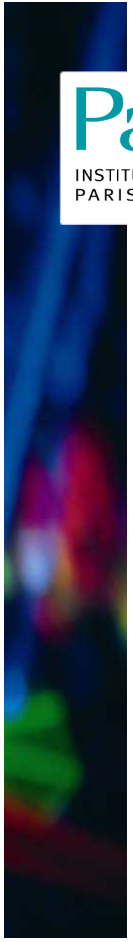
Reduced-order models for linear dynamics and control in aerodynamics

Grégory Dergham



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUT OF TECHNOLOGY

ARTS ET MÉTIERS ParisTech

2011-ENAM-0023

École doctorale n° 432 : Sciences des Métiers de l'Ingénieur

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École Nationale Supérieure d'Arts et Métiers

Spécialité " Mécanique "

présentée et soutenue publiquement par

Grégory DERGHAM

le 23 Juin 2011

Reduced-order models for linear dynamics and control in aerodynamics

Directeur de thèse : **Alain LERAT**
Co-encadrement de la thèse : **Jean-Christophe ROBINET & Denis SIPP**

Jury

M. José Eduardo WESFREID, Directeur de Recherche, PMMH, ESPCI	Président
M. Uwe EHRENSTEIN, Professeur, IRPHE, Université de Provence	Rapporteur
M. Bernd NOACK, Directeur de Recherche, Institut P', CNRS	Rapporteur
M. Peter SCHMID, Directeur de Recherche, LadHyX, École Polytechnique	Examinateur
M. Jérôme HOEPFFNER, Maître de Conférences, IJLDA, Université Pierre et Marie Curie	Examinateur
M. Alain LERAT, Professeur, DynFluid, Arts et Métiers ParisTech	Examinateur
M. Jean-Christophe ROBINET, Maître de Conférences, DynFluid, Arts et Métiers ParisTech	Examinateur
M. Denis SIPP, Maître de Recherche, DAFE, ONERA	Examinateur

T H È S E

Arts et Métiers ParisTech - Centre de Paris
Laboratoire DynFluid

Acknowledgments

First of all I would like to express my gratitude to Professor Alain Lerat, chair of the DynFluid laboratory at Arts et Métiers ParisTech, and to Professor Laurent Jacquin, director of the DAFE at ONERA, for making possible the cooperation between the two laboratories in which the thesis has been carried out. I also would like to express my gratitude to the members of my jury, Professors Uwe Ehrenstein, Bernd Noack, José Eduardo Wesfreid, Peter Schmid and Doctor Jérôme Hoepffner for having accepted to study my manuscript and for their valuable suggestions and constructive criticisms.

My most sincere appreciation to Jean-Christophe Robinet, my supervisor at DynFluid, whose long and fruitful discussions have provided me valuable suggestions, constructive advices, motivation to pursue my research, and mostly for having supported me all along the thesis. Special thanks to Denis Sipp, my supervisor at ONERA, who closely followed my thesis work at DAFE, giving me many helpful suggestions, important advices and constant encouragement during the course of this work. Without your common agreement, Jean-Christophe and Denis, the collaboration between Dynfluid and DAFE would not have been possible, and for that I am so grateful to you.
I also thank José Eduardo Wesfreid, Jean-Luc Aider and Thomas Duriez for the exchange of so much fruitful information on their experimental works at the PMMH laboratory at ESPCI. Sincere gratitude is due to Vincent Brion, Olivier marquet and Alexandre Barbagallo for taking intense academic interest in this study as well as providing valuable suggestions that improved the quality of this thesis.

I express my sincere thanks to all the members at DynFluid and DAFE who welcomed me in the kindest way. Most of all, special thanks go to Guillaume, Jean-Yves, Jérèmie, Olivier and Alexandre for the many fun times, endless discussions on life, who contributed to keep me in a good mood with my thesis work. I also thank my workmates Rocio, Valérie, Laurence and Clément for their support during the painful moments, especially at the end of the manuscript writing.
Mostly, I am grateful to all my family and friends who always encouraged me, for all the good times we spent together. Without the continuous help and encouragement you offered me during these years, this thesis would not have been the same.

Contents

Introduction

Motivation: separated flows in aerodynamics

Separated flows are commonly encountered in aerodynamics as soon as the flow over an obstacle reaches a sufficiently high Reynolds number. Flow separation occurs when a boundary layer is subjected to either a drastic change of the wall curvature or a large enough pressure gradient. In such cases, the shear stress is reduced to zero where separation occurs; the flow becomes detached and exhibits eddies or vortices. This phenomenon intervenes in a majority of industrial applications and is well-known to induce large modifications of the aerodynamic forces acting on vehicles owing to the low-pressure levels within the recirculation. For instance, in automobile applications, the massive recirculation region at the rear of a vehicle is responsible for the predominant part of its drag, see figure 1 (top). Another interesting example concerns the flows around space launchers which exhibit a massive separation occurring at the abrupt change in the geometry of their first stage. In both these cases, flow separation leads to a dramatic increase of the base drag and may represent up to 70% of the total drag. Alternatively, the recirculation bubbles observed over airfoils at high angles of attack result in an increase of the pressure drag and in the worst case in a sudden loss of lift and stall, all of which being undesirable, see figure 1 (bottom). As a result, much effort and research on aerodynamics has gone into the design of surfaces which keep the boundary layer attached as long as possible and delay its separation.

Another essential feature of separated flows is their ability to shed large vortices. In fact, separated boundary layers are often prone to hydrodynamic instabilities that amplify perturbations, leading to unsteadiness and sometimes to turbulence. What particularly interests us in this thesis is the low-frequency shedding of large-scale vortices. Unsteady flow structures characterized by spatial scales of the order of those of the studied object are commonly encountered in aerodynamics. For turbulent flows, such dynamics are associated with the largest flow scales within the Kolmogorov energy cascade and are associated with a production mechanism of the turbulent kinetic energy. Meanwhile, similar low-frequency unsteadiness are equally observed in low Reynolds number and laminar flows, which makes them rather persistent phenomena in aerodynamics, see figure 2. The occurrence of these unsteadiness appears in a wide variety of industrial applications and are usually detrimental to a satisfactory operation, which can be illustrated by some famous examples:
The separated flow over a circular cylinder is a simple prototype of bluff body flow which may account for the flow past a bridge, an undercarriage or a submarine pipe. This flow is typically observed to self-sustains a Bénard-von Kármán vortex street, see figure 3 (left). The shedding of these large vortices is often undesirable as it results in a significant increase of the mean drag and in unsteady lateral loads, see Choi *et al.* (2008). Worst, if the vortex shedding reaches the resonance frequency of the structure, it could cause severe structural failures. Interestingly, this phenomenon may also be observed in very high Reynolds number atmospheric flows such as the flow over an island depicted in figure 3 (right). Axisymmetric afterbody flows also stand for an interesting example. When passing to the transonic regime, space launchers such as Ariane V are subjected to strong vibrations. Such vibrations are generated by the low-frequency periodic shedding of large-scale vortices resulting from the self-interaction of the axisymmetric separated shear-layer. As a result, the massive separated area generates

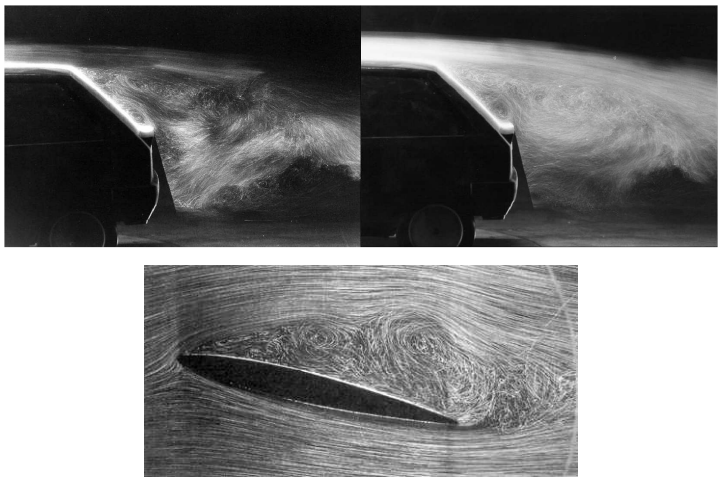


Figure 1: Top: separated flow behind a land vehicle on a wind tunnel. Taken from Aider (2008). Bottom: separated flow over an inclined airfoil. ONERA picture, H. Werlé.

high fluctuating lateral loads which are particularly harmful for the payload, see Deprés *et al.* (2004). The flow over a backward-facing step or a cavity displays an elongated recirculation bubble which is characterized by a single shear layer. This latter may be unstable to one of the most famous shear instability: the Kelvin-Helmholtz instability. In such cases, the rolling-up of the vorticity sheet into well-defined periodic vortices is observed, which generates large-scale flow structures in the downstream of the recirculation zone, see figure 4. This phenomenon is typically encountered in diffuser flows or also in the curved air intakes of modern military airplanes where the vortical structures are known to impact the compressor and decrease the performance of the engine. As well, cavity flows are famous for being the site of violent unsteadiness. The large vortical structures generated from the lip of the cavity travel along the shear layer and hit the downstream edge of the cavity, generating powerful pressure waves, see Rowley *et al.* (2002). For instance, this phenomenon is well-known to occur on flows over bomb bays, open sunroofs or multi-element wings. In the worst cases, cavity flow instabilities may cause severe structural vibrations which increase the cost of vehicles maintenance, or decrease their lifetime, and propagate extensive noise pollution over long distances.
A last example concerns the transonic flow over an airfoil. At the cruise conditions of an airplane, the flow acceleration on the suction side of the wing creates a stationary shock which induces the boundary layer separation. Then, for some Mach numbers and angles of attack, the shock and the recirculation bubble may oscillate at a low-frequency, see Jacquin *et al.* (2009). This phenomenon called buffeting is dangerous for the airplane safety and reduces the flight envelope of commercial airplanes.

For all these reasons, the industrial stakes associated with separated flows are considerably high. Notably, it is of utmost importance to understand, predict and control the onset of these low-frequency unsteadiness in flows. Meanwhile, as soon as complex flows are considered, a parametric analysis to design optimal control devices or to optimize some parameters quickly becomes not tractable with the current computational capabilities. It is thus pivotal to be able to model their dynamics with the least possible number of degrees of freedom so as to faithfully reproduce their essential features. This task, also referred to as Model Reduction, precisely motivated the present research.

Origin of the present work

This thesis results from a collaboration between the DynFluid laboratory of Arts et Métiers ParisTech and the Fundamental and Experimental Aerodynamics Department (DAFE) of ONERA. This research takes place within the framework of a common project of these laboratories aiming at modeling the

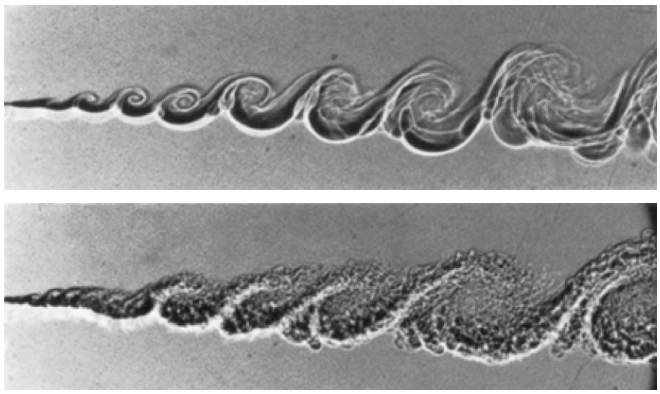


Figure 2: Visualization of planar shear layers in a laminar regime (top) and a turbulent regime (bottom). After Brown & Roshko (1974).

dynamics of separated flows in view of their optimal control. It has been financed for three years by the French Délégation Générale pour l'Armement (DGA).

Originally, the DynFluid laboratory and the DAFE have spent several years in studying the dynamics of fluid flows. Both adopted the hydrodynamics stability theory to investigate, explain and predict some features of aerodynamic flows such as the onset of unsteadiness as well as the frequencies and length scales selected by the flows. To do so, a typical approach consists of studying the linear development of perturbations about a given base flow. The stability analysis then stands for investigating the time evolution of small perturbations, whose dynamics are linear, to assess whether they are amplified or damped. Such stability analyses are referred to as local when the base flows of interest have only one non-homogeneous direction in space. Local stability analyses are usually based on the assumption of a weak non-parallelism of the base flows. For instance, such analyses proved to successfully predict some features of the laminar-turbulent transition in boundary layer flows, shear layer flows or jets, see Schmid & Henningson (2001). Now, as soon as the base flow of interest has two or three non-homogeneous spatial directions, as in the case of separated flows, one may resort to a so-called global stability analysis, see Zebib (1987); Jackson (1987); Theofilis (2000). In this context, both the DynFluid laboratory and the DAFE have acquired valuable knowledge on the linear global stability analysis of fluid flows through several internships and theses. For instance, the global linear dynamics of a laminar incompressible attached and separated boundary layer was investigated in the theses by Alizard (2007) and Cherubini (2010). Additionally, a particular emphasis was given to the numerical methods to solve large linear global stability problems for laminar incompressible and compressible flows in the thesis by Merle (2009).

One of the primary missions of the DAFE in the next years is to understand the origin and control the low-frequency unsteadiness observed in several well-known flow prototypes. This includes the transonic buffeting flows over airfoils, cavities or afterbodies, the trailing wake vortices and the flows over curved air intakes, see figure 5 (left). For instance, the noise of a transonic cavity flow was successfully controlled by Illy (2005) by introducing a small control cylinder at the upstream edge of the cavity. Alternatively, a subsonic separated flow was controlled by means of vortex generators in the thesis by Gardarin (2009) in the S19Ch wind tunnel of ONERA, see figure 5 (right). In all these configurations, the flows are turbulent. Yet, the theoretical investigations of the DAFE have mainly concerned the low Reynolds number laminar-turbulent transition of these flows. This choice was motivated by: (i) the possibility to analyze more easily the low-frequency unsteadiness in these flows to extract physical information, (ii) the availability of tools which do not require to much heavy computations and (iii) the possibility to extend the considered methodologies from laminar to fully turbulent flows by augmenting the Navier-Stokes equations with one or more turbulence model equations. Furthermore, considering

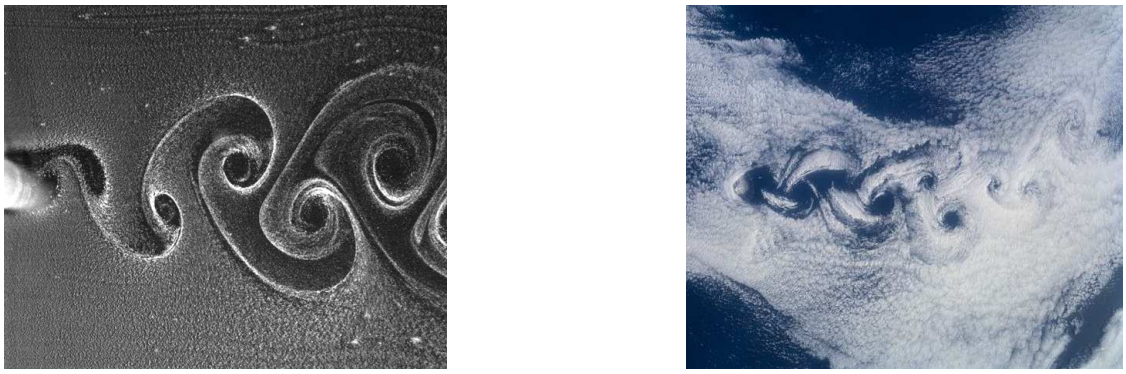


Figure 3: Left: Periodic vortex shedding pattern behind a circular cylinder in water at Reynolds number of 168. The visualization was done using hydrogen bubble technique. After Sanjay Kumar and George Laughlin; Department of Engineering, The University of Texas at Brownsville. Right: Von Kármán vortices off Rishiri Island, Japan. Satellite picture from NASA.

linearized equations offers the possibility to apply some of the modern control techniques such as the linear state space optimal control, see Bewley & Liu (1998); Bewley (2001); Kim (2003); Kim & Bewley (2007). In a general manner, control strategies may be classified into two categories. On the one hand, one refers open-loop or feedforward control where the actuators react in a predefined way to the flow system and to the actuation itself. Noticeably, these types of control have been applied at the DAFE in the theoretical investigations by Marquet (2007); Meliga (2008). On the other hand, closed-loop or feedback control refers to the configurations where the actuation is updated in real-time from some flow measurement so as to reach a given objective. Such a strategy was adopted in the theoretical and numerical thesis by Barbagallo *et al.* (2011). In both cases, the control was meant to suppress or at least reduce the perturbations whose dynamics were assumed to be linear.

Within the framework of these projects, the present thesis aims at finding efficient Reduced Order Models (ROMs) that faithfully capture the dynamics of flows. The idea is to provide a comprehensive study based on the physics of the flows to derive the ROMs. Indeed, these models are meant to preserve the original hydrodynamic stability properties of the flows in view of applying some optimal flow control strategies.

In this context, the distinction introduced by Huerre & Rossi (1998) between oscillator and noise-amplifier flows deserves a particular attention. Oscillator flows are characterized by a self-sustained beating at a very specific frequency. This mechanism is said to be intrinsic since no external disturbances are necessary to keep alive the oscillatory motion of the flow. The onset of an oscillator mechanism is usually well predicted by a global stability analysis through the onset of an unstable global mode. Typical example of oscillator mechanisms have been reported in cylinder flows (Noack & Eckelmann (1994)), backward-facing step flows (Barkley *et al.* (2002); Marquet *et al.* (2009)), the flow over a bump (Gallaire *et al.* (2007); Ehrenstein & Gallaire (2008*b*)) and cavity flows (Sipp & Lebedev (2007); Bres & Colonius (2008); Barbagallo *et al.* (2008)). Alternatively, some flows behave as a noise-amplifier which selectively amplifies in the downstream direction existing upstream noise. In these flows, the emergence of coherence structures is extrinsic and often characterized by a broad range of scales and frequencies. Some of the most famous amplifier flows are the boundary layers (Ehrenstein & Gallaire (2005); Alizard & Robinet (2007); Åkervik *et al.* (2008)), the jets and the backward-facing step flows (Marquet *et al.* (2008); Blackburn *et al.* (2008)).

In fact, the prediction and the control of amplifier flows appears as a very challenging task since their dynamics reflect both the amplifier behavior of the flow and the broadband nature of the upstream noise. For this reason, the physical analysis and model reduction of noise amplifier flows have been privileged in this thesis. More precisely, we considered the separated flow over the backward-facing step depicted

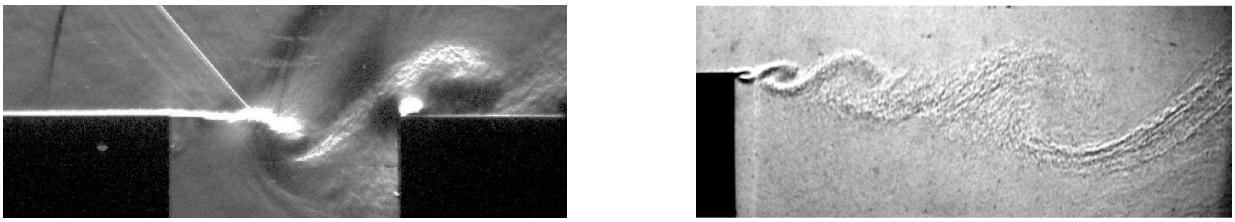


Figure 4: Left: cavity flow observed at the S8Ch wind tunnel of ONERA by strioscopy. After Illy (2005). Right: backward-facing step flow visualized by strioscopy.

in figure 6. This configuration was chosen since the backward-facing step flow accounts for a well-known prototype of amplifier flow where a substantial body of literature exists, such as the experimental works by Armaly *et al.* (1983); Tihon *et al.* (2001); Beaudoin *et al.* (2004); Lee *et al.* (2004) or the numerical investigations by Kim & Moin (1985); Kaiktsis *et al.* (1996); Le *et al.* (1997); Kaiktsis & Monkewitz (2003); Lesieur *et al.* (2003); Aider *et al.* (2007). Additionally, the low Reynolds number flow over this particular step geometry has been investigated in the experimental thesis by Duriez (2009) at the PMMH laboratory. In this latter work, the step shape was chosen circular to allow the displacement of the separation point and mimic some realistic configurations such as the flow over a curved airfoil, the expansion of a diffuser or the rear part of a land vehicle. Notably, the originality of this experimental work was to study the dynamics and to control the low Reynolds number laminar-turbulent transition by means of vortex generators. It is hoped that the present research will contribute to the future experimental investigations on this setup that still exists at the PMMH laboratory.

Objective of this thesis

The objective of this thesis is to design efficient reduced-order models that account for the linear dynamics of large fluid systems. Our procedures are derived on the two-dimensional incompressible flow over the backward-facing step considered in the work by Duriez (2009). Meanwhile, the methodologies investigated here are meant to be applied to any other incompressible open flow such as, for instance, other separated flows, wake flows, boundary layers or vortex flows. The first goal is to find efficient model reduction techniques. Next, a second objective is to illustrate how the use of reduced-order models allows to design an optimal flow control setup.

Brief review on model reduction

Model reduction appears as the essential element of this work and some of its concepts are introduced here. The modeling of complex physical phenomena by means of numerical simulations often requires to solve large-scale dynamical systems. Model reduction consists of constructing a simpler model having a much lower number of degrees of freedom which preserves some properties of the original system. Model reduction has been used for a long time; one should remember the famous dynamical system introduced by Lorenz (1963) to describe the ocean-atmosphere interactions by a system having three degrees of freedom. Since then model reduction has been used in a wide area of applications including oceanography, psychology, biology, probability, statistics or mechanics.
In this thesis, our purpose is to design reduced-order models for aerodynamic applications. Predicting the response of fluid systems subject to a large number of conditions or excitations is often not tractable owing to the complexity of the partial differential equations involved. Hence, model reduction has more recently appeared as a prolific and promising subject in the fluid mechanics community.

There are plenty of model reduction techniques, each of which using the different properties of the original dynamical system. Some of the most famous ones deserve to be mentioned here.

Figure 5: Left: laminar separation from a curved wall in a water tunnel. The flow is visualized by air bubbles and the Reynolds number based on the leading edge (not shown) is 20 000. ONERA picture, H. Werlé (1974). Right: Separated flow observed at the S19Ch wind tunnel of ONERA (2005). The flow is visualized by laser tomoscopy.

For linear systems, ROMs may be obtained by projecting the original equations onto the leading eigenmodes of the dynamical operator (Dowell *et al.* (1997); Åkervik *et al.* (2007)). Such an approach is also referred to as Modal Truncation. For stable linear systems, several techniques based on an input-output formulation of the equations have been introduced. This includes the Balanced Truncation procedure (Moore (1981)), the Krylov approximation methods which are based on the moment matching of the impulse response (Antoulas & Sorensen (2001); Antoulas (2005)) and comprise the Lanczos (Lanczos (1950)) and the Arnoldi procedures (Arnoldi (1951)). One may also refer to the system identification techniques (Ljung (1999)) where the coefficients of the ROMs are identified from some measurements. Some applications of system identification may be found on an oscillating $3D$ delta wing configuration (Tang *et al.* (2001)) or on a cavity flow, see Rowley & Williams (2006). The identification of the coefficients may be computed from the AutoRegressive with eXogenous input technique (ARX), see Huang & Kim (2008), or also the Eigensystem Realization Algorithm (ERA), see Ma *et al.* (2010).
As far as nonlinear systems are considered, the Proper Orthogonal Decomposition (POD) appears as a famous candidate. It consists in projecting the original equations onto a reduced set of dominant, in the energy sense, spatial modes (Aubry *et al.* (1988); Berkooz *et al.* (1993)). An extension was proposed to account for both the spatial modes and their time coefficients in an optimal way. The modes resulting from this technique are called Principal Interaction Patterns (PIPs) in the general case and Principal Oscillation Patterns (POPs) for linear systems, see Hasselmann (1988). Such a model reduction approach was originally developed to study climatic phenomena, see Storch *et al.* (1995). In the case of systems characterized by periodic behaviors, the harmonic balance technique (Hall *et al.* (2002)) provides a very interesting approach since it is based on a Fourier decomposition of the equations. Lastly, one may also refer the A Priori Hyper-Reduction method (Ryckelynck (2002, 2005)) where the original equations are projected onto a set of modes which are derived from an iterative process.

Obviously, there exist many additional model reduction techniques. Owing to the substantial body of literature on the subject, all of them will not be thoroughly detailed in this manuscript. We refer interested readers to the thesis by Placzek (2009) where some of the most recent model reduction procedures used in aerodynamics are reported in detail.

We now give some additional emphasis to the methods which will be considered in this thesis. These are the Modal Truncation, the Proper Orthogonal Decomposition and the Balanced Truncation. All three belong to the category of the so-called projection techniques. In other words, the ROMs are obtained by projecting the original high-dimensional dynamical systems arising from the discretization of the equations governing the fluid motion onto a reduced set of vectors. Such a set comprises flow structures called modes which span the subspace onto which the dynamics are projected. This procedure is called the Galerkin projection when the projection basis is orthogonal and the Petrov-Galerkin projection in the more general case of a bi-orthogonal projection. The projection finally results in a low-order

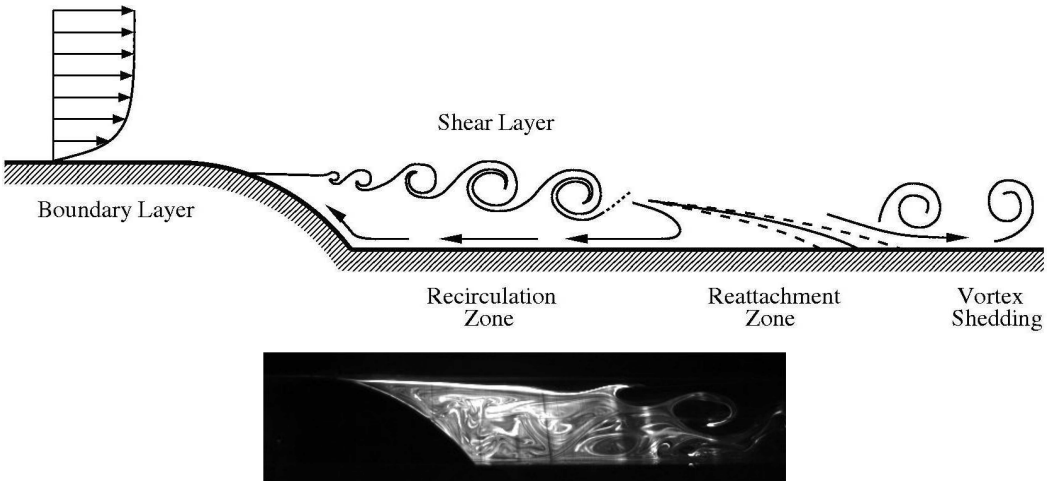


Figure 6: Rounded backward-facing step flow investigated here. Top: sketch of the flow configuration. Bottom: separated flow obtained by Duriez (2009) in a water tunnel at ESPCI (École Supérieure de Physique et de Chimie Industrielles). The flow is visualized by laser induced fluorescence at the Reynolds number of 2800.

dynamical system whose dimension is equal to the number modes in the projection basis. Projection methods may be applied to both linear or nonlinear dynamical systems. Yet, as previously mentioned, the dynamics investigated in this thesis are linear, which means that both the original and the resulting low-order dynamical systems are linear. Naturally, the projection subspace has a central influence on the dynamics returned by the low-order model. In fact, the ROMs are conceptually meant to recover the dynamics of the flow structures used in the projection process. Their choice is thus of pivotal importance to accurately capture the original flow dynamics. The three different projection bases considered in this work are exposed in more detail below.

Modal Truncation

The modal truncation consists of projecting the high-dimensional dynamical operator onto a reduced set of its eigenmodes or global modes. Obviously, the identification of these eigenmodes is only possible for linear and deterministic dynamical systems. Interestingly, each eigenvector is associated with an eigenvalue, which is complex in general, whose real part denotes the growth rate of the mode and whose imaginary part denotes its oscillating frequency. Notably, the modes having a positive real part are unstable and lead to a divergence of the linear dynamics. The global modes stand for the most natural solution of the original equations and they may be ranked according to their frequency and growth rate, which make them a valuable candidate to construct ROMs. The model reduction is obtained by expressing the perturbations into the modal basis and truncating the undesired modes. In practice, the most stable modes, which only contribute to the short-time dynamics, are the ones to be truncated. By this way, Modal Truncation preserves the most unstable part of the original dynamical operator eigenspectrum, and thus its original stability.

In the case where the dynamical operator is normal, namely where its eigenmodes are orthogonal, Modal Truncation appears as an optimal methods to model the dynamics, see Farrell & Ioannou (2001*a*). For this reason it has been extensively used in structure mechanics owing to the normality of the governing equations. Now, regarding fluid mechanics, the governing equations are non-normal and the global modes may provide a very ill-conditioned basis, see Cossu & Chomaz (1997). As a result, their efficiency to construct ROMs is more questionable; it is at least sub-optimal, see Farrell & Ioannou (2001*a*). However, recent articles have reported how the use of a reduced set of global modes allows to efficiently model the transient dynamics of a boundary layer (Åkervik *et al.* (2008)), of the flow past a bump (Ehrenstein

& Gallaire (2008*b*)). Additionally, the modeling and feedback control of a cavity flow has also been reported by Åkervik *et al.* (2007).

Proper Orthogonal Decomposition

The Proper Orthogonal Decomposition, also called POD, is known as one of the most popular model reduction procedure in fluid mechanics. It consists of identifying the set of modes (the POD modes) which optimally reconstruct a given dataset of flow snapshots. Namely, the energy-norm based error between the dataset projected onto the POD modes and the original dataset is minimal. Reduced-order models are then obtained by projecting the high-dimensional dynamical system onto a reduced set of POD modes. Interestingly, the dataset may be taken from either numerical simulations or experiments which thus comprises both linear and nonlinear flow dynamics.
The POD modes may be computed as the eigenvectors of the temporal correlation matrix between the different flow snapshots. They result in an orthogonal basis which most effectively accounts for the subspace spanned by the dataset of interest. The associated POD eigenvalues are positive and ranked according to the fraction of energy captured by the associated modes. As shown by Holmes *et al.* (1997), only a finite number of non-zero POD eigenvalues exists. Mostly, the POD eigenspectrum often fall off quite rapidly, indicating that the leading POD modes capture the principal flow components of the set of snapshots. Owing to this valuable property, the POD has been extensively used in the fluid mechanics community. Originally, POD has been used to identify the coherent structures in turbulent flows, see Lumley (1970); Sirovich (1987); Berkooz *et al.* (1993). Later, the dynamics of coherent structures in the wall region of a turbulent boundary layer has been investigated by Aubry *et al.* (1988) through a reduced-order model of the Navier-Stokes equations obtained by POD. On the other hand, model reduction based on POD does not necessarily preserve the stability of the original dynamical system (Ma & Karniadakis (2002)). For this reason, a considerable effort has been devoted to improve the stability of the POD low-order models. This includes for instance the calibration of the model coefficients (Galletti *et al.* (2004)) but also the addition of a shift mode (Noack *et al.* (2003)), a pressure term (Noack *et al.* (2005)) or a spectral viscosity (Sirisup & Karniadakis (2004)). Regarding flow control, recent advances have been made to include the effects of an action on the POD models, we refer interested readers to the works by Galletti *et al.* (2007); Weller *et al.* (2009*b*); Luchtenburg *et al.* (2009). In particular, successful flow controls based on POD models have been reported in the numerical works by Bergmann *et al.* (2005); Bergmann & Cordier (2008); Weller *et al.* (2009*a*) on cylinder flows. One may also mention the experimental control of a subsonic cavity flow by Samimy *et al.* (2007) where a POD model was obtained from Particle Image Velocimetry (PIV) data.

Balanced Truncation

The last technique considered in this thesis is called Balanced Truncation. It has been originally introduced by Moore (1981) as a model reduction procedure for stable linear input-output systems. Balanced Truncation is meant to capture the dynamics between some inputs, such as actuators, to some outputs, such as sensors. Furthermore, this method displays very attractive features. First, it offers error bounds on the input-output transfer function that are close to the lowest error possible from any reduced-order model (Dullerud & Paganini (2000)). In addition, it is shown to preserve the stability of the original input-output dynamical system. For these reasons, Balanced Truncation clearly appears as a very powerful technique in the control theory community.
Balanced Truncation is based on the concepts of controllability and observability. By definition, controllability refers to the ability of a state to be influenced by the inputs while observability refers to the capacity of a state to be measured by the outputs. The key idea of Balanced Truncation is then is retain the modes which are both highly controllable and observable, called balanced modes, and use them as a projection basis. Technically, the controllability and observability of given states may be measured from two bilinear operators: the so-called controllability and observability Gramians. The balanced modes

then reduce to the leading eigenvectors of the product of the two Gramians, see Moore (1981); Antoulas (2005). Each Gramian may be computed by solving a Lyapunov equation, which is unfortunately not tractable for large systems. This limitation has first restricted the range of Balanced Truncation to rather simple flow configurations, see Joshi *et al.* (1997); Cortelezzi & Speyer (1998); Farrell & Ioannou (2001*a*); Högberg *et al.* (2003). Yet, a recent breakthrough (Willcox & Peraire (2002); Rowley (2005)) allowed to approximate Balanced Truncation for large systems. It consists of approximating the balanced modes by using a snapshot technique, similarly to a POD procedure. In fact, it is shown by Rowley (2005) that Balanced Truncation reduces to a special case of POD which uses a particular dataset (the impulse responses from the inputs) and a particular inner product (the observability Gramian). Accordingly, the approximate procedure derived by Rowley (2005) has been called Balanced Proper Orthogonal Decomposition (BPOD). This latter has been successfully applied to model the input-output dynamics of various fluid flows, see Ilak & Rowley (2008); Barbagallo *et al.* (2009); Bagheri *et al.* (2009*b*); Ahuja & Rowley (2010).

Approach

This thesis is based on theoretical and numerical investigations. Our procedures are derived on the incompressible flow over the two-dimensional ($2D$) backward-facing step depicted in figure 6. Instead of investigating turbulent flows, which would have required much heavier computations, we only considered laminar flows. This choice is also appealing since laminar flows are less complex and better understood than turbulent flows. Consequently, the fluid motions are assumed to be governed by the incompressible Navier-Stokes equations.

As a first step, the linear dynamics of the backward-facing step flow are investigated. In particular, we investigate the dynamics of three-dimensional ($3D$) perturbations linearized about a $2D$ steady solution called the base flow. A global linear stability analysis is derived to assess both the long and short time dynamics of the flow. Next, these results are used to provide the non-reduced linear dynamics.

As a second step the model reduction of the $2D$ linear dynamics of the flow is considered. The three previously introduced projection methods are considered: Modal Truncation, POD and Balanced Truncation. We first study the ability of the least stable global modes to recover the original transient dynamics of the flow. Next, an input-output formulation of the dynamics is adopted to derive low-order models based on POD and Balanced Truncation. Following the BPOD procedure exposed by Willcox & Peraire (2002); Rowley (2005), a snapshot method is derived in the frequency domain to approximate the leading balanced modes. A similar snapshot method is also used to compute both the most controllable and the most observable modes by a POD procedure. The performances of both POD and BPOD models in capturing the input-output dynamics are evaluated and compared. Within this part of the thesis, an important guideline is to successively increase the number of inputs and outputs. As a first step, the ROMs are constructed by considering a single input (actuator) and a single output (sensor). Next, we consider the modeling of the dynamics from all the possible inputs (initial conditions) to the same output (sensor). Lastly, we propose an extension which allows to handle the reduction of large systems where both the inputs and the outputs reduce to the full state space. In this latter case, the POD and balanced modes are approximated by truncating the singular value decomposition of the resolvent operator. Additionally, the most controllable and observable modes are shown to account for the so-called Empirical Orthogonal Functions (EOFs) and Stochastic Optimals (SOs) originally introduced in the analysis of synoptic meteorological data (Lorenz (1956); North (1984)) and also in turbulence modeling (Lumley (1967, 1970)). A particular attention is given to the physical interpretation of the leading EOFs and SOs as the predominant coherent and uncorrelated process of the uncertain flow resulting from a permanent stochastic excitation. The leading EOFs, SOs and balanced modes are also shown to capture the entire flow response from all possible initial conditions, namely the entire flow dynamics.

Finally, the third part of the thesis provides a practical application where a ROM based on POD allows to perform an optimal flow control. To this end, we adopted the LQG (Linear Quadratic Gaussian) closed-loop control to minimize the perturbations of the backward-facing step flow.

Organization of the manuscript

The body of this dissertation is composed of 5 chapters. In the first part of the thesis, chapter 1 presents an overview of the backward facing step flow linear dynamics. In particular, the non-reduced transient dynamics of the flow are exposed. The analysis and discussion on the ability of the global modes to model the $2D$ transient dynamics of the flow is reported in appendix B as an extension of this chapter.

The following part of the thesis is dedicated to the model reduction of the $2D$ input-output linear dynamics of the backward-facing step flow by means of POD and Balanced Truncation. Its comprises three chapters. Chapter 2 first introduces the snapshot technique derived in the frequency domain which allows to compute the modes. The resulting models derived in the case of a single input and a single output are also investigated. Chapter 3 provides an extension of the previous snapshot technique to configurations having a very large number of inputs. Lastly, chapter 4 deals with the modeling of the dynamics from all possible inputs to all possible outputs. It includes a receptivity analysis of the flow based on the singular value decomposition of the resolvent operator as well as a physical interpretation of the leading EOFs and SOs. In addition, the performances of the resulting models are also exposed and assessed.

The question of flow control is ultimately addressed in chapter 5 where the LQG closed-loop control of the backward-facing step flow by means of a reduced compensator is investigated. Since this work was done in collaboration with Dr. Alexandre Barbagallo and Pr. Peter J. Schmid, only a presentation of the flow configuration and a summary of the main results are exposed in the chapter. Yet, the full article resulting from this collaboration is exposed in the appendix section.

Next, we provide a general conclusion on the research exposed in this manuscript. It includes a summary of the major contributions, a discussion of the results as well as recommendations for future investigations. Lastly, it should be mentioned that only few information on the numerical techniques is exposed in the body of the manuscript. However, a review of all the numerical methods and tools has been included in appendix A.

Conferences and journal articles

1. G. Dergham, V. Brion, J.-C. Robinet and D. Sipp
The use of global modes to compute optimal transient growth
7^{th} EUROMECH Fluid Mechanics Conference, September 2008, Manchester

2. G. Dergham, D. Sipp and J.-C. Robinet
Model reduction using Balanced POD with frequential snapshots
IUTAM Symposium on Laminar-Turbulent Transition, June 2009, Stockholm

3. G. Dergham, D. Sipp and J.-C. Robinet
Etude de la dynamique transitoire d'un écoulement de rampe par modèle réduit
Congrès Français de Mécanique, August 2009, Marseille

4. G. Dergham, A. Barbagallo, D. Sipp, J.-C. Robinet and P. J. Schmid
Model reduction and closed-loop control of a convectively unstable flow
8^{th} EUROMECH Fluid Mechanics Conference, September 2010, Bad Reichenhall

5. G. Dergham, D. Sipp, J.-C. Robinet and A. Barbagallo
Model reduction for fluids using frequential snapshots
Accepted for publication in Phys. Fluids (2011)

6. G. Dergham, D. Sipp and J.-C. Robinet
Accurate low dimensional models for deterministic fluid systems driven by uncertain forcing
Accepted for publication in Phys. Fluids (2011)

7. G. Dergham, J.-C. Robinet and D. Sipp
Receptivity of a backward-facing step flow to stochastic perturbations: dynamics decomposition and model reduction
Submitted to J. Fluid Mech. (2011)

8. A. Barbagallo, G. Dergham, D. Sipp, P.J. Schmid and J.-C. Robinet
Closed-loop control of unsteadiness over a rounded backward-facing step
Submitted to J. Fluid Mech. (2011)

Chapter 1

Global stability analysis of the backward-facing step flow

1.1 Introduction

The flow over a backward-facing step stands for an important prototype for studying the effects of flow separation resulting from an abrupt change of geometry. Notably, numerous studies were performed on its dynamics since the thirties, see Goldstein *et al.* (1970); Denham & Patrick (1974); Armaly *et al.* (1983); Adams & Johnston (1988). It is indeed both commonly encountered in many engineering applications which exhibit strong flow separations and used as a reference open flow to perform fundamental studies for flow control, see Chun & Sung (1996). Even if it may be seen as a very simple flow configuration, the understanding of its transition to turbulence is far from being understood. One may refer to the turbulence analysis over separated flows by Le *et al.* (1997). In particular, it also stands for a typical setting in which to understand the hydrodynamic instabilities in separated flows. One may refer to numerous works such as those by Kaiktsis *et al.* (1991); Kaiktsis (1995); Kaiktsis *et al.* (1996); Kaiktsis & Monkewitz (2003); Gresho *et al.* (1993); Fortin *et al.* (1997); Barkley *et al.* (2002); Tylli (2003); Beaudoin *et al.* (2004). In this context, we investigate in this chapter the transition to turbulence of a $2D$ laminar backward facing step flow by numerical means. More precisely, we introduce and analyze the linear dynamics of the flow which will be considered throughout the rest of the manuscript. We proceed by investigating the dynamics of small perturbations about a base flow in order to outline some of its predominant instability mechanisms and possible transition scenarios.

Let us give some insights on the backward-facing step flow dynamics. We have depicted in figure 1.1 a sketch of the flow configuration considered here. It comprises an incoming boundary layer of

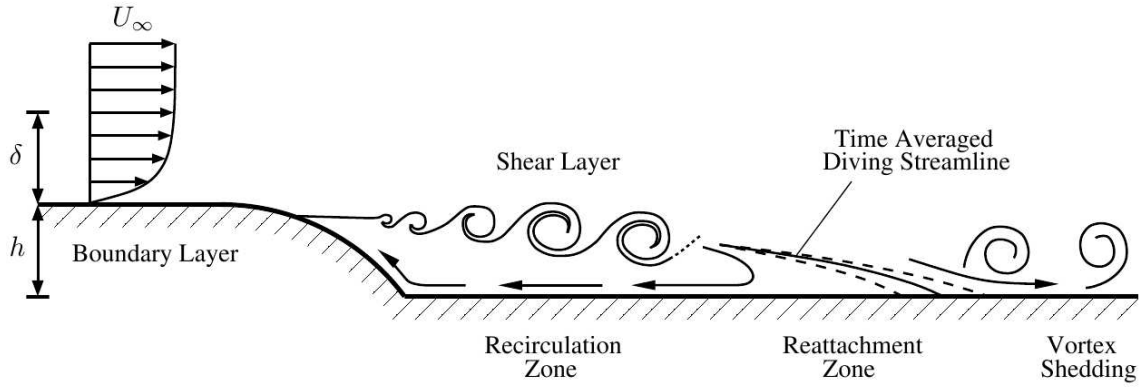


Figure 1.1: Sketch of the backward-facing step flow investigated here.

thickness δ and a step of height h. The flow separates due to the pressure gradient induced by the wall curvature and reattaches farther downstream. The low Reynolds number dynamics are mainly driven by several hydrodynamic instabilities. An important origin of unsteadiness reduce to the inflectional velocity profile in the shear layer responsible for the well known Kelvin Helmholtz instability which often leads to the transition to turbulence. This local instability have a convective nature and leads to small-scale structures. For instance, it yields a typical Strouhal number, based on the shear layer thickness δ_ω and the upstream velocity U_∞ of $St_{\delta_\omega} = 0.2$ (Driver *et al.* (1987)). A second mechanism, characterized by a three-dimensional and stationary structure (Williams & Baker (1997)), has been identified as an unstable global mode of the flow, see Beaudoin *et al.* (2004); Barkley *et al.* (2002); Cherubini *et al.* (2010*b*). This later seems to be related to a centrifugal instability associated with the reattaching streamlines curvature of the flow. Thirdly, a low-frequency unsteadiness characterized by much larger flow structures have also been reported. This phenomenon called as "flapping" results in an harmonic low-frequency displacement of the reattaching point Dogval *et al.* (1994); Ehrenstein & Gallaire (2008*a*); Cherubini (2010). This phenomenon seems to be driven by a more global length scale such as the recirculation height or length. The typical associated Strouhal numbers reported in the literature are observed to be one order of magnitude below those associated with the Kelvin-Helmholtz instability. For instance, the Strouhal number $St_{\delta_\omega} = 0.05$ has been reported by Driver *et al.* (1987). Finally, when increasing the Reynolds number, one may observe more complex and non-linear mechanisms mixing all these different dynamics (Duriez (2009)). Noticeably, one of them is the vortex merging which is characterized by half the frequency of the Kelvin-Helmholtz instability. As the flow becomes turbulent, the accurate identification of all the distinct physical mechanisms is still challenging. In our case, we restrict our analysis to a laminar and low Reynolds number flow and expect to recover some of these flow instabilities.

The geometry of the step, depicted in figure 1.1, comes from the experimental work by Duriez (2009). This later studied the dynamics and control of the flow in both wind and water tunnel experiments. In our case, we considered a $2D$ step (infinite in the spanwise direction) and without a top wall so as to avoid numerical difficulties. Yet, due to the confinement effects induced by the lateral and top walls, our goal is not to get an accurate comparison with the results by Duriez (2009). Instead, we expect to yield a complete study of the linear flow dynamics by using some of the most recent hydrodynamic stability tools. Though, the results presented here will be shown to be consistent with commonly known experimental results. The objective of this chapter is to present a thorough view of the linear mechanisms of the flow dynamics in view of its modeling and control, which is assessed in the next chapters.

The rest of the chapter is organized as follow: In section 1.2 we describe the flow configuration over the backward-facing step along with the governing equations and modeling assumptions. Next, the base flow considered for the stability analysis is presented in section 1.3. In particular, its dependence with the computational domain size and Reynolds number is examined. Then, we study in section 1.4 the long time stability analysis of the base flow through a modal decomposition analysis. Section 1.5 stands for the short time analysis where the ability of the flow to transiently amplify perturbations is assessed. Finally, section 1.6 provides a comparison between the different instability mechanisms and concluding remarks.

1.2 Flow configuration

1.2.1 Geometry

The backward-facing step considered here consists of a fraction of a circular cylinder linked with two flat plates. This geometry has first been investigated in the thesis by Duriez (2009) by experimental means. Figure 1.2(a) represents a sketch of the circular step. Denoting the step height by h, the radius of curvature is given by $R = \frac{5}{2}h$ and the step length is equal to $L = 2h$. The geometry is defined in a cartesian coordinate system (x, y, z), where coordinate x, y and z denote respectively the longitudinal,

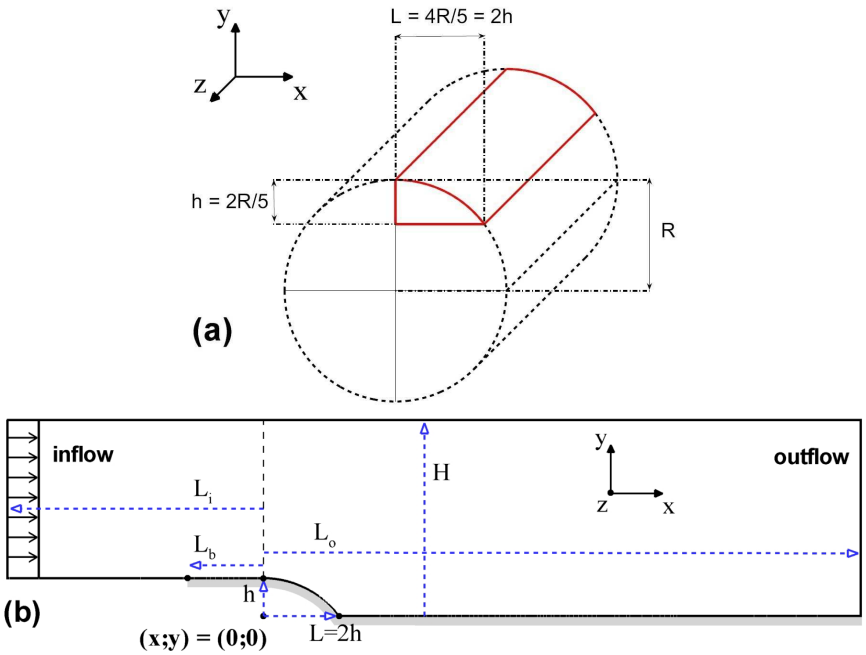


Figure 1.2: (a) Backward-facing step geometry (taken from Duriez (2009)). A sketch of the computational box and its dimensions is provided in (b).

vertical and transverse directions. The streamwise direction is denoted by x and the flow comes from the left, see figure 1.2(b). The step is considered as infinite in the spanwise direction z and without any wall at the top, i.e. in the free-stream.
The origin of the coordinate system is chosen such that the beginning of the step is at $(x = 0; y = h)$ and its ending at $(x = L; y = 0)$. Then, as represented in figure 1.2(b), we introduce the parameters L_i, L_o and H to define the size of the fluid domain. Furthermore, as a fluid flows from the upstream, an incoming boundary layer develops before reaching the step. In the upper flat-plate (at $y = 1$), a free-slip condition is considered for $(x < -L_b)$, whereas the no-slip boundary starts for $(x > -L_b)$. The introduction of this parameter enables to enforce the beginning of the boundary; it has been fixed to $L_b = 2h$ for a computational convenience.

1.2.2 Governing equations

A fluid of kinematic viscosity ν is supposed to enters the domain from the left at $(x = -L_i)$ with a uniform longitudinal velocity U_∞. The reference length and velocity scales are chosen as $h = 1$ and $U_\infty = 1$ in the following. We suppose that the fluid motion is governed by the incompressible Navier-Stokes equations 1.1 given by

$$\begin{cases} \partial_t \boldsymbol{u} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p + Re^{-1} \nabla^2 \boldsymbol{u} \\ \nabla \cdot \boldsymbol{u} = 0 \end{cases} \tag{1.1}$$

Where the vector $\boldsymbol{u} = (u, v, w)$ stands for the velocity field in the cartesian coordinate system, p is the pressure field, and the Reynolds number is defined as $Re = U_\infty h/\nu$.

Regarding the boundary conditions, the uniform velocity $(u = 1; v = 0)$ is prescribed at the inlet boundary $x = -L_i$. The free-slip condition with zero tangential stress $(\partial_y u = 0; v = 0)$ is prescribed on the boundary $(-L_i \leq x \leq -L_b; y = 1)$. The beginning of the no-slip boundary condition at $(x = -L_b; y = 1)$ generates a developing laminar boundary layer and the no-slip condition $(\boldsymbol{u} = 0)$ is imposed on all the remaining $(x > -L_b)$ lower boundary of the domain. At the upper boundary

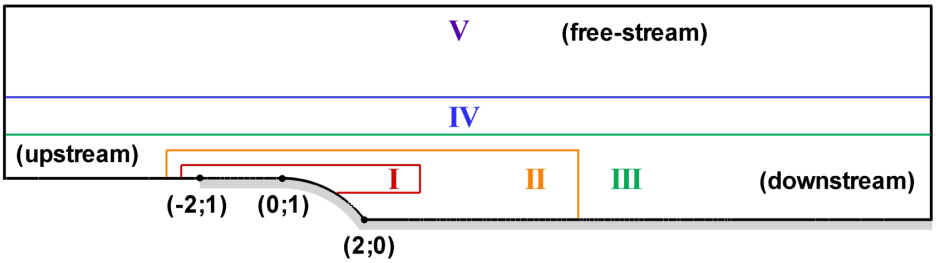


Figure 1.3: Structure of the mesh considered to discretize the fluid domain.

$y = H$, we enforce a symmetry condition ($\partial_y u = 0; v = 0$), which can also be interpreted as a uniform free-stream condition. Finally, a free outflow condition $p\boldsymbol{n} - Re^{-1}(\nabla \boldsymbol{u})\cdot\boldsymbol{n} = 0$ is prescribed at the outlet boundary $x = L_o$, where $\boldsymbol{n}$ stands for the outward normal unitary vector of the boundary.

1.2.3 Linear dynamics

This chapter is devoted to the analysis of the linear dynamics of the backward-facing step flow. It consists in studying the dynamical behavior of small perturbations that develop on a given base flow. Reminding that we study the low Reynolds number fluid mechanics over a transitional separated flow, it is natural to consider a base flow given by a steady solution of the Navier-Stokes equations, see Schmid & Henningson (2001). Next, given the two-dimensionality of the step, we look for a two-dimensional (2D) base flow (with $w = 0$). Even if a three-dimensional steady solution may exist for particular Reynolds numbers (this will be shown next), this choice is motivated since a 2D steady flow yields the solution having the maximum number of symmetry conditions compatible with the governing equations. Mostly, a 2D base flow can be computed in a much simpler way on a 2D mesh. The state flow (u, v, w, p) is decomposed as the sum of the base flow and a small perturbation:

$$(u, v, w, p) = (U, V, 0, P) + \epsilon(u', v', w', p') \tag{1.2}$$

The parameter ε points out that the perturbation is low compared to the base flow. By introducing this decomposition into the Navier-Stokes equations 1.1, and neglecting the ϵ^2 terms, we get at first order the equation governing the base flow:

$$\begin{cases} (\boldsymbol{U}\cdot\nabla)\boldsymbol{U} = -\nabla P + Re^{-1}\,\nabla^2\boldsymbol{U} \\ \nabla\cdot\boldsymbol{U} = 0 \end{cases} \tag{1.3}$$

which defines a 2D steady flow solution. Note that the vector $\boldsymbol{U} = (U, V)$ stand for the velocity field of the base flow. Next, at order ϵ, we recover the equations governing the linear dynamics of the perturbation developing over the base flow, namely the linearized Navier-Stokes equations:

$$\begin{cases} \partial_t\boldsymbol{u} + \boldsymbol{U}\cdot\nabla\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{U} = -\nabla p + Re^{-1}\nabla^2\boldsymbol{u} \\ \nabla\cdot\boldsymbol{u} = 0 \end{cases} \tag{1.4}$$

where we have omitted the primes over the disturbance terms for simplicity. In fact, we now denote by $\boldsymbol{u} = (u, v)$ and p the velocity and pressure fields of the perturbation in the rest of this chapter. The question to known what are the dynamics of such perturbations about the base flow stands for the objective of this chapter. Yet, before proceeding with the stability analysis, the base flow is first presented and investigated in the next section.

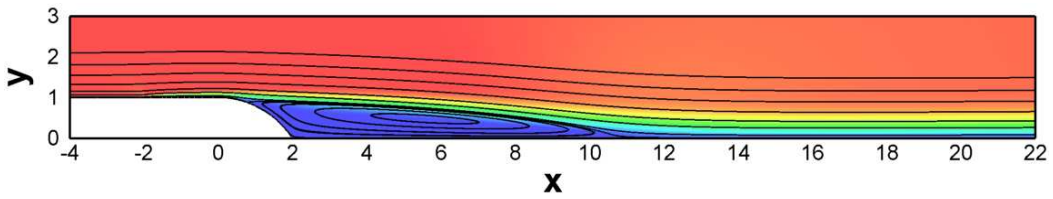


Figure 1.4: Base flow solution at $Re = 600$. The flow is depicted by both its longitudinal velocity and streamlines.

1.3 Base flow

1.3.1 Numerical methods and problem discretization

The 2D steady stationary Navier-Stokes equations 1.3 are solved in the computational domain presented in the previous section. Numerically, this problem is solved by a finite-element formulation using P2-P1 Hood-Taylor elements: six-node quadratic triangular elements with quadratic interpolation for velocities (P2) and three-node linear triangular elements for pressure (P1), see Ding & Kawahara (1999) for more details. The meshes and discrete matrices resulting from the variational formulation of equations 1.3 are generated with the software FreeFem++ (http ://www.freefem.org). The structure of the mesh adopted hereafter has been represented in figure 1.3. It is composed by five zones defined by a constant triangles density. The first zone, having the highest density, is chosen at the upstream in order to include the developing boundary layer and the beginning of the shear layer. This choice is motivated by anticipating that this region will comprise (i) the highest base flow shear and (ii) the smallest perturbation structures. The second zone is chosen to include the whole recirculation region. The third zone extends on the remaining upstream and downstream part of the domain. Finally, the fourth and fifth zone stand for the free-stream regions.

Practically, to compute the base flow, we proceed as in Barkley *et al.* (2002) where a time-dependent simulation of the two-dimensional Navier-Stokes equations is first used to get an approximate solution of the base flow. Next, a Newton iteration method is performed in order to solve the steady solution 1.3, starting the procedure with the approximate steady solution as a guess. This algorithm requires the inversion of non-symmetrical sparse matrices, which is carried out with by direct multifrontal sparse LU solver (MUMPS Amestoy *et al.* (2001)). Solution at larger Reynolds number may be easily followed by continuity; the solution at a smaller Reynolds number yields the new guess to apply the Newton algorithm. Note that this Newton method has been performed with the same finite element approach in the work of Sipp & Lebedev (2007).

1.3.2 Steady solution at $Re = 600$

We present here the base flow solution at $Re = 600$ obtained from the Newton method. Figure 1.4 depicts the streamlines and longitudinal velocity of the resulting steady flow. As expected, the boundary layer start to develop from $x = -2$ until it separates due to the pressure gradient effect induced by the geometry curvature. Beyond separation, an elongated recirculation region is generated. Farther downstream, the boundary layer reattaches the wall and develops until the end of the computational domain.
As mentioned in section 1.1, the backward-facing step flow dynamics may be characterized by two different types of length scales. The first ones, such as the step height or the recirculation length are called "global" whereas the second ones such as the boundary layer or shear layer thickness are said "local". Both depend on the Reynolds number and the starting position of the upstream boundary layer (fixed to $x = -2$ here). These quantities, computed on the base flows, are briefly investigated in the following.

Mesh	d_1	d_2	d_3	d_4	d_5	n_{tri}	n_{vert}	x_s	y_s	x_r	L_r
M_1	50	25	9	3	1	88 181	44 883	0.744	0.887	10.778	10.034
M_2	75	40	15	4	2	229 836	116 240	0.735	0.889	10.806	10.071

Table 1.1: Characteristic of the two meshes considered for the convergence study at $Re = 600$.

1.3.2.1 Recirculation length

A particular attention is paid to the recirculation length, which is used to check the convergence with respect to grid resolution. This choice is motivated since the reattachment length is known to be particularly sensitive to the numerical resolution. How to compute the recirculation length is illustrated in figure 1.5. By computing the skin friction coefficient on the wall, which is defined by

$$C_f(x) = -\frac{2}{Re}\boldsymbol{t}\cdot(\nabla\boldsymbol{U} + \nabla\boldsymbol{U}^t)\cdot\boldsymbol{n} \tag{1.5}$$

where $\boldsymbol{t}$ denotes the unit tangent vector oriented downstream to the wall. The set of x positions where the skin friction is negative defines the recirculation area. Additionally, the first coordinate where the skin friction becomes negative identifies the separation point at $(x_s; y_s)$. Similarly, the reattachment point $(x_r; y_r)$ is obtained from the position where C_f comes back to positive values. The recirculation bubble length L_r is then simply defined by

$$L_r = x_r - x_s \tag{1.6}$$

For the base flow at $Re = 600$, figure 1.5(a) depicts the separation and reattachment points from the computation of the wall skin friction, represented in 1.5(b). The boundary layer separates at $x \approx 0.7$ and reattaches at $x \approx 10.8$, so that $L_r \approx 10.1$.
The convergence with respect to grid resolution has been performed on the meshes denoted by M_1 and M_2 in table 1.1. Note that the five triangles densities (defined by the number of imposed triangle edges per unit length) have been reported in the columns d_1, d_2, d_3, d_4 and d_5. It is noteworthy that these two meshes have been designed in a computational domain such that $L_i = -20$, $L_o = 100$ and $H = 20$, which will be shown, farther in this section, to be sufficient to avoid any confinement effect.
These results highlight that a much finer mesh than M_1 recover the same recirculation length with a 3 digits accuracy. The mesh M_1 is thus considered as sufficiently refined to yield a converged base flow at $Re = 600$.

1.3.2.2 Boundary layer thickness

As a standard way of quantifying the boundary layer thickness, the displacement and momentum thicknesses are considered. The displacement thickness $\delta^*(x)$ is defined as the distance that the surface would have to displace outwards in an inviscid flow so as to yield the same mass flow rate deficit as that existing in the boundary layer. It is expressed by

$$\delta^*(x) = \int_{y_w}^{y_t}(1 - \frac{u}{u_0})dy \tag{1.7}$$

where y_w denotes the lower wall position, y_t stands from a height so that $u_0 = u(y_t)$ denotes a free-stream reference velocity. Usually, y_t is chosen at the maximum available y position so that $u_0 = U_\infty = 1$. However, in our case, the velocity profile may exceed 1 in the boundary layer due to the pressure gradient effects. Consequently, is it more relevant to choose y_t such that $u(y_t)$ is maximum. Since this maximum may not exists for particular x positions, we introduce a more general criterion: for a fixed x position, y_t is chosen such that

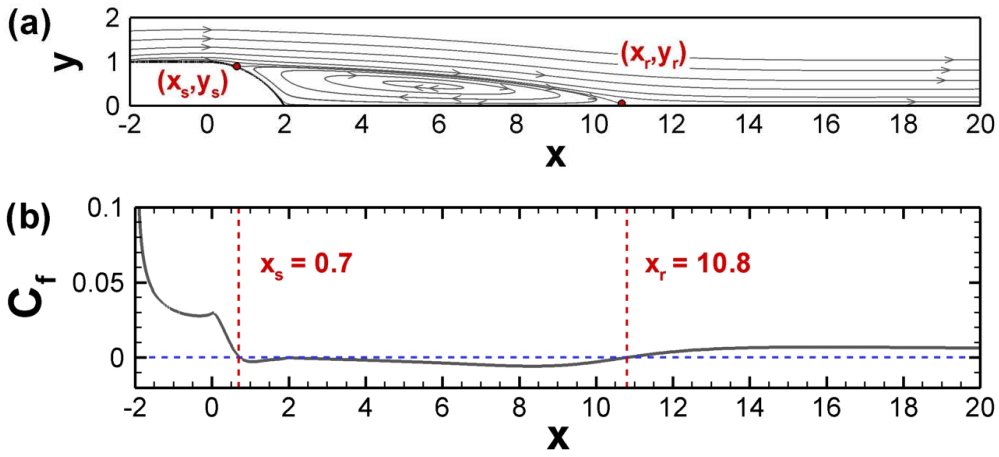


Figure 1.5: (a) Base flow streamlines at $Re = 600$ together with the separation and reattachment points. (b) Skin friction computed on the wall.

$$\partial_y u(y_t) = \frac{1}{100} \max_y \partial_y u \tag{1.8}$$

In fact, the idea is to find the y position at which $\partial_y u \approx 0$. The factor $\frac{1}{100}$ is thus arbitrary, indeed, choosing another relevant factor such as $\frac{1}{1000}$ does not modify the results in our case.
Analogously, we define the momentum thickness $\theta(x)$ by the distance that the wall would have to displace outwards, in an inviscid flow, so as to recover the equal momentum deficit of the mass flow rate. It is computed by a similar formula as

$$\theta(x) = \int_{y_w}^{y_t} \frac{u}{u_0}(1 - \frac{u}{u_0})dy \tag{1.9}$$

Lastly, it is also interesting to compute the shape factor $\mathsf{H}(x) = \delta^*(x)/\theta(x)$. The shape factor is commonly used to characterize the nature of boundary layers. The higher the value of H, the stronger the local adverse pressure gradient, and vice versa. Practically, the standard laminar boundary layer without pressure gradient, also called Blasius boundary layer, is known to display the reference value $\mathsf{H} = 2.59$ while turbulent flows are characterized by much lower values (typically $\mathsf{H} \approx 1.3$).

We have represented in figure 1.6(a) the evolution of the three quantities $\delta^*(x)$, $\theta(x)$, and $\mathsf{H}(x)$ for the upstream boundary layer ($-2 \leq x \leq 0$) and in figure 1.6(b) for the downstream boundary layer ($-20 \leq x \leq 100$). The upstream boundary layer develops from $x = -2$ until it reaches a value $\delta^* = 0.082$ at $x = 0$. The same holds for the downstream boundary layer which is much thicker due to very long computational size and the high diffusion (due to the low Reynolds number). At the outlet position $x = 100$, the boundary layer reaches a displacement thickness of $\delta^* = 0.686$. It should be reminded that spatially developing boundary layers may be subjected to convective instabilities when they reach a sufficient thickness and perturbations are then spatially amplified while being convected downstream. According to local stability analysis, boundary layer may exhibit this instability by amplifying Tollmien-Schlichting waves when the the Reynolds number based on the local displacement thickness Re_{δ^*} is larger than the critical value ≈ 520. In our case, $Re_{\delta^*} = 49.2$ at $x = 0$ and $Re_{\delta^*} = 411.8$ at $x = 100$. Consequently, our set of parameters rules out the potential convective instabilities in both the upstream and downstream boundary layer.
Now, we turn our attention to the shape factor $\mathsf{H}(x)$. In the upstream region, it displays a transient evolution near the starting position $x = -2$ and then approaches the Blasius reference value 2.59 emphasized by a dashed line. The shape factor slowly decrease from $x = -1.5$ to $x = 0$ which is

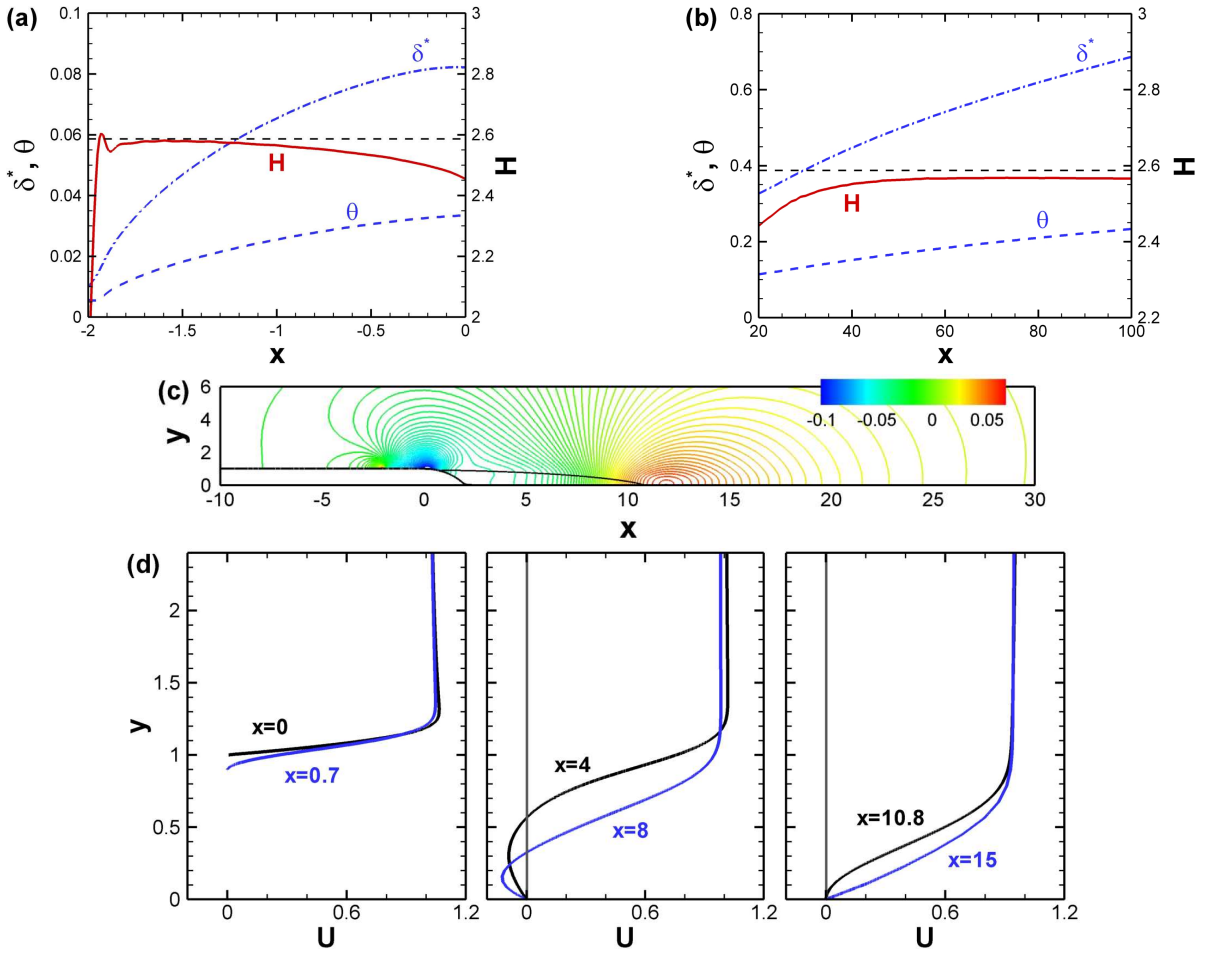


Figure 1.6: (a,b) Evolution of the displacement thickness, the momentum thickness and shape factor in the streamwise direction. (a) stands for the upstream boundary layer while (b) stands for the downstream one. (c) displays the pressure field of the base flow. (d) illustrates the boundary layer profiles at some streamwise positions. These results come from the base flow at $Re = 600$

attributed to pressure gradient effect. This is illustrated in figure 1.6(c) where we have represented the pressure field of the base flow. It is indeed observed that the upstream boundary layer is subjected to a positive (or favorable) pressure gradient on $(-2 \leq x \leq 0)$. Concerning the downstream boundary layer, its shape factor is much closer to the Blasius reference value $\mathsf{H}(x) = 2.59$, see figure 1.6(b). This result is natural and expected since the pressure gradient is observed to be much weaker in this region, see figure 1.6(c).

Beyond the streamwise position $x = 0$, the pressure gradient become strongly negative (adverse) due to the wall curvature induced by the geometry. This naturally leads to the separation of the boundary layer and, farther, to a long developing shear layer. For illustration, some boundary layer profiles $U(y)$ have been depicted on figure 1.6(d). This includes the upstream profiles at $x = 0$, $x = 0.7$, the separated profiles at $x = 4$, $x = 8$ and the reattached profiles at $x = 10.8$ and $x = 15$.

1.3.2.3 Vorticity

Since the base flow is two-dimensional, the vorticity field has only one non-null component. Denoting the x, y and z components of the vorticity field by Ω_x, Ω_y and Ω_z respectively, we have:

$$\Omega_x = \Omega_y = 0 \quad \text{and} \quad \Omega_z = \partial_x V - \partial_y U \tag{1.10}$$

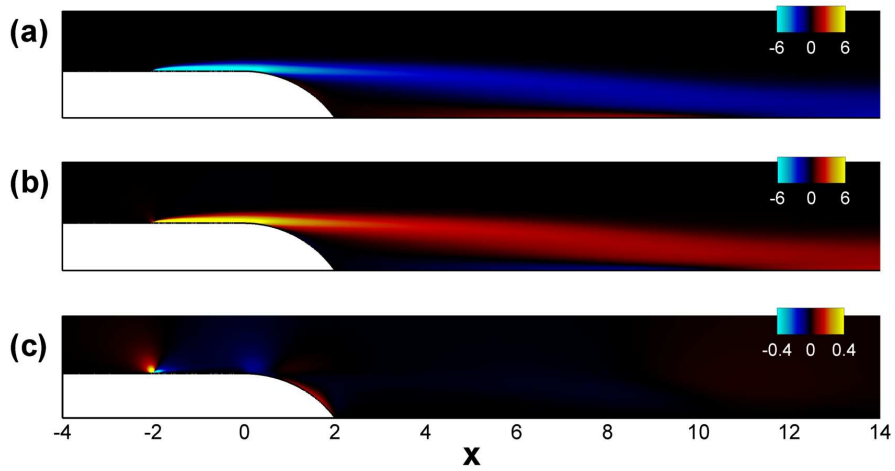


Figure 1.7: (a) Transverse vorticity of the base flow at $Re = 600$ given by $\Omega_z = \partial_x V - \partial_y U$. (b,c) stand for its derivative components $\partial_y U$ and $\partial_x V$ respectively.

We have represented in figure 1.7(a) the transverse component Ω_z of the base flow vorticity at $Re = 600$. As expected, the vorticity is predominantly localized in the upstream and downstream boundary layers, but mostly in the separated shear layer. Actually, a two-dimensional vorticity sheet emerges from the no-slip wall condition and spreads in the downstream direction due to diffusion. Noticeably, the two derivative components $\partial_y U$ and $\partial_x V$ have been shown to play an important role in the amplification of perturbations Schmid & Henningson (2001). It is indeed the base flow gradient term in the linearized Navier-Stokes equations which is responsible for the production of perturbations. We have represented these two derivatives in figure 1.7(b) and (1.7(c) respectively. Interestingly, the term $\partial_y U$ is widely larger than $\partial_x V$. Consequently, the global flow vorticity and shear are predominantly given by the cross-stream variation of the streamwise velocity, analogously to a standard one-dimensional shear layer. As we will see in section 1.5, convective instabilities may indeed arise from the rolling up of the vorticity sheet by the Kelvin-Helmholtz mechanism.

1.3.3 Dependence with the domain size

In this subsection, we briefly investigate the effect of the computational domain on the base flow recirculation length L_r. This is illustrated for the Reynolds number $Re = 600$ where we choose the refinement density of the mesh $M1$ (see table 1.1) which has been shown to be sufficient to yield converged results. The effect of the three parameters L_i, L_o and H are examined here. Note that these quantities have been introduced in figure 1.2 as the space between the inflow, outflow and top boundaries and the origin position $(x = 0; y = 0)$. More particularly, we demonstrate that the previously considered set $L_i = -20$, $L_o = 100$ and $H = 20$ is sufficient to avoid any confinement effect. To this end, we proceed by fixing two of these values and varying the remaining one, while computing the recirculation length of the resulting base flows. We have represented on figure 1.8(a,b) the evolution of the recirculation length as a function of L_i and L_o respectively. We observe that, for the range of parameters $L_i = [3; 20]$ and $L_o = [20; 100]$ investigated, the recirculation length is almost constant. In these cases, no confinement effects could be observed by varying the upstream and downstream boundary positions. On the contrary, the recirculation length clearly depends on the height H of the computational domain. As depicted on figure 1.8(c), the length L_r quickly rises when decreasing the height of the computational box. This observation may be simply interpreted by the flow rate conservation. For low values of H, the free-stream and downstream flow velocities are significantly lower than the prescribed inflow velocity U_∞ so that less inertial energy is brought to the detached boundary layer, which reattaches much farther downstream. However, for $H > 10$, this dependence remains weak which confirms our previous convergence statement. As a

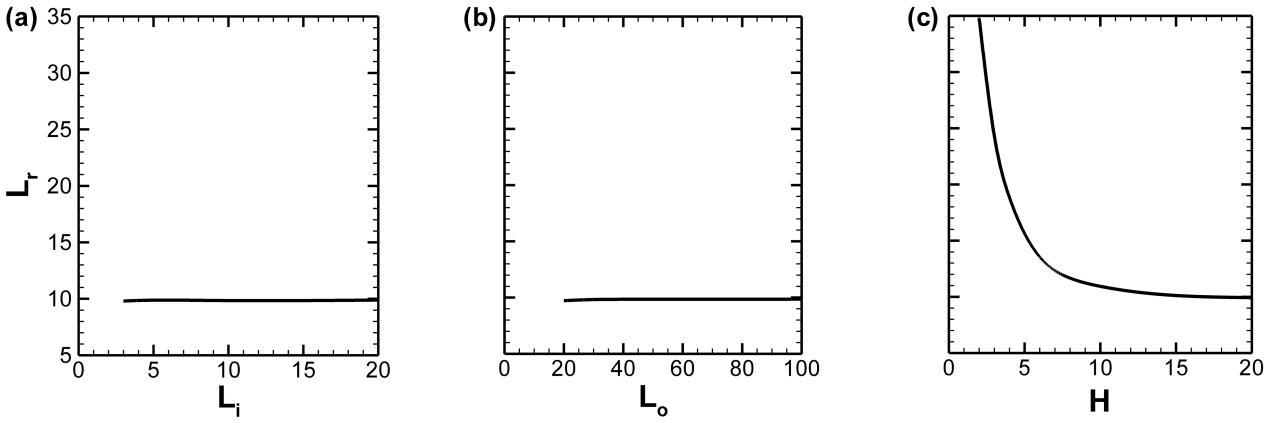


Figure 1.8: (a,b,c) Evolution of the recirculation length L_r of the base flow at $Re = 600$ as a function of L_i, L_o and H respectively.

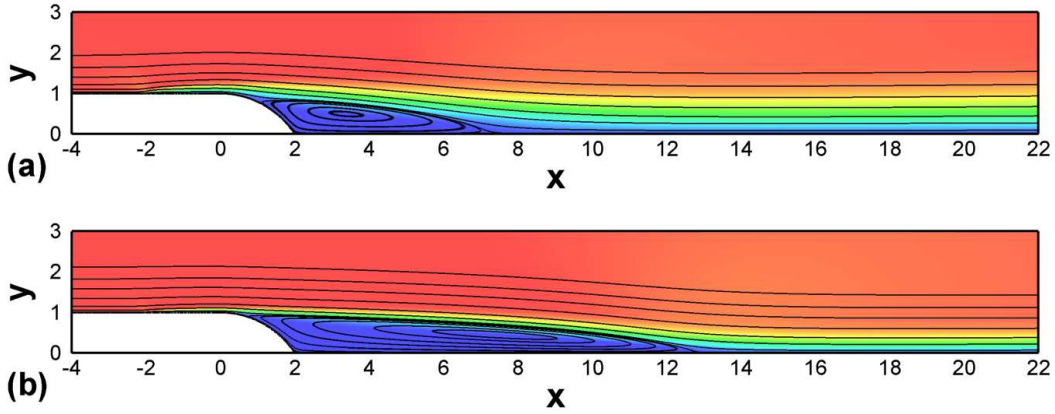


Figure 1.9: (a,b) Longitudinal velocity of the base flows at $Re = 200$ and $Re = 1000$ respectively.

result of this study, we choose hereafter the set of values $L_i = -20$, $L_o = 100$ and $H = 20$ for all the considered computational domains. This choice is indeed sufficient to yield converged results, which also holds (not shown) for the various Reynolds numbers considered in the rest of this chapter.

1.3.4 Dependence with the Reynolds number

We conclude this section by exposing the influence of the Reynolds number on the base flows. We have represented in figure 1.9(a,b) the base flows corresponding to the Reynolds numbers 200 and 1000 respectively. Two important observations are: (i) the thinning of the boundary layer and (ii) the large increase of the recirculation length when increasing the Reynolds number. These two points are quantified in the following.

We have depicted in figure 1.10(a) the evolution of the displacement thickness at $x = 0$ as a function of the Reynolds number. It is naturally observed to fall off when increasing Re. This is easily explained by the weaker viscosity, and therefore, by the weaker diffusion. As a result, the separated shear layer also displays the same behavior. To give an idea, we have computed the shear layer thickness $\delta_\omega(x)$ (also known as the vorticity thickness) which is given by

$$\delta_\omega(x) = \frac{\max_y u(x,y) - \min_y u(x,y)}{\max_y \partial_y u(x,y)} \tag{1.11}$$

We have represented in figure 1.10(b) its evolution at the position $x = 2$ as a function of the Reynolds

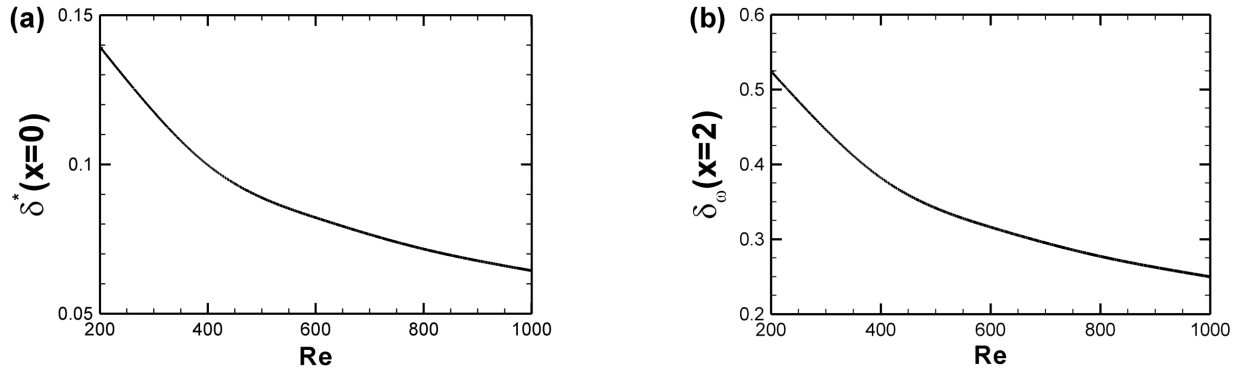


Figure 1.10: (a) Evolution of the boundary layer displacement thickness at (x=0). The variation of the shear layer thickness at (x=2) is displayed in (b).

number. It indeed displays a very similar fall off. Given the results on the stability of shear flows, see Schmid & Henningson (2001), it is reasonable to expect an increasing potential of the flow to spatially amplify perturbations for higher Reynolds numbers. This latter statement will be further highlighted in section 1.5.

Now, we deal with the behavior of the recirculation bubble length L_r as a function of Re, its evolution is represented in figure 1.11(a). We remind that the base flow is steady in time and that the natural viscosity is the only effect which can favorably balance the adverse pressure gradient in the boundary layer. When decreasing the viscous effects by increasing Re, separation occurs earlier and the flow reattaches much farther downstream. This is illustrated in figure 1.11(b) where the coordinates x_s and y_s of the separation point are displayed as a function of the Reynolds number. As expected we notice that x_s decreases while y_s increases for higher Reynolds numbers. Note that the recirculation length reaches quite high values and exceeds 12 when $Re > 1000$. In reality, these base flows may exhibit various instabilities which result in flow unsteadiness. The mean recirculation bubble may then be shortened. As an illustration, we have reported some experimental results from the literature in figures 1.12(a) and 1.12(b). Note that the recirculation lengths reported by the different authors have been extracted from the mean flows on the mid plans ($z = constant$). To give an idea, we have also displayed our computational solution on the same figures by a bold black line. Our results should not be quantitatively compared to the experiments since many factors such as the step shape, the incoming boundary layer thickness, the aspect ratio or the finite transverse length, may largely differ from ours. However, we roughly recover similar recirculation lengths in the low Reynolds number regime. For higher Reynolds numbers, the decrease of the recirculation length observed in figure 1.12(a) comes from the unsteadiness of the flow which acts as an additional viscosity. The analysis presented in the next sections will indeed confirm that the base flows may be subjected to hydrodynamic instabilities for sufficiently high Reynolds numbers. The highest values of L_r computed here may thus be or minor physical or practical relevance.

1.4 Long-time linear stability: a modal analysis

In this section, the linear stability of the base flows is assessed. More precisely, the linear dynamics of three-dimensional perturbations about the $2D$ base flows are investigated for long times. To do so, we classically proceed by computing the leading eigenvalues of the linearized Navier-Stokes operator. Then, the potential occurrence of unstable eigenvalues allows to specify the asymptotic stability of the flows.

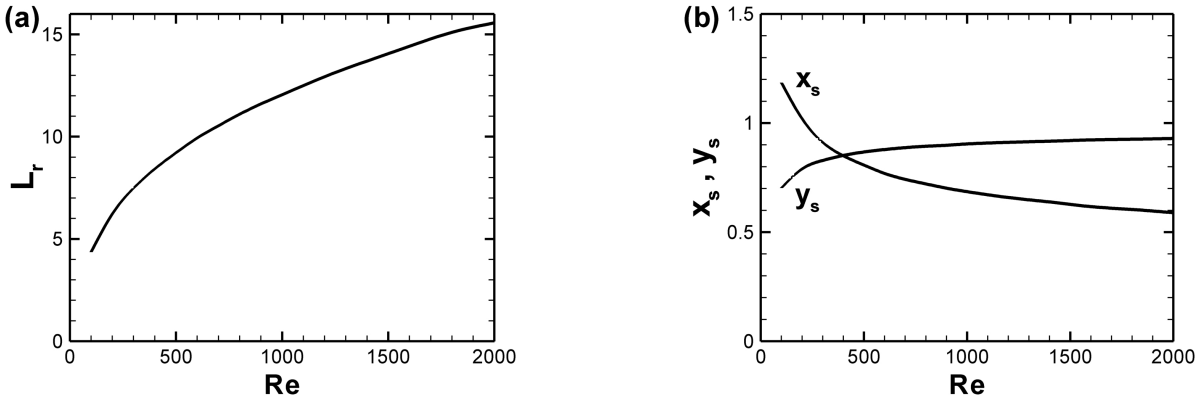


Figure 1.11: (a) Evolution of the recirculation length as a function of the Reynolds number. The variations of the separation point coordinates x_s and y_s are shown in (b).

1.4.1 Problem formulation

We consider the previously defined linearized Navier-Stokes equations (1.4), which yield the dynamics of the $3D$ perturbations (u, v, w, p) about the base flow (U, V, W, P). We denote the $3D$ the perturbation field by $\boldsymbol{q}(x, y, z, t) = (u, v, iw, p)$ where the transverse velocity component has been multiplied by the complex i. This little trick will be shown to lead to purely real equations in the final system formulation, see Theofilis (2000). Since the base flow is $2D$, the perturbation may be decomposed in the form of transverse waves.

$$\boldsymbol{q}(x, y, z, t) = \frac{1}{2}\left[\, \tilde{\boldsymbol{q}}(x, y, t)\, e^{ikz} + c.c. \,\right] \tag{1.12}$$

where $\tilde{\boldsymbol{q}}(x, y, t)$ is the $2D$ component associated with the transverse wave number k, and $c.c.$ denotes the complex conjugate. To be more precise, $\tilde{\boldsymbol{q}}(x, y, t)$ can be written as $\tilde{\boldsymbol{q}}(x, y, t) = (\tilde{u}, \tilde{v}, i\tilde{w}, \tilde{p})$ where all the term with a superscript denote the $2D$ component associated with the wave number k. Now, it is possible to express the linearized Navier-Stokes equations in a matrical and more compact form by

$$\mathsf{B}\, \partial_t \tilde{\boldsymbol{q}} = \boldsymbol{A}\, \tilde{\boldsymbol{q}} \tag{1.13}$$

where we have introduced the two operators $\boldsymbol{A}$ and B which are defined by

$$\boldsymbol{A} = \begin{pmatrix} \mathsf{D} - \mathsf{C} - \partial_x U & -\partial_y U & 0 & -\partial_x \\ -\partial_x V & \mathsf{D} - \mathsf{C} - \partial_y V & 0 & -\partial_y \\ 0 & 0 & \mathsf{D} - \mathsf{C} & k \\ \partial_x & \partial_y & k & 0 \end{pmatrix}, \quad \mathsf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{1.14}$$

where $\mathsf{D} = Re^{-1}\left(\partial_x^2 + \partial_y^2 - k^2\right)$ accounts for the viscous diffusion and $\mathsf{C} = U\partial_x + V\partial_y$ for the advection by the base-flow. Regarding the boundary conditions, they are chosen accordingly with those introduced in 1.2.2. Denoting $\tilde{\boldsymbol{u}} = (\tilde{u}, \tilde{v}, \tilde{w})$, we enforce: (i) a homogeneous Dirichlet condition $\tilde{\boldsymbol{u}} = 0$ on the inlet and wall boundaries, (ii) a symmetry condition $\partial_y \tilde{u} = 0$ and $\tilde{v} = 0$ on the top boundary and on the upstream lower boundary localized at $(x < L_b)$, (iii) the outflow condition $\tilde{p}\boldsymbol{n} - Re^{-1}(\nabla\tilde{\boldsymbol{u}}) \cdot \boldsymbol{n} = 0$ at the outlet. As expected, the particular form of $\tilde{\boldsymbol{q}}$ with the $\tilde{w}$ component in quadrature with the three others allows one to formulate the dynamical system (1.13) with real matrices $\boldsymbol{A}$ and B.

Now we want to decompose the perturbation into global modes. To do so, the common assumption of an exponential time-dependence enables to express the perturbation as

$$\boldsymbol{q}(x, y, z, t) = \frac{1}{2}\left[\, \hat{\boldsymbol{q}}(x, y)\, e^{\sigma t + ikz} + c.c. \,\right] \tag{1.15}$$

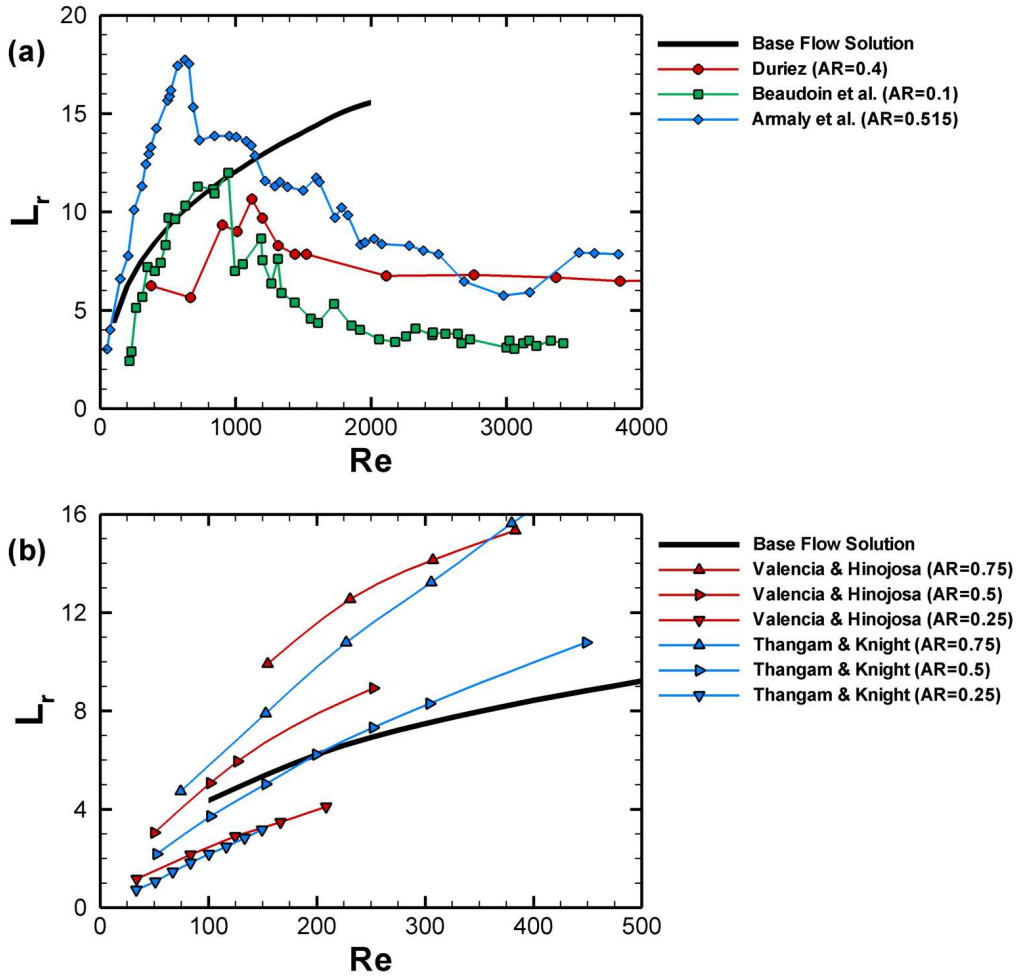


Figure 1.12: Evolution of the mean recirculation length as a function of the Reynolds number for various experimental studies from: Duriez (2009); Beaudoin *et al.* (2004); Armaly *et al.* (1983); Valencia & Hinojosa (1997); Thangam & Knight (1989). The associated aspect ratios $AR = h/H$ have also been reported. Our results, computed from the base flows, have been depicted by a bold black line. Importantly, it should be noted that these experiments have been performed with different conditions (boundary layer thickness, noise level, spanwise confinement) so that an accurate comparison is not relevant. With our choice of parameter $L_b = -2$, our boundary layer thickness at $x = 0$ matches with that obtained in the experiment by Duriez (2009) at $Re = 600$. However, our reciruclation length is found bigger than his at this Reynolds number, which is attributed to the confinement effects induced by both the side and top walls. Interestingly, an increase of the recirculation length is observed for higher aspect ratios in both the experiments by Valencia & Hinojosa (1997) and Thangam & Knight (1989). This latter observation is in agreement with our results exposed in figure 1.8(c).

so that

$$\tilde{\boldsymbol{q}}(x,y,t) = \hat{\boldsymbol{q}}(x,y)\, e^{\sigma t} \tag{1.16}$$

where σ is the complex eigenmode and $\hat{\boldsymbol{q}}$ its associated complex eigenvector. With this definition, the real and imaginary parts of $\sigma = \alpha + i\omega$ are respectively the growth rate and the pulsation of the mode. In other words, if $\alpha < 0$, the global mode is stable whereas if $\alpha > 0$ the mode is unstable. Replacing this expression in the dynamical system (1.13) then defines the following generalized eigenvalue problem

$$\boldsymbol{A}\,\hat{\boldsymbol{q}} = \sigma\, \mathbf{B}\hat{\boldsymbol{q}} \tag{1.17}$$

The modal analysis consists in finding the most unstable eigenmodes (those with the highest value of α). If there is as least one unstable mode, i.e. with a positive growth rate α, then the flow is said globally unstable. Any disturbance would then lead to an exponential increase of the perturbation energy for long times. In reality, the energy would obviously not reach infinite values since non-linear saturation would first intervene and stabilize the system. Now, if there are no unstable modes, the flow is said to be globally stable; this case is dealt with in the next section. In what follows, we briefly present the numerical methods used to solve equation (1.17) and we then present the results of the modal stability analysis.

1.4.2 Numerical methods

Numerically, the discrete matrices $\boldsymbol{A}$ and $\mathbf{B}$ resulting from the variational formulation of the linearized Navier-Stokes equations, given in equation (1.14), are generated with the FreeFem++ software through the same finite-element method. The size of the eigenvalue problem (1.17), i.e. the size of the matrices $\boldsymbol{A}$ and $\mathbf{B}$, is typically of order $O(10^5)$ as soon as two-dimensional base flows are considered. To give an idea, for the mesh previously denoted as M_1, there are $\approx 360\,000$ degrees of freedom. As a result, equation (1.17) cannot be directly solved by classical QR algorithms. Instead, we may only look for the most unstable eigenmodes by using iterative Arnoldi methods. The use of a shift and invert strategy enables to obtain eigenvalues in the vicinity of some given complex shift. In particular, use of a purely imaginary shift enables to track the leading global mode (with the highest values of α). The original eigenvalues problem is converted into:

$$(\boldsymbol{A} - \beta\mathbf{B})^{-1}\mathbf{B}\hat{\boldsymbol{q}} = \zeta\hat{\boldsymbol{q}} \quad \text{with} \quad \zeta = \frac{1}{\sigma - \beta} \tag{1.18}$$

where β is the shift parameter. By using the ARPACK routines, a Krylov subspace is generated by $\mathrm{span}\{\mathbf{C}^k\mathbf{V}\}_{0\leq k\leq m-1}$ with $\mathbf{C} = (\boldsymbol{A} - \beta\mathbf{B})^{-1}\mathbf{B}$, $\mathbf{V}$ being an initial vector and m denoting the dimension of the subspace. A LU decomposition of $(\boldsymbol{A}-\beta\mathbf{B})$ at the beginning of the algorithm allows a fast generation of the Krylov subspace thanks to a successive resolution of the linear system associated with $(\boldsymbol{A} - \beta\mathbf{B})$. A part of the global eigenspectrum can then be constituted by the applying multiple Arnoldi algorithms and moving the shift position in the complex plan. Iterations on Krylov subspaces of dimension 300 allowed to recover a set of 150 eigenvalues with sufficient accuracy.

1.4.3 Stability analysis

The stability analysis is performed by examining the eigenspectrums obtained for different values of the transverse wave number k and Reynolds number Re. A typical eigenspectrum, corresponding to the set of parameters $k = 0$ and $Re = 600$, is depicted in figure 1.13. This one displays some branches and the least stable mode has a null pulsation ω. Note that the spectrum is symmetric since the complex conjugate of any global mode is also a global mode. A large region of the spectrum localized at $\alpha < -0.09$ exhibits a high concentration of modes. This part may be attributed to the discrete representation of the continuous part of the eigenspectrum. Such considerations may be found in the book by Schmid

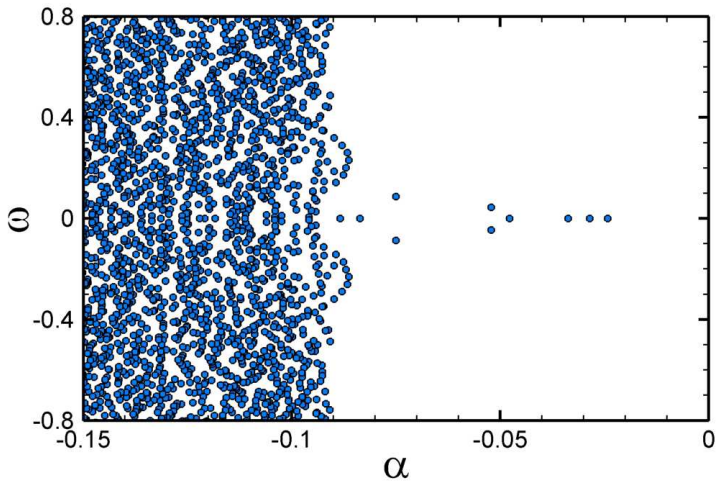


Figure 1.13: Part of the global eigenspectrum computed for the parameters $k = 0$ and $Re = 600$.

& Henningson (2001) where it is shown that the representation of the global convection in open flows induces continuous branches in the eigenspectrum. It should be mentioned here that we did not find any two-dimensional ($k = 0$) unstable global modes for Reynolds numbers up to 2500.

