47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Confere
1 - 4 May 2006, Newport, Rhode Island

AIAA 2006-2145

2nd AIAA Multidisciplinary Design Optimization Specialist Conference, May 1 - 4, 2006, Newport, RI

Multi-point Extended Reduced Order Modeling For Design Optimization and Uncertainty Analysis

G. Weickum*, M.S. Eldred†, and K. Maute‡

For large computational models, standard deterministic optimization approaches can be prohibitively expensive due to the need to repeatedly evaluate the model. This difficulty is amplified when stochastic aspects of the model are included, such as in reliability based design optimization. This work seeks to alleviate the computational costs of analyzing dynamic systems through employing a surrogate model in place of the full model analysis. The surrogate model of interest is a reduced-order model (ROM), which employs a Galerkin projection of the system response using a computed set of basis functions in order to significantly reduce the number of degrees of freedom in the system. The ROM techniques presented will not only be able to approximate the response accurately at the nominal design, with a significant reduction in computation cost, but will also estimate the response due to a change in design or uncertain variable parameters. Two conceptual approaches will be explored: extended reduced order modeling (EROM) and spanning reduced order modeling (SROM). The difference between these methods is an EROM approximates an updated basis (eigenmodes, singular vectors) for each change in parameters while an SROM uses a single spanning basis for the full parameter space. Different techniques for the computation of the EROM and SROM bases are explored, and the best of these methods are incorporated into the optimization and stochastic analysis of a structural problem, demonstrating the benefit of ROMs.

I. Introduction

In the past decade, numerical optimization methods have attracted significant interest from the design engineering community. These methods allow the systematic integration of computational models into the design process exploiting the advances in numerical simulation techniques for design purposes. Design optimization approaches predicting the performance of the engineering systems by high-fidelity nonlinear single and coupled multi-disciplinary models have been applied to a broad range of engineering design problems. Furthermore, stochastic variations of system parameters and operating conditions can be taken into account in the formulation of the design optimization problem. By considering uncertainties directly within computational design, one can make the design either more robust and/or more reliable. The stochastic response of the system can be quantified, for example, by sampling techniques, approximate reliability methods, and stochastic projection schemes. Common to all these methods is they require a repeated evaluation of the simulation models for various configurations of the system.

At the same time as the complexity and accuracy of these models increase, the computational costs for analyzing the systems performance increases. In particular if the stochastic and reliability-based performance criteria are evaluated, using high-fidelity simulation techniques leads to unacceptable computational costs.

*Graduate student, Student Member AIAA. Center for Aerospace Structures, Department of Aerospace Engineering Sciences, University of Colorado at Boulder, Boulder, CO 80309, USA.

†Associate Fellow AIAA. Optimization and Uncertainty Estimation Department, Sandia National Laboratories, Albuquerque, NM 87185-0370. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

‡Assistant Professor, Member AIAA. Center for Aerospace Structures, Department of Aerospace Engineering Sciences, University of Colorado at Boulder, Boulder, CO 80309, USA.

Copyright © 2005 by the American Institute of Aeronautics and Astronautics, Inc. The U.S. Government has a royalty-free license to exercise all rights under the copyright claimed herein for Governmental purposes. All other rights are reserved by the copyright owner.

Copyright © 2006 by Kurt Maute. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission.

Both the optimization procedure and the stochastic analysis method require multiple analyses of altered system states in a parameter space, shown in figure 1. The parameters are either design and/or random variables.

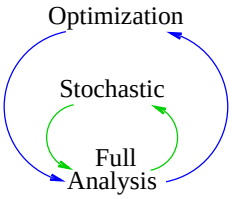

Figure 1. General schematic of an optimization/stochastic process

The key to incorporating a computationally expensive simulation technique into a stochastic design framework is to decrease the expense of analyzing systems altered in the parameter space. Two approaches are frequently used to achieve this goal. Following traditional engineering practice, simplified physical models can be used to reduce the model complexity and the computational costs. For example, a wing can be approximated by a beam model. While this approach is very efficient for predicting some aspects of the system response, simplified models are often not able to predict a broad range of the performance criteria. Therefore, multiple simplified models need to be developed and calibrated to capture all performance criteria of interest, requiring a thorough understanding of the engineering system. For complex systems dominated, for example, by nonlinear phenomena, coupling of multiple physical fields, and/or strong transient effects, simplified physical models often approximate the response inadequately. In addition, engineering intuition is often required to connect the parameters of the simplified models to the design parameters of the real system.

Alternatively, surrogate models have been developed allowing the approximation of the system response as a function of the design parameters based on performance predictions from high-fidelity simulation models. Surrogate models may be broadly characterized as data fit (local, multi-point, or global approximations), multi-fidelity (omitted physics, coarsened discretization or tolerances), or reduced order model (ROM) surrogates. A ROM mathematically reduces the system modeled, while still capturing the physics of the governing partial differential equations (PDEs), by projecting the original system response using a computed set of basis functions. The projection reduces the number of degrees of freedom (DOFs) in a large finite element model ($O(10^4 - 10^9)$ DOF) down to a handful of basis coordinates (typically $O(10^0 - 10^1)$). Thus, the ROM case is distinguished from the data fit case in that it is still intimately tied to the original PDEs and retains their physics, and is distinguished from the multi-fidelity case in it is derived directly from the original high fidelity model and does not require multiple models of differing fidelity. ROM models have proven a successful means of reducing the computational costs of a system's response in time.[1–4] However, ROMs typically approximate the response of only one particular configuration and are therefore of limited use for design optimization and stochastic analysis purposes.

The main emphasis of the paper is the extension of ROMs for capturing the systems response in the parameter space. This paper will first describe the basic framework of ROMs for the dynamic analysis of linear structural response. The approach is based on a Galerkin projection scheme with the basis vectors being proper orthogonal decomposition (POD) modes. The proposed extensions of ROMs are described in Section 3. In Section 4 a numerical study on a connecting rod will be presented comparing the accuracy of various approximation methods. The best of these approaches will be used within optimization and uncertainty quantification to compare performance against the full model.

II. Background

The overall approach is to use the ROM as a surrogate in the place of the full order model. The surrogate is built from information extracted from the full order model, figure 2, and is then interfaced with the optimization or uncertainty quantification method. By using the surrogate, the computational time is significantly reduced.

In general, the surrogate model has a limited range of validity and does not necessarily approximate the whole design space. A trust region framework, such as the ones shown by Giunta and Eldred[5] and Eldred

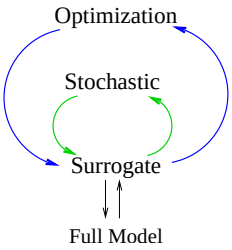

Figure 2. How surrogate model is incorporated into the optimization process

et al.,[6] can be used within optimization to adaptively manage the range of surrogate accuracy. Eq. 1 shows the trust region approximate subproblem:

$$\begin{aligned} \text{minimize} \quad & \hat{f}^k(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u \\ & \parallel \mathbf{x} - \mathbf{x}_c^k \parallel_\infty \leq \Delta^k, \end{aligned} \tag{1}$$

where $\mathbf{x} \in \Re^n$ is the vector of design variables, $\hat{f}^k(\mathbf{x})$ is the surrogate model, $\mathbf{x}_c$ is the center point of the trust region, and Δ^k is the current size of the trust region. After defining an initial trust region, a surrogate is built for the trust region and optimization is performed on the surrogate within the trust region. After completion of the approximate optimization, the ratio of actual improvement in the full order model to the predicted improvement in the surrogate model is calculated. Depending on the ratio, the iteration is accepted or rejected and the trust region is shrunk, expanded, or retained. A new surrogate is then constructed for the new trust region and the process is repeated until convergence.

