PROPER ORTHOGONAL DECOMPOSITION ON COMPRESSED DATA

Oana Marin
Mathematics and Computer Science
Argonne National Laboratory
9700 S Cass Avenue, Lemont IL, USA
oanam@anl.gov

Elia Merzari
Mathematics and Computer Science
Argonne National Laboratory
9700 S Cass Avenue, Lemont IL, USA
emerzari@anl.gov

Philipp Schlatter
Linné FLOW Centre
Royal Institute of Technology
Stockholm 10040, Sweden
pschlatt@mech.kth.se

Andrew Siegel
Mathematics and Computer Science
Argonne National Laboratory
9700 S Cass Avenue, Lemont IL, USA
siegela@anl.gov

ABSTRACT

The main approach to identifying coherent structures in a flow field is the Proper Orthogonal Decomposition, due to its simplicity and effectiveness. However it is a data intensive method which becomes more expensive as the data field increases in size. The difficulty pertains mostly when a three-dimensional decomposition is performed, and the limiting factor is storing and loading large data fields of up to billions of gridpoints. This restriction is a conseuqence of the fact that the I/O bandwidth of supercomputers has not been at the same developmental pace as the CPUs. Lossy compression can reduce the size of the data fields and accelerate the computations. The strategy we suggest here relies on data compression via Discrete Chebyshev Transform (or alternatively Discrete Legendre Transform) which leaves invariant the auto-correlation matrix which lies at the core of the POD method. We show that by discarding over 90% of the data we can still retrieve a good proper ortohonal basis of the data set which deviates from the original by 10^{-2} in the L_2 norm.

INTRODUCTION

Proper Orthogonal Decomposition (POD), and in particular the method of snapshots (Sirovich, 1987), has been widely used to pinpoint coherent structures in highly turbulent flows. The idea is to store a number of decorrelated instantaneous data fields, build the covariance matrix associated with this data set, and decouple the dynamics from the spatial components by performing an eigenvalue decomposition. However with increasing problem size solving the eigenvalue problem is no longer the single most challenge of a POD computation. The data intensive nature of the POD procedure encounters an increasing difficulty due to slow speeds of the I/O on modern architectures. Although current machines have developed considerably in terms of processor speed and intranetwork connectivity, the speeds of the read/write operations have not followed the same trend.

The current work proposes a simple strategy to increase the computational speed of a POD, which can be employed in almost any CFD code, although we showcase it here in Nek5000 (Fischer *et al.*, 2015). The idea is based on compressing/truncating the data using the Discrete Cosine/Chebyshev Transform (DCT) and using the compressed/truncated data set to perform the POD analysis. Compression via data truncation is known as lossy compression. A similar approach has been used on data compressed using wavelets (Uytterhoeven & Roose, 1997), however that strategy required a stage of data decompression in order to retrieve the dominant modes. Since the DCT is an orthogonal transform, such a step is not needed, rendering our approach less prone to errors due to many conversions and also allows us handle better the error incurred through lossy compression.

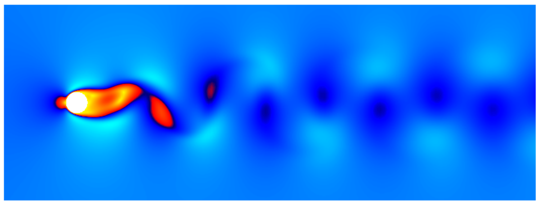

Figure 1. Flow past a cylinder at Reynolds number 100

The case under consideration here is a classical example, the flow past a cylinder in two dimensions at Reynolds number 100. This particular case has been extensively studied and we refer the reader for example to Venturi (2006).

DATA COMPRESSION/TRUNCATION

The DCT has the property of being the optimal approximation of the Karhunen-Loéve Transform(KLT). As is well known in the coherent structures community the Karhunen-Loéve transform is closely related to Principal Component Analysis which laid the grounds for the methods of snapshots of the Proper Orthogonal Decomposition, Sirovich (1987).

