Multi-references acquisition strategy for shape optimization of disc-pad-like mechanical systems

Pradeep Mohanasundaram, Frédéric Gillot, Sébastien Besset, Koji Shimoyama

Noname manuscript No.
(will be inserted by the editor)

Multi-references acquisition strategy for shape optimization of disc-pad like mechanical systems

Pradeep Mohanasundaram · Frédéric Gillot · Sébastien Besset · Koji Shimoyama

Received: date / Accepted: date

Abstract We expose an efficient strategy to deal with shape optimization of dynamical systems exhibiting flutter-type instability induced by friction, such as the considered disc-pad system. The stability of such systems can be analysed through Complex-Eigenvalue Analysis, through which we present a squeal noise criterion to be minimized as a computationally expensive black-box function. The computational domain is discretized through Isogeometric formulation for its advantages in optimization and superior approximation properties which are well studied in structural dynamics. To be computationally efficient with the expensive black-box function, we defined the optimization based on Efficient Global Optimization scheme in the context of multi-objective optimization, with the integration of Isogeometric design-through-analysis methodology. As gradient information is hard to access for such black-box functions, in addition to the presence of constraints, we relied

P. Mohanasundaram
LTDS, Ecole Centrale de Lyon, 36 av. Guy de Colongue, 69134 Écully cedex, France
IFS, Tohoku University, Sendai, Japan
Department of Aerospace Engineering, Tohoku University, Sendai, Japan
ElyTMaX UMI 3757, CNRS, Université de lyon, Tohoku University, international Joint Unit, Tohoku University, Sendai, Japan
E-mail: pradeep.mohanasundaram@ec-lyon.fr

F. Gillot
ElyTMaX UMI 3757, CNRS, Université de lyon, Tohoku University, international Joint Unit, Tohoku University, Sendai, Japan
E-mail: frederic.gillot@ec-lyon.fr

S. Besset
LTDS, Ecole Centrale de Lyon, 36 av. Guy de Colongue, 69134 Écully cedex, France
E-mail: sebastien.besset@ec-lyon.fr

K. Shimoyama
IFS, Tohoku University, Sendai, Japan
ElyTMaX UMI 3757, CNRS, Université de lyon, Tohoku University, international Joint Unit, Tohoku University, Sendai, Japan
E-mail: shimoyama@tohoku.ac.jp

on meta-heuristic approach as a more generic strategy for realizing optimization of such functions in multi-objective context. As one such scheme with its own advantages was observed to provide lack of resolution to define Expected Improvement (EI) with a single reference value, we propose a multi-reference acquisition strategy which can be defined through a fast and efficient algorithm with fewer adaptation to the existing scheme. Results show the efficiency of this approach for our applicative example, which can be extended to other such applications as well.

Keywords Dynamic instability · Friction · Isogeometric analysis · Meta-model · Shape optimization · Bayesian optimization · Multi-objective optimization · Meta-heuristic optimization

1 Introduction

Flutter-type dynamic instability typically defines a self-excitation behaviour in the presence of non-conservative forces. In structural dynamics, this is understood as coalescence of modes, where two modes exist at a same frequency leading to self-excitation between the modes under favorable conditions in the presence of non-conservative forces. We consider shape optimization of braking system through a simple disc-pad representation, where this type of systems can exhibit flutter-type dynamic instability in the presence of friction, perceived as squeal noise (Akay, 2002; Ibrahim, 1994a,b; Kindkaid et al, 2003; Ouyang et al, 2005; Shintani and Azegami, 2014).

Typically, friction induced dynamic instabilities are highly nonlinear phenomena which can be computationally expensive when defined through transient analyses and hence, unrealistic to be considered for optimization. The definition of follower force model for friction makes it possible to define this type of systems as a time-independent linear dynamical system around a fixed point defined through quasi-static hypothesis, which otherwise requires satisfying non-holonomic constraints with strong time dependence (Herrmann, 1971; Hibbit, 1979). Hence, the stability of such linearized systems around a fixed point can be defined through its eigenvalues, commonly known as Complex-Eigenvalue Analysis (CEA) (Abu-Bakar and Ouyang, 2006; Martins et al, 1999; Mottershead and Chan, 1995). Through CEA, we define a black-box function which is adversely expensive for computation, to describe a criterion for stability in shape optimization. Further, for evaluation of the expensive black-box function, we define a parallel computation strategy through dynamic model reduction.

To realize an efficient generic strategy to deal with shape optimization of an arbitrary domain for computationally expensive black-box functions, we encompass Isogeometric approach for discretization (Hughes et al, 2005), and Efficient Global Optimization (EGO) (Jones et al, 1998) approach in the

context of Multi-Objective Optimization (MOO) –commonly known as Multi-Objective Bayesian Optimization (MOBO).

Systems with complex domain are solved through numerical methods like Finite Element Method (FEM) which demands a robust meshing scheme in an optimization loop without human intervention. The limitation of discretisation with classical FEM for optimization is well known. The problem is overcome by Isogeometric approach for discretization where an initial design parameterization and its subsequent analysis-suitable parameterization are defined with in the same parametric space, which avoids the bottle-neck in communication between the two parameterizations and hence, robust meshing could be achieved (Hughes et al, 2005; Nguyen et al, 2011; Nørtoft and Gravesen, 2013; Philipp et al, 2016). Further, with Isogeometric approach, the approximation properties of Non-uniform rational B-spline (NURBS) basis functions lead to better convergence per degree-of-freedom compared to classical finite element class of basis functions, which has been discussed for structural dynamics in (Cottrell et al, 2006). This provides better accuracy to work in high frequency domains which is especially more deemed in vibro-acoustic applications.

Unfortunately numerical methods can still be very computationally expensive in optimization to explore a design space efficiently for convergence considering loop iteration, which is especially more true in MOO. Hence, we define a strategy based on Efficient Global Optimization (EGO) for MOO where the optimal solutions are defined through Pareto-optimal/Non-dominated solutions (NDS). The complexity in accessing gradient information for the black-box function and the presence of constraints make it highly infeasible to define a gradient based optimization scheme and hence, we relied on meta-heuristic approach to define a more generic strategy.

The EGO approach defines optimization through an acquisition function Expected Improvement (EI) which characterizes improvement of a reference value for a given Gaussian prediction from the Gaussian process ($\mathcal{GP}$)/Kriging meta-model (Rasmussen and Williams, 2005) typically used to approximate a computationally expensive function. For a single objective optimization, the reference value corresponds to the known minimum of the function (Utopian value).

The direct extension of this in MOO corresponds to seeking improvement considering an empirical NDS for a multi-variate Gaussian prediction in the objective space, commonly known as Expected hypervolume improvemet ($EHVI$) (Emmerich et al, 2011). The other approach is to define EI independently for all the $\mathcal{GP}$ meta-models with their respective reference value and optimising in the context of MOO, where Utopian value was used as reference in (Jeong and Obayashi, 2006). Though this approach is more simple and versatile to implement, it was found to provide less resolution to define improvement in MOO with a single reference value, detailed in section 5.2.

Hence, with the aim of preserving the simplicity of the approach, we provide an extension of this concept with multiple reference values for a $\mathcal{GP}$ metamodel in MOO, through a simple strategy with fewer adaption to the existing method of EI.

Following introduction, the second section describes the squeal noise problem as our applicative example for flutter-type dynamic instability. The third section introduces Isogeometric formulation as an approach to finite element discretization of computational domains, with a brief description of contact and friction formulation specific for modelling this type of dynamic instability as a time independent definition. In the fourth section, we give a brief explanation about CEA and a stability criterion for shape optimization realized through CEA. The fifth section details optimization setting for shape optimization with Isogeometric approach, given with the objectives for optimization and description of parameterization for the disc-pad system domain. The sixth section provides an overall frame work for MOBO with the definition of a new acquisition function. The proposed approach is then applied for optimization of the disc-pad system and the results are exposed in section seven before concluding in the last section.

2 Model description

Brake squeal phenomenon is an important concern when working on automotive brake designs, where squeal noise can be characterized with a frequency range of 1-16 KHz. The squeal noise can lead to customer complaints which can be detrimental for automotive manufacturers. As a complex phenomenon, many parameters can be studied and optimized for squeal noise during its design phase (Denimal et al, 2018; Nechak et al, 2018). We focus on shape of the system, which has not been widely considered.

Our proposed applicative example is the simple disc-pad system as a brake model. The system consists of a disc defined as a solid annulus geometry fixed at the inner cylindrical face. The pad is constrained to be in contact with the disc and with additional constraints to avoid any rigid body modes in the system. The description of pad shapes in optimization is detailed in section 4. As an initial approximation, the simple system allows to make interesting studies by avoiding the complexity of boundary conditions present in a real braking systems.

2.1 Isogeometric formulation

The Isogeometric approach was developed to merge design and analysis descriptions of a geometry through a same class of basis functions (Bazilevs et al,

2010; Hughes et al, 2005), where the basis functions used for defining a geometry is also used for approximation of solutions in the context of finite element approach. The parameterization of a geometry defined with mapping from a parametric space is achieved through NURBS with a set of basis functions, where its subsequent analysis-suitable parameterization which is usually more refined with a new set of increased number of basis functions is defined with in the same parametric space.

We provide a concise explanation for the finite element approach to approximation and NURBS basis functions in annex section 10.1.

2.2 Continuum description to Isogeometric discretization

The continuum description of the dynamics around a quasi-static fixed point $\tilde{u}$ for the perturbation u can be defined as follows,

$$\begin{aligned} \rho_{\mathrm{k}}\ddot{u}_{\mathrm{k}} + \nabla.\boldsymbol{\sigma}_{\mathrm{k}}(u_{\mathrm{k}}) = 0 \quad \text{in} \quad \Omega_{\mathrm{k}} \\ u_{\mathrm{k}} = 0 \quad \text{on} \quad \Gamma_{D,\mathrm{k}} \subset \partial\Omega_{\mathrm{k}} \\ \boldsymbol{\sigma}_{\mathrm{k}}(u_{\mathrm{k}}).\hat{v}_n = F_{c_{\mathrm{k}}}, \quad \boldsymbol{\sigma}_{\mathrm{k}}(u_{\mathrm{k}}).\hat{v}_t = F_{f_{\mathrm{k}}} \quad \text{on} \quad \Gamma_{c,\mathrm{k}} \subset \partial\Omega_{\mathrm{k}} \end{aligned} \tag{1}$$

where $\boldsymbol{\sigma}(u)$ represents constitutive equations as a function of displacement vector $u : \Omega \rightarrow \mathrm{R}^3$ in an infinitesimal volume with isotropic material properties and strain tensor defined by infinitesimal strain theory, $\partial\Omega$ defines the boundary of Ω and, $\hat{v}_n$ and $\hat{v}_t$ define a normal unit vector and a tangential unit vector respectively on $\partial\Omega$. The subscript k distinguishes the domains in contact.

The parameterization of the domain Ω_{k} as an initial geometric description is defined through NURBS as X_{k}, $X : \hat{\Omega} \rightarrow \Omega$ which defines the mapping from the parametric domain $\hat{\Omega}$ to the physical domain Ω – for simplicity, we consider the parameterization of the domain Ω_{k} through a single patch. The analysis-suitable parameterization of $X \rightarrow \overline{X}$ is achieved through one or several of the refinement methods (h, p and k), where the refinement of X_{k} is defined as $\overline{X}_{\mathrm{k}}$ to take in to account of the modified knot vectors and additional control points – more on parameterization and refinement for the disc-pad system domain is discussed in Section 4.2 .

The Isogeometric approach for approximation of the solution u_{k} is achieved through the same NURBS bases $R_{i,j,k}$ which was used to achieve an analysis-suitable parametrization $\overline{X}$, where $R_{i,j,k}$ defines trivariate spline bases for the approximation of u. And in abstract sense, the bases $R_{i,j,k}$ in parametric space is transformed to the bases $\varphi_{i,j,k}$ in physical space using the push-forward operator $\circ$, where the bases φ_{i} for approximation is defined with the property $\varphi_{\mathrm{i}} = \varphi_{i,j,k} = R_{i,j,k} \circ \overline{X}^{-1}$. Hence, the approximation of a field variable in

Ω_k is defined through $\varphi_{\mathrm{k,i}}, \forall \mathrm{i}$, spanning the finite dimensional function space $\Phi_\mathrm{k} := H^1_{0,\Gamma_{D,\mathrm{k}}}(\Omega_\mathrm{k})$. With the approximation of $u_\mathrm{k} \in \Phi_\mathrm{k}$ as $\overline{u}_\mathrm{k} \approx u_\mathrm{k}$, and with the application of Galerkin's method to Eq. (1) leads to following the weak form:

$$\int_{\Omega_\mathrm{k}} \rho_\mathrm{k} \ddot{\overline{u}}_\mathrm{k}.\varphi_{\mathrm{k,i}}\ d\Omega_\mathrm{k} + \int_{\Omega_\mathrm{k}} \nabla.\boldsymbol{\sigma}_\mathrm{k}(\overline{u}_\mathrm{k}).\varphi_{\mathrm{k,i}}\ d\Omega_\mathrm{k} = 0 \qquad \forall \varphi_{\mathrm{k,i}} \in \Phi_\mathrm{k} \tag{2}$$

The above weak form through expansion of the term $\nabla.\boldsymbol{\sigma}_\mathrm{k}(\overline{u}_\mathrm{k}).\varphi_{\mathrm{k,i}}$ by Green's identity for $\mathrm{n_k}$ domains in contact, is given as

$$\sum_{\mathrm{k}=1}^{\mathrm{n_k}} \Bigg\{ \int_{\Omega_\mathrm{k}} \rho_\mathrm{k} \ddot{\overline{u}}_\mathrm{k}.\varphi_{\mathrm{k,i}}\ d\Omega_\mathrm{k} + \int_{\Omega_\mathrm{k}} \boldsymbol{\sigma}_\mathrm{k}(\overline{u}_\mathrm{k}) : \nabla\varphi_{\mathrm{k,i}}\ d\Omega_\mathrm{k} - \underbrace{\int_{\Gamma_{c_\mathrm{k}}} F_{c_\mathrm{k}}.\varphi_{\mathrm{k,i}}\ d\Gamma_{c_\mathrm{k}}}_{\langle F_{c_\mathrm{k}}, \varphi_{\mathrm{k,i}} \rangle} - \underbrace{\int_{\Gamma_{c_\mathrm{k}}} F_{f_\mathrm{k}}.\varphi_{\mathrm{k,i}}\ d\Gamma_{c_\mathrm{k}}}_{\langle F_{f_\mathrm{k}}, \varphi_{\mathrm{k,i}} \rangle} \Bigg\} = 0 \qquad \forall \varphi_{\mathrm{k,i}} \in \Phi_\mathrm{k} \tag{3}$$

where the inner products $\langle F_{c_k}, \varphi_{k,i} \rangle$ and $\langle F_{f_k}, \varphi_{k,i} \rangle$ define the weak form for contact and friction respectively, which are defined independently on their respective domains. When $F_{c_\mathrm{k}} = F_{f_\mathrm{k}} = 0$, the classical expression for dynamics can be deduced independently for the domain Ω_k as

$$\mathbf{M}_\mathrm{k}\ddot{U}_\mathrm{k} + \mathbf{K}_\mathrm{k}U_\mathrm{k} = 0 \tag{4}$$

where $\mathbf{M}$ and $\mathbf{K}$ represent the classical mass and stiffness matrices. The definition of traction forces require a system level description of the disc-pad ensemble, which is briefed in the next sub-section.

