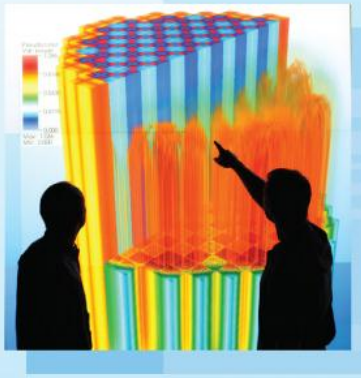
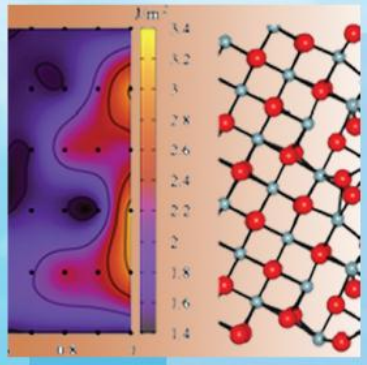
CASL-U-2011-0193-000

Power uprates
and plant life extension



Engineering design
and analysis

Science-enabling
high performance
computing



Fundamental science

Plant operational data

Consortium for Advanced Simulation of LWRs

L3:VUQ.SAUQ.P3.01

Hany Abdel-Khalik
NCSU

Completed: 9/30/11

U.S. DEPARTMENT OF ENERGY | Nuclear Energy

L3 Milestone: **VUQ.SAUQ.Y2.01**

Hybridization Framework for Uncertainty Quantification, Sensitivity Analysis, and Data Assimilation

Responsible PI: Hany Abdel-Khalik, NCSU

Overview:

This activity aims to hybridize the state-of-the-art in uncertainty quantification, sensitivity analysis, and data assimilation techniques to enable their deployment in real-world engineering models. The literature is replete with many techniques that have been developed by researchers from various scientific and engineering communities, each with its own benefits and limitations. For example, on the one hand, the global sampling techniques have proved effective in treating strongly nonlinear model with generalized probability distributions for the input parameters. It is unfortunately limited by the number of input parameters that can be handled. Variational methods, on the other hand, have proven very effective in handling models with many input parameters. They are extremely capable in estimating first order derivatives of responses of interest with respect to all input parameters. The variational estimates are however only locally accurate. Moreover, the computational cost required to extend variational methods to higher order models becomes computationally intractable.

A hybrid variational-sampling approach is therefore introduced in this work to address a real-world engineering problem where the models are generally nonlinear and the number of parameters is very large. It is important to note here that nonlinearity and the explosion in the number of input parameters are two of several other issues that must be addressed when working with real engineering models. Other issues include: how to address models with several physics and scales models that are tightly or loosely coupled? How to account for models relying on a multi-fidelity modeling strategy? Under the same spirit, we recognize that several methods have been developed to address each of these challenges. Our overarching objective is to design a hybrid framework that can combine the advantages and circumvent the limitations of existing methods when applied to a general engineering model.

Proposed Developments:

To achieve the hybridization objective, a multi-phased strategy will be implemented. When the various phases have been completed, a general framework will be proposed (estimated at the end of the second fiscal year) that can be applied to a general engineering model. The phases proposed are as follows:

1. The first phase, representing the outcome of the current work, the hybridization capability will be demonstrated using nonlinear models with many input parameters to complete uncertainty quantification, sensitivity analysis, and data assimilation. The goal is to provide an initial demonstration of the hybridization framework and highlight the advantages and possible limitations of its application.
2. The second phase will extend the hybridization framework to facilitate the transfer of information across multi-physics models. This will be developed by the current PI in the second fiscal year in collaboration with to-be-designated personnel from the VUQ team.
3. The third phase, representing another parallel activity to be completed by Dr. Mike Eldred of the VUQ team, will focus on multi-fidelity models.

The current milestone (L3.VUQ.SAUQ.Y2.01) will focus only on the first phase, that's the hybridization of a global sampling method and a local variational method. In our nomenclature, a global sampling method refers to any method that relies on executing the forward model at a set of points in the input parameters space. The points are often selected randomly or according to some heuristic to optimally cover the parameters space, e.g. random sampling, stratified sampling, sparse-grid sampling, etc. The responses recorded from the various executions are then combined statistically, e.g. using methods like Analysis of Variance, Sobol indices, or via the use of a surrogate, e.g. response surface methods such as polynomial chaos, to capture the relationship between the input parameters and responses. This relationship could then be used to extract key sensitivities, propagate parameters uncertainties, and perform data assimilation. Variationl methods refer to any method that relies on the use of the adjoint model to calculate an adjoint function which can be used in conjunction with the forward solution to calculate first-order response variations with respect to general input parameters perturbations. The advantage is that response variations could be calculated cheaply without having to re-execute the forward model. For higher order variations, additional number of adjoints must be evaluated. The goal of the current milestone is to avoid the numerous forward model executions required by global sampling methods to ensure proper coverage of the parameters space, and the additional adjoint model executions required by variational methods to account for higher order variations.

In particular, we demonstrate the hybridization for three applications as follows:

1. Hybridization of sampling and variational approach for *sensitivity analysis*. In this application, a response surface method is employed with polynomial functions as the basis for the response surface. First order variational methods are employed to reduce the dimensionality of the input parameters before the response surface method is applied. The idea is to employ variational methods to identify the parameters that have the most influence on all model responses. The degree of influence is determined via a user-defined error criterion. All parameters whose associated response variations are above this error criterion are considered influential. The influential parameters are described by a mathematical

subspace such that all parameters perturbations that are orthogonal to this subspace produce response variations that are below the user-defined error criterion. Once the influential parameters are determined, sampling methods are employed to build a surrogate in terms of the influential parameters. When the surrogate is identified, the relationship between the responses and the original parameters could be recovered.

Details on the theory and implementation will be presented at the ANS winter meeting in Washington DC, 2011 under the title: "Response Surface Modeling with Subspace-Enhanced Regression Analysis". The summary may be found in Attachment I of this document.

2. Hybridization of sampling and variational methods for *uncertainty quantification*. In this application, we recognize that the propagated uncertainty for a given response is a composite of two different quantities: a) the uncertainty in the parameters, and b) the sensitivity of the response to the parameters. If variational methods are employed, one approximates the relationship between the responses and input parameters using first order derivatives. If the input parameters uncertainties are Gaussian, one can directly calculate the uncertainties in the responses, also expected to be Gaussian. One can also propagate higher order moments of the parameters probability distributions if so desired. If sampling methods are employed instead, one first identifies a surrogate (generally nonlinear) that approximates as accurately as possible the relationship between responses and parameters. This surrogate is then sampled using a variety of sampling techniques to propagate uncertainties. In both methods, the two different quantities (parameter uncertainty and response sensitivity to the respective parameters) which form the response uncertainties are treated separately, that is first one finds an approximate relationship between responses and parameters which represents sensitivity information, and then samples this relationship using uncertainty information. The proposed approach hybridizes these two methods by identifying a set of influential parameters. The notion of influence here implies that the identified parameters contribute to the responses uncertainties. For example, a parameter than has high sensitivity but very low uncertainty may be considered non-influential since it won't contribute to the response uncertainty. To explain this, consider a simple model with three input parameters, one with high uncertainty and low sensitivity such that their product is below some user-defined error criterion, one with low sensitivity and high uncertainty with the product also below the criterion, and one final parameter with high sensitivity and high uncertainty above the criterion. With variational methods, one tries to approximate the sensitivities of the most dominant parameters. With sampling methods, one propagates uncertainties for the most uncertain parameters. With the new approach however, only one parameter is considered influential. The influential parameters are described by a mathematical subspace. This subspace aggregates all directions in the parameters phase space whose associated uncertainties give rise to responses uncertainties. All directions orthogonal to this subspace are associated with sensitivities and uncertainties whose products are smaller than some user-defined error criterion and hence could be discarded from the variational analysis employed

to determine sensitivities or the sampling analysis employed to propagate uncertainties. The mathematical background is developed in this work to identify this subspace.

