HAL open science

Condition number and clustering-based efficiency improvement of reduced-order solvers for contact problems using Lagrange multipliers

Simon Le Berre, Isabelle Ramiere, Jules Fauque, David Ryckelynck

To cite this version:

Simon Le Berre, Isabelle Ramiere, Jules Fauque, David Ryckelynck. Condition number and clustering-based efficiency improvement of reduced-order solvers for contact problems using Lagrange multipliers. Mathematics , 2022, 10 (9), pp.1495. 10.3390/math10091495 . cea-03714373

HAL Id: cea-03714373

https://cea.hal.science/cea-03714373v1

Submitted on 5 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Condition number and clustering-based efficiency improvement of reduced-order solvers for contact problems using Lagrange multipliers.

Simon Le Berre [1,2], Isabelle Ramière [1,*], Jules Fauque [1] and David Ryckelynck [2]

June 20, 2022

[1] CEA, DES, IRESNE, DEC, SESC, LSC; Cadarache, F-13108 Saint-Paul-Lez-Durance {simon.leberre;isabelle.ramiere;jules.fauque}@cea.fr
[2] PSL Mines-ParisTech, Centre des matériaux, F-91003 Evry; {simon.leberre;david.ryckelynck}@mines-paristech.fr

* corresponding author

Abstract

This paper focuses on reduced-order modeling for contact mechanics problems treated by Lagrange multipliers. The high nonlinearity of the dual solutions lead to poor classical data compression. An hyper-reduction approach based on a reduced integration domain (RID) is considered. The dual reduced basis is then the restriction to the RID of the full-order dual basis, which ensures the hyper-reduced model to respect the non-linearity constraints. However the verification of the solvability condition, associated to the well-posedness of the solution, may induce an extension of the primal reduced basis without guaranteeing accurate dual forces. We highlight the strong link between the condition number of the projected contact rigidity matrix and the precision of the dual reduced solutions. Two efficient strategies of enrichment of the primal POD reduced basis are then introduced. However, for large parametric variation of the contact zone, the reachable dual precision may remain limited. A clustering strategy on the parametric space is then proposed in order to deal with piece-wise low-rank approximations. On each cluster, a local accurate hyper-reduced model is built thanks to the enrichment strategies. The overall solution is then deeply improved while preserving an interesting compression of both primal and dual bases.
Keywords: Hyper-reduction; contact mechanics; Lagrange multipliers; solvability condition; condition number; clustering; local reduced order models

1 Introduction

Variational inequalities (VI) arising from contact mechanics problem are still nowadays computationally expensive to solve. The size of the discrete full-order model (FOM) becomes rapidly prohibitive when accurate solutions are required. The issue of high computational cost has been widely studied for variational equality problems, and a common solution consists in using reduced order modeling (ROM) techniques. But in case of contact VI problems, the application of such techniques is not straightforward. Indeed the non-smoothness of the contact constraints, especially the non-negativity of the contact pressure, remains a challenge for the adaptation of ROM methods.

A number of numerical methods, such as the penalty method or the Augmented Lagrangian (AL) formulation, consist in building unconstrained minimization problems for solving contact. In this particular framework, as the main challenges related to ROM adaptation are circumvented, various reduced-order methods have been successfully applied, see [1, 2, 3] for the penalty method and [4, 5, 6] for the AL (or AL-like) formulation. However these kinds of contact formulations require numerical parameters adjustments to reach convergence. Concerning the penalty method, another disadvantage is that practically this method leads to some approximation of contact conditions (cf. interpenetration).

In counterpart, the Lagrange multipliers approach enable to satisfy precisely surface contact conditions. The resulting contact problem is a saddle-point problem with primal (displacement) and dual (contact pressure) variables. The Lagrange multipliers physically represent the contact pressures (opposite of the contact forces) and must remain non-negative. This constraint avoid the use of the famous Proper Orthogonal Decomposition (POD) snapshots method for the reduction of the dual basis. Hence, very few ROM methods have been applied to such problems. Their features mainly concern the construction of a reliable low-rank dual basis. We can distinguish approaches related to quasi-static contact problems to those related to dynamic contact problems. Concerning the reduction of (quasi-)static contact problems, we can cite the extension of the reduced bases

method [7], where the dual reduced basis is obtained by snapshots selection. This may lead to low compressed reduced basis for large parametric space. In [8], a projected based method has been proposed with a Non-Negative Matrix Factorization (NNMF) to obtain the reduced dual basis. The NNMF satisfies the positivity constraint but doesn't give a high compression rate [9]. In [10] the dual reduced basis is built through a Cone-Projected Greedy algorithm resulting in a better compression rate. However, this approach has not yet been verified outside the training set. Another strategy has been proposed in [9], lying on Hyper-Reduction techniques. Such kind of techniques [11, 12, 13, 14]have shown their efficacy to reduce high-dimensional reconstruction required by non-linearity evaluations. Hence, in [9] an Hyper-Reduction method based on a reduced integration domain (RID) [15] was adapted to contact problems. The low-rank dual basis is then simply defined as the restriction of the full-order dual basis to the RID, ensuring naturally the non-negativity constraint and leading to precise dual forces on the RID. Moreover, this method ensures the verification of the full contact conditions inside the RID.
Concerning dynamic contact problems, the problem of building a reduced dual basis respecting the non-negativity has been recently bypassed [16] through the use of the reduction of the primal variable to compute the Lagrange multipliers in an incremental way or [17] through the use of the Linear Complementarity Programming method where only the primal variable is reduced.

In this work we propose to improve the efficiency of the so-called Hybrid Hyper-Reduced (HHR) method, the hyper-reduced method based on the RID introduced in [9], especially for large parametric variation of the contact zone. The consistency of the HHR approach has been proven in [9], lying on the solvability condition. Since high primal basis compression by POD is performed, this condition may be not initially respected. It has been proposed in [9] to enrich the primal reduced basis with primal finite element (FE) shape functions while not respecting the solvability condition. However, the respect of the solvability condition is generally not sufficient to obtain accurate dual forces. No guideline was proposed in [9] for the choice of the FE shape functions to be added. Practically, for D-dimensional problems ($D > 1$) it was proposed to add all primal FE shapes function lying on the contact zone in the reduced primal basis. For high dimensional problems with large parametric variations, this strategy may compromise data compression. In order to automatically and efficiently extend the reduced primal basis, we first highlight in this contribution the strong relation between the projected contact rigidity matrix condition number and the precision of the dual solutions obtained by the HHR method. Taking this relation into account, we propose two primal basis enrichment strategies aiming at improving the condition number of the projected contact rigidity matrix while respecting the solvability condition. These enrichment strategies are both based on a greedy algorithm. There are very few elements in the literature studying the relation between condition number and precision of a ROM method. To the best of our knowledge, only [13] uses the condition number as a selection criteria for the sampling in the Missing Point Estimation (MPE) method.
Secondly, as the enrichment strategy may increase significantly the size of the primal reduced basis, we propose a clustering strategy derived from [18] to treat contact problems with large parametric spaces. The clustering enables us to build local ROMs and hence to reduce high nonlinear (and non-smooth) problems in a piece-wise low-rank manner. The interested reader can refer to [18, 19, 20, 21, 22] for examples of local ROMs applied to Hyper-Reduction schemes. Here, the local ROMs are obtained by applying on each cluster the HHR method for contact problems with condition number based enrichment introduced in this paper.
The proposed strategies result in highly accurate dual (and primal) solutions with reduced basis of limited sizes even for large variations of the contact zone.

The rest of the article is structured as follows. A brief presentation of the FOM for Lagrange multipliers-based contact problems is witten in Section 2. The HHR model for contact mechanics is also recalled (and few adapted) in this section. In Section 3 we present the correlation between condition number of the projected contact rigidity matrix and precision of dual solutions. We then propose the two primal reduced basis enrichment strategies. In Section 4 the clustering framework for contact mechanics is introduced. Finally in Section 5 numerical experiments are performed allowing to appreciate the efficiency of the proposed strategies.

2 Contact problem, FE and HHR models

2.1 Static unilateral contact probem

We are interested in the solution of elastostatic frictionless unilateral contact problems. The notations we consider are presented in Figure 1. Moreover, we classically denote by u_N the relative normal displacement on the potential contact surface Γ^C, d the initial gap and F_N the normal force component. Then, the Signorini's law for unilateral contact is given on Γ^C by: the non-penetration condition $u_N \leq d$, the non-adhesion condition $F_N \leq 0$ and the complementarity slackness $(u_N - d)F_N = 0$. Under the small strain hypotheses, the strong formulation of the elastostatic frictionless unilateral contact problem can be expressed as [23, 24, 25]:

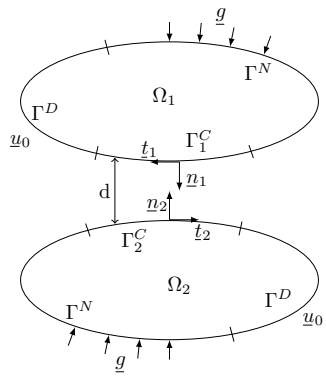


Figure 1: Contact model problem.

$$\begin{cases} -\operatorname{div} \underline{\underline{\sigma}} = \underline{f} & \text{in} \quad \Omega, \\ \underline{\underline{\sigma}} = \underline{\underline{\underline{\underline{C}}}} : \underline{\underline{\varepsilon}} & \text{in} \quad \Omega, \\ \underline{\underline{\varepsilon}} = \dfrac{1}{2}\left(\operatorname{grad} \underline{u} + (\operatorname{grad} \underline{u})^T\right) & \text{in} \quad \Omega, \\ \underline{u} = \underline{u}_0 & \text{on} \quad \Gamma^D, \\ \underline{\underline{\sigma}}\,\underline{n} = \underline{g} & \text{on} \quad \Gamma^N, \\ \underline{\underline{\sigma}}\,\underline{n} = F_N \underline{n} & \text{on} \quad \Gamma^C, \\ u_N \leq d,\ F_N \leq 0,\ (d - u_N)F_N = 0 & \text{on} \quad \Gamma^C, \end{cases} \tag{1}$$

with $\underline{u}$ the displacement field, $\underline{\underline{\sigma}}$ the stress field, $\underline{\underline{\varepsilon}}$ the strain field, $\underline{f}$ the prescribed volume density of forces, $\underline{u}_0$ the imposed displacement on Γ^D, $\underline{g}$ the imposed forces (surface density) on Γ^N and $\underline{n}$ the external unit normal. The tensor $\underline{\underline{\underline{\underline{C}}}}$ denotes the fourth-order elasticity tensor.

The weak formulation derived from problem (1) induces a variational inequality (VI) [23]. This VI, equivalent to a minimization problem, can be classically treated by Lagrange multipliers [24]. The resulting weak formulation is then a saddle-point problem, cf. Eq (2).

$$\begin{cases} \text{Find } (\underline{u}, \lambda) \in \mathcal{V} \times \mathcal{W} \text{ such that} \\ a(\underline{u}, \underline{v}) + b(\underline{v}, \lambda) = l(\underline{v}) & \forall \underline{v} \in \mathcal{V}_0, \\ \quad b(\underline{u}, q - \lambda) \leq \langle d, q - \lambda \rangle_{H^{\frac{1}{2}}(\Gamma^C), H^{-\frac{1}{2}}(\Gamma^C)} & \forall q \in \mathcal{W}, \end{cases} \tag{2}$$

with λ the Lagrange multiplier, $\mathcal{W} = \{q \in H^{-\frac{1}{2}}(\Gamma^C) \ \mid \ q \geq 0\}$, $H^{-\frac{1}{2}}$ the dual of $H^{\frac{1}{2}}$, $b(\underline{v}, q) = \langle v_N, q \rangle_{H^{\frac{1}{2}}(\Gamma^C), H^{-\frac{1}{2}}(\Gamma^C)}$ and $\langle \cdot, \cdot \rangle_{H^{\frac{1}{2}}, H^{-\frac{1}{2}}}$ the dual product,
$\mathcal{V} = \{\underline{v} = \underline{r}(\underline{u}_0) + \tilde{\underline{u}} \ \mid \ \tilde{\underline{u}} \in \mathcal{V}_0\}$, $\underline{r}(\underline{u}_0)$ the lifting of $\underline{u}_0$ over Ω,
$\mathcal{V}_0 = \{\underline{v} \in (H^1(\Omega))^D \ \mid \ \underline{v} = \underline{0} \quad \text{sur} \quad \Gamma^D\}$ and

$$a(\underline{u}, \underline{v}) = \int_\Omega \underline{\underline{\varepsilon}}(\underline{v}) : \underline{\underline{\underline{\underline{C}}}} : \underline{\underline{\varepsilon}}(\underline{u}) \, \mathrm{d}\Omega, \qquad l(\underline{v}) = \int_\Omega \underline{f}^T \underline{v} \, \mathrm{d}\Omega + \int_{\Gamma^N} \underline{g}^T \underline{v} \, \mathrm{d}\Gamma.$$

2.2 Finite element discretisation

The standard FE discretization [25] of saddle-point problem (2) results in system (3).

