Hierarchical Clustering of Modes in Numerical Turbulence Fields

Akifumi OKUNO[1,2]*, Takumi KODAHARA[3], Makoto SASAKI[3,4]

[1] *The Institute of Statistical Mathematics, Tachikawa 190-0014, Japan*
[2] *RIKEN Center for Advanced Intelligence Project, Nihon-bashi 103-0027, Japan*
[3] *College of Industrial Technology, Nihon University, Narashino 274-0072, Japan*
[4] *Research Center for Plasma Turbulence, Kyushu University, Kasuga 816-8580, Japan*

(Received 29 September 2024 / Accepted 9 October 2024)

We propose an automated method for decomposing turbulence, where the modes are derived by applying singular value decomposition to the electrostatic potential field in turbulence simulations. After classifying the modes into zonal flows and turbulence based on azimuthal integration, the turbulence is further decomposed into finer components through hierarchical clustering. This approach systematically breaks down the turbulence into hierarchical levels, providing flexible control over the degrees of freedom in the mode decomposition process.

Keywords: turbulence, electrostatic potential field, data-driven approach, clustering

DOI: 10.1585/pfr.19.1201035

The performance of fusion plasmas is dominated by the nonlinear evolution of turbulence driven by plasma gradients. The multi-scale turbulence structures are generated [1–3], which determine particle and heat transport. Thus, it is crucial to understand the fundamental nonlinear process, while the turbulence has a large degree of freedom, which complicates the physical interpretation. Methods to reduce these degrees of freedom are urgently needed.

There has been rapid progress in data-driven scientific methods, and their application to plasma turbulence has started. The applications of dynamic mode decomposition [4, 5] and of neural networks [6] to the turbulence study have been reported. The orthogonal basis functions can be derived only from data by using singular value decomposition (SVD) so that the energy transfer analysis is possible based on the SVD [7]. Extension of the SVD to multi-fields has been proposed [8, 9]. However, the large number of resulting modes from SVD makes it still challenging to analyze the details of the elementary processes. The methods that can systematically group the essential modes has been required.

This study proposes clustering modes based on their spatial-temporal behavior. We use a set of a turbulence simulation data [10], and apply SVD to the electrostatic potential field, to obtain the initial modes. We systematically and automatically group the modes from SVD by applying hierarchical clustering [11], providing a more automated alternative to the manual classification approach used in [7]. This approach is also anticipated to allow for flexible selection of the degrees of freedom.

As described, our analysis automates and further develops the process presented in [7], which manually classifies the modes obtained through SVD of the electrostatic potential field $\phi(x, y, t)$ into four groups, where (x, y) represent the Cartesian coordinates and t denotes time. The potential field is decomposed as

$$\phi(x, y, t) = \sum_{j \geq 0} s_j \Psi_j(x, y) h_j(t), \tag{1}$$

where $\Psi_j(x, y)$ describes the spatial structure of the jth mode, and $h_j(t)$ captures its temporal evolution. The singular value s_j represents the importance of the contribution of the jth mode $\phi_j(x, y, t) = s_j \Psi_j(x, y) h_j(t)$ to $\phi(x, y, t)$. We assume that s_j is in descending order ($s_0 \geq s_1 \geq s_2 \geq \cdots$), as is common in many computational implementations. Refer to the right section of Fig. 1, which displays the spatial structures $\{\Psi_j\}$.

In our analysis, we first remove the background mode ($j = 0$) having the largest singular value. Subsequently, we classify the remaining modes into zonal flows and turbulence by integrating the absolute differences in the obtained spatial structures along the azimuthal direction. A mode is classified as a zonal flow if the integral yields a small value, while the rest are termed turbulence. Herein, $\mathcal{T} \subset \{0, 1, 2, ...\}$ denotes the mode indices of the turbulence specified.

After identifying the turbulence, we further decompose the overall turbulence indexed by $\mathcal{T}$ into finer components indexed by $\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_C$, where $\{\mathcal{T}_c\}$ satisfies the conditions: $\max\ \mathcal{T}_c < \min\ \mathcal{T}_{c'}$ for any $c < c'$ and $\bigcup_c \mathcal{T}_c = \mathcal{T}$. See Fig. 1 for illustration of the division with $C = 3$. As the singular values represent the contribution of the corresponding structures to the potential field $\phi(x, y, t)$, the turbulence composed of the index subset $\mathcal{T}_c$ is expected to have more significance than that of other subsets $\mathcal{T}_{c'}$ for any $c' > c$.

