

Table 0.1: Table of Notations

R_i	\leq	i
i	\leq	index value
T_c	\leq	
TC	\leq	
<u>Decision Variables</u>		
y_f	=	$\begin{cases} 1, & i \\ 0, & j \end{cases}$

1 Introduction

iugwdpba

2 Modelisation

Finite element methods¹ are widely used numerical methods for approximation of solutions in solving differential equations, where physically a continuum solution on a complex domain Ω is discretised through finite elements, where the discretised domain $\Omega^h \subset \Omega$. This requires the definition of variational/weak formulation for the approximation of a solution, where the solution is defined to be satisfied in an integral sense, which also weakens the continuity requirement of the solution, which otherwise requires the solution to be satisfied in a differential sense known as strong form. There are many methods to arrive at variational formulation such as the principle of stationary action, the principle of virtual work or the Galerkin's approach of weighted residual method. We primarily stick with the view point of Galerkin's approach for defining the weak form where the test functions can also be considered to be the functions associated with virtual work and hence also generalises for the principle of virtual work.

For our application in this thesis which involves contact and friction, we primarily deal with dynamics around a fixed point where perturbations around the fixed point are considered to have no relative change at the contact interface. This means that we largely deal with the assumption of small deformation and no relative sliding, and hence, the spatio-temporal non-linearities associated with the material coordinates are ignored. This also means that there is no interest in the picture of temporal variation for $\Omega \times [0, T]$ and hence, also its subsequent discretisation, are not considered in the following definitions which are given for an arbitrary time. Further, the goal is also to develop weak formulation for contact and friction appropriate for our application.

The continuum description of an initial-boundary value problem in structural dynamics can be expressed as

$$\begin{aligned} \rho \ddot{\mathbf{u}} + \nabla \cdot \boldsymbol{\sigma}(\mathbf{u}) &= \mathbf{f} \quad \text{in } \Omega \\ \mathbf{u} &= 0 \quad \text{on } \Gamma_D \\ \boldsymbol{\sigma}(\mathbf{u}) \cdot \hat{\mathbf{v}}_n &= \mathbf{t}_N \quad \text{on } \Gamma_N \end{aligned} \tag{2.1}$$

where $\mathbf{u} : \Omega^3 \rightarrow \mathbb{R}^3$, $\Gamma_N \cap \Gamma_D = \emptyset$, $\partial\Omega$ defining the boundary of Ω and, $\hat{\mathbf{v}}_n$

¹By the definition of finite element methods, we also include Isogeometric method as a class of finite element methods unless specifically stated as classical finite element methods by which we distinguish the two approaches.

defining the normal unit vector on $\partial\Omega$. Under Isotropic material consideration, the constitutive equations can be defined as

$$\boldsymbol{\sigma} = 2\mu_L \boldsymbol{\varepsilon} + \lambda_L \text{tr}(\boldsymbol{\varepsilon}) \mathbf{I} \quad (2.2)$$

where $\mu_L = \frac{E}{2(1+\nu)}$ and $\lambda_L = \frac{\nu E}{(1+\nu)(1-2\nu)}$ are 3D Lamé parameters expressed in terms of young's modulus E and Poisson's ratio ν . The kinematic relation for the strain tensor $\boldsymbol{\varepsilon}$ under infinitesimal displacement is defined to be the symmetric part of the displacement gradient as

$$\boldsymbol{\varepsilon} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T) \quad (2.3)$$

where $\nabla \mathbf{u}$ is the second-order tensor.

The Eq.(2.1) is multiplied by a weighting function $\delta \mathbf{u}$, which also generalises for the principle of virtual work, as follows

$$\int_{\Omega} \rho \ddot{\mathbf{u}} \cdot \delta \mathbf{u} \, d\Omega + \int_{\Omega} \nabla \cdot \boldsymbol{\sigma}(\mathbf{u}) \cdot \delta \mathbf{u} \, d\Omega = \int_{\Omega} \mathbf{f} \cdot \delta \mathbf{u} \quad \forall \delta \mathbf{u} | \mathbf{u} = 0 \text{ on } \Gamma_D \quad (2.4)$$

Applying Green's theorem for the term $\int_{\Omega} \nabla \cdot \boldsymbol{\sigma}(\mathbf{u}) \cdot \delta \mathbf{u} \, d\Omega$, the weak form of the problem (2.1) can be defined as follows

$$\int_{\Omega} \rho \ddot{\mathbf{u}} \cdot \delta \mathbf{u} \, d\Omega + \int_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : \nabla \delta \mathbf{u} \, d\Omega - \int_{\Gamma_N} \mathbf{t}_N \cdot \delta \mathbf{u} \, d\Gamma_N = \int_{\Omega} \mathbf{f} \cdot \delta \mathbf{u} \quad \forall \delta \mathbf{u} \quad (2.5)$$

where $\nabla \delta \mathbf{u} = \delta \boldsymbol{\varepsilon} + \delta \boldsymbol{\omega}$, with $\boldsymbol{\omega}$ being the anti-symmetric rotation tensor. Since $\boldsymbol{\sigma}$ is symmetric, $\boldsymbol{\sigma}(\mathbf{u}) : \nabla \delta \mathbf{u} = \boldsymbol{\sigma}(\mathbf{u}) : \delta \boldsymbol{\varepsilon}$.

The displacement \mathbf{u} and the stress field $\boldsymbol{\sigma}(\mathbf{u}) \cdot \hat{\mathbf{v}}_n$ on $\partial\Omega$ can be decomposed as

$$\mathbf{u} = u_n \hat{\mathbf{v}}_n + u_t \hat{\mathbf{v}}_t = \mathbf{u}_n + \mathbf{u}_t \quad \text{and} \quad \boldsymbol{\sigma}(\mathbf{u}) \cdot \hat{\mathbf{v}}_n = \sigma_n \hat{\mathbf{v}}_n + \sigma_t \hat{\mathbf{v}}_t = \boldsymbol{\sigma}_n + \boldsymbol{\sigma}_t$$

The above decomposition helps to prescribe normal and tangential stresses on $\partial\Omega$ for Neumann boundary conditions and contact boundary conditions. The contact boundary conditions on $\Gamma_C \subset \partial\Omega : \Gamma_N \cap \Gamma_D \cap \Gamma_C = \emptyset$ will be introduced in the following definitions.

2.0.1 Contact formulation

In this section, we define a short description of the concepts related to contact mechanics, which are important for the formulation in our application. The structural mechanics problem with contact can be viewed as constraints imposed on boundary

of a domain, which leads to the definition of contact boundary conditions. Unlike the classical Dirichlet and Neumann boundary conditions which are known a priori and hence can be prescribed directly in the Eq. (2.1), the contact boundary conditions are unknown a priori. Given in its basic form, it can be seen as a boundary nonlinearity from the non-linear kinematic relations which are also non-smooth multi-valued mapping giving rise to numerical complications. Hence, to satisfy the contact boundary conditions, different formulations exist with diverse approximations based on set of assumptions depending on the application. Nevertheless, we give the basic contact kinematic relations on which the approximations will be defined for our application.

For simplicity, we consider a system with domains Ω_1 and Ω_2 in contact. We start with the definition of gap function defined between the domains as follows,

$$g_n = [\mathbf{X}^{(1)} - \overleftarrow{\mathbf{X}}^{(2)}] \cdot \hat{\mathbf{v}}_n \quad (2.6)$$

where several methods exist for determining $\overleftarrow{\mathbf{X}}^{(2)}$ and $\hat{\mathbf{v}}_n$. The most easiest is to define $\hat{\mathbf{v}}_n$ as outward normal projection from the slave surface $\partial\Omega^{(1)}$ to the master surface $\partial\Omega^{(2)}$ which determines the corresponding $\overleftarrow{\mathbf{X}}^{(2)}$ for any given $\mathbf{X}^{(1)}$. Distinguish between master and slave is made depending on the mesh density where typically slave surface has more elements than the master surface. Classically, the method of closest point projection is widely used where $\overleftarrow{\mathbf{X}}^{(2)}$ is defined as follows

$$\overleftarrow{\mathbf{X}}^{(2)} = \arg \min_{\mathbf{X}^{(2)} \in \partial\Omega^{(2)}} \|\mathbf{X}^{(1)} - \mathbf{X}^{(2)}\| \quad (2.7)$$

where $\hat{\mathbf{v}}_n$ is chosen as an outward normal of $\partial\Omega^{(2)}$. Concerning our application, we mostly deal with contact between flat surfaces with finite deformation, and hence the two approaches result in nearly the same value of $\overleftarrow{\mathbf{X}}^{(2)}$ for a given $\mathbf{X}^{(1)}$, where the problem of non-uniqueness which the closest point projection method suffers doesn't concern us. This is mostly achieved by projection through parameterisation of domains using Isoparametric approach of FEM, where in Isogeometric approach the parametrisation is intrinsic. More on these definitions are discussed in §?? for classical FEM and §4.

Given the definition of gap function, the contact constraints can be defined unilaterally for a domain in contact through the set of following conditions which are commonly known as Signorini or Karush-Kuhn-Tucker (KKT) conditions, as follows

$$g_n \geq 0 \quad (2.8a)$$

$$\sigma_n \leq 0 \quad (2.8b)$$

$$g_n \sigma_n = 0 \quad (2.8c)$$

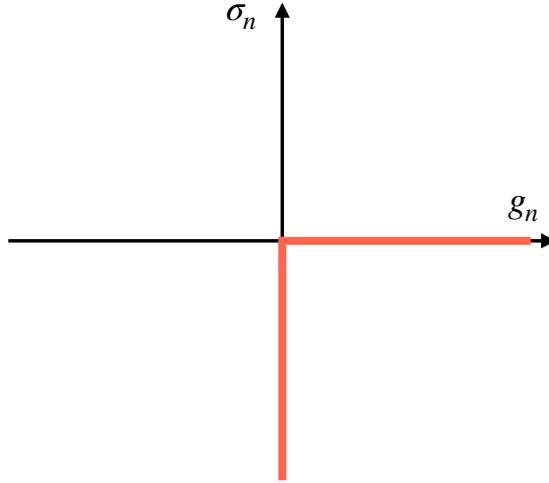


Figure 2.1: Illustration of Signorini conditions

From the conditions, the physical interpretations are apparent, the Eq. (2.8a) states that no penetration is allowed between the domains in contact, while the Eq. (2.8b) states that only compressive stress is allowed at the contact boundary, where the adhesive effects are classically ignored. The Eq. (2.8c) is given as a complementary condition which relates the first two constraints where it can be understood that when the compressive stress is nonzero, the gap function should be zero. It should be noted that the above set of constraints define multi-valued mapping which makes it intrinsically hard to define a generic solution for contact, shown in Fig. 2.1.

For small deformation problems, the gap function (2.9) can be linearised as follows

$$\Delta g_n = [\mathbf{u}^{(1)} - \overline{\mathbf{u}}^{(2)}] \cdot \hat{\mathbf{v}}_n + g_0 = u_n + g_0 \quad (2.9)$$

where g_0 represents the gap function in the reference configuration. Hence, for any incremental time, the linearized expression of the gap function will be used for the following definitions where the condition (2.8a) can be expressed as

$$u_n + g_0 \geq 0 \quad \text{or} \quad u_n - g_0 \leq 0 \quad (2.10)$$

We use the later convention $u_n - g_0 \leq 0$ for the following definitions.

2.0.2 Friction formulation

Friction is defined through Coulomb-Amonton's law where it is based on threshold conditions to define stick and slip characteristics where no motion is allowed until $\|\boldsymbol{\sigma}_t\|$ satisfies the threshold $\mu\|\boldsymbol{\sigma}_n\|$, defined as follows

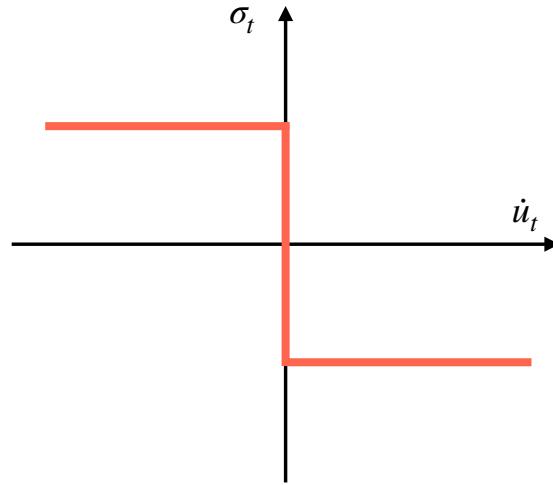


Figure 2.2: Illustration of Coulomb-Amonton's law for friction

$$|\dot{\mathbf{u}}_t| \geq 0 \quad (2.11a)$$

$$||\boldsymbol{\sigma}_t|| - \mu |\sigma_n| \leq 0 \quad (2.11b)$$

$$(||\boldsymbol{\sigma}_t|| - \mu |\sigma_n|) |\dot{\mathbf{u}}_t| = 0 \quad (2.11c)$$

where μ is the classical coefficient of friction. The above conditions can be interpreted as follows, for the stick condition $|\dot{\mathbf{u}}_t| = 0$, $||\boldsymbol{\sigma}_t|| \leq \mu |\sigma_n|$ where $||\boldsymbol{\sigma}_t||$ is inside Coulomb's cone in the space of traction stresses and similarly for the slip condition $|\dot{\mathbf{u}}_t| > 0$, $||\boldsymbol{\sigma}_t|| = \mu |\sigma_n|$ where $||\boldsymbol{\sigma}_t||$ is on the Coulomb's cone. The conditions are graphically shown in Fig. 2.2, similar to Signorini conditions for contact, the conditions define multi-valued mapping.

The initial boundary value problem Eq.(2.1) with unilateral conditions for contact and friction can be given as follows

$$\begin{aligned}
\rho \ddot{\mathbf{u}} + \nabla \cdot \boldsymbol{\sigma}(\mathbf{u}) &= \mathbf{f} \quad \text{in } \Omega \\
\mathbf{u} &= \mathbf{u}_D \quad \text{on } \Gamma_D \\
\boldsymbol{\sigma}(\mathbf{u}) \cdot \hat{\mathbf{v}}_n &= \mathbf{t}_N \quad \text{on } \Gamma_N
\end{aligned}$$

(2.12)

$$\begin{aligned}
g_n \geq 0, \quad \sigma_n \leq 0, \quad g_n \sigma_n = 0 \\
|\dot{\mathbf{u}}_t| = 0 \implies \|\boldsymbol{\sigma}_t\| - \mu |\sigma_n| \leq 0 \\
|\dot{\mathbf{u}}_t| \neq 0 \implies \|\boldsymbol{\sigma}_t\| - \mu |\sigma_n| \frac{\dot{\mathbf{u}}_t}{|\dot{\mathbf{u}}_t|} = 0 \\
&\quad \text{on } \Gamma_C
\end{aligned}$$

Unlike the classical weak form (2.5) which can be obtained through the optimisation of minimising the energy functional, the presence of inequalities from contact and friction expresses the problem in the context of convex optimisation of functional. Hence, the weak form of the problem has the form of variational inequality where the admissible solution $\dot{\mathbf{u}}$ is defined over a convex set, given as follows

$$\begin{aligned}
&\int_{\Omega} \rho \ddot{\mathbf{u}} \cdot (\delta \mathbf{u} - \dot{\mathbf{u}}) \, d\Omega + \int_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : (\nabla \delta \mathbf{u} - \nabla \dot{\mathbf{u}}) \, d\Omega \\
&\quad - \int_{\Gamma_C} \sigma_n (\delta u_n - \dot{u}_n) \, d\Gamma_C - \int_{\Gamma_C} \sigma_t (|\delta \mathbf{u}_t| - |\dot{\mathbf{u}}_t|) \, d\Gamma_C \\
&\quad - \int_{\Gamma_N} \mathbf{t}_N \cdot (\delta \mathbf{u} - \dot{\mathbf{u}}) \, d\Gamma_N - \int_{\Omega} \mathbf{f} \cdot (\delta \mathbf{u} - \dot{\mathbf{u}}) \, d\Omega \geq 0
\end{aligned} \tag{2.13}$$

where the weak form contains the simultaneous presence of two inequalities for contact and friction. To make it complete as an initial boundary value problem, the initial conditions can be defined as \mathbf{u}_0 and $\dot{\mathbf{u}}_0$ which satisfies the above equation at the initial time.

The solution to the above dynamical problem is often discussed in the context of non-smooth mechanics which we do not focus here. The existence and uniqueness of solution to the above problem is conferred under certain conditions through regularisation of the multi-valued mapping from the Signorini conditions (2.8) and the Coulomb-Amonton's law (2.11). Such multi-valued mapping are also seen as fundamental in defining some of the friction-induced dynamic instabilities such as stick-slip. Hence, the model for regularization and the parameters in modelling the regularization are important depending on the hypotheses that model the nature of a given instability.

We focus on modelling flutter-type dynamic instability through classical theories of linear analysis, where the effect of perturbation around a fixed point is analysed. Hence, the stability of the dynamical system with frictional contact (2.13)

can be characterized by determining the fixed point which is typically quasi-static or steady-sliding equilibrium depending on the characteristics of the external forces, and defining the dynamics for the perturbation around the fixed point. For non-linear systems such as the system with frictional contact, the stability can be defined through linearizing the perturbation around a fixed point, which brings the question of modelling the multi-valued mapping to be linear. This is possible through regularization of the multi-valued mapping with functions through normal-compliance approach which will be discussed in detail with the case of steady-sliding equilibrium. As an intermediate step in realizing the stability analysis for the steady-sliding equilibrium, we discuss the quasi-static insight for the above dynamical system (2.13).

The dynamical problem can be expressed as quasi-static problem when the inertial effects can be ignored, given as

$$\begin{aligned} \int_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : (\nabla \delta \mathbf{u} - \nabla \dot{\mathbf{u}}) d\Omega \\ - \int_{\Gamma_C} \sigma_n (\delta u_n - \dot{u}_n) d\Gamma_C - \int_{\Gamma_C} \sigma_t (|\delta \mathbf{u}_t| - |\dot{\mathbf{u}}_t|) d\Gamma_C \\ - \int_{\Gamma_N} \mathbf{t}_N \cdot (\delta \mathbf{u} - \dot{\mathbf{u}}) d\Gamma_N - \int_{\Omega} \mathbf{f} \cdot (\delta \mathbf{u} - \dot{\mathbf{u}}) d\Omega \geq 0 \quad (2.14) \end{aligned}$$

The quasi-static frictional problem characterizes the time-dependent variation of \mathbf{u}_t with Coulomb's conditions, which implies the presence of time-dependent external forces \mathbf{t}_N and \mathbf{f} . Hence, the system can be considered as series of quasi-static equilibrium, where the stability of the dynamical system can be characterized around such equilibrium states. This means that the stability could be defined taking in to account of the time-dependent external forces, or also velocity-dependent friction coefficient, where the history of loading is important for such applications. The solution to the quasi-static problem, either with uniqueness or non-uniqueness is proved to exist under strict conditions.

The quasi-static equilibrium can be expressed as a steady-sliding equilibrium between at least two half-spaces when no net acceleration is present. Similar to quasi-static equilibrium, this can be seen as a series of equilibrium states with respect to time where the equilibrium characteristics remain the same, except for the change in the contact domain. Since the equilibrium characteristics remain the same for all the time, the equilibrium state could be expressed with no time dependent forces but purely by static forces. This also means that the knowledge of $\hat{\mathbf{v}}_t$ for $\sigma_t \hat{\mathbf{v}}_t$ at Γ_C is known a priori, where the general notion of $\hat{\mathbf{v}}_t$ could be given as $\hat{\mathbf{v}}_k$ for the known sliding direction. Hence, the time-independent definition of Coulomb's law could be expressed for the so-called static state as follows

$$|\mathbf{u}_t| \geq 0 \quad (2.15a)$$

$$||\boldsymbol{\sigma}_t|| - \mu|\boldsymbol{\sigma}_n| \leq 0 \quad (2.15b)$$

$$(||\boldsymbol{\sigma}_t|| - \mu|\boldsymbol{\sigma}_n|)|\mathbf{u}_t| = 0 \quad (2.15c)$$

The steady-sliding equilibrium explicitly defines the slip condition where the sliding velocity is constant and hence at the equilibrium, $\sigma_t = \mu\sigma_n$ at Γ_C . For simplicity, we consider only one half-space of the sliding contact where the half-space is considered to be fixed relative to the other half-space, while the other half-space moves parallel to $\hat{\mathbf{v}}_t$ creating a sliding contact. Hence, at any time, the equilibrium state could be expressed as follows

$$\begin{aligned} \int_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : (\nabla \delta \mathbf{u} - \nabla \mathbf{u}) \, d\Omega - \int_{\Gamma_C} \mu \sigma_n (|\delta \mathbf{u}_t| - |\mathbf{u}_t|) \, d\Gamma_C \\ - \int_{\Gamma_N} \mathbf{t}_N \cdot (\delta \mathbf{u} - \mathbf{u}) \, d\Gamma_N - \int_{\Omega} \mathbf{f} \cdot (\delta \mathbf{u} - \mathbf{u}) \, d\Omega \geq 0 \end{aligned} \quad (2.16)$$

Since the knowledge of $\hat{\mathbf{v}}_t$ for $u_t \hat{\mathbf{v}}_t$ at Γ_C is known a priori as $\hat{\mathbf{v}}_k$ and with the slip condition for friction, the above formulation can be defined as follows

$$\begin{aligned} \int_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : (\nabla \delta \mathbf{u} - \nabla \mathbf{u}) \, d\Omega - \int_{\Gamma_C} \mu \sigma_n (\delta \mathbf{u} - \mathbf{u}) \cdot \hat{\mathbf{v}}_k \, d\Gamma_C \\ - \int_{\Gamma_N} \mathbf{t}_N \cdot (\delta \mathbf{u} - \mathbf{u}) \, d\Gamma_N - \int_{\Omega} \mathbf{f} \cdot (\delta \mathbf{u} - \mathbf{u}) \, d\Omega \geq 0 \end{aligned} \quad (2.17)$$

where the multi-valued mapping of friction is replaced by a smooth definition. We now introduce the function space for defining $\delta \mathbf{u}$ and \mathbf{u} , where $\delta \mathbf{u}$ and \mathbf{u} are defined to be from the same space $\mathbf{K} \subset \mathbf{V}$, with \mathbf{K} being the convex subset of \mathbf{V} , given as

$$\mathbf{V} := \{\delta \mathbf{u} \in (H^1(\Omega))^3 | \delta \mathbf{u} = \mathbf{u}_D \text{ on } \Gamma_D\}$$

$$\mathbf{K} := \{\delta \mathbf{u} \in \mathbf{V} | \delta \mathbf{u}_n \leq g_n\}$$

$$\mathbf{f} \in (L^2(\Omega))^3 \text{ and } \mathbf{t} \in (H^{-1/2}(\partial\Omega))^3$$

where $(H^1(\Omega))^3$ is the Sobolev space of functions, given with the property

$$(H^1(\Omega))^3 := \{\delta \mathbf{u} \in (L^2(\Omega))^3, \nabla \delta \mathbf{u} \in (L^2(\Omega))^3\}$$

Hence, the Hilbert space induced by inner product is given through L^2 norm:
 $\|\delta \mathbf{u}\|_{L^2} = \langle \delta \mathbf{u}, \delta \mathbf{u} \rangle = \int_{\Omega} \delta \mathbf{u}^2 \, d\Omega < \infty$. A subspace $(H^{1/2}(\partial\Omega))^3$ can be defined as the restriction of $(H^1(\Omega))^3$ on $\partial\Omega$, where $(H^{-1/2}(\partial\Omega))^3$ is the dual of the space $(H^{1/2}(\partial\Omega))^3$.

Even though \mathbf{t} is typically defined to be in $L^2(\Omega)$, the unknown a priori conditions of \mathbf{t} for σ_n on Γ_C results in the space to be $H^{1/2}(\partial\Omega)$. Unlike the case for the dynamical problem or quasi-static problems which are characterized by the presence of two simultaneous inequalities, the above problem is characterized purely by the inequalities of the Signorini conditions, when friction is defined for the slip condition with equality.

The Signorini and Coulomb's conditions (2.8) (2.11) represent the mathematical model for contact and friction in macroscopic view, where such conditions have been found through experiments to be far from the physical reality. This leads to the view of normal compliance approach to define a better approximation of the physical reality, also through which regularization for the muti-valued mappings is achieved, where σ_n is related as the function of gap function g_n with a set of parameters determined largely by experiments, given as

$$-\sigma_n = c_n(u_n - g_n)_+^{m_n} \quad (2.18)$$

where $(.)_+$ allows only positive value. This can be extended to friction as follows

$$|\dot{\mathbf{u}}_t| \geq 0 \quad (2.19a)$$

$$\|\boldsymbol{\sigma}_t\| - c_t(u_n - g_n)_+^{m_t} \leq 0 \quad (2.19b)$$

$$\|\boldsymbol{\sigma}_t\| - c_t(u_n - g_n)_+^{m_t} |\dot{\mathbf{u}}_t| = 0 \quad (2.19c)$$

where the parameters c_n , m_n , c_t and m_t are determined from the experiments. The normal compliance can also be applied to the previously defined dynamic and quasi-static cases, where existence and uniqueness of solution were proved through normal compliance under certain assumptions. For steady-sliding equilibrium, $|\dot{\mathbf{u}}_t|$ can be expressed as $|\mathbf{u}_t|$.

The Eq. (2.17) can hence be defined through normal compliance as follows

$$\begin{aligned} & \int_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : (\nabla \delta \mathbf{u} - \nabla \mathbf{u}) \, d\Omega - \int_{\Gamma_C} c_t(u_n - g_n)_+^{m_t} (\delta \mathbf{u} - \mathbf{u}) \cdot \hat{\mathbf{v}}_k \, d\Gamma_C \\ & \quad - \int_{\Gamma_N} \mathbf{t}_N \cdot (\delta \mathbf{u} - \mathbf{u}) \, d\Gamma_N - \int_{\Omega} \mathbf{f} \cdot (\delta \mathbf{u} - \mathbf{u}) \, d\Omega \geq 0 \end{aligned} \quad (2.20)$$

The above variational inequality can be expressed as variational equality through the active set strategy, where instead of looking for the admissible displacement field that satisfies the condition $u_n \leq g_n$ on Γ_C , as the minimizer of the above functional in the convex set \mathbf{K} if there exists a unique solution in the set, the set Γ_C is defined a priori, called active set, at least for an incremental time for which $u_n - g_n \rightarrow 0$,

provided that $u_n \leq g_n$ was satisfied prior to the given time. With the knowledge of active set, the displacement field u_n on Γ_C could be satisfied by normal compliance given in Eq. (2.18). Hence, the above functional with inequality can be expressed through equality as follows

$$\int_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : \nabla \delta \mathbf{u} \, d\Omega - \underbrace{\int_{\Gamma_C} c_t (u_n - g_n)_+^{m_t} \delta \mathbf{u} \cdot \hat{\mathbf{v}}_k \, d\Gamma_C}_{\langle \boldsymbol{\sigma}_n, \delta \mathbf{u} \rangle} + \underbrace{\int_{\Gamma_C} c_n (u_n - g_n)_+^{m_n} \delta \mathbf{u} \cdot \hat{\mathbf{v}}_n \, d\Gamma_C}_{\langle \boldsymbol{\sigma}_t, \delta \mathbf{u} \rangle} - \int_{\Gamma_N} \mathbf{t}_N \cdot \delta \mathbf{u} \, d\Gamma_N - \int_{\Omega} \mathbf{f} \cdot \delta \mathbf{u} \, d\Omega = 0 \quad (2.21)$$

where the inner products $\langle \boldsymbol{\sigma}_n, \delta \mathbf{u} \rangle$ and $\langle \boldsymbol{\sigma}_t, \delta \mathbf{u} \rangle$ define the weak form of contact and friction respectively. The definition of Γ_C through the active set strategy means that $\mathbf{t}_N \in (L^2(\Omega))^3$. When $m_n \neq 1$ and $m_t \neq 1$, the steady-sliding equilibrium can be solved through non-linear programming like Newton-Raphson for \mathbf{u}_{eq} .

We do not relate to the experimental determination of the parameters c_n , m_n , c_t and m_t . Hence, c_n is purely given as the penalty parameter p considering numerical stability, which implies that for any $(u_n - g_0) > 0$ defined by $(u_n - g_0)_+$ is penalized by a factor p , where the ideal would be $p \rightarrow \infty$. While c_t is given for the ideal slip state of the steady-sliding equilibrium as $c_t = \mu p$. The normal compliance approach can be viewed in general as modelling springs with certain stiffness c_n which resist penetration at the contact interface. We consider $m_n = m_t = 1$, where the parameters m_n and m_t for any value other than 1 can be physically interpreted as non-linear springs. With the above consideration of the parameters for the normal compliance, the above functional can be expressed as follows

$$\int_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : \nabla \delta \mathbf{u} \, d\Omega - \int_{\Gamma_C} \mu p (u_n - g_n)_+ \delta \mathbf{u} \cdot \hat{\mathbf{v}}_k \, d\Gamma_C + \int_{\Gamma_C} p (u_n - g_n)_+ \delta \mathbf{u} \cdot \hat{\mathbf{v}}_n \, d\Gamma_C - \int_{\Gamma_N} \mathbf{t}_N \cdot \delta \mathbf{u} \, d\Gamma_N - \int_{\Omega} \mathbf{f} \cdot \delta \mathbf{u} \, d\Omega = 0 \quad (2.22)$$

where for finite deformation, the problem can be solved for \mathbf{u}_{eq} with one incremental time step.

The perturbed state of the displacement field for any excitation of the steady-sliding equilibrium can be given as

$$\mathbf{u} = \mathbf{u}_{eq} + \tilde{\mathbf{u}} \quad (2.23)$$

where $\tilde{\mathbf{u}}$ corresponds to the perturbed displacement field. The idea is to analyse the dynamics of the perturbation such that the stability of the dynamical system can be determined through the onset evolution of the dynamics for the perturbation,

where it is hypothesized that the onset evolution of the dynamics can be expressed linearly. This brings the question of linearization for the non-linear frictional contact problem, even with the sufficient approximation made for Eq. (2.22).

As a detour for generalization, we discuss the linearization of the normal compliance terms. The contact and friction terms in Eq. (2.21) with $c_n = p, c_t = \mu p$, $0 < m_n \neq 1$ and $0 < m_t \neq 1$, can be expressed as

$$\langle \boldsymbol{\sigma}_n, \delta \mathbf{u} \rangle_{\Gamma_C} = \int_{\Gamma_C} p(u_n - g_0)_+^{m_n} \delta \mathbf{u} \cdot \hat{\mathbf{v}}_n \, d\Gamma_C \quad (2.24a)$$

$$\langle \boldsymbol{\sigma}_t, \delta \mathbf{u} \rangle_{\Gamma_C} = \int_{\Gamma_C} \mu p(u_n - g_0)_+^{m_t} \delta \mathbf{u} \cdot \hat{\mathbf{v}}_k \, d\Gamma_C \quad (2.24b)$$

The perturbation of the normal compliance can hence be expressed as

$$p(\tilde{u}_n + u_n - g_n)_+^{m_n} - p(u_n - g_0)_+^{m_n} \approx p(\tilde{u}_n)_+^{m_n} \quad (2.25a)$$

$$\mu p(\tilde{u}_n + u_n - g_n)_+^{m_t} - \mu p(u_n - g_0)_+^{m_t} \approx \mu p(\tilde{u}_n)_+^{m_t} \quad (2.25b)$$

if it can be assumed that the parameters of the normal compliance stay the same for the perturbation $\tilde{\mathbf{u}}$ close to the equilibrium \mathbf{u}_{eq} . The linearization for the degree of the polynomial terms at \mathbf{u}_{eq} can be defined as

$$\Delta p(\tilde{u}_n)_+^{m_n}|_{\mathbf{u}_{eq}} = m_n p(\tilde{u}_n)_+^{m_n-1}|_{\mathbf{u}_{eq}} \quad (2.26a)$$

$$\Delta \mu p(\tilde{u}_n)_+^{m_t}|_{\mathbf{u}_{eq}} = \mu m_t p(\tilde{u}_n)_+^{m_t-1}|_{\mathbf{u}_{eq}} \quad (2.26b)$$

where the expressions are still non-linear in the presence of $(.)_+$. It should be noted that the perturbation \tilde{u}_n on Γ_C may not result in the same active set as the equilibrium, due to the possible separation at some parts on Γ_C , which introduces non-linearities defined by $(.)_+$. This is because the separation leads to $u_n - g_0 < 0$ and hence $p(u_n - g_0)_+^{m_n} \rightarrow 0$ which can only be taken into account nonlinearly. This is linearized with the hypothesis that the onset of instability occurs very close to the equilibrium \mathbf{u}_{eq} such that Γ_C remains the same for the perturbation $\tilde{\mathbf{u}}$. This means that if Γ_C must be constant for $\tilde{\mathbf{u}}$, the separation $u_n - g_0 < 0$ must also be penalized which can simply be achieved by evading the operator $(.)_+$. The above hypothesis for the perturbation of \tilde{u}_n on Γ_C would have been impossible to define with Signorini conditions where \tilde{u}_n may lead to strict contact separation on Γ_C , while in the view of normal compliance, the perturbation \tilde{u}_n for no variation of Γ_C can be considered to define the variation of σ_n and hence also its subsequent effect on σ_t for friction.

With the above hypothesis, the linearized weak form of contact and friction for the dynamics of the perturbation $\tilde{\mathbf{u}}$ close to \mathbf{u}_{eq} can be defined as

$$\langle \boldsymbol{\sigma}_n, \delta \mathbf{u} \rangle_{\Gamma_C} = \int_{\Gamma_C} m_n p \tilde{u}_n^{m_n-1} |_{\mathbf{u}_{eq}} \delta \mathbf{u} \cdot \hat{\mathbf{v}}_n \, d\Gamma_C \quad (2.27a)$$

$$\langle \boldsymbol{\sigma}_t, \delta \mathbf{u} \rangle_{\Gamma_C} = \int_{\Gamma_C} \mu m_n p \tilde{u}_n^{m_n-1} |_{\mathbf{u}_{eq}} \delta \mathbf{u} \cdot \hat{\mathbf{v}}_k \, d\Gamma_C \quad (2.27b)$$

It should be noted that the above weak form is different from the classical weak form of contact and friction defined by normal compliance in Eq. (2.21). The above weak form essentially characterises force which is purely displacement-dependent, also known as follower force. The definition of 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 is that the onset of instability occurs at perturbations very close to \mathbf{u}_{eq} such that the non-linearities of the classical contact and friction (2.21) can be effectively ignored and hence, friction phenomenon for a perturbed state $\tilde{\mathbf{u}}$ can be modelled linearly with the contact state of \mathbf{u}_{eq} purely by displacement-dependent forces. Hence, the hypothesis only characterises $\tilde{\mathbf{u}}$ with the linearization of non-linearities very close to \mathbf{u}_{eq} , while the characteristics of non-linearities away from \mathbf{u}_{eq} cannot be foreseen. This essentially reflects on the possibility of utilising linear analysis such as CEA 2.0.2, where CEA can characterise instability of a system through eigenmodes in one computation, which would otherwise require expensive transient analysis to find a limit cycle. This defines the system to be holonomic and autonomous since the nature of follower force depends only on generalized coordinates without explicit time dependence.

