Supplemental Material: Annu. Rev. Fluid Mech. 2014. 46:493–517
doi: 10.1146/annurev-fluid-010313-141253
Adjoint Equations in Stability Analysis
Luchini and Bottaro

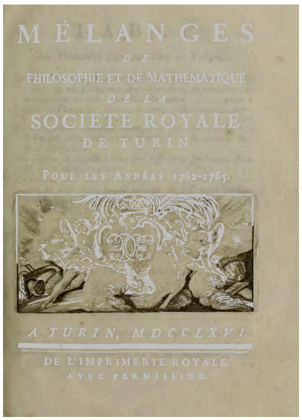
An Introduction to Adjoint Problems

Paolo Luchini[1] and Alessandro Bottaro[2]
[1]DIIN, University of Salerno, Italy
luchini@unisa.it
[2]DICCA, University of Genova, Italy
alessandro.bottaro@unige.it

Supplemental appendix to Adjoint Equations in Stability Analysis, *Annu. Rev. Fluid Mech.* **46**:493–517 (2014)

Contents

First use of adjoint equations by Lagrange, founding member of the Turin Royal Society



MÉLANGES

DE

PHILOSOPHIE ET DE MATHÉMATIQUE

DE LA

SOCIÉTÉ ROYALE

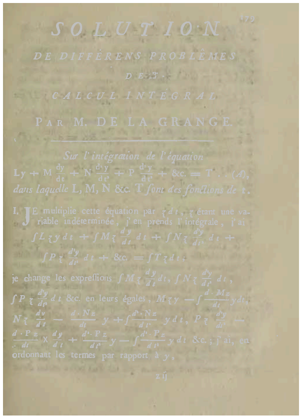
DE TURIN

POUR LES ANNÉES 1762-1765.

A TURIN, MDCCLXVI.

DE L'IMPRIMERIE ROYALE.

AVEC PERMISSION.



179

SOLUTION

DE DIFFÉRENS PROBLÊMES

DE

CALCUL INTÉGRAL

PAR M. DE LA GRANGE.

Sur l'intégration de l'équation

$Ly + M\frac{dy}{dt} + N\frac{d^2y}{dt^2} + P\frac{d^3y}{dt^3} + \&c. = T \;.\;.\; (A),$

dans laquelle L, M, N &c. T *ſont des fonctions de* t.

I. JE multiplie cette équation par $z\,dt$, z étant une variable indéterminée, j'en prends l'intégrale, j'ai

$$\int Lzy\,dt + \int Mz\frac{dy}{dt}dt + \int Nz\frac{d^2y}{dt^2}dt + \int Pz\frac{d^3y}{dt^3}dt + \&c. = \int Tz\,dt;$$

je change les expreſſions $\int Mz\frac{dy}{dt}dt$, $\int Nz\frac{d^2y}{dt^2}dt$, $\int Pz\frac{d^3y}{dt^3}dt$ &c. en leurs égales, $Mzy - \int\frac{d\cdot Mz}{dt}ydt$, $Nz\frac{dy}{dt} - \frac{d\cdot Nz}{dt}y + \int\frac{d^2\cdot Nz}{dt^2}ydt$, $Pz\frac{d^2y}{dt^2} - \frac{d\cdot Pz}{dt}\times\frac{dy}{dt} + \frac{d^2\cdot Pz}{dt^2}y - \int\frac{d^3\cdot Pz}{dt^3}ydt$ &c.; j'ai, en ordonnant les termes par rapport à y,

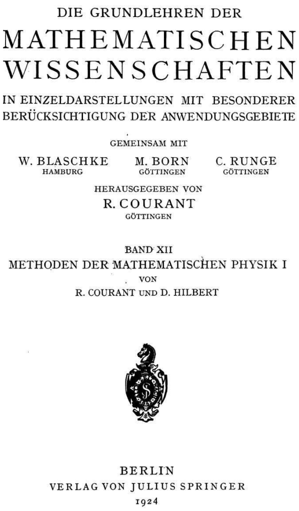
Z ij

1 Historical remarks

The first documented use of adjoint equations goes back to Lagrange (1763), who described a method to lower the order of a general linear ordinary differential equation and applied it to problems as diverse as fluid motion, vibrating strings and the orbit of planets. His method used one or more particular solutions of a certain auxiliary equation obtained through integration by parts, which became known as "équation adjointe" in the French literature. The French term "adjointe" was a century later transliterated into the German "adjungierte" by Fuchs (1873) and the English "adjoint" (after a brief alternation with "adjunct") by Forsyth (1888) and Craig (1889). "Self-adjoint" problems (those for which the direct and adjoint equation coincide) became prominent with the development of the theory of integral equations and Hilbert spaces, and laid out the foundation for quantum mechanics. The qualification "adjungierte" as used in Courant & Hilbert (1924) mostly refers to Fredholm integral equations with mutually transposed kernels — $K(t, s)$ being the transpose of $K(s, t)$ — or to their approximation by mutually transposed matrices, the link with Lagrange's "équation adjointe" being that every reasonably well-behaved differential problem (equation and boundary conditions) can be reformulated as an integral equation and the corresponding adjoints coincide.

In the last half-century, with the advent of computers on one hand and control theory on the other, the adjoints of discrete problems, governed by transposed matrices, acquired an emphasis of their own. It became evident that, in addition to being a powerful analytical tool, the adjoint formulation held numerical value in reducing computation times for a variety of problems in mathematical physics. It also became evident that this value was available whether the original problem was self-adjoint or not. A large number of applications opened up. In preparation we will now introduce adjoints of discrete and continuous systems of increasing complexity, thus reversing, as not rarely happens, the historical order.

Courant–Hilbert original title page



DIE GRUNDLEHREN DER

MATHEMATISCHEN WISSENSCHAFTEN

IN EINZELDARSTELLUNGEN MIT BESONDERER BERÜCKSICHTIGUNG DER ANWENDUNGSGEBIETE

GEMEINSAM MIT

W. BLASCHKE HAMBURG · M. BORN GÖTTINGEN · C. RUNGE GÖTTINGEN

HERAUSGEGEBEN VON

R. COURANT

GÖTTINGEN

BAND XII

METHODEN DER MATHEMATISCHEN PHYSIK I

VON

R. COURANT UND D. HILBERT

BERLIN

VERLAG VON JULIUS SPRINGER

1924

2 The simplest example: transpose of a matrix

Let us consider an *instantaneous*, *discrete* linear system (one in which a finite number of variables depend linearly on a finite number of others with no time delay) described by the explicit expression[1]

$$\mathbf{x} = \mathbf{H}\mathbf{u}, \tag{1}$$

$\mathbf{x}$ being an array of M real or complex variables (an M-vector), $\mathbf{u}$ an N-vector, and $\mathbf{H}$ an $M \times N$-matrix.

If the quantity of interest (the objective) is not the entire state $\mathbf{x}$ but a scalar linear function J of it, expressed through the linear form

$$J = \mathbf{y}\mathbf{x} \tag{2}$$

($\mathbf{y}$ being an M-vector of coefficients), the computation of J would normally entail the sequential evaluation of (1), requiring $\mathrm{O}(MN)$ *flops* (floating-point

[1] The juxtaposition of two vectors (identified by bold small letters) or matrices (identified by bold capital letters) is used here to denote index contraction, with order significant: $\mathbf{vu} = \sum_i v_i u_i$; $(\mathbf{Au})_i = \sum_j A_{ij} u_j$; $(\mathbf{vA})_j = \sum_i v_i A_{ij}$; $(\mathbf{AB})_{ij} = \sum_k A_{ik} B_{kj}$.

operations), and (2), requiring $\mathrm{O}(M)$ flops. However, substituting (1) into (2) gives $\mathbf{yx} = \mathbf{yHu} = \mathbf{vu}$ where $\mathbf{v} = \mathbf{yH}$, or equivalently

$$\mathbf{v} = \mathbf{H}^{\mathrm{T}}\mathbf{y}, \tag{3}$$

superscript $^{\mathrm{T}}$ denoting transposition. Therefore the objective J may also be rewritten in terms of the new N-vector $\mathbf{v}$ as

$$J = \mathbf{vu}. \tag{4}$$

The computational interest of this seemingly trivial relationship is that one can calculate $\mathbf{v}$ from $\mathbf{y}$ just once, at the same $\mathrm{O}(MN)$ cost as a single application of (1), and then obtain J for a number of different inputs at a cost of $\mathrm{O}(N)$ flops for each. The coefficient $\mathbf{v}$, which in this linear example provides the derivative of the objective J with respect to the input $\mathbf{u}$, may be named *sensitivity* of the objective to the input. Equation (3), which we shall name the *adjoint problem* of (1), therefore provides this sensitivity, at a computational advantage with respect to applying (1) to each input in turn.

The hermitian adjoint

A frequent alternate definition of the adjoint equation uses an inner product rather than a linear form. While we do not use an inner product in this article, elucidating the relationship may ease comprehension.

Functional analysis (e.g. Dunford & Schwartz (1958)) distinguishes between a *Banach space*, a vector space of functions endowed with a topology that allows the concepts of continuity and completeness to be defined, and a *Hilbert space*, a Banach space with the addition of an inner product. Using the algebraic properties of a vector space alone, the notion of its *dual* is introduced as the space of all linear functions from a vector to a scalar. The dual and original vector space do not generally overlap: for instance, the dual of a space of continuous and integrable functions may be a space of distributions (the mathematically orthodox way to deal with the Dirac delta function). Adjoint problems properly belong in the dual space. On the other hand when an inner product exists, which we may denote as $s = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle$, it establishes a mapping between the vector $\mathbf{x}_1$ and a linear function from the vector $\mathbf{x}_2$ to the scalar s, i.e. between the original state space and its dual. It then becomes possible to argue whether the problem possibly coincides with its adjoint (is *self-adjoint*) under this mapping.