2.3 Contact and friction formulation to model flutter-type dynamic instability

We provide a brief explanation for the contact and friction formulation defined through a holonomic constraint and a follower force model, and the characteristics of the matrices $\mathbf{K}^c$ and $\mathbf{K}^f$ obtained through such definitions. The definition of a follower force can be largely said as implicit for modelling flutter-type dynamic instability. In the case of flutter-type instability arising from friction between solids, the main hypothesis for CEA is that the onset of instability occurs at perturbations very close to the quasi-static state such that the non-linearities from contact separation can be effectively ignored and hence, the friction phenomenon at the perturbed states can be modelled with the contact conditions of the quasi-static state purely by displacement dependent forces which are also known as follower forces. This defines the system

as holonomic and autonomous since the nature of the follower forces depend only on generalized coordinates without explicit time dependence. Further, we expand the weak formulation of contact and friction. More on the topic of perturbations around a quasi-static state can be found in (Martins et al, 1999).

Even though classically, the constraints are defined independently for contact and friction, the follower force model for friction can be modelled without explicit definition of constraints for friction, but as a direct extension of the contact constraint to be defined. Hence, with the above considered hypothesis, the constraint to model contact and friction is given as holonomic rather than in a conventional sense which are given through inequalities such as Signorini conditions for contact and Coulomb-Amontons law for friction.

We introduce the contact and friction forces F_c and F_f as externally applied traction forces independently on their respective domains. Though the inner products $\langle F_{c_k}, \varphi_{\mathrm{k,i}} \rangle$ and $\langle F_{f_k}, \varphi_{\mathrm{k,i}} \rangle$ are of the linear form, the definition of the forces F_c and F_f through the constraint, with discretization lead to bilinearity. For simplicity, we give a concise explanation for arriving at F_c and F_f through an example by considering two domains Ω_a and Ω_b in contact, with the derivation of traction forces on Ω_a. The holonomic constraint for the domains in contact at $\Gamma_{c_a} : \Omega_a \cap \Omega_b$ is given as

$$(u_a - u_b).\hat{v}_n = 0 \quad \text{on} \quad \Gamma_{c_a} \tag{5}$$

where the holonomic constraint characterizes the property of no contact separation for the perturbed displacement field. The kinematic relation for the contact is satisfied through the outward normal projection $\hat{v}_n$ from Γ_c of one of the domains to the other. The constraint is enforced by penalty method where the degree of violation of the constraint is penalized by a factor $\boldsymbol{\gamma}$, which can be interpreted as normal contact stress F_{c_a} on Γ_{c_a}, given as

$$F_{c_a} = \boldsymbol{\gamma}((u_a - u_b).\hat{v}_n).\hat{v}_n \quad \text{on} \quad \Gamma_{c_a} \tag{6}$$

The value of $\boldsymbol{\gamma}$ is usually determined experimentally rather than conventionally where it is sufficient to be high enough concerning numerical stability. The tangential stress F_{f_a} by friction is defined through follower force model (Hibbit, 1979) with Coulomb's law, where the stress is prescribed in the direction of motion $\hat{v}_t$ which is tangent to the contact stress F_{c_a}, given as follows

$$F_{f_a} = \boldsymbol{\mu}(\boldsymbol{\gamma}(u_a - u_b).\hat{v}_n).\hat{v}_t \quad \text{on} \quad \Gamma_{c_a} \tag{7}$$

where $\boldsymbol{\mu}$ is the coefficient of friction. Through discretization of the displacement field u –similar to arriving at equation (2)– for the continuum description of F_{c_a} and F_{f_a} as $\overline{F}_{c_a}$ and $\overline{F}_{f_a}$ respectively, the weak formulations $\langle \overline{F}_{c_a}, \varphi_{k,i} \rangle$ and $\langle \overline{F}_{f_a}, \varphi_{k,i} \rangle$ can be expressed as

$$\langle \overline{F}_{c_a}, \varphi_{a,\mathrm{i}} \rangle = \int_{\Gamma_{c_a}} (\boldsymbol{\gamma}(\overline{u}_a - \overline{u}_b).\hat{v}_n).\varphi_{a,\mathrm{i}}.\hat{v}_n \; d\Gamma_{c_a} \qquad \forall \varphi_{a,\mathrm{i} \in \Gamma_{c_a}} \in \Phi_a \tag{8}$$

$$\langle \overline{F}_{f_a}, \varphi_{a,\mathrm{i}} \rangle = \int_{\Gamma_{c_a}} \boldsymbol{\mu}(\boldsymbol{\gamma}(\overline{u}_a - \overline{u}_b).\hat{v}_n).\varphi_{a,\mathrm{i}}.\hat{v}_t \; d\Gamma_{c_a} \qquad \forall \varphi_{a,\mathrm{i}\in\Gamma_{c_a}} \in \Phi_a \tag{9}$$

Similarly, the traction forces on Γ_{c_b} can be defined as $F_{c_b} = -F_{c_a}$ and $F_{f_b} = -F_{f_a}$. There are several methods through which the above integrals can be solved which we do not focus here, but the following reference can help with special focus on Isogeometric approach for contact (De Lorenzis et al, 2014).

The above formulations can be extended for the n_{k} domains, where the bilinear form for $\sum_{\mathrm{k}}^{n_{\mathrm{k}}} \langle \overline{F}_{c_{\mathrm{k}}}, \varphi_{\mathrm{k,i}} \rangle$ and $\sum_{\mathrm{k}}^{n_{\mathrm{k}}} \langle \overline{F}_{f_{\mathrm{k}}}, \varphi_{\mathrm{k,i}} \rangle$ define the matrices $\mathbf{K}^c$ and $\mathbf{K}^f$ for contact and friction respectively for the system. The matrix $\mathbf{K}^c$ is symmetric and hence, $\mathbf{K} + \mathbf{K}^c$ leads to a positive definite matrix where the eigenvalues given by CEA mostly define a stable system – with $\mathbf{K}$ being the stiffness matrix of the system for the n_{k} domains. While, the matrix $\mathbf{K}^f$ is non- symmetric from the non-conservative model of friction defined as follower force. Hence, the matrix $\mathbf{K}+\mathbf{K}^c+\mathbf{K}^f$ in a sense models material stiffness along with the augmented stiffness to model contact and friction characteristics for perturbations around a quasi-static state.

3 Stability: analysis and description of criterion

Considering brake squeal, it is well known that squeal noise can be linked to the modal behaviour of a system (Soobbarayen et al, 2013, 2014). Hence, it seems quite obvious to build an optimization criterion to characterize squeal noise on the importance of these unstable modes which defines flutter-type instability. The definition of such a criterion called "Stability criterion" is defined through CEA, detailed in this section as follows.

3.1 Complex-eigenvalue analysis (CEA)

From the expansion of Eq. (3), with the weak form from the Eqs. (8) and (9), the matrix form of the dynamics with contact and friction definition for the considered disc-pad system "$d-p$" is defined as

$$\mathbf{M}_{d-p}\ddot{U}_{d-p} + (\mathbf{K}_{d-p} + \mathbf{K}^c_{d-p} + \mathbf{K}^f_{d-p})U_{d-p} = 0 \tag{10}$$

With the assumption of a feasible solution of the form $Ye^{\lambda t}$ for U, the characteristics eigenvalue problem can be defined as

$$(\lambda^2\mathbf{M}_{d-p} + (\mathbf{K}_{d-p} + \mathbf{K}^c_{d-p} + \mathbf{K}^f_{d-p}))Y = 0 \tag{11}$$

where λ and Y represents an eigenvalue and an eigenvector respectively for the system. It is evident that the value of λ determines the outcome of the

solution U as the power of exponent, hence the stability of the steady-state system. With the definition of non-conservative forces in the model, the value of λ may take a complex value, where the imaginary part $\Im(\lambda)$ defines the oscillatory behaviour of the dynamical system and the real part $\Re(\lambda)$ characterizes the stability of the steady-state.

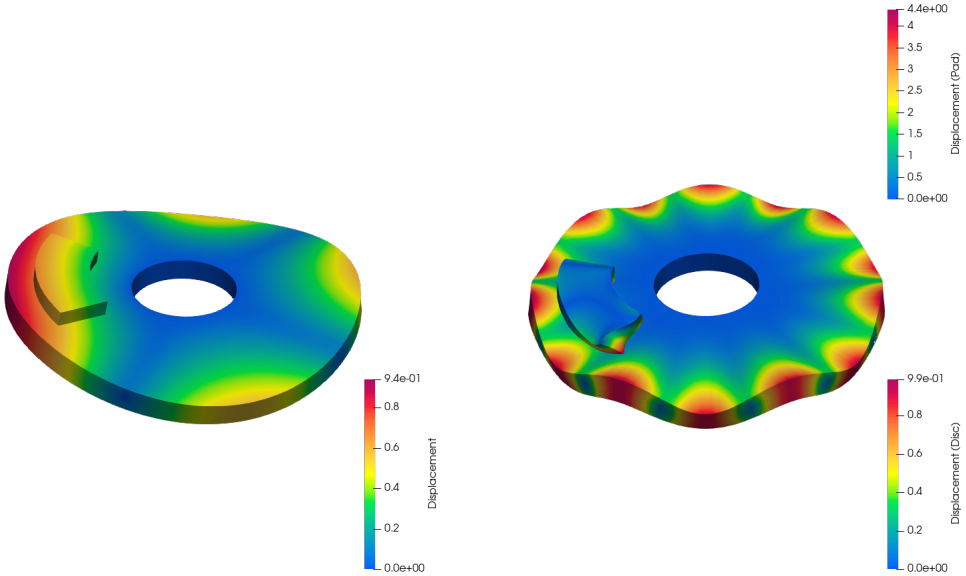
Hence, depending on the value of $\Re(\lambda)$, the stability of the system characterized by a mode of eigenfrequency λ can be categorized into one of the following:

- $\Re(\lambda) > 0$ unstable system, which can indicate the presence of Squeal noise.
- $\Re(\lambda) < 0$ stable system
- $\Re(\lambda) = 0$ neutrally stable. This is usually not a physical steady-state behavior.

In our system, the coefficient of friction $\boldsymbol{\mu}$ is the driving parameter which determines the stability, such that increase in the value of $\boldsymbol{\mu}$ can drive the system from stable to unstable behaviour (Bigoni and Noselli, 2011; Hoffmann et al, 2002; Sinou and Jézéquel, 2007). This type of instability is characterized by Hopf-bifurcation where the presence of a limit cycle is determined by the occurrence of a pair of eigenvalues as $\pm\Re(\widehat{\lambda}_{\boldsymbol{\mu}})+\Im(\widehat{\lambda}_{\boldsymbol{\mu}})$, with $\widehat{\lambda}_{\boldsymbol{\mu}}$ being an eigenfrequency of two modes undergoing coalescence at $\boldsymbol{\mu}$. The presence of this type of limit cycle defines a self-excitation behaviour leading to flutter-type instability. This is physically understood for our system as follows, the bifurcation leads to two modes with $\pm\Re(\widehat{\lambda}_{\boldsymbol{\mu}_o})$ for a same frequency $\Im(\widehat{\lambda}_{\boldsymbol{\mu}_o})$ at the point of bifurcation $\boldsymbol{\mu}_o$ and when $\boldsymbol{\mu} > \boldsymbol{\mu}_o$, the two modes $+\Re(\widehat{\lambda}_{\boldsymbol{\mu}>\boldsymbol{\mu}_o}) + \Im(\widehat{\lambda}_{\boldsymbol{\mu}>\boldsymbol{\mu}_o})$ and $-\Re(\widehat{\lambda}_{\boldsymbol{\mu}>\boldsymbol{\mu}_o}) + \Im(\widehat{\lambda}_{\boldsymbol{\mu}>\boldsymbol{\mu}_o})$ characterize stable and unstable behaviours respectively, leading to mode-coalescence phenomenon. The presence of this type of instability is correlated to the magnitude of brake squeal noise proportional to $+\Re(\widehat{\lambda}_{\boldsymbol{\mu}})$ as a contribution to squeal noise by the unstable mode of frequency $\widehat{\lambda}_{\boldsymbol{\mu}}$ among the other modes in $\Lambda_{\boldsymbol{\mu}}$ which is a set of eigenvalues of the $d-p$ system. Hence, we define a stability criterion C_s through the magnitude of such type of unstable modes $+\Re(\widehat{\lambda}_{\boldsymbol{\mu}})$ in $\Lambda_{\boldsymbol{\mu}}$ for all the values of $\boldsymbol{\mu}$, detailed in 3.2.

Example of mode shapes for the disc-pad system obtained through CEA are shown in the Figures 1 and 2.

From post-processing, the following observations were made: normally at low frequencies, the mode shape of the pad follows the shape of the disc with correspondence in magnitude for displacement field at the contact region. While at higher frequencies, the behaviour is complicated to understand, but relatively larger difference in magnitude for displacement between the disc and pad were observed. We also remind that the results are subjective and depend on the value of $\boldsymbol{\gamma}$ which is usually determined experimentally, where we used an arbitrary suitable value considering numerical stability and convergence. Further, the unstable modes lead to definition of eigenvectors in complex-plane



(a) Mode no. 5, Frequency:1476 Hz (b) Mode no. 26, Frequency:9245 Hz

Fig. 1: Example of disc-pad stable modes

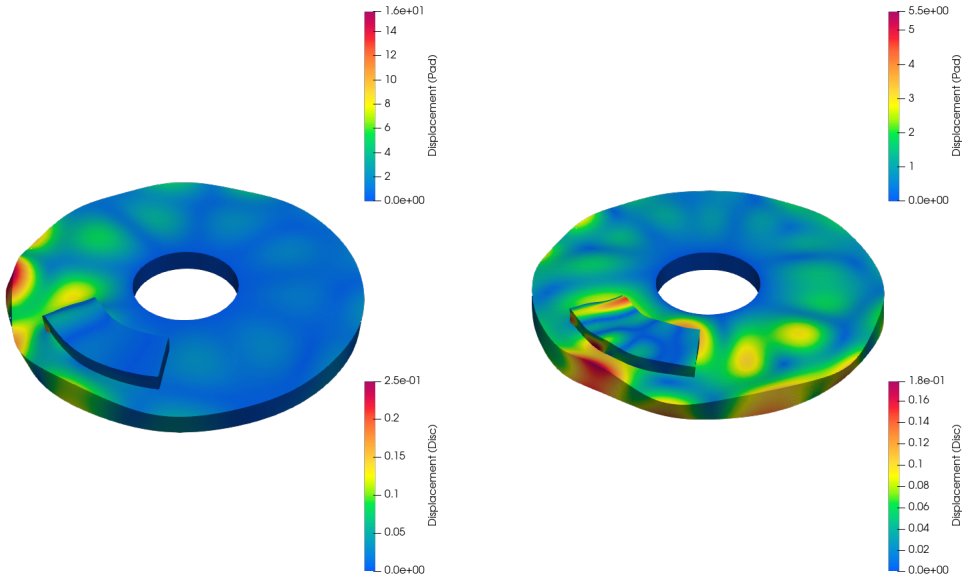
for displacement field to understand its behaviour, which was not considered for representation in Figure 2. For an intuition, a complex eigenvalue, along with the natural frequency defines the phase-lag in the displacement field for an eigenvector and hence the stable equilibrium position of the displacement field for an eigenvector is never achieved simultaneously.