A. Reduced Order Models

In this section, we will discuss the construction of the ROM surrogate model for a dynamic analysis of a finite element model. The governing equation of interest is a structural dynamic response (2), where M is the mass matrix, $\boldsymbol{F}^{int}$ is the internal force, $\boldsymbol{u}$ is the displacement, and the external force is $\boldsymbol{f}^{ext}(\boldsymbol{u}, t)$.

$$\boldsymbol{M}\ddot{\boldsymbol{u}} + \boldsymbol{F}^{int}(\boldsymbol{u}, \dot{\boldsymbol{u}}) = \boldsymbol{f}^{ext}(\boldsymbol{u}, t) \tag{2}$$

The dynamic response is either linear or nonlinear depending on the internal forces $\boldsymbol{F}^{int}$ and the external forcing function, $\boldsymbol{f}^{ext}(\boldsymbol{u}, t)$, which can be dependent on time (t) as well as displacement $(\boldsymbol{u})$.

For large systems, the calculation of the linear dynamic response is costly and the cost is further increased for nonlinear systems. The cost of the dynamic response is reduced by using a reduced order model, which is a low dimensional approximation. Following a Galerkin type projection scheme, the displacements of the system response are approximated by k basis vectors (Φ) and generalized variables η as shown below:

$$\boldsymbol{u}(t) = \sum_{j=1}^{k} \eta_j(t)\phi_{(j)} = \Phi\boldsymbol{\eta}(t) \tag{3}$$

The reduction of the dynamic response is performed by using the approximation from (3) in (2) and pre-multiplying by Φ^T as shown below.

$$\Phi^T\boldsymbol{M}\Phi\ddot{\boldsymbol{\eta}}(t) + \Phi^T\boldsymbol{F}^{int}\Phi(\boldsymbol{\eta}(t), \dot{\boldsymbol{\eta}}(t)) = \Phi^T\boldsymbol{f}^{ext}(\boldsymbol{u}, t) \tag{4}$$

The system, originally $n \times n$, is reduced to a $k \times k$ system, where k are the number of basis vectors used and $k \ll n$. The force vectors are reduced from $n \times 1$ for the full system to $k \times 1$ for the reduced system. The reduced system may be written as:

$$\boldsymbol{M}_R\ddot{\boldsymbol{\eta}} + \boldsymbol{F}_R^{int}(\boldsymbol{\eta}, \dot{\boldsymbol{\eta}}) = \boldsymbol{f}_R^{ext}(\boldsymbol{\eta}, t) \tag{5}$$

where $\boldsymbol{M}_R$, $\boldsymbol{F}_R^{int}$, and $\boldsymbol{f}_R^{ext}$ are dependent upon the basis vectors and the system matrices for a particular design.

For linear systems, $\boldsymbol{F}_R^{int}(\boldsymbol{\eta}, \dot{\boldsymbol{\eta}})$ and $\boldsymbol{f}_R^{ext}(\eta, t)$ are only linear functions of η and can be calculated as follows:

$$\boldsymbol{F}_R^{int}(\boldsymbol{\eta}, \dot{\boldsymbol{\eta}}) = \Phi^T \boldsymbol{K}^{int} \Phi\ \boldsymbol{\eta}(t) + \Phi^T \boldsymbol{D} \Phi\ \dot{\boldsymbol{\eta}}(t) \quad (6)$$

$$\boldsymbol{f}_R^{ext}(\boldsymbol{\eta}, t) = \Phi^T \boldsymbol{f}^{ext}(t) + \Phi^T \boldsymbol{K}^{ext} \Phi\ \boldsymbol{\eta}(t) \quad (7)$$

where $\boldsymbol{K}^{int}$ and $\boldsymbol{K}^{ext}$ are the stiffness matrices associated with the internal and external forces, and $\boldsymbol{D}$ is the damping matrix accounting for viscous damping.

For nonlinear systems, the internal and external forces are dependent on the displacements and the reduced force matrices can be updated in two ways. The first is to evaluate the forces in the full order model, using the approximation of the displacements from (3) which is called an "on-line" approximation. The second method, "off-line" approximation, is to build the force functions so $\boldsymbol{F}_R^{int}(\boldsymbol{\eta}, \dot{\boldsymbol{\eta}})$ and $\boldsymbol{f}_R^{ext}(\eta, t)$ are only functions of η and t. The advantage of using an on-line approximation is the system response is calculated more accurately at the cost of CPU time. Using an off-line approximation, the CPU time is decreased at the cost of accuracy.

In order to augment the ROM in (5) and capture the change of the system response due changes in system parameters, the system matrices, force functions, and the basis vectors need to be approximated as functions of the parameters. For linear systems, it was shown in a previous study by Allen[7] et al. the system matrices should be recalculated at each design point within the optimization process. First and second order Taylor Series approximations of the system matrices were investigated, but were not able to capture the response accurately.

In this study, two means of updating the basis are considered, namely a Taylor Series and combined approximation, where the latter, shown in the same study,[7] was a better approximation. For the Taylor Series approximation, the basis and gradients of the basis are needed from the nominal design. For combined approximation, the gradients are not required, but can be used to improve the approximation of the basis. Once the basis is approximated for a parameter change, a reduced linear dynamic analysis is performed to obtain the output response. For linear systems, eigenmodes are commonly used as basis vectors. Allen et al.[7] studied the approximation of basis vectors using the eigenmodes and their gradients. For nonlinear system and in this study, proper orthogonal decomposition and the method of snapshots are used for basis generation. Different approximation methods will be explored for either the POD modes themselves or the snapshots taken from the full response.

B. Proper Orthogonal Decomposition

The application of POD to develop ROMs has been demonstrated in a variety of other disciplines.[1–4] POD reduces a large set of multidimensional data, often in the millions, to a smaller system, often in the tens or hundreds. The response of interest, the solution of a dynamic system, is projected onto a set of basis functions (or modes) by defining a set of coefficients which define the contributions of the individual basis vectors to the response.

A POD basis, which is used to approximate the system's response (2), is built from snapshots taken from a full transient analysis. When N snapshots are taken from the transient response, and a covariance matrix ($\boldsymbol{R}$) is built, shown below:

$$R_{ij} = \frac{1}{N} \int_\Omega u_i u_j d\Omega \quad (8)$$

or for the discretized case:

$$\boldsymbol{R} = (\boldsymbol{u} - \bar{\boldsymbol{u}})^{\boldsymbol{T}} \tilde{\boldsymbol{M}} (\boldsymbol{u} - \bar{\boldsymbol{u}})$$

$$\bar{\boldsymbol{u}} = \frac{1}{N} \sum u_j$$

where $\boldsymbol{u}$ are the snapshots from the dynamic response, Ω represents the body, and $\tilde{\boldsymbol{M}}$ is the mass matrix with unity density. When finding the covariance matrix, the mean of the snapshots are subtracted from the snapshots. The information of interest is not contained in the mean of the snapshots. The reader may note the covariance matrix defined above is different from the traditional definition, $\boldsymbol{R} = \boldsymbol{u}\tilde{\boldsymbol{M}}\boldsymbol{u}^{\boldsymbol{T}}$, which results in a $n \times n$ matrix. However, the matrix defined in (8) is preferred since its dimension is only $k \times k$, where k represents the number of snapshots considered. An eigen analysis is performed on this smaller covariance

matrix and the modes Φ are solved by:

$$\Phi^K = \sum_{j=1}^{K} a_i^K u_j \qquad K = 1, 2, \cdots, N \tag{9}$$

where a are the eigenvectors of $\boldsymbol{R}$. The POD basis (Φ) is not guaranteed orthogonal to the mass and stiffness matrix, as are the eigenmodes of the system, and are dependent upon the time interval and forcing function utilized in the time integration from which the snapshots were obtained. The POD modes attempt to capture as much of the system response as possible in the sense of a vector norm, and may or may not reflect the eigenmodes.

