

Structure-preserving model reduction for mechanical systems

Sanjay Lall^{a,1}, Petr Krysl^{b,2}, Jerrold E. Marsden^{c,*,3}

^a Aeronautics and Astronautics, Stanford University, Stanford, CA 94305-4035, USA

^b Department of Structural Engineering, Jacobs School of Engineering, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0085, USA

^c Control and Dynamical Systems, 107-81 California Institute of Technology, Pasadena, CA 91125-8100, USA

Dedicated to Alan Newell on the occasion of his 60th birthday

Abstract

This paper focuses on methods of constructing of reduced-order models of mechanical systems which preserve the Lagrangian structure of the original system. These methods may be used in combination with standard spatial decomposition methods, such as the Karhunen–Loève expansion, balancing, and wavelet decompositions. The model reduction procedure is implemented for three-dimensional finite-element models of elasticity, and we show that using the standard Newmark implicit integrator, significant savings are obtained in the computational costs of simulation. In particular simulation of the reduced model scales linearly in the number of degrees of freedom, and parallelizes well.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Model reduction; Karhunen–Loève expansion; Balanced truncation; Lagrangian mechanics

1. Introduction

1.1. Problem description

The problem setting. The problem of constructing simple, yet predictive, models for complex physical systems operating on many length and time scales has a long and distinguished history, from finding finite dimensional Galerkin truncation models to inertial

manifolds and to finding envelope equations, a particular love of Alan's. Another approach, the focus of the present paper, is to use what is variously known as *model reduction*, *the Karhunen–Loève expansion (KLE)*, *empirical eigenfunction*, or *proper orthogonal decomposition method*.

Systems are often modeled by nonlinear partial differential equations that contain phenomena on many scales, which can be both difficult to analyze mathematically and computationally expensive for simulation, design and control problems. The reasons for this difficulty are manifold. For example, the system itself may have complex geometry, such as flow through a jet engine or the dynamic motions of an automobile. Thus, if one attempts to model the fluid equations or those of elasticity in such a complex geometry, the amount of computation will of course be significant, even for simple flows or motions. However, there may

* Corresponding author.

E-mail addresses: lall@stanford.edu (S. Lall), pkrysl@ucsd.edu (P. Krysl), marsden@cds.caltech.edu (J.E. Marsden).

¹ Research partially supported by AFOSR MURI grant F49620-96-1-0471 and NSF/DARPA OPAAL grant DMS-9874082.

² Research partially supported by NSF/DARPA OPAAL grant DMS-9874082.

³ Research partially supported by NSF/ITR grant ACI-0204932 and NSF/DARPA OPAAL grant DMS-9874082.

be intrinsic difficulties in simple geometries as well, such as in turbulent flow with its characteristic feature of a cascade of energy to small scales. It is very important to understand to what extent one has to model the small scale dynamics to achieve accurate models of the large-scale motions. Recent work on large eddy simulation models and averaged fluid equations [10,23,25,26] suggests that indeed one can do this with considerable savings in computational cost. In general, multiscale phenomena, both temporal and spatial, are of great importance as well as the source of many of the difficulties.

Fluid computations. The Karhunen–Loève method is perhaps best known in fluid computations, as described in [11]. The literature is huge in this area and we cite only the recent work of [34,35] as examples. While fluids have a well-known variational structure (see, for instance, [10]), it is more subtle than that of solids because of the large *particle relabeling symmetry group*. Thus, in this paper we have focused on solid mechanical examples.

General goals of this paper. In this paper we focus on the problem of constructing low-dimensional models for *mechanical* systems. Our aim is to develop a general methodology which is applicable to a wide range of mechanical systems, including systems of jointed rigid bodies such as robotic systems, as well as fluid and elastic systems modeled through finite-element analysis.

Such systems are well studied, and a significant amount is known about their mathematical and geometric structure. This geometric structure has fundamental implications for our understanding of the behavior of many mechanical systems. It also leads to computational methods which take advantage of this structure, for example in ensuring that numerical integration methods conserve energy or momentum.

In this paper, our goal is to develop a model reduction procedure that is consistent with and indeed preserves the geometric structure underlying the mechanics, and that ties in with standard computational methods for analysis and simulation of both finite-dimensional and continuum mechanical systems. Our main focus is to develop a basic theory behind this area of *mechanical model reduction*, appli-

cable to nonlinear high-dimensional systems, whose configuration spaces may have constraints; that is, be manifolds. Non-trivial configuration manifolds are, of course, often introduced in holonomic mechanical systems by imposing configuration space constraints, such as those encountered in articulated and robotic systems.

We first discuss our motivation behind the basic problem, explaining why reduced-order models of high-dimensional systems are of both mathematical and computational interest. We then give an overview of the basic theory we develop later in the paper for model reduction of mechanical systems.

1.2. Motivation

The fundamental motivation behind model reduction is that low-dimensional systems should both be simpler to work with analytically, and be faster and more convenient to work with computationally. There is of course a great need for such computational savings in problems of both structural design and control design.

Structural design. Many high-dimensional or continuum mechanical systems exhibit behavior that it is perhaps not unreasonable to expect to be well modeled by appropriate low-dimensional nonlinear systems. An example is given by the dynamic motion of an aircraft wing in flight, where the typical motions of the wing are large-scale and often relatively simple bulk bending dynamics.

The computational advantages are multiplied when considering, for example, performance evaluation of an aircraft wing under dynamic loading. Here finite-element methods are typically used, and performance checked via Monte Carlo sampling. Since a large number of repeated simulation trials must be performed, any reduction in the computational costs per simulation can allow a greater exploration of Monte Carlo space.

Control design. Other applications include control design, where the design of a stabilizing feedback controller may be extremely difficult for a high-dimensional nonlinear system, but much simpler for the low-dimensional system. Model reduction

should preserve the essential features of the system dynamics, and for some control applications this is all that is necessary; the controller does not depend on the highly uncertain fine-scale features of the dynamics. See [3,12,18,30,37].

Systems analysis. Analysis problems of interest include an understanding of the bifurcation structure of the system. A further benefit of model reduction is that often a low-dimensional model can provide qualitative understanding of the phenomena under consideration.

1.3. Previous work

The most basic method of model reduction for linear systems is that of modal expansion, where the phase space of the system is decomposed into subspaces corresponding to an eigendecomposition of the generator of the differential equations. This is a technique of fundamental importance for many applications. However, it is important to realize that the modal decomposition alone does not provide enough information to decide upon a good reduced-order model. What is additionally necessary is a method of deciding *which* modes should be preserved in the model. Typically the low-frequency modes are kept, however, there are many applications where this is not the best choice; in particular, in control systems where the frequency at which an accurate model is necessary is at crossover, and this may not correspond to the low-frequency modes of the system.