Now, we pay attention to the most unstable eigenmode. When varying the parameters in the range $k = [0; 10]$ and $Re = [100; 2000]$, we observed eigenspectrums which are very similar to that depicted in figure 1.13. More particularly, the most unstable eigenmode remains not oscillating (with $\omega = 0$). However, when increasing the Reynolds number, it is observed to possibly become unstable for an interval of wave numbers. We have reported in figure 1.14(a) the evolution of the growth rate of the most unstable eigenmode α_{max} as a function of the wave number k for different Reynolds numbers. For $Re < 400$ the flow is globally stable whereas a band of unstable wave numbers arises for $Re > 600$. This band has been emphasized for $Re = 600$ by dashed lines and is then further broadened for larger Reynolds numbers. The threshold at which $\alpha_{max} = 0$ is of primary interest and corresponds to emergence of the global instability. A $2D$ representation of α_{max} as a function of k and Re has been reported in figure 1.14(b) and the threshold at which $\alpha_{max} = 0$ has been emphasized. The critical Reynolds number above which global instability appears is identified as $Re_c = 526$. At this Reynolds number, the only unstable wave number is found to be $k = 1.125$, which corresponds to the transverse wave length $l_z = \frac{2\pi}{k} = 5.59$. At this stage, we conclude that increasing the Reynolds number leads to global instability through the rising of a stationary and transversally periodic global mode. This one is further presented below.

1.4.4 Description of the unstable eigenmode

To illustrate the dynamics of the eigenmode, we choose the Reynolds number $Re = 600$. The most unstable wave number is then given by $k = 1.08$. According to our notations, the eigenmode may be written in the form $\hat{\boldsymbol{q}}(x, y) = (\hat{u}, \hat{v}, i\hat{w}, \hat{p})$, where we have introduced the three components of the velocity and pressure fields associated with the global mode. Given the decomposition of the perturbation introduced in (1.15), the dynamics of the global mode can be obtained by

$$\boldsymbol{q}(x, y, z, t) = \frac{1}{2} [\, \hat{\boldsymbol{q}}(x, y)\, e^{ikz} + c.c.\,] e^{\alpha_{max} t} \tag{1.19}$$

or, by making explicit the real part of the complex quantity:

$$\boldsymbol{q}(x, y, z, t) = [\hat{\boldsymbol{q}}_r(x, y) \cos(kz) - \hat{\boldsymbol{q}}_i(x, y) \sin(kz)] e^{\alpha_{max} t} \tag{1.20}$$

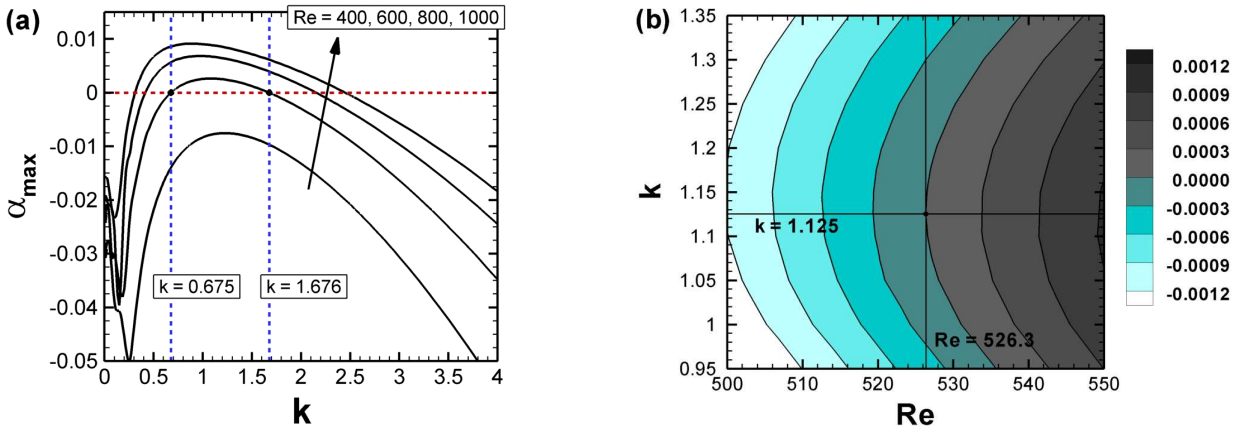


Figure 1.14: (a) Evolution of the most unstable growth rate α_{max} as a function of the wave number k. At $Re = 600$ the unstable wave lengths are $k = [0.675; 1.676]$ and the most unstable one is $k = 1.08$. Figure (b) displays the same quantity as a function of k and Re in a $2D$ plot.

where we have defined the real part $\hat{\boldsymbol{q}}_r$ and imaginary part $\hat{\boldsymbol{q}}_i$ of the eigenmode $\hat{\boldsymbol{q}}$. Note that this equation yields the dynamics of the perturbation state when the initial condition is the global mode itself. The global mode is then observed to amplify exponentially in time at the growth rate α_{max}. We have depicted in figure 1.15(a,b,c) the three velocity components $\hat{u}$, $\hat{v}$ and $\hat{w}$ of the unstable eigenmode. We observe that they remain localized in the recirculation bubble and that the streamwise component is significantly larger than the two others. The spatial structure of the mode displays a long streamwise wave length, as commonly observed for stationary (or low-frequency) structures. The occurrence of such a stationary growing three-dimensional structure has been identified in the literature as a common feature of many low-Reynolds number incompressible separated flows, see Theofilis *et al.* (2000); Barkley *et al.* (2002); Marquet *et al.* (2009); Cherubini *et al.* (2010*b*), but also in the case of the shock-wave/laminar-boundary-layer interaction, see Robinet (2007*a*). In particular, the structure of the present $3D$ eigenmode is very similar to those encountered in these studies.

A $3D$ view of the global eigenmode can be recovered by computing $\hat{\boldsymbol{q}}_r \cos(kz) - \hat{\boldsymbol{q}}_i \sin(kz)$ and its amplitude is arbitrary since it exponentially increases in time. We have displayed in figure 1.16(a) three iso-vorticity sheets of the global eigenmode. As observed in the previously cited references, the mode structure displays rolls extending through the streamlines curvatures within the recirculation bubble. Figure 1.16(b) displays the same quantity from a top view where we have reported by vertical lines the position of the step and the reattachment line of the base flow at $x = 10.8$. This figure further highlights the vorticity rolls induced by the global mode and its confined spatial location within the recirculation region. Furthermore, it should be noticed that the transverse wave length of this mode is $l_z = 5.82$ which stands for a rather large value (the step height is equal to $h = 1$).

Recent works on a backward-facing step flow by Barkley *et al.* (2002) or on the flow behind a bump by Gallaire *et al.* (2007) suggested that this global mode results from an inviscid centrifugal instability linked with the curvature of the closed streamlines in the recirculation region. This statement was argued from the high dependence of the global mode growth rate and the broadening of the unstable transverse wavelengths band as a function of the recirculation length. Mostly, they compared favorably their computational results to the Rayleigh criterion introduced by Bayly (1988) and more recently by Sipp & Jacquin (2000). Such a comparison was also performed successfully in the work by Cherubini *et al.* (2010*b*) by using both a local and global version of the Rayleigh criterion. Due to these previous work, we did not further investigated this aspect of the global mode dynamics.

We now turn our attention to an interesting characteristic of the global mode. We observed in figure 1.15 that the longitudinal velocity component of the unstable global mode is higher than the two others.

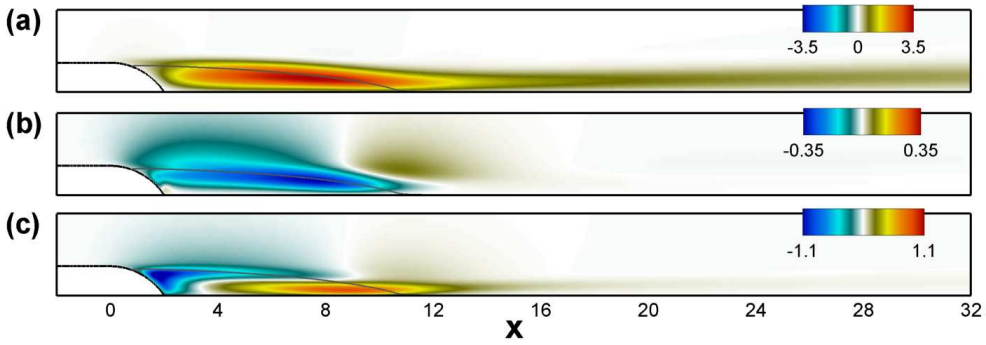


Figure 1.15: (a,b,c) Components $\hat{u}$, $\hat{v}$ and $\hat{w}$ of the unstable eigenmode at $Re = 600$ and for $k = 1.08$.

This observation is reminiscent of the lift-up mechanism which transfer energy from the perturbation streamwise vorticity to its streamwise velocity, see Landahl (1980). Notably, a recent work by Marquet *et al.* (2009) confirmed that the lift-up mechanism plays an important role in the physical mechanism enhancing the unstable global mode of a S-shaped backward-facing step flow. Interestingly, this result holds in our flow configuration. To highlight it, we have depicted in figure 1.17(a) the perturbation longitudinal velocity. It is observed to display alternate high speed and low speed regions inside the recirculation bubble. These are referred to as "high speed and low speed streaks" in the work by Marquet *et al.* (2009) due to its structural analogy with the boundary layer streaks. Furthermore, we have represented in figure 1.18 the longitudinal vorticity of the perturbation in the cross plan $x = 8$ superimposed on the sum of the base flow and perturbation streamwise velocities. We observe that self-sustained streamwise vortices of opposite sign are generated inside the shear layer. These vortices then induce alternated vertical velocity perturbations (shown by arrows). As a result, there is an alternated transport of low speed fluid towards regions of higher speed and inversely within the shear layer of the recirculation bubble. This lift-up mechanism is thus responsible for the alternated high and low speed longitudinal velocity regions observed in figure 1.17(a). The sum of the base flow and the perturbation has been represented in figure 1.17(b) by its longitudinal velocity. We choose the amplitude of the perturbation arbitrary so as to clearly observe its effects on the base flow. Furthermore, the displayed iso-contour, corresponding to the longitudinal velocity -0.001, clearly emphasizes the resulting recirculation bubble. This one is observed to be deformed by the global mode which displaces upstream and downstream the reattachment line. Our results are in agreement with those exposed by Marquet *et al.* (2009) and confirm the role of the lift-up mechanism in the global instability of the flow. We refer interested readers to a recent article by Rodríguez & Theofilis (2010) where the topology of the separated flows reconstructed by linear superposition of a $2D$ base flow and its $3D$ global modes are investigated.

1.5 Short-time linear stability: a non-modal analysis

We observed in the previous section the occurrence of an unstable non oscillating and $3D$ global mode when increasing the Reynolds number. In that case, the dynamics of all other global modes vanish for large times; the flow is then only driven by the unstable global mode. However, some perturbations may be transiently amplified in time before reaching such an asymptotic dynamics. This common feature of open flows has been interpreted by the non-orthogonality of the basis of global modes Schmid & Henningson (2001); Schmid (2007). In particular, it is shown that, even if all the eigenvectors are stable, the energy of the perturbation may reach very large transient values. In reality, these transient energy growths may be sufficiently strong to trigger non-linear effects which often lead to the transition to turbulence. It is thus of primary interest to assess how the flow may undergo such transient growths and by what physical mechanisms. This stands for the concern of this section. More precisely, our goal is to

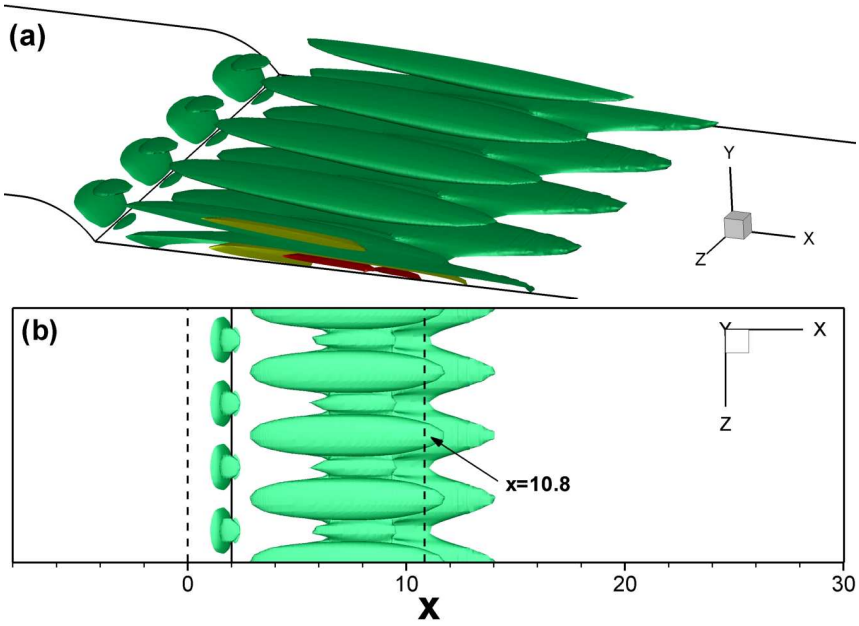


Figure 1.16: (a) Three-dimensional view of the iso-vorticity field of the unstable eigenmode. A top view is provided in (b). The first two vertical lines represent the beginning $x = 0$ and ending $x = 2$ of the step while the last one stands for the reattachment line of the base flow $x = 10.8$.

identify the optimal perturbations, i.e. the perturbations leading to a maximum energy amplification. The analysis of these perturbations and their associated energy gains then allows to assess the different transition scenarios in the flow. Resorting to adjoint equations will be shown to be a necessary step to solve the underlying optimization problem. We begin this section by introducing these adjoint equations. The idea is to define an inner product and to derive the resulting adjoint Navier-Stokes equations. In the remaining of the section, adjoint quantities are used as key mathematical tools to identify the optimal perturbations of the flow.

1.5.1 Adjoint equations

The present formulation follow through the "problem formulation" of section 1.4.1. In particular, we deal with $3D$ transversally periodic perturbations developing on the $2D$ base flow. We start by defining the inner product $\langle\,\rangle$ defined for any $2D$ field $\hat{\boldsymbol{q}}_1(x,y)$ and $\hat{\boldsymbol{q}}_2(x,y)$ by

$$\langle \hat{\boldsymbol{q}}_1, \hat{\boldsymbol{q}}_2 \rangle = \int_\Omega \hat{\boldsymbol{q}}_1^* \, \hat{\boldsymbol{q}}_2 \, d\Omega \tag{1.21}$$

where Ω denotes the entire $2D$ computational domain and * is the transconjugate. Then, we introduce the adjoint operator $\boldsymbol{A}^\dagger$ defined for any $2D$ field $\hat{\boldsymbol{q}}(x,y)$ fulfilling the boundary conditions defined earlier in section 1.4.1 and $\hat{\boldsymbol{q}}^\dagger(x,y)$ satisfying boundary conditions to be determined,

$$\langle \boldsymbol{A}\,\hat{\boldsymbol{q}}, \hat{\boldsymbol{q}}^\dagger \rangle = \langle \hat{\boldsymbol{q}}, \boldsymbol{A}^\dagger\,\hat{\boldsymbol{q}}^\dagger \rangle \tag{1.22}$$

where $\hat{\boldsymbol{q}}^\dagger = (\hat{u}^\dagger, \hat{v}^\dagger, i\hat{w}^\dagger, \hat{p}^\dagger)$ is the adjoint state of $\hat{\boldsymbol{q}}$. It should be noticed that the choice of this scalar product is linked to the perturbation kinetic energy. Indeed, the instantaneous energy density of a $3D$ transversally periodic perturbation $\boldsymbol{q}(x,y,z)$ is given by

$$E(\boldsymbol{q}) = \frac{1}{l_z \Omega} \int_{-l_z/2}^{+l_z/2} \int_\Omega (u^2 + v^2 + w^2)\, dxdydz \tag{1.23}$$

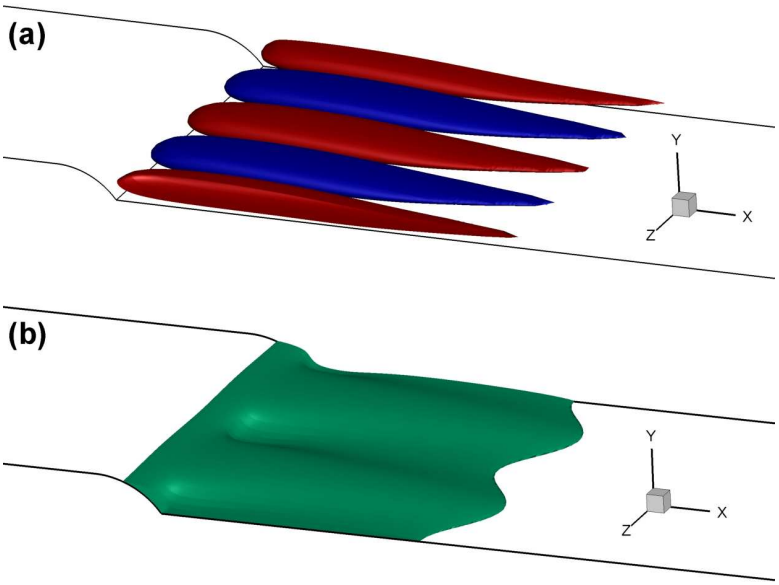


Figure 1.17: (a) Three-dimensional view of the longitudinal velocity of the global eigenmode. The sum of the perturbation (with an arbitrary amplitude) and base flow longitudinal velocity is depicted on (b).

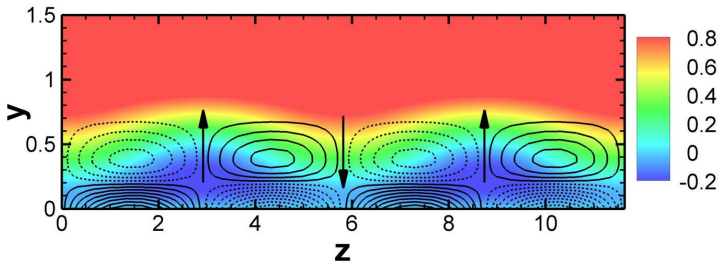


Figure 1.18: Contour of the streamwise vorticity (dotted lines standing for negative values) with superimposed streamwise velocity of the total flow field in the cross plan $x = 8$. The total flow field is the same as that depicted in figure 1.17, the amplitude of the global mode is arbitrary and has been chosen such as to clearly observe its effect on the base flow. The arrows symbolize the vertical momentum induced by the alternate streamwise vortices.

where l_z denotes the spanwise wavelength. By introducing the normal mode decomposition (1.15), the energy can further be formulated as

$$E(\boldsymbol{q}) = \frac{1}{2} E(\hat{\boldsymbol{q}}) e^{2\alpha t} \quad \text{so that} \quad E(\hat{\boldsymbol{q}}) = \langle \hat{\boldsymbol{q}}, \mathbf{B}\hat{\boldsymbol{q}} \rangle \tag{1.24}$$

where the $\frac{1}{2}$ term comes from integral of a periodic quantity in the z direction. Therefore, the energy of any transversally periodic perturbation is proportional to the inner product $\langle \hat{\boldsymbol{q}}, \mathbf{B}\hat{\boldsymbol{q}} \rangle$ which yields the $2D$ kinetic energy of the mode $\hat{\boldsymbol{q}}$.

To derive the adjoint equations, we first express the adjoint operators associated with $\boldsymbol{A}$ and $\mathbf{B}$. We proceed by integrating by parts the left-hand side of (1.22). The resulting adjoint operator $\boldsymbol{A}^\dagger$ is given by

$$\boldsymbol{A}^\dagger = \begin{pmatrix} \mathsf{D}+\mathsf{C}-\partial_x U & -\partial_x V & 0 & -\partial_x \\ -\partial_y U & \mathsf{D}+\mathsf{C}-\partial_y V & 0 & -\partial_y \\ 0 & 0 & \mathsf{D}+\mathsf{C} & k \\ \partial_x & \partial_y & k & 0 \end{pmatrix} \tag{1.25}$$

It is observed to be very similar to the original dynamical operator given in (1.14). More precisely, the difference comes from two distinct terms: (i) the convective terms C have an opposite sign and (ii) the off-diagonal terms $\partial_x V$ and $\partial_y U$ are switched. A common feature of the adjoint equations originating from (i) is the advection of the perturbation in the upstream direction (rather than downstream). Incidentally, the boundary conditions derived from the integration by parts, chosen so as to cancel all the boundary terms, are reversed. More precisely, the inflow condition, now at $x = L_o$, is given by the homogeneous Dirichlet condition $(\hat{u}^\dagger, \hat{v}^\dagger, \hat{w}^\dagger) = 0$ whereas the outflow condition at $x = -L_i$ is given by

$$\hat{p}^\dagger \boldsymbol{n} + Re^{-1}(\nabla \hat{\boldsymbol{u}}^\dagger) \cdot \boldsymbol{n} + \hat{\boldsymbol{u}}^\dagger (\boldsymbol{U} \cdot \boldsymbol{n}) = 0 \tag{1.26}$$

where $\hat{\boldsymbol{u}}^\dagger = (\hat{u}^\dagger, \hat{v}^\dagger, \hat{w}^\dagger)$ denotes the adjoint velocity vector and $\cdot$ stands for the standard scalar product between two vectors. Additionally, the upper boundary and the upstream lower boundary at $(-L_i \leq x \leq -L_b; y = 1)$ are associated with the symmetry condition given by

$$Re^{-1}\partial_y \hat{u}^\dagger + \hat{u}^\dagger V = 0 \quad \text{and} \quad \hat{v}^\dagger = 0 \tag{1.27}$$

Let us first consider the adjoint eigenvalue problem associated with (1.17). Since $\mathbf{B}$ is self-adjoint it reduces to

$$\boldsymbol{A}^\dagger \, \hat{\boldsymbol{q}}^\dagger = \sigma^\dagger \, \mathbf{B} \hat{\boldsymbol{q}}^\dagger \tag{1.28}$$

where $\sigma^\dagger$ and $\hat{\boldsymbol{q}}^\dagger$ stand for the adjoint eigenvalues and adjoint global modes respectively. It can be shown, see Schmid & Henningson (2001), that all the adjoint eigenvalues are the complex conjugate of the direct ones: $\sigma^\dagger = \sigma^*$ and that the adjoint global modes are bi-orthogonal to the direct global modes:

$$\langle \hat{\boldsymbol{q}}_i, \mathbf{B} \hat{\boldsymbol{q}}_j^\dagger \rangle = 0 \quad \text{when} \quad i \neq j \tag{1.29}$$

The operator $\boldsymbol{A}$ is said to be non-normal since it does not commute with its adjoint $\boldsymbol{A}^\dagger$. Therefore, its eigenvectors $\hat{\boldsymbol{q}}$ differs from those $\hat{\boldsymbol{q}}^\dagger$ of $\boldsymbol{A}^\dagger$, namely

$$\langle \hat{\boldsymbol{q}}_j, \mathbf{B} \hat{\boldsymbol{q}}_j^\dagger \rangle \neq 1 \quad \text{for all} \quad j \tag{1.30}$$

Additionally, it is now possible to associate adjoint equations to the linearized Navier-Stokes equations (1.13). They reduce to

$$\mathbf{B} \, \partial_t \tilde{\boldsymbol{q}}^\dagger = \boldsymbol{A}^\dagger \, \tilde{\boldsymbol{q}}^\dagger \tag{1.31}$$

and yield the temporal dynamics of the $2D$ adjoint state $\tilde{\boldsymbol{q}}(x, y, t)$. According to our previous notation, $\tilde{\boldsymbol{q}}$ stands for the modal component associated with the wave number k of the $3D$ adjoint state $\boldsymbol{q}^\dagger(x, y, z, t)$ so that

$$\boldsymbol{q}^\dagger(x, y, z, t) = \frac{1}{2} \left[\, \tilde{\boldsymbol{q}}^\dagger(x, y, t) \, e^{ikz} + c.c. \, \right] \tag{1.32}$$

which yields the adjoint state associated with the original $3D$ perturbation $\boldsymbol{q}(x, y, z, t)$.
An interesting point should be mentioned here. We said that the difference between $\boldsymbol{A}$ and $\boldsymbol{A}^\dagger$ comes from (i) the upstream advection of the perturbation and (ii) the switch in the off-diagonal terms. In fact, (i) is responsible for the so-called "convective non-normality" whereas (ii) induces the "lift-up non-normality", see Cossu & Chomaz (1997); Marquet *et al.* (2009). These two different non-normalities will be shown to play an important role in the adjoint solutions of both (1.28) and (1.31).

1.5.2 Optimal temporal perturbation analysis

1.5.2.1 Problem formulation and governing equations

The first analysis presented here consists in finding the initial perturbation at time $t = 0$ which leads to an optimal energy amplification at time T. The time T is called the horizon time and is a critical parameter of the study. According to our previous formulation, the dynamics of $3D$ perturbations are investigated. The decomposition (1.12) of the perturbation into transverse waves allows to perform the study wave number by wave number. The parameter k thus stands for a second critical quantity.

Let us consider a given value of T and k and a transversally periodic perturbation of the form $\boldsymbol{q}(x, y, z, t) = \frac{1}{2}\,[\,\tilde{\boldsymbol{q}}(x, y, t)\, e^{ikz} + c.c.\,]$. We look for the initial condition $\tilde{\boldsymbol{q}}(0)$ which optimizes the kinetic energy gain

$$\frac{E(\boldsymbol{q}(T))}{E(\boldsymbol{q}(0))} \quad \text{which is also equal to} \quad \frac{E(\tilde{\boldsymbol{q}}(T))}{E(\tilde{\boldsymbol{q}}(0))} \tag{1.33}$$

We introduce the time propagator $\boldsymbol{\Phi}$ which is defined so as to advance the perturbation in time by $\tilde{\boldsymbol{q}}(T) = \boldsymbol{\Phi}\,\tilde{\boldsymbol{q}}(0)$. Then the energy gain can be formulated as

$$\frac{E(\tilde{\boldsymbol{q}}(T))}{E(\tilde{\boldsymbol{q}}(0))} = \frac{\langle \boldsymbol{\Phi}\,\tilde{\boldsymbol{q}}(0), \mathbf{B}\boldsymbol{\Phi}\,\tilde{\boldsymbol{q}}(0)\rangle}{\langle \tilde{\boldsymbol{q}}(0), \mathbf{B}\tilde{\boldsymbol{q}}(0)\rangle} = \frac{\langle \tilde{\boldsymbol{q}}(0), \boldsymbol{\Phi}^\dagger\mathbf{B}\boldsymbol{\Phi}\,\tilde{\boldsymbol{q}}(0)\rangle}{\langle \tilde{\boldsymbol{q}}(0), \mathbf{B}\tilde{\boldsymbol{q}}(0)\rangle} \tag{1.34}$$

where we have introduced the adjoint time propagator $\boldsymbol{\Phi}^\dagger$. Now, it may be easily shown that the optimal initial perturbation and its associated optimal energy gain are given by the largest eigenvector and eigenvalue of the following generalized eigenvalue problem

$$\boldsymbol{\Phi}^\dagger\mathbf{B}\boldsymbol{\Phi}\,\tilde{\boldsymbol{q}}(0) = \gamma\mathbf{B}\tilde{\boldsymbol{q}}(0) \tag{1.35}$$

To solve this problem, we proceed similarly to the work by Blackburn *et al.* (2008); Marquet *et al.* (2008). We start by considering a random initial condition $\tilde{\boldsymbol{q}}_1(0)$ which is evolved forward in time through the Navier-Stokes equations (1.13) to the time T such that $\tilde{\boldsymbol{q}}_1(T) = \boldsymbol{\Phi}\,\tilde{\boldsymbol{q}}_1(0)$. Next, this is immediately followed by evolving the resulting state $\tilde{\boldsymbol{q}}_1(T)$ backward under the adjoint linearized Navier-Stokes equations (1.31) by T time units. The new state denoted by $\tilde{\boldsymbol{q}}_2(0)$ is localized upstream due to the action of the adjoint equations and is such that

$$\boldsymbol{\Phi}^\dagger\mathbf{B}\,\tilde{\boldsymbol{q}}_1(T) = \mathbf{B}\tilde{\boldsymbol{q}}_2(0) \tag{1.36}$$

which may be demonstrated by developing the adjoint Navier Stokes equations. Mostly, this equation leads to

$$\boldsymbol{\Phi}^\dagger\mathbf{B}\boldsymbol{\Phi}\,\tilde{\boldsymbol{q}}_1(0) = \mathbf{B}\tilde{\boldsymbol{q}}_2(0) \tag{1.37}$$

Consequently, by applying successively this procedure, the initial conditions $\tilde{\boldsymbol{q}}_j(0)$ eventually converges to the highest eigenvector of (1.35). The optimal energy gain may then be computed during the last direct time integration. For simplicity we choose a real initial condition $\tilde{\boldsymbol{q}}_1(0)$ so that all the states $\tilde{\boldsymbol{q}}_j(0)$ are real. As a result, the $3D$ associated flow states (u, v, w, p) defined by (1.12) may be recovered from

$$\begin{aligned} u &= \tilde{u}\cos(kz) \\ v &= \tilde{v}\cos(kz) \\ w &= \tilde{w}\sin(kz) \\ p &= \tilde{p}\cos(kz) \end{aligned} \tag{1.38}$$

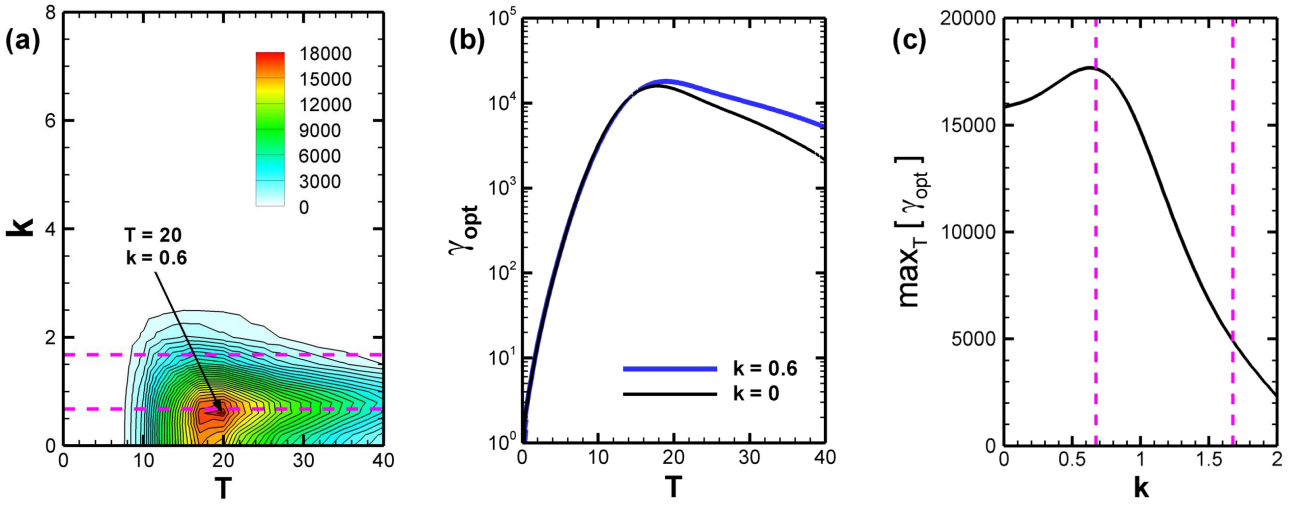


Figure 1.19: (a) Optimal energy gain γ_{opt} as a function of T and k in a $2D$ plot. (b) depicts the evolution of γ_{opt} as a function of T for $k = 0$ and $k = 0.6$. (c) stands for the evolution of maximum value of γ_{opt} over T as a function of k. These results have been computed for $Re = 600$. The band of unstable wave numbers has been emphasized by dashed lines.

1.5.2.2 Results

We apply the iterative direct-adjoint procedure to compute the optimal perturbations on the base flow solution at $Re = 600$. The mesh M_1 is chosen to perform this analysis; it is observed to yield widely converged results in terms of spatial discretization. The resolution of the direct (1.13) and adjoint (1.31) Navier-Stokes equations are performed by time-stepping simulations based on a second-order accurate time marching scheme and a time step $\delta t = 0.002$. We used the Uzawa algorithm which proceeds by computing iteratively the pressure field such that the velocity is divergent free before advancing the velocity field to the next time step. At each time step, the computation of the new pressure is performed by a conjugate gradient method which is accelerated by using the Cahouet-Chabart preconditioner.

The optimal energy gain γ_{opt} and the optimal perturbations are presented next. We have depicted in figure 1.19(a) a $2D$ plot of the optimal gain as a function of T and k. This one highlights potentially high transient growths for the lowest wave numbers k. Especially, the maximum gain is reached for $T = 20$ and $k = 0.6$, which stands for the transverse wave length $l_z = 10.47$. Note that the unstable wave number interval has been denoted by dashed lines. We recall that the optimal energy gain tends to infinity for large times T in this interval due to the exponential increase of the kinetic energy. Concerning two-dimensional perturbations ($k = 0$), an optimal energy gain is reached for $T = 18$. For a clearer representation, we have depicted in figure 1.19(b) the evolution of γ_{opt} as a function of T for two particular wave numbers: the first one $k = 0$ stands for $2D$ perturbations while the second $k = 0.6$ is the wave number at which the maximal energy gain over all possible parameters is found. Interestingly, the two are almost identical for short times ($T < 15$). Furthermore, their maximum energy gains are quite close; the maximum values of γ_{opt} reduce to 15 874 for $k = 0$ and 17 916 for $k = 0.6$. For higher wave numbers, the optimal gains vanish quite quickly. This is illustrated in figure 1.19(c) where we have reported the value $\max_T[\gamma_{opt}(k, T)]$ as a function of k, which is observed to fall off for $k > 1$ (i.e. $l_z < 6.28$). We thus conclude that the most energetic transient growths are related to perturbations having very large transverse wavelengths and $2D$ perturbations.
Let us first describe the $2D$ optimal perturbation ($k = 0$). We have displayed in figure 1.20(a) the optimal initial condition by its longitudinal velocity. Figures 1.20(b-d) stand for the evolution of this perturbation given by the linearized Navier-Stokes equations (1.13). The initial condition is observed to be localized in the upstream boundary layer, near separation, and is convected downstream while

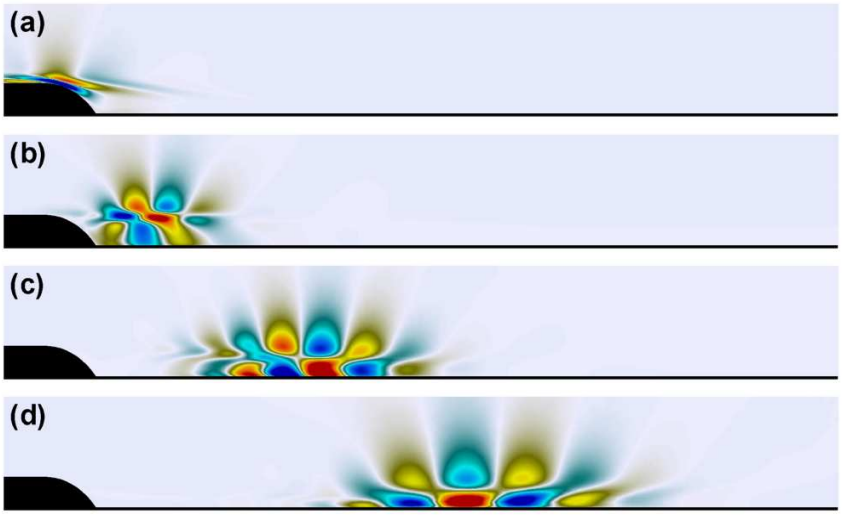


Figure 1.20: (a) Optimal initial perturbation computed for $T = 18$ and $k = 0$. (b,c,d) Evolution of this perturbation at times $t = 6, 18$ and 30 respectively. Flow structures are displayed by their longitudinal velocity.

being amplified in the recirculation bubble. Besides, the energy growth is maximum at $t = 18$ and starts to decrease as soon as the disturbance leaves the recirculation bubble. It should be noticed that the inclination of the initial condition against the shear is reminiscent of the Orr mechanism, see Orr (1907); Butler & Farrell (1992), where the perturbation extracts energy from the mean shear by transporting momentum down the velocity gradient. The shear layer resulting from the flow separation is responsible for the amplification of the perturbation due to the well-known Kelvin-Helmholtz instability. This instability selectively amplifies disturbances, more precisely the thinner the shear layer the smaller the streamwise wavelength of the most amplified disturbance. In our case, the shear layer thickens due to viscosity and the perturbation is thus observed to display increasing streamwise wavelengths while being convected downstream. This latter point is obviously partly attributed to the viscous effects on the perturbation itself. It should be recalled here that the global amplifier behavior of open flows have been recognized to derive from the convective instabilities of the considered flow, see Cossu & Chomaz (1997); Chomaz (2005). In a global framework, convective instabilities originates from the recirculation bubble where the flow may be found convectively unstable from a local stability analysis. Such local analysis have been reported in the work of Robinet (2007*b*) where the most amplified frequencies at each x positions may be found following the work by Chomaz *et al.* (1991).

To highlight the wave packet propagation, we have represented in figure 1.21(a) its kinetic energy envelope through the quantity $E_y(x,t) = \int_y [u(x,y,t)^2 + v(x,y,t)^2 + w(x,y,t)^2]dy$ every $\Delta t = 0.5$. The perturbations at times 10, 16.5 and 30 are furthermore emphasized to point out the spreading of the streamwise spatial support of the wave packet in time. Figure 1.21(b) stands for a $2D$ space-time representation of the wave packet through the quantity $u(x,t)$ extracted at $y = 1.05$. The influence of the non-parallelism of the base flow, associated with the recirculation bubble is noticeable by the curvature of the borders of the wave packet. This is further emphasized by the change of the x/t ray values which are emphasized by dashed lines. The wavefronts are characterized by the phase speeds $v_\phi = \frac{\Delta x}{\Delta t}$ equal to 0.26 and 0.9. Behind reattachment, at $x \approx 11$, the wave packet evolves along constant x/t rays and the associated phase speed reduces to 0.55. This latter point is consistent with the downstream relaxation of the base flow toward a boundary layer.

We now represent the $3D$ optimal perturbation obtained for $T = 20$ and $k = 0.6$. Note that it is recovered from equations (1.38). We have reported in figure 1.22 the longitudinal velocity of the initial

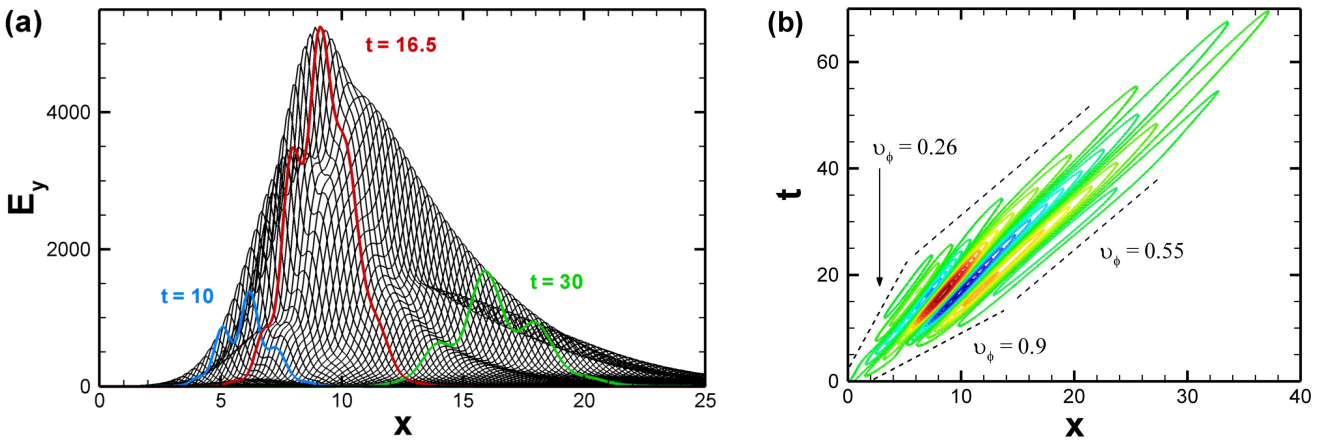


Figure 1.21: Spatial evolution of the wave packet associated with the $2D$ optimal perturbation. (a) The quantity $E_y(x,t)$ is depicted for every time step $\Delta t = 0.5$ as a function of x. (b) depicts the space-time $x-t$ diagram of the quantity $u(x, y = 1.05, t)$.

perturbation along with its evolved optimal state at $t = 20$. Figures 1.22(a,b) stand for a side view whereas figures 1.22(c,d) for a top view. Basically, we get very similar wave packets except that they are, by construction, sinusoidal in the transverse direction. Once more, the spatial streamwise separation between the optimal initial condition and its optimal evolved state is highlighted. The optimal condition is also inclined again the shear and amplified by both the Orr and the Kelvin-Helmholtz instabilities. The dynamics of the $3D$ perturbation seems rather analogous to the optimal $2D$ one. To investigate the role of the transverse velocity component, we have depicted in figure 1.23 the evolution of the energy gain $\gamma(t)$ of the $3D$ optimal perturbation as a function of time together with the contribution from its three velocity components. For instance its longitudinal velocity contribution is equal to

$$\gamma_u(t) = \frac{\int_\Omega \tilde{u}^2(t)dxdy}{\int_\Omega [\tilde{u}^2(0) + \tilde{v}^2(0) + \tilde{w}^2(0)]dxdy} \tag{1.39}$$

and so on for the other two. Interestingly, the w contribution is much smaller than the other two for short times; it is even one order of magnitude below for $t < 2$. By the way, it should be noticed that the v contribution, initially far below the u one, increases very quickly due the the Orr mechanism and the rolling up of the perturbation. For later times, the w component rises and even becomes predominant at $t \approx 17$. Finally, it becomes similar to the u one for larger times ($t > 35$). The exact underlying physical origin of this $3D$ amplifier mechanism has not been further investigated. Yet, due to the increase of the w component in the downstream boundary layer, the optimal final perturbation is reminiscent of the oblique waves possibly sustained in transitional boundary layers, see Monokrousos *et al.* (2010). One may thus attribute the dynamics associated with this $3D$ perturbation to an analogous mechanism.

1.5.2.3 Short time optimal dynamics

We deal here with the particular case of very short time horizons. More precisely, we are looking for the initial perturbation that optimizes the initial energy gain growth rate (or gradient). This analysis is equally performed for the $Re = 600$. Denoting the energy gain at time t by $\gamma(t) = \frac{E(\tilde{\boldsymbol{q}}(t))}{E(\tilde{\boldsymbol{q}}(0))}$, the relative instantaneous energy growth rate $\frac{1}{\gamma(t)}\frac{\partial \gamma(t)}{\partial t}$ may be easily derived into

$$\frac{1}{\gamma(t)}\frac{\partial \gamma(t)}{\partial t} = \frac{\langle \tilde{\boldsymbol{q}}(t), (\boldsymbol{A} + \boldsymbol{A}^\dagger)\tilde{\boldsymbol{q}}(t)\rangle}{\langle \tilde{\boldsymbol{q}}(t), \mathbf{B}\tilde{\boldsymbol{q}}(t)\rangle} \tag{1.40}$$

which may also be solved by considering the generalized eigenvalue problem associated with $(\boldsymbol{A} + \boldsymbol{A}^\dagger)$ and $\mathbf{B}$. In particular, we look for the optimal value of this ratio which also stands for the largest possible

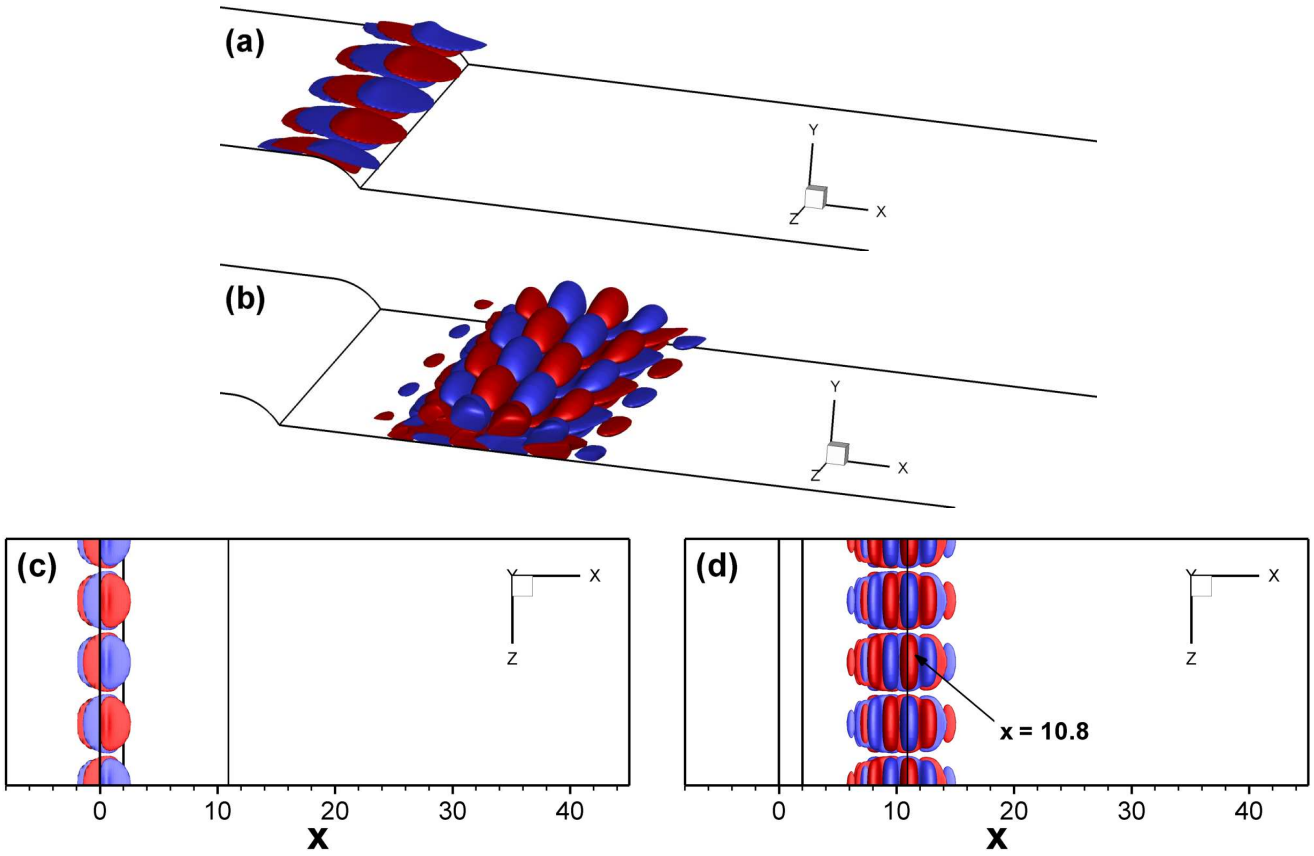


Figure 1.22: Optimal initial perturbation computed for $T = 20$ and $k = 0.6$. (a) Side view of the perturbation and (b) of its evolved state at time $t = 20$. (c,d) Same states displayed with a top view. Flow structures are displayed by their longitudinal velocity.

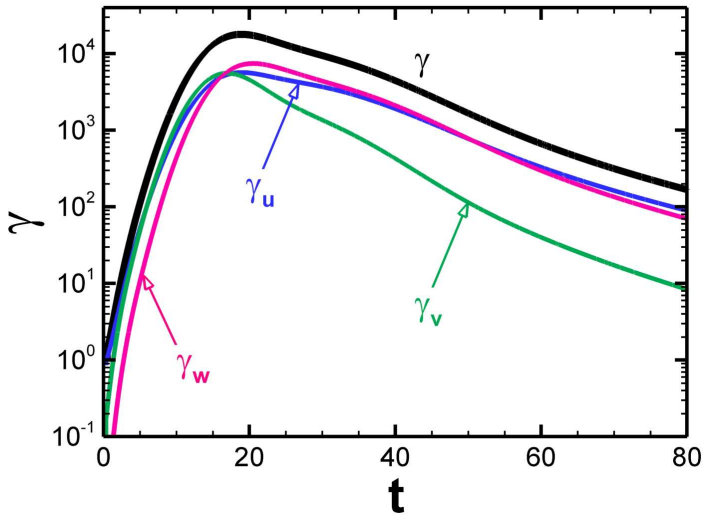


Figure 1.23: Evolution of the energy gain resulting from the $3D$ optimal perturbation along with its u, v and w contributions.

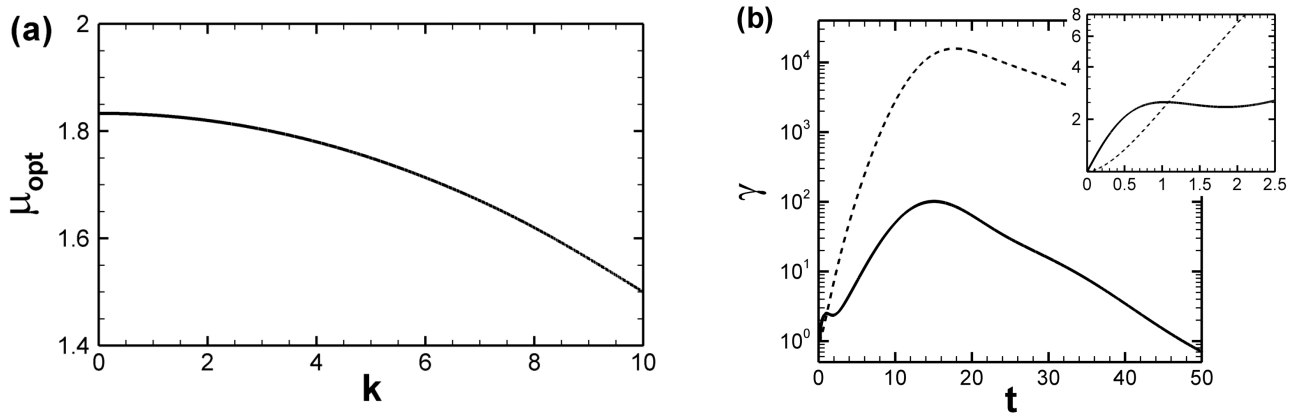


Figure 1.24: (a) Evolution of μ_{opt} as a function of k. (b) Energy gain $\gamma(t)$ of the time integrated optimally growing perturbation (solid line) and the optimal $2D$ perturbation obtained for $T = 18$ and $k = 0$ (dashed line).

instantaneous energy growth rate over all perturbations. This may be achieved by computing the largest eigenvalue μ_{opt} of the eigenvalue problem

$$(\boldsymbol{A} + \boldsymbol{A}^\dagger)\tilde{\boldsymbol{q}} = \mu \mathbf{B}\tilde{\boldsymbol{q}} \tag{1.41}$$

and the optimal associated eigenvector $\tilde{\boldsymbol{q}}$ reduces to the optimally growing perturbation whose energy may grow linearly such that $\gamma(t) = 1 + \mu_{opt}t$ at first order. We have represented in figure 1.24(a) the evolution of μ_{opt} as a function of k. It highlights that the maximum possible growth rate is achieved for $k = 0$, that is for $2D$ perturbations. Figure 1.24(b) depicts the kinetic energy gain resulting from this $2D$ perturbation by a solid line. For comparison, we have reported by a dashed line the gain of the optimal $2D$ perturbation computed for $T = 18$ and $k = 0$. Interestingly, the optimally growing solution is observed to be very quickly amplified in short times. However, the other optimal perturbation quickly becomes more energetic.
To illustrate the structure of the optimally growing perturbation, we have depicted in figure 1.25(a) its longitudinal velocity. Additionally figure 1.25(b) represents its evolved structure at time $t = 15$ where it is optimal, see figure 1.24(b). It is interestingly noticed that the perturbation displays a similar pattern to that of the optimal $2D$ perturbation, except that it has a slightly smaller streamwise wavelength and a different inclination against the shear. This initial disturbance then is then convected and amplified downstream as a similar wave packet. We conclude that the optimally growing perturbation extracts energy from the base flow by the same physical mechanisms, i.e. the Orr and Kelvin Helmholtz instabilities, but with a different contribution from both. More precisely, it preferentially promotes the Orr instability, indeed known as a short time mechanism.

1.5.2.4 Long time optimal dynamics

We now pay attention to very long time horizons. In the case where the flow is globally stable, the energy gain of any perturbation vanishes for long times. On the other hand, if the flow is globally unstable, any disturbance will lead to the growth of the most unstable global mode. We pose the question of which perturbation leads to maximum energy gain for long times. In other words, this stand for finding the initial condition which optimally promotes the most unstable global mode. We consider the flow at $Re = 600$ which has been shown to display a band of unstable transverse wave numbers. Furthermore, we choose the transverse wave number $k = 1.08$ which has been shown to be the most unstable one. For this set of parameters, their is only one unstable global mode, which is the one studied in section 1.4.4. We consider the initial condition $\boldsymbol{q}_0 = \hat{\boldsymbol{q}}_0\, e^{ikz}$ which is further decomposed in the basis of the global eigenmodes:

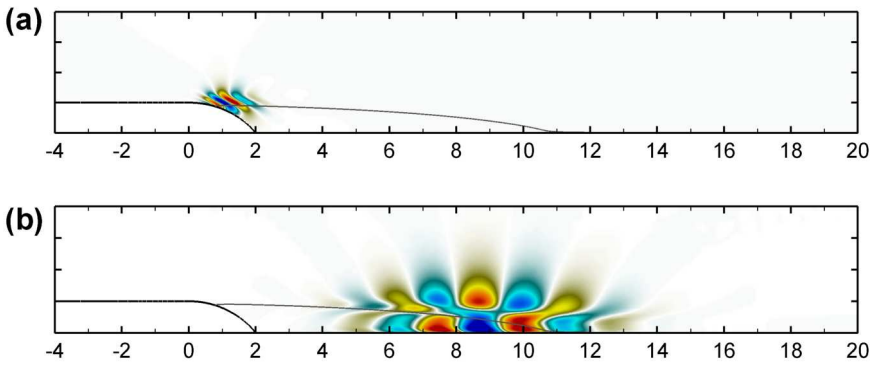


Figure 1.25: (a) Optimally growing perturbation computed for $k = 0$. (b) Evolution of this perturbation to the time $t = 15$.

$$\hat{\boldsymbol{q}}_0 = \sum_j \kappa_j \hat{\boldsymbol{q}}_j \tag{1.42}$$

where $\hat{\boldsymbol{q}}_j$ denotes the j^{th} eigenmode and κ_j are real coefficients. These coefficients can be obtained from the inner product of the initial condition and the adjoint eigenmodes $\hat{\boldsymbol{q}}_j^\dagger$ by $\langle \hat{\boldsymbol{q}}_0, \mathbf{B}\hat{\boldsymbol{q}}_j^\dagger \rangle = \kappa_j$. Consequently, the perturbation state $\boldsymbol{q}(t)$ resulting from the initial condition $\hat{\boldsymbol{q}}_0$ may be written as

$$\boldsymbol{q}(t) = \sum_j \langle \hat{\boldsymbol{q}}_0, \mathbf{B}\hat{\boldsymbol{q}}_j^\dagger \rangle \hat{\boldsymbol{q}}_j \, e^{\sigma_j t + ikz} \tag{1.43}$$

For sufficiently long times, only the component of the unstable global mode $\hat{\boldsymbol{q}}_{max}$ is non-null so that the perturbation dynamics are given by $\boldsymbol{q}(t) = \langle \hat{\boldsymbol{q}}_0, \mathbf{B}\hat{\boldsymbol{q}}_{max}^\dagger \rangle \hat{\boldsymbol{q}}_{max} \, e^{\alpha_{max} t + ikz}$. Next, the energy gain of this perturbation reduces to

$$\gamma(t) = \frac{\langle \hat{\boldsymbol{q}}_0, \mathbf{B}\hat{\boldsymbol{q}}_{max}^\dagger \rangle}{\langle \hat{\boldsymbol{q}}_0, \mathbf{B}\hat{\boldsymbol{q}}_0 \rangle} \, e^{2\alpha_{max} t} \tag{1.44}$$

which is maximum when $\hat{\boldsymbol{q}}_0 = \hat{\boldsymbol{q}}_{max}^\dagger$. This means that the adjoint eigenmode optimally enhances the energy of its direct eigenmode for long times.

The adjoint eigenvalue problem (1.28) is solved by the same numerical methods as for the direct one. The accurate computation of the adjoint modes can be checked by the bi-orthogonal relation (1.29) and the complex conjugate relation $\sigma_j^\dagger = \sigma_j^*$. In our case, we are only interested in the most unstable adjoint global mode, which is associated with the $3D$ periodic unstable mode previously presented. Figures 1.26(a,b,c) represent the three velocity components $\hat{u}^\dagger$, $\hat{v}^\dagger$ and $\hat{w}^\dagger$ respectively of the unstable adjoint eigenmode. They look quite similar to those of the direct eigenmode, see figure 1.15, except that they are localized upstream near separation. This common feature of adjoint modes have been related to the convective non-normality of the linearized Navier-Stokes operator, see Marquet *et al.* (2009). Additionally, the streamwise component is not as predominant as for the direct global eigenmode. This is further quantified in table 1.2 where we have reported the percentage of kinetic energy of the direct and adjoint global modes relative to the three velocity components. Such a clear separation between the direct and adjoint modes velocity components have also been reported in the literature as an effect of the lift-up non-normality, see Marquet *et al.* (2009). The present transfer of energy from the v and w components into the u one is indeed reminiscent of the lift-up mechanism.

Now, we pay attention to the effect of the adjoint global mode onto the base flow. Figure 1.26(a) shows that its u component is maximum near separation while figure 1.26(b) also leads to the same conclusion for the v component, except that is has an opposite sign. As a result, we conclude that the optimal mechanism promoting the direct global mode partly relies on the streamwise displacement of

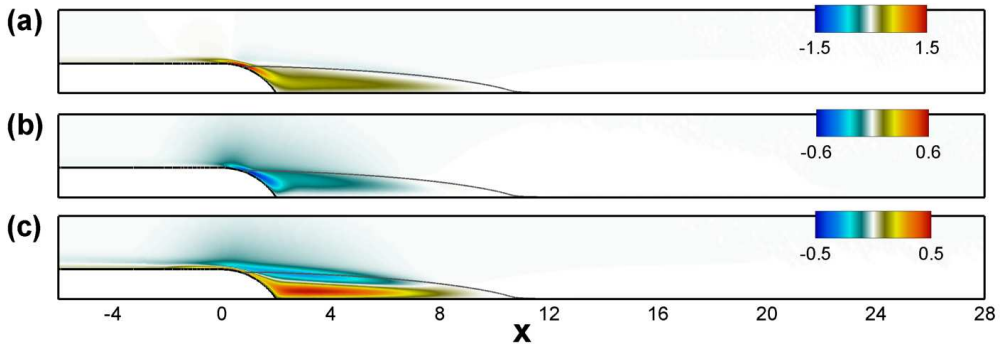


Figure 1.26: (a,b,c) Components $\hat{u}^\dagger$, $\hat{v}^\dagger$ and $\hat{w}^\dagger$ of the unstable adjoint eigenmode at $Re = 600$ and for $k = 1.08$.

% of energy in	u	v	w
Direct global mode	95.08	0.75	4.17
Adjoint global mode	56.34	11.06	32.6

Table 1.2: Percentage of the energy of the direct and adjoint global modes coming from the u, v and w components of the velocity.

the separation point varying sinusoidally along the span. The adjoint global mode acts on the base flow so as to displace the separation point upstream and downstream so that the recirculation region deflates and inflates as we move along the z axis. Noticeably, similar conclusions were found in the work by Marquet *et al.* (2009). For illustration, we have displayed a $3D$ view of the adjoint global mode in figure 1.27(a) by its longitudinal velocity. Furthermore, we have also displayed in figure 1.27(b) the sum of the base flow and adjoint mode (with arbitrary amplitude) iso longitudinal velocity. Similarly to the direct global mode, the amplitude of the adjoint mode is chosen such as to observe its effects on the base flow. The iso value -0.001 allows to clearly emphasizes the distortion of the separation line in the spanwise direction. Note that the $3D$ flow field is recovered by $\hat{\boldsymbol{q}}_r^\dagger \cos(kz) - \hat{\boldsymbol{q}}_i^\dagger \sin(kz)$ from the real part $\hat{\boldsymbol{q}}_r^\dagger$ and imaginary part $\hat{\boldsymbol{q}}_i^\dagger$ of the mode.

Lastly, we consider the energy gain provided by the adjoint global mode. The linearized Navier-Stokes equations (1.13) are solved with the adjoint mode as the initial condition. The resulting energy gain $\gamma(t)$ is depicted in figure 1.28 by a solid line. After a transient growth period for $t < 100$, the energy is observed to rise exponentially in time. Indeed, we showed that the energy increases in time with a factor $e^{2\alpha_{max}t}$ which leads to a linear increase in our log-scale plot with a slope $2\alpha_{max}$. The energy gain of the direct global mode has been computed analytically by $\gamma(t) = e^{2\alpha_{max}t}$ and represented by a dashed line. For large times, the parallelism of the two lines highlights the exponential increase in time of both the adjoint and direct global mode. By construction, the adjoint global mode is the perturbation maximizing the difference $\Delta\gamma$ (represented in figure 1.28) between the two large time energy gains.

1.5.3 Receptivity to harmonic forcing

1.5.3.1 Problem formulation and governing equations

Until now, we investigated the optimal perturbations so as to optimize the energy of the flow at a given time horizon. We now consider another way of analyzing the flow amplifier behavior in the frequency domain. It consists in studying and quantifying the flow response resulting from a permanent harmonic forcing. More precisely, we which to known what is the forcing field, for a frequency excitation ω, which leads to a maximum flow response kinetic energy. Such a study is commonly referred to as a receptivity analysis. The frequency ω stands for a critical parameter of the study, just like T in the previous tem-

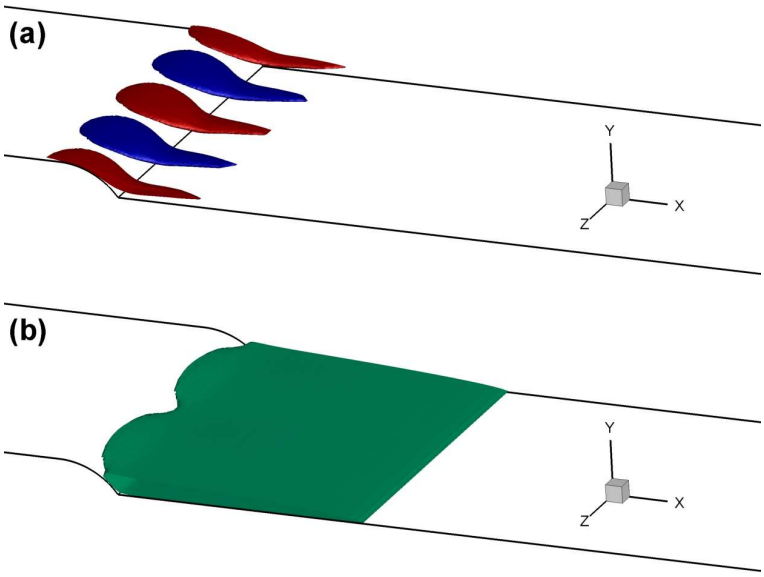


Figure 1.27: (a) Three-dimensional view of the unstable adjoint global mode longitudinal velocity. (b) Sum of the base flow and the adjoint global mode (with arbitrary amplitude) visualized by the iso longitudinal velocity -0.001.

poral analysis. Furthermore, we similarly decompose the $3D$ perturbation into transverse waves, as in equation (1.12) so that the wave number k is also a parameter of the study.

The dynamics of the perturbation $\boldsymbol{q}(x,y,z,t) = (u,v,iw,p)$ is now supposed to be driven an additional excitation. The three momentum conservation equations are forced by the components f_u, f_v and f_w of the $3D$ forcing term $\mathbf{f}(x,y,z,t) = (f_u, f_v, if_w, 0)$. Note that the forcing of the last equation governing the mass conservation has not been investigated for simplicity. We similarly resort to a decomposition into transverse waves to express the forcing as

$$\mathbf{f}(x,y,z,t) = \frac{1}{2}\left[\,\tilde{\boldsymbol{f}}(x,y,t)\,e^{ikz} + c.c.\,\right] \tag{1.45}$$

where, we introduce the $2D$ component $\tilde{\boldsymbol{f}}(x,y,t) = (\tilde{f}_u, \tilde{f}_v, i\tilde{f}_w, 0)$ of the forcing associated with the wave number k. The dynamics of the perturbation $\tilde{\boldsymbol{q}}(x,y,t)$ are then governed by the linearized Navier-Stokes equations with an added forcing term:

$$\mathbf{B}\,\partial_t\tilde{\boldsymbol{q}} = \boldsymbol{A}\,\tilde{\boldsymbol{q}} + \mathbf{B}\tilde{\boldsymbol{f}}(t) \tag{1.46}$$

Importantly, it should be mentioned at this stage that the forcing term and the perturbation are supposed to have the same order of magnitude in order to recover this linearized equation. According to the hypothesis introduced in section 1.2.3, the forcing amplitude is thus supposed to be low compared to the base flow. Due to the linear nature of the equation, we can further decompose both the forcing and the perturbation into time harmonic components by $\tilde{\boldsymbol{q}}(x,y,t) = \hat{\boldsymbol{q}}(x,y)\,e^{i\omega t}$ and $\tilde{\boldsymbol{f}}(x,y,t) = \hat{\boldsymbol{f}}(x,y)\,e^{i\omega t}$ where ω stands for a real frequency while $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{f}}$ may be complex. As a result, the response $\hat{\boldsymbol{q}}$ may be obtained from the forcing $\hat{\boldsymbol{f}}$ by the following expression

$$(i\omega\mathbf{B} - \boldsymbol{A})\hat{\boldsymbol{q}} = \mathbf{B}\hat{\boldsymbol{f}} \tag{1.47}$$

or also, provided that this system is invertible, by $\hat{\boldsymbol{q}} = (i\omega\mathbf{B} - \boldsymbol{A})^{-1}\mathbf{B}\hat{\boldsymbol{f}}$. To simplify the notation, we introduce the resolvent operator defined by $\mathbf{R}(\omega) = (i\omega\mathbf{B} - \boldsymbol{A})^{-1}\mathbf{B}$ which directly links the harmonic forcing to its harmonic response by

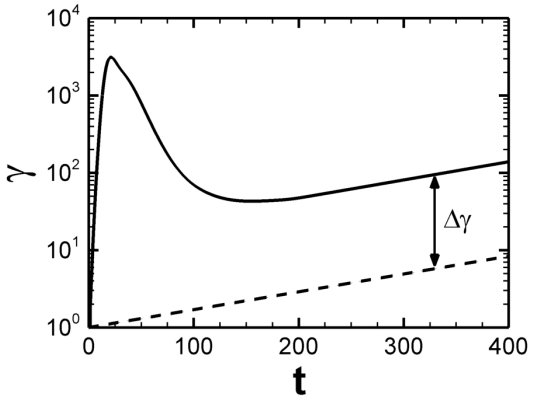


Figure 1.28: Evolution of the energy gain $\gamma(t)$ of the perturbation resulting from the adjoint (solid line) and direct (dashed line) global modes.

$$\hat{\boldsymbol{q}} = \mathbf{R}\,\hat{\boldsymbol{f}} \tag{1.48}$$

Now, our optimal perturbation analysis reads: what is the forcing $\hat{\boldsymbol{f}}$ which optimizes the kinetic energy of the response $\hat{\boldsymbol{q}}$ for the considered values of ω and k. To solve this problem, the energy gain is expressed in the form

$$\frac{E(\hat{\boldsymbol{q}})}{E(\hat{\boldsymbol{f}})} = \frac{\langle \mathbf{R}\,\hat{\boldsymbol{f}}, \mathbf{B}\mathbf{R}\,\hat{\boldsymbol{f}} \rangle}{\langle \hat{\boldsymbol{f}}, \mathbf{B}\hat{\boldsymbol{f}} \rangle} = \frac{\langle \hat{\boldsymbol{f}}, \mathbf{R}^\dagger\mathbf{B}\mathbf{R}\,\hat{\boldsymbol{f}} \rangle}{\langle \hat{\boldsymbol{f}}, \mathbf{B}\hat{\boldsymbol{f}} \rangle} \tag{1.49}$$

where we introduced the adjoint of the resolvent operator $\mathbf{R}^\dagger$. Next, we conclude that the optimal forcing and optimal energy gain may be found as the largest eigenvector and eigenvalue of the following generalized eigenvalue problem

$$\mathbf{R}^\dagger\mathbf{B}\mathbf{R}\,\hat{\boldsymbol{f}} = \lambda\mathbf{B}\hat{\boldsymbol{f}} \tag{1.50}$$

the leading eigenvalue, denoted by λ_{opt}, along with its associated optimal forcing are computed by using the shift and invert Arnoldi method described in section 1.4.2. Finally, note that the $3D$ time dependent perturbation may be recovered in the end by

$$\boldsymbol{q}(x,y,z,t) = [\hat{\boldsymbol{q}}_r\cos(kz) - \hat{\boldsymbol{q}}_i\sin(kz)]\cos(\omega t) - [\hat{\boldsymbol{q}}_i\cos(kz) + \hat{\boldsymbol{q}}_r\sin(kz)]\sin(\omega t) \tag{1.51}$$

where $\hat{\boldsymbol{q}}_r$ and $\hat{\boldsymbol{q}}_i$ denotes the real and imaginary parts of $\hat{\boldsymbol{q}}$. Also, the $3D$ time dependent forcing $\mathbf{f}$ can be recovered by the equivalent expression.

1.5.3.2 Results

As previously done, we expose our results at $Re = 600$. The eigenvalue problem (1.50) has been solved for the parameters $k = [0;8]$ and $\omega = [0;8]$. A particular attention should be paid to the wave number k since the operator $\boldsymbol{A}$ is unstable in the interval $k = [0.675;1.676]$. For unstable wave numbers, the perturbation $\hat{\boldsymbol{q}}$ may not be physically interpreted as the long time response of the flow to the forcing $\hat{\boldsymbol{f}}$, even though the matrix $(i\omega\mathbf{B} - \boldsymbol{A})$ is invertible.
We have represented in figure 1.29(a) a $2D$ plot of the optimal energy gain λ_{opt} as a function of k and ω. Furthermore, we emphasized the unstable wave number interval by two dashed lines. This plot once more highlights the strong potential of the flow to amplify low wave number perturbations. For instance, the maximum gain reached by a $2D$ forcing (with $k = 0$) is reached for $\omega = 0.75$ and equals $\lambda_{opt} = 2.24\,10^6$. Higher values of the gain may also be obtained for $3D$ forcing. The highest value $\lambda_{opt} = 3.05\,10^6$ is

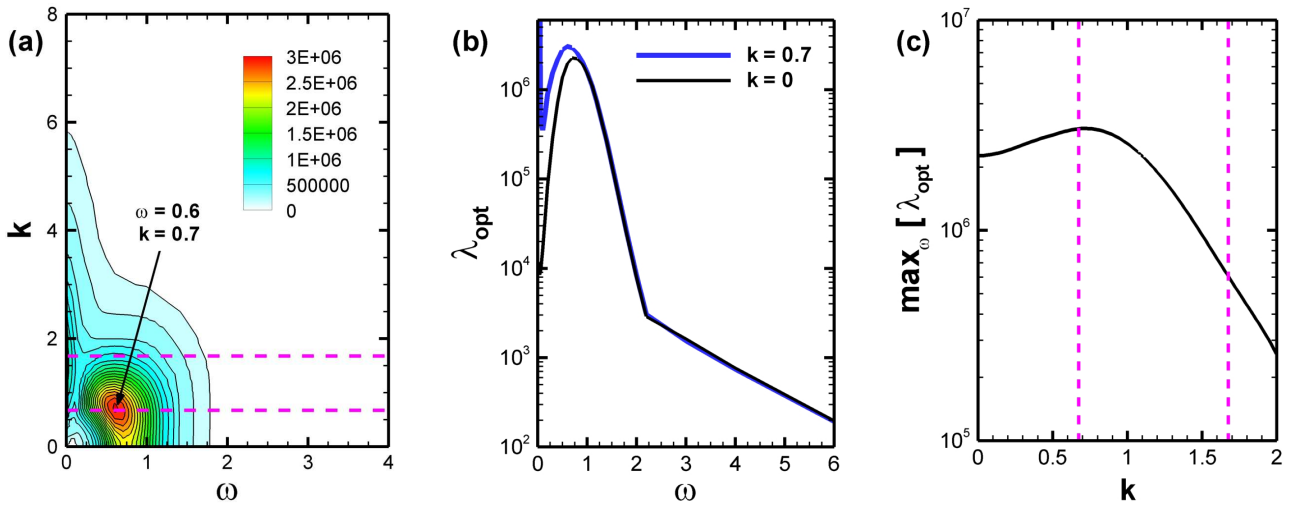


Figure 1.29: (a) Optimal energy gain λ_{opt} as a function of ω and k in a $2D$ plot. (b) depicts the evolution of λ_{opt} as a function of ω for $k = 0$ and $k = 0.7$. (c) stands for the evolution of the maximum value of λ_{opt} over ω as a function of k. These results have been computed for $Re = 600$. The band of unstable wave numbers has been emphasized by dashed lines.

obtained for $\omega = 0.6$ and $k = 0.7$. For a more quantitative representation, we show in figure 1.29(b) the evolution of λ_{opt} as a function of ω for the fixed values $k = 0$ and $k = 0.7$. Both exhibit a similar broad band frequency response, noticeably both are equal for the highest frequencies. Yet, the energy gain for $k = 0.7$ tends to infinity as ω tends to 0, which comes from the instability of $\boldsymbol{A}$ (since $k = 0.7$ is an unstable wave number). The influence of the wave number on the optimal energy gain is further illustrated in figure 1.29(c) where we have represented the evolution of the value $\max_\omega[\lambda_{opt}(k,\omega)]$ as a function of k. Note that the unstable band of wave numbers has also been emphasized by dashed lines. Interestingly, it leads to similar conclusions to those of the optimal temporal perturbation analysis: (i) the optimal energy gain quickly vanishes for small transverse wave lengths, (ii) its maximum is found for the largest wave lengths and (iii) the optimal gain weakly depends on the wave number for $k < 1$.

To illustrate the optimal forcing distributions and their associated responses, we start by considering the solutions obtained for $k = 0$. Figure 1.30 displays the longitudinal velocity of the optimal forcing and response computed for three different frequencies. More precisely, figures 1.30(a,b) represent the solutions for $\omega = 0.25$, figures 1.30(c,d) for $\omega = 0.75$ and figures 1.30(e,f) for $\omega = 2$. In all the cases the optimal harmonic forcing is localized upstream near separation and displays an inclined structure along the shear so as to exploit the Orr mechanism. On the other hand, the associated optimal responses extend farther downstream and display sustained amplified wave packets. Note that the spatial separation between the optimal forcing distributions and their optimal response is known as common feature of non-normal fluid systems resulting from the convective non-normality. Both the optimal forcings and responses outline smaller flow structures for higher frequencies. Similarly, they display much bigger structures for lower frequencies that spatially extend much farther upstream (for the forcing) and downstream (for the response). These results are consistent with the Kelvin-Helmholtz instability illustrated in the previous temporal analysis since high-frequency instabilities are observed where the shear layer is the thinnest whereas the lower ones are observed in the downstream boundary layer where the shear is the thickest.

We now turn our attention to the $3D$ optimal forcings and responses. In the range of wave numbers considered, the real part $\hat{\boldsymbol{q}}_r$ and imaginary part $\hat{\boldsymbol{q}}_i$ of the flow responses are observed to display very similar wave packets, except that they are out of phase in the streamwise direction. Furthermore, this also holds for the real part $\hat{\boldsymbol{f}}_r$ and imaginary part $\hat{\boldsymbol{f}}_i$ of the forcings. This spatial shift is such that, by computing the three-dimensional perturbation by equation (1.51), we get transverse wave packets. This is illustrated in figure 1.31 where we show the longitudinal velocity of a $3D$ optimal harmonic forcing

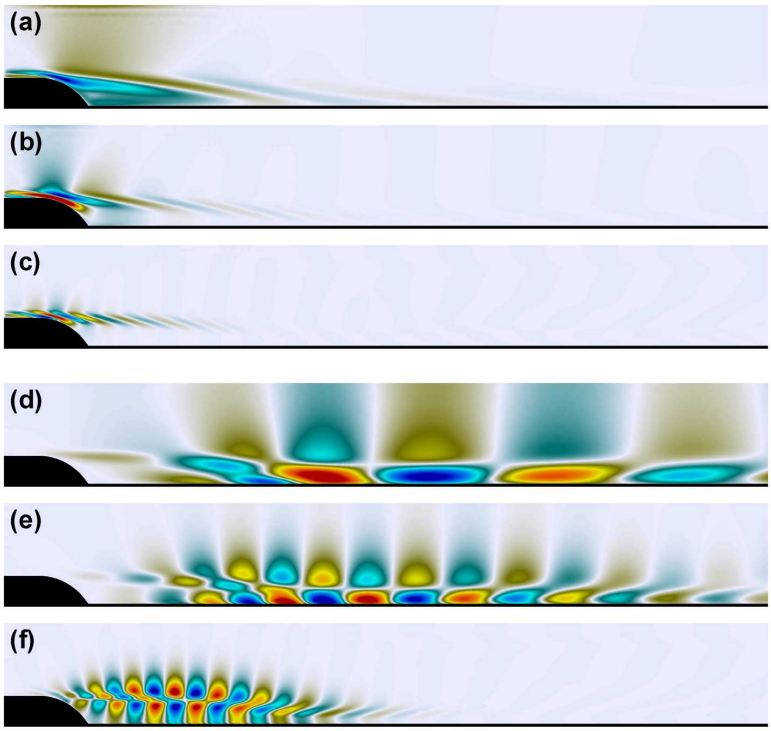


Figure 1.30: (a,b,c) Two-dimensional optimal forcing distributions $\hat{\boldsymbol{f}}$ computed for $\omega = 0.25$, $\omega = 0.75$ and $\omega = 2$ respectively. (d,e,f) Optimal associated responses $\hat{\boldsymbol{q}}$ computed for the same frequencies. Flow structures are displayed by their longitudinal velocity.

along with its associated optimal response. A side view is provided in figures 1.31(a,b) and a top view in figures 1.31(c,d). This particular perturbation has been computed for the set of parameters $\omega = 0.6$ and $k = 0.7$. We interesting observe the appearance of transverse waves that form oblique wave packets. Note that this latter result is much more noticeable for the response than for the forcing. This confirms our previous statement about the possible amplification of oblique waves in the downstream boundary layer. Note that a similar analysis has been reported in the work by Monokrousos *et al.* (2010) on a plat plate boundary layer. It should be recalled that the flow is globally unstable for the wave number $k = 0.7$ at $Re = 600$. However, the $3D$ flow structures computed for other stable wave numbers, or for lower Reynolds number such that the flow is stable, are found to display similar oblique wave packets. Varying the parameters k and ω only alters their transverse wave length $l_z = 2\pi/k$ and their streamwise wave length $l_x = 2\pi v_\phi/\omega$ (where v_ϕ has been introduced as the phase velocity in the streamwise direction). The propagation direction of the oblique wave may then be characterized by the angle $\phi = \arctan(l_x/l_z)$ as shown in figure 1.31(d).

1.5.3.3 Influence of the Reynolds number

To conclude this section we investigate the effect of the Reynolds number on the amplifier behavior of the flow. To do so, we perform the previous receptivity analysis for $k = 0$ while varying the Reynolds number. We have represented in figure 1.32(a) the evolution of the optimal energy gain λ_{opt} as a function of the frequency ω for $Re = 200$, 400, 600, 800 and 1000. These results emphasize two interesting points: (i) the optimal energy gain increases with the Reynolds number whatever the frequency and (ii) the peak of the maximum energy gain is shifted to higher frequencies for higher Reynolds numbers. These two points are further quantified in figures 1.32(b,c). We have represented the peak value $\max_\omega[\lambda_{opt}(k = 0, \omega)]$ of the optimal gain as a function of the Reynolds number in figure 1.32(b). It highlights an exponential increase of the maximum energy gain as a function of Re which is emphasized by the log-scale plot.

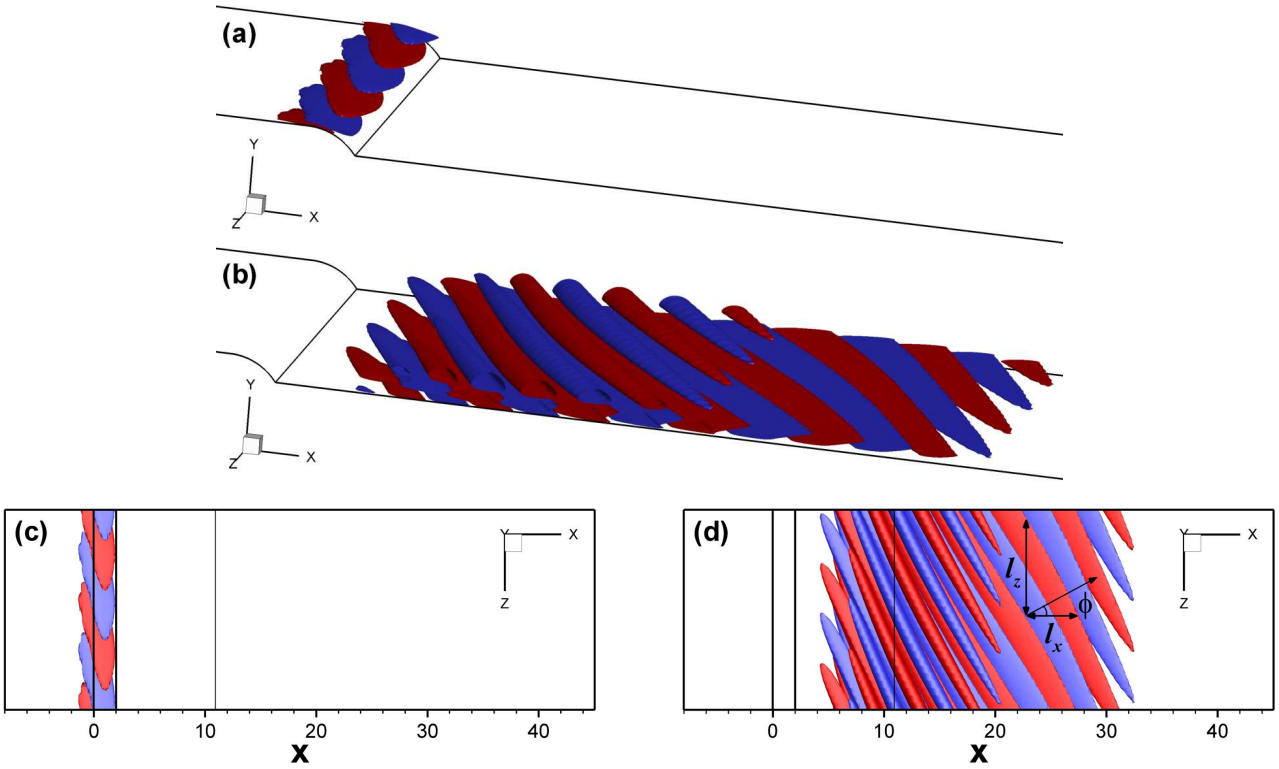


Figure 1.31: Three-dimensional view of a $3D$ optimal forcing (a,c) along with its associated optimal response (b,d). Figures (a,b) display a side view and (c,d) a top view. These flow structures have been computed for the parameters $\omega = 0.6$, $k = 0.7$ and are displayed by their longitudinal velocity.

Such an exponential increase of the flow amplifier behavior is a common feature of open shear flows, see Schmid & Henningson (2001). In particular, it reveals the strong receptivity of the flow and its potential to substantially amplify perturbations, even for moderate Reynolds numbers. Now we look for the frequency ω_{opt} at which the maximum energy gain is reached. This quantity is represented in figure 1.32(c) as a function of the Reynolds number. As previously noticed, the optimal frequency increases for higher Reynolds numbers which is naturally attributed to the thinning of the base flow shear layer. This result is thus consistent with the Kelvin-Helmholtz mechanism. For information, the scale of the associated Strouhal numbers $St_h = \frac{\omega_{opt} h}{2\pi U_\infty}$ have been displayed in a second axis of the plot.

1.6 Overall assessment and concluding remarks

We studied in this chapter the global stability of $3D$ perturbations about the $2D$ base flow over a backward-facing step. An asymptotic stability analysis has revealed the occurrence of a transversally periodic and stationary unstable global mode at the critical Reynolds number $Re_c = 526$. Owing to the literature this mode is known to comes from a centrifugal instability and has also been linked to a lift-up mechanism. On the other hand a short time stability analysis showed the ability of the flow to substantially amplify perturbation by transient growths. The optimal perturbations have been exposed as amplifying wave packets coming from a coupled Orr and Kelvin-Helmholtz instability. The three-dimensional optimal growth mechanism has furthermore been attributed to an inclination of the wave packets in the z direction into oblique waves.
As a summary of these different transition scenarios, we have represented in figure 1.33 the evolution of the time integrated energy gain $\gamma(t)$ for some of the considered optimal perturbations. It includes: (i) the adjoint unstable global mode, (ii) the optimal growth rate perturbation, (iii) the $2D$ optimal perturbation and (iv) the $3D$ optimal perturbation. The representation of all these initial conditions

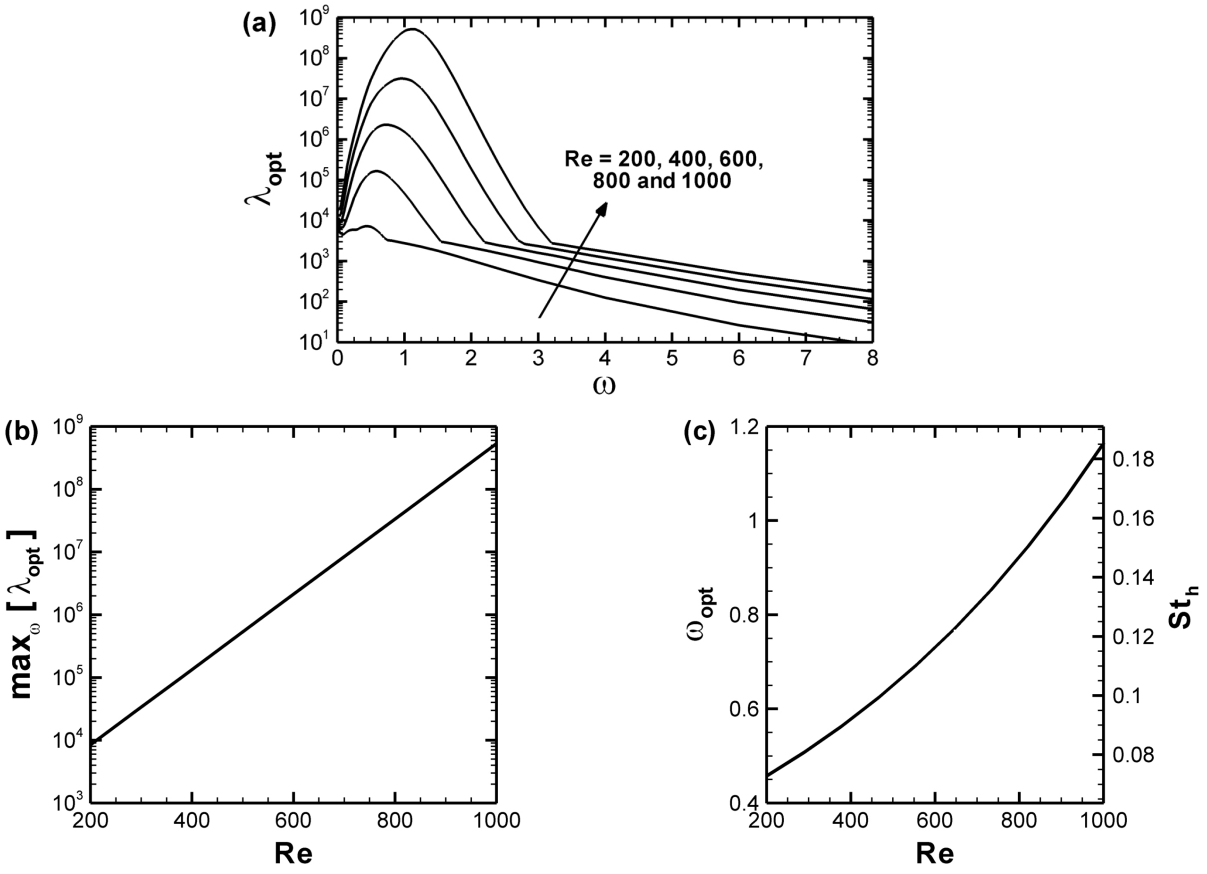


Figure 1.32: (a) Evolution of the optimal energy gain λ_{opt} as a function of ω for $k = 0$ and $Re = 200$, 400, 600, 800 and 1000. (b) represents the optimal energy gain peak $\max_\omega[\lambda_{opt}(k = 0, \omega)]$ as a function of the Reynolds number. (c) depicts the optimal frequency ω_{opt} and associated Strouhal number St_h as a function of the Reynolds number.

clearly allows to compare their energy amplification. For instance, the optimally growing perturbation is observed to be the less energetic one. Regarding the unstable adjoint mode, it yields the maximum energy growth for the largest times but is suboptimal for short times. On the other hand, the $2D$ and $3D$ optimal perturbations lead to a much larger optimal energy gain for short times which exceeds those of the other initial conditions by several orders of magnitude. The difference between the contributions of the $2D$ and $3D$ optimal perturbations is only noticeable for larger times. This plot emphasizes the optimal transition scenario of the flow as a quasi $2D$ amplification of wave packets in the shear layer, see section 1.5.2.2. The most unstable physical mechanism is thus attributed to the Kelvin-Helmholtz instability occurring in the recirculation bubble.

Our results should be carefully compared to experiments since they are based on a number of important assumptions such as: (i) the two-dimensionality of the base flow and mostly (ii) the linear dynamics of the perturbations. As previously mentioned, the recirculation length is obviously not meant to rise indefinitely as shown in figure 1.11(a) and figure 1.12. Owing to the substantial ability of the flow to amplify perturbations, non-linear dynamics are expected to be observed. In fact, the limit at which unsteadiness first appear is not predictable since it only depends on external factors such as the level of background noise, residual turbulence, surface roughness, etc. It should also be mentioned that, according to preliminary results, the three-dimensional confinement induced by lateral walls seems to have a large impact on the stability of the flow. Indeed, the walls creates a drag force that tends to decrease the recirculation length, which stabilizes the flow. Performing a global stability analysis of a $3D$ base flow, with lateral walls, would stand for the natural next step to investigate this aspect of the

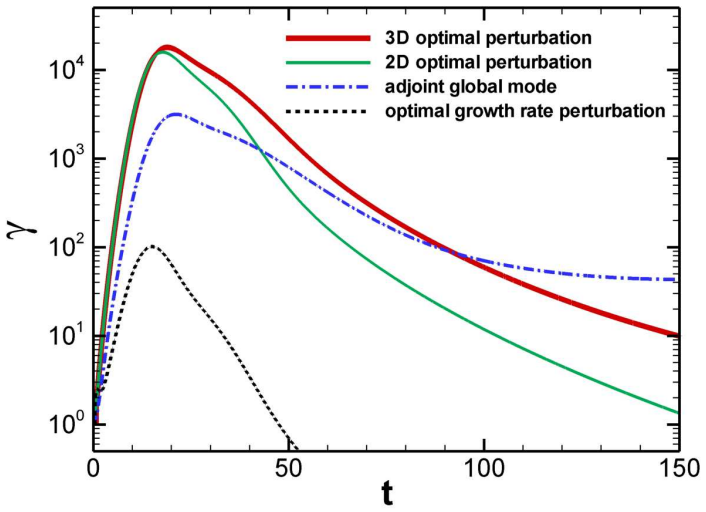


Figure 1.33: Evolution of the time integrated energy gain $\gamma(t)$ for various initial conditions. The black dashed line stand for the adjoint unstable global mode. The blue dash-dotted line represents the perturbation leading to the maximum initial growth rate (which is $2D$ according to our study). Finally, the green and red solid lines stand for the $2D$ and $3D$ optimal perturbations, which have been computed for $(k = 0\,; T = 18)$ and $(k = 0.6\,; T = 20)$ respectively).

dynamics. Interestingly, a three-dimensional and stationary structure reminiscent of our unstable global mode has been reported in the experimental work by Beaudoin *et al.* (2004) on a low Reynolds number backward-facing step flow. Yet, the three-dimensional instabilities found in our work, which have very large transverse wave lengths, may hardly be observed for transversally confined setups.

Our analysis recovers the Kelvin-Helmholtz instability and the associated frequencies have consistent orders of magnitude. For instance, the optimal frequencies from our receptivity analysis are found to increase with the Reynolds number and reach $St_h = 0.185$ at $Re = 1000$. To give an idea, a Kelvin Helmholtz frequency corresponding to $St_h \approx 0.22$ has been found in the experiments by Duriez (2009) at $Re = 1314$. Interestingly, we recover the wave packet phase speed $v_\phi = 55$ which has also been observed in the experiments by Duriez (2009) in the range $Re = [1000; 5500]$.

It should be noticed that the optimal frequencies ω_{opt} computed here may not be directly compared to those of experiments due to the non-linear effects. In our case, the optimal frequency would only be observed at the early development of unsteadiness in a laminar flow. The optimal linear responses of the flow are found in the vicinity of the reattached boundary layer, where the shear layer is relatively thick. Now, in a more complex (and more non-linear situation) the unsteadiness associated with the Kelvin-Helmholtz instability would be characterized by higher Strouhal numbers and structures localized in the very vicinity of the separation point, where the shear layer is much thinner. What could be done to partially alleviate this limitation is to restrict the optimization procedure in the upstream part of the computational domain, where the linear assumption is valid. This would possibly yield the optimal Kelvin-Helmholtz frequency, even if the downstream dynamics are strongly non linear (as in a majority of experiments).

We did not observe in our analysis any low-frequency unsteadiness or "flapping" of the recirculation bubble. Yet, this phenomenon has been linked recently (Ehrenstein & Gallaire (2008*a*); Cherubini (2010)) to a superposition of unstable global modes which gives rise to a low frequency beating due to alternate mode cancelations. In our case, we did not find any unsteady unstable global modes even for much higher Reynolds numbers (up to 2500). We think that unsteady unstable global modes are more likely to be observed (at lower Reynolds numbers) in backward-facing step flows when the incoming boundary layer thickness is relatively large compared to the step height (as in Åkervik *et al.* (2007); Ehrenstein & Gallaire (2008*a*); Cherubini (2010)). A more accurate assessment on this point

would stand for an interesting perspective. Finally, our linear analysis clearly cannot predict non linear mechanisms such as the vortex merging. To alleviate this limitation, a possible way to extend our work would be to include the additional non linear term $\boldsymbol{u} \cdot \nabla \boldsymbol{u}$ in the linearized Navier-Stokes equations (1.4). Indeed, such an extension still allows to perform optimal perturbation analysis, as recently done by Zuccher *et al.* (2006); Cherubini *et al.* (2010*a*). Another possible way would be to compute Direct Numerical Simulations (DNS) to recover the fully $3D$ non-linear flow dynamics, see Cherubini *et al.* (2010*b*); Cherubini (2010).

Interlude

In chapter 1 we analyzed the linear dynamics of a two-dimensional backward-facing step flow. Two major instability mechanisms have been highlighted: (i) the growth of a stationary and transversally periodic unstable global mode and (ii) the transient amplification of upstream perturbations, in the form of wave packets, in the shear layer. The first mechanism was observed to have to a very slow energy growth rates. However, the second one is found to reach substantially high energy gain values for short times. The optimal perturbation has been found to exploit both the Orr and Kelvin-Helmholtz instabilities. It seems that the unstable global mode is quite difficult to observe in experiments, see Beaudoin *et al.* (2004). Suppose we have a quasi two-dimensional flow, at a supercritical Reynolds number, it would be rather long to observe the unstable mode owing to its low energy growth rate. Instead, in a majority of the experiments, the Kelvin-Helmholtz instability is more easily (and often) observed as the first step of the transition to turbulence. The results presented in chapter 1 are in qualitative agreement with experimental results. Yet, including the non-linear effects and studying three-dimensional base flows are obviously the next steps toward a more quantitative comparison. Since we did not have the necessary numerical tools to perform such a study, this issue has not been further investigated.

Now that the linear dynamics of the flow have been thoroughly studied, we turn our attention to its modeling. Owing to our previous results, the most critical dynamics are attributed the transient amplification of perturbations along the shear layer. In particular, we showed that the most energetic perturbations were characterized by very large transverse wave lengths. As a result we choose, for simplicity, to restrict our analysis to the case of two-dimensional perturbations. The purpose behind the next three chapters is to design efficient Reduced Order Models (ROMs) that capture the predominant two-dimensional linear dynamics of the flow. In other words, we wish to find low rank operators that approximate the original linearized Navier-Stokes operator. To do so we choose to use the so-called "projection methods". It consists in projecting the original dynamical operator onto a reduced set of flow structures to obtain the ROMs. Naturally, the choice of the projection basis is fundamental when one attempts to construct efficient ROMs. Mostly, different projection bases may be used depending on what flow properties need to be captured.

