International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering (M&C 2011)
Rio de Janeiro, RJ, Brazil, May 8-12, 2011, on CD-ROM, Latin American Section (LAS) / American Nuclear Society (ANS)
ISBN 978-85-63688-00-2

ON NONLINEAR REDUCED ORDER MODELING

Hany S. Abdel-Khalik
North Carolina State University
Department of Nuclear Engineering, P.O. Box 7909, Raleigh NC 27695
abdelkhalik@ncsu.edu

ABSTRACT

When applied to a model that receives *n* input parameters and predicts *m* output responses, a reduced order model estimates the variations in the *m* outputs of the original model resulting from variations in its *n* inputs. While direct execution of the forward model could provide these variations, reduced order modeling plays an indispensible role for most real-world complex models. This follows because the solutions of complex models are expensive in terms of required computational overhead, thus rendering their repeated execution computationally infeasible. To overcome this problem, reduced order modeling determines a relationship (often referred to as a surrogate model) between the input and output variations that is much cheaper to evaluate than the original model. While it is desirable to seek highly accurate surrogates, the computational overhead becomes quickly intractable especially for high dimensional model, $n \gg 10$.

In this manuscript, we demonstrate a novel reduced order modeling method for building a surrogate model that employs only "local first-order" derivatives and a new tensor-free expansion to efficiently identify all the important features of the original model to reach a predetermined level of accuracy. This is achieved via a hybrid approach in which local first-order derivatives (i.e., gradient) of a *pseudo* response (a *pseudo* response represents a random linear combination of original model's responses) are randomly sampled utilizing a tensor-free expansion around some reference point, with the resulting gradient information aggregated in a subspace (denoted by the active subspace) of dimension much less than the dimension of the input parameters space. The active subspace is then sampled employing the state-of-the-art techniques for global sampling methods. The proposed method hybridizes the use of global sampling methods for uncertainty quantification and local variational methods for sensitivity analysis. In a similar manner to global sampling methods, the developed method searches the space exhaustively for all unique points (e.g., turning points, maxima, minima, and bifurcation points). Moreover, the proposed method identifies only the important sensitivities for the responses as done by local variational methods. Finally, the accuracy of the surrogate model may be estimated using an order statistics approach to determine the minimum number of samples required to reach a desired accuracy.

Key Words: Nonlinear Sensitivity Analysis, Reduced Order Modeling, Tensor-Free Expansions

1. INTRODUCTION

Reduced Order Modeling (ROM) [1] is an essential analytical tool when repeated execution of a computational model is required for engineering-oriented studies such as sensitivity analysis (SA), uncertainty quantification (UQ), identification studies, and model inversion [2]. For most engineering systems exhibiting sufficient real-world complexity, the repeated execution of the forward model becomes computationally intractable. This follows as most real-world complex systems' models (e.g., nuclear reactors) require long execution times, are associated with nonlinear models tightly coupling different physics and scales, and are manipulating millions of

input parameters and output responses. To overcome the challenges of nonlinearity and high dimensionality in engineering-oriented applications such as SA and UQ, this work presents a new hybrid algorithm designed to accurately determine the variations in all model's output responses due to variations in all input parameters in a computationally efficient manner [2]. This is done by hybridizing the use of local variational methods [3] with global sampling methods [4].

Computational scientists have invested in ROM techniques to build surrogates whose predictions are sufficiently close to the original model. Response Surface Methods (RSMs) [5] represent the most common approach for building surrogates. Methods like Gaussian Processes Models, Stochastic Collocation Methods, Polynomial Chaos Methods [see Ref. 4 for a good overview], and Generalized Fourier Expansion methods [6] share a common feature shared by all RSM methods - that is the functional form of the response surface is pre-selected with some undetermined coefficients. The original model is then executed at a number of points in the input parameters phase space and a minimization problem is employed to determine the coefficients by minimizing the discrepancies between the surrogate's and the original model's predictions at the sampled points.

Regardless of the strategy developed to pick the random samples, an important challenge arises when the number of samples is much less than the dimension of the parameters' space which may be described as follows: the structure of the surrogate is unlikely to match that of the original model unless some insight is available a priori (by 'structure' we mean the effective null space and its orthogonal complement). Consequently, one would expect the surrogate to miss some of the influential directions which have a noticeable impact on the responses. Therefore the accuracy of the surrogate becomes difficult to assess outside the range of the samples employed for its construction. Examples are described in subsequent sections.

