Flow control and constrained optimization problems

Laurent Cordier

Institut Pprime, CNRS – Université de Poitiers – ENSMA, UPR 3346, Département Fluides, Thermique, Combustion, CEAT, 43 rue de l'Aérodrome, F-86036 Poitiers Cedex, France

Abstract Constrained optimization is presented as a key enabler for answering numerous important questions in the heart of flow control. These problems range from the extraction of Proper Orthogonal Decomposition modes and tools from linear control theory to optimal control which can be applied to any type of non-linear systems. The determination of optimal growth disturbances is presented as a particular case of constrained optimization. The chapter shall provide a complete description for deriving analytically and solving numerically any specific formulation of constrained optimization.

1 Introduction

The objective of this chapter is to present within the unified framework of constrained optimization problems, different numerical tools which change completely our ideas on flow control in the last decade. We will see in particular that reduced-order modeling based on Proper Orthogonal Decomposition modes (see the contribution by B. Noack et al. in this book), as well as classical techniques of linear control (Linear Quadratic Regulator and Linear Quadratic Gaussian methods) and optimal control, have in common the resolution of a constrained optimization problem. Beyond that, we will also show that the concept of optimal disturbances, introduced in stability theory to explain the transition to turbulence of linearly stable flows, can be also formulated as a constrained optimization problem and, if needed, be solved simultaneously to a control problem. Lastly, we will highlight that inverse methods (model identification or parameter estimation) can be interpreted as a particular constrained optimization problem. The objective is to give the possibility to the interested reader of rapidly developing by

him/her-self the analytical and numerical solutions to the constrained optimization problem of his/her interest. The choice was thus made to detail as much as possible the different stages.

The current chapter is organized as follows: In section 2.1, we introduce the issues of flow control and present, for facilitating future discussions, the different actors on the control scene. Then we introduce the linearized framework, often used in flow control, and finish by formulating a series of questions related directly to different aspects of flow control. In section 2.2, we give some essential elements of linear control theory and continue in section 2.3 by an introduction of model reduction seen under the specific angle of projection methods. In section 3, we focus on the fundamental aspects of optimal control theory. At this stage, the presentation will remain very similar to what can be found in Gunzburger (1997a) and more recently in Gunzburger (2003). Section 4 considers the case of LQR control for a generic system and shows that the solution of a high-dimensional Riccati differential equation is necessary to determine the feedback control law that minimizes the value of the cost function. Section 5 highlights that the determination of optimal disturbances corresponds to a constrained optimization problem for which the control is the initial condition of the dynamical system. Lastly, sections 6 and 7 consider the case where the constraint corresponds to a time-dependent partial differential equation, linear and nonlinear respectively. Section 7 finishes with some numerical results of optimal control for the Burgers equation.

2 Elements of control theory and model reduction

2.1 Flow control

First, in section 2.1.1, we give the scope of flow control and introduce the terminology necessary to present constrained optimization problems as a main topic in modern fluid mechanics. Then, in section 2.1.2, we introduce the linearized framework used in linear control theory. Finally, in section 2.1.3, we list different types of problems which can appear within the framework of flow control while insisting on their similarity.

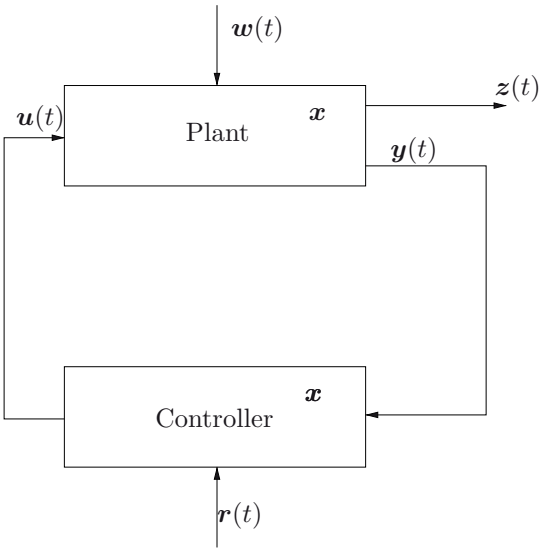
2.1.1 Scope and objectives of flow control

2.1.1.1 General points The goal of a flow control system is to achieve some desired objective by manipulating properly the flow configuration (physical properties, volume forcing or boundary conditions). Based on the type of actuation, either *passive* (no energy expenditure) or *active*, and

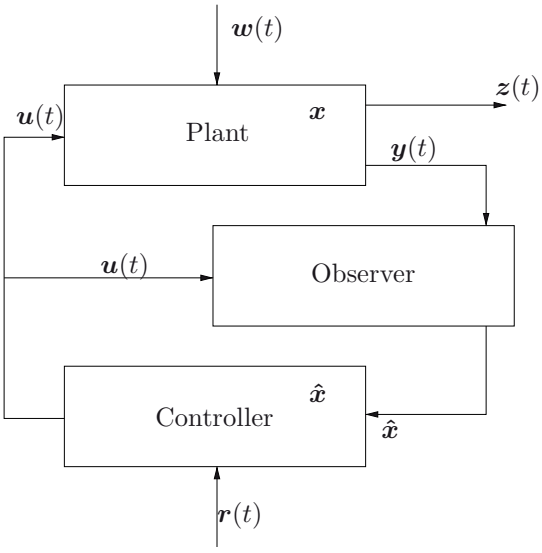
on the means by which the control evolves in response to changes in the flow, *open-loop* or *closed-loop*, different strategies can be considered (see Gad-el-Hak, 2000, for a discussion on this classification). By nature, passive control strategies are similar to shape optimization. Determining the shape that a surface of revolution must have to offer the least resistance to the motion goes back to Newton (end of 17^{th} century) and involved the invention of the calculus of variations. We will see in section 3.1.1 that this question can be formalized as a constrained optimization problem by simply modifying the space on which the solutions are required. In open-loop, the parameters of the actuators are set once for all at the design stage and remain constant throughout the optimization procedure whatever the changes undergone by the flow. With this type of strategy, the sensitivity of the system to external disturbances or to error modeling (change in the parameters of the system) is then important. In addition, stabilizing an unstable solution - what may sometimes be interesting from a point of view of the performances - becomes difficult. For these reasons, we will consider throughout this chapter the case of closed-loop control or feedback control where there exist sensors for measuring at least partially the effects of the control on the system.

2.1.1.2 Terminology In the control literature[1], the mathematical model of the system to be controlled is called *plant*. In general, this model only approximates the behavior of the physical system. We will go back to this point and to the consequences in terms of optimization in section 2.3. The corresponding *state variables* of the plant is noted $\boldsymbol{x}$. The objective of a control system is to make the *reference output* $\boldsymbol{z}$ behave in a desired way by manipulating the *plant input* $\boldsymbol{u}$ (see Fig. 1). The *reference input* $\boldsymbol{r}$ specifies the desired behavior of the reference output. In feedback flow control the *measured plant output* $\boldsymbol{y}$ is fed back into the controller for determining the control. Compared to $\boldsymbol{u}$, the *disturbance input* $\boldsymbol{w}$ consists of those inputs to the plant that are generated by the environment. It includes one contribution coming from the state disturbances $\boldsymbol{w_1}$ and another contribution coming from the measurement noise $\boldsymbol{w_2}$. In the idealized case called *full-state configuration* (see Fig. 1(a)), the entire state $\boldsymbol{x}$ is assumed to be available for the controller. In the general case called *observer-based*

[1] Here, and in the rest of the chapter, we decide to use the standard notations in textbooks of control theory to familiarize the reader coming from fluid mechanics to these notations. Then, otherwise stated, $\boldsymbol{u}$ denotes the control and not a velocity field. Moreover, quantities expressed in boldface correspond to vector quantities.



(a) Full state configuration.



(b) Observer-based configuration.

Figure 1. Typical block diagrams for feedback control.

configuration (see Fig. 1(b)), the plant states that are not measured directly is estimated by an observer. Thereafter, all the quantities with a hat correspond to estimated variables: for instance, $\hat{\boldsymbol{x}}$ are estimated states.

2.1.1.3 Plant modelling The next stage is the determination of the system of equations for the plant (Fig. 2). Starting from a physical system and some measured data, the modelling phase consists of deriving a set of Partial Differential Equations (PDEs) or Ordinary Differential Equations (ODEs). In the first case, after discretization in space of the PDEs with any numerical method (finite element, finite volume, ...), a set of ODEs is obtained. Sometimes the ODEs are discretized in time as well, yielding discrete-time dynamical systems. Here, to simplify the presentation, we will concentrate on continuous-time systems. Finally, since any dynamical system can be reduced to a first-order system of differential equations by changing the set of variables, we obtain a non-linear *state space model* given by

$$\mathcal{S} : \begin{cases} \dot{\boldsymbol{x}}(t) = \boldsymbol{f}(t, \boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{w}(t)), \\ \boldsymbol{z}(t) = \boldsymbol{h}(t, \boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{w}(t)), \\ \boldsymbol{y}(t) = \boldsymbol{g}(t, \boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{w}(t)), \end{cases}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^{n_x}$, $\boldsymbol{u} \in \mathbb{R}^{n_u}$, $\boldsymbol{w} \in \mathbb{R}^{n_w}$, $\boldsymbol{y}(t) \in \mathbb{R}^{n_y}$ and $\boldsymbol{z} \in \mathbb{R}^{n_z}$. The non-linear functions $\boldsymbol{f}$, $\boldsymbol{g}$ and $\boldsymbol{h}$ are defined accordingly.

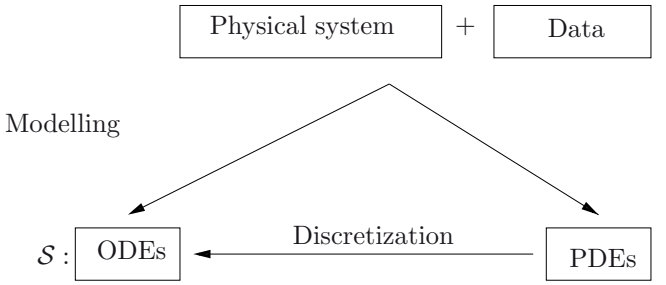


Figure 2. Broad framework of the determination of the plant equations (after Antoulas, 2005).

2.1.2 Linearized framework

Often, in practice, the non-linear system $\boldsymbol{f}$ is linearized around an operating condition of interest. To simplify the future notations, we will assume that the system does not depend explicitly on time and suppress for the moment

the dependance on the external disturbance $\boldsymbol{w}$ writing for the plant equations $\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t))$. Depending on the applications, this operating condition can be a particular solution of the unsteady dynamical system $\boldsymbol{f}$, that is to say $\boldsymbol{f}(\boldsymbol{x}_e(t), \boldsymbol{u}_e(t)) = \boldsymbol{0}$ or an equilibrium point of $\boldsymbol{f}$ characterized by $\dot{\boldsymbol{x}}_e = \boldsymbol{f}(\boldsymbol{x}_e, \boldsymbol{u}_e) = \boldsymbol{0}$. In the domain of flow instabilities, this equilibrium point corresponds to a steady solution of the Navier-Stokes equations.

We then introduce the first-order perturbations $\tilde{\boldsymbol{x}}(t)$ and $\tilde{\boldsymbol{u}}(t)$ such that

$$\boldsymbol{x}(t) = \boldsymbol{x}_e(t) + \tilde{\boldsymbol{x}}(t) \qquad \text{and} \qquad \boldsymbol{u}(t) = \boldsymbol{u}_e(t) + \tilde{\boldsymbol{u}}(t).$$

Expanding $\boldsymbol{f}$ in a Taylor series about $(\boldsymbol{x}_e, \boldsymbol{u}_e)$, we obtain

$$\begin{aligned}\dot{\boldsymbol{x}}_e(t) + \dot{\tilde{\boldsymbol{x}}}(t) =& \boldsymbol{f}(\boldsymbol{x}_e(t), \boldsymbol{u}_e(t)) + J_x(\boldsymbol{x}_e(t), \boldsymbol{u}_e(t))\tilde{\boldsymbol{x}}(t) + J_u(\boldsymbol{x}_e(t), \boldsymbol{u}_e(t))\tilde{\boldsymbol{u}}(t) \\ &+\text{higher order terms}\end{aligned}$$

where J_x (respectively J_u) is the Jacobian matrix of $\boldsymbol{f}$ with respect to $\boldsymbol{x}$ (respectively $\boldsymbol{u}$):

$$(J_x)_{ij} = \frac{\partial f_i}{\partial x_j} \quad \text{with} \quad 1 \leq i \leq n_x \quad ; \quad 1 \leq j \leq n_x$$

and

$$(J_u)_{ij} = \frac{\partial f_i}{\partial u_j} \quad \text{with} \quad 1 \leq i \leq n_x \quad ; \quad 1 \leq j \leq n_u.$$

Neglecting the higher order terms and letting

$$A(t) = J_x(\boldsymbol{x}_e(t), \boldsymbol{u}_e(t)) \quad \text{and} \quad B(t) = J_u(\boldsymbol{x}_e(t), \boldsymbol{u}_e(t))$$

we obtain the linearized state space model

$$\dot{\tilde{\boldsymbol{x}}}(t) = A(t)\tilde{\boldsymbol{x}}(t) + B(t)\tilde{\boldsymbol{u}}(t)$$

where $A(t) \in \mathbb{R}^{n_x \times n_x}$ is the state matrix and $B(t) \in \mathbb{R}^{n_x \times n_u}$ is the input matrix.

Similarly, the nonlinear functions $\boldsymbol{z} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u})$ and $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u})$ may be linearized around the equilibrium point, resulting in a linear, parameter time-varying (LPTV) system given by

$$\mathcal{S}_{LPTV} : \left\{ \begin{array}{l} \dot{\boldsymbol{x}}(t) = A(t)\boldsymbol{x}(t) + B(t)\boldsymbol{u}(t), \\ \boldsymbol{z}(t) = C_1(t)\boldsymbol{x}(t) + D_1(t)\boldsymbol{u}(t), \\ \boldsymbol{y}(t) = C_2(t)\boldsymbol{x}(t) + D_2(t)\boldsymbol{u}(t), \end{array} \right.$$

where for convenience the notation of the fluctuations was removed.

The state model $\mathcal{S}_{LPTV}$ can be further simplified when the system is time-invariant. Adding the linearized contribution from the external disturbances, the system becomes

$$\begin{aligned}\dot{\boldsymbol{x}}(t) &= A\boldsymbol{x}(t) + B_1\boldsymbol{w}(t) + B_2(t)\boldsymbol{u}(t),\\ \boldsymbol{z}(t) &= C_1\boldsymbol{x}(t) + D_{11}\boldsymbol{w}(t) + D_{12}\boldsymbol{u}(t),\\ \boldsymbol{y}(t) &= C_2\boldsymbol{x}(t) + D_{21}\boldsymbol{w}(t) + D_{22}\boldsymbol{u}(t).\end{aligned}$$

This is the more general class of model that we can consider for linear-time invariant (LTI) systems. Throughout this chapter, we will restrict our attention to the simplified[2] linear system

$$\mathcal{S}_{LTI} : \begin{cases} \dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t),\\ \boldsymbol{z}(t) = C_1\boldsymbol{x}(t) + D_1\boldsymbol{u}(t),\\ \boldsymbol{y}(t) = C_2\boldsymbol{x}(t) + D_2\boldsymbol{u}(t),\end{cases} \tag{1}$$

where $C_1 \in \mathbb{R}^{n_z \times n_x}$ and $C_2 \in \mathbb{R}^{n_y \times n_x}$ are the output matrices and where $D_1 \in \mathbb{R}^{n_z \times n_u}$ and $D_2 \in \mathbb{R}^{n_y \times n_u}$ are the input to output coupling matrices. A dynamical system with single input ($n_u = 1$) and single output ($n_y = 1$) is called a SISO (single input and single output) system, otherwise it is called MIMO (multiple input and multiple output) system. When this is not necessary, we will not mention the variable $\boldsymbol{z}$ thereafter.

The advantage of linear systems is that the state, solution of (1), can be found explicitly from the input and the initial conditions (see Zhou et al., 1996):

$$\boldsymbol{x}(t) = e^{At}\boldsymbol{x}(0) + \int_0^t e^{A(t-\tau)}B\boldsymbol{u}(\tau)\ \mathrm{d}\tau$$

where the matrix exponential is defined by the power series:

$$e^{At} = I + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \cdots$$

The reference and measured plant outputs are then generated as a function of the initial conditions and the input:

$$\boldsymbol{z}(t) = C_1e^{At}\boldsymbol{x}(0) + \int_0^t C_1e^{A(t-\tau)}B\boldsymbol{u}(\tau)\ \mathrm{d}\tau + D_1\boldsymbol{u}(t)$$

and

$$\boldsymbol{y}(t) = C_2e^{At}\boldsymbol{x}(0) + \int_0^t C_1e^{A(t-\tau)}B\boldsymbol{u}(\tau)\ \mathrm{d}\tau + D_2\boldsymbol{u}(t).$$

We will see in section 2.2.2 the consequences in terms of observability and controllability of the system $\mathcal{S}$.

[2] $B_1 = D_{11} = D_{21} = 0$, $B \triangleq B_2$, $D_1 \triangleq D_{12}$ and $D_2 \triangleq D_{22}$.

2.1.3 Different types of problems

Within the general framework of flow control, various types of problems can be considered:

Problem 1: How to determine the control law $\boldsymbol{u}$ to apply to the dynamical system $\mathcal{S}$ to minimize a given norm[3] of $\boldsymbol{z}$?

In lack of particular assumption on the model, this problem is designated as optimal control. The model $\boldsymbol{f}$ can then be a Direct Numerical Simulation (Bewley et al., 2001), a Large Eddy Simulation (El Shrif, 2008) or a reduced-order model (see section 2.3) obtained by Proper Orthogonal Decomposition (Bergmann et al., 2005; Bergmann and Cordier, 2008).

Problem 2: Now let us assume that the control system design corresponds to state feedback *i.e.* $\boldsymbol{u} = K\boldsymbol{x}$ for the full state configuration or $\boldsymbol{u} = K\hat{\boldsymbol{x}}$ for the observer-based configuration. Then how to determine the control law $\boldsymbol{u}$, or equivalently the gain matrix K, to apply to $\mathcal{S}$ to minimize a given norm of $\boldsymbol{z}$?

If the system $\mathcal{S}$ is Linear Time Invariant (LTI) then the problem is called Linear Quadratic Regulator or LQR, see section 4 or in Burl (1999).

Problem 3: Let $\hat{\boldsymbol{y}}$ be the estimated value of the output based on the estimated state $\hat{\boldsymbol{x}}$. For an LTI system $\mathcal{S}$, the state space system for the observer is

$$\begin{aligned} \dot{\hat{\boldsymbol{x}}}(t) &= A\hat{\boldsymbol{x}}(t) + B_2\boldsymbol{u}(t) + L\left(\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t)\right), \\ \hat{\boldsymbol{y}}(t) &= C_2\hat{\boldsymbol{x}}(t) \end{aligned} \tag{2}$$

where L is the observer gain matrix.

Then how to determine the gain matrix L so that $\hat{\boldsymbol{x}}$ is roughly equal to $\boldsymbol{x}$? This question corresponds to the observer design. It can be shown (see section 2.2.2) that this problem is dual to the control problem described at the previous item.

Problem 4: How to determine one or more parameters of the system $\mathcal{S}$ knowing the input $\boldsymbol{x}$ and the corresponding output $\boldsymbol{y}$?

Depending on the authors, this question corresponds to the estimation of physical parameters or data inversion (Tarantola, 2005), to systems' identification (Juang and Phan, 2001) or to model calibration (see Cordier et al., 2010, for an application to reduced-order models derived by POD).

[3] An exact definition will be given in section 3.1.2.

Problem 5: The model $\mathcal{S}$ being known, how to determine the input $\boldsymbol{u}$ to apply to $\mathcal{S}$ to obtain given output $\boldsymbol{y}$?

This question, which is very similar to that of the first item, corresponds to a problem of data inversion.

Problem 6: How to determine the initial condition $\boldsymbol{x_0}$ which maximizes the energetic amplification of the dynamical system $\mathcal{S}$?

With this question, we can introduce the concept of optimal disturbances and optimal growth (Schmid and Henningson, 2001). We will see an application in section 5 for the linearized channel flow.

All these problems are sufficiently generals to appear in many scientific disciplines sometimes very distant from each other (engineering, medical or social sciences, ...). In addition, these problems clearly all involve at a different level the resolution of a constrained optimization problem (minimization for the great majority, maximization for the problem of optimal disturbances). The solution of constrained optimization problems will thus be the object of a detailed description in section 3.

2.2 Input-output framework

In section 2.1.2 we learned how, starting from a nonlinear model of dynamics $\mathcal{S}$ resulting from any physical modeling, to determine a linear-time invariant system. Is this step sufficient for control? On one hand, the answer is affirmative because there exist many methods of control dedicated to the linearized systems. On the other hand, we will now see that in general it is necessary to be much more careful since the mapping of measurements $\boldsymbol{y}$ (output) to the control $\boldsymbol{u}$ (input) is crucial to have a chance of success for the control.

2.2.1 Similarity transformations

The objective of this section is to demonstrate that the equations of the state-space system are not unique. Starting from the state-space system (1), reproduced here for convenience:

$$\begin{aligned}\dot{\boldsymbol{x}}(t) &= A\boldsymbol{x}(t) + B_2\boldsymbol{u}(t),\\ \boldsymbol{y}(t) &= C_2\boldsymbol{x}(t) + D_2\boldsymbol{u}(t)\end{aligned}$$

we consider a new state vector

$$\tilde{\boldsymbol{x}}(t) = \mathbb{T}^{-1}\boldsymbol{x}(t)$$

where $\mathbb{T}$ is a constant, invertible transformation matrix. Since $\mathbb{T}$ is invertible, we have $\boldsymbol{x}(t) = \mathbb{T}\tilde{\boldsymbol{x}}(t)$ and $\dot{\boldsymbol{x}}(t) = \mathbb{T}\dot{\tilde{\boldsymbol{x}}}(t)$ ($\mathbb{T}$ independent of time). We

then obtain immediately a new state-space system defined in terms of the state $\tilde{\boldsymbol{x}}$:

$$\begin{aligned}\dot{\tilde{\boldsymbol{x}}}(t) &= \left(\mathbb{T}^{-1}A\mathbb{T}\right)\tilde{\boldsymbol{x}}(t) + \left(\mathbb{T}^{-1}B_2\right)\boldsymbol{u}(t),\\ \boldsymbol{y}(t) &= \left(C_2\mathbb{T}\right)\tilde{\boldsymbol{x}}(t) + D_2\boldsymbol{u}(t)\end{aligned}$$

In summary, the new state-space model is generated by using the following similarity transformations:

$$A \longrightarrow \mathbb{T}^{-1}A\mathbb{T} \quad ; \quad B_2 \longrightarrow \mathbb{T}^{-1}B_2 \quad ; \quad C_2 \longrightarrow C_2\mathbb{T} \quad ; \quad D_2 \longrightarrow D_2.$$

Since there exists an infinite number of state representations for a given system, a natural question is then how we can determine the transformation $\mathbb{T}$ best adapted to control?

2.2.2 Controllability and observability

This section addresses the following fundamental questions:

1. Can we always control a flow?
2. Can the state of a system be estimated from the measurements?

In practice, the answers to these questions provide a guide to the selection of actuators and sensors, and are also useful for developing controllers and observers.

Controllability describes the ability of the control $\boldsymbol{u}$ to influence the state $\boldsymbol{x}$. Conversely, observability describes the ability to reconstruct the state $\boldsymbol{x}$ based on available measurements $\boldsymbol{y}$. To simplify the description, consider $\mathcal{S}_{LTI}$ given by (1) with $D_2 = 0$. In this case, the output $\boldsymbol{y}$ is given (see section 2.1.2) by:

$$\boldsymbol{y}(t) = \underbrace{\int_0^t C_2 e^{A(t-\tau)} B_2 \boldsymbol{u}(\tau)\, \mathrm{d}\tau}_{T_1} + \underbrace{C_2 e^{At}\boldsymbol{x}(0)}_{T_2}.$$

The term T_1 defines a mapping from the space of the control $\boldsymbol{u}$ to the space of the state $\boldsymbol{x}$. Since this map is linear, the image is a subspace of the state-space $\mathbb{R}^{n_x}$ called the controllability subspace. This subspace depends only on the matrices A and B_2, and is denoted $S_{\mathcal{C}}$. Similarly, the term T_2 defines a mapping from the space of the state $\boldsymbol{x}$ to the space of measurement $\boldsymbol{y}$. Since this map is also linear, the image is a subspace of the state-space $\mathbb{R}^{n_y}$ called the observability subspace. This subspace depends only on the matrices C_2 and A, and is denoted by $S_{\mathcal{O}}$. The kernel of this linear map forms a subspace, called the unobservable subspace. Since for these states,

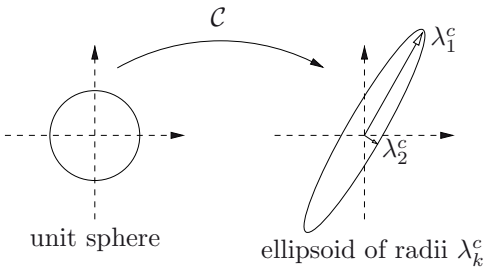


Figure 3. Geometric interpretation of the controllability operator: mapping of unit sphere onto ellipsoid. The direction corresponding to λ_1^c is more controllable than the direction corresponding to λ_2^c.

$\boldsymbol{y} = \mathbf{0}$, it means that the elements of the kernel[4] may be added to any another initial state without changing the output.

2.2.2.1 Controllability Suppose the system defined in (1) is stable. Then, for $\boldsymbol{x}(-\infty) = \mathbf{0}$, the state at time zero $\boldsymbol{x}(0) = \boldsymbol{x_0}$ is given by

$$\boldsymbol{x_0} = \int_{-\infty}^{0} e^{-A\tau} B_2 \boldsymbol{u}(\tau)\, \mathrm{d}\tau.$$

This defines the controllability operator $\mathcal{C}$ by $\boldsymbol{x_0} = \mathcal{C}\boldsymbol{u}$. In geometric terms analogous to the moment of inertia tensor, $\mathcal{C}$ defines a controllability ellipsoid in the state space, with the longest principal axes along the most controllable directions (see Fig. 3).

The controllability gramian is an $n_x \times n_x$ matrix whose eigenvectors span the controllability subspace. It is defined[5] for the system (1) as

$$W_c(t) = \mathcal{C}\mathcal{C}^H = \int_0^t e^{A\tau}\, B_2\, B_2^H\, e^{A^H\tau}\, \mathrm{d}\tau \tag{3}$$

where the exponent H denotes the transconjugate operator (transpose conjugate).

[4] The kernel or null space of a linear transformation is the set of vectors that map to zero. If we associate a matrix $\mathcal{A}$ to the linear transformation, the null space of $\mathcal{A}$ is the set of all vectors x for which $\mathcal{A}x = 0$.

[5] The controllability gramian and later the observability gramian (section 2.2.2.2) can be defined in a more general way by considering a weighted inner product (see appendix A or Ilak 2009).

If the system (1) is stable, we can consider the infinite horizon Gramian ($t \longrightarrow +\infty$) and forget the dependance on time. Since W_c is clearly self-adjoint, it admits a set of real, non-negative eigenvalues λ_k^c and orthonormal eigenvectors $\boldsymbol{x_k^c}$. The eigenvalues are a measure of the amount of control energy required to obtain the corresponding eigenvectors. For two states, $\boldsymbol{x_1^c}$ and $\boldsymbol{x_2^c}$ with $\|\boldsymbol{x_1^c}\|_2 = \|\boldsymbol{x_2^c}\|_2$ where $\|\cdot\|_2$ denote the classical L_2 norm ($\|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^H\boldsymbol{x}$) then if

$$\lambda_1^c = \left(\boldsymbol{x_1^c}\right)^H W_c \boldsymbol{x_1^c} = \|\boldsymbol{x_1^c}\|_{W_c}^2 > \|\boldsymbol{x_2^c}\|_{W_c}^2 = \left(\boldsymbol{x_2^c}\right)^H W_c \boldsymbol{x_2^c} = \lambda_2^c$$

it means that $\boldsymbol{x_1^c}$ is more controllable than $\boldsymbol{x_2^c}$.

When the size of the system $\mathcal{S}$ is not too high, the controllability gramian can be determined[6] directly as the solution of a Lyapunov[7] equation given by:

$$AW_c + W_c A^H + B_2 B_2^H = 0.$$

By definition, the dynamical system (1), or equivalently the pair (A, B_2) is said to be state controllable if and only if, for any initial state $\boldsymbol{x}(0) = \boldsymbol{x_0}$ and any final state $\boldsymbol{x_f}$, there exists an input $\boldsymbol{u}(t)$ such that $\boldsymbol{x}(t_f) = \boldsymbol{x_f}$ for $t_f - t_0 < +\infty$. Unfortunately, this criterion is not very usable. In practice, the controllability of a system will be verified using one or the other of the following equivalent criteria[8] (Lewis and Syrmos, 1995; Zhou et al., 1996; Skogestad and Postlethwaite, 2005):

1. Kalman criterion

$$rank\left(\begin{bmatrix} B_2 & AB_2 & A^2B_2 & \cdots A^{n_x-1}B_2 \end{bmatrix}\right) = n_x.$$

2. $W_c > 0$.
3. W_c is full-rank.
4. $\mathrm{Im}(\mathcal{C}) = \mathbb{R}^{n_x}$.

Finally, let

$$E_u \triangleq \int_{-\infty}^{0} \|\boldsymbol{u}\|_2^2 \, \mathrm{d}t = \int_{-\infty}^{0} \boldsymbol{u}^H(t)\boldsymbol{u}(t) \, \mathrm{d}t,$$

[6] The proof is based on the time differentiation of (3). It can be found in section A7 of Burl (1999).

[7] A common way to solve continuous-time Lyapunov equation is with the function `lyap` of Matlab or with the Slicot library that can be found in `http://www.slicot.net`.

[8] We remind that the rank of a matrix $\mathcal{A}$ corresponds to the maximal number of linearly independent rows or columns of $\mathcal{A}$. Moreover, a symmetric matrix $\mathcal{A}$ is said positive definite (simply denoted $\mathcal{A} > 0$) if $\boldsymbol{x}^H\mathcal{A}\boldsymbol{x} > 0$ for all non-zero vectors $\boldsymbol{x}$. Finally, $Im(f)$ denotes the image of the operator f. If f is a mapping from E to F, then $Im(f) = \{\boldsymbol{y} \in F$ such that $f(\boldsymbol{x}) = \boldsymbol{y}$, for some $\boldsymbol{x} \in E\}$.

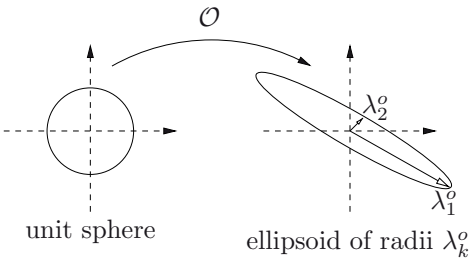


Figure 4. Geometric interpretation of the observability operator: mapping of unit sphere onto ellipsoid. The direction corresponding to λ_1^o is more observable than the direction corresponding to λ_2^o.

with $\boldsymbol{u}(t)$ defined for $t \in]-\infty; 0]$, be the past input energy, it can be shown (Mehrmann and Stykel, 2005) that:

$$E_{u_{min}} = \min_{\boldsymbol{u}} E_u = \boldsymbol{x_0}^H W_c^{-1} \boldsymbol{x_0}.$$

2.2.2.2 Observability We now consider the similar notions as in the previous section but for the output. We will thus follow a similar structure of presentation.

Suppose the system (1) is in some initial state $\boldsymbol{x}(0) = \boldsymbol{x_0}$ and $\boldsymbol{u}(t) = \boldsymbol{0}$ for $t \in [0; +\infty[$. Integrating the dynamics (1), it yields:

$$y(t) = C_2 e^{At} \boldsymbol{x}(0) \tag{4}$$

which defines the observability operator $\mathcal{O}$ by $y(t) = \mathcal{O}\boldsymbol{x_0}$. Similarly to what we have made in section 2.2.2.1 for the controllability, we can analyze this operator in geometric terms (see Fig. 4). Here, $\mathcal{O}$ defines an observability ellipsoid in the state space, with the longest principal axes along the most observable directions.

The observability gramian is an $n_x \times n_x$ matrix whose eigenvectors span the observability subspace. It is defined for the system (1) as

$$W_o(t) = \mathcal{O}^H \mathcal{O} = \int_0^t e^{A^H \tau} \, C_2^H \, C_2 \, e^{A\tau} \, \mathrm{d}\tau. \tag{5}$$

For a stable system, observability can be characterized only by the infinite horizon Gramian ($t \longrightarrow +\infty$) and we can forget the explicit dependance on time in W_o. The eigenvalues λ_k^o of W_o are a measure of the amount of

state energy required to obtain the corresponding eigenvectors $\boldsymbol{x_k^o}$. Obviously, we have the result that for two states, $\boldsymbol{x_1^o}$ and $\boldsymbol{x_2^o}$ with $\|\boldsymbol{x_1^o}\|_2 = \|\boldsymbol{x_2^o}\|_2$ then if

$$\lambda_1^o = \left(\boldsymbol{x_1^o}\right)^H W_o \boldsymbol{x_1^o} = \|\boldsymbol{x_1^o}\|_{W_o}^2 > \|\boldsymbol{x_2^o}\|_{W_o}^2 = \left(\boldsymbol{x_2^o}\right)^H W_o \boldsymbol{x_2^o} = \lambda_2^o$$

it means that $\boldsymbol{x_1^o}$ is more observable than $\boldsymbol{x_2^o}$.

When the dimension of $\mathcal{S}$ is not too high, a common way of determining the observability gramian W_o is to solve the following Lyapunov equation:

$$A^H W_o + W_o A + C_2^H C_2 = 0.$$

By definition, the dynamical system (1), or equivalently the pair (A, C_2) is said to be state observable if and only if, for any time $t_f > 0$, the initial state $\boldsymbol{x}(0) = \boldsymbol{x_0}$ can be determined from knowledge of the input $\boldsymbol{u}(t)$ and output $\boldsymbol{y}(t)$ in the interval $[0; t_f]$. In practice, the observability of a system is verified through one of the following equivalent criteria (Lewis and Syrmos, 1995; Zhou et al., 1996; Skogestad and Postlethwaite, 2005):

1. Kalman criterion

$$rank\left(\begin{bmatrix} C_2 \\ C_2 A \\ \vdots \\ C_2 A^{n_x - 1} \end{bmatrix}\right) = n_x.$$

2. $W_o > 0$.
3. W_o is full-rank.
4. $\ker(\mathcal{O}) = \boldsymbol{0}$.

To conclude this section, let

$$E_y = \int_0^{+\infty} \|\boldsymbol{y}\|_2^2 \, \mathrm{d}t = \int_0^{+\infty} \boldsymbol{y}^H(t)\boldsymbol{y}(t) \, \mathrm{d}t,$$

with $\boldsymbol{y}(t)$ defined for $t \in [0; +\infty[$, be the future output energy, it can be shown easily by substituting (4) in E_y that

$$E_y = \boldsymbol{x_0}^H W_o \boldsymbol{x_0}.$$

2.2.2.3 Duality Duality is an important concept in linear control theory because, used advisedly, it can saved a considerable time in the derivation of properties for the systems under investigation. To go further, we will initially admit that for any primal system defined by (1), that is to say

$$\mathcal{S} : \begin{cases} \dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B_2\boldsymbol{u}(t), \\ \boldsymbol{y}(t) = C_2\boldsymbol{x}(t) \end{cases}$$

we can associate another state-space system, known as dual system, and given by

$$\mathcal{S}_{\text{dual}} : \begin{cases} \dot{\boldsymbol{\xi}}(t) = A^H \boldsymbol{\xi}(t) + C_2^H \boldsymbol{\zeta}(t), \\ \boldsymbol{\eta}(t) = B_2^H \boldsymbol{\xi}(t). \end{cases}$$

Here, $\boldsymbol{\xi}$ is the dual state vector, and $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ contain the dual inputs and outputs. Comparing $\mathcal{S}$ and $\mathcal{S}_{\text{dual}}$ it can be seen that we can deduce the dual system from the knowledge of the primal system with the transformations:

$$A \longrightarrow A^H \qquad \text{and} \qquad B_2 \longrightarrow C_2^H. \tag{6}$$

Duality of controllability and observability From the transformations (6) and the definitions (3) and (5), it is evident that the controllability gramian of the primal system is equal to the observability gramian of the dual system, and vice versa. As a consequence, the following results hold:

1. $\mathcal{S}(A, B_2)$ is controllable if and only if $\mathcal{S}_{\text{dual}}\left(A^H, B_2^H\right)$ is observable,
2. $\mathcal{S}_{\text{dual}}\left(A^H, C_2^H\right)$ is controllable if and only if $\mathcal{S}(A, C_2)$ is observable.

Duality of the control problem and the observer design If we now consider the cost function

$$\mathcal{J}_{\boldsymbol{y}} = \int_0^T \|\boldsymbol{y}\|_2^2 \, \mathrm{d}t$$

and the corresponding cost function

$$\mathcal{J}_{\boldsymbol{\eta}} = \int_0^T \|\boldsymbol{\eta}\|_2^2 \, \mathrm{d}t$$

based on the dual system, it can easily be proved[9] that $\mathcal{J}_{\boldsymbol{y}} = \mathcal{J}_{\boldsymbol{\eta}}$. This property is fundamental in control theory since it can be employed to determine the observer gain matrix L for the observer design (see problem 2.1.3 in section 2.1.3) based on the solution of the dual control problem. Indeed, let $\boldsymbol{x_e}(t) = \boldsymbol{x}(t) - \hat{\boldsymbol{x}}(t)$ be the state error, the main purpose of state observer design is to minimize $\mathcal{J} = \displaystyle\int_0^T \|\boldsymbol{x_e}\|_2^2 \, \mathrm{d}t$ where $\hat{\boldsymbol{x}}$ is given by

[9] Essentially, the proof is based on two results:

1. the transformations (6), and
2. the following equalities

$$\mathcal{J}_{\boldsymbol{y}} = trace\left(C_2 W_c C_2^H\right) = trace\left(B_2^H W_o B_2\right) \quad \text{(see Burl, 1999, p. 113).}$$

(2). An elegant method of determination of the observer gain matrix then consists in minimizing the same functional $\mathcal{J}$ but by introducing the dual problem of the initial system (Huerre, 2006). We then arrive at a Linear Quadratic Regulator problem whose solution is already known (see section 4). Consequently, we will not detail thereafter the observer design (see classical textbooks Zhou et al., 1996; Burl, 1999; Skogestad and Postlethwaite, 2005, for instance) and we will concentrate on the control problem.

2.2.2.4 Balanced truncation The notions of controllability and observability, as defined respectively in sections 2.2.2.1 and 2.2.2.2, give us a means of deciding whether a state affects the system's input-output map: if a state is unobservable, it does not affect the output, and if a state is uncontrollable, it is unaffected by the input. In terms of model reduction dedicated to control (see section 2.3), in opposition to model reduction for physical understanding, it is then capital to preserve controllable and observable modes, but in which proportion? A simple answer was given by Moore (1981) for stable, linear, input-output systems. This method called balanced truncation consists in transforming the state space system into a balanced form whose controllability and observability Gramians become diagonal and equal (balanced realization), together with a truncation of those states that are both difficult to reach and to observe.

Starting from the similarity transformations given in section 2.2.1, it can be easily shown that the controllability and observability gramians become:

$$W_c \longrightarrow \mathbb{T}^{-1} W_c \left(\mathbb{T}^{-1}\right)^H \quad \text{and} \quad W_o \longrightarrow \mathbb{T}^H W_o \mathbb{T}.$$

In the system of coordinates defined by $\mathbb{T}$, we thus have for a balanced realization:

$$\mathbb{T}^{-1} W_c \left(\mathbb{T}^{-1}\right)^H = \mathbb{T}^H W_o \mathbb{T} = \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{n_x} \end{bmatrix}$$

where the Hankel singular values σ_i are real, positive and ordered by convention from largest to smallest. An equivalent way of finding the balancing transformation $\mathbb{T}$ is to compute the eigendecomposition of $W_c W_o$ ($W_c W_o = \mathbb{T}\Sigma^2\mathbb{T}^{-1}$). It can be shown (Burl, 1999) that a balanced realization exists whenever the system is stable and minimal[10]. A geometric interpretation of the balanced truncation is given in Fig. 5.

[10] A state space system is minimal if and only if the system is controllable and observable (Zhou et al., 1996). Moreover, a minimal realization of the system is associated with a matrix A of smallest possible dimension.

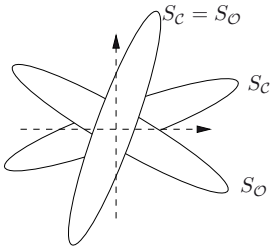


Figure 5. Geometric interpretation of the balanced truncation. $S_{\mathcal{C}}$ and $S_{\mathcal{O}}$ are respectively the controllability and observability subspaces.

An attractive feature of balanced truncation is that there exists a priori error bounds that are close to the lower bound achievable by any reduced-order model (Zhou et al., 1996, for instance). Let G denote the transfer function[11] of the LTI system (1) and G_r the corresponding transfer function of a reduced-order model of order r. It can be proved that, in any reduced-order model, the lower bound for the H_∞ error[12] is

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

and that the upper bound for the error obtained by balanced truncation is given by

$$\|G - G_r\|_\infty \leq 2 \sum_{j=r+1}^{n_x} \sigma_j.$$

If the Hankel singular values are decreasing sufficiently fast, it means that the error norm of the reduced-order model of order r is very close to the lowest possible value.

[11] For a SISO system, the transfer function G from $\boldsymbol{u}$ to $\boldsymbol{y}$ is defined as

$$G(s) = Y(s)/U(s)$$

where $U(s)$ and $Y(s)$ are the Laplace transform of $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$. Moreover, it can be demonstrated that for an LTI system, we have

$$G(s) = C_2 \left(sI - A\right)^{-1} B_2 + D_2$$

where I is the identity matrix.

[12] The H_∞ norm of the system is defined in terms of the transfer function G as:

$$\|G\|_\infty = \sup_\omega \sigma_1 \left(G(j\omega)\right)$$

where $\sigma_1 \left(\mathcal{A}\right)$ corresponds to the maximum singular value of the matrix $\mathcal{A}$ and ω represents frequency.

The procedure of balanced truncation is very attractive in terms of control but the determination of the controllability and observability gramians via the solution of Lyapunov equations is not computationally tractable for very large systems. In addition, the original method suggested by Moore (1981) is limited to the linear systems. These limitations were raised recently by Lall et al. (2002) and then by Rowley (2005) who introduced approximation methods of gramians based only on snapshots of the primal and dual systems (see section 2.2.2.3). The initial method suggested by Lall et al. (2002) was to first estimate the two gramians, and then in a second time to perform the balanced truncation. The main contribution presented in Rowley (2005) is a specific algorithm that can be used to determine the balanced truncation directly from snapshots of the system *i.e.* without needing to compute the gramians themselves. This method is called Balanced POD for deep connections that it shares with POD. The reader will find all the details of the numerical setting in Rowley (2005).

2.3 Model reduction

In section 2.2.2.4, model reduction was already evoked when the least controllable and observable modes of the system were truncated based on the decrease of the Hankel singular values. In this section, we will first justify the interest of reduced-order modeling for flow control (section 2.3.1), and then present in a general way the current methods of model reduction while giving an emphasis on projection-based methods (section 2.3.2).

2.3.1 Need for reduced-order modeling

For a wing considered at cruising flight conditions *i.e.* for a Reynolds number of about 10^7, Spalart et al. (1997) considered that to obtain numerically a converged solution, it is necessary to integrate the Navier-Stokes equations during about 5 10^6 time steps on about 10^{11} grid points. Then, in spite of the recent and considerable progresses of computers, it remains difficult to solve numerically problems where

- either, a great number of resolution of the state equations is necessary (continuation methods, parametric studies, optimization problems or optimal control,...),
- either a solution in real time is searched (active control in closed-loop control for instance).

Not surprisingly, the reduction of the costs of solving nonlinear state equations became a major issue in many scientific disciplines ranging from linear algebra to computer graphics. Sometimes, as it is the case in fluid mechanics/turbulence, model reduction has a long tradition but the objective

is more centered on the improvement of the understanding of the physical mechanisms. Let us quote for example[13]:

- Prandtl boundary layer equations (Schlichting and Gersten, 2003),
- Reynolds-Averaged Navier-Stokes models (Chassaing, 2000),
- Large Eddy Simulation (Sagaut, 2005),
- Low-order dynamical system based on Proper Orthogonal Decomposition (Aubry et al., 1988),
- Reduced-order models based on global modes (Åkervik et al., 2007),

to name a few. Since less than ten years, the methods of model reduction are mainly considered in fluid mechanics for flow control. Lately, these methods progressed considerably under the efforts of the applied mathematicians who were interested in flow control. It is this specific point of view that is retained in the following presentation of the model reduction methods.

2.3.2 Overview of model-reduction methods

Broadly speaking, model order reduction techniques fall into two major categories:

1. projection-based methods,
2. non-projection based methods.

The first group corresponds to the methods that are currently the most used in fluid mechanics. Therefore, this approach will be detailed in section 2.3.2.1. The second group consists mainly of such methods as Hankel optimal model reduction and state-residualization. More information can be found for these methods in Antoulas (2005).

2.3.2.1 Projection-based methods The projection-based methods can be used for dynamical models going from general nonlinear systems given[14] by

$$\mathcal{S} : \left\{ \begin{array}{l} \dot{\boldsymbol{x}}(t) = \boldsymbol{f}(t, \boldsymbol{x}(t), \boldsymbol{u}(t)), \\ \boldsymbol{y}(t) = \boldsymbol{g}(t, \boldsymbol{x}(t), \boldsymbol{u}(t)), \end{array} \right.$$

to LTI models

$$\mathcal{S}_{LTI} : \left\{ \begin{array}{l} E\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B_2\boldsymbol{u}(t), \\ \boldsymbol{y}(t) = C_2\boldsymbol{x}(t) + D_2\boldsymbol{u}(t), \end{array} \right.$$

[13] The traditional numerical methods used to solve partial derivative equations (finite difference, finite volume, finite element, spectral method,...) can also be classified in the framework of reduced-order models since these methods consist in reducing an infinite-dimensional problem to a finite-dimensional one (discretized problems).

[14] To simplify the formulations, we did not consider in this section the contribution of the disturbances $\boldsymbol{w}$ to the models.

written here in the so-called descriptor form. The matrix E is not necessarily invertible but, when it is the case, the traditional LTI formulation is found. For these two systems, the state variables $\boldsymbol{x}$ and output variables $\boldsymbol{y}$ are respectively of size n_x and n_y.

The objective of reduced-order modeling is to determine for $\mathcal{S}$ and $\mathcal{S}_{LTI}$ the corresponding simplified models

$$\widehat{\mathcal{S}} : \left\{ \begin{array}{l} \dot{\widehat{\boldsymbol{x}}}(t) = \widehat{\boldsymbol{f}}(t, \widehat{\boldsymbol{x}}(t), \boldsymbol{u}(t)), \\ \widehat{\boldsymbol{y}}(t) = \widehat{\boldsymbol{g}}(t, \widehat{\boldsymbol{x}}(t), \boldsymbol{u}(t)), \end{array} \right.$$

and

$$\widehat{\mathcal{S}}_{LTI} : \left\{ \begin{array}{l} \widehat{E}\dot{\widehat{\boldsymbol{x}}}(t) = \widehat{A}\widehat{\boldsymbol{x}}(t) + \widehat{B_2}\boldsymbol{u}(t), \\ \widehat{\boldsymbol{y}}(t) = \widehat{C_2}\widehat{\boldsymbol{x}}(t) + \widehat{D_2}\boldsymbol{u}(t) \end{array} \right.$$

where the control $\boldsymbol{u}$ is unchanged. These simplified models are now called reduced-order models since $\widehat{\boldsymbol{x}} \in \mathbb{R}^r$ with $r \ll n_x$ and $\boldsymbol{y} \simeq \widehat{\boldsymbol{y}} \in \mathbb{R}^{n_y}$. A simplified description of model reduction is given in Fig. 6.

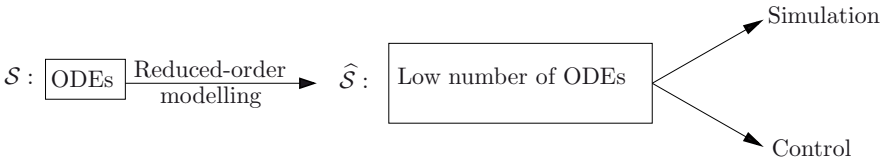


Figure 6. Broad framework of reduced-order modelling (after Antoulas, 2005).

If we want that these reduced-order models can be really usable for the applications concerned, it is necessary that the methods used to derive these simplified models satisfy various constraints:

1. Small approximation error for all admissible input signals $\boldsymbol{u}$ *i.e.*

$$\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\| < \epsilon \times \|\boldsymbol{u}\| \qquad \text{with } \epsilon \text{ a tolerance.}$$

   It means that we need to have estimates of computable error bounds.
2. Stability and passivity (no generation of energy) preserved.
3. Procedure of model reduction numerically stable and efficient.
4. If possible, automatic generation of models.

In what follows we will describe an algorithm that can be used to derive a reduced-order model of any dynamical system. This algorithm, called Petrov-Galerkin projection, is based on a general bi-orthogonal projection

basis. Let V and W be two[15] bi-orthogonal matrices of size $\mathbb{C}^{n_x \times r}$, and $Q \in \mathbb{C}^{n_x \times n_x}$ be the weight matrix such that

$$W^H Q V = I_r$$

where I_r is the identity matrix of size r. In the first step of the algorithm, $\boldsymbol{x}$ is projected on the space spanned by the columns of V *i.e.* $\boldsymbol{x} = V\widehat{\boldsymbol{x}}$. In the second step, this projection is inserted in the dynamical system where we have introduced the residual $\mathcal{R}$ of the state equations. At this stage, we obtain for $\mathcal{S}$

$$\left\{ \begin{array}{l} \boldsymbol{\mathcal{R}} = V\dot{\widehat{\boldsymbol{x}}}(t) - \boldsymbol{f}(t, V\widehat{\boldsymbol{x}}(t), \boldsymbol{u}(t)), \\ \widehat{\boldsymbol{y}}(t) = \boldsymbol{g}(t, V\widehat{\boldsymbol{x}}(t), \boldsymbol{u}(t)), \end{array} \right.$$

and for $\mathcal{S}_{LTI}$

$$\left\{ \begin{array}{l} \boldsymbol{\mathcal{R}} = EV\dot{\widehat{\boldsymbol{x}}}(t) - AV\widehat{\boldsymbol{x}}(t) - B_2\boldsymbol{u}(t), \\ \widehat{\boldsymbol{y}}(t) = C_2 V\widehat{\boldsymbol{x}}(t) + D_2\boldsymbol{u}(t). \end{array} \right.$$

The last step corresponds to a weak projection of the residual on the space spanned by the columns of W *i.e.* $W^H Q\,\boldsymbol{\mathcal{R}} = \mathbf{0}_r$. Finally, we obtain the reduced-order model $\widehat{\mathcal{S}}$ where

$$\widehat{\mathcal{S}} : \left\{ \begin{array}{l} \widehat{\boldsymbol{f}}(t, \widehat{\boldsymbol{x}}(t), \boldsymbol{u}(t)) = W^H Q\, \widehat{\boldsymbol{f}}(t, V\widehat{\boldsymbol{x}}(t), \boldsymbol{u}(t)), \\ \widehat{\boldsymbol{g}}(t, \widehat{\boldsymbol{x}}(t), \boldsymbol{u}(t)) = \boldsymbol{g}(t, V\widehat{\boldsymbol{x}}(t), \boldsymbol{u}(t)), \end{array} \right.$$

and the reduced-order model $\widehat{\mathcal{S}}_{LTI}$ where

$$\begin{array}{ll} \widehat{A} = W^H Q A V, & \widehat{B_2} = W^H Q B_2, \\ \widehat{C_2} = C_2 V, & \widehat{D_2} = D_2, \\ \widehat{E} = W^H Q E V. & \end{array}$$

For the choice of the matrices V and W, various possibilities exist for the linear systems:

1. In the case of Krylov methods (Gugercin and Antoulas, 2006), it corresponds to the projection on the Krylov subspace of the controllability gramian coupled with an identification of the moments of the transfer function.
2. For balanced realizations, this choice corresponds to the projection on dominant modes of the controllability and observability gramians as already discussed in section 2.2.2.4.

[15] When $V \neq W$, it corresponds to an oblique projection, and when $V \equiv W$ it is called Galerkin projection or orthogonal projection.

3. For instabilities, the projection is made on the global and adjoint global modes (Schmid and Henningson, 2001; Barbagallo et al., 2009).
4. Finally, in the case of the Proper Orthogonal Decomposition (Lumley, 1967; Sirovich, 1987), it corresponds to the projection on the subspace determined optimally with snapshots of the system (see the contribution by B. Noack et al. in this book).

For the non-linear systems, the situation is different because, until now, there exists only the Proper Orthogonal Decomposition what explains its intensive use in the past years.

3 Optimal control theory

3.1 Constrained optimization problems

3.1.1 Abstract description

All the constrained optimization problems appearing in fluid mechanics and heat transfers (shape optimization, active flow control, optimal growth, control of thermal systems, ...) can be described mathematically by the following quantities[16] (Gunzburger, 1997a, 2003):

state variables ϕ which describe the flow. Depending on the problem, these variables might be mechanical or thermodynamic, for instance velocity vectors, pressure, temperature, ...

control parameters c**.** In practice, these variables occur as boundary conditions of the state equations[17], when the control is applied at the boundaries of the domain, or directly as a source term in the state equations if the control is distributed inside the domain (volume forcing). In data assimilation (meteorology, oceanography) and for optimal growth (see section 5) these control parameters intervene as initial conditions. According to the application, these parameters might be velocities prescribed at the boundaries (suction/blowing), heat flux or temperature at a wall, or for a shape optimization problem (Mohammadi and Pironneau, 2001), it might be variables allowing to describe

[16] To simplify the presentation, all the variables are here considered as scalars. However, the method extends naturally to the case of vectorial variables. For instance, an optimal control problem is solved for the Linear Quadratic Regulator approach in section 4, and for the three-dimensional Navier-Stokes equations in Bewley et al. (2001) or El Shrif (2008).

[17] Here, we use the traditional terminology in optimal control and call state equations, the equations which govern the dynamics of the system. Other terminologies are primal or direct equations.

geometrically the shape of the boundary. In this last case, the control parameters are rather called design variables.

a cost or objective functional $\mathcal{J}$ which describes a measure of the objectives we wish to achieve. It might be drag minimization, maximization of lift or heat flux, stabilization of a temperature, flow targets, ... This functional $\mathcal{J}$ depends on the state variables ϕ and on the control parameters c, *i.e.* $\mathcal{J}(\phi, c)$.

physical constraints F which represent the evolution of the state variables ϕ in terms[18] of the control parameters c with respect to the physical laws. Mathematically, these constraints are noted:

$$F(\phi, c) = 0.$$

In fluid mechanics, these constraints correspond generally to the Navier-Stokes equations and their associate initial and boundary conditions. If a problem of optimal disturbance is concerned then the initial condition is imposed as a constraint (see section 5). If the control is exerted at the boundaries of the flow domain, the boundary condition can also be included as constraint (see section 6 for an example). Moreover, we will see in section 3.1.2 that an additional constraint must in general be added so that the problem is well posed mathematically.

Finally, the constrained optimization problem can be stated in the following way:

> *determine the state variables ϕ and the control parameters c, such that the objective functional $\mathcal{J}$ is optimal (minimum or maximum according to the case) under the constraints F.*

3.1.2 Ill-posed optimization problem and choice of the cost functional

The choice of the cost functional $\mathcal{J}$ is central in an optimization problem. From a mathematical point of view, the physical quantity to be optimized is represented by

$$\mathcal{J} = \mathcal{M}$$

where $\mathcal{M}$ is an appropriate measure of any physical quantity of interest: drag, lift, disturbance energy, ... The choice of this cost functional is essential in practice so that the optimization problem is well posed. This choice

[18] Rigorously, it would be necessary to note the variables $\phi(c)$ because ϕ depend on the control variables c via the constraints. However, to reduce the notations, we will note the state variables simply as ϕ.

is sometimes difficult to achieve. For instance, it is not obvious to know in advance if it is better to choose as cost functional a measure of the drag to minimize this quantity. In some cases (Bewley et al., 2001; El Shrif, 2008), it seems that it is preferable to minimize the averaged kinetic energy of the flow in order to minimize the drag. In addition, beyond the mathematical difficulty that is raised, we can imagine that the implementation of the control will be eased if the cost functional is based on a relevant quantity for the physics of the problem.

In general, there is no explicit relation between the objective to be reached and the control variable. This can involve that the optimization problem is ill-posed and that its solution is then divergent. To solve this difficulty, the cost of the control should be limited[19]. Let $\mathcal{M}_c$ be a measure of the cost of the control, this limitation can be done:

1. By adding an additional constraint to the physical constraints (F)

   This constraint corresponds to a maximum value which should not be exceeded by the control cost. Let $(\mathcal{M}_c)_{max}$ be an arbitrary positive constant, the problem is then equivalent to impose that $\mathcal{M}_c \leq (\mathcal{M}_c)_{max}$. In optimization, the inequality constraints make intervene optimality conditions known as Karush-Kuhn-Tucker (Bonnans et al., 2003) which are often delicate to take into account. For this reason, it is generally preferred to retain equality type constraints which can be imposed more easily using Lagrange multipliers (see section 3.2). It will thus be sufficient to set an additional constraint of the type $\mathcal{M}_c = \mathcal{M}_c^u$ where $\mathcal{M}_c^u > 0$ is a cost imposed by the user, to do not have to change the nature of the optimization problem to be solved.
2. By modifying the cost functional $\mathcal{J}$

   A possible modification of the cost functional is to consider

   $$\mathcal{J} = \mathcal{M} + \ell \mathcal{M}_c$$

   where ℓ is a positive real constant whose value is fixed by the user according to the importance given to the cost of the control. If the value of the parameter ℓ is low then it means that the cost of the control is not a priority in the practical implementation (low costs of control). On the contrary, if the value of ℓ is high, then the cost of the control is a priority (expensive control). A more thorough discussion is given in section 4 for the LQR control.

[19] Apart from a mathematical justification, a limitation of the control cost is necessary since from an economic point of view the ratio saving/cost is a determining factor.

Since the approach 1 with the inequality constraint is more difficult to implement, the limitation of the cost control is introduced in most of the studies through a modification of the cost functional $\mathcal{J}$. In addition, another interest of the approach is that the penalization parameter ℓ clearly introduces a compromise between the objective to be reached (saving) and the importance of the control (cost).

3.1.3 Three different approaches

The current methods of solving a constrained optimization problem are distinguished in two classes (Gunzburger, 1997a, 2003). The first consists in transforming the original problem of optimization with constraints in an unconstrained optimization problem via the method of Lagrange multipliers (section 3.2) giving optimality conditions of first order. The control is then obtained by resolution of a system of coupled partial derivative equations known as *optimality system*. The second class of methods uses directly an algorithm of optimization (see section 3.3), which then requires the determination of the gradient of the objective functional, or at least of an approximation of this one. Two approaches can be used to evaluate this gradient: approach by the sensitivities described in section 3.3.1 and approach by the adjoint variables developed in section 3.3.2.

3.2 Adjoint or Lagrange multiplier methods

The principle consists to enforce implicitly the constraints of the problem via Lagrange multipliers[20]. A new functional $\mathcal{L}$, known as Lagrange functional, is then introduced to define an unconstrained optimization problem. The validity of such approach can be rigorously demonstrated using theories developed in optimal control by applied mathematicians (see Gunzburger, 1997b, for elements of answers).

3.2.1 Introduction of the Lagrange multiplier

In this section, we follow the procedure outlined in Guegan et al. (2006) and introduce a single vector space $\Theta = \phi \times c \times \xi$ where ξ is the adjoint variable or Lagrange multiplier associated to the constraint $F = 0$. Let $\mathbf{\Phi}^i = \left(\phi^i, c^i, \xi^i\right)$ with $i = I, II$ be two arbitrary elements of Θ, we define a

[20]The Lagrange multipliers are often called adjoint variables. Strictly speaking, this appellation is abusive. A justification will be given *a posteriori* when the adjoint equations of the optimality system will be derived.

generalized inner product as

$$\{\mathbf{\Phi}^I, \mathbf{\Phi}^{II}\} = \left\langle \phi^I, \phi^{II} \right\rangle_s + \left\langle c^I, c^{II} \right\rangle_c + \left\langle \xi^I, \xi^{II} \right\rangle_a , \tag{7}$$

where $\langle \cdot, \cdot \rangle_s$, $\langle \cdot, \cdot \rangle_c$ and $\langle \cdot, \cdot \rangle_a$ are three inner products. These scalar products can be defined in space, in time, in space-time or defined specifically according to the problem which is considered (see later in sections 4, 5 and 6). In the case of optimal disturbances, we will consider in section 5 an energy inner product in order to determine the energy of the initial disturbances.

The constraint F is then enforced by introducing a Lagrangian functional $\mathcal{L}$ defined as:

$$\mathcal{L}(\phi, c, \xi) \triangleq \mathcal{J}(\phi, c) - \langle F(\phi, c), \xi \rangle_a . \tag{8}$$

The new unconstrained optimization problem can then be stated as:

> *determine the state variables ϕ, the control parameters c and the adjoint variables ξ, such as the Lagrangian functional $\mathcal{L}$ reaches an extremum.*

3.2.2 Derivation of the optimality system

The Lagrangian functional $\mathcal{L}$ admits an extremum when $\mathcal{L}$ is rendered stationary. A first-order necessary condition for an extremum of $\mathcal{L}$ is that its first-order variation $\delta\mathcal{L}$ is equal to 0 *i.e.*

$$\delta\mathcal{L} = \frac{\partial \mathcal{L}}{\partial \phi}\delta\phi + \frac{\partial \mathcal{L}}{\partial c}\delta c + \frac{\partial \mathcal{L}}{\partial \xi}\delta\xi = 0.$$

For simplifying further the expression of $\delta\mathcal{L}$, each argument of $\mathcal{L}$ is considered[21] as independent of the others. The necessary condition is then equivalent to

$$\frac{\partial \mathcal{L}}{\partial \phi}\delta\phi = \frac{\partial \mathcal{L}}{\partial c}\delta c = \frac{\partial \mathcal{L}}{\partial \xi}\delta\xi = 0, \tag{9}$$

where the variations $\delta\phi$, δc and $\delta\xi$ are arbitrary.

Equivalently, the stationary points of the Lagrangian $\mathcal{L}$ can be characterized by the gradients of $\mathcal{L}$ with respect to all the variables. By convention, the gradients of $\mathcal{L}$ with respect to ϕ, c and ξ are denoted in the following respectively by $\mathbf{\nabla}_\phi\mathcal{L}$, $\mathbf{\nabla}_c\mathcal{L}$ and $\mathbf{\nabla}_\xi\mathcal{L}$. The stationary points of $\mathcal{L}$ then correspond to:

$$\mathbf{\nabla}_\phi\mathcal{L} = \mathbf{\nabla}_c\mathcal{L} = \mathbf{\nabla}_\xi\mathcal{L} = 0. \tag{10}$$

[21] Note that this is obviously wrong for the original problem involving $\mathcal{J}$ since the variables ϕ and c were constrained to satisfy $F(\phi, c) = 0$.

These gradients are determined as projections of $\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{\Phi})$, gradient of the Lagrangian at point $\boldsymbol{\Phi}$, onto the different subspaces corresponding to each of the variables ϕ, c and ξ. Assuming that $\mathcal{L}$ is Fréchet-differentiable, $\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{\Phi})$ is such that for any variation $\delta\boldsymbol{\Phi}$ we have:

$$\{\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{\Phi}), \delta\boldsymbol{\Phi}\} = \mathrm{d}\ \mathcal{L}|_{\boldsymbol{\Phi}}\,(\delta\boldsymbol{\Phi}), \tag{11}$$

where $\{\cdot,\cdot\}$ is the scalar product introduced in (7). Furthermore, the Gâteau differential $\mathrm{d}\ \mathcal{L}|_{\boldsymbol{\Phi}}$ of the Lagrangian $\mathcal{L}$ evaluated at point $\boldsymbol{\Phi}$ is given by

$$\mathrm{d}\ \mathcal{L}|_{\boldsymbol{\Phi}}\,(\delta\boldsymbol{\Phi}) = \lim_{\epsilon \longrightarrow 0} \frac{\mathcal{L}(\boldsymbol{\Phi} + \epsilon\delta\boldsymbol{\Phi}) - \mathcal{L}(\boldsymbol{\Phi})}{\epsilon}. \tag{12}$$

The expressions (9) and (10) correspond to a necessary and sufficient condition for determining a local extremum of $\mathcal{L}$, but it constitutes only a necessary condition for obtaining a minimum or a maximum. This type of method thus ensures only to obtain a local extremum but not a global one. We then have to keep in mind while using it that the algorithm of optimization may be remain trapped in a local minimum without any physical interest. Obviously, it would be better to use methods of global optimization (genetic algorithms for instance) but those are still too expensive to be used currently in fluid mechanics.

We will now derive the optimality system by setting successively the first variations of $\mathcal{L}$ with respect to the adjoint variable ξ, direct variable ϕ and control variable c equal to zero.

▷ **Determination of $\boldsymbol{\nabla}_{\xi}\mathcal{L}$ or directional derivative in the direction $\delta\xi$:**

A variation $\delta\boldsymbol{\Phi}$ given by $(0, 0, \delta\xi)$ is considered. Using the definition (8) of the Lagrangian functional $\mathcal{L}$, and the definitions (11) and (12) respectively of the Fréchet and Gâteau derivatives, it yields to:

$$\langle \boldsymbol{\nabla}_{\xi}\mathcal{L}, \delta\xi \rangle_a = \lim_{\epsilon \longrightarrow 0} \frac{\mathcal{J}(\phi, c) - \langle F(\phi, c), \xi + \epsilon\delta\xi \rangle_a - \mathcal{J}(\phi, c) + \langle F(\phi, c), \xi \rangle_a}{\epsilon} = 0,$$

i.e. after simplification,

$$\langle F(\phi, c), \delta\xi \rangle_a = 0.$$

Since $\delta\xi$ is arbitrary, it can be deduced that

$$F(\phi, c) = 0, \tag{13}$$

what corresponds to the constraints of the original problem of optimization.

Thus, setting the first variation of $\mathcal{L}$ with respect to the Lagrange multiplier equal to zero gives back the equations of constraints (**state equations**).

▷ **Determination of $\boldsymbol{\nabla}_\phi\mathcal{L}$ or directional derivative in the direction $\delta\phi$**:

In this case, we consider a perturbation $\delta\mathbf{\Phi}$ given by $(\delta\phi, 0, 0)$. Following the same procedure as for the state equations, we obtain:

$$\langle \boldsymbol{\nabla}_\phi\mathcal{L}, \delta\phi\rangle_s = \lim_{\epsilon\longrightarrow 0} \frac{\mathcal{J}(\phi+\epsilon\delta\phi, c) - \langle F(\phi+\epsilon\delta\phi, c), \xi\rangle_a - \mathcal{J}(\phi, c) + \langle F(\phi, c), \xi\rangle_a}{\epsilon} = 0.$$

Introducing the Taylor series of $\mathcal{J}$ and F at the order $\mathcal{O}(\epsilon)$, the previous relation becomes:

$$\lim_{\epsilon\longrightarrow 0}\left(\frac{\partial\mathcal{J}}{\partial\phi}\delta\phi - \langle\frac{\partial F}{\partial\phi}\delta\phi, \xi\rangle_a + \mathcal{O}(1)\right) = 0$$

i.e.

$$\frac{\partial\mathcal{J}}{\partial\phi}\delta\phi - \langle\frac{\partial F}{\partial\phi}\delta\phi, \xi\rangle_a = 0.$$

The first term can be expressed with the inner product $\langle .,.\rangle_a$ yielding to:

$$\langle\frac{\partial\mathcal{J}}{\partial\phi}\delta\phi, 1\rangle_a - \langle\frac{\partial F}{\partial\phi}\delta\phi, \xi\rangle_a = 0.$$

Introducing the adjoint operator (see appendix A) $(.)^+$ with respect to the inner product $\langle .,.\rangle_a$, we can write the previous relation as:

$$\langle\delta\phi, \left(\frac{\partial\mathcal{J}}{\partial\phi}\right)^+\rangle_a - \langle\delta\phi, \left(\frac{\partial F}{\partial\phi}\right)^+\xi\rangle_a = 0.$$

This equality must be verified whatever the variation $\delta\phi$ of ϕ is. We then obtain the **adjoint equations**:

$$\left(\frac{\partial F}{\partial\phi}\right)^+\xi = \left(\frac{\partial\mathcal{J}}{\partial\phi}\right)^+. \tag{14}$$

These equations correspond to the adjoint of the state equations linearized around the state. They are thus linear in the adjoint variables ξ,

thus facilitating their resolution. In addition, the Lagrange multipliers satisfy the equations associated to the constraints with a source term resulting from the cost functional; this justifies the name of adjoint variables given to the Lagrange multipliers.

When the state equations are non-linear (a typical case being the Navier-Stokes equations), the adjoint equations (14) depend on the solution of the state equations (13). The solution of the direct equations is then required before the adjoint equations can be solved. Moreover, it can be shown that for time-dependent problems, the adjoint equations are defined backward in time (see section 4.3.2 for the LQR control). Therefore, in solving the adjoint equations, the calculations start from the terminal condition rather than an initial condition. This characteristic poses a unique challenge for efficient solution of unsteady adjoint equations since the solution of the adjoint equations at each time step requires the solution of the direct equations at the same time step. This leads to a serious demand on computer memory. Some solutions to this problem are proposed in section 3.2.3.

▷ **Determination of $\boldsymbol{\nabla}_c\mathcal{L}$ or directional derivative in the direction δc:**

Here, we consider $\delta\boldsymbol{\Phi} = (0, 0, \delta c)$. We obtain immediately that:

$$\langle \boldsymbol{\nabla}_c\mathcal{L}, \delta c\rangle_c = \lim_{\epsilon \longrightarrow 0} \frac{\mathcal{J}(\phi, c+\epsilon\delta c) - \langle F(\phi, c+\epsilon\delta c), \xi\rangle_a - \mathcal{J}(\phi, c) + \langle F(\phi, c), \xi\rangle_a}{\epsilon} = 0.$$

Then we proceed as we did for the adjoint equations. We first introduce the Taylor series of $\mathcal{J}$ and F at the order $\mathcal{O}(\epsilon)$. It comes

$$\frac{\partial \mathcal{J}}{\partial c}\delta c - \langle \frac{\partial F}{\partial c}\delta c, \xi\rangle_a = 0,$$

i.e. after writing the first term with an inner product

$$\langle \frac{\partial \mathcal{J}}{\partial c}\delta c, 1\rangle_a - \langle \frac{\partial F}{\partial c}\delta c, \xi\rangle_a = 0.$$

Finally, we introduce the adjoint operators to yield

$$\langle \boldsymbol{\nabla}_c\mathcal{L}, \delta c\rangle_c = \langle \delta c, \left(\frac{\partial \mathcal{J}}{\partial c}\right)^+ \rangle_a - \langle \delta c, \left(\frac{\partial F}{\partial c}\right)^+ \xi\rangle_a = 0. \tag{15}$$

Since the variation δc of c is arbitrary, we obtain the **optimality conditions**:

$$\left(\frac{\partial F}{\partial c}\right)^{+} \xi = \left(\frac{\partial \mathcal{J}}{\partial c}\right)^{+}. \tag{16}$$

These optimality conditions are first order. They are satisfied exactly only when an extremum of the cost functional is achieved. One advantage of the Lagrangian-based formulation is to provide not only the optimality condition but also an expression for the gradient of the cost functional $\mathcal{J}$ with respect to the control c. Indeed, in the constrained subspace where $F(\phi, c) = 0$, the gradient of the Lagrangian simply reduces to

$$\boldsymbol{\nabla}_c \mathcal{L} = \boldsymbol{\nabla}_c \mathcal{J}. \tag{17}$$

Starting from (15), we can determine an expression for $\boldsymbol{\nabla}_c \mathcal{L}$ when the relation between the inner products $\langle \cdot, \cdot \rangle_c$ and $\langle \cdot, \cdot \rangle_a$ is known. A striking example can be found in section 5.1.2.3 for the optimal growth perturbation. Finally, the optimality condition (16) must be interpreted as the gap to zero of the gradient of the cost functional $\boldsymbol{\nabla}_c \mathcal{J}$.

The necessary conditions (13), (14) and (16) form a coupled system of partial differential equations called *optimality system*. When the number of unknowns of the optimality system is not too important, a direct method of resolution known as "one shot method" can be used to obtain without iteration the optimal solution (see Galletti et al., 2007, for example, for an application to the calibration of POD reduced order models). Unfortunately, in fluid mechanics, the optimization problems comprise too many degrees of freedom (10^7 for the turbulent channel flow studied by Direct Numerical Simulation in Bewley et al. 2001) to prevent the use of a direct method. It turns out that it is necessary to have recourse to iterative methods for which the optimal control is approximated step by step until convergence. This approach is described in the next section.

3.2.3 Numerical resolution

The optimality system can be solved iteratively in the following manner. The resolution is initialized with a given control $c^{(0)}$ (here and below, the superscripts (n) denote the iteration number). Then, for $n = 0, 1, 2, ...$ and as long as a given criterion of convergence is not satisfied, the following steps are carried out:

Step 1: Solve the state equations (13) forward in time to determine the state variables $\phi^{(n)}$

$$F(\phi^{(n)}, c^{(n)}) = 0.$$

Step 2: Use the state variables computed in step 3.2.3 to solve the adjoint equations (14) backward in time for the adjoint variables $\xi^{(n)}$

$$\left(\frac{\partial F}{\partial \phi}\right)^{+(n)} \xi^{(n)} = \left(\frac{\partial \mathcal{J}}{\partial \phi}\right)^{+(n)}.$$

Step 3: Use the state variables $\phi^{(n)}$ computed in step 3.2.3 and the adjoint variables $\xi^{(n)}$ computed in step 3.2.3 to estimate the optimality conditions (16)

$$\left(\frac{\partial \mathcal{J}}{\partial c}\right)^{+(n)} = \left(\frac{\partial F}{\partial c}\right)^{+(n)} \xi^{(n)}.$$

Step 4: Set $n = n + 1$ and return at step 3.2.3 until a given criterion of convergence is satisfied.

At stage 3.2.3 of the iterative process, a new control can be determined with a gradient type method:

$$c^{(n+1)} = c^{(n)} - \omega^{(n)} \left(\boldsymbol{\nabla}_c \mathcal{J}\right)^{(n)}. \tag{18}$$

The relaxation parameter $\omega^{(n)}$ is given using a line search method (Nocedal and Wright, 1999). It can be shown that this simple iterative method corresponds to a steepest descent algorithm for the unconstrained functional $\mathcal{J}(\phi(c), c)$. Figure 7 represents schematically the above algorithm.

This iterative procedure enables the reduction of the memory required for the resolution of the optimality system. However, for time-dependent problems, the adjoint equations are marched backward in time. For solving the adjoint equations at any time step, it is then necessary to know the solution of the state equations at the same time step. According to the relative importance of CPU and memory in the optimization procedure, several numerical strategies are possible. Let us consider[22] for the discussion that the memory is the limiting criterion in our application. The first method, referred to as instantaneous control in the control literature (see Fig. 8), consists in dividing the time horizon T_o on which optimization is performed

[22] In most of the applications, the memory is indeed the limiting criterion. One exception is the case of real-time flow control where the main objective is to reduce the CPU time necessary for solving the optimality system.

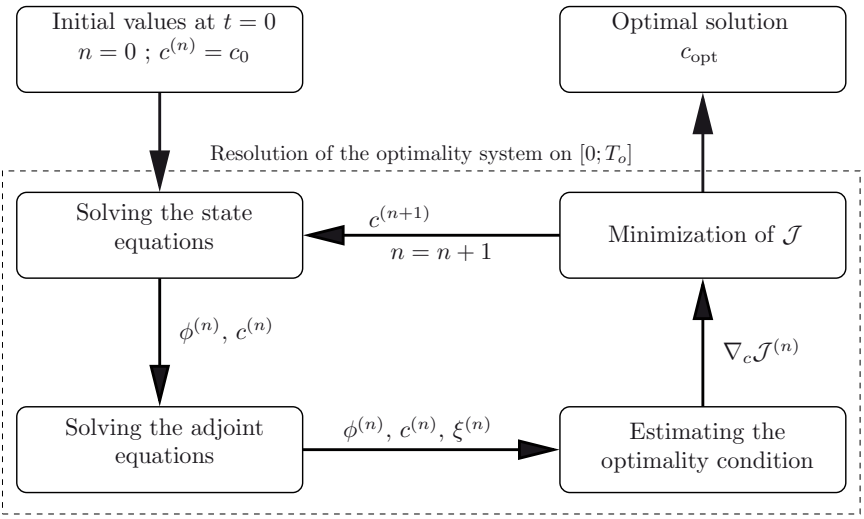


Figure 7. Iterative resolution of the optimality system (schematic representation).

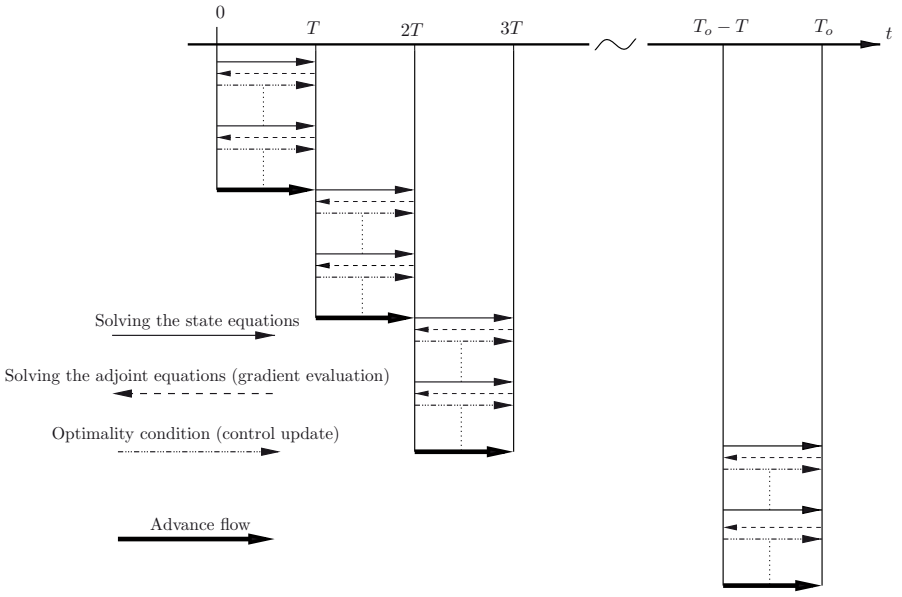


Figure 8. Instantaneous control approach.

in N smaller time window $[T_k; T_{k+1}]$ $(k = 0, \cdots, N)$ of size T. The optimality system is successively solved on each window, where the state reached by the optimized flow at the end of a given interval is taken as guess values for optimization on the following interval (see Fig. 9). Of course, instantaneous control will not lead in general to the same control that would be obtained by optimizing the cost functional over T_o. However, this strategy was used successfully in the past for the turbulent channel flow (Bewley et al., 2001; Chang, 2000; El Shrif, 2008) and for the cylinder wake flow in laminar regime (Protas, 2002; Bergmann et al., 2005). Another method for reduc-

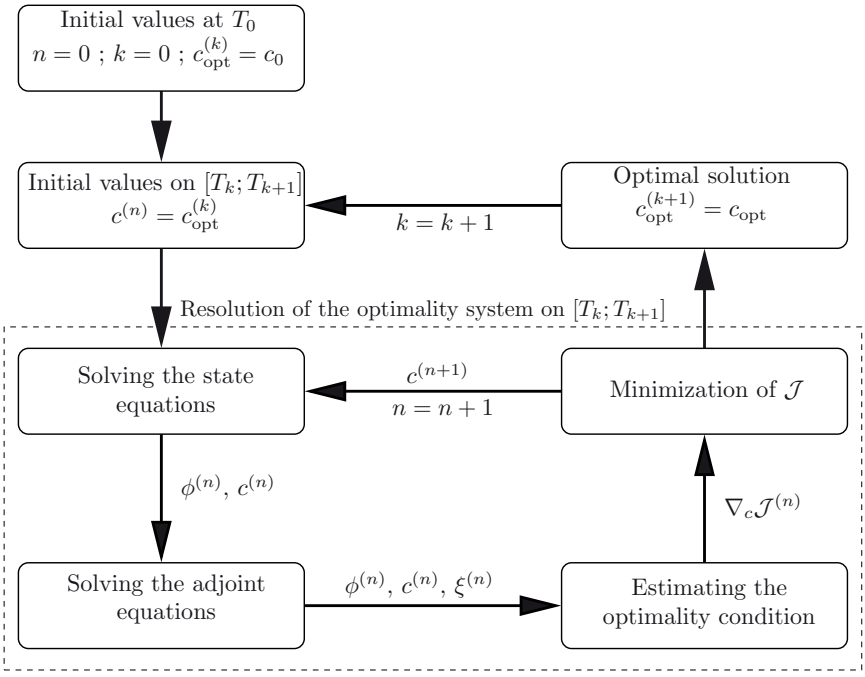


Figure 9. Iterative resolution of the optimality system for the instantaneous control approach (schematic representation).

ing the memory requirement is the use of reduced-order models (see section 2.3) as state equations. Depending on the configurations, these approximation models can be mathematically or physically derived. For instance, in Bergmann et al. (2005), a POD reduced-order model was derived as state equations for the cylinder wake flow and, in El Shrif (2008), a Large Eddy Simulation was used as approximate model for the turbulent channel flow.

Compared to previous studies where a Direct Numerical Simulation was employed as state equations, drastic reductions of the computational costs (memory and CPU) were found. The conceptual drawback of this method is that there is no mathematical assurance that the solution of the optimization algorithm working with the approximation models will correspond to the solution of the optimization problem for the original dynamical system. To circumvent this limitation, one possibility is to embed POD reduced-order modeling in the framework of trust-region optimization as was done with the Trust Region Proper Orthogonal Decomposition algorithm introduced in Fahl (2000). Indeed, with this algorithm, mathematical proofs exist that the solutions converge at least to a local optimum of the original high-fidelity problem (see Bergmann and Cordier, 2008, for numerical evidence in the case of the cylinder wake flow). A third method for simplifying the resolution of the optimality system is to change the time-reversed nature of the adjoint equations. For that, Wang et al. (2008) proposed recently to use a Monte Carlo linear solver for solving forward in time the unsteady adjoint equations. This method was only demonstrated for the Burgers' equation. Many issues remain to be solved before this method can be employed for the Navier-Stokes equations. A last method is using a dynamic checkpointing scheme. The basic idea of checkpointing methods is to solve the state equations first, and store its solution at selected time steps called checkpoints. When the adjoint equations are integrated backward in time, the solutions at corresponding time steps are calculated by re-solving the state equations starting from the nearest checkpoint (Griewank and Walther, 2008). Recently, Wang et al. (2009) suggested a dynamic checkpoint scheme that minimizes the maximum number of recalculations for each time step, and guarantees an efficient calculation of the adjoint equations when memory storage is limited. Moreover, in contrast to previous checkpointing methods, their scheme has provable performance bounds and works for arbitrarily large number of time steps.

3.3 Optimization methods

The second class of methods for solving constrained optimization problems corresponds to the use of optimization algorithms. Many of these algorithms require the gradient of the cost functional with respect to the control parameters or at least an approximation of this gradient. Since the iterative algorithm presented in section 3.2.3 is equivalent to a steepest descent algorithm which does not converge very quickly, it is preferable to use more sophisticated methods of optimization. A typical algorithm of optimization is written as follows:

Start with an initial guess $c^{(0)}$ for the control. Then, for $n = 0, 1, 2, \ldots$ and until a given convergence criterion is achieved, the following phases are carried out:

Step 1: Solve the state equations $F(\phi^{(n)}, c^{(n)}) = 0$ to determine the state variables $\phi^{(n)}$.

Step 2: Compute the gradient of the functional $\mathcal{J}$ with respect to the control variables c: $d\mathcal{J}/dc|_{c^{(n)}}$ or $(\boldsymbol{\nabla}_c \mathcal{J})^{(n)}$.

Step 3: Use the results of stages 3.3 and 3.3 to compute an increment $\delta c^{(n)}$.

Step 4: Determine new control parameters

$$c^{(n+1)} = c^{(n)} + \delta c^{(n)}$$

Step 5: Set $n = n + 1$ and return at stage 3.3.

For each iteration of the optimization algorithm, it is necessary to solve at least one state equation. To reduce the computational costs, it is thus interesting to replace the state equations by reduced-order models (see section 2.3). In addition, many points of this algorithm must be specified:

1. How to determine the gradient of the cost functional at stage 3.3?
2. How to determine the increment of the control at stage 3.3?
3. How to choose the criterion of convergence for the optimization algorithm?

The possible methods that can be used to determine the gradient of the cost functional is discussed in more details in section 3.3.1. For determining the increment of the control, different gradient-based optimization methods (Nocedal and Wright, 1999) can be used such as non-linear conjugate gradient methods (Fletcher-Reeves, Polak-Ribire, Hestenes-Stiefel,...), trust-region methods, quasi-Newton methods (BFGS, DFP, SR1,...). Finally, for the convergence criterion, a stopping test $\| d\mathcal{J}/dc|_{c^{(n)}} \| < \epsilon$ for $\epsilon \longrightarrow 0$ is in general sufficient.

3.3.1 Functional gradients through sensitivities

To obtain the gradient of the cost functional with respect to the control variables at stage 3.3 of the previous algorithm, the chain rule can be applied to $\mathcal{J}(\phi(c), c)$. The following expression is then obtained[23]:

$$\frac{d\mathcal{J}(\phi, c)}{dc} = \frac{\partial \mathcal{J}(\phi, c)}{\partial \phi} \frac{d\phi}{dc} + \frac{\partial \mathcal{J}(\phi, c)}{\partial c}. \tag{19}$$

[23] There will be an equation similar to (19) for each control parameter.

Since the cost functional $\mathcal{J}$ depends explicitly of ϕ and c (see section 3.1.2), the partial derivatives $\dfrac{\partial \mathcal{J}}{\partial \phi}$ and $\dfrac{\partial \mathcal{J}}{\partial c}$ will be "easy" to determine. On the other hand, the implicit dependency of the state variables ϕ on the control variables c renders more delicate the evaluation of the sensitivities $\dfrac{d\phi}{dc}$. In practice, two approaches can however be used:

By finite differences. Indeed, the sensitivities can be approximated by finite difference in the following way:

$$\left.\frac{d\phi}{dc}\right|_{c^{(n)}} \simeq \frac{\phi(c^{(n)} + \Delta c^{(n)}) - \phi(c^{(n)})}{\Delta c^{(n)}}$$

where the step size $\Delta c^{(n)}$ is chosen as small as possible to minimize truncation error but not too small for avoiding that errors due to subtractive cancellation become dominant.

The cost of calculating sensitivities with finite differences is proportional to the number of design variables. This method is expensive numerically since the state equations must be solved for each perturbation of $c^{(n)}$.

By solving linear systems. Another method for calculating the sensitivities consists in differentiating the state equation $F(\phi, c) = 0$. We then obtain:

$$dF = \frac{\partial F}{\partial \phi}\, d\phi + \frac{\partial F}{\partial c}\, dc = 0,$$

i.e.

$$\left(\left.\frac{\partial F}{\partial \phi}\right|_{c^{(n)}}\right) \left.\frac{d\phi}{dc}\right|_{c^{(n)}} = -\left.\frac{\partial F}{\partial c}\right|_{c^{(n)}}. \tag{20}$$

Finally, the sensitivities are obtained by resolution of this linear system. The major drawback of this approach is that it is necessary to solve as many[24] linear systems that there are control parameters. This method is however much more efficient than the approach by finite differences: indeed, the sensitivities are determined exactly by resolution of linear systems.

[24] One can however reduce the computational cost related to this method by noticing that only the right-hand side of (20) depends on a particular control parameter. Then, at a given iteration number n, the left-hand side operator can be discretized once for all and then used to solve all the linear systems.

3.3.2 Functional gradients through adjoint equations

The gradient of the cost functional $\mathcal{J}$ with respect to the control variables c can also be obtained while combining the adjoint equation (14) and the expression (19) giving $\dfrac{d\mathcal{J}}{dc}$.

Indeed, the adjoint of (14) corresponds to

$$\xi^+ \frac{\partial F}{\partial \phi} = \frac{\partial \mathcal{J}}{\partial \phi}. \tag{21}$$

By introducing this equation into the expression (19) of the gradient of the cost functional with respect to the control variables, one obtains:

$$\frac{d\mathcal{J}(\phi, c)}{dc} = \xi^+ \frac{\partial F(\phi, c)}{\partial \phi} \frac{d\phi}{dc} + \frac{\partial \mathcal{J}(\phi, c)}{\partial c}.$$

Finally, using (20), it yields to:

$$\frac{d\mathcal{J}}{dc}(\phi^{(n)}, c^{(n)}) = -\left(\xi^+\right)^{(n)} \left.\frac{\partial F}{\partial c}\right|_{c^{(n)}} + \left.\frac{\partial \mathcal{J}}{\partial c}\right|_{c^{(n)}}. \tag{22}$$

The advantage of this method compared to that of the sensitivities is that it is necessary to solve only one linear system (the adjoint system 21) and that independently of the number of control parameters.

3.4 Differentiation and discretization

Two distinct approaches exist for formulating the adjoint system: continuous and discrete, or using the terminology of Gunzburger (2003) differentiate-then-discretize and discretize-then-differentiate. In the differentiate-then-discretize approach (see Fig. 10), the adjoint problem is derived analytically, based on the original system of partial differential equations, and then discretized using similar numerical methods to those used for discretizing the state equations. In the discretize-then-differentiate approach (see Fig. 11), the continuous direct problem is first discretized and these equations are then differentiated to obtain the discretized adjoint equations. For finite values of the grid sizes, the approximations of the discrete adjoints obtained by the continuous and discrete approaches are different, because the differentiation and discretization steps do not commute. Thus, we have to decide which approach is better for a specific problem.

The main advantage of the discretize-then-differentiate approach is that it yields to the exact gradient (except for round-off errors) of the discretized

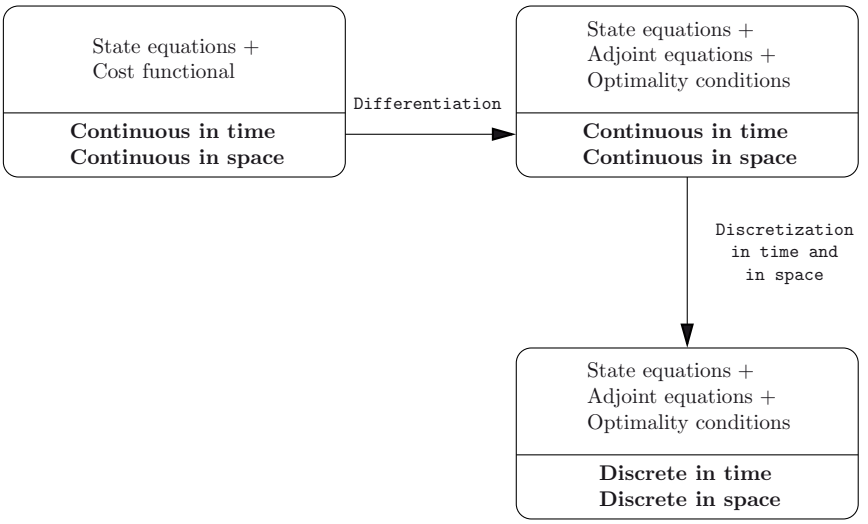


Figure 10. Differentiate-then-discretize approach for adjoint-based optimization methods.

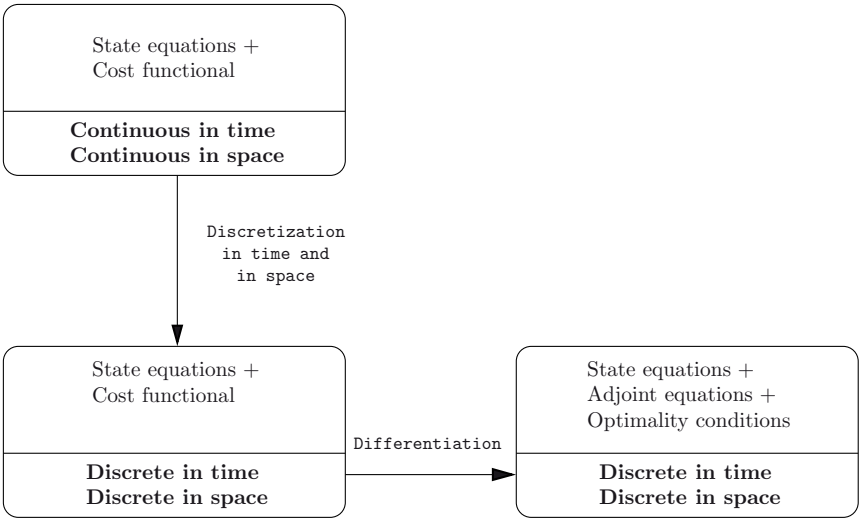


Figure 11. Discretize-then-differentiate approach for adjoint-based optimization methods.

functional. On the other hand, discrete adjoints obtained through the differentiate-then-discretize approach can yield to inconsistent gradients of cost functionals. Indeed, the approximate discrete adjoints are not the exact derivatives of the continuous functional, nor of a discretized functional. If the consistency of functional gradients is the only criterion that is considered then the advantage goes clearly to the discrete approach. However, differentiating by hand the discretized equations may become rapidly a formidable task, all the more if the state equations are strongly non-linear and if the discretization schemes are complex. Fortunately, to simplify the writing of adjoint codes, automatic differentiation software (see the web site `http://www.autodiff.org`) can be used to generate discrete adjoints. However, as one can imagine, adjoint codes created by automatic differentiation tools such as ADIFOR, TAMC, FastOpt, Tapenade, ... usually require more storage and CPU time that those written with the continuous approach. One advantage of using the differentiate-then-discretize approach is the possibility that is offered to design numerical grids that are specifically well suited to the adjoint systems. Indeed, if automatic differentiation tool is used to evaluate the approximate adjoint variables, the same grids are used for the state and adjoint equations what is clearly not optimal for numerical convergence. In the end, the continuous approach seems more natural for deriving the adjoint equations. For this reason, this is the approach that will be used in all the applications described in this chapter.

4 Linear quadratic optimal control

The Linear Quadratic Regulator (LQR) is an optimal control problem where the state equation is linear, the cost functional is quadratic, and the objective of the controller is to regulate, *i.e.* return to zero, some measure of the reference output $\boldsymbol{z}$ without using excessive amounts of control. The cost is evaluated subject to the initial condition $\boldsymbol{x}(0) = \boldsymbol{x_0}$ and with the assumption of no disturbance input $\boldsymbol{w}$. Therefore, we can use for plant model the LTI system given by (1). We must now specify the measure which is used for minimizing the reference output.

4.1 Choice of the cost functional

Let the time horizon for the optimal control be T, a natural choice of the cost functional is

$$\mathcal{J} = \int_0^T \|\boldsymbol{z}\|_2^2 \, \mathrm{d}t.$$

To evaluate $\|\boldsymbol{z}\|_2^2$, we start from the definition of the L^2 norm and consider that the variable $\boldsymbol{z}$ is given by

$$\boldsymbol{z}(t) = C_1\,\boldsymbol{x}(t) + D_{12}\,\boldsymbol{u}(t).$$

We then obtain immediately that:

$$\begin{aligned}\|\boldsymbol{z}\|_2^2 &= \boldsymbol{z}^H\boldsymbol{z} = (C_1\,\boldsymbol{x} + D_{12}\,\boldsymbol{u})^H(C_1\,\boldsymbol{x} + D_{12}\,\boldsymbol{u}) \\ &= (\boldsymbol{x}^H C_1^H + \boldsymbol{u}^H D_{12}^H)(C_1\,\boldsymbol{x} + D_{12}\boldsymbol{u}) \\ &= \boldsymbol{x}^H C_1^H C_1\,\boldsymbol{x} + \boldsymbol{u}^H D_{12}^H D_{12}\boldsymbol{u} + \underbrace{\boldsymbol{u}^H D_{12}^H C_1\,\boldsymbol{x} + \boldsymbol{x}^H C_1^H D_{12}\boldsymbol{u}}_{=2\boldsymbol{x}^H C_1^H D_{12}\boldsymbol{u}}.\end{aligned} \tag{23}$$

The next step is to simplify this expression. A traditional assumption in control theory (Burl, 1999; Zhou et al., 1996) is to postulate that $C_1^H D_{12} = 0$. For that, a natural choice is to consider that

$$C_1 = \begin{pmatrix} Q_x^{1/2} \\ 0 \end{pmatrix} \quad \text{and} \quad D_{12} = \begin{pmatrix} 0 \\ Q_u^{1/2} \end{pmatrix}$$

where $Q_x^{1/2}$ and $Q_u^{1/2}$ are respectively the square root of two positive-definite matrices Q_x and Q_u, to be justified below. Introducing C_1 and D_{12} in the expression of the norm of the reference output $\boldsymbol{z}$, we obtain for the cost functional:

$$\mathcal{J} = \int_0^T \|\boldsymbol{z}\|_2^2\;\mathrm{d}t = \int_0^T (\boldsymbol{x}^H Q_x\,\boldsymbol{x} + \boldsymbol{u}^H Q_u\,\boldsymbol{u})\;\mathrm{d}t.$$

Furthermore, since Q_x and Q_u are positive-definite matrices, the following weighted inner products

$$\|\boldsymbol{x}\|_{Q_x}^2 \triangleq \boldsymbol{x}^H\,Q_x\,\boldsymbol{x} \quad \text{and} \quad \|\boldsymbol{u}\|_{Q_u}^2 \triangleq \boldsymbol{u}^H\,Q_u\,\boldsymbol{u}$$

can be defined, and thus we can rewrite $\mathcal{J}$ as

$$\mathcal{J} = \int_0^T (\|\boldsymbol{x}\|_{Q_x}^2 + \|\boldsymbol{u}\|_{Q_u}^2)\;\mathrm{d}t.$$

A typical choice for the weight matrix Q_u is $Q_u = \ell^2\,\mathrm{Id}$ where ℓ is a real positive number ($\ell \neq 0$) and Id is the identity matrix. Finally, the cost functional may be expressed[25] as

$$\mathcal{J} = \int_0^T \left(\|\boldsymbol{x}\|_{Q_x}^2 + \ell^2\,\boldsymbol{u}^H\,\boldsymbol{u}\right)\;\mathrm{d}t. \tag{24}$$

[25] In this form, the link with the optimal perturbation problem discussed in section 5 becomes obvious. Indeed, when the penalization term ℓ is equal to zero, the expression under the integral is equivalent to the amplification rate of energy G considered in section 5.

The term involving $\|\boldsymbol{x}\|^2_{Q_x}$ is a measure of the energy of the state variable $\boldsymbol{x}$ integrated over the time horizon of optimization. The term involving $\boldsymbol{u}^H\,\boldsymbol{u}$ corresponds to the energy of the control signal. The role of the scalar ℓ^2 is to establish a trade off between two conflicting goals: *i)* decreasing the energy of the state variable which may require a large control signal, and *ii)* a small magnitude of the control which may lead to large amplitude of the state variable. The penalization term ℓ^2 must thus be viewed as a measure of the control cost. When the value of ℓ^2 is large, it means that the practical implementation of the control is "expensive". At the opposite, when the value of ℓ^2 is small, the control is considered as "cheap". Consequently, the minimal value of the cost functional $\mathcal{J}$ is expected when ℓ tends to zero. A first choice for the matrix Q_x and the scalar ℓ^2 is given by the Bryson's rule (Bryson Jr. and Ho, 1975).

4.2 Original problem and Lagrange multipliers

The objective of this section is to formulate the LQR problem in the general framework of optimal control theory (see section 3). For that, in addition to the state variables $\boldsymbol{x}$ and control variables $\boldsymbol{u}$, it is necessary to introduce successively the state equation and the cost functional.

As it was already mentioned at the beginning of section 4, the state equation is written here in the form of an LTI system, *i.e.*

$$\boldsymbol{F}(\boldsymbol{x},\boldsymbol{u}) = \dot{\boldsymbol{x}} - A\,\boldsymbol{x} - B_2\boldsymbol{u} = \boldsymbol{0}.$$

The development of the expression of the cost functional was the subject of section 4.1. On the basis of (24), the objective functional can also be written as

$$\mathcal{J} = \frac{1}{2}\left[\langle C_1\boldsymbol{x}, C_1\boldsymbol{x}\rangle + \ell^2\,\langle \boldsymbol{u},\boldsymbol{u}\rangle\right]$$

where the inner product $\langle\cdot,\cdot\rangle$ is defined as

$$\langle \boldsymbol{a},\boldsymbol{b}\rangle = \int_0^T \boldsymbol{a}^H(t)\boldsymbol{b}(t)\,\mathrm{d}t + \text{complex conjugate.} \tag{25}$$

In the case of the LQR problem, the three inner products $\langle\cdot,\cdot\rangle_s$, $\langle\cdot,\cdot\rangle_c$ and $\langle\cdot,\cdot\rangle_a$ introduced in section 3.2.1 are identical and correspond to (25) since all the direct, adjoint and control variables are only functions of time. This way, the derivation of the optimality condition in section 4.3.3 will be simplified.

The original constrained optimization problem is:

Determine the solution $\boldsymbol{x}(t)$ and the control parameter $\boldsymbol{u}(t)$ such that the cost functional $\mathcal{J}$ reaches a minimum subject to $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{u})=\boldsymbol{0}$.

Following the philosophy of the optimal control theory, we then introduce the Lagrange multiplier or adjoint variable $\boldsymbol{x}^+(t)$ to enforce the constraint $\boldsymbol{F}=\boldsymbol{0}$, and define the Lagrangian functional as

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{x}^+) \triangleq \mathcal{J}(\boldsymbol{x},\boldsymbol{u}) - \left\langle \boldsymbol{F}(\boldsymbol{x},\boldsymbol{u}),\boldsymbol{x}^+\right\rangle .$$

We then consider the unconstrained optimization problem given by:

Determine the solution $\boldsymbol{x}(t)$, the control parameter $\boldsymbol{u}(t)$ and the Lagrange multipliers $\boldsymbol{x}^+(t)$ such that the Lagrangian functional $\mathcal{L}$ reaches a minimum.

Since each argument of $\mathcal{L}$ is supposed to be independent of the others, the first-order necessary conditions for the minimum of $\mathcal{L}$ yield to an optimality system derived in the general case in section 3.2.2.

4.3 Derivation of the optimality system

For deriving the optimality system, we now have to set successively the first variation of $\mathcal{L}$ with respect to $\boldsymbol{x}^+$, $\boldsymbol{x}$ and $\boldsymbol{u}$ equal to zero.

4.3.1 Direct problem

Setting the first variation of $\mathcal{L}$ with respect to the Lagrange multiplier $\boldsymbol{x}^+$ equal to zero is equivalent to the condition:

$$\begin{aligned}
\left\langle \boldsymbol{\nabla}_{\boldsymbol{x}^+}\mathcal{L},\delta\boldsymbol{x}^+\right\rangle &= \lim_{\epsilon\to 0}\frac{\mathcal{L}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{x}^+ +\epsilon\delta\boldsymbol{x}^+)-\mathcal{L}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{x}^+)}{\epsilon} = 0\\
&= \lim_{\epsilon\to 0}\frac{\mathcal{J}(\boldsymbol{x},\boldsymbol{u})-\mathcal{J}(\boldsymbol{x},\boldsymbol{u})}{\epsilon}\\
&- \lim_{\epsilon\to 0}\frac{\left\langle \dot{\boldsymbol{x}}-A\,\boldsymbol{x}-B_2\boldsymbol{u},\boldsymbol{x}^+ +\epsilon\delta\boldsymbol{x}^+\right\rangle - \left\langle \dot{\boldsymbol{x}}-A\,\boldsymbol{x}-B_2\boldsymbol{u},\boldsymbol{x}^+\right\rangle}{\epsilon}\\
&= -\left\langle \dot{\boldsymbol{x}}-A\,\boldsymbol{x}-B_2\boldsymbol{u},\delta\boldsymbol{x}^+\right\rangle = 0.
\end{aligned}$$

Since the variation $\delta\boldsymbol{x}^+$ is arbitrary, we recover the state equation:

$$\dot{\boldsymbol{x}} = A\,\boldsymbol{x}+B_2\boldsymbol{u}.$$

4.3.2 Adjoint problem

Setting the first variation of $\mathcal{L}$ with respect to the state $\boldsymbol{x}$ equal to zero is equivalent to the condition

$$\langle \boldsymbol{\nabla}_{\boldsymbol{x}} \mathcal{L}, \delta \boldsymbol{x} \rangle = \lim_{\epsilon \to 0} \frac{\mathcal{L}(\boldsymbol{x} + \epsilon \delta \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{x}^+) - \mathcal{L}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{x}^+)}{\epsilon} = 0,$$

where the variation $\delta \boldsymbol{x}$ is arbitrary. Substituting $\mathcal{L}$ with its definition, we have

$$\begin{aligned}
\langle \boldsymbol{\nabla}_{\boldsymbol{x}} \mathcal{L}, \delta \boldsymbol{x} \rangle &= \lim_{\epsilon \to 0} \frac{\mathcal{J}(\boldsymbol{x} + \epsilon \delta \boldsymbol{x}, \boldsymbol{u}) - \mathcal{J}(\boldsymbol{x}, \boldsymbol{u})}{\epsilon} \\
&\quad - \lim_{\epsilon \to 0} \frac{\left\langle (\boldsymbol{x} + \epsilon \dot{\delta \boldsymbol{x}}) - A\,(\boldsymbol{x} + \epsilon \delta \boldsymbol{x}) - B_2 \boldsymbol{u}, \boldsymbol{x}^+ \right\rangle - \left\langle \dot{\boldsymbol{x}} - A\,\boldsymbol{x} - B_2 \boldsymbol{u}, \boldsymbol{x}^+ \right\rangle}{\epsilon} \\
&= \underbrace{\frac{\partial \mathcal{J}}{\partial \boldsymbol{x}} \delta \boldsymbol{x}}_{T_I} - \underbrace{\left\langle (\dot{\delta \boldsymbol{x}}), \boldsymbol{x}^+ \right\rangle}_{T_{II}} + \underbrace{\left\langle A \delta \boldsymbol{x}, \boldsymbol{x}^+ \right\rangle}_{T_{III}} = 0. \qquad (26)
\end{aligned}$$

The objective is now to write the terms T_I to T_{III}, appearing in the right-hand side, as a particular inner product utilizing $\delta \boldsymbol{x}$. For T_I, we have by definition

$$T_I = \lim_{\epsilon \to 0} \frac{\mathcal{J}(\boldsymbol{x} + \epsilon \delta \boldsymbol{x}, \boldsymbol{u}) - \mathcal{J}(\boldsymbol{x}, \boldsymbol{u})}{\epsilon} = \frac{1}{2} \left[\langle C_1 \delta \boldsymbol{x}, C_1 \boldsymbol{x} \rangle + \langle C_1 \boldsymbol{x}, C_1 \delta \boldsymbol{x} \rangle \right].$$

Using the symmetry of the inner product (25) and introducing the adjoint C_1^+ of C_1 with respect to (25), T_I becomes

$$T_I = \langle C_1 \delta \boldsymbol{x}, C_1 \boldsymbol{x} \rangle = \left\langle \delta \boldsymbol{x}, C_1^+ C_1 \boldsymbol{x} \right\rangle.$$

For T_{II}, we first use the definition of the inner product and then integrate by part in time. We obtain[26]:

$$\begin{aligned}
T_{II} = \left\langle (\dot{\delta \boldsymbol{x}}), \boldsymbol{x}^+ \right\rangle &= \int_0^T \left(\dot{\delta \boldsymbol{x}} \right)^H \boldsymbol{x}^+ \, \mathrm{d}t + \text{c.c.} \\
&= \left[\delta \boldsymbol{x}^H \boldsymbol{x}^+ \right]_0^T - \int_0^T \delta \boldsymbol{x}^H \dot{\boldsymbol{x}^+} \, \mathrm{d}t + \text{c.c.}
\end{aligned}$$

Since the initial condition $\boldsymbol{x}(0)$ is perfectly known, we have $\delta \boldsymbol{x}(0) = \boldsymbol{0}$. To simplify further T_{II}, we then consider that $\boldsymbol{x}^+(T) = \boldsymbol{0}$. Finally, T_{II} is

[26] The symbol c.c. denotes the complex conjugate.

reduced[27] to

$$T_{II} = \left\langle \delta \boldsymbol{x}, -\dot{\boldsymbol{x}^+} \right\rangle .$$

Lastly, to transform T_{III} we introduce the adjoint matrix A^+ of A with respect to the inner product (25). We thus obtain:

$$T_{III} = \left\langle A\delta \boldsymbol{x}, \boldsymbol{x}^+ \right\rangle = \left\langle \delta \boldsymbol{x}, A^+ \boldsymbol{x}^+ \right\rangle .$$

By gathering the terms T_I to T_{III}, we can then write (26) as

$$\begin{aligned} \langle \boldsymbol{\nabla}_{\boldsymbol{x}} \mathcal{L}, \delta \boldsymbol{x} \rangle &= \left\langle \delta \boldsymbol{x}, C_1^+ C_1 \boldsymbol{x} \right\rangle + \left\langle \delta \boldsymbol{x}, \dot{\boldsymbol{x}^+} \right\rangle + \left\langle \delta \boldsymbol{x}, A^+ \boldsymbol{x}^+ \right\rangle \\ &= \left\langle \delta \boldsymbol{x}, C_1^+ C_1 \boldsymbol{x} + \dot{\boldsymbol{x}^+} + A^+ \boldsymbol{x}^+ \right\rangle = 0. \end{aligned}$$

Since the variation $\delta \boldsymbol{x}$ in the state $\boldsymbol{x}$ is arbitrary, we obtain the adjoint equation

$$-\dot{\boldsymbol{x}^+} = A^+ \boldsymbol{x}^+ + C_1^+ C_1 \boldsymbol{x}.$$

With this definition, the adjoint state must be marched backward in time over the optimization horizon, starting the time integration with the terminal condition

$$\boldsymbol{x}^+(T) = \boldsymbol{0}.$$

4.3.3 Optimality conditions

Setting the first variation of $\mathcal{L}$ with respect to the control $\boldsymbol{u}$ equal to zero is equivalent to the condition

$$\langle \boldsymbol{\nabla}_{\boldsymbol{u}} \mathcal{L}, \delta \boldsymbol{u} \rangle = \lim_{\epsilon \to 0} \frac{\mathcal{L}(\boldsymbol{x}, \boldsymbol{u} + \epsilon \delta \boldsymbol{u}, \boldsymbol{x}^+) - \mathcal{L}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{x}^+)}{\epsilon} = 0,$$

where the variation $\delta \boldsymbol{u}$ of the control $\boldsymbol{u}$ is arbitrary. If we now substitute the Lagrangian $\mathcal{L}$ with its definition, we directly obtain:

$$\begin{aligned} \langle \boldsymbol{\nabla}_{\boldsymbol{u}} \mathcal{L}, \delta \boldsymbol{u} \rangle &= \lim_{\epsilon \to 0} \frac{\mathcal{J}(\boldsymbol{x}, \boldsymbol{u} + \epsilon \delta \boldsymbol{u}) - \mathcal{J}(\boldsymbol{x}, \boldsymbol{u})}{\epsilon} \\ &\quad - \lim_{\epsilon \to 0} \frac{\left\langle \dot{\boldsymbol{x}} - A\,\boldsymbol{x} - B_2(\boldsymbol{u} + \epsilon \delta \boldsymbol{u}), \boldsymbol{x}^+ \right\rangle - \left\langle \dot{\boldsymbol{x}} - A\,\boldsymbol{x} - B_2 \boldsymbol{u}, \boldsymbol{x}^+ \right\rangle}{\epsilon} \\ &= \underbrace{\frac{\partial \mathcal{J}}{\partial \boldsymbol{u}} \delta \boldsymbol{u}}_{T_I} + \underbrace{\left\langle B_2 \delta \boldsymbol{u}, \boldsymbol{x}^+ \right\rangle}_{T_{II}} = 0. \end{aligned} \tag{27}$$

[27] The minus sign, which appeared in front of $\dot{\boldsymbol{x}^+}$ following the temporal integration by part, makes that the adjoint equation is now defined backward in time.

Proceeding similarly as we did for the adjoint equation in section 4.3.2, we immediately find:

$$T_I = \lim_{\epsilon \to 0} \frac{\mathcal{J}(\boldsymbol{x}, \boldsymbol{u} + \epsilon \delta \boldsymbol{u}) - \mathcal{J}(\boldsymbol{x}, \boldsymbol{u})}{\epsilon} = \frac{\ell^2}{2} \left[\langle \boldsymbol{u}, \delta \boldsymbol{u} \rangle + \langle \delta \boldsymbol{u}, \boldsymbol{u} \rangle \right] = \ell^2 \langle \delta \boldsymbol{u}, \boldsymbol{u} \rangle$$

and

$$T_{II} = \left\langle B_2 \delta \boldsymbol{u}, \boldsymbol{x}^+ \right\rangle = \left\langle \delta \boldsymbol{u}, B_2^+ \boldsymbol{x}^+ \right\rangle$$

where B_2^+ is the adjoint matrix of B_2 with respect to the inner product (25).

If we gather the two terms, (27) simplifies to

$$\langle \boldsymbol{\nabla}_{\boldsymbol{u}} \mathcal{L}, \delta \boldsymbol{u} \rangle = \ell^2 \langle \delta \boldsymbol{u}, \boldsymbol{u} \rangle + \left\langle \delta \boldsymbol{u}, B_2^+ \boldsymbol{x}^+ \right\rangle = 0 \qquad \forall \delta \boldsymbol{u}.$$

Finally, since the variation $\delta \boldsymbol{u}$ of $\boldsymbol{u}$ is arbitrary, we obtain the optimality condition

$$B_2^+ \boldsymbol{x}^+ = -\ell^2 \boldsymbol{u},$$

and the gradient of $\mathcal{J}$ with respect to $\boldsymbol{u}$

$$\boldsymbol{\nabla}_{\boldsymbol{u}} \mathcal{L} = B_2^+ \boldsymbol{x}^+ + \ell^2 \boldsymbol{u}.$$

Collecting the results of setting the first variations of the Lagrangian functional to zero yields the optimality system given in Fig. 12. The aim of the next section is to solve this optimality system by utilizing a nonlinear matrix differential equation, known as the Riccati equation.

4.4 Riccati equation

In order to solve the optimality system of the LQR problem (see Fig. 12), we must eliminate two variables among the three unknowns $\boldsymbol{x}$, $\boldsymbol{u}$ and $\boldsymbol{x}^+$. Solving for the optimal control $\boldsymbol{u}$ in the optimality condition yields

$$\boldsymbol{u}(t) = -\frac{1}{\ell^2} B_2^+ \boldsymbol{x}^+(t) \tag{28}$$

where the inverse of ℓ is guaranteed to exist since ℓ is a positive real number (see section 4.1). Eliminating $\boldsymbol{u}$ from the direct and adjoint equations, and combining the resulting equations into a single state equation yields

$$\begin{pmatrix} \dot{\boldsymbol{x}} \\ \dot{\boldsymbol{x}}^+ \end{pmatrix} = \begin{pmatrix} A & -\frac{1}{\ell^2} B_2 B_2^+ \\ -C_1^+ C_1 & -A^+ \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{x}^+ \end{pmatrix} \tag{29}$$

with

$$\boldsymbol{x}(0) = \boldsymbol{x_0} \quad \text{and} \quad \boldsymbol{x}^+(T) = \boldsymbol{0}. \tag{30}$$

State equation:

$$\begin{aligned} \dot{\boldsymbol{x}} &= A\,\boldsymbol{x} + B_2\boldsymbol{u} \\ \boldsymbol{x}(0) &= \boldsymbol{x_0} \qquad \text{(I.C.)} \end{aligned}$$

Adjoint equation:

$$\begin{aligned} -\dot{\boldsymbol{x}^+} &= A^+\boldsymbol{x}^+ + C_1^+ C_1 \boldsymbol{x} \\ \boldsymbol{x}^+(T) &= \boldsymbol{0} \qquad \text{(T.C.)} \end{aligned}$$

Optimality condition:

$$B_2^+ \boldsymbol{x}^+ = -\ell^2 \boldsymbol{u}$$

Cost functional:

$$\mathcal{J} = \frac{1}{2}\int_0^T \left(\boldsymbol{x}^H\, C_1^H C_1\, \boldsymbol{x} + \ell^2\, \boldsymbol{u}^H\, \boldsymbol{u}\right)\, \mathrm{d}t$$

Figure 12. Optimality system for the linear quadratic regulator problem. I.C.: initial condition, T.C.: terminal condition.

The Hamiltonian system (29) represents a set of necessary and sufficient conditions for the control to minimize the cost functional $\mathcal{J}$. This results from the fact that $\mathcal{J}$ is quadratic and from the positivity requirements on the weighting matrix $Q_x = C_1^H C_1$.

By integrating in time the Hamiltonian system (29) subject to the initial and final conditions (30), it can be shown (Burl, 1999, p. 186) that the adjoint variable can be found from the state variable using the linear relationship

$$\boldsymbol{x}^+(t) = \Pi(t)\boldsymbol{x}(t) \tag{31}$$

where $\Pi(t)$ is a square matrix of size n_x. The optimal state-feedback control is found from (28):

$$\boldsymbol{u}(t) = -\frac{1}{\ell^2} B_2^+ \Pi(t)\boldsymbol{x}(t) = K(t)\boldsymbol{x}(t)$$

where $K(t)$ is called the feedback gain matrix. To evaluate the gain matrix K, it is thus necessary to determine the matrix Π. A differential equation for $\Pi(t)$ can then be obtained by first taking the time derivative of (31) and then substituting $\dot{\boldsymbol{x}}$ and $\dot{\boldsymbol{x}^+}$ from (29). After rearrangement, we obtain the Riccati[28] equation:

$$-\dot{\Pi} = A^+ \Pi + \Pi A - \frac{1}{\ell^2} \Pi B_2 B_2^+ \Pi + C_1^+ C_1. \tag{32}$$

This equation depends only on the Hamiltonian matrix introduced in (29). The matrix $\Pi(t)$ is found by solving (32) backward in time from the terminal condition given by:

$$\Pi(T) = 0.$$

In applications where the control is designed to operate for $T \longrightarrow +\infty$, it is reasonable to ignore the transient time of the optimal gains and use steady-state gains instead. The steady-state solution of the Riccati equation is then generated from (31) by setting the derivative to zero. We then obtain the continuous time algebraic Riccati equation (CARE) given by[29]:

$$A^+ \Pi + \Pi A - \frac{1}{\ell^2} \Pi B_2 B_2^+ \Pi + C_1^+ C_1 = 0. \tag{33}$$

The solution of (33) can be used to generate the cost associated with the optimal control. Given $\boldsymbol{x}(0)$, the optimal cost $\mathcal{J}_{\min}$ is given (Burl, 1999) by:

$$\mathcal{J}_{\min} = \boldsymbol{x}^H(0)\Pi(0)\boldsymbol{x}(0).$$

To conclude this section, we give in Fig. 13 a summary of the solutions for the LQR problem considered in the general case where the weighting matrix Q_u is not necessarily equal to $\ell^2\, Id$.

[28] The name Riccati is given to the equation by analogy to the Riccati differential equation: the unknown appears linearly and in a quadratic term (but no higher-order terms).

[29] When the system order is not too high (about 300), (33) can be solved directly with the `care` function under Matlab (Control System Toolbox) or alternatively with the Slicot library. However, numerical algorithms for the solution of large-scale algebraic Riccati equations are still nowadays a topic of research (Benner et al., 2008, for instance). It is then evident that if we want to have a chance to apply sophisticated control algorithms in real systems, strategies of model reduction, such as those discussed in section 2.3, are fundamental.

State-space system:

$$\dot{\boldsymbol{x}}(t) = A\,\boldsymbol{x}(t) + B_2\boldsymbol{u}(t), \qquad \text{with} \quad \boldsymbol{x}(0) = \boldsymbol{x_0}.$$
$$\boldsymbol{z}(t) = C_1\,\boldsymbol{x}(t) + D_{12}\boldsymbol{u}(t), \quad \text{with} \quad C_1 = \begin{pmatrix} Q_x^{1/2} \\ 0 \end{pmatrix} \text{and } D_{12} = \begin{pmatrix} 0 \\ Q_u^{1/2} \end{pmatrix}.$$

Cost functional:

$$\mathcal{J} = \int_0^T (\|\boldsymbol{x}\|^2_{Q_x} + \|\boldsymbol{u}\|^2_{Q_u})\ \mathrm{d}t$$

State feedback:

$$\boldsymbol{u}(t) = K\boldsymbol{x}(t)$$

Optimal feedback gain:

$$K(t) = -Q_u^{-1}B_2^{+}\Pi(t)$$

Riccati equation:

$$-\dot{\Pi}(t) = A^{+}\,\Pi(t) + \Pi(t)\,A - \Pi(t)\,B_2\,Q_u^{-1}\,B_2^{+}\,\Pi(t) + Q_x, \quad \text{with} \quad \Pi(T) = 0.$$

Figure 13. Solutions of the Linear Quadratic Regulator (LQR) problem.

5 Optimal growth perturbation

In stability theory (Schmid and Henningson, 2001), people are interested by the determination of the initial condition which maximizes the energy amplification of the disturbances on a given time horizon $0 \leq t \leq T$. Since stability analysis is concerned with the disturbances around a base flow, we will consider the linearized Navier-Stokes equations. Using compact notations, those equations can be written (Bewley and Liu, 1998, for instance)

as[30]:

$$\boldsymbol{F}(\boldsymbol{q}) = \dot{\boldsymbol{q}} - A\,\boldsymbol{q} = \boldsymbol{0} \quad \text{with} \quad \boldsymbol{q}(\boldsymbol{x}, t=0) = \boldsymbol{q_0}(\boldsymbol{x}) \tag{34}$$

where the matrix A corresponds to the linearized Navier-Stokes operator and $\boldsymbol{x}$ belongs to the spatial domain Ω_x. By definition, the search of optimal disturbances is thus equivalent to an optimization problem. A measure of performance of the optimization can be given by the ratio of disturbance energy at time T to the initial energy *i.e.*

$$\mathcal{J}(\boldsymbol{q}, \boldsymbol{q_0}) = \frac{\|\boldsymbol{q}(\boldsymbol{x}, T)\|_E^2}{\|\boldsymbol{q_0}(\boldsymbol{x})\|_E^2}. \tag{35}$$

The energy inner product $(\cdot,\cdot)_E$ is defined as

$$\left(\boldsymbol{q^I}, \boldsymbol{q^{II}}\right)_E = \int_{\Omega_x} \left(\boldsymbol{q^I}\right)^H M \boldsymbol{q^{II}}\, \mathrm{d}\boldsymbol{x} + c.c. \tag{36}$$

where M is a matrix that is case dependent, see Guegan et al. (2006) or Antkowiak and Brancher (2007) for two typical examples.

The optimization process corresponding to the maximization of $\mathcal{J}$ should respect the constraints given by the linearized Navier-Stokes equations and the specified boundary and initial conditions. This problem is then amenable to the classical framework of constrained optimization problem as discussed in the next section.

5.1 Variational formulation

The objective of this section is to formulate the problem of optimal energy growth in the framework of optimal control theory (see section 3). As we will see thereafter, this problem has many points in common with the LQR control considered in section 4. Indeed, the constraints (34) are also linear and the cost functional (35) is also quadratic. The only differences are that there are now two constraint equations instead of one and that the control corresponds to the initial condition of the linearized system instead of a boundary condition.

[30] In this section, we use the traditional notations in stability theory, $\boldsymbol{q}$ for the perturbation, $\boldsymbol{x}$ for the spatial variable and t for time.

5.1.1 Original problem, inner products and Lagrangian formulation

To enforce the initial condition, we relate the solution $\boldsymbol{q}$ at initial time and $\boldsymbol{q_0}$, the optimal growth perturbation we are looking for, through the relation

$$\boldsymbol{H}(\boldsymbol{q}, \boldsymbol{q_0}) = \boldsymbol{q}(\boldsymbol{x}, 0) - \boldsymbol{q_0}(\boldsymbol{x}) = \boldsymbol{0}. \tag{37}$$

The original constrained optimization problem is then formulated as:

> *Determine the solution* $\boldsymbol{q}(\boldsymbol{x}, t)$ *and the control parameter* $\boldsymbol{q_0}(t)$ *(optimal disturbance) such that the cost functional* $\mathcal{J}$ *reaches a maximum subject to* $\boldsymbol{F}(\boldsymbol{q}) = \boldsymbol{0}$ *and* $\boldsymbol{H}(\boldsymbol{q}, \boldsymbol{q_0}) = \boldsymbol{0}$.

As outlined in section 3.2, the optimal control procedure requires the introduction of Lagrange multipliers or adjoint variables $\widetilde{\boldsymbol{q}}(\boldsymbol{x}, t)$ and $\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}(\boldsymbol{x})$ to enforce respectively the constraints $\boldsymbol{F} = \boldsymbol{0}$ and $\boldsymbol{H} = \boldsymbol{0}$. Furthermore, to lead properly the developments of the optimality system, we need to introduce two additional inner products:

$$\left\langle \widetilde{\boldsymbol{q}}^{\boldsymbol{I}}, \widetilde{\boldsymbol{q}}^{\boldsymbol{II}} \right\rangle = \int_0^T \int_{\Omega_x} \left(\widetilde{\boldsymbol{q}}^{\boldsymbol{I}}\right)^H \widetilde{\boldsymbol{q}}^{\boldsymbol{II}} \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}t + c.c. \tag{38}$$

and

$$\left(\widetilde{\boldsymbol{q}}^{\boldsymbol{I}}, \widetilde{\boldsymbol{q}}^{\boldsymbol{II}}\right) = \int_{\Omega_x} \left(\widetilde{\boldsymbol{q}}^{\boldsymbol{I}}\right)^H \widetilde{\boldsymbol{q}}^{\boldsymbol{II}} \, \mathrm{d}\boldsymbol{x} + c.c. \tag{39}$$

Following the procedure presented in section 3.2.1, a single vector space $\Theta = \boldsymbol{q} \times \boldsymbol{q_0} \times \widetilde{\boldsymbol{q}} \times \widetilde{\boldsymbol{q}}_{\boldsymbol{0}}$ including all the direct and adjoint variables is introduced. Let $\boldsymbol{Q^i} = \left(\boldsymbol{q^i}, \boldsymbol{q_0^i}, \widetilde{\boldsymbol{q}}^{\boldsymbol{i}}, \widetilde{\boldsymbol{q}}_{\boldsymbol{0}}^{\boldsymbol{i}}\right)$ with $i = I, II$ be two arbitrary elements of Θ, an inner product including all the three inner products (36), (38) and (39) is defined as

$$\left\{\boldsymbol{Q}^{\boldsymbol{I}}, \boldsymbol{Q}^{\boldsymbol{II}}\right\} = \left\langle \boldsymbol{q}^{\boldsymbol{I}}, \boldsymbol{q}^{\boldsymbol{II}} \right\rangle + \left(\boldsymbol{q}_{\boldsymbol{0}}^{\boldsymbol{I}}, \boldsymbol{q}_{\boldsymbol{0}}^{\boldsymbol{II}}\right)_E + \left\langle \widetilde{\boldsymbol{q}}^{\boldsymbol{I}}, \widetilde{\boldsymbol{q}}^{\boldsymbol{II}} \right\rangle + \left(\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}^{\boldsymbol{I}}, \widetilde{\boldsymbol{q}}_{\boldsymbol{0}}^{\boldsymbol{II}}\right). \tag{40}$$

The constrained optimization problem is circumvented by introducing the Lagrangian functional

$$\mathcal{L}(\boldsymbol{q}, \boldsymbol{q_0}, \widetilde{\boldsymbol{q}}, \widetilde{\boldsymbol{q}}_{\boldsymbol{0}}) = \mathcal{J}(\boldsymbol{q}, \boldsymbol{q_0}) - \langle \boldsymbol{F}(\boldsymbol{q}), \widetilde{\boldsymbol{q}} \rangle - \left(\boldsymbol{H}(\boldsymbol{q}, \boldsymbol{q_0}), \widetilde{\boldsymbol{q}}_{\boldsymbol{0}}\right),$$

where the constraints have already been included by means of appropriate Lagrange multipliers.

5.1.2 Gradients of the Lagrangian

As outlined in section 3.2.2, determining the stationary points of the Lagrangian $\mathcal{L}$ requires the computation of the gradients of $\mathcal{L}$ with respect to all the variables. By convention, the gradients of $\mathcal{L}$ with respect to $\boldsymbol{q}$, $\boldsymbol{q_0}$, $\widetilde{\boldsymbol{q}}$ and $\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}$ are denoted in the following respectively by $\boldsymbol{\nabla}_{\boldsymbol{q}}\mathcal{L}$, $\boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{L}$, $\boldsymbol{\nabla}_{\widetilde{\boldsymbol{q}}}\mathcal{L}$ and $\boldsymbol{\nabla}_{\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}}\mathcal{L}$. These gradients are determined as projections of $\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{Q})$, gradient of the Lagrangian at point $\boldsymbol{Q}$, onto the different subspaces corresponding to each of the variables $\boldsymbol{q}$, $\boldsymbol{q_0}$, $\widetilde{\boldsymbol{q}}$ and $\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}$. Assuming that $\mathcal{L}$ is Fréchet-differentiable, $\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{Q})$ is such that for any variation $\delta\boldsymbol{Q}$ we have:

$$\{\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{Q}), \delta\boldsymbol{Q}\} = \lim_{\epsilon \longrightarrow 0} \frac{\mathcal{L}(\boldsymbol{Q} + \epsilon\delta\boldsymbol{Q}) - \mathcal{L}(\boldsymbol{Q})}{\epsilon}. \tag{41}$$

5.1.2.1 Determination of $\boldsymbol{\nabla}_{\widetilde{\boldsymbol{q}}}\mathcal{L}$ and $\boldsymbol{\nabla}_{\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}}\mathcal{L}$: direct equations Considering respectively $\delta\boldsymbol{Q} = (0, 0, \delta\widetilde{\boldsymbol{q}}, 0)$ and $\delta\boldsymbol{Q} = (0, 0, 0, \delta\widetilde{\boldsymbol{q}}_{\boldsymbol{0}})$ in (41), we obtain immediately (see section 4.3.1 for similar developments):

$$\left\langle \boldsymbol{\nabla}_{\widetilde{\boldsymbol{q}}}\mathcal{L}, \delta\widetilde{\boldsymbol{q}} \right\rangle = -\left\langle \boldsymbol{F}(\boldsymbol{q}), \delta\widetilde{\boldsymbol{q}} \right\rangle,$$

$$\left(\boldsymbol{\nabla}_{\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}}\mathcal{L}, \delta\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}\right) = -\left(\boldsymbol{H}(\boldsymbol{q}, \boldsymbol{q_0}), \delta\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}\right).$$

At the stationary points of $\mathcal{L}$, these gradients are by definition equal to zero. Since the variations $\delta\widetilde{\boldsymbol{q}}$ and $\delta\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}$ are arbitrary, the constraints $\boldsymbol{F}(\boldsymbol{q}) = \boldsymbol{0}$ and $\boldsymbol{H}(\boldsymbol{q}, \boldsymbol{q_0}) = \boldsymbol{0}$ are recovered.

5.1.2.2 Determination of $\boldsymbol{\nabla}_{\boldsymbol{q}}\mathcal{L}$: adjoint equations For evaluating $\boldsymbol{\nabla}_{\boldsymbol{q}}\mathcal{L}$, $\delta\boldsymbol{Q} = (\delta\boldsymbol{q}, 0, 0, 0)$ is introduced in (41). After some developments which are exactly similar to those presented in section (4.3.2) for the LQR control, we obtain

$$\begin{aligned}\langle \boldsymbol{\nabla}_{\boldsymbol{q}}\mathcal{L}, \delta\boldsymbol{q} \rangle = \frac{2}{\|\boldsymbol{q_0}\|_E^2} \left(\delta\boldsymbol{q}|_T\,, \boldsymbol{q}|_T\right)_E + \left\langle \delta\boldsymbol{q}, \frac{d\widetilde{\boldsymbol{q}}}{dt} + A^+\widetilde{\boldsymbol{q}} \right\rangle - \left(\delta\ \boldsymbol{q}|_T\,, \widetilde{\boldsymbol{q}}|_T\right) \\ - \left(\delta\boldsymbol{q}|_0\,, \widetilde{\boldsymbol{q}}_{\boldsymbol{0}}\right) + \left(\delta\boldsymbol{q}|_0\,, \widetilde{\boldsymbol{q}}|_0\right),\end{aligned} \tag{42}$$

where $\delta\boldsymbol{q}|_t$ stands for $\delta\boldsymbol{q}(\boldsymbol{x}, t)$ with $t = 0$ or T. The expression (42) has to hold true for any $\delta\boldsymbol{q}$ which entails:

1. Adjoint equations
$$\frac{d\widetilde{\boldsymbol{q}}}{dt} = -A^+\widetilde{\boldsymbol{q}}, \tag{43}$$
2. Terminal adjoint condition
$$\widetilde{\boldsymbol{q}}|_T = \frac{2}{\|\boldsymbol{q_0}\|_E^2}\ \boldsymbol{q}|_T\,, \tag{44}$$

3. Compatibility condition

$$\widetilde{\boldsymbol{q}}_{\boldsymbol{0}} = \widetilde{\boldsymbol{q}}|_0 \,. \tag{45}$$

5.1.2.3 Determination of $\boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{L}$: optimality conditions Finally, to determine $\boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{L}$, we consider $\delta\boldsymbol{Q} = (0, \delta\boldsymbol{q_0}, 0, 0)$ in (41). While proceeding similarly that in section 4.3.3, we obtain:

$$\left(\boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{L}, \delta\boldsymbol{q_0}\right)_E = -2\frac{\| \boldsymbol{q}|_T \|_E^2}{\left(\|\boldsymbol{q_0}\|_E^2\right)^2}\left(\boldsymbol{q_0}, \delta\boldsymbol{q_0}\right)_E + \left(\delta\boldsymbol{q_0}, \widetilde{\boldsymbol{q}}_{\boldsymbol{0}}\right). \tag{46}$$

According to definitions (36) and (39) of the inner products, the following expression holds:

$$\left(\delta\boldsymbol{q_0}, \widetilde{\boldsymbol{q}}_{\boldsymbol{0}}\right) = \left(M^{-1}\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}, \delta\boldsymbol{q_0}\right)_E$$

where M^{-1} is the inverse of the matrix M. Consequently, (46) is equivalent to

$$\left(\boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{L}, \delta\boldsymbol{q_0}\right)_E = -2\frac{\| \boldsymbol{q}|_T \|_E^2}{\left(\|\boldsymbol{q_0}\|_E^2\right)^2}\left(\boldsymbol{q_0}, \delta\boldsymbol{q_0}\right)_E + \left(M^{-1}\widetilde{\boldsymbol{q}}_{\boldsymbol{0}}, \delta\boldsymbol{q_0}\right)_E. \tag{47}$$

Since $\delta\boldsymbol{q_0}$ is arbitrary, the gradient of the Lagrangian with respect to the initial perturbation is

$$\boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{L} = -2\frac{\| \boldsymbol{q}|_T \|_E^2}{\left(\|\boldsymbol{q_0}\|_E^2\right)^2}\boldsymbol{q_0} + M^{-1}\widetilde{\boldsymbol{q}}_{\boldsymbol{0}},$$

and the optimality condition corresponds to

$$\boldsymbol{q_0} = \frac{\left(\|\boldsymbol{q_0}\|_E^2\right)^2}{2\| \boldsymbol{q}|_T \|_E^2} M^{-1}\widetilde{\boldsymbol{q}}_{\boldsymbol{0}} \tag{48}$$

where $\widetilde{\boldsymbol{q}}_{\boldsymbol{0}} = \widetilde{\boldsymbol{q}}(\boldsymbol{x}, t = 0)$ due to the compatibility condition (45). If we suppose that M is equal to the identity matrix then the optimality system given in Schmid (2007) on page 145 is recovered. Furthermore, in the constrained subspace where $\boldsymbol{F}(\boldsymbol{q}) = \boldsymbol{0}$ and $\boldsymbol{H}(\boldsymbol{q}, \boldsymbol{q_0}) = \boldsymbol{0}$, the gradient of the Lagrangian simply reduces to

$$\boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{L}(\boldsymbol{Q}) = \boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{J}(\boldsymbol{Q}). \tag{49}$$

The optimality system corresponding to the search of optimal growth perturbation at a fixed time T is given in Fig. 14. This optimality system can be solved iteratively using the procedure described in section 3.2.3 where $\boldsymbol{\nabla}_{\boldsymbol{q_0}}\mathcal{J}$ given by (49) is used for determining the descent direction (see section 7). By varying the time T over which the optimization is performed, the maximum growth curve $G(t)$ (see Fig. 15 for an example) is

State equation:

$$\begin{aligned} \dot{\boldsymbol{q}} &= A\,\boldsymbol{q} \\ \boldsymbol{q}(\boldsymbol{x}, t=0) &= \boldsymbol{q_0}(\boldsymbol{x}) \qquad \text{(I.C.) and control parameter} \end{aligned}$$

Adjoint equation:

$$\begin{aligned} \frac{d\widetilde{\boldsymbol{q}}}{dt} &= -A^{+}\widetilde{\boldsymbol{q}} \\ \widetilde{\boldsymbol{q}}|_T &= \frac{2}{\|\boldsymbol{q_0}\|_E^2}\ \boldsymbol{q}|_T \qquad \text{(T.C.)} \end{aligned}$$

Optimality condition:

$$\boldsymbol{q_0} = \frac{\left(\|\boldsymbol{q_0}\|_E^2\right)^2}{2\|\ \boldsymbol{q}|_T\ \|_E^2} M^{-1}\widetilde{\boldsymbol{q}}_{\mathbf{0}}$$

Compatibility condition:

$$\widetilde{\boldsymbol{q}}_{\mathbf{0}} = \widetilde{\boldsymbol{q}}|_0$$

Cost functional:

$$\mathcal{J}(\boldsymbol{q}, \boldsymbol{q_0}) = \frac{\|\boldsymbol{q}(\boldsymbol{x}, T)\|_E^2}{\|\boldsymbol{q_0}(\boldsymbol{x})\|_E^2}$$

Figure 14. Optimality system for the optimal growth perturbation problem. I.C.: initial condition, T.C.: terminal condition.

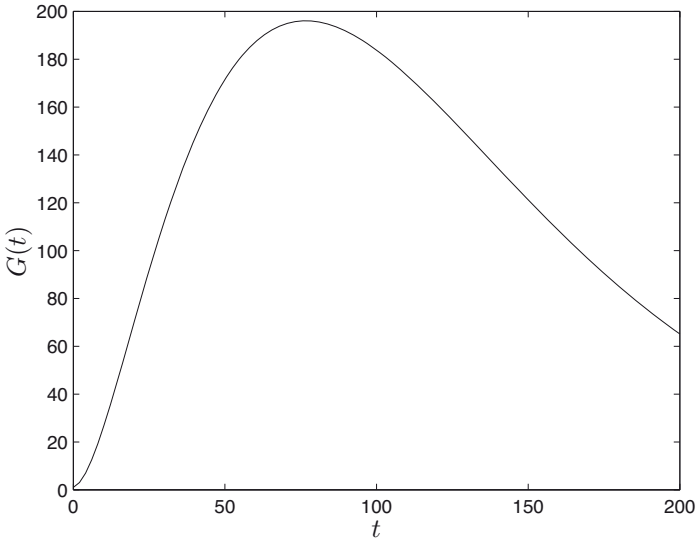


Figure 15. Maximum growth curve $G(t)$ for Poiseuille flow at $Re = 1000$. The streamwise and spanwise wavenumbers are $\alpha = 0$ and $\beta = 2$ (see Cordier, 2009, for the details). The number of Chebyshev points for the discretization is 200.

obtained. A second step consists in seeking the time $t_{\max}$ for which the curve $G(t)$ reached its maximum. The corresponding initial condition $\boldsymbol{q_0}$ is called global optimal perturbation, or, in short, optimal perturbation. An example of optimal perturbation for the Poiseuille flow is given in Fig. 16 and the corresponding solution $\boldsymbol{q}(\boldsymbol{x}, t_{\max})$ displayed in Fig. 17.

Another way of determining the optimal disturbances is based on the use of matrix exponential directly related to the linear system (34). This is the method which is the most used in the literature because, as it will be described in the following section, it is not necessary to use iterative methods to determine the curve of temporal amplification energy.

5.2 Formulation based on matrix exponential

In this section, the same notations as in Schmid and Henningson (2001) are used what means that the disturbance $\boldsymbol{q}$ is now discretized in space. We

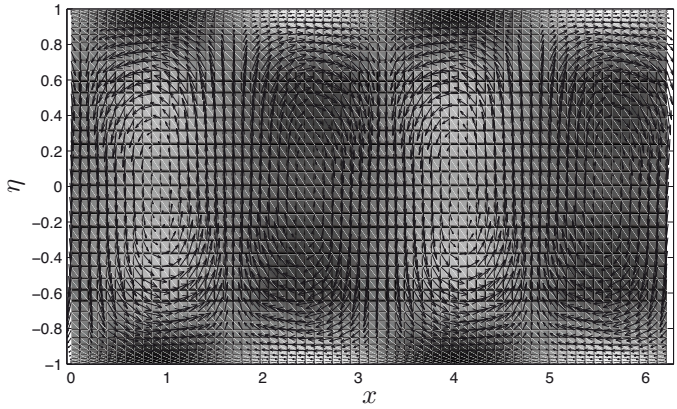


Figure 16. Optimal initial condition for Poiseuille flow at $Re = 1000$. The streamwise and spanwise wavenumbers are $\alpha = 0$ and $\beta = 2$ (see Cordier, 2009, for the details). The number of Chebyshev points for the discretization is 200. x and η are respectively the streamwise and wall-normal non-dimensional coordinates.

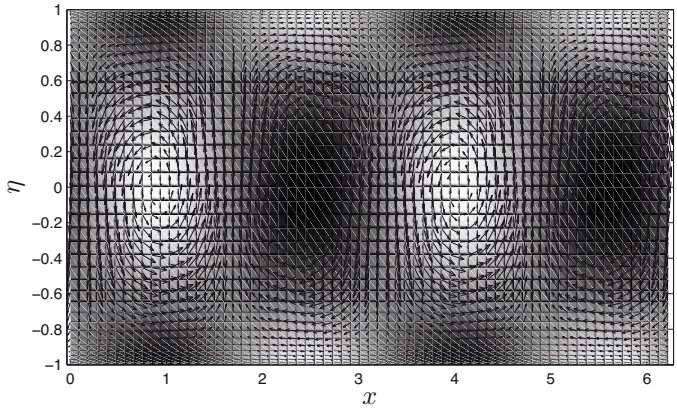


Figure 17. Optimal perturbation for Poiseuille flow at $t = t_{\max}$ and $Re = 1000$. The streamwise and spanwise wavenumbers are $\alpha = 0$ and $\beta = 2$ (see Cordier, 2009, for the details). The number of Chebyshev points for the discretization is 200. x and η are respectively the streamwise and wall-normal non-dimensional coordinates.

will thus have $\boldsymbol{q} \triangleq \boldsymbol{q}(t)$ and no more $\boldsymbol{q} = \boldsymbol{q}(\boldsymbol{x}, t)$ as in section 5.1. The objective is to determine the maximum possible amplification G of initial energy density *i.e.*

$$G(t) = \max_{\boldsymbol{q_0} \neq \boldsymbol{0}} \frac{\|\boldsymbol{q}(t)\|_E^2}{\|\boldsymbol{q_0}\|_E^2} \quad \text{with} \quad \boldsymbol{q_0} = \boldsymbol{q}(0) \tag{50}$$

where the energy inner product is now defined as $\|\boldsymbol{q}(t)\|_E^2 = \boldsymbol{q}(t)^H \, Q_q \, \boldsymbol{q}(t)$ with Q_q an appropriate weighting matrix. For practical reasons, it is more desirable to work with the standard-L_2 norm than with weighting matrices. Using the Cholesky decomposition, we then decompose the weight matrix Q_q as $Q_q = F^H F$. This way, the energy inner product can be written also as

$$\|\boldsymbol{q}(t)\|_E^2 = \boldsymbol{q}(t)^H \, Q_q \, \boldsymbol{q}(t) = \boldsymbol{q}(t)^H \, F^H F \, \boldsymbol{q}(t) = \|F \, \boldsymbol{q}(t)\|_2^2. \tag{51}$$

Furthermore, for a linear state-space model defined by

$$\dot{\boldsymbol{q}} = A \, \boldsymbol{q} \quad \text{with} \quad \boldsymbol{q}(0) = \boldsymbol{q_0}, \tag{52}$$

the solution at all positive times is given by (see section 2.1.2)

$$\boldsymbol{q}(t) = e^{At} \, \boldsymbol{q_0}. \tag{53}$$

Consequently, the maximum energy amplification G may be written also:

$$G(t) = \max_{\boldsymbol{q_0} \neq \boldsymbol{0}} \frac{\|\boldsymbol{q}(t)\|_E^2}{\|\boldsymbol{q_0}\|_E^2} = \max_{\boldsymbol{q_0} \neq \boldsymbol{0}} \frac{\|F \, \boldsymbol{q}(t)\|_2^2}{\|F \, \boldsymbol{q_0}\|_2^2} = \max_{\boldsymbol{q_0} \neq \boldsymbol{0}} \frac{\|F \, e^{At} \, \boldsymbol{q_0}\|_2^2}{\|F \, \boldsymbol{q_0}\|_2^2}.$$

Introducing the change of variable $\boldsymbol{s_0} = F \boldsymbol{q_0}$, the previous expression becomes:

$$G(t) = \max_{\boldsymbol{s_0} \neq \boldsymbol{0}} \frac{\|F \, e^{At} F^{-1} \, \boldsymbol{s_0}\|_2^2}{\|\boldsymbol{s_0}\|_2^2} = \max_{\|\boldsymbol{s_0}\|_2 = 1} \|F \, e^{At} F^{-1} \, \boldsymbol{s_0}\|_2^2 \triangleq \|F \, e^{At} F^{-1}\|_2^2 = \sigma_1^2$$

where σ_1 is the maximum singular value of $F \, e^{At} F^{-1}$. Computationally, it is expensive in general to evaluate e^{At} since, depending on the numerical discretization, A may be a dense matrix. A better solution is to change the formulation of the original problem by writing it in the basis of the eigenvectors of A. Using the eigenvalue decomposition of A, *i.e.* $A = PDP^{-1}$, it is obvious that the linear state-space system (52) simplifies in

$$\dot{\boldsymbol{K}} = D \, \boldsymbol{K} \quad \text{with} \quad \boldsymbol{K}(0) = \boldsymbol{K_0}. \tag{54}$$

Here, $\boldsymbol{K}(t)$ corresponds to the coefficients of the development of $\boldsymbol{q}$ into the eigenvectors of A: $\boldsymbol{q}(t) = P\boldsymbol{K}(t)$. The solution of (54) writes immediately

$$\boldsymbol{K}(t) = e^{Dt}\boldsymbol{K_0}, \tag{55}$$

where D is the diagonal matrix of eigenvalues of A.

With this new formulation, the energy inner product is:

$$\begin{aligned}\|\boldsymbol{q}(t)\|_E^2 = \boldsymbol{q}(t)^H\, Q_q\, \boldsymbol{q}(t) = \boldsymbol{K}(t)^H\, P^H Q_q P\, \boldsymbol{K}(t) = \boldsymbol{K}(t)^H\, F^H F\, \boldsymbol{K}(t)\\ = \|F\, \boldsymbol{K}(t)\|_2^2\end{aligned}$$

where this time the matrix F corresponds to the Cholesky decomposition of $P^H Q_q P$.

Following exactly the same approach as previously in this section, we obtain

$$G(t) = \|F\, e^{Dt} F^{-1}\|_2^2 = \sigma_1^2. \tag{56}$$

This expression corresponds to a convenient and efficient way of computing the maximum transient growth. Indeed, (56) involves the L_2-norm (very convenient) and evaluates the matrix exponential of a diagonal matrix (very efficient).

6 Linearized Burgers equation

In section 4, we considered the LQR control where the state equation was written as a Linear Time Invariant state-space model, the cost objective was quadratic and the control was distributed. Section 5 was dedicated to the study of optimal growth disturbances where, compared to the case of LQR, the control corresponded to the initial condition of the linear state equation and where the objective was to maximize the energy amplitude of the perturbations. In this section, we now consider the case of the boundary control where the state equation is a Partial Differential Equation. This configuration is characteristic from the applications that can be met in fluid mechanics (Bewley et al., 2001; El Shrif, 2008), heat transfers and thermal systems (Müller, 2006). To simplify the presentation, we consider a one-dimensional configuration and take for state equation the linearized Burgers equation. The optimality system for the nonlinear Burgers equation with boundary and distributed control is given in section 7.

We will follow as closely as possible the presentation made in section 5 for the optimal growth perturbations.

6.1 Problem formulation and Lagrangian-based approach

Define $\Xi = \{(x,t)\,|\,(x,t) \in [0,L] \times [t_0,t_f]\}$ as the physical domain of the process. To simplify the future developments, we consider that $\Xi = \Omega_x \times \Omega_t$ where $\Omega_x = [0,L]$ and $\Omega_t = [t_0,t_f]$ and introduce three inner products

$$\langle u,v\rangle_\Xi = \int_\Xi u(x,t)v(x,t)\ \mathrm{d}x\ \mathrm{d}t, \tag{57}$$

$$\langle u,v\rangle_{\Omega_x} = \int_{\Omega_x} u(x,t)v(x,t)\ \mathrm{d}x, \tag{58}$$

$$\langle u,v\rangle_{\Omega_t} = \int_{\Omega_t} u(x,t)v(x,t)\ \mathrm{d}t, \tag{59}$$

where $u(x,t)$ and $v(x,t)$ are two sufficiently regular real-valued functions defined on Ξ.

The linearized Burgers equation is given by

$$F_U(u) = u_t + U(x)u_x - u_{xx} = 0 \quad \text{with} \quad u(x,t_0) = u_0(x), \tag{60}$$

where $u_t = \dfrac{\partial u}{\partial t}$, $u_x = \dfrac{\partial u}{\partial x}$ and $u_{xx} = \dfrac{\partial^2 u}{\partial x^2}$. In (60), $U(x)$ is a real-valued function defined on Ω_x. Furthermore, we consider that at the upper boundary of Ω_x, we have $u(L,t) = 0$.

In this section, our objective is to determine the function $u_w(t) = u(x = 0,t)$ (*i.e.* the temporal disturbance at the lower boundary of the domain) such that the classical L_2 norm of u is minimized at the final time of integration ($t = t_f$). In other words, we seek to minimize the cost functional defined by

$$\mathcal{J}(u,u_w) = \left(\int_{\Omega_x} u^2\ \mathrm{d}x\right)_{t=t_f} = \langle u,u\rangle_{\Omega_x}|_{t=t_f}\,.$$

In addition, the optimization problem must be mathematically well-posed (see the discussion in section 3.1.2). It is then necessary to add a regularization term to the cost functional $\mathcal{J}$. Finally, the cost functional

$$\begin{aligned}\mathcal{J}(u,u_w) &= \left(\int_0^L u^2\ \mathrm{d}x\right)_{t=t_f} + \ell\int_{t_0}^{t_f} u_w^2\ \mathrm{d}t,\\ &= [[u,u]] + \ell\langle u_w,u_w\rangle_{\Omega_t},\end{aligned} \tag{61}$$

is used where the penalty parameter $\ell > 0$ allows us to set the "price" of the control effort. To simplify further the expression of $\mathcal{J}$, an additional inner product was introduced in (61). This scalar product is defined as

$$[[u, v]] = \left(\int_{\Omega_x} u(x,t)v(x,t)\ \mathrm{d}x \right)_{t=t_f} = \langle u, v \rangle_{\Omega_x}|_{t=t_f}\,. \tag{62}$$

Contrary to the case of the optimal disturbances (see section 5.1), the control intervenes as a boundary condition and no more as an initial condition. However, the same formalism remains applicable and the boundary condition can be enforced through the relation:

$$H(u, u_w) = u(0,t) - u_w(t) = 0.$$

The original constrained optimization problem is then stated:

> *Determine the solution $u(x,t)$ and the control parameter $u_w(t)$ (upper boundary condition) such that the cost functional $\mathcal{J}$ reaches a minimum subject to $F_U(u) = 0$ and $H(u, u_w) = 0$.*

The procedure described in section 3.2 is then followed for enforcing the constraints[31]. For that, we introduce a single vector space $\Theta = u \times u_w \times u^+ \times \lambda^+$ where $u^+(x,t)$ and $\lambda^+(t)$ are the Lagrange multipliers corresponding respectively to $F_U = 0$ and $H = 0$. Let $\boldsymbol{Q^i} = \left(u^i, u_w^i, (u^+)^i, (\lambda^+)^i\right)$ with $i = I, II$ be two arbitrary elements of Θ, an inner product including the inner products (57) and (59) is defined as

$$\begin{aligned}\{\boldsymbol{Q^I}, \boldsymbol{Q^{II}}\} = &\left\langle u^I, u^{II} \right\rangle_\Xi + \left\langle u_w^I, u_w^{II} \right\rangle_{\Omega_t} \\ &+ \left\langle \left(u^+\right)^I, \left(u^+\right)^{II} \right\rangle_\Xi + \left\langle \left(\lambda^+\right)^I, \left(\lambda^+\right)^{II} \right\rangle_{\Omega_t}.\end{aligned} \tag{63}$$

The constraints are then enforced by introducing the Lagrangian functional

$$\mathcal{L}(u, u_w, u^+, \lambda^+) = \mathcal{J}(u, u_w) - \langle F_U(u), u^+ \rangle_\Xi - \langle H(u, u_w), \lambda^+ \rangle_{\Omega_t}. \tag{64}$$

The new unconstrained optimization problem can then be stated as:

[31] The constraints of the problem which are not imposed by Lagrange multipliers (initial condition and boundary condition at the upper boundary of the spatial domain) will be enforced a posteriori on the solutions.

Determine the solution $u(x,t)$, the control parameter $u_w(t)$ and the Lagrange multipliers $u^+(x,t)$ and $\lambda^+(t)$ such that the Lagrangian functional $\mathcal{L}$ reaches a minimum.

6.2 Optimality system

As it was discussed in details in section 3.2.2, a necessary condition for obtaining a minimum or maximum of the Lagrangian is to set the gradients of $\mathcal{L}$ with respect to u, u_w, u^+ and λ^+ equal to zero *i.e.*

$$\boldsymbol{\nabla}_u \mathcal{L} = \boldsymbol{\nabla}_{u_w} \mathcal{L} = \boldsymbol{\nabla}_{u^+} \mathcal{L} = \boldsymbol{\nabla}_{\lambda^+} \mathcal{L} = 0.$$

The derivation of the optimality system thus passes by the computation of the gradients of $\mathcal{L}$ with respect to all the variables.

6.2.1 Direct equations

Differentiating (64) with respect to the adjoint variables u^+ and λ^+ yields immediately to

$$\begin{aligned}\langle \boldsymbol{\nabla}_{u^+} \mathcal{L}, \delta u^+ \rangle_\Xi &= \lim_{\epsilon \to 0} \frac{\mathcal{L}(u, u_w, u^+ + \epsilon \delta u^+, \lambda^+) - \mathcal{L}(u, u_w, u^+, \lambda^+)}{\epsilon} \\ &= -\langle F_U(u), \delta u^+ \rangle_\Xi = 0, \end{aligned} \tag{65}$$

and

$$\begin{aligned}\langle \boldsymbol{\nabla}_{\lambda^+} \mathcal{L}, \delta \lambda^+ \rangle_{\Omega_t} &= \lim_{\epsilon \to 0} \frac{\mathcal{L}(u, u_w, u^+, \lambda^+ + \epsilon \delta \lambda^+) - \mathcal{L}(u, u_w, u^+, \lambda^+)}{\epsilon} \\ &= -\langle H(u, u_w), \delta \lambda^+ \rangle_{\Omega_t} = 0. \end{aligned} \tag{66}$$

Since the variations δu^+ and $\delta \lambda^+$ are arbitrary, F_U and H necessarily have to vanish, and we thus recover the constraints.

6.2.2 Adjoint equations

For differentiating (64) with respect to the direct variable u, we consider a variation $\delta \boldsymbol{Q} = (\delta u, 0, 0, 0)$. It yields

$$\langle \boldsymbol{\nabla}_u \mathcal{L}, \delta u \rangle_\Xi = \lim_{\epsilon \to 0} \frac{\mathcal{L}(u + \epsilon \delta u, u_w, u^+, \lambda^+) - \mathcal{L}(u, u_w, u^+, \lambda^+)}{\epsilon} = 0. \tag{67}$$

Substituting $\mathcal{L}$ with its definition, we have immediately due to the linearity of F_U and H:

$$\begin{aligned}\langle \boldsymbol{\nabla}_u \mathcal{L}, \delta u\rangle_\Xi &= \lim_{\epsilon\to 0} \frac{\mathcal{J}(u+\epsilon\delta u, u_w) - \mathcal{J}(u, u_w)}{\epsilon} \\ &- \lim_{\epsilon\to 0} \frac{\langle F_U(u+\epsilon\delta u), u^+\rangle_\Xi - \langle F_U(u), u^+\rangle_\Xi}{\epsilon} \\ &- \lim_{\epsilon\to 0} \frac{\langle H(u+\epsilon\delta u, u_w), \lambda^+\rangle_{\Omega_t} - \langle H(u, u_w), \lambda^+\rangle_{\Omega_t}}{\epsilon} \\ &= 2[[u, \delta u]] - \langle F_U(\delta u), u^+\rangle_\Xi - \langle \delta u(0,t), \lambda^+\rangle_{\Omega_t}. \end{aligned} \tag{68}$$

The objective is now to write the term $\langle F_U(\delta u), u^+\rangle_\Xi$ as a particular inner product utilizing δu. By definitions of the scalar product (57) and of F_U, we have:

$$\langle F_U(\delta u), u^+\rangle_\Xi = \underbrace{\langle (\delta u)_t, u^+\rangle_\Xi}_{T_I} + \underbrace{\langle U(x)(\delta u)_x, u^+\rangle_\Xi}_{T_{II}} - \underbrace{\langle (\delta u)_{xx}, u^+\rangle_\Xi}_{T_{III}}.$$

The simplification of the terms T_I to T_{III} is then equivalent to an exercise of integration by parts. We obtain immediately:

$$T_I = \int_\Xi u^+ \, (\delta u)_t \ \mathrm{d}x \ \mathrm{d}t = \left[\int_{\Omega_x} u^+ \delta u \ \mathrm{d}x\right]_{t=t_0}^{t=t_f} - \langle \delta u, u_t^+\rangle_\Xi,$$

$$T_{II} = \int_\Xi u^+ \, U(x) \, (\delta u)_x \ \mathrm{d}x \ \mathrm{d}t = \left[\int_{\Omega_t} U \, u^+ \delta u \ \mathrm{d}t\right]_{x=0}^{x=L} - \langle \delta u, \left(U \, u^+\right)_x\rangle_\Xi,$$

and

$$T_{III} = \int_\Xi u^+ \ (\delta u)_{xx} \ \mathrm{d}x \ \mathrm{d}t = \left[\int_{\Omega_t} u^+ \, (\delta u)_x \ \mathrm{d}t\right]_{x=0}^{x=L} - \int_\Xi u_x^+ \ (\delta u)_x \ \mathrm{d}x \ \mathrm{d}t$$

where

$$\int_\Xi u_x^+ \ (\delta u)_x \ \mathrm{d}x \ \mathrm{d}t = \left[\int_{\Omega_t} \delta u \, u_x^+ \ \mathrm{d}t\right]_{x=0}^{x=L} - \langle \delta u, u_{xx}^+\rangle_\Xi.$$

By replacing the terms T_I to T_{III} by their expressions, we can write (68) as

$$\begin{aligned} &2\left(\int_{\Omega_x} u \, \delta u \ \mathrm{d}x\right)_{t=t_f} - \int_{\Omega_t} \lambda^+ \, \delta u(0,t) \ \mathrm{d}t + \langle u_t^+ + \left(U u^+\right)_x + u_{xx}^+, \delta u\rangle_\Xi \\ &- \left[\int_{\Omega_t} \left(\left(U u^+ + u_x^+\right) \delta u - u^+ \, (\delta u)_x\right) \ \mathrm{d}t\right]_{x=0}^{x=L} - \left[\int_{\Omega_x} u^+ \delta u \ \mathrm{d}x\right]_{t=t_0}^{t=t_f} = 0. \end{aligned} \tag{69}$$

This last expression has to hold true for any variation δu. That means that (69) must be in particular true for variations presumed unspecified on Ξ, except at its boundaries $\partial\Xi$ where we consider

$$\delta u(x,t) = 0 \qquad \forall (x,t) \in \partial\Xi.$$

Introducing this particular perturbation in (69) yields to

$$\langle u_t^+ + \left(Uu^+\right)_x + u_{xx}^+, \delta u \rangle_\Xi = 0,$$

what leads to the adjoint equation of the linearized Burgers equation

$$F_U^+(u^+) = u_t^+ + (Uu^+)_x + u_{xx}^+ = 0 \qquad \forall (x,t) \in \Xi. \tag{70}$$

Contrary to the direct equation (60), the adjoint equation is now parabolic in decreasing time. As a consequence, it is then necessary to provide (70) with a terminal condition. In addition, since the cost functional $\mathcal{J}$ that we considered is defined locally in time (minimization of the norm of u at final time t_f), the adjoint equation is independent of the optimization problem. This was not the case for the LQR problem (see section 4.3.2) where a source term coming directly from $\mathcal{J}$ appears in the adjoint equation.

According to the adjoint equation, (69) is now equivalent to

$$\begin{aligned} &2\left(\int_{\Omega_x} u\,\delta u \,\mathrm{d}x\right)_{t=t_f} - \int_{\Omega_t} \lambda^+ \,\delta u(0,t)\,\mathrm{d}t \\ &- \left[\int_{\Omega_t} \left(\left(Uu^+ + u_x^+\right)\delta u - u^+\left(\delta u\right)_x\right)\,\mathrm{d}t\right]_{x=0}^{x=L} - \left[\int_{\Omega_x} u^+ \delta u\,\mathrm{d}x\right]_{t=t_0}^{t=t_f} = 0. \end{aligned} \tag{71}$$

In terms of perturbations, the initial and upper boundary conditions of the original problem result in

$$\delta u(x,t_0) = 0 \qquad \text{and} \qquad \delta u(L,t) = 0.$$

Using these conditions, (71) may then be rewritten as

$$\begin{aligned} &\left(\int_{\Omega_x} \left(2u - u^+\right)\,\delta u\,\mathrm{d}x\right)_{t=t_f} + \left(\int_{\Omega_t} u^+\left(\delta u\right)_x\,\mathrm{d}t\right)_{x=L} \\ &\qquad + \left[\int_{\Omega_t} \left(\left(Uu^+ + u_x^+ - \lambda^+\right)\delta u - u^+\left(\delta u\right)_x\right)\,\mathrm{d}t\right]_{x=0} = 0. \end{aligned} \tag{72}$$

Since the perturbation δu may be chosen arbitrarily, we can consider:

- Functions u which are unspecified at $t = t_f$ but with constant values along $x = 0$ and $x = L$. In terms of perturbations, this choice corresponds to

$$\delta u(0,t) = 0 \qquad \text{and} \qquad \delta u(L,t) = 0.$$

Inserting these particular perturbations in (72) lead to the terminal condition of the adjoint equation:

$$u^+(x,t_f) = 2u(x,t_f). \tag{73}$$

- Functions u which are constant everywhere on Ξ except at $x = L$, *i.e.*

$$\delta u(0,t) = 0 \qquad \text{and} \qquad (\delta u)_x(0,t) = 0.$$

This choice yields to a first boundary condition for the adjoint equation:

$$u^+(L,t) = 0. \tag{74}$$

- Finally, we consider functions u which are constant everywhere on Ξ except at $x = 0$, *i.e.*

$$(\delta u)_x(L,t) = 0.$$

This type of perturbation leads to a second boundary condition for the adjoint equation:

$$u^+(0,t) = 0 \tag{75}$$

and to the compatibility condition

$$u_x^+(0,t) - \lambda^+(t) = 0. \tag{76}$$

6.2.3 Optimality conditions

For the differentiation of (64) with respect to the control variable u_w, a variation $\delta \boldsymbol{Q} = (0, \delta u_w, 0, 0)$ is considered. By definition of the Fréchet and Gâteau differentials, we obtain:

$$\langle \boldsymbol{\nabla}_{u_w} \mathcal{L}, \delta u_w \rangle_{\Omega_t} = \lim_{\epsilon \to 0} \frac{\mathcal{L}(u, u_w + \epsilon \delta u_w, u^+, \lambda^+) - \mathcal{L}(u, u_w, u^+, \lambda^+)}{\epsilon} = 0. \tag{77}$$

If we then replace the Lagrangian functional $\mathcal{L}$ by its definition, we obtain immediately by linearizing the expression of $\mathcal{J}$ and by linearity of H that:

$$\begin{aligned} \langle \boldsymbol{\nabla}_{u_w} \mathcal{L}, \delta u_w \rangle_{\Omega_t} &= \lim_{\epsilon \to 0} \frac{\mathcal{J}(u, u_w + \epsilon \delta u_w) - \mathcal{J}(u, u_w)}{\epsilon} \\ &\quad - \lim_{\epsilon \to 0} \frac{\langle H(u, u_w + \epsilon \delta u_w) - H(u, u_w), \lambda^+ \rangle_{\Omega_t}}{\epsilon} \\ &= 2\ell \langle u_w, \delta u_w \rangle_{\Omega_t} + \langle \lambda^+, \delta u_w \rangle_{\Omega_t} = 0. \end{aligned} \tag{78}$$

Since the variation δu_w is arbitrary, we determine the gradient of $\mathcal{J}$ with respect to u

$$\boldsymbol{\nabla}_{u_w}\mathcal{L} = 2\ell u_w + \lambda^+,$$

and the optimality condition

$$-2\ell u_w(t) = \lambda^+(t).$$

Using the compatibility condition (76), the optimality condition simplifies in

$$-2\ell u_w(t) = u_x^+(0,t). \tag{79}$$

The optimality system corresponding to the linearized Burgers equation is given in Fig. 18.

7 Non-linear Burgers equation

In this section, we consider the case of the nonlinear Burgers equation where the control is both distributed and applied at the boundaries of the spatial domain. Since this is a natural extension to the linearized Burgers equation studied in section 6, we will not detail the derivation of the optimality system (see El Shrif, 2008, for that) but we will present some numerical results of optimal control (section 7.2).

7.1 Formulation and optimality system

The Burgers equation that we consider is given by

$$F(u,\Phi) = \frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial u^2}{\partial x} - \nu\frac{\partial^2 u}{\partial x^2} - \Phi = 0 \quad \text{with} \quad u(x,t_0) = u_0(x),$$

where ν is the kinematic viscosity and $\Phi(x,t)$ corresponds to the distributed control. In our application, this equation is solved on a physical domain defined by $\Xi = \{(x,t) \setminus (x,t) \in \Omega_x \times \Omega_t\}$ with $\Omega_x = [0,L]$ and $\Omega_t = [t_0,t_f]$. Furthermore, we suppose that boundary controls are applied at the left and right boundaries of the domain. The boundary conditions are then

$$u(0,t) = \phi_L(t)$$

$$u(L,t) = \phi_R(t)$$

where ϕ_L and ϕ_R correspond to control parameters.

State equation:

$$\begin{aligned} F_U(u) &= u_t + U(x)u_x - u_{xx} = 0 \\ u(x,t_0) &= u_0(x) \qquad \text{(I.C.)} \\ u(0,t) &= u_w(t) \qquad \text{(B.C.) and control parameter} \\ u(L,t) &= 0 \qquad \text{(B.C.)} \end{aligned}$$

Adjoint equation:

$$\begin{aligned} F_U^+(u^+) &= u_t^+ + (Uu^+)_x + u_{xx}^+ = 0 \qquad \forall (x,t) \in \Xi \\ u^+(0,t) &= 0 \qquad \text{(B.C.)} \\ u^+(L,t) &= 0 \qquad \text{(B.C.)} \\ u^+(x,t_f) &= 2u(x,t_f) \qquad \text{(T.C.)} \end{aligned}$$

Optimality condition:

$$-2\ell u_w(t) = u_x^+(0,t)$$

Cost functional:

$$\mathcal{J}(u,u_w) = \left(\int_0^L u^2 \, \mathrm{d}x \right)_{t=t_f} + \ell \int_{t_0}^{t_f} u_w^2 \, \mathrm{d}t$$

Figure 18. Optimality system for the linearized Burgers equation. B.C.: boundary condition, I.C.: initial condition, T.C.: terminal condition.

In this section, we seek to minimize the cost functional defined by

$$\begin{aligned}\mathcal{J}(u,\Phi,\phi_L,\phi_R) =& \frac{\omega_1}{2}\int_{\Omega_t}\int_{\Omega_x}(u-\hat{u})^2\ \mathrm{d}x\ \mathrm{d}t + \frac{\omega_2}{2}\int_{\Omega_x}[u(x,t_f)-\bar{u}(x)]^2\ \mathrm{d}x \\ &+\frac{\ell_1}{2}\int_{\Omega_t}\phi_L^2(t)\ \mathrm{d}t + \frac{\ell_2}{2}\int_{\Omega_t}\phi_R^2(t)\ \mathrm{d}t \\ &+\frac{\ell}{2}\int_{\Omega_t}\int_{\Omega_x}\Phi^2(x,t)\ \mathrm{d}x\ \mathrm{d}t.\end{aligned}$$

The first two terms try to match the solution u respectively on Ξ and over the spatial domain Ω_x at t_f to given functions $\hat{u}$ and $\bar{u}$. These two targets are generally determined based on physical arguments: laminar flow, solutions of minimum drag, unstable steady solutions,... This type of functional corresponds to a target optimization problem. The last three terms are penalty terms that limit the size of the control functions Φ, ϕ_L and ϕ_R. The positive constants ω_1, ω_2, ℓ, ℓ_1 and ℓ_2 are chosen to adjust the relative importance of the five terms in $\mathcal{J}$.

The problem of optimization that we are interested to solve is

> *Determine the solution u and the control parameters Φ, ϕ_L and ϕ_R such that the cost functional $\mathcal{J}$ reaches a minimum.*

This constrained optimization problem is absolutely similar to the one treated in section 6.2 for the linearized Burgers equation. After developments (see El Shrif, 2008, for the details), the optimality system summarized in Fig. 19 is obtained.

7.2 Results of optimal control

7.2.1 Numerical parameters and space-time discretization

For the numerical applications, we consider a simplified version of the original optimization problem described in section 7.1. First, we suppose that the control is not applied at the spatial boundaries *i.e.* $\phi_L = \phi_R = 0$ in Fig. 19. Second, we assume that the targets $\hat{u}$ and $\bar{u}$ are equal to the initial condition u_0 (see below for the expression). Third, we take as weighting coefficients for the two targets $\omega_1 = \omega_2 = 1$. Finally, the cost functional

State equation:

$$F(u) = \frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial u^2}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} - \Phi = 0 \tag{80a}$$
$$u(0,t) = \phi_L \qquad \text{(B.C.)}$$
$$u(L,t) = \phi_R \qquad \text{(B.C.)}$$
$$u(x,0) = u_0(x) \qquad \text{(I.C.)}$$

Adjoint equation:

$$F^+(u^+) = -\frac{\partial u^+}{\partial t} - u\,\frac{\partial u^+}{\partial x} - \nu\,\frac{\partial^2 u^+}{\partial x^2} = \omega_1\,(u - \hat{u}) \tag{80b}$$
$$u^+(0,t) = 0 \qquad \text{(B.C.)}$$
$$u^+(L,t) = 0 \qquad \text{(B.C.)}$$
$$u^+(x,t_f) = \omega_2\,(u(x,t_f) - \bar{u}(x)) \qquad \text{(T.C.)} \tag{80c}$$

Optimality condition:

$$\nabla_\Phi \mathcal{J} = \ell\,\Phi + u^+ \tag{80d}$$
$$\nabla_{\phi_L} \mathcal{J} = \ell_1 \phi_L + \nu\,\frac{\partial u^+}{\partial x}(0,t)$$
$$\nabla_{\phi_R} \mathcal{J} = \ell_2 \phi_R - \nu\,\frac{\partial u^+}{\partial x}(L,t)$$

Cost functional:

$$\mathcal{J} = \frac{\omega_1}{2}\int_{\Omega_t}\int_0^L (u-\hat{u})^2\;\mathrm{d}x\;\mathrm{d}t + \frac{\omega_2}{2}\int_{\Omega_x}[u(x,t_f) - \bar{u}(x)]^2\;\mathrm{d}x$$
$$+\frac{\ell}{2}\int_{\Omega_t}\int_{\Omega_x}\Phi^2(x,t)\;\mathrm{d}x\;\mathrm{d}t + \frac{\ell_1}{2}\int_{\Omega_t}\phi_L^2(t)\,dt + \frac{\ell_2}{2}\int_{\Omega_t}\phi_R^2(t)\;\mathrm{d}t$$

Figure 19. Optimality system for the non-linear Burgers equation. B.C.: boundary condition, IC: initial condition, T.C.: terminal condition.

corresponds to

$$\mathcal{J}(u,\Phi) = \frac{1}{2}\int_{\Omega_t}\int_{\Omega_x}(u-u_0)^2\ \mathrm{d}x\ \mathrm{d}t + \frac{1}{2}\int_{\Omega_x}\left[u(x,t_f)-u_0(x)\right]^2\ \mathrm{d}x + \frac{\ell}{2}\int_{\Omega_t}\int_{\Omega_x}\Phi^2(x,t)\ \mathrm{d}x\ \mathrm{d}t$$

with $\Omega_x = [0,1]$ and $\Omega_t = [0,1]$. Here the initial condition is selected as

$$u_0(x) = \sin\left(\pi\frac{\tan\left(c_s\left(2x-1\right)\right)}{\tan(c_s)}\right)$$

where c_s is a stretching coefficient introduced to represent correctly the boundary layers. In our simulations, we choose $c_s = 1.3$ and $\nu = 0.01$.

To solve numerically the Burgers equation (80a) and the adjoint equation (80b), a numerical scheme known as *Forward Time, Centered Space* (FTCS) is used. This method corresponds to a forward scheme in time and to a centered finite difference sheme of order 2 in space. The main interest of this scheme is its easiness of implementation. The direct and adjoint equations are discretized in time and space on a constant mesh (Δt, Δx). Noting $u_{j,n} = u(j\Delta x, n\Delta t)$, the discretized versions of (80a) and (80b) are respectively

$$u_{j,n+1} = u_{j,n}\left(1-2s\right) + s\left(u_{j+1,n}+u_{j-1,n}\right) - \frac{\Delta t}{2\Delta x}u_{j,n}\left(u_{j+1,n}-u_{j-1,n}\right) + \Phi_{j,n}, \quad (81)$$

and

$$u^+_{j,n+1} = u^+_{j,n}\left(1+2s\right) - s\left(u^+_{j+1,n}+u^+_{j-1,n}\right) - \frac{\Delta t}{2\Delta x}u_{j,n}\left(u^+_{j+1,n}-u^+_{j-1,n}\right) - \omega_1\left(u_{j,n}-\hat{u}_{j,n}\right)\Delta t, \quad (82)$$

where $s = \nu\frac{\Delta t}{\Delta x^2}$.

7.2.2 Optimization procedure

The optimality system given in Fig. 19 is now solved iteratively. The procedure is similar to the one described in section 3.2.3 in the general framework. Let k ($k = 0, \cdots, +\infty$) be the iteration number of the optimization procedure, the gradient $\boldsymbol{g}^{(k)}$ determined from the adjoint formulation can be used

to estimate a descent[32] direction $\boldsymbol{p}^{(k)}$. A strategy for optimization of the control parameters[33] $\boldsymbol{c}^{(k)}$ is then to approach the minimum by a sequence of steps constructed as

$$\boldsymbol{c}^{(k+1)} = \boldsymbol{c}^{(k)} + \alpha^{(k)}\, \boldsymbol{p}^{(k)} \tag{83}$$

where $\alpha^{(k)}$ is a positive scalar called step length.

The computation of $\alpha^{(k)}$ is the linesearch, and may itself be iterative. It can be proved (Gould and Leyffer, 2002) that if the linesearch allows steps that are either too long or too short relative to the amount of decrease that they provide, then wrong convergences may appear. For determining the values of $\alpha^{(k)}$, one possibility is to search for the minimizer of $\mathcal{J}(u, \boldsymbol{c}^{(k)} + \alpha^{(k)}\, \boldsymbol{p}^{(k)})$. Due to the expensive cost of the method, this type of exact linesearch is rarely employed nowadays. Here, we prefer to use the backtracking Armijo method for which there exists some guarantee of sufficient decrease of the cost functional (see Nocedal and Wright, 1999, for the algorithm).

The simplest choice of the descent direction corresponds to the steepest-descent method for which $\boldsymbol{p}^{(k)} = -\boldsymbol{g}^{(k)}$. The drawback of such a steepest descent technique is that the descent direction is only based on local information whereas after some iterations, we have a more global description of the cost functional. For this reason, a particular variant of the conjugate gradient algorithm, referred as the Hestenes-Stiefel method (Nocedal and Wright, 1999), was selected. It is given by

$$\boldsymbol{p}^{(k+1)} = -\boldsymbol{g}^{(k+1)} + \beta^{(k+1)}\, \boldsymbol{p}^{(k)} \quad \text{with} \quad \boldsymbol{p}^{(0)} = -\boldsymbol{g}^{(0)} \tag{84}$$

where the coefficient $\beta^{(k+1)}$ is defined as

$$\beta^{(k+1)} = \frac{\left(\boldsymbol{g}^{(k+1)}\right)^T \left(\boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)}\right)}{\left(\boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)}\right)^T \boldsymbol{p}^{(k)}}.$$

The optimization algorithm is the following:

Step 1: Starting from a guess value $\boldsymbol{c}^{(k)}$ for the distributed control (no control is in general an acceptable guess value for $k = 0$), the discretized version (81) of the direct Burgers equation (80a) is solved forward in time from $t = 0$ to $t = t_f$.

[32] By definition, $\boldsymbol{p}^{(k)}$ is a descent direction if $\left(\boldsymbol{p}^{(k)}\right)^T \boldsymbol{g}^{(k)} < 0$ where $(\cdot)^T$ denotes the transposition.

[33] For the Burgers equation, the control parameter $\boldsymbol{c}$ is Φ.

Step 2: The terminal condition $u^+(x, t_f)$ for the adjoint variable is computed using (80c).

Step 3: The discretized version (82) of the adjoint Burgers equation (80b) is solved backward in time from $t = t_f$ to $t = 0$.

Step 4: The gradient of the cost functional $\mathcal{J}$ with respect to the control variable $\boldsymbol{c}$ is computed using (80d). This gradient $\boldsymbol{g}^{(k)} = (\boldsymbol{\nabla}_{\boldsymbol{c}}\mathcal{J})^{(k)}$ is estimated based on the adjoint variable u^+ determined in step 7.2.2.

Step 5: The gradient-based optimization method (83) is used for updating the control.

Step 5.1: The direction of descent $\boldsymbol{p}^{(k)}$ is determined by (84) based on the gradient $\boldsymbol{g}^{(k)}$ computed in step 7.2.2 and previous descent directions (conjugate gradient).

Step 5.2: An inexact linesearch (backtracking Armijo method) is used to determine the step length $\alpha^{(k)}$.

Step 5.3: The previous estimate of the optimal control is updated by $\boldsymbol{c}^{(k+1)} = \boldsymbol{c}^{(k)} + \alpha^{(k)}\,\boldsymbol{p}^{(k)}$.

Step 6: Return to step 7.2.2 and iterate until a given criterion of convergence is satisfied.

7.2.3 Results

Figure 20 represents at the final time of integration t_f, the solution u obtained by optimal control of the Burgers equation for the three first iterations of the optimization procedure. In accordance with what is expected, u converges rapidly to the target solution u_0. Indeed, we observe in Fig. 21 that about twenty iterations of the iterative procedure are sufficient for the cost functional $\mathcal{J}$ to converge. Finally, Fig. 22 represents the spatial evolution of the distributed control obtained at convergence at four different time instants.

8 Conclusion

In this chapter, we have outlined the interest of constrained optimization for solving different types of problems encountered in fluid mechanics, and particularly in flow control. Indeed, constrained optimization methods are in the heart of reduced-order modeling techniques such as POD, and of data assimilation methods used everyday in weather forecast or, within a different framework, for determining optimal growth perturbations. In addition, constrained optimization appears naturally in control theory, whether in linear control techniques such as LQR and LQG, or in nonlinear approaches such as Model Predictive Control (see the contribution by R. King in this

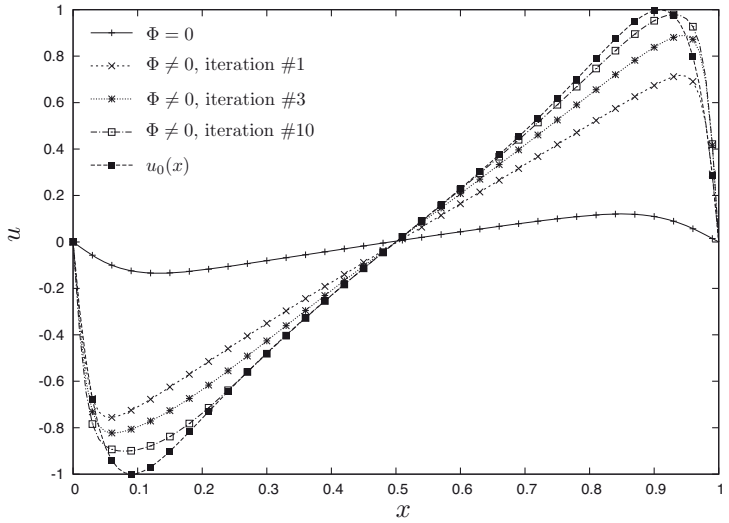


Figure 20. Distributed control of the Burgers equation with $\ell = 0.01$. Comparison of the solutions at final time t_f with and without control.

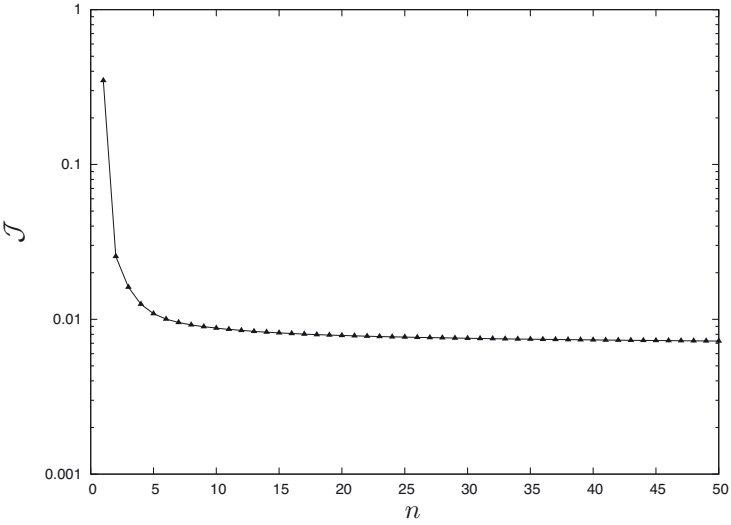


Figure 21. Distributed control of the Burgers equation with $\ell = 0.01$. Decrease of the cost functional $\mathcal{J}$ with the iteration number.

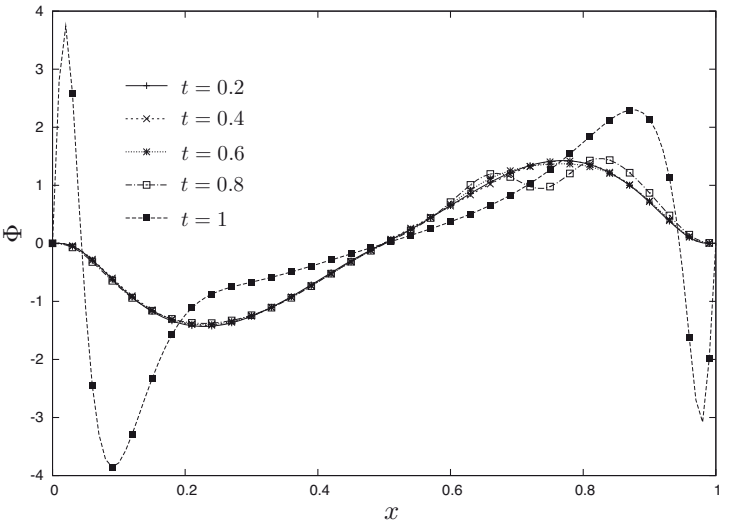


Figure 22. Distributed control of the Burgers equation with $\ell = 0.01$. Distributed control $\Phi(x, t)$ at convergence.

book), frequently used experimentally in process engineering. Lastly, optimal control and shape optimization, often considered in many fields of applied mathematics (computer fluid dynamics, computer graphics, multidisciplinary optimization to name a few), are nothing else than constrained optimization methods adapted to the resolution of a given problem. In section 3, fundamentals aspects of optimal control theory have been explained in details. This should allow an interested reader to derive a corresponding optimality system for his/her own problem of interest. In sections 4 to 6, different constrained optimization problems were formulated for linear 1D configuration. Finally, in section 7, the formalism is extended to a nonlinear 1D PDE equation. The reader is referred to El Shrif (2008) for the derivation of optimality systems for the Navier-Stokes equations where a Direct Numerical Simulation and a Large Eddy Simulation of the flow are both considered. Particularly noteworthy is the similarity of the formulation between the determination of optimal growth perturbations and other methods employed in control.

A Adjoint operator and inner product

Suppose Ω is an Hilbert space, with inner product $\langle \cdot, \cdot \rangle$. Consider a continuous linear operator $\mathcal{N} : \Omega \longrightarrow \Omega$. Using the Riesz representation theorem, one can show that there exists a unique continuous operator $\mathcal{N}^+ : \Omega \longrightarrow \Omega$ with the following property:

$$\langle \mathcal{N}\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \mathcal{N}^+\boldsymbol{y} \rangle \qquad \forall\, \boldsymbol{x},\, \boldsymbol{y} \in \Omega.$$

Here, the symbol + denotes the adjoint, and the operator $\mathcal{N}^+$ is the adjoint of $\mathcal{N}$. Since $\mathcal{N}$ is assumed linear, this definition can be extended directly to the matrix $\mathcal{A}$ associated to $\mathcal{N}$.

It should be noted that in general for a given matrix $\mathcal{A}$, we have $\mathcal{A}^+ \neq \mathcal{A}^H$, the two being equal only when the inner product used to derive the adjoint does not have an associated weight (classical Euclidean inner product). Indeed, if we consider the weighted inner product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ given by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_W = \boldsymbol{x}^H W \boldsymbol{y}$$

where W is positive definite, the continuous adjoint operator of $\mathcal{A}$ with respect to this inner product is defined as:

$$\begin{aligned}
\langle \mathcal{A}\boldsymbol{x}, \boldsymbol{y} \rangle_W &= \langle \boldsymbol{x}, \mathcal{A}^+\boldsymbol{y} \rangle_W \qquad \forall\, \boldsymbol{x},\, \boldsymbol{y} \in \Omega \\
\iff (\mathcal{A}\boldsymbol{x})^H W \boldsymbol{y} &= \boldsymbol{x}^H W \mathcal{A}^+ \boldsymbol{y} \\
\iff \boldsymbol{x}^H \mathcal{A}^H W \boldsymbol{y} &= \boldsymbol{x}^H W \mathcal{A}^+ \boldsymbol{y} \\
\iff \mathcal{A}^H W &= W \mathcal{A}^+ \\
\iff \mathcal{A}^+ &= W^{-1} \mathcal{A}^H W.
\end{aligned}$$

Since W is positive definite, W^{-1} is well defined and we thus have $\mathcal{A}^+ = W^{-1}\mathcal{A}^H W$ *i.e.* $\mathcal{A}^+ = \mathcal{A}^H$ only when W is equal to the identity matrix.

Bibliography

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

A. Antkowiak and P. Brancher. On vortex rings around vortices: an optimal mechanism. *J. Fluid Mech.*, 578:295–304, 2007.

A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems.* Advances in Design and Control. SIAM, 2005.

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

A. Barbagallo, D. Sipp, and P. Schmid. Closed-loop control of an open cavity flow using reduced order models. *J. Fluid Mech.*, 641:1–50, 2009.

P. Benner, J.-R. Li, and T. Penzl. Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems. *Numer. Linear Algebra Appl.*, 15:755–777, 2008.

M. Bergmann and L. Cordier. Optimal control of the cylinder wake in the laminar regime by Trust-Region methods and POD Reduced Order Models. *J. Comp. Phys.*, 227:7813–7840, 2008.

M. Bergmann, L. Cordier, and J.-P. Brancher. Optimal rotary control of the cylinder wake using POD Reduced Order Model. *Phys. Fluids*, 17 (9):097101:1–21, 2005.

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

T. R. Bewley, P. Moin, and R. Temam. DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms. *J. Fluid Mech.*, 447:179–225, 2001.

J.-F. Bonnans, J.-C. Gilbert, C. Lemarchal, and C. A. Sagastizbal. *Numerical Optimization*. Springer, 2003.

A. E. Bryson Jr. and Y. Ho. *Applied Optimal Control: Optimization, Estimation and Control*. Taylor & Francis, 1975.

J. B. Burl. *Linear Optimal Control: H_2 and H_∞ Methods*. Addison-Wesley Publishing, 1999.

Y. Chang. *Approximate models for optimal control of turbulent channel flow*. PhD thesis, universit de Rice, 2000.

P. Chassaing. *Turbulence en mécanique des fluides*. Polytech. Cépadués, 2000.

L. Cordier. Elements of control theory. In *Workshop "Flow Control Methods and Applications"*, Poitiers, December 7-11, 2009.

L. Cordier, B. Abou El Majd, and J. Favier. Calibration of POD Reduced-Order models using Tikhonov regularization. *Int. J. Numer. Meth. Fluids*, 63(2), 2010.

A. El Shrif. *Contrôle optimal d'un écoulement de canal turbulent*. PhD thesis, Institut National Polytechnique de Lorraine, 2008.

M. Fahl. *Trust-Region methods for flow control based on Reduced Order Modeling*. PhD thesis, Trier university, 2000.

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

B. Galletti, A. Bottaro, C.-H. Bruneau, and A. Iollo. Accurate model reduction of transient and forced wakes. *Eur. J. Mech. B/Fluids*, 26(3): 354–366, 2007.

N. I. M. Gould and S. Leyffer. An introduction to algorithms for nonlinear optimization. Technical Report RAL-TR-2002-031, Rutherford Appleton Laboratory, 2002. URL `http://www.numerical.rl.ac.uk/reports/reports.shtml`.

A. Griewank and A. Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, Second Edition*. SIAM, 2008.

A. Guegan, P. J. Schmid, and P. Huerre. Optimal energy growth and optimal control in swept Hiemenz flow. *J. Fluid Mech.*, 566:11–45, 2006.

S. Gugercin and A. C. Antoulas. Model reduction of large-scale systems by least squares. *Linear Algebra and its Applications*, 415(2-3):290–321, 2006.

M. D. Gunzburger. Introduction into mathematical aspects of flow control and optimization. In *Lecture series 1997-05 on inverse design and optimization methods*. Von Kármán Institute for Fluid Dynamics, 1997a.

M. D. Gunzburger. Lagrange multiplier techniques. In *Lecture series 1997-05 on inverse design and optimization methods*. Von Kármán Institute for Fluid Dynamics, 1997b.

M. D. Gunzburger. *Perspectives in flow control and optimization*. SIAM, 2003.

P. Huerre. Optimal control. In *First Summer School on Optimization and Control of Flows and Transfers*, Aussois, 12-17 March, 2006. Lecture notes.

M. Ilak. *Model reduction and feedback control of transitional channel flow*. PhD thesis, Princeton University, 2009.

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

S. Lall, J. E. Marsden, and S. Glavaski. A Subspace Approach to Balanced Truncation for Model Reduction of Nonlinear Control Systems. *International Journal of Robust and Nonlinear Control*, 12(6):519–535, 2002.

F. L. Lewis and V. L. Syrmos. *Optimal Control*. John Wiley and Sons, New York, second edition, 1995.

J. L. Lumley. *Atmospheric Turbulence and Wave Propagation. The structure of inhomogeneous turbulence*, pages 166–178. Nauka, Moscow, 1967.

V. Mehrmann and T. Stykel. Balanced truncation model reduction for large-scale systems in descriptor form. In V. Mehrmann P. Benner and D. Sorensen, editors, *Dimension Reduction of Large-Scale Systems*, volume 45 of *Lecture Notes in Computational Science and Engineering*, pages 83–115. Springer-Verlag, Berlin/Heidelberg, 2005.

B. Mohammadi and O. Pironneau. *Applied Shape Optimization for Fluids*. Oxford University Press, 2001.

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

E. A. Müller. *Optimal control of thermal systems*. PhD thesis, ETH Zurich, 2006.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer series in operations research, 1999.

B. Protas. On the "Vorticity" Formulation of the Adjoint Equations and its Solution Using Vortex Method. *Journal of Turbulence*, 3:48–55, 2002.

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

P. Sagaut. *Large-eddy simulation for incompressible flows - An introduction*. Springer-Verlag, 2005.

H. Schlichting and K. Gersten. *Boundary Layer Theory*. Springer, 8th edition, 2003.

P. J. Schmid. Nonmodal stability theory. *Ann. Rev. Fluid Mech.*, 39:129–162, 2007.

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

L. Sirovich. Turbulence and the dynamics of coherent structures. *Quarterly of Applied Mathematics*, XLV(3):561–590, 1987.

S. Skogestad and I. Postlethwaite. *Multivariable feedback control - Analysis and design, 2nd Edition*. Wiley, 2005.

P. R. Spalart, W. H. Jou, M. Strelets, and S. R. Allmaras. Comments on the Feasibility of LES for Wings, and on a Hybrid RANS/LES Approach. In *1st AFOSR Int. Conf. on DNS/LES*, Ruston, L.A., Aug. 4-8, 1997. In: Advances in DNS/LES, C. Liu and Z. Liu (eds.), Greyden Press, Columbus, OH.

A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.

Q. Wang, D. Gleich, A. Saberi, N. Etemadi, and P. Moin. A Monte Carlo method for solving unsteady adjoint equations. *J. Comp. Phys.*, 227(12): 6184–6205, 2008.

Q. Wang, P. Moin, and G. Iaccarino. Minimal Repetition Dynamic Checkpointing Algorithm for Unsteady Adjoint Calculation. *SIAM J. Sci. Comput.*, 31:2549, 2009.

K. Zhou, J. C. Doyle, and K. Glover. *Robust and optimal control*. Prentice Hall, 1996.