A Hierarchical POD Reduction Method of Finite Element Models with Application to Simulated Mechanical Systems

Martin Björklund

A thesis submitted for the degree of

Master of Science in Engineering Physics

Supervisor: Mats G. Larson
Department of Mathematics and Mathematical Statistics, Umeå Univeristy

Examiner: Daniel Vågberg
Department of Physics, Umeå University

May 7, 2012

Abstract

When simulating mechanical systems the flexibility of the components often has to be taken into account. This is particularly important for simulations when high detailed information is demanded, e.g. to calculate stresses. To this end the Finite Element Method (FEM) is often used. However the models can become very large, containing millions of degrees of freedom. Solving large linear systems are computationally demanding. Therefore ways of reducing the problem is often sought. These reduction does, however, remove much of the details that was to be investigated. In this thesis this problem is addressed by creating a reduction scheme, using Proper Orthogonal Decomposition (POD), that significantly reduces a problem but still captures much of the details.

A novel method for enriching regular POD-based model reduction methods with hierarchically determined enrichment POD-modes is developed. The method is proposed and validated in a FEM application towards dynamical simulation. The enriched method is compared against a regular POD reduction technique. An numerical study is made of a model example of linear elasticity in a gearwheel. The numerical study suggests that the error of displacements is around ten times smaller, on average, when using the enriched basis compared to a reference basis of equal dimensionality consisting of only regular POD modes. Also it is shown that local quantities as the von Mises stress in a gearwheel tooth is preserved much better using the enriched basis. An a posteriori error estimate is proposed and proved for the static case, showing that the error is bound.

Keywords: Model reduction, FEM, POD, enrichment, hierarchy, global and local behaviour

Sammanfattning

När man simulerar mekaniska system så måste man ofta ta hänsyn till de ingående komponenternas flexibilitet. Detta är särskilt viktigt då man gör simuleringar med krav på hög detaljkännedom, såsom mätningar av spänningar i kugghjul etc. Till detta ändamål används ofta en Finit Element Metod (FEM). Dock kan modellerna ofta bli väldigt stora, med över en miljon frihetsgrader. Att lösa linjära system av den storleken är beräkningsmässigt krävande. Därför är det naturligt att försöka reducera problemen. Reduktion innebär dock att information försvinner, i synnerhet de detaljer som skulle beräknas. I detta examensarbete så behandlas problemet genom att skapa en ny metod för reducering av stora finita element modeller. Metoden bygger på tidigare kunskap om Proper Orthogonal Decomposition (POD) som ett sätt att reducera modeller. Den nya metoden reducerar finita ellement modeller samtidigt som den bibehåller hög detalj.

En ny metod utvecklas för att berika en vanlig POD-baserad modellreduktion med hjälp av hieraktiskt bestämda berikningsmoder. Metoden beskrivs och testas i en dynamisk FEM-applikation av elasticitet i ett kugghjul i 2 dimensioner. Metoden för berikning jämförs numeriskt med en metod som använder vanlig POD-reduktion. Körningar visar att felet i den berikade metoden är omkring 10 gånger mindre, i genomsnitt, jämfört med en vanlig metod. Det visas också att spänningar bevaras på ett mycket bra sätt med den nya berikningsmetoden. Dessutom så formuleras och bevisas ett a posteriori estimat för statiska lastfall, vilket innebär att felet i metoden är bundet.

Nyckelord: Modellreduktion, FEM, POD, berikning, hirarki, globalt och lokalt beteende.

Acknowledgements

The work herein is a master thesis in physics engineering at the Department of Physics at Umeå University. The work has been carried out within the group of Computational Mathematics at the Department of Mathematics and Mathematical statistics at Umeå University. A work space was provided by UMIT Research Lab.

There are many people who have helped me and inspired me during my master thesis. First and foremost I would like to thank my supervisor Mats G. Larson for all the inspiration, help and guidance he has given during this thesis. I would also like to thank Håkan Jakobsson for his patience and always answering my endless questions. Also in no apparent order I would like to thank the rest of the people who have been part of the Computational Mathematics research group during my thesis: Karl Larsson, Robert Söderlund, Tor Troeng, Per Vesterlund and Jakob Öhrman. You have all been very helpful and inspiring. I would also like to thank all of the colleagues at UMIT Research Lab who have all contributed to a pleasant working environment.

I also wish to thank my fiancée Anna, our daughters Tilia and Elin, my fience's father Lennart, her mother Inga-Maj, and her siblings Andreas, Annika and Stefan and my parents Ulla and Torbjörn, my brother Marcus and his partner Fredrik for all the love, encouragement and support which without I would not have done so much in my life and certainly not this thesis.

Last but not least I would like to thank all of the close friends I've made during my time studying Engineering Physics at Umeå University, Erik, Jonas, Kristoffer, Tomas, Christofer, Mattias and Mikael.

Contents

1 Introduction

This thesis concerns creating a new method of reducing large finite element models using and hierarchical approach to the Proper Orthogonal Decomposition (POD) for application to simulated mechanical systems. Below is a description of the method, followed by a problem statement together with the thesis objectives and a short summary of results. After that the thesis is outlined as follows: The 2nd chapter presents the continuous model of linear elasticity, then a finite element model for linear elastodynamics is derived by first finding the variational form from the continuous model. The 3rd chapter starts of by describing model reduction and the POD. Then in chapter 4 a proposed method to do adaptive model reduction in a simulation is proposed. In the 5th chapter the gearwheel model problem is described together with numerical studies and results of them. Finally in chapter 6 the results of the evaluation are presented and discussed together with proposed future work.

Approximation of Partial Differential Equation's and Reduction of the Models

Partial Differential Equation's (PDE) occur everywhere in science. Examples are thermodynamics, structural dynamics and fluid dynamics. Trying to solve a PDE on a complex domain, such as a car body for instance, is hard. Therefore different methods are used to approximate the continuous problem with a discrete problem that can be handled by computer software. In this thesis the Finite Element Method(FEM) is used to approximate a time-dependent elasticity problem.

A matter of concern is that even when a problem is approximated it can consist of millions of degrees of freedom, making a simulation very computationally demanding. There is always a struggle between making a simulation as accurate as possible, and running as fast as possible. It is therefore natural to look for ways to reduce the problem to make it easier to solve. This is known as model reduction. One set of these reductions are so called projection reductions. Their aim is to project the problem onto a subspace that is of significantly lower dimension, to reduce the degrees of freedom, but still captures the properties of the problem to a satisfactory degree.

The projection method used in this thesis is POD. The use of POD to reduce a problem is well known [7]. However it is not widely used within the area of elasticity, which this thesis concerns. The POD returns a set of orthonormal modes, or basis vectors. Thus they span a subspace to the original problem. An advantage with the POD is that it captures much of the overall properties of a data set in a few number of modes, or basis-vectors, and also these modes are ordered by significance. This makes a significant reduction possible by creating a basis consisting of the first m modes. How many modes that are to be used depends on how much a given quantity, e.g. energy, needs to be preserved. To preserve local properties, like small variations or high frequencies, most modes need to be used which makes a general POD method badly suited for reduction of systems that still require good accuracy in local quantities. In this thesis a way around this problem is developed by creating a reduced model using a hierarchical approach to create a POD-based basis enriched with POD-modes that captures local behaviour.

When simulating mechanical systems the flexibility of the components often has to be taken into account. This is particularly important for simulations when high detailed information is demanded, e.g. to calculate stresses. The problem studied in the thesis is a $2D$ gearwheel with the coordinate system fixed at the centre. The gear tooth boundary will be under the influence of a sliding contact. This sliding contact is modelled as a non-elastic gearwheel constantly in contact with the wheel considered. The contact force is pointing inwards, normal to the tooth boundary in the point of contact. A force acting on a point is said to be the load case for that point.

Local and Global Displacements

When a force acts on a body it is displaced. If the body is fixed there will be deformation. Note the difference between displacements and deformation. A deformation is a displacement but a displacement does not need to be a deformation. The key difference is that deformations give rise to stresses, pure displacements do not. When a force acts on a gear tooth the tooth will be displaced. It will be indented at the point of contact and also the whole tooth will get bent. In addition the entire gear wheel will also get deformed as the force tries to rotate the wheel around the fixed axis. With this in mind it is possible to classify deformations as global or local. Global deformation would be deformations that occur on a large scale in the gear wheel. In this thesis it is more precisely defined as the deformations that can arise from all forces. In other words if there exists deformations that are common to all the forces acting on the gear boundary, they are to be considered as global. Thus the local deformation would be the rest, the ones that only a subset of all forces contribute to. Examples of global deformation would, in this case, be the deformation of the inner parts of the gear wheel. On the other hand the bending of a gear tooth is not global. Because this deformation is tied only to the forces acting on that specific tooth. This is a hierarchical approach in one step. Where the global deformations would be the *parent* and the local deformation the *children.* Of course it is possible to incorporate more levels of hierarchy where some deformation will be children and parents giving a tree-like structure. However this is depending on the domain and in this case it suffices to use only one level of children.

The Chosen Method of Model Reduction

To do a reduction of this model the method of static load cases is used [2]. The static elasticity problem is solved for a set of load cases that are considered to mimic the expected forces acting on the gear wheel. Therefore there exists as many load cases as gear tooth boundary points. The static solution of the load cases yields a set of static responses, telling how the model gets displaced. Then the POD is invoked on the solution of the static case to generate POD-modes that capture the properties of the static solutions. Global POD modes will be those that captures the global static displacements to a satisfactory degree, generally relatively few modes are needed for this. But to get a good representation of the local displacements almost all remaining POD-modes need to be used. To get around this problem a routine that enables hierarchical adaptive reduction of the problem is created. Essentially there is a global basis created using global POD-modes. This is then used to get an idea of where there are local variations that gets lost. The variations are then clustered, indicating where a reduction refinement is needed. From the clusters enrichment POD-modes are obtained. The enrichment modes will then represent the part that was lost in the global reduction. Then the enrichment POD-modes are appended to the global basis, yielding an enriched basis. Since the clusters are not identical the simulation will be using different bases for reduction during different time-steps. In the case of the model problem studied in this thesis the clusters will correspond to the gear teeth.

All of the reduction refinement is carried out in the off-line phase of the simulation as a pre process step. This means that there are really no constraints on how expensive calculations can be. A FEM time dependent solver is created and a reduction scheme is incorporated into the solver. The scheme dictates when a new basis should be used. This means that the only part of the enriched method that is used in the on-line phase is the re-projection of the system of equations from one basis to another. This re-projection is usually an expensive operation so there is a need to construct the clusters in a way that they are not changed often.

