Analytical and Numerical Methods in Mechanical Design
Vol.1, Issue 1, Winter & Spring 2022, 1- 8

Reconstruction of Gappy Data for Cavity Flow Using Gappy Proper Orthogonal Decomposition

Matin Hoseini [1], Nader Montazerin[2], Ghasem Akbari [3,*]

[1]*Department of Mechanical Engineering, Amirkabir University of Technology, Tehran, Iran*
[2] *Department of Mechanical Engineering, Amirkabir University of Technology, Tehran, Iran*
[3]*Department of Mechanical Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran*
Received: 07 February 2022- Accepted: 11 May 2022
**Corresponding author: gh.akbari@iau.ac.ir*

Abstract

The present study examines the possibility of completing gappy flow fields with the method of gappy proper orthogonal decomposition (GPOD). The procedure is performed on a numerically simulated cavity flow. The DNS data is artificially made incomplete by randomly omitting localized data. Two levels of gappiness are examined to evaluate the GPOD procedure. The results indicate that the relative error between the GPOD estimation and the real field directly depends on the level of gappiness. As the gappiness increases, the prediction accuracy decreases. It is shown that the relative error does not monotonically decrease because of inherent noise in higher-level modes of energy. The optimal count of modes in GPOD procedure is obtained and discussed. The contribution of GPOD procedure on spatial experimental data is also addressed.

***Keywords*:** Proper Orthogonal Decomposition, Cavity Flow, Gappy Data, Reconstruction, Energy mode

1. Introduction

Proper Orthogonal Decomposition (POD) is a method for producing reduced order representations of high order systems. The proper orthogonal decomposition was first suggested by Lumley [1] for fluid mechanics applications. This method is developed by Sirovich, the so-called 'snapshot POD' theory, in the context of marred faces reconstruction [2]. It uses average mode of existing patterns to reconstruct the target pattern.

The POD method is also useful in reconstructing missing data from incomplete data sets (gappy POD approach) [2, 3]. This approach results in superior estimations of the missing data, even when compared to spatial interpolation techniques such as kriging [4-6]. However, the gappy POD methods are faced with several practical limitations. For instance when applied to experimental data, where the true field is unknown a priori, it is difficult to determine a convergence criterion based on the Eigen spectrum of the modes. Moreover, the convergence of the reconstruction is on the basis of the whole field of data, without any consideration to the spatial variations within the field [3].

Implementation of gappy POD theory on the experimental data requires sufficient knowledge on the procedure as well as the optimal count of modes for reconstruction of velocity snapshots. The present study implements gappy POD approach on numerically simulated cavity flow to provide such knowledge. Since the DNS data is complete, some localized data are randomly removed to make a gappy data. The optimum number of modes in gappy POD approach are then calculated and discussed.

The article is organized as follows. Section 2 describes the cavity flow simulation procedure and its considerations. This follows in section 3 by introducing the classical and gappy proper orthogonal decomposition theory and their implementation on the velocity data. Section 4 discusses the results of GPOD approach on the DNS data of cavity flow and suggests the optimal counts for GPOD modes.

2. Direct numerical simulation of cavity flow

It is intended to accurately simulate the transient flow in a planar cavity. The describing flow equations are the continuity and Navier–Stokes equations, for conservation of mass and momentum, respectively. Assuming the properties of fluid to be constant, the non-dimensional form of these equations are expressed as:

$$\nabla \cdot \mathbf{V}^* = 0 , \tag{1}$$

$$\frac{\partial \mathbf{V}^*}{\partial t^*} + \mathbf{V}^* \cdot \nabla \mathbf{V}^* = -\nabla p^* + \frac{1}{\mathrm{Re}} \nabla^2 \mathbf{V}^* \tag{2}$$

where $\mathbf{V}^* = \mathbf{V}/U$ is the normalized velocity vector, $p^* = p/\rho U^2$ is the normalized pressure, $t^* = tU/L$ and $\mathrm{Re} = UL/\nu$ is the Reynolds number. L is the characteristic length scale of cavity and ρ and ν are the fluid density and kinematic viscosity, respectively.