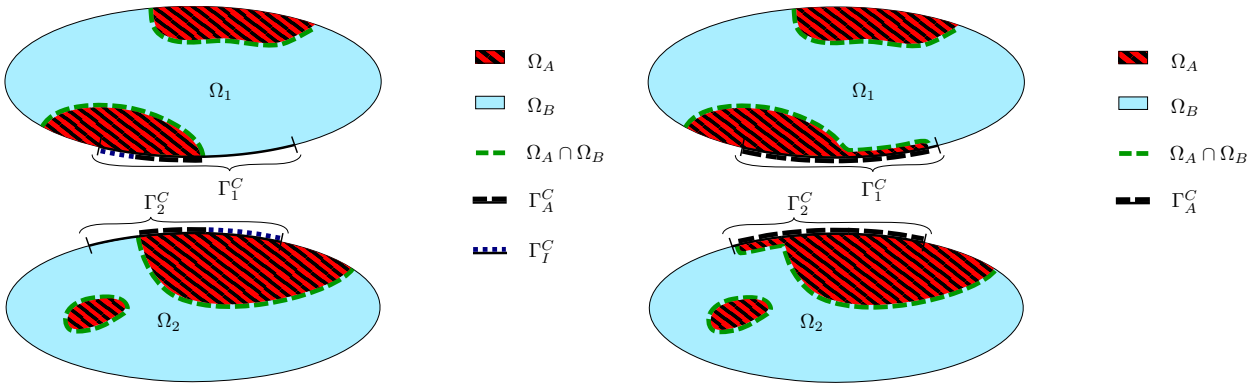
$$\begin{cases} \text{Find } (\underline{U}, \underline{\Lambda}) \in \mathbb{R}^N \times (\mathbb{R}^+)^{N_\lambda} \text{ such that} \\ \quad \underline{\underline{K}}\,\underline{U} + \underline{\underline{B}}^T \underline{\Lambda} = \underline{F}, \\ \qquad\qquad \underline{\underline{B}}\,\underline{U} \leq \underline{D}, \\ \underline{\Lambda}^T (\underline{D} - \underline{\underline{B}}\,\underline{U}) = 0, \end{cases} \tag{3}$$

with $\underline{U}$ being the discretization of the displacement solution in the primal FE basis, $\underline{\Lambda}$ the discretization of the Lagrange multipliers in the dual FE basis, N is the number of primal DoFs and N_λ the number of potential contacts. The matrices or vectors $\underline{\underline{K}} \in \mathbb{R}^{N \times N}$, $\underline{\underline{B}} \in \mathbb{R}^{N_\lambda \times N}$, $\underline{F} \in \mathbb{R}^N$, $\underline{D} \in \mathbb{R}^{N_\lambda}$ are respectively the rigidity matrix, the contact rigidity matrix, the source vector and the initial gap vector.

The discrete saddle-point problem (3) has to respect the classical Ladyzenskaia-Babuška-Brezzi (LBB) [26, 27] or inf-sup condition implying stability of the solution and hence solvability of the problem. For node-to-node contact with linear (or bilinear) finite elements approximation, which are of interest here, this condition is ensured [28].

2.3 Hybrid Hyper-Reduced model

2.3.1 Reduced integration domain for contact problems.



(a) Extended RID interface (Γ_I^C). (b) Extended RID (Ω_A).

Figure 2: Modification of the RID definition for the contact problems.

The RID [15], that will be the support of the hyper-reduced discretized problem, is built from a set of DoFs associated to nodes located of the whole domain Ω. The DoFs selection method usually consists of a sampling method such as the DEIM [12] or an arbitrary selection for areas of interest.
For contact problems, the RID definition must be adapted to ensure computable gaps on the RID. We focus here on node-to-node contact algorithm. In [9], an extension of the RID interface has been proposed, see Figure 2a. The unpaired DoFs on the potential contact surface of the RID were hence removed from the set of available balance equations. In this paper, we propose an extension of the RID to treat contact on each node of the potential contact zone included in the RID, see Figure 2b. Denoting by $\mathcal{L}$ the list of node indices used to build the RID, the complementary contact list $\mathcal{L}'$ reads

$$\mathcal{L}' = \{\text{node } j \text{ on } \Gamma^C; j \notin \mathcal{L} \ \mid \ \exists i \in \mathcal{L} \text{ such that } (i,j) \text{ is a pair of contact nodes}\}. \tag{4}$$

The RID is now defined as:

$$\Omega_A = \left(\bigcup_{i \in \mathcal{L}} \text{supp}(\underline{\varphi}_i) \right) \ \bigcup \left(\bigcup_{j \in \mathcal{L}'} \text{supp}(\underline{\varphi}_j) \right), \tag{5}$$

with $\underline{\varphi}_k$, $k = 1, ..., N$ the FE shape functions.
We denote Ω_B the counterpart of Ω_A such that $\Omega = \Omega_A \cup \Omega_B$. The complementary list $\mathcal{L}'$ plays the role of a zone of interest as defined in [15]. This construction variation allows to deal with more potential contacts, but slightly enlarges the size of the RID.

2.3.2 Reduced bases

Concerning the primal basis reduction, the standard POD-snapshot method [29] is used. The size l of this reduced basis (RB) is automatically obtained according to the approximation error $\epsilon_{\mathcal{POD}}$ on the snapshot matrix. The primal RB is then the matrix $\underline{\underline{V}} \in \mathbb{R}^{N \times l}$ containing the l first left singular vectors of the singular value decomposition (SVD) of the snapshots matrix. The discrete displacement is then projected in this RB $\underline{U} = \underline{\underline{V}}\underline{\gamma}$, $\underline{\gamma} \in \mathbb{R}^l$ being degrees of freedoms (DoFs) associated to the primal RB , $l \ll N$.

In order to avoid a non efficient compression of the dual basis, we use the RID concept to simply reduce the number of dual DoFs by restriction of the dual FOM basis to the RID. This strategy enable precise dual forces to be obtained on the RID, as shown in [9].
In counterpart, the reduced dual basis is of local support only. If the solution outside the RID is of interest, a reconstruction strategy based on the solution of a non-negative least squares (NNLS) problem using the dual snapshots as well as the ROM primal and dual solutions has been proposed in [9].

2.3.3 Hybrid Hyper-Reduced formulation

Complexity reduction of hyper-reduced models accounts for the sparsity of finite element matrices. We introduce the set $\mathcal{A}$ of primal DoFs not connected to Ω_B:

$$\mathcal{A} = \left\{ i \in [\![1,N]\!] \ \mid \ \int_{\Omega_B} \underline{\varphi}_i^T \underline{\varphi}_i d\Omega = 0 \right\}, \tag{6}$$

and the set $\mathcal{I}$ of DoFs connected to $\mathcal{A}$ by the rigidity matrix $\underline{\underline{K}}$ (interface nodes in the discretization scheme sense):

$$\mathcal{I} = \left\{ j \in [\![1,N]\!], j \notin \mathcal{A} \ \mid \ \exists i \in \mathcal{A} \quad \text{s.t.} \quad (\underline{\underline{K}})_{ij} \neq 0 \right\}. \tag{7}$$

Concerning the reduction of the contact conditions, we define the set $\mathcal{A}_c$ of DoFs of $\mathcal{A}$ localised on the potential contact zone and the set $\mathcal{A}_\lambda$ of dual DoFs connected to $\mathcal{A}_c$ by the contact matrix $\underline{\underline{B}}$.

$$\mathcal{A}_c = \{ j \in \mathcal{A} \ \mid \ \exists i \in [\![1,N_\lambda]\!] \quad \text{s.t.} \quad (\underline{\underline{B}})_{ij} \neq 0 \}, \tag{8}$$

$$\mathcal{A}_\lambda = \{ i \in [\![1,N_\lambda]\!] \ \mid \ \exists j \in \mathcal{A}_c \quad \text{s.t.} \quad (\underline{\underline{B}})_{ij} \neq 0 \}. \tag{9}$$

The cardinal of $\mathcal{A}_\lambda$, denoted N_λ^C in the sequel, corresponds to the number of potential contact points in the RID and hence to the size of the dual reduced basis. The Hybrid Hyper-Reduced (HHR) model, introduced in [9], reads:

$$\begin{cases} \text{Find } (\underline{\gamma}, \underline{\Lambda}) \in \mathbb{R}^l \times (\mathbb{R}^+)^{N_\lambda^C} \text{ such that} \\ \underline{\underline{V}}[\mathcal{A},:]^T \underline{\underline{K}}[\mathcal{A}, \mathcal{A} \cup \mathcal{I}] \underline{\underline{V}}[\mathcal{A} \cup \mathcal{I},:] \underline{\gamma} + \underline{\underline{V}}[\mathcal{A}_c,:]^T \underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c]^T \underline{\Lambda}[\mathcal{A}_\lambda] = \underline{\underline{V}}[\mathcal{A},:]^T \underline{F}[\mathcal{A}], \\ \underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c] \underline{\underline{V}}[\mathcal{A}_c,:] \underline{\gamma} \leq \underline{D}[\mathcal{A}_\lambda], \\ \underline{\Lambda}[\mathcal{A}_\lambda]^T (\underline{D}[\mathcal{A}_\lambda] - \underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c] \underline{\underline{V}}[\mathcal{A}_c,:] \underline{\gamma}) = 0, \end{cases} \tag{10}$$

with $\underline{\gamma}$ the primal solution and $\underline{\Lambda}[\mathcal{A}_\lambda]$ the dual solution of the reduced model. As shown on Eq. (10), the HHR formulation respects all the Signorini's conditions inside the RID.

As Problem (3), Problem (10) is a saddle-point problem. It has been proven in [9] that the HHR formulation is consistent with the FE formulation if the solvability condition of Problem (10) is respected. In the setting of (10) it leads to have $\underline{\underline{B}}[\mathcal{A}_{\lambda,\text{active}}, \mathcal{A}_c] \underline{\underline{V}}[\mathcal{A}_c,:]$ of full row rank, $\mathcal{A}_{\lambda,\text{active}}$ being the set of DoFs associated to an active contact (equality constraint), which is a priori unknown.

3 Solvability condition and enrichment of the primal reduced basis

3.1 Extended solvability condition

The set of active contact $\mathcal{A}_{\lambda,\text{active}}$ being a priori unknown, we have to consider a stronger solvability-like condition. We impose the projected contact rigidity matrix including all potential contacts, i.e. $\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c] \underline{\underline{V}}[\mathcal{A}_c,:]$, to be of full row rank. Since $\underline{\underline{B}}$ contains as many lines as there are potential contact points (node-to-node contact), the extended solvability condition on the reduced model (10) imposes

$$\text{rank}(\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c] \underline{\underline{V}}[\mathcal{A}_c,:]) = N_\lambda^C. \tag{11}$$

A necessary condition to verify Eq. (11) is that $N_\lambda^C \leq l$, which is not always respected due to the RID construction. In order to respect condition (11) without modifying the RID, the primal RB has to be enriched.

In [9] it was proposed to enrich the primal RB with FE shape functions related to the nodes on the contact surface. In the case of Lagrange FE, the extended primal RB denoted $\overline{\underline{\underline{V}}}$ then reads:

$$\overline{\underline{\underline{V}}} = \left(\underline{\underline{V}} \ \underline{\underline{I}}_N[\,:, \mathcal{L}_C] \right), \tag{12}$$

with $\underline{\underline{I}}_N$ the identity matrix of size N, $\mathcal{L}_C \subset \mathcal{A}_C$ the node indices of FE shape functions to be added. The primal reduced basis dimension is then $\bar{l} = l + card(\mathcal{L}_C)$. This enrichment is advantageous to accurately solve the solution on the contact surface, but the choice of $\mathcal{L}_C$ may be not trivial. In [9] the enrichment was either done manually with an arbitrary selection (for one-dimensional problems) or it was proposed to add all primal FE shape functions of the potential contact zone in order to simply ensure the full row rank condition, good stability, and accurate solutions on the contact zone. In this latter case $\mathcal{L}_C = \mathcal{A}_C$, which may lead to a significant expansion of the primal RB.

One of the main contribution of this paper is to propose strategies to automatically and wisely obtain the set $\mathcal{L}_C$ in order to optimize the ratio between the primal RB dimension $\bar{l}$ and the precision of the dual reduced solutions.

3.2 Importance of the projected contact rigidity matrix condition number

In the FE model, the condition number of the contact rigidity matrix $\underline{\underline{B}}$, written $\kappa(\underline{\underline{B}})$, is equal to 1. Accordingly to this observation, the more the condition number of the projected contact rigidity matrix of the HHR model, denoted $\kappa(\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c] \underline{\underline{V}}[\mathcal{A}_c,:])$, is close to 1, the more precise are the dual reduced solutions. On the other hand,

the primal reduced solutions are less sensitive to this condition number. An illustration of this behaviour of dual and primal HHR solutions is reported in Figure 3, from the test case of Section 5.2. The relative errors considered are presented in Section 5.1, Equations (17) (primal) and (18) (dual).

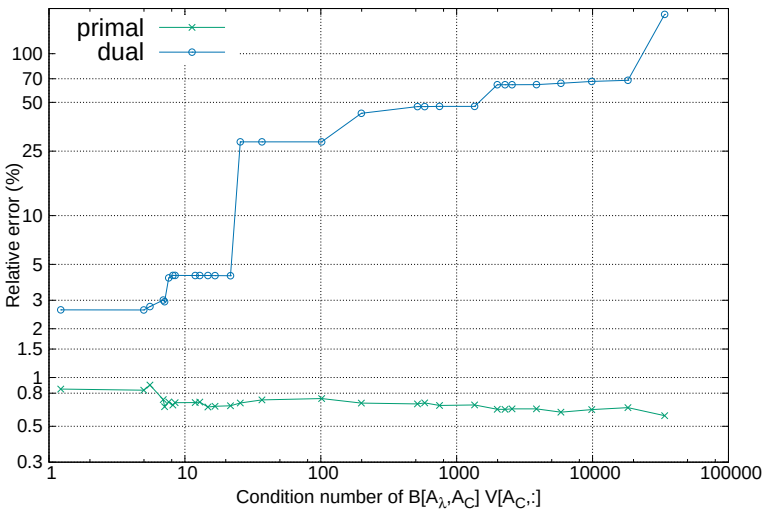


Figure 3: Impact of the condition number of the projected contact rigidity matrix on the dual and primal relative errors.

