Methods for Simulation-based Analysis of Fluid-Structure Interaction

Matthew F. Barone and Jeffrey L. Payne

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Approved for public release; further dissemination unlimited.
Methods for Simulation-based Analysis of Fluid-Structure Interaction

Matthew F. Barone and Jeffrey L. Payne
Aerosciences and Compressible Fluid Mechanics Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0825
mbarone@sandia.gov

Abstract

Methods for analysis of fluid-structure interaction using high fidelity simulations are critically reviewed. First, a literature review of modern numerical techniques for simulation of aeroelastic phenomena is presented. The review focuses on methods contained within the arbitrary Lagrangian-Eulerian (ALE) framework for coupling computational fluid dynamics codes to computational structural mechanics codes. The review treats mesh movement algorithms, the role of the geometric conservation law, time advancement schemes, wetted surface interface strategies, and some representative applications. The complexity and computational expense of coupled Navier-Stokes/structural dynamics simulations points to the need for reduced order modeling to facilitate parametric analysis. The proper orthogonal decomposition (POD)/Galerkin projection approach for building a reduced order model (ROM) is presented, along with ideas for extension of the methodology to allow construction of ROMs based on data generated from ALE simulations.
Contents

1 Introduction .............................................................................................................. 5

2 Literature Review of Coupled Fluid-Structure Simulation Methods ............... 6
   2.1 Overview ............................................................................................................. 6
   2.2 Mesh Movement ............................................................................................... 7
   2.3 Geometric Conservation Law ......................................................................... 10
   2.4 Time Advancement Schemes ........................................................................ 11
   2.5 Interface Strategies ....................................................................................... 14
   2.6 Applications ................................................................................................... 16

3 The Reduced Order Modeling Approach ......................................................... 17
   3.1 The POD/Galerkin Approach ....................................................................... 17
   3.2 POD/Galerkin for Compressible Fluid Mechanics ..................................... 22
   3.3 Coupled Fluid/Structure ROMs .................................................................... 28
   3.4 Sources of Error in Reduced Order Modeling ........................................... 29

4 Plans for Future Work .......................................................................................... 31

References .............................................................................................................. 32

Figures

1 Results of a bibliographic database search on “fluid structure interaction” AND (“numerical” OR “simulation” OR “computation”). Paper counts are scaled by the total number of papers in the database relative to the period 1995-2004. .................................................................................................................. 6

2 Number of POD modes required to describe 95% of the averaged fluctuation energy (solid symbols), and number of flow simulation mesh points (open symbols) for selected fluid flow simulations. .............................. 21
Aerodynamic loading of flexible structures is an important aspect of many engineering disciplines, including aeroelasticity of flight vehicles, and wind turbine aerodynamics. Advances in computational mechanics algorithms, along with continually growing computational resources, are beginning to make feasible the analysis of coupled fluid/structure problems using high-fidelity computer simulation. Simulation of fluid-structure interaction has been a rapidly growing area in the computational sciences arena, as indicated by the results of a recent bibliographic database search shown in Figure 1. Nonetheless, development of the computational building blocks necessary to perform these simulations, and how they are pieced together to enable a coupled physics simulation, are formidable problems. Further, the use of such simulation tools for real engineering systems with complicated geometries cannot be considered routine. Maximum benefit will be derived from the simulations if strategies for their efficient employment are mapped out beforehand.

This document contains two parts aimed at addressing each of these issues. The first part is a focused survey of the current state-of-the-art in coupled fluid/structure simulation capabilities. The primary applications of interest are static and dynamic aeroelastic phenomena for flight vehicles and wind turbine blades at high Reynolds number. The review is, therefore, biased towards methods suitable for simulation of such phenomena. It also ignores much of the important research carried out on aeroelasticity before numerical simulation with the Navier-Stokes equations was even an option. This prior research formed a basis of understanding that is essential for proper interpretation of the results from the high-fidelity simulation methods surveyed here.

The second part of this report addresses reduced order modeling as an enabling technology for high-fidelity simulations. Generally speaking, a reduced order model (ROM) is a relatively inexpensive mathematical model of a physical system that is derived from experimental data or simulation data. In the present context, the ROM is a system of ordinary differential equations, derived from simulation data and the governing partial differential equations (PDEs) of a fluid/structure system, that has many fewer degrees of freedom (DOF) than the discretized PDEs. The goal of the ROM is to accurately reproduce behavior of the original system of PDEs over a range of system input parameters at a small fraction of the cost of solving the PDEs. This section is partly a survey of existing ROM technology and partly an exposition on new ideas and possible research directions for reduced order modeling of fluid/structure systems.
Figure 1. Results of a bibliographic database search on “fluid structure interaction” AND (“numerical” OR “simulation” OR “computation”). Paper counts are scaled by the total number of papers in the database relative to the period 1995-2004.

2 Literature Review of Coupled Fluid-Structure Simulation Methods

2.1 Overview

This review is designed to inform the reader on the existing numerical methods available for simulation of aeroelastic systems. Most of the methods are based on the Arbitrary Lagrangian-Eulerian (ALE) formulation of the equations for the fluid motion [23], and a Lagrangian description of the motion of the structure. The ALE formulation provides for solution of the fluid problem on a moving mesh by recasting the equations in a reference frame relative to the mesh motion. In a finite volume scheme, this means accounting for the changing cell volumes and additional flux terms due to the mesh motion. An alternative to the ALE formulation is to keep the grid fixed in an Eulerian reference frame and only move the fluid-structure interface. This requires a means of applying the interface boundary conditions for arbitrary interface positions, such as the immersed boundary method [59, 53]. The interface itself is tracked by an appropriate technique, such as the level set method [50]. Some adaptive grid capability is needed in conjunction with the im-
mersed boundary method in order to maintain tightly spaced grid points within the boundary layer. Immersed boundary methods for computational aeroelasticity problems are not as mature as ALE-based methods, and are not discussed in this review.

A wide variety of structural models may be used, varying in complexity from simple modal descriptions to highly detailed finite element models. The level of fidelity depends on the problem at hand, namely the geometric complexity, and the degree of linearity of the structural response. The aerodynamic loads may be described by either a linear (e.g. incompressible potential flow) or nonlinear (e.g. the Euler equations) model. Linear aerodynamic models and models based on the transonic small disturbance equations currently comprise most of the workhorse methods in industry-based analyses of aeroelasticity. This review focuses on the emerging technology of coupling fully nonlinear models for the fluid flow (Euler and Navier-Stokes equations) with a structural model. The nonlinear behavior of fluid motion, particularly for compressible flows in the transonic regime, is an essential aspect of many aeroelastic phenomena of interest. Nonlinear aeroelastic analysis can be used to tackle problems such as transonic flutter, limit cycle oscillations (LCO), and aeroservoelastic problems involving control surface deflections.

The remainder of this review is comprised of sections treating the major technical aspects of fluid-structure simulations in the ALE framework: mesh movement strategies, geometric conservation laws, time advancement schemes, and fluid-structure interface strategies. The final section gives an overview of a few recent applications in computational aeroelasticity that represent the current state-of-the-art.

### 2.2 Mesh Movement

The ALE formulation requires a scheme for moving fluid mesh points as the domain boundaries translate, rotate, and deform. There are currently three established categories of mesh movement strategies: 1) the spring analogy, 2) the elastic medium analogy, and 3) remeshing. This section reviews each of the three categories, citing references to important work in development of each of the methods. At the end of the section, an attempt is made to assess the different methods in terms of their cost, complexity, robustness, generality, and parallelizability.

In the spring analogy method, the mesh is viewed as a structural system described by some combination of lineal and/or torsional springs. Batina [8] applied a lineal spring method to solution of the Euler equations for a rigid pitching airfoil. In the lineal spring method, the element or control volume edges are treated as springs with stiffness inversely proportional to the edge length. The static equilibrium equation solves for the grid point displacements whenever the mesh needs to be moved. The lineal spring analogy prevents edges from collapsing, but does not prevent collapse of elements or grid line crossing. To remedy this situation Farhat et al. [26] added torsional springs (in addition to the lineal springs) to the grid nodes, demonstrating improved robustness over the lineal spring method on several two-dimensional model problems. Murayama et al. [56] extended the
torsional spring scheme to three dimensions and modified the implementation in the interest of computational efficiency. To maintain robustness, however, Murayama et al. added heuristic near-surface functions that increase the spring stiffness to infinity near a solid surface. Bartels [7] modified the lineal spring analogy to prevent the case of grid collapse around convex surfaces, demonstrating the improved method on the simulation of spoiler actuation in two dimensions. Bartel’s improved method uses transfinite interpolation as an initial algebraic mesh initialization step, followed by the spring-based smoothing operation. It is, therefore, limited in scope to structured meshes. Löhner and Yang [45] proposed a method related to the spring analogy that is based on a Laplacian smoothing with variable diffusivity based on distance from the surface.