One of the most widely used methods for model reduction of general dynamical systems is to apply a KLE to the state space, and use a Galerkin projection to construct the reduced system. The KLE was introduced by Pearson [32], and developed by several authors, including [14,21]. The use of this method for analysis of turbulent flows was pioneered by Lumley [22].

For linear control systems, the method of balanced truncation due to [29] has proved to be extremely useful in practical applications, as it tends to preserve the dynamics which are important for control. A mathematically precise version of this statement is possible, using the results of [6,9] on the errors obtained using balanced truncation.

The relationship between balancing and the KLE method was developed in the papers by Lall et al. [17,18], where a method of using the KLE in order to construct the balanced truncation of a linear system of n first-order differential equations was constructed. In fact, the standard KLE methods applied to linear systems in first-order form is equivalent to the method known as *input-balancing* for controlled systems with a single-input.

For mechanical systems, many of the above relationships remain to be worked out. A possible approach is to use balancing methods as used in [18] in combination with the methods in this paper to construct reduced-order nonlinear models for a mechanical system.

The reduced-order systems we construct are integrated using globally-supported basis functions, known as Ritz functions, in contrast with the standard locally-supported shape function approach. Previous approaches deriving the Ritz functions from the linearization of the equations of motion were probably initiated by Nickell [31], who extended the use of modal superposition methods for systems with nonlinear dynamics. Wilson et al. [38] later introduced the so-called load-dependent vectors. A recent overview of these techniques was given by Leger and Dussault [19], and this method was also used in [15,16].

2. Overview

2.1. An example: the docking device

We will illustrate the theory with an example of finite-element analysis of nonlinear three-dimensional elasticity, called the *docking device*, shown in Fig. 1. This model consists of two rectangular blocks of rubber, connected by a steel rod. One block is mounted on fixed supports. The other block is clamped to an unsupported rigid and massive steel frame.

The elasticity in this system is modeled by a nonlinear finite-element model with 9600 degrees of freedom. For this example, we construct a reduced-order model with 36 degrees of freedom which captures the essential dynamics of the high-order system.

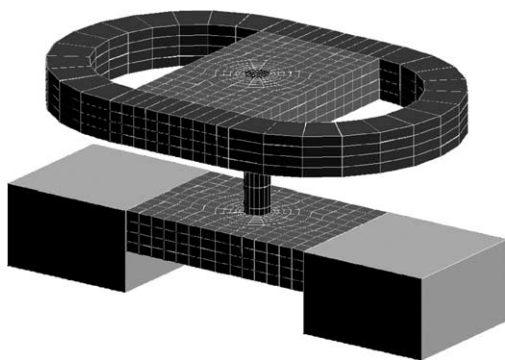


Fig. 1. The docking device.

Clear computational savings are obtained for simulation of this proof-of-concept system. Simulation of the original system for 1 s took about 24 CPU hours, using an implicit Newmark integrator, whereas the simulation time for the reduced system, with the same time-step, was about 10 CPU hours.

In the final section of this paper, we discuss the numerical methods used for simulation, and how the problem of model reduction is fundamentally connected to these methods. We show that the reduced-order models we construct result in significantly reduced computational effort in simulation, and we give some theoretical results showing that these computational benefits increase substantially as problem size increases; in particular, the model reduction affords linear scalability with respect to the size of the finite-element model, and the computations parallelize very well.

2.2. Structure preservation

Mechanical structure. We take the Lagrangian viewpoint on mechanical systems, where the system is defined by a Lagrangian function $L : TQ \rightarrow \mathbb{R}$, where Q is the configuration manifold and TQ the corresponding velocity phase space. Typically, the Lagrangian is of the form kinetic minus potential energy.

This mechanical structure leads to important properties. Notably, mechanical systems, in the absence of forcing and dissipation, conserve energy as well as

quantities associated with the symmetries of the system. The dynamics of a mechanical system also satisfy a variational principle, and the evolution maps consist of symplectic transformations. All of these properties can be viewed as fundamental to any model of a given mechanical system, and the reduced-order system should, if possible, possess them.

The model reduction techniques presented in this paper preserve the mechanical structure underlying the system dynamics, and are applicable for systems on non-Euclidean configuration spaces.

The basic idea of how we achieve this is as follows. Rather than considering the original system as defined by a set of ordinary or partial differential equations, we view it as defined by a configuration space and an associated Lagrangian function. From this starting point, the complete dynamics of the system can be constructed, via the variational principle governing the system behavior. We therefore take the viewpoint that the reduced-order system should also be specified by a Lagrangian function on a new configuration space of small dimension. All of the dynamics of the reduced system are then completely specified by the variational principle. Because of this construction, all of the standard geometrical structure and properties of the original mechanical system are inherited by the reduced system.

Non-Euclidean configuration spaces. A further important feature of mechanical systems is that the configuration space is usually non-Euclidean. For example, the configuration space of the three-dimensional rigid body is the manifold $SO(3)$ or $SE(3)$ and robotic systems consisting of rigid bodies with links have configuration spaces that are products of copies of $SE(3)$ with constraints imposed. Fluids and elasticity problems also have non-flat configuration manifolds, such as defined by incompressibility constraints in the Lagrangian (material) representation of the fluid. These are fundamental constraints due to the physics of the system, which a reduced-order model should also preserve.

Discrete mechanics, numerical methods and reduction. This same point of view regarding the variational principle as basic (rather than the equations themselves) has proven useful in the development

of discrete mechanics (see [27] for an overview and literature). The idea here is to discretize Hamilton's principle rather than standard discretizations of the Euler–Lagrange equations. This approach gives rise to structure-preserving integration algorithms. In fact, the ideas of structure preservation used in this paper are also tied to the numerical techniques we employ; we make use of the Newmark integrator, which has recently been shown to be a variational integrator by Kane et al. [13]; see also [20] for an asynchronous generalization to the PDE context and with applications to nonlinear elastodynamic simulations. As such, the model reduction gives rise to a time-discretized evolution equation which exactly preserves momentum and the symplectic form, and approximately preserves energy.

This variational view for obtaining structure preservation also occurs in the now very popular method of Lagrangian reduction of Marsden and Scheurle (see [4] for a current account and literature).

Given their common variational nature, we expect that these different lines of work can be merged in fruitful and exciting ways.