The flow domain is a $L \times L$ square box, that its respective non-dimensional size is 1×1. The upper boundary is horizontally moving from right to left, with a speed U. The initial and boundary conditions for the normalized equations (1) and (2) are considered as presented in table 1.

The finite difference decomposition is used to simulate the flow field and the projection approach is selected as a time-accurate procedure. It is capable to overcome some difficulties associated with velocity-pressure formulation of Navier-Stokes equations [7, 8]. The order of accuracy of this approach is first-order in time and second-order in space.

Table 1. Initial and boundary conditions for the considered cavity flow.

Initial/Boundary Condition	Location/Explanation
$\mathbf{V}^* = 0$	for stationary boundaries
$\mathbf{V}^* = [1,0,0]^t$	velocity components of moving plate at the top of cavity
$\mathbf{V}^*(\mathbf{X}, t=0) = 0$	initial condition; stationary fluid at the origin.
$\frac{\partial p^*}{\partial n} = 0$	for all boundaries in normal direction (n)

The differential equations plus boundary/initial conditions are discretized on a staggered grid with dimension 128×128. The numerical error criterion in solving the poison equation in each instance is considered to be 1×10^{-6}. The grid dependency test is successfully performed and it is ensured that the implemented grid is sufficiently fine to generate accurate temporal fields.

3. Proper Orthogonal Decomposition

3.1 Simple POD theory on DNS data

Each instantaneous DNS data can be considered as a snapshot of flow field. Consider N snapshot of instantaneous velocity fields $\{ V^{*i} : i = 1,2,...,N \}$. All velocity component of these N snapshots are arranged in a matrix $\mathbf{V}^*$ as

$$\mathbf{V}^* = [V^{*1} V^{*2} \cdots V^{*N}] = \begin{bmatrix} V^{*1}_{1} & V^{*2}_{1} & \cdots & V^{*N}_{1} \\ V^{*1}_{2} & V^{*2}_{2} & \cdots & V^{*N}_{2} \\ \vdots & \vdots & \ddots & \vdots \\ V^{*1}_{M} & V^{*2}_{M} & \cdots & V^{*N}_{M} \end{bmatrix} \tag{3}$$

where *M* is the number of positions of velocity vectors in a given snapshot. The auto covariance matrix (**C**) is defined as

$$C = V^{*T} V^{*} \tag{4}$$

and the corresponding eigenvalue problem

$$CA^{i} = \lambda^{i} A^{i} \tag{5}$$

can then be solved. The solutions are ordered according to the size of eigenvalues

$$\lambda^{1} > \lambda^{2} > \cdots > \lambda^{N} = 0. \tag{6}$$

The eigenvectors of **C** make up a basis for constructing the POD modes, φ^{i} ,

$$\varphi^{i} = \sum_{r=1}^{N} A_{r}^{i} V^{*n} \Big/ \left\| \sum_{r=1}^{N} A_{r}^{i} V^{*n} \right\| \tag{7}$$

where A_{r}^{i} is the *r*th component of the eigenvector corresponding to λ^{i} from **C** and the discrete 2-norm is defined as

$$\| y \| = \sqrt{y_1^2 + y_2^2 + \cdots + y_M^2} \tag{8}$$

Each snapshot can be expanded in a series of the POD modes with expansion coefficients a_i for each POD mode *i*. The coefficients, also called POD coefficients, are determined by projecting the fluctuating part of the velocity field onto the POD modes

$$a^{n} = \psi^{T} V^{*n} \tag{9}$$

where

$$\psi = [\varphi^{1} \varphi^{2} \cdots \varphi^{N}] \tag{10}$$

The expansion of the velocity data of snapshot number *n* is

$$V^{*n} = \sum_{i=1}^{N} a_{i}^{n} \varphi^{i} = \psi a^{n} \tag{11}$$

The magnitude of the *i*th eigenvalue, λ_i , describes the relative importance of the *i*th POD basis vector. This importance is commonly quantified by defining the relative energy content, E_i , for each basis vector *i* as

$$E_i = \lambda_i \Big/ \sum_{j=1}^{m} \lambda_i \tag{12}$$

It can be shown that the amount of total kinetic energy from velocities in the snapshots that are associated to a given POD mode is proportional to the corresponding eigenvalue [9]. Consequently, ordering of eigen values and eigen vectors ensures that the most important modes in terms of energy are the first modes. This usually means that the first modes will be associated with large scale flow structures. In other words, if a flow has dominant flow structures, these are dominantly reflected in the first POD modes and hence a given snapshot can often be reconstructed satisfactorily using only the first few modes [10].