The new method hybridizes the use of global sampling methods and local variational methods. We give a short comparison here between the two classes of methods [3]. Both methods are employed to propagate parameters uncertainties and extract dominant parameters sensitivities on the estimated responses' uncertainties. Global sampling methods employ the forward model and randomly samples all input parameters from their prior probability density functions (PDF). After each run, responses deviations from reference values are recorded and the procedure is repeated with different random samples until a reliable estimate of responses PDFs is obtained. The advantage of this approach lies in its simplicity and ability to obtain detailed (i.e., all moments) PDFs for all responses. This is primarily important for a general nonlinear model and general input parameters PDFs where the responses PDFs are expected to deviate considerably from the Gaussian shape. The disadvantage is that sensitivity information are more difficult to infer (often sought via a response surface approach or an Analysis of Variance approach), and the number of model executions can be too large to render a practical approach for high dimensional models. When the number of input parameters is sufficiently small, one could exhaustively (i.e., via a greedy algorithm) sample the input space to identify important sensitivities, especially when the models are highly nonlinear. With the number of input parameters numbering in the tens of thousands to billions, the greedy search algorithm becomes computationally prohibitive. Computational scientists have devised numerous strategies to limit

the number of samples with considerable success [4]. To our knowledge, however, the state-of-the-art techniques are still limited to models with small number of input parameters, i.e. $n \sim 10$.

Alternatively, the local variational methods trade obtaining detailed responses PDFs for achieving computational efficiency and obtaining more sensitivity information. In their most practical implementation, the model is linearized and only first order derivatives of responses with respect to input parameters are determined. With this information only, the responses' PDFs are approximated as linear combinations of the input parameters' PDFs. If the input parameters uncertainties are described by Gaussian PDFs, the responses will also have Gaussian PDFs. If the model is nonlinear however, responses' PDFs are expected to deviate from the Gaussian shape. The advantage is that sensitivity information can help identify influential input parameters responsible for propagated uncertainties. The disadvantage is that if the model deviates considerably from being linear, this approach will poorly approximate the responses' PDFs. To identify higher order effects, the mathematical framework for variational methods have been extended to nonlinear models. For example, to characterize second order effects, an extra execution of the model is required for each input parameter in order to characterize its second order derivative which captures its interaction with other parameters. Clearly, for a high dimensional model, this approach becomes computationally prohibitive.

While significant advances have been accomplished in these two areas, there has been very little communication between their respective computational scientists and practitioners. Exceptions include gradient-enhanced methods that utilize local gradient information to reduce the number of samples required to construct the surrogate [7]. In this work, we also seek a hybrid method that can take advantages of the combined benefits of variational and sampling methods and circumvents some of their deficiencies. The idea is to employ local variational methods to identify a subspace (denoted hereinafter by the active subspace) that captures all direct and indirect parameters' cross-interaction effects up to all orders inherent in the original model, and then employ global sensitivity methods to build a surrogate that can be exhaustively sampled to determine sensitivity information and propagate parameters' uncertainties. Departing markedly from existing RSM methods, we employ a novel tensor-free expansion with three distinctive features: first, it uses an *infinite* series expansion; second, a *spanning* set rather than a basis is employed; and third, *arbitrary* functions may be used in the expansion. Finally, our approach employs an order-statistics type approach to measure the accuracy of the surrogate model given the number of samples employed to build it.

While the expansion uses initially an *infinite* number of terms to ensure an infinite radius of convergence for the surrogate, it is subsequently truncated based on a select numerical tolerance limit. This limit quantifies an upper bound on the maximum allowed discrepancy between the surrogate's and original model's predictions. This feature distinguishes our approach from RSM methods, where the surface is fixed a priori and then it becomes difficult to assess its accuracy especially when the number of samples that can be afforded is much smaller than the dimension of the input parameters' space.

The second distinctive feature of the new expansion is that it employs a spanning set rather than a basis to approximate the model's responses. Most RSM methods employ basis functions, which by definition are independent of each other. A spanning set, however, may contain

functions that are linearly dependent on other functions in the set. By definition, a basis represents the smallest spanning set. While at a first glance this may be considered an undesired feature, employing a spanning set provides enough flexibility to the surrogate model to tailor itself to match the structure of the original model which, as discussed before, represents the main challenge of RSM methods. We show that the active subspaces of the surrogate and original model could be matched when a spanning set is employed.

Finally, the arbitrariness of the spanning functions reduces the new expansion to various well-known expansions such as the Generalized Fourier expansion with its rich mathematical theory.

