Aerodynamic Data Reconstruction and Inverse Design using Proper Orthogonal Decomposition

T. Bui-Thanh*, M. Damodaran†

Singapore-Massachusetts Institute of Technology Alliance (SMA)

School of Mechanical and Production Engineering

Nanyang Technological University, Nanyang Avenue, Singapore 639798

K. Willcox‡

Aerospace Computational Design Laboratory

Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

The application of proper orthogonal decomposition for incomplete (gappy) data for compressible external aerodynamic problems has been demonstrated successfully in this paper for the first time. Using this approach, it is possible to construct entire aerodynamic flow fields from the knowledge of computed aerodynamic flow data or measured flow data specified on the aerodynamic surface, thereby demonstrating a means to effectively combine experimental and computational data. The sensitivity of flow reconstruction results to available measurements and to experimental error is analyzed. Another new extension of this approach allows one to cast the problem of inverse airfoil design as a gappy data problem. The gappy methodology demonstrates a great simplification for the inverse airfoil design problem, and is found to work well on a range of examples, including both subsonic and transonic cases.

*SMA Graduate Research Student

†Associate Professor, Associate Fellow AIAA

‡Assistant Professor of Aeronautics and Astronautics, Member AIAA

Introduction

The proper orthogonal decomposition (POD), also known as Karhunen Loève expansion and principle components analysis, has been widely used for a broad range of applications. POD analysis yields a set of empirical modes, which describes the dominant behavior or dynamics of given problem. This technique can be used for a variety of applications, including derivation of reduced-order dynamical models,[1] steady analysis and design of inviscid airfoils,[2] image processing[3] and pattern recognition.[4]

Sirovich introduced the method of snapshots as a way to efficiently determine the POD modes for large problems.[5] In particular, the method of snapshots has been widely applied to computational fluid dynamic (CFD) formulations to obtain reduced-order models for unsteady aerodynamic applications.[6–9] A set of instantaneous flow solutions, or "snapshots" are obtained from a simulation of the CFD method. The POD process then computes a set of modes from these snapshots, which is optimal in the sense that, for any given basis size, the error between the original and reconstructed data is minimized. Reduced-order models can be derived by projecting the CFD model onto the reduced space spanned by the POD modes.

Everson and Sirovich[10] have developed a modification of the basic POD method that handles incomplete or "gappy" data sets. Given a set of POD modes, an incomplete data vector can be reconstructed by solving a small linear system. Moreover, if the snapshots themselves are damaged or incomplete, an iterative method can be used to derive the POD basis. This method has been successfully applied for reconstruction of images, such as human faces, from partial data.

In the research described here, the gappy POD approach is extended for application to aerodynamic problems. Incomplete aerodynamic data may arise in several situations. Firstly, a limited set of data may be available from experimental measurements. The gappy POD provides a way to reconstruct full flowfield information, using a combination of the available experimental and supplemental computational data. Secondly, certain data may

be missing because it is not known. For example, one may have a set of snapshots which correspond to a set of airfoil shapes and their respective flowfields. Given a new airfoil shape, the gappy POD provides a way to quickly estimate the corresponding flowfield. Conversely, the gappy POD can be used to solve the problem of inverse design: given a target pressure distribution, the optimal airfoil shape can be determined by appropriate interpolation of known designs. Finally, data may be incomplete due to damage of storage facilities.

Many of the current methods for airfoil design focus on the use of optimization. Lighthill[11] developed pioneering work using the method of conformal mapping, which was later extended to compressible flows by McFadden.[12] By introducing the finite-difference method to evaluate sensitivity derivatives, Hicks and Henne[13] first attempted to solve the airfoil design problem as a constrained optimization. Since then, gradient-based methods have been widely used for aerodynamic design. More recently, Jameson[14] applied control theory to shape design optimization for Euler and Navier-Stokes problems, using an efficient adjoint approach to obtain gradient information. In order to reduce the computational cost of solving the design optimization problem, Legresley and Alonso[2] used the POD technique for both direct and inverse designs. The gappy POD method presented here for the inverse design problem is fast compared with other optimization methods, which is an advantage for routine use in design. In addition, the gappy approach allows for both computational and prior experimental data to be utilized in the design process.

In this paper, the basic POD approach is first outlined, followed by a description of the gappy POD method. A series of examples are then presented which demonstrate how the gappy POD method can be used for reconstruction of flowfield data and extended for airfoil inverse design. Finally, we present some conclusions.

Proper Orthogonal Decomposition Theory and Extensions

Proper orthogonal decomposition

The basic POD procedure is summarized briefly here. The optimal POD basis vectors Φ are chosen to maximize the cost:[1]

$$\max_{\Psi} \frac{\langle|(U,\Psi)^2|\rangle}{(\Psi,\Psi)} = \frac{\langle|(U,\Phi)^2|\rangle}{(\Phi,\Phi)} \tag{1}$$

where (U,Φ) is the inner product of the basis vector Φ with the field $U(x,t)$, x represents the spatial coordinates, t is time, and $<>$ is the time-averaging operation. It can be shown that the POD basis vectors are eigenfunctions of the kernel K given by

$$K(x,x') = \langle U(x,t), U^*(x',t)\rangle \tag{2}$$

where U^* denotes the hermitian of U. The method of snapshots, introduced by Sirovich,[5] is a way of determining the eigenfunctions Φ without explicitly calculating the kernel K. Consider an ensemble of instantaneous field solutions, or "snapshots". It can be shown that the eigenfunctions of K are linear combinations of the snapshots as follows

$$\Phi = \sum_{i=1}^{m} \beta_i U^i \tag{3}$$

where U^i is the solution at a time t_i and the number of snapshots, m, is large. For fluid dynamic applications, the vector U^i contains the flow unknowns at a given time at each point in the computational grid. The coefficients β_i can be shown to satisfy the eigen-problem

$$R\beta = \Lambda\beta \tag{4}$$

where R is known as the correlation matrix

$$R_{ik} = \frac{1}{m}\left(U^i, U^k\right) \tag{5}$$

The eigenvectors of R determine how to construct the POD basis vectors [using (3)], while the eigenvalues of R determine the importance of the basis vectors. The relative "energy" (measured by the 2-norm) captured by the i^{th} basis vector is given by $\lambda_i / \sum_{j=1}^{m} \lambda_j$. The approximate prediction of the field U is then given by a linear combination of the eigenfunctions

$$U \approx \sum_{i=1}^{p} \alpha_i \Phi^i \tag{6}$$

where $p << m$ is chosen to capture the desired level of energy, Φ^i is the i^{th} POD basis vector, and the POD coefficients α_i must be determined as a function of time.

The basic POD procedure outlined above considers time-varying flows by taking a series of flow solutions at different instants in time. The procedure can also be applied in parameter space as in Epureanu et al.,[15] that is, obtaining flow snapshots while allowing a parameter to vary. The parameter of interest could, for example, be the flow freestream Mach number, airfoil angle of attack, or airfoil shape.

POD for reconstruction of missing data

In CFD applications, the POD has predominantly been used for deriving reduced-order models via projection of the governing equations onto the reduced space spanned by the basis vectors. Here, we consider a different application of the method, which is based on the gappy POD procedure developed by Everson and Sirovich[10] for the reconstruction of human face images from incomplete data sets. In this paper, the gappy POD methodology will be extended for consideration of fluid dynamic applications. The gappy POD procedure is first described.

The first step is to define a "mask" vector, which describes for a particular flow vector where data is available and where data is missing. For the flow solution U^k, the corresponding

mask vector n^k is defined as follows:

$$\begin{aligned} n_i^k &= 0 \text{ if } U_i^k \text{ is missing or incorrect} \\ n_i^k &= 1 \text{ if } U_i^k \text{ is known} \end{aligned}$$

where U_i^k denotes the i^{th} element of the vector U^k. For convenience in formulation and programming, zero values are assigned to the elements of the vector U^k where the data is missing, and pointwise multiplication is defined as $\left(n^k, U^k\right)_i = n_i^k U_i^k$. Then the gappy inner product is defined as $(u, v)_n = ((n, u), (n, v))$, and the induced norm is $(\|v\|_n)^2 = (v, v)_n$.

Let $\{\Phi^i\}_{i=1}^m$ be the POD basis for the snapshot set $\{U^i\}_{i=1}^m$, where all snapshots are completely known. Let g be another solution vector that has some elements missing, with corresponding mask vector n. Assume that there is a need to reconstruct the full or "repaired" vector from the incomplete vector g. Assuming that the vector g represents a solution whose behavior can be characterized with the existing snapshot set, an expansion of the form (6) can be used to represent the intermediate repaired vector $\tilde{g}$ in terms of p POD basis functions as follows:

$$\tilde{g} \approx \sum_{i=1}^{p} b_i \Phi^i \tag{7}$$

To compute the POD coefficients b_i, the error, E, between the original and repaired vectors must be minimized. The error is defined as

$$E = \|g - \tilde{g}\|_n^2 \tag{8}$$

using the gappy norm so that only the original existing data elements in g are compared. The coefficients b_i that minimize the error E can be found by differentiating (8) with respect to each of the b_i in turn. This leads to the linear system of equations

$$Mb = f \tag{9}$$

where $M_{ij} = (\Phi^i, \Phi^j)_n$ and $f_i = (g, \Phi^i)_n$. Solving equation (9) for b and using (7), the intermediate repaired vector $\tilde{g}$ can be obtained. Finally, the complete g is reconstructed by replacing the missing elements in g by the corresponding repaired elements in $\tilde{g}$, i.e. $g_i = \tilde{g}_i$ if $n_i = 0$.