Details on the theory and implementation will be presented at the ANS winter meeting in Washington DC, 2011 under the title: "On the Propagation of Uncertainties in High Dimensional Models". The summary may be found in Attachment II of this document.

3. Hybridization of sampling and variational methods for *data assimilation*. Data assimilation answers the following question: given input parameters uncertainties and measured responses, how can one adjust parameters to improve agreement between measurements and predictions for responses of interest. A parameter is adjusted based on three different factors: a) if it impacts the response of interest, i.e. has high sensitivity; b) if it has a noticeable uncertainty; and c) if it impacts a response whose associated measurement has low uncertainty. Condition (a) is important because if the parameter has weak sensitivity, it will have to be adjusted by very large amount to impact the response of interest, which is likely to be outside its uncertainty range. Condition (b) is necessary to ensure adjustments are statistically consistent within their prior uncertainties, i.e. a parameter whose prior uncertainty is very small implies that it is known very well and hence should not be adjusted by large amounts even if it has strong sensitivity. Finally condition (c) implies that responses associated with small measured uncertainties should have more impact on adjusting parameter as compared to measurements with high uncertainties. One can employ the ideas presented earlier to find a subspace of influential parameters where now an influential parameter represents a direction in the parameters space which is associated with high uncertainty, high sensitivity, and can vary a response with low uncertainties. Again a user-defined criterion could be employed to determine the size of that subspace.

   Details on the theory and implementation will be presented at the ANS winter meeting in Washington DC, 2011 under the title: "Dimensionality Reduction in Global Nonlinear Optimization". The summary may be found in Attachment II of this document.

The above demonstrations have been applied to a real engineering model representing a PWR fuel assembly. The cross-sections uncertainties are propagated to calculate uncertainties in the *k*-eigenvalue, the various reaction rates densities, the flux distribution, and the homogenized cross-sections. While serving as a demonstration of the hybrid framework, these results will form the basis for future work focusing on the propagation of neutronics uncertainties in reactor core calculations. This will be an integral part when sources of uncertainties from the various physics are to be combined to address the CASL challenge problems.

ATTACHMENT I

Response Surface Modeling with Subspace-Enhanced Regression Analysis

Youngsuk Bang and Hany S. Abdel-Khalik

Department of Nuclear Engineering, North Carolina State University, Raleigh, NC 27695
ybang@ncsu.edu; abdelkhalik@ncsu.edu

INTRODUCTION

Response surface methodology (RSM) is a mathematical technique employed to approximate the relationship between model's output responses and input parameters, often done by treating the model as a black-box [1]. RSM assumes a functional form for the response surface with some unknown coefficients. A minimization problem determines the coefficients that minimize the discrepancies between the surface's predicted responses and original model's responses generated from multiple model executions. If done properly, the response surface could be employed instead of the original computationally expensive model for engineering studies that require repeated model execution such as sensitivity analysis and uncertainty quantification. In that case, the response surface can be called as a surrogate model or a reduced order model (ROM).

To construct a reliable ROM, the number of original model's execution becomes dependent on the number of input parameters and the order of the surrogate. For very large scale models such as those employed in nuclear reactor calculations, the number of input parameter is often very large; thereby limiting the applicability of RSM to models with small number of parameters only.

In this summary, we present an algorithm to reduce the number of original model's execution in order to render practical the construction of ROM for models with large number of input parameters. We show that the computational cost is dependent on the size of the so-called active subspace rather than the size of the input parameters. The new algorithm requires the local (first-order) derivatives of the responses with respect to input parameters. By sampling the derivatives over several points in the input parameter space, one can identify an active subspace of size *r* that can be employed to capture all possible variations of responses of interest.

MATHEMATICAL DESCRIPTION

Consider a simple k^{th} order polynomial function which is assumed as unknown[1]:

$$y(\vec{x}) = \sum_{i=1}^{k} \left(\vec{a}_i^{\,T} \vec{x}\right)^i \tag{1}$$

where $\vec{a}_i = \begin{bmatrix} a_i^1 & \cdots & a_i^n \end{bmatrix}^T$ and n denotes the dimension of input parameters space. The derivative of Eq. (1) is given by:

$$\frac{dy(\vec{x})}{d\vec{x}} = \sum_{i=1}^{k} \left\{ i \times \left(\vec{a}_i^{\,T} \vec{x}\right)^{i-1} \right\} \vec{a}_i \tag{2}$$

This shows that all possible $\frac{dy(\vec{x})}{d\vec{x}}$ belong to a subspace spanned by the vectors $\{\vec{a}_i\}$. By evaluating derivatives at random points in the input parameters space, and building a matrix of the form:

$$\begin{bmatrix} \left.\frac{dy}{d\vec{x}}\right|_{\vec{x}_1} & \cdots & \left.\frac{dy}{d\vec{x}}\right|_{\vec{x}_r} \end{bmatrix},$$

One can construct the same subspace spanned by the vectors $\{\vec{a}_i\}$. This idea was first introduced to build low rank approximation for first-order derivative operators in Ref. 3. It was later proved that the errors resulting from the low rank approximation could be controlled, see Ref. 4.

Note that any input parameters perturbations that is orthogonal to this subspace will not change the response, hence the notion of the 'active' subspace. After the active subspace is identified, an orthonormal basis $\{\vec{q}_i\}$ is constructed. By limiting input parameters perturbation to the active subspace, one can write $\vec{x}$ as:

$$\vec{x} = \vec{x}_0 + \sum_{i=1}^{r} \alpha_i \vec{q}_i = \vec{x}_0 + \mathbf{Q}\vec{\alpha} \tag{3}$$

where $\vec{\alpha} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_r \end{bmatrix}^T$. Because both $\{\vec{a}_i\}$ and $\{\vec{q}_i\}$ span the same active subspace, one can write:

$$\vec{a}_i = \tilde{a}_i^1 \vec{q}_1 + \cdots + \tilde{a}_i^r \vec{q}_r = \mathbf{Q}\tilde{a}_i \tag{4}$$

where $\tilde{a}_i = \begin{bmatrix} a_i^1 & \cdots & a_i^r \end{bmatrix}^T$. Without loss of generality, $\vec{x}_0$ is assumed to zero for simplicity. Then, Eq. (1) can be written as:

$$y(\vec{\alpha}) = \sum_{i=1}^{k} \left(\tilde{a}_i^{\,T} \vec{\alpha}\right)^i \tag{5}$$

Note that Eq. (5) is transformed into a function of $\vec{\alpha}$ of only r input parameters. If $r \ll n$, RSM could be used to capture the functional form in Eq. (5) in a much more computationally efficient manner than the direct application of RSM to Eq. (1).

[1] This special form for the polynomial is used to simplify the illustration. Application to general polynomial form could be demonstrated using a modified polynomial presentation (see Ref. [2]).

This approach can be extended for models with multi-responses. As an example, consider a model with two responses given by:

$$y_1(\vec{x}) = \sum_{i=1}^{k} \left(\vec{a}_i^{\,T} \vec{x}\right)^i \tag{6}$$

$$y_2(\vec{x}) = C_1 \sin\left(\vec{b}^T \vec{x}\right) \tag{7}$$

Define a pseudo-response and its derivative by:

$$y(\vec{x}) = r_1 y_1(\vec{x}) + r_2 y_2(\vec{x}) \tag{8}$$

$$\frac{dy(\vec{x})}{d\vec{x}} = r_1 \sum_{i=1}^{k} \left\{ i \times \left(\vec{a}_i^{\,T} \vec{x}\right)^{i-1} \right\} \vec{a}_i + r_2 C_1 \cos\left(\vec{b}^T \vec{x}\right) \vec{b} \tag{9}$$

where r_1 and r_2 are scalars which can be chosen randomly except zero.

As done before, all possible variations of $\frac{dy(\vec{x})}{d\vec{x}}$ may be described by a subspace that is spanned by $\{\vec{a}_i\}$ and $\vec{b}$. By sampling $\frac{dy(\vec{x})}{d\vec{x}}$ at multiple points in the input parameters space, the active subspace could be determined in a similar manner [2].