The elastic medium analogy can be viewed as a refinement of the spring analogy. Now the grid is viewed as embedded in a continuous elasto-static medium. The compressive and shear moduli of the medium, along with the numerical method used to solve the elasto-static equilibrium equations, define the grid deformation scheme. Johnson and Tezduyar [40] applied a finite element method to solve the linear elastic equilibrium equations for a finite element fluid mesh. Deformation of small (near body) mesh elements is kept small by scaling the allowable mesh deformation by the element volume. Some examples of a rigid, oscillating 2D airfoil using a hybrid mesh are given. The chord Reynolds number of 1000 is relatively low for aeroelastic computations. For problems with large mesh motions the authors admit a need for re-meshing when the grid quality becomes too poor. Essentially the same dynamic mesh procedure has been applied to much more complicated fluid-structure interaction problems, e.g. the simulation of flow past a parachute and the simulation of large numbers of spheres falling through a viscous medium [41]. For such complex problems the linear elastic method only goes so far before some type of re-meshing is required.

Chen and Hill [19] recognized that the linear elastic static equilibrium equations could be solved more efficiently using a linear boundary element method (BEM). The boundary element method actually provides for both the mesh movement for general grid types and gives the interpolation matrix for the fluid/structure interface. The mesh movement capability is demonstrated on a 3D wing with ten degree twist deformation at the wing tip.

The linear elastic analogy was applied in the context of aerodynamic shape optimization by Nielsen and Anderson [57]. In their work the mesh material stiffness is proportional to the cell aspect ratio so that the near-body mesh does not deform as much as cells further away from the body. The numerical method for solving the pseudo-structural grid system is not described. Improvement in mesh quality and robustness of the movement is demonstrated for several two-dimensional airfoil problems. A three-dimensional problem is presented for deformation of a wing; this example required ten incremental steps to accomplish the mesh motion.

The linear elastic pseudo-structural analogy fails to give valid meshes when the mesh deformations become too large. The method of Bar-Yoseph et al. [6] proposes to solve this issue by making the material properties nonlinear, implying a nonlinear stress-strain relation. In their work, the material properties of the grid medium are functions of the local element quality. The element quality is a scalar quantity that gives some measure
of mesh distortion. The resulting pseudo-structural equations are solved using a finite element method. Examples of mesh movement are given, but the grids appear coarse and the demonstration is unconvincing. Gao et al. [32] also apply a nonlinear structural medium analogy, but solve the resulting equations using a nonlinear boundary element method. The parameters that define the nonlinear stress-strain relation are found by solving an optimization problem. Robustness and grid quality are demonstrated for a 20 degree airfoil pitch problem that was solved incrementally in 5 degree steps. Measures of computational cost are not quoted, but the authors’ comments suggest that the method may be expensive for dynamic mesh applications. Solution of the nonlinear BEM problem also appears to be significantly more complex than solution of the linear BEM problem [19].

Re-meshing of the volume mesh is an alternative to the spring and structural analogy methods. Re-meshing techniques generate a new mesh each time the boundary moves based on the prescribed boundary motion and the geometry of the problem, rather than on any type of structural analogy. Morton et al. [54] applied an algebraic re-meshing scheme for structured, overset meshes that maintains grid line orthogonality near a surface and also maintains grid point position in the grid overlap regions. This method was demonstrated on an elastically mounted, freely vibrating circular cylinder problem. Melville [51] developed a mesh movement scheme based on the proximity of a grid point to nearby surfaces. The surface point movement influences the interior grid point directly, without reference to the underlying grid topology or connectivity. The method relies on several heuristics for determining the influence of the surface movements on the interior mesh point movement. It was successfully demonstrated on Euler [51] and Navier-Stokes [52] aeroelastic simulations of an F-16 fighter plane. Baker [5] used a linear elastic structural analogy to move the mesh, then applied a mesh coarsening/enriching procedure to maintain grid quality. Martineau and Georgala [49] have developed a two-step mesh movement algorithm that first initializes the mesh based on a rigid-body motion procedure, then applies a smoothing operation. This procedure seems cumbersome due to the fact that the distance to the two nearest surfaces must be tracked for each mesh point or element. Although re-meshing may seem inelegant and overly complex at times, it is important to keep in mind that strategies relying solely on spring or structural medium analogies are usually only demonstrated on relatively simple geometries. Complex geometry and/or large deformations often require at least periodic re-meshing. This is especially the case when thin boundary layers are resolved with very tightly spaced grid cells near the fluid-structure boundary.

Some comparisons of the robustness of mesh movement strategies have been published [14, 26, 32, 73], although they are far from comprehensive. The lineal spring analogy is generally discredited on all but inviscid problems with relatively small mesh movements. The torsion spring improvement leads to enhanced robustness, but will still generate meshes of poor quality for large-enough deflections. The linear elastic medium approach generally shows a dramatic improvement over lineal spring analogies and a less pronounced, but measurable, improvement over the torsional/lineal spring method. Addition of variable linear elastic material properties further improves robustness for large rigid body motions. Nonlinear modeling of the grid pseudo-structure can allow for larger mesh deflections but this is gained at an unspecified increase in computational cost.
It is difficult to compare the cost of the different mesh motion strategies, mainly due to the lack of published timings in the literature. It can be reasonably assumed that the cost increases with increasing robustness and effectiveness of the method. One potentially significant cost savings could be realized by solution of the pseudo-structural equilibrium equations by a boundary element method rather than a finite element or other volumetric approach.

Most of the mesh movement methods may be applied to structured or unstructured element types. The exceptions are the methods of [7] and [54], which apply only to structured grids. Ease of parallelization of the methods is difficult to assess. The spring methods require the solution of a large matrix system; parallel methods for such a system are readily available. A method that solves the elasto-static equilibrium or similar equations needs a parallel finite element algorithm or something similar. The BEM methods are, in theory, easily parallelized. They have the advantage of not requiring grid point connectivity information since the grid displacements depend only on a surface integral evaluation. Global parallel re-meshing, particularly using those methods that require tracking of proximity to nearby surfaces, poses formidable algorithmic challenges.

2.3 Geometric Conservation Law

The geometric conservation law (GCL) states that a method for solving a fluid flow on a moving and/or deforming mesh exactly preserves a uniform flow. The corresponding mathematical expression for a cell volume $\Delta V_i$ with cell boundaries moving at grid velocity $w$ is

$$\frac{d}{dt} \int_{\Delta V_i} dV = \int_{\partial V_i} w \cdot dS.$$  \hspace{1cm} (1)

The terminology and idea of geometric conservation was originally due to Thomas and Lombard [70], who used it to construct improved finite difference schemes on moving meshes. Equation (1) may be derived from the flow conservation equations, assuming a uniform flow. In order to satisfy the GCL in a discrete sense, the numerical approximations of the left and right hand sides of (1) should balance exactly.

Although violation of the GCL clearly introduces an additional error related to mesh-movement, it is not immediately clear whether this error should be of concern. A review of the literature on this topic suggests that the treatment or disposal of the GCL can have an impact on accuracy in certain cases, and has a stronger connection to numerical stability of a time advancement scheme. Only a weak theoretical connection has been made concerning accuracy. Guillard and Farhat [33] proved that for a given scheme that is $p$-order time-accurate on a fixed mesh, satisfying the corresponding $p$-discrete geometric conservation law is a sufficient condition for the scheme to be at least first-order time-accurate on a moving mesh. Enhancement of accuracy has been practically demonstrated on test problems, e.g. in [43].

Perhaps more important than accuracy is the relation of the GCL to numerical stability.
Stability analysis of time marching schemes that obey and violate the GCL has shown that, in general, satisfying the GCL is not a necessary or sufficient condition for a time advancement scheme to be stable [31, 11]. These studies used a finite element discretization of a linear advection diffusion model problem, and so considered only the linear stability of the schemes. It was, however, shown in both of these studies that satisfying a first order GCL is a sufficient condition for unconditional stability of the backward Euler implicit scheme. Farhat et al. [28] investigated the nonlinear stability of a nonlinear scalar hyperbolic conservation law in multiple dimensions. The nonlinear stability criterion is based on the maximum principle, which should be satisfied for a monotone flux scheme that is time-integrated in a stable fashion. Satisfying the GCL is a necessary and sufficient condition for nonlinear stability for all the schemes examined, which included an explicit scheme, a first order implicit scheme, and a second order implicit scheme. The stability results are verified for a uniform flow, a 1-D shock tube problem, and the aeroelastic response of a fighter plane configuration. Because of this demonstrated connection to numerical stability (and accuracy), Farhat and co-workers have advocated time-advancement schemes for coupled fluid-structure problems that obey the GCL [43].

### 2.4 Time Advancement Schemes

There are three classes of methods for advancing a time-accurate fluid/structure simulation forward in time: the monolithic approach, the fully coupled approach, and the loosely coupled approach.

In the monolithic approach for aeroelasticity problems [9, 54], the fluid and structure equations of motions are viewed as a single equation set and solved using a unified solver. From the computer code’s point of view, a structural element is differentiated from a fluid element or control volume only by the difference in variables and spatial representation scheme for each type of element. The primary advantage of a monolithic approach is that fully consistent coupling is preserved; that is, the fluid and structure are perfectly synchronized while advancing a single time step. This usually leads to enhanced robustness, stability, and larger allowable time steps.