3.2 Gappy POD theory

The gappy POD procedure uses a POD basis to reconstruct missing, or "gappy" data. This procedure was developed by Everson and Sirovich [2] and can be described as follows. The first step is to define a mask vector, which describes for a particular flow vector where data are available and where data are missing. For the flow solution V^{*i}, the corresponding mask vector n^{i} is defined as follows:

$n^{i} = 0$ if V_{r}^{*i} is missing,

$n^i = 1$ if V_r^{*i} is known.

V_r^{*i} denotes the r^{th} element of the vector V^{*i}. Point-wise multiplication is defined as $\left(n^i, V^{*i}\right)_r = n_r^i V_r^{*i}$. The gappy inner product is then defined as $(u,v)^N = \left((N,u),(N,v)\right)$, and the induced norm is $\left(\|v\|^N\right)^2 = (v,v)^N$ $(\|v\|^N)^2 = (v,v)^N$.

Let $\{\varphi_i\}(i=1,2,...,m)$ be the standard POD basis for the snapshot set $\{V^{*i}\}(i=1,2,...,m)$, where all snapshots are completely known. Let $\{g_i\}(i=1,2,...,m)$ be another set of solution vectors that have some elements missing, with corresponding mask vectors n_r^i. It is intended to reconstruct the full or repaired vector from the incomplete vectors $\{g_i\}(i=1,2,...,m)$. Assuming that the vectors {g□} represent a solution whose behavior can be characterized with the existing snapshot set, the intermediate repaired vector $\{\tilde{g}_i\}(i=1,2,...,m)$ can be expressed in terms of p POD basis functions as follows:

$$\tilde{g}^k \approx \sum_{j=1}^{P} b_j^k \varphi_j, \qquad K = 1,2,\cdots,m \tag{13}$$

To compute the POD coefficients b^r, the error, ε, between the original and repaired vectors must be minimized. The squared error is defined as

$$(\varepsilon^k)^2 = \| g^k - \tilde{g}^k \|_n^2, \qquad K = 1,2,\cdots,m \tag{14}$$

using the gappy-norm such that only the original existing data elements in g are compared. The coefficients b_i^r that minimize the error ε^r can be found by differentiating (11) with respect to each of the b_i^r in turn. This leads to the linear system of equations

$$Mb^k = f^k, \qquad K = 1,2,\cdots,m \tag{15}$$

where

$$M_{ij} = (\varphi^i, \varphi^j)_n, \qquad i,j = 1,2,\cdots,m \tag{16}$$

and

$$f_j^k = (g^k, \varphi_j)_n, \qquad j,k = 1,2,\cdots,m \tag{17}$$

Solving equation (15) for b^r and using equation (16), the intermediate repaired vectors $\{\tilde{g}_i\}(i=1,2,...,m)$ can be obtained. Finally, the complete $\{g_i\}(i=1,2,...,m)$ is reconstructed by replacing the missing elements in g^i by the corresponding repaired elements in $\tilde{g}^i$, i.e. $g_r^i = \tilde{g}_r^i$ if $n_r^i = 0$.

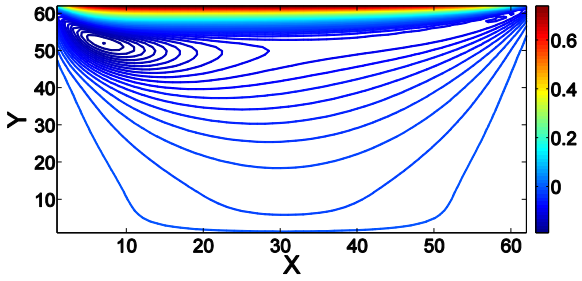
The initial condition used in the gappy region is typically the nonzero average value over all snapshots. However, the present study uses a zero initial condition for all masked points to check the robustness of algorithm in predicting the accurate data [3].