The up-front cost of using POD modes as a basis is the full system analysis required to obtain the snapshots, whereas using eigenmodes requires solving the eigen problem. The choice between which modes to use is dependent upon the type of problem, and the difference in cost between a full system time integration and an eigen analysis. For fluid or multi-physics problems, for example, single time integration is often significantly cheaper than an eigen analysis, so POD modes are preferred.

1. Sensitivity of POD Modes

The derivatives of the POD modes with respect to the design variables are needed to build a first order approximation of a design change from the nominal design. An analytical approach is used to approximate the derivatives of the POD modes with respect the design parameter p. First the POD eigen system is differentiated with respect to the system parameter p:

$$\frac{d}{dp}\left(\mathbf{C} - \omega_i^2 \mathbf{I}\right) \varphi_i = 0 \tag{10}$$

$$\mathbf{C} = \frac{\mathbf{u^T M u}}{N} \tag{11}$$

M is the mass matrix with unity density, ($\mathbf{u}$) is the matrix of snapshots taken from the system response, ω_i^2 are the eigenvalues, φ_i are the eigenvectors, and N is the number of snapshots taken. Applying the product rule, equation (10) results in:

$$\left(\mathbf{C} - \omega_i^2 \mathbf{I}\right) \frac{d\varphi_i}{dp} + \left(\frac{\frac{\partial u^T}{\partial p}\mathbf{M}u + u^T \frac{\partial \mathbf{M}}{\partial p} u + u^T \mathbf{M} \frac{\partial u^T}{\partial p}}{N} - 2\omega_i \frac{\partial \omega_i}{\partial p}\mathbf{I} \right) \varphi_i = 0 \tag{12}$$

Premultiplying (12) by φ_i^T eliminates the first term due to symmetry in C and results in:

$$\varphi_i^T \left(\frac{\partial u^T}{\partial p}\mathbf{M}u + u^T \frac{\partial \mathbf{M}}{\partial p} u + u^T \mathbf{M} \frac{\partial u^T}{\partial p} \right) \varphi_i = 2N\omega_i \varphi_i^T \mathbf{I} \varphi_i \frac{\partial \omega_i}{\partial p} \tag{13}$$

Using (13), the gradients of the eigenvalues with respect to the system parameter are found as:

$$\frac{d\omega_i}{dp} = \frac{\varphi_i^T \left(\frac{\partial u^T}{\partial p}\mathbf{M}u + u^T \frac{\partial \mathbf{M}}{\partial p} u + u^T \mathbf{M} \frac{\partial u^T}{\partial p} \right) \varphi_i}{2N\omega_i \varphi_i^T \mathbf{I} \varphi_i} \tag{14}$$

This solution for the sensitivities of the eigenvalues is substituted into $d\omega_i/dp$ in the first term of (12), where Ω is defined as:

$$\Omega = \left(\frac{\frac{\partial u^T}{\partial p}\mathbf{M}u + u^T \frac{\partial \mathbf{M}}{\partial p} u + u^T \mathbf{M} \frac{\partial u^T}{\partial p}}{N} - 2\omega_i \varphi_i^T \mathbf{I} \varphi_i \frac{\partial \omega_i^2}{\partial p} \right) \varphi_i \tag{15}$$

which results in the following system

$$\left(\mathbf{C} - \omega_i^2 \mathbf{I}\right) \frac{d\phi_i}{dp} = \Omega \tag{16}$$

This system is solved for $d\varphi_i/dp$ by:

$$\frac{d\varphi_i}{dp} = \tilde{\mathbf{C}}^{+} \Omega \tag{17}$$

where $\tilde{\mathbf{C}}^{+}$ is the generalized inverse[8,9] of the singular matrix $\tilde{\mathbf{C}}$:

$$\tilde{\mathbf{C}} = \left(\mathbf{C} - \omega_i^2 \mathbf{I}\right) \tag{18}$$

The last step is to expand $d\varphi_i$ for the approximation of the POD modes $d\phi_i$

$$\frac{d\phi_i}{dp} = \frac{du}{dp}\varphi_i + u_i \frac{d\varphi_i}{dp} \tag{19}$$

C. Combined Approximation

Combined approximation (CA)[10–14] is a reanalysis method used to approximate the basis vectors due to a change in system parameters. The new basis is approximated as a linear combination of another basis:

$$\tilde{\phi}_i(\boldsymbol{p}) = y_1 \mathbf{r_1} + y_2 \mathbf{r_2} + \cdots + y_n \mathbf{r_n} \tag{20}$$

where y_i are constants and $\boldsymbol{r}$ is the basis used for CA. A binomial series expansion about the original design is often chosen as the reduced basis.[10,11] In this study, one of the methods chosen for the CA approach is a Taylor Series approximation of the basis function. The Taylor Series approximation uses the original basis and the first derivatives of the basis with respect to the system parameters to approximate the current design at $\boldsymbol{p}$. Therefore, the CA approximation becomes:

$$\tilde{\phi}_i(\boldsymbol{p}) \approx y_1 \phi_i + y_2 \frac{\partial \phi_i}{\partial p_1} + \cdots + y_{n+1} \frac{\partial \phi_i}{\partial p_n} \tag{21}$$

The reader may note this is equivalent to a first order Taylor series expansion if $y_1 = 1$ and $y_{i>1} = \Delta p_{i-1}$. However, to find the coefficients $\boldsymbol{y}$, the newly assembled system matrices are reduced by the Taylor Series basis $\boldsymbol{r}$. Once the reduced matrix $\mathbf{M_{CA}}$ and $\mathbf{K_{CA}}$ are found, where $\mathbf{M_{CA}} = \mathbf{r^T M r}$ and $\mathbf{K_{CA}} = \mathbf{r^T K r}$, the following eigenvalue problem is solved to find $\boldsymbol{y}$:

$$\mathbf{K_{CA}} \mathbf{y} = \lambda \mathbf{M_{CA}} \mathbf{y} \tag{22}$$

where $\boldsymbol{y}$ are the eigenmodes of (22). One can approximate $\tilde{\phi}_i(\boldsymbol{p})$ via (22) using only the first eigenmode of 22). This method is referred to as single CA in the following. Alternatively, additional eigenmodes from (22) can be used to approximate several basis vectors by considering multiple eigenmodes of (22). This approach is referred to as full CA and is detailed below.

III. Extended Reduced Order Modeling Approaches

ROMs are accurate for the nominal design it was built from. For optimization or uncertainty quantification, the ROM needs to be accurate for the design change. The extension of the ROM into the design space will be further stated as an extended reduced order model (EROM). For the model of interest, the system matrices from (2) are dependent on a design change as well as the basis Φ from (3). As stated earlier, the system matrices are rebuilt at each design. In this section various alternatives will be studied for approximating the basis vectors as a function of system parameters.

The approximation of the basis, POD modes for this study, can be approximated using the basis and/or the gradients of the basis with respect to design and random variables at the nominal design, which is represented in figure 3 with a red cross. The approximation built in the middle is only a local approximation and will most likely not be able to approximation the whole design space (one point EROM). In an attempt to capture more of the design space, a multi-point approximation of the design space is explored. The green crosses in figure 3 represent sample points in the design space at which new snapshots and/or a new basis vector are computed. The idea is the whole design space can be approximated, if a sufficient number of sample points is used.

