

Compressive techniques for accelerated flow analysis

By P.J. Schmid[†], B. Mantravadi[†], A. Glazkov[†], A. Patil AND T. Flint

A randomized projective compression scheme is proposed that greatly accelerates the analysis of flows based on a snapshot sequence. It acts on the spatial direction of each snapshot while leaving the temporal dynamics, encoded in the snapshot sequence, intact within a user-defined distortion factor. Underlying this compression is the Fast Johnson-Lindenstrauss Transform (FJLT) which guarantees an isometric embedding of the snapshot sequence in a restricted subspace. The reduction scheme is demonstrated on Rayleigh-Bénard convection in a spherical shell, where reductions down to $\sim 3.4\%$ of the full degrees of freedom could be reported, without perceptible deterioration of the spatial structures. An extension of the technique to Krylov subspace sequences is discussed as well.

1. Introduction

Over the past decades, simulations of fluid systems have continuously pushed the limits of scale resolution and physical features and have reached an impressive level of detail and complexity. Multiple phases and transitions between them, reactive flows, turbulence, particle-laden configurations, aerothermodynamic effects, radiative transport, and acoustic feedback are but a few complications that require a significant effort to simulate – and an even more substantial effort to analyze. Common to all these simulations is the production of large-scale flow fields in the form of snapshot sequences (for a data-based analysis), or access to discretized governing equations with expansive degrees of freedom (for a model-based approach). As the numerical treatment of systems with a vast range of spatio-temporal scales and multiple, interacting physical constituents are becoming commonplace, the tools to analyze this type of flows to gain insight from the simulated flow fields must follow suit in sophistication and intricacy.

In the analysis of these types of flows it is not unusual that the structures of interest, i.e., structures that are responsible for the bulk of mass, momentum and energy transport, are few in number and compact in space (Taira *et al.* (2017)). The immediate question, prompted by this observation, arises whether a more efficient extraction that bypasses the processing of all degrees of freedom and rather focuses on the essential ones, can be developed and employed. The same question has emerged in and defined the field of compressive sensing (Candes & Wakin (2008)), where signals that are known to be sparsely representable in a suitable basis can be processed more efficiently by taking advantage of their sparse structure. The concerted effort to answer this question has produced some of the most effective and powerful algorithms in the field of signal processing (see, e.g., Candes & Tao (2006)).

Rather than recovering a sparse signal from a compressed representation, the present paper will attempt to compress a rich, but structurally sparse signal into a compact

[†] King Abdullah University of Science and Technology (KAUST), Saudi Arabia

representation without compromising its essential features or its recoverability. This general approach has a long and distinguished history, related to various disciplines such as spectral analysis, model reduction, or feature encoding. The representation of multi-harmonic signals using a Fourier transform, the compression of localized and multi-scale fields using a wavelet transform, or, more recently, the learning of an auto-encoder from processing snapshot sequences are examples of this compressive viewpoint.

Central to a spectral analysis of flow field sequences is the idea of a decomposition into a superposition of spatio-temporal subcomponents (Taira *et al.* (2017)). This approach is akin to the familiar separation-of-variables approach for the solution of partial differential equations, where a linear combination of triplets consisting of spatial components, scalar amplitudes, and temporal elements is posed and determined. In this notion, the spatial structures form the transformation basis, the temporal elements constitute the transformed signals, and the amplitudes reflect the spectrum (i.e., importance) of the various triplets and give rise to a hierarchical representation of the full process. The truncation of this hierarchy yields a reduced system that can be interpreted as a projection of the full system onto the most important subprocesses. This kind of decomposition can be employed on the homogeneous problem or the driven problem: in the former case, we obtain the intrinsic behavior (related to inherent instabilities, in the broadest sense), in the latter case, we obtain the response behavior (related to resonances, in the broadest sense).

In recent years, the merit and potential of randomized projections and expansion bases for the reduction of complex, but compressible (in the information-theoretic sense) signals have come to the forefront of many research fields. In linear algebra (Martinsson & Tropp (2020)), highly efficient algorithms have been developed for approximately rank-bounded matrices (i.e., matrices with a strong drop in their singular values). Linear systems, eigenvalues, singular values, and matrix functions for these types of matrices can now be determined without squandering computational resources on the approximate null-space.