1.1 Problem Statement

When simulating mechanical systems the components flexibility often has to be regarded. This is particularly important for simulations when high detail is demanded, i.e. to calculate stresses. To do this a finite element model is created. However since the model contains a large number of degrees of freedom it is not practical to use it in a dynamical simulation. Therefore it is necessary to reduce the model to a small number of modes that preserves the system properties to a satisfactory degree. Unfortunately a reduction means that information is lost, and in the case of POD the local details that is used to calculate stresses are lost. The problem then becomes to develop a reduction method based on the POD that offers a significant reduction and also preserves local quantities to a satisfactory degree.

1.2 Thesis Objectives

The aim of the master thesis is to develop a new model reduction technique for elastodynamics problems, based on static load case POD. The reduction technique shall better capture local variations than conventional model reduction methods using POD. It will then be implemented in a simulation software and tested against a reference reduction technique.

1.3 Summary of Results

- An enrichment technique using local modes based on local clusters of the error from a global POD projection is shown to generate greatly more accurate solutions in the von Mises stresses and also in the displacements than a reference POD reduction based only global modes.
- An a posteriori error estimate for the static load case is formulated and proved. The a posteriori estimate is used in the clustering routine in the enrichment algorithm.

1.4 Literature

For this thesis several articles and books were studied. The Ph.D. thesis by Jakobsson [2] has been an invaluable resource for this thesis not only for the technical aspects of POD in solid mechanics but also for insights into several reduction methods and research into these. For a deeper understanding of POD the book by Holmes and Lumley [7] was studied. For the use of FEM the book by Larson and Bengzon [6] is used. The article by Kerfriden, Passieux and Bordas[8] was studied to get insights into model reduction, POD and a method of local refinement of reduced models. The use of clustering techniques for arrangement of snapshots was briefly studied through the Ph.D. thesis by Cardoso [1]. Also Chakrabarti and Mehrotra [5] have studied the use of clustering snapshots and global/local reduction of data sets. The study of contact mechanics, dynamics of structures and computational solid mechanics was done using the books by Johnson [4], Shabana [9], Humar [3] and Fung and Tong [10]

2 Model of Linear Elasticity

This chapter involves the continuous model of linear elastodynamics. Then via the variational form a Finite Element Model is presented.

2.1 Continuous Model of Linear Elastodynamics

The linear elastic dynamic problem of a homogeneous isotropic body $\Omega \subset \mathbb{R}^d$ under the assumption of small deformations and strains reads: find the symmetric stress tensor $\boldsymbol{\sigma}(\mathbf{u})$ and the displacement vector $\mathbf{u}$, such that

$$\begin{aligned}
\rho\ddot{\mathbf{u}} - \nabla \cdot \boldsymbol{\sigma} &= \mathbf{f}, \quad x \in \Omega, & t > 0 \qquad (2.1a)\\
\boldsymbol{\sigma}(\mathbf{u}) &= 2\mu\epsilon(\mathbf{u}) + \lambda(\nabla \cdot \mathbf{u})I, \quad x \in \Omega, & t > 0 \qquad (2.1b)\\
\mathbf{u} &= \mathbf{u}_D, \quad x \in \Gamma_D, & t > 0 \qquad (2.1c)\\
\boldsymbol{\sigma}(\mathbf{u}) \cdot \mathbf{n} &= \mathbf{g}, \quad x \in \Gamma_N, & t > 0 \qquad (2.1d)\\
\mathbf{u} &= \mathbf{u}_0, \quad x \in \Omega, & t = 0 \qquad (2.1e)\\
\dot{\mathbf{u}} &= \mathbf{v}_0, \quad x \in \Omega, & t = 0 \qquad (2.1f)
\end{aligned}$$

where ρ is the density. For simplicity $\rho = 1$. $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$, $\mathbf{u}_0$ and $\mathbf{v}_0$ are initial conditions for displacements and velocities. The Cauchy strain tensor is written $\boldsymbol{\epsilon}(\mathbf{u}) = 1/2(\nabla\mathbf{u} + \nabla\mathbf{u}^T)$ and is used in e.q. (2.1b), which is Hooke's law, to get the stress tensor $\boldsymbol{\sigma}$. Also $\mathbf{f}$ is a body force, i.e. acting in every point of the domain Ω, $\mathbf{g}$ a given traction load acting along a segment Γ_N of the boundary. The boundary has outward unit normal $\mathbf{n}$. Assuming that along the rest of the boundary, Γ_D the body is clamped and cannot be displaced then $\mathbf{u}_D = 0$. The elastic properties of the body are given by the positive Lame parameters: $\mu = E[2(1+\nu)]^{-1}$ and $\lambda = E\nu[(1+\nu)(1-\nu)]^{-1}$ where E is Poisson's ratio and ν is Young's modulus. Ω is the continuous domain of the body. In this case Ω is to be partitioned to obtain a the finite dimensional space V^h.

2.2 Weak Formulation

The derivation of a finite element method always starts by rewriting the differential equation as a variational equation. That is done to get the PDE on a weak form from which a discrete form can be obtained. The weak form is obtained by multiplying equation (2.1a) with a test function $\mathbf{v} \in V = \{\mathbf{w} \in [H^1(\Omega)]^d : \mathbf{w}|_{\Gamma_D} = \mathbf{0}\}$ and integrate by parts over Ω[1]. The derivation of the variational form is not included. Assuming that

[1] $H^1(\Omega) = \{v : v \in L^2 \frac{\partial v}{\partial x_i} \in L^2, \forall i = 1, 2..d\}$. $H^1 \subset L^2$ is a Hilbert space. A Hilbert space has a norm and an inner product. All functions $v \in H^1$ require that their partial derivatives exist in L^2. For

$\mathbf{u}_D = 0$ and that $\rho(x)$ is constant the variational form of (2.1a) reads: Find $\mathbf{u} \in V$, such that

$$\int_\Omega \ddot{u}(x)v(x)dx + (\boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\epsilon}(\mathbf{v})) = (\mathbf{f}, \mathbf{v}) + (\mathbf{g}_N, \mathbf{v})_{\Gamma_N}, \quad \forall \mathbf{v} \in V \tag{2.2}$$

The notation $\mathbf{A} : \mathbf{B} = A_{ij}B_{ij}$ is the inner product of two second order tensors, with summation over repeated indices. Formally the variational form (2.2) reads: for $t > 0$ find $\mathbf{u} \in V$, such that

$$(\ddot{\mathbf{u}}, \mathbf{v}) + a(\mathbf{u}, \mathbf{v}) = l(\mathbf{v}), \quad \forall \mathbf{v} \in V \tag{2.3}$$

Which is an abstract weak form of the problem (2.1). Here $a(\mathbf{u}, \mathbf{v}) = (\boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\epsilon}(\mathbf{v}))$ is a a coercive bounded bilinear form. $l(\mathbf{v}) = (\mathbf{f}, \mathbf{v}) + (\mathbf{g}_N, \mathbf{v})_{\Gamma_N}$ is a linear form[6]. A linear form is mapping $V \to \mathbb{R}$ such that

1. $l(u + v) = l(u) + l(v)$
2. $l(xu) = xl(u), x \in \mathbb{R}$

for all $u, v \in V$ A bilinear form is mapping $V \times V \to \mathbb{R}$ such that

1. $a(u + v, w) = a(u, w) + a(v, w)$
2. $a(u, w + v) = a(u, w) + a(u, v)$
3. $a(xu, v) = xa(u, v), x \in \mathbb{R}$
4. $a(u, xv) = xa(u, v), x \in \mathbb{R}$

A bilinear form is said to be symmetric if $a(u, v) = a(v, u)$. If also $a(u, u) \geq 0$ with equality if and only if $u = 0$ then $a(\cdot, \cdot)$ defines a scalar product in V. Given the scalar product a norm can be defined by

$$|||\mathbf{v}|||_E = \sqrt{a(\mathbf{v}, \mathbf{v})} \tag{2.4}$$

This norm is called the energy norm. In this thesis the energy norm is denoted as $|||\cdot|||$, the Euclidean norm[2] is denoted as $\|\cdot\|$. Functions existing in L^2 are measured by the L^2-norm

$$\|\mathbf{v}\|_{L^2} = (\mathbf{v}, \mathbf{v})_{L^2} = \int_\Omega (\mathbf{v}\mathbf{v})^{1/2} dx \tag{2.5}$$

The L^2-norm is denoted $\|\cdot\|_{L^2}$ when used in this thesis.

It can be shown that a solution to the abstract weak problem exists and that it is unique via use of the Lax-Milgram lemma [6].

---

more information see [6] or any other textbook on FEM.

[2]Since the n-dimensional space $\mathcal{R}^n$ is called an *Euclidean space* the intuitive associated norm $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 \ldots + v_n^2}$ R is also known as the *Euclidean norm*

2.3 Finite Element Model

The idea of FEM is to approximate solutions of weak problems in a finite dimensional subspace $V^h \subset V$ of the corresponding infinite dimensional object. This subspace is constructed by partitioning the the domain Ω into elements like for instance triangles in 2 dimensions. Then the finite element approximation of (2.3) can be written: find $\mathbf{U} \in \mathcal{V}^h$ such that

$$(\ddot{\mathbf{U}}, \mathbf{v}) + a(\mathbf{U}, \mathbf{v}) = b(\mathbf{v}), \quad \forall \mathbf{v} \in \mathcal{V}^h \tag{2.6}$$