3.3 FE enrichment of the primal RB

In order to make the condition number $\kappa(\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c]\underline{\underline{V}}[\mathcal{A}_c, :])$ closer to $\kappa(\underline{\underline{B}})$, it seems natural to add to the primal RB $\underline{\underline{V}}$ some FE shape functions located on the potential contact zone of the RID, cf. Eq (12). When $\mathcal{L}_C = \mathcal{A}_C$, we then get $\kappa(\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c]\overline{\underline{\underline{V}}}[\mathcal{A}_c, :]) = \kappa(\underline{\underline{B}})$.
We propose to automatically obtain the set $\mathcal{L}_C \subset \mathcal{A}_C$ through a greedy algorithm based on $\kappa(\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c]\overline{\underline{\underline{V}}}[\mathcal{A}_c, :])$, see Algorithm 1. In this algorithm, $\mathcal{A}_C(p)$ denotes the DoF number associated to the pth element of $\mathcal{A}_C$.

Algorithm 1: Greedy algorithm based on the condition number of the projected contact rigidity matrix to enrich the primal RB.

Data: maximum condition number κ_{max}, matrices $\underline{\underline{B}}$ and $\underline{\underline{V}}$, the set $\mathcal{A}_C$ (primal FE DoFs on the potential contact zone of the RID)
Result: The set of FE DoFs to activate $\mathcal{L}_C \subset \mathcal{A}_C$ and the primal enriched RB $\overline{\underline{\underline{V}}}$

```
L_C = ∅ ; k = 0 ; G = [[1,card(A_C)]] ;
κ_min = κ(B[A_λ, A_c]V[A_c, :]) ;
V̄ = V;
while (κ_min > κ_max or rank(B[A_λ, A_c]V̄[A_c, :]) < N_λ^C) and k < card(A_C) do
    k = k + 1 ;
    for p ∈ G do
        V̄_tmp = (V̄ I_N[:, A_C(p)]) ;
        if κ(B[A_λ, A_c]V̄_tmp[A_c, :]) < κ_min then
            κ_min = κ(B[A_λ, A_c]V̄_tmp[A_c, :]) ;
            p_min = p ;
        end
    end
    G = G \ {p_min} ;
    L_C = {L_C, A_C(p_min)} ;
    V̄ = (V̄ I_N[:, A_C(p_min)]) ;
end
```

It is worth noting that even when the extended solvability condition is respected, the set $\mathcal{L}_C$ could continue to be enlarged in order to improve the dual solutions.

3.4 POD modes and FE shape functions enrichment of the primal RB

The FE shape functions enrichment is a way to extend the primal RB, but it naturally comes to mind to use an enrichment strategy that uses more POD modes (less truncated SVD) for a given RID. In this way, more

global information (from snapshots) can be included in the extended primal RB.
Practically, this strategy is based on a second POD threshold $\epsilon_{\mathcal{POD}2} < \epsilon_{\mathcal{POD}}$ to obtain l_2 supplementary primal POD modes. The primal POD basis considered in the HHR model (10) is now $\underline{\underline{V}} \in \mathbb{R}^{N\times(l+l_2)}$.

The extension of the primal POD RB through the order of singular vectors does not guarantee neither the extended solvability condition nor a satisfying condition number of the projected contact rigidity matrix. This strategy has hence to be combined with FE shape functions enrichment (see Section 3.3 and Algorithm 1) to obtain accurate dual solutions.

The main advantage of this strategy is that it allows us to use a more precise POD RB without impacting the size of the RID. Indeed, in most cases, the RID construction is based on a selection method relying on the primal RB (running the DEIM algorithm on the primal RB for example). Hence, there is a strong connection between the size of the primal RB and the size of the RID. Using the proposed strategy, the RID construction is still based on the POD threshold $\epsilon_{\mathcal{POD}}$ while the HHR model is based on the extended POD threshold $\epsilon_{\mathcal{POD}2}$. Let us underline that $\epsilon_{\mathcal{POD}2}$ must remain large enough in order to avoid overfitting. Hence this strategy is mainly of interest for large parametric spaces.

Remark For both enrichment strategies, one has to verify the basis property of the extended matrix $\overline{\underline{\underline{V}}}$ on the RID. The verification has to be done in the offline part of the method, without impacting the online part.

4 Clustering

In case of highly nonlinear problems such as non-smooth contact problems, clustering strategies may offer the possibly to treat the nonlinearity in a linear piece-wise manner through low-rank (ROM) approximations on different clusters partitioning the parametric space [18]. For large parametric spaces, we then expect to improve the accuracy of the reduced solution for a given dimension of RBs.

4.1 From regular snapshots

The framework of the clustering strategy is presented in Figure 4. The training set $\mathcal{P}_{tr}$ is generated through a regular discretization of the training space $\mathcal{P}$. FOM primal snapshots are calculated on the parametric points of $\mathcal{P}_{tr}$.
We then consider a partition of $\mathcal{P}$ composed of N_s elements relying on the discretization $\mathcal{P}_{tr}$. On each element $j = 1, ..., N_s$, an orthonormal basis $\underline{\underline{V}}_j$ is obtained from POD on the element-wise snapshots matrix. This local snapshots matrix is composed by snapshots associated to points in $\mathcal{P}_{tr}$ belonging to element j.

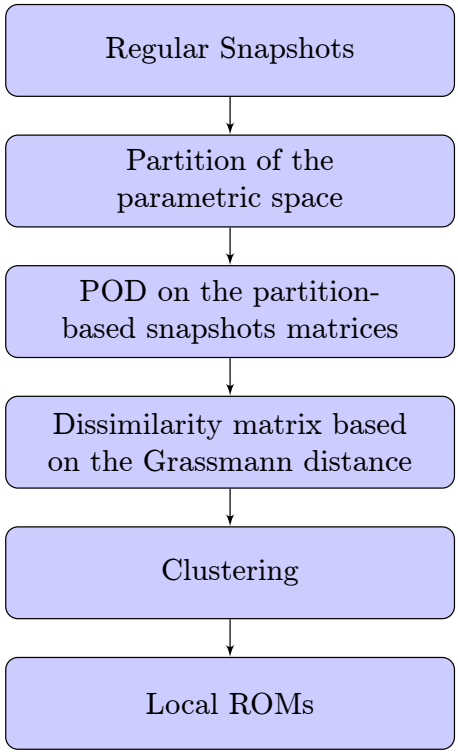


Figure 4: Clustering framework

As suggested in [18, 30], the Grassmann distance $d_{Gr(\infty,\infty)}$ between local RBs is then used for the clustering the simulated data for model order reduction. The dissimilarity matrix $\underline{\underline{\delta}} \in \mathbb{R}^{M_s\times M_s}$ writes:

$$\forall j_1, j_2 \in [\![0, N_s]\!]^2, \qquad (\underline{\underline{\delta}})_{j_1,j_2} = d_{Gr(\infty,\infty)}(\underline{\underline{V}}_{j_1}, \underline{\underline{V}}_{j_2}), \tag{13}$$

with

$$d_{Gr(\infty,\infty)}(\underline{\underline{A}}_1, \underline{\underline{A}}_2) = \left(\frac{\pi^2}{4}|a_1 - a_2| + \sum_{i=1}^{min(a_1,a_2)} \alpha_i^2\right)^{\frac{1}{2}}, \tag{14}$$

where a_1 and a_2 are the dimension of bases $\underline{\underline{A}}_1$ and $\underline{\underline{A}}_2$ respectively. The angles α_i are related to the singular values $\Sigma_{ii}, 1 < i \leq min(a_1, a_2)$, of $\underline{\underline{A}}_1^T \underline{\underline{A}}_2$ by Eq. (15)

$$\Sigma_{ii} = cos(\alpha_i). \tag{15}$$

The smaller $(\underline{\underline{\delta}})_{j_1,j_2}$ the closer the subspace spanned by bases $\underline{\underline{V}}_{j_1}$ and $\underline{\underline{V}}_{j_2}$ separately. A clustering algorithm based on the dissimilarity matrix $\underline{\underline{\delta}}$ is finally applied to build K clusters from the initial partition of $\mathcal{P}$ into N_s elements. In this work we use the k-medoids algorithm [31], to have a focus on compactness.

On each cluster k, $k = 1, ..., K$ the local snapshots matrix gathers the snapshots associated to points of the training set $\mathcal{P}_{tr}$ included into the cluster. From this local snapshots matrix, a local HHR model is built following the strategy described in Section 2.3. Especially, a local RID is considered in each cluster k.
As the clustering aims to reduce the range of the parameter variations, an accurate low-rank approximation of the primal basis can be obtained. Hence, in this case only the enrichment strategy based on FE shape functions

(see Section 3.3) has to be considered in order to ensure both the respect of the local solvability condition as well as accurate reduced dual forces.

Concerning the classification, a trivial strategy based on the clustering partition of the parametric space $\mathcal{P}$ is used. This classification is used online for the local ROM selection associated to given input data (generally outside the training set $\mathcal{P}_{tr}$).

4.2 From irregular snapshots

The regular-based clustering strategy introduced in Section 4.1 may become prohibitive for high-dimensional parametric spaces. In such cases, an irregular distribution of snapshots in the parametric space has to be considered for the training set $\mathcal{P}_{tr}$, generally thanks to greedy algorithms. Then, it is no more suitable to rely on an element-wise partition of the parametric space for the clustering. A simple adaptation of the procedure of Section 4.1 is proposed here. The dissimilarity matrix is now based on the Grassmann distance between each normalized snapshot $\underline{S}_j$, $j = 1, ..., M_s$, with $M_s = card(\mathcal{P}_{tr})$ the number of snapshots. In this case, Eq. (14) reduces to

$$d_{Gr(\infty,\infty)}(\underline{S}_{j_1}, \underline{S}_{j_2}) = \arccos(\underline{S}_{j_1}^T \underline{S}_{j_2}), \qquad \forall j_1, j_2 \in [\![0, M_s]\!]^2. \tag{16}$$

The clustering strategy returns in this case in labelling the snapshots.
Local ROMs are built similarly to Section 4.1 from the cluster-based local snapshots matrix.
The simplification of the clustering phase comes with adding complexity in the classification construction required for the online local ROM selection. Advanced classifiers, see for example [18], have to be considered as no clustering partition of the parametric space is available. In this work, we propose to build offline a supervised learning classification method trained on the snapshots and their cluster label. In the numerical examples, the Random Forest classification method [32, 33] has been performed.

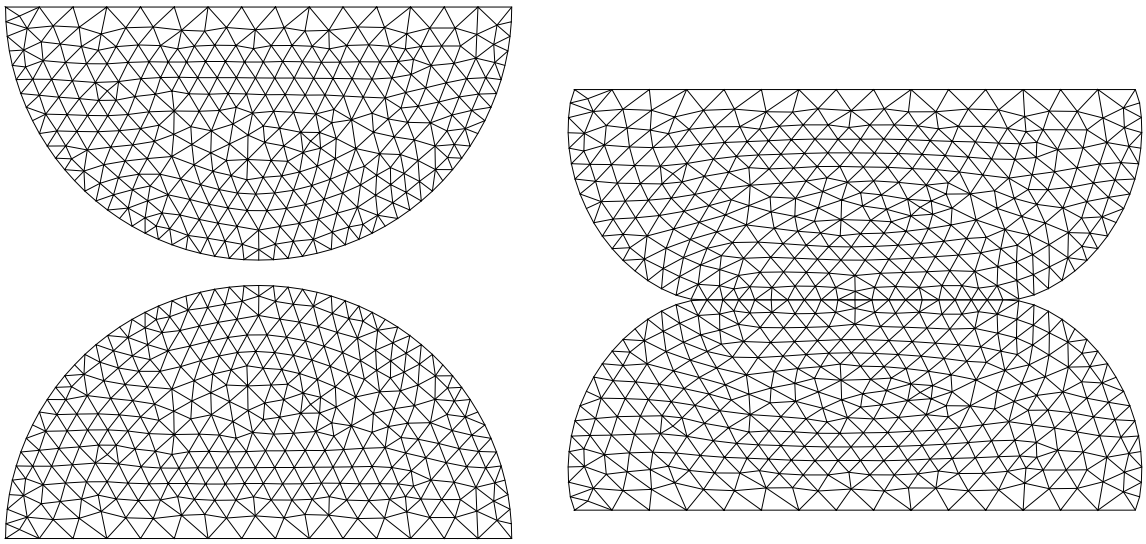
5 Numerical results

In all the following numerical experiments, the FOM solution is obtained thanks to the Cast3M finite elements software (2020 version) [34]. Lagrange P1 or Q1 finite elements have been used, with 3 or 4 integration points respectively for strain/stress calculation.