Generally speaking, any projection is lossy. However, projections can be constructed that preserve, at least within user-specified bounds, certain characteristics of snapshots or operators. One of the most celebrated projections is based on the Johnson-Lindenstrauss theorem (Johnson & Lindenstrauss (1984)) that states that, given a sequence of vectors, the Euclidean distances between *any two* vectors in the sequence are preserved under a projection, up to a small user-specified distortion factor. In essence, this projection produces a second sequence from our original one that is referred to as an isometric embedding (into a reduced space). An application of this compression scheme has been presented in Glazkov & Schmid (2023) for flow through a linear cascade of turbine blades. In this paper, we will explore this decomposition, its implementation, and its effectiveness in accelerating traditional flow analysis of large-scale data sequences. In addition, its influence on iterative techniques for matrix-free algorithms will be assessed as well.

2. Mathematical background

Starting point for our analysis is a sequence of flow-field snapshots that stem from either a numerical simulation or from an experimental effort. The snapshots will be assumed to be sampled with a constant sampling frequency, given by a constant time-step Δt between any two consecutive snapshots. We denote the sequence of N snapshots by $\mathbf{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N\}$. Each individual flow field \mathbf{q}_i contains observables or state-

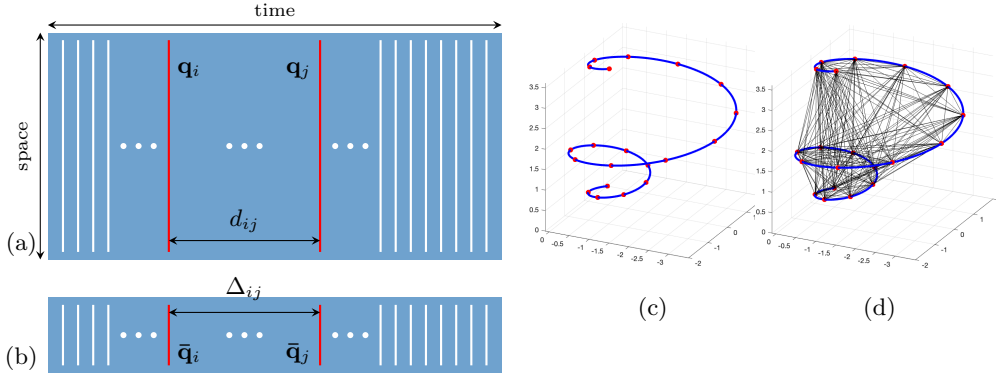


FIGURE 1. Sketch of embedding a snapshot sequence into a lower-dimensional space: an original snapshot sequence (a) is reduced in the spatial (row) direction by a projection Φ to produce a second sequence (b). The distances d_{ij} between any two snapshots $\mathbf{q}_{i,j}$ from the original sequence are nearly identical to the equivalent distances Δ_{ij} between the corresponding compressed snapshots $\bar{\mathbf{q}}_{i,j} = \Phi\mathbf{q}_{i,j}$. In phase space, the snapshot sequence can be represented by a trajectory (c). The projection nearly preserves the Euclidian distances between any two snapshots, indicated by black lines in (d).

vector components evaluated at measurement locations or grid locations. Under these assumptions, the flow field \mathbf{q}_i contains all spatial and physical information. Let us assume that each snapshot contains n components, with $n \gg N$. The number n represents the full number of degrees-of-freedom which we wish to compress; the sequence \mathbf{Q} of snapshots can then be interpreted as a trajectory in an n -dimensional phase space.

