Discovering Multiscale Structure Using Data-Driven Wavelets

Daniel Floryan and Michael D. Graham

Department of Chemical and Biological Engineering, University of Wisconsin-Madison

Multiscale structure is everywhere

Many processes in science and engineering have important features at multiple scales of time and/or space, including biological tissues, active matter, oceans, networks, speech, and images. Identifying the multiscale features of these processes is crucial to our understanding of them, and will ultimately allow us to better control the processes.

Here, we give focus to turbulence, which greatly impacts commercial flight, the transport of oil through pipelines, and weather systems. The image below shows one instance in time of a simulation of turbulence. The colouring shows the kinetic energy of the fluid, with darker colouring corresponding to greater energy. There are apparently localized features at many scales; how can we extract the building blocks of these features in a rational manner?

How can we extract multiscale structure from data?

A primary method is principal components analysis (PCA). Given an ensemble of data, PCA yields an orthogonal basis whose elements are optimally ordered by energy content. The basis elements tend to have global support, whereas we seek localized multiscale features. Thus, PCA seems ill-suited to multiscale data.

Wavelets, on the other hand, produce orthogonal bases whose elements have local support. The basis elements are translations and dilations of a single template called the mother wavelet. Wavelets are localized not only in space, but also in scale, so they seem perfectly suited for the analysis of multiscale data. However, the mother wavelet is prescribed a priori, so traditional wavelets cannot extract structure from data. Can we combine the data- and energy-driven nature of PCA with the multiscale suitability of wavelets?

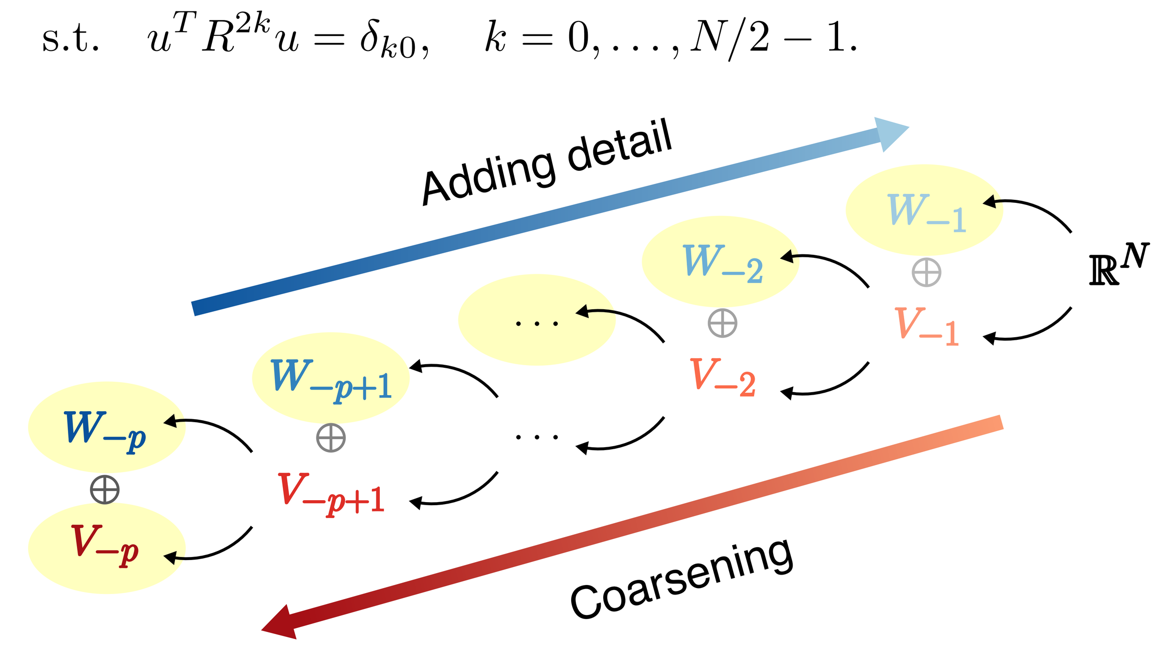
Data-driven wavelets use energy principles and wavelet hierarchies

As sketched below, wavelets split a space into orthogonal subspaces, one that provides a coarse approximation to the space (red), and one that fills in the missing details (blue). The splitting is performed recursively on the coarse subspaces. The original space can be expressed as a direct sum of the coarsest subspace (which usually gives the spatial mean) and all the detail subspaces.