5.1 Test case a: Half disks of Hertz, unidimensional parametric space

This test case is described in [10]. It considers two half disks in a two-dimensional setting, each disk having the same radii $R = 1$ m. The two disks are separated by an initial gap being at minimum $d = 0.1$ m. Both disks have the same material characteristics: a Young modulus $E = 15$ Pa and a Poisson coefficient $\nu = 0.35$. Dirichlet boundary conditions are imposed on the horizontal edges of the disks, with $u_x = 0$ and $u_y = \pm\mu$ with an upward movement for the lower disk and a downward movement for the upper disk. The range values of μ defines the parametric space. A representation of the initial state of the test case can be seen in Figure 5a.

The Finite Element full-order model (FOM) is built on a mesh with 678 nodes. The circular part of both disks, constituting the potential contact zone, contains 55 nodes on each side. The deformation obtained with the FOM for $\mu = 0.3$ is presented in Figure 5b.



(a) Initial state.

(b) Deformation for $\mu = 0.3$.

Figure 5: Half-disks test case. FOM configurations

We consider a parametric space $\mathcal{P} = [0.15, 0.45]$ (m) for μ and build the training set $\mathcal{P}_{tr} = \{0.15 + 0.01i \big| 0 \leq i \leq 30\}$ (m). Hence card($\mathcal{P}_{tr}$) = 31. To assess the accuracy of the methods on this test case we will use three kind of errors, the primal relative error in 2-norm Eq. (17), the dual relative error in 2-norm Eq. (18) and the error on the minimum energy used in [10] Eq. (19)

$$e_u(\mu) = \left(\frac{||\underline{u}_{ROM}(\mu) - \underline{u}_{FOM}(\mu)||^2_{2,\Omega}}{||\underline{u}_{FOM}(\mu)||^2_{2,\Omega}} \right)^{\frac{1}{2}}, \tag{17}$$

$$e_\lambda(\mu) = \left(\frac{||\underline{\lambda}_{ROM}(\mu) - \underline{\lambda}_{FOM}(\mu)||^2_{2,\Gamma^C_A}}{||\underline{\lambda}_{FOM}(\mu)||^2_{2,\Gamma^C_A}} \right)^{\frac{1}{2}}, \tag{18}$$

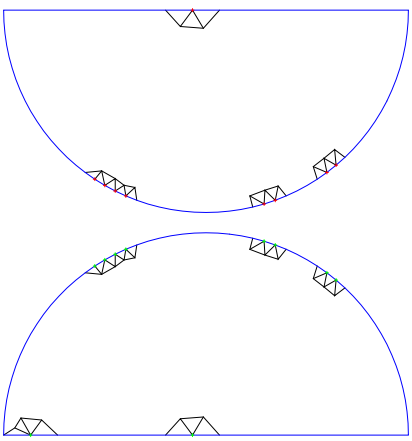
$$e_{ener}(\mu) = \frac{1}{2} \left| a(\underline{u}_{ROM}(\mu), \underline{u}_{ROM}(\mu)) - f(\underline{u}_{ROM}(\mu)) - a(\underline{u}_{FOM}(\mu), \underline{u}_{FOM}(\mu)) + f(\underline{u}_{FOM}(\mu)) \right|, \tag{19}$$

with $(\underline{u}_{ROM}(\mu), \underline{\lambda}_{ROM}(\mu))$ and $(\underline{u}_{FOM}(\mu), \underline{\lambda}_{FOM}(\mu))$ denoting the ROM (here obtained with the HHR model) and FOM solutions (primal and dual) respectively.

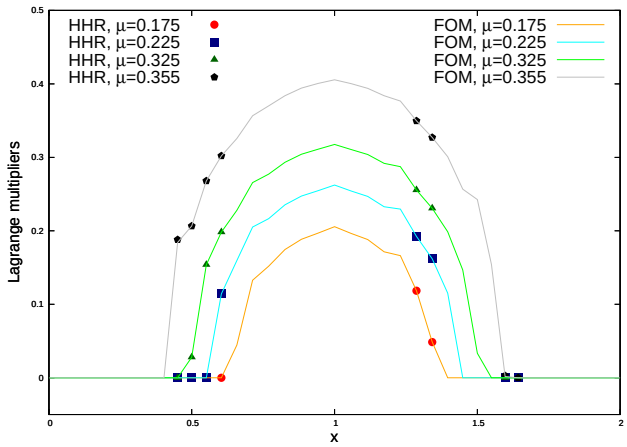
Let us underline that in the case of the HHR method, the error on the Lagrange multipliers (18) is naturally evaluated on the RID. The extension from Γ^C_A to Γ^C could however being performed through the reconstruction strategy introduced in [9]. On the counterpart, the extension to Ω of the displacement field is obvious thanks to the POD basis. Hence, the primal error will always be evaluated on the whole domain Ω.

This test case has the particularity to present a double symmetry (horizontal and vertical) for the dual and the primal solutions. This symmetry added to the relatively small contact surface and the one-dimensional parametric space leads to a test case relatively easy to treat for ROMs. Using the HHR, a small RID is enough to give very good results.

In the following results, the RID has been built from the set $\mathcal{L}$ obtained applying the DEIM to the primal POD reduced basis. This set is supplemented by two nodes on the top and bottom surfaces to be able to treat Dirichlet boundary conditions. As the parametric space is quite small, an accurate error threshold of $\epsilon_{\mathcal{POD}} = 10^{-8}$ is used resulting in a primal RB with 9 vectors ($l = 9$). The RID is reported in Figure 6a. It results in 16 nodes on the contact surfaces (8 on each side), hence $N_\lambda^C = 8$. The solvability condition (cf. Eq. (11) is in this case directly respected. Moreover, the condition number is also initially very good $\kappa(\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c]\underline{\underline{V}}[\mathcal{A}_c, :]) = 3.4$. Hence, this test case does not require any enrichment of the primal RB to obtain accurate results with the HHR method. In Figure 6b, the Lagrange multipliers for the FE and HHR models are reported for different values of $\mu \notin \mathcal{P}_{tr}$. This Figure confirms the very good local accuracy on the Lagrange multipliers.



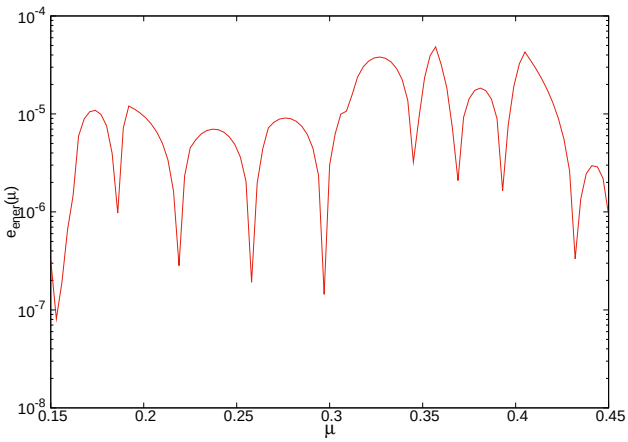
(a) Reduced integration domain (RID).



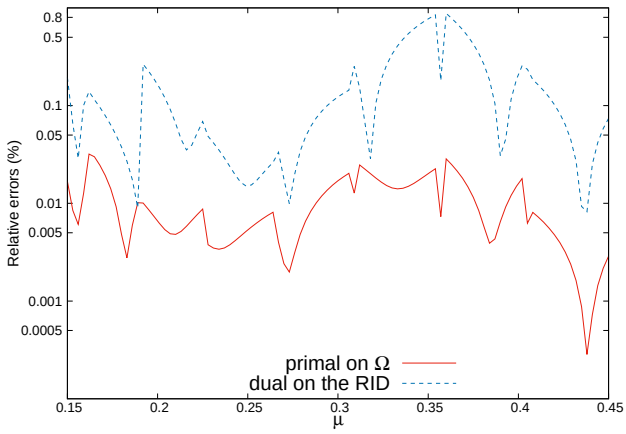
(b) Lagrange multipliers for different values of $\mu \notin \mathcal{P}_{tr}$.

Figure 6: Half-disks test case. HHR model : RID and Lagrange multipliers. $l = 9$, $N_\lambda^C = 8$.

To further evaluate the accuracy the method we compute the errors $e_u(\mu)$, $e_\lambda(\mu)$ and $e_{ener}(\mu)$, either on the RID and on Ω (with the reconstructed dual solution) on a set of 101 points in $\mathcal{P}$ (including the 11 points of the training set $\mathcal{P}_{tr}$). The results are reported in Figure 7.



(a) Error on the minimum energy $e_{ener}(\mu)$.



(b) Primal error on the displacement field $e_u(\mu)$ on Ω and dual error of the Lagrange multipliers $e_\lambda(\mu)$ on the RID.

Figure 7: Half-disks test case. HHR errors evaluations of 101 points. $l = 9$, $N_\lambda^C = 8$

The primal solution is always approximated with accuracy, with relative errors in 2-norm below 0.05% (5.10^{-4}) on all the parametric space. The dual solution is also approximated accurately ranging from 0.1% to a maximum of 0.8%. The error on the minimum energy is very also satisfying, with errors below 10^{-4} on Ω, even outside the training set.

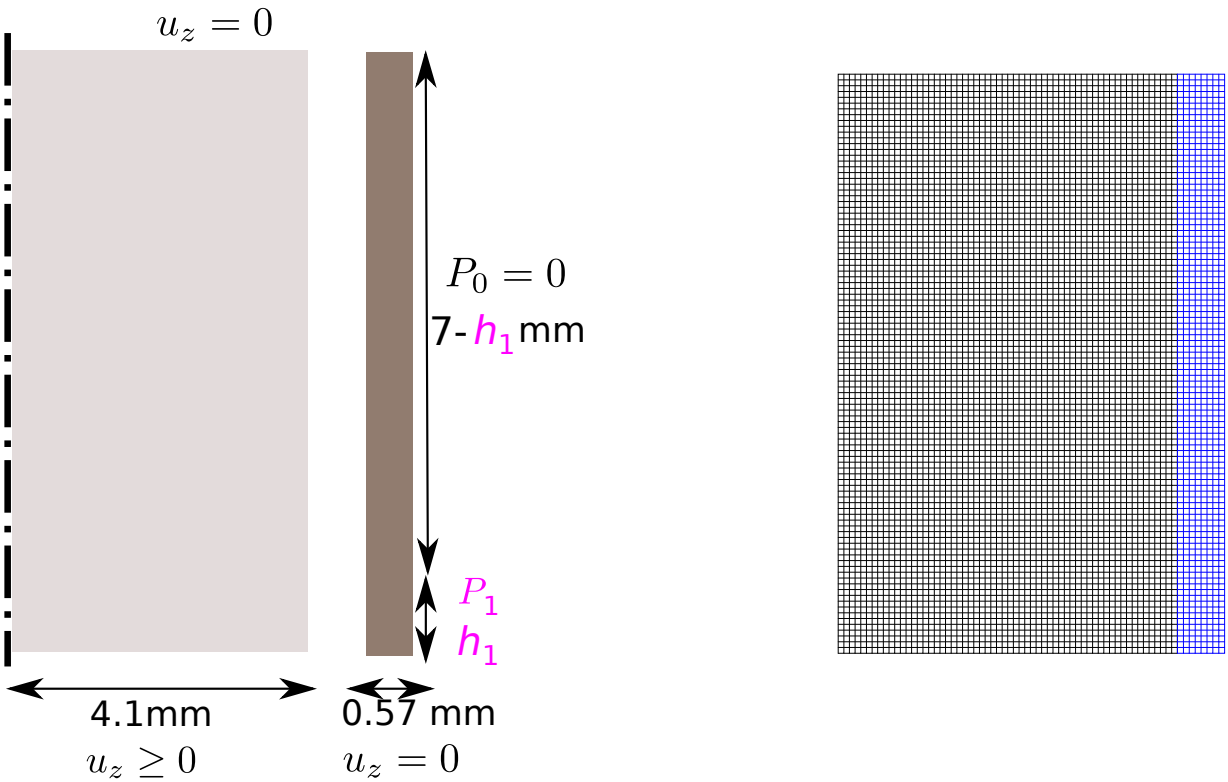
In [10], the CPG (Cone-Projected Greedy) algorithm has been applied on the same problem with the same training set and the same POD truncation ($l = 9$ POD modes). The CPG method yields to errors ranging from 10^{-6} to 10^{-3} for the error on the minimum energy, and from 10^{-5} to 4.10^{-3} for the error on the displacement field. Those errors were obtained on parametric points of the training set only. No errors on Lagrange multipliers

are reported. Compared with the errors shown in Figure 7 of the present paper, it highlights the robustness of the proposed HHR method, that is to say the accuracy of the results on all the parametric space even outside the training set.

5.2 Test case b: 2D axisymmetric contact - bidimensional parametric space

We consider here a two-dimensional axisymmetric test case introduced in [35], representing a simplified contact problem between a nuclear fuel pellet and its surrounding cladding. The test's geometry is sketched in Figure 8a. The two materials follow an elastic isotropic behaviour with a Young modulus and Poisson coefficient of $E = 190$ GPa, $\nu = 0.3$ for the fuel, and $E = 78$ GPa, $\nu = 0.34$ for the cladding. The uniform initial gap between the two solids is $d = 2$ μm. The test case consists in applying a pressure P_1 on a length h_1 on the outside part of the cladding. The range values of parameters P_1 and h_1 form the parametric space $\mathcal{P} = [5, 200]$ MPa$\times[0.6, 6.0]$ mm. This configuration leads to a low displacement along the z axis, allowing a good modeling of the contact with a node-to-node configuration. The parametric variations are challenging for ROM models as it includes strong variations of the contact surface as well as the contact pressure. The FOM is built on a mesh with 7070 nodes, 6161 for the mesh representing the fuel and 909 for the cladding, see Fig. 8b. There are 101 potential contacts ($N_\lambda = 101$).

