Appendix A

Linear Algebra

We recall here a few elements of linear algebra frequently used in applications. Two perspectives are taken, the first one focuses on the decomposition of vectors on eigen-bases, at the heart of linear stability theory (§A.2), the other tackles essentially the same questions but from an "energetic" point of view *via* the definition of appropriate scalar products (§A.3). Beforehand, Section A.1 is devoted to the enumeration of a few elementary properties linked to the matrix representation of operators and changes of bases. Among the many reference books on the subject, let us just mention [Hirsch and Smale (1974)] that fits our purpose particularly well.

A.1 Vector Spaces, Bases, and Linear Operators

We suppose that the definition and immediate properties of vector spaces are known. So, let us consider such a space, noted $\mathbb{X}$ with elements (vectors) $\mathbf{X}$. Scalars x entering linear combinations can be real or complex. To every vector space on $\mathbb{R}$, one can associate a complex extension formed with the same vectors but in which the scalars (and thus also the components of the vectors) are in $\mathbb{C}$.

A vector $\mathbf{X} \in \mathbb{X}$ can be specified by its components in a given basis $\{\mathbf{E}_i, i = 1, \ldots, d\}$:

$$\mathbf{X} = \sum_{i=1}^{d} x_i \mathbf{E}_i \,, \tag{A.1}$$

where d is the dimension of $\mathbb{X}$. A *linear form* is a linear function that associates a scalar to every vector in the space. The set of forms also has a vector space structure called the *dual space*. Forms extracting every component of the vectors on a given basis makes the *dual basis*.

A linear operator $\mathcal{L}$ can be defined by its action on the basis vectors:

$$\mathcal{L}\mathbf{E}_j = \sum_i l_{ij}\mathbf{E}_i\,,$$

which defines a $(d \times d)$, two-dimensional array called a *matrix*, here denoted $[\,\mathrm{L}\,]$; by convention, the first (second) subscript labels the lines (columns). The column vectors of $[\,\mathrm{L}\,]$ are the images of the basis vectors. The result of the action of $\mathcal{L}$ on an arbitrary vector then reads

$$y_j = (\mathcal{L}\mathbf{X})_j = \sum_i l_{ji}x_i\,. \tag{A.2}$$

The representation of a linear operator changes with the basis. Let us define the new basis vectors $\{\mathbf{F}_j, j = 1, \ldots, d\}$ by their components in the old basis through:

$$\mathbf{F}_j = \sum_i t_{ij}\mathbf{E}_i\,. \tag{A.3}$$

This builds a matrix $[\,\mathrm{T}\,]$ with elements t_{ij} representing the operator $\mathcal{T}$ expressing the change of basis. $\mathcal{T}$ is a linear *invertible* operator, whose column vectors are the components of the new basis vectors in the old basis. Components $\{x_i\}$ and $\{x'_i\}$ of some given vector $\mathbf{X}$ in the two bases $\{\mathbf{E}_i\}$ and $\{\mathbf{F}_i\}$ are then related by:

$$x_i = \sum_j t_{ij}x'_j\,, \tag{A.4}$$

while the new components are given as a function of the old ones by:

$$x'_j = \sum_k u_{jk}x_k\,, \tag{A.5}$$

where the u_{ij} are the elements of the matrix $[\,\mathrm{U}\,]$ inverse of $[\,\mathrm{T}\,]$ and denoted $[\,\mathrm{T}\,]^{-1}$. For this reason, the components x_i are said to be *contra-variant*.

Concretely, the inverse of a given square matrix $[\,\mathrm{A}\,]$ can be obtained by hand as the transposed of the matrix of cofactors, divided by the determinant $\det([\,\mathrm{A}\,])$ of the matrix. The cofactor of element a_{ij} is $\alpha_{ij} = (-1)^{i+j}\det([\,\tilde{\mathrm{A}}_{ij}\,])$, where $[\,\tilde{\mathrm{A}}_{ij}\,]$ is the $(d-1)\times(d-1)$ matrix obtained by suppressing line i and column j of matrix $[\,\mathrm{A}\,]$. Matrix $[\,\mathrm{U}\,]$ can be obtained from $[\,\mathrm{T}\,]$ in this way.

