Robust optimal shape design strategies for systems with friction-induced instabilities and uncertain parameters

Achille Jacquemond

To cite this version:

Achille Jacquemond. Robust optimal shape design strategies for systems with friction-induced instabilities and uncertain parameters. Other. Ecole Centrale de Lyon, 2024. English. NNT : 2024ECDL0039 . tel-04920798

HAL Id: tel-04920798
https://theses.hal.science/tel-04920798v1

Submitted on 30 Jan 2025

N° d’ordre NNT : 2024ECDL0039

THESE de DOCTORAT DE L’ECOLE CENTRALE DE LYON
membre de l’Université de Lyon

École Doctorale N°162
MEGA – Mécanique, Énergétique, Génie Civil, Acoustique

Spécialité de doctorat : Génie mécanique

Soutenue publiquement le 17/10/2024, par :

Achille JACQUEMOND

Robust optimal shape design strategies for systems with friction-induced instabilities and uncertain parameters

Devant le jury composé de :

LALLART, Mickaël, Professeur, INSA Lyon	Examinateur, Président
AUGER, Anne, Directrice de recherche, Inria RandOpt Team, CMAP, École Polytechnique	Rapporteure
BERGER, Sébastien, Professeur, INSA Centre Val de Loire	Rapporteur
MAHMOUDI, Sonia, Maitresse de conférences, Advanced Institute for Materials Research, Tohoku University	Examinatrice
SHIMOYAMA, Koji, Professeur, Kyushu University	Examinateur
BESSET, Sébastien, Maitre de conférences HDR, École Centrale de Lyon	Directeur de thèse
GILLOT, Frédéric, Maitre de conférences, École Centrale de Lyon	Co-encadrant de thèse

Unité de recherche : LTDS – Laboratoire de Tribologie et Dynamique des Systèmes

Remerciements

Je tiens à remercier mes encadrants de thèse Sébastien Besset et Frédéric Gillot qui m'ont donné l'opportunité d'effectuer cette thèse sur un sujet très intéressant et dans le cadre d'une collaboration internationale. Je les remercie pour leurs conseils d'un point de vue scientifique, mais aussi pour leur soutien moral, et leur aide précieuse pour permettre les mobilités au Japon. Grâce à leur encadrement, j'ai pu achever ces travaux de thèse avec succès.

Je remercie Koji Shimoyama de m'avoir accueilli dans son laboratoire à l'université de Tohoku pendant trois mois durant ma première année de thèse, ainsi que pour ses conseils scientifiques tout au long de ces travaux, et aussi d'avoir accepté de faire partie du jury.

Je remercie le professeur Obayashi de m'avoir accuelli au sein de son laboratoire à l'université de Tohoku et je remercie toutes les personnes qui m'ont soutenu et aidé au Japon, lors d'une année qui a été particulièrement enrichissante.

Je remercie Anne Auger et Sébastien Berger d'avoir accepté de rapporter ma thèse et pour leurs remarques et questions judicieuses.

Je remercie Mickaël Lallart d'avoir accepté de présider mon jury de thèse.

Je remercie Sonia Mahmoudi d'avoir accepté d'être examinatrice dans mon jury et pour ses questions très intéressantes lors de la soutenance.

Je remercie mes parents et ma sœur qui m'ont soutenu tout au long de mon parcours académique.

Je remercie mes anciens collaborateurs de la société Mecalam et mon ancien patron François, qui m'ont beaucoup encouragé et inspiré à poursuivre cette thèse.

Je remercie le personnel de l'École Centrale et du LTDS, les autres doctorants, et toutes les personnes de l'administration pour leur aide pendant ces trois ans.

Je remercie également mes amis pour leur soutien durant cette expérience.

Abstract

The design of mechanical systems is a challenging subject for industrial manufacturers, with ever-increasing demands for high performance, precision, and speed. Numerical shape optimization is becoming more and more popular for design of mechanical systems for its effectiveness when compared to creating and adjusting designs sequentially, but problems such as absence of gradient, nonlinearity, and computational cost of the performance metrics constitute major difficulties. Moreover, geometric uncertainties can have considerable impact on system performance, and must be mitigated to ensure robustness of the proposed solutions. Disc brakes, used in various types of vehicles, are complex systems which exhibit squeal noise due to friction-induced instabilities. Numerically quantifying squeal noise leads to gradient-free, nonlinear, and costly performance functions which are characteristic of the overall challenges in industrial design of mechanical systems. The presented PhD research introduces new ways of conducting robust shape optimization for industrial design combining modern numerical tools, applied to the case of a disc brake under squeal noise criterion. Based on the Isogeometric Analysis method for shape parametrization and mechanical analysis, two approaches are proposed to search for optimal designs which balance performance and robustness to uncertainties. The first method uses sparse Polynomial Chaos Expansions to get an estimation of output variability to uncertainties and combines it with a genetic algorithm to search for optimal solutions. This approach is shown to provide solutions which balance performance and robustness at a reasonable computational cost and with satisfactory accuracy. The second method uses a variant of the Kriging surrogate modeling method capable of taking into account noisy observations for uncertainty propagation, combined with a Bayesian Optimization-inspired enrichment strategy. This approach shows potential for great computational cost reduction although more efforts are to be made to increases applicability in high dimensions and accuracy of the associated robustness metric. Overall the conducted research demonstrates the difficulties of carrying out shape optimization of mechanical systems while taking into account the impact of uncertainties, with compromise between computational cost and accuracy being the key aspect. Nonetheless, both proposed methods show great potential and open new perspectives in the current context of industrial design of mechanical systems.

Keywords: shape optimization, uncertainty quantification, robustness, mechanical design, multi-objective optimization, blackbox functions, Isogeometric Analysis, Polynomial Chaos Expansions, Kriging, nonlinear mechanics, disc brake, squeal noise.

Résumé en français

La conception de systèmes mécaniques est un défi pour les industriels, avec des exigences croissantes en performance, en precision, et en temps de calcul. L'optimisation de forme numérique devient de plus en plus utilisée de par son efficacité par rapport à la conception par création et modification séquentielle de la géometrie des pièces mécaniques. Cependant, plusieurs problèmes tels que l'absence de gradient, la non-linearité, et le coût de calcul des fonctions numériques utilisées pour quantifier la performance, constituent des difficultés majeures. De plus, la présence d'incertitudes géometriques pouvant avoir un impact considérable sur la performance des systèmes mécaniques, celles-ci doivent être prises en compte. Les freins à disque, presents dans de nombreux types de véhicules, sont des systèmes mécaniques complexes qui ont tendance à émettre un bruit de crissement dû à un phénomène complexe d'instabilités vibratoires induites par frottement. La quantification numérique du bruit de crissement pour sa minimisation mène à une fonction de performance sans gradient, coûteuse en temps de calcul et non-linéaire, qui représente fidèlement les principaux défis en conception optimale de systèmes mécaniques. Le travail de thèse présenté introduit de nouvelles approches pour l'optimisation robuste de forme de systèmes mécaniques, en combinant des méthodes numériques modernes, appliquées au cas d'un frein à disque sous critère de minimisation de bruit. En se basant sur la méthode d'Analyse Isogéometrique pour la paramétrisation de la forme et l'analyse du comportement mécanique, deux approches qui permettent de rechercher des conceptions optimales équilibrant performance et robustesse aux incertitudes, sont proposées. La première approche utilise la méthode du Chaos Polynomial creux afin d'obtenir une estimation de la variabilité de la fonction de performance aux incertitudes géometriques, combinée avec un algorithme génétique pour la recherche de solutions optimales. La capacité de cette approche à fournir des conceptions équilibrant performance et robustesse avec un coût de calcul raisonnable et une précision satisfaisante, est démontrée. La deuxième approche utilise la méthode du Krigeage bruité, méthode de meta-modélisation avec la possibilité de prendre en compte du bruit dans ses observations, pour la quantification d'incertitudes, combinée avec une stratégie d'enrichissement inspirée de l'Optimisation Bayésienne. Les résultats obtenus démontrent le grand potentiel de cette approche à réduire le coût de calcul de l'optimisation robuste, bien que des efforts soient nécessaires pour généraliser la méthode proposée en grande dimension, ainsi que pour assurer la fiabilité de l'estimation de robustesse. Généralement, l'étude présentée illustre les difficultés associées à l'optimisation de forme robuste de systèmes mécaniques complexes, dans le cas du frein à disque sous critère de bruit, l'aspect clé étant le compromis entre coût de calcul et précision des solutions obtenues. Les deux approches proposées constituent une contribution significative en optimisation de forme robuste et ouvrent de nouvelles perspectives dans le domaine.

Mots-clés: optimisation de forme, quantification d'incertitudes, robustesse, conception mécanique, optimisation multi-objectifs, fonction boîte noire, Analyse Isogéométrique, Chaos Polynomial, Krigeage, mécanique non-linéaire, frein à disque, bruit de crissement.

Contents

Nomenclature

Disc brake model

α_d First coefficient for Rayleigh damping

β_d Second coefficient for Rayleigh damping

$\boldsymbol{\mu}$ Range of studied eigenvalues for the stability function

$\boldsymbol{\Phi}$ Vector of modes of the fixed-interface structure

$\boldsymbol{\phi}_{eig}$ System eigenvector

$\boldsymbol{\Psi}$ Vector of static modes

$\boldsymbol{C}$ Damping matrix

$\boldsymbol{F}_C$ Vector of external forces applied to the contact interface degrees of freedom

$\boldsymbol{F}_{ext}$ External force vector after Craig & Bampton reduction

$\boldsymbol{F}_{ext}$ Vector of external forces

$\boldsymbol{F}_{full}$ External force vector of the full system before Craig & Bampton reduction

$\boldsymbol{F}_{NL}$ Vector of non linear contact forces

$\boldsymbol{I}$ Identity Matrix

$\boldsymbol{J}_{F_{NL}}$ Jacobian matrix of the nonlinear forces for the whole system

$\boldsymbol{J}^*_{F_{NL}}$ Jacobian matrix of the nonlinear forces for one contact point pair

$\boldsymbol{K}$ Stiffness matrix after Craig & Bampton reduction

$\boldsymbol{K}_{C,C}$ Full mass matrix's components which link internal degrees of freedom together

$\boldsymbol{K}_{C,I}$ Full mass matrix's components which link contact degrees of freedom to internal degrees of freedom

$\boldsymbol{K}_{d,init}$ Initial stiffness matrix of the disc

$\boldsymbol{K}_d$ Stiffness matrix of the disc after application of displacement interpolation

$\boldsymbol{K}_{full}$ Stiffness matrix of the full system before Craig & Bampton reduction

$\boldsymbol{K}_{I,C}$ Full mass matrix's components which link internal degrees of freedom to contact degrees of freedom

$\boldsymbol{K}_{I,I}$ Full mass matrix's components which link internal degrees of freedom together

$\boldsymbol{M}$ Mass matrix after Craig & Bampton reduction

$\boldsymbol{M}_{C,C}$ Full mass matrix's components which link contact degrees of freedom together

$\boldsymbol{M}_{C,I}$ Full mass matrix's components which link contact degrees of freedom to internal degrees of freedom

$\boldsymbol{M}_{d,init}$ Initial mass matrix of the disc

$\boldsymbol{M}_d$ Mass matrix of the disc after application of displacement interpolation

$\boldsymbol{M}_{full}$ Mass matrix of the full system before Craig & Bampton reduction

$\boldsymbol{M}_{I,C}$ Full mass matrix's components which link internal degrees of freedom to contact degrees of freedom

$\boldsymbol{M}_{I,I}$ Full mass matrix's components which link internal degrees of freedom together

$\boldsymbol{q}_g$ Vector of generalized degrees of freedom associated to each mode of the structure

$\boldsymbol{Q}_{c,tot}$ Total transfer matrix for contact correspondence between disc and pad

$\boldsymbol{Q}_c$ Transfer matrix for contact correspondence between disc and pad

$\boldsymbol{Q}_{red}$ Transfer matrix for Craig & Bampton reduction

$\boldsymbol{R}_{NR}$ Newton Raphson residue vector

$\boldsymbol{S}$ State matrix for the equations of motion

$\boldsymbol{U}$ Vector of the full system's degree of freedom displacements

$\boldsymbol{U}_C$ Vector of the full system's contact interface degree of freedom displacements

$\boldsymbol{U}_I$ Vector of the full system's internal degree of freedom displacements

$\boldsymbol{U}_0$ Equilibrium point displacements

$\boldsymbol{U}_{int}$ Vector of non contact point displacements of the disc

$\boldsymbol{U}_{new}$ Vector of new contact point displacements of the disc

$\boldsymbol{U}_{old}$ Vector of old contact point displacements of the disc

$\boldsymbol{V}$ State variable for the equations of motion

δ Relative displacement between the disc and pad contact point pair

$\delta\boldsymbol{U}$ Small perturbation of the equilibrium point

Γ Displacement interpolation function for contact correspondence between disc and pad

λ System eigenvalue

μ Friction coefficient

ν Poisson coefficient of the pad material

ω_i i-th structural mode frequency

ω_{max} Maximum frequency of the studied interval for damping

ω_{min} Minimum frequency of the studied interval for damping

$\Re^+$ Positive part of the real part operator

ρ Density of the pad material

θ Angle between the friction force direction and $\vec{u}_x$

$\vec{F}_{n,d}$ Vector of the disc contact force normal to the contact surface

$\vec{F}_{n,p}$ Vector of the pad contact force normal to the contact surface

$\vec{F}_{t,d}$ Vector of the disc contact force tangential to the contact surface

$\vec{F}_{t,p}$ Vector of the pad contact force tangential to the contact surface

$\vec{F}_{x,d}$ Disc contact force along $\vec{u}_x$

$\vec{F}_{x,p}$ Pad contact force along $\vec{u}_x$

$\vec{F}_{y,d}$ Disc contact force along $\vec{u}_y$

$\vec{F}_{y,p}$ Pad contact force along $\vec{u}_y$

$\vec{F}_{z,d}$ Disc contact force along $\vec{u}_z$

$\vec{F}_{z,p}$ Pad contact force along $\vec{u}_z$

$\vec{u}_n$ Unit vector in the normal direction to the contact surface

$\vec{u}_t$ Unit vector in the tangential direction to the contact surface

$\vec{u}_x$ Unit vector of the first Cartesian coordinate

$\vec{u}_y$ Unit vector of the second Cartesian coordinate

$\vec{u}_z$ Unit vector of the third Cartesian coordinate

ξ_d Damping rate

C_s Disc brake stability function

C_{NL} Constant for simplification of the non-linear contact force expression

d_0 Shift parameter for the stability criterion

E Young's modulus of the pad material

k_L Linear contact stiffness

k_{NL} Non-linear contact stiffness

n Number of displacement interpolation functions

n_{eig} Number of eigenvalues of the system

n_{modes} Number of structural modes considered for the Craig & Bampton reduction

R_e External radius of the disc geometry

r_e External radius of the pad geometry

R_i Internal radius of the disc geometry

r_i Internal radius of the pad geometry

T Thickness of the disc geometry

t Thickness of the pad geometry

u Brake model control point displacement along the first direction of the Cartesian coordinate system

$U_{n,d}$ Disc contact point displacement along the normal to the contact surface

$U_{n,p}$ Pad contact point displacement along the normal to the contact surface

$U_{x,d}$ Disc contact point displacement along $\vec{u}_x$

$U_{x,p}$ Pad contact point displacement along $\vec{u}_x$

$U_{y,d}$ Disc contact point displacement along $\vec{u}_y$

$U_{y,p}$ Pad contact point displacement along $\vec{u}_y$

$U_{z,d}$ Disc contact point displacement along $\vec{u}_z$

$U_{z,p}$ Pad contact point displacement along $\vec{u}_z$

v Brake model control point displacement along the second direction of the Cartesian coordinate system

V_r Disc-pad relative sliding speed

w Brake model control point displacement along the third direction of the Cartesian coordinate system

x First coordinate of the brake model's control points on the contact surface plane

y Second coordinate of the brake model's control points on the contact surface plane

Isogeometric Analysis

$\boldsymbol{B}$ Strain-displacement matrix

$\boldsymbol{C}_{BC}$ B-spline curve

$\boldsymbol{C}_{NC}$ NURBS curve

$\boldsymbol{C}_{NS}$ NURBS surface

$\boldsymbol{E}$ Material property matrix

$\boldsymbol{J}_{\xi\eta}$ Jacobian matrix for the mapping of the master element to the parametric space

$\boldsymbol{J}_{st}$ Jacobian matrix for the mapping of the physical space to the master element

$\boldsymbol{K}_{init}$ Stiffness matrix of a NURBS model

$\boldsymbol{P}_i$ vector of coordinates of the i-th control point

$\boldsymbol{s}$ NURBS knot vector

$\boldsymbol{W}$ Vector of weights for Gauss quadrature

γ_i Weight associated to i-th control point

k NURBS knot multiplicity

m_c Number of NURBS control points along the second geometric direction

n_c Number of NURBS control points along the first geometric direction

$N_{i,p}$ i-th B-spline basis function of order p

N_{INT} Number of integrations for Gauss quadrature

p Order of NURBS basis functions along the first geometric direction

q Order of NURBS basis functions along the second geometric direction

s_i NURBS knot value

Kriging

$\hat{Y}$ Kriging output prediction

β_j j-th unknown coefficient for the Kriging trend

$\boldsymbol{\Delta}$ Noisy Kriging sample outputs

$\boldsymbol{\eta}$ Main parameter of the Matérn 5/2 correlation function

$\hat{\boldsymbol{\beta}}$ Vector of Kriging coefficients

$\boldsymbol{\lambda}_{\boldsymbol{K}}$ Linear predictor coefficients

$\boldsymbol{f}$ Vector of all regression functions

$\boldsymbol{F}_X$ Vector of regression functions evaluated at all the sample points

$\boldsymbol{R}_X$ Spatial correlation matrix

$\boldsymbol{X}_K$ Matrix of Kriging observations inputs

$\boldsymbol{X}_p$ Prediction input vector

$\mathbf{Y}_K$ Vector of Kriging observation outputs

ϵ_i Noisy Kriging i-th output noise component

η_k Hyperparameter of the Matérn 5/2 correlation function

$\mathbb{E}$ Expected value operator

$\mathbb{MSE}$ Mean square error operator

$\tilde{\mathbf{Y}}_K$ Noisy Kriging sample outputs

σ_K Kriging prediction standard deviation

σ_Z Standard deviation of random process Z

d_X Number of deterministic input parameters of the physical model

f_j j-th fixed regression function for the Kriging trend

f_{cdf} Cumulative distribution function of the normal distribution

f_{pdf} Probability density function of the normal distribution

I_{EI} Expected improvement function

k_r Number of regression functions for the Kriging trend

n_s Number of samples used to build the Kriging model

R Spatial correlation function

R_{M52} Matérn 5/2 correlation function

Y_{min} Minimum value among the Kriging sample outputs

Z Kriging systematic deviation from the linear part

Optimization problem

$\boldsymbol{X}_d$ Vector of pad shape parameters

$\boldsymbol{X}_u$ Vector of pad shape parameters which are considered uncertain

$\overline{X_i^u}$ i-th uncertain parameter average value

σ_u Uncertain parameters' standard deviation

σ_{C_s} Standard deviation of the stability criterion output due to uncertainties

θ_j Angular cylindrical coordinate of the j-th pad control point for shape modification

A Area of the disc-pad contact surface

A_{min} Minimal area of the disc-pad contact surface for the braking capacity constraint

g_1 First objective function of the optimization problem

g_2 Second objective function of the optimization problem

h_1 Constraint function of the optimization problem

n_u Number of uncertain parameters

r_j Radial cylindrical coordinate of the j-th pad control point for shape modification

X_i^u i-th uncertain parameter

Polynomial Chaos Expansion

$\hat{\boldsymbol{a}}$ Final vector of PCE coefficients obtained by least-squares

α Multi-index which identifies the components of the polynomial basis

$\boldsymbol{\Psi}_t$ Vector which gathers orthonormal polynomials Ψ_α

$\boldsymbol{A}$ Regression matrix containing the values of the basis polynomials at the experimental design points

$\boldsymbol{a}$ Vector of variable PCE coefficients

$\boldsymbol{X}$ Random vector representing the uncertain input parameters of the physical model

$\boldsymbol{x}$ Input parameters of the physical model

$\boldsymbol{X}_{exp}$ Set of experimental design points for PCE construction

$\boldsymbol{y}$ Outputs of the physical model

$\boldsymbol{Y}_{exp}$ Set of experimental outputs for PCE construction

$\delta_{\alpha\beta}$ Multidimensional Kronecker delta

δ_{jk} Kronecker delta

ϵ_{emp} Normalized empirical estimation of the generalization error

ϵ_{gen} Generalization error

ϵ_{LOO} Leave-one-out cross-validation error

ϵ_{val} Generalization error evaluated through a validation set

λ_{pen} Penalization term for Least Angle Regression

$\mathscr{A}$ Subset of $\mathbb{N}^M$ which contains the selected multi-indices α

$\mathscr{A}^{M,p,q}$ q-norm truncation scheme basis

$\mathscr{A}^{M,p,r}$ Maximum interaction truncation scheme basis

$\mathscr{A}^{M,p}$ Standard truncation scheme basis

$\mathscr{M}$ Considered physical model

$\mathscr{M}^{PC}$ Truncated polynomial expansion

μ_{PCE} Mean value of the PCE

ϕ_j Univariate orthonormal polynomials

Ψ_α Multivariate polynomials orthonormal with respect to f_X

σ_{PCE} Standard deviation of the PCE

a_α Deterministic coefficients of the PCE

f_X Probability density function of random vector $\boldsymbol{X}$

$f_{x_k^u}$ Probability density function of the k-th PCE uncertain input parameter

M Number of uncertain input parameters

N_{val} Size of the validation set for error estimation

P Number of elements of $\mathscr{A}^{M,p}$

p Maximum polynomial order of the truncated PCE

q Parameter for the q-norm truncation scheme

r Parameter for the maximum interaction scheme

$W^{(i)}$ Gaussian quadrature weights for PCE coefficient calculation

$x^{(i)}$ Gaussian quadrature quadrature points for PCE coefficient calculation

x_{val}^i i-th input validation point for error estimation

Y Random variable representing the model output in the uncertain case

State of the art

$\mathbf{x}_j$ Potential solution of a multi-objective optimization problem

f_i^{obj} Objective function of a multi-objective optimization problem

f_{ls} Limit-state function

F_{st} Safety factor of a structure

L_{st} Load of a structure

P_f Probability of failure

R_{st} Resistance of a structure

S Space of input parameters of a multi-objective optimization problem

List of Figures

List of Tables

Introduction

Optimizing the performance of mechanical systems is a major goal for manufacturers. One important way to affect the performance of any system is to modify the shape of its mechanical parts. In other words, the design process is a key step in the production of mechanical systems. Creating and adjusting the design of mechanical parts sequentially has been the norm for decades, but this type of approach is very restrictive in terms of exploring the full space of possible designs. For this reason, shape optimization processes have become increasingly popular and are being implemented in various industries. Shape optimization consists in finding shapes of mechanical parts which maximize certain predefined performance metrics, using sophisticated numerical techniques. In the industrial context, simulation tools such as FEM are commonly used to quantify performance. The performance functions originating from such simulation tools are often gradient-free, nonlinear, and computationally costly. Optimization methods allowing to handle such complex cases, for example Efficient Global Optimization (EGO), have been the object of abundant research in recent years. Additionally, it is not uncommon to have multiple competing performance metrics in which case Multi-objective Optimization (MOBO) approaches must be considered, which attempt to find sets of solutions representing the best compromises between the different objectives (Pareto optimality). Moreover, the design and manufacturing processes inevitably include various sources of uncertainty which can have a considerable effect on system performance. In particular, performance can be very sensitive to geometrical discrepancies between numerically optimized designs and actual manufactured parts. Under such conditions, robustness of optimal solutions is a predominant challenge associated to shape optimization. Reliability-based Design Optimization (RBDO), which optimizes performance functions subject to probabilistic constraints such as reliability indicators or failure probability, is currently the most commonly applied method to deal with design optimization under uncertainty. Tackling the challenges related to modern design of real mechanical systems, namely the presence of uncertainties, non-linearity, computational cost and gradient-free character of performance metrics, and the multiplicity of these metrics, requires the development of ever more sophisticated and complex numerical methods. Given the current state of research on this topic, there are currently three crucial methodology aspects which will be decisive for the future of design optimization of mechanical systems under uncertainties.

Firstly, how designs are modified forms the basis of the design optimization process. There are currently several ways of describing the shapes of mechanical systems in the aim of optimizing performance. Size optimization modifies the dimensions of geometrical parameters which define the shape. Its main advantage is its ease of implementation, in particular in conjunction with simulation methods such as the Finite Element Method

(FEM) without creating troublesome meshing issues. Its main drawback is the lack of exploration of the design space since the overall shape of the system is predetermined. Topology optimization (TO) is the exact opposite, where the distribution of a material is freely optimized within a defined bounding domain in order to minimize some cost function. In theory, this method is attractive as it allows complete exploration of the design space, however in practice, issues such as how to combine it with Computer-Aided Engineering (CAE) tools, the manufacturability of the obtained designs, and necessity of the cost function's gradient make it a questionable choice in some settings especially for gradient-free design under uncertainties. Generative design, a method similar to topology optimization but which employs artificial intelligence and which is generally used earlier in the design process, suffers from some of the same problems. Shape optimization, which is somewhere in between size optimization and topology optimization, is often combined with FEM where difficulties related to meshing are common during the shape modification process. Nonetheless a relatively recent alternative to FEM called Isogeometric Analysis (IGA) has made it possible to overcome the meshing complications, and presents itself as an interesting candidate for design optimization under uncertainties.

How optimal solutions are searched for is the second key aspect. There are two main categories of methods used to solve continuous optimization problems. Gradient-based methods compute the objective functions' derivatives in order to follow the direction of descent (or ascent) of the considered functions and lead to optimal solutions. Their main drawbacks are that they require access to the gradient and that they do not guarantee global optimal solutions. On the other hand, heuristic (and meta-heuristic) methods do not require the gradient and are capable of finding global optimal solutions, however, they require higher computational cost compared to gradient-based methods. Also, surrogate modeling methods which replace expensive objective functions with cheap approximation models in the optimization process, have gained popularity in recent years for dealing with computationally costly applications.

The third important aspect is taking into account uncertainties for which there are again two main methodologies. The above-mentioned RBDO is based on considering the effect of uncertainties as a limit-state function, in other words as a constraint which must be avoided. The other main method, "robust design optimization", considers the effect of uncertainties as an optimization objective which must be minimized. RBDO has been largely predominant in the past few decades, but is computationally demanding when applied to real engineering problems and is better suited for cases where the gradient is available. Robust optimization has been far less studied, but presents itself as a conceptually attractive method as it has the potential to provide a set of Pareto optimal solutions which balance performance objectives and robustness.

In light of the current state of design optimization of mechanical systems under uncertainties, the presented PhD research places great importance on the following ideas:

- Dealing with the gradient-free character of performance metrics without using gradient approximation methods which are very costly
- Mitigating computational cost while guaranteeing sufficient accuracy of the proposed optimal solutions, in the context of uncertainties which automatically make the computational cost substantially higher
- Proposing methodologies for solving uncertain gradient-free design optimization problems with an efficient combination of modern numerical tools

There is a plethora of mechanical systems used in everyday life which exhibit complex behavior and whose performance must be evaluated using numerical tools leading to costly gradient-free objective functions. The disc brake systems present in automotive, rail and aeronautic vehicles are one such system. Disc brake systems exhibit squeal noise which poses a challenge for manufacturers who have been searching for ways of reducing or eliminating such noises. The vibrational phenomenon which leads to squeal noise, known as friction-induced instability, is extremely complex, highly nonlinear, and has not yet been fully understood although it has been the subject of considerable research efforts for several decades. There are different numerical ways of predicting and quantifying squeal noise, which require intricate modeling strategies and often lead to nonlinear, blackbox, and costly figures of merit. Design optimization of a disc brake system for squeal noise minimization has been studied previously in the deterministic case, but robust optimization has not yet been explored in depth. Studies on the effect of various sources of uncertainty (for instance sensitivity analysis-based) on squeal noise propensity have nonetheless been carried out. The state of current research on numerical modeling of squeal noise in disc brake systems and related uncertainties presents the perfect opportunity to study a full robust shape optimization process under squeal noise minimization criterion, as an application of robust design optimization under nonlinear expensive gradient-free performance metrics.

The presented PhD research is organized as follows. Chapter 1 gathers the literature relevant to our research, which includes background on friction-induced instabilities and disc brake squeal noise modeling techniques, multi-objective optimization of blackbox functions, geometrical description methods for shape optimization, and uncertainty quantification in the context of optimization. Chapter 2 presents the numerical modeling strategy for the disc brake squeal noise problem. First the geometrical model of the disc brake and underlying mathematical framework are presented, followed by the equations of motion necessary for friction-induced instability analysis to lead to the definition of the squeal noise quantifying criterion which will be used as the main figure of merit for the considered application. Additionally, some physical insight and validation elements for the developed model are given. Chapter 3 describes the robust shape optimization problem tackled in our research, including input parameters, objective functions and constraint functions. Chapter 4 introduces a first strategy for robust shape optimization applied to the disc brake problem based on the use of Polynomial Chaos Expansions and specifically their sparse counterpart for uncertainty quantification, combined with a genetic algorithm in order to search for solutions balancing robustness and squeal noise minimization. Chapter 5 introduces an alternative strategy for robust shape optimization, based on a variant of the Kriging surrogate modeling technique originally designed to consider measurement error, in an attempt to take into account geometrical uncertainties, applied to the disc brake problem. Finally, the conclusion discusses important takeaways and perspectives of the presented research.

1 State of the art

This chapter is a summary of the current literature in the fields related to the topic of the PhD research. It contains background on multi-objective optimization of blackbox functions, shape optimization, uncertainty quantification and optimization in the presence of uncertainties, and friction-induced instability phenomena and modeling. It also explains where the presented PhD research is positioned with respect to the current literature.

1.1 Multi-objective optimization of blackbox functions

Optimization problems which involve nonlinear blackbox objective functions which are expensive to evaluate, present several challenging aspects which make them particularly difficult to solve. Blackbox functions have no analytical expression and their derivatives cannot be computed, which eliminates any gradient-based optimization methods. Population-based methods, which require more objective function evaluations to reach global optima, must be considered. However, the high computational cost of each function evaluation is a bottleneck when coupled with such population-based algorithms.

1.1.1 Multi-objective optimization

Multi-objective optimization is an area of mathematical optimization which deals with problems involving the simultaneous minimization or maximization of more than two or more objective functions. The most common way to formulate a multi-objective optimization problem is:

$$\min_{\mathbf{x}_j \in S} (f_1^{obj}(\mathbf{x}_j), f_2^{obj}(\mathbf{x}_j), ..., f_{k_{obj}}^{obj}(\mathbf{x}_j)) \tag{1.1}$$

Where $\mathbf{x}_j$ is a potential solution (usually a vector with size corresponding to the number of input parameters), S is the space of input parameters, and $f_1^{obj}, f_2^{obj}, ..., f_k^{obj}$ are the objective functions. Also, note that maximization of an objective function can always be expressed as the minimization of its negative. The objective functions are usually antagonistic, which means that there is no single solution which optimizes all objectives. Rather, there is a set of solutions which represent different trade-offs between the conflicting objective functions. In order to find the best set of solutions, the notion of domination is used, which allows one to rank different potential solutions to a multi-objective problem. A solution $\mathbf{x}_1$ is said to dominate a solution $\mathbf{x}_2$ if:

$$\begin{aligned} &\forall i \in 1, ..., k, f_i^{obj}(\mathbf{x}_1) \leq f_i^{obj}(\mathbf{x}_2) \\ &\exists i \in 1, ..., k, f_i^{obj}(\mathbf{x}_1) < f_i^{obj}(\mathbf{x}_2) \end{aligned} \tag{1.2}$$

A “Pareto optimal” solution is defined as a solution which is not dominated by any other solution, or in other words, it is a solution which cannot be improved in any of the objective functions without degrading at least one of the other objective functions. The set of all Pareto optimal solutions (made of vectors of input parameter values) is called the “Pareto set”, while the set of objective function outputs corresponding to these inputs is called the “Pareto front”. Constraint functions can also be considered in multi-objective optimization. Solutions which violate at least one of the constraint functions are called “non-feasible” solutions. There are many types of methods for the resolution of multi-objective problems and their classification can be challenging [1]. One classification consists in differentiating two main types of algorithms: gradient-based methods and gradient-free methods. This classification is specifically relevant in this PhD work since we focus on applications with gradient-free objective functions. Gradient-based methods are generally faster than gradient-free methods, but since they require the gradient and do not guarantee convergence to global optimal solutions, we do not consider them in this research work. Among gradient-free methods, meta-heuristic methods have gained popularity in the past years for their efficiency in many types of difficult problems with large and complex search spaces, discontinuous Pareto fronts, and for their ability to reach global near-optimal solutions. The main idea of meta-heuristic methods is to iteratively improve an initial solution, by using a set of rules or “heuristics” to modify the solution. The process is repeated until a satisfactory solution is found, and in practice, several potential solutions are iteratively improved at the same time. Many meta-heuristic methods are inspired by natural processes such as animal behavior, evolution, and genetics. Some examples of popular algorithms include the particle swarm algorithm [2], the bat algorithm [3], Non-dominated Sorting Genetic Algorithm II (NSGAII) [4], artificial bee colony algorithm [5], and simulated annealing [6]. A comprehensive overview of meta-heuristic methods can be found in [7]. In this PhD work, we opt for genetic algorithms for Pareto front search.

1.1.2 Bayesian optimization

Bayesian Optimization (BO), also known as Efficient Global Optimization (EGO) is an approach designed to overcome the issues related to expensive blackbox objective functions, which aims at finding a global optimum in a minimum number of steps through surrogate modeling paired with an appropriate acquisition function. It was first introduced by Mockus [8, 9] and was then popularized as EGO by Jones [10]. It is based on the idea of prior and posterior in probability theory, formalized by Bayes' theorem. The objective function is treated as a random function and a prior belief is placed over it which captures the beliefs about the function's behavior. Then a posterior which better approximates the function is constructed using new samples drawn from the objective function. This mechanism requires some technique to approximate the function based on a restricted number of expensive function evaluations. The approximation model is usually referred to as "surrogate model" or "metamodel". The process is iteratively repeated, and at each iteration new samples drawn from the blackbox function are used to update the surrogate model. These samples are determined through a sophisticated selection method called acquisition function or infill sampling criterion. At each iteration, the optimum as given by the current surrogate model is determined using some optimization algorithm, and the acquisition function is then used to determine new sample points to enrich the surrogate model. In order to get closer to the global optimum at each iteration, the surrogate model must be improved where the quality of solution is high, but also where the objective function has not yet been sampled. The acquisition function is designed to find the compromise between so-called "exploitation" of promising areas and "exploration" of untouched areas of the design space. The different steps of BO can be summarized as follows:

1. Construct an initial surrogate model based on blackbox samples well-distributed throughout the multi-dimensional design space
2. Find the optimum of the surrogate model using some efficient optimization algorithm
3. Determine the next query point(s) using the acquisition function
4. Update the surrogate model with the new query point
5. Repeat steps 2 to 4 until some convergence criterion is met

A popular choice of metamodeling technique for engineering applications is the Kriging method [11, 12, 13] also known as Gaussian process regression (GPR) [14] (depending on the context), which is based on the assumption that there is spatial correlation between function samples, and whose main advantage is its built-in uncertainty estimation for each prediction, which is particularly useful for the acquisition function. Other metamodeling techniques such as polynomial response surfaces [15, 16], artificial neural networks [17, 18], support vector machines [19, 20, 21], have also been somewhat successfully applied to engineering optimization problems for computational cost reduction.

For the selection of new samples, there are three main well-known acquisition functions: upper confidence bound (UCB), probability of improvement (PI), and expected improvement (EI). UCB contains explicit exploitation and exploration terms and is easy to tune towards one component or the other. PI calculates the probability of improving the current best estimate without taking into account the magnitude of the improvement whereas expected improvement uses the expected value of the improvement which contains information about the quality of the mean value of the Gaussian process as well as the associated uncertainty at the considered point. Since being proposed in 1978 [22] and popularized by Jones [10], EI has become the most prevalent acquisition strategy and the standard in BO [23, 24, 25, 26, 27, 28].

1.2 Shape optimization

Shape optimization encompasses methods which aim to find the shape of a system which minimizes a certain cost or performance function, while possibly satisfying some constraints. The way that the geometries of mechanical parts are described and modified is one of the key aspects of shape optimization for mechanical design. The precursor of shape optimization for mechanical design is sizing optimization, where the geometrical dimensions of mechanical parts are taken as optimization parameters. Sizing optimization has been extensively researched for various engineering applications [29, 30, 31, 32, 33]. It has in particular been used for the minimization of squeal noise in disc brake systems [34]. Sizing optimization can at best be considered as the most basic form of shape optimization, since the shape is always predefined and only the dimensions are modified, which highly restricts the range of potential design possibilities. Shape optimization is a class of methods which seek to minimize a certain cost function (while possibly satisfying given constraints) by modifying the shape using smooth modifications of the mechanical parts' domain boundaries. In order to control and affect the shape within the optimization process, parametrization of the geometry is necessary, and needs to be combined with mechanical design and analysis techniques.

1.2.1 Finite Element Method-based shape optimization

So far, the most commonly applied strategy has been to describe the shape within the FEM framework, which is the most prevalent tool for mechanical analysis. The idea in this case is to take the FEM model's nodal locations as geometric parameters and to displace them by scalar amounts in order to modify the shape. Because this approach allows the modification of many individual boundary-defining points, it can potentially yield diverse and original shapes. Typically, this method is performed on the CAD geometry with full automation of the subsequent meshing process.

This strategy has been applied to many structural mechanics engineering applications [35, 36, 37, 38, 39, 40, 41, 42], however, it comes with some major disadvantages. The primary issue is the inconsistency between the design and analysis geometries. The design geometry, usually spline-based, is transformed into a discretized analysis-compatible geometry made of elements and nodes. This difference poses problems in terms of accuracy [43, 44] as any CAD optimized shape is selected based on the performance of the analysis geometry. Another important issue is that re-meshing the analysis geometry is necessary every time the CAD geometry is modified to avoid accuracy problems due to mesh distortion. Since meshing accounts for a large portion of computational cost in FEM, re-meshing becomes a major bottleneck [45, 46]. Other issues linked to mesh topology maintenance [47, 48] and to the order of continuity of approximation functions [47] can lead to undesirable accuracy problems for sensitivity calculations. To address the difficulties related to shape optimization with traditional FEM, various new methods, mostly variants or extensions of FEM-based shape optimization, have been proposed.

Mesh morphing [49, 50] is a technique which resolves re-meshing related problems by defining the shape parametrization after the meshing step, on the analysis geometry. Mesh morphing techniques can be either mesh-based, where computational node locations are updated while maintaining mesh topology, or meshless, where mesh topology is ignored and grid cell points are modified. Several algorithms have been proposed for mesh morphing such as radial basis function, deforming volumes, or pseudosolids. The main advantages of mesh morphing are the gain in computational efficiency, the conservation of mesh quality, and the fact that no nodes or elements are added or removed during the process. Disadvantages include limited shape modification possibilities in the mesh-based case and grid cell distortions in the meshless case. Adaptive mesh refinement (AMR) is a strategy which has been used for decades [51, 52, 53, 54] to avoid mesh distortion and ensuing degradation of analysis results in FEM-based shape optimization processes. Three types of refinement are commonly used, p-, h- and r-refinements. h-refinement controls element size and is applied by decreasing the size of the elements while keeping polynomial order fixed. p-refinement allows more shape distortion and fixed mesh topology by using coarse elements and increasing polynomial order [55]. r-refinement consists in node repositioning and is usually used together with either p- or h-refinement. AMR works best when coupled with error estimation to limit refinement to areas of the geometry where it is needed [56]. The main drawback of AMR is that it requires smooth connection between CAD and analysis geometries which makes it difficult to apply to industrial applications. XFEM is a technique initially proposed to deal with crack problems and model cracks independently from the rest of the mesh to avoid remeshing issues [57]. It was later combined with the level set technique, a numerical moving interface tracking method, to be applied to shape optimization in order to increase the flexibility of the geometric description while avoiding remeshing [58]. Fixed grid techniques are another class of methods used to address mesh distortion issues. The idea is to cover the initial shape of the system with a Finite Element discretized grid of quadrilateral elements, with each element being located either inside, outside or on the boundary of the design space. When the shape is changed, the material properties of the FE elements are altered to reflect the varying boundary of the geometry and how much of the structural shape is contained in each grid cell. Fixed-grid FEM (FG-FEM) [59, 60] and Interface Enriched Generalized FEM (IGFEM) [61, 62, 63] are two examples of such methods. Finally, the Finite Cell Method (FCM) [64, 65] combines a fictitious domain of adaptive quadrature points for geometry description and approximation of the numerical solution with an unfitted structured mesh of higher-order finite elements. Its advantages are that it does not require boundary conforming meshes, and that it can achieve high accuracy with refinement (mesh and polynomial order). It has been shown capable of integrating complex geometric models into FEM analysis.