2. MATHEMATICAL DESCRIPTION OF PROPOSED APPROACH

In order to set the stage for the proposed method, we first present a mathematical description of the model. Few new definitions are then introduced for the active and inactive subspaces for a general nonlinear operator. Finally, the new expansion is presented.

2.1. Definitions

2.1.1 Mathematical Model. Let the original model be described by the following two equations:

$$\vec{\Pi}\left(\vec{\phi},\vec{x}\right)=\vec{\xi}\left(\vec{x}\right),\ \text{and}\ \ \vec{y}=\Sigma\left(\vec{\phi},\vec{x}\right), \tag{1}$$

where $\vec{x}\in\mathbb{R}^n$ is a *n*-component vector representing the *n* input parameters to the model, $\vec{\phi}$ is a vector representing the system's state, and $\vec{\Pi}$ describes the physics models (i.e., constraints) relating input parameters to system's state. The *m* model's responses, represented by a vector $\vec{y}\in\mathbb{R}^m$, are calculated as functions of the state vector and the input parameters. The system of equations in (1) could be compactly represented as follows:

$$\vec{y}=\vec{\Theta}\left(\vec{x}\right) \tag{2}$$

2.1.2. The Inactive Subspace. The inactive subspace of a nonlinear operator $\vec{\Theta}$ identifies the set of directions in the input parameters' space that are considered non-influential with respect to the model's responses. If $\vec{\Theta}$ is a linear operator represented by a matrix $\boldsymbol{\Theta}$, the inactive subspace is analogous to the null space of the matrix, $\mathrm{N}\left(\boldsymbol{\Theta}\right)$.

The descriptive 'non-influential' is employed hereinafter to denote that changes in input parameters along these directions do not lead to *noticeable* changes in the responses. Mathematically any change that is not equal to zero is noticeable. However from an engineering (practical) point of view, a numerical tolerance limit ε is often used to replace the mathematical zero as it is more practical to ignore all variations below a certain cutoff deeming them unimportant. The inactive subspace is represented symbolically by $\mathbb{Z}_{\mathrm{I}}\left(\vec{\Theta}\right)$ with dimension $\ell_{\mathrm{I}}=\dim\left(\mathbb{Z}_{\mathrm{I}}\right)$. This definition will be instrumental in our development as it identifies the directions in the input parameters space that will be excluded from sampling.

2.1.3. The Active Subspace. The active subspace for the nonlinear operator $\vec{\Theta}$, denoted by $\mathbb{Z}_{\mathrm{A}}\left(\vec{\Theta}\right)$, represents the orthogonal complement of the inactive subspace: $\mathbb{Z}_{\mathrm{A}}\left(\vec{\Theta}\right)\oplus\mathbb{Z}_{\mathrm{I}}\left(\vec{\Theta}\right)=\mathbb{R}^n$. If $\vec{\Theta}$ is a linear operator, the active subspace represents $\mathrm{R}\left(\boldsymbol{\Theta}^T\right)$. As will be shown later, this subspace will be built from randomly generated gradient information for pseudo responses. For linear models, the gradient of a response is the same everywhere in the phase space. For nonlinear models however, the gradient changes depending on where it is evaluated in the phase space. We employ a sampling strategy to build a subspace that captures all gradient variations in order to identify all important nonlinear effects. To propagate uncertainties and evaluate sensitivities, the random samples are subsequently constrained to the active subspace. This results in reducing the dimensionality of the problem thus rendering it computationally tractable.

2.2. Tensor-Free Subspace Expansion

2.2.1. Definition. A general nonlinear function $f\left(\vec{x}\right)$ of *n* input parameters represented by a vector $\vec{x}\in\mathbb{R}^n$ may be expanded around a reference point $\vec{x}_0$ as follows (without loss of generality we assume $\vec{x}_0=\vec{0}$ and $f\left(\vec{x}_0\right)=0$) to simplify the expressions:

$$f\left(\vec{x}\right)=\sum_{k=1}^{\infty}\sum_{j_1,..j_l..,j_k=1}^{n}\alpha_{j_1..j_l..j_k}\ \psi_1\left(\vec{\beta}_{j_1}^{(k)T}\vec{x}\right)\ ..\ \psi_l\left(\vec{\beta}_{j_l}^{(k)T}\vec{x}\right)\ ..\ \psi_k\left(\vec{\beta}_{j_k}^{(k)T}\vec{x}\right) \tag{3}$$

where $\left\{\psi_l\right\}_{l=1}^{\infty}$ are arbitrary scalar functions. This expression provides an exact estimate for the function $f\left(\vec{x}\right)$ at every point in the phase space. This could be easily verified by reducing it to Taylor series expansion by assuming: $\psi_l\left(\theta\right)=\theta$. We demonstrate that later in this section.