As a first candidate, we considered the set of the least damped global modes. Indeed, at the beginning of the thesis, this choice seemed to be natural for several reasons. First, global modes are physically significant since they are solution of the linearized Navier-Stokes equations. In particular, they are associated with a growth rate and a frequency which are physically significant and concrete quantities. The part of the eigenspectrum corresponding to the least damped modes can be computed through the iterative Arnoldi technique. Then, the idea is to select the most unstable modes and truncate those with the lowest growth rates. It is then reasonable to expect that the most stable modes, whose amplitude decreases exponentially in time, will only contribute to a minor part of the dynamics. A second motivation for using global modes relies on some results exposed in the literature. For example, it has been demonstrated in the work by Ehrenstein & Gallaire (2005) that a finite number of two-dimensional global modes is able to capture the convective instability behavior of the flat-plate boundary layer. Later, it has been shown by Åkervik *et al.* (2007) that the sum of some global eigenmodes is able to describe

the dynamics of a localized disturbance in a high-aspect-ratio smooth cavity flow. By projection on the least stable global eigenmodes, a low-order model of the flow is obtained and used to design a linear-quadratic-Gaussian controller. This controller is also shown to efficiently damp out the flow global oscillations. It is also worth mentioning the review article by Henningson & Åkervik (2008) and the analysis by Ehrenstein & Gallaire (2008*b*) of the low-frequency unsteadiness called "flapping" by means of a global modes decomposition.
There is hence some evidence that a moderate number of two-dimensional global modes can potentially capture the transient dynamics of wall-bounded open flows. This supports the idea that global modes may provide new possibilities for model reduction with flow-control applications. For these reasons, we attempted to model the dynamics of our backward-facing step flow by means of global modes. Unfortunately, we found in this case that the global modes were unable to recover the convective instability of the flow. To understand this failure, we considered other flow configurations such as the smooth cavity studied by Åkervik *et al.* (2007), a square cavity flow and a lid-driven cavity. We did not have the time to make a complete study and go into much detail but we have reported our findings in appendix B. This appendix should be read as an extension of chapter 1 and as a link motivating the next chapters. To summarize, it is observed that the ability of the global modes to model the dynamics strongly depends on the flow configuration. The main reason for this failure is attributed to the very high non-normality of the global modes. As shown by Trefethen & Embree (2005), the most stable modes are extremely ill-conditioned and thus can barely be computed. One the other hand, depending on the flow configuration, the most stable modes are often seen to play an important role in the transient dynamics. For instance, a recent work by Barbagallo *et al.* (2009), dealing with linear state-space flow control, indicated that the most important global modes to be used in the design of an efficient low-order controller are not the most unstable ones. We refer interested readers to the recent review articles by Bagheri *et al.* (2009*c*) and Sipp *et al.* (2010).

As a result of this failure, we considered another approach. The idea is to define some inputs and outputs on the system and look for models that capture the input-output dynamics. A powerful technique, known as Balanced Truncation has been introduced by Moore (1981) and enables to reach such a goal. It is conceptually based on the concepts of controllability and observability. More precisely, controllable flow states may be easily triggered by the inputs whereas observable flow states yield an important contribution to the outputs. In this context, we start by considering a single input and a single output on the backward-facing step flow. The ability of the most controllable modes and of the most and equally controllable/observable modes (also called balanced modes) to capture the input-output dynamics is investigated in the next chapter. A particular attention is paid to the technique used to compute these modes. We show how the use harmonic flow responses, derived from harmonic excitations at the input or at the output, allows to design efficient ROMs. It should also be mentioned at this stage that the unstable linear dynamics of a square cavity flow are also considered in the next chapter to illustrate the ability of the procedure to handle unstable fluid systems. The results presented in the next chapter have been reported in an article that have been submitted to *Physics of Fluids*.

Chapter 2

Model reduction of the input-output dynamics by means of frequential snapshots

This chapter deals with model reduction of high-order linear systems. An alternative method to approximate Proper Orthogonal Decomposition and Balanced Truncation is exposed here within the framework of the incompressible Navier-Stokes equations. The method of snapshots used to obtain low-rank approximations of the system controllability and observability Gramians is carried out in the frequency domain. Model reduction is thus performed using flow states that are long-time harmonic responses of the flow to given forcings, we call them frequential snapshots. In contrast with the recent works using time-stepping approach, restricted to stable systems, this one can always be computed for systems without marginal modes while it reduces to the same procedure for stable systems. We show that this method is efficient to perform POD and BPOD reduced-order models in both globally stable and unstable flows through two numerical examples: the flow over a backward-facing step and the flow over a square cavity. The first one is a globally stable flow exhibiting strong transient growths as a typical noise amplifier system while the second is a globally unstable flow representative of an oscillator system. In both cases, it is shown that the frequency-based snapshot method yields reduced-order models that efficiently capture the input-output behavior of the system. In particular, regarding the unstable cavity flow, our resulting unstable reduced-order models possess the same unstable global modes and stable transfer functions as those of the full system.

2.1 Introduction

During the last decade, linear state space flow control has become a new and promising research subject, see Kim & Bewley (2007); Sipp *et al.* (2010). Originally introduced for flows governed by linear instability mechanisms (Bewley & Liu (1998); Bewley (2001)), it brought control theoretical tools (Burl (1999); Zhou *et al.* (2002)) such as optimal control within the reach of the fluid mechanics community. Optimal flow control obviously displays attractive features. One may refer to previous works such as the control of a transitional boundary layer (Högberg & Henningson (2002); Chevalier *et al.* (2007)), a transitional channel flow (Högberg *et al.* (2003)) or also turbulent channel flows (Moin & Bewley (1994); Joshi *et al.* (1997)). Yet, as soon as complex or more realistic flows are considered, the direct application of many optimal flow control tools is no more computationally tractable. One may then resort to a modeling of the dynamics by reduced-order models (ROMs) to alleviate this problem. In this context, modeling the dynamics between some particular inputs, such as the actuators, and some particular outputs, such as the sensors, may be sufficient to perform effective optimal flow control. For instance, let us mention the application of LQG closed-loop flow control to more complex flows, see Åkervik *et al.* (2007); Bagheri

et al. (2009*b*); Ahuja & Rowley (2010); Barbagallo *et al.* (2009). The goal of this chapter is to explore an alternative way of performing such ROMs in the context of transitional separated flows.

The Proper Orthogonal Decomposition (POD) method has been used extensively for reduced-order modeling of fluid mechanics problems. It was proposed by Lumley (1970) as an unbiased technique for identifying the most energetic patterns (or structures) in a flow. In particular, POD has been shown to yield a valuable analysis tool to discriminate coherent structures in turbulent flows, see Sirovich (1987); Berkooz *et al.* (1993). Thus, POD modes stand for natural candidates to design ROMs. This may be achieved by projecting the original equations (either linear or not) onto the POD modes. Such a reduced modeling of the Navier-Stokes equations was first performed by Aubry *et al.* (1988). Noticeably, a time-space deterministic version of the POD has been introduced by Aubry *et al.* (1991) as a systematic tool for complex system analysis. This latter consists of a bi-orthogonal decomposition into spatial orthogonal modes (similar to the POD modes) and temporal orthogonal modes (standing for the deterministic time coefficients of the modes). Since then, POD model reduction has spawned a substantial body of literature on low-dimensional models for flows, see Noack *et al.* (2003); Buffoni *et al.* (2006); Galletti *et al.* (2007). Recently, POD based ROMs have also been incorporated into flow control strategies for the flow around a cylinder (Delville *et al.* (1998); Tadmor *et al.* (2004); Bergmann *et al.* (2005)) or the flow over an open cavity, see Samimy *et al.* (2007). Although this model reduction technique is tractable for very large data sets and applicable to complex flows, POD modes may not be the best structures for describing the dynamics, even in a linearized case. Indeed, the truncated low-energy features of a flow may contribute to a significant part of the global dynamics, as in the typical example of acoustic modes in cavity oscillations, see Rowley *et al.* (2004).

Within the context of model reduction, both the controllability, i.e., the ability of the applied forcing to reach flow states, and observability, i.e., the ability of flow states to register at the sensor locations, are equally important. An expansion basis that balances these two concepts was introduced more than two decades ago for stable linear input-output systems by Moore (1981). It is shown that the balanced basis can be computed as the eigenvectors of the product of the so-called controllability and observability Gramians (which measure controllability and observability respectively). Model reduction is then performed by considering a basis where the states that respond most strongly to inputs (most controllable states) are also states that have the most influence on the outputs (most observable states). This powerful technique, commonly applied in control theory, is known as balanced truncation and constitutes a quasi-optimal basis in terms of modeling the input-output dynamics. Further development of the method extended its range to nonlinear control problems, see Scherpen (1993) and Lall *et al.* (2002). An optimal and accurate algorithm for the calculation of this basis has been found by Laub *et al.* (1987) and has been used on some fluid problems of small size (Cortelezzi & Speyer (1998)) but its associated computational effort is rather high for systems of moderate size and quickly becomes unaffordable for systems of large size and realistic complexity (about 5 000 states or more).

However, recent developments by Willcox & Peraire (2002) and Rowley (2005) combining computational methodology from POD modes with a balancing procedure has overcome this difficulty and has brought the model reduction of large-scale control problems within reach of current computational technology. It was shown that the Gramians can be approximated using two series of snapshots resulting from two different numerical simulations and that the algorithm introduced by Laub *et al.* (1987) can be generalized to take into account these approximate Gramians. This new method is referred as Balanced Proper Orthogonal Decomposition (BPOD) due to the use of flow snapshots and the connection to POD established by Rowley (2005). This new technique has been applied to several linearized stable flows: the case of a channel flow (Ilak & Rowley (2006, 2008)), a one-dimensional model equation mimicking an open flow (Bagheri *et al.* (2009*c*)) and a boundary-layer flow (Bagheri *et al.* (2009*b*)).

Regarding the reduction of linearized unstable systems, the original method by Moore (1981) is no longer applicable. To overcome this limitation, an extension was proposed by Zhou *et al.* (1999) by introducing frequency-domain definitions of controllability and observability Gramians. In contrast to Gramians defined in the literature for stable systems, these Gramians can always be computed for systems without

marginal modes and they reduce to the standard controllability and observability Gramians when the systems are stable, see Kim (1998); Antoulas (2005). The model reduction procedure by Zhou *et al.* (1999) essentially decouples the dynamics of the flow on the stable and unstable subspaces, and then truncates the relatively uncontrollable and unobservable modes of each of the two subspaces. Following this idea, the model reduction of the unstable linearized flow over an inclined flat plate was performed by Ahuja & Rowley (2008, 2010) by partitioning the system behavior into stable and unstable dynamics. In this case, the stable subspace was modeled by the standard BPOD algorithm by Rowley (2005) while the unstable dynamics were treated similarly to the work by Åkervik *et al.* (2007) by a projection onto global eigenmodes. The same partitioning has also been carried out by Barbagallo *et al.* (2009) in the case of the unstable dynamics over a square cavity flow. In both studies, the unstable global modes are first computed via a shift-invert Arnoldi technique, and then used to project the series of snapshots required by the BPOD process onto the stable subspace.

In this chapter, we are interested in another method for the balancing and reduction of possibly unstable systems. It consists of using the frequency-domain definitions of the controllability and observability Gramians proposed by Zhou *et al.* (1999) and an approximation of these Gramians with frequential snapshots. This way, the Gramians are no longer approximated by using two series of snapshots arising from time-stepping simulations, but rather by using flow state responses to harmonic forcings. These flow states involved in the process are called frequential snapshots throughout this work due to their natural link with the current time-based snapshot method. The idea to use frequential expressions of the Gramians and to compute frequential snapshots to perform reduced-order models has already been introduced by Willcox & Peraire (2002), although it has never been carried out. The goal of this chapter is to show that the use of these frequential snapshots are able to build efficient reduced-order models for linearized and possibly unstable fluid systems in a global framework. This issue is studied through two examples: a globally stable but convectively unstable flow over a rounded backward-facing step and a globally unstable flow over a square cavity. In both cases, a single actuator (input) and a single sensor (output) are placed near separation and reattachment respectively. The approximate balanced truncation procedure is then derived considering this input and this output using the snapshot method in the frequential framework. At the same time, the POD modes constructed from an impulse released from the input, i.e. the most controllable modes, are also considered and computed through the same frequential snapshots. Both BPOD and POD reduced-order models are eventually evaluated by comparing their impulse and frequency responses to that of their associated full system. In the case of the unstable flow over the square cavity, particular care is given to the reduction performance of the stable and unstable subspaces. Additionally, our results are compared to a previous study by Barbagallo *et al.* (2009) on the same flow configuration where partitioning of the two subspaces and separate reduction was performed. The main contributions of this work comprise (i) an illustration of the ability of harmonic flow states responses (frequential snapshots) to build efficient BPOD/POD reduced-order models, and (ii) a new algorithm to perform model reduction for unstable linear systems without partitioning the stable and unstable subspace, and thus, without computing any global eigenmodes.

The chapter proceeds along the following outline: in section 2.2, we first briefly describe the BPOD and POD model reduction procedure using frequential snapshots within the framework of the linearized Navier-Stokes equations with actuation and sensing. In particular, a comparison of the computational costs of the temporal and frequential domain based snapshot method is assessed. In section 2.3, we present numerical results of ROMs using the example of a globally stable flow over a backward-facing step. In section 2.4, we investigate the case of a globally unstable flow over a square cavity and compare the performance of our ROMs with the previous work by Barbagallo *et al.* (2009). Finally, we conclude with a brief discussion in section 2.5.

2.2 Model reduction methodology

2.2.1 Problem formulation

2.2.1.1 Governing equations

We consider the incompressible Navier-Stokes equations, with a small actuation and a sensing, governing the dynamics of the velocity $\boldsymbol{u}$ and pressure p fields:

$$\begin{cases} \partial_t \boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla p + Re^{-1}\nabla^2\boldsymbol{u} + \varepsilon\mathcal{B}c(t) \\ \nabla\cdot\boldsymbol{u} = 0 \\ m(t) = \mathcal{C}\boldsymbol{u} \end{cases} \tag{2.1}$$

where the only parameter reduces to the Reynolds number Re. The actuation is denoted by the term $\varepsilon\mathcal{B}c(t)$ corresponding to a momentum forcing. The parameter ε indicates that the forcing is small compared to the other terms of the equation and $c(t)$ is the temporal law of the actuation also referred to as the input of the problem. The quantity $m(t)$ measured by the sensor stands for the output of the system, it is expressed as the result of a measure operator $\mathcal{C}$ applied on the velocity field $\boldsymbol{u}$. Note that we have assumed an actuation and sensing based on the velocity components only for simplicity. Considering this problem, we which to design a reduced order model able to capture the linear input-output dynamics of this system.

Before proceeding with model reduction, we first express the governing equations in a linear state-space form. To this end, a base flow $(\boldsymbol{U}, P)$ to linearize about has to be determined. This is accomplished by setting the unsteady terms of equations (2.1) to zero and solving the resulting nonlinear equations:

$$\begin{cases} (\boldsymbol{U}\cdot\nabla)\boldsymbol{U} = -\nabla P + Re^{-1}\nabla^2\boldsymbol{U} \\ \nabla\cdot\boldsymbol{U} = 0 \\ m_s = \mathcal{C}\boldsymbol{U} \end{cases} \tag{2.2}$$

where m_s stands for the measure of the base flow. The flow state can then be decomposed as the sum of the base flow and a small perturbation by $(\boldsymbol{u}, p) = (\boldsymbol{U}, P) + \varepsilon(\boldsymbol{u}', p')$. Similarly, the measure $m(t)$ can be decomposed as the sum of the steady part coming from the base flow and the small contribution from the perturbation as $m(t) = m_s + \varepsilon m'(t)$ where $m'(t) = \mathcal{C}\boldsymbol{u}'$. Substitution of these decompositions into equations (2.1), neglecting the ε^2 term and omitting primes yields the linearized Navier-Stokes equations:

$$\begin{cases} \partial_t \boldsymbol{u} + \boldsymbol{U}\cdot\nabla\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{U} = -\nabla p + Re^{-1}\nabla^2\boldsymbol{u} + \mathcal{B}c(t) \\ \nabla\cdot\boldsymbol{u} = 0 \\ m(t) = \mathcal{C}\boldsymbol{u} \end{cases} \tag{2.3}$$

which govern the linear dynamics about the considered base flow. It is furthermore assumed that the amplitude of the perturbation remains weak compared to the base flow so that the validity of equations (2.3) is guaranteed.

Using a numerical approach, one then proceeds by discretizing the problem on a mesh (ours is based on a finite element method described in the end of this section). If we write the discretized equations (2.3) in matrix form, we obtain

$$\begin{pmatrix} \mathsf{Q}_1 & 0 \\ 0 & 0 \end{pmatrix} \frac{d}{dt}\begin{pmatrix} \mathsf{X}_1 \\ \mathsf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathsf{A}_1 & \mathsf{A}_2^* \\ \mathsf{A}_2 & 0 \end{pmatrix}\begin{pmatrix} \mathsf{X}_1 \\ \mathsf{X}_2 \end{pmatrix} + \begin{pmatrix} \mathsf{Q}_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \mathsf{B}_1 \\ 0 \end{pmatrix} c(t) \tag{2.4a}$$

$$m(t) = \begin{pmatrix} \mathsf{C}_1 & 0 \end{pmatrix}\begin{pmatrix} \mathsf{X}_1 \\ \mathsf{X}_2 \end{pmatrix} \tag{2.4b}$$

where X_1 denotes the velocity fields and X_2 stands for the corresponding pressure field. $(\mathsf{B}_1, 0)$ and $(\mathsf{C}_1, 0)$ denote respectively the vectors of the discretized actuation $\mathcal{B}$ and sensing $\mathcal{C}$ operators. The linearized

Navier-Stokes operator has been decomposed so that A_1 stands for convection and diffusion while A_2 and A_2^* are the parts relative to incompressibility and pressure effect respectively, the superscript * denoting the transconjugate. For the design of reduced-order models we need to reformulate the above equations into a standard state-space form. This is achieved by restricting the system state to its divergence-free velocity field. A few calculations exposed in appendix 2.A demonstrate that equations (2.4) can be formulated as the following standard linear input-output system

$$\frac{d\mathsf{X}_1}{dt} = \mathsf{A}\mathsf{X}_1 + \mathsf{B}c(t) \tag{2.5a}$$
$$m(t) = \mathsf{C}\mathsf{X}_1 \tag{2.5b}$$

where we have introduced $\mathsf{A} = \mathsf{P}_1\mathsf{A}_1$, $\mathsf{B} = \mathsf{P}_1\mathsf{Q}_1\mathsf{B}_1$ and $\mathsf{C} = \mathsf{C}_1$ so that the projection matrix onto the divergence-free space reduces to $\mathsf{P}_1\mathsf{Q}_1$. The dimension of the full system is called hereafter n_1 and denotes the number of degrees of freedom in equation (2.5) or alternatively the size of the discretized state velocity vector X_1. Once defined, our linear input-output state space system has to be associated with scalar products for the input, the output and the state space. Since the input and output are scalars, their associated inner product are naturally chosen as the standard hermitian product so that their associated energy are respectively $|c|^2$ and $|m|^2$. As mentioned in the work by Ilak & Rowley (2008), the choice of inner product on the state space does not intervene in balanced truncation, although it does for POD. We choose the inner product Q_1 that takes into account the numerical discretization by adding a weight matrix to the standard inner product. Its associated energy is $\mathsf{X}_1^*\mathsf{Q}_1\mathsf{X}_1$, i.e. the kinetic energy of the perturbation. This choice is thus intuitively appealing, since POD will capture the true energy of the perturbations.
As recognized in existing literature (Zhou *et al.* (2002); Antoulas (2005); Bagheri *et al.* (2009*c*)), the input-output behavior is the critical quantity that has to be carefully taken into consideration within the framework of linear input-output system dynamics and more particularly in control theory. Mathematically, the relevant quantities to assess this performance are the impulse response and/or the transfer function. As a result, these are the criterions considered in the following sections to quantify the performance of the ROMs.

2.2.1.2 Model reduction phenomenology

Before proceeding with a precise and more technical description of the model reduction procedure, we briefly introduce its concept here. For an easier physical interpretation of the procedure, we first assume that the fluid system is linearly stable; the case of unstable systems is discussed afterwards.

The actuator (input) excites the flow dynamics while the sensor yields a measure corresponding to the output. The present model reduction procedure consists in projecting the original equations (2.5) onto a low number of flow structures. If we denote the projection basis by the matrix T_1 and its bi-orthogonal set by S_1, then the Petrov-Galerkin projection of the original system provided by equations (2.5) leads to the following reduced system

$$\frac{d\mathsf{X}_{1r}}{dt} = \mathsf{A}_r\mathsf{X}_{1r} + \mathsf{B}_r c(t) \tag{2.6a}$$
$$m_r(t) = \mathsf{C}_r\mathsf{X}_{1r} \tag{2.6b}$$

which governs the dynamics of the reduced output signal $m_r(t)$ and the reduced state X_{1r} is such that $\mathsf{X}_1 = \mathsf{T}_1\mathsf{X}_{1r}$ and $\mathsf{X}_{1r} = \mathsf{S}_1^*\mathsf{Q}_1\mathsf{X}_1$. We have also introduced in the above equations the reduced dynamical operator, actuation and sensing denoted by A_r, B_r and C_r respectively. They are given by

$$\mathsf{A}_r = \mathsf{S}_1^*\mathsf{Q}_1\mathsf{A}\mathsf{T}_1 \qquad \mathsf{B}_r = \mathsf{S}_1^*\mathsf{Q}_1\mathsf{B} \qquad \mathsf{C}_r = \mathsf{C}\mathsf{T}_1 \tag{2.7}$$

The performance of such ROMs is then assessed by comparing their input-output behavior to the original one. The choice of the projection basis is obviously of primary importance if one plans to design efficient ROMs. For stable linear systems, it was shown by Moore (1981) that their controllability and observability stand for critical quantities. By definition, a flow structure is said controllable if it may be forced by a small amount of energy from the input. Additionally, a flow structure is said observable if its dynamics yield a high energy contribution to the output. In this context, the two candidates examined here to project the dynamics are: (i) the most controllable modes, denoted as POD modes and (ii) the set of equally (and most) controllable and observable modes, known as balanced modes.

Due to the linear nature of the equations, any harmonic excitation will lead to a harmonic flow response and, consequently, to a harmonic measurement. It is thus natural to resort to a frequency framework when dealing with such systems. Incidentally, the particularity of our work is the introduction of harmonic flow states meant to design efficient ROMs. Let us first describe the procedure to design the POD modes. We consider harmonic excitations from the input so as to get harmonic flow responses. By covering a sufficiently wide frequency range, we expect to capture the most energetic responses from all possible harmonic forcings. Next, based on these harmonic responses, we compute the set of most controllable modes on that frequency range. The resulting so-called POD modes are thus conceptually designed to optimally capture the energy triggered by the input on the particular frequency interval of interest. Now, regarding the balanced modes, we do not only look for structures which are controllable (by the actuator) but also observable by the sensor. Consequently, the balanced modes not only include information from the input but also from the output. As we will see in the rest of this section, it is possible to identify the flow structures to which the sensor is most sensitive through the adjoint Navier-Stokes equations. This is performed analogously in the frequency domain by computing the harmonic flow states which yield the maximum energy contribution to the sensor energy. Such harmonic flow structures, once computed on a frequency interval are included in the balanced truncation procedure in order to find equally controllable and observable modes on the considered frequency interval.
In the end, the POD modes only include information from the actuator and the flow dynamics; they depend on the operators A and B. On the other hand, the balanced modes also rely on the sensor and depend on all three operators: A, B and C. By construction, the balanced modes are expected to be superior to the POD modes in capturing the input-output behavior since they include additional information from the original input-output system. Concerning linear unstable systems, the concepts of controllability and observability are no longer defined, neither are the harmonic responses of the system to harmonic forcings. However, the overall procedure introduced above will be shown to remain tractable, as illustrated later in section 2.4.

How to obtain the POD and balanced modes stands for our next concern. In the following the balanced modes, which rely on both the concepts of controllability and observability, are first introduced. The presentation of the POD modes (most controllable modes) is then given in a second step.

2.2.2 Balanced model reduction

2.2.2.1 Controllability and observability Gramians

Balanced truncation originally emerged from the control theory literature (Moore (1981)) as a way to reduce linear input-output systems such as (2.5) while quasi-optimally preserving its input-output behavior. Indeed, a useful property of balanced truncation is that it yields a priori error bounds on the transfer function that are close to the lower bound achievable by any reduced-order model. It is conceptually expressed in terms of controllability an observability of the modes used in the model reduction process. A specific state is deemed controllable if there exists a control law $c(t)$ which is able to modify the flow from any state toward this specific state. For controllable states, the notion of controllability then quantifies how easy (or difficult) the state can be reached from any state. Similarly,

observability measures how easy (or difficult) a given flow state can be detected by the sensors. At the sensor location almost unobservable flow states leave hardly any footprint behind and are thus nearly invisible to the measurement efforts. The key idea of balanced truncation is to compute, rank and select modes that are equally observable and controllable in order to project the full original system onto them. For a given flow state, the mathematical quantities that enable measuring these two properties are the controllability and observability Gramians G_c and G_o. Considering our input-output system (2.5) with a stable dynamical operator A, these Gramians are defined by:

$$\left\{ \begin{array}{l} \mathsf{G}_c = \int_0^{+\infty} e^{\mathsf{A}t} \mathsf{B}\mathsf{B}^* e^{\mathsf{A}^* t} dt \\ \mathsf{G}_o = \int_0^{+\infty} e^{\mathsf{A}^* t} \mathsf{C}^* \mathsf{C} e^{\mathsf{A}t} dt \end{array} \right. \tag{2.8}$$

For unstable systems, this definition does not hold since the integrals are not convergent anymore as $t \to \infty$. Yet, one may use a frequency domain definition of the Gramians, as proposed by Zhou *et al.* (1999), in order to avoid this problem:

$$\left\{ \begin{array}{l} \mathsf{G}_c = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (j\omega \mathsf{I} - \mathsf{A})^{-1} \mathsf{B}\mathsf{B}^* (-j\omega \mathsf{I} - \mathsf{A}^*)^{-1} d\omega \\ \mathsf{G}_o = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (-j\omega \mathsf{I} - \mathsf{A}^*)^{-1} \mathsf{C}^* \mathsf{C} (j\omega \mathsf{I} - \mathsf{A})^{-1} d\omega \end{array} \right. \tag{2.9}$$

It can be shown by using Parseval's theorem that the two definitions are equivalent for stable systems, see Zhou *et al.* (1999); Antoulas (2005). Furthermore, it has been demonstrated by Zhou *et al.* (1999) that the frequency domain definition (2.9) still works for unstable systems as far as there are no marginal modes and that balanced truncation is then equivalent to separating the stable and unstable parts of the transfer function and doing the balanced realization for both parts separately.

A technique referred to as balancing, consists in finding flow fields with equal emphasis on either controllability and observability property. This latter is mathematically equivalent to finding a transformation basis in which the Gramians G_c and G_o appear diagonal and equal, see Zhou *et al.* (2002). Otherwise, these balanced modes can also be directly computed as the eigenvectors of the product of the two previously defined Gramians, see Moore (1981). However, this is not computationally tractable for the complex flows considered here where the number of degrees of freedom is of order $O(10^{5-6})$. For systems of large dimension such as those encountered here, the Gramians are huge matrices which cannot be easily computed or stored. Instead, the algorithm introduced by Rowley (2005) and referred to as BPOD approximates balanced truncation while remaining tractable even for very large systems and relies on the use of flow state snapshots.

2.2.2.2 Introduction of the frequential snapshots

To compute the balanced modes by a low-cost algorithm, a technique introduced by Rowley (2005) consists of factoring the controllability and observability Gramians using flow state snapshots. Until now, the snapshot technique was achieved by considering the temporal expressions (2.8) of the Gramians, as in the works by Rowley (2005); Ilak & Rowley (2008); Bagheri *et al.* (2009*b*); Ahuja & Rowley (2010); Barbagallo *et al.* (2009). This approach is not adopted in our work. In this chapter, the frequential expressions of the Gramians (2.9) are considered instead. Thus, introducing the flow states $\hat{\mathsf{X}}_1(\omega)$ and $\hat{\mathsf{Y}}_1(\omega)$ defined by:

$$\hat{\mathsf{X}}_1(\omega) = (j\omega \mathsf{I} - \mathsf{A})^{-1} \mathsf{B} \tag{2.10a}$$

$$\mathsf{Q}_1 \hat{\mathsf{Y}}_1(\omega) = (-j\omega \mathsf{I} - \mathsf{A}^*)^{-1} \mathsf{C}^* \tag{2.10b}$$

the frequential expressions of the Gramians (2.9) reduce to:

$$\mathsf{G}_c = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mathsf{X}}_1\hat{\mathsf{X}}_1^*(\omega)d\omega \approx \frac{1}{2\pi}\sum_{i\in\mathbb{Z}} \hat{\mathsf{X}}_1\hat{\mathsf{X}}_1^*(\omega_i)\delta_i \tag{2.11a}$$

$$\mathsf{G}_o = \mathsf{Q}_1\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mathsf{Y}}_1\hat{\mathsf{Y}}_1^*(\omega)d\omega\right)\mathsf{Q}_1 \approx \mathsf{Q}_1\left(\frac{1}{2\pi}\sum_{i\in\mathbb{Z}} \hat{\mathsf{Y}}_1\hat{\mathsf{Y}}_1^*(\omega_i)\delta_i\right)\mathsf{Q}_1 \tag{2.11b}$$

where $\{\omega_i, i \in \mathbb{Z}\}$ is a given set of discrete pulsations and δ_i denotes appropriate quadrature coefficient. Note that $\hat{\mathsf{X}}_1$ and $\hat{\mathsf{Y}}_1$ are complex vector fields such that

$$\overline{\hat{\mathsf{X}}}_1(\omega_i) = \hat{\mathsf{X}}_1(-\omega_i) \tag{2.12a}$$

$$\overline{\hat{\mathsf{Y}}}_1(\omega_i) = \hat{\mathsf{Y}}_1(-\omega_i) \tag{2.12b}$$

where $\bar{}$ denotes the complex conjugate so that we can further expand equations (2.11) to

$$\mathsf{G}_c \approx \frac{1}{\pi}\sum_{i\in\mathbb{N}}[\hat{\mathsf{X}}_{1r}\hat{\mathsf{X}}_{1r}^*(\omega_i) + \hat{\mathsf{X}}_{1i}\hat{\mathsf{X}}_{1i}^*(\omega_i)]\delta_i \tag{2.13a}$$

$$\mathsf{G}_o \approx \mathsf{Q}_1\left(\frac{1}{\pi}\sum_{i\in\mathbb{N}}[\hat{\mathsf{Y}}_{1r}\hat{\mathsf{Y}}_{1r}^*(\omega_i) + \hat{\mathsf{Y}}_{1i}\hat{\mathsf{Y}}_{1i}^*(\omega_i)]\delta_i\right)\mathsf{Q}_1 \tag{2.13b}$$

where we have introduced their real parts $\hat{\mathsf{X}}_{1r}$, $\hat{\mathsf{Y}}_{1r}$ and their imaginary parts $\hat{\mathsf{X}}_{1i}$, $\hat{\mathsf{Y}}_{1i}$. In practice, the set of frequencies $\{\omega_i\}$ is finite so that only a finite interval of frequencies is considered. The numbers of real direct $(\hat{\mathsf{X}}_{1r}, \hat{\mathsf{X}}_{1i})$ and adjoint $(\hat{\mathsf{Y}}_{1r}, \hat{\mathsf{Y}}_{1i})$ flow states used in the process are denoted respectively by n_d and n_a (both are even). As a result, the controllability and observability Gramians may be factored as

$$\mathsf{G}_c \approx \mathbf{X}_1\mathbf{X}_1^* \tag{2.14a}$$

$$\mathsf{G}_o \approx \mathsf{Q}_1\mathbf{Y}_1\mathbf{Y}_1^*\mathsf{Q}_1 \tag{2.14b}$$

by stacking these states as columns of the matrices $\mathbf{X}_1$ and $\mathbf{Y}_1$ as follows:

$$\mathbf{X}_1 = \frac{1}{\sqrt{\pi}}\left[\begin{array}{cccc} \hat{\mathsf{X}}_{1r}(\omega_0)\sqrt{\delta_0} & \hat{\mathsf{X}}_{1i}(\omega_0)\sqrt{\delta_0} & \hat{\mathsf{X}}_{1r}(\omega_1)\sqrt{\delta_1} & \hat{\mathsf{X}}_{1i}(\omega_1)\sqrt{\delta_1} \quad \dots \end{array}\right] \tag{2.15a}$$

$$\mathbf{Y}_1 = \frac{1}{\sqrt{\pi}}\left[\begin{array}{cccc} \hat{\mathsf{Y}}_{1r}(\omega_0)\sqrt{\delta_0} & \hat{\mathsf{Y}}_{1i}(\omega_0)\sqrt{\delta_0} & \hat{\mathsf{Y}}_{1r}(\omega_1)\sqrt{\delta_1} & \hat{\mathsf{Y}}_{1i}(\omega_1)\sqrt{\delta_1} \quad \dots \end{array}\right] \tag{2.15b}$$

where the dimension of $\mathbf{X}_1$ is $n_1 \times n_d$ and that of $\mathbf{Y}_1$ is $n_1 \times n_a$. Interestingly, the flow state $\hat{\mathsf{X}}_1(\omega)$ is solution (see appendix 2.B.1) of the following direct problem:

$$\left[j\omega\begin{pmatrix} \mathsf{Q}_1 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} \mathsf{A}_1 & \mathsf{A}_2^* \\ \mathsf{A}_2 & 0 \end{pmatrix}\right]\begin{pmatrix} \hat{\mathsf{X}}_1 \\ \hat{\mathsf{X}}_2 \end{pmatrix} = \begin{pmatrix} \mathsf{Q}_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \mathsf{B}_1 \\ 0 \end{pmatrix} \tag{2.16a}$$

emphasizing that this state reduces to the harmonic flow response resulting from the forcing of the linear Navier-Stokes equations with an harmonic momentum actuation $(\mathsf{B}_1, 0)$ of pulsation ω. Analogously, it is shown in appendix 2.B.2 that $\hat{\mathsf{Y}}_1(\omega)$ is solution of the following adjoint problem:

$$\left[-j\omega\begin{pmatrix}\mathsf{Q}_1 & 0\\ 0 & 0\end{pmatrix}-\begin{pmatrix}\mathsf{A}_1^* & \mathsf{A}_2^*\\ \mathsf{A}_2 & 0\end{pmatrix}\right]\begin{pmatrix}\hat{\mathsf{Y}}_1\\ \hat{\mathsf{Y}}_2\end{pmatrix} = \begin{pmatrix}\mathsf{Q}_1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\mathsf{Q}_1^{-1}\mathsf{C}_1^*\\ 0\end{pmatrix} \tag{2.17a}$$

pointing out that this adjoint state stands for the harmonic response flow state resulting from the forcing of the adjoint linear Navier-Stokes equations with a harmonic momentum actuation $(\mathsf{Q}_1^{-1}\mathsf{C}_1^*, 0)$ of pulsation ω. Hereafter, these direct and adjoint flow states are referred to as frequential snapshots in order to draw a parallel with existing time-domain definitions of the Gramians and associated temporal snapshots.

2.2.2.3 Computation of the balanced basis

The approximate Gramians (2.14) are not actually computed due to the large storage cost, but the leading modes of the transformation that balances these Gramians are computed using a cost-efficient algorithm introduced by Laub *et al.* (1987) and detailed below. It involves computing the singular value decomposition of the direct $\mathbf{X}_1$ and adjoint $\mathbf{Y}_1$ snapshots cross product $\mathbf{Y}_1^*\mathsf{Q}_1\mathbf{X}_1$, which is of size $n_a \times n_d$:

$$\mathbf{Y}_1^*\mathsf{Q}_1\mathbf{X}_1 = \mathsf{M}\Sigma\mathsf{N}^* \tag{2.18}$$

where M and N are orthogonal matrices ($\mathsf{M}^*\mathsf{M} = \mathsf{I}$, $\mathsf{N}^*\mathsf{N} = \mathsf{I}$) of dimension $n_a \times n_a$ and $n_d \times n_d$ while Σ is diagonal and of size $n_a \times n_d$. For the fluid systems we are interested in, the typical number of snapshots is of order $O(10^{2-4})$, thus resulting in a reasonable computational cost. In a final step, denoting the balanced basis by the matrix T_1 and its bi-orthogonal set by S_1, we have

$$\mathsf{T}_1 = \mathbf{X}_1\mathsf{N}\Sigma^{-1/2} \tag{2.19a}$$
$$\mathsf{S}_1 = \mathbf{Y}_1\mathsf{M}\Sigma^{-1/2} \tag{2.19b}$$

It is easily confirmed that the bi-orthogonality condition $\mathsf{S}_1^*\mathsf{Q}_1\mathsf{T}_1 = \mathsf{I}$ is satisfied and that, once transformed into these bases, the Gramians G_c and G_o appear diagonal and equal to Σ:

$$(\mathsf{Q}_1\mathsf{S}_1)^*\mathsf{G}_c(\mathsf{Q}_1\mathsf{S}_1) = \Sigma \tag{2.20a}$$
$$\mathsf{T}_1^*\mathsf{G}_o\mathsf{T}_1 = \Sigma \tag{2.20b}$$

and also that they are the eigenvectors of their product:

$$\mathsf{G}_c\mathsf{G}_o\mathsf{T}_1 = \mathsf{T}_1\Sigma^2 \tag{2.21a}$$
$$\mathsf{G}_o\mathsf{G}_c(\mathsf{Q}_1\mathsf{S}_1) = (\mathsf{Q}_1\mathsf{S}_1)\Sigma^2 \tag{2.21b}$$

The diagonal entries of the transformed Gramians Σ, called Hankel singular values (HSVs), decrease monotonically and are directly related to the controllability and observability of the corresponding states. It can be shown (Rowley (2005)) that the columns of T_1 form the first columns of the balancing transformation and the columns of $\mathsf{Q}_1\mathsf{S}_1$ constitute the first columns of the inverse transformation.

2.2.3 POD model reduction

For the sake of completeness, POD modes are also considered in this study and their ability to capture the input-output dynamics is evaluated and compared to that of BPOD models. For stable systems, we have previously introduced the POD modes as the most controllable modes relative to the actuator. This means that they stand for the structures the most easily triggered by the actuation. The POD

modes, denoted by the matrix R_1, are given by the eigenvectors of the product $\mathsf{G}_c\mathsf{Q}_1$ (Rowley (2005); Barbagallo *et al.* (2009)) that is

$$\mathsf{G}_c\mathsf{Q}_1\mathsf{R}_1 = \mathsf{R}_1\Lambda \tag{2.22}$$

where the diagonal entries of the matrix Λ are called the POD eigenvalues. The POD eigenvalues decrease monotonically and are directly related to the controllability of the corresponding modes. Their computation is performed by using the classical snapshot method introduced by Sirovich (1987). We have previously shown that $\mathsf{G}_c = \mathbf{X}_1\mathbf{X}_1^*$ where the collected dataset $\mathbf{X}_1$ stands for the direct frequential snapshots. We proceed by computing the eigenvalue decomposition of the product $\mathbf{X}_1^*\mathsf{Q}_1\mathbf{X}_1$, which is of size $n_d \times n_d$, by $\mathbf{X}_1^*\mathsf{Q}_1\mathbf{X}_1 = \mathsf{L}\Lambda\mathsf{L}^*$ where L is orthogonal ($\mathsf{L}^*\mathsf{L} = \mathsf{I}$). The POD modes can then be computed by

$$\mathsf{R}_1 = \mathbf{X}_1\mathsf{L}\Lambda^{-1/2} \tag{2.23}$$

Note that, as expected, the POD modes are orthogonal with respect to the kinetic energy inner product, that is $\mathsf{R}_1^*\mathsf{Q}_1\mathsf{R}_1 = \mathsf{I}$. It should be emphasized at this stage that this definition of the POD modes is not general since the considered dataset $\mathbf{X}_1$ is directly linked to the input $(\mathsf{B}_1, 0)$ location. Reduced-order models are obtained by a Galerkin projection of the initial full system onto the modes with the highest POD eigenvalues. The procedure is similar to that described in section 2.2.1.2 where we substitute T_1 and S_1 by R_1 in equation (2.7).

2.2.4 Discussion

2.2.4.1 Practical considerations

Frequency information may be the most accessible quantity from many mechanical systems. In fluid mechanics, a physical interpretation of the flow behavior often relies on the frequency decomposition of its response to different forcings. For instance, some flows may behave as "oscillators" which impose their own frequency on the intrinsic dynamics, insensitively to external noise. On the other hand, some flows known as "noise-amplifiers" selectively amplify upstream noise, which often leads to a broad-band low-frequency spectrum of the downstream flow response. Due to this natural way to describe unsteady flows, frequency-based methods play a central role in the design of low-order models. Let us mention some examples such as the Dynamic Mode Decomposition (DMD), see Rowley *et al.* (2009); Schmid (2010), which decomposes dynamical modes in the frequency domain or also the partitioning of modes into low, dominant and high frequency in view of designing robust ROMs, see Aubry *et al.* (1988); Noack *et al.* (2003, 2005).

Concerning our linear state space formulation, the excitation by the actuator over a frequency interval is a way of extracting relevant information about the flow dynamics. Similarly, the adjoint frequency snapshots are introduced as a way to extract sensitivity information of the sensor over the same range of frequency. From an experimental point of view, the direct snapshots may be naturally obtained from harmonic excitations from the actuator. Suppose that the actuation law is given by $\mathsf{B}\cos(\omega t)$, then the associated harmonic response of the flow will be given by $\hat{\mathsf{X}}_{1r}(\omega)\cos(\omega t) - \hat{\mathsf{X}}_{1i}(\omega)\sin(\omega t)$. Accordingly, the direct snapshots $\hat{\mathsf{X}}_{1r}$ and $\hat{\mathsf{X}}_{1i}$ may be simply recovered by extracting the flow states at successive times t and $t + T/4$, where $T = 2\pi/\omega$ is the period of the flow. The procedure to approximate the balanced modes requires the knowledge of adjoint states which cannot be directly extracted from experiments. However, we refer interested readers to the recent attempts to determine balanced ROMs without having to resort to adjoint information reported by Or & Speyer (2008) and Ma *et al.* (2010).

An important point should also be noticed at this stage: sections 2.3 and 2.4 include comparisons between POD and BPOD models in terms of modeling the input-output dynamics. As already mentioned earlier,

the balanced modes are superior to the POD modes to reach this goal since they are conceptually designed to do so. However, even if BPOD models may yield the correct input-output dynamics, there is no guarantee that the original flow state may be captured. On the contrary, POD modes are meant to gradually capture the energy of the flow response in order to accurately recover its associated flow field, or at least its predominant energetic patterns. In fact, the superiority of the BPOD models comes from their potential to represent non-physical flow states which yield the correct input-output signal.

2.2.4.2 Computation of the snapshots

As previously mentioned, the snapshot method was achieved by considering the temporal expressions of the Gramians. Technically, the impulse response required to compute these temporal snapshots are computed by a direct and adjoint direct numerical simulation of the linearized Navier-Stokes equations. This time-stepping approach has been used so far in previous works on BPOD model reduction to yield the terms $e^{\mathsf{A}t}\mathsf{B}$ and $e^{\mathsf{A}^*t}\mathsf{C}^*$. On the contrary, in a frequential framework, the snapshots correspond to harmonic responses to harmonic forcings. In this case, one has to invert a direct and adjoint system to get the terms $(j\omega\mathsf{I}-\mathsf{A})^{-1}\mathsf{B}$ and $(-j\omega\mathsf{I}-\mathsf{A}^*)^{-1}\mathsf{C}^*$ for each selected frequency ω. The large-scale matrices associated with these systems, which are of size $n_1 \times n_1$ here, are often sparse (depending on the numerical method used). The availability of efficient methods to inverse these huge systems is thus of utmost importance if one plans to adopt this strategy. Practically, the numerical methods to achieve these inversions can be either direct (Marquet & Sipp (2009); Amestoy *et al.* (2001)) or iterative (Monokrousos *et al.* (2010)). A valuable asset for proceeding in the frequential framework is the possibility to compute separately each snapshot, contrary to the time-stepping approach where the previous snapshots are required to compute the next one. This latter observation makes the frequential approach intrinsically fitted to parallel computation of the snapshots.

2.2.4.3 Fall-off of the snapshots norm

The snapshot method yields an approximation of the temporal 2.8 and frequential 2.9 expressions of the Gramians by discretizing and truncating the unbounded integrals to a maximum time and a maximum frequency respectively. It is thus of primary interest to know how quickly does the norm of the snapshots cross-product within these integrals decrease at long times and high frequencies. It can easily be shown that the temporal snapshots norm decreases exponentially as $O(e^{-at})$ for long time where a is the growth rate of the least damped mode, whereas the frequential snapshots have an algebraic fall-off norm as $O(1/\omega^n)$ for high pulsation ($n = 2$ in the case of the Navier-Stokes equations due to viscous terms). This point leads to the first conclusion that, for any stable system ($a < 0$), the temporal snapshots norm fall-off in time always become faster, for sufficiently long time, than the frequential snapshots norm fall-off in frequency. In spite of this observation, it is not possible to say in general if one approach would require less snapshots than the other one. All the same, it should be emphasized that, in a majority of fluid mechanics problems, the frequential responses norm remains centered and peaked at a well specified frequency corresponding physically to the promoting of natural hydrodynamic instabilities. This robust observation can be supported by mentioning the cases of boundary layers, separated flows or wake flows that are subject respectively to the Tollmien-Schlichting, the Kelvin-Helmholtz and the Benard-von Kàrmàn instabilities. In this framework, the frequential approach may be very appreciative for weakly stable systems, i.e. for systems where a is close to 0. In these cases, the impulse response relaxation time is very high so that the temporal approach would require very long time stepping simulations to compute the snapshots while the frequential responses would remain peaked at a given low frequency.

2.2.5 Numerical methods

Our numerical approach is based on a finite element method. All the equations are first rewritten in a variational formulation and then spatially discretized using a mesh composed of triangular elements. In

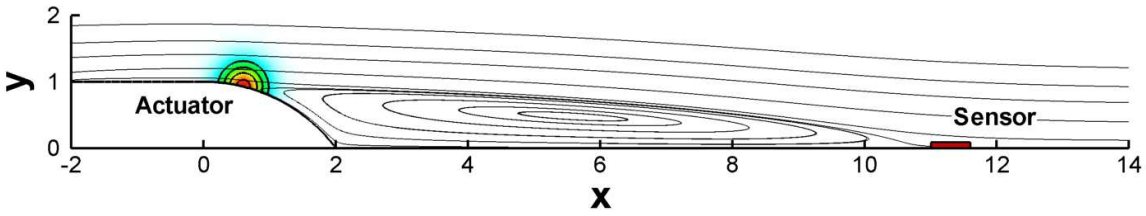


Figure 2.1: Streamlines of the base flow at $Re = 600$. The actuator and sensor locations are also depicted.

particular, the velocity fields are projected onto six-node quadratic triangular elements with quadratic interpolation (P2-elements) whereas the pressure field is discretized using three-node linear triangular elements (P1-elements). All the discrete matrices resulting from the projection of the variational formulations onto the basis of finite elements are sparse, they are built with the FreeFem++ software (http://www.freefem.org).
The frequential snapshots $\hat{\mathsf{X}}_1(\omega_i)$ and $\hat{\mathsf{Y}}_1(\omega_i)$ at discrete pulsations ω_i are obtained by inverting the linear systems (2.16) and (2.17). These snapshots are then recombined to compute the matrices $\mathbf{X}_1$ and $\mathbf{Y}_1$ by using an equidistant spacing $\Delta\omega$ between the snapshots and quadrature coefficients δ_i corresponding to the 4^{th}-order Simpson method. The matrix inversion required to compute these snapshots are handled through a direct multifrontal sparse LU solver (MUMPS), see Amestoy *et al.* (2001). Hence, the cost of this algorithm is approximately given by the cost of the LU decomposition of a large sparse complex matrix since the following successive inverses are cheap.

2.3 Globally stable case: the rounded backward-facing step flow

Here we provide an example of BPOD/POD model reduction using the snapshot method in the frequency domain as described in section 2.2. The case of a globally stable flow is first studied in this section by considering the input-output dynamics over a backward-facing step flow.

2.3.1 Flow configuration

We consider the two-dimensional rounded backward-facing step depicted on figure 2.1. It consists of a circular part designed so that its length is twice its height (this geometry stems from the experimental work by Duriez (2009)). The upstream velocity and the step height are used to make all quantities non-dimensional. The beginning and ending of the step are located at $(x = 0, y = 1)$ and $(x = 2, y = 0)$. The upstream, downstream and upper boundaries are respectively located at $x = -20$, $x = 100$ and $y = 20$. A uniform and unitary velocity field $(u = 1, v = 0)$ is prescribed at the inlet boundary $(x = -20)$ and a laminar boundary layer starts developing on the lower boundary at $(x = -2, y = 1)$. A free-slip condition with zero tangential stress $(\partial_y u = 0, v = 0)$ is prescribed on the boundary $(-20 \leq x \leq -2, y = 1)$. No-slip boundary conditions $(\boldsymbol{u} = 0)$ are imposed on $(-2 \leq x \leq 0, y = 1)$, on the step wall and on the downstream wall $(2 \leq x \leq 100, y = 0)$. Symmetry boundary conditions are used at the upper boundary and a free outflow condition $p\boldsymbol{n} - Re^{-1}(\nabla \boldsymbol{u}) \cdot \boldsymbol{n} = 0$ is used at the outlet ($\boldsymbol{n}$ being the outward normal unitary vector of the boundary).
The resulting base flow is computed by solving a Newton-Raphson method, as in the work by Sipp & Lebedev (2007). We choose a Reynolds number $Re = 600$ where the flow is globally stable to two-dimensional perturbations, its corresponding base flow is represented in figure 2.1. The displacement thickness at $x = 0$ is $\delta^* \approx 0.082$, leading to a Reynolds number based on the displacement thickness of $Re_{\delta^*} \approx 49.2$. This choice of parameters rules out instabilities related to the boundary layer dynamics. The boundary layer separates at $x \approx 0.6$ and reattaches at $x \approx 11$ exhibiting a long shear layer responsible for strong transient growths. The input-output behavior is investigated by introducing one actuator and

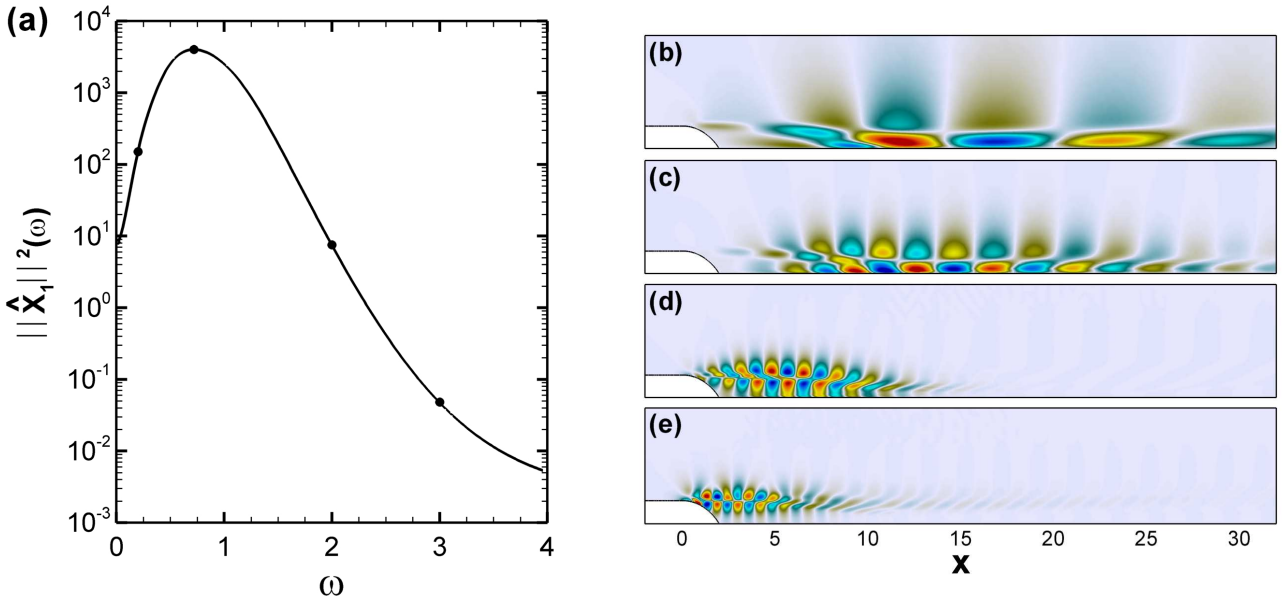


Figure 2.2: The left plot (a) shows the energy of the direct snapshots $\hat{\mathsf{X}}_1$ as a function of ω. Figures (b),(c),(d) and (e) represent the real parts of the longitudinal velocity of the direct snapshots associated with the frequencies $\omega_i = 0.2$, 0.72, 2 and 3 respectively.

one sensor as sketched in figure 2.1. The actuator is located on the step, just before the boundary layer separation, so as to trigger the most efficient response and the sensor is placed downstream, in the vicinity of the reattachment point. Once discretized, the equations governing the dynamics of small perturbations with actuation and sensing are given by the linear input-output system previously introduced in equations (2.5). The actuator B stands for a volume forcing of Gaussian shape on the vertical velocity component centered on the step wall at $x = 0.6$, with a width of 0.6 and a height of 1. As the measured quantity, we consider the wall-normal shear stress evaluated at and integrated over a localized region of the wall (the sensor location) so that $m(t) = \int_{x=11}^{x=11.6} \partial_y u \, dx$, which yields the vector C.
To give an idea, a typical discretization yields $n_1 \approx 360\,000$ degrees of freedom stemming from about 90 000 triangles.

2.3.2 Frequential snapshots

The frequential direct and adjoint snapshots required for the model reduction are computed by solving equations (2.16) and (2.17). We have computed 399 equispaced complex frequential snapshots $\hat{\mathsf{X}}_1$ and $\hat{\mathsf{Y}}_1$ from $\omega = 0$ to $\omega = 4$ (resulting in $n_d = n_a = 798$).
As expressed in section 2.2, the direct snapshots correspond to the long time responses of the system to harmonic forcings of pulsation ω from the input location. We have represented in figure 2.2(a) the evolution of the kinetic energy of the direct snapshots $\|\hat{\mathsf{X}}_1\|^2 = \hat{\mathsf{X}}_1^* \mathsf{Q}_1 \hat{\mathsf{X}}_1$ as a function of ω to highlight the noise amplifier behavior of this flow. The most amplified frequency $\omega \approx 0.72$ is associated with the Kelvin-Helmholtz instability of the separated shear layer. Figures 2.2(b-e) depict the real part of the longitudinal velocity of the direct snapshots $\hat{\mathsf{X}}_1$ for different frequencies. As expected, the corresponding patterns extend downstream from the actuator through the shear layer. The triggered wavepackets shown on these figures exhibit a spatial support that strongly depends on ω. Excitation to higher frequencies leads to smaller flow structures and to a faster diffusion. In particular, the highest frequency responses remain spatially localized in the vicinity of the actuator while the smallest frequency responses are widely extended downstream.
Figure 2.3 is devoted to the same purpose for the adjoint snapshots. In a similar manner, the adjoint

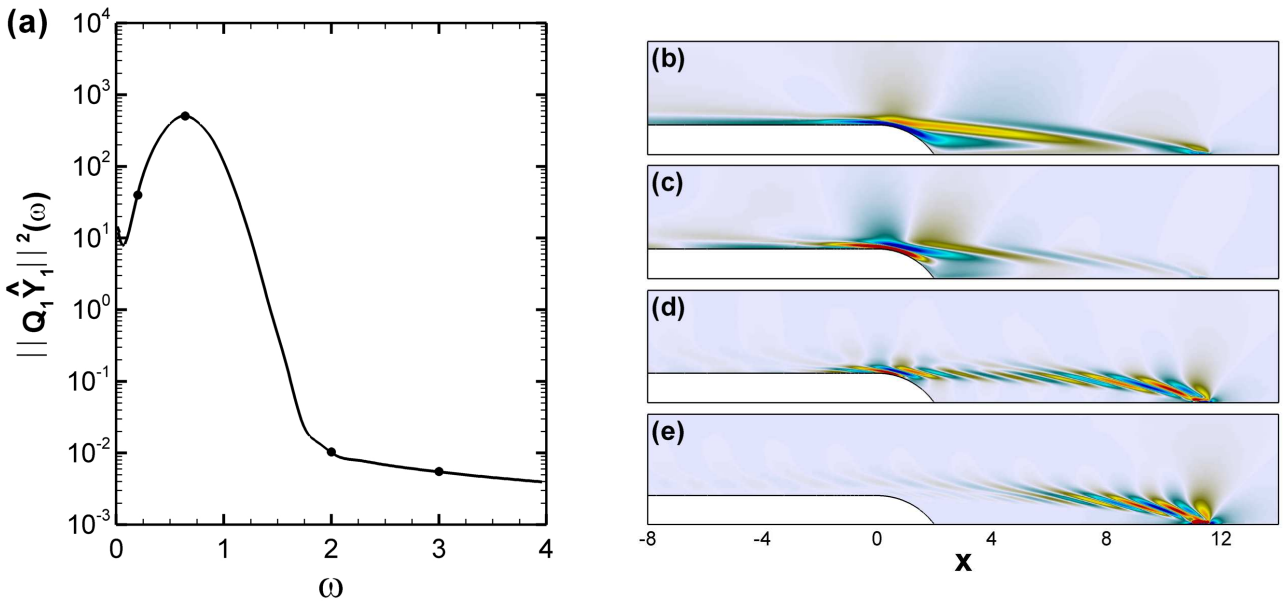


Figure 2.3: The left plot (a) shows the energy of the adjoint snapshots as a function of ω. Note that, owing to the definition (2.10) of the snapshots, the relevant quantity to be measured is $\mathsf{Q}_1\hat{\mathsf{Y}}_1$. Figures (b),(c),(d) and (e) represent the real parts of the longitudinal velocity of the adjoint snapshots associated with the frequencies $\omega_i = 0.2$, 0.64, 2 and 3 respectively.

snapshots correspond to the long time responses of the adjoint system to a harmonic forcing of pulsation ω from the sensor. The energy $\|\mathsf{Q}_1\hat{\mathsf{Y}}_1\|^2$ displays a peak at a nearby frequency of $\omega \approx 0.64$ and, analogously, the adjoint flow patterns are convected backward in time from the sensor to the upstream flow. A physical interpretation of the adjoint snapshots may arise by considering the transfer function of the system (2.5). The transfer function $\hat{\mathsf{G}}(\omega)$ links the fourier transform of the input $\hat{u}(\omega)$ to that of the output $\hat{m}(\omega)$ by

$$\hat{m}(\omega) = \hat{\mathsf{G}}(\omega)\hat{u}(\omega) \tag{2.24}$$

and can be computed by $\hat{\mathsf{G}}(\omega) = \mathsf{C}(j\omega\mathsf{I} - \mathsf{A})^{-1}\mathsf{B}$ or, introducing the adjoint snapshot $\hat{\mathsf{Y}}_1(\omega)$, by

$$\hat{\mathsf{G}}(\omega) = \hat{\mathsf{Y}}_1^*(\omega)\mathsf{Q}_1\mathsf{B} \tag{2.25}$$

This means that, given a frequency ω_0, the component $\hat{m}(\omega_0)$ of the measure is proportional to $\hat{\mathsf{Y}}_1^*(\omega_0)\mathsf{Q}_1\mathsf{B}$, i.e. the inner product between the actuator B and the frequential snapshot $\hat{\mathsf{Y}}_1(\omega_0)$. In other words, the flow structures excited by the output and shown in figure 2.3(b-e) are also the states to which the sensor is the most sensitive at this given frequency excitation. In our case, we observe that these forcing structures leading to maximum measurements at the sensor are upstream-tilted patterns located along the shear layer. They are leant against the shear so as to optimally trigger the Orr mechanism and are then amplified through the shear layer. Analogously to the direct snapshots, the spatial support of the adjoint states strongly depends on ω. For higher frequencies, smaller flow structures are localized in the vicinity of the sensor due to a higher diffusion while smallest frequency responses are largely extended upstream.

2.3.3 Reduced-order models

The previously computed frequential snapshots are used within the procedure outlined in section 2.2 to build BPOD/POD based reduced-order models. We have represented on figure 2.4(a) the first 14 HSVs σ_j. As expressed in section 2.2, the HSVs provide a way to assess the controllability and observability

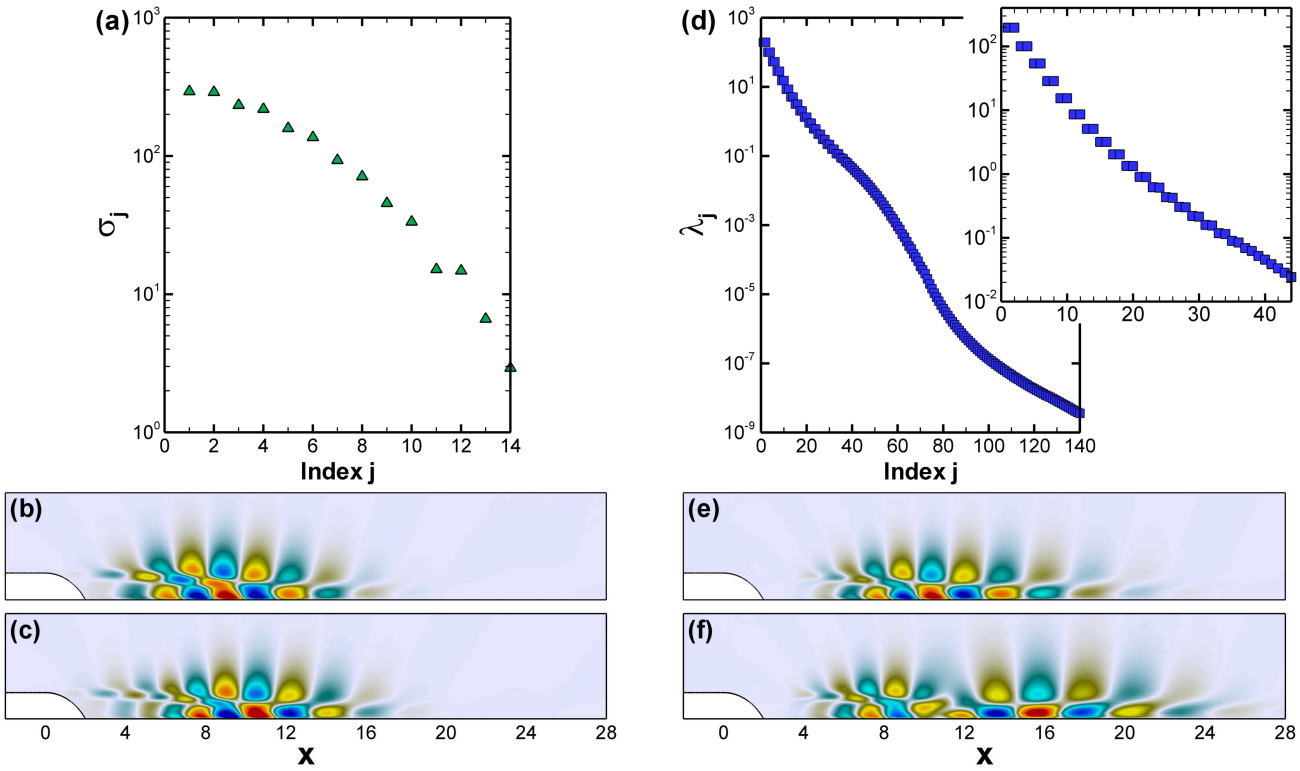


Figure 2.4: (a) First 14 HSVs σ_j and (d) first 140 POD eigenvalues λ_j. (b) and (c) stand for the streamwise velocity component of the first and third BPOD modes. Analogously, (e) and (f) stand for the first and third POD modes.

of its associated modes. This quantity is naturally used to decide on a truncation point and thus on the size of the reduced-order models. A significant drop in the HSVs is observed and justifies the truncation of the balancing basis. Similarly, we have represented on figure 2.4(d) the first 140 POD eigenvalues λ_j. Note that they rank the associated POD modes according to how easily they can be influenced by the input, i.e. their controllability. We have represented in figure 2.4(b,c) the first and third BPOD modes, visualized by their streamwise velocity component. They consist of wavepackets resulting from the amplification of vortices along the shear layer and spatially localized between the actuator and the sensor. Figures 2.4(e,f) are devoted to the same purpose for the first and third POD modes. It should be noted that they reduce to the structures that are the most easily influenced by the input. In other words, low energy is needed to force these large-scale structures downstream owing to the amplification provided by the intrinsic flow dynamics. They also appear as wavepackets but are somewhat more spatially extended downstream where the energy of the response to forcing is the largest. This latter observation is consistent with other recent works, see Bagheri *et al.* (2009*c*,*b*).

Interestingly we notice a pairwise occurrence of the eigenvalues. Looking at figure 2.4(a), the first 6 HSVs come in pairs while it is even more obvious on figure 2.4(d) for the first 30 POD eigenvalues. According to the works by Aubry *et al.* (1991, 1992) based on the time-space bi-orthogonal decomposition of complex signals, the POD modes of a traveling wave consist of degenerated pairs of modes, having the same eigenvalues and having shifted spatial structures of one another. In our case, it is indeed recovered that the eigenmodes of a given pair are a quarter period out of phase. This result holds for both BPOD and POD modes. This confirms that our models are predominantly governed by the dynamics of traveling waves. Recalling that POD models are meant to recover the exact flow field response, contrary to BPOD models (see section 2.2.4.1), it is natural to observe a clearer representation of the original traveling wave packet in the case of POD modes. For higher BPOD or POD modes, the pairwise occurrence of the eigenvalues gradually vanishes. One may attribute this result to a deviation from traveling waves due to

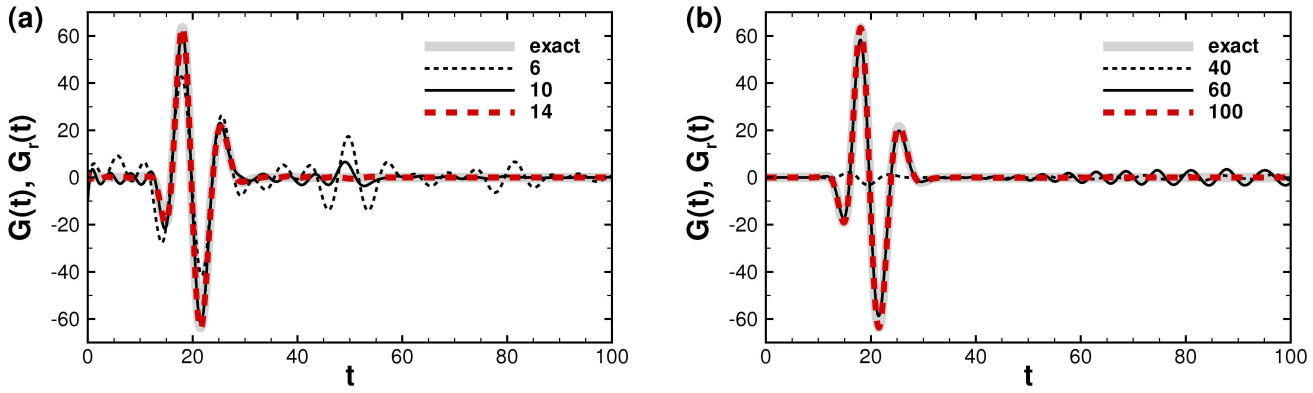


Figure 2.5: (a) Impulse response of the full system $\mathsf{G}(t)$ and of the reduced-order models $\mathsf{G}_r(t)$ for (a) BPOD models and (b) POD models.

the modulation of the wave packet in the streamwise direction, see Aubry & Lima (1995); Aubry *et al.* (2003).
It should be mentioned at this stage that our 399 complex and equally spaced snapshots are sufficient for accurate computation of the BPOD and POD modes, since for a larger number of snapshots, with finer spacing or larger frequency interval, there is no considerable change in the eigenvalue spectrum.

2.3.4 Impulse response and transfer function

By definition, the input-output behavior links the effect of the actuator on the flow to the information extracted by the sensor. It can be described by the impulse response or, equivalently, by the transfer function. The impulse response of a linear system is important, since the response of the system to any input can be found from the convolution of the impulse response with the input. The impulse response $\mathsf{G}(t)$ of the full system is obtained numerically from a time-stepper simulation of the linearized Navier-Stokes equations (2.4) for the control law $c(t) = \delta(t)$, $\delta(t)$ being the dirac function. In other words, the impulse response is $\mathsf{G}(t) = \mathsf{C}\mathsf{X}_1(t)$ where $\mathsf{X}_1(t)$ is solution of the initial-value problem

$$\frac{d\mathsf{X}_1}{dt} = \mathsf{A}\mathsf{X}_1 \tag{2.26a}$$

$$\mathsf{X}_1(t=0) = \mathsf{B} \tag{2.26b}$$

The corresponding simulation is integrated in time using a second-order accurate scheme. The impulse responses of the ROMs $\mathsf{G}_r(t)$ are computed directly by $\mathsf{G}_r(t) = \mathsf{C}_r e^{\mathsf{A}_r t}\mathsf{B}_r$ where A_r, B_r and C_r stand for the matrices A, B and C projected onto r modes. Figures 2.5(a,b) display the impulse response of the full system and those of the BPOD and POD reduced-order models respectively. The impulse response of BPOD models with $r = 6$, $r = 10$ and $r = 14$ are represented on figure 2.5(a) and those of POD models with $r = 40$, $r = 60$ and $r = 100$ are represented on figure 2.5(b). With $r = 14$ BPOD modes or $r = 100$ POD modes, we observe that reduced models register the same signal as the full model. The impulse at the actuator B generate a wavepacket that travels along the shear layer until it reaches the sensor at $t \approx 10$. The effect of this wavepacket is measured by the sensor until $t \approx 35$ and is then further convected downstream to eventually leave the computational domain.
Since all frequencies are equally excited by an initial impulse, a convenient and alternative way of expressing the input-output behavior of a linear system is to switch to the frequency domain. The frequency response of the full system and the ROMs are compared next. The transfer function of the full system is simply given by $\hat{\mathsf{G}}(\omega) = \int_0^\infty e^{-j\omega t}\mathsf{G}(t)dt$ while those of the ROMs are obtained, using equivalent notations, by $\hat{\mathsf{G}}_r(\omega) = \mathsf{C}_r(j\omega\mathsf{I}_r - \mathsf{A}_r)^{-1}\mathbf{B}_r$ where I_r stand for the identity matrix of size r. Note that the exact transfer function is easily computed by measuring (multiplying by C) each already

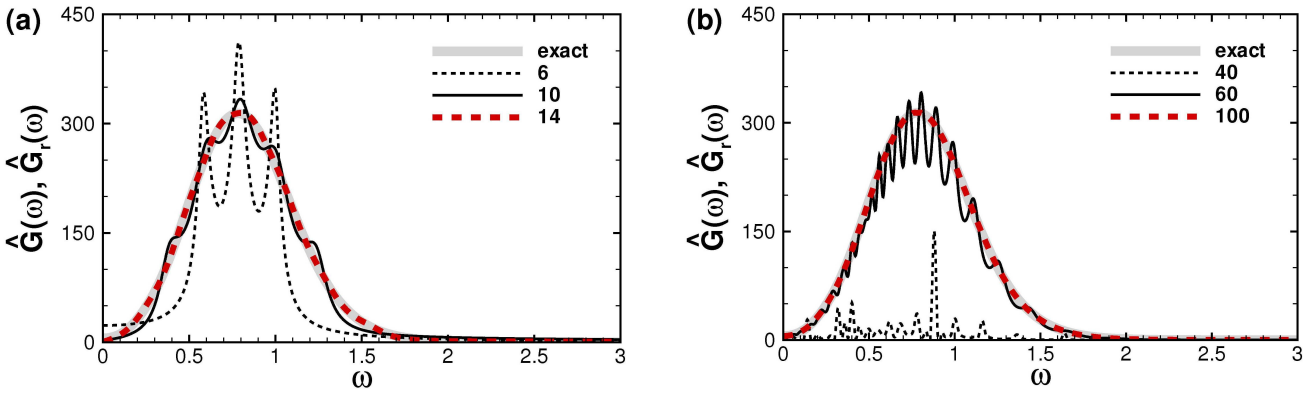


Figure 2.6: Transfer function of the full system $\hat{\mathsf{G}}(\omega)$ and of the reduced-order models $\hat{\mathsf{G}}_r(\omega)$ for (a) BPOD models and (b) POD models.

computed direct snapshots $\hat{\mathsf{X}}_1^*(\omega_i) = (j\omega_i \mathsf{I} - \mathsf{A})^{-1}\mathbf{B}$. Figures 2.6(a,b) depict the absolute values of the transfer functions of the full system together with those of BPOD and POD models respectively. We recover a preferred frequency around $\omega = 0.78$ which corresponds to the pseudo-pulsation observed in figure 2.5. Note that this pulsation is also associated with the amplification of the impulse perturbation through the shear layer due to the Kelvin-Helmholtz instability. In terms of models efficiency, the same conclusion arises since 14 BPOD and 100 POD modes are respectively sufficient to capture the most important trends of the input-output behavior.
For the sake of completeness, we also quantify the performance of the models in capturing the input-output behavior by computing the H_∞ relative norm of the error. For a model if size r, this error, denoted by $e_\infty(r)$ is given by

$$e_\infty(r) = \frac{\max_{\omega\in\mathbb{R}} |\hat{\mathsf{G}}(\omega) - \hat{\mathsf{G}}_r(\omega)|}{\max_{\omega\in\mathbb{R}} |\hat{\mathsf{G}}(\omega)|} \tag{2.27}$$

The choice of this norm is motivated by the availability of theoretical bounds on the discrepancy between the approximate and exact transfer functions for balanced truncation. In figure 2.7(a) and 2.7(b), we have represented the error $e_\infty(r)$ for the BPOD and POD models respectively. As expected, the error falls off in both cases, assessing the efficiency of these models. In particular, we observe that the number of required POD modes to reach a given error e_∞ is significantly higher than the number of required BPOD modes. Indeed, the BPOD error falls quite rapidly and remains bounded between the lower (valid for any ROM) and the upper (valid for balanced truncation models) theoretical bounds (Antoulas (2005); Rowley (2005)) given by:

$$\sigma_{r+1} < ||\hat{\mathsf{G}}(\omega) - \hat{\mathsf{G}}_r(\omega)||_\infty \leq 2 \sum_{j=r+1}^{n_1} \sigma_j \tag{2.28}$$

where n_1 is the dimension of the full system. We conclude that both models succeed in capturing the full input-output behavior of the system with a superiority of BPOD models over POD ones. This latter point is consistent with existing literature, see Barbagallo *et al.* (2009).

2.4 Globally unstable case: the square cavity flow

Here we deal with the case of a two-dimensional square cavity flow. We consider the same configuration, input and output as those used by Barbagallo *et al.* (2009) whose base flow exhibit 8 unstable global modes. Due to the unstable nature of the flow, any initial disturbance is amplified, leading to large-amplitude perturbations for sufficiently large times. Recalling the linear dynamics assumption introduced

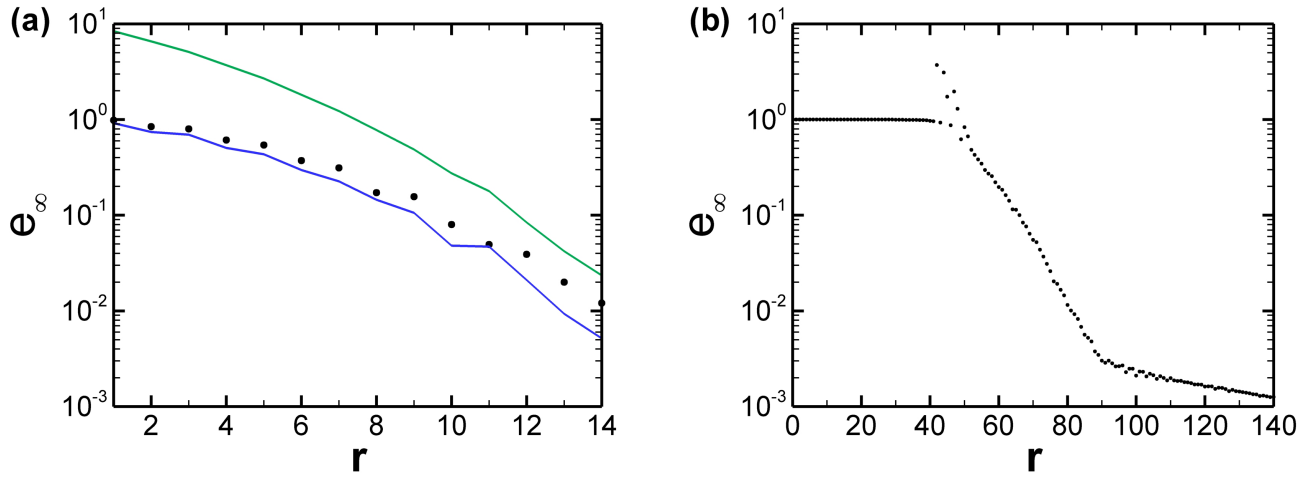


Figure 2.7: H_∞ relative norm of the error e_∞ as a function of the size r of the reduced order models for (a) the BPOD and (b) the POD modes. Note that the upper and lower bounds on the error, computed by equation (2.28), have been reported in (a) by solid lines.

in section 2.2, we are concerned here with the early development of the perturbations, i.e. the stage where the perturbation amplitude remains weak. In the reference work (Barbagallo *et al.* (2009)), the authors have designed reduced-order models for flow control purpose. Particularly, they showed that BPOD and POD modes were successful in capturing the input-output behavior of their flow. They used a snapshot method based on a temporal definition of the Gramians, restricting their approach to the stable subspace dynamics. Consequently they had to model the unstable dynamics separately, using the unstable global modes. The second numerical example presented here is used as a reference case to show that frequential snapshots enable to reduce the system without separating the unstable and stable subspaces and that the resulting ROMs yield the correct full input-output dynamics (meaning both the correct stable and unstable input-output behaviors). To this end, we reduce the system using BPOD and POD modes, as in the previous section and we next compare our results to those obtained by Barbagallo *et al.* (2009).

2.4.1 Flow configuration

We briefly describe the flow configuration studied in this section. The square cavity flow of interest has first been introduced in Sipp & Lebedev (2007) where a more detailed description of the geometry and boundary conditions is available. It consists of a uniform incoming flow over a square cavity. The Reynolds number based on the uniform upstream velocity and cavity depth is fixed to $Re = 7500$. The corresponding base flow, defined as a solution of the steady Navier-Stokes equations, is displayed on figure 2.8 by its streamlines together with the actuator and sensor location. Note that the boundary layer starts developing at $x = -0.4$ (the origin of the coordinate system coincides with the top left corner of the square cavity).

Following the work by Barbagallo *et al.* (2009), we choose an actuator consisting of a parabolic normal velocity blowing near the upstream edge of the cavity, over the streamwise extent $(-7/20 \leq x \leq 0, y = 0)$. Additionally, they adopted a transformation referred to as lifting in order to formulate the problem as a driven homogeneous state space system. We will not go into further details about this matter as it is not our point. Yet, it is important to note that their input B used to perform BPOD or POD ROMs is solution of the steady but inhomogeneous Navier-Stokes equations with boundary conditions given for $(-7/20 \leq x \leq 0, y = 0)$ by the velocity profile

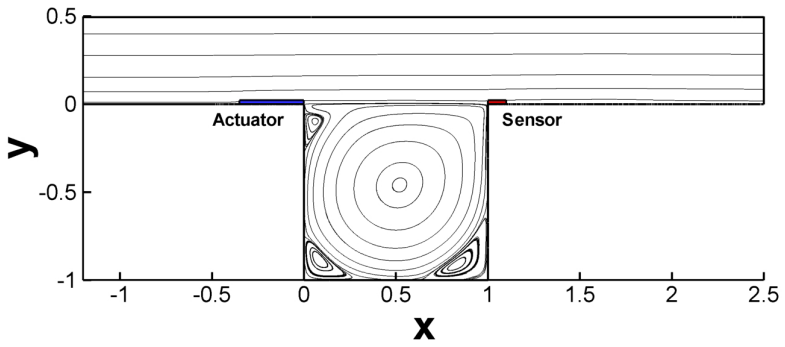


Figure 2.8: Streamlines of the base flow at $Re = 7500$. The actuator and sensor locations are also depicted.

$$u(x, y = 0, t) = 0 \tag{2.29}$$

$$v(x, y = 0, t) = -\frac{x(1600x + 560)}{147} \tag{2.30}$$

The solution of this steady problem and associated inhomogeneous boundary conditions constitutes the actuation which is referred to as the input B on the finite input-output system (2.5). The sensor is located near the downstream edge of the cavity, on the segment $(1 \leq x \leq 1.1, y = 0)$. The output is the wall-normal shear stress evaluated at and integrated over a localized region of the wall:

$$m(t) = \int_{x=1}^{x=1.1} \left.\frac{\partial u}{\partial y}\right|_{y=0} (t)\, dx \tag{2.31}$$

Similarly, this measurement is expressed by the matrix C on the finite system (2.5) resulting from the discretization on a given mesh. Note that we choose the same numerical finite element approach based on the same non-structured triangular mesh as that used by Barbagallo *et al.* (2009) in order to avoid any numerical effect. This one results in $n_1 \approx 780\,000$ degrees of freedom stemming from about 200 000 triangles.

As previously mentioned, the base flow is globally unstable at $Re = 7500$. A global stability analysis of the flow is presented in Barbagallo *et al.* (2009), we briefly summarize their results here. The system of equations (2.5) allows the computation of a temporal global spectrum and associated modes via the common assumption of an exponential time-dependence expressed as $\mathsf{X}_1(x, y, t) = \tilde{\mathsf{X}}_1(x, y)e^{\gamma t}$ where $\gamma \in \mathbb{C}$ is the eigenvalue and $\tilde{\mathsf{X}}_1$ the eigenvector given by

$$\mathsf{A}\tilde{\mathsf{X}}_1 = \gamma\tilde{\mathsf{X}}_1 \tag{2.32}$$

Eigenvalues are decomposed into $\gamma = \alpha + j\omega$ where α is the amplification rate and ω the pulsation so that unstable modes are characterized by $\alpha > 0$. The low-frequency part of the eigenspectrum, corresponding to the most unstable global modes, has been represented on figure 2.9. This stability analysis displays 4 unstable "physical" global modes, i.e. 8 if the complex conjugates are counted. These four unstable modes are denoted by E_{-3}, E_{-2}, E_{-1} and E_0 as in Barbagallo *et al.* (2009). Furthermore, we have also highlighted a stable global mode, denoted by E_2, which displays a prominent contribution to the input-output dynamics, see Barbagallo *et al.* (2009). The role of this particular mode is illustrated below.

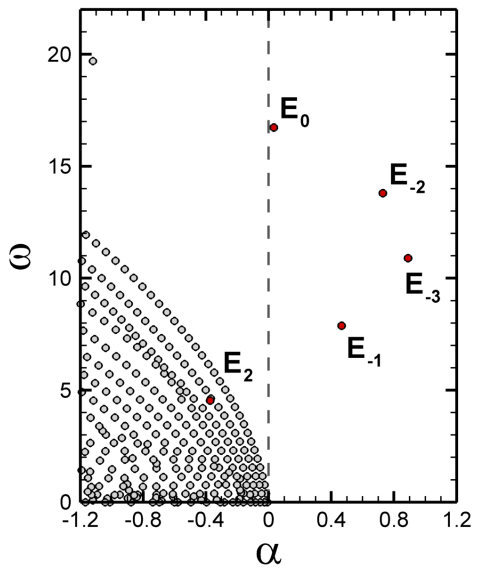


Figure 2.9: Part of the global eigenspectrum of the square cavity flow at $Re = 7500$ (taken from Barbagallo *et al.* (2009)).

2.4.2 Frequential snapshots

Similarly to the flow over the backward-facing step, the frequential direct and adjoint snapshots required for the model reduction are computed by solving equations (2.16) and (2.17). We have computed 499 equispaced complex frequential snapshots $\hat{\mathsf{X}}_1$ and $\hat{\mathsf{Y}}_1$ from $\omega = 0$ to $\omega = 35$, resulting in $n_d = n_a = 998$. Since the full system is unstable, these complex flow states can no longer be interpreted physically as long time responses to harmonic forcings. However, the inversion required to solve equations (2.16) and (2.17) remain tractable as far as there are no marginal modes, which is the case here. We have represented on figure 2.10 the evolution of the snapshots kinetic energy $\|\hat{\mathsf{X}}_1\|^2$ and $\|\mathsf{Q}_1\hat{\mathsf{Y}}_1\|^2$ as a function of ω. We observe that both quantities display 5 peaks and eventually decrease quite abruptly for higher frequencies. These peaks are actually resulting from a pseudo-resonance with nearby global modes. This statement is argued by the representation of the frequencies of the corresponding global modes by dashed lines. To be more precise, the four highest frequencies depicted on figure 2.10 correspond to the four isolated unstable global modes labeled E_{-3}, E_{-2}, E_{-1} and E_0. Within this set, we observe that the modes having the smallest growth rate lead to the highest energy peak, emphasizing the pseudo-resonance phenomenon. As for the first low-frequency peak at $\omega \approx 4.54$, it is related to the stable global mode E_2. This particular mode has been shown (Barbagallo *et al.* (2009)) to have an important contribution to the input-output behavior, more precisely, it is both strongly controllable and observable. Interestingly, we recover a peak of energy for both the direct and adjoint snapshots at its frequency.

2.4.3 Reduced-order models

Following the procedure introduced in section 2.2, the snapshots are used to build BPOD and POD based ROMs as in the case of the backward-facing step flow. The first 40 HSVs σ_j and first 200 POD eigenvalues λ_j computed for the square cavity flow are represented on figures 2.11(a) and (b) respectively. Analogously to the results presented in section 2.3, a significant drop in the HSVs is observed and the first ones are seen to come in pairs due to the representation of traveling structures by the superposition of modes that are 90° out of phase. Moreover, the same observations hold for the POD eigenvalues. Similarly to the previous numerical example, it was found that the 499 computed snapshots are sufficient for an accurate computation of the BPOD/POD modes and eigenvalues.
The resulting ROMs defined by equation (2.6) are found to be unstable with both BPOD and POD

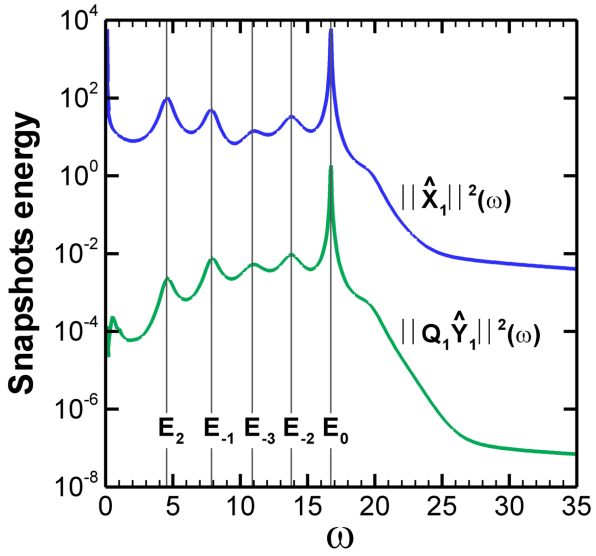


Figure 2.10: Energy of the direct $\hat{\mathsf{X}}_1$ and adjoint $\mathsf{Q}_1\hat{\mathsf{Y}}_1$ snapshots as a function of ω. The lines indicate the frequencies of nearby global modes E_2, E_{-1}, E_{-3}, E_{-2} and E_0 (ordered from left to right).

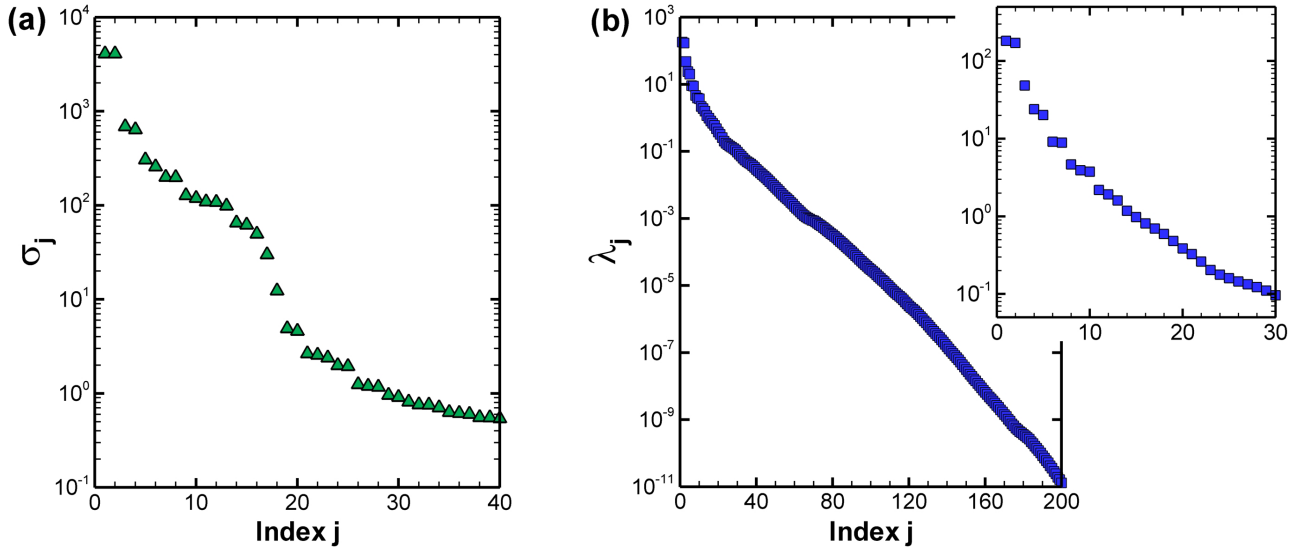


Figure 2.11: (a) First 40 HSVs σ_j and (b) first 200 POD eigenvalues λ_j.

models. In other words, the reduced matrices A_r now possess unstable eigenmodes, which is not surprising since the original system is unstable. For unstable systems, the usual transfer functions are no more defined. Yet, one can still compute the "frequency response" of the full unstable system $\hat{\mathsf{G}}(\omega) = \mathsf{C}(j\omega\mathsf{I} - \mathsf{A})^{-1}\mathsf{B}$ and of the ROMs $\hat{\mathsf{G}}_r(\omega) = \mathsf{C}_r(j\omega\mathsf{I}_r - \mathsf{A}_r)^{-1}\mathsf{B}_r$ as long as there are no marginal modes. This latter quantity may be useful to assess the performance of the ROMs in capturing the full input-output dynamics, all the more so as Zhou *et al.* (1999) demonstrated that the upper theoretical bound on the H_∞ transfer function error $e_\infty(r)$, given in equation (2.28), still holds for the balanced truncation of possibly unstable systems. We have represented on figure 2.12 the relative error $e_\infty(r)$ defined by equation (2.27) in the case of the cavity flow together with theoretical bounds given by (2.28). The error is indeed observed to lie between these bounds and decrease while increasing the size r of the ROMs, linking favorably our results to the theoretical predictions by Zhou *et al.* (1999). This first result illustrates the relevancy and ability of BPOD models to capture the over-all input-output dynamics of our unstable system. However, we next provide a deeper insight into the performance of the BPOD and

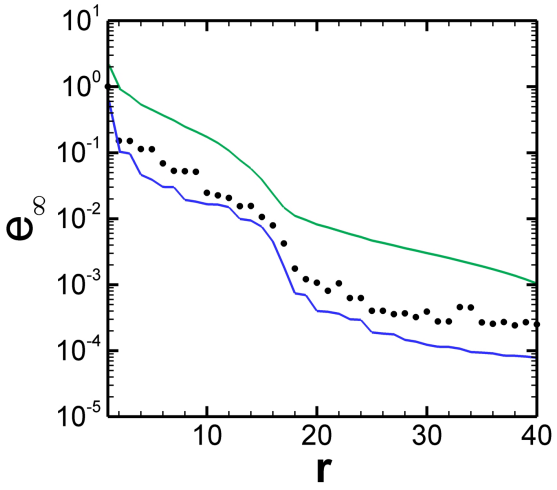


Figure 2.12: Relative error e_∞ as a function of the size r of the ROMs for the BPOD models. The upper and lower bounds on the error, computed by equation (2.28), are also displayed by the upper and lower solid lines.

POD models in modeling the unstable and stable input-output behavior. This issue has been largely investigated by Barbagallo *et al.* (2009) as they showed that the accurate modeling of each subspace is essential when it comes to effective closed-loop control design. As a result, particular attention is given to them in the remaining of this section.

2.4.3.1 Comparison of the unstable subspaces

Because of their ability to model the inherent instabilities and because of their low dimensionality, the unstable global modes come out as the most natural basis of the unstable subspace. Besides, Ahuja & Rowley (2010) and Barbagallo *et al.* (2009) directly used them to represent the dynamics of their unstable subspace, leading to an "exact" model since no modeling assumptions are invoked. The unstable dynamics of the ROMs is thus compared to the original one through their unstable modes.
The eigenspectrum and global modal decomposition of the ROMs is obtained by directly computing the eigenvalues/eigenvectors of the reduced matrix A_r. We have represented on figure 2.13(a) the number of unstable eigenmodes as a function of the size r of the BPOD models. it is observed that models of sufficiently high size eventually exhibit 8 unstable global modes just like the full system. The number of unstable modes rise quickly to reach 8 for a model of size $r = 15$ which is highlighted on the figure by a dashed line. We note that BPOD models of size 24 and 32 possess 9 unstable global modes, which may be attributed to the sensitivity of the procedure to numerical issues. Figure 2.13(b) represents the same information for POD models. Similarly, the number of unstable modes rises until the value of 8 but this increase is much more erratic than for BPOD models. Additionally, the model size required to get the 8 unstable modes is $r = 82$ (depicted by a dashed line on the figure) which is far more important than for BPOD models.
The unstable part ($\alpha > 0$) of the eigenspectrum of the full system and those of several BPOD and POD models are depicted on figure 2.14(a) and 2.14(b) respectively. Since the eigenspectrum is symmetric about the axis ($\omega = 0$), only the upper complex half-plan ($\omega > 0$) is represented. The computation of the full system eigenspectrum (and eigenvectors) is based on a classical shift and invert iterative Arnoldi algorithm. The results on BPOD models, depicted on figure 2.14(a), show that the unstable eigenvalues of the ROMs quickly tend to those obtained by Barbagallo *et al.* (2009) on the full system as the model size r is increased. As for POD models, a similar behavior is observed on figure 2.14(b) though much more modes are required to recover the unstable global modes of the full system.
This convergence of the unstable eigenmodes is further outlined in table 2.1 where we have listed the

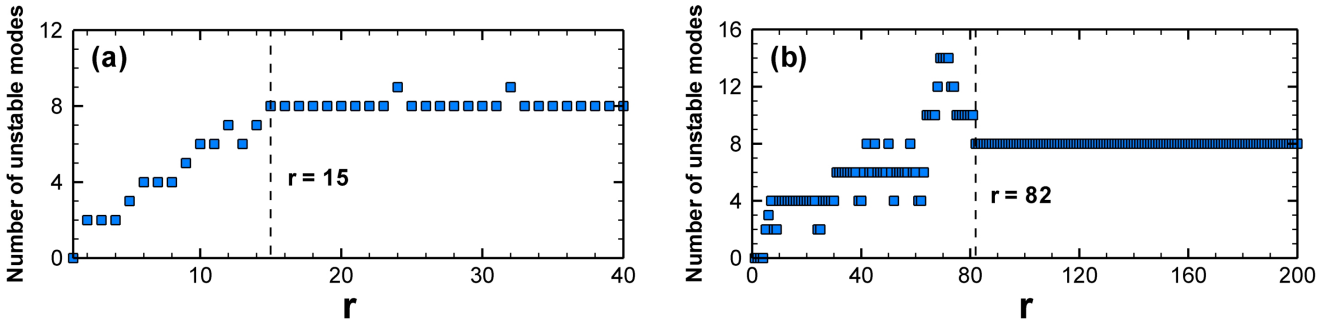


Figure 2.13: Number of unstable modes for (a) the BPOD models and (b) the POD models. Both ROMs exhibits 8 unstable modes from the dashed lines standing for $r = 15$ and $r = 82$ respectively.