We then define a projection Φ acting on the snapshots and taking them from their original n -dimensional space to a k -dimensional space, with $k \ll n$. The projection Φ can then be cast into a $k \times n$ matrix. We seek to construct the projection Φ such that

$$(1 - \varepsilon) \|\mathbf{q}_i - \mathbf{q}_j\|_2 \leq \|\Phi(\mathbf{q}_i - \mathbf{q}_j)\|_2 \leq (1 + \varepsilon) \|\mathbf{q}_i - \mathbf{q}_j\|_2, \quad \forall i, j \in [1, N] \quad (2.1)$$

which means that the Euclidean distance between *any* two snapshots from our sequence is preserved under the projection Φ up to a distortion factor ε . The above expression is sketched in figure 1(a,b): the original snapshot sequence $\mathbf{q}_{1,\dots,N}$ is compressed into a second sequence $\bar{\mathbf{q}}_{1,\dots,N} = \Phi\mathbf{q}_{1,\dots,N}$. The distance between two representative snapshots $\mathbf{q}_{i,j}$ is preserved under the projection: $\Delta_{ij} = \|\Phi(\mathbf{q}_i - \mathbf{q}_j)\|$ is, within a factor of $(1 \pm \varepsilon)$, close to the original distance $d_{ij} = \|\mathbf{q}_i - \mathbf{q}_j\|$.

The Johnson-Lindenstrauss theorem ensures the existence of a projection Φ that satisfies the above condition, referred to as the restricted isometry property, or RIP for short. The projection Φ is random, and the above condition is satisfied in a probabilistic manner: the restricted isometry property holds with a probability above $2/3$. Before addressing the design of the randomized projection Φ , we need to attend to the compression factor n/k . This compression factor depends on the distortion ε we allow for our projection. The Johnson-Lindenstrauss theorem states that $k \sim \varepsilon^{-2}$. For example, for a state-vector with $n = 10^7$ degrees of freedom and a user-defined distortion factor of $\varepsilon = 0.01$, we can expect an embedding dimension of $k \sim 10^4$ or a compression factor of $n/k \sim 10^3$.

It is worth examining why the near-preservation of distances between any two snapshots maintains the temporal dynamics under the projection. This can be seen by sketch-

ing the snapshot sequence as a trajectory in phase space. Figure 1(c) depicts a representative trajectory in three-dimensional phase space: each snapshot with three components constitutes a point in this space, and the temporal evolution contained in the snapshot sequence is encoded in the path from snapshot-point to snapshot-point. Under the restricted isometry property, a projection onto a lower-dimensional space will nearly preserve each distance between any pair of phase-space points (see figure 1(d)). This near preservation imposes $\mathcal{O}(N^2)$ constraints on the positions in the lower-dimensional (projected) phase-space. Under the spatio-temporal decomposition, premultiplied by the matrix representing Φ , the temporal part is thus only minimally modified, and any property related to the temporal dynamics is equally encoded in the projected space.

The remaining question is related to the computation of the projection Φ , which translates the powerful Johnson-Lindenstrauss theorem into an equally powerful algorithm for the dynamics-preserving reduction of snapshot sequences. This translation has been accomplished in strides, where the most common manifestation is referred to as the Fast Johnson-Lindenstrauss Transform, or FJLT, proposed by Ailon & Chazelle (2009, 2010). It consists of a triple product of matrices, two of them randomized, one of them deterministic. It is cast in the form

$$\Phi = PHD, \quad P \in \mathbb{R}^{k \times n}, \quad H \in \mathbb{R}^{n \times n}, \quad D \in \mathbb{R}^{n \times n} \quad (2.2)$$

with P as a randomized, sparse, rectangular matrix, H as a deterministic, square matrix encoding a Walsh-Hadamard transform, and D as a randomized, diagonal matrix consisting of ± 1 on the diagonal. The first $k \times n$ matrix P performs a randomized projection onto a lower-dimensional row-space. Only one entry in each row of P is non-zero; in this sense, P is randomly picking rows from any vector that has been multiplied by H and D . The deterministic $n \times n$ matrix H is responsible for the spectral transform (in our case, a Walsh-Hadamard transform), i.e., a transfer of spatial information into scale (wavenumber) information. In combination, the matrices P and H would represent a randomized pick of spatial scales. If we are processing a flow field with a dominant and/or band-limited spectrum, the sparse nature of the P -matrix would mean that a zero output is the most likely result, since only a few entries after multiplication by H are non-zero. To circumvent this spectral concentration in a narrow band of scales, we need to artificially 'broaden' the spectral content of the incoming flow field: this is the purpose of the D -matrix. It consists of a random assembly of $+1$ and -1 on the diagonal, following a uniform distribution. This pre-multiplication generates a broad spectral content across all scales, while still carrying the spectral content of the original signal, before it is scale-transformed and projected by H and P , respectively.