$\mathbf{U}$ exists in V^h so an ansatz can be made

$$\mathbf{U} = \sum_{j=1}^{N} \xi_j \boldsymbol{\varphi}_i \tag{2.7}$$

where N is the number of degrees of freedom. This together with the fact that $\mathbf{v}$ also exists in V^h and can be written as a linear combination of basis functions φ_i yields the equivalent problem: Find ξ_j, $j = 1, 2..N$ such that

$$\sum_{j=1}^{N} \ddot{\xi}_j(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i) + \sum_{j=1}^{N} \xi_j a(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i) = b(\boldsymbol{\varphi}_i), \quad i = 1, ..., N \tag{2.8}$$

which is the linear system of equations

$$\mathbf{M}\ddot{\boldsymbol{\xi}} + \mathbf{K}\boldsymbol{\xi} = \mathbf{b} \tag{2.9}$$

where $\mathbf{M}$ is the $N \times N$ matrix called the Mass matrix, including the density. $\mathbf{K}$ is the $N \times N$ matrix called the stiffness matrix and $\mathbf{b}$ is the $N \times 1$ vector called the load vector and. $\mathbf{u}$ is the time dependent $N \times 1$ vector containing the nodal coefficients, displacements in the elasticity case. This model has not included a damping, or friction, factor. Usually a damping factor is proportional to a velocity, therefore damping may be introduced into the finite element model in the form of Rayleigh damping [10]

$$\mathbf{C} = c_1\mathbf{M} + c_2\mathbf{K} \tag{2.10}$$

Rayleigh damping is a linear combination of the mass and stiffness matrices. By adding this to (2.9) the FEM is complete

$$\mathbf{M}\ddot{\boldsymbol{\xi}} + \mathbf{C}\dot{\boldsymbol{\xi}} + \mathbf{K}\boldsymbol{\xi} = \mathbf{b} \tag{2.11}$$

This system of second order ordinary differential equation's(ODE) need to be solved for every time-step of the simulation. A solver for first order systems will be used, therefore the system need to be reduced into first order ODE's

$$\mathbf{M}\dot{\boldsymbol{\eta}} + \mathbf{C}\boldsymbol{\eta} + \mathbf{K}\boldsymbol{\xi} = \mathbf{b} \tag{2.12a}$$

$$\boldsymbol{\eta} = \dot{\boldsymbol{\xi}} \tag{2.12b}$$

this system is, in this thesis, solved by the use of the Crank Nicholson time-stepping scheme.

The finite element model for the elastostatic problem is derived in the same way. The derivation is omitted but it is obvious that a model of elastostatics is obtained by omitting the time-dependent terms from equation (2.9). It is presented here because it is used to do the model reduction.

$$\mathbf{K}\boldsymbol{\xi} = \mathbf{b} \tag{2.13}$$

3 Model Reduction

The full FEM dimension, N, might be too large to be used practically. In those cases model reduction, also called model order reduction, is of great use. The goal is to find a subspace $V^{h,m} \subset V^h$, $m << N$, that is sufficiently small but still maintain adequate approximation properties of the original problem.

There are several ways of constructing a low dimensional subspace. In this thesis the method of reduction used is through projection, where the problem is projected onto a subspace, see figure 3.1. In this thesis the Proper Orthogonal Decomposition (POD) is used to find the basis of that subspace. The POD is widely used within fluid dynamics and to some extent within structural dynamics. One downside with the POD is that it does not handle local variations very well. When invoking the POD a set of so called modes are obtained. Modes are basis functions that ideally capture properties of the problem.

Reduction of the model of course introduces errors in the solution. To get control of the error a priori or a posteriori error estimations may be used[6]. An a priori error estimation can be used when the full solution is known beforehand, like for instance when an analytical solution can be obtained. However in general the full solution of a problem is not known and then an a posteriori error estimate is needed. For an a posteriori estimate the residual of the problem is used to estimate the error. The a posteriori error estimate is used in this thesis as an indicator to refine the reduction. Below is a short introduction to model reduction and then a summary of the POD and the POD snapshot method. After that two a posteriori error estimates for the elastostatic problem (2.13) is proposed and proved.

3.1 Reduced Model

The construction of the reduced model from the finite element model is analogous to the construction of the finite element model from the continuous model. Consider the finite

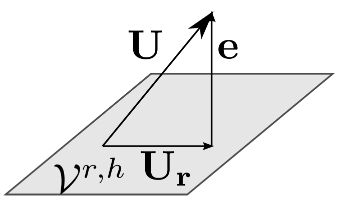

Figure 3.1: Projection of the full finite element solution $\mathbf{U}$ onto a subspace $\mathcal{V}^{r,h}$. Note that the error $\mathbf{e}$ is orthogonal to $\mathcal{V}^{r,h}$ and exists in the original space $\mathcal{V}^h$

element model (2.6). Then the reduced problem reads: Find the reduced displacements $\mathbf{U}^m \in \mathcal{V}^{h,m}$ such that

$$(\ddot{\mathbf{U}}^m, \mathbf{v}) + a(\mathbf{U}^m, \mathbf{v}) = b(\mathbf{v}), \quad \forall \mathbf{v} \in \mathcal{V}^{h,m} \tag{3.1}$$

Given a basis $\{\boldsymbol{\psi}_i\}_{i=1}^m$ in $V^{h,m}$ make the ansatz

$$\mathbf{U}^m = \sum_{j=1}^{m} \xi_j \boldsymbol{\psi}_j \tag{3.2}$$

which yield the $[m \times m]$ linear system of equations

$$\sum_{j=1}^{m} \ddot{\xi}_j (\boldsymbol{\psi}_j, \boldsymbol{\psi}_i) + \sum_{j=1}^{m} \xi_j a(\boldsymbol{\psi}_j, \boldsymbol{\psi}_i) = b(\boldsymbol{\psi}_i), \quad i = 1, 2, .., N \tag{3.3}$$

this together with the fact that $\boldsymbol{\psi}_i$ can be written $\boldsymbol{\psi}_i = \sum_{k=1}^{N} c_{i,k} \boldsymbol{\varphi}_k$, where $c_{i,k} \in \mathbb{R}$ makes it possible to write equation (3.1) like

$$\sum_{j=1}^{m} \ddot{\xi}_j \sum_{k=1}^{N} c_{i,k} \sum_{l=1}^{N} c_{i,l} (\boldsymbol{\varphi}_k, \boldsymbol{\varphi}_l) + \sum_{j=1}^{m} \xi_j \sum_{k=1}^{N} c_{i,k} \sum_{l=1}^{N} c_{i,l} a(\boldsymbol{\varphi}_k, \boldsymbol{\varphi}_l) = \sum_{l=1}^{N} c_{i,l} b(\boldsymbol{\varphi}_k), \quad i = 1, 2, .., N \tag{3.4}$$

By introducing the $[N \times m]$ coefficient matrix

$$\boldsymbol{W} = \begin{bmatrix} c_{1,1} \cdots c_{1,m} \\ \vdots \ddots \vdots \\ c_{N,1} \cdots c_{N,m} \end{bmatrix} \tag{3.5}$$

where c are the reduced basis coefficients. This is from now on referred to as a basis matrix. Inserting it into equation (3.4) the following system is obtained

$$\rho \boldsymbol{W^T M W} \ddot{\boldsymbol{\xi}} + \boldsymbol{W^T K W \xi} = \boldsymbol{W^T b} \tag{3.6}$$

which is the reduced version of equation (2.9). The reduced damping term is just the linear combination of the reduced mass and stiffness matrices. The reduced equation of the static solution is given by removing the time-dependent terms.

$$\boldsymbol{W^T K W \xi} = \boldsymbol{W^T b} \tag{3.7}$$

3.2 The Error of Reduction and the Static Residual

The total error in a reduced model can be described by

$$\mathbf{E} = (\mathbf{u} - \mathbf{U}) + (\mathbf{U} - \mathbf{U^m}) \tag{3.8}$$

where $\mathbf{E_d} = \mathbf{u} - \mathbf{U}$ represents a discretization error and $\mathbf{E_r} = \mathbf{U} - \mathbf{U^m}$ represents a reduction error. The discretization error arises when the continuous problem is discretized using finite elements and it is controlled by refining the mesh where needed. The reduction error arises as a result of the reduction of the finite full dimensional model. The reduction error is controlled by adding modes to the reduced basis and thus increasing the dimensionality of the reduced model. In this thesis the concern is focused only on the reduction error and it will henceforth be referred to as the error $\mathbf{e}$. By subtracting equation (3.1) from equation (2.6) the following expression for the error is obtained

$$\rho(\ddot{\mathbf{e}}, \mathbf{v}) + a(\mathbf{e}, \mathbf{v}) = \mathbf{0}, \quad \forall v \in V^{h,m} \tag{3.9}$$

which is known as Galerkin orthogonality. It states that if the form $\rho(\cdot,\cdot) + a(\cdot,\cdot)$ is considered an inner product then the error in the approximation is orthogonal to the reduced space $V^{h,m}$. In this way it is the best approximation possible [6].

The full solution is not always available to measure the error. In that case the error in a static load can be estimated based on the residual of the static problem (2.13)

$$\mathbf{R_i}(\mathbf{U^r}) = \mathbf{b_i} - \mathbf{K}\mathbf{U^r} \tag{3.10}$$

where $\mathbf{b_i}$ is load and $\mathbf{U^r}$ is the solution to corresponding reduced static problem. The residual will be used in the a posteriori estimate derived in section 3.4.

3.3 Proper Orthogonal Decomposition

The POD, sometimes referred to as Karhunen-Loeve transformation or Principal Component Analysis(PCA), is a method to construct low dimensional representation of high dimensional data. The idea is to create an orthonormal basis that spans a given set of data optimally with regard to the mean square error.

Assume there is an ensemble of functions, $U = \{\mathbf{u_k}\}$. Representations of $\mathbf{u_k}$ will be sought by projecting $\mathbf{u_k}$ onto a basis. Therefore assume that $\mathbf{u_k}$ belong to a Hilbert space V. The POD concerns finding a basis $\{\boldsymbol{\varphi}_j\} \subset V$ that is optimal in the sense that finite dimensional representations on the form

$$\mathbf{u_n}(x) = \sum_{j=1}^{n} (\mathbf{u}, \varphi_{\mathbf{j}})\varphi_{\mathbf{j}} \tag{3.11}$$

describe typical members $\mathbf{u} \in U$ better than representations, of the same dimension, of any other basis [7]. Optimality is formalized by claiming that the mean square error $\langle \|\mathbf{u} - \mathbf{u_N}\|^2 \rangle$ is minimized, $\langle \cdot \rangle$ denote an averaging operation. This leads to a variational eigenvalue problem

$$\langle (\mathbf{u}, \boldsymbol{\varphi})(\mathbf{u}, \mathbf{v}) \rangle = \lambda(\boldsymbol{\varphi}, \mathbf{v}) \tag{3.12}$$

for which the solution yield the sought basis. More details on this are found in the book [7]. Also it is shown in [2] that when V is a general separable Hilbert space the solution to the variational problem is a basis in that space that fulfils the optimality requirement.