3.2 Stability criterion

In the context of shape optimization, the idea is to reduce the influence of the friction parameter $\boldsymbol{\mu}$ in determining the shape of the system to characterize instability, since the parameter $\boldsymbol{\mu}$ is mostly uncertain in the real world and also the instability could be easily avoided at lower values of $\boldsymbol{\mu}$. Hence to define a criterion which characterizes the instability for a geometric shape X of the system independent of the coefficient of friction, we define the criterion as follows

$$C_s(X) = \int_{\boldsymbol{\mu}} \max\{\Re(\Lambda(X, \boldsymbol{\mu}))\} \tag{12}$$

which estimates the maximum of the real part of the eigenvalues of the $d-p$ system at a given value of $\boldsymbol{\mu}$, integrated over all the values of $\boldsymbol{\mu}$.



(a) Mode no. 45, Frequency:12236 Hz (b) Mode no. 82, Frequency:17259 Hz

Fig. 2: Example of disc-pad unstable modes, where the displacement field is considered only for real-part of the eigenvector

When defined through numerical integration, the criterion demands evaluation at several values of $\boldsymbol{\mu}$ and hence is computationally expensive. But this can be evaded through taking advantage of parallel computation with reduced dynamical models using Craig-Bampton reduction (Bampton and Craig, 1968), where the computation of the matrices in physical coordinates followed by dynamic model reduction are achieved on the main core. The only varying parameter in the parallel cores is $\boldsymbol{\mu}$ and hence, the calculation of the reduced friction matrix $\hat{\mathbf{K}}^f_{d-p}$ matrix –i.e., the reduced matrix represented in the Craig-Bampton coordinates– is defined with $\boldsymbol{\mu} = 1$ on the main core and the parallelization is defined for distinct values of $\boldsymbol{\mu}$ for $\boldsymbol{\mu}\hat{\mathbf{K}}^f_{d-p}$. This means that in addition to the definition of $\boldsymbol{\mu}\hat{\mathbf{K}}^f_{d-p}$, the computation on the parallel cores is limited to solving the eigenvalue problem (11) with the reduced dynamical model for evaluation of the criterion (12), which makes it computationally efficient.

4 Optimization setting

In this section, we detail the shape optimization defined through NURBS parameterization of shapes for the pad, with its associated constraints and the objectives for the optimization. We also provide a short description of the pa-

rameterization and refinement strategy for the disc-pad system domain.

4.1 Shape optimization

The optimization is defined for the boundary $\partial\Gamma_{c_{Pad}}$ of the planar surface of the pad $\Gamma_{c_{Pad}}$ which is in contact with the disc, where the thickness of the pad and the design parameters of the disc are constant. The geometry of $\Gamma_{c_{Pad}}$ is parameterized through NURBS as

$$X_s^{Pad}(\xi,\eta) = \sum_{i=0}^{n}\sum_{j=0}^{m} R_{i,p}(\xi)R_{j,q}(\eta)\mathrm{P}_{i,j} \tag{13}$$

Hence, in our setting, the shape optimization is defined for the shape of the NURBS curves $X_c^{(1)}(s), X_c^{(2)}(t), X_c^{(3)}(u)$and $X_c^{(4)}(v)$ which parameterizes $\partial\Gamma_{c_{Pad}}$ that encloses the surface $X_s^{Pad}(\xi,\eta)$, as shown in Figure 3, where the curves can be expressed as

$$\begin{aligned} X_s^{Pad}(\xi,\eta|\xi=0) &= X_c^{(1)}(s)\\ X_s^{Pad}(\xi,\eta|\eta=0) &= X_c^{(2)}(t)\\ X_s^{Pad}(\xi,\eta|\xi=1) &= X_c^{(3)}(u)\\ X_s^{Pad}(\xi,\eta|\eta=1) &= X_c^{(4)}(v) \end{aligned} \tag{14}$$

The parameterization $X_s^{Pad}(\xi,\eta)$ from the above four curves defined for optimization was realized through discrete Coon's patch method, given by the following relation

$$\begin{aligned} X_s^{Pad}(\xi,\eta) = X_c^{(1)}(s)(1-\xi) + X_c^{(3)}(u)(\xi) + X_c^{(2)}(t)(1-\eta) + X_c^{(4)}(v)(\eta)\\ - X_c^{(1)}(0)(1-\xi)(1-\eta) - X_c^{(1)}(1)(1-\xi)(\eta)\\ - X_c^{(3)}(0)(\xi)(1-\eta) - X_c^{(3)}(1)(\xi)(\eta) \end{aligned} \tag{15}$$

The Coon's patch method is an explicit linear method and hence computationally efficient in realizing parameterization, but the method doesn't guarantee injective mapping. In our experience, the shapes realized through Coon's patch method that doesn't satisfy injective mapping are largely too conceptual for pad shapes and hence, given the complexity of realizing parameterization for such shapes, we stick with only the shapes realized through Coon's patch method for which injective parameterization exists.

For simplicity owing to the initial phase of our analysis, in order to limit the parameters in optimization, we restricted the degree of each curve to 2 and hence, this leads to the surface $X_s^{Pad}(\xi,\eta)$ with the property $p = q = 2$, and each curve is defined only through three control points which are just enough to define a curve of degree 2. Then for shape optimization, we constrained

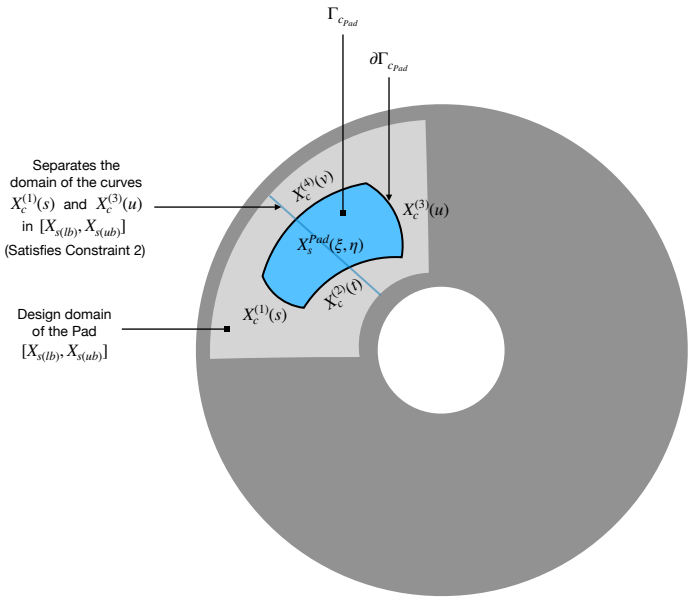


Fig. 3: An illustration describing the parameterization of $\Gamma_{c_{Pad}}$ and $\partial\Gamma_{c_{Pad}}$.

the mid-control point of each curve segment X_c to move only normal to the line joining the end control points of their respective curve segment, such that the initial configuration of a curve aligns with the line joining its end control points and that the mid-control points are also aligned, given as Constraint 1. Furthermore, the optimization is defined only for the position of the control points $\mathrm{P}_{i,j}$ for fixed weights (32), i.e, we considered the optimization of the NURBS geometry only through affine transformation without considering projective transformation.

The injective parameterization can be achieved if the determinant of the Jacobian that expresses a transformation $X : \hat{\Omega} \rightarrow \Omega$ is positive in all $\hat{\Omega}$, where in our case is tested for $X_s^{Pad}(\xi, \eta)$. The set of constraints for testing this condition was realized geometrically in our case where the mid-control points are constrained (Constraint 1), and is given as follows

Constraint set 2:

$$\{X_c^{(1)}(s) \cap X_c^{(2)}(t)\} \cup \{X_c^{(3)}(u) \cap X_c^{(4)}(v)\} \cup \{X_c^{(1)}(s) \cap X_c^{(3)}(u)\} \cup$$
$$\{X_c^{(1)}(s) \cap X_c^{(4)}(v)\} \cup \{X_c^{(2)}(t) \cap X_c^{(3)}(u)\} \cup \{X_c^{(2)}(t) \cap X_c^{(4)}(v)\} = \emptyset$$
$$\forall s, t, u, v \in (0, 1)$$

$$\{X_c^{(1)}(0)(1-\xi) + X_c^{(3)}(0)(\xi)\} \cap \{X_c^{(1)}(0+\Delta s)(1-\xi) + X_c^{(3)}(0+\Delta u)(\xi)\} = \emptyset$$
$$\{X_c^{(1)}(1)(1-\xi) + X_c^{(3)}(1)(\xi)\} \cap \{X_c^{(1)}(1-\Delta s)(1-\xi) + X_c^{(3)}(1-\Delta u)(\xi)\} = \emptyset$$
$$\{X_c^{(2)}(0)(1-\eta) + X_c^{(4)}(0)(\eta)\} \cap \{X_c^{(2)}(0+\Delta t)(1-\eta) + X_c^{(4)}(0+\Delta v)(\eta)\} = \emptyset$$
$$\{X_c^{(2)}(1)(1-\eta) + X_c^{(4)}(1)(\eta)\} \cap \{X_c^{(2)}(1-\Delta t)(1-\eta) + X_c^{(4)}(1-\Delta v)(\eta)\} = \emptyset$$
$$\forall \xi, \eta \in [0, 1] \quad (16)$$

where Δ represents an arbitrary small variation for the given parameters. The first constraint avoids the intersection between the curves except for the end points, where no self intersection of the curves is implicitly possible with the constraints on the mid-control points. Satisfying the first constraint which guarantees a fixed topology does not assure injective parameterization through Coon's patch method, for which the last set of four constraints are necessary.

Further, the definition of the pad surface to be with in the bounds of the disc surface is given through a box constraint as follows

Constraint set 3:

$$(X_{s(lb)} \leq X_s^{Pad}(\xi, \eta) \leq X_{s(ub)}) : \{[X_{c(lb)}^{(i)}, X_{c(ub)}^{(i)}]\} \cap \{[X_{c(lb)}^{(j)}, X_{c(ub)}^{(j)}]\} = \emptyset \quad (17)$$

where the choice of $X_{s(lb)}$ and $X_{s(ub)}$ depends on the design choice for the domain of the disc to be in contact with the pad. Further, the box constraint is adapted to limit the redundancies in geometric description i.e, to limit the scope for a given shape to be represented through more than one way with in the same design space. To avoid this type of redundancy, we restricted the domain through box constraints for at least two curves $X_c^{(i)}(.)$ and $X_c^{(j)}(.)$ of the four curves, such that the intersection of their domains is a null set. This leads to restriction of the design space with compromise for reducing the redundancies. Hence, we avoided some of the redundancies such that the restricted design space has lower chance for more meaningful designs. This maybe an interesting anomaly to investigate since the redundancies may lead to larger design space with more severe multi-modality for a function with shape variables.

We further impose an inequality constraint in order to avoid designs with smaller contact surface, given as

Constraint 4:

$$Area(X_s^{Pad}(\xi,\eta)) \geq A_{min} \tag{18}$$

where $Area(X_s^{Pad}(\xi,\eta)) : \int_\xi \int_\eta |\frac{\partial X_s^{Pad}}{\partial \xi} \times \frac{\partial X_s^{Pad}}{\partial \eta}| d\xi d\eta$ and the choice of A_{min} depends on the minimum contact surface area that is required on the Pareto-front, since the maximization of the contact surface is to be defined as one of the objectives.

The definition of the shape of $X_s^{Pad}(\xi,\eta)$ through this strategy means that there is no requisite for a reference configuration to define the optimization, but instead the pad shapes are created through random generation of curves with C^0 continuity between them. We assume that this restricts bias to any particular configuration and hence encouraging more randomness in defining a meaningful geometry. This highly restricts the use of gradient-based approaches for optimization, since this can cause apparent discontinuities induced by the constraints. Some of the limitations can also be attributed to lack of exploring classical shapes such as the annulus sector pad shapes in our application even though such shapes are already a subset of the the design space defined, which can be otherwise defined through a reference configuration, and also the randomness lead to more probability of failure for the constraints.

Finally, the objectives for the Multi-objective optimization can be posed as optimization of the following functionals:

- Objective 1: $min\, C_s(X_s^{Pad}(\xi,\eta)) \,|\, \Im(\Lambda(X_s^{Pad}(\xi,\eta))) \in [10KHz, 13KHz]$
- Objective 2: $max\, Area(X_s^{Pad}(\xi,\eta))$

where the functional optimization is defined over space of NURBS functions. As aforementioned, we fixed the degrees and the number of control points of the NURBS surface $X_s^{Pad}(\xi,\eta)$ and hence, the optimization is restricted to a fixed number of control points $\mathrm{P}_{i,j}$.

4.2 Isogeometric parameterization and refinement strategies for the disc-pad system domain with contact considerations

For the following, we do not focus on the mesh sensitivity for CEA or the stability criterion, but instead the below refinement strategies can be seen as to realise the classical mesh refinement considerations for a contact problem, where more elements are typically defined on Γ_c and at the vicinity of $\partial\Gamma_c$ to capture more accurately the contact characteristics and the strong solution gradient. This is especially more challenging with NURBS parameterization for local refinement, hence we expose here some strategies to achieve local refinement for such applications. From our observations, the refinement at the Γ_c and around $\partial\Gamma_c$ seems to effect the results of CEA and converges with

sufficient refinement, but a more qualitative assessment of the sensitivity has not been developed here.

The planar parameterization $X_s^{Pad}(\xi,\eta)$ can be easily extended to define the volume $X_v^{Pad}(\xi,\eta,\zeta)$ considering the thickness of the pad, through the tensor product definition (34), given a NURBS line along the thickness to achieve the parameterization of Ω_{Pad}. For the disc, we used multi-patch parameterization for Ω_{Disc} as $X_v^{Disc} := X_v^{Disc(P1)}(\xi,\eta,\zeta) \cup X_v^{Disc(P2)}(\xi,\eta,\zeta)$ to achieve local refinement on $\Gamma_{c_{Disc}}$. The initial surface parameterization for the disc patches: $X_s^{Disc(P1)}(\xi,\eta)$ and $X_s^{Disc(P2)}(\xi,\eta)$ can be achieved through the concept of revolved surfaces which is pretty straightforward, detailed in (Piegl and Tiller, 1996), which assures robust injective mapping since the curve to be revolved is a straight line perpendicular to the disc axis. The planar parameterizations $X_s^{Disc(P1)}(\xi,\eta)$ and $X_s^{Disc(P2)}(\xi,\eta)$ can be extended to $X_v^{Disc(P1)}(\xi,\eta,\zeta)$ and $X_v^{Disc(P2)}(\xi,\eta,\zeta)$ respectively similar to achieving $X_v^{Pad}(\xi,\eta,\zeta)$

For any refinement, the space for parameterization remains the same i.e, $(\xi,\eta,\zeta) : [0,1]^3$ and the refinement is defined only through manipulation or addition of knots and control point to achieve an analysis-suitable parameterization. After an analysis-suitable parameterization, to take in to account of the additional control points and the manipulated knot vectors, X_v^{Pad} and X_v^{Disc} can be expressed as $\overline{X}_v^{Pad}$ and $\overline{X}_v^{Disc}$. Hence, the NURBS bases associated with $\overline{X}_v^{Pad}$ and $\overline{X}_v^{Disc}$ are used to define the space for approximation in isogeometric approach, detailed in Section 2.2.