It is important to convince ourselves of the efficiency of each step in this algorithm. The pre-multiplication by D is equivalent to a randomized sign-switch of the incoming signal, which can be accomplished in $\mathcal{O}(n)$ operations. The next step, the Walsh-Hadamard transform H , can be implemented, similar to an FFT, in $\mathcal{O}(n \log n)$ operations. The final step, the row-selection matrix P , takes k operations: we simply decide on and extract k components from the signal resulting from a multiplication by H and D . All in all, the transformation of a n -component state vector into a k -component compressed vector can be accomplished in $\mathcal{O}(n \log n)$ operations. During these operations, none of the matrices P, H, D are explicitly formed or stored; only their action on the original snapshots \mathbf{q}_i is needed.

While the Johnson-Lindenstrauss theorem covers the embedding based on pairwise distances, the algorithm can be applied to a higher linkage between snapshots. For

example, we can consider quadruplets of snapshots and demand that none of the cumulative distances are to be distorted beyond a given bound. This higher linkage will restrict the embedding beyond the constraints of pairwise matchings, yield a higher probability of compliance (beyond the 2/3-value for pairwise linkage), but will decrease the compressibility n/k of the original data set (i.e., more linked snapshots require a higher-dimensional embedding space, compared to pairwise linked snapshots). The governing parameters are thus: (i) the linkage number ξ , (ii) the number of snapshots N , and (iii) the distortion factor ε . The resulting parameters are given by (see Dasgupta & Gupta (1999) for details)

$$k \geq 4 \left(\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3} \right)^{-1} \log \xi \quad n/k \sim \frac{n\varepsilon^2}{\log \xi} \quad \delta \leq \frac{1}{3} \binom{\xi}{2} \quad (2.3)$$

where δ denotes the failure probability for satisfying the JL-condition.

Higher linkage is a familiar concept when processing snapshot sequences. It is often found in the analysis of sequential data streams in the form of time-delay (or Ruelle-Takens) embedding; see Takens (1981). In this application, original snapshots are linked by stacking ℓ consecutive snapshots to form a new composite snapshot. These composite snapshots, containing a short chain of an evolution process, are then processed in their entirety. Their significant overlap from composite snapshot to composite snapshot enforces temporal coherence and aids in the extraction of important subprocesses. It has to be recognized, however, that time-delay embedding is far weaker than the linkage we enforce based on the Johnson-Lindenstrauss transform. Time-delay Ruelle-Takens embedding establishes a linkage between *consecutive* snapshots, while Johnson-Lindenstrauss linkage enforces a stronger *all-to-all* connectivity between snapshots. Despite these important differences, in what follows, we will compare higher-linkage Johnson-Lindenstrauss cases with the equivalent time-delay embeddings.

3. Rayleigh-Bénard convection in a spherical shell

As a demonstration case, we consider Rayleigh-Bénard convection inside spherical shells in the supercritical regime. This case has been motivated by a study of convection patterns on the Sun. We choose a Rayleigh number of $Ra = 4 \cdot 10^6$, which falls far above the critical Rayleigh number. The governing equations are discretized in a hybrid scheme, where the spherical surface has been treated by a discrete exterior calculus (DEC) framework (Hirani (2003); Desbrun *et al.* (2005)), while the radial coordinate direction uses standard high-order finite differences (Mantravadi *et al.* (2023)). The DEC-approach provides a grid- and geometry-independent formulation of the governing equations and automatically and exactly satisfies differential calculus operations (such as, e.g., the divergence-theorem) on a general surface grid. In this manner, any geometric manipulation, e.g., changing the sphere into an ellipsoid, or any grid reallocation does not require any changes in the discretization schemes.

