equation 2.17, the speed-up ratio of GAKBSM is $\mathcal{P}_s = 9$. Actually, the speed-up ratio is even larger considering the approximation accuracy, because the standard surrogate modeling method requires further sample points to achieve the same quality as GAKBSM. In conclusion, GAKBSM is able to capture the system behavior with much higher efficiency and accuracy. The optimization behavior using the different surrogate models as system equations is discussed in detail in the next chapter.

4.7. Summary of Chapter

In this chapter, a GA assisted knowledge-based surrogate modeling method is designed. First of all, it utilizes the physical knowledge of the engineering problem to generate candidate terms as basic elements of regression functions. Then it uses the searching power of GA to improve the formats of the regression model during the evolutionary process. The GAKBSM is a comprehensive technique applying both mathematical and physical model to achieve better approximation quality than standard surrogate modeling. This method provides a more general solution for high-accuracy surrogate modeling. The performance of GAKBSM is demonstrated with a 50-design variable engineering problem. Much higher accuracy and efficiency is shown compared with the standard surrogate modeling method. However, to obtain good accuracy in design optimization, it is still necessary to improve the quality of the knowledge-based surrogate models locally, which is realized by adaptive surrogate-based design optimization in the next chapter.

5. Parallelized Adaptive Surrogate Based Design Optimization

Multiple-infill criteria is designed in this chapter for adaptive surrogate-based design optimization. Refinement of the quality of surrogate models is possible with the infill points, as more and more information appears after validating the optimization results. Since the evaluation of all sample points and all infill points are completely independent of each other, the adaptive surrogate-based design optimization process can be parallelized.

5.1. Motivation and Overview of the Chapter

After the computationally inexpensive surrogate models are constructed for the system responses, they can be used in design optimization to replace the computationally expensive system simulations. As is introduced in Section 2.3, this process is called surrogate-based design optimization (SBDO). The motivation of SBDO is to ease computational effort in optimization of large and complicate engineering design problems, which is however, at the cost of losing the accuracy at some degree in obtaining the system responses. Therefore, it is important to validate the goodness of surrogate models before and after the optimization process. It is also important to refine the surrogate models whenever it is possible or necessary to ensure the validity of the final optimization results. Although engineering knowledge and genetic algorithm assisted techniques are used to improve the performance of surrogate modeling, there are still problems in validating and refining the surrogate models, which are explained as follows.

- The number of sample points needed to ensure the accuracy of surrogate models are not known a prior. Insufficient sample points will generally result in poor surrogate models. While, redundant sample points will cost too much effort and are, therefore, not the solution to the problem.

- Generating test points by performing extra system simulations to validate the surrogate models is also not desirable. This is also because of the computational expense consideration.

- While in refining the surrogate models with additional system simulations, it is hard to decide which positions in the design space should be evaluated.

In this chapter, the above mentioned problems will be solved by parallelized adaptive surrogate based design optimization (PASBDO). The techniques in PASBDO are explained as follows:

- Adaptive surrogate based design optimization (ASBDO):
  ASBDO starts surrogate modeling with relatively small sample size and then utilizes infill criteria to assess the goodness of surrogate models and to locate additional sample points for refinement.

- Infill criterion:
  Infill criterion uses dimensionless quantity to represent the approximation error of surrogate models in the design space. By finding the maximum value of the infill criterion in the design space, the position where additional data is necessary can be identified. The additional reference data in ASBDO is called infill points.

- Multiple-infill criteria:
  The multiple-infill criteria is developed based on the infill criterion. It uses different infill criteria in different phases of the ASBDO process. It is possible to identify multiple locations of infill points according to the requirments in local or global accuracy of the surrogate models for optimization.

- Parallelized computing:
  PASBDO is the implementation of ASBDO using parallelization techniques, which parallelizes the sampling process and the infilling process.

The discussion of different infill criteria is first made on a one-variable example. A complete ASBDO process contains surrogate modeling before optimization, surrogate-based design optimization and refinement of surrogate models after optimization. Accordingly, there are infill criteria used before and after the surrogate-based design optimization. The infill criterion used before optimization improves the quality of surrogate models globally in the design space. The infill criteria used after surrogate-based design optimization is responsible for the improvement of optimization results in each loop of PASBDO. The procedure and the performance of PASBDO will be demonstrated with fiber angle design optimization of the stiffened composite panel.

5.2. Estimated Root Mean Square Error of Kriging Used as Infill Criterion Before Optimization

5.2.1. Introduction of Infill Criterion

The quality of a surrogate model is unknown unless it is assessed according to a reasonable criterion. Root mean square error [Mood et al. (1913)] is the most frequently used criterion for measuring of error related to multiple points. A typical formulation of root mean square error is given by

$$\varepsilon = \sqrt{\frac{1}{n_t} \sum_{i=1}^{n_t} [\hat{y}^{(i)} - y^{(i)}]^2}, \tag{5.1}$$

where n_t is the number of test points. Restricted by the computational cost, assessment based on a large number of test points generated by expensive system simulations is not practical. As a result, estimated root mean square error (RMS) is used. Unlike ε, the RMS represents the property of a stochastic process and, therefore, is not depend on the test points. The definition of estimated RMS is given by

$$\mathrm{RMS} = \sqrt{\mathrm{E}\left[(\hat{y} - y)^2\right]}. \tag{5.2}$$

Since the surrogate modeling method used in this dissertation is based on Kriging, which contains a stochastic process, the use of RMS as estimation error is possible. According to the introduction of Kriging in Section 2.2.3, the estimated root mean square error at any prediction site x is given as follows

$$\mathrm{RMS}(\boldsymbol{x}) = \sqrt{\sigma^2\left[1 + c(\boldsymbol{x})V^{-1}c(\boldsymbol{x})^T - r(\boldsymbol{x})R^{-1}r(\boldsymbol{x})^T\right]}. \tag{5.3}$$

In Equation 5.3, σ^2 is the maximum likelihood estimate of the variance of the Gaussian process (see Equation 2.10) of Kriging. σ^2 is calculated as follows:

$$\sigma^2 = \frac{1}{N}(Y - F\hat{\boldsymbol{\beta}})^T(Y - F\hat{\boldsymbol{\beta}}). \tag{5.4}$$

$r(\boldsymbol{x})$ is a vector of correlation functions evaluated at the prediction site $\boldsymbol{x}$ and the sample sites $X = \left\{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}, \cdots, \boldsymbol{x}^{(N)}\right\}^T$. This is given by

$$r(\boldsymbol{x}) = \left[\mathcal{R}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{x}, \boldsymbol{x}^{(1)}\right), \mathcal{R}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{x}, \boldsymbol{x}^{(2)}\right), \cdots, \mathcal{R}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{x}, \boldsymbol{x}^{(N)}\right)\right]. \tag{5.5}$$

$c(\boldsymbol{x})$ is a vector evaluated by

$$c(\boldsymbol{x}) = F^T R^{-1} r(\boldsymbol{x}) - f(\boldsymbol{x}), \tag{5.6}$$

where,

$$f(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_p(\boldsymbol{x})], \tag{5.7}$$

which are the regression functions in the regression model of Kriging evaluated at the prediction site $\boldsymbol{x}$. V is a matrix evaluated at all sample sites, given by

$$V = F^T R^{-1} F. \tag{5.8}$$

All symbols in Equation 5.2 to 5.8 are defined the same as in the introduction of Kriging in Section 2.2.3. A flowchart in Figure 5.1 demonstrates how the RMS infill criterion works in refining surrogate models, which is described as follows:

(i) Sample points are generated and initial surrogate models are constructed.

(ii) Position $\boldsymbol{x}^*$ of the maximum RMS in the design space is searched by GA.

(iii) The system responses at $\boldsymbol{x}^*$ are evaluated and used as infill points to refine current surrogate models.

(iv) The above step is repeated until no more significant improvement can be achieved according to a converging tolerance. The total number of infill points and the maximum iteration number are used as alternative stopping criteria.
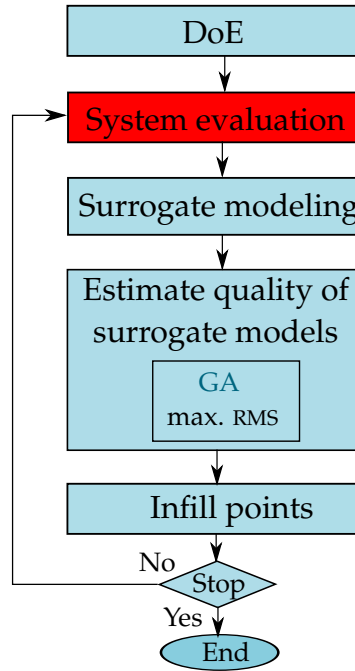
Figure 5.1.: The flowchart of using RMS infill criterion to refine surrogate models.

5.2.2. Performance of the RMS Infill Criterion on a Test Example

A one-variable function is used to investigate the performance of the estimated root mean square error as an infill criterion in refining surrogate models. The one-variable test function is defined in Forrester et al. (2008), which is given as:

$$z(x) = (6x - 2)^2 \sin(12x - 4), \quad x \in [0, 1]. \tag{5.9}$$

This is a typical test function for surrogate modeling and optimization algorithms. By solving the equation with the derivative of $z(x)$ equals zero [Widder (1947)], which is $z'(x) = 0$, it is found that this test function has four stationary points. As is shown in Figure 5.2, those stationary points are a local minimum, a saddle point, a local maximum and a global minimum, respectively. The first three of them have function values very close with each other. The global optimum is easily ignored by poor distribution of sample points or by improperly chosen surrogate models. Therefore, this one-variable function is also a simple but critical function for testing surrogate modeling and optimization methods. In case of surrogate-based design optimization, it usually requires to validate and refine the surrogate models.

In this section, the RMS as an infill criterion for surrogate modeling is tested first. The refinement of surrogate model for the one-variable function is performed iteratively, which is described as follows:

(i) An initial surrogate model is constructed with three Latin hypercube sample points using Kriging.

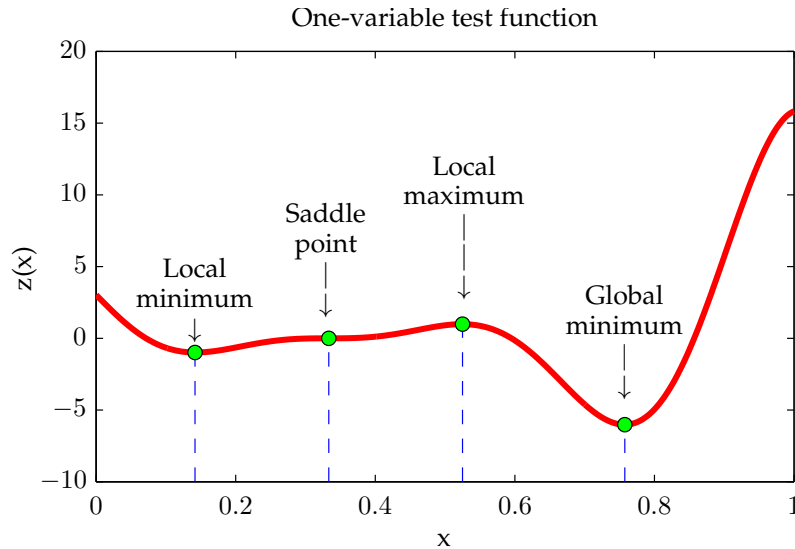(ii) Genetic algorithm is used to find the location of maximum RMS(x).

Figure 5.2.: The one-variable test function has a local minimum, a saddle point, a local maximum and a global minimum.

(iii) Evaluate infill point at the obtained location.

(iv) The infill point is added to the sample points to update the surrogate model. Then it goes to (ii) and continues until the stopping criterion is active. The stopping criterion is as mentioned above that the quality of the surrogate model converges or the maximal number of iterations is reached.

To assess the performance of the infill criteria critically, a group of extremely unlucky initial sample points are used all through this chapter. As can be seen in the first graph of Figure 5.3, the initial sample points completely fail to cover the global-optimum region. The consequence of the poor initial sample points is a poor surrogate model. In **Iteration 1**, the initial surrogate model is shown as a quadratic curve passing through all three sample points, which, however, fails to catch the characteristics of the test function. The refinement of the surrogate model with RMS as infill criterion is carried out iteratively. In each iteration, only one infill point is identified and added to the sample points. The whole process is illustrated iteration by iteration in Figure 5.3, which includes in total eight iterations. It can be seen in the 8th iteration that the refined surrogate model fits well with the true model. The total number of evaluations of the true model is ten, which includes three initial sample points and seven infill points. The remarks for the performance of RMS are listed as follows:

- Although RMS has the advantage that no test points are generated, its ability to identify infill points depends on the quality of the current surrogate models. The information from poor surrogate models can be misleading, and, therefore, it usually takes several iterations to find valuable infill points.

- The magnitude of RMS is not necessarily related to the quality of the surrogate model. As is seen from **Iteration 1** in Figure 5.3, RMS(x) of the surrogate model built with three sample points has a magnitude of around 10^{-15}. The inveracious RMS value is
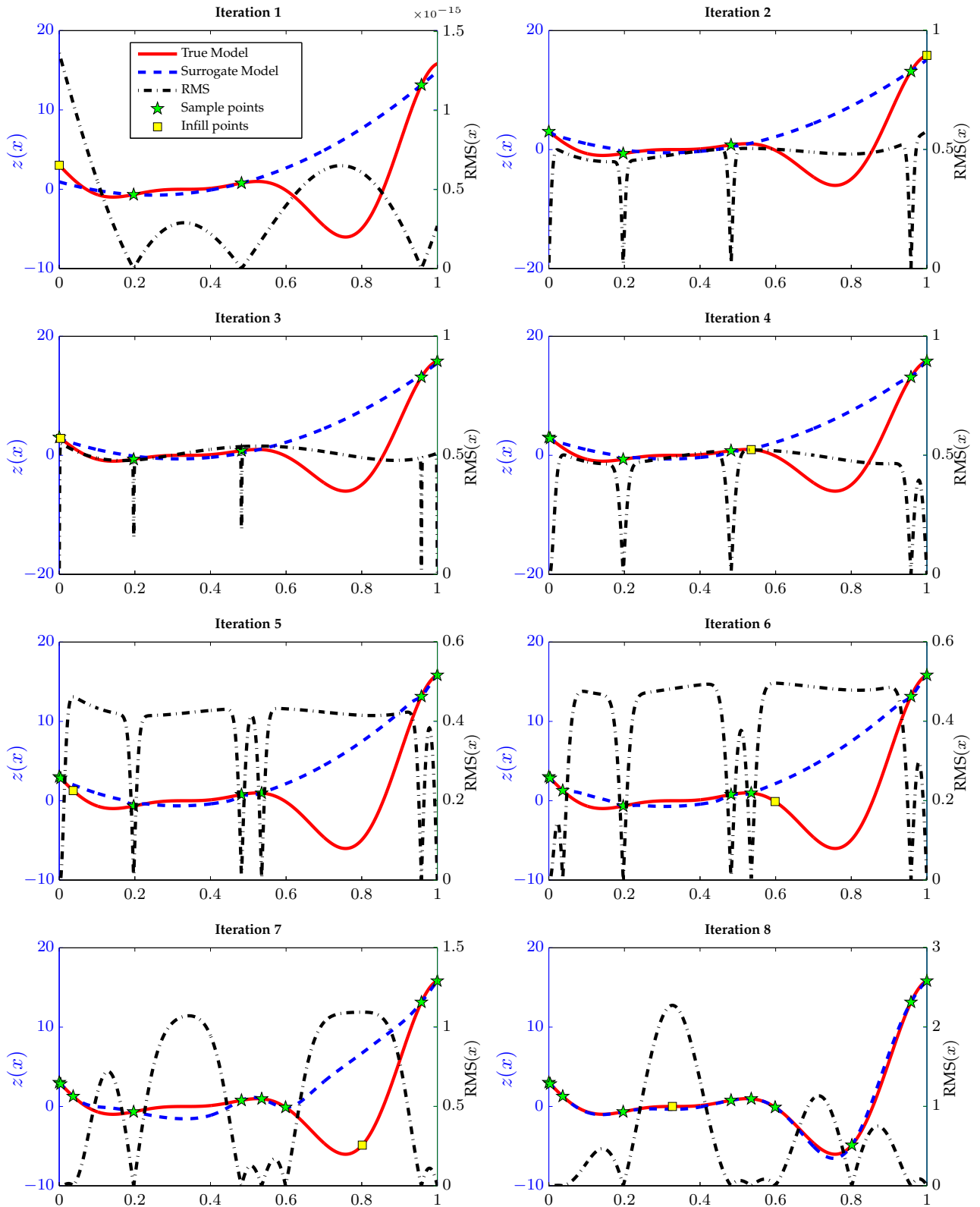
Figure 5.3.: The first eight iterations in refining surrogate model of the one-variable function with estimated root mean square error.

a result of the less of information, when the current regression model in Kriging can fit well with all current sample points. The surrogate model seems to be a perfect approximation of the true model, although it is actually quite inaccurate. Therefore, the magnitude of $RMS(x)$ can not be used as a stopping criterion.

- As a result of the misleading information from too small sample size, it can be seen in **Iteration 1** to **Iteration 3** that the locations of maximum $RMS(x)$ are at the border of the design domain. The $RMS(x)$ function is multimodal, which has the following properties:

    - $RMS(x)$ decreases sharply at the locations of sample points, where information of the true model is provided.

    - The infill points found by the genetic algorithm are usually the global maximum of $RMS(x)$, which tend to appear at the border of the design domain.

    - The local maxima of $RMS(x)$ are usually located in the regions between two sample points. The $RMS(x)$ is usually relatively larger between the two adjacent sample points, which has larger distances than those of the others. The local maxima are good candidates of infill points, as can be seen in all iterations in Figure 5.3.

- As is displayed in **Iteration 5** to **Iteration 8**, after adding several infill points, the RMS criterion is able to assess the quality of surrogate models pretty well. From the fifth iteration on, RMS can correctly locate infill points to improve the surrogate models greatly.

- The RMS infill criterion does not differ between the local and global optima of the test function. As can be seen from **Iteration 5** to **Iteration 8**, the location of infill point switches between local and global optimum of the true model, showing no preference.

According to the remarks mentioned above, the RMS criterion is used as an infill criterion to refine surrogate models before they are used as system equations in optimization. Since it is not known a prior whether the sample size in the initial surrogate models is large enough to prevent misleading infill points, it is better to find locations of both the global and local maxima of the $RMS(x)$ function. Then those locations are evaluated and used as infill points all at once to refine the initial surrogate models. This allows us to start with relatively small sample size in initial surrogate modeling. During the design optimization process, the surrogate models also need to be refined to ensure the accuracy of the optimal results. In the optimization process, infill criterion which is able to distinguish regions of local and global optima is preferred. Therefore, the expected improvement as an infill criterion is required, which is discussed in the next section.

5.3. Expected Improvement of Kriging Used as Infill Criterion in Surrogate-Based Design Optimization

5.3.1. Introduction of the Infill Criterion

As the name indicates, expected improvement (EI), which is introduced in Forrester et al. (2008) is the amount of improvement to the current optimum that is expected by adding infill points. Unlike the RMS, EI does not only quantify the quality of surrogate models by the estimated error, but also scales the error to generate preference in global optimum. The preference is created by taking in the distances between the estimated objective function value and the current optimal. Therefore, the magnitude of the improvement to the current optima is considered in the infill criterion. The location which has the largest EI value will be identified as infill point. The global optimum is thus preferred because it can provide larger improvement. By this means, regions containing local and global optimums are differentiated and, as a result, infill points can be more effectively identified. Using EI infill criterion, it is more likely to obtain the global optimum or a good local optimum. The expected improvement is formulated as follows:

$$
\text{EI}(\boldsymbol{x}) = \begin{cases} [z_{\min} - \hat{z}(\boldsymbol{x})] \, \Phi\left(\dfrac{z_{\min} - \hat{z}(\boldsymbol{x})}{\text{RMS}(\boldsymbol{x})}\right) + \text{RMS}(\boldsymbol{x})\varphi\left(\dfrac{z_{\min} - \hat{z}(\boldsymbol{x})}{\text{RMS}(\boldsymbol{x})}\right), & \text{RMS}(\boldsymbol{x}) > 0 \\[4ex] 0, & \text{RMS}(\boldsymbol{x}) = 0 \end{cases}
$$

(5.10)

where $z_{\min}$ is the current optimal objective value. $\hat{z}(\boldsymbol{x})$ is the estimated value of objective function at any prediction site $\boldsymbol{x}$. $\Phi(x)$ and $\varphi(x)$ are the cumulative distribution function (CDF) and probability density function (PDF) [Kolmogorov (1950)], respectively. $\Phi(x)$ and $\varphi(x)$ in Equation 5.10 are given as follows:

$$
\Phi\left(\frac{z_{\min} - \hat{z}(\boldsymbol{x})}{\text{RMS}(\boldsymbol{x})}\right) = \frac{1}{2}\left[1 + \text{erf}(\frac{z_{\min} - \hat{z}(\boldsymbol{x})}{\sqrt{2}\text{RMS}(\boldsymbol{x})})\right],
$$

(5.11)

where erf represents the Gauss error function,

$$
\varphi\left(\frac{z_{\min} - \hat{z}(\boldsymbol{x})}{\text{RMS}(\boldsymbol{x})}\right) = \frac{1}{\sqrt{2\pi}}e^{\dfrac{-(z_{\min} - \hat{z}(\boldsymbol{x}))^2}{2\text{RMS}^2(\boldsymbol{x})}}.
$$

(5.12)

In EI criterion,

- $z_{\min} - \hat{z}(\boldsymbol{x})$ represents the amount of possible improvement in objective function at location $\boldsymbol{x}$. It is responsible for identifying locations with better objective function value than current optimum.

- RMS$(\boldsymbol{x})$, as introduced in Setion 5.2, is responsible to identifying the unexplored regions.

As a result, EI criterion gives precedence to locations with low objective function value or high approximation uncertainty. Infill points generated with EI criterion are capable of balancing model exploration and optimization efficiency. Figure 5.4 demonstrates the process of using EI infill criterion in ASBDO for refining surrogate models and assisting the searching of optimum. It can be seen that the infill points are identified after surrogate-based optimization, which allows information to be gathered and to assist the identification of infill points.
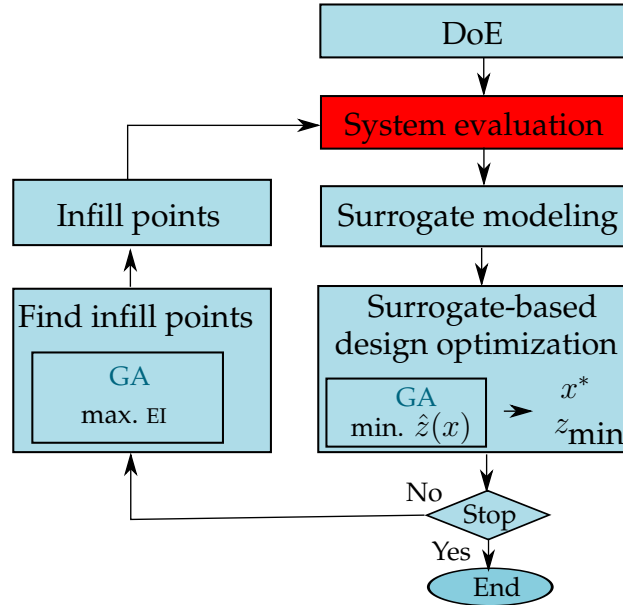
Figure 5.4.: The flowchart of using EI infill criterion to refine surrogate models in ASBDO.

5.3.2. Performance of the EI Infill Criterion on a Test Example

The process of adaptive surrogate-based design optimization with EI used as infill criterion is also demonstrated with the one-variable function. The problem is to find the global minimum of the test function, which is formulated as follows:

$$\text{minimize} \quad z(x) \ = (6x - 2)^2 \sin(12x - 4), \quad x \in [0, 1]. \tag{5.13}$$

The problem is solved by adaptive surrogate-based optimization, using the EI infill criterion. The EI infill process in ASBDO takes the following steps:

(i) An initial surrogate model is constructed using Kriging with three Latin hypercube sample points. The sample points are the same as those in the RMS.

(ii) Perform surrogate-based design optimization and get current best z_{min}. Validate the design, when an optimum is converged, stop, otherwise, continue.

(iii) Genetic algorithm is used to find the location of maximum $\text{EI}(x)$.

(iv) Evaluate infill point at the obtained location.