The fully coupled approach also synchronizes the fluid and structure systems at each time step, but does so using a partitioned scheme. In a partitioned scheme, the fluid and structure code modules are separate, with fluid loads and structural displacements transferred back and forth within a single time step. The solvers for the fluid and structure systems are entirely separate and may be constructed for efficiency in each case. In the fully coupled approach, subiterations are performed until the entire system is fully converged. The fully coupled approach retains the synchronicity property of the monolithic scheme but also has the advantages of a partitioned scheme, namely improved code maintainability and algorithmic flexibility for physically disparate systems.

The loosely coupled approach is similar to the fully coupled approach because it, too, is a partitioned method. However, the fluid/structure system is not sub-iterated to full con-
vergence at each time step. Instead, the fluid and structure system exchange data one, or
maybe two, times within a time step. The fluid and structure solution updates are lagged,
or staggered, resulting in lower computational cost per time step than a fully coupled ap-
proach. The two systems are never fully in phase, and this introduces a temporal error
in addition to the truncation error of the fluid and structure integration schemes. Care
must be taken to maintain both accuracy and stability when constructing a loosely cou-
pled scheme. Bendiksen [9] argues that a lagged approach not only introduces additional
error, but may also result in a system that is not dynamically equivalent to the physical
system. Unless the time lag is sufficiently small, spurious numerical solutions may exist.
However, loosely coupled approaches have been successfully demonstrated on an array
of aeroelasticity problems and the dynamic equivalence argument does not appear to be of
great practical importance. In addition to the already mentioned advantages of a partitioned
approach, the primary (potential) advantage of a loosely coupled scheme is the relatively
small computational expense per time step.

The remainder of this section discusses only fully and loosely coupled approaches. The
SIERRA programming environment [67], in particular, is set up to support a partitioned ap-
proach. The monolithic scheme, while it has certain advantages mentioned above, presents
the difficulty of integrating fluid and structure solvers into a single solver. It is not a sim-
ple task to start from existing single-physics codes and accomplish this integration (and
maintain the resulting software efficiently), although promising research is underway to
alleviate some of the obstacles to this approach [38].

An example of a fully coupled approach is the work of Alonso and Jameson [1], who
coupled a two-dimensional Euler code with a linear modal pitch/plunge structural model of
an airfoil. The Euler equations were solved with a second order implicit temporal scheme
with multigrid acceleration, while the modal equations were advanced using a separate
solver. Information between the two domains was exchanged at the end of each pseudo-
time iteration and the entire system was fully converged at each physical time step. Weer-
atunga and Pramono [72] used a similar method but with the 3D Euler equations and a
beam/shell finite element model for the structure. Several subiterations were employed
for each physical time step rather than requiring full convergence. Savings in computa-
tional cost of more than a factor of three over a standard loosely coupled approach were
demonstrated, due primarily to an increase in the allowable time step for the fully coupled
approach. Recently, Cebral and Löhner [18] review an underrelaxed predictor/corrector
scheme that iterates on fluid/structure solves, passing underrelaxed fluid loads and struc-
ture displacements and velocities between the solvers. Their approach can be viewed as a
Jacobi iteration for solving the complete fluid-structure system.

Analysis of the stability and accuracy of loosely coupled schemes has been carried out
in a series of papers by Charbel Farhat and co-workers [43, 61, 60, 44, 29]. Energy-norm
stability of various schemes was examined in Piperno et al. [61]. The Conventional Se-
rial Staggered (CSS) scheme involves a structural predictor step, followed by a fluid mesh
movement and fluid solve, followed by a structural update. The Conventional Parallel Stag-
gered (CPS) procedure contains no predictor/corrector steps, so that the fluid and structural
solves are performed in parallel. The CPS scheme has the advantage of facilitating coupling on parallel computers, at the expense of accuracy and numerical stability. Piperno et al. [61] also point out that in aeroelasticity problems, the time step required for accurate resolution of fluid phenomena is often less than that required for the structure. This is because flutter and other dynamic aeroelastic modes usually have a relatively low structural frequency. It is sometimes more economical to subcycle the fluid solver within a single structural time step. Care must be taken in averaging the fluid loads from the subcycle solutions and transferring them to the structure in a way that maintains numerical stability.

Two improved loose coupling procedures, the Improved Serial Staggered (ISS) and Improved Parallel Staggered (IPS) were presented in [44] and [29]. The ISS procedure adds a non-trivial structural displacement predictor step, and is constructed so that the fluid and structure states are one-half time step out of phase. The IPS procedure passes fluid loads to the structural solver at one-half time step, then uses updated structural displacements to advance the fluid system to the next full time step (this is done in parallel with a new structural update). The performance of the improved methods are compared to the CSS and CPS methods for a wing flutter prediction validation case. The improved procedures allow a coupling time step that is 20-46 times larger than previously used time steps without degradation of accuracy. The ISS algorithm allowed a time step several times larger than the IPS method for this problem.

The importance of obeying the discrete form of the geometric conservation law, Equation (1), in maintaining temporal accuracy and stability was investigated in Ref. [43]. The discrete geometric conservation law (DGCL) is satisfied by evaluating fluxes at particular intermediate mesh configurations between time level \( n \) and \( n + 1 \). In general, a scheme that obeys the DGCL is more accurate and stable than one that does not. An energy analysis performed on the full coupled fluid/structure/moving mesh transient problem [60] gives the accuracy of various staggered partitioned schemes in terms of the order of the error in energy conservation. The standard CSS scheme is first order energy conserving, while 2nd and 3rd order CSS constructions are possible. The ISS scheme is shown to be 3rd order energy conserving. Energy conservation is of importance in aeroelastic stability analyses, since phenomena like flutter depend on a precise balance in energy transfer between the fluid and structure.

A completely different approach to computational modeling of dynamic aeroelastic phenomena is the harmonic balance method [69]. The harmonic balance method solves the governing fluid dynamic equations using temporal Fourier transforms coupled with a steady state CFD solver. A discrete number of temporal frequencies are considered, usually a fundamental frequency and several harmonics. Methods for solving for the relevant frequency in, say, a flutter analysis are available when the driving frequency is not known \textit{a priori}. Each frequency considered essentially requires the solution to a steady CFD problem, and the solution modes are coupled (in a nonlinear fashion) to construct the complete solution. To date, the harmonic balance fluid solver has been coupled to a simple two-dimensional rigid-body pitch/plunge structural model of an airfoil. While the method may be extended to three-dimensional problems with more sophisticated descriptions of
the structural modes, it is difficult to imagine its extension to transient aeroelastic phenomena where the modal signature of the fluid response is broadband in nature. An example of such a problem would be the response to control surface actuation. Nevertheless, for problems dominated by a single frequency and harmonics, the harmonic balance approach has the distinct advantage of not needing to resolve a transient solution before a periodic or quasi-periodic state is reached, as with time domain simulations. This can lead to significant computational savings. Transforming a time-domain CFD code to a frequency domain solver will lead to significant code development costs, however.

2.5 Interface Strategies

Handling of the interface region, or wetted surface, is the crux of the code coupling problem for fluid and structural analysis codes. The interface boundary conditions dictate that (i) the surface stress must be in equilibrium between the fluid and structure and (ii) the local displacement of the surface results in a corresponding local displacement of the fluid. Further continuity conditions require the fluid surface grid to follow the fluid-structure interface and the fluid mesh velocity at a point on the interface must equal the velocity of the interface itself. The stress boundary condition requires that fluid stresses must be transferred to the structural grid nodes before performing a structural solve. The displacement boundary condition requires that the resulting structural displacements cause a corresponding movement of the fluid mesh boundary at the interface.

The literature on fluid-structure interface strategies is somewhat disjointed. On the one hand, there is a body of work on efficient and general interpolation methods for transfer of data between two surfaces with differing discretizations. Emphasis here is on robustness, accuracy, and efficiency of the interpolations in the presence of complicated geometry and/or widely disparate fluid and structure grids. Then there is a more directed body of work on handling the interface between finite volume/finite element fluid codes and finite element structural codes in an accurate and conservative fashion. A third class of work deals with the details of implementing the methods in a computer code. There is very little cross-referencing in these areas of research, even in review articles, so it is difficult to compare the different strategies. Part of the reason for this is that the preferred interface strategy most likely depends on the application and the fidelity of the structural model. A wing-box model may require a much different interface treatment than a detailed FE model with shell elements. The goals of this review are to identify some of the interface strategies demonstrated on problems in computational aeroelasticity to date, and also identify driving factors in the selection of an interface strategy.

Most fluid/structure interface methods use some form of interpolation. Fluid loads are interpolated from the fluid grid points to the structural grid points, and displacements are interpolated from the structural grid points to the fluid grid, which is then deformed to accommodate the displacements. The performance of such a scheme will depend on the accuracy and robustness of the interpolation scheme [10, 66], as well as the grid densities near the interface. There are many candidate interpolation schemes, some of which are
tested in Smith et al. [66]. A simple and consistent approach for interpolation is to use the underlying finite element representation of the displacements for interpolation [30, 16]. Likewise, the finite volume (or fluid discretization) representation of the fluid loads is used to interpolate the forces.