From (A.1), (A.3) and (A.4), one readily gets:

$$y'_j = \sum_k u_{jk}\sum_l l_{kl}\sum_i t_{li}x'_i = \sum_i l'_{ji}x'_i\,,$$

hence

$$l'_{ji} = \sum_k\sum_l u_{jk}l_{kl}t_{li}\,, \tag{A.6}$$

so that the elements of the matrix representing $\mathcal{L}$ in the new basis are given by $l'_{li} = \sum_k \sum_j u_{lk} m_{kj} t_{ji}$, *i.e.*

$$[\mathrm{L}'] = [\mathrm{T}]^{-1}[\mathrm{L}][\mathrm{T}].$$

Matrices $[\mathrm{L}']$ et $[\mathrm{L}]$ are said to be *similar*. It is easy to see that

$$[\mathrm{L}']^k = [\mathrm{T}]^{-1}[\mathrm{L}]^k[\mathrm{T}].$$

Remark. While developing Quantum Mechanics, Dirac introduced notations useful to linear algebra. In his setting, vectors are denoted as $|\mathbf{X}\rangle$ and linear forms as $\langle\mathbf{Y}|$. Let $\{|\mathbf{E}_i\rangle, i = 1, ..., d\}$ be a basis and $\{\langle\mathbf{E}_i|, i = 1, ..., d\}$ the dual basis. The projector on basis vector $|\mathbf{E}_i\rangle$ reads $|\mathbf{E}_i\rangle\langle\mathbf{E}_i|$, and the fact that the basis is complete leads to $\sum_i |\mathbf{E}_i\rangle\langle\mathbf{E}_i| = \mathcal{I}$, where $\mathcal{I}$ is the identity operator in $\mathbb{X}$. Coming back to (A.1), one gets $|\mathbf{X}\rangle = \left[\sum_i |\mathbf{E}_i\rangle\langle\mathbf{E}_i|\right] |\mathbf{X}\rangle$, so that, by identification, $\langle\mathbf{E}_i|\mathbf{X}\rangle$ is the ith contra-variant component of $|\mathbf{X}\rangle$.

The action of the linear operator $\mathcal{L}$ on a vector $|\mathbf{X}\rangle$ gives the vector

$$|\mathbf{Y}\rangle = \mathcal{L}|\mathbf{X}\rangle = \left[\sum\nolimits_j |\mathbf{E}_j\rangle\langle\mathbf{E}_j|\right] \mathcal{L} \left[\sum\nolimits_i |\mathbf{E}_i\rangle\langle\mathbf{E}_i|\right] |\mathbf{X}\rangle\,,$$

which allows the identification $l_{ji} = \langle\mathbf{E}_j|\mathcal{L}|\mathbf{E}_i\rangle$, cf. (A.2).

In a change of basis, vectors of the new basis are defined from the old one by $|\mathbf{F}_j\rangle = \sum_i |\mathbf{E}_i\rangle\langle\mathbf{E}_i|\mathbf{F}_j\rangle$, which defines operator $\mathcal{T}$. Its inverse $\mathcal{U} = \mathcal{T}^{-1}$ is represented by the matrix with elements $u_{ji} = \langle\mathbf{F}_j|\mathbf{E}_i\rangle$. The transformation (A.4) of the components of vector $|\mathbf{X}\rangle$ is then immediately obtained. In the same way the elements of the matrix representing $\mathcal{L}$ in basis $\{|\mathbf{F}_j\rangle\}$ are given by: $\langle\mathbf{F}_j|\mathcal{L}|\mathbf{F}_i\rangle = \sum_k \sum_l \langle\mathbf{F}_j|\mathbf{E}_k\rangle\langle\mathbf{E}_k|\mathcal{L}|\mathbf{E}_l\rangle\langle\mathbf{E}_l|\mathbf{F}_i\rangle$, which is easily identified to (A.6). At this stage Dirac's notations are essentially of mnemotechnical interest.

MATLAB notations can also be useful. In this framework, all kinds of arrays are placed between square brackets. The comma (or a blank space) is used to separate elements in a line, the semi-colon to indicate the line change (equivalent to a 'carriage return'). A vector, traditionally represented in column form is thus a $(d \times 1)$ array $\mathbf{V} = [V_1; V_2; ...]$. The line representation, transposed of the column representation then reads: $[V_1, V_2, ...] \equiv [V_1; V_2; ...]^{\mathrm{t}}$. With these conventions, the canonical scalar product can be written $\mathbf{V} \cdot \mathbf{W} = [V_1, V_2, ...][W_1; W_2; ...] = \sum V_k W_k$. By contrast the tensorial product of two vectors $\mathbf{V} \otimes \mathbf{W}$ is a $d \times d$ array with elements $V_i W_j$, which one can write $[V_1; V_2; ...][W_1, W_2, ...]$.

A.2 Structure of a Linear Operator

As we have seen in Chapter 2, and more specifically in Section 2.2, the solution to the initial value problem for a linear differential system:

$$\dot{\mathbf{X}} = \mathcal{L}\mathbf{X}\,, \quad \mathbf{X}(t=0) = \mathbf{X}_0\,, \tag{A.7}$$

leads to the evaluation of

$$\mathbf{X}(t) = \exp(t\mathcal{L})\mathbf{X}_0 \tag{A.8}$$

with, by definition:

$$\exp \mathcal{Z} \equiv \mathcal{I} + \sum_{k=1}^{\infty} \frac{1}{k!}\mathcal{Z}^k \tag{A.9}$$

for an arbitrary linear operator $\mathcal{Z}$. The computation of this exponential is made easier by resolving the structure of the operator $\mathcal{Z}$, that is to say by decomposing the full space $\mathbb{X}$ into a direct sum of nontrivial *invariant subspaces*.

