**AFRL-AFOSR-UK-TR-2022-0023**

---

# l1-based sparsification of reduced order models of high Reynolds number turbulent flows

**Da Ronch, Andrea**
**UNIVERSITY OF SOUTHAMPTON**
**UNIVERSITY RD**
**SOUTHAMPTON, HAMPSHIRE, SO17 1BJ**
**GBR**

---

**02/17/2022**
**Final Technical Report**

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515 Box 14, APO AE 09421

# REPORT DOCUMENTATION PAGE

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.**

| | | | |
|---|---|---|---|
| **1. REPORT DATE**<br>20220217 | **2. REPORT TYPE**<br>Final | **3. DATES COVERED**<br>**START DATE**<br>20170815 | **END DATE**<br>20210814 |

**4. TITLE AND SUBTITLE**
l1-based sparsification of reduced order models of high Reynolds number turbulent flows

| | | |
|---|---|---|
| **5a. CONTRACT NUMBER** | **5b. GRANT NUMBER**<br>FA9550-17-1-0324 | **5c. PROGRAM ELEMENT NUMBER**<br>61102F |
| **5d. PROJECT NUMBER** | **5e. TASK NUMBER** | **5f. WORK UNIT NUMBER** |

**6. AUTHOR(S)**
Andrea Da Ronch

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
UNIVERSITY OF SOUTHAMPTON
UNIVERSITY RD
SOUTHAMPTON, HAMPSHIRE SO17 1BJ
GBR

**8. PERFORMING ORGANIZATION REPORT NUMBER**

| | | |
|---|---|---|
| **9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**<br>EOARD<br>UNIT 4515<br>APO AE 09421-4515 | **10. SPONSOR/MONITOR'S ACRONYM(S)**<br>AFRL/AFOSR IOE | **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**<br>AFRL-AFOSR-UK-TR-2022-0023 |

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
A Distribution Unlimited: PB Public Release

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

From a physical standpoint, turbulence is a multi-scale phenomenon exhibiting a wide hierarchy of spatial and temporal scales. This property, coupled with the intrinsic nonlinearity of the governing equations, poses considerable difficulties to its modelling and analysis. One of the major challenges to obtain a satisfactory mathematical description of this phenomenon arises from the fact that the dynamics of flow structures at a particular length or time scale cannot be examined in isolation without also considering at the same time the whole hierarchy of complementary scales [Domaradzki et al., 1994]. In fact, nonlinear interactions between triads of scales play a fundamental role as they are the main driver of energy transfer between coherent structures [Pope, 2001, Moffatt, 2014]. In turn, the organisation of triadic interactions has a direct influence on the physics of a number of flow phenomena, such as direct and inverse energy cascades [Kolmogorov, 1991] or transition to turbulence [Craik, 1971, Rempfer and Fasel, 1994a,b] in different flow configurations [Schmidt, 2020]. Overall, this property makes the development of computationally efficient and physically-interpretable reduced order dynamical models a challenging task.

Historically, the study of triadic interactions has been conducted by first employing an appropriate decomposition technique to educe coherent structures from the turbulent motion and then characterising the intensity of the inter-modal couplings through the perspective of the resulting Galerkin model [Noack et al., 2008]. One of the key findings of such studies is that energy transfers are not uniformly distributed in modal space. In fact, not all interactions have the same importance and energy flows along preferential directions. Specifically, there is evidence suggesting that the nonlinear interaction pattern among coherent structures is often sparse.

The aim of this project has been to develop model order reduction methods that exploit this intrinsic property of turbulent flows. Fundamentally, our objective was to be able to generate models of complex flows that resolve a wide range of scales, but are at the same time computationally efficient and interpretable. To this end, we have borrowed tools and techniques from machine learning and optimisation, to develop two distinct techniques presented in this report.

**15. SUBJECT TERMS**

| | | | | |
|---|---|---|---|---|
| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT**<br>SAR | **18. NUMBER OF PAGES**<br>37 |
| **a. REPORT**<br>U | **b. ABSTRACT**<br>U | **c. THIS PAGE**<br>U | | |

| | |
|---|---|
| **19a. NAME OF RESPONSIBLE PERSON**<br>DOUGLAS SMITH | **19b. PHONE NUMBER** *(Include area code)*<br>314 235 6013 |

# $l_1$-based Sparsification of Reduced Order Models of High-Reynolds-Number Vortical Flows

Andrea da Ronch (PI), Davide Lasagna (Co-I), Riccardo Rubini (PhD)

Grant Number FA9550-17-1-0324

Final Performance Report

February 2022

# Contents

# List of Figures

# List of Tables

# 1 Summary

From a physical standpoint, turbulence is a multi-scale phenomenon exhibiting a wide hierarchy of spatial and temporal scales. This property, coupled with the intrinsic nonlinearity of the governing equations, poses considerable difficulties to its modelling and analysis. One of the major challenges to obtain a satisfactory mathematical description of this phenomenon arises from the fact that the dynamics of flow structures at a particular length or time scale cannot be examined in isolation without also considering at the same time the whole hierarchy of complementary scales [Domaradzki et al., 1994]. In fact, nonlinear interactions between triads of scales play a fundamental role as they are the main driver of energy transfer between coherent structures [Pope, 2001, Moffatt, 2014]. In turn, the organisation of triadic interactions has a direct influence on the physics of a number of flow phenomena, such as direct and inverse energy cascades [Kolmogorov, 1991] or transition to turbulence [Craik, 1971, Rempfer and Fasel, 1994a,b] in different flow configurations [Schmidt, 2020]. Overall, this property makes the development of computationally efficient and physically-interpretable reduced-order dynamical models a challenging task.

Historically, the study of triadic interactions has been conducted by first employing an appropriate decomposition technique to educe coherent structures from the turbulent motion and then characterising the intensity of the inter-modal couplings through the perspective of the resulting Galerkin model [Noack et al., 2008]. One of the key findings of such studies is that energy transfers are not uniformly distributed in modal space. In fact, not all interactions have the same importance and energy flows along preferential directions. Specifically, there is evidence suggesting that the nonlinear interaction pattern among coherent structures is often sparse.

The aim of this project has been to develop model order reduction methods that exploit this intrinsic property of turbulent flows. Fundamentally, our objective was to be able to generate models of complex flows that resolve a wide range of scales, but are at the same time computationally efficient and interpretable. To this end, we have borrowed tools and techniques from machine learning and optimisation, to develop two distinct techniques presented in this report.

## 1.1 Publications resulting from this project

This research project has led to three journal publications, two submitted to the Journal of Fluid Mechanics and one to the Chinese Journal of Aeronautics:

1. Rubini, R., Lasagna, D. & Da Ronch, A. (2020). The l1-based sparsification of energy interactions in unsteady lid-driven cavity flow. Journal of Fluid Mechanics, 905, A15.
   https://doi.org/10.1017/jfm.2020.707

2. Rubini, R., Lasagna, D., Da Ronch, A. (2021). l1-based calibration of POD-Galerkin models of two-dimensional unsteady flows, Chinese Journal of Aeronautics, 34(1)
   https://doi.org/10.1016/j.cja.2020.10.009

3. Rubini, R., Lasagna, D. & Da Ronch, A. (2021). *A-priori* sparsification of Galerkin based reduced order models. Journal of Fluid Mechanics, under review
   https://arxiv.org/abs/2105.06753

Two papers on the "a-posteriori" sparsification, using $l_1$-based regression, have been accepted, while the paper on the "a-priori" sparsification is currently under review. Dr Riccardo Rubini, the PhD student funded by the project, has carried out most of the research and has defended successfully his PhD on December, 08 2021.

# 2 Introduction

In the classical description of developed turbulent flows [Lumley, 1979, Pope, 2001, Jiménez, 2018], energy is transferred across the hierarchy of coherent structures via nonlinear triadic interactions. Implicit in this picture is the fact that not all interactions have the same importance, but they occur in preferential patterns. In fact, extensive numerical evidence suggests that the nonlinear interaction pattern among coherent structures is sparse. The evolution of structures at a certain length scale depends predominantly upon a subset of all other structures [Kraichnan, 1971, Ohkitani, 1990, Brasseur and Wei, 1994] and the influence of interactions with the complementary set of structures can be generally neglected with minor global effects.

Successful attempts to construct a reduced set of equations that exploit this sparsity have been made in the past, often for canonical geometries where triadic interactions are conveniently examined in Fourier space and using a coarse-grained partitioning of the hierarchy of scales. Laval et al. [1999] considered two-dimensional homogeneous decaying turbulence and developed a reduced set of coupled partial differential equations governing the evolution of the large and small scales. In this model, only dominant terms were retained based on observations from direct numerical simulation. With the goal of identifying fundamental mechanisms underlying wall turbulence, Thomas et al. [2015] developed nonlinear reduced models of plane Couette flow directly from the governing equations by first partitioning the flow into a streamwise-averaged mean and a perturbation field, and then neglecting nonlinear interactions among the streamwise varying perturbations, i.e. the perturbation-perturbation nonlinearity [Thomas et al., 2014].

When reduced-order dynamical representations are derived using Galerkin projection on a low-dimensional subspace identified by a set of modal structures [Fletcher, 1984, Rowley and Dawson, 2017], triadic interactions are conveniently studied in modal space by examining a third-order coefficient tensor arising from projection of the basis function on the convective term of the Navier-Stokes equations [Noack et al., 2008, 2011]. Sparsity characteristics have also been observed in this reduced-order setting. Couplet et al. [2003] constructed Galerkin models of the separated turbulent flow past a backward-facing step using Proper Orthogonal Decomposition (POD) modes [Lumley, 1970, Sirovich, 1987] and observed that the energy transfer pattern in modal space shares many properties with its counterpart in isotropic homogeneous three-dimensional turbulent flows [Yeung et al., 1995]. For instance, the authors observed that interactions are local in modal space and that a direct energy cascade exists. Analogously, Rempfer and Fasel [1994a] examined the power budget of POD modes in a transitional boundary layer and observed that interactions in modal space occur predominantly between triads of modes whose sum of modal indices is equal to zero, similar to energy interactions between Fourier modes in homogeneous turbulence.

However, classical model order reduction techniques [Rowley and Dawson, 2017] have not traditionally exploited this feature. In fact, when modal decompositions such as POD are employed, densely-connected models are usually obtained, as the third-order coefficient tensor is dense (i.e. most coefficients are different from zero) for inhomogeneous flows without particular symmetries. This hinders the interpretation of the underlying physics of scales interactions and increases computational costs, since all triadic interactions have to be evaluated for advancing the model forward in time.

## 2.1 Objectives

The overarching aim of the present project is to develop model order reduction techniques that exploit the intrinsic sparsity of scale interactions in fluid flows. Fundamentally, the aim is to generate reduced order models that resolve a wide range of scales while preserving computational efficiency and interpretability. Other researchers have applied sparsity-coding methods on either low-dimensional fluid systems or flows with simple temporal behaviour (e.g. periodic flows), e.g Loiseau and Brunton [2018]. In these systems, interactions between coherent structures are simpler to analyse and rationalise. Since the inception of the project, our interest has been to work on fluid systems characterised by a large number of degrees of freedom and chaotic temporal behaviour. In such systems, scale interactions are more complex, which is ideal to push the envelope of the statistical methods employed. To achieve this goal, our three-year research project has focused on developing two different approaches, borrowing computational methods from statistics and optimisation theory. These techniques are described in section 3. More details can be found in the publications that have been prepared during the course of this project.

# 3 Methods, Assumptions and Procedures

## 3.1 Reduced Order Modelling

Before introducing the two sparsification methods, we report here some of the required notation and background concepts. We consider a space of square integrable velocity vector fields defined over a spatial domain $\Omega$, endowed by the standard inner product

$$(\mathbf{u}, \mathbf{v}) \coloneqq \int_{\Omega} \mathbf{u} \cdot \mathbf{v} \mathrm{d}\Omega, \tag{1}$$

where $\mathbf{u}, \mathbf{v}$ are two elements of such space. The resulting $\mathcal{L}^2(\Omega)$ norm is denoted as $||\mathbf{u}|| = \sqrt{(\mathbf{u}, \mathbf{u})}$. Using the time averaged velocity field $\bar{\mathbf{u}}(\mathbf{x})$ as a base flow, and denoting by $\mathbf{u}'(t, \mathbf{x})$ the velocity fluctuation $\mathbf{u}(t, \mathbf{x}) - \bar{\mathbf{u}}(\mathbf{x})$, an $N$-dimensional expansion expressed by the ansatz

$$\mathbf{u}(t, \mathbf{x}) = \bar{\mathbf{u}}(\mathbf{x}) + \mathbf{u}'(t, \mathbf{x}) = \bar{\mathbf{u}}(\mathbf{x}) + \sum_{i=1}^{N} a_i(t) \boldsymbol{\phi}_i(\mathbf{x}), \tag{2}$$

is introduced to describe the space-time velocity field, where $a_i(t)$ and $\boldsymbol{\phi}_i(\mathbf{x})$, $i = 1, \ldots N$ are the temporal and global spatial modes, respectively, with $\|\boldsymbol{\phi}_i(\mathbf{x})\| = 1$. These modes may be computed a posteriori from numerical or experimental data or a priori from a characteristic operator of the system [Taira et al., 2017] or from completeness considerations [Noack and Eckelmann, 1994]. Reduced order models are then derived by projecting the governing equations onto the subspace defined by the modes [Rowley and Dawson, 2017]. Restricting our analysis to configurations where the boundaries are either no-slip walls or periodic, this procedure results in an autonomous system of coupled nonlinear ordinary differential equations (ODEs)

$$\sum_{j=1}^{N} \mathsf{M}_{ij} \dot{a}_j(t) = \tilde{\mathsf{C}}_i + \sum_{j=1}^{N} \tilde{\mathsf{L}}_{ij} a_j(t) + \sum_{j=1}^{N} \sum_{k=1}^{N} \tilde{\mathsf{Q}}_{ijk} a_j(t) a_k(t), \qquad i = 1 \ldots, N, \tag{3}$$

defining the temporal evolution of the coefficients $a_i(t)$. Here, we only report the definitions of the quadratic coefficients

$$\tilde{\mathsf{Q}}_{ijk} = (\boldsymbol{\phi}_i, \boldsymbol{\phi}_j \cdot \nabla \boldsymbol{\phi}_k), \tag{4}$$

while expressions for the tensors $\tilde{\boldsymbol{C}}$ and $\tilde{\boldsymbol{L}}$ can be found in Noack et al. [2011]. The matrix $\boldsymbol{M}$, with entries $\mathsf{M}_{ij} = (\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$, takes into account the fact that the spatial modes may not be orthogonal and is introduced here for generality.