Several authors have raised the importance of conservation of momentum and energy in the transfer of loads and displacements [4, 3, 13, 30, 34, 19]. It is possible to cast the transfer of structural mesh displacements to fluid grid point displacements in matrix form

\[ \mathbf{u}_f = \mathbf{T} \mathbf{u}_s. \]  

(2)

A scheme for transferring loads from the fluid grid points to the structural grid nodes that ensures a conservative transfer of energy between the two systems must then take the form

\[ \mathbf{f}_s = \mathbf{T}^T \mathbf{f}_f. \]  

(3)


The importance of energy conservation at the fluid-structure interface depends on the problem being solved. Problems where the fluid and/or structural grid are coarse will likely require a conservative scheme for good results. Also, prediction of complex dynamic aeroelastic phenomena like flutter and limit cycle oscillations (LCO) may be sensitive to the conservation properties of the numerical scheme. An imbalance in energy transfer between the fluid and the structure may excite spurious instabilities and, therefore, should be avoided. Note, however, that nonconservative interpolation schemes [34] and schemes designed to conserve momentum, but not necessarily energy [16, 17], have been applied successfully to problems in aeroelasticity.

The proper method for data transfer at the interface is related to the detail of the structural model. The review article by Guruswamy [34] breaks up data transfer algorithms for aeroelasticity problems according to the type of structural model employed: modal shapes, beam elements, wing-box models, plate and shell finite elements, and detailed finite element models of the entire structure. For problems with coarse representations of the structure, the near-interface fluid grid will still be fine enough to capture details of the flowfield in that region. A robust method to transfer displacements from the structural representation to the fluid grid is required. An example is the modeling of a 3D wing by 2D plate elements along the mean chord line. This issue is examined in [13] and examples of interpolation method performance in such situations are presented in [66].

Practical implementation of data transfer methods for general geometric configurations is not trivial. Details of a parallel, pre-processing program that “glues” the fluid grid to the finite element structural model are given in Maman and Farhat [48]. Some algorithmic
details for a similar methodology are given in Cebral and Löhner [16]. Sandia’s Algorithms for Contact in a Multiphysics Environment (ACME) library [12] can also be used to define a fluid-structure interface.

2.6 Applications

Computational analysis of nonlinear aeroelastic systems is still an expensive proposition. The state-of-the-art is represented by the simulations reported by Farhat et al. [27] for aeroelastic analysis of an F-16 fighter. The structural model is a detailed linear finite element model with 168,799 degrees of freedom. The unstructured fluid grid consists of 403,919 nodes. It is not clear from the paper if the Euler or Navier-Stokes equations are being solved, although the grid appears to be more appropriate for solution of the Euler equations. The total CPU time for a single simulation on six processors was 12.8 hours, while on 24 processors it was 3.3 hours. The computing platform was an SGI Origin 2000. Mapping a flutter boundary requires at least several simulations varying the flutter speed index (or altitude), while keeping Mach number fixed. The authors of [27] estimate that less than four days would be required to obtain five flutter boundary solutions (about 20-30 total simulations) using 24 processors. These timings assume that the relevant aeroelastic parameters can be extracted from the time domain solution after running only two cycles of the lowest frequency mode of the structure. This is accomplished in [27] by using a parameter identification algorithm called the Eigensystem Realization Algorithm [42]. These estimates may be optimistic considering the results of Melville [52] for time-domain analysis of a similar F-16 configuration using a Reynolds-Averaged Navier-Stokes code coupled with a modal structural model. Farhat et al. only ran the simulations for a total physical time of about 270 ms and considered only the first bending mode. Melville also shows the response of structural modes three and four, which required at least 1.5 seconds to capture due to the presence of a low frequency modulation of the signal. If these higher modes are of importance, the estimates of computational resources need to multiplied by a factor of about six.

An alternative to a “brute-force” time domain approach is to use CFD simulations to construct a reduced order model (ROM) of the aeroelastic system. The ROM seeks to simplify the problem by identifying important modes of the fluid system and/or the coupled system, basing the analysis on the contribution of these important modes. The computational cost is reduced due to a dramatic reduction in the number of degrees of freedom retained. Specific examples of ROMs, such as eigenmode decomposition and proper orthogonal decomposition (POD), are given in [24]. Construction of ROMs for nonlinear aerodynamic systems, considered in detail in Section 3 of this report, is an active area of research [46, 55].

One way to apply a ROM using CFD is through the harmonic balance approach. Only several structural modes are considered, and the fluid response (which may be nonlinear) is determined by solution in the frequency domain by solving for the driving frequency and up to several harmonics. Thomas et al. [69] have shown how an aeroelastic analysis may be
performed by providing a desired aero-structural mode response amplitude as an input, then solving for the associated Mach number, altitude, primary mode frequency, and other mode response amplitudes as unknowns using an iterative method. The computational efficiency is enhanced by use of a simplified pitch/plunge airfoil model and by applying the harmonic balance method to compute the required fluid responses.

The important lesson from these and other aeroelastic analyses is that a high-fidelity approach alone is not sufficient to provide a useful analysis capability. Either parameter identification methods, reduced order modeling, or a combination of the two is required to make the process useful and applicable over even a modest parameter space.

3 The Reduced Order Modeling Approach

Simulation of three-dimensional unsteady flow at high Reynolds number remains an expensive proposition, even with advances in large eddy simulation (LES) and the continuing rapid increase in available computing capacity. This makes such simulations of limited use to the analyst desiring CFD-based design, optimization, and parametric studies. Current alternatives to computing the time-dependent flow often require modeling assumptions that are too coarse to capture the relevant physics, e.g. Reynolds averaging.

The reduced order modeling approach seeks to derive an approximation to the physical system from some limited results of the high-fidelity simulation model. This approximate model is called the reduced order model (ROM). The reduced order model contains many fewer degrees of freedom (DOF) than the full simulation (where DOF scales with number of mesh points) and is thus much cheaper to compute. The terminology low dimensional model is often used in this same context, although here we use “reduced order” since the number of DOF of a quantitatively accurate model for a complex system could be $O(10^3)$, in which case “low-dimensional” is a poor descriptor. The terminology “low-dimensional” is often appropriate for those models seeking only qualitatively correct descriptions of the flow, for example in the identification of coherent structures in turbulent flow [37], or for short-time prediction of flow features for flow control applications [68].

This section focuses on reduced order modeling techniques appropriate for nonlinear fluid motion, with the ultimate goal of coupling the fluid ROM with a structural dynamics ROM for analysis of fluid/structure interaction.

3.1 The POD/Galerkin Approach

This section describes the POD/Galerkin method for reducing the order of complex physical systems. The approach consists of two steps: calculation of a reduced basis using the proper orthogonal decomposition of an ensemble of flowfield realizations, followed by Galerkin projection of the governing partial differential equations onto the reduced basis.
The first step involves the transfer of kinematic information from the high-fidelity simulation to a relatively small number of modes. The second step involves a translation of the full-system dynamics to the implied dynamics of these modes. When successful, the result of this procedure is a set of time-dependent ordinary differential equations in the modal amplitudes that accurately describes the flow dynamics of the full system of PDEs for some limited set of flow conditions.

### 3.1.1 Proper Orthogonal Decomposition

The Proper Orthogonal Decomposition (POD) is a mathematical procedure that, given an ensemble of data, constructs a basis for the ensemble that is optimal in a well-defined sense. POD is alternatively called Karhunen-Loève decomposition or Principle Components Analysis. It has been used in various scientific disciplines, including image processing, signal analysis, data compression, and oceanography. The mathematical development of POD for fluid flow applications in particular is described in some detail in [47] and [37]. The essentials of this development and the properties of POD most important to reduced order modeling are presented in this section.

Consider an ensemble \( \{u_k(x)\} \) of real vector fields on the domain \( x \in \Omega \). In the present context, the ensemble consists of a set of instantaneous snapshots of a numerical simulation solution field. The \( u \)'s are assumed to belong to an infinite-dimensional Hilbert space \( H(\Omega) \) with associated inner product \( (f, g) \). Following the approach of [64], we will defer the definition of the inner product until a particular application of the POD is considered, requiring only that it obey the usual requirements for an inner product. Note that this results in a general formulation for the POD that differs in some aspects from formulas derived for the \( L^2(\Omega) \) Hilbert space.

The POD basis is a set of functions \( \{\phi_j(x)\} \) that is the “best” linear basis for description of the ensemble. Since the basis is linear, a flowfield \( u \in \text{span}\{\phi_j\} \) can be represented as a linear combination of the POD modes,

\[
u(x,t) = \sum_j a_j(t) \phi_j(x).
\]

The POD modes, or empirical eigenfunctions, are defined by requiring that the averaged projection of the ensemble \( u^k \) onto \( \phi \) is a maximum:

\[
\max_{\phi \in H(\Omega)} \frac{\langle (u, \phi)^2 \rangle}{\|\phi\|^2},
\]

where \( \| \cdot \| \) is the norm generated by the inner product. The averaging operator \( \langle \cdot \rangle \) used in (5) could be an ensemble average over many experimental realizations, or it could be a time-average taken from different samples of a single experiment\(^1\). The main assumption

---

\(^1\)For an ergodic system, the time average and ensemble average will be equal as the number of samples becomes large.
regarding the averaging operator is that it commutes with the inner product. This assumption is shown to hold for the scalar case defined on the Hilbert space $L^2$ under certain conditions on $u$ (see section 3.8.1 of [37]).