4. Results and discussion

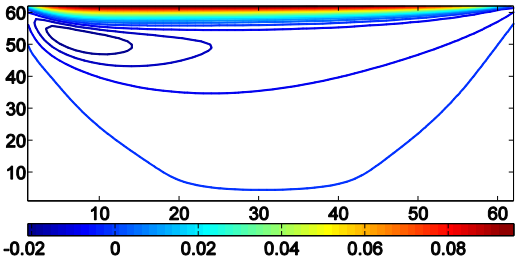
4.1 POD Modes

The POD approach is applied on 100 temporally evolving realizations from DNS of cavity flow. Figure 1 illustrates the streamwise component of the mean flow as well as its first 4 modes. A qualitative similarity can be observed in the patterns of mean flow with that of first mode (for each of planar velocity components). It is consequently expected that the first mode captures a dominant part of kinetic energy. This can be ensured by calculating the relative energy of each mode. According to section 2, this quantity is obtained by dividing the square eigenvalue of the target mode to the total energy of the modes (which is the sum of all square eigenvalues).

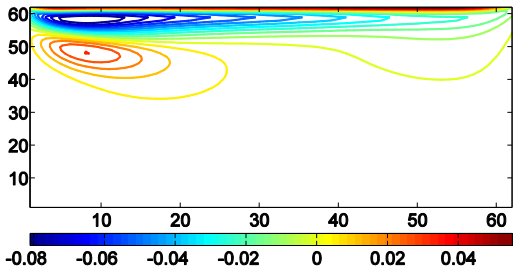
Figure 2 presents the relative energy of the first 10 modes for two velocity components. As is evident, the major share of energy is captured in the first modes and as the mode number increases, its relative energy decreases. There is no particular difference between the relative significance of energy between two velocity components.



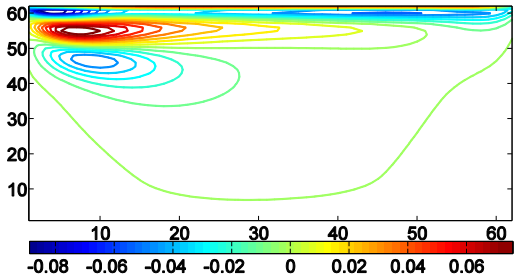
(a)



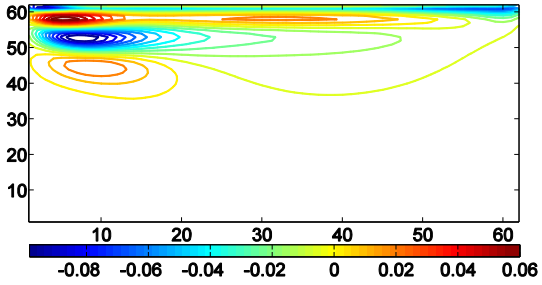
(b)



(c)



(d)



(e)

Fig. 1. Comparison of mean flow pattern for the streamwise component (a) with first four modes: first mode (b), second mode (c), third mode (d), and fourth mode (e).

4.2 Reconstruction with GPOD

The original DNS data is converted into a gappy data after implementing a random discarding procedure. The gappy POD approach is then applied on the flawed data and its function is examined. Figure 3 illustrates the GPOD relative error for streamwise velocity component. In other words, it is a measure of convergence for GPOD algorithm. The relative error is defined as [3]:

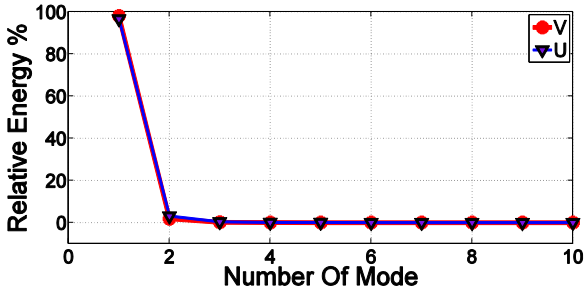
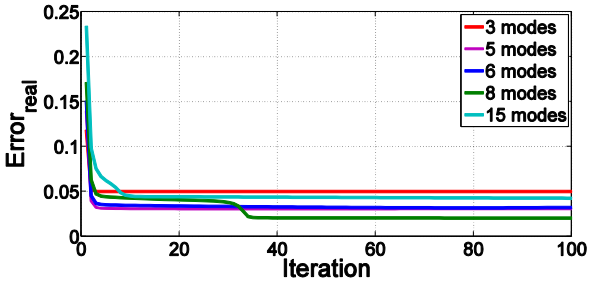
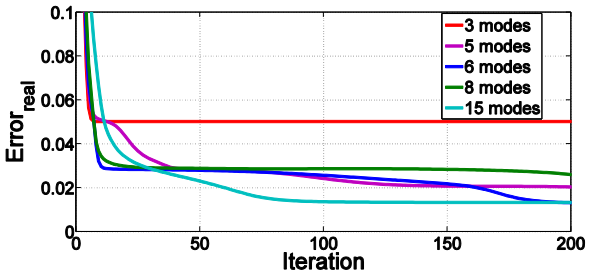


Figure 2. Relative energy of the first 10 modes for two velocity components.



(a)



(b)

Fig. 3. The GPOD relative error for streamwise velocity component for (a) 5% of gappiness, and (b) 30% of gappiness.

$$\frac{\| V_M^{*(i)} - V^* \|_2}{\| V^* \|_2} = \frac{\int\int_{\Omega T} | (\tilde{V}_M^{*(i)}(x,t) - V^*(x,t) | \, dxdt}{\int\int | V^*(x,t) | \, dxdt} \tag{18}$$

where $\tilde{V}_M^{*(i)}(x,t)$ denotes the *M*th mode of reconstructed streamwise component of the velocity field at the *i*th iteration, while $\tilde{V}^*(x,t)$ denotes the original streamwise velocity component that is obtained from the DNS of cavity flow. Parts (a) and (b) of figure 3 correspond to 5% and 30% of gappiness, respectively (the gappiness percentage indicates the ratio of unavailable data to total amount of data). The different curves in figure 3 correspond to different number of implemented modes, i.e. 3, 5, 6, 8 and 15 modes. Figure 3 shows that the relative error in gappy POD approach tends towards a constant value that is not so tiny. The main reason is that the gappy POD is responsible to predict the original DNS field from an incomplete snapshot. However, the GPOD has a particular potential/capacity for prediction of that data and can not have an exact estimation (tiny error). Increasing of the gappiness level intensifies the error and increases the number of required iteration for a converged pattern.

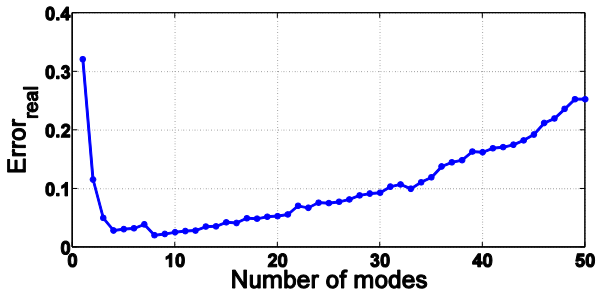
Increasing of the number of modes does not necessarily results in a monotonic decreasing of error. There is an optimum mode number such that adding the modes before it reduces the error and enhances the captured structures, while the later modes deteriorate the prediction accuracy. This is shown in figure 4 that presents variation of relative error with increasing the number of implemented modes for two gappiness level, i.e. 5% and 30%. According to this figure, the minimum relative error for 5% and 30% of gappiness is occurred for $M = 8$ and $M = 9$, respectively. Such behavior is not surprising as including higher-level modes may introduce noise that may degrade the quality of the reconstructed solution.

