INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN BIOMEDICAL ENGINEERING
Int. J. Numer. Meth. Biomed. Engng. (2012)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cnm.1491

Real-time simulation of surgery by reduced-order modeling and X-FEM techniques

S. Niroomandi [1], I. Alfaro [1], D. González [1], E. Cueto [1,*,†] and F. Chinesta [2]

[1] *Aragon Institute of Engineering Research, Universidad de Zaragoza, Zaragoza, Spain*
[2] *EADS Corporate International Chair, Ecole Centrale de Nantes, Nantes, France*

SUMMARY

This paper describes a novel approach for the simulation of surgery by a combined technique of model order reduction and extended finite element method (X-FEM) methods. Whereas model order reduction techniques employ globally supported (Ritz) shape functions, a combination with X-FEM methods on a locally superimposed patch is developed for cutting simulation without remeshing. This enables to obtain models with very few degrees of freedom that run under real-time constrains even for highly non-linear tissue constitutive equations. To show the performance of the technique, we studied an application to refractive surgery in the cornea. Copyright © 2012 John Wiley & Sons, Ltd.

Received 31 May 2011; Revised 31 August 2011; Accepted 9 November 2011

KEY WORDS: real time; surgery; model order reduction; X-FEM

1. INTRODUCTION

Development of surgery simulators for the risk-free training of surgeons has been a very active field of research in the last decades. Of course, this development is a very challenging task because of the very different complexities arising in this kind of simulators [1]. One of the sources of complexity is due to the highly non-linear behavior of soft living tissues, which are frequently modeled under the fiber-reinforced hyperelasticity framework [2]. Other source of complexity comes from the highly restrictive feedback rates imposed by the simulators (25 Hz for visual feedback and some 500 Hz if we want to add haptic feedback to the system). The third source of complexity comes from the multi-physic nature of the phenomena occurring in the actual surgery procedure: non-linear elasticity, contact, cutting, temperature, and so on.

Such simulators should provide a physically more or less accurate response such that, with the use of haptic devices, a realistic feedback is transmitted to the surgeon in terms of both visual feedback and force feedback. By 'accurate response', we mean that an advanced user should not encounter 'unphysical' sensations when handling the simulator. We definitely do not pursue an accurate solution in engineering terms. Following [3], '... the model may be physically correct if it looks right'.

Ayache and co-workers [1] defined the three generations of surgical simulators as those able to, respectively, reproduce accurately the anatomy (geometry), the physics (constitutive equations of soft tissue, temperature), and finally, physiology (blood flow, breathing, etc.). Undoubtedly, no third-generation simulation has ever been developed, and only some rigorous attempts have been made at the second-generation level (see [4, 5], among others). Most of the existing simulators

*Correspondence to: E. Cueto, Aragon Institute of Engineering Research, Universidad de Zaragoza, Edificio Betancourt, Maria de Luna, s.n., E-50018 Zaragoza, Spain.

†E-mail: ecueto@unizar.es

can be classified into the first-generation category, even if they provide haptic feedback, because they only consider linear elastic response, for instance [3, 6–8]. Other are based upon spring and mass systems, which do not even reproduce the equations of linear elasticity (see [9] and references therein).

Incorporating non-linear tissue constitutive equations into real-time frameworks is challenging in itself. Among the various techniques developed to reach this objective, we can distinguish techniques on the basis of the use of explicit finite elements and GPUs (see, for instance [4, 5]). Limitations of explicit finite elements with respect to stability of the result for large time steps are well known, however. Others have preferred to employ reduced-order modeling through proper orthogonal decomposition (POD), also known as principal component analysis (PCA) [10]. In either of these approaches, non-linear must be understood as 'geometrically non-linear', as the aforementioned approaches include simple co-rotational formulations or Saint Venant–Kirchhoff constitutive equations. References [11–13] seem to be among the earliest works that incorporate material (not only geometrical) non-linearity into the real-time framework.

The third source of difficulties, as mentioned earlier, comes from the multi-physic nature of the problem. Although the thermal dependence of the problem is often—if not always—neglected, contact detection and cutting simulation are of utmost importance for a convincing result in terms of both visual and haptic perception.

Numerical simulation of contact mechanics is a well-known field in the computational mechanics community (see, for instance the recent book of Wriggers [14] to acquire an overall impression of the difficulty of the topic). An accurate simulation of the process of contact between surgical tools and organs, and between organs themselves, seems to be at this moment out of reach under real-time requirements. Some simplified algorithms, however, provide very realistic results at high feedback rates (see for instance [15]). This algorithm supports distributed contact detection between a complex object (the surgical tool, for instance) and a deformable body reduced model, at 1 kHz rates.

As mentioned before, cutting simulation is another important source of troubles for real-time modeling of surgery. This is so as it is necessary to modify the geometry and/or the topology of the domain and its associated mesh, and this needs to be carried out without penalizing the computation times of the integration of the equations of motion. A vast corps of literature has been devoted to this end. See, for instance, [5, 16–18], to name a few. All these approaches share the same spirit. All of these works propose more or less sophisticated algorithms that allow to remove finite elements from an existing mesh to create a cut or even to remove a whole part of the organ, if it is being ablated. None of them seem to be aware of the existence of extended finite element method (X-FEM) (see for instance [19] and the subsequent enormous list of papers derived from the seminal work of Moës and colleagues).

Therefore, after a more or less exhaustive review of the existing literature, a conclusion can be drawn. Techniques based upon model order reduction constitute nowadays an appealing choice for an efficient, real-time, integration of the equations of motion [20]. They seem to be unique to efficiently handle material non-linearities such as those arising from state-of-the-art models for living soft tissues [12, 13]. And, finally, they seem to be specially well suited for efficient contact detection algorithms in real time [15].

Reduced-order modeling employs global (Ritz) shape functions to construct the approximation spaces in a Galerkin framework. These shape functions are usually constructed to be optimal under certain requirements. For instance, in POD techniques, these functions are extracted after a statistical treatment of the results of problems similar to the one at hand [21–24]. These techniques are known in many branches of science and technology under a wide variety of names, including the aforementioned POD, PCA, or Karhunen–Loève decomposition. Other techniques are even able to construct these shape functions without any prior knowledge on the results of similar problems (see [25–29]).

However, the use of global, Ritz, shape functions imposes additional limitations. For instance, the use of X-FEM techniques is not straightforward in this context. In this paper, a new method is presented, which combines the features of existing methods based upon model order reduction [12, 13] and the ease of creating cuts and discontinuities without remeshing of the X-FEM technique. As

DOI: 10.1002/cnm

will be noticed, the proposed technique has its origins in the s-FEM techniques by Fish [30] and the multiscale FEM by Rank [31].

The paper is organized as follows. In Section 2, the basics of POD techniques is reviewed in order to show how it can deal with non-linear constitutive equations under real-time requirements. Then, in Section 3, a new technique is developed, which can efficiently create cuts and discontinuities in the reduced model without remeshing and with minor modifications in the problem's stiffness matrix. Finally, in Section 4, an analysis of its performance is performed in an application to refractive surgery of the cornea.