The value of this expansion lies in its ability to identify a reduced order model that captures all the important higher order effects of the original model, including both direct effects of each input parameter, and the indirect effects capturing the cross-interactions between parameters. This may be demonstrated by calculating the derivative of the expansion in Eq. (3) as follows:

$$\vec{\nabla}f\left(\vec{x}\right)=\sum_{k=1}^{\infty}\sum_{j_1,..j_l.,j_k=1}^{n}\alpha_{j_1....j_k}\ \psi_1\left(\vec{\beta}_{j_1}^{(k)T}\vec{x}\right)..\psi_{l-1}\left(\vec{\beta}_{j_{l-1}}^{(k)T}\vec{x}\right)\psi_l'\ \left(\vec{\beta}_{j_l}^{(k)T}\vec{x}\right)\psi_{l+1}\left(\vec{\beta}_{j_{l+1}}^{(k)T}\vec{x}\right)..\psi_k\left(\vec{\beta}_{j_k}^{T}\vec{x}\right)\vec{\beta}_{j_l}^{(k)} \tag{4}$$

where $\psi_l'\ \left(\vec{\beta}_{j_l}^{(k)T}\vec{x}\right)\vec{\beta}_{j_l}^{(k)}$ is the derivative of the term $\psi_l\left(\vec{\beta}_{j_l}^{(k)T}\vec{x}\right)$. Note that each term of this summation represents a vector in the input parameters' space pointing in the direction of $\vec{\beta}_{j_l}^{(k)}$. Accordingly, the gradient of the response function, $\nabla f\left(\vec{x}\right)$, may be viewed as a linear combination of all the $\left\{\vec{\beta}_{j_l}^{(k)}\right\}$ vectors. These vectors are free to select based on the gradient

information from the original model. In our approach, the $\left\{\vec{\beta}_{j_l}^{(k)}\right\}$ vectors will be selected such as the active subspace for the surrogate model matches that of the original model.

Most RSM methods struggle to measure the accuracy of the surrogate model. This is because the active subspace for the surrogate is fixed by the choice of the functions employed to build the surrogate's surface, which do not necessarily match the active subspace for the original model. This implies that there will be directions in the phase space that belong to the active subspace of the original model (i.e. capable of changing the response), but not captured by the surrogate. We demonstrate this with an example: consider a Polynomial Chaos type expansion:

$$f\left(\vec{x}\right)=\sum_{i=1}^{P}c_i\Psi_i\left(\vec{x}\left(\vec{\xi}\right)\right), \tag{5}$$

where $\vec{\xi}$ are the chaos variables; they are related to the input parameters $\vec{x}$ via pre-selected functions; and *P* is the order of the chaos. Differentiating this equation gives:

$$\nabla f\left(\vec{x}\right)=\sum_{i=1}^{P}c_i\nabla\Psi_i\left(\vec{x}\left(\vec{\xi}\right)\right)\frac{d\vec{x}}{d\vec{\xi}} \tag{6}$$

The gradient in this expression represents a linear combination of directions in the input parameters phase space that are fixed by the pre-selected relationship between the chaos variables and the input parameters; denote the subspace spanned by these fixed directions the chaos subspace. In this approach, the chaos subspace is determined blindly without any feedback from the original model. Moreover, in gradient-enhanced Polynomial Chaos methods, one attempts to determine the coefficients c_i by minimizing the differences between the gradient information generated from the original model and the chaos subspace. This has to be done in a least-squares sense as the original model's gradients belong to its active subspace which in general is different from the chaos subspace determined from the selected chaos functions. Therefore, it is expected that some components of the original model's gradient may not be captured correctly by the surrogate model. Moreover, some of the directions in the chaos subspace may belong to the inactive space of the original model. Sampling along the inactive space is wasteful as it does not lead to changes in the response function. In our approach, the active subspace is not fixed a priori, however, it is determined from the original model's gradient information; thereby the inactive subspace for the surrogate and the original model are expected to be the same which ensures an optimum sampling strategy.

Now, back to Eq. (3): as pointed out earlier, while the expansion is allowed to go to infinity, it is truncated based on a given numerical tolerance limit. This is achieved by first identifying the minimum number of gradient evaluations from Eq. (4) that are needed to span the entire active subspace. Gradient information are sampled from Eq. (4) at random points in the input space, which identifies the minimum number of terms to be retained in Eq. (3) in order to satisfy the given tolerance limit.