In order to build the HHR model, 500 snapshots regularly distributed in the parametric space are used for the training set $\mathcal{P}_{tr}$. Due to the large variations of the parametric space, a POD threshold of $\epsilon_{\mathcal{POD}} = 10^{-6}$ is used to initially limit the dimension of the primal RB to $l = 36$ (see the decrease of approximation error on the snapshot matrix in Figure 16). Using the set of points $\mathcal{L}$ obtained from the DEIM applied to the primal RB results in the RID illustrated in Figure 9a where a region of interest corresponding to the Neumann surface has been added. The number of potential contact is the RID is $N_\lambda^C = 12$. This RID gives ride to unsatisfactory primal solutions. We then expand the set $\mathcal{L}$ using a second DEIM on a dual POD RB (with a threshold $\epsilon = 4 \times 10^{-4}$). It worth underlying that this dual POD RB is only used for the RID expansion. The expanded RID is presented in Figure 9b, having also the Neumann surface as zone of interest. In this case $N_\lambda^C = 34$. The advantage of using the DEIM on a dual POD RB is that the selected DoFs will be located on the contact surface.



(a) Geometry of the 2D axisymmetric test case.

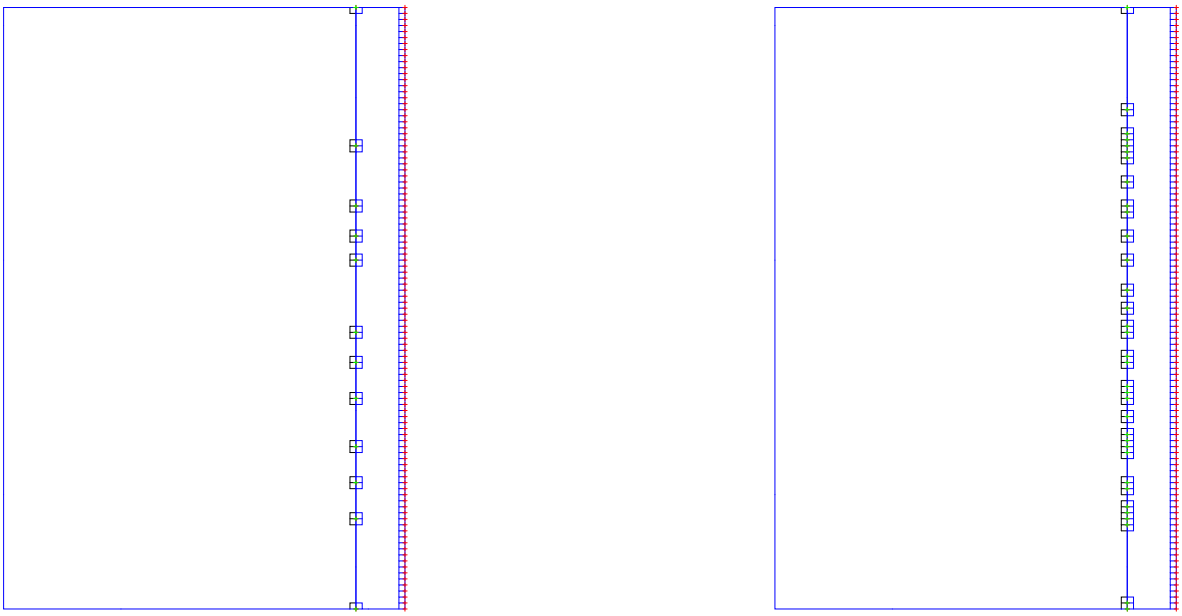
(b) FOM mesh

Figure 8: 2D axisymmetric test case. Sketch of the problem and FOM discretization.

In the following the results presented focus on the expanded RID ($l = 36$, $N_\lambda^C = 34$) and on parameters outside of the training set $\mathcal{P}_{tr}$. In this case, the solvability condition is not initially respected. Hence, the primal RB have to be enriched to solve the HHR problem.

5.2.1 Enrichment strategies

We focus here on the algorithm presented in Section 3.3 for the enrichment of the primal RB with FE shape functions. Figure 10 show the output at the different iterations of the greedy algorithm 1. Although the

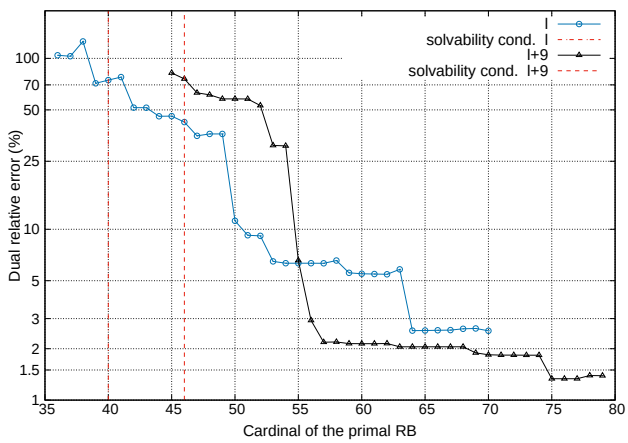


(a) RID obtained with one DEIM. ($l = 36$, $N_\lambda^C = 12$)

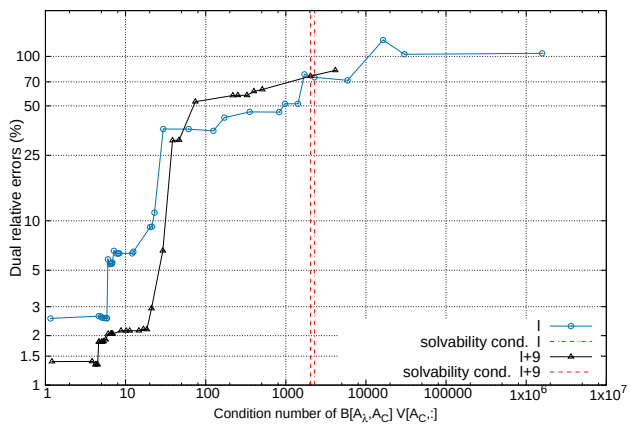
(b) Rid obtained with two DEIM. ($l = 36$, $N_\lambda^C = 34$)

Figure 9: 2D axisymmetric test case. Reduced integration domain (RID)

solvability condition isn't verified initially, a numerical solution can be obtained. However this latter exhibits high dual relative errors. Figure 10 illustrates that the strict respect of the solvability condition (indicated by the dashed red vertical lines) isn't enough to obtain a satisfying accuracy on the dual solutions. With FE shape functions enrichment, the solvability condition is verified for $\bar{l} = 40$ (4 supplementary FE shape functions), but with a condition number of $\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c]\underline{\underline{V}}[\mathcal{A}_c, :]$ superior to 1000, lying to dual relative errors greater than 70%. The further enrichment of the primal RB based on the improvement of this condition number enables to drastically improve the dual relative errors, up to 2% for the parametric point (170 MPa, 3.2mm) considered in Figure 10a. The dual relative error becomes significantly lower for a condition number around 10, see Figure 10b. For $\kappa_{max} = 10$, the greedy algorithm results in card($\underline{\underline{V}}$)=55 and a dual relative error of 6.3%.



(a) Dual relative error regarding the number of FE modes added to the primal POD RB enriched ('l+9') or not ('l').



(b) Relative dual errors regarding the condition number of the projected contact rigidity matrix for the two enrichment strategies.

Figure 10: 2D axisymmetric test case. Study of the two enrichment strategies: by FE modes only ('l') and by POD+FE ('l+9', 9 POD modes added), with $l = 36$ and $N_\lambda^C = 34$. Point: $P_1 = 170$MPa and $h_1 = 3.2$mm

The enrichment strategy of the primal RB with POD and FE modes, introduced in Section 3.4, is now compared to the enrichment strategy based on FE modes only, and previously illustrated. We set $\kappa_{max} = 1$ (maximal enrichment) in Algorithm 1 for both approach. We report in Table 1 the results obtained for different truncature levels for the enrichment by POD modes on two points of the parametric space (strong and weak contact) outside the training set $\mathcal{P}_{tr}$.

The enrichment method by POD+FE modes can enable us to reach dual and primal relative errors lower

Table 1: 2D axisymmetric test case: study of the enrichment strategy by POD modes and FE shape functions ($l = 36$ et $N_\lambda^C = 34$)

					$P_1 = 170$MPa $h_1 = 3.2$mm		$P_1 = 30$MPa $h_1 = 5.25$mm	
					Rel. err.		Rel. err.	
	nb POD mod. add.	$\epsilon_{\mathcal{POD}2}$ $\times 10^{-7}$	$\kappa(\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c]\underline{\underline{V}}[\mathcal{A}_c, :])$ after POD enrich.	$\bar{l} = card(\underline{\underline{V}})$ final	primal	dual	primal	dual
FE Enr.	0	–	1.5 10^6	70	1.0%	2.5%	2.9%	21%
POD+FE Enr.	3	7.5	2.3 10^4	73	2.5%	1.8%	0.5%	26%
POD+FE Enr.	6	5	4.5 10^3	76	0.5%	1.6%	1.6%	16%
POD+FE Enr.	9	3.5	4.1 10^3	79	0.1%	1.3%	3.5%	5.6%
POD+FE Enr.	12	2.7	3.4 10^3	82	1.4%	1.7%	2.5%	6.7%
POD+FE Enr.	15	2.2	878	85	1.9%	1.7%	1.3%	5.8%
POD+FE Enr.	18	1.7	386	88	1.1%	1.4%	0.2%	5.3%
POD+FE Enr.	21	1.3	383	91	7.8%	1.8%	13.4%	9.3%
POD+FE Enr.	23	1.1	330	93	0.8%	1.3%	16%	9.9%
POD+FE Enr.	26	0.9	328	96	12%	1.3%	1.5%	9.5%

than the enrichment method with only FE modes. Using a maximal enrichment $\kappa_{max} = 1$, the enrichment method by extended POD naturally leads to higher cardinal of the primal extended RB (see Table 1).

Regarding the errors, we can see that for a parametric point already well treated by the FE enrichment method (see $P_1 = 170$ MPa and $h_1 = 3.2$ mm), few additional POD modes are enough to lower the dual relative error. However, when the dual relative errors remain high with the FE enrichment method, it is useful to add more POD modes to lower the dual relative error. For example, on the point $P_1 = 30$ MPa and $h_1 = 5.25$ mm, adding 9 POD modes (corresponding to $\epsilon_{\mathcal{POD}2}$=3.5 10^{-7}) allows us to divide the dual relative error by four. We can outline from Table 1 that the choice of $\epsilon_{\mathcal{POD}2}$ remains not trivial as a too low values lead to phenomena associated with over-fitting that degrade the accuracy of the solutions.

Let us notice an enrichment method with POD modes only will not be efficient to lower the condition number of the projected contact rigidity matrix without impacting the precision of the reduced solutions.

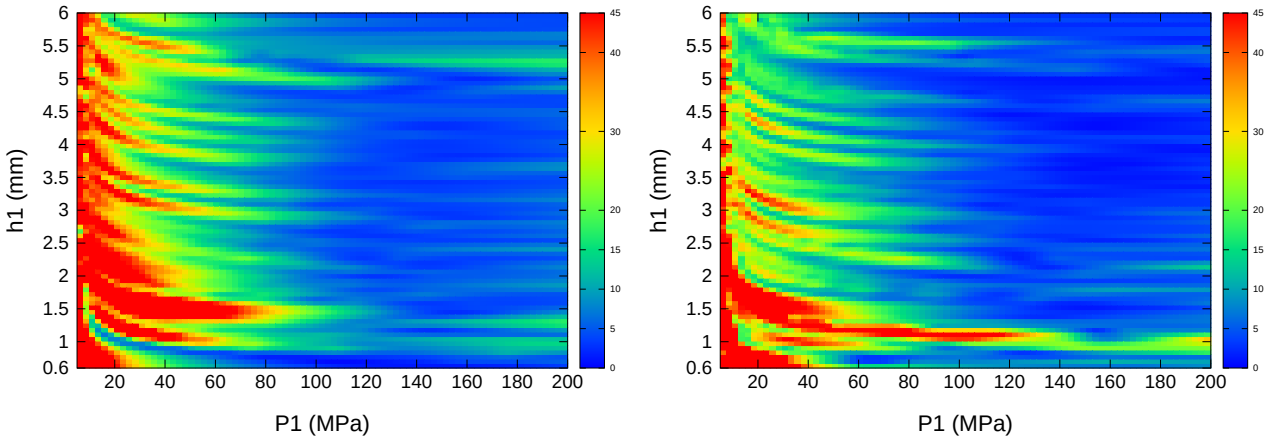
In Figure 10, the output of the greedy algorithm on the primal RB extended with 9 supplementary POD modes (denoted 'l+9' in the legend) are reported. For an equivalent primal RB dimension, the POD+EF enrichment strategy becomes interesting only for low errors, see Figure 10a. However regarding to condition number of the projected contact rigidity matrix, the POD+FE enrichment strategy is quite always improving the dual errors for a given prescribed condition number, see Figure 10b. It is worth noting that the solvability condition is reached for a comparable condition number. This confirms the strong relation between solvability, precision and condition number.