2. MODEL ORDER REDUCTION BY PROPER ORTHOGONAL DECOMPOSITION METHODS

2.1. The Karhunen–Loève decomposition

The technique here considered can be seen as a particular instance of *a posteriori* model order reduction techniques. This means that the process to achieve real-time performance is composed by a sequence of *off-line* simulation, storage of the results, and subsequent analysis, followed by an *on-line* process of real-time simulation that employs the Ritz functions obtained in the off-line phase of the method. Following Cotin and Bro-Nielsen, '... We do not care about the time taken for one-time pre-calculation such as setting up equations, inverting matrices, etc.' [3].

Therefore, we assume that the evolution of a certain (scalar or vectorial) field $u(\boldsymbol{x},t)$, governed by a PDE, is known, typically from existing previous simulations on problems similar to that we are interested in. In practical applications, this field is expressed in a discrete form, that is, it is known at the nodes of a spatial finite element mesh and for some time steps of existing simulations $u(\boldsymbol{x}_i,t^n) \equiv u_i^n$. The same can be written by introducing a time discretization $u^n(\boldsymbol{x}) \equiv u(\boldsymbol{x},t=n\Delta t)$; $\forall n \in [1,\cdots,P]$. The main idea of the Karhunen–Loève (K-L) decomposition is how to obtain the most typical or characteristic structure $\phi(\boldsymbol{x})$ among these $u^n(\boldsymbol{x})$ $\forall n$. This is equivalent to obtaining a function $\phi(\boldsymbol{x})$ that maximizes α defined by

$$\alpha = \frac{\sum_{n=1}^{n=P}\left[\sum_{i=1}^{i=N}\phi(\boldsymbol{x}_i)u^n(\boldsymbol{x}_i)\right]^2}{\sum_{i=1}^{i=N}(\phi(\boldsymbol{x}_i))^2} \tag{1}$$

This maximization ($\delta\alpha = 0$) leads to

$$\sum_{n=1}^{n=P}\left[\left(\sum_{i=1}^{i=N}\tilde{\phi}(\boldsymbol{x}_i)u^n(\boldsymbol{x}_i)\right)\left(\sum_{j=1}^{j=N}\phi(\boldsymbol{x}_j)u^n(\boldsymbol{x}_j)\right)\right] = \alpha\sum_{i=1}^{i=N}\tilde{\phi}(\boldsymbol{x}_i)\phi(\boldsymbol{x}_i);\ \ \forall\tilde{\phi} \tag{2}$$

which is equivalent to

$$\sum_{i=1}^{i=N}\left\{\sum_{j=1}^{j=N}\left[\sum_{n=1}^{n=P}u^n(\boldsymbol{x}_i)u^n(\boldsymbol{x}_j)\phi(\boldsymbol{x}_j)\right]\tilde{\phi}(\boldsymbol{x}_i)\right\} = \alpha\sum_{i=1}^{i=N}\tilde{\phi}(\boldsymbol{x}_i)\phi(\boldsymbol{x}_i);\ \forall\tilde{\phi} \tag{3}$$

Defining the vector $\boldsymbol{\phi}$ such that its ith component is $\phi(\boldsymbol{x}_i)$, Equation (3) takes the following matrix form

$$\tilde{\boldsymbol{\phi}}^T\boldsymbol{c}\,\boldsymbol{\phi} = \alpha\tilde{\boldsymbol{\phi}}^T\boldsymbol{\phi};\ \forall\tilde{\boldsymbol{\phi}} \Rightarrow \boldsymbol{c}\,\boldsymbol{\phi} = \alpha\boldsymbol{\phi} \tag{4}$$

where the two-point correlation matrix is given by

$$c_{ij} = \sum_{n=1}^{n=P}u^n(\boldsymbol{x}_i)u^n(\boldsymbol{x}_j) \Leftrightarrow \boldsymbol{c} = \sum_{n=1}^{n=P}\boldsymbol{u}^n(\boldsymbol{u}^n)^T \tag{5}$$

DOI: 10.1002/cnm

which is symmetric and positive definite. If we define the matrix $\boldsymbol{Q}$ containing the discrete field history

$$\boldsymbol{Q} = \begin{pmatrix} u_1^1 & u_1^2 & \cdots & u_1^P \\ u_2^1 & u_2^2 & \cdots & u_2^P \\ \vdots & \vdots & \ddots & \vdots \\ u_N^1 & u_N^2 & \cdots & u_N^P \end{pmatrix} \tag{6}$$

it is straightforward to verify that the matrix $\boldsymbol{c}$ in Equation (4) results

$$\boldsymbol{c} = \boldsymbol{Q}\ \boldsymbol{Q}^T \tag{7}$$

where the diagonal components are given by

$$c_{ii} = (\boldsymbol{Q}\ \boldsymbol{Q}^T)_{ii} = \sum_{j=1}^{j=P} (u_i^j)^2. \tag{8}$$

Thus, the functions defining the most characteristic structure of $u^n(\boldsymbol{x})$ are the eigenfunctions $\phi_k(\boldsymbol{x}) \equiv \boldsymbol{\phi}_k$ associated with the highest eigenvalues.

2.2. A posteriori reduced-order modeling

As mentioned before, the main ingredient of the POD method is to perform off-line some initial simulations on the complete model(s) that allow to obtain $u(\boldsymbol{x}_i, t^n) \equiv u_i^n$, $\forall i \in [1, \cdots, N]$, $\forall n \in [1, \cdots, P]$, and from these, the r eigenvectors related to the r-highest eigenvalues $\boldsymbol{\phi}_k = \phi_k(\boldsymbol{x}_i)$, $\forall i \in [1, \cdots, N]$, $\forall k \in [1, \cdots, r]$ (with $r \ll N$). From these simulations, the r eigenfunctions can be successfully employed for approximating the solution of a problem slightly different than the one that has served to define $u(\boldsymbol{x}_i, t^n)$. For this purpose, it is necessary to define the matrix $\boldsymbol{A}$

$$\boldsymbol{A} = \begin{pmatrix} \phi_1(\boldsymbol{x}_1) & \phi_2(\boldsymbol{x}_1) & \cdots & \phi_r(\boldsymbol{x}_1) \\ \phi_1(\boldsymbol{x}_2) & \phi_2(\boldsymbol{x}_2) & \cdots & \phi_r(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\boldsymbol{x}_N) & \phi_2(\boldsymbol{x}_N) & \cdots & \phi_r(\boldsymbol{x}_N) \end{pmatrix}. \tag{9}$$

The typical structure of the discretized form of the (non-linear) equations of elasticity is of the type

$$\boldsymbol{K}\ \boldsymbol{U} = \boldsymbol{F}. \tag{10}$$

Obviously, in the case of time-dependent problems, $\boldsymbol{F}$ contains the contribution of the solution at the previous converged time step.

Assuming that the unknown vector contains the nodal (usually, displacement) degrees of freedom, it can be expressed as

$$\boldsymbol{U} = \sum_{i=1}^{i=r} \zeta_i \boldsymbol{\phi}_i = \boldsymbol{A}\ \boldsymbol{\zeta}, \tag{11}$$

which gives, for the complete system of equations,

$$\boldsymbol{K}\ \boldsymbol{U} = \boldsymbol{F} \Rightarrow \boldsymbol{K}\ \boldsymbol{A}\ \boldsymbol{\zeta} = \boldsymbol{F}. \tag{12}$$

Multiplying both terms by $\boldsymbol{A}^T$ gives

$$\boldsymbol{A}^T \boldsymbol{K}\ \boldsymbol{A}\ \boldsymbol{\zeta} = \boldsymbol{A}^T \boldsymbol{F}, \tag{13}$$

which proves that the final system of equations is of low order, that is, the dimensions of $\boldsymbol{A}^T \boldsymbol{K}\ \boldsymbol{A}$ are $r \times r$, with $r \ll N$, and the dimensions of both $\boldsymbol{\zeta}$ and $\boldsymbol{A}^T \boldsymbol{F}$ are $r \times 1$.