NUMERICAL TEST

Consider the following two responses:

$$y_1(\vec{x}) = \vec{a}_1^{\,T} \vec{x} + \left(\vec{a}_2^{\,T} \vec{x}\right)^2 \tag{10}$$

$$y_2(\vec{x}) = \left(\vec{a}_3^{\,T} \vec{x}\right)^3 \tag{11}$$

where $\vec{a}_1$, $\vec{a}_2$, $\vec{a}_3$ and $\vec{x}$ are $\mathbb{R}^{100\times 1}$. Define a pseudo-response by:

$$y(\vec{x}) = \vec{a}_1^{\,T} \vec{x} + \left(\vec{a}_2^{\,T} \vec{x}\right)^2 + \left(\vec{a}_3^{\,T} \vec{x}\right)^3 \tag{12}$$

Then,

$$\frac{dy(\vec{x})}{d\vec{x}} = \vec{a}_1 + 2\left(\vec{a}_2^{\,T} \vec{x}\right)\vec{a}_2 + 3\left(\vec{a}_3^{\,T} \vec{x}\right)^2 \vec{a}_3 \tag{13}$$

Note that $\frac{dy(\vec{x})}{d\vec{x}}$ is spanned by three vectors, $\vec{a}_1$, $\vec{a}_2$, and $\vec{a}_3$. Once the active subspace basis $\mathbf{Q}$ is determined by range identification algorithm [4], the original equations can be transformed into:

$$y_1(\vec{\alpha}) = \tilde{a}_1^{\,T} \vec{\alpha} + \left(\tilde{a}_2^{\,T} \vec{\alpha}\right)^2 \tag{14}$$

$$y_2(\vec{\alpha}) = \left(\tilde{a}_3^{\,T} \vec{\alpha}\right)^3 \tag{15}$$

where $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{a}_3$ and $\vec{\alpha}$ are $\mathbb{R}^{3\times 1}$. A curve of $y(\vec{\alpha})$ for randomly sampled $\vec{\alpha}$ can be fitted by any regression method. Note that in Eq. (10) and Eq. (11), the 300 coefficients should be determined by regression analysis. On the other hand, in Eq. (14) and Eq. (15), only 30 coefficients are to be estimated. This implies that the regression analysis on the transformed equation requires less number of simulations.

In this study, MATLAB built-in function *nlinfit* is used for nonlinear least squares data fitting. Once $\tilde{a}_1$, $\tilde{a}_2$, and $\tilde{a}_3$ are determined, the original equation can be obtained by changing variables; $\vec{a}_1 = \mathbf{Q}\tilde{a}_1$, $\vec{a}_2 = \mathbf{Q}\tilde{a}_2$, $\vec{a}_3 = \mathbf{Q}\tilde{a}_3$, and $\vec{x} = \mathbf{Q}\vec{\alpha}$. The test results are shown in Figure 1 and Figure 2. First, $\vec{a}_1$, $\vec{a}_2$, and $\vec{a}_3$ are selected randomly and the suggested method is implemented. The RMS (root mean square) error describing the discrepancy between the responses evaluated by the original model and those by the model estimated by the new method are found to be 4.4115E-10 and 1.9378E-09, respectively, for the two responses. These are numerically negligible considering the precision of the calculations.

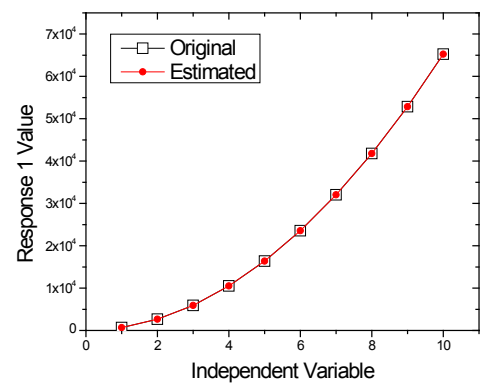


Fig. 1 Response 1 Predicted Variations

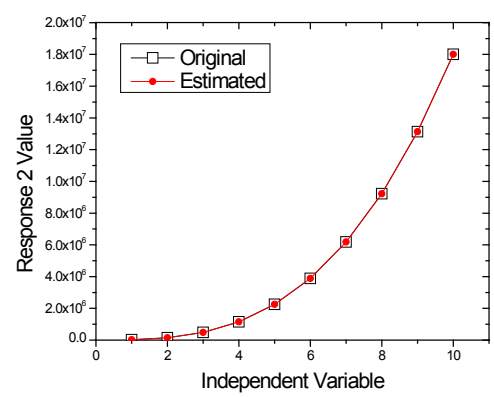


Fig. 2. Response 2 Predicted Variations.

CONCLUSION

Computationally efficient RSM enhanced by aggregating first order derivatives in a subspace is presented. Before conducting a regression analysis, the

equation can be transformed into one with lower dimensions for the input parameters. The method can be easily extended to models with many responses via the use of pseudo responses. Numerical test demonstrates that the suggested algorithm significantly reduces the computational cost without loss of accuracy.

ACKNOWLEDGEMENT

This work was supported by 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. Raymond H. Myer et al, *Response Surface Methodology*, Third Edition, A John Wiley & Sons, Inc., Publication, Hoboken, New Jersey, (2009).
2. Hany S. Abdel-Khalik and Jason M. Hite, "Reduced Order Modeling: Tensor-Free Expansion for Nonlinear Features Identification", Transactions of the American Nuclear Society, Hollywood FL, June, (2011).
3. Abdel-Khalik, H., “Adaptive Core Simulation”, PhD Dissertation, North Carolina State University, (2004).
4. N. Halko, P.G. Martinsson, J. Tropp, “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions.” SIAM Review, 53(2), (2011).

On the Propagation of Uncertainties in High Dimensional Models

Youngsuk Bang, Christopher Kennedy, and Hany S. Abdel-Khalik
Department of Nuclear Engineering, North Carolina State University, Raleigh, NC 27695
{ybang,cbkenned,abdelkhalik}@ncsu.edu

INTRODUCTION

This summary focuses on the propagation of parameter uncertainties in models associated with many input parameters and many output responses. Fundamentally, there are two approaches that have been widely used in the literature to propagate input uncertainties: the forward approach and the adjoint approach [1]. The forward approach is effective for models with few input parameters and many output responses. Conversely, the adjoint approach has proven powerful for models with many input parameters and few responses. A recently introduced reduced order modeling (ROM) approach has hybridized the forward and adjoint approaches to reduce the computational overhead for models with many input parameters and output responses [2]. This summary employs these developments to introduce yet another approach to further reduce the computational overhead required to propagate uncertainties. The focus in this summary will be on linearized models. Extension to nonlinear models represents an ongoing activity and will be discussed in a future summary.

MATHEMATICAL DESCRIPTION

Let the given model by described by:

$$\vec{y} = \vec{\Theta}(\vec{x}) \tag{1}$$

where $\vec{x} \in \mathbb{R}^n$ is a vector representing the n parameters and $\vec{y} \in \mathbb{R}^m$ the m responses. The prior uncertainties about $\vec{x}$ are described by a covariance matrix $\mathbf{C}_x$. One can show using a first-order approximation that the responses uncertainties may also be described by a covariance matrix:

$$\mathbf{C}_y = \mathbf{\Theta}\mathbf{C}_x\mathbf{\Theta}^T \tag{2}$$

where $\mathbf{\Theta}$ is denoted as the sensitivity matrix whose elements are given by:

$$[\mathbf{\Theta}]_{ij} = \frac{\partial y_i}{\partial x_j}$$

Eq. (2) is often denoted as the sandwich rule as the covariance matrix is multiplied from right and left by the sensitivity matrix. There are fundamentally two approaches to evaluate Eq. (2). The first approach, denoted here as the explicit approach, evaluates the full sensitivity matrix using either forward or adjoint sensitivity analysis and then calculates the triple matrix product in Eq. (2). The premise is that the first order derivatives can be used to approximate the action of the original model, i.e., they act as a surrogate that approximates variations in the responses due to variations in input parameters.