If the $N$ modes span collectively an $N$-dimensional subspace, $\mathsf{M}_{ij}$ is invertible and the system (3) can be rearranged as

$$\dot{a}_i(t) = \mathsf{C}_i + \sum_{j=1}^{N} \mathsf{L}_{ij} a_j(t) + \sum_{j=1}^{N}\sum_{k=1}^{N} \mathsf{Q}_{ijk} a_j(t) a_k(t) \qquad i = 1 \ldots, N, \tag{5}$$

with

$$\mathsf{C}_i = \sum_{q=1}^{N} \mathsf{M}^{-1}_{iq} \tilde{\mathsf{C}}_q, \qquad \mathsf{L}_{ij} = \sum_{q=1}^{N} \mathsf{M}^{-1}_{iq} \tilde{\mathsf{L}}_{qj} \quad \text{and} \quad \mathsf{Q}_{ijk} = \sum_{q=1}^{N} \mathsf{M}^{-1}_{iq} \tilde{\mathsf{Q}}_{qjk}. \tag{6}$$

As observed by Rempfer and Fasel [1994a], the infinite dimensional matrix $\mathsf{M}_{ij}$ should be first inverted and then truncated to maintain a good prediction accuracy. For the cases discussed herein, we have not followed this procedure as we observed that the matrix $\mathsf{M}_{ij}$ has a strong diagonal structure. Hence, the error performed by truncating it to size $(N, N)$ and then inverting it can be reasonably assumed to be small.

Since the spatial modes satisfy automatically the boundary conditions, the expansion (2) provides a suitable foundation to examine interactions between coherent structures in complex geometries. Here, we follow established approaches [Rempfer and Fasel, 1994b]) and analyse such interactions by introducing the modal energies $e_i(t) = \frac{1}{2} a_i(t) a_i(t)$, $i = 1 \ldots, N$. The instantaneous rate of change is given by

$$\dot{e}_i(t) = \mathsf{C}_i a_i(t) + \sum_{j=1}^{N} \mathsf{L}_{ij} a_i(t) a_j(t) + \sum_{j=1}^{N}\sum_{k=1}^{N} \mathsf{Q}_{ijk} a_i(t) a_j(t) a_k(t), \qquad i = 1, \ldots, N, \tag{7}$$

obtained by multiplying (5) by $a_i(t)$. Note that, in a general case where the modes do not form an orthonormal set, the domain integral of the kinetic energy of velocity fluctuations is given by

$$E(t) = \frac{1}{2} \int_{\Omega} \mathbf{u}'(t, \mathbf{x})^2 d\Omega = \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{M}_{ij} a_i(t) a_j(t) \tag{8}$$

and not by a straightforward sum of the terms $e_i(t)$. The right hand side of equation (7) is composed of three terms describing energy transfers between the hierarchy of modes. The first two describe variations of energy due to production/dissipation arising from interactions with the mean flow and from viscous effects [Noack et al., 2011]. The third term defines variations of energy arising from inviscid nonlinear interactions between triads of modes. Following Rempfer and Fasel [1994a], these are defined in a time averaged sense by the quadratic interaction tensor $\boldsymbol{N}$ with entries

$$\mathsf{N}_{ijk} = \mathsf{Q}_{ijk} \overline{a_i a_j a_k}, \tag{9}$$

where the overbar denotes temporal averaging. The study of this term is the principal focus of the current analysis.

Spatial modes obtained from classical decompositions have generally global support over the domain (see e.g. Taira et al. [2017]). The result is that the evolution equations (5) are not strictly sparse in the sense employed by Brunton et al. [2016]. In fact, unless particular symmetries apply, the tensor $\boldsymbol{Q}$ is generally dense, *i.e.* most of its entries are different from zero and the right hand side of (5) contains all monomial terms in the modal amplitudes $a_i(t)$ up to order two. However, as anticipated in the introduction, in turbulent realisations of the Navier-Stokes equations only a subset of all triadic interactions contributes to a significant degree to the overall energy budget [Couplet et al., 2003, Rempfer and Fasel, 1994b]. The two methods developed in this work start from this fundamental observation and aim to generate a sparse Galerkin model, defined by a sparse coefficient tensor $\boldsymbol{Q}^s$ that is a good approximation of the original dynamical system in the sense that the mismatch between the transfer tensors $\boldsymbol{N}^s$ and the original $\boldsymbol{N}$ obtained from the definitions (4, 6) is as small as possible across the hierarchy of modes.

## 3.2 A-posteriori sparsification

The first sparsification technique developed during this research borrows data-driven sparse-coding methods recently proposed in the literature [Brunton et al., 2016, Loiseau and Brunton, 2018, Brunton et al., 2019] as a means to identify relevant triadic interactions in Galerkin models of turbulent flows. The approach generates sparse reduced order models by pruning weak interactions that are not relevant for the dynamics. The cornerstone of the proposed approach is $l_1$-regularised regression [Friedman et al., 2008, Tibshirani, 2013], widely used in the statistical community to extract parsimonious representation of complex datasets containing a subset of predominant features. The non-differentiable, yet convex, nature of the $l_1$ regularisation allows transforming the interaction selection problem into a convex optimisation problem that can be solved efficiently, with a unique solution. This technique is equivalent to calibrating all model coefficients by solving a regression problem [Cordier et al., 2010] augmented with a term that penalises the $l_1$ norm of the coefficients vector, to promote sparsity in the solution. The basis functions associated to the model states are never modified by the approach. Therefore, the technique is also referred to as "a-posteriori" sparsification, since it sparsifies an existing Galerkin model derived from a set of predetermined basis functions.

Sparsity-promoting regression techniques have been recently proposed by Brunton and coworkers [Brunton et al., 2016, Kaiser et al., 2018] in the SINDy framework (Sparse Identification of Nonlinear Dynamics), as a means to discover parsimonious dynamical representations of systems whose underlying (but hidden) evolution equations are sparse in the space of possible functions [Brunton et al., 2016]. However, at the time of performing our research, these methods had only been applied to relatively small Galerkin models [Loiseau and Brunton, 2018], and it was not fully understood if these can be utilised to identify and extract relevant interactions in larger models in agreement with the established picture of energy interactions in turbulent flows. In this sense, the driving factor of the present research is closer to the work of Nair and Taira [2015], Taira et al. [2016] and Nair et al. [2018]. These authors employed network-theoretic sparsification approaches [Newman, 2018] to identify key vortex-to-vortex interactions in two-dimensional homogeneous turbulence, obtaining sparse models that capture the essential physics of unsteady fluid flow with a reduced number of interactions between the same large number of states.

### 3.2.1 Description of the approach

In the first step, we assume that $N_t$ snapshots of the velocity field are available from simulation and arrange samples of the temporal coefficients $a_i(t_j)$, $i = 1, \ldots, N$ and $j = 1, \ldots, N_t$, into the data matrix $\boldsymbol{A} \in \Re^{N_t \times N}$, with entries $\mathsf{A}_{ij} = a_i(t_j)$. Similarly, we construct the modal acceleration matrix $\dot{\boldsymbol{A}} \in \Re^{N_t \times N}$, containing the time derivative of the temporal coefficients obtained by projecting the modes $\boldsymbol{\phi}_i(\mathbf{x})$ on snapshots of the Eulerian acceleration field $\partial_t \mathbf{u}(t_j, \mathbf{x})$ and correcting such projections with $\mathsf{M}$ when modes are not orthogonal (see also Rempfer and Fasel [1994b]). We then exploit the polynomial structure of the Galerkin system (5) to construct the database matrix $\boldsymbol{\Theta}(\boldsymbol{A}) \in \Re^{N_t \times q}$

$$\boldsymbol{\Theta}(\boldsymbol{A}) = \begin{pmatrix} 1 & a_1^1 & a_2^1 & \dots & a_N^1 & a_1^1 a_1^1 & \dots & a_N^1 a_N^1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & a_1^{N_t} & a_2^{N_t} & \dots & a_N^{N_t} & a_1^{N_t} a_1^{N_t} & \dots & a_N^{N_t} a_N^{N_t} \end{pmatrix}, \tag{10}$$

called nonlinear feature library in Brunton et al. [2016], where $q = (N+1) + N(N+1)/2$ is the total number of features, the sum of constant, linear and quadratic interactions. The number of quadratic coefficients is only $N(N+1)/2$ because the interaction between mode $i$ and $j$ is considered only once in (10). As discussed later on in the paper, this avoids columns of $\boldsymbol{\Theta}(\boldsymbol{A})$ becoming linearly dependent, which would in turn result in numerical stability issues in the solution regression problem (see e.g. Perret et al. [2006] and Cordier et al. [2010]).

Arranging the projection coefficients tensors $\mathsf{C}$, $\mathsf{L}$ and $\mathsf{Q}$ associated to the $i$-th mode into a coefficient vector $\boldsymbol{\beta}_i \in \Re^q$, the Galerkin system (3) can be equivalently expressed as

$$\dot{\boldsymbol{A}}_i = \boldsymbol{\Theta}(\boldsymbol{A})\boldsymbol{\beta}_i, \quad i = 1, \dots N, \tag{11}$$

where $\dot{\boldsymbol{A}}_i$ is the $i-$th column of the modal acceleration matrix. The key idea is that if some nonlinear interactions are more important than others, then the corresponding entries of the coefficient vector $\boldsymbol{\beta}_i$ can be shrunk to zero with minor effects on the predictive ability of the resulting model. The challenge is to find a systematic method to identify the dominant interactions and prune unnecessary coefficients whilst calibrating the remaining model coefficients such as to preserve the overall energy budget. Here, we adopt an established sparsity-promoting regression technique known as LASSO regression (Least Absolute Shrinkage Selection Operator, see Tibshirani [1996]). In short, it leads to a set of $N$ optimisation problems of the form

$$\min_{\boldsymbol{\beta}_i} ||\boldsymbol{\Theta}(\boldsymbol{A})\boldsymbol{\beta}_i - \dot{\boldsymbol{A}}_i||_2^2 + \gamma_i ||\boldsymbol{\beta}_i||_1, \qquad i = 1, ..., N, \tag{12}$$

one for each mode, where $\|\cdot\|_p$ denotes the $l_p$ norm of a vector. The first term in the objective function in (12) produces calibrated models that have minimum prediction error on the modal acceleration (see discussion in Cordier et al. [2010] and Couplet et al. [2005]). The second term penalises large model coefficients, regularises the regression and encourages sparsity in the solution by shrinking exactly to zero coefficients in $\boldsymbol{\beta}_i$ corresponding to columns of $\boldsymbol{\Theta}(\boldsymbol{A})$ with little dynamical influence. Ideally, to prune unnecessary coefficients, a penalisation term proportional to the cardinality of $\boldsymbol{\beta}_i$, $\mathrm{card}(\boldsymbol{\beta}_i)$, would formally be more correct [Jovanović et al., 2014]. However, the resulting optimisation problem would be computationally intractable even for Galerkin models of modest dimensions. In fact, this penalisation is usually relaxed to the computationally tractable $l_1$ term [Ramirez et al., 2013]. Regardless, the optimisation problems (12) are convex and thus have an unique solution. In addition, the approach lends naturally to parallelisation, since the optimisation problems can be solved independently for each mode. In initial stages of the research, we have found approaches based on sequential thresholded least-squares (Brunton et al. [2016], Zhang and Schaeffer [2019], Loiseau and Brunton [2018]) to be not sufficiently robust. Hence, solutions of (12) have been computed using the `sklearn` (Pedregosa and Varoquaux [2011]) library, which implements a sub-gradient descent algorithm to manage the non differentiability of the $l_1$ norm.

The weights $\gamma_i$ in equation (12) are arbitrary and can be tuned to trade prediction ability (when they are small) for sparsity (when they are large). To formalise these concepts, we introduce the global reconstruction error $\epsilon$

$$\epsilon = \sum_{i=1}^{N} \frac{||\boldsymbol{\Theta}(\boldsymbol{A})\boldsymbol{\beta}_i - \dot{\boldsymbol{A}}_i||_2^2}{||\dot{\boldsymbol{A}}_i||_2^2} \tag{13}$$

and the system density $\rho$

$$\rho = \frac{1}{Nq}\sum_{i=1}^{N}\mathrm{card}(\boldsymbol{\beta}_i). \tag{14}$$

In equation (13), the absolute reconstruction error $||\boldsymbol{\Theta}(\boldsymbol{A})\boldsymbol{\beta}_i - \dot{\boldsymbol{A}}_i||_2^2$ is normalised with the mean squared acceleration $||\dot{\boldsymbol{A}}_i||_2^2$ to balance the global reconstruction error across the hierarchy, which would be otherwise dominated by the most energetic modes. On the other hand, the density $\rho$ ranges from 0, when all interactions have been pruned, to 1, for a fully connected model. Note that for large models, the density is dominated by the quadratic tensor $\mathsf{Q}$. A one-parameter family of models can be generated by varying the regularisation weights $\gamma_i$, producing a Pareto front [Schmidt and Lipson, 2009] on the $\rho$–$\epsilon$ plane. Since only a subset of triadic interactions is relevant, the expectation is that a sweet spot appears on this curve, defining 'optimal' penalisation coefficients $\gamma_i$. It is important to observe that the penalisation coefficient $\gamma_i$ can be chosen independently for each index $i$, implying that reconstruction error and sparsity can be modulated arbitrarily across the spectrum of modes. In our analysis we consider two different modulation strategies. In strategy S1, the weight is constant for all modes, $\gamma_i = \gamma$. This strategy sparsifies more aggressively the equations of motion of low-energy modes, because the $l_1$ penalisation term has a higher importance than the $l_2$ component. In this work we observed that the lowest global reconstruction error is obtained when $\gamma_i$ is kept constant across the modes. We also introduce strategy S2, where the weight is normalised with respect to the mean squared modal acceleration as $\gamma_i = ||\dot{\boldsymbol{A}}_i||_2^2\gamma$. This is equivalent to solving problem (12) using the relative error in (13) as least-squares component of the objective function. This strategy results in a more balanced sparsification across the hierarchy of modes and avoids earlier truncation, i.e. when all coefficients of a high-index mode are set to zero. Other strategies can be, of course, devised. Here, we mention, for instance, the possibility to tune the penalisation coefficients to obtain a uniform sparsification across the spectrum or to obtain a uniform relative reconstruction error. Analysing these strategies is an interesting avenue for future work.

One potential modification of this approach is that discussed in Loiseau and Brunton [2018], namely to enforce that the nonlinear term in the sparsified Galerkin model conserves energy exactly (see e.g. Balajewicz et al. [2013] and Noack et al. [2008] for a formal definition). In practice, this can be achieved by introducing a set of constraints on the coefficients vectors $\boldsymbol{\beta}_i$. The constraints, however, couple together the regression problems of all modes, resulting in one optimisation problem of larger dimension. As we will demonstrate later in the report, the energy conservation error of models obtained from the unconstrained approach is small in relative terms. This occurs because the temporal coefficients in $\boldsymbol{A}$ are originally obtained from an energy conserving nonlinearity, and the regression "discovers" this property from data. Hence, throughout the research we always solved problems (12) independently, without additional constraints.