DOI: 10.1002/cnm

From the paragraphs above, we recover the initial idea exposed in Section 1. The main idea of model order reduction techniques is to employ, in a Ritz framework, a set of global basis that are, in a statistical sense, the best suited to reproduce the just computed, complete models. This is in sharp contrast with the FEM, which employs general purpose, piecewise polynomial shape functions to approximate the solution in a Galerkin framework.

2.3. Limitations of classical model reduction techniques

Although the model order reduction technique reviewed before gives excellent results in some, very useful, situations (see, for instance, [11, 20]) and has given very promising results at real-time feedback rates (see [10]), it still has some limitations. The main limitation deals with the severe limitations imposed by real-time constraints in haptic environments (set to some 0.5–1 kHz). This means that in practical situations, it is not possible to update the stiffness matrix $\boldsymbol{K}$ of Equation (13) in the course of the simulation. For strongly non-linear tissue constitutive equations, this can give errors below the 20% of the solution obtained by classical non-linear finite elements [11], still valid in some situations.

In some previous works of the authors, see [12, 13], a method has been developed and validated that allows for a fully non-linear simulation (in some neighborhood of previous, equilibrated, solutions) without the need of tangent stiffness matrix updating. The method is based on the combination of model order reduction techniques and asymptotic expansions of the solution (the so-called asymptotic numerical method [32–35]).

The second, very important, limitation of classical model reduction techniques (also of more recent versions such as those in [11, 12]) is that the employ of Ritz functions very much complicates the issue of simulating surgical cutting. The appearance of a cut implies a change of topology in the geometry of the domain and also an important change in the stiffness matrix, element connectivities, and so on. X-FEM techniques [19] very much limit these inconveniences for standard finite element simulations, with only minimal changes in the number of degrees of freedom and element connectivity, but these are not readily applicable to reduced-order models employing Ritz basis functions.

In the next section, we briefly review the X-FEM in the framework of standard finite elements, prior to its introduction in conjunction with reduced basis.

3. INTRODUCING CUTS IN THE REDUCED MODEL BY X-FEM TECHNIQUES

3.1. A brief review of the X-FEM

In order to make this work self-contained, a brief review of the well-known X-FEM is included here. The interested reader is referred to [19], for instance, among other classical references in the field.

The basic ingredient of X-FEM methods is to consider a cut or crack Γ_d as a discontinuity in the displacement field. Therefore, by simply enriching those nodes whose shape function's support intersects, the crack with a discontinuous functions will suffice to obtain a conforming discretization without the need of remeshing. Thus, the new approximation of the displacement will be, in the context of X-FEM, for a single crack or cut,

$$\boldsymbol{u}^h(\boldsymbol{x}) = \sum_{i \in I} \boldsymbol{u}_i N_i(\boldsymbol{x}) + \sum_{j \in J} \boldsymbol{b}_j N_j(\boldsymbol{x}) H(\boldsymbol{x}), \tag{14}$$

where $J = \{j \in I : \omega_j \cap D \neq \emptyset\}$, I represents the set of all nodes in the FE mesh, $N_i(\boldsymbol{x})$ represents the ith node shape function, whose support is ω_i, evaluated at $\boldsymbol{x}$, $\boldsymbol{u}_i$ represents node i's displacement vector, and $H(\boldsymbol{x})$ is discontinuous across the crack Γ_d, whose geometry is represented by D. Finally, $\boldsymbol{b}_j$ represents a new set of enriched degrees of freedom that control the magnitude of the displacement discontinuity across the crack.

Classical works of X-FEM in the context of fracture mechanics such as in [19] still add some more degrees of freedom in order to reproduce the linear elastic fracture mechanics asymptotic solution at the crack front. These are not considered here for several reasons. Firstly, for living soft tissues, the

DOI: 10.1002/cnm

form of the asymptotic solution (if any) at the crack front is in general not known. In any case, the gain in accuracy would be, in general, negligible, if compared with all the simplifying assumptions taken up to now in order to reach a real-time performance.

3.2. Coupling reduced models and X-FEM descriptions of surgical cuts

The approach here suggested is composed by the coupling between the (Ritz) reduced basis and a superimposed patch of finite elements where the force exerted by the scalpel exceeds some limit value, thus appearing a cut in the organ. This approach can be seen as a generalization of some existing techniques in the field of finite elements, notably the so-called s-version of the FEM by Fish [30] and the multiscale FEM by Rank [31], among others [36, 37]. The basic of the method is easily understood from Figure 1.

The domain Ω is discretized by finite elements, and a reduced-order model is constructed on top of it that employs Ritz basis functions, as explained before. Once the force exerted by the scalpel reaches a prescribed threshold (see the sections that follow), a cut is supposed to appear at the contact location. Around that region, a superimposed patch Ω_{L} of finite elements is placed around the scalpel tip. In Ω_{L}, the displacement field is approximated as

$$\boldsymbol{u} = \boldsymbol{u}^0 + \boldsymbol{u}^{\mathrm{L}} \text{ in } \Omega_{\mathrm{L}} \tag{15}$$

where, to guarantee the compatibility of the displacement field, we enforce

$$\boldsymbol{u}^{\mathrm{L}} = \boldsymbol{0} \text{ on } \Gamma_{0\mathrm{L}}. \tag{16}$$

The reduced-order model approximates the displacement everywhere, in the spirit of Equation (11),

$$\boldsymbol{u}^0(\boldsymbol{x}) = \sum_{i=1}^{i=r} \zeta_i \boldsymbol{\phi}_i(\boldsymbol{x}), \tag{17}$$

whereas in Ω_{L} a traditional FEM mesh, including discontinuous enrichment, is added:

$$\boldsymbol{u}^{\mathrm{L}}(\boldsymbol{x}) = \sum_{j \in \Omega_{\mathrm{L}}} \boldsymbol{d}_j N_j(\boldsymbol{x}) + \sum_{k \in J} \boldsymbol{b}_k N_k(\boldsymbol{x}) H(\boldsymbol{x}). \tag{18}$$

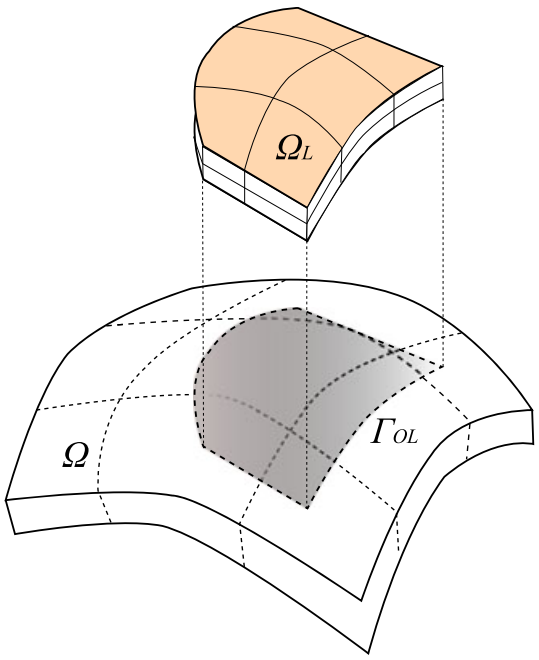

Figure 1. Basics of the method for reduced basis–X-FEM coupling.