POD with an incomplete snapshot set

The gappy POD procedure can be extended to the case where the snapshots themselves are not completely known. In this case, the POD basis can be constructed using an iterative procedure. Consider a collection of incomplete data $\left\{g^k\right\}_{k=1}^m$, with an associated set of masks $\left\{n^k\right\}_{k=1}^m$. The first step is to fill in the missing elements of the snapshots using average values as follows:

$$h_i^k(0) = \begin{cases} g_i^k & \text{if } n_i^k = 1 \\ \bar{g}_i & \text{if } n_i^k = 0 \end{cases} \tag{10}$$

where $\bar{g}_i = \frac{1}{P_i} \sum_{k=1}^m g_i^k$, $P_i = \sum_{k=1}^m n_i^k$ and $h^k(l)$ denotes the l^{th} iterative guess for the vector h^k. A set of POD basis vectors can now be computed for this snapshot set, and iteratively used to refine the guess for the incomplete data. The procedure can be summarized as follows, beginning with $l = 0$:

1. Use the basic POD procedure on the snapshot set $\left\{h^k(l)\right\}_{k=1}^m$ to obtain the POD basis vectors for the current iteration, $\left\{\Phi^k(l)\right\}_{k=1}^m$.

2. Use the first p of these POD basis vectors to repair each member of the snapshot ensemble, as described in the previous section. The intermediate repaired data for the current iteration is given by

$$\tilde{h}^k(l) = \sum_{i=1}^{p} b_i^k(l) \Phi^i(l) \tag{11}$$

3. The values from these intermediate repaired data are now used to reconstruct the

missing data for the next iteration as follows

$$h_i^k(l+1) = \begin{cases} h_i^k(l) & \text{if } n_i^k = 1 \\ \tilde{h}_i^k(l) & \text{if } n_i^k = 0 \end{cases} \tag{12}$$

4. Set $l = l + 1$ and go to step 1.

The above iterative procedure should be repeated until the maximum number of iterations is reached or until the algorithm has converged. When evaluating convergence, one can consider both the POD eigenvalues and the POD eigenvectors as will be demonstrated in the results.

Inverse Design via the Gappy POD Method

We now describe how the gappy POD method can be extended for fluid dynamic applications. In particular, a new variant of the method is proposed to perform inverse design of a two-dimensional airfoil. Typically, given a target pressure distribution P^*, the inverse design problem is to find an optimal airfoil shape whose surface pressure distribution P minimizes the cost

$$J = \|P^* - P\|_2^2 \tag{13}$$

In order to solve this inverse design problem using the gappy POD method, the snapshots are first redefined. Rather than containing only flow variables, each snapshot is augmented to also contain airfoil coordinates. For example, consider the augmented snapshot set, $\{V^i\}_{i=1}^m$, where each snapshot contains a surface pressure distribution P^i and corresponding set of airfoil coordinates C^i:

$$V^i = \begin{bmatrix} C^i \\ P^i \end{bmatrix} \tag{14}$$

The target vector $V^* = [C^{*T}\, P^{*T}]^T$ can then be considered as an incomplete data vector, where P^* is known and C^* must be determined. Thus, the gappy POD procedure can be

used to determine the optimal airfoil shape, using the procedure outlined in the previous section and minimizing the cost in (13) with respect to the gappy norm

$$J = \|V^* - \tilde{V}\|_n^2 \tag{15}$$

where n is the mask vector corresponding to V^* and the intermediate repaired vector $\tilde{V}$ is represented by a linear combination of basis vectors as in (7).

The inverse design problem has thus been converted into a problem of reconstructing missing data. In order to determine the airfoil shape, a system of linear equations must be solved, with size equal to the number of POD basis functions. The gappy POD method will then produce not only the optimal airfoil shape, but also the corresponding surface pressure distribution. If further flowfield information is desired, such as pressure distribution off the surface or other flow variables, this data could also be included in the augmented snapshot set.

The POD eigenvalues give guidance as to how many POD eigenfunctions should be included in the basis. Typically, one will include p basis vectors so that the relative energy captured, $\sum_{i=1}^{p} \lambda_i / \sum_{j=1}^{m} \lambda_j$, is greater than some threshold, typically taken to be 99% or higher. This energy measure determines how accurately a snapshot in the original ensemble can be reconstructed using the POD basis; however, it does not give any information regarding the accuracy of reconstructing a new vector. In the inverse design problem stated above, it is therefore important to monitor the value of the cost function J. One may choose enough POD basis functions to capture 99% or more of the snapshot energy, but the optimal value of J remains unacceptably high. This indicates that the subspace spanned by the chosen snapshots is not sufficiently rich to capture the desired design airfoil. Approaches for addressing this issue will be presented in the results section, which now follows.

Results and Discussion

Results will now be presented to demonstrate how the gappy POD method can be used for fluid dynamic applications. We first consider a series of examples which show how the gappy POD can be used to reconstruct incomplete flowfield data. Secondly, we consider the application of the gappy POD to the problem of inverse design. All flow solutions were obtained from an inviscid steady-state CFD code, which uses a finite volume formulation as presented in Jameson et al.[16] A standard C-grid with 192 x 32 elements is used for all examples.

Reconstruction of flowfield data

The case considered is the NACA 0012 airfoil at a freestream Mach number of 0.8. To create the POD basis, 51 snapshots are computed at uniformly spaced values of angle of attack in the interval $\alpha = [-1.25°, 1.25°]$ with a step of 0.05°. The six dominant POD modes contain respectively 85.76%, 9.31%, 3.26%, 1.09%, 0.38%, and 0.13% of the snapshot ensemble energy. An incomplete flowfield was then generated by computing the flow solution at $\alpha = 0.77°$ (which is not one of the snapshots), and then retaining only the pressure values on the surface of the airfoil. The total number of pressure values in the full flowfield is 6369 and the number of pressure values on the airfoils surface is 121, hence 98.1% of the data is missing. The goal, then, is to reconstruct the full pressure flowfield using the gappy POD method. Such a problem might occur, for example, when analyzing experimental data. Typically, experimental measurements will provide only the airfoil surface pressure distribution while the gappy POD method provides a way to combine this experimental data with computational results in order to reconstruct the entire flowfield. Figure 1(a) shows the points on the NACA 0012 airfoil surface where pressure field values are available. Figures 1(b) and (c) show the reconstructed pressure contours with four and six POD modes, respectively, compared with the original contours of the CFD solution. As expected, the more modes used, the more accurate is the reconstruction. With just limited surface pressure data available, the complete pressure field can be determined very accurately with only six POD

modes, showing that the gappy POD methodology for data reconstruction works effectively for an aerodynamic application.

Sensitivity to available data

In the above example, it has been assumed that surface pressure measurements are available over the entire airfoil surface; however, in practice, the number of available surface measurements is limited to only a few points on the surface. Hence, it is important to assess the sensitivity of the reconstruction result to both the quantity and the locations of these available surface data. In order to select a limited number of pressure measurements, a heuristic approach used by Cohen et al.[17] for unsteady flows is used in this study. This approach is based on an analogy with structural sensing, which suggests that sensors should be located in areas of high modal activity. The POD modes of a flow often exhibit sinusoidal spatial variation, and it has been shown that sensors placed at local POD modal minima and maxima yield effective flow sensing results.

This heuristic procedure is applied here to study the sensitivity of the reconstruction results to the amount of available data. An initial configuration of eleven measurement points corresponding to spatial optima of the first POD mode plus a few points near the leading edge, where all POD modes were seen to vary rapidly, is first chosen. Additional measurements are then considered by adding in turn the spatial optima of modes 2, 3, 4, 5 and 6, resulting in configurations with 15, 21, 29, 31 and 39 sensing points respectively. Figure 2 shows the resulting error between the exact and reconstructed pressure field for each of these configurations and compares it with the case where all 121 surface points have been used. The percentage error is defined as

$$e = \frac{||p - \tilde{p}||}{||p||} \times 100\% \tag{16}$$

where p and $\tilde{p}$ are vectors containing the actual and reconstructed pressure measurements, respectively. For all the cases, six POD modes have been used for the reconstruction using

the gappy system. It can be seen that even with a very small number of sensors, the entire flowfield can be reconstructed very accurately, confirming the effectiveness of the heuristic sensor placement algorithm. The figure also shows that subsequent reductions in the error diminishes as higher modes are considered in the sensing. This is consistent with the fact that subsequent modes constitute progressively less of the total energy. The POD eigenvalues can therefore be used not only to select the number of modes in the gappy system, but also to choose an appropriate number of sensors.