1.2.2 Isogeometric Analysis-based shape optimization

Isogeometric Analysis (IGA) is a variant of FEM proposed by Hughes [66] which smoothly connects the design and analysis processes by employing the shape functions used to describe the CAD geometry for the analysis step as well. Since FEM-based shape optimization suffers from issues due to the lack of a seamless connection between design and analysis, IGA appears as an elegant solution. The building blocks of CAD have been integrated in FEM in the past [67, 68], and the progress in this area of research led to the development of IGA. CAD geometric descriptions usually make use of different types of spline functions, most commonly Non-Uniform-Rational-B-Splines (NURBS). In IGA, these functions are reused in the analysis step to approximate the solution across the domain. In addition to the smooth connection between design and geometry, IGA presents several advantages. The analysis geometry is not an approximation of the real geometry, it is the exact geometry. There is no need for complex meshing procedures, which means that overall modeling time is greatly reduced compared to FEM. Accuracy of solutions is also generally improved compared to FEM due to the higher order of the basis functions. The benefits of IGA not only make it a good alternative to FEM for structural analysis in many engineering applications, but also make it particularly well-suited to shape optimization because the geometry being optimized is directly the design geometry. The use of splines also makes it possible to generate very original yet smooth potential shape solutions. The application of IGA to shape optimization problems has already received considerable attention [69, 70, 71, 72]. The main drawback of IGA is related to mesh refinement and particularly local refinement, which generally requires splitting the IGA geometry into sections called “patches” in order to apply different levels of refinement to different patches. The connection of nonconforming patches then becomes a difficult challenge [73, 74]. An alternative to NURBS called T-splines which allow local refinement without the nonconforming patch issue, appear as an attractive option in this regard [75, 76].

1.2.3 Topology Optimization

Opposite to sizing optimization, Topology Optimization (TO) is a mathematical technique which spatially optimizes the distribution of material within a bounded domain while keeping total freedom in the shape definition. TO is the purest form of shape optimization in the sense that the design space is explored to its fullest and it does not depend on the initial configuration. Since it was initially proposed in 1988 [77], several variants of TO have been proposed, which use different formulations for shape description. Density-based TO applies penalization to the mechanical properties of each element of a FE mesh to find the distribution of material which optimizes the objective function. The Solid Isotropic Material Penalization method (SIMP) [78, 79] is the most common variant of density-based topology optimization and has been implemented in various commercial software [80, 81, 82]. Evolutionary Structural Optimization (ESO) is a different strategy which implements topology optimization using heuristic strategies to progressively remove material volume [83]. Bi-directional Evolutionary Structural Optimization (BESO) is an improved version of ESO where material can also be added thus overcoming some of the shortcomings of ESO [84]. Another class of topology optimization methods is the so-called Level-set Method (LSM) which describes and smoothly modifies the structural boundary of the shape using level set functions and iteratively solving the Hamilton-Jacobi equation to find the optimal configuration [85]. LSM has notably been shown to reduce the number of design variables and mitigate boundary problems compared to density-based topology optimization [86, 87].

In light of its freedom of shape description, TO continues to be applied to engineering problems in diverse fields [88, 89, 90, 91, 92, 93]. Nonetheless, TO suffers from one important drawback which is the difficulty to manufacture the optimized shapes and the necessity for adjustments to improve manufacturability. Although progress in additive manufacturing (AM) [94, 95], which overcomes the limitations of traditional manufacturing by joining material layer by layer, has made it more and more feasible to manufacture shapes originating from TO processes [89], TO is still considered and used as a conceptual design method.

1.3 Uncertainty quantification

The relevant background and literature on uncertainty quantification in the context of optimization in the presence of uncertainties is introduced in the following sub-sections.

1.3.1 Uncertainty classification, quantification and propagation

There are multiple sources of uncertainties in real-world systems and their numerical models. Classifying uncertainties can be a challenging task, but is important in order to apply the appropriate uncertainty quantification and propagation techniques to a certain problem. Mechanical systems can be described by numerical models and associated parameters. However, the value of these parameters is not usually precisely known. Taking into account uncertainties in different parts of a mechanical system can lead to considerable fluctuations in performance predictions, and therefore should be carried out with care. Some examples of common uncertainty sources are:

- model input parameters which are not precisely known or which can vary. In our case geometric shape parameters can vary due to manufacturing errors (between model and reality). The material properties are also an example of uncertain parameters
- experimental uncertainties can arise when uncertain experimental measurements are used in a modeling strategy
- numerical models are based on equations which are approximations of reality and represent uncertainty linked to insufficient knowledge
- numerical uncertainties are introduced in the model results when numerical methods are used to solve the model equations

There are many more ways to describe uncertainty sources, and it is important to understand and characterize the types of uncertainties we are dealing with before selecting the methods used to take them into account. It can also be useful to compare the importance of different uncertainties in order to prioritize the predominant sources.

Generally, uncertainties are classified into two main categories: epistemic and aleatory. Epistemic uncertainties are due to a lack of knowledge, and can be reduced by adding more knowledge to the model, but this can be complex or computationally expensive. On the other hand, aleatory uncertainty is linked to the unpredictable character of outcomes of an event which is treated as random. Aleatory uncertainty can usually be described by a probability distribution function. Figure 1.1 shows an example of uncertainty quantification centered on the distinction between epistemic and aleatory uncertainties. In the context of the probability theory and statistics of uncertainty quantification, there is not necessarily a need to differentiate between these two types of uncertainties because the main available tools are the same in both cases (stochastic modeling and analysis of uncertainty propagation). A more important distinction should be made between parameter uncertainties and modeling uncertainties. Designing a mechanical system usually starts from a conceptual or theoretical system, which is then numerically modeled, and leads to a final real-life manufactured system. Modeling approximations are inevitable and create differences in system behavior between the computational model and the manufactured system. According to [96], approximations inherent to the modeling method (e.g. FEM) should not be considered as uncertainties, however, errors due to model reduction (which is often necessary to reduce computational cost) should be taken into account as uncertainties. On the other hand, parameter uncertainties are directly related to parameters of the numerical model, for example geometric parameters whose value may differ between the model and the real system, due to the manufacturing process. Nonetheless, in general, it is difficult to define precise guidelines on the way to take into account different types of uncertainty, which usually depends on the context of each studied application.

Classical uncertainty quantification and propagation methods are usually classified into two types: probabilistic and non-probabilistic. In both cases the idea is to first characterize the uncertainties, then inject them into the model and propagate them to the model outputs. Probabilistic methods aims at describing the uncertainties with probability distribution functions. Non-probabilistic methods use the interval information of random variables to characterize uncertainty. Non-probabilistic methods are commonly better suited to incomplete, imperfect data and knowledge. For example, interval arithmetic defines the parameters as intervals. Fuzzy logic, possibility theory, and evidence theory are types of non-probabilistic methods.

In the context of computationally expensive models, classical methods can be prohibitive because they often require many model runs, which is why surrogate model-based methods have been introduced and are now widely used for uncertainty propagation. Surrogate-based methods are sometimes presented as separate from classical methods, but they are also either probabilistic or non-probabilistic. Also, certain surrogate modeling techniques contain built-in statistical properties and tools which can be advantageously used for uncertainty propagation.

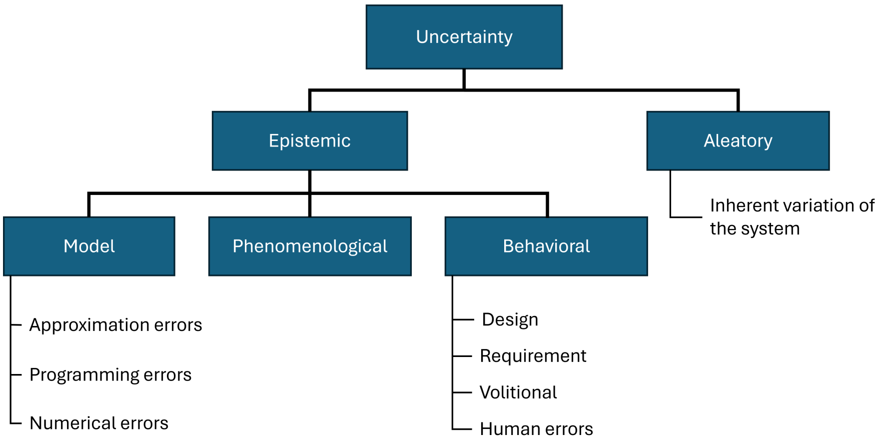


Figure 1.1: Example of uncertainty classification

1.3.2 Reliability-based approaches

Deterministic models are used when uncertainty is relatively small. When uncertainties are large, stochastic models are needed. The deterministic approach requires a safety factor and results in an under- or over-designed system. On the other hand, the stochastic approach uses the statistical properties of the system response (mean value, variance, confidence interval) to describe it comprehensively, and results in a robust system design. Probability theory has been incorporated into the engineering community for complex systems thanks to high-powered computers, which enable the combination of traditional analysis methods with uncertainty quantification. Reliability-based approaches are founded on the concept of reliability which is comprehensively discussed in [97]. Reliability can be simply defined as "the probability that a system will perform its function over a specified period of time and under specified service conditions" [98], and the system design is deemed appropriate when each requirement on the output structural behavior is met within a sufficient degree of certainty.

Structural reliability approaches are based on the calculation of the probability of violation of the limit-state function(s) of a structure during its life, the goal being to minimize this probability through design choices. Such approaches yield structures which account for uncertainties and thus have a lower probability of failure than if designed with deterministic methods, at equivalent cost. The limit-state is defined as the limit of a structure which when exceeded renders the structure unable to perform as required. If the probability of failure is above a predefined required value the structure is deemed unreliable. Any system can have multiple limit-state functions depending on the performance objectives defined for design. There are two main types of limit-states. So-called "ultimate" limit-states are related to a collapse of part or all of the structure, whereas "serviceability" limit-states, if violated, entail only disruption of the normal use of the structure. The limit-state function f_{ls} also indicates the margin of safety between the resistance R_{st} and the load L_{st} of a structure. The probability of violating the limit-state is denoted as the probability of failure P_f. Thus, three regions can be defined according to the value of f_{ls}:

- $f_{ls} > 0$ safe region
- $f_{ls} = 0$ failure surface
- $f_{ls} < 0$ failure region

These three regions are illustrated in figure 1.2. From the mean and standard deviation of function f_{ls}, the reliability (or "safety") index can be computed. This quantity indicates the distance of a design from the failure surface $f_{ls} = 0$ and provides a good measure of reliability. The safety factor $F_{st} = \frac{R_{st}}{L_{st}}$ is another well-known, simpler parameter for reliability analysis.

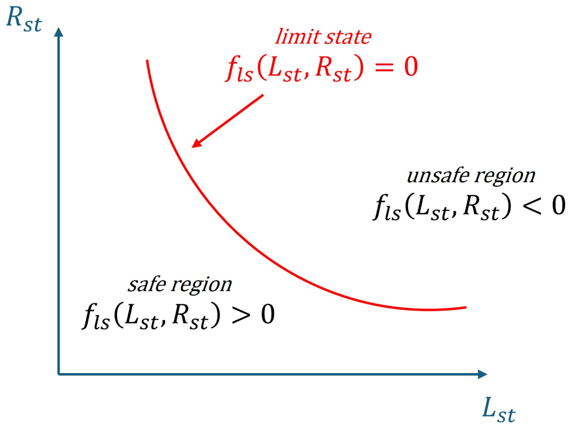


Figure 1.2: Illustration of the limit-state in reliability-based approaches

Random field

There are many sources of uncertainties in engineering structures (loads, materials, geometry, manufacturing, ...) and these uncertainties fluctuate over space and time, which means the responses of the structure fluctuate accordingly. The resulting space and time-varying data must be estimated to be incorporated in realistic simulations which consider uncertainty analysis. The random field is what characterizes the mathematical model of the variability which is parametrized by the correlation between the different locations. Random fields [99] are random functions of one or more variables, which can be used for instance to characterize distributed property inputs in structural problems. The goal is to realistically model the variability of the inputs, in order to facilitate the estimation and analysis of the corresponding random outputs. This should allow to assess the allowable bounds of the inputs to guarantee safety bounds for the outputs. Deterministic analysis is not sufficient to characterize the system under varying input properties. The random field or random process is what characterizes the spatial variability. The main idea when manipulating random processes is that the outcome of one experiment is not one point, but a function over an interval of the domain (or intervals of multiple domains). The resulting function is called a realization of the random process. A collection of realizations is called an ensemble. When the random process is based on a set of samples over an interval, n random variables and their joint probability distributions can be used. The moments of the joint probability distribution are defined (mean, auto-covariance, auto-correlation). The concept of random field is thus very important in reliability-based approaches. A random process can be thought of as a random function which is more difficult to manipulate than random variables. For this reason, in practice, random variables are generated using deterministic functions with random coefficients, which is equivalent to discretizing the corresponding random process, which becomes an indexed set of an infinite number of random variables. The proper discretization, characterization and representation of the statistical correlation between each random variable is critical and can be specified by covariance functions. Many different methods have been proposed to implement the random variable-based approach [100], the most prevalent being Taylor series representation, and orthogonal polynomial expansion.

First and second order reliability methods

The origin of these methods stems from the curse of dimensionality in the calculation of the probability of failure which requires simplification of the analytical integral expression. This can be done by linearizing the limit-state function at the failure surface $f_{ls} = 0$ using Taylor expansions. Initially, this idea led to the implementation of the so-called First Order Second Moment method (FOSM) and the Second Order Second Moment method (SOSM), which respectively use the first-order and second-order Taylor expansions of the limit-state function. Generally, FOSM is easy to implement but it was shown to be inaccurate for small probabilities of failure and for highly nonlinear responses. SOSM is computationally more expensive because of the added term, though the accuracy improvement is in most cases minimal.

FOSM was improved based on a geographical interpretation of the safety index by Hasofer and Lind [101] which lead to the notion of most probable failure point (MPP) by using a transformation of the design variables from the original space to the space of normalized independent variables (HL transformation), which yields the more commonly known First Order Reliability Method (FORM) [102] and Second Order Reliability Method (SORM), which respectively rely on first and second order approximations at the MPP. The basic idea is that the limit-state is approximated by a tangent plane at the MPP, then a bound can be specified based on the probability of failure. If the approximation at the MPP is accurate then the bound will lead to good results. Otherwise, inaccurate approximation can lead to large errors (example: nonlinear failure surface).

There are multiple variants of the FORM method depending on different factors. For instance, several options are possible to solve the reliability index minimization problem (primal methods, penalty methods, Lagrange multipliers, ...). The HL-RF method is designed to deal with cases where the random variables are not randomly distributed by transforming them to an equivalent normal distribution [103, 104, 105]. Two-Point Adaptive Nonlinear Approximation methods (TANA) [106] can help accelerate convergence for nonlinear cases. FORM also includes built-in sensitivity factors which provide interesting insight into the random variables. Firstly, they show the relative contribution of different random variables to the failure probability, and secondly, the sign of the sensitivity factors indicates the relationship between the performance function and each physical variable (increasing or decreasing).

When the failure surface has high nonlinearity in the neighborhood of the design point, FORM can produce inaccurate results [107], in which case SORM [108, 109, 110, 111] may sometimes be more effective. Instead of the space of normalized independent random variables, the “rotated new standard normal” space is used, which facilitates the integration of the probability of failure in the multi-dimensional case. Since this method requires computation of the second derivatives of the limit-state function, it is computationally more expensive. Breitung’s formulation [108] yields the probability of failure from the main curvatures of the failure surface. SORM also has multiple variants depending on the details of the resolution strategy. Tvedt’s formulation [111, 112] is an alternative to Breitung’s formulation based on a more accurate three-term approximation based on a power series expansion. SORM with adaptive approximations [113] alleviates computational cost by using approximations of the second order derivatives, which also makes it suitable for implicit performance functions.

Reliability-based Design Optimization

Reliability-based Design Optimization (RBDO) is an approach which encompasses the reliability concepts presented in the previous sub-sections and uses them to solve optimization problems subject to uncertainties [114, 115, 116, 117, 118]. The main idea is to optimize system performance functions under probabilistic constraints such as a limit on the probability of failure or reliability index. This method was first applied to systems using classical methods for solving constrained optimization. These approaches can be divided into two main groups, direct approaches which handle constraints in an explicit manner (generalized reduced gradient method, feasible directions method), and indirect approaches where the constrained problem is solved as a sequence of unconstrained minimization problems (penalty methods, Lagrange multipliers). RBDO can nonetheless be applied with various methods such as evolutionary algorithms, as long as the constraints for reliability are well taken into account. Compared to the deterministic optimization case, an RBDO solution not only provides an optimized design with respect to the performance functions, but also a design with a high degree of confidence. One of the main drawbacks of RBDO is related to the curse of dimensionality which restricts its use in higher dimensions. Also, calculation of the limit-state is costly and requires the gradient information. Another problem is that even though the retained optimal solutions will have a guaranteed minimal level of reliability, the method does not focus explicitly on finding solutions with the highest level of reliability or robustness.

1.3.3 Sampling methods

Sampling methods consist in the direct use of experiments to obtain mathematical solutions or probabilistic information for difficult-to-solve systems, and are usually paired with the strategies described in sections 1.3.2 and 1.3.4 for improved efficiency.

Monte Carlo Simulation (MCS) [119], also known as "simple random sampling" is the most basic and straightforward sampling method available and consists in computing realizations based on randomly generated large sampling sets for uncertain variables, which became practical for structural numerical analysis with the advent of digital computers. MCS is a powerful tool to determine the approximate probability of a specific event which is the outcome of a series of stochastic processes. It consists in digital generation of random variables for subsequent statistical analysis of trial outputs. The total number of samples and boundary limits are important factors which affect the efficiency of the method. MCS can be used to directly compute the probability of failure in reliability-based methods, for instance as the ratio of the number of failed samples (violated the limit-state) by the total number of samples. It is worth using when the number of trials of simulation is less than the number of integration points required in numerical integration. More generally, MCS is used in optimization as a brute-force reference result with guaranteed converged results given a sufficiently large number of samples. Random selection of points assumes that the function will be well-represented in an unbiased manner. For a large number of random variables, a comparatively large sampling set is required for convergence, which leads to prohibitive computational effort.

To predict risk accurately, especially when the probability of failure is relatively small, MCS becomes inefficient. The tail of the distribution is the most important factor. Concentrating sampling in this part can be a good strategy to decrease execution times and minimize computer storage requirements. Variance reduction techniques (importance sampling, systematic sampling, stratified sampling, ...) [120] are a way to deal with the slow convergence of MCS. Importance sampling, for instance, introduces the importance sampling probability density function (PDF). This function has to be chosen in such a way that an unbiased estimate of the probability of failure can be obtained with many fewer observations than in the case of the regular PDF-based calculation. The ideal choice of importance sampling PDF can be obtained through calculus. Another popular alternative to MCS is Latin Hypercube Sampling (LHS) [120, 121, 122], which is an extension of the Latin Square concept to higher dimension, also known as "stratified sampling". LHS guarantees non-overlapping designs, with a fairly straightforward construction process, based on dividing the distribution of each random variable into equal probability bins. The regularity of the bins ensures that all portions of the range of each random variable are well represented and results in small variance in the response. LHS is less expensive to generate than MCS and allows more flexible sample sizes while ensuring stratified sampling.

1.3.4 Stochastic expansions

Stochastic expansions are a tool which represent uncertainties of systems by using polynomials for the characterization of the stochastic system, and come with useful convergence properties in the context of stochastic analysis [123]. Stochastic expansions originally stem from the study of irregularities in Brownian motion (also called "Wiener process") which lead to the "multiple stochastic integral with homogeneous chaos" method [124]. Then, it was shown that any stochastic process can be described in terms of the Wiener process which is the simplest stochastic process, concept which is at the root of the Polynomial Chaos Expansion (PCE) method. Stochastic expansions can be categorized into two main types, intrusive and non-intrusive formulations. In the intrusive formulation the representation of uncertainty is explicitly expressed within the analysis of the system, and thus requires access to the structural analysis (e.g. FEM). The non-intrusive formulation uses results from the structural analysis to characterize the stochastic system's behavior, and consequently the structural analysis can be considered as a blackbox function in this case.

PCE is nowadays the most well-known non-intrusive stochastic expansion method and uses orthogonal polynomials of random variables, which are most of the time standard normally distributed. PCE is convergent in the mean square sense and enables a simplified calculation of the statistical moments due to the fact that any order PCE consists of orthogonal polynomials. Probabilistic collocation methods are a variant of stochastic expansion-based methods where responses of stochastic systems are projected onto the PCE [125]. Their limitations for large-scale models led to the stochastic response surface method which uses partial derivatives of the model outputs with respect to the model inputs [126]. Stochastic approximation is a non-intrusive approach which uses PCE to create a surrogate model of the stochastic responses. For regression of unknown or complex nonlinear relationships, common polynomial models can be used, but the use of orthogonal polynomials (i.e. PCE) is better in order to avoid ill-conditioned problems. The added advantage of PCE lies in the direct access to various statistics including mean, variance, confidence interval, etc. The main idea of stochastic approximation is to select an appropriate basis function to represent the response of uncertain systems. Cases where the random variables are not standard normal can be handled in two ways. Firstly, the Askey scheme can be used to determine which type of polynomials is more appropriate according to the PDF of the considered random variables [127, 128], as several orthogonal polynomials have orthogonality weighting functions which match well-known PDFs. The Askey scheme classifies the hypergeometric orthogonal polynomials and indicates the limit transition relations between them (Laguerre and Hermite polynomials can be obtained from Jacobi polynomials). The second option is to transform the random input variables into standard normal random variables, then the original PCE can be applied [129, 126]. Determining if a constructed PCE model is acceptable is an important step of the stochastic expansion process and can be done with the help of tools such as analysis of variance (ANOVA) and residual plots. One major drawback of PCE is that it can be computationally prohibitive when there are many input parameters and when the numerical model is expensive to evaluate. So-called sparse PCE methods [130, 131] offer a potential solution to this problem by truncating the PCE to keep only the significant coefficients. Least angle regression PCE (LARS-PCE) is an example of such methods [132]. Figure 1.3 shows an example of the comparison of sparse and non-sparse PCE coefficient distributions [133].

Another prominent type of stochastic expansion is the Karhunen Loeve (KL) expansion which is useful when a large number of highly-correlated and/or superfluous variables are used to describe the stochasticity of a system which causes a decrease in accuracy and reliability of the predictions. The KL expansion provides a way of generating correlated random variables while ensuring reduction of dimensionality and minimal loss in accuracy. The KL transform stems from works by Karhunen, Loeve and Hotelling on the orthogonal transform, the Hotelling transform and principal component analysis (PCA) [134], which all employ similar concepts. After the effectiveness of LHS sampling was displayed [121], various approaches to control correlation structures were proposed [135, 136], and multiple studies showed the effectiveness of the KL transform for generating correlated random variables paired with the LHS sampling scheme [137, 138].

The Spectral Stochastic Finite Element Method (SSFEM) [139] is the most well-known stochastic expansion-based intrusive method, and consists in using PCE to represent the stochastic responses and the KL expansion to represent the random variable inputs. This method has been shown effective for different applications and their associated random behavior. The KL expansion cannot usually be used to represent the response because it requires known covariance functions to be constructed, which is most of the time not the case for response functions.

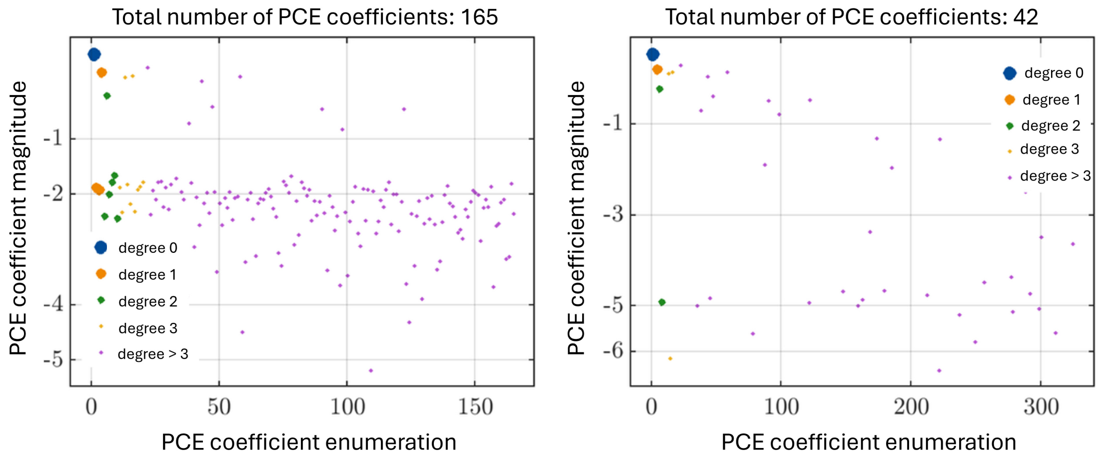


Figure 1.3: Example of comparison between ordinary (left) and sparse (right) PCE coefficient distributions

1.4 Friction-induced instabilities and squeal noise

The necessary background and literature on friction-induced instabilities and squeal noise, including general context, origin, and quantification of squeal noise are introduced in the following sub-sections.

1.4.1 Context

Friction related vibrational and tribological phenomena can lead to many different types of sounds and noises. Although some of these sounds can be intentionally produced, for instance in the case of some musical instruments, many industrial systems exhibit undesirable noises, which need to be studied and modeled in order to reduce them as much as possible. Disc brake systems are a typical example of such systems, predominantly present in the automotive, rail transport, and aviation industries. Disc brakes are a type of braking system which uses friction between different mechanical parts when squeezed together in order to slow down or stop the rotation of the rotor. Through this process the energy due to motion is mostly dissipated as heat.

In the automotive industry, disc brakes include three main components, the rotor disc which rotates with the wheels, the brake pads, which generate friction when pushed against the rotor, and the calipers which hold the pads into place and generate pressure (through actuation of a piston) to clamp the pads against the disc. Disc brake squeal noise in the automotive industry can be a problem in terms of noise pollution, but not to the extent of creating health problems for the general public. Vibrational levels associated to squeal noise in car brakes are also not sufficient to inflict structural damage on the mechanical system. Rather, noise during braking operations in new vehicles is often associated to malfunction by customers, which is the cause of many vehicle returns which are highly undesirable for automobile manufacturers from a sales point of view. Figure 1.4 illustrates the complexity of an automotive disc brake, showing a side view of its main components [140].

Disc brake systems in trains have the same basic components as in cars: a disc rotating with the shaft, brake pads, and a caliper type structure holding and compressing the disc and pads together. One notable difference is in the construction of the brake pads, which are often made of several distinct pieces of friction material (friction pins) fixed to a supporting backplate rather than a single friction material portion. In the rail transport industry, squeal noise due to braking operations can reach extremely high levels (as high as 120 dB in some cases) and is a problem from a health perspective.

In the commercial aircraft industry, more heavy-duty disc brakes are used, made of a series of several discs rotating with the wheel interleaved with stationary discs. The rotating discs and stationary discs are squeezed against each other on the axle by multiple pistons placed around the discs which creates friction and slows down the wheel rotation. Squeal noise in commercial aircraft is not a concern as noise levels from other parts of the aircraft (mostly the engine) are much higher and drown out any noise coming from the brake system.

In order to model and propose ways of reducing squeal noise in disc brake systems, it is necessary to understand the phenomena from which they originate. Friction-induced vibrations have been extensively studied throughout the years, but there is no universally accepted theory concerning the mechanisms giving rise to this phenomenon. Many experimental studies which suggest different interpretations and explanations for the presence of noise and vibrations in various systems with friction interfaces have been summarized [141, 142, 143, 144, 145, 146, 147].

Friction instabilities can arise either from so-called “forced” vibrations or “self-excited” vibrations. In this PhD research we focus on self-excited vibrations which are responsible for noise in brake systems most of the time. Different types of friction-induced noise have been classified and characterized according to their frequency spectrum and their mechanism of appearance [141, 148, 149]. For instance, “groan” is a low frequency noise induced by instabilities due to stick-slip at low rotation speeds or to friction coefficient decrease with sliding speed. “Hum” is another type of noise which appears between 100 and 500 Hz, which originates from coupling of rigid body modes. “Judder”, exhibited at low frequencies and due to forced oscillations induced by a variation of contact force at the interface, is another example of instability-related noise in disc brake systems. Squeal noise, usually exhibited at higher frequencies (1-20 kHz), is known to be linked mostly to coupling of distinct structural modes (see section 1.4.2), and can reach very high levels. In some cases, it can be difficult to distinguish and classify different instability-induced sounds because their mechanisms of appearance are not always clear and their frequency ranges can overlap. The presented applicative work focuses on self-excited friction-induced instabilities linked to mode coupling or in other words, squeal noise inducing instabilities.

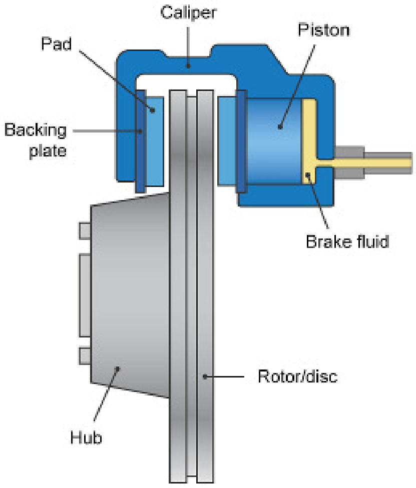


Figure 1.4: Scheme of an automotive disc brake system

1.4.2 Origin of squeal noise

The origin of squeal noise has been researched experimentally, analytically and numerically for decades and is still being studied to this day. Friction-induced instabilities have been extensively studied at different scales, using models and experimental means with greatly varying complexity, and resulting in various explanations for the origin of squeal noise. In particular, a distinction is to be made between phenomena resulting from a tribological analysis of the problem, and phenomena linked to geometrical and structural mechanics' considerations. The diverse research concerning the mechanism responsible for squeal noise occurrences in systems with friction interfaces can be grouped into four main theories.

Firstly “stick-slip”, which arises from a tribological analysis of frictional phenomena, has long been considered the main explanation for squeal noise in frictional systems. Stick-slip results from the distinction of static and dynamic coefficients of friction and consists in the alternation of “sticking” and “slipping” phases between two solids in contact, which results in self-excited vibrations. This behavior has been largely studied experimentally and through analytical models [150, 151, 152, 153, 154]. However, stick-slip has since been shown to be insufficient to explain many instances of squeal noise.

Stick-slip suggests the existence of periods of time where there is no relative movement between the two solids in contact, whereas this is not realistically compatible with the high speed nature of brake squeal phenomena. Also, several experimental studies have shown instabilities in systems with a constant coefficient of friction [155, 156].

Another tribological analysis based theory is the variation of the friction coefficient with sliding speed. Decreasing and continuous variation of the friction coefficient as a function of the sliding speed has been shown equivalent to a negative damping term which causes self-excited vibrations. This has been illustrated on a simple two degree of freedom analytical model [157].

Sprag-slip was introduced in response to the inability of stick-slip and friction coefficient variation to fully explain the instabilities which cause squeal noise, and is a geometrical interpretation of friction-induced instabilities. It was first presented intuitively through a simple system: an inclined flexible bar maintained in contact with a treadmill [144]. With a constant coefficient of friction, the tangential friction force, which approaches infinity when the inclination angle of the bar reaches a specific value, gives rise to an alternation of “sprag” and “slip” periods which in turn cause vibrations.

Lastly, mode coupling [150, 158, 159, 160, 161, 162], is now the most widely accepted theory for the origin of squeal noise. It is characterized by the coupling of the behavior along two separate directions of the system, and results in an asymmetric contribution to the stiffness matrix, which can lead to instability. It can also be interpreted as a generalization of sprag-slip to systems with a large number of degrees of freedom. This type of instability is also known as “mode lock-in” or “flutter” instability. The presented PhD work focuses on this type of instability.

1.4.3 Quantifying squeal noise

There are multiple ways of gaining insight into squeal noise generated by frictional systems, assuming the noise is linked to friction-induced mode coupling behavior. Usually, the equations of motion containing nonlinear friction and contact terms, are linearized around an equilibrium point. During this process the Jacobian of the nonlinear friction and contact terms is computed and taken into account in the stiffness matrix of the whole system, which produces an asymmetry. From this point there are two main methods for gaining information on squeal noise propensity. Firstly, computing the complex eigenvalues of the asymmetric system provides information about the stability of the system which is directly linked to potential squeal noise occurrences. Secondly, the equations of motion can be solved through time integration, and vibrational (i.e. noise) levels can be obtained after the steady state is reached [163, 147, 164, 165, 146]. The next logical step of this type of approach is then to study the acoustic waves created when the vibrations are transmitted to ambient air [166, 167, 146]. There is a crucial difference between the two types of approaches. Stability analysis assumes that the probability that noise will occur is linked to the presence and amount of unstable eigenvalues based on Lyapunov’s theory of dynamical stability [168]. In other words, it yields a level of confidence that noise will occur. On the other hand, time integration-based approaches aim to obtain actual vibrational levels. It is difficult to say that one type of method is superior to the other and the choice of approach depends more on the goals and constraints in the context of each study. One main advantage of stability analysis consists in the low computational cost comparatively to methods which integrate the equations of motion over time, and is one of the reasons why the presented research focuses on stability-based noise prediction.

The term Complex Eigenvalue Analysis (CEA) is used to refer to stability analysis based methods. CEA has repeatedly been successfully used to predict noise in frictional systems [169, 170, 171, 172, 173, 174]. Based on the linearization of the equations of motion which leads to an asymmetrical equivalent stiffness contribution, the eigenvalues are assumed complex, and their real part allows to conclude on the stability of the fixed point: if at least one eigenvalue has a positive real part, the fixed point is unstable. Instability analysis is commonly carried out with respect to a control parameter, often the friction coefficient, damping, or some geometric or material parameter. This leads to the notion of Hopf bifurcation [168], where as a function of the control parameter, the real part of an eigenvalue becomes positive at the bifurcation point, in other words, it is the transition between stable and unstable behavior. At the same time, the corresponding imaginary parts (i.e. frequencies) converge, which characterizes the coupling of initially distinct structural modes. Similar methods like the Routh-Hurwitz criterion can be used to characterize stability of smaller systems [175] but CEA is the most effective when dealing with large Finite Element models. The mathematical background for the notion of stability of a fixed point of a non-linear ordinary differential equation is well known [147] and relies on Lyapunov's theorem which ensures stability of the non-linear system given stability of the linearized system.

While CEA is an excellent tool to identify potentially problematic instabilities in the system, it has some drawbacks which should be mentioned. Since it only provides a potential of noise occurrence, it does not give any information about which modes are actually most harmful (amplitude of vibration in steady state for example) [176, 165, 177, 164]. Also, the importance of certain unstable modes of the system can be overestimated or underestimated, and some unstable modes appearing in the transient state can be absent from CEA. There are additional tools, complementary to CEA, which can be useful to help effectively avoid instabilities. For instance, harmonic balance [178, 179] and shooting methods [165] can be used to obtain stationary solutions without computing the transient state. The modal absorption index method (MAI) can also be mentioned, which eliminates modes less likely to appear in a time simulation [180]. And lastly, modal amplitude stability analysis (MASA) can predict unstable modes not calculated by CEA, but predominant in the time response [181].

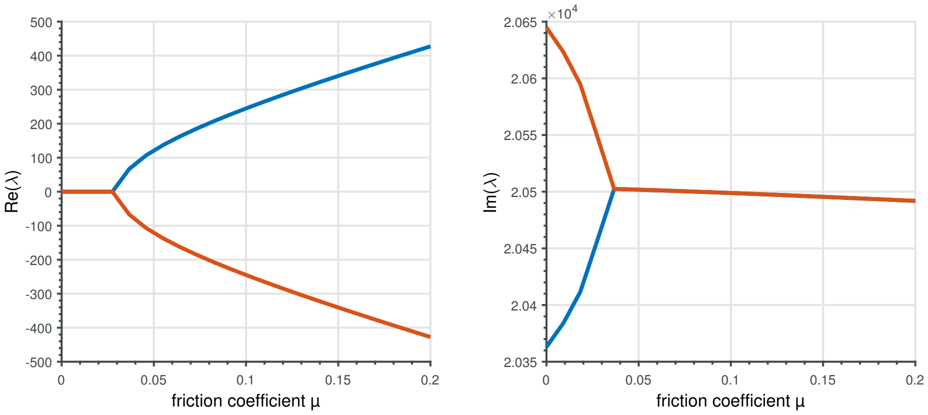


Figure 1.5: Example of a Hopf bifurcation in the case of a disc brake system with the friction coefficient as a control parameter (left-hand side: real parts, right-hand side: imaginary parts)

1.5 Situation of the PhD research with respect to the literature

The presented PhD research focuses on designing robust shape optimization methods for complex mechanical systems with nonlinear, costly, and gradient-free performance metrics, applied to the case of a disc brake system under dynamical criterion. The main quantity of interest in the studied applicative case is the measure of potential squeal noise occurrence due to friction-induced dynamic instabilities. The choice of method for its quantification is a crucial aspect which greatly impacts the behavior of the considered objective function and its computational cost. In light of its computational efficiency compared to time integration-based methods, we have chosen CEA to quantify the squeal noise propensity of the studied system. This method directly links noise propensity to the extent of dynamic instabilities and therefore consists in giving a measure of the possibility of noise occurrence rather than a measure of noise levels. Moreover, the unpredictable behavior of friction-induced instabilities makes for an extremely interesting case of complex nonlinear functions in mechanical systems.

The choice of method for shape parametrization is a key aspect in the studied problem as it does not only impact the diversity of shapes which can be obtained, but also the adequate consideration of geometric uncertainties and the structural analysis results. FEM-based shape optimization techniques greatly suffer from the issues linked to meshing such as the fact that the analysis and design geometries are not strictly the same which means that optimal solutions in terms of analysis geometry may not be optimal in terms of design (and thus manufactured) geometry. For the same reason, taking into account geometric uncertainties which affect the shape of the optimized mechanical parts is difficult if not impossible. Alternative FEM-based methods which mitigate or eliminate mesh issues are an interesting option but difficulty of implementation becomes a disadvantage. Consequently, we have decided to exclude FEM-based shape optimization in the context of this PhD research. TO is an interesting formulation of the shape optimization problem notably because of the extreme freedom for potential shapes which allows one to fully explore the optimization solution space. Nonetheless, the difficulty to manufacture shapes obtained with TO without using subsequent adjustments of the geometry make it difficult to apply in a robust optimization setting where it is highly desirable that the optimized shapes correspond to the final designs. In this context, IGA appears as a good balance between the disadvantages of FEM and TO, where meshing issues are not a problem and freedom of shape definition is moderate compared to TO while allowing more originality than FEM. Additionally, IGA offers the advantage of higher computational efficiency thanks to the elimination of the meshing steps and potentially offers a higher degree of analysis accuracy owing to the higher order of continuity of its basis functions. For these reasons, we have chosen to apply IGA for the robust shape optimization problem presented in this research.

In the context of optimizing quantities of interest in mechanical systems subject to uncertainties, reliability-based methods have been largely prevalent so far. These approaches provide the advantage of having been well-documented and tested in the past, and more importantly it guarantees a minimal level of reliability for the resulting designs. However, these approaches are also susceptible to the curse of dimensionality and in gradient-free cases or cases when the quantities of interest are blackbox functions, it can be more difficult to implement or even infeasible. From the broader perspective of robust optimization, it does not explicitly search for the most robust possible solutions but rather only for the best solutions which comply with a predefined minimal degree of reliability.

Multi-objective methods are another class of approach for optimization problems under uncertainty, where robustness to uncertainties is considered as an additional optimization objective, rather than as a constraint as in the case of RBDO. This type of approach is commonly known as “robust optimization” and yields a Pareto front of solutions which represent compromises between the performance objectives and the robustness of these objectives to uncertainties. Thus, the final design can be designated among the set of compromises, and the balance between high robustness and best performance can be freely chosen, taking into account the operating conditions and risks for the mechanical system under study. This method formulates the robust optimization problem in a way that constitutes an interesting alternative to reliability-based methods. Additionally, it provides more freedom for handling tedious cases with blackbox nonlinear functions, which is the object of our research. The main challenge resides in quantifying the robustness of the system, which requires proper quantification and propagation of the considered uncertainties. In this case the idea is to consider that the quantity of interest has statistical properties which allows one to express its first and second moments as antagonistic functions, according to the information theory (bias-variance tradeoff [182]). The multi-objective approach has been successfully carried out in cases where the gradient of performance functions was available, however, when no gradient is available, the variance of the quantity of interest is more dependent on the computational power required to reach sufficient evaluations during the optimization process. Some recent results point to the use of meta-models able to reconstruct the statistical information of the output, to mimic the classical Efficient Global Optimization scheme (EGO) [10]. Given the advantages of the more novel MOO-type approaches, the PhD research will focus on the robust optimization approach, rather than reliability-based methods.

The undertaken robust optimization problem must balance minimization of the main quantity of interest and maximization of the quantity of interest's robustness to uncertainties. With this in mind, and in the present-day context where production of mechanical systems must be carried out in a way which minimizes cost, the main challenge is to find the right compromise between computational cost and accuracy, which are antagonistic quantities. Methods based on surrogate models, which do not necessarily lead to satisfactory results as could be obtained with extremely expensive Monte-Carlo-type simulations, must nevertheless be explored and analyzed. Some drawbacks of such methods may include: generation of robust solution sets which represent only subsets of the "real" full Pareto front, and inevitable introduction of approximations in the quantification of robustness. Nonetheless, such solutions can still be effectively used for the design of mechanical systems if there is some control over the induced errors.

