

1.5in 0.6in 1.0in 0.8in 20pt 0.25in 9pt 0.3in

UNIVERSITY OF SOUTHAMPTON

Sparsification techniques for Reduced Order Models of turbulent flows

by

Riccardo Rubini

Supervisors:

Dr. Davide Lasagna

Dr. Andrea Da Ronch

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the

Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

25-June-2018

Abstract

The ability to perform fast and accurate simulations is becoming a more and more compelling requirement in the applications concerning fluid flows. Nowadays, the capabilities of the current computers have made the high fidelity simulations more and more widespread. However, performing high fidelity simulations of realistic flow field configurations remains a very demanding endeavour. To overcome these difficulties and reduce the computational time designers have developed reduced order models (ROM) techniques. One of the more promising is the ROM based on the Galerkin projection. It aims to concentrate the most of the information in few degrees of freedom and then to describe complex flow configurations more efficiently but still with high accuracy .However, due to the nonlinear nature of the Navier-Stokes equations the resulting ROM is densely connected. The computational cost of the time integration of a densely connected model grows like the cube of the degree of freedoms taken into consideration. This could make the problem computationally intractable as more degrees of freedom are needed to correctly describe the flow physics.

This work aims to exploit the emerging sparsity structure of high Reynolds number flows in order to generate sparsely connected and computational efficient reduced order models. Well established sparsity-promoting techniques coming from statistical learning will be exploited to identify the sparsity pattern of the corresponding model.

The sparsification procedure will be applied on three test cases of increasing computational and physical complexity, namely : the Kuramoto-Sivashinsky equation, the two dimensional flow inside a closed lid driven cavity and a fully three dimensional turbulent flow over a backward step. Subsequently

once the procedure will be validated, the influence on the Reynolds number on the sparsity pattern of the solution will be investigated.

The preliminary results obtained on the Kuramoto-Sivashinsky model will be then presented. They already show how the increase of the dynamic active scales of motion seems to have an influence on the sparsity pattern of the corresponding ROM and on the accuracy of the solution.

Contents

Abstract	iii
1 Introduction	1
1.1 High Reynolds number flows	1
1.2 Reduced order modelling	2
1.3 Scales interaction in turbulence	4
1.4 Sparse regression	5
1.5 Sparse regression of dynamical systems	7
1.6 Work programme	9
2 Literature Review	11
2.1 Reduced order modelling	11
2.2 Proper orthogonal decomposition	12
2.3 Construction of the reduced order model	13
2.3.1 Pressure term in ROMs models	14
2.3.2 Dissipation in Reduced Order Models	15
2.3.3 Applications of reduced order modelling	16
2.4 Energy flow analysis in POD Galerkin representation of the flow	17
2.5 Scales interaction and energy transfer in turbulence	19
2.6 Sparse identification of dynamical systems	21
2.7 The problem of selecting the relevant features	23
2.7.1 Solution of the LASSO problem	26
2.8 Conclusions	28
3 Sparse Regression of Dynamical Systems	31
3.1 Sparse regression	31
3.2 Sparse regression of dynamical systems	34
3.3 Choice of the sparsification technique	37
3.3.1 Greedy approach	37
3.3.2 Choice of the regularization parameter λ	37

3.4	Example: the Lorenz equations	39
3.5	Sparse POD basis	42
3.5.1	Mathematical formulation	42
3.5.2	Standard POD vs sparse POD	44
3.6	Conclusions	45
4	1-D Turbulence: Kuramoto-Sivashinsky Equation	47
4.1	The mathematical model	47
4.2	Numerical procedure	48
4.3	Energy analysis in Kuramoto-Sivashinsky flow	50
4.4	Regression procedure	55
4.5	Length influence on the sparsity	56
4.5.1	Energy interactions	57
4.5.2	Greedy sparsification	58
4.5.3	l_1 regression with $\lambda_k = \text{const}$	59
4.5.4	l_1 regression with $\lambda_k = <\sqrt{u_k u_k^*}> \lambda$	62
4.6	Comparison between different strategies	64
4.7	Conclusions	65
5	Conclusions	67
6	Future Work	71
6.1	Cross validation	71
6.2	l_1 regression	72
6.2.1	Interface with the DNS code	72
6.2.2	2D cavity flow	73
6.2.3	3D backward facing step	73
A	Planning	75
	Bibliography	79

Chapter 1

Introduction

Nowadays, due to the increasing power and affordability of computational resources, high fidelity simulations are becoming more and more a common practice. Although the general direction of the computational fluid dynamics community seems to move towards large eddy simulations (LES) and direct numerical simulations (DNS), resolving the Navier-Stokes equations at high Reynolds numbers is still a non negligible computational effort. It is not likely that this approach will be introduced as design process at least in the near future. For this reason simplified approaches such as Reynolds-Average-Navier-Stokes (RANS) or DDES (Delayed-Detached-Eddy-Simulations) are still very important tools in the early stages of the design process. However, the performances of many engineering flow systems rely on the description of complex and often unsteady flow phenomenon. The approximations introduced in the RANS and DDES approaches often are not able to provide a satisfactory and general description of the flow.

1.1 High Reynolds number flows

Turbulence plays a very important role in engineering. In fact, the vast majority of flows of practical interest are characterized by high Reynolds number and hence by a fully developed turbulent regime. Although the governing equations that describe a turbulent flow are known since the end of the 18th century their resolution is still a very challenging task both

theoretically and numerically. Turbulence can be seen as the spatial and temporal evolution of field containing a very wide range of different time and space structures. In particular, away from the domain boundary the energy is distributed among the difference flow structures following the Kolgomorov rule [Pope \(2001\)](#). It states that the separation between the biggest and the smallest scales grows with the Reynolds number as $Re^{9/4}$. Hence the need to resolve all the scales implies that high fidelity simulations become not affordable for fairly high Reynolds numbers. The increasing of the number of dynamical active scales is qualitatively sketched in figure 1.1. It shows the energy distribution as a function of the length scale. It is clear how the increasing of the Reynolds number leads to a increasing of scales of motion to take into consideration to fully describe the flow.

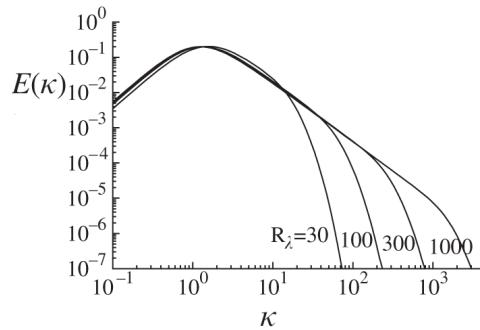


FIGURE 1.1: Different shapes of three different spectra for increasing values of the Reynolds number. In ordinate is reported the kinetic energy of every modes. In abscissa the wavenumber that in proportional to L^{-1}

1.2 Reduced order modelling

A full resolution of all turbulent scales is far beyond the capabilities of the current computers. A promising approach to this problem is Reduce Order Modelling (ROM) technique. The main idea behind Reduced Order Modelling is to project the full Navier-Stokes system onto a low-dimensional basis. Of particular convenience is the model arising from the Galerkin projection of the original system onto the basis obtained by from the Proper Orthogonal Decomposition (POD) of the flow field [Noack et al. \(2011\)](#). This model is appealing mainly because the POD decomposition generate the optimal decomposition of the flow field in term of its energy, hence it is

supposed to provide a optimal low dimensional description of a complex flow field.

The Galerkin projection technique is able to describe complex systems projecting it on a new basis that contains the maximum amount of information. This should allows to study a relatively simpler system respect to the original one that anyway should be able to describe with a good degree of accuracy the physical features of the original system.

Mathematically the procedure to creating a Galerkin reduced order model is the following. It consists firstly to decompose the velocity field in fluctuating and mean field and the fluctuating part is decomposed in POD eigenfunctions that are a spatial shape $\phi(\mathbf{x})$ and a temporal evolution $a(t)$:

$$\mathbf{u}(\mathbf{x}, t) := \mathbf{U}(\mathbf{x}) + \sum_{i=1}^N a_i(t) \phi_i(\mathbf{x}) \quad i = 1, \dots, N, \quad (1.1)$$

where N is the number of modes taken into consideration. Then the Navier-Stokes are projected into the new basis:

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p - \frac{1}{Re} \nabla^2 \mathbf{u} = 0 \quad (1.2)$$

This procedure leads to a system of coupled ordinary differential equations (ODEs):

$$\frac{d}{dt} a_i = C_i + \frac{1}{Re} L_{ij} a_j + Q_{ijk} a_j a_k \quad i = 1, \dots, N \quad (1.3)$$

The computational cost of the time integration of this system grows like N^3 . This is a direct consequence to the fact that the second term at the right hand side of equation 1.3 Q_{ijk} is a tensor of rank three and then the evaluation of $Q_{ijk} a_j a_k$ requires N^3 operations. Moreover, the POD basis functions are globally defined all over the domain, therefore all the entries in the tensor Q_{ijk} are different from zero and need to be computed. This result in a manageable problem for low dimensional systems but could become an issue for complex systems where a high number of modes is needed to describe the physics correctly.

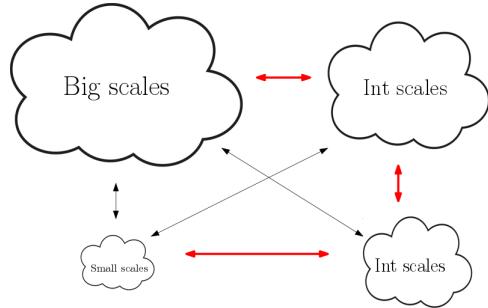
The bottom line is that even if the problem is rewritten on more convenient coordinates system the physical complexity of the original model still influences the new one. However, the physics of high Reynolds number turbulence flows can be exploited to neglect the not necessary terms in the system 1.3 and make the time integration of the ODEs system faster keeping a good accuracy.

1.3 Scales interaction in turbulence

Reduced order models allow to take into consideration less degrees of freedom respect to the original system. However, the intrinsic multiscale nature of turbulence and the mathematical formulation of the Navier-Stokes equations generate fully connected ROMs with a dense nonlinear interaction tensor Q_{ijk} (*i.e.* with all entries different from zero). The numerically evaluation of this tensor becomes computationally expensive.

As explained in section 1.1, high Reynolds number flows are an inherently multi-scale phenomenon. This means that the dynamics is composed by a wide range of spatio-temporal structures with the separation between their spatial scales that grows proportionally to the Reynolds number. The established picture of the physics of turbulent flows states that the dynamics of a certain flow structure is mainly influenced by structures of a similar length and not much by scales considerably smaller or bigger.

This means that it is reasonable to suppose that the most significant energy transfer will occur between the structures of commensurable length scales, while the interactions between the ones of considerably different dimension will not contribute significantly in the dynamics of the structure itself. Hence, energy transfers are sparse and not all the scales contribute equally in defining the dynamics of a certain scale. This process is qualitatively sketched in figure 1.2, where the red arrows represent the strong interactions between structures of similar length while the black arrows show that negligible interactions occur between structures of very different sizes. This process can be mathematically expressed stating that the dynamics of a certain mode is mainly influenced by the mode similar to the mode itself.



$$\frac{\partial}{\partial t} (\text{Scale}) = f(\text{Scale}) + \dots$$

FIGURE 1.2: Schematic representation of the locality of the scale interactions. The red arrows schematically show that only the interactions between modes of commensurable size are important for the dynamics of a certain mode.

The ROM generated from the Galerkin projection of the velocity field onto the POD modes does not exploit this sparsity structure but considers all possible interactions. Hence, the key to generate computationally cheap reduce order models lies in exploiting this sparsity structure of the energy transfer in the flow with at high Reynolds numbers.

1.4 Sparse regression

The objective of the classical regression algorithms is to find a set of coefficients that minimize a certain objective function (usually sum of square error called LSTSQ). In case of systems with a large number of degree of freedom the LSTSQ approach results in solution where all the regression coefficients are different from zero [Tibshirani and Friedman](#). Usually a sparsity promoting regression algorithm is constructed adding to the cost function a penalization term usually in the form of a l_1 norm (sum of the absolute values) of the regression vector, obtaining a regression problem in the form:

$$\tilde{\beta} = \operatorname{argmin}_{\beta} \frac{1}{2n} \{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \}, \quad (1.4)$$

Equation 3.5 shows the two terms that compose the penalty function in the case of Least Absolute Shrinkage Selection operator (LASSO) regression. The first term right hand side is the classical least square objective function. The second term is the penalty function expressed as the sum of the absolute value of the regression coefficients (entries of the vector β).

The role the penalization term can be twofold. Firstly it can work as regularization term in case of ill-posed problem. Secondly it allows solutions where the non relevant regression coefficients are shrunk to zero, providing a sparse solution vector. Figure 1.3 represents the desired result obtained by the sparse regression algorithm i.e. only three terms of the vector β are different from zero. This means that the target function y can be described by only three features (columns) of the matrix X .

$$y = X \times \beta$$

The diagram illustrates the matrix multiplication $y = X \times \beta$. On the left, the vector y is shown as a vertical stack of three colored squares: pink, cyan, and yellow. In the center, the matrix X is depicted as a 3x12 grid of colored squares. To the right of X , a multiplication sign (\times) is followed by a vertical vector β consisting of six colored squares: magenta, white, purple, white, green, and white.

FIGURE 1.3: Qualitatively representation of a sparse system. Only three coefficients are needed to represent the dataset.

A sparse solution makes the resulting model simpler and improves the physical interpretability of the solution itself. In the standard formulation of the *LASSO* regression [Tibshirani \(1996\)](#) the penalization term is usually multiplied by a scalar parameter that weights the importance of the penalization term respect to the least square term.

The structure of the objective function leads to one-parameters family of solutions. When the regularization parameter is large the solution is composed by only few terms it is very parsimonious but, on the other hand, it likely to not be accurate. When the regularization parameter is low the solution is composed by a lot of terms it is likely to be accurate but the interpretability of the solution could be much more difficult.

Hence it is possible to find a certain value of the sparsification parameter that locates the optimum trade off between the accuracy and the computational cost of the solution. In the case of a turbulent flow the location of this optimum point is expected to be a function of the Reynolds number. In particular the separation between scales should allow accurate solutions with all the relevant features condensed in few terms.

1.5 Sparse regression of dynamical systems

The objective of this work is to develop a rigorous approach to select the most important interactions that describe the dynamics of a turbulent flow. The approach proposed is to apply a well established regression technique to the system of ODEs arising from the Galerkin projection of the Navier-Stokes equations into a POD subspace.

The mathematical formulation of the POD reduced order model previously introduced is particular suitable to perform a sparse regression. The resulting Galerkin model 2.8 is a dynamical system that exhibits at most quadratic nonlinearities.

Hence it is possible to express the rate of change of each modal amplitude (left and side of the equation (1.3)) as weighed sum of monomials of the modal amplitude up to a combinations of the second order:

$$\dot{a}_i = \Theta(a)\beta_i \quad (1.5)$$

Where every row of the matrix Θ represent the state of the system at a certain temporal snapshot and is built by the following combination of the monomial a_i

$$\Theta(a) = \begin{bmatrix} 1 & \dots & a_1^0 & \dots & a_m^0 & \dots & a_1^0 a_1^0 & \dots & a_n^2 a_n^2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 1 & \dots & a_1^n & \dots & a_m^n & \dots & a_1^n a_1^n & \dots & a_n^n a_n^n \end{bmatrix}$$

The entries of the matrix β are the weight of the coefficients of the ROM. Theoretically this matrix is fully populated, the objective of the sparsification is to identify the coefficients that actively contribute to the time rate of change of the modal amplitude in question, the others will be shrunk to zero by the l_1 term.

In the previous subsection we explained how the formulation of the *LASSO* [Tibshirani \(1996\)](#) regression introduces an extra parameter that allow to tune the sparsity of the resulting solution. In the frameworks of the ROM sparsification technique the sparsity parameter determines the sparsity of the resulting model. Hopefully the underlying physics of a turbulent flow originates a hierarchy of models as sketched in figure 1.4.

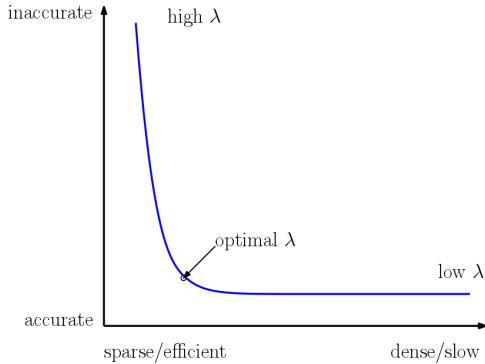


FIGURE 1.4: Expected behavior of the reconstruction error arising from a series of regression in the form 3.5. The parameter of the curve is the value of the regularization weight λ

The left part of the graph 1.4 represents a class of sparse and computationally efficient models, unfortunately this comes at the cost of obtaining a not accurate ROM. On the other hand in the right part of the figure can be located the class of fully-connected models where only few coefficients are zero, this models are accurate but not computationally efficient. The key observation is the existence of a knee point ,hopefully located in the left part of the graph, that identifies the optimal sparse ROM that includes only the really significant interactions. One of the objectives of the following work is to understand how the location of the knee point depends on the features of the flow, in particular we expect to observe a sharper knee point at a decreasing sparsity ratio when the Reynolds number increase, i.e. the separation between large and small scale in the flow increases.