Sensitivity to experimental error

Another practical consideration is the effect of experimental error. In the examples above, CFD data have been used to simulate measurements that might be available from an experiment; however, in practice, these measurements are subject to sensor noise. This effect is simulated artificially in this study by adding a random noise component to each CFD pressure measurement. Maximum noise levels, expressed as a percentage of the freestream pressure value, of $\pm 1\%$, 2%, 5%, 10% and 20% are applied. This yields the results shown in Figure 3, which compares the reconstruction error arising from using all 121 measurement points and with 2, 4, 6 and 8 POD modes. For all the cases, the percentage error shown in the plot is averaged over 200 random trials. A slight modification to the final step of the reconstruction algorithm is applied in this investigation. Rather than replacing only the missing pressure values by the corresponding repaired values, the entire pressure field is represented using the POD expansion given in equation (7). This avoids filling the reconstructed flowfield with noisy data and allows the reconstruction process to act as a filter when high noise levels are present.

Figure 3 shows a number of interesting trends. As expected, as the magnitude of the sensor noise increases, the magnitude of the reconstruction error also increases, even though the prediction is still very good with a relatively high noise level of 10%. Inspecting the points corresponding to the nominal case with no sensor noise, it can be seen that as more POD modes are used in the reconstruction, the error decreases; however, once again, subsequent

gains are reduced due to the small amount of additional flow energy associated with higher modes. In particular, it can be seen that there is little improvement in increasing the number of modes from six to eight (as the first six modes already contain 99.9% energy).

In contrast, the result is very different when sensor noise is considered. Figure 3 shows clearly that the sensitivity to noise is increased as the number of POD modes increases. Using more POD modes in the representation allows the flowfield to be represented more accurately, but it also allows a greater degree of "data matching" as the gappy formulation attempts to minimizes the error between the measured and the reconstructed data. If very high levels of noise are present, the reconstruction with a smaller number of modes yields more accurate results. In most fluid applications, the dominant POD modes tend to correspond to flow patterns with low spatial frequency. These modes do not have the resolution to match the high frequency components of the sensor data, causing them to naturally filter sensor noise. While plots such as Figure 3 are highly problem dependent, in practice one should use the POD eigenvalues to select the minimum possible number of POD modes that will achieve the desired resolution in nominal cases. For example, for the problem considered here, the small additional advantage of including modes 7 and 8 should be regarded with extreme caution. Further, if very high levels of sensor noise are expected, consideration should be given to reducing the number of modes beyond what might be chosen under nominal circumstances.

While Figure 3 shows the trends clearly, it does not convey well the absolute quality of the reconstruction, since the percentage error defined in (16) contains many farfield points with a very small error contribution. The pointwise percentage error can be defined for point i as

$$\hat{e}_i = \frac{|p_i - \tilde{p}_i|}{|p_i|} \times 100\% \tag{17}$$

where p_i and $\tilde{p}_i$ are respectively the actual and the reconstructed pressure values at point i. Table 1 shows the average and maximum pointwise errors for the eight-mode case and compares them to the overall percentage error computed using (16). From the table, it can be seen that the average and overall error quantities are very small, even for very high levels

of noise. The maximum pointwise error grows with noise level at roughly the same rate as the average error, but is much larger in magnitude. As might be expected, the largest pointwise errors are observed on and near the airfoil surface. Table 1 also shows the ability of the gappy reconstruction method to act as a filter for noisy data. If the actual measurement data are used in the final reconstructed pressure field (instead of representing all points by the POD expansion), then the maximum pointwise error is, as expected, observed to be of comparable magnitude to the percentage noise level. Table 1 shows that using the reconstructed data, this error is reduced by roughly a factor of two.

Incomplete snapshot set

In the second example, the creation of a set of POD basis vectors from an incomplete set of snapshots is investigated. This problem may again be of interest if partial flowfield data are available from experimental results. Using the gappy POD methodology, experimental and computational data with differing levels of resolution can be combined effectively to determine dominant flow modes.

Again, we consider the NACA 0012 airfoil at a freestream Mach number of 0.8. A 26-member snapshot ensemble is used, corresponding to steady pressure solutions at angles of attack in the range $\alpha = [0^\circ, 1.25^\circ]$, uniformly spaced with an interval of 0.05°. To create the incomplete snapshot set for this example, 30% of the pressure data of each snapshot is discarded randomly. The algorithm described in the theory is then used to repair the data as follows. By first repairing the missing data points in each snapshot with the average over available data at that point, a new ensemble of data is created that has no missing values. With this new ensemble, a first approximation to the POD basis is then constructed. Then, each snapshot in the ensemble is repaired using the first approximation of the POD basis. This repaired ensemble is then used to construct a second approximation to the POD basis. For the example in this section, the iterative procedure above is stopped after 50 iterations.

In Figure 4, the second snapshot with 30% data missing is repaired by the above procedure with five POD modes, which contain 99.99% of the flow energy. Figure 4(a) shows the

original damaged snapshot. After one iteration, the repaired snapshot in Figure 4(b) begins to resemble the CFD solution; however, a large error remains. Figure 4(c) shows the repaired snapshot after 25 iterations and can be seen to match closely with the original. Figure 5 shows the repairing process for the 23^{rd} snapshot. Compared to the contours with 30% data missing in Figure 5(a), the reconstruction in Figure 5(c) is already close to the CFD result with only seven iterations. It can be seen that the convergence of the reconstruction process depends on the details of the particular snapshot under consideration. In particular, it depends on the structure of the flow snapshot and how much data is missing. For the 23^{rd} snapshot shown in Figure 5, the convergence rate is much faster than for the second snapshot shown in Figure 4.

The convergence of the POD eigenvalue spectrum of the incomplete ensemble is shown in Figure 6. It can be seen that after one iteration the first two eigenvalues have converged, while convergence of subsequent eigenvalues requires more iterations. For example, after 45 iterations, it can be seen that only the first five eigenvalues have converged; however, these five modes account for almost all of the flow energy (99.99%).

The rate of convergence for the construction of the POD modes is also investigated using this incomplete snapshot set. This is done by comparing the POD basis vectors constructed using a complete data set with those constructed with the incomplete set. In Figure 7(a), it can be seen that, after just two iterations, the first mode constructed with incomplete data matches well with the desired result. However, in Figure 7(b), the construction of the second mode after ten iterations still shows significant deviation from the actual mode. At least 31 iterations are needed to obtain the second mode accurately. Figure 7(c) shows that after 50 iterations the estimate of the third POD mode is reasonable but not fully converged.

It is observed that the more energetic a POD mode, the faster the rate of construction from a given partial data set. The first mode, which captures 90.65% energy, requires only two iterations to converge very close to the desired result; however, the second and the third modes, which capture 7.7% and 1.4%, require 31 and 50 iterations, respectively. This trend

is observed for subsequent modes and may be related to the fact that successive POD modes often correspond to higher spatial frequency flow structures. For example, it can be seen in Figure 7 that the second and third modes contain progressively more high-frequency shock structure. It is therefore not surprising that construction of higher modes requires more iterations.

Finally, we comment on the computational cost of the iterative gappy POD procedure. At each iteration, in order to reconstruct m snapshots, m systems of the form (9) must be solved. Furthermore, at each iteration, an eigenvalue problem of size m must be solved to determine the POD modes. Using a 1.6 GHz Pentium 4 personal computer, with six POD modes each iteration took less than two seconds.

Inverse Design Problem

The final set of examples demonstrates how the gappy POD method can be applied to the problem of inverse airfoil design. A collection of snapshots is generated as in (14) by choosing a set of airfoil shapes and obtaining their corresponding surface pressure distributions. (Other aspects of the flowfield could also be included if they are of interest.) While in this paper, CFD results were used to create the snapshots, in practice, the flow data could be obtained from computational simulations, experimental results, tabulated data, or a combination thereof. The goal, then, is to use the gappy POD method to determine the optimal airfoil shape that produces a given target pressure distribution, which is not contained in the snapshot collection.

Subsonic Flow

Snapshots for subsonic airfoil design are created by considering the RAE 2822 airfoil and adding a series of Hicks-Henne bump functions,[13] which make smooth changes in the geometry. The Hicks-Henne bump functions are given by

$$y(x) = A\left\{\sin\left[\pi x^{log(1/2)/log(t_1)}\right]\right\}^{t_2}, \quad 0 \le x \le 1 \tag{18}$$

where A is the maximum bump magnitude, t_2 is used to control the width of the bump and $x = t_1$ is the location of the maximum of the bump. Thirty one bump functions with $t_2 = 4, A = 0.005$ were added to each of the upper and lower surfaces of the RAE 2822 airfoil to create a total of 63 airfoil snapshots, some of which are shown in Figure 8(a). The flow solutions for these airfoils were computed using the CFD model with zero angle of attack and a freestream Mach number of 0.5.

The pressure distribution for the Korn airfoil, whose geometry is also shown in Figure 8(a), is specified as the first design target. It can be seen in Figure 8(a) that, while the Korn airfoil shares some similarities with the RAE 2822-based snapshot set, its camber and thickness distribution are quite distinct. This example thus represents a challenge for the gappy POD inverse design methodology. The first 32 POD eigenvalues corresponding to the snapshot set are shown in Figure 8(b). It can be seen that the first 21 POD modes contain 99% of the system energy. Figure 9(a) shows the points on the Korn airfoil where target pressure values are specified. Using the gappy POD procedure, the corresponding optimal airfoil shape can then be determined. Figures 9(b), (c) and (d) compare the exact Korn airfoil and the target pressure to the POD design results using one, 15 and 29 modes, respectively. It can be seen that as the number of modes is increased, the predicted shape and its pressure distribution agree more closely with the exact solutions. The corresponding values of the cost J are given in Table 2. Using 29 POD modes, which accounts for 99.97% of the snapshot energy, it can be seen that the error in the pressure distribution is very small.

