Model reduction and feedback control of transitional channel flow

Miloš Ilak

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
By the Department of
Mechanical and Aerospace Engineering

Advisor: Clarence W. Rowley

November 2009

Abstract

This dissertation examines the use of reduced-order models for design of linear feedback controllers for fluid flows. The focus is on transitional channel flow, a canonical shear flow case with a simple geometry yet complex dynamics. Reduced-order models of the linearized Navier-Stokes equations, which describe the evolution of perturbations in transitional channel flow, are computed using two methods for snapshot-based balanced truncation, Balanced Proper Orthogonal Decomposition (BPOD) and Eigensystem Realization Algorithm (ERA). The performance of these models in feedback control is evaluated in both linearized and nonlinear Direct Numerical Simulations (DNS) of channel flow.

The first part of the dissertation describes the application of BPOD to very large systems, and the detailed evaluation of the resulting reduced-order models. Exact balanced truncation, a standard method from control theory, is not computationally tractable for very large systems, such as those typically encountered in fluid flow simulations. The BPOD method, introduced by Rowley (2005), provides a close approximation. We first show that the approximation is indeed close by applying the method to a 1-D linear perturbation to channel flow at a single spatial wavenumber pair, for which exact balanced truncation is tractable. Next, as the first application of BPOD to a very high-dimensional linear system, we show that reduced-order BPOD models of a localized 3-D perturbation capture the dynamics very well. Moreover, the BPOD models significantly outperform standard Proper Orthogonal Decomposition (POD) models, as illustrated by a striking example where models using the POD modes that capture most of the perturbation energy fail to capture the perturbation dynamics.

Next, reduced-order models of a complete control system for linearized channel flow are obtained using ERA, a computationally efficient method that results in the same reduced-order models as BPOD. Linear Quadratic Gaussian (LQG) compensators, which include a reduced-order estimator based on a small number of velocity measurements, are designed for these models and used for feedback control of the energy growth of a localized perturbation near the channel wall. The performance of both a localized body-force near the channel wall and wall blowing/suction as actuation mechanisms is first studied in linearized DNS. It is found that the linear compensators are successful in reducing the growth of the perturbation energy, and that the body force actuation results in a larger decrease of the perturbation energy growth than actuation using wall blowing/suction. We then proceed to show that these compensators are also able to prevent transition to turbulence for nonlinear simulations in some cases, despite performance limitations imposed by the spatial separation of the perturbation and the actuator.

Finally, since it is found that a fundamentally nonlinear mechanism of transition is not captured by the linear models, it is of interest to study nonlinear models for flow control. As a first step towards investigating nonlinear balanced truncation

models of channel flow, a method for empirical nonlinear balanced truncation proposed by Lall *et al.* (2002) is tested on a nonlinear 1-D model problem, the Complex Ginzburg-Landau (CGL) equation. The performance of the resulting models is compared to the performance of nonlinear models obtained by projection of the full equation onto modes computed via balanced truncation of the linear part of the CGL equation. It is found that the models obtained by the latter approach are not only able to capture the dynamics of the nonlinear CGL equation, but that they also outperform the models obtained using the empirical nonlinear balanced truncation method.

Acknowledgments

First and foremost, I thank Prof. Clancy Rowley for his guidance and encouragement over the years. He introduced me to the field of model reduction and suggested this project as my dissertation topic. Clancy was willing to spend hours helping me out with details early on, and later his keen insight would always nudge me in the right direction when things seemed to be going nowhere. He also taught me many useful computing practices, and his commitment to constantly improving the great technical resources in the lab kept me productive and motivated.

I would also like to thank Profs. Philip Holmes and Luigi Martinelli for their helpful comments as the readers of this dissertation, and Profs. Alexander Smits and Robert Stengel for being examiners. I have also benefited from interaction with all of them in different capacities over the years as student, teaching assistant, or simply through friendly conversations.

It would not have been possible to complete this work without the support of my labmates. Room H-125 has been an excellent work environment for the past five years thanks to the company of Sunil Ahuja (now at UTRC), Zhanhua Ma, Melissa Green, Juan Melli, Prof. Mingjun Wei (now at New Mexico State University), Steve Brunton, Kunihiko (Sam) Taira, Peter Norgaard, Lauren Padilla, Jonathan Tu, Brandt Belson, and, most recently, Mort. Zhanhua, Sam, and Melissa also provided feedback on drafts of some chapters. I am especially grateful to Sunil and Zhanhua for their friendship and for all the discussions over the years; the outcome of many of them is reflected in this work.

The students of the MAE department, past and present, provided a genuinely fun and supportive environment. I will miss walking over to the faculty lounge with my coffee cup at 4pm, lunch at the big table in the atrium, the happy hours at The Ivy Inn, the annual Festivus celebration and other festive occasions. My dear friend Raghu Kukillaya is especially acknowledged for all the long evenings of preparing together for generals, as well as for all the foosball games over the years.

Very special thanks are due to Jessica O'Leary for all the forms, letters, scheduling, and in general for going out of her way to make life easier for us graduate students. Bill Wichser has been an invaluable source of computing support, offering a lot of useful advice and always responding timely to emergency situations.

Part of the work presented here was performed in close collaboration with Shervin Bagheri during a visit to the Department of Mechanics at the Royal Institute of Technology in Stockholm, thanks to the hospitality of Prof. Dan Henningson and Docent Luca Brandt. Also, many thanks are due to Shervin, Espen Åkervik, Antonios Monokrousos, and the rest of the group for the enjoyable time in Stockholm. *Tack så mycket för allting!*

Outside of E-Quad, it has been a privilege to share both good and bad moments with my dear friends Ronny Luss (who was also a great roommate), David Cham-

pagne, Sonya Nikolova, Najwa Aaraj, Stephane Kena-Cohen, Zubin Jacob, and many others. Very special thanks are also due to Srđan Krstić, Đorđe Milićević, Đorđe Nikolić, Andrea Nedić, Isidora Ljuri, Nebojša Stanković, Jelena Bradić, Milena Stopić, and the rest of the Serbian community for all the good times over the years. I am also grateful to the Bulgarians on campus for all the language tables, and to the Princeton University Brewers for teaching me how to do some fun experimental work in my spare time.

I have been fortunate to have one of my closest friends, Venelin Saltirov, not too far away in New York, and to see him often and share many good times. Thanks to modern technology, my friends back home have remained a part of my life despite the distance that separates us. In particular, late night (for them early morning) chats with Stanislava Todorova and Ognian Kassabov have been a source of encouragement, especially during the writing of this dissertation.

Thanks are also due to everyone at the Engineering and Physics departments at Swarthmore College, where my academic journey started nine years ago, and in particular to Profs. Fred Orthlieb, Erik Cheever, Carr Everbach and Nelson Macken from Engineering, for providing me with the foundations I built upon at Princeton.

Finally, I would not have arrived at the point of writing these acknowledgments without the unwavering support of my family. Being so far away from home has been difficult both for me and for my parents and grandparents, but I have always been able to count on their encouragement, and I am counting on it for my future endeavors as well. This dissertation is dedicated to them.

This work was supported initially by a Princeton University first-year graduate fellowship, and later by National Science Foundation grant CMS-0347239. This dissertation carries the number T-3203 in the records of the Department of Mechanical and Aerospace Engineering.

Dedicated to my parents and grandparents,
and to the loving memory of my grandfather Jerko Ilak.

Mami, tati, Olici, deka Miši i deka Jeri.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Fluid flows are ubiquitous. Significant advances have been made in understanding them, in particular in the past hundred years. The need to manipulate flows to our advantage has been one of the driving forces in fluid mechanics research, and work towards this goal has grown to be a mature discipline known as flow control.

Fluid flows are complex and thus challenging to describe fully and accurately. Nevertheless, it has been observed that their behavior may often be described in simple terms, with only one or a few key mechanisms. Very similar flows can arise in physically very different situations. For example, turbulent jets of gas ejected from galactic centers may look much like jets arising in the laboratory on earth. Behavior dominated by these key mechanisms may often be described as *low-dimensional*, since the number of degrees of freedom that determine it may be quite low.

Besides its intriguing physics, inherent low-dimensional behavior of fluid flows has practical significance — if we are able to isolate and control only the main mechanism, we may not need a full and completely accurate description of the entire flow. This observation is at the heart of the model reduction approach taken in this thesis. The idea is to extract the key features of a highly complex system and represent them in a smaller, simple one, develop ways of controlling the simple system, and apply the results to the original fluid flow. This approach is not without substantial difficulties and is the subject of active research. While this thesis addresses all the stages of the process, the main focus is on the first task — the extraction of reliable models and the verification of their usefulness — followed by a test of the obtained models in the specific context of control of transitional channel flow.

1.2 Current advances in control of shear flows

In the effort to develop flow control strategies, understanding of the underlying physics of the flow and applying the tools of control theory to that flow go hand in hand. While there are many efforts in both areas, exchange of information between the fluid mechanics and control theory communities is very important, as emphasized by Bewley (2001). The key efforts in understanding and controlling shear flows undertaken in the past few decades are outlined below. The focus here is on works relevant to the topic of this dissertation. For a thorough review of flow control efforts in general, the comprehensive book by Gad-el Hak (2000) is recommended.

1.2.1 Results in model reduction

Besides seeking a simple description of a fluid flow, the main motivation for model reduction is the inability to work with tractable models of the full flow. The Navier-Stokes equations, which describe well the vast majority of flows of practical interest, are infinite-dimensional, and their discretization results in very high-dimensional systems. While it has become standard practice to solve these equations using high-performance computers, standard approaches for control design for these equations are still not tractable, and low-order descriptions are sought for that goal. The method of Galerkin projection, in which high-dimensional equations are projected onto an appropriate low-dimensional basis, is the main method used to obtain tractable models of lower dimension. The directions that span the basis are also known as *modes*, and the terms 'set of modes' and 'low-dimensional basis' (or simply 'basis' when it is clear from the context that it is low-dimensional) are used interchangeably in this thesis. We next outline the methods for obtaining such a basis that are used in this thesis. A detailed discussion of these methods is given in next chapter.

Proper Orthogonal Decomposition (POD). POD was introduced in the context of fluid flows by Lumley (1967, 1970) A detailed treatment of the method is given in Holmes *et al.* (1996). The method was initially used to identify coherent structures in flows (see, for example, Moin & Moser, 1989).

Since the behavior of a given flow simulation or a time sequence of experimental data may be described very well in a low-dimensional manner by projection of the *data* onto POD modes, it has been assumed that low-dimensional models may be obtained as well by Galerkin projection of the *dynamics* onto the leading POD modes. Low-dimensional POD models have been shown to be successful at this task in some flows. Examples include cavity flow (Rowley, 2001; Rowley & Juttijudata, 2005), a temporally evolving free shear layer (Wei & Rowley, 2009), flow past a cylinder (Noack *et al.*, 2003), boundary layer flow (Aubry *et al.*, 1988), flow past

an airfoil in 2-D (Luchtenburg *et al.*, 2009) and channel flow in a minimal flow unit (Podvin & Lumley, 1998). On the other hand, POD models often fail to capture some of the most important aspects of the dynamics of the original system, for example in Couette flow (Smith, 2003; Smith *et al.*, 2005) and transitional channel flow (Ilak & Rowley, 2008*b*). A major disadvantage of POD models is that in a feedback control setting they do not capture the inputs and outputs of the original system. This deficiency has been addressed by the introduction of techniques such as shift modes (Noack *et al.*, 2003). Other techniques for improvement of the performance of POD models are reviewed by Siegel *et al.* (2008).

Balanced truncation and BPOD. Balanced truncation, a model reduction method that has become standard in modern control theory, was first introduced by Moore (1981). Standard references on the method include Dullerud & Paganini (1999) and Zhou *et al.* (1996). A detailed introduction of the method will be given in the next chapter. The essence of the method is the inclusion of both inputs and outputs of a control system in the reduced-order model, and balancing of the sensitivity of model states to inputs with their potential to affect outputs, resulting in superior model performance. Balanced truncation of closed-loop control systems was first studied by Jonckheere & Silverman (1983).

While balanced truncation has been applied to fluid flows in 1-D problems (Farrell & Ioannou, 2001), the first computationally tractable approximation of the method for very large problems was recently developed by Rowley (2005). The method, known as Balanced Proper Orthogonal Decomposition (BPOD), has been used successfully to compute reduced-order models and design feedback control in several recent works. Unstable steady states for 2-D flow past a flat plate at an angle of attack were stabilized by Ahuja & Rowley (2008, 2009), and the growth of perturbations in a 2-D linearized Blasius boundary layer was suppressed using a BPOD reduced-order compensator by Bagheri *et al.* (2009*b*). The first applications of BPOD to a moderately high-resolution 3-D system were in Ilak & Rowley (2008*b*,*a*), where the application is modeling and control of transitional flow. The results of these two works will be discussed in detail in the subsequent chapters.

It has been shown recently (Ma *et al.*, 2009*a*) that a well-known method from system identification theory known as Eigensystem Realizaton Algorithm (ERA, see, for example, Juang & Pappa, 1985) is equivalent to balanced truncation. This method results in significant computational savings over both exact balanced truncaiton and BPOD, and will also be used in this thesis. A drawback of the method is that the two sets of modes required to obtain balanced models via Galerkin projection are not computed, only the models themselves are obtained. This method is discussed further in Section 2.3.4.

Balanced truncation for nonlinear systems. While in this thesis we focus on linear techniques, ultimately it is desirable to obtain nonlinear models that will

capture the input/output behavior of the original nonlinear flow. The standard balanced truncation method is defined only for linear systems, and much research has been done on extending the method to nonlinear systems, starting with the work of Scherpen (1993). The first-principles approach taken in that work and in some other more recent works is not immediately applicable to high-dimensional systems, since Hamilton-Jacobi equations need to be solved. More recent work on similar methods includes Fujimoto *et al.* (2002); Verriest & Gray (2004); Fujimoto & Tsubakino (2008). A comparison of linear and nonlinear balancing and an interpretation of linear balancing that extends to nonlinear systems is given by Krener (2006).

On the other hand, an empirical method in the spirit of snapshot-based methods such as BPOD has been introduced by Lall *et al.* (2002). This method will be used in this thesis in an application to the Complex Ginzburg-Landau (CGL) equation. The CGL equation has been studied widely as a basic model for limit cycles such as vortex shedding in fluid flows (see, for example, Chomaz, 2005; Cossu & Chomaz, 1997). Control of the CGL equation has been widely studied (Lauga & Bewley, 2004; Cohen *et al.*, 2005; Bagheri *et al.*, 2009*c*; Aamo *et al.*, 2005). In Chapter 6 we compute nonlinear reduced-order models of the CGL equation both using the method of Lall *et al.* (2002) and via Galerkin projection of the nonlinear equation onto modes from linear balanced truncation.

1.2.2 Optimal control and estimation

In this thesis the reduced-order models are designed for the purpose of feedback control. The approach to control that has become standard over the last several decades is based on the notion of a *state-space*, where a vector of variables defined as the state of the system, along with inputs and outputs, describes the dynamics of the system. The models we compute are thus known as *state-space models*. The problems the control designer needs to solve are the design of the appropriate controller (i.e., computation of feedback gains) and the estimation of the system state, which is typically unavailable to the closed-loop system but is required to compute the feedback control input, from a small number of measurements. Both problems have been thoroughly studied, leading to the development of the standard Linear-Quadratic-Gaussian (LQG) compensator, which we use in this thesis.

Computing a suitable reduced-order model is often the most difficult part of the control design procedure, as the methods for the computation of the LQG compensator are standard and have negligible computational cost for models of low order such as the ones we use. Routines for the solution of the necessary matrix Riccati equations are readily available in MATLAB. The necessary concepts and techniques from modern control theory will be introduced throughout the thesis as needed. Some standard references on the state-space method and modern and

optimal control include Stengel (1994); Ogata (1997); Friedland (1986); Zhou *et al.* (1996); Dullerud & Paganini (1999).

The issue of estimation has been as important as control design, since in the standard state-space approach the entire system state is needed, and in practice one typically needs to recover that information from a small number of measurements. Considerable attention has been devoted recently to the study of estimation in shear flows (Högberg *et al.*, 2003; Hœpffner *et al.*, 2005; Chevalier *et al.*, 2006), and it has been shown that the flow can be reconstructed to a satisfactory degree using measurements of wall skin-friction and pressure (Bewley & Protas, 2004).

Finally, the field of nonlinear control has seen significant development in the recent decades (see, for example, Nijmeier & van der Schaft, 1990; Krstić *et al.*, 1995). In this thesis, we focus on preventing the growth of a small perturbation and controlling it while it is still small and thus described well by linearized equations. The application of nonlinear control methods to nonlinear reduced-order models such as ones we obtain in Chapter 6 is beyond the scope of this work and is the subject of future work.

1.2.3 Understanding transition and turbulence in shear flows

Recent research on transition in shear flows has focused on the large non-normal transient growth of exponentially stable linear perturbations to the laminar flow, which is thought to lead to the so-called 'subcritical' or 'bypass' transition (Schmid & Henningson, 2001; Trefethen *et al.*, 1993; Farrell, 1988; Butler & Farrell, 1992; Bamieh & Dahleh, 2001; Jovanović & Bamieh, 2005; Reddy *et al.*, 1998). A comprehensive treatment of the subject is given by Schmid & Henningson (2001). Some early works on the subject demonstrated that linear growth of energy in perturbations to inviscid flow is possible even if the laminar profile has no inflection points (Ellingsen & Palm, 1975; Landahl, 1980). Later, the view of non-orthogonality of the eigenfunctions of the Orr-Sommerfeld operator, which governs the linear dynamics of perturbations to laminar flow, was established as the likely explanation for subcritical transition. Individually these eigenvectors (also known as eigenmodes) grow or decay in time depending on the Reynolds number (Re) — for most shear flow there is a critical value of Re above which there are exponentially unstable eigenvectors. Due to the non-normality, the norm of the sum of multiple stable eigenmodes, and thus the perturbation energy, may grow in time initially before decay. While a purely linear perturbation will eventually decay at subcritical Re, in the actual flow governed by the full Navier-Stokes equations nonlinear effects may be amplified once the perturbation has grown linearly, leading to transition to turbulence. This process is called 'bypass transition', since transition happens faster than the growth of unstable eigenmodes (which are not present if Re is subcritical), also known as Tollmien-Schlichting waves, which are thus 'bypassed'.

The first highly accurate numerical computations of the eigenvalues of the Orr-Sommerfeld operator Orszag (1971), which showed when transition to turbulence can be expected using standard linear stability theory, and the discrepancy between this finding and experimental and numerical evidence have prompted research into the causes of subcritical transition. In the experiments of Patel & Head (1969) it was found that transition for channel flow may occur at Reynolds numbers as low as 1350, which is considerably lower than the lowest critical value $Re \approx 5772$ computed by Orszag (1971). Other experiments have also found evidence of subcritical transition in shear flows (Nishioka *et al.*, 1975; Kao & Park, 1970). Subcritical transition has also been observed in numerical computations (see, for example, Schmid & Henningson, 1992; Reddy *et al.*, 1998).

In the light of the evidence for its role in transition, the suppression of transient growth is of interest in controlling transition. One of the main potential practical applications is drag reduction, since the drag in laminar flows in significantly lower than drag in turbulent flows. On the other hand, promotion of transition may be desirable in some applications, such as chemical processes where fast mixing is desirable. Farrell & Ioannou (1993, 1996); Trefethen *et al.* (1993) studied the non-normal growth of perturbations to shear flows in detail. Low-dimensional models of transition obtained by inclusion of fundamental physical mechanisms rather than model reduction of the Navier-Stokes equations are reviewed by Baggett & Trefethen (1997). The transition of localized perturbations was studied by Henningson *et al.* (1993), and the transition of some standard classes of linear perturbations in channel flow was studied in detail by Reddy *et al.* (1998).

Another approach to transition and turbulence in shear flows has been the study of the 'edge of chaos'. There is evidence that for some shear flows states exist that are neither laminar nor turbulent, and divide the state-space into two — a region where all solutions will re-laminarize, and a region where the solutions eventually enter a 'non-trivial' attractor, which is interpreted as a signature of turbulence (see, for example, Skufca *et al.*, 2006; Schneider *et al.*, 2008). In Kim & Moehlis (2008), the edge states for a low-dimensional model of Couette flow were studied. This type of study is potentially useful for control design, as it may point out the region in phase space where solutions may be expected to re-laminarize in finite time.

The main flow of interest in this thesis is transitional channel flow. By transitional we mean the regime during which the perturbation is assumed to be small enough for the linearized mechanism to dominate, although nonlinear effects are present as well. Channel flow was chosen since it is a canonical flow and a standard test problem. The simple geometry allows for relatively inexpensive accurate Direct Numerical Simulation (DNS). Channel flow simulations have been standard benchmarks in the study of turbulent flows, starting with the early investigations of Moin & Kim (1980, 1982, 1985); Kim & Moin (1986), through the well-known simulation of Kim *et al.* (1987), and including a more recent simulation at a higher Reynolds number by Moser *et al.* (1999). Transition in channel flow was studied, among others,

by Orszag & Kells (1980); Orszag & Patera (1980), later by Sandham & Kleiser (1992), and in the already mentioned works of Henningson *et al.* (1993) and Reddy *et al.* (1998). Channel flow will be introduced in more detail in Chapter 3.

1.2.4 Control of transition and turbulence in shear flows

The application of control theory and model reduction methods to transitional shear flows has expanded in parallel with the improved understanding of transition. In addition to the work in control already mentioned in Section 1.2.1, other approaches that do not necessarily involve the model reduction methods described in this thesis have been attempted. An overview of the recent progress in flow control using linear control theory is given by Kim & Bewley (2007).

One of the first successful attempts at controlling turbulent channel flow is the work of Choi *et al.* (1994). Later, efforts to control the linearized Navier-Stokes equations were undertaken initially on single-wavenumber perturbations (Joshi *et al.*, 1997; Bewley & Liu, 1998). 1-D perturbations will be mentioned throughout this thesis. The usual approach to treating discretized Navier-Stokes equations for shear flows in 3-D is to employ a Fourier decomposition in the streamwise and spanwise directions, resulting in a set of 1-D problems at each spatial wavenumber pair, as discussed later in Section 3.1.2. The term '1-D perturbation' in this thesis will refer to such a 1-D problem at a particular wavenumber pair. We note that in physical space these perturbations have three-dimensional structure, periodic in both streamwise and spanwise directions.

In the works by Farrell & Ioannou (2001) and Lee *et al.* (2001), balanced truncation was applied to linearized channel flow at particular wavenumber pairs the standard algorithms are applicable, since the full system is 1-D. The controllers of Lee *et al.* (2001) were then tested on nonlinear channel flow. Control and estimation of linearized channel flow was studied by Högberg *et al.* (2003). In that work, wall blowing/suction was used as actuation and the flow field was recreated based on a measurement of wall skin-friction. The controllers were shown to be successful, although they are designed at each streamwise and spanwise spatial wavenumber pair separately, and then combined, and actuation at all spatial wavenumber pairs is assumed. On the other hand, in this thesis, the controllers are designed for the 3-D flow without a similar decomposition into wavenumber pairs, and only localized actuators are considered.

The input/output behavior of the linearized operator was studied by Bamieh & Dahleh (2001) and Jovanović & Bamieh (2005), and it was shown that the energy amplification due to the non-normality of the underlying operator can be on the order of $\mathcal{O}(Re^3)$. The latter work demonstrates that the streamwise velocity is most receptive to perturbations and thus correlates best to perturbation kinetic energy. This finding is used to guide output selection in Chapter 5.

More recently, Åkervik *et al.* (2007) and Barbagallo *et al.* (2008) studied control of a cavity flow using the eigenfunctions of the full 2-D linearized Navier-Stokes equations, also known as *global eigenmodes*. Although successful models can be designed using these modes, the modes need to be selected carefully so that the model captures the inputs and outputs of the system (Hœpffner, 2006). BPOD models, on the other hand, are designed to capture well system inputs and outputs, as will be demonstrated in this thesis.

Other techniques for control have been applied to channel flow recently, such as motion planning (Cochran & Krstić, 2009), as well as boundary control for 2-D channel flow (Aamo *et al.*, 2003) and 3-D channel flow (Cochran *et al.*, 2006).

1.3 Organization and contributions

Overall, the main contributions of this work are: the first computation of reduced-order BPOD and ERA models for a 3-D flow with localized actuation without decomposition into 1-D perturbations, the use of these models for feedback control in both linear and nonlinear DNS, and a preliminary investigation of balanced truncation for a 1-D nonlinear model problem for fluid flows. Specifically, the contributions of each chapter are outlined below.

Chapter 2. The main idea of model reduction via Galerkin projection, along with a discussion of the choices of basis for projection, is introduced in Section 2.1. A brief introduction to POD reduced-order models is given in 2.2. Next, balanced truncation and its computation for large systems is introduced in Section 2.3. Eigensystem Realization Algorithm (ERA), a method recently shown to be equivalent to balanced truncation, but without the need for adjoint simulations, is introduced in Section 2.3.4. Extensions of balanced truncation to nonlinear systems are introduced in Section 2.3.5.

Chapter 3. The governing equations for channel flow are introduced in Section 3.1, along with the appropriate linearization and the adjoint equations derived with the appropriate inner product. The numerical computations involved in model reduction, including steps for verification of results, are described in Section 3.2. The software tools used in the thesis are also introduced in 3.3, as they are tightly coupled to the underlying flow physics.

Chapter 4. A detailed investigation of BPOD and a comparison to POD is described in Chapter 4. The choices of actuators and measurements are discussed in Section 4.1. A detailed comparison of BPOD to exact balanced truncation and POD for a 1-D system, similar to the 1-D systems used in previous works (see, for example, Farrell & Ioannou, 2001), is shown in Section 4.2. Next, the first application of BPOD to modeling of a large 3-D system is discussed in Section 4.3, where the reduced-order models of a localized perturbation are presented. An analysis of the

dynamics of a POD model described in Section 4.3, inspired by the recent work of Mezić (2004) in graph decomposition analysis of dynamical systems, is given in Section 4.4. This chapter investigates only linearized channel flow.

Chapter 5. Next, balanced models are used to design both feed-forward and feedback controllers, first for linearized, and then for nonlinear DNS runs. For these cases, ERA is used to compute models. Two choices of actuation — a velocity distribution corresponding to a body force and localized wall blowing/suction on the lower channel wall — are considered. It is shown that significant reduction in the perturbation energy can be obtained simply by designing a controller that minimizes the streamwise velocity component of the perturbation at a particular location. Preventing transition to turbulence for full nonlinear Navier-Stokes equations is also achieved in some nonlinear DNS runs.

Chapter 6. Finally, although linear models can be useful, it is ultimately desirable to develop nonlinear models of fluid flows, and this chapter describes our efforts towards that goal applied to the Complex Ginzburg-Landau equation. This work was done mostly during a visit to the Mechanics Department of the Royal Institute of Technology in September–December, 2007, in close collaboration with Shervin Bagheri, and under the joint guidance of Prof. Dan Henningson, Docent Luca Brandt and Prof. Clarence Rowley.

Chapter 7. The conclusions of this work and the many possible directions for future work are outlined in the final chapter.

Some technical details and examples of computational procedures not essential to the main flow of the presentation have been included in appendices. Most of the results described in this thesis and preliminary studies leading up to these results have been published in Rowley & Ilak (2006); Ilak & Rowley (2006, 2008*b*) and Ilak & Rowley (2008*a*), and two articles with the results of Chapters 5 and 6 respectively are in preparation.

Chapter 2

Reduced-order models and balanced truncation

Model reduction is at the heart of the approach to flow control taken in this thesis. Therefore, we begin by introducing the idea of projecting a high-dimensional system onto a low-dimensional space and we review standard methods for it, with a detailed description of the particular methods used in this work. We focus on methods that are tractable for very large systems typical for high-resolution discretization of the Navier-Stokes equations for simulation of fluid flows. Such systems can have dimension of $\mathcal{O}(10^{5-9})$, rendering many standard control design methods intractable and necessitating the use of reduced-order models to design controllers.

A standard approach for obtaining a low-dimensional description of a system is Galerkin projection onto a low-dimensional subspace, which has been a standard method, especially for projecting onto Proper Orthogonal Decomposition (POD) modes (Holmes *et al.*, 1996). The projection, however, need not be orthogonal, and a second set of modes, bi-orthogonal[1] to the original set, can be used. This method, known as Petrov-Galerkin projection, is used in balanced truncation, a standard method that will be shown in subsequent chapters to outperform POD significantly.

Balanced truncation (Moore, 1981) is a well-established method in linear control theory. Compared to other methods for model reduction, balanced truncation takes into account both the inputs and the outputs of the system, while also capturing the underlying system dynamics that is relevant to those inputs and outputs. Snapshot-based balanced truncation (Rowley, 2005), also known as BPOD in some cases, which allows for the computation of balanced reduced-order models for very large systems, is the key method used in this thesis. We also discuss the recently discovered equivalence between the Eigensystem Realization Algorithm (ERA) (Ma *et al.*, 2009*a*; Juang & Pappa, 1985; Juang & Phan, 2001) and balanced truncation,

[1]See Eq. 2.5 for a definition of bi-orthogonality.

resulting in tremendous computational savings for computing balanced models. This method has been used for the results that will be presented in Chapter 5.

Finally, Petrov-Galerkin projection can be used for both linear and nonlinear systems. In particular, nonlinear equations can be projected onto a basis computed for linear equations, or alternatively, a basis of modes obtained from nonlinear simulations, such as that proposed by Lall *et al.* (2002), described later in this chapter. In Chapter 6 we compute nonlinear models using modes obtained from snapshot-based balanced truncation of a linear system, first for the nonlinear Complex-Ginzburg Landau (CGL) equation.

It would not be possible to thoroughly discuss all the mathematical and control theoretic aspects of model reduction within the confines of a dissertation chapter. Therefore, for an in-depth understanding, the reader is referred to standard references, for example the textbooks of Zhou *et al.* (1996) and Dullerud & Paganini (1999) for state-space systems and balanced truncation, and Holmes *et al.* (1996) for Galerkin projection and POD.

2.1 Reduced-order models

A reduced-order model of a high-dimensional system has low dimension (i.e., a smaller number of degrees of freedom) while retaining desirable properties of the original systems. For control systems, typically the most desirable property to be preserved is the input-output behavior. A reduced-order model for control design may be obtained by Galerkin projection onto a set of modes.

2.1.1 Galerkin projection

The idea of Galerkin projection is, given a system:

$$\dot{x} = f(x), \qquad x(t) \in \mathcal{X}, \tag{2.1}$$

where $\mathcal{X}$ is a high-dimensional Hilbert space (for example, $\mathcal{X} = \mathbb{R}^n$, where n is large), to project onto a low-dimensional subspace $T \subset \mathcal{X}$, which can alternatively be thought of as the span of the most important[2] basis functions (modes) for the state space $\mathcal{X}$. We start by noting that we can represent the dynamics of $x(t)$ in a given basis $\{\theta_j\}$ as:

$$x(t) = \sum_{j=1}^{n} a_j(t)\theta_j, \tag{2.2}$$

[2]What is meant by 'important' modes is different for different methods, and the criteria for choosing the basis will become apparent from the discussion later in this chapter.

where θ_j are time-independent basis functions and $a_j(t)$ are the corresponding time coefficients. If the basis functions form an orthonormal set with respect to an inner product defined by $\langle \cdot, \cdot \rangle$, meaning that:

$$\langle \theta_i, \theta_j \rangle = \delta_{ij}, \tag{2.3}$$

where δ_{ij} is the Kronecker delta, a set of ODEs for the time coefficients can be computed using:

$$\dot{a}_j = \langle \theta_j, f(x) \rangle \,. \tag{2.4}$$

Galerkin projection is widely used in numerical computations, for example in finite element methods, where appropriate basis functions (modes) are defined on the domain of computation, and the PDE of interest is projected onto a subset of a finite number of those functions (since a PDE is infinite-dimensional, it may be expanded into an infinite sum of basis functions). If we consider the projection of the infinite-dimensional PDE onto the entire subset of n modes required for an accurate solution as the 'full system', a reduced-order model of order r can be obtained as a set of ODEs for the time evolution of the first r coeffecients by projecting the original system onto the first r modes only. This means including only the first r modes in the expansion in 2.2, where typically $r \ll n$, so that $T = \mathbb{R}^r$. Galerkin projection is applicable to both linear and nonlinear systems, and often the basis is chosen to be POD modes, although other choices can be made, as will be discussed in subsequent sections.

We remark here that a low-dimensional description of the evolution of a system for a given initial condition and forcing can be obtained by projecting stored 'snapshots' of the system state onto modes, which is different from actually integrating the ODEs in (2.4). This is a key observation, since a good basis for the former may not be suitable for the latter. We also remark that computationally, in the context of fluid flows, Galerkin projection can be done in two ways – using the exact 'right-hand side' of the governing equations, which may involve double or triple expansion sums for nonlinear systems, or, for a linear system, using a 'wrapper' around a DNS solver (see Bagheri *et al.*, 2009*b*, for details of this approach). If the time step in the latter approach is small enough, the two methods are equivalent. In this thesis Galerkin projection is computed using the exact right-hand side of the governing equations.

2.1.2 Choice of basis and direction of projection

The choice of basis for Galerkin projection in model reduction of fluid flows is limited, and here we discuss three common choices. The first to be used historically, the method of Proper Orthogonal Decomposition (POD) is the most widely used for obtaining the low-dimensional basis (Sirovich, 1987; Aubry *et al.*, 1988; Holmes *et al.*, 1996; Podvin & Lumley, 1998; Smith, 2003; Smith *et al.*, 2005; Noack *et al.*, 2003). Recently, global eigenmodes of the Navier-Stokes equations linearized about

equilibria have been used to obtain reduced-order models for control (Åkervik *et al.*, 2007; Barbagallo *et al.*, 2008). A thorough review of this approach is given by Henningson & Åkervik (2008). A major disadvantage of POD and global eigenmodes is that there is no straightforward way of capturing accurately actuation and measurements in the models. Finally, snapshot-based balanced truncation, also known as Balanced Proper Orthogonal Decomposition (BPOD), a method that captures the input/output behavior well, introduced to the fluid mechanics community by Rowley (2005), has gained popularity in the recent few years. It has been applied to flow past a plate at an angle of attack (Ahuja & Rowley, 2008, 2009), transitional channel flow (Ilak & Rowley, 2008*b*,*a*, 2006; Rowley & Ilak, 2006) and the Blasius boundary layer (Bagheri *et al.*, 2009*b*).

An important difference between POD and the two latter techniques is that global eigenmodes and balanced truncation involve non-orthogonal Galerkin projection using adjoint modes, also known as Petrov-Galerkin projection. Figure 2.1 illustrates the two ways of projecting the dynamics $f(x)$ from (2.1) evolving in a three-dimensional space onto a two-dimensional subspace T. The projection P_T is orthogonal, while the projection $\tilde{P}_T$ is non-orthogonal as illustrated by the red line, which is not orthogonal to the subspace T. The red line is orthogonal to a different subspace denoted by S. The modes spanning S are usually called *adjoint modes*, and their significance for balanced truncation will be discussed later in this chapter. The two sets are bi-orthogonal:

$$\langle \psi_i, \phi_j \rangle = \delta_{ij}, \tag{2.5}$$

where we denote the modes spanning T as ϕ and the modes spanning S as ψ. It is evident from Fig. 2.1 that the choice of direction of projection can make a lot of difference in capturing the system dynamics even when projecting on the same basis, as will be shown in subsequent chapters.

2.2 Proper Orthogonal Decomposition (POD)

Proper Orthogonal Decomposition, also known as Karhunen-Loève expansion, was first introduced to the fluids community by Lumley (1967, 1970) in order to extract low-dimensional behavior of the flow characterized by its *coherent structures*. More details about POD can be found in standard references (Sirovich, 1987; Holmes *et al.*, 1996); here we only focus on the properties that are relevant for the numerical computations and control design procedure used in the subsequent chapters.

Mathematically, POD modes are the eigenfunctions of the autocorrelation matrix of the states of a system given by (2.1) integrated over time (Holmes *et al.*, 1996). Here we will focus on POD for a system that is discretized in both space and time. The modes can then be computed by stacking simulation snapshots $x(t_k)$ at some

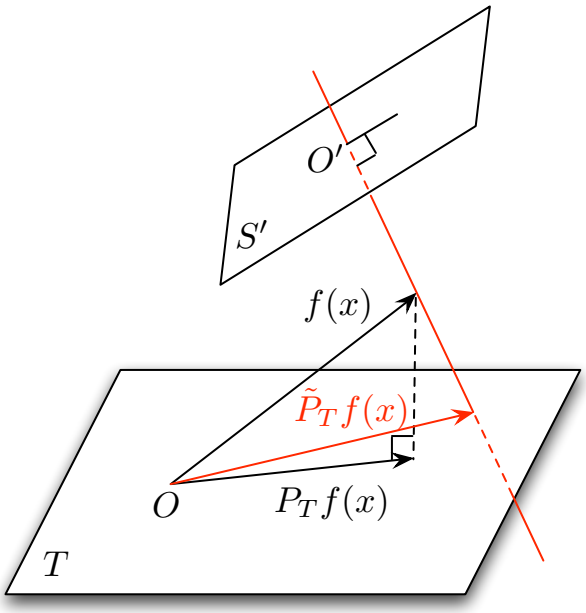


Figure 2.1: An illustration of Galerkin and Petrov-Galerkin projection onto a two-dimensional subspace of a three-dimensional Hilbert space. The direction of the red line is the nullspace of the non-orthogonal projection onto the subspace T. The subspace S, which contains the origin O, and which is orthogonal to the nullspace, is spanned by the adjoint modes. For clarity of the illustration, a two-dimensional space S' with origin O', parallel to S, is shown in the figure instead of S. A translation that maps O' to O would map S' to S. See text for details.

times t_k into a matrix X with appropriate time quadrature weights δ[3]:

$$X = [x(t_1)\sqrt{\delta_1} \;\; x(t_2)\sqrt{\delta_2} \;\; \ldots \;\; x(t_m)\sqrt{\delta_m}], \tag{2.6}$$

where m is the total number of snapshots, and solving the $n \times n$ eigenvalue problem

$$XX^T\Theta = \Theta\Lambda, \tag{2.7}$$

where Λ is the diagonal matrix of eigenvalues. This method is known as the *direct method* for computing POD and it becomes computationally intractable when the number of states n is very large, since the $n \times n$ matrix XX^T becomes too large to store in memory. Sirovich (1987) has shown that the first m POD modes can also be obtained using the solutions of the considerably smaller $m \times m$ eigenvalue problem

$$X^TXU = U\Lambda, \tag{2.8}$$

the orthonormal modes being the columns of the matrix

$$\Theta = XU\Lambda^{-1/2}. \tag{2.9}$$

[3]The time quadrature weights are necessary, since the integral of the autocorrelation matrix XX^T is computed from snapshots at discrete times, i.e., using quadrature. These weights change if the number of snapshots or the spacing in time between two consecutive snapshots changes in order to approximate the continuous integral appropriately (see Appendix B).

This method is numerically tractable for very large systems, since the number of snapshots is typically much smaller than the dimension of the system. It has become known as the *method of snapshots* and it is key for most POD calculations, both from simulations and from experimental data, as well as for other methods, such as empirical balanced truncation, where a low-dimensional basis is computed from simulation results. The method of snapshots was used for all POD calculations in this thesis except for some 1-D cases for which the computational cost is negligible.

POD modes have a particularly intuitive meaning for fluid flows, since they represent the most energetic structures in a given simulation if velocity snapshots are taken — the set of POD modes is the optimal solution to the problem of finding a low-order basis of given dimension that captures the largest fraction of the kinetic energy in the simulation snapshots (Holmes *et al.*, 1996). The energy contained in the modes is measured by the eigenvalues which are the entries of the diagonal matrix Λ, and the sum of the eigenvalues is equal to the total energy in the snapshots. For flows in limited spatial domains and dominated by coherent structures with low-dimensional behavior, the leading few POD modes often capture well over 95% of the flow energy. While POD was initially most often used to analyze the behavior of coherent structures (Lumley, 1967), for example in turbulent channel flow (Moin & Moser, 1989), it was also realized that the basis could be suitable for Galerkin projection in order to obtain a reduced-order model of the system dynamics (Sirovich, 1987; Aubry *et al.*, 1988; Holmes *et al.*, 1996). It has however been shown that the dynamics of the flow is not necessarily captured by leading POD modes (Smith, 2003; Smith *et al.*, 2005; Ilak & Rowley, 2008*b*).

For linear state-space systems, the POD modes of data from the response of the state of the system to an impulsive input (to be discussed below) are the modes which are most sensitive to inputs, also known as the most controllable modes (Rowley, 2005). However, both controllability and observability, which is the potential of states to affect system outputs, are important for the input-output behavior of a system, and POD often fails to capture highly observable modes. On the other hand, balanced truncation does take into account both of these properties, and we next describe this method.

2.3 Balanced truncation

2.3.1 Exact balanced truncation

Balanced truncation is a standard model reduction method (Moore, 1981; Dullerud & Paganini, 1999; Zhou *et al.*, 1996) originally developed for LTI (Linear Time-

Invariant) stable input-output systems of the form

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx, \end{aligned} \tag{2.10}$$

where $u \in \mathcal{U} = \mathbb{R}^p$ is the vector of inputs, $y \in \mathcal{Y} = \mathbb{R}^q$ is the output, $x \in \mathcal{X} = \mathbb{R}^n$ is the state vector (although in general all three spaces can be complex as well), and A, B, and C are matrices of appropriate dimension. The idea of balancing is to find a change of coordinates in which the controllability and observability Gramians, defined by

$$W_c = \int_0^\infty e^{At} B B^+ e^{A^+ t}\, dt, \quad W_o = \int_0^\infty e^{A^+ t} C^+ C e^{At}\, dt, \tag{2.11}$$

are equal and diagonal, their entries being known as Hankel singular values (HSVs). The matrices A^+, B^+ and C^+ in (2.11) define the corresponding adjoint system. It should be noted that in general $A^+ \neq A^T$, the two being equal only when the inner product used to derive the adjoint does not have an associated weight (see Appendix A). It can be shown that balanced truncation does not depend on the choice of the inner product on the state space $\mathcal{X}$, although it does depend on the choices of inner products for $\mathcal{U}$ and $\mathcal{Y}$. This is discussed in more detail in Appendix A.

The system in this new set of coordinates where the Gramians are equal and diagonal can be thought of having a balance between the sensitivity of the states to inputs (controllability) and their potential to affect the system outputs (observability). One then truncates the least controllable and observable modes, corresponding to the smallest eigenvalues of the Gramians. A graphical illustration of the method for a 2-D system is shown in Fig. 2.2. The two ellipsoids are level sets of the quantities $x^T W_c x$ and $x^T W_o x$ and can be thought of as geometric measures of controllability and observability for a given (two-dimensional) state x of the system. For unit magnitudes of the inputs and outputs, the directions for which $x^T W_c x = 1$ and $x^T W_o x = 1$ are the most controllable and most observable directions respectively, and these directions (states) are typically different.

An additional feature of balanced truncation is that some predictions about model performance can be made. In control design we are often interested in the worst-case error between the reduced-order model and the full simulation, which is known as the infinity error norm of the system (see Appendix D for a definition) and for which balanced truncation has apriori error bounds (Dullerud & Paganini, 1999). The H_∞ lower bound for the error in any reduced-order system is

$$\|G - G_r\|_\infty \geq \sigma_{r+1}, \tag{2.12}$$

where $G(s)$ is the transfer function of the full system, $G_r(s)$ is the transfer function of a reduced-order model with state dimension r, and σ_j is the j-th Hankel singular value (in decreasing order). The upper bound for the error for balanced truncation

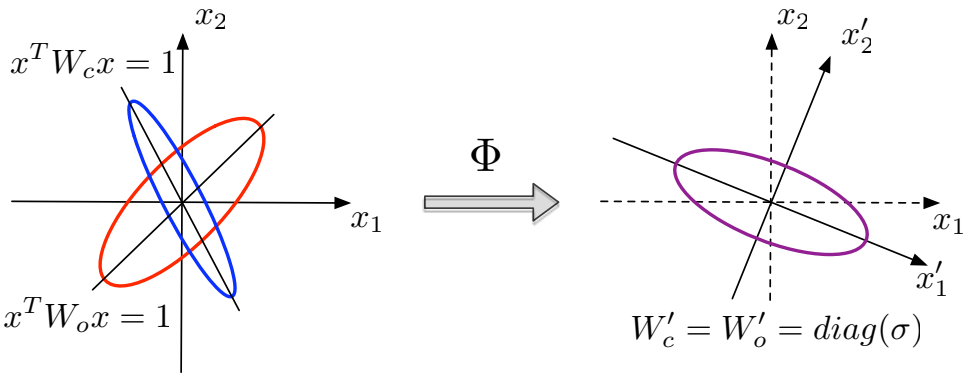


Figure 2.2: A graphical illustration of the balancing transformation. The most controllable states and the most observable directions are not aligned on the left, while in the balanced coordinates they coincide. This implies that, while neither direction can be truncated on the left-hand side, the x_2' direction on the right-hand side has little influence on the input-output behavior of the system.

is given by

$$\|G - G_r\|_\infty \leq 2\Sigma_{j=r+1}^{n} \sigma_j. \tag{2.13}$$

The upper bound on the error can be very close to the lower bound if the HSVs decrease fast, meaning that the error norm of the models is very close to the lowest possible value, as will be shown in the subsequent chapters for models of transitional channel flow. A detailed proof of the error bounds for balanced truncation, along with a discussion of the related Hankel operator and an alternative model reduction approach known as Hankel norm approximation, is given in Zhou *et al.* (1996).

