Modelling from Measurements

FINAL REPORT

Mazahir Saleem
11023230 | PHD IN ELECTRICAL ENGINEERING
POLITECNICO DI MILANO (POLIMI)

Introduction to KolSol

KolSol is a pseudospectral solution created especially to model the classical fluid dynamics issue known as Kolmogorov flow. The Fourier-Galerkin approach, as Canuto [1] explains, is used by this solver to solve partial differential equations, like the Navier-Stokes equations. A sinusoidal forcing term is applied to the system in Kolmogorov flow, which propels the fluid's development of turbulent behavior. KolSol employs a spectral approach to convert the governing equations into Fourier space, which facilitates faster computations, particularly for periodic boundary conditions.

KolSol provides both **NumPy** and **PyTorch** implementations, giving users the option to select between a PyTorch workflow that is autograd-compatible and a high-performance FFT computation using NumPy. Since gradients of the solution with respect to distinct parameters are frequently needed in machine learning framework integration, the PyTorch version allows for automatic differentiation.

In order to compute spatial derivatives and nonlinear terms, the solver discretises the velocity fields and evolves them over time using Fourier transforms. The two-dimensional Navier-Stokes equations control the system's evolution in the presence of an external sinusoidal forcing term. KolSol uses a time-stepping approach to simulate the flow dynamics and manages the pseudospectral decomposition of the velocity fields.

Data Generation

KolSol generates numerical solutions to the two-dimensional incompressible Navier-Stokes equations. Specifically, the solver computes the velocity components of fluid flow over time for each grid point in a two-dimensional domain. In this configuration, the x-component (u) and y-component (v) of the velocity fields are computed.

KolSol uses a grid of size $N \times N$, where N denotes the number of discretized points in each spatial dimension. This discretization is critical for understanding the flow dynamics in a two- or three-dimensional simulation.

For each time step, the solver determines the nonlinear interaction between velocity components, as well as the effects of viscosity and external forcing. These computations are carried out in Fourier space, utilizing the efficiency of Fast Fourier Transforms (FFT), and then converted back to physical space for interpretation.

The simulation creates two arrays of velocity fields, one for each component (u and v), that evolve over a set number of time steps. This data can be used to analyze flow characteristics, visualize system behavior, or provide input for machine learning models in scientific computing workflows.

Note: you may find the code and visualization of data generation with the file name "DataGeneration_KolSol.ipynb".

Methodology

Several approaches can be used to analyze and handle the fluid dynamics data supplied by KolSol. Some of these are:

- Singular Value Decomposition (SVD)
- Randomized SVD
- Dynamic Mode Decomposition (DMD)

- Extended Dynamic Mode Decomposition (EDMD)
- Sparse Identification of Nonlinear Dynamics (SINDy)
- Gaussian Mixture Models (GMMs)

Each approach will be discuss with the results and commentary here after.

Singular Value Decomposition (SVD)

SVD is an exceptionally efficient method for reducing dimensionality and extracting features. It decomposes the data matrix into singular values and vectors, allowing dominant modes to be identified. This is especially beneficial for simplifying large-scale fluid dynamics simulations while preserving important flow structures.

SVD can be employed on velocity fields to derive low-rank approximations that capture the most important flow features (e.g., coherent structures) while decreasing noise. It is Simple, computationally efficient for medium-sized datasets, and generally applicable but SVD does not explicitly incorporate dynamics, it cannot directly capture time-varying patterns.

Note: you may find the code and results/plots with the file name “SVD_DR.ipynb”.

The first plot depict the X-velocity (u) and Y-velocity (v) components at the final time step. Both velocity components have complicated flow topologies, which suggest turbulent flow. The colour distribution indicates regions of high and low velocities, with clear spatial fluctuations indicating precise and realistic simulations. The inferno color map efficiently emphasizes magnitude disparities in the field.

The plot of Singular Values depicts the strength of each mode obtained from the SVD. Singular values represent the amount of variance that each mode adds to the entire dataset. The abrupt decrease of singular values indicates that the dataset is extremely compressible, which implies that the system's behavior is governed by a few key modes.