One says that a vector subspace $\mathbb{X}'$ of $\mathbb{X}$ is invariant for $\mathcal{L}$ if the image of every $\mathbf{X} \in \mathbb{X}'$ is in $\mathbb{X}'$. Trivial invariant subspaces are $\mathbb{X}$ itself, the null space $\mathbb{O}$, the image of $\mathbb{X}$ by $\mathcal{L}$ (set of vectors $\mathbf{Y} = \mathcal{L}\mathbf{X}$ for all $\mathbf{X} \in \mathbb{X}$) and the kernel of $\mathcal{L}$ (set of vectors $\mathbf{Z} \in \mathbb{X}$ such that $\mathcal{L}\mathbf{Z} = \mathbf{0}$).

The kernel of $\mathcal{L}$ is reduced to $\mathbb{O}$ when $\mathcal{L}$ is invertible, which is the case if its determinant is nonzero. One then says that the matrix has 'maximal rank'. (The rank of a general matrix, *i.e.* not necessarily square, is the dimension of the largest square sub-matrix with nonzero determinant that can be extracted from it by suppressing lines and columns.)

The simplification brought by the so-obtained decomposition of space $\mathbb{X}$ is apparent through the change of basis that makes it explicit: Let $\mathbb{X} = \mathbb{X}_1 \oplus \mathbb{X}_2$ and $\mathcal{L}$ the operator for which $\mathbb{X}_1$ is invariant, then for $\mathbf{X}_1 \in \mathbb{X}_1$, $\mathcal{L}\mathbf{X}_1 \in \mathbb{X}_1$ and, in a basis of $\mathbb{X}$ formed with a basis of $\mathbb{X}_1$ appropriately completed by a basis in the supplementary space $\mathbb{X}_2$, $\mathcal{L}$ takes on a block structure:

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{O} & \mathcal{L}_{22} \end{bmatrix},$$

where $\mathcal{O}$ is the null operator ($\mathcal{O}\mathbf{X} = \mathbf{0}$ whatever $\mathbf{X}$). If in addition $\mathbb{X}_2$ is

also invariant, $\mathcal{L}$ acquires a block-diagonal structure:

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{O} \\ \mathcal{O} & \mathcal{L}_{22} \end{bmatrix} .$$

The eigen-direction attached to an eigenvalue s of $\mathcal{L}$, such that:

$$\mathcal{L}\mathbf{X} = s\mathbf{X}$$

is the prototype of the sought invariant subspaces (hence the kernel of operator $\mathcal{M} = \mathcal{L} - s\mathcal{I}$). The eigenvalues are the roots of the *characteristic polynomial* obtained by expanding the determinant:

$$\det(\mathcal{L} - s\mathcal{I}) = 0 .$$

The *fundamental theorem of algebra* asserts that, in dimension d, this degree-d polynomial has d roots in $\mathbb{C}$, possibly degenerate. One can thus write:

$$\det(\mathcal{L} - s\mathcal{I}) = a_0 + a_1 s + \ldots + a_{d-1}s^{d-1} + s^d = 0 = \prod_j (s - s_j)^{d_j} ,$$

where d_j is the multiplicity of eigenvalue s_j, with $\sum_j d_j = d$.

In order to treat complex eigenvalues ($s \in \mathbb{C}$) of a linear operator acting in a vector space on $\mathbb{R}$, one must work with its *complex extension.* For example a rotation in a two-dimensional subspace is associated to a pair of simple conjugate purely imaginary roots, solutions to $s^2 + \omega^2 = 0$, and, though there are no real eigenvectors, there are two eigen-directions in the complex extension.

In the general case, at least one eigenvector $\widehat{\mathbf{X}}_j$ can be found for a given (real or complex) eigenvalue s_j in the complex extension. If s_j is non degenerate, the associated eigen-subspace is one-dimensional. It is generated by the eigenvector $\widehat{\mathbf{X}}_j$ solution to $(\mathcal{L} - s_j\mathcal{I})\widehat{\mathbf{X}}_j = \mathbf{0}$ (see Exercise A.4).

If s_j is degenerate, which corresponds in dynamics to a linear resonance condition, the problem is more complicated and one must search for a special basis in which the matrix representing the restriction of the operator to this subspace is in its *normal* form.