2.3.2 Empirical balanced truncation

In order to compute the balancing transformation, the Gramians are simultaneously diagonalized. This procedure is not computationally tractable for very large systems, as it requires the solution of very large matrix equations, known as Lyapunov equations. A computationally tractable procedure for finding the leading modes of the balancing transformation using the method of snapshots is given in Rowley (2005). Here we outline the key steps. Let the system in (2.10) evolve in a state space $\mathcal{X} = \mathbb{R}^n$. The response of the state of the system in (2.10) to an impulsive input is defined as $x(t) = e^{At}B$ (in contrast to the usual impulse response given by $y(t) = Ce^{At}B$). One begins by computing the snapshots of the impulse-state response of both the original system and the corresponding adjoint system

$$\begin{aligned} \dot{z} &= A^+ z + C^+ v \\ w &= B^+ z, \end{aligned} \tag{2.14}$$

and stacking the direct and adjoint snapshots as columns of matrices X and Y with appropriate quadrature weights (Rowley, 2005):

$$X = [x(t_1)\sqrt{\delta_1} \dots x(t_m)\sqrt{\delta_m}], \qquad Y = [z(t_1)\sqrt{\delta_1} \dots z(t_m)\sqrt{\delta_m}]. \tag{2.15}$$

Although in the above equation it is assumed that the direct and adjoint snapshots are taken at exactly the same times, the number of snapshots and the quadrature weights in X and Y are generally different. One can show that the Gramians in (2.11) may then be approximated by *empirical Gramians* (Lall *et al.*, 2002) $W_{c,e}$ and $W_{o,e}$, as

$$W_c \approx W_{c,e} = XX^+, \qquad W_o \approx W_{o,e} = YY^+. \tag{2.16}$$

The key idea in the method of snapshots is to compute the transformation that balances the empirical Gramians (or at least the dominant directions of this transformation) without actually computing the Gramians themselves, whose dimension is large, resulting in significant computational savings. In this respect, this method resembles the method of snapshots for POD. To compute the balancing transformation, one computes the singular value decomposition (SVD) of the matrix Y^+X (see Appendix A for a discussion of Y^+):

$$Y^+X = U\Sigma V^T, \tag{2.17}$$

where U and V are unitary matrices, and Σ is a diagonal matrix. The balancing transformation Φ and its inverse Ψ are then found by

$$\Phi = XV\Sigma^{-1/2}, \quad \Psi = YU\Sigma^{-1/2}. \tag{2.18}$$

The columns of Φ are the *balancing modes* (alternatively referred to as the direct modes) and the columns of Ψ are the *adjoint modes*, and the two bi-orthogonal sets of modes are used for the Petrov-Galerkin projection. The matrix Y^+X is also known as the Hankel matrix. The entries of the diagonal matrix Σ are the Hankel singular values. As shown by Rowley (2005), the balancing modes are the leading columns of the balancing transformation and the adjoint modes are the leading rows of its inverse. If n snapshots are taken, all rows and columns of the balancing transformation can be computed, although usually numerical roundoff errors arise in computing the modes that correspond to very small HSVs. A balanced reduced-order model is then given by:

$$\begin{aligned} \dot{a} &= \Psi_1^+ A\Phi_1 a + \Psi_1^+ Bu \\ y &= C\Phi_1 a, \end{aligned} \tag{2.19}$$

where the matrices $\Psi_1 : \mathbb{R}^r \to \mathcal{X}$ and $\Phi_1 : \mathbb{R}^r \to \mathcal{X}$, which contain the first r columns of the balancing transformation and its inverse, respectively, are $n \times r$. We will refer to r, the number of states we want to retain in the system, as the rank of the model throughout this thesis.

A different procedure for approximating balancing transformations has also been used in Willcox & Peraire (2002), in which the Gramians are separately reduced (that is, low-rank approximations of $W_{c,e}$ and $W_{o,e}$ are first constructed, and then

the balancing transformation for the rank-reduced Gramians is computed by an unspecified algorithm). However, this procedure is more computationally intensive than BPOD, and has been shown to give worse results (Rowley, 2005), since almost-uncontrollable modes may be strongly observable and should thus not be truncated.

An approach for obtaining empirical balanced models without adjoint simulations, using the Moore-Penrose inverse (pseudoinverse) of the balancing modes instead of adjoint modes has been proposed recently by Or *et al.* (2008). However it was demonstrated by Ma *et al.* (2009*a*) that the resulting models are not a true approximation of balanced truncation.

2.3.3 Output projection and BPOD

If we are interested in the full flow field in a fluid problem, for example to accurately capture the energy growth of a perturbation, the number of outputs can be very large ($n = q$). In this case, the computation of the adjoint simulations of the system given by (2.14) may not be tractable, since one simulation is needed for each component of the output. A way to reduce the number of system outputs is to first project the output onto a low-dimensional subspace, i.e., taking $\tilde{y} = P_s C x$, where P_s is an orthogonal projection onto a s-dimensional subspace of $\mathcal{Y}$, as suggested in Rowley (2005). The system is now of the form:

$$\begin{aligned} \dot{x} &= Ax + Bu \\ \tilde{y} &= P_s C x, \end{aligned} \tag{2.20}$$

where s is the rank of the output projection. The projection P_s that minimizes the 2-norm of the difference between the original transfer function and the output-projected transfer function is given simply by the POD of the set of impulse-state responses (Rowley, 2005). This projection can be written as $P_s = \Theta_s \Theta_s^T$, where columns of $\Theta_s : \mathbb{R}^s \to \mathcal{Y}$ are POD modes. Another way to write the system is as follows:

$$\begin{aligned} \dot{x} &= Ax + Bu \\ \hat{y} &= \Theta_s^T C x. \end{aligned} \tag{2.21}$$

Here, the outputs of the system are just the coefficients of the POD modes of the system impulse response and $\hat{y} \in \mathbb{R}^s$. This s-dimensional output carries the same information as the n-dimensional output $\tilde{y}$, which is easily shown using the fact that for any projection P, $P^2 = P$ (see, for example, Trefethen & Bau, 1997, Lecture 6):

$$\|\tilde{y}\|^2 = x^T C^T P_s^T P_s C x = x^T C^T P_s C x = x^T C^T \Theta_s \Theta_s^T C x = \|\hat{y}\|^2. \tag{2.22}$$

The corresponding adjoint system can now be written as:

$$\begin{aligned} \dot{z} &= A^+ z + (\Theta_s^T C)^+ v \\ w &= B^+ z. \end{aligned} \tag{2.23}$$

Note that if the output is the full state, so that $C = I_n$, and the adjoint is defined with respect to the standard L_2 inner product, the initial conditions of the adjoint simulations are just the POD modes (columns of Θ_s). In practical computations, depending on the choice of inner product used in defining the adjoint system, and on the numerical quadrature method (for example, if the computations are done using Chebyshev polynomials) the matrix $(\Theta_s^T C)^+$ is usually just the matrix Θ_s pre-multiplied by a matrix of inner product weights.

The idea of Balanced POD (BPOD) is to compute the snapshot-based balanced truncation of the system (2.21) instead of (2.10), so that only s adjoint simulations are needed. It is easily shown that the systems (2.20) and (2.21) have the same observability Gramian, again using the fact that for any projection P, we have $P^2 = P$. Transforming (2.21) to balanced coordinates and writing $x = \Phi_1 a$, analogously to (2.19), a reduced-order model is obtained as follows:

$$\begin{aligned} \dot{a} &= \Psi_1^+ A \Phi_1 a + \Psi_1^+ B u \\ y_s &= \Theta_s^T C \Phi_1 a. \end{aligned} \tag{2.24}$$

Note that $r \leq p$, where p is the number of non-zero HSVs. If we assume that $C = I_n$ (this is the case in fluid simulations where we need to know the entire flow field), we can represent the output of (2.24) as $y_s = \Theta_s^T \Phi_1 z$, which is now the vector of time coefficients of the s standard POD modes obtained from the impulse response of the system. For fluid flows the full field output of the model can be recovered from these coefficients and the corresponding modes. For a given dimension of the output projection, all BPOD models will have s outputs regardless of the model rank r, while the number of POD model outputs is equal to r at each rank. The effect of output projection on model performance will be illustrated in Chapters 4 and 5.

Finally, if a system has a small number of measurement outputs (such as when using an estimator), corresponding to a matrix C_1 with a small number of rows, and in addition the main output of interest is the energy of the entire flow field, output projection is used for the energy part and the model reduction is performed as

$$\begin{aligned} \dot{x} &= Ax + Bu & & \dot{a} = \Psi_1^+ A \Phi_1 a + \Psi_1^+ B u \\ y &= \begin{bmatrix} C_1 \\ I_n \end{bmatrix} x & \Rightarrow & \quad y_s = \begin{bmatrix} C_1 \\ \Theta_s^T \end{bmatrix} \Phi_1 a. \end{aligned} \tag{2.25}$$

A summary of BPOD algorithm is given below. We remark here that, although the acronym BPOD was introduced by Rowley (2005) in the context of systems with a large number of outputs where output projection is necessary, it has also been used for systems with a small number of outputs where this step is not necessary.

1. Simulate the system (2.10) and form the matrix of snapshots X as given by (2.15).

2. If the number of outputs is very large, such as in the case where $C = I_n$, compute the POD modes of the set of snapshot outputs. If $C = I_n$, the matrix X from the previous step is used, if $C \neq I_n$, the matrix of snapshots for POD will be CX.

3. Simulate the adjoint system (2.14), with the columns of the C^+ matrix as initial conditions, or, if the number of outputs is large, with the POD modes computed in the previous step as initial condition and form the matrix of snapshots Y as given by (2.15).

4. Compute the SVD of the matrix $Y^+X = U\Sigma V^T$ and the direct and adjoint modes:
$$\Phi = XV\Sigma^{-1/2}, \quad \Psi = YU\Sigma^{-1/2} \tag{2.26}$$

5. The system (2.10) in balanced coordinates is then given by (2.19) or alternatively by (2.24) or (2.25) if the system has a very large number of outputs. The corresponding reduced-order model of rank r is obtained by using the first r columns of Φ and Ψ. If the initial condition of the original system is given by the vector x_0, the initial condition in the reduced-order model is its projection $\Psi_r^+ x_0$.

Extensions of balanced truncation, both exact and snapshot-based, have been introduced for time-periodic systems (Ma *et al.*, 2009*b*) and unstable systems (Zhou *et al.* (1999) for exact method and Ahuja & Rowley (2009) for snapshot-based method). In this thesis we only consider balancing of stable LTI systems and snapshot-based balancing of nonlinear systems, which will be introduced later in this chapter.

2.3.4 Eigensystem Realization Algorithm

It has recently been observed that a well-known method from system identification, the Eigensystem Realization Algorithm (ERA) (Juang & Pappa, 1985; Juang & Phan, 2001) generates models theoretically identical to BPOD models (Ma *et al.*, 2009*a*). In particular, it is shown that the Hankel matrix Y^+X can be formed only using a set of measurements from an experiment, or output signals extracted from a simulation of the full field. The advantage of this method is that adjoint simulations are not required, resulting in tremendous computational savings in terms of CPU time and memory requirements, as well as avoiding the challenges that can sometimes arise in derivation and implementation of the adjoint equations for a given problem. Also, reduced-order models can be obtained from experimental measurements using ERA. A disadvantage, however, is that the method only computes models, and does not return the direct and adjoint modes, which are useful for investigating parameter variations in the models and for obtaining non linear models.

ERA was used for the results presented in Chapter 5, as it allows a fast computation of models for different choices of measurement, thereby making a detailed comparison of different output variables and locations feasible.

2.3.5 Balanced truncation of nonlinear systems

Balanced truncation has been shown to be an excellent method for model reduction of linear systems, including linearized Navier-Stokes equations, which will be discussed in the next chapter. Unfortunately, the evolution of linearized perturbations is typically restricted to a small region of validity in the neighborhood of the corresponding equilibrium state of the nonlinear system, and the nonlinear dynamics is omitted altogether in linear models. It is therefore desirable to obtain nonlinear models of the flows of interest. POD has traditionally been used for this purpose, since the application of the snapshot method is identical for linear and nonlinear systems and models can be obtained in a straightforward fashion, but, as mentioned earlier, the method has been shown in many cases to fail to capture dynamics accurately. On the other hand, balanced truncation in its precise formulation described in Section 2.3.1 is only defined for linear systems, and there is no procedure for nonlinear systems that is its direct equivalent. Still, the appeal of the theory of linear balancing and the successful application of the method to many problems has spurred extensive research in the area of nonlinear balancing.

Although extensive theory has been developed from 'first principles', i.e., using the extensions of the definitions of controllability and observability to nonlinear systems, the size of systems that most methods for nonlinear balancing can be applied to is modest at best. The original procedure proposed by Scherpen (1993) involves the solution of Hamilton-Jabobi PDEs in order to obtain the controllability and observability functions for a nonlinear system, which is computationally not feasible for large systems. Other methods, which involve a degree of complexity, such as sliding interval balancing (Verriest & Gray, 2004) or Taylor series expansion (Fujimoto & Tsubakino, 2008) have also been applied to small systems. On the other hand, the empirical snapshot-based method proposed by Lall *et al.* (2002) is more straightforward to apply.

The essence of the method in Lall *et al.* (2002) is that the space of the possible inputs is sampled over a number of impulse response magnitudes in order to compute an empirical controllability Gramian, and the state space is sampled with a number of amplitudes of the basis vectors spanning the space in order to compute an empirical observability Gramian. For nonlinear systems, the adjoint system is not defined as precisely as for linear systems, although nonlinear adjoint operators derived using Hamiltonian extensions of the original nonlinear system have been proposed (Fujimoto *et al.*, 2002). Lall *et al.* (2002) propose simulating the system with all the basis vectors of the state space as initial conditions for the

empirical observability Gramian, since in the linear case this empirical Gramian is equal to the exact snapshot-based observability Gramian. This makes the method computationally expensive for large systems, since the number of the required simulations is of the order of the number of the states. For the discussion that follows we assume a general form of a nonlinear input-output system:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t)). \end{aligned} \tag{2.27}$$

The following definitions are based on the ones given in Lall *et al.* (2002), with the notation adepted to match the conventions used in this thesis. We start with the empirical controllability Gramian:

$$\hat{W}_c = \sum_{l=1}^{r} \sum_{m=1}^{s} \sum_{i=1}^{p} \frac{1}{rsc_m^2} \int_0^\infty \mathcal{F}^{ilm}(t) dt, \tag{2.28}$$

where the matrix $\mathcal{F}^{lm}(t) \in \mathbb{R}^{n \times n}$ is given by:

$$\mathcal{F}^{ilm}(t) := (x^{ilm}(t) - \bar{x}^{ilm})^*(x^{ilm}(t) - \bar{x}^{ilm}), \tag{2.29}$$

where $x^{ilm}(t)$ the state response corresponding to the impulsive input $u(t) = c_m T_l e_i \delta(t)$, where the matrix T_l is a $p \times p$ rotation matrix in the input space, and the vectors e_i are the basis vectors of the input space. The overbar indicates a time average.

The empirical nonlinear observability Gramian, as defined by Lall *et al.* (2002), is given by:

$$\hat{W}_o = \sum_{l=1}^{r} \sum_{m=1}^{s} \frac{1}{rsc_m^2} \int_0^\infty T_l \mathcal{G}^{lm}(t) T_l^* dt, \tag{2.30}$$

where the entries of the matrix $\mathcal{G}^{lm}(t) \in \mathbb{R}^{n \times n}$ are

$$\mathcal{G}_{ij}^{lm}(t) := (y^{ilm}(t) - \bar{y}^{ilm})^*(y^{jlm}(t) - \bar{y}^{jlm}) \tag{2.31}$$

and $y^{ilm}(t)$ is the output of the original system corresponding to the initial condition $x_0 = c_m T_l e_i$ with zero input. Here e_i are the basis vectors of the state space. We assume that the input is related to the state through the linear input matrix B.

Figure 2.3 illustrates the method of Lall *et al.* (2002) graphically for a sample system with two actuators and a state of dimension three. In Figure 2.3a, the set of rotation matrices $\mathcal{T}$, with dimension 2×2 in this example, can be thought as determining the direction of the unit basis vectors of the input space e_1 and e_2, and the amplitude c gives their magnitude. Similarly, in Figure 2.3b, the set $\mathcal{T}$ of 3×3 rotation matrices can be thought of as determining the direction of each of the states used as initial conditions, with c being their amplitudes.

An open question is the selection of the range of initial condition amplitudes c_m for both the controllability and observability Gramians. In fluid problems, typically the linearized problem governs the evolution of a small disturbance about a base flow. Beyond a certain magnitude of the perturbation, the nonlinear effects

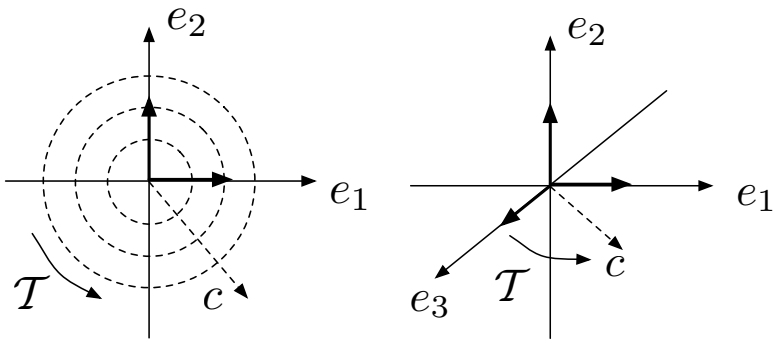


Figure 2.3: A graphical illustration of the method of Lall *et al.* (2002). (a) Sampling of an example two-dimensional input space for the empirical controllability Gramian, where c is the magnitude of an impulse, and $\mathcal{T}$ is a rotation matrix in the coordinate system e_1, e_2. (b) Sampling of an example three-dimensional state space for the empirical observability Gramian. Again $\mathcal{T}$ is a rotation matrix, and c is the corresponding amplitude.

dominate. If for example it is desired to study weakly nonlinear effects in transitional flow, only small amplitudes should be used, while if fully turbulent flow is considered, a range of large perturbation amplitudes should be selected.

Once the empirical Gramians are obtained, they are diagonalized simultaneously, and Galerkin projection is performed as described in Section 2.1.1.

2.4 Summary

This chapter has introduced the standard approach for obtaining reduced-order models of a large-scale dynamical system using Galerkin or Petrov-Galerkin projection onto an appropriate low-dimensional basis. The focus is on the method of empirical balanced truncation, in which reduced-order models that capture well the input-output behavior of the full system are computed using simulation snapshots. Thus, solving matrix equations of dimension too large for most modern computers is avoided. In this thesis, reduced-order models are obtained using POD, exact balanced truncation, snapshot-based balanced truncation with output projection (BPOD) and Eigensystem Realization Algorithm (ERA), and the corresponding abbreviations will be used in later chapters. In addition, nonlinear models will be obtained using the method of Lall *et al.* (2002) and projection of the nonlinear system onto balancing modes computed from balanced truncation of a linear system.

Chapter 3

Numerical simulation and model reduction of channel flow

The main effort in this thesis is the application of the model reduction methods introduced in the previous chapter to modeling and feedback control of transitional channel flow. Besides being a proof-of-concept study for application to more complex geometries found in practical flow situations, channel flow is itself a problem with complex dynamics, and it has become a canonical case for studying transitional dynamics. In particular, we are interested in delaying or preventing transition to turbulence by controlling small perturbations to the stable laminar state. The dynamics of these perturbations can be described by the linearized Navier-Stokes equations and is characterized by large transient growth, thought to be the key mechanism behind 'bypass transition' in shear flows (Schmid & Henningson, 2001), as described in Section 1.2. Since the required computations are closely related to the underlying theory, this chapter summarizes the most relevant topics from both theory and computations.

Plane channel flow belongs to the class of shear flows, for which the velocity profile is determined by the shear stress that the fluid experiences due to the presence of a wall. Fully developed channel flow is a parallel shear flow, i.e., the mean velocity profile is independent of the streamwise direction. Due to its simple geometry yet fairly complex dynamics, channel flow has become a canonical problem for numerical investigation of both fully turbulent and transitional flow (Moin & Kim, 1980; Kim *et al.*, 1987; Schmid & Henningson, 1992; Henningson *et al.*, 1993), and recently for development of flow control strategies (Choi *et al.*, 1994; Lee *et al.*, 2001; Högberg *et al.*, 2003; Ilak & Rowley, 2008*b*). Of particular significance to this work is the study of Kim *et al.* (1987), in which an important benchmark Direct Numerical Simulation (DNS) was performed and the flow statistics were carefully studied.

A DNS code based on the method described in Kim *et al.* (1987) is used both to obtain the simulation snapshots for the computation of the reduced-order models

and for their testing in the nonlinear case. Since adjoint simulations are required in order to obtain the snapshots for the computation of the BPOD basis, appropriate adjoint equations are derived for the linearized case. The adjoint equations can be defined using different inner products, and we choose an inner product that is convenient for the numerical simulations.

Although it is technically a postprocessing stage, the computation of the reduced-order models is tightly coupled to the direct numerical simulation, since many of the same numerical methods are used. Therefore we also describe some of the challenges in model computation and some approaches to addressing those challenges, in particular modular software design.

3.1 Plane channel flow

In this section we introduce channel flow and the appropriate linearization used to obtain reduced-order models. A detailed treatment of shear flows and channel flow in particular is given in standard textbooks, for example Pope (2000) or White (1991). More details about the linearized Navier-Stokes equations for shear flows can be found in Schmid & Henningson (2001).

3.1.1 Governing equations and dynamics

Channel flow is the flow between two infinite parallel flat plates, as illustrated by Fig. 3.1. In order to derive the governing equations for channel flow, we start with the incompressible Navier-Stokes equations:

$$\frac{\partial \tilde{u}_i}{\partial \tilde{t}} + \tilde{u}_j \frac{\partial \tilde{u}_i}{\partial \tilde{x}_j} = -\frac{1}{\tilde{\rho}} \frac{\partial \tilde{p}}{\partial \tilde{x}_i} + \tilde{\nu} \frac{\partial^2 \tilde{u}_i}{\partial \tilde{x}_j \tilde{x}_j} \tag{3.1}$$

$$\tilde{\nabla} \cdot \tilde{\mathbf{u}} = 0, \tag{3.2}$$

with no-slip boundary conditions (i.e., all velocity components are zero at the walls). All quantities with a tilde denote dimensional quantities, $\tilde{\rho}$ is the fluid density and $\tilde{\nu}$ is the kinematic viscosity of the fluid. Here $x_{j=1,3}$ corresponds to the x, y, z axes and $u_{j=1,3}$ corresponds to the u, v, w (streamwise, wall-normal and spanwise) velocity components in Fig. 3.1. We apply a non-dimensionalization given by:

$$u = \frac{\tilde{u}}{U_c}, \quad x = \frac{\tilde{x}}{\delta}, \quad t = \tilde{t}\frac{U_c}{\delta}, \quad \rho = \frac{\tilde{\rho}}{\rho_0}, \quad p = \frac{\tilde{p} - p_0}{\rho_0 U_c^2}. \tag{3.3}$$

Here U_c is the centerline velocity of the laminar flow with the same average mass flux as the turbulent flow, δ is one-half of the channel thickness, ρ_0 is a characteristic density, and p_0 is a reference pressure. For incompressible flow, we can set $\tilde{\rho} = \rho_0$ since the density is constant. Equation (3.1) may then be re-written in a non-

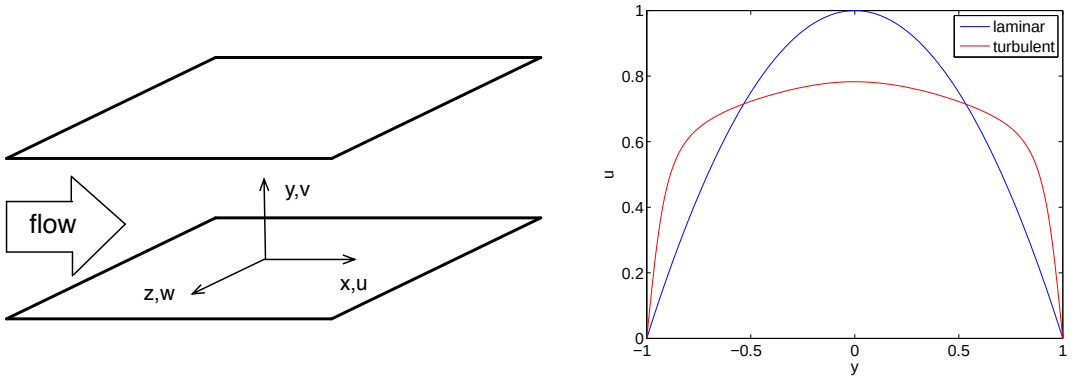


Figure 3.1: Left: Channel flow geometry. The fields $u(x, y, z)$, $v(x, y, z)$ and $w(x, y, z)$ are the streamwise, wall-normal and spanwise components of the flow velocity. Right: A comparison of the laminar parabolic profile $u = 1 - y^2$ and an average profile calculated from a converged turbulent simulation with the same mass flux.

dimensional form as:

$$\frac{\partial u_i}{\partial t} + \tilde{u_j}\frac{\partial \tilde{u}_i}{\partial \tilde{x}_j} = -\frac{\partial p}{\partial x_i} + \frac{1}{Re_c}\frac{\partial^2 \tilde{u}_i}{\partial \tilde{x}_j \tilde{x}_j}, \tag{3.4}$$

where the centerline Reynolds number is defined as:

$$Re_c = \frac{U_c \delta}{\nu}. \tag{3.5}$$

Physically, the shear forces at the walls cause the flow to lose energy. It can be shown that, both for fully developed turbulent flow, and laminar flow, the left-hand side of (3.4) is zero (see Appendix B). Therefore, in order to maintain fully developed flow, a favorable streamwise pressure gradient that exactly balances the shear forces is required. This pressure gradient can either be kept constant or dynamically adjusted in order to enforce a constant mass flux. For simulations, the latter approach results in a faster convergence to a statistically steady state (see Appendix B).

If the channel walls are chosen to be at $y = \pm 1$, for the non-dimensionalization chosen here it can be shown (see Appendix B) that the laminar profile is given by:

$$U(y) = -\frac{dp}{dx}\frac{Re_c}{2}(1 - y^2), \tag{3.6}$$

while a mean turbulent profile (see Fig. 3.1) is computed from averaging snapshots of the velocity field in streamwise and spanwise directions in space, as well as in time (*ensemble average*), also known as Reynolds averaging. The Reynolds number commonly used to characterize the fully turbulent simulation is the frictional Reynolds number, defined as:

$$Re_\tau = \frac{\rho u_\tau D}{\mu},$$

where $\mu = \nu\rho$ is the dynamic viscosity and u_τ is a *frictional velocity*, defined as:

$$u_\tau = \sqrt{\frac{\tau_w}{\rho}},$$

with the wall shear defined as $\tau_w = \mu \frac{\partial u}{\partial y}|_{wall}$ and computed from an ensemble average profile of a converged simulation. An alternative non-dimensionalization of the Navier-Stokes equations using u_τ as the characteristic velocity is often used in studying turbulent channel flow.

The key parameter in the simulations in this work is the centerline Reynolds number Re_c, and from now on it will be assumed that $Re = Re_c$ unless specified otherwise. This is the appropriate Reynolds number to characterize the linearized flow regime (Schmid & Henningson, 1992; Högberg *et al.*, 2003), and since our focus will be on perturbations which are small enough for the flow to remain in the neighborhood of the linear regime, we will use Re_c. On the other hand, Re_τ is usually the characterizing parameter of the turbulent regime.

3.1.2 Linearized equations

We next describe how the linearized evolution of a perturbation to the laminar profile can be described by a state-space system in the form described in Chapter 2, thereby fitting into a suitable modeling and control framework. For shear flows, the linearized equations may be conveniently written in terms of the wall-normal velocity v and the wall-normal vorticity η (see, for instance, Schmid & Henningson, 2001). We recall that the vorticity field is defined as:

$$\omega = \nabla \times \mathbf{u}, \tag{3.7}$$

and $\eta = \omega_2$. The other flow variables (e.g., streamwise and spanwise velocities u and w) may then be computed using the continuity equation $\partial_x u + \partial_y v + \partial_z w = 0$ and the definition of wall-normal vorticity. In these coordinates, the linearized (non-dimensional) equations have the form:

$$\begin{aligned} \left[(\partial_t + U\partial_x)\Delta - U''\partial_x - \frac{1}{Re}\Delta^2\right] v &= 0 \\ \left[\partial_t + U\partial_x - \frac{1}{Re}\Delta\right] \eta &= -U'\frac{\partial v}{\partial z}, \end{aligned} \tag{3.8}$$

where $U(y)$ is the laminar profile, $\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$ is the Laplacian and the prime indicates differentiation with respect to y. The first equation is the Orr-Sommerfeld equation and the second one is known as the Squire equation. It was first shown numerically by Orszag (1971) that the Orr-Sommerfeld equation for channel flow is stable up to $Re \approx 5772$, when an exponentially unstable eigenmode first arises. The Squire equation has stable eigenmodes for all values of Re. Still, complex behavior due to the non-normality exists for stable eigenmodes. The term

on the right hand side of the Squire equation represents tilting of the spanwise component of the mean flow vorticity (which here is just U') by the strain rate $\partial v/\partial z$, giving rise to wall-normal vorticity (Butler & Farrell, 1992). In the limit of high Reynolds number, the perturbation growth is dominated by this process, in particular for streamwise-constant perturbations. While the system also exhibits phenomena such as degeneracies and resonances (Gustavsson, 1986; Henningson & Schmid, 1992), non-normality has been shown to have a dominating effect on the energy growth (Reddy & Henningson, 1993).

In operator form, we can represent the equations using more compact notation as follows:

$$\frac{\partial}{\partial t}\begin{bmatrix}-\Delta & 0\\ 0 & I\end{bmatrix}\begin{bmatrix}v\\ \eta\end{bmatrix} = \begin{bmatrix}L_{OS} & 0\\ -U'\partial_z & L_{SQ}\end{bmatrix}\begin{bmatrix}v\\ \eta\end{bmatrix}, \tag{3.9}$$

where

$$L_{OS} = U\partial_x\Delta - U''\partial_x - \frac{1}{Re}\Delta^2$$
$$L_{SQ} = -U\partial_x + \frac{1}{Re}\Delta$$

are the Orr-Sommerfeld and Squire operators, respectively. If we define the matrix operator A as:

$$A = \begin{bmatrix}-\Delta & 0\\ 0 & I\end{bmatrix}^{-1}\begin{bmatrix}L_{OS} & 0\\ -U'\partial_z & L_{SQ}\end{bmatrix}, \tag{3.10}$$

with no-slip boundary conditions, we can write the system in standard state-space form:

$$\begin{aligned}\dot{x} &= Ax + B_1u_1 + B_2u_2\\ y &= Cx,\end{aligned} \tag{3.11}$$

where $x = [v;\eta]^T$, B_1 and B_2 represent the spatial (time-independent) distributions of the perturbations and actuators respectively, $u_1(t)$ and $u_2(t)$ are the corresponding input vectors (the time-dependent amplitudes of the columns of B_1 and B_2), y is the vector of system outputs, related to the state by the matrix C. The actuation and the disturbances are equivalent mathematically as they are both inputs to the system.

It is important to note that, in order to obtain the POD basis needed for output projection in BPOD, we simulate the system given by (3.9) with a given perturbation or actuator as initial condition until the response has decayed to negligible levels. The matrix XX^+ that can then be formed from the snapshots will closely approximate the controllability Gramian given by (2.11), where the integral extends to infinite time. Likewise, to obtain the Y matrix of adjoint snapshots from (2.15), we simulate the adjoint system until the response has decayed to negligible levels.

Most investigations of linearized channel flow consider each streamwise and spanwise spatial wavenumber pair separately after Fourier transforming the equation in streamwise and spanwise directions. In this formulation, the three-dimensional velocity field is represented as:

$$q(x,y,z,t) = \text{Real}\{\hat{q}(y,t)e^{i(\alpha x+\beta z)}\}, \tag{3.12}$$

where $\hat{q}(y,t) = [\hat{v}(y,t)\ \hat{\eta}(y,t)]^T$, and α and β are the streamwise and spanwise wavenumbers, respectively. The equations for the evolution of $\hat{q}$ are:

$$\begin{aligned}\left[(\partial_t + i\alpha U)(\mathcal{D}^2-k^2)^2 - i\alpha\mathcal{D}^2U - \frac{1}{Re}(\mathcal{D}^2-k^2)^2\right]\hat{v} &= 0\\ \left[\partial_t + i\alpha U - \frac{1}{Re}(\mathcal{D}^2-k^2)^2\right]\hat{\eta} &= -i\beta\mathcal{D}U\hat{v},\end{aligned} \tag{3.13}$$

where $k^2 = \alpha^2+\beta^2$, $\mathcal{D}$ denotes differentiation in the y-direction, and, with a slight abuse of notation, U is the 1-D laminar profile (Schmid & Henningson, 2001). While this formulation has been very useful in studying the linearized flow dynamics, the focus of this thesis is on the general form of the equations, as given by (3.8). The single-wavenumber approach is used only in Chapter 4 for a thorough validation of the BPOD procedure on a 1-D problem resulting from this formulation, for which exact balanced truncation can easily be computed for purposes of comparison. We note here that for all 1-D perturbations considered in this thesis $k^2 \neq 0$, and all 3-D perturbations have no $k^2 = 0$ component.

3.1.3 Derivation of the adjoint equations

To determine the corresponding adjoint equations, one first needs to define an inner product on the vector space $\mathcal{X}$ of flow variables (v,η). Since balanced truncation is independent of the choice of inner product used to define the adjoint (see Appendix A), we may choose an inner product that is convenient for numerical computations. Let us define the inner product:

$$\langle(v_1,\eta_1),(v_2,\eta_2)\rangle_M = \frac{1}{V}\int_\Omega(-v_1\Delta v_2 + \eta_1\eta_2)\,dx\,dy\,dz, \tag{3.14}$$

where V denotes the fluid volume and Ω denotes the computational box. Note that, letting $M : \mathcal{X}\to\mathcal{X}$ denote the matrix operator on the left hand side of Eq. (3.9), this is just the L_2 inner product of (v_1,η_1) with $M(v_2,\eta_2)$.

The inner product in (3.14) is related to the energy inner product commonly used in analyzing single-wavenumber perturbations after Fourier decomposition (Gustavsson, 1986; Butler & Farrell, 1992), given by:

$$\langle(\hat{v}_1,\hat{\eta}_1),(\hat{v}_2,\hat{\eta}_2)\rangle_E = \frac{1}{8k^2}\int_{y=-1}^{y=1}(-\hat{v}_1(\mathcal{D}^2-k^2)\hat{v}_2 + \hat{\eta}_1\hat{\eta}_2)\,dy. \tag{3.15}$$

Using continuity, it can be shown that at a particular wavenumber pair (α, β) the following relation holds for $k \neq 0$:

$$\begin{pmatrix} \hat{u} \\ \hat{v} \\ \hat{w} \end{pmatrix} = \begin{bmatrix} i\alpha\mathcal{D}/k^2 & -i\beta/k^2 \\ 1 & 0 \\ i\beta\mathcal{D}/k^2 & i\alpha/k^2 \end{bmatrix} \begin{pmatrix} \hat{v} \\ \hat{\eta} \end{pmatrix}, \tag{3.16}$$

where $\hat{u}(y)$ and $\hat{w}(y)$ are the streamwise and spanwise velocity components at the particular wavenumber, respectively (Hœpffner, 2006). Using this relation, it can be shown that the norm corresponding to the inner product in (3.15) is exactly the kinetic energy of the perturbation. The inner product in (3.15) is different from the one we have introduced above, as there is no re-scaling at each wavenumber in (3.14). In the case of a single-wavenumber perturbation, the two inner products differ only by a constant factor, but for a general 3-D field with contributions at all wavenumbers this is not the case.

With the definition of the inner product from (3.14), the adjoint equations corresponding to (3.9) are easily found by integration by parts:

$$\frac{\partial}{\partial t} \begin{bmatrix} -\Delta & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} v \\ \eta \end{bmatrix} = \begin{bmatrix} L^*_{OS} & U'\partial_z \\ 0 & L^*_{SQ} \end{bmatrix} \begin{bmatrix} v \\ \eta \end{bmatrix} \tag{3.17}$$

where

$$L^*_{OS} = -U\partial_x\Delta - 2U'\partial_x\partial_y - \frac{1}{Re}\Delta^2$$
$$L^*_{SQ} = U\partial_x + \frac{1}{Re}\Delta.$$

The adjoint operator thus becomes:

$$A^+ = \begin{bmatrix} -\Delta & 0 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} L^*_{OS} & U'\partial_z \\ 0 & L^*_{SQ} \end{bmatrix}. \tag{3.18}$$

3.1.4 Inner product on the output space

Although the time evolution of the linearized disturbances is fully described by the wall-normal velocity-vorticity formulation, the output of the system may be chosen to be in different variables. When using POD, the choice of inner product can have a large impact on the results. For our system, since the other two velocity components (u, w) are easily be recovered using continuity and the definition of vorticity, we can choose the full velocity field to be the output, and use the kinetic energy inner product given by

$$\langle \mathbf{u_1}, \mathbf{u_2} \rangle = \frac{1}{V} \int_\Omega (u_1 u_2 + v_1 v_2 + w_1 w_2)\, dx\, dy\, dz. \tag{3.19}$$

This choice is intuitively appealing, since the POD modes for the output projection will capture the true kinetic energy of the perturbation. We therefore define the

output space $\mathcal{Y}$ in our system as the space $\mathbb{R}^n$ (recall than n is the number of grid points in the simulation), together with the inner product defined by (3.19). We note here that the space $\mathcal{X}$ is also $\mathbb{R}^n$, though endowed with a different inner product (the M-inner product described in the previous section).

3.2 Computation of models

The computation of reduced-order models of linearized channel flow is carried out using snapshots obtained from linearized DNS of the systems (3.9) and (3.17). Nonlinear DNS, which typically require very high resolution, is used in this thesis only to test the performance of the linear controllers in delaying or preventing transition to turbulence. While the computation of modes and models using the algorithms described in Chapter 2 is straightforward, some computational practices resulting in more efficient calculations are discussed in this section, along with guidelines for verification of the model computations.

3.2.1 Computation of balancing and adjoint modes

Although the grid resolution required for the linear simulations of (3.9) and (3.17) is usually considerably lower than for turbulent runs, the computation of modes and models may still require substantial resources, both in terms of CPU time and memory. If the entire snapshot matrices X and Y in (2.26) are loaded in memory when computing the direct and adjoint modes in BPOD, and the linearized DNS has a high resolution (for example, $\mathcal{O}(10^5)$ grid points) and a large number of snapshots ($\mathcal{O}(10^3)$), the available memory may be insufficient. The amount of required memory is easily reduced significantly, since a computation of modes using (2.26) does not require loading of the entire snapshot matrices X and Y and snapshots can be read in one by one or in smaller sets. This saving in memory may be achieved at a cost in computation time, since typically loading small numbers of data files may take more time than loading an entire set of snapshots at once.

Next, the largest part of the computational cost in computing BPOD models in terms of time is the calculation of inner products for the Hankel matrix, particularly if the inner product is different from a standard weighted inner product. This is the case for the inner product described in Section 3.1.3, due to the computation of the Laplacian, which involves spectral derivatives. It turns out that a significant decrease in the time required to compute the inner products can be achieved by exploiting the structure of the Hankel matrix Y^+X. While in this thesis we treat the linearized systems (3.9) and (3.17) as continuous-time systems, in this section we take a discrete-time approach for illustration purposes.

To describe the algorithm, we begin with the case where all snapshots are evenly spaced in time. Let us define the discrete-time matrices:

$$\tilde{A} = e^{A\delta t}, \quad \tilde{A}^+ = e^{A^+\delta t}, \tag{3.20}$$

where A and A^+ are defined in Eqs. (3.10) and (3.18), respectively, and δt is a fixed time step[1]. The evolution of the states of the direct and adjoint systems from time t_k to t_{k+1} is then:

$$x_{k+1} = \tilde{A}x_k \tag{3.21}$$
$$z_{k+1} = \tilde{A}^+ z_k, \tag{3.22}$$

where we have used the compact notation $x(t_k) = x_k$. Furthermore, the evolution of the system from t_k to t_{k+n} for any integer n is given by:

$$x_{k+n} = \tilde{A}^n x_k \tag{3.23}$$
$$z_{k+n} = (\tilde{A}^+)^n z_k, \tag{3.24}$$

Therefore, if we have m snapshots of a simulation of the direct system (3.9) and p snapshots of a simulation of the adjoint system (3.17), the snapshot matrices defined by (2.15) become:

$$X = \left[x_0, \tilde{A}x_0, \tilde{A}^2 x_0 \ldots \quad , \tilde{A}^{m-1}x_0\right] \tag{3.25}$$
$$Y = \left[z_0, \tilde{A}^+ z_0, (\tilde{A}^+)^2 z_0 \ldots \quad , (\tilde{A}^+)^{p-1} z_0\right]. \tag{3.26}$$

We note that the difference $\delta t = t_{k+1} - t_k$ is typically larger than the time step of the DNS simulation, and may often also be larger than the gap in time between two consecutive snapshots written to disk (this gap is usually at least several time steps of the DNS simulation, depending on simulation parameters and the dynamics of the simulated flow, see Appendix C for a discussion of snapshot spacing). The Hankel matrix Y^+X of inner products may then be written as:

$$Y^+X = \begin{bmatrix} \langle y_0, x_0\rangle & \langle y_0, x_1\rangle & \ldots & \langle y_0, x_{m-1}\rangle \\ \langle y_1, x_0\rangle & \langle y_1, x_1\rangle & \ldots & \langle y_1, x_{m-1}\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle y_{p-2}, x_0\rangle & \langle y_{p-2}, x_1\rangle & \ldots & \langle y_{p-2}, x_{m-1}\rangle \\ \langle y_{p-1}, x_0\rangle & \langle y_{p-1}, x_1\rangle & \ldots & \langle y_{p-1}, x_{m-1}\rangle \end{bmatrix}, \tag{3.27}$$

or equivalently:

[1] See (Stengel, 1994, Chapter 2) for a detailed discussion of the relationship between a continuous-time system and the corresponding discrete-time system.

$$Y^+X = \begin{bmatrix} \langle y_0, x_0 \rangle & \left\langle y_0, \tilde{A}x_0 \right\rangle & \dots & \left\langle y_0, \tilde{A}^{m-1}x_0 \right\rangle \\ \left\langle \tilde{A}^+ y_0, x_0 \right\rangle & \left\langle \tilde{A}^+ y_0, \tilde{A}x_0 \right\rangle & \dots & \left\langle \tilde{A}^+ y_0, \tilde{A}^{m-1}x_0 \right\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \left\langle (\tilde{A}^+)^{p-2} y_0, x_0 \right\rangle & \left\langle (\tilde{A}^+)^{p-2} y_0, \tilde{A}x_0 \right\rangle & \dots & \left\langle (\tilde{A}^+)^{p-2} y_0, \tilde{A}^{m-1}x_0 \right\rangle \\ \left\langle (\tilde{A}^+)^{p-1} y_0, x_0 \right\rangle & \left\langle (\tilde{A}^+)^{p-1} y_0, \tilde{A}x_0 \right\rangle & \dots & \left\langle (\tilde{A}^+)^{p-1} y_0, \tilde{A}^{m-1}x_0 \right\rangle \end{bmatrix}. \tag{3.28}$$

From the property of the adjoint:

$$\left\langle \tilde{A}x_1, x_2 \right\rangle = \left\langle x_1, \tilde{A}^+ x_2 \right\rangle, \tag{3.29}$$

we see that there are only $p+m-1$ unique elements of the matrix, and therefore we only need to compute $p+m-1$ inner products instead of $p \times m$ inner products. The number of the inner products to be computed can be orders of magnitude smaller for a large number of snapshots. For example, for a thousand direct snapshots and thousand adjoint snapshots, 1999 inner products need to be computed instead of a million, resulting in computational savings of a factor of 500 in CPU time (depending on how the reading in of data is managed, these savings may be lower, although they are still at least one order of magnitude). When using ERA, the computational savings are even greater, as there are no inner products involved, however the present method is useful when it is desirable to have the bases of balancing and adjoint modes in addition to reduced-order models.