Among the MOO-type approaches, there are multiple options which are worth implementing and testing for the disc brake squeal noise application, depending on the way the uncertainties are considered or on the type of surrogate model. Firstly, the PhD research follows work on the same applicative system in the deterministic case [34, 183]. This research suggested that a Kriging-type metamodel was useful in describing the squeal noise propensity function derived using FEM, and could be used in the multi-objective Bayesian shape optimization scheme for minimization of squeal noise and maximization of contact surface area, with an associated enrichment strategy. Even though uncertainties were not taken into account, a Kriging-based approach is an interesting perspective for our robust optimization research in the case of a nonlinear expensive blackbox figure of merit. Particularly, the ability of so-called "noisy" Kriging [184] to take into account heterogeneous noise in the samples used to build the Kriging model, is a potentially promising prospect for treating uncertainties. Secondly, the aforementioned stochastic expansion method can be used in its non-intrusive version, as a surrogate model of the stochastic performance function and simultaneously provides statistical information on the outputs. Integrating this method into a multi-objective optimization scheme may be one of the best options for robust optimization in such conditions. Intrusive stochastic expansion approaches could be used to quantify and propagate the uncertainties directly from inside the structural mechanics numerical model. This approach has been applied to FEM models in the past but only recently in the context of isogeometric analysis [185].

2 Friction-induced instability model

In the presented PhD research, the main goal is to design robust shape optimization strategies for complex mechanical systems under gradient-free, nonlinear and costly performance metrics, applied to a simplified disc brake system under dynamical criterion. This chapter presents the modeling process used to represent the disc brake system and associated friction-induced instabilities, including elementary mathematical background of IGA, the description of the simplified disc-pad structure's IGA model, contact and friction formulation, model reduction, damping consideration, and the mechanical equations which lead to mode coupling. The criterion used to evaluate instability and noise, which is subsequently used as the main performance metric in the robust optimization process, is then introduced, followed by insight into the physical behavior and computational cost of the developed model.

2.1 Disc-pad IGA model

This section presents the geometrical model of the disc brake, based on the IGA method.

2.1.1 IGA formulation

As explained in section 1.2.2, IGA fills the gap between design and analysis in traditional FEM and provides flexibility in shape definition. In this research we use the most common formulation of IGA, which employs Non-Uniform-Rational-B-Splines (NURBS) as basis functions. Other types of basis functions such as T-splines can also be used [186]. We briefly recall the basic principles of NURBS below, more relevant details can be found in [66, 187]. IGA uses B-spline basis functions which are characterized by their order, a set of knots, and associated control points. The control points are a set of points which when linked by piecewise linear interpolation define what is called the control mesh, which can then be used to control the shape of the system. The control points also represent degrees of freedom. The physical mesh is distinct from the control mesh and consists of "knot spans" and "patches". Each patch represents a portion of the computational domain and can be broken down into knot spans which are points, lines, or surfaces depending on whether the considered geometry is in 1D, 2D or 3D. Elements of the physical mesh are defined by inserting knots into each knot span. The basis functions have C^{p-k} continuity across each knot, k being the knot multiplicity. The so-called "knot vector" is defined as:

$$\mathbf{s} = [s_0, s_1, ..., s_{n_c+p}] \tag{2.1}$$

where n_c and p are respectively the number of control points (or number of basis functions) and the order of the basis functions. One knot vector defines what is called a "patch", and each interval $[s_i, s_{i+1}]$ of a knot vector defines a "knot span". A B-spline basis function is defined recursively (note that it is necessary to consider that $\frac{0}{0} := 0$ for the definition to work properly):

$$
\begin{aligned}
N_{i,0}(s) &= \begin{cases} 1 & s_i \leq s < s_{i+1}, \\ 0 & otherwise \end{cases} \\
N_{i,p}(s) &= \frac{s - s_i}{s_{i+p} - s_i} N_{i,p-1}(s) \\
&+ \frac{s_{i+p+1} - s}{s_{i+p+1} - s_{i+1}} N_{i+1,p+1}(s)
\end{aligned} \tag{2.2}
$$

A B-spline curve of order p is expressed using the coordinates of the control points $\mathbf{P}_i$:

$$\mathbf{C}_{BC}(s) = \sum_{i=1}^{n_c} N_{i,p}(s) \mathbf{P}_i \tag{2.3}$$

NURBS are a generalization of B-splines which are commonly used for IGA, and which employ rational B-spline functions with weights γ_i associated to each control point. NURBS curves are defined by:

$$\mathbf{C}_{NC}(s) = \sum_{i=1}^{n_c} \frac{\gamma_i N_{i,p}(s)}{\sum_{j=1}^{n_c} \gamma_j N_{j,p}(s)} \mathbf{P}_i. \tag{2.4}$$

The NURBS surface is written similarly as:

$$\mathbf{C}_{NS}(s) = \sum_{i=1}^{n_c} \sum_{j=1}^{m_c} \frac{\gamma_{ij} N_{i,p}(s) N_{j,q}(t)}{\sum_{k=1}^{n_c} \sum_{l=1}^{m_c} \gamma_{kl} N_{k,p}(s) M_{l,q}(t)} \mathbf{P}_{ij}. \tag{2.5}$$

The NURBS volume description is expressed in a similar manner, by generalizing the above equations to the 3D case.

Properties of B-spline and NURBS basis functions include partition of unity ($\sum_{i=1}^{n} N_{i,p}(s) = 1$), local support ($N_{i,p}(s) = 0$ for $s \notin [s_i, s_{i+p+1}]$), non-negativity ($N_{i,p}(s) \geq 0$) and high order of continuity (C^{p-k}).

Three mesh refinement techniques are possible in IGA. "Knot insertion" consists in inserting additional knots in knot spans and can be used to split the domain into smaller parts while keeping inter-element continuity. Inserting knots implies an increment in the number of control points and basis functions. It is similar to h-refinement in classical FEM. "Degree elevation" consists in increasing the polynomial order of the basis functions and is equivalent to p-refinement in FEM. When increasing the polynomial order, knot multiplicity must be increased (adding existing knot values in the knot vector) accordingly in order to maintain the initial continuity of the basis functions across element boundaries. The third type of refinement is known as “k-refinement” and combines polynomial order elevation with subsequent knot insertion. k-refinement has no counterpart in classic FEM and has been shown to potentially outperform p-refinement [188]. The advantage of global mesh refinement in IGA compared to classic FEM is that it does not affect the geometry itself. The main disadvantage of NURBS-based IGA with respect to refinement is that it does not enable straightforward local refinement. T-spline-based IGA is a more recent strategy which enables smooth local refinement in any geometry [186].

The basic main steps of the IGA modeling procedure are as follows:

1. Construct the CAD geometry
2. Apply global refinement
3. Map the parametric space (where the basis functions are defined) to a master element
4. Apply numerical integration to obtain the characteristic matrices of the problem, using Gauss quadrature
5. Set boundary conditions (uniform displacement and traction applied similarly to FEM)

The numerical integration for the stiffness contribution $\mathbf{K}_{init}$ is briefly described below:

$$\begin{aligned}
\mathbf{K}_{init} &= \int \mathbf{B}^T\mathbf{E}\mathbf{B}\, d\Omega && (2.6)\\
&= \int_{t_j}^{t_{j+1}} \int_{s_i}^{s_{i+1}} \mathbf{B}^T\mathbf{E}\mathbf{B}|\mathbf{J}_{dt}|\, dsdt && (2.7)\\
&= \int_{-1}^{1} \int_{-1}^{1} \mathbf{B}^T\mathbf{E}\mathbf{B}|\mathbf{J}_{dt}||\mathbf{J}_{\xi\eta}|\, d\xi d\eta && (2.8)\\
&= \sum_{k=1}^{N_{INT}} \mathbf{B}_k^T\mathbf{E}\mathbf{B}_{\mathbf{k}}|\mathbf{J}_{dt}|_k|\mathbf{J}_{\xi\eta}|_k\mathbf{W}_k && (2.9)
\end{aligned}$$

where $\mathbf{E}$ is the material property matrix, $\mathbf{B}$ is the strain-displacement matrix, $\mathbf{W}$ are weights for Gauss quadrature, $\mathbf{J}_{st}$ and $\mathbf{J}_{\xi\eta}$ are the Jacobian matrices which respectively map the physical space to the master element and the master element to the parametric space (more details in [187]).

2.1.2 Simplified disc-pad model

The disc brake model is constructed based on an open source Matlab code called NLIGA [189] which implements IGA on a variety of continuum mechanics example problems applied to relatively basic geometries. The disc and pad geometries are constructed based on the provided example cases. The main objective of the presented work being the robust optimization of gradient-free nonlinear functions, applied to disc brake friction-induced instabilities, we focus on simplifying the disc brake model as much as possible while providing faithful representation of said instabilities from a physical point of view. Nonetheless, the simplifications applied throughout the modeling process will be properly justified. The first important simplification is that only two mechanical parts are modeled, the disc and one pad. Most disc brake systems usually consist of two pads pressing on either side of the disc. However, in the aim of modeling friction-induced instabilities, it is not necessary to model both pads, which would further complexify the model and increase computational cost. Also, any surrounding mechanical parts which are present in real disc brake systems, such as calipers which hold the pads in place, the hydraulic system used to apply pressure to the pads or the rotor shaft, are not modeled but are replaced by appropriate boundary conditions. Also note that the shape optimization problem considered in the presented PhD research is limited to the pad geometry which means the shape of the pad will be varied throughout the optimization process but the disc shape will remain fixed.

Disc brake systems have been modeled with analytical two-dimensional plate formulations [190, 191] which turned out to be insufficient for break squeal analysis since modal coupling links modes inside and outside the plane [192]. Thus, three-dimensional models were introduced and particularly using FEM [193], able to take into account complex geometries and multiple phenomena. In this research we construct a three-dimensional model using three translational degrees of freedom per control point.

Disc

The disc is a cylindrical geometry characterized by thickness $T = 2cm$, defined by internal radius $R_i = 4cm$ and external radius $R_e = 16cm$. When constructing the disc model, the method to enable contact between the disc and the pad must be considered. The contact formulation described later in section 2.1.4 employs a node-to-node strategy, which implies the number of contact points is dependent on the level of refinement of the disc NURBS geometry. For this reason, we opted for a highly refined disc in order to be able to increase the number of contact points somewhat freely if needed. A highly refined disc also has the added advantage of higher accuracy of the analysis results. The challenging aspect when applying refinement to the disc is that only the portion of the disc where the pad comes into contact actually needs to be refined. Applying refinement to the whole disc may be inefficient. Multiple options were considered for this issue:

- Make a single cylindrical NURBS geometry for the disc and refine the whole disc.

In this case global deformation will also be well represented but most of the disc will be over-refined, which is not ideal in terms of computational cost

- Divide the disc into two NURBS patches: a highly-refined quarter cylinder where it is assumed the pad will come in contact, and a coarse 3/4 cylinder where it is assumed there is no contact. The two non-conforming patches must then be connected using an adequate nonconforming patch coupling method
- Use T-spline based IGA instead of NURBS and make a single geometry for the disc, with a highly refined portion for contact (T-Splines have the advantage of allowing local refinement without having to construct and connect different patches).
- Instead of modeling a full-disc, only model a quarter-disc with NURBS and refine the whole geometry

For the gain in terms of computational cost and ease of implementation, restricting the disc model to a quarter-disc was retained. Considering the fact that properly modeling friction-induced instabilities (in particular the nonlinear complex character of the resulting performance metric) is the primary purpose of the disc brake model, modeling the full disc appears as counter-productive given the necessity to consider non-conforming mesh coupling or T-spline strategies in this case. Figure 2.1 shows a comparison of the full disc NURBS geometry vs the retained quarter-disc NURBS geometry.

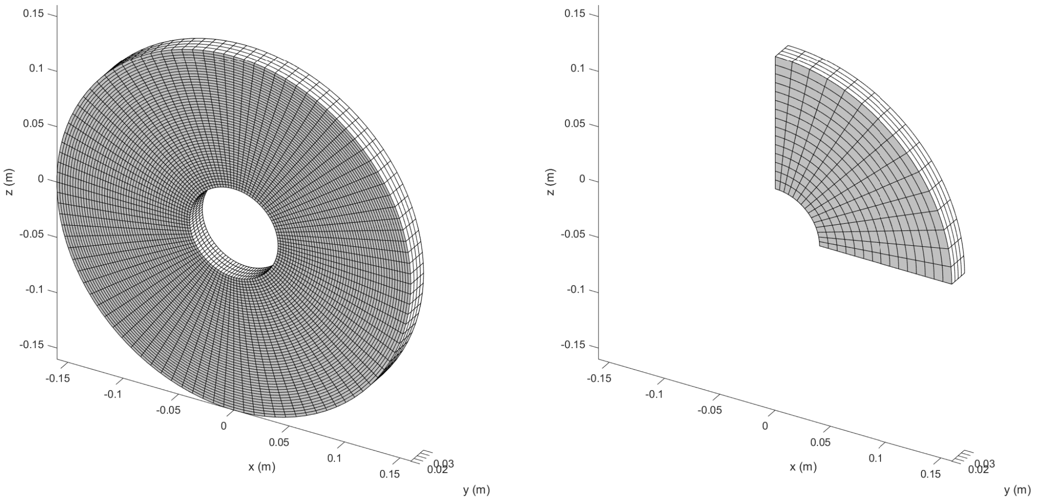


Figure 2.1: NURBS full disc geometry (left) and quarter-disc geometry (right) (the NURBS control points are not displayed)

Pad

The initial pad geometry is designed using the same principle as the disc geometry, with thickness $t = 1.5cm$, internal radius $r_i = 9cm$, external radius $r_e = 12cm$ and an angular length of 50°. Since the considered shape optimization problem focuses on varying the pad geometry, the parametrization which allows us to modify the pad geometry is a key aspect. Similarly to the disc geometry, refinement must also be applied to the pad in order to model contact. The initial default geometry before and after refinement is shown in figure 2.2. Figure 2.3 shows the quarter-disc geometry and the pad geometry together. In our shape optimization context, the pad geometry will be modified by changing the position of certain well-chosen control points, in order to access a nearly topological design space for the pad.

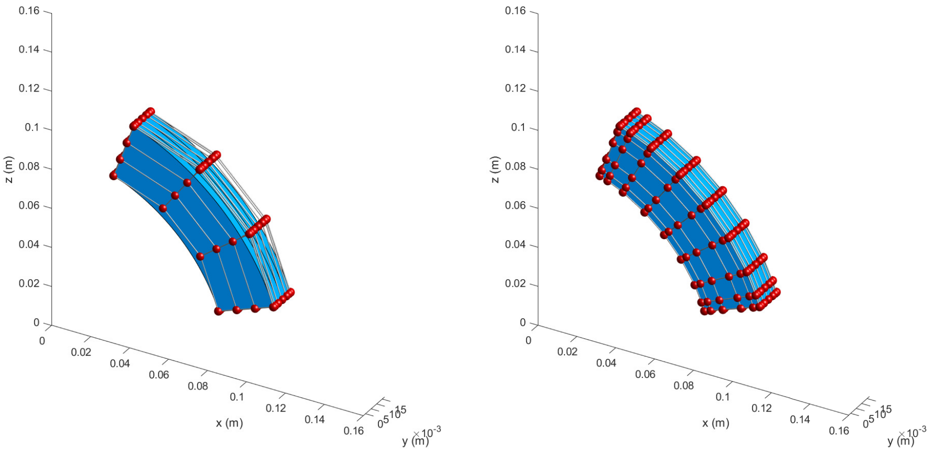


Figure 2.2: NURBS pad geometry (left:non-refined, right: refined)

2.1.3 Boundary conditions

Boundary conditions are applied to each part of the system in order to represent a real disc brake's behavior as accurately as possible, given the simplified two-part system. Fixed degrees of freedom are imposed directly to the NURBS control points, as zero displacement boundary conditions do not require the control points to interpolate the geometry. In order to model a fixed link between the inner surface of the disc and the wheel shaft which is not modeled in our case, all degrees of freedom of the innermost control points along the radial direction of the disc geometry are fixed. The two translational degrees of freedom of all control points in the plane of the top surface of the pad geometry are fixed as well. The remaining degrees of freedom normal to the pad surface are left free in order to model the contact forces. Moreover, a uniform pressure is

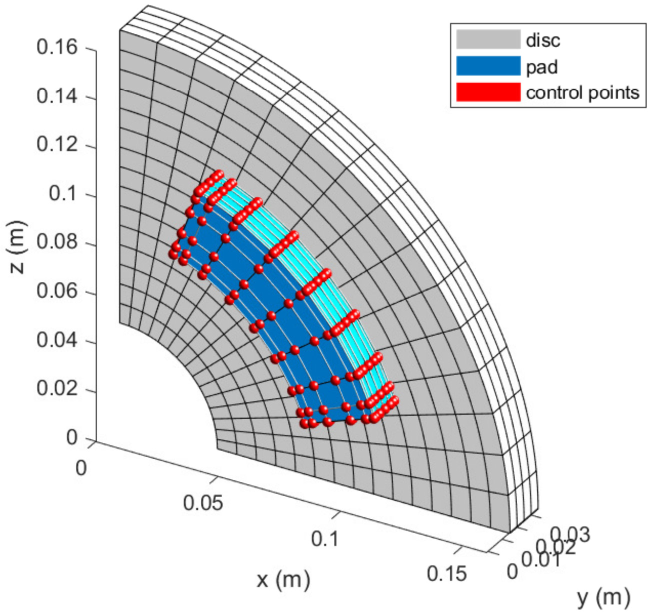


Figure 2.3: NURBS quarter-disc and pad geometries

applied to the face of the pad opposite to the contact face, by distributing an equivalent external force between the control points of the surface.

2.1.4 Contact formulation

This section presents the contact formulation used to model contact and friction between the disc and pad, starting from a brief overview of contact methods in IGA, and then explaining the adopted method in the specific case of the disc brake system.

Contact in IGA

Classical FEM approaches have been tested to enable contact between isogeometric elements [194]. Many classical approaches have been transferred from FEM to IGA, for instance the mortar approach [187]. There are three main contact strategies in FEM which can be potentially adapted for IGA:

- Node to node (NTN)
- Node to surface (NTS)
- Surface-to-surface (STS) (Gauss-point-to-surface and mortar method)

The NTN enforces contact directly between the FEM nodes of the mesh [195, 196]. Its main advantage is that it can accurately model the effect of the contact pressure.

However, it cannot be used for large sliding problems where the contact nodes are greatly displaced from their initial position. Also, it has been scarcely applied to IGA due to the fact that the control points (equivalent to FEM nodes) often do not interpolate the geometry.

The NTS method [197] ("node-to-segment" in 2D) enforces the contact constraints between a node of one contact surface, designated as the slave surface, and the corresponding surface on the other contact surface, the master. This means that the contact integrals are collocated at the slave's nodes. In IGA, the challenge of this method is to collocate the contact contribution at appropriate physical nodes, knowing that the control points are not interpolatory. Two main methods have been implemented in previous research. The first, "single-pass" [198], collocates the contact integrals at physical points in one-to-one correspondence with the control points associated to the surface. The second, "double-pass" [199], approximates the slave surface with interpolation elements and the master surface also, or taken as the actual NURBS surface and using a closest point projection algorithm. The main problems of NTS are inability to pass the contact patch test, strong dependency of results on discretization for the single-pass, and over-constraint for the double-pass.

The STS method (STS), does not collocate the contact integrals at the slave nodes, but rather enforces the contact constraints in an integral manner. One common version of STS is the mortar method [200, 201] which evaluates the weak form of contact constraints by numerical integration, rather than using collocation. In IGA, the contact constraints are projected to the degrees of freedom of the slave surface so that the right number of constraints is obtained. The mortar method satisfies both the patch test and the "inf-sup" stability requirements. Its main disadvantages are computational cost (computation and storage of mortar integrals) and its difficulty to implement.

Choice of method for the studied case

The proposed method must be adapted to the three-dimensional problem with friction taken into account. However, we have a simplified problem compared to most general contact problems in IGA. In our case:

- both contact surfaces are plane,
- the IGA mesh control points on the contact surfaces interpolate the geometry (except for some close to the boundary), so we can enforce contact in a regularized sense at control points directly,
- we do not need to consider the time-dependent problem or simulate relative motion between the two parts, since we are only interested in conducting Complex Eigenvalue Analysis of the system in a frictional contact state.

Given these assumptions NTN adapted to the case of IGA appears as the most straightforward and appropriate technique to apply contact constraints in our case. Applying NTS or STS, whose implementation is more difficult, would be counter-productive since modeling relative motion is not needed. In order to apply NTN contact to our numerical model, the first main challenge is connecting the two contact surfaces at points which make sense physically, in other words dealing with the non-conforming IGA meshes of the disc and the pad. After one-to-one correspondence between the disc and pad contact points is enabled, the penalty method is applied in order to enforce the contact constraints.

Enabling one-to-one correspondence between disc and pad contact points

First, the coordinates of control points located on the bottom surface of the pad (which is assumed in contact with the disc) are identified. The points which define the edges in the radial direction are discarded because they do not interpolate the geometry. Then, these contact points are "projected" onto the top surface of the disc, in reality the coordinates of the obtained points on the disc along the normal direction to the contact surface are exactly the same as those of the pad contact points since the bottom surface of the pad is situated on the same plane as the top surface of the disc. However, in the characteristic matrices of the disc-pad system, these points will appear twice, on one hand those of the disc and on the other hand those of the pad. For each contact point on the pad, the closest control point on the disc is found and the idea is to displace this point on the disc surface so that it has the same coordinates as the pad point. No new points are added to the disc, certain existing points are selected and displaced to make them line up (in terms of coordinates in the contact plane) with the contact points of the pad. Figure 2.4 shows a schematic view of how the closest disc control points are displaced towards each pad contact point on the contact surface.

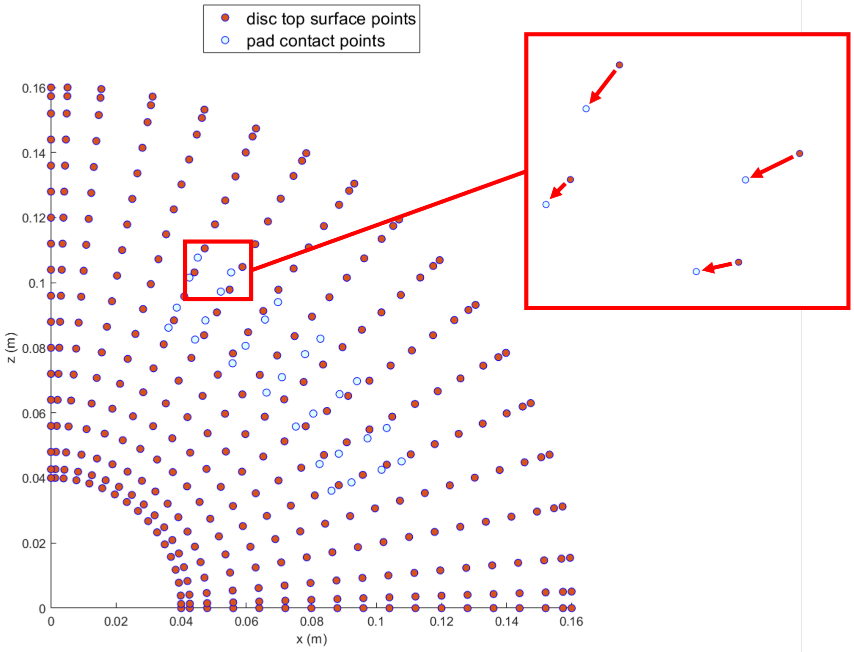


Figure 2.4: Illustration of the pad contact points being displaced towards the closest disc top surface contact points to enable one-to-one correspondence

To achieve this, we interpolate between the displacements of the "initial" disc contact points $(u_1\ u_2\ ...\ u_n)^T$ to obtain the displacements of the new replacement points $(u_{n+1}\ u_{n+2}\ ...\ u_{2n})^T$. The displacement of any point can be expressed as a combination of those of the initial points:

$$u(x,y) = u_1\Gamma_1(x,y) + u_2\Gamma_2(x,y) + ... + u_n\Gamma_n(x,y) \tag{2.10}$$

where Γ_k , $(k = \{1, ..., n\})$ are displacement interpolation functions chosen such that their evaluation at the initial points' coordinates (x_i, y_i) is:

$$\Gamma_k(x_i, y_i) = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases} \qquad \forall i, k \in \{1, ..., n\} \tag{2.11}$$

At an arbitrary new point j, that is $\forall\, j \in \{n+1, ..., 2n\}$, equation 2.10 yields:

$$u_j(x_j, y_j) = \begin{bmatrix} \Gamma_1(x_j, y_j) & \Gamma_2(x_j, y_j) & \dots & \Gamma_n(x_j, y_j) \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{2.12}$$

Then, the displacements of the new points can be expressed from those of the initial points in matrix form:

$$\begin{bmatrix} u_{n+1} \\ u_{n+2} \\ \vdots \\ u_{2n} \end{bmatrix} = \begin{bmatrix} \Gamma_1(x_{n+1}, y_{n+1}) \\ \Gamma_2(x_{n+2}, y_{n+2}) \\ \vdots \\ \Gamma_n(x_{2n}, y_{2n}) \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{2.13}$$

Thus, a transfer matrix Q_c is defined between initial points and new points :

$$Q_c = \begin{bmatrix} \Gamma_1(x_{n+1}, y_{n+1}) \\ \Gamma_2(x_{n+2}, y_{n+2}) \\ \vdots \\ \Gamma_n(x_{2n}, y_{2n}) \end{bmatrix} \tag{2.14}$$

And a total transfer matrix is defined in order to apply the transformation to the disc global matrices:

$$Q_{c,tot} = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & Q_c \end{bmatrix} \tag{2.15}$$

The non-contact displacements of the disc U_{int} and the new contact point displacements U_{new} can be expressed in terms of themselves and the initial disc contact points (the closest points to each pad contact point):

$$\begin{bmatrix} U_{int} \\ U_{new} \end{bmatrix} = Q_{c,tot} \begin{bmatrix} U_{int} \\ U_{old} \end{bmatrix} \tag{2.16}$$

The new mass and stiffness matrices of the disc K_d and M_d can be computed from their initial versions $K_{d,init}$ and $M_{d,init}$ with the old points as follows:

$$\begin{aligned} K_d &= Q_{c,tot}^T K_{d,init} Q_{c,tot} \\ M_d &= Q_{c,tot}^T M_{d,init} Q_{c,tot} \end{aligned} \tag{2.17}$$

Equations 2.10 to 2.14 describe the procedure for one degree of freedom per control point. However, our model uses three degrees of freedom (translations) per control point, in which case equation 2.12 can be extended to three degrees of freedom $(u(x_j,y_j)\ v(x_j,y_j)\ w(x_j,y_j))^T$:

$$\begin{bmatrix} u(x_j, y_j) \\ v(x_j, y_j) \\ w(x_j, y_j) \end{bmatrix} = \begin{bmatrix} \Gamma_1(x_j, y_j) & 0 & 0 & \Gamma_2(x_j, y_j) & 0 & 0 & \dots & \Gamma_n(x_j, y_j) & 0 & 0 \\ 0 & \Gamma_1(x_j, y_j) & 0 & 0 & \Gamma_2(x_j, y_j) & 0 & \dots & 0 & \Gamma_n(x_j, y_j) & 0 \\ 0 & 0 & \Gamma_1(x_j, y_j) & 0 & 0 & \Gamma_2(x_j, y_j) & \dots & 0 & 0 & \Gamma_n(x_j, y_j) \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ w_1 \\ u_2 \\ v_2 \\ w_2 \\ \vdots \\ u_n \\ v_n \\ w_n \end{bmatrix} \tag{2.18}$$

The choice of interpolation functions is not trivial. In order to provide a full rank transfer matrix, the order of the polynomials has to be high enough, more particularly its number of terms needs to be at least equal to the number of contact points on the disc. For instance, for our test case which consisted of 14 contact points, selecting a polynomial of degree 2 or 3 was not sufficient, but degree 4 provided enough terms (in 2D).

Contact and friction laws

In this section the contact and friction formulation is described for one pair of contact points. The non linear contact forces applied at the interface are separated into normal contact forces $\vec{F}_{n,d}$, $\vec{F}_{n,p}$ and tangential contact forces (friction) $\vec{F}_{t,d}$, $\vec{F}_{t,p}$ (subscript d designates the force applied to the disc while subscript p indicates the pad). To enforce contact constraints, we apply a penalty-based regularized strategy. Regularized laws do not respect the no-penetration condition but this can be interpreted as asperities being crushed at the microscopic scale. They are less realistic than non regular laws (where contact points are not fixed), however, they are less time-consuming and easier to implement. The applied regularized law is based on the force/displacement curve of a pad compressed onto a disc [179, 177]:

$$\vec{F}_{n,d} = \begin{cases} (k_L\delta + k_{NL}\delta^3)\vec{u}_n & \delta < 0 \\ \vec{0} & \delta \geq 0 \end{cases} \tag{2.19}$$

$$\vec{F}_{n,p} = -\vec{F}_{n,d} \tag{2.20}$$

where k_{NL} is the non-linear stiffness, k_L is the linear stiffness, $\delta = U_{n,d} - U_{n,p}$ is the relative displacement between the disc and the pad along the normal to the contact surface, and $\vec{u}_n$ is the unit vector in the normal direction to the contact surface.

When using penalty-based contact enforcement, care must be taken in the choice of stiffness parameters such as k_{NL} and k_L. Low stiffnesses will lead to behavior similar to that of a simple spring-mass (often used to explain harmonic motion) characterized by low frequency oscillations. On the other hand, increasing contact stiffnesses too much will tend to resemble a fixed link and will not allow relative motion along the normal to the contact surface between the two parts in contact.

For friction, the well-known Coulomb law is sufficient to model mode coupling based instabilities which is our main interest in this research. More elaborate laws with dependency of the coefficient of friction on sliding speed are out of the scope of our model (need to be derived through experimental tests and are more adapted to describe stick-slip type instabilities at low speeds). The radial component of the friction force is neglected. Coulomb's law in our case yields the following expressions for the friction force on the disc and the pad:

$$\vec{F}_{t,d} = \|\vec{F}_{n,d}\| sign(V_r)\vec{u}_t \tag{2.21}$$

$$\vec{F}_{t,p} = -\vec{F}_{t,d} \tag{2.22}$$

where μ is the friction coefficient, $sign(V_r)$ is the sign of the relative sliding speed, which we set as positive, and $\vec{u}_t$ is the unit vector in the tangential direction to the contact surface.

Figure 2.5 shows a schematic view of the contact and frictional forces for one contact pair.

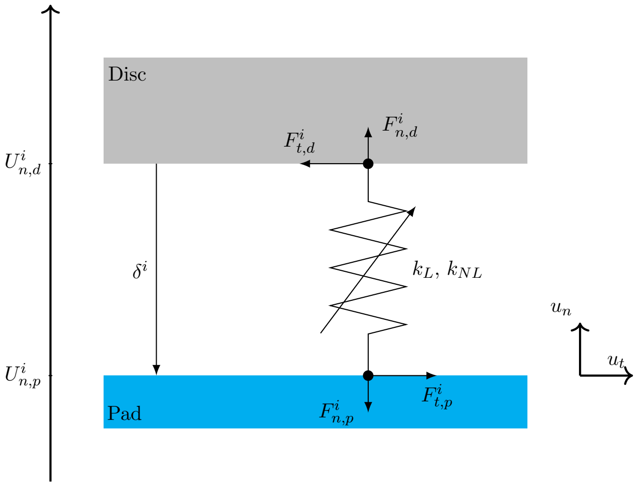


Figure 2.5: Scheme of the disc-pad contact model for one contact pair

2.2 Model reduction

This section presents the theory behind Craig & Bampton model reduction for systems containing contact interfaces, and its application in the case of the simplified disc brake model.

2.2.1 Craig & Bampton

In the context of Finite Element modeling, computational cost can become very large, especially when high mesh refinement is needed for modeling precise phenomena, as is the case with our model due to contact and friction. Additionally, in optimization problems which require multiple calls to the objective function, Finite Element models with a large number of degrees of freedom can be a bottleneck. An efficient way to mitigate this problem is to use model reduction techniques, which consist in projecting the model onto a basis made of fewer physical degrees of freedom, thus drastically reducing the problem size. Nevertheless, the choice of the basis is a crucial aspect because it conditions the accuracy of the reduced model with respect to the original. In our case, in addition to the computational cost problem, the model reduction helps limit the problem to the interface degrees of freedom as explained below. Ritz methods are the most well-known reduction methods, which decompose the solutions into a linear combination of shape functions weighted by so-called generalized degrees of freedom. This is equivalent to projecting the solutions on a vectorial space which dimension depends on the choice of shape functions. The most common version of Ritz methods is the projection onto the modal basis which uses the mode shapes as shape functions.

When a model includes contact and friction, simple Ritz methods are not adequate because the displacements of the contact interface degrees of freedom $\mathbf{U}_C$ need to be treated separately from those of the rest of the degrees of freedom $\mathbf{U}_I$, called internal degrees of freedom. This differentiation of the degrees of freedom is called sub-structuring and each substructure can be reduced separately. These methods can be divided into two main types: fixed-interface which are more adapted to systems where the contact interface is rigid and free-interface which are used when the contact interface is flexible.

In this case we use the most well-known fixed interface reduction method: Craig & Bampton method. The following equations describe the general procedure for this method. First, the displacements are rearranged by splitting the degrees of freedom into contact interface degrees of freedom and internal degrees of freedom:

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_C \\ \mathbf{U}_I \end{bmatrix} \tag{2.23}$$

Then, the rearrangement of the system's full mass and stiffness matrices comes naturally:

$$\mathbf{M}_{full} = \begin{bmatrix} \mathbf{M}_{C,C} & \mathbf{M}_{C,I} \\ \mathbf{M}_{I,C} & \mathbf{M}_{I,I} \end{bmatrix} \quad \mathbf{K}_{full} = \begin{bmatrix} \mathbf{K}_{C,C} & \mathbf{K}_{C,I} \\ \mathbf{K}_{I,C} & \mathbf{K}_{I,I} \end{bmatrix} \tag{2.24}$$

The idea is to remove the internal degrees of freedom and replace them with some equivalent contribution of much smaller size. Thus, the problem size will be that of the contact degrees of freedom plus that of the equivalent contribution. The displacements of the internal degrees of freedom are projected onto a combination of fixed-interface modes and the static behavior of the interface:

$$\mathbf{U}_I = \mathbf{\Phi}\mathbf{q}_g + \mathbf{\Psi}\mathbf{U}_I \tag{2.25}$$

where $\mathbf{\Phi} = \begin{bmatrix}\Phi_1\Phi_2 \ldots \Phi_{n_{modes}}\end{bmatrix}$ are the n_{modes} first modes of the structure with fixed interface degrees of freedom, obtained by solving the following equation:

$$(\mathbf{K}_{I,I} - \omega_i^2\mathbf{M}_{I,I})\Phi_i = \mathbf{0} \quad \forall i \in \{1...n_{modes}\} \tag{2.26}$$

where $\mathbf{q}_g$ contains the generalized degrees of freedom associated to each mode in $\mathbf{\Phi}$, also called modal contributions. $\mathbf{\Psi}$ is calculated by solving the static problem at the interface:

$$\begin{bmatrix} \mathbf{K}_{I,I} & \mathbf{K}_{C,I} \\ \mathbf{K}_{I,C} & \mathbf{K}_{I,I} \end{bmatrix} \begin{bmatrix} \mathbf{U}_C \\ \mathbf{U}_I \end{bmatrix} = \begin{bmatrix} \mathbf{F}_C \\ \mathbf{0} \end{bmatrix} \tag{2.27}$$

The second line of this matrix system leads to the expression of the internal degrees of freedom as a function of the interface degrees of freedom, and to the extraction of the static modes $\mathbf{\Psi}$ (see equation 2.28). This is equivalent to moving each degree of freedom of the interface and observing how the rest of the structure deforms. Furthermore, special attention needs to be taken with the external forces, as the external forces applied to the internal degrees of freedom must be zero, or the internal degrees of freedom where force is applied must artificially be added to the interface degrees of freedom and treated as such. In our case we apply the second option for the pad which has pressure applied to the face opposite to the contact face.

$$\begin{aligned} &\mathbf{K}_{I,C}\mathbf{U}_C + \mathbf{K}_{I,I}\mathbf{U}_I = \mathbf{0} \\ &\mathbf{U}_I = (-\mathbf{K}_{I,I}^{-1}\mathbf{K}_{I,C})\mathbf{U}_C = \mathbf{\Psi}\mathbf{U}_C \end{aligned} \tag{2.28}$$

Finally, the initial degrees of freedom are re-written as a function of the reduced degrees of freedom, and a transfer matrix Q is introduced:

$$\begin{bmatrix} \mathbf{U}_C \\ \mathbf{U}_I \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{\Psi} & \mathbf{\Phi} \end{bmatrix} \begin{bmatrix} \mathbf{U}_C \\ \mathbf{q}_g \end{bmatrix} = \mathbf{Q}_{red} \begin{bmatrix} \mathbf{U}_C \\ \mathbf{q}_g \end{bmatrix} \tag{2.29}$$

And the original full system matrices $\mathbf{M}_{full}, \mathbf{K}_{full}, \mathbf{F}_{full}$ can be reduced into matrices $\mathbf{M}, \mathbf{K}, \mathbf{F}_{ext}$:

$$\begin{aligned} &\mathbf{M} = \mathbf{Q}_{red}^T \mathbf{M}_{full} \mathbf{Q}_{red} \\ &\mathbf{K} = \mathbf{Q}_{red}^T \mathbf{K}_{full} \mathbf{Q}_{red} \\ &\mathbf{F}_{ext} = \mathbf{Q}_{red}^T \mathbf{F}_{full} \end{aligned} \tag{2.30}$$

2.2.2 Application to the disc-pad model

As explained in section 2.1.2, the disc model is highly refined in order to be able to impose contact in a one-to-one manner between disc and pad contact points even when the pad is also relatively highly refined (in which case the number of contact points is high). The Craig & Bampton method is applied in order to reduce the total number of degrees of freedom. The main idea is to reduce the disc and pad at the same time using the Craig & Bampton method, after having identified the contact points. This reduction uses only the contact points as junction points. However, the main purpose of the disc brake model is to be used inside an optimization process where the shape of the pad is varied many times. Taking this into account, the reduction can only be done once the contact points on the pad have been identified, i.e. inside the optimization loop (because the contact points' position varies with the pad shape). The disc reduction using only its contact points also follows this rule since the location of disc contact points is dependent on those of the pad. Nonetheless, because of the large number of degrees of freedom of the refined disc, reducing it inside the optimization loop is not feasible since the calculation of the static junction modes is too time-consuming and presents a bottleneck in terms of computational cost. For this reason, we reduce the disc brake model in two steps:

- First, Craig & Bampton is applied to the disc only, using the points of the potential contact surface of the disc (control points located on the top surface) as junction points. This can be thought of as a preliminary calculation, which is carried out once and never takes place inside the optimization loop.
- Then, Craig & Bampton is applied to the once-reduced disc and the pad at the same time, using the identified pad contact points and the corresponding disc contact points as the junction (inside the optimization loop).

This process was found to be the best way to incorporate the highly refined disc model inside the calculation of the dynamical objective while keeping computational cost to a minimum.

To verify that the reduction process works, a certain number of vibrational modes before and after the both model reductions are compared and show good correspondence. The second reduction in particular is interesting to verify since it uses non physical degrees of freedom resulting from the first reduction as inputs. The frequencies of the first 10 modes before and after the second Craig & Bampton reduction were computed ($f_1 = 0\,Hz$, $f_2 = 0\,Hz$, $f_3 = 0\,Hz$, $f_4 = 930.8\,Hz$, $f_5 = 960.7\,Hz$, $f_6 = 1041.1\,Hz$, $f_7 = 1057.1\,Hz$, $f_8 = 1237.3\,Hz$, $f_9 = 1606.9\,Hz$, $f_{10} = 2133.8\,Hz$). The frequencies before and after reduction were found to be identical to one decimal place, which shows that the reduction works well. Note that the first three modes have a zero frequency because they are rigid body modes which originate from the pad model. The boundary conditions applied to the pad do not block all its movements, however, after applying the contact constraints between disc and pad contact points, these rigid body modes disappear (the remaining free movements of the pad are eliminated by linking it to the disc).

2.3 Damping

Damping is an important and complex aspect of structural dynamics modeling. It corresponds to the physical phenomenon of energy dissipation through various media, for example through heat. There are three main well-known methods to model damping:

Rayleigh damping, modal damping, and hysteretic damping. Hysteretic damping is very simple but not applicable to numerical methods since it involves a complex imaginary term. Modal damping is very effective to apply more or less damping to different modes, but can be relatively time-consuming to implement because it requires an eigenvalue computation. Rayleigh damping applies the same level of damping to a well-chosen range of frequencies, and is very simple to implement. In the case of the simplified disc brake system, Rayleigh damping is applied. It consists in a damping matrix **C**, made of a linear combination of the stiffness matrix **K** and the mass matrix **M**, as shown in 2.31:

$$\begin{aligned}
&\mathbf{C} = \alpha_d \mathbf{K} + \beta_d \mathbf{M} \\
&\alpha_d = 2\xi_d \frac{\omega_{min}\omega_{max}}{\omega_{min} + \omega_{max}} \\
&\beta_d = 2\xi_d \frac{1}{\omega_{min} + \omega_{max}}
\end{aligned} \tag{2.31}$$

where ω_{min} and ω_{max} are the minimum and maximum frequencies of the studied interval, and ξ_d is the damping rate applied at ω_{min} and ω_{max}. The matrix **C** is then incorporated into the system equations (see equation 2.36).

2.4 Equations of motion & Complex Eigenvalue Analysis

This section presents the equations of motion which lead to the calculation of coupled complex eigenvalues due to friction-induced instability.