3.2. Anatomy of the Expansion. To explain the new expansion, we start with some simplified examples: consider a linear model given by:

$$f(\vec{x}) = b_1 x_1 + b_2 x_2 + ... + b_n x_n , \tag{7}$$

which may be rewritten in vector form as: $f(\vec{x}) = \vec{\beta}^T \vec{x}$, where $\vec{\beta} = \begin{bmatrix} b_1 & ... & b_n \end{bmatrix}^T \in \mathbb{R}^n$. In this representation, any change in the parameters that is orthogonal to the vector $\vec{\beta}$ will result in no change in the response. Therefore it would be reasonable to limit the random samples to be parallel to the vector $\vec{\beta}$. The vector $\vec{\beta}$ is readily available if one have the capability to calculate local gradient at reference point $\vec{x}_0$. In this case, the active subspace is given by: $\mathbb{Z}_{\mathrm{A}} = \mathrm{span}\{\vec{\beta}\}$, and has a single dimension, $\ell_{\mathrm{A}} = 1$.

Now, consider a second order model:

$$f(\vec{x}) = \sum_{i=1}^{n} b_i x_i + \sum_{i,j=1}^{n} c_{ij} x_i x_j \tag{8}$$

In vector form, this equation may be rewritten as: $f(\vec{x}) = \vec{\beta}^T \vec{x} + \vec{x}^T \mathbf{H} \vec{x}$, where $\mathbf{H}$ is the Hessian matrix which defines the second order derivatives, i.e., $[\mathbf{H}]_{ij} = (1/2)\partial^2 f / \partial x_i \partial x_j$. The Hessian matrix is symmetric and may be decomposed using the Singular Value Decomposition (SVD) [8]: $\mathbf{H} = \mathbf{V}\mathbf{S}\mathbf{V}^T$ which reduces the model's equation to:

$$f(\vec{x}) = \vec{\beta}^T \vec{x} + \frac{1}{2} \sum_{i=1}^{r} s_i \left[\vec{\upsilon}_i^T \vec{x} \right]^2 , \tag{9}$$

where s_i are the diagonal entries of the diagonal matrix $\mathbf{S}$ and $\{\vec{\upsilon}_i\}$ are the singular vectors representing the columns of the matrix $\mathbf{V}$. Note that the summation goes up to r (the rank of $\mathbf{H}$) only, where $r \le n$.

Following the argument presented earlier, one should limits parameters' perturbations to the active subspace $\mathbb{Z}_{\mathrm{A}} = \mathrm{span}\{\vec{\beta}, \vec{\upsilon}_1, ..., \vec{\upsilon}_r\}$ with dimension $\ell_{\mathrm{A}} = r + 1$.

Now recall from matrix theory that there are an infinite number of sets that may span a particular subspace. For example, any set of vectors $\{\vec{\nu}_1, ..., \vec{\nu}_{\ell_{\mathrm{A}}}\}$ that satisfies the condition below could be used as a spanning set for the active subspace:

$$\vec{\nu}_j = \lambda_1 \vec{\beta} + \sum_{i=2}^{\ell_{\mathrm{A}}} \lambda_i \vec{\upsilon}_i , \tag{10}$$

where $\{\lambda_i\}_{i=1}^{\ell_A}$ are randomly selected. One can show that with computer-generated random numbers, the vectors $\{\vec{v}_1,....,\vec{v}_{\ell_A}\}$ are independent and may therefore be employed to span the active subspace.

Now, consider the derivative information of the model given by:

$$\nabla f(\vec{x}) = \vec{\beta} + \sum_{i=1}^{r} s_i \left[\vec{\upsilon}_i^T \vec{x}\right] \vec{\upsilon}_i \tag{11}$$