Using the same ensemble of snapshots as in the previous case, three different target pressure distributions were considered, corresponding to the NACA 63212, Quabeck 2.0/10 R/C sailplane HQ 2010, and GOE 117 airfoils. The resulting inverse design results are shown in Figure 10. Although 29 POD modes are used for the NACA 63212 in Figure 10(a), there is still a small region on the upper surface near the leading edge which cannot be resolved accurately. The situation is worse for the HQ 2010 airfoil. As shown in Figure 10(b), there are some regions on the upper and lower surfaces need to be improved. In Figure

10(c), the design airfoil is still far away from the exact GOE 117 airfoil, indicating that this geometry, which differs considerably from the baseline RAE 2822 airfoil, is not contained in the subspace spanned by the snapshot set considered. From Table 2, the values of the cost J can be seen to be larger than for the Korn airfoil, especially in the case of the HQ 2010 and GOE 117 airfoils. Therefore, a way to improve the design airfoil needs to be developed.

One approach to improve the inverse design results is to increase the richness of the subspace spanned by the POD basis vectors. This can be achieved by including more snapshots in the ensemble. The exact airfoil could be obtained if further snapshots were added to the set; however, this implies some a priori knowledge of the desired result so that appropriate snapshots may be chosen. A better way to improve the design airfoil is proposed in Legresley and Alonso,[2] where an available design airfoil at some iteration is used as an intermediate baseline, to which some bump functions are added to generate a new set of snapshots. This new snapshot collection is then used to compute a new set of POD modes and thus restart the design procedure. Here, we make use of a similar method for the gappy POD procedure. For the case of the NACA 63212 above, the design airfoil with 29 POD modes is used as an intermediate baseline airfoil, to which 60 bump functions with $t_2 = 4, A = 0.005$ are added to obtain a new collection of 61 snapshots. A similar procedure is used for the intermediate design airfoils shown in Figure 10 for the HQ 2010 and GOE 117 cases, respectively. It can be seen in Figure 11 that the design airfoils now match the exact airfoils very well. In order to obtain a cost on the order 10^{-4}, 43, 32 and 39 POD modes were required for the NACA 63212, HQ 2010, and GOE 117 airfoils, respectively. The costs are shown in Table 2. The results from this restarted gappy POD procedure are much better than those obtained from using the 60-snapshot ensemble above. Moreover, by allowing multiple restarts, this procedure enables the consideration of an inverse design whose geometry differs significantly from that of the baseline airfoil.

Transonic Flow

All the examples in the previous section are at a freestream Mach number of $M = 0.5$, hence the flows are shock free. It is expected that in the transonic regime, the inverse design will be more difficult due to the presence of shocks.

The initial snapshot ensemble from the previous section is used, consisting of 63 airfoils based on the RAE 2822 airfoil; however, the flow solutions are now computed at a freestream Mach number of $M = 0.8$. Figures 12 (a) and (b) show the results for an inverse design target corresponding to the Korn airfoil pressure distribution at $M = 0.8$. It can be seen that, in the transonic case, even when 36 POD modes, which capture 99.997% of the energy, are used, the lower surface cannot be resolved accurately and the cost J of 0.0230 is still large. However, the design airfoil can be improved by restarting the gappy POD method as described in the previous section. The design airfoil with 36 POD modes is now used as the intermediate baseline and 60 bump functions with $t_2 = 4, A = 0.001$ are added to create a new set of snapshots. Figures 12(c) and (d) show the design airfoil using POD modes created from the new ensemble. The design airfoil with 20 POD modes is significantly improved in Figure 12(c), while using 29 modes allows the Korn airfoil to be captured almost exactly as shown in Figure 12(d). This example shows that the design of a transonic airfoil is more difficult than a subsonic airfoil. The restarted gappy POD procedure can be used to obtain more accurate results, i.e. smaller J; however, as more restarts are used, the design procedure becomes more expensive.

In general, several restarts of the gappy POD could be required in order to obtain the desired target pressure distribution, especially if the design airfoil is significantly different from those included in the snapshot set, or if significant nonlinearities exist in the flow. For each restart, the CFD solver must be used to obtain the flow solutions corresponding to the new set of parameterized airfoils - this is the most expensive part of the computation. One could also utilize different approaches to generate the new airfoil set. For example, the parameters A, t_1 and t_2 for the bump functions could be varied. Considering more

bump functions would result in a larger snapshot set, but may reduce the number of restarts required.

Conclusion

For the first time, the gappy POD has been applied to a number of steady aerodynamic applications. In particular, the method has been shown to be very effective for reconstructing flow fields from incomplete aerodynamic data sets. This approach is useful for many real applications where there may be a need to combine experimental and computational results. A heuristic approach of placing sensors at the locations of POD modal optima has been shown to be very effective at yielding accurate reconstruction results with low sensitivity to the number of available measurement points. While results are problem dependent, a general strategy for reducing the sensitivity of the reconstruction results to experimental error has also been proposed. In the near future, based on this initial investigation, ongoing research will address the question of developing improved sensor placement strategies in the context of gappy framework, and the extension of the reconstruction approach to unsteady flows. A new extension of the methodology has also been proposed for the inverse design of airfoil shapes. Given a database of airfoil shapes and pressure distributions, it has been shown that the gappy POD approach can be used to design an airfoil to match a specified pressure target. Even when the target airfoil is different to those contained within the original database, a systematic restart procedure can be used to improve the accuracy of the design results.

References

[1]Holmes, P., Lumley, J., and Berkooz, G., *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, UK, 1996.

[2]LeGresley, P. and Alonso, J., "Investigation of Non-Linear Projection for POD Based Reduced Order Models for Aerodynamics," AIAA Paper 2001-0926, presented at 39th Aerospace Sciences Meeting and Exhibit, Reno, NV, 2001.

[3]Sirovich, L. and Kirby, M., "Low-dimensional procedure for the characterization of human faces." Jounal of the Optical Society of America A, 4(3):519-524, 1987.

[4]Fukanaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1972.

[5]Sirovich, L., "Turbulence and the Dynamics of Coherent Structures. Part 1 : Coherent Structures," *Quarterly of Applied Mathematics*, Vol. 45, No. 3, October 1987, pp. 561–571.

[6]Dowell, E., Hall, K., Thomas, J., Florea, R., Epureanu, B., and Heeg, J., "Reduced Order Models in Unsteady Aerodynamics," AIAA Paper 99-1261, 1999.

[7]Hall, K. C., Thomas, J. P., and Dowell, E. H., "Reduced-Order Modeling of Unsteady Small-Disturbance Flows Using a Frequency-Domain Proper Orthogonal Decomposition Technique," AIAA Paper 99-0655, 1999.

[8]Romanowski, M., "Reduced Order Unsteady Aerodynamic and Aeroelastic Models using Karhunen-Loéve Eigenmodes," AIAA Paper 96-194, 1996.

[9]Beran, P. and Silva, W., "Reduced-Order Modeling: New Approaches for Computational Physics," AIAA Paper 2001-0853, 2001.

[10]Everson, R. and Sirovich, L., "The Karhunen-Loeve Procedure for Gappy Data," J.Opt.Soc.Am., 12: 1657-1664, 1995.

[11]Lighthill, M., *A New Method of Two-dimensional Aerodynamic Design*, Aeronautical Research Council, United Kingdom, 1945.

[12]McFadden, G., "An Artificial Viscosity Method for the Design of Supercritical Airfoils," New York University report No. C00-3077-158, 1979.

[13]Hicks, R. and Henne, P., "Wing Design by Numerical Optimization," *Journal of Aircraft*, Vol. 15, 1978, pp. 407–412.

[14]Jameson, A., "Aerodynamic Design via Control Theory," *Journal of Scientific Computing*, Vol. 2, 1988, pp. 233–260.

[15]B. I. Epureanu, Dowell, E. H. and Hall, K., ""A Parametric Analysis of Reduced Order

Models of Potential Flows in Turbomachinery Using Proper Orthogonal Decomposition," 2001-GT-0434, Proceedings of ASME TURBO EXPO 2001, New Orleans, Louisiana, June 2001.

[16]Jameson, A., Schmidt, W., and Turkel, E., "Numerical Solutions of the Euler Equations by Finite Volume Methods Using Runge-Kutta Time-Stepping Schemes," AIAA Paper 81-1259 14th Fluid and Plasma Dynamics Conference, Palo Alto, California, 1981.

[17]Cohen, K., Siegel, S., and McLaughlin, T., "Sensor placement based on proper orthogonal decomposition modeling of a cylinder wake," AIAA Paper 2003-4259, 2003.