Normally across the boundary $\partial\Gamma_c$ of a contact domain Γ_c, there is drastic change in the solution gradient and hence, the parameterization needs special attention owing to the continuity of the NURBS bases. The tensor product property of the NURBS gives further challenge for local refinement which is usually desired on Γ_c. These challenges can be largely overcome by adaptation of NURBS bases to T-splines (Bazilevs et al, 2010) and THB-splines (Giannelli et al, 2012), but requires extensive adaptation. Hence, we defined a multi-patch parameterization strategy through collocation and projection of the properties defined on control points between two merging surfaces, which was simple and efficient for our application with fewer adaptation. Even though the considered multi-patch approach only considers C^0 solution continuity between the patches, the post-processing of the mode shapes show sufficient smoothness in displacement field across the patches for the disc, shown in Figure 4.

The multi-patch parameterization for Ω_{Disc} to break the tensor product definition of NURBS is shown in Figure 5, where one patch $\overline{X}_v^{Disc(P1)}$ contains the contact domain $\Gamma_{c_{Disc(P1)}}$ defined through a fine mesh by h-refinement and the other patch $\overline{X}_v^{Disc(P2)}$ with a relatively coarse mesh sufficient to capture

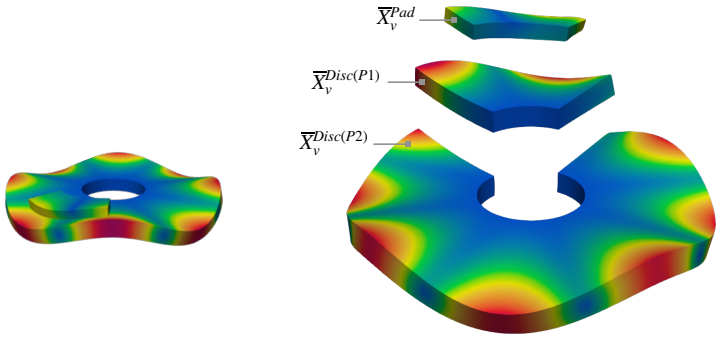


Fig. 4: Anatomy of parameterization for the disc-pad system with arbitrary dimensions, shown here for Mode 9, Frequency: 3630 Hz

the required dynamic properties. And different strategies were used to reduce the solution smoothness induced by the continuity of the NURBS approximation across the boundary $\partial\Gamma_{c_{Disc(P1)}}$ where typically strong solution gradient exists. For pad shapes where the knot lines on $\overline{X}_v^{Disc(P1)}$ can be aligned with $\partial\Gamma_{c_{Disc(P1)}}$, h-refinement can be used with finer refinement around $\partial\Gamma_{c_{Disc(P1)}}$, while the contact domain $\Gamma_{c_{Disc(P1)}}$ itself is discretized by h-refinement through a relatively coarse mesh compared to the refinement around $\partial\Gamma_{c_{Disc(P1)}}$, but finer than the rest of the domain. For pad shapes where the knot lines on $\overline{X}_v^{Disc(P1)}$ cannot be aligned with the boundary $\partial\Gamma_{c_{Disc(P1)}}$, we purely relied on h-refinement with much finer refinement. For the shape optimization, we used the later strategy due to random definition of shapes.

Further, in the optimization loop, the dimension of the disc patches were adapted for changes in $\Gamma_{c_{Disc(P1)}}$ to realize a more restrictive local refinement on $\Gamma_{c_{Disc(P1)}}$ and across $\partial\Gamma_{c_{Disc(P1)}}$. Hence, the knot vectors were also adapted to have uniform spacing of knots for a given region irrespective of the change in the dimension of the patches.

5 Optimization process

The stability criterion as previously defined does not allow gradient based optimization schemes and hence, we purely rely on metaheuristic approaches. The following section aims at describing a strategy for optimization using EGO for expensive black-box functions and its extension to MOO using a metaheuristic approach. Further, we describe a new strategy to improve resolution for the improvement to be defined with the Bayesian view for MOBO, with fewer adaptation to existing well-established methods and with focus on a

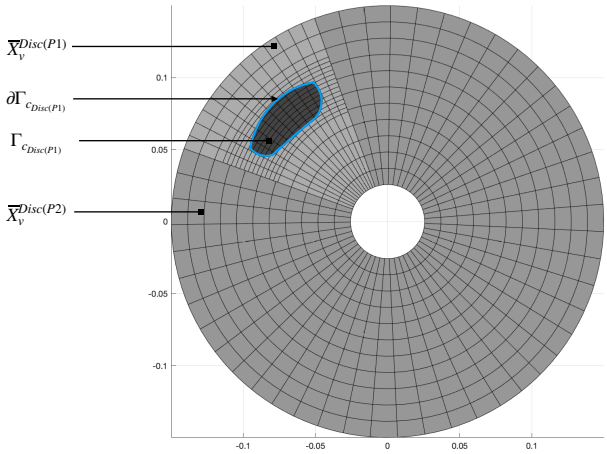


Fig. 5: Multi-patch parameterization of Ω_{Disc} as $\overline{X}_v^{Disc} := X_v^{Disc(P1)}(\xi,\eta,\zeta) \cup X_v^{Disc(P2)}(\xi,\eta,\zeta)$, with h-refinement at the contact region $\Gamma_{c_{Disc(P1)}}$.

simple and fast algorithm to implement.

Bayesian optimization is an effective strategy for optimising computationally expensive objective functions (Shahriari et al, 2016). The idea is based on Bayes rule where the prior knowledge $\mathcal{P}(\mathcal{H})$ of the hypothesis $\mathcal{H}$ and the likelihood of the evidence $\mathcal{E}$ given the hypothesis $\mathcal{P}(\mathcal{E}|\mathcal{H})$ are used to infer the posterior knowledge of the hypothesis given the evidence $\mathcal{P}(\mathcal{H}|\mathcal{E})$, where the proportionality is expressed as follows

$$\mathcal{P}(\mathcal{H}|\mathcal{E}) \propto \mathcal{P}(\mathcal{E}|\mathcal{H})\mathcal{P}(\mathcal{H}) \tag{19}$$

In our setting, the hypothesis $\mathcal{H}$ corresponds to a function $f(x)$ and the evidence $\mathcal{E}$ to $\mathcal{F} : \{f(x_1), f(x_2), \ldots, f(x_{\mathtt{n}})\}$ where the function was sampled on $\mathcal{D} : \{x_1, x_2, \ldots, x_{\mathtt{n}}\}$. The posterior knowledge can be then used to infer the optimum of the function where the expensive computational model needs to be evaluated. The new evaluation is then used to update the belief of the prior, and with the likelihood to infer a new posterior. The process is run subsequently for optimization with the expectation of reaching the global optimum for the function. The most common method to model the prior of a function is through Gaussian process, which also infers the posterior as Gaussian. This is more efficient since this presents the prediction and the uncertainty of the prediction, which provides a decisive knowledge to construct an acquisition function to sample more efficiently for optimization.

The modelling of prior, and the inference of posterior, for a function to be approximated takes the role of surrogate modelling commonly known as Gaussian process regression or Kriging. There are different context through which the Gaussian process regression could be presented owing to its diverse origins. We present in the context of function space view of Gaussian processes, where with relation (19), we defined the probability directly over the space of functions. The interested readers can refer to the following articles for more details (Rasmussen and Williams, 2005)(Forrester et al, 2006). Followed by, we present an acquisition function where we adapted Expected Improvement(EI) that defines EGO (Jones et al, 1998).

A Gaussian process can be viewed as distribution over a function contrary to a Gaussian distribution which is distribution over a random variable or in the case of a multi-variate gaussian distribution over random variables. Hence, any sample from a Gaussian process ($\mathcal{GP}$) is a function, where the distribution constitutes a mean function $\mu(x)$ and a covariance function $k(x, x^{'})$ given as follows

$$f(x) \approx \mathcal{GP}(\mu(x), k(x, x^{'})) \tag{20}$$

In other words, the existence of a function is treated in a mere probabilistic sense. The $\mathcal{GP}$ prior $\mathcal{P}(f(x))$ is defined for modelling $f(x)$ through a mean function which is usually a constant or a polynomial trend function. This can be seen as the deterministic part which captures the general trend of the function, while the covariance function $k(x, x^{'})$ models the stochastic trend which is the spatial correlation between any $f(x) - \mu(x)$ and $f(x') - \mu(x')$. The spatial correlation is modelled by hyperparameters which are the constants in a covariance function $cov(f(x) - \mu(x), f(x') - \mu(x'))$. To define the prior over $k(x, x^{'})$, the hyperparameters are estimated a priori, which is usually done by optimising the likelihood function for $arg\,max_{\theta}\ L(\mathcal{F}|\theta)$.

The conditioning of $\mathcal{P}(f(x))$ with $\mathcal{F}$ for the sampled arguments $\mathcal{D}$ as $\mathcal{P}(f(x)|\mathcal{D}, \mathcal{F}, \theta)$ results in a $\mathcal{GP}$ posterior which can be viewed in a finite dimensional sense as the posterior joint Gaussian distribution of

$\mathcal{P}(f(x_1^*)), \mathcal{P}(f(x_2^*)), \ldots, \mathcal{P}(f(x_{\mathtt{m}}^*))$ across rest of the function where its arguments $x_{\mathtt{i}}^*$ has not been sampled, i.e. $x_{\mathtt{i}}^* \notin \mathcal{D}$. While, the $\mathcal{P}(f(x_{\mathtt{i}}^*)|\mathcal{D}, \mathcal{F}, \theta)$ for any $x_{\mathtt{i}}^*$ can be obtained by marginalizing the posterior joint Gaussian distribution.

To move on from the abstractness of a $\mathcal{GP}$ to Gaussian distribution, the properties of Multi-variate Gaussian distributions allow to isolate a part of $\mathcal{GP}$ to define a joint Gaussian distribution of only the sampled arguments $\mathcal{D}$ and an argument x^* where we want to predict, where the joint distribution is expressed as

$$\begin{bmatrix} \mathcal{P}(f(x_1)) \\ \vdots \\ \mathcal{P}(f(x_n)) \\ \mathcal{P}(f(x_{\mathfrak{i}}^*)) \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mu(x_1) \\ \vdots \\ \mu(x_{\mathfrak{n}}) \\ \mu(x_{\mathfrak{i}}^*) \end{bmatrix}, \begin{bmatrix} k(x_1,x_1) & \dots & k(x_1,x_{\mathfrak{n}}) & k(x_1,x_{\mathfrak{i}}^*) \\ \vdots & \ddots & \vdots & \vdots \\ k(x_{\mathfrak{n}},x_1) & \dots & k(x_{\mathfrak{n}},x_{\mathfrak{n}}) & k(x_{\mathfrak{n}},x_{\mathfrak{i}}^*) \\ k(x_{\mathfrak{i}}^*,x_1) & \dots & k(x_{\mathfrak{i}}^*,x_{\mathfrak{n}}) & k(x_{\mathfrak{i}}^*,x_{\mathfrak{i}}^*) \end{bmatrix} \right) \tag{21}$$

The above joint distribution can be disassembled to define the mean and the covariance matrices for the sampled arguments and an argument to be predicted as

$$\mathcal{K} = \begin{bmatrix} k(x_1,x_1) & \dots & k(x_1,x_{\mathfrak{n}}) \\ \vdots & \ddots & \vdots \\ k(x_{\mathfrak{n}},x_1) & \dots & k(x_{\mathfrak{n}},x_{\mathfrak{n}}) \end{bmatrix}, \mathfrak{K}^* = \begin{bmatrix} k(x_1,x_{\mathfrak{i}}^*) \\ \vdots \\ k(x_{\mathfrak{n}},x_{\mathfrak{i}}^*) \end{bmatrix}, \mu(\mathcal{D}) = \begin{bmatrix} \mu(x_1) \\ \vdots \\ \mu(x_{\mathfrak{n}}) \end{bmatrix} \tag{22}$$

where $k(x_{\mathfrak{i}}, x_{\mathfrak{j}}) = cov(\mu(x_{\mathfrak{i}}) - f(x_{\mathfrak{i}}), \mu(x_{\mathfrak{j}}) - f(x_{\mathfrak{j}}))$ is defined through a covariance function with the prior estimation of the hyperparameters. We used Matern 5/2 kernel as the covariance function, since it is more versatile to characterize spatial correlation in our application and robust in the context of Cholesky decomposition.

The conditioning of the joint distribution Eq.21 defined by the prior knowledge of θ with the sampled data $\mathcal{D}$ and $\mathcal{F}$ gives the prediction for any x_i^* as follows

$$\mathcal{P}(f(x_i^*)|\mathcal{F},\mathcal{D},\theta) = \mathcal{N}(\underbrace{\mu(x_i^*) + \mathcal{K}^{-1}\mathfrak{K}^*(\mathcal{F}-\mu(\mathcal{D}))}_{\hat{\mu}(x_i^*)}, \underbrace{k(x_i^*,x_i^*) - \mathfrak{K}^*\mathcal{K}^{-1}\mathfrak{K}^{*T}}_{\hat{\sigma}^2(x_i^*)}) \tag{23}$$

The conditioning of the prior $\mathcal{P}(f(x_i^*))$ with the likelihood $\mathcal{F}$ to infer the posterior $\mathcal{P}(f(x_i^*)|\mathcal{F},\mathcal{D},\theta)$ gives the underlying Bayesian view $\mathcal{P}(\mathcal{H}|\mathcal{E})$.

5.1 Acquisition function definition

While we can infer prediction and the variance of a prediction, to sample, it is more efficient to quantify with a single scalar value which combines both to get a better inference. This is achieved through Acquisition function, where the goal for sampling can be defined based on the optimization context, for which a wide range of functions exist. In the scope of "exploration and exploitation" principle for optimization, we used Expected improvement (EI) criterion which is simply the expectation $E(I(x))$ of the improvement $I(x) = f(x)^+ - \hat{f}(x)$, where $\hat{f}(x) : \mathcal{N}(\hat{\mu}(x), \hat{\sigma}(x))$ as an outcome of the $\mathcal{GP}$ prediction and $f(x^+)$ being the reference value to be improved. The EI function is given as follows

$$EI(x) = \int_{-\infty}^{f(x^+)} I(x)\text{PDF}(\hat{f}(x))\ d\hat{f}(x) \tag{24}$$

where PDF(.): Probability density function. For a single objective Bayesian optimization, the problem leads to finding the $arg\,max_x EI(x)$, which is essentially the approach of EGO (Jones et al, 1998). For MOBO, the concept can be extended through pareto-optimal solutions, where one or several of the objectives could be maximising the EI of their respective functions, which was first proposed in (Jeong and Obayashi, 2006). More strategies and acquisition functions for MOBO, each with its respective advantages, are given in the following references (Knowles, 2006)(Luo et al, 2015).