For the multi-point approximations, three different methods are used where either the POD modes or the snapshots are taken from each design point. Within each of the three methods several alternative approaches will be studied. Method M-1 finds the basis due to a design change with no derivative information. Option M-1a refers to the approximation of the basis vectors by a single CA, Option M-1b by a full CA. For a single CA, each of the individual modes is approximated using information only from like modes. An example

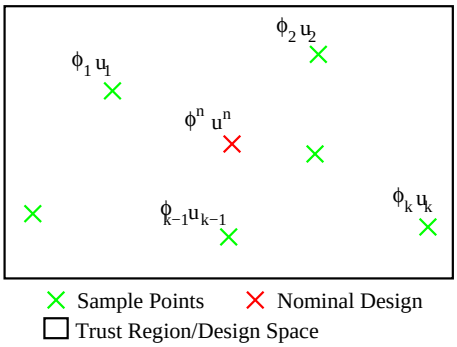

Figure 3. Schematic of how multi-point approximation will be implemented

is approximating the first mode using only the first modes at each sample point. The first modes from each design will be used in CA to find the approximated first mode at the design change and like wise for the preceding modes. For a design with four sample points, the first POD mode at the design change is approximated using the first POD mode at each sample point in CA. This is also done for preceding modes. In the eigen analysis from (22), a four by four eigen analysis is performed, where only the first eigenvector is used in (20). For a full CA, all of the modes are used to approximate all of the other modes due to a design change. In the previous example, assuming each of the four design points have three individual modes, a twelve by twelve eigen analysis is performed where all the modes from (22) are used in the approximation of the new basis. Since all of the eigenvectors are used, this increases the number of bases in ROM. In the single CA, four new modes are produced where twelve are produced for the full CA. Once the modes are found, they are used in the ROM to predict the response at the design change.

In option M-1c, a single CA on the snapshots is performed where each snapshot is approximated using like snapshot from each of the design points in CA. There are as many eigen analyses as there are snapshot for this technique where only the first eigenvector from (22) is used in (20). Once the snapshots are approximated, a POD analysis is performed to find the modes at the design change. If two design points are used, and only ten snapshots are taken from each design, the first snapshot is predicted by taking the first snapshot from each design. This produced a reduced mass and stiffness matrix which is two by two and in turn produced a two by two eigen analysis. Only the first eigenvector is used in (20) for the approximation of the first snapshot. This is done for all ten snap shots.

In option M-1d all of the snapshots from each sample are taken into account. The snapshots are put collectively together and a POD analysis is performed on the whole set of snapshots. Once the POD modes are built a full CA is performed to find the new basis vectors.

In method M-2, all of the same options are used, but derivative information is used in the approximation. However, in option M-2d, the mode and its corresponding derivatives are used in a single CA and done for all preceding modes. For method M-2e, a full CA is performed where all modes and derivatives are used to approximate modes at the design change.

Method M-3 collects two approaches referred to as spanning ROMs (SROMs). Option M-3a collectively put the bases from each sample point together and M-3b collectively puts the snapshot together from each design, perform a POD analysis and the modes acquired will be used to approximate the design space.

The difference between SORM (M-3) and EROM (M-1 and M-2) is once the basis vectors are built for SORM, there is no update of the bases as the system parameters change. In the EROM methods, the basis vectors are updated at each different design through CA on either the basis (M-1a,b and M-2a,b), the snapshots (M-1c and M-2c) or the combination of snapshots (M-1d and M-2d,e) from each sample point. For options M-1d and M-2d,e, once the POD modes are found, there is no need to rebuild the POD modes again because it would produce the same POD modes each time as the snapshots are not changing. The actual approximation comes from performing CA on the POD modes from the combination of snapshots.

The various methods studied are summarized below.

1. EROM approximation using no derivative information (method 1): Figures 6

a) Single CA of the basis

$$\tilde{\phi}_i(\boldsymbol{p}) \approx y_1\phi_i^{(1)} + \cdots + y_k\phi_i^{(k)}$$

b) Full CA of method 1a

c) Single CA of the snapshots

$$\tilde{u}_i(\boldsymbol{p}) = y_1 u_i^{(1)} + \cdots + y_k u_i^{(k)} => POD => \tilde{\Phi}$$

d) Full CA on the collective set of snapshots

$$U_{Total} = [\boldsymbol{u}^{(1)}|\cdots|\boldsymbol{u}^{(k)}] => POD => \tilde{\Phi}$$

2. EROM approximation using derivative information (method 2): Figures 7

a) Single CA of basis and derivatives

$$\tilde{\phi}_i(\boldsymbol{p}) \approx y_1\phi_i^{(1)} + \cdots + y_{j+1}\frac{\partial\phi_i^{(1)}}{\partial p_j} + \cdots + y_m\phi_i^{(k)} + y_{m+1}\frac{\partial\phi_i^{(k)}}{\partial p_j}$$

b) Full CA of method 2a

c) Single CA on the snapshots and derivative

$$u_i = y_1 u_i^{(1)} + y_2\frac{du_i^{(1)}}{dp_j} + \cdots + y_k u_i^{(k)} + y_{k+1}\frac{du_i^{(k)}}{dp_j} => POD => \tilde{\Phi}$$

d) Single CA on the collective set of snapshots and derivative

$$U_{Total} = [\boldsymbol{u}^{(1)}|\cdots|\boldsymbol{u}^{(k)}] => POD => \tilde{\Phi}$$

$$\Phi_{snap}, [\frac{d\boldsymbol{u}^{(1)}}{dp}|\cdots|\frac{d\boldsymbol{u}^{(k)}}{dp}] => \frac{d\Phi_{snap}}{dp}$$

e) Full CA of method 2d

3. SROM approximation (method 3): Figures 8

a) Combination of bases

$$\tilde{\Phi}(\boldsymbol{p}) \approx [\Phi^{(1)}|\cdots|\Phi^{(k)}]$$

b) Combination of snapshots

$$U_{Total} = [\boldsymbol{u}^{(1)}|\cdots|\boldsymbol{u}^{(k)}] => POD => \tilde{\Phi}$$

In the above list, $\tilde{\phi}_i(\boldsymbol{p})$ is the approximated basis at a particular configuration and ϕ_i are the basis at the sampled design. y are constants found in CA and the super-script denotes information from the k^{th} sample point.

There are differences which should be noted for each method. The first two methods, EROM approximations, use CA to predict the modes at a design change. This is an actual approximation of the modes at a particular configuration. Method M-3 combines information from each of the designs in the expectation the information taken from each design will be able to predict a change in design. This method is not an actual approximation of the reduced model at a particular configuration but is expected information taken at each sampling point will be sufficient enough at predicting the modes at another configuration.

For the single CA an eigen analysis is performed for each individual basis approximation. If, for example, there are three sample points, and four bases are needed to approximate the system response at a particular configuration, then four, three by three eigen analyses are performed where only the first eigenvector is used in the basis approximation. If a full CA is used a single twelve by twelve eigen analysis is performed where all the eigenvectors are used. Also when running the actual ROM, the number of degrees of freedom in the system is four and twelve respectively for the example above. If CA is performed on the snapshots and only the first eigenmode is used in CA, there are as many eigen analyses as there are snapshots.

Comparing the computation cost of each method, the full analysis for method (recalibration point) M-1 and M-3 are equal. The full analysis of method two is more expensive because gradient information is required. The approximation of the modes is done for only method M-1 and M-2, making these methods more expensive. Performing the reduced dynamic analysis, a single CA is less expensive then a full CA because more modes are used in the full CA of the dynamic analysis. A summary of the computation cost is shown below.

$$\begin{gathered}\textbf{FullAnalysis}\\ Method \quad 1 == Method \quad 3\\ Method \quad 1 < Method \quad 2\\ \\ \textbf{ROMAnalysis}\\ Method \quad 3 < Method \quad 1 < Method \quad 2\\ where\\ Single \quad CA < Full \quad CA\end{gathered} \tag{23}$$

IV. Computational Experiments: Connecting Rod

The various methods and options previously introduced are studied with a typical mechanical component: a connecting rod.[15,16] This example was used in a previous study[7] optimizing the design based on an EROM approximation.