We consider the case of $m_n = 1$ and $m_t = 1$ for the following discussions, where the linearization of Eqs. (2.24) can simply be expressed as

$$\langle \boldsymbol{\sigma}_n, \delta \mathbf{u} \rangle_{\Gamma_C} = \int_{\Gamma_C} p \tilde{u}_n \delta \mathbf{u} \cdot \hat{\mathbf{v}}_n \, d\Gamma_C \quad (2.28a)$$

$$\langle \boldsymbol{\sigma}_t, \delta \mathbf{u} \rangle_{\Gamma_C} = \int_{\Gamma_C} \mu p \tilde{u}_n \delta \mathbf{u} \cdot \hat{\mathbf{v}}_k \, d\Gamma_C \quad (2.28b)$$

The initial-boundary value problem for the dynamics of the perturbation can be given as

$$\begin{aligned} \rho^{(k)} \ddot{\tilde{\mathbf{u}}}^{(k)} + \nabla \cdot \boldsymbol{\sigma}^{(k)}(\tilde{\mathbf{u}}^{(k)}) &= 0 && \text{in } \Omega^{(k)} \\ \tilde{\mathbf{u}}^{(k)} &= \tilde{\mathbf{u}}_D^{(k)} && \text{on } \Gamma_D^{(k)} \\ \sigma_n^{(k)}(\tilde{\mathbf{u}}^{(k)}) \hat{\mathbf{v}}_n &= -p \tilde{u}_n^{(k)} \hat{\mathbf{v}}_n, & \sigma_t^{(k)}(\tilde{\mathbf{u}}^{(k)}) \hat{\mathbf{v}}_k &= \mu p \tilde{u}_n^{(k)} \hat{\mathbf{v}}_k && \text{on } \Gamma_C^{(k)} \end{aligned} \quad (2.29)$$

It should be noted that for linear case, the problem can be explicitly defined without the fore knowledge of \mathbf{u}_{eq} , by defining Γ_C explicitly. For simplicity, we

consider the parameters p and μ to be constant between the contact interface of all the domains in contact. The subscript k distinguishes the domains in contact. Similar to Eq. (2.5), the weak form of the differential formulation can be expressed as

$$\int_{\Omega^{(k)}} \rho^{(k)} \ddot{\tilde{\mathbf{u}}}^{(k)} \cdot \delta \mathbf{u}^{(k)} d\Omega^{(k)} + \int_{\Omega^{(k)}} \boldsymbol{\sigma}^{(k)}(\tilde{\mathbf{u}}^{(k)}) : \nabla \delta \mathbf{u}^{(k)} d\Omega^{(k)} - \int_{\Gamma_C^{(k)}} \mathbf{t}_C^{(k)} \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)} = 0 \quad (2.30)$$

The traction force $\mathbf{t}_C \in H^{-1/2}(\Gamma_C)$ can be decomposed as follows

$$\begin{aligned} \int_{\Gamma_C^{(k)}} \mathbf{t}_C^{(k)} \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)} &= \int_{\Gamma_C^{(k)}} (\sigma_n^{(k)} \hat{\mathbf{v}}_n + \sigma_t^{(k)} \hat{\mathbf{v}}_t) \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)} \\ &= \int_{\Gamma_C^{(k)}} \sigma_n^{(k)} \hat{\mathbf{v}}_n \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)} + \int_{\Gamma_C^{(k)}} \sigma_t^{(k)} \hat{\mathbf{v}}_t \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)} \end{aligned} \quad (2.31)$$

Hence, the Eq. (2.30) for the n_k domains in contact can be defined as

$$\sum_{k=1}^{n_k} \left\{ \int_{\Omega^{(k)}} \rho^{(k)} \ddot{\tilde{\mathbf{u}}}^{(k)} \cdot \delta \mathbf{u}^{(k)} d\Omega^{(k)} + \int_{\Omega^{(k)}} \boldsymbol{\sigma}^{(k)}(\tilde{\mathbf{u}}^{(k)}) : \nabla \delta \mathbf{u}^{(k)} d\Omega^{(k)} \right. \\ \left. - \underbrace{\int_{\Gamma_C^{(k)}} \sigma_n^{(k)} \hat{\mathbf{v}}_n \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)}}_{\langle \sigma_n^{(k)}, \delta \mathbf{u}^{(k)} \rangle_{\Gamma_C^{(k)}}} - \underbrace{\int_{\Gamma_C^{(k)}} \sigma_t^{(k)} \hat{\mathbf{v}}_t \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)}}_{\langle \sigma_t^{(k)}, \delta \mathbf{u}^{(k)} \rangle_{\Gamma_C^{(k)}}} \right\} = 0 \quad (2.32)$$

where $\boldsymbol{\sigma}_n, \boldsymbol{\sigma}_t \in (H^{-1/2}(\Gamma_C))^3$. The inner products can be expressed in $(H^{1/2}(\Gamma_C))^3$ space which is the restriction of $\mathbf{u} \in H^1(\Omega) \rightarrow H^{1/2}(\Gamma_C)$ through the normal compliance approach, where the above equation can be defined as

$$\sum_{k=1}^{n_k} \left\{ \int_{\Omega^{(k)}} \rho^{(k)} \ddot{\tilde{\mathbf{u}}}^{(k)} \cdot \delta \mathbf{u}^{(k)} d\Omega^{(k)} + \int_{\Omega^{(k)}} \boldsymbol{\sigma}^{(k)}(\tilde{\mathbf{u}}^{(k)}) : \nabla \delta \mathbf{u}^{(k)} d\Omega^{(k)} \right. \\ \left. + \underbrace{\int_{\Gamma_C^{(k)}} p \tilde{u}_n^{(k)} \hat{\mathbf{v}}_n \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)}}_{\langle \boldsymbol{\sigma}_n^{(k)}, \delta \mathbf{u}^{(k)} \rangle_{\Gamma_C^{(k)}}} - \underbrace{\int_{\Gamma_C^{(k)}} \mu p \tilde{u}_n^{(k)} \hat{\mathbf{v}}_t \cdot \delta \mathbf{u}^{(k)} d\Gamma_C^{(k)}}_{\langle \boldsymbol{\sigma}_t^{(k)}, \delta \mathbf{u}^{(k)} \rangle_{\Gamma_C^{(k)}}} \right\} = 0 \quad (2.33)$$

For simplicity in expansion of the contact and friction compliance terms, we consider contact between two domains $\Omega^{(a)}$ and $\Omega^{(b)}$, with the derivation of traction forces on $\Omega^{(a)}$. Hence, the inner products $\langle \boldsymbol{\sigma}_n^{(a)}, \delta \mathbf{u}^{(a)} \rangle_{\Gamma_C^{(a)}}$ and $\langle \boldsymbol{\sigma}_t^{(a)}, \delta \mathbf{u}^{(a)} \rangle_{\Gamma_C^{(a)}}$ can be expressed as

$$\langle \boldsymbol{\sigma}_n^{(a)}, \delta \mathbf{u}^{(a)} \rangle_{\Gamma_C^{(a)}} = \int_{\Gamma_C^{(a)}} p \tilde{u}_n^{(a)} \hat{\mathbf{v}}_n \cdot \delta \mathbf{u}^{(a)} d\Gamma_C^{(a)} = \int_{\Gamma_C^{(a)}} p [(\tilde{\mathbf{u}}^{(a)} - \tilde{\mathbf{u}}^{(b)}) \cdot \hat{\mathbf{v}}_n] \delta \mathbf{u}^{(a)} \cdot \hat{\mathbf{v}}_n d\Gamma_C^{(a)} \quad (2.34)$$

$$\langle \boldsymbol{\sigma}_t^{(a)}, \delta \mathbf{u}^{(a)} \rangle_{\Gamma_C^{(a)}} = \int_{\Gamma_C^{(a)}} \mu p \tilde{u}_n^{(a)} \hat{\mathbf{v}}_n \cdot \delta \mathbf{u}^{(a)} d\Gamma_C^{(a)} = \int_{\Gamma_C^{(a)}} \mu p [(\tilde{\mathbf{u}}^{(a)} - \tilde{\mathbf{u}}^{(b)}) \cdot \hat{\mathbf{v}}_n] \delta \mathbf{u}^{(a)} \cdot \hat{\mathbf{v}}_k d\Gamma_C^{(a)} \quad (2.35)$$

The traction forces on $\Omega^{(b)}$ can similarly be defined from the conservation of momentum as $\boldsymbol{\sigma}_n^{(a)} = -\boldsymbol{\sigma}_n^{(b)}$ and $\boldsymbol{\sigma}_t^{(a)} = -\boldsymbol{\sigma}_t^{(b)}$. We defined a suitable function space $\mathbf{V} \subset (H^1(\Omega))^3$ on which the solution \mathbf{u} of the functional in modelling the continuum is presumed to exist. The arbitrary variation $\delta \mathbf{u}$ in an infinite-dimensional function space \mathbf{V} can be expressed as

$$\delta \mathbf{u} = \sum_i^{\infty} \delta \boldsymbol{\vartheta}_i \mathbf{v}_i \quad \forall \mathbf{v}_i \in \mathbf{V} \quad (2.36)$$

Hence, the Eq. (2.0.2) can be defined for a system of two domains $\Omega^{(a)}$ and $\Omega^{(b)}$ as

$$\sum_k^{a,b} \left\{ \left(\int_{\Omega^{(k)}} \rho^{(k)} \ddot{\tilde{\mathbf{u}}}^{(k)} \cdot \mathbf{v}_i^{(k)} d\Omega^{(k)} + \int_{\Omega^{(k)}} \boldsymbol{\sigma}^{(k)}(\tilde{\mathbf{u}}^{(k)}) : \nabla \mathbf{v}_i^{(k)} d\Omega^{(k)} \right. \right. \\ \left. \left. + \int_{\Gamma_C^{(k)}} p [(\tilde{\mathbf{u}}^{(k)} - \tilde{\mathbf{u}}^{(\sim k)}) \cdot \hat{\mathbf{v}}_n] \mathbf{v}_i^{(k)} \cdot \hat{\mathbf{v}}_n d\Gamma_C^{(k)} - \int_{\Gamma_C^{(k)}} \mu p [(\tilde{\mathbf{u}}^{(k)} - \tilde{\mathbf{u}}^{(\sim k)}) \cdot \hat{\mathbf{v}}_n] \mathbf{v}_i^{(k)} \cdot \hat{\mathbf{v}}_k d\Gamma_C^{(k)} \right) \delta \boldsymbol{\vartheta}_i \right\} = 0 \quad (2.37)$$

where the equation must be satisfied for any arbitrary variation of $\delta \boldsymbol{\vartheta}$. $\sim k$ defines the domain which is not k , i.e., $k = a$ implies $\sim k = b$ and vice-versa.

With the finite element approach, we define a finite-dimensional function space ${}_h \mathbf{V} \subset \mathbf{V}$ and hence, there is some bound for ${}_h \mathbf{v}_i \in {}_h \mathbf{V}$. We define the approximation of \mathbf{u} in the same space ${}_h \mathbf{V}$ as $\mathbf{u} \approx {}_h \mathbf{u} \in {}_h \mathbf{V}$, where the above equation can be defined as

$$\sum_k^{a,b} \left\{ \int_{\Omega^{(k)}} \rho^{(k)} {}_h \ddot{\tilde{\mathbf{u}}}^{(k)} \cdot {}_h \mathbf{v}_i^{(k)} d\Omega^{(k)} + \int_{\Omega^{(k)}} \boldsymbol{\sigma}^{(k)}({}_h \tilde{\mathbf{u}}^{(k)}) : \nabla_h \mathbf{v}_i^{(k)} d\Omega^{(k)} \right. \\ \left. + \int_{\Gamma_C^{(k)}} p [({}_h \tilde{\mathbf{u}}^{(k)} - {}_h \tilde{\mathbf{u}}^{(\sim k)}) \cdot \hat{\mathbf{v}}_n] {}_h \mathbf{v}_i^{(k)} \cdot \hat{\mathbf{v}}_n d\Gamma_C^{(k)} - \int_{\Gamma_C^{(k)}} \mu p [({}_h \tilde{\mathbf{u}}^{(k)} - {}_h \tilde{\mathbf{u}}^{(\sim k)}) \cdot \hat{\mathbf{v}}_n] {}_h \mathbf{v}_i^{(k)} \cdot \hat{\mathbf{v}}_k d\Gamma_C^{(k)} \right\} = 0 \\ \forall_h \mathbf{v}_i^{(k)} \in {}_h \mathbf{V}^{(k)} \quad (2.38)$$

We explain the definition of space ${}_h\mathbf{V}$ for classical finite elements in , but we mainly explain in detail for Isogeometric methods with focus on contact in , where the Isogeometric approach is new for the given application. But given a suitable choice of finite-dimensional space ${}_h\mathbf{V}$, the above equation can be defined in matrix form as

$$\mathbf{M}^{(a-b)} \ddot{\tilde{\mathbf{U}}}^{(a-b)} + (\mathbf{K}^{(a-b)} + \mathbf{K}_C^{(a-b)} + \mathbf{K}_F^{(a-b)}) \tilde{\mathbf{U}}^{(a-b)} = 0 \quad (2.39)$$

The properties of the matrices can be given as follows, where \mathbf{M} and \mathbf{K} are symmetric and positive definite. \mathbf{K}_C is also symmetric which essentially defines the conservation of momentum at the interface, while \mathbf{K}_F is non-symmetric with the non-conservative nature of friction at slip state. Often damping matrix \mathbf{C} is also added purely in numerical context, typically through modal or Rayleigh damping. In the scope of modelling rotational inertia effects, gyroscopic matrix \mathbf{G} can also be defined, where \mathbf{C} and \mathbf{G} are both velocity dependent. We do not consider damping and gyroscopic effects for the following discussions, where we mainly focus on contact and friction definitions.

With the definition of finite-dimensional space ${}_h\mathbf{V}$, the steady-sliding problem can be expressed as

$$\sum_k^{a,b} \left\{ \int_{\Omega} \boldsymbol{\sigma}^{(k)}({}_h\mathbf{u}^{(k)}) : \nabla {}_h\mathbf{v}_i^{(k)} d\Omega - \int_{\Gamma_C} \mu p({}_h u_n^{(k)} - {}_h g_n)_+ {}_h\mathbf{v}_i^{(k)} \cdot \hat{\mathbf{v}}_k d\Gamma_C \right. \\ \left. + \int_{\Gamma_C} p({}_h u_n^{(k)} - {}_h g_n)_+ {}_h\mathbf{v}_i^{(k)} \cdot \hat{\mathbf{v}}_n d\Gamma_C - \int_{\Gamma_N} {}_h\mathbf{t}_N^{(k)} \cdot {}_h\mathbf{v}_i^{(k)} d\Gamma_N - \int_{\Omega} {}_h\mathbf{f}^{(k)} \cdot {}_h\mathbf{v}_i^{(k)} d\Omega = 0 \right\} \\ \forall {}_h\mathbf{v}_i^{(k)} \in {}_h\mathbf{V}^{(k)} \quad (2.40)$$

where the solution to the problem corresponds to \mathbf{u}_{eq} . The matrix form of the problem can be expressed as

$$(\mathbf{K}^{(a-b)} + \mathbf{K}_C^{(a-b)} + \mathbf{K}_F^{(a-b)}) \mathbf{U}_{eq}^{(a-b)} = \mathbf{F}^{(a-b)} \quad (2.41)$$

For the ease of computation, the dynamical model (2.39) can be reduced through model reduction where we used Craig & Bampton reduction. This is essentially to take advantage of sub-structuring in shape optimisation, where only the matrices of the sub-structures defined for shape optimisation need to be computed in every iteration. Craig & Bampton (C&B) reduction essentially captures the properties at the contact interface along with the internal dynamic properties. In accordance to the method, for a domain $\Omega^{(k)}$, the vector $\mathbf{U}^{(k)}$ and the matrices $\mathbf{K}^{(k)}$ and $\mathbf{M}^{(k)}$ are split between internal (u) and interface (v) degrees of freedom as follows

$$\mathbf{U}^{(k)} = \begin{bmatrix} \mathbf{U}_u^{(k)} \\ \mathbf{U}_v^{(k)} \end{bmatrix} \quad (2.42)$$

$$\mathbf{K}^{(k)} = \begin{bmatrix} \mathbf{K}_{uu}^{(k)} & \mathbf{K}_{uv}^{(k)} \\ \mathbf{K}_{vu}^{(k)} & \mathbf{K}_{vv}^{(k)} \end{bmatrix} \quad (2.43)$$

$$\mathbf{M}^{(k)} = \begin{bmatrix} \mathbf{M}_{uu}^{(k)} & \mathbf{M}_{uv}^{(k)} \\ \mathbf{M}_{vu}^{(k)} & \mathbf{M}_{vv}^{(k)} \end{bmatrix} \quad (2.44)$$

The transformation matrix \mathbf{T} can be constructed from the composition of static transformation $\Theta_s = \mathbf{K}_{uu}^{-1}\mathbf{K}_{uv}$ and the matrix of eigenvectors $\Theta_d = [\Theta_1 \ \Theta_2 \ \dots \ \Theta_H]$ obtained by solving the problem $(\mathbf{K}_{uu} - \lambda_i^2 \mathbf{M}_{uu})\Theta_i = 0$ for the first H modes, where the choice of H depends on the number of modes to be represented in the problem. The matrix \mathbf{T} can hence be defined as

$$\mathbf{T}^{(k)} = \begin{bmatrix} \Theta_d^{(k)} & \Theta_s^{(k)} \\ \mathbf{0} & \mathbb{I} \end{bmatrix} \quad (2.45)$$

The transformation to define the reduced mass $\hat{\mathbf{M}}^{(k)}$ and stiffness $\hat{\mathbf{K}}^{(k)}$ matrices can be realized by $\hat{\mathbf{M}} = \mathbf{T}^T \mathbf{M} \mathbf{T}$ and $\hat{\mathbf{K}} = \mathbf{T}^T \mathbf{K} \mathbf{T}$, where the reduced matrices are expressed in the reduced coordinates $\mathbf{Z} = \begin{bmatrix} \mathcal{M} \\ \mathbf{U}_v \end{bmatrix}$, with \mathcal{M} representing the modal coordinates.

The composition of the matrices $\hat{\mathbf{M}}$ and $\hat{\mathbf{K}}$ can be expressed as

$$\hat{\mathbf{M}} = \begin{bmatrix} \hat{\mathbf{M}}_{uu} & \hat{\mathbf{M}}_{uv} \\ \hat{\mathbf{M}}_{vu} & \hat{\mathbf{M}}_{vv} \end{bmatrix} \quad (2.46)$$

$$\hat{\mathbf{K}} = \begin{bmatrix} \hat{\mathbf{K}}_{uu} & \hat{\mathbf{K}}_{uv} \\ \hat{\mathbf{K}}_{vu} & \hat{\mathbf{K}}_{vv} \end{bmatrix} \quad (2.47)$$

where for $\hat{\mathbf{M}}$, the sub-matrices can be expressed as $\hat{\mathbf{M}}_{uu} = \mathbb{I}$, $\hat{\mathbf{M}}_{uv} = \hat{\mathbf{M}}_{vu}^T = \Theta_d^T (\mathbf{M}_{uv} - \mathbf{M}_{uu} \mathbf{K}_{uu}^{-1} \mathbf{K}_{uv})$ and $\hat{\mathbf{M}}_{vv} = \mathbf{M}_{vv} - \mathbf{M}_{vu} \mathbf{K}_{uu}^{-1} \mathbf{K}_{uv} - \mathbf{K}_{vu} \mathbf{K}_{uu}^{-1} \mathbf{M}_{uv} + \mathbf{K}_{vu} \mathbf{K}_{uu}^{-1} \mathbf{M}_{uu} \mathbf{K}_{uu}^{-1} \mathbf{K}_{uv}$. Similarly for $\hat{\mathbf{K}}$, the sub-matrices can be expressed as $\hat{\mathbf{K}}_{uu} = \Theta_d^T \mathbf{K}_{uu} \Theta_d = \hat{\mathcal{I}}$, $\hat{\mathbf{K}}_{uv} = \hat{\mathbf{K}}_{vu} = 0$ and $\hat{\mathbf{K}}_{vv} = \mathbf{K}_{vv} - \mathbf{K}_{vu} \mathbf{K}_{uu}^{-1} \mathbf{K}_{uv}$, where $\hat{\mathcal{I}} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_H \end{bmatrix}$.

Essentially, the interface degrees of freedom \mathbf{U}_v for a given domain is defined by contact and friction degrees of freedom, where the reduced contact $\hat{\mathbf{K}}_C$ and friction

$$\hat{\mathbf{K}}_F \text{ matrices defined by the coordinates } \mathbf{U}_v^{(a-b)} = \begin{bmatrix} \mathbf{U}_v^{(a)} \\ \mathbf{U}_v^{(b)} \end{bmatrix}, \text{ has the form} \\ \hat{\mathbf{K}}_C^{(a-b)} = \begin{bmatrix} \hat{\mathbf{K}}_C^{(a)} & \hat{\mathbf{K}}_C^{(a,b)} \\ \hat{\mathbf{K}}_C^{(b,a)} & \hat{\mathbf{K}}_C^{(b)} \end{bmatrix} \quad (2.48)$$

$$\hat{\mathbf{K}}_F^{(a-b)} = \begin{bmatrix} \hat{\mathbf{K}}_F^{(a)} & \hat{\mathbf{K}}_F^{(a,b)} \\ \hat{\mathbf{K}}_F^{(b,a)} & \hat{\mathbf{K}}_F^{(b)} \end{bmatrix} \quad (2.49)$$

The reduced mass matrix of the system, $\hat{\mathbf{M}}^{(a-b)}$, can simply be expressed as

$$\hat{\mathbf{M}}^{(a-b)} = \begin{bmatrix} \hat{\mathbf{M}}^{(a)} & 0 \\ 0 & \hat{\mathbf{M}}^{(b)} \end{bmatrix} \quad (2.50)$$

where the mass matrix is essentially uncoupled between the domains. But it should be noted that, for $\hat{\mathbf{M}}^{(k)}$, the internal \mathbf{U}_u and interface \mathbf{U}_v degrees of freedom are coupled, i.e., $\hat{\mathbf{M}}_{uv}^{(k)} = (\hat{\mathbf{M}}_{vu}^{(k)})^T \neq 0$. This is not true in case of the reduced stiffness matrix of the system, $\hat{\mathbf{K}}^{(a-b)}$, where coupling exists through contact and friction. Hence, the sub-structuring of the domains are essentially coupled through the interface degrees of freedom \mathbf{U}_v , where the stiffness matrix with the definition of contact and friction can be expressed with the matrix $\hat{\mathbf{K}}^{(a-b)}$ in reduced coordinates as

$$\hat{\mathbf{K}}_{\cup CF}^{(a-b)} = \begin{bmatrix} \hat{\mathcal{I}}^{(a)} & 0 & 0 & 0 \\ 0 & \hat{\mathbf{K}}_{vv}^{(a)} + \hat{\mathbf{K}}_C^{(a)} + \hat{\mathbf{K}}_F^{(a)} & 0 & \hat{\mathbf{K}}_C^{(a,b)} + \hat{\mathbf{K}}_F^{(a,b)} \\ 0 & 0 & \hat{\mathcal{I}}^{(b)} & 0 \\ 0 & \hat{\mathbf{K}}_C^{(b,a)} + \hat{\mathbf{K}}_F^{(b,a)} & 0 & \hat{\mathbf{K}}_{vv}^{(b)} + \hat{\mathbf{K}}_C^{(b)} + \hat{\mathbf{K}}_F^{(b)} \end{bmatrix} \quad (2.51)$$

where the coordinates of the matrix is defined by $\mathbf{Z}^{(a-b)} = \begin{bmatrix} \mathbf{Z}^{(a)} \\ \mathbf{Z}^{(b)} \end{bmatrix}$, which is also true for $\hat{\mathbf{M}}^{(a-b)}$. Unlike $\hat{\mathbf{M}}^{(k)}$, for $\hat{\mathbf{K}}^{(k)}$ there is no coupling between $\mathbf{U}_u^{(k)}$ and $\mathbf{U}_v^{(k)}$, but essentially it is only $\hat{\mathbf{K}}^{(a-b)}$ that couples the domains.

Brake squeal

The problem of brake squeal analysis has been studied for a very long time with one of the early reviews from North and is still a challenging issue due to the immense complexities involved. Some of the complexities can be attributed to modelling friction, where models based on even simple macroscopic view (Coulomb's law) can give rise to non-smooth boundary non-linearities, which can be hard to model analytically and numerically. The early analyses were largely based on lumped models (mass-spring) where the discussions were primarily based on analytical solutions. These models were quite sophisticated and some with large degrees of freedom, which greatly improved the understanding of the friction induced phenomena. One of the conclusive evidence was that the instability occurs even at constant coefficient of friction from the coupling of two degrees of freedom by friction force.

Physical description

Considering brake squeal, it is well known that squeal noise can be linked to modal behaviour. Hence, it seems quite obvious to build an optimisation criterion to characterise squeal noise on the presence of unstable modes which define flutter-type instability. Given a linear system, the most efficient is to use eigenvalue analysis where we can characterise modes through eigenvalues in one computation. The topic of linearisation of such systems has been the discussion with previous explanations, where we achieve linearisation of contact and friction around a fixed point.

For the following definitions, we consider an applicative example of disc-pad system d – p as a simplified brake system, which essentially consists of two domains: a disc $\Omega^{(d)}$ and a pad $\Omega^{(p)}$. The disc is defined as a solid annulus geometry fixed at the inner cylindrical face. The pad is constrained to be in contact with the disc defined by \mathbf{u}_{eq} 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. The idea of defining a simple system at least initially is to make focused studies on individual problems by avoiding the complexity of boundary conditions present in a real braking system, where we focus on defining a frame-work for shape optimisation of such systems. For the following explanations, we express the matrices in normal coordinates as supposed to reduced coordinates where we later define a strategy for sub-structuring with reduced coordinates in optimisation. The problem (??) for d – p system can be defined as

$$\mathbf{M}^{(d-p)} \ddot{\tilde{\mathbf{U}}}^{(d-p)} + (\mathbf{K}^{(d-p)} + \mathbf{K}_C^{(d-p)} + \mathbf{K}_F^{(d-p)}) \tilde{\mathbf{U}}^{(d-p)} = 0 \quad (2.52)$$

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

$$(\lambda^2 \mathbf{M}^{(d-p)} + (\mathbf{K}^{(d-p)} + \mathbf{K}_C^{(d-p)} + \mathbf{K}_F^{(d-p)})) \Theta = 0 \quad (2.53)$$

It is evident that the value of λ determines the state of the perturbed solution $\tilde{\mathbf{U}}$. Since \mathbf{K}_F is non-symmetric from the definition of friction at the slip state, this can lead to λ and Θ being a complex value and a complex vector respectively. For a complex value of λ , the imaginary part $\Im(\lambda)$ defines the oscillatory behaviour of $\tilde{\mathbf{U}}$ and the real part $\Re(\lambda)$ characterizes the stability of $\tilde{\mathbf{U}}$.

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

- $\Re(\lambda) > 0$ unstable mode
- $\Re(\lambda) < 0$ stable mode
- $\Re(\lambda) = 0$ neutral

In our system, the coefficient of friction μ is the driving parameter which determines stability, such that increase in the value of μ can drive the system from stable to unstable behaviour. This type of instability is characterised by Hopf-bifurcation where the presence of a limit cycle is determined by the occurrence of a pair of eigenvalues $\pm \Re(\hat{\lambda}_\mu) + \Im(\hat{\lambda}_\mu)$, with $\hat{\lambda}_\mu$ being the eigenfrequency of two modes undergoing coalescence at μ . 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 $\pm \Re(\hat{\lambda}_{\mu_o})$ for a given frequency $\Im(\hat{\lambda}_{\mu_o})$ at the point of bifurcation μ_o and when $\mu > \mu_o$, the two modes $+\Re(\hat{\lambda}_{\mu>\mu_o}) + \Im(\hat{\lambda}_{\mu>\mu_o})$ and $-\Re(\hat{\lambda}_{\mu>\mu_o}) + \Im(\hat{\lambda}_{\mu>\mu_o})$ characterize stable and unstable behaviours respectively, leading to mode-coalescence phenomenon. The presence of this type of instability in the range of brake squeal frequency can indicate the presence of brake squeal noise. But, it should be noted that the magnitude of $+\Re(\hat{\lambda})$ defines the rate of divergence which may not quantify noise level.

Example of mode shapes for the disc-pad system obtained through CEA are shown in the Figures 2.3 and 2.4.

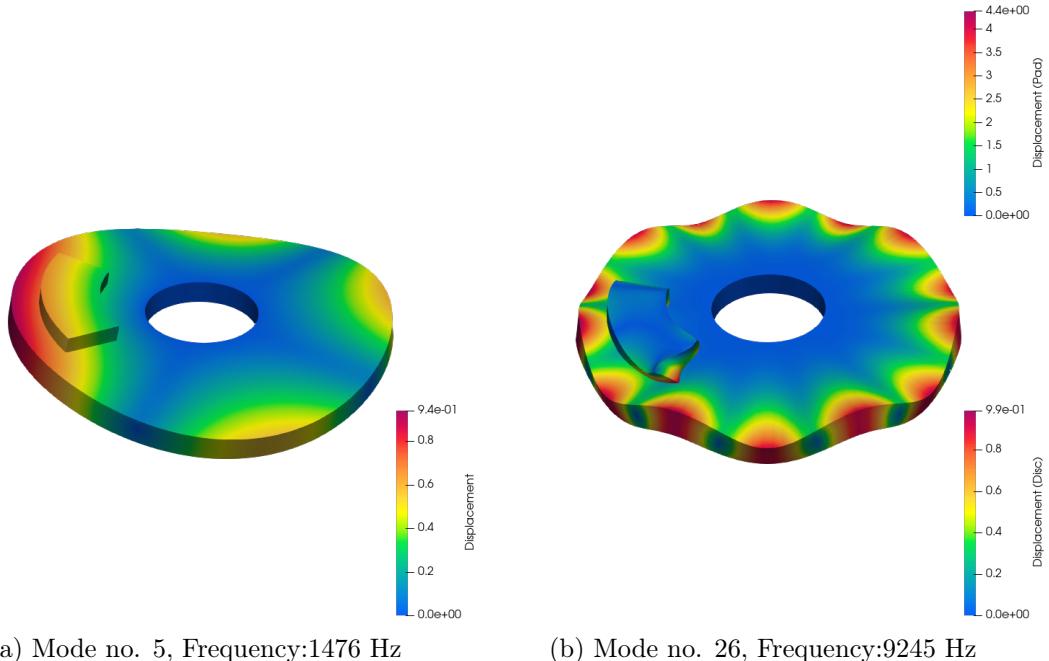
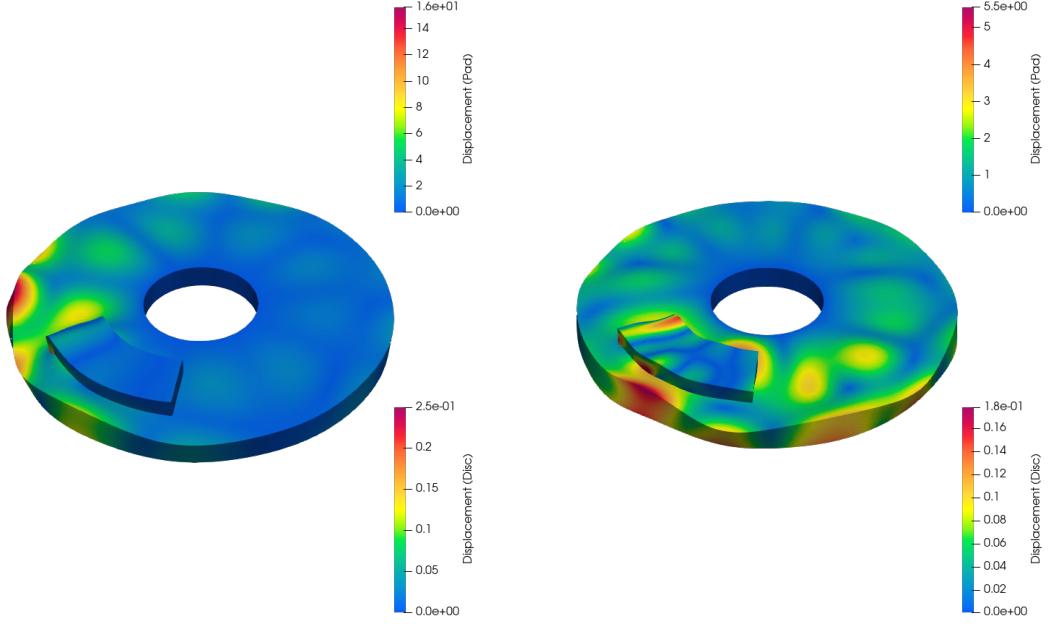


Figure 2.3: Example of disc-pad stable modes

We discuss the empirical observations from the post-processing of mode shapes, even though the results are highly subjective and depend on the value of p which is typically determined from experiments in the light of normal compliance. Typically at low frequencies, the mode shape of the pad follows the shape of the disc with correspondence in displacement field at the contact interface. While at higher frequencies,



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

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

Figure 2.4: Example of disc-pad unstable modes, where the displacement field is considered only for real-part of the eigenvector

the behaviour is complicated to understand, but relatively large difference in magnitude of displacement field between the disc and the pad was observed. Further, the unstable modes lead to definition of eigenvectors in complex-plane for displacement field, which was not considered for representation in Figure 2.4. For intuition, a complex eigenvalue with non-zero real and imaginary parts, 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.

Optimization criterion definition

In the context of shape optimization, the idea is to define a criterion for optimization independent of the coefficient of friction μ , such as to reduce the influence of μ in determining the shape. This is because the parameter μ is mostly uncertain in real world and also instabilities could be easily averted at lower values of μ . Hence, to define a criterion which characterizes instability for a geometric shape \mathbf{X} independent of μ , we define the criterion as follows

$$C_s(\mathbf{X}) = \int_{\mu} \max\{\Re(\Lambda(\mathbf{X}, \mu))\} d\mu \quad (2.54)$$

where $\Lambda = \{\lambda_1 \dots \lambda_{(.)}\}$ is a set of eigenvalues of the system. The criterion is essentially a black-box function defined by the maximum of the real part in $\Lambda(\mathbf{X})$ at

a given value of μ , integrated over μ . Typically, it can be too unrealistic or optimistic to minimise the criterion over the whole range of squeal frequency (1 to 16 KHz) and hence, the set Λ can be chosen for a specific range of frequency. This can also be a better strategy in defining meta-model for C_s , since the meta-model can be more accurate in characterising the behaviour of modes over a specific range of frequency than the whole range. Even though, no correlation can be implied between $\max\{\Re(\Lambda(\mathbf{X}, \mu))\}$ and noise level, choosing $\max\{\Re(\Lambda(\mathbf{X}, \mu))\}$ is essential to define some smoothness for C_s in optimisation. Hence the choice of $\max\{\Re(\Lambda(\mathbf{X}, \mu))\}$ does not necessarily characterise noise level but the presence of instability which can be accounted for squeal noise. Nevertheless, the Utopian goal of $C_s = 0$ defines lack of instabilities and hence characterising noise level may not be a concern if such a goal could be reached in optimisation.