## 3.3 A-priori sparsification

The major downside of the "a-posteriori" approach discussed in the previous section is that a rigorous connection between the modal structures and the projection model is necessarily lost, since the procedure involves calibrating model coefficients without modifying the basis functions. Hence, the analysis of energy paths in the model is not necessarily consistent with the spatial structure of the basis functions. Therefore, we attempted to bridge this gap by proposing a second, novel sparsification method in which the goal is to seek directly modal structures that capture energy transfer mechanisms efficiently. More precisely, the method consists in seeking a new basis set that produces a sparse quadratic coefficient tensor regulating triadic interactions, without the need for an *a posteriori* tuning based on $l_1$-based regression. We refer to this method as *a-priori* sparsification, since sparse characteristics are obtained before the projection step, not after. The computational approach is inspired by the subspace rotation technique of Balajewicz et al. [2013, 2016], where a small rotation of the POD subspace was sought to absorb the unresolved dissipative scales into the basis functions and stabilise the long-term behaviour without the need for empirical eddy-viscosity terms. Here, the key idea is to seek a small rotation of the original POD subspace to alter and sparsify energy transfer paths. In practice, the rotation is found by solving a constrained optimisation problem, minimising the loss of energy optimality subject to a constraint on the sparsity of the quadratic coefficient tensor.

### 3.3.1 Description of the approach

To identify the new set of modal structures, we utilise a subspace rotation technique introduced in the context of stabilisation of Galerkin-based reduced order models by Balajewicz et al. [2013]. Geometrically, this technique consists in seeking a rotation of a $N$-dimensional POD subspace within a larger POD subspace of dimension $M$. The rotation is defined by a transformation matrix $\boldsymbol{X} \in \Re^{M \times N}$, satisfying $\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{I}$ to ensure that the rotated spatial basis functions remain an orthonormal set. The rotated basis functions and the associated temporal coefficients, denoted in what follows with a tilde, are expressed as a linear combination of the original POD spatial and temporal modes as

$$\tilde{\boldsymbol{\phi}}_i(\mathbf{x}) = \sum_{j=1}^{M} \mathsf{X}_{ji} \boldsymbol{\phi}_j(\mathbf{x}), \tag{15a}$$

$$\tilde{a}_i(t) = \sum_{j=1}^{M} \mathsf{X}_{ji} a_j(t). \tag{15b}$$

It is worth pointing out that finding a new set of modal structures directly would be a much higher dimensional problem to tackle. The number of unknowns would be proportional to the numbers of modes sought for multiplied by the number of degrees of freedom of the problem at hand. Seeking new modal structures as a linear combination of POD modes represents a significant reduction in complexity, controllable by varying the dimension $M$. Using POD modes as building blocks also has the advantage of producing a basis with good energy reconstruction properties.

The linear and quadratic coefficients of the Galerkin system (3) obtained by projection on the rotated subspace are then given by the following matrix expressions

$$\tilde{\boldsymbol{C}} = \boldsymbol{X}^\top \boldsymbol{C} \quad \text{and} \quad \tilde{\boldsymbol{L}} = \boldsymbol{X}^\top \boldsymbol{L} \boldsymbol{X}, \tag{16}$$

while the quadratic coefficients are cubic polynomial functions of the entries of the rotation matrix

$$\tilde{\mathsf{Q}}_{ijk} = \sum_{p,q,r=1}^{M} \mathsf{Q}_{pqr} \mathsf{X}_{pi} \mathsf{X}_{qj} \mathsf{X}_{rk}, \quad i,j,k = 1,\dots,N. \tag{17}$$

In these expressions, the tensors $\boldsymbol{C} \in \Re^M$, $\boldsymbol{L} \in \Re^{M \times M}$ and $\boldsymbol{Q} \in \Re^{M \times M \times M}$ are the Galerkin coefficient tensors obtained from the $M$-dimensional set of original POD modes. Our goal is to seek a rotation matrix for which the rotated quadratic interaction coefficient tensor $\tilde{\boldsymbol{Q}}$ has a sparse structure, i.e. where as many as possible of the quadratic interaction coefficients are identically zero.

At this stage, it is worth noting that any rotation is necessarily accompanied by a loss of average fluctuation kinetic energy reconstructed by the new basis. The energy reconstructed by a set of $P$ POD modes can be quantified by utilising the average modal energies $\lambda_i = \overline{a_i a_i}$, $i = 1, \dots, P$.

Arranging them into a diagonal matrix $\boldsymbol{\Lambda}_P \in \Re^{P\times P}$, the trace $Tr(\boldsymbol{\Lambda}_P)$ defines an upper bound for the reconstructed energy for any $P$-dimensional set of basis functions, due to well-known optimality properties of POD. Similarly, the energy reconstructed by the rotated basis can be expressed with the average modal energies $\tilde{\lambda}_i = \overline{\tilde{a}_i\tilde{a}_i},\ i = 1, \dots N$, of the rotated temporal coefficients (17) and arranging them into the diagonal matrix $\tilde{\boldsymbol{\Lambda}}_N = \boldsymbol{X}^\top\boldsymbol{\Lambda}_M\boldsymbol{X} \in \Re^{N\times N}$. The loss of reconstructed average fluctuation kinetic energy with respect to an $N$-dimensional POD subspace is then quantified as

$$\mathcal{J}(\boldsymbol{X}) = Tr(\boldsymbol{\Lambda}_N - \boldsymbol{X}^\top\boldsymbol{\Lambda}_M\boldsymbol{X}). \tag{18}$$

As observed in Balajewicz et al. [2013], this quantity is necessarily non-negative due to the optimality of the original POD basis, i.e.

$$\overline{\int_\Omega \|\mathbf{u}'(t,\mathbf{x})\|^2 \mathrm{d}\Omega} = Tr(\boldsymbol{\Lambda}_\infty) > Tr(\boldsymbol{\Lambda}_N) \geq Tr(\tilde{\boldsymbol{\Lambda}}_N), \tag{19}$$

where the last equality holds for $M > N$ in trivial cases only. In addition, the quantity (18) is also always identically zero when $M = N$, for any $\boldsymbol{X}$, since any linear combination of $N$ POD modes necessarily spans the same original $N$-dimensional subspace.

To measure sparsity of the quadratic interaction coefficient tensor, we use the $l_1$ norm operator, denoted as $\|\cdot\|_1$, in light of the practical intractability of the zero norm in optimisation [Jovanović et al., 2014]. Then, the trade-off between energy optimality and sparsity is expressed by formulating the following constrained optimisation problem

$$\begin{aligned} \min_{\boldsymbol{X}} \quad & Tr(\boldsymbol{\Lambda}_N - \boldsymbol{X}^\top\boldsymbol{\Lambda}_M\boldsymbol{X}) && \text{(20a)}\\ \text{subject to} \quad & ||\tilde{\boldsymbol{Q}}||_1 \leq ||\boldsymbol{Q}||_1/\xi, && \text{(20b)}\\ & \boldsymbol{X}^\top\boldsymbol{X} = \boldsymbol{I}_{N\times N}, && \text{(20c)} \end{aligned}$$

referred to as problem P1 in what follows. The role of the objective function (20a) is to favour transformation matrices that minimise the loss of energy optimality, producing a set of basis function with good energy representation ability. On the other hand, the constraint (20b) encourages sparse solutions, because some entries of $\tilde{\boldsymbol{Q}}$ are shrunk to zero during the solution of (20) by the non-differentiability of the $l_1$ norm [Tibshirani, 2013, Friedman et al., 2008]. Sparsification of the linear coefficients tensor $\tilde{\boldsymbol{L}}$ should not be expected. Then, the weight $\xi$ is an arbitrary penalisation parameter that controls the relative $l_1$ norm of the rotated quadratic coefficient tensor and the sparsity of the resulting Galerkin model. This parameter ranges from 1 to arbitrarily large values. When it is equal to 1, the $l_1$ norm of the rotated tensor is not affected and no sparsification is obtained. Conversely, using larger $\xi$ constrains the norm of the rotated tensor to decrease. This promotes sparsity due to the non-differentiability of the $l_1$ norm, with a similar mechanism to well-known LASSO methods [Tibshirani, 1996]. In addition, different definitions of the sparsification constraint (20b) can be used to highlight different physical aspects of the resulting bases.

It is worth noting that using the *a-posteriori* LASSO-based sparsification method discussed in the previous section, all quadratic coefficients could, in principle, be set to zero by using a large regularisation weight in the LASSO optimisation problem. This is because the constant, linear and quadratic coefficients of the Galerkin model are directly the optimisation variables. In the present case, the model coefficient cannot be modified directly, but the model tuning is performed indirectly through the rotation matrix $\boldsymbol{X}$, which is the actual optimisation variable of the problem. The important consequence is that it might not always be possible to find a rotation that sets an arbitrarily large number of model coefficients to zero. This indicates that problem (20) might not have a feasible solution if the penalisation weight is too large.

An important characteristic of optimisation problem (20) is that while the objective (20a) is convex, the sparsity-promoting constraint (20b) is not, as it involves cubic polynomials in the optimisation variables, the entries of the transformation matrix $\boldsymbol{X}$. Consequently, the solution might not be unique and several local minima, corresponding to different sets of basis functions, may be obtained by starting the optimisation from different initial guesses. However, starting the optimisation from small random perturbations of the original POD basis (justified by the need to retain good energy reconstruction properties) consistently produced the same optimal solutions, which will be presented in the following sections.

Assuming a feasible solution of problem (20) can be found, the Galerkin model constructed from projection onto the optimal rotated basis does not necessarily possess better long-term temporal stability characteristics than the original POD model. In fact, it is well-known that POD-Galerkin models exhibit long-term instability because of the deficit of energy dissipation attributed to the truncation of small dissipative scales (Noack et al. [2008], Schlegel and Noack [2015]; see also the recent work of Grimberg et al. [2020] for a point against this argument). In the present case, the transformation $\boldsymbol{X}$ obtained from solution of (20) does not necessarily result in an improved description of dissipative processes. Classically, this issue is cured by introducing, *a posteriori*, an eddy-viscosity-type term in the Galerkin model [Galletti et al., 2004, Noack et al., 2005, Östh et al., 2014]. However, an *a posteriori* correction would not remain in the spirit of the present work. We thus favour the subspace-rotation-based stabilisation approach proposed by Balajewicz et al. [2013, 2016], which can be introduced naturally in the present formulation. In practice, we augment problem (20) with the additional implicit constraint

$$Tr(\tilde{\boldsymbol{L}}) = Tr(\boldsymbol{X}^\top \boldsymbol{L} \boldsymbol{X}) = -\eta \tag{21}$$

where the auxiliary variable $\eta \in \Re^+$ is chosen such that

$$\chi(\eta) = \frac{\overline{E(t)} - \overline{E_{\text{DNS}}(t)}}{\overline{E_{\text{DNS}}(t)}} = 0, \tag{22}$$

i.e. that the relative difference of the average fluctuation kinetic energies from direct numerical simulation (DNS) and from numerical simulation of the new model vanishes. The variable $\eta$ controls dissipation mechanisms in the Galerkin model by altering the spectrum of $\tilde{\boldsymbol{L}}$ and ensures long-term stability. As observed by Balajewicz et al. [2013], $\eta$ is not known a priori, but can be found in an inner optimisation loop to ensure that the excess average fluctuation kinetic energy defined by $\chi(\eta)$ is zero. With this additional constraint, problem (20) becomes

$$\begin{aligned}
&\min_{\boldsymbol{X}} && Tr(\boldsymbol{\Lambda}_N - \boldsymbol{X}^\top \boldsymbol{\Lambda}_M \boldsymbol{X}) && \text{(23a)}\\
&\text{subject to} && ||\tilde{\boldsymbol{Q}}||_1 \leq ||\boldsymbol{Q}||_1/\xi, && \text{(23b)}\\
& && Tr(\boldsymbol{X}^\top \boldsymbol{L} \boldsymbol{X}) = -\eta \quad \text{with} \quad \chi(\eta) = 0, && \text{(23c)}\\
& && \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{I}_{N\times N}. && \text{(23d)}
\end{aligned}$$

In this formulation, denoted as P2 henceforth, there is still only one free parameter, $\xi$. The additional constraint guarantees long-term stability but it can only be satisfied (i.e. the problem is feasible) if $M > N$. In fact, any rotation $\boldsymbol{X} \in \Re^{N\times N}$ cannot alter the subspace spanned by the original $N$ POD modes and the associated description of dissipation mechanisms captured by the model. In what follows, we consider models with ratio $M/N = 2$ and 3.

In practice, a small random perturbation of the original $N$-dimensional POD basis was used as initial guess for the optimisation and successive optimisation problems for different penalisations $\xi$ were started from the solution of the previous problem. Problem (23) was solved with the open-source package for nonlinear and non-convex optimisation `NLopt` [Johnson, 2014]. We utilised a solver implementing the Method of Moving Asymptotes (MMA) algorithm [Svanberg, 2014] which requires the gradient of the objective function and of the constraints. The key element to make the procedure viable is to evaluate the sparsity-promoting constraint and its gradient as efficiently as possible. A naive implementation requires $\mathcal{O}(M^3N^3)$ operations for the evaluation of the sparsity promoting constraint and $\mathcal{O}(M^4N^4)$ for the evaluation of its gradient with respect to the rotation $\boldsymbol{X}$, and costs become quickly intractable. A significantly more efficient algorithm to compute these two quantities with costs scaling as $\mathcal{O}(MN^3 + M^2N^2 + M^3N)$ for the evaluation of the sparsity-promoting constraint and its gradient can be derived and is key to make the procedure viable (see discussion of the method in the appendix of [Rubini et al., 2021]).

One further technical remark is that the sparsity-promoting constraint is a non-smooth function of the transformation matrix $\boldsymbol{X}$, posing difficulties for the utilisation of gradient-based optimisation algorithms. One approach is to implement a subgradient descent method, often used for the solution of the LASSO [Friedman et al., 2008]. However, this algorithm can only be implemented if an analytical solution of the optimisation problem is known, which is not the case here. In this work, we used a manual soft-thresholding approach where entries of the rotated tensor $\tilde{\boldsymbol{Q}}$ smaller that the numerical tolerance specified to the gradient-based optimiser (typically $tol = 10^{-5}$) are set to zero at the end of the optimisation.

The Galerkin models obtained from solution of (23) are then characterised by examining the density of the rotated triadic interaction tensor, defined as

$$\rho = \frac{||\tilde{\boldsymbol{Q}}||_0}{||\boldsymbol{Q}||_0}, \tag{24}$$

where the $l_0$ norm $||\cdot||_0$ counts the nonzero elements of a tensor. The density can also be expressed as the average

$$\rho = \frac{1}{N}\sum_{i=1}^{N} \rho_i \tag{25}$$

with the modal densities $\rho_i = ||\tilde{\boldsymbol{Q}}_i||_0/||\boldsymbol{Q}_i||_0$, $i = 1, \ldots, N$ being the relative number of non-zero coefficients in the slices $\boldsymbol{Q}_i$ of the quadratic coefficient tensor associated to each modal index. To express the energy captured by the $N$-dimensional rotated basis, we also introduce the global energy reconstruction factor

$$e_N = Tr(\tilde{\boldsymbol{\Lambda}}_N)/Tr(\boldsymbol{\Lambda}_\infty), \tag{26}$$

a quantity always strictly lower than one. The density and the reconstruction factor of the rotated Galerkin models depend on the dimensions $M$ and $N$ and on the penalisation weight $\xi$. To characterise the effects of these parameters, we visualise the rotated systems on the $\rho - e_N$ plane, similar to the approach used in Rubini et al. [2020].