1.6 Work programme

The structure of the works is composed by analysis of three different test cases of increasing computational and physical complexity:

- Kuramoto Sivashinsky equation [Hyman and Nicolaenko \(1986\)](#): This model is composed by a one dimensional PDE. The model was originally developed to study the velocity of propagation of a laminar flame in Bunsen burner. However, over the years this model has gained much more interest because of its mathematical behavior. In fact, this is one of the simplest PDES that exhibits chaotic solutions. This is due to the mathematical structure of the equation. Thus a chaotic solution of the KS equation exhibits a very similar qualitative behaviour to a turbulent solution of a Navier-Stokes equation with the same separation of scales arising from the increasing of the length of the domain. Moreover because of the extremely simple topology of the domain, it is possible to solve numerically the model through a spectral method obtaining an ODEs system where the unknown are the temporal amplitude of the Fourier modes. This feature allows us to study in the most appropriate an elegant way the interactions between scales of different wavelength.
- Cavity Flow: The second test case will be the two dimensional flow into a closed cavity with a moving lid. This system exhibits increasingly complex flow structures with the Reynolds number. However, the two dimensional nature of the flow makes it computational affordable even to run multiple test cases. Moreover, due to the boundary condition the construction of the ROM model of the flow is particularly easy. This makes this problem the perfect test case to build the POD framework that will be later applied to a more demanding test case.
- Three dimensional backward step: The final configuration we will consider will be a fully three dimensional separated turbulent flow over a step. This test case present all the complexities that characterize a fully three dimensional turbulent flow.

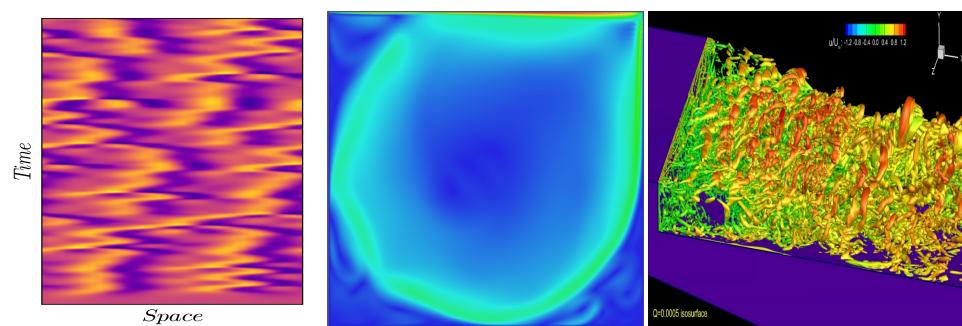


FIGURE 1.5: Example of the test cases that will be taken into consideration. From left to right: 1D solution of Kuramoto-Sivashinsky flow, 2D flow in a lid driven cavity at $Re = 10^4$ and fully turbulent flow over a step

Chapter 2

Literature Review

In this chapter all the relevant works that can be the ground for the present work will be listed. Firstly, we will focus on the construction of POD based reduced order model and the technical problems arising from this kind of procedure. Subsequently the energy transfer between modes in turbulent flows will be analysed. In particular we will report the results obtained in the framework of Fourier analysis for isotropic turbulence and in the framework of POD for a more geometrically complex configuration. The last part of the chapter takes into considerations the most well established techniques, developed in the field of the statistical learning, to extract the most relevant information from complex datasets. Finally a paragraph will be dedicated to how this techniques can be extended to the framework of the dynamical systems and in particular to the ROM Galerkin models.

2.1 Reduced order modelling

Reduced order modelling techniques have gained significance importance both in industrial and academic environments. One of the most appealing techniques is the POD Galerkin model. This technique consist to project the Navier-Stokes equations into the subspace obtained by a POD decomposition of the flow. All the technical details regarding the mathematical formulation are provided in [Volkwein \(2011\)](#), here we will report only the basic idea of the procedure.

2.2 Proper orthogonal decomposition

The proper orthogonal decomposition (POD) is a method originated from the statistical analysis of data. In particular, considering a scenario where the same phenomenon is measured n times and every measurement is composed by a vector y_i of a certain number of entries. The objective is to discover interdependency within the data and then reduce the data to a smaller number of parameters. Mathematically this problem can be written as the maximization of :

$$\underset{\phi}{\text{maximize}} \quad \sum_{i=1}^n |(y_i, \phi)|^2 \quad \|\phi\|^2 = 1 \quad (2.1)$$

this physically means to maximize the projections of the objective function ϕ over the dataset y_i . In the case of discrete sampling the measurement can be arranged in the snapshots matrix. It is constructed appending the snapshots of the system at every time step column by column:

$$Y = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ y_0 & y_1 & \vdots & y_n \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

where y_i is the state of the system at a certain time. This is equivalent to solve the eigenvalue problem:

$$\mathbf{R}\phi = \sigma\phi, \quad (2.2)$$

where \mathbf{R} is the correlation matrix defined as $\mathbf{R} = YY^T$ with Y the snapshots matrix. The results of the whole procedure is a new set of functions ϕ and their eigenvalues σ ranked according their energy content. Then it is possible to decompose the flow as following:

$$\mathbf{u}(\mathbf{x}, t) := \mathbf{U}(\mathbf{x}) + \sum_{i=1}^N a_i(t)\phi_i(\mathbf{x}) \quad i = 1, \dots, N. \quad (2.3)$$

These basis functions posses a lot of interesting properties that make them suitable for the Galerkin procedure. They are orthonormal and they are ranked according to the energy contained in every eigenvector ϕ_i . The energy optimality of the POD eigenmodes makes possible to describe a great percentage of the kinetic energy of the flow with just few basis functions. This is sketched in figure 2.1 where the spectrum of a POD analysis of a flow inside a cavity is plotted.

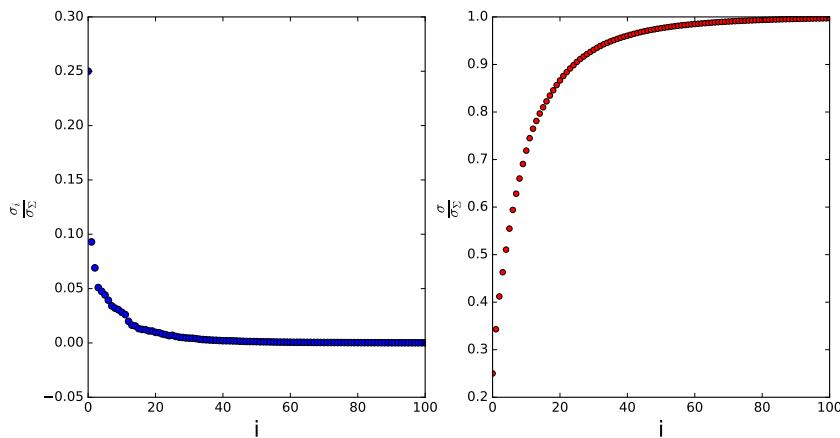


FIGURE 2.1: Normalized spectrum left and normalized cumulative spectrum right for the 2D cavity flow at $Re = 10^4$

The left figure shows how the most part of the energy is concentrated in the first few modes of the spectrum. The right figure instead shows how with just 40 modes it is already possible to capture more than 90 percent of the total energy of the system.

2.3 Construction of the reduced order model

Once the POD basis has been generated it is possible to construct the corresponding reduced order model (see for [Noack et al. \(2011\)](#) [Noack et al. \(2005\)](#) for technical details). The starting point is defining an inner product in the space of square integrable vector fields on the domain V :

$$(\mathbf{u}, \mathbf{v})_V := \int_V dV \mathbf{u} \cdot \mathbf{v}, \quad (2.4)$$

and the corresponding vector norm:

$$\|\mathbf{u}\| = \sqrt{(\mathbf{u}, \mathbf{u})}. \quad (2.5)$$

The residual of the Navier-Stokes equation is then defined:

$$\mathbf{R}(\mathbf{u}) = \partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p - \frac{1}{Re} \nabla^2 \mathbf{u} \quad (2.6)$$

and the projected onto the new basis function defined in 2.3 through the inner product 2.4:

$$(\mathbf{R}(\mathbf{u}), \phi)_V = 0 \quad (2.7)$$

The result is a system of coupled ODEs in the following form:

$$\dot{a}_i = C_i + \frac{1}{Re} L_{ij} a_j + Q_{ijk} a_j a_k + (-\nabla p, \phi) \quad i = 1, \dots, N \quad (2.8)$$

where C_i arises from the interaction with the mean flow. $L_{ij} = (\phi_i, \nabla^2 \phi_j)$ arises from the dissipative viscous term and $Q_{ijk} = (\phi_i, \phi_j \cdot \nabla \phi_k)_V$ represents the non linear interactions. The last term $(-\nabla p, \phi)$ is the result of the pressure term and needs to be modeled.

2.3.1 Pressure term in ROMs models

A compelling problem is related to the modelling of the pressure term arising from the Galerkin projection of the pressure filed onto the POD basis; $(\phi, \nabla p)$. This term cannot be simply expressed as a combination of the of POD modes so it remains unknown. In literature several approaches have been presented. The one proposed by [Galletti et al. \(2004\)](#) where the pressure term is modeled as a linear combination of the POD modes : $(\mathbf{u}, \nabla p) = C_{ij} a_j(t)$, the coefficients of the matrix C_{ij} are found through an

optimization procedure that aims to match the predictions of the reduced order model with the dataset from which has been generated.

A different approach is proposed by [Bergmann et al. \(2009\)](#). They try to take into account of the pressure term formulating a pressure extend ROM [Bergmann et al. \(2008\)](#). In this new formulation a new set of basis functions for the pressure field are introduced such that the pressure term can be directly evaluated through the pressure mode.

Finally, a deep analysis about the problematics involved by the pressure term and its modelling is provided by [Noack et al. \(2005\)](#). In particular the authors investigate the role of the pressure term in a shear layer configuration. They show how neglecting the contribution of the pressure term can lead to a non negligible amplitude errors even for a very low dimensional system. Most importantly, these errors cannot be compensated increasing the number of modes taken into consideration.

2.3.2 Dissipation in Reduced Order Models

One of the most important features of the *POD – ROMs* is that, due to the energetic optimality of the POD it is possible to capture a very high percentage of the kinetic energy of the flow with only few modes. This means that in the reduced order model only the bigger and most energetic structures will be taken into consideration. However, viscous dissipation is necessary to provide the right physical description of the flow. The viscous dissipation takes place in the smaller and unresolved scales of the flow. This implies that a ROM built up with only few modes will not be able to dissipate the correct amount of kinetic energy. This extra energy present in the flow could lead both to inaccurate results or, in numerical instabilities.

Several efforts have been done during the years to workaround this issue. [Bergmann et al. \(2009\)](#) proposed two different approaches. In the first, called Residual Based Stabilization, the goal is to approximate the fine scales of the flow with some adapted basis functions. The second method is to update the POD database on the fly during the simulation of the ROM itself.

A more classical approach based on the concept of eddy viscosity has been proposed by [Östh et al. \(2014\)](#). The authors present different formulations

of increasing complexity to describe the unresolved scales in the flow. Then they tested these model around a Ahmed body configuration.

Finally, a conceptually different solution has been proposed by [Balajewicz et al. \(2013\)](#). In this work, the authors reproduce the right amount of dissipation in the model adding an additional constraint to the original proper orthogonal decomposition algorithm. The main idea of this procedure is to rotate the original set of basis functions into a new, more dissipative configuration. The authors then have tested this approach over a high Reynolds number cavity flow proving its effectiveness. This approach is iterating because it is a very natural way to impose any kind of constraint to the POD basis that is generated.

2.3.3 Applications of reduced order modelling

Reduced order modelling is a powerful technique that allows the generation of computationally cheap representations of large multiscale systems resulting from the discretization of the PDEs. Hence, the affordable computational cost of the ROMs makes them an essential component to perform real time simulations and control of complex PDEs systems. In particular, reduced order model techniques have been proposed by [Ravindran \(2000\)](#) [Tonn et al. \(2010\)](#) in the field of optimal control and optimal shape design [Bui-Thanh et al. \(2008\)](#) of flow devices. An interesting application in the field of flow structure interaction has been proposed by [Lassila et al. \(2012\)](#). In this case they exploit the computational affordability of the ROM models to compute iteratively a large number of fluid solutions and reach the convergence of the fluid-structure interactions. Because the ability of the ROM models to condensate the most important features of the flow fields in few degrees of freedom, they have been used to perform also more fundamental studies about the fundamental physics involved. [Alizard and Robinet \(2011\)](#) have successfully used the ROM approach to perform the stability analysis of a flat plate boundary layer. [Kitsios et al. \(2011\)](#) uses reduced order modelling approach to study the stability of the flow structure generated by the leading edge separation in NACA 0015 profile. Finally it is worth to mention that the reduced order model techniques have been applied by [Nguyen et al. \(2014\)](#) to the framework of the reacting flow models.

2.4 Energy flow analysis in POD Galerkin representation of the flow

Energy transfer in turbulent flows has always been a topic of primary importance to understand the physics underlying the process. Historically, the energy flow analysis has been performed using Fourier decomposition, it provides a very clear and elegant description on the physical scales involved in the interaction. The major downside is that Fourier analysis can be used only for homogeneous flows, limiting its application to quite simple situations. The POD decomposition overcomes this limitation and try to extend the energy flow analysis to more complex flow configurations. The general and mathematical framework is developed in the work by [Noack et al. \(2005\)](#). The approach is the following: once the velocity field is decomposed in a mean and fluctuating part $\mathbf{u}(x, t) = \mathbf{U}(x) + \tilde{\mathbf{u}}(x, t)$, the total kinetic energy associated to the fluctuations per mass unit is defined as :

$$K(t) = \frac{1}{2} \|\tilde{\mathbf{u}}\|^2, \quad (2.9)$$

or equivalently as a function of the POD decomposition: $K(t) = \sum_i K_i(t)$ where $K_i(t) = \frac{1}{2} a_i(t)^2$. Subsequently an equation formally similar to [2.8](#) describing the evolution of K_i :

$$\dot{K}_i = a_i \dot{a}_i = C_i a_i + L_{ij} a_j a_i + Q_{ijk} a_j a_k a_i \quad (2.10)$$

A work concerning the application of this procedure has been done by [Couplet et al. \(2003\)](#). These authors perform a quantitative accurate analysis of the energy transfer between POD modes obtained decomposing a fully turbulent flow past a backward facing step. In order to understand the relative importance of the interaction between structures they define the influence of the j_{th} mode on the energy of the i_{th} as :

$$\Pi(i|j) = \sum_{k=1}^j Q_{ijk} a_k a_j a_i \quad (2.11)$$

The absolute value of this function is plotted in 2.2 in logarithmic scale.

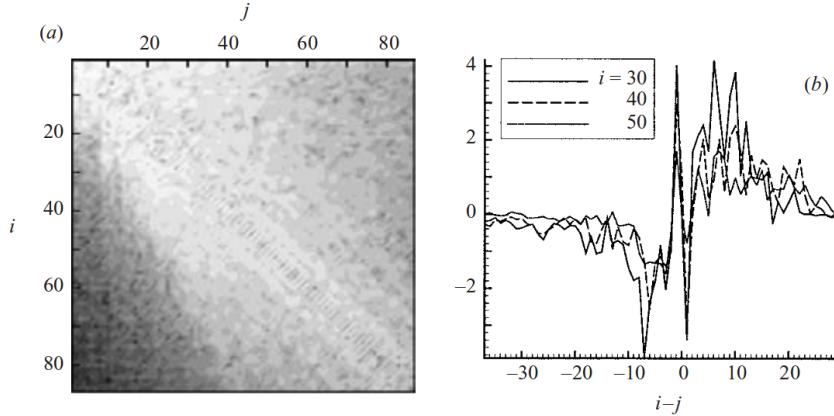


FIGURE 2.2: a) Map of $\log(\Pi(i|j))$ white maximum and black minimum.
b) $\Pi(i|j)$ as a function of $i-j$ for three different i . Figure from [Couplet et al. \(2003\)](#)

The major finding of this works can be list in the following points:

- From a theoretical point of view: as opposed to the Fourier representation of the flow all the triadic interaction in $Q_{ijk}a_ja_k$ (the matrix represented in figure 2.2 is fully populated) contribute to energy transfer in the flow, while the mean energy transfer between couple of modes is zero because of the orthogonality condition between modes.
- The observed that the energy transfer between the POD modes is local, in particular the mean energy transfer between two modes a_i and a_j is negligible when $|i - j| > 25$ as shown in figure 2.2 b) . This is a confirmation that the energy transfer between modes is local as already demonstrated for the Fourier modes, however it was not clear a priori in the case of a flow field decomposed in POD modes.
- The energy flow exhibits qualitative similar features as the one observed for the Fourier modes. The mean energy flow is directed from the big to the small scales (forward energy transfer) with only some spots of backward transfer.

The physical explanation of this behaviour can be qualitatively explained looking to the flow structures associated with the POD modes plotted in

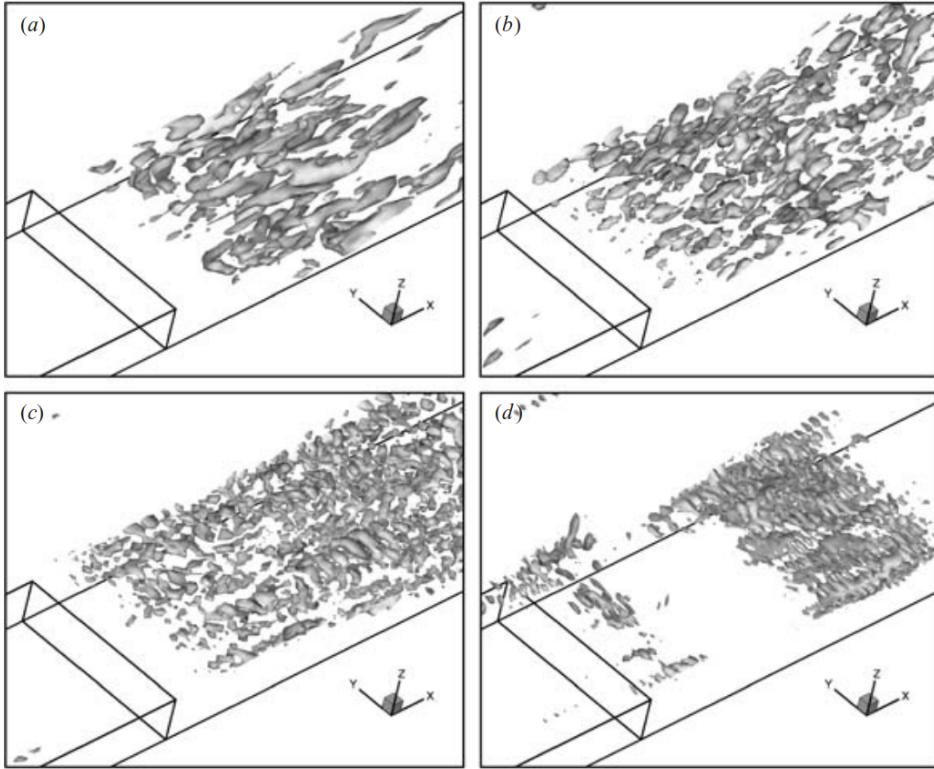


FIGURE 2.3: iso contour of the Q criterion for some POD modes: a) ϕ_1
b) ϕ_{20} c) ϕ_{42} d) ϕ_{87} . Figure from [Couplet et al. \(2003\)](#)

figure 2.3. Since the POD modes are sorted by decreasing order of kinetic energy the higher is the index the smaller is the flow structure associated with the mode itself. In other words the POD modes converge to the Fourier mode for high indexes. The Kolmogorov's local isotropy hypothesis is still valid for the less energetic POD modes. This allow to extend the main conclusions from the Fourier analysis to the POD framework.