The rod is clamped at the inner circumference of the left hole, and a transient force is applied to the inner circumference of the right hole. The rod has an overall length of $42mm$, a thickness of $3mm$, a Poisson ratio of 0.3, and a Young's modulus of $E = 7.2 \times 10^5 N/mm^2$. A linear elastic response and light damping is assumed using a Raleigh model, with $\alpha = 10^{-5}$ and $\beta = 10^{-5}$. The beam is discretized by 400 isoparametric 4 node plane-stress elements, resulting in a total of 936 degrees of freedom. All computations are performed within MatLab utilizing the CALFEM finite element toolbox.[17] Two geometric parameters, p_1 and p_2, control the horizontal position of the center hole, as depicted in figure 4.

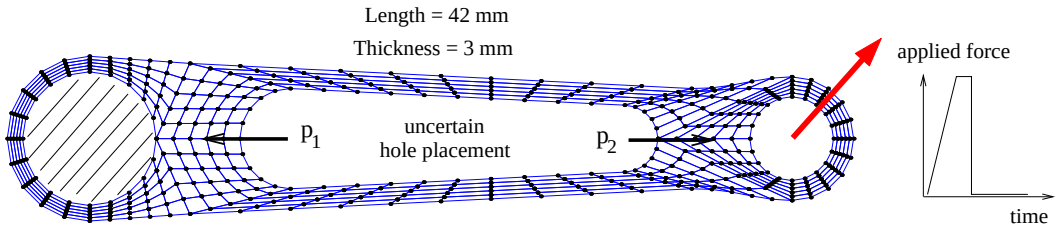

Figure 4. Model of the connecting rod

For illustration purposes, the following optimization problem is studied: find the values of the parameters p_1 and p_2 such the dissipation energy is minimized.

$$\begin{gathered}\min_{\boldsymbol{s}} \quad (Dissipation \quad Energy)\\ subject \quad to: \quad -4 \le s_1 \le 4\\ subject \quad to: \quad -4 \le s_2 \le 4\end{gathered} \tag{24}$$

A. EROM of Connecting Rod

For the rod, each method (M-1, M-2 and M-3) for approximating the dissipation energy for a particular configuration was analyzed. The options were tested for a one, two and three point EROM approximation. The best estimations of the design space will be implemented into the optimization of the rod.

The top left of figure 5 shows the full order response where the contours represent the dissipation energy. The contours from method M-1 and M-3 are the same using a one-point approximation with the exception

of method M-1b and M-1d which produce similar results, shown on the top right of 5. This is equivalent to approximating each design with the same modes. The middle left plot is a one point EROM, which is the same for all of the techniques in method except M-2b and M-2e. Using the full CA, the design space is captured better, but still not satisfactory. To see the effects of the multi-point approximations, each of the methods were tested and the contours plotted of the dissipation energy. For each method, two and three points in the design space were used for the approximation.

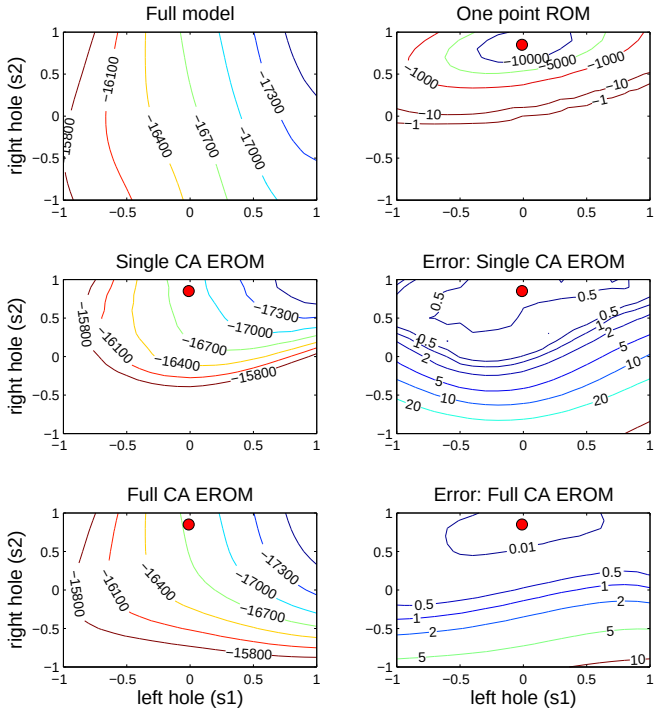

Figure 5. Comparison of full, one point ROM, EROM and full CA on the EROM

Figure 6 shows the approximation of method M-1 (no derivative information). The plots on the left of figures 6-8) show a two point sampling method and the right shows a three point sampling method. Each of the figures shows the plots in the order the techniques stated above. Figure 6 shows the contours without using derivative information. Figure 7 shows the contours using derivative information and the third figure shows the SROMS approximations.

The errors measured by the L^2 norm of the three point estimation are shown in figures 9. Method M-2d, is a fair approximation and needs to be studied further to see if this technique will produce accurate results. When a two-point approximation was used in M-2, the trend of dissipation energy was captured but there is a clear discrepancy with respect to the contours. If an optimization algorithm was run to find the minimum of the dissipation energy, it would have converged to the correct position, given the design space. The three-point approximation changed the trend of the design space. An optimization algorithm would converge to the incorrect optimum, but comparing the contours lines, it more accurately represents the position of the contour, especially near the left of the design space. To see if the accuracy can be improved, more sample point could be taken.

Three of the options resulted in poor approximations, namely method M-1d, M-3a and M-3b. The other methods were able to approximate the design space, but some better than others. In any of the techniques using a full CA, the accuracy of the response was increased. While the accuracy of the response was increased, the computational time was also increased. Comparing the results of methods M-1 and M-2 showed using gradient information always increases the accuracy of the approximation. Again this is at the cost of having to calculate derivatives at recalibration points and, if the full CA is used, then the computational time of