We note that sometimes the dynamics of an impulse response is such that the snapshot spacing in time can be varied in order to save memory, for example more snapshots may be taken during an initial transient with fast dynamics, while fewer snapshots may be needed for a later stage of the simulation. In this case, more inner products would need to be computed than in the above case, although still much fewer than $p \times m$, since snapshots for intermediate steps between t_k and t_{k+1} need to be used (recall that δt is often larger than the time between two consecutive snapshots).

3.2.2 Verification of results

Since many steps are involved in the computation of models, errors can be introduced into the computations and propagated through subsequent steps. While one usually cannot verify the computation of new models against known results, as many sources of error as possible should be eliminated from the computations. Basic checks that were performed for the computations in this thesis are outlined here. It is essential that these checks be included in any model reduction procedure that employs the methods described in this thesis.

For POD computations:

- The orthogonality of the leading POD modes was verified.

- The sum of the total energy in the snapshots must be equal to the sum of the POD eigenvalues. The sum of the diagonal entries of the matrix XX^T is the integral of the kinetic energy over the entire dataset, and by a standard matrix property $\mathrm{Tr}(XX^T) = \sum_j \lambda_j$, where λ are the eigenvalues. A discrepancy between the actual energy in the dataset and $\mathrm{Tr}(XX^T)$ may indicate an error in the implementation of correct grid weights or time quadrature weights.

- A convergence study of POD eigenvalues and the corresponding modes for different numbers of snapshots and spacings between snapshots was performed. In general, the results of a POD computation should be trusted only if a computation with a larger number of snapshots does not produce significantly different results.

An example of the application of the above procedure to the computation of POD modes of a localized perturbation studied in the next chapter is described in Appendix C.

Similarly, for BPOD:

- The bi-orthogonality of balancing and adjoint modes must be verified. Depending on the choice of the inner product and the discretization of the adjoint operator, the accuracy may not be as high as for orthogonality of POD modes. For example, computing derivatives for the inner product defined by 3.14 introduces small numerical errors. We also note that due to the fact that a discrete adjoint does not exactly satisfy the property (3.29), using the method described in Section (3.2.1) can slightly impact the accuracy of the mode computation, although this effect was not found to have significant effects on the results reported here.

- A convergence study of Hankel singular values and modes for different numbers of snapshots and spacings between snapshots was performed. While this step is not always practical for large-scale computations, it should always be performed for smaller calculations. In general, results should be trusted only if a larger number of snapshots does not produce significantly different results.

For both POD and BPOD, since the modes are linear combinations of simulation snapshots, it was verified that the modes satisfy the same boundary conditions as the snapshots. A visual inspection of modes should always be done for resolution, symmetry or boundary condition issues, and to confirm that the structures are reasonable from a physical point of view.

3.3 Software tools

When computing reduced-order models, we require a set of software tools that allows flexibility in the computations. To that end, here we discuss some computational approaches to the problems described in this thesis.

3.3.1 The DNS solver

The basic nonlinear DNS code used in this thesis was written by C.W. Rowley, following Kim *et al.* (1987). The code is pseudospectral, meaning that it employs Fourier transforms in the streamwise and spanwise directions and a Chebyshev transform in the wall-normal direction, but the nonlinear terms are computed in physical space. The FFTW library (Frigo & Johnson (2005), also see `http://www.fftw.org`) was used for the Fourier transforms. The discretization in time is performed using the second-order Adams-Bashforth scheme for the nonlinear terms and the Crank-Nicholson scheme for the linear terms (the overall scheme is sometimes denoted as ABCN2). The code was written in Fortran 90. The contributions to the code related to this thesis are listed below.

1. Implementation and testing of an adjoint solver, using the equations (3.17).
2. Implementation of constant mass flux (see Appendix B) for the nonlinear simulations.
3. Addition of diagnostic outputs such as velocity probes, measurement outputs and the energy norm.
4. Implementation of feedback control and estimation modules for testing of control designs.
5. Re-organization of most modules to make the code more readable and more easily modified. The addition of new features required new modules and the division of old ones into functional units.
6. Implementation of the HDF5 data format for input/output. HDF5 (Hierarchical Data Format, see `http://www.hdfgroup.org`) is a standard data format that has been implemented on various platforms and with support for many languages including MATLAB, Fortran, and Python. Loaders for the format are also available in commercial visualization software such as Tecplot.

The code was originally parallelized by Mingjun Wei using MPI, and all new features are enabled in the parallel version. Previous to this work, the original version of the channel code was used in the investigation of Lagrangian Coherent Structures (Green *et al.*, 2007), and a modification of the code was used to study reduced-order models for Couette flow (Smith, 2003; Smith *et al.*, 2005). The linearized version (including the adjoint code) was used for the studies reported

in this thesis (Ilak & Rowley, 2008*b*,*a*). A validation study of the channel code for the turbulent case is presented in Appendix B.

3.3.2 A modular tool for computing modes and models

Since most of the steps in the computation of modes involve postprocessing of DNS simulation data, some of the features of the DNS code can be used for that purpose. These common features between the DNS code and the postprocessing codes can be thought as 'building blocks' — for example, 'black box' Fourier and Chebyshev transforms, a Poisson boundary-value problem solver or input/output routines.

Modular design. The goals of having a highly optimized fast parallel solver for large problems and a flexible postprocessing tool using the same 'building blocks' can sometimes be incompatible. While a Fortran 90 code whose structure was based on the DNS code data structure was used initially for postprocessing, the need for a flexible tool for the later work led to the development of an object-oriented tool with a modular structure. Python (`http://www.python.org`) was chosen for this task. This language interpreted, object-oriented language with fairly simple syntax has recently gained popularity in scientific computing. The main goal in developing the new tool was that the combination of easy use and the object-oriented organization structure would allow for fast creation of scripts, rather than more complex and time-consuming implementation of new features in a Fortran code. The object-oriented framework allows for easy inclusion of 'black-box' elements for the different computations. The tool is similar to a simple language, as many tasks can be performed using short scripts that call the 'black-box' elements.

Testing. Another important feature of the object-oriented approach is the easy integration of tests in the code due to its modularity. In addition, testing frameworks make incorporating automated testing into codes written in Python or C++ a simple process. Thus, it is possible at any moment during the development process to test the performance of the different building blocks and their interaction. Also, problems with migration from one computer to another can be easily identified this way, since a self-test of the code after each new installation will reveal errors immediately. Of course, one should still carefully examine results of all simulations, since tests can only discover problems which have occurred previously in the code and do not guarantee that the code is completely bug-free. The Python tool is still in its development stages; however, it has been used successfully for postprocessing for the results presented in Chapters 5 and 6.

3.3.3 Other tools

MATLAB was used for many of the computations, in particular the flow control design. Many of the postprocessing steps for the simulation data can be done

in MATLAB as well on machines with sufficient memory. For example, all the postprocessing for the channel code validation in Appendix B was done using MATLAB. The 1-D computations in Section 4.2 were done entirely in MATLAB; The Matrix Differentiation Suite (Weideman & Reddy, 2000) was used for some of the computations. Visualization of 3-D fields was accomplished using Tecplot, although MATLAB can be used for this purpose as well.

3.4 Summary

Transitional channel flow, which we attempt to control using reduced-order models in this thesis, is described by the incompressible Navier-Stokes equations. A DNS solver is used for both the linear and nonlinear simulations used in this work. The evolution of linear perturbations is described by the linearized Navier-Stokes equations in velocity-vorticity formulation, and the corresponding adjoint equations have been derived in order to obtain the necessary snapshots for computing BPOD from adjoint simulations. An existing parallel DNS code was thoroughly tested and considerably upgraded. The inner product used in deriving the adjoint equations was chosen so that minimal changes in the DNS solver were required. The computation of modes, models and controllers requires a combination of different software tools, and a modular set of tools for computing modes and models was developed in Python.

Chapter 4

Modeling of transitional channel flow

When reduced-order models are designed for use in feedback control, it is desirable to investigate their properties and performance in detail. A logical first step in this direction is to evaluate the performance of the model in capturing the dynamics of the original system without control feedback (also known as *open-loop*), which is what we undertake in this chapter. While good open-loop performance may not always be necessary for good closed-loop performance (a model may fail to capture all the relevant dynamics of the full system, but it may capture enough of the input/output dynamics for the closed-loop performance to be satisfactory), a careful investigation of open-loop performance still provides valuable insight into the model reduction techniques used, in this case BPOD and POD. Therefore, in this chapter we investigate the performance of BPOD and POD reduced-order models of perturbations to linearized channel flow. The use of balanced reduced-order models for feedback control of both linearized flow and nonlinear DNS simulations will be presented in the next chapter.

A comparison of BPOD to exact balanced truncation for a system where the latter is tractable was performed by Rowley (2005) for streamwise-constant linear perturbations to channel flow. Here we make a more detailed comparison for a one-dimensional single-wavenumber perturbation with optimal energy growth, which indicates that BPOD is indeed a very close approximation to balanced truncation, and that the difference between the two methods becomes significant only for higher modes, which typically have very little influence on capturing the system dynamics. This analysis also confirms the conclusion of Rowley (2005) that BPOD outperforms POD, which often fails to capture the correct input/output dynamics of the original system.

Next, balanced truncation of a 3-D system without decomposition into single-wavenumber pairs is performed for the first time. Balanced models of a localized perturbation show good performance, and significantly outperform POD. The emphasis for this system is on capturing the non-normal growth of the perturbation

energy, which is why the output is chosen to be the full velocity field, and an output projection is necessary in order to compute the models. Some useful performance measures of the models are described in detail, in particular the impulse response, frequency response, off-design condition performance and capturing of actuation. It is shown that BPOD models perform well according to those measures, making them potentially suitable for developing closed-loop controllers.

As a result of a detailed investigation of the performance of the POD models of the 3-D perturbation, a striking example was found in which modes containing negligible energy are very significant for the system dynamics. A close look at the dynamics of a model containing these modes, inspired by the recent work of Mezić (2004) in graph decomposition analysis of dynamical systems, is presented at the end of the chapter. Most of the results of this chapter have been published in Ilak & Rowley (2008*b*).

4.1 Choice of actuators and measurements

The choice of actuators and measurements is key in any control system. In order to motivate the selection of inputs and outputs in this chapter, we first introduce the overall feedback control strategy used in this work. Here we have two outputs — an output of interest y_1, which we typically desire to either drive to zero or make follow a particular reference input (this is known as a regulator problem), and a measurement output y_2 used for the computation of the control feedback. The two inputs are a disturbance which we would like to suppress (B_1) and which we have no control over, and an actuator which we use to control the system (B_2):

$$\begin{aligned} \dot{x} &= Ax + B_1 u_1 + B_2 u_2 \\ y_1 &= C_1 x \\ y_2 &= C_2 x. \end{aligned} \tag{4.1}$$

This system (same as system (3.11), just with a second output added) belongs to the class of MIMO (multiple-input multiple-output) systems, and a graphical illustration for the system is shown in Fig. 4.1, along with a schematic of the inputs and outputs for channel flow with localized inputs and outputs. The term 'plant' is a standard term from control theory and it encompasses the underlying dynamical system along with its actuators and measurements, while the term 'compensator' stands for the device (or devices) used to achieve the desired effect on the system output. The control designer's task is to design a compensator for a given plant. The different options for inputs and outputs are discussed below. It is important to select control inputs (actuators) that are going to affect the flow most favorably and outputs which will be most useful for the determination of the system performance and for computing the control input.

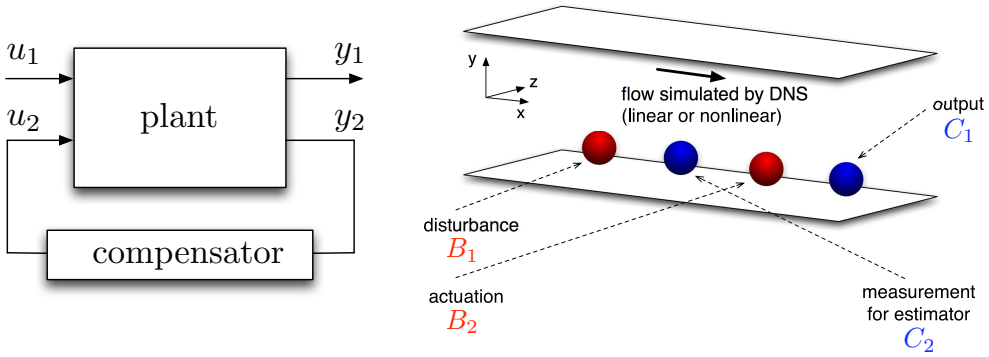


Figure 4.1: A schematic of the overall feedback control system setup. Left: the basic control system setup. Right: a schematic of channel flow with localized disturbance, measurement, actuation and output.

4.1.1 Choice of actuation

Two main mechanisms for actively manipulating wall-bounded flows that have been considered in numerical and experimental studies are body force actuation and wall blowing/suction. The former choice is attractive due to easier implementation in numerical studies, in particular when reduced-order models are developed, while the latter is thought to be more relevant from a practical standpoint (Lundell, 2003). The potential of plasma actuators, whose effect on the flow can be modeled by body forces, has been discovered through experimental studies in the past decade (for a review, see Corke *et al.*, 2009). While these actuators may offer an advantage over wall blowing/suction, as wall actuation may be limited due to uncontrollability of modes which correspond to flow structures far away from the wall (Kim & Bewley, 2007), recent results in experiments with wall actuation using suction slots have also shown promise (Lundell, 2007; Lundell *et al.*, 2009).

The inclusion of wall blowing/suction in modeling and simulation comes with added computational complexity. While it is straightforward to include inhomogeneous boundary conditions in the DNS simulation, including them in model computation as control inputs is more complex, as will be described in the next chapter. For this reason, body forces were initially used to characterize the performance of balanced truncation models, such as in the models described in this chapter, as well as, for example, in the work of Bagheri *et al.* (2009*b*).

4.1.2 Choice of measurements

The key quantities of interest in shear flow diagnostics include the pressure field or localized pressure measurements, measurements of wall shear (corresponding to drag), and measurements of full velocity fields or from localized velocity probes. While all these quantities are typically available in simulations, in practical flow

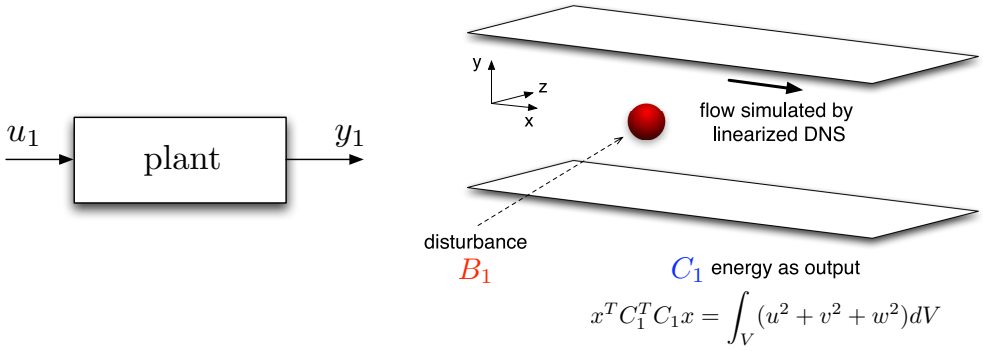


Figure 4.2: A schematic of the simplified control setup for the open-loop investigation in this chapter. The only input is a perturbation, in this case localized.

control applications this is most often not the case. It is therefore important to design control strategies using physically realizable sensing mechanisms, and to model those mechanisms correctly in simulations. Wall shear measurements are especially attractive since they can be non-intrusive and sensors can be wall-mounted, but a potential drawback, as for other wall measurements, is that they may not capture the influence of flow structures that are far from the wall and effectively unobservable (Kim & Bewley, 2007).

While it is feasible to measure velocity, pressure and wall shear in experiments, the measurement of the full field in real time is still impracticable, especially in three dimensions. In the initial study undertaken in this chapter, the output is the kinetic energy of the entire perturbation flow field. The reason for this choice was twofold — to evaluate the capabilities of the balanced models to capture the large transient energy growth of linear perturbations, and to test carefully the output projection discussed in Chapter 2. This choice of input and output for a localized perturbation in the center of the channel, described later in Section 4.3.1, is graphically depicted in Fig. 4.2. The use of different measurements and actuators for feedback control is described in the next chapter.

4.2 Validation of BPOD on a 1-D problem

We start by investigating the system given by (4.1) only in the presence of disturbances (without the Bu_2 term) and without the measurement term $C_2 x$. In order to validate the numerical methods, we first obtain BPOD models from three-dimensional simulations of simple and well-known single-wavenumber perturbation cases, described by Butler & Farrell (1992) and also investigated by Schmid & Henningson (2001). As discussed in Section 3.1.1, the general form of such disturbances is given by

$$q(x, y, z, t) = \hat{q}(y, t) e^{(i\alpha x + i\beta z)}, \tag{4.2}$$

with $\hat{q}(y,t) = [\hat{v}(y,t)\ \hat{\eta}(y,t)]^T$. The standard approach to such perturbations is to compute the time evolution of $\hat{q}(y,t)$, which fully describes the system, since the velocity components u and w can easily be computed from this data. For this one-dimensional problem, standard algorithms for computing balanced truncation are computationally tractable. Therefore, we are able to compare the models resulting from exact balanced truncation[1], to BPOD models obtained from three-dimensional simulations of the real part of the full field, $Re\{q(x,y,z,t)\}$ at a particular wavenumber pair (α,β) on a large grid, similar to the comparison made by Rowley (2005) for a streamwise-constant perturbation. We note that for a given wavenumber pair the comparison between BPOD and exact balanced truncation can be computed using only 1-D simulations, but we also performed 3-D simulations in order to verify our codes. We also note that, since the outputs of the output-projected system and the reduced-order models are coefficients of POD modes, the C_1 matrix in (4.1) was modified so that the output of the full system is expressed in the POD basis as well. In this way, a meaningful comparison between the balanced truncation of the full system and BPOD is obtained.

The perturbations were chosen so that they exhibit the largest, or optimal (alternatively, 'worst-case'), energy growth. They were computed using the method described by Reddy & Henningson (1993) and their energy growth was verified against values reported in that work. While streamwise-constant perturbations exhibit the largest energy growth, perturbations where both $\alpha \neq 0$ and $\beta \neq 0$ exhibit more interesting dynamics. We focus on the $\alpha = 1, \beta = 1$ perturbation at $Re = 1000$, whose energy growth and streamwise velocity profile are shown in Fig. 4.3. The computational grid used in the three-dimensional simulation was $16 \times 65 \times 16$, corresponding to 33280 states in the system given by (3.11). Balanced truncation of the 1-D problem with 65 Chebyshev modes is easily and accurately computed using the algorithm described in Rowley (2005) so that BPOD performed on the large system can be compared to exact balanced truncation.

4.2.1 Mode subspaces

It was found that 500 equally spaced snapshots are sufficient for accurate computation of the POD modes, since for a larger number of snapshots with finer spacing there is no considerable change in the eigenvalue spectrum or the corresponding modes. We see from Fig. 4.4 that the most significant eigenvalues and the corresponding modes typically come in pairs, representing traveling structures that are 90 degrees out of phase. The first pair of modes contains 90.45% of the energy, while the first three pairs contain 99.6% of the energy. For the balanced POD models, four-mode and eight-mode output projections were chosen (see Section 2.3.3), cor-

[1]For the 1-D case, exact balanced truncation can easily be computed in MATLAB using standard algorithms, such as the one in Laub *et al.* (1987)

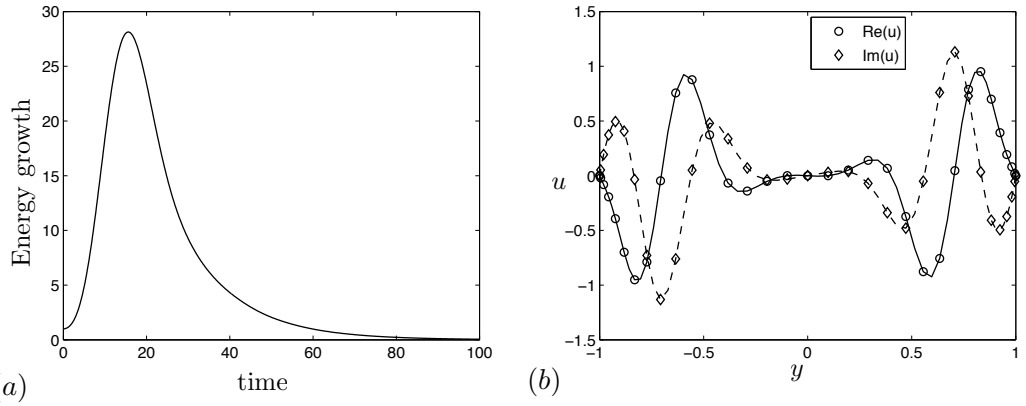


Figure 4.3: (a) Kinetic energy growth for the optimal perturbation at wavenumber $\alpha = 1, \beta = 1$ at $Re = 1000$. (b) The $\alpha = 1, \beta = 1$ optimal perturbation, showing streamwise velocity u (complex).

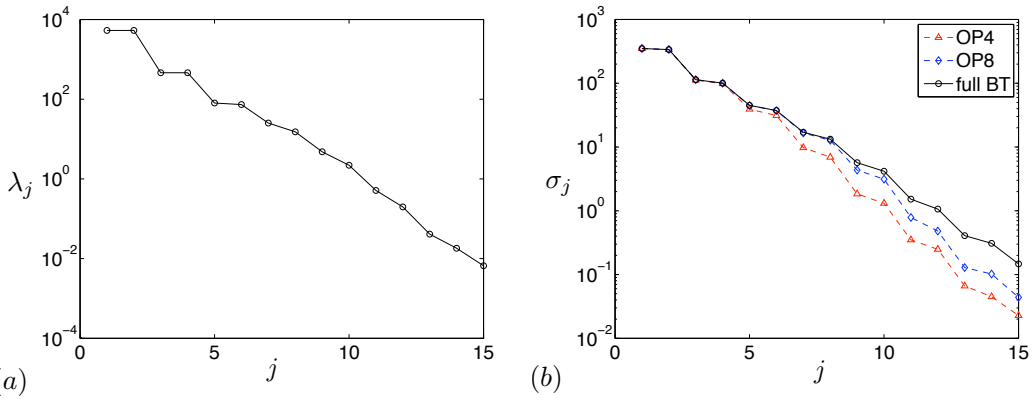


Figure 4.4: (a) The first 15 POD eigenvalues for $\alpha = 1, \beta = 1$ initial perturbation at $Re = 1000$. (b) The first 15 Hankel singular values (HSVs) for: four-mode ($\triangle$) and eight-mode ($\diamond$) output projections and full balanced truncation ($\circ$) for the same case.

responding to respectively 98.3% and 99.9% of total energy contained in the POD modes.

We also notice that the HSVs (Fig. 4.4) come in pairs, indicating that the most significant modes in the BPOD mode basis are again traveling structures similar to the POD modes. It is important to include these pairs of modes in the reduced-order models, as stability of the models for balanced truncation is guaranteed only if $\sigma_r > \sigma_{r+1}$ where r is the rank of the model (Dullerud & Paganini, 1999). While for POD modes there is no such requirement, mode pairs should always be included in the models on physical grounds. We also notice that, for each output projection rank s (see Section 2.3.3), approximately the first s HSVs for BPOD and exact balanced truncation are equal. The same observation was made by Rowley (2005), although there is no proof of this property at this point.

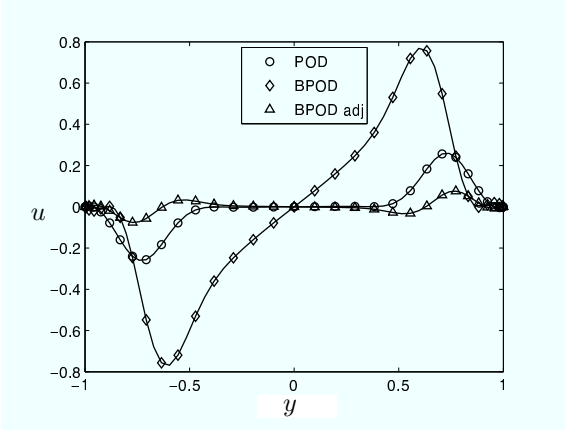


Figure 4.5: Streamwise velocity for the first POD mode, balancing mode and adjoint mode for the $\alpha = 1, \beta = 1$, $Re = 1000$ initial condition.

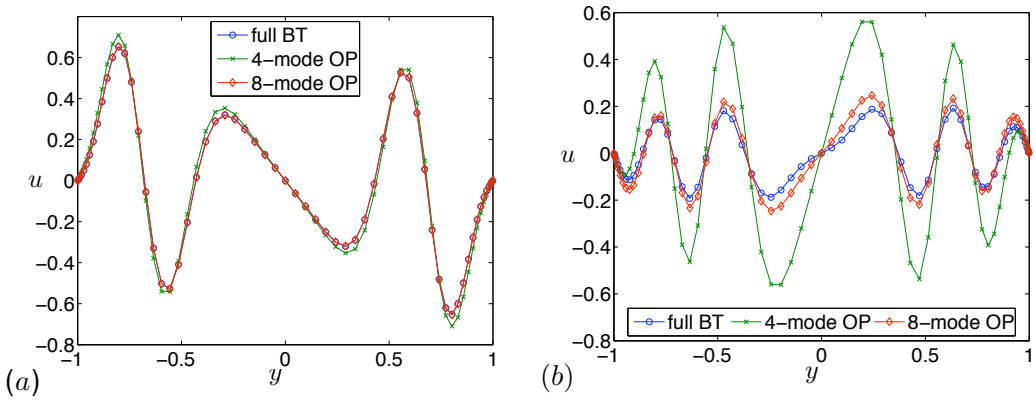


Figure 4.6: Streamwise velocity for (a) the sixth balancing mode and (b) the tenth balancing mode for BPOD with two different output projections and for full balanced truncation.

The first POD mode is shown in Fig. 4.5 together with the first balancing and adjoint modes from a four-mode output projection. Figure 4.6 shows the streamwise velocity of the sixth and tenth balancing modes, illustrating the effect of the choice of output projection rank. The first four balancing modes from BPOD look identical for both output projections, while the sixth mode is not very accurately captured by a four-mode output projection. Both output projections do not capture very accurately the higher modes such as the tenth mode, which is also illustrated by the HSVs in Fig. 4.4. As we show below, this inaccuracy does not significantly affect model performance, since these higher modes are not very significant dynamically.

In this single-wavenumber case, the exact eigenvalue spectrum of the A matrix from (4.1) at a given Reynolds number can easily be computed. We note here that the eigenvalues of the matrix A and therefore the poles of the corresponding transfer function are independent of the initial condition (which is just the B matrix for our impulse response simulations). Figure 4.7 shows the spectra of the full operator and

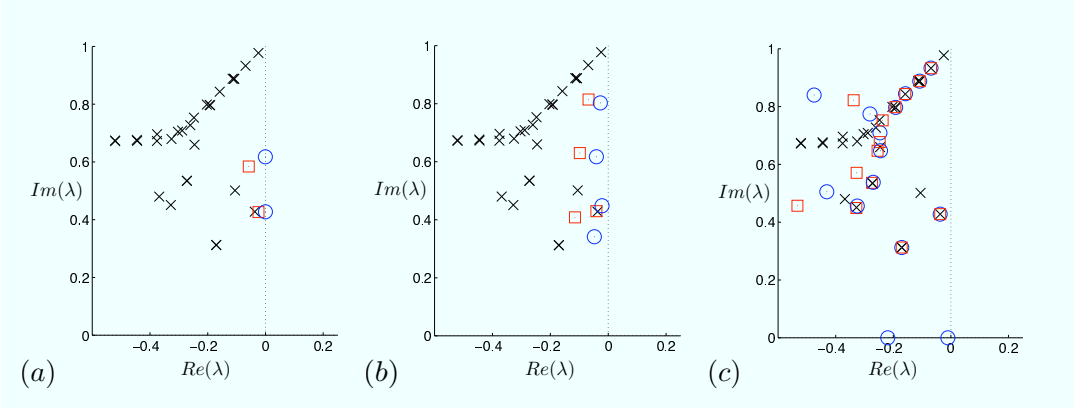


Figure 4.7: Spectrum of the full operator and reduced-order models for rank (a) 4, (b) 8, (c) 30. The BPOD modes are from the eight-mode output projection. Symbols: BPOD (□), POD (◯), full operator (X). Only the most important part of the full spectrum is shown (i.e., the part closest to the imaginary axis). Some of the eigenvalues correspond to uncontrollable eigenmodes and cannot be captured by the models (see text for explanation). The 30-mode BPOD model includes some spurious eigenvalues on the real axis, which were not found to affect the model performance.

three reduced-order models of different rank for the $\alpha = 1, \beta = 1$ perturbation. Since the spectra are symmetric about the real axis, we only show the upper half of the complex plane. We see that, while the representation of the full spectrum improves for both methods as the rank increases, BPOD captures more accurately some of the most slowly decaying eigenvalues, which have the most influence on the dynamics of the system. For the rank four model, the POD model appears to be marginally stable, while the BPOD model closely approximates the eigenvalue closest to the origin. At higher order, the POD models improve and capture approximately the same eigenvalues as the BPOD models of the same rank. It is also important to notice that some of the eigenvalues of the full system are never captured by the reduced-order models. These eigenvalues correspond to uncontrollable eigenmodes of the full system, and can never be excited by this particular perturbation.

4.2.2 Impulse response

We next compare the impulse response of the system to that of the reduced-order models. 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. Figure 4.8 shows the capturing of the growth of kinetic energy by POD and BPOD models, as well as the first two outputs of the reduced-order models. The poor performance of low order POD models for the traveling structure perturbation is evident. Even the eight-mode POD model, which captures the energy growth well, does not accurately capture the phase of the oscillations at later times.

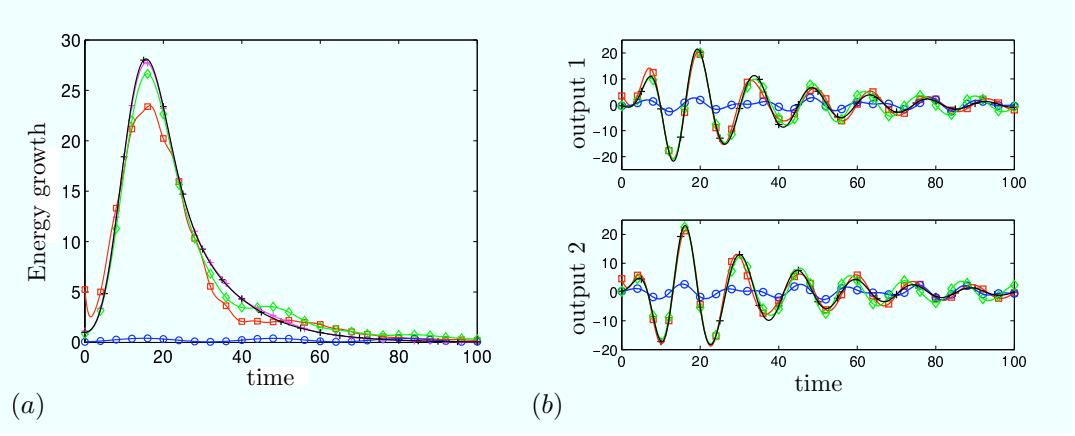


Figure 4.8: (a) $\alpha = 1, \beta = 1$ optimal perturbation at $Re = 1000$, eight-mode output projection, 4-mode and 8-mode models. Full simulation (+), 4-mode POD (◯), 4-mode BPOD (□), 8-mode POD (⋄), 8-mode BPOD (+). The 8-mode BPOD model impulse response is almost indistinguishable from the full system's impulse response. (b) First two outputs, symbols as defined in (a).

Figure 4.9 shows the 2-norm of the error between the impulse response of the reduced-order model and the full simulation, given by $\|G - G_r\|_2$, normalized by the 2-norm of the impulse response of the full simulation $\|G\|_2$, as a function of the model rank r. A discussion of the 2-norm and other transfer function norms is given in Appendix D.

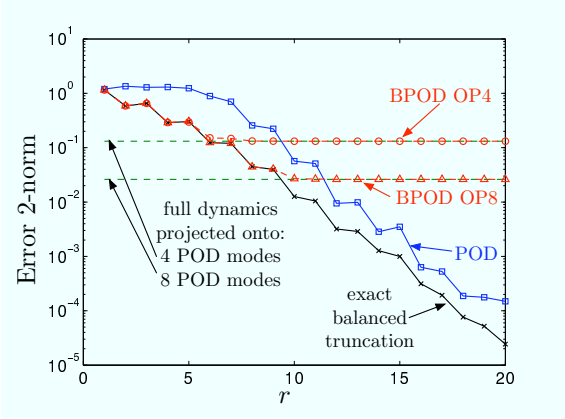


Figure 4.9: Error 2-norms for the $\alpha = 1, \beta = 1, Re = 1000$ perturbation for full balanced truncation, POD and BPOD at two output projections.

This figure provides a clear demonstration of the effect of output projection. A four-mode output projection means that we are effectively performing balanced truncation on the dynamics of the first four POD modes of the full system. The dashed lines in the figure indicate the 2-norms of the error between the full dynamics of the output-projected system and the full system. As the rank of the BPOD models is increased, the dashed lines, which are the limit of accuracy, are reached

fast. As already seen in Fig. 4.8, for very low-order models, POD is clearly outperformed by balanced POD. A ten-mode POD model is necessary in order to capture fully the dynamics of the first four POD modes (i.e., to cross the upper dashed line), while a BPOD model with a four-mode output projection (i.e., designed to capture well the dynamics of the first four POD modes) captures the dynamics of the first four POD modes quite well already at rank six. The slow improvement in POD model performance indicates that the dynamics of the perturbation cannot be represented only by retaining the first few POD modes (which in this case are also the most controllable modes). Adding new BPOD modes beyond rank eight and ten (for four- and eight-mode output projections, respectively) does not improve the model performance noticeably, since the dynamics of the output-projected system is already captured fully. It is also important to note that the performance of the BPOD models is identical to that of full balanced truncation almost until the rank at which BPOD model error norms level off due to the output projection. This indicates that the higher balancing modes which are not computed accurately due to the approximation inherent in the output projection (such as those shown in Fig. 4.6) do not significantly influence the reduced-order model performance, the main limitation being the capturing of the full system by the output projection.

4.2.3 Frequency Response

The frequency response encompasses system behavior over the complete range of possible forcing, and is perhaps the best indication of overall system performance. Therefore, from the control designer's point of view, having a low-order model that represents well the frequency response of the original system is of key importance. Frequency response of single-wavenumber perturbations was investigated by (Schmid & Henningson, 2001) using the resolvent norm (see Appendix D), where at each frequency the maximum amplification over all initial conditions is computed. Here the frequency response of the system with a given actuator or perturbation is of interest.

A standard way of representing synthesized frequency response for MIMO (multiple input multiple output) systems is a plot of the maximum singular value of the transfer function matrix $\max(\sigma(H(i\omega)))$ as a function of frequency, also known as a singular value Bode plot. The highest peak of this plot is the infinity norm of the transfer function (see Appendix D). Fig. 4.10 shows such plots for the $\alpha = 1, \beta = 1$ perturbation and clearly demonstrates the advantages of BPOD for capturing the dynamics of the system. We see that even for a two-dimensional model the resonant peak is captured well by the model, while for POD the peak is very narrow, with very low response at other frequencies. This behavior is typical of balanced truncation, as shown in Dullerud & Paganini (1999) — the first modes to be captured are the ones which are most significant dynamically, while the correct response is gradually built up in less significant frequency bands as more modes are added. For POD

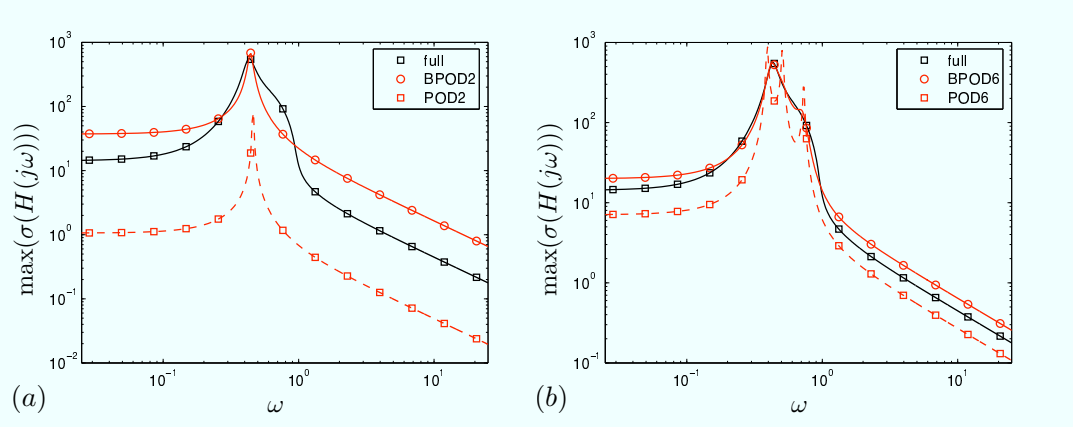


Figure 4.10: Frequency response of the models for the $\alpha = 1, \beta = 1$ perturbation. (a) 2-mode POD (dashed), 2-mode BPOD (full), (b) 6-mode POD (dashed), 6-mode BPOD (full).

models, on the contrary, the response improves incrementally at all frequencies as more modes are added and a higher number of modes is needed to accurately capture the resonant peak. For ten-mode models, both POD and BPOD perform well (not shown in the figure), with BPOD frequency response being almost indistinguishable from the frequency response of the the full system. POD frequency response also includes spurious non-physical peaks at low order, which correspond to eigenvalues very close to the imaginary axis for low order of truncation, as seen in Fig. 4.7 (a).

We are also interested in the worst-case error between the reduced-order model and the full simulation (the infinity error norm), which was described in Section 2.3.1. Figure Fig. 4.11 shows the infinity norm of the error transfer function between the full system and the reduced-order model as a function of model rank for the first fifteen orders of truncation. The infinity norms for exact balanced truncation lie within the theoretical bounds given by Eqns. (2.12) and (2.13), while for BPOD, for each of the two output projections, the norms stay within those bounds up to approximately the rank of the output projection, analogous to the two-norm behavior shown in Fig. 4.9. The infinity norms for POD at low rank are considerably higher than those for balanced truncation and BPOD, corresponding to the frequency responses shown in Fig. 4.10 (a) and (b).

4.2.4 Variation of Reynolds number

Another very desirable feature of a reduced-order model is good performance for off-design values of the system parameters. We would like the models to remain valid for a wide range of the model parameters, or at least for the range appropriate for the physical application of the model. The only parameter we are considering in our models is the Reynolds number, so the response of models was compared to the full simulation when Re is changed. Separating the operators from (3.9) into

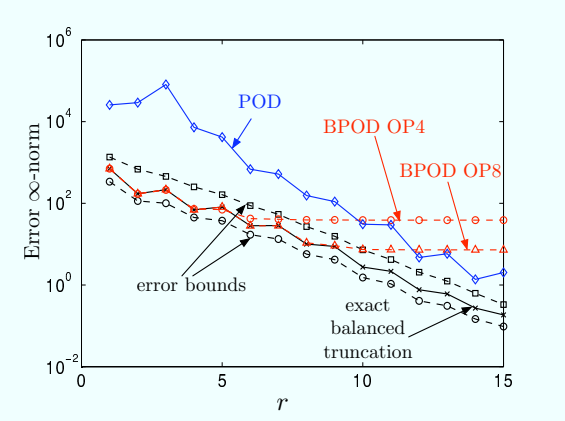


Figure 4.11: Infinity error norms for POD, exact balanced truncation and BPOD with the infinity error bounds.

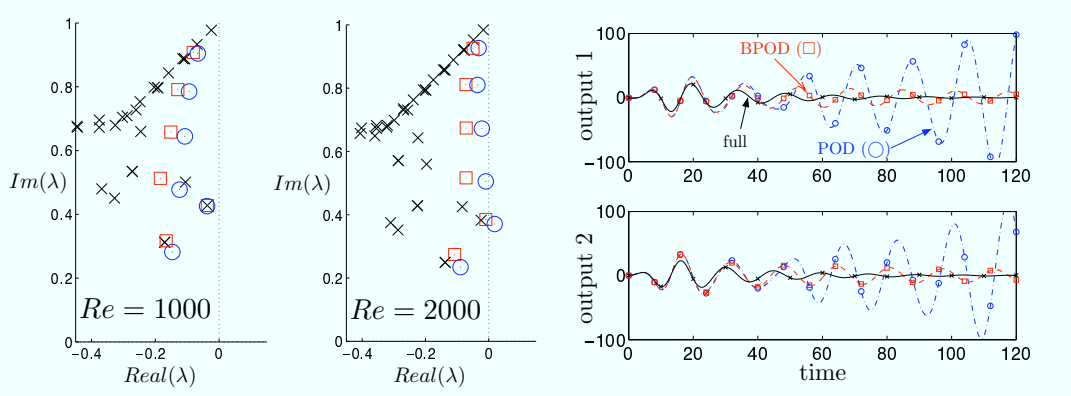


Figure 4.12: Left: The comparison of spectra of the full operator at $\alpha = 1, \beta = 1$ to the spectra of rank 12 models as the value of Re is increased to 2000. Right: The performance of corresponding reduced-order models at $Re = 2000$ - first two outputs. See text for detailed description. Symbols: full (×), BPOD (□), POD (◯).

convective and diffusive parts, we can re-write the state-space equation as

$$\dot{x} = A_{\text{conv}}x + \frac{1}{Re}A_{\text{diff}}x + Bu, \tag{4.3}$$

where:

$$A_{\text{diff}} = \begin{bmatrix} -\Delta & 0 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} -\Delta^2 & 0 \\ 0 & \Delta \end{bmatrix}, \tag{4.4}$$

with all the other terms in (3.8) contained in A_{conv}. We can then separately project the matrices A_{conv} and A_{diff} as in Eq. (2.19) at any Reynolds number onto the POD and BPOD modes obtained at $Re = 1000$, the B matrix being just the initial condition at $Re = 1000$. Figure 4.12 shows the performance of 12-mode POD and BPOD models when the value of Re in (4.3) was changed to 2000 and the impulse response of the resulting models was compared to the impulse response of the full system. This rank of the model was chosen since both models perform well at the

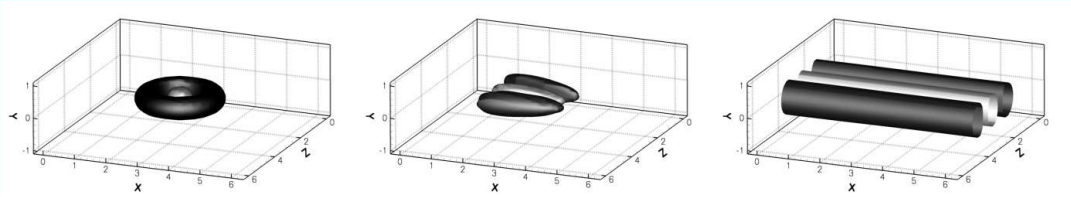


Figure 4.13: The development of the wall-normal velocity of the perturbation given by Eq. (4.5) at $t = 0$ (left), $t = 14$ (middle) and $t = 160$ (right) which corresponds to the maximum energy growth. The isosurface of half of the maximum value of v is light, and isosurface of half the minimum of v is dark.

design condition of $Re = 1000$. We see that the BPOD model eigenvalues stay closer to the full simulation eigenvalues (which move as well), and also remain in the left half of the complex plane, while for $Re = 2000$ the POD model becomes unstable due to one pair of eigenvalues crossing into the right half of the complex plane. This indicates a greater range of validity for BPOD models and better stability at off-design conditions than POD.