The two enrichment methods are now compared on the whole parametric space. For that purpose, HHR solutions are calculated on 6400 points regularly distributed in the parametric space. We set the maximum condition number to $\kappa_{max} = 10$ which seems to be a good compromise between precision and dimension of the extended primal RB.

The maps of the dual relative errors are reported in Figure 11 for the FE enrichment strategy (Fig.11b) and POD+EF enrichment strategy with 9 suuplementary POD modes (Fig. 11a). It clearly shows an improvement due to the POD+EF enrichment method. The dimension of the final extended primal RB ($\bar{l}$) is in this case 7 modes larger (instead of 9), which implies that the improvement of the condition number is slightly facilitated by the addition of POD modes.

Whatever the enrichment strategy, it remains high dual relative errors on the left side of the maps, for low pressures P_1. This highlights the complexity to efficiently reduced a problem on such parametric space, where the contact can be strong, weak or even absent, and on an large or small surface.
This naturally open the way to study clustering strategies on the parametric space.

Before presenting numerical results for the clustering strategies, we first study here the impact of the size of the RID over the enrichment strategies. We hence consider a larger extended RID with more potential contact points. This RID is simply obtain considering a higher dual POD RB threshold of $\epsilon = 1 \times 10^{-4}$ for the DEIM point selection, see Section 5.2. In this case, the RID contains $N_\lambda^C = 60$ potential contacts with always the same primal RB of $l = 36$ POD modes, see Figure 12.
As a consequence of the augmentation of N_λ^C, the solvability condition requires a larger enriched primal RB. For this RID, the FE enrichment method reaches the solvability condition for $\bar{l} = 61$, that is to say 35 supplementary FE modes. But as previously shown, for this extended basis, the dual reduced solutions are not enough satisfying. Further enrichment based on the condition number of the projected contact rigidity matrix is required.
In Figure 13, we report dual relative error maps using both FE and POD+FE enrichment with a maximal



(a) FE enrichment. $\bar{l} = 55$.

(b) POD+FE enrichment (l+9). $\bar{l} = 62$.

Figure 11: 2D axisymmetric test case. Dual relative errors maps (in %) for the HHR model with $l = 36$, $N_\lambda^C = 34$ on a test space of 6400 points. Enrichment strategies applied with $\kappa_{max} = 10$.

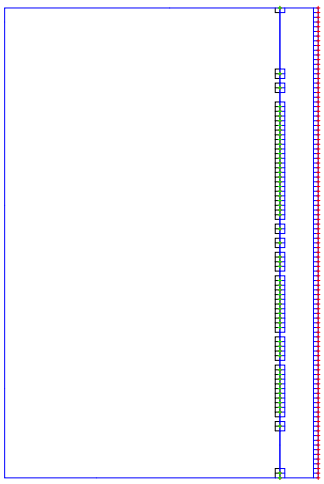
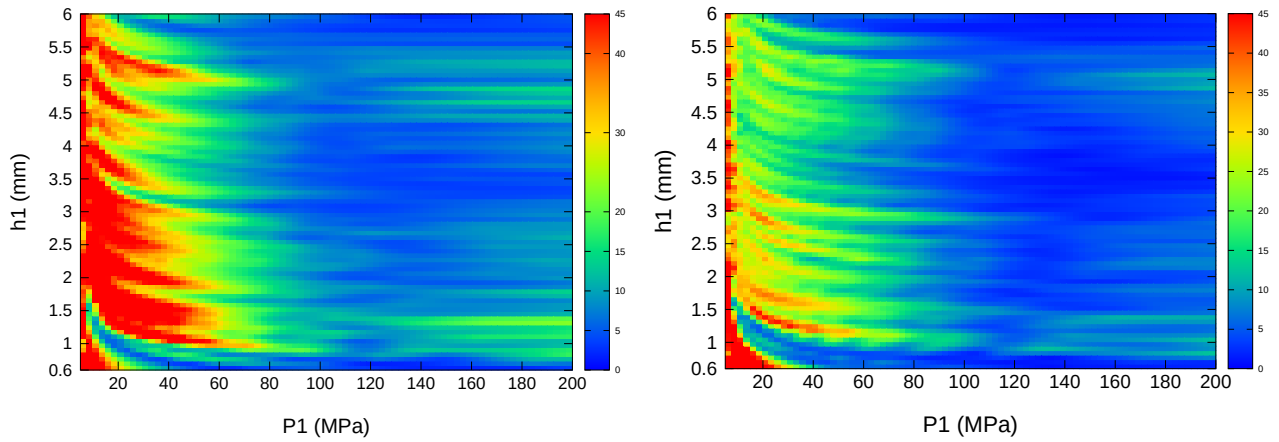


Figure 12: 2D axisymmetric test case. Larger extended RID. $l = 36$, $N_\lambda^C = 60$

condition number $\kappa_{max} = 10$. This results in a primal RB with a cardinal of $\bar{l} = 83$ for the FE enrichment method and of $\bar{l} = 91$ for the POD+FE enrichment method with 9 supplementary POD modes ($l + 9$). Compared to Figure 11a, the first conclusion to be drawn is that the increase of N_λ^C does not improve the error maps for the FE enrichment method, see Figure 13a. In counterpart, we can see a significant improvement for the POD+FE enrichment method, cf. Figures 11b and 13b .

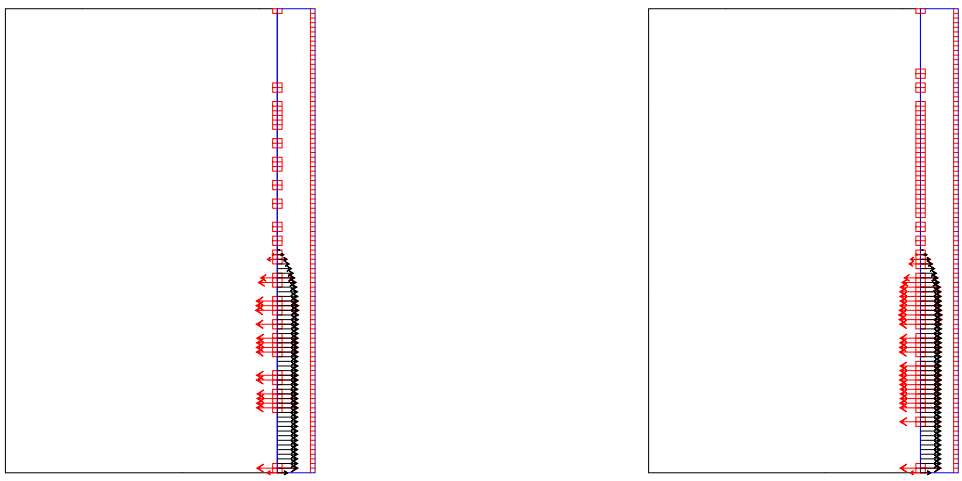


(a) FE enrichment. $\bar{l} = 83$.

(b) POD+FE enrichment (l+9). $\bar{l} = 91$.

Figure 13: 2D axisymmetric test case. Dual relative errors maps (in %) for the HHR model with $l = 36$, $N_\lambda^C = 60$ on a test space of 6400 points. Enrichment strategies applied with $\kappa_{max} = 10$.

An important point here is to remember that the RID contains more Lagrange multipliers when $N_\lambda^C = 60$ than when $N_\lambda^C = 34$, meaning that the computation of the dual relative error as defined by (18) implies more forces in the case $N_\lambda^C = 60$. Contact forces for the parametric point $P_1 = 170$MPa and $h_1 = 3.2$mm for both configuration of N_λ^C with full FE enrichment ($\kappa_{max} = 1$) are reported in Figure 14. In both situations, the dual forces are locally very accurate on the RID (2.6% for $N_\lambda^C = 60$, and 2.5% for $N_\lambda^C = 34$). However, having more potential contact points in the RID, and therefore more contact forces solved is obviously an advantage both for the physical interpretation of the simulation and the reconstruction of the dual solution, but the drawback is the dimension of the primal RB that grows with N_λ^C.
This motivates furthermore to look forward clustering approaches to limit the increase of the primal RB cardinal while reaching the same accuracy.



(a) HHR with $l = 36$, $N_\lambda^C = 34$ and $\bar{l} = 70$.

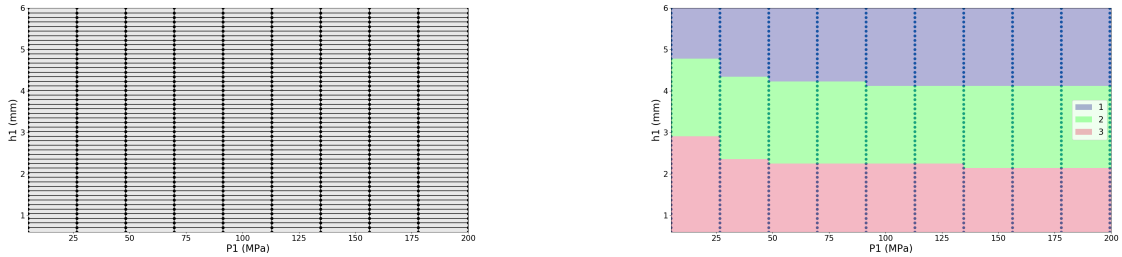
(b) HHR with $l = 36$, $N_\lambda^C = 60$ and $\bar{l} = 96$.

Figure 14: 2D axisymmetric test case. Contact forces for $P_1 = 170$MPa and $h_1 = 3.2$mm. HHR with FE enrichment (red arrows) and FOM (black arrows). (forces superposed on the right side).

5.2.2 Clustering strategies

From regular snapshots

We focus here on the clustering strategy proposed in Section 4.1. The same regular training set $\mathcal{P}_{tr}$ composed by 500 snapshots is used. The partition of parametric space is built through groups of 4 snapshots forming rectangles in the 2D parametric space representation, see Figure 15a. In Figure 15b we can see the results of the clustering algorithm for $K = 3$ clusters, each point representing a snapshot and the background color representing a cluster.



(a) Partition of the parametric space.

(b) Clusters for $K = 3$.

Figure 15: 2D axisymmetric test case. Partition and clustering of the parametric space from regular snapshots.

The clusters contain respectively 170, 194 and 170 snapshots in their local snapshots matrices. Figure 15b shows that the clustering strategy on the parametric space naturally focuses on the position h_1 of the potential contact zone. It confirms that a large variation of the potential contact zone is the preponderant parameter of nonlinearity.

As the parametric space is locally reduced into the cluster, the local HHR models are obtained from a primal POD threshold $\epsilon_{\mathcal{POD}} = 10^{-8}$. The choice of this threshold still enables a low-rank approximation as confirmed by Figure 16.

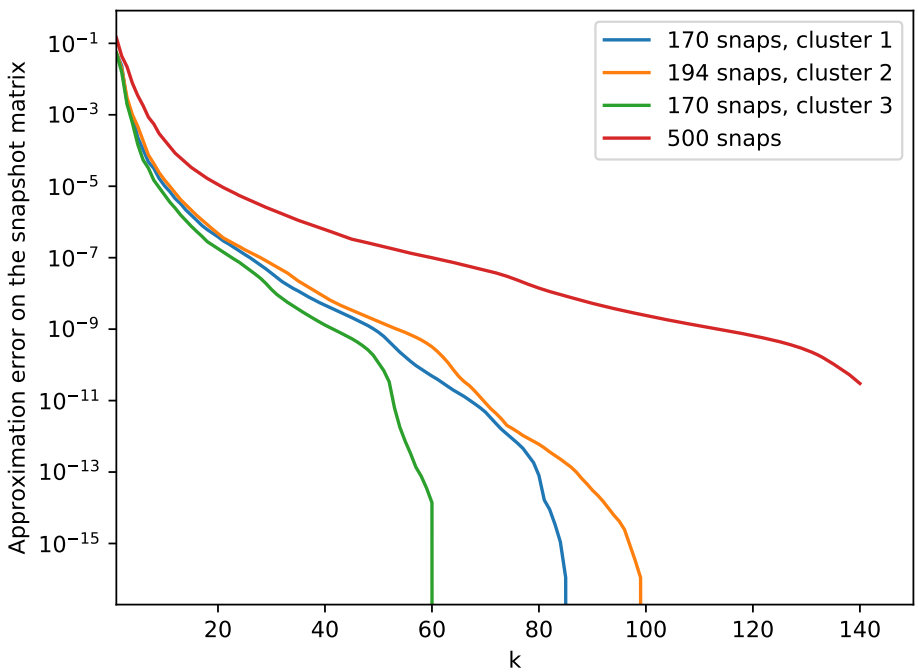
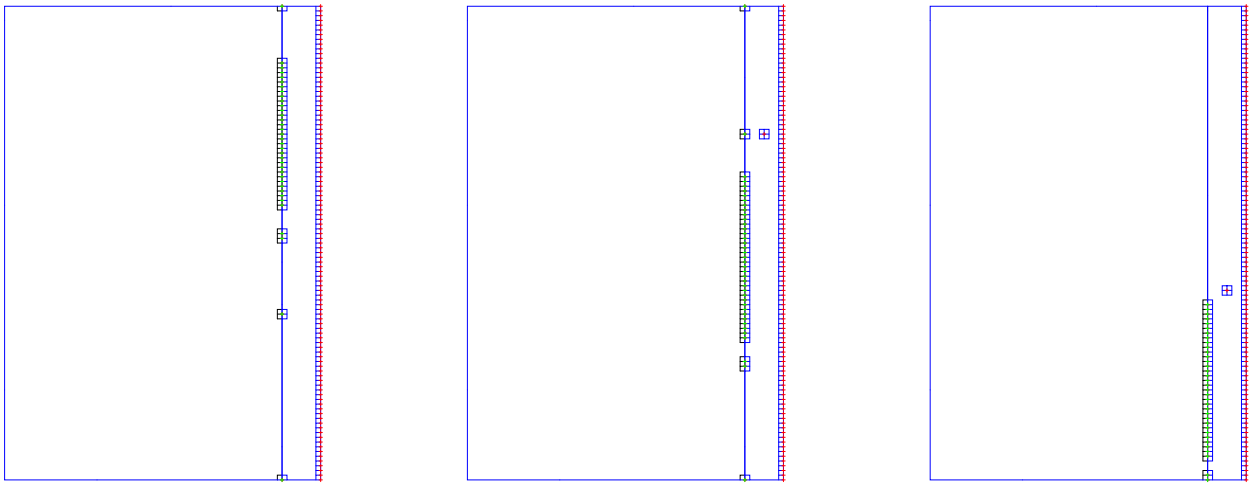


Figure 16: 2D axisymetric test case. Approximation error on the snapshot matrices for the case with (K=3) and without clustering.