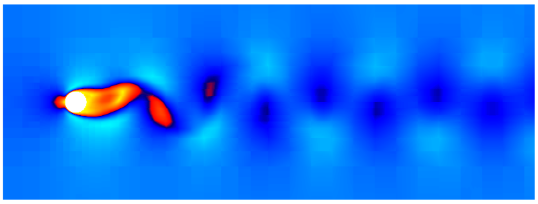

Figure 2. Flow past a cylinder at Reynolds number 100, data field compressed at 97% and error 10^{-2}

The DCT has also been the preferred method for image compression, serving as the basis for the JPEG format Ahmed *et al.* (1974).There exists a proof that the KLT is optimal for the truncation of a signal in *rms* norm, however we could show that this is simlar in L_2 norm. Recently we developed a data compression algorithm for scientific data Marin *et al.* (2016) which we equipped with an a priori error estimator and observed that by discarding up to 97% of the data there is no qualitative information loss.

The DCT is an orthogonal transform which has optimal energy compactness properties, i.e. given a signal it reshuffles the information such that the highest energy modes are stored in the first few components, thus allowing for an efficient truncation of the data.

The orthogonality property of the DCT transform, T, reads

$$T^T T = T T^T = I \ . \tag{1}$$

Given a velocity field $\mathbf{u}$ the energy norm becomes discretely

$$\int \mathbf{u}^2 \approx \underline{u}^T B \underline{u} \tag{2}$$

where B is diagonal matrix constructed in a three dimensional set up as the tensor product of one dimensional mass matrices B_1, i.e. $B = B_1 \otimes B_1 \otimes B_1$. On a Chebyshev grid the mass matrix $B_1 = \frac{\pi}{n} I$, where n is the number of grid points within one mesh element in any direction, and I is the identity matrix. We can now easily note that the energy norm in real space is equivalent discretely to the energy norm in DCT space

$$\int \mathbf{u}^2 = (T\underline{u})^T B (T\underline{u}) = \underline{u}^T T^T B T \underline{u} = \underline{u}^T B \underline{u} \tag{3}$$

where we used the fact that the matrix yielding the integration weights is a diagonal factor of the identity matrix, and thus $T^T B T = B T^T T = B$.

Now consider a truncated signal $\tilde{u}$ where we removed a certain amount of data such that $\int (\mathbf{u} - \tilde{\mathbf{u}})^2 \leq \varepsilon$ with an ε prescribed by the user. From Equation 3 we have that also for the truncated quantity $\underline{\tilde{u}}^T T^T B T \underline{\tilde{u}} = \underline{\tilde{u}}^T B \underline{\tilde{u}}$.

These observations facilitated the derivation of a straight forward a priori error estimator by which we can easily assess the error committed via truncation. The claim that 97% of data truncation (at an error of 10^{-2} in energy norm) preserves the relevant data is illustrated in Figure 2 where the original behaviour of the cylinder wake is preserved even by retaining only 3% of the original signal in Figure 1.

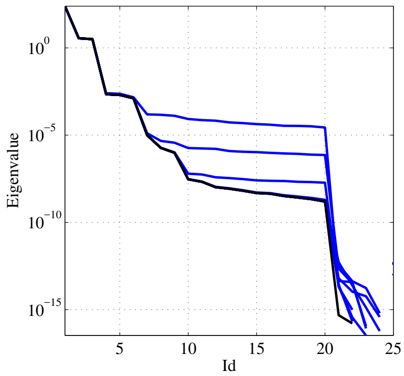

Figure 3. Eigenvalues decay: (blue) various truncation levels, (black) original matrix

PROPER ORTHOGONAL DECOMPOSITION

The underlying assumption of the POD is that the data field $\mathbf{u}$ can be decomposed as

$$\mathbf{u} = \sum_k a_k(t) \phi_k(t) \ . \tag{4}$$

To determine both a_k and ϕ_k, and the subsequent flow dynamics, the POD requires a number M of snapshots which are assembled into the covariance matrix, i.e.:

$$C_{ij} = \int \mathbf{u}(\mathbf{x}, t_i) \mathbf{u}(\mathbf{x}, t_j) d\Omega, \ \mathbf{x}, \mathbf{y} \in \Omega \tag{5}$$

where $i, j = 1, M$. Discretely this matrix can be represented as $C_{ij} = \underline{u}_i^T B \underline{u}_j$.