DOI: 10.1002/cnm

The discrete equilibrium equations can be obtained after the weak form of the problem, namely, *find* $\boldsymbol{u} \in \boldsymbol{U} = \{\boldsymbol{u} : \boldsymbol{u} \in \mathcal{H}^1, \boldsymbol{u} = \overline{\boldsymbol{u}} \text{ on } \Gamma_u\}$ *such that*

$$\int_\Omega \boldsymbol{\varepsilon}^* : \boldsymbol{\sigma}\, d\Omega = \int_{\Gamma_t} \boldsymbol{u}^* \boldsymbol{t}\, d\Gamma + \int_\Omega \boldsymbol{u}^* \boldsymbol{b}\, d\Omega, \quad \forall \boldsymbol{u}^* \in \boldsymbol{U}_0 = \{\boldsymbol{u}^* : \boldsymbol{u}^* \in \mathcal{H}_0^1, \boldsymbol{u} = \boldsymbol{0} \text{ on } \Gamma_u\}, \tag{19}$$

where, as usual, Γ_u represents the Dirichlet (essential) part of the boundary of the domain, $\overline{\boldsymbol{u}}$ represents the value of the prescribed displacement at that location, $\boldsymbol{t}$ is the prescribed vector of traction at the boundary, and finally, Γ_t is the Neumann (natural) part of the boundary Γ. Admissible variations of the displacement are calculated as

$$\boldsymbol{u}^* = \left(\boldsymbol{u}^0\right)^* + \left(\boldsymbol{u}^{\mathrm{L}}\right)^*. \tag{20}$$

By substituting the displacement field approximations (17) and (18) into the weak form (19), and after invoking the arbitrariness of admissible displacements $\boldsymbol{u}^*$, we arrive at a discrete system of equations of the type

$$\begin{pmatrix} \boldsymbol{K}^{\zeta\zeta} & \boldsymbol{K}^{\zeta d} & \boldsymbol{K}^{\zeta b} \\ \boldsymbol{K}^{d\zeta} & \boldsymbol{K}^{dd} & \boldsymbol{K}^{db} \\ \boldsymbol{K}^{b\zeta} & \boldsymbol{K}^{bd} & \boldsymbol{K}^{bb} \end{pmatrix} \begin{pmatrix} \boldsymbol{\zeta} \\ \boldsymbol{d} \\ \boldsymbol{b} \end{pmatrix} = \begin{pmatrix} f^{\zeta} \\ f^{\boldsymbol{d}} \\ f^{\boldsymbol{b}} \end{pmatrix}, \tag{21}$$

where the stiffness matrices $\boldsymbol{K}^{\alpha\beta}$ are given by

$$\boldsymbol{K}_{IJ}^{\alpha\beta} = \int_{\Omega_i} (\boldsymbol{B}_I^{\alpha})^T \boldsymbol{D}\, \boldsymbol{B}_J^{\beta}\, d\Omega \tag{22}$$

where matrix $\boldsymbol{B}_I^{\alpha}$, $\alpha = \zeta, d, b$ represent the standard shape function derivative matrix for any of the three types of approximation functions here considered, namely, the Ritz functions $\phi(\boldsymbol{x})$, the standard finite element shape functions $N(\boldsymbol{x})$, or the discontinuous enrichment shape functions, $N(\boldsymbol{x})H(\boldsymbol{x})$. Ω_i represents the domain of integration, either Ω or Ω_{L}, respectively. Finally, $\boldsymbol{D}$ represents the consistently linearized constitutive tensor. As mentioned before, because of the severe restrictions placed by real-time requirements in haptic environments, it is not possible, in general, to update this tangent stiffness tensor. In this work, for the sake of simplicity in the exposition of the method, an approach similar to that in [11] is employed, in which no updating is accounted for. This leads, of course, to higher errors in the results, still acceptable in virtual surgery environments. If more accurate results are needed, an alternative formulation can be used [13], which employs asymptotic expansions of the displacement field and, without any updating, allows to closely follow complex non-linear force-displacement paths. The extension of the technique developed in some of our former works [13] to this framework is a work in progress.

In order to evaluate the previous integrals leading to the stiffness matrix of the problem, no special integration procedure is employed, in sharp contrast with the original works of Fish [30], as we assume that the superimposed mesh conforms with the existing one. It is necessary, however, to perform some form of tailored numerical integration in those elements enriched by discontinuous displacements, following the standard procedures of X-FEM.

3.3. Simplified physics of the cutting procedure

The last ingredient in the method is related to the placement of the patch Ω_{L} during the surgery. Once contact between the scalpel and the organ has been detected by a suitable contact algorithm (see [15] for instance, for a valid contact criterion in reduced model settings), a criterion must be set in order to determine if cutting appears, thus generating a new boundary in the domain, or not. We follow closely the criteria set in [38]. Although greatly simplified, these criteria have demonstrated to provide realistic enough results in haptic environments. A scalpel cuts along its blade, so a decomposition of the acting force as in Figure 2 is employed:

$$\boldsymbol{F}_{\text{ext}} = \boldsymbol{F}_{\perp} + \boldsymbol{F}_{\parallel} = \boldsymbol{F}_{\perp} + \boldsymbol{F}_a + \boldsymbol{F}_n \tag{23}$$

DOI: 10.1002/cnm

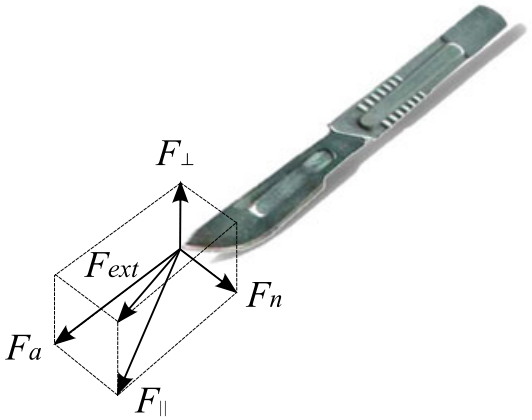

Figure 2. Force decomposition at the scalpel point of contact.

A threshold value of the force F_{cut} is considered such that lower modulus of the force $\boldsymbol{F}_{\parallel}$ produces static friction, but no cut. Once $\|\boldsymbol{F}_{\parallel}\|$ exceeds F_{cut}, the finite element patch Ω_{L} is added to the scalpel tip. In this work, a value $F_{\text{cut}} = 3N$ has been considered.

In order to simplify the process and to make it simpler and (notably) faster, once the threshold value F_{cut} is reached, a whole finite element is then cut. No cut of length smaller than the typical element size h is considered. If the finite element mesh is dense enough, this limitation does not very much affect the results. Remember that the size of the global finite element mesh, N, does not affect the size of the reduced model, r (see Equation (13).

3.4. Speeding up the X-FEM

The standard X-FEM technique, as presented before, provides excellent results but is, in general, somewhat *heavy* for real-time simulation. In this section, several assumptions are taken into account in order to alleviate these limitations.