2.4.1 Treatment of the nonlinear forces

Similarly to section 2.1.4, the nonlinear forces are described for only one contact point pair. By projecting the friction forces (see section 2.1.4) along the Cartesian coordinates in the plane of the contact interface along $\vec{u}_x$ and $\vec{u}_z$, in addition to the contact forces along $\vec{u}_y$, the components of the non linear frictional contact forces can be expressed in the Cartesian coordinate system (equation 2.32). Angle θ_F which is the angle between the friction force direction and $\vec{(u)}_x$, is schematically represented in figure 2.6.

$$\begin{aligned}
&\vec{F}_{y,d} = \vec{F}_{n,d} = \left[k_L(U_{n,p} - U_{n,d}) + k_{NL}(U_{n,p} - U_{n,d})^3\right]\vec{u}_y \\
&\vec{F}_{y,p} = -\|\vec{F}_{y,d}\|\vec{u}_y \\
&\vec{F}_{x,d} = -\mu cos(\theta_F) \times \|\vec{F}_{y,d}\|\vec{u}_x \\
&\vec{F}_{z,d} = \mu sin(\theta_F) \times \|\vec{F}_{y,d}\|\vec{u}_z \\
&\vec{F}_{x,p} = \mu cos(\theta_F) \times \|\vec{F}_{y,d}\|\vec{u}_x \\
&\vec{F}_{z,p} = -\mu sin(\theta_F) \times \|\vec{F}_{y,d}\|\vec{u}_z
\end{aligned} \tag{2.32}$$

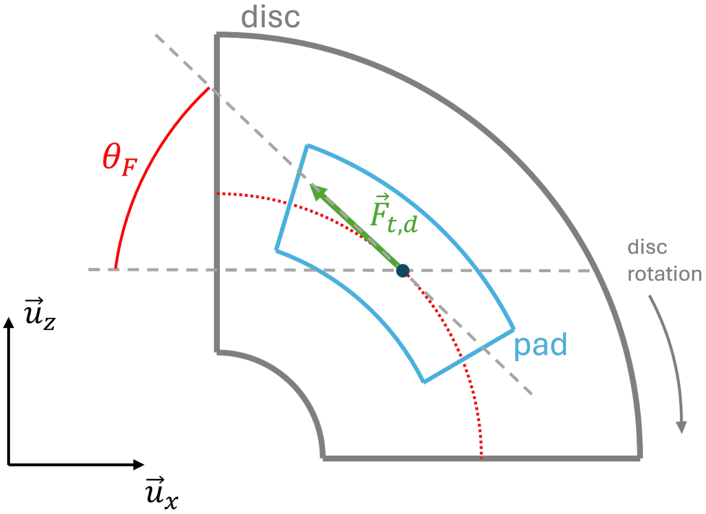


Figure 2.6: Scheme representing angle θ_F and the friction force for one contact point on the simplified disc brake system

The Jacobian matrix of the non linear forces can then be expressed on one contact pair as:

$$\mathbf{J}^*_{F_{NL}} = \begin{bmatrix} \frac{\partial F_{x,d}}{\partial U_{x,d}} & \frac{\partial F_{x,d}}{\partial U_{y,d}} & \frac{\partial F_{x,d}}{\partial U_{z,d}} & \frac{\partial F_{x,d}}{\partial U_{x,p}} & \frac{\partial F_{x,d}}{\partial U_{y,p}} & \frac{\partial F_{x,d}}{\partial U_{z,p}} \\ \frac{\partial F_{y,d}}{\partial U_{x,d}} & \frac{\partial F_{y,d}}{\partial U_{y,d}} & \frac{\partial F_{y,d}}{\partial U_{z,d}} & \frac{\partial F_{y,d}}{\partial U_{x,p}} & \frac{\partial F_{y,d}}{\partial U_{y,p}} & \frac{\partial F_{y,d}}{\partial U_{z,p}} \\ \frac{\partial F_{z,d}}{\partial U_{x,d}} & \frac{\partial F_{z,d}}{\partial U_{y,d}} & \frac{\partial F_{z,d}}{\partial U_{z,d}} & \frac{\partial F_{z,d}}{\partial U_{x,p}} & \frac{\partial F_{z,d}}{\partial U_{y,p}} & \frac{\partial F_{z,d}}{\partial U_{z,p}} \\ \frac{\partial F_{x,p}}{\partial U_{x,d}} & \frac{\partial F_{x,p}}{\partial U_{y,d}} & \frac{\partial F_{x,p}}{\partial U_{z,d}} & \frac{\partial F_{x,p}}{\partial U_{x,p}} & \frac{\partial F_{x,p}}{\partial U_{y,p}} & \frac{\partial F_{x,p}}{\partial U_{z,p}} \\ \frac{\partial F_{y,p}}{\partial U_{x,d}} & \frac{\partial F_{y,p}}{\partial U_{y,d}} & \frac{\partial F_{y,p}}{\partial U_{z,d}} & \frac{\partial F_{y,p}}{\partial U_{x,p}} & \frac{\partial F_{y,p}}{\partial U_{y,p}} & \frac{\partial F_{y,p}}{\partial U_{z,p}} \\ \frac{\partial F_{z,p}}{\partial U_{x,d}} & \frac{\partial F_{z,p}}{\partial U_{y,d}} & \frac{\partial F_{z,p}}{\partial U_{z,d}} & \frac{\partial F_{z,p}}{\partial U_{x,p}} & \frac{\partial F_{z,p}}{\partial U_{y,p}} & \frac{\partial F_{z,p}}{\partial U_{z,p}} \end{bmatrix} \tag{2.33}$$

The highlighted elements are the only non-zero elements of the Jacobian, since all the non linear force components depend only on the displacements of the disc and pad along the $\vec{u}_y$ direction. Thus, we can see the asymmetry of the Jacobian which is due to the derivatives of the tangential forces (friction). When detailing the expressions of the different non-zero partial derivative terms of the Jacobian we notice that they all have similar expressions. In particular, if we define a number C_{NL}:

$$C_{NL} = -k_L + k_{NL}(-3U_{y,d}^2 - 3U_{y,p}^2 + 6U_{y,p}U_{y,d}) \tag{2.34}$$

The partial derivatives are expressed below:

$$\begin{aligned}
\frac{\partial F_{x,d}}{\partial U_{y,d}} &= \frac{\partial F_{x,p}}{\partial U_{y,p}} = -C_{NL}\mu cos(\theta_F) \\
\frac{\partial F_{y,d}}{\partial U_{y,d}} &= \frac{\partial F_{y,p}}{\partial U_{y,p}} = C_{NL} \\
\frac{\partial F_{z,d}}{\partial U_{y,d}} &= \frac{\partial F_{z,p}}{\partial U_{y,p}} = C_{NL}\mu sin(\theta_F) \\
\frac{\partial F_{x,p}}{\partial U_{y,d}} &= \frac{\partial F_{x,d}}{\partial U_{y,p}} = C_{NL}\mu cos(\theta_F) \\
\frac{\partial F_{y,p}}{\partial U_{y,d}} &= \frac{\partial F_{y,d}}{\partial U_{y,p}} = -C_{NL} \\
\frac{\partial F_{z,p}}{\partial U_{y,d}} &= \frac{\partial F_{z,d}}{\partial U_{y,p}} = -C_{NL}\mu sin(\theta_F)
\end{aligned} \tag{2.35}$$

2.4.2 Equation of motion

The equation of motion of the brake system with friction interface is the following:

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{F}_{ext} + \mathbf{F}_{NL}(\mathbf{U}) \tag{2.36}$$

2.4.3 Equilibrium point

The equilibrium point is defined by:

$$\mathbf{K}\mathbf{U_0} = \mathbf{F}_{ext} + \mathbf{F}_{NL}(\mathbf{U_0}) \tag{2.37}$$

$\mathbf{U_0}$ is determined iteratively using the Newton Raphson method until the components of the residue vector $\mathbf{R}_{NR}$ are close enough to zero. The equilibrium point at each iteration ${\mathbf{U_0}}^{i+1}$ is determined from the equilibrium point at the previous iteration ${\mathbf{U_0}}^{i}$, and the residue is computed:

$$\begin{aligned}
&{\mathbf{U_0}}^{i+1} = {\mathbf{U_0}}^{i} - (\mathbf{K} - \mathbf{J}_{F_{NL}}({\mathbf{U_0}}^{i}))^{-1}(\mathbf{F}_{ext} + \mathbf{F}_{NL}({\mathbf{U_0}}^{i})) \\
&\mathbf{R}_{NR}({\mathbf{U_0}}^{i}) = \mathbf{K}{\mathbf{U_0}}^{i} - \mathbf{F}_{ext} - \mathbf{F}_{NL}
\end{aligned} \tag{2.38}$$

with $\mathbf{J}_{F_{NL}}({\mathbf{U_0}}^{i}) = \frac{\partial \mathbf{F}_{NL}}{\partial \mathbf{U}}({\mathbf{U_0}}^{i})$.

2.4.4 Linearization

The non linear friction forces can be linearized around the equilibrium, using the 1st order Taylor expansion:

$$\begin{aligned}
\mathbf{F}_{NL}(\mathbf{U_0} + \delta\mathbf{U}) &\approx \mathbf{F}_{NL}(\mathbf{U_0}) + \frac{\partial \mathbf{F}_{NL}}{\partial \mathbf{U}}(\mathbf{U_0})\delta\mathbf{U} \\
&\approx \mathbf{F}_{NL}(\mathbf{U_0}) + \mathbf{J}_{F_{NL}}(\mathbf{U_0})\delta\mathbf{U}
\end{aligned} \tag{2.39}$$

where $\delta\mathbf{U}$ is a small perturbation of the equilibrium point. The linearized global equation becomes:

$$\mathbf{M}\delta\ddot{\mathbf{U}} + \mathbf{C}\delta\dot{\mathbf{U}} + (\mathbf{K} - \mathbf{J}_{F_{NL}})\delta\mathbf{U} = 0 \tag{2.40}$$

2.4.5 Eigenvalue problem

Expressing the perturbation as:

$$\begin{aligned} \delta\mathbf{U} &= \phi_{eig}e^{\lambda t} \\ \delta\dot{\mathbf{U}} &= \phi_{eig}\lambda e^{\lambda t} \\ \delta\ddot{\mathbf{U}} &= \phi_{eig}\lambda^2 e^{\lambda t} \end{aligned} \tag{2.41}$$

The following eigenvalue problem is obtained:

$$(\lambda^2\mathbf{M} + \lambda\mathbf{C} + (\mathbf{K} - \mathbf{J}_{F_{NL}}))\phi_{eig} = 0 \tag{2.42}$$

2.4.6 State system

In order to solve the eigenvalue problem stated above, which takes into account damping, state variable $\mathbf{V}$ is introduced:

$$\mathbf{V} = \begin{bmatrix} \delta\mathbf{U} \\ \delta\dot{\mathbf{U}} \end{bmatrix} \tag{2.43}$$

Equation 2.40 can then be re-written under the so-called state system form:

$$\begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \delta\dot{\mathbf{U}} \\ \delta\ddot{\mathbf{U}} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & -\mathbf{K} \\ \mathbf{K} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \delta\mathbf{U} \\ \delta\dot{\mathbf{U}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{2.44}$$

which then yields the first-order differential equation $\dot{\mathbf{V}} = \mathbf{S}\mathbf{V}$, with:

$$\mathbf{S} = -\begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} & -\mathbf{K} \\ \mathbf{K} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix} \tag{2.45}$$

The unstable modes can then be identified by calculating directly the eigenvalues of matrix $\mathbf{S}$.

2.5 Stability function

As explained in section 1.4.3, there are multiple methods that can be used to predict the noise propensity of frictional systems. The most common way is to use time integration to obtain the vibrational (and thus noise) levels. However, this approach is usually very time consuming. In the interest of providing a time-efficient quantification of noise, we select a CEA-based approach. This method is different from time integration and similar methods where the goal is to determine the level of noise generated, as the noise levels are not obtained. The idea is rather to predict the possibility of noise occurrence by assuming noise occurrence is directly related to the number and magnitude of positive real parts of complex eigenvalues. According to Lyapunov's theory of dynamical stability, instability can be identified when the real part of complex eigenvalues becomes positive. In our case, the asymmetry in the stiffness matrix caused by the friction force terms leads to complex eigenvalues and mode coupling phenomenon, which induces half of the coupled real parts to become positive. In general, the complex eigenvalues are computed over a range of a certain control parameter in order to get an overall idea of the extent of instability. This leads to the notion of Hopf bifurcation. Here we use the friction coefficient as a control parameter, which means we can quantify instability and therefore noise propensity for a range of friction coefficients. This means the measure of noise is representative of several operating conditions rather than one nominal case. The proposed noise quantification criterion is defined as:

$$C_s(\mathbf{X}) = \sum_{i=1}^{n_{eig}} \left(\int_{\boldsymbol{\mu}} \max\left(\Re\{\lambda_i(\mathbf{X}_d, \boldsymbol{\mu})\} + d_0, 0\right) \right), \tag{2.46}$$

where n_{eig} is the number of eigenvalues, μ is the range of friction coefficients, $\mathbf{X}_d$ is the set of parameters which describe the shape of the disc brake system (explained in detail in chapter 3), λ_i is the i-th eigenvalue, and $\Re^+$ is the positive part of the real part operator. This criterion basically translates to taking the sum of all the areas under the positive real part curves on the stability (real part) diagram. Figure 2.7 illustrates the criterion in the most simple case with one coupled mode pair, and no damping. Note that parameter d_0, which is by default set to zero, can be used to shift the whole stability diagram along the y-axis, which can be useful in order to differentiate between several very low instability cases in terms of potential noise propensity. Furthermore, the noise criterion only takes into account modes in the audible frequency range (20Hz-20kHz), as modes outside this range will not produce any human-audible sound. From now on and until the end of this dissertation, the criterion defined in equation 2.46 will be named "stability function", with minimization of the stability function being equivalent to minimization of noise occurrence.

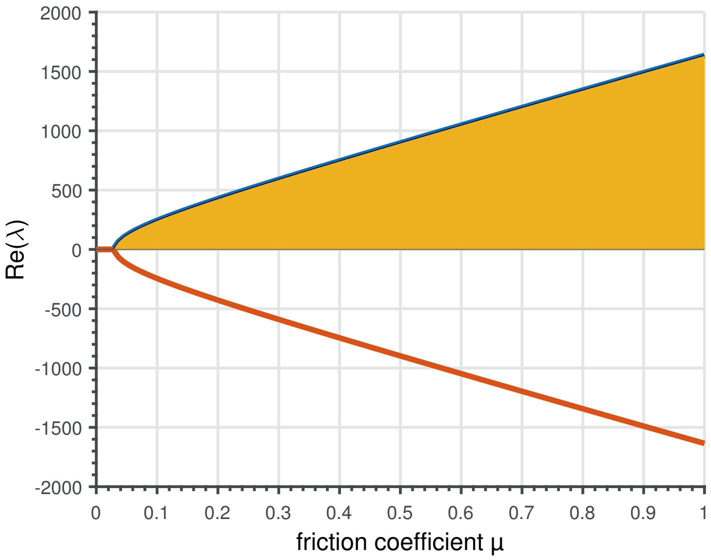


Figure 2.7: Graphical illustration of the stability function (in orange) in a simple case with only one unstable mode pair and no damping

2.6 Physical insight and model validation results

This section presents friction-induced instability results obtained with the model described in the previous sub-sections. These results evidence the capability of our model to reproduce friction-induced instability behavior, and give insight into the physical effect and computational impact of various parameters such as damping, the number of contact points, and material properties.

2.6.1 Hopf bifurcation & dynamical instability identification

Figure 2.8 shows the variation of the real part of the eigenvalues of the three first pairs of unstable modes as a function of the friction coefficient. The Hopf bifurcation obtained at a certain value of friction coefficient for each mode pair is a clear indication of the unstable behavior. Amongst the three first unstable mode pairs, only one is located in the audible frequency range. Note that for the prediction of squeal noise, the two mode pairs whose frequency is above 20 kHz would not be taken into account.

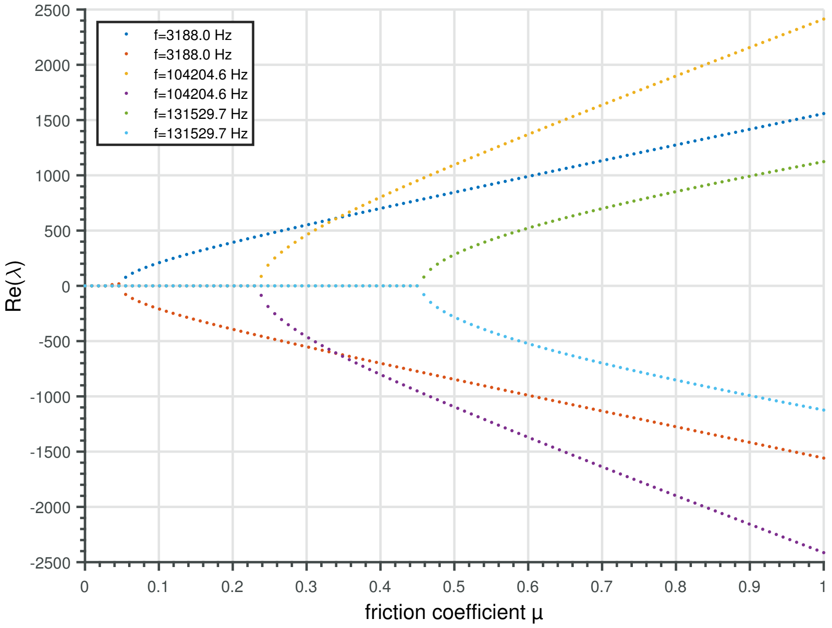


Figure 2.8: Real part of the eigenvalues of three unstable mode pairs as a function of the friction coefficient μ

Figure 2.9 shows the variation of the real part and the imaginary part of the eigenvalues of the first unstable mode pair as a function of the friction coefficient, as well as the imaginary part as a function of the real part. These plots are a good example of typical friction-induced instabilities.

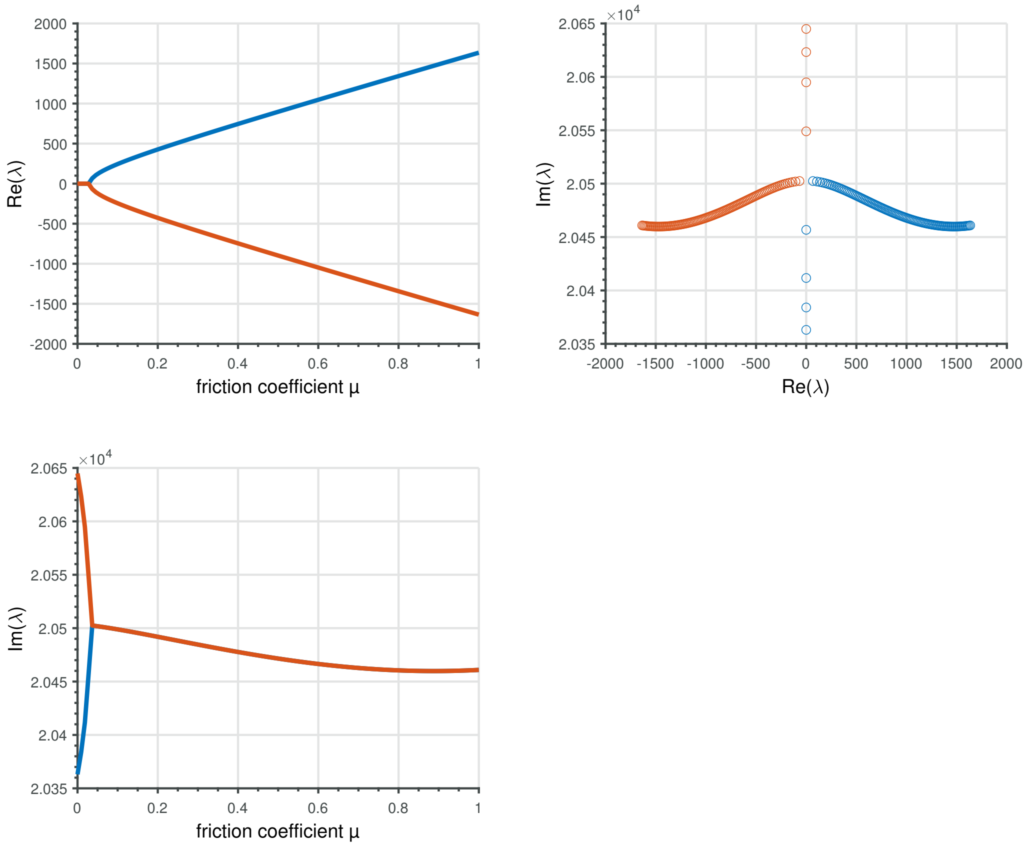


Figure 2.9: Eigenvalues of the first unstable mode pair (top left: real part verses friction coefficient, bottom left: imaginary part verses friction coefficient, top right: imaginary part verses real part)

2.6.2 Effect of damping

Figure 2.10 shows the effect of damping on one unstable mode pair. As expected, adding damping to the structure shifts the real part of the unstable pair towards negative values, thus reducing instability. Also, the bifurcation appears at a slightly lower value of friction coefficient when considering damping, and the initial real parts of the two coalescent eigenvalues are not exactly at the same value as they are in the case without damping. These effects are consistent with what is observed in previous studies of this phenomenon [202, 175]. The settings used for the Rayleigh damping (described in section 2.3) are : $\omega_1 = 2e4 \, rad/s$, $\omega_2 = 7e5 \, rad/s$, $\xi = 5\,\%$. Taking into account damping does not considerably affect the computational cost of determining unstable modes.

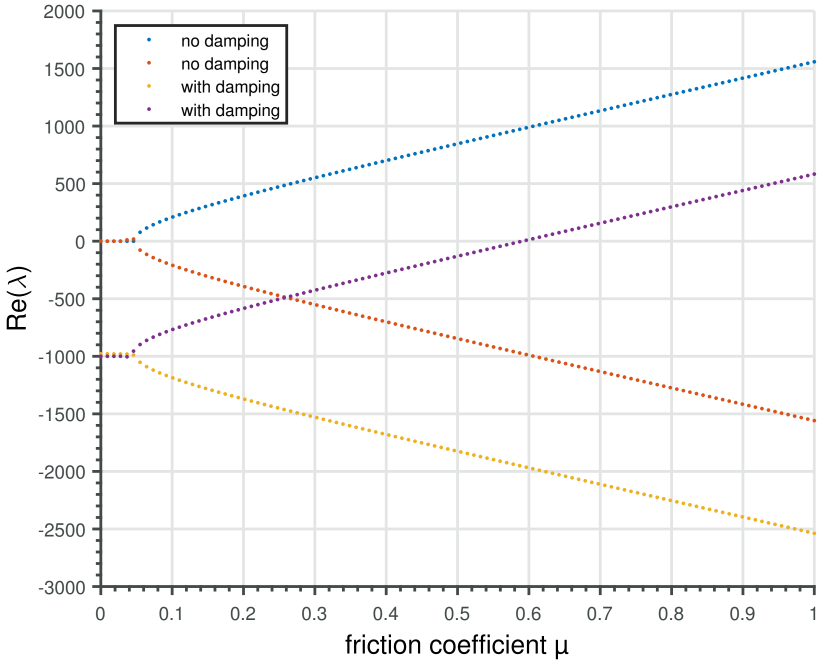


Figure 2.10: Effect of damping on the real part of the first pair unstable modes

2.6.3 Linear vs nonlinear contact stiffness

The computational cost being a crucial aspect in an optimization problem, it has to be taken into account at every step of the implementation process. In the optimization scheme, since the calculation of the stability function is potentially carried out thousands of times, the computational cost of the function for a single pad geometry must be reduced as much as possible. In this context, the question of including contact non-linearities is important. Given that the friction-induced instability is not due to non-linearities (it is due to the asymmetric friction terms in the stiffness matrix), and that modeling non-linear contact has considerable computational cost (because the equilibrium point needs to be calculated) the prospect of limiting the model to linear contact is worth investigating.

The results of this section were computed with the following values for the linear and non-linear contact stiffness constants: $k_L = 1\text{e}7\ N/m$ and $k_{NL} = 1\text{e}14\ N/m$ (refer to section 2.1.4 for more details on how these two constants are taken into account). If we compare the results using these settings with a case where only linear stiffness terms are taken into account (see figure 2.11), we observe that even a high value of nonlinear stiffness has minimal effect on the unstable behavior. This makes sense because when using non-linear contact stiffnesses, at equal linear and non-linear stiffness constants, the cubic stiffness term (which is a function of the δ introduced in section 2.1.4) is negligible compared to the linear stiffness term. Thus, the effect of the nonlinear term is low enough to consider that the completely linear model is sufficient in our case, given that we are only using it for stability analysis through eigenvalue determination. We recall that taking non-linear contact stiffnesses into account is crucial in more sophisticated models used to describe the behavior of the system over time since the non-linearities are needed to obtain a non-zero amplitude level in the time response.

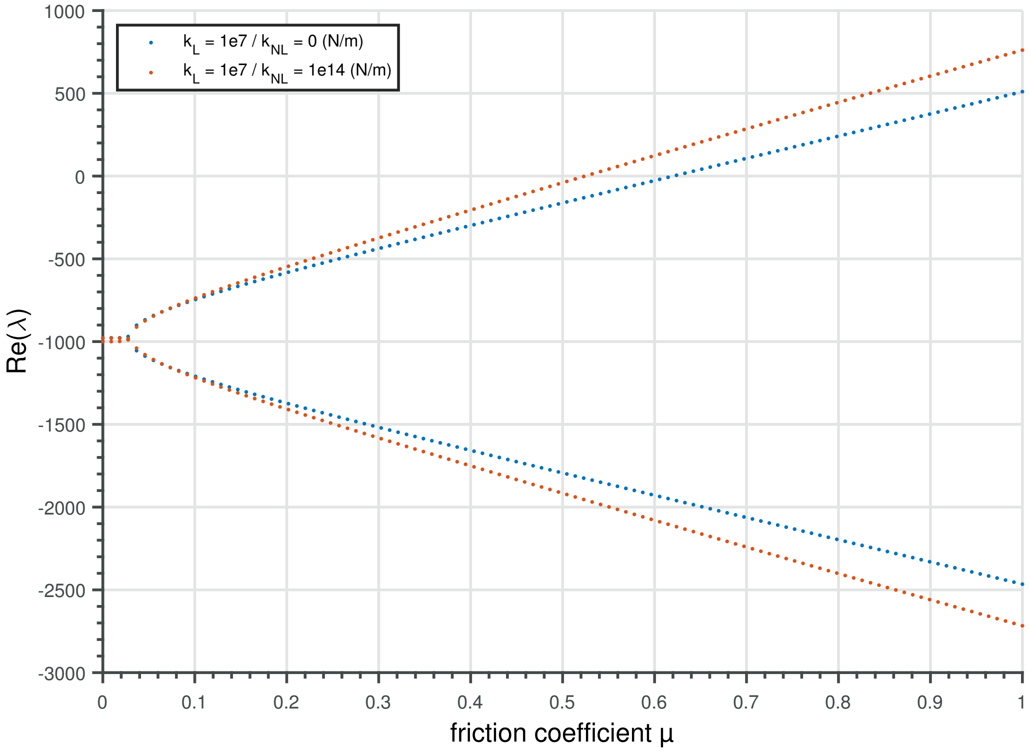


Figure 2.11: Effect of including a non-linear contact stiffness constant compared to the linear case, on the real part of the first unstable mode as a function of the friction coefficient μ

Using only a linear contact stiffness, the computational cost was reduced to approximately 25 seconds, compared to 109 seconds with the linearization process. Analyzing the distribution of computation cost to different tasks, it appears that approximately 80% of computation time corresponds to the search for the equilibrium point whereas the remaining 20% is attributed to the other tasks (contact point determination, disc matrix transformation, Craig & Bampton reduction, eigenvalue calculation, etc.). In light of these results, the linear stiffness contact model is used for the optimization problem & the rest of the presented research. Note that the computational cost results in this sub-section were obtained using a machine equipped with a 16-core 2.25 GHz processor and 94 GB of RAM.

2.6.4 Effect of the number of contact points

The number of contact points is important when modeling friction-induced instabilities and friction in general, and we can speculate that more contact points provide a more precise model. However, the overall objective of this work being to design efficient robust optimization approaches, the precision of the contact model is not paramount, whereas computational cost is a key point of the disc brake model. Since the computation time increases substantially with the number of contact points, we choose to limit ourselves to a model with fewer contact points, the primary aim for the model being that it reproduces instability. We also verify the effect of the number of contact points on dynamic instability behavior. Figure 2.12 shows the imaginary parts as a function of the real parts for one coupled mode pair (still over a range of friction coefficients), which is a good way of visualizing both the frequency-related and stability related information in one graph, for different numbers of contact points. As the number of contact points increases we notice a convergence phenomenon. Furthermore, figure 2.13 shows the evolution of the computational cost of evaluating the stability function as the number of contact points is increased, carried out with a standard computer (64-bit, 4-core 1.50GHz base speed CPU, 16 GB RAM). For the rest of the disc brake study we use 28 contact points (unless specified otherwise), which is a number of points where the convergence phenomenon is well-advanced while computational cost is mitigated.

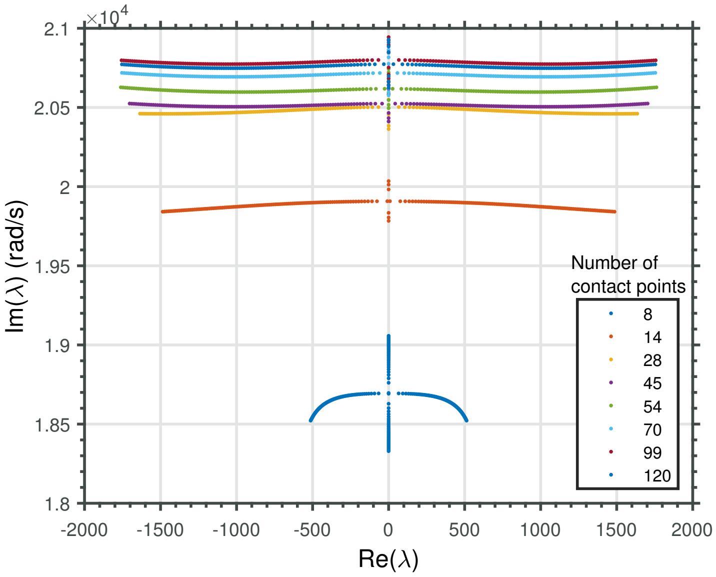


Figure 2.12: Stability plot convergence for one pair of coupled modes as the number of contact points is increased

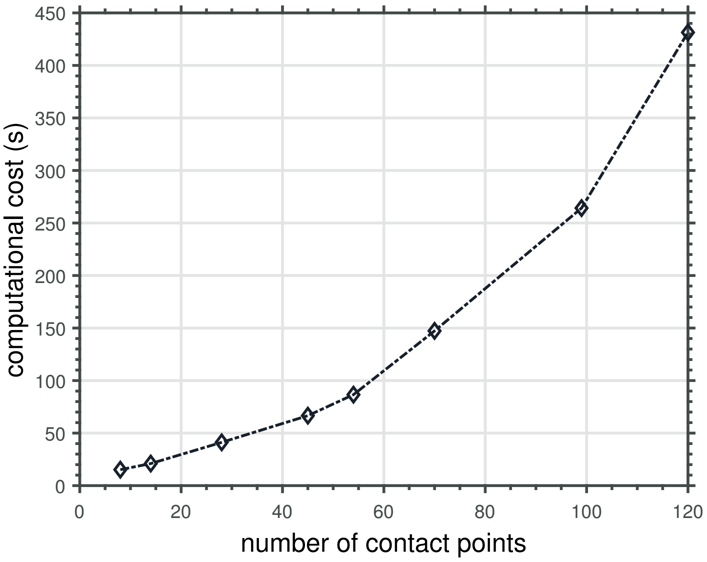


Figure 2.13: Evolution of the computational cost of one stability function evaluation as the number of contact points is increased

2.6.5 Material and pad thickness considerations

While steel is widely used for the disc material in brake systems, many friction materials are available for the brake pad which differ in their mechanical properties. Friction materials are specifically designed to produce friction when in contact with other solids in order to stop or control motion in mechanical systems. Friction materials can be made of various substances such as metals, ceramics, rubber or resins. These materials tend to wear down with use which makes analyzing brake pad thickness an important aspect. This section presents a parametric study of the effect of the pad's material properties on squeal noise occurrence according to the stability function presented in section 2.5. The Poisson coefficient ν (ranging from 0.17 to 0.28), Young modulus E (ranging from 2.5 to $16.22GPa$), and density ρ (ranging from 1210 to $2700kg/m3$) of 21 different friction materials were collected from two industrial friction material manufacturers [203, 204]. The detailed properties of the studied materials are listed in Table 2.1. First the noise propensity of all 21 materials was evaluated. Secondly, parametric studies of ν, E and ρ independently, around nominal material properties corresponding to 4 of the 21 different friction materials, are carried out. The goal is to gain insight into the effect of each parameter, even though in an industrial setting material properties cannot be varied in a continuous manner. Finally, the effect of pad thickness on squeal noise for the same 4 materials is also analyzed. The thickness cannot be optimized, as it is progressively reduced through wear during brake usage, but the effect of thickness is relevant to overall disc brake design and maintenance.

The stability function calculated for the 21 pad friction materials is displayed in figure 2.14. Material MI00950 yields the lowest stability function value at around 929 while material MI00704 yields the highest value at approximately 1235. Moreover, materials MI00950, FTL087, MI00608 and FTL180, which exhibit relatively diverse noise outputs, are used for the parametric study.

The variation of the stability function as a function of E, ν, ρ and the pad thickness, around nominal material properties of 4 friction materials (FTL180, FLT087, MI00620, MI00950) is shown in figure 2.15. Firstly, we observe that the variation with respect to each studied material property always has the same global trend irrespective of the studied material. This suggests that the effect of each material property is independent from the others. The variation with respect to E consists of a steep decreasing trend below $0.4GPa$ and a much flatter decreasing trend for $E > 0.4GPa$. The variation as a function of ν is small (less than 30 in the studied range), with a linear increasing trend. The variation of the stability function as the density increases is a relatively steep decreasing trend with a discontinuity at $2000kg/m^3$. Although the stability function does not seem much affected by variations in the Poisson coefficient, the Young's modulus and density do have a substantial effect. Generally, the obtained results suggest that friction materials with a higher Young's modulus and with a lower density are less prone to friction induced instabilities and squeal noise occurrences. An increase in thickness of the pad geometry from $1mm$ to $5mm$ causes a steep increase in stability function while in the $1mm$ to $5mm$ range the stability function decreases slowly. This suggests that with repeated braking operations, as the pad wears down to $5mm$ thickness, squeal noise may occur slightly more frequently, whereas if the pad wears down further, approaching $1mm$, there may be a sudden decrease in noise occurrence.

Table 2.1: List of pad friction materials and their properties

Material reference	Young's modulus (GPa)	Poisson coefficient	Density (kg/m^3)
FTL087	5.5	0.24	1720
FTL098	5.2	0.25	2200
FTL171	13.4	0.252	1570
FTL171M	8.5	0.19	1420
FTL172	7.52	0.17	2000
FTL175	2.54	0.21	1850
FTL180	7.29	0.28	1210
FTL180HF	7.3	0.28	1280
MI00850	7.042	0.19	1900
MI00201	5.2	0.25	1850
MI00401	13.354	0.255	1600
MI00402	8.432	0.19	1450
MI00501	8.4	0.24	1900
MI00601	2.5	0.18	1800
MI00604	3.896	0.27	1800
MI00606	2.413	0.23	1850
MI00608	5.506	0.24	1750
MI00620	9.19	0.24	1800
MI00704	7.26	0.27	1200
MI00801	5.381	0.24	2100
MI00950	16.22	0.22	2700

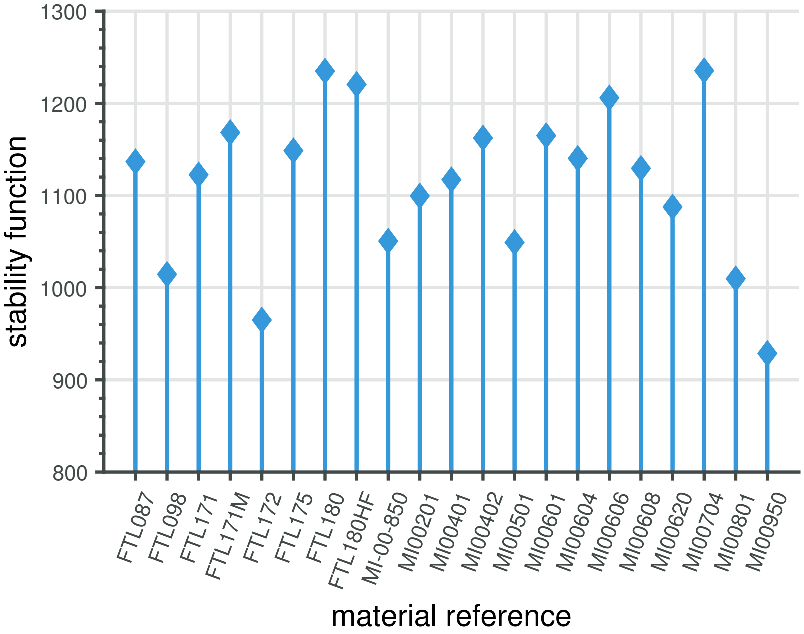


Figure 2.14: Stability function output for 21 different pad friction materials

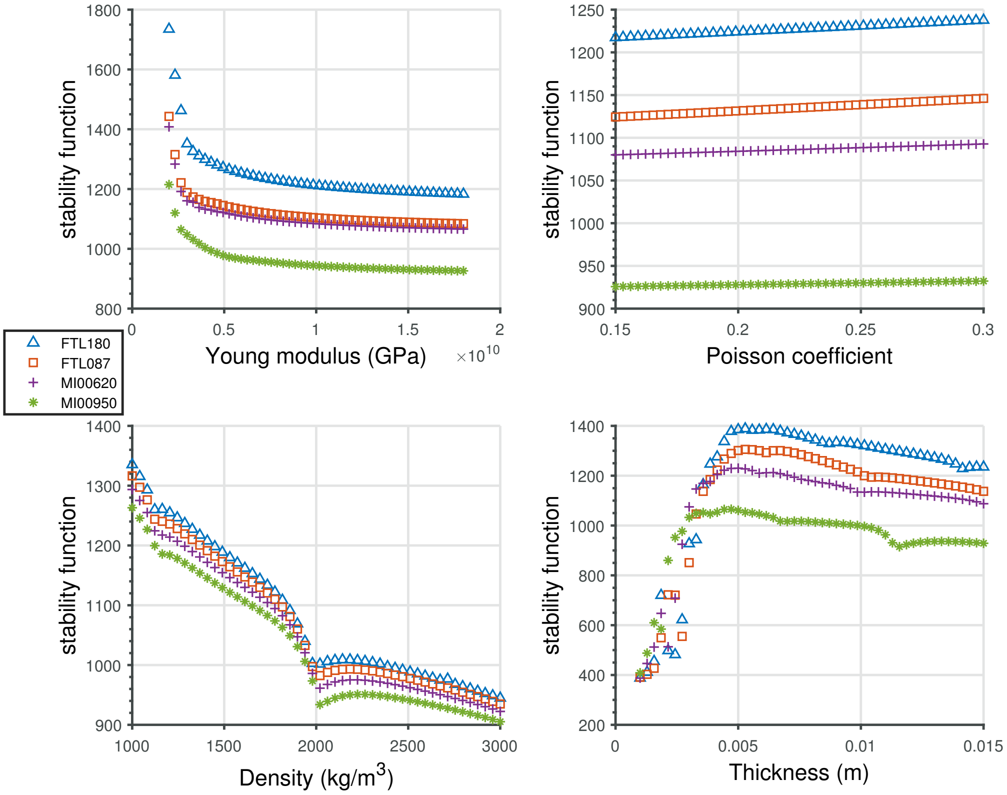


Figure 2.15: Variation of the stability function as a function of E (top left), ν (top right), ρ (bottom left) and the thickness (bottom right) around nominal properties of 4 friction materials

2.7 Important takeaways from the disc brake modeling process

This chapter constitutes the basis for numerically modeling the application studied in this research work on robust shape optimization. The disc brake model is used to quantify squeal noise occurrence in a simplified disc brake system based on dynamic instability and yields a gradient-free performance function which is the central aspect of the studied uncertain optimization problem. The performance function is based on CEA, in a similar manner to a few past studies. The unavailability of this function's gradient, its non-negligible computational cost, and its potentially complex nonlinear behavior represent important challenges in view of integrating it in a shape optimization process which considers uncertainties.

The presented modeling strategy is focused on correctly reproducing friction-induced instability behavior while mitigating the computational cost as much as possible and allowing flexible geometric parametrization in the context of robust shape optimization, by concentrating on a few key aspects which also make the modeling strategy unique when compared to previous research. First, we employ the IGA method for its convenience in shape optimization problems and for its computational cost-saving characteristics. Our model is one of the first IGA-based friction-induced instability disc brake models. Secondly, the simple penalty-based contact formulation which takes advantage of the studied problem's unique configuration by applying contact constraints directly between IGA control points, demonstrates the ability to reproduce friction-induced instabilities with fairly basic tools and is one of the first implementations of node-to-node contact strategies in IGA. Also, the Craig & Bampton reduction, applied in two successive steps, is another key aspect which further reduces the computational cost of the stability function.

Initial tests suggest that the friction-induced instability phenomenon is well represented. Analyzing the effect of damping, the number of contact points, the effect of nonlinear contact stiffnesses, and the effect of material parameters allows to deepen our understanding of such parameters' importance when modeling friction-induced instabilities, and to offer insight into how to select such parameters.