2.5 Scales interaction and energy transfer in turbulence

In the previous section some works regarding the energy analysis in the POD framework have been presented. However, the scale interaction in homogeneous turbulence is a very deeply studied topic. In particular a rich

literature is available for test case like homogeneous turbulence in a periodic box.

It is a common belief that energy transfers, originated by the non linear term in the Navier- Stokes equations, is the principal physical process that affects the evolution of a turbulent flow field. Although, the increasing possibility to perform more and more detailed simulations the elementary process involved in the energy transfer are still not completely understood. These elementary processes are mainly the energetic interactions between three different wavenumber in the Fourier representation of the Navier-Stokes equations, called triadic interactions. One of the most debated subject is the relative contribution of the local and non local energy transfer across the spectrum to the temporal evolution of a certain mode. One of the main difficulties is that there is still not a clear definition of what local and nonlocal means. It is accepted that the energy transfers occurring between scales that are roughly smaller than two times and bigger of the half of the considered scale can be considered local, while all the interaction with the bigger or smaller scales are nonlocal.

In [Tennekes and Lumley \(1972\)](#) the authors underlines how the inertial subrange in the Kolmogorov spectrum is a direct consequences of the local interactions between triads. In [Brasseur and Wei \(1994\)](#), relying on results coming from numerical simulations, noticed that there is not appreciable energy transfer between wavenumber widely separated in scales. Similar conclusions have been drawn by [Domaradzki and Adams \(2002\)](#) that showed that the contribution to the energy budget of a certain scale comes mainly from scales of comparable size and therefore can be depicted as local. All cited works have been performed for fairly low Reynolds number and can't be rigorously extended to high Reynolds number flows. However some works such as [Itsweire et al. \(1986\)](#) show that the same mechanism holds, at least from a qualitative point of view, for high Reynolds numbers as well.

The authors previously mentioned ([Tennekes and Lumley \(1972\)](#),[Brasseur and Wei \(1994\)](#)) suggest how in this context is very important to distinguish between two different concepts:

- Energy transfer: it involves two different scales of motion. It can be local if the two scales are similar or if nonlocal

- Triad interaction: it represents the interactions between three different scales of motion. It is local if all three scales are similar and nonlocal if one of the three scales is bigger or smaller respect to the other two.

In the context of the triad interactions the concept of local or nonlocal energy transfer is complex. In fact, if local triad interaction always implies local energy transfer the opposite is not always true. It is possible to have local energy transfer caused by nonlocal triads if the energy is exchanged between two large wavenumber (long legs) in the triad without affecting the small wavenumber.

An interesting result was found by [Domaradzki and Rogallo \(1990\)](#) analyzing the transfer term involved in the energy equation. They found out that the energy transfer in homogeneous, isotropic turbulence is mainly local occurring between similar scales of motion. However, the triads interactions responsible for this process could be non local occurring between separated scales of motions. These results have been obtained with low Reynolds number turbulence but are likely to be extended to high Reynolds number as well. However the local energy transfer seems to dominate the dynamics of turbulent flows. The conclusions of [Domaradzki and Rogallo \(1990\)](#) are that although they are unable to give satisfactory description of the role played by the bigger scales in the energy transfer the local energy transfer plays a predominant role in the energy transfer in particular in the vicinity of the spectral cut off. The major takeaway of these works is that if this dynamics could be proved true it could indicate the possibility to simulate the biggest scales independently or more likely simulate the dynamics of certain scales taking into account only the scales of commensurable size.

2.6 Sparse identification of dynamical systems

Sparsity promoting techniques along with machine learning have proven to be a very powerful tools in many field of the engineering and science such as image recognition and signal analysis.

Only recently engineers have understood how these concepts can be extended to a more physical framework, in particular for the identification of dynamical systems.

As explained in section 2.2 the Galerkin Reduced Order Model of a flow is mathematically equivalent to a dynamical system. Then the sparsification techniques can be extended to this kind of system in order to either improve the performances of the reduced order model itself and to test the true capabilities of this techniques from a more physical point of view. The basic steps of this procedure are explained in figure 2.4. Firstly the data are collected from measurement or numerical simulation. Then the dataset is cast in a form suitable for performing the regression. Finally the regression reconstructs the model that better describes the temporal evolution contained in the dataset.

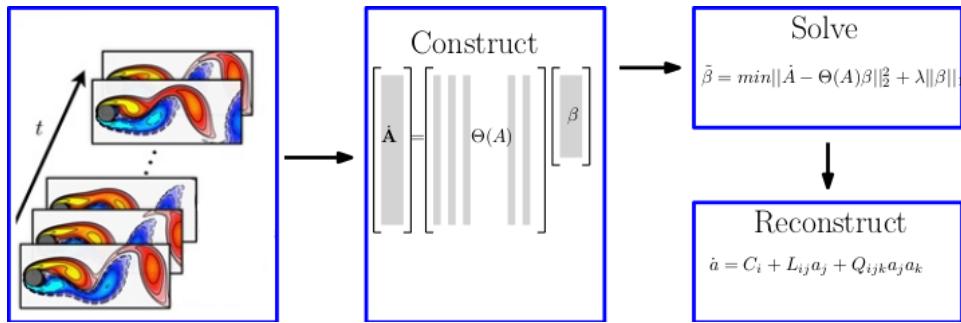


FIGURE 2.4: Dynamical system reconstruction procedure: 1) The data are collected from simulation or experiments. 2) the matrix of the time derivative \hat{A} and the library $\Theta(A)$ are constructed. 3) the LASSO is performed. 4) the dynamical system is identified

Significant effort in this direction have been made by [Brunton et al. \(2015\)](#), [Brunton et al. \(2016\)](#). They developed the SINDy framework that exploits the sparsity promoting techniques such LASSO regression and subset selection to discover nonlinear governing equations from noisy measurements. The method has been successfully tested over different nonlinear chaotic systems such as the Lorenz attractor and the POD Galerkin reduced order model of a laminar flow around a cylinder.

The results reported in [Brunton et al. \(2015\)](#) show how the LASSO algorithm is able to identify the correct structure of the dynamical system such as limit cycles and attractors. These works are of key importance for the extension

of the sparsification algorithm to a more physical framework, showing that the sparsification algorithm are able to reconstruct and identify the correct physics of the system.

A further improvement of the SINDy algorithm has been done in [Loiseau and Brunton \(2018\)](#). In this work the authors noticed that the sparse regression technique is just a mathematical procedure performed on a certain dataset but the conservation of the physical properties of the system must be ensured.

The authors then present a way to impose that the energy is preserved by the nonlinear term in the ROM representation of the Navier-Stokes system formulating the LASSO problem as a minimization problem with additional constraints in the convex optimization framework cvxpy [Diamond and Boyd \(2016\)](#).

The energy conservation of the nonlinear term of the Navier-Stokes equations is a quite important issue in order to preserve the numerical stability of the ROM and to ensure that the physics of the original flow will not be modified once the sparsification is terminated. Both of these issues are considered by [Balajewicz et al. \(2013\)](#) where a mathematical reformulation of the POD is proposed in order to obtain a reduced order model with the right amount of dissipation and with the right energy transfer between the scales.

2.7 The problem of selecting the relevant features

Machine Learning techniques born in a more IT oriented environment recently have begun to move towards more physical and engineering oriented applications, such as fundamental physics and fluid dynamics. Nowadays Machine Learning aims to address larger and larger datasets so the ability to identify the most relevant information in a potentially overwhelming quantity of data has become an important feature to embed in the algorithms. An interesting overview over this problematic is done by [Blum and Langley \(1997\)](#). They explain both at a very conceptual and practical level the problem of the irrelevant features and the concept of relevance in the data mining framework. The bottom line is that at a practical level we seek

for an algorithm that 'scale well' with domain with many irrelevant features, in other words the number of training samples needed to reach a certain level of accuracy should grow slowly with the number of features present in the dataset.

A very good example of feature selection model is explained in [Roig et al. \(2009\)](#), in this work the authors present a novel point of view to create a correspondence between two different images. It is interesting to notice how the problem can be reformulated as a minimization problem of an energy function that represents the distance from a certain subspace. They also apply three different approaches to solve the minimization problem showing that the quadratic programming approach provides the best results.

The previous examples although very interesting have the common property to not have any clear structure in the original dataset. However the Galerkin model of a fluid flow has at most a quadratic dependency from the temporal amplitude of the modes. The particular structure of the data allows us to restrict the searching area implementing a linear regression on a library decided a priori (see [Tibshirani and Friedman](#).) The main reasons why this kind of methods are appealing are : they are fairly simple and are able to provide adequate accuracy and interpretability of the results, moreover sometimes for models where the number of features is bigger than the number of training samples (situation likely to occur in fluid dynamics applications) they are able to provide even better results than more complex algorithms. A good overview over the linear regression techniques to data mining is provided by [Roig et al. \(2009\)](#). In this work the authors first explain how a simple least square regression is not satisfying mainly because in case of system with a very large number of features the least square tends to produce results with non zeroed coefficients that makes the results difficult interpretation. A good alternative to pure least square is the ridge regression [Tibshirani and Friedman](#) where the coefficients are shrunk to zero by a quadratic penalization, although this kind of regression is more stable and provides statistically better results respect the pure least square it still produces dense results and then hardly interpretable. An other good alternative is the subset selection, this regression is obtained adding to the least square function an l_∞ penalization term. The major drawback of this technique is that a certain feature is kept or eliminated from the model,

then there is the possibility to obtain very different models with slightly changes in the training set.

The optimal trade off between the stability of the ridge regression and the selection of an optimal set of coefficient is represented by the LASSO regression. The LASSO does not focus on a certain subset of coefficients but defines a continuous shrinking operation that can produce coefficients that are exactly zero.

Moreover the authors [Roig et al. \(2009\)](#) provide the relative merits and the drawbacks of the previous techniques in the different scenarios: the results can be easily wrap up in the following table 2.1

	Few features with large effects	Medium features with Medium effects	Large features with small effects
Subset selection	Good	Medium	Bad
LASSO	Medium	Good	Good
RIDGE	Bad	Medium	Good

TABLE 2.1: Performance of the three different algorithms for datasets with different features

As explained in the introduction the penalization term is usually multiplied by a arbitrary scalar coefficient that weights the penalization term respect the least square part. This produce a family of solutions that depend from the choice of this parameter. The choice of the regularization parameter is usually left to the experience of the user. However in the years some techniques to find the optimal regularization parameter in a systematic way. The most common are the cross-validation and the K-fold validation [Golub et al. \(1979\)](#).

The main objective of the cross-validation technique in to try to create a model as much as possible independent from the training dataset. This is done splitting the dataset in N folds, then use some of them for training the algorithm and the other to test it. And then to iterate the process through all the folds [2.5](#).

The concept behind the cross-validation technique can be easily explained using figure [2.6](#) where is sketched the qualitative trend of the training error



FIGURE 2.5: Conceptual sketch of the dataset splitting in the cross-validation procedure.

and of the test error as a function of the complexity of the model that in case of the penalized regression is proportional to the inverse of the penalization parameter.

The left part of the graph represent a class of very simple models, the training error and the cross-validation error are both very high because the model is too simple to describe the dataset. On the other hand, the right part of the figure contains more complex models. In this scenario the training error is low because the model is able to reconstruct the training dataset very precisely. However this result is not satisfactory because the model is too particularized for the specific dataset on which has been trained. It doesn't perform well on a different test set. The main aims of the cross validation is to find the penalization parameters (or model complexity) that represent the best trade off between model accuracy and generalization, the vertical line in figure 2.6.

2.7.1 Solution of the LASSO problem

The solution of a constrained regression involves the solution of an optimization problem in the following form:

$$\tilde{\beta} = \operatorname{argmin}_{\beta} \frac{1}{2n} \{F(\beta) + \lambda G(\beta)\} \quad (2.12)$$

where F and G can be two arbitrary functions. Mathematically speaking this kind of optimization belongs to the class of convex problems, a lot of

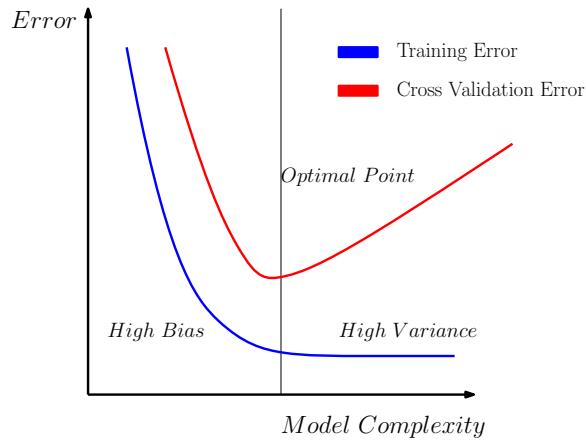


FIGURE 2.6: Qualitative trend of the training error and the cross-validation error as a function of the model complexity. The vertical line indicate the optimal value of model complexity and divide the zone of overfitting right part of the diagram and under fitting left part

literature and algorithms have been developed to solve this class of problems, such that the libraries developed by [Pedregosa et al. \(2011\)](#) and [Diamond and Boyd \(2016\)](#).

Although there are a lot of well established algorithms developed to specifically solve this kind of problems, solve a large scale regression problems remains a challenge. The choose of the best algorithm possible for the application taken into account is an important step of the process.

One of the most widely applied technique is the ADMM (Alternate Method of Multipliers) algorithm. It is basically a gradient descent where the objective function in equation (3.1) is split in two separate function of different variables.

$$\tilde{\alpha}, \tilde{\beta} = \underset{\beta, \alpha}{\operatorname{argmin}} \frac{1}{2n} \{ F(\beta) + \lambda G(\alpha) \} \quad (2.13)$$

In the case of the LASSO regression $F(\beta) = \|X\beta - y\|_2^2$ and $G(\alpha) = \|\alpha\|_1$. The equivalence between the two problems is recovered adding and additional constrain, $\alpha = \beta$. Although this could seem an over complication it actually enables a very fast and efficient implementation of the gradient descent algorithm and most importantly allows more degrees of freedom on the mathematical formulation of the problem extending the form of F, G

to non-smooth and non convex functions [Benning et al. \(2015\)](#), [Liu et al. \(2017\)](#). It is finally worth to mention that a lot of effort has been done in the field of signal processing and computer vision, where a new reformulation of the ADMM algorithm has been proposed by [Bibi et al. \(2017\)](#) with the aim to increase its scalability to very big datasets. These techniques although developed in different fields could be ported to fluid dynamics applications.

2.8 Conclusions

The major takeaways from the literature review are the followings:

- In the field of the reduced order modelling a lot of effort has been dedicated to construct more efficient technique and formalize the mathematical background. Moreover a lot of authors have analysed quite deeply the most common problems relative to the reduction of order such as: the problem of the dissipation and the closure model of the pressure term.
- There is a lot of detailed literature concerning the interactions between scales in turbulent flows. The vast majority of this work are carried out in the framework of Fourier analysis; luckily this aspect should not represent a too big limitation because as explained in Sec (2.2.3) the energy flux in the POD representation and in the Fourier have a qualitatively similar behaviour, this is particularly true for the inertial part of the spectrum.
- Only recently the concept of sparse regression has been extended to the framework of the dynamical systems. These techniques have been applied only on simple system but they showed to be able to reconstruct completely the system including the linear and the nonlinear part. Although the systems taken into consideration are quite simple the approach seems very promising.
- The problem of selecting the most relevant features is a well known problem in the scientific community. During the years a lot of effort has been dedicated to improve and formalize the sparse regression

techniques. So we can relies on a wide set of well known and tested regression techniques as well as validation techniques such as cross validation and K-fold validations. Moreover significant effort has been done to develop algorithms that provide good performance even on very large datasets.

The biggest open questions that this project aims to answer are:

- Although the mathematical framework and the numerical techniques related to the sparse regression are well known and validated they have only recently extended to the framework of the dynamical systems. Albeit the LASSO regression seems to perform very well in reconstructing the nonlinear features of systems like the Lorenz attractor and the flow past a cylinder it is still not clear how these kind of methods scale with the number of degrees of freedom of the system. Moreover it is not proved yet if the resulting sparsified reduced order model is somehow the results of the underlying physics of the interaction between scales in the flow or it the result of a mathematical procedure decoupled from the physics.
- If the algorithm is able to describe the right physics of the flow what is the right sparsification approach to choose and how to decide the value of the sparsity parameter λ in order to favour some aspects or others in the resulting reduced order model.
- It is not clear if the algorithm is coupled with the physics of the flow, in particular if there is any correlations between the terms eliminated by these regression methods and the value of the energy transfer related to these terms.
- Lastly the effect of the Reynolds number on the sparsity of the system. In particular how the separation of scales emerging increasing the Reynolds number affects the sparsity pattern of the solution and consequently the efficiency and the accuracy of the solution.