This essentially means that it is easy to create a basis $\{\boldsymbol{\varphi}_j\}_{j=1}^m$ for the low dimensional space $V^{h,m} \subset V^h$ that will contain a good approximation of the original problem. When the POD is invoked on a data set with dimension $[N \times M]$ the result is M modes obtained, arranged by significance. This means that the first mode, e.g. the mode of lowest rank, is the most significant. So by including as many modes m as necessary a basis $\{\boldsymbol{\varphi}_j\}_{j=1}^m$ is the optimal basis of dimension m[1]. This calls for a definition of what *necessary* means. In this context it would be a relative measure of some quantity of the problem being preserved when invoking the POD as a reduction method.

3.3.1 Snapshot POD of Static Deformations

For transient processes $\mathbf{u}(\mathbf{x}, t)$ a widely used application of the POD would be to first collect M samples of the process from a full dimensional simulation. The samples could be, in the case of linear elasticity, the N nodal displacements that are the solution of (2.9). These samples are called snapshots and are arranged into a $[N \times M]$ matrix. By invoking the POD on the snapshot matrix a set of M so called POD-modes are obtained, together with singular values for each mode. These modes are used to create a new basis $V^{h,m} \subset V$, $m << N$ and used for a simulation. This method is commonly referred to as the snapshot POD.

Many applications run a short full scale, thus slow, simulation to capture all the possible, or probable, dynamics of the system and save the states of the system in the snapshot matrix. Another method could be to collect real experimental data in the snapshot matrix. Then the basis can be constructed and used for the reduced simulation. However this might not be desirable as it might be too time consuming to run a full scale simulation or hard to obtain real and accurate measurements. Also there is no known method of choosing snapshots that are considered good. Usually snapshots are chosen in an arbitrary way by for instance saving every, or every other time step of a simulation. There are methods as in [1] that aim at getting the snapshots evenly spaced in state-space by using a k-means clustering technique.

The snapshot POD method is used in this thesis, but the snapshots are captured in a different way. A data set of static load deformations is created by first setting up a load matrix $\mathbf{B}$. The columns $\mathbf{b}_i$ are the different load cases that act on the body, i.e. traction loads along the teeth of a gear wheel. Then the static load deformations $\mathbf{u}_i$ are determined by solving the static linear equations of elasticity for each load case $\mathbf{b}_i$. By collecting all the static deformations $\mathbf{u}_i$ a $[N \times M]$ matrix $\mathbf{U}$ is obtained and is the snapshot matrix of static deformations. By invoking the POD on this matrix a reduced basis can be constructed that is based on the possible static deformations modes of the body. This method differs in one fundamental way from the method above, in that a full scale simulation, or experimental measurement, is not needed to be done. Instead the method of static load cases requires good knowledge of the forces that act on the body.

In [2] the eigenvalue problem of a snapshot POD with static load cases is derived

$$\frac{1}{M}\mathbf{U}^T\mathbf{Z}\mathbf{U}\mathbf{c} = \boldsymbol{\Sigma}\mathbf{c}, \tag{3.13}$$

[1] In the sense argued around (3.11)

where $\mathbf{\Sigma} = [\sigma_{ii}]$ is a diagonal matrix of singular values and $\mathbf{Z}$ is a matrix used to base the inner product on, for instance by setting Z as the identity matrix the POD is based on the Euclidean inner product. If Z is set to the stiffness matrix $\mathbf{K}$ or the mass matrix $\mathbf{M}$ then the POD is based on the energy inner product or the L^2 inner product respectively.

3.3.2 Optimality and Preservation of Properties

In [7] it is shown that the first n POD basis functions capture more energy on average than the first n basis functions of any other basis. This is however based on the fact that $\mathbf{u}_i$ are velocities and then the singular values translates to kinetic energy content of their corresponding mode. In the case of deformations the energy analogy is lost, however in [2] it is shown that the energy analogy can be retained by basing the POD on the energy inner product. In that case the quantity preserved is the strain energy. The linear strain energy in elastic displacement $\mathbf{u} \in H^1(\mathbf{\Omega})^3$ is given by the energy functional

$$S(\mathbf{u}) = \frac{1}{2}a(\mathbf{u}, \mathbf{u}) \tag{3.14}$$

where $a(v, v)$ is the energy inner product. The total strain energy in the projection is then given by

$$\langle S(\mathbf{U}_n)\rangle = \frac{1}{2}\langle a(U_n, U_n)\rangle = \frac{1}{2}\sum_j \sigma_j \tag{3.15}$$

where U_n is from equation (3.11). This shows that the singular values are proportional to the energy content in the corresponding POD mode. This makes it possible to state a measurement for the relative energy preservation in a projection

$$Q = \frac{\sum\limits_{j=1}^{n} \sigma_j}{\sum\limits_j \sigma_j}. \tag{3.16}$$

It is important to remember that this holds, for elastic deformations, as an average energy preservation if the POD is based on the energy inner product. If some other inner product,i.e. the Euclidean, is used the energy analogy is again lost. However it can be argued that this quantity still is an indicator of how well a POD basis captures the properties of the problem but there cannot be any conclusion drawn based upon physical quantities.

3.4 A Posteriori Error Estimation

Two a posteriori error estimates are proved. The first theorem provides a way to find the norm of the error exactly and the second estimates it. The first theorem make use of the Cholesky factorization of $\mathbf{A}$. $\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$ where $\mathbf{L}$ is a lower triangular matrix.

Also in both theorems, since the static residual is used, the matrix $\mathbf{A}$ corresponds to the stiffness matrix, making this an error estimate in the energy norm[2].

Theorem 1 *Let* $\mathbf{R}(\mathbf{U}^r)$ *be the static residual* (3.10). *If* $\mathbf{A}$ *a* $[N \times N]$ *symmetric positive definite matrix and* $\mathbf{L}$ *is the square root of the stiffness matrix* $\mathbf{A}$ *then the error* $\mathbf{e}$ *is bound:* $|||e||| = \|\mathbf{L^{-1}R(U^r)}\|$

Proof. The proof starts by first showing that $|||\mathbf{e}||| \leq \|\mathbf{L}^{-1}\mathbf{R}(\mathbf{U^r})\|$ and then continue by proving that in fact $|||\mathbf{e}||| = \mathbf{L^{-1}R(U^r)}$

$$
\begin{aligned}
|||\mathbf{e}|||^2 =&\mathbf{e^T A e} && (3.17a)\\
=&\mathbf{e^T A(U - U^r)} && (3.17b)\\
=&\mathbf{e^T R(U^r)} && (3.17c)\\
=&\mathbf{e^T L L}^{-1}\mathbf{R(U^r)} && (3.17d)\\
\leq&\|\mathbf{e^T L}\|\|\mathbf{L}^{-1}\mathbf{R(U^r)}\| && (3.17e)\\
=&(\mathbf{e^T L L^T e})^{1/2}\|\mathbf{L}^{-1}\mathbf{R(U^r)}\| && (3.17f)\\
=&|||\mathbf{e}|||\|\mathbf{L}^{-1}\mathbf{R(U^r)}\| && (3.17g)\\
=&\|\mathbf{L}^{-1}\mathbf{R(U^r)}\| && (3.17h)
\end{aligned}
$$

which concludes the first part. The second part shows that $|||\mathbf{e}||| = \|\mathbf{L}^{-1}\mathbf{R}(\mathbf{U}^r)\|$

$$
\begin{aligned}
\|\mathbf{L}^{-1}\mathbf{R(U)}\| =&\|\mathbf{L}^{-1}\mathbf{L L^T (U - U^r)}\| && (3.18a)\\
=&\|\mathbf{L^T e}\| && (3.18b)\\
=&(\mathbf{e^T L L^T e})^{1/2} = |||\mathbf{e}||| && (3.18c)
\end{aligned}
$$

which concludes the proof. ■

The first part make use of the Cholesky factorization and Cauchy-Schwarz inequality and then the norm is explicitly calculated. The second part shows that there is in fact equality with the error in the theorem when the equation is solved.

Theorem 2 *Let* $\mathbf{R(U^r)}$ *be the static residual* (3.10). *If* $\mathbf{A}$ *is a* $[N \times N]$ *symmetric positive definite matrix and* $\boldsymbol{\lambda} = [\lambda_o, \lambda_1 ... \lambda_N]$ *are the eigenvalues of* $\mathbf{A}$ *then the error is bound by:* $|||e||| \leq \frac{1}{\lambda_{min}^{1/2}}\mathbf{R(U^r)}$

Proof. As in the above proof assume that $\mathbf{L}$ is the square root of $\mathbf{A}$. The proof starts at the end of part 1 in the proof above.

$$
\begin{aligned}
\|\mathbf{L}^{-1}\mathbf{R(U)}\| =&\|\mathbf{A}^{-1/2}\mathbf{R(U^r)}\| && (3.19a)\\
\leq&\|\mathbf{A}^{-1/2}\|\|\mathbf{R(U^r)}\| && (3.19b)\\
=&\frac{1}{\lambda_{min}^{1/2}}\|\mathbf{R(U^r)}\| && (3.19c)
\end{aligned}
$$

[2]The energy norm is defined as $|||\mathbf{v}||| = (\mathbf{v^T A v})^{1/2}$ where $\mathbf{A}$ is the positive definite stiffness matrix. The regular 2-norm of a matrix is denoted $\|\cdot\|$.

■

An explanation of the above proof may be needed. The first step is simply Cauchy-Schwarz inequality. The second step makes use of the fact that $\mathbf{A}^{-1}$ is symmetric and can be eigen-decomposed. For a symmetric N by N matrix the following holds:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} \tag{3.20a}$$

$$\mathbf{A}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^{-1} \tag{3.20b}$$

$$[\Lambda]_{ii}^{-1} = \frac{1}{\lambda_i} \tag{3.20c}$$

where $\mathbf{Q}$ is a square matrix N by N that contain the eigenvectors column-wise, $\mathbf{\Lambda}$ is the diagonal matrix with N eigenvalues λ as entries. Also for the Euclidean norm of a symmetric matrix it holds that:

$$\|\mathbf{A}\| = |\lambda_{max}| \tag{3.21}$$

where λ_{max} is the greatest eigenvalue of $\mathbf{A}$. Thus the greatest eigenvalue of $\mathbf{A}^{-1}$ is the smallest eigenvalue of $\mathbf{A}$.

4 Method of Basis Enrichment using Hierarchical POD

In this chapter a hierarchical method to enrich a regular POD model reduction method is presented. First there is a general outline then it follows by more detailed explanations of the parts involved in the algorithm.