The first one is related to the support of enriched degrees of freedom. The so-called shifted enrichment functions [39] allow to minimize the support of enriched approximation functions and therefore the width of the superimposed finite element mesh. These functions have the simple form

$$\phi_i(\boldsymbol{x}) = \frac{H(\boldsymbol{x}) - H_i}{2} \tag{24}$$

where H_i represents the value of $H(\boldsymbol{x})$ at the ith node.

The second big problem of the use of X-FEM for real-time simulation is that of the integration of the resulting weak form. The discontinuity forces to explicitly take into account the two resulting elements from a cut element and to apply standard Gauss integration on each volume. Location of the integration points in these resulting volumes is often a very heavy task. For that purpose, in this work, an alternative approach has been employed. It is based on the use of linear tetrahedra (in spite of their well-known numerical stiffness), for which exact, closed-form expressions exist for the stiffness matrix. Hexahedral meshes in the vicinity of the cut are then split into tetrahedra. Thus, the resulting stiffness matrix of an element traversed by the cut will be

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{K}^{uu} & \boldsymbol{K}^{ua} \\ \boldsymbol{K}^{au} & \boldsymbol{K}^{aa} \end{pmatrix} \tag{25}$$

where $\boldsymbol{K}^{uu}$ corresponds to the standard, unenriched element stiffness matrix, and $\boldsymbol{K}^{ua}$, $\boldsymbol{K}^{au}$, and $\boldsymbol{K}^{aa}$ correspond to the enriched degrees of freedom. As the linear tetrahedra have constant strain (and therefore constant derivatives matrix $\boldsymbol{B}$), we arrive at [40]

DOI: 10.1002/cnm

$$\boldsymbol{K}_{ij}^{uu} = \boldsymbol{K}_{ij} \tag{26}$$

$$\boldsymbol{K}_{ij}^{ua} = \left(\frac{V_a}{V} \Psi_{aj} + \frac{V_b}{V} \Psi_{bj} \right) \boldsymbol{K}_{ij} \tag{27}$$

$$\boldsymbol{K}_{ij}^{au} = \left(\frac{V_a}{V} \Psi_{ai} + \frac{V_b}{V} \Psi_{bi} \right) \boldsymbol{K}_{ij} \tag{28}$$

$$\boldsymbol{K}_{ij}^{aa} = \left(\frac{V_a}{V} \Psi_{ai} \Psi_{aj} + \frac{V_b}{V} \Psi_{bi} \Psi_{bj} \right) \boldsymbol{K}_{ij} \tag{29}$$

where $\boldsymbol{K}^{uu}$ is identical to that of the non-enriched tetrahedron, and $\Psi_{a(b)}$ represents the enrichment functions above (respectively, below) the cut plane. In the same spirit, V, V_a, and V_b represent the volume of the tetrahedron, or the volume above or below the cut, respectively.

The formulation presented before is apt for linear FEM. In the case of large displacements or strains, more sophisticate approaches should be used. This would be imperative in the case of resections, with large displacements of the resected part, for instance. However, for the application pursued in this paper, and for the sake of speed of simulation, it is maintained as presented before. As will be shown, the results are good in spite of this formal contradiction.

4. APPLICATION TO THE SIMULATION OF CORNEAL REFRACTIVE SURGERY

It must be highlighted at this point that the purpose of this work is not to validate the use of reduced-order models in the context of real-time applications. This has been performed in many previous works (see for instance [10–13], among others). Another purpose of this work is also not to validate X-FEM technique to accurately simulate moving discontinuities in solid mechanics problems (for instance, see, among others, [19, 41–44]). However, the combination of both provides a unique set of features that makes the resulting technique an appealing choice. Among these features, the feasibility of real-time simulation of complex non-linear tissues (including, for instance, fiber-reinforced hyperelastic tissue, see [11]) combined with the possibility of simulating surgical cutting without remeshing (at haptic feedback rates, it is obvious that for realistic rendering, the cut must be represented accurately, but only at some 30-Hz feedback rates and not at 500 Hz). To this end, the case of corneal refractive surgery is analyzed here under the framework of the proposed method.

Astigmatism is a refractive error due to the non-spherical shape of the cornea, that is, the refractive power is not uniform in all meridians. Refractive surgery techniques are used to modify the curvature in order to repair the refractive error of the eye [2]. This defect may be corrected by making the cornea as spherical as possible, through the application of cuts.

In addition, properties of the cornea can be quite different between patients with the same level of pathology; therefore, the technique presented before can be seen as an efficient means to plan a patient-specific surgery that minimizes uncertainty in the results, providing the surgeon with the sensations and results he will obtain later in the true surgery.

4.1. Finite element model of the cornea

The model of the cornea employed here is based upon that presented in [2]. The cornea was meshed using trilinear hexahedral elements. It consisted of 8514 nodes and 7182 elements. The mesh is shown in Figure 3 in two views. The cornea is clamped at its base, giving a dome-like problem that showed buckling under some types of loads.

The cornea was assumed to be, without loss of generality, hyperelastic, with a Kirchhoff–Saint Venant behavior. More sophisticated material behaviors can also be efficiently tackled with this technique, as in [11], where a two-families of fibers reinforced hyperelasticity model was successfully employed.

The Kirchhoff–Saint Venant model is characterized by the energy function given by

$$\Psi = \frac{\lambda}{2} (\mathrm{tr}(\boldsymbol{E}))^2 + \mu \boldsymbol{E} : \boldsymbol{E} \tag{30}$$

DOI: 10.1002/cnm

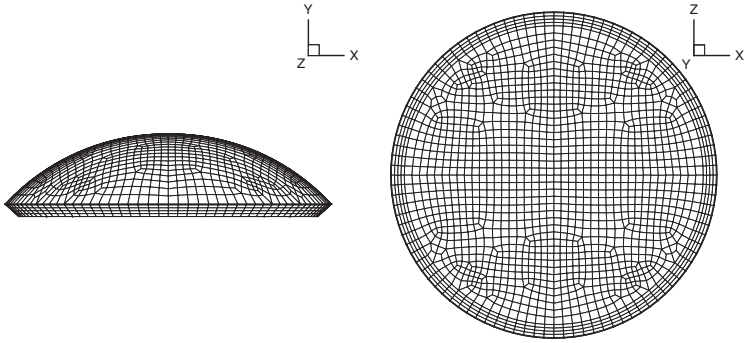

Figure 3. Geometry of the finite element model for the human cornea.