Chapter 3

Sparse Regression of Dynamical Systems

In this chapter will be briefly explained the mathematical formulation that allows us to apply regression on dynamical systems. Then the focus will be moved of the sparse regression techniques, the main target will be to analyze the different parts of the objective function that should lead to a sparse solution and try to explain in particular why the l_1 norm tends to set the least important coefficients exactly to zero. Subsequently we will briefly discuss the different sparsification techniques that can be obtained with different choices of the regularization parameter λ . Finally a quick example on the Lorenz system will be presented.

3.1 Sparse regression

In the section 2.7 we stressed out the property of the LASSO regression to perform well on dataset of medium/big size with features with medium/low importance and to provide a stable and reliable solution thanks to the property of continuously shrinking to zero the coefficients of the regression. It is obtained adding to the least square objective function a penalization term of the regression coefficients in the form of a l_1 norm. Given a objective vector y a database matrix X and a dataset vector β , the standard lagrangian formulation of the LASSO is:

$$\tilde{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2n} \{ \|X\beta - y\|_2^2 + \lambda \|\beta\|_1 \} \quad (3.1)$$

Figure 3.1 shows qualitatively the effect of two different kinds of norms for $\beta \in \mathbb{R}^2$. The ellipses show the location of the least square regression that is obviously not sparse (none of the two degrees of freedom is zero). When a penalization term is added the location of the minimum can be graphically located where the isolines of the least square and of the penalization are tangent. On the left we can see the effect of a l_2 penalization term that tends to shrink the value of the coefficients but has no particular tendency to push them, or at least one of them, towards zero.

Conversely the particular shape of the l_1 norm's isolines increase the likelihood that the minimum can be located exactly over one axis (and so the correspondent coefficient set exactly to zero) as depicted in the right part of Fig.3.1.

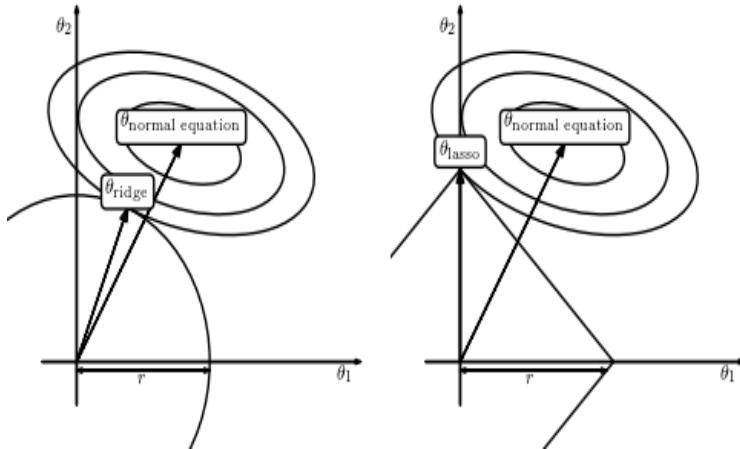


FIGURE 3.1: Qualitative effect of two different kind of norms on the location of the minimum in the parameter space θ_1 - θ_2

As mentioned in the introduction the extra degrees of freedom added by the presence of λ can be used to tune the sparsity of the model and create a family of different model for the same dataset. To understand how the process of shrinkage takes place could be useful to analyze a very simple two dimensional system and see how different values of the regularization weight affects the location of the optimal solution.

Increasing the value of λ in 3.1 the effect is that the relative weight of the l_1 norm in the objective function became more and more important and the coefficients will be progressively shrunk to zero. This process is shown in the following figure where is depicted the value of the least square part (continuous line) the l_1 part (dashed line) and the identified minimum of the least square (green square) and of the lasso penalty function (red triangle)

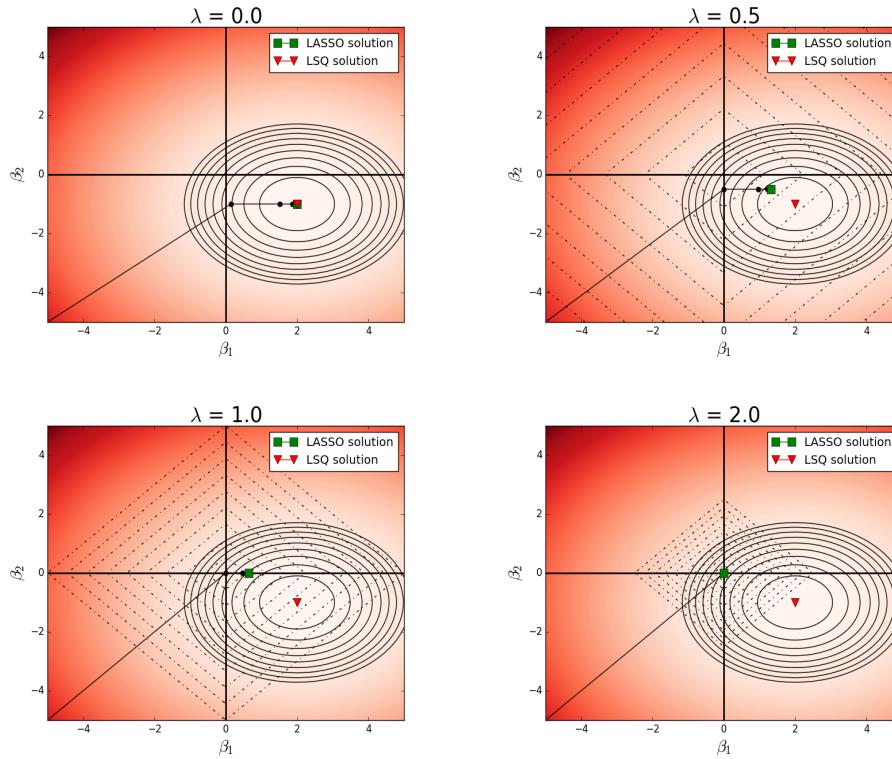


FIGURE 3.2: Qualitative behavior of the location of the minimum of 3.1
as the sparsity parameter increases

It is interesting to observe the effect of the sparsity parameter on the location of the minimum. In particular:

- $\lambda = 0$ there is no difference between the LASSO an the ordinary least square and the location is the same.
- $\lambda = 0.5$ there is a small effect of the l_1 and the value of β_2 is pushed towards the β_1 axis.
- $\lambda = 1$ the effect of l_1 has became strong enough to shrink one parameter to zero, this is the process of feature selection

- $\lambda = 2$ for too high values of λ the effect of the l_1 part is too strong and all the coefficients are set to zero.

Although simple the previous example shows two important aspects. Firstly, the mechanism behind the sparsity promoting feature of the LASSO algorithm. Secondly, the importance of performing a parametric study on the effect of lambda on the sparsity features and on the predictive capabilities of the model in order to find the optimal trade off between good prediction but complex model (first picture) and bad prediction and simple model (last picture).

In this section we presented this simple example in order to introduce the key aspects of the LASSO regression. All the considerations we did are easily generalizable for higher dimensional systems.

3.2 Sparse regression of dynamical systems

A dynamical system can be formally defined as rule that describe the evolution of a point in a certain state space. It can be mathematically described as :

$$\dot{a}_i = C_i + L_{ij}a_j + Q_{ijk}a_ja_k, \quad (3.2)$$

Where \mathbf{f} can be an arbitrary complex function of the vector state \mathbf{x} . It is possible to recast a general dynamical system to a form that can be solved using a convex l_1 regression and extract the most relevant terms of the RHS that contribute to the dynamics represented by the LHS.

In this work we will follow the formalism introduced by [Loiseau and Brunton \(2018\)](#). The first is to collect a time series of data and its derivative and cast them into two different matrices A and \dot{A} . The lower index represents the temporal evolution of the system (it spans from 0 to n) while the upper index represents the number of measurements for a given snapshots (it ranges from 0 to m)

$$A = \begin{bmatrix} a_1^0 & \dots & a_n^0 \\ \vdots & & \vdots \\ a_1^m & \dots & a_n^m \end{bmatrix}$$

$$\dot{A} = \begin{bmatrix} \dot{a}_1^0 & \dots & \dot{a}_n^0 \\ \vdots & & \vdots \\ \dot{a}_1^m & \dots & \dot{a}_n^m \end{bmatrix}$$

Subsequently through a systematic procedure a matrix, called library matrix, of candidate functions is created from the matrix A . It can be as complex as the user wants, however for fluid dynamics applications due to the fact that it is known that the Navier-Stokes equations are only quadratic in the state vector the database will be generated consequently.

$$\Theta = \begin{bmatrix} 1 & \dots & a_1^0 & \dots & a_m^0 & \dots & a_1^0 a_1^0 & \dots & a_n^2 a_n^2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 1 & \dots & a_1^n & \dots & a_m^n & \dots & a_1^n a_1^n & \dots & a_n^n a_n^n \end{bmatrix}$$

Every row of Θ contains all the possible quadratic combination of the modes a_i (called features) present in the system 3.2 for a certain temporal snapshot.

Then the system 3.2 can be equivalently rewritten in a different formulation, with β the vector of the coefficients to be optimized to match the value of .

$$\dot{A}_i = \Theta(A)\beta_i \quad (3.3)$$

This procedure can be described as follows. The columns of \dot{A} represent the temporal dynamic of a mode i . Θ contains the temporal evolution of all the features (quadratic combination of the modes) present in the system. Finally, the matrix β contains the relative weights of every feature in the dynamic of the modes i .

Due to the structure of the problem represented in 3.3 it is possible to perform the sparse regression on every mode separately.

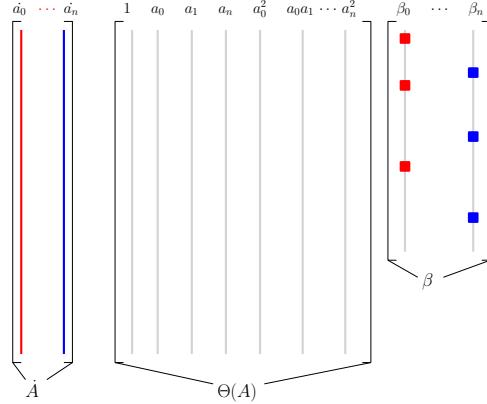


FIGURE 3.3: Sparsification procedure applied to a dynamical system. The blue and red lines represent the temporal evolution of the system. The red and blue dots in the matrix β represent the features (column of the matrix Θ) that contribute the most to the dynamic of the corresponding mode

$$\tilde{\beta}_i = \underset{\beta}{\operatorname{argmin}} \left\{ \|\Theta\beta_i - \dot{A}_i\|_2^2 + \lambda \|\beta_i\|_1 \right\} \quad (3.4)$$

$$\begin{bmatrix} \dot{a}_0 \\ \vdots \\ \dot{a}_n \end{bmatrix} = \begin{bmatrix} 1 & a_0 & a_1 & \dots & a_0^2 & \dots & a_n^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_n \end{bmatrix}$$

FIGURE 3.4: Feature selections applied on the dynamics of one mode

Figure 3.4 shows the sparsification procedure applied to a single mode. The left hand side contains the temporal evolution of the first mode. The database matrix Θ contains the temporal evolution of the features that contribute to the dynamic of the selected mode. The feature selection procedure

consists in the optimization of the coefficients of the vector β_0 according to the objective 3.4. The effect of the l_1 norm should be to shrink to zero the coefficients of β_0 relative to a feature not relevant for the dynamic of a_0 . In figure 3.4 is show how the final result is that the temporal dynamic of the derivative of a_0 can be describe with as $\dot{a} = C + \beta_0^1 a_1 + \beta_0^{n+1} a_0^2$.

3.3 Choice of the sparsification technique

The addition of the regularization weight in the regression problem introduces an extra degree of freedom. In this section we will explain how the choice of this parameter can be used to generate sparse model with different properties. In particular we will propose three different sparsification techniques. One based on the magnitude of the energetic interactions between modes. The others based on different choice of the value of the regularization weight .

3.3.1 Greedy approach

The first sparsification approach presented here will be the so called greedy sparsification approach. This approach lies on the simple idea to rank the interactions of every mode with the others in base of the value of its energetic interaction. Subsequently the mode with less energetic interaction will be simply neglected in the dynamic of a certain mode. This procedure does not involve any optimization procedure on the coefficients of the ODEs system.

3.3.2 Choice of the regularization parameter λ

The sparse regression procedure presented in the previous section boils down to solve an optimization problem in the form :

$$\tilde{\beta}_i = \underset{\beta}{\operatorname{argmin}} \frac{1}{2n} \left\{ \|\Theta\beta - \dot{A}\|_2^2 + \lambda \|\beta\|_1 \right\} \quad (3.5)$$

All the term of the expression except λ are either known or unknown to be computed, this gives us an extra degree of freedom in order to create a family of model as explained in 1.4.

However, the framework of the dynamical systems a regression in the form ?? is performed on each mode separately this allow us to choose a different sparsification parameter λ for each mode.

This is actually a very delicate point and involve the difference between truncating a system and sparsifying a system. The concept can be qualitatively expressed with the help of the following figure 3.5. To compare the graph to the scale interaction we can think that every black dots represent the dynamic of a certain mode, the a_i , while the lines represent the interaction with the other modes that contribute to the dynamic of the mode i , the term $Q_{ijk}a_ja_k$ in the reduced order model.

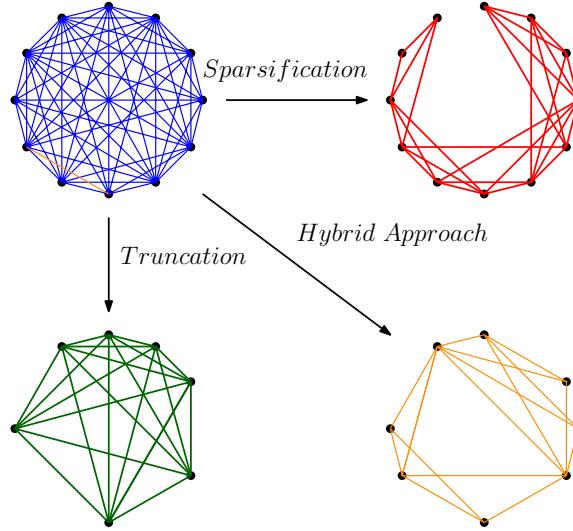


FIGURE 3.5: Example of two different approach on graph sparsification: full graph on the upper left, sparsified graph on the upper right and truncated graph on the lower left and hybrid approach in the lower right

With the term sparsification we intend that process that keeps all the modes in the system (black dots in upper right part) but neglects the non necessary interactions (blue lines) and keep only the important interaction (red thick lines). The objective in the process no information about any of the modes will be lost but just the one regarding the interactions. With pure truncation instead, we intend that process in which we choose to approximate

the system with a smaller number of modes, in other words we keep the full amount of interactions (green lines in the lower left) but we choose to completely eliminate one or more modes (black dots). The hybrid approach instead is a trade off between the two previously explained: it eliminates both modes and interaction from the system.

The role of λ in the problem 3.5 is simply to shrink to zero the coefficients that have the least contribution to the dynamic of a certain mode. However we don't know a priori which of the two approaches will be predominant if sparsification, truncation or an hybrid approach.

For these reasons in this work we will propose two different sparsification techniques obtaining changing the value of λ between the modes.

- The first is keeping the value of λ constant across the spectrum, we expect that this strategy will bring to a results in the middle between the two aforementioned approaches. So the final results will probably be the complete elimination of the least energetic modes and the sparsification of the most energetic ones.
- An other approach is somehow try to scale the value of λ mode by mode with the value of the mean energy of the following mode $\lambda \rightarrow < a_i a_i^* > \lambda$. With this approach we expect to avoid the complete elimination of any mode but instead the elimination of the not important interaction that contribute to the dynamic inside a certain mode.

Finally the two previous techniques will be evaluated between themselves and with a so called 'Greedy'. Then the terms corresponding the low energy contribution will be simply set to zero. We expect that this technique, although computational cheap since does not involve any optimization process, will provide a higher reconstruction error and then much worse performances.

3.4 Example: the Lorenz equations

In this section we will discuss, as an example, the application of the previously discussed method on the Lorenz equations, introduced in [Lorenz](#)

(1963). This example is particularly interesting because it is simple enough to provide a solution of simple interpretability. However, it already presents complex features, such as nonlinear dynamics and chaotic behavior, that are characteristic of turbulent flow models.

The Lorenz equations are:

$$\dot{a}_1 = \sigma(a_2 - a_1) \quad (3.6)$$

$$\dot{a}_2 = -a_1a_3 + ra_1 - a_2 \quad (3.7)$$

$$\dot{a}_3 = a_1a_2 - ba_3. \quad (3.8)$$

The database matrix Θ has been generated as explained in the previous section and the optimization problem in the form (3.4) has been solved using the convex optimization framework cvxpy developed by [Diamond and Boyd \(2016\)](#).

A series of sparse regressions has been performed with value of λ ranging from 10^{-8} to 10^{-1} . The trend of the coefficient of the system and of the reconstruction error defined as $\epsilon = \|\dot{A} - \dot{\tilde{A}}\|_2^2$. $A = [a_1 \ a_2 \ a_3]$ is the matrix of the original system and \tilde{X} is the one relative to the reconstructed dataset. The norm is calculated as the Frobenius norm.

It is possible to see from the right picture how increasing the sparsification parameter, *i.e.* moving toward the left side of the image, the non necessary coefficients, red dashed lines, are progressively shrunk to zero. It is also interesting to notice how the elimination of these coefficient from the model has a little or no effect on the error (left image). The error starts to grow only when the value of λ reach a value that starts to affect the physical solution as well. Finally we can observe that, for this particular system, a choice of $\lambda \sim 10^{-3}$ could be the optimal one in fact all the non necessary interaction has been eliminated from the model and the global reconstruction error is not affected.

In is interesting now to compute the solution of the reconstructed system of ODEs from a slightly different initial condition. The point of this is to see if the reconstructed system has the same physical and mathematical properties of the original model.