2.3. Basis selection

In order to perform model reduction, the original dynamics must in some sense be represented in a new set of basis functions. However, it is important to understand that the techniques in this paper do not simply 'project the equations' onto a reduced-basis as in the standard linear setting.

However, although the details are very different, model reduction of mechanical systems requires solution of an optimization problem in order to select an appropriate reduced-order configuration space. In this paper, we show how even for nonlinear problems on non-Euclidean configuration spaces, it is possible to embed the system into a linear space in such a way that standard optimization methods can be used for the basis selection, and the reduced system still inherits the natural mechanical structure of the original system. However, the way this basis is used is different from the usual Galerkin projection or singular perturbation approach.

There exist many methods for basis selection; equivalently, in the context of finite-element models of elasticity, we may view this as mesh selection or element selection. These methods can be directly plugged into the theoretical framework in this paper. Known methods for basis selection include the following.

The Karhunen–Loève expansion. The docking-device example in this paper makes use of the KLE to compute the reduced-basis. This is a data-based method for selecting a subspace of given dimension, optimal in the sense of least-squares; it is also known as the proper orthogonal decomposition (POD) or method of empirical eigenfunctions. The KLE method is widely used in fluid mechanics to find large-scale coherent structures.

Balancing. The method of *balanced truncation* was introduced by Moore [29] in the context of realization theory for linear control systems. The paper of Lall et al. [18] showed that the process of balancing and truncation could be separated, with balancing providing a method for selecting the basis in which to best capture the dynamics relevant to the input–output behavior of a coupled *nonlinear* system. This method of balancing can be used in exactly the same way as the KLE to provide a mechanically consistent method of model reduction for controlled mechanical systems, as well as for mechanical systems in which coupled-dynamics play a role. An example would be tying together a reduced-order model of an elastic wing with a full flight simulation.

Wavelet decomposition and error bounds. Wavelet decomposition techniques are naturally hierarchical and multiresolutional, and provide a mathematical framework on which to form adaptive basis methods for finite-element elasticity. In fact, these methods are strongly tied to new subdivision methods for representing the dynamics of thin shells [5]. Note that the model reduction problem for traditional dynamic thin-shell models is more complicated than that for three-dimensional elasticity, due to the underlying geometry of the configuration space and the way essential boundary conditions are treated. The subdivision thin-shell models are an exception to this rule, because their configuration spaces are Euclidean [5]. We do not give a discussion of error bounds in this

work; this is an important but separate issue. However, following the ideas in [33], we hope to address this issue in a future publication.

Goals. The goal of this paper is to establish the technique of *structure-preserving schemes* for *mechanical model reduction* as a *proof-of-concept*; the examples we give are not the ultimate examples, but show the main ideas. We are building a systematic framework to understand coarsening and refinement of models and as better hierarchical methods are developed, such as those of Krysl et al. [15,16], and symmetry methods are better understood (see, for example, [36]), these methods will become more practical. We return to some of these issues at the end of the paper.

3. Model reduction of mechanical systems

In this paper, our goal is to develop model reduction methods for mechanical systems. We now make precise our notion of mechanical system, following the modern approach to mechanics; see, for example, the treatment in [24]. Let the space of configurations of a mechanical system be a differentiable manifold Q . The Lagrangian is a function $L : TQ \rightarrow \mathbb{R}$, where TQ is the tangent bundle of Q , typically given by the difference between the kinetic and potential energy of the system:

$$L(q, \dot{q}) = T(q, \dot{q}) - V(q).$$

The equations of motion of this system are given by the Euler–Lagrange equations, which in generalized coordinates q^i are

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0. \quad (1)$$

A traditional approach. A procedure often used for model reduction is to rewrite the Euler–Lagrange equations in first-order form, as

$$\dot{x}(t) = f(x(t)), \quad (2)$$

where $x(t)$ is the $2n$ -dimensional state of the system, consisting of the phase variables (p, q) . The model reduction problem is then defined to be the problem of

constructing a lower dimensional dynamical system

$$\dot{y}(t) = g(y(t)), \quad (3)$$

where $y(t)$ is k -dimensional for some $k \ll n$, and a reconstruction function $h : y \mapsto x$ such that given $y(t)$ the reconstruction $\tilde{x}(t) = h(y(t))$ is, in some appropriate sense, a good approximation to $x(t)$ for a certain range of system behaviors.

Mechanical problem formulation. However, since the dynamics of a mechanical system are specified entirely by the Lagrangian L and the configuration space Q , it is natural to use these as our starting point for model reduction. Instead of the above problem definition, we formulate the model reduction problem as follows:

- (i) Given the configuration space Q , find a submanifold $Q_r \subset Q$.
- (ii) Construct a new mechanical system whose configuration space is Q_r .

This separation of the procedure into these two stages is very important, since it separates the configuration space selection from the dynamic reconstruction. The state space of this reduced system will be TQ_r , a submanifold of the original state space TQ .

3.1. Construction of the reduced dynamics

First-order systems. For first-order systems of the form (2) evolving on Euclidean space \mathbb{R}^n , one approach is to find a set of basis functions to describe the state evolution, and perform a linear decomposition

$$x(t) = \sum_{i=1}^k a_i(t) \phi_i,$$

where the basis vectors ϕ_i lie in Euclidean space \mathbb{R}^n . This is equivalent to specifying a subspace $S \subset \mathbb{R}^n$ of the original state space, on which the important dynamics evolve.

The next stage of the traditional process is the actual construction of the reduced dynamics on this subspace; many different standard methods exist, for example Galerkin projection and singular perturbation. The Galerkin projection simply projects

the vector field locally onto the subspace S , leading to a reduced-dimension set of ordinary differential equations.

Mechanical systems. Application of the Galerkin projection to a mechanical system does not in general preserve the mechanical structure. A better approach for mechanical systems is to use *constraints*. Rather than projecting the dynamics onto the constraint manifold, one instead applies mechanical constraints to the system so that the dynamics naturally evolve on the constraint submanifold.

We propose the following procedure. The original Lagrangian L is restricted to the constraint submanifold Q_r to give a new Lagrangian $L|_{Q_r}$ on Q_r . This new Lagrangian is then used to construct Euler–Lagrange equations for the reduced dynamics on Q_r , in the usual way. These reduced dynamics will then be truly mechanical, and satisfy all of the structural properties of such systems.

In this way, the construction of a new mechanical system on Q_r is enabled. In order to complete the construction of the reduced-order model, all that remains is to make a good choice of the submanifold Q_r .