List of Tables

1 Pressure reconstruction errors for various levels of sensor noise with eight POD modes in the reconstruction. Shown are percentages for the average pointwise error, the maximum pointwise error, and the 2-norm of the overall relative error. 24
2 Optimal cost versus number of POD modes for subsonic inverse design cases. 25

Noise level (%)	average $\hat{e}$ (%)	max $\hat{e}$ (%)	e (%)
1	0.03	0.60	0.0814
2	0.06	1.01	0.15
5	0.1437	2.4228	0.355
10	0.2867	4.5145	0.7115
20	0.5715	9.7506	1.4058

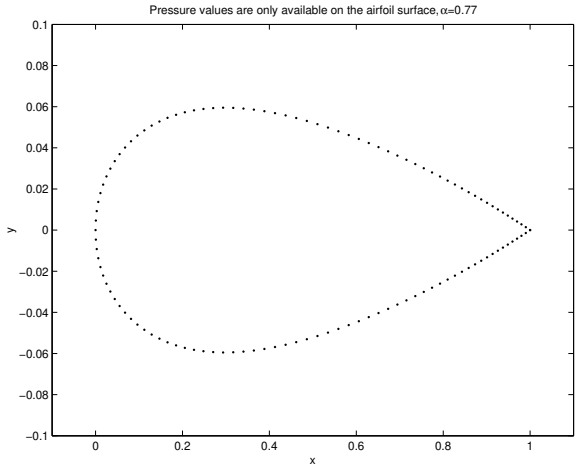
Table 1 Pressure reconstruction errors for various levels of sensor noise with eight POD modes in the reconstruction. Shown are percentages for the average pointwise error, the maximum pointwise error, and the 2-norm of the overall relative error.

Airfoil	Number of POD modes	Cost, J
Korn	1	0.022
Korn	15	0.0047
Korn	29	2.9426e-004
NACA 63212	29	6.2673e-004
HQ 2010	29	0.0061
GOE 117	29	0.0056
NACA 63212 with restart	43	1.7435e-004
HQ 2010 with restart	32	2.0622e-004
GOE 117 with restart	39	6.8127e-004

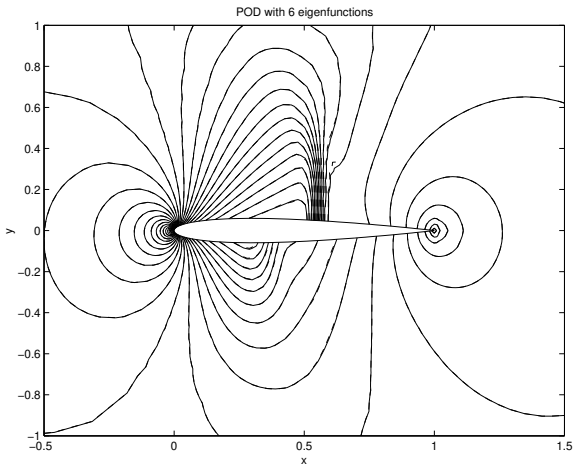
Table 2 Optimal cost versus number of POD modes for subsonic inverse design cases.

List of Figures

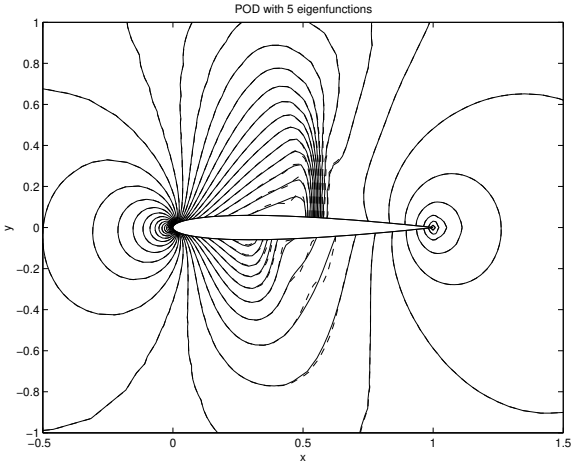
1 The reconstruction of the pressure field from the airfoil surface pressure distribution (dash), compared with the original contours (solid). 27
2 Percentage pressure reconstruction error versus the number of available surface measurement points. Six POD modes were used in the reconstruction. . 28
3 Percentage pressure reconstruction error versus experimental error in surface pressure measurements. The noise level in the measurements is given as a percentage of the freestream pressure value. 29
4 Reconstruction of the second snapshot (dash), compared with the original contours (solid). 30
5 Reconstruction of the 23^{rd} snapshot (dash), compared with the original contours (solid). 31
6 The eigenvalue spectrum for construction of the POD basis from an incomplete snapshot set. Shown are the POD eigenvalues at various stages in the iterative process. 32
7 Construction of POD modes from an incomplete snapshot set. 33
8 The airfoil snapshots and corresponding POD eigenvalues. 34
9 Inverse design of the Korn airfoil using gappy POD. $M = 0.5$, snapshots based on RAE 2822. 35
10 Inverse design of the NACA 63212, HQ 2010 and GOE 117 airfoils using gappy POD. $M = 0.5$, snapshots based on RAE 2822. 36
11 Inverse design of the NACA 63212, HQ 2010 and GOE 117 airfoils using restarted gappy POD. $M = 0.5$, original snapshots based on RAE 2822, restarted snapshots based on intermediate airfoils shown in Figure 10. 37
12 Inverse design of the Korn airfoil using gappy POD with and without restarts. $M = 0.8$. 38



a) The points on the airfoil where pressure values are available.



b) The reconstruction with four POD modes.



c) The reconstruction with six POD modes.

Figure 1 The reconstruction of the pressure field from the airfoil surface pressure distribution (dash), compared with the original contours (solid).

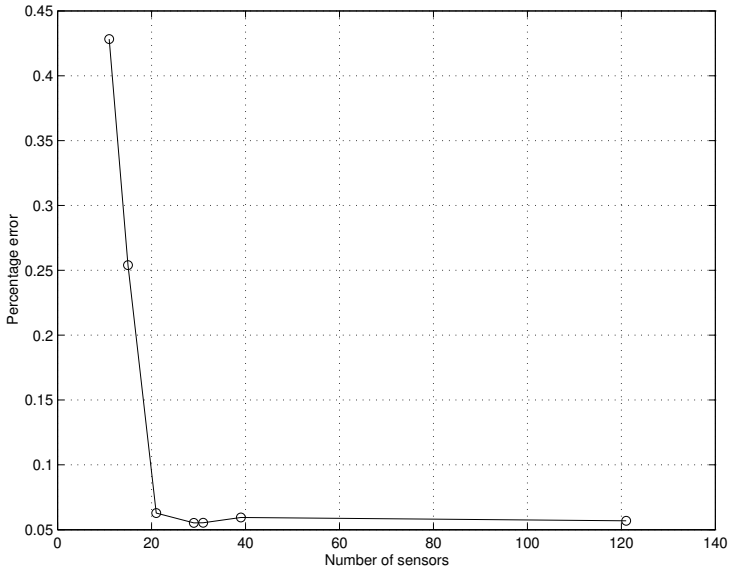


Figure 2 Percentage pressure reconstruction error versus the number of available surface measurement points. Six POD modes were used in the reconstruction.

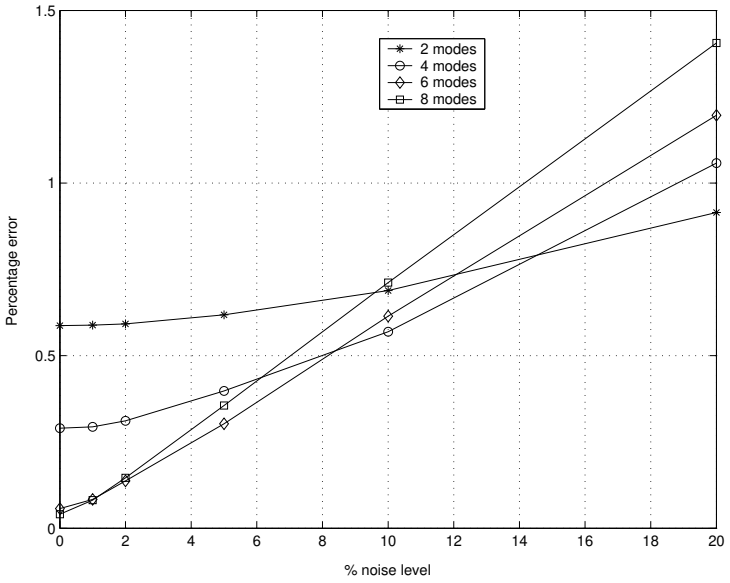
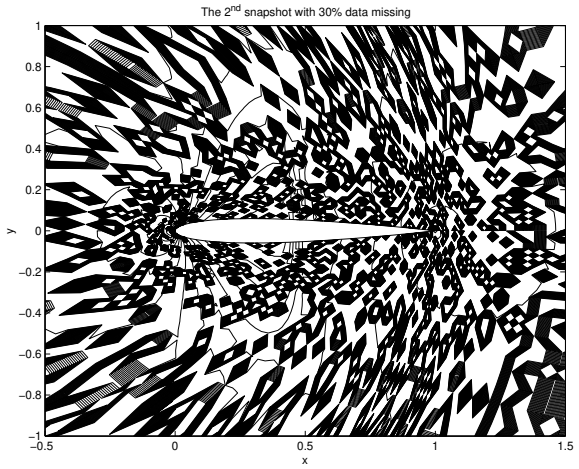
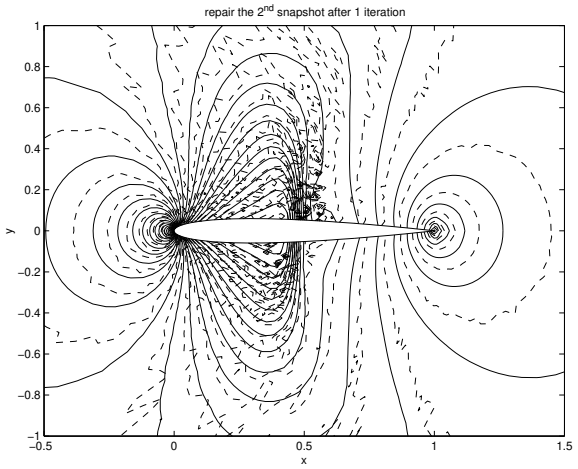