The first few singular values are substantially larger than the remainder, indicating that dimensionality reduction can be effectively implemented. The majority of the dynamics are captured in the first few settings. The quick decline of singular values is a common feature in fluid simulations like this one, where the first few modes account for most of the variance, while the remaining modes account for finer details or noise.

The reconstructed fields retain the overall structure and behavior of the original velocity fields, particularly the prominent flow patterns. The finer nuances in the flow, however, are smoothed down because just ten modes are used to depict the system. The reduction in dimensionality is likely to result in the loss of some small-scale turbulence and complicated flow patterns.

Randomized Singular Value Decomposition (RSVD)

Randomized SVD is a faster, approximate version of classical SVD, making it more suited for huge datasets. Randomized SVD can be used to reduce high-dimensional flow data generated across multiple time steps quickly and accurately. It can help to compress data for storage or analysis. Randomized SVD is faster for large datasets and preserves the benefits of SVD while being slightly less accurate due to approximation.

Note: you may find the code and results/plots with the file name “RSVD.ipynb”.

The plot of singlular values indicates the importance or "strength" of each component. Typically, the first few components have greater singular values, representing the most of the volatility in the data. This means they represent the most important patterns in the velocity field. As the singular values decline, the components capture finer, less dominating information, which may represent noise or small-scale flow changes.

The results of the Randomized SVD on the u and v components of the velocity field show that the first five dominating components capture the fundamental flow patterns across the grid. The leading components exhibit large-scale structures that reflect significant flow directions and consistent vortical patterns, with u components indicating horizontal or directed flows and v components indicating vertical variations. Each succeeding component captures finer characteristics, such as smaller-scale turbulence or periodic flow changes.

Dynamic Mode Decomposition (DMD)

DMD is an excellent technique for recording time-evolving patterns and breaking down the flow into spatial modes with appropriate temporal dynamics. It finds prevalent patterns in the flow that change linearly over time.

DMD can be used to detect dominating flow features and track their temporal evolution in Kolmogorov flows. It can also be used to analyze recurrent behavior in the flow field and as a technique for model reduction. Dynamically meaningful decomposition captures temporal dynamics but It assumes linearity in time and may miss sophisticated nonlinear interactions.

Note: you may find the code and results/plots with the file name “DMD_kol.ipynb”.

Mode 1 in the u-component captures prominent and large-scale structures in the horizontal flow field (along the u-axis). This mode most likely represents the horizontal flow's strongest and most dominant component.

Mode 2 in the u-component catches smaller, less dominating flow characteristics. These could include horizontal oscillations, smaller ripples, or higher-dynamic localized zones that are not as steady or vast as those catches in Mode 1.

Mode 1 in the v-component catches the major vertical flow structures, which are similar to Mode 1 in the u-component but in the vertical direction. This mode most likely represents the biggest vertical motions or the most stable component of vertical dynamics.

Mode 2 catches finer vertical dynamics, such as oscillations, as well as smaller vertical structures like vortices or vertical eddy. This mode is typically associated with less prominent but nonetheless substantial structures.

Extended Dynamic Mode Decomposition (EDMD)

Extended DMD improves regular DMD by adding nonlinear effects utilizing a larger set of observables, making it better suited to capturing nonlinear fluid dynamics. Extended DMD is effective for Kolmogorov flow data with strong nonlinear interactions. It is more capable to adequately capturing the intricacy of turbulent flows than standard DMD.

Extended DMD is more effective at handling nonlinear dynamics than DMD. However, the computational cost is higher, and the expanded observables may need to be carefully tuned.

Note: you may find the code and results/plots with the file name “ExtendedDMD_kolsol.ipynb”.

EDMD can identify both linear and nonlinear dynamics in the u and v components. EDMD captures horizontal flow structures in u, and vertical structures in v. The polynomial basis enables EDMD to uncover more complicated temporal behaviors and interactions among different segments of the flow.

In both cases (u and v), the first mode captures the most dominant, large-scale flow patterns, whereas the second mode focusses on smaller, transitory, or fluctuating dynamics. Together, these modes give a complete picture of the system dynamics.