POD-modes are used to create a basis which spans a reduced space, $V^{r,h} \subset V^h$. Normally a snapshot POD captures global behaviour to a high degree and the local behaviour gets lost. For instance in a gearwheel the contact between two gears occur in a point in the $2D$ case, and in a line in the $3D$ case. Due to this contact there will be a smooth large scale deformation of the gearwheel and sharp small scale deformations that are localized around the point of contact. The large scale and small scale deformations are henceforth called global and local behaviour respectively. Much of the local behaviour disappear when a POD based model reduction is used, which can be seen in figure 4.1. This might not be desirable. For instance if it is important to measure the stress in a gear tooth then a regular POD based model reduction is not recommended as stresses are highly dependant on the local behaviour. Then the problem is to find a basis that captures both global and local behaviour. To this end the snapshot POD is used in a hierarchical setting. In this case a snapshot is the same as the static solution to a load case. A central idea is that the point of contact will be localized. Therefore it should be possible to capture the local variations in a small number of enrichment modes.

First POD-modes are obtained from all snapshots and thus capture the global behaviour. The local behaviour corresponds to the error from the reduction, e.g the information in each snap shot that was not captured by the reduction, see fig 4.2 and fig 3.1. Therefore the error in the reduced model is studied to determine which snapshots that loose too much information in the reduction. These snapshots are then clustered to find groups of snapshots to create an enriched basis for. More on the clustering is found in section 4.3. The error in the clusters are used as a new set of snapshots on which the POD is invoked. This set of POD modes capture the overall behaviour within a given cluster. The behaviour in a cluster will then correspond to local behaviour. Thus a set of regular POD modes, $\mathbf{W}_g = \{\boldsymbol{\varphi}_j^g\}_{j=1}^{m_g}$, capturing global behaviour and a set of enrichment modes,$\mathbf{W}_l = \{\boldsymbol{\varphi}_j^l\}_{j=1}^{m_l}$, capturing local behaviour are obtained. m_g and m_l denote the number of global and enrichment modes.

Appending the enrichment modes onto the regular modes yields an enriched basis that can be used to reduce the model and still capture both global and local behaviour, within the snapshots forming the cluster used for enrichment. It is important to clarify that the enrichment modes still have global support on the domain.

The enrichment process is done in an offline phase of the simulation, therefore it is not

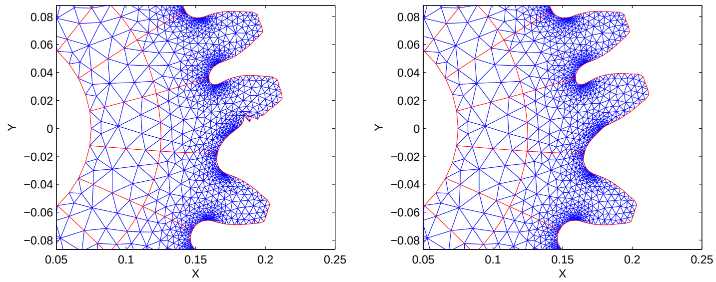

Figure 4.1: Shows the solution after ten time-steps of a gear that's been subjected to a sliding contact. The left figure shows the full finite element solution and the right figure shows the solution when the model has been reduced using regular POD. In both cases the gear is bent but in the reduced version the local variations, especially near the point of contact disappear.

of priority to make it as fast as possible. However there will be changes of basis during the online phase. The change of basis means that the equations need to be re-projected, which is an expensive operation.

4.1 Algorithm for Basis Enrichment

The POD reduction method for static load cases uses the solutions of all the predefined static load cases as a snapshot matrix. Thus each snapshot represents the static solution to a specific load. It is then possible to identify snapshots that does not get well represented with the global reduction. This is done by using one of the a posteriori estimates from section 3.4 as an indicator. Provided the load cases have been ordered in a chronological way, there will appear to be clusters in the indicator, representing sequences of snapshots that need to be refined, illustrated in figure 4.4. When the snapshots needing refinement have been identified the error in those snapshots is used to create local enrichment POD-modes that capture the error. Appending the enrichment modes to the global modes yield the enriched basis. In addition a scheme describing when the clusters should be activated during runtime is created in the same process. This is a hierarchical enriched model reduction method and it is described in algorithm 1.

In the on-line phase, during runtime, of the simulation there will be instances when the reduced space changes. This occurs when clusters are activated and deactivated. This requires the equations to be re-projected which is a costly operation. Also the solution from the previous time-step needs to be re-projected. Due to the cost of the operation and the fact that re-projection introduces errors the implementation must se to it that there are not too many cluster changes. The damping of the model will neutralize error arising from cluster change within a certain amount of time steps, therefore the damping should decide the minimum amount of time steps that need to pass before a cluster change can be allowed.

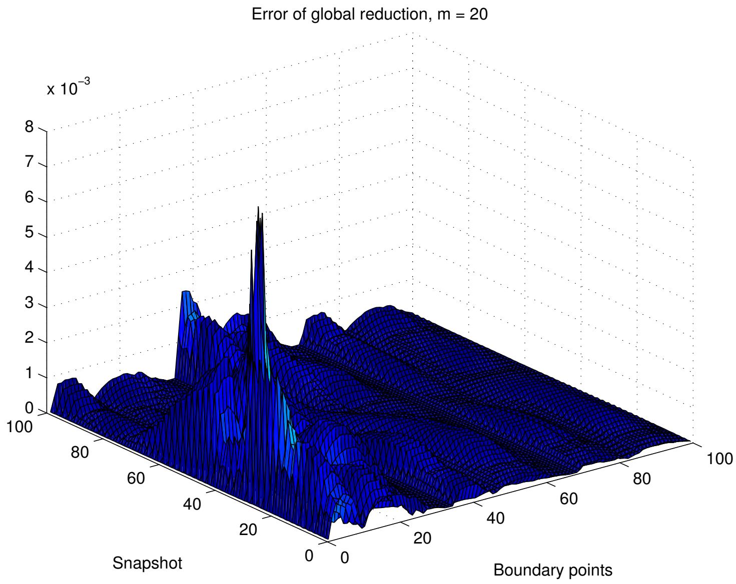

Figure 4.2: Example vector field of the error in the first 100 snapshots. The domain is in this case only represented by the first 100 boundary nodes. The reduction is based on 20 global modes.

Algorithm 1

1: Define a set of appropriate load cases $\mathbf{B}$.
2: Solve for the static deformation modes $\mathbf{U} = \mathbf{K^{-1}B}$.
3: Generate modes $\{\boldsymbol{\varphi}\}_i$ by invoking POD on the static deformations $\mathbf{U}$.
4: Identify and select the m_g global modes with largest singular values and create a new global basis $\mathbf{W}_g = \{\boldsymbol{\varphi}\}_{i=1}^{m_g}$.
5: Construct a reduced representation, $\mathbf{U^r}$, of the statics solution using equation (3.7). Then obtain the error snapshot matrix $\mathbf{E}$ by removing the global representation from the original snapshot matrix $\mathbf{E} = \mathbf{U} - \mathbf{U}^r$
6: Use a clustering algorithm to determine which snapshots in $\mathbf{E}$ that correspond to a local variation that was not captured by the global approximation.
7: Create new local snapshot matrices $\mathbf{E}_{l_i}$ for each cluster i.
8: Invoke POD on each local cluster to generate local POD modes $\boldsymbol{\varphi}_{l_i}$ with global domain support.
9: Create an enriched basis $\mathbf{W}_e$ by selecting appropriate modes from $\boldsymbol{\varphi}_{l_i}$ and appending them to $\mathbf{W}_g$
10: Use the enriched basis in transient simulation

Basically algorithm 1 consists of five parts: Defining load cases, finding the global modes, clustering the indicator values, selecting the local enrichment modes. Finally use the information to create locally enriched bases and a scheme that change basis in a transient simulation. Assuming the load cases are known the algorithm takes three parameters: Static load cases $\mathbf{B}$, relative tolerance for error clustering ζ and an approximative tolerance, η, determining how many modes are included in the global base. To make the method as generic as possible it is important to find a method to identify global modes that works. It is of even greater importance that correct clusters of snapshots are found with the indicator.

4.2 Identifying Global Modes

In this thesis it is argued that there should be a transition between global and local modes in a POD. This argument is based on the fact that the lowest order modes from a POD represents global behaviour and the local behaviour is captured within the high order modes. To identify a transition between global and local behaviour the property conservation, Q, in equation (3.16) might be used. This is however not a complete indication of the transition, since it requires a knowledge of how much of the quantity is contained within global behaviour.

In this thesis the argument is that an indicator of a transition between global and local modes can be seen as the points where there are significant change in the singular values. I.e. points where the derivative of the singular values are large in comparison to its neighbours. This is illustrated in figure 4.3. Of course there will be large derivatives for the first few singular values. Therefore it is suggested to use this indicator together with Q to find a transition. A basic algorithm is stated below.

Algorithm 2

1: Find the number of modes, m, satisfying $Q \geq \eta$
2: Look for large derivatives in the singular values beyond mode m
3: IF a mode found: set it as the last global mode, m_g
4: ELSE: set $m_g = m$

This routine includes a vague notion of the derivative being large that needs to be specified during implementation. Also Q needs to be large enough too, however now only an estimate is needed. In figure 4.3 a distinct drop in significance is shown around mode 20 and in table 4.1 it is stated that more than 99% of the models properties are captured in 20 global modes. It should be mentioned that there might not exist a distinct transition between global and local modes. In that case Q will determine where to introduce a cut off for global modes.

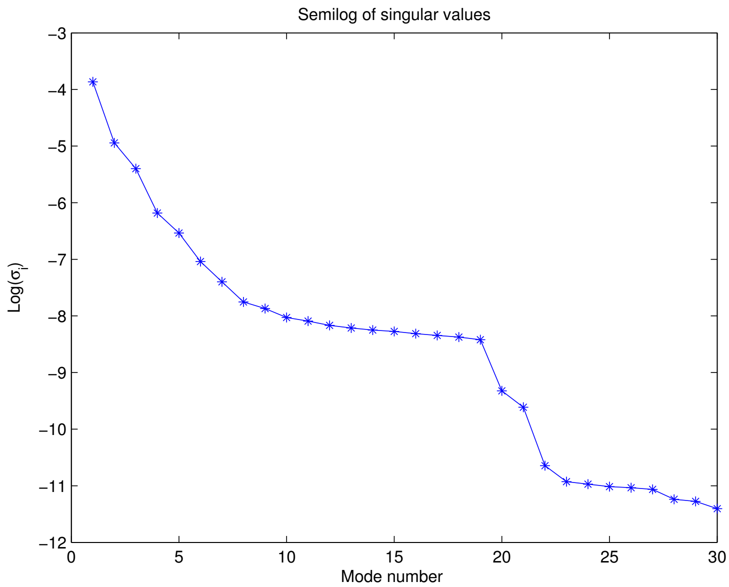

Figure 4.3: Semi-log of the singular values for the first 30 modes, illustrates a distinct drop in significance between modes 19 and 22.