Evaluation of C_s can be computationally expensive, but with the aid of model reduction and parallel computing, it can be made to be efficient. In the following, we give the general frame-work for evaluating C_s .

The reduced stiffness matrices of the multi-patch disc can be defined with C&B method as

$$\hat{\mathbf{K}}^{(d)} = \begin{bmatrix} \hat{\mathcal{I}}^{(d)} & 0 \\ 0 & \hat{\mathbf{K}}_{vv}^{(d)} \end{bmatrix} \quad (2.55)$$

where the matrix is defined by the coordinates $\mathbf{Z}^{(d)} = \begin{bmatrix} \mathcal{M}^{(d)} \\ \mathbf{U}_v^{(d)} \end{bmatrix}$, with $\mathbf{U}_v^{(d)}$

defining the degrees of freedom on $\Gamma_C^{(d)}$ ². The matrix is essentially the same in optimisation, when the optimisation is defined only for Ω^p . Hence, for a given definition of shape, the following matrix is computed

$$\hat{\mathbf{K}}^{(p)} = \begin{bmatrix} \hat{\mathcal{I}}^{(p)} & 0 \\ 0 & \hat{\mathbf{K}}_{vv}^{(p)} \end{bmatrix} \quad (2.56)$$

The evaluation of C_s demands the definition of $\hat{\mathbf{K}}^{(d-p)}$ for several values of μ . Since μ is the property of interface, the characteristics at the interface can be decoupled as interface degrees of freedom \mathbf{U}_v in C&B reduced coordinates. For the definition of $\hat{\mathbf{K}}_F^{(a-b)}$, μ can be factored out as $\mu \hat{\mathbf{K}}_F^{(a-b)}|_{\mu=1}$, where in this case with μ factored out, $\hat{\mathbf{K}}_F^{(a-b)}|_{\mu=1}$ can be interpreted as $\hat{\mathbf{K}}_F^{(a-b)}$ computed with $\mu = 1$. The idea is that for the evaluation of C_s , with numerical integration defined over μ , the matrix $\hat{\mathbf{K}}_F^{(a-b)}$ does not need to be evaluated for discrete values of μ , but instead μ can be defined as factor of $\hat{\mathbf{K}}_F^{(a-b)}|_{\mu=1}$, where $\hat{\mathbf{K}}_F^{(a-b)}$ in this case is computed only once with $\mu = 1$.

²It should be noted that in the context of multi-patch parameterisation, detailed in §, $\Gamma_C^{(d)}$ corresponds to $\Gamma_C^{(d_1)}$, where the contact interface is defined to be on $\Omega^{(d_1)}$.

The computational cost of evaluating C_s with numerical integration for discrete values of μ can be reduced through parallel computation. The only varying parameter for parallelisation is μ for the definition of $\hat{\mathbf{K}}_F^{(a-b)}$, hence the computation of the matrices $\hat{\mathbf{M}}^{(p)}$, $\hat{\mathbf{K}}^{(p)}$, $\hat{\mathbf{K}}_C^{(d-p)}$ and $\hat{\mathbf{K}}_F^{(d-p)}|_{\mu=1}$ are achieved on single core. Further with $\hat{\mathbf{M}}^{(d)}$ and $\hat{\mathbf{K}}^{(d)}$ already computed, the reduced matrices of $\hat{\mathbf{M}}^{(d-p)}$ and $\hat{\mathbf{K}}^{(d-p)}$ can also be defined on single core as

$$\hat{\mathbf{M}}^{(d-p)} = \begin{bmatrix} \hat{\mathbf{M}}^{(d)} & 0 \\ 0 & \hat{\mathbf{M}}^{(p)} \end{bmatrix} \quad (2.57)$$

$$\hat{\mathbf{K}}_{\cup C}^{(d-p)} = \begin{bmatrix} \hat{\mathcal{I}}^{(a)} & 0 & 0 & 0 \\ 0 & \hat{\mathbf{K}}_{vv}^{(a)} + \hat{\mathbf{K}}_C^{(d)} & 0 & \hat{\mathbf{K}}_C^{(d,p)} \\ 0 & 0 & \hat{\mathcal{I}}^{(p)} & 0 \\ 0 & \hat{\mathbf{K}}_C^{(p,d)} + \hat{\mathbf{K}}_F^{(p,d)} & 0 & \hat{\mathbf{K}}_{vv}^{(p)} + \hat{\mathbf{K}}_C^{(p)} \end{bmatrix} \quad (2.58)$$

where the matrices are expressed in coordinates $\mathbf{Z}^{(d-p)}$. Similarly, the matrix $\hat{\mathbf{K}}_F^{(d-p)}|_{\mu=1}$ can be expressed in $\mathbf{Z}^{(d-p)}$ coordinates as

$$\hat{\mathbf{K}}_{\cup F}^{(d-p)}|_{\mu=1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \hat{\mathbf{K}}_F^{(d)}|_{\mu=1} & 0 & \hat{\mathbf{K}}_F^{(d,p)}|_{\mu=1} \\ 0 & 0 & 0 & 0 \\ 0 & \hat{\mathbf{K}}_F^{(p,d)}|_{\mu=1} & 0 & \hat{\mathbf{K}}_F^{(p)}|_{\mu=1} \end{bmatrix} \quad (2.59)$$

The evaluation of C_s with numerical integration can be expressed as

$$\int_{\mu} \max\{\Re(\Lambda(\mathbf{X}, \mu))\} d\mu \approx \sum_{\mu_i \in [0,1]} \max\{\Re(\Lambda(\mathbf{X}, \mu_i))\} w_i \quad (2.60)$$

where μ_i can be spaced evenly in the interval $[0, 1]$. Hence, on each parallel core, the matrix $\hat{\mathbf{K}}_{\cup CF}^{(d-p)}$ can be computed as

$$\hat{\mathbf{K}}_{\cup CF}^{(d-p)} = \hat{\mathbf{K}}_{\cup C}^{(d-p)} + \mu_i \hat{\mathbf{K}}_{\cup F}^{(d-p)}|_{\mu=1} \quad (2.61)$$

Along with the definition of $\hat{\mathbf{K}}_{\cup CF}^{(d-p)}$, the computation in parallel cores is defined for Eq. (2.53) in reduced coordinates, where in each core, the computation of the characteristics eigenvalue problem in reduced coordinates can be expressed as

$$(\lambda^2 \hat{\mathbf{M}}^{(d-p)} + \hat{\mathbf{K}}_{\cup C}^{(d-p)} + \mu_i \hat{\mathbf{K}}_{\cup F}^{(d-p)}|_{\mu=1}) \Theta = 0 \quad (2.62)$$

This eventually leads to the evaluation of $\max\{\Re(\Lambda(\mathbf{X}, \mu_i))\}$ in each parallel core and hence with the evaluation of $\max\{\Re(\Lambda(\mathbf{X}, \mu_i))\}$ on all parallel cores, C_s can be computed from $\sum_{\mu_i \in [0,1]} \max\{\Re(\Lambda(\mathbf{X}, \mu_i))\} w_i$.

3 FEM modelisation

In classical FEM, typically the elements are constructed from Lagrange polynomials, known as Lagrange elements, where the interpolating polynomials define unity at the nodes. This property means that the nodes lie on the surface discretised by Lagrange elements, which brings the intuition of Node-to-Node contact where the contact is defined between the nodes of conforming meshes at the contact interface. Node-to-Node contact can be essentially considered as the collocation method where the strong form of contact or friction definition is satisfied at the nodes in Γ_C . The area effects can also be partly considered through Isoparametric mapping even though it may not be precise.

Given that we focus on shape optimisation, Node-to-Node contact can place severe constraints in meshing, where structured meshing must be preferred to define confirming meshes at the contact interface. Depending on the design space in shape optimisation, this may not be a very robust strategy, since structured meshing can severely restrict mesh adaptation to avoid distorted elements that can effect the Isoparametric definition of integrals. On the other hand, it is well known that unstructured mesh definition may not also lead to robust meshing with classical FEM. Due to such complications with meshing, Isogeometric approach can be considered by choosing a robust parameterization strategy which would be rather difficult with classical FEM. But also the Node-to-Node contact can not be explicitly defined with Isogeometric approach, since the control points which correspond to nodes in classical FEM may not lie on the surface. Rather, the collocation may not have to be on the nodes which is typically preferred in classical FEM, but else where on the surface. Hence, the collocation scheme in classical FEM typically defined with Node-to Node contact takes a different form with Isogeometric approach, where the collocation can be defined on the surface or implicitly in the knot span. We also remind that defining an initial Parametrisation may also be cumbersome with Isogeometric approach but given a well-defined initial parameterisation, Isogeometric approach can achieve robust refinement. Although the complications with meshing for classical FEM can be avoided with contact formulations that do not demand confirming meshes at the interface, developing such formulations with Isogeometric approach can be even more advantageous in shape optimisation.

The effect of contact formulation to the prediction of instabilities with CEA has not been largely studied, where the interest is on the contact characteristics pertaining to the dynamics of the perturbation $\tilde{\mathbf{u}}$ rather than \mathbf{u}_{eq} (2.23). Nevertheless, we develop a more rigorous formulation with Isogeometric approach in §, purely for

its advantages in optimisation, where it does not require confirming meshes at the interface – since also defining confirming meshes can be even more cumbersome with Isogeometric approach due to the tensor product nature of NURBS.

We define optimisation of simple shapes through classical FEM discretisation, where the definition of simple shapes vow for a robust structured meshing and hence Node-to-Node contact can be preferred for its simplicity. With the following explanations, we detail the Node-to-Node collocation method to model the contact and the friction definitions in the problem (2.29), where we also show the relation of collocation method to the weak form of contact and friction terms in (4.7). It should be noted that this type of contact and friction definitions are defined with approximations specific for modelling flutter-type dynamic instability with CEA, detailed in §. Considering Eq. (4.7), for classical FEM with Lagrange elements, the space ${}_h\mathbf{V}$ is defined by the bases ${}_h\mathbf{v}_i$ of Lagrange polynomials.

Contact formulation

The contact definition of the initial-boundary value problem Eq. (2.29) can be defined in finite element context as

$${}_h\sigma_n^{(k)}({}_h\tilde{\mathbf{u}}^{(k)}) = -p_h \tilde{u}_n^{(k)} \quad \text{on} \quad \Gamma_C^{(k)} \quad (3.1)$$

where for a system with two domains $\Omega^{(a)}$ and $\Omega^{(b)}$ in contact. With confirming mesh at $\Gamma_C^{(a)}$ and $\Gamma_C^{(b)}$, it can be said that for any node $i \in \Gamma_C^{(a)}$, there exists a unique node $j \in \Gamma_C^{(b)}$ that forms contact. Hence, the contact force for a given node $i \in \Gamma_C^{(a)}$ in contact with a node $j \in \Gamma_C^{(b)}$ can simply be expressed with Node-to-Node contact as

$$p({}_h\tilde{\mathbf{u}}^{(a)} - {}_h\tilde{\mathbf{u}}^{(b)}).\hat{\mathbf{v}}_n \Big|_{\substack{{}_h\mathbf{v}_i^{(a)} = {}_h\mathbf{v}_i^{(b)} \\ = 1}} = p(\tilde{\mathbf{u}}_i^{(a)} - \tilde{\mathbf{u}}_j^{(b)}).\hat{\mathbf{v}}_n = {}_h t_n^{(k)} \quad (3.2)$$

where ${}_h\tilde{\mathbf{u}}^{(a)} = \sum_{\forall i \in \Omega^{(a)}} {}_h\mathbf{v}_i^{(a)} \tilde{\mathbf{u}}_i^{(a)}$ and ${}_h\tilde{\mathbf{u}}^{(b)} = \sum_{\forall j \in \Omega^{(b)}} {}_h\mathbf{v}_j^{(b)} \tilde{\mathbf{u}}_j^{(b)}$. Since the collocation is defined at the nodes itself, the bases ${}_h\mathbf{v}_i^{(a)} = {}_h\mathbf{v}_i^{(b)} = 1$ for Lagrange elements or typically for elements in classical FEM. The above equation is stated specifically for linear case, where normal compliance terms are expressed to be linear.

As an alternate interpretation, the collocation method can also be defined from the weak form of contact (2.34) as

$$\begin{aligned} \langle {}_h\boldsymbol{\sigma}_n^{(a)}, {}_h\mathbf{v}_i^{(a)} \rangle_{\Gamma_C^{(a)}} &= \int_{\Gamma_C^{(a)}} p[({}_h\tilde{\mathbf{u}}^{(a)} - {}_h\tilde{\mathbf{u}}^{(b)}).\hat{\mathbf{v}}_n] {}_h\mathbf{v}_i^{(a)}.\hat{\mathbf{v}}_n \, d\Gamma_C^{(a)} \rightarrow \\ &\quad p(\tilde{\mathbf{u}}_i^{(a)} - \tilde{\mathbf{u}}_j^{(b)}).\hat{\mathbf{v}}_n \end{aligned} \quad (3.3)$$

if the weighting function ${}_h\mathbf{v}_i^{(a)}$ of the weak form is defined to be the Dirac-delta function ($\delta_D(\cdot)$) as ${}_h\mathbf{v}_i^{(a)} = \delta_D(\tilde{\mathbf{u}}_i^{(a)} - \tilde{\mathbf{u}}^{(a)})$ ¹.

As aforementioned, we do not focus on the physical characteristics of contact stiffness in the light of normal compliance, and its subsequent effect on CEA, where we define contact stiffness purely as a penalty coefficient p . But to proceed with the following discussions, we must make a strong assumption on the existence of global contact stiffness (p_G) such that this property is independent of the contact interface area. This is because, if p is a parameter of differential area determined from experiments, it cannot be taken in to account with Node-to-Node contact unless area is implicitly considered. This is largely the case if contact stiffness has correlation with contact pressure at the interface. It is only safe to say that experimental studies can alone determine if p is a global or local parameter, such that to eliminate any bias of contact stiffness in defining instability for shape optimisation. Nevertheless, with this strong proposition, contact stiffness between any two spatially corresponding nodes is defined by local contact stiffness p_l , defined as

$$p_l = \frac{p_G}{m} \quad (3.4)$$

where m is the number of contact nodes at the interface.

Similarly, the friction definition of the initial-boundary value problem Eq. (2.29) can be defined in finite element context as

$${}_h\sigma_t^{(k)}({}_h\tilde{\mathbf{u}}^{(k)}) = \mu p {}_h\tilde{u}_n^{(k)} \quad \text{on } \Gamma_C^{(k)} \quad (3.5)$$

Similar to the definition of contact force, the tangential force of friction can be defined with Node-to-Node collocation as

$$\mu p({}_h\tilde{\mathbf{u}}^{(a)} - {}_h\tilde{\mathbf{u}}^{(b)}).{}_{\hat{\mathbf{v}}_n} \Big|_{{}_{h\mathbf{v}_i^{(a)}} = {}_{h\mathbf{v}_j^{(b)}} = 1} = \mu p(\tilde{\mathbf{u}}_i^{(a)} - \tilde{\mathbf{u}}_j^{(b)}).{}_{\hat{\mathbf{v}}_n} \quad (3.6)$$

where vectorially, the friction force can be expressed as $\mu[p(\tilde{\mathbf{u}}_i^{(a)} - \tilde{\mathbf{u}}_j^{(b)}).{}_{\hat{\mathbf{v}}_t}]{}_{\hat{\mathbf{v}}_t}$, with $\hat{\mathbf{v}}_t$ for a given node being known a priori. For the $d-p$ system, the friction force can be resolved in to tangential and radial components relative to the disc axis. Since, the relative sliding velocity is negligible in the radial direction, the frictional force component of the radial direction is also negligible and hence often ignored.

Mesh sensitivity

In relation to the expensive evaluation of the stability criteria, mesh convergence study was performed for the influence of mesh at the contact interface and outside the contact interface. The eigenmodes reflecting maximum instability as predicted

¹ $\delta_D(\tilde{\mathbf{u}}_i^{(a)} - \tilde{\mathbf{u}}^{(a)}) : {}_h\mathbf{v}_i^{(a)} = 0, \forall \tilde{\mathbf{u}}_i^{(a)} \neq \tilde{\mathbf{u}}^{(a)}$ and ${}_h\mathbf{v}_i^{(a)} = 1$ if $\tilde{\mathbf{u}}_i^{(a)} = \tilde{\mathbf{u}}^{(a)}$

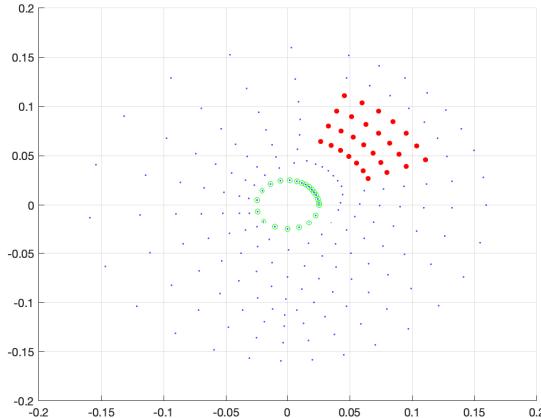


Figure 3.1: Node plot for a coarse mesh of the disc geometry with contact nodes represented in red

by the maximum positive real part of the complex eigenvalues are only of interest to us and hence taken in to account to check for convergence. It should be noted that the contact formulation can have some influence on mesh convergence especially for CEA, where no clear studies has been performed. In this case, we are specific for convergence with for node-to-node formulation.

The dynamic properties with respect to maximum instability show little variation for change in mesh size out of the contact region for a given shape and number of contact points. The comparison is shown through a model with highly coarse mesh as in Fig. 3.1 against a model with the same shape but for a relatively fine mesh as in Fig. 3.2, while maintaining the same number and position for the contact points. The mismatch in frequencies between the models is shown in Fig. 3.3 where the range for frequency is zoomed to a pair of frequencies which induce mode coalescence predicted to cause the maximum instability. The shift in unstable frequencies and the point of bifurcation of the maximum real part as shown in Fig. 3.4 are observed to be very low for change in mesh size.

Though the variation of the mesh out of the contact interface is shown to have a little influence on the maximum instability, the variation of mesh at the contact interface is observed to have a considerable effect on the dynamic properties. This can be seen by comparing results of the models in Fig. 3.1 and Fig. 3.2 against the models in Fig. 3.5 and Fig. 3.6 which are of the same shape. Hence, it is intuitive to assume a large number of contact points, since the contact interface is a continuum after all. The convergence is shown with large number of contact points in Fig. 3.5 and Fig. 3.6 with plot for bifurcation of the real part in Fig. 3.8) and frequencies inducing maximum instability in Fig. 3.7.

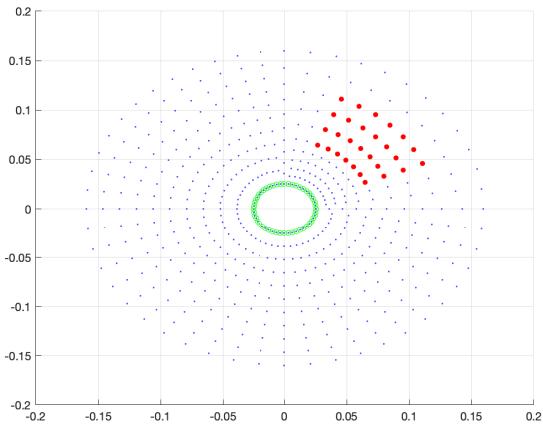


Figure 3.2: Node plot for a relatively fine structured mesh of the disc geometry with contact nodes represented in red

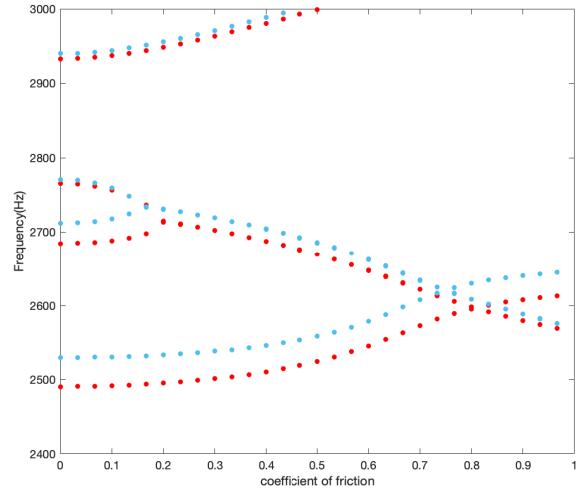


Figure 3.3: Plot of Frequency vs Friction coefficient, of modes showing maximum instability; Blue represents for the model in 3.1; Red represents for the model in 3.2

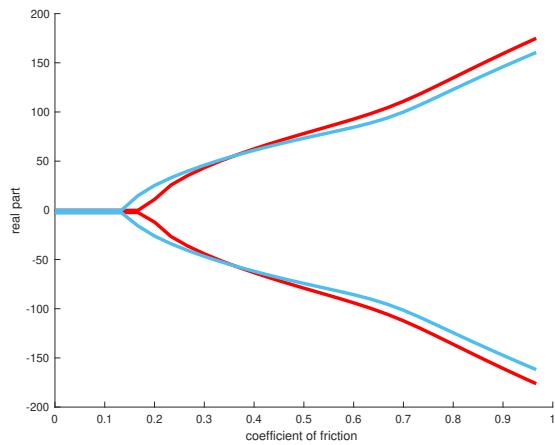


Figure 3.4: Plot of Real part of the complex eigenvalues vs Friction coefficient, of modes showing maximum instability; Blue represents for the model in 3.1; Red represents for the model in 3.2

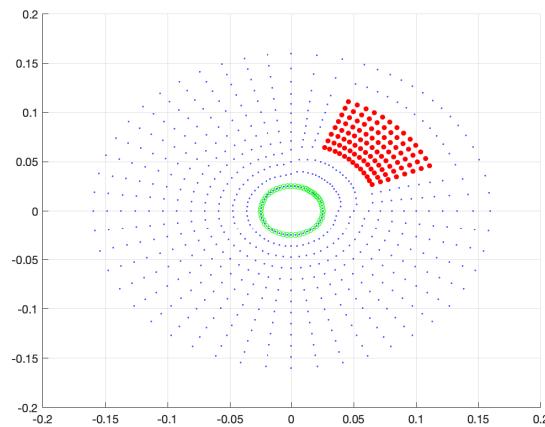


Figure 3.5: Node plot for a fine mesh with contact nodes represented in red

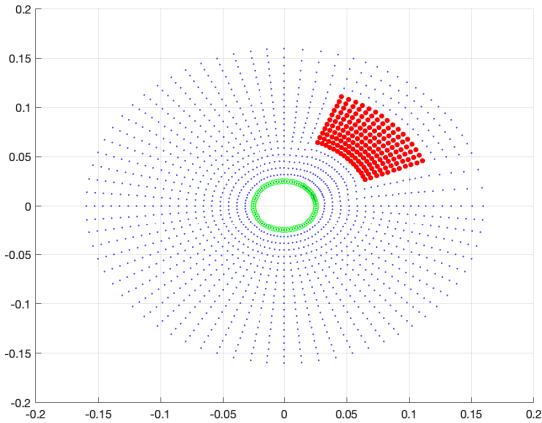


Figure 3.6: Node plot for a relatively finer mesh compared to 3.5, especially on the contact interface with contact nodes represented in red

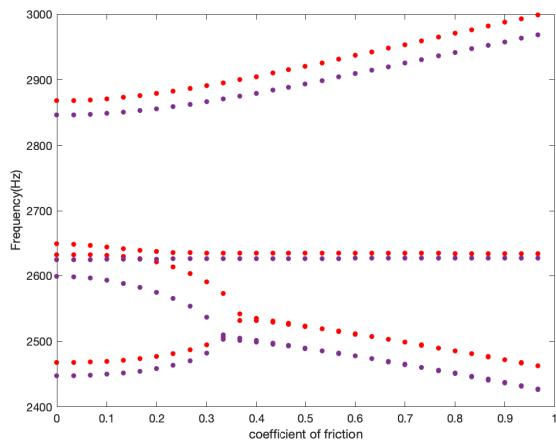


Figure 3.7: Plot of Frequency vs Friction coefficient, of modes showing maximum instability; Red represents the plot for the model in 3.5; Violet represents the plot for the model in 3.6

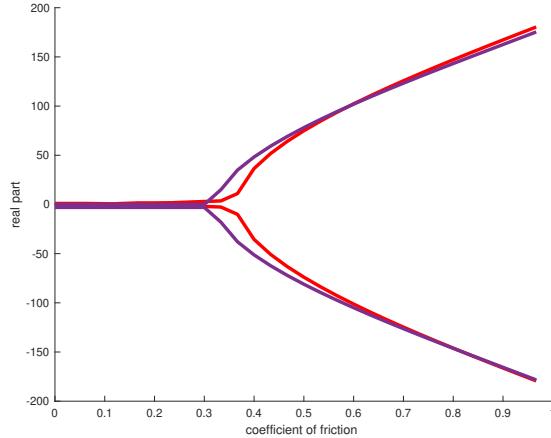


Figure 3.8: Plot of Real part of the complex eigenvalues vs Friction coefficient, of modes showing maximum instability; Red represents the plot for the model in 3.5; Violet represents the plot for the model in 3.6

The definition of the contact points can be observed to have significant influence on the maximum real part of the complex eigenvalues and hence also the stability criteria, which demands a good mesh definition to define a robust stability criteria in optimisation. For this reason, we defined a structured mesh as in Fig. 3.9 where linear hexahedral elements are largely used through out the model with smaller elements at the contact interface, while the region outside of contact interface is defined by larger elements. The difference in mesh density is compromised by introducing 3D wedge elements to maintain a structured mesh.

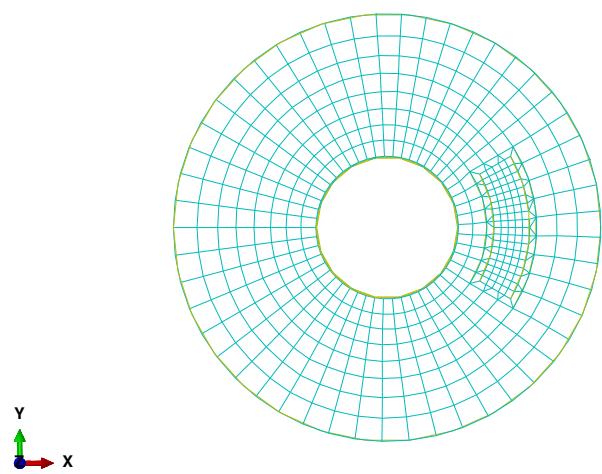


Figure 3.9: Considered mesh definition

4 IGA formulation

[1] We start with the definition of B-spline functions with extension to B-spline curves, from which NURBS curves are introduced with the definition of a weighing parameter and a non-uniform knot vector. 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,

$$N_{i,0}(\xi) = \begin{cases} 1 & \xi_i \leq \xi < \xi_{i+1} \\ 0 & otherwise \end{cases} \quad (4.1)$$

$$N_{i,p}(\xi) = \frac{\xi - \xi_i}{\xi_{i+p} - \xi_i} N_{i,p-1}(\xi) + \frac{\xi_{i+p+1} - \xi}{\xi_{i+p+1} - \xi_{i+1}} N_{i+1,p-1}(\xi) \quad (4.2)$$

where p is defined recursively for $p > 0$ to obtain a curve of degree p , which starts with a piecewise constant at $p = 0$. Naturally, a uniform knot vector can be defined as $\boldsymbol{\xi} = \{\xi_1, \xi_2, \dots, \xi_{n+p+1}\}$, where any $\xi_i - \xi_{i+1}$ is uniformly spaced. For a uniform knot vector, the bases span with continuity C^{p-1} between the knots, where it satisfies partition of unity $\sum_{i=1}^n N_{i,p}(\xi) = 1$ for $[\xi_p, \xi_{n+1}]$, with n being the number of control points. Further, the span of any $N_{i,p}(\xi)$ is defined in $[\xi_i, \xi_{i+p+1}]$, and $N_{i,p}(\xi) \geq 0, \forall \xi$.

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 a non-uniform knot vector. The multiplicity $\mathcal{M} = p$ for the first knot and the last knot defines an open knot vector, where the basis functions model interpolation between the first and the last knots. The basis functions defined with an open knot vector satisfies partition of unity $\forall \xi$. Through B-spline basis functions and a knot vector $\boldsymbol{\xi} = \{\xi_1, \dots, \xi_{n+p+1}\}$, a B-spline curve can be defined with coefficients of the basis functions as follows

$$\mathbf{X}_c(\xi) = \sum_{i=1}^n N_{i,p}(\xi) \mathbf{P}_i \quad (4.3)$$

where with an open knot vector for a curve, the ends of the curve are C^0 . The coefficients $\mathbf{P}_i \in \mathbb{R}^d$ are the control points, with d being the dimension of the space. The definition of a weighing parameter $w_i > 0$ associated with a respective basis function N_i , normalized defines rational B-splines where it respects the partition of

unity, given as follows

$$\mathbf{X}_c(\xi) = \sum_{i=1}^n \underbrace{\frac{w_i N_{i,p}(\xi)}{\sum_{i=0}^n w_i N_{i,p}(\xi)}}_{R_{i,p}} \mathbf{P}_i \quad (4.4)$$

The parameter w_i provides a new dimension for controlling the geometry through projective transformation, while the affine transformation is achieved by \mathbf{P}_i . Hence, the combination of non-uniform knot vectors and rational basis functions define NURBS. Further, if all weights are the same, NURBS is simply a B-spline with non-uniform knot vector.

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

$$\mathbf{X}_s(\xi, \eta) = \sum_{i=1}^n \sum_{j=1}^m R_{i,p}(\xi) R_{j,q}(\eta) \mathbf{P}_{i,j} \quad (4.5)$$

which is supported by knot vectors $\boldsymbol{\xi} = \{\xi_1, \dots, \xi_{n+p+1}\}$ and $\boldsymbol{\eta} = \{\eta_1, \dots, \eta_{m+q+1}\}$, for the domain $[\xi_1, \xi_{m+q+1}] \times [\eta_1, \eta_{m+q+1}]$, with $n \times m$ net of control points $\mathbf{P}_{i,j}$. Similarly, to define volume, the tensor product NURBS volume is defined as follows

$$\mathbf{X}_v(\xi, \eta, \zeta) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l \underbrace{R_{i,p}(\xi) R_{j,q}(\eta) R_{k,r}(\zeta)}_{R_{i,j,k}(\boldsymbol{\Xi})} \mathbf{P}_{i,j,k} \quad (4.6)$$

where the knot vectors are given as $\boldsymbol{\xi} = \{\xi_1, \dots, \xi_{n+p+1}\}$, $\boldsymbol{\eta} = \{\eta_1, \dots, \eta_{m+q+1}\}$ and $\boldsymbol{\zeta} = \{\zeta_1, \dots, \zeta_{l+r+1}\}$.

The above expression can be simply expressed in matrix form as $\mathbf{X}_v(\boldsymbol{\Xi}) = \mathbf{R}(\boldsymbol{\Xi})\mathbf{P}$. where,
 $\mathbf{R}(\boldsymbol{\Xi})$

IGA discretization

We defined the general view of the space ${}_hV$ and now we give a more precise definition of the space with the Isogeometric approach. The main idea with Isogeometric approach is to define ${}_hV$ as the space of the NURBS basis functions which also parameterizes the geometry. The parameterization of a domain $\Omega \in \mathbb{R}^3$ as an initial geometric description through NURBS can be defined as $\check{\mathbf{X}}_v^{(k)}(\check{\boldsymbol{\Xi}}^{(k)}) = \check{\mathbf{R}}^{(k)}(\check{\boldsymbol{\Xi}}^{(k)})\check{\mathbf{P}}^{(k)}$, $\mathbf{X} : \hat{\Omega} \rightarrow \Omega$, where \mathbf{X} 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: $[\xi_1, \dots, \xi_{n+p+1}] \times [\eta_1, \dots, \eta_{m+q+1}] \times [\zeta_1, \dots, \zeta_{l+r+1}]$. The

analysis-suitable parameterization \mathbf{X} ¹ can be achieved through the refinement of $\check{\mathbf{X}} \rightarrow \mathbf{X}$ with one or several of the refinement methods (h , p and k), where \mathbf{X} can be defined as $\mathbf{X}_v(\boldsymbol{\Xi}) = \mathbf{R}(\boldsymbol{\Xi})\mathbf{P}$ to take in to account of the modified knot vectors and control points – more on parameterization and refinement for our applicative example of disc-pad system is discussed in.