A sequence of 568 snapshots has been generated, and each snapshot contains $n = 1,638,720$ degrees of freedom. Our original data matrix is thus $1,638,720 \times 568$ in dimension. To this matrix, we apply the FJLT with $\varepsilon = 10^{-2}$ and $\xi = 2$ which allowed us to reduce the data matrix to a $55,448 \times 568$ matrix without distorting the temporal dynamics by more than one percent (with a probability above 0.667). The compressed dimension stems from taking the equality on the bounds expressed in (2.3).

We perform a POD analysis (Berkooz *et al.* (1993)), which is equivalent of decomposing

the snapshot data matrix according to a singular value decomposition. With \mathbf{Q} as the data matrix, we have

$$\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \quad (3.1)$$

where \mathbf{U} contains the spatial modes, $\mathbf{\Sigma}$ the singular values, and \mathbf{V} the temporal dynamics. Applying our FJLT-projection $\Phi = \text{PHD}$ to the above expression, we obtain

$$\Phi\mathbf{Q} = \underbrace{\Phi\mathbf{U}}_{\bar{\mathbf{U}}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^H \quad (3.2)$$

where $\bar{\mathbf{U}}$ denotes the compressed spatial structures (in the embedding phase space), and $\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{V}}$ the corresponding amplitudes and temporal signal, respectively. Since the temporal dynamics is not altered under the action of Φ (up to a small distortion factor), we can take $\bar{\mathbf{\Sigma}}\bar{\mathbf{V}} \approx \mathbf{\Sigma}\mathbf{V}$. Under these assumptions we can solve for \mathbf{V} in the original and compressed case and recover the spatial structures of the original set from the compressed one. Mathematically, we have

$$\mathbf{U} \approx \mathbf{Q}\bar{\mathbf{V}}\bar{\mathbf{\Sigma}}^{-1}. \quad (3.3)$$

The computation of $\bar{\mathbf{V}}$ and $\bar{\mathbf{\Sigma}}$ is performed on the compressed dataset and is hence far more efficient than the same computation on the full dataset.

Applying this compression technique to 568 snapshots taken from numerical simulations of Rayleigh-Bénard convection in a spherical shell, a representative POD mode (i.e., the second most dominant) is shown in figure 2(a) and juxtaposed to the equivalent mode from the full dataset (figure 2(b)). Hardly any difference in the structural composition of the mode (visualized by the radial velocity component) is discernible. Nevertheless, the compressed mode (figure 2(a)) has been computed in a fraction of the time, compared to the mode from the full snapshot sequence: while the full analysis of the data sequence took 157 seconds, the compressed analysis could be accomplished in about 4 seconds.

This also demonstrates that with a compression ratio of 29.55 in the spatial direction, the dominant POD modes are highly compressible (in the information-theoretic sense) and only require $\sim 3.4\%$ of the degrees of freedom available in one snapshot. This statement is not only true for the extraction of a single mode, but also holds for the POD analysis of the full data set. The first three POD modes of the Rayleigh-Bénard convection problem are shown in figure 3, again visualized by contours of the radial velocity.

With a selected linkage parameter of $\xi = 2$ (pairwise linkage) and a maximum distortion factor of $\varepsilon = 0.01$ (one percent), we achieved a compression of 29.55 and a computational speedup of 43. More specifically, only 55,448 degrees of freedom ($\sim 3.4\%$ of the full 1,638,720 degrees of freedom) are necessary to extract the full modal structure of the snapshot sequence. For this low-linkage case, the bounds can only be guaranteed for two-thirds of all cases. Statistical tests, however, show that the actual failure probability is lower, rendering the two-thirds value a rather conservative estimate. Moreover, the distortion in the failed cases rarely exceeds double the defined ε (in our case, two percent); see Glazkov & Schmid (2023) for details. For higher-linkage cases, the achievable compression will diminish; however, the failure probability of satisfying the Johnson-Lindenstrauss condition decreases precipitously. Tables 1 and 2 present the compression and computational speed-up factors as a function of the linkage parameter ξ .