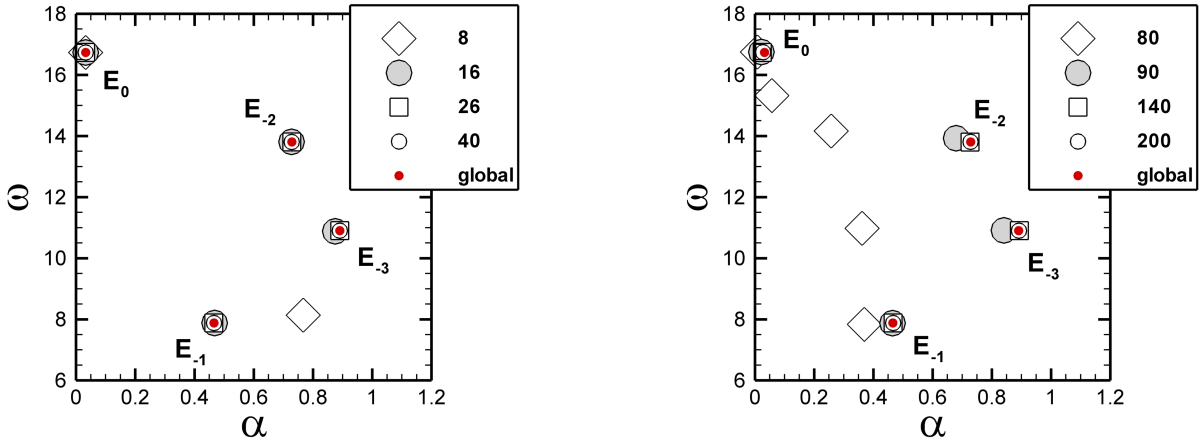
		E_{-3}		E_{-2}		E_{-1}		E_0	
	Size	α	ω	α	ω	α	ω	α	ω
BPOD	15	0.874806	10.77852	0.723728	13.8188	0.464693	7.884405	0.032403	16.73146
	20	0.889216	10.8997	0.728018	13.8042	0.465734	7.88121	0.0324203	16.7315
	30	0.890282	10.9011	0.728619	13.8037	0.465488	7.88134	0.0324236	16.7315
	40	0.890204	10.9004	0.728317	13.8039	0.465577	7.88133	0.0324238	16.7315
POD	82	0.540170	10.93990	0.417714	14.20645	0.401845	7.842179	0.020019	16.74988
	120	0.915394	10.8931	0.748506	13.7754	0.47076	7.88286	0.0303051	16.7293
	160	0.88925	10.9022	0.722727	13.8058	0.466002	7.88255	0.0215329	16.7345
	200	0.891552	10.9028	0.727882	13.8094	0.466655	7.88173	0.0275948	16.7415
Full System	778 410	0.890451	10.9008	0.728513	13.8037	0.465557	7.88173	0.032426	16.7315

Table 2.1: Growth rates α and pulsation ω of the unstable modes labeled E_{-3}, E_{-2}, E_{-1} and E_0 of the full system and those of several ROMs. The size column stand for the size r of the ROMs and is equal to n_1 for the full system.

growth rate α and pulsation ω of the unstable modes of the full system and those of several ROMs. Note that the column titled "Size" stands for the eigenvalue problem size, that is r for the ROMs and n_1 for the full system. We considered that the unstable modes labeled E_{-3}, E_{-2}, E_{-1} and E_0 could be clearly identified for BPOD models with $r \geq 15$ and POD models with $r \geq 82$ as their unstable global modes frequency approximates correctly those of the full system with a 2 digits accuracy. The superiority of BPOD models over POD ones is more quantitatively illustrated by these results. Note that some further information on the eigenspectrum of POD models is provided in appendix 2.C.
We now turn our attention to the unstable subspaces. In our case, they are low dimensional (8 in the case of the full system) and their most natural basis is simply given by the unstable eigenvectors. By computing the unstable eigenvectors of the ROMs, we come to the conclusion that, for models of sufficient size, they match those of the full system once rebuilt in their original basis. This is illustrated in figure 2.15(a,b) where we depict the real part of the longitudinal velocity of the most unstable global mode E_{-3} associated, respectively, with the full system and a POD model of size 150. Note that the mode built from the POD ROM is computed by $\mathsf{R}_1\tilde{\mathsf{X}}_{1r}$ where $\tilde{\mathsf{X}}_{1r}$ is the reduced unstable eigenvector. The two resulting flow structures are observed to be almost indistinguishable. Furthermore, this latter observation holds (not shown here) for the 3 other unstable modes and for BPOD ROMs (computed by $\mathsf{T}_1\tilde{\mathsf{X}}_{1r}$). In this subsection, we have thus demonstrated that both the BPOD and POD procedures (for models of sufficient size) are successful in modeling the original unstable subspace insofar as the unstable subspace of the ROMs possesses the same natural basis.

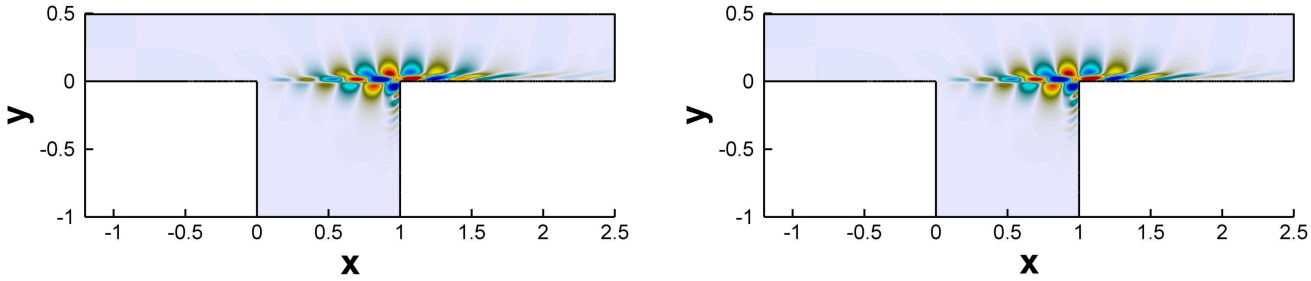
2.4.3.2 Comparison of the stable input-output dynamics

The stable subspace is, as often, high dimensional and the main effort in reducing the system's dimension is expended reducing the dynamics in the stable subspace. Furthermore, it contains substantial physical



(a) Eigenspectrum of BPOD models. *(b) Eigenspectrum of POD models.*

Figure 2.14: Unstable eigenspectrum of the full system versus those of BPOD/POD reduced models.



(a) Solution of a shift and invert Arnoldi algorithm. *(b) Solution of the reduced-order model built with 150 POD modes.*

Figure 2.15: Real part of the longitudinal velocity of the most unstable eigenvector E_{-3}.

information about the input-output behavior we are interested in. Similarly to previous studies (Ahuja & Rowley (2010); Barbagallo *et al.* (2009)), we assess the accuracy of the ROMs in modeling the exact stable input-output dynamics by computing their stable transfer function.
The stable transfer function of the full system can be computed either by $\hat{\mathsf{G}}_s(\omega) = \mathsf{C}(j\omega\mathsf{I}-\mathsf{A}_s)^{-1}\mathsf{B}$, where A_s is the restriction of A on its stable subspace, or by the Fourier transform of the stable impulse response defined by $\mathsf{G}_s(t) = \mathsf{C}e^{\mathsf{A}_s t}\mathsf{B}$. Here, the exact transfer function is taken from the work by Barbagallo *et al.* (2009) where they adopted the second solution. As for the ROMs, the stable transfer functions $\hat{\mathsf{G}}_{rs}(\omega)$ are directly computed by $\hat{\mathsf{G}}_{rs}(\omega) = \mathsf{C}_r(j\omega\mathsf{I}_r - \mathsf{A}_{rs})^{-1}\mathsf{B}_r$ where A_{rs} is the restriction of A_r on its stable subspace. Figure 2.16(a,b) depicts the transfer function of the full system superimposed on those of several BPOD and POD models respectively. It is observed to be very well approximated in both cases for ROMs of sufficiently high size. We recover a preferred frequency around $\omega = 4.6$, which corresponds to the frequency of the stable mode E_2 and to the first peak observed in figure 2.10. For a model of size r, this performance is quantified by computing the H_∞ relative norm of the error $e_{\infty s}(r)$ defined by:

$$e_{\infty s}(r) = \frac{\max_{\omega\in\mathbb{R}} |\hat{\mathsf{G}}_s(\omega) - \hat{\mathsf{G}}_{rs}(\omega)|}{\max_{\omega\in\mathbb{R}} |\hat{\mathsf{G}}_s(\omega)|} \tag{2.33}$$

This error is plotted on figure 2.17(a) and 2.17(b) for BPOD and POD ROMs respectively. It is meant to convey the convergence behavior of the transfer functions of the ROMs as a function of the size r of the models. Similarly to the backward-facing step flow case, the error converges to zero as the number of modes increases and a faster error decrease is observed for BPOD models compared to POD ones.
Comparing these error convergences to those observed by Barbagallo *et al.* (2009), we notice that both our BPOD/POD models require more modes to reach a given error $e_{\infty s}$. This is not surprising as our

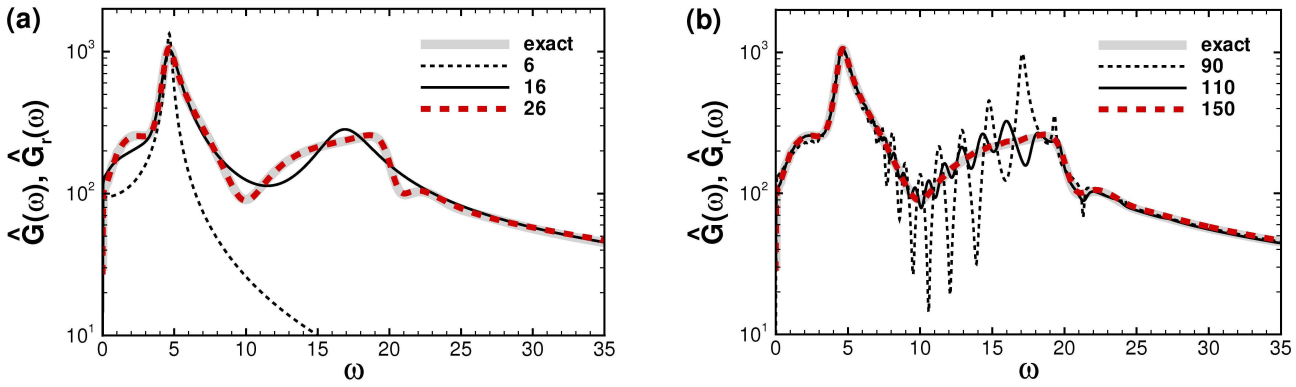


Figure 2.16: Transfer function of the full system stable part compared to those of (a) BPOD and (b) POD models.

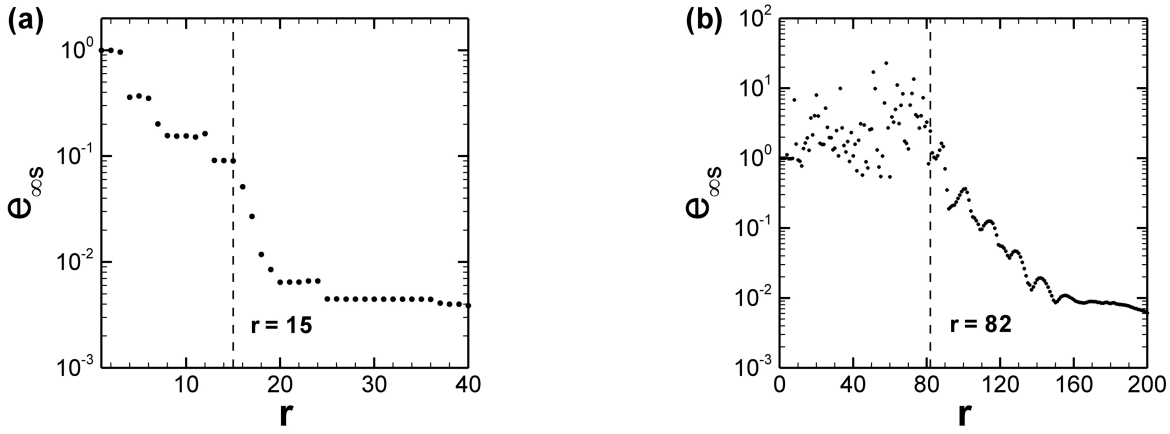


Figure 2.17: Relative norm of the error $e_{\infty s}$ as a function of the size r of the ROMs for (a) the BPOD and (b) the POD models. The limits from which the ROMs exhibit 8 unstable global modes are depicted by dashed lines standing for $r = 15$ and $r = 82$ respectively.

procedures perform the reduction of both the stable and unstable dynamics at the same time while theirs are focused on the stable dynamics. Given this fact, the performance of our models in capturing the correct stable input-output behavior is delayed compared to theirs due to the constraint of modeling the unstable subspace at the same time. This assertion is further argued on figure 2.17(a,b) where we have drawn by dashed lines the limits $r = 15$ and $r = 82$ from which the models possess 8 unstable global modes. We clearly notice a sudden fall-off on the error starting from these limits, that is, a sudden improvement of the models in capturing the stable input-output dynamics.

2.4.3.3 Assessment

We conclude from this section that both ROMs not only succeed in capturing the full unstable subspace, but also the stable input-output dynamics. Furthermore, BPOD models proved to capture these dynamics with less modes, which is in agreement with our previous results on the backward-facing step flow.

It should be mentioned here that the present BPOD procedure based on frequential definitions of the Gramians reduces the stable and unstable subspaces simultaneously, see Zhou *et al.* (1999). In other words, it is equivalent to separating the stable and unstable parts of the transfer function and performing the model reduction for both parts separately. This may explain the quick ability of BPOD ROMs to capture simultaneously and gradually the unstable (see figure 2.13(a) and 2.14(a)) as well as the stable input-output dynamics (see figure 2.17(a)).

Concerning POD models, the same idea does not hold and, as a result, their ability to model both dynamics is not as progressive. Indeed, observing figure 2.13(b), we note that the modeling of the unstable subspace is not gradual and, most importantly, we show on figure 2.17(b) that the POD ROMs are completely unable to model the correct stable dynamics until the unstable subspace is well captured from $r \approx 82$. This latter point may indicate that a bad modeling of the unstable subspace involves the incapability of POD ROMs to correctly model the stable dynamics.
Finally, it should be emphasized that our frequential approach to perform BPOD and POD model reduction is not based on a partition of the unstable and stable subspaces contrary to the previous works by Ahuja & Rowley (2010); Barbagallo *et al.* (2009). In their case, they had to: (i) compute the global eigenmodes by a shift and invert Arnoldi algorithm, (ii) perform a direct and adjoint time-stepping simulation while projecting at each time step (with the global modes) the resulting states onto the stable subspace. Their unstable dynamics are then modeled by the unstable global modes while their snapshots arising from the direct and adjoint simulations are used to build BPOD models (Ahuja & Rowley (2010); Barbagallo *et al.* (2009)) and POD models (Barbagallo *et al.* (2009)) of the stable input-output dynamics. The square cavity flow configuration studied in this section has thus illustrated the ability of the frequential snapshots to build BPOD and POD ROMs of an unstable flow system without separating the unstable and stable subspaces. This prove to be a valuable asset insofar as we neither had to compute any global modes nor to perform any projection onto the stable subspace.

2.5 Conclusion

In this chapter, we have described how the use of frequential responses of a flow to a given actuator enables to compute the basis of the most controllable modes (POD modes). Analogously, the harmonic flow states yielding the maximum contribution to the sensor energy have been introduced to compute the most and equally controllable and observable modes: the balanced modes (BPOD modes). ROMs have been designed by the projection of the full original system of equations onto BPOD/POD modes computed from the so-called frequency snapshots.
As a first step, the whole procedure has been carried out on a stable linear system: the flow over a rounded backward-facing step. The first example stands for a well-known noise amplifier flow in the sense that small perturbations can be strongly amplified through the shear layer. The computation of the frequency snapshots highlighted the frequency selection process of the flow. In particular, an energy peak of the flow response to harmonic actuation has been observed nearby the frequency associated with the Kelvin-Helmholtz instability. These frequential snapshots have then been used to build the BPOD and POD models that proved to be effective in modeling the linear input-output behavior of the flow. Both the impulse response and transfer function are recovered by the ROMs and a better efficiency of BPOD models to do so was observed as expected from the literature.
As a second step, we moved on to the case of a linear unstable system. To assess and quantify the ability of our procedure to reduce unstable systems, we applied the same technique on a well-known oscillator system: the flow over a square cavity. On the one hand, the resulting ROMs were shown to capture the same unstable global modes as those of the original system and, on the other hand, an accurate modeling of the stable dynamics has also been recovered by investigating the stable transfer functions. Similarly, a superiority of BPOD models over POD ones was noticed. Contrary to the work by Barbagallo *et al.* (2009) where the authors made a partition to model the stable (using BPOD/POD ROMs based on temporal snapshots) and unstable (using global modes) subspaces separately, our model reduction technique has proved to model efficiently at the same time both dynamics.
In summary, we proved the possibility and efficiency of frequential snapshots to yield BPOD and POD models that are: (i) identical to those computed from temporal snapshots for stable systems, (ii) operational for unstable systems without separating the unstable and stable subspaces. Therefore, this contribution on model reduction seems to be a promising alternative tool to compute POD modes or to approximate balanced truncation. We hope that the present contribution will somehow aid in the

design of frequency-based ROMs in view of building efficient closed-loop flow controllers.

Chapter appendix

2.A Formulation of the incompressible linearized Navier-Stokes equations as a standard state-space system

The design of reduced-order models requires to reformulate the linearized Navier-Stokes equations into a standard state-space form. Yet, this is not trivial since the matrix before the time derivative is not invertible because of incompressibility. To this end, we proceed as in Barbagallo *et al.* (2009). We first multiply the momentum equation by $\mathsf{A}_2\mathsf{Q}_1^{-1}$, which yields, assuming that $\mathsf{A}_2\dot{\mathsf{X}}_1 = 0$, an expression for the pressure in terms of the velocity field

$$\mathsf{X}_2 = -\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_2^*\right)^{-1}\left[\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_1\right)\mathsf{X}_1 + \mathsf{A}_2\mathsf{B}_1 c(t)\right] \tag{2.34}$$

This relation can be used to eliminate the explicit divergence constraint and allows us to write the governing linearized equations in the desired form

$$\frac{d\mathsf{X}_1}{dt} = \mathsf{P}_1\mathsf{A}_1\mathsf{X}_1 + \mathsf{P}_1\mathsf{Q}_1\mathsf{B}_1 c(t) \tag{2.35a}$$

$$m(t) = \mathsf{C}_1\mathsf{X}_1 \tag{2.35b}$$

where

$$\mathsf{P}_1\mathsf{Q}_1 = \mathsf{I} - \mathsf{Q}_1^{-1}\mathsf{A}_2^*\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_2^*\right)^{-1}\mathsf{A}_2 \tag{2.36}$$

is the projection matrix onto the divergence-free space. It is noteworthy that P_1 is a Hermitian operator, so that we can take advantage of the relation $\mathsf{P}_1^* = \mathsf{P}_1$. By defining $\mathsf{A} = \mathsf{P}_1\mathsf{A}_1$ and $\mathsf{B} = \mathsf{P}_1\mathsf{Q}_1\mathsf{B}_1$ and $\mathsf{C} = \mathsf{C}_1$, we recover equation (2.5).

2.B Expressing the frequential snapshots in relation to the linearized Navier-Stokes equations

2.B.1 Direct snapshots

The direct flow states $\hat{\mathsf{X}}_1$ involved in the snapshots method are defined in equations (2.10). Substituting A and B by their expression calculated in appendix 2.A leads to:

$$(j\omega\mathsf{I} - \mathsf{P}_1\mathsf{A}_1)\hat{\mathsf{X}}_1 = \mathsf{P}_1\mathsf{Q}_1\mathsf{B}_1 \tag{2.37}$$

If we introduce the pressure $\hat{\mathsf{X}}_2$ associated to $\hat{\mathsf{X}}_1$ by

$$\hat{\mathsf{X}}_2 = -\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_2^*\right)^{-1}\left[\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_1\right)\hat{\mathsf{X}}_1 + \mathsf{A}_2\mathsf{B}_1\right] \tag{2.38}$$

we get

$$j\omega\hat{\mathsf{X}}_1 - \mathsf{Q}_1^{-1}\mathsf{A}_1\hat{\mathsf{X}}_1 - \mathsf{Q}_1^{-1}\mathsf{A}_2^*\hat{\mathsf{X}}_2 = \mathsf{B}_1 \tag{2.39}$$

so that we recover equation (2.16):

$$\left[j\omega\begin{pmatrix}\mathsf{Q}_1 & 0\\ 0 & 0\end{pmatrix} - \begin{pmatrix}\mathsf{A}_1 & \mathsf{A}_2^*\\ \mathsf{A}_2 & 0\end{pmatrix}\right]\begin{pmatrix}\hat{\mathsf{X}}_1\\ \hat{\mathsf{X}}_2\end{pmatrix} = \begin{pmatrix}\mathsf{Q}_1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\mathsf{B}_1\\ 0\end{pmatrix} \tag{2.40a}$$

The finite number of flow fields $\hat{\mathsf{X}}_1(\omega_i)$ at discrete pulsations ω_i can be obtained by inverting this linear system and are stacked as columns of the matrix $\mathbf{X}_1$. The above demonstration shows that the controllability Gramian G_c, defined in (2.14) can be thought of as the spatial correlation matrix for the harmonic responses to harmonic forcings at the actuator location.

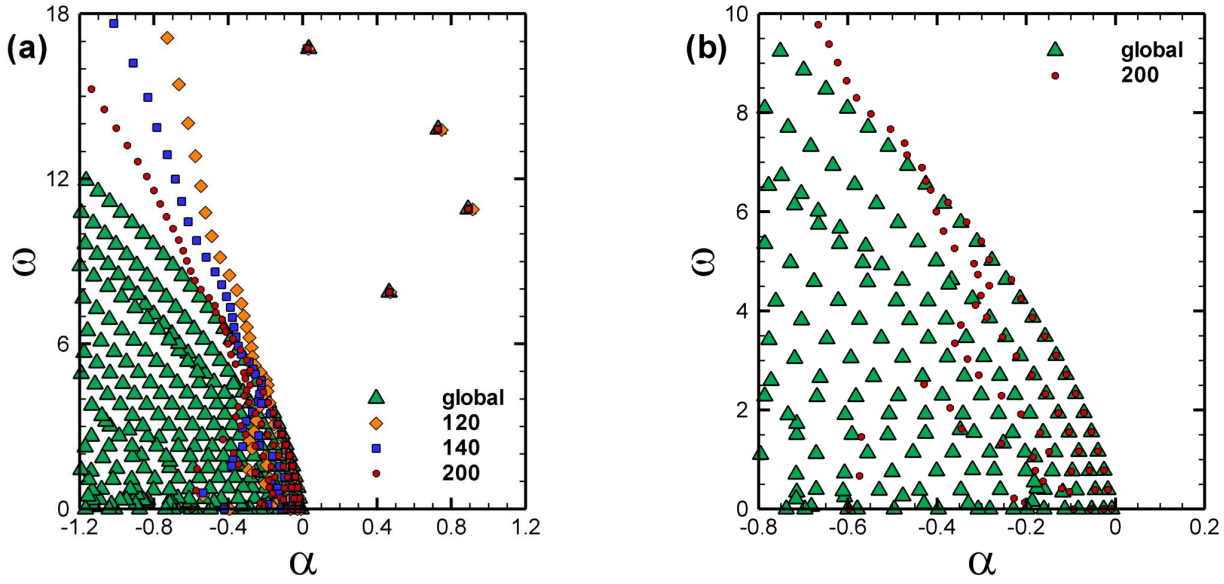


Figure 2.18: (a) Part of the eigenspectrum of POD models of size 120, 140 and 200 superimposed onto that of the full system. A closer view is depicted for the model of size 200 on (b).

2.B.2 Adjoint snapshots

The adjoint flow states $\hat{\mathsf{Y}}_1$ involved in the snapshot method are defined in equations (2.10). In a similar manner, a substitution of the expressions of A and C calculated in appendix 2.A leads to:

$$(-j\omega\mathsf{I} - \mathsf{P}_1\mathsf{A}_1^*)\mathsf{P}_1\mathsf{Q}_1\hat{\mathsf{Y}}_1 = \mathsf{P}_1\mathsf{Q}_1\left(\mathsf{Q}_1^{-1}\mathsf{C}_1^*\right) \tag{2.41}$$

If we omit the incompressibility constraint $\mathsf{P}_1\mathsf{Q}_1$ on $\hat{\mathsf{Y}}_1$, which is already applied through the evolution operator $(-j\omega\mathsf{I} - \mathsf{P}_1\mathsf{A}_1^*)$, we have:

$$(-j\omega\mathsf{I} - \mathsf{P}_1\mathsf{A}_1^*)\hat{\mathsf{Y}}_1 = \mathsf{P}_1\mathsf{Q}_1\left(\mathsf{Q}_1^{-1}\mathsf{C}_1^*\right) \tag{2.42}$$

Similarly, if we introduce the adjoint pressure $\hat{\mathsf{Y}}_2$ associated to $\hat{\mathsf{Y}}_1$ by

$$\hat{\mathsf{Y}}_2 = -\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_2^*\right)^{-1}\left[\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_1^*\right)\hat{\mathsf{Y}}_1 + \mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{C}_1^*\right] \tag{2.43}$$

we get

$$-j\omega\hat{\mathsf{Y}}_1 - \mathsf{Q}_1^{-1}\mathsf{A}_1^*\hat{\mathsf{Y}}_1 - \mathsf{Q}_1^{-1}\mathsf{A}_2^*\hat{\mathsf{Y}}_2 = \mathsf{Q}_1^{-1}\mathsf{C}_1^* \tag{2.44}$$

so that we recover equation (2.17):

$$\left[-j\omega\begin{pmatrix}\mathsf{Q}_1 & 0\\ 0 & 0\end{pmatrix} - \begin{pmatrix}\mathsf{A}_1^* & \mathsf{A}_2^*\\ \mathsf{A}_2 & 0\end{pmatrix}\right]\begin{pmatrix}\hat{\mathsf{Y}}_1\\ \hat{\mathsf{Y}}_2\end{pmatrix} = \begin{pmatrix}\mathsf{C}_1^*\\ 0\end{pmatrix} \tag{2.45a}$$

The finite number of flow fields $\hat{\mathsf{Y}}_1(\omega_i)$ at discrete pulsations ω_i can be obtained by inverting this linear system and are stacked as columns of the matrix $\mathbf{Y}_1$. Analogously, The observability Gramian G_o can be

thought of as the spatial correlation matrix for the harmonic responses of the adjoint system to harmonic forcings at the sensor location.

2.C Eigenspectrum of POD models

This appendix is devoted to showing an interesting observation on the eigenspectrum of POD models. We have represented in figure 2.18(a) a larger part of the eigenspectrum corresponding to models of size $r = 120$, $r = 140$ and $r = 200$ together with the spectrum of global eigenmodes. As detailed in section 2.4, the unstable eigenvalues of POD models tend to those of the full system. Here we take a closer look at the stable part of the eigenspectrum ($\alpha < 0$). The global eigenspectrum, depicted by triangles, is characterized by eigenmodes organized in branches. We notice that each POD eigenspectrum displays a branch of modes that tends to the most unstable branch of the global eigenspectrum as the model size r is increased.
Figure 2.18(b) represents the same global eigenspectrum focused on the region $-0.8 < \alpha < 0.2$ and that of a POD model of size 200. It is strikingly noticed that the first unstable POD eigenvalues superimpose onto the first global eigenmodes. Note that BPOD models do not yield the same observation (not shown here). This result suggests that the POD procedure somehow rebuilds the original spectrum of the original full system.

Interlude

We showed in chapter 2 how to use a set of frequential snapshots to construct Reduced Order Models (ROMs) that efficiently capture the input-output dynamics. In the case of the backward-facing step flow, we successfully modelled the dynamics from an actuator, located in the vicinity of the flow separation, to a sensor located near reattachment. Since we considered only one actuator (input) and only one sensor (output), our problem may be referred to as a Single Input Single Output (SISO) configuration. We have represented in figure 2.19 a sketch of the SISO dynamics investigated so far. The input $\mathsf{B}c(t)$ acts on the flow dynamics which yield the measured quantity $m(t)$ by means of the sensing C. A critical point to understand is that the ROMs designed in our SISO configuration are not meant to capture the dynamics from another input B to another output C. In particular, these ROMS do not model the flow dynamics, i.e. they neither approximate the dynamical operator A nor the time propagator which may be formally written as $e^{\mathsf{A}t}$. Instead, as exposed in chapter 2, they are shown to recover the original impulse response $\mathsf{C}e^{\mathsf{A}t}\mathsf{B}$. Generally speaking, a ROM which captures the input-output dynamics does not necessarily accurately model the flow states driven by the input. For instance the Balanced models presented in the previous chapter are not able to do so. On the other hand, the ROMs based on the most controllable modes (referred to as POD models) have been observed to recover the original flow response $e^{\mathsf{A}t}\mathsf{B}$, namely the entire flow dynamics excited by the input.

Now, in more realistic situations, a ROM may be required to model the dynamics from Multiple Inputs to Multiple Outputs (which is commonly referred to as MIMO configurations). The snapshot method presented in the previous chapter can easily be extended to a case with a moderate number of inputs and outputs, see Rowley (2005). To do so, the controllable or balanced modes are computed from additional sets of snapshots (one additional set for each new input and output). As a result, this procedure quickly becomes not tractable for large systems as soon as the number of inputs and/or outputs becomes large. Nevertheless, it is of pivotal importance for practical applications to design ROMs that capture a large number of inputs and outputs. For instance, we have seen that the backward-facing step flow is a noise amplifier whose dynamics are driven by external disturbances. As a result, one can expect the flow to be driven by much more than one input. Mostly, realistic external disturbances act as an unknown noise so that one cannot predict a priori which inputs are to be modelled by the desired ROM. Alternatively, as soon as considering a $3D$ set-up, namely a $3D$ base flow, many applications would require much more than one sensor. Let us for instance consider the design of a closed-loop flow controller in the case of

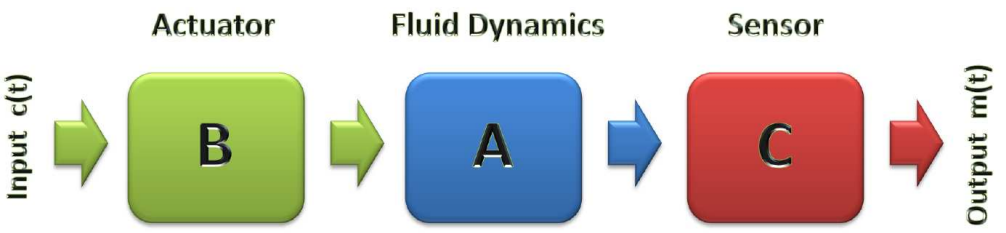


Figure 2.19: Sketch of the Single Input Single Output (SISO) dynamics investigated in chapter 2.

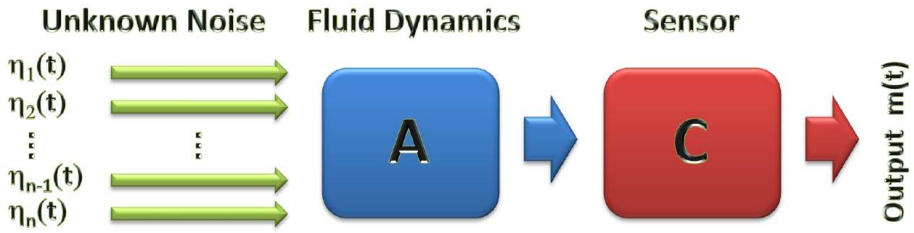


Figure 2.20: Sketch of the Multiple Inputs Single Output (MISO) dynamics investigated in chapter 3.

our backward facing step flow. In a $3D$ set-up with lateral walls, we would rather expect to require one or several arrays of sensors (namely outputs) in the transverse direction to design an accurate low-order controller.

In this context, the next chapter deals with a case with a large number of inputs. The backward-facing step flow, along with the same single sensor, are considered. We suppose that the flow is forced by an unknown noise and we look for ROMs which capture the exact measurement from the sensor. To do so, we choose to consider all the degrees of freedom of the state-space as the inputs. Then, two projection bases are considered and investigated to design the ROMs. The first set consists of the most observable modes which are the flow structures leading to a maximum energy contribution to the sensor measurement. In other words, they are the states to which the sensor is the most sensitive. The second basis consists of balanced modes which are computed by an approximate Balanced Truncation procedure. This approximate procedure will be shown to be nearly optimal to capture the input-output dynamics. Owing to an analogous procedure introduced by Rowley (2005) and called "output projection", ours will be referred to as "input projection". The next chapter gives a particular emphasis to the numerical techniques which allow to compute these balanced modes. Furthermore, a performance comparison between the ROMs resulting from the different projection bases is assessed. Analogously to figure 2.19, we have presented a sketch of the input-output dynamics of interest in figure 2.20. In this case the input $\mathsf{B}c(t)$ is replaced by the unknown noise $\eta(t)$ which has a very large number of degrees of freedom. Such a set-up may accordingly be referred to as a Multiple Inputs Single Output (MISO) configuration. The results exposed in this chapter have also been reported in an article submitted to *Physics of Fluids*.

Chapter 3

Model reduction of the dynamics between uncertain forcing and a sensor

This chapter addresses the model reduction of high-order linear systems within the framework of the incompressible Navier-Stokes equations. We look for reduced-order models that capture the response of some specific sensor whatever the initial flow condition and in the presence of any time-dependent external forcing. Namely, this work deals with the accurate modeling of the input-output dynamics of a fluid system when considering each degree of freedom of the system as an input, and the given measurement as the output. In the case of complex or realistic flows, the number of inputs is too large to apply the standard balanced truncation procedure. To alleviate this problem, we introduce a method called input projection. Input projection is shown to be analogous to the output projection procedure introduced by Rowley (2005). To illustrate the model reduction, we consider the dynamics of a globally stable flow over a rounded backward-facing step. Reduced-order models are obtained by projecting the full original system onto: (i) the basis of the leading balanced modes computed from the input-projected systems and (ii) the most observable modes. The balanced models are observed to accurately capture the transient growths along the separated flow whatever the input while outperforming the models based on the most observable modes.

3.1 Introduction

A variety of open flows, such as boundary layers, mixing layers or separated flows, are subject to convective instabilities which amplify upstream low-level noise. In fact, even if globally stable, these flows may sustain large flow unsteadiness characterized by broadband frequencies due to the presence of permanent upstream noise, see Kaiktsis *et al.* (1996). These flows are often called selective noise-amplifiers and may be highly sensitive to small upstream perturbations (residual turbulence, noise, surface roughness). In particular, the dynamics of such flows strongly depend on the characteristics of the upstream noise, which are unknown in practice. As a result, modeling the dynamics of amplifier flows remains a challenging task which, however, is crucial in many industrial applications. Typical examples are: (i) the closed-loop control of the laminar-turbulent transition in boundary layer flows (Bagheri *et al.* (2009*b*); Bagheri & Henningson (2011); Semeraro *et al.* (2011)) and (ii) the design of near optimal state estimators in meteorology (Farrell & Ioannou (2001*b*,*a*)). In both cases, due to the large size of the fluid systems, reduced-order models (ROMs) of the flow dynamics which capture the noise are required.

In order to capture such an unspecified noise, one solution consists of building ROMs which capture all the possible inputs. As a prototype example, we consider the flow over a rounded backward-facing step shown in figure 3.1. This flow is assumed to be driven by an uncertain forcing at the upstream. Our objective is then to design low-order models that capture the linear dynamics on a single specific sensor, for all possible forcing configurations. Such a ROM may be very useful in the context of closed-loop

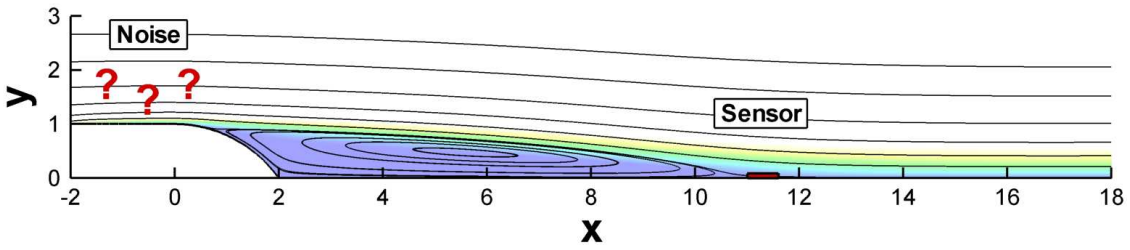


Figure 3.1: Backward-facing step flow investigated here. The steady-state base flow at $Re = 600$ is depicted by its streamlines and longitudinal velocity. The upstream, downstream and upper boundaries are respectively located at $x = -20$, $x = 100$ and $y = 20$. The exact position of the sensor is also displayed.

control. For example, Bagheri *et al.* (2009*b*) assumed that the upstream noise was known in the design of their ROM. This noise was taken as a single spatial structure continuously forced in time by a white noise signal. The associated LQG compensator will therefore only work for this specific spatial structure and location of the noise. Now, in real experiments, the noise is unknown and a model capturing all possible inputs would overcome this limitation.

A model reduction technique called Balanced Truncation is particularly suited for stable linear input-output systems, see Moore (1981). It consists of finding a basis of equally controllable and observable modes ranked according to these properties. ROMs are obtained by the projection of the original system onto the leading balanced modes. Balanced Truncation is said to be quasi-optimal to capture the dynamics from the inputs to the outputs since it offers theoretical bounds on the transfer function error which are close to the lower bound achievable by any reduced-order model, see Dullerud & Paganini (2000). This useful property makes it a very efficient and popular technique which brought the use of modern optimal flow control tools (Burl (1999); Bewley (2001); Zhou *et al.* (2002); Kim & Bewley (2007); Sipp *et al.* (2010)) within reach of current computational technology. Previous works on Balanced Truncation comprise the extension to non linear problems by Scherpen (1993) and Lall *et al.* (2002), the introduction of an algorithm by Laub *et al.* (1987) for the accurate and optimal computation of the balanced basis and also the extension to unstable linear systems proposed by Zhou *et al.* (1999).

Technically, the balanced basis can be computed as the eigenvectors of the product of two operators, namely the so-called controllability and observability Gramians, see Kim (1998); Antoulas (2005). For systems of moderate size, with up to $O(10^3)$ degrees of freedom, the balanced modes can be computed directly by computing explicitly the Gramians (Farrell & Ioannou (2001*b*,*a*); Joshi *et al.* (1997); Högberg *et al.* (2003)). However, an approximate procedure is required for larger systems since this computation quickly becomes untractable. A recent extension proposed by Willcox & Peraire (2002) and Rowley (2005) has overcome this difficulty by approximating the balanced modes without computing the Gramians, by using a snapshot method. This technique, called Balanced Proper Orthogonal Decomposition (BPOD), has been performed on both stable (Ilak & Rowley (2008); Bagheri *et al.* (2009*b*); Dergham *et al.* (2011)) and unstable flow configurations, see Ahuja & Rowley (2008); Barbagallo *et al.* (2009); Ahuja & Rowley (2010).

Balanced Truncation designs efficient low-order models which capture the dynamics from specific inputs to specific outputs. However, it may be completely ineffective to deal with the inputs and the outputs that have not been explicitly taken into account. In particular, it is not meant to model the true flow perturbation triggered from all the possible initial conditions unless both the input and the output spaces are chosen as the full state space. For large systems, using the method of snapshots requires the computation of one direct and one adjoint simulation per input and output. Consequently, choosing a large number of inputs and outputs quickly becomes unaffordable with the standard snapshot method. Though, Rowley (2005) has introduced the technique called output projection to approximate the balanced modes for systems having a very large number of outputs. Output projection has been applied for

the model reduction of a channel flow (Ilak & Rowley (2006, 2008)) and to perform closed-loop control on the flow past an inclined flat plate (Ahuja & Rowley (2008, 2010)) by choosing the total kinetic energy of perturbations as the control objective to minimize.
Alternatively, another model reduction procedure is worth mentioning. The eigenvectors of the controllability and observability Gramians constitute two sets of orthogonal modes that may be considered to perform model reduction. The leading eigenmodes of the two Gramians account for the most controllable and observable states respectively, see Bagheri *et al.* (2009*b*). These modes have already been used to build reduced-order models for fluid systems, see Rowley (2005); Ilak & Rowley (2008); Barbagallo *et al.* (2009); Bagheri *et al.* (2009*b*). In these cases, they are referred to as Proper Orthogonal Decomposition (POD) models since they are computed from the classical method of snapshots introduced by Sirovich (1987). By construction, controllable modes optimally capture the energy triggered by the input while observable modes are the flow structures leading to maximum output energy. In spite of these properties, the resulting ROMs are known to be sub-optimal for capturing the input-output dynamics, and may even be ineffective, see Rowley *et al.* (2004).
In this chapter, we are concerned with Balanced Truncation of stable linear input-output systems having a very large number of inputs. Particular attention is given to the case where the input is the full state, which stands for finding low-order models that capture the dynamics from any initial condition or any forcing distribution to some given output. The procedure developed in this chapter consists of an approximate Balanced Truncation and an extension of the output projection procedure introduced by Rowley (2005). In particular, the resulting ROMs are meant to quasi-optimally capture the input-output dynamics. Owing to its technical analogy with the output projection procedure, we call the present method input projection. The procedure is illustrated on the incompressible flow over a two-dimensional backward-facing step, which only accounts for a prototype of noise-amplifier flow. The formalism introduced here do not rely on physical insight into this particular flow configuration and can be applied to any other stable open flow configuration. The guideline of this chapter is very similar to that adopted by Ilak & Rowley (2008) where the performance of the output projection procedure is assessed on a channel flow. Notably, both the ROMs based on the most observable modes and the leading balanced modes are investigated. Their performance are evaluated and also compared. The goal of this chapter is twofold: (i) show the ability of input projection to accurately model the "full-input to single output" transfer function of a large fluid system and (ii) to illustrate its quasi-optimality.
The rest of the chapter proceeds along the following outline: in section 3.2 we start by introducing the model reduction procedures within the framework of the incompressible Navier-Stokes equations. These techniques are then applied to the two-dimensional flow over a backward-facing step, which is described in section 3.3. Results are presented in section 3.4 where the performance of the ROMs is assessed. A critical assessment of the input projection technique is provided in section 3.5 and concluding remarks are presented in section 3.6.

3.2 Model reduction methodology

3.2.1 Problem formulation

3.2.1.1 Governing equations

We consider the incompressible Navier-Stokes equations excited by a small noise (the input) together with a sensing (the output). The equations governing dynamics of the velocity $\boldsymbol{u}$ and pressure p fields are given by

$$\begin{cases} \partial_t \boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla p + Re^{-1}\nabla^2\boldsymbol{u} + \varepsilon\eta(t) \\ \nabla\cdot\boldsymbol{u} = 0 \\ m(t) = \mathcal{C}\boldsymbol{u} \end{cases} \tag{3.1}$$

where Re denotes the Reynolds number and $\eta(t)$ stands for a field modeling the noise. The parameter

ε points out that the noise is small compared to the other terms of the equation. Thus, $\varepsilon\eta(t)$ acts as a small time dependent forcing on the momentum equation. As in a practical case, we consider that the spatial and temporal distributions of the noise $\eta(t)$ are unknown. Since the velocity field has a divergence-free constraint, it can easily be shown that only the divergence-free part of $\eta(t)$ will have an effect on the system, so that we can suppose that $\nabla\cdot\eta(t) = 0$. This forcing term $\eta(t)$ is considered as the input of our problem in the following.
The measure denoted by $m(t)$ represents a quantity extracted from the flow by some sensors. Mathematically, it is expressed as the result of a measure operator $\mathcal{C}$ applied to the velocity field $\boldsymbol{u}$. For simplicity, we will consider in this work a single sensor, so that $m(t)$ is a scalar. This measurement is also referred to as the output of our system. In this context, we wish to design ROMs capable of capturing the linear input-output dynamics of this system. In other words, we wish to model accurately these dynamics whatever the noise distribution.
First, we linearize the dynamics given by the equations (3.1) at the order ε about a base flow. This base flow is chosen here to be the solution of the associated steady Navier-Stokes equations with $\varepsilon = 0$. The velocity of the base flow is denoted by $\boldsymbol{U}$ in the following and we adopt, for simplicity, the notation $(\boldsymbol{u}, p)$ for the perturbation field and $m(t)$ for the measure of the perturbation. The resulting set of equations reads

$$\begin{cases} \partial_t \boldsymbol{u} + \boldsymbol{U}\cdot\nabla\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{U} = -\nabla p + Re^{-1}\nabla^2\boldsymbol{u} + \eta(t) \\ \nabla\cdot\boldsymbol{u} = 0 \\ m(t) = \mathcal{C}\boldsymbol{u} \end{cases} \tag{3.2}$$

which accounts for the linear dynamics of the perturbations about the base flow. Within a numerical approach, equations (3.2) are discretized on a mesh. It can be shown (see appendix 3.A) that we can write the resulting discretized equations in the following matrix form

$$\frac{d\mathsf{X}}{dt} = \mathsf{A}\mathsf{X} + \eta(t) \tag{3.3a}$$
$$m(t) = \mathsf{C}\mathsf{X} \tag{3.3b}$$

where X and C denote the vectors resulting from the discretization of $\boldsymbol{u}$ and $\mathcal{C}$. Additionally, we assumed that $\eta(t)$ and $m(t)$ maintain the same notation once discretized and the matrix A represents the linearized Navier-Stokes operator which is supposed to be stable in the following. Equations (3.3) constitute the stable linear input-output state-space system considered in the following. The size of this system, denoted by n, is also its number of degrees of freedom. Thus, the matrix A is of size $n \times n$, the states X and $\eta(t)$ are of size $n \times 1$ and C is of size $1 \times n$.

3.2.1.2 Definition of the norms

Mathematically, there are two relevant and equivalent quantities for assessing the input-output behavior of system 3.3: the impulse response $\mathsf{g}(t) = \mathsf{C}e^{\mathsf{A}t}$ and the transfer function $\mathsf{G}(\omega) = \mathsf{C}(i\omega\mathsf{I} - \mathsf{A})^{-1}$. To quantify performance of the ROMs, we compute the error on the transfer function, which requires the definition of some norms.

First, let us introduce an inner product, denoted by $\langle\,\rangle$, on the state space. This inner product is chosen as the standard energy inner product. It is defined for two states Z_1 and Z_2, corresponding to the velocities $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, by

$$\langle \mathsf{Z}_1, \mathsf{Z}_2 \rangle = \int_\Omega \boldsymbol{u}_1.\boldsymbol{u}_2 \, d\Omega \tag{3.4}$$

Where Ω is the fluid volume. We also suppose that the input space, which is equal to the state space, is endowed with the same inner product. It is then possible to define the adjoint operator $\mathsf{A}^\dagger$ by

$$\langle \mathsf{Z}_1, \mathsf{A}\mathsf{Z}_2 \rangle = \langle \mathsf{A}^\dagger \mathsf{Z}_1, \mathsf{Z}_2 \rangle \tag{3.5}$$

for all possible states Z_1 and Z_2. Furthermore, we define the adjoint of the state X and of the operator C by

$$\langle \mathsf{X}, \mathsf{Z} \rangle = \mathsf{X}^\dagger \mathsf{Z} \quad \text{and} \quad \mathsf{C}\mathsf{Z} = \langle \mathsf{C}^\dagger, \mathsf{Z} \rangle \tag{3.6}$$

for all possible states Z. Note that $\mathsf{g}(t)$ and $\mathsf{G}(\omega)$ are vectors of size $1 \times n$; their adjoint are defined similarly to that of C. It is straightforward to show that $\mathsf{G}^\dagger(\omega) = (-i\omega \mathsf{I} - \mathsf{A}^\dagger)^{-1}\mathsf{C}^\dagger$ and $\mathsf{g}^\dagger(t) = e^{\mathsf{A}^\dagger t}\mathsf{C}^\dagger$, which are both of size $n \times 1$.

The norm associated with the inner product $\langle\,\rangle$ will be denoted by $||\,||$ in the following. It is defined for a state X by $||\mathsf{X}|| = \sqrt{\langle \mathsf{X}, \mathsf{X} \rangle}$. From this inner product, it is possible to define two standard norms to quantify the transfer function. The 2-norm of the transfer function G is defined by either a frequency or a time integral:

$$||\mathsf{G}||_2 = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{+\infty} ||\mathsf{G}^\dagger(\omega)||^2 d\omega} = \sqrt{\int_0^{+\infty} ||\mathsf{g}^\dagger(t)||^2 dt} \tag{3.7}$$

Furthermore, we also consider the ∞-norm of the transfer function which is defined by

$$||\mathsf{G}||_\infty = \max_\omega ||\mathsf{G}^\dagger(\omega)|| \tag{3.8}$$

Note that both norms are satisfactorily defined since we assumed that A is stable. In addition, the two norms $||\,||_2$ and $||\,||_\infty$ should not be confused with the norm $||\,||$ that measures a state.

3.2.2 Balanced Truncation

Balanced Truncation is a well-known model reduction technique used for stable linear input-output systems of the form (3.3). It relies on the concept of controllability and observability of the flow states. Controllability quantifies how easy a state can be reached from any other state while observability quantifies the amount of measure triggered by a given flow state, see Moore (1981). The key idea of Balanced Truncation is to find a basis of equally controllable and observable modes and project the original equations onto the set of the most controllable/observable modes. This basis, which is called the balanced basis, may be computed as the eigenvectors of the product of the controllability and observability Gramians defined respectively by

$$\mathsf{G}_c = \int_0^{+\infty} e^{\mathsf{A}t} e^{\mathsf{A}^\dagger t} dt \qquad \mathsf{G}_o = \int_0^{+\infty} e^{\mathsf{A}^\dagger t} \mathsf{C}^\dagger \mathsf{C} e^{\mathsf{A}t} dt \tag{3.9}$$

For large systems, a procedure known as Balanced POD (BPOD) yields a snapshot-based approach to approximate Balanced Truncation, see Rowley (2005). It relies on the computation of the matrices $\mathsf{X}(t) = e^{\mathsf{A}t}$ and $\mathsf{Y}(t) = e^{\mathsf{A}^\dagger t}\mathsf{C}^\dagger$ for a discrete set of times in order to stack them (with appropriate quadrature weights) as columns of a matrix $\mathbf{X}$ and $\mathbf{Y}$ such that the Gramians may be factored as

$$\mathsf{G}_c \approx \mathbf{X}\mathbf{X}^\dagger \qquad \mathsf{G}_o \approx \mathbf{Y}\mathbf{Y}^\dagger \tag{3.10}$$

Next, one can compute the balanced basis from the singular value decomposition $\mathbf{Y}^\dagger\mathbf{X} = \mathsf{M}\Sigma\mathsf{N}^*$, where * denotes the transconjugate. The balanced basis T and its inverse S are found by

$$\mathsf{T} = \mathbf{X}\mathsf{N}\Sigma^{-1/2} \qquad \mathsf{S} = \mathbf{Y}\mathsf{M}\Sigma^{-1/2} \tag{3.11}$$

where the balanced modes are the columns of the matrix T while its biorthogonal basis constitutes the columns of S. Note that their orthogonality relation reads $\mathsf{S}^\dagger\mathsf{T} = \mathsf{I}$. The entries of the diagonal matrix Σ

are known as the Hankel singular values (HSVs) and refer to the equal controllability and observability of the associated balanced modes.

Using the snapshot method introduced by Rowley (2005), the computation of $\mathsf{Y}(t) = e^{\mathsf{A}^\dagger t}\mathsf{C}^\dagger$ requires one adjoint simulation since there is only one output. However, the computation of $\mathsf{X}(t) = e^{\mathsf{A}t}$ is not tractable since it would require n direct simulations, one for each degree of freedom of the system. The computation of the exact balanced basis associated with system 3.3 is thus not possible in our case. Yet, we alleviate this problem by considering instead: (i) the observable modes, which are already known in the literature, and (ii) the balanced modes computed by using the technique introduced here and called "input projection".

3.2.3 Observable modes

Firstly, we consider an orthogonal basis of modes ranked according to their observability. These modes are obtained as the leading eigenvectors of the observability Gramian G_o. The resulting flow structures are by definition the most observable states and are ranked according to their contribution to the output energy, see Bagheri *et al.* (2009*b*).
For large systems, the leading observable modes can be computed by using the snapshot method introduced by Rowley (2005). The first step is to compute the flow states $\{\mathsf{Y}(t) = e^{\mathsf{A}^\dagger t}\mathsf{C}^\dagger\,;\, t>0\}$ with an adjoint simulation to build the matrix $\mathbf{Y}$. Next, we compute the singular value decomposition $\mathbf{Y}^\dagger\mathbf{Y} = \mathsf{L}\Lambda\mathsf{L}^*$. The leading observable modes are then obtained as the columns of the matrix R given by

$$\mathsf{R} = \mathbf{Y}\mathsf{L}\Lambda^{-1/2} \tag{3.12}$$

Note that the orthogonality of these modes reads $\mathsf{R}^\dagger\mathsf{R} = \mathsf{I}$. It should also be noticed that the observable modes reduce to the POD modes of the dataset $\{\mathsf{Y}(t)\,;\, t>0\}$ computed with the energy inner product. The diagonal matrix Λ yields the leading eigenvalues of G_o which account for the observability of the corresponding eigenvectors.
Reduced-order models can then be obtained by projecting the full system (3.3) onto the most observable modes. The choice of such ROMs is intuitively motivated insofar as it is based on the modes having the highest contribution to the sensor energy. However, contrary to balanced models, these ROMs are sub-optimal (and may be ineffective) for capturing the true input-output dynamics.

3.2.4 Input projection

The second approach considered here is the method called input projection. The idea is to project the input $\eta(t)$ on a low-dimensional subspace while optimally preserving, in the 2-norm sense, the original transfer function. If we introduce an orthogonal projection P_s on a s-dimensional subspace of the input space, the new input-output system reduces to

$$\frac{d\mathsf{X}}{dt} = \mathsf{A}\mathsf{X} + \mathsf{P}_s\eta(t) \tag{3.13a}$$
$$\tilde{m}(t) = \mathsf{C}\mathsf{X} \tag{3.13b}$$

where $\tilde{m}(t)$ is the output of the input-projected system. We look for the projection P_s that minimizes the 2-norm error between the original impulse response $\mathsf{G}(\omega)$ and that of the input-projected system, which is given by $\tilde{\mathsf{G}}(\omega) = \mathsf{G}(\omega)\mathsf{P}_s$. This error can expressed by

$$||\mathsf{G} - \mathsf{G}\mathsf{P}_s||_2 = \sqrt{\int_0^{+\infty} ||\mathsf{g}^\dagger(t) - \mathsf{P}_s\mathsf{g}^\dagger(t)||^2 dt} \tag{3.14}$$

since $\mathsf{P}_s^\dagger = \mathsf{P}_s$. Furthermore, noting that $\mathsf{g}^\dagger(t) = \mathsf{Y}(t)$, we infer that the projection that minimizes this error stands for the projection onto the first s POD modes of the dataset $\{\mathsf{Y}(t)\,;\, t>0\}$. In other words, P_s reduces to the orthogonal projection onto the most observable modes which were introduced in section 3.2.3. Consequently, P_s can be written as

$$\mathsf{P}_s = \mathsf{R}\mathsf{R}^\dagger \tag{3.15}$$

where R is the matrix of size $n \times s$ whose columns are the first s observable modes. Thus P_s is a matrix of size $n \times n$ and of rank s.

By construction, only the s most important degrees of freedom of the input $\eta(t)$ have an effect on the output $\tilde{m}(t)$. By considering $\tilde{\eta}(t) = \mathsf{R}^\dagger\eta(t)$ as a new input of size $s \times 1$, the system (3.13) can then be written as

$$\frac{d\mathsf{X}}{dt} = \mathsf{A}\mathsf{X} + \mathsf{R}\tilde{\eta}(t) \tag{3.16a}$$
$$\tilde{m}(t) = \mathsf{C}\mathsf{X} \tag{3.16b}$$

This new input-output system possess the same controllability and observability Gramians as those of system (3.13). It is referred to as the s input-projected system hereafter. Interestingly, the computation of the balanced modes of this new system becomes affordable since it only requires s direct simulations, one for each column of R to get $\mathsf{X}(t) = e^{\mathsf{A}t}\mathsf{R}$. The balanced basis of system (3.16) will be denoted by T and S in the rest of the chapter. It should also be emphasized that these balanced modes depend on the parameter s and that they approximate the balanced modes of the original system (3.3).

Note that the idea to project the input state is analogous to the method called output projection introduced by Rowley (2005) and performed by Ilak & Rowley (2008) and Ahuja & Rowley (2010). In their case, the output of the system is the entire state space so that they optimally project the output on a low-dimensional subspace. In fact, the input projection introduced here can be interpreted as the output projection on an adjoint system. This observation is presented in more detail in appendix 3.C.

3.2.5 Model reduction

Reduced-order models are obtained by projecting the input-projected equations (3.13) onto the basis of its leading balanced modes. Keeping the first r balanced modes (columns of T and S) the dynamics of the reduced state X_r, which is of size $r \times 1$, are given by:

$$\frac{d\mathsf{X}_r}{dt} = \mathsf{A}_r\mathsf{X}_r + \mathsf{B}_r\eta(t) \tag{3.17a}$$
$$m_r(t) = \mathsf{C}_r\mathsf{X}_r \tag{3.17b}$$

where $m_r(t)$ is the output of the reduced-order model and the above-introduced matrices are defined by:

$$\mathsf{X}_r = \mathsf{S}^\dagger\mathsf{X} \qquad \mathsf{A}_r = \mathsf{S}^\dagger\mathsf{A}\mathsf{T} \qquad \mathsf{B}_r = \mathsf{S}^\dagger\mathsf{P}_s \qquad \mathsf{C}_r = \mathsf{C}\mathsf{T} \tag{3.18}$$

where A_r is of size $r \times r$, B_r is of size $r \times n$ and C_r is of size $1 \times r$. The corresponding transfer function is then defined by $\tilde{\mathsf{G}}_r(\omega) = \mathsf{C}_r(i\omega\mathsf{I}_r - \mathsf{A}_r)^{-1}\mathsf{B}_r$ where I_r stands for the identity matrix of size r.

ROMs based on the most observable modes are also considered in the following. These models are obtained by using the same procedure, by replacing T and S by R.

3.3 Application to a backward-facing step flow

3.3.1 Flow configuration

3.3.1.1 Geometry

We consider a two-dimensional rounded backward-facing step of circular geometry, see figure 3.1, and an incoming flow from the left. This geometry was originally studied by Duriez (2009). The upstream velocity and the step height are used to make all quantities non-dimensional. The beginning and end of the step are located at $(x = 0, y = 1)$ and $(x = 2, y = 0)$ respectively. The boundary conditions are the following: (i) a uniform and unitary velocity field $(u = 1, v = 0)$ is prescribed at the inlet boundary $x = -20$. (ii) A free-slip condition with zero tangential stress $(\partial_y u = 0, v = 0)$ is prescribed on the boundary $(-20 \leq x \leq -2, y = 1)$. (iii) A laminar boundary layer starts developing on the lower boundary at $x = -2$ as no-slip boundary conditions $(\boldsymbol{u} = 0)$ are imposed on $(-2 \leq x \leq 0, y = 1)$, on the step wall and on the downstream wall $(2 \leq x \leq 100, y = 0)$. (iv) Symmetry boundary conditions are used at the upper boundary $y = 20$ and (v) a free outflow condition $p\boldsymbol{n} - Re^{-1}(\nabla \boldsymbol{u}) \cdot \boldsymbol{n} = 0$ is used at the outlet $x = 100$ ($\boldsymbol{n}$ being the outward normal unitary vector of the boundary).

3.3.1.2 Problem discretization

We used a finite element approach to discretize the problem. The variational formulation of the governing equations is spatially discretized using a mesh composed of triangular elements. The velocity fields are projected onto six-node quadratic triangular elements with quadratic interpolation (P2-elements) whereas the pressure field is discretized using three-node linear triangular elements (P1-elements). The matrices resulting from the projection of the variational formulations onto the basis of finite elements are sparse and are built with the FreeFem++ software (http://www.freefem.org). The mesh considered in this work yields $n \approx 360\,000$ degrees of freedom stemming from about 90 000 triangles.

3.3.1.3 Base flow

The base flow considered here is computed by using a Newton-Raphson method for a Reynolds number $Re = 600$. The solution is depicted on figure 3.1. At this Reynolds number, the flow is observed to be globally stable; the matrix A does not have any unstable eigenvalues. The displacement thickness at $x = 0$ is $\delta^* \approx 0.082$, leading to a Reynolds number based on the displacement thickness of $Re_{\delta^*} \approx 49.2$. The boundary layer separates at $x \approx 0.6$ and reattaches at $x \approx 11$. Even if this base flow is globally stable, it may sustain flow unsteadiness due to the transient energy growth of perturbations. The Kelvin-Helmholtz instability selectively amplifies the upstream low-level noise along the shear layer to drive the flow dynamics downstream. As shown on figure 3.1 by the question-marks, the noise distribution is assumed to be unknown. The sensor is placed downstream in the vicinity of the reattachment point. The measured quantity is chosen as the wall-normal shear stress evaluated at and integrated over a localized region of the wall, namely $m(t) = \int_{x=11}^{x=11.6} \partial_y u \, dx$. This choice of measurement also defines the output operator C.

3.3.2 Computation of the modes

The balanced and observable modes are computed from the singular value decomposition of $\mathbf{Y}^\dagger\mathbf{X}$ and $\mathbf{Y}^\dagger\mathbf{Y}$ respectively. The computation of the matrices $\mathbf{X}$ and $\mathbf{Y}$ is based on the frequential expressions of the Gramians, see Willcox & Peraire (2002), which are derived from Parseval's theorem:

$$\begin{cases} \mathsf{G}_c = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{+\infty} (j\omega\mathsf{I}-\mathsf{A})^{-1}\mathsf{R}\mathsf{R}^\dagger(-j\omega\mathsf{I}-\mathsf{A}^\dagger)^{-1}d\omega \\ \\ \mathsf{G}_o = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{+\infty} (-j\omega\mathsf{I}-\mathsf{A}^\dagger)^{-1}\mathsf{C}^\dagger\mathsf{C}(j\omega\mathsf{I}-\mathsf{A})^{-1}d\omega \end{cases} \tag{3.19}$$

By considering these expressions, we compute the matrices $\hat{\mathsf{X}}(\omega) = (j\omega\mathsf{I}-\mathsf{A})^{-1}\mathsf{R}$ and $\hat{\mathsf{Y}}(\omega) = (-j\omega\mathsf{I}-\mathsf{A}^\dagger)^{-1}\mathsf{C}^\dagger$ for a discrete set of frequencies in order to stack them as the columns of the matrices $\mathbf{X}$ and $\mathbf{Y}$. To this end, we used 399 equidistant frequencies ranging from $\omega = 0$ to $\omega = 4$ and quadrature coefficients corresponding to the 4^{th}-order Simpson method. This choice proved to be sufficient for an accurate computation since, for a larger number of snapshots with a finer spacing or a larger frequency interval, there is no considerable change in the singular values and modes considered here. For each frequency, $\hat{\mathsf{Y}}(\omega)$ is computed by one matrix inversion since $\mathsf{C}^\dagger$ is a vector. Likewise, s inversions are required to get $\hat{\mathsf{X}}(\omega)$; one for each column of R. These matrix inversions are performed through a direct multifrontal sparse LU solver (MUMPS), see Amestoy *et al.* (2001). Further details on the discretized formulation of the Gramians resulting from our choice of inner product are available in appendix 3.B.
It should be mentioned that the temporal expressions of the Gramians are usually adopted to factor the Gramians in (3.10), see Ilak & Rowley (2008); Bagheri *et al.* (2009*b*); Barbagallo *et al.* (2009); Ahuja & Rowley (2010). The present frequential approach has been proved to be equivalent to the temporal one (Zhou *et al.* (1999); Willcox & Peraire (2002)) and it has been applied to the same backward facing step flow configuration by Dergham *et al.* (2011). Whether using the temporal or the frequential approach leads in the end to the same projection basis and to the same ROMs; one may simply choose between the two according to the available computational tools, see Dergham *et al.* (2011).

3.4 Results

3.4.1 Observable modes

Using the procedure outlined in section 3.2.3, we compute the most observable modes by using equation (3.12). The singular values λ_j, ranking the observability of associated states, are represented on figure 3.2(a). The observability of the modes is seen to fall off quite rapidly indicating that only the first modes will have a significant contribution to the measured energy. We have represented in figure 3.2(b,c,d,e) the first, second, third and 12^{th} observable modes, visualized by their streamwise velocity component. The first eigenvalues are in pairs, indicating that the most significant modes are traveling structures that are 90° out of phase, figures 3.2(b,c) illustrate this statement.
As one could expect, the flow structures which yield a maximum energy on the sensor are located upstream. The first modes are spatially located in the vicinity of the separation point while higher modes, as the 12^{th}, have a more extended spatial support. These most observable structures are tilted in the upstream direction, leaning against the shear layer. This result is consistent with other recent works (Bagheri *et al.* (2009*c*,*b*)) where it is interpreted as a way to extract energy from the mean shear by transporting momentum down the velocity gradient by the so-called Orr mechanism, see Butler & Farrell (1992). For the sake of simplicity, the ROMs obtained by projection onto the most observable modes are referred to as "observable models" in the following.

3.4.2 Balanced modes

Next, we compute the balanced modes of the input-projected system (3.16). Several values of the rank s of input-projection, ranging from 1 to 20, have been considered. The balanced basis $(\mathsf{T};\mathsf{S})$ and its associated HSVs σ_j depend on this rank. Figure 3.3 depicts the HSVs of input projected systems of rank 4, 10, 16 and 20. It is observed that increasing the rank of the input projection leads to a convergence of

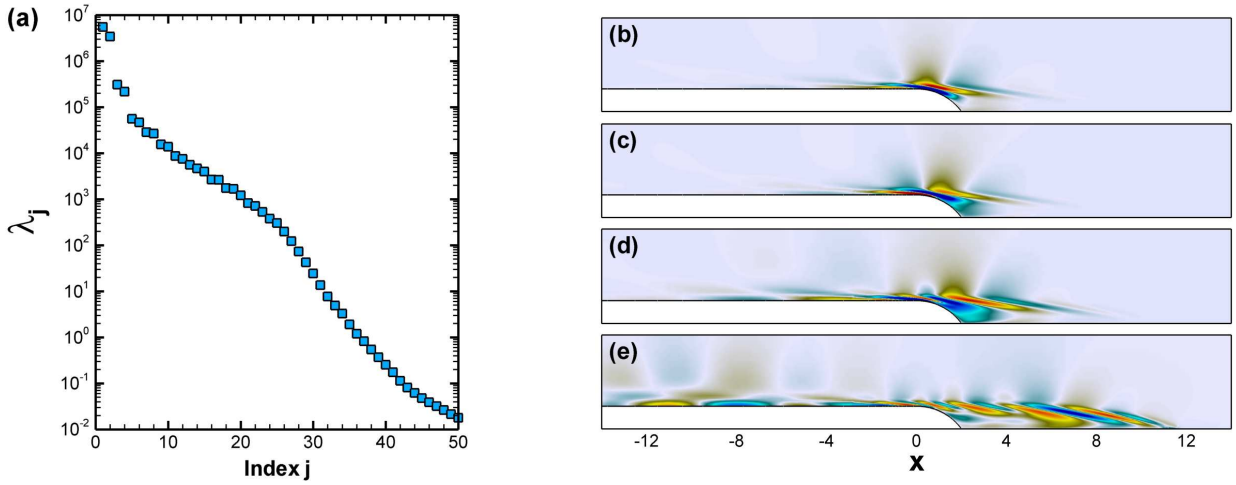


Figure 3.2: (a) First 50 observable eigenvalues λ_j. Figures (b),(c),(d) and (e) represent the longitudinal velocity of the first, second, third and 12^{th} observable modes.

the leading HSVs. In addition, the number of converged HSVs for each input projection is approximately equal to the input-projection rank s, which was also noticed in previous works using the output projection procedure, see Rowley (2005); Ilak & Rowley (2008); Ahuja & Rowley (2010). It should be emphasized that this convergence is also noted for the associated leading balanced modes (not shown here). The first Hankel singular values also come in pairs, indicating that the most significant balanced modes are traveling structures. The HSVs, assessing the controllability/observability of associated states are also observed to fall off quite rapidly, indicating that only the first modes will have a significant contribution to the overall input-output behavior.
We have represented in figures 3.4(a-f) the first, third and 12^{th} balanced modes computed for the case $s = 20$. Figures 3.4(a,c,e) depict the balanced modes (columns of T) while figures 3.4(b,d,f) stand for their associated adjoint (columns of S). We observe that the leading balanced modes appear as wavepackets that are somewhat more spatially extended than the observable modes. We also note a spatial separation between the balanced modes and their adjoint. Essentially, this separation highlights the non-orthogonality of the balanced modes. This point can be interpreted as a consequence of the convective nature of the instabilities along the shear layer in which disturbances grow in amplitude as they are convected in the downstream direction, see Ilak & Rowley (2008); Bagheri *et al.* (2009*b*). As a result, the controllable and observable subspaces are separated in the streamwise direction which implies that the distribution of both the input (the full state) and the output (the sensor) may hardly be captured by an orthogonal projection onto the leading modes of only one subspace. The ROMs resulting from the projection onto the leading balanced modes are called the "balanced models" hereafter.

3.4.3 Performance of the ROMs

The leading observable and balanced modes are used to build reduced order models. Their performance in capturing the original dynamics of the full system (3.3) is assessed by scrutinizing how the input-output transfer function is captured. To that purpose, we compute the relative errors of the reduced transfer functions. Both the 2-norm and the ∞-norm have been considered. Their associated relative errors are denoted by e_2 and e_∞ respectively. They are defined for a model of size r by

$$e_2(r) = \frac{||\mathsf{G} - \tilde{\mathsf{G}}_r||_2}{||\mathsf{G}||_2} \qquad e_\infty(r) = \frac{||\mathsf{G} - \tilde{\mathsf{G}}_r||_\infty}{||\mathsf{G}||_\infty} \tag{3.20}$$

where the full transfer function is computed from the snapshots $\hat{\mathsf{Y}}(\omega)$ by $\mathsf{G}(\omega) = \hat{\mathsf{Y}}^\dagger(\omega)$, see appendix 3.B.3. The evolution of these two errors with the size of the ROMs is presented on figures 3.5(a) and

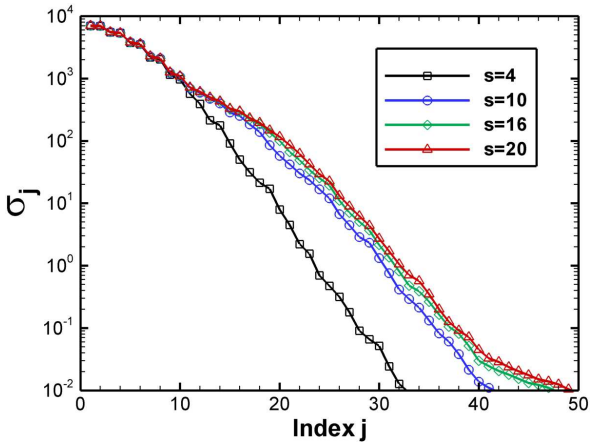


Figure 3.3: First 50 HSVs σ_j corresponding to the input-projected systems $s = 4, 10, 16, 20$.

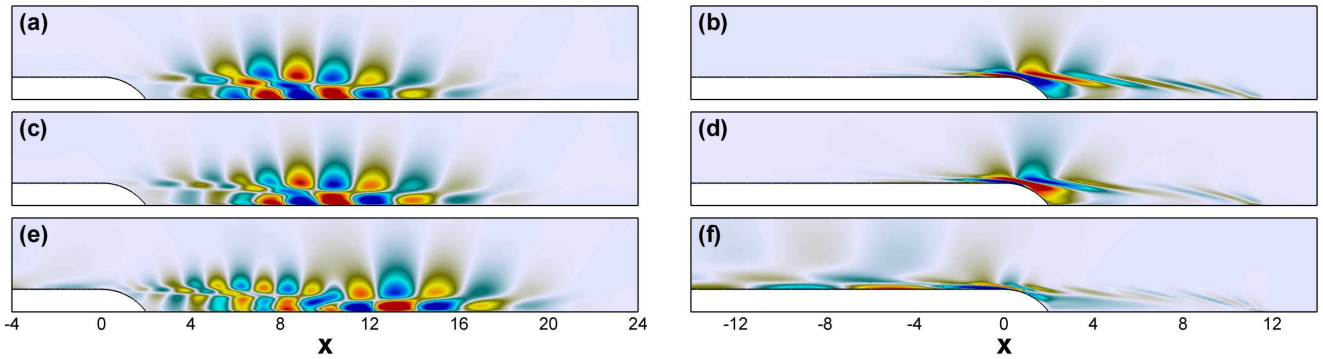


Figure 3.4: Figures (a) and (b) represent the longitudinal velocity of the first balanced mode and its adjoint respectively. (c) and (d) depict analogously the third modes while (e) and (f) represent the 12^{th}. These modes have been computed from an input-projected system with $s = 20$.

3.5(b) respectively. Both yield a good illustration of the effect of the input projection on the performance of the models.

First let us examine the evolution of the 2-norm error of the balanced and observable models on figure 3.5(a). When increasing the size of the models, the relative error is observed to decrease much faster for the balanced models than for the observable ones. Note that some points of the observable models are missing which means that their associated relative error considerably exceeds 100%. Notably, owing to the input projection, a limit of accuracy is reached by the balanced models as the model size r is increased. Indeed, the 2-norm error between the original impulse response $\mathsf{G}(\omega)$ and that of an input-projected system $\tilde{\mathsf{G}}(\omega)$ of rank s, which was given in equation 3.14, can also be expressed as a function of the observable eigenvalues λ_j by

$$||\mathsf{G} - \tilde{\mathsf{G}}||_2^2 = \sum_{j=s+1}^{n} \lambda_j \tag{3.21}$$

The resulting relative errors have been plotted on figure 3.5(a) by dots as a function of the input projection rank s. In other words, for an input projection of rank s, the associated dot yields the limit of accuracy of the corresponding balanced model. This statement is illustrated by dashed lines which account for the error associated with the input projections of rank $s = 4, 10, 16, 20$. Note that the observable eigenvalues can be directly used to compute the limit of accuracy of the balanced models. Interestingly, provided that the limit is not reached, the performance of the balanced models are the

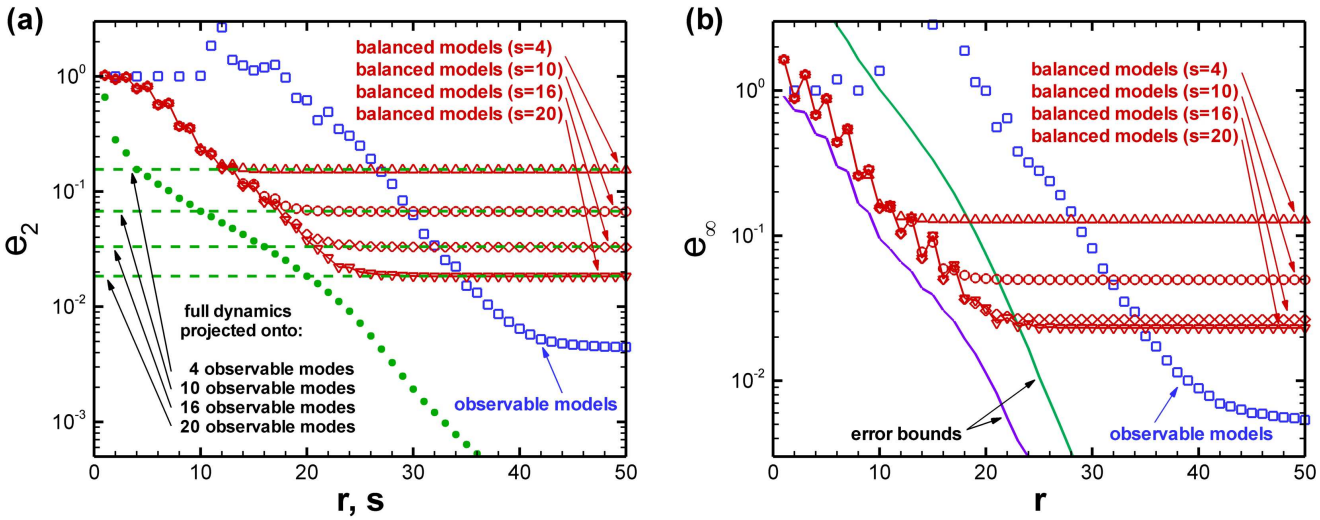


Figure 3.5: Relative error norm of observable and balanced models as a function of the models size r. (a) represents the 2-norm while (b) accounts for the ∞-norm. The 2-norm error limits due to input projection are depicted on (a) as a function of the projection rank s by dots. The dashed lines illustrate these error limits for the ranks $s = 4, 10, 16, 20$. The ∞-norm error bounds, computed on a $s = 20$ input-projected system, are represented on (b) by solid lines.

same whatever the model size r. This observation is consistent with our previous results where we have shown that the first s HSVs and balanced modes are converged when using an input projection of rank s. Consequently, our procedure closely approximates the exact balanced truncation on the original system (without input projection) until the limit of accuracy due to the projection is reached. Increasing the rank of the input projection then delays this limit.

Examining the evolution of the relative ∞-norm error on figure 3.5(b), the same conclusions arise. Yet, it is interesting to consider this norm because of the availability of theoretical bounds on the discrepancy between the approximate and exact transfer functions in the case of the balanced models. The transfer function $\tilde{\mathsf{G}}_r$ of the ROM approximates that of the input projected system $\tilde{\mathsf{G}}$ while guaranteeing (Rowley (2005)) the relation

$$\sigma_{r+1} < ||\tilde{\mathsf{G}} - \tilde{\mathsf{G}}_r||_\infty \leq 2 \sum_{j=r+1}^{n} \sigma_j \tag{3.22}$$

based on the associated HSVs σ_j. The lower bound is valid for any ROM whereas the upper bound is valid for models based on balanced truncation. Assuming the HSVs decrease rapidly, as in the present flow configuration, the upper bound is close to the lower bound achievable by any reduced-order model and the procedure is said to be quasi-optimal. Considering the convergence of the HSVs previously mentioned, it is reasonable to consider these bounds as approximate bounds on $||\mathsf{G} - \tilde{\mathsf{G}}_r||_\infty$. The representation of the associated relative errors are depicted on figure 3.5(b) by solid lines. As these bounds are converged for small ROM sizes (in terms of input projection rank s), the error of the balanced models is observed to lie between these bounds until the limit of accuracy is reached. Note that these error bounds are computed from the HSVs of a $s = 20$ input-projected system. Interestingly, the performance of the balanced models is very close to the lower bound achievable by any ROM.

We conclude that both models succeed in capturing the full input-output behavior of the original system. Observable models are clearly outperformed by balanced models. In addition, the performances of the balanced models are quasi-optimal and subject to theoretical bounds, while those of observable models are not.

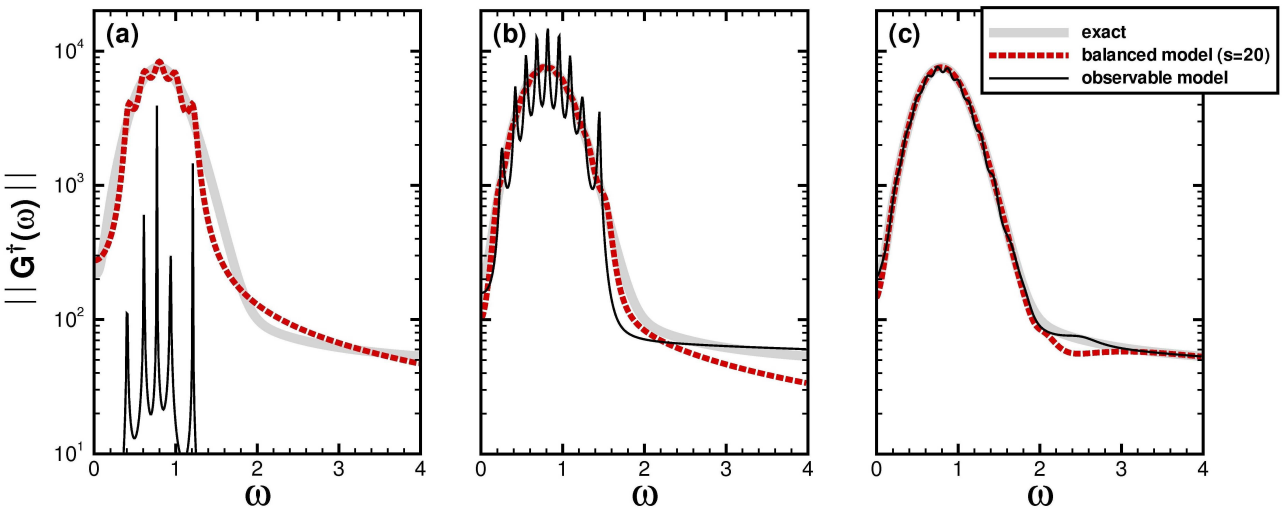


Figure 3.6: Singular value Bode plot of the full system compared to those of balanced and observable models. Figures (a,b,c) account for ROMs of size 10, 20 and 30 respectively. The balanced models are computed on a $s = 20$ input-projected system.

3.4.4 Frequency response

In the case of a multiple input multiple output (MIMO) system, a standard way of quantifying the synthesized frequency response is to compute the maximum singular value of the transfer function. For the full system, it is simply computed by the norm introduced in 3.2.1.2 by $||\mathsf{G}^\dagger(\omega)|| = \sqrt{\mathsf{G}(\omega)\mathsf{G}^\dagger(\omega)}$. For a ROM of size r, it is computed similarly by using the corresponding model transfer function $\tilde{\mathsf{G}}_r(\omega)$ (see appendix 3.B.3).
The representation of this quantity, see figure 3.6, is known as the singular value Bode plot. We have depicted on figures 3.6(a,b,c) the frequency responses of balanced and observable models of size 10, 20 and 30 respectively. Note that the balanced models are based on balanced modes computed for an input projection rank $s = 20$. The exact frequency response clearly highlights a preferred frequency around $\omega = 0.79$ which corresponds to the amplification of perturbations through the shear layer due to the Kelvin-Helmholtz instability. These plots are a clear demonstration of the advantages of balanced models for capturing the dynamics of the system. We see that even a balanced model of size 10 roughly approximates the original response while a 20 sized model yields a very good approximation of the frequential response peak. Meanwhile, observable models are completely unable to recover the most important trend of the input-output behavior for these model sizes. Observable models eventually capture the exact transfer function for models of size higher than approximately 30, see figure 3.6(c).

3.4.5 Modeling the dynamics of some localized actuators

In this last part, we illustrate the ability of the ROMs to capture the dynamics from any input to the sensor. To this end, we arbitrarily choose four actuators in the upstream part of the flow and investigate their impulse response on the sensor. Their spatial distribution, denoted by B_1, B_2, B_3 and B_4, are chosen of Gaussian shape on the vertical momentum component, see figure 3.7. These Gaussians have a width (full width at half maximum) of 0.4, a height of 1 and are centered at $(-0.3; 1.2)$, $(4; 0)$, $(6; 0.7)$ and $(5; 1.5)$ respectively.
Impulse responses of the full system are computed by time-stepper simulations of the linearized Navier-Stokes equations by taking these inputs as initial conditions, namely $\mathsf{X}(t = 0) = B_i$ for $i = [1; 4]$. Impulse responses of the ROMs are also considered, they are directly computed by $\mathsf{C}_r e^{\mathsf{A}_r t}\mathsf{S}^\dagger\mathsf{B}_i$. Both are represented on figures 3.8(a-d) for the four inputs. The balanced and observable models have a size $r = 20$ and the balanced model was designed from an input projection of rank $s = 20$. As expected, we

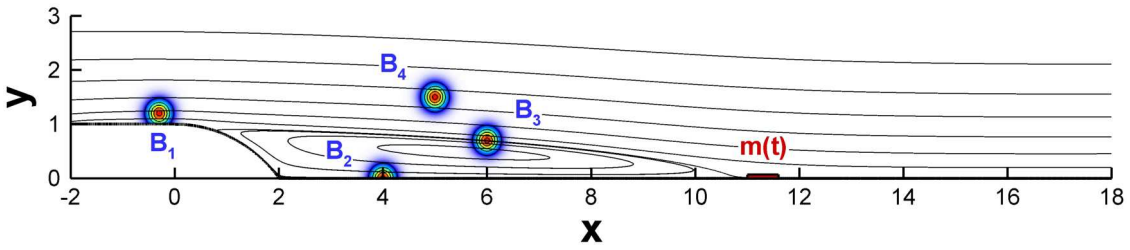


Figure 3.7: Representation of the four chosen inputs by their vertical component. The streamlines of the base flow and the position of the sensor are also displayed.

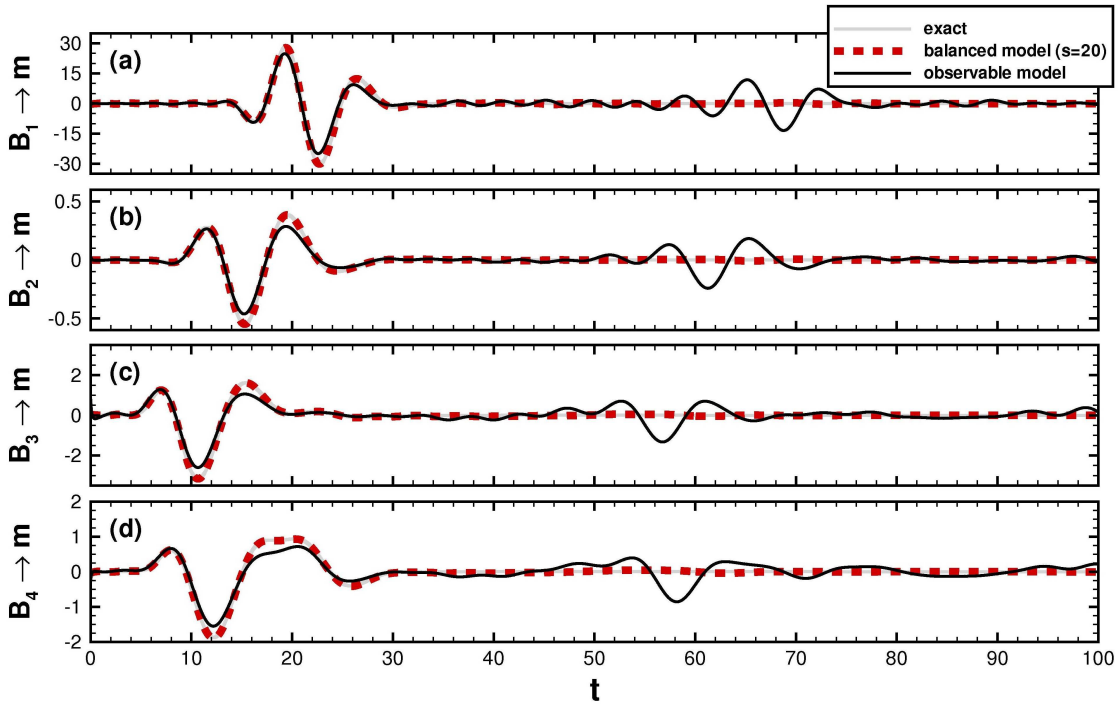


Figure 3.8: Figures (a-d) represent the impulse responses of the four different inputs. The exact solutions come from direct numerical simulations where the number of degrees of freedom is about $n \approx 360\,000$. The dashed and solid lines account for balanced and observable models of size 20 respectively. Note that the balanced model has been computed with an input projection of rank $s = 20$.

observe that all the inputs investigated here are accurately captured by the balanced model. Meanwhile, the observable model is clearly not as efficient at modeling the main pulses and even sustains undesired oscillations at larger times.

These results are consistent with our previous observations shown on figures 3.5 and 3.6 since the balanced model has already a low error for a size $r = 20$ which is not the case for the observable model. These examples illustrate the ability of the ROMs to represent the dynamics of localized inputs that do not intervene in the model reduction process but also the superiority of the balanced models to reach this goal.

3.5 Critical assessment

The results obtained in the present study illustrate the efficiency of the ROMs in capturing the original transfer function. However, the flow configuration considered here is relatively simple; the dynamics are

two-dimensional and there is only one sensor. Thus we raise questions regarding the applicability of input projection in realistic flows.
Let us first describe a situation where the present techniques are readily usable to design a low-order closed-loop controller. A flat plate boundary layer is subject to two-dimensional instabilities (TS waves). We need one upstream sensor for the estimation (just upstream of branch I), one actuator (near branch I), and one sensor for the control objective (near branch II). If the noise triggering the transition is unknown (which is usually the case), then a ROM describing the dynamics from the whole input space and the actuator to the two sensors is required. Hence, the input projection technique applied with two sensors should be effective for this purpose. Note that Bagheri *et al.* (2009*b*) assume that the upstream noise is known, which is unphysical in some sense. Also, in the control objective we only require the measurement at the downstream sensor to be reduced, not the whole perturbation energy. Indeed, this would have required the output projection technique, which is not tractable here since the number of inputs is too high.

As soon as considering three-dimensional perturbations or when increasing the Reynolds number, the flow dynamics may become much more complex. In addition, a larger number of sensors may be required to accurately estimate the flow, see Semeraro *et al.* (2011). In these cases, the use of input projection raises some practical issues.
First, the present techniques still hold for three-dimensional flows but with three-dimensional operators and discretizations. Since the resulting dynamical systems are much bigger, iterative techniques rather than direct inversion techniques are required to compute the snapshots. For example, one could use the temporal expressions of the Gramians and solve the direct and adjoint linearized Navier-Stokes equations to compute the snapshots and approximate the Gramians, as done by Semeraro *et al.* (2011).
Secondly, the number of direct and adjoint simulations may be much larger. Indeed, one adjoint simulation is required for each additional sensor. Furthermore, the flow dynamics may be significantly more complex so that it may not be represented by as few number of degrees of freedom as in the numerical example exposed here. As a result, we expect that more observable modes (i.e. additional direct simulations) would be required for an accurate input projection. In addition, the singular value decompositions of the matrices $\mathbf{Y}^\dagger\mathbf{Y}$ and $\mathbf{Y}^\dagger\mathbf{X}$ may even become challenging. For these reasons, the possibly large number of sensors and of required observable modes clearly constitutes the bottleneck of the input projection procedure.
Lastly, the model reduction of more complex dynamics would result in a slower fall-off of the observable eigenvalues and HSVs. Consequently, more observable and balanced modes would be required to design efficient ROMs. In particular, these ROMs may be too large to be used in the design of real-time closed-loop controllers. However, one has to keep in mind that the input projection technique is quasi-optimal in capturing the dynamics from unknown forcing and no other technique would perform better.
We voluntarily considered in this work all the degrees of freedom as inputs in view of designing ROMs which are robust with respect to uncertainties on the noise. Thus, if the input projection procedure fails to design sufficiently low-order models, then the alternative solution would be to gain insight into the particular flow physics. For instance, one may resort to experimental investigations or to receptivity analyses to gain additional information on the noise that one desires to reproduce.

3.6 Conclusion

In this chapter, we have described how to approximate Balanced Truncation for large linear systems when the number of inputs is large. The procedure has been derived on the two-dimensional Navier-Stokes equations subject to an unknown noise (the inputs) and with a single sensor (the output). We introduced a technique called input projection which consists of projecting the input space onto a low-dimensional subspace while optimally preserving the original transfer function. The optimal projection appears as the orthogonal projection onto the most observable modes, namely the leading eigenvectors of the

observability Gramian. Connections with the already existing output projection technique introduced by Rowley (2005) are also highlighted.

The whole procedure is applied to a stable linear system: the flow over a rounded backward-facing step. A time-dependent forcing term, viewed as a noise, is assumed to act equally on each degree of freedom of the flow state while the sensor is placed near the reattachment point. As a first step, the most observable modes are computed and seen to extend in the upstream part of the flow. Then, the balanced modes of the input-projected systems are computed and observed to converge when increasing the rank of the input projection. Reduced-order models are obtained by projecting the input-projected systems onto the leading balanced and observable modes. These ROMs are evaluated by examining their synthesized frequency response. For both projection bases, the frequency selection process of the original system is recovered as we accurately model the frequency response peak associated with the Kelvin-Helmholtz instability. To quantify this performance, the 2-norm and ∞-norm of the error between the reduced and exact transfer function were computed. We found that the balanced models have an error fall off that decreases quickly until a limit fixed by the rank of the input projection. Mostly, the performances of the balanced models are subject to theoretical bounds and observed to be quasi-optimal. On the other hand, observable models are suboptimal in fulfilling the same objective and do not yield any theoretical bounds on this error. Nevertheless, it is observed to tend to zero when increasing the size of the models. A better efficiency of the balanced models to reach a desired accuracy is clearly identified, which is consistent with the existing literature.

In summary, the balanced truncation of input-projected systems has been shown to yield efficient ROMs which capture the dynamics from any input to a given output. This technique is (i) quasi-optimal in achieving this goal and (ii) subject to theoretical bounds. This contribution may be a promising tool in view of designing reduced models of systems where only a few flow measurements are available while the position of the forcing is unknown. Therefore, it may be a useful addition to the tools of modern flow control theory to design real-time closed-loop controller.

Chapter appendix

3.A Expressing the linearized Navier-Stokes equations as an input-output state-space system

Considering a numerical approach, we proceed by discretizing equations (3.2). The system may be written in a finite-sized matrix form as

$$\begin{pmatrix} \mathsf{Q}_1 & 0 \\ 0 & 0 \end{pmatrix} \frac{d}{dt} \begin{pmatrix} \mathsf{X}_1 \\ \mathsf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathsf{A}_1 & \mathsf{A}_2^* \\ \mathsf{A}_2 & 0 \end{pmatrix} \begin{pmatrix} \mathsf{X}_1 \\ \mathsf{X}_2 \end{pmatrix} + \begin{pmatrix} \mathsf{Q}_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta(t) \\ 0 \end{pmatrix} \tag{3.23a}$$

$$m(t) = \begin{pmatrix} \mathsf{C}_1 & 0 \end{pmatrix} \begin{pmatrix} \mathsf{X}_1 \\ \mathsf{X}_2 \end{pmatrix} \tag{3.23b}$$

where X_1 denotes the velocity field and X_2 represents the corresponding pressure field. The linearized Navier-Stokes operator has been decomposed so that A_1 accounts for the convection and diffusion while A_2 and A_2^* are the parts relative to the incompressibility and pressure effects respectively. Q_1 is the mass matrix, it also stands for the inner product associated with the kinetic energy of perturbations, namely $\mathsf{X}_1^*\mathsf{Q}_1\mathsf{X}_1 = \int_\Omega \boldsymbol{u}^2 \, d\Omega$ where Ω is the fluid volume. $(\mathsf{C}_1, 0)$ denotes the vector of the discretized sensing $\mathcal{C}$ operator.

To reformulate the above equations (3.23) into a standard state-space form, we restrict the system state to its divergence-free velocity field. We first multiply the momentum equation by $\mathsf{A}_2\mathsf{Q}_1^{-1}$, which yields, assuming that $\mathsf{A}_2\dot{\mathsf{X}}_1 = 0$, an expression for the pressure in terms of the velocity field

$$\mathsf{X}_2 = -\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_2^*\right)^{-1}\left[\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_1\right)\mathsf{X}_1 + \mathsf{A}_2\eta(t)\right] \tag{3.24}$$

We can then eliminate the explicit divergence constraint to write the system equations in the desired form

$$\frac{d\mathsf{X}_1}{dt} = \mathsf{P}_1\mathsf{A}_1\mathsf{X}_1 + \mathsf{P}_1\mathsf{Q}_1\eta(t) \tag{3.25a}$$

$$m = \mathsf{C}_1\mathsf{X}_1 \tag{3.25b}$$

where we have introduced the matrix P_1 defined by

$$\mathsf{P}_1\mathsf{Q}_1 = \mathsf{I} - \mathsf{Q}_1^{-1}\mathsf{A}_2^*\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_2^*\right)^{-1}\mathsf{A}_2 \tag{3.26}$$

Notably, $\mathsf{P}_1\mathsf{Q}_1$ reduces to the projection matrix onto the divergence-free space and P_1 is a Hermitian operator ($\mathsf{P}_1^* = \mathsf{P}_1$). Since we assumed that $\eta(t)$ is a divergence-free state, we have $\mathsf{P}_1\mathsf{Q}_1\eta(t) = \eta(t)$. Finally, by defining $\mathsf{A} = \mathsf{P}_1\mathsf{A}_1$ and $\mathsf{C} = \mathsf{C}_1$, we recover equation (3.3):

$$\frac{d\mathsf{X}}{dt} = \mathsf{A}\mathsf{X} + \mathsf{B}\eta(t) \tag{3.27a}$$

$$m(t) = \mathsf{C}\mathsf{X} \tag{3.27b}$$

3.B Introducing the inner products in the discretized equations

Adjoint operators denoted by the superscript † are different in general from the standard transconjugate *. We chose the inner product associated with the kinetic energy for both the states and the inputs, namely $\langle \mathsf{Z}_1, \mathsf{Z}_2 \rangle = \mathsf{Z}_1^*\mathsf{Q}_1\mathsf{Z}_2$ for all states or inputs Z_1 and Z_2. As a result, the mass matrix Q_1 intervenes in the definition of the adjoint operators.

Considering the definition of $\mathsf{A}^\dagger$ and $\mathsf{C}^\dagger$ introduced in section 3.2.2, it is easily found that $\mathsf{A}^\dagger = \mathsf{Q}_1^{-1}\mathsf{A}^*\mathsf{Q}_1$ and $\mathsf{C}^\dagger = \mathsf{Q}_1^{-1}\mathsf{C}^*$. The operator $\mathsf{P}_s = \mathsf{R}\mathsf{R}^\dagger$ has been introduced as the orthogonal projection onto the first s observable modes acting on any flow state or input. Analogously, considering our choice of inner product, it can be expressed as $\mathsf{P}_s = \mathsf{R}\mathsf{R}^*\mathsf{Q}_1$, so that $\mathsf{R}^\dagger = \mathsf{R}^*\mathsf{Q}_1$.

3.B.1 Observability Gramian

Observable modes are the eigenvectors associated with the largest eigenvalues of the observability Gramian:

$$\mathsf{G}_o = \frac{1}{2\pi}\int_{-\infty}^{+\infty}(-j\omega\mathsf{I}-\mathsf{A}^\dagger)^{-1}\mathsf{C}^\dagger\mathsf{C}(j\omega\mathsf{I}-\mathsf{A})^{-1}d\omega \tag{3.28}$$

Replacing the adjoint quantities by their explicit expression, the observability Gramian may be written as

$$\mathsf{G}_o = \mathsf{Q}_1^{-1}\left(\frac{1}{2\pi}\int_{-\infty}^{+\infty}(-j\omega\mathsf{I}-\mathsf{A}^*)^{-1}\mathsf{C}^*\mathsf{C}(j\omega\mathsf{I}-\mathsf{A})^{-1}d\omega\right) \tag{3.29}$$

and the flow states $\hat{\mathsf{Y}}(\omega)$ introduced in section 3.3.2 are then defined by $\mathsf{Q}_1\hat{\mathsf{Y}}(\omega) = (-j\omega\mathsf{I}-\mathsf{A}^*)^{-1}\mathsf{C}^*$. Finally, the procedure outlined in section 3.2 to compute the observable modes can be performed by replacing $\hat{\mathsf{Y}}^\dagger(\omega)$ by $\hat{\mathsf{Y}}^*(\omega)\mathsf{Q}_1$. The observability Gramian then reads $\mathsf{G}_o \approx \mathbf{Y}\mathbf{Y}^*\mathsf{Q}_1$.