The Isogeometric approach for approximation of the solution \mathbf{u}_k is achieved through the same NURBS bases $R_{i,j,k}$, where for the vector-valued function space $_h\mathbf{V}$, the vectorial definition of the bases $\mathbf{R}_{i,j,k} \in \mathbb{R}^3$ can be defined as

$$\left\{ \begin{bmatrix} R_{i,j,k} \\ 0 \\ 0 \end{bmatrix} \right\} \cup \left\{ \begin{bmatrix} 0 \\ R_{i,j,k} \\ 0 \end{bmatrix} \right\} \cup \left\{ \begin{bmatrix} 0 \\ 0 \\ R_{i,j,k} \end{bmatrix} \right\}$$

where in matrix form, $\mathbf{R}_{i,j,k}(\boldsymbol{\Xi}) := \begin{bmatrix} R_{i,j,k}(\boldsymbol{\Xi}) & 0 & 0 \\ 0 & R_{i,j,k}(\boldsymbol{\Xi}) & 0 \\ 0 & 0 & R_{i,j,k}(\boldsymbol{\Xi}) \end{bmatrix}$

which is taken in to account through the definition of the matrix $\mathbf{R}(\boldsymbol{\Xi})$ and \mathbf{P} as

$$\mathbf{R}(\boldsymbol{\Xi}) = [\mathbf{R}_{1,1,1}(\boldsymbol{\Xi}) \quad \cdots \quad \mathbf{R}_{n,m,l}(\boldsymbol{\Xi})]$$

$$\mathbf{P} = [P_{1,1,1}^x \quad P_{1,1,1}^y \quad P_{1,1,1}^z \quad \cdots \quad P_{n,m,l}^x \quad P_{n,m,l}^y \quad P_{n,m,l}^z]^T$$

In a abstract sense, the bases $\mathbf{R}_{i,j,k}(\boldsymbol{\Xi})$ in parametric space is transformed to the bases $\phi_{i,j,k}(x, y, z)$ in physical space using the push-forward operator \circ , where the bases ϕ_i is defined with the property $\phi_i = \phi_{i,j,k}(\mathbf{X}) = \mathbf{R}_{i,j,k}(\boldsymbol{\Xi}) \circ \mathbf{X}^{-1}$. Hence, the approximation of a field variable on Ω is defined through all the bases ϕ_i spanning the finite-dimensional function space Φ . Considering Eq. with the Isogeometric approach, the finite-dimensional space $_h\mathbf{V} \rightarrow \Phi$, and its associated bases $_h\mathbf{v}_i \rightarrow \phi_i$. The approximation of $\tilde{\mathbf{u}} \in \Phi$ can be defined as $_h\tilde{\mathbf{u}} = \sum_{\forall i \in \Omega} \phi_i \tilde{\mathbf{u}}_i$, expressed in matrix form as $_h\tilde{\mathbf{u}} = \mathbf{N}(\mathbf{X})\mathbf{U}$, where

$$\mathbf{N}(\mathbf{X}) = [\phi_i(\mathbf{X}) \quad \cdots \quad \phi_{n \times m \times l}(\mathbf{X})]$$

$$\mathbf{U} = [U_i^x \quad U_i^y \quad U_i^z \quad \cdots \quad U_{n \times m \times l}^x \quad U_{n \times m \times l}^y \quad U_{n \times m \times l}^z]^T$$

When $\Gamma_C = \emptyset$, the Eq. can be simply expressed in matrix form as

¹For simplicity of the notation, we define \mathbf{X} to be the default notation for analysis-suitable parameterization of a domain $\Omega \in \mathbb{R}^3$

$$\sum_{k=1}^{n_k} \left\{ \int_{\Omega^{(k)}} \rho^{(k)} {}_h \tilde{\mathbf{u}}^{(k)} \cdot {}_h \mathbf{v}_i^{(k)} + \int_{\Omega^{(k)}} \boldsymbol{\sigma}^{(k)}({}_h \tilde{\mathbf{u}}^{(k)}) : {}_h \mathbf{v}_i^{(k)} \right. \\ \left. + \int_{\Gamma_C^{(k)}} p[({}_h \tilde{\mathbf{u}}^{(k)} - {}_h \tilde{\mathbf{u}}^{(\sim k)}) \cdot \hat{\mathbf{v}}_n] {}_h \mathbf{v}_i^{(k)} \cdot \hat{\mathbf{v}}_n - \int_{\Gamma_C^{(k)}} \mu p[({}_h \tilde{\mathbf{u}}^{(k)} - {}_h \tilde{\mathbf{u}}^{(\sim k)}) \cdot \hat{\mathbf{v}}_n] {}_h \mathbf{v}_i^{(k)} \cdot \hat{\mathbf{v}}_k \right\} = 0 \quad \forall_h \mathbf{v}_i \in {}_h \mathbf{V} \quad (4.7)$$

For the following explanation, we consider contact between two domains Ω_a and Ω_b , where the formulation for contact and friction are given in Eq.(2.34) and Eq.(2.35) respectively. The parameterization of the domains $\Omega^{(a)}$ and $\Omega^{(b)}$ defined through NURBS can be expressed as $\mathbf{X}^{(a)} = \mathbf{R}^{(a)}(\boldsymbol{\Xi}^{(a)})$ and $\mathbf{X}^{(b)} = \mathbf{R}^{(b)}(\boldsymbol{\Xi}^{(b)})$. For the perturbed displacement field $\tilde{\mathbf{u}}$ around a equilibrium \mathbf{u}_{eq} , Γ_C was hypothesized to be stationary, where the effect of $\tilde{\mathbf{u}}$ for a stationary Γ_C was modelled through the normal compliance approach. Hence, Γ_C is known a prior from the solution \mathbf{u}_{eq} in solving for an equilibrium configuration. Further, $\Gamma_C : g_n = 0$, i.e., $\mathbf{X}^{(a)} \cdot \hat{\mathbf{v}}_n = \overleftarrow{\mathbf{X}}^{(b)} \cdot \hat{\mathbf{v}}_n$, where $\hat{\mathbf{v}}_n$ in this case is taken to be the outward normal projection from the slave side $\Gamma^{(a)}$ to the master side $\Gamma^{(b)}$. This means that $\overleftarrow{\mathbf{X}}^{(b)} : \overleftarrow{\mathbf{X}}^{(b)}(\mathbf{X}^{(a)})$, where for $\mathbf{X}^{(a)}$ that parametrizes $\Gamma_C^{(a)}$, a projection exists that maps $\mathbf{X}^{(a)}$ on $\Gamma_C^{(b)}$ as $\overleftarrow{\mathbf{X}}^{(b)}$. For the following explanations, we detail the derivation of traction forces on $\Gamma^{(a)}$ which also similarly applies for $\Gamma^{(b)}$. The approximation of $\langle h\boldsymbol{\sigma}_n^{(a)}, \phi_i^{(a)} \rangle_{\Gamma_C^{(a)}}$ and $\langle h\boldsymbol{\sigma}_t^{(a)}, \phi_i^{(a)} \rangle_{\Gamma_C^{(a)}}$ in the function space Φ can be defined as

$$\langle h\boldsymbol{\sigma}_n^{(a)}, \phi_i^{(a)} \rangle_{\Gamma_C^{(a)}} = \\ \int_{\Gamma_C^{(a)}} p[(\mathbf{N}^{(a)}(\mathbf{X}^{(a)})\mathbf{U}^{(a)} - \mathbf{N}^{(b)}(\overleftarrow{\mathbf{X}}^{(b)})\mathbf{U}^{(b)}) \cdot \hat{\mathbf{v}}_n] \phi_i^{(a)} \cdot \hat{\mathbf{v}}_n \, d\Gamma_C^{(a)} \quad \forall \phi_i^{(a)} \in \Phi^{(a)} \quad (4.8)$$

$$\langle h\boldsymbol{\sigma}_t^{(a)}, \phi_i^{(a)} \rangle_{\Gamma_C^{(a)}} = \\ \int_{\Gamma_C^{(a)}} \mu p[(\mathbf{N}^{(a)}(\mathbf{X}^{(a)})\mathbf{U}^{(a)} - \mathbf{N}^{(b)}(\overleftarrow{\mathbf{X}}^{(b)})\mathbf{U}^{(b)}) \cdot \hat{\mathbf{v}}_n] \phi_i^{(a)} \cdot \hat{\mathbf{v}}_t \, d\Gamma_C^{(a)} \quad \forall \phi_i^{(a)} \in \Phi^{(a)} \quad (4.9)$$

where in matrix form,

$$\phi_i \cdot \hat{\mathbf{v}} := \begin{bmatrix} \phi_{i,j,k}(\mathbf{X}) \hat{v}^x & 0 & 0 \\ 0 & \phi_{i,j,k}(\mathbf{X}) \hat{v}^y & 0 \\ 0 & 0 & \phi_{i,j,k}(\mathbf{X}) \hat{v}^z \end{bmatrix}$$

The expression for $\langle h\boldsymbol{\sigma}_n^{(a)}, \phi_i^{(a)} \rangle_{\Gamma_C^{(a)}}$ and $\langle h\boldsymbol{\sigma}_t^{(a)}, \phi_i^{(a)} \rangle_{\Gamma_C^{(a)}}$ can be further expanded as

$$\begin{aligned} \langle {}_h\boldsymbol{\sigma}_n^{(a)}, \boldsymbol{\phi}_i^{(a)} \rangle_{\Gamma_C^{(a)}} = & \\ & \int_{\Gamma_C^{(a)}} p[(\boldsymbol{\phi}_i^{(a)} \cdot \hat{\mathbf{v}}_n) (\mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_n) - (\boldsymbol{\phi}_i^{(a)} \cdot \hat{\mathbf{v}}_n) (-\mathbf{N}^{(b)} \cdot \hat{\mathbf{v}}_n)] \mathbf{U}^{(a,b)} d\Gamma_C^{(a)} \\ & \forall \boldsymbol{\phi}_{i \in \Gamma_C^{(a)}} \in \Phi^{(a)} \quad (4.10) \end{aligned}$$

$$\begin{aligned} \langle {}_h\boldsymbol{\sigma}_t^{(a)}, \boldsymbol{\phi}_i^{(a)} \rangle_{\Gamma_C^{(a)}} = & \\ & \int_{\Gamma_C^{(a)}} \mu p[(\boldsymbol{\phi}_i^{(a)} \cdot \hat{\mathbf{v}}_t) (\mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_t) - (\boldsymbol{\phi}_i^{(a)} \cdot \hat{\mathbf{v}}_t) (-\mathbf{N}^{(b)} \cdot \hat{\mathbf{v}}_t)] \mathbf{U}^{(a,b)} d\Gamma_C^{(a)} \\ & \forall \boldsymbol{\phi}_{i \in \Gamma_C^{(a)}} \in \Phi^{(a)} \quad (4.11) \end{aligned}$$

where

$$\mathbf{N} \cdot \hat{\mathbf{v}} := [\phi_1(\mathbf{X}) \cdot \hat{\mathbf{v}} \quad \cdots \quad \phi_{n \times m \times l}(\mathbf{X}) \cdot \hat{\mathbf{v}}]$$

$$\mathbf{U}^{(a,b)} = \begin{bmatrix} \mathbf{U}^{(a)} \\ \mathbf{U}^{(b)} \end{bmatrix}$$

We expand the terms of the form $\int_{\Gamma_C^{(a)}} \boldsymbol{\phi}_i^{(a)} \mathbf{N}^{(b)}(\mathbf{X}^{(b)}) d\Gamma_C^{(a)}$ ² in Eq. (4.10) and Eq. (4.11), given as

$$\begin{aligned} \int_{\Gamma_C^{(a)}} \boldsymbol{\phi}_i^{(a)} \mathbf{N}^{(b)}(\mathbf{X}^{(b)}) d\Gamma_C^{(a)} = & \\ & \left[\int_{\Gamma_C^{(a)}} \boldsymbol{\phi}_i^{(a)}(\mathbf{X}^{(a)}) \cdot \boldsymbol{\phi}_1^{(b)}(\overleftarrow{\mathbf{X}}^{(b)}(\mathbf{X}^{(a)})) d\Gamma_C^{(a)} \quad \cdots \right. \\ & \left. \int_{\Gamma_C^{(a)}} \boldsymbol{\phi}_i^{(a)}(\mathbf{X}^{(a)}) \cdot \boldsymbol{\phi}_{n \times m \times l}^{(b)}(\overleftarrow{\mathbf{X}}^{(b)}(\mathbf{X}^{(a)})) d\Gamma_C^{(a)} \right] \quad (4.12) \end{aligned}$$

where the integral is simultaneously defined over the bases of the two contact domains, since $\boldsymbol{\phi}^{(a)} \in H^{-1/2}(\Gamma^{(a)})_C$ and $\boldsymbol{\phi}^{(b)} \in H^{-1/2}(\Gamma^{(b)})_C$. Even though the definition of integral is possible for $\boldsymbol{\phi}_1^{(b)}(\overleftarrow{\mathbf{X}}^{(b)}(\mathbf{X}^{(a)}))$ on $\Gamma_C^{(a)}$, for dissimilar meshes at the contact interface, the definition of numerical quadrature scheme for the integral demands domain decomposition to find the common span: $\boldsymbol{\phi}_{i \in \Gamma_C^{(a)}}^{(a)} \cap \boldsymbol{\phi}_{j \in \Gamma_C^{(b)}}^{(b)}$. This means that

²For simplicity of the expansion, we ignore the unit vectors $\hat{\mathbf{v}}_n$ and $\hat{\mathbf{v}}_t$

the integral can only be defined through a quadrature scheme specific on the span of $\phi_{i \in \Gamma_C^{(a)}}^{(a)}$ or $\phi_{j \in \Gamma_C^{(b)}}^{(b)}$ for which the projection $\phi_{i \in \Gamma_C^{(a)}}^{(a)} \phi_{j \in \Gamma_C^{(b)}}^{(b)} \neq 0$. Alternatively, this can be viewed as the projection of $\phi_{i \in \Gamma_C^{(a)}}^{(a)}$ on $\phi_{j \in \Gamma_C^{(b)}}^{(b)}$ for which the relation of weak sense should hold, given as

$$\int_{\Gamma_C^{(a)}} [\phi_i^{(a)} \cdot \phi_1^{(a)} + \dots + \phi_i^{(a)} \cdot \phi_{n \times m \times l}^{(a)}] d\Gamma_C^{(a)} = \int_{\Gamma_C^{(a)}} [\phi_i^{(a)} \cdot \phi_1^{(b)} + \dots + \phi_i^{(a)} \cdot \phi_{n \times m \times l}^{(b)}] d\Gamma_C^{(a)} \quad (4.13)$$

where it verifies the conservation of momentum at the contact interface. We satisfy the relation in an approximate sense, where we consider the integral $\int_{\Gamma_C} \phi_i d\Gamma_C$ on one of the domains – in this case $\Gamma^{(a)}$ – through collocation defined as $\int_{\Gamma_C^{(a)}} \phi_i^{(a)} d\Gamma_C^{(a)} \rightarrow \sum_{\forall i \in \mathbf{I}^{(a)}} {}^i \phi_i^{(a)}$ where $\mathbf{I}^{(a)}$ is the set of points on $\Gamma_C^{(a)}$ which depends on the collocation scheme. Hence, for Eq. (4.13), the integral for the projection of $\phi_i^{(a)}$ on the bases in $H^{-1/2}(\Gamma_C^{(a)})$ and $H^{-1/2}(\Gamma_C^{(b)})$ can be given through collocation as

$$\sum_{\forall i \in \mathbf{I}^{(a)}} [{}^i \phi_i^{(a)} \cdot {}^i \phi_1^{(a)} + \dots + {}^i \phi_i^{(a)} \cdot {}^i \phi_{n \times m \times l}^{(a)}] = \sum_{\forall i \in \mathbf{I}^{(a)}} [{}^i \phi_i^{(a)} \cdot {}^i \phi_1^{(b)} + \dots + {}^i \phi_i^{(a)} \cdot {}^i \phi_{n \times m \times l}^{(b)}] \quad (4.14)$$

where ${}^i \phi^{(a)} := \phi^{(a)}({}^i \mathbf{X}^{(a)})$ and ${}^i \phi^{(b)} := \phi^{(b)}(\overleftarrow{\mathbf{X}}^{(b)}({}^i \mathbf{X}^{(a)}))$. This implicitly satisfies the conditions for conservation of momentum even though the integral $\int_{\Gamma_C^{(a)}} (\phi_i^{(a)})(\phi_1^{(b)}) d\Gamma_C^{(a)}$ may not be defined accurately. But this can effect the continuity of the solution, which is typically verified through Patch-test. For any i , the following relation also holds

$$[{}^i \phi_i^{(a)} \cdot {}^i \phi_1^{(a)} + \dots + {}^i \phi_i^{(a)} \cdot {}^i \phi_{n \times m \times l}^{(a)}] = [{}^i \phi_i^{(a)} \cdot {}^i \phi_1^{(b)} + \dots + {}^i \phi_i^{(a)} \cdot {}^i \phi_{n \times m \times l}^{(b)}] = {}^i \phi_i^{(a)} \quad (4.15)$$

This means that any quantity defined through collocation at i over ${}^i \phi_i^{(a)}$ is projected equally over the bases in $H^{-1/2}(\Gamma_C^{(a)})$ and $H^{-1/2}(\Gamma_C^{(b)})$. It should be noted that the collocation strategy can be replaced by a numerical quadrature scheme as $\int_{\Gamma_C^{(a)}} \phi_i^{(a)} d\Gamma_C^{(a)} \approx \sum_{\forall i \in \mathbf{I}^{(a)}} {}^i w {}^i \phi_i^{(a)}$ where $\mathbf{I}^{(a)}$ in this case corresponds to the quadrature points with ${}^i w$ being the quadrature weights, but the notion of ${}^i w$ on $\phi^{(b)} \in H^{-1/2}(\Gamma_C^{(b)})$ may not be realistic when ${}^i w$ is defined for $\phi^{(a)} \in H^{-1/2}(\Gamma_C^{(a)})$.

The Eq. (4.10) and Eq. (4.11) defined through collocation can be expressed as

$$\begin{aligned} \langle {}_h\boldsymbol{\sigma}_n^{(a)}, \boldsymbol{\phi}_i^{(a)} \rangle_{\mathbf{I}^{(a)}} = & \\ \sum_{\forall i \in \mathbf{I}^{(a)}} p[(\boldsymbol{\phi}_i^{(a)} \cdot \hat{\mathbf{v}}_n)(\mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_n) - (\boldsymbol{\phi}_i^{(a)} \cdot \hat{\mathbf{v}}_n)(-\mathbf{N}^{(b)} \cdot \hat{\mathbf{v}}_n)] \mathbf{U}^{(a,b)} d\Gamma_C^{(a)} \\ \forall \boldsymbol{\phi}_{i \in \Gamma_C^{(a)}} \in \Phi^{(a)} \quad (4.16) \end{aligned}$$

$$\begin{aligned} \langle {}_h\boldsymbol{\sigma}_t^{(a)}, \boldsymbol{\phi}_i^{(a)} \rangle_{\mathbf{I}^{(a)}} = & \\ \sum_{\forall i \in \mathbf{I}^{(a)}} \mu p[(\boldsymbol{\phi}_i^{(a)} \cdot \hat{\mathbf{v}}_t)(\mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_n) - (\boldsymbol{\phi}_i^{(a)} \cdot \hat{\mathbf{v}}_t)(-\mathbf{N}^{(b)} \cdot \hat{\mathbf{v}}_n)] \mathbf{U}^{(a,b)} d\Gamma_C^{(a)} \\ \forall \boldsymbol{\phi}_{i \in \Gamma_C^{(a)}} \in \Phi^{(a)} \quad (4.17) \end{aligned}$$

Or alternatively can be expressed as

$$\mathbf{K}_C^{(a)} = \sum_{\forall i \in \mathbf{I}^{(a)}} p[({}^i \mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_n)^T ({}^i \mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_n) - ({}^i \mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_n)^T (-{}^i \mathbf{N}^{(b)} \cdot \hat{\mathbf{v}}_n)] \mathbf{U}^{(a,b)} \quad (4.18)$$

$$\mathbf{K}_F^{(a)} = \sum_{\forall i \in \mathbf{I}^{(a)}} \mu p[({}^i \mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_t)^T ({}^i \mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_n) - ({}^i \mathbf{N}^{(a)} \cdot \hat{\mathbf{v}}_t)^T (-{}^i \mathbf{N}^{(b)} \cdot \hat{\mathbf{v}}_n)] \mathbf{U}^{(a,b)} \quad (4.19)$$

$$\text{where } {}^i \mathbf{N} := [{}^i \boldsymbol{\phi}_1 \quad \dots \quad {}^i \boldsymbol{\phi}_{n \times m \times l}]$$

Similar to the Isoparametric approach in the classical FEM, the integral is defined over the parametric domain $\hat{\Omega}$, where the above expressions can be defined as

$$\mathbf{K}_C^{(a)} = \sum_{\forall i \in \mathbf{I}^{(a)}} [p[({}^i \mathbf{R}^{(a)} \cdot \hat{\mathbf{v}}_n)^T ({}^i \mathbf{R}^{(a)} \cdot \hat{\mathbf{v}}_n) - ({}^i \mathbf{R}^{(a)} \cdot \hat{\mathbf{v}}_n)^T (-{}^i \mathbf{R}^{(b)} \cdot \hat{\mathbf{v}}_n)] | {}^i \mathbf{J}^{(a)} |] \mathbf{U}^{(a,b)} \quad (4.20)$$

$$\mathbf{K}_F^{(a)} = \sum_{\forall i \in \mathbf{I}^{(a)}} [\mu p[({}^i \mathbf{R}^{(a)} \cdot \hat{\mathbf{v}}_t)^T ({}^i \mathbf{R}^{(a)} \cdot \hat{\mathbf{v}}_n) - ({}^i \mathbf{R}^{(a)} \cdot \hat{\mathbf{v}}_t)^T (-{}^i \mathbf{R}^{(b)} \cdot \hat{\mathbf{v}}_n)] | {}^i \mathbf{J}^{(a)} |] \mathbf{U}^{(a,b)} \quad (4.21)$$

where ${}^i\mathbf{R} := \mathbf{R}({}^i\boldsymbol{\Xi})$, with ${}^i\boldsymbol{\Xi}$ being the collocation point in the parametric space. While, ${}^i\boldsymbol{\Xi}^{(b)}$ is the corresponding map of $\overleftarrow{\mathbf{X}}^{(b)}$ in the parametric space, which can be determined through Newton-Raphson method in solving for $\mathbf{X}^{(b)}({}^i\boldsymbol{\Xi}^{(b)}) = \overleftarrow{\mathbf{X}}^{(b)}(\mathbf{X}^{(a)}({}^i\boldsymbol{\Xi}^{(a)}))$. Hence, there exists a mapping ${}^i\boldsymbol{\Xi}^{(a)} \rightarrow {}^i\boldsymbol{\Xi}^{(b)}$ for which $\mathbf{X}^{(a)}({}^i\boldsymbol{\Xi}^{(a)}) = \mathbf{X}^{(b)}({}^i\boldsymbol{\Xi}^{(b)})$.

From the conservation of momentum at the interface, the following relation holds $\boldsymbol{\sigma}_n^{(a)} = -\boldsymbol{\sigma}_n^{(b)}$ and $\boldsymbol{\sigma}_t^{(a)} = -\boldsymbol{\sigma}_t^{(b)}$. Hence, the traction stresses at the interface $\Gamma^{(b)}$ can be similarly defined as

$$\mathbf{K}_C^{(a)} = \sum_{\forall i \in \mathbf{I}^{(a)}} [p[({}^i\mathbf{R}^{(b)} \cdot \hat{\mathbf{v}}_n)^T (-{}^i\mathbf{R}^{(a)} \cdot \hat{\mathbf{v}}_n) \quad ({}^i\mathbf{R}^{(b)} \cdot \hat{\mathbf{v}}_n)^T ({}^i\mathbf{R}^{(b)} \cdot \hat{\mathbf{v}}_n)] | {}^i\mathbf{J}^{(a)} |] \mathbf{U}^{(a,b)} \quad (4.22)$$

$$\mathbf{K}_F^{(a)} = \sum_{\forall i \in \mathbf{I}^{(a)}} [p[({}^i\mathbf{R}^{(b)} \cdot \hat{\mathbf{v}}_t)^T (-{}^i\mathbf{R}^{(a)} \cdot \hat{\mathbf{v}}_n) \quad ({}^i\mathbf{R}^{(b)} \cdot \hat{\mathbf{v}}_t)^T ({}^i\mathbf{R}^{(b)} \cdot \hat{\mathbf{v}}_n)] | {}^i\mathbf{J}^{(a)} |] \mathbf{U}^{(a,b)} \quad (4.23)$$

Hence, the contact stiffness matrix \mathbf{K}_C and the friction matrix \mathbf{K}_F for the system can be defined as

$$\mathbf{K}_C = \begin{bmatrix} \mathbf{K}_C^{(a)} \\ \mathbf{K}_C^{(b)} \end{bmatrix} \text{ and } \mathbf{K}_F = \begin{bmatrix} \mathbf{K}_F^{(a)} \\ \mathbf{K}_F^{(b)} \end{bmatrix}$$

It should be noted that, for the Eqns. (4.22) and (4.23) even though the integral should be defined over $\Gamma^{(b)}$ as $\langle h\boldsymbol{\sigma}^{(b)}, \boldsymbol{\phi}_i^{(b)} \rangle_{\Gamma^{(b)}}$, the collocation points $\mathbf{I}^{(a)}$ are determined only on $\Gamma^{(a)}$, where its corresponding projection on $\Gamma^{(b)}$ is defined through the projection $\overleftarrow{\mathbf{X}}^{(b)}$. This is commonly also known as one-pass. The Eqns. (4.22) and (4.23) can be further simplified based on the relation (4.15), where the following could be stated

$${}^i\boldsymbol{\phi}_i^{(a)} \cdot {}^i\boldsymbol{\phi}_i^{(a)} = \sum_{\forall \boldsymbol{\phi}_{i \in \Gamma_C^{(a)}}} {}^i\boldsymbol{\phi}_i^{(a)} \cdot {}^i\boldsymbol{\phi}_j^{(a)} = {}^i\boldsymbol{\phi}_i^{(a)} \quad (4.24)$$

$${}^i\boldsymbol{\phi}_i^{(b)} \cdot {}^i\boldsymbol{\phi}_i^{(b)} = \sum_{\forall \boldsymbol{\phi}_{i \in \Gamma_C^{(b)}}} {}^i\boldsymbol{\phi}_i^{(b)} \cdot {}^i\boldsymbol{\phi}_j^{(b)} = {}^i\boldsymbol{\phi}_i^{(b)} \quad (4.25)$$

This is similar to the lumping approach where the off-diagonal terms of a row is summed to the diagonal. It should be noted that, also the following relation holds

$${}^i\boldsymbol{\phi}_i^{(a)} \cdot {}^i\boldsymbol{\phi}_i^{(a)} = \sum_{\forall \boldsymbol{\phi}_{i \in \Gamma_C^{(a)}}} {}^i\boldsymbol{\phi}_i^{(a)} \cdot {}^i\boldsymbol{\phi}_j^{(b)} = {}^i\boldsymbol{\phi}_i^{(a)} \quad (4.26)$$

$${}^i\phi_i^{(b)} \cdot {}^i\phi_i^{(b)} = \sum_{\forall \phi_{i \in \Gamma_C^{(b)}}} {}^i\phi_i^{(b)} \cdot {}^i\phi_j^{(a)} = {}^i\phi_i^{(b)} \quad (4.27)$$

but only the inner product of the bases from the same space can be lumped, i.e, even though the above relations hold, it can not be lumped since the inner product of the bases are from different space and hence not possible to collocate on to a diagonal term of any particular basis. Moreover, it will not satisfy conservation of momentum with the above case but the relations implicitly define the conservation of momentum at the interface and hence cannot be lumped. Further, the analyses show nearly no variation between the lumping approach and the default formulation, hence we use this property as an advantage where only off diagonal terms have to be stored in memory.

As a side note, It is well known that the choice of master and slave can lead to bias with on-pass. The bias can be eliminated by the so called two-pass formulation where after one-pass, the role of the master and the slave is switched, and the average of the projections is taken in to account, given for $\Gamma_C^{(a)}$ as

$$\langle {}_h\sigma_n^{(a)}, \phi_i^{(a)} \rangle_{I^{(a,b)}} = \frac{1}{2} \langle {}_h\sigma_n^{(a)}, \phi_i^{(a)} \rangle_{I^{(a)}} + \frac{1}{2} \langle {}_h\sigma_n^{(a)}, \phi_i^{(a)} \rangle_{I^{(b)}} \quad (4.28)$$

$$\langle {}_h\sigma_t^{(a)}, \phi_i^{(a)} \rangle_{I^{(a,b)}} = \frac{1}{2} \langle {}_h\sigma_t^{(a)}, \phi_i^{(a)} \rangle_{I^{(a)}} + \frac{1}{2} \langle {}_h\sigma_t^{(a)}, \phi_i^{(a)} \rangle_{I^{(b)}} \quad (4.29)$$

and could be defined otherwise for $\Gamma_C^{(a)}$. The difference with one-pass and two-pass was found to have no effect in our application and hence we stick with the one-pass formulation.

We explore the contact patch test for different numerical quadrature schemes on various discretisation settings between two blocks in contact with each other. The lower block is fixed on the bottom face, while a uniform load of 1KN

5 Classical optimisation scheme for FEM model

Sensitivity analysis

The global sensitivity analysis for the involved shape parameters was performed using the Variance-based method which comes from Hoeffding-Sobol decomposition. This method is based on decomposing the variance of a function to its variance associated with the parameters and the interaction between the parameters. Hence, higher the variance in output of a function induced by a parameter infers higher sensitivity. The method is applied through Monte-Carlo based estimation defined by latin hypercube sampling for efficiency. In effect, to evaluate the global behaviour and to increase the accuracy for the given Monte-Carlo based estimation on the presumed asymptotic case demands a large computation of design points, which is simply impossible to converge with a reasonable time given the computation cost to evaluate the stability criteria. Hence, a surrogate model based on Gaussian process regression was used, detailed in section . The stability criteria given by the surrogate model is defined as \hat{C}_s , where $C_s \approx \hat{C}_s$.

To understand the effect of the shape parameters on the stability criteria, the first-order and the total-order sensitivity indices are computed. The first-order indices define the contribution of a given parameter to the change in unconditional variance $V(\hat{C}_s)$, while the total-order indices add to it the contribution of all the higher-order interactions on the given parameter. The general expression for the first-order index S_i and the total-order index S_{Ti} can be given as

$$S_i = \frac{V_{X_i}(E_{X_{\sim i}}(\hat{C}_s|X_i))}{V(\hat{C}_s)} \quad (5.1)$$

$$S_{Ti} = 1 - \frac{V_{X_{\sim i}}(E_{X_i}(\hat{C}_s|X_{\sim i}))}{V(\hat{C}_s)} \quad (5.2)$$

where $V_{X_i}(E_{X_{\sim i}}(\hat{C}_s|X_i))$ is the variance of the conditional expectation on the function of the stability criteria \hat{C}_s evaluated by conditioning the parameter X_i for several values across the bounded design space and similarly, $V_{X_{\sim i}}(E_{X_i}(\hat{C}_s|X_{\sim i}))$ is the variance of the conditional expectation obtained by conditioning all parameters except for X_i .

The described probability measures are estimated based on the estimators proposed in . The Monte-Carlo based estimation for the given estimators require two matrices Y_A and Y_B of equal size with rows and columns representing the design points and the parameters respectively. To evaluate the first order index of the i th parameter, all the parameters of Y_B are unchanged except for the i th parameter (i th column of the matrix) which is replaced by the i th parameter of Y_A to obtain the matrix Y_{Bi} . Similarly, to evaluate the total-order index of the i th parameter, all the parameters of Y_B are changed with the parameters of Y_A except for the i th parameter to obtain the matrix Y_{Bti} . Hence, the matrices Y_{Bi} and Y_{Bti} represent the conditioning of the parameters with respect to the matrix Y_A , which in a sense is used to evaluate the conditional probability terms and also to describe the effective unconditional variance, as given by the estimators. For n parameters and p design points where the parameters are to be conditioned, it requires an estimation of $(n + 1)p$ design points to evaluate the first order index or the total order index of all the parameters.

The sensitivity analysis was performed on seven shape parameters as in table ?? describing the complete geometry of the considered model. The value of p as described in section ?? is chosen to be 1500 and hence evaluating a total of 24000 design points with meta-model to evaluate the first-order and the total-order indices. The evaluation was repeated for different sample sets to check for convergence, which is seen to be not difficult with the chosen p value and with an estimated standard error for the indices of no more than 0.02.

The description of the parameters are as in

As it can be seen, the first-order indices show relatively high values for the parameters X1 - thickness of the disc and X4 - thickness of the pad. The total-order indices also increase relatively for the two parameters. But the global variance of the stability criteria can be largely attributed to independent effects from the parameters rather than interaction between them.

Further, the results are also shown with closed second-order indices, combining the independent effects and the interaction between any two parameters.

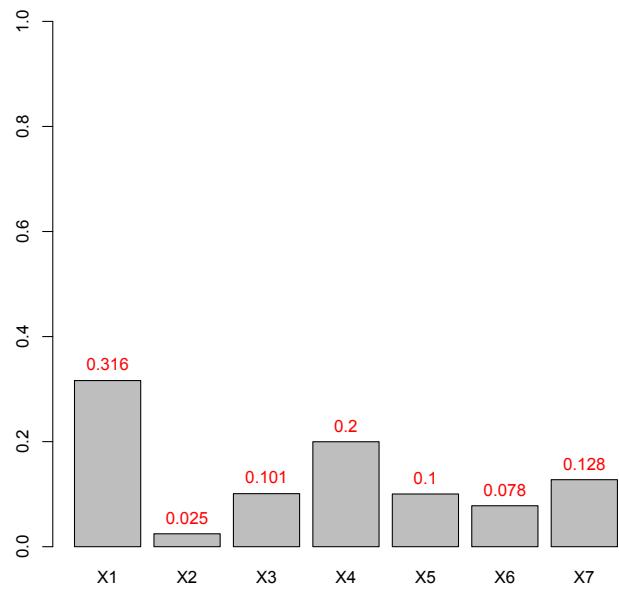


Figure 5.1: First-order Sobol indices

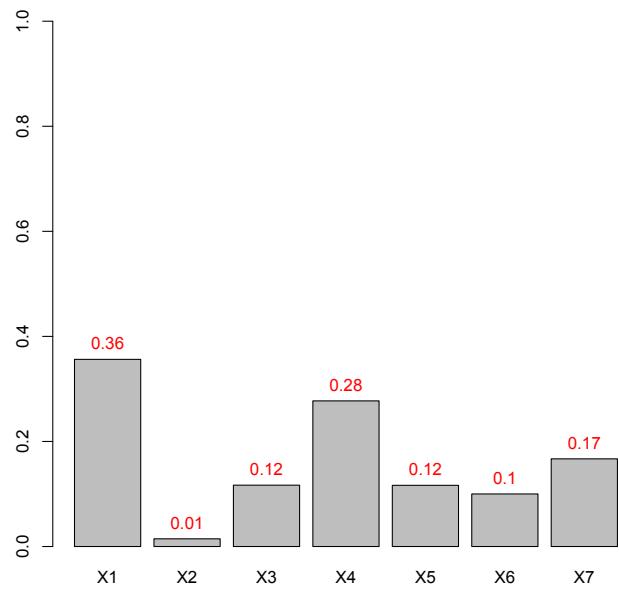


Figure 5.2: Total-order Sobol indices

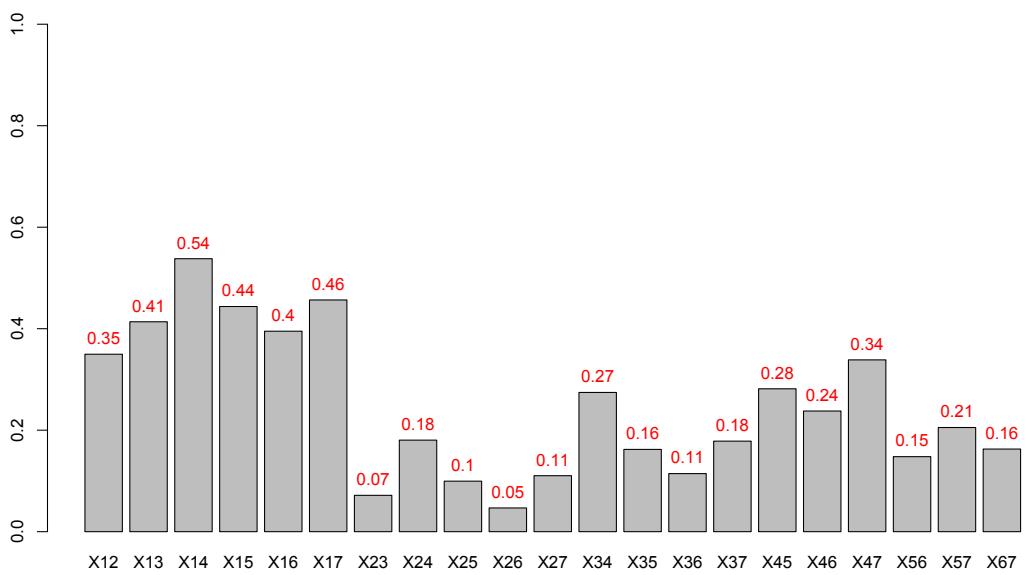


Figure 5.3: Closed second-order Sobol indices

6 IGA based optimisation scheme

6.0.1 IGA optimization setting

In this section, we detail the shape optimization defined through NURBS parameterization of shapes for the pad, with its associated constraints and objectives for optimization. We also provide a short description of parameterization and refinement strategy for the disc-pad system domain $\Omega^{(d-p)}$.