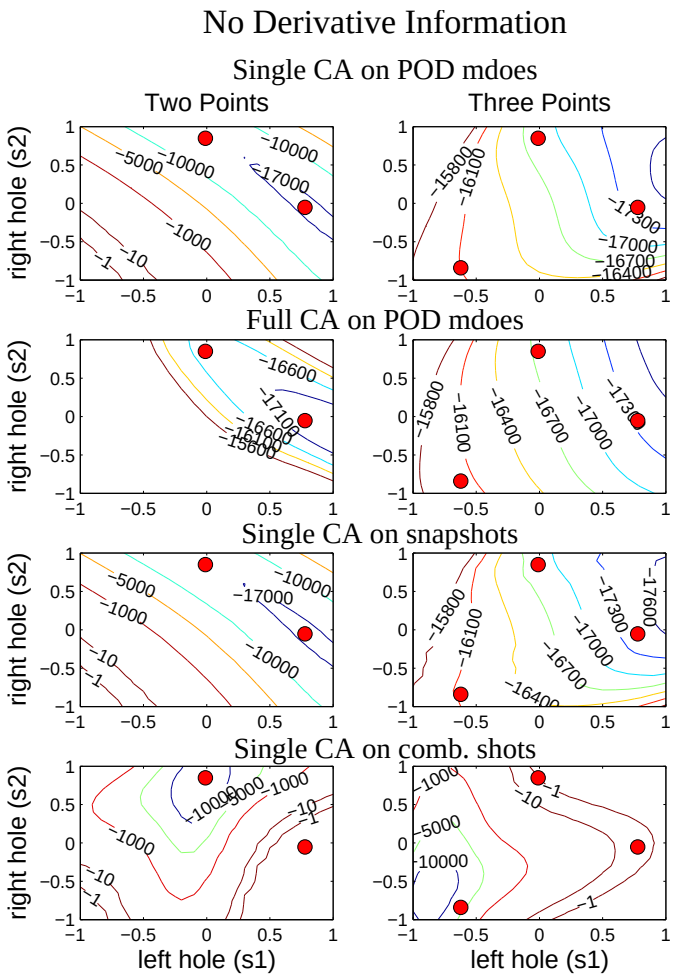

Figure 6. Two- and three-point approximations of method 1

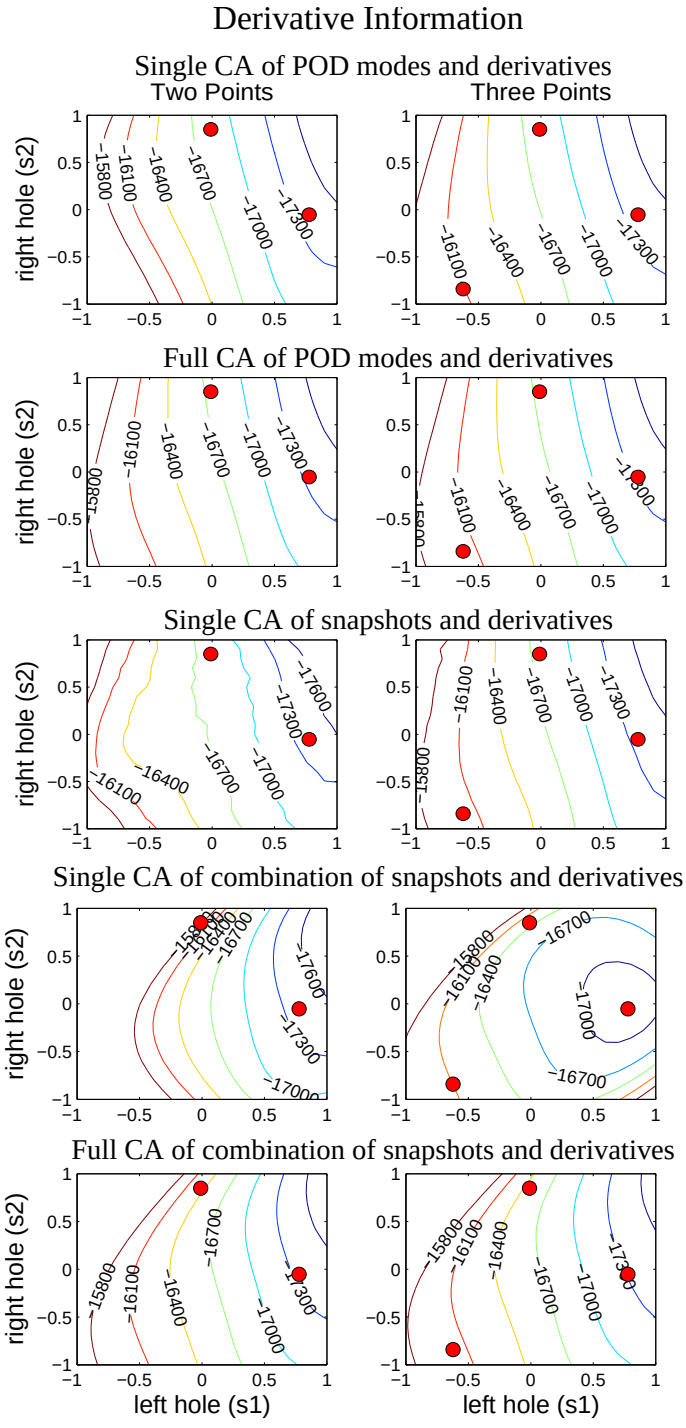

Figure 7. Two- and three-point approximations of method M-2

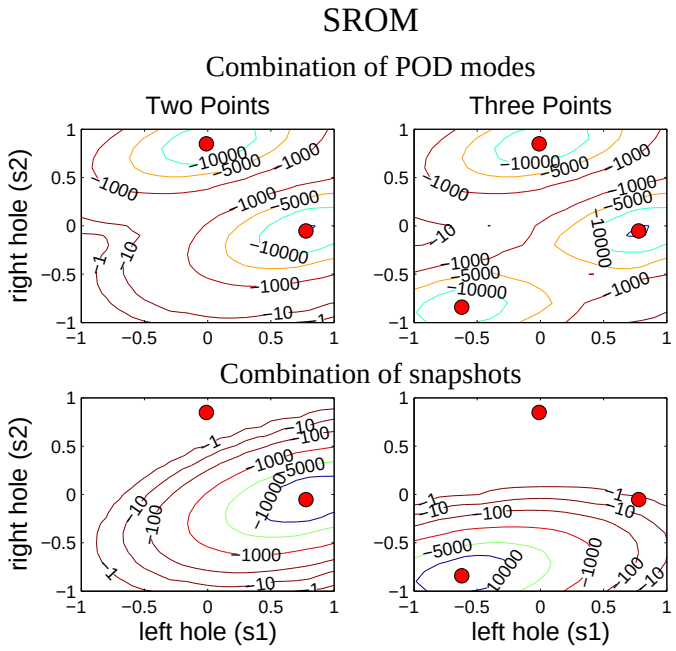

Figure 8. Two- and three-point approximations of method M-3

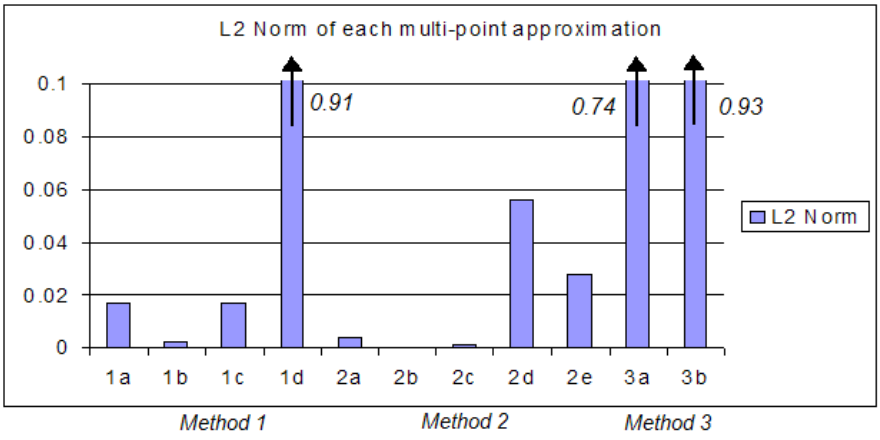

Figure 9. L^2 norm of each of the techniques from the three-point approximation

analyzing the ROM is further increased due to the increased number of degrees of freedom in the ROM. The method which performed the best was method M-2b. This method used derivatives of the POD modes in a full CA estimation.

B. Optimization of Rod

The accuracy of the multi-point approximations was shown in the above section and will be tested in the optimization of the rod. The starting point for the optimization process is at (0,0), and the contour of the dissipation energy of the full model is shown in 10. The deterministic optimum from the full model is also show in the figure.

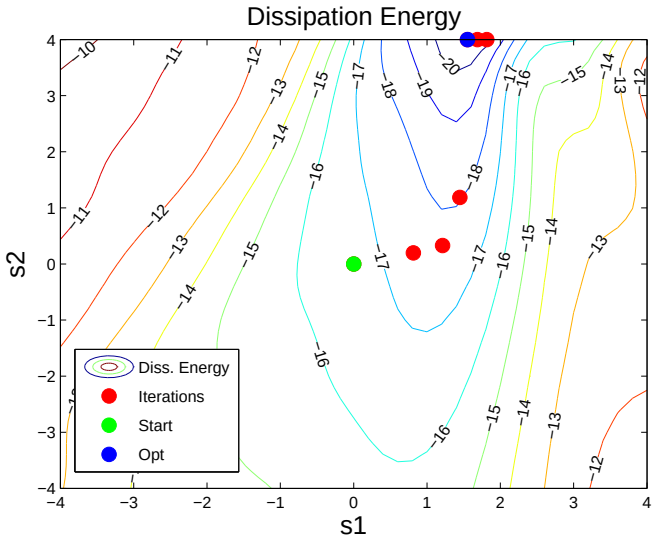

Figure 10. Contour plot minimum dissipation energy in the design space

The optimization of the rod will be performed using the three options M-1b, M-2a and M-2b. Options M-1b and M-2b showed the best accuracy with in each method and option M-2a is a benchmark to show the improvement in using multi-point approximations from a previous study.[7] While Allen et al.[7] used eigenmodes as bases, in this study, POD modes are used.