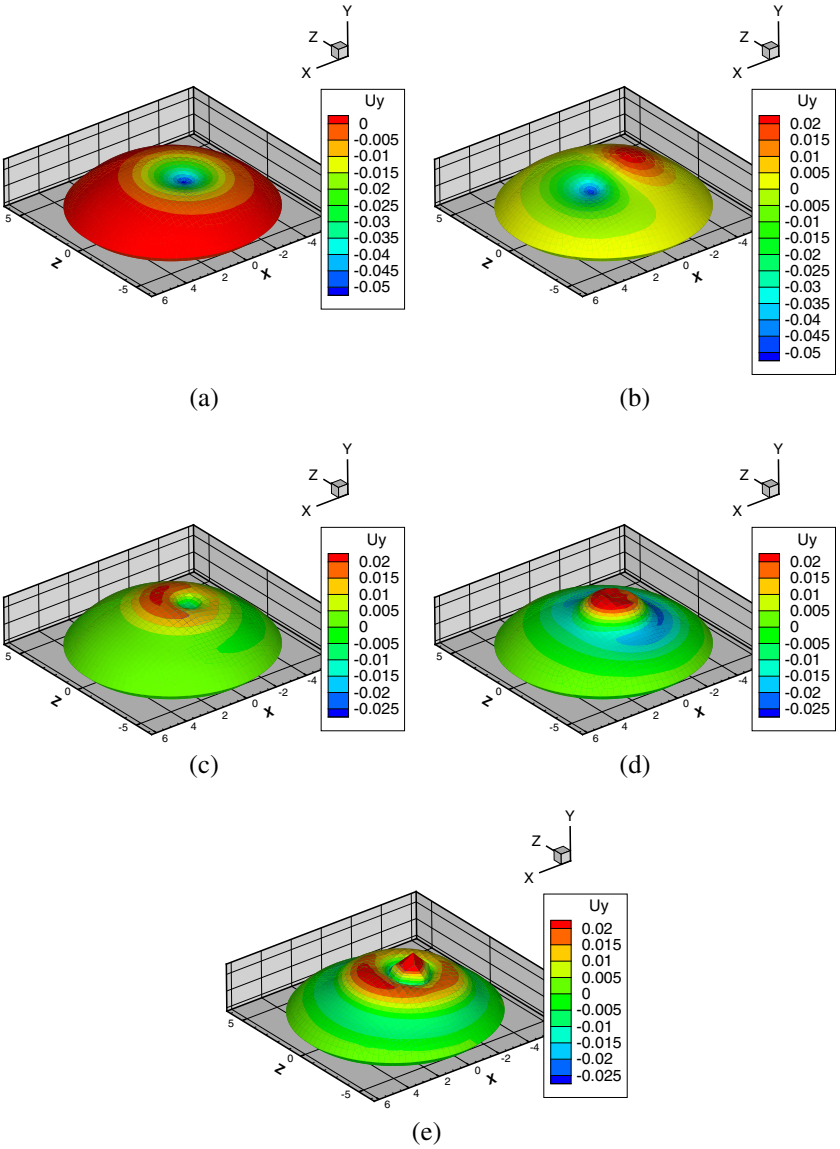

Figure 4. Five most important eigenmodes for the simulation of the cornea. The corresponding eigenvalues are 10.77 (a), 0.014 (b), $2.58 \cdot 10^{-4}$ (c), $6.53 \cdot 10^{-7}$ (d), and $4.43 \cdot 10^{-11}$ (e).

DOI: 10.1002/cnm

where λ and μ are Lame's constants. The Green–Lagrange strain tensor, $\boldsymbol{E}$, has the form

$$\boldsymbol{E} = \frac{1}{2}(\boldsymbol{F}^T\boldsymbol{F} - 1) = \gamma_l(\boldsymbol{u}) + \gamma_{nl}(\boldsymbol{u}, \boldsymbol{u}) \tag{31}$$

where $\boldsymbol{F} = \nabla\boldsymbol{u} + \boldsymbol{I}$ is the gradient of deformation tensor. The second Piola–Kirchhoff stress tensor can be obtained by

$$\boldsymbol{S} = \frac{\partial\Psi(\boldsymbol{E})}{\partial\boldsymbol{E}} = \mathbf{C} : \boldsymbol{E} \tag{32}$$

in which $\mathbf{C}$ is the fourth-order constitutive (elastic) tensor.

The Kirchhoff–Saint Venant hyperelastic model possesses well-known limitations, particularly some instabilities when subjected to pure compression. Nevertheless, it remains to be interesting for some applications. Noteworthy, in real-time simulation environments (see [10] and references therein, for instance) it is among the state-of-the-art models that can be computed under the severe limitations that real-time frameworks impose. The material properties of the cornea were assumed to be $E = 2$ MPa and $\nu = 0.48$ [2].

For a punctual load of increasing value, the complete model gave the five modes that capture 99% of the energy of the model. These modes are shown in Figure 4. Therefore, the reduced-order model has a stiffness matrix of 15×15 (five modes for each component of the displacement vector).

With these five modes, standard reduced-order models provide an error under 20% for loads placed at different positions to the one considered in the evaluation of the modes in Figure 4 and with no updating of the stiffness matrix. This level of error is considered valid in many real-time applications [11]. If more accurate simulations are needed, the method proposed in [13] provides a nearly exact simulation for displacement under the scalpel tip up to 1 mm (possibly much more than that, but at the scale of the cornea, 1-mm displacement involves very large strains and no reference finite elements results could be obtained to compare with).

4.2. Simulating limbal relaxing incisions

Limbal relaxing incisions are one of the three main types of cuts made in corneal refractive surgery. These incisions are made near the outer edge of the iris and are used to correct minor astigmatism (typically less than 2 diopters). In this case, a cut is made roughly at this position. In Figure 5, a

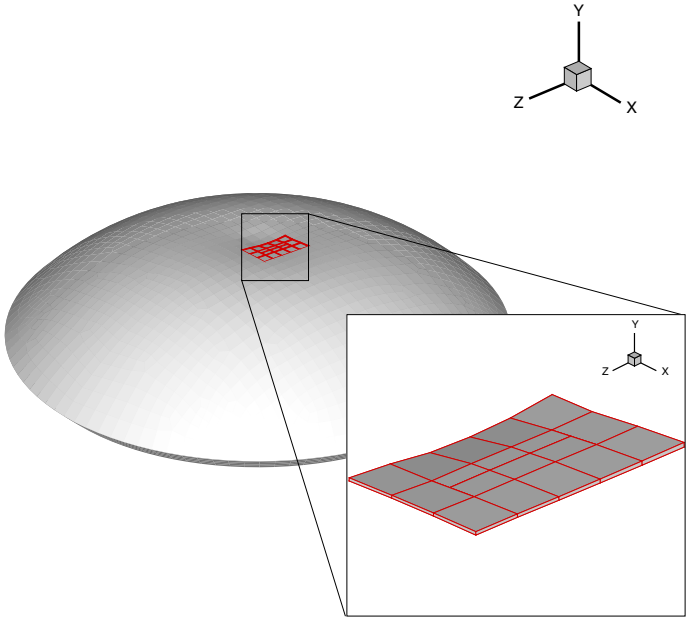

Figure 5. Superimposed X-FEM mesh at the position of limbal relaxing incision.

DOI: 10.1002/cnm

detailed view of the deformed cornea, together with the superimposed X-FEM mesh, is presented. For post-processing purposes, those elements cut by the scalpel are represented as two different finite elements in the figure, although no such elements exist in the simulation, as explained before.

The obtained displacement field is shown in Figure 6. A detailed view of the x-direction displacement field is shown in Figure 7, where the appearance of the cut can be noticed, in spite of the low magnitude of the displacement between crack lips.

In order to see the difference that a practitioner would see when dealing with the simulator, in Figure 8, a comparison is made between the solutions obtained by employing the complete model, explicitly meshing the cut by separate hexahedral finite elements, and that obtained by the combined POD-X-FEM model. The practically indistinguishable displacement field that is obtained (both are superimposed in Figure 8, in practice no difference can be noticed visually) is noticeable.