Evaluating this gradient at a random point in the input space is equivalent to creating a random linear combination of the vectors: $\{\vec{\beta}, \vec{\upsilon}_1,....,\vec{\upsilon}_r\}$. The use of random numbers in Eq. (11) guarantees that a new gradient direction is created in the active subspace [9]. When this process is repeated ℓ_A times, a complete basis for the active subspace is obtained. The ℓ_A may be determined employing the following criterion:

$$\left|\nabla_{\ell_A+1} f(\vec{x}) - \sum_{j=1}^{\ell_A} \mu_j \nabla_j f(\vec{x})\right| < \varepsilon, \tag{12}$$

where ε is a numerical tolerance limit which may represent the machine precision, or the desired accuracy of the calculation. If ε is selected to be exactly equal to zero, this condition implies that the ℓ_A+1 randomly generated gradient is linearly dependent on the first ℓ_A vectors and hence must belong to the active subspace. Since it is not computationally necessary to reach a zero for most practical engineering applications, a numerical tolerance limit is used instead. This would imply that a sufficiently acceptable portion of the ℓ_A+1 randomly generated gradient lies in the active subspace. In our implementation, an order statistics approach is employed to satisfy the condition in Eq. (12) in a statistical sense given a user defined confidence interval and associated probability of the tolerance limit being within the selected interval, see Ref [2] for implementation details.

Now, we extend the above analysis to models with multiple responses as follows. Define the q^{th} response's function and gradient as:

$$f_q(\vec{x}) = \vec{\beta}_q^T \vec{x} + \frac{1}{2} \sum_{i=1}^{r_q} s_{i,q} \left[\vec{\upsilon}_{i,q}^T \vec{x}\right]^2, \tag{13}$$

$$\nabla f_q(\vec{x}) = \vec{\beta}_q + \sum_{i=1}^{r} s_{i,q} \left[\vec{\upsilon}_{i,q}^T \vec{x}\right] \vec{\upsilon}_{i,q} \tag{14}$$

Given that reference gradients and the Hessians' singular vectors are generally different for different responses, the active subspace is selected to span the union of all these vectors.

The total number of reference gradients and Hessians' singular vectors for *m* responses is:

$$m+\sum_{q=1}^{m} r_q \,, \tag{15}$$

where r_q is the rank of the q^{th} response's Hessian matrix. Clearly, this number could be larger than the dimension of the parameters' space $\mathbb{R}^n$, implying that they must be linearly dependent since one cannot have more than *n* independent vectors in an *n* dimensional space. To identify the minimum number of vectors required to span the active subspace of dimension ℓ_{A}, a random linear combination of these vectors is created and the process is repeated until the entire subspace is spanned. To linearly combine, with random weights, the reference gradients and hessian singular vectors for different responses, we define a pseudo response given by [10]:

$$f^{\text{pseudo}}\left(\vec{x}\right)=\sum_{q=1}^{m}\pi_q f_q\left(\vec{x}\right), \tag{16}$$

where $\left\{\pi_q\right\}$ are randomly generated. The gradient of the pseudo response is given by:

$$\nabla f^{\text{pseudo}}\left(\vec{x}\right)=\sum_{q=1}^{m}\pi_q \nabla f_q\left(\vec{x}\right)=\sum_{q=1}^{m}\pi_q\vec{\beta}_q+\sum_{q=1}^{m}\sum_{i=1}^{r_q} s_{i,q}\left[\vec{\upsilon}_{i,q}^{T}\vec{x}\right]\vec{\upsilon}_{i,q} \tag{17}$$

which would create a random linear combination of all reference gradient vectors $\left\{\vec{\beta}_q\right\}$ and the Hessians' singular vectors $\left\{\vec{\upsilon}_{i,q}\right\}$, if it is evaluated at a random point in the parameters space $\vec{x}$. Like before, this process is to be repeated ℓ_{A} times until the next direction $\ell_{\text{A}}+1$ is linearly dependent on the first ℓ_{A} directions, i.e.,

$$\left|\nabla f_{\ell_{\text{A}}+1}^{\text{pseudo}}\left(\vec{x}\right)-\sum_{j=1}^{\ell_{\text{A}}}\mu_j\nabla f_j^{\text{pseudo}}\left(\vec{x}\right)\right|<\varepsilon \tag{18}$$

Note that the above analysis identifies the active subspace without having to extract the reference gradients and the Hessian's singular vectors. Once the active subspace is determined, these vectors could be recovered via a minimization of dimension ℓ_{A}.