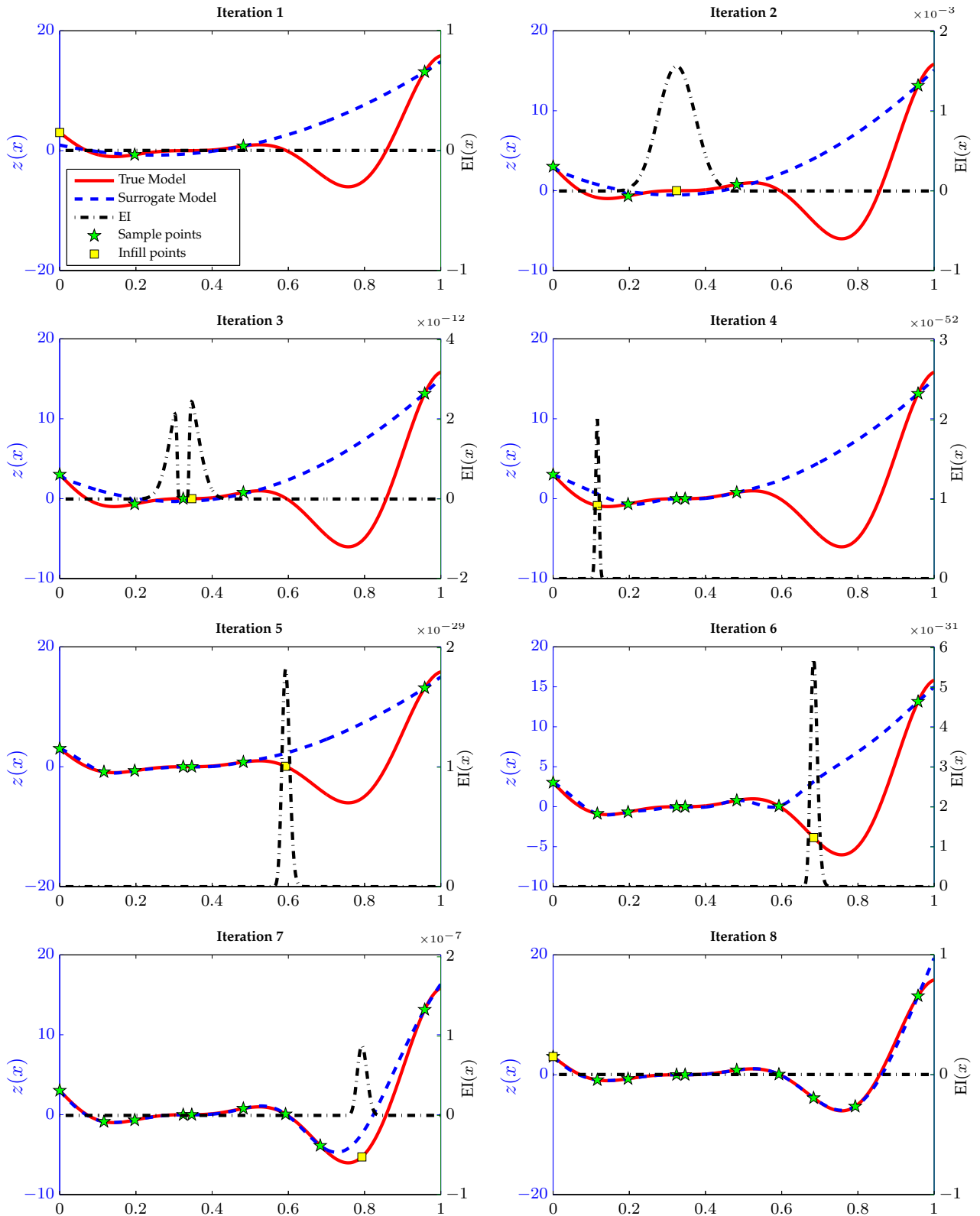Figure 5.5.: The first eight iterations in adaptive surrogate-based optimization of the one-variable test problem. Expected improvement (EI) is used as infill criterion.

(v) The infill point is added to update surrogate models. Then go to (ii).

The infill process in ASBDO of the one-variable test example is shown in Figure 5.5, which also has eight iterations. As is demonstrated, in the last iteration, the surrogate model can represent the true model well. Actually, in the 7th iteration, the optimum and the infill point are already close to the true optimum. With some tolerance the minimization of the one-variable function problem can be completed with all together nine function evaluations: three initial sample points and six infill points. Remarks for the expected improvement as infill criterion in ASBDO according to the results on the test problem are listed as follows.

- The magnitude of EI is also not necessarily relevant with the quality of the surrogate model. As is seen in **Iteration 1** and **Iteration 8** of Figure 5.5, $EI(x) = 0$ in both iterations while not both surrogate models are good.

- The main difference between EI and RMS is: $RMS(x)$ contains much more peaks than EI. Which is to say that although EI can also be multimodal, it tends to be concentrated only on the specific locations and makes those locations extremely distinctive. Such behavior gives the EI the advantage that important infill points can be identified more efficiently.

- The global optimal region starts to be identified by EI when either it is already covered in initial surrogate models or after the local optimal regions are explored. As is shown in Figure 5.3, with the extremely unlucky initial sample points, the global optimal region is totally unexplored. As a result, from **Iteration 2** to **Iteration 5** only the local optimal region is refined. Therefore, it is to be noted that a good initial surrogate model is very important for the identification ability of the EI infill criterion.

According to the above characteristics of EI and the RMS discussed earlier, it can be concluded that the two infill criteria are complementary. The using of RMS on the initial surrogate model in order to explore the system globally will generate a good starting surrogate model for EI. EI will then be able to zoom fast into the interested regions for design optimization.

5.4. Constrained Expected Improvement of Kriging Used as Infill Criterion in Surrogate-Based Design Optimization

5.4.1. Introduction of Infill Criterion

Most if not all engineering design problems are constrained optimization problems, which has to satisfy a number of requirements. For surrogate modeling of constrained optimization, it is more important to acquire good accuracy of the feasible domain. Therefore, infill points are better located in the feasible domain or near the boundary of the feasible area. For the Gaussian process in Kriging, the possibility of feasibility $PF(x)$ can be estimated. The

constrained expected improvement CEI, which finds the intersection of $\text{EI}(\boldsymbol{x})$ and $\text{PF}(\boldsymbol{x})$ is used. The CEI in this dissertation is altered from the definition of Forrester et al. (2008) in order to adapt with engineering problems. This is defined as:

$$\text{CEI}(\boldsymbol{x}) = \text{EI}(\boldsymbol{x})\text{PF}(\boldsymbol{x}), \tag{5.14}$$

where PF estimates the possibility of feasibility of any location in the design domain, which is given as:

$$\text{PF}(\boldsymbol{x}) = \frac{1}{2}\left[1 + \text{erf}(\frac{-\left\|g_{\text{lim}} - \hat{g}(\boldsymbol{x})\right\|}{\sqrt{2}\text{RMS}(\boldsymbol{x})})\right], \tag{5.15}$$

where g_{lim} is the boundary value of the feasible domain of the constraint function $g(\boldsymbol{x})$. According to the standard definition of a constraint optimization problem, it is required that $g(\boldsymbol{x}) \leq g_{\text{lim}}$. $\hat{g}(\boldsymbol{x})$ is the surrogate model of the constraint function. Therefore, the term $\left\|g_{\text{lim}} - \hat{g}(\boldsymbol{x})\right\|$ in Equation 5.15 is the absolute distance of the surrogate constraint function value and the critical constraint value. The smaller the distance is, the closer the current surrogate constraint value is to the boundary of constraints.

The intension of the absolute-distance treatment is to concentrate on both feasibility and activation of constraints. The former is mathematically expressed as $g(\boldsymbol{x}) - g_{\text{lim}} < 0$. The later is $\left\|g(\boldsymbol{x}) - g_{\text{lim}}\right\| < \xi$, where ξ is a predefined small value, which is very close to zero in accordance with the tolerancce in the engineering problem. In Equation 5.15 the absolute distance is used so that the designs near both sides of the constrain boundary are acceptable as long as the distance is within engineering tolerance. When $\left\|g_{\text{lim}} - \hat{g}(\boldsymbol{x})\right\| = 0$, it means the current design is on the feasible boundary and the constraint function is strictly active. If the surrogate constraint function value is approaching the boundary from the feasible domain, which is

$$\hat{g}(\boldsymbol{x}) < g_{\text{lim}}, \quad \left\|g_{\text{lim}} - \hat{g}(\boldsymbol{x})\right\| \to 0, \tag{5.16}$$

then the constraint function is about to be activated. Such process is usually connected with better objective function value. If the surrogate constraint value is approaching the constraint boundary from the infeasible domain, which is

$$\hat{g}(\boldsymbol{x}) > g_{\text{lim}}, \quad \left\|g_{\text{lim}} - \hat{g}(\boldsymbol{x})\right\| \to 0, \tag{5.17}$$

then the feasibility of the current design improves. As the surrogate model improves gradually with the CEI infill criterion, the estimated feasible boundary approaches the true feasible boundary. In a word, the constrained expected improvement tends to balance the preference of better design by EI and feasible design by PF. In other word, the CEI defined in Equation 5.14 identifies the locations with better objective function values according to EI and with constraint function values close to g_{lim} according to PF.

5.4.2. Performance of the CEI Infill Criterion on a Test Example

The characteristics of CEI is also demonstrated with the one-variable test function as objective function, and with a constraint function. The constrained test problem is formulated as:

$$
\begin{aligned}
\text{minimize} \quad & z(x) \; = (6x - 2)^2 \sin(12x - 4), \quad x \in [0, 1] \\
\text{such that} \quad & g(x) \; = 25x - 22 - z(x) < 0.
\end{aligned}
\tag{5.18}
$$

It follows the same steps as that of EI with only the infill criterion changed to CEI. The process of refining the surrogate model of the one-variable function with CEI is illustrated in Figure 5.6. The hatched line represents the boundary of the constraint function with the hatched side being infeasible. It can be seen that the optimal result is obtained within six iterations. In **Iteration 6**, the infill point is found close to the constraint boundary and has the best objective function value. A total number of nine sample points is used, which include three initial samples and six infill points. Properties of the CEI infill criterion are discussed as follows:

- The magnitude of CEI is also not necessarily relevant with the quality of the surrogate model. The order of the magnitude of CEI is generally much smaller than that of RMS and EI. This is because that CEI is the production of two small quantities. The relative difference of CEI along the design space is important in identifying locations of infill point. Therefore, it requires the computer to have high operation precision, which is generally available in the computers today.

- In **Iteration 2** and **Iteration 3**, it is shown that the infill points found by CEI shift towards the boundary of constraint function compared with those by EI. This indicates that in these iterations the possibility of feasibility dominates. In **Iteration 4** and **Iteration 5** infill points appear near the optima of the current surrogate models. This indicates that the expected improvement dominates. In **Iteration 6** infill point is close to the boundary of constraint function. Beside this one, all infill points are feasible.

- The number of peaks in CEI is far fewer than that of RMS. This shows that CEI can concentrate on the interested positions.

- A strong oscillation in CEI near the constraint boundary in **Iteration 6** is a result of the balance between EI and PF, since neither is able to overwhelming the other in this region.

The efficiency of CEI in identifying the critical infill points also depends on the initial surrogate model. To pursue further improvement in efficiency and accuracy, a combination of RMS and CEI as an infill criterion is preferred, which will be implemented in the next section.
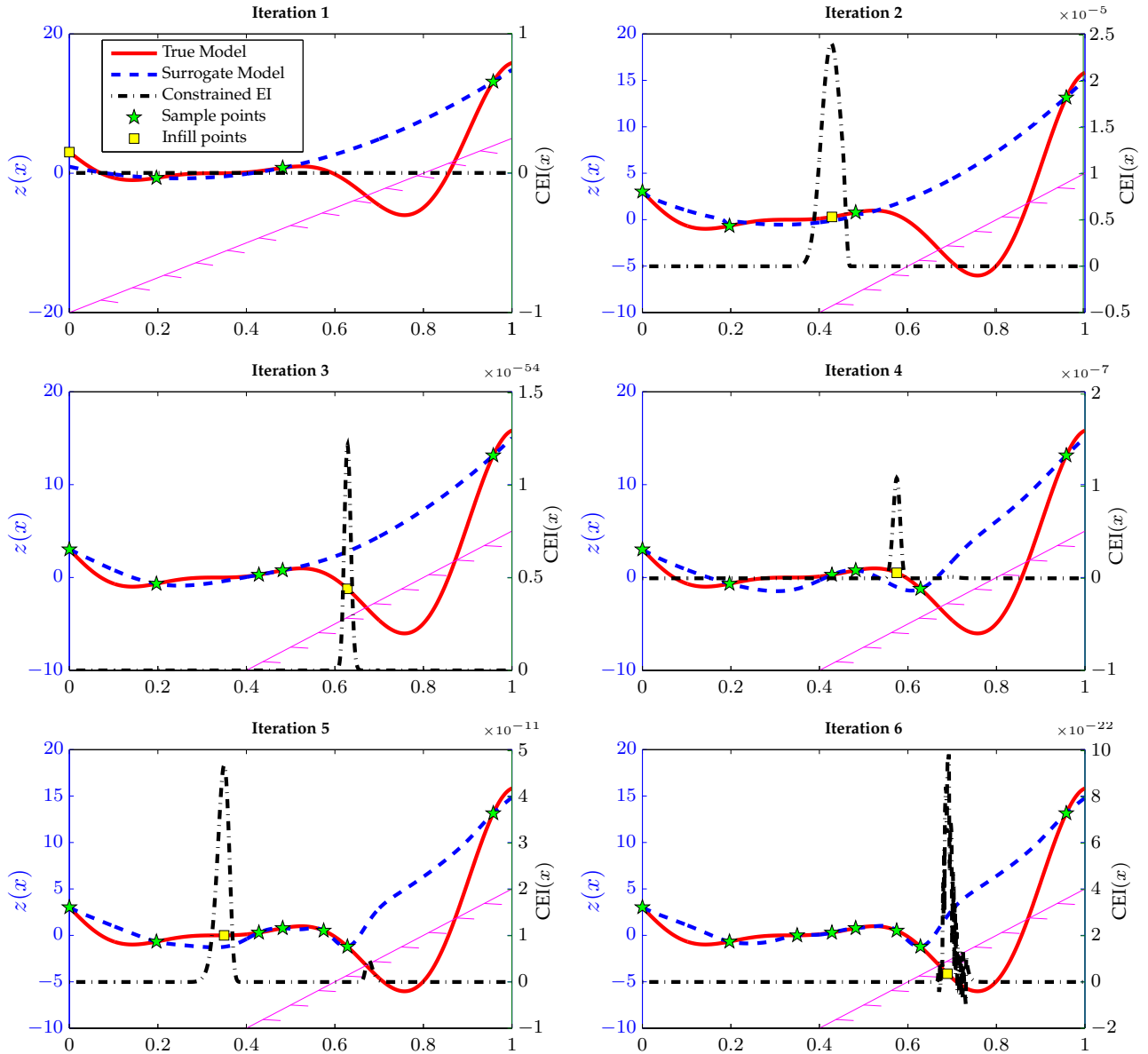
Figure 5.6.: The first six iterations of adaptive surrogate-based optimization of the constrained one-variable function. Constrained expected improvement (CEI) is used as infill criterion.

5.5. Adaptive Surrogate-Based Design Optimization Using Multiple-Infill Criteria

5.5.1. Introduction of Infill Criterion, Optimization method and Validation conditions

Multiple-infill criteria are proposed in this dissertation. Instead of searching for a single infill point per iteration, the multiple-infill criteria can identify many infill positions at once. Advantages of generating multiple infill points are described as follows:

- The initial surrogate model can be refined faster. Fewer iterations are required to achieve the same accuracy.

- The quality of the surrogate models can be explored more widely, and more regions to be refined can be identified at once.

- Optimal design can be found more efficiently. Since more regions are explored, local and global optimal regions can be identified earlier.

- Fewer infill points are required to complete ASBDO, because once the desired region is discovered, no more effort will be spent on the inferior regions.

Considering that the RMS, EI and CEI are more or less multimodal, multiple infill points should be found using both global searching algorithm and local searching algorithm with multiple starting points. The global search is realized by the GA, while the multiple local searches are performed with sequential quadratic programming (SQP) starting from a number of distributed initial designs. The local optimization for searching multiple infill points is noted as N_P SQP in this dissertation, where N_P is the number of starting points. SQP is a gradient and Hessian [Hesse (1897)] based method, which uses Lagrangian functions for the optimization of nonlinear and continuous system equations. For example, considering a nonlinear contrained optimization problem:

$$
\begin{aligned}
\text{minimize} \quad & z(\boldsymbol{x}) \\
\text{such that} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq 0 \\
& \boldsymbol{h}(\boldsymbol{x}) = 0.
\end{aligned}
\tag{5.19}
$$

The Lagrangian function for this problem is formulated as:

$$
\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = z(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \lambda_j h_j(\boldsymbol{x}),
\tag{5.20}
$$

where $\boldsymbol{\lambda}$ is a vector of the Lagrange multipliers [Lagrange (1815)]. The SQP method iteratively finds the optimal search direction $\boldsymbol{s}_k$ (k: iteration index) of the following optimization problem:

$$
\begin{aligned}
\text{minimize} \quad & \mathcal{L}_k + \nabla \mathcal{L}_k^T \cdot \boldsymbol{s}_k + \frac{1}{2} \boldsymbol{s}_k^T \cdot \mathbf{H}_k \cdot \boldsymbol{s}_k \\
\text{such that} \quad & \boldsymbol{g}_k + \nabla \boldsymbol{g}_k^T \cdot \boldsymbol{s}_k \leq 0 \\
& \boldsymbol{h}_k + \nabla \boldsymbol{h}_k^T \cdot \boldsymbol{s}_k = 0.
\end{aligned}
\tag{5.21}
$$

The optimization problem in Equation 5.21 is composed of quadratic objective and linearized constraint functions of s_k. $\mathbf{H}$ is the Hessian matrix, which is a square matrix of the second-order derivatives of $\mathcal{L}(\boldsymbol{x})$. For maximization problem, it is only necessary to use $-z(\boldsymbol{x})$ instead of $z(\boldsymbol{x})$. A more detailed introduction of SQP and the implementation of SQP can be found in Powell (1978) and Baier et al. (1994).

The whole process with GA and N_P SQP in this dissertation is called multiple-searching process, which is capable of identifying multiple infill points for improving the surrogate models. It should be noticed that during the multiple-searching process, it is possible that identical points or very closely adjacent locations are obtained. Therefore, the distances between all the locations of infill points are checked. According to these distances, redundant points are identified and deleted so that no repetitive high-fidelity evaluations are carried out. The ASBDO process of a constrained optimization problem using the multiple-searching infill criteria is listed as follows:

(i) Initial DoE is obtained with Latin hypercube sampling.

(ii) Initial surrogate models are constructed using Kriging.

(iii) GA and N_P SQP are used to find multiple infill locations with $\mathrm{RMS}(x)$.

(iv) Infill points are evaluated and surrogate models are updated.

(v) Surrogate-based design optimization is executed to get the best design x^* and $\hat{z}_{\min}$. Check whether this design is a feasible optimum. Stop the whole process if it is satisfied, otherwise, continue.

(vi) Use GA and N_P SQP to find infill points with $\mathrm{CEI}(x)$ and go to (iv).

In step (v) of the above process, the Karush-Kuhn-Tucker (KKT) optimality conditions are checked for the current design. The KKT conditions introduced by Karush (1939), Kuhn and Tucker (1951) for a general optimization problem in Equation 5.19 are described as follows:

$$
\begin{aligned}
&\nabla z(\boldsymbol{x}^*) + \sum_i \nabla g_i(\boldsymbol{x}^*) \cdot \lambda_i + \sum_j \nabla h_j(\boldsymbol{x}^*) \cdot \lambda_j = 0 \\
&\boldsymbol{g}(\boldsymbol{x}^*) \le 0 \\
&\boldsymbol{h}(\boldsymbol{x}^*) = 0 \\
&\sum_i g_i(\boldsymbol{x}^*) \cdot \lambda_i = 0 \\
&\boldsymbol{\lambda} \ge 0.
\end{aligned}
\tag{5.22}
$$

For engineering problems, usually the above conditions do not have to be strictly satisfied, as long as they are within the predefined tolerance of the sytem. By allowing some engineering tolerance, the optimization converging process can be greatly speeded, while the final design can still be very close to the theoretical optimum. To validate the surrogate-based optimization results, z, $\boldsymbol{g}$ and $\boldsymbol{h}$ in Equation 5.22 are replaced by $\hat{z}$, $\hat{\boldsymbol{g}}$ and $\hat{\boldsymbol{h}}$, respectively. Besides, the surrogate-based optimal designs need to be evaluated with the high-fidelity models, so as to get the real objective and constraint function values: $z(\boldsymbol{x}^*)$, $\boldsymbol{g}(\boldsymbol{x}^*)$ and $\boldsymbol{h}(\boldsymbol{x}^*)$.

Then the following conditions are checked:

$$\begin{aligned}
\|z(\boldsymbol{x}^*) - \hat{z}(\boldsymbol{x}^*)\| &\leq \xi_z \\
\|g_i(\boldsymbol{x}^*) - \hat{g}_i(\boldsymbol{x}^*)\| &\leq \xi_g \\
\left\|h_j(\boldsymbol{x}^*) - \hat{h}_j(\boldsymbol{x}^*)\right\| &\leq \xi_h.
\end{aligned} \tag{5.23}$$

ξ_z, ξ_g and ξ_h in Equation 5.23 are the accuracy tolerances of the objective funtion and constraint functions defined for the surrogate models. It has to be noticed that in engineering problems, not all system equations are necessarily approximated. Only those system responses, which are obtained through computationally expensive models are approximated. Therefore, the conditions in Equation 5.23 are also only necessary for the approximated system equations.

In the multiple-searching process, N_P is a very importance parameter, which affects the efficiency in the searching of infill points, as well as the efficiency in finding the optimal design. General rules for adjusting N_P are discussed as follows:

- N_P should increase with the dimension of design space.

- For highly-nonlinear or oscillating system responses, larger N_P is desired.

- For initial surrogate models, N_P should be set relatively large. As the quality of surrogate models improves, N_P can be set smaller and smaller.

5.5.2. Performance of the Multiple Infill Criterion on a Test Example

To demonstrate how the multiple-searching method works in ASBDO, the constrained one-variable test example in Equation 5.18 is used. First, the setting $N_P = 7$ is made and the multiple-searching method is performed using $RMS(x)$ as infill criterion. Then, for this one-dimensional problem, $N_P = 0$ is set and $CEI(x)$ is used as infill criterion. The multiple-searching process is shown in Figure 5.7. It can be seen that the optimum of the constrained one-variable test problem is found within three iterations. The total number of sample points used to solve this problem is eight, including three initial samples and five infill points. Remarks of multiple-searching process are as follows:

- In **Iteration 1** one global and two local maxima of the $RMS(x)$ function are identified with the multiple-searching method. Compared with the single infill point found in **Iteration 1** of Figure 5.3, the multiple-searching method allows to explore the system more thoroughly. This behavior is very important for refining of the initial surrogate models, which are usually extremely rough because of limited information.

- With the surrogate model refined after adding the multiple-infill points from **Iteration 1**, the $CEI(x)$ infill criterion is able to identify the global optimum of the constrained problem very efficiently in the next iterations. In **Iteration 3** the infill point is found exactly at the optimum.
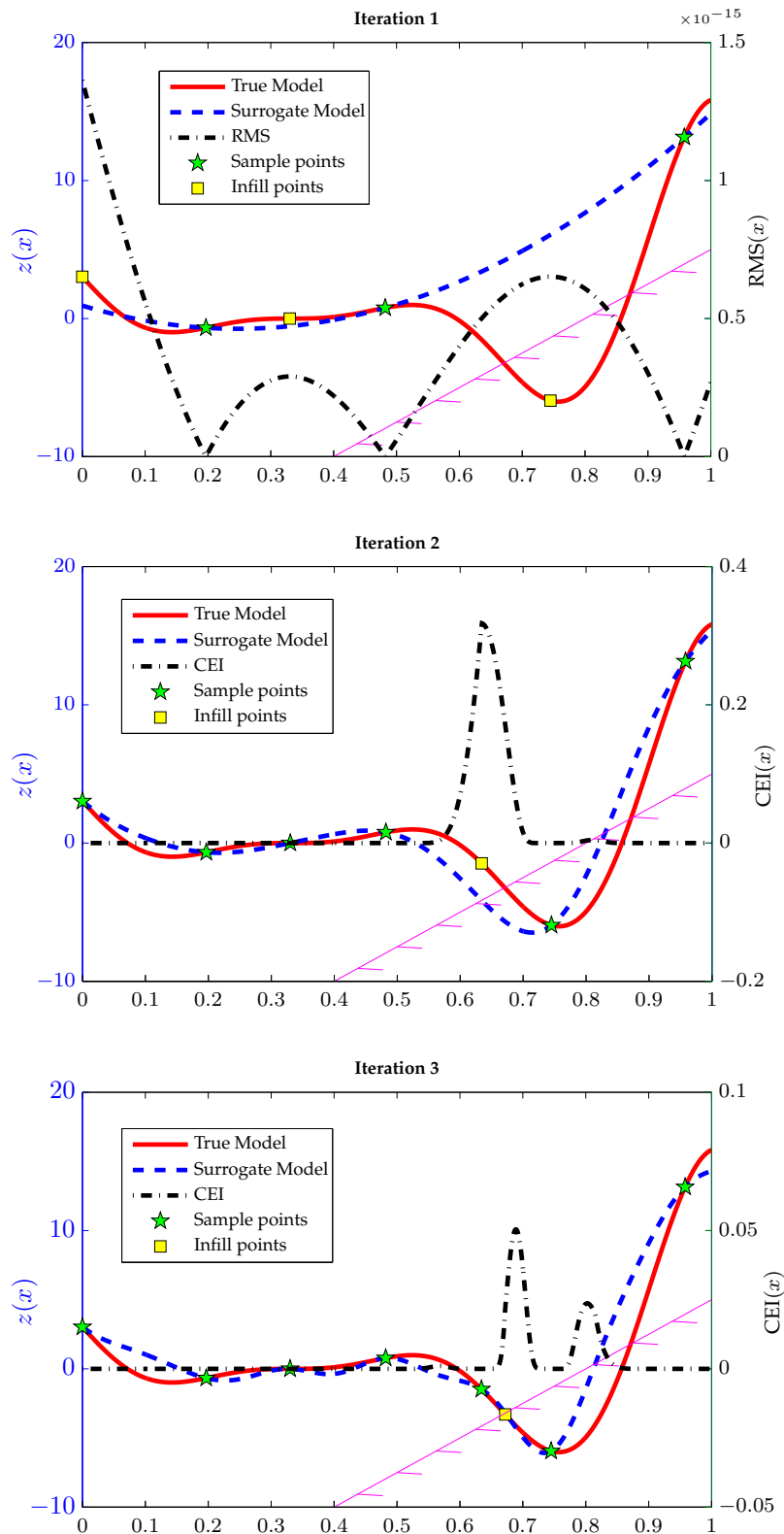
Figure 5.7.: Three iterations of adaptive surrogate-based optimization of the constrained one-variable problem. Multiple-searching process is used to identify infill points using RMS as infill criterion in **Iteration 1** and CEI as infill criterion in the following iterations.

By using the multiple-infill criteria, no effort is wasted in wondering round the local valleys of the test model after the first iteration. It can be seen that although the final surrogate model are quite inaccurate at many positions, the optimum result is accurate enough. That is to say, in surrogate-based optimization problems, the surrogate models do not have to be accurate all over the design space. For engineering problems with large design space and very complicate system responses, it is even more important to focus the computation effort on the critical sites. Therefore, the multiple-searching method accompanied with the RMS and CEI infill criteria is suitable for large engineering problems.