3.2. Submanifold selection

This is the critical part of the model reduction process. We would like to find a submanifold Q_r which contains those dynamics of the system which we would like to model.

One possible approach would be to try to directly perform some kind of expansion in a more general nonlinear context using manifolds. While theoretically appealing, it does not appear likely that this approach would lead to computationally tractable problems.

The alternative approach we present in this paper is therefore to embed Q in a linear space V of higher dimension, and apply a linear subspace decomposition in V . Having chosen some subspace $V_r \subset V$, we can construct Q_r as the intersection $Q \cap V_r$.

The particular embedding to be used here is open to choice. In fact, for a given problem there is often a natural linear space in which to embed the dynamics. This may be motivated both by the physics of the

problem, as well as by the application to which the reduced model will be put. We now give some examples for the choice of linear space.

Generalized coordinates for rigid bodies. Generalized coordinates provide a *local* parameterization of the configuration space for a system of coupled-rigid bodies. An example is given by the Euler angle system used to describe the orientation of a single rigid body in space. This choice of embedding has the disadvantage that an arbitrary choice must be made in the location of the singularity. These coordinates also present problems in accurate integration of systems of coupled-rigid bodies; the amount of time spent by numerical codes in changing coordinates can be comparable to that used in the rest of the integration process. For this reason, we favor other parameterizations described below for rigid systems.

Trace norm. Material coordinates specify the location in space of each rigid body relative to a specified reference configuration. This system of coordinates has the significant advantage in numerical simulation that the inertia tensor is invariant over the configuration space.

For coupled-rigid bodies, it is sufficient to specify the location in \mathbb{R}^3 of three distinct points in each rigid body, with additional constraints on their relative position. The embedding can thus be viewed as a relaxation of the natural constraints imposed by the rigidity of the physical system. If the reduced-order model is required for simulation, then these constraints can be reimposed by the integration algorithm.

For either rigid or flexible structures, one can extend the notion of material coordinates, in the standard method used in finite-element codes, to a specification of the individual particle locations that make up the system.

If we define an inner product on this space as the L_2 inner product between particles, then the inner product can be expressed in terms of the position and orientation as follows. Consider two configurations $q_1, q_2 \in \text{SE}(3)$ of a rigid body D , given by the location of the center of mass $c_i \in \mathbb{R}^3$ and the orientation $R_i \in \text{SO}(3) \subset \mathbb{R}^{3 \times 3}$, with the natural embedding of $\text{SO}(3)$ in the space of 3×3 matrices. Then the L_2 inner

product is given by

$$\begin{aligned} \langle q_1, q_1 \rangle &= \int_{x \in D} (R_1 x + c_1)^* (R_2 x + c_2) dx \\ &= \int_{x \in D} c_1^* R x dx + \int_{x \in D} x^* R_2 c_2 dx \\ &\quad + \int_{x \in D} c_1^* c_2^* dx + \int_{x \in D} R_{ij} R_{ik} x_j x_k dx. \end{aligned} \quad (4)$$

If we define the constant matrix Q by

$$Q_{ij} = \int_{x \in D} x_i x_j dx$$

and let v be the volume of D , then the above inner product becomes

$$\langle q_1, q_1 \rangle = v c_1^* c_2 + \text{Trace}(R_1 Q R_2^*).$$

In this way, any system of (possibly connected) rigid bodies can be embedded in a space consisting of products of $\text{SO}(3)$ and \mathbb{R}^3 , and the above inner product used to compute a subspace decomposition with respect to the L_2 distance between configurations.

Other possible embeddings of $\text{SO}(3)$ in \mathbb{R}^m include the use of quaternions, which we can view as $\text{SU}(2)$, a subset of \mathbb{C}^2 equal to the unit ball in \mathbb{R}^4 . We can also map rotations to the space of skew-symmetric matrices in $\mathbb{R}^{3 \times 3}$, via exponential coordinates, and use the standard inner product on that space.

Constraints. Mechanical systems often involve constraints. In the case of the docking-device example, the base is rigidly clamped. These constraints are *homogeneous*; that is, in displacement coordinates for the finite-element representation, they simply restrict some of the configuration variables to be zero. These are thus linear constraints, which are simple to construct in the linear space V in which the configuration manifold is embedded.

4. Finite-element models of elasticity

The model reduction described below is specialized to three-dimensional elastodynamics, but applies equally well to Lagrangian finite-element models of other types of solids independently of the constitutive equation used.

4.1. System formulation

For finite-element elasticity in three dimensions, the configuration space of spatially discretized mechanics consists of n copies of \mathbb{R}^3 (where n is the number of finite-element nodes). Normally, material coordinates are used for computation, and so the Lagrangian has the form

$$L(q, \dot{q}) = \frac{1}{2} \dot{q}^* M \dot{q} - V(q), \quad (5)$$

where M is the constant mass matrix and V the potential energy. The corresponding Euler–Lagrange equations can be written as

$$M \ddot{q} = -DV(q). \quad (6)$$

Form of the reduced system. We can characterize Q_r as the span of orthonormal basis vectors $\{\phi_1, \dots, \phi_k\}$, and define the $k \times r$ matrix P whose rows are the ϕ_i . Then the constraint is that $(I - P^* P)q = 0$, and the restricted Lagrangian is

$$L_r = \frac{1}{2} \dot{y}^* P M P^* \dot{y} - V(P^* y),$$

where y are the corresponding generalized coordinates on the constraint surface. This leads to the Euler–Lagrange equations for the reduced system

$$P M P^* \ddot{y} = -P D V(P^* y). \quad (7)$$

Note that these are not the same equations as would be obtained by first converting (1) to first-order form (2) on phase space and then applying the Galerkin projection. In fact, the projection operator is not defined on phase space at all; it is only defined on the embedded configuration space.

4.2. Computation

So far, we have not discussed how having a reduced-order model will increase the computational efficiency of simulation. In the case of local material coordinates, the equations of motion for the reduced-order model have the form of Eq. (7).

Both the Euler–Lagrange equations of the original system and those of Eq. (7) contain the forcing term DV . Also, when the nonlinear system of equations is

linearized to use the Newton iteration for solution, the forcing term derivative (Hessian) needs to be evaluated in both the original and reduced system. A concern is therefore that the integration of the reduced system will be at least as computationally intensive as integration of the original system. If the nonlinearities are of known functional form (for example, a quartic radiation term in heat transfer), some of these constructions may be performed only once, ‘by hand’, and it may be possible for instance to explicitly write down an expression for the function $P \circ DV \circ P^*$; see, for instance [1,39].