The optimization is defined for the boundary $\partial\Gamma_C^{(p)}$ of the planar surface $\Gamma_C^{(p)}$ of the pad which is in contact with the disc, where the thickness of the pad and the design parameters of the disc are set to be constant. The geometry of $\Gamma_C^{(p)}$ can be parameterized through NURBS as

$$\check{\mathbf{X}}_s^{(p)}(\xi, \eta) = \sum_{i=0}^n \sum_{j=0}^m R_{i,p}(\xi) R_{j,q}(\eta) \mathbf{P}_{i,j} \quad (6.1)$$

Hence, in this setting, the shape optimization is defined for the shape of the NURBS curves $\mathbf{X}_c^{(1)}(s), \check{\mathbf{X}}_c^{(2)}(t), \check{\mathbf{X}}_c^{(3)}(u)$ and $\mathbf{X}_c^{(4)}(v)$ which parameterizes $\partial\Gamma_C^{(p)}$ that encloses the surface $\check{\mathbf{X}}_s^{(p)}(\xi, \eta)$, as shown in Figure 6.1, where the curves can be expressed as

$$\begin{aligned} \check{\mathbf{X}}_s^{(p)}(\xi, \eta | \xi = 0) &= \mathbf{X}_c^{(1)}(s) \\ \check{\mathbf{X}}_s^{(p)}(\xi, \eta | \eta = 0) &= \mathbf{X}_c^{(2)}(t) \\ \check{\mathbf{X}}_s^{(p)}(\xi, \eta | \xi = 1) &= \mathbf{X}_c^{(3)}(u) \\ \check{\mathbf{X}}_s^{(p)}(\xi, \eta | \eta = 1) &= \mathbf{X}_c^{(4)}(v) \end{aligned} \quad (6.2)$$

This leads to the problem of defining the parameterisation $\check{\mathbf{X}}_s^{(p)}(\xi, \eta)$ given the four parametric curves $\mathbf{X}_c^{(1)}(s), \mathbf{X}_c^{(2)}(t), \mathbf{X}_c^{(3)}(u)$ and $\mathbf{X}_c^{(4)}(v)$. The parameterisation should be characterized by injective mapping which is ensured if the Jacobian does not vanish. For $\mathbf{X} : \hat{\Omega} \rightarrow \Omega$, verifying Jacobian on $\hat{\Omega}$ in transfinite sense would be impossible, for which it can be verified in a finite sense with the property of determinant-Jacobian function for a NURBS parameterisation. The definition of determinant of Jacobian for a NURBS parameterisation can be expressed as a function of higher-order NURBS to the NURBS parameterisation, given as

$$|\mathbf{J}(\check{\mathbf{X}}_s(\xi, \eta))| = \left| \begin{bmatrix} \frac{\partial \check{\mathbf{X}}_s}{\partial \xi} & \frac{\partial \check{\mathbf{X}}_s}{\partial \eta} \end{bmatrix} \right| = \sum_{i=1}^{2n-1} \sum_{j=1}^{2m-1} R_{i,2p-1}(\xi) R_{j,2q-1}(\eta) \mathbf{O}_{i,j} \quad (6.3)$$

The condition for injective mapping $|\mathbf{J}(\check{\mathbf{X}}_s(\xi, \eta))| > 0$ for $(\xi, \eta) \in [0, 1]^2$ can be said to be satisfied if $\mathbf{O}_{i,j} > 0$, which is a sufficient condition but not a necessary

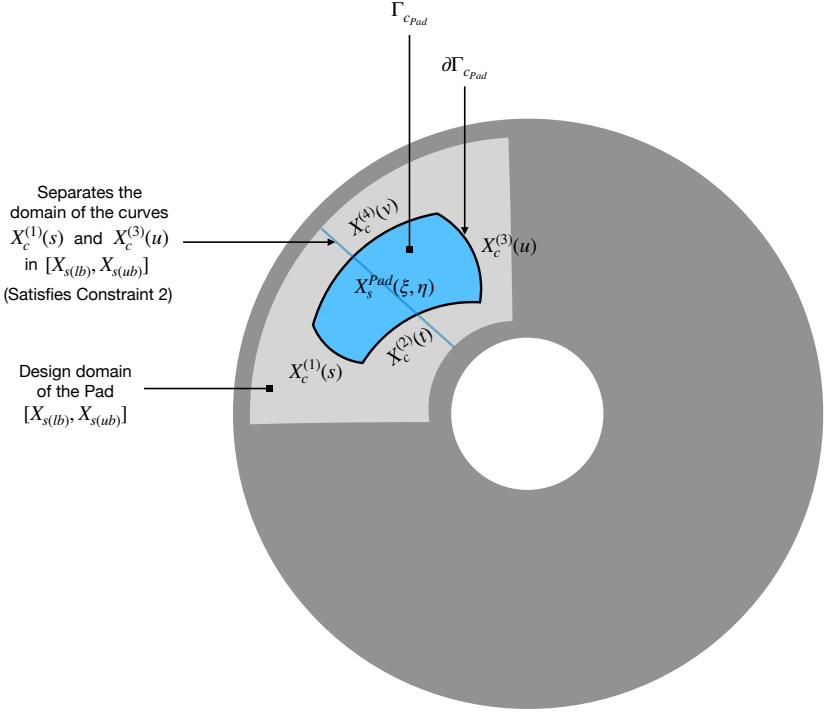


Figure 6.1: An illustration describing the parameterisation of Γ_{cPad} and $\partial\Gamma_{cPad}$.

one. This is because, if $|\mathbf{J}(\check{\mathbf{X}}_s(\xi, \eta))| = 0$ for any point on boundary, even though $|\mathbf{J}(\check{\mathbf{X}}_s(\xi, \eta))| > 0$ on $(0, 1)^2$, $\mathbf{O}_{i,j} < 0$. Nevertheless, $\mathbf{O}_{i,j}$ is often considered to check the validity of a parameterisation for injectivity, especially in the scope of defining optimisation to achieve an injective parameterisation.

The general idea behind Isogeometric approach is that given an initial parameterisation $\check{\mathbf{X}}$ of a domain Ω with NURBS, analysis-suitable parameterisation \mathbf{X} can be achieved with in the same parametric space, through addition or manipulation of knots and control points. But achieving initial parameterisation with injective mapping can be a difficult challenge especially for arbitrary definition of shapes in an optimisation, where to achieve a quality parameterisation with injective mapping can be even more challenging. The problem of defining $\check{\mathbf{X}}$ is more related to computer-aided design (CAD), where the role of CAD is typically not focused on defining \mathbf{X} . This is because, for the illustration of CAD, there can be multiple ways to parametrise a domain, that may not necessarily be suited for defining an approximation space ${}_h\mathbf{V}$ in the context of isogeometric analysis as Φ ¹. A typical approach in CAD is that a complicated domain being defined as a trimmed domain² (Fig. 6.2) or union of several trimmed domains.

¹To distinguish the space ${}_h\mathbf{V}$ in the context of isogeometric approach, we define Φ to be ${}_h\mathbf{V}$

²The definition of Ω as a trimmed domain can be defined as $\Omega \subset \hat{\Omega}_\oplus$, where $\hat{\Omega} \rightarrow \Omega_\oplus$

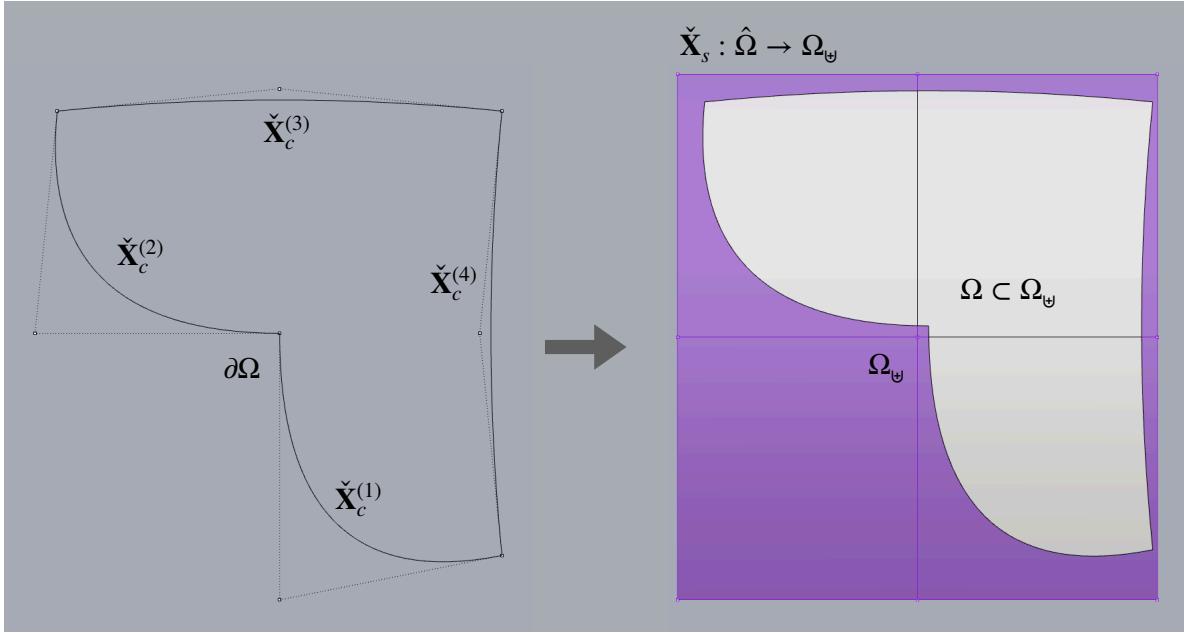


Figure 6.2: Parameterisation with trimmed domain

For the definition of a trimmed domain, since only essential part of the mapping that defines the domain from the parametric space to the physical space is considered, it does not place severe restriction over the complete parametric space to be mapped to the domain, illustrated in. This can be better in the context of designing where a surface can be loosely defined to contain a closed curve, but may not be suitable for defining Φ . One could say that for $\Omega \subset \Omega_\psi$, with Ω_ψ parameterised by \mathbf{X} , and given the bases $\phi_i := \mathbf{R}_{i,j,k}(\Xi) \circ \mathbf{X}^{-1}$, only the bases ϕ_i defining Ω can be considered for approximation. This is essentially the approach of the immersed methods, where typically the bases ϕ_i defining Ω is distinguished with material properties at the quadrature points, along with local refinement at the boundary of $\Omega \subset \Omega_\psi$ through hierarchical refinement. Though immersed methods can have more flexibility in defining ${}_h\mathbf{V}$, we do not focus on such approaches owing to its novelty which can require immense time to develop. We purely focus on defining $\mathbf{X} : \hat{\Omega} \rightarrow \Omega$ which can be called body-fitted parameterisation. The point is that to achieve body-fitted injective parameterisation for any shape as initial parameterisation can be too demanding from purely the perspective of CAD (Fig. (Fig. 6.3)).

For complex shapes, it is typically preferred to define analysis-suitable parameterisation directly, rather from a prior definition of initial parameterisation, where analysis-suitable parameterisation with sufficient refinement can be well suited to define injective parameterisation. From the perspective of CAD, this makes no difference as long as body-fitted injective parameterisation is achieved and hence, sometimes, no distinction can necessarily be made between initial and analysis-suitable parameterisations (Fig. 6.4).

Nevertheless, given the complexity of defining initial or analysis-suitable param-

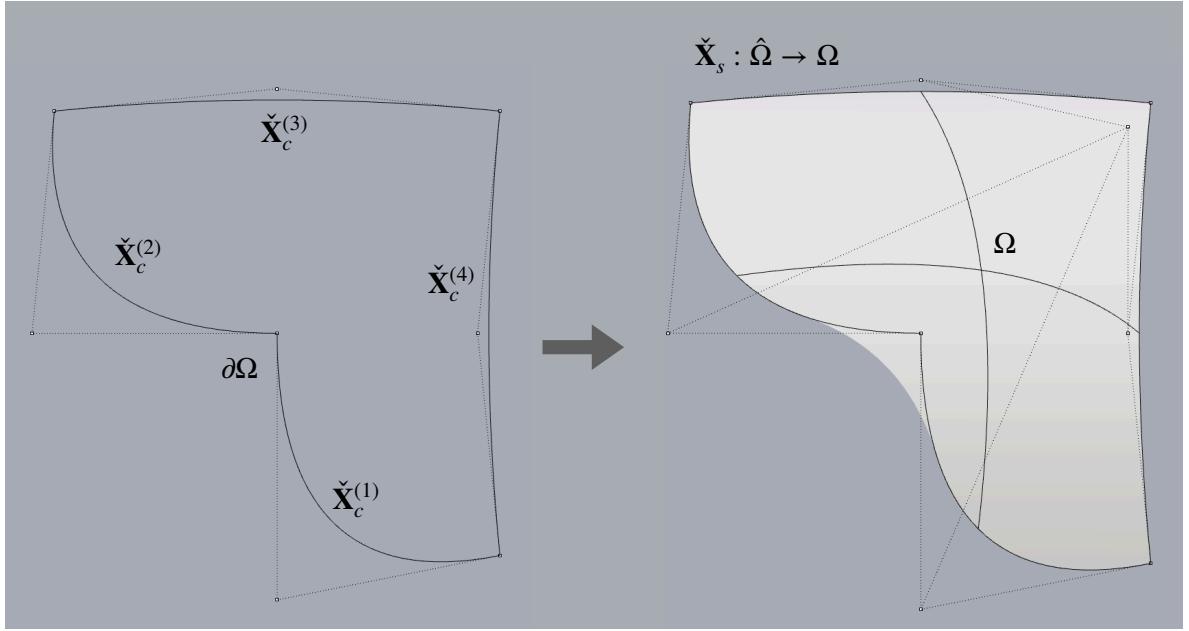


Figure 6.3: Body-fitted non-injective parameterisation

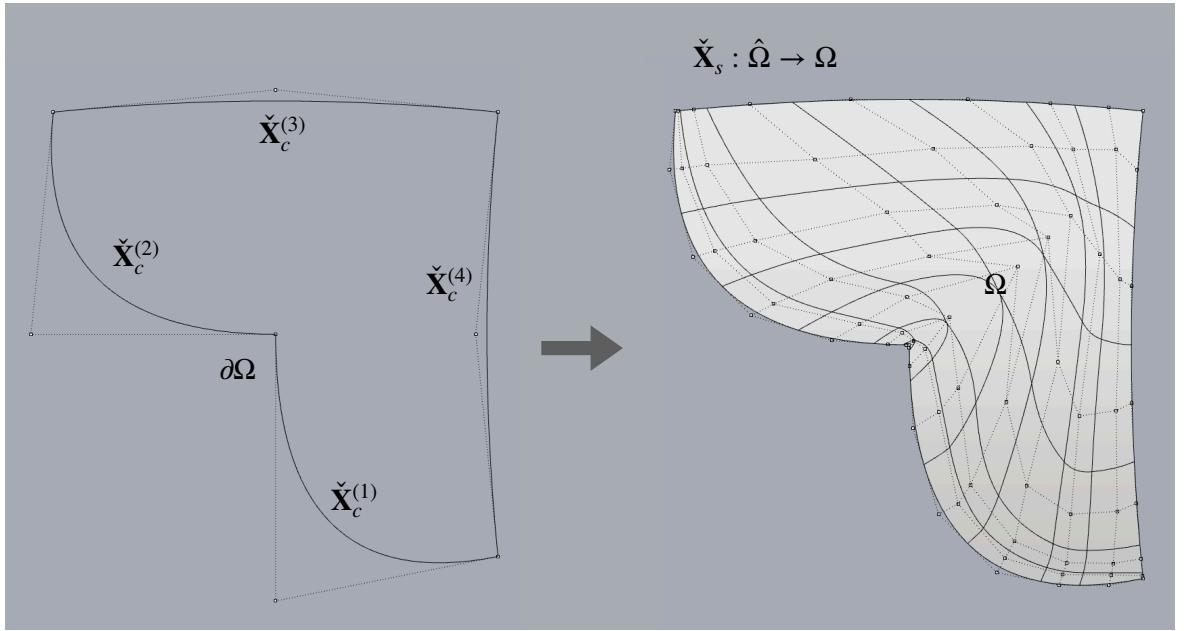


Figure 6.4: Body-fitted injective parameterisation

parameterisation, it is seen as a more robust strategy to define ${}_h\mathbf{V}$ compared to classical FEM. It should also be reminded that a complex domain can also be defined through multiple patches, where each patch corresponds to body-fitted parameterisation. But this also requires a more robust strategy to split a complex domain in to patches for any arbitrary definition of shape, at least for a fixed topology. Nevertheless, we adapt

multi-patch parameterisation as a strategy for local refinement and sub-structuring in optimisation.

For the scope of this thesis, we consider discrete Coon's patch method as a preliminary approach for the parameterisation part of the Bayesian shape optimisation framework. The idea is to adapt a more advanced parameterisation strategy for future evolution of the framework.

The parameterization of $\check{\mathbf{X}}_s^{(p)}(\xi, \eta)$ with the above four curves (6.2) by discrete Coon's patch method can be given as

$$\begin{aligned}\check{\mathbf{X}}_s^{(p)}(\xi, \eta) = & \check{\mathbf{X}}_c^{(1)}(s)(1 - \xi) + \mathbf{X}_c^{(3)}(u)(\xi) + \check{\mathbf{X}}_c^{(2)}(t)(1 - \eta) + \mathbf{X}_c^{(4)}(v)(\eta) \\ & - \check{\mathbf{X}}_c^{(1)}(0)(1 - \xi)(1 - \eta) - \mathbf{X}_c^{(1)}(1)(1 - \xi)(\eta) \\ & - \check{\mathbf{X}}_c^{(3)}(0)(\xi)(1 - \eta) - \check{\mathbf{X}}_c^{(3)}(1)(\xi)(\eta)\end{aligned}\quad (6.4)$$

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

In the scope of shape optimisation, $\check{\mathbf{X}}_s^{(p)}(\xi, \eta)$ can be defined as the function to be optimised, on which constraints can be imposed. We define Constraint set 1, which contains constraints intrinsic of the boundary curves (6.2). For simplicity owing to the preliminary definition of framework, in order to limit the parameters in optimization, we restricted the degree of each curve to 2 and hence, this leads to the surface $\check{\mathbf{X}}_s^{(p)}(\xi, \eta)$ 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. Furthermore, the optimisation is defined only for the position of the control points $\mathbf{P}_{i,j}$ for $w_{i,j} = 1$ (4.4), i.e, we considered the optimization of the NURBS geometry only through affine transformation without considering projective transformation.

While the end control points will be constrained relative to the disc domain, given in Constraint set 3, we impose constraint on the mid-control point of each curve segment. The control points for any curve segment can be expressed as \mathbf{P}_1 , \mathbf{P}_2 and \mathbf{P}_3 , where \mathbf{P}_1 and \mathbf{P}_3 define the end control points. If the initial configuration of the curve can be expressed as a line segment $\overline{\mathbf{P}_1 \mathbf{P}_3}$ with $\mathbf{P}_2 = \frac{\mathbf{P}_1 + \mathbf{P}_3}{2}$, then the constraint on \mathbf{P}_2 can be expressed relative to the initial configuration as $\mathbf{P}_2 \perp \overline{\mathbf{P}_1 \mathbf{P}_3}$

Constraints between curve segments are given as Constraint set 2 which contains constraints to confirm injective mapping, which also implicitly preserves topology.

Injective parameterisation can be said to be achieved if $|\mathbf{J}(\check{\mathbf{X}}_s(\xi, \eta))|$ does not vanish on all $\hat{\Omega}$. With the mid-control points constrained, the set of constraints for testing this condition was realised geometrically, given as

Constraint set 2:

$$\begin{aligned} & \{\mathbf{X}_c^{(1)}(s) \cap \check{\mathbf{X}}_c^{(2)}(t)\} \cup \{\mathbf{X}_c^{(3)}(u) \cap \check{\mathbf{X}}_c^{(4)}(v)\} \cup \{\mathbf{X}_c^{(1)}(s) \cap \check{\mathbf{X}}_c^{(3)}(u)\} \cup \\ & \{\mathbf{X}_c^{(1)}(s) \cap \check{\mathbf{X}}_c^{(4)}(v)\} \cup \{\mathbf{X}_c^{(2)}(t) \cap \check{\mathbf{X}}_c^{(3)}(u)\} \cup \{\mathbf{X}_c^{(2)}(t) \cap \check{\mathbf{X}}_c^{(4)}(v)\} = \emptyset \\ & \forall s, t, u, v \in (0, 1) \\ \\ & \{\mathbf{X}_c^{(1)}(0)(1 - \xi) + \mathbf{X}_c^{(3)}(0)(\xi)\} \cap \{\mathbf{X}_c^{(1)}(0 + \Delta s)(1 - \xi) + \mathbf{X}_c^{(3)}(0 + \Delta u)(\xi)\} = \emptyset \\ & \{\mathbf{X}_c^{(1)}(1)(1 - \xi) + \mathbf{X}_c^{(3)}(1)(\xi)\} \cap \{\mathbf{X}_c^{(1)}(1 - \Delta s)(1 - \xi) + \mathbf{X}_c^{(3)}(1 - \Delta u)(\xi)\} = \emptyset \\ & \{\mathbf{X}_c^{(2)}(0)(1 - \eta) + \mathbf{X}_c^{(4)}(0)(\eta)\} \cap \{\mathbf{X}_c^{(2)}(0 + \Delta t)(1 - \eta) + \mathbf{X}_c^{(4)}(0 + \Delta v)(\eta)\} = \emptyset \\ & \{\mathbf{X}_c^{(2)}(1)(1 - \eta) + \mathbf{X}_c^{(4)}(1)(\eta)\} \cap \{\mathbf{X}_c^{(2)}(1 - \Delta t)(1 - \eta) + \mathbf{X}_c^{(4)}(1 - \Delta v)(\eta)\} = \emptyset \\ & \forall \xi, \eta \in [0, 1] \quad (6.5) \end{aligned}$$

where Δ represents an arbitrary small variation. The first constraint avoids intersection between the curves except for the end points. Satisfying the first constraint which guarantees a fixed topology does not assure injective parameterisation through Coon's patch method, for which the last set of four constraints are necessary. The last set of four constraints implicitly avoid concave intersection between the curves. Given that the curves do not intersect except for convex intersection at the end points, and with constraints on the mid-control points, injective parameterisation can be achieved with Coon's patch method.

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:

$$(\mathbf{X}_{s(lb)} \leq \mathbf{X}_s^{(p)}(\xi, \eta) \leq \mathbf{X}_{s(ub)}) : \{[\mathbf{X}_{c(lb)}^{(i)}, \mathbf{X}_{c(ub)}^{(i)}]\} \cap \{[\mathbf{X}_{c(lb)}^{(j)}, \mathbf{X}_{c(ub)}^{(j)}]\} = \emptyset \quad (6.6)$$

where the choice of $\mathbf{X}_{s(lb)}$ and $\mathbf{X}_{s(ub)}$ 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 defined in 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 $\mathbf{X}_c^{(i)}(.)$ and $\mathbf{X}_c^{(j)}(.)$ 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 on reducing the redundancies. Hence, we avoided some of the redundancies on empirical notion, such that the restricted design space has lesser meaningful designs. This maybe an interesting anomaly to investigate, since the redundancies may lead

to larger design space with more severe multi-modality.

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

Constraint 4:

$$Area(\mathbf{X}_s^{(p)}(\xi, \eta)) \geq A_{min} \quad (6.7)$$

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

The definition of the shape of $\check{\mathbf{X}}_s^{(p)}(\xi, \eta)$ through this strategy means that there is no requisite for a reference configuration to define optimization, but instead the pad shapes are defined through random generation of curves with C^0 continuity between them. We assume that this restricts bias to any specific shape and hence encouraging more randomness in defining a meaningful geometry. This highly restricts the use of gradient-based approaches for optimization, since the constraints are also black-box and may have discontinuities. 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. The randomness in the definition of shapes can lead to higher probability of failure for the constraints, and hence more constraint evaluation in optimisation.

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

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

where the optimisation of the functionals are defined over the space of NURBS functions. Since we fixed the order and the number of control points of the NURBS surface $\check{\mathbf{X}}_s^{(p)}(\xi, \eta)$, the optimisation is restricted to a fixed number of control points.

6.0.2 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 C_s , 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 local refinement for NURBS parameterization, hence we expose here some strategies to achieve local refinement. Empirically, the refinement at 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, since it requires a detailed study of not only the refinement but also the contact formulation and the nature of modelling contact stiffness.

The planar parameterization $\check{\mathbf{X}}_s^{(p)}$ can be easily extended to define $\Omega^{(p)}$ as $\mathbf{X}_v^{(p)}$ considering the thickness of the pad through the tensor product definition (4.6), given a NURBS line along the thickness. The disc domain $\Omega^{(d)}$ was realised by multi-patch parameterization $\mathbf{X}_v^{(d)} := \mathbf{X}_v^{(d_1)} \cup \mathbf{X}_v^{(d_2)}$ to achieve local refinement on $\Gamma_C^{(d_2)}$. The surface parameterization for the disc patches $\check{\mathbf{X}}_s^{(d_1)}$ and $\check{\mathbf{X}}_s^{(d_2)}$ can be achieved through the concept of revolved surface, detailed in Pieg1995TheNB, which assures robust injective parameterisation since the curve to be revolved is a straight line perpendicular to the disc axis, given that the straight line does not pass through the axis. The planar parameterizations $\check{\mathbf{X}}_s^{(d_1)}$ and $\check{\mathbf{X}}_s^{(d_2)}$ can be extended to $\mathbf{X}_v^{(d_1)}$ and $\mathbf{X}_v^{(d_2)}$ respectively, similar to achieving $\mathbf{X}_v^{(p)}$.

For any refinement, the space for parameterisation remains the same i.e, $(\xi, \eta, \zeta) \in [0, 1]^3$ and the refinement is defined only through manipulation or addition of knots and control point to achieve an analysis-suitable parameterisation. After an analysis-suitable parameterisation, to take in to account of the additional control points and the manipulated knot vectors, $\mathbf{X}_v^{(p)}$ and $\mathbf{X}_v^{(d)}$ can be expressed as $\bar{\mathbf{X}}_v^{(p)}$ and $\bar{\mathbf{X}}_v^{(d)}$. Hence, the NURBS bases associated with $\bar{\mathbf{X}}_v^{(p)}$ and $\bar{\mathbf{X}}_v^{(d)}$ are used to define the space for approximation in isogeometric approach, detailed in Section ???. It should be noted that often analysis-suitable parameterisation is achieved directly in the scope of defining an injective parameterisation, since for complex shapes, analysis-suitable parameterisation with sufficient refinement can be well suited to define injective parameterisation. Hence, often no distinguish can be made between initial and analysis-suitable parameterisations, since the definition of body-fitted initial parameterisation in the context of CAD may already demand sufficient refinement to achieve even an elementary injective parameterisation. Since we only choose designs for which injective parameterisation exists with Coon's patch method, it is safe to say that for any analysis-suitable injective parameterisation achieved through Coon's patch method, initial injective parameterisation can also be achieved by Coon's patch method. Hence, we talk in the context that initial parameterisation of CAD does not have sufficient refinement to define the space ${}_h\mathbf{V}$ and hence we explicitly define analysis-suitable parameterisation through refinement.

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 ow-

ing 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 BAZILEVS2010229 or THB-splines GIANNELLI2012485, but requires extensive adaptation. Hence, we defined a multi-patch parameterization strategy through collocation and projection of 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, shown in Figure 6.5.

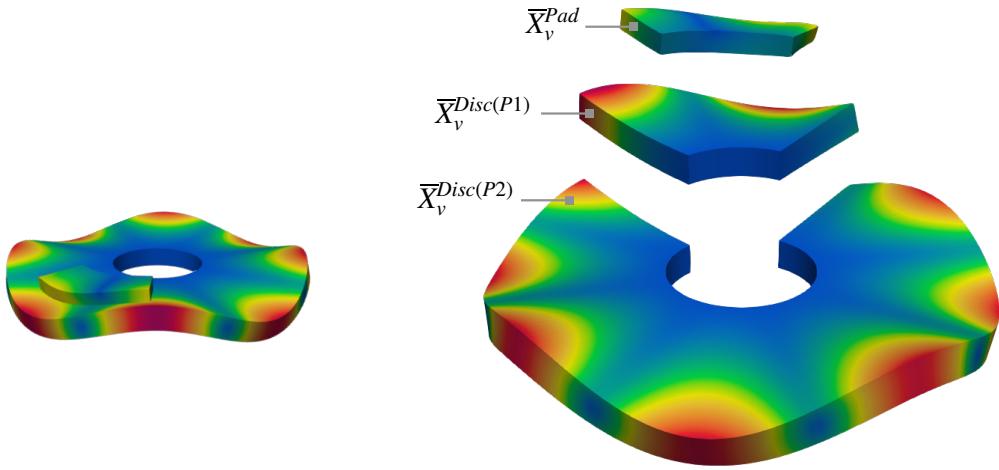


Figure 6.5: Anatomy of parameterization for the disc-pad system with arbitrary dimensions, shown here for Mode 9, Frequency: 3630 Hz

The multi-patch parameterization of $\Omega^{(d)}$ to break the NURBS tensor product definition is shown in Figure 6.6, where one patch $\overline{X}_v^{(d_1)}$ contains the contact domain $\Gamma_C^{(d_1)}$ defined through a fine mesh by h -refinement and the other patch $\overline{X}_v^{(d_2)}$ with a relatively coarse mesh sufficient to capture 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 $\partial\Gamma_C^{(d_2)}$ where typically strong solution gradient exists. For pad shapes where the knot lines on $\overline{X}_v^{(d_1)}$ can be aligned with $\partial\Gamma_C^{(d_1)}$, h -refinement can be used with finer refinement around $\partial\Gamma_C^{(d_1)}$, while the contact domain $\Gamma_C^{(d_1)}$ itself is discretized by h -refinement through a relatively coarse mesh compared to the refinement around $\partial\Gamma_C^{(d_1)}$, but finer than the rest of the domain. For pad shapes where the knot lines on $\overline{X}_v^{(d_1)}$ cannot be aligned with the boundary $\partial\Gamma_C^{(d_1)}$, 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.

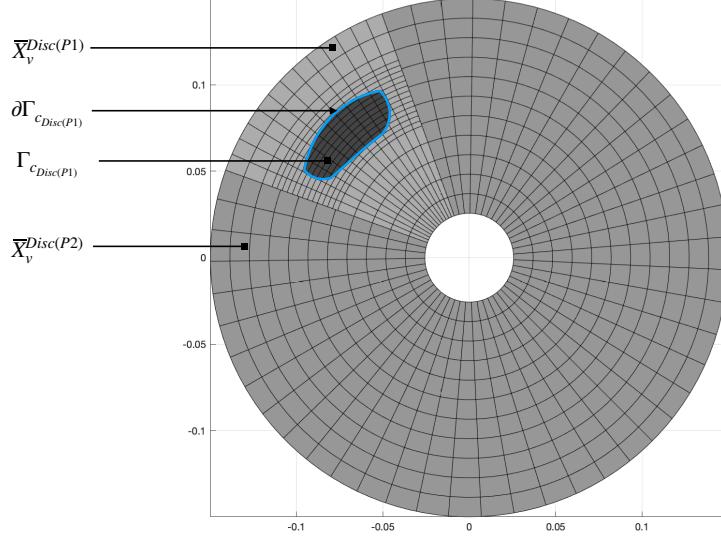


Figure 6.6: Multi-patch parameterization of Ω_{Disc} as $\overline{X}_v^{Disc} := X_v^{Disc(P1)}(\xi, \eta, \zeta) \cup X_v^{Disc(P2)}(\xi, \eta, \zeta)$, with h -refinement at the contact region $\Gamma_{c_{Disc}(P1)}$.

6.0.3 Bayesian optimization

Bayesian optimization is an effective strategy for optimising computationally expensive objective functions [2].

We begin the following explanations without defining the specifics of modelling the probability \mathcal{P} which is given as knowledge and considering the optimisation of a single function $f(\mathbf{x})$ for $\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$. 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

$$\mathcal{P}(\mathcal{H}|\mathcal{E}) \propto \mathcal{P}(\mathcal{E}|\mathcal{H})\mathcal{P}(\mathcal{H}) \quad (6.8)$$

where $\mathcal{P}(\mathcal{H}|\mathcal{E})$ defines Bayesian inference. In our setting, the hypothesis \mathcal{H} corresponds to the function $f(\mathbf{x})$ and the evidence \mathcal{E} to $\mathcal{F}_{1:n} : \{f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)\}$ where $f(\mathbf{x})$ is sampled on $\mathcal{X}_{1:n} : \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, with $\mathcal{D}_{1:n} : \{\mathcal{X}_{1:n}, \mathcal{F}_{1:n}\}$. This is typically known as function-space view, since the probability is defined on the space of functions. It can be hard to conceptualise such view with functions, but it is possible if one can imagine the existence of a function in a mere probabilistic sense such that

the random draw from a probability distribution is a function. The relation (6.9) can be expressed in this case as

$$\mathcal{P}(f(\mathbf{x})|\mathcal{D}_{1:n}) \propto \mathcal{P}(\mathcal{D}_{1:n}|f(\mathbf{x}))\mathcal{P}(f(\mathbf{x})) \quad (6.9)$$

The prior over a function, $\mathcal{P}(f(\mathbf{x}))$, is typically modelled through spatial correlation which is assumed to be known a priori, where the hypothesis is that a given function exhibits certain characteristics of spatial correlation which can be generalized globally. In other words, a prior belief is defined over the space of functions, such that the functions in the space largely exhibit certain characteristics of spatial correlation. With the prior $\mathcal{P}(f(\mathbf{x}))$ defined, and given the likelihood of the points sampled on the function, $\mathcal{P}(\mathcal{D}_{1:n}|f(\mathbf{x}))$, the posterior knowledge of the function, $\mathcal{P}(f(\mathbf{x})|\mathcal{D}_{1:n})$, can be inferred from the relation (6.9). The posterior knowledge $\mathcal{P}(f(\mathbf{x})|\mathcal{D}_{1:n})$ is then used to infer the next point \mathbf{x}_{n+1} to be sampled, depending on the strategy set for sampling in optimization. The sampled point \mathbf{x}_{n+1} is then used to update the belief of the prior $\mathcal{P}(f(\mathbf{x}))$ in the light of $\mathcal{D}_{1:n+1}$, and with the likelihood $\mathcal{P}(\mathcal{D}_{1:n+1}|f(\mathbf{x}))$ to infer a new posterior $\mathcal{P}(f(\mathbf{x})|\mathcal{D}_{1:n+1})$, which characterizes active learning. The process is run subsequently with the prospect of finding the global optimum for the function through active learning, defines Bayesian optimization.

With the general idea defined for Bayesian optimization, at least in the context of optimizing a single function, we can now define the notion of modelling \mathcal{P} which is typically defined through Gaussian process (\mathcal{GP}). While a Gaussian distribution defines distribution over a random variable or in the case of a multi-variate Gaussian distribution over random variables, a \mathcal{GP} defines distribution over a function, such that each draw from a \mathcal{GP} is a function. For some intuition of the following explanations, this can be thought in a discrete sense as all the points, of a function drawn from a \mathcal{GP} as being related through a dependent multi-variate Gaussian distribution such that each point is a univariate Gaussian distribution over a value of the function.