To given an overview of all the different infill methods discussed in this chapter, their performance on the one-variable test problem is shown in Table 5.1. As can be seen that the multiple-searching infill criterion has the best efficiency in the one-variable test problem. It requires the fewest number of iterations and high-fidelity evaluations. Besides, it is able to search first widely in the design space and then focus on the feasible and global optimal regions. The summation in Table 5.1 is supposed to be taken as a general reference for

Table 5.1.: Properties of different infill criteria in the one-variable test problem.

Infill criterion	RMS	EI	CEI	Multiple-searching
Number of iterations	8	7	6	3
Number of evaluations	10	9	9	8
Range of exploration	Wide	Focused	Focused	Wide&Focused
Deal with constraints	No	No	Yes	Yes
Aim at optimization	No	Yes	Yes	Yes

choosing infill criteria according to the description of the tasks. This is discussed as follows:

- If the task is to approximate the system responses as accurately as possible in the whole design space, then RMS is preferred.

- If the task is to perform surrogate-based optimization and only the accuracy of the optimum is cared about, then the EI, CEI and multiple-searching infill criteria can be considered.

- For constrained optimization problem, CEI and multiple-searching infill criteria are the choice.

- For constrained optimization with large dimensional design space or multimodal system responses, the multiple-searching as infill criterion is recommended in adaptive-surrogate-based design optimization.

Last but not least, it should be noted that these data are based on the one-variable test example and the numbers in Table 5.1 do not indicate absolute advantage of one method over another. However, the test function is a typical and critical problem for surrogate modeling and optimization, especialy when the sample points are extremely unlucky, which is the

case tested in this chapter. With this one-variable function, it is not only able to illustrate the infill processes, but also possible to show the characteristics of the different infill criteria. Besides, the infill criteria are also tested and shown to be very efficient on a two-variable multimodal function and a 15-variable aircraft wingbox optimization problem in Xu et al. (2012). Further, in Section 5.7 of this chapter, the multiple-searching infill criterion is tested in surrogate-based optimization for the stiffened composite panel introduced in Section 4.6.1. In Chapter 6 and 7, this infill criterion is used in ASBDO to solve two large-scale engineering problems.

5.6. Parallel Computing in Adaptive Surrogate-Based Design Optimization

Parallel computing techniques have been developed with the rapid progress in computer science and hardware. With parallel computing programs, many calculations can be processed simultaneously, which allows large speed-up in mission execution compared with sequential processing. Therefore, parallel computing are preferred in problems with time-consuming numerical simulations, such as computational fluid dynamics (CFD), fluid-structural interaction (FSI) and crashworthiness simulations.

A basic requirement to perform parallel computing is that a task should be able to be divided into independent subtasks. Those subtasks are then assigned to parallel processing units. There are many different levels and types of parallelization techniques available, e.g. multicore computing, cluster computing, massive parallel processing (MPP), etc. Yet there is no infinite speed-up with parallel computing. For one thing, even when a task is broken into many small parts, there are still serial commands to be processed. For another thing, time is also spent in the transmission of data and communication for assignment of tasks between processing units. Therefore, the decision on parallelization should be made based on the given tasks and the available resources. In adaptive surrogate-based design optimization, several tasks can be broken into independent subtasks, which are:

(1) evaluation of sample points (parallelized),

(2) multiple-searching of infill points with GA and N_P SQP (not parallelized),

(3) surrogate-based design optimization with GA and N_P SQP (not parallelized),

(4) evaluation of infill points (parallelized),

(5) validation of surrogate-based optimal designs (parallelized).

All these five tasks can be parallelized, yet not all of them can benefit from parallelization. In (1), (4) and (5), parallel computing is used since high-fidelity system simulations for each individual point are usually the most time-consuming parts in ASBDO. No parallelization is necessary in (2) and (3), where surrogate models are evaluated.

According to the resources available in LLB, cluster computing is used for parallelization. The architecture of the cluster in LLB is shown in Figure 5.8. The cluster in LLB has
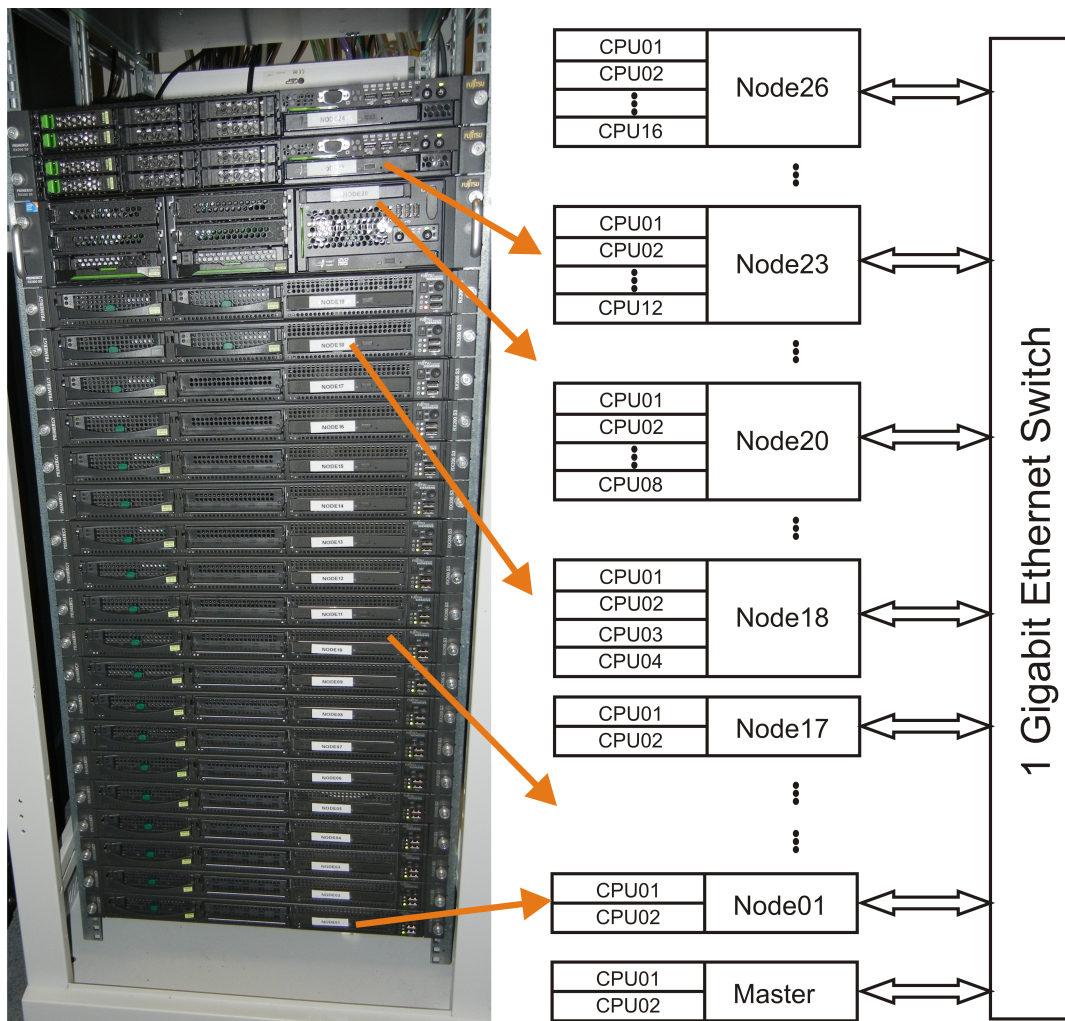
Figure 5.8.: The architecture of the cluster in LLB.

been developing to provide more and more nodes and faster calculation ability. As can be seen, the cluster currently contains one master node and another 26 nodes. Each node has a number of CPUs, ranging from two for most of the nodes, to 16 for a few large nodes. The total number of CPUs is 132, which means theoretically, 132 tasks can be processed at the same time. The number of CPUs N_c, which are currently free for submitting tasks is called the capability of the cluster. All nodes in the cluster are connected with the one Gigabit Ethernet switch, which allows communication between all nodes.

Distributed computing program is used for implementation of the parallel computing on the cluster. First, one logs in to one of the node on the cluster and prepare subtasks. Then, the multiple subtasks are submitted and dispatched to different CPUs on different nodes by a queuing system through the communication between switches and nodes. Also, by communicating with the switch, running status of the jobs on different nodes can be checked. Finally, calculation results from different nodes are gathered to the original master node. The program for distributed computing has been developed in LLB since 2004 by Langer (2007). The program was originally written to speed up optimization with evolutionary algorithm (EA), where fitness evaluation of individuals in a generation are parallelized. The

program was then used by Binder and Wehrle (2012) in LLB for parallelized sample points evaluation in surrogate modeling of crashworthiness problems.

In this dissertation, the program will be used to parallelize the evaluation of sample points, infill points and optimal designs in ASBDO. The speed-up ratio r_s of parallelization over sequential processing is defined:

$$\mathcal{P}_p = \frac{\sum_{i=1}^{N} T_i + T_s}{T_p + T_A + T_s},$$ (5.24)

where,

T_i: time required for evaluation of each single subtask,

T_s: time spent on serial commends of a task, which are not parallelized,

T_A: administration time, which is the cost of communication between nodes on the cluster,

T_p: time used by parallelized computing of all N subtasks.

For computationally expensive simulation problems, T_s and T_A can be neglected compared with T_i. The ratio of speed-up is limited by T_p, which equals to $\max(T_i)$, $i = 1, 2, \cdots, N$. It is the time spent on the slowest subtask if the number of subtasks does not exceed the capability (N_c) of the cluster. If $N > N_c$, then the exceeded number of jobs have to wait until there are CPUs free on the cluster. In that case T_p is larger than $\max(T_i)$. Therefore, both limitations and benefits have to be kept in mind while implementing parallelization in order to make best use of the hardware and software resources.

The parallelized process of adaptive surrogate-based design optimization is noted as PASBDO in this dissertation. The flowchart of PASBDO is demonstrated in Figure 5.9. The double arrows ⇊ refer to the parallelized high-fidelity simulations. It is shown that in PASBDO, parallelization is made for the high-fidelity evaluation of sample points from DoE, infill points and validation of optimization results. The searching process with infill criteria and the surrogate-based optimization are performed on the computationally inexpensive models. GA and multiple-starting point SQP (N_P SQP) are used as optimization algorithms.

In PASBDO, one loop is a complete surrogate-based optimization process, which includes surrogate modeling, surrogate-based optimization, validation of surrogate optmial results, searching with infill criteria and the evaluation of infill points. The concept of the loop in PASBDO should be differed from the iteration in an optimization alqorithm. In the first loop of PASBDO, initial sample points are parallelly evaluated, initial surrogate models are constructed, and RMS is used as infill criterion to improve the surrogate models widely in the design space. Afterwords, surrogate models are updated and multiple surrogate-based optimizations are performed with GA and SQP from N_P starting points. The surrogate-based optimization results are evaluated and validated. With the information from the validation, GA and N_P SQP are used again to focus the search of infill points with CEI for better optimal designs. It is required that by the multiple-search with RMS in the first loop, the initial surrogate models should already be globally explored, and, therefore,
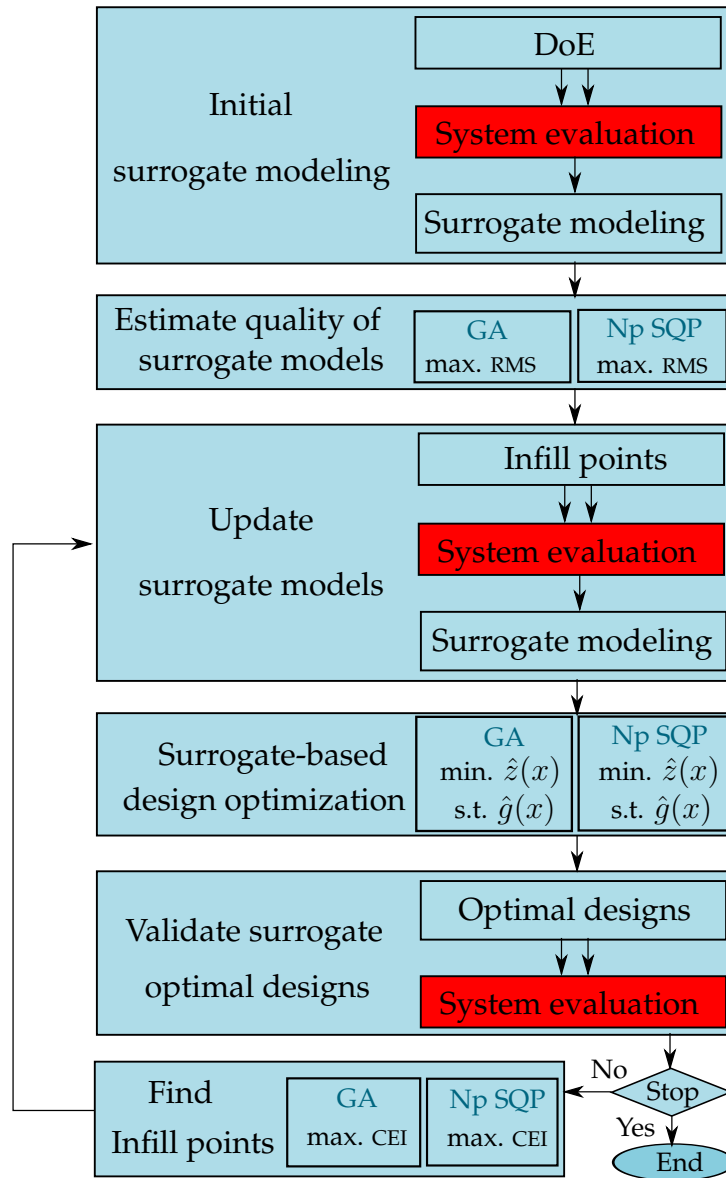
Figure 5.9.: Flowchart of parallelized adaptive surrogate-based design optimization.

from the second loop on, the multiple-search with only CEI is performed.

The maximum number of infill points obtained in one multiple-searching process can be as many as $N_P + 1$: one from GA and N_P from SQP. Since those infill points are parallelly evaluated, another rule for setting N_P is to be considered: it is better to make sure that N_P does not exceed the available computational resources N_c on the cluster. One should be aware that once the number of infill points to be parallelly computed exceeds N_c, all CPUs will be occupied and the waiting time can be as long as the time one high-fidelity simulation takes.

Last but not least, although it is not shown in the flowchart of Figure 5.9, engineering knowledge and GA assisted knowledge-based surrogate modeling can take part in all surrogate modeling modules for further higher efficiency and accuracy.

5.7. Demonstrative Example: Parallelized Adaptive Surrogate Based Design Optimization for Stiffened Composite Panel

In this section, the fiber angle design optimization problem for the stiffened composite panel, as introduced in Section 4.6, is solved with three methods. First, the optimization problem is conducted without using surrogate modeling techniques. This method is noted as direct optimization (DO) in this dissertation. Then, with the knowledge-based surrogate models obtained from Section 4.6.7, a surrogate-based design optimization (SBDO) is carried out. Finally, the parallelized adaptive surrogate-based design optimization (PASBDO) is implemented. The optimization results using these three methods are compared and their performances are discussed.

5.7.1. Task Definition for Fiber Angle Design Optimization of Stiffened Composite Panel

As is introduced in Section 4.6, there are 50 design variables for the fiber angle optimization problem, which includes 20 fiber orientation angles for the skin and 30 for the webs of the stiffeners. The buckling load factor is to be maximized under the given loads, while the strength constraint should be satisfied. The optimization problem is formulated as follows:

$$\begin{aligned} \text{minimize} \quad & z(\boldsymbol{\alpha}) = -\lambda_b(\boldsymbol{\alpha}) \\ \text{such that} \quad & g(\boldsymbol{\alpha}) = \text{F}_{\text{TsH}}(\boldsymbol{\alpha}) - 1 \leq 0 \\ \text{and} \quad & -90 \leq \boldsymbol{\alpha} \leq 90. \end{aligned} \tag{5.25}$$

According to the energy method described in Timoshenko (1961), the critical buckling load is determined by minimizing the total potential energy Π of a structural system. A structure is in stable equilibrium when the total potential energy is at a minimum, that is $\delta\Pi = 0$ and $\delta^2\Pi > 0$. Otherwise, the structure is in unstable or neutral equilibrium and a buckling mode is formed. Π is composed of the strain energy U and the external work V, which are calculated on a virtual displacement. The strain energy depends on the stiffness of the structure, while the external work is propotional to the external loads. When the fiber orientation changes, the stiffness of the structure is adjusted, and, therefore, the critical buckling load is also altered. Significant changes in different buckling modes can be seen according to different levels of strain energy. In other word, when a higher level of strain energy is reached by adjusting fiber orientations, a higher buckling load factor will appear after a previously converged buckling load factor. The optimization problem is, therefore, rather chanlleging and computationally extensive.

Knowledge-based surrogate models for both the objective function and the constraint function are constructed, and the surrogate-based design optimization problem is formulated as:

$$\begin{aligned} \text{minimize} \quad & \hat{z}(\boldsymbol{\alpha}) = -\hat{\lambda}_b(\boldsymbol{p}_1(\boldsymbol{\alpha})) \\ \text{such that} \quad & \hat{g}(\boldsymbol{\alpha}) = \hat{\text{F}}_{\text{TsH}}(\boldsymbol{p}_2(\boldsymbol{\alpha})) - 1 \leq 0 \\ \text{and} \quad & -90 \leq \boldsymbol{\alpha} \leq 90, \end{aligned} \tag{5.26}$$

where $\boldsymbol{p}_1(\boldsymbol{\alpha})$ and $\boldsymbol{p}_1(\boldsymbol{\alpha})$ are the knowledge-based properties of the composite panel, which are obtained in Section 4.6.7. The task is, therefore, to accomplish the fiber angle design optimization problem while keeping the computational effort as small as possible.

5.7.2. Optimization of the Stiffened Composite Panel with System Equations From High-Fidelity Simulations

For the stiffened composite panel, FEM is used for structural analysis, which includes a static structural analysis and a buckling analysis. Since the execution expense for this example is not too large (20s per evaluation), a directly optimization is sustainable although not as efficient as using surrogate models. Direct optimization is performed to provide a benchmark, which is used for comparison with PASBDO. In direct optimization, the high-fidelity FEM simulations are used to solve the system equations. SQP (see Section 5.5.1) is selected as the optimization alqorithm because of its ability of search in the gradient direction of the system equations. This feature make SQP more efficient than the genetic algorithm. However, as is shown in the parameter study in Section 4.6.5, the system responses are multimodal. Therefore, a number of feasible starting points are generated for SQP in order to reduce the risk of trapping in local optima. The converging process of the best design from those starting points is shown in Figure 5.10.

Detailed information of several iterations are marked in Figure 5.10, where n_{It} stands
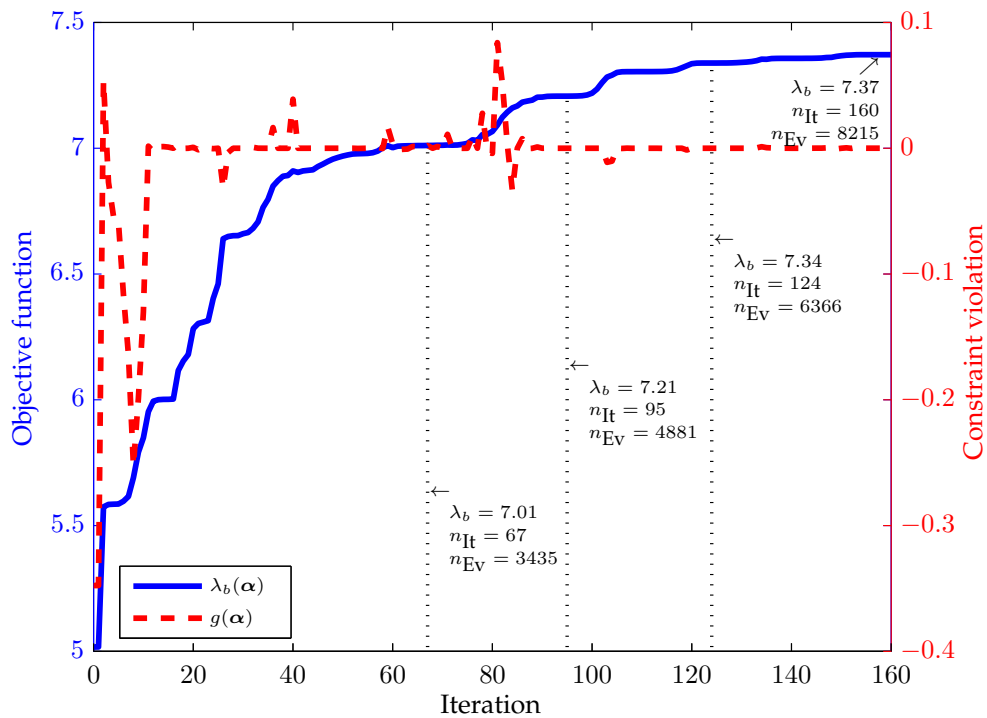

Figure 5.10.: The optimization converging process of the fiber angle design with high-fidelity system equations.

for the number of iterations in the optimization process and n_{Ev} represents the number of evaluations of the high-fidelity system simulation. It is seen that the converging process is composed of several subordinate steps, which is controlled by the magnitude of termination tolerance defined for convergence. Smaller tolerance is helpful for identifying better optima, while it usually costs more iterations.

For the different buckling load factors which are converged, different energy levels of buckling modes are obtained with by designs of the fiber orientations. The objective function λ_b converges first at the 15th iteration with the value 6.00, and the number of high-fidelity evaluations to achieve this level is 772. Then a higher energy level is obtained and λ_b converges at the 29th iteration with the value 6.65. Afterwords, at the 67th iteration $\lambda_b = 7.01$, at the 95th iteration $\lambda_b = 7.21$, and at the 124th iteration $\lambda_b = 7.34$ are obtained. It can be seen that the increase of objective function becomes slower, which means extremely large computational effort is required for very small improvement in the structure. The optimization is stopped at the 160th iteration with $\lambda_b = 7.34$. It has to be noticed that the gradient and Hessian calculation in SQP for this problem is realized by finite differencing. Since this problem contains 50 design variables, 51 system evaluations are performed in one iteration for gradient calculation, and it is even more when Hessian is required. In some FEM codes, analytical or semianalytical gradient calculation are provided, where the number of function evaluation in each iteration is remarkably smaller. However, in this dissertation, focus is on the reduction of computational effort by surrogate modeling instead of the study of FEM solutions. Further, there are engineering problems where analytical and semianalytical gradients are very difficult to obtain.

A summary of the discussion about the performance of direct optimization is listed as follows:

- The optimization starts from a feasible design, then varies to slightly violated designs, but finally converges at feasible designs.

- Buckling modes with higher and higher energy level are slowly identified in the optimization process.

- Although the DO converges relatively fast for optimization within each energy level, the total number of iterations is very large to get high buckling modes.

- If the termination tolerance is looser, then the optimization can converge earlier at slightly lower buckling load factors. For example, at the 67th iteration $\lambda_b = 7.01$, where the number of high-fidelity simulation is 3435.

5.7.3. Surrogate-Based Design Optimization of the Stiffened Composite Panel

In this section, instead of calling the high-fidelity system simulations, the optimization is performed on the surrogate models. First, the surrogate models for λ_b and FI, which are constructed in Section 4.6.7 with GA assisted knowledge-based surrogate modeling method are used for optimization. No adaptive strategy nor infill criterion is used here, which means

the optimization result depends completed on the quality of the surrogate models. The approximation errors on λ_b and FI according to the validation data in Section 4.6.8 are 4.71% and 2.20%, respectively, which are pretty good. Also, since only 150 sample points are involved in the surrogate models, the number of high-fidelity evaluation in this test is 150 as well.

The converging process of SBDO is demonstrated in Figure 5.11. It is shown that the optimization converges at the 20th iteration. The solid line shows the objective function evaluated on surrogate model $\hat{\lambda}_b$ in each iteration. The true objective function value λ_b in each iteration is also given, shown with the dash-dotted line.
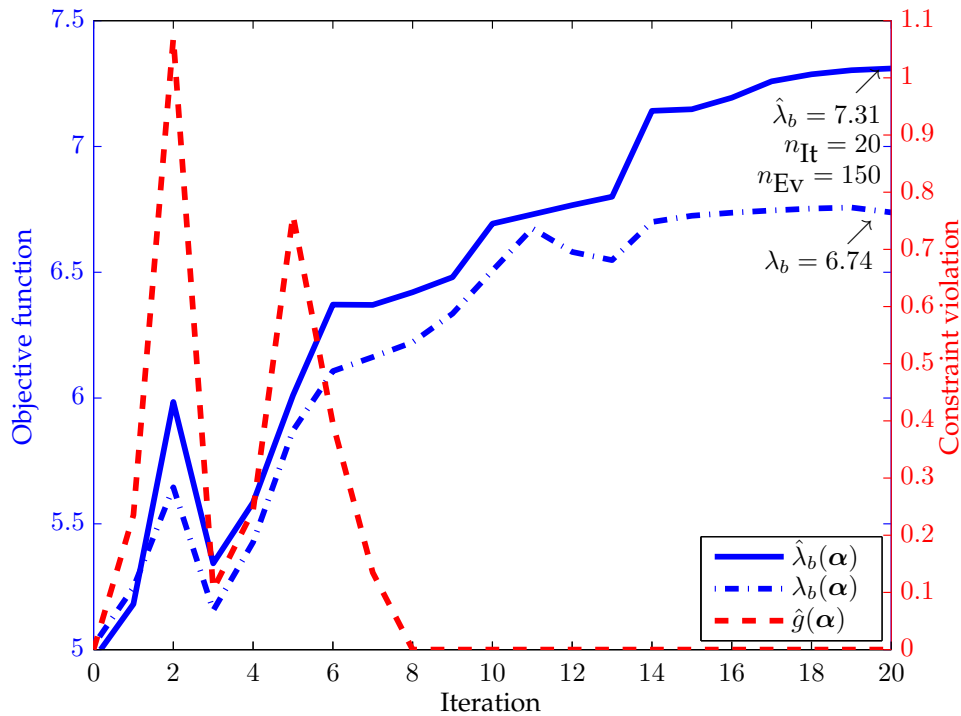

Figure 5.11.: The converging process of SBDO with 150 sample points for the fiber angle design of the stiffened composite panel.