The constrained optimization problem (5) with constraint $\|\phi\| = 1$ reduces to the eigenvalue problem

$$R\phi = \lambda \phi, \quad (6)$$

where

$$R\phi \equiv \langle u^k(u^k,\phi) \rangle. \quad (7)$$

The operator $R$ is self-adjoint and non-negative definite; if we further assume that $R$ is compact, then there are exists a countable set of non-negative eigenvalues $\lambda_i$, with associated eigenfunctions $\phi_i$. The eigenfunctions, appropriately normalized, form an orthonormal subspace of $H$, i.e. $(\phi_i, \phi_j) = \delta_{ij}$. The notions of compactness of operators and spaces, as well as the theory of self-adjoint operators, come from the mathematical discipline of functional analysis; see, e.g., [71]. For more detail on the compactness of $R$ and the required assumptions, see section 3.8.2 of [37].

The POD modes are the eigenfunctions $\phi_i$ associated with nonzero $\lambda_i$. Taking the inner product of (6) with $\phi$, it is straightforward to show that $\langle (u^k, \phi_i)^2 \rangle = \lambda_i$. In other words, the magnitude of the eigenvalue is equivalent to the average energy of the projection of the ensemble onto the associated eigenfunction, where the square of the inner product is interpreted as an energy measure. The POD modes may be ordered according to the magnitude of their eigenvalue, with $\lambda_1, \phi_1$ equal to the eigenvalue/eigenfunction pair with the largest eigenvalue, $\lambda_N$ equal to the smallest non-zero eigenvalue, and $\lambda_1 > \lambda_2 > \ldots > \lambda_n > \ldots > \lambda_N$. In building reduced order models one is interested in truncating the POD basis and retaining only the $K < N$ most energetic modes. It can be shown that the sequence of truncated POD bases form an optimal set, in the sense that a POD basis comprised of $K$ modes describes more energy (on average) of the ensemble than any other linear basis of the same dimension $K$. This compression of the ensemble energy into a minimum number of modes makes the POD basis attractive for reduced order modeling.

The span of the POD basis is not complete in $H(\Omega)$, but it is complete in the sense that, on average, any snapshot used to construct it can be represented, i.e. $\langle \|u^k - \sum_j (u^k, \phi_j) \phi_j\| \rangle = 0$. Conversely, each POD mode can be reconstructed as a linear combination of the observations used to construct the basis (section 3.3.1 of [37]). Thus,

$$\phi(x) = \sum_{k=1}^{M} a_k u^k(x). \quad (8)$$

Equation (8) is a mathematical expression of the intuitive argument that the POD basis contains only information on the kinematics of the flowfield that were already encoded in the observations. Further down the road in the reduced order modeling process, the dynamical features of a model will depend critically on the observations used to construct the reduced basis. A consequence of (8) is that the POD eigenfunctions share any closed linear property shared by all the ensemble members $u^k$. Examples of such properties are
the divergence-free property of incompressible flow and satisfaction of linear boundary conditions such as the no-slip surface condition.

In practice, the $u_k$ are vectors of state variables at discrete grid point locations, each containing a single solution from the numerical simulation. They will have length $NL$, where $N$ is the total number of grid points and $L$ is the number of dependent variables describing the flow state. Thus, the discretized version of (6) will be an eigenvalue problem of order $NL$. For $N \gg M$, where $M$ is the number of flowfield snapshots used, this procedure is costly and, it turns out, inefficient.

Sirovich [65] showed how the eigenvalue problem (6) can be reduced to order $M$, resulting in a much more efficient procedure for $N \gg M$. Assume that the averaging operator $\langle \cdot \rangle$ is a time average over a finite number of samples, so that $\langle f \rangle = 1/M \sum_{k=1}^{M} f^k$. Now, substitute the modal decomposition (4) into (6) to obtain

$$\frac{1}{M} \sum_{i=1}^{M} u^i \left( \sum_{k=1}^{M} a_k(t)u^k \right) = \lambda \sum_{k=1}^{M} a_k u^k. \quad (9)$$

Using the property $(x + y, z) = (x, z) + (y, z)$,

$$\frac{1}{M} \sum_{i=1}^{M} u^i \left( u^i, \sum_{k=1}^{M} a_k u^k \right) = \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{M} a_k \left( u^i, u^k \right)$$

$$= \sum_{i=1}^{M} \left[ \sum_{k=1}^{M} \frac{1}{M} (u^i, u^k) a_k \right] u^i = \lambda \sum_{k=1}^{M} a_k u^k. \quad (11)$$

A sufficient condition for the solution of (6) is then

$$\sum_{k=1}^{M} \frac{1}{M} (u^i, u^k) a_k = \lambda a_i; \quad i = 1, \ldots, M. \quad (12)$$

Equation (12) is one row of a new eigenvalue problem with row index $i$ and column index $k$. Once the eigenvectors for (12) are computed, the POD modes are computed using (8). This is the so-called “method of snapshots” for computing a POD basis.

The potential reduction in DOF when applying POD to nonlinear fluid flow problems can be estimated by examining reported results in the literature. Figure 2 gives results from several applications of POD, showing the number of grid points $N_{\text{grid}}$ required to perform a flow simulation along with the dimension of the POD basis $K_{\text{POD}}$ required to represent 95% of the fluctuation kinetic energy. Results from four separate studies are presented. Reference [21] examined the two-dimensional laminar flow past a circular cylinder. Reference [15] computed the two-dimensional transitional flow in a driven cavity. Reference [22] solved for the two-dimensional transitional flow past an airfoil. Reference [20] used POD to create ROMs for a three-dimensional large eddy simulation of the flow over a backward facing step. Reynolds numbers are based on cylinder diameter, airfoil chord, cavity depth, or step height. POD compresses the DOF by about three orders of magnitude for the two-dimensional laminar and transitional flows, and about four orders of magnitude for the single LES result.
3.1.2 Galerkin Projection

The second step for constructing the reduced order model is to project the governing PDEs onto the POD basis. A Galerkin projection is possible if the basis functions satisfy the boundary conditions of the PDE. In the previous section it was mentioned that the POD modes satisfy any linear boundary condition satisfied by the ensemble of flow-fields.

Consider a generic nonlinear PDE containing quadratic and cubic nonlinearities that governs the behavior of a time-dependent vector field $u(x,t)$,

$$
\frac{\partial u}{\partial t} = L u + N_2(u,u) + N_3(u,u,u). \tag{13}
$$

The operator $L$ is a linear operator, $N_2$ is a quadratic nonlinear operator, and $N_3$ is a cubic nonlinear operator. The Galerkin projection of equation (13) onto each POD mode $\phi_j$ is

$$
\left( \frac{\partial u}{\partial t}, \phi_j \right) = (L u, \phi_j) + (N_2(u,u), \phi_j) + (N_3(u,u,u), \phi_j). \tag{14}
$$
Substituting the POD decomposition for $u$ into (14), applying the algebraic rules of inner products along with orthogonality of the POD basis gives

$$\frac{da_k}{dt} = \sum_l d_l (\phi_k, L(\phi_l)) + \sum_{l,m} d_l a_m (\phi_k, N_2(\phi_l, \phi_m)) + \sum_{l,m,n} d_l a_m a_n (\phi_k, N_3(\phi_l, \phi_m, \phi_n)).$$

(15)

This is the reduced order model for equation (13) by the POD/Galerkin method. It is a time-dependent system of ODE’s of order equal to the number of retained POD modes $K$, with $k = 1, 2, \ldots, K$. The inner products in (15) are functionals of the known, time-independent POD modes $\phi(x)$, and may be precomputed before integration of the ROM. The nature of the nonlinearities present in the original equations strongly affects the cost of solving the reduced system. For example, the cost of evaluating a cubic term in a $K$-mode reduced order model is $K$ times that of evaluating a quadratic term for the same model.

It is customary in fluid mechanics applications of POD/Galerkin to build the reduced order model in terms of fluctuations about a mean state. Denoting the mean by $\bar{u}$ and the fluctuation by $\tilde{u}$, the state is written as

$$u(x, t) = \bar{u}(x) + \tilde{u}(x, t).$$

(16)

The assumption implicit in (16) is that the mean state is time-independent or at least changes very slowly relative to the time rate-of-change of the fluctuations. In this case the ROM contains additional terms related to the mean flow.