For finite-dimensional vectors, such as appear in all numerical approximations, an element of the dual space is a linear form like (2); a general inner product can be defined as

$$\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \mathbf{x}_1^* \mathbf{Q} \mathbf{x}_2$$

where * denotes complex conjugation and $\mathbf{Q}$ is an arbitrary positive-definite (hermitian) matrix. (2) and (4) may then be alternately written as $J = \langle \bar{\mathbf{y}}, \mathbf{x} \rangle = \langle \bar{\mathbf{v}}, \mathbf{u} \rangle$, where

$$\mathbf{y} = \bar{\mathbf{y}}^* \mathbf{Q}, \qquad \mathbf{v} = \bar{\mathbf{v}}^* \mathbf{Q}.$$

Both $\bar{\mathbf{v}}$ and $\mathbf{v}$ are suitable definitions of an adjoint vector, which can be distinguished by naming $\bar{\mathbf{v}}$ the *hermitian adjoint*, and are easily converted to one another where both apply. The hermitian adjoint definition is convenient when the emphasis is on self-adjoint problems (§8.1), since a mapping between the state space and its dual (and the choice of a specific $\mathbf{Q}$) is needed for such a notion to make sense. On the other hand the definition, used everywhere in this article, of adjoint equation as the one yielding the sensitivity $\mathbf{v}$ applies to a larger class of linear problems (including those that are not posed in a Hilbert space) and naturally extends to nonlinear ones (§12). In computation it dispenses with matrix $\mathbf{Q}$ and a few complex conjugations.

More generally, one may be interested in computing P objective scalars J_p for Q different input vectors $\mathbf{u}_q$ (with both $P \leq M$ and $Q \leq N$ for them to be linearly independent). The important balance to be kept in mind is that the advantage of using an adjoint depends on whether there are more inputs or outputs. In this example it's easy to work out that if $P/M < Q/N$, computing the adjoint first is faster, whereas if $P/M > Q/N$ the direct system gives a quicker answer.

The adjoint advantage: with fewer inputs than outputs direct computation is faster, with fewer outputs than inputs adjoint computation is.

3 What are adjoints used for?

As the previous example has shown, adjoint equations yield sensitivity fields. In the following sections we shall consider several variations of discrete and continuous systems and define an adjoint problem for each. The common thread will be that for every linear problem one can, at least in principle, define an input-output relationship like (1) mediated by a matrix (or, in a continuous setting, by an operator) $\mathbf{H}$, named its *resolvent*, and all the adjoint problems we shall define are, in this sense, equivalent to (3). However, explicitly calculating the resolvent may not always be a practical thing to do. Just as replacing (1)–(2) by (3)–(4) involves a computational economy when fewer outputs than inputs are used, calculating the adjoint of each individual step that defines the system is generally more efficient than calculating an all-encompassing resolvent and applying (1)–(3) to it. This will become clearer as we add more examples along the way.

The adjoint formulation is useful when one is seeking to obtain one or a few outputs of a system for a wide range of possible inputs. There are several such cases in fluid mechanics (and other disciplines), in particular in stability theory as will be expounded in this article, but the greatest advantage is obtained in optimization. In fact, the typical optimization problem has a single objective function (possibly combining multiple objectives through suitable weights) that has to be minimized or maximized with respect to a large number, or even a continuum, of input variables: a perfect application for adjoints!

4 Adjoint of a system of linear algebraic equations

Let us now examine an instantaneous linear system where the relationship between state $\mathbf{x}$ and input $\mathbf{u}$ is implicit, namely of the form

$$\mathbf{A}\mathbf{x} = \mathbf{u} \tag{5}$$

with $\mathbf{A}$ denoting a non-singular $M \times M$ square matrix. One obvious way to devise an adjoint of this problem is to convert (5) to (1) by using the inverse $\mathbf{H} = \mathbf{A}^{-1}$ and then proceed as in §2; however a more powerful alternative is to write the objective J defined by (2) in the form (4), with an as yet unknown coefficient vector $\mathbf{v}$, and then substitute (5) into it:

$$J = \mathbf{v}\mathbf{u} = \mathbf{v}\mathbf{A}\mathbf{x} = \mathbf{y}\mathbf{x}. \tag{6}$$

It is evident by inspection that this becomes an identity for any $\mathbf{x}$ (and consequently for any $\mathbf{u}$) if

$$\mathbf{A}^{\mathrm{T}}\mathbf{v} = \mathbf{y}, \tag{7}$$

which deserves to be named the adjoint system of (5). The advantage is once again computational: obtaining $\mathbf{v}$ from (7) only requires the solution of a linear system with a specific r.h.s., which is classically known to be a less costly numerical algorithm than constructing the inverse $\mathbf{H} = \mathbf{A}^{-1}$ (and even more so if $\mathbf{A}$ happens to be a sparse or banded matrix).

The adjoint of a system of linear equations also finds a useful role in defining the *compatibility condition* that must be satisfied when the coefficient matrix is singular, for the system to admit a solution: if $\mathbf{A}$'s determinant is zero, both the direct and transposed systems have a nonzero homogeneous solution, as is known from linear algebra, and system (5) only admits solutions under a suitable compatibility condition. If $\mathbf{v}$ is any nonzero solution of the homogeneous adjoint system, $\mathbf{vA} = 0$, contracting both sides of (5) with $\mathbf{v}$ gives

$$\mathbf{vu} = 0 \tag{8}$$

which is thus a nontrivial necessary condition on $\mathbf{u}$ for the solution $\mathbf{x}$ to exist.

Compatibility condition: for an inhomogeneous singular linear system to admit a solution, the product of its r.h.s. and its adjoint homogeneous solution must be zero.

5 Entering dynamics: the adjoint of a discrete-time dynamical system

Let us now complicate matters slightly by adding dynamics (i.e., time dependence) in its simplest form, that of a discrete-time dynamical system described by the state-transition equation

$$\mathbf{x}_{n+1} = \mathbf{A}_n\mathbf{x}_n \tag{9}$$

with state dimension M, $M \times M$ transition matrix $\mathbf{A}_n$, initial condition $\mathbf{x}_0$, discrete time n running from 0 to $N-1$, and an objective depending linearly on $\mathbf{x}_N$:

$$J = \mathbf{y}\mathbf{x}_N. \tag{10}$$

Discrete-time dynamical system: system characterized by a numerable sequence of states each of which is completely determined by its predecessor.

The direct computation of J involves $\mathrm{O}(NM^2)$ flops for (9) and $\mathrm{O}(M)$ flops for (10). We can also unroll the process in the form of the explicit product chain

$$J = \mathbf{y}(\mathbf{A}_{N-1}(\mathbf{A}_{N-2}(\cdots \mathbf{A}_1(\mathbf{A}_0\mathbf{x}_0)))).$$

Now, let's associate the products to the left rather than to the right:

$$J = ((((\mathbf{y}\mathbf{A}_{N-1})\mathbf{A}_{N-2}\cdots)\mathbf{A}_1)\mathbf{A}_0)\mathbf{x}_0.$$

Clearly we have again obtained a compact expression of the objective as a function of the input, namely the linear form $J = \mathbf{v}_0\mathbf{x}_0$ where

$$\mathbf{v}_n = \mathbf{A}_n^{\mathrm{T}}\mathbf{v}_{n+1} \tag{11}$$

and $\mathbf{v}_N = \mathbf{y}$. Equation (11) defines for all purposes a new dynamical system, which can be denoted as adjoint of (9), with the peculiarity that discrete time n runs backwards from $N-1$ down to 0, and that instead of an initial it has a "terminal" condition $\mathbf{v}_N = \mathbf{y}$. Notice also that with no difficulty the transition matrix $\mathbf{A}$ has been allowed to change with n; time-invariant systems are of course included as a special case.

Just as in §2, the computation of the objective J for several different initial conditions $\mathbf{x}_0$ may be obtained by applying a single sequence of (11), at the same cost as a single sequence of (9), and then calculating $J = \mathbf{v}_0\mathbf{x}_0$ at a cost of $\mathrm{O}(M)$ flops for each initial condition. This procedure only involves matrix-vector products, and is much less computationally expensive than calculating the resolvent $\mathbf{H} = \mathbf{A}_{N-1}\mathbf{A}_{N-2}\cdots\mathbf{A}_1\mathbf{A}_0$ through matrix-matrix products. In addition, the same value of J repeatedly appears at every intermediate step of (9) and (11):

$$J = \mathbf{v}_n\mathbf{x}_n = \text{const.} \quad \text{for } n = 0\ldots N; \tag{12}$$

this property turns up very useful when testing an adjoint computer program (§14).

General dynamical system with a time-varying input term and an objective depending on the state at all times.

The evolution equation of a more general dynamical system with a time-varying input term $\mathbf{u}_n$ weighted by an input matrix $\mathbf{B}_n$ may be written as

$$\mathbf{x}_{n+1} = \mathbf{A}_n\mathbf{x}_n + \mathbf{B}_n\mathbf{u}_n \qquad (n = 0\ldots N-1). \tag{13}$$

Correspondingly, a more general objective depending on the state at all times is

$$J = \sum_{n=0}^{N}\mathbf{y}_n\mathbf{x}_n. \tag{14}$$

Obtaining an adjoint for this more general case is no more difficult than in the previous example. It only requires recursively eliminating (in the sense of Gaussian elimination) each $\mathbf{x}_{n+1}$ for $n = N-1\ldots 0$. First, inserting $\mathbf{x}_N$ from (13) into (14) gives

$$J = \mathbf{y}_N(\mathbf{A}_{N-1}\mathbf{x}_{N-1} + \mathbf{B}_{N-1}\mathbf{u}_{N-1}) + \sum_{n=0}^{N-1}\mathbf{y}_n\mathbf{x}_n =$$

$$= \mathbf{v}_N\mathbf{B}_{N-1}\mathbf{u}_{N-1} + \mathbf{v}_{N-1}\mathbf{x}_{N-1} + \sum_{n=0}^{N-2}\mathbf{y}_n\mathbf{x}_n$$

where $\mathbf{v}_N = \mathbf{y}_N$ and $\mathbf{v}_{N-1} = \mathbf{A}_{N-1}^{\mathrm{T}}\mathbf{v}_N + \mathbf{y}_{N-1}$. By recursion

$$J = \sum_{n=0}^{N-1}(\mathbf{v}_{n+1}\mathbf{B}_n)\mathbf{u}_n + \mathbf{v}_0\mathbf{x}_0 \tag{15}$$

where

$$\mathbf{v}_n = \mathbf{A}_n^{\mathrm{T}}\mathbf{v}_{n+1} + \mathbf{y}_n \qquad (n = N-1\ldots 0). \tag{16}$$

This is the equation adjoint to (13).

An alternate way to derive the same result is to observe that (13) and (14) may be formally reduced to a particular case of (9) and (10) by defining an *augmented state* that includes in a single much larger vector the components of all states and all inputs at all times. If sparseness of the transition matrices is suitably taken into account, (15) and (16) are obtained again.