To make a fair comparison between DO and SBDO, the same feasible initial design is used here. The final objective function value of SBDO ($\hat{\lambda}_b = 7.31$) is also very close with DO ($\lambda_b = 7.34$). However, when looking at the true system response ($\lambda_b = 6.74$) of this design, there is a difference, which means the design obtained by SBDO is actually different from DO. It is seen in Figure 5.11 that during the optimization, the $\hat{\lambda}_b$ and λ_b have generally the same trend, although not exactly the same. The distance between $\hat{\lambda}_b$ and λ_b grows as the value of λ_b increases. This is a result of the approximation error. As one can see from the consistency check in Figure 4.17, the difference is also larger for higher λ_b. Besides, the number of sample points at the region where $\lambda_b > 6.5$ is quite small, which also causes the relatively high prediction error for large λ_b. Since the objective of this optimization problem is to maximize λ_b, it is necessary to refine the surrogate models for regions of larger

λ_b values. This proves the importance of validating the surrogate designs and adaptive surrogate modeling, even when the quality of initial surrogate models are already pretty good. Remarks about the performance of SBDO are summarized as follows:

- With the same starting point as in DO, SBDO also varies at the beginning between infeasible and feasible designs, but it converges much faster.

- The accuracy of the surrogate model for λ_b reduces as the value of λ_b increases.

- The maximum relative approximation error is 8.46% during the optimization process.

- The number of evaluations in SBDO is 150 and during the optimization only surrogate models are evaluated, which is extremely fast to calculate.

It is shown that when non-ignorable approximation error exists in the important design domains, the surrogate models need to be refined for more reliable optimization results. However, whether this problem can be efficiently solved by SBDO with more sample points
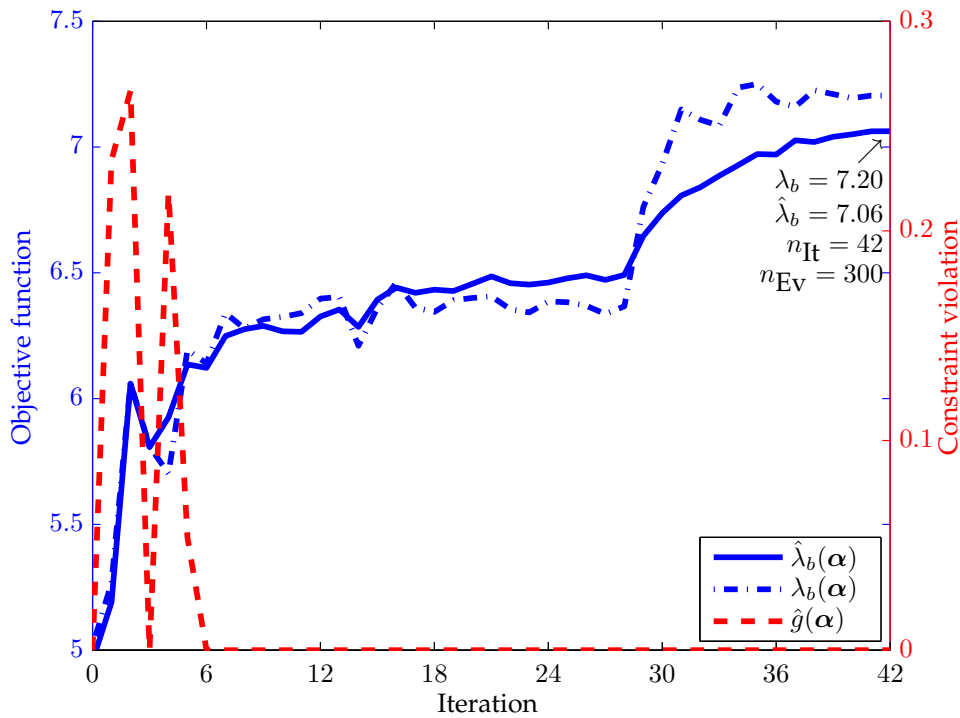

Figure 5.12.: The converging process of SBDO with 300 sample points for the fiber angle design of the stiffened composite panel.

is to be investigated. To demonstrate how much improvement in the performance of optimization can be obtained by using more sample points. Surrogate models constructed with 300 sample points is then tested in SBDO. The same GA assisted knowledge-based surrogate modeling method is used. The approximation errors for the two system responses λ_b and FI are 4.57% and 2.04%, respectively. Compared with 4.71% and 2.20% in the surrogate models

with 150 sample points, the approximation accuracy is slightly improved. The same starting point is again used for fair comparison. The optimization converging process of SBDO with the new surrogate models is shown in Figure 5.12.

In this case, the number of high-fidelity evaluations is 300 and the number of iterations with surrogate models is 42. It can be seen in Figure 5.12 that the surrogate models can approximate the high-fidelity models better in the optimization process. Although the optimization result on the surrogate models still differs from the results on the high-fidelity models, the difference between them is smaller. At the final design $\hat{\lambda}_b = 7.06$ and $\lambda_b = 7.20$, the relative difference of which is 1.94%. The maximum relative approximation error is 4.00% during the whole SBDO process, which is a great improvement compared with 8.46% in the surrogate model with 150 sample points. Besides the improvement in approximation quality, the objective funtion value is also improved, which means better design is obtained. However, the final design is still not accurate enough and the final design is not as good as the benchmark result in DO. Such results suggest that the verification of the final design is necessary after SBDO. ASBDO is preferred because additional sample points are located at selected positions from infill criteria right after verification of the design each time.

5.7.4. Parallelized Adaptive Surrogate-Based Design Optimization of the Stiffened Composite Panel

In this section, the PASBDO method is applied to the optimization problem of the stiffened composite panel. The initial surrogate model is the one constructed with 150 sample points, as used in section 5.7.3. Then the surrogate model is refined with multiple-infill points identified according to the RMS infill criterion. After that, the fiber angle design optimization is performed on the updated surrogate models and the CEI infill criterion is used to find multiple-infill points, which are aimed at increasing the accuracy of the surrogate models near the optimal regions. The converging process of PASBDO of the stiffened composite panel model is shown in Figure 5.13.

One has to be aware that the process of PASBDO is composed of several loops. In each loop, surrogate-based design optimization is performed. Therefore, each loop of PASBDO contains the iterations of a complete optimization process. The convergence of PABDO has two conditions:

- convergence has to be achieved through the loops of PABDO, and

- convergence has to be achieved through the iterations of optimization algorithm inside each loop of PASBDO.

It is shown in the left graph of Figure 5.13 that a design is converged with 8 loops of PASBDO. The graph on the right side shows the iterations of an optimization in the last loop of PASBDO.

Each loop of PASBDO is a complete multiple-searching process, which contains surrogate modeling, search of infill points, refinement of surrogate models, SBDO and verification of optimal results. Since the surrogate-based optimal results are validated and corrected in each loop of PASBDO, all objective function values in the left graph are exactly
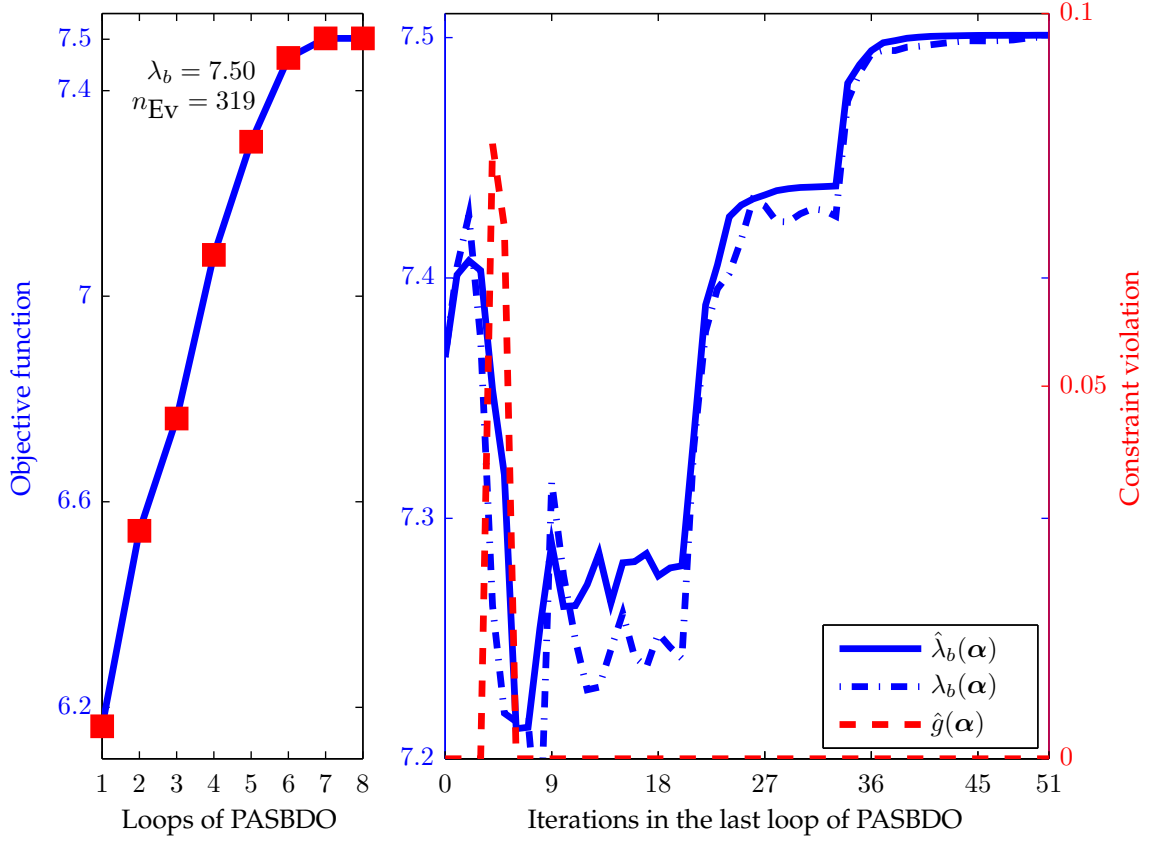
Figure 5.13.: Left: the converging process of PASBDO for the fiber angle design of the stiff-ened composite panel; right: the converging process of the SBDO in the last (8th) loop of PASBDO.

the same as in the high-fidelity models, which is $\hat{\lambda}_b = \lambda_b$. The final objective function value is $\hat{\lambda}_b = \lambda_b = 7.50$, which is the best compared with those obtained from DO and SBDO in previous sections. The total number of high-fidelity evaluations is 319, which includes the evaluation of the 150 initial sample points, all infill points and the optimal designs in each loop of PASBDO. Besides, one can see that the objective funtion value is already better in the first loop of PASBDO ($\lambda_{b0} = 6.16$) than the starting points in DO and SBDO ($\lambda_{b0} = 5.02$). This is because in the first loop of PASBDO, instead of immediately performing SBDO, the initial surrogate models are globally explored and updated with RMS infill criterion first. With the improved surrogate models, SBDO is able to find better optimal designs as the result for the first loop of PASBDO.

In the right graph of Figure 5.13, a SBDO process in the last loop of PASBDO is show. Both the $\hat{\lambda}_b$ and λ_b through the iterations are plotted. It is seen that the accuracy of the surrogate models in the converging process is much better than those in the previous section. The maximum relative difference between $\hat{\lambda}_b$ and λ_b is only 0.67%. Compared with 8.46% in SBDO with 150 sample points and 4.00% in SBDO with 300 sample points, the quality of surrogate models in PASBDO are much better. Therefore, it is shown in the fiber angle

design problem of the composite panel that the PASBDO method has the best performance. The accuracy of the surrogate model is vastly improved in the optimization process with the help of the multiple-searching infill criteria. Further, it has greatly increased the efficiency of the optimization problem.

5.7.5. Comparison of the Optimization Performances in Fiber Angle Design of the Stiffened Composite Panel

In this section, the performance of different methods in the fiber angle design optimization problem is summarized. In Table 5.2 the performance data are listed for the three different methods, which are design optimization (DO) via high-fidelity system equations surrogate-based design optimization (SBDO) and parallelized adaptive surrogate-based design optimization (PASBDO), respectively. It shows that the PASBDO has the large advantages in

Table 5.2.: Performances of different methods in the fiber angle design optimization problem.

Performance of methods	DO	SBDO	PASBDO
Number of high-fidelity evaluations	3435	150	319
Buckling load factor	7.01	6.74	7.50
Maximum approximation error	–	8.46%	0.67%
Surrogate modeling speed-up ratio	–	–	10.77
Parallelization speed-up ratio	–	–	1.55
Total speed-up ratio	–	–	16.69

approximation accuracy and the optimization efficiency over the DO and SBDO methods. Compared with DO, great reduction in computational effort is obtained by PASBDO. Only 10% of the evaluations of high-fidelity models in DO is required in PASBDO. Compared with SBDO, the PASBDO method is capable of increasing the accuracy in the surrogate models and the optimization results, with a relatively small increase in the number of evaluations and time expense. The maximum approximation error in the optimization process with SBDO and PASBDO is 8.46% and 0.67%, respectively. Compared with DO and SBDO, PASBDO gives the best optimization result. The speed-up ratio of using surrogate models in the optimization problem over direct optimization is $\mathcal{P}_s = \dfrac{3435}{319} = 10.77$. Further, the parallelization speed-up ratio for this problem is $\mathcal{P}_p = 1.55$, which means the solution is obtained 1.55 times faster in PASBDO than without any parallelization. Therefore, the total speed-up ratio of PASBDO for this fiber angle design problem is

$$\mathcal{P} = \mathcal{P}_s \mathcal{P}_p = 16.69. \tag{5.27}$$

The speed-up ratio of PASBDO is the production of the speed-up ratio by using surrogate models and the speed-up ratio by using parallelization. It has to be noticed that the fiber angle design problem of this stiffened composit panel is a medium-size optimization problem.

The time expense of high-fidelity model for this problem is around 20 seconds, which is of the same magnitude as the time cost in the serial commends of PASBDO. The serial commends of PASBDO include the SBDO optimization process and the multiple-searching process of infill points, where thousands of evaluations of the surrogate models are performed. As a result, the parallelization speed-up ratio for this medium size problem is small, which is only 1.55. The parallelized computation is more beneficial for large-scale structural optimization problems, which take hours or even more time per evaluation. In the following two chapters, the application of PASBDO in two typical large-scale structural optimization problems are given.

The buckling modes of the final optimal results obtained by DO, SBDO and PASBDO are shown in Figure 5.13. There are three graphs in Figure 5.13, which shows from left to right, the buckling mode of the design via DO, SBDO and PASBDO, respectively. In each graph,
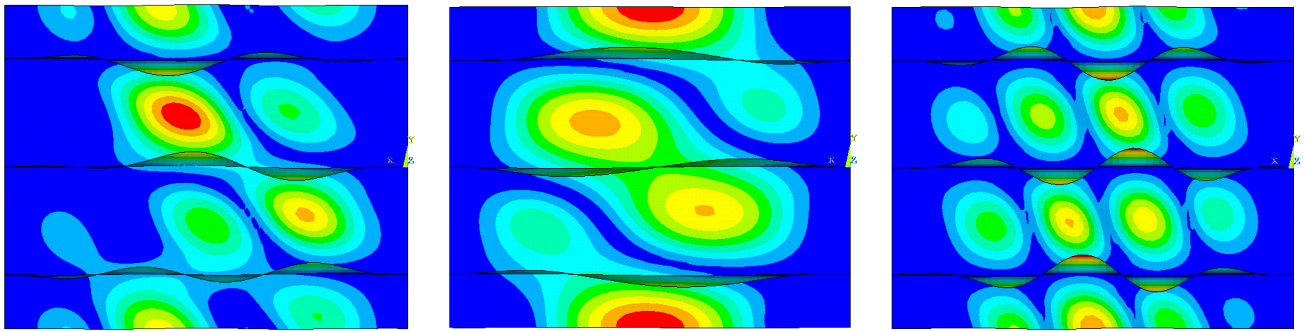

Figure 5.14.: Left: buckling mode of the optimal design obtained by design optimization via high-fidelity system equations (DO); middle: via surrogate-based design optimization (SBDO); right: via parallelized adaptive surrogate-based design optimization (PASBDO).

a bottom view of the stiffened composite panel is shown, which is composed of three stiffeners and one panel as the skin. The three stiffeners are the three curves going horizontally through each graph, and rest part is the skin of the panel. It can be seen that the buckling mode shape of the design obtained by PASBDO has the most local buckling for both the stiffeners and the skin. In the local buckling mode, the deformation in the structure has the smallest magnitude. Such deformation is safer compared with those with large global buckling. This further proves that the PASBDO method is capable of generating better designs than the DO and SBDO methods.

5.8. Summary of Chapter

In this chapter, a parallelized adaptive surrogate-based design optimization method is developed to perform the optimization tasks with high accuracy and efficiency. The PASBDO method start surrogate modeling with relatively small sample size, and then uses multiple-searching criteria to identify infill points, which bring refinement to the current surrogate

models. With the multiple-searching criteria, it is able to estimate the goodness of surrogate models without calculating on further testing points through high-fidelity simulations. Infill points, which are responsible for globally exploring, as well as those aiming at approaching the optimal designs are found in PASBDO. Besides, the PASBDO method chooses a number of best sample points as multiple initial designs for optimization, which also enlarges its ability of identifying better designs for multi-modal optimization problems. The use of parallelization through cluster computation further explores the advantages in efficiency, since all expansive modules in PASBDO can be devided into independent subtasks and then calculated seperately at the same time. In this chapter, the performance of PASBDO is validated with a medium size problem, which is the fiber angle design problem of a stiffened composite panel from an aircraft wing. In the following chapters, the PASBDO method will be applied to solve large-scale problems in both aerospace and automobile engineering design.

6. Parallelized Adaptive Surrogate Based Design Optimization of an Aircraft Wingbox Under Aeroelastic Load

It is important to consider the aeroelastic effects in the process of aircraft design. Computational aeroelastic analysis is used in the early design stage, which involves coupled analysis of the structural and aerodynamic displines. It is well known that one challenge in the analysis with fluid-structure interaction is the high computational cost and time expense, especially in design optimization problems with large amount of design variables. In order to avoid substantive evaluations of computationally expensive system equations, it is necessary to use an effective and reliable method. In this chapter, the design optimization problem of an aircraft wingbox under aeroelastic load is solved with the parallelized adaptive surrogate-based design optmization method.

6.1. Motivation and Overview of the Chapter

The concept of aeroelasticity can be demonstrated in Figure 6.1 with the famous Collar's aerolastic triangle as introduced in Collar (1978). The dynamic aeroelastic analysis considers all three displines, which are aerodynamic, elastic and inertia forces. In this dissertation,
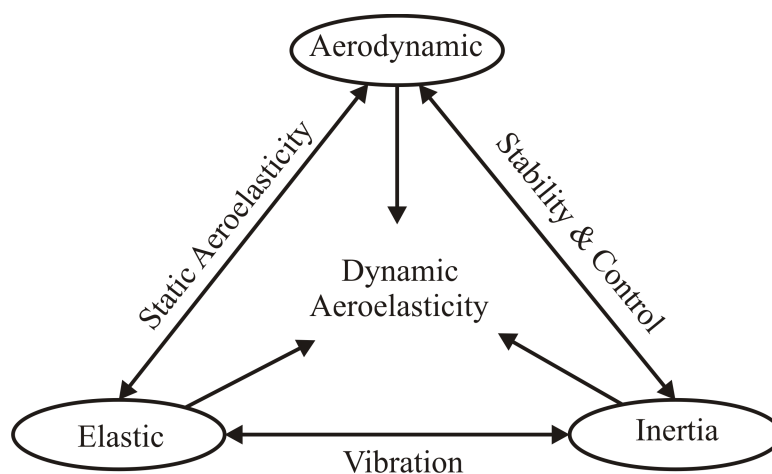

Figure 6.1.: Collar's aeroelastic triangle.

only static aeroelasticity and vibration are analyzed. The stability and control of an aircraft is not considered, so that no coupling of aerodynamic and inertia is necessary. However, the coupling of elastic and aerodynamic has to be performed. Although the studies and practices of aeroelsticity have been developping for a century, there are still challenges to be faced. For example, problems in the computational modeling and simulation of aeroelasticity are listed as follows:

- The aeroelastic analysis of transonic flow fields and complex configurations can only be accurately represented with high fidelity CFD simulations [Raveh (2005)]. Although linear panel aerodynamic models with simplified geometries have shown good accuracy and fast computing feature in performing subsonic and supersonic aeroelastic analysis, the prediction is unreliable when it comes to transonic flow and complex configurations.

- For structures involving complex and large number of modes, it is also difficult to obtain reliable prediction with linear aeroelastic method [Wright and Cooper (2007)] or reduced order model method [Lucia et al. (2004)].

- A higher level of automation in multidisciplinary design optimization (MDO) under aeroelastic analysis is required. Commercial softwares have rapid development and matured performance in many types of aerolastic analyses, such as static aeroelastic trimming, flutter, divergence and gust response analysis [Schuster et al. (2003)]. However, their ability in automatic MDO is still restricted by the complexity of aeroelasticity under variable geometries and flight conditions.

- Last but not least, the high-fidelity CFD simulations require massive computational cost. Therefore, system exploration, parameter study and MDO process are usually blocked by large number of system evaluations in aeroelastic problems.

The problem to be solved in this chapter is an aircraft wingbox under aeroelastic loads, which is a large-scale structural design optimization. The filght condition of the aircraft is descirbed by the following: the maximum flight speed is $V = 236$ m/s (871 km/h) and the flight hight is 11 km. Transonic flow with mach number $Ma = 0.82$ is analysed. In transonic fields, the predition of pressure distribution with the linear aeradynamic models is poor. High-fidelity aerodynamic analysis with CFD and aeroelastic analysis with fluid-structure interaction (FSI) is, however, very time consuming and computationally expensive. High-fidelity CFD coupled with structural finite element analysis is used for computational aeroelasticity, which needs long simulation time for each evaluation. This optimization problem contains large number of design variables, including the thicknesses of the skins, the ribs, the spars and the stringers. Since the design space has large dimension, a great number of search iterations would be required to find the optimal design.

To cope with the above problems, surrogate modeling strategies are favored, mainly for two reasons. First, surrogate models are much more computationally efficient compared with the high-fidelity aeroelastic analysis. Surrogate models are merely simple mathematical approximation of the complex physical behaviors, which can be constructed by exploring

information from adequate number of the high-fidelity sample points. Second, surrogate models provide the posssibility to integrate different simulation codes of different engineering disciplines into a common computing enviroment, using single programming language. By doing this, it automates the evaluation of system equations in MDO.

However, the high dimension of design space also requires that the sample size for surrogate modeling to be extremely large. To avoid this, the PASBDO method is used to ease the problems. Besides mathematical tools for construction and refinement of surrogate models, the adoption of physical understanding of the wingbox structure and the aeroelactic phenomenons to asssit surrogate modeling is important, which is described in details in this chapter. An overview of the rest of this chapter is given as listed:

- First, the aircraft wingbox design optimization problem is described. The design variables, system responses and the high-fidelity simulations are introduced.

- Then, task definition of the optimization problem is given. Here both the definition of the originial design optimization problem and the surrogate-based design optimization problem are shown.

- The physical understanding of the wingbox structure and analytical theory of aeroelasticity are adopted to generate knowledge-based terms, which are the basis in GA assisted knowledge-based surrogate modeling.

- The PASBDO method is used to solve the design optimization problem adaptively and efficiently.

- Finally, the performance of the proposed strategy for solving this problem is discussed, which includes the approximation accuracy of surrogate models and the computational effort.

6.2. Introduction of the Aircraft Wingbox Design Optimization

The aircraft wing studied in this chapter is refered to a passenger aircraft based on the Airbus A320 geometry. The airfoil shape is formed according to the DLR-F4 aerodynamic configuration given in Redeker (1994). A CAD model of the aircraft wing is shown in Figure 6.2, which is constructed in LLB by Armanini (2011). It is shown that the wing is composed of wingbox, flaps, ailerons and engine. Fuel and fuel tanks, which are components inside the wingbox, are not explicitely shown. The structure of the wingbox includes upper and lower wing skins, ribs, stringers, front and rear spars and spars connecting with the fuselage. Wingbox of an aircraft wing is the main structure that sustains aeroelastic loads, while the flaps and ailerons are mainly used for manoeuvring. Therefore, in structural optimization, only the geometric parameters of the wingbox are considered as design variables. To simulate the aeroelastic loads, finite element models for structural analysis and aerodynamic analysis are constructed, which will be introduced as an example of computationally expensive engineering problem in the following sections.
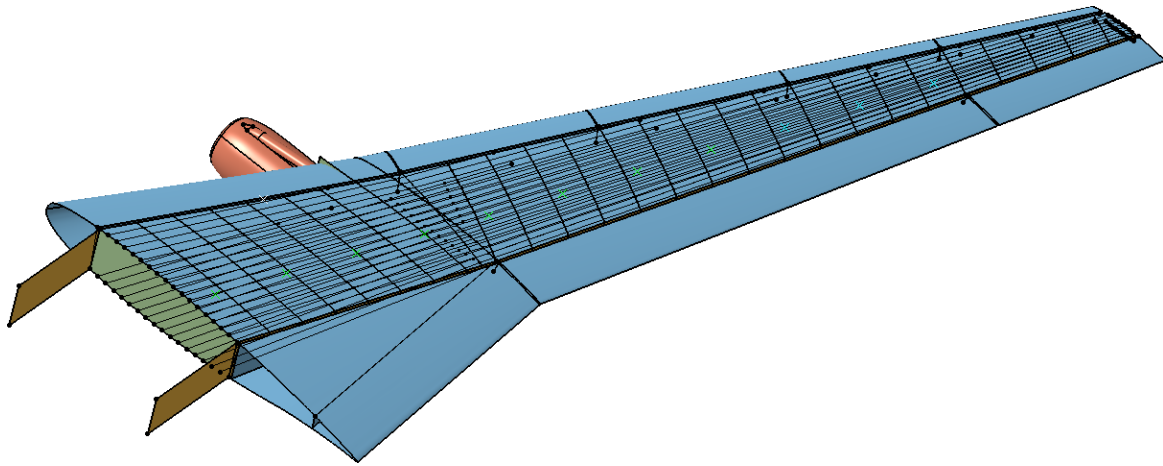
Figure 6.2.: CAD model of an aircraft wing, the wingbox structure (only the parts between front and rear spars) of which is to be optimized.