For general high-dimensional problems, typically computed by discretization methods such as finite elements, the situation is more difficult; the forcing function DV is not known analytically, but instead is evaluated by numerical quadrature in an assembly process at each time-step of the integration.

In fact, even though at first glance the reduced-order system appears at least as complex as the original system, simulation of the reduced system requires significantly less computational effort. We now show how integration methods are tied into the model reduction framework, and how this leads to significant computational savings.

4.3. Newmark integrator

The classical Newmark algorithm is the following. Let γ and β be real numbers, usually chosen between 0 and 1. Then the Newmark method is the map from (q_n, \dot{q}_n) to (q_{n+1}, \dot{q}_{n+1}) , defined by

$$q_{n+1} = q_n + h\dot{q}_n + \frac{1}{2}(h^2)((1 - 2\beta)\ddot{q}_n + 2\beta\ddot{q}_{n+1}), \quad (8)$$

$$\dot{q}_{n+1} = \dot{q}_n + h((1 - \gamma)\ddot{q}_n + \gamma\ddot{q}_{n+1}), \quad (9)$$

wherein q_n , \dot{q}_n and \ddot{q}_n are the discrete configurations, velocities, and accelerations, respectively. The discrete accelerations \ddot{q}_n are defined by the Euler–Lagrange equations

$$M\ddot{q}_n = -DV(q_n). \quad (10)$$

Here h is the time-step of the algorithm. The Newmark algorithm is second-order accurate if and only if $\gamma =$

$1/2$, otherwise the algorithm is only consistent, and so for the remainder of this paper we choose $\gamma = 1/2$.

The Newmark integrator has recently been shown to be *variational*, for all values of β , by Kane et al. [13]. Because of the variational structure, this numerical integration method is given by the solution to a variational principle in discrete-time, and it has the property that it exactly conserves momentum and the symplectic form.

For explicit integration, significant computational savings may be obtained simply by the increase in stable time-step allowed by the reduced system, in which the often high-frequency and fine-scale dynamics of the original system have been removed. However, for implicit integration, computational savings are obtained more directly, and it is this case we consider in the next section.

4.4. Implicit integration

At each time-step, given (q_n, \dot{q}_n) we have to solve the implicit equation

$$R(q_{n+1}) = 0,$$

where the *residual* $R(q_{n+1})$ is defined by

$$R(q_{n+1}) := \frac{1}{h^2\beta}Mq_{n+1} + DV(q_{n+1}) - \frac{1}{h^2\beta}M(q_n + h\dot{q}_n) + \frac{1 - 2\beta}{2\beta}DV(q_n). \quad (11)$$

As is standard, the solution of this equation is approximated via Newton–Raphson iteration. We denote the approximation to q_{n+1} at each iteration by q_{n+1}^i , and write the iteration as a map from q_{n+1}^i to q_{n+1}^{i+1} , with initial guess $q_{n+1}^0 = q_n$. The Newton–Raphson iteration procedure at each time-step of the Newmark integrator is then

$$q_{n+1}^{i+1} = q_{n+1}^i + \Delta q_{n+1}^i,$$

where Δq_{n+1}^i is the solution to the linear equation

$$DR(q_{n+1}^i)\Delta q_{n+1}^i = -R(q_{n+1}^i). \quad (12)$$

The gradient $DR(q_{n+1}^i)$ is called the *effective stiffness* and is given by

$$DR(q) = \frac{1}{h^2\beta}M + D^2V(q)$$

for $q \in Q$.

Computational requirements for the original system. In general, for each q^i this matrix is sparse, and therefore the asymptotic growth rate of computation time with respect to the number of degrees of freedom n for solution of the linear system of equations (12) is $O[s(n)]$, where $s(n)$ depends on the solver type and the particular sparseness properties of M and D^2V .

Several other computations are required; the stiffness D^2V and the internal forcing term DV are both assembled elementwise by the numerical algorithm, and the predictor equation (9) is evaluated. All of these steps are asymptotically linear in the number of degrees of freedom. Therefore the total asymptotic computational cost is dominated by the linear system solver.

Computational requirements for the reduced system. The Lagrangian for the reduced-order system has the form of Eq. (5), with the corresponding Euler–Lagrange equations of the form of (7). Application of the Newmark integration algorithm to this equation gives the residual

$$\begin{aligned} \tilde{R}(y_{n+1}) := & \frac{1}{h^2\beta}PMP^*y_{n+1} + PDV(P^*y_{n+1}) \\ & - \frac{1}{h^2\beta}PMP^*(y_n + h\dot{y}_n) \\ & + \frac{1 - 2\beta}{2\beta}PDV(P^*y_n). \end{aligned} \tag{13}$$

The corresponding Newton–Raphson iteration requires solution of the linear equation

$$D\tilde{R}(y_{n+1}^i)\Delta y_{n+1}^i = -\tilde{R}(y_{n+1}^i) \tag{14}$$

with the effective stiffness given by

$$D\tilde{R}(y) = \frac{1}{h^2\beta}PMP^* + PD^2V(P^*y)P^*$$

for $y \in Q_r$.

The effective stiffness of the reduced system is now only a $(k \times k)$ -dimensional matrix, where k is the dimension of the reduced system. However, it is now a

Table 1
Asymptotic computational costs for integration

Operation	Original system	Reduced system
Force assembly	$O(n)$	$O(n) + O(nk)$
Effective stiffness assembly	$O(n)$	$O(n) + O(nk^2)$
Linear solver	$O[s(n)]$	$O(k^3)$

full matrix, and so the asymptotic cost for solution of this linear equation grows as $O(k^3)$.

The effective stiffness matrix is still assembled elementwise, so there is no need to assemble and store the matrices M and D^2V and then perform the conjugate transformation associated with P . In this way, storage requirements are reduced if k^3 is less than the number of elements of M , which is typical. However, assembly of the elementwise effective stiffness has asymptotic computational time growth of $O(nk^2)$. Computation of the internal forcing term $PDV(P^*y)$ has asymptotic growth as $O(nk)$, due to the multiplication by P and P^* .

A summary of the asymptotic costs for the original and reduced systems is shown in Table 1. This table shows that the computational cost of simulating the full-order system grows as $O[s(n)] \geq O(n)$. For instance, the cost of direct solution of a banded system grows as $s(n) = n^2$. For iterative solvers, each iteration costs $O(n)$, and the number of iterations depends on n , so again $s(n) \geq n$. On the other hand, the computational cost of simulating the reduced-order system grows *linearly* in n . Thus, for larger systems, we expect to obtain considerable computational savings.