3.B.2 Controllability Gramian

Let us consider the controllability Gramian of the input-projected system

$$\mathsf{G}_c = \frac{1}{2\pi}\int_{-\infty}^{+\infty}(j\omega\mathsf{I}-\mathsf{A})^{-1}\mathsf{R}\mathsf{R}^\dagger(-j\omega\mathsf{I}-\mathsf{A}^\dagger)^{-1}d\omega \tag{3.30}$$

Making explicit the adjoint operators, we obtain the new expression:

$$\mathsf{G}_c = \left(\frac{1}{2\pi}\int_{-\infty}^{+\infty}(j\omega\mathsf{I}-\mathsf{A})^{-1}\mathsf{R}\mathsf{R}^*(-j\omega\mathsf{I}-\mathsf{A}^*)^{-1}d\omega\right)\mathsf{Q}_1 \tag{3.31}$$

The input projection procedure previously introduced in section 3.2 can then be derived by using the states $\hat{\mathsf{X}}(\omega) = (j\omega\mathsf{I}-\mathsf{A})^{-1}\mathsf{R}$ and its adjoint $\hat{\mathsf{X}}^\dagger(\omega) = \hat{\mathsf{X}}^*(\omega)\mathsf{Q}_1$. The resulting factored form of the controllability Gramian can be written as $\mathsf{G}_c \approx \mathbf{X}\mathbf{X}^*\mathsf{Q}_1$.

3.B.3 Generalized transfer functions

We see in this appendix how to compute the transfer functions and their associated norms. First let us consider the full transfer function of system (3.3). It is defined by

$$\mathsf{G}(\omega) = \mathsf{C}(i\omega\mathsf{I}-\mathsf{A})^{-1} \tag{3.32}$$

which is a vector of size $1\times n$. $\mathsf{G}(\omega)$ can also be expressed as a function of the flow states $\hat{\mathsf{Y}}(\omega)$ by

$$\mathsf{G}(\omega) = \hat{\mathsf{Y}}^\dagger(\omega) = \hat{\mathsf{Y}}^*(\omega)\mathsf{Q}_1 \tag{3.33}$$

The transfer function can be quantified by its associated standard 2-norm and ∞-norm given by

$$||\mathsf{G}||_2 = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{+\infty}||\mathsf{G}^\dagger(\omega)||d\omega} \qquad ||\mathsf{G}||_\infty = \max_\omega ||\mathsf{G}^\dagger(\omega)|| \tag{3.34}$$

Practically, we can easily compute these norms from the flow states $\hat{\mathsf{Y}}(\omega)$ since

$$||\mathsf{G}^\dagger(\omega)|| = \sqrt{\mathsf{G}(\omega)\mathsf{G}^\dagger(\omega)} = \sqrt{\mathsf{G}(\omega)\mathsf{Q}_1^{-1}\mathsf{G}^*(\omega)} = \sqrt{\hat{\mathsf{Y}}^*(\omega)\mathsf{Q}_1\hat{\mathsf{Y}}(\omega)} \tag{3.35}$$

Now we consider the transfer function of the reduced systems obtained from the input projected system (3.13). It has been defined by

$$\tilde{\mathsf{G}}_r(\omega) = \mathsf{C}_r(i\omega\mathsf{I}_r - \mathsf{A}_r)^{-1}\mathsf{B}_r \tag{3.36}$$

where A_r, B_r and C_r have been introduced in section 3.2.5. The 2-norm and ∞-norm of this transfer function can be computed similarly to that of the full system by replacing $||\mathsf{G}^\dagger(\omega)||$ by $||\mathsf{G}_r^\dagger(\omega)|| = \sqrt{\mathsf{G}_r(\omega)\mathsf{G}_r^\dagger(\omega)}$ in equations (3.34). The quantity $||\mathsf{G}_r^\dagger(\omega)||$ accounts for the synthesized frequency response of the ROMs. It can be computed analogously to that of the full system by replacing $\hat{\mathsf{Y}}(\omega)$ in equation (3.35) with its equivalent on the reduced system $\hat{\mathsf{Y}}_r(\omega)$ which is defined by

$$\hat{\mathsf{Y}}_r(\omega) = \mathsf{B}_r^\dagger(-i\omega\mathsf{I}_r - \mathsf{A}_r^*)^{-1}\mathsf{C}_r^* \tag{3.37}$$

Note that B_r is equal to $\mathsf{S}^\dagger\mathsf{R}\mathsf{R}^\dagger$ for balanced models and reduces to $\mathsf{R}^\dagger$ for observable models. Notably, an equivalent of equation (3.33) can be expressed for the transfer function of the reduced systems by

$$\tilde{\mathsf{G}}_r(\omega) = \hat{\mathsf{Y}}_r^\dagger(\omega) = \hat{\mathsf{Y}}_r^*(\omega)\mathsf{Q}_1 \tag{3.38}$$

3.C Link with output projection

In this appendix, we show that the input projection described in this chapter is equivalent to the output projection introduced by Rowley (2005) when applied to the adjoint system. First, we consider the adjoint problem associated with equations 3.3. It can be written as

$$\frac{d\mathsf{Y}}{dt} = \mathsf{A}^\dagger\mathsf{Y} + \mathsf{C}^\dagger\upsilon(t) \tag{3.39a}$$
$$\zeta(t) = \mathsf{Y} \tag{3.39b}$$

where Y is the adjoint state, $\upsilon(t)$ and $\zeta(t)$ denote the adjoint input and output respectively. Using the definition of the Gramians previously introduced, one can observe that the controllability Gramian of the direct problem is equal to the observability Gramian of the adjoint one and vice versa. Note that, compared to the direct system, input and output sizes have been exchanged. In particular, the output space of this new system becomes the full state space.

In the output projection method introduced by Rowley (2005), the output dimension is very large and an output projected system is considered. For the present adjoint system, the projection that minimizes the 2-norm error between the original transfer function and the output-projected transfer function is given by the POD of the dataset $\{\mathsf{Y}(t)\,;\,t>0\}$. In other words, it stands for the projection onto the most controllable modes of the adjoint system, or equally to the projection onto the most observable modes of the direct system. As a result, the output projected system can be expressed as a function of the matrix R by

$$\frac{d\mathsf{Y}}{dt} = \mathsf{A}^\dagger\mathsf{Y} + \mathsf{C}^\dagger\upsilon(t) \tag{3.40a}$$
$$\tilde{\zeta}(t) = \mathsf{R}^\dagger\mathsf{Y} \tag{3.40b}$$

where we have introduced the low-dimensional projected output $\tilde{\zeta}(t)$. Returning to the direct problem, the input-output system reduces to

$$\frac{d\mathsf{X}}{dt} = \mathsf{A}\mathsf{X} + \mathsf{R}\tilde{\eta}(t) \tag{3.41a}$$

$$\tilde{m}(t) = \mathsf{C}\mathsf{X} \tag{3.41b}$$

which governs the dynamics of the input-projected system introduced in section 3.2.4.

Interlude

In chapter 3, we investigated the ability of the observable and balanced modes to design ROMs that capture the dynamics from any forcing to a given sensor. A significant part of the chapter was devoted to comparing the performance of the two resulting ROMs. In particular, balanced models are shown to be nearly optimal for capturing the input-output behavior.

In fact, the superiority of the balanced models has been noticed in both the SISO configuration of chapter 2 and the MISO configuration of chapter 3. These balanced ROMs are indeed supposed to approximate the original Balanced Truncation procedure whose very purpose is to model the input-output dynamics. On the other hand, it is not surprising that the ROMs based on the most controllable and observable modes are sub-optimal to do so since since they have a different purpose. For example the controllable modes are meant to gradually capture the most energetic patterns triggered by the inputs and do not include information from the outputs. Likewise, the most observable modes rank the states according to their contribution to the future outputs and are defined independently of the inputs.

An important feature of all the ROMs presented until now should be stressed out. The higher the number of inputs and/or outputs the higher are the sizes required by the ROMs to recover the original input-output dynamics. This intuitive result is illustrated by comparing the performance of the balanced models exposed in the two previous chapters. In a more general manner, a common feature of these ROMs is that the higher their requirement the higher their design complexity and their required size.

In the next chapter, we are concerned with the case where we neither define any actuators nor any sensors. More precisely, we look for models that capture the entire flow response whatever the initial condition and/or the external forcing. Within our input-output formalism, this reduces to considering all the degrees of freedom of the state-space as the inputs and the outputs, namely $\mathsf{B} = \mathsf{I}$ and $\mathsf{C} = \mathsf{I}$. Indeed, the input-output behavior is then characterized by the impulse response which may be formally written as $e^{\mathsf{A}t}$, which is also the time propagator associated with the original flow dynamics. Analogously, the transfer function reads $(i\omega\mathsf{I} - \mathsf{A})^{-1}$ which also reduces to the resolvent operator associated with the flow dynamics.

The derivation of such models that capture the entire flow dynamics may be of great technological interest. Let us mention the linear state-space flow control, such as the Linear Quadratic Gaussian (LQG) control. When designing a closed-loop flow controller, these ROMs would allow to both capture the effect of any unknown noise, as it would be in real experiments, but also to directly target a minimization of the perturbation kinetic energy rather than some sensor measurement. Another application concerning the state estimation of large-scale flows is worth mentioning. For instance, it was shown by Farrell & Ioannou (2001*b*) how to use reduced-order models to obtain near-optimal low-order state estimators. They constructed a reduced-order Kalman filter for the purpose of state identification in a time-dependent quasigeostrophic storm track model. The ROMs used in the process were required to capture the predominant part of the entire system dynamics.

How to design such ROMs has already been reported in the literature, see Farrell & Ioannou (2001*b*,*a*). However, the considered flow configurations had sufficiently low degrees of freedom so that the controllability and observability Gramians could be computed directly from two Lyapunov equations. As soon

as the number of degrees of freedom is higher than $O(10^5)$, such a computation becomes not tractable. Furthermore, one cannot either extend the output or input projection procedures since they are technically based on the small number of inputs or outputs. Instead, we present in the next chapter a new way of approximating the controllable, observable and balanced modes in a case where the inputs and outputs stand for the whole state-space. The key idea is to express a decomposition of the Gramians in the frequency domain as a function of the resolvent operator. For each frequency, the resolvent operator is then approximated by truncating its singular value decomposition.

An important part of the next chapter is devoted to the interpretation of the projection bases as deterministic flow processes deriving from the stochastically forced flow. Indeed, assuming that the flow is forced white in space and time, we show that the controllable modes reduce to the states accounting for a maximum contribution to the flow response sustained variance. These are the so-called POD modes, see Lumley (1970), which reduce to the largest eigenvectors of the covariance matrix of the system. Note that the most controllable modes investigated in chapter 2 are called POD modes owing to the technique used to compute them. Strictly speaking, one may only use the terminology POD for the present modes which are known to order the coherent process of the uncertain flow response, see Aubry *et al.* (1988); Berkooz *et al.* (1993); Holmes *et al.* (1997). Alternatively, these modes are also referred to as the Empirical Orthogonal Functions (EOFs) in the literature, see North (1984). On the other hand the most observable modes are shown to rank the forcing structures according to their effectiveness in producing the statistically maintained variance of the flow response; they are called the Stochastic Optimals (SOs). The first part of the next chapter assess the decomposition of the backward facing step flow dynamics in the frequency domain through the singular value decomposition of the resolvent. Secondly, assuming the flow is stochastically forced with a white in space and time disturbance, we rank the Empirical Orthogonal Functions and the Stochastic Optimals as deterministic flow states and illustrate their ability to design efficient ROMs. Furthermore, balanced models are considered and their performance is also evaluated. The results outlined in the next chapter have been reported in an article that have been submitted to the *Journal of Fluid Mechanics*.

Chapter 4

Frequency decomposition and model reduction of the full dynamics

Methods for approximating stable linear fluid dynamics by a system of lower order are examined. The case of a two-dimensional backward facing-step flow is considered throughout of the study within the framework of the incompressible Navier-Stokes equations. We design reduced-order models which are meant to capture its global dynamics on a given frequency interval. To this end, a singular value decomposition and truncation of the resolvent is performed in order to identify a sub-dynamics spanned by the leading optimal harmonic forcing and responses. Three bases are then computed on this sub-dynamics: the Empirical Orthogonal Functions, the Stochastic Optimals and the Balanced modes. Reduced-order models are obtained by projecting the original equations onto these three bases. The resulting models are evaluated by their performance in capturing the original time propagator and resolvent. These models are shown to accurately recover the optimal perturbations (the leading singular vector of the time propagator) but also the input-output behavior of the flow (the leading singular vectors of the resolvent). In general, provided that a flow dynamics are predominantly represented on a particular frequency range, these models can potentially recover the accurate flow responses from any forcing or initial condition.

4.1 Introduction

A typical feature of a wide variety of open flows is their very unstable nature and high sensitivity to background perturbations. This behavior is classically observed on boundary layers, mixing layers, jets or separated flows, even at low Reynolds numbers. Such a behavior is known to come from convective instabilities where the disturbances amplitude grows while being advected downstream. These flows are called selective noise-amplifier due to their ability to preferentially amplify some particular frequencies. Modelling such flows is thus rather challenging since it requires to capture the effect of an upstream and potentially complex noise, which is unknown in practice. Yet, designing low-order models for such flows is crucial for many engineering applications which includes flow control, climate modeling or forecasting in meteorology.

Owing to these flow's ability to selectively amplify disturbances, it is natural to resort to an input-output formulation of their dynamics. Thus, the purpose of an efficient model is to capture the effects (outputs) of the flow when subjected to some perturbations or forcing (inputs). For example, the recent work of Bagheri *et al.* (2009*b*) relies on a Reduced Order Model (ROM) which captures the linear dynamics between some inputs (actuators) to some outputs (sensors) in view of performing a closed-loop control of a spatially developing boundary layer. In this case, a successful stabilization of the flow was observed owing to the design of an efficient Linear Quadratic Gaussian compensator which was required to minimize the measurement energy at a sensor (output) in the presence of a noise taken as a single input. Now, it would be of great technological interest to possess a model which could accurately

recover the entire flow response (all the outputs) from all possible forcing (all the inputs). Such an improvement would allow to both (i) capture the effect of any unspecified noise, as it would be in real experiments, but also to directly target a minimization of the global perturbation kinetic energy rather than some measurement signal. Additionally, in a more general manner, it is also of fundamental interest to determine the minimum number of flow structures which represent the dynamics of a flow. The design of such low-order models motivated this present work.

We are concerned here with the model reduction of large linear fluid systems and, more particularly, in the design of accurate ROMs which capture the dynamics from all the possible inputs to all the possible outputs. As a prototype of amplifier flow, a two-dimensional backward facing step flow is considered, see figure 4.1. This globally stable flow is assumed to be driven by uncertain external disturbances at the upstream which are selectively amplified in the shear layer. The ROMs designed here are meant to accurately recover the resulting flow response or, in other words, its global amplifier behavior. The model reduction is performed by projecting the full original equations onto a reduced basis of modes. The choice of relevant fluid structures, yet far from being obvious, is of utmost importance; a significant part of this chapter is devoted to this important question.

Based on the previous studies of Cossu & Chomaz (1997); Chomaz (2005) on the space-time dynamics of open flows, it is well recognized that a global amplifier behavior could derive from convective instabilities. More precisely, the possibly large transient energy growths could be attributed to the nonorthogonality of the set of global eigenmodes associated with the flow linear dynamics. Based on this observation, Ehrenstein & Gallaire (2005); Alizard & Robinet (2007) investigated the global convective dynamics of a two-dimensional flat-plate boundary layer by means of an appropriate superposition of global eigenmodes. They showed that the optimal amplifier scenario based on a reduced basis of Tollmien-Schlichting type modes gives rise an upstream localized wavepacket that grows while being advected downstream. This work was latter revisited by Åkervik *et al.* (2008); Henningson & Åkervik (2008) where the initial perturbation leading to a maximum energy growth was found as an upstream located wavepacket inclined against the shear. This latter observation demonstrated the ability of the global modes to recover the initial Orr mechanism exposed by Butler & Farrell (1992) in shear flows where perturbations rise to an upright position while borrowing energy from the mean shear. Since then, global modes were considered as a good candidate to provide ROMs for flow control applications. For instance, a successful stabilization of the unsteadiness over an open-cavity flow was performed by Åkervik *et al.* (2007) by using a ROM based on a global modes expansion of the perturbation. However, such a success was later shown to strongly dependent on the dynamics of the specific flow situation. For example, it has been illustrated on the Ginzburg-Landau equation by Bagheri *et al.* (2009*c*) that the spatial support of the direct and adjoint global modes may be strongly separated from the actuators and sensors location. Later, the work by Barbagallo *et al.* (2009) pointed out the poor efficiency of a ROM based on global modes to capture the stable input-output dynamics of a square cavity flow. This inability was attributed to the lack of controllability (ability of the applied forcing to reach flow states) and/or observability (ability of flow states to register at the sensor locations) of the global modes.

Since both controllability and observability are critical quantities in the design of ROMs, an expansion basis that balances the two concepts has been introduced by Moore (1981) for stable linear input-output systems. The model reduction associated with the projection onto the set of the most and equally controllable and observable structures is known as Balanced Truncation. Interestingly, this technique is shown to quasi-optimally capture the input-output transfer function since it offers theoretical bounds on its error which are close to the lower bound achievable by any ROM, see Dullerud & Paganini (2000). Owing to this useful property, Balanced Truncation yields an efficient and popular model reduction process. For instance, it has been used to perform closed-loop control on simple one-dimensional amplifier flows, see Joshi *et al.* (1997); Högberg *et al.* (2003).

The computation of the balanced modes is based on the eigenvalue decomposition of the product of two operators: the controllability and observability Gramians. However, both are solution of a Lyapunov equation which cannot be solved for large systems. To alleviate this problem, a method was introduced

by Willcox & Peraire (2002) in order to approximate the Gramians by using flow state snapshots. Furthermore, the work by Rowley (2005) allowed to approximate directly the balanced modes with a snapshot method without computing the Gramians; this technique is called Balanced Proper Orthogonal Decomposition (BPOD). This new tool brought the model reduction of large-scale control problems within reach of current computational technology. As a result, several recent works were based on BPOD to perform the closed-loop control of more complex flows such as the stabilization of a square cavity flow by Barbagallo *et al.* (2009), of a flat-plate boundary layer by Bagheri *et al.* (2009*b*), or of the flow over an inclined airfoil by Ahuja & Rowley (2010). Even if BPOD yields a very effective technique to capture the dynamics from a moderate number inputs to a few outputs on large systems, it is not meant to model accurately the exact flow response (all outputs) from all possible inputs. The technique introduced by Rowley (2005) called output projection partially alleviates this problem by optimally projecting the output space so as to capture the dynamics from a few specific actuators to all possible outputs. Such models allow to capture the entire flow state response resulting from the forcing at these specific inputs. Output projection has been successfully applied in the works by Ilak & Rowley (2008) and Ahuja & Rowley (2010). However, since these models are based on a moderate number of inputs, they are unable to recover the dynamics resulting from all possible forcing or from an unspecified noise.

Another way of reducing the dynamics consists of exciting the system with an unbiased random forcing and projecting the dynamics onto the resulting most energetic structures. These modes are computed as the leading eigenvectors of the spatio-temporal covariance of the observed fluid fields; they are known as Proper Orthogonal Decomposition (POD) modes or also Empirical Orthogonal Functions (EOFs). Such modes are popular in the design of ROMs since they constitute energy-ranked coherent structures of the flow. Thus, model reduction based on POD has been pursued by a great many researchers. For instance, one may refer to the works by Hasselmann (1988); North (1984) in the context of climate modeling and by Lumley (1970); Sirovich (1987); Berkooz *et al.* (1993); Holmes *et al.* (1997) in turbulence modeling. In these studies, the EOFs yield a well-suited basis for the representation of the evolved flow behavior. However, as underlined by Farrell & Ioannou (1993*a*) and Farrell & Ioannou (1993*b*), this basis is sub-optimal for capturing the transient dynamics of a flow. These authors further argued that not only the set of most energetic structures (the EOFs) need to be captured for an accurate representation of the dynamics but also the set of optimally growing perturbations at initial time, which are called Stochastic Optimals (SOs). This statement is even illustrated by Farrell & Ioannou (2001*a*) where the authors have compared the efficiency of various projection bases to recover the dynamics of a one-dimensional Couette flow. On the one hand, they found that both the EOFs and SOs lead to accurate ROMs. On the other hand, they showed that the best performance is obtained from balanced modes; by retaining the structures representing both the optimals and the evolved optimals in a balanced manner. This latter technique consists of considering each degree of freedom of the flow state as an input and an output and applying the Balanced Truncation procedure introduced by Moore (1981). Even if the flow structures exposed by Farrell & Ioannou (2001*a*) yield promising projection bases to design accurate ROMs, their computation remains not tractable for large fluid systems, even if using a snapshot method.

Based on the encouraging results by Farrell & Ioannou (2001*a*), this chapter aims at providing a new tool to approximate the EOFs, the SOs and the balanced modes for large fluid systems. As an illustration, the procedure is applied to the two-dimensional backward-facing step flow. The present technique is not based on any particular input (forcing) or output (sensing) on the flow. Instead, it is conceptually meant to capture the predominant dynamics on a given frequency interval. The key idea of this chapter relies on the singular value decomposition and truncation of resolvent operator in the frequency domain. Indeed, the eigenanalysis of the resolvent operator is shown to identify the most energetic part of the map linking all possible harmonic forcing to their associated responses. This decomposition is used to approximate the leading EOFs, SOs and balanced modes. Notably, a particular attention is paid to dynamics decomposition of the backward-facing step flow in the frequency domain by studying its optimal harmonic forcing and responses. The three projection bases are then analyzed in light of this frequency decomposition. Lastly, the resulting low-order models are shown to accurately capture the

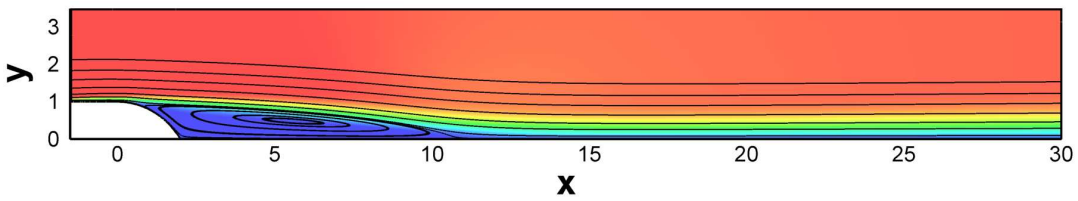


Figure 4.1: Backward-facing step flow investigated here. The steady-state solution at $Re = 600$ is depicted by its streamlines and longitudinal velocity.

original dynamics.
The chapter is organized along the following outline: the backward facing step flow is presented in section 4.2 where we introduce the base flow and the non-reduced transient dynamics. Then, section 4.3 is devoted to the decomposition in the frequency domain of the input-output dynamics. In particular, the optimal harmonic forcing and responses of the flow are investigated. Next, the leading EOFs, SOs and balanced modes are computed and exposed in section 4.4. The performance of the ROMs in capturing the original dynamics is assessed in section 4.5 for all three projection bases. Finally, concluding remarks and future prospects are given in section 4.6.

4.2 Flow configuration and modeling

4.2.1 Base flow

We are concerned with the global dynamics and model reduction of the incompressible flow over a two-dimensional backward-facing step. The geometry of this step, which is depicted in figure 4.1, was originally studied by Duriez (2009). It is made up with two flat plates linked by a circular arc extending from $(x = 0, y = 1)$ to $(x = 2, y = 0)$. The step height h and the incoming flow velocity U_∞ are chosen as the characteristic length and velocity scales so that the Reynolds number is defined by $Re = \frac{U_\infty h}{\nu}$ where ν is the kinematic viscosity.
The flow dynamics are linearized about a base flow chosen as the steady solution of the Navier-Stokes equations at $Re = 600$. To compute the base flow, the domain is chosen large enough so as to obtain a uniform free-stream. The flow enters the domain with the constant velocity U_∞ from the left at $x = -20$. We imposed a free-slip condition on the upstream part of the lower boundary $(-20 \leq x \leq -2, y = 1)$. The beginning of a laminar boundary layer is then enforced at $x = -2$ by imposing a no-slip condition on the remaining lower boundary between $(-2 \leq x \leq 100)$. A symmetry condition is implemented on the top boundary at $y = 20$ and a standard outflow condition is prescribed at the outlet $x = 100$. The base flow is computed by using a Newton method similar to that used by Sipp & Lebedev (2007).
The base flow solution at $Re = 600$ is depicted in figure 4.1 by its streamlines and longitudinal velocity. The flow displays an elongated recirculation bubble between separation at $x \approx 0.6$ and reattachment at $x \approx 11$. The boundary layer at $x = 0$ has a displacement thickness $\delta^* = 0.082$ leading to a Reynolds number based on the displacement thickness of $Re_{\delta^*} = 49.2$.

4.2.2 Linear dynamics

The dynamics of small perturbations $(\boldsymbol{u}, p)$ about the base flow, denoted by $\boldsymbol{U}$, are governed by the linearized Navier-Stokes equations

$$\begin{cases} \partial_t \boldsymbol{u} + \boldsymbol{U} \cdot \nabla \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \boldsymbol{U} = -\nabla p + Re^{-1} \nabla^2 \boldsymbol{u} \\ \nabla \cdot \boldsymbol{u} = 0 \end{cases} \tag{4.1}$$

This equation is spatially discretized using a finite elements approach with Taylor-Hood elements (P2-P2-P1) and implemented using the FreeFem++ software, see Hecht *et al.* (2005). Once discretized, it is

shown in appendix 4.A that equations 4.1 can be written in the following dynamical system form

$$\dot{\mathsf{x}} = \mathsf{A}\,\mathsf{x} \tag{4.2a}$$

where x is the vector of the discretized velocities $\boldsymbol{u}$ and the matrix A stands for the discretized Navier-Stokes operator. This discretization has been performed on a smaller mesh extending from $(-1.5 \leq x \leq 30)$ and with $(y \leq 3,5)$. This mesh proved to be sufficiently large to accurately capture the perturbation dynamics along the shear layer. The new computational domain is the one depicted in figure 4.1. Homogeneous Dirichlet boundary conditions are used at the inflow $x = -1.5$ and on the wall. A symmetry condition is adopted at the upper boundary $y = 3.5$ and a free outflow condition $p\boldsymbol{n} - Re^{-1}(\nabla \boldsymbol{u}) \cdot \boldsymbol{n} = 0$ is used at the outlet $x = 30$, where $\boldsymbol{n}$ is the outward unit normal vector. Note that the base flow has been interpolated on this new mesh. The number of degrees of freedom of the resulting problem is $n = 170\,260$, which also accounts for the dimension of the vector x or the matrix A in equation (4.2).

In this context, the goal of this chapter is to reduce as much as possible this finite-sized system while preserving the original fluid dynamics. Before proceeding with model reduction, the non-reduced dynamics need to be introduced together with the means to describe it.

4.2.3 Amplifier behavior

The perturbations amplitude is quantified by its kinetic energy contained in the entire domain. We introduce the associated inner product $\langle\,\rangle$, which is defined for any states x_1 and x_2 by

$$\langle \mathsf{x}_1, \mathsf{x}_2 \rangle = \mathsf{x}_1^\dagger \mathsf{x}_2 = \int_\Omega \boldsymbol{u}_1.\boldsymbol{u}_2 \, d\Omega \tag{4.3}$$

where Ω is the fluid volume, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ denote the associated velocity fields and the superscript † denotes the adjoint of a state. For the chosen Reynolds number $Re = 600$, the flow is observed to be globally stable since the matrix A does not display any unstable eigenvalues. Consequently, the energy of any initial condition will decay exponentially to zero for sufficiently large times. However, perturbations may be transiently amplified owing to the non-normality of A, see Trefethen & Embree (2005); Chomaz (2005). The potential of the flow to amplify upstream perturbations is illustrated here.

The explicit solution of the dynamical system (4.2) may be expressed as

$$\mathsf{x}(t) = e^{\mathsf{A}t}\mathsf{x}_0 \tag{4.4}$$

which means that the state solution x at time t can be directly computed from the initial condition x_0 at time $t = 0$ from the time propagator $e^{\mathsf{A}t}$. Then, the perturbation energy gain in time may be written as

$$\frac{\langle \mathsf{x}(t), \mathsf{x}(t) \rangle}{\langle \mathsf{x}_0, \mathsf{x}_0 \rangle} = \frac{\langle \mathsf{x}_0, e^{\mathsf{A}^\dagger t} e^{\mathsf{A}t} \mathsf{x}_0 \rangle}{\langle \mathsf{x}_0, \mathsf{x}_0 \rangle} \tag{4.5}$$

where we have introduced the adjoint operator $\mathsf{A}^\dagger$ defined from the inner product $\langle\,\rangle$. Hence, the maximum energy gain over all possible initial conditions is computed by the maximum eigenvalue of $e^{\mathsf{A}^\dagger t} e^{\mathsf{A}t}$ and the optimal initial condition by its leading eigenvector. These quantities have been computed from the same iterative algorithm and numerical methods used by Marquet *et al.* (2008). We have represented in figure 4.2(a) the optimal energy gain, denoted by the L_2 matrix norm as $||e^{\mathsf{A}t}||_2^2$, as a function of time. We observe that the potential of the flow to amplify perturbation is substantial; energy gain may exceed 10^4. The maximum gain is reached for an optimal time $t = 18$. Additionally, the optimal initial perturbation computed for that optimal time and the perturbation into which it evolves

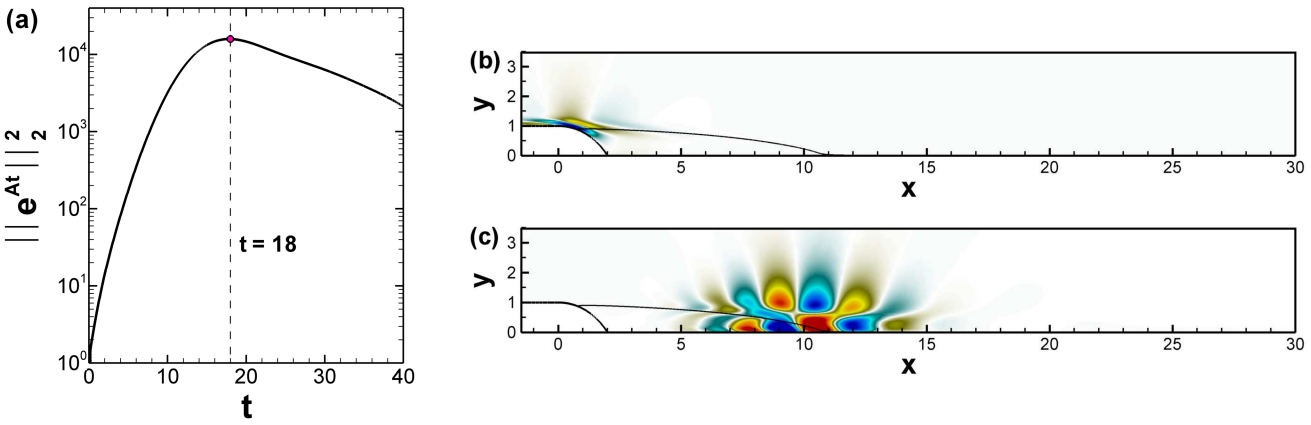


Figure 4.2: (a) Optimal perturbation gain $||e^{\mathsf{A}t}||_2^2$ at time t. (b) Optimal initial condition computed for $t = 18$. Figure (c) accounts for its evolved optimal state at $t = 18$. Both fields are represented by their longitudinal velocity.

are displayed in figure 4.2(b,c). The optimal initial perturbation shown in figure 4.2(b) is localized upstream and is inclined along the shear. On the other hand, the final optimal perturbation shown in figure 4.2(c) is located just beyond reattachment. These results are consistent with the existing literature, see Marquet *et al.* (2008); Blackburn *et al.* (2008), emphasizing the ability of the perturbations to be amplified along the recirculation bubble. Notably, the inclination of the initial condition highlights the well known Orr mechanism as a way to extract energy from the mean shear by transporting momentum down the velocity gradient (Butler & Farrell (1992)).
The modeling of the original linear flow dynamics reduces to capturing its amplifier behavior. As a result, the dynamics to be captured by the ROMs consist of the map linking all possible initial conditions x_0 to their future state $\mathsf{x}(t)$. As a result, the objective of the model reduction is to approximate as closely as possible the original time propagator $e^{\mathsf{A}t}$ for all positive times.

Analogously, the transient flow dynamics may also be investigated by considering the responses of the flow to permanent forcing. Owing to the linear nature of the equations, this analysis can be performed by considering frequency-separated harmonic forcing. Assuming that system (4.2) is forced by the harmonic term $\hat{\mathsf{f}}e^{i\omega t}$, the equations become

$$\dot{\mathsf{x}} = \mathsf{A}\mathsf{x} + \hat{\mathsf{f}}e^{i\omega t} \tag{4.6a}$$

which has the harmonic solution $\mathsf{x}(t) = \hat{\mathsf{x}}e^{i\omega t}$, where $\hat{\mathsf{x}}(\omega) = (i\omega\mathsf{I}-\mathsf{A})^{-1}\hat{\mathsf{f}}$. The operator $\mathsf{R}(\omega) = (i\omega\mathsf{I}-\mathsf{A})^{-1}$ is defined as the resolvent and links the harmonic forcing to its associated response. Given that formulation, the flow dynamics can be analogously defined as the map linking all possible forcing to their associated response. Consequently, the reduced-order models are also required to capture the original resolvent operator for all frequencies.

This objective is rather challenging and it is practically expected that the ROMs only approximate the resolvent on a finite frequency interval where the dynamics is predominantly comprised. For that reason, a decomposition of the input-output dynamics in the frequency domain is performed in the next section in order to derive the projection bases and the ROMs.

4.3 Dynamics decomposition in the frequency domain

Studying the responses of a flow subjected to harmonic forcing is known as a receptivity analysis. Such an approach is commonly used to predict in which frequency bands flow instabilities are more likely to

be observed when exposed to permanent external forcing, see Alizard *et al.* (2009); Monokrousos *et al.* (2010); Sipp *et al.* (2010). In this section, the sets of the leading optimal forcing and responses are investigated to identify the predominant part of the input-output dynamics.

4.3.1 Singular Value Decomposition of the resolvent

The resolvent $\mathsf{R}(\omega) = (i\omega\mathsf{I} - \mathsf{A})^{-1}$ has been introduced so as to link an harmonic forcing $\hat{\mathsf{f}}$ to its associated response $\hat{\mathsf{x}}$. Consequently, by using the relation $\hat{\mathsf{x}} = \mathsf{R}(\omega)\hat{\mathsf{f}}$, the kinetic energy gain produced by a forcing may be expressed as

$$\frac{\langle \hat{\mathsf{x}}, \hat{\mathsf{x}} \rangle}{\langle \hat{\mathsf{f}}, \hat{\mathsf{f}} \rangle} = \frac{\langle \hat{\mathsf{f}}, \mathsf{R}^\dagger(\omega)\mathsf{R}(\omega)\hat{\mathsf{f}} \rangle}{\langle \hat{\mathsf{f}}, \hat{\mathsf{f}} \rangle} \tag{4.7}$$

where we have introduced the adjoint operator $\mathsf{R}^\dagger(\omega)$ derived from the inner product $\langle\,\rangle$. The matrix $\mathsf{R}^\dagger(\omega)\mathsf{R}(\omega)$ is a hermitian matrix so that its eigenvectors are orthogonal. It is thus possible to span the forcing space with the orthogonal basis $\hat{\mathsf{f}}_i$ derived from

$$\mathsf{R}^\dagger(\omega)\mathsf{R}(\omega)\hat{\mathsf{f}}_i = \lambda_i^2 \hat{\mathsf{f}}_i \tag{4.8}$$

where λ_i^2 denote the energy gains induced by the forcing $\hat{\mathsf{f}}_i$. Furthermore, the highest eigenvalue λ_1^2 denotes the maximum energy gain over all possible forcing, which is reached by the optimal forcing $\hat{\mathsf{f}}_1$. More generally, the leading forcing terms $\hat{\mathsf{f}}_i$ constitute the set of orthogonal functions leading to an optimal and ranked energy amplification. We call them optimal harmonic forcing. Their corresponding optimal responses $\hat{\mathsf{x}}_i$ can be obtained by solving

$$\hat{\mathsf{x}}_i = \mathsf{R}(\omega)\hat{\mathsf{f}}_i \tag{4.9}$$

which also form an orthogonal basis spanning the response space. In fact, it is noteworthy that $\hat{\mathsf{f}}_i(\omega)$ and $\hat{\mathsf{x}}_i(\omega)$ are the right and left singular vectors of $\mathsf{R}(\omega)$ associated with its singular values λ_i. The leading optimal forcing/responses and singular values are investigated next.

Equation (4.8) is a large eigenvalue problem of size n which cannot be solved directly. Instead, we used an iterative Arnoldi algorithm based on Krylov subspaces in order to compute the leading eigenvalues/eigenvectors, see Lehoucq & Scott (1997). It is based on the successive inversion of the sparse matrix $(i\omega\mathsf{I} - \mathsf{A})$ which is handled through a direct multifrontal sparse LU solver (MUMPS, see Amestoy *et al.* (2001)). Hence, the cost of this algorithm stems on the the cost of the LU decomposition of a large sparse complex matrix. In our case, an accurate computation of the leading 40 eigenvalues/eigenvectors has been performed for frequencies ranging from $\omega = 0$ to $\omega = 6$.
The optimal energy gains λ_i^2 are represented in figure 4.3 as a function of the frequency. In other words, there are 40 points represented on the graph for each frequency which account for $\lambda_1^2, \ldots, \lambda_{40}^2$. What is striking is the existence of branches, which one displays a broad and high peak. The description of these branches and their physical interpretation is presented next.

4.3.2 First branch: the shear layer dynamics

Let us start with the higher and curved branch denoted on the graph as the branch $A1$. This one comprises the first singular value λ_1 which largely prevails for low frequencies until $\omega \approx 2.6$. In this range, the largest energy gain λ_1^2 displays a broadband response typical of amplifier flows that reach a maximum of $2.2\,10^6$ at $\omega \approx 0.75$. Yet, for higher frequencies, this branch crosses other branches and is then no more represented by the first singular value.
To interpret this branch, we investigate the forcing and responses, namely the right and left singular vectors of $\mathsf{R}(\omega)$, associated with its singular values. We have represented in figure 4.4 these singular vectors. On the left part, figures 4.4(a,c,e,g) represent the optimal harmonic forcing at the frequencies

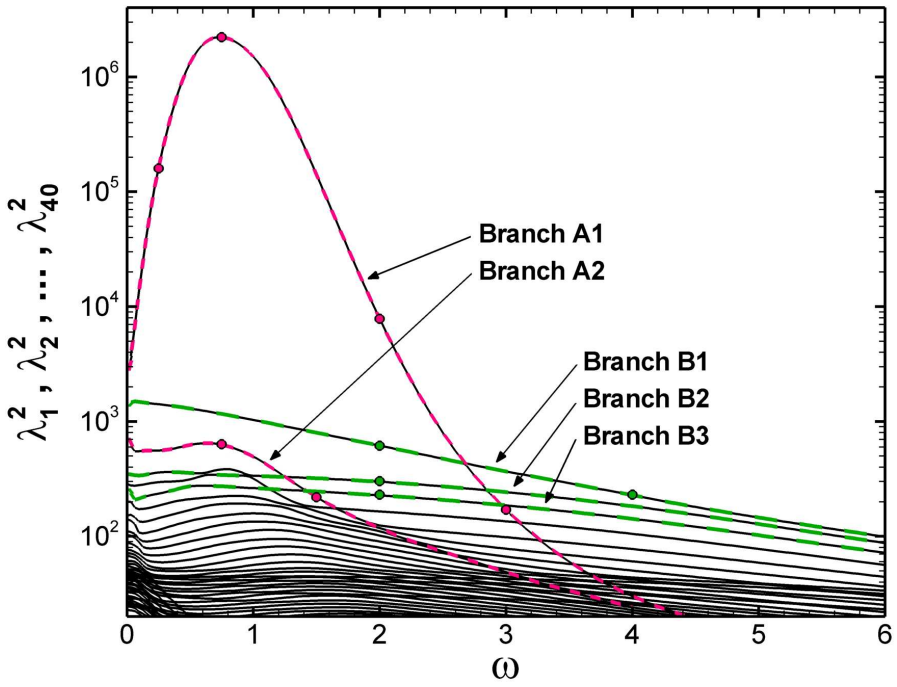


Figure 4.3: First 40 optimal harmonic gains $\lambda_1^2, \ldots, \lambda_{40}^2$ versus frequency ω.

$\omega = 0.25, 0.75, 2$ and 3 respectively. Analogously, figures 4.4(b,d,f,h) stand for the associated optimal harmonic responses at the same frequencies. Note that these flow structures are displayed by their longitudinal velocity. We observe that the optimal forcing fields are localized upstream, near separation, and display inclined patterns along the shear so as to exploit the Orr mechanism. Regarding the associated responses, they display typical wave packet flow structures whose spatial support extends farther downstream. Excitation at higher frequencies leads to smaller flow structures and the resulting responses remain spatially localized at the upstream part of the separated shear layer. On the contrary, for lower frequencies, the resulting flow structures are much bigger and the responses extend much farther until the downstream part of the boundary layer. These results are consistent since the high-frequency instabilities are observed where the shear layer is the thinnest whereas the low-frequency ones are observed where the boundary layer and the local shear are the thickest. High-frequency responses, such as those depicted in figure 4.4(f) and figure 4.4(h), highlight a Kelvin-Helmholtz amplification mechanism since perturbations grow in the shear layer as a wave of corotative rolls. These flow structures are furthermore characterized by a phase speed $v_\phi \approx 0.55$. Alternatively, the low-frequency responses such as those depicted in figure 4.4(b) and figure 4.4(d) are reminiscent of the global modes associated with a convective Tollmien-Schlichting instability. Such modes have been carefully investigated in the previous works by Ehrenstein & Gallaire (2005); Alizard & Robinet (2007); Åkervik *et al.* (2008) and their phase speed is rather close to $v_\phi \approx 0.4$.

All the optimal forcing and responses on branch $A1$ have a common point, they physically account for the amplification and advection of a wave packet along the recirculation bubble and the downstream shear. Thus, this branch may be called a "shear layer" branch.

4.3.3 Second branch: the free-stream dynamics

Now, we are concerned with the second most energetic branch which is called $B1$ in figure 4.3. Beyond $\omega \approx 2.6$, branch $A1$ is observed to be no more predominant and the energy gains are relatively much lower. The first singular values λ_1^2 then belong to branch $B1$. To further interpret the dynamics contribution of this branch, we have represented two forcing and response fields obtained on that branch

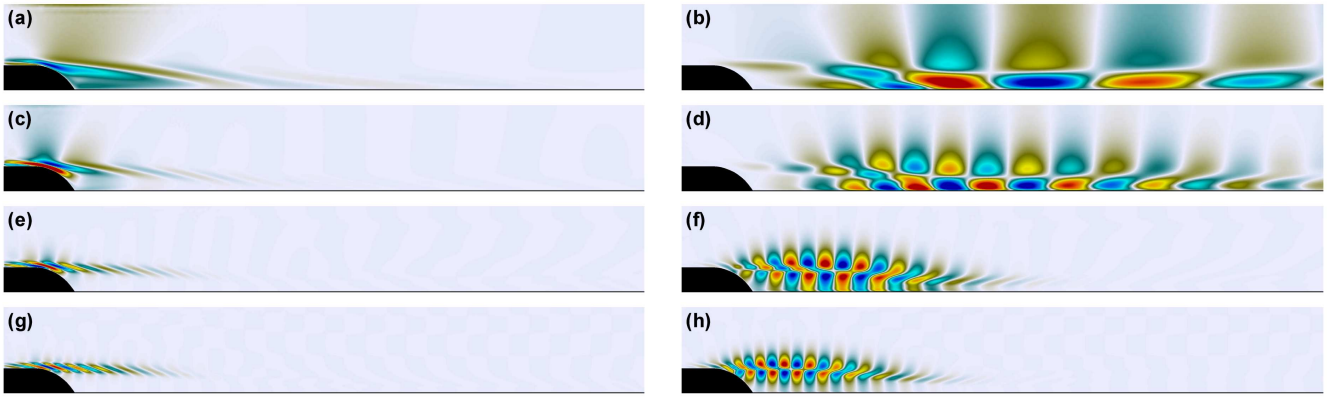


Figure 4.4: (a,c,e,g) Optimal forcing computed for $\omega = 0.25, 0.75, 2$ and 3 respectively. (b,d,f,h) represent the associated optimal responses at the same frequencies.

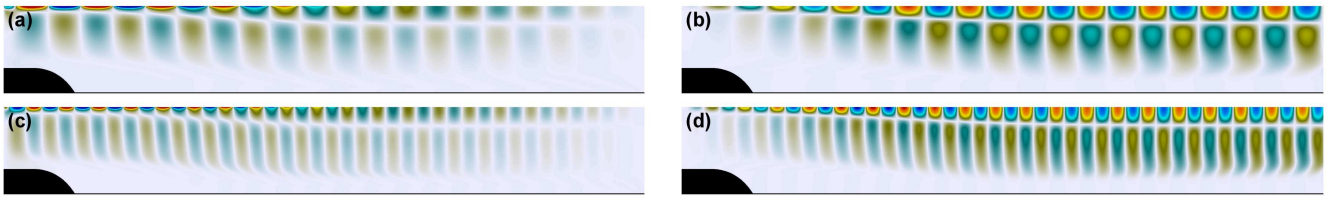


Figure 4.5: (a,c) Optimal forcing computed on branch $B1$ for $\omega = 2$ and 4 respectively. (b,d) represent the associated optimal responses at the same frequencies.

in figure 4.5. Figures 4.5(a,c) represent the optimal forcing at the frequencies $\omega = 2$ and 4 respectively while figures 4.5(b,d) account for the associated optimal responses. Similarly to the previous branch, the forcing fields are located upstream whereas the responses extend downstream. However, both the forcing and the responses reduce here to streamwise oscillating waves spreading over the free-stream part of the flow. These waves are modulated in the y direction and have two maxima. The only, yet most important, difference between these modes when changing the frequency is the streamwise wave length of the flow structures. Increasing the frequency leads to a smaller streamwise wave length for both the forcing and responses. Interestingly, their phase velocity is observed to be constant with the frequency and equal to $v_\phi \approx 0.95$.
As a result, the optimal forcing and responses on this branch are interpreted as a part of the dynamics governing the free-stream advection of perturbations at the velocity ≈ 1. This branch is thus called a "convective" branch.

4.3.4 Other branches: sub-optimal modes

Concerning the other branches, they can also be classified as either a convective or shear layer one. As an illustrative example, we have depicted on figure 4.6(a,b) the two harmonic responses taken from the branches labeled $B2$ and $B3$ respectively at the frequency $\omega = 2$. These responses are very similar to those observed on branch $B1$ except that they display smaller cross-stream wave lengths. More precisely, the response chosen on $B2$ displays three maxima in the y direction while that on $B3$ has four. These results are reminiscent of the fact that the optimal responses are orthogonal. Generally speaking, the convective branches are meant to convect all possible upstream forcing. The decomposition in such branches then highlights the contribution of all possible streamwise and cross-stream modulated forcing individually.
In an analogous way, we have represented in figure 4.7 the optimal responses taken on the shear layer branch $A2$ and computed for $\omega = 0.75$ and 1.5 respectively. Both display two streamwise separated

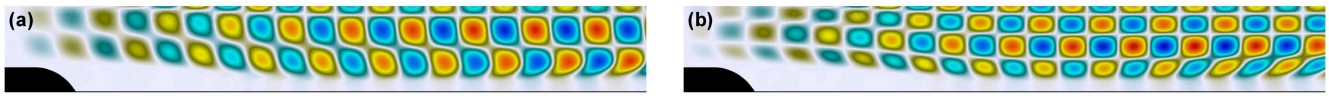


Figure 4.6: (a) Optimal response at the frequency $\omega = 2$ chosen on branch $B2$. (b) represents the optimal response at the same frequency chosen on branch $B3$.

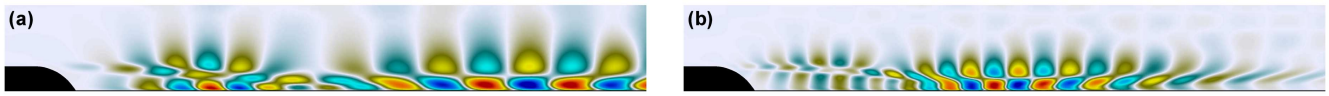


Figure 4.7: Optimal responses chosen on branch $A2$ at the frequencies $\omega = 0.75$ (a) and $\omega = 1.5$ (b).

wave packets localized in the shear layer. In addition, their streamwise wave length are the same as those of the responses observed on branch $A1$ at the same frequencies. This latter observation is also reminiscent of the orthogonality of the set of optimal responses. This second shear layer branch is much less energetic than the primary one and is even less than the first convective branch.

The frequency decomposition exposed in this section has emphasized a clear separation between: (i) upstream forcing and downstream responses but mostly (ii) a free-stream convective dynamics and an amplifier dynamics along the shear layer. The most energetic dynamics originate from shear instabilities and are predominant for the frequencies lower than $\omega \approx 2.6$. What is of utmost importance is to note that the singular value decomposition of the resolvent leads to a decomposition in the frequency domain which optimally orders the set of forcing (inputs) and associated responses (outputs). Notably, this decomposition highlights the most energetic part of the map linking all possible inputs to all possible outputs in the frequency interval of interest. Thus, it provides a milestone in order to design relevant projection bases for the model reduction.

4.4 Projection bases

The previous ranking of optimal forcing and responses for each frequency does not yield a ready-made basis of projection. To derive such a basis, we consider here the case where all the degrees of freedom and frequencies are equally forced to extract (i) the set of the most energetic responses and (ii) the set of the most amplified structures. Furthermore, (iii) the most and equally growing and energetic flow structures, also called balanced modes, are also considered. Technically, the idea is to decompose the dynamics by ranked stochastic processes which are derived from the previous frequency decomposition. This section is devoted to the introduction of these three bases of projection.

4.4.1 First basis: Empirical Orthogonal Functions

Since many flows exhibit few dominant instability mechanisms, the most energetic processes often give a good idea of the flow structure. Truncating a flow dynamics to its most energetic patterns is common in studies of turbulence (see Lumley (1970); Holmes *et al.* (1997)) and is known as the Proper Orthogonal Decomposition (POD). The most energetic patterns are usually referred to as POD modes or also Empirical Orthogonal Functions (EOFs).
The idea consists of artificially forcing system (4.2) by a stochastic and unbiased process in order to extract dynamical information from the flow response. This response can then be decomposed and ranked by the leading energetic coherent processes. We consider the dynamics of the stochastically forced system given by

$$\dot{\mathsf{x}} = \mathsf{A}\,\mathsf{x} + \mathsf{F}\xi(t) \tag{4.10}$$

where F is a matrix of size $n \times n$ governing the structure of the forcing. Furthermore, the vector $\xi(t)$ of size $n \times 1$ is a δ-correlated white noise process with zero mean. Namely, we have

$$[\xi_i(t)] = 0 \quad \text{and} \quad [\xi_i(t)\xi_j^\dagger(t')] = \delta_{ij}\delta(t-t') \tag{4.11}$$

where [] denotes the ensemble average and δ_{ij} the Kronecker symbol. The i^{th} component $\xi_i(t)$ of the noise is defined so as to force the i^{th} column of the matrix F. We assume that F is a unitary matrix ($\mathsf{F}^\dagger\mathsf{F} = \mathsf{I}$) so that its columns are orthogonal; the forcing is then defined as the sum of an orthonormal set of uncorrelated processes.

System 4.10 yields the dynamics of $\mathsf{x}(t)$, which is a random flow field, when subjected to a permanent random excitation. Since the flow is globally stable (A is stable), the solution of this problem tends to a statistically steady flow state for sufficiently large times. Furthermore, as shown in Farrell & Ioannou (1993*a*, 1996), this statistically steady flow state can be characterized by its covariance matrix $\mathcal{P} = [\mathsf{x}(t)\mathsf{x}(t)^\dagger]$ which is independent of F and reduces to

$$\mathcal{P} = \int_0^{+\infty} e^{\mathsf{A}t} e^{\mathsf{A}^\dagger t} dt \tag{4.12}$$

This matrix is hermitian so that its eigenvectors form an orthogonal basis. These eigenvectors are the so-called EOFs. By construction, the leading EOFs span most concisely the flow structures accounting for the variance of the statistically steady state, see Farrell & Ioannou (1993*a*). Indeed, assuming that the i^{th} EOF φ_i is associated with the eigenvalue β_i and that the random state $\mathsf{x}(t)$ is decomposed in the sum

$$\mathsf{x}(t) = \sum_{i=1}^{n} \alpha_i(t)\varphi_i \tag{4.13}$$

then it can be shown (Hœpffner (2006)) that the expansion coefficients $\alpha_i(t)$ are scalar uncorrelated random variables:

$$[\alpha_i(t)\alpha_j(t')] = \delta_{ij}\delta(t-t')\beta_i \tag{4.14}$$

Notably, the variance of the coefficients α_i are equal to the eigenvalues of φ_i. Therefore, the trace of the covariance matrix reduces to the total variance maintained in the statistically steady state. Since the space-time dynamics of the EOFs are uncorrelated from each other, they are commonly interpreted as coherent processes of the flow field. In other words, the random flow field $\mathsf{x}(t)$ with known covariance is decomposed in equation (4.13) as a family of coherent processes evolving in parallel. Energy is furthermore mostly carried by the leading eigenvectors which make them a preferred basis of functions for the truncation of dynamical systems.
A deterministic interpretation of the covariance matrix $\mathcal{P}$ is worth mentioning at this point. As shown in Luenberger (1969); Farrell & Ioannou (2001*a*), the minimum past forcing (in terms of energy) required to reach a given state x_0 is equal to $\mathsf{x}_0^\dagger\mathcal{P}^{-1}\mathsf{x}_0$. The operator $\mathcal{P}$ thus measures the controllability of the flow states and is also known as the controllability Gramian. The leading EOFs reduce to the flow structures the most easily triggered by the flow, or equivalently, that require the least energy to be forced. Consequently, the leading EOFs may also be referred to as the most controllable modes.

4.4.1.1 Computation

For a problem of small size, the covariance matrix $\mathcal{P}$ can be directly computed from a Lyapunov equation, see Farrell & Ioannou (1993*a*, 2001*a*). However, in our case, solving such an equation is not tractable. An alternative approach has to be employed to compute the leading EOFs without requiring to compute

or even store the matrix $\mathcal{P}$. The technique introduced here enables to overcome this difficulty and is based on the frequency domain decomposition of the dynamics presented in section 4.3.
First, $\mathcal{P}$ is formulated by a frequency integral by using Parseval's theorem:

$$\mathcal{P} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathsf{R}(\omega)\mathsf{R}^{\dagger}(\omega)d\omega \tag{4.15}$$

where the resolvent operator has replaced the time propagator. As a second step, the frequency integral is discretized in order to obtain the sum

$$\mathcal{P} = \frac{1}{2\pi}\sum_{i\in\mathbb{Z}} \mathsf{R}(\omega_i)\mathsf{R}^{\dagger}(\omega_i)\delta_i \tag{4.16}$$

where $\{\omega_i, i \in \mathbb{Z}\}$ is a given set of discrete frequencies and δ_i denotes appropriate quadrature coefficients. $\mathcal{P}$ reduces to the sum of the matrices $\mathsf{R}(\omega_i)\mathsf{R}^{\dagger}(\omega_i)$, which cannot be computed or stored. To alleviate this limitation, an approximation is performed. These matrices are replaced by their truncated eigenvalue decomposition:

$$\mathsf{R}(\omega_i)\mathsf{R}^{\dagger}(\omega_i) \approx \mathsf{X}\Lambda^2\mathsf{X}^{\dagger}(\omega_i) \tag{4.17}$$

where X and Λ^2 are the matrices of leading eigenvectors and eigenvalues respectively. For an approximation of rank k, X and Λ are of size $n \times k$ and $k \times k$ respectively. The rank k is thus a fundamental parameter to assess the efficiency of this approximation. Note that the columns of $\mathsf{X}(\omega)$ reduce to the left singular vectors of $\mathsf{R}(\omega)$ and mostly to the optimal responses previously denoted by $\hat{\mathsf{x}}_i(\omega)$. Furthermore, the diagonal elements $\lambda_i(\omega)$ of matrix $\Lambda(\omega)$ are the leading singular values of $\mathsf{R}(\omega)$. By using the decomposition (4.17), we can factor the covariance matrix $\mathcal{P}$ in

$$\mathcal{P} = \frac{1}{2\pi}\sum_{i\in\mathbb{Z}} \mathsf{X}\Lambda^2\mathsf{X}^{\dagger}(\omega_i)\delta_i = \mathbf{X}\mathbf{X}^{\dagger} \tag{4.18}$$

where the matrix $\mathbf{X}$ is obtained by stacking the flow states $\mathsf{X}\Lambda$ for each frequency ω_i with appropriate quadrature coefficients. For a set of n_ω frequencies, $\mathbf{X}$ is of size $n \times kn_\omega$. It is then possible to recover the leading eigenvalues/vectors of $\mathcal{P}$ by using the classical snapshot technique introduced by Sirovich (1987). First, we perform the eigenvalue decomposition of $\mathbf{X}^{\dagger}\mathbf{X}$ as

$$\mathbf{X}^{\dagger}\mathbf{X} = \mathsf{H}\mathsf{B}\mathsf{H}^{*} \tag{4.19}$$

where the diagonal matrix B contains the leading eigenvalues β_i associated with the EOFs, namely the leading eigenvalues $\mathcal{P}$. The leading EOFs φ_i are computed by

$$\varphi_i = \frac{1}{\sqrt{\beta_i}}\mathbf{X}\mathbf{H}_i \tag{4.20}$$

where $\mathbf{H}_i$ denotes the i^{th} column of the matrix $\mathbf{H}$. Note that the EOFs are orthogonal by construction. It should be emphasized that the present procedure is meant to approximate the EOFs from the subset of the k leading optimal harmonic responses.

4.4.1.2 results

The computation of the EOFs is based on the optimal harmonic responses of the flow, which are computed using the numerical method introduced in section 4.3. To discretize the frequency-domain integral defined in equation (4.15), an equidistant spacing between 129 frequencies has been adopted and quadrature coefficients corresponding to the 4^{th}-order Simpson method have been used. Note that the rank k of truncation has to be carefully taken into consideration to assess the accuracy of the approximation.

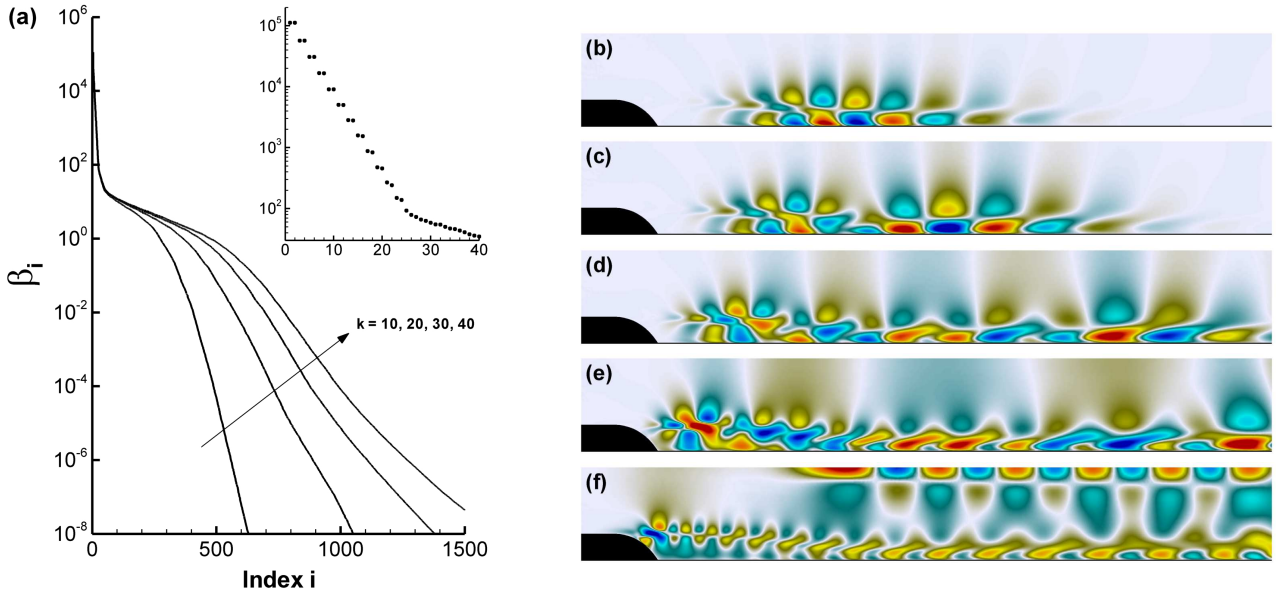


Figure 4.8: (a) first 1500 eigenvalues β_i. (b-f) depict the longitudinal velocity of the first, third, 9^{th}, 18^{th} and 40^{th} Empirical Orthogonal Functions.

We have represented in figure 4.8(a) the first 1500 eigenvalues β_i for various values of the approximation rank $k = 10, 20, 30$ and 40. The leading eigenvalues are observed to be converged when increasing the value of k and this also holds for the associated eigenvectors (not shown here). Additionally, figure 4.8(a) provides a zoom on the first 40 eigenvalues, which are converged in terms of the rank k. The first eigenvalues come in pair which indicates that the most significant EOFs are traveling structures. Furthermore, the leading eigenvalues fall off quite rapidly, indicating that the flow response to stochastic forcing is only driven by a low number of coherent processes. We have represented in figures 4.8(b-f) the longitudinal velocity of the first, third, 9^{th}, 18^{th} and 40^{th} EOFs respectively. They appear as wave packets that are spatially extended downstream, where the response energy to forcing is the largest. Contrary to the optimal responses exposed in section 4.3, the EOFs are not associated with a particular frequency. Rather, they account for the most energetic patterns emerging from a white noise forcing which excites all the frequencies and spatial wavelengths in an equal balance. Interestingly, the first EOF is characterized by the same streamwise wavelength as that of the optimal flow response at $\omega = 0.75$. This observation confirms that the forcing of all space-time scales results in a preferred frequency of $\omega = 0.75$ in the flow response, see figure 4.3. As for higher EOFs, their spatial support is observed to extend farther downstream. Furthermore, they display more complex structures characterized by more than one particular wavelength. In particular, smaller structures are observed near separation while larger structures are mainly localized downstream, see figures 4.8(e,f). This result is consistent since we compute the EOFs from a linear combination of the leading optimal harmonic responses taken over a frequency range, see equation (4.20).
As previously mentioned, these modes represent uncorrelated energy-ranked flow states so that the j^{th} EOF contributes to $\beta_j / \sum_{i=1}^{n} \beta_i \times 100\%$ of the total sustained variance. As an indicative illustration, the first 4 EOFs contain 71% of the total variance while the first 10 ones contain 94% of it (independently of the rank k).

4.4.2 Second basis: the Stochastic Optimals

The EOFs previously described consist of the set of optimal responses of the flow when subjected to an external stochastic forcing. However, owing to the non-normality of the base flow, the optimal responses are structurally distinct from the corresponding optimal excitations. In particular, the set of

forcing distributions yielding a maximum contribution to the variance of the statistically steady state differs from the EOFs. We call them the Stochastic Optimals (SOs). As mentioned by Farrell & Ioannou (2001*a*), an accurate representation of the original dynamics requires to capture both the set of optimally growing perturbations and those into which they evolve. Consequently, building ROMs by using the set of Stochastic Optimals stands for a natural alternative to the use of EOFs.
Considering system (4.10), the maintained variance can be expressed, see Farrell & Ioannou (1993*a*), by

$$[x^\dagger x] = \mathrm{trace}(\mathsf{F}^\dagger \mathcal{Q} \mathsf{F}) \tag{4.21}$$

where the matrix $\mathcal{Q}$ is defined by

$$\mathcal{Q} = \int_0^{+\infty} e^{\mathsf{A}^\dagger t} e^{\mathsf{A}t} dt \tag{4.22}$$

This matrix is also hermitian so that its eigenvectors are orthogonal. We call these eigenvectors the Stochastic Optimals (SOs). According to equation (4.21), they rank the forcing structures, i.e. the columns of F, by their contribution to exciting the maintained variance. Let ϕ_i be the i^{th} eigenvector of $\mathcal{Q}$ with its eigenvalue γ_i and κ_i be a set of scalars, then a forcing given by $\sum_{i=1}^n \kappa_i \phi_i \xi_i(t)$ will lead, for large times, to a statistically steady flow state of variance equal to $\sum_{i=1}^n \kappa_i \gamma_i$. In other words, the variance sustained by each eigenvector at statistical equilibrium is equal to its associated eigenvalue and the total variance is given by the trace of $\mathcal{Q}$. Consequently, any random forcing can be decomposed as a family of uncorrelated and orthogonal coherent processes ranked by their contribution to the evolved flow state sustained variance. Since energy is mostly triggered by these eigenvectors (the SOs), they also constitute a preferred basis of functions to design reduced-order models.
Interestingly, a deterministic interpretation of $\mathcal{Q}$ also exists. If we consider an initial state x_0, the time-integrated kinetic energy of its evolved state can be computed by

$$\int_0^{+\infty} \mathsf{x}^\dagger(t)\mathsf{x}(t)dt = \mathsf{x}_0^\dagger \left(\int_0^{+\infty} e^{\mathsf{A}^\dagger t} e^{\mathsf{A}t} dt \right) \mathsf{x}_0 = \mathsf{x}_0^\dagger \mathcal{Q} \mathsf{x}_0 \tag{4.23}$$

The operator $\mathcal{Q}$ thus measures the observability of the flow states and is also known as the observability Gramian. Its leading eigenvectors reduce to the flow structures leading to maximum time-integrated kinetic energy. For that reason, the leading SOs may also be referred to as the most observable modes.

4.4.2.1 Computation

The computation of the leading eigenvectors of $\mathcal{Q}$ involves the same numerical issue as that of $\mathcal{P}$. The computational technique exposed here is thus mainly similar. The matrix $\mathcal{Q}$ is written and discretized in the frequency domain

$$\mathcal{Q} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathsf{R}^\dagger(\omega)\mathsf{R}(\omega)d\omega = \frac{1}{2\pi} \sum_{i \in \mathbb{Z}} \mathsf{R}^\dagger(\omega_i)\mathsf{R}(\omega_i)\delta_i \tag{4.24}$$

where $\{\omega_i, i \in \mathbb{Z}\}$ is a given set of discrete frequencies and δ_i denotes appropriate quadrature coefficients. In this form, $\mathcal{Q}$ is formulated as the sum over the frequencies ω_i of $\mathsf{R}^\dagger(\omega_i)\mathsf{R}(\omega_i)$. Similarly to the computation of the EOFs, these matrices are approximated by their truncated eigenvalue decomposition:

$$\mathsf{R}^\dagger(\omega_i)\mathsf{R}(\omega_i) \approx \mathsf{Y}\Lambda^2\mathsf{Y}^\dagger(\omega_i) \tag{4.25}$$

where Y and Λ^2 are the matrices of leading eigenvectors and eigenvalues respectively. For an approximation of rank k, the matrices Y and Λ are of size $n \times k$ and $k \times k$ respectively. Analogously, we recover the diagonal elements $\lambda_i^2(\omega)$ which are the optimal energy gains at the frequency ω. In addition, the associated eigenvectors (columns of $\mathsf{Y}(\omega)$) reduce to the optimal harmonic forcing previously denoted by $\hat{\mathsf{f}}_i(\omega)$. Introducing this decomposition in equation (4.25) then leads to a factorization of $\mathcal{Q}$ in the form

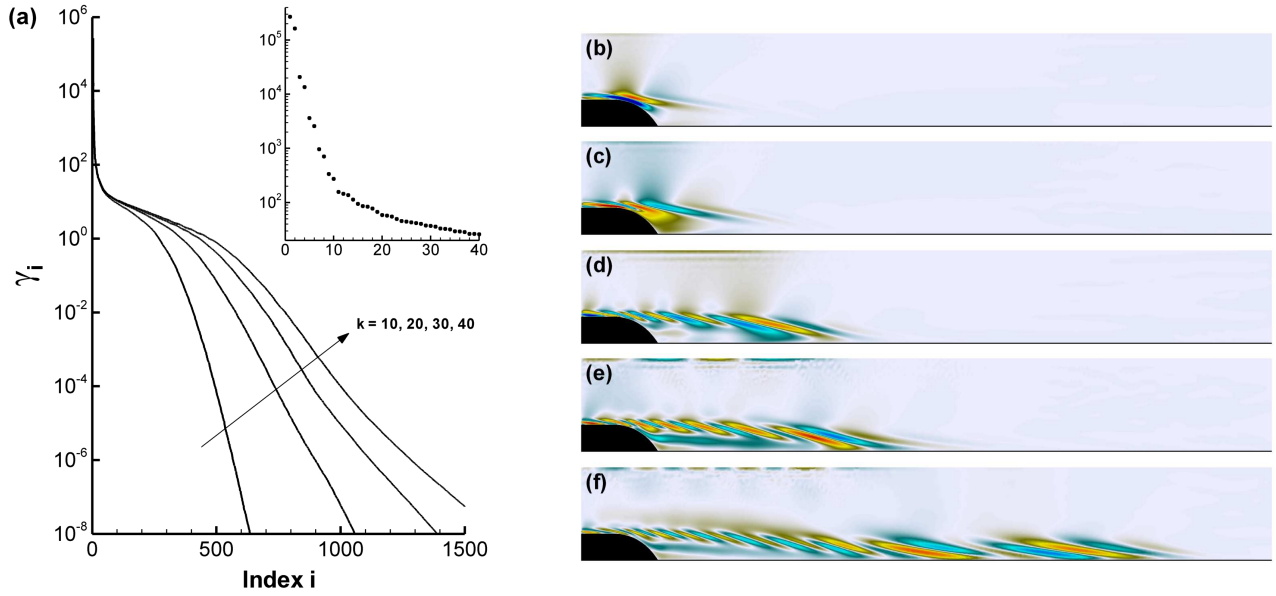


Figure 4.9: (a) first 1500 eigenvalues γ_i. (b-f) depict the longitudinal velocity of the first, third, 9^{th}, 16^{th} and 30^{th} Stochastic Optimals.

$$\mathcal{Q} = \frac{1}{2\pi} \sum_{i \in \mathbb{Z}} \Upsilon \Lambda^2 \Upsilon^\dagger(\omega_i)\delta_i = \mathbf{Y}\mathbf{Y}^\dagger \tag{4.26}$$

where the matrix $\mathbf{Y}$ is built by stacking the states $\Upsilon\Lambda$ for each frequency ω_i with appropriate quadrature coefficients. Similarly, for a set of n_ω frequencies, $\mathbf{Y}$ is of size $n \times k n_\omega$. To compute the leading eigenvalues and eigenvectors of $\mathcal{Q}$, we proceed analogously by using the eigenvalue decomposition of $\mathbf{Y}^\dagger\mathbf{Y}$ as

$$\mathbf{Y}^\dagger\mathbf{Y} = \mathsf{L}\Gamma\mathsf{L}^* \tag{4.27}$$

where the diagonal matrix Γ contains the leading eigenvalues of $\mathcal{Q}$, which are denoted by γ_i in the following. The leading associated eigenvectors, i.e. the Stochastic Optimals, are denoted by ϕ_i and are obtained from

$$\phi_i = \frac{1}{\sqrt{\gamma_i}} \mathbf{Y}\mathbf{L}_i \tag{4.28}$$

where $\mathbf{L}_i$ represents the i^{th} column of the matrix $\mathbf{L}$. Note that, by construction, these modes are also orthogonal. By this way, equation (4.28) allows to approximate the Stochastic Optimals from a linear combination of the leading optimal harmonic forcing over the frequency range of interest.

4.4.2.2 results

The computation of the Stochastic Optimals relies on the optimal harmonic forcing $\hat{\mathsf{f}}_i(\omega)$ introduced in section 4.3. They are computed similarly to the EOFs, by using the same frequency discretization.
We have depicted in figure 4.9(a) the first 1500 eigenvalues γ_i for different values of the approximation rank k. The cases $k = 10, 20, 30$ and 40 have been investigated. Analogously, the leading eigenvalues are observed to be converged when increasing the parameter k. Furthermore, the leading eigenvalues also come in pair and fall off rapidly. Consequently, (i) the first Stochastic Optimals are traveling patterns that are 90° out of phase and (ii) only a low number of forcing structures have a significant contribution to the evolved sustained variance. Figures 4.9(b-f) represent the first, third, 9^{th}, 16^{th} and 30^{th} Stochastic Optimals respectively. As one could expect, the leading SOs reduce to upstream located structures. This

result is consistent and further illustrates the spatial separation between optimal forcing and optimal responses in open flows, see Cossu & Chomaz (1997). Notably, they display flow structures that are inclined against the shear so as to make use of the Orr mechanism to optimally extract energy from the base flow. The first SO is mainly localized near the separation point and looks extensively similar to the optimal harmonic forcing obtained at $\omega = 0.75$, see figure 4.4(c). This observation confirms that the maximum variance possibly sustained by a white noise stochastic forcing exploits the preferred frequency $\omega = 0.75$ which has been exposed in section 4.3 as the most receptive frequency of the flow. Regarding the higher Stochastic Optimals, see figures 4.9(e,f), they display a much more extended spatial support. Reminding that the SOs yield a basis of uncorrelated forcing ranked by their energy contribution, it is interesting to assess the part of the total energy resulting from the leading ones. For the four values of k investigated, the first 4 modes lead to 97% of the total evolved variance while the first 10 ones yield about 99% of it.

4.4.3 Third basis: the Balanced modes

The first two bases introduced in this section have been designed so as to optimally capture the most energetic flow responses and the optimal excitations respectively. However, an accurate model reduction of the dynamics requires the faithful representation of both these flow structures, see Farrell & Ioannou ($2001a$). As a result, representing the dynamics of both the optimals and the evolved optimals in a balanced manner is of relevant interest.
The technique called Balanced Truncation has been introduced to this end for stable linear input-output systems, see Moore (1981); Glover (1984); Zhou & Doyle (1998). It consists of finding a projection basis of equally and optimally controllable and observable modes. If we denote the Balanced modes by t_i, this statement can be formulated by

$$\frac{1}{\mathsf{t}_i^\dagger \mathsf{P}^{-1} \mathsf{t}_i} = \mathsf{t}_i^\dagger \mathsf{Q} \mathsf{t}_i = \sigma_i \tag{4.29}$$

where the first term assesses the controllability of t_i while the second denotes its observability. Furthermore, both terms are equal to the quantity σ_i which is referred to as the Hankel Singular Value (HSV). The key idea of Balanced Truncation is then to discard the modes which are both weakly energetic (controllable) and weakly amplified (observable) in the flow. The remaining Balanced modes, associated with the highest HSVs, are then used as a projection basis. By construction, the leading Balanced modes are meant to capture the input-output dynamics rather than the most energetic forcing or responses. Balanced Truncation is known to be quasi-optimal to capture the input-output transfer function insofar as it yields theoretical bounds on its error which are close to the lower bound achievable by any ROM, see Dullerud & Paganini (2000). This useful property makes it a popular technique and motivated the choice of this third basis.

4.4.3.1 Computation

Technically, these modes are found as the leading eigenvectors of the product $\mathcal{P}\mathcal{Q}$ of the controllability and observability Gramians, see Moore (1981). Such a computation has been done in the case of small-sized systems by Farrell & Ioannou ($2001a$). In this latter work, the authors computed directly $\mathcal{P}$ and $\mathcal{Q}$ by solving two Lyapunov equations.

In the present chapter, we are concerned with a large fluid system and such a direct computation is not tractable. Thereby, a similar approximation based on the singular value decomposition of the resolvent is performed to overcome this limitation. We consider the approximations of P and Q which were introduced as

$$\mathcal{P} \approx \mathbf{X}\mathbf{X}^\dagger \quad \text{and} \quad \mathcal{Q} \approx \mathbf{Y}\mathbf{Y}^\dagger \tag{4.30}$$

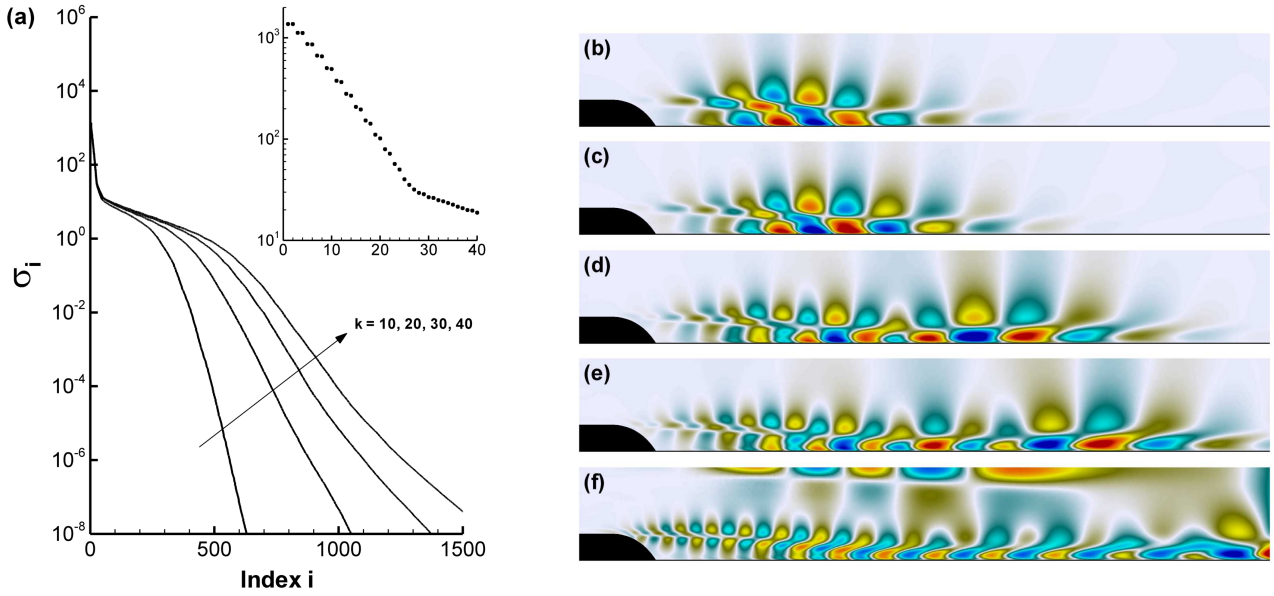


Figure 4.10: (a) first 1500 Hankel Singular Values σ_i. (b-f) display the longitudinal velocity of the first, third, 9^{th}, 14^{th} and 32^{th} Balanced modes.

where $\mathbf{X}$ and $\mathbf{Y}$ contain the normalized first k optimal harmonic forcing and responses of the flow on a discrete set of frequencies. It is then possible to find the Balanced basis from the algorithm introduced by Laub *et al.* (1987). This algorithm relies on the singular value decomposition of the cross product

$$\mathbf{Y}^{\dagger}\mathbf{X} = \mathsf{M}\Sigma\mathsf{N}^{*} \tag{4.31}$$

where the diagonal matrix Σ contains the Hankel Singular Values σ_i. The Balanced modes t_i and its bi-orthogonal set s_i are then recovered from

$$\mathsf{t}_i = \frac{1}{\sqrt{\sigma_i}}\mathbf{X}\mathsf{N}_i \quad \text{and} \quad \mathsf{s}_i = \frac{1}{\sqrt{\sigma_i}}\mathbf{Y}\mathsf{M}_i \tag{4.32}$$

4.4.3.2 results

The Balanced modes are computed for the approximation ranks $k = 10, 20, 30$ and 40 by using the same set of discrete frequencies ω_i. Figure 4.10(a) depicts the first 1500 Hankel singular values σ_i corresponding to the approximation ranks k investigated. The leading HSVs are observed to converge when increasing the approximation rank k. Analogously, they come in pair which highlights the pairwise occurrence of Balanced modes as a representation of traveling structures. Furthermore, owing to the HSVs fall off, it is reasonable to expect that the predominant part of the input-output dynamics may be captured by the leading modes.
We have depicted in figures 4.10(b-f) the first, third, 9^{th}, 14^{th} and 32^{th} Balanced modes t_i by their longitudinal velocity. They appear as downstream wavepackets similar to EOFs. On the other hand, their spatial support is localized in a slightly more upstream position, which is particularly noticeable for the first and third modes. This latter observation is reminiscent of their equal controllability and observability. Indeed, the Balanced modes are less controllable than the EOFs (and are not localized as much downstream) whereas they are also less observable than the SOs (and are not localized as much upstream). Regarding higher Balanced modes, this statement is less obvious, see figures 4.10(e,f). In addition, similarly to the EOFs and the SOs, higher modes display a much more extended spatial support.

4.5 Model reduction

In the previous section, we have introduced a way to approximate the leading EOFs, SOs and Balanced modes. For an approximation rank $k = 40$, we observed a convergence of approximately the first 100 modes, see figures 4.8(a), 4.9(a) and 4.10(a). The higher modes which are not converged in terms of the rank k may not be interpreted as either EOFs, SOs or Balanced modes. However, they also stand for good candidates to project the dynamics since they are conceptually designed from the leading optimal harmonic forcing and responses of the flow. Consequently, all the leading 1500 modes of the three different bases are considered and used to build the ROMs. Studying the performance of the ROMs in capturing the original amplifier dynamics is the purpose of this last section.

4.5.1 Projection of the original fluid dynamics

The model reduction is obtained by a Galerkin projection of the original dynamics (4.2). We denote by T the matrix whose columns are the leading r modes of interest (either φ_i, ϕ_i or t_i) and by S the matrix which contains its bi-orthogonal basis (namely φ_i, ϕ_i or s_i). The parameter r thus denotes the dimensionality of the projection basis such that T and S are of size $n \times r$. In addition, the bi-orthogonality condition reads

$$\mathsf{S}^\dagger \mathsf{T} = \mathsf{I}_r \tag{4.33}$$

where I_r denotes the identity of size $r \times r$. The flow state x is then expressed as a linear combination of the expansion basis by

$$\mathsf{x} = \mathsf{T}\mathsf{z} \quad \text{and} \quad \mathsf{z} = \mathsf{S}^\dagger \mathsf{x} \tag{4.34}$$

where the coefficients of this combination are contained in the vector z of size $r \times 1$ which is called the reduced state in the following. The Petrov-Galerkin projection then yields the reduced system

$$\dot{\mathsf{z}} = \mathsf{A}_r \, \mathsf{z} \tag{4.35}$$

where the reduced dynamical operator is obtained by $\mathsf{A}_r = \mathsf{S}^\dagger \mathsf{A} \mathsf{T}$ and is of size $r \times r$.

4.5.2 Time propagation

First, we examine the ability of the ROMs to recover the original time propagator. This performance is evaluated by considering the optimal energy gain $||e^{\mathsf{A}t}||_2^2$ over all possible initial conditions at time t. To compute this quantity on a ROM of size r, the low-rank time propagator $\mathsf{T}e^{\mathsf{A}_r t}\mathsf{S}^\dagger$ is considered instead of $e^{\mathsf{A}t}$. Note that the optimal energy gain has already been considered as a criterion to assess the accuracy of ROMs on the previous works by Åkervik *et al.* (2007, 2008); Henningson & Åkervik (2008). Results are exposed in figure 4.11 for models based on (a) EOFs, (b) SOs and (c) Balanced modes. In all cases, we observe a convergence to the original optimal curve displayed by bold dashed lines. To give an idea, the optimal energy gain obtained at time $t = 18$ differs from less than 2% from all three models of size 1500. Interestingly, the Balanced models yield the best performance for the smallest sizes r. This observation is illustrated by the performance of the ROMs of size 100 in figure 4.11. This result is consistent since Balanced Truncation is supposed to be quasi-optimal to model the dynamics. Furthermore, Balanced models have a monotonic increasing performance while those of the two other models is more erratic.

To further illustrate the dynamical behavior of the ROMs, the time evolution of the optimal perturbation on a 1500 EOFs model is displayed in figure 4.12. The choice of a model based on EOFs is arbitrary since the two other models yield very close results. The optimal initial condition computed for the optimal time $t = 18$ is displayed in figure 4.12(a) by its streamwise velocity. This perturbation is located in the vicinity of the separation point and is tilted against the shear. The time evolving optimal perturbation

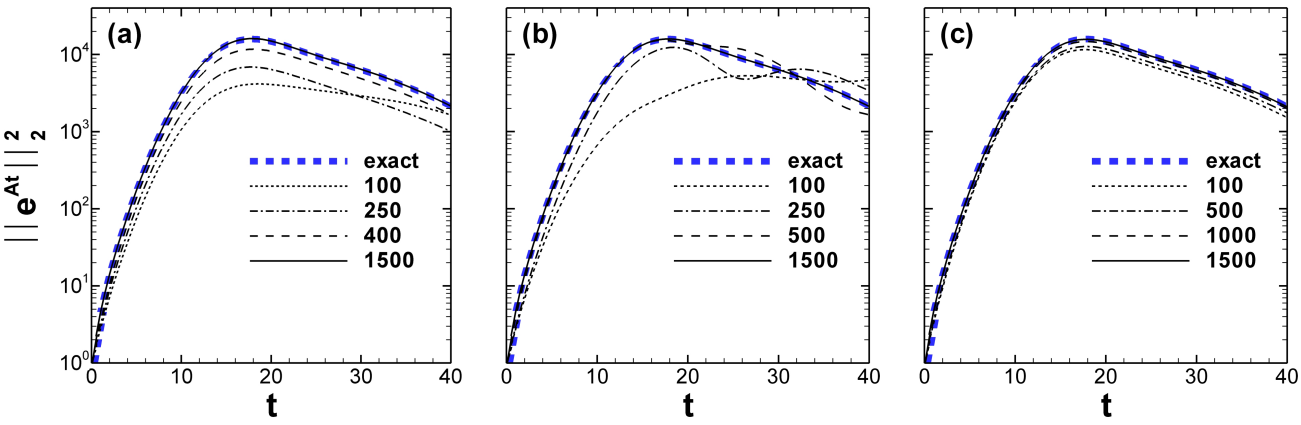


Figure 4.11: Optimal perturbation gain $||e^{At}||_2^2$ at time t for ROMs based on (a) EOFs, (b) SOs and (c) Balanced modes.

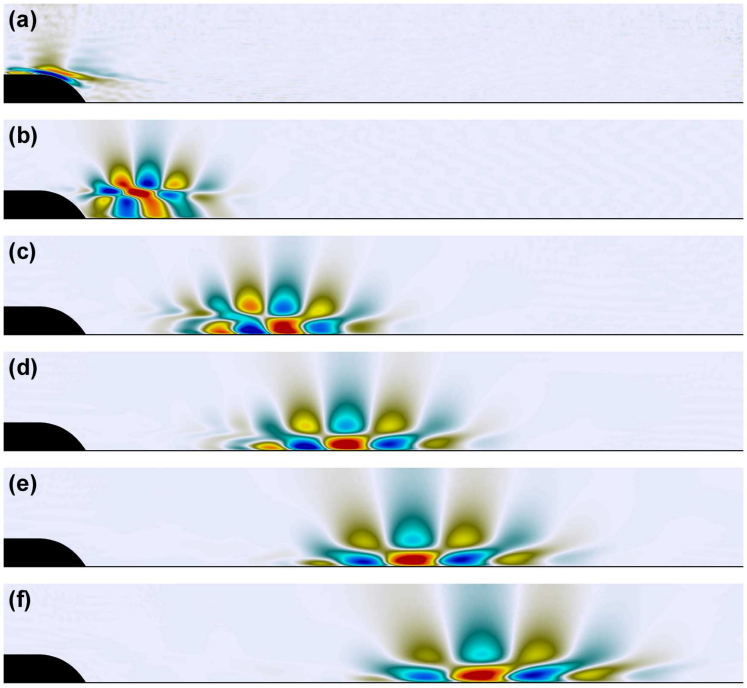


Figure 4.12: (a) Optimal initial perturbation computed for $t = 18$ on a ROM based on 1500 EOFs. (b-f) Evolution of this perturbation computed from the associated low-order time propagator for the times $t = 6, 18, 24, 30$ and 36 respectively. Flow structures are displayed by their longitudinal velocity.

computed from the low-order time propagator $\mathsf{T}e^{\mathsf{A}_r t}\mathsf{S}^\dagger$ is represented in figure 4.12. Figures 4.12(b-f) represent the perturbation at the times $t = 6, 18, 24, 30$ and 36 respectively. The perturbation is observed to evolve along the shear layer and then along the downstream boundary layer as a localized wave packet. We conclude that the considered ROM based on EOFs faithfully captures the original optimal perturbation dynamics, see figure 4.2. Notably, both the Orr and Kelvin-Helmholtz mechanisms are accurately recovered. This latter conclusion also holds for SOs and Balanced models.

4.5.3 Frequency response

As a second step, the performance of the models is evaluated in the frequency domain. We investigate their ability to capture the original resolvent $\mathsf{R}(\omega)$. To do so, we consider the optimal energy gain over all

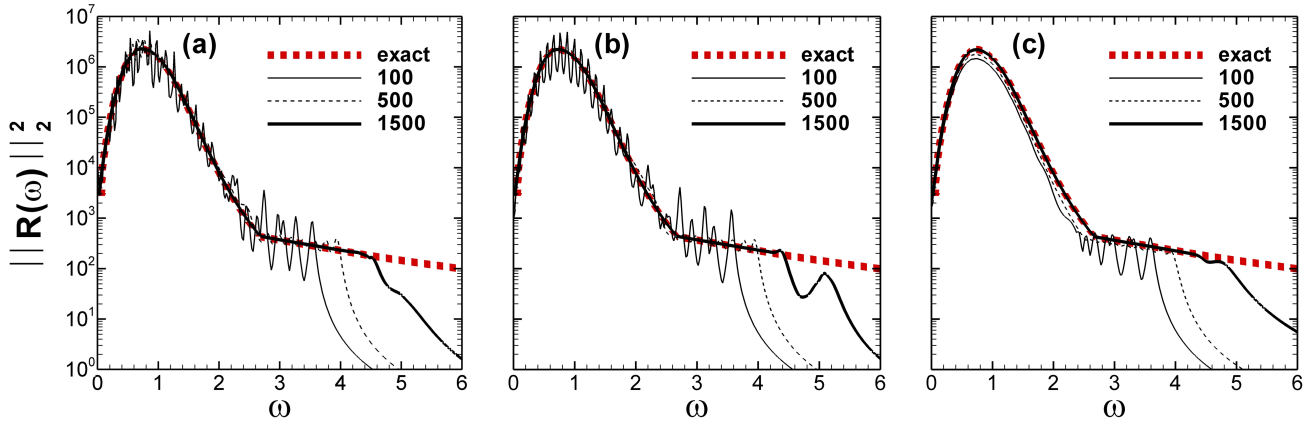


Figure 4.13: Optimal perturbation gain $||\mathsf{R}(\omega)||_2^2$ at frequency ω for ROMs based on (a) EOFs, (b) SOs and (c) Balanced modes.

possible harmonic forcing, namely $||\mathsf{R}(\omega)||_2^2$. This quantity is computed on the ROMs of size r by using the low-rank resolvent $\mathsf{T}(i\omega\mathsf{I}_r - \mathsf{A}_r)^{-1}\mathsf{S}^\dagger$ instead of $\mathsf{R}(\omega)$. The results corresponding to models based on EOFs, SOs and Balanced modes are reported in figures 4.13(a,b,c) respectively. Similarly, a convergence to the original optimal gain is observed for all three models. The first energy peak (associated with shear instabilities) is well captured for all ROMs even though the convergence is more erratic for EOFs and SOs models than for Balanced models. The high-frequency dynamics are also well captured for all ROMs until a cut-off frequency $\omega \approx 4$.

4.5.3.1 Error quantification

For the sake of completeness, the ability of the ROMs to capture the leading optimal energy gain is quantified. To this end, we introduce the relative ∞-norm of the error defined by

$$e_\infty(r) = \frac{\max_\omega |\lambda_1(\omega) - \lambda_{1r}(\omega)|}{\max_\omega |\lambda_1(\omega)|} \tag{4.36}$$

where $\lambda_1 = ||\mathsf{R}(\omega)||_2$ is the leading singular value of the resolvent and λ_{1r} represents the same quantity computed from the ROMs. Note that this error measures the relative difference of the optimal energy gain peak values. Figure 4.14 represents this quantity as a function of the size r of the models. We observe that the error tends to 0 for all three models. This fall off is somehow erratic for the EOFs and SOs models but more continuous and monotonic for the Balanced models. As reminded in section 4.4, exact Balanced Truncation is quasi-optimal to capture the input-output transfer function, see Moore (1981); Farrell & Ioannou (2001*a*). Since we consider here all the degrees of freedom of the full system as inputs and outputs, the input-output transfer function reduces to the resolvent operator. Accordingly, Balanced models of small size are much more accurate than those based on EOFs and SOs. Their error falls off quite rapidly until $r \approx 100$ which corresponds to the limit below which the Balanced modes are converged in terms of the approximation rank k, see figure 4.10(a). For models of higher size, the superiority of the Balanced models declines until being even inferior to the SOs models. This may be attributed to the difference between the exact higher Balanced modes and those approximated and used in the present models. A more precise understanding of the ranking performance between the three models is out of concern here.

These results show that all three models recover very accurately the optimal harmonic responses of the original system. For $r = 1500$, the error reduces to 2%, 0.6% and 0.55% for the EOFs, SOs and Balanced models respectively. Furthermore, it should be mentioned that the same analysis has been performed by considering the relative 2-norm error and that the same conclusions were observed.

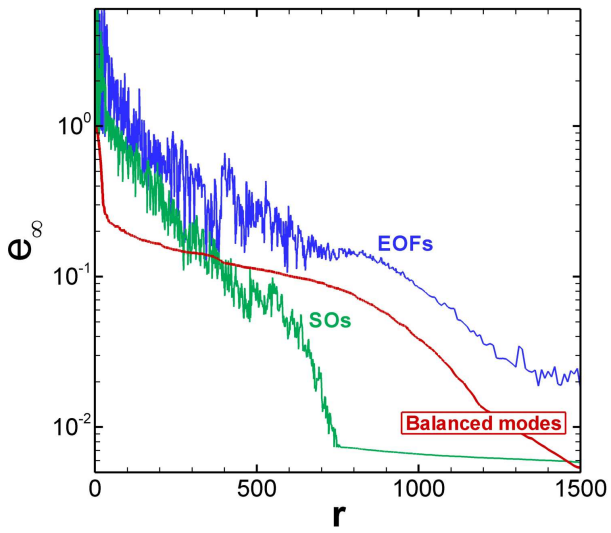


Figure 4.14: Relative ∞-norm error of the low-order models as a function of their size r.

4.5.3.2 Sub-optimal frequency responses

The most energetic part of the input-output dynamics has been highlighted in section 4.3 by ranking the leading optimal harmonic forcing/responses over a frequency interval. To further assess the dynamics of the ROMs, the same analysis is performed by investigating how their leading 40 optimal energy gains compare to the original ones. Note that the optimal energy gains of the ROMs are directly computed as the singular values of the low-order resolvent operator.
Results are exposed here in the case of EOFs models. The leading optimal energy gains $\lambda_1^2, \ldots, \lambda_{40}^2$ are depicted in figure 4.15 as a function of frequency. The results from EOFs models of size $r = 500$, 1000 and 1500 are represented in figures 4.15(a,b,c) respectively. The singular values computed from EOFs models are observed to converge to the exact ones when increasing the size of the models. Furthermore, this result also holds of the associated optimal forcing and responses (not shown here). Not only the "shear layer" branches are captured by the models but also the "convective" ones, see section 4.3. All the leading singular values are well captured until the cut-off frequency $\omega \approx 4$ which is also the limit considered to approximate the EOFs. Furthermore, it should be emphasized that we arbitrarily chose EOFs to expose these results since investigating SOs and Balanced models leads to the same conclusions. Our results demonstrate the ability of the ROMs to recover the input-output dynamics on the considered frequency interval. Interestingly, the high-frequency dynamics which are not captured by the ROMs reduce to the convection of small-scale structures in the free-stream. As a result, the ROMs are expected to accurately capture the exact flow response from all possible low-frequency forcing.