6.2.1. Simplification in Model Construction of the Aircraft Wing

To facilitate the automatic optimization process, parameterized models are necessary, which are constructed in this chapter with ANSYS Parametric Design Language (APDL). The parameters of the aircraft wing model can be easily altered, followed with automatic reconstruction, remesh and reanalyse. The analysis of the aircraft wing under aeroelastic loads requires two models, one is the structural model and the other is the aerodynamic model. The two models are analysed independently with different types of analyses, but their analysis results are coupled so that their outputs act as inputs of each other. Because of the difference in the analysis, the emphasized components and the requiements in the mesh are different in the two models. For the structural mechanic analysis, the stiffness of the model is the main factor that affects the analysis results, so that the details about the size (thickness), materials, element types of the structural components should be considered. For the fluid dynamic analysis, the outer shape or the configuration should be carefully modeled, while the inner structure can be neglected. Simplification in the model construction of the aircraft wing is necessary, so that computational structural mechanics and fluid dynamics simulations can focus on their own main effects, and the trivial details are ignored. In the following, the simplified structural mechanical model and the aerodynamic model are introduced respectively.

The finite element model for structural mechanic analysis is shown in Figure 6.3. The upper wing skin is not shown on purpose so that the inner structure of the aircraft wing can be seen. In this model, only the wingbox is considered and all flaps are not shown. The engine, the fuel and tanks in the wingbox contribute little in sustaining aerodynamic loads, so that they are all simplified to mass points. The wing skin, spars and ribs have much larger inplane size than the thickness, so that they are all modeled as 2D shell elements. The stringers supporting both the upper band lower wing skin have very large slenderness ratio, so that they are all modeled as 1D beam elements, with a T-cross section.

The aerodynamic model requires the outer shapes instead of the inner structures, so that
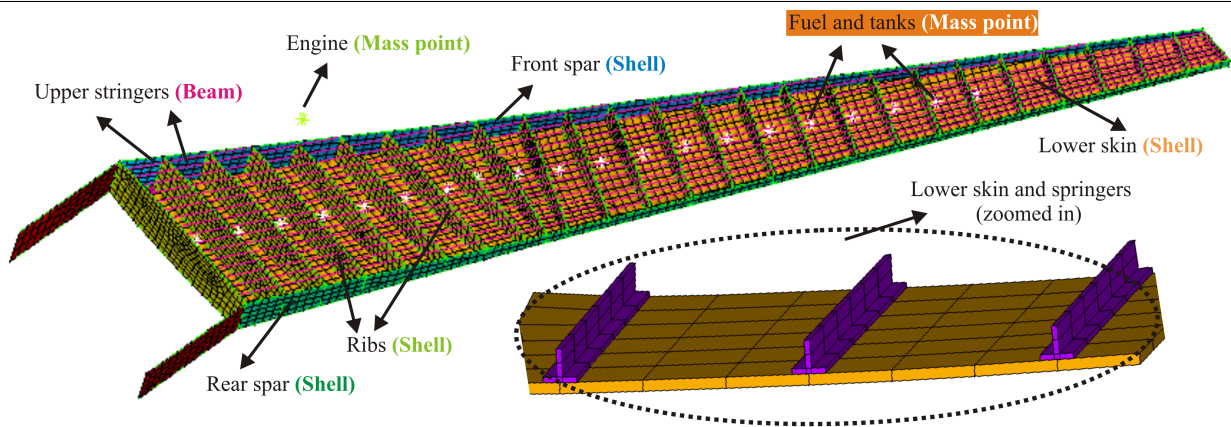
Figure 6.3.: Finite element model of the wingbox for structural analysis (the upper wing skin is not shown).

the shape of the whole wing has to be modeled, as well as the shape of the fuselage. The shape of the engine also plays a role but much smaller, which is for simplification reason not modeled. The airplane has a symmetric configuration, as a result, only half of the fuselage and the right wing are to be considered, and symmetric boundary conditions are given. To model the air around the components, a volume surround the components has to be models, which are given the properties of the air at a certain flight height and speed. The volume has to be kept as small as possible so that compuational effort can be saved. However, the volume of air also needs to be large enough to consider the change of flow around the wing. The volume which contains air surrounding the aircraft is called control volume. The size of
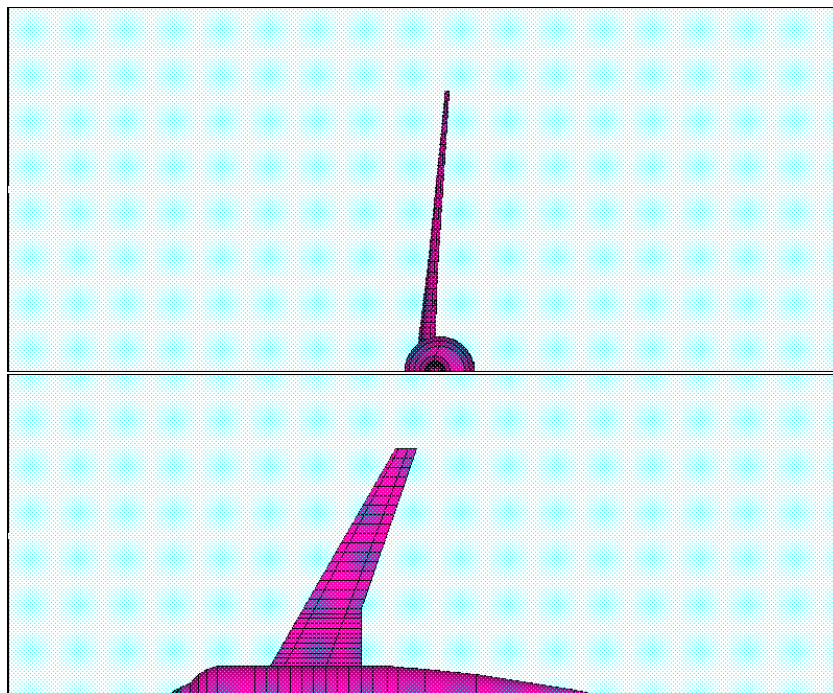

Figure 6.4.: Control volume for flow field in aerodynamic analysis.

the volume is determined refering to engineering experience and by trying and adjusting. The control volume is shown in Figure 6.4, which is demonstrated with two different angles of view.

Once the control volume is defined, it is meshed to form the finite element model for aerodynamic simulation. The number of elements could be extremely large since the control volume is relatively large compared with the airplane. The mesh around the airplane should be very fine because the boundary laryer is created in this region. While the mesh further away from the airplane body can be relatively sparse.
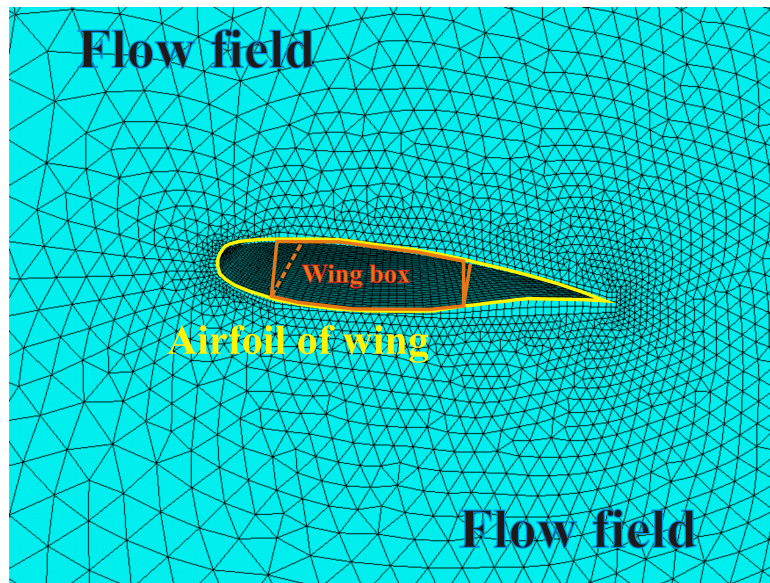

Figure 6.5.: Finite element model (3D) of the aircraft wing for aerodynamic analysis.
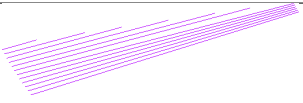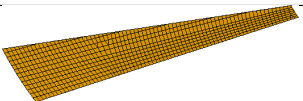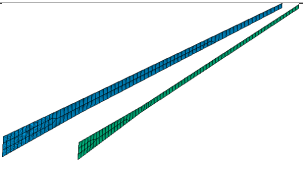
6.2.2. Design Variables
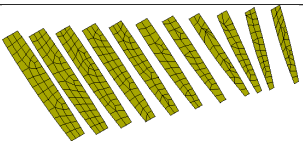
The details of the mesh around the aircraft wing finite element model is shown in Figure 6.5, where the regions of flow field, the wing and wingbox are marked.

The wingbox model contains a total number of 344 design variables, which are the sizing parameters of the stringers, the skins, the spars and the ribs. The parameters are variable spanwisely and chordwisely, and they are also varying according to upper and lower positions for the skin. The parameters are noted as xg for the stringers, xk for the skins, xp for the spars and xr for the ribs, repectively. A list of the components with the defined design variables can be found in Table 6.1.

6.2.3. System Responses

The system responses considered in this dissertation include stresses σ and displacements δ of the wingbox structure under the aeroelastic load, as well as the resonance frequence f of the structure. The system responses are obtained by performing the computational structural mechanical (CSM) analysis with the finite element method, in which the loads

Table 6.1.: Design variables of the aircraft wingbox model.

Components	Design variables	Description
	$xg_1, xg_2, \cdots, xg_{112}$	Thicknesses, widths and heights of the T cross section of stringers
	$xk_1, xk_2, \cdots, xk_{156}$	Thicknesses of skin panels, varing spanwisely and chordwisely
	$xp_1, xp_2, \cdots, xp_{50}$	Thicknesses of the spars, varing along the spanwise direction
	$xr_1, xr_2, \cdots, xr_{26}$	Thicknesses of the ribs

(pressure distribution) on the structure are obtained from CFD simulation. The CSM and CFD analyses are coupled so that the aerodynamic model gets updated deformation from the structural model and transfers the updated pressure distribution to the structure. Both the CSM and CFD are solved based on the FEM method. For the given wingbox structural model, the CSM perfoms a nonlinear static analysis while considering the large-deflection effects. The CFD for the wing model is a three dimensional turbulent compressiable fluid dynamics analysis. Both analyses include a number of iterative processes for solving the system equations and are rather computationally expensive. The process of the coupled analyses is demonstrated in Figure 6.6.

As is discussed in section 6.2.2, the structural model and the aerodynamic model are meshed differently, and thus the positions of nodes are not identical in two models. Therefore, the coupling of two analyses is realized by an interpolation method. The deformation of the nodes in the structural model is interpolated to the nodes in the aerodynamic model, and the pressure distribution in the aerodynamic model is interpolated to the structural model. Both the positions of the nodes on the upper and the lower skins in the structural model and the aerodynamic model are saved to MATLAB as matrices. The pressure distributed on the whole aircraft wing is obtained with the CFD simulation in ANSYS, and the pressure data with respect to the node positions in the aerodynamic model are saved as vectors to MATLAB. The displacements of the nodes in the structural model through CSM analysis are also saved as vectors to MATLAB. Linear interpolation of the pressure and displacement data between the nodes matrices of two models is performed. With the parameterized models and the interpolation between the CSM and CFD analyses, an automatic aeroelastic analysis is possible. By performing such high-fidelity simulations iteratively, the required system responses of the aircraft wing under aeroelastic loads are obtained.
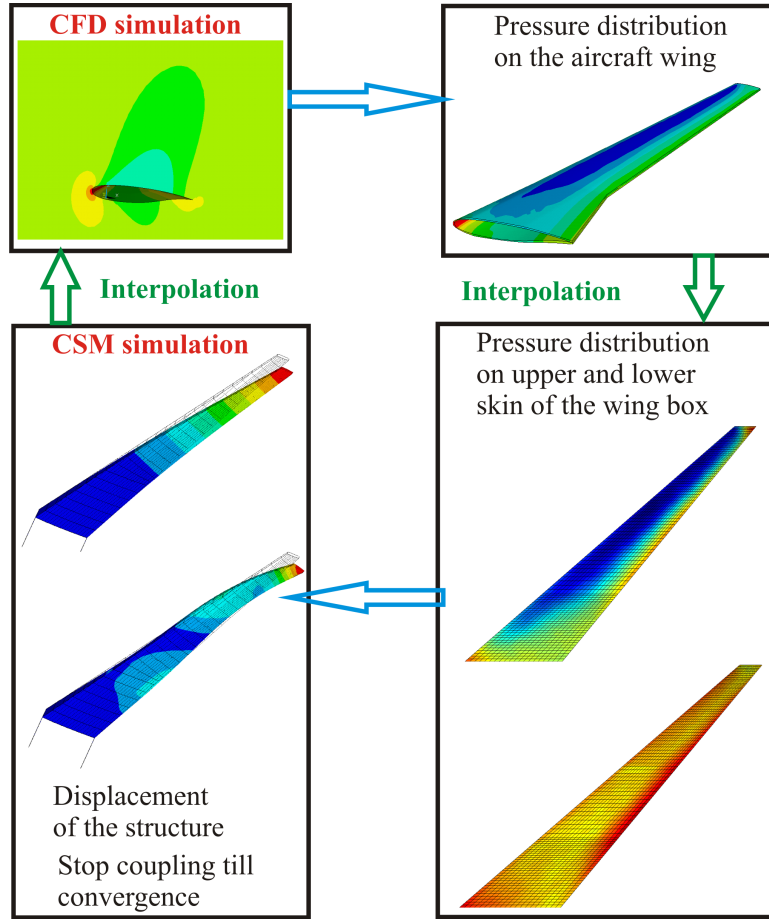
Figure 6.6.: High-fidelity analysis of the aircraft wing under aeroelastic, with coupled CSM and CFD simulations.

6.3. Task Definition

In the design optimization of the aircraft wingbox, the structural mass of the wingbox is to be minimized, while a number of constraints in strength, stiffness and stability have to be satisfied [Niu (1988)]. This optimization problem is mathematically formulated as

$$
\begin{aligned}
\text{minimize} \quad & m(\boldsymbol{x}) \\[2ex]
\text{such that} \quad & \boldsymbol{g}_{\sigma}(\boldsymbol{x}) = \frac{\boldsymbol{\sigma}(\boldsymbol{x})}{\sigma_{\text{cr}}} - 1 \leq 0 \\[2ex]
& \boldsymbol{g}_{\delta}(\boldsymbol{x}) = \frac{\boldsymbol{\delta}(\boldsymbol{x})}{\delta_{\text{cr}}} - 1 \leq 0 \\[2ex]
& g_{f}(\boldsymbol{x}) = 1 - \frac{f(\boldsymbol{x})}{f_{\text{cr}}} \leq 0 \\[2ex]
\text{and} \quad & \boldsymbol{x}_{l} \leq \boldsymbol{x} \leq \boldsymbol{x}_{u}.
\end{aligned}
\tag{6.1}
$$

In Equation 6.1, the $\boldsymbol{g}_\sigma$ represents a vector of constraints related to the stress. The stress constraints requires the maximum stresses on all components of the wingbox to not exceed the critical stress σ_{cr} according to the material strength. Similarly, the $\boldsymbol{g}_\delta$ represents a vector of constraints related to the displacements of the structure. All components of the wingbox should not have a deformation larger than the critical displacement value δ_{cr}. $g_f(\boldsymbol{x})$ refers to the first resonance frequency of the structure, which should not be smaller than a predefined value f_{cr} for stability consideration.

To avoid computational expensive and time-consuming high-fidelity simulations during the optimization process, surrogate-based design optimization is performed. The mathematical formulation of the design optimization using surrogate models replacing the high-fidelity system equations are given as

$$
\begin{aligned}
\text{minimize} \quad & m(\boldsymbol{x}) \\[2mm]
\text{such that} \quad & \boldsymbol{g}_{\hat{\sigma}}(\boldsymbol{x}) = \frac{\hat{\boldsymbol{\sigma}}(\boldsymbol{x})}{\sigma_{\mathrm{cr}}} - 1 \le 0 \\[2mm]
& \boldsymbol{g}_{\hat{\delta}}(\boldsymbol{x}) = \frac{\hat{\boldsymbol{\delta}}(\boldsymbol{x})}{\delta_{\mathrm{cr}}} - 1 \le 0 \\[2mm]
& g_{\hat{f}}(\boldsymbol{x}) = 1 - \frac{\hat{f}(\boldsymbol{x})}{f_{\mathrm{cr}}} \le 0 \\[2mm]
\text{and} \quad & \boldsymbol{x}_l \le \boldsymbol{x} \le \boldsymbol{x}_u.
\end{aligned}
\tag{6.2}
$$

In Equation 6.1 and 6.2, $\sigma_{\mathrm{cr}} = 200\,\mathrm{Mpa}$, $\delta_{\mathrm{cr}} = 2.2\,\mathrm{m}$ and $f_{\mathrm{cr}} = 3\,\mathrm{Hz}$ are constant values refering to the critical constraints. $\hat{\boldsymbol{\sigma}}$, $\hat{\boldsymbol{\delta}}$ and $\hat{f}$ are the surrogate models of $\boldsymbol{\sigma}$, $\boldsymbol{\delta}$ and f. Those surrogate models are to be constructed with the combination of mathematic methods and engingeering knowledge. The task in this chapter is to solve the design optmization problem of the aircraft wingbox with high efficiency and accuracy. The performance of the PASBDO method proposed in this dissertation is to be validated in solving this large-scale engineering problem.

6.4. Parallelized Adaptive Surrogate Based Design Optimization of Aircraft Wingbox under Aeroelastic Loads

High-fidelity analysis as solver of system equations is too much time consuming that the surrogate models of system equations are favored. As is discussed, standard surrogate modeling method requires large number of sample points, and still has difficulties to approximate the system responses of the aircraft wing problem within acceptable accuracy. The use of engineering knowledge combined with a genetic algorithm in Kriging model is designed to deal with the accuracy requirments, as well as to reduce the demanded amount

of sample points in constructing surrogate models. In this section, PASBDO method is used for design optimization of the aircraft wingbox. In each loop of PASBDO, the knowledge-based surrogate models are updated and a complete surrogate-based optimization process is performed. Then the optimization results and the accuracy of the knowledge-based surrogate models are checked and refined adaptively with a couple of infill points. Finally, it is possible to solve the aircraft wingbox design optimization problems with good efficiency and high approximation accuracy.

6.4.1. Knowledge-based Surrogate Modeling

As is disscussed in Chapter 3 and Chapter 4, the first and most important step in knowledge-based surrogate modeling is to analyse and decompose the structural system. By doing this, engineering knowledge-based terms can be generated and the dimension of the surrogate modeling problem can be reduced. To do this, a cross section of the wingbox at a specific y-coordinate is represented in Figure 6.7. As is known from mechanic engineering theory
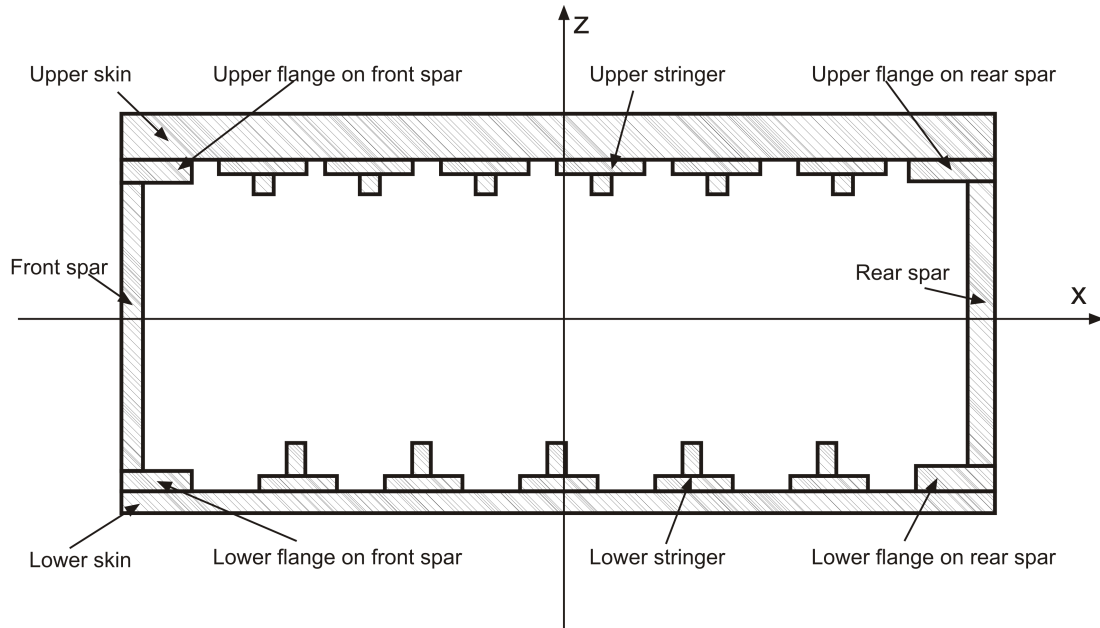

Figure 6.7.: Cross section of the wingbox at a specific y-coordinate.

for beams and shells described in Section 3.3, the cross-sectional properties representing the axial, bending, shear and torsional stiffnesses of the structure are the knowledge-based terms we are looking for. Those terms are able to act as the bridge between design variables and the system responses so as to reduce the dimension of modeling space and nonlinearity of the relations to be approximated. Those properties of the cross section in Figure 6.7 can be derived explicitly with respect to the design variables and form the engineering knowledge-based terms. Engineering knowledge and experience of the aircraft wingbox is then used to choose proper modeling space and modeling responses for surrogate modeling. This is explained as follows:

- For modeling space, the geometric properties of the cross sections of 26 segments along the spanwise direction are derived. Those include the cross-sectional area A, the moment of inertia I_{xx} and I_{zz}, the rotational rigidity J_{yy}, as well as the thickness to surface area ratio $\frac{t}{A_s}$ of the shells. For the term of thickness to surface area ratio $\frac{t}{A_s}$, only the thickness t needs to be considered since the surface area A_s is not changing with respect to the design variables.

- For the modeling responses, instead of stress $\boldsymbol{\sigma}$ and displacement $\boldsymbol{\delta}$, their inverses $\frac{1}{\sigma}$ and $\frac{1}{\delta}$ are approximated. This is because that the system responses $\boldsymbol{\sigma}$ and $\boldsymbol{\delta}$ are inversely related to the design variables.

After the knowledge-based terms are formed, the regression model of the Kriging model is to be constructed. To formulate the regression model, either the traditional quadratic regression functions can be used, or the genetic algorithm can be used to search for the best regression functions. The functions are selcted from a number of candidate formats, which is given in Chapter 4. With either method, it is found that using all the 26 spanwise sections can generate ill-conditional regression functions. The reason for the ill-condition is that the knowledge-based terms of a section are too close with those of the neighboring sections. As a result, nearly identical columns are formed in a matrices describing the structure of the regression functions. While performing Cholesky factorization or inverse of these matrices, ill condition is a problem that prevent the further proceeding of the construction process. Even when the numerical tolerance is set very large to allow the proceeding of the calculation, the ill-conditioned matrices can affect the approximation accuracy. By studying the conditional values of the matrices, which contain the geometric properties of the 26 sections for knowledge-based surrogate modeling, it is suggested that neighbor sections be grouped and share the same design variables. Therefore, the wingbox model is spanwisely grouped into five segments, as shown in Figure 6.8.
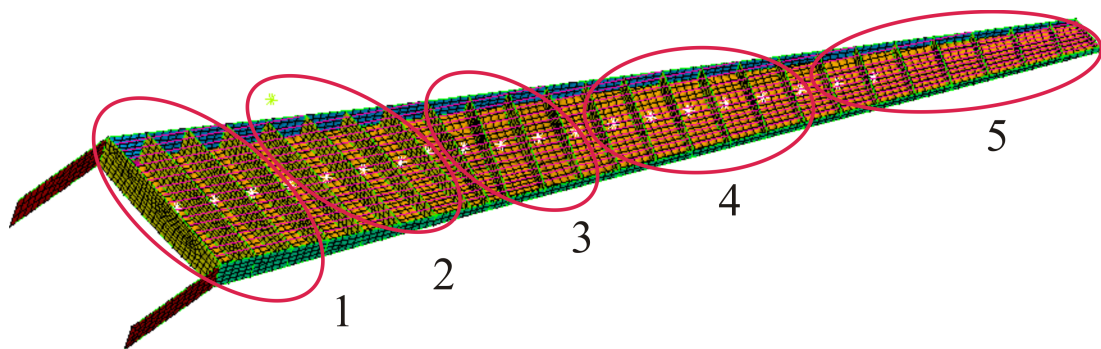
Figure 6.8.: Grouping of spanwise sections in to five segments, which contain four, four, four, six and eight sections, respectively.

The 26 spanwisely distributed sections are so grouped that the first four sections forms the first group, and then the next four, four, six and eight sections form the other four sections. In total, there are five segments (see Figure 6.8), which are studied for knowledge-based surrogate modeling. There are more sections in the segments near the tip than those

near the root. This is because that the regions near the root are generally more critical than the tip regions, and, therefore, they should contain fewer sections and should be studied in more details. Within one segment, the design variables do not change spanwisely, which are, however, still chordwisely variable. Therefore, the total number of design variables reduces from 344 to 74. Knowledge-based terms to form the modeling space of surrogate modeling are generated for the five segments as follows:

- average geometrical properties for the cross sections in the five segments, A_i, $I_{xx,i}$, $I_{zz,i}$, $J_{yy,i}$, $i = 1, 2, .., 5$,

- thicknesses of the ribs in the five segments, t_i, $i = 1, 2, .., 5$.