3 Optimization problem setting

The aim of this PhD research is to provide ways of optimizing the shape of mechanical structures under a performance criterion while taking into account and minimizing the impact of geometric uncertainties on said performance criterion. Before attempting to solve this optimization problem, it is important to clearly define its characteristics. Multi-objective constrained optimization problems are commonly defined using three main components: inputs parameters, objective functions and constraint functions. In the case of robust optimization, it is additionally necessary to clearly define which parameters are considered uncertain, in addition to the deterministic input parameters. This chapter describes the different components of the robust shape optimization problem in the case of the simplified disc brake system.

3.1 Input parameters

This section presents the considered optimization problem's input parameters, with a distinction between deterministic and uncertain input parameters.

3.1.1 Deterministic shape parameters

The input parameters have been chosen to effectively control the shape of the studied mechanical structure. As presented in section 2.1.2, the disc brake numerical model is made of two main parts which are the disc and the pad. The shape optimization problem focuses on modifying the shape of the pad only. In order to affect the IGA model's geometry, the coordinates of the IGA control points are directly modified. However, the pad geometry is refined in order to increase the number of contact points with the disc, which means that the number of control points increases substantially. For this reason, we use the control points before refinement (only six points) in order to modify the pad geometry. The cylindrical coordinates (radial and angular) in the plane of contact are used which makes a total of twelve input parameters $\mathbf{X}_d = [r_1, ...r_6, \theta_1, ..., \theta_6]^T$. Figure 3.1 shows a schematic view of the IGA control point coordinates used to affect the initial shape. Figure 3.2 shows examples of 10 IGA pad geometries (after refinement) obtained from a Latin Hypercube Sampling over the 12 design parameters.

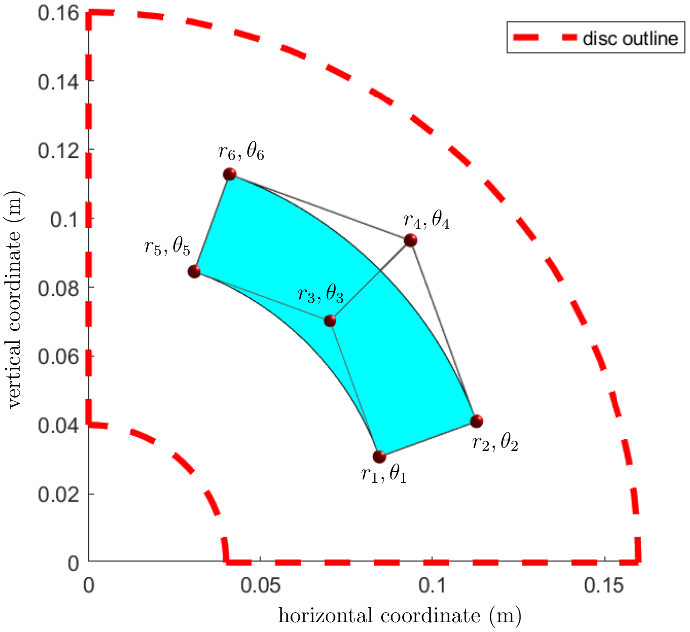


Figure 3.1: Scheme of the initial unrefined IGA pad model with the control points used for shape modification and their corresponding coordinates

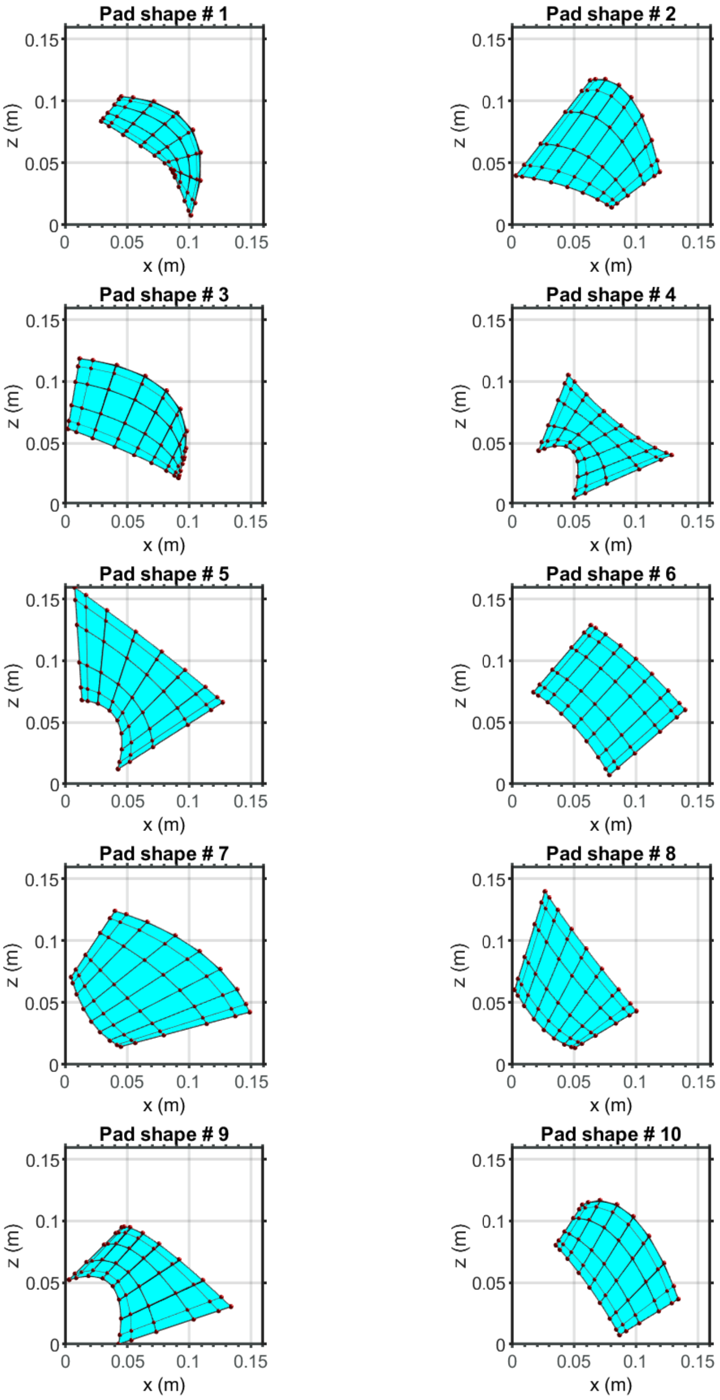


Figure 3.2: Example of 10 pad geometries obtained with the proposed IGA parametrization

3.1.2 Uncertain parameters

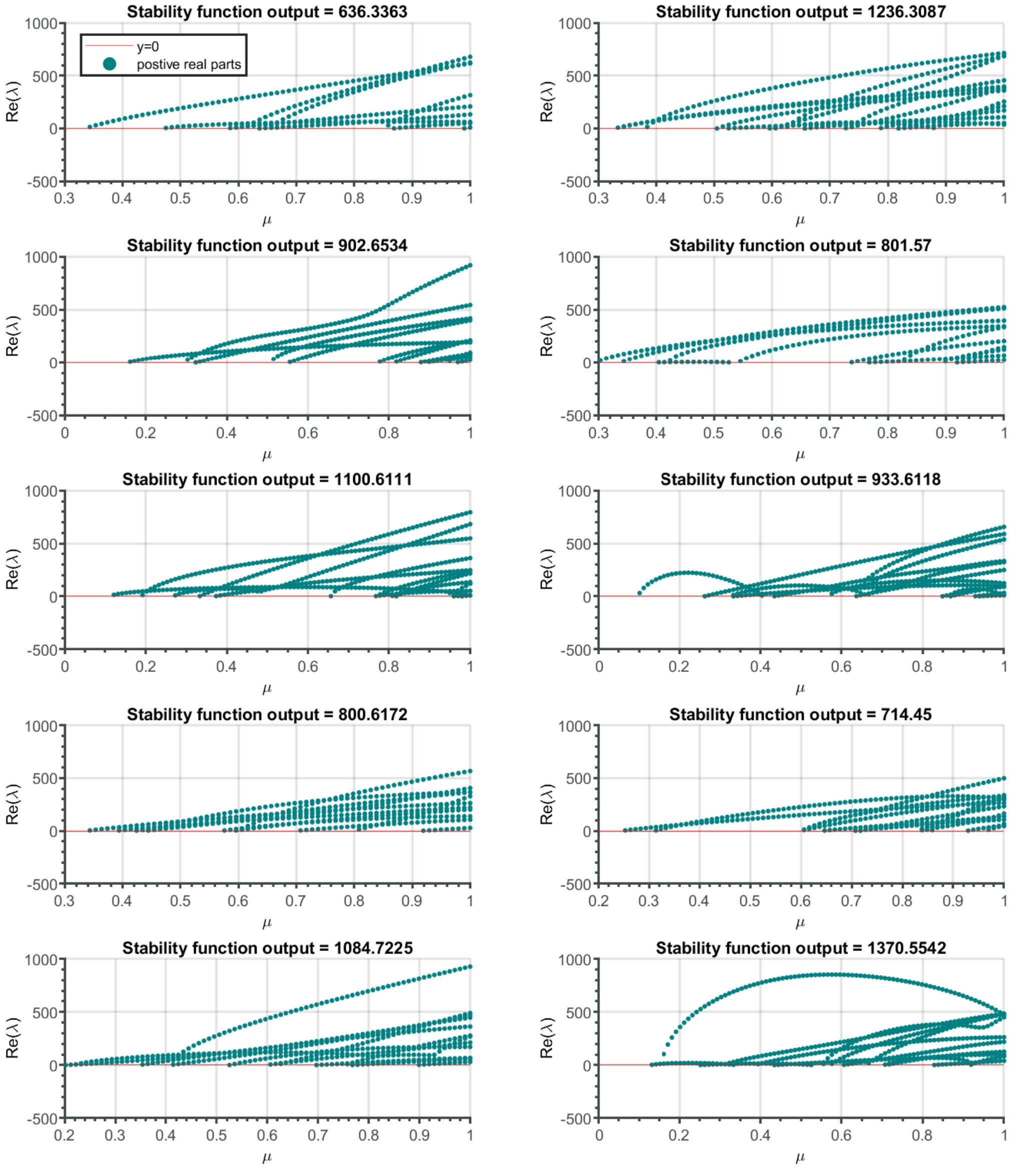
There are many possible sources of uncertainty in the numerical modeling process of mechanical systems (see section 1.3.1). In the presented PhD research, we focus on geometric uncertainties, commonly related to manufacturing processes which can affect the geometry of the numerically modeled mechanical parts and in turn influence the performance in real operating conditions as compared to what is predicted by simulation. However, in the context of the studied application (squeal noise phenomenon exhibited in disc brakes) and taking into account the fact that a number of assumptions were made during the modeling process, we will consider that the modeled geometric uncertainties encompass errors due to manufacturing processes as well as errors due to the inaccuracies in the modeling process. In order to quantify and define these geometric uncertainties, we opt to consider uncertainties directly on the deterministic parameters described in the previous section which are the cylindrical coordinates of the IGA control points before refinement. A set of n_u uncertain parameters $\mathbf{X}_u = [X_1^u, ..., X_{n_u}^u]^T$ is thus considered. Each uncertain parameter X_i^u follows a Gaussian distribution around an average value $\overline{X}_i^u$ and all uncertain parameter values lie in the interval $\overline{X}_i^u \pm 3\sigma_u$. This corresponds to geometric tolerancing as described for instance in the ISO-8015 standards. The specific parameters that are considered uncertain are determined in chapter 4 through sensitivity analysis.

3.2 Objective functions

This section describes the objective functions considered for the robust optimization problem, in the case of the disc brake model. Two objective functions g_1 and g_2 are considered.

3.2.1 Performance function

The first objective function g_1 and main performance function is the stability function C_s defined in section 2.5, which varies with parameters $\mathbf{X}_d$. This function is to be minimized in order to minimize squeal noise occurrence. This function's blackbox character, its relatively high computational cost and its nonlinear behavior make it particularly difficult to study in the context of robust shape optimization. Figure 3.3 shows examples of performance function values and the associated real part stability diagrams (note that only the positive real parts are plotted in this case) for the 10 different pad shapes displayed previously in figure 3.2.


Stability function output = 636.3363

Stability function output = 1236.3087

Stability function output = 902.6534

Stability function output = 801.57

Stability function output = 1100.6111

Stability function output = 933.6118

Stability function output = 800.6172

Stability function output = 714.45

Stability function output = 1084.7225

Stability function output = 1370.5542

Figure 3.3: Stability function values and the associated positive real part stability diagrams for 10 different pad shape designs

3.2.2 Robustness to uncertainties

The second optimization function g_2 is centered around minimizing the impact of geometric uncertainties on the stability function. In order to quantify this, we opt to compute the variance of the performance function output $\sigma^2_{C_s}(\mathbf{X}_d)$ due to the input uncertainties. How this variance is numerically computed is one of the main challenges of this PhD research and will be addressed in the following chapters.

3.3 Constraint functions

In the context of disc brake design, braking capacity is an important aspect. Since the main objective is to minimize squeal noise by affecting the shape of the pad, shapes with a very small surface area may appear during the optimization process. However, such shapes are highly undesirable due to their low braking capacity. For this reason we consider one constraint on the pad surface area (equivalent to the contact surface area) $A(\mathbf{X}_d)$ which much be greater than a predefined minimum value A_{min}.

3.4 Optimization problem recap

The optimization problem applied to the robust disc brake squeal noise minimization case can be summarized as follows:

- **Objectives:**
    1. Minimization of the stability function: $g_1(\mathbf{X}_d) = C_s(\mathbf{X}_d)$,
    2. Minimization of stability function variance with respect to uncertainties (maximization of robustness): $g_2(\mathbf{X}_d) = \sigma^2_{C_s}(\mathbf{X}_d)$.
- **Constraint:**
    1. Insure good braking capacity: $h_1(\mathbf{X}_d) = A(\mathbf{X}_d) \geq A_{min}$.
- **Input parameters:** The (r, θ) cylindrical coordinates of 6 pad shape-defining NURBS control points (12 parameters total, see figure 3.1), $\mathbf{X}_d = [r_1, ...r_6, \theta_1, ..., \theta_6]^T$.
- **Uncertain parameters:** Each uncertain parameter X_i^u follows a Gaussian distribution around an average value $\overline{X}_i^u$ and all uncertain parameter values lie in the interval $\overline{X}_i^u \pm 3\sigma_u$ (ISO-8015 standards).

4 PCE-based strategy

In order to carry out robust shape optimization of complex mechanical structures, multiple challenging aspects must be addressed. In particular, the computational cost of the main performance function and the quantification of its variance due to geometric uncertainties are key aspects to consider when selecting an optimization strategy. In order to tackle these difficulties, we first propose a robust optimization strategy based on sparse Polynomial Chaos Expansions for uncertainty quantification, applied to the simplified disc brake model. This chapter presents the necessary background on PCE, followed by a description of the proposed strategy, and the obtained results in the case of the optimization problem described in chapter 3.

4.1 Polynomial chaos mathematical background

This section presents mathematical background on Polynomial Chaos Expansions, how they are constructed and how their outputs can be used for robust optimization.

4.1.1 Polynomial chaos expansion

First we consider a deterministic physical model $\mathscr{M}$ with finite variance which maps various input parameters $\mathbf{x}$ (geometric, material properties, etc.) to a set of outputs $\mathbf{y}$ such that $\mathbf{y} = \mathscr{M}(\mathbf{x})$. The outputs $\mathbf{y}$ are also referred to as model response and represent any output quantity of interest of the physical model. In the probabilistic context, that is when $\mathbf{x}$ is considered uncertain, the inputs are represented as a random vector $\mathbf{X}$ with associated PDF $f_X(\mathbf{x})$. In this case the response is also a random variable $Y = \mathscr{M}(\mathbf{X})$. In the following we will consider the simpler case where Y is a scalar (only one quantity of interest) which is the case of our studied application. Note that the derivations in the scalar case can be somewhat extended component-wise to the case of a vector-valued response. The finite variance of the model is expressed with the following equation:

$$\mathbb{E}[Y^2] = \int \mathscr{M}^2(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} < \infty \tag{4.1}$$

The polynomial expansion of Y is then carried out by expanding Y onto an orthogonal polynomial basis:

$$Y = \mathscr{M}(\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^M} a_\alpha \Psi_\alpha(\mathbf{X}) \tag{4.2}$$

where $\alpha \in \mathbb{N}^M$ is a multi-index which identifies the components of the polynomial basis, Ψ_α are multivariate polynomials orthonormal with respect to f_X, and a_α are the corresponding unknown deterministic coefficients. Notice that the above expansion is an infinite sum, however, for practical applications it must be truncated to a finite sum, the so-called truncated polynomial expansion $\mathscr{M}^{PC}$, defined as:

$$\mathscr{M}(\mathbf{X}) \approx \mathscr{M}^{PC}(\mathbf{X}) = \sum_{\alpha \in \mathscr{A}} a_\alpha \Psi_\alpha(\mathbf{X}) \tag{4.3}$$

where $\mathscr{A}$ is a subset of $\mathbb{N}^M$ which contains the selected multi-indices α. The multivariate polynomials $\Psi_\alpha(X)$ are built starting from a set of univariate orthonormal polynomials $\phi_k^{(i)}(x_i)$. The univariate polynomials are associated to an inner product which defines their orthogonality properties. A pair of polynomials is orthogonal if:

$$\langle \phi_j^{(i)}(x_i), \phi_k^{(i)}(x_i) \rangle \overset{\text{def}}{=} \int_{D_{X_i}} \phi_j^{(i)}(x_i) \phi_k^{(i)}(x_i) f_{X_i}(x_i) dx_i = 0 \tag{4.4}$$

A sequence of orthonormal polynomials verifies:

$$\langle \phi_j^{(i)}(x_i), \phi_k^{(i)}(x_i) \rangle = \delta_{jk} \tag{4.5}$$

where δ_{jk} is the Kronecker delta, i identifies the input variable with respect to which the polynomials are orthogonal and the corresponding polynomial family, j, k are the associated polynomial degrees, $f_{X_i}(x_i)$ is the i-th input marginal distribution, and $\langle ., . \rangle$ can be interpreted as the expectation value of the product of its components. The multivariate polynomials $\Psi_\alpha(X)$ are then assembled as the tensor product of their univariate counterparts:

$$\Psi_\alpha(\mathbf{x}) \overset{\text{def}}{=} \prod_{i=1}^{M} \phi_{\alpha_i}^{(i)}(x_i) \tag{4.6}$$

Orthonormality of the univariate polynomials leads to orthonormality of the multivariate polynomials:

$$\langle \Psi_\alpha(\mathbf{x}), \Psi_\beta(\mathbf{x}) \rangle = \delta_{\alpha\beta} \tag{4.7}$$

where $\delta_{\alpha\beta}$ is the multidimensional Kronecker delta.

There are multiple families of univariate orthogonal polynomials which are orthogonal to a certain type of probability distribution function. Thus, the univariate polynomials used to build the multivariate polynomial basis can be selected in order to correspond to the probability distribution of the associated univariate random variable. Table 4.1 shows the correspondence between the orthogonal polynomial families and the associated types of probability distributions.

Table 4.1: Common probability distributions and their associated orthogonal polynomial families

Probability distribution	Polynomial family
uniform	Legendre
gaussian	Hermite
gamma	Laguerre
beta	Jacobi

4.1.2 Basis truncation schemes

The previous section explained the need to truncate the PCE to a finite sum in practice. Multiple methods, so-called "truncation schemes", are available to carry out the truncation. The most straightforward method, called the standard truncation scheme, reduces the set $\mathbb{N}^M$ into subset $\mathscr{A}^{M,p}$ as follows:

$$\mathscr{A}^{M,p} = \left\{\alpha \in \mathbb{N}^M : |\alpha| \leq p\right\} \tag{4.8}$$

The number of elements of $\mathscr{A}^{M,p}$ is then:

$$card\mathscr{A}^{M,p} \equiv P = \begin{bmatrix} M+p \\ p \end{bmatrix} \tag{4.9}$$

The maximum interaction truncation scheme is another method which keeps a subset of the standard truncation terms such that the α's have at most r non-zero elements:

$$\mathscr{A}^{M,p,r} = \left\{\alpha \in \mathscr{A}^{M,p} : ||\alpha||_0 \leq r\right\} \tag{4.10}$$

Another common truncation scheme is the *q-norm* truncation scheme (see equation 4.11) (or hyperbolic truncation scheme) which uses the *q-norm* defined in 4.12 to determine the truncation.

$$\mathscr{A}^{M,p,q} = \left\{\alpha \in \mathscr{A}^{M,p} : ||\alpha||_q \leq p\right\} \tag{4.11}$$

$$||\alpha||_q = \left(\sum_{i=1}^{M} \alpha_i^q\right)^{1/q} \tag{4.12}$$

4.1.3 Basis-adaptive PCE

Basis-adaptive PCE is a process through which the best basis of polynomials is chosen from a set of possible bases, by checking which basis yields the best accuracy given the available experimental design samples. Usually this process involves gradually increasing the maximum polynomial degree (in a predefined range), and thus the number of basis elements, until an acceptable value of generalization error is reached. The generalization error is commonly approximated using one of several methods described in section 4.1.4 which are designed to retrieve the error without using any additional experimental samples. Two main types of basis-adaptive processes are commonly used: degree-adaptivity and q-norm-adaptivity, where the available data drives respectively the maximum degree and maximum q-norm of the PCE.

4.1.4 A-posteriori error approximation

A-posteriori error evaluation is an essential aspect when constructing PCE models. Firstly, as stated in the previous sub-section, error approximation is necessary for basis-adaptive PCE, where it conditions the choice of the best candidate PCE. More generally, in surrogate modeling, error estimation is crucial for verifying predictive accuracy, and methods which do not require additional function evaluations compared to those used for surrogate model construction are particularly convenient for computationally expensive models. The so-called generalization error, which is commonly evaluated or approximated in order to quantify the accuracy of the surrogate model, can be defined as follows:

$$\epsilon_{gen} = \frac{\mathbb{E}\left[(\mathscr{M}(\mathbf{X}) - \mathscr{M}^{PC}(\mathbf{X}))^2\right]}{Var(Y)} \tag{4.13}$$

In some cases, for instance when comparing different types of surrogate models built on the same training data, a validation set of function evaluations may be available in which case the generalization can be evaluated directly as:

$$\epsilon_{val} = \frac{N_{val} - 1}{N_{val}} \frac{\sum_{i=1}^{N_{val}} \left(\mathscr{M}(x^i_{val}) - \mathscr{M}^{PC}(x^i_{val})\right)^2}{\sum_{i=1}^{N_{val}} \left(\mathscr{M}(x^i_{val}) - \frac{1}{N_{val}} \sum_{i=1}^{N_{val}} \mathscr{M}(x^i_{val})\right)^2} \tag{4.14}$$

However, in most cases, high computational cost of the original model prevents use of a validation set, and other methods must be used. A normalized empirical estimation of the generalized error can be used, which evaluates how accurately the surrogate model predicts experimental samples:

$$\epsilon_{emp} = \frac{\sum_{i=1}^{N_{val}} \left(\mathscr{M}(x^i) - \mathscr{M}^{PC}(x^i)\right)^2}{\sum_{i=1}^{N_{val}} \left(\mathscr{M}(x^i) - \frac{1}{N_{val}} \sum_{i=1}^{N_{val}} \mathscr{M}(x^i)\right)^2} \tag{4.15}$$

Another, more popular option is the leave-one-out cross-validation error which consists in creating surrogate models using a reduced experimental set which excludes one point from the original full set, and then quantifying the prediction error at the excluded point. This is done successively excluding each point of the experimental set once, and yields the following expression for the approximated generalized error or "leave-one-out cross-validation error":

$$\epsilon_{LOO} = \frac{\sum_{i=1}^{N_{val}} \left(\mathscr{M}(x^i) - \mathscr{M}^{PC\setminus i}(x^i)\right)^2}{\sum_{i=1}^{N_{val}} \left(\mathscr{M}(x^i) - \frac{1}{N_{val}} \sum_{i=1}^{N_{val}} \mathscr{M}(x^i)\right)^2} \tag{4.16}$$

A generalized version of leave-one-out cross-validation, known as leave-k-out cross validation can also be used, where k points (instead of 1 point) are successively excluded from the experimental set. Over-fitting problems with the normalized empirical estimate have led to use of leave-one-out cross-validation in most cases. There are also techniques which can help to avoid underestimation of the error estimation in the case of leave-one-out and normalized empirical errors by applying a correction factor to the original error to obtain a corrected error.

4.1.5 Calculation of the PCE coefficients

This section presents the different available methods for the calculation of the PCE coefficients.

Projection method

The most direct way to determine the coefficients from the definition of the PCE, is to multiply equation 4.2 by Ψ_β and take the expected value of the result:

$$a_\alpha = \mathbb{E}\left[\Psi_\alpha \cdot \mathscr{M}(\mathbf{X})\right] = \int_{D_X} \mathscr{M}(\mathbf{x})\Psi_\alpha(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} \tag{4.17}$$

The integral is then usually computed using Gaussian quadrature:

$$a_\alpha \approx \sum_{i=1}^{N} W^{(i)} \mathscr{M}(x^{(i)}) \Psi_\alpha(x^{(i)}) \tag{4.18}$$

where $W^{(i)}$ are weights and $x^{(i)}$ are the quadrature points, which represent the roots of the corresponding univariate polynomial basis functions. The main drawback of this calculation method is related to the fact that the number of model evaluations needed is equal to the number of quadrature points, and thus increases rapidly with the number of input variables. To alleviate this drawback, a more advanced quadrature method requiring fewer integration points, Smolyak sparse quadrature, can be used. An estimate of the quadrature error on the PCE coefficients can also be obtained.

Regression methods (Ordinary least-squares)

The second main method to determine the PCE coefficients is least-squares regression which consists in formulating the coefficients' calculation as a least-squares minimization problem of the residual between the infinite series $\mathscr{M}(\mathbf{X})$ introduced in equation 4.2 and its truncated version $\mathscr{M}^{PC}$ from equation 4.3:

$$\hat{\mathbf{a}} = \arg\min\ \mathbb{E}\left[\left(\mathscr{M}^{PC}(\mathbf{X}) - \mathscr{M}(\mathbf{X})\right)^2\right] \tag{4.19}$$

which can be written more conveniently as:

$$\hat{\mathbf{a}} = \arg\min\ \mathbb{E}\left[\left(\mathbf{a}^T \mathbf{\Psi}_t(\mathbf{X}) - \mathscr{M}(\mathbf{X})\right)^2\right] \tag{4.20}$$

where $\hat{\mathbf{a}}$ is the final vector of PCE coefficients, $\mathbf{a}$ is the vector of variable PCE coefficients, $\mathbf{\Psi}_t$ is the vector which gathers the orthonormal polynomials Ψ_α, and P is as introduced in equation 4.9.

The minimization problem is most commonly solved using Ordinary Least-Squares (OLS), in which case the solution is given directly by:

$$\hat{\mathbf{a}} = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T Y_{exp} \tag{4.21}$$

where $\mathbf{Y}_{exp} = \{y_e^{(1)}, ..., y_e^{(N)}\}$ is a set of experimental outputs corresponding to a set of N inputs $\mathbf{X}_{exp} = \{x_e^{(1)}, ..., x_e^{(N)}\}$, and $\mathbf{A}$ is the "regression matrix" made up of the values of the basis polynomials at the experimental design points i.e., $A_{ij} = \Psi_j(x_e^{(i)}),\ i = 1, ..., N,\ j = 0, ..., P-1$

Sparse regression methods

Sparse regression methods aim at constructing sparse polynomial chaos expansions, i.e. limiting the number of expansion elements, while keeping high prediction accuracy, in high dimension. The idea is to retain only the most significant polynomial terms in the expansion. Various sparse algorithms have been proposed, the most popular method being Least Angle Regreession (LAR) which is described in the following section. Other recently popular algorithms can be found in [205]. LAR is based on a reformulation of the minimization problem introduced in equation 4.22 by adding a penalization term $\lambda_{pen}||\mathbf{a}||_1 = \sum_{\alpha \in \mathscr{A}} |a_\alpha|$ which favors low-rank solutions:

$$\hat{\mathbf{a}} = \arg\min\ \mathbb{E}\left[\left(\mathbf{a}^T \Psi(\mathbf{X}) - \mathscr{M}(\mathbf{X})\right)^2\right] + \lambda_{pen}||\mathbf{a}||_1 \tag{4.22}$$

The LAR algorithm operates on the idea of moving potential sets of coefficients or "regressors" between a "candidate set" and an "active set". In particular, it uses correlation with the current residual to select new regressors and equi-correlation with the current residual (through analytical relations) to determine the best set of coefficients for each active set. A-posteriori error estimation (leave-one-out) is necessary to select the best sparse polynomial basis once all the iterations are finished. Disadvantages of the LAR method are that it requires re-building models to compute the leave-one-out error, and only non-constant regressors are handled. Hybrid-LAR, which uses OLS for the surrogate model building portions of the algorithm, was introduced to overcome these drawbacks, and is now commonly used. In terms of performance, LAR has been shown to deliver particularly accurate PCE models even with relatively modest experimental designs.

4.1.6 Moments of a PCE and interest for uncertainty quantification

The orthonormality property of the polynomials used to construct a PCE model is extremely interesting from an uncertainty quantification point of view, as it means that the mean value and the variance of the modeled function can be obtained directly from the coefficients of the PCE. They are given by:

$$\mu_{PCE} = \mathbb{E}\left[\mathscr{M}^{PC}(\mathbf{X})\right] = a_0 \tag{4.23}$$

$$\sigma^2_{PCE} = \mathbb{E}\left[\left(\mathscr{M}^{PC}(\mathbf{X}) - \mu_{PCE}\right)^2\right] = \sum_{\alpha\in\mathscr{A}\backslash 0} a^2_\alpha \tag{4.24}$$

The total variance of the model due to uncertainties defined by a probability distribution function, can be obtained through the coefficients of a PCE model constructed with a relatively small sample set, in other words it can be obtained at a much lower cost compared to the traditionally required Monte-Carlo simulation. PCE can also be effectively used for sensitivity analysis, for instance with the Sobol method [206], where the Sobol indices can also be obtained directly from the PCE coefficients.

4.2 Application to the disc brake optimization problem

The goal is to evaluate the ability of the proposed PCE-based approach to solve a robust shape optimization problem under gradient-free nonlinear performance metric, by applying it to the disc brake optimization problem formalized in chapter 3. The main idea is to use the uncertainty quantification capabilities of the PCE method to quantify the effect of uncertainties on the stability function output. Note that implementation of PCE was carried out using UQLab [207], which is an open-source Matlab framework for uncertainty quantification.

4.2.1 Problem setting and resolution

First we recall the optimization problem described in chapter 3 while integrating the components used for the proposed PCE-based approach. The N_d input parameters $\mathbf{X}_d = [x^d_1, ..., x^d_{n_d}]^T$ are the coordinates of the IGA control points $x_i = P_{i,j}$, $i = 1...n_d$ chosen for shape modification. $j = 1...d_s$ identifies the components of each coordinate vector, d_s being the dimension of the shape description. Among these parameters, n_u parameters $\mathbf{X}_u = [x^u_1, ..., x^u_{n_u}]^T$ are considered uncertain and each parameter's uncertainty is defined through a PDF $f_{x^u_k}$, $k = 1...n_u$ around the deterministic nominal parameter value. Objective function g_1 is the studied system's main performance function which is the stability function $C_s(\mathbf{X}_d)$. In the proposed PCE-based approach we do not attempt to relieve the computational cost of the stability function and thus use it directly in the optimization process without applying any metamodeling strategy. Objective function g_2 quantifies the variability of function g_1 to uncertainties. Function g_2 is evaluated using the output variance (see equation 4.24) given by a local PCE model constructed around each potential uncertain input associated to the joint uncertain input PDF $f_{\mathbf{X}_u}$:

$$g_2(\mathbf{X}_d) = \sigma^2_{PCE}(\mathbf{X}_u) \tag{4.25}$$

Constraint function h_1 guarantees sufficient braking capacity by constraining the contact surface area, and is taken into account as an inequality $h_1(\mathbf{X}_d) = A(\mathbf{X}_d) \geq A_{min}$.

The bi-objective optimization problem is finally solved using the NSGAII [208] method. Figure 4.1 schematically shows how the different components of the proposed strategy are combined to obtain a Pareto front of robust optimal solutions. The IGA model described in chapter 2 yields parameters which describe the shape of the system which are then used as input parameters for the optimization problem. It also yields the analysis of the system which in turn provides the main performance function which can then be evaluated for different input parameter combinations (objective function g_1). A PCE model is constructed to obtain the variance to uncertainties around each potential optimization solution (objective function g_2). The optimization problem described in section 3 is then solved using NSGAII [4] to obtain a Pareto front of solutions which balance performance and robustness of the performance function to uncertainties. The parameters of NSGAII used for resolution are summarized in table 4.2. Constraints are not represented on figure 4.1 but are considered through NSGAII.

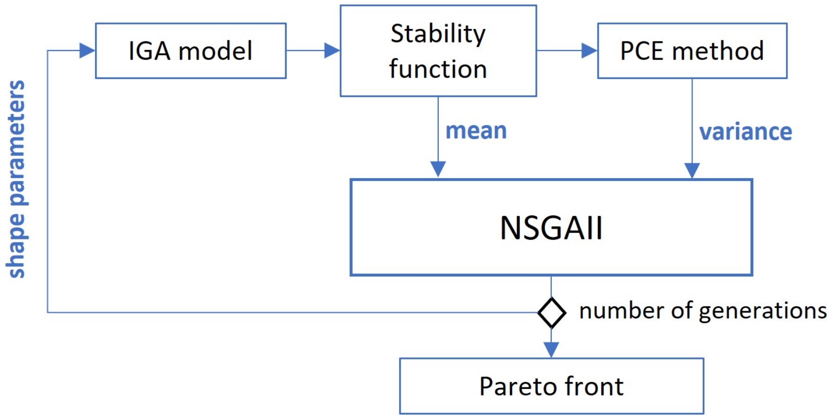


Figure 4.1: Scheme of the resolution process used to obtain a Pareto front of robust optimal solutions with the PCE-based approach

Table 4.2: NSGA2 parameters

Parameter	Parameter value
crossover probability	0.9
mutation probability	0.5
crossover distribution index	10
mutation distribution index	20

Sensitivity analysis

For the proposed PCE-based approach, we consider $N_u = 4$ uncertain parameters. In order to select which out of the 12 shape parameters have the most effect on the response when considered uncertain, we conduct sensitivity analysis. We opt for the Sobol total indices calculated through PCE approximation as sensitivity indicators. Details on the Sobol total indices can be found in [206]. The uncertain parameters must be varied during the optimization process, which means that the sensitivity of the output to each parameter cannot be measured for uncertainties around only one nominal value for each parameter. Rather, the sensitivity to uncertainties should be taken around multiple deterministic samples representative of the whole design space. For this reason, 100 samples of the 12 deterministic parameters are drawn across the design space, using an LHS sampling strategy, and the sensitivity due to uncertainties around each of these samples and associated to each individual shape parameter is evaluated through Sobol total indices. The Sobol indices are calculated by constructing a PCE model based on a Gaussian distribution of uncertainties around each sample. The Sobol total indices are then stacked (summed) for each of the 12 input parameters. Figure 4.2 shows the obtained result. The result shows that all 12 parameters have substantial effect, with parameters r_2, r_6, θ_1, and θ_6 having overall the most effect on the response. These 4 parameters are thus chosen as the uncertain parameters in the robust optimization problem. Note that all PCE models used for the computation of the Sobol indices were constructed with a sufficient number of experimental data (for representation of the uncertainties) in order to guarantee leave-one-out error not greater than 1%.

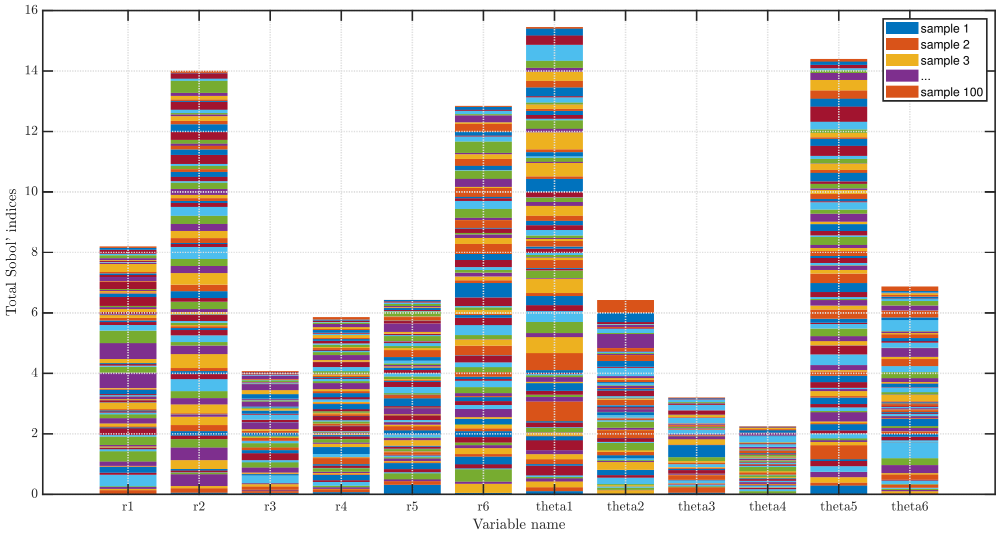


Figure 4.2: Sensitivity analysis of the stability function with respect to the twelve optimization parameters, over 100 experimental samples representative of the design space, summed for each input parameter

4.3 Results

This section presents the main results obtained using the PCE-based optimization approach.

4.3.1 PCE output variance convergence

In this section we observe the evolution of the stability function output variance as predicted by the PCE model as the number of function samples used to build the PCE model is progressively increased, and we compare it to a reference Monte-Carlo-based variance value, in order to get an idea of the attainable precision on the variance and of the number of function samples necessary to obtain said precision. Figure 4.3 shows the stability function's PCE model output variance due to input uncertainties on the four selected uncertain parameters, as a function of the number of experimental samples for one arbitrarily selected pad geometry design. This is plotted in the case of the OLS (top left) and LARS (bottom left) methods, and in each case, PCE model error indicators are also plotted as a function of the number of samples (top right and bottom right). A reference value for the output variance, computed using 10000 Gaussian-distributed uncertain Monte-Carlo (MC) experimental samples, is also displayed. For both PCE methods general convergence of the output variance towards the reference value is observed, however, the results for the arbitrarily selected geometry do indicate that the LARS-PCE model's variance may converge somewhat faster than that of the OLS-based model, which would suggest better overall performance for LARS compared to OLS at an equivalent number of training samples. Moreover, it can be seen that the output variance and the error indicator variations are not smooth with respect to the number of experimental samples, which makes it difficult to select a number of samples which guarantees sufficient convergence of the variance estimation for any given pad geometry.

Figure 4.4 shows an analysis of the coefficient distribution for the OLS and the LARS PCE models used in the above variance convergence test, in the case of 100 training samples. The PCE-coefficient analysis consists in plotting the magnitude of the PCE coefficients (in log scale) as a function of their enumeration (i.e. a_α in equation 4.3). The coefficients with the lowest polynomial degree appear on the left-hand side of the graph whereas the polynomial degree gets higher towards the right-hand side. Since the basis of polynomials is orthonormal, it is possible to compare the coefficient magnitudes. In this case we can observe that the OLS-PCE model actually uses a relatively restricted number of coefficients, with a polynomial degree of at most 2, which would suggest there is no need to use sparse PCE. However, in the case of the LARS-PCE model, the distribution of coefficients is significantly different. The number of degree 2 coefficients decreases from 10 to 3, and 8 coefficients of degree 3 with a significant magnitude appear. This suggests that the LARS algorithm is able to find a more effective set of PCE coefficients than the basic OLS method, which seemingly increases the accuracy of the output function's variance prediction.