4.3 Models of a 3-D localized perturbation

We next consider an actuator with a velocity distribution that cannot be described as a one-dimensional problem, corresponding to a localized body force in the center of the channel. This case corresponds to (4.1) without the $B_2 u_2$ term, with the input matrix B_1 representing the velocity field in Fig. 4.13. Individual localized disturbances to channel flow were investigated by Henningson *et al.* (1993). Since balanced truncation involves the approximation of the system's Gramians (although in BPOD we do not actually compute the Gramians explicitly), we are interested in following both the direct and adjoint impulse-state responses until all transients have completely decayed. The computational box necessary for following individual localized disturbances long enough in time would be very large, and we instead consider a periodic array of small disturbances in the channel center. It should be noted that the behavior of this periodic array can be quite different from the behavior of a single localized disturbance, in particular in terms of the energy growth, since the periodic array quickly develops into a streamwise-constant vortex. The exact form of the initial condition considered here is

$$v(x, y, z, 0) = A \left(1 - \frac{r^2}{\alpha^2}\right) e^{(-r^2/\alpha^2 - y^2/\alpha_y^2)} (\cos(\pi y) + 1) \tag{4.5}$$

where $x_c, 0, z_c$ are the coordinates of the center of the computational domain and $r^2 = (x - x_c)^2 + (z - z_c)^2$. The wall-normal vorticity perturbation is zero. This form was picked in order to satisfy the condition that the mean perturbation velocity is zero at each wavenumber. The $(\cos(\pi y) + 1)$ term was added to make the field satisfy exactly the boundary conditions $v(\pm 1) = v_y(\pm 1) = 0$. The amplitude A was

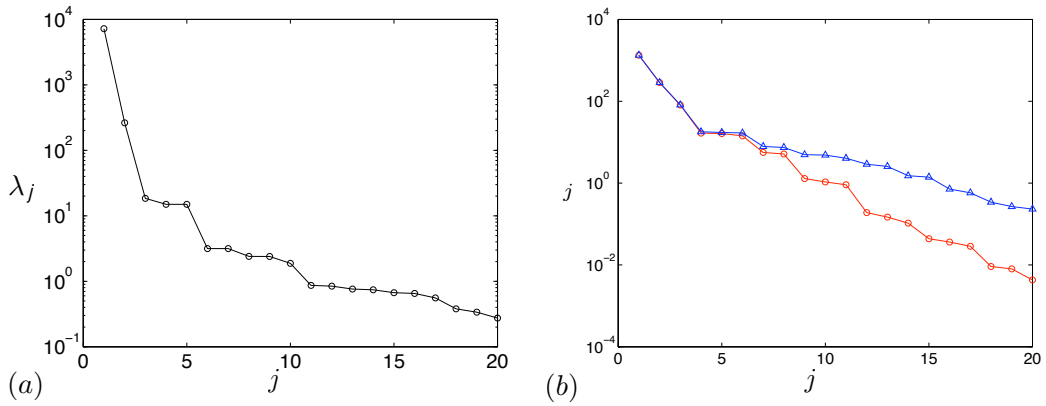


Figure 4.14: (a) The first 20 POD eigenvalues for the Gaussian-like disturbance impulse response. (b) The first 20 HSVs for five-mode (∘) and ten-mode (△) output projections.

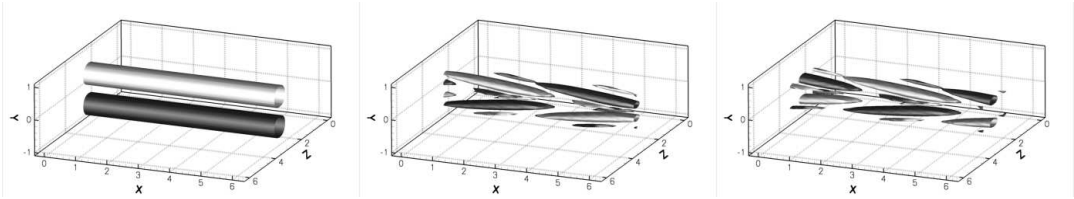


Figure 4.15: Left to right: the first, fourth and fifth POD modes for the localized actuator, showing streamwise velocity. The isosurface of half of the maximum value of u is light, and isosurface of half the minimum of u is dark.

set to 1 for this simulation (this has no impact on the results, since the simulation is linear), and the parameters α and α_y were set to $\alpha = 0.7$ and $\alpha_y = 0.6$. The Reynolds number chosen for this simulation was $Re = 2000$. The traveling structure rapidly develops into a streamwise-constant form, since the growth of wall-normal vorticity results in the development of streamwise streaks (see Fig. 4.13).

The grid size was $32 \times 65 \times 32$, corresponding to 133,120 states for the full (v, η) system. The simulation was run for 1200 dimensionless time units ($t = t^d U_c/\delta$), and the timestep used was $\Delta t = 0.004$. During this time, the energy of the initial

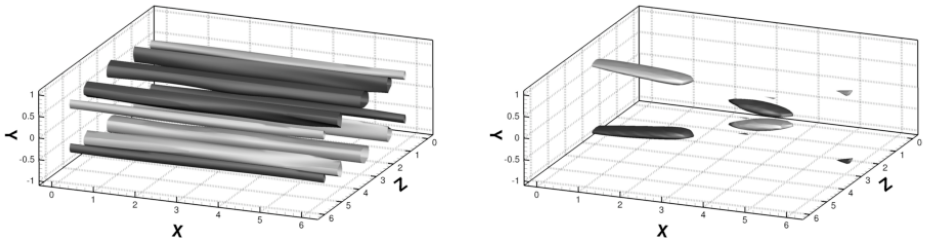


Figure 4.16: Left: POD mode 10, right: POD mode 17 for the localized perturbation, corresponding to 0.025% and 0.0074% of the total energy. The isosurface of half of the maximum value of u is light, and isosurface of half the minimum of u is dark.

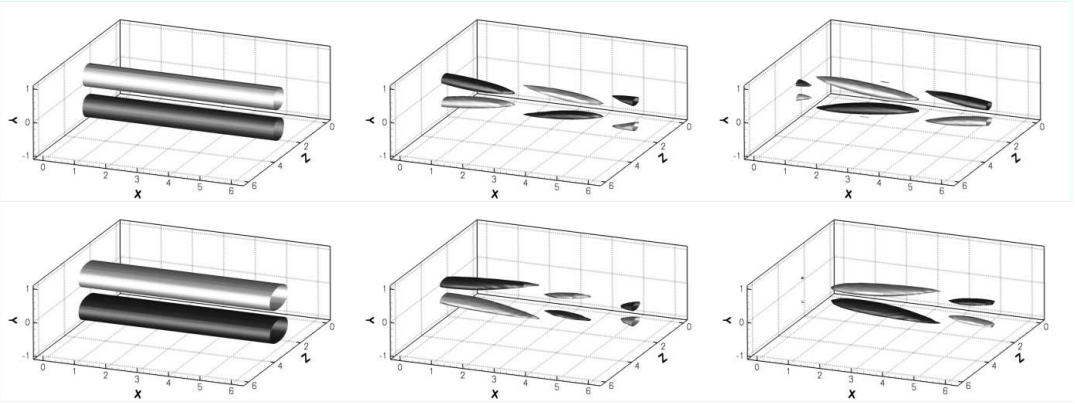


Figure 4.17: Top row: balancing modes one, four and five (left to right) from balanced POD, showing streamwise velocity for the localized perturbation. The modes are from a five-mode output projection. The isosurface of half of the maximum value of u is light, and isosurface of half the minimum of u is dark.. Bottom row: the corresponding adjoint modes. Note the similarity between the balancing modes and the corresponding POD modes in Fig. 4.15.

disturbance decayed to approximately 1.5 percent of its initial value. The POD modes were taken over 1000 snapshots, with fine spacing between snapshots for the initial period in order to capture the traveling structures well and larger spacing once the streamwise structure was developed, after it was verified that POD eigenvalues and the corresponding modes do not change significantly if more snapshots are used (see Appendix C). Fig. 4.14 (a) shows the POD eigenvalues of the impulse response. The first five modes contain 99.72% of the perturbation energy, and the first ten modes contain 99.9% of the energy. In this case the spectrum contains both streamwise-constant (and nearly-streamwise constant modes) as well as traveling structures due to the initial transient. The first three modes are streamwise-constant structures, while the fourth and the fifth modes correspond to a traveling structure, which accounts for only 0.40% of the total energy. Modes one, four and five are shown in Fig. 4.15.

Next, the adjoint simulations were computed and the BPOD procedure was performed on a five-mode output projection, containing only the most important traveling structure, as well as on a ten-mode output projection. These ranks were chosen due to large drops in energy significance after the fifth and tenth mode, as shown in Fig. 4.14 (a). Fig. 4.14 (b) shows the HSVs for these two output projections. We notice that the HSVs are equal for the pairs of modes 4–5 and 7–8 for five-mode output projection, corresponding to traveling structures in the basis of BPOD modes. Even more interestingly, for the ten-mode output projection, HSVs for the modes 4–6 are equal. Although the stability of balanced truncation models is guaranteed only when $\sigma_{r+1} < \sigma_r$, where r is the number of states retained (Dullerud & Paganini, 1999), 4-mode, 5-mode and 7-mode models for both output projections were found to be stable. The model error for impulse response, however, decreases significantly if both modes corresponding to a traveling structure are included, as

will be shown in Sec. 4.3.1 (see Fig. 4.19). Balanced POD modes one, four, and five, and the corresponding adjoint modes for the five-mode output projection are shown in Fig. 4.17. Note that the structure of modes four and five in Fig. 4.17 is almost identical, except for a spatial phase shift of exactly one quarter of the periodic domain.

4.3.1 Impulse Response

Figure 4.18 (a) shows the perturbation energy growth as captured by three different POD models. It was observed that the inclusion of modes which come in pairs (see Fig. 4.14 (a)) in the basis used to form the reduced order models does not change the system behavior appreciably — the response of a model including modes 1–9 (not shown in figure) is virtually indistinguishable from the response of the model including only the first three modes. Hence, the traveling structure modes do not contribute significantly to the dynamics of this perturbation. The inclusion of the tenth mode, which is streamwise-constant, improves the performance significantly, and the model composed of only the first three modes and the tenth mode performs as well as one including the first ten modes. In the same fashion, including the mode pairs 11–12, 13–14 and 15–16 does not affect the model performance. Including the seventeenth mode, which is also a streamwise-constant mode, improves the performance further. The tenth and the seventeenth mode (see Fig. 4.16) correspond to 0.025% and 0.0074% of the total energy. The low-order POD models were found to capture poorly the initial condition of the full simulation (this will be discussed in more detail in Sec. 4.3.4), so they were also started from different initial conditions at later times (before or around the peak energy growth), when the projection of the simulation onto POD modes is close to the full simulation data, and they still did not capture the correct peak and the subsequent decay of the energy.

On the other hand, the performance of very low-order BPOD models is significantly better. Figure 4.18 (b) shows the perturbation energy growth as captured by three different models. Although the two-mode model does not accurately capture the initial condition, it does represent the energy growth at later stages reasonably well. A three-mode BPOD model captures the kinetic energy of the full simulation very well except for the initial period. While more modes are needed to capture the initial transient exactly, if only the energy growth is of interest, the three-mode model is sufficient. This striking difference is an illustration of the advantage of balanced truncation — for POD it is difficult to know apriori which modes will be important for the system dynamics, as demonstrated above, and a good low-order model was found only after a careful examination of the mode basis which provided some insight into the underlying physics.

Figure 4.19 shows the error $\|G-G_r\|_2$ for BPOD using the two output projections and for POD. The error for the BPOD models quickly reaches the asymptotic values

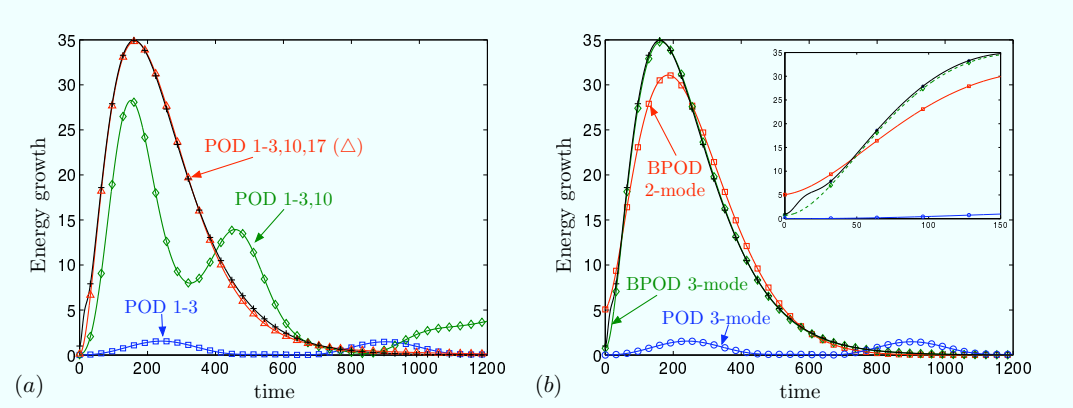


Figure 4.18: (a) Three, four and five-mode POD models formed from the indicated modes (b) Two-mode and three-mode BPOD models. The very low-order BPOD models do not capture very well the initial transient, as shown in the inset. The BPOD models are from the ten-mode output projection. The full simulation is marked by the black line (+) in both (a) and (b).

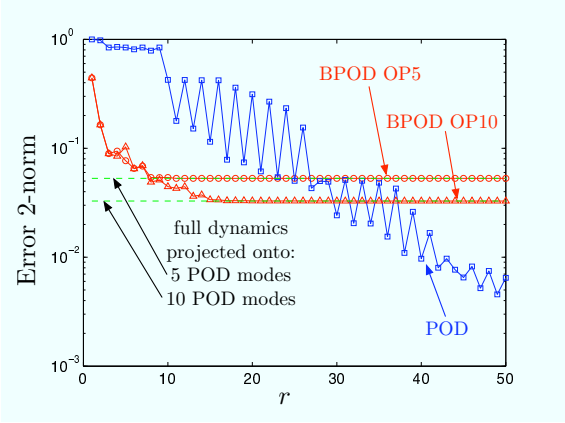


Figure 4.19: Error 2-norms for localized actuator, showing POD models and BPOD at two output projections.

dictated by the output projection, although more modes are needed compared to the optimal perturbation case described in the previous section due to the more complex dynamics. POD starts to match the performance of the ten-mode output projection BPOD only around rank 30, and also varies a lot with the model rank. This corresponds to the already observed fact that dynamically important POD modes are not highly ranked in terms of energy. Whenever the POD modes come in pairs, including only one of the modes results in unphysical oscillations and incorrect capturing of the dynamics. Eventually POD has better performance than BPOD, however recall that these POD models have r outputs, where r is the model rank, while the BPOD models have only s outputs and thus capture only the dynamics of the first s POD modes (recall that s is the rank of the output projection, defined in Section 2.3.3). It should also be noted that some POD models exhibit sustained or very slowly decaying oscillations, and that the corresponding two-norms are in

fact infinite, since the system 2-norm by definition corresponds to integration of the output from $t = 0$ to $t = \infty$ (see Appendix D). Since the final simulation time for Fig. 4.19 is finite, the 2-norms computed from simulation data for such models appear to be large but finite as well. Although the four-mode BPOD models are stable for both output projections, including just one of the modes corresponding to a pair of equal HSVs 4–5 for the five-mode output projection deteriorates model performance, while we see a large decrease in the error when the fifth mode is included, as well as when we include subsequent pairs. For the ten-mode output projection, there are three equal HSVs 4–6, and a significant decrease in the error is seen only when we include all three of those modes (in particular, the error norm of the five-mode BPOD model is significantly larger than that of the four-mode model).

4.3.2 Frequency Response

Figure 4.20 shows the singular value Bode plots of POD and ten-mode output projection BPOD models for the localized disturbance. The frequency response of the 50-mode BPOD model, which is a very close approximation of the frequency response of the actual disturbance, has the shape of a low-pass filter with a break frequency of 0.01 rad/s with two resonant peaks near 1 rad/s, which are similar to the peak observed for the single-wavenumber traveling structure perturbation in the previous section. We see that POD models again have spurious peaks at low model ranks. The addition of mode pairs corresponding to the traveling structures is necessary in order to reproduce the peaks around 1 rad/s for both POD and BPOD, however BPOD captures those peaks with only the triple 4–6 and the mode pair 7–8, as well as modes 9 and 10 (Figure 4.20 (b)) while all POD modes 1–17 are needed to reproduce the same peaks and there are still spurious peaks. Since the peaks correspond to the low-energy traveling structures, it is not surprising that only a three-mode BPOD model performs so well in capturing the kinetic energy of the full simulation, as shown in the previous section. On the other hand, if the frequency response of the actuator around the frequency of 1 rad/s needs to be captured accurately, the higher BPOD modes need to be included.

4.3.3 Variation of Reynolds number

Figure 4.21 shows some of the eigenvalues of the 17-mode POD and BPOD models as the Reynolds number is increased. As in Section 4.2.4, we use the modes from the design condition ($Re = 2000$ in this case) and form the models using (4.3). Both POD and BPOD models have eigenvalues on the real axis very close to the origin, which remain stable and correspond to the slow evolution of the streamwise-constant structures. At each Re, the eigenvalues of both models move towards the right half of the complex plane and while the BPOD model always remains stable,

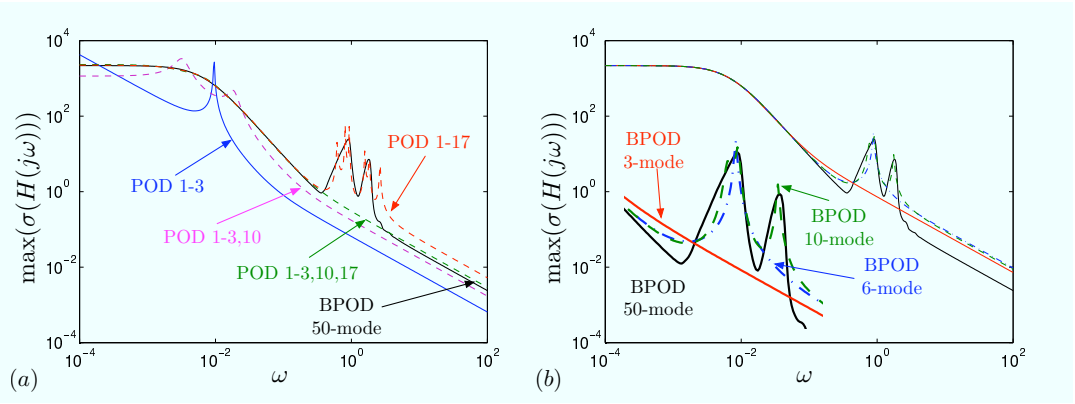


Figure 4.20: Singular value Bode plots for POD (a) and BPOD (b) models for the localized disturbance. The models are compared to a 50-mode BPOD model. The close-up in (b) shows that a six-mode BPOD model is needed to capture the larger resonant peak, and a ten-mode BPOD model captures both peaks.

the POD model first appears marginally stable at $Re = 2500$ and then unstable at $Re = 3000$. The effect of the eigenvalues that move to the right half of the complex plane is clearly seen in Fig. 4.21 (b). A model that includes modes 1–17 grows unstable at $Re = 3000$, showing that inclusion of modes that at design condition do not contribute to the overall dynamics can significantly deteriorate the performance of the model at off-design condition. This can also be seen from the frequency responses shown in the previous section — even at the design condition, the spurious high peaks correspond to marginally stable modes. Although stable, the 1–3,10,17 POD model is highly inaccurate at $Re = 3000$, with large successive peaks in the kinetic energy which decay very slowly, indicating the high sensitivity of those POD models which remain stable to a change in the Reynolds number. On the other hand, the three-mode BPOD model is still remarkably close to the full system.

It is important to note here that as the Reynolds number is increased, the nonlinearity will have a stronger effect on the development of the disturbance and the linearization may not be valid for higher Re in the first place. The comparison of the linear perturbation growth with a full nonlinear DNS solution is essential for a true validation of the models for control applications, since we may be modeling the linearized flow well, but the linearized flow may not be a good approximation to the actual flow. This comparison will be presented in Chapter 5.

4.3.4 Capturing of actuation

An important property of a reduced-order model is how well it captures the effects of the actuator in the original system, especially for models that are intended for developing controllers. In order for a reduced-order model to capture the effect of

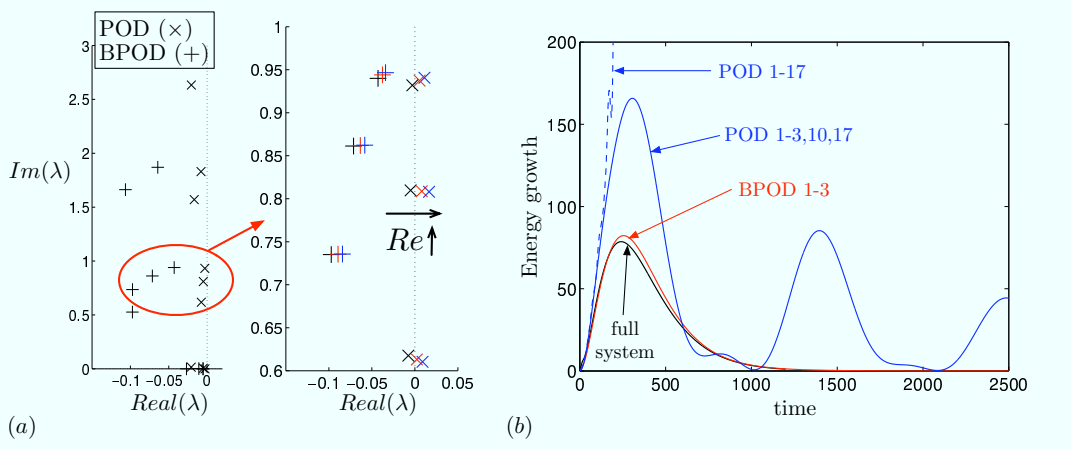


Figure 4.21: (a) The eigenvalues of 17-mode models at $Re = 2000$, $Re = 2500$, and $Re = 3000$. (b) The performance of the models at the off-design condition of $Re = 3000$.

an actuator, it is necessary at a minimum for the input term in the equations ($B_2 u_2$ in (3.11)) to be contained in the subspace used for projecting the equations. Note that here, even for the POD case, the effect of the actuator is partially included, since the dataset used for POD is generated by an impulsive input. One way to measure the degree to which the input 'directions' are captured by the modes used in the model is to compute the projection of the columns of the input matrix B_2 in (3.11) onto the basis modes. In the system we are considering here, $B = B_2$ is a single column vector, representing the initial disturbance given to the system (or actuation via a body force in the center of the domain). Fig. 4.22 shows the norm of the projection $\|P_r B\|/\|B\|$ of the POD and BPOD modes onto the input vector B, which is just the initial condition for each simulation. The inset in Fig. 4.22(a) again shows a geometric illustration of the non-orthogonal projection as seen in Fig. 2.1, where $\mathcal{N}(P)$ denotes the nullspace of a projection operator P. The balancing modes clearly capture the input direction with many fewer modes than POD: even very low-order models have a significant norm after projection, and in fact the norm of B after projection is almost always greater than the norm of B due to the non-orthogonal projection, as shown in Fig. 4.22. Any orthogonal projection such as P must satisfy $\|Px\| \leq \|x\|$, while for a non-orthogonal projection we may have $\|P'x\| > \|x\|$, which is the case for the first several BPOD modes. Clearly, the B matrix has a very small projection onto the POD modes unless many modes are taken, so it is impossible for very low-order POD models to capture the response of an actuator without introducing more modes (such as the B matrix itself, Krylov subspaces, or shift modes (Noack *et al.*, 2003)).

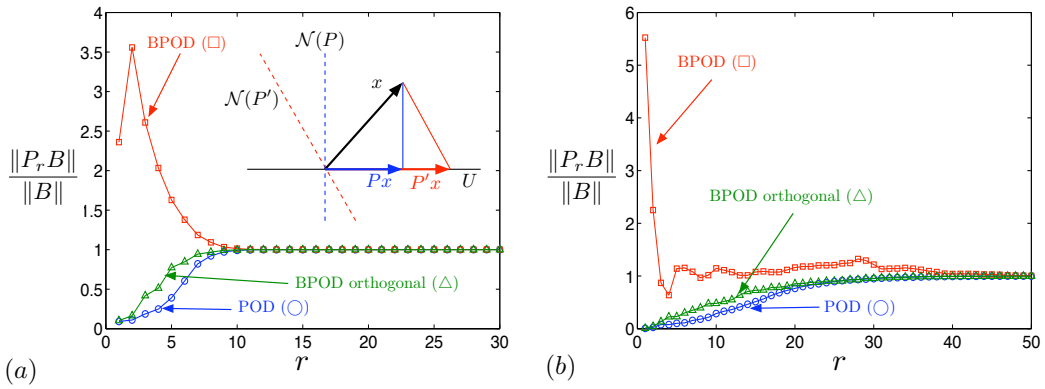


Figure 4.22: Norm of the projection of the B matrix (a single column vector) onto subspaces used for reduced-order models, (a) $\alpha = 1, \beta = 1$, (b) localized disturbance. The diagram in (a) illustrates the non-orthogonal projection used in BPOD.

4.3.5 Subspace comparison

The BPOD procedure uses both a different projection and a different set of modes in order to form reduced-order models, and we next look at a comparison of the two mode subspaces. A way to compare two subspaces is to compute $Tr(P_A P_B P_A) = T$, where P_A and P_B are the corresponding projection operators (Everson & Sirovich, 1995). The trace of the matrix $Tr(P_A P_B P_A)$ as a function of the subspace rank is shown in Fig. 4.23 for the five-mode and ten-mode output projections where P_A and P_B are the orthogonal projectors onto the POD and BPOD subspaces respectively. The value of the trace T is the same as the subspace rank r at low order, indicating very similar modes (see Figs. 4.15 and 4.17). For both POD and five-mode OP BPOD, modes four and five are a pair of structures, and including both modes from the pair brings the value of s to almost exactly 5. It is interesting to observe that for the five-mode output projection, $r = T$ exactly at $r = 5$, while for the ten-mode output projection $r = T$ at $r = 10$, and that above those values the value of the trace is lower than the rank. This can be explained by the fact that BPOD is attempting to approximate the output projection of the data of the given rank. It is interesting to note that the subspaces including the first three POD and BPOD modes are virtually identical, indicating that the non-orthogonal Petrov-Galerkin projection via adjoint modes makes the enormous difference that we have seen in the performance of the corresponding models. As mentioned above, the POD basis is the basis of the most controllable modes, and is indeed optimal in capturing a given dataset, but as we have shown, it can fail to capture the dynamics correctly.

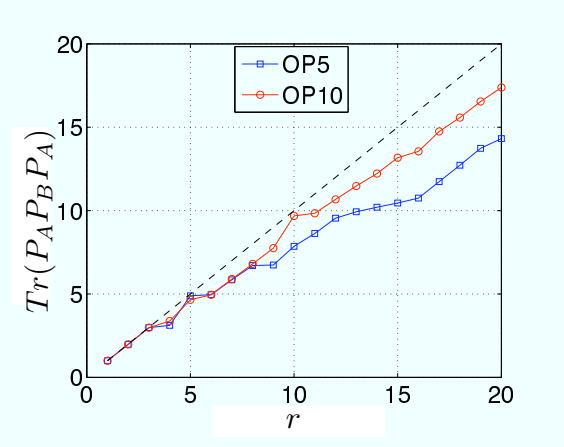


Figure 4.23: Plots of $\mathrm{Tr}(P_A P_B P_A)$ as a function of the subspace rank r for the localized perturbation for both output projections.

4.4 Alternative view of POD model performance

We next examine in more detail the performance of the POD model from Section 4.3.1 in the spirit of (Mezić, 2004); in particular, we represent the dynamical system as a network and utilize the horizontal-vertical decomposition described in that work to investigate the importance of the low-energetic modes 10 and 17 (recall that they contain 0.025% and 0.0074% of the total energy respectively). In order to introduce the concept of 'production lines' in dynamical systems which can be identified using graph representation, we first consider the very simple two-dimensional dynamical system:

$$\dot{x} = Ax, \tag{4.6}$$

where

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The solution of the system for initial condition $x_0 = [c\,;\,0]$ is

$$\begin{aligned} x_1(t) &= ct \\ x_2(t) &= c \end{aligned},$$

where c is a constant. The evolution of the state variable x_1 depends on x_2, while x_2 is independent and decouples. A graph representation of this system is shown Fig. 4.24, according to the theorem on horizontal-vertical decomposition given in Mezić (2004) — the lowest level is the second state, which 'feeds' the first state and thus determines its behavior. The dynamics of this system is very simple — x_1 grows without bound, and the 'production line' consists of x_2 'feeding' x_1. Horizontal-vertical decompositions such as the one shown here can be used for more complicated systems, as will be shown later in this section.

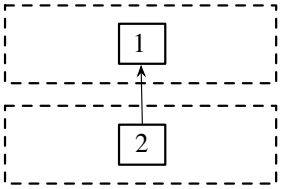
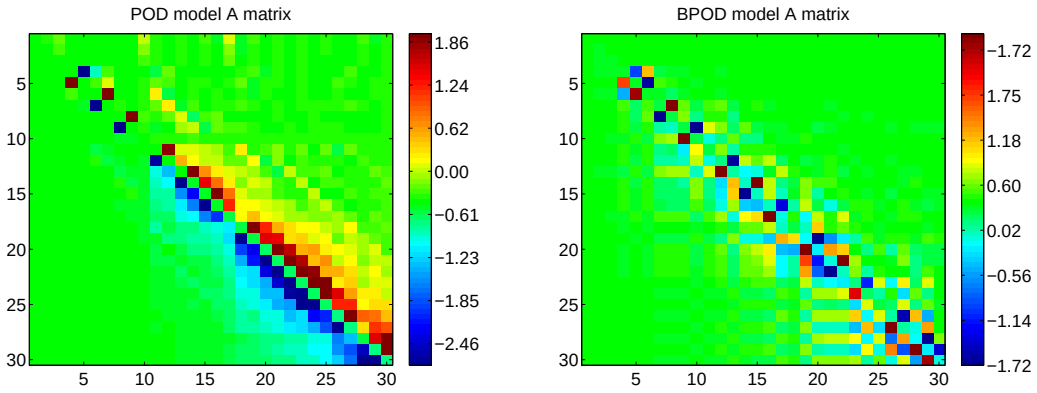
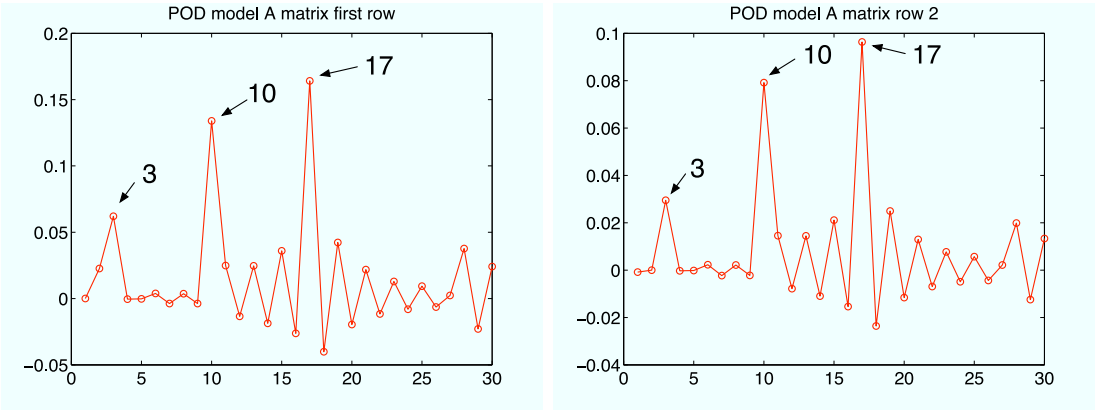


Figure 4.24: Graphical representation of dependence in the simple example in Eq. 4.6.



(a) (b)

Figure 4.25: (a) A plot of the values of POD model A matrix entries. (b) A plot of the values BPOD model A matrix entries. See text for discussion of the intermodal coupling.



(a) (b)

Figure 4.26: The first two rows of the 30-mode A matrix for the POD model (the matrix in Fig. 4.25a). The growth of the first mode (a) depends strongly on modes 2,3,10,17, and the growth of the second mode (b) depends strongly on modes 3,10,17. The contributions of the other modes, which come in pairs and correspond to traveling structures, cancel themselves approximately.

	1	2	3	10	17
1	1.1949e-05	2.2610e-02	6.2066e-02	1.3396e-01	1.6413e-01
2	-8.2676e-04	-2.0015e-05	2.9477e-02	7.9151e-02	9.6336e-02
3	-1.6218e-04	-2.1053e-03	-1.0597e-04	4.4604e-02	6.3601e-02
10	-3.5614e-05	-5.7532e-04	-5.0699e-03	-6.6598e-03	1.8059e-02
17	-1.3563e-05	-2.2811e-04	-2.7021e-03	-2.5312e-02	-4.9944e-02

Figure 4.27: A matrix for model 1–3,10,17. Note that all the entries below the diagonal are negative — they correspond to negative feedback, while the entries above the diagonal are positive — they denote growth. The largest entries are the coefficients of modes 10 and 17 in the equation for the first mode.

We next consider the POD model consisting of modes 1-3,10,17, discussed in Section 4.3.1. The entries of the A matrices of POD and BPOD models for the perturbation from Section 4.3.1, each computed using the 30 leading modes, are shown in Fig. 4.25. The strong dependence of the leading POD mode evolution on modes 2,3,10,17 can also be seen in Fig. 4.26, which shows the first two rows of the POD model A matrix. On the other hand, there are no significant coefficients of higher modes in the BPOD model. Fig. 4.25 indicates that the oscillating mode pairs in the POD model are strongly coupled, as expected, and they are coupled only amongst themselves, which is apparent from the figure, as green color indicates zero or values very close to zero. The streamwise-constant modes 1,2,3 and 10 are coupled only amongst themselves as well. In particular, modes 1-3 strongly depend on modes 10 and 17 (the dependence of the first two modes is clearly shown in Fig. 4.26). The coupling of mode 17 to some of the oscillating modes may be related to its importance in capturing correctly the dynamics of the full system. Fig. 4.27 shows the entries of the A matrix for the POD model with the five important modes included. A simple order-of-magnitude analysis of mode coefficients can be used to draw an approximate graph for the interconnection of the key model in the model. Following Mezić (2004), a network can be represented by a matrix M, whose elements are given by

$$\begin{aligned} M_{ij} &= \frac{1}{l_i} \quad \text{if} \quad |A_{ij}| > \epsilon \\ M_{ij} &= 0 \quad \text{otherwise.} \end{aligned} \tag{4.7}$$

Here l_i is the number of non-zero entries in a matrix row. We note here that in Mezić (2004) the matrix M is defined as a *stochastic matrix*, $\epsilon = 0$, and the general case of a system with uncertainty is considered. In order to identify a 'production line' in the model, using an order-of-magnitude approximation, and picking $\epsilon = 2 \times 10^{-2}$, we may neglect all negative feedbacks on and below the diagonal in Fig. 4.27, except for the dependence of mode 17 on mode 10 and on itself. The matrix M for the

POD model with all five modes is then:

$$M_5 = \begin{bmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix},$$

while the matrix for a three-mode model excluding modes $2, 3$ would be represented by

$$M_3 = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \\ 0 & 1/2 & 1/2 \end{bmatrix},$$

Of course, since no terms of the A matrix shown in Fig. 4.27 are exactly zero, this is a crude approximation. Moreover, in the actual model, the negative feedbacks in the lower-diagonal part and on the diagonal of the matrix ultimately cause the energy to decay — were it not for these feedbacks, the solution would grow without bound, which is unphysical. Therefore, this picture, which excludes that important feature of the dynamics, is meant only to illustrate qualitatively the energy growth mechanism in the model, but not the complete dynamics. The upper-diagonal structure of the matrix corresponds to an energy transfer pathway, similar to the one described by Eisenhower & Mezić (2007). One can now draw an approximate graph for the interconnection of the key modes in the model (Fig. 4.28). The key feature of the model is the almost purely vertical structure — the low-energy modes are feeding the leading mode with energy, and a 'production line' passes through modes 2 and 3 from the tenth and seventeenth mode to the first mode. Note that modes 10 and 17 are on the same horizontal level, since there is a two-way connection between them, i.e. there is significant dependence of one on the other.

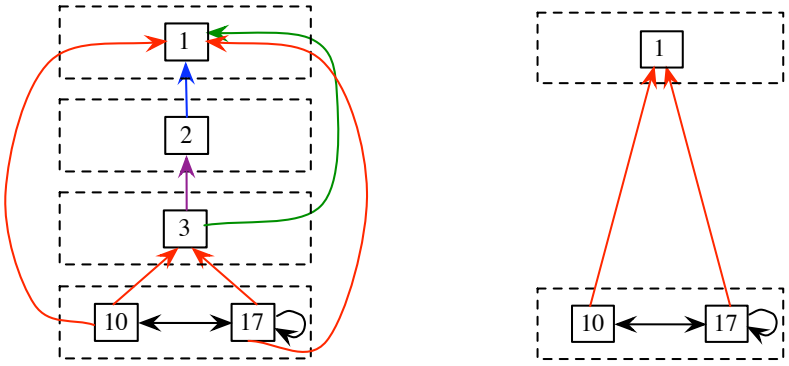


Figure 4.28: Left: Horizontal-vertical decomposition for model including modes 1,2,3,10–17, right: horizontal-vertical decomposition for model including modes 1,10,17.

In order to examine the importance of the tenth and seventeenth modes, we exclude modes 2 and 3 from the model, since as Figs. 4.25 and 4.26 show, their influence on the first mode is much smaller. We would expect the dynamics to

remain 'structurally' similar to the model including all five important modes — mode 1 will still grow, followed by decay. Fig. 4.29 shows the performance of the models and the coefficient of the first mode for the different models. We note that, although most negative feedbacks and diagonal elements in the A matrix of the model have been neglected for the graph decomposition, they have been included in the model, as otherwise there would be unbounded growth of the energy. It turns out that the model with modes $1, 10, 17$ is inaccurate in terms of capturing the growth, but unlike the model with modes 1–3, the dynamics is indeed qualitatively similar to the dynamics of the full system — growth followed by monotonic decay once the first mode grows large enough, as opposed to the unphysical oscillations in the energy captured by only the first three modes.

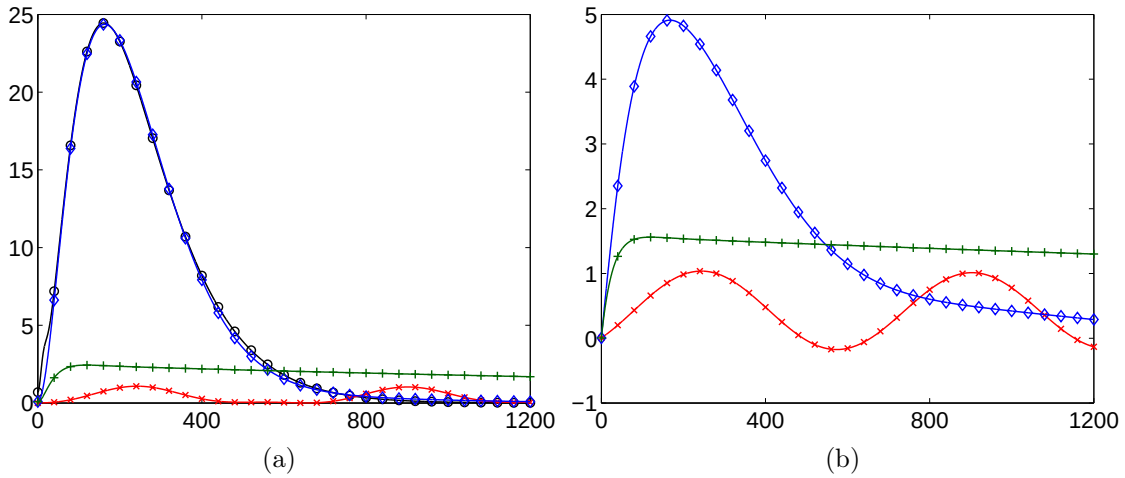


Figure 4.29: (a) The performance of POD reduced-order models in capturing the energy of the perturbation. Symbols: full simulation (◯), POD 1–3 (×), POD 1–3,10,17 (⋄), POD 1,10,17 (+). (b) The coefficient of the first mode. Symbols same as on left, full simulation not plotted.

This analysis offers another view of the POD and BPOD models — a key feature of the input/output behavior, shown by the 'production line', is not captured correctly in a reduced-order model unless the important POD modes are included. From the plot of the entries of the matrix representing the POD model dynamics in Fig. 4.25, we have seen that many of the modes with energy content that is high compared to the energy content of modes 10 and 17 are not coupled to the first three modes, and therefore do not contribute to the overall energy growth — they represent traveling structures, as described earlier in this chapter. The qualitative behavior of the model is captured by a simplified model that retains the same vertical structure, but does not include the second and third modes, indicating the importance of the low-energetic modes 10 and 17, although the result is much different qualitatively. On the other hand, the important input/output dynamics is captured in the dynamically much simpler BPOD modes, and just the leading modes are able to reproduce the energy growth of the perturbation. Of course, it should be kept in mind that neglecting the negative feedbacks and most of the

diagonal entries of the A matrix of the POD model when representing the model as a graph with vertical structure is a major approximation, and the present approach can only be used to gain qualitative insight into the dynamics of the system.

4.5 Summary

In this chapter the theoretical and computational methods described in the previous two chapters were applied to both 1-D and 3-D localized perturbations to linearized channel flow. Snapshot-based balanced truncation with output projection (BPOD) was thoroughly compared to POD and exact balanced truncation and it was demonstrated that BPOD produced models with superior performance. The specific contributions of the work presented in this chapter are:

- BPOD and exact balanced truncation were computed for a 1-D system for which the latter method is feasible, and a detailed comparison similar to the one in Rowley (2005) was done. It was shown that the modes and HSVs are captured well, and that a fairly low rank of output projection is sufficient to capture well the dynamics of the system.
- The first calculation of balanced truncation models for a large-scale 3-D system was performed — a localized perturbation on a $32 \times 65 \times 32$ computational grid, corresponding to 133,120 states for the full (v, η) system.
- Model performance such as impulse response, frequency response and capturing of actuation in the models, was investigated in detail, and BPOD modes were consistently found to outperform POD at comparable truncation rates. It was also found that the BPOD models can perform well at an off-design value of the Reynolds number, while POD models have a much smaller range of validity with respect to the Reynolds number. In addition, a striking example of important dynamics being concealed by low-energy POD modes was found (1–3,10,17 POD model in Section 4.3.1).
- A comparison of the subspaces of balancing and POD modes was performed, indicating that the two bases are quite similar, and that the non-orthogonal Petrov-Galerkin projection using adjoint modes is responsible for the difference in the models' behaviors.
- A more detailed investigation of the 1–3,10,17 POD model inspired by the recent work of Mezić (2004) reveals a simple vertical structure which provides qualitative insight about the model dynamics.

Chapter 5

Feedback control of transitional channel flow

The main application of the reduced-order models discussed in the previous chapters is the design of linear controllers using reduced-order models. As discussed in Chapter 1, it has been argued that the growth of perturbations is governed by an essentially linear process. It is therefore expected that the energy growth may be suppressed using the tools of linear control theory. Specifically, while the perturbation is small, the linear energy growth dominates, and a linear controller should therefore be able to prevent transition. While the results presented so far have been for the linearized system only, at the end of this chapter we apply the linear compensator designed using balanced reduced-order models to nonlinear DNS of channel flow.

The control design process we adopt in this work has four main stages, starting with the choice of the control system setup and the simulations to obtain reduced-order models. Once the models are obtained, a compensator that consists of a controller and a state estimator is first designed at the reduced-order model level. The computational cost of this part of the process is negligible, and typically the control designer tries many controllers until the desired result is achieved. After that, the controller is tested first on the full-scale linear system (linearized Navier-Stokes equations), and then on the nonlinear system (full nonlinear DNS).

In this chapter, we first demonstrate the use of balanced reduced-order models, obtained using the Eigenstate Realization Algorithm (ERA), introduced in Section 2.3.4, in reducing the energy growth of a localized perturbation near the channel wall in linearized DNS simulations. The size of the computational box and the centerline Reynolds number used in this chapter are the same as in Kim *et al.* (1987), for both linearized simulations and nonlinear ones used in the last section. ERA was used due to the tremendous computational savings over BPOD. In Section 5.2 we find that, using a localized body force as actuation, a compensator

designed to minimize a single velocity output is also able to reduce the perturbation energy significantly in linearized DNS. The reduction of energy growth in this case is comparable to the reduction achieved by using the entire flow field energy as the output of interest. An estimator based on two velocity measurements is incorporated into the compensator. Compensators using wall blowing/suction as actuation are found not to perform as well, but still achieve energy growth reduction (Section 5.3).

Finally, the compensators are also tested on nonlinear simulations with the initial amplitude of the perturbation high enough for transition to happen in the uncontrolled case (Section 5.4). The Re_τ in the nonlinear case for these computations is lower than in the simulation of Kim *et al.* (1987) due to the imposed constant mass flux. Since the flow is not expected to transition for simulations in which a successful controller is applied and the dynamics is expected to remain very close to the linear dynamics, we are not concerned with the details of the turbulent flow once transition becomes irreversible, and only examine the initial stages of the simulations. It is found that re-laminarization is possible using a body force actuator only if the streamwise separation of the perturbation and the actuator is small.

5.1 Control design and implementation

There are four steps in the control design process:

1. Select system inputs and outputs, perform simulations and obtain reduced-order models.

2. Design the compensator (i.e., compute the estimator and controller gains) at the reduced-order model level and test its performance.

3. Test the compensator at the linearized DNS level — if the reduced-order model captures the linearized dynamics well, the compensator should perform well.