As a result, there are only 25 knowledge-based terms. Surrogate modeling is then constructed with those knowledge-based terms instead of 74 design variables. Therefore, the modeling space is greatly reduced according to the engineering knowledge of the system. Latin hypercube sampling method is used for design of experiments. 360 sample points are generated for surrogate model construction, which means 360 runs of the high-fidelity simulations are performed. Now that the modeling space and modeling responses are decided, and sample points are generated, the regression functions can be searched with GA. The initial knowledge-based surrogate models are constructed with Kriging, which, however, should be refined adaptively in order to asure the accuracy of the optimization results.

6.4.2. Parallelized Adaptive Surrogate Based Design Optimization and The Results

After the initial knowledge-based surrogate models are determined, the PASBDO is used to refine the surrogate models with the help of multiple infill criteria and parallel computing. With the loops of PASBDO, the quality of surrogate models for system responses are gradually improved. Eventually, by using both GA for global search and SQP with multiple starting points for local search on the surrogate models, the optimization results can be obtained very efficiently. The procedure of PASBDO is introduced in Chapter 5, and details of the settings of PASBDO for the aircraft wingbox design problem are listed as follows:

- 360 initial sample points are generated by LHS method for the construction of initial surrogate models.

- Both the heuristic optimization method GA and the gradient and Hessian based optimization method SQP are used to search for the optimal design on the knowledge-based surrogate models.

- The total number of optimization processes in one loop of PASBDO is set to 20, which include one GA and 19 SQP. The number 20 is selected according to the dimension of modeling space and the available CPU resourses on the cluster. Setting larger number of parallelized processes is recommended as long as enough CPU resourses are available.

- The 20 different starting points are seleted from the sample points, which have the best objective function values considering their feasibility. Since the starting points are chosen from sample points generated with LHS, they are also well distributed in the design space.

- The minimum allowable distance between sample points is set to 0.1. This means if the distance between two points is smaller than 0.1, then one of them is considered to be redundant and is deleted. The remained infill points are evaluated with the high-fidelity models using parallelized computing technique.

- The maximum number of loops of PASBDO is set to ten as a stopping criterion.

- The tolerance of the change in design variables, objective function, constraint functions between two loops and the maximum allowable constaint violation are set to 10^{-2}, which are the convergence criteria in addition to the stopping criterion.

After two loops of PASBDO, a design is already converged, which satisfies the optimality conditions defined in Equation 5.22 and 5.23 for SBDO. The final design for the components of the wingbox is shown in Figure 6.9. The varying of thicknesses of the wingbox components along spanwise segments and the stringers along the chordwise direction is demonstrated. The results in Figure 6.9 are interpreted as follows.

- The left subgraph of Figure 6.9 demonstrates the optimized thicknesses of components in the five segments along the spanwise direction. Those components are the front spar, the rear spar and the ribs. The observations are listed below.

    - For both the front and rear spars, the inboard segments (No. 1 and 2) are thicker than the outboard segments (No. 4 and 5). This is clear because the inboard segments have to withstand greater stresses than the outboard segments.

    - For both spars and ribs, the middle segments (No. 3) have the largest thickness. As can be seen in Figure 6.8, the third segment is located at the position where the orientation of the ribs changes. This results in larger distances between ribs and longer spar components than those of the inner segments (No. 1 and 2). As longer shells are more susceptible to local buckling that it is reasonable to have larger thicknesses there.

    - Generally, the ribs have much smaller thickness than the spars, except for the the ribs at the tip segments, where both are very thin. The function of ribs is to transfer loads between upper and lower skins and spars. The ribs are mainly under shear stress and can be thinner, as is also suggested by the results. Usually the ribs are designed with lightening holes [Niu (1988)].

- In the right-top graph of Figure 6.9, the optimized thicknesses of wing skins are shown. More specifically, they are the upper and the lower wing skin close to the front spar along the five spanwise segments. For them the observations are also listed in the following:
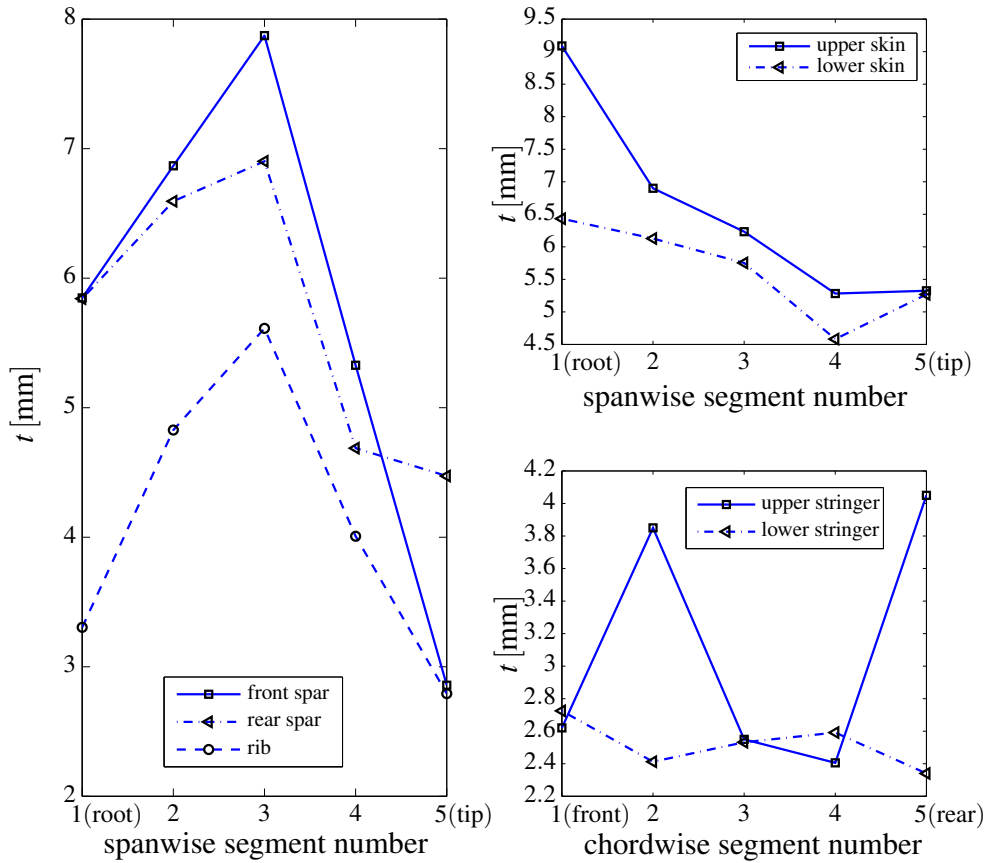
Figure 6.9.: Results of the design variables obtained by PASBDO for the aircraft wingbox design optimization problem [Xu et al. (2013)].

  – The thicknesses of both the upper and lower skins are generally decreasing from the root to the tip. This is obvious since no matter under bending or torsion or both, the root segments have to sustain higher stresses than the tip segments.

  – The upper skin elements have larger thicknesses than the lower ones. This is because of the lift on the wing that the upper skin is under compression while the lower skin is under extension.

• The thicknesses of the flanges of the stringers along the chordwise direction is shown in the right bottom graph of Figure 6.9. No clear trend is seen chordwisely for the stringers and the thickness values are much smaller than components shown in the other two subgraphs. This is because of the using of knowledge-based terms, which calculate the cross-sectional properties. Since each stringer contributes a much smaller portion compared with the other components, the surrogate models are unable to approximate the effects of those design variables accurately. A detailed design of stringers with variable sizing parameters or even shapes and materials can performed seperately following the designs obtained in this chapter.

Performances of the PASBDO method for the aircraft wingbox design problem can be seen in Table 6.2, which contains information in each loop of PASBDO. The initial model in

Table 6.2.: Information in each loop of PASBDO for the aircraft wingbox design problem [Xu et al. (2013)].

	n_{samp}	$\max(\epsilon_{\text{rrms}})$	n_{feas}	m_{opt}	g_{max}
Initial model	360	3.25%	0	1347.6	0.3183
1st loop model	20 infill points	2.08%	2	1318.5	-0.0302
2nd loop model	20 infill points	1.99%	15	1302.6	-0.0172

the table means the initial surrogate models constructed by GA assisted knowledge-based surrogate modeling method with 360 sample points. The 1st and 2nd loop models refer to the knowledge-based surrogate models refined by infill points recognized during the loops of PASBDO. Only two loops of PASBDO is necessarily executed before a design is converged. $\max(\epsilon_{\text{rrms}})$ is the maximum relative root mean square error of the approximation, which is cross validated in the process of PASBDO. n_{feas} is the number of feasible designs obtained from the 20 optimization processes with different starting points. g_{max} is the maximum constraint violation value, of which only nonpositive value indicates feasible designs.

It can be seen that the initial knowledge-based surrogate models have good accuracy ($\max(\epsilon_{\text{rrms}}) = 3.25\% < 5\%$), which are then improved in two loops, each generates 20 infill points. However, it should be noted that no feasible design can be guaranteed by optimizing on the initial surrogate models. For this large dimensional design problem, although the accuracy of initial approximations obtained with knowledge-based surrogate modeling are not bad, it is not good enough for optimization. This is because that in optimization, focus should be put on the critical regions and the approximation quality in such regions should be very high. In the searching process of SQP, it is quite possible that the search goes trapped in unsampled regions of surrogate models. Those regions might containing local minima of the surrogate models but are actually infeasible points when checked with high-fidelity system equations. Therefore, it is very important to generate infill points in PASBDO, so that those regions are validated and clarified. Afterwords, better and feasible design regions can be identified. Finally, it is proper to use the refined surrogate models as system equations for design optimization.

6.5. Discussion of Computational Effort

The main purpose for developing the PASBDO method is to reduce the computational effort in design optimization for problems with computationally expensive system equations. The aircraft wingbox design optimization problem is one of the typical large-scale problems for validating PASBDO. To demonstrate the performance of the PASBDO, it is necessary to discuss the computational effort of other methods, which are listed as follows:

- Gradient-based optimization algorithm with finite differencing

The aircraft wingbox design optimization problem has large dimensionality, which contains 74 design variables after the grouping of sections. Let n_{DV} be the number of design vaiables. The calculation of gradients with finite differencing requires $n_{DV} + 1$ evaluations of the system equations in each iteration of the optimization. The number of evaluations can be even higher when the calculation of Hessian matrix are necessary. Therefore, even just for e.g., six iterations, which is a typical small number for the convergence of gradient-based algorithms, 450 system evaluations are needed. Normally much more iterations are used for searching in high-dimensional design space.

- Gradient-based optimization algorithm with analytical gradient calculation
  This is one of the most efficient optimization method, which is however, not possible for the complicate CSM and CFD system simulations of the aircraft wing.

- Gradient-based optimization algorithm with semianalytical gradients
  This is a very promising method that does not only reduce computational effort, but also maintain good accuracy in system evaluations. This method, however, requires to explicitly export the global stiffness matrix and the aerodynamic influence coefficients [Wright and Cooper (2007)] of the aircraft wing system. Since such requirments can not be fulfilled by ANSYS, other FEM codes should be considered. However, this needs further time and effort for programming and validation. As a result, the performance of this technique is not possible to be shown for comparison in this dissertation, but it is highly recommanded.

- Evolutionary algorithms
  Such methods require either a large population or a large number of iterations (generations) for the searching of high-dimensional design space. A large population means a great number of system evaluations and large number of iterations will also lead to soaring computational efforts. For example, for $n_{DV} = 74$, a reasonable population size would be at least 300, which means even only two generations of searching would result in 600 system evaluations.

- Standard surrogate modeling based design optimization
  This method has to face the difficulty of DOE for large dimensionality. To ensure good distribution and well coverage of the design space, the number of sample points should be reasonably large. Besides, if using a standard quadratic polynomial model, ≥ 2850 sample points are required for least-square determination of the regression coefficients. To reach good local approximation quality at critical design domains, even more sample points are necessary. Further, the approximation quality can not be guaranteed with larger sample size, since the formats of the surrogate models are even more important factors.

Having discussed the computational efforts required in these methods, the advantages of the PASBDO method over most of them is clear. Reasons for the good performance of the PASBDO method in the aircraft wingbox design problem are explained in details.

First of all, the contribution for good efficiency and accuracy comes from the knowledge-based surrogate modeling. It reduced the dimension of the modeling space to from 74 to 25,

which greatly relieves the pressure in DOE. The 25 knowledge-based terms represent the mechanical properties of the structure, which make it possible to approximate the modeling responses with simpler formulations of regression functions. Therefore, it requires fewer sample points to determine regression coefficients of the surrogate models and it can achieve better approximation precision.

The use of estimated root mean square error and constrained expected improvement as infill criteria is also a reason for the good performance of PASBDO. The multiple searching based on these infill criteria provided the desired infill points, which are either in the regions close to the optima or in the unexplored regions of the design space. In other words, this strategy is capable of identifying locations on current surrogate models, which are most necessarily to be improved. It refines the surrogate models with very small effort, i.e. only a couple of infill points are evaluated in each loop of PASBDO.

A number of surrogate-based design optimization processes are executed with different starting points. This is beneficial for three reasons. First, the starting points are a number of best sample points. By starting from those positions, it has higher chance to discover the optima or at least good designs than starting from the boundary of design space or a random position. Second, the starting points are well distributed in the design space, so that it can reduce the chance of trapping in a local optimum. Last but not least, a number of optimization processes with different starting points are performed on surrogate models, which are computationally very efficient. It would take too much effort to do this on the computationally expensive system equations.

With all the above strategies, the required number of system evaluations to solve the optimization problem is only 400. Besides, the parallelized computing technique can further reduce the time expense. The use of parallelization provides a speed-up ratio of $\mathcal{P}_p = 18.2$ for the aircraft wingbox design optimization problem.

6.6. Summary of Chapter

Design optimization of an aircraft wingbox under aeroelastic loads is introduced and solved with PASBDO in this chapter. The high-fidelity evaluations of the system equations requires the coupling of CSM and CFD simulations, which are computationally expensive. The PAS-BDO method is used to accelerate the optimization process. PASBDO starts with initial surrogate models, which are constructed with the combination of mathmetical tools and mechanical engineering knowledge. Infill criteria are used to varify the quality of surrogate models and make refinements at the most promising positions. Multiple surrogate-based optimization processes are implemented in each loop of PASBDO to enhance the ability of identifying good designs. The use of parallelization technique further increases the efficiency in solving the wingbox optimization problem. In conclusion, the PASBDO method has shown its capability in solving this large-scale structural design optimization problem. It is demonstrated that by using the PASBDO, high efficiency in optimization and good approximation accuracy in surrogate models can be achieved.

7. Parallelized Adaptive Surrogate Based Design Optimization of an Automobile Front Crash System Under Impact Load

Crashworthiness simulations are used in the automobile industry to analyse the behavior of a vehicle or its components under impact loads, so as to assess the safety of the vehicle and guide the structural design. A high-fidelity crashworthiness simulation is usually extremely time consuming since a large number of finite elements and discrete time intervals are involved [Bois (2011)]. Therefore, design optimization, which usually requires enormous evaluations of the system equations are hindered. In this chapter, the PASBDO method is applied to the design optimization problem of an automobile front crash system. It is shown that the PASBDO method is a robust method, which is capable of sovling the optimization problem with high accuracy and efficiency.

7.1. The Front Crash System of Electric Car MUTE

MUTE is the name of a two-seat energy-efficient city car developed at the Technische Universität München (TUM). The project was carried out through the cooperation of 20 departments of TUM, of which the department LLB was responsible for structural design of the space frame. The structural design optimization of the MUTE space frame is carried out in LLB by Erich Wehrle and other colleague, more introduction of which can be found in Wehrle et al. (2012) and Fuchs et al. (2012). The design is aimed at structural weight reduction, while maintaining the crashworthiness requirements in structual strength, stiffness and stability, as well as costs and other industry standards, as one can found in Wehrle (2013). The study of the MUTE front crash system is an important part of the space-frame design, as shown in Figure 7.1. In this dissertation, the application of PASBDO method is performed on the MUTE front crash system design problem, which contains computationally expensive system equations and nonsmooth system responses. Sizing and shape design optimization of the front crash system under impact loads is performed with the assistance of the surrogate modeling techniques to ease the burden from time-consuming crash simulations.
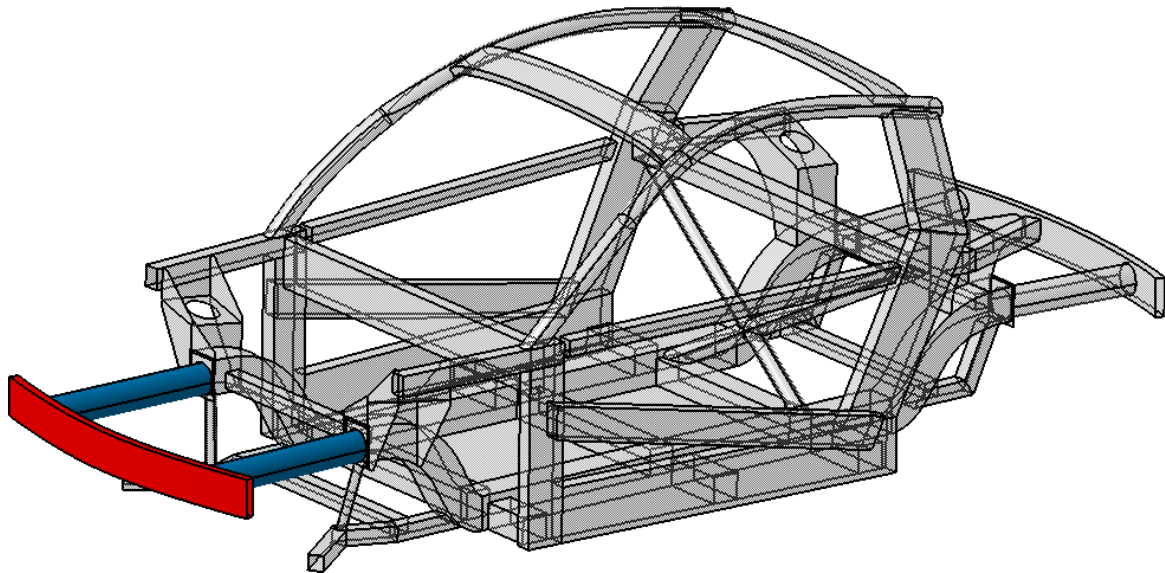
Figure 7.1.: MUTE with the front crash system to be optimized.

7.1.1. Simplification in Model Construction of the Front Crash System

In design optmization of structures under crash simulation, it is beneficial to focus on the main load bearing elements, so that the model size is smaller and the simulation time is smaller. The constructed simulation model of the front crash system is, therefore, simplified. The simplified front crash system is composed of a bumper and two crash tubes, as shown in Figure 7.2.
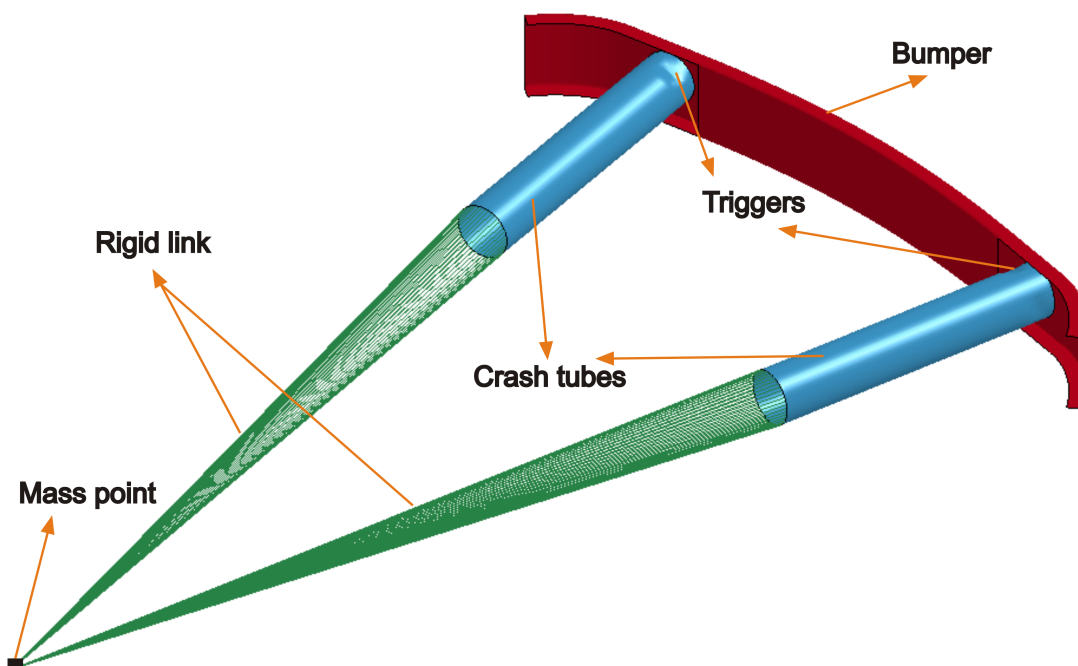

Figure 7.2.: Simplified configuration of the front crash system.

A mass point and rigid links between the mass point and the crash tubes represent the simplified other parts of the vehicle.

The bumper has a smooth surface, extended from a rectangular cross section along a designer-controlled curve. The leading edge of the bumper is the part of the system that touches crash barrieres. The other side of the bumper has two open slots on both left and right edges to avoid unnecessary weight. The left and right crash tubes are connected with the leading edge of the bumper to allow direct pathing routes for the impact loads. There are triggers at the front of both tubes to activate the expected folding mode of the tubes when crash starts. In the front crash system, the structure of the whole vehicle is simplified into a mass point, which has the mass and inertia of the complete vehicle. Since the mass of the bumper and tubes only share around 1% of the vehicle, the mass of the complete vehicle is assumed to be constant during design optimization of the front crash system for simplification. The total mass m_c is 0.7 Kg. The inertial properties assigned to the mass point is calculated as follows by assuming the vehicle to be a solid cubic.

$$
\begin{aligned}
I_{xx} &= \frac{1}{12} m_c (w^2 + d^2) \\
I_{yy} &= \frac{1}{12} m_c (l^2 + d^2) \\
I_{zz} &= \frac{1}{12} m_c (w^2 + l^2) \\
I_{xy} &= I_{yz} = I_{xz} = 0
\end{aligned}
\tag{7.1}
$$

where w, d, l are the width, depth, and length of the vehicle, respectively. The rigid links between the mass point and the tubes represent the connection between the other parts of the vehicle and the front crash system. The crash tubes and bumper are modeled as thin-wall shell elements. With such construction, the model of the front crash system is simple enough to allow reasonably fast crash simulation, while still being a complete entity to represent the feature of the front-crash of the vehicle.

7.1.2. Design Variables

There are in total seven design variables in the design optimization of the front crash system, as shown in Figure 7.3. The meaning and domain of all design variables are given as follows:

$$
\begin{aligned}
x_1 &: \text{angle of the crash tubes,} & [0, 10]\,(^\circ) \\
x_2 &: \text{radius of the crash tubes,} & [35, 75]\,(mm) \\
x_3 &: \text{thickness of the crash tubes,} & [1, 4]\,(mm) \\
x_4 &: \text{offset of the bumper center line,} & [-50, 100]\,(mm) \\
x_5 &: \text{height of the bumper,} & [30 + 2x_2, 50 + 2x_2]\,(mm) \\
x_6 &: \text{width of the bumper,} & [30, 75]\,(mm) \\
x_7 &: \text{thickness of the bumper,} & [1, 4]\,(mm)
\end{aligned}
\tag{7.2}
$$

In Figure 7.3, examples of different settings of design variables are demonstrated. The bottom-left graph shows that when $x_1 = 0$ and $x_4 = -20$, the tubes of the front crash system are parallel with the forward driving direction, and the bumper has a curve going towards
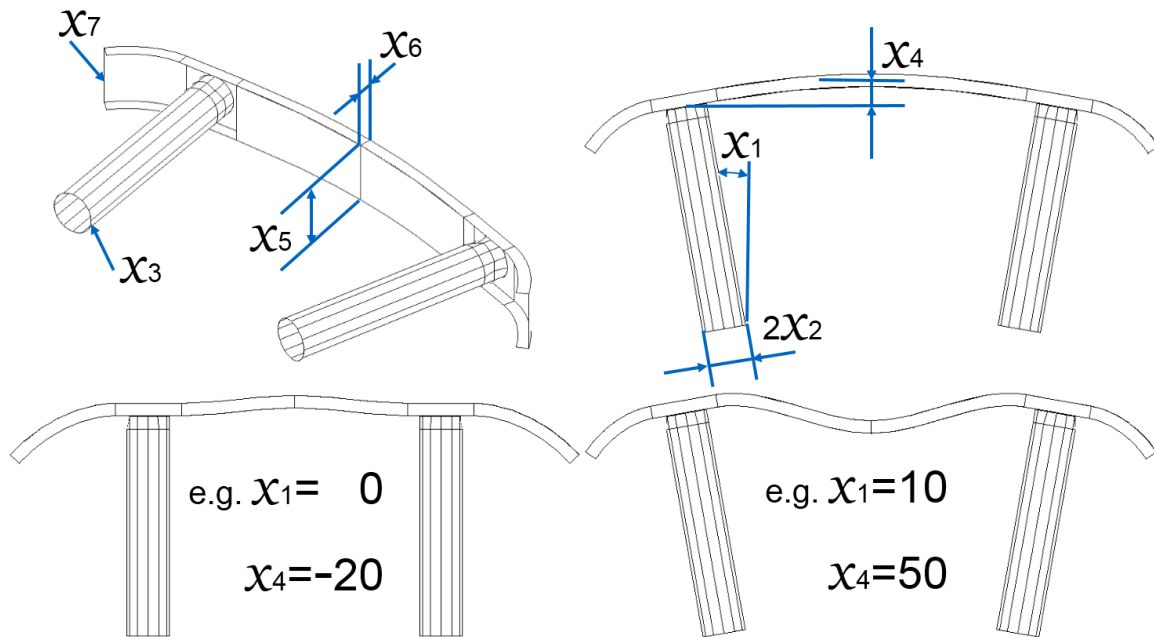
Figure 7.3.: Design variables of the front crash system.

the forward direction forming a C-shape curve. Such shape is beneficial when a front complete crash happens, as shown in Figure 7.4, because the tubes have the direct path to take the impact loads. It is shown in the bottom-right graph of Figure 7.3 that, when $x_1 = 10$ and $x_4 = -20$, the tubes are no more parallel-located and the center of the bumper goes backwords, forming a M-shape curve. Such design is benifical when a front side crash happens, as shown in Figure 7.5, because the crashed and uncrashed sides are isolated to prevent extensive damage. The seven design variables control the shape and size of the front crash system together. A best setting of the design variables is to be searched in order to minimize the structural mass, while satisfying all design requirements.