A.2.1 *Jordan normal form*

The main tool to determine the Jordan normal form of an operator is the theory of operator polynomials, *i.e.* polynomials in the form $\mathcal{Q} = \sum_{n=0}^{n_{\max}} a_n \mathcal{L}^n$. The decomposition of the vector space $\mathbb{X}$ into a direct sum of

invariant subspaces rests on the determination of the kernels of operators $(\mathcal{L} - s_j\mathcal{I})^{d'}$, where d' is a trial dimension, as a generalization of the eigen-direction defined as the kernel of $\mathcal{L} - s_j\mathcal{I}$. These kernels are called the *generalized eigen-subspaces* associated to eigenvalue s_j.

Without entering the derivation, we now give some practical results about them. First they form a series of embedded subspaces of increasing dimensions, the kernel of $(\mathcal{L} - s\mathcal{I})^2$ containing that of $\mathcal{L} - s\mathcal{I}$, *etc.* The largest of them, called the *principal subspace*, has dimension d_j equal to the multiplicity of the eigenvalue, it is the kernel of $(\mathcal{L} - s_j\mathcal{I})^{d_j}$. Next, the Cayley–Hamilton theorem stipulates that $\mathcal{L}$ fulfills its own characteristic polynomial, that is:

$$\prod_j (\mathcal{L} - s_j\mathcal{I})^{d_j} \equiv 0\,,$$

which can be understood as the most direct translation of this decomposition into a direct sum of principal subspaces.

The numerical approach to multiples eigenvalues as it is implemented in usual computer routines can lead to difficulties in concrete cases. It may therefore be interesting to present the main analytical steps of the computation. Let $\mathbb{X}_j$ be the principal subspace associated to s_j, with dimension d_j, and $\mathcal{L}_j$ the restriction of $\mathcal{L}$ to $\mathbb{X}_j$. Define also $\mathcal{M}_j = \mathcal{L}_j - s_j\mathcal{I}_{d_j}$, where $\mathcal{I}_{d_j}$ is the identity operator in $\mathbb{X}_j$. Then:

- The index of a principal vector $\mathbf{X}$ $(\in \mathbb{X}_j)$ is defined as the smallest nonnegative integer m such that $\mathcal{M}_j^m\mathbf{X} = 0$. This index m is such that $1 \leq m \leq d_j$; the lower bound $m = 1$ corresponds to the case of the eigenvector $(\mathcal{M}\mathbf{X} = 0)$; the upper bound d_j is because $\mathbf{X}$ belongs to $\mathbb{X}_j$ by assumption $(\mathcal{M}_j^{d_j}\mathbf{X} = 0)$. It can easily be seen that when $\mathbf{X}$ is a vector with index m then $\mathcal{M}_j^n\mathbf{X}$ as index $m - n$ for all $n < m$.

- Exploring $\mathbb{X}_j$, one begins with finding the maximum index $m_{\max}$ of vectors in this subspace by computing the kernel of $\mathcal{M}^{d'}$ for increasing trial dimensions d'. When $m_{\max} = d_j$, one gets a vector with maximal index $\mathbf{Y}_{d_j}$ by explicitly solving the linear problem

$$(\mathcal{L}_j - s_j\mathcal{I}_{d_j})^{d_j}\mathbf{X} = 0\,.$$

- One then considers the series of d_j vectors $\mathbf{Y}_{d_j}$, $\mathbf{Y}_{d_j-1} = \mathcal{M}_j\mathbf{Y}_{d_j}$, $\mathbf{Y}_{d_j-2} = \mathcal{M}_j\mathbf{Y}_{d_j-1} = \mathcal{M}_j^2\mathbf{Y}_{d_j}$, $\ldots$, $\mathbf{Y}_1 = \mathcal{M}_j\mathbf{Y}_2 = \mathcal{M}_j^{d_j-1}\mathbf{Y}_{d_j}$, that one can write in the more condensed form:

$$\left\{\mathbf{Y}_{d_j-n} = \mathcal{M}_j^n\mathbf{Y}_{d_j};\quad n = 0, 1, \ldots, d_j - 1\right\}\,,$$

noticing that $\mathbf{Y}_0 = \mathcal{M}_j \mathbf{Y}_1 = \mathcal{M}_j^{d_j} \mathbf{Y}_{d_j} \equiv 0$ since $\mathbf{Y}_{d_j}$ has maximal index by assumption. The series of vectors $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_{d_j}\}$ is an acceptable basis for the principal subspace $\mathbb{X}_j$ attached to s_j since one has d_j linearly independent vectors. This property can easily be checked by showing that any linear combination $\mathbf{Z} = \sum_{n=1}^{d_j} \mu_n \mathbf{Y}_n$ cancels if and only if all its coefficients cancel. Computing first $\mathcal{M}_j^{d_j-1}\mathbf{Z}$, one indeed gets a single non-identically zero term $\mu_{d_j}\mathbf{Y}_{d_j}$, which implies $\mu_{d_j} = 0$. Next, the computation of $\mathcal{M}^{d_j-2}\mathbf{Z}$ gives $\mu_{d_j-1} = 0$, and so on, down to μ_1.