Note that for the second order model, the active subspace consists of two sets of directions, the first set describes the reference gradients which are responsible for the first order effects of the parameters on the responses, and the second set is the singular vectors of the Hessians which quantifies the second order effects. Extending this idea to a model of the n^{th} order, one can view the active subspace as an aggregate of the directions representing the various orders. This allows one to represent the function via the following expansion:

$$f\left(\vec{x}\right)=\sum_{i_1=1}^{n}\left(\vec{\beta}_{i_1}^{(1)T}\vec{x}\right)+\sum_{i_1,i_2=1}^{n}\left(\vec{\beta}_{i_1}^{(2)T}\vec{x}\right)\left(\vec{\beta}_{i_2}^{(2)T}\vec{x}\right)+\sum_{i_1,i_2,i_3=1}^{n}\left(\vec{\beta}_{i_1}^{(3)T}\vec{x}\right)\left(\vec{\beta}_{i_2}^{(3)T}\vec{x}\right)\left(\vec{\beta}_{i_3}^{(3)T}\vec{x}\right)+.... \quad (19)$$

where $\left\{\vec{\beta}_i^{(k)}\right\}_{i=1}^{n}$ represents a spanning set for the k^{th}-order effects, denoted hereinafter by the k^{th}-order span. In a compact form the above expansion becomes:

$$f\left(\vec{x}\right)=\sum_{k=1}^{\infty}\sum_{i_1,...i_l,...i_k=1}^{n}\left(\vec{\beta}_{i_1}^{(k)T}\vec{x}\right)..\left(\vec{\beta}_{i_l}^{(k)T}\vec{x}\right)..\left(\vec{\beta}_{i_k}^{(k)T}\vec{x}\right), \quad (20)$$

and the associated gradient is given by:

$$\nabla f\left(\vec{x}\right)=\sum_{k=1}^{\infty}\sum_{i_1,...i_l,...i_k=1}^{n}\left(\vec{\beta}_{i_1}^{(k)T}\vec{x}\right)..\left(\vec{\beta}_{i_{l-1}}^{(k)T}\vec{x}\right)\left(\vec{\beta}_{i_{l+1}}^{(k)T}\vec{x}\right)..\left(\vec{\beta}_{i_k}^{(k)T}\vec{x}\right)\vec{\beta}_{i_l}^{(k)} \quad (21)$$

In this representation, the active subspace is spanned by *all*-order span vectors. As described before, these vectors must be linearly dependent as there can only be *n* independent vectors in an *n* dimensional space. Identifying the minimum number of independent vectors to span the active subspace may be achieved by creating ℓ_{A} random linear combinations of *all*-orders spans' vectors. For models with *m* responses, this is equivalent to creating ℓ_{A} random evaluations of the gradient of pseudo responses, which is given by:

$$\nabla f^{\text{pseud}}\left(\vec{x}\right)=\sum_{q=1}^{m}\pi_q\sum_{k=1}^{\infty}\sum_{i_1,...i_l,...i_k=1}^{n}\left(\vec{\beta}_{i_1,q}^{(k)T}\vec{x}\right)..\left(\vec{\beta}_{i_{l-1},q}^{(k)T}\vec{x}\right)\left(\vec{\beta}_{i_{l+1},q}^{(k)T}\vec{x}\right)..\left(\vec{\beta}_{i_k,q}^{(k)T}\vec{x}\right)\vec{\beta}_{i_l,q}^{(k)}, \quad (22)$$

where the q subscripts distinguishes the order-span vectors for the various responses.

3. CASE STUDY

A test problem, based on Morris's experiment [11], is employed (Eq. (5)). As implemented, this experiment involves single-output test function with one hundred different inputs, with first- through fourth-order interaction terms. The coefficients for the various terms are either fixed or randomly selected with large variations to simulate the relative importance of various terms. The computational model of this experiment is given by:

$$\begin{aligned} y=&\sum_{i=1}^{r_1}\alpha_i\left(\vec{\beta}_{1i}^{T}\vec{x}\right)+\sum_{i,j=1}^{r_2}\alpha_{ij}\left(\vec{\beta}_{2i}^{T}\vec{x}\right)\left(\vec{\beta}_{2j}^{T}\vec{x}\right)+\\ &+\sum_{i,j,k=1}^{r_3}\alpha_{ijk}\left(\vec{\beta}_{3i}^{T}\vec{x}\right)\left(\vec{\beta}_{3j}^{T}\vec{x}\right)\left(\vec{\beta}_{3k}^{T}\vec{x}\right)+\sum_{i,j,k,l=1}^{r_4}\alpha_{ijkl}\left(\vec{\beta}_{4i}^{T}\vec{x}\right)\left(\vec{\beta}_{4j}^{T}\vec{x}\right)\left(\vec{\beta}_{4k}^{T}\vec{x}\right)\left(\vec{\beta}_{4l}^{T}\vec{x}\right) \end{aligned} \quad (23)$$

Select all $\{\bar{\beta}\}$ vectors to be independent and assume $r_2 = .. = r_4 = 5$ to simulate that the 2nd, 3rd, and 4th order tensors are rank-deficient. In this case, there is a total of 16 different directions in the phase space, along which input data perturbations can change the output response.