4. Test the compensator at the nonlinear DNS level — if the previous steps in the design are successful, and if the nonlinear dynamics remains in the neighborhood of the fixed point, the compensator should perform well; however, there are no guarantees for this.

In the previous chapters we described in detail the first step, which is often the most complex one. In this section we describe the second step of the process — the compensator design — in some detail, before presenting the results for feedback control of a localized perturbation in the subsequent sections.

5.1.1 Inputs and outputs

Here we continue the discussion of inputs and outputs started in the previous chapter, now with the focus on the inputs and outputs for closed-loop control. In Section 4.3 (also see Ilak & Rowley, 2008*b*), a localized actuator/disturbance was the only input to the system and the output y_1 was the full state, approximated in the reduced-order models using output projection. In this work, we consider models that incorporate both the disturbance we wish to suppress and the actuator as inputs and the measurement and full-state output as outputs. The model is therefore MIMO (multiple-input-multiple-output), and the B and C matrices are given by:

$$B = \begin{bmatrix} B_1 & B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}. \tag{5.1}$$

The choice of inputs and outputs is guided by practical considerations. While channel flow is an idealized canonical case, the actuators (B_2), outputs ($y_1 = C_1 x$) and measurements for the estimator ($y_2 = C_2 x$) are designed to be close to realistic mechanisms. The two different actuators chosen are a body force near the channel wall and a wall blowing/suction actuator, and the performance of each one is tested separately.

Wall-normal and streamwise velocity were measured at different locations both as outputs to be minimized and as measurements for the estimator. The models were computes using ERA, and thus the localized measurements could be extracted from single grid points without the need to model them as 3-D fields for adjoint simulations. If it is desired to compute modes, the velocity measurement may be represented by a smooth 3-D Gaussian function $C(x, y, z)$ centered at the point of interest in the flow (Bagheri *et al.*, 2009*c*), so that the corresponding output of the system is computed as:

$$y(t) = \int_V C(x, y, z) v(x, y, z, t)\, dV, \tag{5.2}$$

where V is the volume of the computational box. The same function used to model the perturbation and the actuator may be used to model localized measurements if it is desired to run adjoint simulations and compute balancing and adjoint modes.

While the measurement of perturbation energy would not be available in an experiment, it is available in DNS simulations, and may be included in models as the norm of the output y_1 using output projection, as shown in Section 4.3. An optimal controller will then minimize the energy of the perturbation, as will be shown in Section 5.1.3. In this chapter we compare the performance of controllers that minimize a single velocity signal downstream of the initial perturbation to controllers that minimize the perturbation energy, as we may consider the latter to be the most desirable scenario, although impracticable in an experiment.

Since the variables that fully define the flow are v and η, a measurement of v would be convenient since the computation of the other two velocity components is not necessary. The input/output behavior at different wavenumber pairs was investigated by Jovanović & Bamieh (2005), and it was shown that the streamwise velocity response is very well correlated to the total energy in the flow field. Moreover, the largest input/output norm was computed for the case of a perturbation in the wall-normal direction and measurement of streamwise velocity. Therefore, it is expected that the compensators designed with streamwise velocity as output have better performance than those designed with other velocity components as output to be minimized.

5.1.2 Treatment of the inhomogeneous boundary conditions for wall blowing/suction actuation

While the implementation of wall blowing/suction into a DNS simulation is fairly straightforward since it only involves setting non-zero boundary conditions in the wall-normal direction, the computation of the corresponding reduced-order models for control design requires some care. A standard method for treatment of inhomogeneous boundary conditions when designing the control gains is known as *lifting* (see, for example, Högberg *et al.*, 2003). In this approach, the solution is represented as the sum $x = x_h + x_p$, where the homogeneous solution x_h satisfies homogeneous boundary conditions ($u = v = v_y = 0$ at walls), and x_p is a particular solution satisfying the desired inhomogeneous boundary condition. In our reduced-order models, we expand x_h in terms of modes that satisfy the homogeneous boundary conditions, and write $x_p = Z\phi$, where $\phi(t)$ is the amplitude of the wall blowing/-suction, and Z is chosen to be a steady-state solution of the linearized equations with boundary condition $v = 1$ at the wall, as in Högberg *et al.* (2003). For a single wavenumber pair this solution is easily computed using a spectral collocation method. For a localized actuator on the channel wall, Z may also be computed by starting the simulation impulsively with the wall blowing and suction actuator as an initial condition and letting the solution reach steady state, although the first method is much faster.

After augmenting the full state with the control input ϕ, the system in (3.11) becomes:

$$\begin{aligned} \begin{bmatrix} \dot{x}_h \\ \dot{\phi} \end{bmatrix} &= \begin{bmatrix} A & AZ \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_h \\ \phi \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u_1 + \underbrace{\begin{bmatrix} -Z \\ I \end{bmatrix}}_{B_2} \dot{\phi} \\ y &= \begin{bmatrix} C & CZ \end{bmatrix} \begin{bmatrix} x_h \\ \phi \end{bmatrix}, \end{aligned} \tag{5.3}$$

where the input is now $u_2 = \dot{\phi}$, where the dot indicates a time derivative. Some of the alternative approaches to lifting for treating inhomogeneous boundary conditions

are actuation modes (Noack *et al.*, 2004) or using a weak formulation (Camphouse, 2005). In the former approach, the actuation is not captured by the reduced-order models in a systematic fashion, rather it is treated separately from the system dynamics, while in the latter approach the modes have inhomogeneous boundary conditions, which may cause difficulties in applying feedback control (Camphouse, 2005).

Only the homogeneous part of the augmented system (5.3) is included in the reduced-order models. When reduced-order models are obtained from modes in usual balanced truncation, it is of particular importance to work with modes that have homogeneous boundary conditions, since expressing model solutions as a linear combination of modes would result in incorrect boundary conditions for the model solution. The input for the homogeneous part of the augmented system in (5.3) has inhomogeneous boundary conditions, and using it as an initial condition results in the first snapshot having an inhomogeneous boundary condition, which would be undesirable for the computation of modes, since modes are linear combinations of snapshots. Spatial filtering of the initial condition for the impulse response of Z is thus required. One approach is to use an ad-hoc Laplacian filter (Ilak & Rowley, 2008*a*), or, alternatively, the DNS solver itself may be used as the filter (Bagheri *et al.*, 2009*a*), since advancement of the initial condition by a small DNS time step will enforce the correct boundary conditions. The latter approach is also employed if a particular form of the localized perturbation or actuator needs to be placed near the wall with appropriate boundary conditions, as later in Section 5.2.

5.1.3 LQR design

A Linear Quadratic Regulator (LQR) was used for feedback control. LQR design is described in detail in standard references on control theory (for example Stengel, 1994), and here we only outline the essence of the method. In control-theoretic terms, an LQR regulator is a proportional controller, resulting in a control law given by $u = -Kx$, where the gain matrix K is computed so that it minimizes the objective function:

$$J = \int_0^\infty (x^T Q x + u^T R u)\, dt. \tag{5.4}$$

This type of controller is called a regulator, since it regulates the departure from $x = 0$, in which case the first term of the integrand in the cost function would be zero. The control input may be confused with the streamwise velocity u, as the use of the letter is established in both fluid mechanics and control theory with different meanings. In this chapter it is indicated in the text what is meant each time the symbol is used and it is not obvious from the context. Given the state-space system and the weight matrices Q and R, the gain:

$$K = R^{-1} B^T P \tag{5.5}$$

is computed from the solution of an algebraic matrix Riccati equation given by [1]:

$$A^T P + PA - PB_2 R^{-1} B_2^T P + Q = 0. \tag{5.6}$$

The weight matrix Q is typically chosen to be $Q = C_1^T C_1$, since $x^T C_1^T C_1 x = \|y_1\|^2$, so that the cost function is then:

$$J = \int_0^\infty (y_1^T y_1 + u^T R u)\, dt. \tag{5.7}$$

This new cost function now minimizes the sum of the norm of a vector of a small number of outputs of interest (in practice these are the outputs we are able to measure in an experiment), and the control input over time. We note that the controller penalty matrix R is a scalar when we have just one control input, as is the case for all simulations in this thesis. We also note that, if the output $C_1 = I$, i.e. it is the full state, the controller will minimize the 2-norm of the total flow field, which in the case of fluid systems corresponds to the total energy. In the reduced-order models in which output projection is used to approximate the energy of the full field, the weight matrix becomes $Q = \Theta_s \Theta_s^T$, where Θ_s are the first s POD modes.

Computation of the control gain matrix K for the full system (4.1) using (5.6) is clearly not feasible for 3-D fields in transitional channel flow, as the required matrices cannot even be stored in memory for a large system such as a linearized DNS. This is one of the main reasons for using reduced-order models to design the controllers, as the gains for the models are computed at very low computational cost. Instead of the matrices (A, B, Q), in this case we use:

$$A_r = \Psi_r^+ A \Phi_r, \quad B_{2,r} = \Psi_r^+ B_2, \quad Q_r = \Phi_r^T C_1^T C_1 \Phi_r. \tag{5.8}$$

This computation may be done using, for example, the standard `lqr` command in MATLAB. These gains are then applied to simulations of the full non-linear system.

For the case of wall blowing/suction actuation the gains are designed for the augmented reduced-order model, with the matrices:

$$\hat{A} = \begin{bmatrix} \Psi_r^+ A \Phi_r & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{B}_2 = \begin{bmatrix} -Z \\ I \end{bmatrix}, \quad \hat{Q} = \begin{bmatrix} Q_r & Q_r Z \\ Z^T Q_r & Z^T Q_r Z \end{bmatrix}. \tag{5.9}$$

The last row and column of the matrix $\hat{Q}$ (two rows and columns if there is actuation on both walls) take into account the effect of the wall blowing/suction. Thus, the feedback control gains in this case are obtained by solving a $(r+1) \times (r+1)$ or $(r+2) \times (r+2)$ Riccati equation.

As described by Högberg *et al.* (2003), the contribution of the homogeneous part of the flow to the control gains at the walls needs to be subtracted from the

[1] The algebraic Riccati equation is a special case of the differential Riccati equation, which also contains a time-dependent term. For Linear Time-Invariant (LTI) systems such as those considered in this thesis, and for *infinite time-horizon*, i.e. integration until $t = \infty$ as in (5.4), the Riccati equation reduces to the form given here (see, for example, Stengel, 1994).

computed values. The gains may also be written as $\hat{K} = [K_h \quad K_\phi]$, and, since:

$$\dot{\phi} = -K_h x + K_h Z\phi - K_\phi \phi = -K_h x_h - (K_{end} - K_h Z)\phi, \tag{5.10}$$

we see that if the gains are to be used in a DNS, the last gain needs to be modified by subtracting $K_{end} - K_h Z$ from the value obtained. We emphasize again that the control input u_2 is the time derivative of the wall blowing and suction, which is integrated in time. The time advancement of the control input is computed at each step in the DNS simulations, using the same numerical scheme as the one used in advancing the DNS solution in time.

5.1.4 LQE design

The control input term $u_2 = -Kx$ may only be computed if the state x is known. The states of the system are often not available (i.e., cannot be measured), or, in the case of fluid simulations, the dimension of the state is too large for the computation of the control input (also known as *full-state feedback*) to be practical. Instead, often an estimator (also known as *observer*) provides an *estimate* of the state, which is then fed back to the controller for computing the control input.

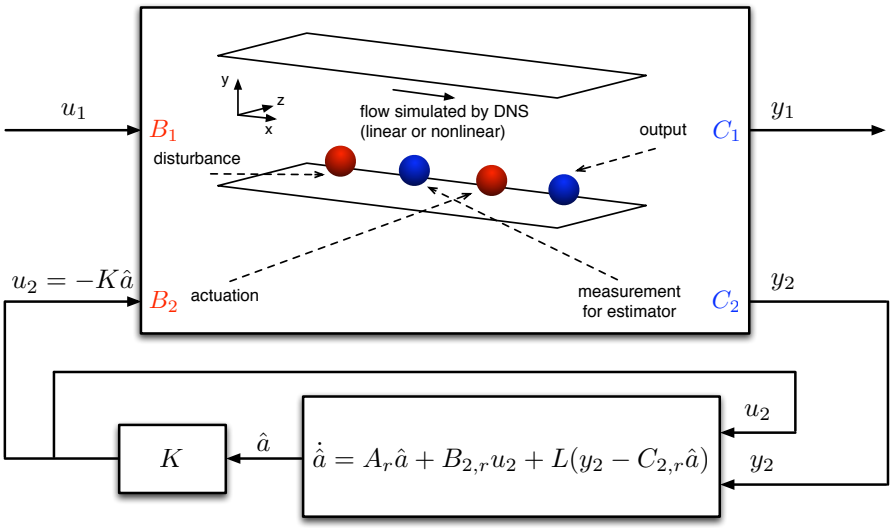


Figure 5.1: A schematic of the control setup with reduced-order estimator in the loop. The configuration of inputs and outputs in this figure is representative of the configurations we use later in this chapter. The hat indicates estimated state.

The estimator is essentially a system with the same dynamics as the original system, but with the error between the true measurement and the estimated measurement as the quantity that needs to be minimized. Thus, the problem of estimation is dual to the problem of optimal control, and an optimal estimator, also known as a Linear Quadratic Estimator (LQE) is designed by solving Riccati equations as for LQR problem.

If we introduce noise both in the state equation (also known as 'process noise') and in the measurements that the estimate is based on, the system becomes:

$$\begin{aligned} \dot{x} &= Ax + B_1 u_1 + B_2 u_2 + Gw \\ y_1 &= C_1 x \\ y_2 &= C_2 x + \alpha v. \end{aligned} \tag{5.11}$$

Here w is the process noise, and the matrix $Q = ww^T$ is known as the process noise covariance matrix. G is the sensitivity of the state to the process noise, v is Gaussian measurement noise (not to be confused with the wall-normal velocity component v) with amplitude α. The Riccati equation to be solved is now given by:

$$AP + PA^T - \frac{1}{\alpha^2} PC_2^T C_2 P + GQG^T = 0. \tag{5.12}$$

The corresponding estimation gains given by:

$$L = -\frac{1}{\alpha^2} PC_2^T, \tag{5.13}$$

minimize the error between the measured and estimated states.

We next discuss the choice of G and Q. If we set $G = I$ and $Q = B_1 B_1^T$, $GQG^T = B_1 B_1^T$. In Doyle & Stein (1979), it is shown that robustness of the estimator is increased by adding 'fictitious noise' in Q, so that:

$$Q = Q_0 + qBVB^T, \tag{5.14}$$

where Q_0 is the noise covariance for the plant, and V any positive-definite symmetric matrix. In our case, $V = I$, $Q_0 = 0$ and $q = 1$, which would correspond to no noise in the actual plant. The final Riccati equation we solve in order to design the estimator is thus:

$$AP + PA^T - \frac{1}{\alpha^2} PC_2^T C_2 P + B_1 B_1^T = 0, \tag{5.15}$$

α being the only tuning parameter. A low value of α corresponds to less noise in the measurement, and therefore better estimator performance.

A LQR controller that incorporates an LQE estimator is also known as a Linear Quadratic Gaussian (LQG) compensator. By the separation principle (see, for example, Friedland, 1986), the estimator and the controller are designed separately, and if both are stable, the entire compensator will be stable. We note that it has been shown (Doyle, 1978) that there are no guaranteed stability margins for a system with an estimator, i.e., in some cases the resulting closed-loop system may become unstable under arbitrarily small perturbations, which should be kept in mind by the control designer. A schematic of the full compensator is shown in Fig. 5.1.

5.2 Feedback control of linearized flow using localized body force as actuation

We next demonstrate the performance of the controllers designed using the procedure described above. The main goal is to reduce the energy growth of a localized perturbation placed near the wall. The perturbation is very similar in shape to the one described in Section 4.3, although it is now placed near the lower wall at the streamwise location $x = \pi$ (see Fig. 5.3a). The resolution of the linearized DNS was $64 \times 65 \times 32$. Although the energy growth of this perturbation is considerably lower than the typical growth of optimal perturbations (see, for example, Butler & Farrell, 1992), it was still observed to transition in nonlinear simulations, as will be shown in Section 5.4. In this section and next section, we design compensators for the linearized flow and test them in linearized DNS simulations.

That the exact form of the perturbation B_1 is:

$$v(x, y, z, 0) = A \left(1 - \frac{r^2}{\alpha^2}\right) e^{(-r^2/\alpha^2 - (y-y_0)^2/\alpha_y^2)}, \tag{5.16}$$

where $(x_c, 0, z_c)$ are the coordinates of the center of the computational domain and $r^2 = (x - x_c)^2 + (z - z_c)^2$. The wall-normal vorticity is zero. We note that the term $(\cos(\pi y) + 1)$, added in Eq. (4.5) in order satisfy the boundary conditions $v(\pm 1) = v_y(\pm 1) = 0$ for a perturbation in centered in the channel mid-plane, places the perturbation away from the wall even if y_0 is chosen very close to the wall. Therefore, this term was not included for the perturbations used in this chapter, and filtering using the DNS was employed instead, i.e., the perturbation given by (5.16) was advanced by a short linearized DNS time step, and the result was used as the perturbation B_1 in the simulations.

In this section we use a localized body force actuator, identical in shape to the disturbance B_1, and at the same distance from the wall, but placed downstream, at $x = 2\pi$, as shown in Fig. 5.3(b). Figure 5.2 shows the schematics of the two setups used in this section. In studies of modeling and control of spatially developing boundary layers (see Bagheri *et al.*, 2009*b*), actuators are typically located downstream from the perturbation, and we use a similar setup in this study. We remark again (recall Section 4.3) that due to the periodicity of the computational box in the streamwise direction, once the perturbation 'hits its own tail', the dynamics is considerably different from the dynamics of a perturbation in an infinitely long box. In particular, the energy growth is larger due to the streamwise structures that develop sooner than they would have developed in a longer domain. Thus, the physical equivalent of this setup is a periodic array of such perturbations and actuators.

Streamwise and/or wall-normal velocity signals are used for both the estimator measurements y_2 and as the output to be minimized by the controller y_1, and the

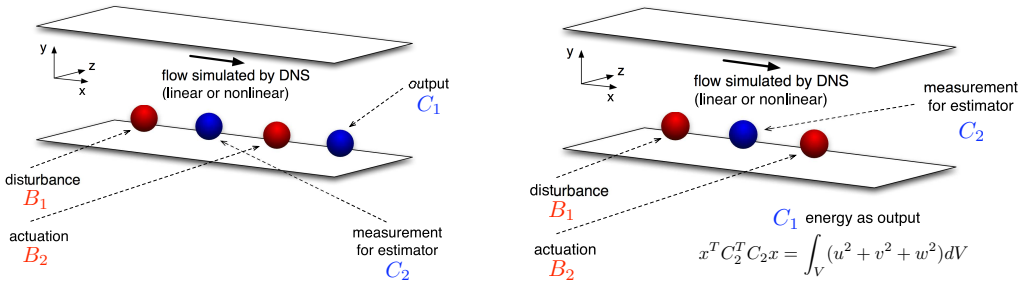


Figure 5.2: The control system setup for the cases considered in this section: localized body force actuation and localized measurement as output y_1 (a) and localized body force actuation and full field output y_2, corresponding to minimizing perturbation energy (b).

energy of the full field is also used as the output y_1 for some cases. As discussed earlier, the latter choice is not practically realizable, but it does provide a useful comparison with the cases in which a single signal is used as output y_1.

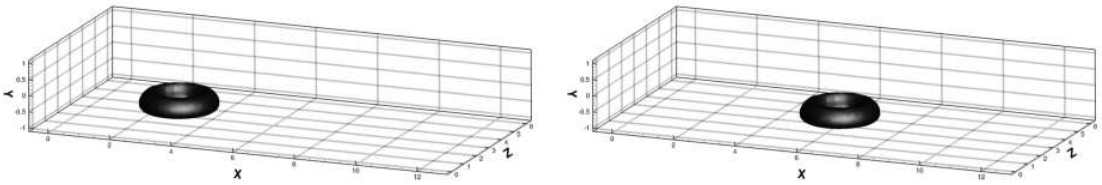
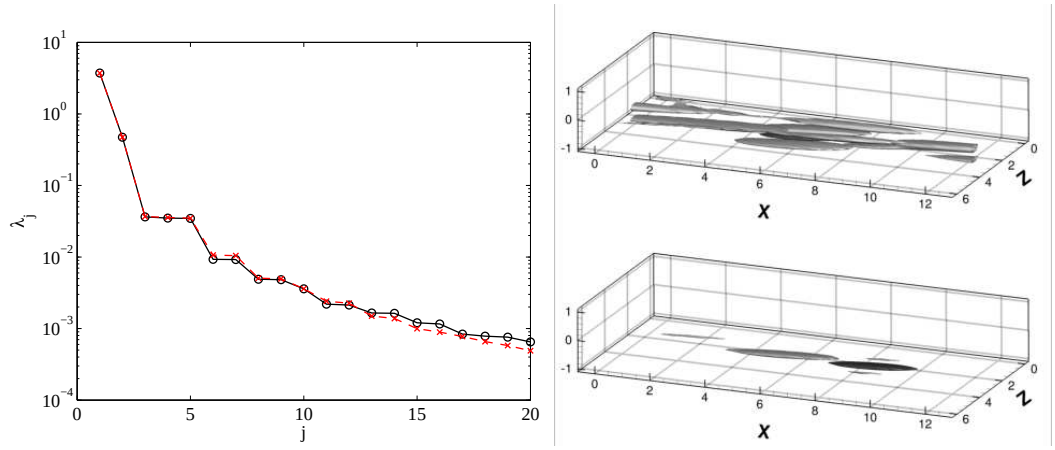


Figure 5.3: (a) The disturbance (B_1 matrix). (b) The body-force actuator (B_2 matrix).

5.2.1 POD modes

In order to select the location of measurements, some knowledge of the dynamics of the modeled system is needed. Measurements should be placed near highly controllable directions, since these directions are most easily excited by perturbations, and thus the measurement would extract most information. As shown in Chapter 2, the POD modes of the impulse response of the perturbation are the most controllable modes of the system. Therefore, we examine them in order to determine suitable velocity probe locations. These modes will also be needed later for output projection.

The POD modes taken here include data from impulse responses of both the disturbance B_1 and the actuator B_2. Although the energy evolution in time is identical for the actuator and the disturbance, the actual snapshots are different due to the spatial shift, and it is thus necessary to take into account both sets.



(a) (b)

Figure 5.4: (a) A comparison of eigenvalues for the disturbance only multiplied by two (X) and for POD with both disturbance and actuator snapshots included (◯). (b) Isosurfaces of the streamwise velocity of the sixth mode for POD of the impulse response of just B_1 (top) and for impulse responses of both B_1 and B_2 (bottom).

The leading modes are identical to the POD modes of B_1 only, as indicated by the eigenvalues in Fig. 5.4a, but some of the higher modes are not, which is seen from the eigenvalues plotted in Fig. 5.4. Since the dynamics here is similar to the dynamics of the perturbation described in Section 4.3, the modes were again computed using uneven snapshot spacing, with a total of 2181 snapshots during the time interval (0,3200). In Fig. 5.4 we show the a comparison of the POD eigenvalues when just the perturbation impulse response snapshots are included and when both impulse responses are included. The POD eigenvalues for B_1 only are multiplied by two in Fig. 5.4, since the total energy in the snapshots is exactly half of the energy of the two sets of snapshots taken together. The leading mode (see Fig. 5.5) is streamwise constant, and the 1-D profiles of the streamwise and spanwise velocities at $z = \pi$ are shown. Two measurement locations for the estimator are selected, one near the wall at wall-normal coordinate $y = -0.96$, referred to hereinafter as y_{21}, and one where the velocities have maximum absolute value, which for both the streamwise velocity u and the wall-normal velocity v is at approximately $y = -0.60$, referred to hereinafter as y_{22}[2]. We also use a signal at the location $(x, y, z) = (3\pi, -0.60, 3\pi/2)$ for the output y_1 when not using the full field.

5.2.2 Model performance - minimizing streamwise velocity

We next discuss the performance of compensators designed for the reduced-order models. Two measurements were used for the estimator, as described above. We

[2]Again, the notation may be potentially ambiguous due to the use of the same letter as the wall-normal coordinate in fluid mechanics terminology and output in the language of control theory.

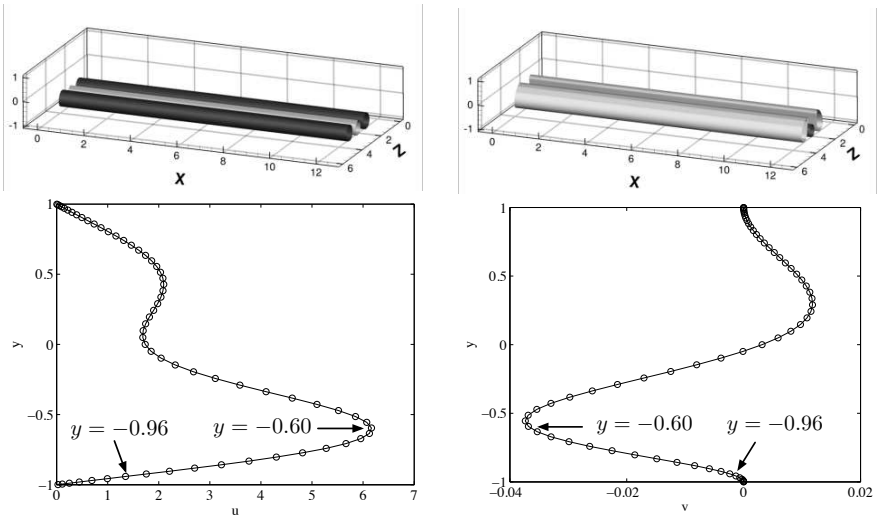


Figure 5.5: The leading POD mode for snapshots from both the localized perturbation and the localized actuator. The 1-D profiles are shown for the streamwise (left) and wall-normal (right) velocity, indicating the points where the velocity outputs are chosen.

first consider a single signal of streamwise velocity as y_1, since we may expect streamwise velocity to be well-correlated to the perturbation energy, which we desire to minimize. The transfer function from the inputs to the outputs is a 3×2 matrix. In this section and the next section, we refer to the uncontrolled linearized DNS (impulse response the system with B_1 as input matrix) as the 'full system' when compared to reduced-order model outputs or controlled runs. We note that in all simulations the input signal $u_1(t)$ is just $u_1(t) = \delta(t = 0)$, i.e., we start the system with B_1 as initial condition. We do not consider disturbance inputs at later times, such as, for example, a random forcing in time for the duration of the simulation, as in Bagheri *et al.* (2009*b*).

In order to compute balanced modes, ERA was used here. The Hankel matrix was 5958×5292, which is equivalent to taking 1986 adjoint snapshots and 2646 direct snapshots in BPOD. Fig. 5.6 shows that a 30-mode ERA model captures very well all outputs. We note that in a spatially developing flow, such as the boundary layer, there would be no signal from the actuator to the output y_{21}, since the information cannot travel upstream, but since channel flow is periodic, that signal is non-zero here. While the energy of the perturbation was found to be captured very well by as few as three modes for the localized perturbation discussed in Section 4.3, a 30-mode model was required to capture the dynamics very well in this case, since initial transient dynamics is of importance, and there are multiple inputs and outputs.

A key parameter for the compensator design is the value of the control input penalty in LQR. Since there is only one control input in all cases considered in this thesis, R is always a scalar. Values of $R = 0.1$ and $R = 0.001$ were picked, the latter being for a very aggressive controller. Fig. 5.7 shows the output y_1 and the

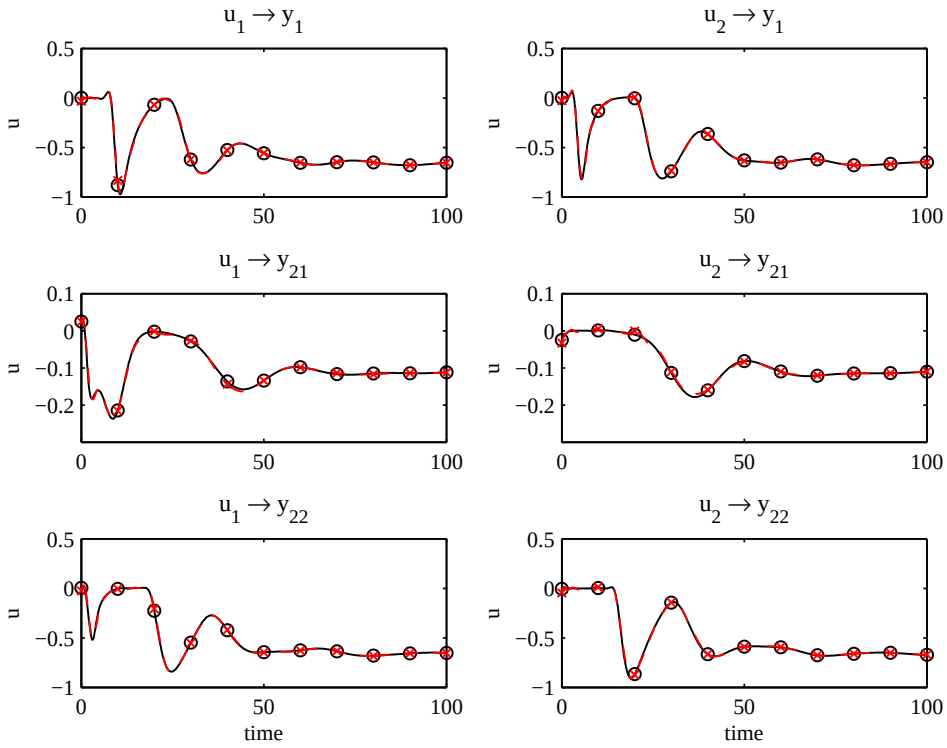


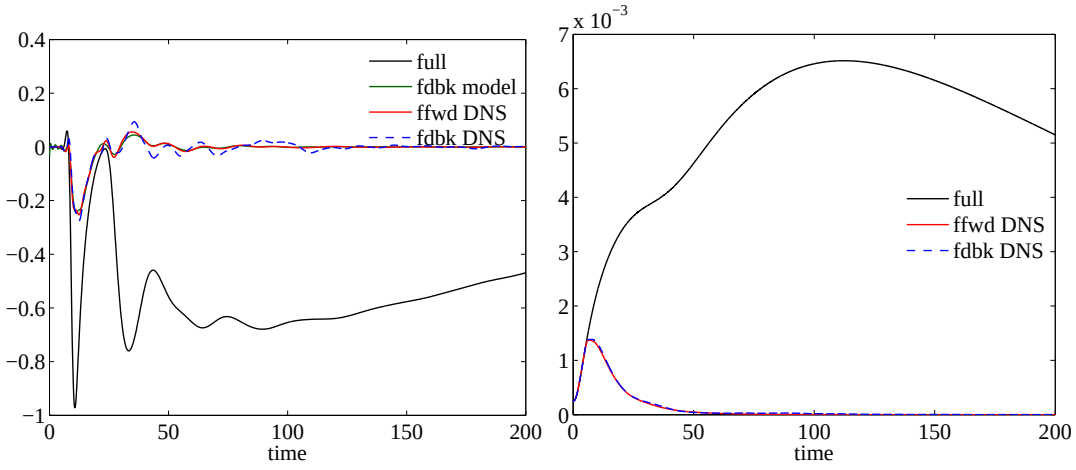
Figure 5.6: A comparison of the impulse responses of the uncontrolled linearized DNS ('full system', ◯) and the uncontrolled ERA model (X), using body force actuation, u measurement at $x = 3\pi$ as y_1, and two measurements of u at $x = 3\pi/2$ as y_2. The six panels represent the six elements of the 3×2 transfer function matrix for the model (three outputs and two inputs). The two entries of the output vector y_2 are denoted by y_{21} and y_{22}. Refer to Fig. 5.2 (a) for a schematic of the control setup. Note that u_1 and u_2 are control inputs, while u is a streamwise velocity signal.

energy for the case with $R = 0.1$. In addition to the stages of testing the controller described in 5.1, another stage was included here. The control input time history obtained for the reduced-order model was fed to the linearized DNS as a *feed-forward* input[3], and the results were compared to those using the full compensator (with estimation and feedback). We note that full-state feedback may be used as an additional test of the compensator when adjoint modes are available[4], as in Ilak & Rowley (2008*a*). The estimator was designed as described in Section 5.1.4, using a very clean measurement $\alpha = 0.001$ ($R = 1 \times 10^{-6}$). This corresponds to an 'aggressive' estimator, i.e., fairly high values of the estimator gains L, but no effect on the stability of the closed-loop system was observed.

We notice that the prediction of the model for the closed-loop performance and the output for the case of open-loop forcing are almost exactly overlapping,

[3]The term feed-forward means that a pre-computed control input time history is fed to the system, i.e. the control is not updated in real-time.

[4]In this case control is updated in real-time without an estimator, since the state of the reduced-order model is obtained at each DNS time step by projecting the full field onto balancing modes.



(a) (b)

Figure 5.7: (a) Single output y_1 for body force actuation with u measurements, comparing the uncontrolled impulse response of B_1 ('full') with controlled response of the reduced model, feedforward controlled linearized DNS and linearized DNS with feedback control. (b) Perturbation energy for the same run.

indicating that the model does capture very well the dynamics of the system. While the outputs of the open-loop and closed-loop cases are slightly different, the corresponding energy, which has been decreased significantly, is almost identical in the two cases. It was found that there is no appreciable difference in the performance for $R = 0.001$ in LQR (see Section 5.4 for a discussion of the performance limitations in the system).

5.2.3 Model performance - minimizing wall-normal velocity

The wall-normal velocity was also considered, as both y_1 and y_2. Figure 5.8 shows the performance of the uncontrolled 30-mode model. We notice that the transfer function from the actuator to the measurement near the wall $u_2 \rightarrow y_{21}$ is not captured very well initially, which is not important for the controller performance, as the entire measurement vector y_2 is upstream of the actuator.

A comparison of the performance of feedback controllers designed for this model with values of the controller penalty $R = 0.1$, $R = 0.01$ and $R = 0.001$ is shown in Fig. 5.9. We see that an increasingly aggressive controller is more successful at reducing the energy, approaching the effectiveness of the compensator based on the streamwise velocity, although not reaching it (see Fig. 5.11).

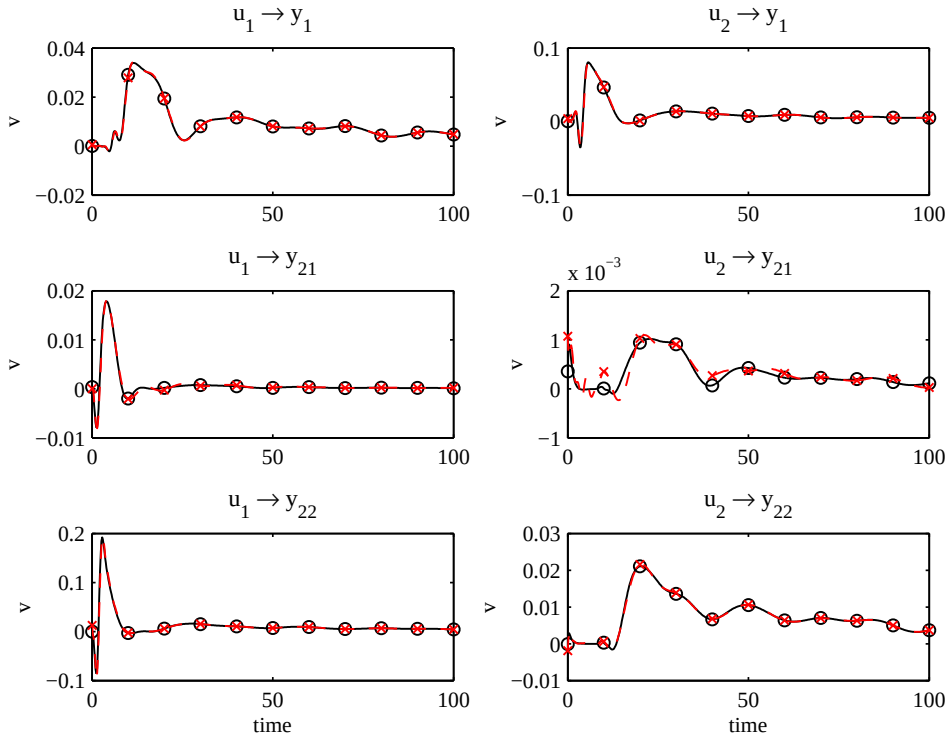


Figure 5.8: A comparison of the impulse responses of the uncontrolled linearized DNS ('full system', ◯) and the uncontrolled ERA model (X), using body force actuation, v measurement at $x = 3\pi$ as y_1, and two measurements of v at $x = 3\pi/2$ as y_2. The six panels represent the six elements of the 3×2 transfer function matrix for the model (three outputs and two inputs). The two entries of the output vector y_2 are denoted by y_{21} and y_{22}. Refer to Fig. 5.2 (a) for a schematic of the control setup.

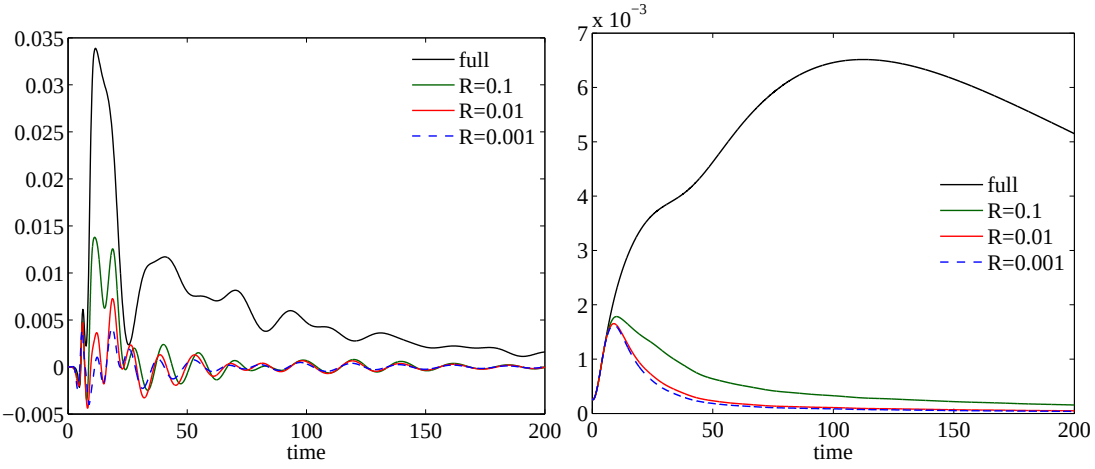


Figure 5.9: (a) Single output y_1 for body force actuation with v measurements, comparing an uncontrolled simulation to controlled simulations at three different values of the parameter R in LQR. (b) Perturbation energy for the same run.

5.2.4 Model performance - minimizing energy

We have seen that significant reduction in perturbation energy growth is achieved by a compensator that is designed to reduce the streamwise velocity as output y_1. It is of interest to compare the performance of this compensator to the performance of a compensator with the full field as output y_1, i.e. with a penalty on the total perturbation energy in the LQR cost function, since minimizing the total perturbation energy is potentially the most desirable option in suppressing transition. Measurements at a particular point in the flow such as the ones considered in the previous two subsections are more readily available in an experiment than measurements of the full velocity field. Therefore, we are interested in learning how close the controller performance using these measurements is to performance for minimizing the total energy.

In order to approximate the full field output at the reduced-order model level, output projection is used, as described in Chapter 2. Thus, when evaluating the performance of the reduced-order model, we compare the norm $\|\hat{y}\|^2$ (2.21) to the energy of the full field. If the rank of output projection is high enough, i.e., the reduced-order model is designed to capture well enough of the leading POD modes, the norm $\|\hat{y}_1\|^2$ will be very close to the overall perturbation energy, and we may then design a compensator for the reduced-order model to minimize this norm.

The advantage of ERA is particularly significant here, since models with high dimension of output projection are easily computed once the POD modes are available. In this case, only the projection of the impulse response snapshots onto the POD modes is needed to form the part of Hankel matrix Y^+X corresponding to the output C_1 (for details on forming the Hankel matrix using output signals, see Ma *et al.*, 2009*a*). On the other hand, the computation of the same models using BPOD would require many adjoint simulations. High rank of output projection may be necessary in order to correctly capture the transient dynamics of the system, as important dynamics of the initial transient is often not captured by the most energetic POD modes, as shown in Chapter 4.

Fig. 5.10 shows a comparison of the impulse response of the uncontrolled linear perturbation evolution and the impulse response of the reduced-order model. We note that the energy time histories for both inputs here are identical, since the actuator and the disturbance are identical, just with a phase shift.

Fig. 5.11a shows a comparison of the energy for linearized DNS with feedback control for the three different choices of the output y_1. Using full field as output indeed produces results very close to the results using only streamwise velocity, which is in agreement with the findings of Jovanović & Bamieh (2005) that the energy contained in just the streamwise velocity component of the perturbation is close to the full perturbation energy. Fig. 5.11b shows that the control input for minimizing the energy has a lower peak, i.e., less control effort is needed than when

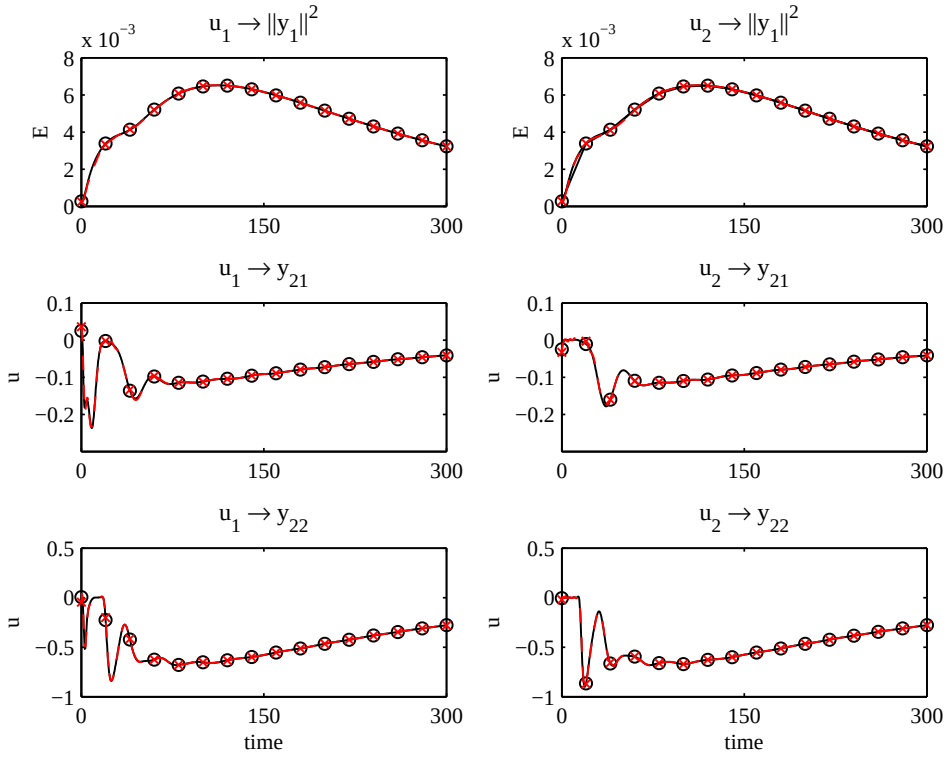


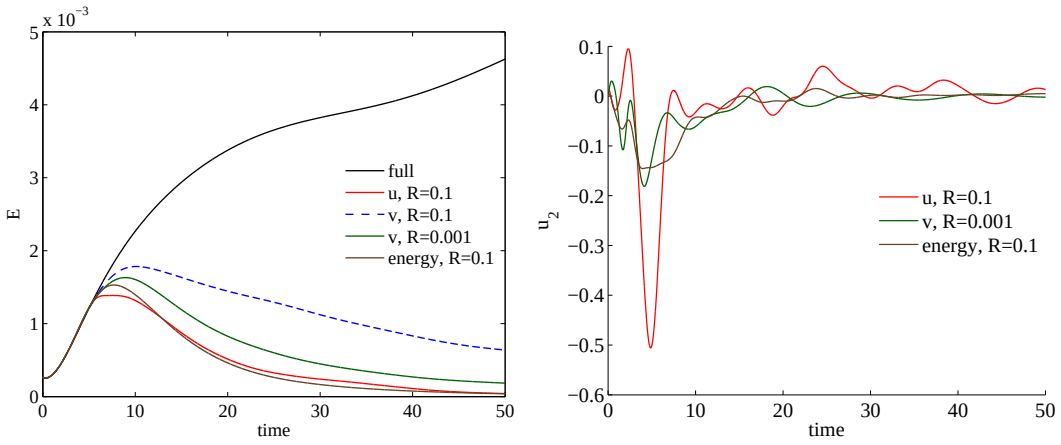
Figure 5.10: A comparison of the impulse responses of the uncontrolled linearized DNS ('full system', ◯) and the uncontrolled ERA model (X), using body force actuation, full field measurement as y_1, and two measurements of u at $x = 3\pi/2$ as y_2. The six panels represent the six elements of the 3×2 transfer function matrix for the model (three outputs and two inputs). The two entries of the output vector y_2 are denoted by y_{21} and y_{22}. Refer to Fig. 5.2 (b) for a schematic of the control setup. Note that u_1 and u_2 are control inputs, while u is a streamwise velocity signal.

the velocity signal is minimized. We notice that in all cases the energy is not affected at all by the application of the control until a certain time. This is due to a time delay in the system introduced by the spatial separation of the actuator and the disturbance (see Section 5.4). In Ilak & Rowley (2008*a*) the actuator was placed at exactly the initial location of the disturbance, and the controller was able to affect the energy growth immediately.