- The Jordan normal form directly derives from the matrix representation of $\mathcal{L}_j$ in this basis: Vector $\mathbf{Y}_1$ (index 1) is along eigenvector $\widehat{\mathbf{X}}_j$ and, by construction of the $\mathbf{Y}_n$, one gets:

$$\mathbf{Y}_{n-1} = (\mathcal{L}_j - s_j \mathcal{I}_{d_j})\mathbf{Y}_n \,,$$

that is:

$$\mathcal{L}_j \mathbf{Y}_n = s_j \mathbf{Y}_n + \mathbf{Y}_{n-1} \,.$$

The expression of the elementary Jordan block directly derives from this expression:

$$\begin{array}{c} \\ n-1 \\ n \\ n+1 \end{array} \begin{array}{c} \begin{array}{ccc} n-1 & n & n+1 \end{array} \\ \begin{pmatrix} 0 & \vdots & \\ 1 & 0 & \vdots \\ s_j & 1 & 0 \\ 0 & s_j & 1 \\ \vdots & 0 & s_j \\ & \vdots & 0 \end{pmatrix} \end{array}$$

Introducing the operator $\mathcal{N}_{d_j}$ represented in this basis by a matrix with a series of ones just above the diagonal and zeroes everywhere else, one gets

$$\mathcal{L}_j = s_j \mathcal{I}_{d_j} + \mathcal{N}_{d_j} \,.$$

- When the maximal index is such that $m_{\max} < d_j$, one first constructs the subspace generated by the $\mathbf{Y}_n$ just found. This subspace has dimension $m_{\max}$, and one has to redo the same work in the supplementary subspace with dimension $d_j - n_{\max}$, *i.e.* find the maximal index of vectors in this new

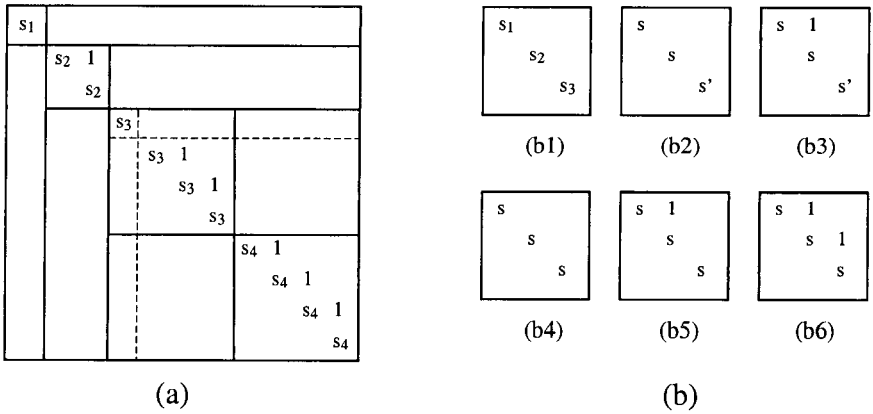

Fig. A.1 (a) Jordan block structure in the basis of generalized eigenvectors, example: s_1 is non-degenerate; s_2 is double but without diagonal form; s_3 is with multiplicity four, decomposed into two subspaces with dimensions one and three; s_4 is also with multiplicity four (zero entries are left blank). (b) Possible cases in dimension $d = 3$: Three distinct eigenvalues, possibly only in the complex extension if two eigenvalues out of the three are complex conjugate (b1). Two real distinct eigenvalues, in diagonal form (b2) or not (b3). Eigenvalue with multiplicity three and vectors with maximum index $n_{\max} = 1$, 2, or 3, giving three Jordan blocks of order 1 (b4), one of order 1 and one of order 2 (b5), or one of order 3 (b6).

subspace, find one vector with maximal index, build the associated basis, *etc.* up to the point where the full subspace $\mathbb{X}_j$ is decomposed. Using the so-obtained partial bases to build the basis of $\mathbb{X}_j$ one can represent $\mathcal{L}_j$ as a direct sum of Jordan blocks aligned along the diagonal. This result is illustrated in Figure A.1(a). In applications, one rarely deals with problems in a dimension higher than three analytically. The work involved in this systematic approach is therefore never as fastidious as it could seem. Figure A.1(b) illustrates all the possible cases for $d = 3$. Notice also that in case (b1), when the three distinct roots are s_1 plus a pair of complex eigenvalues $s_\pm = \sigma \pm i\omega$, the diagonal form holds in the complex extension only, otherwise one has a square block, *e.g.*

$$\begin{bmatrix} s_1 & 0 & 0 \\ 0 & \sigma & -\omega \\ 0 & +\omega & \sigma \end{bmatrix}.$$

The two-dimensional case was illustrated in Chapter 2 without mentioning the index concept. This gap is filled in Exercise A.4.2.