Both tables contain additional information. Table 1 also contains a column (fourth column) that reports the compression factor for an equivalent time-delay embedding.

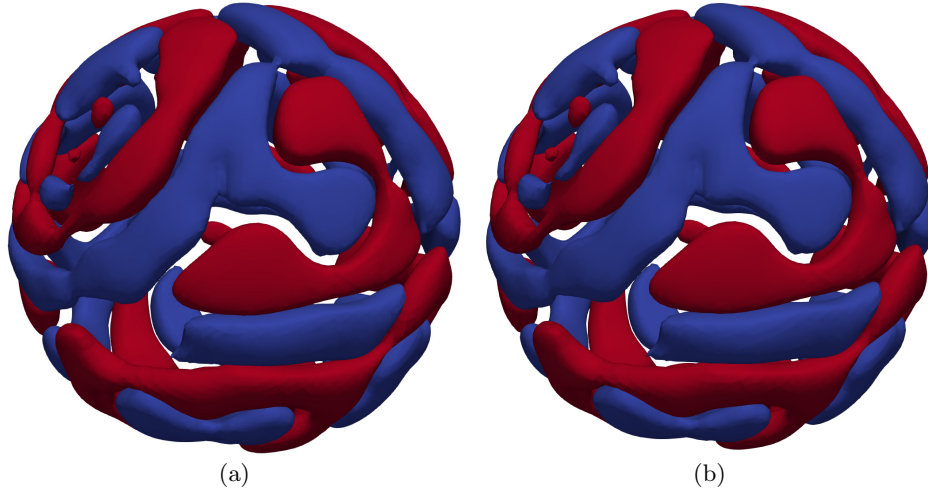


FIGURE 2. Rayleigh-Bénard convection in a spherical shell at $Ra = 4 \cdot 10^6$. Comparison of the second most dominant POD modes (a) from an FJLT-compressed analysis with the equivalent mode (b) from a full analysis. A distortion factor of $\epsilon = 0.01$ has been selected, and a linkage parameter of $N = 2$ has been applied.

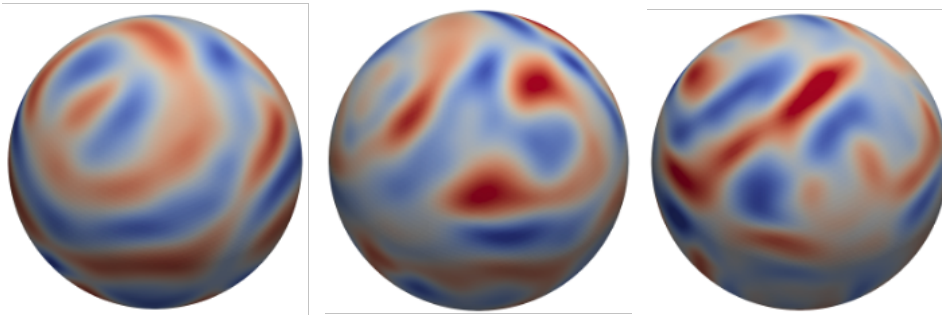


FIGURE 3. Rayleigh-Bénard convection in a spherical shell at $Ra = 4 \cdot 10^6$: the first three dominant POD modes, visualized by the radial velocity, obtained from an FJLT-accelerated analysis.

linkage	latent-space dimension	compression factor	compression factor (delay embedded)
2	55,448	29.55	59.11
4	110,896	14.78	59.11
6	143,336	11.43	68.60
8	166,352	9.85	78.81
12	198,792	8.24	98.92
16	221,800	7.39	118.21
24	254,240	6.45	154.69
32	277,256	5.91	189.14

TABLE 1. Compression ratios for different FJLT embedding dimensions (linkage parameters).

linkage	speed-up (POD)	speed-up (DMD)
2	43.03	42.74
4	19.56	20.51
6	14.99	16.00
8	12.88	13.66
12	10.30	10.40
16	8.80	9.43
24	7.74	7.80
32	7.72	7.40

TABLE 2. Speed-up for POD and DMD analysis with different linkage parameters.