6 From discrete to continuous time

The adjoint of a system described through differential equations can also be described through differential equations, and the constructive procedure discovered in his time by Lagrange allows us to derive such adjoint equations.

Continuous-time dynamical system and objective.

Let us consider the linear, continuous-time dynamical system

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{17}$$

with initial condition $\mathbf{x}(0) = \mathbf{x}_0$ and objective

$$J = \int_0^T \mathbf{y}(t)\mathbf{x}(t)\,\mathrm{d}t. \tag{18}$$

Useful information about the sensitivity of J to the initial condition and input may be gathered by looking at a discrete approximation of (17).

The most basic (first-order, explicit) numerical discretization of (17) with a stepsize $\Delta t = T/N$ is a discrete dynamical system of the general form (13), namely

$$\mathbf{x}_{n+1} = (1 + \Delta t\,\mathbf{A})\,\mathbf{x}_n + \Delta t\,\mathbf{B}\mathbf{u}_n. \tag{19}$$

The corresponding adjoint equation, according to §5, is

$$\mathbf{v}_n = (1 + \Delta t\,\mathbf{A}^{\mathrm{T}})\,\mathbf{v}_{n+1} + \Delta t\,\mathbf{y}_n, \tag{20}$$

to be marched backwards in n from the terminal condition $\mathbf{v}_N = 0$. The look of equation (20) suggests that it might be the numerical discretization of the new differential equation

$$-\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \mathbf{A}^{\mathrm{T}}\mathbf{v} + \mathbf{y}. \tag{21}$$

That this is indeed the adjoint of (17) can be proved in a similar manner as (6) was derived. We first build a linear form by contracting both sides of (17) with an as yet unknown vector-valued function $\mathbf{v}(t)$ and integrating:

$$\int_0^T \mathbf{v}\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\,\mathrm{d}t = \int_0^T \mathbf{v}\,(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u})\,\mathrm{d}t.$$

This identity may then be led to only contain $\mathbf{x}$ and not its derivative through an integration by parts:

$$\Big[\mathbf{v}\mathbf{x}\Big]_0^T - \int_0^T \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\,\mathbf{x}\,\mathrm{d}t = \int_0^T \mathbf{v}\,(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u})\,\mathrm{d}t, \tag{22}$$

whence, by collecting the coefficients of $\mathbf{x}$ and comparing with (18),

$$J = \int_0^T \left(-\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} - \mathbf{v}\mathbf{A}\right)\mathbf{x}(t)\,\mathrm{d}t = \mathbf{v}(0)\,\mathbf{x}(0) + \int_0^T \mathbf{v}\mathbf{B}\mathbf{u}\,\mathrm{d}t \tag{23}$$

provided $\mathbf{v}$ is chosen to be a solution of (21) with the terminal condition $\mathbf{v}(T) = 0$. That differential equation for $\mathbf{v}$, to be integrated backwards in time just like (20), thus provides the explicit expression of the objective as a linear function of the initial condition and input data and therefore deserves to be called the adjoint equation of (17).

Remarkably the adjoint equation is endowed with a similar structure and computational difficulty as the original direct problem, as clearly appears from the comparison of the corresponding discretizations (19) and (20). Just like the discrete systems of the previous section, (21) must be solved by marching backwards in time with the terminal condition[2] $\mathbf{v} = 0$ at $t = T$, in order to render the expression (23) of J independent of $\mathbf{x}(T)$ and a function of $\mathbf{x}(0)$ and $\mathbf{u}$ only. In addition, the relationship between the direct and adjoint problem's eigenvalues, which will be seen in §8, is such that if the direct problem is well-posed in forward time (for instance, it is the semi-discretization of a parabolic problem with a forward direction of stable evolution), the adjoint problem is well-posed in backward time.

In the context of stability and receptivity analysis, early applications of the adjoint equation (21) to a parabolic problem can be found in Luchini & Bottaro (1996, 1998), who calculated the receptivity of the Görtler instability, and Airiau et al. (2002), who developed an adjoint parabolized stability equation.

7 Connection between adjoint and resolvent

The solution of the linear system (17), with initial condition $\mathbf{x}(0) = \mathbf{x}_0$ and input $\mathbf{u}(t)$, may always be formally written as a linear combination of homogeneous solutions, i.e. in matrix notation,

$$\mathbf{x}(t) = \mathbf{H}(t,0)\mathbf{x}_0 + \int_0^t \mathbf{H}(t,\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\,\mathrm{d}\tau \tag{24}$$

where $\mathbf{H}(t,\tau)$ is the solution in the $[\tau, t]$ interval of the homogeneous matrix differential equation

Resolvent, also known as Green's function or propagator.

$$\frac{\mathrm{d}\mathbf{H}}{\mathrm{d}t} = \mathbf{A}(t)\mathbf{H}(t,\tau) \tag{25}$$

with initial condition[3] $H_{ij} = \delta_{ij}$ at $t = \tau$. $\mathbf{H}(t,\tau)$ is by definition the resolvent introduced in §3 (also known as *Green's function* or *propagator*) of (17).

Inserting (24) into the definition of the objective (18) (while for simplicity assuming $\mathbf{x}_0 = 0$ and $\mathbf{H}(t,\tau) = 0$ for $t < \tau$) gives

$$J = \int_0^T \int_0^T \mathbf{y}(t)\mathbf{H}(t,\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\,\mathrm{d}\tau\,\mathrm{d}t;$$

reversing the order of integration and comparing to (23) then yields

$$\mathbf{v}(\tau) = \int_\tau^T \mathbf{y}(t)\mathbf{H}(t,\tau)\,\mathrm{d}t, \tag{26}$$

account having been taken that $\mathbf{H}(t,\tau) = 0$ for $t < \tau$. It may be noticed that (24) is analogous to (1), and (26) to (3), of §2.

[2] A nonzero terminal condition may optionally be introduced to account for the presence of an additional term proportional to $\mathbf{x}(T)$ in (18). Alternately, $\mathbf{v}(T)$ may be assumed to be always zero and such a term included in $\mathbf{y}$ as a delta function.

[3] δ_{ij} being Kronecker's symbol, i.e. 1 for $i = j$ and 0 for $i \neq j$.

It follows that the transposed resolvent $\mathbf{H}^{\mathrm{T}}(t,\tau)$, if regarded as a function of its second argument τ in the $[0,t]$ interval, is the resolvent of the adjoint problem and obeys the homogeneous adjoint differential equation

$$-\frac{\mathrm{d}\mathbf{H}^{\mathrm{T}}}{\mathrm{d}\tau}=\mathbf{A}^{\mathrm{T}}(\tau)\mathbf{H}^{\mathrm{T}}(t,\tau), \tag{27}$$

to be integrated backwards from $\tau=t$ to $\tau=0$ with the terminal condition $H^{\mathrm{T}}_{ij}=\delta_{ij}$ at $\tau=t$.

As an easy to remember conclusion, the resolvent obeys in symmetrical fashion the homogeneous direct equation as a function of final time t and the homogeneous adjoint equation as a function of initial time τ. However, as the previous examples have shown, the resolvent matrix is mostly a useful theoretical tool whereas calculating either the direct or the adjoint vector is computationally faster.

7.1 The resolvent of a time-invariant problem

When $\mathbf{A}$ is independent of time, the resolvent is classically represented as[4]

$$\mathbf{H}(t,\tau)=\exp\left[\mathbf{A}\ (t-\tau)\right]\equiv\sum_{m=1}^{M}e^{\sigma_m\,(t-\tau)}\,\mathbf{u}_m\otimes\mathbf{v}_m \tag{28}$$

where σ_m are $\mathbf{A}$'s eigenvalues (assumed to be distinct for simplicity here) and $\mathbf{u}_m$, $\mathbf{v}_m$ are $\mathbf{A}$'s right and left eigenvectors. That (28) is a solution of both (25) and (27) should be self-evident, as σ_m are the eigenvalues of $\mathbf{A}$ and $-\sigma_m$ the eigenvalues of $-\mathbf{A}^{\mathrm{T}}$.

The sum in (28) can be cut down to a single dominant mode when the latter is exponentially amplified. As such, it plays an important role in the temporal receptivity analysis of global modes (Hill 1992) and in the spatial receptivity analysis of parallel flows (Fedorov 1984, Hill 1995, Tumin & Fedorov 1984). Its extension through an inhomogeneous multiple-scale WKBJ approximation provides the receptivity of quasi-parallel flows (Zuccher 2002, Zuccher & Luchini 2013). It must be remarked that this single-mode approximation (very common in the investigation of boundary-layer instabilities) is the only example in this review where computing the resolvent is computationally faster than resorting to adjoint equations.

8 Adjoint of a boundary-value problem or of an eigenvalue problem

It may have been noticed that the finite term arising from the integration by parts in (22) contains exactly twice as many addends as there are initial conditions for $\mathbf{x}$, and that the terms not used by such initial conditions have been eliminated by choosing the terminal conditions for $\mathbf{v}$ appropriately. In fact this is a rule: for a more general boundary-value problem (b.v.p.) specifying P conditions at $t=0$ and $M-P$ conditions at $t=T$ (where the independent variable t does not necessarily have to represent time) it is always possible to choose $M-P$ conditions at $t=0$ and P conditions at $t=T$ for $\mathbf{v}$ in such a way that

Adjoint of a boundary-value problem: specify $M-P$ conditions at one end and P conditions at the other end in such a way that the explicit expression of the objective only contains the original M boundary data.

[4] $\otimes$ denoting the dyadic product of two vectors to produce a matrix.

(23) only contains the original boundary data. Therefore the adjoint of a b.v.p. will be another b.v.p.

An additional complication is that the solution of a b.v.p., contrary to the initial-value problem to which Picard's classical existence and uniqueness theorem applies, cannot *a priori* be granted to exist and/or be unique. Indeed another fruitful application of adjoints, although one that will not be addressed here, is precisely in devising existence and uniqueness proofs for such solutions.

If the b.v.p. is homogeneous and contains a parameter σ, it becomes an eigenvalue problem, one that generally admits only the null (trivial) solution except for special values of σ which are named its *eigenvalues.* Its adjoint problem (equation and boundary conditions) can be determined just as for any other b.v.p., and will depend on σ too. Under reasonable regularity conditions (essentially requiring that a so called *normal form* — a form explicit in the state derivatives such as (17) — exists and is not singular, e.g. Tricomi (1957)), differential eigenvalue problems share the same properties as matrix eigenvalue problems. In particular, the direct and adjoint problem have the same eigenvalues and mutually orthogonal eigenfunctions.

For a (generalized) matrix eigenvalue problem of the form

$$\mathbf{A}\mathbf{x} = \sigma\mathbf{B}\mathbf{x}, \tag{29}$$

with adjoint

$$\mathbf{v}\mathbf{A} = \sigma\mathbf{v}\mathbf{B}, \tag{30}$$

the existence of a nontrivial solution of (30) for the same σ where one of (29) exists is a consequence of the fact that the direct and transposed matrix have the same determinant. Orthogonality is classically proven by considering an eigenvector $\mathbf{x}_1$ of (29) corresponding to an eigenvalue σ_1 and an adjoint eigenvector $\mathbf{v}_2$ of (30) corresponding to a different eigenvalue σ_2. Contracting (29) with $\mathbf{v}_2$ and (30) with $\mathbf{x}_1$, and subtracting, gives

$$\mathbf{v}_2(\mathbf{A}\mathbf{x}_1) - (\mathbf{v}_2\mathbf{A})\mathbf{x}_1 = \mathbf{v}_2(\sigma_1\mathbf{B}\mathbf{x}_1) - (\sigma_2\mathbf{v}_2\mathbf{B})\mathbf{x}_1$$

i.e.

$$0 = (\sigma_1 - \sigma_2)\mathbf{v}_2\mathbf{B}\mathbf{x}_1$$

whence $\mathbf{v}_2\mathbf{B}\mathbf{x}_1 = 0$ (by a similar token, also $\mathbf{v}_2\mathbf{A}\mathbf{x}_1 = 0$) follows unless $\sigma_1 = \sigma_2$. Furthermore, if the eigenvectors form a complete set (a *basis*) and $\mathbf{v}_i\mathbf{B}\mathbf{x}_i \neq 0$ for all $i = 1 \dots M$, an arbitrary vector $\mathbf{x}$ can be expanded as

$$\mathbf{x} = \sum_{i=1}^{M} \frac{\mathbf{v}_i\mathbf{B}\mathbf{x}}{\mathbf{v}_i\mathbf{B}\mathbf{x}_i}\mathbf{x}_i. \tag{31}$$

Projector: the adjoint eigenvector multiplied by **B** projects an arbitrary vector onto the corresponding direct eigenvector.

This formula may be read to say that, with proper normalization, $\mathbf{v}_i\mathbf{B}$ is the *projector* that extracts the component of $\mathbf{x}$ along $\mathbf{x}_i$.

For a differential eigenvalue problem in normal form, similar properties are proven by first reformulating it as a Fredholm integral equation governed by a compact operator (this is where the normal form is required) and then showing that a compact operator can be uniformly approximated by a matrix to any desired precision. In addition, if the problem is self-adjoint (see below) the set of eigenvectors is complete (Courant & Hilbert 1924, Tricomi 1957). Notice, however, that if boundary conditions are given over an infinite interval (the operator is no longer compact) completeness may be lost or may involve continuous ranges of eigenvalues (a *continuous spectrum*).

8.1 Self-adjoint problems

When equations (29) and (30) coincide (or, for complex-valued matrix elements, when the second is the complex conjugate of the first, denoted by a * superscript), the problem is qualified *self-adjoint*. This occurs when both matrices $\mathbf{A}$ and $\mathbf{B}$ are hermitian: $\mathbf{A}^{\mathrm{T}} = \mathbf{A}^*$ and $\mathbf{B}^{\mathrm{T}} = \mathbf{B}^*$. Indeed, in many differential-equations and mathematical-physics textbooks adjoint problems are introduced mostly in preparation for the special properties of self-adjoint ones, namely:

Self-adjoint problem: problem that coincides with its adjoint in a suitable inner-product space.

1. for each eigenvector $\mathbf{x}_n$, $\mathbf{v}_n = \mathbf{x}_n^*$;
2. for each eigenvalue σ_n, $\sigma_n = \sigma_n^*$ (all eigenvalues are real);
3. with suitable normalization, eigenvectors are orthonormal with respect to the hermitian inner product $\mathbf{x}_m^* \mathbf{B} \mathbf{x}_n = \delta_{mn}$ (in the case of functional spaces, to the $\mathcal{L}_2$ inner product $\int_0^T \mathbf{x}_m^* \mathbf{B} \mathbf{x}_n \,\mathrm{d}t = \delta_{mn}$);
4. if $\mathbf{B}$ is non-singular, the set of eigenvectors always forms a basis.

For instance, every second-order linear differential equation can be put in a *Sturm-Liouville* form (Tricomi 1957), which makes it self-adjoint provided the inner product is defined through a suitable weight function.

Here we shall mostly dispense with these properties, as this review purports to highlight the use of adjoints in the general non-self-adjoint case. A theoretical detail, however, has to be highlighted: whereas the definition of adjoint equation as the one providing a sensitivity only involves a linear form, the idea of a self-adjoint problem requires the additional existence of an inner product, so that direct and adjoint eigenfunctions live in the same Hilbert space and can be compared. The inner product may be defined in more than one way, e.g. with the intervention of a weight function as in Sturm-Liouville problems, and a problem that is self-adjoint under a certain inner product will not be so under another. Some problems may even turn out to be self-adjoint under an operation that lacks the algebraic properties of an inner product (an example will appear in §10.3).

9 Structural sensitivity of an eigenvalue problem

As a particular application of the adjoint, the objective whose sensitivity is sought may be chosen to be the eigenvalue itself. This will become useful every time the growth rate of an instability is being investigated as a function of the physical parameters of interest.

In the context of the finite-dimensional system (29) we can consider how the eigenvalue changes under an infinitesimal perturbation $\mathrm{d}\mathbf{A}$ of matrix $\mathbf{A}$. This is named a *structural perturbation*, meaning a perturbation of the *structure* (the coefficients) rather than the forcing (the r.h.s.) of a linear system. Taking the differential of both sides of (29) gives

Structural perturbation: perturbation of the coefficients rather than the forcing of a linear system.

$$\mathrm{d}[(\mathbf{A} - \sigma\mathbf{B})\,\mathbf{x}] = (\,\mathrm{d}\mathbf{A} - \,\mathrm{d}\sigma\mathbf{B})\,\mathbf{x} + (\mathbf{A} - \sigma\mathbf{B})\,\mathrm{d}\mathbf{x} = 0.$$

The eigenvector perturbation $\mathrm{d}\mathbf{x}$ can be algebraically eliminated from this equation by contracting it with the adjoint eigenvector $\mathbf{v}$. Recalling (30) in fact provides

$$\mathrm{d}\sigma = \frac{\mathbf{v}\ \mathrm{d}\mathbf{A}\,\mathbf{x}}{\mathbf{v}\,\mathbf{B}\,\mathbf{x}}. \tag{32}$$

The reader may notice that this formula is also encountered in numerical matrix algebra (Golub & Van Loan 1983), where it is used to define the condition number of an eigenvalue through its sensitivity to roundoff errors.

If the perturbation $\mathrm{d}\mathbf{A}$ has only one nonzero element, say $\mathrm{d}A_{mn}$, (32) can be further simplified to give

$$\mathrm{d}\sigma = \frac{v_m x_n}{\mathbf{v}\,\mathbf{B}\,\mathbf{x}}\ \mathrm{d}A_{mn}. \tag{33}$$

This expression shows that the sensitivity of the eigenvalue to an infinitesimal disturbance $\mathrm{d}A_{mn}$ of an individual matrix element is proportional to the product between the components x_n and v_m of, respectively, the direct and adjoint eigenvector.

Equation (32) is the basis for the concepts of base-flow sensitivity and optimal base-flow perturbation of a stability problem (Bottaro et al. 2003). Equation (33) was used to introduce the structural-sensitivity map of a global mode as a means to determine its *wavemaker* (Giannetti & Luchini 2003, 2007).

10 Adjoints in optimization and control

The simplifying power of adjoints is not limited to linear objectives. They reappear naturally in the optimization of quadratic forms.

In a discrete setting a positive-definite quadratic objective function, defined as

$$J = \Re\left(\frac{1}{2}\mathbf{x}^*\mathbf{Q}\mathbf{x} - \mathbf{b}^*\mathbf{x}\right) \tag{34}$$

with arbitrary coefficients $\mathbf{b}$ and $\mathbf{Q}$ (where $\mathbf{Q}$ can be assumed to be hermitian with no loss of generality), is minimized when its differential is zero:

$$\mathrm{d}J = \Re\left(\mathbf{y}\,\mathbf{dx}\right) = 0 \quad \text{where} \quad \mathbf{y} = \left(\mathbf{Q}\mathbf{x} - \mathbf{b}\right)^*.$$

When the system is instantaneous, i.e. responds with no time delay and its state doubles up as the variable to be optimized, the minimum of (34) can be attained by, either directly or iteratively, inverting the linear system $\mathbf{Q}\mathbf{x}-\mathbf{b} = 0$. For the iterative solution of this optimization problem a number of classical *gradient-based* iteration algorithms have been developed (gradient, conjugate gradient, BFGS, etc., see e.g. Press et al. (1999)) which rely on the sensitivity $\mathbf{y}$ (another name of which is the *gradient*) being computed at each iteration.

When the system is dynamic, one can in principle replace it by its resolvent and proceed in the same way, or exploit the fact that computing a sensitivity is exactly what adjoint equations are best at. If, for instance, the state $\mathbf{x}$ is indirectly determined by the discrete-time dynamical system of §5 and J is a quadratic form in the final state $\mathbf{x}_N$, the objective (10) becomes replaced by

$$\mathrm{d}J = \Re\left(\mathbf{y}\,\mathbf{dx}_N\right) \quad \text{with} \quad \mathbf{y} = \left(\mathbf{Q}\mathbf{x}_N - \mathbf{b}\right)^*. \tag{35}$$

The calculation of $\mathbf{v}$ from this $\mathbf{y}$ proceeds exactly as in §5, but a new $\mathbf{y}$ must be considered every time $\mathbf{x}_N$ changes. The result is what is often called *direct–adjoint iteration*:

1. Start with a (possibly zero) guessed value for the initial condition $\mathbf{x}_0$.
2. Perform a *direct* solution of system (9) from time 0 to time N.
3. Calculate $\mathbf{y}$ from $\mathbf{x}_N$ according to (35).
4. Perform an *adjoint* solution of system (11) from $\mathbf{v}_N = \mathbf{y}$ at time N to get the sensitivity (the gradient) $\mathbf{v}_0$ at time 0.
5. Improve the value of $\mathbf{x}_0$ using $\mathbf{v}_0$ with one of the gradient-based optimization algorithms and loop back to step 2 until convergence.

Direct–adjoint iteration: to alternate solutions of the direct and adjoint problem in an iterative optimization loop.

A similar direct–adjoint iteration can be applied to optimize an objective depending on the state at all times of any of the discrete or continuous problem types considered in §2–11. Since the adjoint problem has the same computational cost as the direct problem, each iteration costs twice as much as a single direct solution with specified initial conditions and forcing.

By contrast, the alternate route of preliminarily calculating a resolvent matrix and then optimizing requires a number of direct-solution times equal to the dimension M of $\mathbf{x}_0$. The balance falls in favour of direct–adjoint iteration when the expected number of iterations to convergence is smaller than $M/2$.

10.1 Optimal perturbation

In investigating the transient growth of boundary-layer disturbances, a key role is played by the initial condition that maximizes the energy amplification of a system of the general form (9), i.e. an objective defined as $J = E_{\text{out}}/E_{\text{in}}$ where the input and output energies are quadratic forms $E_{\text{in}} = \mathbf{x}_0^*\mathbf{Q}_{\text{in}}\mathbf{x}_0$ and $E_{\text{out}} = \mathbf{x}_N^*\mathbf{Q}_{\text{out}}\mathbf{x}_N$ characterized by two, possibly different, hermitian matrices $\mathbf{Q}_{\text{in}}$ and $\mathbf{Q}_{\text{out}}$. The extremality condition for this Rayleigh ratio (or equivalently for E_{out} under the constraint $E_{\text{in}} = 1$) may be written in terms of the resolvent $\mathbf{H}$ as

$$\mathbf{H}^{\text{H}}\mathbf{Q}_{\text{out}}\mathbf{H}\mathbf{x}_0 = J\,\mathbf{Q}_{\text{in}}\mathbf{x}_0 \tag{36}$$

(where superscript $^{\text{H}}$ denotes the hermitian transpose, i.e. the complex-conjugate transpose). Since both $\mathbf{H}^{\text{H}}\mathbf{Q}_{\text{out}}\mathbf{H}$ and $\mathbf{Q}_{\text{in}}$ are hermitian matrices, this can be recognized to be a self-adjoint eigenvalue problem with all the properties listed in §8.1, in particular with all real eigenvalues of J. Therefore its maximum eigenvalue and corresponding optimal initial condition can be determined by a simple power iteration:

$$\tilde{\mathbf{x}}_0 = \mathbf{Q}_{\text{in}}^{-1}\mathbf{H}^{\text{H}}\mathbf{Q}_{\text{out}}\mathbf{H}\mathbf{x}_0^{(i)}; \quad J = \sqrt{\tilde{\mathbf{x}}_0^*\mathbf{Q}_{\text{in}}\tilde{\mathbf{x}}_0}; \quad \mathbf{x}_0^{(i+1)} = \tilde{\mathbf{x}}_0/J \tag{37}$$

(where superscript $^{(i)}$ counts iterations and $\tilde{\mathbf{x}}_0$ is an intermediate variable). In practice, (37) can be implemented in the form of a direct–adjoint iteration without explicit use of the resolvent, if only it is recognized that multiplication by $\mathbf{H}$ is equivalent to one application of the direct and multiplication by $\mathbf{H}^{\text{H}}$ to one application of the adjoint equations.

Direct-adjoint iteration in the form (37) was instrumental to the computation of optimal inflow disturbances for a Blasius boundary layer by Andersson et al. (1998, 1999) and Luchini (1997, 2000), who achieved the quantitative calculation of non-parallel spatial transient growth in such a flow (a parallel approximation of which had been studied by Butler & Farrell (1992)). The more general gradient-based algorithm was used in the open-loop optimal control of linear (Cathalifaud & Luchini 2000) and nonlinear (Zuccher et al. 2004) transient growth through blowing and suction at the wall.

10.2 Optimal control

In another example, the optimization at hand might be a closed-loop optimal-control problem. In the so-called *linear-quadratic regulator* (LQR) setting (e.g. Kim & Bewley (2007)), this is usually formulated as follows:

LQR optimal control problem: For every given initial condition find the input that minimizes a weighted sum of the state and control energies.

> For a dynamical system of equation (17), and a given initial condition $\mathbf{x}(0) = \mathbf{x}_0$, find the input $\mathbf{u}(t)$ that minimizes the quadratic form $J = \frac{1}{2}\int_0^T (\mathbf{x}^*\mathbf{Q}\mathbf{x} + \mathbf{u}^*\mathbf{R}\mathbf{u})\,\mathrm{d}t$

(where $\mathbf{Q}$ and $\mathbf{R}$ can be assumed to be hermitian with no loss of generality). Considering the first variation of the objective,

$$\delta J = \int_0^T \Re\left(\mathbf{x}^*\mathbf{Q}\boldsymbol{\delta}\mathbf{x} + \mathbf{u}^*\mathbf{R}\boldsymbol{\delta}\mathbf{u}\right)\,\mathrm{d}t,$$

and applying (21)–(23) with $\mathbf{y} = \mathbf{x}^*\mathbf{Q}$ yields the adjoint equation

$$-\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \mathbf{A}^{\mathrm{T}}\mathbf{v} + \mathbf{x}^*\mathbf{Q}. \tag{38}$$

and the explicit expression of the objective variation

$$\delta J = -\Re\Big[\mathbf{v}\boldsymbol{\delta}\mathbf{x}\Big]_0^T + \int_0^T \Re\left[(\mathbf{u}^*\mathbf{R} + \mathbf{v}\mathbf{B})\boldsymbol{\delta}\mathbf{u}\right]\,\mathrm{d}t.$$

The variation δJ is then made to vanish by: 1) letting $\boldsymbol{\delta}\mathbf{x}(0) = 0$ (since the initial condition $\mathbf{x}(0)$ is imposed); 2) letting $\mathbf{v}(T) = 0$ (the terminal condition for the backward integration of (38)); 3) letting

$$\mathbf{u} = -\mathbf{R}^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{v}^*. \tag{39}$$

Hamiltonian matrix: combining the direct and adjoint problem together, shares the same symmetry as the Hamilton equations of analytical mechanics.

At this stage there are two possibilities. One is to write the system of equations (17) (with $\mathbf{u}$ replaced by (39)) and (38) together as the $2M \times 2M$ linear problem

$$\frac{d}{dt}\begin{bmatrix}\mathbf{x}\\ \bar{\mathbf{v}}\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\mathrm{H}}\\ \mathbf{Q} & -\mathbf{A}^{\mathrm{H}}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \bar{\mathbf{v}}\end{bmatrix} \tag{40}$$

where $\bar{\mathbf{v}} = -\mathbf{v}^*$ and boundary conditions are $\mathbf{x}(0) = \mathbf{x}_0$, $\bar{\mathbf{v}}(T) = 0$. This boundary-value problem, which shares the same symmetry as the Hamilton equations of analytical mechanics and is therefore also termed *hamiltonian*, may then be dealt with *as is* or transformed into the so-called *matrix Riccati equation* (Kim & Bewley 2007). Its solution is classically obtained by applying the eigenvector expansion (28) of the resolvent, with the hamiltonian matrix in the place of $\mathbf{A}$. This is a standard procedure in control theory, for which easy to use computer libraries are available; however the resolvent matrix is not sparse, even when $\mathbf{A}$ is, and therefore its use may become impractical for very large state vectors such as obtained from the discretization of partial differential equations (§11).

An alternative is to apply direct–adjoint iteration, by repeatedly solving (17) forward in time with a tentative $\mathbf{u}$ and (38) backward in time for the sensitivity $\mathbf{u}^*\mathbf{R} + \mathbf{v}\mathbf{B}$, using then one of the gradient-based optimization algorithms to produce an improved $\mathbf{u}(t)$ for the next step (Pralits & Luchini 2010).

10.3 Adjoint of the direct–adjoint

The full purpose of an optimal-control problem is not so much to solve the optimization problem (40) for *one* initial condition $\mathbf{x}_0$, as to determine the matrix (*feedback kernel*) $\mathbf{K}$ that provides the feedback control $\mathbf{u}(0)$ for *any* $\mathbf{x}_0$ in the form of $\mathbf{u}(0) = \mathbf{K}\mathbf{x}_0$. It is sufficient for this purpose to obtain $\mathbf{u}$ at time $t = 0$, because $\mathbf{u}$ for a later time will be calculated from the state at that time. That such a $\mathbf{K}$ exists is ensured by the problem (40) being linear.

Notice that $\mathbf{K}$ is an $N \times M$ matrix, if M is the dimension of the state and N of the input. A possible direct way to determine $\mathbf{K}$ one column at a time is:

1. choose a basis of M independent M-vectors for $\mathbf{x}_0$ (e.g., the trivial basis $(1, 0, 0, \dots)$, $(0, 1, 0, \dots)$, $\dots$);

2. for each $\mathbf{x}_0$ solve (either directly or iteratively) (40) and extract the corresponding $\bar{\mathbf{v}}(0)$ (also an M-vector);

3. calculate $\mathbf{u}(0)$ (an N-vector) from (39) and store it in the appropriate column of $\mathbf{K}$.

The above procedure requires (40) to be solved M times. However if $N \ll M$ (a frequent occurrence), a much more efficient route exists: to calculate $\mathbf{K}$ one *row* at a time by applying adjoint analysis a second time over! Since (40) is already a direct–adjoint problem, we name this approach ADA (*adjoint of the direct–adjoint*) (Pralits & Luchini 2010).

At first thought it may seem like a strange idea to take the adjoint twice, but if the compounded hamiltonian problem (40) is regarded as just another linear system with $\mathbf{x}(0)$ as input and $\mathbf{u}(0)$ as output, the way to proceed becomes straightforward. Contracting each side of the direct system (40) with a new $2M$-vector which may be denoted $[\mathbf{x}^{\dagger^*}, \mathbf{v}^{\dagger^*}]$ and integrating by parts gives the adjoint system

$$\frac{d}{dt}\begin{bmatrix}\mathbf{x}^\dagger \\ \mathbf{v}^\dagger\end{bmatrix} = \begin{bmatrix}-\mathbf{A}^{\mathrm{H}} & -\mathbf{Q} \\ -\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\mathrm{H}} & \mathbf{A}\end{bmatrix}\begin{bmatrix}\mathbf{x}^\dagger \\ \mathbf{v}^\dagger\end{bmatrix} \tag{41}$$

(account having been taken that both $\mathbf{Q}$ and $\mathbf{R}$ are hermitian) and boundary conditions $\mathbf{v}^\dagger(0) = -\mathbf{y}^*$, $\mathbf{x}^\dagger(T) = 0$, where we have assumed an objective of the form $J = \mathbf{y}\bar{\mathbf{v}}(0)$. According to (39), we can now identify this objective with one selected component of $\mathbf{u}(0)$ by letting $-\mathbf{v}^{\dagger^*}(0)$ equal the corresponding row of $\mathbf{R}^{-1}\mathbf{B}^{\mathrm{H}}$. The solution of the adjoint hamiltonian problem will then provide the sensitivity $\mathbf{x}^{\dagger^*}(0)$ of this component of $\mathbf{u}(0)$ to the entire initial condition $\mathbf{x}(0)$, i.e. one row of $\mathbf{K}$.

In fact, the process becomes even simpler because (41) reduces easily to the same form as (40). (The hamiltonian problem is, in a suitable sense, self-adjoint, which is not so surprising because it already contains both direct and adjoint components.) The two problems become identical, as can be easily verified, under the *symplectic transformation* (well known from analytical mechanics, e.g. Goldstein (1980) p. 343)

$$\mathbf{v}^\dagger = -\mathbf{x}, \quad \mathbf{x}^\dagger = \bar{\mathbf{v}}.$$

Therefore, whatever algorithm is already available to solve the hamiltonian problem (40) it may be reused for its adjoint, by just assigning $\mathbf{y}^*$ as the initial condition for $\mathbf{x}(0)$ and taking $\bar{\mathbf{v}}^*(0)$ as the result.

ADA: adjoint of the direct–adjoint hamiltonian system.

In conclusion we arrive at the following ADA procedure:

1. choose a basis of N independent M-vectors for the row space of $\mathbf{R}^{-1}\mathbf{B}^{\mathrm{H}}$ (in practice, take each row of this matrix in turn);
2. using each M-vector as the initial condition $\mathbf{x}^*(0)$, solve (either directly or iteratively) (40) and extract the corresponding $\bar{\mathbf{v}}(0)$;
3. assign $\bar{\mathbf{v}}^*(0)$ (also an M-vector) to the corresponding row of $\mathbf{K}$.

This algorithm achieves to calculate $\mathbf{K}$ by rows, rather then by columns, using N (the dimension of the input) rather than M (the dimension of the state) independent solutions of (40).

The ADA algorithm is part of an ongoing effort to devise Riccati-less procedures to solve control problems, i.e. iterative algorithms that can exploit sparseness and do not require the solution of a matrix Riccati equation. While, in fact, the standard approach is adequate for problems with, say, a thousand state variables, the discretization of a fluid flow problem easily contains millions or hundreds of millions of variables, and a full matrix of a million by a million elements is not a viable tool. The frequently adopted route of model reduction (replacing the original problem by a model having a smaller number of state variables) in a sense defeats the purpose of optimization, unless the model itself is optimized which is a task of comparable difficulty to optimizing the controller. On the other hand, an iterative algorithm can deal with such problems without any model reduction. An overview of several suitable iterative approaches (one of which is ADA) is given by Bewley et al. (2013). The dual version of ADA for an estimation problem (every control problem has a dual estimation problem, see e.g. Kim & Bewley (2007)) is described under the acronym of AAD by Semeraro et al. (2013).

AAD Adjoint of the adjoint-direct hamiltonian problem.

11 The space continuum: adjoint of a partial differential equation

The procedure outlined in §6–8 generalizes unchanged to linear partial differential equations involving a function f of time (which may be renamed x_0) and H other (typically spatial) independent variables $x_1 \ldots x_H$. Formally, the equation (or system of equations) is multiplied by an as yet unknown function (or vector of such functions) v and integrated over the solution domain Ω. Subsequently each term involving a derivative of f is integrated by parts one or more times until the original unknown f can be factored out and only derivatives of v appear.

The multidimensional equivalent of integration by parts is known as the Lagrange-Green identity (e.g., John (1982) p. 79). For a single term of the form $a\,\partial b/\partial x_h$ this identity (a consequence of Gauss' theorem applied to the integral of $\partial(a\,b)/\partial x_h$) may be written as

Lagrange–Green identity.

$$\underbrace{\int_\Omega a\,\frac{\partial b}{\partial x_h}\,\mathrm{d}\Omega}_{1} = \underbrace{\oint_{\partial\Omega} a\,b\,n_h\,\mathrm{d}\partial\Omega}_{2} \underbrace{-\int_\Omega \frac{\partial a}{\partial x_h}\,b\,\mathrm{d}\Omega}_{3}\,. \tag{42}$$

where n_h is the h-th component of the outward normal versor $\mathbf{n}$. Repeatedly applying this transformation as many times as needed to each and every derivative in the partial differential equation

$$\mathcal{L}(f) = u \tag{43}$$

(where $\mathcal{L}$ is a differential operator representing a linear combination of partial derivatives of its argument) leads to an identity which can be formally written as

$$\int_\Omega v\mathcal{L}(f)\,\mathrm{d}\Omega = \oint_{\partial\Omega} \mathcal{M}(v,f,\mathbf{n})\,\mathrm{d}\partial\Omega + \int_\Omega \mathcal{L}^\dagger(v)f\,\mathrm{d}\Omega \tag{44}$$

where the boundary integral collects all terms denoted as type 2 in (42) ($\mathcal{M}$ being linear in each of its three arguments) and $\mathcal{L}^\dagger$ (also a differential operator representing, like $\mathcal{L}$, a linear combination of partial derivatives of its argument) all terms of type 3.

Finally, replacing $\mathcal{L}(f)$ by u and equating the last term of (44) to the objective

$$J = \int_\Omega yf\,\mathrm{d}\Omega \tag{45}$$

yields the adjoint differential equation

$$\mathcal{L}^\dagger(v) = y \tag{46}$$

and the explicit expression of the objective as a function of the problem's r.h.s. and boundary data:

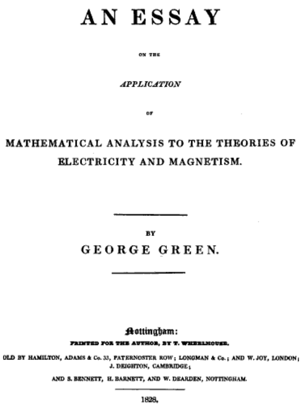
$$J = \int_\Omega vu\,\mathrm{d}\Omega - \oint_{\partial\Omega} \mathcal{M}(v,f,\mathbf{n})\,\mathrm{d}\partial\Omega. \tag{47}$$

The only peculiar difficulty (which is sometimes turned into an advantage) arises in specifying suitable functional spaces to which f and v must belong in order to ensure that all intermediate expressions exist. Algebra defines the concept of *dual* space as the vector space of all linear forms over a given vector space. Therefore, *a priori* f must belong to a (typically Sobolev) space regular enough that all required derivatives exist and v to its dual space. This is referred to as the *strong formulation* of the differential problem. However, after the integration by parts only derivatives of v appear; v is then required to be sufficiently regular whereas regularity requirements on f can be relaxed. The transformation to adjoint form allows the differential problem to be defined also for functions which are not as highly differentiable, and provides what is called its *weak formulation.*

Dual space: vector space of all linear forms over a given space.

11.1 The example of Green's formula

George Green's privately published comprehensive essay



AN ESSAY

ON THE

APPLICATION

OF

MATHEMATICAL ANALYSIS TO THE THEORIES OF ELECTRICITY AND MAGNETISM.

BY

GEORGE GREEN.

Nottingham:
PRINTED FOR THE AUTHOR, BY T. WHEELHOUSE.
OLD BY HAMILTON, ADAMS & Co. 33, PATERNOSTER ROW; LONGMAN & Co.; AND W. JOY, LONDON;
J. DEIGHTON, CAMBRIDGE;
AND S. BENNETT, H. BARNETT, AND W. DEARDEN, NOTTINGHAM.

1828.

Probably the most classical example of adjoint of a partial differential equation is Green's formula for the Poisson equation (Green 1828), together with the related reciprocity theorems for the wave equation and electromagnetic fields.

Starting from the two-dimensional Poisson equation as an example, namely

$$\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} = u, \tag{48}$$

we multiply both sides by v and integrate. Then (42) is applied twice to the first and twice to the second term. There results the familiar Green's identity that is usually expressed through the ∇ operator as

$$\int_\Omega v\nabla^2 f \,\mathrm{d}\Omega = \oint_{\partial\Omega} \left(v\frac{\partial f}{\partial n} - f\frac{\partial v}{\partial n} \right) \mathrm{d}\partial\Omega + \int_\Omega f\nabla^2 v \,\mathrm{d}\Omega. \tag{49}$$

Defining an objective J according to (45) and comparing (49) with (44) then yields the explicit expression of the objective

$$J = \int_\Omega vu \,\mathrm{d}\Omega - \oint_{\partial\Omega} \left(v\frac{\partial f}{\partial n} - f\frac{\partial v}{\partial n} \right) \mathrm{d}\partial\Omega$$

and the adjoint differential equation

$$\nabla^2 v = y.$$

This is again a Poisson equation formally similar to (48), as originally discovered by Green, a fact which is today denoted by qualifying the Poisson problem as self-adjoint. The availability of an adjoint (be it coincident or not with the direct differential equation) implies that the objective J can be calculated multiple times as an integral over the direct problem's forcing and boundary data after (46) has been solved just once. For the Poisson equation, Green's result is most frequently applied with y equal to a Dirac delta function (e.g. Greenberg (1988)) so that the objective becomes the value of f at a target point, but the potentiality of the adjoint also remains open for any other objective and its corresponding y.

12 Adjoint of a nonlinear problem

Sensitivity: derivative of the objective function with respect to the forcing and boundary conditions.

As was remarked in §2 the sensitivity, or the adjoint, we have until now spoken of is defined as the derivative of the objective with respect to the input data of the problem. While for a linear problem this derivative is constant and can be related to the transpose of a matrix, for a more general nonlinear problem the adjoint will depend on the state itself and lose its linear-algebra interpretation. Nonetheless it will continue to make perfect sense as a derivative, something we have already made use of in §10.

The nonlinear generalization of the discrete-time system (9) (which includes (13) as remarked at the end of §5) can be written as a state-transition equation of the form $\mathbf{x}_{n+1} = \mathbf{f}_n(\mathbf{x}_n)$ (where $n = 0 \ldots N-1$), with initial condition $\mathbf{x}_0$ and objective $J = F(\mathbf{x}_N)$. The derivative of J with respect to the initial condition is given by the chain rule of differentiation as

$$\frac{\mathrm{d}J}{\mathrm{d}\mathbf{x}_0} = \frac{\partial J}{\partial \mathbf{x}_N}\frac{\partial \mathbf{x}_N}{\partial \mathbf{x}_{N-1}}\cdots\frac{\partial \mathbf{x}_2}{\partial \mathbf{x}_1}\frac{\partial \mathbf{x}_1}{\partial \mathbf{x}_0},$$

where each $\partial\mathbf{x}_{n+1}/\partial\mathbf{x}_n$ is a square jacobian matrix. In close analogy to (11), this repeated product may be associated to the left and written as a backwards recursion:

$$\mathbf{v}_n = \left(\frac{\partial \mathbf{x}_{n+1}}{\partial \mathbf{x}_n}\right)^{\mathrm{T}} \mathbf{v}_{n+1} \qquad (n = N-1\dots 0), \tag{50}$$

with terminal condition $\mathbf{v}_N = \partial J/\partial\mathbf{x}_N$ and eventual result $\mathbf{v}_0 = \mathrm{d}J/\,\mathrm{d}\mathbf{x}_0$. Since each $\mathbf{v}_n$ is a vector, not a matrix, (50) is no more computationally cumbersome than (11) and can exploit sparseness; the same remains true for other discrete or continuous problem variations.

In many practical applications, the transition function $\mathbf{f}_n$ will be such that only a few components of the state $\mathbf{x}$ actually change in each step. The direct and adjoint systems may then be usefully rewritten in the form

Nonlinear discrete-time dynamical system and its adjoint.

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{g}_n(\mathbf{x}_n) \qquad (n = 0\dots N-1) \tag{51}$$

$$\mathbf{v}_n = \mathbf{v}_{n+1} + \left(\frac{\partial \mathbf{g}_n}{\partial \mathbf{x}_n}\right)^{\mathrm{T}} \mathbf{v}_{n+1} \qquad (n = N-1\dots 0), \tag{52}$$

where $\mathbf{f}_n = \mathbf{x}_n + \mathbf{g}_n$ and just the nonzero components of $\mathbf{g}_n$ and $\partial\mathbf{g}_n/\partial\mathbf{x}_n$ need be acted upon. Notice that the adjoint vector $\mathbf{v}$ is sparsely altered, not in those components whose state changes in each step, but in those which the state components that change depend upon.

The only significant computational aggravation of the nonlinear adjoint is in its memory requirements: $\partial\mathbf{g}_n/\partial\mathbf{x}_n$ is a function of $\mathbf{x}_n$, and since the loop in (52) runs opposite to the loop in (51), the state $\mathbf{x}_n$ has to be permanently stored (or at least *checkpointed*, see Griewank & Walther (2008)) for it to be accessible in reverse order.

13 Adjoint of the Navier–Stokes problem

The most general fluid-mechanics problem we shall consider in this review is one governed by the Navier–Stokes equations for an incompressible fluid with constant viscosity:

$$\frac{\partial \vec{v}}{\partial t} + \nabla\cdot(\vec{v}\otimes\vec{v}) + \nabla p = \frac{1}{Re}\nabla^2\vec{v} + \vec{f}, \tag{53a}$$

$$\nabla\cdot\vec{v} = \dot{m}. \tag{53b}$$

Here $\vec{v}$ is the dimensionless velocity vector[5], p is the reduced pressure, Re is the Reynolds number and $\vec{f}$, $\dot{m}$ are an external volume force and mass source[6]

[5] The $\vec{\ }$ modifier is used to denote vectors in physical space, as distinguished from numerical vectors of arbitrary dimension. The scalar product of two space vectors is explicitly indicated by a dot, $\nabla\cdot$ denotes the divergence operator.

[6] Notice that in the presence of a mass source the conservative and convective forms of the momentum equation are no longer equivalent.

which will serve the role of the r.h.s. u in (43) or (48). No obstacle prevents us from applying adjoint techniques to compressible, non-newtonian or even more general fluids, but for definiteness we shall limit ourselves to (53).

Equations (53) can be linearized by letting $\vec{v} \leftarrow \vec{v} + \vec{\delta v}$ and similarly for the other variables, where symbols prefixed by a δ are infinitesimal in the sense of variational calculus. This results in the linearized Navier–Stokes equations (LNSE):

LNSE: Linearized Navier-Stokes equations

$$\frac{\partial \vec{\delta v}}{\partial t} + \nabla \cdot \left(\vec{v} \otimes \vec{\delta v} + \vec{\delta v} \otimes \vec{v} \right) + \nabla \delta p - \frac{1}{Re} \nabla^2 \vec{\delta v} = \vec{\delta f}, \tag{54a}$$

$$\nabla \cdot \vec{\delta v} = \delta \dot{m}. \tag{54b}$$

An arbitrary, generally nonlinear, functional J of the velocity and pressure fields may be introduced as the objective. Examples of such functionals are the integral of the kinetic-energy density v^2 over the volume Ω or part of it (possibly the inlet or outlet surface), the lift or drag (momentum flux through the solid boundaries), etc. The first variation δJ of the objective functional will be a linear functional of $\vec{\delta v}$ and δp, expressible in the integral form

$$\delta J = \int_0^T \int_\Omega \left(\mathcal{D}_{\vec{v}} J \cdot \vec{\delta v} + \mathcal{D}_p J \; \delta p \right) \mathrm{d}\Omega \, \mathrm{d}t$$

where $\mathcal{D}_{\vec{v}} J$ and $\mathcal{D}_p J$ (possibly distributions) are named *Fréchet derivatives* of J with respect to velocity and pressure.

In order to apply the Lagrange-Green identity (44), we can multiply (54a) and (54b) by two as yet unspecified functions (whose meaning will become clear later) $\vec{f}^\dagger$ and $\dot{m}^\dagger$, and integrate by parts as many times as needed to let derivatives of $\vec{\delta v}$ and δp disappear. Doing so yields the adjoint equations

Adjoint Navier–Stokes equations.

$$\frac{\partial \vec{f}^\dagger}{\partial t} + \vec{v} \cdot \left(\nabla \vec{f}^\dagger + \nabla \vec{f}^{\dagger \mathrm{T}} \right) + \nabla \dot{m}^\dagger + \frac{1}{Re} \nabla^2 \vec{f}^\dagger = -\mathcal{D}_{\vec{v}} J, \tag{55a}$$

$$\nabla \cdot \vec{f}^\dagger = -\mathcal{D}_p J, \tag{55b}$$

plus the explicit expression of the objective variation as a function of the forcing and boundary conditions:

$$\begin{aligned} \delta J = \int_0^T \int_\Omega \left(\vec{f}^\dagger \cdot \vec{\delta f} + \dot{m}^\dagger \delta \dot{m} \right) \mathrm{d}\Omega \, \mathrm{d}t - \left[\int_\Omega \vec{f}^\dagger \cdot \vec{\delta v} \, \mathrm{d}\Omega \right]_0^T + \\ - \int_0^T \oint_{\partial\Omega} \Bigg[\left(\vec{f}^\dagger \cdot \vec{\delta v} \right) \vec{v} + \left(\vec{f}^\dagger \cdot \vec{v} + \dot{m}^\dagger \right) \vec{\delta v} + \\ + \frac{1}{Re} \left(\nabla \vec{f}^\dagger \cdot \vec{\delta v} - \nabla \vec{\delta v} \cdot \vec{f}^\dagger \right) + \vec{f}^\dagger \; \delta p \Bigg] \cdot \vec{n} \, \mathrm{d}\partial\Omega \, \mathrm{d}t. \end{aligned} \tag{56}$$

Just as in §8, the boundary conditions for (55) must be chosen by the rule that, term by term and portion by portion of the boundary, (56) should only depend on the physically assigned boundary conditions for (54) (for instance, $\vec{f}^\dagger$ must be zero at final time if $\vec{\delta v}$ is assigned at initial time, or both must be periodic if an oscillating solution is sought). It then becomes apparent from the first term of (56) that the adjoint variables $\vec{f}^\dagger$ and $\dot{m}^\dagger$ have the meaning of Fréchet derivatives of J with respect to volume force and mass injection rate, while the remaining terms provide the sensitivities of J to the boundary conditions, which are of importance in the *receptivity* analysis of boundary layers. In a problem with no volume force or mass injection, only the sensitivities at the boundaries are used but the meaning of $\vec{f}^\dagger$ and $\dot{m}^\dagger$ remains unchanged, just as in Green's formula for the Laplace rather than the Poisson equation.

Variables of the adjoint Navier–Stokes problem: Fréchet derivatives of the objective with respect to volume force and mass injection rate.

14 Writing an adjoint computer program: adjoint of the discretization versus discretization of the adjoint

As clearly appears from the considerations at the beginning of §6, there is a complete correspondence between the adjoint of an integral or differential problem and the adjoint of a discrete approximation of the same, provided the approximation is convergent. Therefore in numerical applications a choice can be operated between writing an adjoint differential equation such as (46) or (55) and then discretizing it (discretization of the adjoint, DA) or starting from a discretization of the direct problem in the form (5) and obtaining its adjoint (adjoint of the discretization, AD) as in (7). For most purposes the choice is a matter of personal preference, however a few subtle points are involved:

DA: Discretization of the Adjoint.
AD: Adjoint of the Discretization (also used for Automatic Differentiation, with essentially the same meaning).

1. Although DA and AD converge to the same result (when they converge at all, which case by case may or may not be an easy proposition to prove mathematically), they do not necessarily share the same discretization error for nonzero step size.

2. Convergence may be in a weak (integral) sense rather than pointwise, especially when derivatives are involved; the evaluation of the sensitivity to boundary conditions requires special care.

3. For efficiency the AD approach should not be taken literally to mean that an overall resolvent matrix must be built and then transposed, but rather that each intermediate stage of the numerical procedure must be taken an adjoint of. Fortunately the process is fairly straightforward, and can even be automated through automatic differentiation software (Griewank & Walther 2008).