4.6 Conclusion

The two-dimensional incompressible flow over a rounded backward facing-step has been considered as a prototype of open flow subject to hydrodynamic instabilities. The main objective of this work was to design accurate low-order models of its input-output dynamics. To reach that goal, the dynamics has been decomposed into a set of optimal harmonic forcing and responses. The input-output dynamics, seen as the mapping between all possible forcing and their associated responses, has been truncated on a given frequency interval. Next, three sets of modes have been computed to project the original equations and design the ROMs: the leading EOFs, SOs and Balanced modes.
For the considered backward-facing step flow, the singular value decomposition of the resolvent has clearly highlighted two distinct dynamics. The first ones represent the amplification and advection of

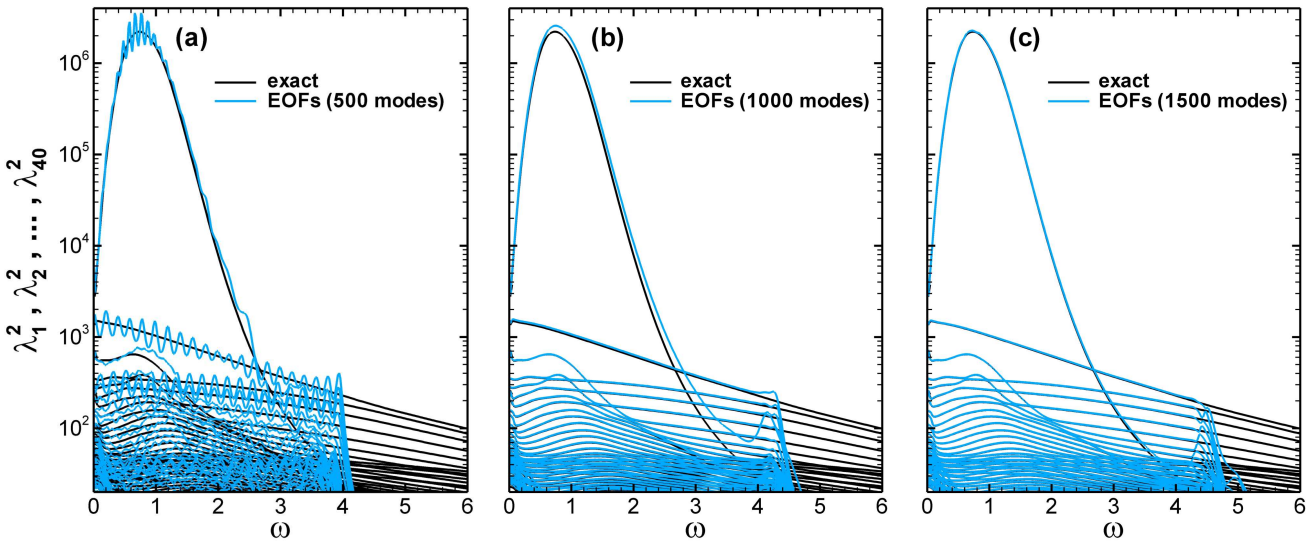


Figure 4.15: First 40 optimal energy gains $\lambda_1^2, \ldots, \lambda_{40}^2$ versus frequency ω computed on the full system and EOFs models. (a,b,c) represent the results from models based on 500, 1000 and 1500 modes respectively.

perturbations as wave packets along the separated shear layer and the downstream boundary layer. The second ones account for the free-stream convection of perturbations. One should expect that the balance between these two dynamics depends on the size of the computational domain. Actually, preliminary results performed on a bigger domain indicate that the free-stream convection dynamics are more energetic while the shear layer ones remain almost the same. In such a case, a larger number of optimal forcing/responses would be required to capture the predominant input-output dynamics. We thus conclude that designing accurate ROMs may be more challenging when considering much bigger computational domains.

A truncation of the leading 40 optimal forcing and responses on the frequency range $\omega = [0; 4]$ has been adopted to compute the projection bases. This choice proved to be sufficient for an accurate computation of the leading 100 EOFs, SOs and Balanced modes. Nevertheless, the higher modes have been also considered to build the ROMs. We showed that all three models recover the optimal perturbation of the system (the leading singular value of the time propagator) and the leading 40 optimal harmonic forcing/responses (the leading singular values of the resolvent). The original system of size $n = 170\,260$ has thus been reduced to models of size 1500 that accurately capture the dynamics.

Practically, such ROMs are not based on any input or output. Instead, the methodology introduced here is conceptually meant to capture the dynamics on a given frequency interval. In the case where the most energetic dynamics are predominantly driven by some known instability at a given frequency, as in the present study, the ROMs may capture the predominant input-output dynamics. In other words, they may recover the original flow response from all the possible initial conditions or external excitations. This work thus yields a new and promising contribution to the design of ROMs for fluid flows. It may be a valuable asset in the design of Kalman filter estimators for large-scale flows where the initial conditions are often uncertain or varying. Alternatively, in the context of flow control, such models would allow to consider systems where the number of inputs/outputs is very large or unknown. In particular, in the context of closed-loop flow control, one possible prospect would be to consider a large number of actuators and sensors in order to identify their optimal position.

Chapter appendix

4.A Formulating the Navier-Stokes equations as a standard dynamical system

The linear flow dynamics with external forcing have been assumed to be governed by the following system

$$\dot{\mathsf{x}} = \mathsf{A}\,\mathsf{x} + \mathsf{f} \tag{4.37}$$

where x is the discretized velocity field. This appendix is devoted to showing how to get this formulation. We start by considering the linearized Navier-Stokes equations (4.1) with a forcing term on the momentum equations. Once discretized, these equations may be written in the following finite-sized matrix form:

$$\begin{pmatrix} \mathsf{Q}_1 & 0 \\ 0 & 0 \end{pmatrix} \frac{d}{dt} \begin{pmatrix} \mathsf{x}_1 \\ \mathsf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathsf{A}_1 & \mathsf{A}_2^* \\ \mathsf{A}_2 & 0 \end{pmatrix} \begin{pmatrix} \mathsf{x}_1 \\ \mathsf{x}_2 \end{pmatrix} + \begin{pmatrix} \mathsf{Q}_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathsf{f}_1(t) \\ 0 \end{pmatrix} \tag{4.38}$$

where x_1 and x_2 denote the velocity and pressure fields, and $\mathsf{f}_1(t)$ is the forcing term. The operator A_1 has been introduced as the part of the dynamics governing the convection and diffusion. Analogously, A_2 and A_2^* have been introduced as the parts relative to the incompressibility and pressure effects respectively. The mass matrix Q_1 accounts for the inner product associated with the kinetic energy of perturbations, namely $\mathsf{x}_1^*\mathsf{Q}_1\mathsf{x}_1 = \int_\Omega \boldsymbol{u}^2\, d\Omega$ where Ω is the fluid volume.
The equation on the pressure is meant to enforce incompressibility on the velocity field. We thus eliminate it and restrict the velocity field to its divergence-free part. First, we multiply the momentum equation by $\mathsf{A}_2\mathsf{Q}_1^{-1}$ which leads an expression for the pressure as a function of the velocity field

$$\mathsf{x}_2 = -\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_2^*\right)^{-1}\left[\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_1\right)\mathsf{x}_1 + \mathsf{A}_2\mathsf{f}_1(t)\right] \tag{4.39}$$

The explicit divergence constraint can then be eliminated by writing

$$\frac{d\mathsf{x}_1}{dt} = \mathsf{P}_1\mathsf{A}_1\mathsf{x}_1 + \mathsf{P}_1\mathsf{Q}_1\mathsf{f}_1(t) \tag{4.40}$$

where

$$\mathsf{P}_1\mathsf{Q}_1 = \mathsf{I} - \mathsf{Q}_1^{-1}\mathsf{A}_2^*\left(\mathsf{A}_2\mathsf{Q}_1^{-1}\mathsf{A}_2^*\right)^{-1}\mathsf{A}_2 \tag{4.41}$$

is the projection matrix onto the divergence-free space. It is noteworthy that P_1 is a Hermitian operator ($\mathsf{P}_1^* = \mathsf{P}_1$). We can then introduce the linearized Navier-Stokes operator as $\mathsf{A} = \mathsf{P}_1\mathsf{A}_1$ and the divergence-free part of the forcing as $\mathsf{f} = \mathsf{P}_1\mathsf{Q}_1\mathsf{f}_1$ so that we recover equation (4.37).

4.B Computation of the optimal harmonic forcing and responses

The starting point to approximate the leading EOFs, SOs and Balanced modes is to compute the leading optimal harmonic forcing and responses of the flow. This appendix shows how the choice of the inner product intervenes in their computation. We consider the forcing $\mathsf{f}_1(t)$, which may not be divergence-free, so that the system is written as

$$\dot{\mathsf{x}} = \mathsf{A}\,\mathsf{x} + \mathsf{P}_1\mathsf{Q}_1\mathsf{f}_1(t) \tag{4.42}$$

where the operator $\mathsf{P}_1\mathsf{Q}_1$ projects the forcing onto the divergence-free space. For an harmonic forcing $\mathsf{f}_1(t) = \hat{\mathsf{f}}e^{i\omega t}$, the harmonic long time response of the flow is given by $\mathsf{x}(t) = \hat{\mathsf{x}}e^{i\omega t}$ with $(i\omega\mathsf{I}-\mathsf{A})\hat{x} = \mathsf{P}_1\mathsf{Q}_1\hat{f}$ which leads to $\hat{x} = (i\omega\mathsf{I} - \mathsf{A})^{-1}\mathsf{P}_1\mathsf{Q}_1\hat{f}$. The resolvent operator is then defined by $\mathsf{R} = (i\omega\mathsf{I} - \mathsf{A})^{-1}\mathsf{P}_1\mathsf{Q}_1$.

Now, considering the inner product Q_1 based on the kinetic energy, the adjoint of the resolvent operator $\mathsf{R}^\dagger(\omega)$ is defined by

$$\mathsf{R}^\dagger(\omega) = \mathsf{Q}_1^{-1}\mathsf{R}^*(\omega)\mathsf{Q}_1 \tag{4.43}$$

where the superscript * denotes the complex conjugate transpose. We showed that the optimal harmonic forcing $\hat{\mathsf{f}}_i$ and responses $\hat{\mathsf{x}}_i$ are obtained from $\mathsf{R}^\dagger(\omega)\mathsf{R}(\omega)\hat{\mathsf{f}}_i = \lambda_i^2\hat{\mathsf{f}}_i$ and $\mathsf{R}(\omega)\mathsf{R}^\dagger(\omega)\hat{\mathsf{x}}_i = \lambda_i^2\hat{\mathsf{x}}_i$. As a result, making explicit the expression of the adjoint operator leads to the following equations

$$\mathsf{R}^*(\omega)\mathsf{Q}_1\mathsf{R}(\omega)\hat{\mathsf{f}}_i = \lambda_i^2\mathsf{Q}_1\hat{\mathsf{f}}_i \quad \text{and} \quad \mathsf{R}(\omega)\mathsf{Q}_1^{-1}\mathsf{R}^*(\omega)\mathsf{Q}_1\hat{\mathsf{x}}_i = \lambda_i^2\hat{\mathsf{x}}_i \tag{4.44}$$

Practically, we only solve the first one with a classical Krylov-Arnoldi algorithm to compute the leading forcing terms $\hat{\mathsf{f}}_i$. The associated responses can then be recovered in a simpler way by $\hat{\mathsf{x}}_i = \mathsf{R}(\omega)\hat{\mathsf{f}}_i$.

Interlude

The previous chapter concludes the part of the thesis devoted to Reduced Order Models (ROMs). In this one, the resolvent operator has been introduced as a key element in both the analysis of the flow dynamics and the computation of the projection bases. We showed the ability of the Empirical Orthogonal functions (EOFs), the Stochastic Optimals (SOs) and balanced modes to recover the original flow dynamics. In fact, the common point between the three previous chapters is the shift into the frequency domain that allows to compute frequential snapshots which are used in the model reduction process. Over the three previous chapters, the backward-facing step flow has been investigated from an input-output point of view. We considered the case of a single input and a single output in chapter 2. The case of all possible inputs and the same output has been investigated in chapter 3. In the end, we have shown how to consider all the state-space as both the inputs and the outputs in chapter 4. As an assessment from these chapters, we can say that we have increased the requirement of the models, namely the number of inputs and outputs. Yet, it is naturally observed that the size of the required ROMs increases accordingly. Let us give an illustrative example. Suppose we look for a model with more than 95% of accuracy on the input-output transfer function with the H_∞ norm. If we consider the balanced models computed in chapters 2, 3 and 4, the required sizes of the ROMs are 11, 16 and 930 respectively. Now, one should naturally choose a model reduction among those presented in this thesis in relation to the desired expectations.

We now move on to the final chapter of this thesis. Its purpose is to illustrate some of the possible applications of the low-order models investigated so far. Here, we considered the design of a low-order closed-loop flow controller. We considered the two-dimensional backward-facing step flow along with an additional upstream noise distribution (input) which acts as a random forcing and induces flow unsteadiness. Next, we have introduced an actuator (input) along with several sensors (outputs). The idea is then to extract information from the flow via a measurement in order to provide an accurate real-time feedback actuation and suppress the perturbations induced by the noise. To solve this problem, we adopted an optimal state-space flow control technique: the Linear Quadratic Gaussian control (LQG). It consists in computing a control law based on a Linear description of the system in order to minimize a Quadratic cost functional when the system is driven by Gaussian noise.

In this context, the computation of a Reduced Order Model appears as an indispensable step in the design of the feedback flow controller. Indeed, the first step of the LQG control is to compute a so-called compensator which takes the measurement as an input and returns the control law governing the actuation, see Burl (1999); Zhou *et al.* (2002). To compute the compensator, one need to solve two Riccati equations which have the size of the linear input-output system of interest. In our case, the number of degrees of freedom is about $O(10^5)$ which makes the resolution of the Riccati equations impossible. Now, given the distribution of the inputs and outputs, the control design process amounts to identifying the input signals when output signals are given. Consequently, the design of the compensator does not require the complete knowledge of the full dynamics but only the part governing the relationship between input and output signals. Consequently, the compensator may be designed based on a low-order model which captures the original input-output dynamics. Mostly, the predominant reason to resort to ROMs is that a low-order compensator is more likely to yield an actuation law in real-time for

practical applications. For these reasons, model reduction appears as a key element to design effective compensators.

The next chapter presents the feedback control of the two-dimensional backward-facing step flow by means of an LQG compensator based on a reduced-order model. This work has been done in collaboration with Dr. Alexandre Barbagallo and Pr. Peter J. Schmid. Mostly, it should be noticed that my contribution concerns the dynamics of the non-reduced flow dynamics as well as the model reduction. On the other hand, the analyses related to the flow control were handled by Alexandre Barbagallo. The results of this work have been exposed in an article that has been submitted to the *Journal of Fluid Mechanics*. This latter consists of a comprehensive study of various components of the LQG-control design applied to the backward-facing step flow. The major contribution of this paper is the introduction of tools intended to design effective closed-loop control strategies for amplifier flows. Since a large part of this work was performed by Alexandre Barbagallo, we chose to report this article in the appendix section. Consequently, the next chapter is very short and only includes a brief description and summary of the main results. I would highly recommend interested readers to consider both the following summary and the full article presented in the appendix section.

Chapter 5

Closed-loop control of unsteadiness over the backward-facing step

5.1 Introduction: the LQG control

This last chapter of the thesis is devoted to the feedback control of the two-dimensional backward-facing step flow. We considered a linear state-space flow control approach: the LQG control. LQG control means Linear Quadratic Gaussian control. It stands for computing a control law based on a (L)inear description of the system such as to minimize a given (Q)uadratic cost functional, the system being forced by a (G)aussian noise. In this context, the objective of the control is to suppress or at least to reduce the perturbations induced by the noise. Since the Navier-Stokes equations are non-linear, the first step is to identify a linear state-space model representative of the system of interest. This model is called the plant. As a second step, we introduce an actuator to act on the flow (input) and a sensor to extract real time information from the flow (output). Next, the design of the LQG control consists of the computation of a compensator which takes the measurement as an input and yields the actuation law as an output. For illustration, a sketch of the key element of the compensated system is depicted in figure 5.1. We define the flow state $\mathsf{X}(t)$ whose dynamics are governed by the plant, with the initial condition $\mathsf{X}(t_0)$. We furthermore assume that the plant is subject to an external random noise. The sensor yields a measurement $m(t)$ of the flow state $\mathsf{X}(t)$ which is given to the compensator. This latter then compute the control law $u(t)$ (i.e. the temporal behavior of the actuator) such as to minimize a given cost functional.

The compensator is composed of two distinct components: the estimator and the controller. The controller provides to the actuator a control law equal to $u(t) = \mathsf{K}\mathsf{X}(t)$, namely the multiplication of the state $\mathsf{X}(t)$ by a matrix K called the control gain. It is shown in the standard reference, see Burl (1999), that the control gain matrix which allows to minimize a given quadratic cost functional may be obtained by solving a Riccati matrix equation. The controller requires the knowledge of the flow state $\mathsf{X}(t)$ which is not an available quantity in practical situations. This is where the estimator intervenes. Its purpose is to compute an estimated state $\mathsf{Y}(t)$ from the time history of the measurement. This estimated state is then used by the controller instead of $\mathsf{X}(t)$. Similarly, the estimator may be computed by solving a Riccati matrix equation. We recall that the computations of both the controller and the estimator are based on a linear description of the system of interest, namely the plant. Additionally , both can be designed independently owing to the separation principle, see Burl (1999). We refer interested readers to the book by Zhou *et al.* (2002) for additional information on optimal state-space flow control.

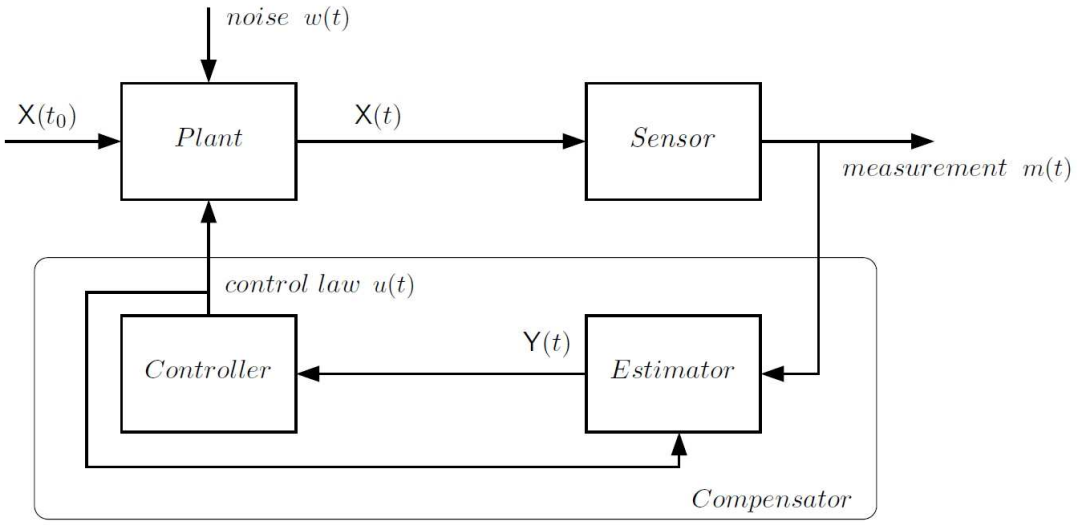


Figure 5.1: Sketch of a typical Linear Quadratic Gaussian control setup.

5.2 Flow configuration

We consider the two-dimensional incompressible flow over the backward-facing step investigated in the previous chapters. The Reynolds number, based on the step height and incoming velocity is fixed to $Re = 600$. Since we are attempting to control the flow by a LQG strategy, a linear model of the flow dynamics is required. We naturally choose to linearize the Navier-Stokes equations about a base flow (chosen as a steady solution). This stands for studying the early stages of development of the perturbations to be controlled since the nonlinear effects are negligible for small perturbations. We furthermore consider only two-dimensional perturbations. The base flow solution at $Re = 600$ has been represented in figure 5.2 by its streamlines.
As shown in chapter 1, the flow is globally stable at this Reynolds number, however, it exhibits a substantial ability to transiently amplify perturbations. As a results, the objective of the control is to reduce the perturbations possibly induced by an external noise which excites the flow. We assume that the flow unsteadiness (or perturbations) are triggered by a noise source modeled as a gaussian vertical body force B_1, see figure 5.2. The actuator is also chosen as gaussian vertical body force and is located in the vicinity of the base flow separation, see B_2 in figure 5.2. The sensor C_p yields a measure of the shear-stress integrated in a part of the downstream wall; it is the same sensor as the one considered in chapter 2 and chapter 3. The measurement from the sensor C_p will be a quantity to be minimized by the controller. Additionally, equivalent shear-stress sensors denoted as C_1, C_2, C_3 and C_4, see figure 5.2, will be considered to yield the measurement used by the estimator. The linearized Navier-Stokes equations are discretized with the finite element method; the resulting linear input-output finite system reads

$$\begin{aligned} \mathsf{Q}\frac{d\mathsf{X}}{dt} &= \mathsf{A}\mathsf{X} + \mathsf{Q}\mathsf{B}_1 w(t) + \mathsf{Q}\mathsf{B}_2 u(t) \\ m(t) &= \mathsf{C}_i\mathsf{X} \quad \text{with} \quad i = 1, 2, 3 \text{ or } 4 \end{aligned} \tag{5.1}$$

where the state vector $\mathsf{X}(t)$ is composed of the longitudinal and vertical velocity and the pressure on all the degrees of freedom of the mesh. The matrices Q, A, B_1, B_2 and C_i denote respectively a weight matrix arising from the discretization, the linearized Navier-Stokes operator, the noise matrix, the control matrix and the measurement matrices. Furthermore, we have introduced the control law $u(t)$, the measurement $m(t)$ as well as the noise $w(t)$. For simplicity, the noise is assumed to be white in time with zero mean and a variance equal to the parameter W^2. Furthermore, we suppose that the measurements C_i used by the estimator are corrupted with a noise $g(t)$ of variance G^2.

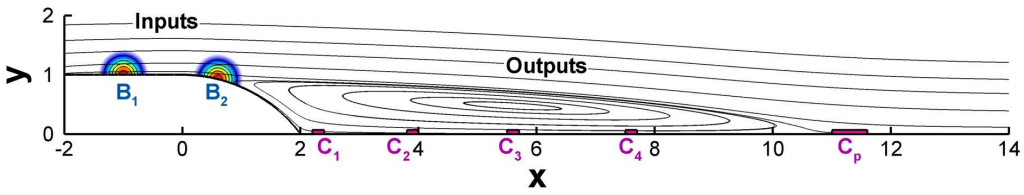


Figure 5.2: Streamlines of the base flow solution at $Re = 600$. The location of the inputs and outputs are also displayed.

5.3 Reduced-order compensator

As previously mentioned, the design of the controller and estimator is not tractable for very large systems such as the one considered here. The resolution of the two Riccati equations involved in the process is only tractable for systems of moderate dimensions ($\approx 10^3$). Here, the spatial discretization leads to a system with $\approx 360\,000$ degrees of freedom so that a reduced-order model of the flow becomes mandatory for the computation of the compensator. The reduced-order model is only required to capture the input-output dynamics. There are only two inputs: the noise B_1 and the actuator B_2. On the other hand, since the perturbation kinetic energy will be considered as a cost functional in the following, we wish to capture the entire flow response driven by these two inputs. At this stage of the thesis, we had not investigated all the model reduction techniques presented in the previous chapters. Therefore, we adopted a simple POD model. To be more precise, it reduces to the same POD model investigated in chapter 2, except that there are two inputs. In other words, we compute the set of the modes the most easily influenced by the two inputs; the most controllable modes. The original input-output system is projected onto the reduced POD basis to obtain the following low-order system.

$$\begin{aligned} \frac{d\hat{\mathsf{X}}}{dt} &= \hat{\mathsf{A}}\hat{\mathsf{X}} + \hat{\mathsf{B}}_1 w(t) + \hat{\mathsf{B}}_2 u(t) \\ m(t) &= \hat{\mathsf{C}}_i\hat{\mathsf{X}} \quad \text{with} \quad i = 1, 2, 3 \text{ or } 4 \end{aligned} \tag{5.2}$$

where the superscript $\hat{}$ represents the reduced order quantities (projected onto the reduced basis of modes). It has been found that a model of size 150 is sufficient to recover the entire dynamics driven by the two inputs. Naturally, using more advanced methods, such as the Balanced Truncation along with the output projection procedure, would have resulted in a much smaller ROM. Yet, building the smallest model is not the point of this work; we only need an accurate low-order representation of the input-output dynamics to investigate the feedback control performance.

Let us introduce the reduced-order estimator. Its task is to construct an approximate flow state $\hat{\mathsf{X}}$ from partial information from the sensor $\hat{\mathsf{C}}_i$. The estimated state $\hat{\mathsf{Y}}$ is supposed to satisfy a set of equations similar to those governing the original system:

$$\frac{d\hat{\mathsf{Y}}}{dt} = \hat{\mathsf{A}}\hat{\mathsf{Y}} + \hat{\mathsf{B}}_2 u(t) - \hat{\mathsf{L}}(m - \hat{\mathsf{C}}_i\hat{\mathsf{Y}}) \tag{5.3}$$

where the noise term $\hat{\mathsf{B}}_1 w(t)$ has been replaced by the forcing term $-\hat{\mathsf{L}}(m - \hat{\mathsf{C}}\hat{\mathsf{Y}})$. This forcing term represents the difference between the true measurement signal $\hat{\mathsf{C}}_i\hat{\mathsf{X}}$ and the estimated measurement signal $\hat{\mathsf{C}}_i\hat{\mathsf{Y}}$ multiplied by the matrix $\hat{\mathsf{L}}$. In fact, this forcing term is meant to drive the estimated state $\hat{\mathsf{Y}}$ toward the true one $\hat{\mathsf{X}}$. The design of the estimator consists of the computation of the matrix $\hat{\mathsf{L}}$ called the Kalman gain which is performed from a constrained optimization problem in which the estimation error $\hat{\mathsf{Z}} = \hat{\mathsf{X}} - \hat{\mathsf{Y}}$ is to be minimized. Practically, $\hat{\mathsf{L}}$ may be computed by solving a Riccati matrix equation, see Burl (1999). Importantly, the two white noise source $g(t)$ and $w(t)$ are taken into account in this Riccati equation. Indeed, the ratio of their variance G/W appears as a parameter in the design of the estimator

that accounts for the noise-to-signal ratio of the sensor, or equivalently to the speed of estimation. It should be noticed that the Riccati equation may be easily solved for our system of size 150.
Regarding the reduced-order controller, it is designed to compute the control law $u(t)$ so as to minimize a predefined cost functional such that $u(t) = \hat{\mathsf{K}}\hat{\mathsf{X}}(t)$. Two cost functional are considered in this work. The first one represents the energy of the measurement extracted from the sensor C_p which may be written as

$$J_m = \int_0^\infty (\hat{\mathsf{X}}^*\hat{\mathsf{C}}_p^*\hat{\mathsf{C}}_p\hat{\mathsf{X}} + l^2u^2)dt \tag{5.4}$$

and the second represents the perturbation kinetic energy contained in the entire domain:

$$J_e = \int_0^\infty (\hat{\mathsf{X}}^*\hat{\mathsf{Q}}\hat{\mathsf{X}} + l^2u^2)dt \tag{5.5}$$

where * denotes the transconjugate, $\hat{\mathsf{C}}_p\hat{\mathsf{X}}$ is the measured signal and $\hat{\mathsf{X}}^*\hat{\mathsf{Q}}\hat{\mathsf{X}}$ is the perturbation energy computed in the ROM. Note that the parameter l has been introduced in either case. It is referred to as the control cost and quantifies the user-specified weighting of the control energy compared to the energy to be minimized. The design of the controller reduces to the computation of the reduced control gain $\hat{\mathsf{K}}$. Analogously, the matrix $\hat{\mathsf{K}}$ is obtained by solving a Riccati matrix equation, see Burl (1999).
In the end, the reduced-order compensator consists of the successive action of the reduced-order estimator and the reduced-order controller. Once we have chosen the values of l and G/W, we compute the associated matrices $\hat{\mathsf{K}}$ and $\hat{\mathsf{L}}$. Next, the measure $m(t)$ is used to build the estimated state $\hat{\mathsf{Y}}(t)$ from which the controller provides the control law $u(t) = \hat{\mathsf{K}}\hat{\mathsf{Y}}(t)$.

5.4 Summary of the results

The main concern of this work is to give a physical insight into the feedback control of the backward-facing step flow. To some extent, the idea is to address practical questions in view of getting closer to an experimental implementation. The results of this work have been organized in four quasi-independent sections that are summarized below. We refer interested readers to the full article exposed in appendix C where a thorough report of all the results is available.

5.4.1 Part I: Design of an efficient estimator

The first part of this work is devoted to the estimation process and more particularly to the placement of the sensor. As a result, the controller is turned off; we wish to construct an accurate representation of the state triggered by the noise $\hat{\mathsf{B}}_1$. We investigate the estimation performance when choosing the four different sensors $\hat{\mathsf{C}}_i$ and varying the noise-to-signal ratio G/W.
Choosing large values of G/W simulates the case where the signal received by the sensor is highly corrupted. In such a case, it is naturally observed that the estimator is unable to yield a correct approximation of the flow state. On the other hand, for low values of G/W, the signal is weakly affected by the noise and an accurate estimation may be recovered. Recalling that the flow is subject to convective instabilities, the perturbation triggered by $\hat{\mathsf{B}}_1$ is amplified during its advection. For this reason, it is necessary to locate the estimation sensor in a sufficiently downstream position so as to measure a weakly corrupted signal. However, we show that the estimation error $\hat{\mathsf{Z}}(t)$ only start to fall off once the disturbance triggered by $\hat{\mathsf{B}}_1$ reaches the estimation sensor. This is illustrated in figure 5.3 where we represent the estimation error $\|\hat{\mathsf{Z}}\|_2^2 = \hat{\mathsf{Z}}^*\hat{\mathsf{Q}}\hat{\mathsf{Z}}$ as a function of time where we have triggered an impulse at the input $\hat{\mathsf{B}}_1$. We have also represented by a bold and grey line the energy of the original state $\|\hat{\mathsf{X}}\|_2^2 = \hat{\mathsf{X}}^*\hat{\mathsf{Q}}\hat{\mathsf{X}}$. It is observed that the error start to decrease as soon as the triggered wave packet reaches the estimators $\hat{\mathsf{C}}_i$. As a results, we conclude that the sensor used for the estimation should be

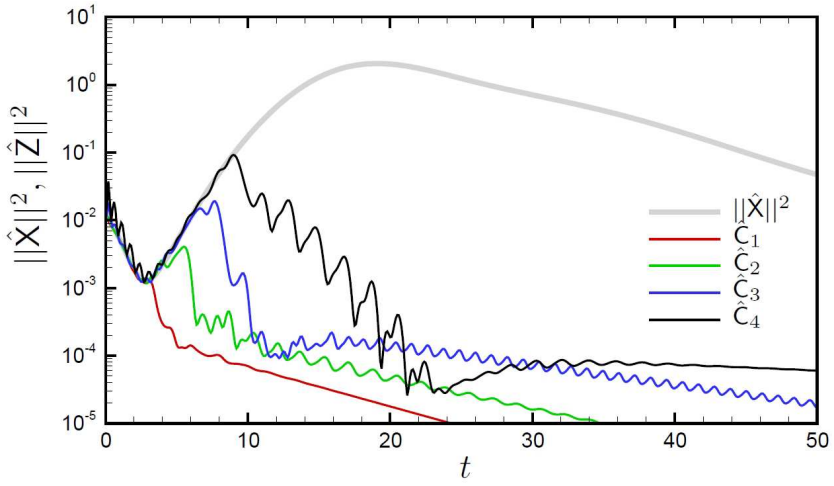


Figure 5.3: Energy of the estimation error $\|\hat{\mathsf{Z}}\|_2^2$ versus time for the four different estimation sensors. The energy of the original state $\|\hat{\mathsf{X}}\|_2^2$ is also represented by a bold grey line.

placed by considering a balance between the speed of estimation (where it should be located upstream) and the noise-to-signal ratio (where it should be located rather downstream).

5.4.2 Part II: Influence of the cost functional on the control performance

As a second step, the controller is turned on and we investigate the effect of the cost functional on the control performance. We start by considering the case where we attempt to minimize the energy measured by the sensor C_p; namely the cost functional J_m. It is shown that, for low values of the control cost l and noise ratio G/W (which is referred to as the large gain limit), we observe an increase of the total kinetic energy of the perturbation even though the energy measured from the sensor C_p is decreased. This observation has been interpreted to an extreme sensitivity of the compensators that are designed in the vicinity of the large gain limit. In such a case, we observed that the control law $u(t)$ provided by the compensator yields strong control actions at high frequencies. Next, it is shown that this behavior may be suppressed by choosing the cost functional J_e based on the total kinetic energy of the flow state. In this latter case, we succeeded in decreasing both the total kinetic energy of the flow state and the measurement energy from the sensor C_p.

5.4.3 Part III: Effect of the plant deviation on the compensated system

In the third part of this work we investigate how the performance of the feedback control is altered by the model reduction used to design the compensator. In other words, we study how an error between the original plant and the reduced-order model based on which the compensator is designed affects the control performance. It is shown that the most critical quantity to be captured by the reduced-order models is the input-output behavior between the actuator B_2 and the estimation sensor (chosen as C_1 here). As a result, the error between the original and reduced plant is quantified by the relative H_2 error of the impulse response from B_2 to C_1. This error, denoted by err, is then considered to account for the plant deviation.

As a first step, we choose a ROM of size 150 as the reference plant. Consequently, we can design a "perfect" compensator based on this plant which corresponds to the case where $err = 0$. Next, we artificially degrade the ROMs used to construct the compensators by decreasing their size. By this way, various compensators characterized by different plant deviations are investigated. The resulting compensated systems are observed to be unstable in the vicinity of the large gain limit. Furthermore, when increasing the plant deviation err, the unstable region in the $(l, G/W)$ plan grows so that the range of

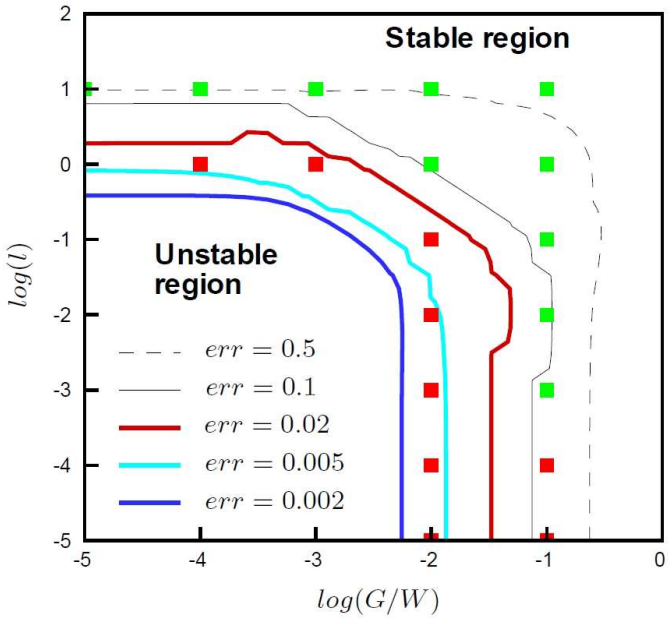


Figure 5.4: Regions of unstable compensated systems in the $(l, G/W)$ plan for different values of the plant deviation err. The red and green squares represent unstable and stable compensated systems respectively in the case where the plant consists of the full linearized Navier-Stokes equations.

the usable parameters l and G/W is reduced. This result is illustrated in figure 5.4 where the regions in the $(l, G/W)$ plan leading to unstable compensated systems have been reported. The five contour lines separating the unstable regions from the stable ones correspond to the errors $err = 0.5, 0.1, 0.02, 0.005$ and 0.002.

As a second step, the feedback control of the full linearized Navier-Stokes equations with a compensator based a ROM of size 150 has been investigated. The considered values of the parameters l and G/W have been reported in figure 5.4 by red and green symbols which denote the unstable and stable simulations respectively. Interestingly, the plant deviation reduces to $err = 0.02$ in this case. The unstable behavior of some of the compensated systems is thus in agreement with our results obtained with the low-order reference plant.
Finally, we conclude that the more the ROM differs from the plant to be controlled, the more restricted is the range of usable parameters l and G/W to design an effective compensator. This result is important since the model reduction, which is required for the practical design of a compensator, inevitably induces a loss of information and a deviation of the plant.

5.4.4 Part IV: Feedback control of the non-linear Navier-Stokes equations

The fourth and last part is devoted to the feedback control of the non-linear Navier-Stokes equations by means of a linear compensator. The variance of the random noise is varied such as to investigate various degrees of non-linearity. In particular, three cases are examined: a quasi-linear, a weakly non-linear and a strongly non-linear development of the perturbations.
In the case where the dynamics of the perturbation are quasi-linear, the control is found to be very effective. However, its efficiency is observed to decease as non-linear effects become stronger. In the weakly non-linear situation, the dynamics of the perturbation is still linear close to the sensor used for estimation but becomes non-linear after the sensor. In this case, a good control effort is also observed. Now, in the strongly non-linear case, the non-linear effects are noticeable upstream of the sensor and a clear decrease of the control efficiency is observed. The effect of the control in the strongly non-linear

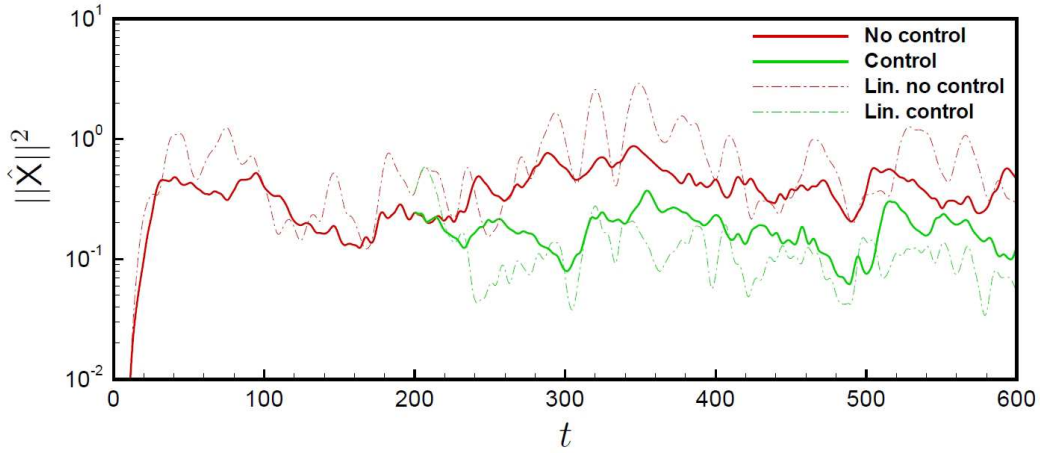


Figure 5.5: Energy of the perturbation versus time for the strongly non-linear simulations (thick lines) and for the linear simulations (dashed lines). Note that the same random sequence has been used for the noise signal in all the cases.

simulation is illustrated in figure 5.5 where the kinetic energy of the perturbation is depicted as a function of time. The thick lines account for the strongly non-linear simulations whereas the dashed lines account for the linear simulations. Note that the same random noise has been used in all these simulations. Noticeably, a significant decrease of the perturbation kinetic energy is observed in both cases. For even stronger non-linear simulations, the compensated system is found to possibly diverge. In the end, one may remember that a linear compensator has a positive action on the non-linear flow dynamics for a significant range of perturbation amplitudes.

5.5 Concluding remarks

The two-dimensional backward-facing step flow has been successfully controlled by a feedback LQG control strategy. The use of a reduced-order model accounting for the linear dynamics of the flow has been used as a key element to approximate a low-order plant and to design a low-order compensator. Important issues related to the practical implementation of such control strategies have been addressed. Owing to this work and to the results exposed by Barbagallo *et al.* (2009), we conclude that the design of a closed-loop control set-up is more challenging in the case of an amplifier flow (such as the backward-facing step flow) than for an oscillator flow (such as the cavity flow). Indeed, oscillator flows display instabilities which are narrow-banded in frequency and thus more easily controllable and observable. On the other hand, noise-amplifier flows are very sensitive to external excitations and exhibit instabilities characterized by a rather broadband of frequencies. We demonstrated that a sensitivity analysis of the compensator performance with respect to various noise sources is a key element for a successful closed-loop control design. We hope that the present work constitutes an additional and valuable tool to overcome some of the practical implementation challenges that arise in the design of a closed-loop flow control. Finally, we refer interested readers to the appendix C where the full article is exposed.

Conclusion

The research presented in this thesis contributes to the study of the model reduction of the linear dynamics of fluid flows. Our main objective was to investigate how to design Reduced Order Models (ROMs) by projection methods that accurately capture the linear two-dimensional input-output dynamics of a backward-facing step flow. Our first concern was to investigate the linear dynamics of the flow by means of a global stability analysis. Next, various sets of modes were considered to construct ROMs which model the original dynamics. In particular, some of the most famous projection methods used to reduce the dynamics of fluid flows have been investigated: the projection on the basis of the global eigenmodes, the Proper Orthogonal Decomposition (POD) and the Balanced Truncation. Finally, the last chapter was devoted to a practical aerodynamic application where model reduction appeared as a key element: the LQG closed-loop control of the backward-facing step flow based on a reduced-order compensator.

Summary of the results

The linear and incompressible dynamics of the backward-facing step flow has been introduced in chapter 1. It has been found that the flow becomes globally unstable at $Re = 526$. The unstable global mode has been identified as a stationary and transversally periodic structure similar to those encountered in various different separated flows, see Theofilis *et al.* (2000); Barkley *et al.* (2002); Marquet *et al.* (2009); Cherubini *et al.* (2010*b*); Robinet (2007*a*). In particular, this instability has been attributed to both lift-up and centrifugal mechanisms owing to the previous work by Marquet *et al.* (2009) and Cherubini *et al.* (2010*b*). The transient dynamics of the flow has been investigated through an optimal perturbation analysis together with a receptivity analysis. We identified the disturbances which most effectively lead the flow to turbulence. The most unstable mechanisms were attributed to coupled Kelvin-Helmholtz and Orr instabilities which were shown to lead to substantial transient energy growths. Additionally, the optimal transient mechanism, leading to a maximum kinetic energy growth, has been related to the development of oblique waves in the shear layer, similarly to the optimal transition scenario observed by Monokrousos *et al.* (2010) in a flat-plate boundary-layer. This chapter thus provides a rather thorough analysis of the linear global stability of a $2D$ backward-facing step flow where we used some of the most recent hydrodynamic stability tools.
The use of the global eigenmodes to capture the transient dynamics of the flow has been briefly addressed in appendix B. We showed that their ability to capture the optimal perturbations strongly depends on the flow configuration. Mostly, computational limitations were attributed to the very bad conditioning of the linearized Navier-Stokes operator owing to the strong streamwise non-normality of the equations. These results are in agreement with those recently exposed by Barbagallo *et al.* (2009); Sipp *et al.* (2010).

An input-output formalism of the equations was then privileged: we looked for low-order models which capture the dynamics from some inputs to some outputs. To begin with, a single input and a single output where considered in chapter 2. We showed how to compute the most controllable modes (the structures the most easily influenced by the input) by a POD procedure. Alternatively, the most and equally controllable and observable modes, referred to as the balanced modes have also been considered. We showed that both bases lead to ROMs that accurately model the original input-output dynamics

with a superiority of the balanced models. These results are in agreement with the previous works by Rowley (2005); Ilak & Rowley (2008); Barbagallo *et al.* (2009). The major contribution of this chapter is the introduction of a frequency-based snapshots method used to compute both the most controllable modes (POD modes) and the balanced modes. This approach offers the possibility to design ROMs from the flow state responses to harmonic excitations. This alternative technique is shown to display two main advantages. First, frequential information of given flow responses may be a much more accessible to experimental measurements and/or to a physical interpretation. Secondly, this procedure is shown, see Zhou *et al.* (1999), to be directly applicable to unstable systems without separating the stable and unstable dynamics. To illustrate this valuable property, the procedure has been derived successfully on both the backward-facing step flow and the unstable flow over a square cavity.

In the next part of the thesis, our frequency-based snapshot technique was extended to flow configurations having more than a moderate number of inputs and outputs. We considered in chapter 3 the modeling of the dynamics from all the possible inputs to a single output. The most observable modes (the structures which optimally excites the output) have been computed by a POD procedure by using the frequential snapshot method. Projecting the linearized Navier-Stokes equations on this basis proved to lead to efficient ROMs. Yet, the major contribution of this chapter was the introduction of the input-projection technique which allows to approximate balanced modes. Indeed, these balanced modes have been used to design more efficient ROMs. Noticeably, this new procedure stems from the works by Rowley (2005); Ilak & Rowley (2008) where an analogous technique called output-projection has been exposed. We illustrated the superiority and the quasi-optimality of the balanced models obtained with the input projection procedure on the backward-facing step flow. The inputs were chosen as all the degrees of freedom of the state space discretization and the output as a single sensor. By construction, the ROMs presented in this chapter may be particularly useful for flows which are continuously excited by an unknown noise (input) and where there are a moderate number of outputs.

In chapter 4, a consistent framework has been developed in order to extend the model reduction to fluid systems with a very large number of inputs and outputs. Low-order models that capture the dynamics from all the possible inputs to all the possible outputs have been computed on the backward-facing step flow configuration. A major emphasis has been given to the formalism and methodology which allows to derive the projection bases and the resulting ROMs. We followed through the previous work by Farrell & Ioannou ($2001a$) where the authors constructed ROMs by using the most controllable modes (called Empirical Orthogonal Functions), the most observable modes (called Stochastic Optimals) as well as balanced modes. In the case of our $2D$ backward-facing step flow, the computation of these modes was not tractable due to the large size of the system. Hence, the major contribution of this chapter is to propose a way of approximating the leading EOFs, SOs and balanced modes for large fluid systems. The key idea was to approximate the resolvent operator by truncating its singular value decomposition. The leading singular vectors of the resolvent have been computed on a finite frequency range and used to approximate the leading EOFs, SOs and balanced modes. Next, the resulting projection bases have been used to design ROMs which were shown to successfully capture the predominant part of the original flow dynamics.

Mostly, the procedure used in chapter 4 proved to be useful in view of analyzing the linear dynamics of amplifier flows. Indeed, the singular value decomposition of the resolvent operator allowed to decompose the global linear dynamics of the flow by ranking the leading optimal harmonic forcing and their associated optimal response. Noticeably, owing to this decomposition, we identified two separate dynamics: (i) the shear layer dynamics which are characterized by broadband frequency responses and an amplification peak at the Kelvin-Helmholtz frequency and (ii) the dynamics related to the convection of the perturbations in the free-stream. In addition, as exposed in Farrell & Ioannou ($1993a$), both the EOFs and the SOs have a significant physical meaning. Assuming the flow is excited by a white in space and time random noise, the EOFs would represent the structures that optimally account for the variance sustained by the statistically steady flow response. Interestingly, the EOFs also reduce to the POD modes used to identify the coherent structures in turbulent flows, see Lumley (1970); Aubry *et al.*

(1988); Berkooz *et al.* (1993). Similarly, the SOs stand for the forcing structures yielding a maximum contribution to the variance of the resulting statistically steady flow response. As a result, we expect that both may be used to rank the leading energetic coherent processes of a flow for post-processing purposes.

Chapter 5 has put emphasis on the question of the closed-loop control of the backward-facing step flow. A Linear Quadratic Gaussian control based on a reduced-order compensator has been adopted. The same model reduction procedure and LQG control have been performed recently by Bagheri *et al.* (2009*b*) on a flat-plate boundary layer flow. Yet, our contribution relies on the analysis and discussion on some issues that inevitably arise in the practical implementation of the compensator for amplifier flows. We considered a POD model of size 150 to design the compensator. This latter was observed to successfully control the flow by reducing the perturbations amplitude. Owing to the noise amplifier behavior of the flow, sensitivity appeared as a key concept in the design and performance evaluation of the compensators. For that reason, we investigated the behavior of various compensated systems. In particular, the effects of the sensor placement, of the cost functional, of the plant deviation and of the nonlinearities were investigated. Last but not least, we successfully controlled the fully non-linear Navier-Stokes equations by using a linear low-order compensator and obtained a significant reduction of the perturbations kinetic energy.

Suggestions for future works

This thesis provides physical and methodological insights into the linear dynamics, model reduction and closed-loop control over a backward-facing step flow. A special care was taken to remain as close as possible to practical applications. However, with the current available numerical tools, the direct implementation of the methods exposed in this dissertation is still premature for highly complex flows. Meanwhile, these methods are expected to be extended in future works to deal with more complicated flow configurations such as those encountered in industrial applications.

All the flow configurations investigated in this thesis were two-dimensional since the base flows were supposed to be non-homogeneous in only two directions. Yet, the extension of the concepts introduced here (stability analysis, model reduction and LQG control) may be naturally extended to three-dimensional base flows. In fact, the critical limitations are only computational ones; one may encounter either memory requirements or CPU time issues. In this thesis, we adopted numerical methods based on the direct inversion of large and sparse matrices. For instance, these inversions intervened in the computation of the base flow, of the snapshots, of the global eigenmodes or of the optimal forcings and responses. These were handled by a sparse scalable direct lower-upper (LU) solver, see Amestoy *et al.* (2001). Preliminary results on three-dimensional flow configurations have been obtained by using the same numerical methods and the same finite element solver. In such a case, an illustration of the computational requirements on $3D$ configurations may be found in the article by Sipp *et al.* (2010). The computations remain relatively short in time but require a much larger amount of memory. Noticeably, the sparsity of the matrices are different for $3D$ settings: there are a significantly larger number of non zero coefficients per line as compared to a $2D$ setting with the same number of degrees of freedom. We expect that domain decomposition methods may improve the scalability of the large-scale linear problems when using parallel computation with a high number of processors.
Otherwise, a significant alternative is to resort to matrix-free methods in which the jacobian matrix (standing for the linearized Navier-Stokes operator) is never explicitly formed. In particular, time-stepping methods allow to perform linear stability analyses by using collecting flow state snapshots from linearized simulations. For instance, it has been shown in the literature how to compute the global modes or to perform an optimal growth analysis of a flow, see Tuckerman (1998); Tuckerman & Barkley (2000); Barkley *et al.* (2008). More recently, time-stepping methods have been used by Monokrousos *et al.* (2010); Bagheri *et al.* (2009*a*) in view of using only a linear or non-linear Navier-Stokes solver to

perform a global stability analysis. Such methods seem rather promising to handle very large or complex flow configurations. For instance, we refer interested readers to the global stability analysis of the $3D$ base flow over a spheroid by Tezuka & Suzuki (2006) or that of a jet in crossflow by Bagheri *et al.* (2009*d*). Equally, time-stepping methods may be considered to compute the time response from some inputs to design reduced order models. For example, one may refer to the work by Semeraro *et al.* (2011) where the authors designed a feedback control of a transitional flat-plate boundary layer by means of a reduced-order LQG compensator. In this case, they considered a $3D$ base flow with a row of actuators (inputs) and a row of sensors (outputs) and used a time-stepping method to construct the desired ROMs.

The dynamics investigated here have been linearized about a base flow chosen as the steady solution of the Navier-Stokes equations. However, when increasing the Reynolds number, the computation of the base flows becomes difficult. The size of the meshes required to capture all the flow structures may become prohibitively large. Mostly, for very high Reynolds number flows, their is no guarantee in general that a steady solution of the equations exists. On the one hand, globally stable flows at high Reynolds number may be so much sensitive that the accurate computation of a steady solution is not possible. On the other hand, globally unstable flows undergo one of numerous bifurcations when increasing the Reynolds number such as a steady solution of the equations can no longer be considered as a physically relevant base flow to linearize about.

When considering high Reynolds number and turbulent flows, an alternative approach deserves a particular attention. It consists of augmenting the Navier-Stokes equations by a turbulence model. The dynamics of the flow are then governed by the so-called URANS equations: Unsteady Reynolds Average Navier-Stokes equations. Such models are based on the assumption that the dynamics of the small and large scales are decoupled. More precisely, the small scales, characterized by high frequencies, are supposed to be accounted for by the turbulence model whereas the dynamics of the large scales, associated with low frequencies, are governed by the time integration of the URANS equations. By this way, it is possible to extend the concepts introduced in this thesis to turbulent flows by considering the URANS equations instead of the Navier-Stokes equations. The base flow may be naturally chosen as a steady solution of the URANS equations. The linear dynamics of perturbations about this base flow then stand for the early development of large-scale (and low-frequency) structures within the turbulent flow. We expect that URANS models allow to extend the model reduction techniques and LQG control investigated here to fully turbulent flows. Noticeably, a global stability analysis including the Spalart Allmaras turbulence model was performed by Crouch *et al.* (2007) to investigate the onset of transonic shock-buffeting on an airfoil. Additionally, the same strategy has been used more recently by Cossu *et al.* (2009) to identify the streaks in a turbulent boundary layer. Lastly, this approach has also been considered by Luchtenburg *et al.* (2009) to design reduced-order models that recover the dynamics of a turbulent flow. In this latter work, the authors computed URANS simulations with a k-ω turbulence model to construct a POD-Galerkin ROM that captures the effect of an high-frequency actuation on the mean flow.

Finally, we would like to suggest the use of different model reduction techniques for future works. Firstly, using system identification methods seems a rather promising way to extend our results to more realistic flow configurations. Indeed, the most observable modes or the balanced modes computed in our work are based on adjoint snapshots. Consequently, their associated ROMs cannot be computed in a real setting where adjoint information is not accessible. Meanwhile, the Eigensystem Realization Algorithm (ERA), which stands for a system identification method, has been shown by Ma *et al.* (2010) to possibly construct balanced models without resorting to adjoint information. This valuable property allows to compute the same models as those presented in this thesis while extending their range to experimental implementation and practical applications.

When using projection methods, the ROMs are meant to represent the dynamics of the data on which they have been designed. Now, in practical situations, a ROM is often desired to model various flow

configurations. For instance it may be required to model the different flight conditions of an airplane. Instead of increasing the size of a ROM to extend its effective range, a possible solution would be to adjust it in real time to the flow operating condition. Such an improvement is possible when using system identification methods by updating the coefficients of the ROM as soon as its performances start to fall off. By using these successive adjustments, it is even conceivable to update a linear ROM such as to effectively model the non-linear dynamics of a flow.

Lastly, we would like to mention some extensions concerning the model reduction of non-linear systems. In fact, the Proper Orthogonal Decomposition has been extensively used in the past few decades to model the non-linear dynamics of fluid flows, see Aubry *et al.* (1988); Noack *et al.* (2003); T. R. Smith & Holmes (2005). It simply consists of collecting the snapshots from nonlinear simulations, computing the POD modes and projecting the original equations onto them to obtain a nonlinear low-order model. Yet, it is generally acknowledged that such models capture the most energetic patterns of the flow instead of its dynamics and that they require considerable fine tuning. We expect that extending the Balanced Truncation procedure to non-linear systems might solve these problems. The Balanced Truncation of nonlinear systems has first been introduced by Scherpen (1993) and investigated later in extensive works, see Lall *et al.* (2002); Fujimoto & Tsubakino (2008). In particular, a recent article by Ilak *et al.* (2010) is worth mentioning. This latter assesses the model reduction of the complex non-linear Ginzburg-Landau equations which mimic the behavior of a convectively or globally unstable fluid flow. In this work, the authors proceeded by projecting the non-linear equations onto a set of balanced modes obtained from the linear part of the dynamical operator. Interestingly, they showed that the resulting ROMs outperform the standard non-linear POD model reduction. In other words, they demonstrated that the superiority of balanced models over POD ones extends to the case of a nonlinear system. These results are promising but the modeling of more realistic flow dynamics, governed by the Navier-Stokes equations for instance, would be of relevant interest.

Appendices

Appendix A

Numerical methods and tools

This appendix is devoted to the description of the numerical tools used throughout this thesis. It includes a presentation of:

(i) The finite element method used to spatially discretize the equations

(ii) The algorithm used to solve the time-integrated Navier-Stokes equations

(iii) the Newton method used to compute the base flows.

(iv) The Arnoldi algorithm used to solve the large generalized eigenvalue problems.

Some parts of this appendix are inspired from the recent theses by Marquet (2007) and Brion (2009) where the authors have used very similar numerical techniques.

A.1 The finite element method

All the Partial Differential Equations (PDE) encountered in the thesis have been solved by using a finite element method. It consists of two steps: (i) define the variational (or weak) formulations of the PDE to be solved and (ii) discretize in space these formulations. The space discretization has been performed by using the freely available software *FreeFem++* developed by Frédéric Hecht, Olivier Pironneau, Antoine Le Hyaric and Kohji Ohtsuka, see Hecht *et al.* (2005). For additional information on the finite element method, we refer interested readers to the book by Girault & Raviart (1986) or the review article by Glowinski & Pironneau (1992).

A.1.1 Spatial discretization

The first step of the space discretization is to define a mesh. Let us consider a computational domain Ω and its border denoted by Γ. We used the finite element mesh generator of *FreeFem++* which yield an automatic triangulation of the domain Ω. The triangulation yields triangle elements whose apices are called vertices. The number of triangles, namely the mesh refinement, is adjusted by fixing the number of apices on each edge of the domain.

A.1.2 Finite elements

Given the mesh, the flow fields (such as the velocity or the pressure) are discretized on a finite dimensional approximation of the state-space. The definition of the discretized state-spaces is not straightforward, see Girault & Raviart (1986). In particular, the variational formulations may be ill-posed if we consider

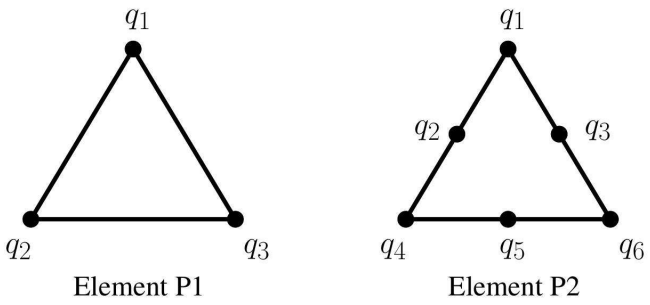


Figure A.1: Schematic of the elements $P2$ and $P1$ along with their nodes.

the same discretized state-space for the velocity and pressure fields. We thus adopted the *inf-sup* condition (or LBB condition), see Girault & Raviart (1986), which have been shown to be well-suited for generalized Stokes problems. Several state-spaces respect this condition, such as the $P1 - P0$ elements, see Fortin (1981) or also the Taylor-Hood elements $P(k+1) - Pk$ (with $k \geq 1$), see Hood & Taylor (1973). We choose the finite elements $P2 - P1$, namely the velocity fields are projected onto six-node quadratic triangular elements with quadratic interpolation ($P2$-elements) whereas the pressure field is discretized using three-node linear triangular elements ($P1$-elements). To further illustrate these elements, a schematic of the $P2$ and $P1$ elements (and their nodes) is provided in figure A.1. The basis functions $\psi_i(x, y)$ of the state-space $P1$ are polynomial functions of order 1 on each triangle. Suppose the k^{th} triangle is characterized by the nodes q_i, as shown in figure A.1, then we have

$$\begin{aligned} \psi_i(x,y) = a_i^k + b_i^k x + c_i^k y \quad \text{for} \quad (x,y) \in T_k \\ \psi_i(q_i) = 1 \quad \text{and} \quad \psi_i(q_j) = 0 \quad \text{if} \quad i \neq j \end{aligned} \tag{A.1}$$

Similarly, The basis functions $\phi_i(x, y)$ of the state-space $P2$ are polynomial functions of order 2 on each triangle such that

$$\begin{aligned} \phi_i(x,y) = a_i^k + b_i^k x + c_i^k y + d_i^k x^2 + e_i^k xy + f_i^k y^2 \quad \text{for} \quad (x,y) \in T_k \\ \phi_i(q_i) = 1 \quad \text{and} \quad \phi_i(q_j) = 0 \quad \text{if} \quad i \neq j \end{aligned} \tag{A.2}$$

As a result, the pressure is represented by 3 degrees of freedom on each triangle while a component of the velocity has 6 degrees of freedom. The discretization of the Navier-Stokes equations into the function bases ϕ_i and ψ_i is derived in the following.

A.2 Numerical solution of the incompressible Navier-Stokes equations

We present here the numerical methods used to solve the time-integration of the Navier-Stokes equations. This concerns the $2D$ linearized Navier-Stokes equations solved in chapter 1 and the $2D$ non-linear Navier-Stokes equations considered in chapter 5. As a first step, we first expose the procedure on the non-linear equations; the case of the linearized equations is discussed afterwards. We start by describing the time scheme and we introduce the variational formulation of the equations. The resulting problem is discretized by the finite element method. We then present the Uzawa algorithm used to advance the solution forward in time.

A.2.1 Time scheme

We consider the non-linear Navier-Stokes equations

	α_0	α_1	α_2	β_1	β_2
BDF_1	1	-1		1	
BDF_2	3/2	-2	1/2	2	-1

Table A.1: Coefficients of the BDR_r methods where r is the order of the scheme, see Kress & Lötstedt (2004).

$$\begin{aligned} \frac{\partial \boldsymbol{u}}{\partial t} + \mathcal{N}(\boldsymbol{u}) + \mathcal{L}(\boldsymbol{u}, p) &= 0 \\ \nabla \cdot \boldsymbol{u} &= 0 \end{aligned} \tag{A.3}$$

where $\mathcal{N}(\boldsymbol{u}) = (\boldsymbol{u} \cdot \nabla)\boldsymbol{u}$ denotes the non-linear (and convective) terms while $\mathcal{L}(\boldsymbol{u}) = \nabla p - Re^{-1} \nabla^2 \boldsymbol{u}$ denotes the linear terms. This equation is discretized in time and we look for the solution $(\boldsymbol{u}^{n+1}, p^{n+1})$ at time t^{n+1} as function of the previous solutions $(\boldsymbol{u}^n, p^n)$, $(\boldsymbol{u}^{n-1}, p^{n-1})$, etc. The time step is denoted by $\Delta t = t^{n+1} - t^n$ and is chosen as a fixed value. The non-linear terms are treated explicitly so that equations (A.3) qualitatively transform into

$$\begin{aligned} \frac{1}{\Delta t}\boldsymbol{u}^{n+1} + \mathcal{L}(\boldsymbol{u}^{n+1}, p^{n+1}) &= \mathbf{f}^n_{\text{explicit}} \\ \nabla \cdot \boldsymbol{u}^{n+1} &= 0 \end{aligned} \tag{A.4}$$

where all the explicit terms are gathered in the term $\mathbf{f}^n_{\text{explicit}}$. On the other hand the linear terms are treated semi-implicitly by using the Backward Differentiation Formula scheme (BDF) which leads to

$$\begin{aligned} \sum_{j=0}^{r} \frac{\alpha_j}{\Delta t}\boldsymbol{u}^{n+1-j} + \mathcal{L}(\boldsymbol{u}^{n+1}, p^{n+1}) &= -\sum_{j=1}^{r} \beta_j \mathcal{N}(\boldsymbol{u}^{n+1-j}) \\ \nabla \cdot \boldsymbol{u}^{n+1} &= 0 \end{aligned} \tag{A.5}$$

where the coefficient α_j and β_j are detailed in table A.1 and r is the order of the scheme. Note that the explicit term has been obtained by extrapolating the solution from the previous time steps with a $(r-1)^{th}$ order polynomial. At order 1, the non-linear term extrapolation simply reduces to $(\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^n$. At second order it becomes $2(\boldsymbol{u}^n \cdot \nabla)\boldsymbol{u}^n$-$(\boldsymbol{u}^{n-1} \cdot \nabla)\boldsymbol{u}^{n-1}$. Regarding the time derivative term, it is approximated at order 1 by

$$\frac{\partial \boldsymbol{u}}{\partial t} \approx \frac{\boldsymbol{u}^{n+1} - \boldsymbol{u}^n}{\Delta t} \tag{A.6}$$

while it is given at order 2 by

$$\frac{\partial \boldsymbol{u}}{\partial t} \approx \frac{3\boldsymbol{u}^{n+1} - 4\boldsymbol{u}^n + \boldsymbol{u}^{n-1}}{2\Delta t} \tag{A.7}$$

If we gather in equations (A.5) the terms at t^{n+1} at the left hand side and the terms at t^n at the right hand side, we have the following equation

$$\begin{aligned} \frac{\alpha_0}{\Delta t}\boldsymbol{u}^{n+1} + \mathcal{L}(\boldsymbol{u}^{n+1}, p^{n+1}) &= -\sum_{j=1}^{r} [\frac{\alpha_j}{\Delta t}\boldsymbol{u}^{n+1-j} + \beta_j \mathcal{N}(\boldsymbol{u}^{n+1-j})] \\ \nabla \cdot \boldsymbol{u}^{n+1} &= 0 \end{aligned} \tag{A.8}$$

This one is solved at order $r = 1$ for the first time steps (about the first 10) and then at order $r = 2$.

A.2.2 Variational formulation

The variational formulations associated with equations (A.8) are derived and discretized by the finite element method. We call $\mathcal{X} = \mathsf{H}_1(\Omega)$ and $\mathcal{M} = \mathsf{L}_2(\Omega)$ the continuous state-spaces in the physical domain Ω. We suppose that the velocity $\boldsymbol{u}$ and pressure fields p belong to $\mathcal{X}^2$ and $\mathcal{M}$ respectively (since we consider the two components of the velocity field). Equations (A.8) are multiplied by the test functions $\boldsymbol{v} \in \mathcal{X}^2$ and $q \in \mathcal{M}$ which gives after integration and application of the Green's formula

$$\begin{array}{rrcl} & \frac{\alpha_0}{\Delta t}\int_\Omega \boldsymbol{v}\cdot\boldsymbol{u}^{n+1}d\Omega & & \\ -\int_\Omega (\nabla\cdot\boldsymbol{v})\,p^{n+1}d\Omega \quad & +\frac{1}{Re}\int_\Omega \nabla\boldsymbol{v}:\nabla\boldsymbol{u}^{n+1}\,d\Omega & & \\ +\int_\Gamma \boldsymbol{v}\cdot(p\,\boldsymbol{n})\,d\Gamma \quad & -\frac{1}{Re}\int_\Gamma \boldsymbol{v}\cdot(\nabla\boldsymbol{u}^{n+1}\cdot\boldsymbol{n})\,d\Gamma & = & -\frac{\alpha_j}{\Delta t}\int_\Omega \sum_{j=1}^{r}\boldsymbol{v}\cdot\boldsymbol{u}^{n+1-j}\,d\Omega \\ & & & -\beta_j\int_\Omega \sum_{j=1}^{r}\boldsymbol{v}\cdot[(\boldsymbol{u}^{n+1-j}\cdot\nabla)\boldsymbol{u}^{n+1-j}]\,d\Omega \\ & \int_\Omega q\,(\nabla\cdot\boldsymbol{u}^{n+1})\,d\Omega & = & 0 \qquad \forall\boldsymbol{v}\in\mathcal{X}^2 \text{ and } \forall q\in\mathcal{M} \end{array} \tag{A.9}$$

where $\boldsymbol{n}$ is the outward normal unitary vector of the boundary. Owing to our choice of boundary conditions, the terms defined by an integration on the border Γ vanish. This statement is further discussed later in the paragraph dealing with the boundary conditions. We then write this equation in a more concise form as

$$\begin{array}{rcll} \frac{\alpha_0}{\Delta t}a(\boldsymbol{v},\boldsymbol{u}^{n+1}) + \frac{1}{Re}d(\boldsymbol{v},\boldsymbol{u}^{n+1}) + b(\boldsymbol{v},p^{n+1}) & = & a(\boldsymbol{v},\mathbf{f}) & \forall\boldsymbol{v}\in\mathcal{X}^2 \\ b(\boldsymbol{u}^{n+1},q) & = & 0 & \forall q\in\mathcal{M} \end{array} \tag{A.10}$$

where we have denoted the explicit non-linear terms by $\mathbf{f}$ and introduced and the three following operators

$$a(\boldsymbol{v},\boldsymbol{u}) = \int_\Omega \boldsymbol{v}\cdot\boldsymbol{u}\,d\Omega \qquad d(\boldsymbol{v},\boldsymbol{u}) = \int_\Omega \nabla\boldsymbol{v}:\nabla\boldsymbol{u}\,d\Omega \qquad b(\boldsymbol{v},p) = -\int_\Omega (\nabla\cdot\boldsymbol{v})\,p\,d\Omega \tag{A.11}$$

A.2.3 Space discretization

The computational domain is divided into finite elements. The discretized state-spaces become $\mathcal{X}_h$ and $\mathcal{M}_h$ for the velocity and pressure fields respectively. We denote by $\boldsymbol{u}_h \in \mathcal{X}_h^2$, $p_h \in \mathcal{M}_h$ and $\mathbf{f}_h \in \mathcal{X}_h$ the discretized forms of the previous variables. Suppose $I = \dim(\mathcal{X}_h)$ and $J = \dim(\mathcal{M}_h)$, these quantities are decomposed into the function bases $\boldsymbol{\phi}_i$ and ψ_i by

$$\begin{array}{c} \boldsymbol{u}_h(x,y,t) = \sum_{i=1}^{I}\boldsymbol{u}_i(t)\boldsymbol{\phi}_i(x,y) \qquad p_h(x,y,t) = \sum_{i=1}^{J}p_i(t)\psi_i(x,y) \\ \mathbf{f}_h(x,y,t) = \sum_{i=1}^{I}\mathbf{f}_i(t)\boldsymbol{\phi}_i(x,y) \end{array} \tag{A.12}$$

By introducing these decompositions into equations (A.10), we have

$$\begin{array}{rcll} \frac{\alpha_0}{\Delta t}a(\boldsymbol{v}_h,\boldsymbol{u}_h^{n+1}) + \frac{1}{Re}d(\boldsymbol{v}_h,\boldsymbol{u}_h^{n+1}) + b(\boldsymbol{v}_h,p_h^{n+1}) & = & a(\boldsymbol{v}_h,\mathbf{f}_h) & \forall\boldsymbol{v}_h\in\mathcal{X}_h^2 \\ b(\boldsymbol{u}_h^{n+1},q_h) & = & 0 & \forall q_h\in\mathcal{M}_h \end{array} \tag{A.13}$$

In the end, we can express these equations as a function of the matrices A, D, B and the vector F defined by

$$A_{ij} = \int_\Omega \boldsymbol{\phi}_i \cdot \boldsymbol{\phi}_j \, d\Omega \qquad D_{ij} = \int_\Omega \nabla\boldsymbol{\phi}_i \cdot \nabla\boldsymbol{\phi}_j \, d\Omega \qquad B_{ij} = -\int_\Omega \psi_i \, (\nabla\cdot\boldsymbol{\phi}_j) \, d\Omega$$
$$F_i = a(\boldsymbol{\phi}_i, \mathbf{f}_h) \quad \text{or also} \quad A\,\mathbf{f}_h = F \tag{A.14}$$

We introduce the matrix $M = (\alpha_0/\Delta t)A + Re^{-1}D$ such as to express the system in a more concise form:

$$\begin{pmatrix} M & B^t \\ B & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_h^{n+1} \\ p_h^{n+1} \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix} \tag{A.15}$$

The flow state solution at t^{n+1} may then be computed by solving this system. To do so, we used the iterative algorithm described below.

A.2.4 Uzawa algorithm

We used an Uzawa method to solve equations (A.15). This algorithm allows to compute the flow state $(\boldsymbol{u}_h^{n+1}, p_h^{n+1})$ iteratively. Suppose we know the flow state solutions up to time t^n, we proceed as follows:

(i) We start by computing the explicit term $\mathbf{f}_h$ and the vector F

(ii) We compute the velocity field $\boldsymbol{u}_h^n$ by solving the following problem

$$\boldsymbol{u}_h^n = M^{-1}(F - B^t p_h^n) \tag{A.16}$$

where the inversion of the matrix M is handled by using the UMFPACK library, see Davis (1993).

(iii) We then define the new pressure p_h^{n+1} by

$$p_h^{n+1} = p_h^n + \rho_n B\boldsymbol{u}_h^n \tag{A.17}$$

and restart the procedure at step (ii) by replacing p_h^n by the pressure p_h^{n+1} computed by equation (A.17).

In fact, system (A.15) is considered as a constrained minimization problem: we look for the pressure and velocity fields satisfying the Navier-Stokes equations and which minimize the divergence of the velocity field. The pressure field is identified by using a gradient method with a constant step size. In other words, we choose a fixed value for the step size ρ_n for each time step. When this algorithm converges, the pressure field is such that $B\boldsymbol{u}_h^n = 0$ which enforce a divergence-free velocity field. As a result, the final solution $(\boldsymbol{u}_h^{n+1}, p_h^{n+1})$ satisfies equations (A.15). Since this method converges quite slowly, it requires preconditioning. Here, we used the Cahouet Chabart preconditioner.

A.2.5 Implementation of the boundary conditions

In both the wall and inflow boundary conditions, the two components of the velocity field are fixed, which may be written as $\boldsymbol{u} = \boldsymbol{u}_{bc}$. These two boundary conditions are referred to as Dirichlet conditions. They are imposed by a penalty method: in equation (A.15), we affect the value $F_i = \boldsymbol{u}_{bc} \times 10^{30}$ and the coefficient $M_{ii} = 10^{30}$ at the nodes i associated with these boundary conditions. The two other boundary conditions considered in the thesis are the outflow and symmetry conditions. The outflow condition is chosen as

$$p\,\boldsymbol{n} - Re^{-1}(\nabla\boldsymbol{u})\cdot\boldsymbol{n} = 0 \tag{A.18}$$

which allows to cancel the boundary integrals in equation (A.9). The symmetry condition may be expressed as

$$\partial_y u = 0 \quad \text{and} \quad v = 0 \tag{A.19}$$

where u and v are the two components of the velocity field. The first condition $v = 0$ is a Dirichlet condition which is imposed by the penalty method and the second one $\partial_y u = 0$ is such that both cancel the boundary integrals in equation (A.9).

A.2.6 Case of the linearized Navier-Stokes equations

The only difference between the linear and non-linear Navier-Stokes equations is the convection term. As a result, only the explicit term $\mathbf{f}_h$ is modified. Denoting the base flow by $\boldsymbol{U}_0$, the non-linear term $(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}$ is replaced by $(\boldsymbol{U}_0\cdot\nabla)\boldsymbol{u}+(\boldsymbol{u}\cdot\nabla)\boldsymbol{U}_0$. The first term represents the convection of the perturbation by the base flow while the second is a production term assessing the deformation of the perturbation by the base flow strain. Since the base flow is frozen, these two terms can be computed once and for all before the march in time of the simulation. For this reason, linear simulations are much faster than non-linear ones.

A.3 Computation of the base flow

A.3.1 The Newton method

The base flow $(\boldsymbol{U}, P)$ is obtained by solving the steady Navier-Stokes equations

$$\begin{aligned} (\boldsymbol{U}\cdot\nabla)\boldsymbol{U} &= -\nabla P + Re^{-1}\,\nabla^2\boldsymbol{U} \\ \nabla\cdot\boldsymbol{U} &= 0 \end{aligned} \tag{A.20}$$

with associated boundary conditions. Owing to the convection term $(\boldsymbol{U}\cdot\nabla)\boldsymbol{U}$, these equations are non linear. We used a Newton method to solve iteratively these equations. We start the procedure from an approximate solution $(\boldsymbol{U}_0, P_0)$ satisfying the appropriate boundary conditions. This state is not solution of equations (A.20), which may be formally written as

$$\mathrm{NS}(\boldsymbol{U}_0, P_0) \neq 0 \tag{A.21}$$

where the Navier-Stokes operator is defined by

$$\mathrm{NS}(\boldsymbol{U}_0, P_0) = \begin{pmatrix} (\boldsymbol{U}_0\cdot\nabla)\boldsymbol{U}_0 + \nabla P_0 - Re^{-1}\,\nabla^2\boldsymbol{U}_0 \\ \nabla\cdot\boldsymbol{U}_0 \end{pmatrix} \tag{A.22}$$

The idea is then to look for the modification $(\delta\boldsymbol{U}, \delta P)$ such that the flow state $(\boldsymbol{U}_0 + \delta\boldsymbol{U}, P_0 + \delta P)$ is solution of equation (A.20), namely $\mathrm{NS}(\boldsymbol{U}_0+\delta\boldsymbol{U}, P_0+\delta P) = 0$. To do so, we linearize the Navier-Stokes equations about the approximate solution, which gives

$$\mathrm{LNS}_{(\boldsymbol{U}_0,P_0)}(\delta\boldsymbol{U}, \delta P) = -\mathrm{NS}(\boldsymbol{U}_0, P_0) \tag{A.23}$$

where $\mathrm{LNS}_{(\boldsymbol{U}_0,P_0)}$ denotes the Navier-Stokes equations linearized about the state $(\boldsymbol{U}_0, P_0)$ such that

$$\mathrm{LNS}_{(\boldsymbol{U}_0,P_0)}(\delta\boldsymbol{U}, \delta P) = \begin{pmatrix} (\boldsymbol{U}_0\cdot\nabla)\delta\boldsymbol{U} + (\delta\boldsymbol{U}\cdot\nabla)\boldsymbol{U}_0 + \nabla(\delta P) - Re^{-1}\,\nabla^2(\delta\boldsymbol{U}) \\ \nabla\cdot\delta\boldsymbol{U} \end{pmatrix} \tag{A.24}$$

As a result, $(\delta\boldsymbol{U}, \delta P)$ can be computed by solving the linear system resulting from equations (A.23). Next, we replace the initial state $(\boldsymbol{U}_0, P_0)$ by the new one $(\boldsymbol{U}_0 + \delta\boldsymbol{U}, P_0 + \delta P)$ and apply the procedure iteratively until $\mathrm{NS}(\boldsymbol{U}_0, P_0) = 0$. In the end, the final state $(\boldsymbol{U}_0, P_0)$ reduces to the base flow solution.

The successive resolution of the linear system (A.23) is handled by the UMFPACK library. To give an idea, about 5 to 20 iterations are typically required for the flow configurations investigated.

A.3.2 Numerical solution: spatial discretization

We now show how to derive and discretize the variational formulation associated with equation (A.23). This equation may be written in the form

$$\begin{pmatrix} (\boldsymbol{U}_0\cdot\nabla)\delta\boldsymbol{U} + (\delta\boldsymbol{U}\cdot\nabla)\boldsymbol{U}_0 + \nabla(\delta P) - Re^{-1}\,\nabla^2(\delta\boldsymbol{U}) \\ \nabla\cdot\,\delta\boldsymbol{U} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_0 \\ g_0 \end{pmatrix} \tag{A.25}$$

where the left-hand side term stands for the linearized Navier-Stokes operator and the right-hand side one is a known source term. Basically, we wish to compute $(\delta\boldsymbol{U},\delta P)=(\delta U,\delta V,\delta P)$ from the base flow $\boldsymbol{U}_0$ and the source term $(\mathbf{f}_0,g_0)$. We proceed similarly to the time-integrated Navier-Stokes equations. The first equation is multiplied by the test function $\boldsymbol{v}\in\mathcal{X}^2$ and the second one by the test function $q\in\mathcal{M}$. Next, the two resulting equations are integrated in the domain Ω, which gives

$$\begin{aligned} \int_\Omega \boldsymbol{v}\cdot[(\boldsymbol{U}_0\cdot\nabla)\delta\boldsymbol{U}]\,d\Omega \quad &+\int_\Omega \boldsymbol{v}\cdot[(\delta\boldsymbol{U}\cdot\nabla)\boldsymbol{U}_0]\,d\Omega & & & \\ +\int_\Omega \boldsymbol{v}\cdot\nabla(\delta P)\,d\Omega \quad &-Re^{-1}\int_\Omega \boldsymbol{v}\cdot\nabla^2(\delta\boldsymbol{U})\,d\Omega &=& -\int_\Omega \boldsymbol{v}\cdot\mathbf{f}_0\,d\Omega & \forall\boldsymbol{v}\in\mathcal{X}^2 \\ & \int_\Omega q\,(\nabla\cdot\,\delta\boldsymbol{U})\,d\Omega &=& -\int_\Omega q\,g_0\,d\Omega & \forall q\in\mathcal{M} \end{aligned} \tag{A.26}$$

We then apply the Green's formula to the first equation so that

$$\begin{aligned} \int_\Omega \boldsymbol{v}\cdot[(\boldsymbol{U}_0\cdot\nabla)\delta\boldsymbol{U}]\,d\Omega \quad &+\int_\Omega \boldsymbol{v}\cdot[(\delta\boldsymbol{U}\cdot\nabla)\boldsymbol{U}_0]\,d\Omega & & & \\ -\int_\Omega (\nabla\cdot\boldsymbol{v})\,\delta P\,d\Omega \quad &+Re^{-1}\int_\Omega \nabla\boldsymbol{v}\cdot\nabla(\delta\boldsymbol{U})\,d\Omega & & & \\ +\int_\Gamma \boldsymbol{v}\cdot(\delta P\,\boldsymbol{n})\,d\Gamma \quad &-Re^{-1}\int_\Gamma \boldsymbol{v}\cdot[\nabla(\delta\boldsymbol{U})\cdot\boldsymbol{n}]\,d\Gamma &=& -\int_\Omega \boldsymbol{v}\cdot\mathbf{f}_0\,d\Omega & \forall\boldsymbol{v}\in\mathcal{X}^2 \\ & \int_\Omega q\,(\nabla\cdot\,\delta\boldsymbol{U})\,d\Omega &=& -\int_\Omega q\,g_0\,d\Omega & \forall q\in\mathcal{M} \end{aligned} \tag{A.27}$$

The boundary conditions imposed on the fields $(\delta U,\delta V,\delta P)$ are very similar to those used for the time-integrated Navier-Stokes equations. For instance, we impose $\delta\boldsymbol{U}=0$ for the Dirichlet boundary conditions such as to fix the two components of the velocity on the associated boundaries. Concerning the symmetry condition, it reads

$$\partial_y(\delta U)=0 \quad \text{and} \quad \delta V=0 \tag{A.28}$$

and the free outflow condition is chosen as

$$\delta P\,\boldsymbol{n} - Re^{-1}[\nabla(\delta\boldsymbol{U})]\cdot\boldsymbol{n}=0 \tag{A.29}$$

Consequently, whatever the boundary condition, the boundary integrals obtained in equations (A.27) vanish. The next step consists of discretizing the resulting equations by using the finite element method. We consider the previously introduced discretized state-spaces $\mathcal{X}_h$ and $\mathcal{M}_h$ along with their basis $\{\phi_i\}_{i=1\ldots I}$ and $\{\psi_i\}_{i=1\ldots J}$. The velocity and pressure fields are discretized in the following form

$$\delta \boldsymbol{U}_h(x,y,t) = \sum_{i=1}^{I} \delta \boldsymbol{U}_i(t)\boldsymbol{\phi}_i(x,y) \qquad \delta P_h(x,y,t) = \sum_{i=1}^{J} \delta P_i(t)\psi_i(x,y) \tag{A.30}$$

As well, the base flow and the source terms are discretized, the resulting states being denoted by $\boldsymbol{U}_{0h}$, $\mathbf{f}_{0h}$ and g_{0h}. Consequently, we are left with

$$\begin{aligned}
\int_\Omega \boldsymbol{v}_h \cdot [(\boldsymbol{U}_{0h}\cdot\nabla)\delta\boldsymbol{U}_h]\,d\Omega \quad &+\int_\Omega \boldsymbol{v}_h \cdot [(\delta\boldsymbol{U}_h\cdot\nabla)\boldsymbol{U}_{0h}]\,d\Omega \\
-\int_\Omega (\nabla\cdot\boldsymbol{v}_h)\,\delta P_h\,d\Omega \quad &+Re^{-1}\int_\Omega \nabla\boldsymbol{v}_h : \nabla(\delta\boldsymbol{U}_h)\,d\Omega \;=\; -\int_\Omega \boldsymbol{v}_h\cdot\mathbf{f}_{0h}\,d\Omega \qquad \forall \boldsymbol{v}_h \in \mathcal{X}_h^2 \\
&\int_\Omega q_h\,(\nabla\cdot\delta\boldsymbol{U}_h)\,d\Omega \;=\; -\int_\Omega q_h\,g_{0h}\,d\Omega \qquad \forall q_h \in \mathcal{M}_h
\end{aligned} \tag{A.31}$$

To simplify these equations, we introduce the matrices P and C along with the vectors F_0 and G_0 defined by

$$\begin{aligned}
P_{ij} &= \int_\Omega \boldsymbol{\phi}_i\cdot[(\boldsymbol{\phi}_j\cdot\nabla)\boldsymbol{U}_0]\,d\Omega & C_{ij} &= \int_\Omega \boldsymbol{\phi}_i\cdot[(\boldsymbol{U}_0\cdot\nabla)\boldsymbol{\phi}_j]\,d\Omega \\
F_{0,i} &= \int_\Omega \boldsymbol{\phi}_i\cdot\mathbf{f}_{0h}\,d\Omega & G_{0,i} &= \int_\Omega \psi_i\,g_{0h}\,d\Omega
\end{aligned} \tag{A.32}$$

so that we can derive a more concise expression. By introducing the matrix $N = P + C + Re^{-1}D$, the system (A.31) reads

$$\begin{pmatrix} N & B^t \\ B & 0 \end{pmatrix}\begin{pmatrix} \delta\boldsymbol{U}_h \\ \delta P_h \end{pmatrix} = \begin{pmatrix} F_0 \\ G_0 \end{pmatrix} \tag{A.33}$$

The unknown state ($\delta\boldsymbol{U}_h$,δP_h) may then be computed by solving this linear system. It should be noticed that the matrix associated with this system is very large. The size of the state vector ($\delta\boldsymbol{U}_h$,δP_h) is typically of order $n \approx O(10^5)$ in our flow configurations. The size of the linear system is thus of size $n^2 \approx O(10^{10})$. Yet, the matrix associated with this system is sparse which offers advanced numerical techniques to inverse it. In our cases, this inversion is handled by using the direct multifrontal sparse LU solver called MUMPS, see Amestoy *et al.* (2001).

An important point should be mentioned here. We just showed how to solve equation (A.23) in order to iteratively compute a steady solution of the Navier-Stokes equations (namely a base flow). Basically, we have derived the variational formulation associated with the linearized Navier-Stokes equations and discretized it by the finite element method. We would like to emphasize that the large matrix introduced in equation (A.33) accounts for the discretized Navier-Stokes operator. The computation of this matrix is the first step to identify the large generalized eigenvalue problems encountered in this thesis. Such eigenvalue problems intervene in both the computation of the global modes (as in chapter 1) or the computation of the optimal harmonic forcing and responses (see chapter 1 and chapter 4). Owing to the very similar variational formulations involved in both cases, we chose to not further expose their calculations.

A.4 Solving the large generalized eigenvalue problems

Some flow states introduced in chapter 1 and chapter 4, such as the leading global modes, the optimal harmonic forcing or the optimal harmonic responses, are computed by solving a large generalized eigen-

value problem. We briefly describe here how to handle such problems. Let us consider as a prototype example the eigenvalue problem

$$\boldsymbol{A}\,\mathsf{x} = \sigma\,\mathbf{B}\mathsf{x} \tag{A.34}$$

where $\boldsymbol{A}$ and $\mathbf{B}$ are two operators, σ is the eigenvalue and x the eigenvector. Note that all these quantities may be complex in general. With our finite element discretization, both the matrices $\boldsymbol{A}$ and $\mathbf{B}$ are very large and sparse. To give an idea, for a matrix $\boldsymbol{A}$ of size $n \times n$ with $n \approx 10^6$, the number of non-zero entries is about 10^7. The matrix $\mathbf{B}$ is not diagonal so that equation (A.34) is referred to as a generalized eigenvalue problem. Furthermore, $\mathbf{B}$ is hermitian and non-invertible. We wish to compute the eigenvalues and eigenvectors of this problem. The standard QR methods are not able to deal with such large systems since we have typically $n \approx O(10^5)$ degrees of freedom. Instead, we can apply iterative techniques to approximate a part of the eigenspectrum. The best algorithm known to date to compute the eigensolutions of such systems is the Implicitly Restarted Arnoldi Method (IRAM) which is described in the following.

A.4.1 The shift and invert transformation

Basically, the standard Arnoldi algorithm allows to compute the largest eigenvalues of a matrix. Yet, the Arnoldi algorithm combined with a *shift and invert* transformation allows to compute the eigenvalues which are the closest to any complex value. Suppose we look for the eigenvalues σ which are the closest to the complex β, which is referred to as the shift. The initial problem is formulated as

$$(\boldsymbol{A} - \beta\mathbf{B})\,\mathsf{x} = (\sigma - \beta)\,\mathbf{B}\mathsf{x} \tag{A.35}$$

which leads to the following equivalent eigenvalue problem

$$\mathbf{C}\,\mathsf{x} = \kappa\,\mathsf{x} \tag{A.36}$$

where the eigenvalue becomes $\kappa = (\sigma - \beta)^{-1}$ and $\mathbf{C} = (\boldsymbol{A} - \beta\mathbf{B})^{-1}\mathbf{B}$. The resulting system defines a standard eigenvalue problem. Mostly, the highest eigenvalues κ of the problem (A.36) becomes the eigenvalues σ of the initial problem which are the closest to the shift β. Thus, applying the Arnoldi method to the problem (A.36) allows to compute the eigenvalues σ in a desired part of the complex plane by varying the shift position.

A.4.2 The Arnoldi method

The basic idea of the Arnoldi method can be understood by considering the power iteration method. Suppose a state x is decomposed in the basis of the n eigenvectors x_i of $\mathbf{C}$ as

$$\mathsf{x} = \sum_{i=1}^{n} \varsigma_i\,\mathsf{x}_i \tag{A.37}$$

If we denote the associated eigenvalues κ_i in decreasing order ($\kappa_1 > \kappa_2 > \ldots > \kappa_n$) and multiply m times the previous relation by the matrix $\mathbf{C}$, we have

$$\mathbf{C}^m\mathsf{x} = \sum_{i=1}^{n} \varsigma_i \mathbf{C}^m\,\mathsf{x}_i = \sum_{i=1}^{n} \varsigma_i \kappa_i^m\,\mathsf{x}_i \tag{A.38}$$

Consequently, if we choose a sufficiently large number m of iterations and if the difference between the two largest eigenvalues κ_1 and κ_2 is large enough, then we have $\mathbf{C}^m\mathsf{x} = \varsigma_1\kappa_1^m\,\mathsf{x}_1$ from which the eigenvalue κ_1 may be calculated by the Rayleigh quotient

$$\kappa_1 = \frac{\mathsf{x}_1^* \mathbf{C} \mathsf{x}_1}{\mathsf{x}_1^* \mathsf{x}_1} \tag{A.39}$$

where * denotes the transconjugate. This ability of the power method to converge the largest eigenvalue of the problem $\mathbf{C}\mathsf{x} = \kappa\mathsf{x}$ is used in the Arnoldi algorithm. More precisely, the Arnoldi algorithm allows to compute more than one eigenvalue. The idea is to project the original problem (A.36) onto a Krylov subspace of low dimension and solve the resulting low-order problem. Let us consider the Krylov subspace of dimension k associated with the initial vector w

$$\mathcal{K}_k = \text{span}\{\mathsf{w}, \mathbf{C}\mathsf{w}, \mathbf{C}^2\mathsf{w}, \ldots, \mathbf{C}^{k-1}\mathsf{w}\} \tag{A.40}$$

The purpose of the Arnoldi algorithm is to find the better approximation of the leading eigenvalues and eigenvectors of $\mathbf{C}$ in this subspace. To do so, an orthogonalization of the Krylov subspace is performed by a Gram-Schmidt process. The resulting orthogonal basis v_i is such that $\mathcal{K}_k = \text{span}\{\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_k\}$. The Arnoldi algorithm proceed by computing iteratively the following expression

$$\mathbf{C}\mathbf{V}_k = \mathbf{V}_k\mathbf{H}_k + \mathbf{f}_k\mathbf{e}_k^t \tag{A.41}$$

where $\mathbf{V}_k$ is a matrix of size $n \times k$ whose columns are the orthogonal vectors v_i, $\mathbf{H}_k$ is an upper Hessenberg matrix of size $k \times k$ and $\mathbf{f}_k\mathbf{e}_k^t$ is a residual. Note that $\mathbf{f}_k$ and $\mathbf{e}_k$ are vectors of size n and k respectively. The calculation of the vectors v_i is done iteratively at each addition of a new member in the Krylov subspace, see Lehoucq & Scott (1997) for details. The procedure is meant to cancel the residual $\mathbf{f}_k\mathbf{e}_k^t$ so that the eigenvalues of the matrix $\mathbf{H}_k$ reduce to the leading eigenvalues of $\mathbf{C}$.

The eigenvalues θ_i and associated eigenvectors s_i of $\mathbf{H}_k$ may be computed easily by a standard QR method. As soon as the residual term $\mathbf{f}_k\mathbf{e}_k^t$ becomes negligible, the Krylov subspace $\mathcal{K}_k$ yields a good approximation of the leading eigenspace of $\mathbf{C}$ and the vectors v_i span this eigenspace. In such a case, the eigenvalues θ_i are good approximations to the eigenvalues κ_i of $\mathbf{C}$. Furthermore, The associated eigenvectors x_i may be recovered from those of $\mathbf{H}_k$ by

$$\mathsf{x}_i = \mathbf{V}_k\mathsf{s}_i \tag{A.42}$$

The accuracy of the approximation increases with the Krylov subspace size k. However, it is not possible to know the size k required to reach a given accuracy. Mostly, the convergence speed of the algorithm strongly depends on the initial vector w used to build the Krylov subspaces. For instance, if some of the eigenvectors x_i we are looking for are nearly orthogonal to the Krylov subspaces, the convergence speed becomes very low. As a result, a bad choice of the initial vector w may lead to a substantial increase in the number of required iterations to have an accurate approximation of the eigenvalues. For this reason, instead of increasing k until convergence, a possible improvement is to restart the Arnoldi algorithm with the same value of k but with a different initial vector. The different restart techniques may be classified in two categories: the explicit methods ERAM (Explicitly Restarted Arnoldi Method) and the implicit methods IRAM (Implicitly Restarted Arnoldi Method). For additional information on the restart techniques, we refer interested readers to the work by Sedrakian (2000).

A.4.3 Practical solution

The resolution of equation (A.34) is handled by the ARPACK library, see Lehoucq & Scott (1997). It is based on an IRAM procedure developed by Sorensen and allows to solve both real and complex generalized eigenvalue problems. It requires the successive inversion of linear systems of the form

$$(\boldsymbol{A} - \beta\mathbf{B})\,\mathsf{y} = \mathsf{z} \tag{A.43}$$

as well as matrix-vector products of the form $\mathsf{z} = \mathbf{B}\mathsf{x}$. The linear systems inversion is handled by the direct multifrontal sparse LU solver called MUMPS, see Amestoy *et al.* (2001). Most of our eigenvalue problems are complex so that we used the complex driver *zndrv4* of ARPACK. We specify to the driver the following information:

(i) The type of eigenvalue search. We used the "LM" mode corresponding to eigenvalues with largest magnitude.

(ii) The desired number of eigenvalues

(iii) The maximum size k of the Krylov subspaces

(iv) The maximum number of iteration of the algorithm

(v) The complex guess (or shift) value

Interestingly, the software provides a measure of the error. For each eigenvalue/eigenvector (σ_i, x_i), we have access to the residual norm

$$\|(\boldsymbol{A} - \sigma_i \mathbf{B})\,\mathsf{x}_i\|_2 = \|\boldsymbol{\xi}\|_2 = \sqrt{\sum_{j=1}^{n} \boldsymbol{\xi}_j^2} \tag{A.44}$$

where $\boldsymbol{\xi}_j$ denotes the components of the residual vector $\boldsymbol{\xi}$. In all our computations, we checked that $\|\boldsymbol{\xi}\|_2$ does not exceed 10^{-12} to have converged results.

Appendix B

Reduced-order models based on global modes

In this appendix, we deal with the ability of the global modes to design efficient Reduced Order Models (ROMs). We proceed by projecting the original dynamics onto the set of the least stable global modes and we analyze the transient dynamics of the resulting models. Such an analysis has already been performed in some previous works such as those by Ehrenstein & Gallaire (2005); Åkervik *et al.* (2007); Henningson & Åkervik (2008); Ehrenstein & Gallaire (2008*b*). In all these works, the performance of the global modes to model the transient flow dynamics is assessed through an optimal perturbation analysis. The authors proceeded by computing the optimal energy gain possibly triggered by all the possible perturbations within the subspace spanned by the set of the least damped global modes. While increasing the number of global modes (and the size of the models) they observed a convergence of the optimal energy gains. This indicated that the ROMs could potentially capture the most amplified perturbations of the original flow. The evolved optimal perturbations computed on their ROMS are also been observed to display amplified wave packets advected in the downstream direction, which is consistent with our expectations.

Contrary to these previous works, we performed an additional optimal perturbation analysis on the full original flow. This allowed to quantify more accurately the difference between the optimal energy gains computed on the ROMs and on the full system. To do so, we used exactly the same numerical technique described in chapter 1 and referred to as a "direct-adjoint" procedure. How to perform the model reduction and the computation of the optimal perturbations in the reduced models are presented next.

B.1 Model reduction: governing equations

Analogously to the above-mentioned works, we restrict our analysis to the dynamics of two-dimensional perturbations ($k = 0$) without transverse velocity component. Based on the formalism of chapter 1, the perturbation may be written $\boldsymbol{q}(x, y, t) = (u, v, p)$. It is then decomposed in the form of a global mode:

$$\boldsymbol{q}(x, y, t) = \frac{1}{2} \left[\hat{\boldsymbol{q}}(x, y)\, e^{\sigma t} + c.c. \right] \tag{B.1}$$

where $\hat{\boldsymbol{q}}$ is the global eigenvector and $\sigma = \alpha + i\omega$ denotes its associated eigenvalue. Note that both are complex. As previously seen in chapter 1, the global modes are solution of the large scale generalized eigenvalue problem

$$\boldsymbol{A}\, \hat{\boldsymbol{q}} = \sigma\, \mathbf{B} \hat{\boldsymbol{q}} \tag{B.2}$$

which is solved by an iterative shift-invert Arnoldi technique. By applying several times the procedure, with different shift values, we compute the part of the eigenspectrum corresponding to the most unstable modes. Next, the idea is to decompose the flow state $\boldsymbol{q}$ into a sum of global modes:

$$\boldsymbol{q} = \mathbf{X}\boldsymbol{q}_r \tag{B.3}$$

where $\mathbf{X}$ is a matrix whose columns are the global modes and $\boldsymbol{q}_r$ reduces to the coordinates of $\boldsymbol{q}$ in the global modes basis. If we truncate $\mathbf{X}$ to the most unstable global modes, then $\boldsymbol{q}_r$ stands for a reduced-order flow state. Let us give some insight on the size of these quantities. We denote the number of degrees of freedom of the full system by n so that the flow state $\boldsymbol{q}$ is of size $n \times 1$. When computing the complex eigenmodes, we pay attention to only compute the eigenvalues located on the upper half complex plan ($\omega \geq 0$) since those located on the lower plan simply reduce to their complex conjugate. If we compute n_c complex eigenmodes on the half complex plan and if there are n_s stationary modes ($\omega = 0$), then we get in the end $n_r = 2n_c - n_s$ linearly independent complex global modes. These n_r modes constitutes the columns of the matrix $\mathbf{X}$ which is of size $n \times n_r$ and the reduced-order state $\boldsymbol{q}_r$ is of size $n_r \times 1$. Note that the global modes selected to build the models are always chosen to be the least stable ones, as in the above-mentioned articles.