7.1.3. Load cases

Two load cases are analyzed to study the front crash behavior of the system. One is according to the United States new car assessment program (US NCAP) and the other is according to the european new car assessment program(EU NCAP). In both cases, the out of plane deformation of the mass point is constrained in crash simulation.

- US NCAP: As is shown in Figure 7.4, according to US NCAP, a full-width frontal crash simulation is performed. The vehicle goes forward with a velocity of 56 km/h and the whole head crashes into a rigid wall.
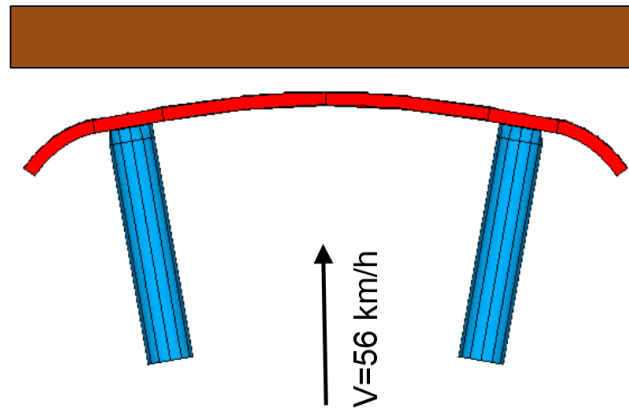
Figure 7.4.: Crash simulation of MUTE front crash system under US NCAP.

- EU NCAP: Based on the EU NCAP, a frontal offset impact crash simulation is carried out, as is show in Figure 7.5. 40% of the head crashes into the rigid wall, while the vehicle is going forward with 64 km/h velocity. In this load case, it is assumed that only half of the kinetic energy is to be absorded by the front crash system, and, therefore, the simulated velocity of the structure is 45 km/h.
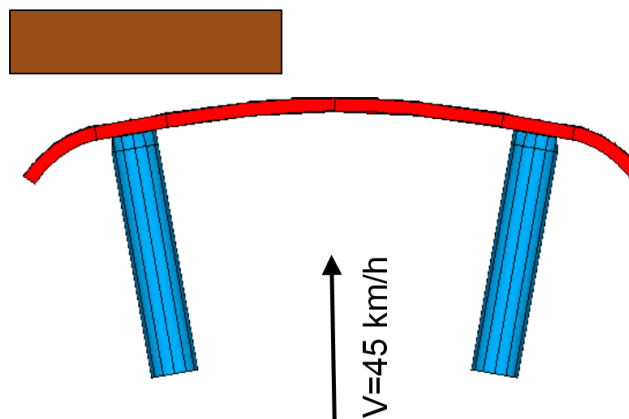

Figure 7.5.: Crash simulation of MUTE front crash system under EU NCAP.

7.1.4. Material Properties

The material of the front crash system is extrusion aluminium alloy, EN AW-6060, the material data of which are based on Zarei (2008). Those material properties of EN AW-6060 are listed in Table 7.1, which are the density ρ, the elastic modulus E, the poisson ratio v and the yield stress σ_y. Since elasto-plastic deformation of the structure is involved in crash process, the stress-strain relation of the material is given, which is demonstrated in Figure 7.6.

7.1.5. System Responses

There are in totol six system responses, which are used to form the objective and constraint functions of the front crash system design optmization problem. The first system response

Table 7.1.: The material properties of extrusion aluminium alloy, EN AW-6060.

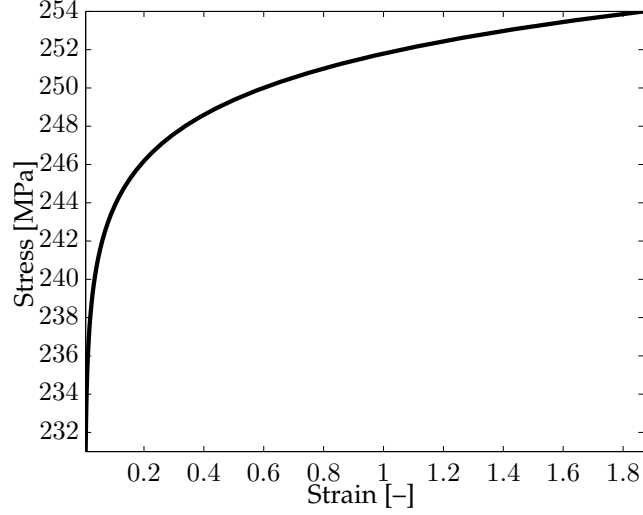| Density	Elastic modulus	Poisson ratio	Yield stress
2.63e3 kg/m^3	70000 MPa	0.33	231 MPa

Figure 7.6.: The plastic stress-strain relationship of the extrusion aluminium alloy, EN AW-6060 for the front crash system.

is the mass of the front crash system m, which is the objective function. Once the design variables are given, m can be calculated analytically, and, therefore, no computationally expensive simulations are necessary for m. The other five system responses are calculated to form constaint functions, which are introduced in the following. After the model is constructed, and load cases are defined, explicit finite element analyses are carried out for the crash simulations of the front crash system. The crash responses are obtained, which include the compression length L_{cr} of the crash tubes and the maximum resulted section force on the end of the crash tube F_{sec} during crash simulation, as demonstrated in Figure 7.7. The two load cases have the same definition of system responses, but different values, noted as L_{cr}^{US}, F_{sec}^{US}, L_{cr}^{EU}, F_{sec}^{EU}, respectively.

The crash process involves the folding of the crash tubes, which is a progressively plastic deformation procedure. The propagation of energy and force in the tubes produces oscillated section forces. The section forces recorded in the crash simulation process contain noise, and, therefore, a Butterworth filter [Butterworth (1930)] is used to smooth the obtained data, which is implemented by Fellner and Wehrle (2013). In Figure 7.8, an example of the unfiltered and filtered section forces in the simulation process is demonstrated. The maximum section force is then obtained from the filtered data, and applied as a crash safty constraint.

$L_{cr} = L_1 - L_2 = U_1 - U_2$: crash length

F_{sec} : resulted section force
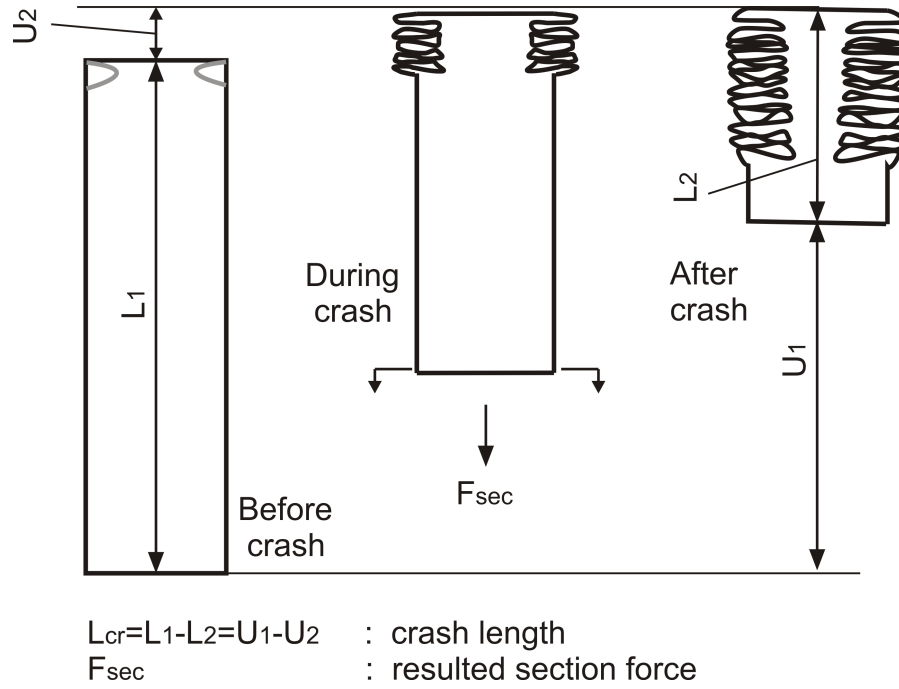
Figure 7.7.: System crash responses: compression length and section force of the crash tubes.
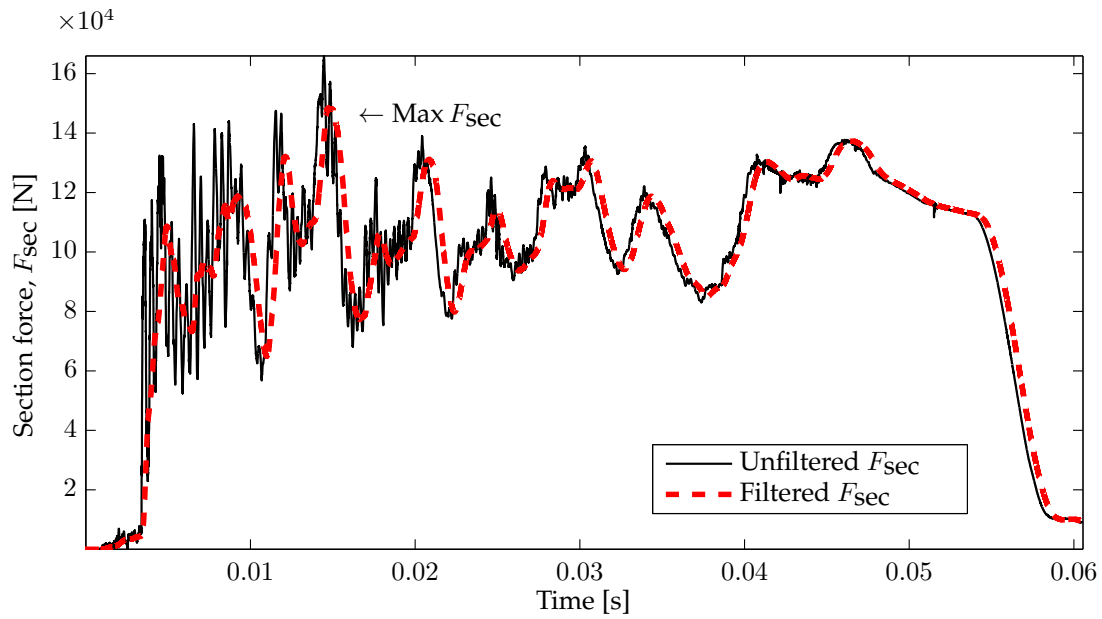
Figure 7.8.: An example of the unfiltered and filtered section forces in the crash simulation.

The last system response is the thickness to radius ratio $\dfrac{x_3}{x_2}$ of the crash tubes. This ratio is a very important geometrical contraint. The critical plastic buckling stress σ_{cr} is depend on the thickness to radius ratio of the crash tubes. The critical thickness to radius ratio γ_{cr} seperates the elastic and plastic buckling of the crash tube [Abramowicz and Jones (1997)], which is, therefore, derived by

$$\sigma_{cr} = \sigma_y, \tag{7.3}$$

where

$$\sigma_{\text{cr}} = \frac{n^2 \gamma_{\text{cr}}^2 E}{12(1 - v^2)}.$$ (7.4)

In Equation 7.4, n is a constant, which equals to the number of triggers on each crash tube. Physical experiments for the folding modes of the crash tubes under quasi-static compression loads are carried out in LLB to test this design concept. One test sample is shown in Figure 7.9. In the experiments, concerns are the folding modes affected by the position, the amount and the geometrical parameters of the triggers on the crash tubes, etc. Acoording to the research and experiment results from Wehrle (2013)], the number of triggers is chosen, $n = 3$, which shows robust effects in triggering the same folding mode for energy absorbing. To sum up, the six system responses are listed as follows:
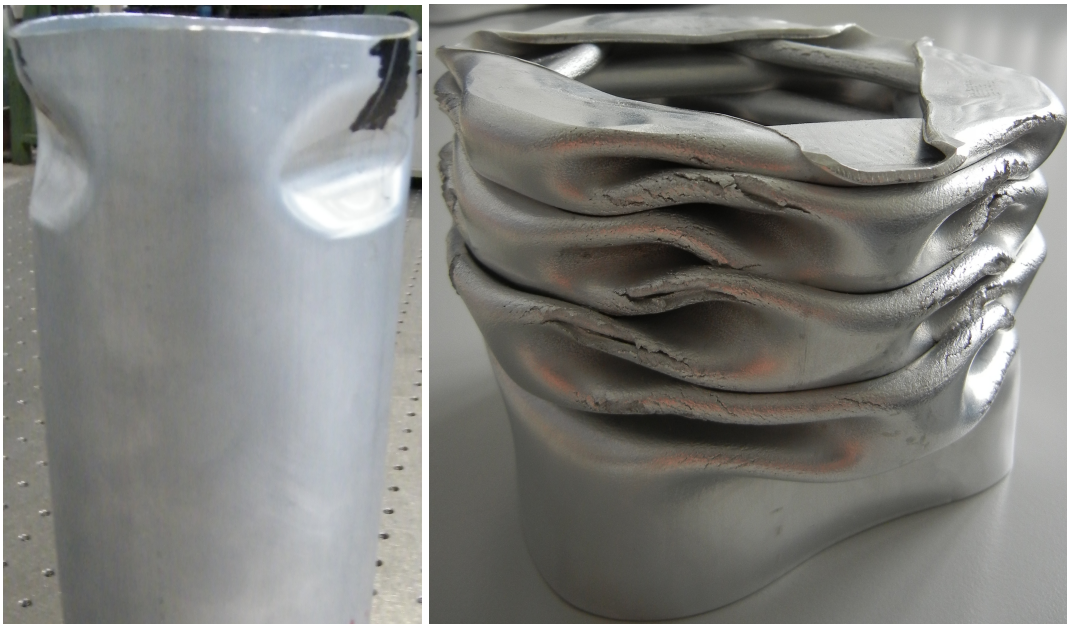
Figure 7.9.: An example of the physical experiments for the folding modes of the crash tubes under quasi-static compression loads [Wehrle (2013)].

m: mass of the front crash system (crash tubes and bumper),

L_{cr}^{US}: compression length in the US NCAP load case,

F_{sec}^{US}: maximum section force on crash tubes during crash in the US NCAP load case,

L_{cr}^{EU}: compression length in the EU NCAP load case,

F_{sec}^{EU}: maximum section force on crash tubes during crash in the EU NCAP load case,

$\frac{x_3}{x_2}$: thickness to radius ratio of the crash tubes.

7.2. Task Definition

The design optimization of the front crash system of MUTE under impact loads requires the structural mass to be minimized, while all constraints to be satisfied. The constraints are discribed as follows:

- compression length in both the US NCAP and EU NCAP load cases should not exceed 450 mm,

- resulted maximum section forces on crash tubes during crash in both load cases should not exceed 150 KN,

- thickness to radius ratio of the crash tubes should be so designed, that the critical plastic buckling stress does not exceed the yield stress.

The mathematical formulation of the front crash system design problem is given in Equation 7.5.

$$\text{minimize} \quad m(\boldsymbol{x})$$

$$\text{such that} \quad g_1(\boldsymbol{x}) = \frac{L_{\text{cr}}^{US}}{450} - 1 \le 0$$

$$g_2(\boldsymbol{x}) = \frac{F_{\text{sec}}^{US}}{1.5 \times 10^5} - 1 \le 0$$

$$g_3(\boldsymbol{x}) = \frac{L_{\text{cr}}^{EU}}{450} - 1 \le 0 \tag{7.5}$$

$$g_4(\boldsymbol{x}) = \frac{F_{\text{sec}}^{EU}}{1.5 \times 10^5} - 1 \le 0$$

$$g_5(\boldsymbol{x}) = \frac{n^2 \gamma_{\text{cr}}^2 E}{12(1 - v^2)\sigma_y} - 1 \le 0$$

$$\text{and} \quad \boldsymbol{x}_l \le \boldsymbol{x} \le \boldsymbol{x}_u, \ \boldsymbol{x} = [x_1, x_2, \cdots, x_7].$$

To avoid computationally expensive and time-consuming crash simulations during the optimization process, surrogate models are constructed for the system responses. The mathe-

matical formulation of the design optimization using surrogate models are given as:

$$\text{minimize} \quad m(\boldsymbol{x})$$

$$\text{such that} \quad \hat{g}_1(\boldsymbol{x}) = \frac{\hat{L}_{\text{cr}}^{US}}{450} - 1 \leq 0$$

$$\hat{g}_2(\boldsymbol{x}) = \frac{\hat{F}_{\text{sec}}^{US}}{1.5 \times 10^5} - 1 \leq 0$$

$$\hat{g}_3(\boldsymbol{x}) = \frac{\hat{L}_{\text{cr}}^{EU}}{450} - 1 \leq 0 \tag{7.6}$$

$$\hat{g}_4(\boldsymbol{x}) = \frac{\hat{F}_{\text{sec}}^{EU}}{1.5 \times 10^5} - 1 \leq 0$$

$$g_5(\boldsymbol{x}) = \frac{n^2 \gamma_{\text{cr}}^2 E}{12(1 - v^2)\sigma_y} - 1 \leq 0$$

$$\text{and} \quad \boldsymbol{x}_l \leq \boldsymbol{x} \leq \boldsymbol{x}_u, \ \boldsymbol{x} = [x_1, x_2, \cdots, x_7].$$

Comparing Equation 7.5 and 7.6, it is seen that the only difference between them is the use of $\hat{L}_{\text{cr}}^{US}$, $\hat{F}_{\text{sec}}^{US}$, $\hat{L}_{\text{cr}}^{US}$ and $\hat{F}_{\text{sec}}^{US}$ in Equation 7.6. Those are the surrogate models for L_{cr}^{US}, F_{sec}^{US}, L_{cr}^{US} and F_{sec}^{US}, respectively. The task in this chapter is to solve the design optmization problem of the front crash system with the assistant of surrogate models. High efficiency and accuracy is expected with the PASBDO method.

7.3. Parallelized Adaptive Surrogate Based Design Optimization of the MUTE Front Crash System

The crash simulation of each load case for the MUTE front crash system costs around 30 minites on a CPU with current average performance. Since two load cases are considered in this dissertation, roughly one hour is required in total for each evaluation of the system equation.

The buckling mode of thin-walled column is influenced by many factors. Besides the thickness to radius ratio and the radius to length ratio of the column, other factors such as the smallest element size in the structure, the use of mass scaling in crash simulation and the simulation time step, etc., also play important roles. The setting for those parameters are tested and determined in LLB by Tischer and Wehrle (2012). Generally, the crash simulation has a rather unstable nature. Slightly change in design variables can result in the change of folding mode, and further affect the smoothness of the system responses. Parameter studies of the front crash system performed in Binder and Wehrle (2012) and Zhang and Wehrle (2013) show oscillating behavior in both the compression length and the section force with respect to the thickness and radius of the crash tubes.

The oscillating behavior causes difficulties in optimization. As is shown in the works of Binder and Wehrle (2012) and Zhang and Wehrle (2013) that convergence is hard to achieve during optimization with the crash simulation models. Besides, large time and compuational cost have to be dealt with.

The above mentioned problems motivate the use of parallelized adaptive surrogate-based design optimization in the MUTE front crash system. The PASBDO follows the steps demonstrated in Figure 5.9 of Section 5.6. Details of the PASBDO process for the front crash system are discussed.

36 initial sample points are generated by LHS method for the construction of initial surrogate models of L_{cr}^{US}, F_{sec}^{US}, L_{cr}^{US} and F_{sec}^{US}. Since there are seven design variables in this problem, 36 is the number of coefficients in a full quadratic regession function. With 36 sample points, an initial Kriging model containing a second-order polynomial regression function is constructed for each response.

The number of infill points to be searched for each response is set to eight. Since there are four responses to be approximated, 32 infill points are found in total, with the multiple-searching RMS infill criterion. The minimum allowable distance between sample points is set to 0.1. Distances between all infill points are calculated. If the distance between two points is smaller than 0.1, then one of them is considered to be redundant and is deleted. The remained infill points are evaluated on the crash simulation model parallelly using cluster computing technique.

The maximum number of iterations of PASBDO is set to ten as one stopping criterion. When more than ten loops of PASBDO are needed, the total number of infill points to be evaluated with high-fidelity simulations can be as many as 300, and, therefore, the efficiency is no more good. The tolerance of the change in design variables, objective function, constraint functions in two iterations and the maximum allowable constaint violation are set to 10^{-3}, which are the convergence criteria in addition to the stopping criterion.

Surrogate-based design optimization is performed, which include one GA optimization for global searching and 35 gradient-based optimizations for local searching. All nonredundant optimal results are validated simultaneously. The number of gradient-based optimizations changes adaptively in the PASBDO process, which increases by 35 in each iteration when convergence of the PASBDO results is not obtained.

Infill points aiming at increasing the accuracy of surrogate models near the optimal regions are identified with the multiple-searching CEI infill criterion. All nonredundant infill points are evaluated simultaneously. Surrogate models are refined, the minimum allowable distance between sample points decreases and the loops of PASBDO continues until an optimal design is converged. Following the above procedures, an optimal design is converged in the fifth loop of PASBDO. The converging process is shown in Figure 7.10. Details of results in PASBDO for optimization of the MUTE front crash system are listed in Table 7.2. The meanings of the symbols in Table 7.2 are explained as follows:

n_{samp}: the number of sample points. This is only valid for the zeroth loop, where initial sampel points are generated with LHS for construction of initial surrogate models.

n_{infill}: the number of infill points found in each loop by multiple-searching criteria.
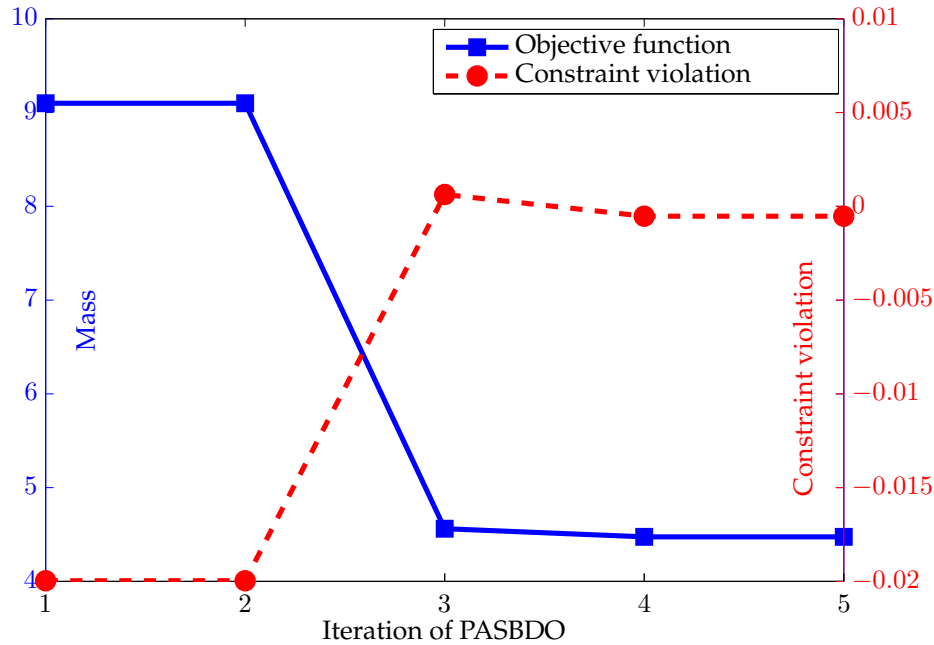
Figure 7.10.: The converging process of PASBDO for the optimization of the MUTE front crash system.

n_{feas}: the number of feasible designs found in each loop. The number of feasible designs is an important sign of the quality of surrogate models in regions containing optimal designs.

m_{opt}: the objective function value of the optimal designs found in each loop. It is the mass of the front crash system, which is to be minimized.

g_{max}: the maximum violation of constraints. The design is infeasible when $g_{\text{max}} > 0$, and feasible when $g_{\text{max}} \leq 0$. The closer the g_{max} is to 0, the nearer the optimal design is located at the constraint border. The value of g_{max} does not only represent the feasibility of the design, but also indicate the quality of the surrogate models at the optimal design. Only when the quality of the surrogate models is good enough near the optimal regions, the final design can be pushed to the constraint border during the optimization process to discover better designs.

From Figure 7.10 and Table 7.2, it can be concluded that the PASBDO method is capable of converging efficiently. The ability of identifying optimal designs accurately is discussed later. With a few number of infill points in each loop pf PASBDO, the surrogate models are refined in the critical design regions and the final optimal design is identified. The mass of the front crash system is reduced to less than 4.5 kg, with four of the five constraints are driven to the border. The total number of loops of PASBDO is five and the total number of high-fidelity evaluations is 110.

Besides, the PASBDO is robust because its performance does not depend on the starting vectors as many gradient-based optimization methods. For PASBDO, no initial starting vec-

Table 7.2.: Results in PASBDO for optimization of the MUTE front crash system .