With data-driven wavelets, we optimize every subspace while encouraging the basis elements to have local support. Since large scales tend to have the most energy, we maximize the amount of a dataset's energy contained in the coarse subspaces. We cast the optimization problem at each scale in terms of the wavelet generator, u, the data at that scale, Z, and the shift operator, R. What results is a combination of the hierarchical nature of wavelets with the energy principle of PCA.

$$\max_u \quad u^T A u - \lambda^2 \mathrm{Var}(u), \quad A = \frac{1}{\|Z\|_F^2} \sum_{k=0}^{N/2-1} R^{-2k} Z Z^T R^{2k}$$

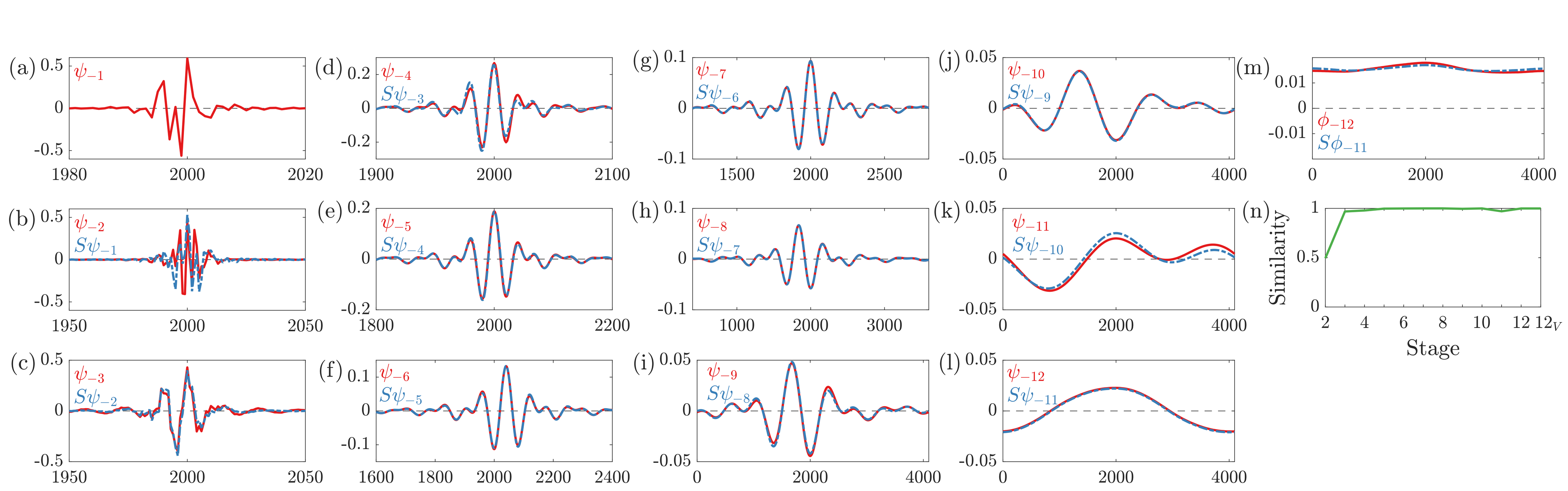
$$\text{s.t.} \quad u^T R^{2k} u = \delta_{k0}, \quad k = 0, \ldots, N/2-1.$$



In turbulence, data-driven wavelets reveal a range of self-similar scales

We have computed data-driven wavelets for a turbulence dataset, among others. Turbulence is notoriously difficult to analyze, so this dataset provides a strenuous test of our method. The dataset consists of the fluid velocity along randomly sampled lines in the cube of turbulence, each sample being a vector of length 4096, yielding 12 dyadic scales.

We show the wavelets at each scale (red) as well as re-scaled versions of wavelets from the preceding scale (blue). The computed wavelets are remarkable for the fact that the middle scales are nearly self-similar. Importantly, self-similarity is not built into the method; it is extracted from the data. Self-similarity in an intermediate range of scales was hypothesized 80 years ago, and this is the first time that localized, self-similar structures have been extracted from turbulence. We think of these structures as the building blocks of turbulence.



For more details, see arXiv:2009.00682

Data obtained from the JHTDB at http://turbulence.pha.jhu.edu

Funded by AFOSR grant FA9550-18-1-0174 and ONR grant N00014-18-1-2865 (Vannevar Bush Faculty Fellowship)