Given the definition (B.1), we known analytically how the global modes evolve in time. As a result, the dynamics of the state $\boldsymbol{q} = \mathbf{X}\boldsymbol{q}_r$ can be written as

$$\boldsymbol{q}(t) = \mathbf{X}e^{\boldsymbol{\Sigma}t}\boldsymbol{q}_{r0} \tag{B.4}$$

where $\boldsymbol{q}_{r0} = \boldsymbol{q}_r(t = 0)$ and $\boldsymbol{\Sigma}$ is the diagonal matrix whose elements are the complex eigenvalues σ_i associated with the eigenvectors constituting the columns of $\mathbf{X}$. The expression (B.4) allows to compute analytically the time evolution of the low rank flow state $\boldsymbol{q}(t)$. This valuable property then allows to compute the optimal perturbations on the ROM in a straightforward manner. According to the formalism introduced in chapter 1, the optimal perturbation is the state $\boldsymbol{q}(0)$ which optimizes the kinetic energy gain γ such that

$$\boldsymbol{q}^\dagger(T)\mathbf{B}\boldsymbol{q}(T) = \gamma\, \boldsymbol{q}^\dagger(0)\mathbf{B}\boldsymbol{q}(0) \tag{B.5}$$

where the adjoint quantities, denoted by the superscript †, are such that the kinetic energy of the state $\boldsymbol{q}$ is equal to $E(\boldsymbol{q}) = \langle \boldsymbol{q}, \mathbf{B}\boldsymbol{q} \rangle = \boldsymbol{q}^\dagger\mathbf{B}\boldsymbol{q}$. Next we replace $\boldsymbol{q}(T)$ in equation (B.5) by its decomposition (B.4). Notably, the adjoint of $\boldsymbol{q}(T)$ is given by

$$\boldsymbol{q}^\dagger(T) = [\mathbf{X}e^{\boldsymbol{\Sigma}T}\boldsymbol{q}_{r0}]^\dagger = \boldsymbol{q}_{r0}^* e^{\boldsymbol{\Sigma}^* T}\mathbf{X}^\dagger \tag{B.6}$$

where * denotes the transconjugate. In the end, we are left with a similar but smaller generalized eigenvalue problem

$$[e^{\boldsymbol{\Sigma}^* T}\mathbf{X}^\dagger\mathbf{B}\mathbf{X}e^{\boldsymbol{\Sigma}T}]\boldsymbol{q}_{r0} = \gamma_r\, [\mathbf{X}^\dagger\mathbf{B}\mathbf{X}]\boldsymbol{q}_{r0} \tag{B.7}$$

whose largest eigenvalue γ_r and associated eigenvector yield the optimal energy gain and optimal perturbation possibly sustained in the reduced basis of global modes. Note that the term $\mathbf{X}^\dagger\mathbf{B}\mathbf{X}$ stands for the cross product matrix of the global modes with each other. Indeed, the inner product between the i^{th} and j^{th} global modes (denoted by $\mathbf{X}_i$ and $\mathbf{X}_j$) is given by $\langle \mathbf{X}_i, \mathbf{B}\mathbf{X}_j \rangle$ and is also equal to the (i, j) component of the matrix $\langle \mathbf{X}, \mathbf{B}\mathbf{X} \rangle = \mathbf{X}^\dagger\mathbf{B}\mathbf{X}$.

B.2 Results: application to the rounded backward-facing step flow

The previous procedure is first applied on the backward-facing step flow investigated in chapter 1. The geometry of the computational domain and associated boundary conditions may be found there. We

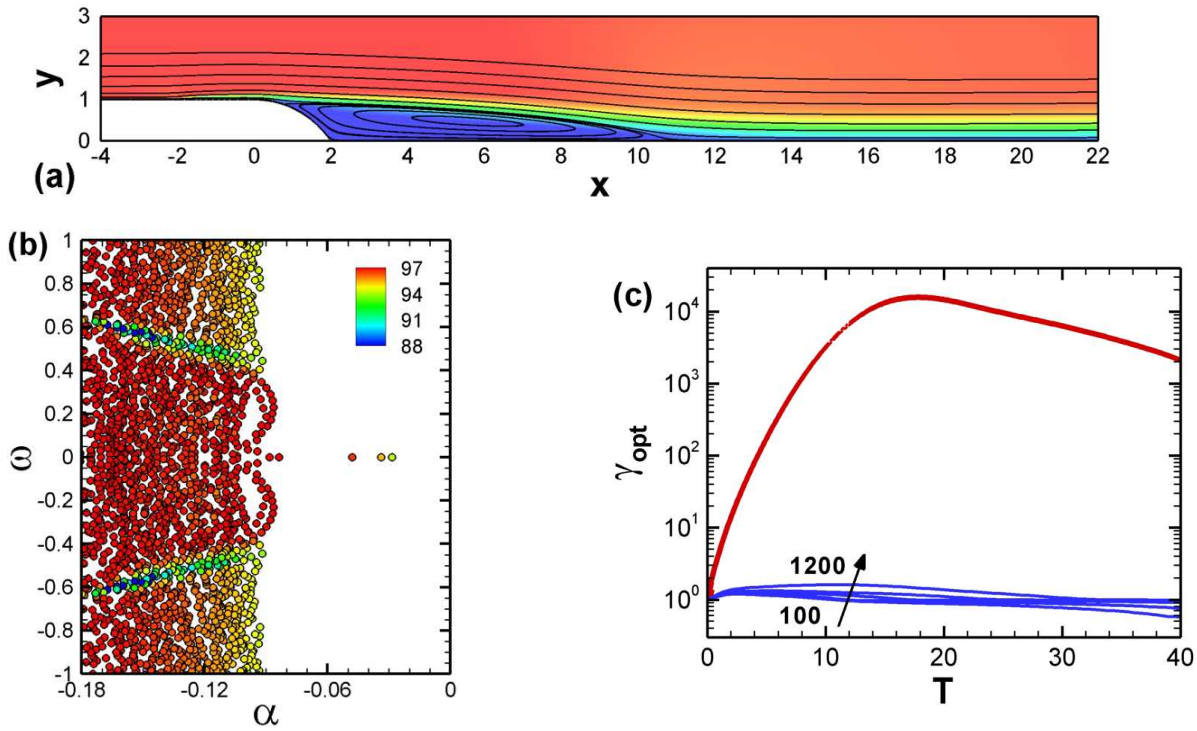


Figure B.1: (a) Base flow solution at $Re = 600$ depicted by its streamlines and longitudinal velocity. (b) Eigenspectrum of the flow. The eigenvalues are colored by the kinetic energy barycenter x_c of their associated eigenvector. (c) Evolution of the optimal energy gain γ_{opt} as a function of the horizon time T. The red curve represents the solution computed on the full system. The set of blue curves represent the solutions from the ROMs with $n_c = 100, 200, 400, 800$ and 1200.

choose to fix the Reynolds number to $Re = 600$. The corresponding base flow has been depicted in figure B.1(a) by its streamlines and longitudinal velocity. Note that a typical discretization yields $n \approx 400\,000$ degrees of freedom stemming from about 90 000 triangles. The eigenspectrum associated with this base flow has been exposed in figure B.1(b). Furthermore, the eigenvalues have been colored by the kinetic energy barycenter x_c of their associated eigenvector, which is computed by

$$x_c = \frac{\int_\Omega x\,[\hat{u}^2 + \hat{v}^2]dxdy}{\int_\Omega [\hat{u}^2 + \hat{v}^2]dxdy} \tag{B.8}$$

This quantity has been introduced to give an indication on the mean longitudinal location of the global modes. Interestingly, we observe that a majority of the global modes are localized in the very vicinity of the downstream boundary of the computational domain. Indeed, the downstream boundary is located at $x = 100$ and the major part of the computed eigenvalues are characterized by $x_c > 96$.
The evolution of the optimal energy gains γ_{opt} as a function of T are displayed in figure B.1(b). The red curve represents the solution on the full original system whereas the blue curves depict the solution computed on the reduced basis of global modes. We observe that increasing the size of the ROMs, from $n_c = 100$ to 1200, barely increases the associated optimal energy gains. In all the cases, the ROMs clearly cannot capture the optimal perturbations of the original system. This result is not surprising since they are localized upstream, near separation, whereas the projection basis consists of structures localized in the very downstream.
To check the accuracy of the global modes computation, we considered the residual $\boldsymbol{\xi} = \boldsymbol{A}\,\hat{\boldsymbol{q}} - \sigma\,\mathbf{B}\hat{\boldsymbol{q}}$. It was observed that the 2 norm $\|\boldsymbol{\xi}\|_2 = \sqrt{\sum_{i=1}^{n} \boldsymbol{\xi}_i^2}$ does not exceed 10^{-12} for all the considered global modes, which indicates that our computation is consistent. The spatial localization of the modes in the very downstream boundary seems rather related to the very high non-normality of the global modes,

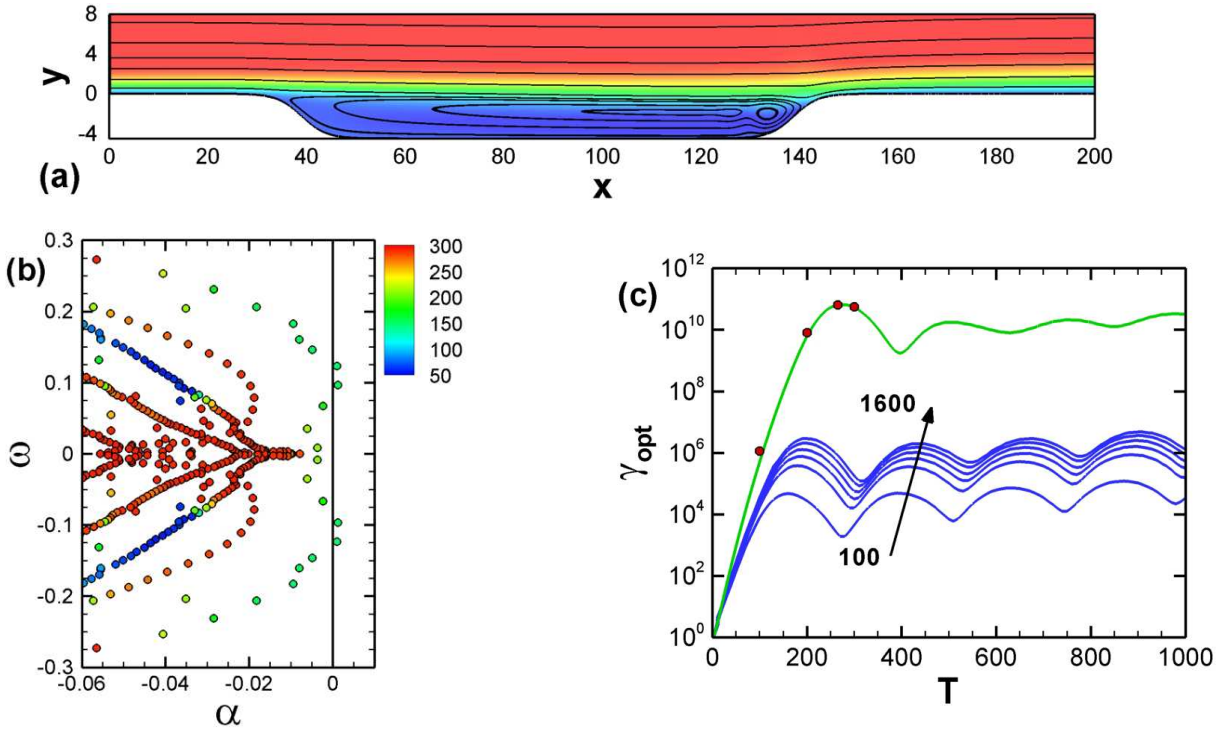


Figure B.2: (a) Base flow solution at $Re = 350$ depicted by its streamlines and longitudinal velocity. (b) Eigenspectrum of the flow. The eigenvalues are colored by the kinetic energy barycenter x_c of their associated eigenvector. (c) Evolution of the optimal energy gain γ_{opt} as a function of the horizon time T. The red dots represent four solutions computed on the full system for $T = 100, 200, 265$ and 300. The green line depicts the time-integrated optimal perturbation computed for $T = 265$ on the full system. The blue lines stand for the optimal perturbations computed on the ROMs with $n_c = 100, 200, 400, 800, 1200$ and 1600.

and of the Navier-Stokes operator. This feature has been exposed by Cossu & Chomaz (1997) as a result of the convective non-normality. Our results show that the most stable modes, truncated here, may be of utmost importance to describe the short-time dynamics of this particular flow. Unfortunately, the more stable the eigenvalues the more ill-conditioned are the associated eigenvectors, see Trefethen & Embree (2005). As a result, the computation of a larger part of the eigenspectrum quickly becomes not tractable. For interested readers, a critical assessment on the use of global modes to design ROMs may be found in the articles by Barbagallo *et al.* (2009); Ehrenstein *et al.* (2010); Alizard & Robinet (2010); Sipp *et al.* (2010).

To further investigate the ability of the global modes to capture the optimal perturbations of a flow, we choose to test the procedure on other flow configurations. These different cases are briefly investigated and exposed in the remaining of this appendix.

B.3 Case of a high-aspect-ratio smooth cavity flow

The second considered configuration reduces to the high-aspect-ratio smooth cavity flow originally investigated by Åkervik *et al.* (2007). We adopted exactly the same geometry, boundary conditions and Reynolds number as in this reference article. The geometry of the cavity and the associated base flow have been depicted in figure B.2(a). The smooth cavity is symmetric with respect to its center at $x_e = 89$ and its upstream wall boundary is defined by $\Gamma(x) = -2.25(\tanh(a(x-b))+1)$ for $0 \leq x \leq x_e$ with $a = 0.2$ and $b = 39$, such as to match smoothly the flat plate upstream and downstream. All variables are made non-dimensional with the displacement thickness δ^* and the free-stream velocity U_∞ at the inflow

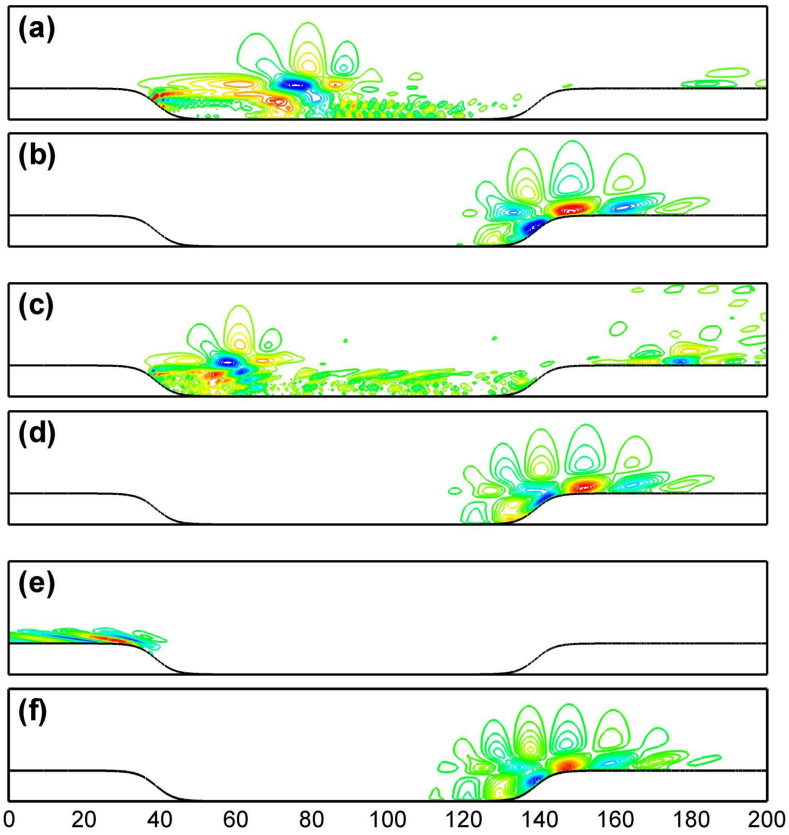


Figure B.3: (a,c,e) Optimal perturbations computed on the ROMs with $n_c = 100, 1600$ and on the full system respectively. (b,d,f) represent their corresponding optimal responses, which are reached for $T = 160, 200$ and 265 respectively. Flow structures are displayed by their longitudinal velocity.

$x = 0$, where a Blasius profile is prescribed. The Reynolds number is defined, for this configuration, by $Re = U_\infty \delta^*/\nu$ and is fixed to 350. Following the procedure used by Åkervik *et al.* (2007), we first computed the base flow corresponding to the domain $0 \leq x \leq 400$, $\Gamma(x) \leq y \leq 80$. Then we interpolated the resulting steady flow into the domain $0 \leq x \leq 300$, $\Gamma(x) \leq y \leq 75$ where stability calculations were carried out. To give an idea, the second domain discretization lead to $n \approx 390\,000$ degrees of freedom stemming from about 86 000 triangles.

Figure B.2(b) displays the eigenspectrum of the flow. It is similar to that obtained in the reference work by Åkervik *et al.* (2007). Notably, it exhibit two unstable and oscillating global modes. Similarly to the previous case, we have reported the values of the kinetic energy barycenter x_c of the associated eigenvectors. Even if a lot of global modes are localized in the very vicinity of the downstream boundary (at $x = 300$), a branch of modes is observed to be localized within the shear layer. These modes are characterized by a streamwise position such that x_c is comprised, roughly, between 50 and 150.

The evolution of the optimal energy gains are represented in figure B.2(c). The red dots stand for the solutions computed on the full system for four different horizon times. The green line represents the evolution of the energy gain $\gamma(t)$ triggered by the optimal perturbation which is computed for $T = 265$ on the full system. Finally, the optimal gains computed on the ROMs are represented by the set of blue lines. Interestingly, the curves obtained from the different ROMs display the same trend: a fast initial transient energy growth is followed by a global cycle with a period of approximately 300 time units. Even if the optimal energy gains possibly sustained by the ROMs seem to converge while increasing the number of modes, they stay nearly four orders of magnitude below the solution computed on the full system.

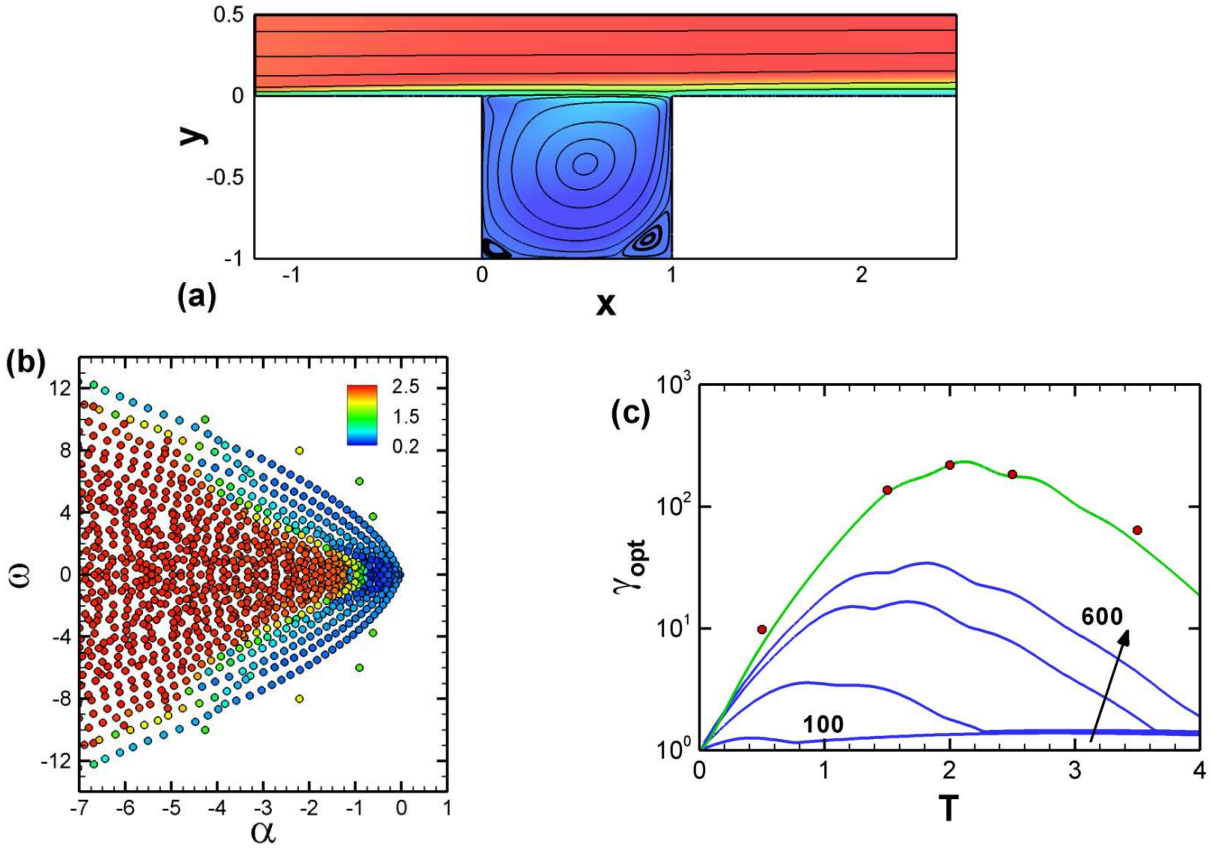


Figure B.4: (a) Base flow solution at $Re = 2000$ depicted by its streamlines and longitudinal velocity. (b) Eigenspectrum of the flow. The eigenvalues are colored by the kinetic energy barycenter x_c of their associated eigenvector. (c) Evolution of the optimal energy gain γ_{opt} as a function of the horizon time T. The red dots represent five solutions computed on the full system for $T = 0.5, 1.5, 2, 2.5$ and 3.5. The green line depicts the time-integrated optimal perturbation computed for $T = 2$ on the full system. The blue lines stand for the optimal perturbations computed on the ROMs with $n_c = 100, 200, 400$ and 600.

Now, let us show the spatial structure of the optimal perturbations. We have represented in figures B.3(a,c,e) the optimal perturbations computed on the ROMs with $n_c = 100, 1600$ and on the full system respectively and for the horizon times $T_{opt} = 160, 200$ and 265 respectively. Note that we choose the horizon times at which the first energy gain maxima are found, see figure B.2(c). Additionally, we have represented in figures B.3(b,d,f) their associated optimal responses at their respective horizon times, namely $\boldsymbol{q}(T_{opt})$. Note that we computed these optimal responses on the global modes basis by equation (B.4). We observe that increasing the number of global modes (and the size of the ROMs) allows to move the optimal perturbation in the upstream direction. On the other hand the optimal flow response at the time T_{opt} is captured for very low-order models. These results are more encouraging than those found on the backward-facing step flow. This improvement is clearly attributed to the better spatial location of the global modes, see figure B.2(b).

B.4 Case of a square cavity flow

Based on the results on the previous cavity flow, we found interesting to study a case where the computation domain is much smaller. We thought that this would, in some way, enforce a better spatial location for the least damped global modes. As a result we investigated the present flow configuration. The geometry consists of a square cavity of length $l_e = 1$ and the computational boundaries are chosen to be located at $x = -1.2$, $x = 2.5$ and $y = 0.5$, see figure B.4(a). A Blasius boundary layer profile, with

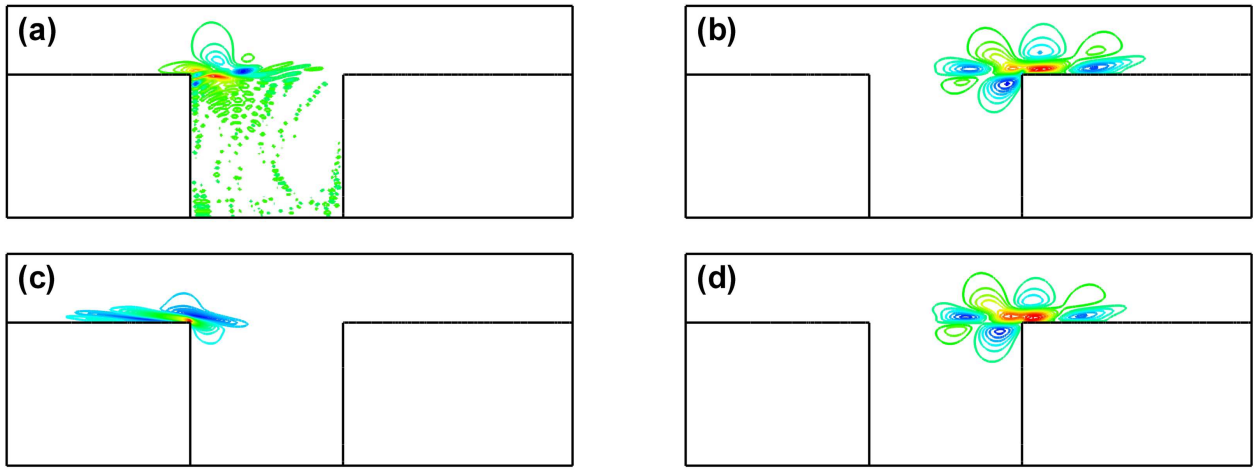


Figure B.5: (a,c) Optimal perturbations computed on the ROM with $n_c = 600$ and on the full system respectively. (b,d) represent their corresponding optimal responses, which are reached for $T = 1.8$ and 2 respectively. Flow structures are displayed by their longitudinal velocity.

a displacement thickness $\delta^* = 0.0172$ and a free-stream velocity $U_\infty = 1$, is prescribed at the inflow. The Reynolds number is then defined by $Re = U_\infty l_e/\nu$ and is fixed to $Re = 2000$. A wall condition is imposed on the lower boundaries and a symmetry condition for the upper boundary. Finally, the free outflow condition $p\boldsymbol{n} - Re^{-1}(\nabla \boldsymbol{u}) \cdot \boldsymbol{n} = 0$ is prescribed at the outlet, where $\boldsymbol{n}$ stands for the outward normal unitary vector of the boundary. The resulting base flow is represented in figure B.4(a) by its streamlines and longitudinal velocity. Note that the number of degrees of freedoms is $n \approx 390\,000$ stemming from about 86 000 triangles.
The results are now quickly presented. For simplicity, they have been exposed by analogous plots to those of the previous cavity flow. The flow eigenspectrum is shown in figure B.4(b) and the evolution of the optimal energy gains in figure B.4(c). This time, the eigenspectrum displays a large number of modes which are spatially localized within the shear layer. In particular, these modes may be identified in figure B.4(b) by their streamwise mean position x_c ranging from 0.2 to 1. Furthermore, the optimal energy gain computed on a ROM with $n_c = 600$ is only one order of magnitude below the original one. As expected, the performance of the ROMs based on global modes are better in this flow configuration. For the sake of completeness, we have represented in figure B.5 the optimal perturbations and optimal responses of the flow computed on both the ROM with $n_c = 600$ and the full system.

B.5 Case of a lid-driven cavity flow

The last configuration investigated here is the lid-driven cavity flow. This flow may be called a "closed" flow as opposed to "open" flows in the sense that it has no inflow and no outflow. Interestingly, the non-normality of the Navier-Stokes operator is widely higher for open flows than for closed flows. Consequently, we expect that the global modes of the lid driven cavity flow may have a better conditioning. We conclude this appendix by showing that the global modes are indeed far more efficient to model the transient dynamics in the case of a closed flow.
The geometry consists of a square of length $l_e = 1$, see figure B.6(a). The cavity is driven by imposing the velocity $u = U_\infty = 1$ and $v = 0$ on the upper boundary and the three other boundaries are walls ($u = 0$ and $v = 0$). The Reynolds number is defined by $Re = U_\infty l_e/\nu$ and is fixed to $Re = 2000$. The streamlines and velocity magnitude of the resulting base flow is represented in figure B.6(a). The considered mesh has $n \approx 180\,000$ degrees of freedom and is composed of about 41 000 triangles.
The results are also quickly exposed by using analogous plots to those of the two previous cavity flows. Similarly, the eigenspectrum of the flow is represented in figure B.6(b) while the evolution of the optimal

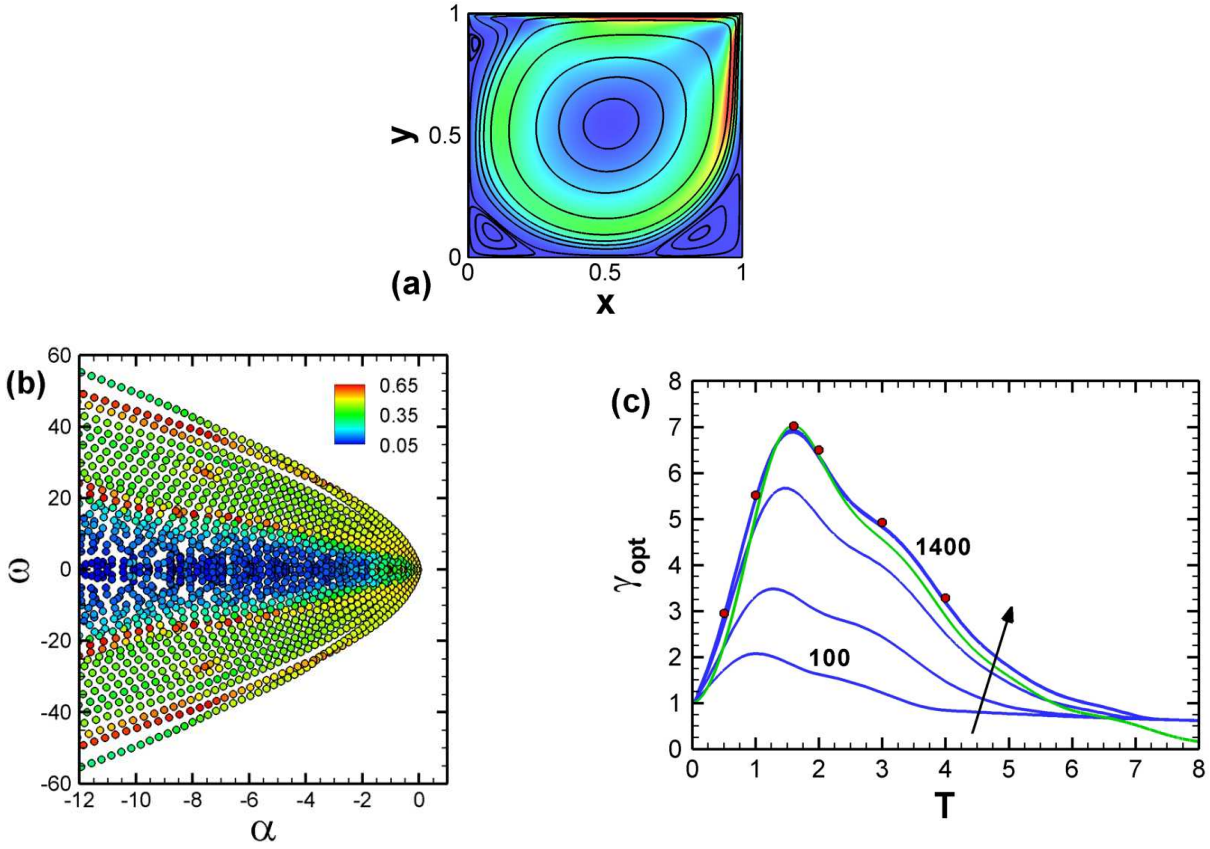


Figure B.6: (a) Streamlines and velocity magnitude of the base flow solution at $Re = 2000$. (b) Eigenspectrum of the flow. The eigenvalues are colored by the kinetic energy barycenter x_c of their associated eigenvector. (c) Evolution of the optimal energy gain γ_{opt} as a function of the horizon time T. The red dots represent six solutions computed on the full system for $T = 0.5, 1, 1.6, 2, 3$ and 4. The green line depicts the time-integrated optimal perturbation computed for $T = 1.6$ on the full system. The blue lines stand for the optimal perturbations computed on the ROMs with $n_c = 100, 200, 400, 800, 1200$ and 1400.

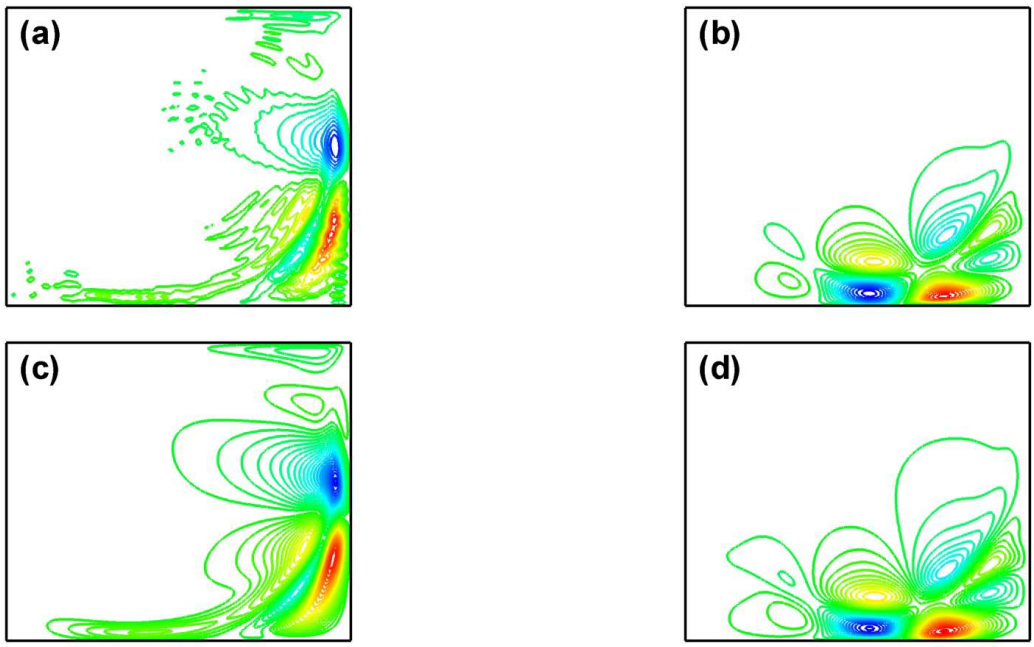


Figure B.7: (a,c) Optimal perturbations computed on the ROM with $n_c = 1400$ and on the full system respectively. (b,d) represent their corresponding optimal responses, which are both reached for $T = 1.6$. Flow structures are displayed by their longitudinal velocity.

energy gains are shown in figure B.4(c). We eventually found in this flow configuration a convergence of the optimal energy gains. Indeed, the ROMs with $n_c = 800, 1200$ and 1400 lead to nearly identical optimal energy gains which are also very close to those computed on the full system. To further confirm the accuracy of the largest ROMs, we have represented in figure B.7 the optimal perturbations and associated optimal responses computed on both the ROM with $n_c = 1400$ and the full system. This time, not only the optimal flow response, but also the optimal perturbation are well captured by the ROM. We conclude that the basis of the least stable global modes are far more efficient to model the transient dynamics of perturbations in the case of such a closed flow.

Appendix C

Article on flow control

Abstract

We consider a two-dimensional incompressible flow over a rounded backward-facing step at Reynolds number $Re = 600$. This configuration is characterized by a detachment of the flow close to the step followed by a recirculation zone; even though the flow is globally stable, perturbations are amplified as they are convected along the shear-layer. The presence of upstream random white noise renders the flow unsteady with a broadband spectrum. The article aims at suppressing the unsteadiness by an anti-phase controller, which converts a shear-stress measurement taken from a wall sensor into a control law that is fed into an actuator. A comprehensive study of various components of closed-loop control design — covering sensor placement, choice and influence of the cost functional, accuracy of the reduced-order model, compensator stability and performance — shows that a successful control of this flow requires a judicious balance between estimation speed and estimation accuracy and between stability limits and performance requirements. The inherent amplification behavior of the flow can be reduced by an order of magnitude, if the above-mentioned constraints are observed. In particular, for good perfomance of the controller, it has been argued that the actuator (resp. the performance sensor) should be placed in regions where the optimal forcings (resp. optimal responses) are strong while the estimation sensor should be placed upstream near the actuator to favour estimation speed. Also, if high performance compensators are sought, a very accurate reduced-order model is required especially for the dynamics between the actuator and the estimation sensor: very minute errors even at low-energetic and high frequencies may render the large-scale compensated linearized simulation unstable. Finally, coupling the linear compensator to nonlinear simulations shows a gradual deterioration in control performance as the amplitude of the noise increases.

Closed-loop control of unsteadiness over a rounded backward-facing step

By ALEXANDRE BARBAGALLO[1,2], GREGORY DERGHAM[3], DENIS SIPP[1], PETER J. SCHMID[2] AND JEAN-CHRISTOPHE ROBINET[3]

[1]ONERA-the French Aerospace Lab, 8 rue des Vertugadins, 92190 Meudon, France
[2]Laboratoire d'Hydrodynamique (LadHyX), CNRS-Ecole Polytechnique, 91128 Palaiseau, France
[3]DynFluid Laboratory, Arts et Metiers ParisTech, 151 Bd. de l'Hopital, 75013 Paris, France

(Received 5 May 2011)

We consider a two-dimensional incompressible flow over a rounded backward facing step at Reynolds number $Re = 600$. This configuration is characterized by a detachment of the flow close to the step followed by a recirculation zone; even though the flow is globally stable, perturbations are amplified as they are convected along the shear-layer. The presence of upstream random white noise renders the flow unsteady with a broadband spectrum. The article aims at suppressing the unsteadiness by an anti-phase controller, which converts a shear-stress measurement taken from a wall sensor into a control law that is fed into an actuator. A comprehensive study of various components of closed-loop-control design — covering sensor placement, choice and influence of the cost functional, accuracy of the reduced-order model, compensator stability and performance — shows that a successful control of this flow requires a judicious balance between estimation speed and estimation accuracy and between stability limits and performance requirements. The inherent amplification behavior of the flow can be reduced by an order of magnitude, if the above-mentioned constraints are observed. In particular, for good perfomance of the controller, it has been argued that the actuator (resp. the performance sensor) should be placed in regions where the optimal forcings (resp. optimal responses) are strong while the estimation sensor should be placed upstream near the actuator to favour estimation speed. Also, if high performance compensators are sought, a very accurate reduced order model is required especially for the dynamics between the actuator and the estimation sensor: very minute errors even at low-energetic and high frequencies may render the large-scale compensated linearized simulation unstable. Finally, coupling the linear compensator to nonlinear simulations shows a gradual deterioration in control performance as the amplitude of the noise increases.

1. Introduction

Many industrial fluid devices are afflicted by undesirable flow behavior — such as unsteadiness, separation, instabilities, and transition to turbulence — which limits performance, endangers safe operation or is detrimental to structural components. Flow control is quickly becoming a key technology in engineering design to overcome inherent limitations, to advance into unexplored parameter regimes, to extend safety margins and to ensure operation under optimal conditions. A prototypical and much-studied example is the compressible flow over a cavity which is characterized by instabilities that manifest themselves in a buffeting motion, in induced drag (Gharib & Roshko 1987) and in in-

tense noise emission (see Rossiter 1964). In air intakes of aircraft engines separated flow can act as an amplifier for incoming perturbations causing unsteadiness which in turn results in loss of performance and material fatigue. Transitional and turbulent boundary layers have long attracted attention from the flow control community (see, e.g., Joslin 1998; Saric *et al.* 2003; Kim 2003; Boiko *et al.* 2008; Archambaud *et al.* 2008), mainly due to their ubiquity in vehicle aerodynamics and their central role as the source of skin friction. For all three examples, flow control techniques that effectively eliminate instabilities, efficiently reduce noise amplification or successfully diminish drag are essential in maintaining desired flow conditions.

Control strategies greatly vary in complexity, in expended energy, but also in their ability to achieve prescribed control objective under realistic conditions. Passive control devices, which aim at modifying the mean flow, are popular in many industrial applications due to their simplicity and reliability. For example, vortex generators can be effective in energizing boundary layers and thus preventing or delaying flow separation (see Lin 2002; Choi 2008). Constant blowing or suction at the wall has similar effects, but the amount of required energy to modify the mean flow is generally high, which is acknowledged as the principal drawback of passive control devices. An alternative strategy directly targets the perturbation dynamics without altering the mean flow. This active approach has received significant interest in industry and academia, and feedback control methods have been developed and applied to a great many generic flow configurations.

The design procedure of flow control laws critically depends on the nature of the flow to be controlled. Oscillator-type flows which are defined by a global instability resulting in self-sustained oscillatory fluid behavior are more easily controlled, since the flow is dominated by a limited number of structures of well-defined frequencies. Sensitivity to noise is comparatively low, and the estimator and controller can simply reconstruct the flow state from measurements and act upon it according to the control objective. A second type of flow behavior, referred to as noise-amplifiers, is substantially more challenging to control. This type of flow is globally stable but is characterized by a strong propensity to amplify noise and a broadband spectrum of responsive frequencies. These characteristics make the flow and the control performance highly sensitive not only to physical noise sources and uncertainties, but also to methodological approximations, modeling inaccuracies and truncation errors that inevitably arise during the estimator and control design. The propagation of small perturbations, whether of physical or computational origin, is appropriately tracked and quantified by frequency-based transfer functions which reveal preferred frequencies or confirm the successful reduction of the flow's amplification potential.

Optimal flow control techniques have been widely applied for active control purposes. In particular, the linear quadratic Gaussian (LQG) framework has been adopted for the control of small-amplitude perturbations in oscillator and amplifier flows. Examples of oscillator flows include, among others, the supercritical flow over a shallow or deep cavity (Åkervik *et al.* 2007; Barbagallo *et al.* 2009), the flow over a shallow bump (Ehrenstein *et al.* 2010) and the flow past a flat plate (Ahuja & Rowley 2010). In all cases, stabilization of the flow by feedback control strategies could be accomplished. Amplifier flows are dominated by convective and transient processes, and successful control is defined by a marked reduction of the flow's inherent amplification potential. Control of amplifier flows using LQG techniques has first been attempted for very idealized geometries (see, e.g., Joshi *et al.* 1997; Bewley & Liu 1998; Chevalier *et al.* 2007), namely, in simple, one-dimensional configurations. For more complex and higher-dimensional flows, direct application of the LQG-framework becomes prohibitively expensive, and reduced-order models have to be introduced for the practical design of the compensator. LQG-based

compensators using reduced-order models have been applied by Bagheri *et al.* (2009) to control the amplification of perturbations in a spatially developing boundary layer and by Ilak & Rowley (2008) to control transitional channel flow. In Bagheri & Henningson (2010), strong emphasis has been put on the model reduction technology; in particular, it has been demonstrated that the reduced-order model had to accurately capture the input-output behavior between actuators and sensors to ensure a positive compensator performance (Kim & Bewley 2007). Despite first successful attempts at applying LQG-feedback control to amplifier flows, many questions remain open about the design and practical implementation of compensators for this type of flow. Owing to the flow's tendency to transiently amplify perturbations, sensitivity becomes the key concept in the design and performance evaluation of compensators: sensitivity to sensor and actuator placement, sensitivity to the accuracy of the reduced-order model, sensitivity to nonlinear effects. Some of these issues have been addressed using an idealized (parallel base flow) model problem in Ilak (2009); a comprehensive analysis of closed-loop control for amplifier flows, however, is missing.

The goal of the present study is an identification of the various environmental and procedural factors and the assessment of their influence on the performance of the compensator for a specific amplifier flow: the case of a two-dimensional, laminar flow over a rounded backward-facing step (see Marquet *et al.* 2008; Blackburn *et al.* 2008; Dergham *et al.* 2011). This configuration is characterized by a detachment of the flow close to the step followed by a recirculation zone; even though the flow is globally stable, perturbations are amplified as they are convected along the shear-layer. This is due to a Kelvin-Helmholtz-type instability which creates a convectively unstable region, extending approximately from the detachment to the reattachment point. If upstream residual noise is present, for example white-Gaussian-noise, then the flow is unsteady with a broadband spectrum in the bubble. The control objective consists in suppressing this unsteadiness thanks to an anti-phase controller, which converts a measurement signal obtained from a wall-sensor into a control law that is fed into an actuator. We believe that the results obtained for this specific configuration also hold, at least qualitatively, for other convectively unstable flows, such as, e.g., a supercritical boundary layer developing over a flat plate subject to Tollmien-Schlichting instabilities. Discussion about similarities and differences between the two configurations are postponed to §2.

The present study is structured as follows. After a brief description of the flow configuration, its noise amplification behavior and the basic principles of LQG-control and model reduction (§2), we start by considering the estimation problem (§3), address sensor placement and estimation speed and establish the noise-to-signal ratio as a critical parameter. In §4 the controller will be introduced, performance limitations of the compensator will be discussed and the influence of the choice of control objective will be assessed. §5 will present the application of the LQG-controller to linearized numerical simulations; specifically the sensitivity to model inaccuracies and its relation to stability margins for the compensated system will be treated. In §6 we apply linear control to a nonlinear simulation and discuss the validity range of the linear compensator. A summary of results and conclusions are given in §7.

2. Configuration and mathematical model

2.1. *Flow configuration and governing equations*

We study the laminar and incompressible flow over a two-dimensional rounded backward-facing step which is sketched in figure 1 together with the geometric measures, the base

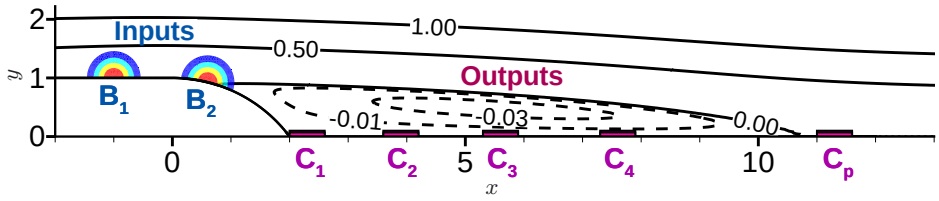


FIGURE 1. Sketch of the geometry for flow over a rounded backward-facing step, showing the stream-function of the base flow at $Re = 600$ (dashed-values refer to negative values). The stream-function is set to zero at the lower boundary of the computational domain. The upstream, downstream and top boundaries are respectively located at $x = -20$, $x = 100$ and $y = 20$. A typical mesh yields $n \approx 360\,000$ degrees of freedom from about 90 000 triangles. The positions of the input ($\mathsf{B}_{1,2}$) and output ($\mathsf{C}_{1,2,3,4,p}$) devices are also shown.

flow stream-function and the setup of control inputs and sensor outputs. Only a reduced part of the computational domain is shown. The step height h and the inflow velocity U_∞ are chosen as the characteristic length and velocity scales of the problem. The rounded part of the step consists of a circular arc extending from $(x = 0, y = 1)$ to $(x = 2, y = 0)$. The flow enters the computational domain from the left (at $x = -20$) with a constant streamwise velocity. A free-slip condition is imposed on the upstream part of the lower boundary ($-20 \leqslant x \leqslant -2, y = 1$) beyond which a laminar boundary layer starts to develop; no-slip conditions are enforced on the remaining lower boundary given by $-2 \leqslant x \leqslant 100$. On the top part of the computational domain, at $y = 20$, a symmetry condition is implemented, and a standard outflow condition is prescribed at the outlet ($x = 100$).

The Reynolds number based on the step height and inflow velocity is chosen as $Re = 600$ and held constant throughout our study. For this Reynolds number, the flow separates at $x \approx 0.6$ and reattaches at $x \approx 11$, forming an elongated recirculation bubble. The displacement thickness of the incoming boundary layer at $x = 0$ is equal to $\delta^* \approx 0.082$ which yields a Reynolds number based on the displacement thickness of $Re_{\delta^*} \approx 49.2$. The base flow, a solution of the nonlinear steady Navier-Stokes equations, is visualized by streamlines in figure 1.

Flow over a rounded backward-facing step is a prototypical example for an amplifier flow since small upstream perturbations may be selectively amplified in the shear layer due to a Kelvin-Helmholtz instability (see next section for details). Characteristic unsteadiness arises from low-level noise via a linear amplification mechanism, which subsequently saturates nonlinearly once sufficiently high amplitudes have been reached. It is the goal of this article to devise and assess an active feedback control strategy that decreases the convective amplification of random perturbations. This strategy is designed for and operates within the linear regime which justifies using the Navier-Stokes equations linearized about the base flow as a mathematical model. The governing equations are spatially discretized using finite elements of Taylor-Hood type (P2-P2-P1) and implemented using the FreeFem++ software (see Hecht *et al.* 2005). In matrix form, these read

$$\mathsf{Q}\frac{d\mathsf{X}}{dt} = \mathsf{A}\mathsf{X} \tag{2.1}$$

where X denotes the state vector containing the velocity and pressure fields, A represents the Navier-Stokes operator linearized around the base-flow (shown in figure 1) and Q stands for the mass matrix, which simultaneously defines the perturbation kinetic energy according to $E = \mathsf{X}^*\mathsf{Q}\mathsf{X}$.

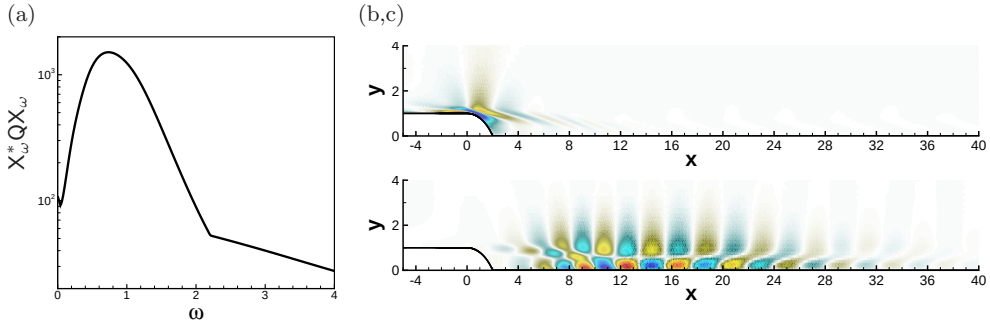


FIGURE 2. (a) Optimal energy response of the linear system to a harmonic forcing of frequency ω. (b) Optimal forcing visualized by contours of the streamwise velocity. (c) Optimal response visualized by contours of the streamwise velocity.

2.2. *Noise amplification behavior*

For the chosen Reynolds number of $Re = 600$ the flow is globally stable and the matrix A does not show any unstable eigenvalues. Yet, sustained unsteadiness may arise from the continuous excitation of the flow by upstream noise. This noise amplifier behavior can be analyzed and quantified in terms of the optimal harmonic forcing and its response in the frequency domain (see Alizard *et al.* 2009; Sipp *et al.* 2010). For a given frequency ω, a periodic forcing of the form $\mathsf{F}_\omega \exp(i\omega t)$ yields a corresponding response $\mathsf{X}_\omega \exp(i\omega t)$ where X_ω is given by $\mathsf{X}_\omega = (i\omega\mathsf{Q} - \mathsf{A})^{-1}\mathsf{Q}\mathsf{F}_\omega$. The optimal forcing is defined as the forcing F_ω of unit energy (i.e. $\mathsf{F}^*_\omega \mathsf{Q} \mathsf{F}_\omega = 1$) which maximizes the energy of the response. The corresponding response X_ω is referred to as the optimal response; its energy $\mathsf{X}^*_\omega \mathsf{Q} \mathsf{X}_\omega$ is the optimal energy gain due to external forcing at a prescribed frequency ω. We will use the subscript $_\omega$ to indicate quantities defined in the frequency domain.

In figure 2(a), the optimal energy gain is presented as a function of the forcing frequency. The semi-logarithmic graph shows a parabolic curve with the highest response to forcing around a frequency of $\omega = 0.8$. For frequencies above $\omega \approx 2$, the energy gain decays monotonically. In figures 2(b) and (c), the spatial shapes of the optimal forcing and response (real part only) taken at the highest energy gain are visualized by contours of the streamwise component. In accordance with the convective nature of the flow, we observe that the optimal forcing is concentrated upstream near the step while the associated response is located further downstream. Hence, a noise source situated near the rounded step can efficiently trigger perturbations whose maximum amplitudes are attained near $x \approx 12$. Based on this result, we model the noise as a Gaussian-shaped momentum forcing located at $(x = -1, y = 1)$ with a width of 0.6 and a thickness of 1. After discretization, this forcing appears in form of the matrix B_1 in the following linear system

$$\mathsf{Q}\frac{d\mathsf{X}}{dt} = \mathsf{A}\mathsf{X} + \mathsf{Q}\mathsf{B}_1 w(t) \tag{2.2}$$

with $w(t)$ describing the temporal behavior of the noise. For the sake of simplicity, the noise will be taken as white-in-time with zero mean $\langle w \rangle = 0$ and variance $\langle w^2 \rangle$ denoted by W^2.

In continuing to set up our flow control problem, an appropriate objective or cost functional has to be specified. To this end, two quantities will be considered. The first quantity consists of the shear stress measured at the wall and is computed following $m_p = \mathsf{C}_p\mathsf{X} = \int_{x=11}^{x=11.6} \partial_y u' \, dx$. The placement of this sensor has been motivated by the

location of maximum response of the flow to harmonic excitation (see figure 2(c)). Note that for higher frequencies ($\omega \geqslant 2$), the optimal responses move upstream toward the detachment point and are not seen anymore by m_p; for lower frequencies, the spatial support of the optimal responses extends both upstream and downstream (the most energetic region staying near the reattachment point). Hence, sensor m_p will detect all frequencies of the perturbation satisfying $\omega \leqslant 2$. The second quantity of interest is the total kinetic perturbation energy contained in the entire domain; it is given by $E = \mathsf{X}^*\mathsf{Q}\mathsf{X}$. The control to be designed will aim at diminishing either m_p or E. It is interesting to note that, under random forcing, the two quantities of interest, m_p and E, display a frequency response strikingly similar to the one given in figure 2(a); the spatial structure of the stochastic response resembles the one given in figure 2(c) (the reader is referred to figure 8 for verification).

2.3. *Linear Quadratic Gaussian (LQG) control*

A closed-loop control strategy is considered in order to weaken or suppress the amplification of perturbations. In contrast to open-loop control strategies, this method extracts information from the system via measurements which is then processed to apply real-time actuation. This technique allows flow manipulation with rather low expended energy and permits the application and adaptation of control laws to a variety of flow situations, provided the model is representative of and robust to physical and parametric changes. The approach taken in our study is based on a compensator designed within the Linear Quadratic Gaussian (LQG) control framework (see Burl 1999). The actuator through which control efforts are exerted on the flow consists of a body force acting on the vertical momentum component; the location, shape and type of the actuator is summarized in the matrix B_2 (see figure 1). The control law $u(t)$ which describes the temporal behavior of the actuator is based on real-time measurements of the flow from sensors located at various positions along the wall. These sensors extract either shear stress or pressure information. The governing system of equations, including the actuators and sensors, can be cast into the familiar state-space form

$$\mathsf{Q}\frac{d\mathsf{X}}{dt} = \mathsf{A}\mathsf{X} + \mathsf{Q}\mathsf{B}_1 w + \mathsf{Q}\mathsf{B}_2 u, \qquad m = \mathsf{C}\mathsf{X} + g(t) \tag{2.3}$$

where $g(t)$ is a zero-mean measurement noise of variance G^2, which contaminates the measure m. The link between the measurement signal m and the actuation law u is provided by the compensator. Figure 3 presents a sketch of a typical LQG-control setup, including the system to be controlled as well as the two components of the compensator: the estimator and the controller. The module labeled "plant" represents our fluid system whose flow characteristics we wish to modify; it is given mathematically by (2.3). The plant depends on the initial condition $\mathsf{X}(t_0)$, the noise input $w(t)$ and the control law $u(t)$ and provides as an output the state vector X. A measurement signal m can be extracted at all times from the state vector which is then passed to the compensator. In a first step, the estimator will reconstruct an estimated state $\hat{\mathsf{Y}}(t)$ from the measurement m which is, in a second step, used by the controller to compute the control law $u(t)$. More details about the design of the estimator and the controller will be given below.

It is important to stress that the placement of the actuator and sensor is critical for the success of closed-loop control. In our case, the actuator is positioned near the separation point (see control input B_2 in figure 1) which corresponds to the location where the optimal forcing structures are most prominent whatever the frequency (see figure 2(b) which illustrates the case $\omega = 0.8$). This placement optimally exploits the sensitivity of the flow to external forcing and suggests that low-amplitude control at this location

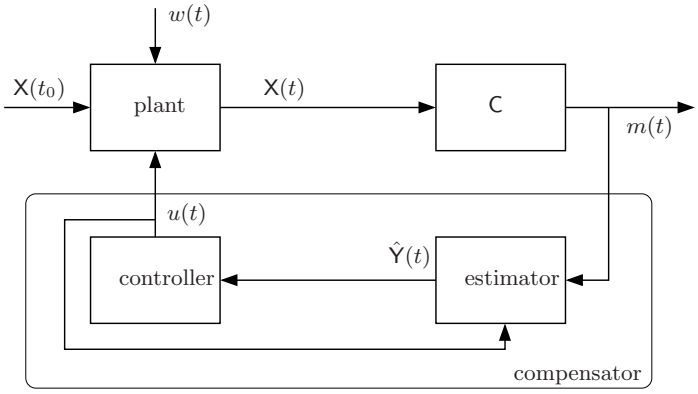


FIGURE 3. Sketch of a typical LQG-control setup.

may exert sufficient influence on the flow behavior to accomplish our control objective. In other words, the chosen actuator location should ensure low control gains.

Analogously, the placement of the sensor requires care and thought. Commonly, measurements are taken at locations where the flow feature we wish to suppress is particular prevalent. Recalling the spatial structure of the most amplified flow response to optimal forcing (see figure 2(c)), this would suggest a sensor placement downstream of the reattachment point near $x \approx 12$. Nevertheless, we will demonstrate that this particular choice does not yield an efficient and effective closed-loop control, and we will methodically explore the estimator performance based on sensors placed further upstream in the recirculation bubble (see figure 1). In particular, four discrete sensor locations, denoted by $\mathsf{C}_{1,2,3,4}$, will be assessed; these are distinct from the performance sensor C_p.

More generally, to ensure low control gains and a physically relevant control objective the actuator should be placed at the transition location from the stable to the convectively unstable flow regime (branch I) and the performance sensor C_p at the transition location from convectively unstable to the stable flow regime (branch II). Indeed, in a convectively unstable flow, the instability is cumulative in the downstream direction: branch I is the point where an action has largest effect on the perturbation while branch II is the point where the perturbation is strongest. The location of branch I and branch II usually depend on the frequency of the perturbation; in the present flow over a backward-facing step, however, branch I and branch II are located for all frequencies near the flow separation and reattachment points (not shown here). Hence, a single actuator and a single performance sensor (plus one estimation sensor) are sufficient to stabilize the flow for all frequencies. In the case of a convectively unstable boundary layer developing over a flat plate, the location of branch I and branch II strongly depend on the perturbation frequency. Hence, multiple input-output triplets (each consisting of actuator, performance sensor, estimation sensor) are necessary, where each triplet is designed to stabilize the flow in a bounded area in the downstream direction and in a restricted frequency band.

2.4. *Reduced-order model based on proper orthogonal decomposition*

The design of the estimator and controller involves the numerical solution of two Riccati equations for the Kalman and control gain, respectively. The numerical effort is proportional to the dimension of the system matrix A, which makes the direct solution of the Riccati equation excessively expensive or even impossible. It is thus necessary and common practice to substitute the full system by an equivalent system of considerably smaller dimensions and to compute the two gains based on this reduced-order model

of the flow. A standard technique to arrive at a reduced-order model of the flow uses a Galerkin projection of the governing equations onto proper orthogonal decomposition (POD) modes (see Sirovich 1987). This method proves to be efficient (see Bagheri *et al.* 2009; Barbagallo *et al.* 2009; Bagheri & Henningson 2010) in capturing the main characteristics of the original system required for closed-loop control, namely the dynamics between the inputs (given by B_1 and B_2) and the outputs (given by sensors $\mathsf{C}_{1,2,3,4}$ and the control objectives E and C_p). The governing equation for the reduced-order model is similar to (2.3) and is given by

$$\frac{d\hat{\mathsf{X}}}{dt} = \hat{\mathsf{A}}\hat{\mathsf{X}} + \hat{\mathsf{B}}_1 w + \hat{\mathsf{B}}_2 u, \qquad m = \hat{\mathsf{C}}\hat{\mathsf{X}} \tag{2.4}$$

where the superscript $\hat{\cdot}$ indicates reduced quantities. The energy of the state E is simply given by $\|\hat{\mathsf{X}}\|^2 = \hat{\mathsf{X}}^*\hat{\mathsf{X}}$ since the projection basis is orthonormal with respect to the energy inner product. The use of a reduced-order model decreases the dimension of our system from $\mathcal{O}(10^6)$ to ~ 150 degrees of freedom and thus allows the application of standard direct algorithms for LQG-control design. The choice of POD modes as a projection basis has essentially been motivated by the requirement of capturing the energy output. A standard balanced-truncation technique (see Bagheri *et al.* 2009; Barbagallo *et al.* 2009; Ahuja & Rowley 2010) is not able to cope with large-dimensional outputs, unless the "output projection technique" is used (Rowley 2005); this technique however is beyond the subject of this study. In appendix A, we show that all transfer functions from w and u to $m_{1,2,3,4,p}$ and $E(t)$ can be captured with an $\mathcal{H}_2$-error of less than 1%. This error is quite small which should ensure at first glance that a model study of compensated systems with the "ROM as a plant" is representative of compensated systems with the "linearized DNS as a plant". This viewpoint is assumed in §3 for the estimation problem and in §4 for the complete estimation-control problem. The above mentioned error of 1% is nonetheless appreciable. §5 is then devoted to the question: is this 1% error between the linearized simulation and the POD model really sufficiently small to insure in all cases that a model study with a ROM is representative of the large-scale compensated case? One of the objectives of the present article is precisely to quantitatively assess the influence of the ROM's quality on the performance of the compensator. Such questions become even more pertinent under more realistic conditions (e.g., in experimental realizations) as errors inevitably corrupt the quality of a ROM. It is essential to understand the minimum requirements on the quality of the ROM for acceptable or required control performance; it is equally essential to know how the compensated system will fail, if these requirements are not met. In addition, should specific inputs and/or outputs be more accurately captured than others? In what follows we will try to address some of these points.

3. Estimation and sensor placement

As a first step of the full control design and performance evaluation process, we concentrate on the estimator, in particular its performance with respect to the location of the sensors.

3.1. *Presentation of the estimator*

In general, the estimator's task is the approximate reconstruction of the full state vector using only limited information from the measurement. This approximate state vector will then be used by the controller to determine a control strategy that accomplishes our objective. The estimated state $\hat{\mathsf{Y}}$ is assumed to satisfy a set of equations similar to the

one governing the original system (2.4). We have

$$\frac{d\hat{\mathsf{Y}}}{dt} = \hat{\mathsf{A}}\hat{\mathsf{Y}} + \hat{\mathsf{B}}_2 u(t) - \hat{\mathsf{L}}(m - \hat{\mathsf{C}}\hat{\mathsf{Y}}) \tag{3.1}$$

where the original noise term $\hat{\mathsf{B}}_1 w(t)$ has been replaced by the forcing term $-\hat{\mathsf{L}}(m-\hat{\mathsf{C}}\hat{\mathsf{Y}})$. The latter term represents the difference between the true measurement signal $m(t) = \hat{\mathsf{C}}\hat{\mathsf{X}}$ and the estimated measurement signal $\hat{\mathsf{C}}\hat{\mathsf{Y}}$ and is applied as a forcing term premultiplied by $\hat{\mathsf{L}}$. This term is to drive the estimated state $\hat{\mathsf{Y}}$ toward the true state $\hat{\mathsf{X}}$. In the forcing term, the so-called *Kalman gain* $\hat{\mathsf{L}}$ can be computed from an optimization problem in which the cost functional is taken as the error between the full and estimated state, i.e., $\hat{\mathsf{Z}} = \hat{\mathsf{X}} - \hat{\mathsf{Y}}$, and is subsequently minimized. The resulting optimality condition yields a Riccati equation, from which the Kalman gain $\hat{\mathsf{L}}$ follows (see Burl 1999). Commonly, the energy of the estimation error is formulated in the time domain; it will prove advantageous in our case, though, to recast the energy in the frequency domain. Using Parseval's theorem we obtain for the objective functional

$$\bar{\mathsf{Z}} = \int_{-\infty}^{\infty} \|\hat{\mathsf{Z}}_\omega\|^2 \, d\omega \tag{3.2}$$

where $\hat{\mathsf{Z}}_\omega$ denotes the Fourier-transform of the error $\hat{\mathsf{Z}}$. Two sources of noise — both assumed as white in time — are taken into account in the derivation and solution of the Riccati equation: a plant noise $w(t)$ of variance W^2 driving the dynamics of the original system (2.4) and a measurement noise $g(t)$ with variance G^2 contaminating the measurement $m(t)$. The ratio of the two standard deviations, i.e., G/W, can be taken as a parameter that governs the speed of the estimation process, but can also be interpreted as the noise-to-signal ratio of the sensor. For example, considering a constant standard deviation W of the plant noise, the parameter G/W represents a quality measure of the sensor. Large values of G/W indicate that the measurement noise $g(t)$ is too high to ensure a correct signal; the corresponding Kalman gain $\hat{\mathsf{L}}$ tends to zero. Consequently, the forcing term in (3.1) has a negligible effect on the system which, in turn, leads to a poorly performing estimator. This parameter regime is referred to as the small gain limit (SGL). Contrary to the small gain limit, for $G/W \ll 1$ the corruption of the measurement signal by noise is low compared to the stochasticity arising from the system itself; as a consequence, the estimation process becomes highly effective due to the substantial Kalman gains. This parameter regime, referred to as the large gain limit (LGL), comprises the most performing estimators for a given configuration.

By construction, the performance of the estimator crucially depends on details related to the measurement signal, and the type of sensor (in terms of its noise-to-signal ratio) as well as its location have to be chosen judiciously if overall success of the closed-loop control effort is to be expected. In what follows, we will consider four sensors $\hat{\mathsf{C}}_{1,2,3,4}$ that are identical in type but placed at four different positions within the recirculation bubble, and assess their capability of efficiently estimating the flow state. Special emphasis will also be directed toward the quantitative influence of the noise-to-signal ratio G/W introduced above.

3.2. *Performance of the estimator*

In this paragraph, we further elaborate on estimating the flow state $\hat{\mathsf{X}}$ governed by (2.4). The estimation problem is decoupled from the control problem (see Burl 1999). For this reason, we can set the control law to zero ($u(t) = 0$) without loss of generality and continue our study of the estimation problem without actuation. The system is driven

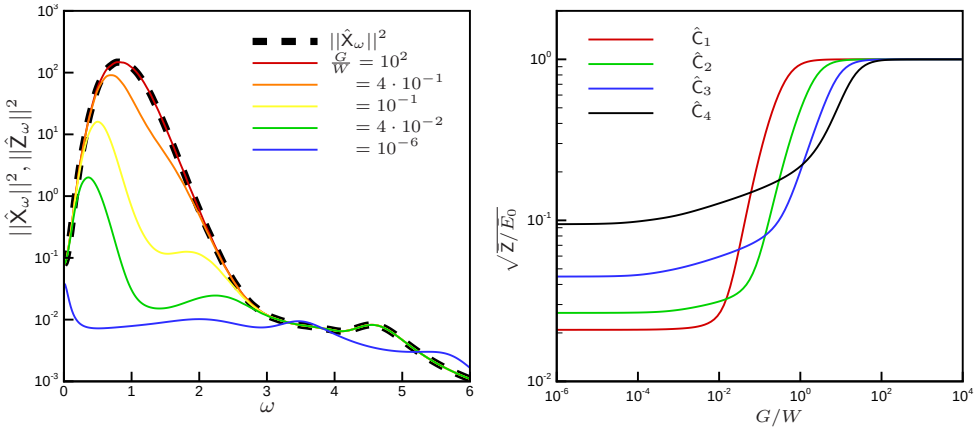


FIGURE 4. Performance of the estimator versus frequency using four different sensors and selected values of the estimation (noise-to-signal) parameter G/W. (a) sensor $\hat{\mathsf{C}}_1$. Normalized performance of the estimator, integrated over all frequencies, versus the estimation parameter G/W for four different sensor locations; (b) measuring shear-stress.

by white noise represented by $w(t)$; but, owing to the linearity of (2.4), the performance of the estimator can equivalently be studied by considering harmonic forcings $w(t) = \exp(i\omega t)$ of a given frequency ω. It is then convenient to reformulate the coupled plant/estimator system in the frequency domain and state the governing equations for the harmonic response as

$$\begin{pmatrix} \hat{\mathsf{X}}_\omega \\ \hat{\mathsf{Y}}_\omega \end{pmatrix} = \begin{pmatrix} i\omega\mathsf{I} - \hat{\mathsf{A}} & \hat{\mathsf{0}} \\ \hat{\mathsf{L}}\hat{\mathsf{C}} & i\omega\mathsf{I} - \left(\hat{\mathsf{A}} + \hat{\mathsf{L}}\hat{\mathsf{C}}\right) \end{pmatrix}^{-1} \begin{pmatrix} \hat{\mathsf{B}}_1 \\ \hat{\mathsf{0}} \end{pmatrix}. \tag{3.3}$$

where, as before, the subscript $_\omega$ indicates variables defined in the frequency domain. The estimation error in frequency space is given as $\hat{\mathsf{Z}}_\omega = \hat{\mathsf{X}}_\omega - \hat{\mathsf{Y}}_\omega$.

In figure 4(a) the frequency dependence of the estimation error $\|\hat{\mathsf{Z}}_\omega\|^2$ is displayed for the shear-stress sensor placed at C_1 (see figure 1) and for selected values of the noise-to-signal parameter G/W. For comparison, the norm of the state vector $\|\hat{\mathsf{X}}_\omega\|^2$, which is similar to figure 2(a), is included as a dashed line. In the small gain limit (red lines), the estimator, as expected, does not succeed in identifying the state, producing an error as large as the norm of the original state. As the parameter G/W decreases though, the estimation process improves due to a less contaminated input from the sensors and the estimation error is reduced — mainly at frequencies where the system reacts strongly to external excitations. As the parameter G/W approaches the large gain limit (blue lines), the error curve eventually converges to the lowest possible value. This curve then defines the best attainable performance for sensor $\hat{\mathsf{C}}_1$.

This general behavior is observable for each of the four sensors (not shown here). The final errors in the large gain limit, however, are not identical for all sensors: the best performance is achieved by sensor $\hat{\mathsf{C}}_1$. As the location of the sensor is moved further downstream in the separation bubble (considering successively the sensors $\hat{\mathsf{C}}_1$, $\hat{\mathsf{C}}_2$, $\hat{\mathsf{C}}_3$ and $\hat{\mathsf{C}}_4$), the frequencies which are naturally amplified by the system are less well predicted. This suggests that the estimator based on $\hat{\mathsf{C}}_1$ will be more efficient in accurately determining the flow state.

An instructive way of assessing the performance of an estimator over all frequencies is

to directly compute the cost functional $\bar{\mathsf{Z}}$ (see eq. (3.2)) normalized by the energy of the state. We thus introduce $\sqrt{\bar{\mathsf{Z}}/\bar{E}_0}$ with $\bar{E}_0 = \int_{-\infty}^{\infty} \|\hat{\mathsf{X}}_\omega\|^2 \, d\omega$ and $\hat{\mathsf{X}}_\omega = (i\omega\mathsf{I} - \hat{\mathsf{A}})^{-1}\hat{\mathsf{B}}_1$. When this quantity is close to 1, the estimation process has failed with a 100% estimation error; the smaller the value, the better-performing the estimator. This quantity is displayed in figure 4(b) versus the estimation parameter G/W for each of the four sensors. The red curve represents the estimator performance based on sensor $\hat{\mathsf{C}}_1$. For this sensor location and for noise-to-signal ratios above 1, the sensor noise prohibits a correct estimation resulting in an estimation error of 100%. As the noise-to-signal ratio diminishes further, the performance of the estimator progressively increases until it reaches the large gain limit for values of G/W less than approximately 10^{-2}. Below this value of G/W, the estimator performs at its optimum. Similar behavior can be observed for the remaining sensor locations given by $\hat{\mathsf{C}}_2$, $\hat{\mathsf{C}}_3$ and $\hat{\mathsf{C}}_4$: the small-gain-limit regime is clearly detectable at high values of G/W. However, the exact values for which the estimator reaches the large gain limit becomes less sharply defined as the sensor location is moved further downstream in the separation bubble. Comparing the performance of estimators based on different sensors, we conclude, in agreement with figure 4, that the performance in the large gain limit is best for the sensor $\hat{\mathsf{C}}_1$ and decreases as the sensor is moved further downstream. It is surprising though that the estimator based on $\hat{\mathsf{C}}_4$, which performs worst in the large gain limit, shows better performance at high values of the noise-to-signal ratio G/W. For example, if we consider the value $G/W = 10^0$, the estimator based on $\hat{\mathsf{C}}_4$ displays a relative error of 20% while the estimator based on $\hat{\mathsf{C}}_1$ still shows an error of 100%. If a constant noise level of the system-generated signal is assumed ($W = \text{const.}$), this implies that the estimator based on $\hat{\mathsf{C}}_4$ can cope with higher levels of measurement noise than the estimator based on $\hat{\mathsf{C}}_1$. In practice, this means that less-quality sensors can be used as long as they are placed further downstream; this point will be discussed further in the next section.

To conclude, the estimation errors are rather small ($< 10^{-1}$) for all sensors in the large gain limit. Yet, the estimator based on $\hat{\mathsf{C}}_1$ is most efficient with a performance measure of $2 \cdot 10^{-2}$, while the estimator based on $\hat{\mathsf{C}}_4$ only reaches a value of 10^{-1} in the large gain limit. At first glance, this difference in performance may seem small and insignificant, but it will be shown below (section § 4) that it nonetheless has a strong influence on the efficiency of the compensator. But first, the next section will offer a physical explanation for the observed loss of estimation performance by analyzing the above results in the time domain rather than the frequency domain.

3.3. *Interpretation in the time domain*

Even though a formulation of the estimation problem in the frequency domain is the proper choice for designing closed-loop control strategies for amplifier flows, it nevertheless remains challenging to attach physical meaning to the frequency-based results; an interpretation of our findings in the time domain seems more intuitive. The main result — the estimator's performance deteriorates as the sensor is gradually moved from the upstream $\hat{\mathsf{C}}_1$ to the downstream $\hat{\mathsf{C}}_4$ position — suggests that the travel time of a perturbation, before it is detected by the sensor, plays a critical role. To validate this proposition, we start by rewriting the estimation performance parameter in the time domain using Parseval's theorem. We obtain $\int_{-\infty}^{\infty} \|\hat{\mathsf{Z}}_\omega\|^2 d\omega = 2\pi \int_0^{\infty} \|\hat{\mathsf{Z}}\|^2 dt$, with $\hat{\mathsf{Z}} = \hat{\mathsf{X}} - \hat{\mathsf{Y}}$ and $\hat{\mathsf{X}}$ and $\hat{\mathsf{Y}}$ satisfying the following system of equations in the time domain

$$\frac{d}{dt}\begin{pmatrix}\hat{\mathsf{X}}\\ \hat{\mathsf{Y}}\end{pmatrix} = \begin{pmatrix}\hat{\mathsf{A}} & \hat{0}\\ -\hat{\mathsf{L}}\hat{\mathsf{C}} & \hat{\mathsf{A}} + \hat{\mathsf{L}}\hat{\mathsf{C}}\end{pmatrix}\begin{pmatrix}\hat{\mathsf{X}}\\ \hat{\mathsf{Y}}\end{pmatrix}, \qquad \begin{pmatrix}\hat{\mathsf{X}}\\ \hat{\mathsf{Y}}\end{pmatrix}_{t=0} = \begin{pmatrix}\hat{\mathsf{B}}_1\\ \hat{0}\end{pmatrix}. \tag{3.4}$$

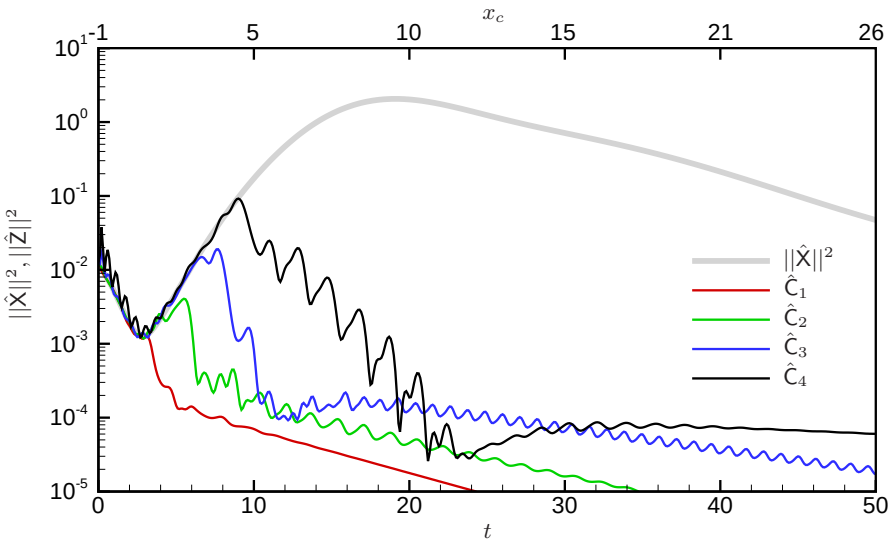


FIGURE 5. Temporal evolution of the energy of the error vector $\|\hat{Z}\|^2$ for four different sensor locations $\hat{C}_{1,2,3,4}$, and energy of the (uncontrolled) state $\|\hat{X}\|^2$ (in gray).

The above system determines an impulse response triggered by the noise term $\hat{B}_1$: the initial condition $\hat{X}_{t=0} = \hat{B}_1$ is advected downstream while being amplified along the shear-layer of the recirculation bubble. The energy $\|\hat{X}\|^2$ of this perturbation is displayed versus time by a gray thick line in figure 5. In addition to the time axis, a second axis is displayed at the top of the figure where the location of the advected wave packet, evaluated by the energy-weighted x-centroid x_c defined as $x_c(t) = \int x k'(x,y,t) dx\, dy / \int k'(x,y,t) dx\, dy$ with $k'(x,y,t) = (u'^2 + v'^2)/2$ as the pointwise kinetic-energy at time t, is shown. The initial condition is associated with the non-zero state energy $\|\hat{X}\|^2 = \|\hat{B}_1\|^2$ at $t = 0$. The energy then decreases for $0 < t \lessapprox 1.5$ as the perturbation traverses the stable region of the flow between the noise location and the separation point. Beyond the separation point of the shear layer, the wave packet enters the convectively unstable region and its energy grows until the perturbation reaches the attachment point. At time $t \approx 19$ the energy reaches a maximum; the corresponding energy-weighted x-centroid x_c is located at $x \approx 9.5$. The perturbation continues through a convectively stable region and the state energy decreases accordingly. During this advective process the estimator tries, in real-time, to reconstruct the actual state from the information provided by one sensor, and the estimation performance parameter is given by the integral in time of the actual estimation error $\hat{Z}$. In the following analysis, all estimators ($\hat{C}_1$ to $\hat{C}_4$) will operate at their respective large gain limit, which ensures the best attainable performance for each estimator.

In figure 5 the thin solid lines display the energy of the estimation error $\|\hat{Z}\|^2$ as a function of time for all four sensors. The red curve traces the estimation error associated with sensor $\hat{C}_1$. For short times ($0 < t \lessapprox 3$), the estimation error energy is comparable to the state energy, indicating a relative estimation error of 100%. Starting at $t \approx 3$, the error drops abruptly by one order of magnitude before a more gradual decrease sets in for $t \gtrapprox 6$. The estimator becomes effective as soon as the error curve clearly detaches from the state energy curve (gray thick solid line); the state is hence well estimated beyond $t \approx 4$ using sensor $\hat{C}_1$. The estimation error curves (green, blue and black curves) for the

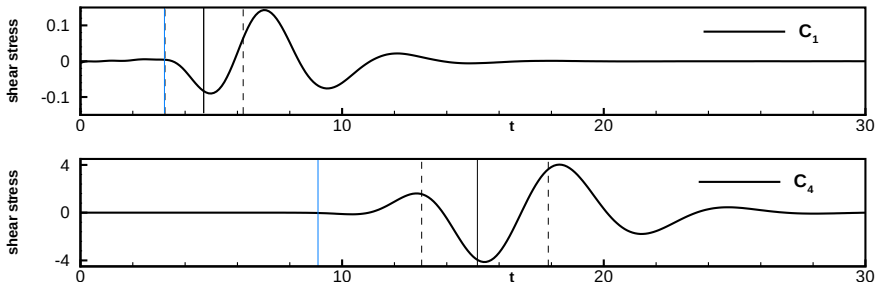


FIGURE 6. Measured impulse response at the four sensor locations, $\hat{\mathsf{C}}_{1,4}$ (top to bottom) The blue vertical lines correspond to the times when $\|\hat{\mathsf{Z}}\|^2/\|\hat{\mathsf{X}}\|^2 = 0.9$ (see figure 5). The black vertical lines indicate the times when the energy-based x-centroid of the wave packet x_c reaches the sensor; the dash lines give the times when $x_c \pm \sigma$ reaches the sensor, with σ denoting the standard deviation of the wave packet.

remaining sensors $\hat{\mathsf{C}}_{2,3,4}$ display a similar behavior: a relative estimation error of 100% for early times, followed by a pronounced drop after a critical time and finally a gradual decay. The abrupt decline in the estimation error energy, however, occurs considerably later than for sensor $\hat{\mathsf{C}}_1$, and this delay increases steadily as the sensor location is moved further downstream. Nevertheless, in all cases the estimation error ultimately decreases, and the flow state appears to be well-estimated for large times. We thus conclude that the principal difference between the estimators is the time at which they start to become effective: sensor $\hat{\mathsf{C}}_1$, located furthest upstream, yields the earliest accurate estimates of the state, followed by $\hat{\mathsf{C}}_2$, $\hat{\mathsf{C}}_3$ and finally $\hat{\mathsf{C}}_4$.

More insight is gained by displaying the measurements from the different sensors for the above impulse-response simulation (see figure 6). For sensor $\hat{\mathsf{C}}_1$ (top figure) we observe a quiet phase ($0 < t \lessapprox 4$), after which a sinusoidal signal, the footprint of the wave packet traveling downstream in the shear layer above the sensor, is detected. The measurement returns to zero for $t \gtrapprox 16$. Similar features can be observed for the other three sensors and in particular for sensor $\hat{\mathsf{C}}_4$ (lower figure); however, the time of first detection is delayed and the amplitude of the signal is substantially increased (by nearly forty times between $\hat{\mathsf{C}}_1$ and $\hat{\mathsf{C}}_4$) as we move the sensor location further downstream. The time delay in detecting the wave packet is closely linked to the overall performance of the four estimators: early detection yields better results. The detection times in figure 6 (blue vertical bars) correspond to the critical times in figure 5. After the wave packet has been captured by the sensor, the estimation proceeds rapidly due to the large Kalman gains (large gain limit). If noise is generated at $\hat{\mathsf{B}}_1$, the $\hat{\mathsf{C}}_4$-estimator for example is able to identify the associated response in the region $x > 5$ but is incapable of detecting any response in the region $-1 < x < 5$. The performance of the estimator is thus less determined by the quality of the reconstructed state — all energy curves in figure 5 tend to zero — than by its reaction time which translates into a spatial range where state responses to noise are detectable. It is this distinction that will have a marked impact on the performance of closed-loop control of amplifier systems (see § 4) and will reveal the effectiveness of seemingly performing estimators when incorporated into a compensator.

Larger amplitudes are detected at sensors located further downstream which stems from the amplification of the wave packet due to a convective instability along the shear layer of the separation bubble; the four sensors $\hat{\mathsf{C}}_{1,2,3,4}$ capture the wave packet at various stages of this instability. This difference in amplitude also explains some features observed

in figure 4(b). Comparing the curves associated with the different sensors $\hat{\mathsf{C}}_{1,2,3,4}$, we notice that the estimation process becomes effective for different values of the noise-to-signal ratio G/W. For example, sensor $\hat{\mathsf{C}}_1$ starts to perform well for $G/W < 1.25$, while sensor $\hat{\mathsf{C}}_4$ only requires $G/W < 52$, which leads us to conclude that higher-quality sensors are required when one plans to place them further upstream where signals are generally weaker.

In summary, two competing mechanisms have been isolated in the estimation process for amplifier flows: (i) for an effective estimator, the sensor has to be located sufficiently upstream to allow a *rapid* identification of the perturbation; (ii) on the other hand, the noise-to-signal ratio G/W has to be sufficiently small to enable an *accurate* estimate, thus favoring or forcing the placement of noisy sensors further downstream where the signal amplitudes are higher. In short, a balance between speed and accuracy of the estimation process has to be struck. Whereas the upstream placement of the sensors runs somewhat counter to the intuitive placement of the sensor near the reattachment-line, it will be shown that, for our prototypical configuration, the speed of estimation appears more critical for a successful compensator performance than the capture of highly accurate measurements.

4. Closed-loop control based on reduced-order model

After our analysis of the estimator and its performance, we now direct our attention to the complete compensator. After a brief presentation of the controller and its design steps, we investigate the performance of the compensator built on the four sensors $\hat{\mathsf{C}}_{1,2,3,4}$. Two objective functionals for the controller will be studied: (i) the square of the measurement based on $\hat{\mathsf{C}}_p$ and integrated over time, and (ii) the time-integral of the entire perturbation energy. Within this section, the plant is modeled by the reduced-order model introduced in §2.

4.1. *Presentation of the controller*

We will aim at suppressing perturbations in our fluid system by employing an optimal control strategy which will be designed to minimize a predefined cost functional. In mathematical terms, a control law of the form $u(t) = \hat{\mathsf{K}}\hat{\mathsf{X}}$ will be assumed where the control gain $\hat{\mathsf{K}}$ arises from the solution of a Riccati equation (see Burl 1999). Traditionally, the cost functional is related to a quantity measuring the energy of the state, but also takes into account the control effort in terms of its expended energy.

In our study, two measures of the state will be considered in the minimization process: the energy contained in the measurement extracted at location $\hat{\mathsf{C}}_p$ (see figure 1) yielding a cost functional of the form

$$J_m = \int_0^\infty \left(\hat{\mathsf{X}}^* \hat{\mathsf{C}}_p^* \hat{\mathsf{C}}_p \hat{\mathsf{X}} + l^2 u^2 \right) dt, \tag{4.1}$$

or the perturbation energy contained in the entire domain leading to a cost functional of the form

$$J_e = \int_0^\infty \left(\hat{\mathsf{X}}^* \hat{\mathsf{X}} + l^2 u^2 \right) dt. \tag{4.2}$$

The parameter l appears in either choice and is referred to as the cost of control, as it quantifies the user-specified weighting of the control energy compared to the quantity to be minimized. Similar to the estimator, a small-gain-limit (resp. large-gain-limit) parameter regime exists where the controller exerts nearly no action (resp. maximum action)

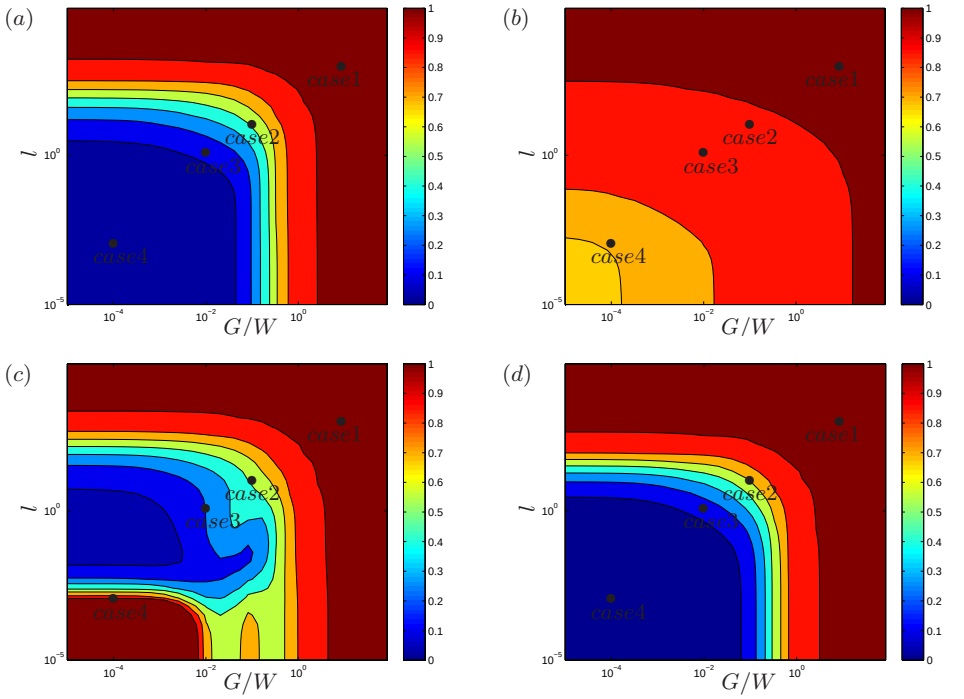


FIGURE 7. Contours of the performance measures P_m and P_e as a function of noise-to-signal ratio G/W and control cost parameter l. (a):$P_m|_{cf:m}$ using estimation sensor $\hat{\mathsf{C}}_1$. (b):$P_m|_{cf:m}$ using $\hat{\mathsf{C}}_4$. (c):$P_e|_{cf:m}$ using $\hat{\mathsf{C}}_1$. (d):$P_e|_{cf:e}$ using $\hat{\mathsf{C}}_1$.

on the flow. Invoking the separation principle (see Burl 1999), the controller design can be carried out independent of the estimator design.

The performance assessment of the compensator will follow the frequency-based framework for amplifier flows outlined in §2 and §3. Considering the state-space system (2.4) driven by a harmonic excitation $w(t) = \exp(i\omega t)$ the response of the compensated system reads

$$\begin{pmatrix} \hat{\mathsf{X}}_\omega \\ \hat{\mathsf{Y}}_\omega \end{pmatrix} = \begin{pmatrix} i\omega\mathsf{I} - \hat{\mathsf{A}} & -\hat{\mathsf{B}}_2\hat{\mathsf{K}} \\ \hat{\mathsf{L}}\hat{\mathsf{C}} & i\omega\mathsf{I} - \left(\hat{\mathsf{A}} + \hat{\mathsf{B}}_2\hat{\mathsf{K}} + \hat{\mathsf{L}}\hat{\mathsf{C}}\right) \end{pmatrix}^{-1} \begin{pmatrix} \hat{\mathsf{B}}_1 \\ \hat{0} \end{pmatrix}. \tag{4.3}$$

The above equation will form the basis for our performance analysis of the compensated system, where we will focus on the influence of controllers designed with J_m or J_e as well as on the impact of the control cost l and the noise-to-signal ratio G/W.

4.2. *Performance of the compensator using a cost functional based on the measurement*

We start by considering a compensator whose controller has been designed using the cost functional J_m, i.e., the state is measured by the energy output of the sensor $\hat{\mathsf{C}}_p$ and the resulting controller aims at minimizing the energy of the measurement $m_p = \hat{\mathsf{C}}_\mathsf{p}\hat{\mathsf{X}}$.

4.2.1. *Effect on the perturbation measure m_p*

Figures 7(a) and (b) show the performance of the compensator designed to minimize the measurement where each subplot displays results using respectively sensor $\hat{\mathsf{C}}_1$ and

$\hat{\mathsf{C}}_4$ for the estimator. Results related to sensors $\hat{\mathsf{C}}_2$ and $\hat{\mathsf{C}}_3$ display intermediate results and are not shown here. The measurement-based performance P_m of the compensator is defined as

$$P_m = \left[\frac{\int_{-\infty}^{+\infty} |\hat{\mathsf{C}}_p \hat{\mathsf{X}}_\omega|^2 d\omega}{\bar{m}_0^2} \right]^{1/2} \tag{4.4}$$

where $\hat{\mathsf{X}}_\omega$ is given by (4.3) and $\bar{m}_0^2$ is the measurement energy related to the uncontrolled case. Each plot shows iso-contours of P_m in the $(G/W, l)$-plane. Contours with hot colors (red) indicate parameter combinations where the control has been ineffective in reducing the measurement energy; contours with cold colors (blue) point to values of $(G/W, l)$ where the perturbation measure has been reduced successfully. The convergence of the performance P_m towards one (ineffective closed-loop control) is common to all sensor configurations as either the control cost l or the noise-to-signal ratio G/W exceeds a critical value. This parameter regime corresponds to small-gain-limit situations where either the control gain $\hat{\mathsf{K}}$ or the Kalman gain $\hat{\mathsf{L}}$ approach zero. For small noise-to-signal ratios G/W and small control parameters l (inexpensive control), the estimator provides an accurate approximation of the state which is subsequently multiplied by a non-zero control gain to obtain a positive action on the perturbation. As a result, the performance measure P_m is rather small in this parameter regime since both the estimator and the controller reach their large gain limit and behave at their best. The compensator based on sensor $\hat{\mathsf{C}}_1$ is, by a considerable margin, the most efficient with a performance parameter P_m equal to 0.026 in the large gain limit; this means that only 2.6% of the uncontrolled measurement energy remains after control is applied. As the sensor location for the estimator is moved further downstream though, the performance parameter P_m rises substantially in the large-gain region: 17%, 36% or 71% of the uncontrolled measurement energy could not be removed by the compensator using sensors $\hat{\mathsf{C}}_2$, $\hat{\mathsf{C}}_3$ or $\hat{\mathsf{C}}_4$, respectively. This exercise clearly demonstrates that an actuator placed near the edge of the step requires a sensor located in its vicinity, if satisfactory performance of the compensator is to be expected; alternatively, a sensor further downstream (e.g., $\hat{\mathsf{C}}_4$) requires an actuator in its upstream neighborhood, but this layout will produce larger control gains and ultimately less performance compared to the upstream configuration (e.g. $\hat{\mathsf{C}}_1$). In what follows, we will concentrate on sensor $\hat{\mathsf{C}}_1$ and further probe its performance behavior and limitations.

More physical insight into the compensated system can be gained by computing the transfer function between the noise w and the performance measurement m_p. Four cases, labeled accordingly in figure 7(a), are analyzed in detail: case 1 is representative of an ineffective compensator in the small gain limit, both cases 2 and 3 characterize a system with average performance while case 4 corresponds to a compensator operating in the large gain limit. The governing parameters, i.e., the noise-to-signal ratio and control cost, for theses cases are summarized in table 1 (second and third column) together with values of various performance measures.

In figure 8(a) the magnitude of the transfer function from $w(t)$ to $m_p(t)$ for each of the four cases is displayed, and results pertaining to the uncontrolled system are overlaid in black symbols. As expected, the compensator operating in the small gain limit (case 1, shown in red) does not act on the flow and the transfer function is identical to the uncontrolled one. By progressively reducing the noise-to-signal ratio and the control parameter (light blue and dark blue curve), the most amplified frequencies are considerably reduced and the compensator becomes effective over a wider range of frequencies, even though frequencies above $\omega \approx 1.5$ are slightly more amplified compared to the uncontrolled case.

Case	l	G/W	$P_m\|_{cf:m}$	$P_e\|_{cf:m}$	$P_m\|_{cf:e}$	$P_e\|_{cf:e}$
1	10^3	10^1	1.000	1.000	1.000	1.000
2	10^1	10^{-1}	0.566	0.574	0.784	0.707
3	10^0	10^{-2}	0.135	0.254	0.292	0.157
4	10^{-3}	10^{-4}	0.026	1.084	0.037	0.005

TABLE 1. Performance measures based on measurement energy (subscript $_m$) or global perturbation energy (subscript $_e$). The compensator has been designed using on a cost functional based on measurement energy (subscript $_{cf:m}$) or on global perturbation energy (subscript $_{cf:e}$). Four selected cases, ranging from the small gain limit to the large gain limit, are presented.

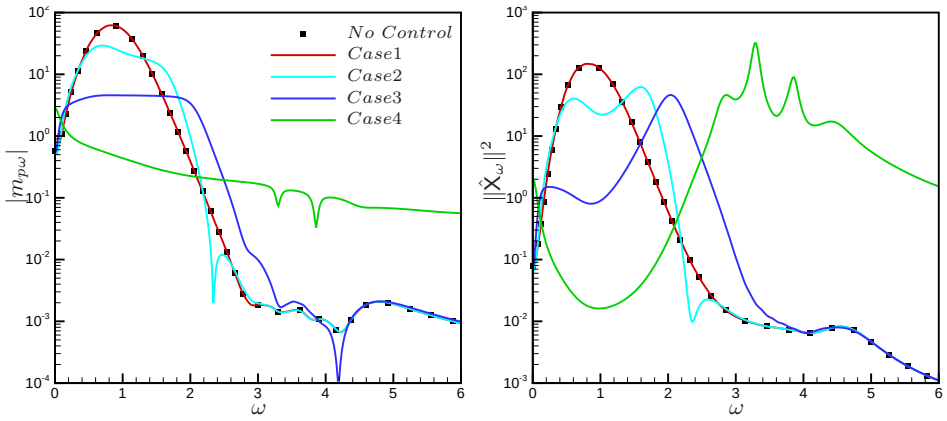


FIGURE 8. (a) Magnitude of the transfer function $|m_{pw}|$ for four different $(G/W, l)$-parameter settings, as well as the uncontrolled case. (b) Transfer function between noise and global energy for the same four parameter settings.

This tendency continues until the large gain limit (in green) is reached: the low frequencies ($\omega < 2$) which would be naturally amplified by the uncontrolled system have been successfully suppressed, which explains the very good performance of the compensator with $P_m = 0.026$ (see table 1, fourth column).