5.2 Multiple reference values acquisition strategy

We extend the work of (Jeong and Obayashi, 2006), by defining multiple reference values with EI in the context of MOO to achieve MOBO. For a single objective optimization, we look for the improvement $EI(x|f(x^+))$ where $f(x^+)$ usually corresponds to the Utopian value. While, in the MOO context, the Utopian value known for a function to seek improvement can sometimes be too unrealistic on some parts of the objective space. This demands the improvement to be defined locally in an objective space through NDS. As well-known, a single objective EI can be extended to MOO through $EHVI$ where the improvement for a multi-variate gaussian prediction[1] is defined through the NDS in an objective space rather than a specific value from the NDS, which leads to definition of improvement in the hypervolume metric. But the evaluation of integration to define the $EHVI$ for a multi-variate gaussian prediction is cumbersome and nearly impossible for large number of objectives. Hence, we follow the approach similar to defining $max\ EI$ independently for $\mathcal{GP}$ meta-models for MOBO through non-dominated principle of genetic algorithms like Non-dominated sorting genetic algorithm-2 (NSGA-2), but with muliple reference values rather with a single reference value. This means that different regions of an objective space can constitute its own reference value as goal for improvement. This requires a precise definition for realistic goal/reference value to seek improvement, for which we define in a probablistic sense with the following explanation.

The proposed approach is generic, since it preserves the generic characteristics of NSGA-2 for MOO and the analytical evaluation of EI criterion, even though we make modifications for EI with the new approach to be efficient in MOO. Constraints can also be handled easily as part of NSGA-2, based on ranking for the degree of violation of the constraints or even completely removing the individuals violating the constraints from the population and

[1] Multi-variate Gaussian prediction is obtained through defining a joint distribution of independent univariate Gaussian predictions from $\mathcal{GP}$ meta-models in MOO

hence discouraging such individuals for future generations.

While a pareto front could be achieved with a single reference value $f(x^+)$ for $max\, EI(x|f(x^+))$ as one of the objectives, the resolution to define improvement is only efficient for a subset of the decision space depending on the $\mathcal{GP}$ prediction relative to $f(x^+)$. This can be seen through the following cases where $\underset{minf(x)}{EI}$ is defined for minimization of $f(x)$, as follows:

$$\begin{aligned} &\textbf{Case 1} : For\ \hat{\mu}(x) \pm \hat{\sigma}(x) \gg f(x^+),\ \underset{minf(x)}{EI}(x|f(x^+)) \approx 0 \\ &\textbf{Case 2} : For\ \hat{\mu}(x) \pm \hat{\sigma}(x) \ll f(x^+),\ \underset{minf(x)}{EI}(x|f(x^+)) \approx f(x^+) - \hat{\mu}(x) \end{aligned} \tag{25}$$

We here consider that $f(x)$ is an objective of a MOO and hence for MOBO, the optimization is defined for $\underset{minf(x)}{EI}$ in the place of $f(x)$. **Case 1** shows that the choice of the utopian value $f(x^+)$ to seek improvement for a subset of x in some regions of the objective space where $\hat{\mu}(x) \pm \hat{\sigma}(x) \gg f(x^+)$ can have no probabilistic chance for improvement and hence, to seek for improvement in this case can be said as being too greedy. While, this is insignificant in the context of single objective optimization where these measures can be ignored, the zero or the infinitesimal values provide less resolution for comparison to define NDS for MOO.

Case 2 shows that the measure of improvement is simply given by the distance between the prediction $\hat{\mu}(x)$ and the reference $f(x^+)$, which is only acceptable in cases where there is no realistic reference to seek improvement. This is possible when choosing the Nadir value as reference for improvement, which can be said as being too pessimistic to seek improvement. The pessimistic sense of seeking improvement can be seen as lack of risk for exploration in the regions of the objective space where $\hat{\mu}(x) \pm \hat{\sigma}(x) \ll f(x^+)$ and hence the absence of the uncertainty term $\hat{\sigma}(x)$ in evaluating $\underset{minf(x)}{EI}$. The above two cases show the limitation of using a single reference value for EI and hence, it is efficient to define improvement locally considering the NDS for MOO. The above cases are graphically shown in Figure 6.

Choosing the reference value $f(x^+)$ to define EI depending on where $\hat{f}(x)$ lies in the objective space makes the improvements hard to be compared since there is not a fixed reference value on the span of $\hat{f}(x)$ for the comparisons to be based on. Hence, we augment the EI to have a common frame of reference for comparison, where we consider the axes of the objective space itself as the frame of reference for comparison. The comparison is essential for defining NDS for optimization in the context of MOO. Hence, the augmentation of EI defined for a function $f(x)$ with $f(x^+)$ as reference value is given through the criterion Expected Value (EV) as follows

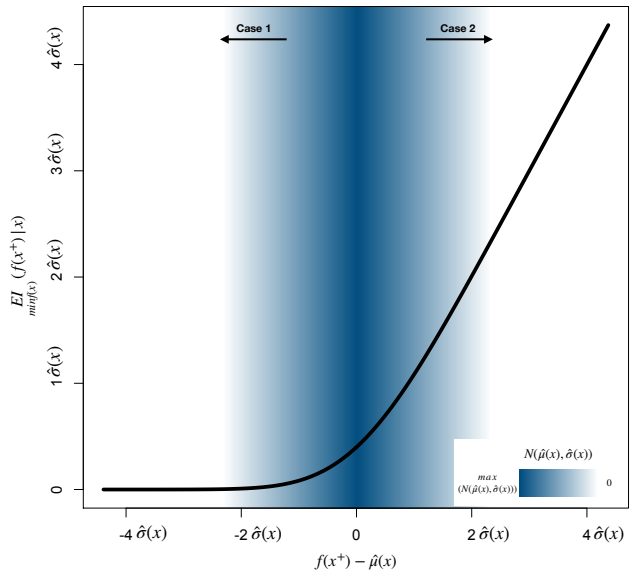


Fig. 6: Variation of $\underset{minf(x)}{EI}(f(x^+)|x)$ for change in reference value $f(x^+)$

$$\underset{minf(x)}{EV}(x|f(x^+)) = f(x^+) - \underset{minf(x)}{EI}(x|f(x^+)) \tag{26}$$

The definition of EV avoids the problem of comparing improvements with several reference values on the span of $\hat{f}(x)$. But the above two cases take a different role for the EV criterion (see Figure 7), given as follows

$$\begin{aligned} &\textbf{Case 1}: For\ \hat{\mu}(x) \pm \hat{\sigma}(x) \gg f(x^+),\ \underset{minf(x)}{EV}(x|f(x^+)) \approx f(x^+) \\ &\textbf{Case 2}: For\ \hat{\mu}(x) \pm \hat{\sigma}(x) \ll f(x^+),\ \underset{minf(x)}{EV}(x|f(x^+)) \approx \hat{\mu}(x) \end{aligned} \tag{27}$$

For **Case 1**, in the context of EI, the improvement becomes infinitesimal or zero for comparison. While in the context of EV, this leads to a problem of "over-estimation" for improvement, where the value of EV becomes the reference value itself and hence, an unrealistic reference value can lead to overestimation of improvement.

Similarly, the difference in context between EI and EV can be seen for **Case 2**. Here, we give the explanation in avoiding the above cases regarding the EV criterion since it is much easier to work with the coordinates of the objective space. As defined before, the **Case 1** can be avoided by seeking for

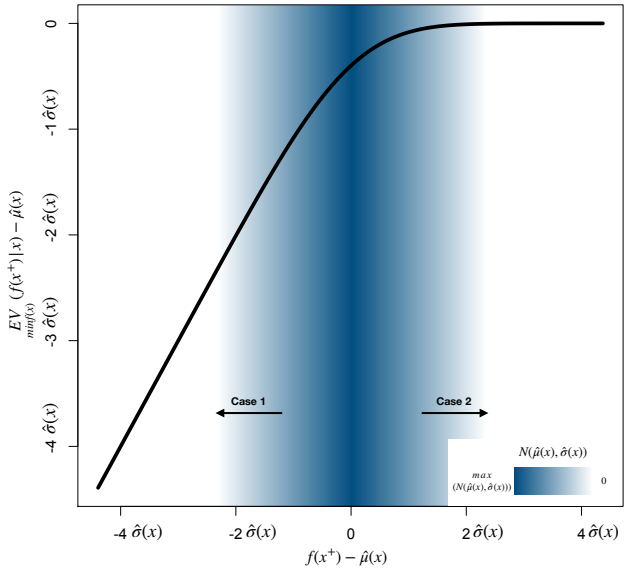


Fig. 7: Variation of $\underset{minf(x)}{EV}(f(x^+)|x)$ for change in reference value $f(x^+)$

improvement in regard to a reference value which is more realistic for improvement rather than being too greedy. The realistic scope of improvement considering $f(x^+)$ can be defined through the CDF $\mathcal{P}[\hat{f}(x) \leq f(x^+)]$ of the $\mathcal{GP}$ prediction $\hat{f}(x) : \mathcal{N}(\hat{\mu}(x), \hat{\sigma}(x))$. We here remind that the given CDF is the same as Probability of Improvement criterion (PI). Hence, the CDF constituting to zero for any reference value can be said as being too greedy for improvement, where there can be a limit set for the CDF value to be considered with compromise on greed. For the objective $minf(x)$, the limit could be set in terms of $\hat{\mu}(x) - \epsilon\hat{\sigma}(x)$ where we seek for $\mathcal{P}[\hat{f}(x) \leq f(x^+)] > \mathcal{P}[\hat{f}(x) \leq \hat{\mu}(x) - \epsilon\hat{\sigma}(x)]$, with ϵ being the parameter to be defined. A higher value of ϵ means higher the greed to seek improvement, but the balance to set the limit could be otherwise seen as the acceptable risk that can be considered for exploration. The higher risk with considering a high value of ϵ for $\hat{\mu}(x) - \epsilon\hat{\sigma}(x)$ may reap higher benefits but this could be otherwise since there is an equal probability for $\hat{f}(x) > \hat{\mu}(x) + \epsilon\hat{\sigma}(x)$. Hence, this requires a right balance for the choice of ϵ with acceptable risk for exploration.

For MOO, the most feasible improvement that we can look for is based on empirical NDS/pareto-optimal solutions $\mathscr{P} := \{\mathrm{p}_1, \mathrm{p}_2, \ldots, \mathrm{p}_x\}$ of the sampled arguments in the objective space, where $\mathrm{p_i}$ is an argument in the NDS. With $f(x)$ constituting the objective space, there can be multiple NDS that satisfy the condition $f(\mathrm{p_i}) \leq \hat{\mu}(x) - \epsilon\hat{\sigma}(x)$, where we choose the $\mathrm{p_i}$ which gives the

minimum value for $\underset{minf(x)}{EV}$. The choice of the minimum value of $\underset{minf(x)}{EV}$ for a function to be minimised intrinsically avoids the **Case 2** unless it is impossible when no suitable reference value exists. This avoids the choice of the reference value which can be too pessimistic for improvement. In overall, this provides a balance between greed and being pessimistic for improvement, defined by ϵ. The above definitions could be expressed as follows

$$\begin{aligned} \underset{min\, f(x)}{EV}(x|\mathscr{P},\epsilon) = \underset{\mathrm{p_i}\in\mathscr{P}_s}{min}\,(\underset{minf(x)}{EV}(f(\mathrm{p_i})|x)) \\ \mathscr{P}_s := \{f(\mathrm{p_i}) \leq \hat{\mu}(x) - \epsilon\hat{\sigma}(x), \forall \mathrm{p_i} \in \mathscr{P}\} \end{aligned} \tag{28}$$

where the discrete optimization of $\underset{\mathrm{p_i}\in\mathscr{P}_s}{min}\,(\underset{minf(x)}{EV}(f(\mathrm{p_i})|x))$ can be implicitly satisfied when defining NDS with $\underset{min\, f(x)}{EV}(x|\mathscr{P},\epsilon)$ as one of the objectives in the MOO to be defined. Alternatively, from the nature of proportionality for $EV(f(\mathrm{p_i})|x) \propto f(\mathrm{p_i})$, shown in Figure 7, the most suitable reference value $f(\mathrm{p_i}^\star)$ to achieve $\underset{\mathrm{p_i}\in\mathscr{P}_s}{min}\,(\underset{minf(x)}{EV}(f(\mathrm{p_i})|x))$ can be simply given as $f(\mathrm{p_i}^\star) = \underset{\mathrm{p_i}\in\mathscr{P}_s}{min} f(\mathrm{p_i})$. In short, we define the Eq. (28) as EV criterion for the following discussions.

The given definitions allow to define a EV criterion in the context of MOO for a $\mathcal{GP}$ meta-model. Hence, the infill points for MOBO can be obtained from the NDS achieved through the MOO of EV criteria, where each objective in the MOO is a EV criterion defined for a computationally expensive function to be optimized. The MOO can be achieved through algorithms like NSGA. The independent definition of improvement for functions to be optimized through EV in a MOO allows to work with only univariate Gaussian distributions and hence, large number of EV criteria could be optimized with ease.

Once NDS are obtained for optimization of EV criteria in MOO, the problem leads to choosing the infill points among the NDS. This can be dealt with clustering in the objective space through K-means of the NDS and a point can be sampled from each cluster (Jeong et al, 2006), to achieve better diversification also the possibility for parallelization. Considering simplicity, we chose the most uncertain point in each cluster given by the uncertainty of the $\mathcal{GP}$ models, where this can achieve exploration. Hence, this strategy tries to reduce uncertainty in the parts of the $\mathcal{GP}$ meta-models involved in a MOO, which are presumed to define NDS in the objective space rather than to seek for improvement through any specific metric. But care should be taken, since the adjacent points between two adjacent clusters can have the same measure of uncertainty and hence can lead to samples for parallelization from the same part of the design space.

We here give a further possibility to choose the infill points especially with the EV criterion. Since we define EV in the coordinates of the objective space

itself, there is also possibility to seek for Hyper-volume improvement (HVI). This is computationally inexpensive since only the hypervolume calculation of the NDS obtained through the optimization of the infill points taking into account the empirical NDS of the sampled points in the objective space need to be evaluated.

6 Results and discussions

All the results presented in the following discussion were obtained for the objectives defined in section 4. The $\mathcal{GP}$ meta-model was defined for the computationally expensive objective function C_s where we used a linear polynomial trend function to define the prior mean and the covariance function was defined by Matern 5/2 kernel considering anisotropic spatial correlation. Hence, in the context of MOBO, the infill points were determined primarily for the evaluation of the function C_s. Even though in our case, the $\mathcal{GP}$ model for MOBO was defined only for one of the objectives, it can be extended for MOBO with multiple $\mathcal{GP}$s.