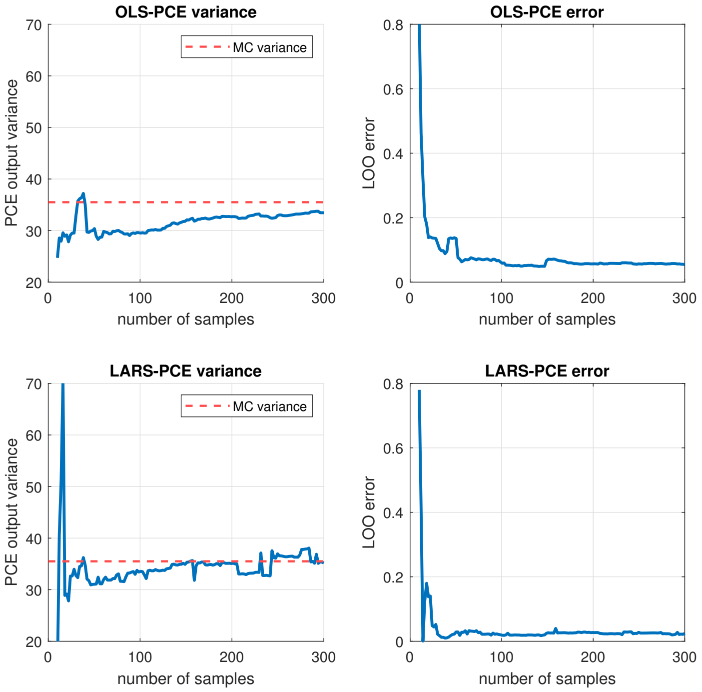
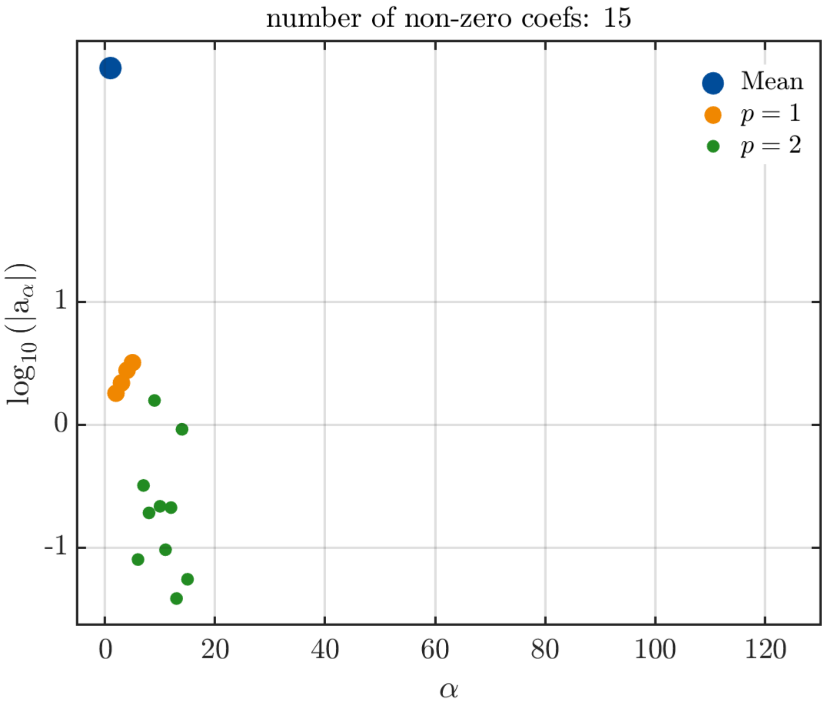
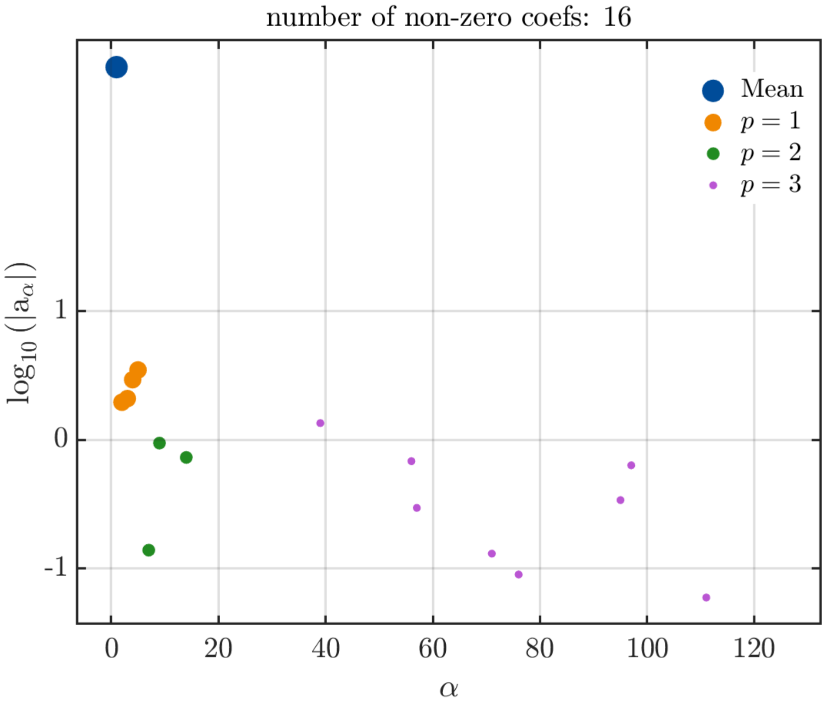


Figure 4.3: Evolution of the stability function PCE model variance (top) and LOO error (bottom), for the OLS (left) and LARS (right) methods, as a function of the number of experimental samples, for one arbitrary pad shape design



(a) OLS



(b) LARS

Figure 4.4: Coefficient distribution for a PCE model of the stability function constructed with OLS (top) and with LARS (bottom)

4.3.2 Bi-objective optimization results: case 1

The proposed PCE-based strategy is first tested with two different algorithms for the PCE coefficient search: OLS and LARS. The idea is to compare a non-sparse and a sparse regression method in terms of accuracy for the prediction of robustness to uncertainties. In this case, the material parameters used for the pad part are: $E = 2GPa$, $\nu = 0.1$, $\rho = 2500kg/m^3$. For each PCE-model construction, 100 experimental samples are used, knowing that one PCE model per potential pad geometry is necessary to evaluate the stability function's variance to uncertainties.

Figures 4.5 and 4.6 show the Pareto front after 10 NSGAII generations respectively when using the OLS method and the LARS method for the output variance estimation. The Pareto front contains the set of non-dominated solutions after 10 generations with a population of 30 individuals per generation, and represents different levels of compromise between the stability function minimization (i.e. noise occurrence minimization) and stability function variance minimization (i.e. maximization of robustness to uncertainties). The full optimization process, in the case of each coefficient calculation method, required roughly 9 days using a computational machine with a 32-core 2.895 GHz processor and 1 TB of RAM, and using 12 workers on Matlab's parallel pool. In each case we computed validation points for the variance estimation (three for the OLS method and four for the LARS method), using Monte Carlo simulations of 10000 samples for each point as a reference, in order to verify the accuracy of each PCE method in predicting the output variance to uncertainties. The number of validation points is limited by the fact that each point requires considerable computational cost. In the case of the OLS method (see figure 4.5), all three validation points' variances are badly predicted, which means that the obtained Pareto front is extremely unreliable. Notice that the variance is underestimated for all three points which may explain the low number of total individuals in the Pareto front, as entire portions of the real Pareto front may be eliminated by the individuals with non-converged underestimated variance. On the other hand, the LARS method (see figure 4.6) delivers a more regular Pareto front with a wider range of stability function and variance values. All four validation points are predicted well enough to consider that the obtained Pareto front approaches to some extent the "real" Pareto front. However validation points 2 and 3 (pink and red) do suggest that the variance estimation could still be improved. Overall, the LARS method delivers better results than the OLS method, which is coherent with the result of section 4.3.1, although this result was only based on a single potential solution (i.e. one pad shape).

Figure 4.7 shows the disc brake pad shapes corresponding to different optimization solutions in the Pareto front obtained with the LARS-PCE method for the variance estimation (not all the solutions' shapes are plotted to avoid overcrowding the figure), with the corresponding pad surface areas summarized in table 4.3. The first observation is that there are three main types of obtained shapes, which correspond to three "clusters" of solutions within the Pareto front. Secondly, among these three types of shapes, we can observe that shapes with higher contact surface area are more prone to noise but also more robust to uncertainties, which is intuitively a physically sound result. Shapes with lower contact surface area are conversely less to prone to noise but more sensitive to uncertainties. Finally, it is important to highlight the originality and diversity of shapes enabled by the IGA shape description.

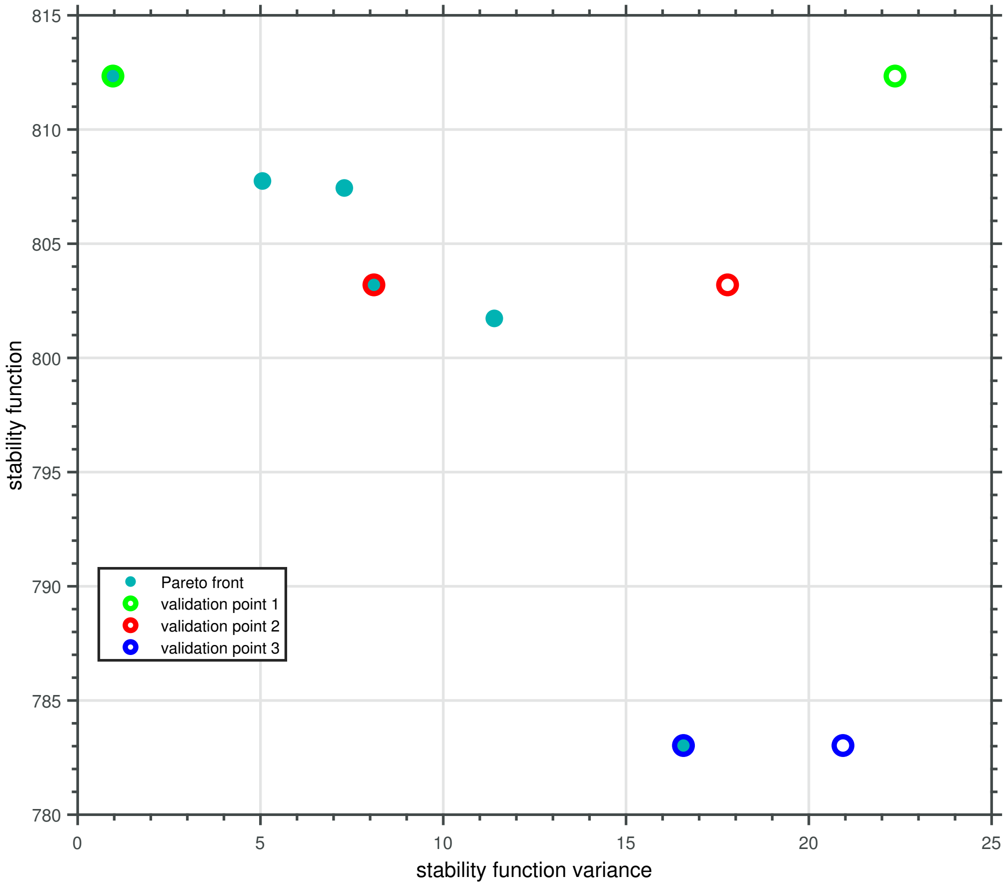


Figure 4.5: Pareto front of solutions which balance minimization of the stability function and its variance to uncertainties, obtained after 10 NSGAII generations, using the OLS-PCE method for variance estimation, and with three Monte-Carlo validation points for variance estimation verification (case 1)

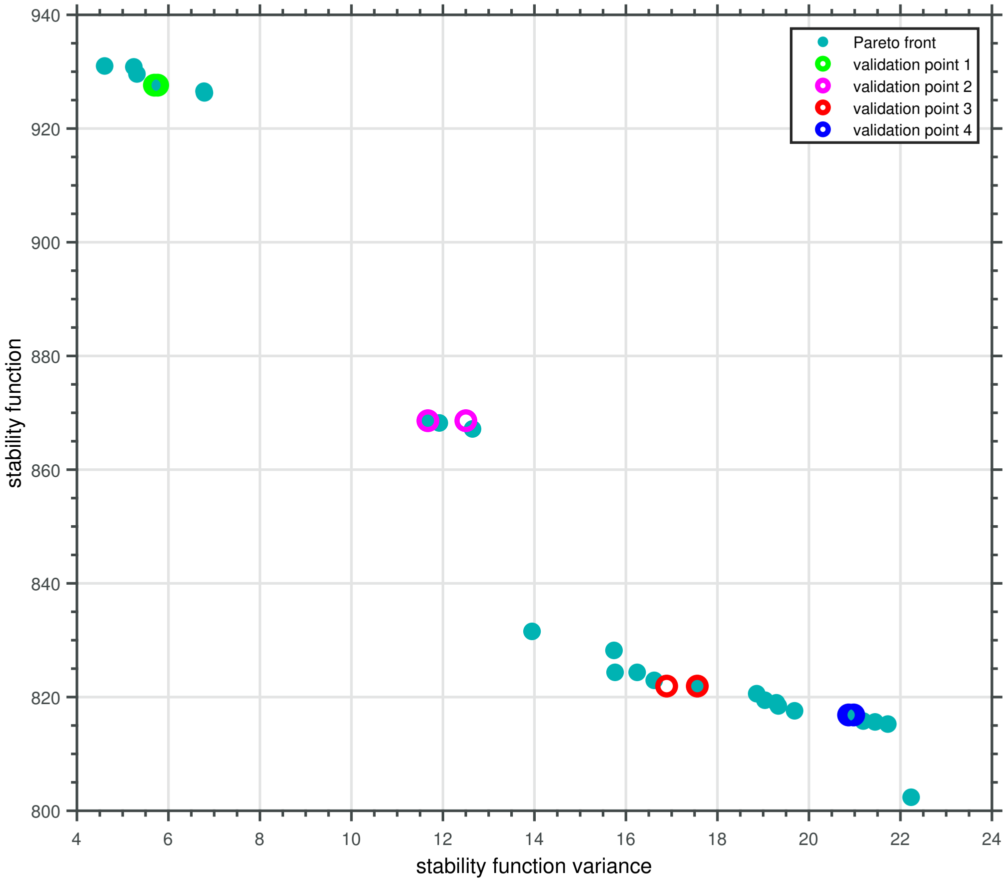


Figure 4.6: Pareto front of solutions which balance minimization of the stability function and its variance to uncertainties, obtained after 10 NSGAII generations, using the LARS-PCE method for variance estimation, and with four Monte-Carlo validation points for variance estimation verification (case 1)

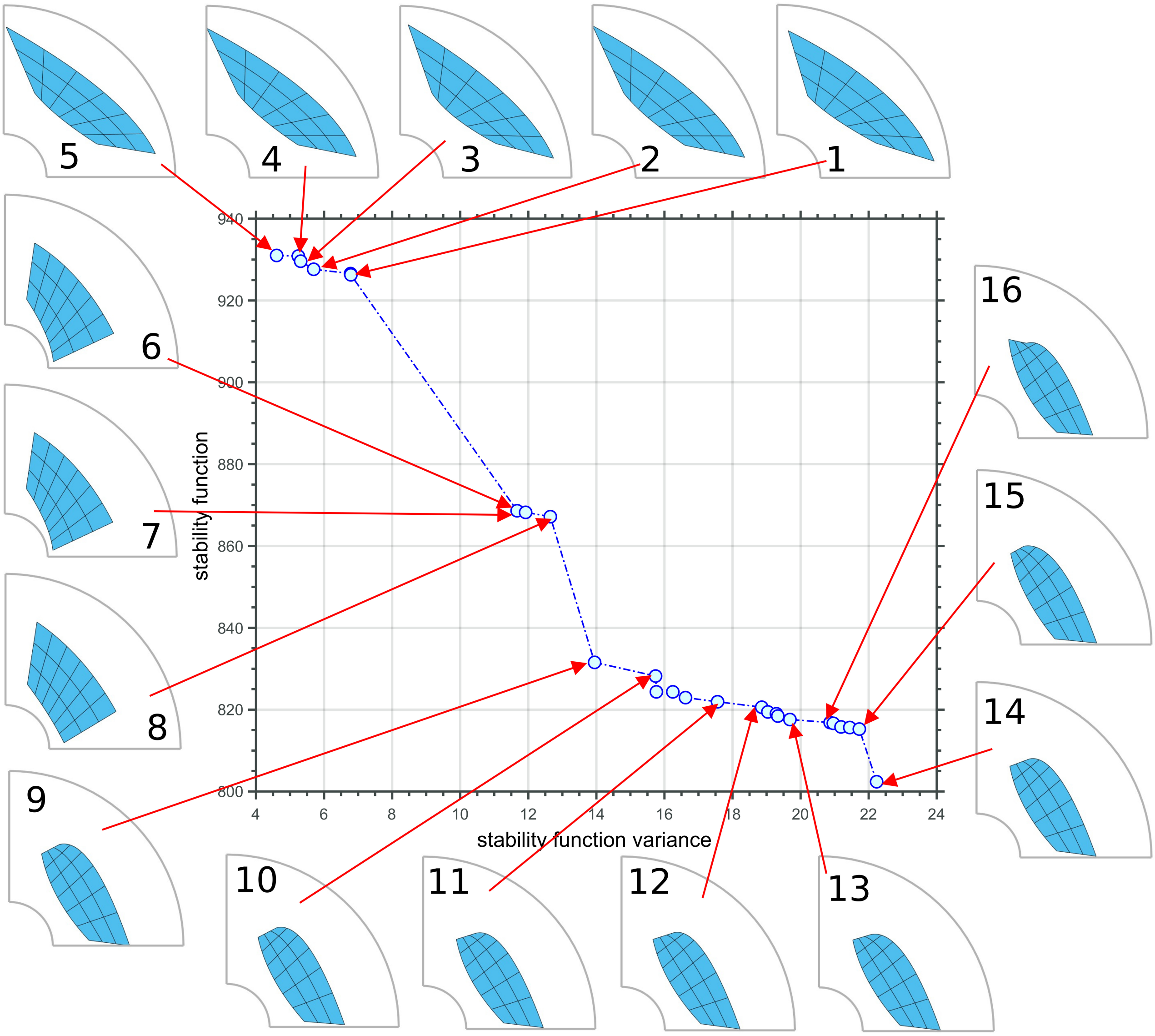


Figure 4.7: Pareto front of solutions which balance minimization of the stability function and its variance to uncertainties, obtained after 10 NSGAII generations, using the LARS-PCE method for variance estimation, with the corresponding pad shape designs for 16 out of 26 solutions (case 1)

Table 4.3: Pad shape solution surface areas for optimization case 1

Shape number	Surface area (cm^2)
1	57.13
2	56.51
3	53.93
4	54.66
5	53.85
6	45.69
7	45.63
8	45.83
9	35.83
10	38.71
11	36.80
12	36.63
13	35.96
14	34.53
15	36.00
16	35.77

4.3.3 Bi-objective optimization results: case 2

In this section we test the same PCE-based strategy with the LARS method only on a slightly different optimization problem where the pad material has been changed compared to section 4.3.2. The material parameters used for the pad in this section are: $E = 16.22GPa$, $\nu = 0.22$, $\rho = 2700kg/m^3$. For each PCE-model construction, 100 experimental samples are used similarly to the previous section.

The Pareto front of solutions which balance low noise potential and robustness to geometric uncertainties after 15 NSGA2 generations with 50 individuals per generation, is shown in figure 4.8. Note that the settings used for NSGA2 are the same as in case 1 (see table 4.2). The full optimization process required roughly 21 days using the same computational machine as in case 1 (see section 4.3.2). Firstly, we notice that validation points 1 to 4, all located in the low variance region of the obtained Pareto front, are very well predicted by LARS-PCE. Then, as we progress along the Pareto front towards solutions with higher variance and less potential for noise, the validation points' prediction accuracy appears to deteriorate. The obtained result suggests that LARS-PCE is more efficient in predicting low variance solutions, which may seem slightly counter-intuitive as one may expect higher variance solutions to be easier to predict. Overall, even though the poorly predicted variance suggests that the proposed LARS-PCE-based method may be inaccurate in some cases, the obtained result is still very positive in the context of industrial mechanical design, as considerable gains in terms of computational cost are obtained compared to directly applying Monte-Carlo for the variance prediction inside the optimization loop. However, the lack of accuracy observed for some of the Pareto front solutions suggests that the proposed method should be applied in conjunction with Monte-Carlo validation if applied in an industrial context.

For analysis of the Pareto front from a physical point of view, figure 4.9 shows 13 out of the 49 optimal solution shapes, while table 4.4 lists the corresponding pad surface areas. Similarly to case 1, the shapes with higher surface area tend to be more prone to noise but more robust to uncertainties while the shapes with lower surface area are less prone to noise but less robust to uncertainties. This is physically consistent with what is expected. Again, the diversity and originality of shapes obtained through IGA is confirmed with at least four different pad shape types in the Pareto front in this case. Note that the overall values of the stability function are lower than in case 1, which is consistent with the change of pad material and the results on the impact of the pad material of section 2.6.5.

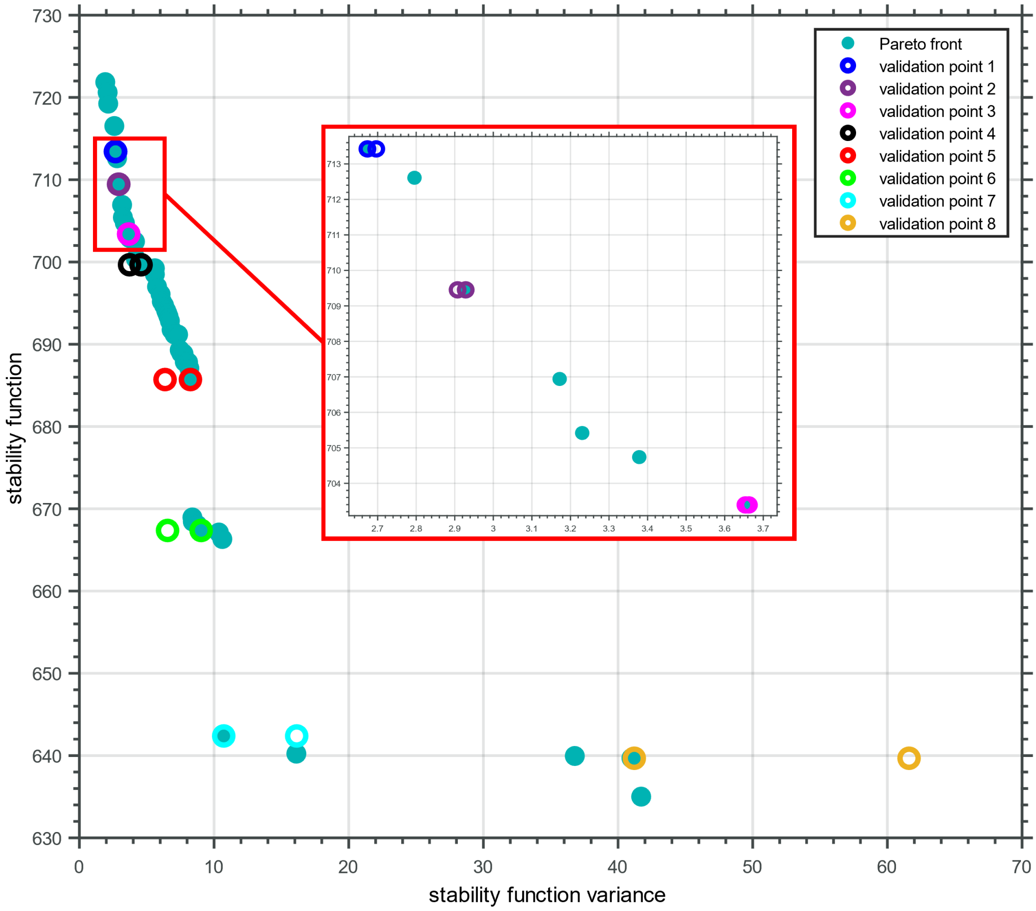


Figure 4.8: Pareto front of solutions which balance minimization of the stability function and its variance to uncertainties, obtained after 15 NSGAII generations, using the LARS-PCE method for variance estimation, and with eight Monte-Carlo validation points for variance estimation verification (case 2)

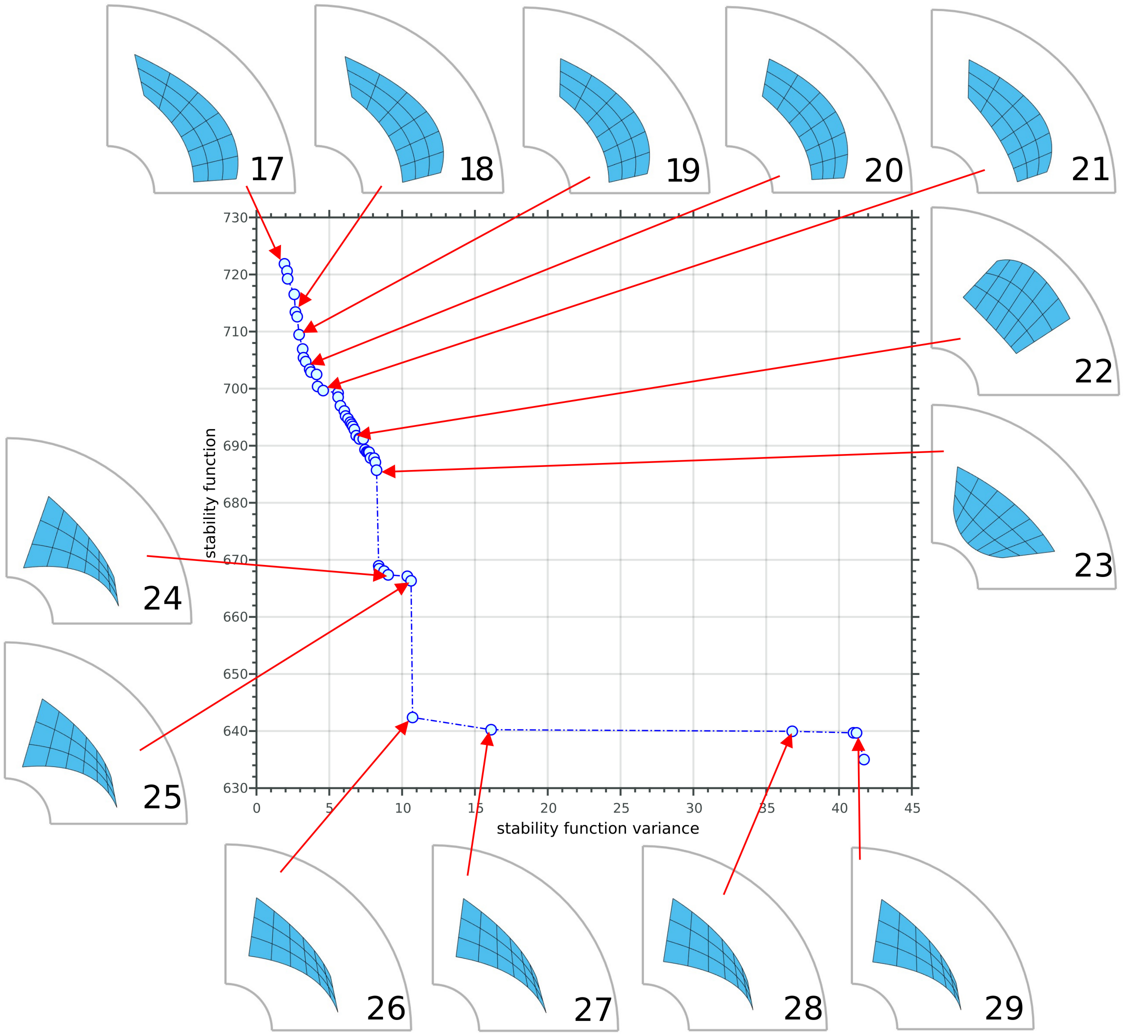


Figure 4.9: Pareto front of solutions which balance minimization of the stability function and its variance to uncertainties, obtained after 15 NSGAII generations, using the LARS-PCE method for variance estimation, with the corresponding pad shape designs for 13 out of 49 solutions (case 2)

Table 4.4: Pad shape solution surface areas for optimization case 2

Shape number	Surface area (cm^2)
17	41.24
18	39.79
19	39.79
20	38.44
21	39.09
22	41.30
23	38.75
24	28.31
25	29.30
26	24.36
27	24.77
28	27.66
29	26.85

4.4 Important takeaways from the PCE-based strategy

This chapter presented a full robust shape optimization methodology, based on a combination of sparse polynomial chaos expansions to address the robustness aspect and a genetic algorithm to search for optimal solutions, applied to the disc brake instability problem.

The proposed approach was shown effective in providing a Pareto front of solutions which balance low noise potentiality (low instability) and high robustness to geometric uncertainties. The quantification of robustness through estimation of output variance to uncertainties was achieved at a significantly lower cost using PCE compared to classical Monte Carlo simulation which requires thousands of output evaluations, even though some loss in precision is inevitable.

The comparison of two variants of the PCE method, OLS and LARS, showed the high efficiency of LARS which was able to approach true output variances to a satisfactory extent in the pursuit of the bi-objective Pareto front, while OLS was insufficient in this aspect, at an equivalent number of training samples. The superiority of LARS compared to OLS is coherent with the literature on sparse polynomial chaos methods. However, the LARS method was found to produce polynomial expansions with a larger number of coefficients than with OLS which yields a relatively low number of coefficients, which goes against what is expected when comparing sparse and non-sparse PCE. It was observed that some of the degree 2 coefficients appearing in the OLS distribution are replaced by degree three contributions with similar magnitude in the LARS case. These coefficients are most likely the source of the higher accuracy achieved by LARS in terms of output variance prediction.

Inaccuracy in the prediction of variance to geometric uncertainties using LARS-PCE is noticed for some Pareto front solutions, which suggests that the proposed method should be used in conjunction with Monte-Carlo validation in the context of industrial design.

Additionally, physical insight on the disc brake robust noise minimization problem was obtained, namely that pad geometries with higher contact surface area are more prone to noise but also more robust to uncertainties, and vice versa. The efficiency of the IGA formulation for shape description, particularly in comparison to classical FEM, is to be noted even though it is not clearly demonstrated in the results. Nonetheless, its ability to produce diverse and original shapes is shown to some extent.

The obtained results also highlight three main issues linked to the high computational burden involved in robust optimization problems for complex gradient-free functions. Firstly, the available computational resources combined with the complexity of the problem restricts the number of generations, the population sizes (for genetic optimization) and the number of reference points that can be used. Higher computational resources would allow to perform the optimization process with more generations, individuals and reference points, and would give further insight into the efficiency of the proposed approach. Secondly, the variance estimations were all performed using PCE models constructed with a fixed number of experimental samples. In the interest of reducing computational cost and potentially increasing accuracy, it would be interesting to attempt progressively increasing the number of experimental samples until some tolerance on the PCE model error (e.g. LOO error) is reached. The main difficulty in this case would be the definition of the tolerance. Finally, from a broader perspective, in the context of robust optimization, a clear takeaway of this chapter is that taking into account uncertainties in an optimization process requires considerable compromise between computational cost and accuracy. Although the proposed PCE-based variance estimation is much less costly than Monte Carlo simulation, the gain in cost comes at the price of a decrease in accuracy on the estimated variances. When considering uncertainties in complex optimization problems, computational cost related aspects such as those highlighted above will continue to be an inevitable challenge, at least until substantial advances are made in the development of more powerful computational means.

5 Kriging-based strategy

This chapter presents an alternative robust optimization strategy based on Bayesian Optimization and the Kriging metamodeling technique. The idea is to replace the evaluation of the costly objective function by a less expensive metamodel, and then use concepts from BO in order to enrich the metamodel in regions of the design space which are the most interesting in terms of the optimization objectives. One of the main advantages of the Kriging method for metamodeling is that it includes an estimation of the prediction error through output variance. For the quantification of the variance due to uncertainties we attempt to use a variant of the Kriging model called noisy Kriging originally designed to take into account measurement errors into the Kriging variance in addition to the effect of the prediction error. This chapter first introduces the theoretical background of the Kriging method and its "noisy" counterpart, before presenting the proposed strategy for robust optimization and the obtained results, in the case of the simplified disc brake system.

5.1 Kriging background

Kriging was originally used in the geostatistical and mining field and mathematically formalized by Matheron in 1963 [209] following the empirical work of Danie G. Krige aiming at determining the geographical distribution of gold locations based on a restricted number of boreholes. Kriging has since been largely extended to diverse engineering problems and in higher dimensions. In the machine learning context it is also known as "Gaussian Process Regression" (GPR).

Kriging can be thought of as spatial prediction, that is, predicting values at locations for which we do not have observations based on available samples. The main idea is that the way that a process varies across space can be characterized by two main components: a large scale variation called the trend, and a small scale spatial auto-correlation called the error term. It can also be interpreted as a signal-noise type of decomposition where the signal captures the large scale variation and the noise is what's left. Under suitable assumptions, Kriging is the Best Linear Unbiased Predictor (BLUP). This notion arises when the assumption that the variance is constant and that there is no covariance between the different observations, is false. A "linear" predictor is one where the prediction is a linear combination of the observation outputs. "Unbiased" means that the predictor has no systematic error and that the mean prediction error is zero, or in other words that the expected value is on average equal to the true value. The term "best" has to do with the precision of the prediction, in other words the variability of the prediction error. Among all linear unbiased predictors, Kriging gives the most precise prediction. To obtain the BLUP, the prediction is expressed as a weighted sum of the observations. An optimality criterion is used which guarantees that on average the deviation between the observed value and the predicted value is zero (unbiasedness) and that the variance of the prediction error is the smallest possible. The specificity of Kriging with respect to generic weighted interpolation is that the weights are connected to the spatial covariance. The BLUP is used by exploiting the covariance matrix between all the observed locations as well as the covariance between the new location and the observed locations. The basic idea is to use the fact that there is a structure in the covariation between pairs of locations and that this structure depends on the distance between the locations. In addition to the predicted values, Kriging also provides an explicit expression for the prediction error, formalized as the prediction variance, which indicates the quality of the prediction. Since two observations which are further apart from each other have smaller covariance, the improvement to the precision is greater from observations which are close together.

There are different main variants of the Kriging method. The classification depends not on the error term but rather on the way that the large scale variation (trend) is modeled. The three main types of Kriging are:

- Simple Kriging
- Ordinary Kriging
- Universal Kriging

Simple Kriging assumes that there is no large scale variation, i.e., that the mean is constant and known, so it does not need to be estimated. Ordinary Kriging also assumes a constant mean but unknown, in other words it needs to be estimated. The model in this case focuses on prediction of deviations from the unknown mean. In practice the mean is not constant, it varies with the location, in which case Universal Kriging is used. In this case the model focuses on the covariances among the errors terms of the regression model using the residuals.

5.1.1 Kriging formalism

The main equations for the Kriging prediction are detailed in [210]. A brief summary is repeated here. Equation 5.1 shows the two parts of the Kriging model, the first is the simple linear regression of the data, or “trend” and the second is the systematic deviation from the linear part:

$$\hat{Y}(\mathbf{X}_K) = \sum_{j=1}^{k_r} \beta_j f_j(\mathbf{X}_K) + Z(\mathbf{X}_K) \tag{5.1}$$

where $Z(\mathbf{X}_K)$ is a random process assumed stationary and with zero mean. Its covariance is given in the following:

$$cov(\mathbf{X}_1, \mathbf{X}_2) = \sigma_Z^2 R(\mathbf{X}_1 - \mathbf{X}_2) \tag{5.2}$$

where σ_Z^2 is the variance of $Z(\mathbf{X}_K)$, and R is the spatial correlation function which controls the smoothness, the influence of nearby points, and the differentiability of the model. The choice of correlation function R is important and must be made carefully depending on the application at hand. There are several common types of correlation functions which are summarized in [211]. In this study, we apply the Matérn 5/2 correlation function, which we chose through trial-and-error on the simplified disc brake model. The Matérn 5/2 correlation function is defined below:

$$R(\mathbf{X}_1 - \mathbf{X}_2) = R_{M52}(\boldsymbol{\eta}) = \prod_{k=1}^{d_X} \left(1 + \frac{\sqrt{5}|\eta_k|}{\tau_k} + \frac{5\eta_k^2}{3\tau_k^2}\right) e^{\left(-\frac{\sqrt{5}|\eta_k|}{\tau_k}\right)} \tag{5.3}$$

where $\mathbf{X}_1 - \mathbf{X}_2 = \boldsymbol{\eta} = \{\eta_1, ..., \eta_{d_X}\}$, d_X is the size of $\mathbf{X}_1$ and $\mathbf{X}_2$ (number of deterministic input parameters of the physical model), R_{M52} is the Matérn 5/2 correlation function, and the τ_k values are hyperparameters of the Matérn 5/2 correlation function commonly called "characteristic length scales" which are tuned during the Kriging model construction process.

In order to construct a Kriging model, a set of n_s samples or "observations" of the modeled function are needed, characterized by inputs $\mathbf{X}_K$, and the corresponding outputs $\mathbf{Y}_K$:

$$\begin{aligned} \mathbf{X}_K &= \{\mathbf{X}_1, ..., \mathbf{X}_{n_s}\} \\ \mathbf{Y}_K &= \{Y_1, ..., Y_{n_s}\} \end{aligned} \tag{5.4}$$

The Kriging model is first and foremost a linear predictor, that is each prediction is a linear combination of the observation outputs:

$$\hat{Y}(\mathbf{X}_K) = \lambda_K^T(\mathbf{X}_K)\mathbf{Y}_K \tag{5.5}$$

The specificity of the Kriging approach is to find the BLUP considering $\hat{Y}(\mathbf{X}_K)$ as a random variable and by searching for the optimal set of "weights" $\boldsymbol{\lambda}_K$. The "best" predictor is found by minimizing the mean square error of the prediction (see equation 5.6) while guaranteeing unbiasedness with the constraint on the expected value in equation 5.7.

$$\mathbb{MSE}\left[\hat{Y}(\mathbf{X}_K)\right] = \mathbb{E}\left[\boldsymbol{\lambda}_K^T(\mathbf{X}_K)\mathbf{Y}_K - \hat{Y}(\mathbf{X}_K)\right]^2 \tag{5.6}$$

$$\mathbb{E}\left[\boldsymbol{\lambda}_K^T(\mathbf{X}_K)\mathbf{Y}_K\right] = \mathbb{E}\left[\hat{Y}(\mathbf{X}_K)\right] \tag{5.7}$$

A vector $\mathbf{F}_X$ of the set of regression functions evaluated at all the observation points is defined, $\mathbf{f}$ being the vector of all regression functions:

$$\mathbf{F}_X = \begin{bmatrix} f_1(\mathbf{X}_1) \\ \vdots \\ f_k(\mathbf{X}_1) \\ f_1(\mathbf{X}_2) \\ \vdots \\ f_k(\mathbf{X}_2) \\ \vdots f_1(\mathbf{X}_{n_s}) \\ \vdots \\ f_k(\mathbf{X}_{n_s}) \end{bmatrix} = \begin{bmatrix} \mathbf{f}(\mathbf{X}_1) \\ \mathbf{f}(\mathbf{X}_2) \\ \vdots \\ \mathbf{f}(\mathbf{X}_{n_s}) \end{bmatrix} \tag{5.8}$$

The correlation matrix, which is constituted of the evaluation of the spatial correlation function at all pairs of known points is also defined:

$$\mathbf{R}_X = \begin{bmatrix} R(\mathbf{X}_1, \mathbf{X}_1) & \dots & R(\mathbf{X}_1, \mathbf{X}_{n_s}) \\ \vdots & \ddots & \vdots \\ R(\mathbf{X}_{n_s}, \mathbf{X}_1) & \dots & R(\mathbf{X}_{n_s}, \mathbf{X}_{n_s}) \end{bmatrix} \tag{5.9}$$

This matrix is positive semi-definite and has ones on the diagonal.
Finally, vector $\mathbf{r}$ contains the correlation function evaluated between each observation point and a new unknown point which we want to predict:

$$\mathbf{r}(\mathbf{X}_p) = \begin{bmatrix} R(\mathbf{X}_p, \mathbf{X}_1) \\ \vdots \\ R(\mathbf{X}_p, \mathbf{X}_{n_s}) \end{bmatrix} \tag{5.10}$$

By solving the constrained minimization problem (see equations 5.6 and 5.7), the BLUP is obtained as:

$$\hat{Y}\mathbf{X}_p = \mathbf{f}^T(\mathbf{X}_p)\hat{\beta} + \mathbf{r}^T(\mathbf{X}_p)\mathbf{R}_X^{-1}\left(\mathbf{Y}_K - \mathbf{F}_X\hat{\beta}\right) \tag{5.11}$$

with:

$$\hat{\beta} = \left(\mathbf{F}_X^T\mathbf{R}_X^{-1}\mathbf{F}_X\right)^{-1}\mathbf{F}_X^T\mathbf{R}_X^{-1}\mathbf{Y}_K \tag{5.12}$$

Finally, the mean square error of the estimate is:

$$\mathbb{MSE}\left[\hat{Y}(\mathbf{X}_p)\right] = \sigma_Z^2\left(1 - \begin{bmatrix}\mathbf{f}^T(\mathbf{X}_p) & \mathbf{r}^T(\mathbf{X}_p)\end{bmatrix}\begin{bmatrix} 0 & \mathbf{F}_X^T \\ \mathbf{F}_X & \mathbf{R}_X \end{bmatrix}\begin{bmatrix}\mathbf{f}(\mathbf{X}_p) \\ \mathbf{r}(\mathbf{X}_p)\end{bmatrix}\right) \tag{5.13}$$

Note that $\hat{Y}(\mathbf{X}_p)$ is also known as the Kriging prediction mean which represents the prediction's nominal value, while the mean square error of the estimate is also known as the Kriging variance (square of the Kriging standard deviation which we denote σ_K) which represents an estimation of the error on each Kriging prediction.

5.1.2 Steps of Kriging

The process for constructing a Kriging model can be divided into the following steps:

1. Gather observations with inputs $\mathbf{X}_K$ and outputs $\mathbf{Y}_K$
2. Select the set of regression functions which define the process mean over the domain
3. Select the correlation function used for the random process portion of the model
4. Optimize the correlation function hyperparameters through optimization of the Maximum Likelihood Estimate (MLE) [212] and deduce the regression coefficients
5. Construct the Kriging model based on the minimization of the mean square error of the prediction and the unbiasedness constraint, which yields the expression for the BLUP
6. Assess the capability of the Kriging model to predict the original model using validation

5.2 Noisy Kriging

Noisy Kriging is a variant of Kriging which was designed for the case of non-exact observations, which can arise for instance when there are measurement errors, or when the observations rely on non-converged numerical models. A comprehensive description of noisy Kriging can be found in [184]. For noisy Kriging, the observations no longer conform to equation 5.4 because the outputs $\tilde{\mathbf{Y}}_K$ contain heterogeneous noise components $\{\epsilon_1, ..., \epsilon_{n_s}\}$:

$$\begin{aligned}\mathbf{X}_K &= \{\mathbf{X}_1, ..., \mathbf{X}_{n_s}\} \\ \tilde{\mathbf{Y}}_K &= \{Y_1 + \epsilon_1, ..., Y_{n_s} + \epsilon_{n_s}\}\end{aligned} \tag{5.14}$$

Thus, in the noisy case, the equations are slightly modified compared to the description in section 5.1.1 by adding the diagonal matrix $\mathbf{\Delta}$ of noise components $\{\epsilon_1, ..., \epsilon_{n_s}\}$ to the correlation matrix $\mathbf{R}_X$. The noisy Kriging model does not interpolate the observations and its variance is non-zero at these points. The idea is then to use the ability of noisy Kriging to take into account heterogeneous noise in its observations in order to model the variability of the studied function induced by uncertainties. This is detailed in the next sections.