The prior over a function can hence be defined as \mathcal{GP} prior, where the advantage of modelling the prior as Gaussian means that it preserves the conditioning of the Gaussian prior given the likelihood to infer the posterior as Gaussian as well. This is advantageous for Bayesian optimisation, since inferring the posterior as Gaussian presents the prediction as mean and the uncertainty of the prediction as variance, which provides a decisive knowledge to construct an acquisition function to sample more efficiently. The \mathcal{GP} posterior defined through Bayesian inference from conditioning a \mathcal{GP} prior given the likelihood of the sampled points over the function, characterizes a regression model, known as \mathcal{GP} regression. Hence, the meta-modelisation of the function $f(\mathbf{x})$ can be defined through \mathcal{GP} regression.

The \mathcal{GP} prior $\mathcal{P}(f(\mathbf{x}))$ over $f(\mathbf{x})$ can be expressed as

$$f(\mathbf{x}) \approx \mathcal{P}(f(\mathbf{x})) = \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')),$$

$$\mu(\mathbf{x}) = \text{E}[f(\mathbf{x})], \quad k(\mathbf{x}, \mathbf{x}') = \text{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))] \quad (6.10)$$

where the distribution constitutes a mean function $\mu(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$. The function $\mu(\mathbf{x})$ can be seen as the deterministic part which captures the general trend of $f(\mathbf{x})$, while the covariance function $k(\mathbf{x}, \mathbf{x}')$ models the stochastic trend which is the spatial correlation between any $f(\mathbf{x}) - \mu(\mathbf{x})$ and $f(\mathbf{x}') - \mu(\mathbf{x}')$. The deterministic part $\mu(\mathbf{x})$ is largely modelled as a constant or through polynomials, which is demanding to estimate a priori and also higher degree polynomial trend functions can lead to overfitting over the sampled points. Hence, care should be taken in defining the general trend such that some spatial correlation exists with respect to the trend. Recently, focus has also been in defining $\mu(\mathbf{x})$ with Polynomial chaos expansion approach.

The spatial correlation is modelled by hyperparameters $\boldsymbol{\theta}$ which are the constants known a priori in a covariance function $\text{cov}(f(\mathbf{x}) - \mu(\mathbf{x}), f(\mathbf{x}') - \mu(\mathbf{x}'))$, where the choice of the covariance function depends on the application. Even though the prior knowledge is defined to be known, it is often determined from the light of the sampled points. Hence, to define the prior over $k(\mathbf{x}, \mathbf{x}')$, the hyperparameters are estimated a priori from the sampled points, which is usually achieved by optimising the likelihood function for $\arg \max_{\boldsymbol{\theta}} L(\mathcal{F}|\boldsymbol{\theta})$. More on optimising for hyperparameters will be detailed in the upcoming explanations, where $\boldsymbol{\theta}$ often contains parameters to model $\mu(\mathbf{x})$ in addition to the hyperparameters.

With $\boldsymbol{\theta}$ determined, the \mathcal{GP} prior $\mathcal{P}(f(\mathbf{x}))$ can now be defined. The conditioning of $\mathcal{P}(f(\mathbf{x}))$ with the likelihood of the sampled points $\mathcal{D}_{1:n}$ results in a \mathcal{GP} posterior $\mathcal{P}(f(\mathbf{x})|\mathcal{D}_{1:n}, \boldsymbol{\theta})$ which can be viewed in a finite-dimensional sense as the posterior joint Gaussian distribution of $\mathcal{P}(f(\mathbf{x}_1^*)), \mathcal{P}(f(\mathbf{x}_2^*)), \dots, \mathcal{P}(f(\mathbf{x}_n^*))$ across rest of the function where its arguments \mathbf{x}_i^* has not been sampled, i.e. $\mathbf{x}_i^* \notin \mathcal{X}_{1:n}$.

To move on from the abstractness of \mathcal{GP} to a practical finite-dimensional Gaussian distribution useful for making inference at an arbitrary point $\mathbf{x}^* \in \mathcal{X}$, given the sampled points $\mathcal{X}_{1:n}$, the properties of multi-variate Gaussian distribution allow to isolate a part of the \mathcal{GP} prior $\mathcal{P}(f(\mathbf{x}))$ to define a joint Gaussian distribution of only the sampled points $\mathcal{X}_{1:n}$ and an argument \mathbf{x}^* where the inference is to be made, where the joint distribution can be expressed as

$$\begin{bmatrix} \mathcal{P}(f(\mathbf{x}_1)) \\ \vdots \\ \mathcal{P}(f(\mathbf{x}_n)) \\ \mathcal{P}(f(\mathbf{x}^*)) \end{bmatrix} = \mathcal{N} \left(\begin{bmatrix} \mu(\mathbf{x}_1) \\ \vdots \\ \mu(\mathbf{x}_n) \\ \mu(\mathbf{x}^*) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_n) & k(\mathbf{x}_1, \mathbf{x}^*) \\ \vdots & \ddots & \vdots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & \dots & k(\mathbf{x}_n, \mathbf{x}_n) & k(\mathbf{x}_n, \mathbf{x}^*) \\ k(\mathbf{x}^*, \mathbf{x}_1) & \dots & k(\mathbf{x}^*, \mathbf{x}_n) & k(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix} \right) \quad (6.11)$$

The above joint distribution can be partitioned to define the mean and the covariance for the sampled points and the point to be inferred as

$$\boldsymbol{\Sigma} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & \dots & k(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix}, \boldsymbol{\Sigma}^* = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}^*) \\ \vdots \\ k(\mathbf{x}_n, \mathbf{x}^*) \end{bmatrix}, \boldsymbol{\mu}(\mathcal{X}) = \begin{bmatrix} \mu(\mathbf{x}_1) \\ \vdots \\ \mu(\mathbf{x}_n) \end{bmatrix} \quad (6.12)$$

The conditioning of the joint distribution Eq.(6.11) defined by the prior knowledge of $\boldsymbol{\theta}$ with the sampled data \mathcal{D} gives the prediction for \mathbf{x}^* as follows

$$\mathcal{P}(f(\mathbf{x}_i^*)|\mathcal{D}, \boldsymbol{\theta}) = \mathcal{N}(\underbrace{\mu(\mathbf{x}_i^*) + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^*(\mathcal{F} - \boldsymbol{\mu}(\mathcal{X}))}_{\hat{\mu}(\mathbf{x}^*)}, \underbrace{k(\mathbf{x}^*, \mathbf{x}^*) - \boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{*T}}_{\hat{\sigma}^2(\mathbf{x}^*)}) \quad (6.13)$$

where the function $f(\mathbf{x})$ approximated by the Gaussian process regression model can be defined as $\hat{f}(\mathbf{x}) := \mathcal{N}(\hat{\mu}(\mathbf{x}), \hat{\sigma}(\mathbf{x}))$.

The choice of the covariance function and the estimation of the hyperparameters in defining spatial correlation are important since they are the determining factors that distinguish the above distribution for a given observation. The hyperparameters parameterizes spatial correlation through smoothness or correlation length or sometimes both ³, for which a large class of covariance functions exist to choose from depending on the application. The most commonly used in engineering optimisation are the Gaussian and the Matérn class of covariance functions, where for isotropic correlation, the Gaussian covariance function can be defined as

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\theta^2}||\mathbf{h}||^2\right) \quad (6.14)$$

where $\mathbf{h} := [h_1, h_2, \dots, h_l]$, $h = (f(x) - \mu(x)) - (f(x') - \mu(x'))$ and \mathbf{x} is considered to be in \mathbb{R}^l . This is defined with only a single hyperparameter θ since it assumes the spatial correlation to be isotropic. The anisotropic consideration of spatial correlation can be defined as

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{k=1}^l \frac{1}{2\theta_k^2} |h_k|^2\right) \quad (6.15)$$

where it leads to determining l no. of θ . The Gaussian covariance function models spatial correlation only with correlation length defined through the factor $\frac{1}{2\theta^2}$, while

³It should be noted that the parameters modelling smoothness and correlation length are not independent, but rather interdependent such that parameters modelling smoothness has influence on correlation length and vice-versa. But largely, smoothness parameters can be said to quantify the gradient factor for the variation of h , while correlation length parameters can be said to quantify the influence of the points on each other for the variation of h

the smoothness for the variation of h is defined for a fixed power 2. The Matérn class of covariance functions provide flexibility in modelling smoothness through a predefined parameter v , where the function can be expressed for anisotropic variation as

$$k(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^l \sigma^2 \frac{2^{(1-v)}}{\text{G}(v)} \left(\frac{\sqrt{2v}|h_k|}{\theta_k} \right)^v \text{B} \left(\frac{\sqrt{2v}|h_k|}{\theta_k} \right) \quad (6.16)$$

where G and B are the Gamma function and the Bessel function of order v respectively. The value of v is typically defined to be $5/2$ or $3/2$, where as $v \rightarrow \infty$, it converges to squared exponential function and for $v = 1/2$, it simply characterizes an exponential function. The Gaussian covariance makes strong smoothness assumption with the infinite differentiability of the function which can be unreal and hence, Matérn class of functions are typically preferred which are $v - 1$ times differentiable. We used Matérn with $v = 5/2$, expressed as

$$k(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^l \sigma^2 \left(1 + \frac{\sqrt{5}|h_k|}{\theta_k} + \frac{5h_k^2}{3\theta_k^2} \right) \exp \left(-\frac{\sqrt{5}|h_k|}{\theta_k} \right) \quad (6.17)$$

With the definition of a covariance function, the optimisation of the hyperparameters to model the prior (6.11) is given by $\arg \max_{\boldsymbol{\theta}} L(\mathcal{F}|\boldsymbol{\theta})$, where $L(\mathcal{F}|\boldsymbol{\theta})$ defines the likelihood of the observed data given the hyperparameters, defined by the joint probability as

$$L(\mathcal{F}|\mu(\boldsymbol{\theta}), \sigma(\boldsymbol{\theta})) = \frac{1}{\sqrt{(2\pi\sigma^2)^n |\boldsymbol{\Sigma}|}} \exp \left[-\frac{(\mathcal{F} - \boldsymbol{\mu}(\mathcal{X}))^T \boldsymbol{\Sigma}^{-1} (\mathcal{F} - \boldsymbol{\mu}(\mathcal{X}))}{2\sigma^2} \right] \quad (6.18)$$

In optimising the above function for maximum likelihood, the function can be simplified by taking the natural logarithm while preserving the monotonicity of the function as

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \ln|\boldsymbol{\Sigma}| - \frac{(\mathcal{F} - \boldsymbol{\mu}(\mathcal{X}))^T \boldsymbol{\Sigma}^{-1} (\mathcal{F} - \boldsymbol{\mu}(\mathcal{X}))}{2\sigma^2} \quad (6.19)$$

The definition of the logarithm preserves the monotonicity of the function and hence also the optimum point of the function. Mean can be defined as $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\theta}_{\boldsymbol{\mu}}$ when modelled as regression with hyperparameters $\boldsymbol{\theta}_{\boldsymbol{\mu}}$ ⁴, where \mathbf{X} defines a matrix of size $n \times p$, with p being the number of linear combination of functions defined for regression. The maximum likelihood estimate of $\boldsymbol{\mu}$ in this case is simply the maximum likelihood estimate of $\boldsymbol{\theta}_{\boldsymbol{\mu}}$ which can be deduced from $\frac{\partial \ln(L)}{\partial \boldsymbol{\theta}_{\boldsymbol{\mu}}} = 0$ as

$$\check{\boldsymbol{\theta}}_{\boldsymbol{\mu}} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathcal{F} \quad (6.20)$$

⁴ $\boldsymbol{\theta} := \{\boldsymbol{\theta}_{\boldsymbol{\mu}}\} \cup \{\boldsymbol{\theta}_{\boldsymbol{\Sigma}}\}$, where $\boldsymbol{\theta}_{\boldsymbol{\mu}}$ and $\boldsymbol{\theta}_{\boldsymbol{\Sigma}}$ correspond to the hyperparameters of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

which is simply the minimiser for generalized least-squares. This is apparent, since the minimiser of $\boldsymbol{\theta}_\mu$ can be viewed as the minimisation of the generalised least-squares problem given as

$$\check{\boldsymbol{\theta}}_\mu = \arg \min_{\boldsymbol{\theta}_\mu} \frac{(\mathcal{F} - \boldsymbol{\mu}(\mathcal{X}, \boldsymbol{\theta}_\mu))^T \boldsymbol{\Sigma}^{-1} (\mathcal{F} - \boldsymbol{\mu}(\mathcal{X}, \boldsymbol{\theta}_\mu))}{2\sigma^2} \quad (6.21)$$

Similarly, the maximum likelihood estimate for σ can be defined from $\frac{\partial \ln(L)}{\partial \sigma} = 0$ as

$$\check{\sigma} = (\mathcal{F} - \mathbf{X}\boldsymbol{\theta}_\mu)^T \boldsymbol{\Sigma}^{-1} (\mathcal{F} - \mathbf{X}\boldsymbol{\theta}_\mu) \quad (6.22)$$

Substituting the maximum likelihood estimates of $\check{\boldsymbol{\theta}}_\mu$ and $\check{\sigma}$ in to Eq.(6.19), and with the constants removed as affine terms, one obtains

$$\ln(L(\boldsymbol{\theta})) \approx -\frac{n}{2} \ln(\check{\sigma}^2(\boldsymbol{\theta})) - \frac{1}{2} \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta}_\Sigma)| \quad (6.23)$$

The above function can be typically expected to be multimodal and hence the optimisation is typically achieved with Genetic algorithm (GA) for global convergence, followed by the best individuals from the GA as seeds for the quasi-newton algorithms like BFGS for local convergence. As a variation, quasi-newton schemes are also applied with in GA for the best individuals in each generation to define the parent population of the next generation.

The Bayesian inference $\mathcal{P}(f(\mathbf{x}_i^*) | \mathcal{D}_{1:n}, \boldsymbol{\theta})$ can be used in sampling for optimisation from the inference of the prediction $\hat{\mu}(\mathbf{x}^*)$ and the uncertainty $\hat{\sigma}^2(\mathbf{x}^*)$ of the prediction. Naturally, question arises for the goal of sampling in balance between exploration and exploitation. Exploration can viewed as the means to gain more knowledge about the function especially where high uncertainty is reflected by the \mathcal{GP} posterior. But with pure exploration, it diverts the goal in search for a global optimum with the consequence of reducing the uncertainty over the knowledge of the function, unless reducing the uncertainty also exposes the global optimum as a repercussion which happens rarely. In contrast, exploitation focuses on parts of the function which is inferred to define optimum through the prediction from the \mathcal{GP} posterior. Pure exploitation in optimisation can be viewed as more optimistic with the predictions and hence can underestimate the uncertain parts reflected by the \mathcal{GP} posterior.

This is where sampling through an acquisition function plays an important role in guiding the search for optimisation where the construction of the acquisition function can be adapted to set the balance between exploration and exploitation depending on the objective, for which a wide range of acquisition functions exist. In general, acquisition functions define improvement with respect to a reference value $f(\mathbf{x}^+)$ through a probabilistic metric, where $f(\mathbf{x}^+)$ typically corresponds to the utopian

value⁵ of the function, at least in the context of single objective optimisation. While in multi-objective optimisation the definition of utopian value corresponds to empirical Pareto front which will be discussed later. We will introduce some of the acquisition functions for single-objective optimisation which will be referred for the upcoming explanations. As defined before, we consider the case of optimising $f(\mathbf{x})$ for $\min_{\mathcal{X}} f(\mathbf{x})$ for the following explanations and hence the utopian value $f(\mathbf{x}^{++})$ ⁶ is defined by $\mathbf{x}^{++} = \arg \min_{\mathbf{x}_i \in \mathcal{X}_{1:n}} f(\mathbf{x}_i)$.

Probability of Improvement (PI) for any \mathcal{GP} outcome $\hat{f}(\mathbf{x})$ is given as

$$PI(\mathbf{x}) = \mathcal{P}(f(\mathbf{x}) \leq f(\mathbf{x}^{++})) \quad (6.24)$$

where,

$$\mathcal{P}(f(\mathbf{x}) \leq f(\mathbf{x}^+)) = CDF\left(\frac{\hat{\mu}(\mathbf{x}) - f(\mathbf{x}^{++})}{\hat{\sigma}(\mathbf{x})}\right) \quad (6.25)$$

PI gives more weight on exploitation than exploration in optimisation. This can be seen with the following case, where for a point with low variance for $\hat{\mu} < f(\mathbf{x}^{++})$ reflects more scope for improvement than for a point with the same $\hat{\mu}$ but larger variance, where PI reflects more focus on exploitation which leads to highly exhaustive search locally. PI is the same when both points have $\hat{\mu} = f(\mathbf{x}^{++})$ and when $\hat{\mu} > f(\mathbf{x}^{++})$, the point with larger variance has larger PI , where PI reflects focus on exploration. This means that PI focuses on exploration unless there is no possibility for exploitation, where in real world, this leads to exhaustive search locally around the best points before moving on to the next exploration search. To overcome this effect, a trade-off parameter $\mathcal{E} \geq 0$ is introduced, given as

$$PI(\mathbf{x}) = \mathcal{P}(f(\mathbf{x}) \leq f(\mathbf{x}^{++}) + \mathcal{E}) \quad (6.26)$$

where typically \mathcal{E} is set to be higher initially in an optimisation to drive exploration, and decreases to zero by the end of the optimisation to drive exploitation. PI clearly lacks a good balance between exploration and exploitation for which Expected Improvement EI is typically deemed to be effective.

EI is the expectation of the improvement, $E(I(\mathbf{x}))$, where the improvement is typically defined with respect to the utopian value $f(\mathbf{x}^{++})$ as $I(\mathbf{x}) = f(\mathbf{x}^{++}) - \hat{f}(\mathbf{x})$. The expectation of the improvement, $EI(\mathbf{x}|f(\mathbf{x}^{++}))$ ⁷ can hence be expressed as

⁵Utopian value is the observed optimum value of the function

⁶We use the notation $f(\mathbf{x}^{++})$ to define the Utopian value as reference value and $f(\mathbf{x}^+)$ to define an arbitrary reference value

⁷For simplicity, we define $EI(\mathbf{x}|f(\mathbf{x}^{++}))$ as $EI(\mathbf{x})$ unless we want to emphasize the use of $f(\mathbf{x}^{++})$

$$EI(\mathbf{x}) = E(I(\mathbf{x})) = \int_{-\infty}^{f(\mathbf{x}^{++})} I(\mathbf{x}) PDF\left(\frac{f(\mathbf{x}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) df(\mathbf{x}) \quad (6.27)$$

For intuition, the above expression can be given as

$$EI(\mathbf{x}) = \underbrace{\left(f(\mathbf{x}^{++}) - \frac{\int_{-\infty}^{f(\mathbf{x}^{++})} f(\mathbf{x}) PDF\left(\frac{f(\mathbf{x}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) df(\mathbf{x})}{CDF\left(\frac{f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right)} \right)}_{f_{cen}} CDF\left(\frac{f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) \quad (6.28)$$

where f_{cen} is the first moment of area/centroid of the $PDF\left(\frac{f(\mathbf{x}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) \in (-\infty, f(\mathbf{x}^{++})]$ on the axis of $f(\mathbf{x})$. Hence, EI can be understood as the measure of f_{cen} with respect to the reference value $f(\mathbf{x}^{++})$, given as $f(\mathbf{x}^{++}) - f_{cen}$, weighted by the $CDF\left(\frac{f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right)$ which is simply $PI(\mathbf{x})$. The term $\int_{-\infty}^{f(\mathbf{x}^{++})} f(\mathbf{x}) PDF\left(\frac{f(\mathbf{x}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) df(\mathbf{x})$ can be seen as the measure of $E(\hat{f}(\mathbf{x}))$ in the interval $(-\infty, f(\mathbf{x}^{++})]$, where it defines the expected value rather than the expected improvement defined by EI with respect to a reference value. We elaborate these definitions since it will be useful for definitions to extend EI to MOO.

Overall, EI provides better trade-off between exploration and exploitation, unlike the greedy nature of PI which primarily focuses on exploitation. This is because the term $f(\mathbf{x}^{++}) - f_{cen}$ weighted by PI provides the additional factor in EI to balance the search for exploration. The above expression of EI can be simplified as

$$EI(\mathbf{x}) = (f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x})) CDF\left(\frac{f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) + \hat{\sigma}(\mathbf{x}) PDF\left(\frac{f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) \quad (6.29)$$

Similar to Eq. (6.26), it is possible to control the trade-off between exploration and exploitation by introducing $\mathcal{E} \geq 0$ to the above expression as follows

$$EI(\mathbf{x}) = (f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x}) - \mathcal{E}) CDF\left(\frac{f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) + \hat{\sigma}(\mathbf{x}) PDF\left(\frac{f(\mathbf{x}^{++}) - \hat{\mu}(\mathbf{x}))}{\hat{\sigma}(\mathbf{x})}\right) \quad (6.30)$$

The other common acquisition function is defined with the bound of $\hat{f}(\mathbf{x})$, where for a function to be minimised, the lower confidence bound can be defined as

$$LCB(\mathbf{x}) = \hat{\mu}(\mathbf{x}) - \epsilon\hat{\sigma}(\mathbf{x}) \quad (6.31)$$

where $\epsilon \geq 0$. It is quite intuitive to think that LCB favours exploration over exploitation. It is apparent that each acquisition function will give rise to distinct sampling behaviour and hence, the choice of the acquisition function depends on the goal for sampling. Similar to the approach of maximising Eq. (??), the optimisation of the acquisition functions can be achieved with the combination of GA and quasi-newton algorithms to determine the infill point \mathbf{x}^t .

Multi-objective optimisation

For MOO, the problem can be formulated as

$$\min_{\mathbf{x} \in \mathcal{X}} \mathbf{f}(\mathbf{x}) \quad (6.32)$$

where $\mathbf{f} = [f_1(\mathbf{x}), \dots, f_m(\mathbf{x})]$, $\mathbf{f} : \mathcal{X} \subset \mathbb{R}^l \rightarrow \mathcal{S} \subset \mathbb{R}^m$, with \mathcal{S} being the objective space. The definition of optimality in single objective optimisation is defined with minimum or maximum of a function. While in MOO for a set of functions to be minimised, where typically the functions can not be minimised simultaneously without conflict between the objectives, where minimising one function can implicitly maximise the other. Hence, the definition of optimality in MOO is given by Pareto optimality which defines optimality considering the best compromise between the objectives.

Hence, any two vectors $\mathbf{x}_a, \mathbf{x}_b \in \mathcal{X}$, and $\mathbf{x}_a \neq \mathbf{x}_b$, the following conditions can be stated for Pareto-dominance to define Pareto-optimality on $\min_{\mathbf{x} \in \mathcal{X}} \mathbf{f}(\mathbf{x})$, given as

- $\mathbf{x}_a \preceq \mathbf{x}_b$ (\mathbf{x}_a weakly dominates \mathbf{x}_b) i.f.f. $\forall i, f_i(\mathbf{x}_a) \leq f_i(\mathbf{x}_b)$
- $\mathbf{x}_a \prec \mathbf{x}_b$ (\mathbf{x}_a dominates \mathbf{x}_b) i.f.f. $\mathbf{x}_a \preceq \mathbf{x}_b \ \& \ \exists i \ s.t \ f_i(\mathbf{x}_a) < f_i(\mathbf{x}_b)$
- $\mathbf{x}_a \sim \mathbf{x}_b$ (neither dominates the other) i.f.f. $\mathbf{x}_a \not\prec \mathbf{x}_b$ and $\mathbf{x}_b \not\prec \mathbf{x}_a$

where $i \in \{1, \dots, m\}$. Any $\mathbf{x}_a \in \mathcal{X}$ can be said as Pareto-optimal i.f.f. $\nexists \mathbf{x}_b \in \mathcal{X} \ s.t. \ \mathbf{x}_a \prec \mathbf{x}_b$, given $\mathbf{x}_b \neq \mathbf{x}_a$. Hence, the Pareto-optimal set/Non-dominated solutions(NDS) \mathcal{P} and Non-dominated points (NDS)/Pareto-front $\mathbf{f}(\mathcal{P})$ can be defined as follows

- $\mathcal{P} := \{\mathbf{x}_a \in \mathcal{X} : \mathbf{x}_a \neq \mathbf{x}_b, \nexists \mathbf{x}_b \in \mathcal{X} \ s.t. \ \forall i, f_i(\mathbf{x}_a) \leq f_i(\mathbf{x}_b) \ \& \ \exists i \ s.t \ f_i(\mathbf{x}_a) < f_i(\mathbf{x}_b)\}$
- $\mathbf{f}(\mathcal{P}) := \{\mathbf{f}(\mathbf{x}_a) \in \mathcal{S} : \mathbf{x}_a \neq \mathbf{x}_b, \nexists \mathbf{x}_b \in \mathcal{X} \ s.t. \ \forall i, f_i(\mathbf{x}_a) \leq f_i(\mathbf{x}_b) \ \& \ \exists i \ s.t \ f_i(\mathbf{x}_a) < f_i(\mathbf{x}_b)\}$

Hence, Pareto-front in the context of MOO corresponds to the optimal solution of a function. Often distinguish is made between the true Pareto-front which corresponds to the true optimal of a function in single objective case and the observed Pareto-front known as empirical Pareto-front which corresponds to the known optimal of a function in single objective case. For the following definitions, we define NDS as empirical Pareto-front unless otherwise specified. Similar to the Utopian point in Single objective optimisation, Utopian point in MOO corresponds to a vector of all the known optimal solutions in \mathcal{S} . While the Nadir point defines the opposite, if the optimal points are defined to be the minimum of functions in MOO, Nadir point defines the maximum of all the functions in the NDS.

Before moving on to the Multi-objective Bayesian optimisation (MOBO), we introduce some of the standard approaches in MOO, since most of the approaches in MOBO has influence of the standard approaches where in MOBO typically acquisition functions are replaced for expensive functions and optimised in the context of MOO. Broadly MOO approaches can be classified in to two, where one approach typically converts MOO in to a set of single objective optimisations and optimise with classical single objective strategies. While the other approach works directly in the context of MOO, where Pareto-dominance measures are used directly in optimisation. The former approach is typically based on aggregation procedures, where a vector of objectives are scalarized by summation with assigning weights to each objective. Hence optimising the scalar function with different set of weights, one can obtain a set of Pareto-optimal points. The simplest of them can be given as $\sum_i^m w_i f_i(\mathbf{x})$, where $w_i \geq 0$ and $\sum_i^m w_i = 1$. The major drawback is that due to linear scalarization, only convex parts of the Pareto-front can be found. To over come this problem, the optimisation can be defined for any one function with other functions as constraints, or through modifying the scalarization as $\sum_i^m w_i (f_i(\mathbf{x}))^p$, where the parameter p should be defined a priori, which demands a priori knowledge of the Pareto-front. The other approach to generalize for non-convex type of Pareto-front is based on Tchebychev aggregartion which is simply the weighted norm in L_p metric as $p \rightarrow \infty$, given as

$$\max_i w_i (|f_i(\mathbf{x}) - f_i(\mathbf{x}^{i++})|) \quad (6.33)$$

where $f_i(\mathbf{x}^{i++})$ is the Utopian value of f_i . Normally the above term is augmented to avoid weakly Pareto-optimal solutions as

$$\max_i w_i (f_i(\mathbf{x}) - f_i(\mathbf{x}^{i++})) + \beta \sum_i^m |f_i(\mathbf{x}) - f_i(\mathbf{x}^{i++})| \quad (6.34)$$

where β can be any sufficiently small value. The main drawback with scalarization of the MOO problem is that it is hard to define a generic frame work considering flexibility, where it is highly sensitive to the parameters and need to be defined or actively learned with prior knowledge of the optimisation problem, where such prior

knowledge are typically unknown. The other major drawback is also with enhancing diversity which cannot be often explicitly achieved with a priori definition of weight vectors. This is also a set back in applications to define a focused search on the Pareto-front. This can be overcome with penalty boundary insertion strategies which further require a good prior knowledge of parameters to initialize.

The other class of approaches in dealing with MOO are nature-inspired approaches which are meta-heuristic, where we focus on GA which belongs to the class of Evolutionary Algorithms. GAs are inspired from concepts based on evolution such as fitness, natural selection, cross-over and mutation to guide the search for optimisation. GAs are typically more robust, and can handle discontinuities well. Constraint definitions are also more easy and generic with GAs. GAs developed for MOO share the same advantages along with other major advantages well-suited for MOO, where it allows the possibility to utilise the optimality measure in MOO such as Pareto-dominance and diversity directly. There are several methods exist under the concept multi-objective GAs, where typically the difference comes in defining the fitness measures and niching. We use NSGA-2 where the fitness measures are defined based on the ranking of the Pareto-front and the crowding distance. Some other typical variation of fitness measures are defined based on quality indicators of the Pareto-front such as Hypervolume (also known as \mathcal{S} -metric) and epsilon indicators.

Multi-objective Bayesian optimisation

Similar to the MOO problem, the Multi-objective Bayesian optimisation (MOBO) can be defined as through the scalarization approach of converting a multi-objective problem in to an aggregate set of single objective problems, where typically Augmented Tchebycheff aggregation is used. In [Knowles 2006](#), \mathcal{GP} for the scalarized function is fitted and the subsequent definition of EI is optimised. [Liu 2008](#) defined a variation of the approach, where the \mathcal{GP} models are independently fitted to each objective function and the EI of the \mathcal{GP} meta-models are scalarized, where the optimization of EI is defined in parallel for the aggregation. Further variations with scalarization approach typically involves the Penalty boundary insertion to maintain diversity of the NDS.

A more direct extension of the MOBO comes form the extension of the concept of EI to MOO in a direct or indirect sense. EI for a single objective optimization can be seen as the measure of improvement in lesbegue norm, and hence, a more direct extension of the EI for MOO leads to the defintion of improvement with lesbegue measure in higher dimensions, also known as S-metric or Hypervolume metric. This approach is first defined in [Emmerich 2008](#) which can be seen as the extension of $SMS - EMOEA$ [Emmerich 2005](#). Similar to the definition of improvement for a single objective case with respect to Utopian value, in MOO, the improvement is defined with respect to the empirical Pareto-front typically bounded by Nadir point.

The hypervolume (HV) can be given as

$$HV(\mathbf{f}(\mathcal{P}_S)|\mathbf{R}) = \mathcal{L}\left(\bigcup_{\mathbf{f}(\mathbf{p}_i) \in \mathbf{f}(\mathcal{P}_S)} \{\mathbf{f}(x) : \mathbf{f}(\mathbf{p}_i) \preceq \mathbf{f}(x) \preceq \mathbf{R}\}\right) \quad (6.35)$$

where \mathcal{L} is the lebesgue measure in \mathbb{R}^m when $\mathbf{f}(x) \in \mathbb{R}^m$, bounded by the reference point $\mathbf{R} \in \mathbb{R}^m$. The hypervolume improvement (HVI) can be defined as the hypervolume dominated by a point \mathbf{f}_o with respect to the Pareto-front $\mathbf{f}(\mathcal{P}_S)$ bounded by \mathbf{R} , given as

$$HVI(\mathbf{f}_o|\mathbf{f}(\mathcal{P}_S), \mathbf{R}) = HV(\mathbf{f}(\mathcal{P}_S) \cap \mathbf{f}_o|\mathbf{R}) - HV(\mathbf{f}(\mathcal{P}_S)|\mathbf{R}) \quad (6.36)$$

Similar to defining the expectation of improvement $E(I(\mathbf{x}))$ in single-objective case, for the multi-objective case, the expectation is defined over HVI as $E(HVI(\mathbf{x}))$, given as

$$EHVI(\mathbf{x}|\mathbf{f}(\mathcal{P}_S), \mathbf{R}) = \int_{\mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathcal{P}_S)|\mathbf{R}} HVI(\mathbf{f}(\mathbf{x})|\mathbf{f}(\mathcal{P}_S), \mathbf{R}).PDF(\hat{\mathbf{f}}(\mathbf{x}))d\mathbf{f}(\mathbf{x}) \quad (6.37)$$

where $\hat{\mathbf{f}} : \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ is the prediction defined by multivariate independent normal distribution ⁸. Even though there exists some correlation between the predictions through which Pareto-optimal solutions are presumed to exist, the multi-variate Gaussian prediction is defined to be independent to avoid complexity, i.e., $\hat{\boldsymbol{\Sigma}}$ is diagonal. And hence for the joint distribution $\hat{\mathbf{f}}(\mathbf{x})$, $PDF(\hat{\mathbf{f}}(\mathbf{x})) := \prod_{i=1}^m PDF\left(\frac{\hat{f}_i - \mu_i}{\sigma_i}\right)$. The maximum of $EHVI$ can be chosen as the infill point for an iteration in MOBO where batch selection cannot be defined since the scalar value of $EHVI$ contains no metric to compare for diversity. Any target based improvement is defined through weights or truncation applied to the $EHVI$ Auger 2009, Palar 2018. But the main disadvantage of this method is that it requires the computation of m dimensional hypervolume, where the integration is typically achieved by expensive Monte-Carlo methods. But recently, a formula has been proposed for any number of dimensions but the complexity still increases exponentially with number of objectives. Another approach similarly defined with S -metric is based on LCB (6.31) but still requires the expensive evaluation of HVI for LCB prediction to optimise for the maximum of HVI to choose the infill points. Further if the LCB is dominated, penalty value is assigned to avoid plateaus of the criterion.

The other approach is given in Keane 2006 for two objective case, where this approach interprets the geometric nature of EI in single-objective case – detailed in (6.28) – and extends to multi-objective case to define an equivalent metric. The improvement $I(\mathbf{x})$ is modelled implicitly by the Euclidean distance (L_2 norm) between the centroid \mathbf{f}_{cen} and the nearest point $\mathbf{f}_{near} \in \mathbf{f}(\mathcal{P}_S)$ to \mathbf{f}_{cen} . Hence, $E(I(\mathbf{x}))$ in this

⁸The multi-variate Gaussian prediction is obtained through defining a joint distribution of independent univariate Gaussian predictions from \mathcal{GP} meta-models

case is given as the product of $\|\mathbf{f}_{cen} - \mathbf{f}_{near}\|$ and the probability $\mathcal{P}(\mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathcal{P}_S) | \mathbf{R})$ for the given prediction $\hat{\mathbf{f}}(\mathbf{x})$ that can dominate $\mathbf{f}(\mathcal{P}_S)$ bounded by \mathbf{R} . \mathbf{f}_{cen} can be calculated from $\mathbf{f}_{cen} = \frac{E(\mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathcal{P}_S) | \mathbf{R})}{\mathcal{P}(\mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathcal{P}_S) | \mathbf{R})}$. For any prediction that does not dominate the Pareto-front, the improvement is given by so called augmented improvement. With the evaluation of the integral for higher dimensional $\hat{\mathbf{f}}(\mathbf{x})$, this method can become cumbersome.

EI has been extended to targeted MOO simply as the product of all the *EIs* of each objective, given as

$$mEI(x|\mathbf{R}) = \prod_i^m EI_i(x|\mathbf{R}_i) \quad (6.38)$$

This approach is computationally efficient for a targeted search, since it only deals with univariate distributions and hence, analytical evaluation of the criteria and gradient is possible for many objectives. One of the main concerns with this approach is choosing the reference vector \mathbf{R} considering the target which is presumed to be part of the true Pareto-front, and hence, the main emphasis of this approach is defining a proper \mathbf{R} for the target. In fact this approach is equivalent to *EHVI* under some hypothesis when \mathbf{R} is also a non-dominated point. As a default, the preference to the reference value is given to the centre of the Pareto-front, where the centre is identified by the closest point in Euclidian measure on the Utopian-Nadir line. With the user provided aspiration point \mathbf{R}_o , \mathbf{R} is adapted dynamically by projecting the closest point of the empirical Pareto-front to the broken line joining \mathbf{R}_o to the Utopian and the Nadir point.