If we compute the covariance matrix on the truncated signal then we have

$$\tilde{C}_{ij} = \underline{\tilde{u}}_i^T B \underline{\tilde{u}}_j = \underline{u}_i^T B \underline{u}_j - \Sigma_{ij} = C_{ij} - \Sigma_{ij} \tag{6}$$

where $\tilde{\lambda}_i$ are the eigenvalues of the truncated matrix $\tilde{C}$ and $\Sigma_{ij} \leq \varepsilon$ for all $i, j = 1, \ldots M$.

The central observation is that the eigenvalues of $\tilde{C}$, as a symmetric matrix, must obey

$$tr(\tilde{C}) = \sum \tilde{\lambda}_i. \tag{7}$$

If we consider a perturbation matrix which enjoys the same properties as the covariance matrix (which Σ does by construction) we can say even more about the eigenvalues of both perturbed and unperturbed matrices as will soon be shown.

Lossy compression of the data translates, in the case of the covariance matrix, to a perturbation stemming from the amount of information truncated. Thus the more we compress the data the higher the perturbation in the covariance matrix. However we know this perturbation since it was imposed by he user. We aim to assess here how much truncation/compression is permissible in order to obtain the same POD decomposition.

Assessing the impact of the perturbation of the covariance matrix on its eigenvalues falls into the filed of matrix perturbation theory, Stewart & Sun (1990), which provides clear boundedness results for symmetric matrices such as the covariance matrix. We cite in particular Weyl's theorem, Weyl (1912), which states that for a matrix $C = \tilde{C} + \Sigma$ the eigenvalues will be bounded by the norm of the perturbation matrix Σ

$$|\tilde{\lambda}_k - \lambda_k| \leq \|\Sigma\|_2 \,, \tag{8}$$

where the matrix 2-norm is given by the maximum eigenvalue $\|\Sigma\|_2 = \lambda_{max}$. In our case the norm of the matrix Σ is easy to compute since every entry $\Sigma_{ij} \leq \varepsilon$ where ε is imposed by the user a priori. Also Σ is a symmetric positive definite matrix and shares the same properties as the underlying covariance matrix C.

In the same vein a bound on the of sum all eigenvalues is given be the Weilandt-Hoffman theorem

$$\sum_k^M (\tilde{\lambda}_k - \lambda_k)^2 \leq \|\Sigma\|_F . \tag{9}$$

where $\|\Sigma\|_F$ is the Frobenius norm.

In Figure 3 the decay of the eigenvalues for compressions ranging from 97% down to no compression is displayed. The relationship between error and compression ratio is illustrated in Figure 4. To read these plots in conjunction we take the example of the most severe compression 97% and in Figure 4 note that this incurs a tolerated error of 10^{-2}, however in Table 1 the committed error in practice was even lower at 4.83^{-3}. This translates into a perturbation in the covariance matrix and in Figure 3 the eigenvalues display the highest deviation, especially after the 6-th mode. If we consider a compression of approximately 66% this gives an error of 1.24^{-5} and we see in Figure 3 that the fourth level of compression yields a decay similar to that of the original unperturbed matrix. We note than anything lower than an error of 10^{-5} gives already a decay indistinguishable from that of the original matrix.

In Figure 5 we illustrate the reliability of Weyl's result in giving an upper bound that is not overestimating the error. The deviation of the perturbed eigenvalues (black) from the estimate based on the norm of the perturbation matrix (red) is unnoticeable in the regime of $10^{-3} - 10^{-9}$. For values lower than 10^{-9} the round off errors have a higher impact which is to be expected, and the eigenvalues are far closer to the ones of the original matrix. Imposing a tolerance of 10^{-9} would have almost no impact on the covariance and the modes.

Noting now using Equation 3 that the entries of the matrix C are invariant to the DCT transform, i.e. operating with $\underline{u}$ or its counterpart $T\underline{u}$ in DCT space yields the same result.

The second step in determining the dominant modes is to perform an eigenvalue decomposition of the matrix $\tilde{C}$ and note how they deviate form the original matrix C.