A trust region framework[5,6] will be used in the optimization process. The initial bounds of the trust region were from -2 to 2 for both design variables (global bound from -4 to 4 for both design variables). Two single-point approximations throughout the optimization process will be used to find the minimum of the dissipation energy (method M-2a,b). Three different multi-point approximations are looked into for the optimization of the rod. The first was appending to the basis (method M-2b) after convergence was reached in the trust region. Therefore, the first trust region step in the optimization process was a single-point approximation of the design space, and each trust region step after was a multi-point approximation. The last two optimizations are multi-point approximations of the trust region using method M-1b and M-2b where four and two points are used respectively.

To compare each of the methods, function evaluations of the full and reduced model are going to be compared. The number of full function evaluation is shown in figure 11 and the ROM evaluations are shown in 12. Figure 13 compares the cost of each of the methods with reference to the number of function evaluation needed by scaling the ROM evaluation comparable to the full model evaluations.

In the case of the rod example, the costs for a ROM evaluation are roughly one fourth of the one for the full model, so the number of ROM evaluations was divided by four and added to the number of full evaluations. This can vary significantly from one model to the next, where for most cases the ROM will be significantly cheaper than the full model analysis. The optimization with the EROM showed it would be more costly to use a one point EROM. The method using a multi-point approximation without derivatives of a full CA provided to be roughly the same as running the full model in the optimization process. The two methods, which provided to be computationally cheaper than using the full-order model, are using a

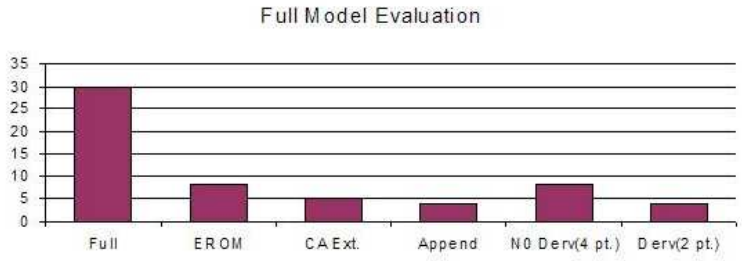

Figure 11. Total number of full function evaluations for the optimization of the connecting rod

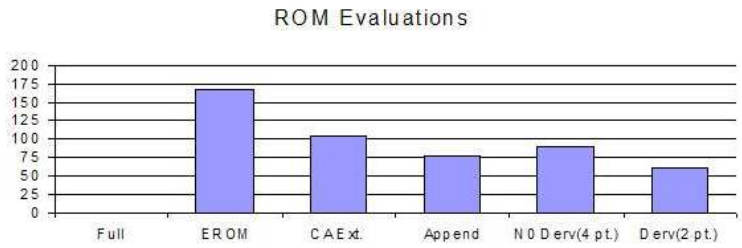

Figure 12. Total number of ROM evaluations in the optimization of the connecting rod

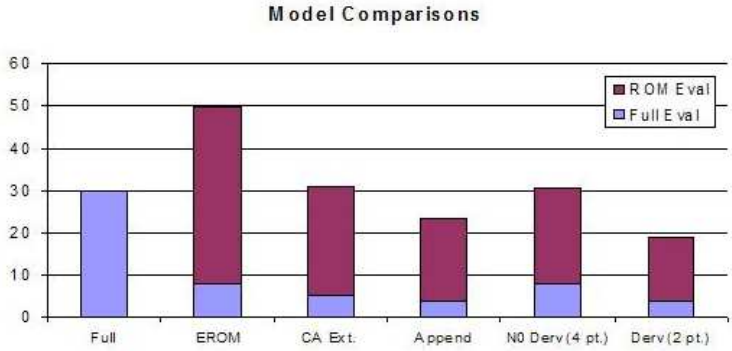

Figure 13. Comparing the total cost by scaling the ROM evaluations comparable to the full model evaluation in the optimization of the connecting rod

multi-point approximation with derivatives and appending basis information after each trust region step. The savings for the ROM in this example was one fourth the cost of the full model. This is not a significant reduction in cost. Clearly, there is a dependence on the ratio of the full to ROM evaluation time, where depending on the ratio it might be more beneficial to use the full model over the ROM model. In most cases though, the ROM is significantly cheaper than the full model, and would be more beneficial to run any of the methods used above for optimization.

The multi-point approximations might be more beneficial for design problems which do not have many design variables. The benefit of using multi-point approximations is the trust region can be increased and a sample point could get close to the optimum which will provide a quick convergence. In addition, as some of the sample points might not be beneficial for finding the optimum, the results show that appending to the basis information after each trust region step helps to overcome such situations.

C. Stochastic Analysis

To illustrate the utility of the EROM in uncertainty analysis, the rod in Figure 4 is used, and the placement of the center hole is modeled as a manufacturing uncertainty by the two uncertainty variables p_1 and p_2. The intended placement of the hole is treated as the mean design, and a normal distribution is assigned to the horizontal position of both ends of the hole, each with a standard deviation of $0.2mm$. As a performance measure, the amount of energy dissipated from the system between $0.4ms$ and $1.0ms$ is measured and utilized to evaluate altered design states.

The most general uncertainty analysis technique is Monte Carlo Simulation (MCS). In general, MCS is prohibited for most realistic dynamic systems due to the computational costs of each simulation, and the high number of computations required for an accurate solution. This cost is significantly reduced by utilizing the EROM due the reduced costs of time integration. However, the EROM proposed still requires assembly of the altered system matrices, which are expensive for a large number of samples. MCS is performed here not as a proposed solution of alleviating the computational burden, but as a means of demonstrating the effectiveness of the EROM in the uncertainty space.

A Monte Carlo analysis is performed on both the full order and the EROM (one point M-1b). 6500 samples were taken in all, and the same sample points were used for both the full order model and EROM. Each sample point represents one particular realization of the uncertainty parameters, for which a dynamic analysis is carried out and the energy dissipated in the system between $0.4ms$ and $1.0ms$ is recorded. To test the framework in terms of an uncertainty analysis, a failure surface is created by picking a critical energy dissipation level of $-16.5mJ$. If a design failed to dissipate at least $16.5mJ$ of energy, that is $E_d \geq -16.5mJ$, then the design is considered unsafe. The EROM, which was calibrated at the initial design, was first tested within the range of possible sample points, the mean design plus or minus 4 standard deviations, and the results were sufficiently accurate.

The results of the Monte Carlo analyses are illustrated in Figure 14. The plots on the left in Figure 14 represent the sample points and resulting probability density function (PDF) for the MCS with the full order system analyses, whereas the plots on the right represent the samples points and resulting PDF for the MCS with EROM system analyses. The sample points that failed in Figure 14 are shown in red (darker in color), while the points whose energy dissipation was sufficient are shown in green (lighter in color). The reader may note, through the MCS estimate of the failure surface is identified as the border between the two colors.

Qualitatively, the failure surfaces and PDFs of the full order analysis and EROM analysis appear very similar. Quantitative results are shown in Table 1. The MCS using the full order system analysis returned a probability of failure of 20.24 %, and the MCS utilizing the EROM returned a probability of failure of 20.32 %. The difference between the two methods is 0.4 %, The percent error associated with a Monte Carlo analysis, based on a 95 % confidence interval, is:

$$PercentError = 200\sqrt{\frac{1 - P_f}{nP_f}} \tag{25}$$

where P_f is the probability of failure and n is the number of samples. For the given example, the percent error is approximately 0.4%.