5.3 Feedback control of linearized flow using wall blowing/suction as actuation

Next, a localized blowing/suction actuator on the lower wall was used. The zero-net-mass-flux wall blowing/suction in the form of a circular 2-D 'sombrero' [5] function

[5]Name inspired by the shape of the function when plotted.



(a) (b)

Figure 5.11: Energy comparison for uncontrolled impulse response of the perturbation, feedforward control and feedback control for the localized body force actuator (a) and comparison of control input histories $u_2(t)$ for three different outputs y_1 (b).

given by:

$$v(x,-1,z,0) = \left(1 - \frac{r^2}{\alpha^2}\right) e^{-r^2/\alpha^2}, \tag{5.17}$$

where $r^2 = (x-x_c)^2 + (z-z_c)^2$. The value $\alpha = 0.7$ was used. This function because its integral over the entire lower channel wall is zero and thus no net flux of wall-normal velocity is added to the flow. The boundary conditions for this localized function are not periodic in x, z and periodicity was enforced by adding copies of the function centered at $(x_c \pm 4\pi, z_c \pm 2\pi)$. Without this correction, the non-periodicity was found to introduce spurious high-wavenumber components. Rather than using a distributed actuator as in previous work (Högberg *et al.*, 2003; Lee *et al.*, 2001), this actuator is localized and thus potentially closer to a practical device.

Since in the previous section it was shown that the measurements of wall-normal velocity result in worse performance than streamwise velocity or energy, here we only compare the performance of the models for streamwise velocity and energy as C_1. The particular solution $B_2 = -Z$ (after regularization by a short DNS time step in order to enforce the correct boundary conditions) is shown in Fig. 5.13.

5.3.1 Model performance - minimizing streamwise velocity

It was found that in this case a 40-mode model is necessary in order to capture well the dynamics of the uncontrolled system (Fig. 5.14). Fig. 5.15a shows the output signal y_1 for uncontrolled and controlled runs and the energy of the simulation. We note that the initial energy growth is even larger than in the uncontrolled case — with this choice of actuator, minimizing u does not minimize the energy. It was

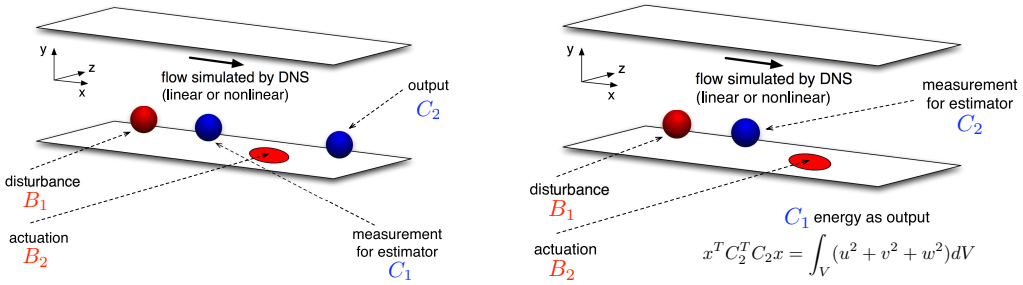


Figure 5.12: The control system setup for the cases considered in this section: wall blowing/suction as actuation and localized measurement as output y_1 (a) and wall blowing/suction as actuation and full field as output y_1, corresponding to minimizing perturbation energy (b).

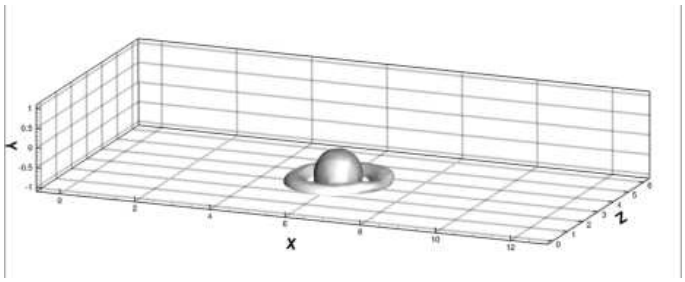


Figure 5.13: An isosurface of the half of the maximum value of the v component of $-Z$. Recall from Section 5.1.2 that $B_2 = [-Z; 1]$.

found that the aggressive controller use here ($R = 0.001$) even increases the overall energy growth significantly, although it successfully reduces the single output signal that it was designed to reduce. This controller was tested only in a feed-forward run, i.e., the control input was pre-computed for the reduced-order model and used in the DNS.

5.3.2 Model performance - minimizing energy

We next consider the full field as output y_1. In this case, the POD modes for output projection were computed using snapshots from both impulse responses, which now look quite different. The leading mode is shown in Fig. 5.16. It was found that there is no appreciable difference in the performance of models with measurements at the location corresponding to the maximum values of these modes and the performance using the measurements at the same locations as in the body force case.

Fig. 5.17 shows a comparison of the impulse responses of the uncontrolled model and uncontrolled linearized DNS for this case, showing that the dynamics is captured well by the reduced-order model. It should be noted that the energy of the impulse response of the actuator $B_2 = -Z$ exhibits no growth at all.

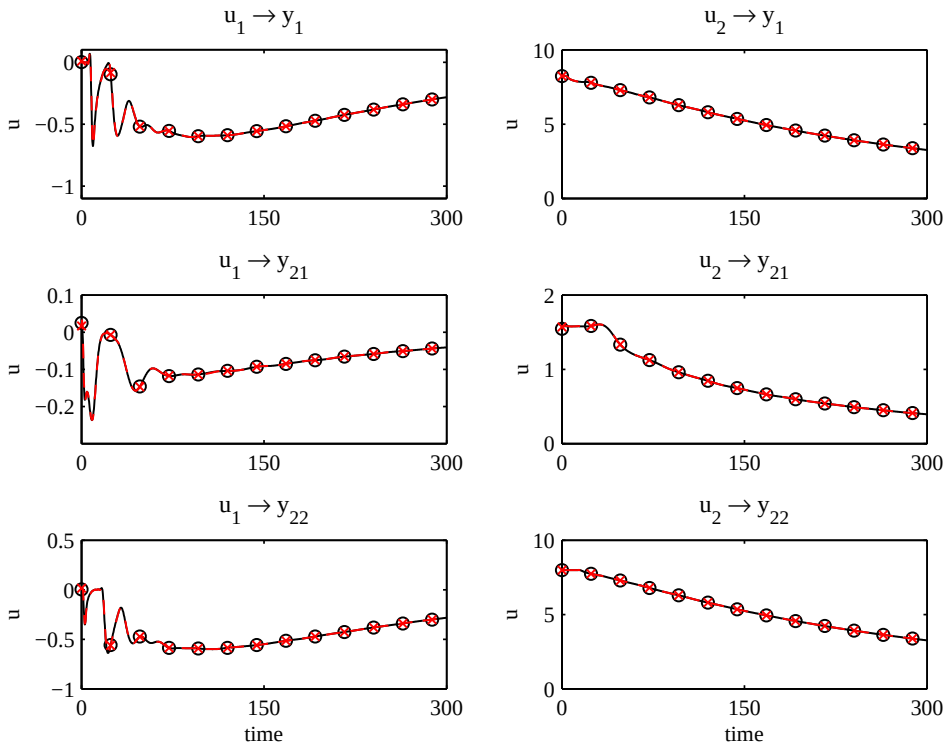
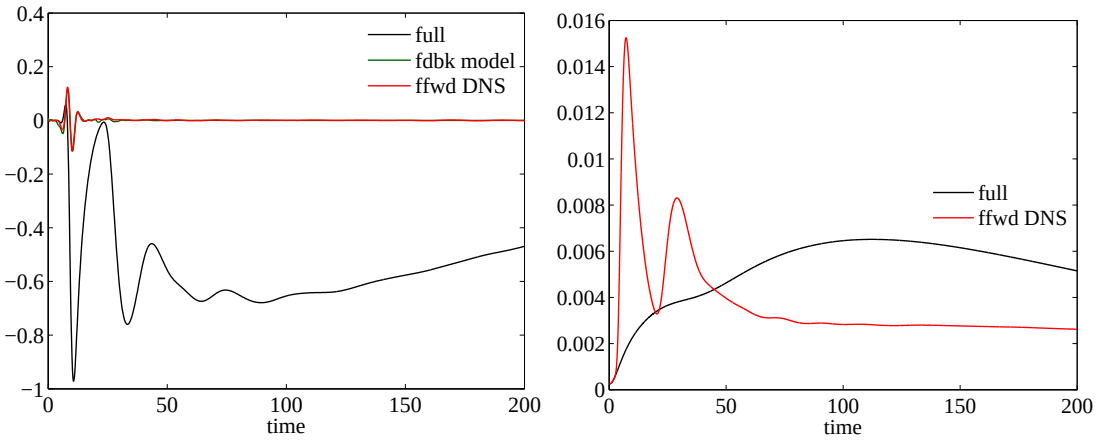


Figure 5.14: A comparison of the impulse responses of the uncontrolled linearized DNS ('full system', ◯) and the uncontrolled ERA model (X), using wall blowing/suction actuation, streamwise velocity u measured at $x = 3\pi$ as y_1, and two measurements of u at $x = 3\pi/2$ as y_2. The six panels represent the six elements of the 3×2 transfer function matrix for the model (three outputs and two inputs). The two entries of the output vector y_2 are denoted by y_{21} and y_{22}. Refer to Fig. 5.12 (a) for a schematic of the control setup.



(a) (b)

Figure 5.15: Output y_1 for wall blowing/suction actuation with u measurements (a) for uncontrolled and controlled runs. Perturbation energy for the same case (b).

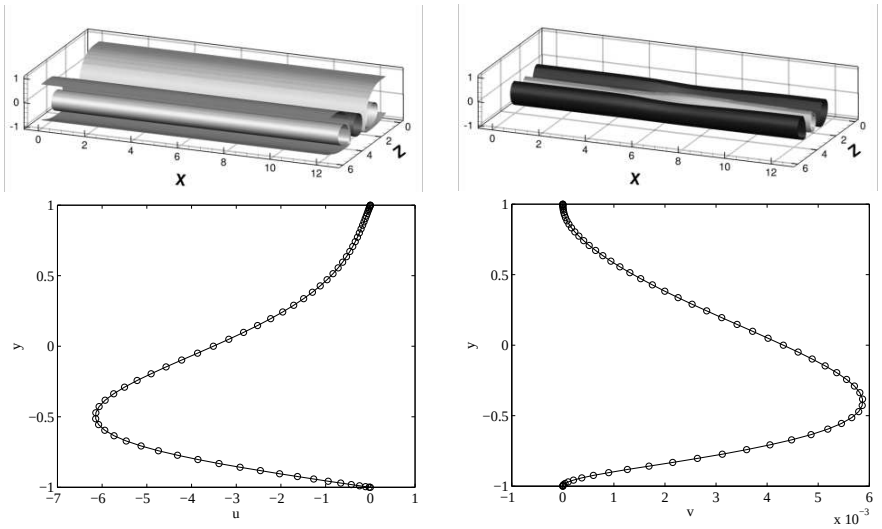


Figure 5.16: The leading POD mode including both snapshots of the localized perturbation and the actuator $B_2 = -Z$. Left: streamwise velocity u isosurface (top) and 1-D profile (bottom), right: wall-normal velocity v.

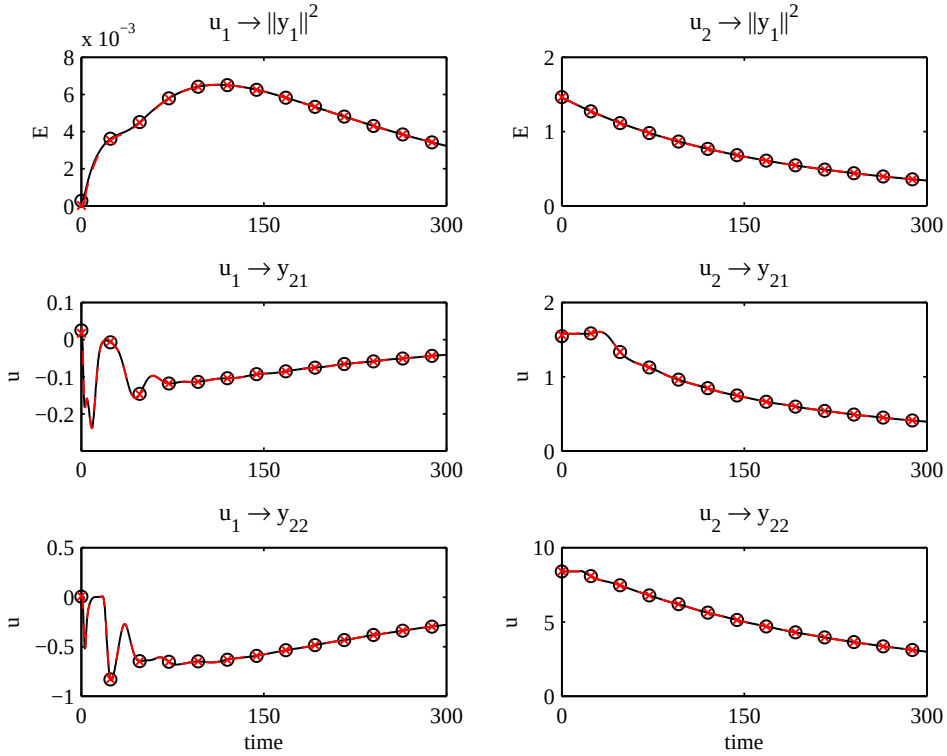


Figure 5.17: A comparison of the impulse responses of the uncontrolled linearized DNS ('full system', ◯) and the uncontrolled ERA model (X), using wall blowing/suction actuation, full field as y_1, and two measurements of u at $x = 3\pi/2$ as y_2. The six panels represent the six elements of the 3×2 transfer function matrix for the model (three outputs and two inputs). The two entries of the output vector y_2 are denoted by y_{21} and y_{22}. Refer to Fig. 5.12 (b) for a schematic of the control setup. Note that u_1 and u_2 are control inputs, while u is a streamwise velocity signal.

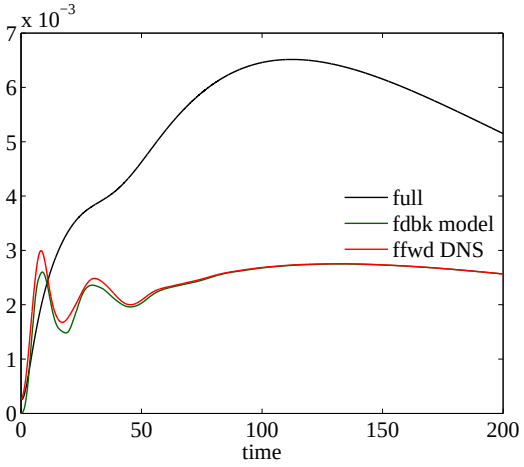


Figure 5.18: Perturbation energy for wall blowing/suction actuation with u measurements and energy output.

Table 5.1: Summary of results for feedback control of linearized DNS runs. Note that the quantity u_2 is the scalar control input, not to be confused with the streamwise velocity u. Also note that the integral of the actuation over time for the localized body force actuation and the wall blowing/suction actuation are not directly comparable due to the different actuators.

Actuation (B_2)	C_1	C_2	R in LQR	E_{max}/E_f (%)	$\int_0^{t_{end}} u_2^2\, dt$
localized near wall	u	u	0.1	21.3	23.18
	v	v	0.1	27.4	1.62
	v	v	0.01	25.4	3.24
	v	v	0.001	25.0	4.12
	Energy	u	0.1	23.5	5.60
wall blowing/suction	u	u	0.1	56.7	28.5
	Energy	u	0.001	11.1	4.8

Fig. 5.18 shows the perturbation energy growth for the simulation with the compensator on. The reduction in energy growth is not as significant as in the case of the body force actuator, but the actuator is successful, and no large overshoot in the energy as in Fig. 5.15. This controller was also tested only in a feed-forward run, i.e., the control was pre-computed for the reduced-order model and used in the DNS. The results of this section and the previous section are summarized in Table 5.1. The ratio E_{max}/E_f of the maximum perturbation energy observed in a controlled tun to the maximum energy of the uncontrolled perturbation impulse response E_f is used as a performance measure. The integral in time of the control input u_2 is also shown for the different runs.

5.4 Compensator performance in nonlinear DNS

Next, compensators designed for the linearized flow were tested in nonlinear DNS simulations. We first studied the uncontrolled nonlinear evolution of the perturbation from the previous two sections at different amplitudes. Fig. 5.19 shows a comparison in energy growth between linearized DNS and nonlinear DNS at different initial energy values. The resolution for all nonlinear runs in this section is $192 \times 129 \times 192$, which is slightly higher than the resolution in Kim *et al.* (1987). As before, the box size is $4\pi \times 2 \times 2\pi$, and $Re_c = 3300$. Constant mass flux of $\dot{m} = 4/3$ was imposed (see Appendix B). The ratio E_0/E_{lam} of the perturbation initial energy (in our framework this is the energy norm of B_1) and the energy per unit volume of the laminar flow (see Appendix B) above which transition occurs[6] was found to be approximately 6.85×10^{-5} (a simulation at 5.81×10^{-5} did not transition, while a simulation at 7.35×10^{-5} did). This very low value of the transition threshold is comparable in magnitude to the thresholds for some of the perturbations at $Re = 3000$ found by Reddy *et al.* (1998).

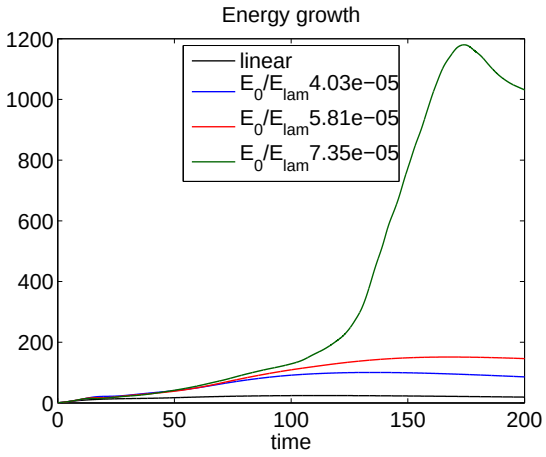


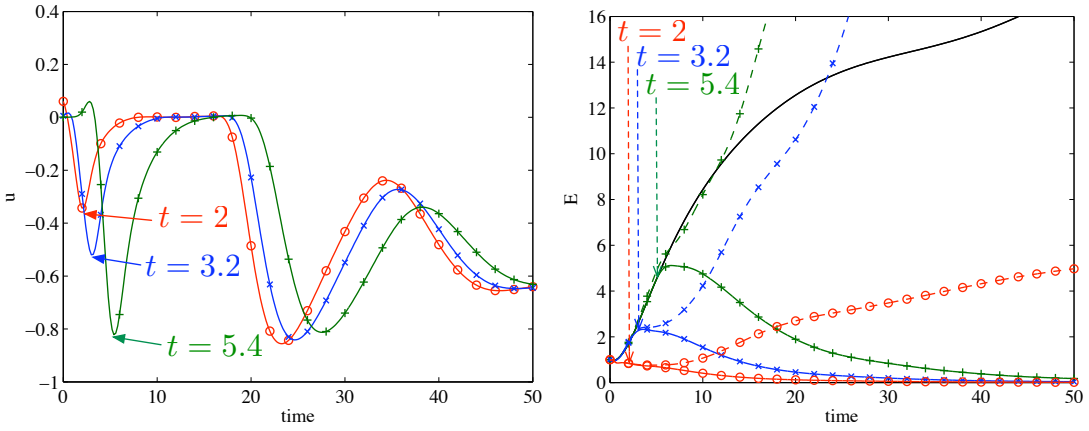
Figure 5.19: A comparison of energy growth for the linear evolution of the uncontrolled perturbation and three uncontrolled nonlinear simulations. The threshold for transition is at between values of 5.81×10^{-5} and 7.35×10^{-5} for the amplitude A in (5.16).

Since transition is observed in nonlinear channel flow with finite amplitude of the perturbation as initial condition transitions, we apply a compensator from Section 5.2 in an attempt to prevent it. The first investigation of balanced model performance in closed-loop control of transitional channel flow was undertaken by Ilak & Rowley (2008*a*), where closed-loop control with full-state feedback for BPOD models was used (the state of the reduced-order model was computed by projection of the DNS solution onto balanced modes at each time step). In that work, the study was performed at a lower Reynolds number ($Re = 2000$), and

[6]This value is also known as the transition threshold for a given perturbation (see Reddy *et al.*, 1998).

at a low resolution grid for the nonlinear simulations, which was still sufficient to capture the initial stages of transition. It was shown that transition could be delayed successfully by placing the actuator at the same location as the perturbation, in that case the center of the channel. Here we repeat this investigation for the current control setup, which is more realistic due to the spatial separation of the perturbation and actuator.

A velocity signal at the location of the body force actuator from Section 5.2 is plotted in Fig. 5.20 for three different positions of the actuator, $x = 2\pi$, $x = 3/2\pi$ and $x = 5/4\pi$, corresponding to streamwise separation of π, $\pi/2$ and $\pi/4$ respectively. A close-up of the perturbation energy for both uncontrolled and controlled cases is shown in the same figure, indicating that the controller only starts affecting the energy growth once the perturbation reaches it. This delay is imposed by the geometry of the problem. The same kind of time delay was observed by Bagheri *et al.* (2009*a*). In Ilak & Rowley (2008*a*), the actuator was located at the center of the channel, the same location as the disturbance, and was thus able to immediately suppress the energy growth. The spatial separation in the current setup (see Figs. 5.2 and 5.3), however, is more likely to be encountered in physical applications. In each case, feed-forward control was used to test the actuators.



(a) (b)

Figure 5.20: (a) A single-point measurement of the u velocity at the point where the actuator is centered for three different streamwise locations (uncontrolled runs), corresponding to spatial separation in streamwise direction of: π (+), $\pi/2$ (X), $\pi/4$ (◯). (b) Energy growth of the full linearized simulation (black solid line), and energy growth of controlled runs using the three setups. Same symbols as on left, solid lines represent linear feedback control runs, and dashed lines represent nonlinear feed-forward controlled runs. The only nonlinear simulation that eventually re-laminarizes is the one corresponding to spatial separation of $\pi/4$ (◯, dashed line).

The time delay imposes a performance limitation on the model (Skogestad & Postlethwaite, 2005, Chapter 5). The time delay imposes a minimum value of the cost function in the LQR problem, i.e., the overall energy of the controlled flow can never be zero, no matter how low the value of the control penalty R is, while in the

absence of such a limitation an arbitrarily 'cheap' controller (i.e., no penalty on the controller, meaning that infinite control inputs are allowable) would immediately kill the energy growth. We next explore the effect of this performance limitation on delaying transition in the nonlinear simulations.

A mechanism of transition in localized perturbations in channel and boundary layer flows was described as 'β-cascade' by Henningson *et al.* (1993), due to the rapid introduction of spanwise wavenumbers (corresponding to β in the Fourier expansion $u = u(y)\exp(i(\alpha x + \beta z))$). Since the perturbations that exhibit most linear energy growth are streamwise-constant (see, for example, Butler & Farrell, 1992; Schmid & Henningson, 2001), in a localized perturbation with components at many wavenumbers, it can be expected that over time streamwise-constant ($\alpha = 0$) components will dominate, and thus the introduction of new wavenumbers will be primarily at low streamwise wavenumber α and higher β.

Fig. 5.21 shows the 2-D Fourier transform of the v velocity at $y = -0.60$ (transformed both in x and z) at $t = 5.4$, and at a later time, $t = 12$, for linear and nonlinear uncontrolled evolution of the perturbation, as well as linear and nonlinear controlled cases. This figure demonstrates that the linear controller does not act early enough to affect the development of the β-cascade.

Figure 5.22 shows a time history of the 2-D Fourier transform of the v velocity for the case where the separation in the streamwise direction between the perturbation and the actuator was $\pi/4$. This simulation (red lines in Fig. 5.20) was found to relaminarize, reaching peak energy growth of 6.46 at $t = 94$. As seen in Fig. 5.22, there is still a β-cascade mechanism, although the initial action of the controller reduces the magnitudes at the high spanwise wavenumbers to low magnitudes. Due to the action of the controllers, the amplitudes of the newly arising spanwise components are low and the overall energy growth of the perturbation is small enough for it to re-laminarize.

5.5 Summary

In this chapter, ERA was used to compute balanced reduced-order models of a system with two inputs and two outputs. Optimal LQG compensators were designed for the models, and tested in both linear and nonlinear DNS runs. The Reynolds number and the box size in both linearized and nonlinear DNS were the same as in the benchmark work of Kim *et al.* (1987), and the resolution in the nonlinear case was slightly higher than in that work. ERA was used, since balanced models are obtained much faster than computing BPOD.

It was found in Section 5.2 that, using localized body force actuation, a compensator designed to minimize a signal of the streamwise velocity u work as well as compensator designed to minimize the total perturbation energy, although at the

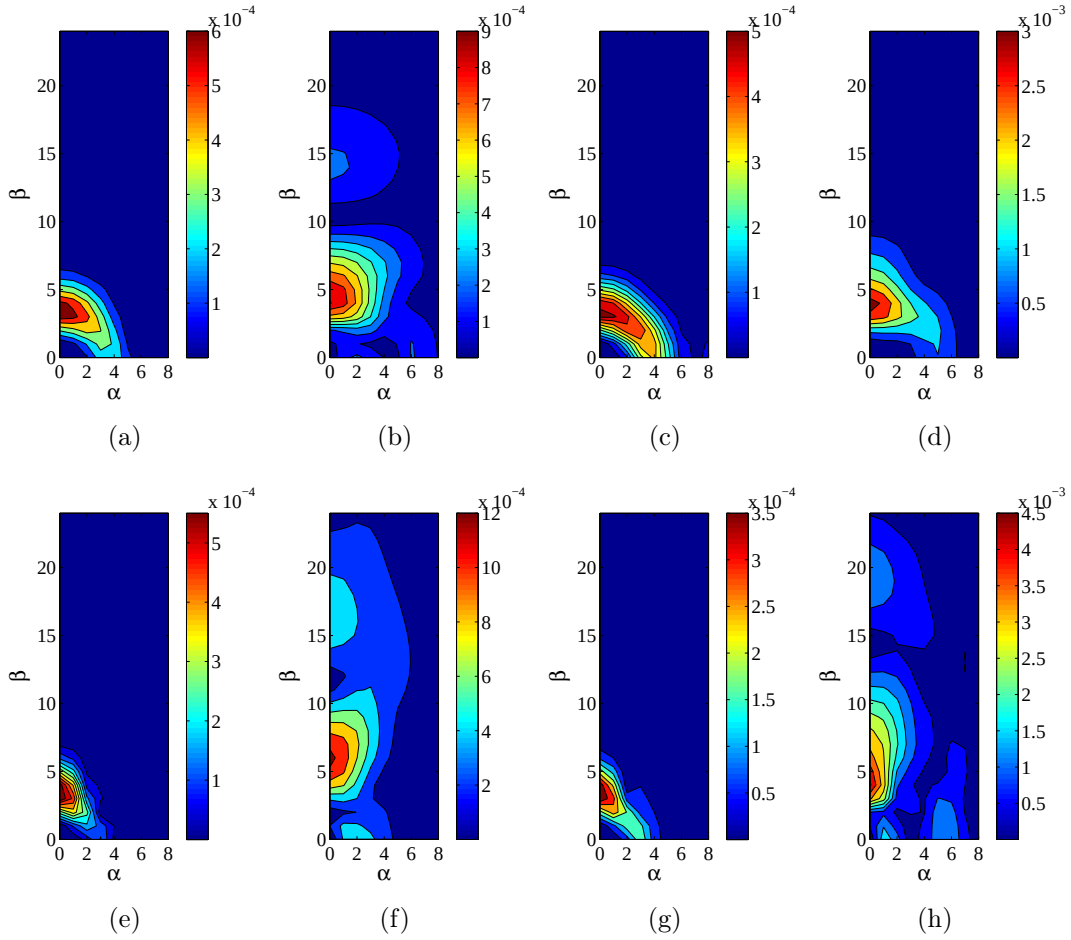


Figure 5.21: Spatial Fourier transform for the x, z-plane at $y = -0.60$ at $t = 5.4$ (a-d) and $t = 12$ (e-h). From left to right: linear evolution (a,e), nonlinear evolution (b,f), linear evolution with compensator on (c,g) and nonlinear evolution with compensator on (d,h). While it appears that there are no high spanwise wavenumbers introduced in the nonlinear controlled simulation at $t = 5.4$ (d) when the controller starts to act (recall Fig. 5.20), a β-cascade is still apparent later at $t = 12$ (h), and the linear controller thus fails to prevent transition. The ratio E_0/E_{lam} for the nonlinear simulations is 7.35×10^{-5}.

expense of some control effort. In 5.3, using wall blowing/suction as actuation, the same kind of compensator introduced a significant energy overshoot in the total perturbation energy. The overshoot was not present when the compensator with wall blowing/suction is designed to minimize the total perturbation energy. It was also found that, due to the time delay introduced in the system by the spatial separation of the perturbation, the measurement and actuator imposes performance limitations on the compensator, and using more aggressive controllers (corresponding to a smaller weight R in the LQR problem) does not improve the performance.

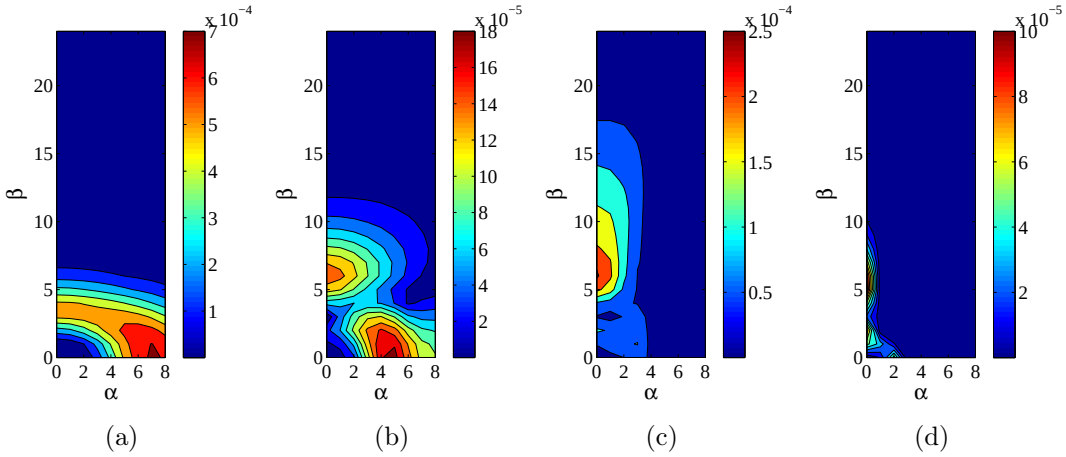


Figure 5.22: A time evolution of the spatial Fourier transform for the x, z-plane at $y = -0.60$, plotted at $t = 0.8$ (a), $t = 4$ (b), $t = 12$ (c), $t = 120$ (d). The ratio E_0/E_{lam} is 7.35×10^{-5}.

It was found that transition in nonlinear simulations for an initial perturbation amplitude above the transition threshold can only be prevented by the linear controllers if the actuation is placed close to the location in the initial perturbation in the streamwise direction. While controllers with actuators that were placed at π and $\pi/2$ spatial distance units downstream from the peturbation were not successful at reducing transition growth, an the actuator was placed at $\pi/4$ units downstream was found to suppress transition at a value of E_0/E_{lam} above the transition threshold. It was observed that in all cases the β-cascade mechanism arises despite the action of the compensator, but in the last case the compensator is able to act early enough in suppressing spanwise wavenumbers, so that the amplitude of the introduced wavenumbers is very small and they decay without transition.

Chapter 6

Towards nonlinear reduced-order models

The linear models of transitional channel flow were found in the previous chapters to be successful at capturing and controlling the linear dynamics of localized perturbations. It was shown that a nonlinear mechanism that governs the interaction of linearly growing perturbations at different spatial wavenumber pairs, known as the β-cascade (Henningson *et al.*, 1993), is not captured by the linear models, which do not include interaction among wavenumber pairs. It is therefore desirable to obtain nonlinear reduced-order models that may capture this mechanism.

The empirical method of Lall *et al.* (2002), described in Chapter 2, which in this thesis we refer to as nonlinear balanced truncation, is not practical for very large systems, since the number of simulations required to compute the empirical observability Gramian is too large. Therefore, the Complex Ginzburg-Landau (CGL) equation, a 1-D complex PDE, is used as a model problem. This equation has been widely studied recently, both in as a model problem for flows such as flow past a cylinder (Chomaz, 2005), and in the context of control design (Lauga & Bewley, 2004; Bagheri *et al.*, 2009*c*). In the latter work balanced truncation models for the linear part of the equation were found to perform well when used for feedback control, motivating the present extension of the investigation to nonlinear models.

Nonlinear models are also computed using balancing and adjoint modes computed using standard balanced truncation of the linear part of the equation. This approach avoids the computational expense and difficulty of nonlinear balanced truncation. It was shown in the previous chapters that linear balanced truncation and BPOD models capture well the linear dynamics of perturbations. Thus, we may expect that, if the form of the nonlinear terms is known exactly, which is the case for the CGL equation (and also the Navier-Stokes equations), nonlinear models formed by projection onto these modes may also be able to capture the nonlinear dynamics.

We show that this is indeed the case, and that the results warrant an investigation of nonlinear models of transitional channel flow.

The work presented in this chapter was started during a visit to the Mechanics Department at the Royal Institute of Technology, Stockholm in September–December, 2007, in close collaboration with Shervin Bagheri, and under the guidance of Prof. Dan Henningson, Docent Luca Brandt and Prof. Clarence Rowley.

6.1 The Complex Ginzburg-Landau equation

We begin by briefly introducing the CGL equation. Details can be found in Chomaz (2005); Cossu & Chomaz (1997); Bagheri *et al.* (2009*c*). The CGL equation is a 1-D PDE of convection-diffusion type, with an extra term representing linear exponential disturbances, defined on the infinite domain $x \in \{-\infty, \infty\}$, and it is given by:

$$\frac{\partial q}{\partial t} = Aq + f(q) \tag{6.1}$$

where the linear part is given by:

$$Aq = (-\nu\frac{\partial}{\partial x} + \gamma\frac{\partial^2}{\partial x^2} + \mu)q. \tag{6.2}$$

Since the spatial domain is infinite, there are no boundary conditions, but the complex solution $q(x,t)$ is required to remain finite as $x \to \pm\infty$. The complex convective velocity is defined as $\nu = U + 2ic_u$, $\gamma = 1 + ic_d$ is a diffusion parameter, and the parameter μ is defined as:

$$\mu(x) = (\mu_0 - c_u^2) + \mu_2\frac{x^2}{2}. \tag{6.3}$$

The spatially varying parameter $\mu(x)$ models the presence of exponential disturbances, and in this form defines a region:

$$-\sqrt{-2(\mu_0 - c_u^2)/\mu_2} < x < \sqrt{-2(\mu_0 - c_u^2)/\mu_2},$$

in which disturbances are amplified. The upstream and downstream ends of the unstable region are called branch I and branch II, respectively. Details on reduced-order modeling and control of the linear part of the CGL equation are given in Bagheri *et al.* (2009*c*). The nonlinear term $f(q)$ is given by:

$$f(q) = -a\|q\|^2 q, \tag{6.4}$$

where a is a real parameter. This nonlinear term corresponds to the normal form for a Hopf bifurcation, and also appears in Landau amplitude equations. It determines the stability of the solutions that branch from the trivial solution as the bifurcation parameter μ_0 varies, as derived by Chomaz (2005). For $a > 0$, the solution of the CGL equation saturates to a limit cycle in the cases where the linear part of the problem is unstable. This represents a 1-D model of a cylinder wake. An

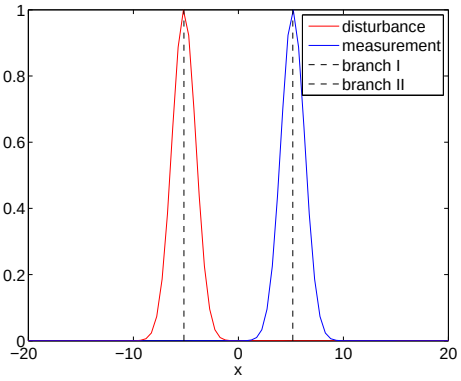


Figure 6.1: The disturbance (B) and measurement (C) for the 1-D domain of the CGL equation.

important feature of this nonlinear term is that, for real a, it does not introduce new frequencies, which is easily shown. Rather, it behaves as a nonlinear damping mechanism, the damping being proportional to the norm $\|q\|^2$ at each point in space. Therefore, when the system is forced at a given frequency, the steady-state response will have exactly the same frequency if the linear part is stable.

Introducing actuation and measurement, we can write the CGL equation in a state-space form:

$$\begin{aligned} \dot{q} &= Aq + f(q) + Bu \\ y &= Cq \end{aligned} \tag{6.5}$$

We assume a linear disturbance term, described by B, which would correspond to B_1 in the notation of the previous chapters. We do not consider actuation, which would be represented by a B_2 term. The output is related to the state by a linear measurement matrix C. The system is therefore single-input single-output (SISO). The setup of the control problem is shown in Figure 6.1. The disturbance is Gaussian and is located at branch I, since this is the beginning of the region receptive to disturbances. We note that the maximum of the absolute value of an optimal ('worst-case') disturbance is located near branch I (Bagheri *et al.*, 2009*c*). The measurement is also a Gaussian function in x with the same parameters located at branch II, as this is the location where the perturbation grows largest before starting to decay in space in the convectively unstable case. The exact form of the function used for the B and C matrices is:

$$g(x) = e^{-(x-x_0)^2/\sigma^2}, \tag{6.6}$$

with x_0 being the location of branches I and II for B and C respectively, and with $\sigma = 1.6$. The first step in Galerkin projection is the expansion of the field q onto a set of modes as $q = \Phi z$. Then, after taking the inner product of the equations with the inverse transformation Ψ, the CGL equation is projected onto a basis of modes

as follows:

$$\begin{aligned} \dot{z} &= \Psi^T A \Phi z - a\big(\Psi^T |\Phi z|^2 (\Phi z)\big) + \Psi^T B u \\ y &= C \Phi z, \end{aligned} \tag{6.7}$$

where we have used Eq. 6.4 to expand the nonlinear term in the basis Φ. From the above expression it may be expected that any basis that captures the linear dynamics and the actuation correctly (assuming zero initial condition on the state) will reproduce the dynamics of the state successfully, since the form of the nonlinear term is known exactly. If, in addition, the projection basis captures correctly the output C, the output of the reduced-order model will match the output of the full system. This means that modes from linear balanced truncation, which typically capture well the linear dynamics and actuation (see Chapter 4), are expected to provide a suitable basis for Galerkin projection of the nonlinear CGL equation.

6.2 Reduced-order models of the CGL equation

Reduced-order models of the CGL equation were computed using linear balanced truncation and the method of Lall *et al.* (2002), introduced in Chapter 2, for three different sets of parameters. The nonlinear term is stabilizing, as described above. If the linear part is locally convectively unstable, the nonlinear solution also exhibits only a convective instability, with the nonnormal growth reduced significantly compared to the linear solution due to the damping effect of the nonlinearity. We have labeled this case as Case A, and we have investigated it for a value of the parameter μ_0 only slightly below the critical value of μ_{cr}, specifically at 95% of its value. We also investigate two cases with $\mu_0 > \mu_{cr}$, labeled B and C, with μ_0 at 105% and 110% of the critical value respectively. In all simulations, $U = 2.0$, $c_u = 0.3$, $c_d = -1.0$, $\mu_2 = -0.08$, $\mu_{critical} = 1.132$ and $a = 0.1$.

The CGL equation was discretized pseudospectrally using Hermite polynomials, and the Crank-Nicolson scheme was used for time advancement (for details, see Bagheri *et al.*, 2009*c*). All simulations in this paper have been performed with $N = 100$, i.e. the full system has 100 complex states, corresponding to 200 degrees of freedom. Two sets of ten amplitudes c_m (see Chapter 2) each were chosen for the initial conditions in Case A: a set of low amplitudes between 0.01 and 0.3 on a logarithmic scale, and a set of higher amplitudes between 0.3 and 3.0, also on a logarithmic scale. For Cases B and C, five amplitudes were used in each set of initial conditions. All the cases investigated are summarized in Table 6.1. For case C, narrow Gaussians were used as initial conditions for the impulse responses of the basis vectors of the state space for the empirical observability Gramian in nonlinear balanced truncation. This was necessary, since it was found that numerical artifacts due to the discontinuity in setting just one point to 1 and all other points to zero are amplified in the unstable region, leading to unphysical limit cycle oscillations even for the case where the initial condition is downstream from the unstable region.

Table 6.1: Summary of the different cases for model reduction of the CGL equation.

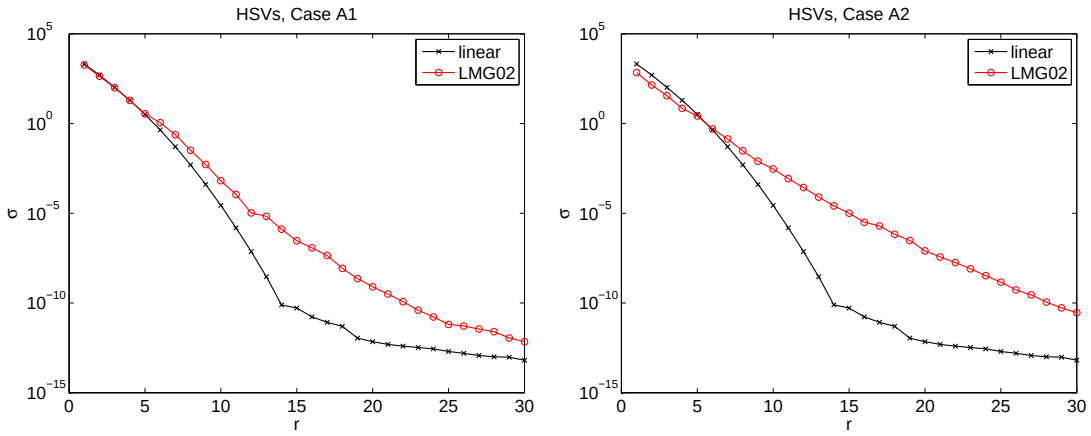
case	u amplitudes	μ_0/μ_{cr}	stability of linear part
$A1$	{0.01, ..., 0.3}	0.95	stable
$A2$	{0.3, ..., 3.0}	0.95	
$B1$	{0.01, ..., 0.3}	1.05	unstable
$B2$	{0.3, ..., 3.0}	1.05	
$C1$	{0.01, ..., 0.3}	1.10	
$C2$	{0.3, ..., 3.0}	1.10	

Since the system has a single input, the set of rotation matrices $\mathcal{T}$ (recall the definitions from Section 2.3.5) consists only of ± 1, and the set $\mathcal{E}$ consists of just $e = 1$, i.e. both sets consist of scalars instead of matrices. The number of simulations for the empirical controllability Gramian $\hat{W}_C$ is therefore equal to twice the number of amplitudes c_m which we choose. Both positive and negative amplitudes of the initial conditions were used, corresponding to $T = I$ and $T = -I$ in the Lall *et al.* (2002) method. We note here that in the linearly unstable cases B and C all simulations saturate to the limit cycle, and the amplitude of the initial condition plays a role only in the transient behavior before the steady-state oscillation is reached. The amplitude of the initial condition in the convectively unstable Case A has more influence on the dynamics, since the damping due to the nonlinear term will be stronger for higher amplitudes.

6.3 Case A

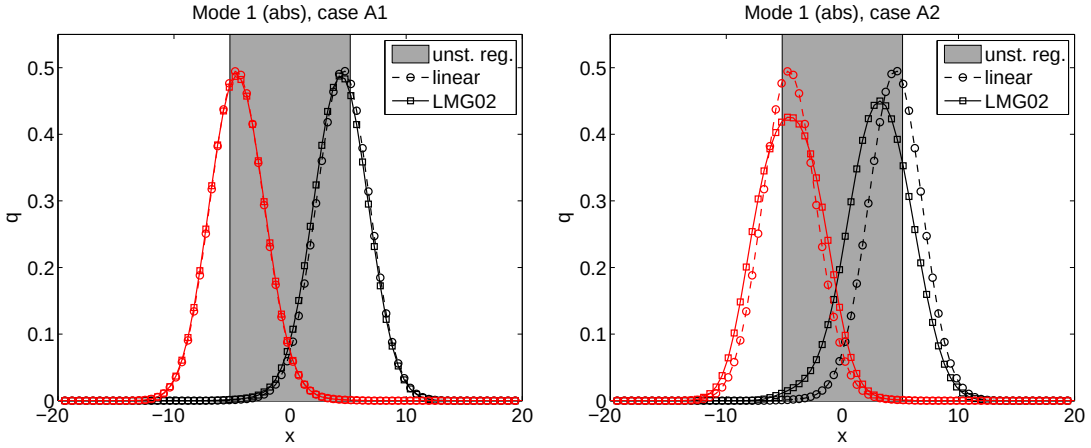
Figure 6.2 shows the Hankel singular values corresponding to the two sets of modes for each of the initial condition amplitude sets. As expected, in Case A1, the leading HSVs from linear balanced truncation are very close to the HSVs from the nonlinear simulations. The first mode for linear balanced truncation and the method of Lall *et al.* (2002) for Case A is shown in Figure 6.3. The spatial support of the nonlinear balancing and adjoint modes is very similar to that of the linear ones. The peaks of the adjoint modes are localized near branch I, while the peaks of the balancing modes are localized near branch II. The overlap of the balancing and adjoint modes is not very large, although it is sufficient in the first mode for the flow to be susceptible to convective instabilities (see Chomaz, 2005).

Figure 6.4 shows the impulse responses of rank 5 models for the Lall *et al.* (2002) method and linear balanced truncation. It is interesting to note that the performance of the models based on the linear modes is very good for all the simulations, but that the models based on nonlinear balanced truncation for case A2 do not capture the output as well. We note here that at a higher rank of the



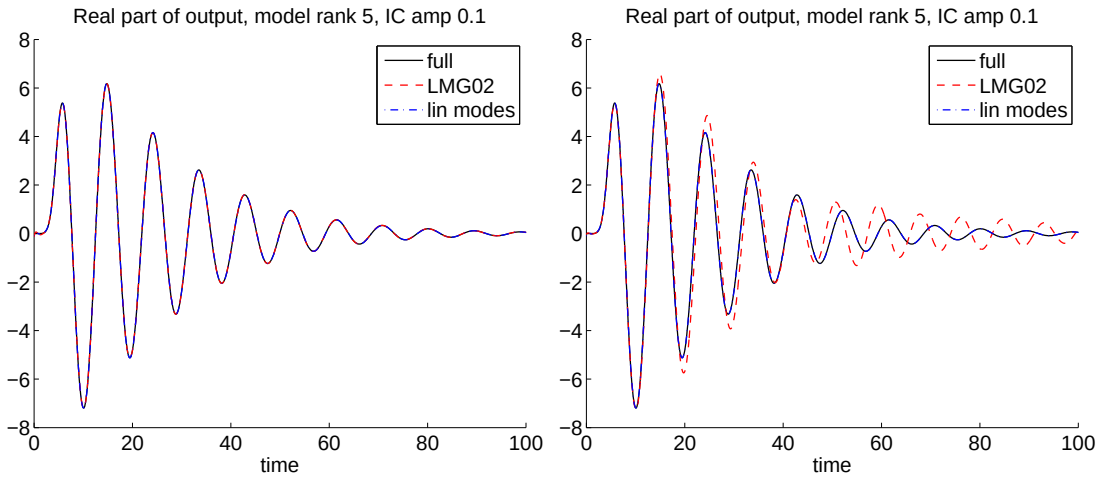
(a) (b)

Figure 6.2: The Hankel singular values for high (a) and low (b) amplitudes of the initial condition for Case A (cases A1 and A2 respectively). The abbreviation LMG02 refers to results of the empirical nonlinear balanced truncation method of Lall *et al.* (2002) in all figures in this chapter.



(a) (b)

Figure 6.3: Absolute value of first balancing mode (black) and adjoint mode (red) for linear balanced truncation (circles) and nonlinear balanced truncation (squares) for Case A1 (a) and Case A2 (b). The grey shading indicates the unstable region. The peaks of the nonlinear balancing modes for the large IC amplitudes are shifted slightly upstream with respect to the linear mode. The modes are normalized by their L_2 norm.



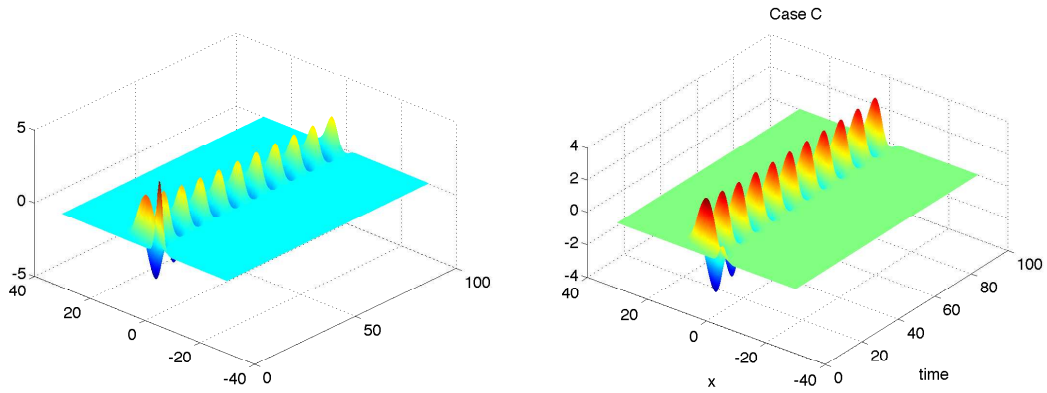
(a) (b)

Figure 6.4: Impulse responses for case A, with rank 5 models and initial condition amplitude 0.1. (a) Response of models using modes from case A1. (b) Response of models using modes from case A2. The response of the models that use modes from linear balancing almost exactly overlaps with the response of the full system.

reduced-order model, the model based on the method of Lall *et al.* (2002) also performs well.

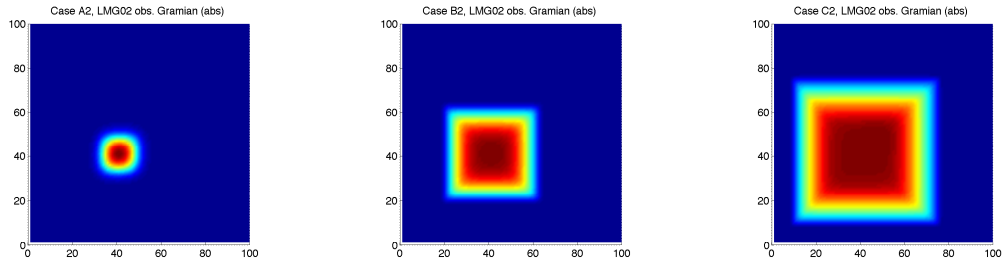
The unit impulse response amplitude of 0.1 used in all of the simulations in Figure 6.4 is in the lower amplitude range, so one may expect nonlinear balanced truncation for Case A2 to have inferior performance for this initial condition. However, the models based on nonlinear balanced truncation modes from case A2 perform poorly even for high amplitudes of the initial condition (not shown). The performance of the linear modes is superior, and the nonlinear modes computed from the simulations with low impulse response amplitudes are more similar to the linear modes, as shown above, which explains the difference in performance.

The good performance of linear balanced truncation and the similarity of the linear and nonlinear modes may be explained in terms of nonlinear systems analysis. We may think of the linear part of the CGL equation as evolving on a 'flat' manifold of the linear state space tangent to the 'curved' manifold of the nonlinear state space. For this linearly stable case, the further the initial condition is from the equilibrium, the faster it gets attracted to the equilibrium by the cubic nonlinearity. Hence most of the dynamics of the impulse responses 'happens' on the 'flat' manifold near the equilibrium, and that is what the snapshots in nonlinear balanced truncation capture. It is therefore not surprising that the linear modes capture well the dynamics of the nonlinear system and that the modes obtained using the method of Lall *et al.* (2002) are very similar to the modes obtained from linear balanced truncation. We may conclude that modeling this case (and similar more complex flows without saturation) using linear balanced truncation is sufficient and there is



(a) (b)

Figure 6.5: The real part of the response of the nonlinear CGL equation for different values of the parameter μ_0: (a) Case B: $\mu_0 = 1.188$ (105% μ_c), (b) Case C: $\mu_0 = 1.245$ (110% μ_c).



(a) (b) (c)

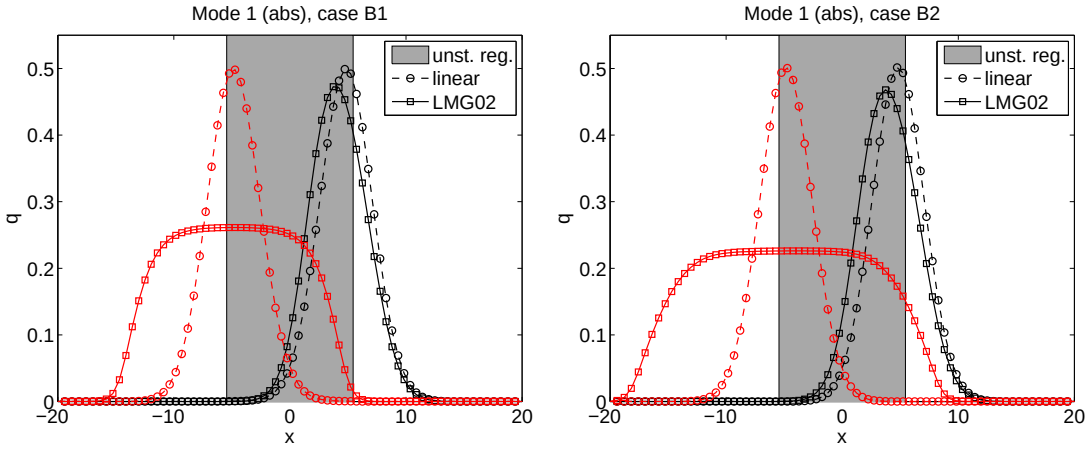
Figure 6.6: Absolute value of the entries of the empirical observability Gramians from nonlinear balanced truncation for cases A2 (a), B2 (b) and C2 (c). See definitions in Section 2.3.5. Red indicates large values.

no need to apply the more complicated nonlinear method. This is unlikely to be the case for more complex systems.

6.4 Cases B and C

For cases B and C, the equilibrium at $q = 0$ is unstable, and methods for balanced truncation of unstable systems need to be used. Since the system here is small, we have used the method of Zhou *et al.* (1999). We note that for large systems where the exact Gramians are not available, this method is not tractable, an extension of the method for large systems by Ahuja & Rowley (2008) may be used.

The empirical observability Gramians for the are shown in Fig. 6.6. As the value of the critical parameter is increased, the regions of the domain that affect the



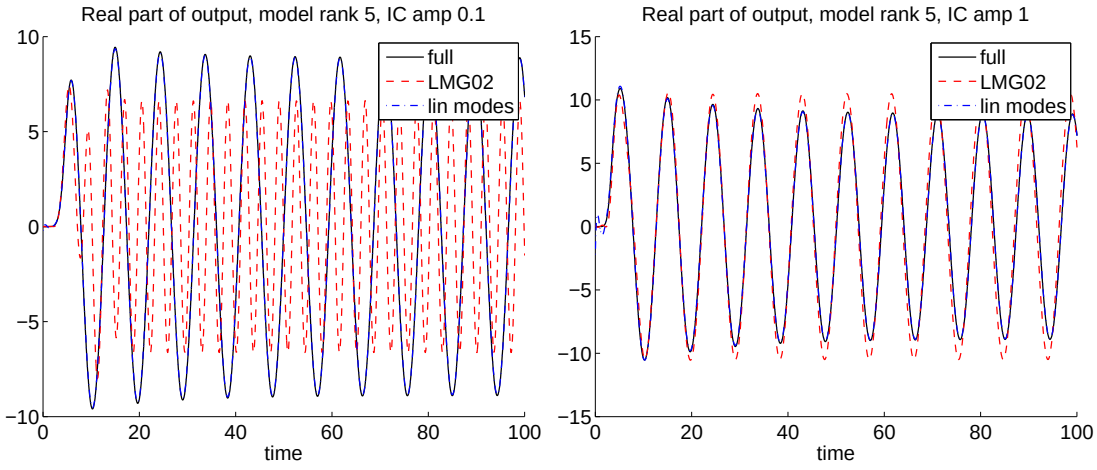
(a) (b)

Figure 6.7: Absolute value of first balancing mode (black) and adjoint mode (red) for linear BT (circles), nonlinear balanced truncation (squares) and approximate Lall method (x) for Case B, for small (a) and large (b) amplitudes of the initial condition. The grey shading indicates the unstable region. The modes are normalized by their L_2 norm.

output strongly (in case B and C this implies saturation of the limit cycle) become larger, which can also be seen from the plots of the leading modes in the cases. The empirical controllability Gramians (not shown) look very similar in all three cases. Physically, this corresponds to the higher value of the critical parameter which makes the upstream part of the domain more receptive to perturbations, and thus makes the highly observable part of the domain larger. On the other hand, the controllability in the three cases, which is a measurement of how well states respond to inputs, does not change considerably.

In the case of saturated oscillations (Fig. 6.5), the usual infinite-time Gramians cannot be computed as in Case A and therefore finite-time Gramians need to be taken into account. In this case, the duration of the simulation is of key importance, since the initial transient behavior becomes less important as the simulations become longer. In this work the Gramians were computed using relatively short simulations where some of the transient can be expected to play a role, although for the set of parameters investigated, the transient is very fast. More precise determination of appropriate selection of snapshots that ensures capturing the transient is a subject of future work.

The leading mode for case B is shown in Fig. 6.7. While the leading mode from nonlinear balanced truncation has similar support as the linear mode, and looks very similar to the nonlinear mode in case A, the adjoint modes from nonlinear balanced truncation have much wider spatial support than both the linear and nonlinear adjoint modes shown in Fig. 6.3. This corresponds to the plot of the empirical observability Gramian shown in Fig. 6.6b. It is interesting to note that for case B2,



(a) (b)

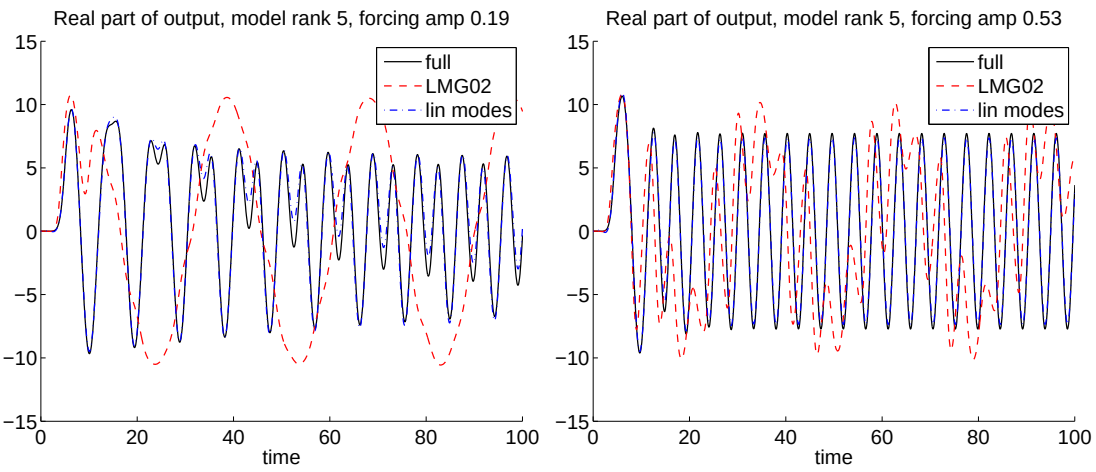
Figure 6.8: Impulse responses for case B2, with rank 5 models. Responses for initial condition amplitude 0.1, which was not included in the sampling for the nonlinear balanced truncation, are shown in (a), and responses for initial condition amplitude 1.0, which was included in the sampling, are shown in (b). The response of the models that use modes from linear balancing almost exactly overlaps with the response of the full system. At higher model rank, both methods perform really well.

the adjoint mode has much wider spatial support than in case B1, which can be explained by the easier excitation of the unstable region at higher initial condition amplitudes.

As shown by the results in Fig. 6.8, the linear modes capture the dynamics of the full system very well, even better than the empirical nonlinear modes, for a value of the parameter μ_0 close to the critical value. This may be expected in the light of the results for Case A. At higher model ranks (not shown), nonlinear balanced truncation models also capture well the output of the original system.

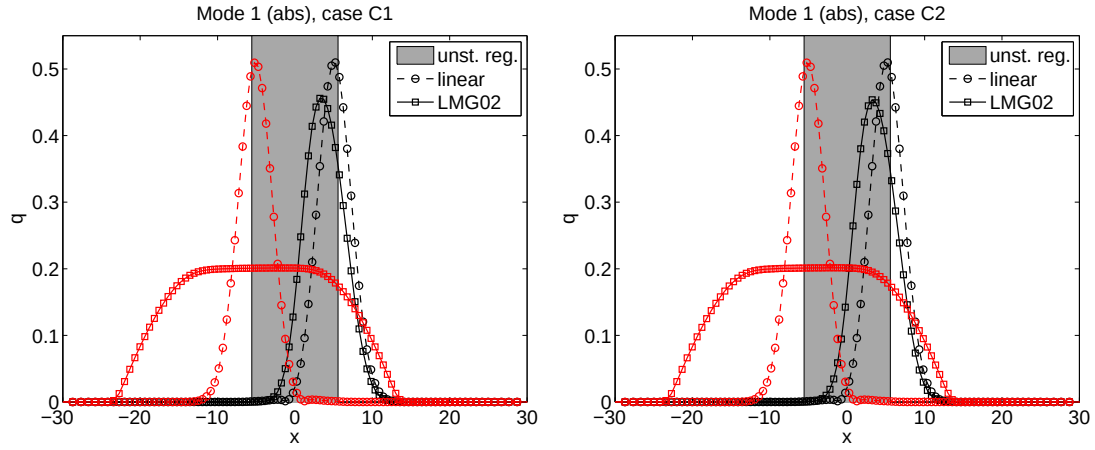
We have also investigated the forcing of the CGL equation at a particular frequency. The nonlinear modes were obtained by forcing the equation at the same frequency, but different amplitudes. Figure 6.9 shows the responses from the resulting models based on amplitudes from case B1 at two amplitudes, a lower one where both the driving frequency and the frequency of the limit cycle can be seen in the response, and a higher one where the driving frequency dominates. Surprisingly, the models based on linear modes are able to capture the behavior well at a low rank, while nonlinear balanced truncation models can perform poorly. At higher ranks of the models (not shown), both capture the output really well.

Figure 6.10 shows the first mode for cases C1 and C2. We note that the nonlinear balancing modes from the two subcases look quite similar, as opposed to cases A and B, where the modes computed using impulse responses at different amplitudes are different. This can be expected, since as the value of the parameter μ_0 increases,



(a) (b)

Figure 6.9: Responses for case B1, models with rank 6, forced at frequency $\omega = 1.35$, with amplitude of forcing 0.19 (a) and amplitude of forcing 0.53 (b). The response of the models that use modes from linear balancing almost overlaps with the response of the full system. At higher ranks of the models, both linear and nonlinear balanced truncation models capture the response perfectly.



(a) (b)

Figure 6.10: Absolute value of first balancing mode (black) and adjoint mode (red) for linear BT (circles) and nonlinear balanced truncation (squares) for Case C. The grey shading indicates the unstable region. Case C1 is shown in (a), and case C2 is shown in (b). The modes are normalized by their L_2 norm. Note the change in scale of the x-axis compared to Figs. 6.3 and 6.7.

the initial condition amplitude has less influence on the impulse response, as the saturation happens immediately. Similar to cases A and B, the performance of the models based on linear modes was found to be superior for both impulse responses and forcing at a specific frequency (results not shown).

It was also found that the models, all of which were obtained for the amplitude of the nonlinear term $a = 0.1$, also perform well if the amplitude in the model is changed to a different value (results not shown), again with the linear modes performing better at low rank than then nonlinear balanced truncation.

6.5 Summary

In this chapter, we applied the method of Lall *et al.* (2002) for empirical nonlinear balanced truncation to the Complex Ginzburg-Landau equation and compared the performance of the resulting nonlinear reduced-order models with the performance of nonlinear models obtained from using linear balanced truncation. We found that nonlinear models obtained via Galerkin projection of the CGL equation onto a basis of balancing modes computed for the linearized system using standard linear balanced truncation can perform very well, while using the proposed method for nonlinear balancing often results in models that do not capture the key dynamics. These results indicate that linear balanced truncation, which involves considerable fewer computational challenges, can potentially be used successfully in modeling of some nonlinear systems.

Further efforts in this direction would include the investigation of nonlinear models using linear balanced truncation or BPOD modes for larger problems such as transitional channel flow, or a cylinder in crossflow, which has already been investigated using POD by Noack *et al.* (2003). In these cases, the behavior of the nonlinear system can be much more complex than the simple nonlinear saturation observed and the method of Lall *et al.* (2002) or a similar method for computing nonlinear reduced-order models may be required to obtain good performance, despite the necessary computational expense.

Chapter 7

Conclusions and future work

7.1 Conclusions

This thesis presented the application of balanced truncation and Balanced Proper Orthogonal Decomposition (BPOD) to reduced-order modeling of linearized channel flow, and the use of the resulting models for feedback control. It was demonstrated that BPOD models capture the dynamics of linearized channel flow very well, and in particular much better than Proper Orthogonal Decomposition (POD). Furthermore, it was shown that controllers designed using BPOD models are successful in significantly reducing the energy growth of localized perturbations in linearized DNS simulations. However, the same controllers fail to capture a key nonlinear mechanism in nonlinear DNS simulations, and are effective only in a limited set of cases. While more detailed studies of actuator/measurement placement and model performance limitations are likely to improve the effectiveness of the linear controllers, it is also desirable to investigate nonlinear reduced-order models that would fully capture the transition mechanism. To that end, an empirical method for balanced truncation of nonlinear systems proposed by Lall *et al.* (2002) was evaluated by computing reduced-order models of the Complex Ginzburg-Landau equation, a 1-D model of fluid flows, as a first step towards similar investigations of channel flow.

The key feature of BPOD — capturing of effects of actuation and measurement in reduced-order models, along with a priori error bounds and its tractability for large systems, makes it the method of choice for model reduction in this work. Some of the technical details of the application of balanced truncation to channel flow, including the derivation of the appropriate adjoint equations, were presented in Chapters 2 and 3.

The thorough investigation of the capturing of the uncontrolled dynamics of linearized channel flow in Chapter 4 demonstrated first on a 1-D system that BPOD

is an excellent approximation for exact balanced truncation, thus reinforcing the findings of Rowley (2005). It was then shown in Chapter 4 that some important features from the control designer's point of view for the 3-D system, such as impulse response, frequency response, capturing of actuation and performance at off-design Reynolds number, were captured very well by the balanced reduced-order models.

It was also found that, while the leading POD modes capture very well the energy of the perturbation, the corresponding models do not capture the dynamics well: in particular, POD models fail to reproduce the energy growth of the perturbation, which was demonstrated by the particularly striking example for a POD model including low-energetic modes in Section 4.3.1. It was shown that the energy growth of the perturbation is captured only if modes with very low energy content are included in the POD models, while BPOD models that include only the leading balancing modes performed very well. The dynamics of this model, which includes modes (1-3,10,17), was further investigated in Section 4.4 using a graph decomposition method proposed by Mezić (2004).

Linear Quadratic Gaussian (LQG) compensators were designed in Chapter 5 using the Eigensystem Realization Algorithm (ERA), which is equivalent to BPOD and yields the same models at a small fraction of the computational cost, as shown by Ma *et al.* (2009*a*). These compensators were then tested, first on linear, and then on nonlinear runs. It was found that a compensator designed to reduce the streamwise velocity of the perturbation at a single point in the flow reduced the perturbation energy growth as much as a compensator designed to reduce the full perturbation energy. These results are in line with the findings of Jovanović & Bamieh (2005) about the correlation between streamwise velocity component of the perturbation and total perturbation energy. The LQG compensators include reduced-order estimators based on the ERA models, which were found to perform very well based on just two point measurements of streamwise or spanwise perturbation velocity.

A wall blowing/suction actuator, typically thought to be more practicable than body forces inside the flow, was found to have more severe performance limitations in comparison to a localized body force actuator. This limitation may have been expected, since actuation only at the wall has a more limited effect on the flow inside the channel than body force actuation, and further investigation in order to improve the performance of wall blowing/suction actuators is warranted. However, recent experimental work on introduction of body forces through plasma actuators (Corke *et al.*, 2009) indicates the potential of such devices in practical applications, and also warrants further studies using body force actuation.

The spatial separation of the perturbation and the actuator in the feedback control setup introduces an effective time delay in the system. In Ilak & Rowley (2008*a*), it was shown that the placing the actuator at the exact location of the perturbation results in successful suppression of transition for the nonlinear DNS,

since the controller is able to act immediately. A more realistic setup with separation in the streamwise directon between the actuator and the initial perturbation was used in Chapter 5, as we typically do not know where in the flow a perturbation will arise. As demonstrated by the study of perturbation energy growth and transition suppression for different spatial separation between the disturbance and the actuator, the β-cascade mechanism is not captured by the linear models, and thus only controllers that reduce the energy in low spanwise wavenumbers quickly are able to completely prevent transition. The results presented in Chapters 4 and 5 are the first BPOD computations for 3-D linearized flow, and for a large-scale system.

Since a key mechanism for coupling of linearly growing perturbation components at different spatial wavenumbers in transition is inherently nonlinear, and cannot be captured by linear models, resulting in linear controllers with limited effectiveness, nonlinear reduced-order models were also studied. A proposed method for empirical nonlinear balanced truncation was evaluated on a 1-D model vortex shedding, the Complex Ginzburg-Landau equation, in Chapter 6. It was found that reduced-order models computed using the method of Lall *et al.* (2002) do not capture well the dynamics of the CGL, and often perform worse than nonlinear models computed using modes from linear balanced truncation. Since linear balanced truncation is much simpler to compute than nonlinear balanced truncation, this finding indicates that nonlinear models of more complicated flows, such as channel flow, may be obtained at a relatively low computational expense.

7.2 Future work

The study performed here may be considered 'proof-of-concept,' as it uses a canonical test case, but it opens the door to many opportunities for future work, some of which we outline below.

Further work on channel flow. While perhaps the ultimate goal in flow control research is to extend the methods that are proven effective to flows encountered in practical applications, channel flow still remains an attractive problem for further development of flow control strategies. The simple geometry and relatively low cost of simulation, compared to, for example, the boundary layer, or 3-D flows in more complicated geometries, makes channel flow attractive for initial studies of new flow control methods. Besides linearized channel flow in the case of transition mechanisms, the minimal flow unit (Jiménez & Moin, 1991; Podvin & Lumley, 1998) has been shown to be a good test case for turbulent flows, as key dynamics is exhibited in a fairly low spatial domain, and advanced control strategies, nonlinear models and controllers, and other future approaches may be tested on this flow at relatively low computational expense.

Extension to other flows. It is of interest to extend the techniques demonstrated here to other flows, in particular ones with more practical relevance, such as the flat plate boundary layer. Efforts for this flow have already been undertaken (Bagheri *et al.*, 2009*b*) by applying the BPOD method. Control of flow past a flat plate in 2-D using BPOD was investigated by Ahuja & Rowley (2009), and potential extensions of this work for airfoils, first in 2-D and then in 3-D would be useful in bringing the flow control approach presented here closer to practical applications in aircraft. The use of ERA allows for computation of models with a larger number of inputs and outputs easily and quickly, and allows for exploration of different designs more efficiently for truly high-dimensional problems.

Balanced truncation for nonlinear systems. Both theoretical work and work on implementation of existing proposed methods is needed in order to obtain models that describe well the nonlinear dynamics, and in particular transition. The attractiveness of the snapshot-based method of Lall *et al.* (2002) makes it a potential candidate for large systems. The work presented in this thesis has identified some difficulties with that method, but further investigation, and in particular extensions for very large systems would be desirable.

Actuator and sensor placement. The placement of sensors and actuators is still somewhat ad-hoc, and relies on insights about flow dynamics. For example, in Chapter 5 POD modes were used to find a suitable location for measurements. Using ERA, models are computed easily once impulse response snapshots are available, and techniques using 'ERA-in-the-loop' as part of an optimization iteration procedure may potentially be designed in order to obtain truly optimal models.

Performance limitations and robustness. The performance limitations of the controllers based on the reduced-order models were demonstrated in Chapter 5. The detailed treatment of the subject in Skogestad & Postlethwaite (2005) and in works such as Qiu & Davison (1993) should be used as a starting point for predicting and evaluating performance limitations of models. Based on this work, techniques that guide actuator and sensor placement for control design may be developed. In existing systems where the placement is limited by other factors, these techniques may help predict the system performance and make the control designer aware of the limitations.

Another important property of both models and compensators is their robustness. In this thesis we have considered only localized perturbations of known structure introduced at a particular location and a particular point in time, i.e., as initial conditions to the problem. The performance of reduced-order compensators subject to additional disturbances, such as free stream turbulence in boundary layers, or surface roughness on the walls, was not investigated. This would be an important further step for the practical application of controllers designed using the models developed in this thesis. Balanced models were shown to be promising in this

respect, as demonstrated by the performance of the models in Chapter 4 at off-design Reynolds number.

Edge of chaos. Recent work on computation of edge states in turbulent flows (see, for example, Schneider *et al.*, 2008) has shown that such states exist in shear flows. Low-dimensional descriptions of such states can be given by reduced-order models(Kim & Moehlis, 2008) . The edge of chaos may be of interest to the control designer, since it may indicate regions in the state space where controllers will be successful, thus guiding control design.

Appendix A

BPOD using a continuous adjoint

When computing the exact balanced truncation, the balancing transformation is found from the eigenvalue problem $W_c W_o T = T\Sigma^2$ where W_c and W_o are the controllability and observability Gramians with $A^+ = A^T$. We show here that, although the product $W_c W_o$ does not depend on the inner product on the state space used to define the adjoint system, the appropriate weight M must be included in the computations via the method of snapshots.

We can represent the weighted inner product of two vectors q_1 and q_2 as:

$$\langle q_1, q_2 \rangle_M = \int_\Omega q_1^* M q_2 \, d\Omega, \tag{A.1}$$

where the domain of integration Ω is the Hilbert space itself. The star denotes the complex conjugate transpose. The inner product weight M is part of the definition of the Hilbert space itself. We define the so-called continuous adjoint of an operator A with respect to this inner product as:

$$\langle Aq_1, q_2 \rangle_M = \left\langle q_1, A^+ q_2 \right\rangle_M . \tag{A.2}$$

We use the symbol + in order to distinguish the adjoint from the standard matrix transpose A^T. From this definition it is easily shown that the adjoint system (2.14) is derived from the system (2.10) as follows:

$$\begin{aligned} \langle Ax, z \rangle_M = \left\langle x, A^+ z \right\rangle_M \quad &\Rightarrow \quad A^+ = M^{-1} A^T M \\ \langle Bu, x \rangle_M = \left\langle u, B^+ x \right\rangle \quad &\Rightarrow \quad B^+ = B^T M \\ \langle Cx, y \rangle = \left\langle x, C^+ y \right\rangle_M \quad &\Rightarrow \quad C^+ = M^{-1} C^T . \end{aligned} \tag{A.3}$$

In the above, we have assumed that the input and output spaces use the standard (unweighted) inner product. Next, we obtain for the Gramians:

$$\begin{aligned} G_c &= \int_0^\infty e^{At} B B^+ e^{A^+ t} dt \\ &= \int_0^\infty e^{At} B B^T M M^{-1} e^{A^T t} M dt \\ G_o &= \int_0^\infty e^{A^+ t} C^+ C e^{At} dt \\ &= \int_0^\infty M^{-1} e^{A^T t} M M^{-1} C^T C e^{At} dt, \end{aligned} \tag{A.4}$$

where G_c and G_o denote the Gramians obtained with the weighted inner product. Since the matrices M and M^{-1} are constant, we can take them out of the integrals, obtaining:

$$G_c G_o = W_c W_o. \tag{A.5}$$

Thus, we have shown that balanced truncation does not depend on the choice of the inner product used to derive the adjoint system, and this allows us to use a convenient inner product. (In numerical simulations the 'simple' discrete adjoint $A^+ = A^T$ may in fact be more difficult to compute than a continuous adjoint which may retain a similar form of the equations; for instance, this is the case for linearized channel flow).

Next, we consider the computation of balancing and adjoint modes via the method of snapshots. From the definition of the empirical Gramians (Eq. (2.16)) it is easily shown that $Y^+ = Y^T M$ (recall that the snapshots of the adjoint simulations, which are the columns of Y, are given by $z(t) = e^{A^+ t} C^+$). Thus, we can write the SVD in Eq. (2.17) as:

$$Y^T M X = U \Sigma V^T. \tag{A.6}$$

If we define the inverse of the balancing transformation as $\Psi^T = \Sigma^{-1/2} U^T Y^T$ we can easily compute the adjoint modes just from the SVD and from the adjoint snapshots. Recall that the *columns* of Ψ give the adjoint modes. The two sets of modes will now be bi-orthogonal with respect to the M inner product, so that $\Psi^T M \Phi = I$.

An alternative, more intuitive explanation is that, since both the direct and the adjoint snapshots 'live' in the state space, the correct inner product is that including the weight M (which is a part of the definition of the Hilbert space in which they reside). It is therefore this weighted inner product that should be used for forming the matrix for the SVD. Furthermore, the balancing and adjoint modes are bi-orthogonal with respect to this weighted inner product.

Appendix B

Validation of turbulent channel DNS code

Here we summarize the definitions of relevant Reynolds numbers and their relationship for laminar and turbulent channel flow, the non-dimensionalization used in the simulations, the driving of the flow in simulations, as well as some useful quantities in laminar flow that are useful when investigating the linear evolution of perturbations. Finally, some results of converged turbulent runs using the DNS code discussed in Chapter 3 are given.

B.1 Reynolds number definitions and relations

The Reynolds number is defined as the ratio of the inertial to viscous forces in the flow:

$$Re = \frac{\rho U D}{\mu}, \tag{B.1}$$

where ρ is the fluid density, U is a characteristic velocity and D is a characteristic length scale. The length scale for channel flow is typically taken to be half of the channel width, $\delta = 1$, but different choices of velocity can be made. The Reynolds numbers used are:

- *Centerline Reynolds number*, Re_c, which is defined as:

  $$Re_c = \frac{\rho U_c D}{\mu},$$

  where Re_c is the centerline velocity of the flow.

- *Frictional Reynolds number*, Re_τ, defined as:

  $$Re_\tau = \frac{\rho u_\tau D}{\mu},$$

where u_τ is a *frictional velocity*, defined as:

$$u_\tau = \sqrt{\frac{\tau_w}{\rho}},$$

with the wall shear defined as $\tau_w = \mu \frac{\partial u}{\partial y}|_{wall}$, where y is the wall-normal direction.

It can easily be shown that the relationship between Re_c and Re_τ for laminar flow is:

$$Re_\tau = \sqrt{2Re_c}.$$

The relationship for turbulent channel flow will be discussed below.

B.2 Nondimensionalization

The non-dimensionalized momentum equations were derived in Section 3.1.1. We note here that an alternative non-dimensionalization is often used for turbulent flow (Kim *et al.*, 1987), where the characteristic velocity is u_τ defined above, and the characteristic length scale is still the half-thickness of the channel δ:

$$u = \frac{\tilde{u}}{u_\tau}, \quad x = \frac{\tilde{x}}{\delta}, \quad t = \tilde{t}\frac{u_\tau}{\delta}.$$

The relevant Reynolds number in Kim *et al.* (1987) is thus Re_τ, and the non-dimensionalized equations are:

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{1}{Re_\tau}\frac{\partial^2 u_i}{\partial x_j x_j}. \tag{B.2}$$

Here, however, we will use the non-dimensionalization by U_c and δ. We can define the velocity field as:

$$u(t) = \bar{u} + u'(t), \tag{B.3}$$

where $\bar{u}$ is the mean velocity. In this context, the averaging is over both x, z and time, i.e., it is an ensemble average, as discussed by Pope (2000), and $u'(t)$ is the fluctuation part. The overbar over any other quantity also indicates this same ensemble average.

B.3 Driving the flow

B.3.1 Constant pressure gradient

One way to drive the flow is to use a constant pressure gradient in the streamwise direction. At the wall, the mean wall shear stress is exactly balanced by the pressure

gradient in the streamwise direction, so keeping a constant pressure gradient imposes a constant value of Re_τ.

In order to show the relation between the pressure gradient and the wall shear stress, we start with the Reynolds-averaged momentum equation in the x-direction:

$$\overline{u}\frac{\partial \overline{u}}{\partial x} + \overline{v}\frac{\partial \overline{u}}{\partial y} + \overline{w}\frac{\partial \overline{u}}{\partial z} = -\frac{\partial \overline{p}}{\partial x} - \left(\frac{\partial \overline{u'^2}}{\partial x} + \frac{\partial \overline{u'v'}}{\partial y} + \frac{\partial \overline{u'w'}}{\partial z}\right) + \frac{1}{Re_c}\nabla^2\overline{u}. \tag{B.4}$$

This is the equation for the mean flow, but involves time averages of fluctuations. For channel flow, all terms on the LHS are zero, since there is no mean spanwise or wall-normal velocity, and no streamwise dependence of $\overline{u}$. Since there is also no variation in x and z of the fluctuation terms, the equation reduces to:

$$0 = -\frac{\partial \overline{p}}{\partial x} - \frac{\partial \overline{u'v'}}{\partial y} + \frac{1}{Re_c}\nabla^2\overline{u}. \tag{B.5}$$

It is shown by (Pope, 2000) that the shear stress can be written as:

$$\tau(y) = \frac{1}{Re_c}\frac{d\overline{u}}{dy} - \overline{u'v'}, \tag{B.6}$$

so that Eq. (B.5) can be written as:

$$\frac{d\tau}{dy} = \frac{d\overline{p}}{dx}. \tag{B.7}$$

It now becomes apparent that the shear stress in the flow is exactly balanced by the pressure gradient, and fixing one of the two quantities will determine the other one. Also, both sides have to be constant since τ depends only on y and $\overline{p}$ depends only on x. It is also shown by Pope that the pressure gradient is uniform across the flow (including, of course, the wall). It can also be shown that the shear stress is given by:

$$\tau(y) = \tau_w(1 - y), \tag{B.8}$$

where τ_w is the shear stress at the lower wall ($y = -1$). Also, at the wall, using our non-dimensionalization,

$$-\tau_w = \frac{d\overline{p}}{dx}. \tag{B.9}$$

Alternatively, this relationship may be derived from a momentum balance in the streamwise direction. Recalling the definition of u_τ, and remembering that as part of the non-dimensionalization we divided the Navier-Stokes equations by ρ, we now have:

$$u_\tau = \sqrt{\tau_w}. \tag{B.10}$$

Finally, it follows from the definitions of Re_c and Re_τ that, for flow forced at constant pressure gradient:

$$Re_\tau = Re_c\sqrt{-\frac{d\overline{p}}{dx}}, \tag{B.11}$$

which for laminar flow can be shown to be exactly $Re_\tau = \sqrt{2Re_c}$.

B.3.2 Constant mass flux

A constant mass flux (corresponding to a constant bulk velocity) can also be used to drive the flow. This type of simulation converges to a statistically steady state much faster than a simulation with constant pressure gradient, as will be shown below. In this case, one needs to solve for the pressure gradient that would be required to impose the desired mass flux at each time step. Thus, the relation (B.11) is not useful here due to the varying pressure gradient. For this case, there is a correlation given in (Chevalier *et al.*, 2007):

$$Re_\tau \approx 0.116 Re_c^{0.88}. \tag{B.12}$$

This is a modified form of a correlation given in Pope (2000), which in turn is based on the correlation of Dean (1978). Here Re_c is the centerline Reynolds number of the *laminar flow* with the desired mass flux. The true (turbulent) centerline Reynolds number is, like Re_τ, also a result of the simulation, and typically it is about 0.77 of Re_c (Philipp Schlatter, private communication).

Since the streamwise mass flux $\dot{m}$ is a linear function of the pressure gradient, the pressure gradient is updated in the DNS at each time step as follows.

1. Calculate the mass flux $\dot{m}_1$ with the current pressure gradient $(dp/dx)_1$.
2. Calculate the mass flux $\dot{m}_2$ with twice the current pressure gradient. The factor of two is arbitrarily chosen, since the mass flux is a linear function of the pressure gradient.
3. Calculate the pressure gradient for the next step necessary to enforce the desired ('target') mass flux $\dot{m}_T$ from

$$\frac{dp}{dx} = \left(\frac{dp}{dx}\right)_1 \left(1 + \frac{\dot{m}_T - \dot{m}_1}{\dot{m}_2 - \dot{m}_1}\right).$$

B.3.3 Summary

It should be kept in mind that, for constant-pressure gradient flow, there is one 'free parameter' for a channel flow simulation out of the three values: Re_c, Re_τ and the pressure gradient. That means that choosing a combination of any two determines the third one from Eq. (B.11). Typically one chooses Re_τ and Re_c. On the other hand, for constant mass flux, the two parameters are the mass flux and Re_τ. In this case, in a converged simulation Re_τ is a statistical average, which can be computed from simulation data, or alternatively from the correlation (B.12), which should be used to verify simulation results. Re_c is the laminar value which directly corresponds to the chosen mass flux, but the turbulent centerline Reynolds number can also be computed from a converged simulation, and is usually 0.77 of Re_c.

B.4 Laminar flow

It is easily shown using (B.4) that the laminar velocity profile is:

$$U(y) = -\frac{dp}{dx}\frac{Re_c}{2}(1-y^2). \tag{B.13}$$

Now we can define some relevant quantities that are useful for simulations:

- The *centerline velocity* U_c. When constant mass flux is imposed, the value is typically chosen so that $U_c = 1$ for laminar flow with that mass flux (U_c will fluctuate in a turbulent simulations), which corresponds to setting the initial pressure gradient to $-\frac{dp}{dx} = \frac{2}{Re_c}$ (this is just a guess, the pressure gradient will adjust in order to satisfy the mass flux requirement).

- The *half-width* δ is the half of the distance between the channel walls. This is typically set to $\delta = 1$. Note that this is a non-dimensional δ (abuse of notation here since δ was used above).

- The *mass flux* for the non-dimensionalized equations is defined as:

$$\dot{m} = \frac{1}{2L_xL_z}\int_0^{L_x}\int_0^{L_z}\int_{-1}^{1} U(x,y,z)\,dx\,dy\,dz = \frac{1}{2}\left|y - \frac{y^3}{3}\right|_{-1}^{1} = 4/3. \tag{B.14}$$

  So, for laminar flow with $U_c = 1$ and $\delta = 1$ the mass flux is 1.333333333.

- The *laminar flow energy* is defined as the kinetic energy of the flow per unit volume:

$$E_K = \frac{1}{2L_xL_z}\int_0^{L_x}\int_0^{L_z}\int_{-1}^{1} U^2(x,y,z)\,dx\,dy\,dz = \frac{1}{2}\left|y - \frac{2y^3}{3} + \frac{y^5}{5}\right|_{-1}^{1} = 4/15. \tag{B.15}$$

  So, for laminar flow with $U_c = 1$ and $\delta = 1$ the kinetic energy per unit volume is 0.266666667.

- The *bulk velocity* U_{bulk} would be the velocity for a uniform profile ($U = const$) with the same mass flux as the laminar profile. The bulk velocity is defined as:

$$U_{bulk} = \frac{1}{2\delta}\int_{-1}^{1} U(y)\,dy = 2/3. \tag{B.16}$$

  So, for laminar flow with $U_c = 1$ and $\delta = 1$, U_{bulk} is 0.666666667. In turbulent channel simulations, setting $U_{bulk} = 2/3$ is equivalent to setting the mass flux to be 4/3.

- Another quantity that is often used is the *bulk Reynolds number*:

$$Re_b = \frac{\rho U_{bulk} D}{\mu},$$

  which for laminar flow is easily shown to be $Re_b = 2/3 Re_c$.

B.5 Example results

We show results of two simulations, both with box size $4\pi \times 2 \times 2\pi$ and $Re_c = 3300$, one ran at constant mass flux of 4/3 and the other one at fixed dp/dx corresponding to $Re_\tau = 180$. In both cases the initial condition is the perturbation from Chapter 5 at an amplitude above its transition threshold.

B.5.1 Constant mass flux

For the constant mass flux simulation (Fig. B.1), it was found that Re_τ is in excellent agreement with the correlation (B.12), which predicts $Re_\tau = 144.79$. The averaging is performed between t=400 and t=774.2, with a total of 1871 snapshots. The turbulent Re_c was found to be $Re_{c,t} \approx 2584$, or 0.78 of Re_c, as predicted.

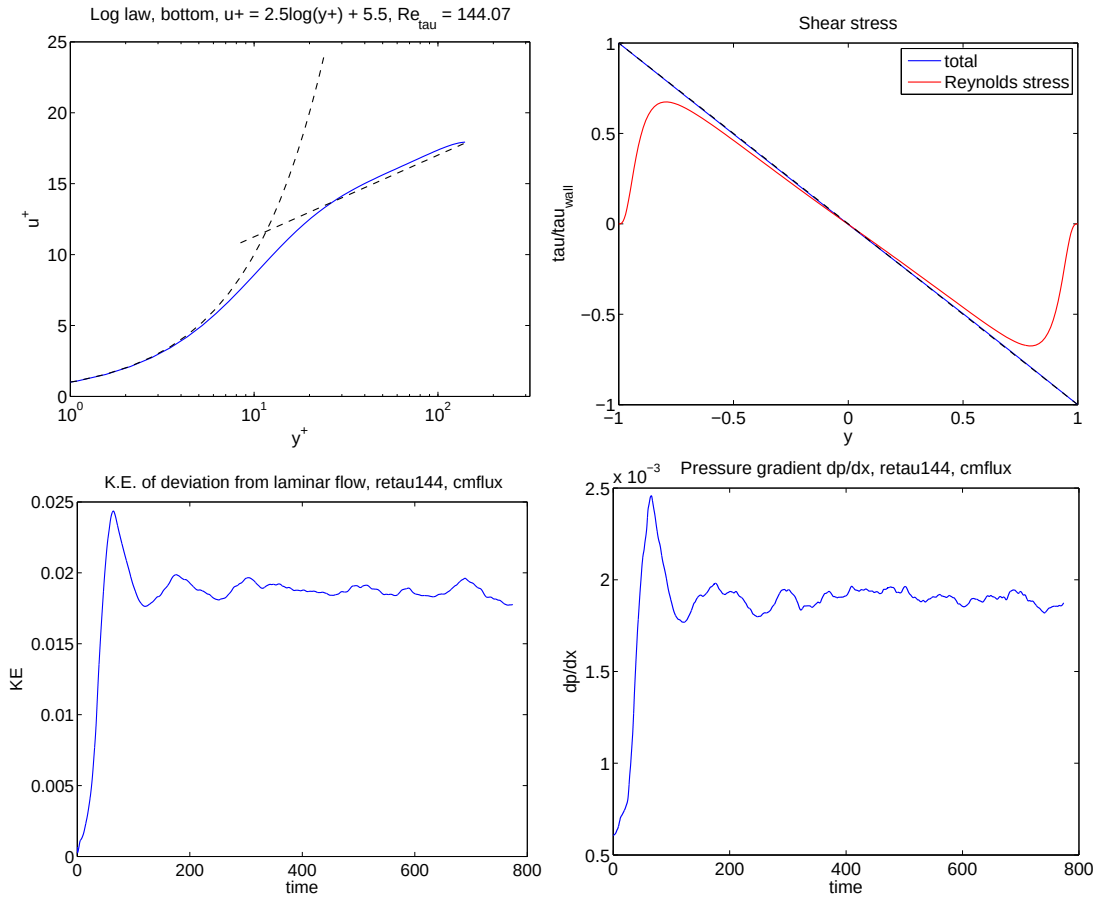


Figure B.1: Log law, shear stress, perturbation energy and pressure gradient history for constant mass flux value of 4/3 at $Re_c = 3300$. The log law curves for the top (not shown) and bottom walls are identical, and $Re_\tau = 144.07$ for both. The perturbation energy (lower left panel) and pressure gradient (lower right panel) history indicates that the simulation quickly converges to a mean energy and pressure gradient values.

B.5.2 Constant pressure gradient

The constant pressure gradient simulation takes a considerably longer time to converge. The value of Re_τ in the example shown has leveled off at approximately $Re_\tau \approx 180$ after 800 time units. Plots of the log law and total and Reynolds stresses is shown in Fig. B.2. The averaging in this figure is performed between $t = 400$ and $t = 800$, with a total of 1000 snapshots. The pressure gradient here is $(180/3300)^2 = 0.002975$.

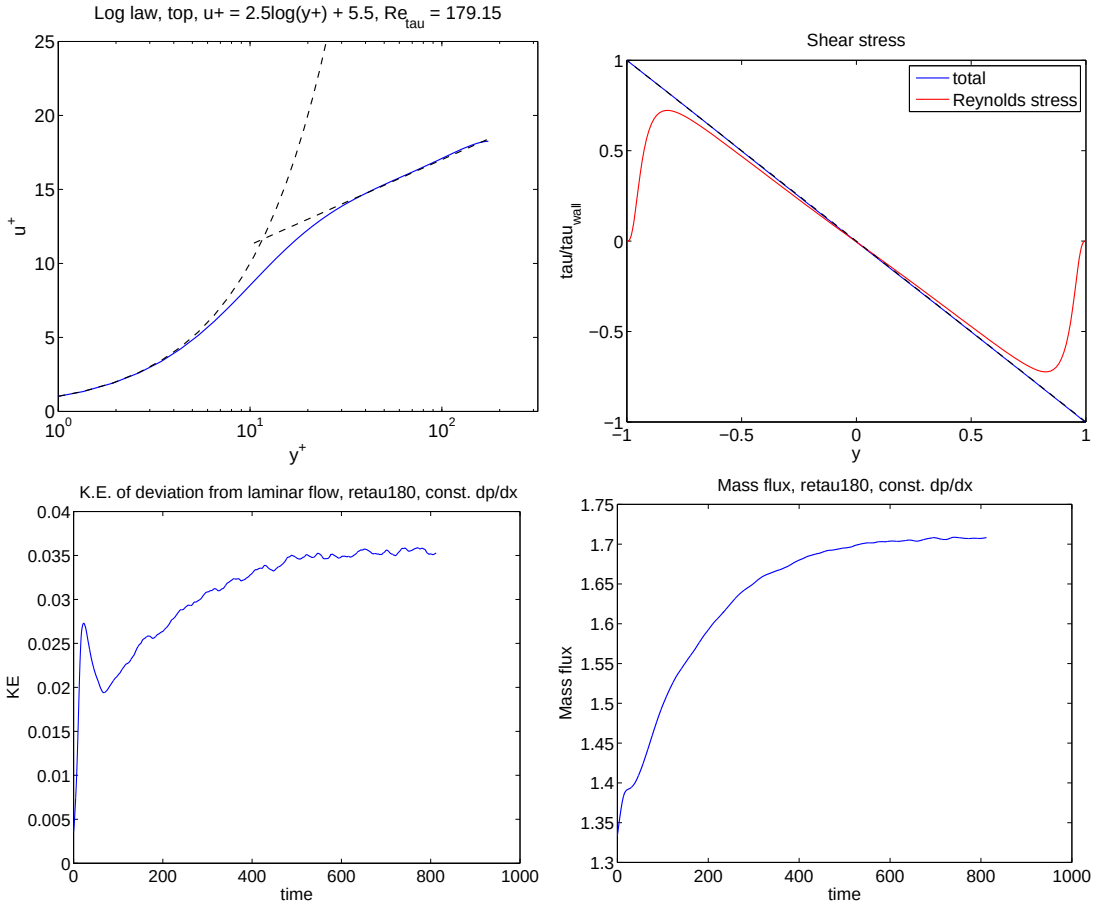


Figure B.2: Log law, shear stress, energy and mass flux history for constant pressure gradient at $Re_c = 3300$. The bottom log law curve (not shown) is almost identical to the top one, and the corresponding $Re_\tau = 178.85$ for the bottom. The perturbation energy (lower left panel) and mass flux (lower right panel) histories show that the simulation needs a much longer time to converge.

B.5.3 Comparison

We next compare the Reynolds stresses and velocity profiles in inner coordinates for the two cases (see Fig. B.3). Fig. B.4 compares the root-mean-square fluctuations for the two runs, and the one for $Re_\tau = 180$ matches Fig. 6 in Kim *et al.* (1987) really well.

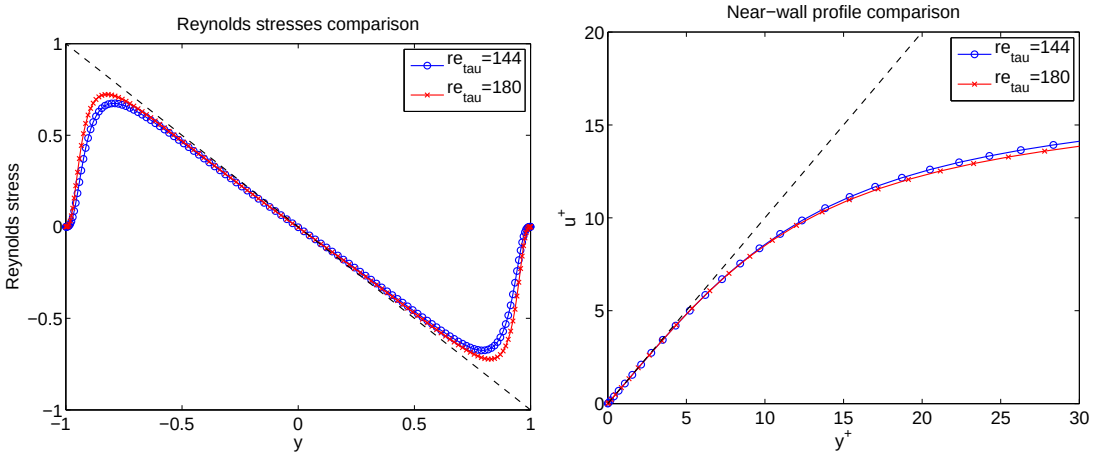


Figure B.3: Left: comparison of Reynolds stresses for the two runs, constant mass with $Re_\tau = 144$ and constant pressure gradient with $Re_\tau = 180$. Right: comparison of velocity profiles for the two simulations in inner coordinates.

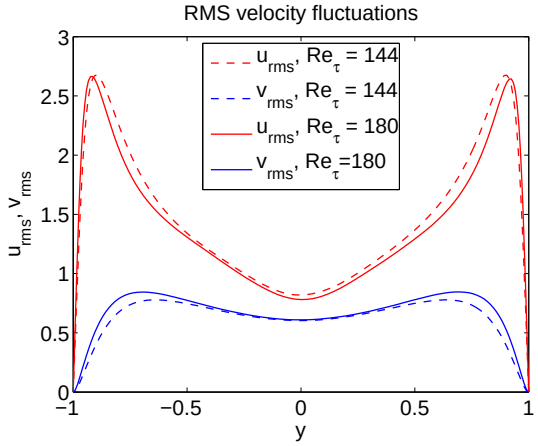


Figure B.4: RMS velocity fluctuations for the two cases, constant mass flux with $Re_\tau = 144$ and constant pressure gradient with $Re_\tau = 180$. v_{rms} appears to have higher peaks for $Re_\tau = 180$ than for $Re_\tau = 144$, which can be expected for the higher value of Re_τ.

Appendix C

Validation of POD computations

Here we describe the verification of convergence of POD modes and eigenvalues for the localized perturbation modeled in Section 4.3. Fig. C.1 shows the leading POD eigenvalues for four different choices of snapshot number and spacing. The largest computation was for 3000 evenly spaced snapshots over 1200 dimensionless time units. The first few hundred POD eigenvalues are identical to the POD eigenvalues obtained from 2000 evenly spaced snapshots (not shown), which establishes convergence of the computation. Thus, the results using 3000 snapshots will be regarded as the 'true' eigenvalues and modes.

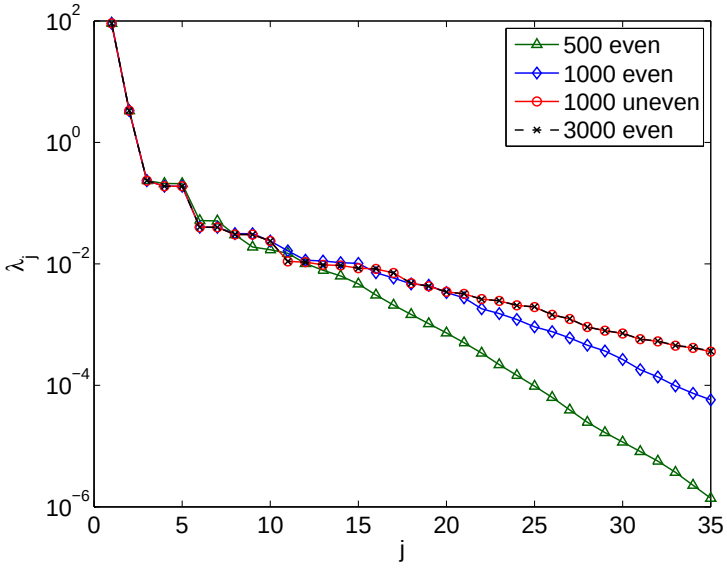


Figure C.1: Convergence of POD eigenvalues for a perturbation.

While the modes from the calculation with the highest available resolution should be used whenever possible, if computations are to be repeated, for example for different perturbations with very similar dynamics, or using with a different inner

product, it may be desirable to take a smaller number of snapshots. We next study the possibility for this type of computational savings.

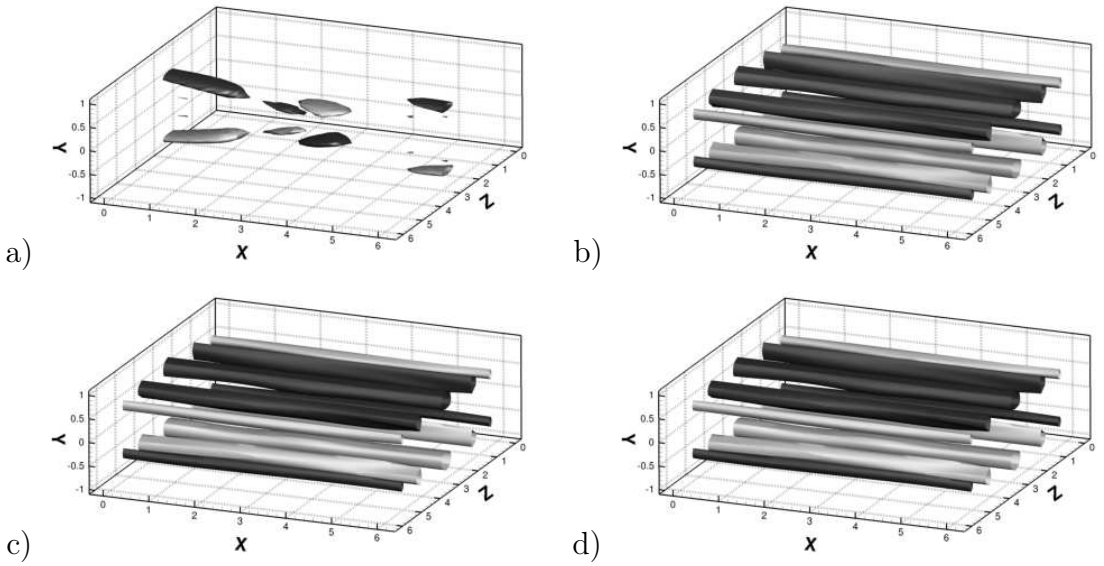


Figure C.2: A comparison of the tenth POD mode (streamwise velocity) for the four computations. a) 500 snapshots, b) 1000 snapshots with even spacing, c) 1000 snapshots with uneven spacing, d) 3000 snapshots.

Although the integral of the kinetic energy over time, which is equal to the sum of the POD eigenvalues, is almost equal for all four simulations (within about 10^{-3} %), the eigenvalues are different. Taking five hundred snapshots turns out to be insufficient to match the finest computation. Taking a thousand evenly spaced snapshots results in better match of the first nine eigenvalues, but does not converge for higher modes. These higher modes are desirable if one is to compute models of higher rank. Savings can still be achieved by using unevenly spaced snapshots, if the dynamics of the perturbation is known, as in this case, where there is fast initial dynamics due to traveling structures which grow into streamwise structures. Thus, taking smaller snapshot spacing during the initial transient and larger spacing during the later part may be able to capture all the relevant energy structures. Fig. C.1 shows that this is indeed the case. The 'uneven' spacing is 500 snapshots with a gap of 0.2 time units, for the first 100 time units and 500 snapshots with a gap of 2.2 time units for the period, during which the perturbation dynamics is a fairly slow growth and decay. The slower decay of eigenvalues for the computations with more snapshots indicates that more energetically significant modes are captured in the basis than in the case for a small number of snapshots.

Fig. C.2 shows the streamwise velocity for the tenth mode in the same four computations of Fig. C.1. For 500 equally spaced snapshots, the tenth mode looks completely different from the correct one, indicating that the flow structure in this computation are not captured in the correct order of importance according

to kinetic energy. For 1000 evenly spaced snapshots, the mode is visually almost indistinguishable from the 'true' mode, but the eigenvalue still has a different value, as seen in Fig. C.1.

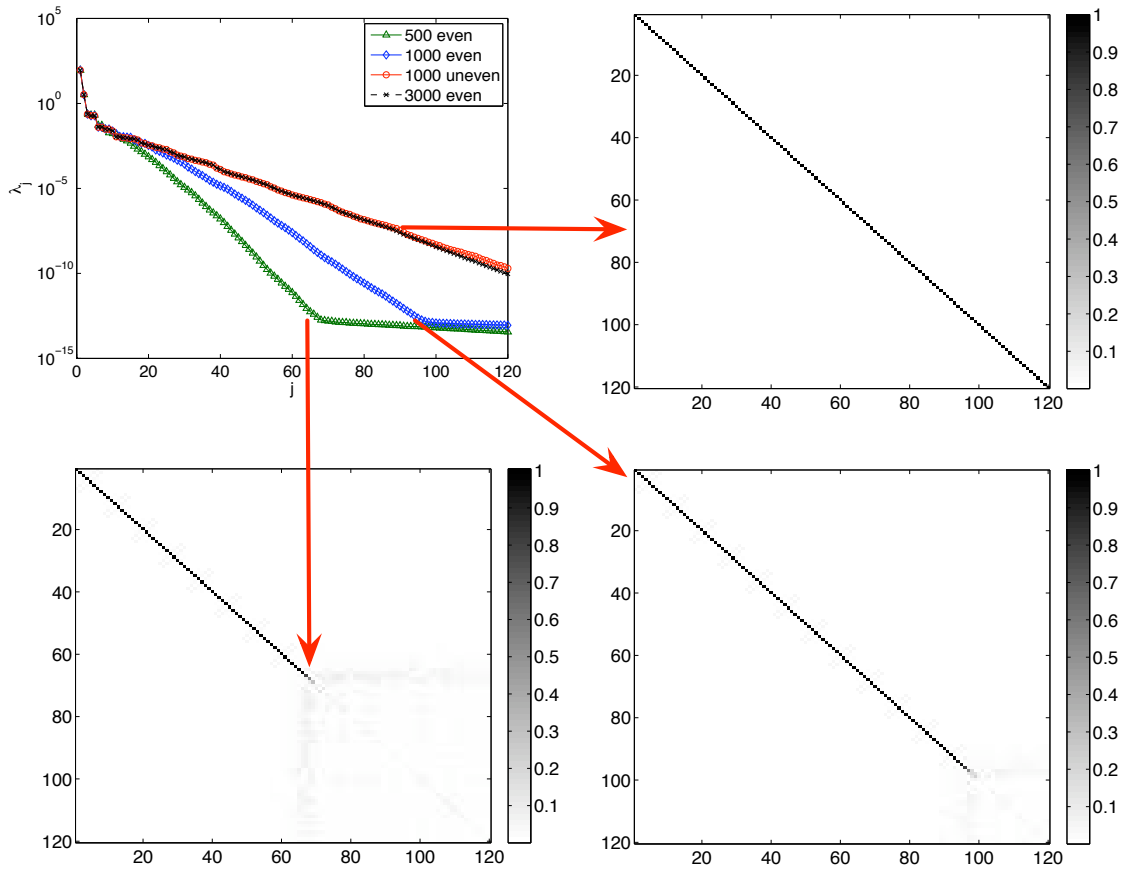


Figure C.3: The eigenvalues for the four calculations for up to $j = 120$, and the corresponding orthogonality matrices, as indicated by the arrows. The unresolved computations result in a fewer number of 'useful' modes due to numerical roundoff error buildup for very small values of λ_j.

The orthogonality of the modes was also verified. Fig. C.3 shows the eigenvalues for the four calculations again (plotted up to a higher j) together with the absolute values of the entries of the corresponding matrix O for which $O_{ij} = \langle \theta_i, \theta_j \rangle$ for each computation. The POD modes should be orthogonal as long as the corresponding eigenvalues are larger than about $\lambda_1/10^{15}$, beyond which numerical roundoff errors start to dominate the computations. The orthogonality 'breaks' first for the case with 500 snapshots, since fewer structures are captured in the POD modes and the eigenvalues decay faster. We see that for 1000 uneven snapshots, the first 120 modes are perfectly orthogonal and thus useful for computing reduced-order models or output projection in BPOD. In most cases, however, the use of more than 50-60 POD modes for either task is not practical, as the resulting models are not truly low-dimensional anymore.

Appendix D

Transfer function norms for linear systems

Here we define the transfer function norms mentioned in Chapter 2 and used in Chapter 4 for evaluating model performance. For an in-depth discussion, see Dullerud & Paganini (1999) or Zhou *et al.* (1996). The *transfer function* of a control system is simply a function relating the inputs and outputs of the system. For a single-input single-output (SISO) system, it is defined as:

$$G(s) = \frac{\hat{y}(s)}{\hat{u}(s)}, \tag{D.1}$$

where $\hat{y}(s)$ and $\hat{u}(s)$ are the output and the input, respectively, after the Laplace transform of the system has been taken. The absolute value of the transfer function for a SISO system can be thought of as the gain of the system in control theory terms. It is easily shown that for the state space system given by:

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du, \end{aligned} \tag{D.2}$$

the transfer function is defined by:

$$G(s) = C(sI - A)^{-1}B + D. \tag{D.3}$$

In this thesis we do not deal with the feed-through term D, which has been included for completeness here. For a multiple-input multiple-output (MIMO) system, the transfer function is a thus a matrix with as many rows as the number of outputs and as many columns as the number of inputs. The following norms of this matrix can be computed:

- The 2-norm $\|G(s)\|_2$ is also known as the H_2 norm in control system terminology. For any matrix G the 2-norm is defined by:

$$\|G\|_2 = \sqrt{max(\lambda(G^T G))}, \tag{D.4}$$

where λ are the eigenvalues of the matrix. For stable systems, this norm physically corresponds to integrating the impulse response in time from $t = 0$ to $t = \infty$, which is why we find it useful for evaluating the capturing of the impulse response in Chapter 4. We also use the error norm between two systems, $\|G_1(s) - G_2(s)\|_2$, which in the physical domain corresponds to the integral in time of the difference between the outputs of two systems with transfer functions $G_1(s)$ and $G_2(s)$, in order to evaluate how close to each other two systems are. In our case, we compare the transfer function of the full system to the reduced-order model. We note that, in order for this error norm to be defined, the number of outputs of the two systems that are being compared needs to be the same, which is true for the reduced-order models in Chapter 4 and the output-projected system.

- The infinity norm $\|G(s)\|_\infty$ is also known as the H_∞ norm in control system terminology. By definition, the infinity norm of any matrix is defined as the maximum singular value of the matrix. Physically, it corresponds to the magnitude of the response of the system at the frequency where its frequency response has the peak, i.e. this is the maximum value of the transfer function over all possible frequencies. The error norm $\|G_1(s) - G_2(s)\|_\infty$ is analogous to the 2-error norm.

The resolvent norm for a linear system whose dynamics is governed by a matrix A is defined as $\|(sI - A)^{-1}\|_2$, where $s = j\omega$, ω being the frequency of forcing. This quantity was used by Schmid & Henningson (2001) to study the behavior of perturbations at different frequencies, but it is not a transfer function norm, as it does not include the effects of the B and C matrices. In Chapter 4 we study the true frequency response of linear channel flow, corresponding to the actual transfer function.

Bibliography

Åkervik, E., Hœpffner, J., Ehrenstein, U. & Henningson, D. S. 2007 Optimal growth, model reduction and control in a separated boundary-layer flow using global eigenmodes. *Journal of Fluid Mechanics* **579**, 305–314.

Aamo, O. M., Krstić, M. & Bewley, T. R. 2003 Control of mixing by boundary feedback in 2d channel flow. *Automatica* **39**, 1597–1606.

Aamo, O. M., Smyshlyaev, A. & Krstić, M. 2005 Boundary control of the linearized Ginzburg-Landau model of vortex shedding. *SIAM Journal on Control and Optimization* **43** (6), 1953–1971.

Ahuja, S. & Rowley, C. W. 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. W. 2009 Feedback control of unstable steady states of flow past a flat plate using reduced-order estimators. *Journal of Fluid Mechanics* (submitted).

Aubry, N., Holmes, P., Lumley, J. L. & Stone, E. 1988 The dynamics of coherent structures in the wall region of a turbulent boundary layer. *Journal of Fluid Mechanics* **192**, 115–173.

Baggett, J. S. & Trefethen, L. N. 1997 Low-dimensional models of subcritical transition to turbulence. *Physics of Fluids* **9** (4).

Bagheri, S., Åkervik, E., Brandt, L. & Henningson, D. S. 2009*a* Matrix-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. *Journal of Fluid Mechanics* **620**, 263–298.

Bagheri, S., Henningson, D. S., Hœpffner, J. & Schmid, P. J. 2009*c* Input-output analysis and control design applied to a linear model of spatially developing flows. *Applied Mechanics Reviews* **62** (2).

Bamieh, B. & Dahleh, M. 2001 Energy amplification in channel flows with stochastic excitation. *Physics of Fluids* **13** (11), 3258–3269.

Barbagallo, A., Sipp, D., Jacquin, L. & Schmid, P. J. 2008 Control of an incompressible cavity flow using a reduced order model based on global modes. AIAA Paper 2008-3904.

Bewley, T. R. 2001 Flow control: new challenges for a new renaissance. *Progress in Aerospace Sciences* **37**, 21–58.

Bewley, T. R. & Liu, S. 1998 Optimal and robust control and estimation of linear paths to transition. *Journal of Fluid Mechanics* **365**, 305–349.

Bewley, T. R. & Protas, B. 2004 Skin friction and pressure: the "footprints" of turbulence. *Physica D. Nonlinear Phenomena* **196**, 28–44.

Butler, K. & Farrell, B. 1992 Three-dimensional optimal perturbations in viscous shear flow. *Physics of Fluids A* **4**, 1637–1650.

Camphouse, R. C. 2005 Boundary feedback control using proper orthogonal decomposition models. *Journal of Guidance, Control, and Dynamics* **28** (5), 931–938.

Chevalier, M., Hœpffner, J., Bewley, T. R. & Henningson, D. S. 2006 State estimation in wall-bounded flow systems. Part 2. Turbulent flows. *Journal of Fluid Mechanics* **552**, 167–187.

Chevalier, M., Schlatter, P., Lundbladh, A. & Henningson, D. S. 2007 SIMSON: A pseudo-spectral solver for incompressible boundary layer flows. Technical Report TRITA-MEK 2007:07. KTH Mechanics.

Choi, H., Moin, P. & Kim, J. 1994 Active turbulence control for drag reduction in wall-bounded flows. *Journal of Fluid Mechanics* **262**, 75–110.

Chomaz, J. 2005 Global instabilities in spatially developing flows: Non-normality and nonlinearity. *Annual Review of Fluid Mechanics* **37**, 357–392.

Cochran, J. & Krstić, M. 2009 Motion planning and trajectory tracking for three-dimensional Poiseuille flow. *Journal of Fluid Mechanics* **626**, 307–332.

Cochran, J., Vazquez, R. & Krstić, M. 2006 Backstepping boundary control of Navier-Stokes channel flow: a 3D extension. In *Proceedings of the 2006 American Control Conference*.

Cohen, K., Siegel, S., McLaughlin, T., Gillies, E. & Myatt, J. 2005 Closed-loop approaches to control of a wake flow modeled by the Ginzburg-Landau equation. *Computers and Fluids* **34**, 927–949.

Corke, T., Post, M. & Orlov, D. 2009 Single dielectric barrier discharge plasma enhanced aerodynamics: physics, modeling and applications. *Experiments in Fluids* **46**, 1–26.

Cossu, C. & Chomaz, J. 1997 Global measures of local convective instabilities. *Physical Review Letters* **78** (23), 4387–4390.

Dean, R. 1978 Reynolds number dependence of skin friction and other bulk flow variables in two-dimensional rectangular duct flow. *Journal of Fluids Engineering* **100**, 215–223.

Doyle, J. C. 1978 Guaranteed margins for LQG regulators. *IEEE Transactions on Automatic Control* **23** (4), 756–757.

Doyle, J. C. & Stein, G. 1979 Robustness with observers. *IEEE Transactions on Automatic Control* **24** (4), 607–611.

Dullerud, G. E. & Paganini, F. 1999 *A Course in Robust Control Theory: A Convex Approach*, *Texts in Applied Mathematics*, vol. 36. Springer-Verlag.

Eisenhower, B. & Mezić, I. 2007 A mechanism for energy transfer leading to conformation change in networked nonlinear systems. 46th IEEE Conference on Decision and Control.

Ellingsen, T. & Palm, E. 1975 Stability of linear flow. *Physics of Fluids* **18** (4), 487–488.

Everson, R. & Sirovich, L. 1995 Karhunen-Loève procedure for gappy data. *Journal of the Optical Society of America A* **12** (8), 1657–1664.

Farrell, B. 1988 Optimal excitation of perturbations in viscous shear flow. *Physics of Fluids* **31**, 2093–2102.

Farrell, B. F. & Ioannou, P. J. 1993 Stochastic forcing of the linearized Navier-Stokes equations. *Physics of Fluids A* **5** (11), 2600–2609.

Farrell, B. F. & Ioannou, P. J. 1996 Turbulence suppression by active control. *Physics of Fluids* **8** (5), 1257–1268.

Farrell, B. F. & Ioannou, P. J. 2001 Accurate low-dimensional approximation of the linear dynamics of fluid flow. *Journal of the Atmospheric Sciences* **58**, 2771–2789.

Friedland, B. 1986 *Control system design: An Introduction to State-Space Methods*. McGraw-Hill.

Frigo, M. & Johnson, S. G. 2005 The design and implementation of FFTW3. *Proceedings of the IEEE* **93** (2), 216–231, special issue on "Program Generation, Optimization, and Platform Adaptation".

Fujimoto, K., Scherpen, J. M. A. & Gray, W. 2002 Hamiltonian realizations of nonlinear adjoint operators. *Automatica* **38**, 1769–1775.

Fujimoto, K. & Tsubakino, D. 2008 Computation of nonlinear balanced realization and model reduction based on taylor series expansion. *Systems and Control Letters* **57**, 283–289.

Green, M. A., Rowley, C. W. & Haller, G. 2007 Detection of Lagrangian coherent structures in 3D turbulence. *Journal of Fluid Mechanics* **572**, 111–120.

Gustavsson, L. H. 1986 Excitation of direct resonances in plane Poiseuille flow. *Studies in Applied Mathematics* **75**, 227–248.

Gad-el Hak, M. 2000 *Flow Control: Passive, Active, and Reactive Flow Management*. Cambridge University Press.

Henningson, D. S. & Åkervik, E. 2008 The use of global modes to understand transition and perform flow control. *Physics of Fluids* **20** (031302).

Henningson, D. S., Lundbladh, A. & Johansson, A. V. 1993 A mechanism for bypass transition from localized disturbances in wall-bounded shear flows. *Journal of Fluid Mechanics* **250**, 169–207.

Henningson, D. S. & Schmid, P. J. 1992 Vector eigenfunction expansions for plane channel flows. *Studies in Applied Mathematics* **87**, 15–43.

Hœpffner, J. 2006 Stability and control of shear flows subject to stochastic excitations. PhD thesis, Royal Institute of Technology, Stockholm.

Hœpffner, J., Chevalier, M., Bewley, T. R. & Henningson, D. S. 2005 State estimation in wall-bounded flow systems. Part 1. Perturbed laminar flows. *Journal of Fluid Mechanics* **534**, 263–294.

Högberg, M., Bewley, T. R. & Henningson, D. S. 2003 Linear feedback control and estimation of transition in plane channel flow. *Journal of Fluid Mechanics* **481**, 149–175.

Holmes, P., Lumley, J. L. & Berkooz, G. 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge, UK: Cambridge University Press.

Ilak, M. & Rowley, C. W. 2006 Reduced-order modeling of channel flow using traveling POD and balanced POD. AIAA Paper 2006-3194, 3^{rd} AIAA Flow Control Conference, Washington, California.

Ilak, M. & Rowley, C. W. 2008*a* Feedback control of transitional channel flow using balanced proper orthogonal decomposition. AIAA Paper 2008-4230, 5^{th} AIAA Theoretical Fluid Mechanics Conference, Seattle, Washington.

Ilak, M. & Rowley, C. W. 2008*b* Modeling of transitional channel flow using balanced proper orthogonal decomposition. *Physics of Fluids* **20** (034103).

Jiménez, J. & Moin, P. 1991 The minimal flow unit in near-wall turbulence. *Journal of Fluid Mechanics* **225**, 213–240.

Jonckheere, E. A. & Silverman, L. M. 1983 A new set of invariants for linear systems—application to reduced order compensator design. *IEEE Transactions on Automatic Control* **28** (10), 953–964.

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. *Journal of Fluid Mechanics* **332**, 157–184.

Jovanović, M. R. & Bamieh, B. 2005 Componentwise energy amplification in channel flows. *Journal of Fluid Mechanics* **534**, 145–183.

Juang, J.-N. & Pappa, R. S. 1985 An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics* **8** (5), 620–627.

Juang, J.-N. & Phan, M. Q. 2001 *Identification and control of mechanical systems*. Cambridge University Press.

Kao, T. W. & Park, C. 1970 Experimental investigations of the stability of channel flows. Part 1. Flow of a single liquid in a rectangular channel. *Journal of Fluid Mechanics* **43**, 145–164.

Kim, J. & Bewley, T. R. 2007 A linear systems approach to flow control. *Annual Review of Fluid Mechanics* **39**, 383–417.

Kim, J. & Moin, P. 1986 The structure of the vorticity field in turbulent channel flow. Part 2. Study of ensemble-averaged fields. *Journal of Fluid Mechanics* **162**, 339–363.

Kim, J., Moin, P. & Moser, R. 1987 Turbulence statistics in fully-developed channel flow at low Reynolds number. *Journal of Fluid Mechanics* **177**, 133–166.

Kim, L. & Moehlis, J. 2008 Characterizing the edge of chaos for a shear flow model. *Physical Review E* **78** (036315).

Krener, A. J. 2006 Model reduction for linear and nonlinear control systems. In *45th Conference on Decision and Control, San Diego, CA*. Bode Lecture.

Krstić, M., Kanellakopoulos, I. & Kokotović, P. 1995 *Nonlinear and Adaptive Control Design*. John Wiley and Sons, Inc.

Lall, S., Marsden, J. E. & Glavaški, S. 2002 A subspace approach to balanced truncation for model reduction of nonlinear control systems. *International Journal of Robust and Nonlinear Control* **12**, 519–535.

Landahl, M. 1980 A note on an algebraic instability of inviscid parallel shear flows. *Journal of Fluid Mechanics* **98**, 243–251.

Laub, A. J., Heath, M. T., Page, C. C. & Ward, R. C. 1987 Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Transactions on Automatic Control* **32**, 115–122.

Lauga, E. & Bewley, T. R. 2004 Performance of a linear robust control strategy on a nonlinear model of spatially developing flows. *Journal of Fluid Mechanics* **512**, 343–374.

Lee, K. H., Cortelezzi, L., Kim, J. & Speyer, J. 2001 Application of reduced-order controller to turbulent flows for drag reduction. *Physics of Fluids* **13** (5), 1321–1330.

Luchtenburg, D. K., Günther, B., Noack, B. R., King, R. & Tadmor, G. 2009 A generalized mean-field model of the natural and high-frequency actuated flow around a high-lift configuration. *Journal of Fluid Mechanics* **623**, 283–316.

Lumley, J. L. 1967 The structure of inhomogeneous turbulence. In *Atmospheric Turbulence and Wave Propagation* (ed. A. Yaglom & V. Tatarski), pp. 166–78. Moscow: Nauka.

Lumley, J. L. 1970 *Stochastic Tools in Turbulence*. New York: Academic Press.

Lundell, F. 2003 Pulse-width modulated blowing/suction as a flow control actuator. *Experiments in Fluids* **35**, 502–504.

Lundell, F. 2007 Reactive control of transition induced by free-stream turbulence. *Journal of Fluid Mechanics* **585**, 41–71.

Lundell, F., Monokrousos, A. & Brandt, L. 2009 Feedback control of boundary layer bypass transition: Experimental and numerical progress. 47th Aerospace Sciences Meeting, AIAA Paper 2009-612.

Ma, Z., Ahuja, S. & Rowley, C. W. 2009*a* Reduced order models for control of fluids using the eigensystem realization algorithm. *Theoretical and Computational Fluid Dynamics* (Submitted).

Ma, Z., Rowley, C. W. & Tadmor, G. 2009*b* Snapshot-based balanced truncation for linear time-periodic systems. *IEEE Transactions on Automatic Control* (Submitted).

Mezić, I. 2004 Coupled nonlinear dynamical systems: Asymptotic behavior and uncertainty propagation. In *Proceedings of the 43rd IEEE Conference on Decision and Control*.

Moin, P. & Kim, J. 1980 On the numerical solution of time-dependent viscous incompressible fluid-flows involving solid boundaries. *Journal of Computational Physics* **35**, 381–392.

Moin, P. & Kim, J. 1982 Numerical investigation of turbulent channel flow. *Journal of Fluid Mechanics* **118**, 341–377.

Moin, P. & Kim, J. 1985 The structure of the vorticity field in turbulent channel flow. Part 1. Analysis of instantaneous fields and statistical correlations. *Journal of Fluid Mechanics* **155**, 441–464.

Moin, P. & Moser, R. 1989 Characteristic-eddy decomposition of turbulence in a channel. *Journal of Fluid Mechanics* **200**, 471–509.

Moore, B. C. 1981 Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* **26** (1), 17–32.

Moser, R. D., Kim, J. & Mansour, N. N. 1999 Direct numerical simulation of turbulent channel flow up to Re_τ=590. *Physics of Fluids* **11** (4), 943–945.

Nijmeier, H. & van der Schaft, A. 1990 *Nonlinear Dynamical Control Systems*. Springer-Verlag.

Nishioka, M., Iida, S. & Ichikawa, Y. 1975 An experimental investigation of the stability of plane poiseuille flow. *Journal of Fluid Mechanics* **72**, 731–751.

Noack, B., Afanasiev, K., Morzyński, M., Tadmor, G. & Thiele, F. 2003 A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics* **497**, 335–363.

Noack, B., Tadmor, G. & Morzyński, M. 2004 Low-dimensional models for feedback flow control. Part I: Empirical Galerkin models. In *2nd AIAA Flow Control Conference*. Portland, Oregon, U.S.A., June 28 – July 1, AIAA Paper 2004-2408 (invited contribution).

Ogata, K. 1997 *Modern Control Engineering*, 3^{rd} edn. Prentice-Hall.

Or, A. C., Speyer, J. L. & Carlson, H. A. 2008 Model reduction of input-output dynamical systems by proper orthogonal decomposition. *Journal of Guidance, Control, and Dynamics* **31** (2), 322–328.

Orszag, S. A. 1971 Accurate solution of the Orr-Sommerfeld stability equation. *Journal of Fluid Mechanics* **50**, 689–703.

Orszag, S. A. & Kells, L. C. 1980 Transition to turbulence in plane Poiseuille and plane Couette flow. *Journal of Fluid Mechanics* **96**, 159–205.

Orszag, S. A. & Patera, A. T. 1980 Subcritical transition to turbulence in plane channel flows. *Physical Review Letters* **45** (12), 989–993.

Patel, V. & Head, M. 1969 Some observations on skin friction and velocity profiles in fully developed pipe and channel flows. *Journal of Fluid Mechanics* **38**, 181–201.

Podvin, B. & Lumley, J. L. 1998 A low-dimensional approach for the minimal flow unit. *Journal of Fluid Mechanics* **362**, 121–155.

Pope, S. B. 2000 *Turbulent Flows*. Cambridge University Press.

Qiu, L. & Davison, E. 1993 Performance limitations of non-minimum phase systems in the servomechanism problem. *Automatica* **29** (2), 337–349.

Reddy, S. C. & Henningson, D. S. 1993 Energy growth in viscous channel flows. *Journal of Fluid Mechanics* **252**, 209–238.

Reddy, S. C., Schmid, P. J., Baggett, J. S. & Henningson, D. S. 1998 On stability of streamwise streaks and transition thresholds in plane channel flows. *Journal of Fluid Mechanics* **365**, 269–303.

Rowley, C. W. 2001 Modeling, simulation, and control of cavity flow oscillations. PhD thesis, California Institute of Technology.

Rowley, C. W. 2005 Model reduction for fluids using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos* **15** (3), 997–1013.

Rowley, C. W. & Ilak, M. 2006 Reduced-order models of linearized channel flow using balanced truncation. In *Proceedings of the 14th Mediterranean Conference on Control and Automation*. Ancona, Italy.

Rowley, C. W. & Juttijudata, V. 2005 Model-based control and estimation of cavity flow oscillations. In *Proceedings of the 44th IEEE Conference on Decision and Control*. Seville, Spain.

Sandham, N. D. & Kleiser, L. 1992 Late stages of transition to turbulence in channel flow. *Journal of Fluid Mechanics* **245**, 319–348.

Scherpen, J. M. A. 1993 Balancing for nonlinear systems. *Systems and Control Letters* **21** (2), 143–153.

Schmid, P. J. & Henningson, D. S. 1992 A new mechanism for rapid transition involving a pair of oblique waves. *Physics of Fluids A* **4** (9), 1986–1989.

Schmid, P. J. & Henningson, D. S. 2001 *Stability and Transition in Shear Flows*, *Applied Mathematical Sciences*, vol. 142. Springer-Verlag.

Schneider, T. M., Gibson, J. F., Lagha, M., De Lillo, F. & Eckhardt, B. 2008 Laminar-turbulent boundary in plane Couette flow. *Physical Review E* **78** (037301).

Siegel, S. G., Seidel, J., Fagley, C., Luchtenburg, D. M., Cohen, K. & McLaughlin, T. 2008 Low-dimensional modelling of a transient cylinder wake using double proper orthogonal decomposition. *Journal of Fluid Mechanics* **610**, 1–42.

Sirovich, L. 1987 Turbulence and the dynamics of coherent structures, parts I–III. *Quarterly of Applied Mathematics* **XLV** (3), 561–590.

Skogestad, S. & Postlethwaite, I. 2005 *Multivariable Feedback Control: Analysis and Design*, 2nd edn. John Wiley and Sons.

Skufca, J. D., Yorke, J. A. & Eckhardt, B. 2006 Edge of chaos in a parallel shear flow. *Physical Review Letters* **96** (174101).

Smith, T. R. 2003 Low-dimensional models of plane Couette flow using the proper orthogonal decomposition. PhD thesis, Princeton University.

Smith, T. R., Moehlis, J. & Holmes, P. 2005 Low-dimensional models for turbulent plane couette flow in a minimal flow unit. *Journal of Fluid Mechanics* **538**, 71–110.

Stengel, R. F. 1994 *Optimal Control and Estimation*. Dover.

Trefethen, L. N. & Bau, D. I. 1997 *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics.

Trefethen, L. N., Trefethen, A. E., Reddy, S. C. & Driscoll, T. A. 1993 Hydrodynamic stability without eigenvalues. *Science* **261**, 578–584.

Verriest, E. I. & Gray, W. 2004 Nonlinear balanced realizations. In *Proceedings of the 43rd IEEE Conference on Decision and Control*.

Wei, M. & Rowley, C. W. 2009 Low-dimensional models of a temporally evolving free shear layer. *Journal of Fluid Mechanics* **618**, 113–134.

Weideman, J. A. C. & Reddy, S. C. 2000 A matlab differentiation matrix suite. *ACM Transactions on Mathematical Software* **26** (4), 465–519.

White, F. M. 1991 *Viscous Fluid Flow*, 2nd edn. McGraw-Hill.

Willcox, K. & Peraire, J. 2002 Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal* **40** (11), 2323–2330.

Zhou, K., Doyle, J. C. & Glover, K. 1996 *Robust and Optimal Control*. New Jersey: Prentice-Hall.

Zhou, K., Salomon, G. & Wu, E. 1999 Balanced realization and model reduction for unstable systems. *International Journal of Robust and Nonlinear Control* **9** (3), 183–198.