Another approach was introduced in [Jeong](#) where *EIs*⁹ are treated as objectives and optimised in multi-objective context typically with MOEA algorithms like NSGA-2. The main advantage of this approach is that it is generic, since it preserves the generic characteristics of NSGA-2 in MOO and the analytical evaluation of *EI* criterion. Further the optimisation of *EIs* with NSGA-2 results in Pareto-optimal solutions of *EI* criteria where it expresses the optimality directly in the context of MOO. This is unlike the previous approaches where optimality is expressed through a scalar value, which does not provide much information in multi-objective context. This is efficient since we are dealing with a set of Pareto-optimal set of solutions from which the infill points can be chosen for diversity or targeted search, or even it provides the choice to scalarize with quality indicators. 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 hence discouraging such individuals for future generations.

The main drawback with this approach is defining the reference value for *EI* in the context of MOO, where typically Utopian or Nadir value is chosen as reference.

⁹Set of *EI* criterion

This can work in some cases, but largely choosing Utopian or Nadir value can be too optimistic or pessimistic to define Pareto-optimal measure of EI . For single objective optimisation, we look for the improvement $EI(\mathbf{x}|f(\mathbf{x}^+))$ where $f(\mathbf{x}^+)$ usually corresponds to the Utopian value $f(\mathbf{x}^{++})$. While in the context of MOO, the Utopian value of 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 the objective space. As well-known, single objective EI can be extended to MOO through $EHVI$ where the improvement for a multi-variate gaussian prediction is defined with respect to the empirical Pareto-front rather than a specific reference value. But the evaluation of integration to define the $EHVI$ for a multi-variate gaussian prediction can be cumbersome for large number of objectives. Hence, we extend the work of Jeong by defining multiple reference values in the MOO of EIs to define infill points for MOBO, also while preserving the generic characteristics of the approach. This means that different regions of the 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 probabilistic sense with the following explanation.

The MOO of EIs for $\min_{\mathbf{x} \in \mathcal{X}} \mathbf{f}(\mathbf{x})$ can be defined as

$$\max_{\mathbf{x} \in \mathcal{X}} \left[EI_{\min f_1(\mathbf{x})}(\mathbf{x}|f_1(\mathbf{x}^{1+})) \cdots EI_{\min f_m(\mathbf{x})}(\mathbf{x}|f_m(\mathbf{x}^{m+})) \right] \quad (6.39)$$

We consider the perspective from a single-objective $\max EI(\mathbf{x}|f_i(\mathbf{x}^{i+}))$ in the above MOO, so as to detail the independent effect of defining a reference value $f_i(\mathbf{x}^{i+})$ for $\max EI(\mathbf{x}|f_i(\mathbf{x}^{i+}))$ in characterising the Pareto-optimal solutions. While a Pareto-front could be achieved with a single reference value $f_i(\mathbf{x}^{i+})$ for $\max EI(\mathbf{x}|f_i(\mathbf{x}^{i+}))$, the resolution to define improvement is only efficient for a subset of the decision space depending on the \mathcal{GP} prediction relative to $f_i(\mathbf{x}^{i+})$. This can be seen through the following cases for $EI_{\min f_i(\mathbf{x})}$ ¹⁰ as follows:

Case 1 : For $\hat{\mu}_i(\mathbf{x}) \pm \hat{\sigma}_i(\mathbf{x}) \gg f_i(\mathbf{x}^{i+})$, $EI_{\min f_i(\mathbf{x})}(\mathbf{x}|f_i(\mathbf{x}^{i+})) \approx 0$ (6.40)

Case 2 : For $\hat{\mu}_i(\mathbf{x}) \pm \hat{\sigma}_i(\mathbf{x}) \ll f_i(\mathbf{x}^{i+})$, $EI_{\min f_i(\mathbf{x})}(\mathbf{x}|f_i(\mathbf{x}^{i+})) \approx f_i(\mathbf{x}^{i+}) - \hat{\mu}_i(\mathbf{x})$

Case 1 shows that the choice of the Utopian value $f_i(\mathbf{x}^{i+})$ to seek improvement for a subset of \mathbf{x} on parts of the objective space where $\hat{\mu}_i(\mathbf{x}) \pm \hat{\sigma}_i(\mathbf{x}) \gg f_i(\mathbf{x}^{i+})$ 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, zero or infinitesimal values provide less resolution for comparison to define NDS for MOO.

¹⁰The following cases are shown specifically for the case of $\min f_i(\mathbf{x})$, while for $\max f_i(\mathbf{x})$, the relations are inversed

Case 2 shows that the measure of improvement is simply given by the distance between the prediction $\hat{\mu}_i(\mathbf{x})$ and the reference value $f_i(\mathbf{x}^{i+})$, which can only be 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 parts of the objective space where $\hat{\mu}_i(\mathbf{x}) \pm \hat{\sigma}_i(\mathbf{x}) \ll f_i(\mathbf{x}^{i+})$ and hence the absence of the uncertainty term $\hat{\sigma}_i(\mathbf{x})$ in evaluating $\frac{EI}{\min f_i(\mathbf{x})}$. The above two cases show the limitations of using a single reference value for EI and hence, it is efficient to define improvement locally in the objective space. The above cases are graphically shown in Fig. 6.7

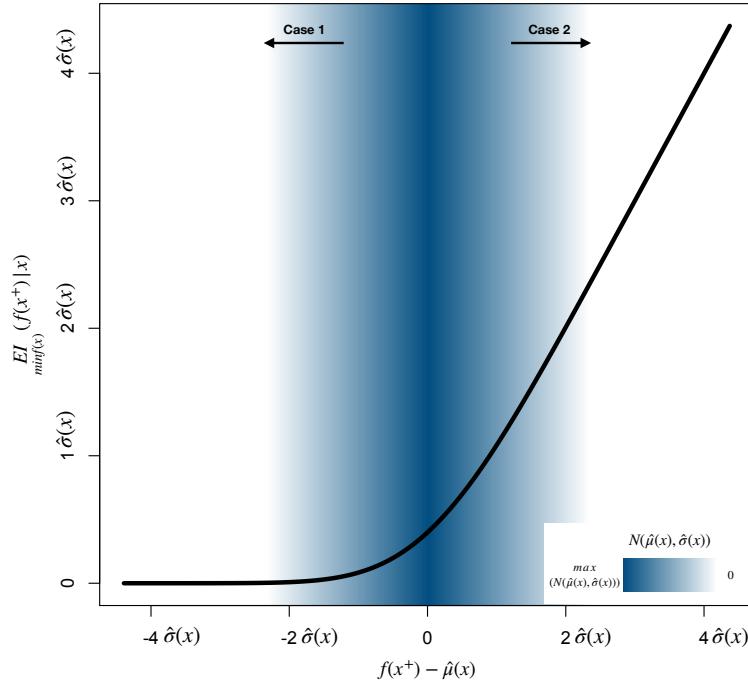


Figure 6.7: Variation of $\frac{EI}{\min f_i(\mathbf{x})} (f(\mathbf{x}^+) | \mathbf{x})$ for change in reference value $f(\mathbf{x}^+)$

As a first thought, this can be overcome by choosing an appropriate reference value $f_i(\mathbf{x}^{i+})$ depending on the prediction $\hat{f}_i(\mathbf{x})$. But choosing a reference value $f_i(\mathbf{x}^{i+})$ to define EI depending on where $\hat{f}_i(\mathbf{x})$ lies in the objective space can make the improvements hard to be compared, since different predictions can have different definition of improvement depending on the choice of the reference value. The comparison is also essential for defining NDS for optimization in the context of MOO. 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 augmentation of EI with $f_i(\mathbf{x}^{i+})$ as reference value is given through the criterion

expected value (EV)¹¹ as follows

$$EV_{\min f_i(\mathbf{x})}(\mathbf{x}|f_i(\mathbf{x}^{i+})) = f_i(\mathbf{x}^{i+}) - EI_{\min f_i(\mathbf{x})}(\mathbf{x}|f_i(\mathbf{x}^{i+})) \quad (6.41)$$

The definition of EV avoids the problem of comparing improvements with several reference values, but the above two cases take a different role for the EV criterion, given as

$$\begin{aligned} \text{Case 1 : For } \hat{\mu}_i(\mathbf{x}) \pm \hat{\sigma}_i(\mathbf{x}) &\gg f_i(\mathbf{x}^{i+}), \quad EV_{\min f_i(\mathbf{x})}(\mathbf{x}|f_i(\mathbf{x}^{i+})) \approx f_i(\mathbf{x}^{i+}) \\ \text{Case 2 : For } \hat{\mu}_i(\mathbf{x}) \pm \hat{\sigma}_i(\mathbf{x}) &\ll f_i(\mathbf{x}^{i+}), \quad EV_{\min f_i(\mathbf{x})}(\mathbf{x}|f_i(\mathbf{x}^{i+})) \approx \hat{\mu}_i(\mathbf{x}) \end{aligned} \quad (6.42)$$

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 the problem of overestimation, 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**, where EV simply converges to $\hat{\mu}_i(\mathbf{x})$. This means that two predictions with the same $\hat{\mu}_i(\mathbf{x})$ but different $\hat{\sigma}_i(\mathbf{x})$ will have the same preference. The above cases are graphically shown in Fig. 6.8

We give the explanation in avoiding the above cases with EV criterion since it is much easier to work in the coordinates of the objective space. As defined before, **Case 1** can be avoided by seeking improvement with respect to a reference value which is more realistic for improvement rather than being too greedy. The realistic scope of improvement considering $f_i(\mathbf{x}^{i+})$ can be defined through the CDF : $\mathcal{P}(f_i(\mathbf{x}) \leq f_i(\mathbf{x}^{i+}))$ for any \mathcal{GP} outcome $\hat{f}_i(\mathbf{x}) : \mathcal{N}(\hat{\mu}_i(\mathbf{x}), \hat{\sigma}_i(\mathbf{x}))$. We here remind that the given CDF is essentially the 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 measure to be considered with compromise on greed. The limit could be set in terms of $\hat{\mu}_i(\mathbf{x}) - \epsilon \hat{\sigma}_i(\mathbf{x})$ ¹² where we seek for $\mathcal{P}(f_i(\mathbf{x}) \leq f_i(\mathbf{x}^{i+})) \geq \mathcal{P}(f_i(\mathbf{x}) \leq \hat{\mu}_i(\mathbf{x}) - \epsilon \hat{\sigma}_i(\mathbf{x}))$, 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 higher value of ϵ for $\hat{\mu}_i(\mathbf{x}) - \epsilon \hat{\sigma}_i(\mathbf{x})$ may reap higher benefits but this could be otherwise, since there is equal probability for $f_i(\mathbf{x}) \geq \hat{\mu}_i(\mathbf{x}) + \epsilon \hat{\sigma}_i(\mathbf{x})$. Hence, this requires the right balance for the choice of ϵ with acceptable risk for exploration.

¹¹We here remind that the expected value can also be otherwise defined as $E(\hat{f}(\mathbf{x}))$ in the interval $(-\infty, f(\mathbf{x}^{++}))$ (detailed in (6.28)) contrary to EI which is defined as $E(f(\mathbf{x}^{++}) - \hat{f}(\mathbf{x}))$, but considering $E(\hat{f}(\mathbf{x}))$ leads to similar properties as EI in comparing with several reference values

¹²We give the relation in terms of LCB , since for any $f(\mathbf{x}^+) : \mathcal{P}(f(\mathbf{x}) \leq f(\mathbf{x}^+)) \geq \mathcal{P}(f(\mathbf{x}) \leq \hat{\mu}(\mathbf{x}) - \epsilon \hat{\sigma}(\mathbf{x}))$, the following relation holds $f(\mathbf{x}^+) \geq \hat{\mu}(\mathbf{x}) - \epsilon \hat{\sigma}(\mathbf{x})$

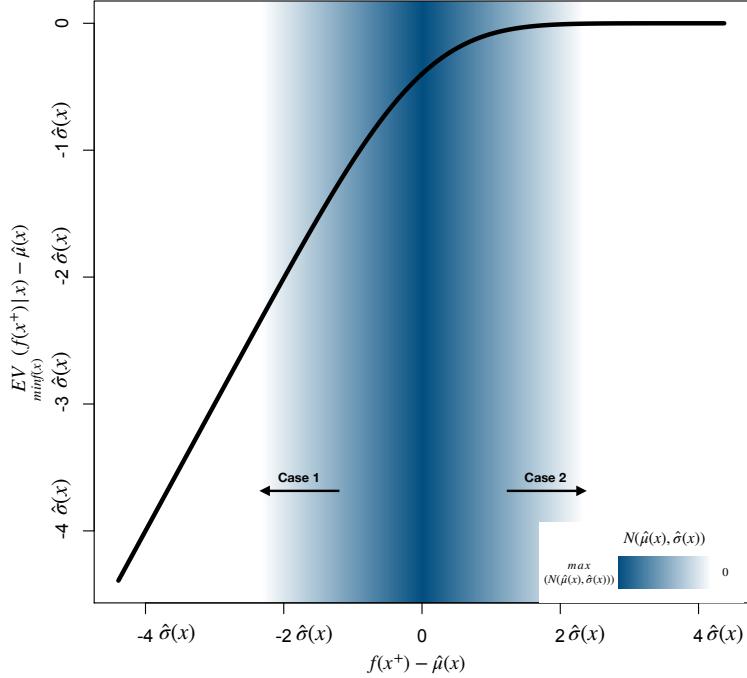


Figure 6.8: Variation of $\underset{\min f_i(\mathbf{x})}{EV} (f(\mathbf{x}^+) | \mathbf{x})$ for change in reference value $f(\mathbf{x}^+)$

For MOO, the most feasible improvement that we can look for is with respect to the empirical Pareto-front $f_i(\mathcal{P}_S) := \{f_i(\mathbf{p}_1), f_i(\mathbf{p}_2), \dots, f_i(\mathbf{p}_{(.)})\}$ of the observed samples. But, there can be multiple $\mathbf{p}_i : f_i(\mathbf{p}_i) \leq \hat{\mu}_i(\mathbf{x}) - \epsilon \hat{\sigma}_i(\mathbf{x})$, where we choose the \mathbf{p}_i which gives the minimum value of $\underset{\min f_i(\mathbf{x})}{EV}$. The choice of the minimum value of $\underset{\min f_i(\mathbf{x})}{EV}$ for a function to be minimised intrinsically avoids **Case 2** which defines the pessimistic choice of reference value, unless it is not possible when no suitable reference value exists. In overall, the above definitions provide the balance between greed and being pessimistic for improvement through the parameter ϵ . The above definitions could be expressed as follows

$$\begin{aligned} \underset{\min f_i(\mathbf{x})}{EV} (\mathbf{x} | \mathcal{P}_S, \epsilon) &= \min_{\mathbf{p}_i \in \mathcal{P}_s} (\underset{\min f_i(\mathbf{x})}{EV} (f_i(\mathbf{p}_i) | \mathbf{x})) \\ \mathcal{P}_s &:= \{f_i(\mathbf{p}_i) \geq \hat{\mu}_i(\mathbf{x}) - \epsilon \hat{\sigma}_i(\mathbf{x}), \forall \mathbf{p}_i \in \mathcal{P}_S\} \end{aligned} \quad (6.43)$$

where the discrete optimisation of $\min_{\mathbf{p}_i \in \mathcal{P}_s} (\underset{\min f_i(\mathbf{x})}{EV} (f_i(\mathbf{p}_i) | \mathbf{x}))$ can be implicitly satisfied through defining NDS with $\underset{\min f_i(\mathbf{x})}{EV} (\mathbf{x} | \mathcal{P}_S, \epsilon)$. Alternatively, from the nature of proportionality for $\underset{\min f_i(\mathbf{x})}{EV} (f_i(\mathbf{p}_i) | \mathbf{x}) \propto f_i(\mathbf{p}_i)$, shown in Fig. 6.8, the most suitable reference value $f_i(\mathbf{p}_i^*)$ to achieve $\min_{\mathbf{p}_i \in \mathcal{P}_s} (\underset{\min f_i(\mathbf{x})}{EV} (f_i(\mathbf{p}_i) | \mathbf{x}))$ can be simply given as

$f(\mathbf{p}_i^*) = \min_{\mathbf{p}_i \in \mathcal{P}_s} f(\mathbf{p}_i)$. In short, we define the Eq. (6.43) as EV criterion for the following discussions.

The given definitions allow to define EV criterion in the place of EI , where the problem (6.39) can be redefined as follows

$$\min_{\mathbf{x} \in \mathcal{X}} \left[\min_{f_1(\mathbf{x})} EV(\mathbf{x} | \mathcal{P}_S, \epsilon) \cdots \min_{f_m(\mathbf{x})} EV(\mathbf{x} | \mathcal{P}_S, \epsilon) \right] \quad (6.44)$$

The MOO can be achieved through NSGA-2, which leads to Pareto-optimal solutions ($\mathcal{P}_{\hat{S}}$). The independent definition of improvement allows to work with only univariate Gaussian distributions and hence, large number of EV criteria could be optimised with relatively ease.

The primary goal for choosing $f_i(\mathbf{p}_i) \in f_i(\mathcal{P}_S)$ independently for each $\min_{f_i(\mathbf{x})} EV(\mathbf{x} | \mathcal{P}_S, \epsilon)$ is that it can lead to choosing reference points $\mathbf{f}(\mathbf{p}_i) \in \mathbf{f}(\mathcal{P}_S)$, where the intention is to define improvements $\mathcal{J}_{\mathbf{f}(\mathbf{p}_i)}$ ¹³ (Fig. 6.9). The choice of $f_i(\mathbf{p}_i)$ independently for each $\min_{f_i(\mathbf{x})} EV(\mathbf{x} | \mathcal{P}_S, \epsilon)$ means that it can implicitly lead to choosing reference points $[\mathbf{g}_1 = f_1(\mathbf{p}_i) \in f_1(\mathcal{P}_S), \dots, \mathbf{g}_m = f_m(\mathbf{p}_i) \in f_m(\mathcal{P}_S)]$ ¹⁴ $\in \mathbb{R}^m$ from a grid of points $\mathcal{G} := \{f_1(\mathcal{P}_S) \times \dots \times f_m(\mathcal{P}_S)\}$. This means that $\mathbf{f}(\mathcal{P}_S) \subset \mathcal{G}$, where \mathcal{G} also contains points which dominate or be dominated by points in the set $\mathbf{f}(\mathcal{P}_S)$, which includes the Utopian point: $\mathbf{g}_o \preceq \mathbf{f}(\mathcal{P}_S)$ ¹⁵ and the Nadir point: $\forall i, \mathbf{f}(\mathbf{p}_i) \preceq \mathbf{g}_o$. This means that improvements can not only be defined to dominate any empirical Pareto-front point $\mathbf{f}(\mathbf{p}_i) \in \mathbf{f}(\mathcal{P}_S)$, but also multiple Pareto-front points in the set $\mathbf{f}(\mathcal{P}_S)$.

With the population based evolutionary strategy of NSGA-2, for a given population, the optimisation is simultaneously defined with several reference points in the objective space. This means that the individuals in a given generation of NSGA-2 are defined with their respective reference point and since the improvements are represented as expected values in the objective space, comparisons can be made to define non-dominated sorting and niching operations. This implicitly models optimisation of several EIs in the objective space, with each EI defined with a unique reference point, illustrated in 6.11.

An example for the optimisation of (6.44) is shown in Fig. , for a simple MOO test functions set: ZDT1 of two dimensions ($m = 2$). In this case, $\mathbf{f}(\mathcal{P}_S) = \{\mathbf{R}_1, \mathbf{R}_6, \mathbf{R}_{11}, \mathbf{R}_{16}\}$, $\mathcal{G} = \{\mathbf{R}_1, \dots, \mathbf{R}_{16}\}$. The Pareto front $\mathbf{f}(\mathcal{P}_{\hat{S}})$ is defined by the reference points $\mathbf{R}_2, \mathbf{R}_6, \mathbf{R}_{10}, \mathbf{R}_{11}, \mathbf{R}_{16}$, where two of the reference points \mathbf{R}_6 and \mathbf{R}_{11} belong to \mathcal{P}_S . This means that the improvements $\mathcal{J}_{\mathbf{R}_6}$ and $\mathcal{J}_{\mathbf{R}_{11}}$ independently define

¹³ $\mathcal{J}_{\mathbf{f}^*}(\mathbf{x}) \in \mathcal{J}_{\mathbf{f}^*} = \left\{ \left[\min_{f_1(\mathbf{x})} EV(\mathbf{x} | f_1^*) \cdots \min_{f_m(\mathbf{x})} EV(\mathbf{x} | f_m^*) \right] \mid \forall i, f_i^* = \min_{\mathbf{p}_i \in \mathcal{P}_s} f_i(\mathbf{p}_i), \forall \mathbf{x} \in \mathcal{X} \right\}$,
 $\mathbf{f}^* = [f_1^* \cdots f_m^*]$

¹⁴ It should be noted that the choice of \mathbf{p}_i is independent for each f_i

¹⁵ \mathbf{g}_o is an arbitrary point in \mathcal{G} with index o

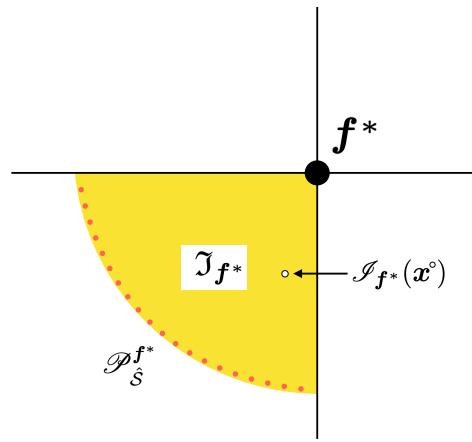


Figure 6.9: Illustration of the set $\mathcal{I}_{f(p_i)}$ and its associated subsets

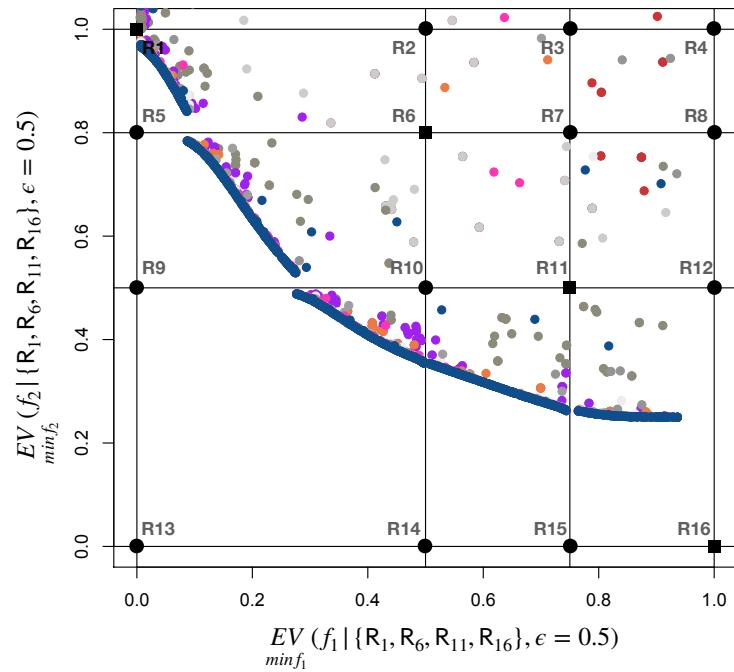


Figure 6.10: An example for optimization of EV

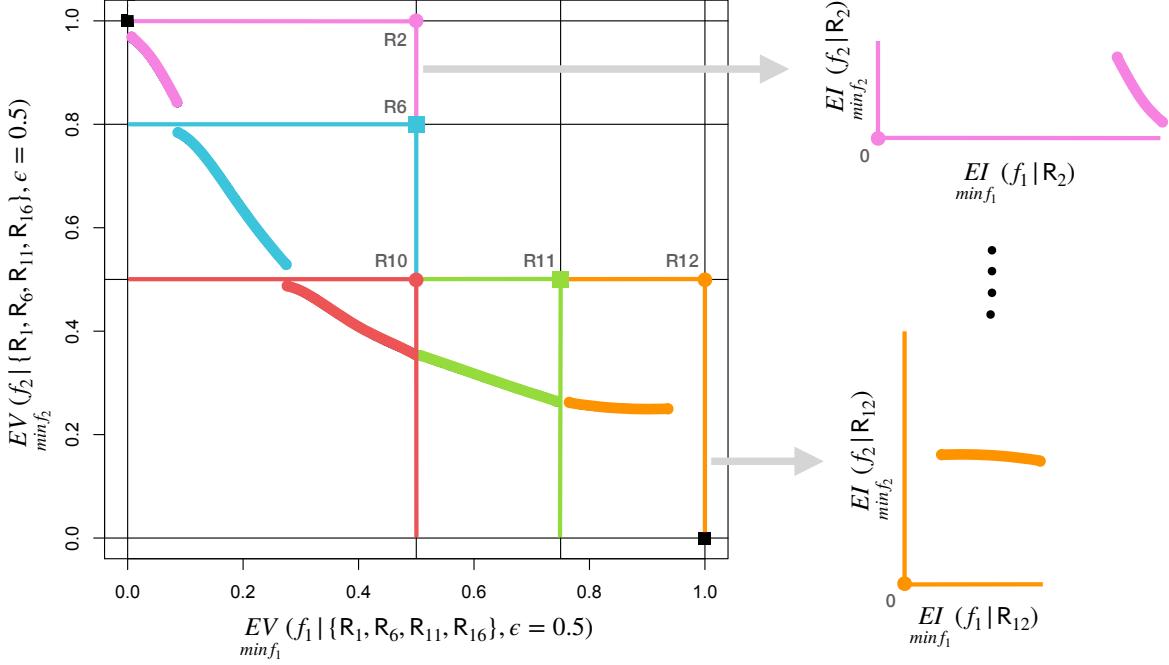


Figure 6.11: Relation of EV to reference point

domination over \mathbf{R}_6 and \mathbf{R}_{11} respectively, as $\mathcal{I}_{\mathbf{R}_6} \preceq \mathbf{R}_6$ and $\mathcal{I}_{\mathbf{R}_{11}} \preceq \mathbf{R}_{11}$. For \mathbf{R}_{10} , something interesting happens, where $\mathbf{R}_{10} \notin \mathcal{P}_S$ but is purely an outcome of the implicit definition of grid \mathcal{G} . In this case, $\mathbf{R}_{10} \preceq \{\mathbf{R}_6, \mathbf{R}_{11}\}$, hence also the improvements $\mathcal{I}_{\mathbf{R}_{10}} \preceq \{\mathbf{R}_6, \mathbf{R}_{11}\}$, i.e., any improvement $\mathcal{I}_{\mathbf{R}_{10}} \in \mathcal{I}_{\mathbf{R}_{10}}$ dominates the two Pareto-front points \mathbf{R}_6 and \mathbf{R}_{11} . Hence, \mathcal{G} consists of reference points to define improvements with respect to any $\mathbf{f}(p_i)$ or combination of several $\mathbf{f}(p_i)$. \mathcal{G} also contains points $\mathbf{g}_o : \mathbf{f}(\mathcal{P}_S) \preceq \mathbf{g}_o$ where the improvements defined with such points are essential for guiding the solutions to define $\mathcal{P}_{\hat{S}}$. The reference points \mathbf{R}_2 and \mathbf{R}_{12} are also weakly dominated by the points in the set $\mathbf{f}(\mathcal{P}_S)$, where in this case, the improvements $\mathcal{I}_{\mathbf{R}_2}$ are presumed to define the intermediate points between \mathbf{R}_1 and \mathbf{R}_6 , and similarly the improvements $\mathcal{I}_{\mathbf{R}_{12}}$ are presumed to define the intermediate points between \mathbf{R}_{11} and \mathbf{R}_{16} . In this example, the Utopian point corresponds to $\mathbf{R}_{12} \preceq \mathbf{f}(\mathcal{P}_S)$ and the Nadir point corresponds to \mathbf{R}_4 where all the points in \mathcal{G} except \mathbf{R}_4 dominates \mathbf{R}_4 .

Once $\mathcal{P}_{\hat{S}}$ is obtained optimising Eq. (6.44) with NSGA-2, the problem leads to choosing the infill points among $\mathcal{P}_{\hat{S}}$. Firstly, we discuss the preference of choosing an improvement $\mathcal{I}_{\mathbf{f}^*} \in \mathcal{I}_{\mathbf{f}^*}$, where the reference point \mathbf{f}^* can be considered to define a part of $\mathbf{f}(\mathcal{P}_{\hat{S}})$, i.e., $\{\mathcal{I}_{\mathbf{f}^*} \cap \mathbf{f}(\mathcal{P}_{\hat{S}})\} \neq \emptyset$. The part of the Pareto-front defined by \mathbf{f}^* can hence be expressed as $\mathcal{P}_{\hat{S}}^{\mathbf{f}^*} = \{\mathcal{I}_{\mathbf{f}^*} \cap \mathbf{f}(\mathcal{P}_{\hat{S}})\}$, where the choice of an infill point in the set $\mathcal{I}_{\mathbf{f}^*}$ can be narrowed down to choosing a point $\mathcal{I}_{\mathbf{f}^*} \in \mathbf{f}(\mathcal{P}_{\hat{S}}^{\mathbf{f}^*})$. A suitable measure to quantify $\mathcal{I}_{\mathbf{f}^*}$ with respect to \mathbf{f}^* can be given through hypervolume metric

as

$$HVI(\mathcal{I}_{f^*} | \mathbf{f}^*) = \prod_{i=1}^m f_i^* - \frac{EV}{\min f_i(\mathbf{x})}(\mathbf{x} | f_i^*) \quad (6.45)$$

Hence, the most suitable choice of $\mathcal{I}_{f^*} \in \mathbf{f}(\mathcal{P}_{\hat{\mathcal{S}}}^{f^*})$ is which maximises $HVI(\mathcal{I}_{f^*} | \mathbf{f}^*)$, expressed as $\max_{\mathcal{I}_{f^*} \in \mathbf{f}(\mathcal{P}_{\hat{\mathcal{S}}}^{f^*})} HVI(\mathcal{I}_{f^*} | \mathbf{f}^*)$. It should be noted that, in this case, $HVI(\mathcal{I}_{f^*} | \mathbf{f}^*)$

is equivalent to $EHVI(\mathbf{x} | \mathbf{f}^*)$, which is also similar to the criteria mEI detailed in , since $f_i^* - \frac{EV}{\min f_i(\mathbf{x})}(\mathbf{x} | f_i^*) = \frac{EI}{\min f_i(\mathbf{x})}(\mathbf{x} | f_i^*)$. The calculation of HVI in this case is analytical and simple irrespective of the dimensions.

The choice of $\max_{\mathcal{I}_{f^*} \in \mathbf{f}(\mathcal{P}_{\hat{\mathcal{S}}}^{f^*})} HVI(\mathcal{I}_{f^*} | \mathbf{f}^*)$ essentially expresses the best HVI with respect to \mathbf{f}^* , but it makes no relation of HVI relative to $\mathbf{f}(\mathcal{P}_S)$. While HVI relative to $\mathbf{f}(\mathcal{P}_S)$ can simply be expressed as $HVI(\mathcal{I} | \mathbf{f}(\mathcal{P}_S), \mathbf{R})$ ¹⁶, this is quite an expensive strategy to pick an infill point from $\mathcal{P}_{\hat{\mathcal{S}}}$, since this requires evaluation of $HVI(\mathcal{I} | \mathbf{f}(\mathcal{P}_S), \mathbf{R})$ for all $\mathcal{I} \in \mathcal{P}_{\hat{\mathcal{S}}}$. Since HVI with respect to \mathbf{f}^* is known, it can be said that the relation of \mathcal{I}_{f^*} with respect to \mathcal{P}_S can be defined by counting the $HVI(\mathbf{f}^* | \mathbf{f}(\mathcal{P}_S), \mathbf{R})$, expressed as

$$HVI_{\oplus}(\mathcal{I}_{f^*} | \mathbf{f}(\mathcal{P}_S), \mathbf{R}) = HVI(\mathcal{I}_{f^*} | \mathbf{f}^*) + HVI(\mathbf{f}^* | \mathbf{f}(\mathcal{P}_S), \mathbf{R}) \quad (6.46)$$

The idea is that instead of defining $HVI(\mathcal{I} | \mathbf{f}(\mathcal{P}_S), \mathbf{R})$ for all $\mathcal{I} \in \mathcal{P}_{\hat{\mathcal{S}}}$, the computation of higher complexity $HVI((.) | \mathbf{f}(\mathcal{P}_S), \mathbf{R})$ is limited to the reference points $\mathbf{g}_o \in \mathcal{G}$ for which $\{\mathcal{I}_{g_o} \cap \mathcal{P}_{\hat{\mathcal{S}}}\} \neq \emptyset$. With the definition of $HVI((.) | \mathbf{f}(\mathcal{P}_S), \mathbf{R})$ for a \mathbf{g}_o , the definition of HVI for any improvement $\mathcal{I}_{f^*} \in \mathcal{I}_{f^*}$ is given by simple product relation in (6.45). The infill point from $\mathcal{P}_{\hat{\mathcal{S}}}$ to define the maximum HVI relative to \mathcal{P}_S can be expressed as

$$\mathbf{x}' = \arg \max_{\mathbf{x} \in \mathcal{X}} \left\{ HVI_{\oplus}(\mathcal{I}_{g_o}(\mathbf{x}) | \mathbf{f}(\mathcal{P}_S), \mathbf{R}) \mid \{\mathcal{I}_{g_o} \cap \mathbf{f}(\mathcal{P}_{\hat{\mathcal{S}}})\} \neq \emptyset, \forall \mathbf{g}_o \in \mathcal{G}, \forall \mathbf{x} \in \mathcal{X} \right\} \quad (6.47)$$

Essentially, we scalarise $\mathcal{I} \in \mathcal{P}_{\hat{\mathcal{S}}}$ to find a single infill point \mathbf{x}' . But this necessarily does not have to be the case, since we have the optimal measure of \mathcal{I} in multiobjective context as $\mathcal{P}_{\hat{\mathcal{S}}}$, this gives way to lot of possibilities of choosing infill points from $\mathcal{P}_{\hat{\mathcal{S}}}$, considering simultaneous improvement in hypervolume with batch-selection and enhancing diversity. We do not yet focus on defining batch-selection in this thesis, exploiting the property of \mathcal{G} .