A.2.2 *Exponential of a matrix*

As far as the linear dynamics evolution governed by (A.7) is concerned, the aim of all that precedes is to turn operator $\mathcal{L}$ into the best adapted form in view of the computation of its exponential (A.8) defined as a power series (A.9). In the basis where $\mathcal{L}$ is in its Jordan form one gets:

$$\exp(t\mathcal{L}_j) = \exp\left[t\left(s_j\mathcal{I}_{d_j} + \mathcal{N}_{d_j}\right)\right] = \exp\left(ts_j\mathcal{I}_{d_j}\right)\exp\left(t\mathcal{N}_{d_j}\right) .$$

Here the exponential of the sum is simply the product of the exponentials because $\mathcal{I}_{d_j}$ commutes with any operator and thus also with $\mathcal{N}_{d_j}^n$. One obviously gets $\exp\left(ts_j\mathcal{I}_{d_j}\right) = \exp(ts_j)\,\mathcal{I}_{d_j}$, and one is left with the computation of the second exponential, which is an easy matter since operator $\mathcal{N}_{d_j}$ is *nilpotent*, *i.e.*, $\left(\mathcal{N}_{d_j}\right)^{d_j} = \mathcal{O}$, so that we are left with a finite number of terms in the power series. Considering the case $d_j = 3$ as an example and dropping all useless indices, we get:

$$\mathcal{N} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathcal{N}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathcal{N}^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

so that

$$\exp(t\mathcal{N}) = \mathcal{I} + t\,\mathcal{N} + \tfrac{1}{2}t^2\mathcal{N}^2 = \begin{bmatrix} 1 & t & \frac{1}{2}t^2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}$$

and thus

$$\exp(t\mathcal{L}) = \begin{bmatrix} \exp(st) & t\,\exp(st) & \frac{1}{2}t^2\,\exp(st) \\ 0 & \exp(st) & t\,\exp(st) \\ 0 & 0 & \exp(st) \end{bmatrix}.$$

The terms $(t^j/j!)\,\exp(st)$ are said to be *secular*. They account for a slow algebraic drift with respect to a dominant exponential behavior.

A.2.3 *Perturbation of a linear problem*

Degeneracy of eigenvalues is a singularity that usually demands to be raised by introducing perturbations in the most general way. It turns out that, the most general perturbation $\delta\mathcal{L}$ brought to some operator $\mathcal{L}$, *i.e.* $\mathcal{L}' = \mathcal{L} + \delta\mathcal{L}$ may be decomposed in two parts, one that does not change its structure and another one ("true" perturbations) that does something nontrivial, either

by changing its eigenvalues or the fact that it can be turned into diagonal form or not.

Examples of operations that do nothing to a linear system are changes of bases, and more generally invertible transformations $\mathcal{T}$, such that $\mathcal{L}' = \mathcal{T}^{-1}\mathcal{L}\mathcal{T}$, in which case $\mathcal{L}$ and $\mathcal{L}'$ are said to be *similar*. In such operations, the characteristic polynomial is indeed left invariant since:

$$\begin{aligned}\det\left[\left(\mathcal{T}^{-1}\mathcal{L}\mathcal{T}\right) - s\mathcal{I}\right] &= \det\left[\mathcal{T}^{-1}(\mathcal{L} - s\mathcal{I})\mathcal{T}\right] \\ &= \det\left[\mathcal{T}\mathcal{T}^{-1}(\mathcal{L} - s\mathcal{I})\right] = \det(\mathcal{L} - s\mathcal{I})\,.\end{aligned}$$

Only the part of perturbation $\delta\mathcal{L}$ that makes $\mathcal{L}'$ depart from staying similar to $\mathcal{L}$ is of interest. The complementary part, which leaves them similar, is sought for as a transformation $\mathcal{T}$ close to identity:

$$\mathcal{T} = \mathcal{I} + \Gamma\,,$$

i.e., with γ_{ij} small so that $\mathcal{T}$ remains invertible. The transformed operator is then given by

$$\mathcal{L}' = \mathcal{T}^{-1}\mathcal{L}\mathcal{T} = (\mathcal{I} + \Gamma)^{-1}\,\mathcal{L}\,(\mathcal{I} + \Gamma)\,.$$