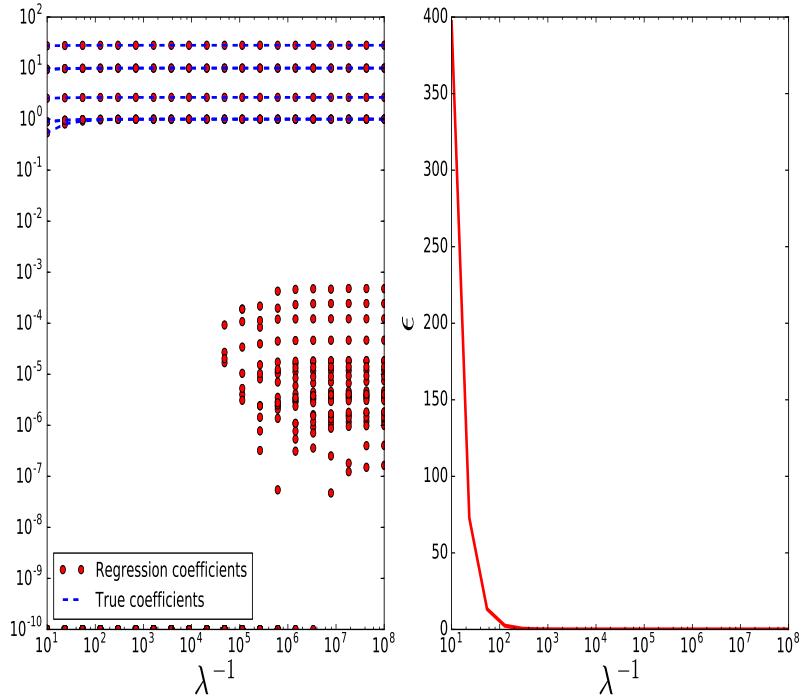


FIGURE 3.6: Left: Trend of the coefficients as a function of the inverse of the sparsification parameter. Right: trend of the error as a function of the inverse of the sparsification parameter

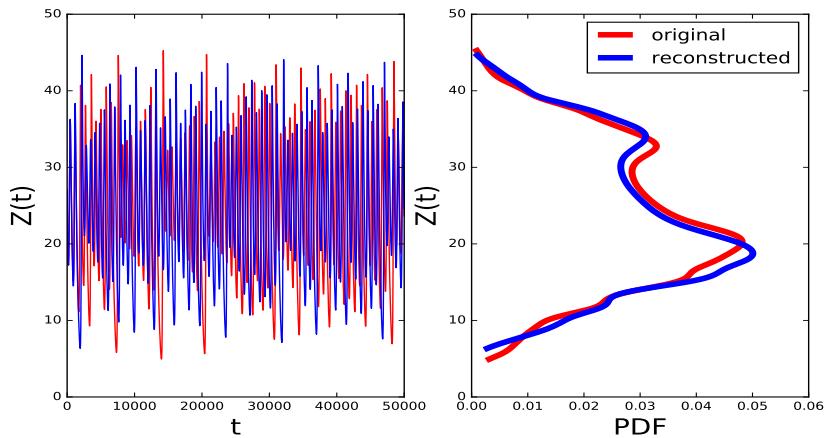


FIGURE 3.7: Temporal dynamics of the original solution and of the reconstructed one for a different initial condition

In figure 3.7 left a comparison between the evolution of the z coordinate of the original solution and the reconstructed one starting from different initial condition is reported. The main target is to check that the statistical properties of the system are conserved.

In particular in the right part of figure 3.7 is plotted the probability distribution function of the trajectory of the z coordinate. It is clear that the two trajectories although not the same share the same probability distribution. As final remark it is important to stress that all the observation done for the reconstructed system will be an important check for the more complex cases that will be analyzed later on in order to ensure that the right physics will be conserved in the sparsification process.

3.5 Sparse POD basis

In this section, we will present a novel approach that aims to generate a set of basis functions that naturally include the sparsity structure of the energetic interactions.

3.5.1 Mathematical formulation

In order to develop this new technique we follow the idea introduced by [Balajewicz et al. \(2013\)](#) and [Jovanović et al. \(2014\)](#). The problem that they face in these works is to find a set of POD modes that naturally incorporate the energy conservativeness of the nonlinear term of the Navier-Stokes equations.

We will try to formulate the problem in a similar fashion. Calling $\tilde{\phi}$ and \tilde{a} are the new sparse set of POD spatial and temporal modes to be computed. We can formulate a new optimization problem that incorporates the sparsity constraint in the following form.

$$\begin{cases} \underset{\tilde{u}, \tilde{a}}{\operatorname{argmin}} \left\{ \int_V |u(x, t) - u_0(x) - \sum_{i=1}^N \tilde{a}(t) \tilde{u}(t)|^2 dV \right\} \\ \int_V \tilde{u}_i \cdot \tilde{u}_j dV = \delta_{ij} \\ \|Q_{ijk}(\tilde{u})\|_1 < c, \end{cases}$$

where $u(x, t)$ and $u_0(x, t)$ are respectively the DNS solution and its mean. $Q_{ijk}(\tilde{u})$ is the nonlinear term of the Navier Stokes equations expressed as a function of the new set of basis functions. c is an extra parameter introduced in the problem that has the aims to wight the contribution of the l_1 term in the solution of the problem. The idea is to exploit the sparse properties of the l_1 norm to find a solution where the matrix associated with the nonlinear part of the Navier-Stokes has the minimum amount of not zero coefficients.

The link between the standard set of POD modes (u and a) and the new set of POD modes is the linear transformation represented by the matrix X that satisfy the following relations:

$$\begin{cases} \tilde{u}_i = X_{ij} u_j \\ \tilde{a}_i = X_{ij} a_j \end{cases}$$

The problem 3.5.1 can be then reformulated as a function of the unknown matrix X

$$\begin{cases} \underset{X}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \lambda_i - (X \Lambda X^T)_{ii} \right\} \\ X X^T = \delta_{ij} \\ \|Q_{ijk}(X)\|_1 < \epsilon, \end{cases}$$

It is worth to notice that the first two equations in previous problem represent the classical POD decomposition. The idea is that the introduction of the l_1 penalization terms induces a slightly rotation in order to bring the original system in a configuration where the sparsity structure of the interaction could emerge. The main problem of this kind of approach from the implementation point of view in fact the problem 3.5.1 has constraints that are non linear. For these reason the well established techniques of the convex optimization can not be used. Anyway a successful approach could be the one used by [Balajewicz et al. \(2013\)](#) that exploited the sequential quadratic programming technique (SQP) implemented in MATLAB in order to reduce the non convex problem to a series of locally convex problems. Similar results could be obtained using the python library pyOpt.

3.5.2 Standard POD vs sparse POD

The conceptual differences between the classical approach to generate sparse reduced order model and the one bases on the a priori generation of a POD basis that incorporate already sparse features is plotted in 3.8.

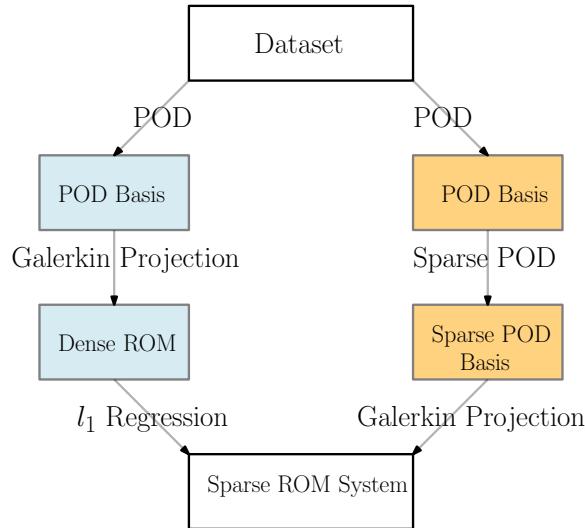


FIGURE 3.8: Conceptual differences between the two approaches. Left: classical approach, Right: sparse proper orthogonal decomposition approach

The main outlines of the classical approach are illustrated in figure 3.8 in the left path, and are the followings:

- Once the dataset is generated a standard proper orthogonal decomposition is performed and the spatial and temporal modes are obtained. These modes are defined globally over the whole domain.
- Then the Reduced Order Model is computed from the POD modes. As mentioned previously these modes are defined globally over the whole domain then there is no reason to believe that the corresponding reduced order model could incorporate any sparsity structure but will be likely composed by fully populated matrices.
- Once the ROM has been generated thanks to the technique introduced in chapter 2 the l_1 regression on dynamical systems is performed and

the sparsity structure of the ROM is extracted obtaining a computational efficient sparse ROM.

The classic procedure although with solid conceptual basis need the introduction of the step related to the l_1 regression. This step is introduced a posteriori to filter the irrelevant information from the original POD basis.

The sparse POD approach, illustrated in the right path of figure 3.8, aims to incorporate the effects of the l_1 regression generating a basis function for the POD system that already incorporate the sparsity structure of the flow. The main objective is then to develop a rigorous mathematical framework that enable the generation of sparse reduced order models directly from the level of the POD decomposition without the need of l_1 regression a posteriori. The step of this new procedure can be explained as follow:

- Generation of the energy optimal POD basis
- From the original POD basis is possible to generate a new set of functions that have the characteristic to minimize the number of active interaction in the nonlinear term of the Navier-Stokes equations. This is achieved through a slightly rotation of the original POD basis as formulated in equation 3.5.1.
- The Galerkin Projection of the Navier-Stokes performed of the new basis functions should generate reduced order model that naturally include the sparsity structure of the flow.

3.6 Conclusions

In this chapter we have introduced the general procedure to cast a dynamical system in a form that can be suitable to perform a regression. We discussed briefly three different types of sparsification approaches that can be adopted. As a demonstration we tested the algorithm on a simple problem namely the Lorenz attractor of which the final solution is known. The sparse regression performed well identifying both the correct terms and the correct numerical

value of them. Finally we have verified that the reconstructed system has the same statistical properties of the original one.

Although the procedure seems reliable it is worth pointing out that in this example the number of term are quite low, then the task of the algorithm was more identifying the underlying already sparse system. The target of this work is more to find a rigorous procedure to sparsify a dense system. To address this task a more complex system will be introduced in the next chapter.

In the final part we have introduce the mathematical framework to generate a new set of basis functions that naturally incorporate the sparsity of the interaction present in a fully turbulent flow. The implementation and the validation of this new procedure will be the focus of the following months of research.

Chapter 4

1-D Turbulence: Kuramoto-Sivashinsky Equation

In this chapter the Kuramoto-Sivashinsky model will be analysed. This model is one of the few 1D models to present chaotic behaviour and, more interestingly, it is possible to change the number of dynamical active modes in the system changing the length of the domain. Thus is possible to use this model to study the relative importance of the interactions for increasingly more complex systems.

4.1 The mathematical model

Originally the Kuramoto-Sivashinsky model arises from the stability analysis of a laminar flame front. However, during the years this model has gained great interest of mathematicians and physicist for its property of being one of the simplest one dimensional system to exhibit chaotic behaviour. The standard mathematical formulation of the problem describe the evolution of the velocity field $u = u(x, t)$ in case of unitary viscosity ($\nu = 1$) and on a periodic domain of length $L = 2\pi\tilde{L}$:

$$\begin{cases} u_t = -uu_x - u_{xx} - u_{xxxx} \\ u(t, 0) = u(t, L) \\ u(0, x) = u_0(x), \end{cases}$$

where the subscripts indicate the partial derivative. The interesting dynamics arises from the three terms at right hand side, namely:

- the second derivative with a 'negative' viscosity $-u_{xx}$ injects energy at small wavenumbers.
- the fourth derivative of the velocity field $-u_{xxxx}$ the hyper-viscosity dissipative term, dissipates the energy at the level of the smallest scales of motion
- the quadratic non linearity $-uu_x$, same as in the Navier-Stokes equations, transfers energy from the large structures to the small ones.

Then dynamics of the solution of the Kuramoto-Sivashinsky equation is very similar to the one of a turbulent solution of the Navier-Stokes equations with energy injection at the biggest scales of motion, energy cascade towards the smallest scales till the viscous dissipation.

Moreover, this model has the interesting characteristic that the range of dynamical active scales increases as the length of the domain increases. Hence, it is possible to reproduce the separation between scales observed in a real turbulent flow in this simple one dimensional system simply varying the length of the domain. This make this model the perfect test case to validate the algorithms that will be used later on in more complex configurations.

4.2 Numerical procedure

Due the particular topology of the domain it is possible to solve the equation in the Fourier space. This property allows us not only to use a very fast and accurate numerical technique but also to write the system in a coordinate basis that allows a very natural and elegant analysis of the interaction

between scales. Introducing $\tilde{L} = \frac{L}{2\pi}$ and $\alpha = \frac{2\pi}{L}$ the velocity field can be decomposed as :

$$u(t, x) = \sum_{k=-N}^N u_k(t) e^{i\alpha kx}, \quad (4.1)$$

where N is the number of Fourier modes taken into consideration.

Then the 1D PDE can be replaced by a set of N ODEs in the following form:

$$\dot{u}_k = (\alpha^2 k^2 - \alpha^4 k^4) u_k - \frac{1}{2} i\alpha k \sum_{p=-N}^N u_p u_{k-p} \quad k = -N, \dots, N \quad (4.2)$$

Then the system has been solved using a fourth order explicit Runge-Kutta time marching technique and the Fast Fourier transform to evaluate the nonlinear term. All the numerical results presented in this chapter are obtained for a domain length equal or bigger than $\tilde{L} = \frac{L}{2\pi} = \frac{22}{2\pi}$, because as pointed out by Cvitanović et al. (2010) this is the minimum length to obtain a self sustained chaotic solution. The number of modes chosen is $N = L$ in order to obtain a fully resolved spectrum.

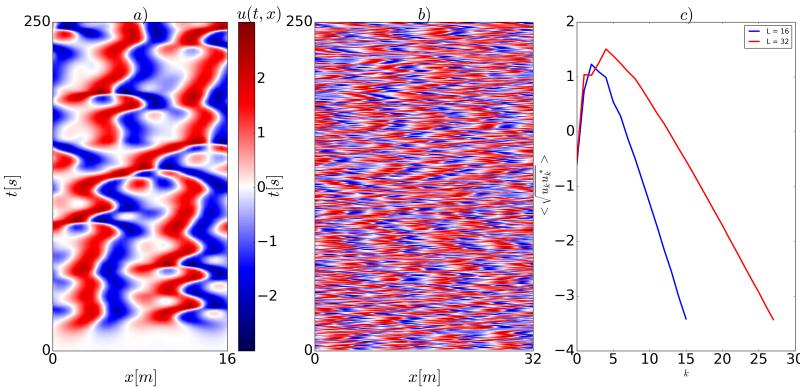


FIGURE 4.1: Chaotic solution of Kuramoto-Sivashinsky equation for two different domain length. L=16 a) b) L=32. In figure c) the correspondent spectrum are plotted

Figure 4.1 shows two different solution of equation (4.1). It is possible to notice how increasing the length the number of structure contained in it increases as well (figure a) and b)). This can be visualized in figure c) where the two spectra show that for the case of longer domain more Fourier modes are needed to reach the target resolution.

4.3 Energy analysis in Kuramoto-Sivashinsky flow

One of the main target of this work is to provide a complete framework and a rigouros procedure to perform sparsification of reduced order models. In particular, one of the targets is to show that the l_1 sparsification procedure is somehow coherent with the physics of the flow. We expect that the interaction that will be selected by the sparsification will be the ones that have the bigger energy transfer in the energy balance of the system. The energy analysis will be performed directly in the Fourier space (see Pope (2001) for all the technical aspects). The equation describing the evolution of the energy is obtained by multiplying equation (4.2) by its complex conjugate (indicated with the * superscript) and summing up, leading to:

$$\begin{cases} (u_{-k})(u_k = (\alpha^2 k^2 - \alpha^4 k^4)u_k - \frac{1}{2}i\alpha k \sum_{p=-N}^N u_p u_{k-p} \\ + \\ (u_k)(\dot{u}_{-k} = (\alpha^2 k^2 - \alpha^4 k^4)u_{-k} - \frac{1}{2}i\alpha(-k) \sum_{p=-N}^N u_p u_{-k-p}) \end{cases}$$

Considering the symmetry $u_k = u_{-k}^*$ and defining the energy within one mode as $E_k(t) = \frac{1}{2}u_k(t)u_k(t)^*$ it is possible to describe its evolution as:

$$\dot{E}_k = T_k + (\alpha^2 k^2 - \alpha^4 k^4)E_k \quad (4.3)$$

where T_k represents the sum of all triadic interaction between three different harmonics k, p, q such that $k = p + q$, it is expressed as:

$$T_k(t) = (\alpha k)Re \left\{ i \sum_{p+q=k} r_{pqk} \right\} = (\alpha k)Re \left\{ i \sum_p u_p u_{k-p} u_k^* \right\}, \quad (4.4)$$

where in this case Re indicates the real part of a complex function. Could be shown that T_k satisfies the energy preserving property $\sum_k T_k = 0$

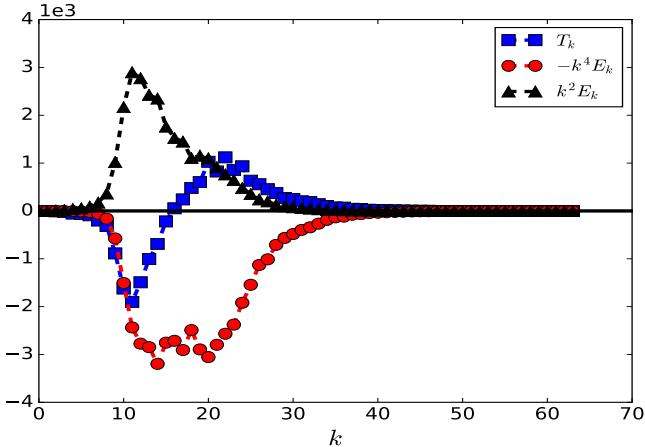


FIGURE 4.2: Trend of the temporal average of the terms T_k , $\alpha^2 k^2 E_k$ and $-\alpha^4 k^4 E_k$ in the equations 4.3. Solution obtained on a domain of length $L = 64$ and $N = 64$ modes

In figure 4.2 are reported the trend of the production $\alpha^2 k^2 E_k$ dissipation $-\alpha^4 k^4 E_k$ and transfer term T_k in the equation 4.3. The different location of the peaks of production and dissipation is due to the fact that the energy is injected in the big scales and dissipated in the small scale.