### 3.2 POD/Galerkin for Compressible Fluid Mechanics

#### 3.2.1 Equations

The compressible Navier-Stokes equations are usually solved in strong conservation law form in order to numerically conserve mass, momentum, and energy, a consideration particularly important in capturing shock waves. However, the conservative form of the equations is not convenient for applying Galerkin projection, since the flux terms cannot be written as simple products of the conservative state variables. For example, the momentum flux $\rho u^2$ cannot be written using products of $\rho$ and $\rho u$. It is useful to recast the equations in terms of primitive variables, so that all terms can be expressed as products of linear functions of the evolving primitive state variables.

First, we write the governing equations for a compressible, viscous fluid evolving on a moving mesh. The Arbitrary Lagrangian-Eulerian (ALE) approach is used, whereby the mesh state $x_i$, $i = 1, 2, 3$, moving at velocity $w_i$, is described relative to a static reference mesh state $\xi_i$. The Jacobian determinant,

$$J = \det \left( \frac{\partial x_i}{\partial \xi_j} \right),$$

(17)
links the “mixed” coordinates \( x_i \) to the “material” coordinates \( \xi \).

The non-dimensional, conservative form of the compressible Navier-Stokes equations for an ideal gas on a moving mesh is

\[
\frac{\partial (J \rho)}{\partial t} + J \frac{\partial}{\partial x_j} \left( \rho \left( u_j - w_j \right) \right) = 0 \tag{18}
\]

\[
\frac{\partial (J \rho u_i)}{\partial t} + J \frac{\partial}{\partial x_j} \left( \rho u_i \left( u_j - w_j \right) + p \delta_{ij} - \frac{1}{Re} \tau_{ij} \right) = 0 \tag{19}
\]

\[
\frac{\partial (J \rho E)}{\partial t} + J \frac{\partial}{\partial x_j} \left( \rho E \left( u_j - w_j \right) + pu_j - \frac{\gamma - 1}{Re} u_i \tau_{ij} + \frac{\gamma}{Re Pr} q_j \right) = 0, \tag{20}
\]

where \( \rho \) is the fluid density, \( u_i \) the fluid velocity, \( p \) the fluid pressure, \( E \) the total fluid energy, \( \tau_{ij} \) the viscous stress tensor, and \( q_j \) the fluid heat flux. The equations are non-dimensionalized using a reference length \( L \), reference density \( \rho_0 \), reference speed \( u_0 \), reference viscosity \( \mu_0 \), and reference coefficient of thermal conductivity \( \kappa_0 \). The Reynolds number is given by \( Re = \rho_0 u_0 L / \mu_0 \) and the Prandtl number is \( \mu_0 C_p / \kappa_0 \).

Using the identity

\[
\frac{\partial J \chi}{\partial t} = J \left( \frac{\partial \chi}{\partial t} + \frac{\partial (\chi w_j)}{\partial x_j} \right), \tag{21}
\]

Equations (18)–(20) may be written in terms of primitive state variables. Choosing the state vector as \( [\zeta \; u_1 \; u_2 \; u_3 \; p]^T \), where \( \zeta = 1/\rho \), and assuming constant \( \mu \), leads to a set of equations with only quadratic nonlinearities [39]:

\[
\frac{\partial \zeta}{\partial t} - \xi \frac{\partial u_j}{\partial x_j} + (u_j - w_j) \frac{\partial \zeta}{\partial x_j} = 0 \tag{22}
\]

\[
\frac{\partial u_i}{\partial t} + (u_j - w_j) \frac{\partial u_i}{\partial x_j} + \xi \frac{\partial p}{\partial x_j} + \delta_{ij} = \xi \frac{\partial \tau_{ij}}{\partial x_j} \tag{23}
\]

\[
\frac{\partial p}{\partial t} + (u_j - w_j) \frac{\partial p}{\partial x_j} + \gamma p \frac{\partial u_j}{\partial x_j} = \gamma - 1 \frac{\partial (u_i \tau_{ij})}{\partial x_j} - \frac{\gamma}{Re Pr} \frac{\partial q_j}{\partial x_j} \tag{24}
\]

### 3.2.2 Inner Products for Compressible Flow

For incompressible flow, it is customary to work with velocity fields that belong to the \( L^2(\Omega) \) Hilbert space. The \( L^2(\Omega) \) inner product is a natural choice for incompressible flow, since it is equal to twice the integrated kinetic energy over \( \Omega \).

For compressible flow, there is no obvious choice for the inner product. Rowley et al. [64] argued for the use of an energy-based inner product; they showed that an energy-based inner product preserves the stability of a stable fixed point at the origin. However, such an inner product is not easily defined for the full compressible flow equations without
seriously complicating the form of the equations and the resulting projections. Rowley et al. applied the isentropic flow assumption before the Galerkin projection step, so that the equations can be written in terms of the state vector \( \mathbf{q} = [u_1 \ u_2 \ u_3 \ a] \), where \( a \) is the isentropic speed of sound. In this case, an inner product related to the stagnation enthalpy (a conserved quantity in steady, inviscid flow) is

\[
(q^1, q^2) = \int_\Omega \left( u_1^1 u_2^2 + u_2^1 u_2^2 + u_3^1 u_3^2 + \frac{2\alpha}{\gamma - 1} a^1 a^2 \right) \, dV.
\] (25)

If \( \alpha = 1 \), the norm implied by the inner product corresponds to the stagnation enthalpy, while taking \( \alpha = 1/\gamma \) gives a norm corresponding to the stagnation energy.

For strongly irreversible flows, including most turbulent flows, the isentropic assumption is poor and the use of isentropic flow equations is questionable. However, the inner product given by equation (25) could still be used in conjunction with the full Navier-Stokes equations written in the primitive form of equations (22)–(24). In this case the inner product can be re-written as

\[
(q^1, q^2) = \int_\Omega \left( u_1^1 u_2^2 + u_2^1 u_2^2 + u_3^1 u_3^2 + \frac{2\alpha \gamma}{\gamma - 1} \left[ p^1 \xi^2 + p^2 \xi^1 \right] \right) \, dV.
\] (26)

Iollo et al. [39] improved stability of reduced order models for compressible, viscous flow by working in the Sobolev space \( H^1(\Omega) \). Using non-dimensional variables, they compared results using an \( L^2 \) norm with those using a Sobolev space with inner product

\[
(q^1, q^2) = \int_\Omega \left( q^1 \cdot q^2 + \epsilon (\nabla q^1 \cdot \nabla q^2) \right) \, dV.
\] (27)

The constant \( \epsilon \) is chosen as proportional to \( T/Re \), where \( T \) is some time scale. The major weakness of this inner product is that the choice of \( T \) is arbitrary; further guidance for specification of \( T \) based on supporting analysis is needed.

### 3.2.3 Subgrid Modeling of Turbulence

To date, there have been two main thrusts in application of the POD/Galerkin methodology to fluid flow. The first, and original, use of the technique seeks to build truly low-dimensional models of transitional or turbulent flow, derived from direct numerical simulation (DNS) data. The primary goal of this line of work is to gain a better understanding of the dominant physical mechanisms that either govern turbulent motion or cause transition to turbulence. Therefore, the use of DNS is appropriate and, for most situations, necessary in order to faithfully describe the relevant flow physics.

The second use of the POD/Galerkin technique is to build a reduced order model, using relatively few DOF, that accurately mimics the results of an expensive simulation. In this
case the goal is to reproduce results from a flow simulation, which may itself use physical modeling assumptions. The utility of the model is closely tied to its robustness in successfully mimicking the full model results over a range of input parameters of interest. Note that the success in the reduced order modeling process is tied to the accurate reproduction of the simulation model from which the reduced model is derived. Further, the reduced order model can be no more accurate than the simulation(s) used to construct it.

Most reduced order models for fluid flow have been constructed for laminar or transitional flow using DNS. DNS is limited to relatively low Reynolds numbers due to the well-known scale separation phenomenon that becomes more pronounced as Reynolds number increases. LES has been developed to deal with this limitation. In LES, only the larger, most energetic flow structures (or eddies) are resolved with the simulation grid, while the effect of the smaller scale fluid motions on the energetic eddies is modeled with a sub-grid scale (SGS) model. The cost benefits of LES are very large for free shear flows, where the cost of LES for a given level of resolution (of turbulent flow energy) is independent of the Reynolds number, for large Reynolds number. For wall-bounded flows the situation is less favorable due to the persistence of energetic motion at small scales very close to the wall. For wall-bounded LES with resolution of near-wall flow structure, the cost scales at least with \( R_{e}^{1.76} \). However, with resolution of only the mean near-wall flow structure the cost scales with \( \ln(R_{e}) \), and with wall-function modeling of the turbulent boundary layer the cost again becomes independent of Reynolds number [62].

Those who have sought to use DNS to build reduced order models of a turbulent flow often incorporate some form of turbulence model in the reduced order model. This is because the modal representation of the flow is truncated at some point, so that the dissipative action of the small-scale modes is neglected. This dissipative action is essential in setting the proper rate of energy transfer from large scales to small scales and must be modeled in some way. An alternative to this approach is to perform the simulations using LES, which already contains a model for dissipation of turbulent kinetic energy. The reduced order model is then built using the LES simulation data as well as the LES equations (i.e. the filtered Navier-Stokes equations), which contain the subgrid model terms.