The second approach, denoted by the implicit approach, avoids the explicit formation of the sensitivity matrix by taking advantage of the low rank structure of the sensitivity matrix and/or the parameters covariance matrix. The most prominent approach here is the Karhunen-Loeve (KL) expansion, which expands the parameters covariance matrix using singular value decomposition [3]. Assuming the effective rank of the covariance matrix is r_x, the KL approach requires only r_x forward model executions to directly evaluate Eq. (2).

Another implicit approach, presented in an earlier work, takes advantage of the low rank structure of the sensitivity matrix by employing a randomized approach to calculate a rank revealing decomposition for the sensitivity matrix [4]. This approach requires r_Θ adjoint and r_Θ forward model executions. This approach is similar to the surrogate-based approach described earlier but completed with less computational overhead. It is advantageous when $r_\Theta \ll \min(r_x, m, n)$; the physical rational behind it is to propagate uncertainties for parameters or combination of parameters that have strong sensitivities. This is sound since parameters with very low sensitivities do not affect the response and hence are not expected to affect the responses uncertainties. Alternatively, the KL approach is more advantageous when $r_x \ll \min(r_\Theta, m, n)$ which physically attempts to propagate uncertainties for parameters that have high uncertainty. The rationale here is that parameters with very small uncertainties will not affect the responses uncertainties since they are known with high precision. Note that in both of these approaches it is essential to select a large enough rank for the sensitivity matrix or the covariance matrix to avoid situations in which parameters have very small uncertainties and very strong sensitivities, or alternatively very small sensitivities and very high uncertainties.

In this summary, a new approach is proposed. Instead of building a surrogate to approximate all possible responses variations with respect to input parameters, the surrogate will be confined to predict variations in responses resulting from input parameters uncertainties. This implies that if a given response is very sensitive to input parameters variations but its uncertainty is zero or

below some user-defined tolerance, the surrogate will predict a zero variation in that response due to input parameters variations.

A recently developed ROM approach is employed to build the surrogate [2]. This approach hybridizes variational methods and sampling methods to identify a subspace in the input parameters space that is employed to construct the surrogate. The premise is that parameters variations that belong to this subspace give rise to uncertainties in the responses; whereas parameters variations that are orthogonal to this subspace do not give rise to uncertainties in the response of interest.

It is important to note that several hybrid methodologies have been proposed by other researchers but with different objectives. To avoid confusion, we note briefly the differences between these methodologies and the proposed approach. Methods like gradient-enhanced regression [5] and adjoint-based interpolation [6] build surrogates for the responses by assuming a response surface in the response space and employ derivatives to reduce the number of model runs required to approximate the surrogate. The surrogate is then used to propagate parameters uncertainties using a sampling method. In our approach, the derivatives are employed to identify a subspace in the input parameters space. The objective here is to identify the directions in the input parameters space responsible for the propagated uncertainties. Therefore the surrogate generated is pre-optimized for uncertainty analysis.

Two variations of this ROM approach are introduced here. The first one emulates the KL-based approach; it is re-iterated here for comparison to the new algorithm [4]: re-write Eq. (2) as follows:

$$\mathbf{C}_y = \mathbf{\Theta}\mathbf{W}_x\mathbf{\Sigma}_x^2\mathbf{W}_x\mathbf{\Theta}^T = \left(\mathbf{\Theta}\mathbf{W}_x\mathbf{\Sigma}_x\right)\left(\mathbf{\Theta}\mathbf{W}_x\mathbf{\Sigma}_x\right)^T$$

where $\mathbf{W}_x\mathbf{\Sigma}_x^2\mathbf{W}_x$ is the SVD of $\mathbf{C}_x$. We assume that the rank of $\mathbf{C}_x$ is given by r_x. Now, notice that:

$$\mathbf{\Theta}\mathbf{W}_x\mathbf{\Sigma}_x = \mathbf{\Theta}\left[\sigma_{x,1}\vec{w}_{x,1} \quad \dots \quad \sigma_{x,r}\vec{w}_{x,r_x}\right]$$

These matrix vector products could be approximated using a matrix-free approach as follows:

$$\mathbf{\Theta}\sigma_{x,j}\vec{w}_{x,j} \simeq \vec{\Theta}\left(\vec{x}_0 + \sigma_{x,j}\vec{w}_{x,j}\right) - \vec{\Theta}\left(\vec{x}_0\right) \quad j = 1,\dots,r_x$$

This of course assumes that the perturbations along the $\vec{w}_{x,j}$ vectors, sometimes referred to as the principal directions of the covariance matrix, are small enough to ensure the linearity assumption is valid. This algorithm requires r_x forward model executions only. Notice that the sensitivity matrix $\mathbf{\Theta}$ is not constructed explicitly.

The second ROM-based algorithm which represents the contribution of this summary takes advantage of the fact that both the sensitivity matrix and the covariance matrix may be ill-conditioned.

Before introducing the algorithm, we motivate the idea first. Consider a linear model with three input parameters: x_1, x_2, and x_3 with normal distributions and uncorrelated standard deviations: σ_1, σ_2, and σ_3. Assume that $\sigma_2 = 0$ and $\partial y/\partial x_1 = 0$. In this case, the propagated uncertainty in y is only determined by: $\sigma_y = \left|\sigma_3\left(\partial y/\partial x_3\right)\right|$. This implies that only the parameter x_3 which has non-zero sensitivity and non-zero uncertainty impacts the response's uncertainty. Such parameter from an uncertainty propagation point of view is considered influential. This is not to be confused with influential parameters from sensitivity analysis point of view. Identifying influential parameters for uncertainty propagation is important for data assimilation as well. This follows as any robust data assimilation procedure attempts to adjust parameters that have both high uncertainty and strong sensitivity. Adjusting parameters with weak sensitivity does not change the response of interest, and adjusting parameters with low uncertainties does not lead to a physical adaption. Therefore, it is the product of the two that is important for both uncertainty propagation and data assimilation.

DETAILS OF IMPLEMENTATION

We introduce below a hybrid ROM-based UQ algorithm to determine the influential directions in the parameters space. The algorithm proceeds as follows:

1. Set $k = 1$
2. Generate a random vector $\vec{\alpha}_k$
3. Construct a pseudo response: $y_{pseudo} = \vec{\alpha}_k^T\vec{y}$. A pseudo response is a linear combination of all model's responses. This allows the algorithm to explicitly form an adjoint solution for each individual response.
4. Employing the adjoint model, calculate: $\dfrac{\partial y_{pseudo}}{\partial \vec{x}}$.
5. Update: $\mathbf{G} = \left[\left.\dfrac{\partial y_{pseudo}}{\partial \vec{x}}\right|_{\vec{\alpha}_1} \quad \dots \quad \left.\dfrac{\partial y_{pseudo}}{\partial \vec{x}}\right|_{\vec{\alpha}_k}\right] \in \mathbb{R}^{n\times k}$
6. Using range finding algorithm, calculate the rank r_Θ of the sensitivity matrix $\mathbf{\Theta}$. The range finding algorithm identifies a low rank approximation for $\mathbf{G}$ such that the matrix reconstruction error measured in an induced L_2 norm meets some user-defined error tolerance. It is important to select this error tolerance to not exceed the accuracy of derivatives. For example, if the derivatives are accurate to only 5 significant digits, employing a tighter tolerance will not improve the accuracy of the low-rank approximation. Moreover, since forward sensitivity analysis is employed by this algorithm, the tolerance should be consistent with the accuracy expected for forward-based sensitivity analysis. This is often in the order of 2 to 3 significant digits.