A simple implementation to decompose the turbulence

*Corresponding author's e-mail: okuno@ism.ac.jp

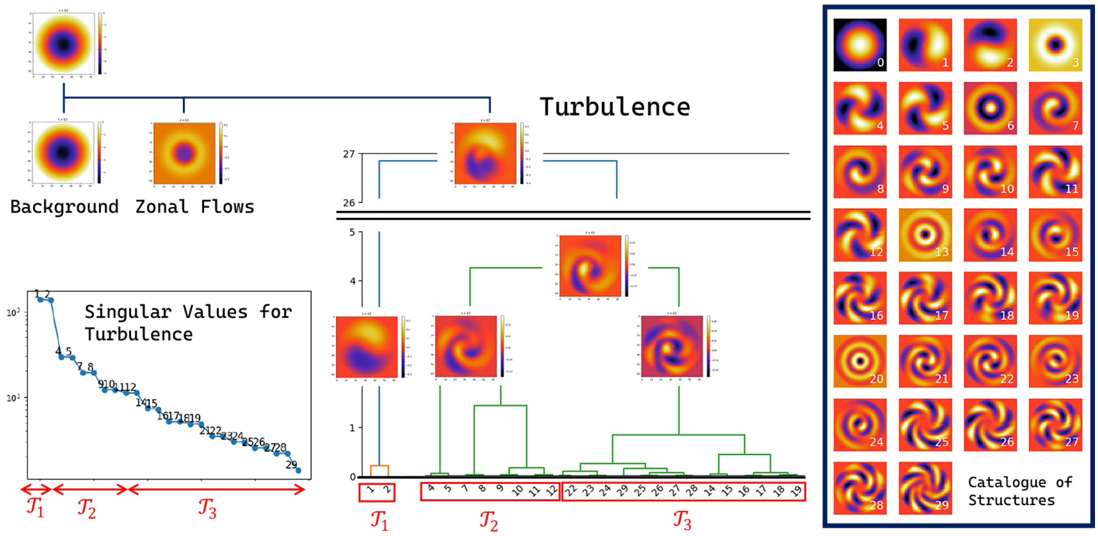

Fig. 1. Decomposition of turbulence via hierarchical clustering.

(indexed by $\mathcal{T}$) into such finer components (indexed by $\mathcal{T}_1, \mathcal{T}_2, ...\mathcal{T}_C$) can be achieved by applying hierarchical clustering with the Ward method [11] to the dissimilarity matrix $D = (d_{i_1 i_2})$, where $d_{i_1 i_2}$ denotes the absolute difference between singular values:

$$d_{i_1 i_2} = |s_{i_1} - s_{i_2}|, \quad i_1, i_2 \in \mathcal{T}. \tag{2}$$

This hierarchical clustering method successively merges or splits clusters based on the dissimilarity matrix $D = (d_{i_1 i_2})$, resulting in a tree-like dendrogram that reveals relationships among data points at various levels of granularity. Refer to the dendrogram in Fig. 1, which clearly demonstrates the division of turbulence into multiple levels.

Finally, for each subset $\mathcal{T}_c$, we visualize the accumulated modes $\tilde{\phi}_c(x, y, t) := \sum_{j \in \mathcal{T}_c} s_j \Psi_j(x, y) h_j(t)$ in Fig. 1. This method allows for a hierarchy of turbulence in terms of the number of degrees of freedom that we want to decompose. In conventional turbulence studies, turbulence is often modeled with three degrees of freedom (background field, zonal flow, and turbulence) [1]. By using the presented method, turbulence can be further divided into a hierarchical structure. In the case of Fig. 1, the turbulence can be divided into linear unstable modes $\mathcal{T}_1$ and others. The other nonlinear modes are decomposed into $\mathcal{T}_2$ and $\mathcal{T}_3$, whose combination represents the generation and annihilation of spiral structures. Although Fig. 1 illustrates the decomposition of turbulence into $C = 3$ modes, we can further subdivide it into finer modes by following the tree downwards. In addition to these visualizations, animation of the decomposed turbulence and the source codes to reproduce the experimental results are available at [12].

In summary, the automated method for decomposing turbulence is proposed, where the modes are derived by applying the SVD to the simulation data. The turbulence is decomposed into finer components through hierarchical clustering. This approach systematically breaks down the turbulence into hierarchical levels with any number of degrees of freedom.

This work was partly supported by JSPS KAKENHI Grant Nos. JP21K03509, 24K06997, 21K17718, 22H05106, the “Strategic Research Projects” grant 2022-SRP-13 from Research Organization of Information and Systems (ROIS), the collaboration programs of NIFS (NIFS23KIIT002, NIFS24KIPS005, NIFS23KIST044), and the Collaborative Research Program of Research Institute for Applied Mechanics, Kyushu University (2024S2-CD-2).

[1] P.H. Diamond *et al.*, Plasma Phys. Control. Fusion **47**, R35 (2005).
[2] T. Yamada *et al.*, Nat. Phys. **4**, 721 (2008).
[3] H. Arakawa *et al.*, Sci. Rep. **6**, 33371 (2016).
[4] M. Sasaki *et al.*, Plasma Phys. Control. Fusion **61**, 112001 (2019).
[5] A. Kusaba *et al.*, Jpn. J. Appl. Phys. **61**, SA1011 (2021).
[6] Y. Jajima *et al.*, Plasma Phys. Control. Fusion **65**, 125003 (2023).
[7] M. Sasaki *et al.*, Plasma Phys. Control. Fusion **63**, 025004 (2020).
[8] T. Kodahara *et al.*, Plasma Fusion Res. **18**, 1202036 (2023).
[9] G. Yatomi *et al.*, Plasma Phys. Control. Fusion **65**, 095014 (2023).
[10] M. Sasaki *et al.*, Phys. Plasmas **26**, 042305 (2019).
[11] J.H. Ward Jr., J. Am. Stat. Assoc. **58**, 236 (1963).
[12] A. Okuno, https://doi.org/10.5281/zenodo.13916918