# 4 Results and Discussion

## 4.1 Benchmark problem and modal decomposition

The large majority of the research performed during the project has focused on two-dimensional unsteady flow in a lid-driven square cavity. This is an established test case for the development and validation of model order reduction techniques [Cazemier et al., 1998, Terragni et al., 2011, Balajewicz et al., 2013, Arbabi and Mezić, 2017, Fick et al., 2018], and we have thus considered it as a suitable benchmark problem to demonstrate and analyse the proposed sparsification techniques.

The Reynolds number is defined as $Re = LU/\nu$ where $L$, $U$ are the cavity dimension and the (uniform) lid velocity, respectively, while $\nu$ is the kinematic viscosity. We investigated the flow regime establishing at $Re = 2 \times 10^4$, where the flow evolves in a chaotic manner (Auteri et al. [2002], Peng et al. [2003]). The chaotic nature of the problem ensures that the frequency spectrum of velocity fluctuations is continuous and energy transfers are scattered in modal space. This features is ideal for the application of sparsification methods. By contrast, in periodic or quasi-periodic regimes developing after the first few bifurcations from the steady laminar flow, the frequency spectrum and energy interaction between modes are highly organised and already show a large degree of sparsity. The domain is defined by the nondimensional Cartesian coordinates $\mathbf{x} = (x, y)$ and the velocity vector $\mathbf{u}(t, \mathbf{x})$ is defined by the components $u(t, \mathbf{x})$ and $v(t, \mathbf{x})$. For visualisation purposes, we introduce the out-of-plane vorticity $\omega = \partial v/\partial x - \partial u/\partial y$. Numerical simulations were performed in `OpenFOAM` with the incompressible flow solver `icofoam`. Three snapshots of the vorticity field obtained from these simulations are shown in figure 1. Most of the dynamically interesting features in this regime originate at the bottom-right corner of the cavity. Specifically, the secondary vortex in the recirculation zone is shed erratically, producing wave-like disturbances advected along the shear layer bounding the primary vortex.

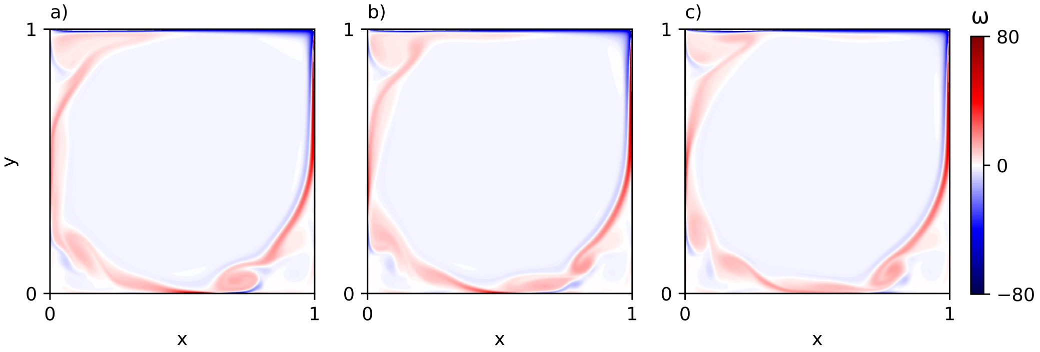

Figure 1: Vorticity field of three different snapshots separated by one non-dimensional time unit, increasing from left to right.

As a starting point, we considered Galerkin models generated using Proper Orthogonal Decomposition (POD) modes. POD produces economic models, but has the well-known shortcoming of mixing together fluid motions at different temporal/spatial scales [Mendez et al., 2019]. Hence, we also considered models generated from modes oscillating at a single frequency obtained from a procedure that is equivalent to a Discrete Fourier Transform (DFT) of the velocity snapshots. In this report, only results obtained on POD-Galerking models will be presented. Additional detail can be found in Rubini et al. [2020]. We denote the normalised cumulative sum of the eigenvalues $\lambda_i$ of the snapshots temporal correlation matrix as

$$e(n) = \sum_{i=1}^{n} \lambda_i / \sum_{i=1}^{N_t} \lambda_i, \tag{27}$$

| $n$ | 1 | 5 | 10 | 15 | 20 | 26 | 50 | 75 | 80 | 95 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| POD | 0.26 | 0.74 | 0.85 | 0.89 | 0.9 | 0.95 | 0.98 | 0.99 | 0.995 | 0.998 | 0.999 |

Table 1: Normalised cumulative energy distribution $e(n)$ for POD modes.

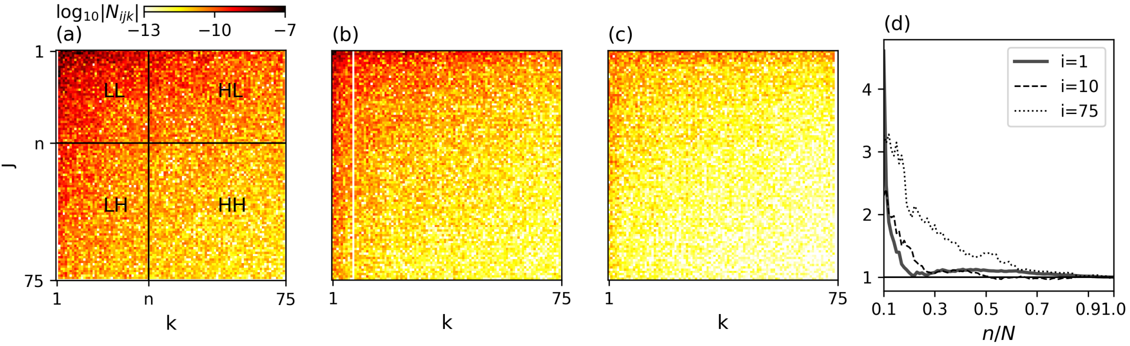

Figure 2: Magnitude of the average interaction tensor coefficients $\mathsf{N}_{ijk}$ for three POD modes across the spectrum, $i = 1, 10$ and 75 in panel (a), (b) and (c) respectively, for a model resolving 99% of the fluctuation kinetic energy. Panel (d) shows the coefficient $\chi_i(n)$ as a function of the normalised cutoff $n$ for the same three modes.

describing the fraction of the fluctuation kinetic energy captured by the first $n$ elements of the expansion (2). This fraction is reported in table 1 for reference.

## 4.2 Energy Analysis

To provide a more robust foundation to understand the sparsification results reported in later sections, we first focus on the analysis of the average energy interactions. The structure of the interaction tensor $\boldsymbol{N}$ for a large POD-based model with $N = 75$, reconstructing more than 99% of the fluctuation kinetic energy, is reported in figure 2, showing the magnitude of the interactions for three slices for $i = 1, 10$ and 75, in panels (a), (b) and (c), respectively. All entries of the tensor $\boldsymbol{N}$ are generally nonzero, although the strength of the interactions varies across several orders of magnitude. This is a combined result of the projection coefficients tensor $\boldsymbol{Q}$, whose entries are typically non zero, and of the complex spectral structure of the temporal coefficients $a_i(t)$. The most important feature of figure 2 is that interactions are highly organised and there exists a subset of interactions that are more active. Specifically, for any mode $i$, triadic interactions can be classified as illustrated in panel (a) in four different categories by introducing a cutoff modal index $n$. The subset of interactions denoted as LL corresponds to nonlinear energy transfer involving pairs of low index modes, HL and LH denote interactions involving high-low/low-high index modes, while HH denotes the subset of interactions involving pairs of high index modes. We observe that the areas corresponding to LL and HL/LH are the most active. If we map low/high modal indices to large/small scales, this result is in agreement with the picture of energy transfer between scales in homogeneous isotropic two-dimensional turbulence [Ohkitani, 1990, Laval et al., 1999], where the large scales interact with the small ones in a non-local fashion. In addition, interactions are not symmetric with respect to a swap of indices $j, k$. This can be quantified

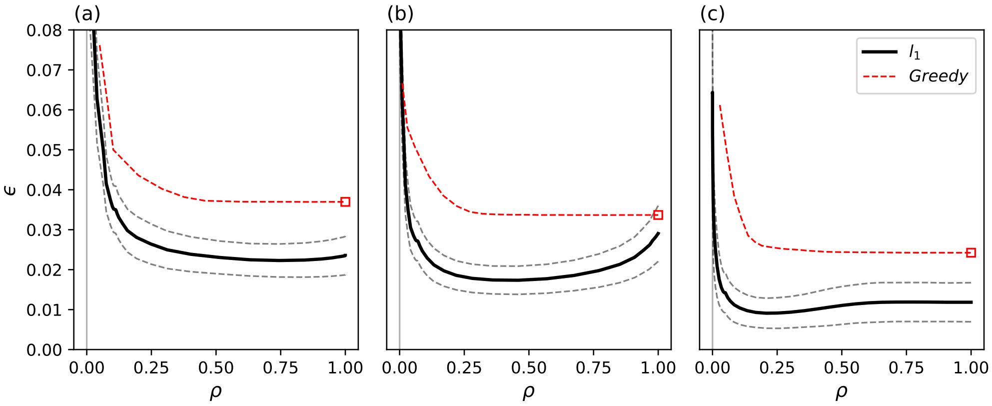

Figure 3: $\rho - \epsilon$ curves for three POD models resolving 90%, 95% and 99% of the kinetic energy in panels (a), (b) and (c), respectively. The black line represents the cross-validated error averaged over K = 10 folds. The dashed grey lines represent plus/minus one standard deviation of the cross-validated error calculated over the folds. The red dashed line shows the reconstruction error obtained with the greedy approach. The squares indicate the global reconstruction error of the Galerkin model obtained directly from projection.

by computing the coefficient

$$\chi_i(n) = \sum_{j=1}^{n}\sum_{k=1}^{N} \mathsf{N}_{ijk} / \sum_{j=1}^{N}\sum_{k=1}^{n} \mathsf{N}_{ijk}, \tag{28}$$

representing the relative dynamical importance of the subset of interactions LL + HL and LL + LH. Panel (d) of figure 2 shows $\chi_i$ for $i = 1, 10$ and 75 as a function of the normalised cutoff $n$. The interaction subset HL is up to four times more important than the subset LH. This is a consequence of the asymmetry of the projection coefficients $\mathsf{Q}_{ijk}$, which arise from the fact that the convective transport of structure $\boldsymbol{\phi}_k(\mathbf{x})$ operated by the structure $\boldsymbol{\phi}_j(\mathbf{x})$ is more intense when the modal structure $\boldsymbol{\phi}_j(\mathbf{x})$ describes large-scale flow features.

## 4.3 A-posteriori sparsification

We now apply the “a posteriori” methodology presented in section 3.2 to three POD-based models resolving 90%, 95% and 99% of the kinetic energy, respectively (see Table 1 for details). Because the size of the database matrix $\boldsymbol{\Theta}(\boldsymbol{A})$ grows quadratically with the number of modes, the number of possible interactions $q$ can easily become larger than the number of available snapshots $N_t$, resulting in an underdetermined regression problem and overfitting. This is a well understood issue in data analysis and requires cross validation techniques to ensure the statistical reliability of the result [Friedman et al., 2008]. In this work, we employed K−fold cross validation, using typically K = 10. Briefly, the database is first divided into K folds. The model is trained using K − 1 blocks and the reconstruction error $\epsilon$ of equation (13) is obtained from the fold that was left out. This procedure is iterated over all folds, obtaining the mean and the standard deviation of the reconstruction error.

Figures 3-(a,b,c) show the sparsification curves on the $\rho - \epsilon$ plane for the three POD models considered. The mean of $\epsilon$ across the folds is displayed as a thick black line, while the grey dashed line indicates plus or minus one standard deviation. These curves have been obtained by solving problem (12) using strategy S1 and progressively increasing the regularisation weight. When low weights are used, dense systems with good prediction accuracy are obtained. The opposite is true for large weights, identifying points in the left part of the graph characterised by low density and poor prediction accuracy. The curves show a sweet spot at around $\rho \approx 0.2$, displaying a plateau for $\rho \gtrsim 0.2$, while the error $\epsilon$ grows quickly when $\rho \lesssim 0.2$. These results indicate that it is possible to prune about 80% of the quadratic interactions in model (3) without influencing the average prediction accuracy. The red dashed line represents the reconstruction error obtained with a naive sparsification approach. The approach consists in pruning coefficients of $\mathsf{Q}_{ijk}$ in the area denoted as HH in figure 2-(a). The approach exploits the structure of $\mathsf{N}_{ijk}$ and is therefore referred to as "greedy". By varying $n \in (1, N)$, models with different sparsity are obtained, with $n = N$ corresponding to the original projection model projection (indicated as a red square in figure 3). The shape of the sparsification curves for the greedy approach are similar to those obtained with the $l_1$ regression. This is a direct consequence of the existence of a subset of most relevant energy interactions. However, the reconstruction error obtained from the greedy method is generally higher than that obtained by the $l_1$ regression, since the optimisation procedure involved in the $l_1$ approach modulates the strength of the remaining interactions by tuning the active quadratic coefficients, minimising the prediction error. As we show later in section 4.3.2 dedicated to analysing the model performance in time integration, this difference will have a marked effect on the long-term temporal stability of the models.

Regardless of the approach, the mean reconstruction error decreases as the resolved energy increases, moving from panel (a) to panel (c), as more modes participate in capturing the dynamics of velocity fluctuations. In addition, larger models can be more effectively sparsified, as the sparsification curve drops more rapidly. This results from the non-local structure of energy interactions shown in figure 2. When one additional low-energy mode is included, the number of relevant interactions to be retained in the model only grows as $\mathcal{O}(N)$ and not as $\mathcal{O}(N^2)$, i.e. all non-local interactions with the rest of the hierarchy denoted as $\mathrm{LL}, \mathrm{LH}$ and HL in figure 2-(a). Since the total number of possible interactions grows as $\mathcal{O}(N^3)$, larger models can be more effectively sparsified. This is conceptually in agreement with the observations of Taira et al. [2016] on the sparsification properties of discrete vortex models. We also observe that the mean prediction error does not necessarily decrease monotonically when the density increases. This phenomenon is particularly visible for the model in panel (b) but all models reproduce the same behaviour. This is a symptom that the number of available snapshots (1500) is potentially not large enough for the number of coefficients ($q = N \times (N+1)/2 + N + 1 = 2926$ for the model in panel (c)) and overfitting would have occurred if no cross-validation had been performed.