Starting from the last point, every computer program is a dynamical system of the form (51), where n numbers the individual steps of program execution (i.e. program statements or blocks of statements in the order they are executed, inclusive of the effect of any branches and loops, as opposed to the order in which they are written) and $\mathbf{x}_n$ is the set of all variables the program keeps in memory at the beginning of each step. The key to, either manually or automatically, writing an AD program is to run through the same sequence of steps backwards, each time updating by (52) the sensitivity of those variables only on which the present step depends. The result will be an adjoint computer program that has a comparable computation time to (exactly the same number of steps as) the direct program.

Convergence of AD is not an issue when the optimal value of an objective is being determined. In this case, in fact, the sensitivity must be driven to zero, and doing so will provide the exact optimum of the discretized objective. If the direct discretization makes any sense, this is the desired result.

When, on the other hand, the sensitivity itself is the target of the computation, the choice between AD and DA is less clear-cut. An example may serve to clarify the potential pitfalls involved.

If the integral (18) defines the objective of an instantaneous system (whose state $\mathbf{x}$ coincides for simplicity with its input $\mathbf{u}$), the sensitivity of J to $\mathbf{x}$ is, trivially, $\mathbf{y}$. If now this integral is discretized by a first-order rectangle rule,

$$J_{\mathrm{r}} = \Delta t \sum_{n=0}^{N-1} \mathbf{y}(n\Delta t)\mathbf{x}(n\Delta t),$$

the derivative of J_{r} with respect to $\mathbf{x}(n\Delta t)$ is $\mathbf{y}(n\Delta t)\Delta t$, proportional to a first-order discretization of the exact result as may be expected (but notice that when the discretization is not uniform the factor Δt must be handled properly). In order to obtain a higher precision, one may decide to discretize the integral by the composite Simpson rule,

$$J_{\mathrm{S}} = \Delta t \sum_{n=0}^{N} w_n \mathbf{y}(n\Delta t)\mathbf{x}(n\Delta t),$$

Composite Simpson rule and its adjoint.

where the weights w_n equal 1/3 for $n = 0$, $n = N$ and otherwise 4/3 for n odd, 2/3 for n even.

The derivative of J_{S} with respect to $\mathbf{x}(n\Delta t)$ will then turn out to be $w_n\mathbf{y}(n\Delta t)\Delta t$, a function oscillating from one grid point to the next which does not converge at all (in a pointwise sense) to the desired $\mathbf{y}(n\Delta t)$. On the other hand, the integral of the product between $\mathbf{y}$ and any smooth function of compact support converges with 4^{th}-order or better accuracy, which shows that convergence can be recovered by a suitable smoothing (or by adopting a non-oscillatory quadrature formula, e.g. Luchini (1994)).

In quite a few problems of mathematical physics, among which the Navier–Stokes equations (53), it so occurs that the highest-derivative terms form a self-adjoint system (the steady Stokes problem is self-adjoint under the transformation $\vec{f}^\dagger = \vec{v}$, $-\dot{m}^\dagger = p$, as may easily be verified by dropping the convective terms in (53) and (55), a fact which probably inspired Hill (1992) and other authors to adopt the names of *adjoint velocity* and *adjoint pressure* for these quantities). In this case AD and DA have equivalent smoothing properties: paying attention to choose a self-adjoint discretization one can get the best of both worlds. For instance, many finite-element discretizations are self-adjoint by construction (they were originally designed for problems that admit a variational principle, which is another facet of being self-adjoint). Finite differences are frequently self-adjoint on a uniformly spaced grid but may require special care on non-uniform ones. Essentially non-self-adjoint, on the other hand, are problems involving inviscid flow, especially in the presence of shocks; these are typical cases where the difference between AD and DA must be given the greatest consideration (Jameson 2003).

An easy test to validate an adjoint program (or part of it) is to verify that the defining identity between (2) and (4) be satisfied for two randomly generated vectors $\mathbf{u}$ and $\mathbf{y}$. This test just requires running the direct program with $\mathbf{u}$ as data, the adjoint program with $\mathbf{y}$, and then comparing the two linear forms $\mathbf{yx}$ and $\mathbf{vu}$. When AD is adopted, the values of J obtained must match within machine accuracy independently of the discretization stepsize, which is very unlikely to happen by chance in the presence of programming errors. Even more forcefully (12) must be satisfied at every intermediate step, which allows a faulty step to be quickly identified. The availability of this test alone is, in our experience, sufficient motivation to favour AD over DA whenever applicable.

Validating an adjoint program: run the direct program with a random $\mathbf{u}$, the adjoint program with a random $\mathbf{y}$, and then compare the values of $\mathbf{yx}$ and $\mathbf{vu}$.

References

Airiau C, Walther S, Bottaro A. 2002. Boundary layer sensitivity and receptivity. *C. R. Mécanique* 330:259–265

Andersson P, Berggren M, Henningson DS. 1998. Optimal disturbances in boundary layers. In *Proc. AFOSR Workshop on Optimal Design and Control, Arlington VA, USA, 30 Sept–3 Oct 1997*, eds. JT Borggaard, J Burns, E Cli, S Schreck. Birkhauser

Andersson P, Berggren M, Henningson DS. 1999. Optimal disturbances and bypass transition in boundary layers. *Phys. Fluids* 11:134–150

Bewley TP, Luchini P, Pralits J. 2013. Methods for the solution of large optimal control problems that bypass open-loop model reduction. *IEEE Trans. Autom. Control* submitted, available at `http://fccr.ucsd.edu/pubs/BLP_MCE_ADA_OSSI.pdf`

Bottaro A, Corbett P, Luchini P. 2003. The effect of base flow variation on flow stability. *J. Fluid Mech.* 476:293–302

Butler KM, Farrell BF. 1992. Three-dimensional optimal perturbations in viscous flows. *Phys. Fluids A* 4:1637–1650

Cathalifaud P, Luchini P. 2000. Algebraic growth in boundary layers: optimal control by blowing and suction at the wall. *Europ. J. Mech. B/Fluids* 19:469–490

Courant R, Hilbert D. 1924. *Methoden des mathematischen Physik, I.* Verlag von Julius Springer

Craig T. 1889. *A treatise on linear differential equations.* John Wiley & Sons

Dunford N, Schwartz JT. 1958. *Linear Operators part I: General Theory.* Interscience

Fedorov AV. 1984. Excitation of Tollmien-Schlichting waves in a boundary layer by a periodic external source located on the body surface. *Fluid Dyn.* 19:888–893

Forsyth AR. 1888. Invariants, covariants, and quotient derivatives associated with linear differential equations. *Phil. Trans. R. Soc.* 179:377–489

Fuchs L. 1873. Über relationen, welche für die zwischen je zwei singulären Punkten erstreckten Integrale der Lösungen linearer Differentialgleichungen stattfinden. *Journal für die reine angewandte Mathematik* 76:177–213

Giannetti F, Luchini P. 2003. Receptivity of the circular cylinder's first instability. In *The 5th Euromech Fluid Mechanics Conference, Toulouse, France, August 24-28, 2003.* IMFT

Giannetti F, Luchini P. 2007. Structural sensitivity of the first instability of the cylinder wake. *J. Fluid Mech.* 581:167–197

Goldstein H. 1980. *Classical Mechanics.* Addison–Wesley

Golub GH, Van Loan CF. 1983. *Matrix Computations.* Johns Hopkins University Press

Green G. 1828. *An essay on the application of mathematical analysis to the theories of electricity and magnetism.* Printed for the author by T. Wheelhouse

Greenberg MD. 1988. *Advanced engineering mathematics.* Prentice-Hall

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

Hill DC. 1992. A theoretical approach for analyzing the restabilization of wakes. In *AIAA paper 1992-0067*

Hill DC. 1995. Adjoint systems and their role in the receptivity problem for boundary layers. *J. Fluid Mech.* 292:183–204

Jameson A. 2003. Aerodynamic shape optimization using the adjoint method. In *Lecture Notes, von Karman Institute for Fluid Dynamics, Brussels, Belgium, Feb. 2003*

John F. 1982. *Partial Differential Equations.* Springer–Verlag

Kim J, Bewley TR. 2007. A linear systems approach to flow control. *Annu. Rev. Fluid Mech.* 39:383–417

Lagrange JL. 1763. Solution de différens problêmes de calcul intégral. *Miscellanea Taurinensia* 3:179–380. [also in Oeuvres de Lagrange, J.A. Serret, ed., Vol. 1, 1867, pp. 471-668]

Luchini P. 1994. End-correction integration formulae with optimized terminal sampling points. *Comput. Phys. Commun.* 83:236–244

Luchini P. 1997. Effects on a flat–plate boundary layer of free–stream longitudinal vortices: optimal perturbations. In *EUROMECH 3rd European Fluid Mechanics Conference, Göttingen 15-18 Sept. 1997.* DLR

Luchini P. 2000. Reynolds-number-independent instability of the boundary layer over a flat surface: Optimal perturbations. *J. Fluid Mech.* 404:289–309

Luchini P, Bottaro A. 1996. A time-reversed approach to the study of Görtler instabilities. In *Advances in Turbulence VI, Proc. 6th Europea n Turbulence Conference, Lausanne 2-5 July 1996*, eds. S Gavrilakis, L Machiels, PA Monkewitz. Kluwer

Luchini P, Bottaro A. 1998. Görtler vortices: a backward-in-time approach to the receptivity problem. *J. Fluid Mech.* 363:1–23

Pralits JO, Luchini P. 2010. Riccati-less optimal control of bluff-body wakes. In *Proc. 7th IUTAM Symposium on Laminar-Turbulent Transition, June 23-26, 2009, Stockholm, Sweden.*, eds. P Schlatter, DS Henningson. Springer

Press WH, A.Teukolsky S, Vetterling WT, Flannery BP. 1999. *Numerical Recipes.* Cambridge University Press

Semeraro O, Pralits JO, Rowley CW, Henningson DS. 2013. Riccati-less approach for optimal control and estimation: an application in 2D boundary layers. *J. Fluid Mech.* submitted

Tricomi FG. 1957. *Integral Equations.* Interscience

Tumin AM, Fedorov AV. 1984. Instability wave excitation by a localized vibrator in the boundary layer. *J. Appl. Mech. Tech. Phys.* 25:867–873

Zuccher S. 2002. *Receptivity and control of flow instabilities in a boundary layer.* Ph.D. thesis, Politecnico di Milano, Italy

Zuccher S, Luchini P. 2013. Boundary-layer receptivity to external disturbances using multiple scales. *Meccanica* in press

Zuccher S, Luchini P, Bottaro A. 2004. Algebraic growth in a Blasius boundary layer: optimal and robust control by mean suction in the nonlinear regime. *J. Fluid Mech.* 513:135–160