### 3.2.4 POD for Moving Fluid Meshes

Coupled fluid/structure interaction problems with moving boundaries require special numerical schemes such as the ALE-based schemes reviewed in Section 2. In the ALE formulation the fluid mesh deforms with the fluid boundary, which in turn is driven by some description of the structural deformation. These considerations necessitate a new look at the POD methodology, which is typically derived for, and applied to, static domains and static fluid meshes.

Perhaps the only serious look at POD on arbitrarily moving meshes is the thesis of Anttonen [2]. In this work the POD procedure is defined \textit{a priori} and in a discrete sense, so that the inner product is the discrete \( L^2 \) inner product. Flowfield snapshots are taken as the
simulation proceeds and the fluid mesh deforms. The POD modes are computed using the usual method of snapshots, without reference to the changing mesh point coordinates. This effectively changes the definition of the inner product in continuous space, as the definition in discrete coordinates remains fixed relative to the mesh motion. The result is a valid POD basis that may be used to reconstruct the sampled flowfield, but the POD representation is not as efficient as the static grid case.

Reference [2] then examines the error introduced in a POD-based ROM when the ROM is used to predict a flow where the grid deformation is different than that used to construct the POD basis. The grid deformation is thus viewed as a parameter in the ROM; the robustness of the ROM is tied to its accuracy in predicting flows where this parameter is “off-design.” It proves difficult to construct an accurate blended POD in this manner, where the POD ensemble is taken from several simulations, each with different grid deformations. The solution adopted in [2], termed MULTI-POD, is to construct several POD ROMs and switch from one to the other based on a measure of how closely the current mesh configuration matches that of each ROM. Alternative methods for building ROMs for moving boundary problems are proposed later in this document in Section 3.3.

### 3.2.5 Modeling Boundary Conditions

In most applications of POD/Galerkin to fluid flows, the boundary conditions are steady, and are satisfied by all of the POD basis functions. This allows a straightforward Galerkin projection that results in an initial value problem for the modal amplitudes that does not contain any boundary terms. An exception to this occurs in low-dimensional modeling of turbulent boundary layers, where the domain normal to the wall is often truncated within the boundary layer and the resulting boundary term requires some modeling.

In the applications considered here, the ROM is specified over the full computational domain so that boundary truncation is not an issue. However, it would be very useful for many types of analysis to consider boundary condition-driven problems, where the boundary conditions may serve as parameters in the problem. Examples of this in a fluid/structure interaction setting are changes in the far-field angle of attack or yaw angle as a function of time, or time-dependent wall velocities resulting from motion of the structure. In this case the Galerkin projection can be carried out such that certain terms in the inner product evaluation are manipulated to generate boundary integral terms in the ROM. The boundary conditions can then be specified within the integrand of these surface integrals.

Of interest are far-field velocity boundary conditions and solid wall boundary conditions that may either be steady but differ from problem to problem, or unsteady. Let \( \mathbf{u}_\infty \) be the specified velocity vector at the far-field boundary \( \Gamma_\infty \), and let \( \mathbf{u}_{wi} \) be the specified wall boundary velocity at the solid surface boundary \( \Gamma_w \). For simplicity, consider the case where the grid is not moving, \( i.e. \ w_i = 0 \) in Equations (22)–(24)). The continuity equation, Equation (22), contains the quadratic term \( \zeta \frac{\partial u_i}{\partial x_j} \). The Galerkin projection of this term...
performed in Equation (15) will result in the following integral,

\[ \int_{\Omega} \phi_j \zeta \frac{\partial u_k}{\partial x_k} \, d\Omega, \]  

(28)

where \( \phi_j \) is a scalar variable that depends upon the definition of the inner product. This integral can be re-written as

\[ \int_{\Omega} \phi_j \zeta \frac{\partial u_k}{\partial x_k} \, d\Omega = \int_{\Omega} \left( \frac{\partial (\phi_j \zeta u_k)}{\partial x_k} - u_k \frac{\partial (\phi_j \zeta)}{\partial x_k} \right) \, d\Omega. \]  

(29)

Application of the divergence theorem gives

\[ \int_{\Omega} \left( \frac{\partial (\phi_j \zeta u_k)}{\partial x_k} - u_k \frac{\partial (\phi_j \zeta)}{\partial x_k} \right) \, d\Omega = \int_{\partial \Omega} \phi_j \zeta u_k n_k \, dS - \int_{\Omega} u_k \frac{\partial (\phi_j \zeta)}{\partial x_k} \, d\Omega. \]  

(30)

The surface integral in Equation (30) is retained in the \( j \)th ROM equation, with the wall-normal velocity \( u_k n_k \) now specified as a function of time. The volume integral term is transformed to a term that is nonlinear in the modal amplitudes \( a(t) \), in the manner of equation (15). This boundary condition is imposed on \( \Gamma_w \) for an inviscid calculation. On \( \Gamma_\infty \) the problem is still underconstrained, since only the normal component of the far-field boundary velocity can be specified. The other components can be specified by considering the convective terms of the momentum equation, (23), which appear during the projection step as

\[ \int_{\Omega} \phi_j u_k \frac{\partial u_i}{\partial x_k} \, d\Omega, \]  

(31)

where \( \phi_j \) is the scalar component of the \( j \)th mode corresponding to the \( i \)th velocity component. Again applying the differential chain rule followed by the divergence theorem gives

\[ \int_{\Omega} \phi_j u_k \frac{\partial u_i}{\partial x_k} \, d\Omega = \int_{\partial \Omega} \phi_j u_i u_k n_k \, dS - \int_{\Omega} u_i \frac{\partial (\phi_j \zeta u_k)}{\partial x_k} \, d\Omega. \]  

(32)

The surface integral is retained in the ROM, allowing the entire velocity vector \( u_{\infty} \) to be specified on \( \Gamma_\infty \). The wall velocity for a no-slip wall could also be specified using this method. An alternative method for specifying the velocity for a no-slip wall is to appeal to the viscous stress terms in the momentum equation. This strategy is not pursued here, but may lead to a more physically consistent means for applying the boundary condition.

This technique for ROM boundary conditions was applied by Ref. [58] in the context of reduced order modeling of the Boussinesq equations. The application was thermal convection in a cavity driven by a non-homogenous heat flux boundary condition. Of critical importance for ROM accuracy and robustness is that the POD basis spans the solution set for all anticipated boundary conditions. This means that the ROM must be trained using an ensemble of snapshots that captures the anticipated boundary conditions. This is accomplished in [58] by considering a Fourier decomposition of boundary heat flux distributions.
and collecting a series of snapshots for each of a range of Fourier wave numbers and amplitudes. The analogue for boundary motion in a fluid/structure interaction problem is to form a POD basis from a range of individual structural mode deflections and amplitudes. For a problem where far-field flow direction changes, solutions over the range of far-field conditions are needed to construct the POD basis.

3.3 Coupled Fluid/Structure ROMs

The tools described in the previous sections for fluid ROM construction can be combined with a structural dynamics model to form a coupled fluid/structure ROM. The approach taken here is to keep the development of the fluid and structural ROMs segregated, then combine them into a single ROM using the appropriate boundary conditions. The reasons for this are two-fold. First, developing the fluid ROM independently keeps the number of parameters required in the ROM to a minimum. For example, the two primary governing parameters for a compressible, viscous fluid are the Mach number and Reynolds number. One can expect a fluid model in isolation to be sensitive to changes in these parameters. The structural response to a fluid flow, however, also depends on the dynamic pressure, with which the aerodynamic loading of the structure scales. A coupled ROM constructed from the beginning using fluid and structural responses would require the introduction of the dynamic pressure as an additional parameter. Second, calculations of structural eigenmodes, at least in the case of small amplitude (linear) displacements, is a routine function of structural dynamics codes. It is prudent to take advantage of this existing analysis capability in formulating the ROM.

A finite element (FE) model of the structural equations of dynamic equilibrium can be written

\[ M\ddot{u}_s + C\dot{u}_s + Ku_s = f_A(u_s, \mathbf{q}), \]  

(33)

where \( u_s \) is the vector of structural displacements, and \( M, C, \) and \( K \) are the mass, damping, and stiffness matrices. \( f_A \) is the vector containing aerodynamic forces, which are functions of the displacements and the fluid state \( \mathbf{q} \). In the linear case, the structural dynamics can often be described by a small number of eigenmodes \( \xi_i \), and the displacement vector can be written in terms of the modes using the decomposition \( u_s = \sum_{i=1}^{S} b_i \xi_i \), where \( S \) is the dimension of the structural modal basis.

Two approaches to construction of the coupled fluid/structure ROM are outlined here. In Approach I, the structural displacements are assumed to be small, so that a linear structural dynamics model is appropriate. As a consequence of the small structural displacements, the mesh motion is also small. The important structural dynamics modes are computed for the structure of interest. Impulse responses of the fluid system to the structural modes are simulated and used to form the ensemble from which the POD basis is extracted. In this case, since the mesh motion is small, the inner products used to form the POD basis and the Galerkin projection are computed with respect to a mean static mesh state. The
effect of surface deformation is accounted for in the ROM using a transpiration boundary condition on a static mesh, Equation (30) or (32). The aerodynamic force vector $f_A$ is computed using the boundary stresses predicted by the fluid ROM. The structural dynamics equations are either loosely or tightly coupled to the fluid ROM, and the entire system is integrated forward in time. This approach takes full advantage of the reduction in DOF provided by Galerkin projection, but is limited by the restriction to small mesh motion.