#### 4.3.1 Energy interactions identified by the regression and conservation properties

To visualise the sparsity pattern identified by the regression as the regularisation weight in equation (12) is increased (constant for all modes in strategy S1), we introduce the tensor $\boldsymbol{\gamma}$ with entries $\gamma_{ijk}$ defined as the value of the regularisation weight at which the corresponding coefficient $\mathsf{Q}_{ijk}$ is shrunk to zero by the LASSO. Figures 4-(a,b,c) show three slices of $\boldsymbol{\gamma}$ for modes $i = 1, 10$ and 75, respectively, for the largest POD model considered, capturing 99% of the total fluctuation kinetic energy. The first interactions to disappear are the small-scale/small-scale interactions. Increasing the penalisation, interactions that are local in modal space are progressively pruned, leaving only non-local interactions involving triadic exchanges with the low-index modes for large penalisations. Interestingly, this pattern does not change qualitatively nor quantitatively as the modal index $i$ increases. In fact, a comparable

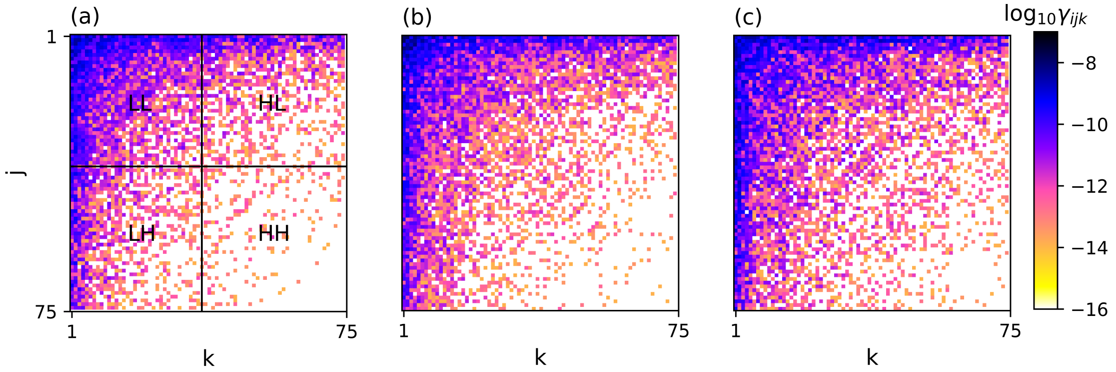

Figure 4: Distribution of the base ten logarithm of $\gamma_{ijk}$ for $i = 1, 10$ and $75$, in panels (a), (b) and (c), respectively, for the POD model resolving 99% of the fluctuation kinetic energy.

number of interactions is retained across the hierarchy and the governing equations of all modes are sparsified by an equal amount. Hence, sparsification has not produced mode truncation, which would have occurred if all coefficients of some low-energy modes had been shrunk to zero by the LASSO. This behaviour can be justified by noting that the mean square acceleration $||\dot{\boldsymbol{A}}_i||_2^2$ of the POD amplitude coefficients varies weakly with $i$. In fact, the sparsification pattern does not change significantly when strategy S2 is used. In figures 5-(a,b,c) the base ten logarithm of the mean energy interaction tensor $\mathsf{N}^s_{ijk}$ computed as in (9) with the sparse coefficient tensor $\boldsymbol{Q}^s$ is shown for the same three modal indices as in figure 4. The data refer to sparse models with $\rho = 0.3$, located nearby the sweet spot of the curves in figure 3-(c). It can be observed that the sparsified model has a pattern of interactions resembling that of the dense model in figure 2. However, weak interactions and the associated flow physics have been pruned. It is also clear that the asymmetry of the interaction pattern observed in figure 2 and the physical mechanism that originates it are invisible to the regression and the interaction pattern in figure 5 is now symmetric with respect to a swap of the indices $j, k$.

Despite the aggressive pruning, the sparse models reproduce fairly accurately the overall structure of the intermodal energy budgets. In the present flow configuration, the convective nonlinearity is energy conserving and Galerkin models should obey the relation $\sum_{i=1}^{N} T_i = 0$ as $N \to \infty$, with $T_i = \sum_{j=1}^{N}\sum_{k=1}^{N} \mathsf{N}_{ijk}$ the time averaged energy transfer rate to/from mode $i$. For finite-dimensional approximations, this property is not satisfied exactly and the residual of the summation can be taken as a measure of the overall energy balance. Figure 6-(a) shows such residual for the $l_1$ sparsified models (empty circles) and for the models obtained with the greedy approach (empty squares), as a function of the density. The residual is normalised by the root mean square value of the rate of change of the integral fluctuation energy. Note that the greedy model at $\rho = 1$ is the model obtained directly from projection. It can be observed that the energy conservation error is relatively small, in the order of $10^{-3} \div 10^{-4}$. For large densities, it is larger than that of the projection model, because the regression tunes model coefficients to minimise the mean square error on the modal accelerations and does not enforce this physical constraint directly. The energy conservation residual decreases for

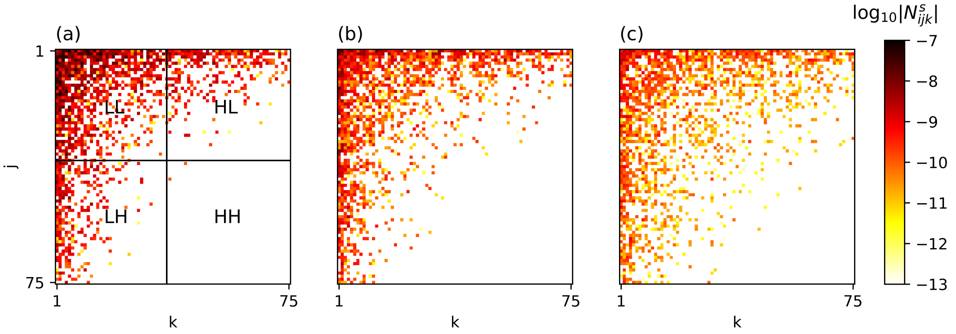

Figure 5: Base ten logarithm of the sparsified interaction tensor $\mathsf{N}^s_{ijk}$ for $i = 1, 10$ and 75 in panels (a), (b) and (c), respectively, for a POD model model resolving 99% of the fluctuation kinetic energy and $\rho = 0.3$.

sparser models and is ten times smaller than the projection model, owing to the lower number of active coefficients that participate in the regression. Figure 6-(b) shows the distribution of the time averaged energy transfer rate associated to mode $i$ for the $l_1$ sparsified model at $\rho = 0.3$ (red crosses) and the dense model obtained from projection (empty circles). Data is reported every second mode. For the projection model, the net energy transfer is negative for the first few modes and changes sign at $i \sim 10$. Physically, this trend suggests that the first few modes extract energy from the mean flow and feed the dissipative high-index modes via triadic interactions. The $l_1$ sparsified model correctly reproduces this global trend, even though no constraints have been introduced [Loiseau and Brunton, 2018]. It is argued that this is a general properties of data-driven techniques relying on optimisation ideas, such as the LASSO, which naturally reproduce invariants and conservation properties embedded in the data to a level defined by noise levels. For instance, Taira et al. [2016] used network-theoretic ideas to sparsify connections in a discrete vortex model and observed that sparsification conserves the invariants of discrete vortex dynamics.

#### 4.3.2 Long-term temporal behaviour of the sparse systems

We now turn our attention to the long-time behaviour of the sparsified models under temporal integration. We consider results for models resolving 95% of the turbulent kinetic energy as an illustrative example, and use the projections of the POD modes onto one of the DNS snapshots to obtain initial conditions. Results are shown in figure 7. Panel-(a) shows the temporal evolution of the integral fluctuation kinetic energy $E(t)$, as defined in (7), for the dense projection model and the sparse models obtained with the $l_1$ and greedy approach, with same density $\rho = 0.3$. The temporal evolution obtained with these models is compared against the fluctuation kinetic energy from DNS. For the projection model, the integral kinetic energy grows substantially in the first 40 time units and then saturates on a value that is about two orders of magnitude larger than what observed in DNS. The over-prediction occurs because the truncation of the small scales in the ansatz (2) leads to a significant imbalance of the production–dissipation budget within the model [Balajewicz et al., 2013, Noack et al., 2005, 2008, 2011]. A qualitatively similar behaviour, if not worse, is then necessarily observed for the model

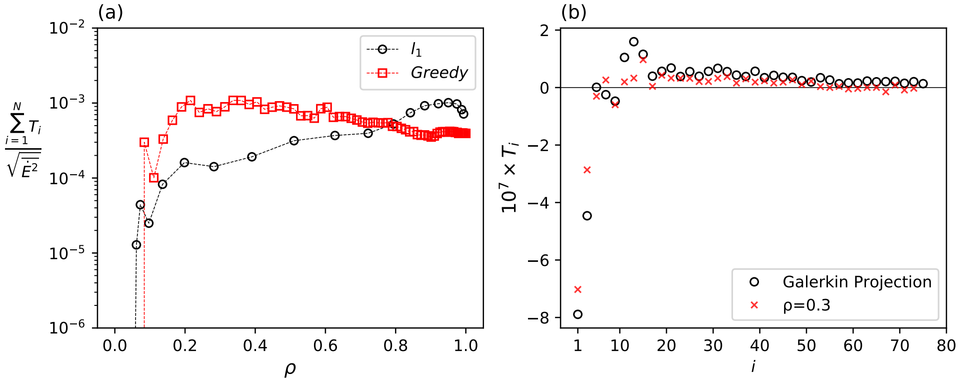

Figure 6: Panel (a): the normalised energy conservation error as a function of the density $\rho$, for the greedy sparsification approach (empty squares) and the $l_1$ based approach (empty circles). Panel (b): the total net energy transfer rate $T_i$ as a function of the modal index $i$ for two POD models resolving 99% of the kinetic energy, with coefficients identified from projection and for a $l_1$ sparse model. One every two data points is reported.

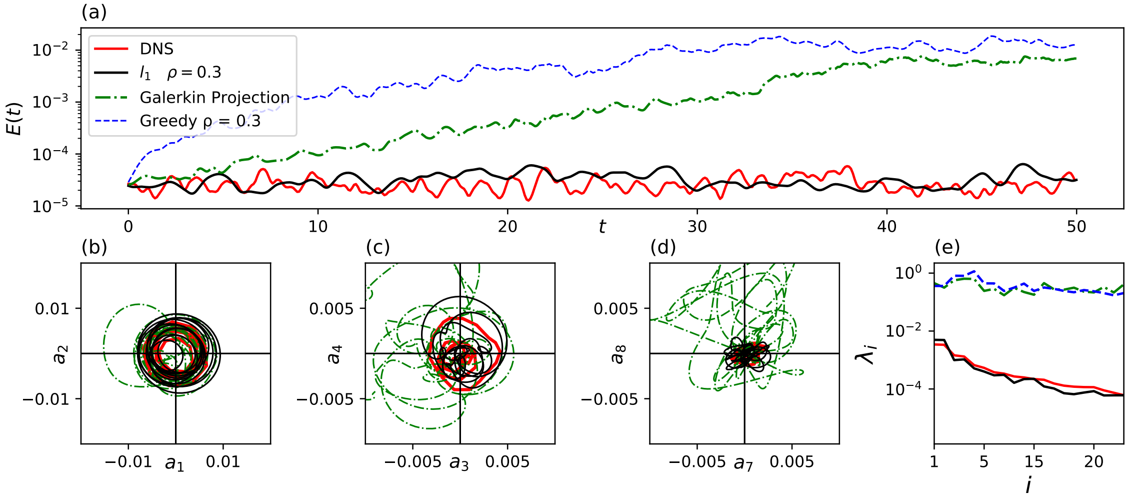

Figure 7: Panel (a): temporal evolution of the turbulent kinetic energy $E(t)$ from DNS compared to that obtained from temporal integration of the $l_1$, greedy and Galerkin projection models. Sparse models have $\rho = 0.3$. Panels (b) to (d): state space projections onto three different mode pairs. Data for the greedy model is omitted as the trajectory quickly leaves the visible range. Panel (e): average modal energy ($\lambda_i = \overline{a_i a_i}$) from DNS and long-time integration of the models considered in panel (a).

sparsified with the greedy approach, since neglecting weak interactions alone does not cure the original dissipation problems. Conversely, the $l_1$ model is able to predict the correct average fluctuation kinetic energy and has excellent long-term stability properties, despite this being not enforced in the regression procedure (see Fick et al. [2018]). We argue that this is due to the fact that the $l_1$ procedure performs a "prune-then-calibrate" approach, where weak interactions are first pruned and the remaining active coefficients are then tuned in the optimisation involved in (12) to match the reference dynamics. It is evident from these results that this second step is key to obtain accurate long-term behaviour. Panels (b) to (d) of figure 7 show a shorter segment of state space trajectory from these models projected onto three different pairs of modes. We omit the data for greedy approach since orbits quickly drift out of the visible range and is thus qualitatively similar to that from the dense model. It is clear that the trajectory of the $l_1$ model remains in the same volume of state space occupied by the DNS projections, while the projection model drifts away to a different region of state-space, over-predicting the integral kinetic energy. This is more effectively visualised in panel (e), displaying the average modal energy $\lambda_i = \overline{a_i a_i}$ as a function of the modal index. Data is obtained from averaging long trajectories after the initial transient in panel (a) has completed. The projection and greedy models predict much larger energies across the entire spectrum, while the $l_1$ correctly predicts the correct decay of modal energies. It is, of course, not possible to guarantee that $l_1$ sparsified models of generic turbulent flows will have good long-term stability [Schlegel and Noack, 2015], but the present results constitute evidence that this is realistically possible on a non-trivial problem. Finally, we have observed in animations of the reconstructed flow fields using the spatial modes and the temporal coefficients from numerical integration that characteristic flow features, such as the erratic burst of corner vortices and the evolution of coherent structures in the shear layer bounding the main vortex, are also well reproduced by the $l_1$ model.

### 4.4 A-priori sparsification

We now report key results obtained by applying the "a-priori" sparsification method to POD-based models.

We consider models with dimension $N = 30$, for ratios $M/N = 2$ and 3 and examine in more detail the effect of the sparsity-promoting constraint (23b). For each ratio, a family of models with different density and reconstructed average kinetic energy is generated by increasing the penalisation parameter $\xi$. Optimal solutions are displayed on the plane $1/\xi$ – $||\tilde{\boldsymbol{Q}}||_1/||\boldsymbol{Q}||_1$ in figure 8-(a). The red dashed line separates solutions that satisfy the sparsity-promoting constraint (23b) (white area feasibility region) from solutions that do not (red area). Note that the stability constraint is satisfied for all points reported in this figure, as we have noted that the optimiser is still able to satisfy (23c) when it first fails to satisfy the sparsity-promoting constraint (23b).