Table 4.1: The quantity Q for different sizes of the basis

n	Q
1	0.5049
5	0.8912
10	0.9408
15	0.9725
20	0.9929

4.3 Clustering

In this implementation the pre-defined load cases are arranged chronologically. The a posteriori estimate (2) is used as an indicator, $\boldsymbol{\nu}$, to find clusters of snapshots that are to be refined using enrichment. To this end the max-norm of $\boldsymbol{\nu}$ is used together with a tolerance parameter, ζ. Then it is determined that the $i:th$ snapshot need refinement if the following inequality is satisfied $\nu_i > \zeta\|\boldsymbol{\nu}\|_{Max}$. In the implementation the inequality is rewritten as $\frac{\nu_i}{\|\boldsymbol{\nu}\|_{Max}} > \zeta$ where the left-hand side is called the relative indicator and the right hand side the tolerance. In figure 4.4 an example of the relative indicator together with a tolerance is shown.

If there is a sequence that is long enough consisting of snapshots needing refinement then a cluster is found corresponding to that sequence. The notion of a sequence being wide enough is connected to the fact that a cluster cannot be too small, since this would mean changing clusters too often in the simulation.

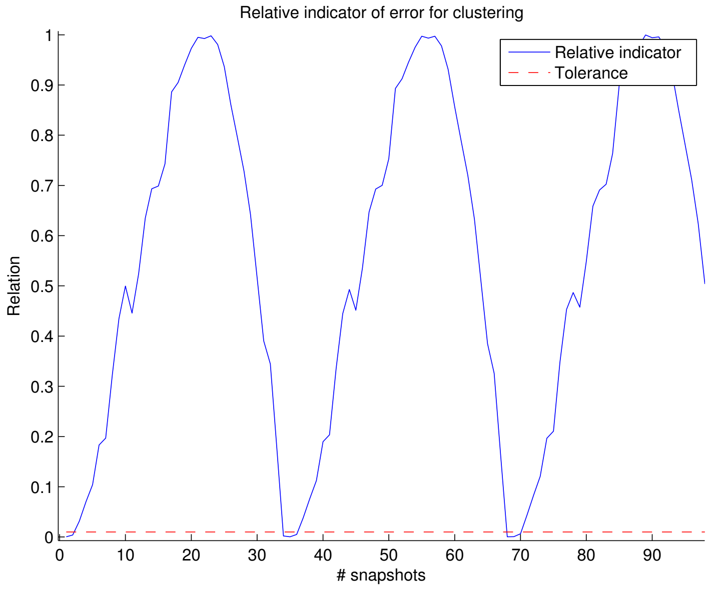

Figure 4.4: The relative error indicator and the tolerance ζ of the first 100 snapshots. 20 global modes and 10 enrichment modes are used and the tolerance is set to 0.01, meaning the error indicator in one snapshot should not exceed 1 percent of the max error. Clusters appear as the snapshots with error greater than the tolerance.

In this implementation there will be snapshots in between clusters that do not need refinement, depending on the tolerance. This causes a problem of small clusters that only contain global modes. The implementation is chosen so that the clusters are extended towards each other so that there are no snapshots in between. In other words there will always be and enriched basis active. One could chose to use non enriched bases when the enrichment is not needed however in this case the clusters containing only global modes would be so small that there would be a lot of re-projections of the equations going on.

Also if for some reason the load cases are not ordered in a chronological order one simply needs to sort them before applying the simple linear clustering technique described.

4.4 Transient Simulation

When a new basis is to be used the system of equations need to be re-projected. The new reduced form of the mass and stiffness matrices are obtained by projecting the original matrices onto the new basis as described in section 3.1. However to solve the reduced system (3.6) in time-step $t = t_n$ requires the solution, from $t = t_{n-1}$. Thus the solution vector needs to be projected from the current space to the next space being used. Let $\boldsymbol{\xi}^a$

be the solution vector at time-step $t = t_n$ belonging to the space $V^{a,r}$ and let $\boldsymbol{\xi}^b$ be the solution vector at the same time-step belonging to the space $V^{b,r}$. Also let $\mathbf{W}^a$ and $\mathbf{W}^b$ denote the basis matrices[1] that span the spaces $V^{a,r}$ and $V^{b,r}$ respectively. Equation (3.2) states the full finite dimensional representation, $\boldsymbol{U}^r$ of the reduced solution. Then the full dimensional solution for $\boldsymbol{\xi}^a$ is found

$$\mathbf{U}^{r,a} = \mathbf{W}^a \boldsymbol{\xi}^a. \tag{4.1}$$

When changing basis the the full dimensional solution should approximately be the same when originating from either $\boldsymbol{\xi}^a$ or $\boldsymbol{\xi}^b$. Then the problem becomes:

$$\mathbf{W}^b \boldsymbol{\xi}^b \approx \mathbf{W}^a \boldsymbol{\xi}^a. \tag{4.2}$$

Projecting onto the second subspace and solving for $\boldsymbol{\xi}^b$ yields

$$\boldsymbol{\xi}^b \approx ((\mathbf{W}^b)^T \mathbf{W}^b)^{-1} (\mathbf{W}^b)^T \mathbf{W}^a \boldsymbol{\xi}^a \tag{4.3}$$

It is intuitive that there is an approximative relation in the above equation since there is projection done with reduced bases, meaning that there will be reduction errors. Therefore cluster transitions should occur when the displacements are as small as possible, making the error small. There might be ways to overcome this error or at least dampen the effect of it by introducing overlapping clusters or activating several clusters at the same time.

[1] A basis matrix was introduced in section 3.1 as the matrix of the mode coefficients of the basis

5 A Gearwheel Model for Validation

As stated in the introduction the problem concerns finding a method that handle sliding contacts. To this end a gearwheel model problem is set up in Matlab R2011a. In the first section the gearwheel model is presented. In the second section the simulations run to collect data and comparisons made are presented.

5.1 The Gearwheel

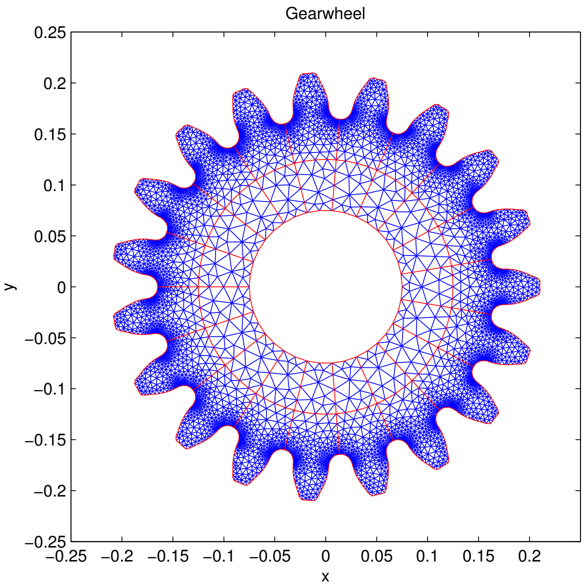

Figure 5.1: The gearwheel mesh used to solve the FEM problem on

The gearwheel domain is a triangular mesh of a gearwheel contour consisting of 19 gear teeth. The inner boundary is considered clamped by the method of penalty in the stiffness matrix. The body-force is $\mathbf{f} = \mathbf{0}$. Then a set of traction loads are applied to the gears in such a way that it would simulate the gear rotating counter-clockwise and being in constant contact with another inelastic gearwheel. Since it is a $2D$ problem the contact is always in one point. The mesh will not rotate to simulate the gears moving, instead the point of contact will be sliding over the boundary. In figure 5.2 the forces defining the traction loads are visualized on the boundary of the three topmost gear teeth. During a transient simulation the forces will act in sequence from right to left.

The forces used are in not the result of an accurate contact simulation. Instead they are artificially constructed. The direction of the force is normal to the boundary in the point of contact. It is computed as a mean value of the normals to the boundary segments neighbouring the point. The magnitude of the force is smallest at the bottom and top of each tooth and largest at the middle of the tooth. This is a crude model but suffices for validation of the enriched reduction method. During the dynamical simulation the forces are not dynamic, instead the sliding contact is simply simulated as a set of static forces being applied in a chronological order. This essentially means that a force in a point of contact does not depend on any prior deformation in its neighbouring points. To do real force dynamics would involve solving a contact problem. All in all the model consists of approx. 650 load cases, meaning that the POD will create approx. 650 modes ordered by significance. The finite element model of the elastic gearwheel is presented in chapter 2 and damping is included. The mesh consists of around 8000 points, resulting in around 16000 degrees of freedom.

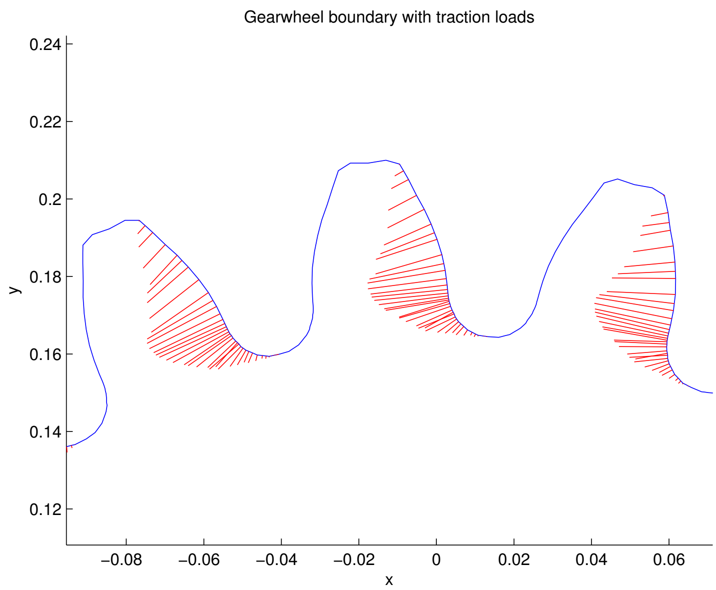

Figure 5.2: The boundary of the three topmost gear teeth together with the applied traction forces

5.2 Numerical Studies

Several simulations are run to get results and validate the enriched POD based reduction method. First a full finite element simulation is run to determine a solution that is considered true. The aim with most simulations is to compare the error of the enriched POD based reduction with the error of a reference POD based reduction method. In this case a reference method means that it should consist of only regular, e.g. global, POD-modes and the basis should be of the same dimension as the enriched basis. In this section the numerical studies and their results are presented. Conclusions on the results are found in the next chapter.