7. If the rank has not been reached yet, increase k and return to step 2.
8. Decompose $\mathbf{G}$ using a rank revealing decomposition, e.g., QR factorization $\mathbf{G} = \mathbf{QR}$, where $\mathbf{Q} \in \mathbb{R}^{n \times r_\Theta}$
9. Re-write Eq. (2) as follows:

$$\mathbf{C}_y = \mathbf{\Theta Q Q}^T \mathbf{C}_x \mathbf{Q Q}^T \mathbf{\Theta}^T = \mathbf{\Theta Q} \tilde{\mathbf{C}}_x \mathbf{Q}^T \mathbf{\Theta}^T$$

$$\mathbf{C}_y = \left(\mathbf{\Theta Q} \tilde{\mathbf{W}}_x \tilde{\Sigma}_x\right)\left(\mathbf{\Theta Q} \tilde{\mathbf{W}}_x \tilde{\Sigma}_x\right)^T$$

10. As done before, the last term can be evaluated as follows:

$$\mathbf{\Theta Q} \tilde{\mathbf{W}}_x \tilde{\Sigma}_x = \mathbf{\Theta}\left[\tilde{\sigma}_1 \bar{\tilde{w}}_1 \quad \ldots \quad \tilde{\sigma}_r \bar{\tilde{w}}_r\right]$$

$$\mathbf{\Theta Q} \tilde{\sigma}_j \bar{\tilde{w}}_j = \mathbf{\Theta}\left[\bar{q}_1 \quad \ldots \quad \bar{q}_{r_\Theta}\right] \tilde{\sigma}_j \bar{\tilde{w}}_j \quad j = 1, \ldots, r$$

$$= \mathbf{\Theta} \tilde{\sigma}_j \sum_{k=1}^{r_\Theta} \tilde{w}_{j,k} \bar{q}_k \simeq \bar{\Theta}\left(\bar{x}_0 + \tilde{\sigma}_j \sum_{k=1}^{r_\Theta} \tilde{w}_{j,k} \bar{q}_k\right) - \bar{\Theta}\left(\bar{x}_0\right)$$

This approach calculates an effective covariance matrix for the input parameters $\tilde{\mathbf{C}}_x$ whose SVD (i.e. singular vectors, also referred to as principal directions) point towards the directions that have both strong sensitivities as determined by the matrix $\mathbf{Q}$ and high uncertainties as determined by $\mathbf{C}_x$. Based on user-defined tolerance, the rank of this matrix is determined as r. Notice that in this approach the adjoint model is executed in the order of r times to identify the matrix $\mathbf{Q}$ (the exact value is $r + s$, see below for the range-finding algorithm); the full sensitivity matrix $\mathbf{\Theta}$ is however never explicitly calculated.

Employing arguments from linear algebra, one can show that the effective range for $\tilde{\mathbf{C}}_x$ has a dimension r that satisfies [7]: $r \le \min\left(r_\Theta, r_x\right)$ due to the following identify, valid for any general matrices **A** and **B**: $\dim(\mathrm{R}(\mathbf{AB})) \le \dim(\mathrm{R}(\mathbf{A}))$.

Finally, for the sake of a complete discussion, we describe briefly the range-finding algorithm employed to find an effective rank/decomposition for the matrix $\mathbf{G}$ [8].

1. Given user-defined tolerance ε
2. Given matrix $\mathbf{G} \in \mathbb{R}^{n \times k}$
3. Given a small number s, e.g., $s = 10$
4. Evaluate the derivatives s extra times to form:

$$\bar{f}_i = \left.\frac{\partial y_{pseudo}}{\partial \bar{x}}\right|_{\bar{\alpha}_i}, \quad i = 1, \ldots, s.$$

Note that these vectors, denoted as oversamples, are generated once and are independent of the k vectors in the matrix $\mathbf{G}$; they are employed to verify if the user-defined tolerance is met.

5. Calculate: $\mathbf{G} = \mathbf{QR}$
6. Let r_0 be initial estimate of the rank.
7. Let $\mathbf{Q}_{r_0} = \mathbf{Q}\left(:, 1 : r_0\right)$; this matrix contains the first r_0 of the matrix $\mathbf{Q}$.
8. Calculate: $z_i = \left\|\left(\mathbf{I} - \mathbf{Q}_{r_0}\mathbf{Q}_{r_0}^T\right) f_i\right\|, \quad i = 1, \ldots, s$
9. If the condition $10\sqrt{2/\pi} \max_{i=1,\ldots,s} z_i \le \varepsilon$ is not satisfied, increase r_0 and go back to step 7. If the condition is satisfied, one can show that with probability $1 - 10^{-s}$ that the matrix reconstruction error is bounded by:

$$\left\|\left(\mathbf{I} - \mathbf{Q}_{r_0}\mathbf{Q}_{r_0}^T\right)\mathbf{G}\right\| \le \varepsilon.$$

The total cost for this algorithm is $r + s$ adjoint model evaluations.

CASE STUDIES

The first case study is employed as a sanity check to compare the uncertainties evaluated by the various approaches. Note that employing an adjoint algorithm for a single response would require a single adjoint and a single forward model evaluation. Thus, clearly if only the uncertainty of a single response is required, the adjoint-based UQ would be the most suitable approach. This case study employs a TSUNAMI-2D cell model [9] with 132 input parameters representing the fission cross sections of U234, U235 and U238 with 44 energy groups. The responses are defined as follows:

$$R_j^g = \frac{\varphi_j^g}{\sum \varphi_j^g}$$

where φ_j^g denoted as the total flux in mixture j and group g is defined by[1]:

$$\varphi_j^g = \sum_{i \in mixture\ \mathrm{j}} \phi_i^g V_i^{cell}$$

and ϕ_i^g is the flux in cell i and group g and V_i^{cell} is the volume of ith cell.

A manufactured covariance matrix $\mathbf{C}_x \in \mathbb{R}^{132 \times 132}$ with rank $r_x = 50$ is employed. The first ROM-based UQ approach is employed with machine precision tolerance to propagate all 50 principal directions of the covariance matrix. In this case, 50 forward model executions are required. Table I compares the k-eigenvalue uncertainty calculated by the ROM-based algorithm, the conventional adjoint-based algorithm and also a sampling-based approach. Results are within a few pcm of each other, which is within the accuracy of the first-order approximation.

Table I. Comparison of Uncertainty Propagation

Method	Std. Dev. [pcm]
Adjoint UQ (Sandwich Rule)	461.9848
ROM-based UQ	458.9782
Sampling (1000 samples)	460.4756
Sampling (500 samples)	461.0369
Sampling (100 samples)	467.9175

[1] In SCALE, a mixture refers to all regions in the space that have the same composition.

The purpose of the second case study is to show that the rank of the effective covariance matrix is less than the minimum of the ranks of the sensitivity matrix and the parameters covariance matrix which is the reason behind the power of the proposed approach. This case study employs an HTGR infinite prismatic lattice modeled by the deterministic package NEWT/TRITON [10,11]. The model was depleted to 40 GWd/MTHM to ensure all actinides are present in the model. The input parameter space was of dimension n = 1584 representing the fission cross-sections of several actinides in 44 group energy format. Numerical tolerances for forward calculations were set to 10^{-6} for both k-eff and the flux. The covariance matrix was manufactured with singular values graphed in Fig 1 (legend 'Cx').

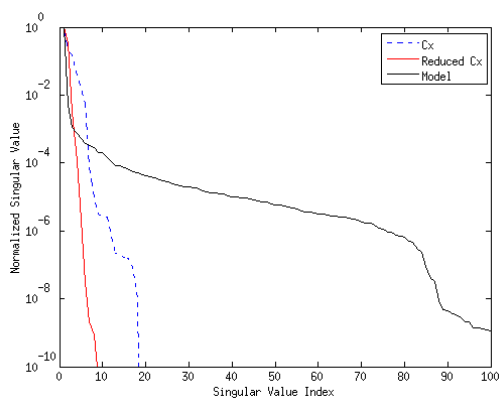


Fig. 1. Singular Value Spectra.

For comparison, the first 100 singular values of the $\mathbf{G}$ matrix are graphed as well (legend 'Model'). Finally, the singular values of the effective covariance matrix $\tilde{\mathbf{C}}_x$ are plotted on the same figure. Notice that the singular values for $\tilde{\mathbf{C}}_x$ drop off much quicker than those of $\mathbf{C}_x$ and $\mathbf{G}$; explaining the potential savings realized by the proposed approach. Based on the selected tolerances, the rank of the sensitivity matrix is $r_\Theta = 20$, and the rank of prior covariance matrix is $r_x = 90$. The rank for the effective covariance matrix is $r = 10$, implying 10 + s adjoint and 10 forward model executions.

The third test case demonstrates the ability to propagate parameters uncertainties to estimate all responses uncertainties. This case employs an assembly model; it is an OECD/NEA benchmark of a stand-alone single assembly mini-core model [12]. The scattering, fission, (n,2n), and capture cross-sections of four uranium isotopes are employed as 704 input parameters. The responses are defined as in the first test case. A total of 396 responses are calculated in 9 different mixtures. The reference responses are shown in Fig. 2. Each section on the graph represents the responses in 44-group in the respective mixture. The absolute standard deviations propagated by the new approach are compared in Fig. 3 to those calculated by the sandwich rule where the sensitivity matrix is explicitly determined.

For this case, the rank of the effective covariance matrix is found to be r = 28 based on a tolerance of 10^{-2} for the range finding algorithm for the matrix $\mathbf{G}$.

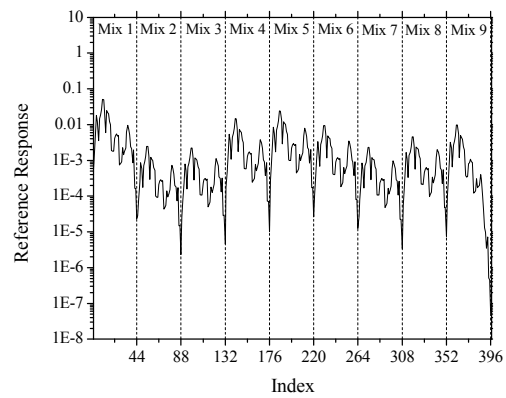


Fig. 2. Reference Responses

This tolerance implies the derivatives will be reconstructed to two significant digits, which is the same accuracy expected with forward-based sensitivity analysis.

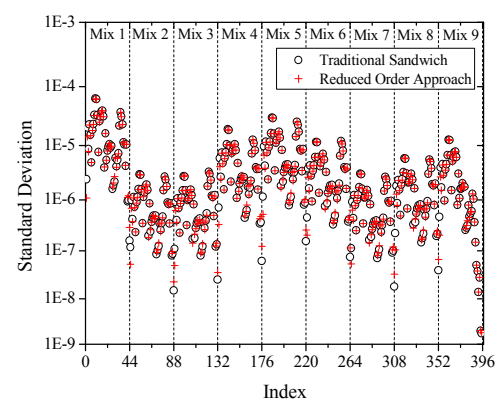


Fig. 3. Responses Absolute Standard Deviations

The associated computational cost is 28+s adjoint and 29 forward model evaluations. Note that, for the adjoint-based approach, 396 adjoint model evaluations and a single forward model evaluation would be required to propagate uncertainties for all 396 responses and, for the KL-based approach, the computational cost would be 704 forward model evaluation assuming a full rank prior covariance matrix.

CONCLUSIONS

A new ROM-based UQ approach is employed to exploit the ill-conditioning of both the covariance matrix and sensitivity matrix in order to reduce the number of adjoint and forward model executions required to propagate uncertainties for all model responses. When $r_x \ll r_\Theta$ (r_x the effective rank of the prior covariance matrix, and r_Θ effective rank of the sensitivity matrix), it is shown that only r_x forward model evaluations are needed. When r_Θ is less than or of the same size as r_x, one can reduce the number of model executions by finding an effective covariance matrix $\tilde{\mathbf{C}}_x$ that captures directions in the input parameters space that contribute the most to the propagated uncertainties. This requires executing the adjoint model r times and the forward model r times, where r is the rank of $\tilde{\mathbf{C}}_x$, another matrix that is defined by the proposed algorithm. This matrix is denoted as the effective prior covariance matrix, it captures directions in the input parameters space that give rise to the propagated responses uncertainties. The principal directions of this matrix are associated with parameters that have both strong sensitivities and high uncertainties. Using linear algebra, one can show that $r \le \min(r_x, r_\Theta)$. Three prototype models are used in this summary to demonstrate the proposed approach. Future work will focus on applying the proposed methodology to typical reactor calculations to propagate multi-group cross-sections through lattice calculations to the few-group cross-sections. Moreover, the proposed algorithm will be extended to nonlinear models where the first-order approximation is no longer acceptable to relate responses and input parameters variations. We believe this is possible as the algorithm employed for ROM construction is applicable to nonlinear models as shown in previous work [2].

REFERENCES

1. D. CACUCI, *Sensitivity and Uncertainty Analysis,* Chapman & Hall, CRC Press, (2003).
2. H. ABDEL-KHALIK, "On Nonlinear Reduced Order Modeling", *Proceedings of the International Conference on Mathematics and Computational Methods Applied in Nuclear Science and Engineering*, Rio De Janeiro, Brazil, May 8-12, (2011).
3. H. STARK, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice-Hall, Inc., (1986).
4. H. ABDEL-KHALIK, Adaptive Core Simulation, PhD Dissertation, North Carolina State University, (2004).
5. B. LOCKWOOD, et. al., "Gradient-Enhanced Universal Krieging for Uncertainty Propagation," Preprint ANL/MCS-P1808-1110, (2010).
6. Q. WANG, *Uncertainty Quantification for Unsteady Fluid Flow using Adjoint-based Approaches*, Ph.D. thesis, Stanford University, Stanford, CA, (2009).
7. G. GOLUB, *Matrix Computations* (3rd ed.), John Hopkins, (1996).
8. N. HALKO, P.G. MARTINSSON, J. TROPP, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions." SIAM Review, 53(2), (2011).
9. SCALE 6.1, "A Modular Code System for Performing Standardized Computer Analyses for Licensing Evaluations", Beta Version, Private communication, ORNL, (2010).
10. M. DEHART, "Benchmark Specification for HTGR Fuel Element Depletion," NEA/NSC/DOC(2009)13, (2009).
11. M. DEHART, "NEWT: A New Transport Algorithm for Two-Dimensional Discrete Ordinates Analysis in Non-Orthogonal Geometries," ORNL/TM-2005/39 (2009).
12. K. IVANOV, "Benchmark for Uncertainty Analysis in Modeling (UAM) for Design, Operation and Safety Analysis of LWRs", Vol. I : Specification and Support Data for Neutronics Cases (Phases I), Version 1.0, OECD Nuclear Energy Agency, (2007).

ATTACHMENT III

Dimensionality Reduction in Global Nonlinear Optimization

Jason M. Hite and Hany S. Abdel-Khalik

Department of Nuclear Engineering, North Carolina State University, Raleigh, NC 27695
jmhite@ncsu.edu, abdelkhalik@ncsu.edu

INTRODUCTION

Numerical optimization plays a key role in a broad range of science and engineering problems, both in research and in industry. An important challenge that faces many of the state-of-the-art techniques is the so-called curse of dimensionality. Two classes of methods can be identified: gradient-based methods and forward exploration methods. Gradient methods employ the gradient of the objective function to advance the search [1], whereas forward exploration methods employ only the forward model and use a predetermined criterion to advance the search. Under the forward exploration category, two methods are widely used: heuristic search methods and random search methods [2]. As an example of heuristic methods is the Nelder-Mead simplex algorithm, well-known to the Applied Mathematics community as the algorithm that powers MATLAB's *fminsearch* [3], but which is only reliable for a small number of variables [4]. As an example of random search methods is the Markov-Chain Monte Carlo method popularized by data assimilation applications requires many model executions even with a few number of input parameters [5]. For example, to optimize only 10 input parameters, it is common to require on the order of 10^4 model executions.

To solve optimization problems with many input parameters ($> 10^6$) forward exploration methods have proved to be the most practical - an example is the simulated annealing approach which employs the Metropolis-Hasting algorithm to explore the input space. Although not guaranteed to find the global minimum, simulated annealing has been employed in many engineering fields with considerable success, in particular in the nuclear engineering community [6]. Given the huge number of forward model executions required, a reduced order modeling (ROM) approach is often employed to find another model of reduced complexity that is inexpensive to run. If the fidelity of the inexpensive model is acceptable, it can be used with confidence to search the input space with a manageable computational cost.

This summary proposes a new approach in which an ROM philosophy is employed to find a reduced *input space* rather than a model with reduced complexity. The reduced input space is then explored using a user-defined optimization algorithm. The idea is to identify all directions in the input space that change the objective function by values that are less than a prescribed tolerance, most obviously chosen to match machine precision. By eliminating these directions from the input space, a significant reduction in the effective dimensionality of the input space could be achieved. The algorithm establishes a strong error bound on the maximum change in the objective function resulting from the reduction of the input space. The benefit of this approach is that the original model is still employed instead of a lower-fidelity approximation, but with a constrained search space.

The new algorithm employs derivative information to find the reduced input space. Unlike gradient-based methods, derivative information is not used to advance the search; it is only used to identify a subspace in the input space to which the random samples are constrained.

Despite their widespread use, theoretical results regarding the behavior of heuristic methods like Nelder-Mead have proven elusive [7]. The behavior of many heuristic methods in more than two dimensions is scarcely understood, thus we limit our discussion in this summary to random search methods only. This is due to their widespread use and reliable performance for a range of engineering models, namely simulated annealing and

MCMC algorithms.

The reduced input space is identified via a hybrid algorithm that statistically samples sensitivity information and aggregates it in a subspace that is subsequently used to reduce the dimensionality of the minimization problem. Notably different from other reduction algorithms, recent developments in random matrix theory are employed to set an upper bound on the error resulting from the dimensionality reduction [8]. By matching this error bound to the numerical precision of the function evaluation, an appropriate size for the subspace can be determined.

BACKGROUND

This approach is based on the *Efficient Subspace Method* as outlined in [9, 10], henceforth abbreviated as ESM. We use the algorithm described in [9] to construct a matrix $\mathbf{Q}$ that describes the active subspace, $\mathbb{S}$, of the objective function such that $\text{Range}(\mathbf{Q}) = \mathbb{S}$. $\mathbf{Q}$ can then be used to constrain the search path of the chosen optimization algorithm to lie only along the directions that produce a change in the objective function. $\mathbf{Q}$ is constructed via evaluating the derivative of the objective function at a randomly generated set of points in the input space. Prior work (see [9]) has shown that some common engineering models can achieve significant reduction in problem dimensionality with a very high probability that the reduced function still spans the full range space of the original function ([8] has some discussion on this probability). Additionally, it is only necessary to construct this basis once, after which it can be reused with any new runs of an algorithm that might benefit from reducing the dimensionality of the objective function[1].

One of the appeals of simulated annealing is that it does not require derivative information to perform the optimization. If derivative information is available, however, it is not easily able to take advantage of the added information. The ESM-based approach introduced here hopes to remedy this; given a small number[2] of gradient evaluations we can construct an approximation for the original function that is of reduced dimensionality but which closely approximates the range space of the original function. This will be most beneficial when the dimension of the active subspace is significantly smaller than the input space of the original model as a relatively small number of gradient evaluation can provide a large reduction in the number of direct function evaluations required.

METHODOLOGY

We base our demonstration on Rosenbrock's function, which is commonly used for benchmarking numerical optimization algorithms [11], given as[3]

$$\mathcal{R}([\,x\;\;y\,]^{\mathsf{T}}) = (1-x)^2 + 100(y-x^2)^2 \tag{1}$$

The novelty of (1) lies in the difficulty of finding the minimum, which is easily shown to occur at $(0,0,0)$, but lies in a thin banana-shaped valley. To reach this valley is trivial but to find the minimum once there proves difficult for many algorithms. We wish to extend this problem to a higher-dimensional one, but we want to maintain an active subspace that is of dimension 2. We will adopt the approach presented in [10], namely we will consider 2 vectors $\bar{x}$ and $\bar{y}$, with $\bar{x},\bar{y} \in \mathbb{R}^r$ and two constant vectors $\bar{\beta}_1$ and $\bar{\beta}_2$ whose entries are randomly drawn from the standard normal distribution and $\bar{\beta}_1, \bar{\beta}_2 \in \mathbb{R}^r$. We then substitute into (1) as follows:

$$\begin{aligned} f([\,\bar{x}\;\;\bar{y}\,]) &= \mathcal{R}\left(\left[\,\bar{\beta}_1^{\mathsf{T}}\bar{x}\;\;\bar{\beta}_2^{\mathsf{T}}\bar{y}\,\right]^{\mathsf{T}}\right) \\ &= (1-\bar{\beta}_1^{\mathsf{T}}\bar{x})^2 + 100(\bar{\beta}_2^{\mathsf{T}}\bar{y} - (\bar{\beta}_1^{\mathsf{T}}\bar{x})^2)^2 \end{aligned} \tag{2}$$

to obtain our objective function, $f([\,\bar{x}\;\;\bar{y}\,])$. Simply from observing (2) it should be clear that f only has two effective degrees of freedom, which lie in the directions parallel to $\bar{\beta}_1$ and $\bar{\beta}_2$. Thus, the dimension of the active subspace of f is only 2 and we can fully characterize f with a lower-dimensional function f^* that is a function of two (scalar) variables using a basis $\mathbf{Q}$ for the active subspace of f, such that $f^*(\alpha_1,\alpha_2) = f(\mathbf{Q}\bar{\alpha})$, where $\bar{\alpha} = [\alpha_1\;\alpha_2]^{\mathsf{T}}$. For the following experiments we chose $r = 6$, implying that f is a function of $2r = 12$ scalar variables in total, but with an active subspace which is still only 2 dimensions.

[1] Provided, of course, that the objective function does not change.

[2] "Small" being equal to the dimension of the active subspace of the original objective function, determined within a given error tolerance using an order-statistics based framework.

[3] Here we have written Rosenbrock's function somewhat differently from what is typical, representing x and y as the components of a vector. This is to avoid an inconsistency in notation for the objective function defined later on.

The rationale behind the choice of Rosenbrock's function may not be clear as its relevance to nuclear models is not immediately apparent. In part, we chose it to demonstrate the approach because it proves difficult to optimize for many optimization algorithms, which become trapped in the valley surrounding the minimum. This demonstrates that the method presented in this summary is not sensitive to the actual model form - if the chosen optimization algorithm is capable of reaching a minimum in the full variable space it will also reach a minimum by searching the reduced space. In truth, the given analysis can be applied to almost any functional form to yield a suitable objective function. In this context, Rosenbrock's function can also be considered as a stand-in for an arbitrary high-dimension function that has strongly correlated inputs. Other work has shown that application of a related approach to methods for uncertainty quantification and sensitivity analysis yields excellent results [9], the goal is to develop a similar approach for problems in nonlinear optimization. In future work, it is desired to examine the application of this method to reactor models, in particular fuel loading optimization.

This extension brings some extra challenges, the first of which is that the problem no longer has a unique solution for the global minimum. Instead of the global minimum being at the origin as is the case with the normal Rosenbrock function, it now lies on the curve along which $\bar{\beta}_1^\intercal \bar{x} = 0$ and $\bar{\beta}_2^\intercal \bar{y} = 0$. However, it can be seen that the points that lie on this curve do still achieve the condition that $f([\,\bar{x}\ \bar{y}\,]) = 0$, so they still produce the same minimum value as the original function. We also note that as stated, this method provides a solution in the *reduced* space and it is necessary to find the corresponding solution in the original input space. The full variable space optimum is given in the same way as the search points, namely by forming the product $\mathbf{Q}\bar{\alpha}^*$, where $\bar{\alpha}^*$ represents the reduced space solution, yielding the full-space solution corresponding to the reduced-space solution. Note that in the following demonstration section we verify that all solutions found by exploring the reduced space and then projecting into the full space are indeed true minima of the objective function.

Trivially, we could construct $\mathbf{Q}$ using the β-vectors since they are known, (they are generated at random a priori in our experiment) but in general the various $\bar{\beta}$'s are unknown as they represent the coefficients in a special expansion of f, as shown in [10]. Instead, we determine $\mathbf{Q}$ by taking the QR decomposition of a matrix $\mathbf{G}$, whose rows are the gradient of f evaluated at random points in the input space, and then discarding the $\mathbf{R}$ matrix. We note that this algorithm will only require 2 evaluations of $\bar{\nabla} f$ since random sampling from the active subspace $\mathbb{S} \subseteq \mathbb{R}^2$ can yield no more than 2 linearly independent vectors. For an arbitrary model, the number of necessary gradient evaluations is set according to a user specified level of confidence that the full variation of the function has been captured. For more elaboration on the theoretical underpinnings and practical implementation of this, see [8] and [12].

It is important to remark here that the proposed methodology does not offer any guarantees regarding the convergence of the search algorithm to the global minimum; it only provides a means to reduce the effective dimensionality of the search space. Qualitatively, one can describe the reduction algorithm as follows: for a given user-defined tolerance, reduce the effective dimensionality of the input space by excluding all directions that change the function by amounts smaller than the specified tolerance. Clearly, for functions that are sufficiently complex excluding any direction in the phase space could lead to missing the global minimum. For practical problems, however, one is often limited by the numerical precision of the function evaluation procedure. One can therefore take advantage of this by setting the tolerance to be equal to the numerical precision of the function evaluation. We address this issue in the demonstration section by applying the given approach to Rosenbrock's function, a common benchmark for optimization algorithms.

DEMONSTRATION

We demonstrate this approach using Mathematica's built-in command *NMinimize*, which encloses 4 separate algorithms: Nelder-Mead, differential evolution, simulated annealing and random search[4], and examine the number of steps required by each algorithm when optimizing f directly versus when using the reduced model f^*. Nelder-Mead, differential evolution and random search are only included for reference; we do not wish to

[4]"Random search" is Mathematica's terminology for a method that begins at a collection of random starting points and employs local optimization techniques at each point to find a global minimum.

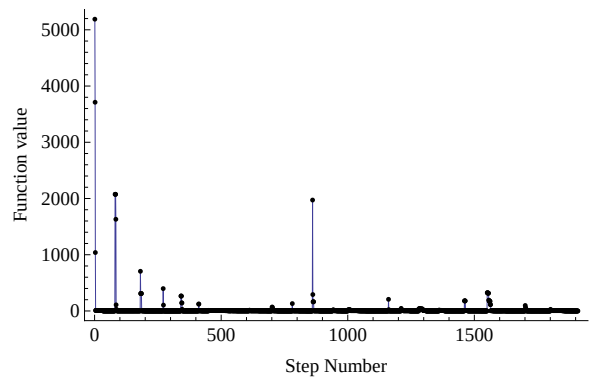
make any predictions about the effect of the dimensionality reduction on these algorithms. We note that all results were verified to minimize f. Table 1 summarizes these results[5]:

Algorithm	Steps (f)	Steps (f^*)
Nelder-Mead	109	82
Differential Evolution	84	105
Simulated Annealing	1654	224
Random Search	359	63

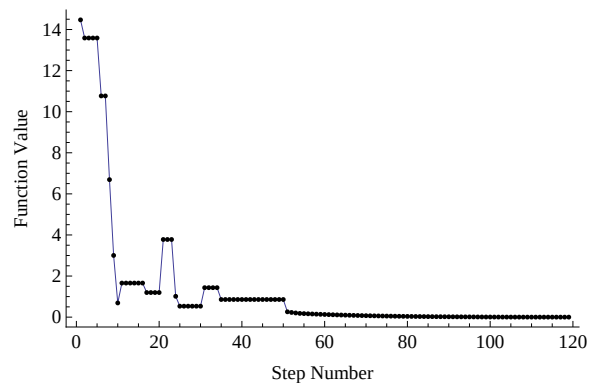
Table 1: Number of function evaluations required by each algorithm to find a minimum

Figure 1 shows a comparison of the path taken by the optimization for simulated annealing. Note the difference in scale between figures 1(a) and 1(b); with the subspace constrained function we do not see the large jumps occurring as the method explores the input space that we do when optimizing the objective function directly.

Figure 2 shows a comparison of the number of function evaluations required by the simulated annealing routine to find a minimum when searching the full input space versus searching the constrained space as the number of input variables is increased. We observe that the number of evaluations required for the full space search increases rapidly while the required evaluations for the constrained search is more or less constant as the input dimension increases.



(a) Without ESM



(b) With ESM

Figure 1: Comparison of simulated annealing with and without subspace surrogate

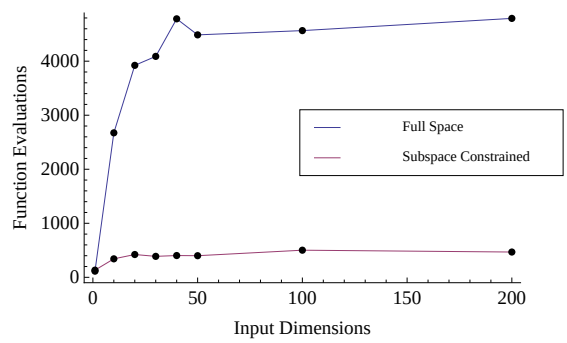


Figure 2: Comparison of constrained versus unconstrained search for simulated annealing

[5]Note that these methods are random in nature and give a varying number of steps; results shown are representative of several runs.

CONCLUSIONS

We can see in table 1 that we achieved an improvement in the number of objective function evaluations required of almost 750%. Using random search we also realized an improvement of approximately 570%. Interestingly, for the other methods we did not see any significant improvement; in fact, differential evolution seems to have actually performed appreciably worse. We prefer not to offer any comment on these methods as we have not undertaken any analysis of these results, though we note that further experiments with Nelder-Mead and differential evolution that are not shown in this exhibited very unusual behavior. The algorithms both gave erratic results for experiments with more than 10 variables, leading us to suspect that they may not be able to cope with so many variables for the particular objective function under study.

The results from these exploratory experiments are very promising, at least for the Metropolis-based methods, and the next logical step would be to test this approach with a realistic physical model. While the objective function in this experiment was constructed expressly to demonstrate the effect of this method, we believe that the same approach will work well with realistic models as well. Simulated annealing showed exceptionally promising results and is common in nuclear engineering, thus pairing ESM with a simulated annealing-based reactor core model optimization problem is clearly the next step.

ACKNOWLEDGEMENTS

This work was supported by a professional development grant funded by the Nuclear Regulatory Commission.

REFERENCES

[1] Bertsekas, D. (1999). Nonlinear programming. Athena Scientific Belmont, MA.

[2] Lewis, R., Torczon, V., and Trosset, M. (2000). Direct search methods: then and now. Journal of Computational and Applied Mathematics, 124(1-2):191207.

[3] *fminsearch*, MATLAB R2011a Documentation, The MathWorks, Inc. 3 Apple Hill Drive Natick, MA 01760-2098. USA.

[4] Olsson, D. and Nelson, L. (1975). The Nelder-Mead simplex procedure for function minimization. Technometrics, 17(1):4551.

[5] W. R. Gilks, S. Richardson and D. J. Spiegelhalter. (1996). Markov chain Monte Carlo in practice. Chapman & Hall/CRC.

[6] Jung, W. and Cho, N. (1993). Determination of design alternatives and performance criteria for safety systems in a nuclear power plant via simulated annealing. Reliability Engineering & System Safety, 41(1):7194.

[7] Lagarias, J. C. et. al. (1998). Convergence properties of the Nelder-Mead simplex algorithm in low-dimensions. Siam Journal of Optimization, 9(1):112-147.

[8] Halko, N., Martinsson, P., and Tropp, J. (2009). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. Arxiv preprint arXiv:0909.4061.

[9] H. Abdel-Khalik. (2011). On nonlinear reduced order modeling. Int. Conf. Math. Comp. Nucl. Sci. Eng., Brazil, (2011).

[10] H. S. Abdel-Khalik and J. M. Hite, Reduced order modeling: tensor-free expansion for nonlinear features identification. (2011). Trans. Am. Nucl. Soc., Hollywood FL, June, 104, 2011

[11] H. Rosenbrock. (1960). An automatic method for finding the greatest or least value of a function. The Computer Journal, 3(3):175.

[12] J. M. Hite and H. S. Abdel-Khalik. (2011). Heuristic Approach for ESM-based Reduced Order Modeling. Trans. Am. Nucl. Soc., Washington DC, Oct, 2011. *Accepted for publication.*