We focus on the approach of clustering improvements in $\mathcal{P}_{\hat{\mathcal{S}}}$ through K -means and a point can be sampled from each cluster to enhance diversity, which also defines

¹⁶ \mathcal{I} is an arbitrary improvement point where we do not focus on the reference point with which it is defined, contrary to \mathcal{I}_{f^*}

the possibility of parallelization. Considering simplicity, we chose the most uncertain point in each cluster, given by the uncertainty of the \mathcal{GP} models. Hence, this strategy tries to reduce uncertainty on the \mathcal{GP} meta-models for the points in $\mathcal{P}_{\mathcal{S}}$, rather than to seek improvement through quality indicators. But care should be taken, since adjacent points between 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 show the effect of ϵ through the following plots, where the idea is to show the behaviour for a range of predictions. The range is defined for $\hat{\mu}(\mathbf{x}) \in [50, 200]$, distinguished with color scale and $\hat{\sigma}(\mathbf{x}) \in [5, 50]$ defined on the vertical axis. The horizontal axis defines the *EV* criteria with four possible reference values $\{62.5, 75, 100, 150\}$ to choose for any prediction. Firstly, to have a general idea, we give the plot of $\hat{\mu}(\mathbf{x})$ against $\hat{\sigma}(\mathbf{x})$

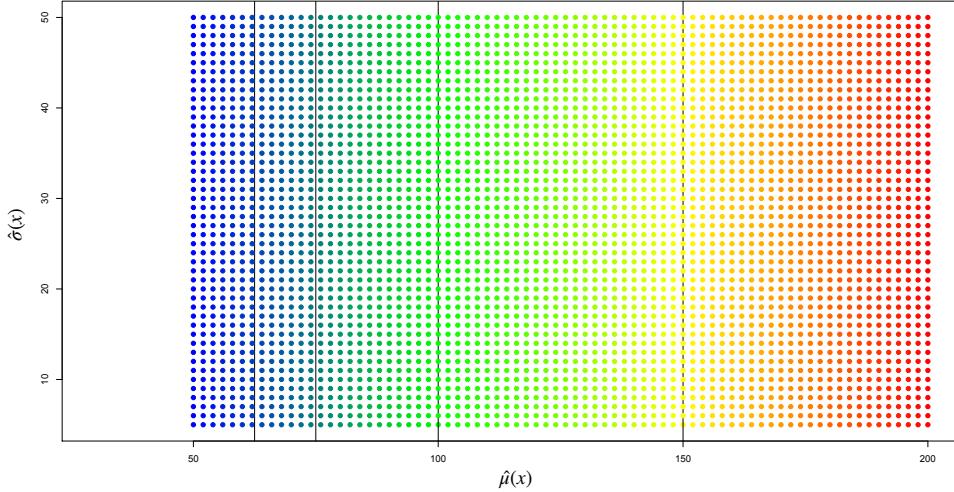
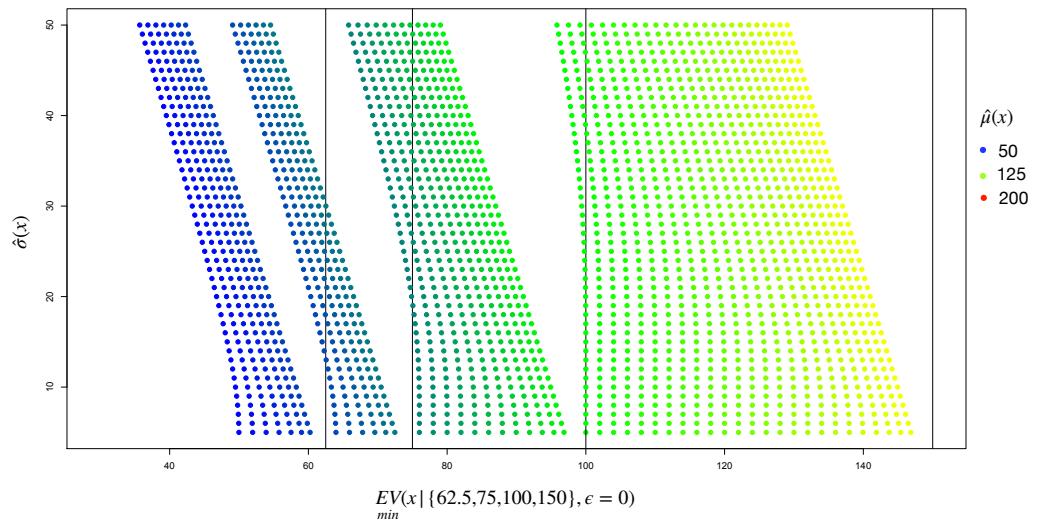
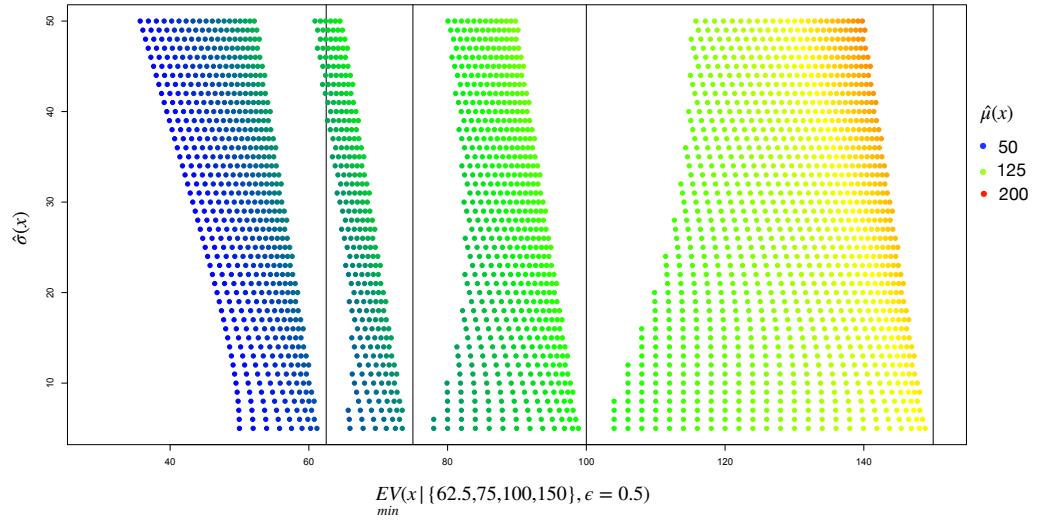


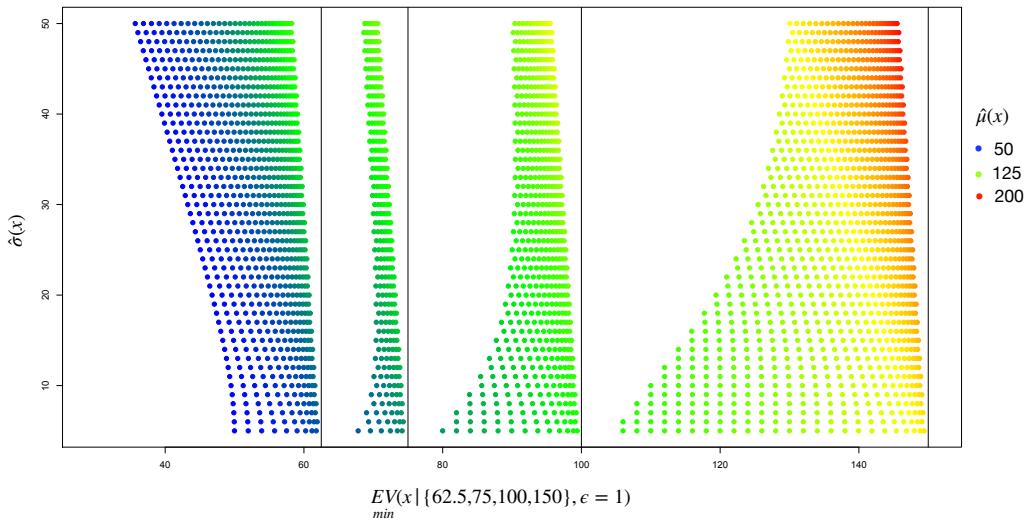
Figure 6.12: Plot showing the characteristics of choosing



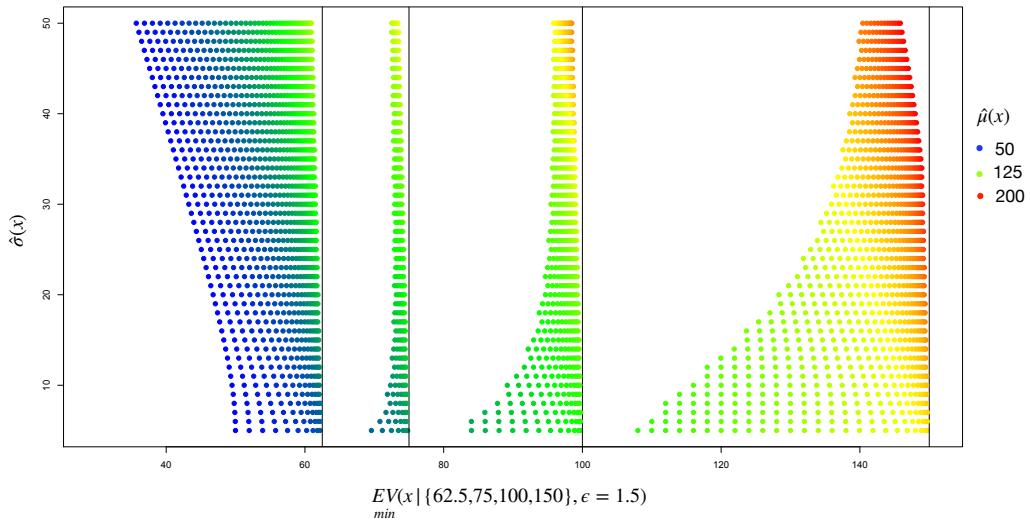
(a) $\epsilon = 0$



(b) $\epsilon = 0.5$



(a) $\epsilon = 1$



(b) $\epsilon = 1.5$

Some of the special cases where it converges to the cases of Nadir to Utopian value of reference is shown below

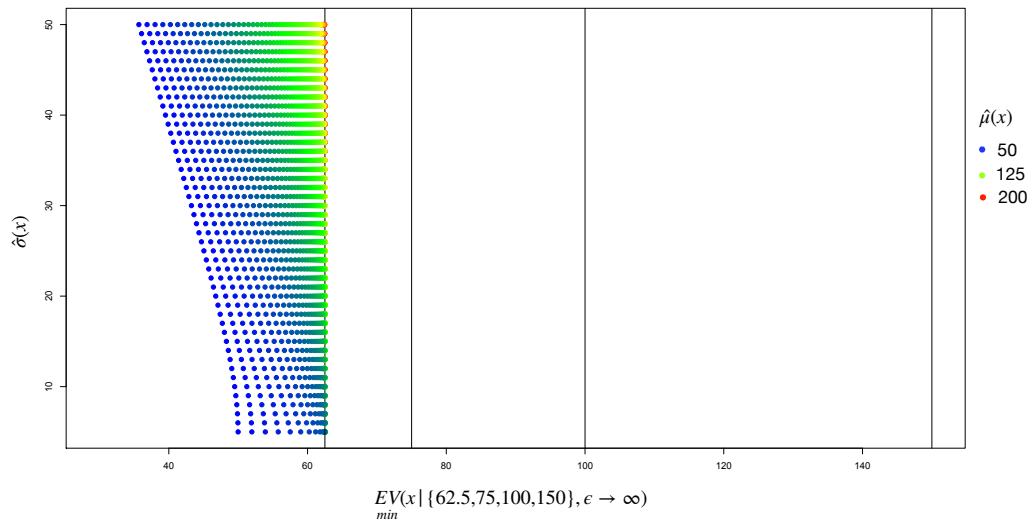


Figure 6.13: $\epsilon \rightarrow \infty$

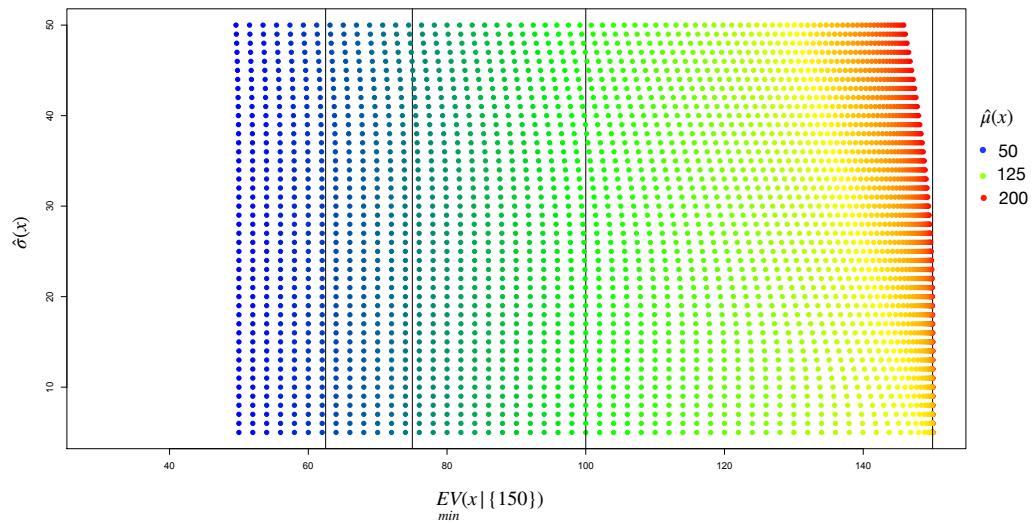


Figure 6.14: EV with Nadir value as reference

7 IGA based shape optimisation

All the results presented in the following discussion were obtained for the objectives defined in section ???. 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{GPs} .

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 7.1 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 which use several reference values.

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

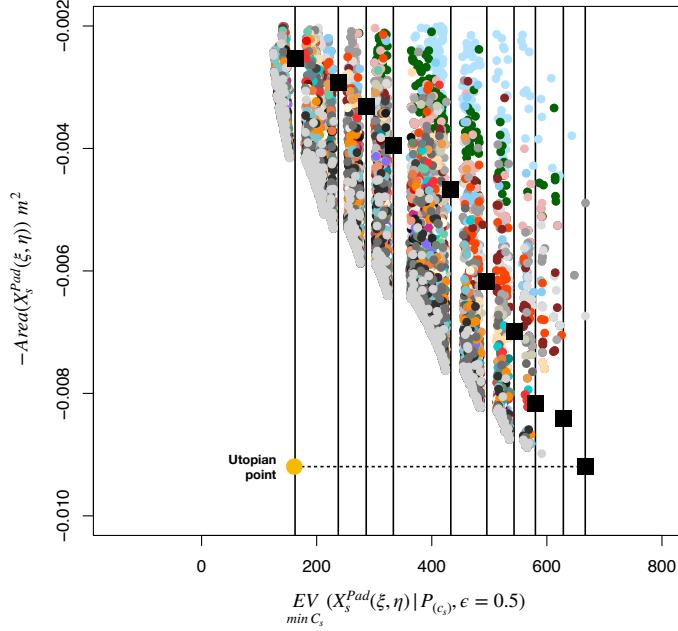
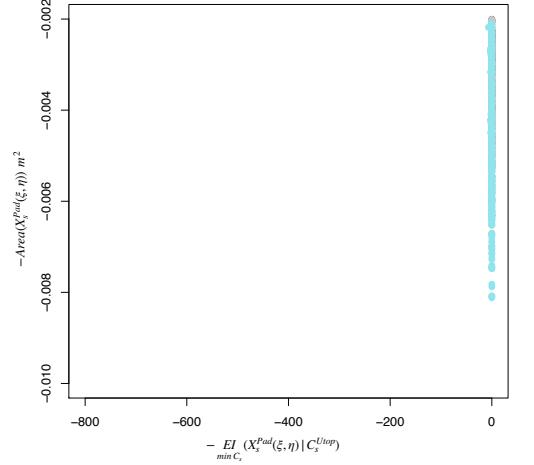


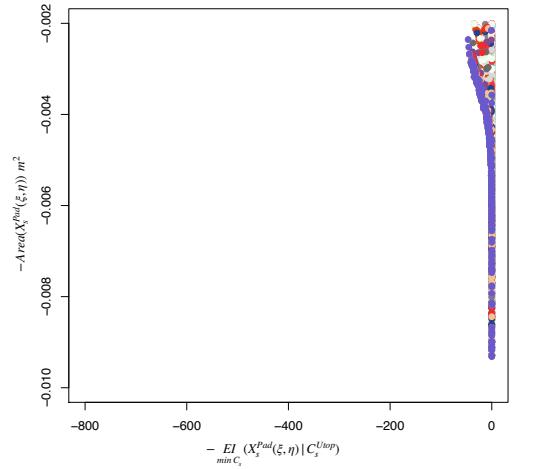
Figure 7.1: Optimization for infill points with EV criterion and $-Area(X_s^{Pad}(\xi, \eta))$. The colored points represent the population, with each color representing a generation obtained through NSGA-2. The black squared points are the NDS ($P_{(c_s)}$) in the objective space and the vertical lines 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.

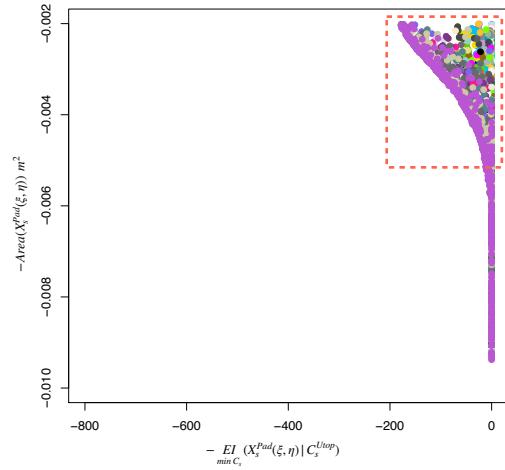
In general, with EV criterion, the initial generation starts with largely individuals whose eligibility for reference value will be far from the Utopian value –where the eligibility depends on the parameter ϵ in (6.43)– and through the progression of the generations will reach the individuals which will be eligible for improvement regarding the Utopian value. This is shown in Figure 7.3 for the EV criterion with $\epsilon = 1$, along with comparison for EI criterion in Figure 7.2 with Utopian value as reference, where a same \mathcal{GP} model was used for both the cases.



(a) Iteration 4:

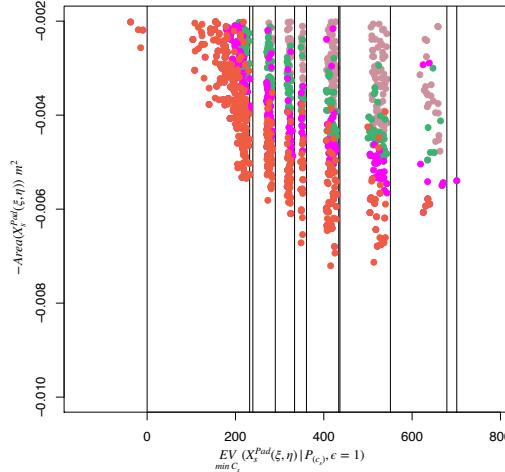


(b) Iteration 22:

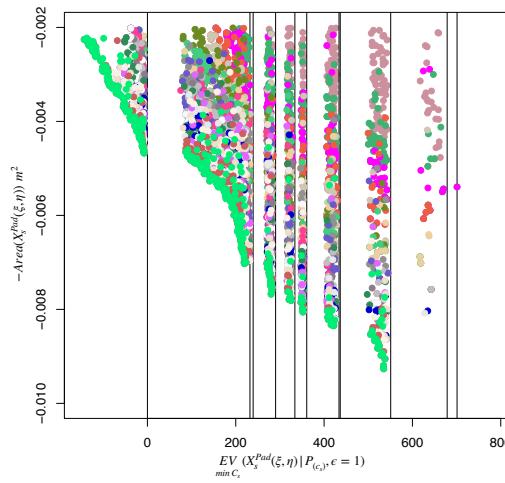


(c) Iteration 50:

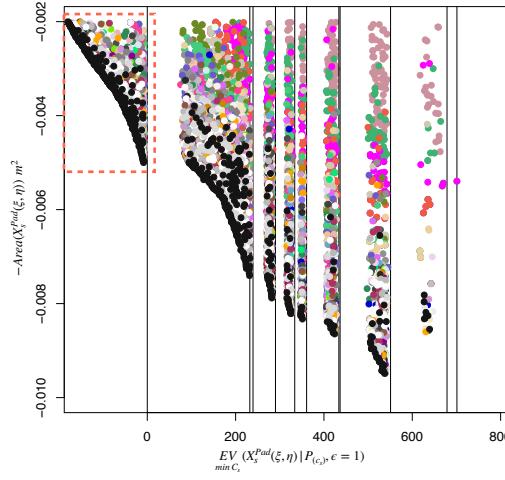
Figure 7.2: Optimization for infill points with EI criterion and $-Area(X_s^{Pad}(\xi, \eta))$. The colored points represent the population with each color representing a generation obtained through NSGA-2. The red highlight at the iteration 50 is to be compared with the iteration 50 from figure 7.3. 83



(a) Iteration 4:



(b) Iteration 22:



(c) Iteration 50:

Figure 7.3: 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 7.2.

When comparing Figure 7.2 and Figure 7.3, 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 .

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 (6.43) 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 ?? 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 7.4.

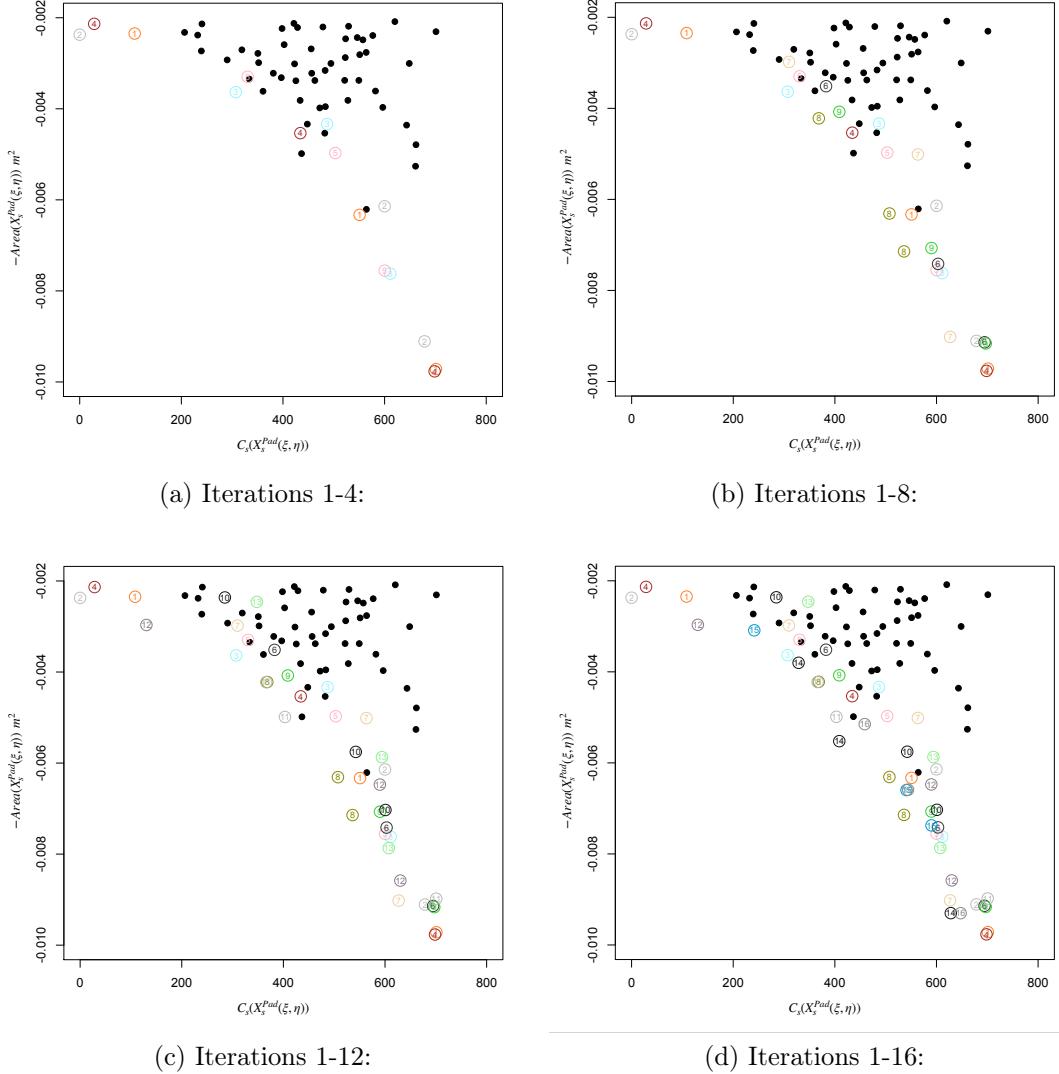


Figure 7.4: Objective space : $C_s(X_s^{Pad}(\xi, \eta)) \mid \Im(\Lambda(X_s^{Pad}(\xi, \eta))) \in [10KHz, 13KHz]$ and $-Area(X_s^{Pad}(\xi, \eta))$. Addition of infill points with progression of iteration for MOBO with *EV* criterion. Black points represent the initial samples. The colored circles 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 7.4) 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 $\text{Area}(X_s^{\text{Pad}}(\xi, \eta))$, which is due to very few designs in this region and hence, all the designs with larger $\text{Area}(X_s^{\text{Pad}}(\xi, \eta))$ 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 7.5, 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 . 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 uncertainty, as detailed in the section .

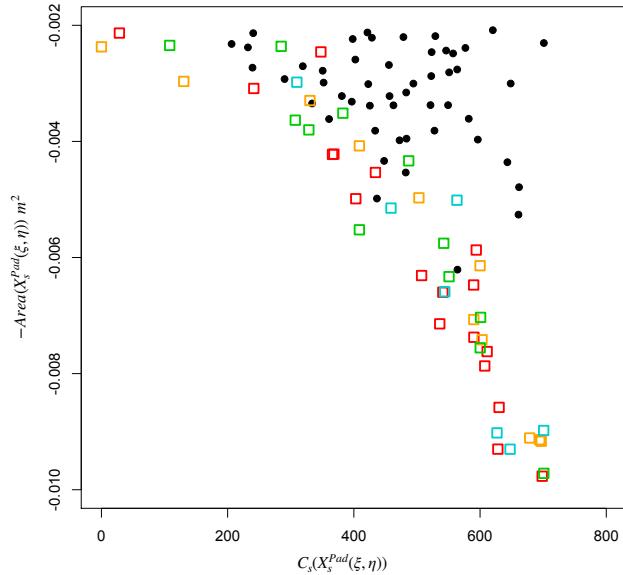


Figure 7.5: Objective space : $C_s(X_s^{\text{Pad}}(\xi, \eta)) \mid \Im(\Lambda(X_s^{\text{Pad}}(\xi, \eta))) \in [10\text{KHz}, 13\text{KHz}]$ vs $-\text{Area}(X_s^{\text{Pad}}(\xi, \eta))$. Clustering of the infill points, with each color representing a cluster and black points representing the initial samples

We expose some of the shapes from the objective space in Figure 7.6. It can be observed that, largely for a given $\text{Area}(X_s^{\text{Pad}}(\xi, \eta))$ to achieve a low value of $C_s(X_s^{\text{Pad}}(\xi, \eta))$, the pad shapes prefer to align more radially to 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 $\text{Area}(X_s^{\text{Pad}}(\xi, \eta))$ is larger can be attributed to lack of design space where this explains the lack of improvement related to $C_s(X_s^{\text{Pad}}(\xi, \eta))$ in this region. While, in the region of the objective space for smaller $\text{Area}(X_s^{\text{Pad}}(\xi, \eta))$, there is more flexibility in defining the pad designs $X_s^{\text{Pad}}(\xi, \eta)$ and hence, better improvement in solutions relative to $C_s(X_s^{\text{Pad}}(\xi, \eta))$ were obtained. We also obtained an interesting solution from this region with no instability as it can be observed, i.e., $C_s(X_s^{\text{Pad}}(\xi, \eta)) \mid \Im(\Lambda(X_s^{\text{Pad}}(\xi, \eta))) \in [10\text{KHz}, 13\text{KHz}] = 0$

It is interesting and important to note that the obtained optimal solutions should not have been obtained using classical shape optimization (ie. without isogeometric elements). 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.

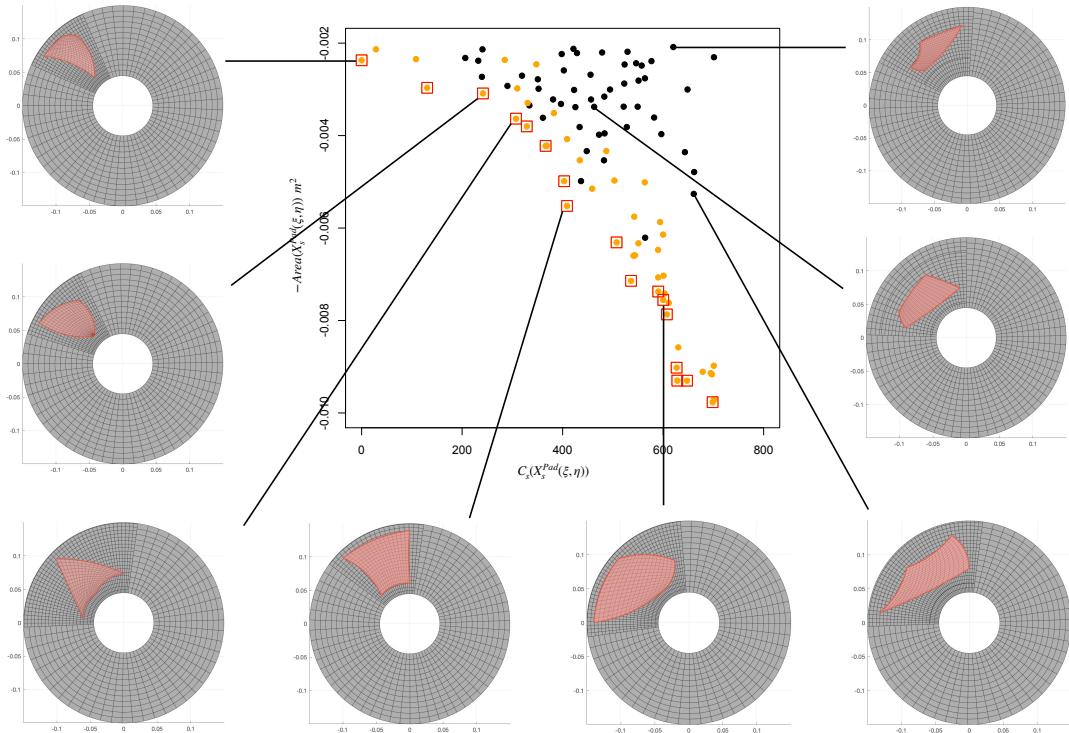


Figure 7.6: A sample of shapes from the objective space are shown. Black points represent the initial samples, Orange points represent the infill points and the NDS obtained after 16 iterations of MOBO are highlighted in red.

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 7.7, 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 7.8, which was considered with infill points from 15 iterations.

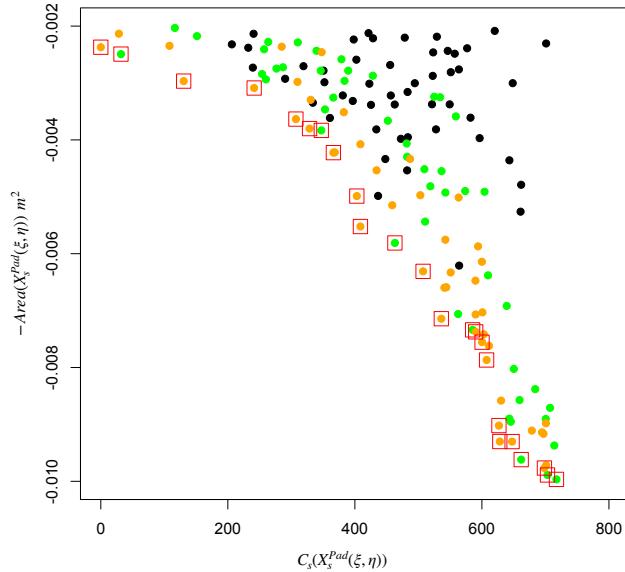


Figure 7.7: Objective space : $C_s(X_s^{Pad}(\xi, \eta)) \mid \Im(\Lambda(X_s^{Pad}(\xi, \eta))) \in [10KHz, 13KHz]$ vs $-Area(X_s^{Pad}(\xi, \eta))$. Comparison of infill points obtained between *EV* and *EI* criteria for an initial sample. Black points represent the initial samples. Orange and green 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 optimisation was performed with 5 different initial samples for 15 iterations, with 3 infill points per iteration. The hypervolume improvement comparison is shown in Figure 7.8 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 7.9.

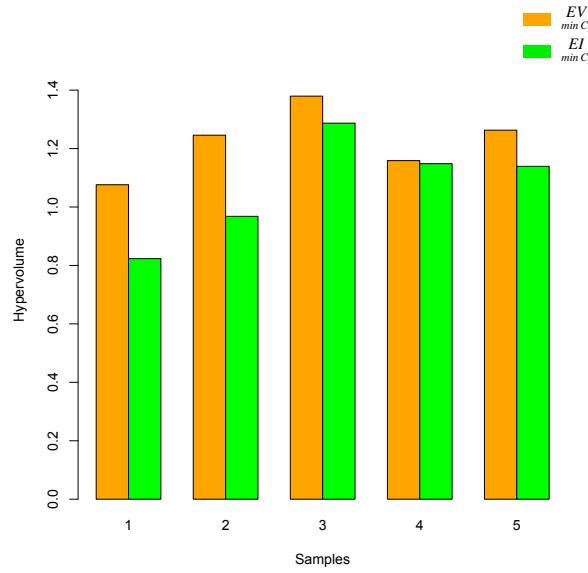


Figure 7.8: Comparison between EV and EI criteria for hypervolume improvement of NDS linked with the NDS of the initial samples, shown for a five set of initial samples.

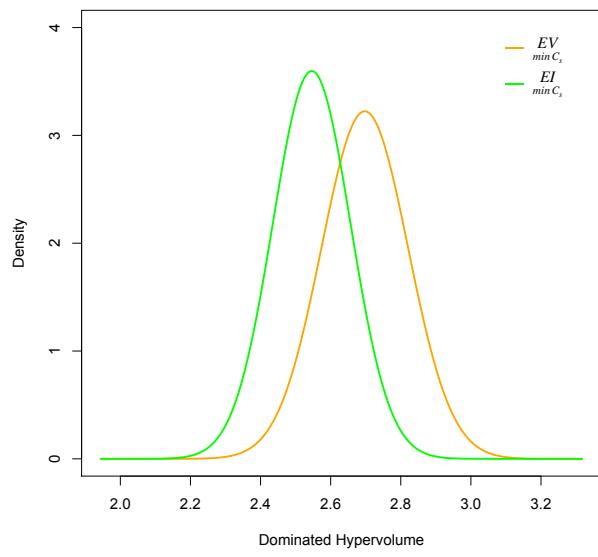


Figure 7.9: The variation of the dominated hypervolume for the NDS with initializing samples, obtained for the same set of initial samples in Figure 7.8.

Bibliography

- [1] Youn Doh Ha. Generalized isogeometric shape sensitivity analysis in curvilinear coordinate system and shape optimization of shell structures. Structural and Multidisciplinary Optimization, 52(6):1069–1088, Dec 2015.
- [2] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. Taking the human out of the loop: A review of bayesian optimization. Proceedings of the IEEE, 104(1):148–175, 2016.