Fig. 1 compares Sobol's index [12] (a measure of the importance of a variable to the total variance of the response) for some input variable using both the new method, and a global sampling method. It is noticed that with only 160 samples (10 times the number of reduced variables) one can obtain the same estimate obtained with 1000 random samples using the existing random approach. The 160 samples are selected such as they are constrained to the subspace spanned by the vectors $\{\bar{\beta}\}$.

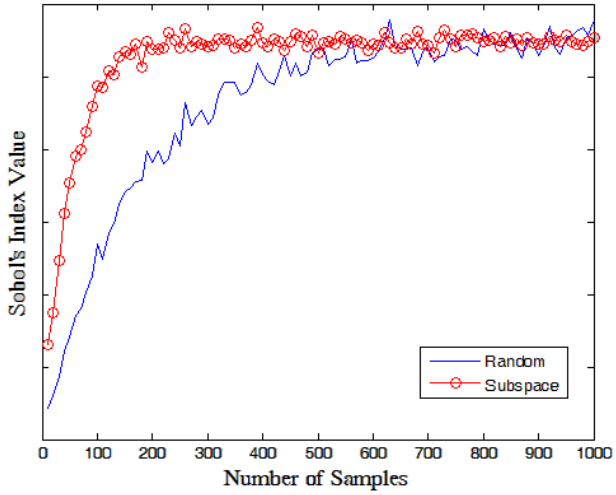

Fig. 1 Sobol's index for an input variable

4. CONCLUSIONS

A new approach has been developed to allow sampling of nonlinear models with many input data. The approach exploits the advantages of sampling methods in determining detailed PDFs for the responses when the models are nonlinear, and sensitivity-based methods for reducing the number of samples for computational efficiency. The approach samples local sensitivity information to identify a subspace that captures all parameters interactions that are responsible for the nonlinear behavior. A tensor-free expansion of the computational model could be used to prove mathematically that the sampling of local sensitivity information allows one to identify with confidence important parameters interactions. This manuscript demonstrated a proof of principle for the proposed approach. A following journal article will provide rigorous mathematical proof for the construction of the subspace.

ACKNOWLEDGMENTS

This work was supported in part by the DOE Nuclear Energy Advanced Modeling and Simulation (NEAMS) program, and the Consortium for Advanced Simulation of Light Water Reactors (www.casl.gov), an Energy Innovation Hub (http://www.energy.gov/hubs) for Modeling and Simulation of Nuclear Reactors under U.S. Department of Energy Contract No. DE-AC05-00OR22725.

REFERENCES

1. Wilhelmus H.A. Schilders, Henk A. van der Vorst, Joost Rommes, editors, *Model Order Reduction: Theory, Research Aspects and Applications*, Berlin, Springer, (2008).
2. H. Abdel-Khalik, "On Nonlinear Reduced Order Modeling," submitted to Nuclear Science and Engineering, December (2010).
3. Dan Cacuci, *Sensitivity and Uncertainty Analysis*, Chapman & Hall, CRC Press, (2003).
4. B. M. Adams, W. J. Bohnhoff, et al., DAKOTA: A multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 5.0 users manual, Tech. Rep. SAND2010-2183, Sandia National Laboratories, December 2009.
5. Box, G. E. P. and Draper, Norman. 2007, *Response Surfaces, Mixtures, and Ridge Analyses*, Second Edition, Wiley, (1987).
6. Lighthill, M. J., *Introduction to Fourier Analysis and Generalised Functions*, Cambridge, England: Cambridge University Press, (1958).
7. Lockwood, B. A., and Anitescu, M., "Gradient-Enhanced Universal Krieging for Uncertainty Propagation," Preprint ANL/MCS-P1808-1110, November, (2010).
8. Golub, G. H., Van Loan, C. F., *Matrix Computations* (3rd ed.), John Hopkins, (1996).
9. H. Abdel-Khalik, *Adaptive Core Simulation*, PhD Dissertation, North Carolina State University, December 2004 (see Appendix).
10. H. Abdel-Khalik, "Adjoint-Based Sensitivity Analysis for Multi-Component Models," Int. Conf. Math. Comp. Nucl. Sci. Eng., Brazil, (2011).
11. M. Morris, Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, **33**(2), (1991).
12. I. SOBOL, Sensitivity Estimates for Nonlinear Mathematical Models, *Mathematical Modelling and Computational Experiments*, **1**(4), (1993).