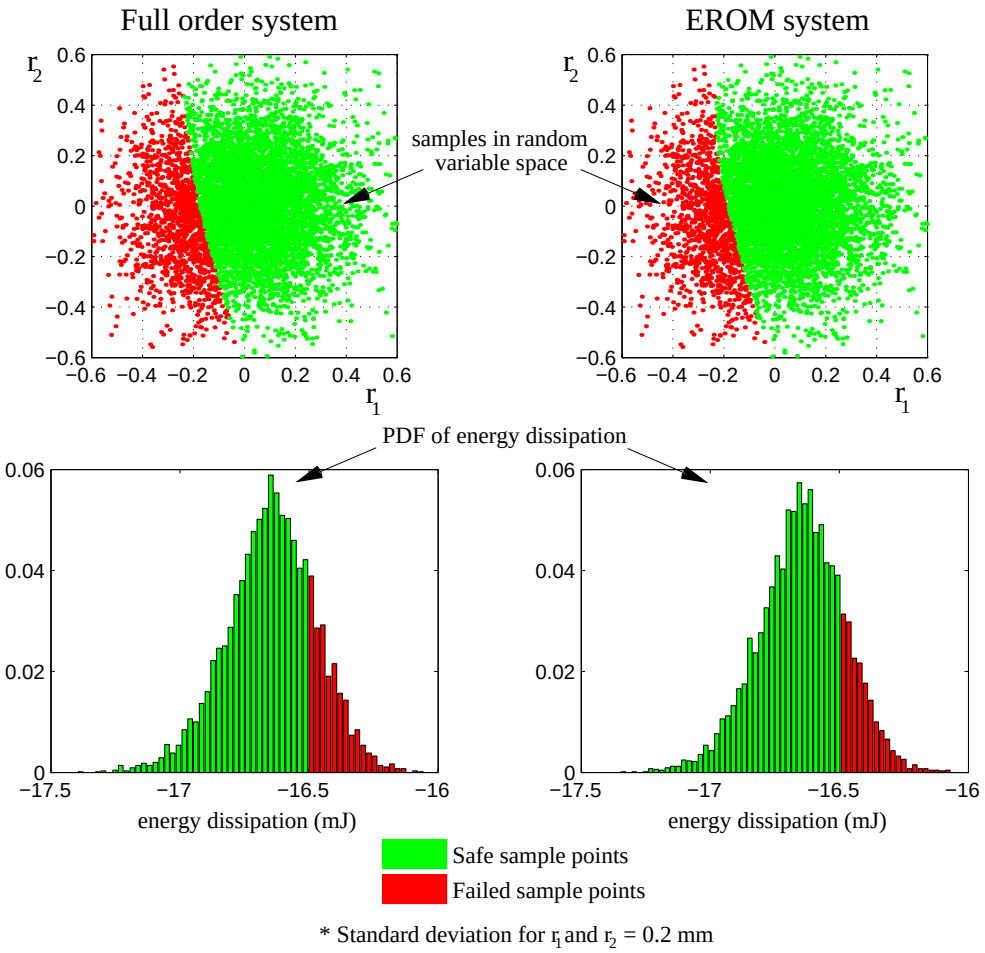

Figure 14. Monte Carlo analysis results for full and EROM systems

Table 1. Monte Carlo results of full and EROM model analysis

Analysis method	Probability of failure	MCS error
Full model	20.24 %	3.97 %
EROM	20.32 %	3.96 %

Conclusions

The paper has introduced a novel concept of reduced order modeling techniques for design optimization and stochastic analysis purposes. A Galerkin projection scheme is used to reduce the number of degrees of freedom in discretized PDE models. An extended reduced order models (EROMs) was proposed which approximates the basis vectors at a particular configuration by a combined approximation scheme. It was shown by numerical studies the EROM approach is accurate and reduces the cost of predicting the response. Various approximation methods have been studied using either multiple sampling points and/or gradient information. To increase the accuracy for large parameter changes, derivatives and/or multiple sampling points are beneficial. The numerical studies suggest approximating the actually basis, opposed to approximating the snapshots, leads to a better accuracy over wider parameter range. Using EROMs within a design optimization process, it was shown to gradually increase the set of basis vectors by including information from previous design points significantly the accuracy.

Acknowledgments

The first author would like to acknowledge the support through a summer internship at Sandia National Laboratories at Albuquerque. The first and third author would like to acknowledge the support of the National Science Foundation through grant DMI-0300539. The opinions and conclusions presented are those of the authors and do not necessarily reflect the views of the sponsoring organizations.

References

[1]Ravindran, S., "Proper Orthogonal Decomposition in Optimal Control of Fluids," Tech. rep., NASA TM-1999-209113, 1999.

[2]Thomas, J., Dowell, E., and Hall, K., "Three-Dimensional Transonic Aeroelasticity Using Proper Orthogonal Decomposition Based Reduced Order Models," *AIAA Paper 2001-1526*, 2001.

[3]Legresley, P. and Alonso, J., "Investigation of nonlinear projection for POD based reduced order models for aerodynamics," *AIAA 2001-16737, 39th Aerospace Sciences Meeting & Exhibit, January 8-11, 2001, Reno, NV*, 2001.

[4]Willcox, K. and Peraire, J., "Balanced model reduction via the proper orthogonal decomposition," *AIAA 2001-2611. 15th AIAA Computational Fluid Dynamics Conference, June 11-14, Anaheim, CA*, 2001.

[5]Giunta, A. and Eldred, M., "Implementation Of A Trust Region Model Management Strategy in the DAKOTA Optimization Toolkit," *Symposium on Multidiscipinary Analysis and Optimization, Long Beach, CA*, AIAA/USA/NASA/ISSMO, September 2000.

[6]M. Eldred, A. G. and Collis, S., "Second-Order corrections for Surrogate-Based Optimization with Model Hierarchies," 2004.

[7]M. Allen, G. W. and Maute, K., "Application of Reduced Order Models for the Stochastic Design Optimization of Dynamic Systems," *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, AIAA/ISSMO, Aug. 30 - Sept. 1 2004.

[8]Adelman, H. and Haftka, R., "Sensitivity analysis of discrete structural systems," *AIAA Journal*, Vol. 24, 1986, pp. 823–832.

[9]Dailey, R., "Eigenvector derivatives with repeated eigenvalues," *AIAA Journal*, Vol. 27, 1989, pp. 486–491.

[10]Kirsch, U., "A unified reanalysis approach for structural analysis, design, and optimization," *Structural and Multidisciplinary Optimization*, Vol. 25, No. 1, 2002, pp. 67–85.

[11]Kirsch, U., "Approximate vibration reanalysis of structures," *AIAA Journal*, Vol. 41, No. 3, 2003, pp. 504–511.

[12]Kirsch, U., "Exact and accurate solutions in the approximate reanalysis of structures," *AIAA Journal*, Vol. 39, No. 11, 2001, pp. 2198–2205.

[13]Kirsch, U., "Efficient, accurate reanalysis for structural optimization," *AIAA Journal*, Vol. 37, No. 12, 1999, pp. 1663–1669.

[14]Kirsch, U., "Combined approximations - a general reanalysis approach for structural optimization," *Structural and Multidisciplinary Optimization*, Vol. 20, No. 2, 2000, pp. 97–106.

[15]Bennet, J. and Botkin, M., "Structural shape optimization with geometric description and adaptive mesh refinement," *AIAA Journal*, Vol. 23, 1985, pp. 458–464.

[16]Zhang, W.-H., , Beckers, P., and Fleury, C., "A unified parametric design approach to structural shape optimization," *International Journal for Numerical Methods in Engineering*, Vol. 38, 1995, pp. 2283–2292.

[17]"CALFEM: A finite element toolbox to MATLAB Version 3.3," http://www.byggmek.lth.se/Calfem/index.htm, Structural Mechanics, LTH, Sweden, 1999.