For small penalisation weights, the optimisation problem has feasible solutions that fall on the boundary of the feasibility region. This suggests that rotations of the original POD basis that minimise the energy loss are found on the boundary, i.e. that there is a competition between sparsification and energy representation. The key feature of figure 8-(a) is that there exists a threshold value $\xi_t$ above which the optimisation problem terminates unsuccessfully in the unfeasible region, i.e. no rotation exists that can reduce the $l_1$ norm of the rotated quadratic coefficient tensor below $\|\boldsymbol{Q}\|_1/\xi_t$. This is manifested in panel (a) by a sudden turn of the solution traces from the feasible region boundary upwards into the red region. This occurs because, for a given dimension $M$, there is only a limited number of coefficients in the tensor $\tilde{\boldsymbol{Q}}$ that can be shrunk to zero by any rotation of the basis functions. The threshold value increases with the ratio $M/N$, i.e. the $l_1$ norm of the rotated coefficient tensor $\tilde{\boldsymbol{Q}}$ can be decreased further when larger dimensions $M$ are used. This is arguably a consequence of the fact that higher ratios $M/N$ corresponds to more degrees of freedom available to the optimiser to ensure that the sparsity constraint is satisfied. It is worth noting that, in LASSO-based sparsification

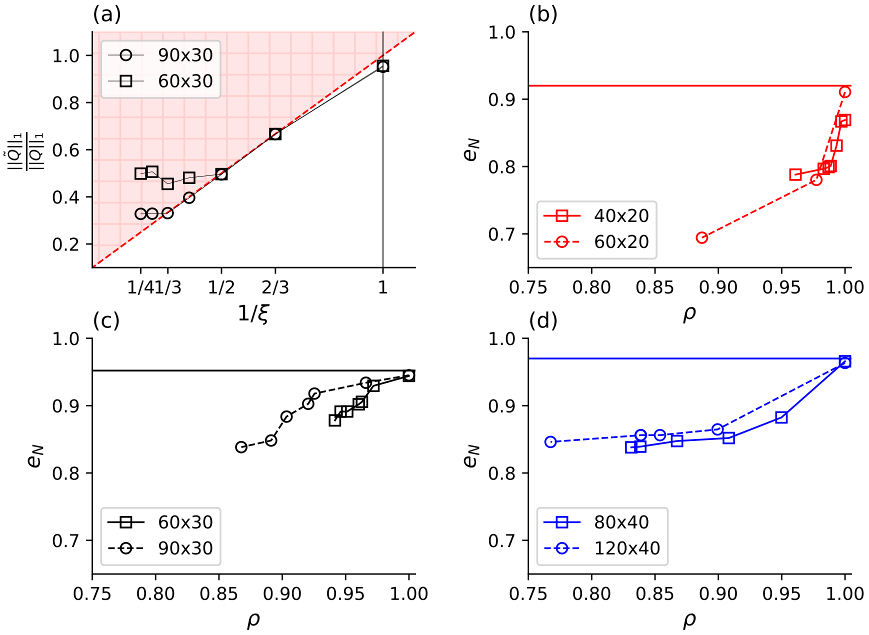

Figure 8: Panel (a): the sparsity-promoting constraint (23b) on the plane $(1/\xi - ||\tilde{\boldsymbol{Q}}||_1/||\boldsymbol{Q}||_1)$. The red region denotes the infeasibity set. Panels (b) to (d): $\rho - e_N$ curves for families of models with dimension $N = 30, 20, 40$ and ratios $M/N = 2, 3$ (open squares and open circles, respectively). The labels in the legend are defined with key $M \times N$.

methods just discussed, the optimiser operates directly on the model coefficients and feasible solutions can always be found, with all coefficients shrunk to zero in the limiting case. However, highly sparsified models are observed to have little physical significance and poor temporal behaviour. Here, all feasible systems with varying sparsity/energy reconstruction properties are temporally stable and provide physically consistent predictions. This result is an effort to try to develop a methodology to ensure global stability in data-driven models, extending the work done by Kaptanoglu et al. [2021].

In panels (c, b d) of figure 8 the trade-off between sparsity and the energy reconstruction properties of the rotated basis is presented on the $\rho - e_N$ plane for models constructed with $N = 20, 30, 40$, respectively. The horizontal line in each panel corresponds to the fraction of reconstructed energy of the original dense POD-Galerkin model. In these panels, squares are used to denote data for $M/N = 2$, while circles denote data for $M/N = 3$. It can be first observed that systems for $\rho = 1$ ($\xi = 1$) do not reconstruct the entire average fluctuation kinetic energy captured by the original POD basis. This is due to the stability constraint (23c), producing a small rotation of the optimal POD basis so that dissipative mechanisms in the Galerkin models to ensure long-term stability are better resolved. Second, when the penalisation $\xi$ is increased, sparser models are obtained, with higher ratios $M/N$ enabling further reduction in density. This, however, comes at the cost of decreasing the energy optimality of the rotated basis. More importantly, the larger the model size, the more the model can be sparsified without significantly affecting the ability of the new basis to reconstruct the fluctuation kinetic energy. This suggests that the sparsification technique becomes more effective as the model complexity and the range of scales resolved by the basis, increases. This appears to be a general trend,

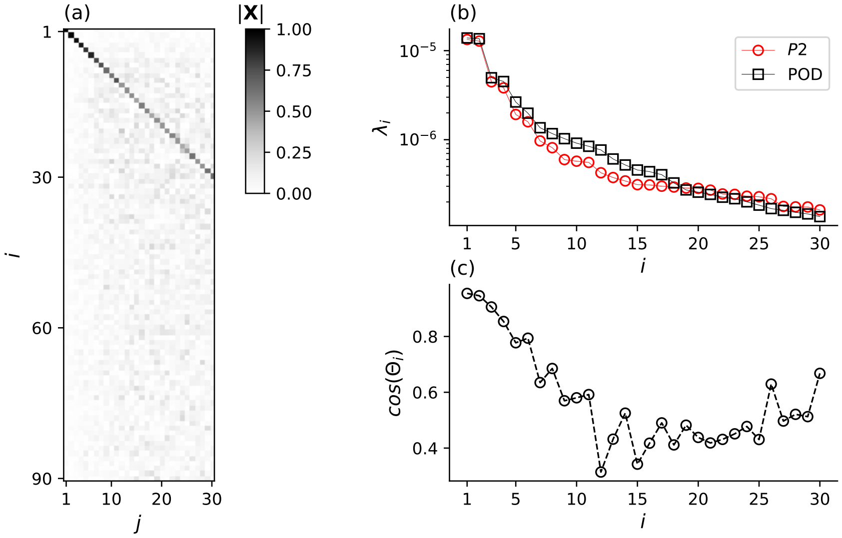

Figure 9: Optimisation results for the case $M/N = 3$, $N = 30$, $\xi = 3$ (with density $\rho = 0.87$). Panel (a): magnitude of the entries of the transformation matrix $\boldsymbol{X}$. Panel (b): distribution of the average modal energies of the original (square) and rotated (circles) basis functions. Panel (c): cosine of the angle between the modes of the original and rotated basis.

since a similar behaviour is observed using the LASSO-based *a-posteriori* sparsification.

### 4.4.1 Analysis of the rotated modal structures

We now move to the analysis of the rotated spatial and temporal basis functions. A model with $N = 30$, $M/N = 3$ and density $\rho = 0.87$, obtained for $\xi = 3$ just before the solution falls into the infeasible region in figure 8-(a) is analysed. Panel (a) of figure 9 shows the magnitude of the entries of the rotation matrix $\boldsymbol{X}$, found from the solution of (23). Panel (b) of the same figure compares the modal energies of the original POD temporal coefficients with those of the rotated modes. Panel (c) shows the cosine of the angle between each pair of original and rotated spatial modes, $\cos(\theta_i) = (\boldsymbol{\phi}_i, \tilde{\boldsymbol{\phi}}_i)$, which is clearly also the diagonal of $\boldsymbol{X}$, because of the orthogonality constraint (23d). The transformation matrix $\boldsymbol{X}$ has large diagonal entries, but significant off-diagonal terms can be observed for $i, j \gtrsim 5$. This indicates that the rotated basis functions bear a strong resemblance to the original POD modes, but that the optimisation has introduced into the new basis small-scale, low-energy features to both stabilise and sparsify the rotated Galerkin model. It is observed that the high-energy temporal and spatial modes are not significantly affected by the rotation and do not differ significantly from the original POD modes. For instance, the first pair of modal energies, corresponding to the dominant fluid oscillation in the cavity, is virtually unchanged. Conversely, high-index, low-energy modes are more significantly rotated away from the corresponding original POD mode and more significant relative differences of the modal energies are observed. We argue that this behaviour derives from the formulation of problem (23), constructed with the aim of generating a basis that minimises the energy loss with respect to the energetically optimal POD. As a result, the optimisation leaves mostly unchanged the most energetic modes that contribute more pronouncedly to the overall energy and rotates by a larger extent the less energetic modes to gain in sparsity and to achieve stability.

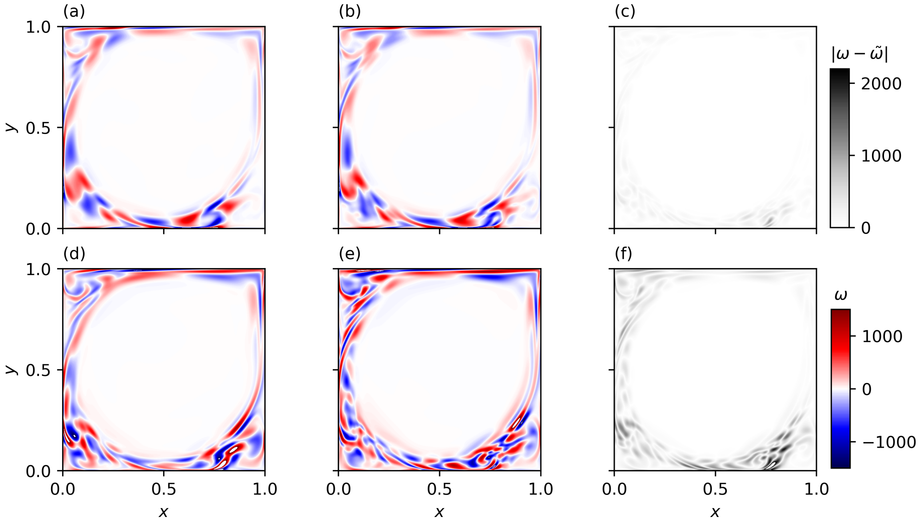

Figure 10: Vorticity fields of the first original POD mode, panel (a), the first rotated mode (denoted as $\tilde{\omega}$), panel (b), and the absolute value of their difference, panel (c). Panels (d), (e) and (f) show the same quantities for mode $i = 19$.

Figure 10 shows the out-of-plane vorticity component $\omega$ for the original POD spatial mode, panels (a,d), and the rotated mode, panels (b,e), for indices $i = 1$ and 19, respectively. The third column, panels (c,f), shows the absolute value of the difference between the original and rotated modes. As expected, the spatial structure of the first mode is not considerably changed by the rotation. On the other hand, mode $i = 19$ is more pronouncedly affected by the rotation, with small-scale vorticity features appearing all along the shear layer. We argue that the introduction of small scale features is a combined effect of the stability constraint (23c), which enhances dissipation in the system [Balajewicz et al., 2016], and of the sparsity-promoting constraint (23b) since spatially fluctuating modes, with stronger gradients, are likely more effective to produce globally zero spatial averages involved in the projection coefficients (4)-(c).

The temporal modes are affected in a similar way. This is illustrated in figure 11 showing the temporal evolution of modes $a_1(t)$ and $a_{19}(t)$ over twenty time units, panels (a) and (b), and their amplitude spectra in panels (c) and (d), respectively. Since the first column of $\boldsymbol{X}$ is close to zero except for $\mathsf{X}_{11}$, mode $a_1(t)$ and its spectral content is not appreciably affected by the rotation, except for a small general decrease of the amplitude due to decrease in energy content (see figure 9-(b)). Conversely, the spectral content of mode $a_{19}(t)$ is remodulated by the rotation by introducing higher energy at high frequency components, consistent with the introduction of small-scale features into the corresponding spatial mode.

### 4.4.2 Interactions identified in the sparse model

The structure of the spatial and temporal modes is only weakly modified by the optimisation, but this is sufficient to introduce sparsity in the rotated quadratic coefficient tensor $\tilde{\boldsymbol{Q}}$ and in the rotated average triadic interaction tensor $\tilde{\boldsymbol{N}}$.

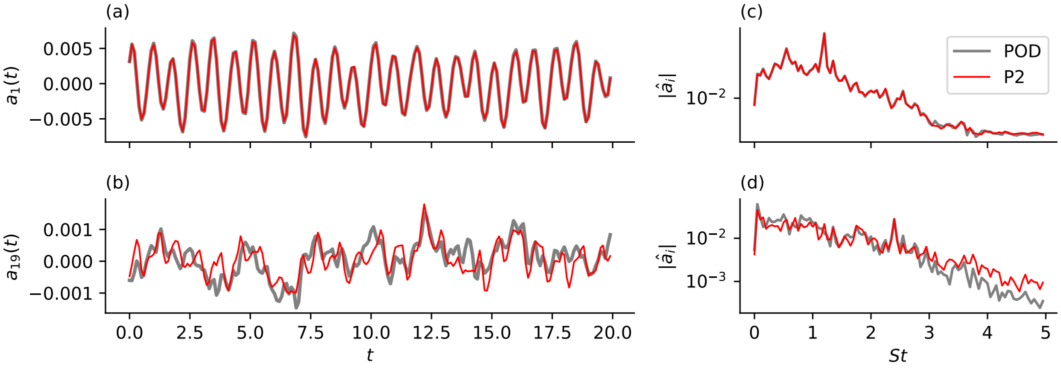

Figure 11: Temporal evolution and amplitude spectrum of $a_1(t)$ in panels (a) and (c), respectively, and for $a_{19}(t)$ in panels (b) and (d). Data is reported for the original POD temporal modes and for the rotated modes of the sparse system obtained from formulation P2.

To visualise how sparsity in these tensors varies when the penalisation weight $\xi$ is increased, we introduce the tensor $\boldsymbol{\xi}$ with entries $\xi_{ijk}$ defined as the value of $\xi$ in (23) at which the corresponding coefficient $\tilde{\boldsymbol{Q}}_{ijk}$ is first shrunk to zero. Figure 12 shows two slices of $\boldsymbol{\xi}$ for $i = 1$ and $i = 30$, for a model with $N = 30$ and $M/N = 3$. Results in panels (a), (b) and (c) are obtained with the complete formulation P2, including both the stability and sparsity promoting constraint, while those in panels (d), (e) and (f) are obtained with formulation P1, which only includes the sparsity promoting constraint. Panels (c) and (f) show the modal density $\rho_i$ as a function of the global density $\rho$ for four modes across the hierarchy. By using formulation P2, which includes the stability constraint (23c), we observe that small-scale/small-scale interactions disappear first, for moderate penalisations, especially for the high-index modes, but generally across the entire hierarchy of modes. Increasing the penalisation, coefficients corresponding to interactions that are local in modal space are progressively pruned, leaving only coefficients capturing non-local interactions with the low-index modes for large penalisations. The key remark is that the structure of the sparsified quadratic coefficient tensor $\tilde{\boldsymbol{Q}}$, solution of the proposed optimisation approach, follows a pattern where coefficients corresponding to energetically weak interactions are pruned first and only relevant interactions are preserved. Interestingly, figure 12-(c) indicates that high-index modes can be sparsified more efficiently. This might be related to the fact that high-index modes are also rotated more aggressively during the optimisation, to ensure that the sparsity promoting constraint (23b) is satisfied and with minor effect of the overall energy reconstruction ability.