In Approach II, the mesh motion is unrestricted, and either a linear or nonlinear description of the structural system is employed. Again, the POD ensemble is formed using fluid responses to structural modal impulses. However, the mesh velocity $w_i$ in the ALE equations for the fluid is retained in the Galerkin projection step. Further, the inner products are computed with respect to the reference state of the mesh $\xi_i$. Recall the definition of the Jacobian determinant in Equation (17), which relates the mixed coordinates of a current mesh state to the reference coordinates. A differential volume of the mesh is related to the reference state by $dV = J(t) dV_{\xi}$. The inner product is now defined relative to the reference state as $\int \ldots J(t) dV_{\xi}$. Also, the appearance of the mesh velocity as a parameter in the governing PDEs results in new contributions to the linear term in the ROM, Equation (15). All the inner products in Equation (15) are now functions of time which, strictly speaking, must be computed at each time step of the ROM. Since the computation of the inner products is $O(N)$, where $N$ is the number of full simulation grid points, this leads to a substantial increase in the cost of integrating the ROM. Some of this additional cost may be eliminated by re-evaluating the inner products every $P$ time steps of the ROM, rather than every time step. An inexpensive error indicator, such as the maximum relative change in cell volume over the domain since the last inner product update, could be used to initiate a new update. Approach II is a new concept that should be validated on simple problems before consideration for large scale fluid/structure applications.

3.4 Sources of Error in Reduced Order Modeling

In the previous sections we emphasized that the worth of a reduced order model is measured by its ability to reproduce the behavior of a high-fidelity simulation model at a much lower cost. A useful definition of the error of a reduced order model is then the difference between the solution of the reduced order model and the solution that would be obtained by running a high-fidelity simulation for the same set of input parameters, initial conditions, and boundary conditions. It is useful to identify and classify possible sources of error in the reduced order model so that appropriate strategies for their mitigation can be developed. Several such error sources are listed in this section, along with possible mitigation strategies.

Truncation of the POD basis: The POD/Galerkin strategy derives its efficiency benefits from truncation of the POD basis to a “reasonable” dimension. However, enough POD modes must be retained so that the dynamics of the ROM are both qualitatively and quantitatively similar to those of the full simulations. Qualitative similarity means that the topology of the phase space of the ROM is the same as that of the simulations. For example, if
the important dynamics of a system are described by a stable limit cycle that is captured by the simulation model, then the ROM phase space should include this limit cycle with the proper stability attributes. Qualitative similarity can be missed even when the retained POD modes completely represent the observed PDE solution. A simple example of this behavior is given in [63]. This type of inaccuracy in the ROM, discussed in more detail in the discussion of insufficient sampling to follow, can be mitigated by sampling over a larger region of phase space. Quantitative similarity means that the ROM reproduces a time history of the flowfield that is an accurate representation of the full simulation solution. This error is usually a strong function of the number of POD modes retained. Note, however, that for a turbulent flow it may not be possible for a ROM to perfectly track the time history of a DNS or LES simulation. In this case the flow statistics of the ROM can still mimic the behavior of the statistics derived from the simulation.

The question of how many POD modes to retain in order to maintain qualitative dynamical similarity and quantitative accuracy is flow-dependent and largely unanswered. A common recipe is to set an energy threshold, say 95%, and keep enough POD modes so that the averaged energy of the truncated basis exceeds this threshold. Use of POD/Galerkin models for quantitative predictions should be accompanied by convergence studies for different numbers of retained modes. However, note that performance of the ROM may actually deteriorate with retention of higher modes due to insufficient sampling (discussed next), insufficient spatial resolution of the higher modes, or roundoff error from the eigenproblem solution used to construct the POD. Systematic procedures for assessing the required number of modes for a given application are needed.

**Insufficient sampling:** A large enough sample size must be computed to allow computation of enough POD modes to form an accurate basis. In general, the larger the number of samples, the better, since more of the system dynamics will be represented in the ensemble and transferred to the POD modes. Perhaps even more important than the number of samples is the type of samples that are taken. For some flows it is necessary to sample both on the solution trajectories of interest as well as away from them. An example is an oscillatory bluff-body wake. If one samples only from the periodic wake state, the POD modes will be able to reproduce the wake shedding behavior but a ROM derived from them may not possess the shedding solution as a stable state. This is because the limit cycle of shedding is stable in the simulation (trajectories near the limit cycle converge to the limit cycle at large times), but unstable in the ROM (the limit cycle exists, but nearby trajectories diverge from it). However, if nearby trajectories are included in the sampling space, qualitative similarity is more likely. This can be done in a simple, ad hoc way by including the transient portion of the simulation (the convergence of the solution to the limiting trajectory, or “attractor”) in the ensemble used to construct the POD.

**Numerical integration of inner products:** In this work the POD/Galerkin process is defined for a continuous system, then discretized. This involves both discretization of the system of PDEs for the simulation, as well as numerical integration of the inner products used to form the POD modes and the Galerkin projection. Clearly, the accuracy of the numerical integration can affect the resulting model fidelity. One way to reduce this source
of error is to use high-order numerical integration techniques to form the inner products.

**Numerical integration of the ROM**: Once the ROM is constructed it is integrated using an appropriate time-advancement technique. Errors in the temporal discretization of the ROM may lead to inaccurate or spurious behavior. This error can be quantified and minimized using careful time step convergence studies.

## 4 Plans for Future Work

Section 3 outlined some approaches for building ROMs of fluid/structure systems that accommodate nonlinear fluid behavior. This is an active area of research, and several new approaches were outlined that have not yet been validated. These include using the POD/Galerkin procedure to model changes in boundary conditions, and using POD/Galerkin on moving mesh problems.

Examination of these new approaches should be carried out in phases, moving from simple model problems to more complicated and realistic problems. A relatively simple one-dimensional problem that exhibits interesting nonlinear behavior is the modeling of a tubular reactor. The behavior of a tubular reactor can be described by a system of convection-diffusion-reaction (CDR) equations [36]. The CDR equations model similar physical processes (convection, diffusion) to those found in the Navier-Stokes equations. The POD/Galerkin procedure can be applied to such a system and the strategies for accommodation of mesh motion can be applied in a one-way coupled setting, where the mesh motion is driven by an autonomous external source.

A simple computer code has been written to solve the CDR equations. This code can be used to generate snapshots for computation of a POD basis using the RBGEN module of the Trilinos linear algebra library [25]. Following the successful demonstration of the ROM methodology to the one-dimensional CDR equations, the focus of this work will shift to multi-dimensional fluid flow applications. Of particular interest is the separated flow over stalled wings and wind turbine blades, which can lead to complicated aeroelastic phenomena such as stall flutter. Efforts will initially be concentrated on reduced order modeling of the flow-field, followed by coupling of the fluid ROM to a structural dynamics model.
References


DISTRIBUTION:

1 MS 0384  
Art Ratzel, 1500

1 MS 0824  
Wahid Hermina, 1510

10 MS 0825  
Matthew Barone, 1515

1 MS 0825  
Basil Hassan, 1515

1 MS 0825  
Jeffrey Payne, 1515

1 MS 0825  
Walt Wolfe, 1515

1 MS 0825  
Larry DeChant, 1515

1 MS 0847  
Pete Wilson, 1520

1 MS 0555  
Mark Garrett, 1522

1 MS 0893  
John Pott, 1523

1 MS 0847  
Jim Redmond, 1524

1 MS 0557  
Daniel Segalman, 1524

1 MS 0557  
Thomas Baca, 1525

1 MS 0372  
Rodney May, 1526

1 MS 0372  
Joseph Jung, 1527

1 MS 0821  
Anthony Thornton, 1530

1 MS 0821  
Lou Gritzo, 1532

1 MS 1135  
Steve Heffelfinger, 1534

1 MS 0384  
Hal Morgan, 1540

1 MS 0382  
Steve Gianoulakis, 1541

1 MS 0382  
Steve Bova, 1541

1 MS 0380  
Ken Alvin, 1542

1 MS 0382  
Jim Stewart, 1543

1 MS 0316  
Curtis Ober, 1433

1 MS 0316  
Thomas Smith, 1433

1 MS 0382  
Alfred Lorber, 1541

1 MS 1110  
Scott Collis, 1414

1 MS 0708  
Paul Veers, 6214
| MS 0708 | Dale Berg, 6214 |
| MS 0708 | Jose Zayas, 6214 |
| MS 1162 | Walt Rutledge, 5422 |
| MS 9018 | Central Technical Files, 8945-1 |
| MS 0899 | Technical Library, 4536 |
| MS 0123 | D. Chavez, LDRD Office, 1011 |