When comparing the efficiency and effectiveness of Dynamic Mode Decomposition (DMD) and Extended Dynamic Mode Decomposition (EDMD) , DMD outperforms EDMD in terms of computational efficiency, providing faster processing and lower resource demand while capturing the system's dominant linear dynamics. However, EDMD, despite being computationally more expensive due to its polynomial basis expansion, provides a richer, more thorough insight of the system's behavior, particularly in capturing nonlinear interactions and finer-scale structures. For systems with considerable nonlinear dynamics, such as the KOLSOL data, EDMD provides greater accuracy and depth, making it the preferable option when a thorough understanding of the system's dynamics is required.

Sparse Identification of Nonlinear Dynamics (SINDy)

SINDy is useful for determining the fundamental governing equations of a dynamical system in sparse form. It creates interpretable models by selecting the most significant terms to describe the system. SINDy can be used with Kolmogorov flow data to generate a reduced-order Navier-Stokes model. It's very useful for determining the main physics governing the flow. SINDy generates interpretable models that are useful for system identification, although it may struggle with very high-dimensional or noisy data.

Note: you may find the code and results/plots with the file name “SINDy.ipynb”.

The SINDy results for the Kolmogorov flow data demonstrate the model's ability to represent the sparse, underlying dynamics that control fluid movement. SINDy isolates the most relevant flow patterns by finding essential terms in differential equations, making it easier to depict complicated, nonlinear behaviors. This reduction displays a distinct structure in Kolmogorov flow dynamics, demonstrating the model's ability to identify the primary forces and interactions that shape the u and v velocity components.

The identified equations for u and v describe the main components of the Kolmogorov flow, such as convection and diffusion terms, which are critical for characterizing fluid dynamics. The simplicity of the obtained equations implies that SINDy was successful in removing less relevant terms, leaving just those with a significant influence. This result gives a compact and interpretable model of the Kolmogorov flow, allowing for future investigation and perhaps reducing computational needs in fluid simulations.

Gaussian Mixture Models (GMMs)

Gaussian Mixture Models (GMMs) are probabilistic models that find clusters or components in data. They can simulate complex distributions and handle uncertainty. GMM can be used to group distinct flow patterns or regimes in data, such as identifying turbulent vs. laminar regions of flow. GMMs are flexible and can accommodate mixed distributions, but they are computationally expensive and sensitive to initialization.

Note: you may find the code and results/plots with the file name "GMMs.ipynb".

The Gaussian Mixture Model (GMM) clustering findings for the velocity field data exhibit diverse patterns and insights as the number of clusters is increased.

With 3 clusters, the plot depicts relatively large, contiguous patches of identical colors. The grouping appears to encompass extensive areas in the velocity field, most likely representing large-scale structures or discrete flow regions with comparable properties. This clustering method is appropriate for finding generic zones of various flow patterns, such as regions with high or low velocities in the u and v components.

The figure with 5 clusters displays smaller, more detailed regions than the 3-cluster result. More subtle fluctuations in the velocity field are captured, most likely representing finer flow patterns or variations. This clustering level can aid in detecting more localized locations with similar flow behavior within the overall patterns seen in the 3-cluster model.

Increasing the number of clusters to 10 improves the clustering results by increasing the granularity of the areas. Each cluster now represents smaller zones with similar velocity characteristics. This level of clustering can expose sub-regions within flow patterns, allowing for better details of turbulence or other small-scale features in the velocity field.

With 20 clusters, the plot becomes quite granular, with several small, dispersed regions of various colors. The clustering now catches extremely tiny, local differences in the velocity field. This large number of clusters may represent temporary or modest fluctuations in flow, which is useful for comprehensive examination of local flow aspects. However, the significance of each cluster may diminish, as some clusters may represent noise or very small fluctuations rather than cohesive flow patterns.

References

[1] Canuto, C. (1988) Spectral methods in fluid dynamics. New York: Springer-Verlag.

…………………………………………..……*Thanks* ☺…………………………………………………….