5.3 Bayesian Optimization background

In the context of Kriging-based BO, the idea is to sequentially update a surrogate model with new function observations in order to get high quality predictions in regions of the design space which are interesting in terms of the optimization objectives, while regions which show low potential are minimally enriched. The main goal of such a strategy is to approach the global minimum faster and to minimize calculation of unnecessary costly observations. The enrichment process is usually carried out iteratively until some stopping criterion is met. One challenge of this process resides in the smart selection of the new observation points used to update the surrogate model, the goal being to simultaneously exploit parts of the design space where the optimization objective is of good quality (the lowest in a minimization problem) and sample at locations of the design space where uncertainty is high which has the advantage of exploring new potentially promising areas. The methods used to find new sampling points are called acquisition functions, or infill sampling criteria. Proposed in 1978 [22] and popularized by Jones [10], Expected Improvement is the most prevalent acquisition function in Kriging-based BO [23, 24, 25, 26]. EI uses the expected value of the improvement which contains information about the quality of the mean value of the Kriging model as well as the associated uncertainty at the considered point.

5.3.1 Theoretical background of EI

In a minimization problem, solved through BO, $Y_{min} = min\,(Y_1, ..., Y_{n_s})$ represents the best function value among all observations at the current iteration. The "improvement" $I_{EI}(\mathbf{X}_p)$ at unsampled point $\mathbf{X}_p$ is defined as:

$$I_{EI}(\mathbf{X}_p) = max(Y_{min} - \hat{Y}(\mathbf{X}_p), 0) \tag{5.15}$$

Expected improvement is then just the expected value operator applied to the improvement:

$$\mathbb{E}\left[I_{EI}(\mathbf{X}_p)\right] = \mathbb{E}\left[max(Y_{min} - \hat{Y}(\mathbf{X}_p), 0)\right] \tag{5.16}$$

which can be expressed in an integral form and then simplified to a closed form using integration by parts:

$$\mathbb{E}\left[I_{EI}(\mathbf{X}_p)\right] = \left(Y_{min} - \hat{Y}(\mathbf{X}_p)\right) f_{pdf}\left(\frac{Y_{min} - \hat{y}}{\sigma_K}\right) + \sigma_K f_{cdf}\left(\frac{Y_{min} - \hat{y}}{\sigma_K}\right) \tag{5.17}$$

where f_{pdf} and f_{cdf} are respectively the PDF and the cumulative distribution function (CDF) of the normal distribution $\mathscr{N}(0,1)$:

$$\begin{aligned} f_{pdf}(\zeta) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta^2}{2}} \\ f_{cdf}(\zeta) &= \frac{1}{2}\left[erf(\frac{\zeta}{\sqrt{2}}) + 1\right] \end{aligned} \tag{5.18}$$

In order to determine the next infill point at a certain iteration, the expected improvement is then maximized over the design space. The closed form of EI is practical to optimize since it is cheap-to-evaluate and its derivatives can be determined.

5.4 Noisy Kriging based robust optimization strategy

The overall objective of this research is to design a robust shape optimization methodology under uncertain input parameters by taking into account robustness as an optimization objective. This method proposes to take advantage of the properties of noisy Kriging to retrieve the prediction of the performance function as well as a measure of its robustness to uncertainties using a single Kriging model and a suitable enrichment strategy. The general concept of BO is to construct a metamodel of the quantity to be optimized and iteratively add observations in order to increase prediction accuracy in regions of the design space which are most promising in terms of the optimization objectives, and progressively approach the global optimum while mitigating the number of necessary observations. In our case, we take the idea of BO and adapt it to the robust optimization case with the noisy Kriging metamodel.

5.4.1 Noise computation

We recall that the considered uncertain shape parameters follow a Gaussian distribution around an average value $\overline{X}_i^u$ and that all uncertain parameter values lie in the interval $\overline{X}_i^u \pm 3\sigma_u$ (see section 3.1.2). In this way, the variability of the output function to such uncertainties can be quantified using the variance of the output corresponding to a Gaussian distribution of each input parameter with average value the nominal value of said parameter. To approximate this variance in a computationally reasonable way, we reuse the sparse PCE method used in chapter 4, with 100 observations per potential solution. The obtained variance is then directly used as observation noise ϵ_i associated to each deterministic observation in the noisy Kriging formulation.

5.4.2 Enrichment strategy

Using only classical acquisition functions like EI may be ineffective because the main goal is to retrieve a measure of robustness from the noisy Kriging model, in which case the enrichment process must also target robust regions of the design space, where observation noise would be the lowest. Thus, enrichment of the noisy Kriging model is carried out along two components:

1. Maximization of EI for exploitation and exploration in terms of Kriging prediction mean
2. Maximization of robustness through minimization of Kriging prediction standard deviation

The two components are then combined by solving a two-objective optimization problem with NSGAII [208], and selecting three relevant points on the obtained Pareto front as new infill points, in order to have two points which respectively favor each objective, and one point which represents a compromise. A scheme of the proposed enrichment and optimization process is shown in figure 5.1. An optimization constraint is also added in order to prevent new infill points from being too close to existing observations, which may otherwise happen in some cases because of the second objective (minimization of the standard deviation).

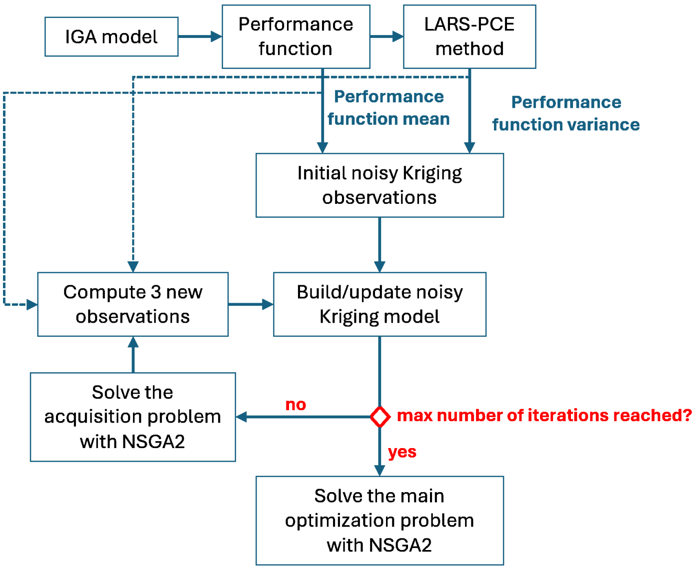


Figure 5.1: Basic diagram of the proposed enrichment and optimization process

The robustness criterion is then taken directly as the enriched noisy Kriging prediction standard deviation. In non-enriched regions of the design space this measure of robustness is very inaccurate since in these regions the portion of the standard deviation due only to the Kriging approximation (not to the noise) is predominant. However, since enrichment targets highly robust regions, the accuracy of the non-enriched regions is not important, given that the presence of uncertainty due to the Kriging approximation ensures that any inaccurate prediction of robustness should most of the time underestimate robustness. In other words, the proposed robustness should mostly be an approximation by excess of the noise level, and the idea is to minimize the error of this approximation in regions of the design space which show good potential for robustness.

5.5 Implementation and results

In this section the noisy Kriging strategy is applied to the simplified disc brake model in the case of only one deterministic input parameter (r_3). This allows to test the proposed strategy on a simple case and easily observe the behavior of the noisy Kriging model which becomes more difficult in higher input parameter dimension. The geometrical parameters considered uncertain are the same as in chapter 4. Note that for implementation of noisy Kriging, we used the Python machine learning library scikit-learn [213].

5.5.1 Comparison between Kriging and noisy Kriging

First we carried out a comparison of regular Kriging (non-noisy) and noisy Kriging models, using 10 stability function observations obtained through LHS, in the case of a single input shape parameter. The comparison is displayed in figure 5.2, with the top graph showing the regular Kriging while the bottom graph shows the noisy Kriging model. In both graphs, the red points represent the stability function observations used to construct the Kriging model. However, the observations in the noisy case come with error bars which display the observation noise under the form of a standard deviation, used to model the variability of the stability function due to geometric uncertainties. The dark blue line represents the Kriging model prediction mean, while the light blue area represents the Kriging model standard deviation which in simple terms is an estimation of the error committed on the Kriging prediction mean.

Two main differences are observed between both types of Kriging models. Firstly, the Kriging model standard deviation is much larger (wider confidence interval) in the case of noisy Kriging. This is due to the presence of noise, in which case the Kriging standard deviation not only contains a contribution due to lack of observations, but also a contribution due to the presence of noise, and both contributions are blended into one mathematical quantity. The purpose of the proposed enrichment strategy is to add observations in promising regions of the design space in order to reduce the Kriging standard deviation to only the noise-induced contribution, and thus use it as an estimate of variability due to uncertainties. The second main difference between the two graphs is that in the noisy case the Kriging prediction mean is also considerably affected compared to the non-noisy case, and it does not interpolate the observations.

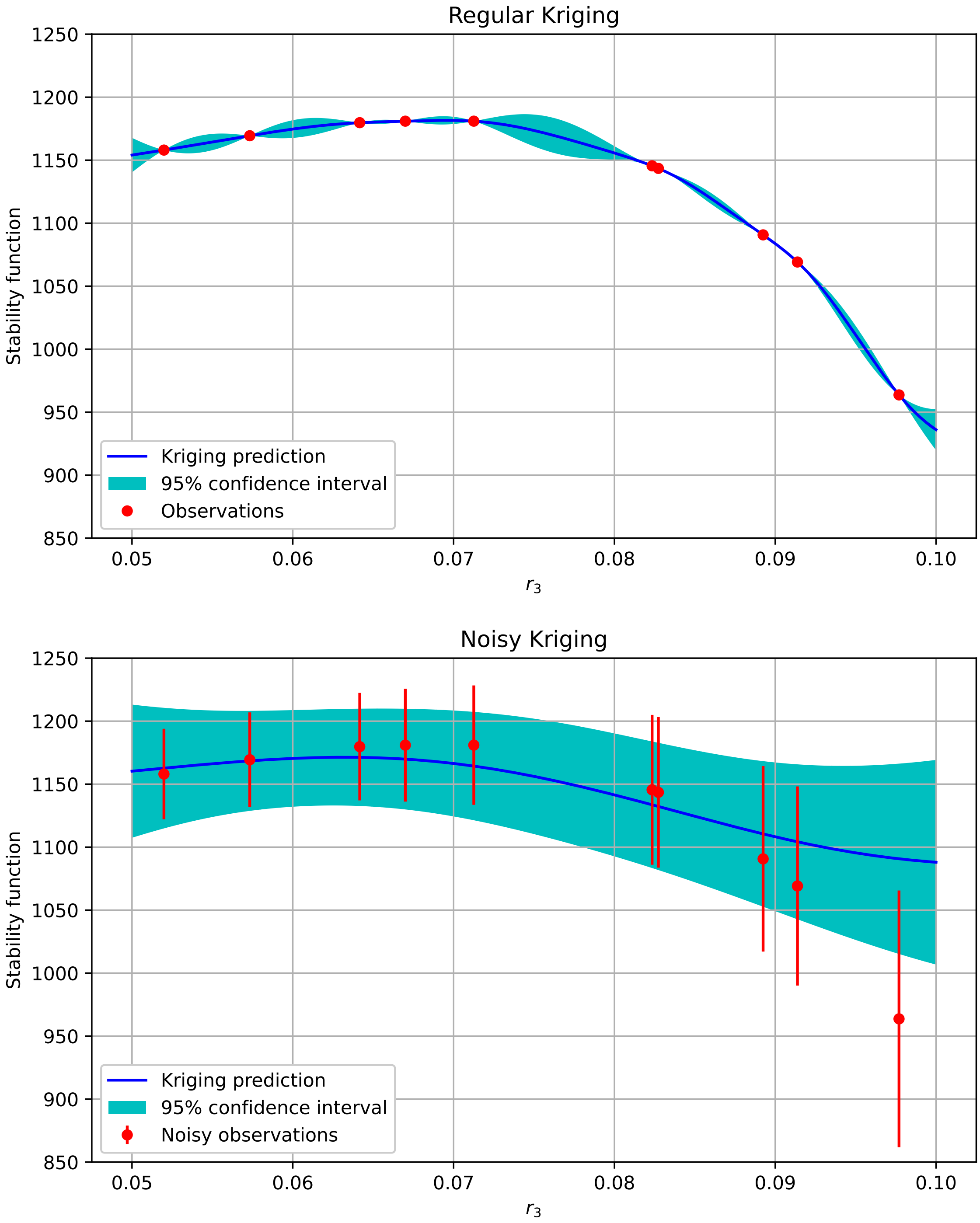


Figure 5.2: Comparison of regular (top) and noisy (bottom) Kriging models of the disc brake stability function for one input parameter, constructed with 10 observations

5.5.2 Results of the enrichment strategy

In this section we compare two noisy Kriging models, one constructed with 40 noisy observations determined by LHS, and the other based on 10 initial LHS observations to which are added 30 observations obtained through 10 iterations of the proposed enrichment strategy (3 observations per iteration) described in section 5.4.2. Both models are displayed in figure 5.3. The aim of this comparison is to observe whether the enrichment strategy is effective in producing a noisy Kriging model which is adapted for the purpose of subsequently performing robust optimization.

Compared to the LHS-based model, the enriched model's graphical representation is very different and can be clearly divided into three parts exhibiting different characteristics and behavior, with notably two intervals specifically targeted for enrichment:

- $0.05 < r_3 < 0.065$: this region is relatively highly enriched with 13 out of 30 enrichment observations included, most likely due to the fact that it exhibits low Kriging variance overall, which is one of the objectives of the enrichment strategy.
- $0.065 < r_3 < 0.084$: this region contains no enrichment samples, which makes sense because it does not have any particular potential in terms of stability function minimization (included in EI maximization) or in terms of low Kriging standard deviation. Even though it is an unexplored region, the exploration component of EI did not lead to any observations being added to it.
- $0.084 < r_3 < 0.1$: this region is highly enriched (17 out of 30 enrichment observations), most likely due to the fact that it shows potential in terms of stability function minimization. The increase in observation noise as r_3 increases appears to translate to an increase in the width of the confidence interval (light blue area) which is promising as it suggests that adding noisy observations tends to make the surrounding Kriging standard deviation proportional to the actual variance due to uncertainties, in the enriched region. This region is also characterized by a considerable gap between noisy Kriging prediction mean and the observations' nominal value for $0.095 < r_3 < 0.1$.

Moreover, several more general observations can be made. Firstly, there is a clear difference in the width of the confidence interval between regions $0.05 < r_3 < 0.065$ and $0.084 < r_3 < 0.1$, which appears somewhat proportional to the difference in observation noise between these regions. This suggests that the Kriging standard deviation could potentially be used as an approximation of variability to uncertainties when surrounding enrichment is sufficient. Secondly, the fact that region $0.065 < r_3 < 0.084$ is not enriched while the two other regions are highly enriched means that when applying robust optimization with an enriched noisy Kriging model, optimal solutions which strongly favor one of the optimization objectives will be advantaged, while solutions which are very balanced in terms of optimization objectives will most likely not be found. Thirdly, the extremities of a noisy Kriging model, even when enriched, may be affected by typical

Kriging extrapolation behavior at the extremities where the confidence interval typically becomes wider (e.g. $r_3 < 0.053$). Also note that highly enriched regions are characterized by relatively evenly spaced observations which is due to the constraint on the spacing between adjacent observations applied during the enrichment process (see section 5.4.2). Finally, building the enriched noisy Kriging model with a total of 40 noisy observations required roughly 30 hours, using the same computational machine as in chapter 4 (32-core 2.895 GHz processor, 1 TB of RAM, 12 workers on Matlab's parallel pool). The computational cost required to build the enriched noisy Kriging was only marginally higher than for the LHS-based model, because the enrichment process uses objective functions which are based on the existing surrogate model and is thus cheap to evaluate. For the same reason, although the full robust optimization process was not carried out following the construction of the noisy Kriging model, it is clear that this step would be very computationally efficient.

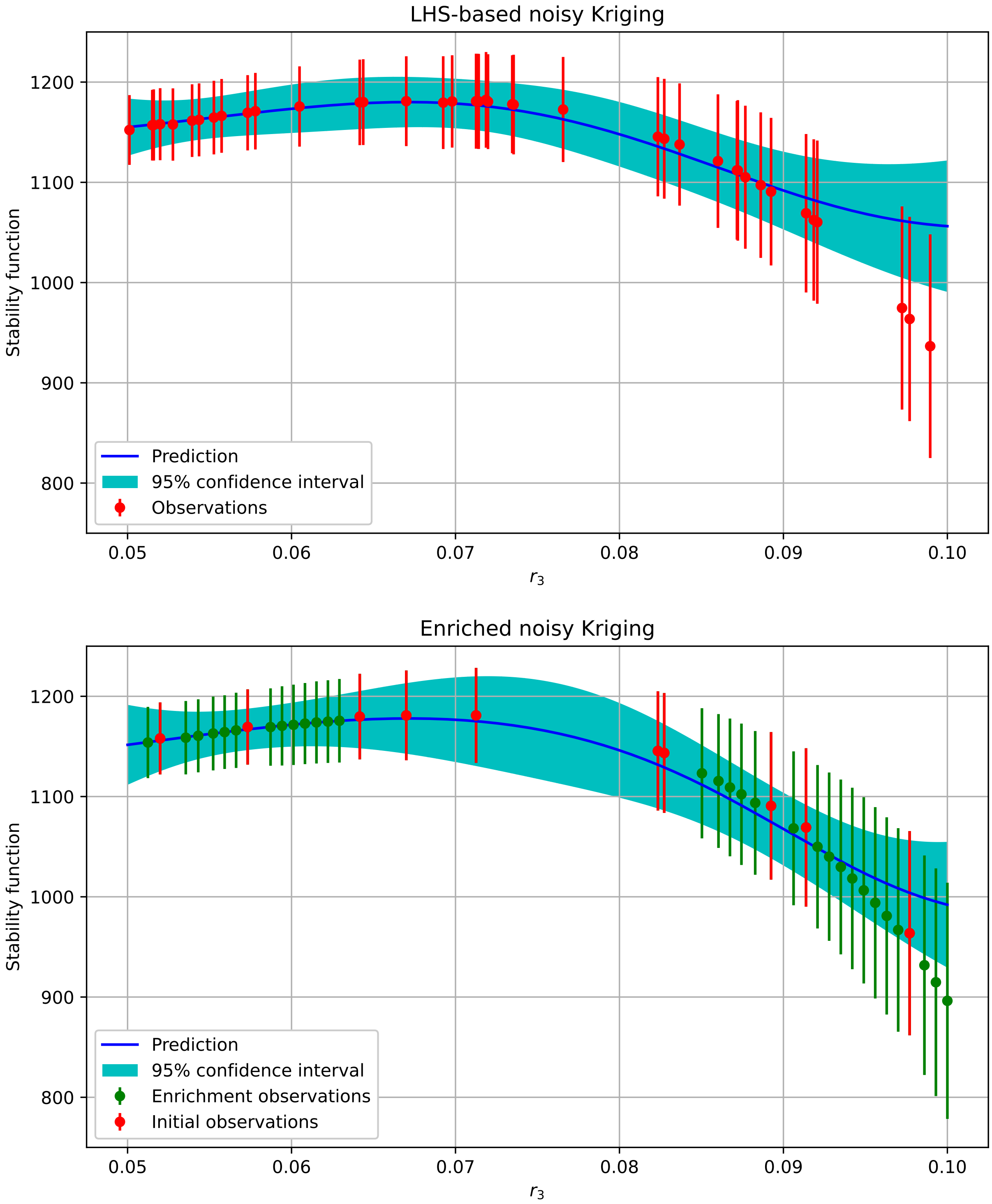


Figure 5.3: Comparison of a noisy Kriging model constructed with 40 LHS observations (top) and one constructed with 10 initial LHS observations and 30 observations determined using the proposed enrichment strategy (bottom), of the disc brake stability function for one input parameter

5.6 Kriging limitations in high dimensions

This section investigates the limitations of the Kriging method applied to the simplified disc brake system in higher dimension of input parameters.

5.6.1 Kriging validation

Although the enriched noisy Kriging strategy shows potential for robust shape optimization of the simplified disc brake model, it shows limitations in higher dimensions, i.e., when increasing the number of deterministic input parameters. These limitations, which have prevented us from applying the noisy Kriging strategy and full robust optimization to the full 12 shape parameter case, are described in this section. Validation of non-noisy Kriging models in the case of 1, 2, 6, and 12 input shape parameters was carried out for an increasing number of stability function observations (LHS). The rule of thumb for constructing Kriging models is to take a number of function observations equal to ten times the number of input parameters [214]. For the simplified disc brake stability function, we purposely overestimated this number and constructed Kriging models based on a number of observations equal to 10, 100, and 1000 times the number of input parameters. 1000 stability function observations were computed for validation, and their predictions by the constructed Kriging models were plotted against the true values. Figure 5.4 compiles the validation results. The closer each point is to the identity line (abscissa=ordinate line), the better the accuracy of the Kriging model. We first notice that as the number parameters is increased, the validation points are more and more spread-out around the identity line, which clearly shows degradation of the Kriging model's prediction accuracy as the number of input parameters is increased. Secondly, in the cases of 6 and 12 input parameters, even though the number of observations is drastically increased, the improvement in prediction accuracy is not satisfactory in a context where quantification of robustness to uncertainties is a major challenge. Overall, the obtained validation results suggest that Kriging-based strategies are not scalable to higher input dimension in the case of the studied simplified disc brake model stability function, most likely due to the complexity and nonlinearity of the underlying physical phenomenon, coupled with the curse of dimensionality, a well-known effect where computational cost required for building surrogate models or for processing data in general, increases exponentially when the input dimensions are increased [215].

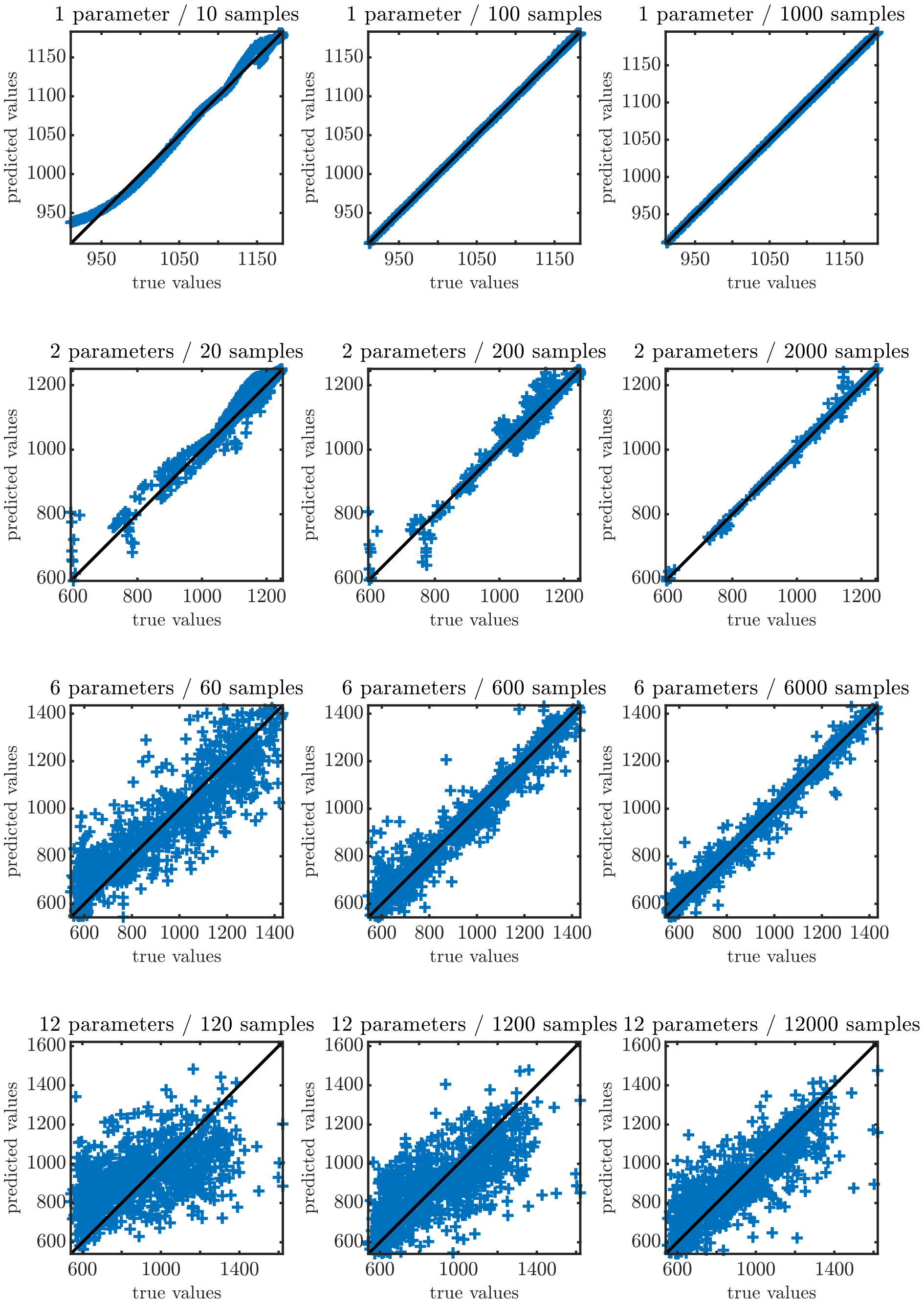


Figure 5.4: Kriging prediction values versus true values of the disc brake stability function, for an increasing number of observations, and for an increasing number of input parameters (1, 2, 6, 12)

5.6.2 Testing a deep learning tool for comparison

This section presents an attempt to apply an alternate surrogate modeling tool to the studied simplified disc brake model. The main goal is to see if a different surrogate modeling technique yields better prediction accuracy than Kriging in high dimensions. The considered surrogate modeling tool is Deep Neural Networks (DNNs), which have gained great popularity and have shown to be a powerful tool in the field of artificial intelligence in recent years. This section is not an in-depth study of DNNs applied to our disc brake model, but rather an opening to the Kriging-based strategy in the aim of providing insight on the possibility of creating models which accurately predict the studied complex disc brake stability function, using modern numerical tools. As such, this section does not include any in-depth mathematical background on DNNs. First, we provide general background and important concepts for DNNs, followed by a description of the main parameters and settings which we focused on in order to tune DNNs to our applicative function, and finally we expose the obtained prediction validation results.

Deep neural networks

A Neural Network (NN) (also known as Artificial Neural Network) is a type of model used in machine learning inspired by biological neural networks in the brain. Neurons in the brain are modeled by “nodes” while synapses which connect the neurons are modeled by “edges”. The nodes are organized into layers, with the first layer being the “input layer” and the last layer being the “output layer”. Between the input layer and the output layer, there are intermediate layers of nodes called “hidden layers”. The signal which travels between neurons is a real number which passes between nodes through the edges. As the signal passes through different layers, the corresponding real number is transformed by a nonlinear function, called “activation function” which determines the output of a node from the sum of its inputs. “Weights” at each connection are used to adjust the strength of the signal between two nodes. The type of layer of a node also affects the type of calculation performed on the signals which pass through it. A Deep Neural Network is a NN which has more than one hidden layer. Figure 5.5 shows the basic structure and elements of a NN. There are many types of NNs suited to different problems and applications [216]. In our case, we study what is known as a “regression” problem in artificial intelligence, where the goal is to build a model able to predict continuous real-valued outputs. For the studied disc brake application, we use feedforward NNs where information only travels in one direction, from the input node through the hidden layer nodes, to the output nodes. Since the stability function only has one output, we consider NNs with only one output node.

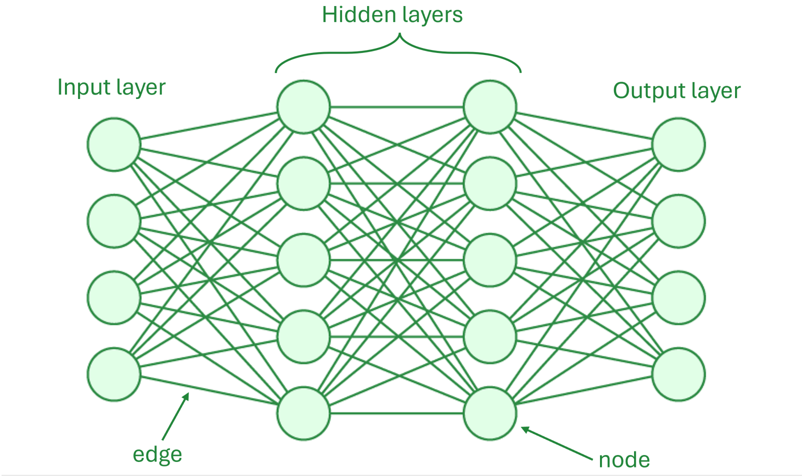


Figure 5.5: Basic structure of a Neural Network

Important parameters and settings

For the application of DNNs to the studied disc brake stability function, we have chosen to use the Python tool Tensorflow [217] with Keras [218], which is the associated high-level API for building NNs. With this tool, there are several different settings and parameters that can be modified to tune the NN to the studied application. We chose to build a fully connected model, where each node from the previous layer is connected to each node of the current layer which allows this type of model to capture complex patterns in training data [219]. With fully connected networks, no assumptions need to be made on the structure of the inputs, which makes them more broadly applicable to diverse problems. On the other hand, they may yield lower accuracy than networks specially selected and calibrated to the structure of a certain problem's inputs. Moreover, the main settings of the constructed NNs which were tuned for application to our disc brake stability function are listed and described below:

- The activation function determines whether or not a node is activated by computing the weighted sum of its inputs and adding bias to it. As stated earlier, the role of the activation function is to introduce nonlinearity into the output of a node. The activation functions available in Keras are listed and described in [220]. For application to the disc brake stability function, the activation function is determined by trial-and-error for each case in terms of number of input parameters and number of observations.

- The loss function measures how well the NN models the training data during the training process by comparing true and predicted output values. We have chosen to use the mean squared error loss function which computes the average of the squared differences between true and predicted outputs.

- The loss function optimizer determines how the loss function is minimized throughout the training process by adjusting the weights at the connections between nodes. There are two main types of optimizers: gradient-based and so called "adaptive" optimizers. The optimizers available in Keras are described in [221]. In our case, the loss function optimizer is determined by trial-and-error.

- The number of hidden layers determines the depth of the NN. The optimal number of hidden layers is a subject of debate. It is application-dependent, and most of the time determined by trial-and-error, although recently more sophisticated methods have been proposed [222]. For simplicity, in our applicative case, we determine the number of hidden layers by trial-and-error.

- The number of "epochs" is the number of training iterations carried out through the training data, with every sample from the training data being used once per epoch.

- Early stopping is a process through which the training process is stopped after a certain number of epochs when the model stops improving, in order to prevent overfitting. In addition, early stopping can help mitigate model training cost.

Application to the simplified disc brake model

This section presents the results of applying NNs for prediction of the disc brake stability function. NNs were built for each of the same cases as in the Kriging validation section 5.6.1, i.e., for 1, 2, 6, and 12 input shape parameters and for an increasing number of function observations (10, 100, 1000 times the number of input parameters). For each case, the NN settings mentioned in the previous section (5.6.2) (activation function, number of hidden layers, loss function optimizer, number of epochs, early stopping) were selected by trial-and-error. The settings used for each case are summarized in table 5.1. Figure 5.6 shows the validation results using the same validation set as for the Kriging validation. We observe the same general trends as in the case of the Kriging model, namely that as the number of input parameters increases the prediction accuracy deteriorates substantially, and that as the number of observations in increased the prediction accuracy improves. Overall, the constructed DNN models do not perform substantially better than the Kriging models. A slight improvement in performance is noticeable in the case of 6 and 12 input parameters and for a high number of observations (respectively 6000 and 12000), however, prediction accuracy is not nearly sufficient to contemplate application to robust optimization in high dimensions of input parameters. Moreover, the trial-and-error process used to determine the NN settings gives some elementary insight into which settings are adapted to the studied disc brake application. Activation functions "gelu", "swish", "selu" and "elu" yield better prediction accuracy among all the activation functions available in Keras. Loss function optimizers "Adam" and "RMSprop" show the best performance with "Adam" being slightly more efficient in most cases. In general, an increase in the number of hidden layers provides better prediction accuracy as the number of input parameters is increased, to the exception of the 12 parameter case, where most of the NN settings have only modest impact on the validation results, which incidentally upholds the idea that building surrogate models of the disc brake stability function is more and more difficult as the number of input parameters increases.

Table 5.1: List of parameters used for building NNs applied to the disc brake stability function, for different numbers of input parameters and observations

Number of input parameters	Number of function observations	Activation function	Number of hidden layers	Loss function optimizer	Number of epochs	Early stopping
1	10	“gelu”	4	“RMSprop”	500	no
	100	“gelu”	3	“Adam”	500	no
	1000	“gelu”	2	“Adam”	500	no
2	20	“gelu”	5	“Adam”	500	no
	200	“gelu”	4	“Adam”	500	no
	2000	“swish”	5	“Adam”	500	no
6	60	“selu”	8	“Adam”	500	no
	600	“swish”	7	“Adam”	1000	yes
	6000	“swish”	5	“Adam”	500	no
12	120	“elu”	2	“Adam”	500	no
	1200	“selu”	5	“Adam”	1000	yes
	12000	“gelu”	4	“Adam”	1000	yes

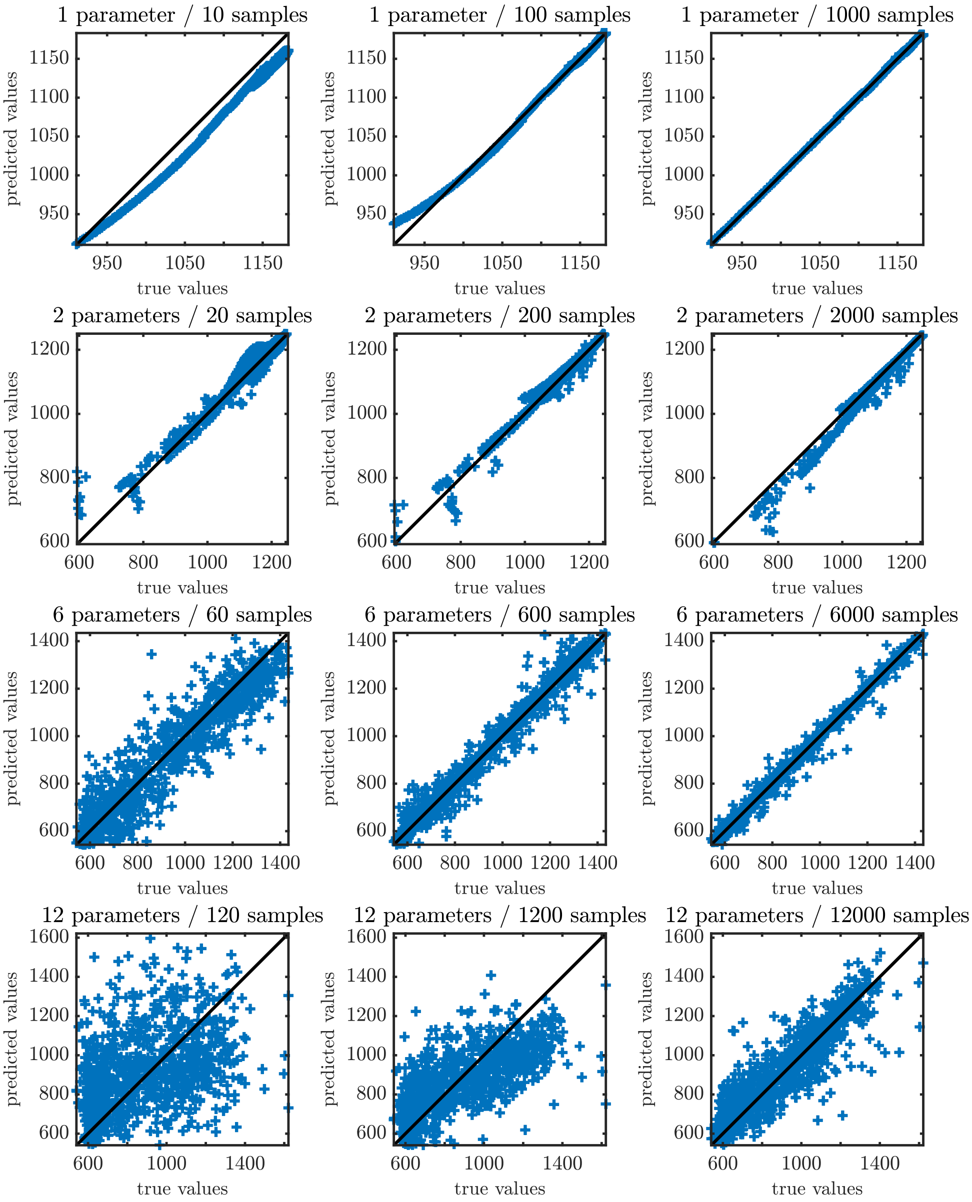


Figure 5.6: Deep Neural Network prediction values versus true values of the disc brake stability function, for an increasing number of observations, and for an increasing number of input parameters (1, 2, 6, 12)

5.7 Important takeaways from the Kriging-based strategy

The proposed noisy Kriging-based approach shows potential in the particularly difficult case tackled in this PhD research where the blackbox stability function which quantifies squeal noise propensity provides no gradient, is computationally expensive, and proves troublesome to approximate using surrogate modeling. As shown in this chapter, the proposed method succeeds in enriching regions of the noisy Kriging model which show potential in terms of high robustness and stability function minimization. Furthermore, in the highly enriched regions, the Kriging standard deviation somewhat follows the trend in observation noise variations, which suggests that the Kriging standard deviation in these regions may be used as a coarse approximation of the variability to uncertainties. Nonetheless, the proposed approach may lead to optimal solutions which represent only a subset of the "true" full Pareto front in some cases, because regions of the design space which balance the optimization objectives equally may not be enriched. Also, the approach is not scalable to higher dimensions in the case of the disc brake squeal noise minimization problem, because of the degradation in prediction accuracy of the Kriging method as the number of input parameters grows, even with a very high number of observations. Attempting to metamodel the disc brake stability function with DNNs mostly confirmed that the inability to construct accurate metamodels for this application in high dimensions is due to the stability function's complex behavior. Incidentally, this attempt also provides initial insight into tuning NNs for this type of application.

The results obtained with the noisy Kriging based strategy introduce interesting outlooks for further research in this direction. A two-dimensional study of the proposed method, which would provide a larger design space and more possibilities in terms of enrichment diversity, may give more insight into the efficiency of the proposed method, even though visualization and analysis of the Kriging approach become more challenging when more than one input parameters are considered. The presented enrichment procedure may also potentially benefit from some adjustments. For instance, investigating the effect of the initial observation set, or testing different enrichment objectives, could provide insight into increasing efficiency. Finally, the main concept of the noisy Kriging method being non-specific to the studied application, it may be worthwhile to test it on different applicative systems.

Overall, the undertaken robust optimization problem must balance minimization of the main quantity of interest and maximization of the quantity of interest's robustness to uncertainties. With this in mind, and in the present-day context of production of mechanical systems requiring increasing precision and speed, the problem from a broader perspective becomes a compromise between computational cost and accuracy, which are highly antagonistic. Methods such as noisy Kriging, which do not necessarily lead to satisfactory results as could be obtained with extremely expensive Monte-Carlo-type simulations, must nevertheless be considered.

Conclusion

The presented PhD research has successfully proposed two new methodologies which combine modern numerical tools for robust shape optimization of mechanical systems under gradient-free, nonlinear, and costly performance functions, applied to the case of the disc brake squeal noise minimization problem. An efficient modeling strategy for the simplified disc brake system, which takes advantage of the benefits of IGA for shape optimization, combined with straightforward yet effective modeling of friction-induced instabilities, provides the main figure of merit which constitutes the basis of the robust shape optimization problem and of the two resolution strategies. Both strategies were designed with uncertainty quantification as a central aspect, as it poses the biggest challenges in terms of computational cost and accuracy.

The idea of the first strategy was to capitalize on the built-in output variance estimation of the PCE method, in order to obtain a relatively cheap estimation of output variability to geometric uncertainties. The search for optimal solutions which balance stability function minimization and robustness to uncertainties was then carried out by building local PCE models of the disc brake stability function for uncertainty quantification inside a genetic optimization loop. The superiority of sparse PCE compared to non-sparse PCE in terms of output variance estimation was as expected, although analysis of the polynomial coefficient distributions revealed an unexpected result with the sparse method containing a larger number of coefficients. Implementation showed that although using a relatively restricted number of function samples for PCE model construction, very reasonable estimations of output variance to uncertainties were provided. The number of function samples appears as the key parameter to balance computational cost and accuracy of the obtained solutions. Nonetheless, in an industrial design process, using the proposed method paired with Monte-Carlo validation of the optimal solutions is recommended. Moreover, use of IGA to generate potential solution shapes was extremely smooth in terms of implementation and combination with numerical analysis, even though this does not appear explicitly in the results, and it provided a great variety of original shapes, as expected. Lastly, analysis of the obtained optimal solutions was coherent with expectations in terms of physical behavior.

The underlying concept for the second main strategy was to apply surrogate modeling over the whole design space in order to reduce overall computational cost, while exploiting Kriging's built-in error estimation metric and the ability of its noisy counterpart to take into account measurement error. Combining noisy Kriging with a Bayesian Optimization-inspired enrichment strategy would allow to pinpoint regions of the design space with high potential in terms of robustness to uncertainties and performance.