It is interesting to notice how the term T_k changes sign along the spectrum, this is the mathematical consequence to the fact that the energy is transferred from the bigger modes, low k and T_k negative, to the smaller modes, high k and T_k positive.

Then the physical role of T_k is to transfer energy from the bigger scales to the smaller ones through the interaction between three different modes. The aim is to analyze the numerical value of the single triadic interaction $r_{pqk} = (\alpha k) \text{Re} \left\{ i \sum_{p+q=k} u_p u_q u_k^* \right\}$ where we expect to observe the phenomenon of separation of scales explained by other author in the first chapter.

Figure 4.3 shows respectively two examples of local and nonlocal interactions for one spatial dimension and two spatial dimension. An interaction is defined local when the correspondent triad form a triangle that as much as possible close to be equilateral. As the triangle became far from the equilateral shape the interaction become nonlocal.

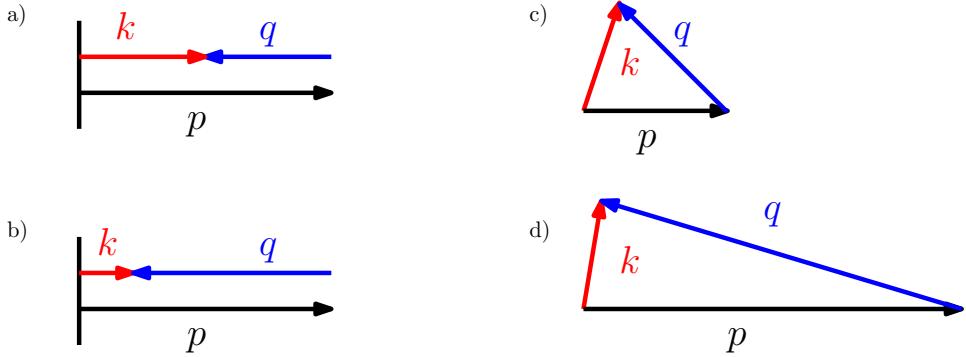


FIGURE 4.3: Schematic example of the different nature of the interactions.
a) local interaction in 1D and b) distant interaction in 1D. The images c)
d) shows respectively the same concept of local and distant interactions
generalized to 2 or 3 dimensional wave vectors

The structure of the Kuramoto-Sivashinsky model in the Fourier space (4.2) has already a sparse structure. In fact the linear term ($\alpha^2 k^2 - \alpha^4 k^4$) is a diagonal matrix. The non linear term associated with the dynamics of a certain mode k , $\sum_{p+q=k} u_p u_q$, takes into consideration only the interactions between modes such that $p + q = k$. From a formal point of view the non-linear term can be expressed as $\sum_p \sum_q Q_{kpq} a_p a_q$, where Q_{kpq} has a banded sparse structure. In particular, fixing the value of k , the set of points different from zero are the one on the line $q = k - p$ in the pq plane. All these lines are parallel so if we collapse them on a single plane they will not overlap. Hence, we can generate diagrams like the one in figure 4.4 where this procedure is graphically explained. No information is lost in the process.

Due to the already sparse structure of the Fourier representation of the Kuramoto-Sivashinsky equation, it is possible to create a 2D diagram that includes all the triadic interaction of the nonlinear term.

In figure 4.5 are plotted the two spectral coordinates of the system in the Fourier space p and q . Because of the sparse structure of the problem all the coefficients for every mode are collected on a diagonal $p + q = k$ then it is possible to collapse a third order tensor on a 2D plane. Moreover it is possible to define two new axis that represent the actual mode $k = p + q$ and the distance of the interaction $\eta = p - q$. This representation of the system makes natural define the interaction as local and nonlocal according their position on the pq or $k\eta$ plane. In particular we will define as :

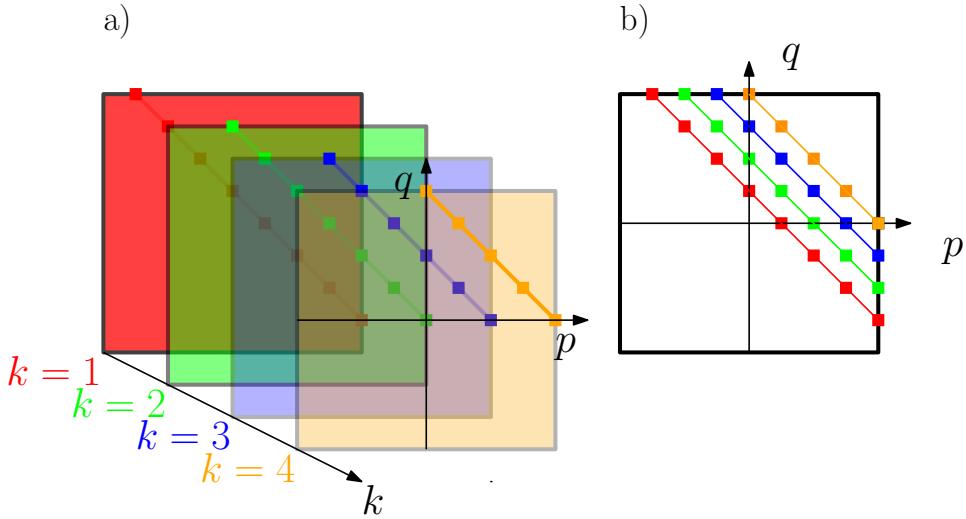


FIGURE 4.4: a) Structure of the tensor Q_{kpq} where the different colour represent the different value of index k , while the indexes (pq) represents the the coordinates inside every squares. The sub diagonals in every square in figure a) are the interactions active for a certain value of k . Because of this structure the tensor Q_{kpq} can be collapsed on a single plane (figure b). The final results is a diagram (figure (b)) where the colours represent the interactions active for different modes

- local interaction : all the point in the plane such as $p \sim q \sim k$, in other words all the points near the axis k i.e. characterized by a low value of η
- distant interaction : the point in the plane such as $p \gg q \sim k$ or $p \sim q \gg k$, in other words all the points far from the axis k i.e. characterized by a high value of η

The aim is to use this kind of diagrams to visualize the sparsity pattern of the solution of the Kuramoto-Sivashinsky flow.

Figure 4.3 left shows one example of graphical visualization on a logarithmic scale of the average of the absolute value of the terms $r_{pqk} = \text{Re} \{i(\alpha k) u_p u_q u_k^*\}$ in the energy equation. It is clear to see how the most energetic terms are the ones correspondent to the modes at low k and with low values of the coordinate $p - q$, that means that the biggest energy transfer is concentrated in the biggest modes while for interaction that does not differ to much as wavenumber.

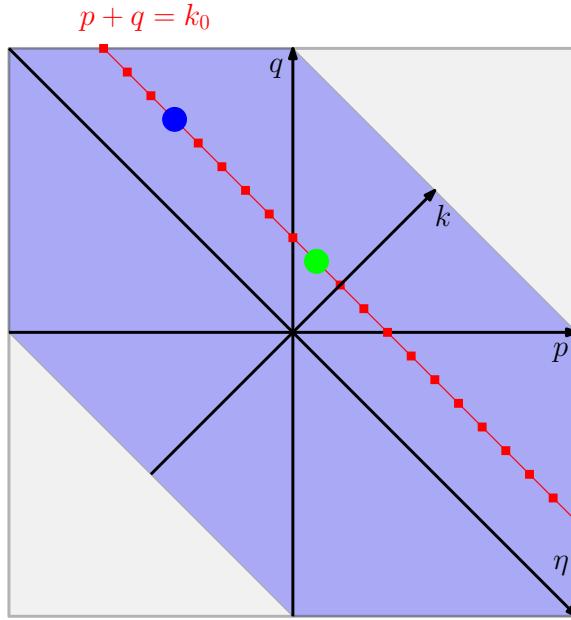


FIGURE 4.5: 2D diagram to visualize the triadic interaction in the Kuramoto-Sivashinsky model. The blue area represents the totality of the interactions presents in the nonlinear term of the equation 4.2. The red squares on the red diagonal line represents all the interactions for the mode taken into consideration ($k_0 = 3$ in this case). The blue dot represent a distant interaction while the green dot represent a local interaction according to the definition given in figure (4.3)

The right visualization of the energy interactions shows the location of the most important interaction from a global point of view, however we lose all the information about the relative importance of the interaction inside a single mode.

Figure 4.3 right is shown the logarithmic diagram of the average energy transfer term r_{pqk} normalized respect to the global energy transfer term within a single mode T_k . In this way it is possible to visualize the relative importance of the interactions inside the single modes.

This two diagrams will be utilized later when we will compare the two different sparsification strategies varying the definition of the parameter λ as previously explained in sec 1.2.2.

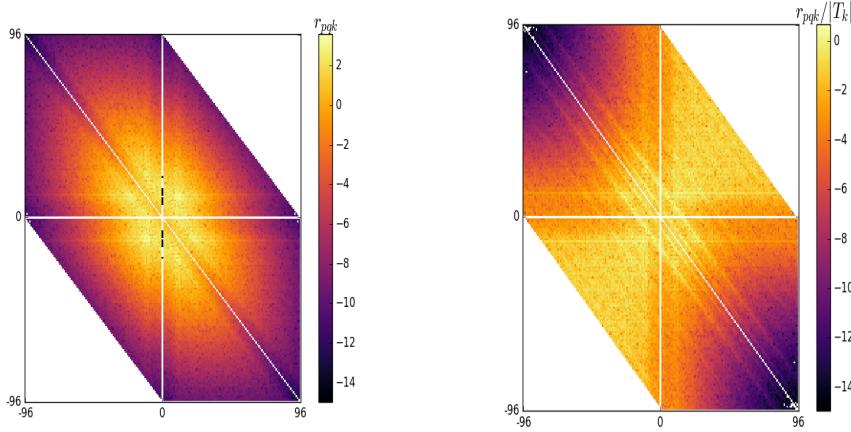


FIGURE 4.6: Temporal average of : left) absolute value of the energy interactions $r_{pqk} = \operatorname{Re} \left\{ i \tilde{k} u_p u_q u_k^* \right\}$ right) relative value of the energy interaction $\frac{r_{pqk}}{|T_k|}$

4.4 Regression procedure

The Fourier representation of the Kuramoto-Sivashinsky equation is very convenient from a computational point of view and for the simplicity of the physical interpretation of the result. However the complex nature of the Fourier transform need a bit of rework to make the dataset usable for the LASSO regression in the form 3.4. Since the LASSO is not easily generalizable to the complex numbers is more convenient to decouple the real and the imaginary part in equation (4.2) rewriting the complex Fourier coefficient $u_k = r_k + i s_k$. Then the system 4.2 is equivalent to

$$\begin{cases} \dot{r}_k = (\alpha^2 k^2 - \alpha^4 k^4) r_k - \frac{1}{2} (\alpha k) \sum_{p=-N}^N (r_p s_{k-p} + s_p r_{k-p}) \\ \dot{s}_k = (\alpha^2 k^2 - \alpha^4 k^4) s_k - \frac{1}{2} (\alpha k) \sum_{p=-N}^N (r_p r_{k-p} - s_p s_{k-p}) \end{cases}$$

Then two different LASSO regression in the form 3.4 will be performed separately on the real and the imaginary part.

4.5 Length influence on the sparsity

In this section we will tackle to core problem of this work: i.e. understand how the number of degrees of freedom of the system affect the sparsity pattern and the reconstruction error of the system itself.

In the first paragraph has been explained the Kuramoto-Sivashinsky model has the peculiarity to increase the number of dynamically active range of motion as the length of the domain increases. In this work three different test case will be taken into considerations with the respective length of $L = 32, 64, 96$. To obtain a fully resolved solution we choose a number of degrees of freedom N equal to the length of the domain.

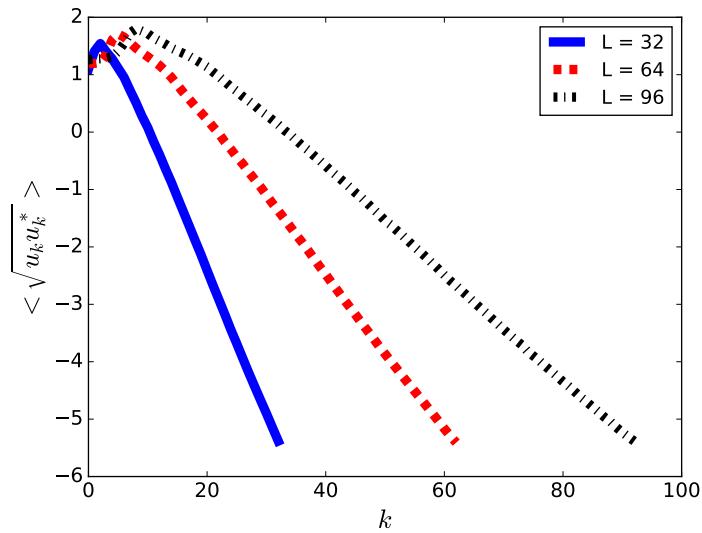


FIGURE 4.7: Energy spectrum for three different domain length

In figure 4.7 the spectrum for the three different solutions of the Kuramoto-Sivashinsky equation is reported. From the spectrum it is easy to observe the phenomenon of the separation of scales explained previously. It shows how increasing the length of the domain the number of scales necessary to obtain the same spectral resolution increases leading to a bigger separation between the large and the small scales inside the system.

4.5.1 Energy interactions

In chapter three we explained that it is possible to generate different sparsification approaches changing the value of the regularization weight λ . Here we will show that the aim of this procedure is to generate a sparse pattern that reproduce the absolute value or the relative value of the energetic interactions.

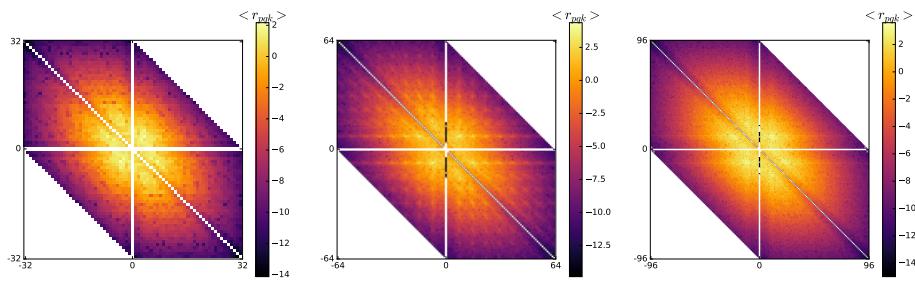


FIGURE 4.8: Temporal average of the absolute value of the real part of r_{pqk}

Figure 4.8 show the value of the term $r_{pqk} = \text{Re} \{ i(\alpha k) u_p u_q u_k^* \}$ in logarithmic scale plotted on the diagram explained in figure 4.5. From the figure it is clear how for all the three different domains the most energetic interactions are located in the area of low wavenumber. We expect that the approach at $\lambda = \text{const}$ across the spectrum will generate a similar sparsity pattern.

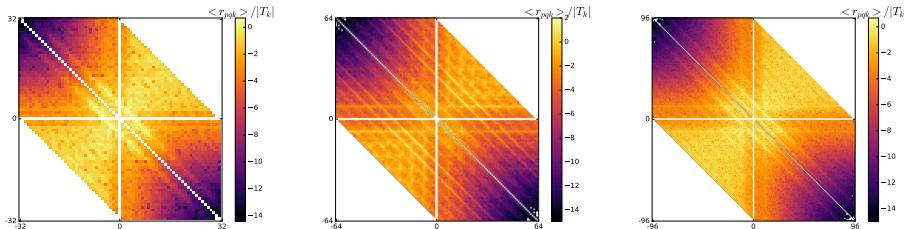


FIGURE 4.9: Temporal average of the absolute value of r_{pqk} divided by the absolute value of the total energy transfer for every mode T_k

Figure 4.9 show the value of $r_{pqk} = \text{Re} \{ i\alpha k u_p u_q u_k^* \} / |T_k|$ in logarithmic scale. This quantity is representative of the relative importance of the energetic interaction within the mode itself. We expect that the sparsification technique with $\lambda = \langle \sqrt{u_k u_k^*} \rangle \lambda$ reproduces a similar sparsification pattern.

4.5.2 Greedy sparsification

In this section we will briefly present the results obtained with the greedy technique. We remind that this approach simply remove from the system the term correlated to the non relevant interactions plotted in the figure 4.8. No optimization process is performed on the coefficients. The reconstruction error of the sparsified system is defined as :

$$\epsilon = \frac{\sum_k (u_k - \tilde{u}_k)^2}{\sum_k u_k^2}, \quad (4.5)$$

where the summation is extended to all the modes present in the original system and the tilde (\tilde{u}_k) represents the reconstructed system. Figure 4.10 shows the isolevels of ϵ for the greedy technique. The colour on the graph represents the value of ϵ when a certain interaction $u_p u_q$ is eliminated from the system.

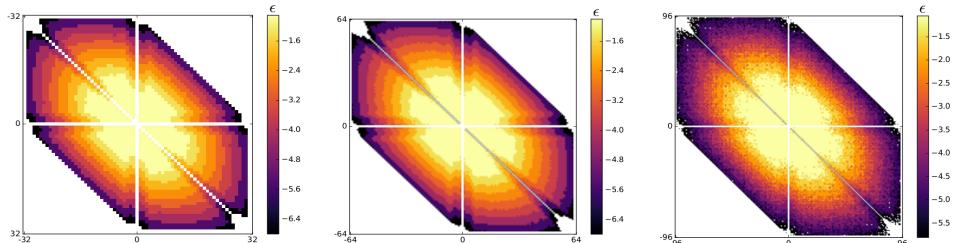


FIGURE 4.10: Absolute value of the reconstruction error ϵ for different lengths of the domain: $L = 32$ left , $L = 64$ centre and $L = 96$ right

To better visualize the effects of the length of the domain on the sparsity pattern of the solution it is useful to construct a plot where on the abscissa is reported the density of the system $\rho = 1 - \frac{N_{zero}}{N_{tot}}$ where N_{zero} is the number of elements set to zero and N_{tot} is the total number of elements. Then $\rho \in [0, 1]$, $\rho = 1$ represents the completely full system while $\rho = 0$ represents a system where all the interactions are set to zero. In ordinate is plotted the value of ϵ defined in (4.5.2).