Loops of PASBDO	n_{samp}	n_{infill}	n_{feas}	m_{opt}	g_{max}
0	36	0	0	7.3909	0.0237
1	0	32	1	9.0988	-0.0196
2	0	17	1	9.0988	-0.0196
3	0	10	6	4.5621	6.3004e-4
4	0	13	7	4.4750	-5.2623e-4
5	0	2	7	4.4750	-5.2623e-4

tor need to be provided previously. The PASBDO method is able to identify the regions containing the best optimal designs based on the distributed sample points and infill points. All starting vectors for SQP are selected from the DoE. As long as the sample points from DoE have good distribution, the risk of trapping in local optimal regions is small. The PASBDO is tested five times with different initial sample points from LHS, and all five tests converge to the same design. The design variables, system responses and constraint function values of the final design obtained with PASBDO are given in Table 7.3. It is seen that the angle of

Table 7.3.: The design variables, system responses and constraint function values of the final design obtained with PASBDO for the MUTE front crash system.

x_1	x_2	x_3	x_4	x_5	x_6	x_7
9.9983	52.4874	3.3195	-49.9713	154.9704	30.2499	1.0004

System responses	m	L_{cr}^{US}	F_{sec}^{US}	L_{cr}^{EU}	F_{sec}^{EU}	$\frac{x_3}{x_2}$
	4.4750	449.3460	1.4311×10^5	406.9220	1.4835×10^5	0.0632

Constraint functions	$g_1(\boldsymbol{x})$	$g_2(\boldsymbol{x})$	$g_3(\boldsymbol{x})$	$g_4(\boldsymbol{x})$	$g_5(\boldsymbol{x})$
	-0.0015	-0.0460	-0.0957	-0.0110	-5.2623e-4

the crash tubes x_1 is close to the upper boundary $10°$. This result indicates that nonparallel tubes are preferred, which is also a sign that the constraints from the EU NCAP is more critical than those from US NCAP in our design problem. The offset of the bumper center line x_3 is approaching the lower boundary -50 mm, which forms a C-shape bumper curve. This shape of bumper is a result of the compensation for the US NCAP, since the design of x_1 takes side in the EU NCAP. Both the width x_6 and the thickness x_7 of the bumper are at the lower boundary. The design results of the bumper indicate that the stiffness of the bumper should be small so that the impact energy can be directly taken by the tubes, which are the

energy absorbers. From the results of the system responses, it can be seen that the crash length of the tube in US NCAP case L_{cr}^{US} is close to the boundary 450 mm. The section force in EU NCAP F_{sec}^{EU} is close to approach the boundary 150 KN. The thickness to radius ratio of the crash tubes activates the critical buckling stress, as is seen by $g_5(\boldsymbol{x}) = -5.2623e^{-4} \to 0$. Besides, the other constraint $g_4(\boldsymbol{x}) = -0.0110$ is also feasible and close to be activated and $g_3(\boldsymbol{x}) = -0.0460$ is although not active, but is feasible and not far away from the boundary.

With the optimal design, the final states of the crash simulations for the US NCAP and EU NCAP load cases are shown in Figure 7.11. It can be seen that the folding modes in both load
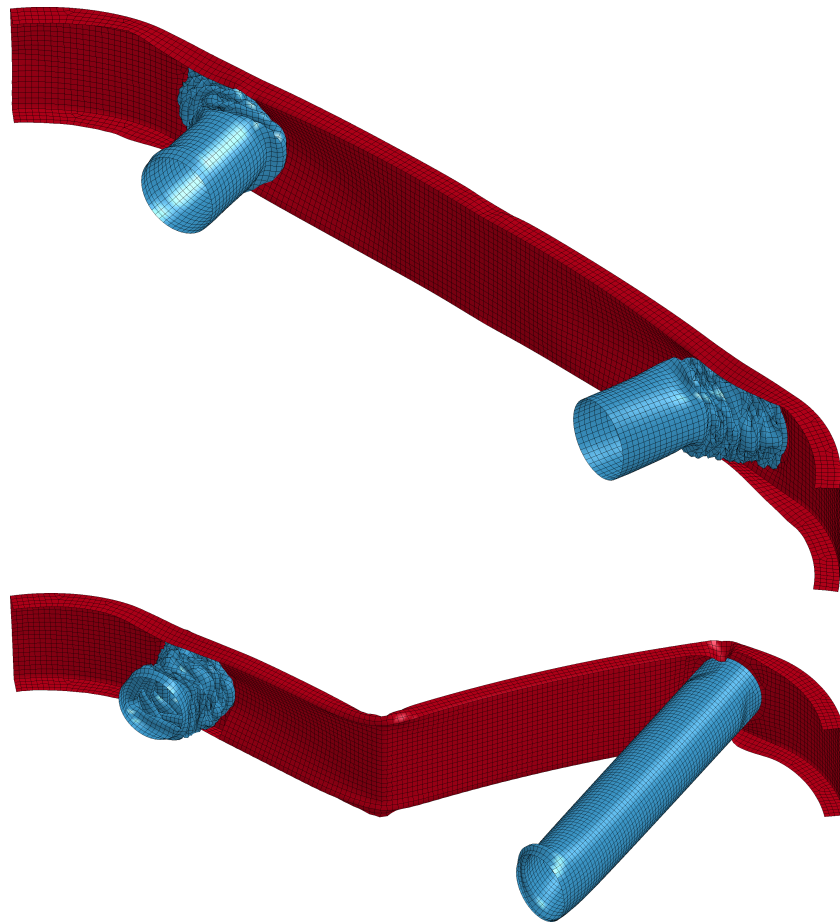

Figure 7.11.: The final states of the crash simulations for the US NCAP (top) and EU NCAP (bottom) load cases with the optimal design.

cases are stable and regular, which are important in absorbing the kinetic energy during the vehicle front crashing process. Since both the compression length of tubes and the maximum section force are under the allowable limits, the final design satisfies the crashworthiness reqguirements, and the structural mass is greatly reduced.

7.4. Discussion of Computational Effort

The structural design optimization problem of the MUTE front crash system studied in this dissertation involves seven design variables. The non-smooth feature of the system

responses of the MUTE front crash system produces difficulties of convergence in the optimization process. Computational efforts required by different methods are discussed. Because of the non-smooth responses and convergence problem, the optimzation result by a gradient-based method for the same front crash system in not yet satisfactory. Either the obtained design is infeasible or the design produces much larger structural mass. The design optmization by a genetic algorithm faces large computational effort becaused of the convergence problem. It is shown in Binder and Wehrle (2012) that for optimization of the front crash system with only 5 design variables:

- when no surrogtae models are constructed, 487 evaluations are required using a genetic algorithm developed in LLB named GAME Langer (2007);

- when standard surrogate-based design optimization is used, 400 evaluations are carried out to generate sample points, yet feasibility of the final design is not ensured;

- when 800 sample points are used for surrogate-based design optimization, it is possible to find a feasible design. However, the constraints are not driven to the boundary, which means that further computational effort is required to obtain better designs.

The PASBDO method, on the other hand, overcomes the convergence problem. It successfully finds the optimum in five loops, with only 110 evaluations of the high-fidelity crash simulations. The design obtained by PASBDO is feasible, while driving the constraint functions to the boundaries. Meanwhile, the structural mass has been greatly reduced. With the use of parallelized computing technique, the time expense of PASBDO on this problem is less than eight hours. The parallelization speed-up ratio in this problem is $\mathcal{P}_p = 9.2$.

From the above discussions, the advantage of PASBDO method in reducing computational effort can be clearly demonstrated. Besides high efficiency, further advantages of the PASBDO are good accuracy and robust performance, which are not shown in other methods for the optimization of the MUTE front crash system under crashworthiness constraints.

7.5. Summary of Chapter

In this chapter, design optimization of the MUTE front crash system under impact loads is introduced. The challenges in this problem includes computationally expensive system equations and non-smooth system responses, and, therefore, convergence is hard to obtain. Further, no analytical gradient information is available for the crash simulation, and, therefore, gradient-based optimization method cannot optimize efficiently. Besides, the gradient-based optimization method has high risk of trapping into infeasible domains. With the PASBDO method proposed in this dissertation, the problems are overcame. Optimal design of the front crash system are obtained by PASBDO with high efficiency, good accuracy and robust performance.

8. Conclusion

Owing to the large computational time expense in high-fidelity simulations, surrogate-based modeling and optimization has become a state of the art in solving engineering design optimization problems of computational expensive system equations. Since surrogate models are computationally cheaper approximations of the engineering system equations, it is much more efficient to search for the optimal solution(s) based on surrogate models, especially when a large number of iterations are required in optimization. The goal in surrogate-based modeling and optimization of the engineering problems is to maintain the necessary degree of accuracy of the optimized solution while using as fewer high-fidelity simulations as possible. Although the advantage in surrogate-based modeling and optimization is clear, it is not always easy to achieve the goal. Both the requirements in efficiency and in accuracy can be failed, when the system equations to be approximated are high dimensional or highly-nonlinear. To cope with these problems, the PASBDO method is proposed in this dissertation, which are validated with a number of mathematical and engineering examples. Among them, two typical computational expensive engineering problems are introduced, which are the aircraft wing design with simulation of fluid structural interaction and the automobile front crash system design with crash simulation. The optimization of these two problems are both solved with the proposed PASBDO method. It is demonstrated that the PASBDO method is capable of providing high-accuracy solutions efficiently in engineering optimization. The strategies in the PASBDO are concluded as follows:

- It reduces the dimension of surrogate modeling space using engineering knowledge of a system, which, more specifically, forms the geometrical and mechanical properties of the mechanical structural models.

- By properly choosing the parameters to be approximated, it reduces the complexity of the input-output relations for approximation, and, therefore, increases the efficiency and accuracy in surrogate modeling.

- Genetic algorithm is used to search for the best formats of the approximation functions with respect to the knowledge-based terms. Therefore, the surrogate modeling process is a combination of the heuristic mathematical searching tool, the full utilization of engineering knowledge and the Kriging data-mining technique.

- Multiple-searching infill criteria are designed to refine the surrogate models adaptively. The multiple-searching infill criteria identify the positions, which are able to contribute the most in improving the quality of surrogate models and the fitness of the optimization results.

- Optimization with both evolutionary strategy and multi-starting points for SQP are used to explore the surrogate models in order to search for the optimal design(s). The well distributed sample points in the design space and the infill points are selected to be the initial designs for the multi-starting points SQP optimization. Such strategy enables the PASBDO to explore the design space more thoroughly.

- Further, since the evaluation of all sample points and infill points are independent of each other, parallelized computing technique is used in PASBDO.

The advantages of PASBDO demonstrated in the examples of this dissertation are listed below.

- It is shown that vastly fewer sample points are required by PASBDO.

- Much higher accuracy in both the surrogate models and the optimization results is obtained by PASBDO compared with standard surrogate-based design optimization (SBDO) method.

- Compared with the direct optimization (DO) method, the PASBDO is remarkably faster for both the aircraft wingbox design and automobile design problems.

- The PASBDO method shows higher chance to find better designs than the direct optimization method and standard surrogate-based optimization method.

- The PASBDO is also shown to be a robust optimization tool for the large-scale and challenging structural engineering design problems.

However, drawbacks of the PASBDO method and possible directions for further improving this method should also be mentioned. The effects in dimension reduction of surrogate modeling space depends on the available engineering knowledge of the specific problem. Therefore, the PASBDO is not a generalized method for all computational expensive optimization tasks. It is more a general idea, which is very important and should always be kept in mind while solving large-scale structural engineering optimization problems. When the number of design variables is extremely large and the existing engineering knowledge for reduction of dimension is limited, the required number of sample points could still be rather large. Therefore, the expense for the construction of surrogate models will be too high to show its advantage in efficiency. The PASBDO method is designed for optimization with continuous design variables, which means the reduction of computational expense for problems involving integer programming is not considered yet. To add an interface for discrete design variables in the Kriging model and the infill criterion is a possible developing direction.

Finally, it should be pointed out that surrogate modeling is only one category of the methods dealing with the computational expansive system equations in engineering design optimization. PASBDO is shown to have overcome some main limitations in traditional surrogate modeling techniques and have achieved high efficiency and accuracy in surrogate-based modeling and optimization. However, that is not to say the PASBDO surpasses other

categories of methods, which also have great performance in reducing computational efforts. For example, the model order reduction (MOR) methods transform large matrices of a system to lower order matrices while keeping the accuracy of the eigen-values and vectors, so that the solving time is greatly reduced. The semi-analytical gradient calculation has also been implemented and frequently used, which is very efficient in gradient based optimization of large-scale structural engineering problems. Finally, it is recommended that combinations of different model reduction and approximation methods be used to accelerate engineering design optimization, which are originally prohibitive because of computationally expensive simulations.

A. Bibliography

W. Abramowicz and N. Jones. Transitionfrom initial global bending to progressive buckling of tubes loaded statically and dynamically. *International Journal of Impact Engineering*, 19: 415–437, 1997.

D.M. Allen. The prediction sum of squares as a criterion for selecting predictor variables. Technical report, Department of Statistics, University of Kentucky, 1971.

L.F. Alvarez, V.V. Toropov, D.C. Hughes, and A.F. Ashour. Approximation model building using programming methodology: Applications. In *Second ISSMO, AIAA Internet Conference on Approximations and Fast Reanalysis in Engineering Optimization*, 2000.

S. Armanini. Construction of a finite element reference model of a wing. Master's thesis, Institute of Lightweight Structures, Technical University of Munich, Germany, 2011.

J. S. Arora. *Introduction to Optimum Design*. Elsevier Academic Press, 2004.

H. Baier, C. Seeßelberg, and B. Specht. *Optimierung in der Strukturmechanik*. Vieweg+Teubner Verlag, Springer, 1994. doi: 10.1007/978-3-322-90700-4.

J.W. Bandler, Q.S. Cheng, S.A. Dakroury, and M.H. Bakr. Space mapping: The state of the art. *IEEE Transactions on Microwave Theory and Techniques*, 1:337–361, 2001.

N. A. Barricelli. Symbiogenetic evolution processes realized by artificial methods. *Methodos*, 9:143–182, 1957.

G. Baudat and F. Anouar. Kernel-based methods and function approximation. In *International Joint Conference on Neural Networks*, 2001.

B.D.Agarwal and L.J. Broutman. *Analysis and Performance of Fiber Composites*. John Wiley & Sons, 1980.

D. Binder and E. Wehrle. Parallelized optimization mmethods in the synthesis of structures under crash loading. Master's thesis, Institute of Lightweight Structures, Technical University of Munich, Germany, 2012.

R. Blumhardt. *Numerische Optimierung des Crashverhaltens von Fahrzeugstrukturen und -komponenten*. PhD thesis, Institute of Lightweight Structures, Technical University of Munich, Germany, 2001.

P.D. Bois. Ls-dyna analysis for structural mechanics, an overview of the core analysis features used by ls-dyna to simulate highly nonlinear transient behavior in engineered structures and systems. Technical report, Alter Engineering, 2011.

A.J. Booker, J.E. Dennis, P.D. Frank, D.B. Serafini, and V. Torczon. Optimization using surrogate objectives on a helicopter test example. In *SSG Applied Research Technology, SSGTECH-97-027, Boeing Shared Services Group (SSG), The Boeing Company, Seattle, WA*, 1997.

G.E.P. Box and K.B. Wilson. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society Series B Method*, 13:1–45, 1951.

S. Butterworth. On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer*, 7:536–541, 1930.

P.P. Camanho, C.G. Davila, S.T. Pinho, L. Iannucci, and P. Robinson. Prediction of in situ strength and matrix cracking in composites under transverse tension and in-plane shear. *Composites Part A: Applied Science and Manufacturing*, 37:165–176, 2006.

A.R. Collar. The first fifty years of aeroelasticity. *Aerospace (Royal Aeronautical Society Journal)*, 2:12–20, 1978.

O. Ewan. *Evolving Turing's Artificial Neural Networks*. PhD thesis, University of Canterbury, Department of Physics and Astronomy, 2010.

F. Fellner and E.J. Wehrle. Entwurfsoptimierung aufprallenergieabsorbierender strukturen mithilfe vereinfachter modellierung. Master's thesis, Institute of Lightweight Structures, Technical University of Munich, Germany, 2013.

R.A. Fisher. *Statistical methods for research workers*. Oliver and Boyd, Edinburgh, 1925.

R.A. Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh, 1st ed. edition, 1935.

A. Forrester, A. Sóbester, and A. Keane. *Engineering Design via Surrogate Modelling*. John Wiley & Sons, 2008.

J. Forsberg and L. Nilsson. On polynomial response surfaces and kriging for use in structural optimization of crashworthiness. *Structural and Multidisciplinary Optimization*, 29(3):232–243, 2005.

S. Fuchs, E.J. Wehrle, S. Matz, and M. Lienkamp. Mute - ein effizientes und marktfähiges elektrofahrzeug durch intelligenten leichtbau. In *5. Symposium Faszination Karosserie und Fahrzeugkonzepte, Wolfsburg*, 2012.

A.A. Giunta and M.S. Eldred. Implementation of atrust region model management strategy in the dakota optimization toolkit. In *8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, CA*, 2000.

T. Goel. *Multiple Surrogates and Error Modeling in Optimization of Liquid Rocket Propulsion Components*. PhD thesis, University of Florida, 2007.

D. Gorissen. *Grid–enabled Adaptive Surrogate Modeling for Computer Aided Engineering*. PhD thesis, Ghent University, Faculty of Engineering Sciences, Department of Information Technology, 2009.

K.K. Gupta and J.L.Meek. *Finite Element Multidisciplinary Analysis*. AIAA Education Series, 2003.

K. Hamza and K. Saitou. Vehicle crashworthiness design via a surrogate model ensemble and a co–evolutionary genetic algorithm. In *ASME International Design Engineering Technical Conferences & Computers and Informaion in Engineering Conference, September 24-28, Long Beach, California, USA*, 2005.

T. Hastie, R. Tibshirani, and J. Friedman. *Data Mining, Inference, and Prediction*. Springer, 2008.

G. Hellard. Composites in airbus: A long story of innovations and experiences. In *Global Investor Forum by Dr Roland Thevenin*, 2008.

L.O. Hesse. *Ludwig Otto Hesse's Gesammelte Werke*. Chelsea Publishing Company Bronx New York, 1897.

R. Hill. A user-friendly theory of orthotropic plasticity in sheet metals. *International Journal of Mechanical Sciences*, 35:19–25, 1993.

M.J. Hinton, A.S. Kaddour, and P.D. Soden, editors. *Failure Criteria in Fibre-Reinforced-Polymer Composites: The World-Wide Failure Exercise*. Elsevier, 2004.

Z. Huang, C. Wang, J. Chen, and H. Tian. Optimal design of aeroengine turbine disc based on kriging surrogate models. *Computers & Structures*, 89:23–37, 2011.

R.M. Jones. *Mechanics of Composite Materials*. Taylor & Francis, 1998.

W. Karush. Minima of functions of several variables with inequalities as side constraints. Master's thesis, Deptartment of Mathematics, University of Chicago, 1939.

H. Khodaparast, G. Georgiou, J.E. Cooper, L.Riccobene, S.Ricci, G.A.Vio, and P.Denner. Efficient worst case 1 – cosine gust loads prediction. In *International Forum on Aeroelasticity and Structural Design, Paris, France*, 2011.

A.N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea publishing company, 1950.

S. Koziel and L. Leifsson, editors. *Surrogate-Based Modeling and Optimization, Applications in Engineering*. Springer, New York, 2013.

D.G. Krige. A statistical approach to some mine valuations and allied problems at the witwatersrand. Master's thesis, University of Witwatersrand, 1951.

H.W. Kuhn and A.W. Tucker. Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, pages 481–492, 1951.

J.L Lagrange. *Mécanique analytique*. Paris, Ve Courcier, 1815.

H. Langer. *Extended Evolutionary Algorithms for Multiobjective and Discrete Design Optimization of Structures*. PhD thesis, Institute of Lightweight Structures, Technical University of Munich, Germany, 2007.

J. Laurenceau, M. Meaux, M. Montagnac, and P. Sagaut. Comparison of gradient–based and gradient–enhanced response–surface–based optimizers. *AIAA JOURNAL*, 48(5):981, May 2010.

A.M. Legendre. *Nouvelles Méthodes pour la Détermination des Orbites des Comètes, Appendix on a least-squares, p.72-80*. Firmin-Didot, Paris, 1805.

L. Leifsson, S. Koziel, E. Jonsson, and S. Ogurtsov. *Chapter title, Aerodynamic Shape Optimization by Space Mapping; Book title, Surrogate-Based Modeling and Optimization, Applications in Engineering*. Springer, New York, 2013.

D.J. Lucia, P.S. Beran, and W.A. Silva. Reduced-order modeling: New approaches for computational physics. *Progress in Aerospace Sciences*, 40:51–117, 2004.

U. Luther and K. Rost. Matrix exponentials and inversion of confluent vandermonde matrices. *Electronic Transactions on Numerical Analysis*, 18:91–100, 2004.

A.M. Mood, F.A. Graybill, and D.C. Boes. *Introduction to the theory of statistics*. McGraw-Hill, Inc, 1913.

P.B. Nair. Physics-based surrogate modeling of parameterized pdes for optimization and uncertainty analysis. In *Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference,Denver, CO*, 2002.

M.C.Y. Niu. *Airframe Structural Design*. Conmilit Press, Hong Kong, 1988.

M. Palanterä and A. Mönicke. Integrated composites design with esacomp and hyperworks. In *2013 Americas Altair Technology Conference*, 2013.

Ö. Petersson, F. Stroscher, and H. Baier. Multidisciplinary optimisation of aircraft wings including gust loads. In *2nd RAeS Aircraft Structural Design Conference, London*, 2010.

R.R. Picard and R.D. Cook. Cross validation of regression models. *American Statistical Association*, 79(387):575–583, September 1984. URL `http://www.jstor.org/stable/2288403`.

M.J.D. Powell. Algorithms for nonlinear constraints that use lagrangian functions. *Mathematical Programming, North-Holland Publishing Company*, 14:224–248, 1978.

N.T. Quan. The prediction sum of squares as a general measure for regression diagnostics. *Business & Economic Statistics*, 6(4):501–504, October 1988. URL `http://www.jstor.org/stable/1391469`.

D. E. Raveh. Computational-fluid-dynamics-based aeroelastic analysis and structural design optimization-a researcher's perspective. *Comput.Methods Appl. Mech. Engrg*, 194: 3453–3471, 2005.

J.N. Reddy and A.K. Pandey. A first-ply failure analysis of composite laminates. *Computers & Structures*, 25:371–393, 1987.

G. Redeker. Dlr–f4 wing–body configuration. *AGARD, 7 rue Ancelle, 92200 Neuilly–sur–Seine France*, 2:AGARD–AR–303, 1994.

K.L. Reifsnider, editor. *Damage in Composite Materials: Basic Mechanisms, Accumulation, Tolerance, and Characterization*. ASTM International, 1982.

C. Runge. über empirische funktionen und die interpolation zwischen äquidistanten ordinaten. *Zeitschrift für Mathematik und Physik*, 46:224–243, 1901.

J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design analysis of computer experiments. *Stat Sci*, 4:409–423, 1989.

D.M. Schuster, D.D. Liu, and L.J. Huttsell. Compuational aeroelasticity: Succesee, progress, challenge. *Journal of Aircraft*, 40, No.5:843–856, 2003.

K. Sieberz, D. Bebber, and T. Hochkirchen. *Statistische Versuchsplanung*. Springer, Berlin, 2010.

S. M. Stigler. Gergonne's 1815 paper on the design and analysis of polynomial regression experiments. *Historia Mathematica*, 1(4):431–439, November 1974.

M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147, 1974. URL `http://www.jstor.org/stable/2984809`.

S. Timoshenko. *Theory of Elastic Stability*. McGraw-Hill, Inc, 1961.

M. Tischer and E.J. Wehrle. Untersuchung des aufprallverhaltens eines elektrokleinstfahrzeuges anhand eines parametrisierten modells mit expliziter finite element analysis. Master's thesis, Institute of Lightweight Structures, Technical University of Munich, Germany, 2012.

A. Todoroki and M. Sekishiro. Stacking sequence optimization to maximize the buckling load of blade-stiffened panels with strength constraints using the iterative fractal branch and bound method. *M. Compos B*, 39(5):842–850, 2008.

S.W. Tsai and E.M. Wu. A general theory of strength for anisotropic materials. *Journal of Composite Materials*, 5(1):58–80, 1971. doi: 10.1177/002199837100500106.

S. Ulaganathan and N. Asproulis. *Chapter title, Surrogate Models for Aerodynamic Shape Optimization; Book title, Surrogate-Based Modeling and Optimization, Applications in Engineering.* Springer, New York, 2013.

S.M. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

H. Wang and G. Li. Min-median-max metamodel-based unconstrained nonlinear optimization problems. *Struct Multidisc Optim*, 45(3):401–415, 2012.

E.J. Wehrle. Design optimization of lightweight structures for electrical vehicles including crashworthiness. In *Leichtbau-Colloquium, Landshut*, 2013.

E.J. Wehrle, B. Sauerer, R. Wehrle, and H. Baier. Ganzheitliche strukturoptimierung von fahrzeugkarosserien der elektromobilität. Technical report, Fakultätsseminar Mobilität, TUM Maschinenwesen, Garching bei München, 2012.

D.V. Widder. *Advanced Calculus*. Prentice-Hall, INC., 1947.

J.R. Wright and J.E. Cooper. *Introduction to Aircraft Aeroelasticity and Loads*. John Wiley & Sons, Ltd, 2007.

Q. Xu, E.J. Wehrle, and H. Baier. Adaptive surrogate-based design optimization with expected improvement used as infill criterion. *Optimization, A Journal of Mathematical Programming and Operations Research, Taylor & Francis*, 61:661–684, 2012. doi: 10.1080/02331934.2011.644286.

Q. Xu, E.J. Wehrle, and H. Baier. *Chapter title, Knowledge-Based Surrogate Modeling in Engineering Design Optimization; Book title, Surrogate-Based Modeling and Optimization, Applications in Engineering*. Springer, New York, 2013. doi: 10.1007/978-1-4614-7551-4. URL `http://link.springer.com/book/10.1007/978-1-4614-7551-4/page/1`.

H. Zarei. *Experimental and Numerical Investigation of Crash Structures using Aluminium Alloys*. PhD thesis, GottfriedWilhelm Leinbiz Universitaet Hannover, 2008.

Y.T. Zhang and E.J. Wehrle. Parametrized geometric modeling of space frames for use in multidisciplinary design optimization. Master's thesis, Institute of Lightweight Structures, Technical University of Munich, Germany, 2013.