4.3. Timings

The results presented in this chapter have been obtained with a MacBook Pro laptop, running Matlab 2010a with an Intel Core 2 duo processor at 2.80 GHz and 4 Gb of DDR3 RAM memory

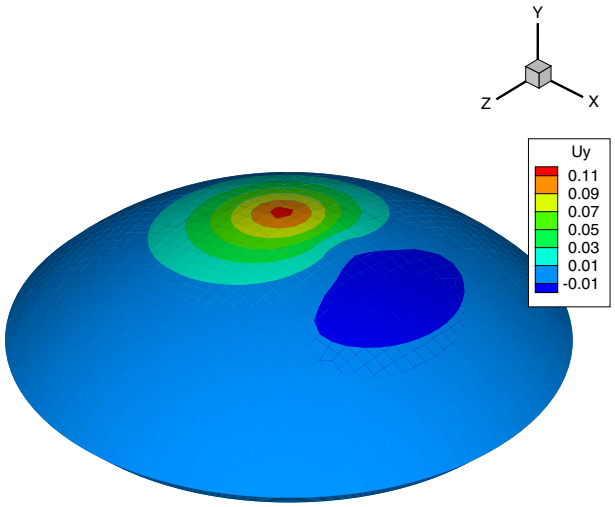

Figure 6. Displacement field (y-direction) in the cutting procedure.

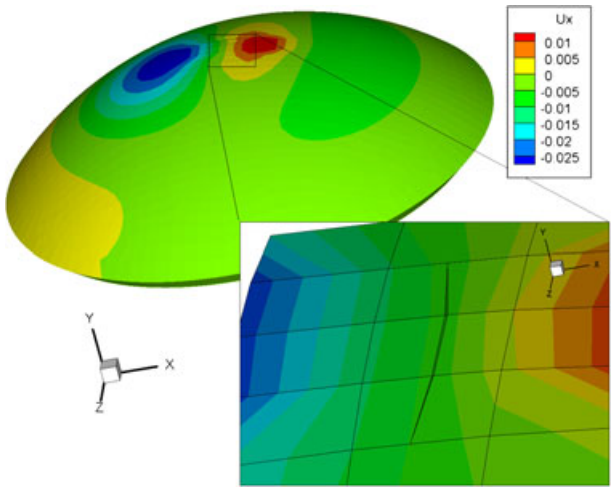

Figure 7. Cutting procedure. Displacement field (x-direction). Enrichment degrees of freedom are magnified 10 times to highlight the magnitude of the cut.

DOI: 10.1002/cnm

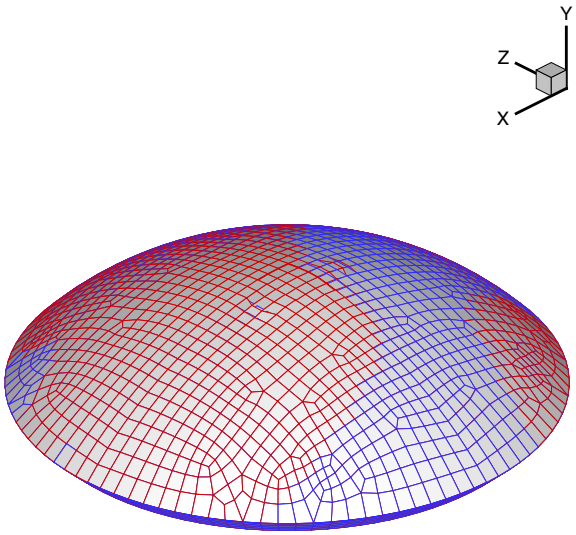

Figure 8. Cutting procedure. Comparison among the complete (red) and reduced (blue) models. Both results are superimposed, highlighting the good correspondence between each other.

(The MathWorks, Inc. Natick, Massachusetts, U.S.A.). Despite the use of rude Matlab code prototypes, the examples of the cornea ran at some 25 Hz, which is enough for visual real-time requirements.

The use of more powerful computers, possibly running in parallel, because many of the procedures here described can easily be programmed to run in that way, or the use of GPUs seem to be a feasible alternative to improve these promising results, is still compatible with the state-of-the-art techniques in the field (see for instance [40]).

5. CONCLUSIONS

In this paper, a novel method for the real-time simulation of surgery in haptic environments has been presented. The method is based on the use of reduced-order modeling. Reduced models are, to the author's knowledge, the only technique able to simulate at real-time feedback rates, highly complex constitutive models for living tissues (fiber-reinforced hyperelastic models, for instance). The problem with such an approach lies in the introduction of surgical cuts. The global character of the approximation functions (Ritz) precludes the possibility of employing standard methods in the literature that use very efficient algorithms for partitioning or eliminating the mesh in the neighboring of the cut. The approach here developed is based on the use of X-FEM techniques, coupled with the existing reduced model through a multi-scale-like method. Thus, the superimposed finite element mesh is capable of reproducing the displacement discontinuities produced by the scalpel, whereas the underlying reduced model is able to accurately reproduce the global behavior of the organ.

The proposed method runs at feedback rate, thus allowing to take part in a surgery simulator, together with some specialized contact detection algorithm. This opens the possibility to incorporate complex, state-of-the-art soft tissue constitutive laws into real-time simulation of surgery, thus leading the possibility of making a true second-generation simulator at hand.

REFERENCES

1. Delingette H, Ayache N. Hepatic surgery simulation. *Communications of the ACM* 2005; **48**:31–36.
2. Alastrue V, Calvo B, Pena E, Doblare M. Biomechanical modeling of refractive corneal surgery. *Journal of Biomechanical Engineering-Trasactions of the ASME* 2006; **128**:150–160.
3. Bro-Nielsen M, Cotin S. Real-time volumetric deformable models for surgery simulation using finite elements and condensation. *Computer Graphics Forum* 1996; **15**(3):57–66.