Combinations of Regular and Enrichment Modes

There are many combinations of global and local modes that add up to a basis of the same dimension. In chapter 4 a method of choosing a good global mode number is presented and it is tested here.

The runs are made with the global modes varying from 1 to 30 and local modes varying from 0 to 20. The runs are made over 200 time-steps, involving the sliding of the contact over 6 gears. A time-mean error of the displacements is measured in the L^2-norm and illustrated as a surface, see figure 5.3.

Displacements over Time

The time mean value of displacements do not really indicate how good the approximation is, only that it is good on average. In order to get an idea of how the enriched reduction method behaves compared to a reference reduction the absolute error is measured for a specific combination of global and local modes and the error in displacement is shown over time, figure 5.4.

Von Mises Stress

The reason for enrichment was to find a way to preserve local quantities. The von Mises stress is such a quantity. It is used as a way to measure stress. It is usually mentioned in connection with deformation and plasticity. More about von Mises stress can be found in the book [10]. The von Mises stress is measured and visualized at a specific time-step when the sliding contact is in the middle of a gear tooth. Figures 5.5, 5.6 and 5.7 and 5.8 show the results. Note that the mesh is refined to get a higher resolution, roughly the number of modes are doubled.

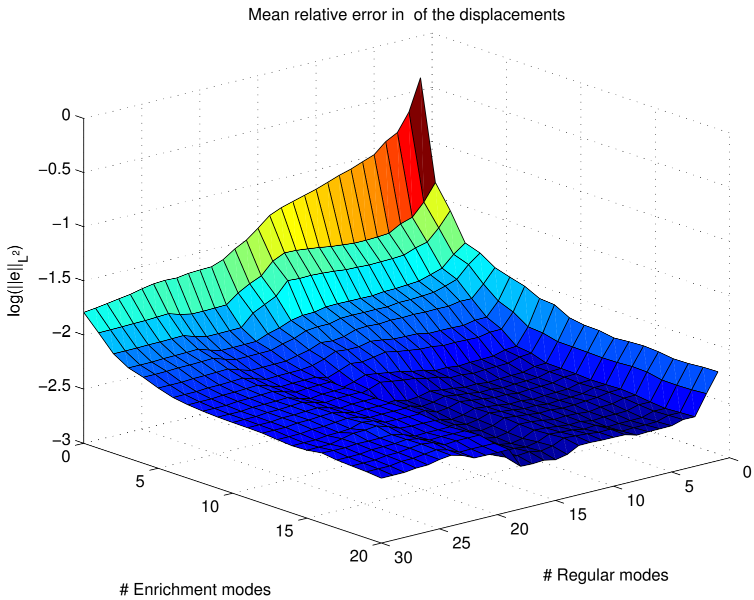

Figure 5.3: Semi-log of the time-mean relative error of solutions with the number of global modes varying between 1 and 30 and the local number of modes varying between 0 and 20.

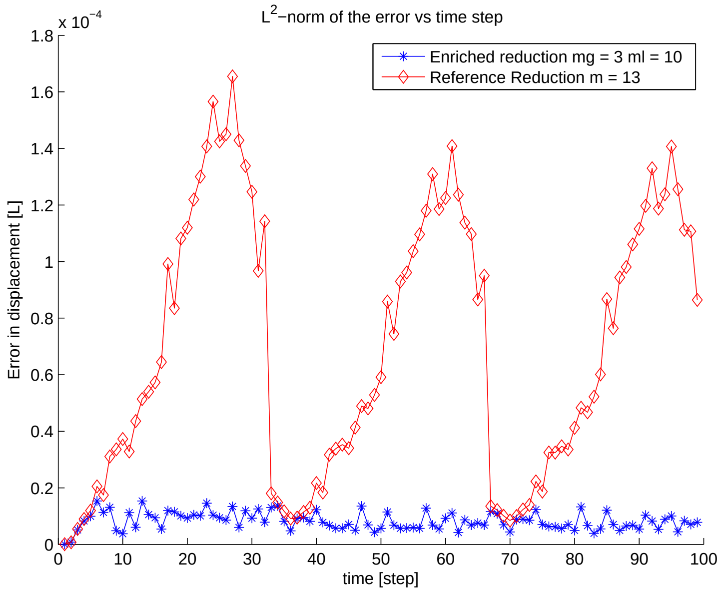

Figure 5.4: L^2-norm of the absolute error in displacement for the enriched method and a reference method. Using 20 global modes and 10 local modes. First 100 time-steps, almost covering the contact sliding over three gears.

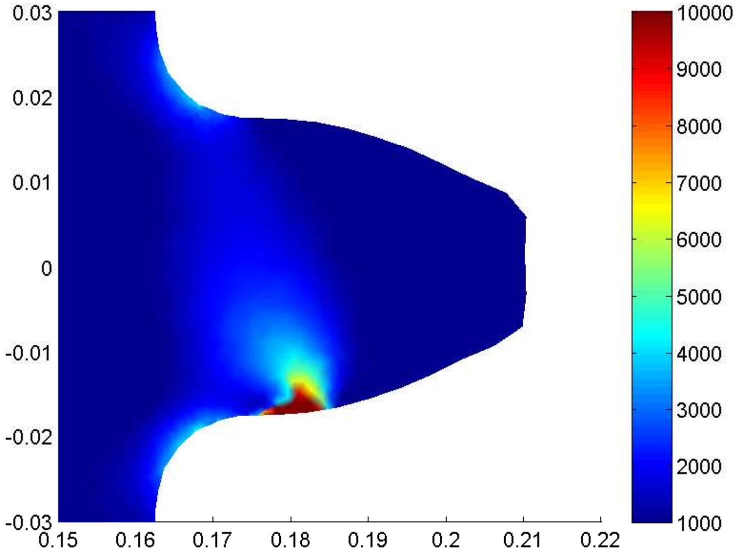

Figure 5.5: The von Mises stress for the full finite element solution.

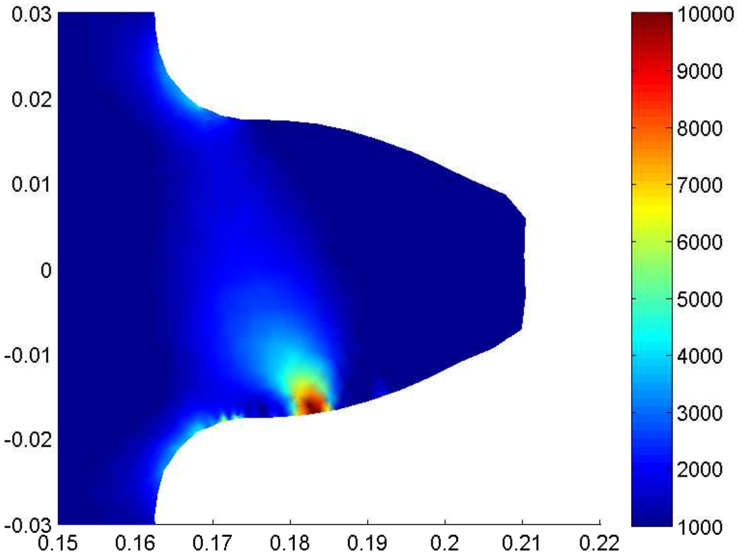

Figure 5.6: The von Mises stress for the enriched reduction method. Using 3 global modes and 20 local modes. Corresponding to around 30% of the enrichment modes available in this cluster.

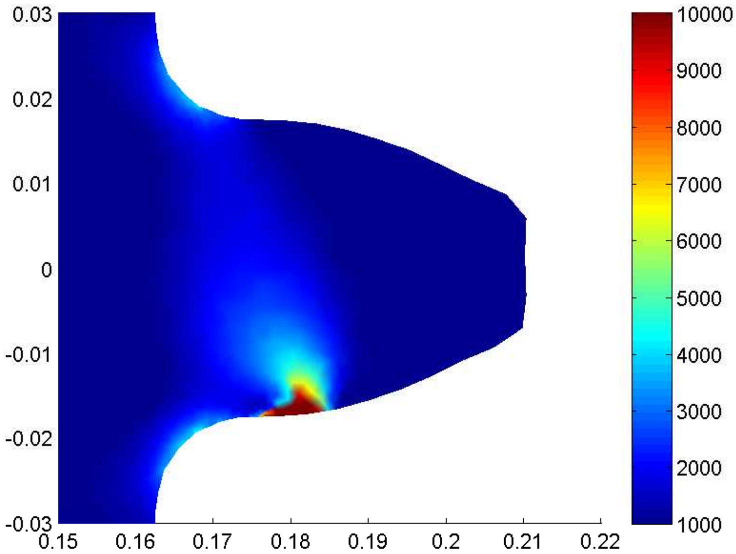

Figure 5.7: The von Mises stress for the enriched reduction method. Using 3 global modes and 45 local modes. Corresponding to around 66% percent of the enrichment modes available in this cluster.

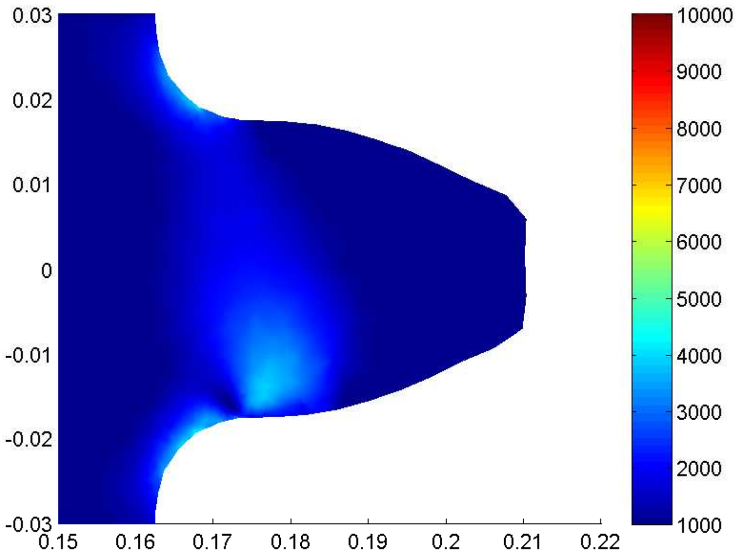

Figure 5.8: The von Mises stress for the reference solution, using 48 regular POD modes.