Nonetheless, considering now the solution obtained from formulation P1 in the bottom panels of figure 12, it is clear that many quadratic coefficients are indeed shrunk to zero during the optimisation and similar global densities are obtained. However, the sparsity pattern does not have a clear relation with the original structure of energy interactions. By contrast, coefficients corresponding to important energy interactions have been shrunk to zero since the optimisation problem P1 is entirely driven by the sparsity-promoting constraint. While in both formulations the same penalisation on the $l_1$ norm of the rotated tensor $\tilde{\boldsymbol{Q}}$ is used, formulation P1 lacks any information regarding the dynamics and temporal evolution of the Galerkin model and the structure of inter-modal energy transfers. This results in an unphysical equal contraction of all quadratic coefficients, across all modes and independently of the strength of the energy interactions they represent. The effect of this behaviour can also be noticed in the $\rho_i - \rho$ curves, panel (f), where different individual modes are sparsified by a similar

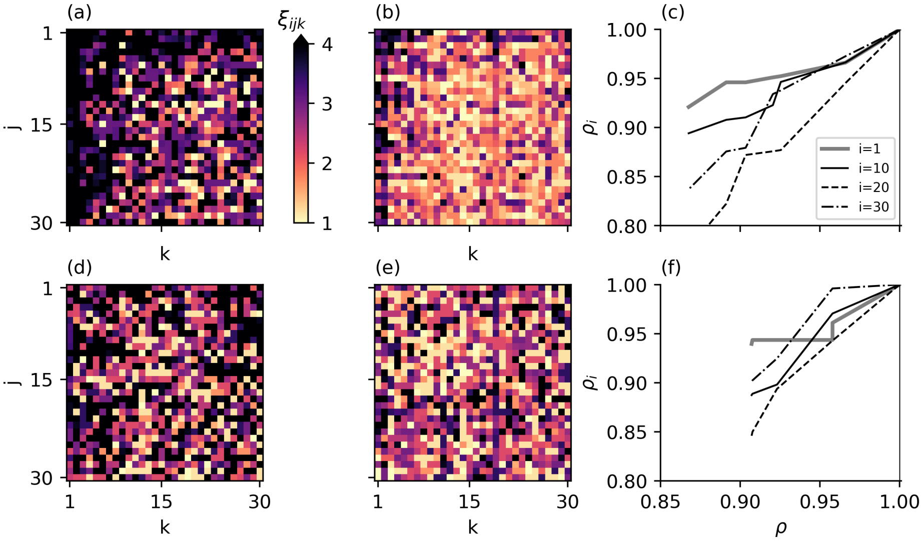

Figure 12: Entries of the tensor $\boldsymbol{\xi}$ for $i = 1, 30$ obtained by solving problem P2, in panels (a) and (b), respectively. In panel (c): modal densities $\rho_i$ as a function of $\rho$. In panels (d), (e) and (f), the same quantities are shown from solution of the optimisation problem P1.

amount. Nevertheless, no significant differences are observed between models obtained from the two formulations on the $\rho - e_N$ plane.

### 4.4.3 Temporal integration and energy analysis of the sparsified system

In this section, we analyse the temporal behaviour of the sparse reduced order models obtained by solving problems P1 and P2 and we consider their rotated triadic interactions tensors $\tilde{\boldsymbol{N}}$. The same minimum-density configuration studied in previous sections, with $M/N = 3$, $N = 30$ and $\xi = 3$, is considered. This configuration is used here as a representative example of the sparse models obtained with the proposed approach. In fact, the spatio-temporal statistical behaviour of these models is only weakly affected by the regularisation parameter $\xi$, but is mostly dominated by the inclusion of the stability constraint, regardless of the sparsity. Models are integrated forward in time with an implicit time stepping scheme for $T = 500$ time units, from an initial condition obtained from one of the snapshots. Figure 13-(a) shows the first one hundred time units of the temporal evolution of the fluctuation kinetic energy for these two models, compared with the evolution from DNS and from the original POD-Galerkin model. Figure 13-(b,c) show the probability density function of the same quantity, computed over a longer time span. As expected, the deficit of dissipation in the original POD-Galerkin model produces fluctuation kinetic energy levels about two orders of magnitude larger than the reference value from DNS. This behaviour is well known (see e.g. Östh et al. [2014], Noack et al. [2016]). The sparse Galerkin model obtained without the stability constraint (problem P1) also reproduces this behaviour. Conversely, the sparsified and stabilised model obtained from problem P2 has realistic temporal behaviour and resolves correctly the average energy and its fluctuation observed in DNS. This suggests that including in the optimisation information on how modal structures are supposed to interact with each other to satisfy the overall power budget is key to obtain sparse Galerkin models with adequate predictive ability. Note that the long-term performance of the models obtained

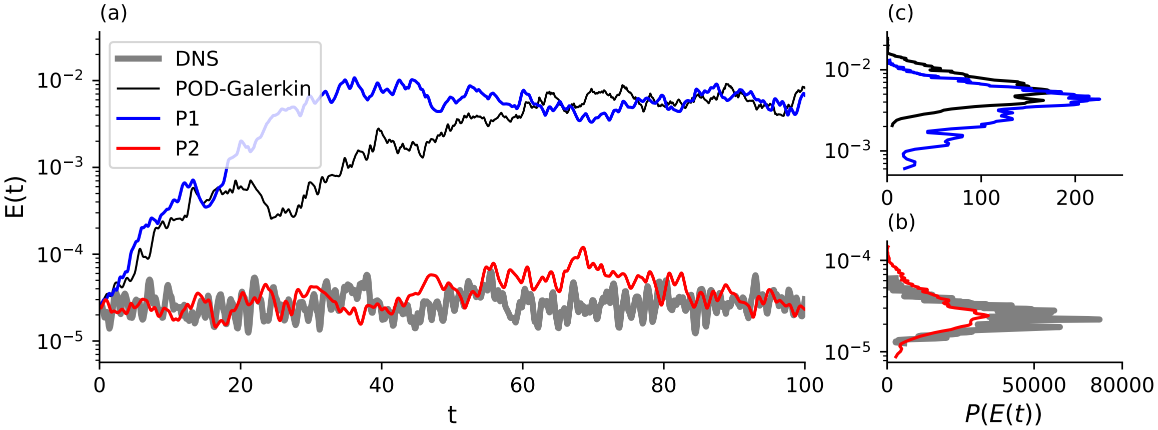

Figure 13: Temporal evolution of the integral fluctuation kinetic energy $E(t)$. The temporal performance of the two sparse models obtained by solving problem (20) are compared against the DNS and the dense model obtained with Galerkin projection. Panels (b,c) show the probability distribution of the energy $E(t)$, obtained for long-time integration of the Galerkin models.

in the present work is generally superior to that of models sparsified *a posteriori*, using a LASSO-based approach, as discussed previously. More specifically, LASSO-based models have been found to be temporally accurate over a time span comparable to that of the data used for the sparsification. Conversely, the present *a priori* sparsified systems are temporally stable for much longer time horizons. In addition, numerical solutions converge to the asymptotic attractor for much larger perturbations of the initial conditions. We argue that this is likely the consequence of enforcing a stronger consistency between the modal structures and the corresponding Galerkin model.

The structure of the nonlinear energy transfer rate tensor $\tilde{\boldsymbol{N}}$ for the rotated Galerkin models is reported in figure 14, showing two slices for $i = 1$ and 30 (first and second row, respectively). Panels (a) and (d) refer to the model obtained from Galerkin projection onto the original POD subspace and will be used as reference. Here the temporal coefficients are the projections onto the DNS snapshots. Data for the model obtained from problem P2 is shown in panels (b) and (e) while data for the model obtained with the P1 formulation is shown in panels (c) and (f). For these models, temporal coefficients are obtained from forward integration of the Galerkin models. It can be observed that the model from formulation P2 displays a pattern of interactions consistent with the DNS data projected onto the original POD modes, both in terms of organisation across modes as well as in strength. The organisation of the interactions is similar to that in the dense Galerkin model obtained using the *a-priori* stabilisation method of [Balajewicz et al., 2013], although a more aggressive reduction in the strength of the interactions between the high-index modes is observed. By contrast, energy interactions in the model obtained from the formulation P1 are orders of magnitude more intense, across all triads. This is a manifestation of the lack of dissipation and the consequential over-prediction of energy, across all modes. Interestingly, it can be observed that the HL/LH asymmetry observed in figure 2 is preserved throughout the sparsification. Conversely, this physical feature is lost in the *a posteriori* approach, for numerical reasons. It is arguable that the *a priori* approach is preferable for systems where $\boldsymbol{N}$ has a complex structure.

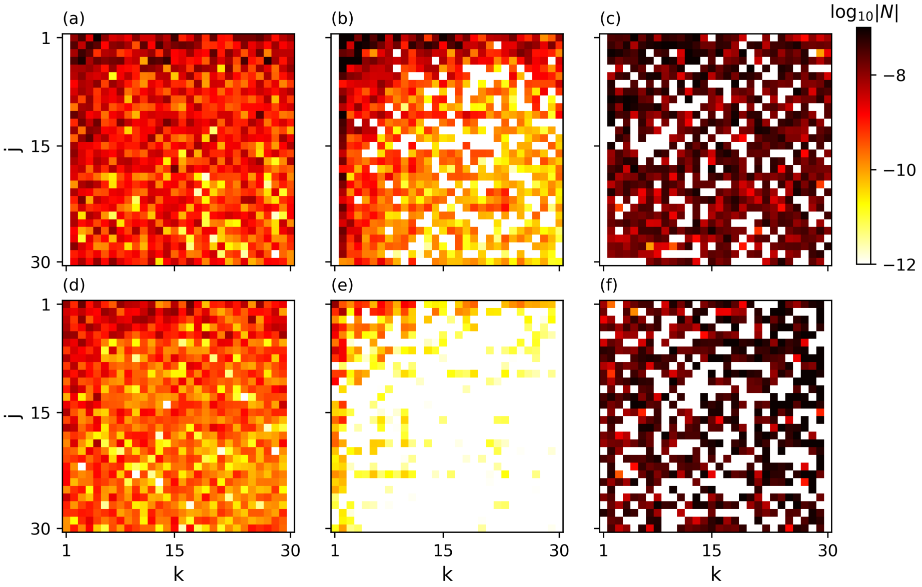

Figure 14: Intensity of the average nonlinear energy transfer rate $\tilde{\boldsymbol{N}}$ for mode $i = 1$ in models obtained from projection and formulations P2 and P1 in panels (a), (b) and (c), respectively. Panels (d), (e) and (f) show the same quantity for mode $i = 30$.

## 5 Conclusions

In this project, we developed and applied data-driven techniques for the identification of sparse projection-based reduced order models of two-dimensional unsteady flows. Our project was motivated by the fact that scale interactions in turbulent flows are sparse, and the motion at any given length scale depends most prominently on the dynamics of a subset of all other scales. Classical reduced order modelling method do not leverage these characteristics. Here, the goal has been to derive sparse large reduced order models covering a wide range of length scales, providing further insight into complex nonlinear mechanism and facilitating the interpretation of scale interactions.

We have proposed two separate sparsification methodologies, named “a-priori” and “a-posteriori” sparsification, differing by the order in which sparsification and order reduction are performed. These methodologies utilise advanced tools from optimisation and/or machine learning and have been shown to scale reasonably well to fluid models composed of tens/hundreds of modal structures. The common trait of these techniques, like any other sparse coding method, is that they invariably include an arbitrary tuning constant. This constant defines a family of models varying in accuracy and sparsity, i.e. tracing a curve on a sparsity/accuracy plane. Increasing sparsity beyond a threshold is often accompanied by a lower prediction accuracy, although a “sweet spot” on this curve often exists, whereby great sparsification can be achieved for a small reduction in prediction abilities.

Throughout the research, we have considered two-dimensional lid-driven cavity flow at a Reynolds number ($Re = 2 \times 10^4$) where the motion is chaotic. Unlike in simpler periodic flows, scale interactions in this flow problem are scattered in modal space, representing a more challenging benchmark case for the two proposed techniques. The analysis of the average energy transfer rates across modal structures in this flow has shown that a subset of relevant nonlinear interactions exists, in agreement with the established picture of scale interactions in two-dimensional flows.

Regarding the “a-posteriori” sparsification method, the main conclusions are as follows:

1. The sparse models preserve to a good degree of accuracy the non-local nature of triadic interactions and the conservation properties of the convective term of the Navier-Stokes equations. In principle, energy conservation could be enforced exactly [Loiseau and Brunton, 2018], although we have not found this to be necessary to obtain satisfactory temporal stability characteristics. In fact, unlike dense models obtained directly from the projection, the $l_1$ sparsified POD-based models have excellent long-term stability properties.

2. Sparse models constructed by a naive procedure where weak interactions are pruned, referred to as greedy approach in Rubini et al. [2020], do not enjoy the same robustness and have worse performance than the dense models. This indicates that, once coefficients corresponding to weak interactions have been shrunk to zero by the $l_1$ penalisation, re-balancing the remaining coefficients with the least-squares term in the LASSO problem (12) is key to preserve accuracy.

3. We have also observed that the effectiveness of the sparsification grows with the number of modes utilised in the projection (energy resolution). This is a result of the non-local nature of scale interactions in two-dimensional flows, where the dynamics of small-scale features is dominated by the advection of the large modes, rather than by the small-scale/small-scale nonlinearity. The interesting consequence is that, while the total number of quadratic interactions grows cubically with the number of modes, the number of relevant interactions does not grow as quickly. Hence, our expectation is that sparsification becomes more effective as the Reynolds number increases, as a result of the increased scale separation.

4. The sparsification of models constructed using POD and DFT modes has shown that the sparsity of energy interactions is not necessarily invariant when analysed on different subspaces. This results stimulated the development of the “a-priori” sparsification technique, aiming at finding a set of basis functions (the projection subspace) for which maximum sparsity can be achieved, i.e. a set of “least interacting” modes.

Regarding the “a-priori” sparsification method, the main conclusions are as follows:

1. As opposed to the “a-posteriori” sparsification approach, the “a-priori” methodology maintains the exact link between the (sparse) Galerkin model and the modal structures utilised for the projection. This can facilitate the interpretation of the energy flows and the spatial structure of the basis functions.