Implementation on the disc brake stability function in a simple one-dimensional case revealed that the proposed strategy behaves satisfactorily, with targeted regions of the design space being enriched and resulting in a potentially exploitable estimation of robustness to uncertainties. On the other hand, the inability to construct accurate Kriging models in higher dimensions of input parameters was a surprising result which suggests very high complexity of the studied application's behavior. This was confirmed by unsuccessfully testing a different, Neural Network-based surrogate modeling strategy on the same application system.

The proposed research and obtained results bring forward several prospects for future development. Regarding the PCE-based strategy, attempting to apply the method with a higher number of uncertain parameters is the next step in terms of robust optimization problem complexity. Moreover, tuning the number of samples for construction of each PCE model is a promising aspect since it could potentially allow further reduction in computational cost. Considering a tolerance on the a-posteriori error (LOO error for instance) is the first possibility which comes to mind. Also, an in-depth study of the difference between sparse and non-sparse PCE when applied to the disc brake application would allow to understand why the LARS algorithm yielded more polynomial coefficients than the non-sparse variant, and testing other sparse methods for the calculation of the coefficients could provide some interesting insight. In addition, although in this PhD research we have only used non-intrusive Polynomial Chaos methods, the application of intrusive methods (see section 1.3.4) could be potentially beneficial in terms of uncertainty quantification accuracy and computational effectiveness. With respect to the noisy Kriging-based method, there is considerable room for improvement. First, testing the method with two input parameters or in general on a broader region of the design space where there are more stability function variations would provide better understanding of the enrichment strategy's efficiency. Another possibility consists in verifying whether the enriched noisy Kriging's standard deviation can be used as a reliable indicator of robustness to uncertainties. If these aspects prove successful, the next obvious step is considering resolution of the robust optimization problem based on an enriched noisy Kriging model. More generally, the inability to build accurate surrogate models of the considered stability function remains an open problem which deserves more investigation. Finally, applying and adapting the two proposed strategies to different application systems is a major prospect, in order to further study their effectiveness in the field of engineering in a broader sense.

The proposed methods constitute two new and original ways of conducting shape optimization in the presence of uncertainties and in the difficult case of gradient-free, nonlinear, and computationally costly performance metrics. Contrarily to the so far predominant RBDO approach which is centered around determining probability of failure, the presented methods adopt the less common robust optimization scheme thus providing Pareto optimal solutions which balance robustness and performance. With respect to robust optimization, where the challenge resides in accurately estimating the output variability to uncertainties, each proposed method introduces a new way of dealing with this issue by leveraging the specific properties of sophisticated surrogate modeling techniques for uncertainty quantification. The advantageous combination of IGA for shape manipulation and mechanical analysis, surrogate modeling for uncertainty quantification and computational cost reduction, and a genetic algorithm for gradient-free exploration of optimal solutions, brings a novel contribution to the field of design optimization for industrial systems in the presence of uncertainties. Regarding the studied applicative system, the presented research constitutes the first full robust shape optimization process applied to the complex friction-induced instability phenomenon responsible for undesirable squeal noise in industrial disc brake systems. Overall, this research project confirms the idea that taking into account uncertainties in the optimization of mechanical systems is a tremendous challenge in terms of added computational cost. Efforts made to minimize this cost inevitably lead to deterioration in accuracy of the proposed solutions which causes the problem to shift towards finding the best possible compromise between cost and accuracy given the industrial context and constraints of the studied application. Only radical improvements in computational power can bring ground breaking change to the paradigm of robust design optimization of industrial systems.

Bibliography

[1] J. C. Bansal, P. Bajpai, A. Rawat, and A. K. Nagar. Sine cosine algorithm for multi-objective optimization. In *Sine Cosine Algorithm for Optimization*, pages 35–63. Springer, 2023.

[2] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. ieee, 1995.

[3] X.-S. Yang and A. Hossein Gandomi. Bat algorithm: a novel approach for global engineering optimization. *Engineering computations*, 29(5):464–483, 2012.

[4] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

[5] D. Karaboga. Artificial bee colony algorithm. *scholarpedia*, 5(3):6915, 2010.

[6] D. Bertsimas and J. Tsitsiklis. Simulated annealing. *Statistical science*, 8(1):10–15, 1993.

[7] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah. Metaheuristic algorithms: A comprehensive review. *Computational intelligence for multimedia big data on the cloud with engineering applications*, pages 185–231, 2018.

[8] J. Močkus. On bayesian methods for seeking the extremum. In *Optimization techniques IFIP technical conference*, pages 400–404. Springer, 1975.

[9] J. Mockus, V. Tiesis, and A. Zilinskas. The application of bayesian methods for seeking the extremum. *Towards global optimization*, 2(117-129):2, 1978.

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

[11] N. Cressie. Spatial prediction and ordinary kriging. *Mathematical geology*, 20:405–421, 1988.

[12] N. Cressie. The origins of kriging. *Mathematical geology*, 22:239–252, 1990.

[13] J. P. Kleijnen. Kriging metamodeling in simulation: A review. *European journal of operational research*, 192(3):707–716, 2009.

[14] J. Quinonero-Candela and C. E. Rasmussen. A unifying view of sparse approximate gaussian process regression. *The Journal of Machine Learning Research*, 6:1939–1959, 2005.

[15] S. Hosder, L. T. Watson, B. Grossman, W. H. Mason, H. Kim, R. T. Haftka, and S. E. Cox. Polynomial response surface approximations for the multidisciplinary design optimization of a high speed civil transport. *Optimization and Engineering*, 2:431–452, 2001.

[16] E. Immonen. Shape optimization of annular s-ducts by cfd and high-order polynomial response surfaces. *Engineering Computations*, 35(2):932–954, 2018.

[17] J. P. Janet, S. Ramesh, C. Duan, and H. J. Kulik. Accurate multiobjective design in a space of millions of transition metal complexes with neural-network-driven efficient global optimization. *ACS central science*, 6(4):513–524, 2020.

[18] P. Koratikere, L. T. Leifsson, L. Barnet, and K. Bryden. Efficient global optimization algorithm using neural network-based prediction and uncertainty. In *AIAA SCITECH 2023 Forum*, page 2683. 2023.

[19] A. Basudhar, S. Lacaze, and S. Missoum. Constrained efficient global optimization with probabilistic support vector machines. In *13th AIAA/ISSMO multidisciplinary analysis optimization conference*, page 9230. 2010.

[20] A. Basudhar, C. Dribusch, S. Lacaze, and S. Missoum. Constrained efficient global optimization with support vector machines. *Structural and Multidisciplinary Optimization*, 46:201–221, 2012.

[21] D. Wang, C. Wang, J. Xiao, Z. Xiao, W. Chen, and V. Havyarimana. Bayesian optimization of support vector machine for regression prediction of short-term traffic flow. *Intelligent data analysis*, 23(2):481–497, 2019.

[22] J. Mockus, V. Tiesis, and A. Zilinskas. The application of Bayesian methods for seeking the extremum. *Towards Global Optimization*, 2(117-129):2, 1978.

[23] H.-M. Gutmann. A radial basis function method for global optimization. *Journal of global optimization*, 19(3):201–227, 2001.

[24] K. Holmström. An adaptive radial basis algorithm (arbf) for expensive black-box global optimization. *Journal of Global Optimization*, 41(3):447–464, 2008.

[25] H. Jie, Y. Wu, and J. Ding. An adaptive metamodel-based global optimization algorithm for black-box type problems. *Engineering optimization*, 47(11):1459–1480, 2015.

[26] J. Müller and C. A. Shoemaker. Influence of ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization problems. *Journal of Global Optimization*, 60(2):123–144, 2014.

[27] R. G. Regis and C. A. Shoemaker. Constrained global optimization of expensive black box functions using radial basis functions. *Journal of Global optimization*, 31(1):153–171, 2005.

[28] Suprayitno and J.-C. Yu. Evolutionary reliable regional kriging surrogate for expensive optimization. *Engineering Optimization*, 51(2):247–264, 2019.

[29] L. Cavagna, S. Ricci, and L. Riccobene. A fast tool for structural sizing, aeroelastic analysis and optimization in aircraft conceptual design. In *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference 17th AIAA/ASME/AHS Adaptive Structures Conference 11th AIAA No*, page 2571. 2009.

[30] L. Simões and J. Negrão. Sizing and geometry optimization of cable-stayed bridges. *Computers & structures*, 52(2):309–321, 1994.

[31] A. Kaveh and V. Kalatjari. Size/geometry optimization of trusses by the force method and genetic algorithm. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik: Applied Mathematics and Mechanics*, 84(5):347–357, 2004.

[32] G. Rozvany and M. Zhou. The coc algorithm, part i: cross-section optimization or sizing. *Computer Methods in Applied Mechanics and Engineering*, 89(1-3):281–308, 1991.

[33] Z. Tian and D. Jiao. Discrete sizing optimization of stepped cylindrical silo using pso method and implicit dynamic fe analysis. *Engineering with Computers*, 37(2):1015–1047, 2021.

[34] P. Mohanasundaram, F. Gillot, K. Shimoyama, and S. Besset. Shape optimization of a disc-pad system under squeal noise criteria. *SN Applied Sciences*, 2(4):547, 2020.

[35] S. Y. Woon, O. M. Querin, and G. P. Steven. Structural application of a shape optimization method based on a genetic algorithm. *Structural and Multidisciplinary Optimization*, 22:57–64, 2001.

[36] P. Fourie and A. A. Groenwold. The particle swarm optimization algorithm in size and shape optimization. *Structural and Multidisciplinary Optimization*, 23:259–267, 2002.

[37] N. Camprubí, M. Bischoff, and K.-U. Bletzinger. Shape optimization of shells and locking. *Computers & structures*, 82(29-30):2551–2561, 2004.

[38] M. Edke and K. Chang. Shape optimization of heavy load carrying components for structural performance and manufacturing cost. *Structural and Multidisciplinary Optimization*, 31:344–354, 2006.

[39] D. Peng and R. Jones. An approach based on biological algorithm for three-dimensional shape optimisation with fracture strength constrains. *Computer methods in applied mechanics and engineering*, 197(49-50):4383–4398, 2008.

[40] D. Brujic, M. Ristic, M. Mattone, P. Maggiore, and G. P. De Poli. Cad based shape optimization for gas turbine component design. *Structural and Multidisciplinary Optimization*, 41:647–659, 2010.

[41] K. Shintani and H. Azegami. Shape optimization for suppressing brake squeal. *Structural and Multidisciplinary Optimization*, 50(6):1127–1135, 2014.

[42] L. Jiang, W. Zhang, G. Ma, and C. Wu. Shape optimization of energy storage flywheel rotor. *Structural and multidisciplinary optimization*, 55:739–750, 2017.

[43] U. Schramm and W. D. Pilkey. The coupling of geometric descriptions and finite elements using nurbs—a study in shape optimization. *Finite elements in analysis and design*, 15(1):11–34, 1993.

[44] Y.-D. Seo, H.-J. Kim, and S.-K. Youn. Shape optimization and its extension to topological design based on isogeometric analysis. *International Journal of Solids and Structures*, 47(11-12):1618–1640, 2010.

[45] R. Haftka. Structural optimization with aeroelastic constraints: a survey of us applications. *International Journal of Vehicle Design*, 7(3-4):381–392, 1986.

[46] G.-W. Jang, Y. Y. Kim, and K. K. Choi. Remesh-free shape optimization using the wavelet-galerkin method. *International journal of solids and structures*, 41(22-23):6465–6483, 2004.

[47] D. Lacroix and P. Bouillard. Improved sensitivity analysis by a coupled fe–efg method. *Computers & structures*, 81(26-27):2431–2439, 2003.

[48] K. K. Choi and N.-H. Kim. *Structural sensitivity analysis and optimization 1: linear systems.* Springer Science & Business Media, 2004.

[49] M. L. Staten, S. J. Owen, S. M. Shontz, A. G. Salinger, and T. S. Coffey. A comparison of mesh morphing methods for 3 d shape optimization. In *Proceedings of the 20th international meshing roundtable*, pages 293–311. Springer, 2012.

[50] B. Saalfeld, M. Rütten, S. Saalfeld, and J. Künemund. Improved mesh morphing based on radial basis functions. In *ECCOMAS*, pages 10–14. 2012.

[51] J. Bennett and M. Botkin. Structural shape optimization with geometric description and adaptive mesh refinement. *AIAA journal*, 23(3):458–464, 1985.

[52] N. Kikuchi, K. Y. Chung, T. Torigaki, and J. E. Taylor. Adaptive finite element methods for shape optimization of linearly elastic structures. *The Optimum Shape: Automated Structural Design*, pages 139–169, 1986.

[53] J. Canales, J. Tárrago, and A. Hernández. An adaptive mesh refinement procedure for shape optimal design. *Advances in Engineering Software*, 18(2):131–145, 1993.

[54] P. Mohite and C. Upadhyay. Adaptive finite element based shape optimization in laminated composite plates. *Computers & Structures*, 153:19–35, 2015.

[55] R. R. Salagame and A. D. Belegundu. Shape optimization with p adaptivity. *AIAA journal*, 33(12):2399–2405, 1995.

[56] A. Schleupen, K. Maute, and E. Ramm. Adaptive fe-procedures in shape optimization. *Structural and Multidisciplinary Optimization*, 19:282–302, 2000.

[57] N. Moës, J. Dolbow, and T. Belytschko. A finite element method for crack growth without remeshing. *International journal for numerical methods in engineering*, 46(1):131–150, 1999.

[58] L. Noël and P. Duysinx. Shape optimization of microstructural designs subject to local stress constraints within an xfem-level set framework. *Structural and Multidisciplinary Optimization*, 55:2323–2338, 2017.

[59] N. H. Kim and Y. Chang. Eulerian shape design sensitivity analysis and optimization with a fixed grid. *Computer methods in applied mechanics and engineering*, 194(30-33):3291–3314, 2005.

[60] S. Woon, O. Querin, and G. Steven. On improving the ga step-wise shape optimization method through the application of the fixed grid fea paradigm. *Structural and multidisciplinary optimization*, 25:270–278, 2003.

[61] S. Soghrati, A. M. Aragón, C. Armando Duarte, and P. H. Geubelle. An interface-enriched generalized fem for problems with discontinuous gradient fields. *International Journal for numerical methods in engineering*, 89(8):991–1008, 2012.

[62] A. R. Najafi, M. Safdari, D. A. Tortorelli, and P. H. Geubelle. A gradient-based shape optimization scheme using an interface-enriched generalized fem. *Computer Methods in Applied Mechanics and Engineering*, 296:1–17, 2015.

[63] A. R. Najafi, M. Safdari, D. A. Tortorelli, and P. H. Geubelle. Shape optimization using a nurbs-based interface-enriched generalized fem. *International Journal for Numerical Methods in Engineering*, 111(10):927–954, 2017.

[64] J. Parvizian, A. Düster, and E. Rank. Finite cell method: h-and p-extension for embedded domain problems in solid mechanics. *Computational Mechanics*, 41(1):121–133, 2007.

[65] D. Schillinger and M. Ruess. The finite cell method: A review in the context of higher-order structural analysis of cad and image-based geometric models. *Archives of Computational Methods in Engineering*, 22:391–455, 2015.

[66] T. J. Hughes, J. A. Cottrell, and Y. Bazilevs. Isogeometric analysis: Cad, finite elements, nurbs, exact geometry and mesh refinement. *Computer methods in applied mechanics and engineering*, 194(39-41):4135–4195, 2005.

[67] P. Kagan, A. Fischer, and P. Z. Bar-Yoseph. New b-spline finite element approach for geometrical design and mechanical analysis. *International Journal for Numerical Methods in Engineering*, 41(3):435–458, 1998.

[68] P. Kagan, A. Fischer, and P. Z. Bar-Yoseph. Integrated mechanically based cae system. In *Proceedings of the fifth ACM symposium on Solid modeling and applications*, pages 23–30. 1999.

[69] W. A. Wall, M. A. Frenzel, and C. Cyron. Isogeometric structural shape optimization. *Computer methods in applied mechanics and engineering*, 197(33-40):2976–2988, 2008.

[70] H. Azegami, S. Fukumoto, and T. Aoyama. Shape optimization of continua using nurbs as basis functions. *Structural and Multidisciplinary Optimization*, 47:247–258, 2013.

[71] K. Kostas, A. Ginnis, C. Politis, and P. Kaklis. Ship-hull shape optimization with a t-spline based bem–isogeometric solver. *Computer Methods in Applied Mechanics and Engineering*, 284:611–622, 2015.

[72] L. Chamoin and H. Thai. Certified real-time shape optimization using isogeometric analysis, pgd model reduction, and a posteriori error estimation. *International Journal for Numerical Methods in Engineering*, 119(3):151–176, 2019.

[73] X. Du, G. Zhao, W. Wang, and H. Fang. Nitsche's method for non-conforming multipatch coupling in hyperelastic isogeometric analysis. *Computational Mechanics*, 65:687–710, 2020.

[74] M. Chasapi, W. Dornisch, and S. Klinkel. Patch coupling in isogeometric analysis of solids in boundary representation using a mortar approach. *International Journal for Numerical Methods in Engineering*, 121(14):3206–3226, 2020.

[75] Y. Bazilevs, V. M. Calo, J. A. Cottrell, J. A. Evans, T. J. R. Hughes, S. Lipton, M. A. Scott, and T. W. Sederberg. Isogeometric analysis using t-splines. *Computer methods in applied mechanics and engineering*, 199(5-8):229–263, 2010.

[76] M. Guo, G. Zhao, W. Wang, X. Du, R. Zhang, and J. Yang. T-splines for isogeometric analysis of two-dimensional nonlinear problems. *Computer Modeling in Engineering & Sciences*, 123(2):821–843, 2020.

[77] M. P. Bendsøe and N. Kikuchi. Generating optimal topologies in structural design using a homogenization method. *Computer methods in applied mechanics and engineering*, 71(2):197–224, 1988.

[78] M. Zhou and G. I. Rozvany. The coc algorithm, part ii: Topological, geometrical and generalized shape optimization. *Computer methods in applied mechanics and engineering*, 89(1-3):309–336, 1991.

[79] M. P. Bendsoe and O. Sigmund. *Topology optimization: theory, methods, and applications.* Springer Science & Business Media, 2003.

[80] K. D. Tsavdaridis, J. J. Kingman, and V. V. Toropov. Application of structural topology optimisation to perforated steel beams. *Computers & structures*, 158:108–123, 2015.

[81] W.-h. Choi, J.-m. Kim, and G.-J. Park. Comparison study of some commercial structural optimization software systems. *Structural and multidisciplinary optimization*, 54:685–699, 2016.

[82] T. Van Truong, U. Kureemun, V. B. C. Tan, and H. P. Lee. Study on the structural optimization of a flapping wing micro air vehicle. *Structural and Multidisciplinary Optimization*, 57:653–664, 2018.

[83] Y. M. Xie and G. P. Steven. A simple evolutionary procedure for structural optimization. *Computers & structures*, 49(5):885–896, 1993.

[84] X. Y. Yang, Y. M. Xie, G. P. Steven, and O. Querin. Bidirectional evolutionary method for stiffness optimization. *AIAA journal*, 37(11):1483–1488, 1999.

[85] M. Yulin and W. Xiaoming. A level set method for structural topology optimization and its applications. *Advances in Engineering software*, 35(7):415–441, 2004.

[86] W. Zhang, Y. Zhou, and J. Zhu. A comprehensive study of feature definitions with solids and voids for topology optimization. *Computer Methods in Applied Mechanics and Engineering*, 325:289–313, 2017.

[87] Y. Zhou, W. Zhang, J. Zhu, and Z. Xu. Feature-driven topology optimization method with signed distance function. *Computer Methods in Applied Mechanics and Engineering*, 310:1–32, 2016.

[88] J.-H. Zhu, W.-H. Zhang, and L. Xia. Topology optimization in aircraft and aerospace structures design. *Archives of computational methods in engineering*, 23:595–622, 2016.

[89] S. Guanghui, G. Chengqi, Q. Dongliang, W. Dongtao, T. Lei, and G. Tong. An aerospace bracket designed by thermo-elastic topology optimization and manufactured by additive manufacturing. *Chinese Journal of Aeronautics*, 33(4):1252–1259, 2020.

[90] D. Jankovics and A. Barari. Customization of automotive structural components using additive manufacturing and topology optimization. *IFAC-PapersOnLine*, 52(10):212–217, 2019.

[91] J. L. Jewett and J. V. Carstensen. Topology-optimized design, construction and experimental evaluation of concrete beams. *Automation in Construction*, 102:59–67, 2019.

[92] S. Wu, Y. Zhang, and S. Liu. Topology optimization for minimizing the maximum temperature of transient heat conduction structure. *Structural and Multidisciplinary Optimization*, 60:69–82, 2019.

[93] Y. Li, Y. Lai, G. Lu, F. Yan, P. Wei, and Y. M. Xie. Innovative design of long-span steel–concrete composite bridge using multi-material topology optimization. *Engineering Structures*, 269:114838, 2022.

[94] U. M. Dilberoglu, B. Gharehpapagh, U. Yaman, and M. Dolen. The role of additive manufacturing in the era of industry 4.0. *Procedia manufacturing*, 11:545–554, 2017.

[95] A. Paolini, S. Kollmannsberger, and E. Rank. Additive manufacturing in construction: A review on processes, applications, and digital planning methods. *Additive manufacturing*, 30:100894, 2019.

[96] C. Soize. *Uncertainty quantification.* Springer, 2017.

[97] I. Elishakoff. *Safety Factors and Reliability: Friends or Foes?: Friends Or Foes?* Springer Science & Business Media, 2004.

[98] S.-K. Choi, R. A. Canfield, and R. V. Grandhi. *Reliability-Based Structural Optimization.* Springer, 2007.

[99] E. Vanmarcke. *Random fields: analysis and synthesis.* World scientific, 2010.

[100] H. G. Matthies, C. E. Brenner, C. G. Bucher, and C. G. Soares. Uncertainties in probabilistic numerical analysis of structures and solids-stochastic finite elements. *Structural safety*, 19(3):283–336, 1997.

[101] A. M. Hasofer and N. C. Lind. Exact and invariant second-moment code format. *Journal of the Engineering Mechanics division*, 100(1):111–121, 1974.

[102] M. Hohenbichler and R. Rackwitz. First-order concepts in system reliability. *Structural safety*, 1(3):177–188, 1982.

[103] M. Hohenbichler and R. Rackwitz. Non-normal dependent vectors in structural safety. *Journal of the Engineering Mechanics Division*, 107(6):1227–1238, 1981.

[104] M. Rosenblatt. Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3):470–472, 1952.

[105] B. Fiessler, H.-J. Neumann, and R. Rackwitz. Quadratic limit states in structural reliability. *Journal of the Engineering Mechanics Division*, 105(4):661–676, 1979.

[106] L. Wang and R. V. Grandhi. Improved two-point function approximations for design optimization. *AIAA journal*, 33(9):1720–1727, 1995.

[107] R. E. Melchers and A. T. Beck. *Structural reliability analysis and prediction.* John wiley & sons, 2018.

[108] K. Breitung *et al.* Asymptotic approximations for multinormal integrals. *Journal of Engineering Mechanics*, 110(3):357–366, 1984.

[109] G. Cai and I. Elishakoff. Refined second-order reliability analysis. *Structural Safety*, 14(4):267–276, 1994.

[110] M. Hohenbichler and R. Rackwitz. Improvement of second-order reliability estimates by importance sampling. *Journal of Engineering Mechanics*, 114(12):2195–2199, 1988.

[111] L. Tvedt. Two second-order approximations to the failure probability. *Veritas report RDIV/20-004083*, 1983.

[112] L. Tvedt. Distribution of quadratic forms in normal space—application to structural reliability. *Journal of engineering mechanics*, 116(6):1183–1197, 1990.

[113] A. Der Kiureghian, H.-Z. Lin, and S.-J. Hwang. Second-order reliability approximations. *Journal of Engineering mechanics*, 113(8):1208–1225, 1987.

[114] H. Agarwal and J. Renaud. Reliability based design optimization using response surfaces in application to multidisciplinary systems. *Engineering Optimization*, 36(3):291–311, 2004.

[115] I. Enevoldsen and J. D. Sørensen. Reliability-based optimization in structural engineering. *Structural safety*, 15(3):169–196, 1994.

[116] Y. Aoues and A. Chateauneuf. Benchmark study of numerical methods for reliability-based design optimization. *Structural and Multidisciplinary Optimization*, 41(2):277–294, 2010.

[117] A. J. Torii, R. H. Lopez, and L. F. F. Miguel. A general rbdo decoupling approach for different reliability analysis methods. *Structural and Multidisciplinary Optimization*, 54(2):317–332, 2016.

[118] X. Zhang. Review of reliability-based design optimization approach and its integration with bayesian method. *IOP Conference Series: Earth and Environmental Science*, 128(1):012109, 2018.

[119] I. M. Sobol. *A primer for the Monte Carlo method.* CRC press, 2018.

[120] G. D. Wyss and K. H. Jorgensen. A users guide to lhs: Sandias latin hypercube sampling software. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 1998.

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

[122] J.-S. Park. Optimal latin-hypercube designs for computer experiments. *Journal of statistical planning and inference*, 39(1):95–111, 1994.

[123] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, pages 385–392, 1947.

[124] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.

[125] M. A. Tatang. Direct incorporation of uncertainty in chemical and environmental engineering systems. 1996.

[126] S. S. Isukapalli. *Uncertainty analysis of transport-transformation models*. Rutgers The State University of New Jersey-New Brunswick, 1999.

[127] D. Xiu and G. E. Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644, 2002.

[128] D. Xiu, D. Lucor, C.-H. Su, and G. E. Karniadakis. Stochastic modeling of flow-structure interactions using generalized polynomial chaos. *J. Fluids Eng.*, 124(1):51–59, 2002.

[129] L. Devroye. Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pages 260–265. 1986.

[130] C. Hu and B. D. Youn. Adaptive-sparse polynomial chaos expansion for reliability analysis and design of complex engineering systems. *Structural and Multidisciplinary Optimization*, 43(3):419–442, 2011.

[131] N. Luthen, S. Marelli, and B. Sudret. Sparse polynomial chaos expansions: Literature survey and benchmark. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):593–649, 2021.

[132] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of computational Physics*, 230(6):2345–2367, 2011.

[133] S. Chair of Risk and E. Z. Uncertainty Quantification. Pce metamodeling: comparison of pce calculation strategies. `https://www.uqlab.com/pce-calculation-strategies`, 2024. Accessed on May 14, 2024.

[134] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[135] R. L. Iman and W.-J. Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics-Simulation and Computation*, 11(3):311–334, 1982.

[136] M. Stein. Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987.

[137] A. Olsson, G. Sandberg, and O. Dahlblom. On latin hypercube sampling for structural reliability analysis. *Structural safety*, 25(1):47–68, 2003.

[138] E. J. Pebesma and G. B. Heuvelink. Latin hypercube sampling of gaussian random fields. *Technometrics*, 41(4):303–312, 1999.

[139] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach.* Courier Corporation, 2003.

[140] I. The MathWorks. Disc brake. `https://fr.mathworks.com/help/sdl/ref/discbrake.html`, 2019. Accessed on May 14, 2024.

[141] A. Akay. Acoustics of friction. *The Journal of the Acoustical Society of America*, 111(4):1525–1548, 2002.

[142] R. A. Ibrahim. Friction-Induced Vibration, Chatter, Squeal, and Chaos—Part I: Mechanics of Contact and Friction. *Applied Mechanics Reviews*, 47(7):209–226, 1994.

[143] R. A. Ibrahim. Friction-Induced Vibration, Chatter, Squeal, and Chaos—Part II: Dynamics and Modeling. *Applied Mechanics Reviews*, 47(7):227–253, 1994.

[144] R. T. Spurr. A theory of brake squeal. *Proceedings of the Institution of Mechanical Engineers: Automobile Division*, 15(1):33–52, 1961.

[145] J.-J. Sinou. *Synthèse non-linéaire des systèmes vibrants : Application aux systèmes de freinage.* Ph.D. thesis, Ecole Centrale de Lyon, 2002. Thèse de doctorat dirigée par Jézéquel, Louis science. Mécanique Ecully, Ecole centrale de Lyon 2002.

[146] G. Corradi. *Approches numérique et expérimentale pour des structures à interfaces frottantes : Application au crissement de frein.* Ph.D. thesis, Ecole Centrale de Lyon, 2019. Thèse de doctorat dirigée par Sinou, Jean-Jacques et Besset, Sébastien Mécanique, Energétique, Génie Civil et Acoustique Lyon 2019.

[147] K. Soobbarayen. *Dynamique non linéaire et rayonnement acoustique de structures comportant des interfaces frottantes.* Ph.D. thesis, Ecole Centrale de Lyon, 2014. Thèse de doctorat dirigée par Sinou, Jean-Jacques et Besset, Sébastien Mécanique Ecully, Ecole centrale de Lyon 2014.

[148] A. Elmaian, F. Gautier, C. Pezerat, and J.-M. Duffal. How can automotive friction-induced noises be related to physical mechanisms? *Applied Acoustics*, 76:391–401, 2014.

[149] N. Kinkaid, O. O'Reilly, and P. Papadopoulos. Automotive disc brake squeal. *Journal of Sound and Vibration*, 267(1):105–166, 2003.

[150] J. Oden and J. Martins. Models and computational methods for dynamic friction phenomena. *Computer Methods in Applied Mechanics and Engineering*, 52(1):527–634, 1985.

[151] C. Gao, D. Kuhlmann-Wilsdorf, and D. D. Makel. The dynamic analysis of stick-slip motion. *Wear*, 173(1):1–12, 1994.

[152] S. Antoniou, A. Cameron, and C. Gentle. The friction-speed relation from stick-slip data. *Wear*, 36(2):235–254, 1976.

[153] A. Banerjee. Influence of kinetic friction on the critical velocity of stick-slip motion. *Wear*, 12(2):107–116, 1968.

[154] N. Hinrichs, M. Oestreich, and K. Popp. On the modelling of friction oscillators. *Journal of Sound and Vibration*, 216(3):435–459, 1998.

[155] S. Liu, J. T. Gordon, and M. Ozbek. Nonlinear model for aircraft brake squeal analysis: Model description and solution methodology. *Journal of Aircraft*, 35:623–630, 1996.

[156] G. R. P. Jarvis and A. M. B. Mills. Vibrations induced by dry friction. *Proceedings of the Institution of Mechanical Engineers*, 178(1):847–857, 1963.

[157] K. Shin, M. Brennan, J.-E. Oh, and C. Harris. Analysis of disc brake noise using a two-degree-of-freedom model. *Journal of Sound and Vibration*, 254(5):837–848, 2002.

[158] N. Hoffmann, M. Fischer, R. Allgaier, and L. Gaul. A minimal model for studying properties of the mode-coupling type instability in friction induced oscillations. *Mechanics Research Communications*, 29(4):197–205, 2002.

[159] M. Bengisu and A. Akay. Stability of friction-induced vibrations in multi-degree-of-freedom systems. *Journal of Sound and Vibration*, 171(4):557–570, 1994.

[160] J. Flint and J. Hulten. Lining-deformation-induced modal coupling as squeal generator in a distributed parameter disc brake model. *Journal of sound and vibration*, 254(1):1–21, 2002.

[161] U. von Wagner, D. Hochlenert, and P. Hagedorn. Minimal models for disk brake squeal. *Journal of Sound and Vibration*, 302(3):527–539, 2007.

[162] G. Spelsberg-Korspeter. Eigenvalue optimization against brake squeal: Symmetry, mathematical background and experiments. *Journal of Sound and Vibration*, 331(19):4259–4268, 2012.

[163] G. V. Des Roches. *Frequency and time simulation of squeal instabilities. Application to the design of industrial automotive brakes.* Ph.D. thesis, Ecole Centrale Paris, 2011.

[164] S. Oberst and J. Lai. Nonlinear transient and chaotic interactions in disc brake squeal. *Journal of Sound and Vibration*, 342:272–289, 2015.

[165] L. Charroyer. *Méthodes numériques pour le calcul des vibrations auto-entretenues liées au frottement: Application au bruit de crissement ferroviaire.* Ph.D. thesis, Lyon, 2017.

[166] H. Lee and R. Singh. Determination of sound radiation from a simplified disk-brake rotor by a semi-analytical method. *Noise Control Engineering Journal*, 52(5):225–239, 2004.

[167] S. Oberst, J. Lai, and S. Marburg. Guidelines for numerical vibration and acoustic analysis of disc brake squeal using simple models of brake systems. *Journal of Sound and Vibration*, 332(9):2284–2299, 2013.

[168] B. D. Hassard, N. D. Kazarinoff, and Y.-H. Wan. *Theory and applications of Hopf bifurcation*, volume 41. CUP Archive, 1981.

[169] S.-W. Kung, K. B. Dunlap, and R. S. Ballinger. Complex eigenvalue analysis for reducing low frequency brake squeal. *SAE transactions*, pages 559–565, 2000.

[170] P. Liu, H. Zheng, C. Cai, Y. Wang, C. Lu, K. Ang, and G. Liu. Analysis of disc brake squeal using the complex eigenvalue method. *Applied acoustics*, 68(6):603–615, 2007.

[171] A. R. AbuBakar and H. Ouyang. A prediction methodology of disk brake squeal using complex eigenvalue analysis. *International Journal of Vehicle Design*, 46(4):416–435, 2008.

[172] H. Lü and D. Yu. Brake squeal reduction of vehicle disc brake system with interval parameters by uncertain optimization. *Journal of Sound and Vibration*, 333(26):7313–7325, 2014.

[173] H. Lü and D. Yu. Stability optimization of a disc brake system with hybrid uncertainties for squeal reduction. *Shock and Vibration*, 2016, 2016.

[174] Z. Chu, F. Zheng, L. Liang, H. Yan, and R. Kang. Parameter determination of a minimal model for brake squeal. *Applied Sciences*, 8(1):37, 2018.

[175] J.-J. Sinou and L. Jézéquel. Mode coupling instability in friction-induced vibrations and its dependency on system parameters including damping. *European Journal of Mechanics-A/Solids*, 26(1):106–122, 2007.

[176] A. R. AbuBakar and H. Ouyang. Complex eigenvalue analysis and dynamic transient analysis in predicting disc brake squeal. *International Journal of Vehicle Noise and Vibration*, 2(2):143–155, 2006.

[177] J.-J. Sinou. Transient non-linear dynamic analysis of automotive disc brake squeal–on the need to consider both stability and non-linear analysis. *Mechanics Research Communications*, 37(1):96–105, 2010.

[178] N. Coudeyras, J.-J. Sinou, and S. Nacivet. A new treatment for predicting the self-excited vibrations of nonlinear systems with frictional interfaces: The constrained harmonic balance method, with application to disc brake squeal. *Journal of sound and vibration*, 319(3-5):1175–1199, 2009.

[179] N. Coudeyras, S. Nacivet, and J.-J. Sinou. Periodic and quasi-periodic solutions for multi-instabilities involved in brake squeal. *Journal of Sound and vibration*, 328(4-5):520–540, 2009.

[180] J. Brunetti, F. Massi, Y. Berthier, *et al.* A new instability index for unstable mode selection in squeal prediction by complex eigenvalue analysis. *Journal of Sound and Vibration*, 377:106–122, 2016.

[181] S. Nacivet and J.-J. Sinou. Modal amplitude stability analysis and its application to brake squeal. *Applied Acoustics*, 116:127–138, 2017.

[182] R. Kohavi, D. H. Wolpert, *et al.* Bias plus variance decomposition for zero-one loss functions. In *ICML*, volume 96, pages 275–83. 1996.

[183] P. Mohanasundaram, F. Gillot, S. Besset, and K. Shimoyama. Multi-references acquisition strategy for shape optimization of disc-pad-like mechanical systems. *Struct. Multidiscip. Optim.*, 64(4):1863–1885, 2021.

[184] V. Picheny, T. Wagner, and D. Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626, 2013.

[185] H. Zhang and T. Shibutani. Development of stochastic isogeometric analysis (siga) method for uncertainty in shape. *International Journal for Numerical Methods in Engineering*, 118(1):18–37, 2019.

[186] Y. Bazilevs, V. Calo, J. Cottrell, J. Evans, T. Hughes, S. Lipton, M. Scott, and T. Sederberg. Isogeometric analysis using t-splines. *Computer Methods in Applied Mechanics and Engineering*, 199(5):229–263, 2010. Computational Geometry and Analysis.

[187] J.-Y. Kim and S.-K. Youn. Isogeometric contact analysis using mortar method. *International Journal for Numerical Methods in Engineering*, 89(12):1559–1581, 2012.

[188] J. Cottrell, T. Hughes, and A. Reali. Studies of refinement and continuity in isogeometric structural analysis. *Computer methods in applied mechanics and engineering*, 196(41-44):4160–4183, 2007.

[189] X. Du, G. Zhao, W. Wang, M. Guo, R. Zhang, and J. Yang. Nliga: A matlab framework for nonlinear isogeometric analysis. *Computer Aided Geometric Design*, 80:101869, 2020.

[190] S. Chan, J. Mottershead, and M. Cartmell. Parametric resonances at subcritical speeds in discs with rotating frictional loads. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 208(6):417–425, 1994.

[191] J. Kang, C. M. Krousgrill, and F. Sadeghi. Dynamic instability of a thin circular plate with friction interface and its application to disc brake squeal. *Journal of Sound and Vibration*, 316(1-5):164–179, 2008.

[192] K. I. Tzou, J. A. Wickert, and A. Akay. In-plane vibration modes of arbitrarily thick disks. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 80432, page V01DT20A008. American Society of Mechanical Engineers, 1997.

[193] G. D. Liles. Analysis of disc brake squeal using finite element methods. *SAE transactions*, pages 1138–1146, 1989.

[194] L. De Lorenzis, P. Wriggers, and T. J. Hughes. Isogeometric contact: a review. *GAMM-Mitteilungen*, 37(1):85–123, 2014.

[195] A. Francavilla and O. Zienkiewicz. A note on numerical computation of elastic contact problems. *International Journal for Numerical Methods in Engineering*, 9(4):913–924, 1975.

[196] W. Xing, J. Zhang, C. Song, and F. Tin-Loi. A node-to-node scheme for three-dimensional contact problems using the scaled boundary finite element method. *Computer Methods in Applied Mechanics and Engineering*, 347:928–956, 2019.

[197] G. Zavarise and L. De Lorenzis. The node-to-segment algorithm for 2d frictionless contact: classical formulation and special cases. *Computer Methods in Applied Mechanics and Engineering*, 198(41-44):3428–3451, 2009.

[198] P. Papadopoulos and R. L. Taylor. A mixed formulation for the finite element solution of contact problems. *Computer Methods in Applied Mechanics and Engineering*, 94(3):373–389, 1992.

[199] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*, volume 15. Springer Science & Business Media, 2012.

[200] C. Bernardi, N. Debit, and Y. Maday. Coupling finite element and spectral methods: First results. *Mathematics of Computation*, 54(189):21–39, 1990.

[201] T. McDevitt and T. Laursen. A mortar-finite element formulation for frictional contact problems. *International Journal for Numerical Methods in Engineering*, 48(10):1525–1547, 2000.

[202] G. Fritz, J.-J. Sinou, J.-M. Duffal, and L. Jézéquel. Effects of damping on brake squeal coalescence patterns–application on a finite element model. *Mechanics Research Communications*, 34(2):181–190, 2007.

[203] I. Freni. Friction materials. `https://www.imafreni.it/en/friction-materials/`, 2022. Accessed on October 1, 2023.

[204] F. Technology. Data sheets. `https://www.frictiontechnology.co.uk/data-sheets/`, 2019. Accessed on October 1, 2023.

[205] N. Lüthen, S. Marelli, and B. Sudret. Sparse polynomial chaos expansions: Literature survey and benchmark. *arXiv preprint arXiv:2002.01290*, 2020.

[206] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.

[207] S. Marelli and B. Sudret. Uqlab: A framework for uncertainty quantification in matlab. In *Vulnerability, uncertainty, and risk: quantification, mitigation, and management*, pages 2554–2563. 2014.

[208] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

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

[210] J. D. Martin and T. W. Simpson. A study on the use of kriging models to approximate deterministic computer models. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 37009, pages 567–576. 2003.

[211] O. Roustant, D. Ginsbourger, and Y. Deville. Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of statistical software*, 51:1–55, 2012.

[212] J.-S. Park and J. Baek. Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram. *Computers & Geosciences*, 27(1):1–7, 2001.

[213] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[214] J. L. Loeppky, J. Sacks, and W. J. Welch. Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, 51(4):366–376, 2009.

[215] N. Altman and M. Krzywinski. The curse (s) of dimensionality. *Nat Methods*, 15(6):399–400, 2018.

[216] C. C. Aggarwal. *Neural Networks and Deep Learning.* Springer, 2018.

[217] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[218] F. Chollet *et al.* Keras. `https://keras.io`, 2015.

[219] B. Ramsundar and R. B. Zadeh. *TensorFlow for deep learning: from linear regression to reinforcement learning.* " O'Reilly Media, Inc.", 2018.

[220] Module: tf.keras.activations. `https://www.tensorflow.org/api_docs/python/tf/keras/activations`, 2015. Accessed on May 22, 2024.

[221] Module: tf.keras.optimizers. `https://www.tensorflow.org/api_docs/python/tf/keras/optimizers`, 2015. Accessed on May 22, 2024.

[222] M. Ibnu Choldun R, J. Santoso, and K. Surendro. Determining the number of hidden layers in neural network by using principal component analysis. In *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2*, pages 490–500. Springer, 2020.