6 Conclusions

A method to enrich a regular POD based reduction method with local POD modes has been developed, implemented and validated. An a posteriori error estimate for the static load case has been stated and proved. A numerical study is done on several combinations of global and local modes showing faster convergence of the error when using a hierarchical enriched POD. It is shown numerically that the enriched method captures local variations, such as von Mises stress, to a much greater degree than a reference method, as was expected. More detail of conclusions and issues of concern and possible solutions to them together with suggested future work is presented below.

6.1 Transition from Global to Local Behaviour

The discussion in this section concerns the results shown in figure 5.3. From this it is evident that introducing local clusters and increasing the number of local modes in them increases the precision of the reduction method to a higher degree than simply adding global modes. It was theorized that a threshold for global modes could be found by looking at table 4.1 for property conservation[1], together with the sharp decrease of the magnitude of the singular values around mode number $19-22$ illustrated by figure 4.3. However when using enrichment it is clear from figure 5.3 that a lot fewer global modes might be needed. In fact it seems that it would suffice with three global modes and as little as ten local modes to get very good results as a time average.

This indicates that the original assumption about global threshold is incorrect. By looking at the shape of the first 6 modes, figure 6.1, one can see that the first mode concerns rotation and the second and third translation of the entire gearwheel. The modes 4-6 and 7-22, figure 6.2, show modes that affect the gear teeth. There is one mode associated to each gear tooth but this is not evident by inspection. These modes are L^2-orthogonal. One important conclusion here is that to get all gear teeth to behave nicely with a general POD reduction all the 22 modes are needed. 3 for the overall behaviour and one for each of the 19 gears. On the other hand the enriched reduction method get a very good average error when using as little as 13 modes, in other words less that is actually needed to get a working method with a regular POD. This is a great result. As of this result the numerical study of the displacements over time, and stresses, is carried out with 3 global modes and 10 enrichment modes as a first step.

The average error increases when more than 20 global modes are added. This is attributed to a systematic error in the clustering regime, where the transition between

[1]Table 4.1 states that 20 modes should preserve 99.3% of the properties

clusters are not set correctly. This fault generates a spike in the error when there is a change of basis, for combinations with more than 20 global modes.

6.2 Behaviour, Change of Clusters and Preservation of Local Properties

This section concerns the results and conclusions around the simulations run to test the behaviour of the error of displacement, and the meassurement of stress. In figure 5.4 it can be seen that the enriched solution seem to generate solutions with much smaller errors in displacement than the reference method. It was said in section 4.4 that the change of clusters would introduce errors. In figure 5.4 the cluster changes takes place at time-steps 36 and 70. It is not evident from this study that there are any sudden spikes in the error, which is good.

The preservation of local properties, in this case the von Mises stress, is illustrated in a good way in figures 5.5, 5.6, 5.7 and 5.8. Notice that the reference solution seem to capture some behaviour quite well. That is the stresses going through the gear tooth and the stresses in the valleys. The enriched method does however capture the sharper local deformations, especially near the point of contact, much better. Note how much larger the stresses are that are captured by the enriched method. Also note that there appear to be some artificial stresses around the valley in the enriched method in figure 5.6. It has been observed that the artificial stresses also arise in a regular POD based reduction. They disappear if more modes are added. It has also been observed that they disappear in the enriched method if just a few more modes are used in the enrichment, as in figure 5.7. This is good and shows that the artificial stresses are part of the POD and not an effect due to the enrichment.

To summarize the enriched method behaves well and captures local variations very nicely using a lot of fewer modes than a regular reduction method. The enrichment method gives a reduction that is around 10 times more accurate than a conventional POD reduction with the same dimension. This will of course lead to computational efficiency increasing with at least a factor of 10. The use of this reduction method seems very appealing in industrial software and will be implemented and tested i SKF's software BEAST.

6.3 Future Work

There are many things that can be done to continue the work on this subject, a few are listed.

- Investigate problems arising when changing the basis, such as the numerical error in the projection.
- The enriched and global parts are currently not orthogonal to each other. Suggested solutions are to orthogonalize the the basis or to solve the global and local parts one at a time and then adding the solutions together.

- Study the method when damping is decreased to investigate how vibrations in previously loaded gears are captured.
- Implementation in SKF's software BEAST to test the enrichment method on better constructed models in $3D$.
- Research other areas of implementation. How can the technique be generalized to cover other applications or even multiphsyic problems?
- implement the a posteriori indicator to automatically add enough enrichment modes to keep the error below a certain tolerance in the clusters.
- Implement generic clustering and test it on a generic irregular domain. Also more levels of hierarchy should be implemented to gain further efficiency.

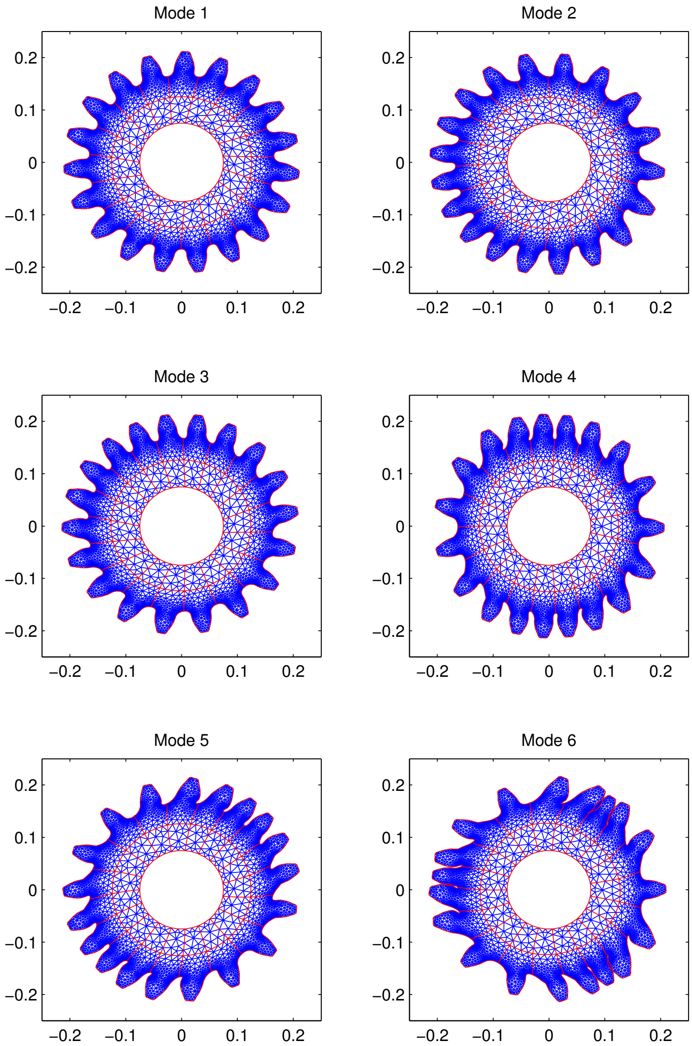

Figure 6.1: Visualisation of the first 6 POD modes. Indicating that the three first POD modes might concern the rotation and translation of the whole of the gearwheel. The mode displacements are exaggerated.

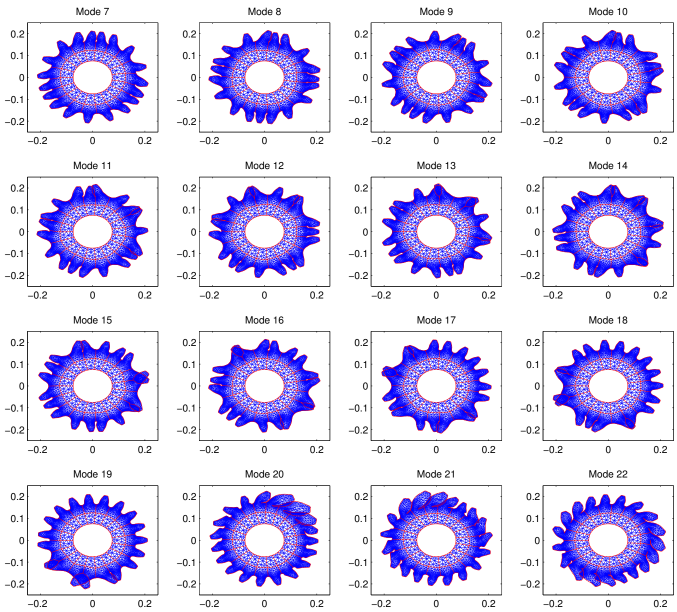

Figure 6.2: Visualisation of POD modes 7 to 22.

7 References

Bibliography

[1] Marco Antonio Cardoso. *Development and Application of Reduced-Order Modeling Procedures for Reservoir Simulation*. PhD thesis, Stanford University, 2009.

[2] Håkan Jakobsson. *Adaptive Reduction of Finite Models in Computational Solid Mechanics*. PhD thesis, Umeå University, 2011.

[3] J.L.Humar. *Dynamics of structures 2nd Edition*. Balkema, 2002.

[4] K.L. Johnson. *Contact Mechanics*. Cambridge University Press, 2003.

[5] S. Mehrotra K. Chakrabarti. Local dimensionality reduction: A new approach to indexing high dimensional spaces. 2000.

[6] F. Bengzon M. G. Larson. *The Finite Element Method - Theory, Implementation and Practice*. Springer, 2012.

[7] G. Berkooz P. Holmes, J. L. Lumley. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, 1998.

[8] S.P.A. Bordas P. Kerfriden, J.C. Passieux. Local/global model order reduction strategy for the simulation of quasi-brittle fracture. 2011.

[9] Ahmed A. Shabana. *Dynamics of Multibody Systems*. Cambridge University Press, 2005.

[10] Pin Tong Y.C. Fung. *Classical and Computational Solid Mechanics*. World Scientific, 2005.