This information relates to the approximate equivalence of an increased linkage with an equivalent Ruelle-Taken embedding, even though we stress again that the linkage (increased ξ) refers to the more stringent all-to-all connectivity.

The second table also contains information for a DMD-analysis (e.g., Rowley *et al.* (2009); Schmid (2010, 2022)). While we concentrated on a POD analysis to demonstrate the compression algorithm, other data-sequence-based decompositions can equally benefit from the FJLT. As an example, we applied the sparsity promoting DMD (Jovanović *et al.* (2014)) to the compressed data sequence to extract the DMD-spectrum. The recovery of the modal structures requires more computational care, the underlying concepts carry over from the POD analysis above. Other decompositions, such as sPOD, ICA, or resolvent analysis may be amenable to the same compressive treatment.

4. From modal analysis to Krylov subspaces

The above analysis has been data-based, relying on a sequence of flow fields generated by either a numerical simulation or experimental measurements. Alternative to this approach, quantitative flow analysis is often performed on the (linearized) operator that maps a given flow field onto its time rate of change. Analyses of this sort often employ iterative Krylov subspace techniques, such as the Arnoldi, GMRES, or BiCGStab algorithm.

In these efforts, Krylov subspaces provide an operator-aware basis for a subsequent projection onto a reduced state. The resulting operator in this reduced space is then taken as a proxy for the full operator, and all ensuing analysis is performed on the reduced operator. Krylov subspaces are generated by repeated application of a linear operator to a given starting flow field. We have

$$\mathcal{K} = \text{span} \{ \mathbf{q}_0, \mathbf{A}\mathbf{q}_0, \mathbf{A}^2\mathbf{q}_0, \dots, \mathbf{A}^{\ell-1}\mathbf{q}_0 \} \quad (4.1)$$

as a ℓ -dimensional Krylov subspace with \mathbf{A} symbolizing the linearized governing equations. Traditionally, this flow field sequence is orthonormalized and then used as a projection basis. Recognizing that \mathcal{K} consists of a snapshot sequence where the mutual relationship of its members encodes the spectral properties of \mathbf{A} , we can apply the FJLT, compress the spatial dimensions and preserve the spectral characteristics. Underlying this effort is the Rayleigh-Ritz procedure – a Galerkin approximation for eigenvalue problems – where the full column-space of the operator \mathbf{A} is approximated by a reduced basis set, say \mathbf{B} . The approximation is then given by the minimization problem

$$\min_{\mathbf{y} \in \mathbb{C}^n, \lambda \in \mathbb{C}} \|\mathbf{A}\mathbf{B}\mathbf{y} - \lambda\mathbf{B}\mathbf{y}\|_2 \quad \text{subject to} \quad \|\mathbf{B}\mathbf{y}\|_2 = 1. \quad (4.2)$$

For any eigenstructure $(\lambda_*, \mathbf{y}_*)$ of an associated matrix \mathbf{M}_* we have the equivalence

$$\|\mathbf{A}\mathbf{B}\mathbf{y}_* - \lambda_*\mathbf{B}\mathbf{y}_*\|_2 = \|(\mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{M}_*)\mathbf{y}_*\|_2. \quad (4.3)$$

where the matrix \mathbf{M}_* is the solution of the final optimization problem

$$\min_{\mathbf{M} \in \mathbb{C}^{n \times n}} \|\mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{M}\|_F. \quad (4.4)$$

From there the FJLT-reduced optimization problem can simply be stated as

$$\min_{\mathbf{M} \in \mathbb{C}^{n \times n}} \|\Phi(\mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{M})\|_F \quad (4.5)$$

and solved via standard linear-algebra routines. Starting from this above expression, a reduced Arnoldi algorithm can be formulated that produces modal structures and spectral information in the embedded subspace.