The results above confirm the successful manipulation of the inherent amplification behavior (see figure 2(a)) of the uncontrolled flow: the pronounced response to low frequencies has been strongly reduced by the compensator.

4.2.2. *Response in the frequency domain: effect on the perturbation energy*

Even though the controller is designed based on the measurement m_p only, the performance of the compensator can also be evaluated by the reduction of the perturbation energy in the entire domain. This is possible in the present case since our reduced-order model is based on POD modes, and therefore accurately captures the energy of the original system. This point has particular implications for experimental control setups where reduced-order models are typically obtained by identification techniques based on input and output data. By construction, such models cannot express or capture state information, and the question arises whether targeting the measurement energy in the cost functional produces commensurate reductions in the entire perturbation energy. While the control gain is still based on the measurement-based cost functional J_m, we therefore

evaluate the global energy performance measure

$$P_e = \frac{\int_{-\infty}^{+\infty} \|\hat{\mathsf{X}}_\omega\|^2 d\omega}{\bar{E}_0} \tag{4.5}$$

with $\hat{\mathsf{X}}_\omega$ from (4.3) and $\bar{E}_0$ as the energy based on the uncontrolled system. The ratio P_e of the perturbation energy for the compensated case to the perturbation energy for the uncontrolled case is depicted in the $(G/W, l)$-plane in figure 7(c). This plot shows similar characteristics than figure 7(a) but also displays important differences. For large values of G/W and l (the small gain limit for estimator and controller) the control action is negligible and no reduction in the perturbation energy is achieved ($P_e = 1$). For moderate estimation and control-cost parameters, the performance parameter P_e decreases, and it appears that a reduction in measurement energy (measured by P_m) brings about a proportional reduction in the overall perturbation energy. However, as the large gain limit is approached (case 4), the value of P_e increases again, even above one, indicating that the perturbation energy of the controlled case exceeds the energy of the uncontrolled case. The measurement energy, however, is efficiently reduced, as by design.

As before, the transfer function between the noise and the energy in the domain $\|\hat{\mathsf{X}}_\omega\|^2$ provides more details of the observed behavior (see figure 8(b)). Case 1 (in red) represents the small gain limit, and the transfer function coincides with the one for the uncontrolled flow since no control action is exerted on the flow. As G/W and l are reduced (light blue curve), the dominant, inherently amplified frequencies around $\omega = 1$ are reduced by the compensator but higher frequencies appear near $\omega = 1.8$. Nonetheless, the energy in the entire domain diminishes ($P_e = 0.458$), see values in table 1 (fifth column). This trend continues as the governing parameters are further decreased (dark blue curve). Finally, in the large gain limit (green curve), the dominant frequencies of the uncontrolled system ($\omega \approx 1$) have been reduced by four decades, while the energy in higher frequencies ($\omega > 2$) has been amplified by four orders of magnitude. This amplification outweighs the control effort on the lower frequencies, thereby leading to an increase in the overall energy of the system ($P_e = 1.085$) and a failure of the compensator when measured in the global energy norm. The reason for this behavior may stem from the location of the spatial supports of the optimal responses at high frequencies. As mentioned in §2.2, this support is upstream of sensor m_p for $\omega \geqslant 2$. Hence, m_p is not able to detect the naturally amplified perturbations (the optimal responses) in this frequency range. This may explain the failure of the compensator to stabilize the energy in high frequencies, since the objective functional of the compensator is precisely the measurement energy based on m_p.

4.3. *Performance of the compensator using a cost functional based on the energy*

We conclude by basing the controller design on the cost functional J_e and by directly targeting the perturbation energy in the entire domain. In figure 7(d), we show in the $(G/W, l)$ plane, the performance parameter P_e. We observe that in the large-gain limit (small values of G/W and l), the compensator now efficiently reduces the perturbation energy $P_e = 0.005$. From table 1 (column 6), we also note that the perturbation measurement is significantly reduced, i.e., $P_m = 0.037$.

5. Closed-loop control using linearized Navier-Stokes simulations

In the last section, we have studied the performance of the designed compensator with the reduced-order-model ROM_{150} as the plant. We will now assess the performance of

this compensator in the case of the large-scale linearized direct numerical simulation (LDNS), i.e., the full, unreduced model for flow over a rounded backward-facing step, as the plant. In many flow control studies, this represents the essential performance test of the compensator. As the reduced order model reproduces all input-output dynamics from the LDNS with an error less than 1%, we expect that the results of the model study presented in §3 and §4 remain valid in the case of the compensated LDNS. We will see that this is indeed true in the case of small and medium gains (§5.1). As a result, the compensated LDNS exactly reproduces the behavior of the compensated ROM. In the medium gain case 3, we will in particular assess the cost related to the present control strategy with an energy-based analysis (§5.2). In the case of larger gains (§5.3), we will show that the residual 1% error between the transfer functions of the LDNS and the ROM may render the compensated LDNS unstable. A sensitivity study, based on minimizing the stability margin of the compensated system, will be conducted to quantify and understand this phenomenon.

The instability problem discussed above is in fact a generic problem which may be encountered as soon as model reduction techniques are combined with optimal control tools. Compensators based on accurate reduced-order models are expected to perform well when employed directly on the high-dimensional plant. If discrepancies between the full and reduced-order transfer function prevail, however, a satisfactory performance of the compensator is no longer guaranteed, nor can any bounds on the decline in performance be given. In any model reduction effort, approximation errors are present and can potentially degrade or ruin the compensator's efficacy. One of the objectives of this section is an assessment of the compensator's sensitivity to deviations of the reduced-order input-output behavior from the one of the full system.

5.1. *Analysis of case 3 ($l = 10^0$ and $G/W = 10^{-2}$)*

We first consider the medium gain case 3 ($l = 10^0$ and $G/W = 10^{-2}$) and use the cost-functional J_e as our objective for the controller. We choose again the best-performing estimator based on sensor C_1. The plant is given by the discretized system of equations (2.3); a second-order scheme is used for the time integration. A white-Gaussian noise is fed into $w(t)$, and a statistically steady state is obtained after some time. The performance measures P_m and P_e defined in (4.4) and (4.5) may be evaluated in the simulations by time-averaging $|\mathsf{C}_p\mathsf{X}|^2$ and $\mathsf{X}^*\mathsf{Q}\mathsf{X}$, respectively. The results should be comparable to those of the ROM-study (§4) since the simulation has been given a white-Gaussian noise in time, which statistically feeds all frequencies by an equal amount.

Figure 9(a) juxtaposes the temporal evolution of energy for the uncontrolled simulation (in red) and the evolution of energy for the compensated simulation (in black), where the same excitation sequence $w(t)$ has been used to ensure a fair comparison. Starting from a zero initial state, a transient phase is observed that quickly evolves into a statistically stationary state. The time for a particle to be convected from the actuator location B_2 to the reattachment point (sensor C_p) is approximately 20 time units. The total simulation time $T = 1800$ is almost 100 times larger, thus ensuring adequate convergence of the statistical properties of the flow. At $t = 200$ the compensator is switched on (black curve). The energy rapidly decreases to levels nearly one decade smaller than in the uncontrolled simulation. For completeness, the measurement signal $m_p(t)$ is displayed in figure 9(b) for the uncontrolled (red) and compensated (black) simulation. Again, a distinct reduction in variance of the signal can be observed.

The third column of table 2 contains the performance measures P_e and P_m corresponding to the above simulation. The values are nearly identical to the ones obtained for the reduced-order model (the numbers from table 1 have been reproduced in the sec-

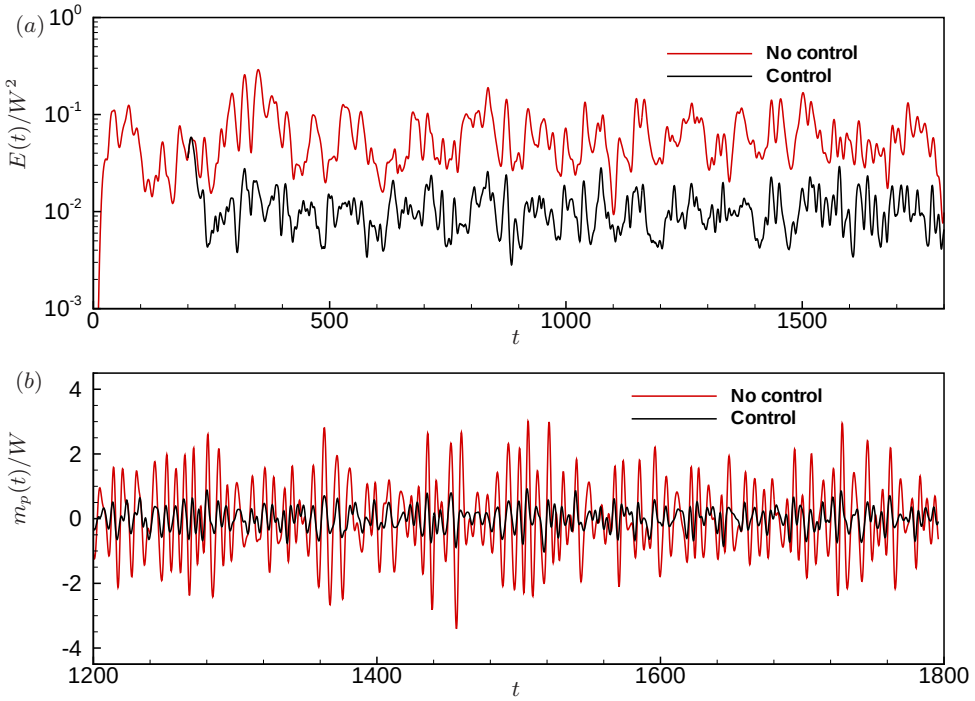


FIGURE 9. Temporal evolution of the energy based on a LDNS continuously forced by random noise. The compensator has been designed under linear assumptions (case 3). (a) Perturbation energy for the uncontrolled (red) and controlled (black) case, (b) measurement energy for the uncontrolled (red) and controlled (black) case.

	ROM	LDNS ($W=1$)	DNS ($W=0.1$)	DNS ($W=1$)	DNS ($W=\sqrt{10}$)
P_e	0.157	0.17	0.17	0.41	0.63
P_m	0.292	0.30	0.31	0.64	0.84
$\overline{k'}_{max}/U_\infty^2$ (%)		0.42	0.0042	0.31	1.96
$\overline{k'}_{max}\big\vert_c/U_\infty^2$ (%)		0.07	0.0007	0.12	1.34
$\overline{k'}_{max}\big\vert_c/\overline{k'}_{max}$		0.16	0.16	0.37	0.69

TABLE 2. Performance evaluation of linear and nonlinear simulations. Column 2: based on the reduced-order model (§4), column 3: based on LDNS with random noise (§5), columns 4, 5, 6: based on DNS with random noise with $W=0.1$, $W=1$ and $W=\sqrt{10}$ (§6).

ond column). It is evident that the overall perturbation energy P_e and the perturbation measurement energy P_m have been reduced by a factor of 0.17 and 0.30, respectively. These results underline the validity of the overall procedure followed within this article.

Finally, figure 10 displays contours of the pointwise perturbation mean-kinetic-energy (normalized by the variance of the input noise) $\overline{k'}/W^2$ for the uncontrolled (a) and controlled (b) simulations. The pointwise kinetic energy of the perturbation is taken as

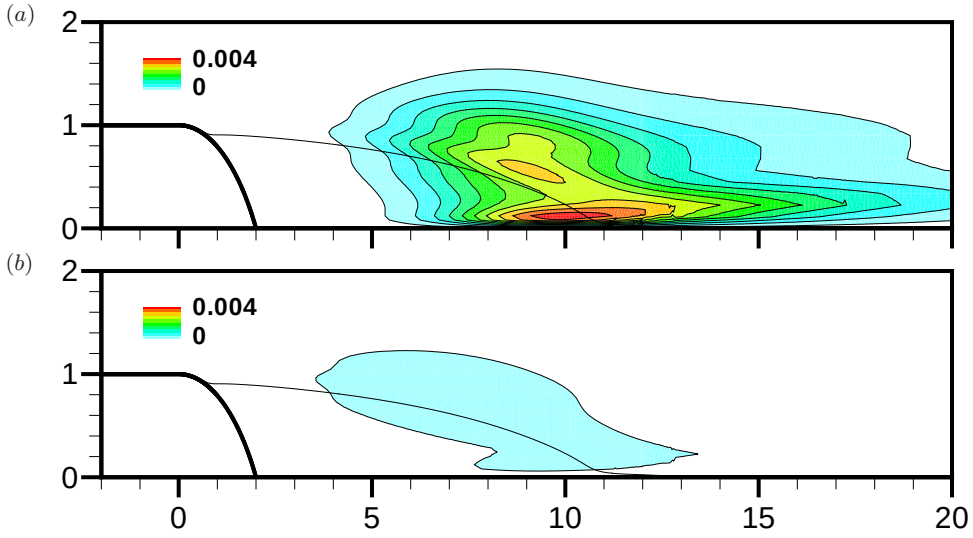


FIGURE 10. Pointwise mean perturbation kinetic energy $\overline{k'}/W^2$ of simulations with random forcing. (a) Uncontrolled LDNS; (b) compensated LDNS (case 3).

$k'(x,y,t) = (u'^2 + v'^2)/2$, and $\overline{(\cdot)}$ designates the time average. The perturbation energy increases along the shear-layer due to a convective instability; its maximum is reached near $x = 10$. This pointwise mean-kinetic-energy distribution for the uncontrolled case is closely related to the optimal response given in figure 2(c). The mean-kinetic-energy contours for the controlled simulation deviate significantly, as the effectiveness of the compensator is clearly demonstrated by the greatly reduced energy levels in the shear-layer region. Table 2 (column 3) shows that the maximum (over space) perturbation mean kinetic energy is lowered by a factor of $1/0.16 \approx 6$. Figure 10 is of interest to experimental studies of flow control as it provides a direct comparison of local turbulence levels throughout the recirculation bubble for the uncontrolled and controlled case.

5.2. *Analysis of control cost*

We will next assess the characteristics of the control cost. We first note that the penalization term u^2, appearing in the objective functionals J_m and J_e introduced in §4, is not representative of the true energy spent by the user for control. We therefore resort to an analysis of the energy budget: evaluate the user-supplied power to the system associated with $u(t)$ and compare it to a physical quantity that describes the realized gain by controlling the flow. To this end, we recall the linearized Navier-Stokes equations (2.3) in continuous form

$$\partial_t \boldsymbol{v}' + \boldsymbol{V}\cdot\boldsymbol{\nabla}\boldsymbol{v}' + \boldsymbol{v}'\cdot\boldsymbol{\nabla}\boldsymbol{V} = -\boldsymbol{\nabla}p' + Re^{-1}\boldsymbol{\nabla}^2\boldsymbol{v}' + w(t)\boldsymbol{b}_1 + u(t)\boldsymbol{b}_2, \qquad \boldsymbol{\nabla}\cdot\boldsymbol{v}' = 0 \quad (5.1)$$

where $\boldsymbol{v}'$ is the perturbation velocity and $\boldsymbol{V}$ the base flow. Taking the inner product of this equation with $\boldsymbol{v}'$, averaging over time $\overline{(\cdot)}$ and integrating in space over a domain Ω (which includes the spatial support of b_2, but excludes the one of b_1) yields the following

equation:

$$\int_{\partial\Omega}\left(\overline{k'}\mathbf{V}\cdot\mathbf{n}+\overline{p'\boldsymbol{v}'}\cdot\mathbf{n}-2Re^{-1}\overline{\mathcal{D}'\cdot\boldsymbol{v}'}\cdot\mathbf{n}\right)ds=\iint_{\Omega}\left(\overline{u(t)\mathbf{b}_2\cdot\boldsymbol{v}'}-\overline{\mathcal{R}'}:\mathcal{D}-2Re^{-1}\overline{\mathcal{D}':\mathcal{D}'}\right)d\Omega. \tag{5.2}$$

where : denotes the contraction operator or Frobenius inner product, defined as $\mathsf{A}:\mathsf{B}=\sum_{ij}A_{ij}B_{ij}=\mathrm{trace}(\mathsf{A}^*\mathsf{B})$. This equation governs the flux of the mean kinetic perturbation energy $\overline{k'}$ across the boundary $\partial\Omega$, given by $\int_{\partial\Omega}\overline{k'}\mathbf{V}\cdot\mathbf{n}\,ds$. It is written in conservative form, and various volume source terms appear on the right-hand-side of the equation which contribute to this flux. The mean external power $\overline{u(t)\mathbf{b}_2\cdot\boldsymbol{v}'}$ represents the mean power supplied by the user during the control effort. The production term $-\overline{\mathcal{R}'}:\mathcal{D}$ involves the Reynolds stress tensor $\mathcal{R}'$ and the strain rate tensor $\mathcal{D}$of the base flow. In the case of an amplifier flow, this term will supply the main power (drawn from the base flow $\mathbf{V}$) which in turn will trigger high values of the mean fluctuating kinetic energy flux. Even though this term may be positive or negative, in the case of an amplifier flow it should be predominantly amd strongly positive. The dissipation term $2Re^{-1}\overline{\mathcal{D}':\mathcal{D}'}$ contains the strain rate tensor $\mathcal{D}'$ of the perturbations and accounts for the mean power lost due to viscous stresses. Lastly, the power supplied to the system does not only enter the mean fluctuating kinetic energy flux, there are two other terms: the mean velocity-pressure correlation $\int_{\partial\Omega}\overline{p'\boldsymbol{v}'}\cdot\mathbf{n}\,ds$ and the viscous diffusion term $\int_{\partial\Omega}-2Re^{-1}\overline{\mathcal{D}'\cdot\boldsymbol{v}'}\cdot\mathbf{n}\,ds$.

We have evaluated the various integrals appearing in (5.2) for the case described in §5.1. For Ω we chose the domain $-0.1\leqslant x\leqslant 11.3$. The left boundary at $x=-0.1$ is downstream of $\mathbf{b}_1$, but upstream of $\mathbf{b}_2$, such that the domain Ω comprises the actuator $\mathbf{b}_2$ and extends downstream beyond the reattachment point. The line integrals along the wall and in the far-field yield zero contributions, and we are thus only left with contributions from the line integrals at $x=-0.1$ and $x=11.3$. In table 3, the numerical values of the non-zero line integrals at $x=-0.1$ and $x=11.3$ as well as the various source terms in $-0.1\leqslant x\leqslant 11.3$ are listed for the uncontrolled and controlled simulations. For the uncontrolled case we identify the kinetic energy flux at the boundaries as the dominant term among the line integrals, with a pronounced increase between the inlet ($x=-0.1$) and the outlet ($x=11.3$). This feature is linked to the convective Kelvin-Helmholtz instability which gives rise to strong disturbance growth along the shear layer of the separation bubble. Closer inspection of the source terms shows that this growth in kinetic energy flux is mainly due to the production term which is more than twice the dissipation term in magnitude (while the power term is zero due to the absent control input). The production term is the "engine" of the instability, located along the shear layer of the separation bubble (see figure 11). Clearly, any effective control effort should target this term in order to reduce the amount of unsteadiness in the flow. For the controlled simulation, similar features can be observed; yet the increase in kinetic energy flux between the boundaries of the domain Ω is significantly smaller (by a factor of 10, as shown in column 5 of table 3). The source terms (columns 8,9,10) show that the production and dissipation terms have decreased accordingly, while the user-supplied power term from the control effort remains negligible in the process. For a minute cost of control (for our case, a value of 0.004) we have attained a substantial reduction of kinetic energy flux between the uncontrolled and controlled simulation (in our case, $1.50-0.16=1.34$) — indeed a highly efficient effort. Of course, this efficiency is due to the fact that we control an instability, i.e., the strong convective instability as the source of unsteadiness is taken advantage of to actively suppress it.

Control	At $x=-0.1$			At $x=11.3$			Sources in $-0.1 \leqslant x \leqslant 11.3$		
	$\overline{k'}\mathbf{V}$	$\overline{p'\boldsymbol{v}'}$	Visc.	$\overline{k'}\mathbf{V}$	$\overline{p'\boldsymbol{v}'}$	Visc.	Power	Prod.	Dissip.
OFF	-0.0058	-0.00082	-0.000025	1.5	0.69	0.0017	0	3.8	-1.7
ON	-0.0058	-0.00079	-0.000025	0.16	0.068	0.0023	0.004	0.56	-0.34

TABLE 3. Evaluation of the control cost in the linearized simulations. All values are to be multiplied by 10^{-3}. The compensator has been designed for $l = 10^0$, $G/W = 10^{-2}$ (case 3).

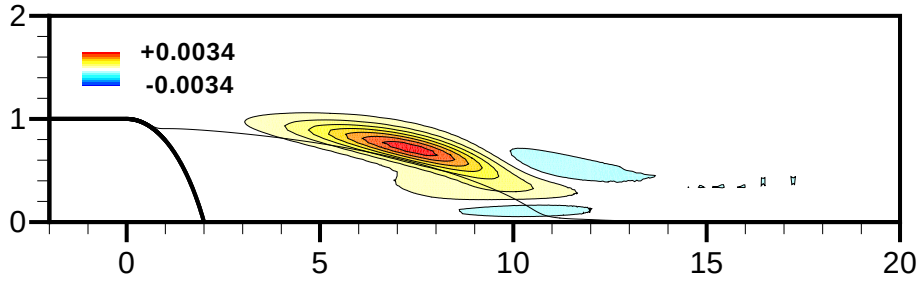


FIGURE 11. Iso-contours of the production term $(-\overline{\mathcal{R}'} : \mathcal{D})/W^2$ for the uncontrolled linear DNS.

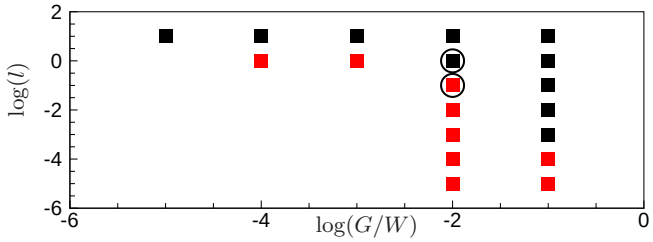


FIGURE 12. Instabilities in the compensated linearized numerical simulation. The red squares (resp. black squares) represent unstable (resp. stable) results. The circles refer to the "medium gain" and "large gain" cases studied in depth in §5.3.1 and 5.3.2.

5.3. *Behavior of compensated systems for larger gains*

This section is devoted to a performance evaluation of the compensator for the case of larger gains. Simulations with the randomly forced, linearized Navier-Stokes equations for a range of control costs l and noise-to-signal ratios G/W have shown that strong instabilities can arise in the compensated system. Figure 12 reports stable (black symbols) and unstable (red symbols) parameter combinations. For sufficiently large l or G/W (black symbols), the response behavior of the linearized simulation is close to the response of model $\texttt{ROM}_{150}$ (as observed in the medium-gains case in §5.1). The linearized DNS then reaches the same performance measures as the ones displayed in figure 7(d). Cases 1, 2 and 3, studied in §3 and §4, belong to this category as well, with equivalent performances of the linearized DNS. On the other hand, we also encountered cases in which the linearized direct numerical simulation exhibited instabilities as a compensator was applied (indicated by red symbols in figure 12). These cases concentrate at the large gain limit in the $(G/W, l)$-plane; case 4, studied in §3 and §4, falls into this category.

5.3.1. *Presentation of sensitivity analysis*

A discrepancy between the model of the plant on which a compensator is based and the model of the plant the compensator is intended to ultimately control can cause a decline in compensator performance or, in more severe cases, an instability in the compensated system. The dynamics of the compensated system is given by

$$\frac{d}{dt}\begin{pmatrix}\mathsf{X}\\ \hat{\mathsf{Y}}\end{pmatrix} = \begin{pmatrix}\mathsf{A} & \mathsf{B}_2\hat{\mathsf{K}}\\ -\hat{\mathsf{L}}\mathsf{C} & \hat{\mathsf{A}}+\hat{\mathsf{B}}_2\hat{\mathsf{K}}+\hat{\mathsf{L}}\hat{\mathsf{C}}\end{pmatrix}\begin{pmatrix}\mathsf{X}\\ \hat{\mathsf{Y}}\end{pmatrix} + \begin{pmatrix}\mathsf{B}_1\\ \hat{0}\end{pmatrix} w(t) \tag{5.3}$$

where the superscript $\hat{\cdot}$ refers to matrices of the $\texttt{ROM}_{150}$-system while matrices without the superscript are associated with the "true" plant model. We will attempt to understand the rise and origin of instabilities in such systems (shown in figure 12). In this effort, the matrix

$$\mathsf{M} = \begin{pmatrix}\mathsf{A} & \mathsf{B}_2\hat{\mathsf{K}}\\ -\hat{\mathsf{L}}\mathsf{C} & \hat{\mathsf{A}}+\hat{\mathsf{B}}_2+\hat{\mathsf{L}}\hat{\mathsf{C}}\end{pmatrix}. \tag{5.4}$$

is of critical importance, since an eigenvalue of M in the unstable half-plane is evidence of an instability in the compensated system. Considering components of the "true" plant model, the matrix M only depends on A, B_2 and C, which describe the input-output dynamics between $u(t)$ and $m_1(t)$; it does, however, not depend on B_1 and Q (note that the gains $\hat{\mathsf{K}}$ and $\hat{\mathsf{L}}$ implicitly depend on B_1 and Q through the Riccati equations, but this dependency has not been accounted for in the present analysis and the gains have been set to constants). This is to say that the input-output dynamics between $w(t)$ and $m_1(t)$, between $w(t)$ and $E(t)$ and between $u(t)$ and $E(t)$ have no influence on the stability of the compensated-system matrix M. Hence, we will focus on the dynamic link between $u(t)$ and $m_1(t)$ and introduce an error measure that describes the disparity between the transfer function of the original plant model and of the $\texttt{ROM}_{150}$ model. We have

$$err = \left[\frac{\int_{-\infty}^{\infty}|m_{1\omega}-\hat{m}_{1\omega}|^2 d\omega}{\int_{-\infty}^{\infty}|\hat{m}_{1\omega}|^2 d\omega}\right]^{1/2} \tag{5.5}$$

with $\hat{m}_{1\omega} = \hat{\mathsf{C}}(i\omega\mathsf{I}-\hat{\mathsf{A}})^{-1}\hat{\mathsf{B}}_2$ and $m_{1\omega} = \mathsf{C}(i\omega\mathsf{I}-\mathsf{A})^{-1}\mathsf{B}_2$.

We will then determine the minimal distance to instability of M by adjusting the components A, B_2 and C of the plant model (taken as matrices of size 150×150, 150×1 and 1×150, respectively) while maintaining a given error err between the respective transfer functions. We proceed in the following manner. We first note that a marginal eigenvalue $s = i\omega$ of the matrix M is equivalent to the largest singular value λ_ω of the resolvent matrix tending to infinity. The resolvent matrix corresponds to the matrix R^{-1} with $\mathsf{R} = i\omega\mathsf{I}-\mathsf{M}$. Formulating an optimization problem, we will try to maximize, for a given frequency ω, the functional $\lambda_\omega^2(\mathsf{A},\mathsf{B}_2,\mathsf{C})$ as a function of A, B_2 and C. The optimum will be sought under the constraint that the transfer function error $err(\mathsf{A},\mathsf{B}_2,\mathsf{C})$ maintains a prescribed error ϵ. The gradients with respect to A,B_2 and C of the above objective functional as well as the constraint condition are given in appendix B. The constrained optimization problem is transformed into an unconstrained problem via a penalization technique, and a Polak-Ribière conjugate-gradient method coupled to a line-search algorithm is used to determine the optimum.

5.3.2. *Results*

Two cases will be analysed in the following (see circles in figure 12): a stable case $l = 10^0, G/W = 10^{-2}$ (referred to as the *medium-gain* case in the following) and an unstable case $l = 10^{-1}, G/W = 10^{-2}$ (the *large-gain* case in the following). Figures 13(a,b) present

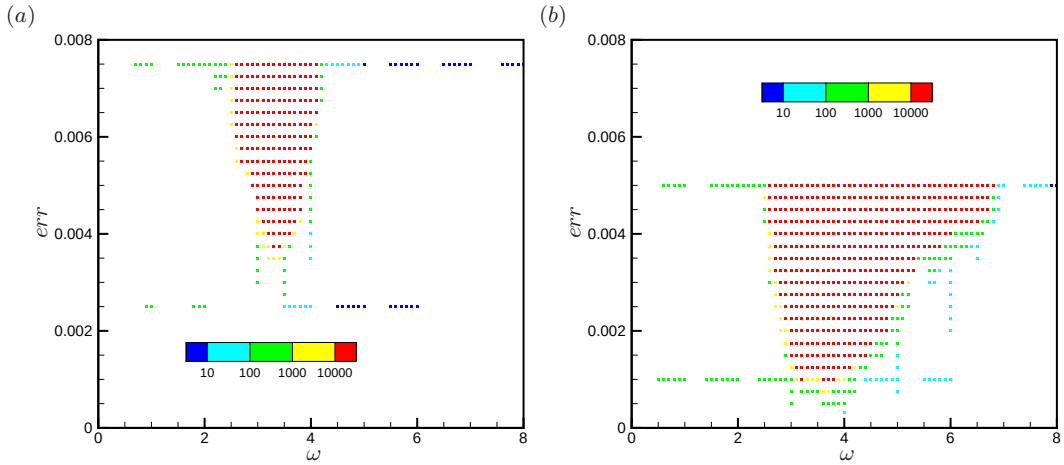


FIGURE 13. Results of the sensitivity analysis, visualized by the largest singular value of the resolvent matrix in the (ω, err)-plane. (a): Medium-gain case ($l = 10^0$, $G/W = 10^{-2}$, case 3). (b): Large-gain case ($l = 10^{-1}$, $G/W = 10^{-2}$). The symbols are colored according to the value of λ_ω.

symbols colored according to the value of λ_ω (based on the exponential color map) in a (ω, err)-plane. Hot colors (red) indicate that the optimization algorithm found marginally unstable compensated systems with an eigenvalue of M in the close vicinity of $s = i\omega$ and, equivalently, the largest singular values λ_ω of R^{-1} above 10^4. Cold colors (blue) signify that the resolvent matrix has a largest singular value λ_ω below 10^2 on the neutral line, indicating that M has no eigenvalue in the close neighborhood of $s = i\omega$. We observe that the transition from unstable to stable parameter combinations is rather sharp, which allows the definition of a clear unstable region. In the medium-gain case (figure 13(a)) the unstable area starts at a transfer function error $err \approx 0.00375$ with a frequency near $\omega = 3.3$. For increasing values of err, the unstable frequency band grows: for $err = 0.0075$, instabilities in the compensated system occur for frequencies within the range $2.6 \leqslant \omega \leqslant 4.1$. For the large-gain case (figure 13(b)), similar features are observed, except that the compensated systems become unstable for considerably smaller errors: as the error exceeds $err \approx 0.001$, frequencies within the range $3.2 \leqslant \omega \leqslant 3.8$ may render the compensated system unstable. For even larger errors err, the unstable frequency band extends to even higher frequencies: for $err = 0.005$, the entire frequency band $2.6 \leqslant \omega \leqslant 6.8$ displays instabilities.

These results support the observations summarized in figure 12. For larger gains, the error err between the plant and the reduced-order model has to be sufficiently small before stability of the compensated system can be ensured. In the present study, the reduced-order-model $\mathtt{ROM}_{150}$ apparently exhibits an error that is acceptable for the medium-gain case (case 3) but too high for the large-gain case. Moreover, the frequency range which is responsible for these instabilities covers the interval $2.6 \leqslant \omega \leqslant 6$. For the design of the reduced-order model, it is thus not sufficient to capture only the energetic, lower frequencies around $\omega \approx 0.8$; rather, it is essential to also account for higher frequencies, despite their low energy content.

Figure 14 shows a typical example from our sensitivity analysis where the compensated system becomes unstable for a particularly small error err of the transfer function. For the

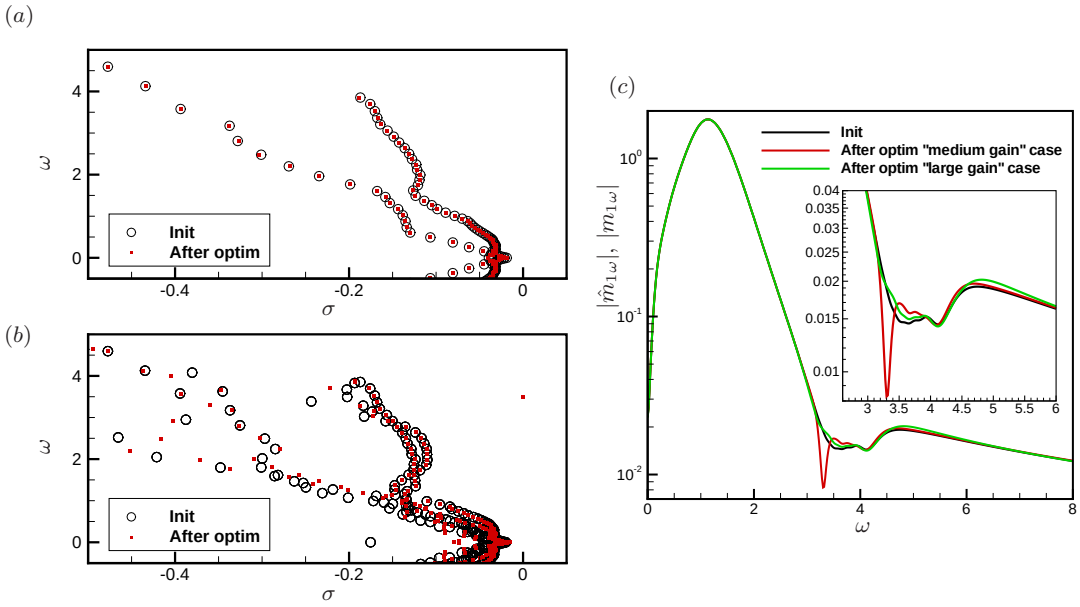


FIGURE 14. Example from the sensitivity analysis. (a,b): Spectra of plant A and compensated system M for the medium-gain case. The circles indicate the plant based on $(\hat{\mathsf{A}}, \hat{\mathsf{B}}_2, \hat{\mathsf{C}})$ while the red squares indicate the plant governed by the optimal $(\mathsf{A}, \mathsf{B}_2, \mathsf{C})$ found by the optimization algorithm. (c): Modulus of the transfer functions $\hat{m}_{1\omega}(\omega)$ and $m_{1\omega}(\omega)$ for medium-gain case ($l = 10^0$, $G/W = 10^{-2}$, $err = 0.00375$, $\omega_{opt} = 3.3$) and the large-gain case ($l = 10^{-1}$, $G/W = 10^{-2}$, $err = 0.00125$, $\omega_{opt} = 3.3$).

medium-gain case, the optimization process, performed for $\omega_{opt} = 3.3$ and $err = 0.00375$, yields a nearly unmodified spectrum for A (figure 14(a)) but strong modifications of the compensated spectrum for M (figure 14(b)). In both figures, the circles represent the spectrum of the plant based on $(\hat{\mathsf{A}}, \hat{\mathsf{B}}_2, \hat{\mathsf{C}})$ while the red squares pertain to a plant governed by the optimal $(\mathsf{A}, \mathsf{B}_2, \mathsf{C})$ determined by the optimization algorithm. The compensated system clearly displays a marginal eigenvalue at the optimization frequency $\omega_{opt} = 3.3$ (see figure 14(b)) which is in accordance with the high value of $\lambda_\omega = 7.5 \cdot 10^4$ found for the singular value of R^{-1}. Further insight can be gained from a comparison of the transfer functions $\hat{m}_{1\omega}(\omega)$ and $m_{1\omega}(\omega)$; the modulus of the respective transfer functions is depicted in figure 14(c). It is evident that the response behavior $u \to m_1$ for the optimized plant (red line) exhibitis a small error near the optimization frequency $\omega_{opt} = 3.3$. For comparison the transfer function modulus for the large-gain case, obtained for $\omega_{opt} = 3.3$ and $err = 0.00125$, is also displayed (in green). The large-gain transfer function is even closer to the original transfer function (black curve) but, nonetheless, displays a very high value for the largest singular value of R^{-1}, namely, $\lambda_\omega = 2 \cdot 10^4$. This latter finding corroborates, for the large-gain case, the high sensitivity of the compensated system to minute variations in the transfer function $u \to m_1$ in the frequency band $\omega \approx 2.6 - 6$. The sensitivity is less pronounced for the medium-gain case.

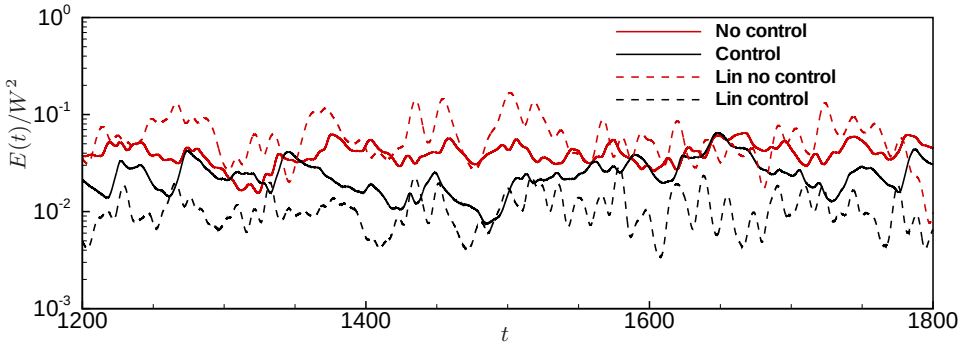


FIGURE 15. Temporal evolution of the mean perturbation energy based on a DNS continuously forced by finite-amplitude noise $W = \sqrt{10}$. The compensator has been designed under linear assumptions (case 3). The linear case has been added by dashed lines and scaled to the initial energy of the nonlinear simulations.

6. Nonlinear effects

The controller has been designed in the case of perturbations governed by linear dynamics. With non-linear governing equations, this situation is recovered if the amplitude of the perturbations is very small in the whole domain. This may be achieved with Navier-Stokes equations with an upstream very small noise amplitude $W \ll 1$. We have seen in the last section that small-amplitude noise may be suppressed by the linearly-designer controller. In the present section, we will gradually increase the amplitude of the incoming perturbations and the amplitude W; nonlinear effects will progressively be introduced and their impact on the overall performance of the linear compensator under off-design conditions can be studied. A similar test has been conducted by Ilak (2009) for impulsive initial conditions.

We have tested three different amplitudes: $W = 0.1$ (see movie in the online version of the paper), $W = 1$ and $W = \sqrt{10}$. The performance measures P_e and P_m corresponding to the three cases are reported in table 2 (columns 4, 5 and 6). The simulations have been run up to $t = 1800$, as in the linear simulation of §5. The random noise sequence from the linear case has been used, which allows a direct comparison with the linear results (reported in column 3). In table 2, the maximum pointwise mean-kinetic-energy of the uncontrolled and controlled (indicated by $|_c$) simulations are listed together with their ratio. For a small noise amplitude $W = 0.1$ the evolution of the uncontrolled perturbation is quasi-linear, and the performance of the compensator is nearly the same as in the linear case (compare columns 3 and 4 of table 2). The performance measures P_e and P_m however gradually deteriorate as the noise amplitude W increases (compare columns 4, 5 and 6 of table 2).

In figure 15, the temporal evolution of the mean-perturbation-energy is displayed for the large amplitude case $W = \sqrt{10}$; the uncontrolled simulation is shown in red, the compensated one in black. The oscillations of the energy curves for the nonlinear case (solid curve) are less pronounced than in the linear case (dashed line). The same observation can be made for the compensated simulation. It is seen that the mean energy level of the perturbation is reduced by the linear compensator, but a noticeable loss in performance is apparent.

As before, the pointwise perturbation mean-kinetic-energy $\overline{k'}/W^2$ gives further insight into the temporal evolution of the energy as the disturbances advect along the shear

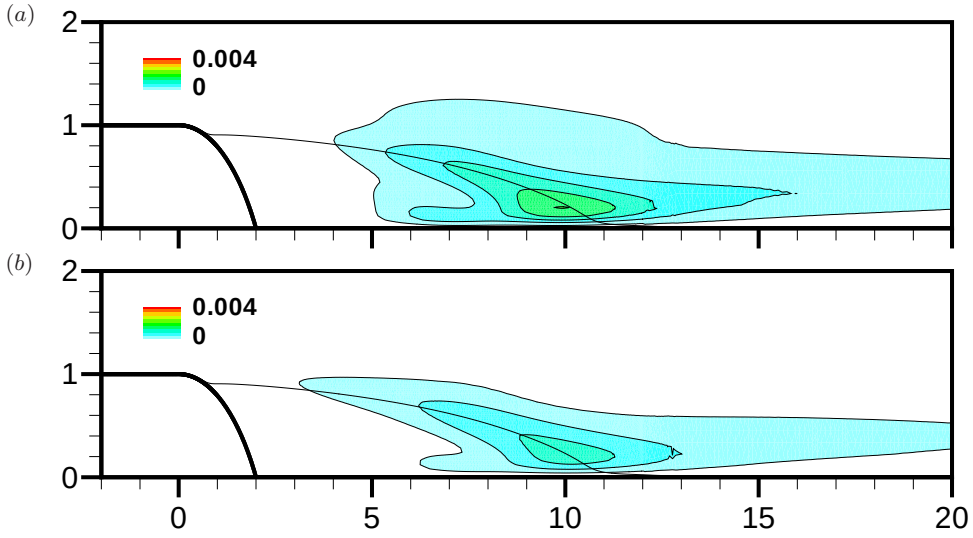


FIGURE 16. Pointwise perturbation mean-kinetic-energy $\overline{k'}/W^2$ of simulations with random forcing for $W = \sqrt{10}$. (a) Uncontrolled DNS. (b) Compensated DNS (case 3).

layer. In figure 16(a) and (b) the pointwise perturbation mean-kinetic-energy from the nonlinear simulation with random noise is displayed for the case with $W = \sqrt{10}$. The color map has been adjusted to permit direct comparison with the equivalent linear results given in figure 10; significant differences to the linear case can be observed. Figure 16(a), illustrating the uncontrolled case, displays saturation effects caused by nonlinearities; the overall shape of the maximum energy contours are, however, similar to the linear case — concentrating on the dynamics in the shear layer and the reattachment area. For the compensated case (figure 16(b)) the energy is maximal in the shear-layer but is, again, lower than in the uncontrolled case; in contrast, the energy is higher than in the linear case for $x > 10$, indicating the appearance of nonlinear structures which are more difficult to control. We notice that the energy could be less attenuated by control efforts than in the linear case: in table 2 (column 6), a maximum perturbation mean-kinetic-energy of 1.96% is observed in the uncontrolled simulation which reduces to 1.34% as the compensator is switched on. The perturbation mean-kinetic-energy is thus diminished by a modest factor of $1/0.69 \approx 1.5$ (compared to a factor of $1/0.17 \approx 6$ for the small amplitude simulation with $W = 0.1$, see column 4). As stronger nonlinearities set in, the linear-based compensator becomes ever less effective.

7. Summary and conclusions

Two-dimensional incompressible flow over a rounded backward-facing step, a canonical configuration showing noise-amplifying behavior, has been controlled by feedback control strategies. Specifically, the LQG-framework has been employed in conjunction with POD-based reduced-order models for the plant. Similar techniques have previously been studied (see Ilak & Rowley 2008; Bagheri & Henningson 2010), but with the main emphasis on model reduction aspects. Important issues related to the practical imple-

mentation of feedback control laws for amplifier flows have been left unaddressed, which motivated this present investigation.

The analysis of the feedback control setup first concentrated on the estimation process. A placement of sensors throughout the convectively unstable region of the flow revealed a distinct advantage of measurement input from the most upstream sensor. It further showed that the speed of estimation is more important than the accuracy of the estimation, while keeping in mind that upstream sensor measurements are more easily corrupted by noise since the signal has not yet been amplified and filtered by the flow. Low-quality sensors should be placed further downstream where the amplitude of the detected signal prevails over the added inherent measurement noise; the resulting delay in estimation, however, will ultimately cause a loss in compensator performance. The noise-to-signal ratio of the sensor thus plays an important role and has been linked to the estimation parameter for the computation of the Kalman gain.

Continuing with the best (most upstream) sensor, the performance of the compensated system has been studied under the idealistic assumptions that the reduced-order model accurately mimics the plant. The noise-to-signal ratio (or estimation parameter) G/W and the cost-of-control parameter l have been varied to cover a range of control scenarios from the small gain limit (SGL), where the compensator is ineffective, to the large gain limit (LGL), where the compensator operates at its maximal performance. An excess in the total energy can be observed when the measurement energy is the control objective; this phenomenon may be traced back to the blindness of sensor m_p to high frequency ($\omega \geqslant 2$) optimal responses. This shortcoming can be overcome by basing the cost functional for the controller design on the total energy (J_e).

A compensator designed with J_e has then been applied to the full, linearized Navier-Stokes equations driven by continuous stochastic forcing. In the case of medium gains, the compensated system reacted as predicted by the ROM-based model study, and a substantial reduction in perturbation energy could be accomplished. A detailed, term-by-term analysis of contributions to the energy budget showed that the control action was extremely efficient: a very weak user-supplied actuation power generated a substantial (order one) gain in the kinetic energy flux near the reattachment location. For compensators designed with larger gains, instabilities in the compensated linearized Navier-Stokes equations arose: even if the compensator is stable for the model it was designed for, its stability can no longer be implied or assured when applied to a slightly different plant. This loss of stability has been formalized and corroborated by a sensitivity analysis that quantitatively determined the stability margins and established that the POD model was particularly sensitive to high frequencies (around $\omega \approx 3-6$). Hence, for a compensator to work efficiently with large gains (in order to achieve maximal performance), the reduced order model (upon which the compensator was designed on) is required to be very precise also in low energetic, high frequencies.

Nonlinear effects have been reintroduced to the closed-loop control problem by applying the compensator, designed under linear assumptions, to the nonlinear Navier-Stokes equations and attempting to suppress continuous noise sources of progressively higher amplitude W. Minor deviations have been detected for small-amplitude noise; in the large-amplitude case, however, the compensator performance deteriorated due to the appearance of nonlinearly triggered structures.

It can be concluded from our study that designing closed-loop control strategies for amplifier flows is significantly more involved than the equivalent design for oscillator flows. While in this latter case instabilities are generally narrow-banded in frequency and thus more easily detectable/controllable, a noise-amplifier produces more broadband signals and magnifies physical and non-physical perturbations alike. For this latter reason, a

	C_1	C_2	C_3	C_4	C_p	E
B_1	0.99	0.60	0.62	0.66	0.74	0.77
B_2	0.61	0.34	0.33	0.32	0.30	0.28

TABLE 4. Relative $\mathcal{H}_2$-errors (in %) for the various input-output combinations. The error on the energy has been evaluated using $[\int_{-\infty}^{+\infty}[(\hat{\mathsf{X}}_\omega^*\hat{\mathsf{X}}_\omega)^{1/2} - (\mathsf{X}_\omega^*\mathsf{Q}\mathsf{X}_\omega)^{1/2}]^2 d\omega / \int_{-\infty}^{+\infty}\mathsf{X}_\omega^*\mathsf{Q}\mathsf{X}_\omega d\omega]^{1/2}$.

comprehensive study of the sensitivity of the compensator performance with respect to various noise sources is imperative for a successful closed-loop control design. Transfer functions, i.e., frequency-based input-output relations, are particularly helpful in pinpointing strong sensitivities, in placing sensors efficiently, and in avoiding undesirable parameter regimes. It is hoped that the present study has introduced and demonstrated effective tools that — despite the inherent challenges – aid in the design of effective closed-loop control strategies for amplifier flows.

Appendix A. Reduced-order model

In view of our control objectives, the reduced-order model should be able to accurately capture the dynamic response of the output variables (given by sensors $\mathsf{C}_{1,2,3,4,p}$ and the perturbation energy $E(t)$) to forcings by the two input variables (given by $\mathsf{B}_{1,2}$). We choose a model-reduction technique based on standard proper orthogonal decomposition (POD), which transforms flow-state snapshots, arising from harmonic forcings introduced via the terms B_1 and B_2, into a finite-dimensional orthogonal basis onto which the full flow equations are projected. An analysis of the reduced-order model confirmed a satisfactory balance between accuracy and size. Details of this type of analysis are given in Dergham *et al.* (2011) for the same backward-facing step configuration using a single input B_2.

In our case, a reduced-order model based on a Galerkin projection of the full system onto the first 150 POD modes has been selected owing to its good performance and comparatively small size. The transfer functions of the full system have been computed by Fourier-transforming signals from the impulse responses of the linearized DNS. The errors between the transfer functions of the reduced-order models and the linearized DNS have been measured using the $\mathcal{H}_2$-norm.

Table 4 presents these errors for different combinations of input triggering and output sensing. We note that the impulse responses triggered by B_2 are more accurately captured than those triggered by B_1. All errors are below 1%.

Appendix B. Sensitivity analysis

We denote by λ_ω the largest singular value of R^{-1} where $\mathsf{R} = i\omega\mathsf{I} - \mathsf{M}$ and M is the matrix defined in equation (5.4). We furthermore introduce $(\mathsf{F},\mathsf{G})^T$ and $(\mathsf{X},\mathsf{Y})^T$ as the principal right and left singular vectors associated with λ_ω, which satisfy

$$\mathsf{R}^{-1*}\mathsf{R}^{-1}\begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix} = \lambda_\omega^2\begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix}, \qquad \begin{pmatrix}\mathsf{X}\\ \mathsf{Y}\end{pmatrix} = \mathsf{R}^{-1}\begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix}. \tag{B 1}$$

The sensitivity analysis is formulated as an optimization problem for determining the minimal stability margin for a specified discrepancy between two transfer functions. We

begin by posing the following Lagrangian

$$L = \lambda_\omega^2 - \begin{pmatrix}\tilde{\mathsf{Y}}\\ \tilde{\mathsf{X}}\end{pmatrix}^* \left[\mathsf{R}\begin{pmatrix}\mathsf{X}\\ \mathsf{Y}\end{pmatrix} - \begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix}\right] - \begin{pmatrix}\tilde{\mathsf{F}}\\ \tilde{\mathsf{G}}\end{pmatrix}^* \left[\lambda_\omega^2 \mathsf{R}^* \begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix} - \begin{pmatrix}\mathsf{X}\\ \mathsf{Y}\end{pmatrix}\right]. \tag{B 2}$$

Rendering this Lagrangian stationary with respect to the variables $(\mathsf{F},\mathsf{G})^T$, $(\mathsf{X},\mathsf{Y})^T$ and λ_ω^2 yields the following expressions for the Lagrange multipliers (denoted by $\tilde{.}$).

$$\begin{pmatrix}\tilde{\mathsf{F}}\\ \tilde{\mathsf{G}}\end{pmatrix} = \mathsf{R}^{-1}\begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix}, \qquad \begin{pmatrix}\tilde{\mathsf{Y}}\\ \tilde{\mathsf{X}}\end{pmatrix} = \lambda_\omega^2\begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix}, \qquad \begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix}^*\begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix} = 1. \tag{B 3}$$

Setting the first variation of the Lagrangian with respect to the control variable R to zero results in an expression for the variation of the squared singular value given by

$$\delta\lambda_\omega^2 = -2\lambda_\omega^2 \mathrm{Re}\left[\begin{pmatrix}\mathsf{F}\\ \mathsf{G}\end{pmatrix}^* \delta\mathsf{R}\begin{pmatrix}\mathsf{X}\\ \mathsf{Y}\end{pmatrix}\right]. \tag{B 4}$$

Using equation (5.4), this expression for $\delta\lambda_\omega^2$ may be expressed in terms of variations in the matrices A, B_2 and C. We obtain

$$\delta\lambda_\omega^2 = 2\lambda_\omega^2 \mathrm{Re}(\mathsf{F}\mathsf{X}^*) : \delta\mathsf{A} + 2\lambda_\omega^2 \mathrm{Re}(\mathsf{F}\mathsf{Y}^*\hat{\mathsf{K}}^*) : \delta\mathsf{B} - 2\lambda_\omega^2 \mathrm{Re}(\mathsf{X}\mathsf{G}^*\hat{\mathsf{L}}) : \delta\mathsf{C}^* \tag{B 5}$$

where : denotes the contraction operator or Frobenius inner product, defined as $\mathsf{A} : \mathsf{B} = \sum_{ij} A_{ij}B_{ij} = \mathrm{trace}(\mathsf{A}^*\mathsf{B})$.

The constraint of a user-prescribed error γ (measured in the $\mathcal{H}_2$-norm) between the full and reduced transfer function may be approximated in the following manner:

$$\mu_\gamma(\mathsf{A},\mathsf{B}_2,\mathsf{C}) = -\gamma^2 + \sum_j |\underbrace{\mathsf{C}(i\omega_j\mathsf{I}-\mathsf{A})^{-1}\mathsf{B} - \hat{\mathsf{C}}(i\omega_j\mathsf{I}-\hat{\mathsf{A}})^{-1}\hat{\mathsf{B}}}_{\phi_j}|^2\Delta\omega_j. \tag{B 6}$$

In the above expression $\Delta\omega_j$ are quadrature coefficients. In our case, we chose a simple first-order approximation of the integral with $\Delta\omega_j = 0.02$ and integration limits $\omega_{min} = -10, \omega_{max} = 10$. The first variation of this functional reads

$$\delta\mu_\gamma = \left(2\sum_j \mathrm{Re}\left[\phi_j(-i\omega_j\mathsf{I}-\mathsf{A}^*)^{-1}\mathsf{C}^*\mathsf{B}^*(-i\omega_j\mathsf{I}-\mathsf{A}^*)^{-1}\right]\Delta\omega_j\right) : \delta\mathsf{A} \tag{B 7}$$

$$+ \left(2\sum_j \mathrm{Re}\left[\phi_j(-i\omega_j\mathsf{I}-\mathsf{A}^*)^{-1}\mathsf{C}^*\right]\Delta\omega_j\right) : \delta\mathsf{B} + \left(2\sum_j \mathrm{Re}\left[\phi_j^*(i\omega_j\mathsf{I}-\mathsf{A})^{-1}\mathsf{B}\right]\Delta\omega_j\right) : \delta\mathsf{C}^*.$$

Lastly, we combine the above expressions and consider the following unconstrained minimization problem

$$J_{\omega,\gamma}(\mathsf{A},\mathsf{B}_2,\mathsf{C}) = -\arctan(\ln(\lambda_\omega^2)/\alpha) + \beta\mu_\gamma^2 + \arctan(\ln(\nu_\omega^2)/\alpha) \tag{B 8}$$

with β as the penalty coefficient. In our computations, we used $\alpha = 10$ and $10^3 \leqslant \beta \leqslant 10^{10}$. Note that the third term in (B 8) penalizes the largest singular value ν_ω of S^{-1} where $\mathsf{S} = i\omega\mathsf{I} - \mathsf{A}$. This is simply to avoid convergence of the optimization algorithm towards solutions where $s = i\omega$ is an eigenvalue of A, which is an undesirable solution. With H and Z as the principal right and left singular vectors of S corresponding to ν_ω and satisfying $\mathsf{S}^*\mathsf{S}\mathsf{H} = \nu_\omega^2\mathsf{H}$ and $\mathsf{Z} = \mathsf{S}\mathsf{H}$ with $\mathsf{H}^*\mathsf{H} = 1$, we can express the first variation of ν_ω^2 in the form $\delta\nu_\omega^2 = 2\nu_\omega^2\mathrm{Re}(\mathsf{H}\mathsf{Z}^*) : \delta\mathsf{A}$.

REFERENCES

Ahuja, S. & Rowley, C. W. 2010 Feedback control of unstable steady states of flow past a flat plate using reduced-order estimators. *J. Fluid Mech.* **645**, 447–478.

Åkervik, E., Hœpffner, J., Ehrenstein, U. & Henningson, D. S. 2007 Optimal growth, model reduction and control in a separated boundary-layer flow using global modes. *J. Fluid Mech.* **579**, 305–314.

Alizard, F., Cherubini, S. & Robinet, J.-Ch. 2009 Sensitivity and optimal forcing response in separated boundary layer flows. *Phys. Fluids* **21**, 064108.

Archambaud, J.P., Arnal, D., Hein, S., de Souza, J. Melo, Hanifi, A., Goddard, J.L., Krier, J. & Donelli, R. 2008 Use of laminar flow technologies for supersonic drag reduction - results of FP project SUPERTRAC. *5th Eur. Cong. Comp. Meth. Appl. Sci. Eng (ECCOMAS)* .

Bagheri, S., Brandt, L. & Henningson, D. S. 2009 Input-output analysis, model reduction and control of the flat-plate boundary layer. *J. Fluid Mech.* **620**, 263–298.

Bagheri, S. & Henningson, D. S. 2010 Simulation of transition delay in boundary layers. *Phil. Trans. Roy. Soc. A* (submitted).

Barbagallo, A., Sipp, D. & Schmid, P. J. 2009 Closed-loop control of an open cavity flow using reduced-order models. *J. Fluid Mech.* **641**, 1–50.

Bewley, T. R. & Liu, S. 1998 Optimal and robust control and estimation of linear paths to transition. *J. Fluid Mech.* **365**, 305–349.

Blackburn, H. M., Barkley, D. & Sherwin, S. J. 2008 Convective instability and transient growth in flow over a backward-facing step. *J. Fluid Mech.* **603**, 271–304.

Boiko, A.V., Dovgal, A.V. & Hein, S. 2008 Control of a laminar separating boundary layer by induced stationary perturbations. *Eur. J. Mech. B/Fluids* **27 (4)**, 466–476.

Burl, J. B. 1999 *Linear optimal control. $\mathcal{H}_2$ and $\mathcal{H}_\infty$ methods*. Addison-Wesley.

Chevalier, M., Hœpffner, J., Åkervik, E. & Henningson, D.S. 2007 Linear feedback control and estimation applied to instabilities in spatially developing boundary layers. *J. Fluid Mech.* **588**, 163–187.

Choi, H. 2008 Control of flow over a bluff body. *Ann. Rev. Fluid Mech.* **40**, 113–139.

Dergham, G., Sipp, D., Robinet, J.-C. & Barbagallo, A. 2011 Model reduction for fluids using frequential snapshots. *Accepted for publication in Phys. Fluids* .

Ehrenstein, U., Passaggia, P.-Y. & Gallaire, F. 2010 Control of a separated boundary layer: Reduced-order modeling using global modes revisited. *Theor. Comp. Fluid Dyn.* pp. DOI 10.1007/s00162–010–0195–5.

Gharib, M. & Roshko, A. 1987 The effect of flow oscillations on cavity drag. *J. Fluid Mech.* **177**, 501–530.

Hecht, F., Pironneau, O., Hyaric, A. Le & Ohtsuka, K. 2005 *Freefem++, the book*. UPMC-LJLL Press.

Ilak, M. 2009 Model reduction and feedback control of transitional channel flow. PhD thesis, Princeton University, Mechanical and Aerospace Engineering.

Ilak, M. & Rowley, C.W. 2008 Modeling of transitional channel flow using balanced proper orthogonal decomposition. *Phys. Fluids* **20**, 034103.

Joshi, S.S., Speyer, J.L. & Kim, J. 1997 A systems theory approach to the feedback stabilization of infinitesimal and finite-amplitude disturbances in plane Poiseuille flow. *J. Fluid Mech.* **332**, 157–184.

Joslin, R.D. 1998 Overview of laminar flow control. *NASA TP* **208705**.

Kim, J. 2003 Control of turbulent boundary layers. *Phys. Fluids* **15**, 1093–1105.

Kim, J. & Bewley, T.R. 2007 A linear systems approach to flow control. *Ann. Rev. Fluid Mech.* **39**, 383–417.

Lin, J.C. 2002 Review of research on low-profile vortex generators to control boundary-ayer separation. *Prog. Aerosp. Sci.* **38**, 389–420.

Marquet, O., Sipp, D., Chomaz, J.-M. & Jacquin, L. 2008 Amplifier and resonator dynamics of a low-Reynolds number recirculation bubble in a global framework. *J. Fluid Mech.* **605**, 429–443.

Rossiter, J.E. 1964 Wind tunnel experiments on the flow over rectangular cavities at subsonic and transonic speeds. *Aero. Res. Counc. R& M* **3438**.

Rowley, C.W. 2005 Model reduction for fluids using balanced proper orthogonal decomposition. *Int. J. Bifurc. Chaos* **15**, 997–1013.

Saric, W.S., Reed, H.L. & White, E.B. 2003 Stability and transition of three-dimensional boundary layers. *Ann. Rev. Fluid Mech.* **35**, 413–440.

Sipp, D., Marquet, O., Meliga, P. & Barbagallo, A. 2010 Dynamics and control of global instabilities in open flows: a linearized approach. *Appl. Mech. Rev.* **63**, 030801.

Sirovich, L. 1987 Turbulence and the dynamics of coherent structures. *Quart. Appl. Math.* **45**, 561–590.

Bibliography

Adams, E. W. & Johnston, J. P. 1988 Effects of the separating shear-layer on the reattachment flow structure part 2: reattachment length and wall shear-stress. *Experiments in Fluids* **6**, 493–499.

Ahuja, S. & Rowley, C. 2008 Low-dimensional models for feedback stabilization of unstable steady states. *AIAA Paper 2008-553, 46th AIAA Aerospace Sciences Meeting and Exhibit* .

Ahuja, S. & Rowley, C. 2010 Feedback control of unstable steady states of flow past a flat plate using reduced-order estimators. *J. Fluid Mech.* **645**, 447–478.

Aider, J.-L. 2008 *Contrôle découlement décollé et aérodynamique automobile*. Research Habilitation (HDR), Pierre et Marie Curie University, France.

Aider, J. L., Danet, A. & Lesieur, M. 2007 Large-eddy simulation applied to study the influence of upstream conditions on the time-dependant and averaged characteristics of a backward-facing step flow. *J. Turbulence* **8** (51), 1–30.

Åkervik, E., Ehrenstein, U., Gallaire, F. & Henningson, D. 2008 Global two-dimensional stability measures of the flat plate boundary-layer flow. *Eur. J. Mech. B/Fluids* **27** (5), 501–513.

Åkervik, E., Hœpffner, J., Ehrenstein, U. & Henningson, D. S. 2007 Optimal growth, model reduction and control in a separated boundary-layer flow using global modes. *J. Fluid Mech.* **579**, 305–314.

Alizard, F. 2007 *Etude de stabilité linéaire globale d'écoulement fortement décollé de couche limite de plaque plane*. Phd thesis, École Nationale Supérieure d'Arts et Métiers, France.

Alizard, F., Cherubini, S. & Robinet, J.-C. 2009 Sensitivity and optimal forcing response in separated boundary layer flows. *Phys. Fluids* **21**, 064108.

Alizard, F. & Robinet, J.-C. 2007 Spatially convective global modes in a boundary layer. *Phys. Fluids* **19** (11), 114105.

Alizard, F. & Robinet, J.-C. 2010 Modeling of optimal perturbations in flat plate boundary layer using global modes: benefits and limits. *Theoret. Comput. Fluid Dynamics* .

Amestoy, P. R., Duff, I. S., Koster, J. & L'Excellent, J.-Y. 2001 A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.* **23** (1), 15–41.

Antoulas, A. C. 2005 *Approximation of Large-Scale Dynamical Systems*. SIAM Publishing, Philadelphia.

Antoulas, A. C. & Sorensen, D. C. 2001 Approximation of large-scale dynamical systems: an overview. *Int. J. Appl. Math. Comput. Sci.* **11** (5), 1093–1121.

Armaly, B. F., Durst, F., Pereira, J. C. F. & Schönung, B. 1983 Experimental and theoretical investigation of backward-facing step flow. *J. Fluid Mech.* **127**, 473–496.

Arnoldi, W. E. 1951 The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.* **9**, 17–29.

Aubry, N., Guyonnet, R. & Lima, R. 1991 Spatio-temporal analysis of complex signals: Theory and applications. *Journal of Statistical Physics* **64**, 683–739.

Aubry, N., Guyonnet, R. & Lima, R. 1992 Spatio-temporal symmetries and bifurcations via bi-orthogonal decompositions. *Journal of Nonlinear Science* **2**, 183–215.

Aubry, N., Holmes, P., Lumley, J. L. & Stone, E. 1988 The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.* **192**, 115–173.

Aubry, N. & Lima, R. 1995 The dynamics of spatio-temporal modulations. *Chaos* **5**, 578–588.

Aubry, N., Lima, R. & Rahibe, M. 2003 Breaking of space-time symmetries in modulated traveling waves. *Chaos* **13**, 541–551.

Bagheri, S., Akervik, E., Brandt, L. & Henningson, D. S. 2009*a* Free methods for the stability and control of boundary layers. *AIAA Journal* **47** (5), 1057–1068.

Bagheri, S., Brandt, L. & Henningson, D. S. 2009*b* Input-output analysis, model reduction and control of the flat-plate boundary layer. *J. Fluid Mech.* **620**, 263–298.

Bagheri, S. & Henningson, D. S. 2011 Transition delay using control theory. *Phil. Trans. Roy. Soc. A* **369** (1940), 1365–1381.

Bagheri, S., Hœpffner, J., Schmid, P. J. & Henningson, D. S. 2009*c* Input-output analysis and control design applied to a linear model of spatially developing flows. *Appl. Mech. Rev.* **62** (2), 020803.

Bagheri, S., Schlatter, P., Schmid, P. J. & Henningson, D. S. 2009*d* Global stability of a jet in crossflow. *J. Fluid Mech.* **624**, 33–44.

Barbagallo, A., Dergham, G., Sipp, D., Schmid, P. & Robinet, J.-C. 2011 Closed-loop control of unsteadiness over a rounded backward-facing step. *Submitted to J. Fluid Mech.* .

Barbagallo, A., Sipp, D., Jacquin, L. & Schmid, P. J. 2008 Control of an incompressible cavity flow using a reduced model based on global modes. *AIAA Paper 2008-3904, 5th AIAA Theoretical Fluid Mechanics Conference* .

Barbagallo, A., Sipp, D. & Schmid, P. 2009 Closed-loop control of an open cavity flow using reduced-order models. *J. Fluid Mech.* **641**, 1–50.

Barkley, D., Blackburn, H. & Sherwin, S. 2008 Direct optimal growth analysis for time-steppers. *Int. J. Num. Meth. Fluids* **57**, 1435–1458.

Barkley, D., Gomes, M. G. & Henderson, R. D. 2002 Three-dimensional instability in flow over a backward-facing step. *J. Fluid Mech.* **473**, 167–190.

Bayly, B. J. 1988 Three-dimensional centrifugal-type instabilities in inviscid two-dimensional flows. *Phys. Fluids* **31**, 56–64.

Beaudoin, J.-F., Cadot, O., Aider, J.-L. & Wesfreid, J. E. 2004 Three-dimensional stationary flow over a backward-facing step. *Eur. J. Mech. B/Fluids* **23**, 147–155.

Bergmann, M. & Cordier, L. 2008 Optimal control of the cylinder wake in the laminar regime by trust-region methods and pod reduced-order models. *J. Comp. Phys.* **227**, 7813–7840.

Bergmann, M., Cordier, L. & Brancher, J.-P. 2005 Optimal rotary control of the cylinder wake using pod reduced-order model. *Phys. Fluids* **17**, 305–314.

Berkooz, G., Holmes, P. & Lumley, J. L. 1993 The proper orthogonal decomposition in the analysis of turbulent flows. *Ann. Rev. Fluid Mech.* **25**, 539–575.

Bewley, T. R. 2001 Flow control: new challenges for a new renaissance. *Progr. Aerosp. Sci.* **37**, 21–58.

Bewley, T. R. & Liu, S. 1998 Optimal and robust control and estimation of linear paths to transition. *J. Fluid Mech.* **365**, 305–349.

Blackburn, H. M., Barkley, D. & Sherwin, S. J. 2008 Convective instability and transient growth in flow over a backward-facing step. *J. Fluid Mech.* **603**, 271–304.

Bres, G. A. & Colonius, T. 2008 Three-dimensional instabilities in compressible flow over open cavities. *J. Fluid Mech.* **599**, 309–339.

Brion, V. 2009 *Stabilité des paires de tourbillons contra-rotatifs: application au tourbillon de jeu dans les turbomachines*. Phd thesis, École Polytechnique, France.

Brown, G. L. & Roshko, A. 1974 On density effects and large structures in turbulent mixing layers. *J. Fluid Mech.* **64**, 775–816.

Buffoni, M., Camarri, S. & Iollo, A. 2006 Low-dimensional modelling of a confined three-dimensional wake flow. *J. Fluid Mech.* **569**, 141–150.

Burl, J. B. 1999 *Linear optimal control. $\mathcal{H}_2$ and $\mathcal{H}_\infty$ methods*. Addison Wesley Longman, Inc. Menlo Park, CA.

Butler, K. M. & Farrell, B. F. 1992 Three-dimensional optimal perturbations in viscous shear flow. *Phys. Fluids A* **4**, 1637–1650.

Cherubini, S. 2010 *Linear and non-linear global instability of attached and separated boundary-layer flows over a flat plate*. Phd thesis, École Nationale Supérieure d'Arts et Métiers, France.

Cherubini, S., Palma, P. D., Robinet, J.-C. & Bottaro, A. 2010*a* A rapid path to transition through non-linear, localized, optimal perturbations in a boundary-layer flow. *Physical Review E* **82**, 066302.

Cherubini, S., Robinet, J.-C., Palma, P. D. & Alizard, F. 2010*b* The onset of three-dimensional centrifugal global modes and their non-linear development in a recirculating flow over a flat surface. *Phys. Fluids* **22**, 114102.

Chevalier, M., Hœpffner, J., Åkervik, E. & Henningson, D. 2007 Linear feedback control and estimation applied to instabilities in spatially developing boundary layers. *J. Fluid Mech.* **588**, 163–187.

Choi, H., Jeon, W.-P. & Kim, J. 2008 Control of flow over a bluff body. *Annual Review of Fluid Mechanics* **40**, 113–139.

Chomaz, J.-M. 2005 Global stabilities in spatially developing flows: non-normality and nonlinearity. *Ann. Rev. Fluid Mech.* **37**, 357–392.

Chomaz, J.-M., Huerre, P. & Redekopp, L. M. 1991 A frequency selection criterion in spatially developing flows. *Study In Applied Mathematics* **84**, 119–144.

Chun, K. B. & Sung, H. J. 1996 Control of turbulent separated flow over a backward-facing step by local forcing. *Experiments In Fluids* **21**, 417–426.

Cortelezzi, L. & Speyer, J. 1998 Robust reduced-order controller of laminar boundary layer transitions. *Phys. Rev. E* **58** (2), 1906–1910.

Cossu, C. & Chomaz, J.-M. 1997 Global measures of local convective instability. *Phys. Rev. Lett.* **78**, 4387–4390.

Cossu, C., Pujals, G. & Depardon, S. 2009 Optimal transient growth and very large-scale structures in turbulent boundary layers. *J. Fluid Mech.* **619**, 79–94.

Crouch, J., Garbaruk, A. & Magidov, D. 2007 Predicting the onset of flow unsteadiness based on global instability. *J. Comp. Phys.* **224**, 924–940.

Davis, T. 1993 User's guide for the unsymmetric-pattern multifrontal package (umfpack). *Tech. Rep.* 1993-020. Computer and Information Sciences Dept., Univ. of Florida.

Delville, J., Cordier, L. & Bonnet, J. 1998 Large-scale structure identification and control in turbulent shear flows. In *Flow Control: Fundamentals and Practice*, pp. 199–273. Springer-Verlag, Berlin.

Denham, M. K. & Patrick, M. A. 1974 Laminar flow over a downstream-facing step in a twodimensional flow channel. *Trans. Inst. Chem. Engrs* **52**, 361–367.

Deprés, D., Reijasse, P. & Dussauge, J. P. 2004 Analysis of unsteadiness in afterbody transonic flows. *AIAA Journal* **42** (12), 2541–2550.

Dergham, G., Sipp, D., Robinet, J.-C. & Barbagallo, A. 2011 Model reduction for fluids using frequential snapshots. *Phys. Fluids* **23**, 064101.

Ding, Y. & Kawahara, M. 1999 Three-dimensional linear stability analysis of incompressible viscous flows using the finite element method. *Int. J. Numer. Mech. Fluids* **31**, 451–479.

Dogval, A. V., Kozlov, V. V. & Michalke, A. 1994 Laminar boundary layer separation: instability and associated phenomena. *Progress in Aerospace Sciences* **30**, 61–94.

Dowell, E. H., Hall, K. C. & Romanowski, M. C. 1997 Eigenmode analysis in unsteady aerodynamics: reduced order models. *Applied Mechanics Reviews* **50** (6), 371–386.

Driver, D. M., Seegmiller, H. L. & Marvin, J. G. 1987 Time-dependent behavior of a reattaching shear-layer. *AIAA Journal* **25** (7).

Dullerud, G. E. & Paganini, F. 2000 A course in robust control theory: A convex approach. In *Texts in Applied Mathematics*, , vol. 36. Springer-Verlag, New York.

Duriez, T. 2009 *Application des générateurs de vortex au contrôle d'écoulements décollés*. Phd thesis, Paris Diderot University, France.

Ehrenstein, U. & Gallaire, F. 2005 On two-dimensional temporal modes in spatially evolving open flows. *J. Fluid Mech.* **536**, 209–218.

Ehrenstein, U. & Gallaire, F. 2008*a* Optimal perturbations and low-frequency oscillations in a separated boundary-layer flow. *AIAA Paper 2008-4232, 5th AIAA Theoretical Fluid Mechanics Conference* .

Ehrenstein, U. & Gallaire, F. 2008*b* Two-dimensional global low-frequency oscillations in a separating boundary-layer flow. *J. Fluid Mech.* **614**, 315–327.

Ehrenstein, U., Passaggia, P.-Y. & Gallaire, F. 2010 Control of a separated boundary layer : Reduced-order modeling using global modes revisited. *Theoret. Comput. Fluid Dynamics* .

Farrell, B. & Ioannou, P. 1993*a* Stochastic forcing of the linearized Navier-Stokes equations. Phys. Fluids A **5** (11), 2600–2609.

Farrell, B. & Ioannou, P. 1996 Generalized stability theory. Part I: Autonomous operators. *J. Atmos. Sci.* **53** (14), 2025–2040.

Farrell, B. F. & Ioannou, P. 1993*b* Stochastic dynamics of baroclinic waves. *J. Atmos. Sci.* **50** (24), 4044–4057.

Farrell, B. F. & Ioannou, P. 2001*a* Accurate low-dimensional approximation of the linear dynamics of fluid flow. *J. Atmos. Sci.* **58**, 2771–2789.

Farrell, B. F. & Ioannou, P. J. 2001*b* State estimation using a reduced order kalman filter. *J. Atmos. Sci.* **58**, 3666–3680.

Fortin, A., Jardak, M., Gervais, J. J. & Pierre, R. 1997 Localization of hopf bifurcations in fluid flow problems. *Int. J. Numer. Meth. Fluids* **24**, 1185–1210.

Fortin, M. 1981 Old and new finite elements for incompressible flows. *Int. J. Num. Meth. Fluids* **1**, 347–364.

Fujimoto, K. & Tsubakino, D. 2008 Computation of nonlinear balanced realization and model reduction based on taylor series expansion. *Systems Control Lett.* **57**, 283–289.

Gallaire, F., Marquillie, M. & Ehrenstein, U. 2007 Three-dimensional transverse instabilities in detached boundary layers. *J. Fluid Mech.* **571**, 221–233.

Galletti, B., Bottaro, A., Bruneau, C. & Iollo, A. 2007 Accurate model reduction of transient and forced wakes. *Eur. J. Mech. B/Fluids* **26**, 354–366.

Galletti, B., Bruneau, C., Zannetti, L. & Iollo, A. 2004 Low-order modelling of laminar flow regimes past a confined square cylinder. *J. Fluid Mech.* **503**, 161–170.

Gardarin, B. 2009 *Contrôle par générateur de vortex d'un écoulement turbulent décollé*. Phd thesis, École Polytechnique, France.

Girault, V. & Raviart, P.-A. 1986 Finite element methods for navier-stokes equations: theory and algorithms. In *Springer Series in Computational Mathematics*, , vol. 5. Springer-Verlag.

Glover, K. 1984 An optimal hankel-norm approximation of linear multivariable systems and their L^∞-error bounds. *Int. J. Control* **39**, 1115–1193.

Glowinski, R. & Pironneau, O. 1992 Finite element methods for navier-stokes equations. *Ann. Rev. Fluid Mech.* **24**, 167–204.

Goldstein, R. J., Eriksen, V. L., Olson, R. M. & Eckert, E. R. G. 1970 Laminar separation, reattachment, and transition of the flow over a downstream-facing step. *Trans. ASME D: J. Basic Engng* **92**, 732–739.

Gresho, P. M., Gartling, D. K., Torczynski, J. R., Cliffe, K. A., Winters, K. H., Garratt, T. J., Spence, A. & Goodrich, J. W. 1993 Is the steady viscous incompressible two-dimensional flow over a backward-facing step at re = 800 stable? *Int. J. Numer. Meth. Fluids* **17**, 501–541.

Hall, K. C., Thomas, J. P. & Clark, W. S. 2002 Computation of unsteady nonlinear flows in cascades using a harmonic balance technique. *AIAA Journal* **40** (5), 879–886.

Hasselmann, K. 1988 Pips and pops: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.* **93** (D9), 11 015–11 021.

Hecht, F., Pironneau, O., Hyaric, A. L. & Ohtsuka, K. 2005 *Freefem++, the book*. UPMC-LJLL Press.

Henningson, D. S. & Åkervik, E. 2008 The use of global modes to understand transition and perform flow control. *Phys. Fluids* **20**, 031302.

Hœpffner, J. 2006 *Stability and control of shear flows subject to stochastic excitations*. Phd thesis, KTH Mechanics, Sweden.

Högberg, M., Bewley, T. R. & Henningson, D. S. 2003 Linear feedback control and estimation of transition in plane channel flow. *J. Fluid Mech.* **481**, 149–175.

Högberg, M. & Henningson, D. S. 2002 Linear optimal control applied to instabilities in spatially developing boundary layers. *J. Fluid Mech.* **470**, 151–179.

Holmes, P. J., Lumley, J. L., Berkooz, G., Mattingly, J. C. & Wittenberg, R. W. 1997 Low-dimensional models of coherent structures in turbulence. *Phys. Rep.* **287** (4), 337–384.

Hood, P. & Taylor, C. 1973 A numerical solution of the navier-stokes equations using the finite element technique. *Comp. and Fluids* **1**, 73–100.

Huang, S.-C. & Kim, J. 2008 Control and system identification of a separated flow. *Phys. Fluids* **20** (10), 101509.

Huerre, P. & Rossi, M. 1998 *Hydrodynamics and nonlinear instabilities*. Cambridge University Press.

Ilak, M., Bagheri, S., Brandt, L., Rowley, C. W. & Henningson, D. S. 2010 Model reduction of the nonlinear complex ginzburg-landau equation. *SIAM J. Applied Dynamical Systems* **9** (4), 1284–1302.

Ilak, M. & Rowley, C. 2006 Reduced-order modeling of channel flow using travelling POD and balanced POD. *AIAA Paper 2006-3194, 3rd AIAA Flow Control Conference* .

Ilak, M. & Rowley, C. 2008 Modeling of transitional channel flow using balanced proper orthogonal decomposition. *Phys. Fluids* **20**, 034103.

Illy, H. 2005 *Contrôle de l'écoulement au-dessus d'une cavité en régime transsonique*. Phd thesis, École Centrale de Lyon, France.

Jackson, C. P. 1987 A finite-element study of the onset of vortex shedding in flow past variously shaped bodies. *J. Fluid Mech.* **182**, 23–45.

Jacquin, L., Molton, P., Deck, S., Maury, B. & Soulevant, D. 2009 Experimental study of shock oscillation over a transonic supercritical profile. *AIAA Journal* **47** (9), 1985–1994.

Joshi, S., Speyer, J. & Kim, J. 1997 A systems theory approach to the feedback stabilization of infinitesimal and finite-amplitude disturbances in plane poiseuille flow. *J. Fluid Mech.* **332**, 157–184.

Kaiktsis, L. 1995 *Instability classification and transition in flow over a backward-facing step*. PhD Thesis, Institut für Energitechnik.

Kaiktsis, L., Karniadakis, G. & Orszag, S. 1996 Unsteadiness and convective instabilities in two-dimensional flow over a backward-facing step. *J. Fluid Mech.* **321**, 157–187.

Kaiktsis, L., Karniadakis, G. E. & Orszag, S. A. 1991 Onset of the three-dimensionality, equilibria, and early transition in flow over a backwardfacing step. *J. Fluid Mech.* **231**, 501–528.

Kaiktsis, L. & Monkewitz, P. A. 2003 Global destabilistion of flow over a backward-facing step. *Phys. Fluids* **15** (12), 3647–3658.

Kim, J. 2003 Control of turbulent boundary layers. *Phys. Fluids* **15**, 1093–1105.

Kim, J. & Bewley, T. 2007 A linear systems approach to flow control. *Ann. Rev. Fluid Mech.* **39**, 383–417.

Kim, J. & Moin, P. 1985 Application of a fractional-step method to incompressible navier-stokes equations. *J. Comput. Phys.* **59**, 308–323.

Kim, T. 1998 Frequency-domain karhunen-loeve method and its application to linear dynamic systems. *AIAA Journal* **36** (11), 2117–2123.

Kress, W. & Lötstedt, P. 2004 Time step restrictions using semi-implicit methods for the incompressible navier-stokes equations. *Tech. Rep.* 2004-030. Department of Information Technology, Uppsala University.

Lall, S., Marsden, J. E. & Glavaski, S. 2002 A subspace approach to balanced truncation for model reduction of nonlinear control systems. *Int. J. Robust Nonl. Contr.* **12**, 519–535.

Lanczos, C. 1950 An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bureau Stan.* **45**, 255–282.

Landahl, M. T. 1980 A note on an algebraic instability of inviscid parallel shear flows. *J. Fluid Mech.* **98**, 243–251.

Laub, A., Heath, M., Paige, C. & Ward, R. 1987 Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Trans. Aut. Contr.* **32** (2), 115–122.

Le, H., Moin, P. & Kim, J. 1997 Direct numerical simulation of turbulent flow over a backward-facing step. *J. Fluid Mech.* **330**, 349–374.

Lee, I., Ahn, S. K. & Sung, H. J. 2004 Three-dimensional coherent structure in a separated and reattaching flow over a backward-facing step. *Experiments in Fluids* **36**, 373–383.

Lehoucq, R. & Scott, J. 1997 Implicitly restarted Arnoldi methods and subspace iteration. *SIAM J. Matrix Anal. Appl.* **23**, 551–562.

Lesieur, M., Begou, P., Briand, E., Danet, A., Delcayre, F. & Aider, J. L. 2003 Coherent vortex dynamics in large-eddy simulations of turbulence. *J. Turbulence* **4**, 016.

Ljung, L. 1999 *System Identification: Theory for the User*. Prentice-Hall, New Jersey.

Lorenz, E. N. 1956 Empirical orthogonal functions and statistical weather prediction. Statistical Forecasting Project Scientific Report No. 1. Massachusetts Institute of Technology, Department of Meteorology.

Lorenz, E. N. 1963 Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20** (2), 130–141.

Luchtenburg, D. M., Günther, B., Noack, B. R., King, R. & Tadmor, G. 2009 Generalized mean-field model of natural and actuated flows over a high-lift configuration. *J. Fluid Mech.* **623**, 283–316.

Luenberger, D. G. 1969 *Optimization by Vector Space Methods*. John Wiley & Sons, Inc.

Lumley, J. L. 1967 The structure of inhomogeneous turbulent flows. In *Atmospheric Turbulence and Radio Wave Propagation*, p. 166. Nauka, Moscow.

Lumley, J. L. 1970 *Stochastic Tools in Turbulence*. Academic Press, New York.

Ma, X. & Karniadakis, G. E. 2002 A low-dimensional model for simulating three-dimensional cylinder flow. *J. Fluid Mech.* **458**, 181–190.

Ma, Z., Ahuja, S. & Rowley, C. W. 2010 Reduced order models for control of fluids using the eigensystem realization algorithm. *Theoret. Comput. Fluid Dynamics* .

Marquet, O. 2007 *Stabilité globale et contrôle d'écoulements de recirculation*. Phd thesis, University of Poitiers, France.

Marquet, O., Lombardi, M., Chomaz, J.-M., Sipp, D. & Jacquin, L. 2009 Direct and adjoint global modes of a recirculation bubble: lift-up and convective nonnormalities. *J. Fluid Mech.* **622**, 1–21.

Marquet, O. & Sipp, D. 2009 Global sustained perturbations in a backward-facing step flow. In *Proceedings of the Seventh IUTAM Symposium on Laminar-Turbulent Transition* (ed. P. Schlatter & D. S. Henningson), pp. 525–528. IUTAM, Springer, Dordrecht, The Netherlands.

Marquet, O., Sipp, D., Chomaz, J.-M. & Jacquin, L. 2008 Amplifier and resonator dynamics of a low-Reynolds number recirculation bubble in a global framework. *J. Fluid Mech.* **605**, 429–443.

Meliga, P. 2008 *A theoretical approach for the onset and control of unsteadiness in compressible afterbody flows*. Phd thesis, École Polytechnique, France.

Merle, X. 2009 *Résolution des équations de stabilité globale en régimes incompressible et compressible avec une méthode aux différences finies de haute précision*. Phd thesis, École Nationale Supérieure d'Arts et Métiers, France.

Moin, P. & Bewley, T. 1994 Feedback control of turbulence. *Appl. Mech. Rev.* **47**, 3–13.

Monokrousos, A., Åkervik, E., Brandt, L. & Henningson, D. S. 2010 Global three-dimensional optimal disturbances in the blasius boundary-layer flow using time-steppers. *J. Fluid Mech.* **650**, 181–214.

Moore, B. 1981 Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Autom. Contr.* **26**, 17–32.

Noack, B., Afanasiev, K., Morzynski, M., Tadmor, G. & Thiele, F. 2003 A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. Fluid Mech.* **497**, 335–363.

Noack, B. & Eckelmann, H. 1994 A global stability analysis of the steady and periodic cylinder wake. *J. Fluid Mech.* **270**, 297–330.

Noack, B. R., Papas, P. & Monkewitz, P. A. 2005 The need for a pressure-term representation in empirical galerkin models of incompressible shear flow. *J. Fluid Mech.* **523**, 339–365.

North, G. R. 1984 Empirical orthogonal functions and normal modes. *J. Atmos. Sci.* **41** (5).

Or, A. C. & Speyer, J. L. 2008 Model reduction of input-output dynamical systems by proper orthogonal decomposition. *Journal of Guidance, Control, and Dynamics* **31–2**, 322–328.

Orr, W. M. 1907 The stability or instability of the steady motions of a perfect liquid and of a viscous liquid. *Proc. Roy. Irish Acad. Sec. A* **27**, 69–138.

Placzek, A. 2009 *Contruction de modèles d'ordre réduit non-linéaires basés sur la décomposition orthogonale propre pour l'aéroélasticité*. Phd thesis, Conservatoire National des Arts et Métiers, France.

Robinet, J.-C. 2007*a* Bifurcations in shock-wave/laminar-boundary-layer interaction: global instability approach. *JFM* **579**, 85–112.

Robinet, J.-C. 2007*b* *Étude des instabilités dans des écoulements spatialement inhomogènes*. Research Habilitation (HDR), Pierre et Marie Curie University, France.

Rodríguez, D. & Theofilis, V. 2010 Structural changes of laminar separation bubbles induced by global linear instability. *J. Fluid Mech.* **655**, 280–305.

Rowley, C. 2005 Model reduction for fluids using balanced proper orthogonal decomposition. *Int. J. Bifurc. Chaos* **15**, 997–1013.

Rowley, C., Colonius, T. & Murray, R. 2004 Model reduction for compressible flows using pod and galerkin projection. *Physica D* **189**, 115–129.

Rowley, C. & Williams, D. 2006 Dynamics and control of high-Reynolds number flow over open cavities. *Ann. Rev. Fluid Mech.* **38**, 251–276.

Rowley, C. W., Colonius, T. & Basu, A. J. 2002 On self-sustained oscillations in two-dimensional compressible flow over rectangular cavities. *J. Fluid Mech.* **455**, 315–346.

Rowley, C. W., Mezić, I., Bagheri, S., Schlatter, P. & Henningson, D. S. 2009 Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127.

Ryckelynck, D. 2002 Réduction a priori de modèles thermomécaniques. *Comptes Rendus Mécanique* **330** (7), 499–505.

Ryckelynck, D. 2005 A priori hyperreduction method: an adaptative approach. *Journal of Computational Physics* **202** (1), 346–366.

Samimy, M., Debiasi, M., Caraballo, E., Serrani, A., Yuan, X., Little, J. & Myatt, J. 2007 Feedback control of subsonic cavity flows using reduced-order models. *J. Fluid Mech.* **579**, 315–346.

Scherpen, J. M. 1993 Balancing for nonlinear systems. *Sys. Contr. Lett.* **21**, 143–153.

Schmid, P. & Henningson, D. 2001 *Stability and Transition in Shear Flows*. Applied Mathematical Sciences 142, Springer.

Schmid, P. J. 2007 Nonmodal stability theory. *Ann. Rev. Fluid Mech.* **39**, 129–162.

Schmid, P. J. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28.

Sedrakian, A. 2000 *Vers une aide à la décision pour les méthodes itératives hybrides parallèles réutilisables*. Phd thesis, Pierre et Marie Curie University, France.

Semeraro, O., Bagheri, S., Brandt, L. & Henningson, D. S. 2011 Feedback control of three-dimensional optimal disturbances using reduced-order models. *J. Fluid Mech.* .

Sipp, D. & Jacquin, L. 2000 Three-dimensional centrifugal-type instabilities of two-dimensional flows in rotating systems. *Phys. Fluids* **12**, 1740–1748.

Sipp, D. & Lebedev, A. 2007 Global stability of base and mean flows: a general approach and its applications to cylinder and open cavity flows. *J. Fluid Mech.* **593**, 333–358.

Sipp, D., Marquet, O., Meliga, P. & Barbagallo, A. 2010 Dynamics and control of global instabilities in open flows: a linearized approach. *Applied Mechanics Reviews* **63**, 030801.

Sirisup, S. & Karniadakis, G. E. 2004 A spectral viscosity method for correcting the long-term behavior of pod models. *J. Comp. Phys.* **194**, 92–116.

Sirovich, L. 1987 Turbulence and the dynamics of coherent structures. *Quart. Appl. Math.* **45**, 561–590.

Storch, H. V., Bürger, G., Schnur, R. & Storch, J.-S. V. 1995 Principal oscillation patterns: a review. *Journal of Climate* **8** (3), 377–400.

T. R. Smith, J. M. & Holmes, P. 2005 Low-dimensional models for turbulent plane couette flow in a minimal flow unit. *J. Fluid Mech.* **538**, 71–110.

Tadmor, G., Noack, B., Morzynski, M. & Siegel, S. 2004 Low-dimensional models for feedback flow control. Part II. Control design and dynamical estimation. *AIAA Paper 2004-2409* .

Tang, D., Kholodar, D., Juang, J.-N. & Dowell, E. H. 2001 System identification and proper orthogonal decomposition method applied to unsteady aerodynamics. *AIAA Journal* **39** (8), 1569–1576.

Tezuka, A. & Suzuki, K. 2006 Three-dimensional global linear stability analysis of flow around a spheroid. *AIAA Journal* **44** (8).

Thangam, S. & Knight, D. 1989 Effect of stepheight on the separated flow past a backward facing step. *Phys. Fluids A* **1**, 604–606.

Theofilis, V. 2000 Advances in global linear instability analysis of nonparallel and three-dimensional flows. *Prog. Aerosp. Sci.* **39**, 249–315.

Theofilis, V., Hein, S. & Dallmann, U. 2000 On the origin of unsteadiness and three-dimensionality in a laminar separation bubble. *Phil. Trans. R. Lond. A* **358**, 3229–3246.

Tihon, J., Legrand, J. & Legentilhomme, P. 2001 Near-wall investigation of backward-facing step flows. *Experiments in Fluids* **31**, 484–493.

Trefethen, L. & Embree, M. 2005 *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press.

Tuckerman, L. S. 1998 Bifurcation analysis for timesteppers. Dynamics Days. Durham, NC, USA, January 9.

Tuckerman, L. S. & Barkley, D. 2000 Bifurcation analysis for timesteppers. In *Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems* (ed. E. Doedel).

Tylli, N. 2003 *Induced global unsteadiness and sidewall effects in the backward-facing step flow: Experiments and numerical simulations*. PhD Thesis, Institut für Energitechnik.

Valencia, A. & Hinojosa, L. 1997 Numerical solutions of pulsating flow and heat transfer characteristics in a channel with a backward-facing step. *Heat Mass Transfer* **32**, 143–148.

Weller, J., Camarri, S. & Iollo, A. 2009*a* Feedback control by low-order modelling of the laminar flow past a bluff body. *J. Fluid Mech.* **634**, 405–418.

Weller, J., Lombardi, E. & Iollo, A. 2009*b* Robust model identification of actuated vortex wakes. *Physica D* **238**, 416–427.

Willcox, K. & Peraire, J. 2002 Balanced model reduction via proper orthogonal decomposition. *AIAA J.* **40**, 2323–2330.

Williams, P. T. & Baker, A. J. 1997 Numerical simulations of laminar flow over a 3D backward-facing step. *Int. J. Numer. Meth. Fluids* **24**, 1159–1183.

Zebib, A. 1987 Stability of a viscous flow past a circular cylindar. *J. Engng. Math.* **21**, 155–165.

Zhou, K. & Doyle, J. 1998 *Essentials of Robust Control*. Prentice-Hall, New Jersey.

Zhou, K., Salomon, G. & Wu, E. 1999 Balanced realization and model reduction for unstable systems. *Int. J. Robust Nonl. Contr.* **9**, 183–198.

Zhou, K., Salomon, G. & Wu, E. 2002 *Robust and Optimal Control*. Prentice-Hall, New Jersey.

Zuccher, S., Bottaro, A. & Luchini, P. 2006 Algebraic growth in a blasius boundary layer: Nonlinear optimal disturbances. *Eur. J. Mech. B/Fluids* **25**, 1–17.

Modèles réduits pour la dynamique linéaire et le contrôle en aérodynamique

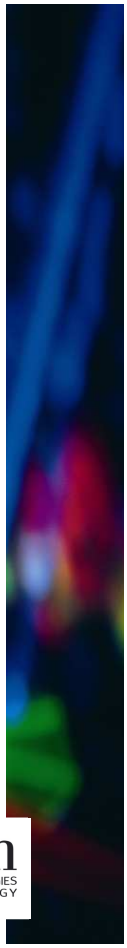
RÉSUMÉ : En aérodynamique, les écoulements décollés sont souvent sujets à de fortes instabilités qui provoquent l'apparition de grosses structures tourbillonnaires. Ces écoulements caractérisés par des instationnarités à basses fréquences sont couramment observés dans les applications aéronautiques et entraînent des effets néfastes tels que d'importantes vibrations des structures ou la génération de bruit. Cette thèse a pour objectif de fournir des modèles d'ordre réduit de tels écoulements aérodynamiques dans le but de concevoir des dispositifs de contrôle optimaux. Un écoulement transitionnel de marche descendante est considéré comme prototype d'écoulement décollé instable. Dans un premier temps, la dynamique linéaire de l'écoulement est étudiée à l'aide d'une analyse de stabilité globale. Nous montrons que l'écoulement amplifie de manière sélective le bruit amont par l'instabilité de Kelvin-Helmholtz. Ensuite, nous utilisons des méthodes de projection pour construire des modèles d'ordre réduit de la dynamique linéaire bidimensionnelle de l'écoulement. Trois approches sont étudiées : (i) l'utilisation des modes globaux les moins stables, (ii) la Décomposition Orthogonale Propre (POD) et (iii) la troncature équilibrée. Cette thèse introduit une méthode des clichés dans le domaine fréquentiel pour calculer les modes contrôlables, observables et équilibrés dominants, ainsi que des techniques pour traiter les systèmes fluides de grande taille. Finalement, nous traitons la question du contrôle en boucle fermée de l'écoulement. Une réduction conséquente des perturbations est obtenue en utilisant une commande Linéaire Quadratique Gaussienne conçue à partir d'un modèle POD.

Mots clés : Instabilités globales, Écoulements décollés, Réduction de modèle, Décomposition Orthogonale Propre, Troncature équilibrée, Contrôle des écoulements.

Reduced-order models for linear dynamics and control in aerodynamics

ABSTRACT : In aerodynamics, separated flows are often subject to strong instabilities which result in the shedding of large-scale vortices. Such low-frequency unsteadiness are commonly encountered in aeronautical applications and lead to detrimental effects such as severe structural vibrations or the generation of extensive noise pollution. This thesis aims at providing low-order models of such aerodynamic flows in order to design optimal control devices. The transitional backward-facing step flow is considered as a prototype of unstable separated flow. Firstly, the linear flow dynamics are examined using a global stability analysis. The flow is found to selectively amplify the upstream noise through the Kelvin-Helmholtz instability. Next, we use projection methods to construct low-order models of the linear two-dimensional dynamics of the flow. Three approaches are investigated: (i) the use of the least damped global modes, (ii) the Proper Orthogonal Decomposition (POD) and (iii) the balanced truncation. This thesis introduces a snapshot method in the frequency domain to compute the leading controllable, observable and balanced modes, as well as techniques to handle large fluid systems. Lastly, the question of the closed-loop control of the flow is addressed. An effective reduction of the perturbations is obtained by using a Linear Quadratic Gaussian compensator designed from a POD model.

Keywords : Global instability, Separated flows, Model reduction, Proper Orthogonal Decomposition, Balanced truncation, Flow control.



ARTS ET MÉTIERS ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY