Reduced order modeling of complex systems

JOHN BURKARDT†, QIANG DU‡, MAX GUNZBURGER† & HYUNG-CHUN LEE⋆1

†School for Computational Science and Information Technology, Florida State University, Tallahassee FL 32306-4120, USA
burkardt@csit.fsu.edu

‡Department of Mathematics, Penn State University, University Park PA 16802, USA
qdu@math.psu.edu
gunzburg@csit.fsu.edu

&

⋆Department of Mathematics, Ajou University, Suwon 442-749, Korea
hclee@ajou.ac.kr

1 Reduced-order modeling

Solutions of (nonlinear) complex systems are expensive with respect to both storage and CPU costs. As a result, it is difficult if not impossible to deal with a number of situations such as: continuation or homotopy methods for computing state solutions; parametric studies of state solutions; optimization and control problems (multiple state solutions); and feedback control settings (real-time state solutions). Not surprisingly, a lot of attention has been paid to reducing the costs of the nonlinear state solutions by using reduced-order models for the state; these are low-dimensional approximations to the state. Reduced-order modeling has been and remains a very active research direction in many seemingly disparate fields. We will focus on three approaches to reduced-order modeling: reduced basis methods; proper orthogonal decomposition (POD); and centroidal Voronoi tessellations (CVT). Before describing the three approaches, we first discuss what we exactly mean by reduced-ordering modeling and make some general comments that apply to all reduced-order models. For a state simulation, a reduced-order method would proceed as follows. One first chooses a reduced basis \( \{u_i, i = 1, \ldots, n\} \), where \( n \) is hopefully very small compared to the usual number of functions used in a finite element approximation or the number of grid points used in a finite difference approximation. Next, one seeks an approximation \( \tilde{u} \) to the state of the form \( \tilde{u} = \sum_{i=1}^{n} c_i u_i \in V \equiv \text{span}\{u_1, \ldots, u_n\} \). Then, one determines the coefficients \( c_i, i = 1, \ldots, n \), by solving the state equations in the set \( V \), e.g., one could find a Galerkin solution of the state equations in a standard way, using \( V \) for the space of approximations. The cost of such a computation would be very small if \( n \) is small (ignoring the cost of the off-line determination of the reduced basis \( \{u_1, \ldots, u_n\} \)). In control or optimization settings, one is faced with multiple state solves.
or real-time state solves. If one approximates the state in the reduced, low-dimensional set $V$, then state solutions will be relatively very cheap. In an adjoint or sensitivity equation-based optimization method, one could also employ the adjoint equations for the low-dimensional discrete state equations; thus, if $n$ is small, the cost of each iteration of the optimizer would be very small relative to that using full, high-fidelity state solutions. In a feedback control setting, the approximate state equations in the low-dimensional space could possibly be solved in real time. Does reduced-order modeling work? It is clear that reduced-order methods should work in an interpolatory setting. In a simulation setting, if the state can be approximated well in the reduced basis $V$, then one should expect that things will work well. If the optimal solution and the path to the optimal solution can be well approximated in the reduced basis $V$, then one should expect that things will work well in an optimal control or design setting. If all the states determined by the feedback process can be well approximated in the reduced basis $V$, then again one should expect that things will work well in a feedback control setting. Thus, the reduced basis $V$ should be chosen so that it contains all the features, e.g., the dynamics, of the states encountered during the simulation or the control process. This, of course, requires some intuition about the states to be simulated or about where in parameter space the optimal set of parameters are located.

What happens in an extrapolatory setting is not so clear. Most reduced-order control computations have been done in an interpolatory regime. It is obvious that if the reduced set $V$ does not contain a good approximation to the solution one is trying to obtain, then one cannot hope to successfully determine that solution.

1.1 Common features shared by reduced-order methods

All reduced bases require the solution of high-fidelity and therefore very expensive discrete state and/or sensitivity equations and/or adjoint equations. The idea is that these expensive calculations can be done offline before a state simulation or the optimization of the design parameters or feedback control is attempted. Moreover, one hopes that a single reduced basis can be used for several state simulations or in several design or control settings.

All reduced-basis sets are global in nature, i.e., the support of the basis functions is global. Therefore, solving the state or sensitivity or adjoint equations with respect to any of the reduced bases requires the solution of dense linear and nonlinear systems. Thus, unless the dimension of a reduced basis is “small,” it cannot be used without some further processing. Unfortunately, in order to obtain meaningful approximations, it is often the case that the use of reduced bases requires the use of a relatively large number of basis functions. However, it is often the case that reduced bases contain “redundant” information in the sense that the dynamics of the state should be well approximated by a set of functions of much lower dimension. The question then arises: how can one extract a reduced basis of smaller dimension that contains all the essential information of a reduced basis of larger dimension? This is where POD and CVT come in and, in this sense, they are reduced-reduced basis methods.

Unfortunately, there is no adequate theoretical foundation for reduced-order methods, even in state simulation settings. However, it is certain that without an inexpensive method for reducing the cost of state computations, it is unlikely that the solution of
three-dimensional optimization and control problems involving complex systems, e.g., the Navier-Stokes system, will become routine anytime soon. Thus, it is also certainly true that these methods deserve more study from the computational and theoretical points of view.

2 Reduced-basis methods

All reduced-order methods are reduced basis methods. However, there is a class of methods that use Lagrange bases, Hermite bases, Taylor bases, and snapshot bases (or more precisely, snapshot sets) that have come to be known as reduced-basis methods. 

**Lagrange bases** consist of state solutions corresponding to several different values of the parameters (Reynolds number, design parameters, etc.). These solutions are obtained by standard (and expensive) techniques such as finite element or finite volume methods. For example, if one has the design parameters \( \{ \alpha_j \}_{j=1}^J \), one obtains \( n \) approximate state solutions for \( n \) sets of parameter values to form the \( n \)-dimensional Lagrange reduced basis.

**Hermite bases** consist of the state variables and the first derivatives of the state variables with respect to parameters (the sensitivities) determined for different values of the parameters. The state and sensitivity approximations are obtained through standard (and expensive) techniques such as finite element or finite volume methods. Thus, again, if one has the design parameters \( \{ \alpha_j \}_{j=1}^J \), one chooses \( M \) sets of parameter values and then one obtains the corresponding \( M \) approximate state solutions and the corresponding \( M(J + 1) \) sensitivity derivative approximations. The \( n = M(J + 1) \) state and sensitivity approximations form the Hermite reduced basis of dimension \( n \). 

**Taylor bases** consist of the state and derivatives of the state with respect to parameters (sensitivities and higher-order sensitivities) determined for a fixed set of design parameters. The state and derivative approximations are obtained through standard (and expensive) techniques such as finite element or finite volume methods. The Taylor basis may be somewhat complicated to program due to the complexity of the partial differential equations that determine the higher-order sensitivities. In addition, the number of higher-order derivatives grows very rapidly with the number of design parameters, e.g., if one has 10 design parameters, there are 55 different second derivative sensitivities. Thus, the dimension of the Taylor reduced basis grows quickly with the number of parameters and the number of derivatives used.

See [13, 20, 21] for more details and for examples of the use of reduced-basis methods for simulation and optimization problems.

2.1 Snapshot sets

The state of a complex system is determined by parameters that appear in the specification of a mathematical model for the system. Of course, the state of a complex system also depends on the independent variables appearing in the model. **Snapshot sets** consist of state solutions corresponding to several parameter values and/or evaluated at several values of one or more of the dependent variables, e.g., steady-state solutions corresponding to several sets of design parameters or a time-dependent state solution for a fixed set of design parameter values evaluated at several time instants during the evolution process or several state solutions corresponding to different sets of parameter values evaluated at
several time instants during the evolution process. Snapshot sets are often determined by solving the full, very large-dimensional discretized system obtained by, e.g., a finite volume or finite element discretization. Experimental data have also been used to determine a snapshot set. Snapshot sets often contain “redundant” information; therefore, snapshot sets must usually be post-processed to remove as much of the redundancy as possible before they can be used for reduced-order modeling. POD and CVT may be viewed as simply different ways to post-process snapshot sets.

Since snapshot sets are the underpinning for POD and CVT, we briefly discuss how they are generated in practice. At this time, the generation of snapshot sets is an art and not a science; in fact, it is a rather primitive art. The generation of snapshot sets is an exercise in the design of experiments, e.g., for stationary systems, how does one choose the sets of parameters at which the state (and sensitivities) are to be calculated (using expensive, high-fidelity computations) in order to generate the snapshot set? Clearly, some a priori knowledge about the types of states to be simulated or optimized using the reduced-order model is very useful in this regard. The large body of statistics literature on the design of experiments has not been used in a systematic manner.

For time-dependent systems, many (ad hoc) measures have been invoked in the hope that they will lead to good snapshot sets. Time-dependent parameters (e.g., in boundary conditions) are used to generate states that are “rich” in transients, even if the state of interest depends only on time-independent parameters. In order to generate even “richer” dynamics, impulsive forcing is commonly used, e.g., starting the evolution impulsively with different strength impulses and introducing impulses in the middle of a simulation. In the future, a great deal of effort needs to be directed towards developing and justifying methodologies for generating good snapshot sets. After all, a POD or CVT basis is only as good as the snapshot set used to generate it.

3 Proper orthogonal decompositions (POD)

Given $n$ snapshots $\tilde{x}_j \in \mathbb{R}^N, j = 1, \ldots, n$, set

$$x_j = \tilde{x}_j - \tilde{\mu}, \quad j = 1 \ldots, n,$$

where $\tilde{\mu} = \frac{1}{n} \sum_{j=1}^{n} \tilde{x}_j$.

The set $\{x_j\}_{j=1}^n$ are the modified snapshots. Let $d \leq n$. Then, the POD basis $\{\phi_i\}_{i=1}^d$ of cardinality $d$ is found by successively solving, for $i = 1, \ldots, d$, the problem

$$\lambda_i = \max_{|\phi_i|=1} \frac{1}{n} \sum_{j=1}^{n} |\phi_i^T x_j|^2 \quad \text{and} \quad \phi_i^T \phi_\ell = 0 \quad \text{for } \ell \leq i - 1.$$

If $n \geq N$, this decomposition is known as the direct method. If $n < N$, it is known as the snapshot method; we will only consider the latter case.

Let $A$ denote the $N \times n$ snapshot matrix whose columns are the modified snapshots $x_j$, i.e.,

$$A = (x_1, x_2, \ldots, x_n) = (\tilde{x}_1 - \tilde{\mu}, \tilde{x}_2 - \tilde{\mu}, \ldots, \tilde{x}_n - \tilde{\mu}).$$
Let $K$ denote the $n \times n$ (normalized) correlation matrix for the modified snapshots, i.e.,

$$K_{ij} = \frac{1}{n} x_j^T x_i \quad \text{or} \quad K = \frac{1}{n} A^T A.$$ 

Let $\chi_i$ with $|\chi_i| = 1$ denote the eigenvector corresponding to the $i$-th largest eigenvalue $\lambda_i$ of $K$. Then, the POD basis is given by $\phi_i = \frac{1}{\sqrt{n \lambda_i}} A \chi_i$. The POD basis is orthonormal, i.e., $\phi_i^T \phi_j = 0$ for $i \neq j$ and $\phi_i^T \phi_i = 1$.

POD is closely related to the statistical methods known as Karhunen-Loève analysis or the method of empirical orthogonal eigenfunctions or principal component analysis. POD is also closely related to the singular value decomposition (SVD) of the modified snapshot matrix $A$. Let $A = U \Sigma V^T$ denote the SVD of $A$; then, $\sigma_i^2 = n \lambda_i$ for $i = 1, \ldots, n$, where $\sigma_i$ is the $i$-th singular value of $A$ and $\lambda_i$ is the $i$-th largest eigenvalue of $K = \frac{1}{n} A^T A$. The POD basis vectors are the first $n$ left singular vectors of the snapshot matrix $A$, i.e., $\phi_i = u_i$ for $i = 1, \ldots, n$.

The POD basis is optimal in the following sense. Let $\{\psi_j\}_{i=1}^n$ denote an arbitrary orthonormal basis for the span of the modified snapshot set $\{x_j\}_{j=1}^n$. Let $P_{\psi,d} x_j$ denote the projection of the modified snapshot $x_j$ onto the $d$-dimensional subspace spanned by $\{\psi_i\}_{i=1}^d$. Clearly we have, for each $j = 1, \ldots, n$,

$$P_{\psi,d} x_j = \sum_{i=1}^d c_{ij} \psi_i \quad \text{where} \quad c_{ij} = \psi_i^T x_j \quad \text{for} \quad i = 1, \ldots, d.$$ 

Let the error be defined by

$$E = \sum_{j=1}^n |x_j - P_{\psi,d} x_j|^2.$$ 

Then, the minimum error is obtained when $\psi_i = \phi_i$ for $i = 1, \ldots, d$, i.e., when the $\psi_i$'s are the POD basis vectors. The connection between POD and SVD makes it easy to show that the error of the $d$-dimensional POD subspace is given by

$$E_{\text{pod}} = \sum_{j=d+1}^n \sigma_j^2 = n \sum_{j=d+1}^n \lambda_j \quad \text{where} \quad n = \text{number of snapshots}$$

$$d = \text{dimension of the POD subspace}.$$ 

If one wishes for the relative error to be less than a prescribed tolerance $\delta$, i.e., if one wants $E_{\text{pod}} \leq \delta \sum_{j=1}^n |x_j|^2$, one should choose $d$ to be the smallest integer such that

$$\sum_{j=1}^d \sigma_j^2 / \sum_{j=1}^n \sigma_j^2 = \sum_{j=1}^d \lambda_j / \sum_{j=1}^n \lambda_j \geq \gamma = 1 - \delta.$$ 

There have also been several variations introduced in attempts to “improve” POD. For details on POD and its variants and its use in flow simulation and control problems, see, e.g., [1, 2, 3, 5, 6, 10, 11, 12, 15, 16, 17, 18, 19, 22, 23, 24, 25, 26, 27, 28].

### 4 Centroidal Voronoi tessellations (CVT)

Given a discrete set of modified snapshots $W = \{x_j\}_{j=1}^n$ belonging to $\mathbb{R}^N$, a set $\{V_i\}_{i=1}^k$ is a tessellation of $W$ if $\{V_i\}_{i=1}^k$ is a subdivision of $W$ into disjoint, covering subsets, i.e.,
\(V_i \subset W\) for \(i = 1, \ldots, k\), \(V_i \cap V_j = \emptyset\) for \(i \neq j\), and \(\bigcup_{i=1}^{k} V_i = W\). Given a set of points \(\{z_i\}_{i=1}^{k}\) belonging to \(\mathbb{R}^N\) (but not necessarily to \(W\)) the Voronoi region corresponding to the point \(z_i\) is defined by

\[
\hat{V}_i = \{ x \in W : |x - z_i| \leq |x - z_j| \quad \text{for} \quad j = 1, \ldots, k, \ j \neq i \},
\]

where equality holds only for \(i < j\). The set \(\{\hat{V}_i\}_{i=1}^{k}\) is called a Voronoi tessellation or Voronoi diagram of \(W\) corresponding to the set of points \(\{z_i\}_{i=1}^{k}\). The points in the set \(\{z_i\}_{i=1}^{k}\) are called the generators of the Voronoi diagram \(\{\hat{V}_i\}_{i=1}^{k}\) of \(W\). Given a density function \(\rho(y) \geq 0\), defined for \(y \in W\), the mass centroid \(z^*\) of any subset \(V \subset W\) is defined by

\[
\sum_{y \in V} \rho(y)|y - z^*|^2 = \inf_{z \in V^*} \sum_{y \in V} \rho(y)|y - z|^2,
\]

where the sums extend over the points belonging to \(V\). The set \(V^*\) can be taken to be \(V\) or it can be an even larger set such as all of \(\mathbb{R}^N\). In case \(V^* = \mathbb{R}^N\), \(z^*\) is the ordinary mean

\[
z^* = \frac{\sum_{y \in V} \rho(y)y}{\sum_{y \in V} \rho(y)}.
\]

In this case, \(z^* \not\in W\) in general.

If \(z_i = z_i^*\) for \(i = 1, \ldots, k\), where

\[
\{z_i\}_{i=1}^{k}
\]

is the set of generating points of the Voronoi tessellation \(\{\hat{V}_i\}_{i=1}^{k}\)

\[
\{z_i^*\}_{i=1}^{k}
\]

is the set of mass centroids of the Voronoi regions \(\{\hat{V}_i\}_{i=1}^{k}\),

we refer to the Voronoi tessellation as being a Centroidal Voronoi tessellation (CVT). The concept of CVT’s can be extended to more general sets, including regions in Euclidean space, and to more general metrics. CVT’s are useful in a variety of applications, including optimal quadrature rules, covoolume and finite difference methods for PDE’s, optimal representation, quantization, and clustering, cell division, data compression, optimal distribution of resources, territorial behavior of animals, optimal placement of sensors and actuators, grid generation in 2D, 3D, and on surfaces, mesh free methods, clustering of gene expression data, image segmentation. See, e.g., [7] for details.

CVT’s are optimal in the following sense. Given the discrete set of points \(W = \{x_j\}_{j=1}^{n}\) belonging to \(\mathbb{R}^N\), we define the error of a tessellation \(\{V_i\}_{i=1}^{k}\) of \(W\) and a set of points \(\{z_i\}_{i=1}^{k}\) belonging to \(\mathbb{R}^N\) by

\[
\mathcal{F}((z_i, V_i), i = 1, \ldots, k) = \sum_{i=1}^{k} \sum_{y \in V_i} \rho(y)|y - z_i|^2.
\]

Then, it can be shown that a necessary condition for the error \(\mathcal{F}\) to be minimized is that the pair \(\{z_i, V_i\}_{i=1}^{k}\) form a CVT of \(W\).

CVT’s of discrete sets are closely related to optimal \(k\)-means clusters so that Voronoi regions and centroids can be referred to as clusters and cluster centers, respectively. The error \(\mathcal{F}\) is also often referred to as the variance, cost, distortion error, or mean square error.
There are several algorithms known for constructing Centroidal Voronoi tessellations of a given set. Lloyd’s method is a deterministic algorithm which is the obvious iteration between computing Voronoi diagrams and mass centroids, i.e., a given set of generators is replaced in an iterative process by the mass centroids of the Voronoi regions corresponding to those generators. MacQueen’s method is a very elegant probabilistic algorithm which divides sampling points into \( k \) sets or clusters by taking means of clusters. We have developed a new probabilistic method which may be viewed as a generalization of both the MacQueen and Lloyd methods and is amenable to efficient parallelization. See [14] for details.

### 4.1 CVT’s and model reduction

CVT’s have been successfully used in data compression; one particular application was to image reconstruction. Therefore, it is natural to examine CVT’s in another data compression setting, namely reduced-order modeling. The idea, just as it is in the Pod setting, is to extract, from a given set of (modified) snapshots \( \{x_j\}_{j=1}^n \) of vectors in \( \mathbb{R}^N \), a smaller set of vectors also belonging to \( \mathbb{R}^N \). In the POD setting, the reduced set of vectors was the \( d \)-dimensional set of POD vectors \( \{\phi_j\}_{j=1}^d \). In the CVT setting, the reduced set of vectors is the \( k \)-dimensional set of vectors \( \{z_k\}_{k=1}^k \) that are the generators of a centroidal Voronoi tessellation of the set of modified snapshots. See [4, 8, 9] for details. Just as POD produced an optimal reduced basis in the sense that the error \( \mathcal{E} \) is minimized, CVT produces an optimal reduced basis in the sense that the error \( \mathcal{F} \) is minimized. One can, in principle, determine the dimension \( d \) of an effective POD basis, e.g., using the eigenvalues of the correlation matrix. Similarly, one can, in principle, determine the dimension \( k \) of an effective CVT basis by examining the (computable) error \( \mathcal{F}(\cdot) \).

A natural question is: why should one use CVT instead of POD? Although justifications have to be substantiated through analyses and extensive numerical experiments, heuristically, one can make some arguments. CVT naturally introduces the concept of clustering into the construction of the reduced basis. CVT is “cheaper” than POD; POD involves the solution of an \( n \times n \) eigenproblem, where \( n \) is the number of snapshots; CVT requires no eigenproblem solution. CVT can handle many more snapshots. Adaptively changing the reduced basis is much less expensive with CVT.

Actually, one does not have to choose between POD and CVT. They may in fact be combined to define hybrid methods which take advantage of the best features of both methods. See [9].

### 4.2 Computational experiments using CVTs for model reduction

Several computational examples of the use of CVTs for model reduction in flow simulation problems are given in [4]. Here, we provide some of the results given in that paper. The setting is the Navier-Stokes system in a box with inflow at the lower left-hand corner and outflow at upper right-hand corner. Snapshots (500 of them) were generated by sampling a finite element approximation at 500 evenly-spaced times in the interval \([0, 5]\). The inflow velocity was impulsively changed at \( t = 0 \) and \( t = 2.5 \). CVT reduced bases were determined from the 500 (modified) snapshots. See [4] for details.
We tested the accuracy resulting from the use of CVT reduced basis for flow simulations with the two inflow conditions depicted in Figure 1. Figure 2 provides plots of the $L^2$ error vs. time for the CVT reduced-basis solutions for three different cardinalities for the bases. The results of Figure 2 indicate that, at least for this example, CVT-based model reduction is effective. Of course, much more computational experimentation, including comparison with POD-based results, is necessary in order to document the efficacy of CVT-based reduced-order modeling. These experiments are the focus of our current efforts in CVT-based model reduction.

![Figure 1: Two velocity boundary conditions used in testing CVT bases.](image1)

![Figure 2: $L^2$ errors vs. time in CVT reduced-basis solution vs. time; left: case 1; right: case 2.](image2)

References