We present a brief description on the characteristics of the Pareto-front obtained for MOBO through an EV criterion. The discontinuities appear in the Pareto-front of Figure 8 due to local definition of improvement, since the optimization is defined with multiple reference values where the discontinuities occur, but with in the same optimization setting to define NDS through NSGA-2. This means that the individuals in a given generation of NSGA-2 are defined with their respective reference values to seek improvement, but the non-dominated sorting and niching are defined on the objective space as a whole –which is less computationally expensive given the simultaneous definition of improvement with several reference values. For EI criterion, the complexity of the algorithm can be given as $O(\mathscr{N})$, where $\mathscr{N}$ is the number of evaluations of EI for an iteration in NSGA-2, while for EV, the complexity increases by $O(\mathscr{N}\mathscr{M})$, with $\mathscr{M}$ being the number of pareto-optimal solutions in the set $\mathscr{P}$. In our case this is outweighed by the complexity in constraint evaluations and iterations for finding children that satisfy the constraints to support the population size.

The optimization for the infill points through NSGA-2 was defined with the following parameters:

- Population size: 200
- No. of generations: 50
- Crossover definition: Simulated Binary crossover (SBX)
- SBX distribution index: 8
- Crossover probability: 0.9
- Mutation definition: Polynomial Mutation
- Polynomial Mutation index: 20
- Mutation probability: 0.2

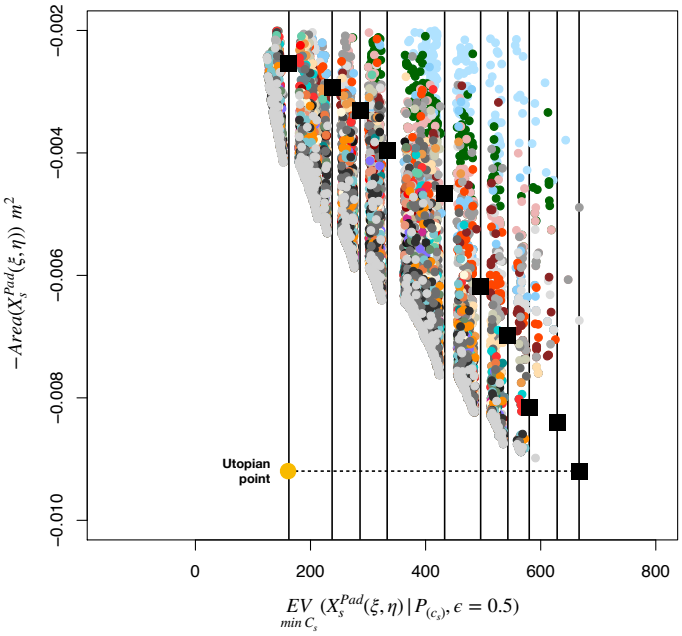
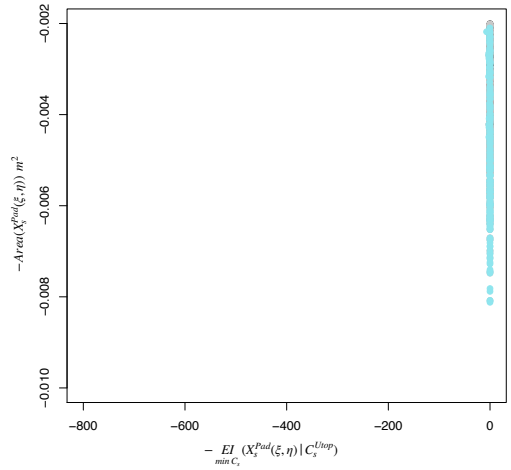


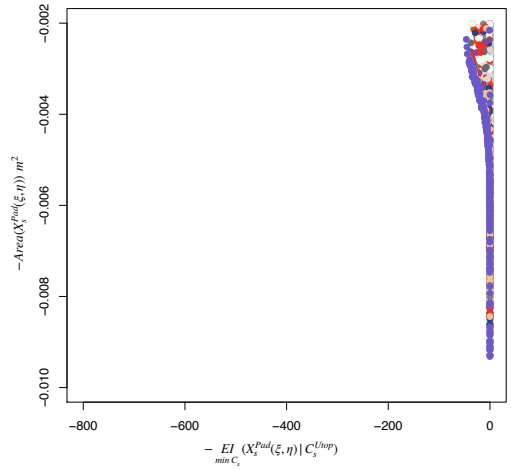
Fig. 8: Optimization for infill points with EV criterion and $-Area(X_s^{Pad}(\xi,\eta))$. The colored points represent the population, with each color representing a generation obtained through NSGA-2. The black squared points are the NDS ($P_{(c_s)}$) in the objective space and the vertical lines correspond to the NDS which are used as reference values for the EV criterion. (We here remind that the coordinates for the objective space and with EV are the same)

Mutation probability was restricted due to more probability for failure with constraints in defining $X_s^{Pad}(\xi,\eta)$. This is balanced with increase in the population size and the number of generations for convergence.

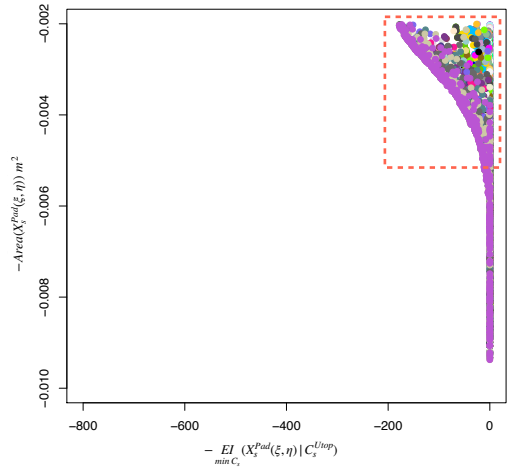
In general, with EV criterion, the initial generation starts with largely individuals whose eligibility for reference value will be far from the Utopian value –where the eligibility depends on the parameter ϵ in (28)– and through the progression of the generations will reach the individuals which will be eligible for improvement regarding the Utopian value. This is shown in Figure 10 for the EV criterion with $\epsilon = 1$, along with comparison for EI criterion in Figure 9 with Utopian value as reference, where a same $\mathcal{GP}$ model was used for both the cases.



(a) Iteration 4:

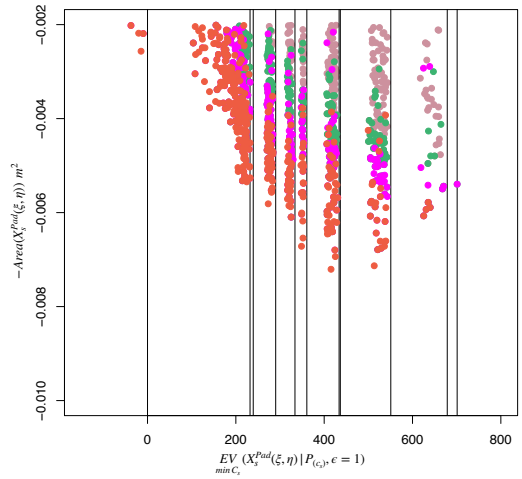


(b) Iteration 22:

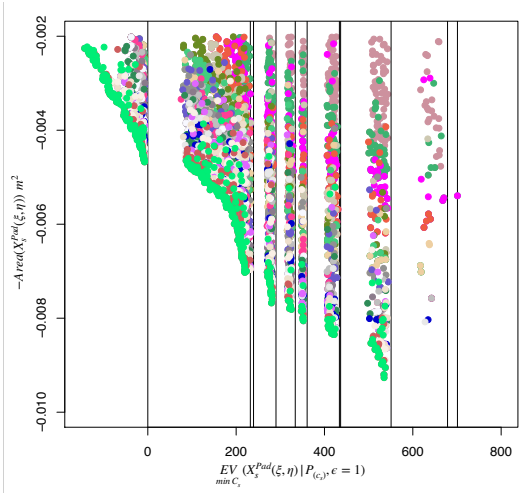


(c) Iteration 50:

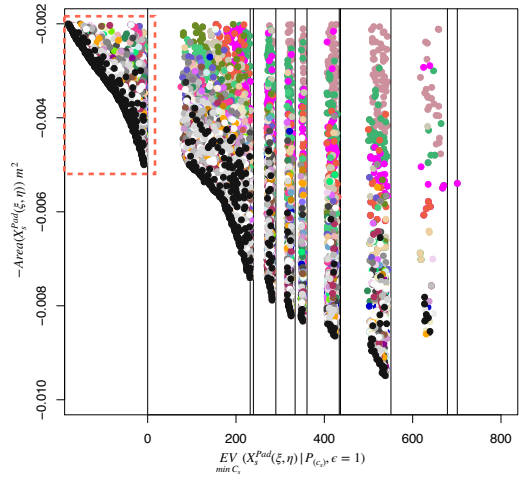
Fig. 9: Optimization for infill points with EI criterion and $-Area(X_s^{Pad}(\xi,\eta))$. The colored points represent the population with each color representing a generation obtained through NSGA-2.The red highlight at the iteration 50 is to be compared with the iteration 50 from figure 10.



(a) Iteration 4:



(b) Iteration 22:



(c) Iteration 50:

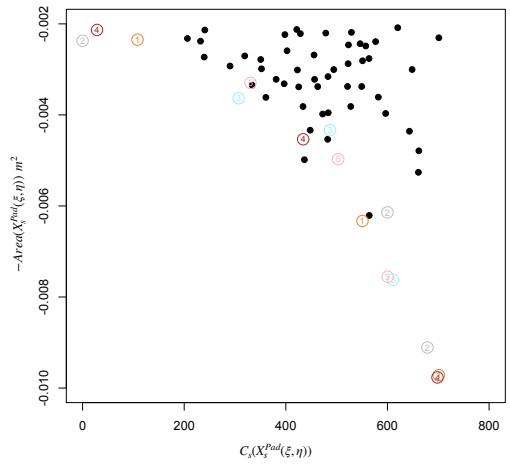
Fig. 10: Optimization for infill points with EV criterion and $-Area(X_s^{Pad}(\xi,\eta))$. Optimization of acquisition function defined by EV criterion to determine infill points. Population with each color representing a generation obtained through NSGA-2 for Min. EV. The red highlight at the iteration 50 is to be compared with the iteration 50 from figure 9.

When comparing Figure 9 and Figure 10, the optimization of EI with Utopian value as reference has some correspondance to the optimization of EV with Utopian value as one of the reference values, which can be seen at the Iteration 50 for both the cases. This is quite expected since the improvement related to the Utopian value is the same for both the criteria atleast in the region where some probability for improvement exists given through the parameter ϵ for the EV criterion. In the region where improvement was not possible to be defined for the EV given an Utopian value, the resolution for EI was very poor, which happens as $Area(X_s^{Pad}(\xi, \eta))$ increases. But for the EV, for a larger $Area(X_s^{Pad}(\xi, \eta))$, the optimization was defined with a reference value more probable. The above comparison shows the limitation of defining improvement based on a single reference value where we chose Utopian value to define the EI criterion, and hence justifies the definition of improvement with EV criterion in our application. The limitation of a single reference value can also be shown through Nadir value which we did not expose here, but a theoretical justification to this limitation was given in 5.2.

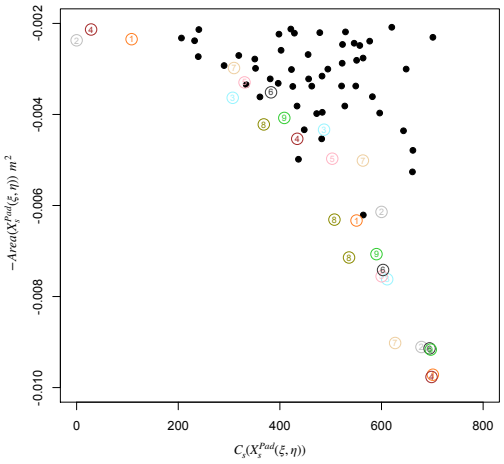
Though in the above case there was a possibility to define some improvement linked with the Utopian value for the EI criterion, but it cannot be in general. This is possible when the Utopian value may correspond to the extremum of the function, where the EI with Utopian value as reference can fail to define improvement completely since there can be no possibility to define improvement related to a given extremum. But the EV criterion can adapt to define improvement with respect to the next possible reference value.

Though we have not provided results to show the effect of the parameter ϵ, it is also possible to infer through (28) that the discontinuities in the NDS will become stronger with increase in the value of ϵ and converges to the case similar to EI with Utopian value as reference, except in different coordinates. The opposite is true with decrease in ϵ where the EV converges to EI with Nadir value as reference.

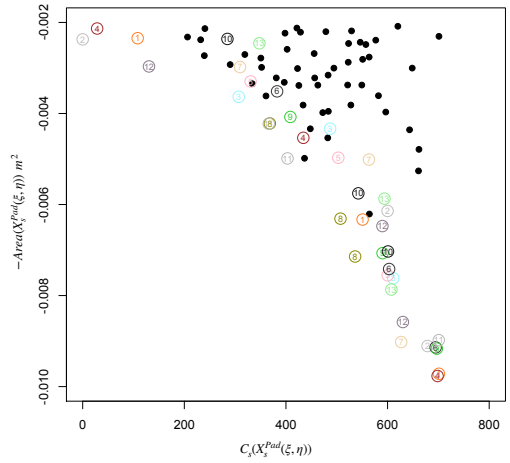
We discussed the optimization to determine the infill points where the choice of the infill points from the NDS is detailed in the section 5.2 and now we show its consequence in the objective space for improvement in NDS of the sampled arguments. The improvement of the NDS for the sampled arguments in the objective space through the progression of iterations for MOBO with the EV criterion is shown in the Figure 11.



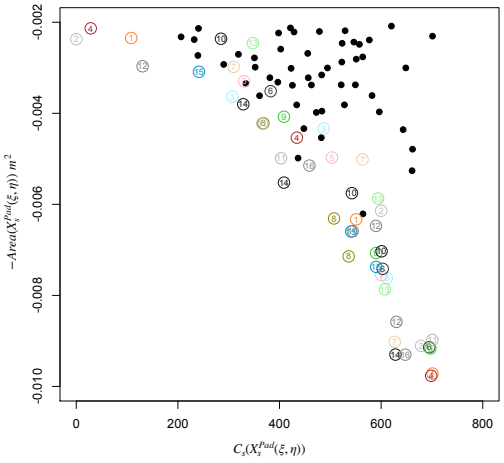
(a) Iterations 1-4:



(b) Iterations 1-8:



(c) Iterations 1-12:



(d) Iterations 1-16:

Fig. 11: Objective space : $C_s(X_s^{Pad}(\xi,\eta)) \,|\, \Im(\Lambda(X_s^{Pad}(\xi,\eta))) \in [10KHz, 13KHz]$ and $-Area(X_s^{Pad}(\xi,\eta))$. Addition of infill points with progression of iteration for MOBO with EV criterion. Black points represent the initial samples. The colored circles represent the Infill points, with 3 infill points per iteration, where each color represents an iteration with the iteration number.

The MOBO optimization (shown in Figure 11) was initiated with an initial sample size of 50 which satisfy the constraints after Latin-Hypercube sampling (LHS) and an addition of 48 infill points were added, with 3 points per iteration for a total of 16 iterations. As it can be seen, the initial samples did not cover any designs with larger $Area(X_s^{Pad}(\xi,\eta))$, which is due to very few designs in this region and hence, all the designs with larger $Area(X_s^{Pad}(\xi,\eta))$ were purely obtained in the process of optimization. The improvement in the NDS can also be seen where all the NDS in the final iteration were obtained through the infill points. We also see some overlap of infill points but clustering through K-means in the design space shows that the overlapping points mostly correspond to completely different clusters, as seen in Figure 12, which also indicates the multi-modality of the function C_s. The multi-modality maybe also due to the redundancies in the design space as explained in section 4. The overlapping can also be attributed to the choice of infill points from the NDS obtained for the optimization of the infill points, where we did not choose any metric for diversification but with the metric of uncertainity, as detailed in the section 5.2.