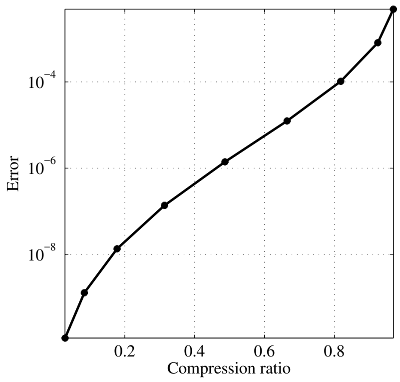

Figure 4. Error versus compression ratio

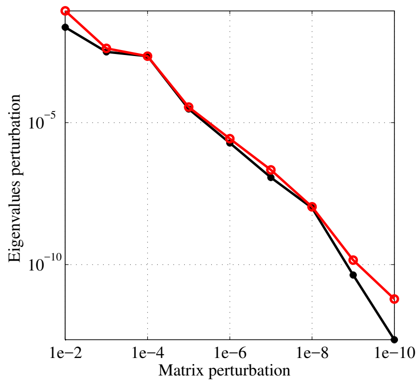

Figure 5. Perturbation in eigenvalues $|\tilde{\lambda}_k - \lambda_k|$ (here for the maximum eigenvalue) as a function of the norm of the perturbation matrix $\|\Sigma\|_2$. (black) eigenvalue deviation, (red) Weyl's estimate

Table 1. Error versus Compression ratio.

Threshold	Error	Compression ratio
10^{-2}	4.83^{-3}	0.97
10^{-3}	8.12^{-4}	0.92
10^{-4}	1.02^{-4}	0.81
10^{-5}	1.24^{-5}	0.66
10^{-6}	1.40^{-6}	0.48
10^{-7}	1.36^{-7}	0.31
10^{-8}	1.35^{-8}	0.17
10^{-9}	1.29^{-9}	0.08
10^{-10}	1.15^{-10}	0.02

RESULTS

After performing a POD analysis using the method of snapshots on data truncated by 97% we compare in Figure 6 and Figure 7 the first modes between original data and compressed data, and in Figure 8 and Figure 9 the second modes. There is no notable difference although on closer scrutiny there is an error between the modes proportional to the truncation error due to compression, visible especially in the tail of the wake and across element boundaries. By assessing the decay of the eigenvalues which are a stand-in for the energy decay corresponding to each mode we observe in Figure 3 that the first 6 modes which encapsulate in this case around 98.995% of the total energy have an identi-

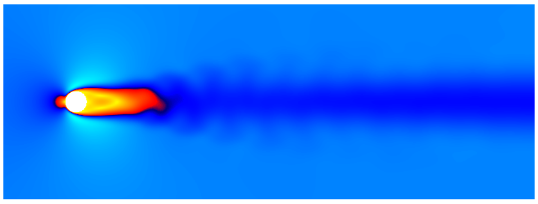

Figure 6. Mode 1 for unperturbed original matrix

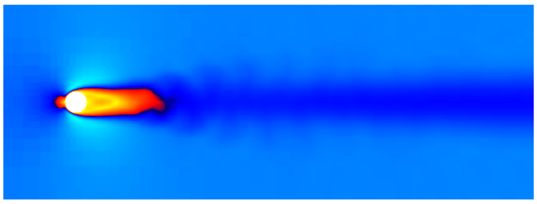

Figure 7. Mode 1 for data compressed at 97% and error 10^{-2}

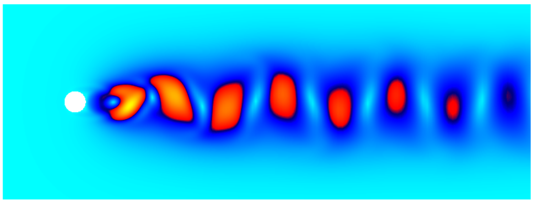

Figure 8. Mode 2 for unperturbed original matrix

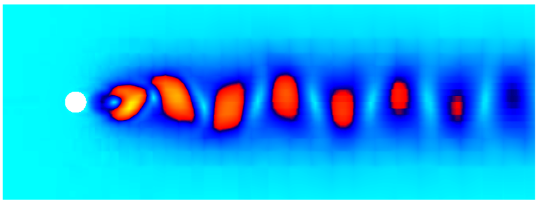

Figure 9. Mode 2 for data compressed at 97% and error 10^{-2}

cal behaviour, however the eigenvalues deviate significantly after mode 6.