Figure 4.11 shows that the curves collapse one over the other for the three different domain lengths. Moreover, the three curve grows monotonically as soon the density of the system ρ moves from 1 to a lower value. Hence the

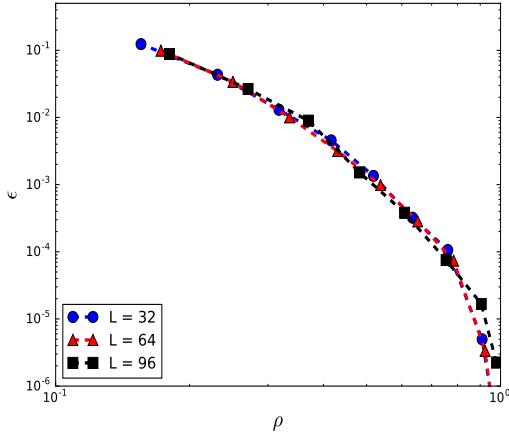


FIGURE 4.11: Reconstruction error for the three different domain lengths obtained with the greedy technique

fact that there is no plateau in the right part of the figure 4.11 means this is not a good sparsification technique because it is not possible to eliminate terms from the system and keep the reconstruction error almost constant at the same time. Moreover it is interesting to notice how this kind of approach seems not to show any effect of the length of the domain (number of degrees of freedom) on the sparsity pattern of the solution.

In this report for time reason has been considered only the case where the cut off is proportional to the absolute value of the energetic interactions 4.8. The same procedure can be performed considering the sparsity pattern arising from the relative intensity of the interactions 4.9. This approach will be analysed in the continuation of the project.

4.5.3 l_1 regression with $\lambda_k = \text{const}$

The philosophy of this approach is to keep the sparsification parameter constant along all the modes, i.e $\lambda_k = \lambda_0$, with λ_0 we intend the coefficient used for the first Fourier mode.

Unfortunately due to time issue we will not be able to present the results for the largest domain ($L = 96$). Hence, we will discuss only the results

related to the cases with $L = 32$ and $L = 64$. The results from the sparsification procedure show how the emerging structure is very similar to the one emerging from the absolute value of the energy transfer term r_{pqk} plotted in 4.8.

In particular in figure 4.12 are reported the isolevels of the reconstruction error ϵ that is made when the coefficient of the corresponding colour is eliminated from the system.

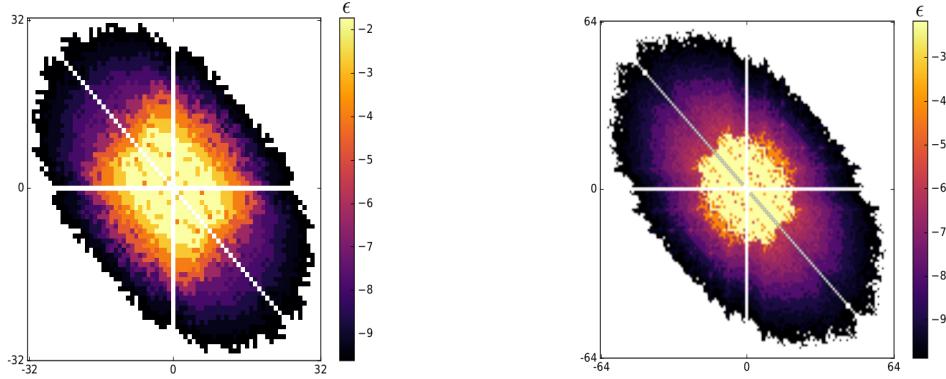


FIGURE 4.12: Absolute value of ϵ made when the corresponding coefficient is eliminated from the system. Domain length of: $L = 32$ left , $L = 64$ right

Figure 4.13 shows the isolevels of the penalization weight λ . In particular, the colour corresponds to the value of λ at which the corresponding coefficient is shrunk to zero.

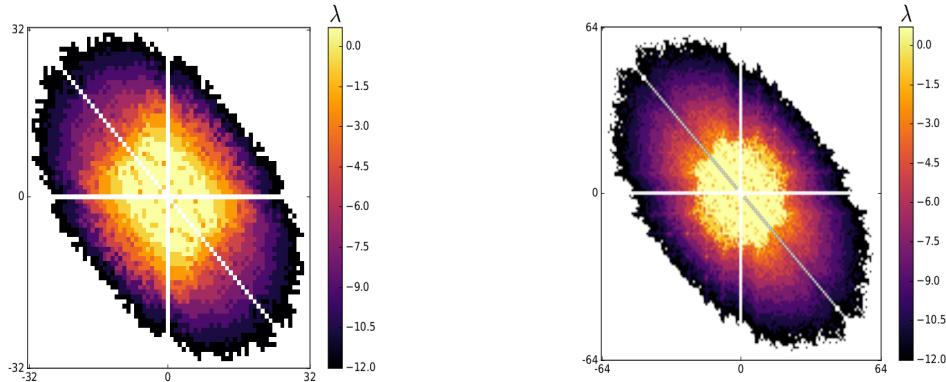


FIGURE 4.13: Value of the sparsification parameter λ for whom the correspondent coefficient disappear from the system. Domain length of: $L = 32$ left and $L = 64$ right

The results obtained are pretty reasonable, the sparsification parameter λ in the can be interpreted as a cut off on the amplitude of the modes, keeping λ constant along the spectrum will lead to a solution where the mode of smaller amplitude will be penalized much more respect to the ones of larger amplitude. This lead to a sparsification pattern that is a compromise between the two approaches presented in the section 1.2.2 of the introduction i.e.:

- The most energetic modes are sparsified i.e the distant interaction, points located in the corners of the diagram 4.5, are eliminated from the dynamics of the mode itself.
- The less energetic modes are completely eliminated from the system.

We can then visualize the reconstruction error on a graph (figure 4.14) where in ordinate is plotted the reconstruction error as defined in equation (4.5.2) and the density of the system ρ in abscissa.

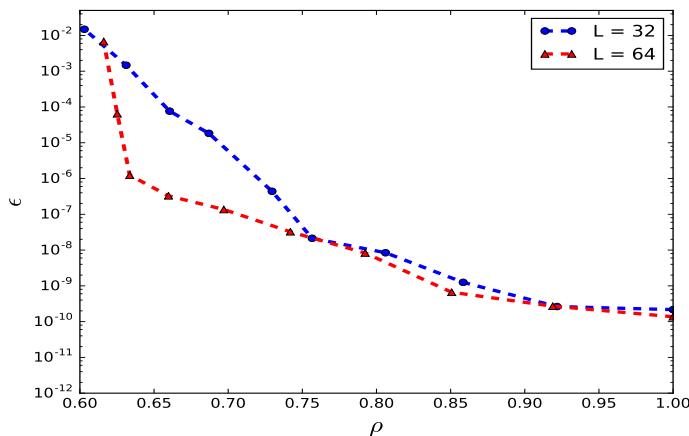


FIGURE 4.14: Trend of the reconstruction error ϵ vs the density ρ for short ($L = 32$) and intermediate ($L = 64$) domain length

The curves in figure 4.14 have been generated varying the value of λ from very low values (right part of the figure) to fairly high values (left part of the picture). Every point in figure corresponds to the system generated by a certain value of λ . The figure shows how for the longer domain it is possible to reach lower values of density for the same reconstruction error ϵ .

Moreover, it is interesting to notice that the two curves are almost horizontal for high level of density, this means that is possible to remove the distant interactions without affect at all the reconstruction error of the system. An other interesting feature is the fact that the two curves change slope for different values of density. This could be somehow related to the movement of the knee point to regions of higher sparsity (and lower density ρ) we would like to observe increasing the number of degrees of freedom of the system. All these results are promising and seem to show a behaviour qualitatively similar to the one expected. We are currently working on a third simulation with a larger domain in order to confirm this trend.

4.5.4 l_1 regression with $\lambda_k = \langle \sqrt{u_k u_k^*} \rangle \lambda$

The approach of keeping a value of λ constant along the whole spectrum leads to a final system that is a trade off between sparsification and truncation. This could provide good results in terms of the reconstruction error ϵ . However, the main philosophy of the sparsification procedure is to generate a less computational expensive system without eliminating any mode but only the interactions that are not important for the dynamics of the mode itself.

In figure 4.9 is reported the average value of the energy transfer term r_{pqk} normalized for the total energy transfer within every mode T_k . This is a way to visualize the most energetic interactions inside every modes. The normalization of the regularization weight λ should bring to a sparsity pattern that consider the relative importance of the interactions within every mode.

In this sparsification approach the value of the sparsification parameters λ is not kept constant but is scaled respect to the value of the mean energy of every mode. $\lambda \rightarrow \langle \sqrt{u_k u_k^*} \rangle \lambda_0$ where λ_0 is the reference value for the zero mode.

As done for the previous approach explained in section 4.5.3, figure 4.15 reports the value of the error ϵ obtained when the interactions of the corresponding colour are eliminated from the system. While figure 4.16 shows the value of λ_0 at which the coefficient of the corresponding colour is eliminated

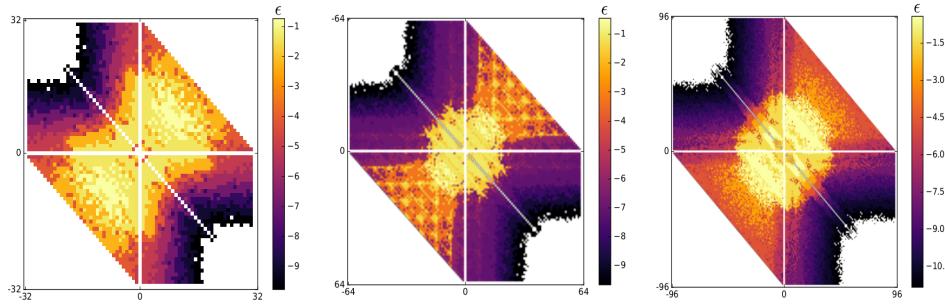


FIGURE 4.15: Value of ϵ when the coefficient of the corresponding colour is eliminated from the system for increasing length of the domain: $L = 32$ left , $L = 64$ center and $L = 96$ right

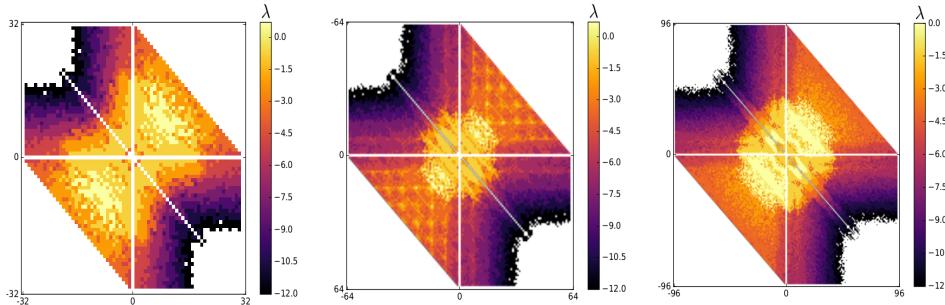


FIGURE 4.16: Value of the sparsification parameter λ_0 at which the corresponding interactions are eliminated from the system. Lengths of the domain: $L = 32$ left , $L = 64$ center and $L = 96$ right

from the system by the LASSO regression. It is interesting to notice how in this case the shape of the sparsity pattern is closer to the pattern of the relative energy transfer plotted in 4.9. This means that a scaling of the sparsification parameters proportional to the energy of the corresponding mode is a reasonable assumption in order to obtain a procedure more oriented towards the sparsification approach respect to the truncation approach.

Finally we would like to underline how the choice of the scaling of the sparsification parameter λ is totally arbitrary, here we have just tested one of the possible solutions.

Figure 4.17 has been generated varying the value of the regularization weight λ in the same way as 4.14. It shows the trend of the reconstruction error for the three different domain lengths taken into consideration. It is interesting to notice how the reconstruction error ϵ corresponding to a certain density of the system becomes lower increasing the length of the domain. This

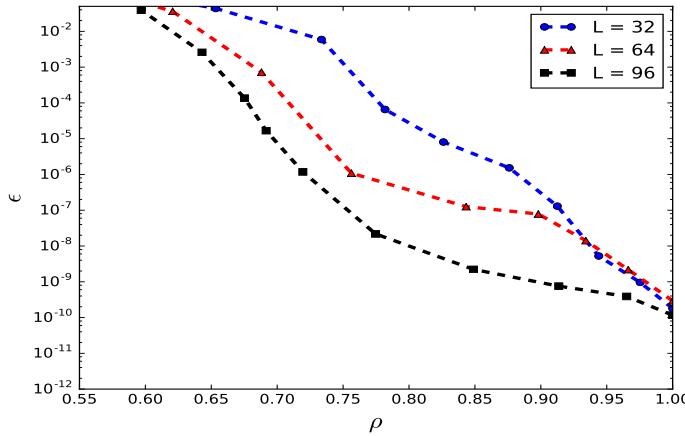


FIGURE 4.17: Trend of ϵ and density ρ of the system for the three different domains length

is a further confirmation that the optimization procedure involved in the l_1 regression is able to capture the effect of the separation between scales emerging increasing the length of the domain.

4.6 Comparison between different strategies

The last section of the chapter is dedicated to briefly compare the results obtained with the two different approaches of the l_1 regression.

Figure 4.18 shows mainly two aspects. The first is the dependency of the reconstruction error ϵ from the length of the domain. A larger domain allows to obtain the same reconstruction error ϵ for a lower value of density ρ of the system. This behaviour could be due to the fact that in this kind of system increasing the length of the domain increase the number of dynamical active modes. Due to the locality of the energetic interactions the dynamics of a certain structure is only weakly dependent from the structures that are much smaller. This allows a greater sparsification of the system. The second aspect is that the error for the regression performed keeping the value of the regularization weight constant provide better performances if the objective is to minimize the reconstruction error ϵ . Moreover, for the case with $\lambda = c$ the slope of the curves is much lower respect the case at

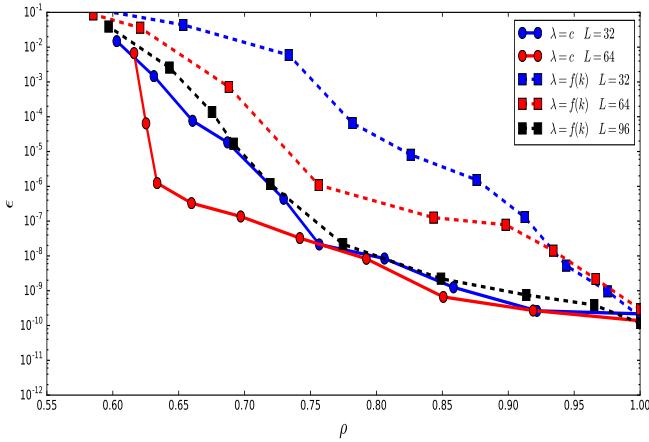


FIGURE 4.18: Trend of ϵ against ρ for all the cases taken into consideration. The squares linked with dotted line represent the test case where the regression has been performed with $\lambda_k = \langle \sqrt{u_k u_k^*} \rangle / \lambda$ (section 4.6.2). The circles linked by continuous line represent the regressions performed keeping the value of λ constant along the spectrum (section 4.6.1)

λ variable. This means that it is possible to remove interactions without affecting the reconstruction error of the system.

4.7 Conclusions

The results obtained from this preliminary study can be divided into two categories: the first regarding the choice of the sparsification procedure and the second regarding the performance between the different approaches. Regarding the choice of the sparsification technique three different techniques have been proposed:

- The greedy approach: the interactions are ranked according to the absolute energy transfer associated and then the least energetic ones are neglected. This technique although computationally efficient seems to perform quite poorly for the system taken into analysis. The reconstruction error grows exponentially as the density of the system decrease (figure 4.11). Moreover this approach seems not able to describe the effect that the separation between scales is expected to have on the sparsity pattern.

- LASSO regression with $\lambda = c$: this approach seems to reproduce the sparsity pattern associated with the absolute value of the energetic interactions (compare figures 4.8 left and 4.12). The result is the truncation of the smaller scales and the sparsification of the largest scales. An effect of the domain length on the sparsity pattern of the solution seems to be present.
- LASSO regression with $\lambda_k = \langle \sqrt{u_k u_k^*} \rangle / \lambda$: this approach reproduces the sparsity pattern associated with the relative energy transfer within every mode (compare figures 4.9 right and 4.15). The result is closer to pure sparsification when only the interaction inside every mode are neglected without the complete elimination of any mode. The same effect of the domain length on the sparsity pattern seems to be present even if less important respect the previous case.

About the different performances between the two different approaches obtained with the LASSO regression a possible explanation could be found in the Fourier representation of the model. The representation in the Fourier space generates a system that has already a sparse structure of the nonlinear interactions matrix. It is reasonable to suppose that, because of this already sparse structure, in order to obtain a lower reconstruction error could be more convenient to eliminate a mode with low energy than eliminate interactions in the dynamic of a mode with high energy.

All the preliminary results presented so far seem to lead in the direction indicated by the works related to the locality of the scale interaction in turbulence. However, in order to confirm this results the next immediate steps will be: perform regression on biggest domains in order to confirm these trends. Implement a cross validation procedure in order to establish the statistical relevance of the results.

Chapter 5

Conclusions

This chapter will be dedicated to recap all the findings and conclusions obtained so far and how these findings will be useful in the continuation of the project. The conclusions will be divided into two sections. The first section dedicated to a brief discussion about the results found in the literature on previous studies. The latter dedicated to a brief discussion of the results obtained so far and how this results can be useful for the continuation of the project.