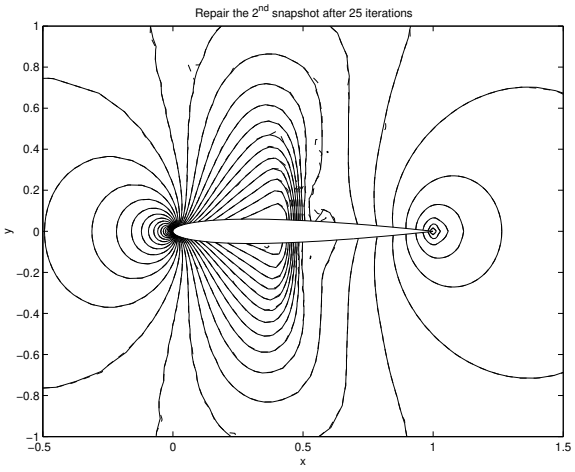
Figure 3 Percentage pressure reconstruction error versus experimental error in surface pressure measurements. The noise level in the measurements is given as a percentage of the freestream pressure value.



a) The second snapshot with 30% data missing.

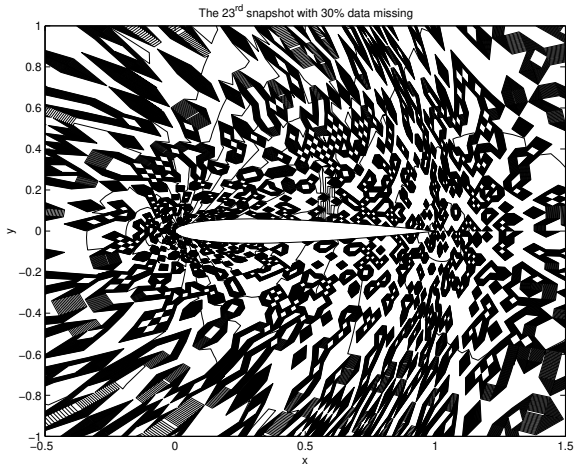


b) Reconstruction of the second snapshot after one iteration.

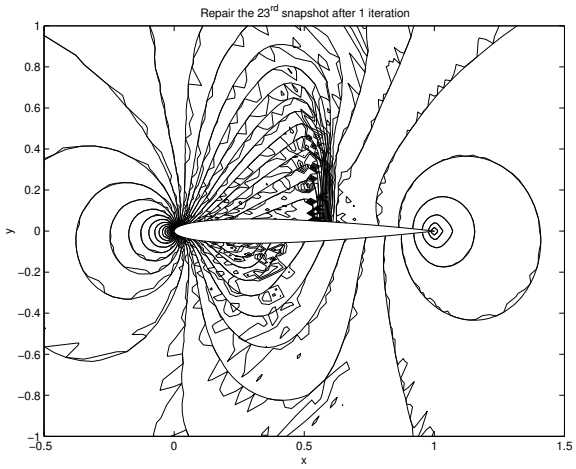


c) Reconstruction of the second snapshot after 25 iterations.

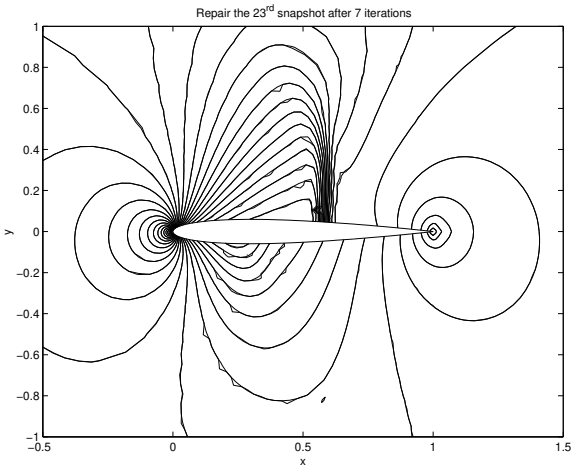
Figure 4 Reconstruction of the second snapshot (dash), compared with the original contours (solid).



a) The 23^{rd} snapshot with 30% data missing.



b) Reconstruction of 23^{rd} snapshot after one iteration.



c) Reconstruction of 23^{nd} snapshot after seven iterations.

Figure 5 Reconstruction of the 23^{rd} snapshot (dash), compared with the original contours (solid).

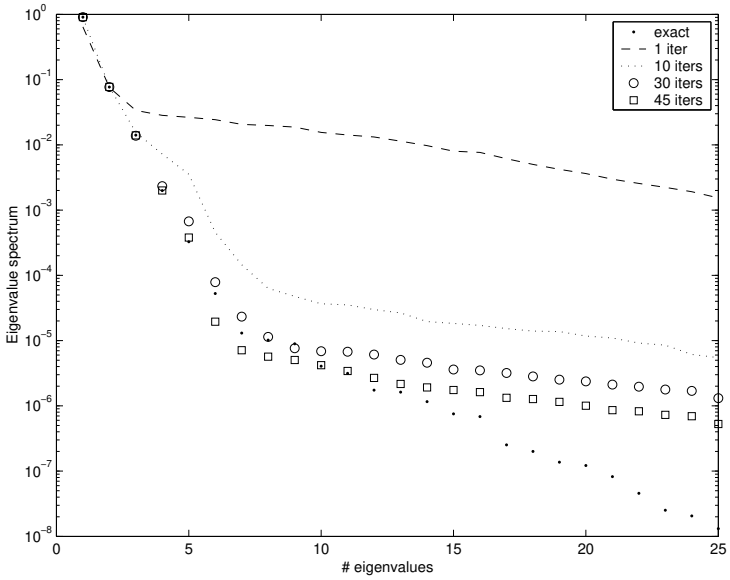
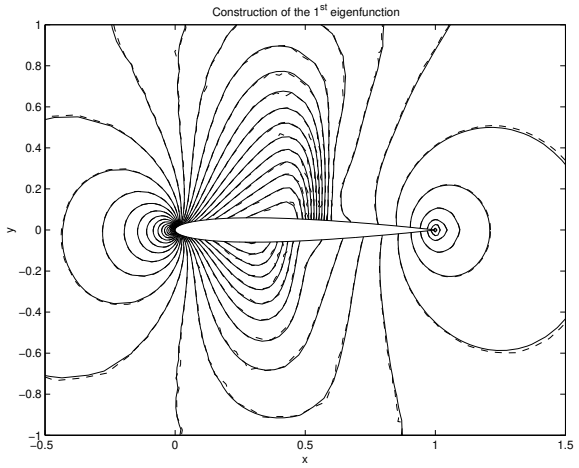
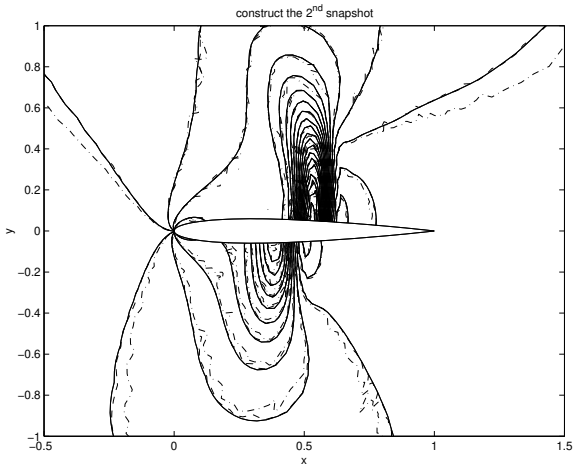


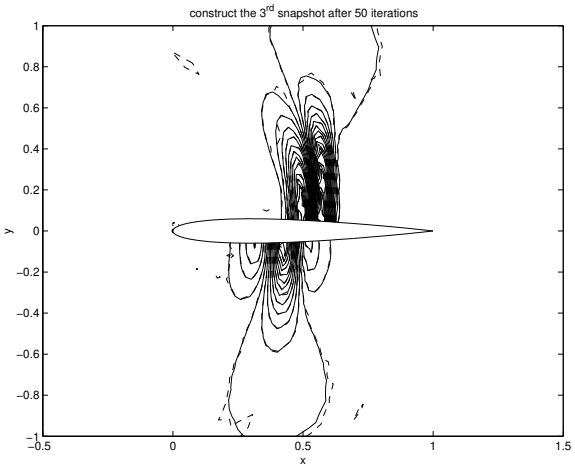
Figure 6 The eigenvalue spectrum for construction of the POD basis from an incomplete snapshot set. Shown are the POD eigenvalues at various stages in the iterative process.



a) Construction of the first mode after two iterations (dash) compared with the exact contours (solid).

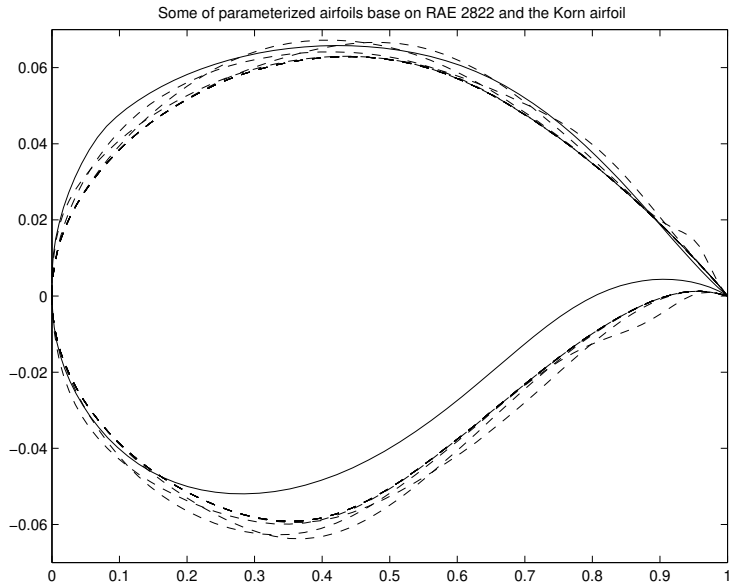


b) Construction of the second mode after ten iterations (dot-dash), after 31 iterations (dash), and the exact contours (solid).

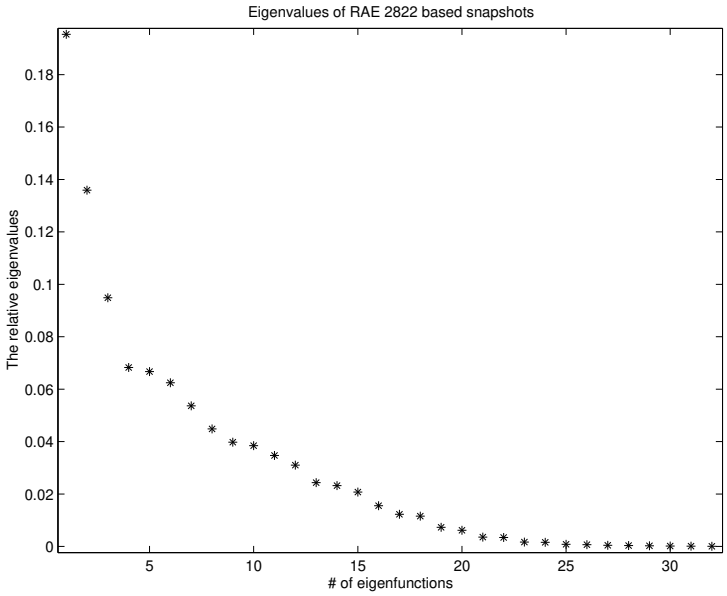


c) Construction of the third mode after 50 iterations (dash) compared with the exact contours (solid).

Figure 7 Construction of POD modes from an incomplete snapshot set.

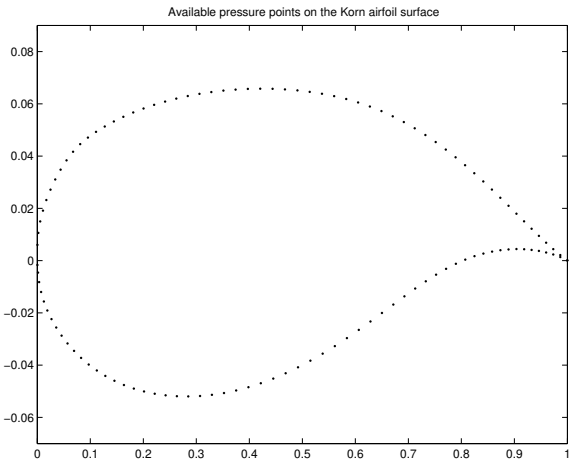


a) Parameterized airfoils based on RAE 2822 (dash) and the Korn airfoil (solid).

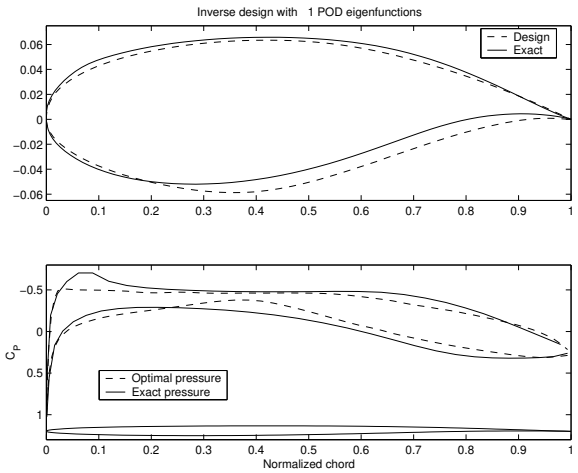


b) The relative eigenvalues from RAE 2822 based snapshots.

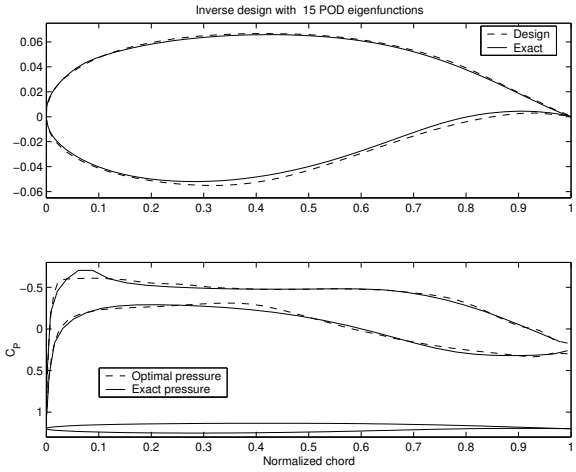
Figure 8 The airfoil snapshots and corresponding POD eigenvalues.



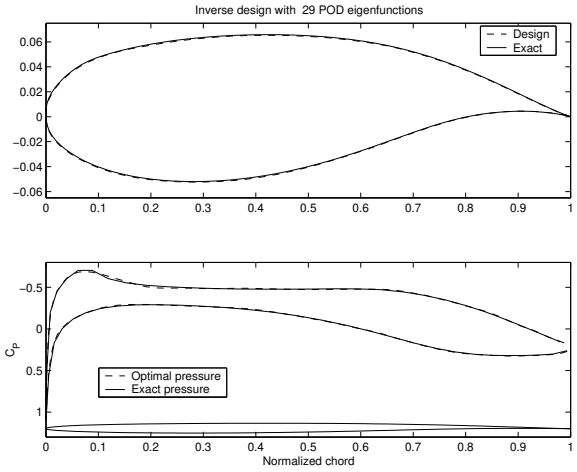
a) The available pressure points on the surface of the Korn airfoil.



b) The exact Korn airfoil (solid) and the design airfoil (dash) with one mode.

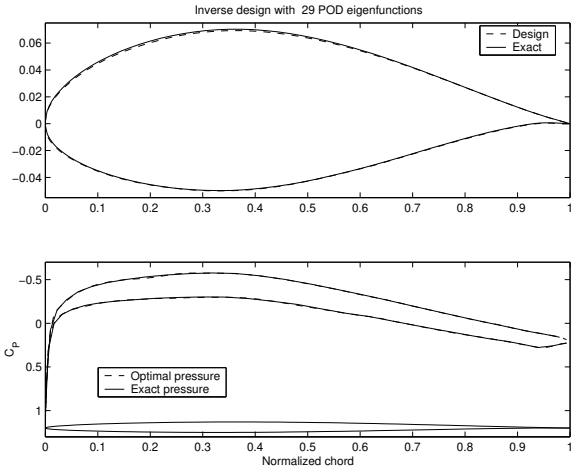


c) The exact Korn airfoil (solid) and the design airfoil (dash) with 15 modes.

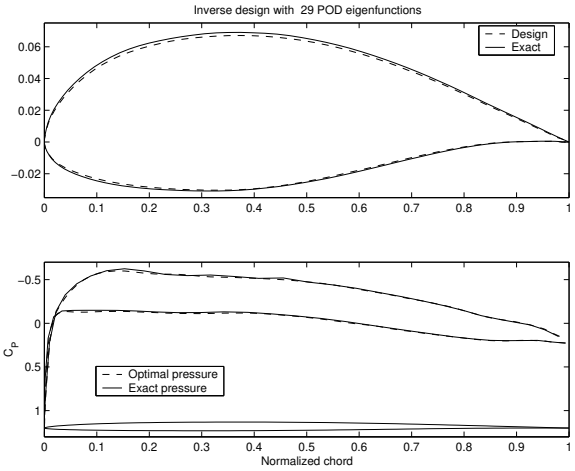


d) The exact Korn airfoil (solid) and the design airfoil (dash) with 29 modes.