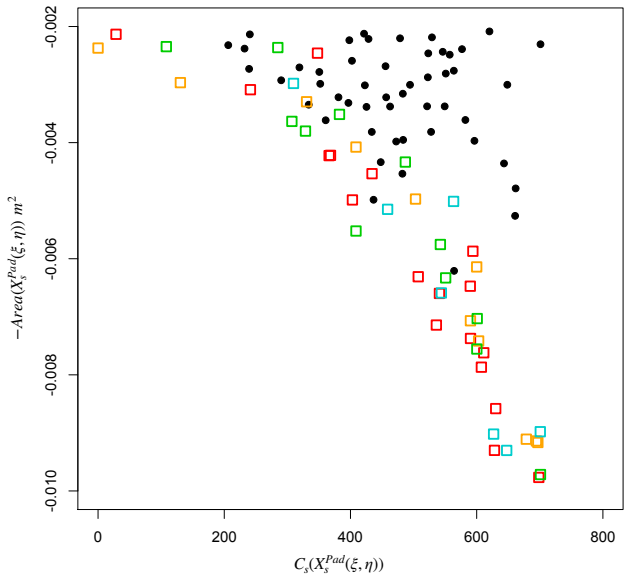


Fig. 12: Objective space : $C_s(X_s^{Pad}(\xi,\eta)) \,|\, \Im(\Lambda(X_s^{Pad}(\xi,\eta))) \in [10KHz, 13KHz]$ vs $-Area(X_s^{Pad}(\xi,\eta))$. Clustering of the infill points, with each color representing a cluster and black points representing the initial samples

The following measures of time are given in an approximate sense for a general notion of the time involved in the given MOBO. We used a 10 physical cores ($x86\_64$) machine, equipped with $252Gb$ of ECC RAM. It took a total of 53 hours to reach the given empirical Pareto-front, for which 36 hours were spent on optimisation of the infill points, which was necessary given the design space and the constraints involved. Each iteration of NSGA-2 took approximately 150 seconds, where most of the time was spent on constraint evaluation, and hence 50 iterations in NSGA-2 took 2 hours to find the converged pareto-front for MOO of the infill points. The evaluation of the stability criteria C_s took around 22 minutes with parallelization, where the parallelization was initialized for 20 friction coefficients between 0 to 1. With the availability of 10 cores, the parallelization was achieved in two batches, with a minute taken for each batch. This would have otherwise taken an additional 20 minutes to evaluate C_s. Since the parallelization was defined for the evaluation of C_s, this restricted the use of parallelization in MOBO, at least given the resources. After analysis of the time involved in each step of the MOBO, we realized that it would have been better to utilize parallelization for MOBO rather than evaluating C_s, since it would have cut time in our case to 47 hours, given that three infill points were chosen for evaluation of C_s per iteration in MOBO.

We expose some of the shapes from the objective space in Figure 13. It can be observed that, largely for a given $Area(X_s^{Pad}(\xi,\eta))$ to achieve a low value of $C_s(X_s^{Pad}(\xi,\eta))$, the pad shapes prefer to align more radially to the the disc than rather tangentially. It also seems that the pad shapes prefer to achieve a shape with three vertices, even though the shapes are defined with four vertices i.e., four curves of C^0 continuity between them, where one of the vertices is smoothed out with some continuity or one of the edges was defined to be very small to mimic a three vertices configuration. The existence of very few solutions when $Area(X_s^{Pad}(\xi,\eta))$ is larger can be attributed to lack of design space where this explains the lack of improvement related to $C_s(X_s^{Pad}(\xi,\eta))$ in this region. While, in the region of the objective space for smaller $Area(X_s^{Pad}(\xi,\eta))$, there is more flexibility in defining the pad designs $X_s^{Pad}(\xi,\eta)$ and hence, better improvement in solutions relative to $C_s(X_s^{Pad}(\xi,\eta))$ were obtained. We also obtained an interesting solution from this region with no instability as it can be observed, i.e., $C_s(X_s^{Pad}(\xi,\eta)) \,|\, \Im(\Lambda(X_s^{Pad}(\xi,\eta))) \in [10KHz, 13KHz] = 0$.

The proposed optimization leads to quite original shapes and the reader should understand that these optimal solutions do not take into account the possibility of producing mechanical parts, which is of course an important design issue but is not the purpose of the paper. Thus, the chosen optimal design has to deal with external constraints which should prevent some of the proposed solutions to be considered.

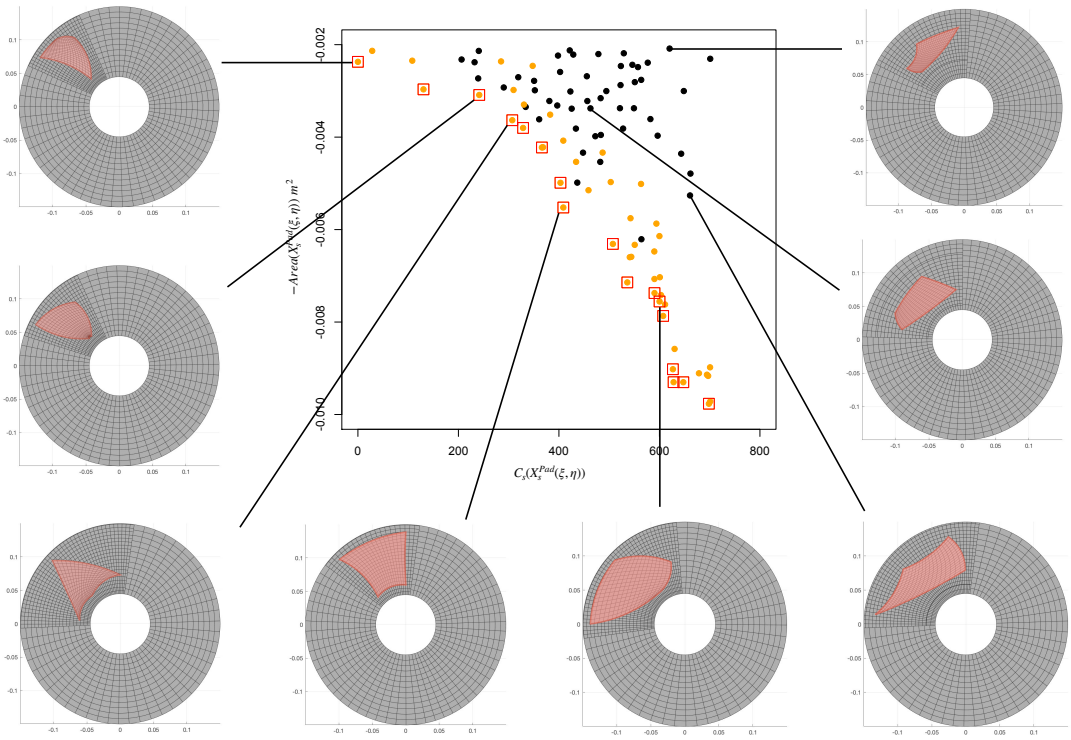


Fig. 13: A sample of shapes from the objective space are shown. Black points represent the initial samples, Orange points represent the infill points and the NDS obtained after 16 iterations of MOBO are highlighted in red.

For a given initial sample, we provide a comparison in the objective space for the infill points obtained with the EV criterion against the same number of infill points obtained through the EI criterion, given in the Figure 14, where more NDS were found to be obtained through the EV criterion. The hypervolume comparison between the two cases corresponds to sample 1 in Figure 15, which was considered with infill points from 15 iterations.

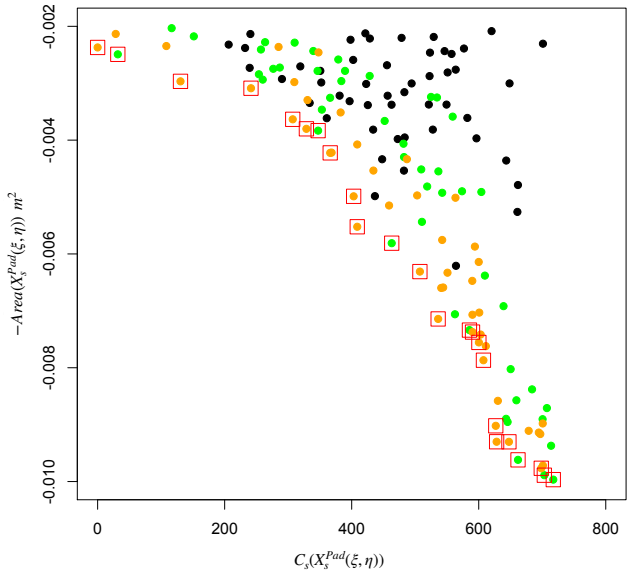


Fig. 14: Objective space : $C_s(X_s^{Pad}(\xi,\eta)) \,|\, \Im(\Lambda(X_s^{Pad}(\xi,\eta))) \in [10KHz, 13KHz]$ vs $-Area(X_s^{Pad}(\xi,\eta))$. Comparison of infill points obtained between EV and EI criteria for an initial sample. Black points represent the initial samples. Orange and green points represent infill points obtained by EV and EI respectively, with red highlights for the NDS.

In order to show the effect of initial samples to the improvement achieved through MOBO with EI and EV criteria in the objective space, the optimization was performed with 5 different initial samples for 15 iterations, with 3 infill points per iteration. The hypervolume improvement comparison is shown in Figure 15 where EV criterion outperforms for all initial samples. We here remind that the infill points were chosen not with the goal of HVI but to reduce the uncertainty of the solutions from each cluster of the NDS obtained for the optimization of infill points, but consequently, this shows better improvement with EV criterion because of better resolution in defining the NDS. Further, the overall variation of the dominated hypervolume for each criterion, with comparison between the criteria are shown in Figure 16.

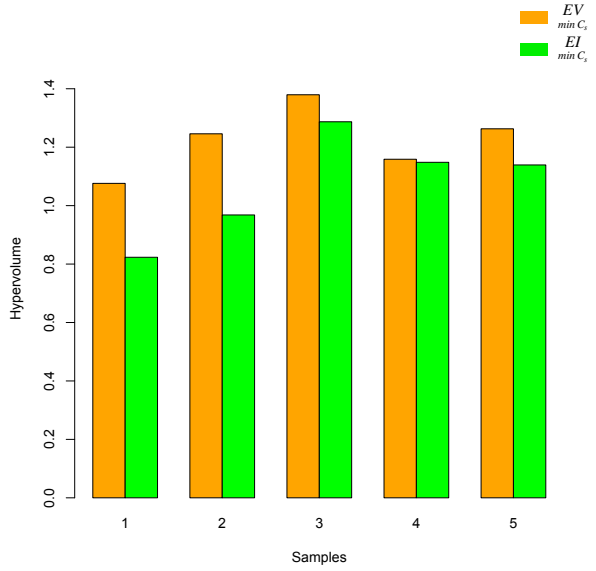


Fig. 15: Comparison between EV and EI criteria for hypervolume improvement of NDS linked with the NDS of the initial samples, shown for a five set of initial samples.

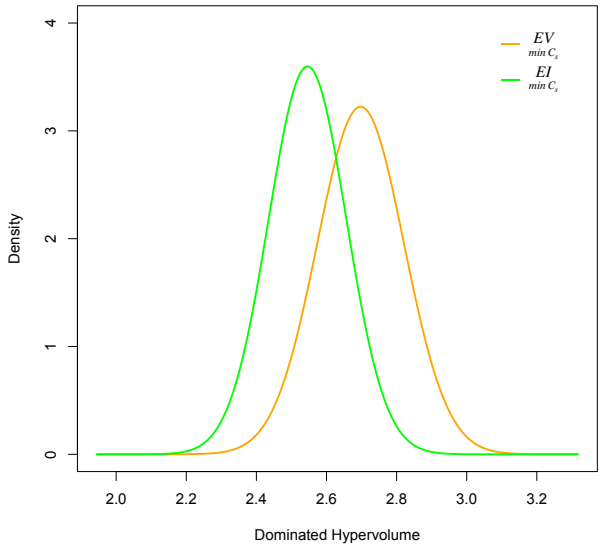


Fig. 16: The variation of the dominated hypervolume for the NDS with initializing samples, obtained for the same set of initial samples in Figure 15.

7 Conclusion

We proposed an efficient strategy to deal with the shape optimization of brake systems through a simple disc-pad representation for squeal noise reduction. The weak formulation of contact and friction specific for modelling friction induced instabilities through CEA was defined, with isogeometric approach for discretisation. This kind of study may be considered for braking systems as found in automotive or aeronautic industry, but also for many other complex structures including friction phenomena, such as clutch. Through CEA, the stability criterion was defined as a black-box function to characterize the instabilities independent of friction coefficient for the shape optimization, where parallel computation and dynamic model reduction techniques were used to reduce the computational cost of the stability criterion.

In parameterizing the computational domain for the disc-pad system with NURBS, a multi-patch parameterization strategy was realized for the disc domain to achieve local refinement at the contact interface. While, the parameterization of the pad shapes was achieved through the discrete Coon's patch method. Hence, the design space in shape optimization was constrained to the pad shapes for which injective parametrization exists with the Coon's patch

method. This was found to provide satisfactory shapes which are less conceptual for a pad design at the reduced cost of more advanced parameterization techniques which require expensive optimization, though we are interested in such conceptual shapes for future works. The injective parameterization obtained through the Coon's patch method was also found to have a good quality parameterization required for CEA, where it can be difficult to define a robust meshing for even such shapes with the classical finite elements without severe element distortion. Though we are aware that the meshing can be achieved irrespective of the domain with the classical finite elements when robustness is not questioned, which is simply not possible with the body-fitted NURBS parameterization. Nevertheless, provided an intial parameterization for a design description, the sub-sequent analysis-suitable parameterization can be easily achieved which is the underpinning of the Isogeometric approach, but defining an injective initial parameterization can be cumbersome at least for a shape optimization with arbitrary definition of shapes, or given a parameterization scheme, the question also arises for the bounds of the design space where a good quality parameterization exists.

Moving on to MOBO, the expensive stability criterion was approximated through a $\mathcal{GP}$ meta-model and a MOO was defined for minimizing the instability and maximizing the contact area. The NDS were obtained through MOBO for a given number of iterations where a new criterion EV was defined as an extension of the classical EI criterion for better resolution in MOO. The infill points obtained at each iteration of the MOBO were observed to provide consistent improvement in the NDS. Further, the hypervolume metric of the NDS was found to be better with the EV criterion compared to the EI criterion for the considered MOO, where the comparison was made with five random initializations. Some of the shapes from the NDS and the worst performing shapes were shown for comparison, where some empirical observations were discussed. It was also clear from the optimization that the instabilities at some narrow frequency range as much as 3 KHz can be reduced or completely eliminated through shape optimization irrespective of the friction coefficient. We are aware that real-life braking systems can involve more complex non-linearities, with more complex domain rather than a simple disc-pad representation which can effect the understanding of the dynamics. Nevertheless, the above frame work with CEA and the disc-pad system can be used for preliminary understanding of the dynamics purely linked with shapes in a design process, where CEA is already widely used in industries for squeal analyses. The optimization framework of MOBO could also be extended with more complex considerations for which the meta-heuristic approach is more generic irrespective of the objective functions and the constraints present, where we only tested for a bi-objective case with $\mathcal{GP}$ meta-modelisation for one of the objectives.