Here we make the reasonable assumption that k does not grow with n . This is consistent with our intuitive understanding of the ‘mode shapes’ described by P ; it is reasonable to suppose that, as n increases, these mode shapes converge to appropriate basis functions for the continuum system.

Since a very fine discretization may be necessary to capture the effects of the fine-scale dynamics on these coarse scale dynamics, a high-resolution model may be needed for the original system. This does *not* mean, however, that the reduced-order model need be of high dimension. The number of states required to approximately describe the important dynamics of the

system is a function of the dynamics, not of the discretization of the full-order system.

Note that computational costs are reduced for only part of the integration; it is only the linear solver where time is saved. For the effective stiffness assembly, the computational costs of the reduced system are in fact slightly increased. However, the savings more than outweigh the extra costs, and we will use the docking-device example to illustrate this.

4.5. Explicit integration

As mentioned above, computational savings can be achieved when using explicit time integration through an increase in the stable time-step. For the original system, the accelerations \ddot{q}_n are solved using Eq. (10) (for explicit integration, M is a diagonal matrix). For the reduced model, the accelerations \ddot{q}_n are solved using

$$\ddot{q}_n = -P\{[(PMP^*)^{-1}P^*]DV(q_n)\}, \quad (15)$$

where the parentheses indicate the order of evaluation. The matrix $(PMP^*)^{-1}P^*$ is precomputed before the time-stepping starts.

4.6. Parallel processing

It is worth noticing that the reduced model may be parallelized very effectively. All the finite-element operations (evaluation of the matrix $M + D^2V$ and computation of the unbalanced force vector) may be performed in parallel using either shared- or distributed-memory paradigms. The solution of the linear equations may then be performed serially, or a parallel dense matrix solver may be used. In any case, since the solution of the linear system represents only a very small fraction of the total CPU time during each step, the parallel efficiency and scalability is likely to be high.

4.7. Dynamic update of the basis

If the basis set does not change during the time integration, i.e., if no basis function is added to or subtracted from the set of basis vectors $\{\phi_1, \dots, \phi_k\}$, the

solution $q_n, \dot{q}_n, \ddot{q}_n$ at any time instant may be written as a linear combination of the basis functions ϕ_j . Then the time integration algorithm may be formulated either in terms of the configuration variables q , or in terms of the generalized coordinates y . Note that, because of the orthonormality of the set of ϕ_j , the transformation between the q variables and the y variables is unique.

Changing the basis at any time instant effectively leads to an introduction of time-dependent constraints. The above correspondence between the configuration coordinates q and the generalized coordinates y no longer holds. Moreover, the effect the time-dependent constraints have on the mechanical properties of the reduced system needs to be carefully considered. In particular, even if the energy of full system is conserved, the energy of the reduced-order model may now vary over time.

One possible benefit of a changing basis is that it allows use of an adaptive procedure, which would select from the complete set of basis functions ϕ_j only those that lead to an optimal cost vs. accuracy ratio, or which would generate additional functions on demand as indicated by error analysis.

Note that the basis might be changed either at the beginning of a time-step, or each equilibrium iteration (14) may be performed with a different basis. If a varying basis is used, the time integration is probably best carried in the configuration variables q .

4.8. Computational details for the docking device

As described in Section 1, the docking device is an example of finite-element analysis for a problem involving three-dimensional nonlinear elasticity. Both the steel and the rubber are handled as hyperelastic materials. The St. Venant constitutive equation adopted here is a reasonable approximation for the small to moderate strains occurring during the simulation. Two faces of the rubber block are clamped (homogeneous boundary condition). The massive frame is given initial velocity at 45° to the major axis of the assembly.

The original finite-element model has 9600 degrees of freedom, and a reduced system was constructed via the KLE consisting of 36 degrees of freedom. A 1 s

simulation was run, which corresponds to about seven swings of the frame.

The implicit Newmark method was used for the full finite-element model, with $\gamma = 1/2$, $\beta = 1/4$ and a time-step of 1.4×10^{-3} . The average computation time per time-step was 120 s, with the residual and the effective stiffness assembly, and the linear equation solver using per iteration 0.3, 12, and 22 s, respectively. The total time for the simulation for 1 s of motion was 24 CPU hours.

The reduced model with 36 modes was also integrated with an implicit Newmark algorithm, using the same time-step as the full system. The residual assembly took 0.4 s, effective stiffness assembly took 14.9 s, and the solver took 0.001 s; the other operations were essentially identical to those for the full system. The total time for a 1 s simulation was 10 CPU hours. As can be seen, the computation time has been reduced by roughly 60% by almost totally eliminating the solver cost. It should be realized that the finite-element model described here is rather small; still, its solution represents a major expense. The asymptotic estimates of Table 1 indicate that the larger the full finite-element model, the more significant the savings due to the use of a reduced model.

5. Future opportunities

5.1. Improvements

It is clear that the procedures outlined here can be improved in many ways that need further exploration for future applications.

Symmetry and travelling bases. Experience has shown that if a system has symmetry, then using travelling Karhunen–Loève bases can dramatically cut down the dimension of the reduced system; see, for example, [8,36] and references therein. It is likewise well understood that discrete symmetries are also important to take into account when building bases. Thus, for a problem with symmetry there is plenty of opportunity for further reductions. Even if the problem does not have symmetry, one can expect that travelling bases for propagating structures and

approximate discrete symmetries to also be useful. This approach has proven to be very powerful for other symmetries as well, including important scaling symmetries; cf. [2]. (These remarks are based on work of Kevrekidis, Rowley and others in progress.)

Space–time decompositions. More generally, one can expect that any time there is more structure in a part of a system (in a space–time sense), one should put more effort into construction of the basis in that part. The theory and implementation for procedures of this sort are yet to be worked out. These ideas are also consistent with parallel efforts to develop space–time adaptive numerical codes such as those that are built around multisymplectic integrators [23].

5.2. Conclusions, comments, and future directions

In this paper we have focussed on giving a context in which model reduction and mechanics are compatible. One could also ask why we did not do this for other reduction techniques, such as inertial manifolds. This is easy: for Hamiltonian or nearly Hamiltonian PDEs, such as occur in ocean dynamics, the spectrum is on or near the imaginary axis and so any inertial manifold would be either infinite or very high dimensional. Despite this there may be large coherent structures and one wants to capture them with low-dimensional models. The distinction between singular values and eigenvalues in this situation plays, as is well appreciated in the literature, a critical role.