DOI: 10.1002/cnm

4. Taylor ZA, Comas O, Cheng M, Passenger J, Hawkes DJ, Atkinson D, Ourselin S. On modelling of anisotropic viscoelasticity for soft tissue simulation: numerical solution and GPU execution. *Medical Image Analysis* 2009; **13**(2):234–244. Includes Special Section on Functional Imaging and Modelling of the Heart.
5. Courtecuisse H, Jung H, Allard J, Duriez C, Lee DY, Cotin S. Gpu-based real-time soft tissue deformation with cutting and haptic feedback. *Progress in Biophysics and Molecular Biology* 2010; **103**(2-3):159–168. Special Issue on Biomechanical Modelling of Soft Tissue Motion.
6. Wang P, Becker AA, Jones IA, Glover AT, Benford SD, Greenhalgh CM, Vloeberghs M. Virtual reality simulation of surgery with haptic feedback based on the boundary element method. *Computers and Structures* 2007; **85**(7-8):331–339.
7. Cotin S, Delingette H, Ayache N. Real-time elastic deformations of soft tissues for surgery simulation. In *IEEE transactions on visualization and computer graphics*, Vol. 5 (1), Hagen H (ed.). IEEE Computer Society: Washington, DC, 1999; 62–73.
8. Basdogan C. Real-time simulation dynamically deformable finite element models using modal analysis and spectral Lanczos decomposition methods. In *Medicine meets virtual reality 2001*. IOS Press Amsterdam, 2001; 46–52.
9. Delingette H, Ayache N. Soft tissue modeling for surgery simulation. In *Computational Models for the Human Body*, Ayache N (ed.), Handbook of Numerical Analysis (Ph. Ciarlet, Ed.) Elsevier: Amsterdam, 2004; 453–550.
10. Barbič J, James DL. Real-time subspace integration for St. Venant–Kirchhoff deformable models. *ACM Transactions on Graphics (SIGGRAPH 2005)* 2005; **24**(3):982–990.
11. Niroomandi S, Alfaro I, Cueto E, Chinesta F. Real-time deformable models of non-linear tissues by model reduction techniques. *Computer Methods and Programs in Biomedicine* 2008; **91**(3):223–231.
12. Niroomandi S, Alfaro I, Cueto E, Chinesta F. Model order reduction for hyperelastic materials. *International Journal for Numerical Methods in Engineering* 2010; **81**(9):1180–1206.
13. Niroomandi S, Alfaro I, Cueto E, Chinesta F. Accounting for large deformations in real-time simulations of soft tissues based on reduced-order models. *Computer Methods and Programs in Biomedicine* 2010. DOI: 10.1016/j.cmpb.2010.06.012. In Press, Corrected Proof:–.
14. Wriggers P. *Computational Contact Mechanics*. Wiley: Chichester, 2002.
15. Barbič J, James D. Time-critical distributed contact for 6-DoF haptic rendering of adaptively sampled reduced deformable models. In *Symposium On Computer Animation 2007: ACM SIGGRAPH/Eurographics Symposium Proceedings*, Metaxas D, Popovic J (eds): 1515 Broadway, New York, NY 10036-9998 USA, 2007; 171–180. ACM SIGGRAPH; Eurog Assoc, Assoc computing machinery. Symposium on Computer Animation, San Diego, CA, AUG 03-04, 2007.
16. Zhang J, Gu L, Li X, Fang M. An advanced hybrid cutting method with an improved state machine for surgical simulation. *Computerized Medical Imaging and Graphics* 2009; **33**(1):63–71.
17. Lee B, Popescu DC, Ourselin S. Topology modification for surgical simulation using precomputed finite element models based on linear elasticity. *Progress in Biophysics and Molecular Biology* 2010; **103**(2-3):236–251. Special Issue on Biomechanical Modelling of Soft Tissue Motion.
18. Meier U, Lopez O, Monserrat C, Juan MC, Alcaniz M. Real-time deformable models for surgery simulation: a survey. *Computer Methods and Programs in Biomedicine* 2005; **77**(3):183–197.
19. Sukumar N, Moës N, Moran B, Belytschko T. Extended finite element method for three-dimensional crack modelling. *International Journal for Numerical Methods in Engineering* 2000; **48**:1549–1570.
20. Krysl P, Lall S, Marsden JE. Dimensional model reduction in non-linear finite element dynamics of solids and structures. *International Journal for Numerical Methods in Engineering* 2001; **51**:479–504.
21. Lorenz EN. *Empirical Orthogonal Functions and Statistical Weather Prediction*. MIT, Department of Meteorology, Scientific Report Number 1, Statistical Forecasting Project, 1956.
22. Park HM, Cho DH. The use of the Karhunen–Loève decomposition for the modeling of distributed parameter systems. *Chemical Engineering Science* 1996; **51**(1):81–98.
23. Brooks GP, Powers JM. A Karhunen-Loeve Galerkin Technique with Shock Fitting for Optimization of a Blunt Body Geometry. *AIAA-2002-3861, 38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Indianpolis, Indiana, July 2002.
24. Karhunen K. Uber lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae scientiarum Fennicae. Series A1. Mathematica - Physica* 1946; **37**:1–79.
25. Ryckelynck D. A priori hyperreduction method: an adaptive approach. *Journal of Computational Physics* 2005; **202**(1):346–366.
26. Ryckelynck D, Chinesta F, Cueto E, Ammar A. On the a priori model reduction: overview and recent developments. *Archives of Computational Methods in Engineering* 2006; **12**(1):91–128.
27. Ammar A, Mokdad B, Chinesta F, Keunings R. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. *Journal of non-Newtonian fluid Mechanics* 2006; **139**:153–176.
28. Ammar A, Mokdad B, Chinesta F, Keunings R. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. Part II: transient simulation using space–time separated representations. *Journal of non-Newtonian fluid Mechanics* 2007; **144**:98–121.
29. Chinesta F, Ammar A, Cueto E. Recent advances in the use of the proper generalized decomposition for solving multidimensional models. *Archives of Computational Methods in Engineering* 2010; **17**(4):327–350.
30. Fish J. The s-version of the finite-element method. *Computers and Structures* 1992; **43**(3):539–547.

DOI: 10.1002/cnm

31. Rank E, Krause R. A multiscale finite-element method. *Computers and Structures* 1997; **64**(1–4):139–144. Computational Structures Technology.
32. Cochelin B, Damil N, Potier-Ferry M. Asymptotic-numerical methods and Padé approximants for non-linear elastic structures. *International Journal for Numerical Methods in Engineering* 1994; **37**:1187–1213.
33. Abichou H, Zahrouni H, Potier-Ferry M. Asymptotic numerical method for problems coupling several nonlinearities. *Computer Methods in Applied Mechanics and Engineering* 2002; **191**(51-52):5795–5810.
34. Cao H-L, Potier-Ferry M. An improved iterative method for large strain viscoplastic problems. *International Journal for Numerical Methods in Engineering* 1999; **44**:155–176.
35. Cochelin B, Damil N, Potier-Ferry M. The asymptotic numerical method: an efficient perturbation technique for nonlinear structural mechanics. *Revue Europeenne des Elements Finis* 1994; **3**:281–297.
36. Ammar A, Pruliere E, Ferec J, Chinesta F, Cueto E. Coupling finite elements and reduced approximation bases. *European Journal of Computational Mechanics* 2009; **18**(5–6):445–463.
37. Ammar A, Chinesta F, Cueto E. Coupling finite elements and proper generalized decompositions. *International Journal for Multiscale Computational Engineering* 2011; **9**(1):17–33.
38. Bielser D, Gross MH. Interactive simulation of surgical cuts. In *Proceedings of the 8th Pacific Conference on Computer Graphics and Applications*, PG '00. IEEE Computer Society: Washington, DC, USA, 2000; 116–125.
39. Zi G, Belytschko T. New crack-tip elements for XFEM and applications to cohesive cracks. *International Journal for Numerical Methods in Engineering* 2003; **57**(15):2221–2240.
40. Jerabkova L, Kuhlen T. Stable cutting of deformable objects in virtual environments using XFEM. *IEEE Computer Graphics and Applications* 2009; **29**(2):61–71.
41. Combescure A, Gravouil A, Gregoire D, Rethore J. X-FEM a good candidate for energy conservation in simulation of brittle dynamic crack propagation. *Computer Methods in Applied Mechanics and Engineering* 2008; **197**(5):309–318. Enriched Simulation Methods and Related Topics.
42. Meschke G, Dumstorff P. Energy-based modeling of cohesive and cohesionless cracks via X-FEM. *Computer Methods in Applied Mechanics and Engineering* 2007; **196**(21-24):2338–2357.
43. Daux C, Moes N, Dolbow J, Sukumar N, Belytschko T. Arbitrary branched and intersecting cracks with the extended finite element method. *International Journal for Numerical Methods in Engineering* 2000; **48**(12):1741–1760.
44. Stolarska M, Chopp DL, Moes N, Belytschko T. Modelling crack growth by level sets in the extended finite element method. *International Journal for Numerical Methods in Engineering* 2001; **51**(8):943–960.

DOI: 10.1002/cnm