The literature shows that reduced order modelling technique is a very well studied topic. In particular several works have been found concerning how to treat all the problematics emerging from a low order approximation of the governing equations. Many authors agree that two of the most compelling are: the loss of numerical stability related to an insufficient viscous dissipation in the smaller scales of the model and the modeling of the term related to the pressure field in the Navier-Stokes equations.

Regarding the sparse identification of dynamical systems several efforts have been dedicated to construct the theoretical background for the sparse regression techniques. Recent works show how these set of techniques can be successfully extended to the framework of the dynamical systems. However, it has not been shown yet if these methods are able to describe the physics of the interactions in a turbulent flow.

In literature the locality of the interactions between modes has been assessed for low Reynolds number flows, anyway, these results seems likely to be valid for higher Reynolds regime as well. In particular, increasing the Reynolds number a more evident sparse structure of the system seems to emerge. Hence, we expect that the sparse regression techniques are able to capture this physical process allowing and increasing sparsity structure as the Reynolds number increases.

The first test case take into consideration, the Kuramoto-Sivashinsky model, has the interesting characteristic to increase the range of dynamical active structures as the the length of the domain increase. This property somehow reproduces the separations of scales observed in 3D turbulent flows as the Reynolds number increases.

The analysis performed on this model shows how the separation of scales arising from the increasing of dynamical active modes in the model allows to generate sparse models that achieve the same performances with a lower density ρ of the system. This is in accordance with the initial hypothesis about the effect of the separation of scales on the performance of the sparse model.

The sparsity pattern generate by the sparse regression is in accordance with the magnitude of the energy interactions between the modes. Moreover, three different sparsification approaches has been proposed, namely:

- Greedy approach : the idea is to neglect the least important energetic interactions from the system without perform any optimization procedure on the coefficients
- l_1 with $\lambda = \lambda_0$: this procedure generates a sparsity pattern that reproduce the pattern of the absolute value of the energetic interactions. In particular, the final result is trade off between the two conjoint effect of : truncation of the small modes and sparsification of the bigger modes.
- l_1 with $\lambda = \lambda < \sqrt{u_k u_k^*} >$: we have proposed a normalization of the sparsification parameter proportional to value of the mean kinetic

energy energy of every mode. This approach generates a sparsity pattern that is in agreement with the relative intensity of the energetic interactions within every mode.

The performances of these three different approaches have been compared. The greedy approach doesn't provide good results. It is not a suitable approach to perform sparsification because the reconstruction error increases exponentially as the sparsity of the system increases. Between the two approaches involving a l_1 regression the approach with a constant sparsification parameter provides better performances. This is probably due to the fact that for this kind of system due to the Fourier decomposition of the solution the nonlinear interaction matrix is already sparse. Hence an approach that privilege the truncation respect to the sparsification is to be favoured. The great differences of performances between the greedy approach and the one based on the l_1 regression is probably due the fact that the optimization procedure change the value of the coefficients of the system in order to compensate the elimination of the less important terms.

It is worthy to stress out how this system in the Fourier space presents already a sparse structure of the interaction. The main target of this analysis was to validate the procedure on a system where the value of the coefficients is known. However, because of the structure of the system we did not expect to reach high sparsity ratio with a reasonably low value of the reconstruction error. Nevertheless, the system shows the expected qualitatively behaviour. We expect that the same behaviour of the reconstruction error will be preserved for the POD Galerkin systems as well. Moreover, due to the full density of the original system we will expect to reach good performance for much more sparse system.

Chapter 6

Future Work

In this chapter we will explain the next steps that will be undertaken in the project. In particular, after the first application of the procedure on the Kuramoto-Sivashinsky model the project will be divided in two main branches. The first is the extension of the standard l_1 regression approach to more geometrically complex test cases, namely the 2D cavity flow and the 3D backward facing step.

6.1 Cross validation

The first step to conclude the analysis related to the Kuramoto-Sivashinsky model is to implement a cross validation procedure in order to check the statistical reliability of the solution. As explained in chapter 2 the cross validation is a step of key importance to analyse the results coming from statistical analysis of the data and in particular linear regression. The cross validation procedure will be implemented using the python machine learning library [Pedregosa et al. \(2011\)](#). Subsequently the cross validation will be performed on all the results coming from the sparse regression.

6.2 l_1 regression

The main idea behind the sparsification procedure has been explained in the current report. The next step is to extend this procedure to more complex test case. In the following subsection we will briefly explain the main point of the working programme and the problems that will be more likely to be faced.

6.2.1 Interface with the DNS code

In the present work the dataset comes from a direct numerical simulation performed with the pseudo spectral finite element code nektar++ [Cantwell et al. \(2015\)](#). The finite element approach guarantee enough flexibility and accuracy to handle fairly complex domains. This advantages come to an increasing complexity of the proper orthogonal decomposition procedure and of the reduced order model generation. In particular the target of this phase of the project is to generate a working interface that allow us to compute the correlation matrix necessary to compute the correlation matrix in the proper orthogonal decomposition. In the literature review we explained how the perform ROM the construction of the correlation matrix \mathbf{R} is needed. The general definition of \mathbf{R} is :

$$R_{ij} = (\mathbf{u}_i, \mathbf{u}_j)_V = \int_V \mathbf{u}_i \cdot \mathbf{u}_j dV \quad (6.1)$$

This integral must be evaluated on the whole domain and on a general mesh, the target is to use the already implemented library used by the DNS code in order to compute this kind of integrals.

Similarly the same interface can be used to compute the matrices involved in the reduced order model, that are in the same form.

This procedure it is highly recommended because leads to a direct advantage. It is possible to use already well validated and tested routine in order to compute the integral and the derivatives necessary to construct the reduced order model [2.8](#). Moreover, the high order finite elements nature of the numerical method implemented in the code leads to the introduction of

extra degrees of freedom inside the element itself. It is probably easily and more accurate to manage them through their own library instead of write a new routine from scratch.

This step is necessary to extend the analysis to the next test cases : 2D cavity flow and 3D backward facing step.

6.2.2 2D cavity flow

Once the interface with the DNS code has been created the first test case to be analysed will be the flow inside a 2D square cavity. We consider this a good test case because it is not too computational expensive but the flow generated is fairly complex to be a good test case to test our methodology. Moreover due to the fact that the flow inside a cavity is a closed flow the development of the corresponding reduced order model is easier.

The complete analysis on this flow will be performed in the following step:

- Creation of the dataset from the DNS simulations in nektar++.
- Proper orthogonal decomposition of the flow , statistical convergence of the POD modes. Generation of the fully connected Galerkin ROM that will be the reference for both accuracy and computational cost.
- Energy flow analysis in the POD space and implementation of the different sparsification approaches presented in this report
- Parametric analysis of the effects of the Reynolds number on the separation between scales an sparsity pattern.

6.2.3 3D backward facing step

The last test case that will be taken into consideration will be the separated flow around a backward facing step. The procedure will be exactly the same explained for the 2D cavity flow. All the routines will have already been validated and fully tested on simplest test cases. The major difficulties of this test case will be related to computational effort needed to generate the dataset from the DNS simulations.

Appendix A

Planning

The main task that will be accomplished in the second year of research will be:

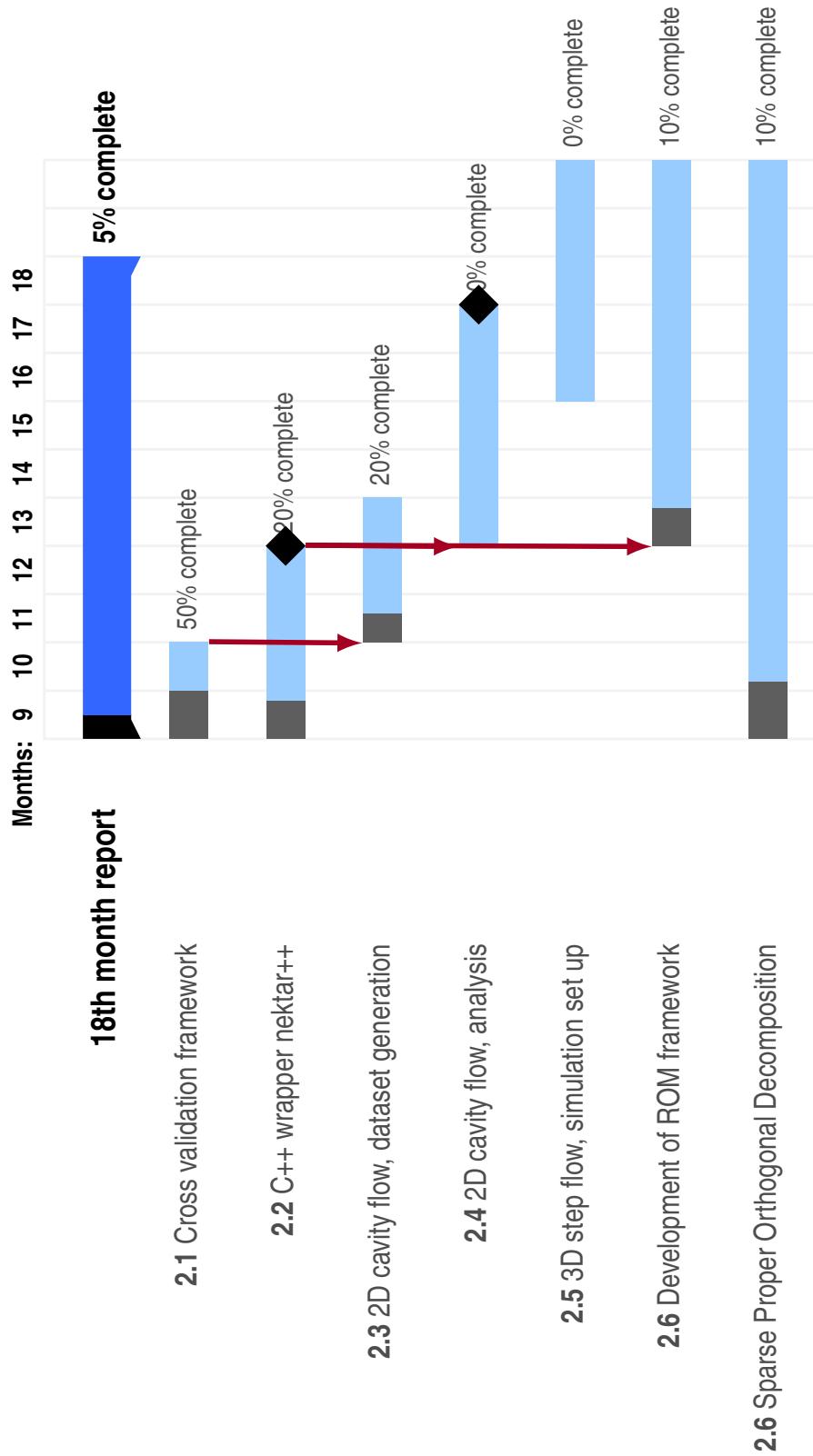
- Implementation of the cross validation procedure
- Developing of c++ wrappers to use the routine implemented in the DNS code
- Generation of the dataset for the 2D cavity flow. The dataset will consist of a set of simulations at different values of the Reynolds number.
- Analysis of the sparsity pattern in the 2D cavity flow.
- Starting to set up the simulation for the 3D backward facing step

In the plan there are two important milestones. The first one is the development of the c++ wrappers. The achievement of this step is necessary to compute integrals and derivatives on non uniform meshes in order to generate the ROMs. The second one in the analysis of the 2D cavity flow. This milestone comprehends the sparse regression, energy analysis and effect of the Reynolds number on the sparsity pattern. These results will be an important part of the 18th month report.

Task 2.5 probably will be not part of the second progression review but it would be important to start the set up of the simulations and of the

sparsification procedure already during the second year of research. The same is valid for the task 2.6 that starts during the second year but continues for the all duration of the project.

Alongside the main project a part of the total time will be dedicated to the development of the framework related to the Sparse Proper Orthogonal Decomposition.



Bibliography

Frédéric Alizard and Jean-Christophe Robinet. Modeling of optimal perturbations in flat plate boundary layer using global modes: benefits and limits. *Theoretical and Computational Fluid Dynamics*, 25(1-4):147–165, 2011.

Maciej J Balajewicz, Earl H Dowell, and Bernd R Noack. Low-dimensional modelling of high-reynolds-number shear flows incorporating constraints from the navier–stokes equation. *Journal of Fluid Mechanics*, 729:285–308, 2013.

Martin Benning, Florian Knoll, Carola-Bibiane Schönlieb, and Tuomo Valkonen. Preconditioned admm with nonlinear operator constraint. In *IFIP Conference on System Modeling and Optimization*, pages 117–126. Springer, 2015.

M Bergmann, C-H Bruneau, and A Iollo. Improvement of reduced order modeling based on pod. In *Computational Fluid Dynamics 2008*, pages 779–784. Springer, 2009.

Michel Bergmann, Charles-Henri Bruneau, and Angelo Iollo. **Improvement of Reduced Order Modeling based on Proper Orthogonal Decomposition**. Research Report RR-6561, INRIA, 2008.

Adel Bibi, Hani Itani, and Bernard Ghanem. Fftlasso: Large-scale lasso in the fourier domain. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, 2017.

Avrim L Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2):245–271, 1997.

- James G Brasseur and Chao-Hsuan Wei. Interscale dynamics and local isotropy in high reynolds number turbulence within triadic interactions. *Physics of Fluids*, 6(2):842–870, 1994.
- Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data. *arXiv preprint arXiv:1509.03580*, 2015.
- Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
- Tan Bui-Thanh, Karen Willcox, and Omar Ghattas. Parametric reduced-order models for probabilistic analysis of unsteady aerodynamic applications. *AIAA journal*, 46(10):2520–2529, 2008.
- Chris D Cantwell, David Moxey, A Comerford, A Bolis, G Rocco, Giandomenico Mengaldo, Daniele De Grazia, Sergey Yakovlev, J-E Lombard, D Ekelschot, et al. Nektar++: An open-source spectral/hp element framework. *Computer Physics Communications*, 192:205–219, 2015.
- 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*, 491:275–284, 2003.
- Predrag Cvitanović, Ruslan L Davidchack, and Evangelos Siminos. On the state space geometry of the kuramoto–sivashinsky flow in a periodic domain. *SIAM Journal on Applied Dynamical Systems*, 9(1):1–33, 2010.
- Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.
- J Andrzej Domaradzki and Robert S Rogallo. Local energy transfer and nonlocal interactions in homogeneous, isotropic turbulence. *Physics of Fluids A: Fluid Dynamics*, 2(3):413–426, 1990.
- JA Domaradzki and Nikolaus A Adams. Direct modelling of subgrid scales of turbulence in large eddy simulations. *Journal of Turbulence*, 3(024):1, 2002.

- B Galletti, CH Bruneau, Luca Zannetti, and Angelo Iollo. Low-order modelling of laminar flow regimes past a confined square cylinder. *Journal of Fluid Mechanics*, 503:161–170, 2004.
- Gene H Golub, Michael Heath, and Grace Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.
- James M Hyman and Basil Nicolaenko. The kuramoto-sivashinsky equation: a bridge between pde’s and dynamical systems. *Physica D: Nonlinear Phenomena*, 18(1-3):113–126, 1986.
- EC Itsweire, KN Helland, and CW Van Atta. The evolution of grid-generated turbulence in a stably stratified fluid. *Journal of Fluid Mechanics*, 162:299–338, 1986.
- Mihailo R Jovanović, Peter J Schmid, and Joseph W Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.
- Vassili Kitsios, Laurent Cordier, J-P Bonnet, Andrew Ooi, and Julio Soria. On the coherent structures and stability properties of a leading-edge separated aerofoil with turbulent recirculation. *Journal of Fluid Mechanics*, 683:395–416, 2011.
- Toni Lassila, Alfio Quarteroni, and Gianluigi Rozza. A reduced basis model with parametric coupling for fluid-structure interaction problems. *SIAM Journal on Scientific Computing*, 34(2):A1187–A1213, 2012.
- Qinghua Liu, Xinyue Shen, and Yuantao Gu. Linearized admm for non-convex non-smooth optimization with convergence analysis. *arXiv preprint arXiv:1705.02502*, 2017.
- Jean-Christophe Loiseau and Steven L Brunton. Constrained sparse galerkin regression. *Journal of Fluid Mechanics*, 838:42–67, 2018.
- Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.
- V.B. Nguyen, M. Buffoni, K. Willcox, and B.C. Khoo. **Model reduction for reacting flow applications.** *International Journal of Computational Fluid Dynamics*, 28(3-4):91–105, 2014.

- Bernd R Noack, Marek Morzynski, and Gilead Tadmor. *Reduced-order modelling for flow control*, volume 528. Springer Science & Business Media, 2011.
- Bernd R Noack, Paul Papas, and Peter 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.
- Jan Östh, Bernd R Noack, Siniša Krajnović, Diogo Barros, and Jacques 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.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Pas-sos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- Stephen B Pope. Turbulent flows, 2001.
- Sivaguru S Ravindran. A reduced-order approach for optimal control of fluids using proper orthogonal decomposition. *International journal for numerical methods in fluids*, 34(5):425–448, 2000.
- Gemma Roig, Xavier Boix, and Fernando De la Torre. Optimal feature selection for subspace image matching. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 200–205. IEEE, 2009.
- Hendrik Tennekes and John Leask Lumley. *A first course in turbulence*. MIT press, 1972.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- Trevor Hastie Robert Tibshirani and Jerome Friedman. *The Elements of Statistical Learning*. Springer.

Timo Tonn, Karsten Urban, and Stefan Volkwein. Optimal control of parameter-dependent convection-diffusion problems around rigid bodies. *SIAM Journal on scientific computing*, 32(3):1237–1260, 2010.

Stefan Volkwein. Model reduction using proper orthogonal decomposition. *Lecture Notes, Institute of Mathematics and Scientific Computing, University of Graz*. see <http://www.uni-graz.at/imawww/volkwein/POD.pdf>, 1025, 2011.