For a simple study of the modes and the system behaviour a higher truncation seems to be satisfactory, however for constructing a reduced order model we indeed need to have as high accuracy as possible. For this particular case we note from the eigenvalues decay in Figure 3 an imposed tolerance of 10^{-5}, corresponding to 66% compression is sufficient for removing any deviation even in the higher modes. In Figure 10-11 we note that the behaviour is unchanged and no minor artifacts are present (as was the case with 10^{-2} errors), even down to mode 6 (Figure 12) which has a very low energy contribution of 0.002 still the mode is fully retrieved with no modifications.

CONCLUSIONS AND OUTLOOK

The POD analysis performed on compressed data of up to 97% (i.e. disregarding 97% of the information) has been shown to yield an error of approximately 10^{-2} in energy norm, which is undetectable from both a visual perspective as well as in terms of preserving the relevant information.

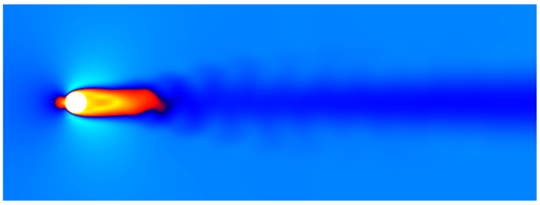

Figure 10. Mode 1 for data compressed at 66% with an error of 10^{-5}

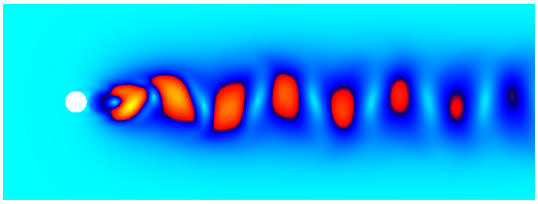

Figure 11. Mode 2 for data compressed at 66% with an error of 10^{-5}

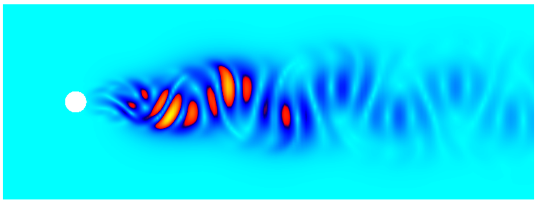

Figure 12. Mode 6 for data compressed at 66% with an error of 10^{-5}

The approach suggested here to use compressed data for POD analysis relies on this qualitative behaviour, as well as a proof that the most relevant modes are preserved thanks to the properties of the compression strategy via DCT transforms. To assess what is the tolerable deviation in eigenvalues and the subsequent reconstruction of the modes we invoked results from matrix perturbation theory and showed that the deviation in eigenvalues is given by the norm of the perturbation matrix, which in turn, is controlled by the user via the imposed tolerance.

We have shown that for reduced order models it is possible to compress at 66% ratio and there is essentially no data loss even for the lowest insignificant modes. The simplicity of the approach and the computational gain it brings about render this strategy for evaluating the dominant modes in a simulation a suitable feature in any CFD code.

REFERENCES

Ahmed, N., Natarajan, T. & Rao, K. R. 1974 Discrete cosine transform. *IEEE Transactions on Computers* **C-23** (1), 90–93.

Fischer, Paul, Lottes, James, Kerkemeier, Stefan, Marin, Oana, Heisey, Katherine, Obabko, Aleks, Merzari, Elia & Peet, Yulia 2015 Nek5000: User's manual (ANL/MCS-TM-351).

Marin, Oana, Schanen, Michel & Fischer, Paul 2016 Large-scale lossy data compression based on an a priori error estimator in a spectral element code (ANL/MCS-P6024-0616).

Sirovich, Lawrence 1987 Turbulence and the dynamics of coherent structures part i: coherent structures. *Quarterly of applied mathematics* **45** (3), 561–571.

Stewart, G.W. & Sun, J. 1990 *Matrix Perturbation Theory*. Academic Press.

Uytterhoeven, Geert & Roose, Dirk 1997 Experimemts with a wavelet-based approximate proper orthogonal decomposition (TW263).

Venturi, Daniele 2006 On proper orthogonal decomposition of randomly perturbed fields with applications to flow past a cylinder and natural convection over a horizontal plate. *Journal of Fluid Mechanics* **559**, 215254.

Weyl, Hermann 1912 Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen* **71** (4), 441–479.