The above analysis is beneficial for spatially measuring techniques such as particle image velocimetry. Those techniques acquire instantaneous velocity field in a planar region. Because of experimental uncertainties and insufficient signal to noise ratio, a few velocity vectors are recognized as invalid/uncertain data and are removed from the snapshots [11]. The resulting vector map is consequently a gappy data that can be completed by the gappy POD algorithm. The present study developed understanding on the optimal number of modes for GPOD on the basis of a complete DNS data set. This issue shall be followed in our future studies on PIV measurements of complicated flow fields.

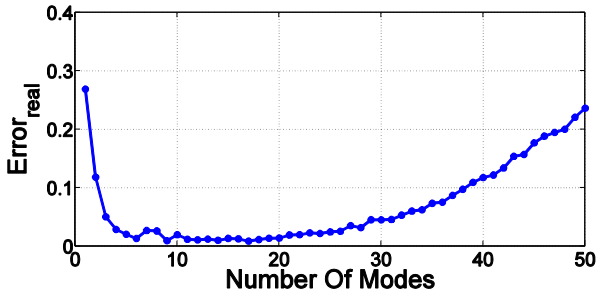
5. Conclusions

The cavity flow is simulated numerically using the projection scheme to evaluate the performance of gappy proper orthogonal decomposition in reconstructing time-dependent gappy data. Although the gappy data is artificially generated by randomly omitting data points from DNS snapshots, the same procedure can be applied to spatially measuring data such as particle image velocimetry.

The implemented procedure can predict the missing data with approximately 1% error for gappiness up to 30%. Examination of GPOD error with the number of considered modes shows that the error is not monotonically decreases with increasing the count of modes. The higher-level modes bring noise into the GPOD procedure and deteriorate the accuracy of missing data prediction. The optimal count of modes for artificial gappy data of cavity flow is obtained to be 8 and 9 for the gappiness level of 5% and 30%, respectively. Elevation of the gappiness level increases the count of required iteration to obtain a convergent data. The optimizations of the present study can be a basis for compensating gappy experimental data in an efficient manner.



(a)



(b)

Fig. 4. Variation of relative error with increasing the count of used modes for two gappiness levels: (a) 30%, and (b) 5%.

References

[1] Zeman, O., and Lumley, J.L., 1974. "Modeling buoyancy driven mixed layers". J. Atm. Sci., 33, pp. 1974-1988.

[2] Everson, R. and Sirovich, L., 1995. "Karhunen-Leoeve procedure for gappy data". J. Opt. Soc. Am., 12, pp. 1657-1664.

[3] Venturi, D. and Karniadakis, G.E., 2004. "Gappy data and reconstruction procedures for flow past a cylinder". J. Fluid Mech., 519, pp. 315-336.

[4] Gunes, H., Sirisup, S., and Karniadakis, G.E., 2006. "Gappy data: To krig or not to krig?". J. Comput. Phys., 212, pp. 358-382.

[5] Venturi, D., 2006. "On proper orthogonal decomposition of randomly perturbed fields with applications to flow past a cylinder and natural convection over a horizontal plate". J. Fluid Mech., 559, pp. 215-254.

[6] Druault, P., and Chaillou, C., 2007. "Use of proper orthogonal decomposition for reconstructing the 3D in cylinder mean-flow field from PIV data". C. R. Mecanique, 335, pp. 42-47.

[7] Peyret, R., and Taylor, T.D., 1983. "Computational methods for fluid flow". Springer Press.

[8] Nobari, M.R.H., Ahrabi, B.R., and Akbari, G., 2009. "A numerical analysis of developing flow and heat transfer in a curved annular pipe". Int. J. Therm. Sci., 48, pp. 1542-1551.

[9] Fukunaga, K., 1990. "Introduction to statistical pattern recognition". 2nd edition.

[10] Meyer, K.E., Pederson, J.M., and Oktayozcan, 2007. "A turbulent jet in cross flow analyzed with proper orthogonal decomposition". J. Fluid Mech., 583, pp. 199-227.

[11] Akbari, G., Montazerin, N., and Akbarizadeh, M., 2012. "Stereoscopic particle image velocimetry of the flow field in the rotor exit region of a forward-blade centrifugal turbomachine". J. Power Energy, 226, pp. 163-181.