2. There is a competition between sparsification and energy representation and this trade-off depends on the model size $N$ and the ratio $M/N$. In particular, larger models can be sparsified more aggressively with less impact on the energy reconstruction ability. On the other hand, higher sparsity can be obtained by rotating the same $N$ POD basis functions within larger subspaces (a higher ratio $M/N$), since the increased flexibility allows shrinking to zero a larger fraction of quadratic interaction coefficients.

3. A physically-consistent organisation of the interactions and a stable long-term behaviour can only be obtained by augmenting the sparsification procedure with the temporal stability constraint. In fact, models obtained without such constraint inherit well-known temporal stability issues displayed by dense POD-Galerkin models

## 5.1 Similarities and differences

In the course of the research, similarities and differences between the two approaches have been observed. First of all, both approaches have been proved able to recover the correct physics of the energy interactions. The structure of the sparse systems shows that the relevant energy pattern is correctly identified by both case, and energy conservation properties of the non linear interactions are preserved. Generally, it has been observed that both approaches perform better in case of larger systems (corresponding to a larger amount of reconstructed turbulent kinetic energy) that can be sparsified more aggressively with less impact both on the reconstruction error. This is promising, especially for application to flows at higher Reynolds number. In addition, all sparse systems have been shown to provide quite accurate predictions of the reconstructed flow field. These are interesting results, showing that the implicit tuning present in both approaches can balance both the unresolved physics associated with the smaller scales and the unmodelled pressure term for open flows.

On the other hand, the "a-posteriori" formulation resulted in much higher sparsities, with little impact on the prediction accuracy. It is arguable that this difference is due to the fact that the "a-priori" approach has a stronger tie back to the physics of the flow. In addition, for the "a-posteriori" sparsification approach the regularisation weight can be chosen to be arbitrarily large resulting in systems with little physical meaning or practical utility. By contrast for the "a-priori" sparsification technique, the optimally sparse system can be identified very clearly as the system with the lowest density that still satisfies the sparsity and stability constraints in the optimisation problem.

Lastly, few considerations about the computational complexity and the computational cost of the two algorithms are in order. One of the key aspects of the "a-posteriori" sparsification consists in recasting the heuristic problem of selecting the most relevant interactions in Galerkin models into a convex optimisation problem. Crucially, this problem is characterised by a unique minimum that can be efficiently found with gradient-based techniques. Solution of this problem is feasible for models with few hundreds modes. Conversely, the formulation of the "a-priori" sparsification framework requires a much higher computational effort. In the present work, we have been able to sparsify systems composed of up to fifty modes. However, it is the authors opinion that there is still room for improvement to scale this framework up to computational performances comparable with regression based techniques.

## 5.2 Outlook

There are two main avenues of research that need to be further explored to obtain a more complete understanding of sparsification techniques for Galerkin models of fluid flows. From a computational point of view the "a-posteriori" sparsification technique is quite well understood. The mathematical framework of $l_1$-based regression is well tested and computationally efficient algorithms for its solution are available in the vast majority of machine learning libraries. Conversely, regarding the "a-priori" approach several aspects deserve further investigation. For instance, the optimisation problem to identify suitable basis functions for the projection is highly non-convex and many local optima, i.e. many different sets of basis functions, exist. This is in stark contrast to most (if not all) existing modal decomposition methods (i.e. POD and SPOD, DMD, Resolvent Analysis), which are formulated as a convex optimisation problem or are the results of a matrix decomposition. It is likely that other mathematical formulations exist that lead to a clearer and more succinct description of scale interactions, especially in low dimensional flow considered in the literature. The authors, and the reviewers of our publications [Rubini et al., 2020, 2021], feel that our proposed methods have only scratched the surface and that more work on this front is warranted.

From a more physical perspective, it has been shown that for two-dimensional flows, where interactions are strongly non-local, the sparsification can generate reduced order models with computational complexity scaling as $N^a$ with $a < 3$ instead of $N^3$, as for dense models, with $N$ the number of modes. It would be of great interest to quantify the scaling of computational complexity in three-dimensional flows, where triadic interactions are local in modal space. The interaction between sparsification, flow physics and computational costs of the reduced order models is an aspect that we have only been able to touch briefly, but deserves certainly more investigation, especially for three-dimensional turbulent flows in realistic configurations.

# References

J. A. Domaradzki, W. Liu, C. Härtel, and L. Kleiser. Energy transfer in numerically simulated wall-bounded turbulent flows. *Physics of Fluids*, 6(4):1583–1599, 1994.

S. B. Pope. Turbulent flows, 2001.

HK Moffatt. Note on the triad interactions of homogeneous turbulence. *Journal of Fluid Mechanics*, 741, 2014.

A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 434(1890):9–13, 1991.

A. Craik. Non-linear resonant instability in boundary layers. *Journal of Fluid Mechanics*, 50(2): 393–413, 1971.

D. Rempfer and H. F Fasel. Dynamics of three-dimensional coherent structures in a flat-plate boundary layer. *Journal of Fluid Mechanics*, 275:257–283, 1994a.

D. Rempfer and H. F. Fasel. Evolution of three-dimensional coherent structures in a flat-plate boundary layer. *Journal of Fluid Mechanics*, 260:351–375, 1994b.

O. T. Schmidt. Bispectral mode decomposition of nonlinear flows. *Nonlinear Dynamics*, 102(4): 2479–2501, 2020.

B. R. Noack, M. Schlegel, B. Ahlborn, G. Mutschke, M. Morzyński, P. Comte, and G. Tadmor. A finite-time thermodynamics of unsteady fluid flows. *Journal of Non-Equilibrium Thermodynamics*, 33(2):103–148, 2008.

J. L. Lumley. Computational modeling of turbulent flows. In *Advances in applied mechanics*, volume 18, pages 123–176. Elsevier, 1979.

J. Jiménez. Coherent structures in wall-bounded turbulence. *Journal of Fluid Mechanics*, 842, 2018.

R. H. Kraichnan. Inertial-range transfer in two-and three-dimensional turbulence. *Journal of Fluid Mechanics*, 47(3):525–535, 1971.

K. Ohkitani. Nonlocality in a forced two-dimensional turbulence. *Physics of Fluids A: Fluid Dynamics*, 2(9):1529–1531, 1990.

J. Brasseur and C. Wei. Interscale dynamics and local isotropy in high Reynolds number turbulence within triadic interactions. *Physics of Fluids*, 6, 1994.

J. P. Laval, B. Dubrulle, and S. Nazarenko. Nonlocality of interaction of scales in the dynamics of 2D incompressible fluids. *Physical Review Letters*, 83(20):4061, 1999.

V. L. Thomas, B. F. Farrell, P. J. Ioannou, and D. F. Gayme. A minimal model of self-sustaining turbulence. *Physics of Fluids*, 27(10):105104, 2015.

V. L. Thomas, B. K. Lieu, M. R. Jovanović, B. F. Farrell, P. J. Ioannou, and D. F. Gayme. Self-sustaining turbulence in a restricted nonlinear model of plane Couette flow. *Physics of Fluids*, 26 (10):105112, 2014.

C. A. J. Fletcher. *Computational Galerkin Methods*. 1st edn. Springer, 1984.

C. W. Rowley and S. T. M. Dawson. Model reduction for flow analysis and control. *Annual Review of Fluid Mechanics*, 49:387–417, 2017.

B. Noack, M. Morzynski, and G. Tadmor. *Reduced-order modelling for flow control*, volume 528. Springer Science & Business Media, 2011.

M. Couplet, P. Sagaut, and C. Basdevant. Intermodal energy transfers in a Proper Orthogonal Decomposition–Galerkin representation of a turbulent separated flow. *Journal of Fluid Mechanics*, 491, 2003.

J. L. Lumley. Stochastic tools in turbulence. Technical report, 1970.

L Sirovich. Turbulence and the dynamics of coherent structure. Part i, ii, iii. *Quat. Appl. Math.*, 3: 583, 1987.

P. K. Yeung, J. G. Brasseur, and Q. Wang. Dynamics of Direct Large-Small Scale Couplings in Coherently Forced Turbulence: Concurrent Physical- and Fourier-Space Views. *Journal of Fluid Mechanics*, 283:43–95, January 1995.

J. C. Loiseau and S. L. Brunton. Constrained sparse Galerkin regression. *Journal of Fluid Mechanics*, 838:42–67, 2018.

K. Taira, S. L. Brunton, S. T. M. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S Ukeiley. Modal analysis of fluid flows: An overview. *AIAA Journal*, pages 4013–4041, 2017.

B. R. Noack and H. Eckelmann. A Global Stability Analysis of the Steady and Periodic Cylinder Wake. *Journal of Fluid Mechanics*, 270:297–330, January 1994.

S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113 (15), 2016.

S. L. Brunton, B. R. Noack, and P. Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52, 2019.

J. Friedman, T. Hastie, and R. J. Tibshirani. *The Elements of Statistical Learning*. Springer, 2008.

R. J. Tibshirani. The LASSO problem and uniqueness. *Electronic Journal of Statistics*, 7:1456–1490, 2013.

L. Cordier, B. A. El Majd, and J. Favier. Calibration of POD reduced-order models using Tikhonov regularization. *International Journal for Numerical Methods in Fluids*, 63(2):269–296, 2010.

E. Kaiser, J. N. Kutz, and S. L. Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proc. Math. Phys. Eng. Sci.*, 474(2219):20180335, November 2018.

A. G. Nair and K. Taira. Network-theoretic approach to sparsified discrete vortex dynamics. *Journal of Fluid Mechanics*, 768:549–571, 2015.

K. Taira, A. G. Nair, and S. L. Brunton. Network structure of two-dimensional decaying isotropic turbulence. *Journal of Fluid Mechanics*, 795, 2016.

A G Nair, S L. Brunton, and K Taira. Networked-oscillator-based modeling and control of unsteady wake flows. *Phys. Rev. E*, 97:063107, Jun 2018.

M. Newman. *Networks*. Oxford university press, 2018.

L. Perret, E. Collin, and J. Delville. Polynomial identification of POD based low-order dynamical system. *Journal of Turbulence*, (7):N17, 2006.

R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, B*, 58(1), 1996.

M. Couplet, C. Basdevant, and P. Sagaut. Calibrated reduced-order POD-Galerkin system for fluid flow modelling. *Journal of Computational Physics*, 207(1):192–220, 2005.

M. R. Jovanović, P. J. Schmid, and J. W. Nichols. Sparsity-promoting Dynamic Mode Decomposition. *Physics of Fluids*, 26(2):024103, 2014.

C. Ramirez, V. Kreinovich, and M. Argaez. Why $l_1$ is a good approximation to $l_0$: A geometric explanation. *Journal of Uncertain Systems*, 7(3):203–207, 2013.

L. Zhang and H. Schaeffer. On the convergence of the SINDy algorithm. *Multiscale Modeling & Simulation*, 17(3):948–972, 2019.

F. Pedregosa and G. et al. Varoquaux. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2011.

M. Schmidt and H. Lipson. Distilling Free-Form Natural Laws from Experimental Data. *Science*, 324 (5923):81–85, 2009.

M. Balajewicz, E. Dowell, and B. Noack. Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier–Stokes equation. *Journal of Fluid Mechanics*, 729, 2013.

M. Balajewicz, I. Tezaur, and E. Dowell. Minimal subspace rotation on the Stiefel manifold for stabilization and enhancement of projection-based reduced order models for the compressible Navier–Stokes equations. *Journal of Computational Physics*, 321:224–241, 2016.

M. Schlegel and B. R. Noack. On long-term boundedness of Galerkin models. *Journal of Fluid Mechanics*, 765:325–352, 2015.

S. Grimberg, C Farhat, and N. Youkilis. On the stability of projection-based model order reduction for convection-dominated laminar and turbulent flows. *Journal of Computational Physics*, 419: 109681, 2020. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2020.109681. URL `https://www.sciencedirect.com/science/article/pii/S0021999120304551`.

B. Galletti, C.H. Bruneau, L. Zannetti, and A. Iollo. Low-order modelling of laminar flow regimes past a confined square cylinder. *Journal of Fluid Mechanics*, 503:161–170, 2004.

B. R. Noack, P. Papas, and P. A. Monkewitz. The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows. *Journal of Fluid Mechanics*, 523:339–365, 2005.

J. Östh, B. R. Noack, S. Krajnović, D. Barros, and J. Borée. On the need for a nonlinear subscale turbulence term in POD models as exemplified for a high-Reynolds-number flow over an Ahmed body. *Journal of Fluid Mechanics*, 747:518–544, 2014.

S. G. Johnson. The nlopt nonlinear-optimization package, 2014.

K. Svanberg. A class of globally convergent optimization methods based on conservative convex separable approximations. *SIAM Journal on Optimization*, pages 555–573, 2014.

R. Rubini, D. Lasagna, and A. Da Ronch. *A-priori* sparsification of galerkin based reduced order-models. *Journal of Fluid Mechanics*, under review, 2021.

R. Rubini, D. Lasagna, and A. Da Ronch. The l1-based sparsification of energy interactions in unsteady lid-driven cavity flow. *Journal of Fluid Mechanics*, 905, 2020.

W. Cazemier, R. W. C. P. Verstappen, and A. E. P. Veldman. Proper Orthogonal Decomposition and low-dimensional models for driven cavity flows. *Physics of fluids*, 10(7):1685–1699, 1998.

F. Terragni, E. Valero, and J. M. Vega. Local POD plus Galerkin projection in the unsteady lid-driven cavity problem. *SIAM Journal on Scientific Computing*, 33(6):3538–3561, 2011.

H. Arbabi and I. Mezić. Study of dynamics in post-transient flows using Koopman mode decomposition. *Physical Review Fluids*, 2(12):124402, 2017.

L. Fick, Y. Maday, A. T. Patera, and T. Taddei. A stabilized POD model for turbulent flows over a range of reynolds numbers: Optimal parameter sampling and constrained projection. *Journal of Computational Physics*, 371:214 – 243, 2018.

F. Auteri, N. Parolini, and L. Quartapelle. Numerical investigation on the stability of singular driven cavity flow. *Journal of Computational Physics*, 183(1):1–25, 2002.

Y. H. Peng, Y. H. Shiau, and R. R. Hwang. Transition in a 2-D lid-driven cavity flow. *Computers & Fluids*, 32(3):337 – 352, 2003.

M. A. Mendez, M. Balabane, and J.-M. Buchlin. Multi-scale proper orthogonal decomposition of complex fluid flows. *Journal of Fluid Mechanics*, 870:988–1036, 2019. doi: 10.1017/jfm.2019.212.

A. A. Kaptanoglu, J. L. Callaham, C. J. Hansen, A. Aravkin, and S. L. Brunton. Promoting global stability in data-driven models of quadratic nonlinear dynamics. *arXiv preprint arXiv:2105.01843*, 2021.

B. R. Noack, W. Stankiewicz, M. Morzyński, and P. J. Schmid. Recursive dynamic mode decomposition of transient and post-transient wake flows. *Journal of Fluid Mechanics*, 809:843–872, November 2016.