The primal RB size in each cluster is $l = \{31, 39, 36\}$. A local RID is built for each local model. As previously a DEIM on the primal RB and a contribution from a DEIM on a dual POD RB (with a threshold $\epsilon = 10^{-6}$) is used to obtain the set of nodes $\mathcal{L}$. The resulting RIDs are reported in Figure 17. The number of potential contacts treated in each RID is $N_\lambda^C = \{36, 40, 35\}$.

As a direct results of the clusters shapes, the RID of the local reduced order models are much more localized. Figure 17 confirms the observation made from Figure 15b, as the three RIDs finally form a domain decomposition of the FOM potential contact zone. For each cluster, the RID is localized around the possible values of h_1,



(a) RID for the ROM associated with cluster 1.($l = 31$, $N_\lambda^C = 36$)

(b) RID for the ROM associated with cluster 2. ($l = 39$, $N_\lambda^C = 40$)

(c) RID for the ROM associated with cluster 3. ($l = 36$, $N_\lambda^C = 35$)

Figure 17: 2D axisymmetric test case.

which represents the zone of status change (contact/gap).

On each cluster, the primal RB extension is performed through the FE enrichment strategy only (since $\epsilon_{\mathcal{POD}}$ is already low). The results obtained with a maximum enrichment ($\kappa_{max} = 1$) for the two points studied in Table 1 are reported in Table 2.

Table 2: 2D axisymmetric test case. Study of the clustering strategy with full local FE enrichment.

	Cluster label	$\kappa(\underline{\underline{B}}[\mathcal{A}_\lambda, \mathcal{A}_c]\underline{\underline{\overline{V}}}[\mathcal{A}_c, :])$ before FE enrich.	$l = card(\underline{\underline{V}})$ initial	$\bar{l} = card(\underline{\underline{\overline{V}}})$ final	Primal Rel. err.	Dual Rel. err.
$P_1 = 170$MPa $h_1 = 3.2$mm	2	$1.5\ 10^5$	39	79	0.03%	2.0%
$P_1 = 30$MPa $h_1 = 5.25$mm	1	$2.2\ 10^4$	31	67	0.02%	2.4%

For the point $P1 = 170$MPa and $h_1 = 3.2$mm, already accurately treated by the enrichment strategies performed on the whole parametric space, similar dual errors are obtained (with a significant improvement on the primal error).
For the point $P1 = 30$MPa and $h_1 = 5.25$mm the results are much better than the ones obtained on the whole parametric space whatever the enrichment strategy (see Table 1). Moreover, the size of the primal RB remains limited ($\bar{l} = 67$).

Generally speaking, the proposed clustering strategy helps smoothening the dual relative errors. It particularly improves the accuracy on the parametric points where the relative errors are high without the clustering strategy, see Figure 18.
Concerning the RB dimensions, this clustering strategy leads to comparable sizes than the strategies introduced in Section 5.2.1. However the dual errors are greatly improved by the clustering compared to Figure 11.
The error maps of Figure 18 are even better than the ones obtained on an extended RID with the POD+EF enrichment strategy, see Figure 13b. In this latter case, the bases dimensions were almost 40% larger.

We can underline, that even with the clustering strategy, it remains difficult to accurately approximate very low contact pressure parametric zones.
The number of clusters K can not be too much increased. Indeed, the more cluster the smaller the local parametric space becomes. But each cluster must have enough snapshots to build a representative POD RB and then an accurate HHR model.
More snapshots could be initially used, but another interesting way to investigate would be to consider a recursive clustering strategy (with possibly adding snapshots) inside clusters having the least accurate HHR solutions.

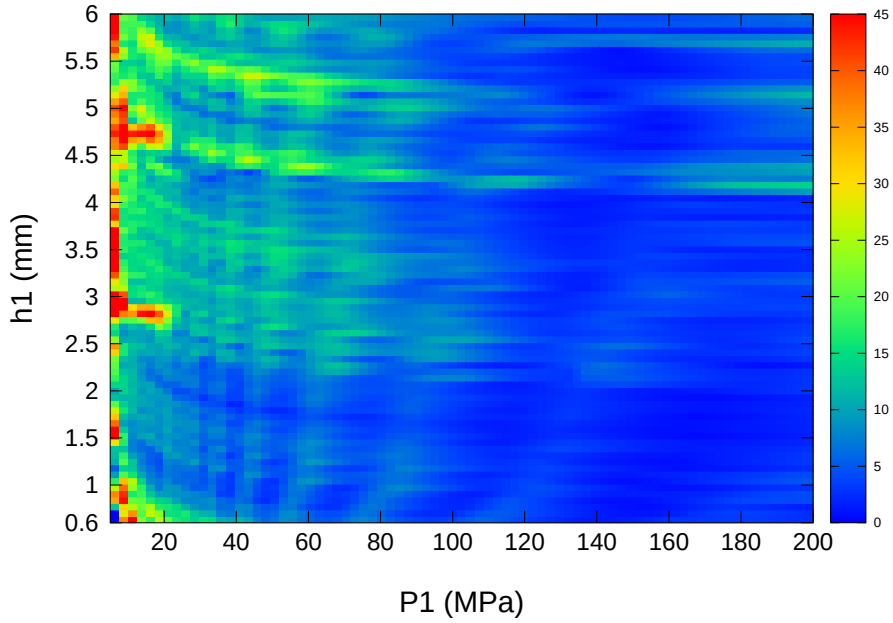
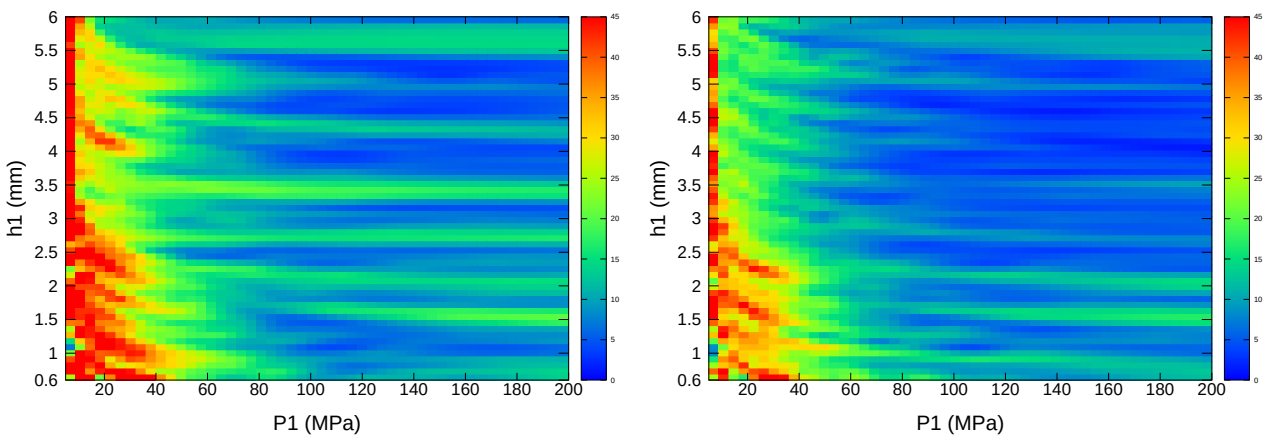


Figure 18: 2D axisymmetric test case. Dual relative errors map (in %) for the clustering strategy with $K = 3$ clusters. Test space of 6400 points. FE enrichment method with $\kappa_{max} = 10$. HHR models with $l = \{31, 39, 36\}$, $N_\lambda^C = \{36, 40, 35\}$ and $\bar{l} = \{54, 67, 57\}$

From irregular snapshots

We focus here on an irregular distribution of the snapshots in the parametric space $\mathcal{P}$ obtained through a snapshot selection by the greedy algorithm introduced in [9]. The resulting training set $\mathcal{P}_{tr}$ contains 200 snapshots, which can be visualized in Figure 20a. Let us notice that they are mainly located around small values of contact pressure P_1.

In order to appreciate the solutions obtained by the clustering algorithm introduced in Section 4.2, we first briefly present some key results obtained by the HHR method with enrichment trained on irregular snapshots of $\mathcal{P}_{tr}$ (without clustering).
The HHR model is built using primal POD thresolhold $\epsilon_{\mathcal{POD}} = 10^{-6}$, resulting in a primal RB of dimension $l = 33$. The RID is obtained as the one shown in Figure 12 (DEIM on the primal RB, and a contribution from a DEIM on a POD dual RB obtained with the treshold $\epsilon = 10^{-4}$). It finally contains $N_\lambda^C = 57$ potential contacts. For the POD+FE enrichment method an error threshold $\epsilon_{\mathcal{POD}2} = 3.5 \times 10^{-7}$ is used (as in Figs. 13b and 11b) resulting in a primal RB of dimension $(l + 5) = 38$ before enrichment. Both enrichment strategies are moreover applied setting $\kappa_{max} = 10$ as a maximum condition number in Algorithm 1. The FE enrichment method results in a primal RB of cardinal $\bar{l} = 60$, while the POD+FE method results in a primal RB of cardinal $\bar{l} = 82$. We present in Figure 19 the dual relative errors on the parametric space for both reduced models.



(a) FE enrichment. $\bar{l} = 60$.

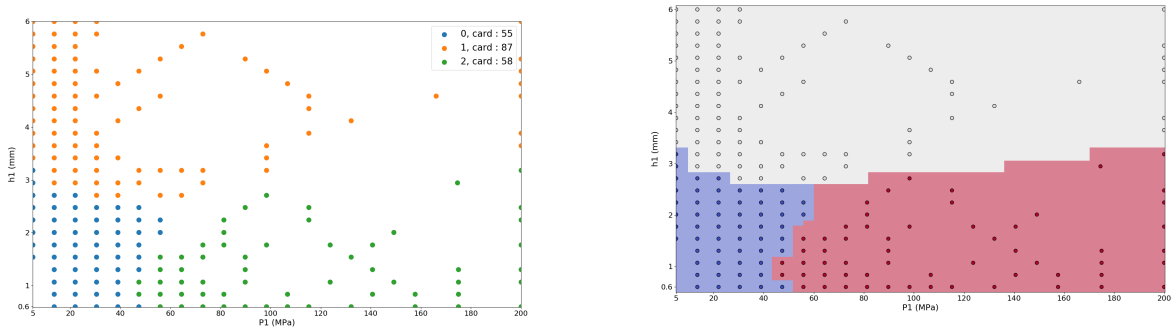
(b) POD+FE enrichment (l+5). $\bar{l} = 82$.

Figure 19: 2D axisymmetric test case. Irregular snapshots in the training set. Dual relative errors maps (in %) for the HHR model with $l = 33$, $N_\lambda^C = 57$ on a test space of 6400 points. Enrichment strategies applied with $\kappa_{max} = 10$.

The errors are in good agreement with those obtained with 500 regular snapshots on a RID of similar size, see

Figure 13. It is worth underlying that better results are globally reached (especially for low contact pressure) and with a more compact primal RB (lower dimension). This confirms the interest of greedy-selected snapshots [9].

In order to evaluate the clustering strategy based on irregular snapshots (cf. Section 4.2) on this training set, we build $K = 3$ clusters through the k-medoid algorithm (as for the previous clustering strategy based on regular snapshots). In Figure 20a, the cluster labels are reported. The Random Forest classifier is then trained to obtain a prediction function, represented in Figure 20b as a partition of the parametric space. Let us underline, that in this case the clusters are not uniquely defined thanks to the contact position h_1 but also according to the pressure values P_1 for small h_1. It confirms that a small contact pressure applied to a reduced contact area is difficult to capture.