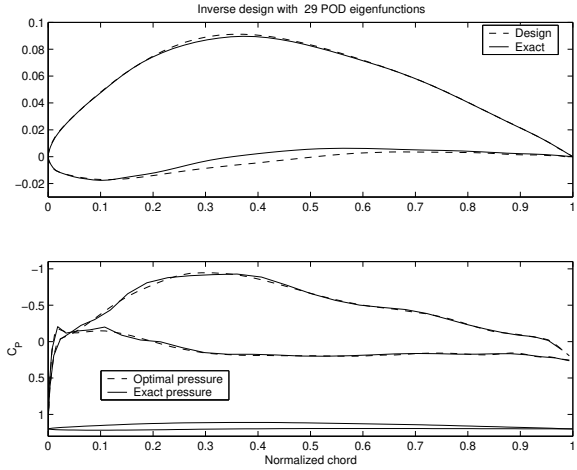
Figure 9 Inverse design of the Korn airfoil using gappy POD. $M = 0.5$**, snapshots based on RAE 2822.**



a) The exact NACA 63212 airfoil (solid) and the design airfoil (dash) with 29 modes.

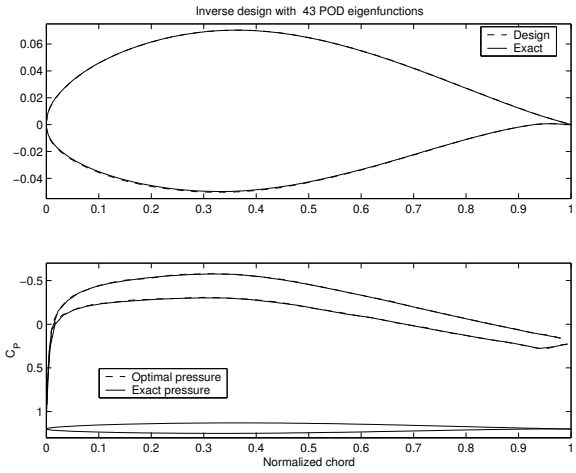


b) The exact HQ 2010 airfoil (solid) and the design airfoil (dash) with 29 modes.

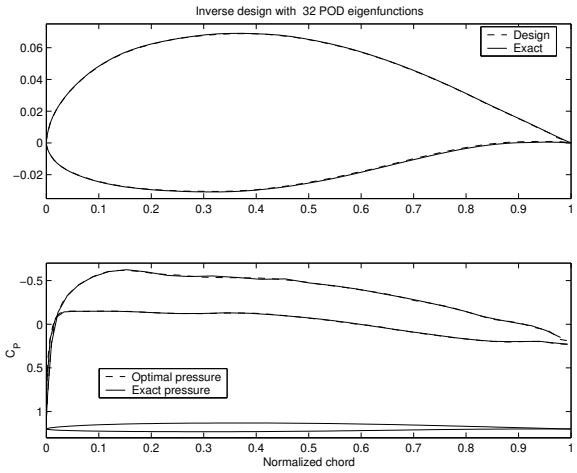


c) The exact GOE 117 airfoil (solid) and the design airfoil (dash) with 29 modes.

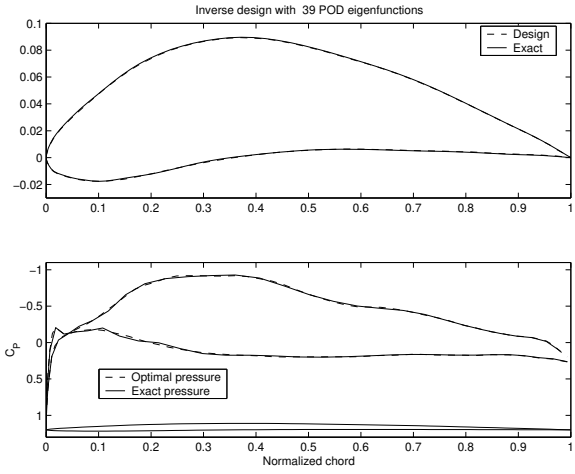
Figure 10 Inverse design of the NACA 63212, HQ 2010 and GOE 117 airfoils using gappy POD. $M = 0.5$, snapshots based on RAE 2822.



a) The exact NACA 63212 airfoil (solid) and the design airfoil (dash) with 43 modes.

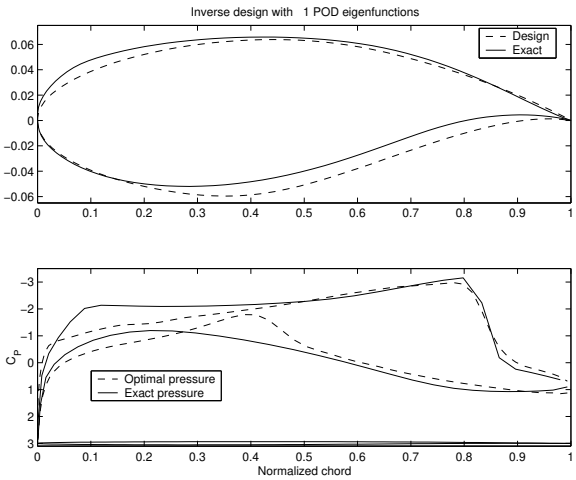


b) The exact HQ 2010 airfoil (solid) and the design airfoil (dash) with 29 modes.

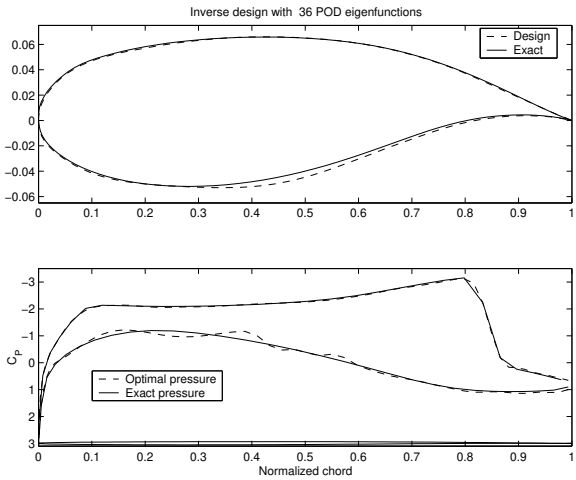


c) The exact GOE 117 airfoil (solid) and the design airfoil (dash) with 34 modes.

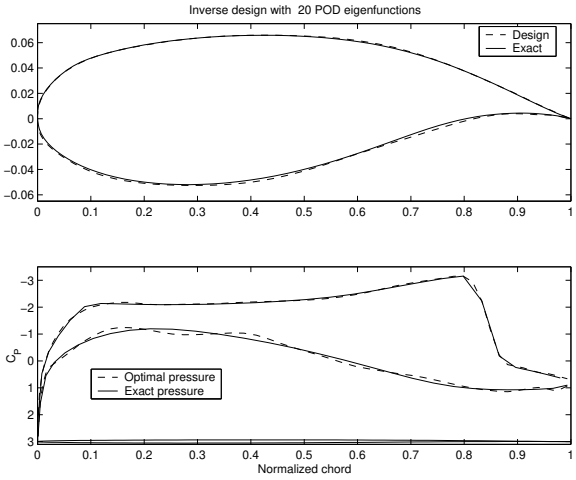
Figure 11 Inverse design of the NACA 63212, HQ 2010 and GOE 117 airfoils using restarted gappy POD. $M = 0.5$**, original snapshots based on RAE 2822, restarted snapshots based on intermediate airfoils shown in Figure 10.**



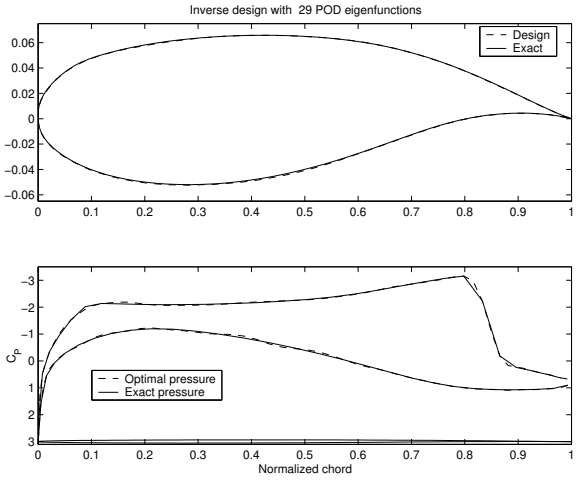
a) The exact Korn airfoil (solid) and the design airfoil (dash) with one mode.



b) The exact Korn airfoil (solid) and the design airfoil (dash) with 36 modes.



c) The exact Korn airfoil (solid) and the design airfoil (dash) with 20 restarted modes.



d) The exact Korn airfoil (solid) and the design airfoil (dash) with 29 restarted modes.

Figure 12 Inverse design of the Korn airfoil using gappy POD with and without restarts. $M = 0.8$**.**