Note that [28] has shown the sense in which Hamiltonian structures restricted to center manifolds retain this structure. Also, note that when Hamiltonian systems governed by an action principle are averaged, as in averaged fluid models, then one gets a simpler model that also retains a Hamiltonian or variational structure (see, for instance, [7,29]). However, the link with either model reduction or with center manifold reduction is not clear. Alan Newell already raised questions like this over 15 years ago.

There are a number of interesting avenues to explore from the foundation we have laid. In particular, application of the geometric methods in this paper to two-dimensional thin-shell problems, in order to achieve a mechanically consistent reduced-order

model, is still a challenging computational task with standard meshing techniques. This is an area where tying together subdivision surfaces and model reduction may lead to significant benefits.

Related to balancing, model reduction which can incorporate the dependence of system behavior on optimization parameters allows integration with optimal parametric design methods.

A fundamental next step is to understand how to select the number of modes for the reduced model: when is the response accurate enough for a given design purpose, in the sense of similarity of the full and reduced models? For control, some of these issues have been studied in detail, although as yet there is little research connecting these ideas to continuum mechanics.

We would also like to make use of model reduction techniques as a way of implementing mesh refinement and coarsening methods. Further research is also necessary in order to understand how to take account of other system features, such as symmetry, which may also lead to greater computational savings.

It would be quite interesting to put the theory here into a more hierarchical and adaptive context by making use of the recent developments of PDE asynchronous, multisymplectic integrators given in [20] and the CHARMs (conforming hierarchical adaptive refinement methods) methodology given in [15,16]. This would hopefully lead to a SPAHMR (structure-preserving adaptive hierarchical model reduction; pronounced “Spammer”) methodology that would also automatically deal with any symmetries that are present in a given problem.

Acknowledgements

We thank Ronald Coifman, John Doyle, Darryl Holm, Yannis Kevrekidis, Clancy Rowley, and Peter Schröder for helpful comments and inspiration.

Appendix A. The Karhunen–Loève decomposition

In this appendix, we give for completeness a description of the KLE, which was used for the model

reduction example in this paper. This is only one of the several possible approximation or optimization methods that may be used in conjunction with the model reduction procedure in this paper; other such methods include balancing transformations and wavelet expansions.

The KLE provides an optimization-based method for finding a subspace on which to construct reduced-order dynamics. The method is known in the literature by several names, including principal component analysis, factor analysis, and total least-squares estimation. The method has been extensively analyzed in the literature, although the original concept goes back to Pearson [32]. We give here a brief outline of the method; for details, see [11].

In practical application, the method makes essential use of *empirical data*, taken either from experiments or from numerical simulation.

For systems of the form (2), the data consists of sampled measurements $\{x^{(1)}, \dots, x^{(N)}\}$ of $x(t)$. For mechanical systems, we make use of the embedding procedure, and each measurement of $q \in Q$ corresponds to a unique $x \in \mathbb{R}^n$.

The next step is to perform a principal component analysis of this data, to find how well it may be approximated by projection onto a k -dimensional subspace of the original n -dimensional state space.

We can characterize the subspace $S \subset \mathbb{R}^n$ by the projection operator Q mapping \mathbb{R}^n onto S . We would like to find Q to minimize

$$H(Q) = \sum_{i=1}^N \|x^{(i)} - Qx^{(i)}\|_2^2$$

the total squared perpendicular distance of the points from the k -plane. The following result is standard.

Theorem A.1 (Total least-squares). *Let R be the correlation matrix of the data, defined by*

$$R := \sum_{i=1}^N x^{(i)} x^{(i)*}$$

and let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ be the ordered eigenvalues of R . Then

$$\min_Q H(Q) = \sum_{i=n-k+1}^n \lambda_i,$$

where the minimum is over all rank k projections Q .

In general, R may not have rank n , if the given data lies within a strict subspace of \mathbb{R}^n . Let $s = \text{Rank } R$, and let $\phi_1, \phi_2, \dots, \phi_s$ be orthonormal eigenvectors of R , corresponding to the non-zero λ_i . Each $x^{(i)}$ can be written as

$$x^{(i)} = \sum_{j=1}^s a_{ij} \phi_j,$$

where $a_{ij} = \langle x^{(i)}, \phi_j \rangle$, and $\langle \phi_i, \phi_j \rangle = \delta_{ij}$. The optimal k -dimensional subspace approximant is given by

$$\hat{x}^{(i)} = \sum_{j=1}^k a_{ij} \phi_j.$$

Denote by P the $k \times n$ matrix whose rows are ϕ_1, \dots, ϕ_k , so that $PP^* = I$. The projected approximant to x is given by $P^*Px \in S$, and $y = Px$ is a representation in terms of the new coordinates ϕ_i on S . This subspace approximant is then optimal, in the sense that the total energy (2-norm) in the subspace is given by

$$\sum_{i=1}^N \|Px^{(i)}\|_2^2 = \sum_{j=1}^k \lambda_j$$

and this is the maximum achievable by any k -plane.

The above procedure finds the optimal subspace, passing through the origin defined by the inner product. To find the optimal *affine* subspace, we use the fact that this must pass through the mean of the data, and normalize the data by subtracting the mean before constructing the correlation matrix.

The eigenvalues of R tell us how close an approximation of the data is provided by a k -dimensional subspace; the goal is to choose k such that the fraction of the total 2-norm in the subspace

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i}$$

is close to 1, yet k is sufficiently small. Clearly this will not always be possible, with models which are

better approximated by low-dimensional systems having relatively few large eigenvalues.

The Karhunen–Loève method therefore picks the subspace containing the interesting dynamics of the system. In general, we would expect that the more eigenvectors we keep, the better approximation we will obtain.

Computationally, application of this method requires only standard matrix computations, despite its application to nonlinear systems.

We can expect such a procedure to work well for model reduction of the system within some given region of state space, and it is within such a specific region that data should be collected. In doing this, one must pay careful attention to symmetry, as in [8] and references therein.