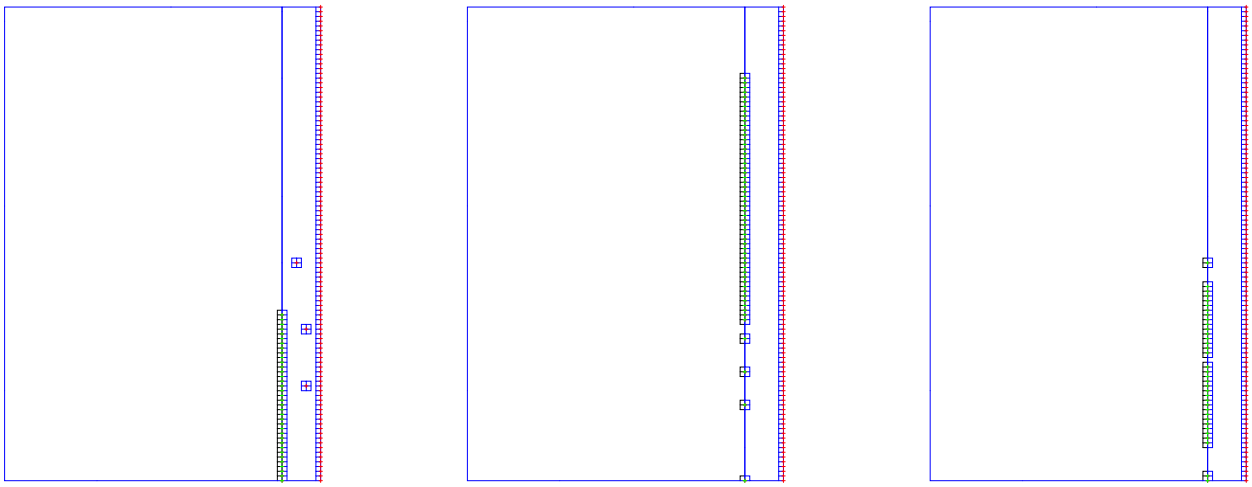


(a) Cluster labels on the irregular snapshots.

(b) Prediction function obtained with the Random Forest classifier.(label 0:blue, label 1:grey, label 2:red).

Figure 20: 2D axisymmetric test case. Clustering strategy for irregular snapshots. $K = 3$ clusters

The local reduced order models are built with the same strategy than for the clustering on regular snapshots. The associated local RIDs are shown in Figure 21. As for the regular snapshots case the RIDs are localized around the potential contact surface of the cluster (h_1 parameter). RIDs of cluster 0 and 2 are quite similar.



(a) Cluster 0.
($l = 27$, $N_\lambda^C = 36$)

(b) Cluster 1.
($l = 35$, $N_\lambda^C = 56$)

(c) Cluster 2.
($l = 23$, $N_\lambda^C = 35$)

Figure 21: 2D axisymmetric test case. Irregular snapshots. RIDs associated to the local ROMs.

The dual errors map is shown in Figure 22. The clustering strategy helps smoothening the dual relative errors, as observed previously. The errors are considerably reduced for RB of comparable sizes to those without clustering (cf. Figure 19), even for low contact pressures.

As improvement, we can see that an adaptive snapshots strategy into cluster could be interesting since big clusters (see cluster 1) may lack of snapshots and since the cluster delimitation could be improved where snapshots are sparsely distributed.

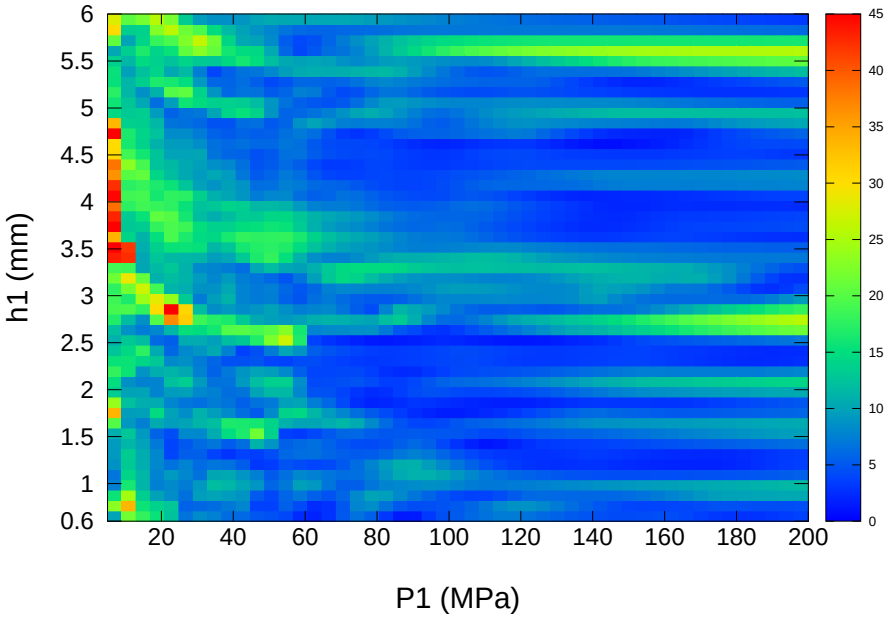


Figure 22: 2D axisymmetric test case. Irregular snapshots. Dual relative errors map (in %) for the clustering strategy with $K = 3$ clusters. FE enrichment method with $\kappa_{max} = 10$. HHR models with $l = \{27, 35, 23\}$, $N_\lambda^C = \{36, 56, 35\}$ and $\bar{l} = \{52, 77, 48\}$

6 Conclusion

Efficient strategies to improve the dual solution precision of reduced order modeling of contact mechanics problems treated by Lagrange multipliers have been proposed. We have focused on the Hybrid Hyper-Reduced (HHR) method where the dual reduced basis returns to a restriction on a Reduced Integration Domain of the full order dual basis.
First, the strong correlation between the condition number of the projected contact rigidity matrix and the accuracy of the dual solution has been highlighted. Two strategies of enrichment of the primal POD reduced basis have been introduced. These strategies allows us to verify the solvability condition of the reduced saddle-point problem (if not respected initially) but also to improve the accuracy of the method, especially the dual solution. Secondly for large parametric space, especially for large variation of the contact zone where the non-smoothness of the contact conditions makes difficult low rank dual approximations, clustering-based strategy have been proposed. Precise local enriched HHR models are then built leading to globally enhance the accuracy of the reduced model using limited sizes of reduced bases. Clustering strategies have been presented for regular or irregular snapshots.

These different strategies have been applied on 2-dimensional test cases demonstrated the very good accuracy of the method on small parametric spaces, and the usefulness contribution of the enrichment and clustering strategies on large parametric spaces. The HHR method is able to accurately treat non-smooth cases where the contact forces can be strong or weak, and very localized or extended.

Future works will focus on the extension of the HHR to more complex geometries where node-to-segment contact detection is required.

7 Acknowledgements

The authors are grateful to the PLEIADES project, financially supported by the CEA, EDF (Électricité de France) and FRAMATOME, that funded this research work.

References

References

[1] M. Balajewicz and J. Toivanen. Reduced order models for pricing european and american options under stochastic volatility and jump-diffusion models. *Journal of Computational Science*, 20:198–204, 2017. 1

[2] E. Bader, Z. Zhang, and K. Veroy. An empirical interpolation approach to reduced basis approximations for variational inequalities. *Mathematical and Computer Modelling of Dynamical Systems*, 22(4):345–361, 2016. 1

[3] D. Scheffold, C. Bach, F. Duddeck, G. Müller, and M. Buchschmid. Vibration frequency optimization of jointed structures with contact nonlinearities using hyper-reduction. *IFAC-PapersOnLine*, 51(2):843–848, 2018. 1

[4] J. Ballani, D. B. P. Huynh, D. J. Knezevic, L. Nguyen, and A. T. Patera. A component-based hybrid reduced basis/finite element method for solid mechanics with local nonlinearities. *Computer Methods in Applied Mechanics and Engineering*, 329:498–531, 2018. 1

[5] A. Giacoma, D. Dureisseix, A. Gravouil, and M. Rochette. A multiscale large time increment/fas algorithm with time-space model reduction for frictional contact problems. *International Journal for Numerical Methods in Engineering*, 97(3):207–230, 2014. 1

[6] A. Giacoma, D. Dureisseix, and A. Gravouil. An efficient quasi-optimal space-time pgd. application to frictional contact mechanics. *Advanced Modeling and Simulation in Engineering Sciences*, 3(1):1–17, 2016. 1

[7] B. Haasdonk, B. Wohlmuth, and J. Salomon. A reduced basis method for parametrized variational inequalities. *SIAM Journal on Numerical Analysis*, 50(5):2656–2676, October 2012. 2

[8] M. Balajewicz, D. Amsallem, and C. Farhat. Projection-based model reduction for contact problems. *International Journal for Numerical Methods in Engineering*, 106(8):644–663, 2016. 2

[9] J. Fauque, I. Ramière, and D. Ryckelynck. Hybrid hyper-reduced modeling for contact mechanics problems. *International Journal for Numerical Methods in Engineering*, 115(1):117–139, July 2018. 2, 4, 5, 9, 18, 19

[10] A. Benaceur, A. Ern, and V. Ehrlacher. A reduced basis method for parametrized variational inequalities applied to contact mechanics. *International Journal for Numerical Methods in Engineering*, 121(6):1170–1197, October 2019. 2, 9, 10

[11] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004. 2

[12] S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010. 2, 4

[13] P. Astrid, S. Weiland, K. Willcox, and T. Backx. Missing point estimation in models described by proper orthogonal decomposition. *IEEE Transactions on Automatic Control*, 53(10):2237–2251, 2008. 2

[14] C. Farhat, P. Avery, T. Chapman, and J. Cortial. Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency. *International Journal for Numerical Methods in Engineering*, 98(9):625–662, 2014. 2

[15] D. Ryckelynck. Hyper-reduction of mechanical models involving internal variables. *International Journal for Numerical Methods in Engineering*, 77(1):75–89, 2009. 2, 4

[16] E. Delhez, F. Nyssen, J. C. Golinval, and A. Batailly. Reduced order modeling of blades with geometric nonlinearities and contact interactions. *Journal of Sound Vibration*, 500:116037, 2021. 2

[17] D. Manvelyan, B. Simeon, and U. Wever. An efficient model order reduction scheme for dynamic contact in linear elasticity. *Computational Mechanics*, 68(6):1283–1295, 2021. 2

[18] T. Daniel, F. Casenave, N. Akkari, and D. Ryckelynck. Model order reduction assisted by deep neural networks (rom-net). *Advanced Modeling and Simulation in Engineering Sciences*, 7(1):16, 2020. 2, 7, 8

[19] D. Amsallem, M.J. Zahr, and K. Washabaugh. Fast local reduced basis updates for the efficient reduction of nonlinear systems with hyper-reduction. *Adv Comput Math.*, 41(5):1187–1230, 2015. 2

[20] M. Redeker and B. Haasdonk. A pod-eim reduced two-scale model for crystal growth. *Advances in Computational Mathematics*, 41(5):987–1013, 2015. 2

[21] B. Peherstorfer, D. Butnaru, K. Willcox, and H.-J. Bungartz. Localized discrete empirical interpolation method. *SIAM Journal on Scientific Computing*, 36(1):A168–A192, 2014. 2

[22] S. Grimberg, C. Farhat, R. Tezaur, and C. Bou-Mosleh. Mesh sampling and weighting for the hyperreduction of nonlinear petrov–galerkin reduced-order models with local reduced-order bases. *International Journal for Numerical Methods in Engineering*, 122(7):1846–1874, 2021. 2

[23] G. Duvaut and J.-L. Lions. *Inequalities in mechanics and physics.* Springer Science & Business Media, 1976. 3

[24] N. Kikuchi and J. Oden. Contact problems in elasticity: A study of variational inequalities and finite element methods. 1987. 3

[25] P. Wriggers. *Computational Contact Mechanics*. Springer Science, 2006. 3

[26] I. Babuška. The finite element method with lagrangian multipliers. *Numer. Math. 20, 179-192*, 1973. 3

[27] F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from lagrangian multipliers. *ESAIM : Mathematical Modelling andNumerical Analysis - Modélisation Mathématique et Analyse Numérique*, 1974. 3

[28] Nagi El-Abbasi and Klaus-Jürgen Bathe. Stability and patch test performance of contact discretizations and a new solution algorithm. *Computers and Structures*, 79(16):1473–1486, 2001. 3

[29] S. Sirovich. Turbulence and the dynamics of coherent structures. *Parts I-III. Q Appl Math.*, 1987. 4

[30] D. Ryckelynck, T. Goessel, and F. Nguyen. Mechanical dissimilarity of defects in welded joints via Grassmann manifold and machine learning. *Comptes Rendus. Mécanique*, 348(10-11):911–935, 2020. 7

[31] H. Park and C. Jun. A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications*, 36(2, Part 2):3336–3341, 2009. 7

[32] L. Breiman. Random forests. *Machine Learning*, 45, 2001. 8

[33] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. 8

[34] Cast3M. `http://www-cast3m.cea.fr`. 9

[35] H. Liu, I. Ramière, and F. Lebon. On the coupling of local multilevel mesh refinement and zz methods for unilateral frictional contact problems in elastostatics. *Computer Methods in Applied Mechanics and Engineering*, (323):1–26, 2017. 11

A Approximation error

The approximation error is obtained by using the energy rate carried by the k first singular vectors of the SVD on the snapshot matrices:

$$\mathcal{E}(k) = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{d} \sigma_i^2} = 1 - \frac{\sum_{i=k+1}^{d} \sigma_i^2}{\sum_{i=1}^{d} \sigma_i^2} \tag{20}$$