8 Conflict of interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

9 Replication of results

The proposed approach has been coded within Matlab and R. Upon demand we can provide the code as well as results files for provided examples in this paper.

Acknowledgements This work was mainly achieved MNRT grant French research ministry. The authors would like to thank the financial support provided by Ingnierie@Lyon, member of the Carnot institutes network.

References

Abu-Bakar AR, Ouyang H (2006) Complex eigenvalue analysis and dynamic transient analysis in predicting disc brake squeal. International Journal of Vehicle Noise and Vibration 2, DOI https://doi.org/10.1504/IJVNV.2006.011051

Akay A (2002) Acoustics of friction. The Journal of the Acoustical Society of America 111:1525–48, DOI https://doi.org/10.1121/1.1456514

Bampton MCC, Craig J R R (1968) Coupling of substructures for dynamic analyses. AIAA Journal 6(7):1313–1319, DOI 10.2514/3.4741

Bazilevs Y, Calo V, Cottrell J, Evans J, Hughes T, Lipton S, Scott M, Sederberg T (2010) Isogeometric analysis using t-splines. Computer Methods in Applied Mechanics and Engineering 199(5):229 – 263, DOI https://doi.org/10.1016/j.cma.2009.02.036, computational Geometry and Analysis

Bigoni D, Noselli G (2011) Experimental evidence of flutter and divergence instabilities induced by dry friction. Journal of the Mechanics and Physics of Solids 59:2208–2226, DOI https://doi.org/10.1016/j.jmps.2011.05.007

Cottrell J, Reali A, Bazilevs Y, Hughes T (2006) Isogeometric analysis of structural vibrations. Computer Methods in Applied Mechanics and Engineering 195(41):5257 – 5296, DOI https://doi.org/10.1016/j.cma.2005.09.027

De Lorenzis L, Wriggers P, Hughes T (2014) Isogeometric contact: A review. GAMM-Mitteilungen 37, DOI https://doi.org/10.1002/gamm.201410005

Denimal E, Nechak L, Sinou J, Nacivet S (2018) A novel hybrid surrogate model and its application on a mechanical system subjected to friction-induced vibration. Journal of Sound and Vibration 434:456–474, DOI https://doi.org/10.1016/j.jsv.2017.08.005

Emmerich M, Deutz A, Klinkenberg J (2011) Hypervolume-based expected improvement: Monotonicity properties and exact computation. 2011 IEEE Congress of Evolutionary Computation (CEC) pp 2147–2154, DOI https://doi.org/10.1109/CEC.2011.5949880

Forrester A, Keane A, Bressloff N (2006) Design and analysis of "noisy" computer experiments. AIAA Journal 44:2331–2339, DOI https://doi.org/10.2514/1.20068

Giannelli C, Jüttler B, Speleers H (2012) Thb-splines: The truncated basis for hierarchical splines. Computer Aided Geometric Design 29(7):485 – 498, DOI https://doi.org/10.1016/j.cagd.2012.03.025

Herrmann G (1971) Dynamics and Stability of Mechanical Systems with Follower Forces. NASA Contractor Report, National Aeronautics and Space Administration

Hibbit H (1979) Some follower forces and load stiffness. International Journal for Numerical Methods in Engineering 14(6):937–941, DOI https://doi.org/10.1002/nme.1620140613, https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.1620140613

Hoffmann N, Fischer M, Allgaier R, Gaul L (2002) A minimal model for studying properties of the mode-coupling type instability in friction in-

duced oscillations. Mechanics Research Communications 29:197–205, DOI https://doi.org/10.1016/S0093-6413(02)00254-9

Hughes T, Cottrell J, Bazilevs Y (2005) Isogeometric analysis: Cad, finite elements, nurbs, exact geometry and mesh refinement. Computer Methods in Applied Mechanics and Engineering 194(39-41):4135–4195, DOI https://doi.org/10.1016/j.cma.2004.10.008

Ibrahim R (1994a) Friction-induced vibration, chatter, squeal and chaos part 1: mechanics of contact and friction. Applied Mechanics Review 47:209–226, DOI https://doi.org/10.1115/1.3111079

Ibrahim R (1994b) Friction-induced vibration, chatter, squeal and chaos part 2: Dynamics and modelling. Applied Mechanics Review 47:227–253, DOI https://doi.org/10.1115/1.3111080

Jeong S, Obayashi S (2006) Multi-objective optimization using kriging model and data mining. International Journal of Aeronautical and Space Sciences 7:1–12, DOI https://doi.org/10.5139/IJASS.2006.7.1.001

Jeong S, Minemura Y, Obayashi S (2006) Optimization of combustion chamber for diesel engine using kriging model. Journal of Fluid Science and Technology 1(2):138–146, DOI https://doi.org/10.1299/jfst.1.138

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

Kindkaid N, O'Reilly O, Papadopoulos P (2003) Automative disc brake squeal. Journal of Sound and Vibration 267:105–166, DOI https://doi.org/10.1016/S0022-460X(02)01573-0

Knowles J (2006) Parego: A hybrid algorithm with on-line landscape approximation for expension multiobjective optimization problems. Evolutionary Computation, IEEE Transactions on 10:50 – 66, DOI https://doi.org/10.1109/TEVC.2005.851274

Luo C, Shimoyama K, Obayashi S (2015) A study on many-objective optimization using the kriging-surrogate-based evolutionary algorithm maximizing expected hypervolume improvement. Mathematical Problems in Engineering 2015, DOI https://doi.org/10.1155/2015/162712

Martins J, Barbarin S, Raous M, Pinto da Costa A (1999) Dynamic stability of finite dimensional linearly elastic systems with unilateral contact and coulomb friction. Computer Methods in Applied Mechanics and Engineering 177(3):289 – 328, DOI https://doi.org/10.1016/S0045-7825(98)00386-7

Mottershead JE, Chan SN (1995) Flutter Instability of Circular Discs with Frictional Follower Loads. Journal of Vibration and Acoustics 117(1):161–163, DOI https://doi.org/10.1115/1.2873860, https://asmedigitalcollection.asme.org/vibrationacoustics/article-pdf/117/1/161/5825298/161_1.pdf

Nechak L, Besset S, Sinou J (2018) Robustness of stochastic expansions for the stability of uncertain nonlinear dynamical systems – application to brake squeal. Mechanical Systems and Signal Processing 111:194–209, DOI https://doi.org/10.1016/j.ymssp.2018.01.021

Nguyen D, Nguyen M, Evgrafov A, Gersborg A, Gravesen J (2011) Isogeometric shape optimization of vibrating membranes. Computer Methods in Applied Mechanics and Engineering 200, DOI https://doi.org/10.1016/j.cma.2010.12.015

Nørtoft P, Gravesen J (2013) Isogeometric shape optimization in fluid mechanics. Structural and Multidisciplinary Optimization 48, DOI https://doi.org/10.1007/s00158-013-0931-8

Ouyang H, Nack W, Yuan Y, Chen F (2005) Numerical analysis of automative disc brake squeal: a review. International Journal of Vehicle Noise and Vibration 1:207–231, DOI https://doi.org/10.1504/IJVNV.2005.007524

Philipp B, Breitenberger M, D'Auria I, Wüchner R, Bletzinger K (2016) Integrated design and analysis of structural membranes using the isogeometric b-rep analysis. Computer Methods in Applied Mechanics and Engineering 303:312 – 340, DOI https://doi.org/10.1016/j.cma.2016.02.003

Piegl L, Tiller W (1996) The NURBS Book. Monographs in Visual Communication, Springer Berlin Heidelberg

Rasmussen CE, Williams CKI (2005) Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press

Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N (2016) Taking the human out of the loop: A review of bayesian optimization. Proceedings of the IEEE 104(1):148–175, DOI https://doi.org/10.1109/JPROC.2015.2494218

Shintani K, Azegami H (2014) Shape optimization for suppressing brake squeal. Structural and Multidisciplinary Optimization 50(6):1127–1135, DOI https://doi.org/10.1007/s00158-014-1102-2

Sinou JJ, Jézéquel L (2007) 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, DOI https://doi.org/10.1016/j.euromechsol.2006.03.002

Soobbarayen K, Besset S, Sinou J (2013) Noise and vibration for a self-excited mechanical system with friction. Applied Acoustics 74:1191–1204, DOI https://doi.org/10.1016/j.apacoust.2013.03.008

Soobbarayen K, Sinou J, Besset S (2014) Numerical study of friction-induced instability and acoustic radiation - effect of ramp loading on the squeal propensity for a simplified brake model. Journal of Sound and Vibration 333:5475–5493, DOI https://doi.org/10.1016/j.jsv.2014.05.037

10 Annexe

10.1 Isogeometric formulation

The Isogeometric formulation is a class of finite element method with main difference in the choice of the class of basis functions used for approximating a continuum solution. In general terms, for finite element method, to solve a governing differential equation $\mathfrak{D}(f) = 0$ in $\Omega \to \mathrm{R}^d$, with $f = 0$ on $\partial\Omega$, an unknown function f is defined in a finite dimensional space $\Phi : \Omega \to \mathrm{R}^d$ whose bases are a set of functions $\varphi_\mathrm{i} : \mathrm{i} \in Z$, to obtain an approximation for f as $\widehat{f} = \sum_{\forall \mathrm{i} \in \Phi} \mathrm{f_i} \varphi_\mathrm{i}$. Further, $\Phi \subset \mathbf{\Phi}$, where $\mathbf{\Phi}$ is a Hilbert space where the solution of the continuum problem belongs. The Hilbert space induced by inner product is given through the L^2 norm, where for the function $f \in L^2(\Omega)$ is defined by $\|f\|_{L^2}^2 = \langle f, f \rangle = \int_\Omega f^2 d\Omega < \infty$.

For second order differential equations, the solutions are sought in a weak sense from the Sobolev space $H^1(\Omega)$ where the inner product forms a Hilbert space for Ω as

$H^1(\Omega) = \{\varphi \in L^2(\Omega), \partial_\mathrm{l}\varphi \in L^2(\Omega), 1 \leq \mathrm{l} \leq d\}$,
$H^1_{0,\partial\Omega}(\Omega) = \{\varphi \in H^1(\Omega), \varphi = 0 \text{ in } \partial\Omega\}$

The approximation of function f in the finite dimensional space Φ as $\widehat{f}$ for the governing equation $\mathfrak{D}(f) = 0$ results in a residual $\mathfrak{D}(\widehat{f}) \neq 0$. The best hope for finding the best projection $\widehat{f} \in \Phi$ is to define the residual as orthogonal to each of the basis φ_i, which guarantees a minimum error in L^2 norm, defined as follows

$\langle \mathfrak{D}(\widehat{f}), \varphi_\mathrm{i} \rangle = 0 \qquad \forall \varphi_\mathrm{i} \in \Phi$,

where the problem leads to finding the coefficients $\mathrm{f_i}$ for all the bases φ_i with set of equations obtained with the above inner product for all the bases in the space Φ. While in classical finite element method, the space of Φ is given through polynomials like Lagrange, Hermite and Serendipity, the Isogeometric method employs NURBS basis functions which are typically used for parameterizing geometries in Computer-Aided Design.

We start with a brief description of NURBS basis functions, on which the Isogeometric framework is principally based on. We start with a definition of B-spline functions with extension to B-spline curves, from which NURBS curves are introduced with the definition of a weighing parameter. This is followed by description of higher dimensional geometries through extension by tensor product definition.

The B-spline basis functions can be defined by Cox de Boor's formula as follows,

$$N_{i,0}(\xi) = \begin{cases} 1 & \xi_i \leq \xi < \xi_{i+1} \\ 0 & otherwise \end{cases} \tag{29}$$

$$N_{i,p}\left(\xi\right) = \frac{\xi - \xi_i}{\xi_{i+p} - \xi_i} N_{i,p-1}\left(\xi\right) + \frac{\xi_{i+p+1} - \xi}{\xi_{i+p+1} - \xi_{i+1}} N_{i+1,p-1}\left(\xi\right) \tag{30}$$

where p is defined recursively for $p > 0$ to obtain a curve of degree p, which starts with a piecewise constant at $p = 0$. The knot vector on which the bases span with continuity C^{p-1} between the knots are defined through an open uniform knot vector as $\Xi = \{\xi_1, \xi_2, \cdots, \xi_{n+p+1}\}$, where it satisfies partition of unity between the knots ξ_p and ξ_{n+1}, with n being the number of control points. The knot vector need not be equidistant and the multiplicity of a knot ξ_i by $\mathcal{M}$ in the knot vector decreases the continuity by $C^{p-\mathcal{M}}$ across the knot ξ_i, which defines non-uniform B-splines. The multiplicity is usually defined at the ends of a curve where the continuity requirement beyond C^0 is not desired, where $\mathcal{M} = p$. The continuity of C^0 means that the control point is clamped to the curve, and with this at the end points, the curve can be seen as the interpolation between the end points through B-spline basis functions. Hence, through B-spline basis functions and a knot vector, a B-spline curve can be defined through the coefficients of the basis functions which are the control points P_i, as follows

$$X_c(\xi) = \sum_{i=1}^{n} \mathrm{P}_i N_{i,p}(\xi) \tag{31}$$

where $\mathrm{P}_i \in \mathrm{R}^d$, with d being the dimension of the space. The definition of a weighing parameter $w_i > 0$ associated with its respective basis function N_i, normalized defines rational B-splines where it respects the partition of unity, given as follows

$$X_c(\xi) = \sum_{i=1}^{n} \underbrace{\frac{w_i N_{i,p}(\xi)}{\sum_{i=0}^{n} w_i N_{i,p}(\xi)}}_{R_{i,p}} \mathrm{P}_i \tag{32}$$

The parameter w_i provides a new dimension for controlling the geometry, through projective transformation, while the affine transformation is achieved by P_i . Hence, the combination of non-uniform knot vectors and rational basis functions define NURBS.

The higher dimensional NURBS are a natural extension of its 1-dimensional precursor through tensor product definition where the order of the tensor is the same as the dimension of the geometry. For a 2-dimensional geometry, the tensor product NURBS surface is defined as follows

$$X_s(\xi, \eta) = \sum_{i=1}^{n} \sum_{j=1}^{m} R_{i,p}(\xi) R_{j,q}(\eta) \mathrm{P}_{i,j} \tag{33}$$

which is supported by a knot vector in all parametric directions. Similarly, to define volume, the tensor product NURBS volume is defined as

$$X_v(\xi,\eta,\zeta)=\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l}\underbrace{R_{i,p}(\xi)R_{j,q}(\eta)R_{k,r}(\zeta)}_{R_{i,j,k}(\boldsymbol{\Xi})}\mathrm{P}_{i,j,k} \tag{34}$$

which is supported by a knot vector in all parametric directions. For a more detailed explanation of NURBS, the interested readers can refer to (Piegl and Tiller, 1996).