We need the inverse of $\mathcal{I} + \Gamma$ but, by identification, it can be checked that

$$(\mathcal{I} + \Gamma)^{-1} = \mathcal{I} - \Gamma + \Gamma^2 - \cdots + (-1)^k\Gamma^k + \ldots$$

so that neglecting all terms beyond first order in Γ, by substitution and in full generality we get:

$$\mathcal{L}' = \mathcal{L} + (\Gamma\mathcal{L} - \mathcal{L}\Gamma)\,.$$

The operator $\Gamma\mathcal{L} - \mathcal{L}\Gamma = [\,\Gamma, \mathcal{L}\,]$ is called the *commutator* of $\mathcal{L}$ and Γ. The "true" perturbation is therefore $\delta\mathcal{L} - [\,\Gamma, \mathcal{L}\,]$ and one must next ask for the number of independent parameters on which it depends.

Nontrivial answers are already found in dimension two. Let

$$\mathcal{L}' = \mathcal{L} + \delta\mathcal{L}\,,$$

where $\delta\mathcal{L}$ is represented by the most general matrix:

$$\delta\mathcal{L} = \begin{bmatrix}\delta l_{11} & \delta l_{12} \\ \delta l_{21} & \delta l_{22}\end{bmatrix} = \delta l_{11}[\mathrm{U}_1] + \delta l_{12}[\mathrm{U}_2] + \delta l_{21}[\mathrm{U}_3] + \delta l_{22}[\mathrm{U}_4]\,,$$

where

$$\left\{[\mathrm{U}_1] = \begin{bmatrix}1 & 0 \\ 0 & 0\end{bmatrix};\quad [\mathrm{U}_2] = \begin{bmatrix}0 & 1 \\ 0 & 0\end{bmatrix};\quad [\mathrm{U}_3] = \begin{bmatrix}0 & 0 \\ 1 & 0\end{bmatrix};\quad [\mathrm{U}_4] = \begin{bmatrix}0 & 0 \\ 0 & 1\end{bmatrix}\right\}$$

form a canonical basis of the space of (2×2) matrices. Operator $\mathbf{\Gamma}$ is further defined by

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, \tag{A.10}$$

whose coefficients are adjustable parameters (by contrast with the δl_{ij} which are given).

Let us first consider the case when $\mathcal{L}$ can be turned into diagonal form:

$$\mathcal{L} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}.$$

Computation of the commutator leads to:

$$[\mathbf{\Gamma}, \mathcal{L}] = \begin{bmatrix} 0 & \gamma_{12}(s_2 - s_1) \\ \gamma_{21}(s_1 - s_2) & 0 \end{bmatrix},$$

and therefore

$$\delta\mathcal{L} - [\mathbf{\Gamma}, \mathcal{L}] = \begin{bmatrix} \delta l_{12} & \delta l_{12} - \gamma_{12}(s_2 - s_1) \\ \delta l_{21} - \gamma_{21}(s_1 - s_2) & \delta l_{22} \end{bmatrix}.$$

When $s_1 = s_2$, $\mathcal{L}$ is a multiple of $\mathcal{I}$ and commutes with any $\mathbf{\Gamma}$. The part of $\delta\mathcal{L}$ that derives from a similarity is thus identically zero so that any perturbation $\delta l_{jj'} \neq 0$ modifies the dynamics, generically by raising the degeneracy (as an exercise one can look for cases where only the diagonalizable character is broken)

By contrast when $s_1 \neq s_2$, there exists a continuous two-parameter family of perturbations that gives an operator similar to the original one since one can choose $\gamma_{jj'} = \delta l_{jj'}/(s_{j'} - s_j)$ for $(jj') = (12)$ or (21). True perturbations form a complementary two-parameters family $(\delta l_{11}, \delta l_{22})$ that modifies each eigenvalue separately.

Things are a little less trivial when $\mathcal{L}$ cannot be turned to diagonal form, but only into the Jordan form:

$$\mathcal{L} = \begin{bmatrix} \bar{s} & 1 \\ 0 & \bar{s} \end{bmatrix}$$

The commutator then reads:

$$[\mathbf{\Gamma}, \mathcal{L}] = \begin{bmatrix} -\gamma_{21} & \gamma_{11} - \gamma_{22} \\ 0 & \gamma_{21} \end{bmatrix}$$

and we see that γ_{12} has disappeared, so that there remains only two parameters γ_{21} and $\gamma_{11} - \gamma_{22}$ to cancel as many terms as possible in:

$$\delta\mathcal{L} - [\Gamma, \mathcal{L}] = \begin{bmatrix} \delta l_{11} + \gamma_{21} & \delta l_{12} - (\gamma_{11} - \gamma_{22}) \\ \delta l_{21} & \delta l_{22} - \gamma_{21} \end{bmatrix}.$$