What remains to be determined is the choice of basis vectors, i.e., columns of \mathbf{B} . These vectors will be drawn from a Krylov subspace; however, in contrast to a standard Arnoldi method, we only *partially* orthogonalize the basis matrix \mathbf{B} . To this end, each incoming vector of the Krylov sequence is only orthonormalized against the last two (or more) vectors in \mathcal{K} . This last step, together with the FJLT-compression, is responsible for a fast build up of an approximate Krylov sequence and the fast execution of the Rayleigh-Ritz procedure. A more detailed analysis of this sketched procedure is required at this moment and will be postponed for a future effort. Preliminary results, however, show considerable promise and establish the FLJT as a prominent component in an accelerated flow analysis. The impact of FJLT (and variants thereof) on computations of eigenvalue and the solution of linear systems will be assessed in a subsequent investigation, with particular emphasis on implicit restarting and the use of effective preconditioners.

5. Conclusions and outlook

An isometric embedding technique has been proposed and explored as a tool to accelerate quantitative flow analysis based on snapshot sequences from simulations or experiments. The embedding takes advantage of the Johnson-Lindenstrauss lemma, together with an effective implementation, which reduces the spatial dimension of large-scale flow fields while preserving the temporal dynamics up to a user-specified distortion factor. The novel technique has been applied to Rayleigh-Bénard convection in a spherical shell at supercritical Rayleigh numbers and was able to perform a full POD-analysis based on only 3.4% of the degrees of freedom. Significant speedups could be reported, for both a POD- and a DMD-analysis of the underlying evolution problem. Future efforts will focus on applications to Krylov subspace techniques.

REFERENCES

- TAIRA, K., BRUNTON, S.L., DAWSON, S.T.M., ROWLEY, C.W., COLONIUS, T., MCKEON, B.J., SCHMIDT, O.T., GORDEYEV, S. THEOFILIS, V. & UKEILEY, L.S. 2017 Modal analysis of fluid flows: An overview. *AIAA J.* **55**(12), 4013–4041.
- CANDES, E.J. & WAKIN, M.B. 2008 An introduction to compressive sampling. *IEEE Signal Proc. Mag.* **25**(2), 21–30.

- CANDES, E.J. & TAO, T. 2006 Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory* **52**(12), 5406–5425.
- MARTINSSON, P.G. & TROPP, J.A. 2020 Randomized numerical linear algebra: foundations and algorithms. *Acta Numerica* **29**, 403–572.
- GLAZKOV, A. & SCHMID, P.J. 2023 Dynamics-preserving compression for modal flow analysis. *J. Fluid Mech.*, (in press).
- JOHNSON, W.B. & LINDENSTRAUSS, J. 1984 Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math.* **26**, 189–206.
- AILON, N. & CHAZELLE, B. 2009 The fast Johnson-Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.* **39**(1), 302–322.
- AILON, N. & CHAZELLE, B. 2010 Faster dimension reduction. *Comm. of ACM* **53**(2), 97–104.
- DASGUPTA, S. & GUPTA, A. 1999 An elementary proof of the Johnson-Lindenstrauss lemma. *UC Berkeley Tech. Rep.* **99-006**.
- TAKENS, F. 1981 Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence* 366–381.
- HIRANI, A.N. 2003 *Discrete exterior calculus*. PhD thesis, California Institute of Technology.
- DESBRUN, M., HIRANI, A.N., LEOK, M. & MARSDEN, J.E. 2005 Discrete exterior calculus. *arXiv preprint math/0508341*.
- MANTRAVADI, B., JAGAD, P. & SAMTANEY, R. 2023 A hybrid discrete exterior calculus and finite difference method for Boussinesq convection in spherical shells. *J. Comp. Phys.* **491**, 112397.
- BERKOOZ, G., HOLMES, P. & LUMLEY, J. 1993 The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* **25**, 539–575.
- C.W. ROWLEY AND I. MEZIĆ AND S. BAGHERI AND P. SCHLATTER AND D.S. HENNINGSON 2009 Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127.
- SCHMID, P.J. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28.
- SCHMID, P.J. 2010 Dynamic mode decomposition and its variants. *Annu. Rev. Fluid Mech.* **54**, 225–254.
- JOVANOVIĆ, M., SCHMID, P.J. & NICHOLS, J.W. 2014 Sparsity-promoting dynamic mode decomposition. *Phys. Fluids* **14**, 024103.