References

- [1] S. Banerjee, J.V. Cole, K.F. Jensen, Nonlinear model reduction strategies for rapid thermal processing systems, *IEEE Trans. Semicond. Manuf.* 11 (1998) 266–275.
- [2] G.I. Barenblatt, *Scaling, Self-similarity, and Intermediate Asymptotics*, Cambridge University Press, Cambridge, 1996.
- [3] A. Bloch, Estimation, principal components, and Hamiltonian systems, *Syst. Contr. Lett.* 6 (1985) 103–108.
- [4] H. Cendra, J.E. Marsden, T.S. Ratiu, *Lagrangian Reduction by Stages*, vol. 152 of *Memoirs*, American Mathematical Society, Providence, RI, 2001.
- [5] F. Cirak, M. Ortiz, P. Schröder, Subdivision surfaces: a new paradigm for thin-shell finite element analysis, *Int. J. Numer. Meth. Eng.* 47 (2000) 2039–2072.
- [6] D. Enns, *Model reduction for control system design*, Ph.D. Thesis, Stanford University, 1984.
- [7] N. Ercolani, M.G. Forest, D.W. Mclaughlin, R. Montgomery, Hamiltonian structure of modulation equation for the sine-Gordon equation, *Duke Math. J.* 55 (1987) 949–983.
- [8] S. Glavaški, J.E. Marsden, R.M. Murray, Model reduction, centering and the Karhunen–Loève expansion, *Proc. CDC* 37 (1998) 2071–2076.
- [9] K. Glover, All optimal Hankel-norm approximations of linear multivariable systems and their L-infinity error bounds, *Int. J. Contr.* 39 (1984) 1115–1193.
- [10] D.D. Holm, J.E. Marsden, T.S. Ratiu, The Euler–Poincaré equations and semidirect products with applications to continuum theories, *Adv. Math.* 137 (1998) 1–81.
- [11] P.J. Holmes, J.L. Lumley, G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, 1996.
- [12] C. Jen, D.A. Johnson, R. Gorez, A reduced-order dynamic model for end-effector position control of a flexible robot arm, *Math. Comput. Simul.* 41 (1996) 539–558.

- [13] C. Kane, J.E. Marsden, M. Ortiz, M. West, Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems, *Int. J. Numer. Meth. Eng.* 49 (2000) 1295–1325.
- [14] K. Karhunen, Zur spektraltheorie stochastischer Prozesse, *Ann. Acad. Sci. Fenn.* 34 (1946).
- [15] P. Krysl, S. Lall, J.E. Marsden, Dimensional model reduction in non-linear finite element dynamics of solids and structures, *Int. J. Numer. Meth. Eng.* 51 (2001) 479–504.
- [16] P. Krysl, E. Grinspun, P. Schröder, Natural hierarchical refinement for finite element methods, *Int. J. Numer. Meth. Eng.* 56 (2003) 1109–1124.
- [17] S. Lall, J.E. Marsden, S. Glavaski, Empirical model reduction of controlled nonlinear systems, *Proc. IFAC World Congr. F* (1999) 473–478.
- [18] S. Lall, J.E. Marsden, S. Glavaski, A subspace approach to balanced truncation for model reduction of nonlinear control systems, *Int. J. Robust Nonlinear Contr.* 12 (2002) 519–535.
- [19] P. Leger, S. Dussault, Nonlinear seismic response analysis using vector superposition methods, *Earthquake Eng. Struct. Dyn.* 21 (1992) 163–176.
- [20] A. Lew, J.E. Marsden, M. Ortiz, M. West, Asynchronous variational integrators, *Arch. Rat. Mech. Anal.* 167 (2003) 85–146.
- [21] M. Loève, Fonctions aléatoire de second order, *C.R. Acad. Sci. Paris* 220 (1945).
- [22] J.L. Lumley, The structure of inhomogeneous turbulence, in: A.M. Yaglom, V.I. Tatarski (Eds.), *Atmospheric Turbulence and Wave Propagation*, Nauka, Moscow, 1967, pp. 166–178.
- [23] J.E. Marsden, G.W. Patrick, S. Shkoller, Multisymplectic geometry, variational integrators and nonlinear PDEs, *Commun. Math. Phys.* 199 (1998) 351–395.
- [24] J.E. Marsden, T.S. Ratiu, *Introduction to Mechanics and Symmetry*, Texts in Applied Mathematics, vol. 17, 2nd ed., 1994, Springer, Berlin, 1999.
- [25] J.E. Marsden, T. Ratiu, S. Shkoller, The geometry and analysis of the averaged Euler equations and a new diffeomorphism group, *Geom. Funct. Anal.* 10 (2000) 582–599.
- [26] J.E. Marsden, S. Shkoller, The anisotropic averaged Euler equations, *Arch. Rat. Mech. Anal.* 166 (2003) 27–46.
- [27] J.E. Marsden, M. West, Discrete mechanics and variational integrators, *Acta Numer.* 10 (2001) 357–514.
- [28] A. Mielke, *Hamiltonian and Lagrangian Flows on Center Manifolds, with Applications to Elliptic Variational Problems*, Springer Lecture Notes in Mathematics, vol. 1489, 1991.
- [29] B. Moore, Principal component analysis in linear systems: controllability, observability, and model reduction, *IEEE Trans. Automat. Contr.* 26 (1981) 17–32.
- [30] R. Murray, S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Boca Raton, FL, 1994.
- [31] R.E. Nickell, Nonlinear dynamics by mode superposition, *Comput. Meth. Appl. Mech. Eng.* 7 (1976) 107–129.
- [32] K. Pearson, On lines and planes of closest fit to points in space, *Philos. Mag.* 2 (1901) 609–629.
- [33] R. Radovitzky, M. Ortiz, Error estimation and adaptive meshing in strongly nonlinear dynamic problems, *Comput. Meth. Appl. Mech. Eng.* 172 (1999) 203–240.
- [34] C.W. Rowley, T. Colonius, A.J. Basu, On self-sustained oscillations in flows over rectangular cavities, *J. Fluid Mech.* 455 (2002) 315–346.
- [35] C.W. Rowley, T. Colonius, R.M. Murray, POD based models of self-sustained oscillations in the flow past an open cavity, in: *Proceedings of the Sixth AIAA/CEAS Aeroacoustics Conference*, June 2000, AIAA Paper No. 2000-1969.
- [36] C.W. Rowley, J.E. Marsden, Reconstruction equations and the Karhunen–Loève expansion for systems with symmetry, *Physica D* 142 (2000) 1–19.
- [37] J.M. Scherpen, H -infinity balancing for nonlinear systems, *Int. J. Robust Nonlinear Contr.* 6 (1996) 645–668.
- [38] E. Wilson, M.W. Yuan, J.M. Dickens, Dynamic analysis by direct superposition of Ritz vectors, *Earthquake Eng. Struct. Dyn.* 10 (1982) 813–821.
- [39] M. Yao, Nonlinear structural dynamic finite element analysis using Ritz vector reduced basis method, *Shock Vib.* 3 (1996) 259–268.