The term δl_{12} can always be suppressed by choosing $\gamma_{11} - \gamma_{22}$ appropriately, and there is enough freedom to cancel one of the two other terms, either δl_{11} or δl_{22}, whereas term δl_{21} can never be suppressed. The most general perturbation that *unfolds the singularity* of this degenerate problem can thus be taken as

$$\mathcal{L}' = \mathcal{L} - \eta\mathcal{U}_3 - \eta'\mathcal{U}_4 \equiv \begin{bmatrix} \bar{s} & 1 \\ -\eta & \bar{s} - \eta' \end{bmatrix}, \tag{A.11}$$

i.e. $\delta l_{21} = -\eta$ and $\delta l_{22} = -\eta'$ where the minus signs are introduced for convenience. The characteristic polynomial then reads:

$$(s - \bar{s})(s - \bar{s} + \eta') + \eta = s^2 - (2\bar{s} - \eta')s + \bar{s}(\bar{s} - \eta') + \eta = 0\,.$$

The discriminant of this quadratic polynomial is $\Delta = \eta'^2 - 4\eta$. How the degeneracy is raised is illustrated in Figure A.2, and one can see that, depending on the path in the parameter space, one can get either real or complex eigenvalues, which account for the dynamics in the vicinity of an improper node (Figure 2.7(b), p. 41): trajectories resemble those close to a focus as well as close to a node (Figure 2.5(a) and Fig. 2.6(a)).

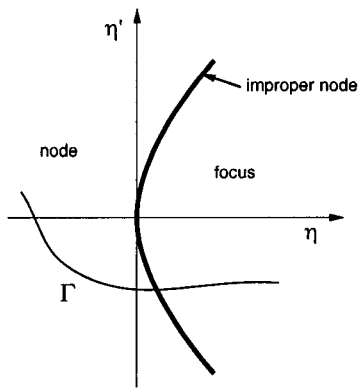

Fig. A.2 Raising the degeneracy at a double non-diagonalizable eigenvalue. The most general perturbation depends on parameters η and η' ($\ll |s|$). Γ is a path in parameter space.

The generalization in dimension d:

$$\begin{bmatrix} \bar{s} & 1 & 0 & \dots & 0 & 0 \\ 0 & \bar{s} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \bar{s} & 1 \\ -\eta_1 & -\eta_2 & -\eta_3 & \dots & -\eta_{d-1} & \bar{s}-\eta_d \end{bmatrix}$$

is called the Jordan–Arnold normal form.

A.3 Metric Properties of Linear Operators

A.3.1 *Scalar products, adjoints, normal and non-normal operators*

The definition of a *scalar product* adds an Euclidean structure to vector spaces on $\mathbb{R}$. A scalar product is a bilinear form with general expression

$$\mathcal{E}(\mathbf{X},\mathbf{Y}) = \sum\nolimits_i \sum\nolimits_j g_{ij} x_i y_j \,,$$

that is symmetric ($g_{ij} = g_{ji}$) and such that $\mathcal{E}(\mathbf{X},\mathbf{X})$ is definite positive, *i.e.* $\mathcal{E}(\mathbf{X},\mathbf{X}) \geq 0$ and $\mathcal{E}(\mathbf{X},\mathbf{X}) = 0$ if and only if $\mathbf{X} = \mathbf{0}$.

The canonical scalar product relative to a given basis simply reads:

$$\mathcal{E}(\mathbf{X},\mathbf{Y}) = \sum\nolimits_i x_i y_j \,.$$

By construction, the basis is orthonormal with respect to the so-defined scalar product since one has: $\mathcal{E}(\mathbf{E}_i,\mathbf{E}_j) = \delta_{ij}$, where δ_{ij} is the Kronecker symbol ($\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$).

The elements of the matrix $[\,\mathrm{T}\,]$ representing the change from an orthonormal basis $\{\mathbf{E}_i\}$ to another $\{\mathbf{F}_j\}$ are given by $t_{ij} = \mathcal{E}(\mathbf{E}_i,\mathbf{F}_j)$, those of the inverse matrix $[\,\mathrm{U}\,]$ by

$$u_{ij} = \mathcal{E}(\mathbf{F}_i,\mathbf{E}_j) = \mathcal{E}(\mathbf{E}_j,\mathbf{F}_i) = t_{ji}.$$

Such matrices are said to be *orthogonal*.

In the case of vector spaces on $\mathbb{C}$, the scalar product is defined through a Hermitian form

$$\mathcal{H}(\mathbf{X},\mathbf{Y}) = \sum\nolimits_i \sum\nolimits_j g_{ij} x_i^* y_j \,,$$

whose coefficients g_{ij} now fulfill $g_{ij} = g_{ji}^*$, so that

$$\mathcal{H}(\mathbf{X}, \mathbf{Y}) = \mathcal{H}(\mathbf{Y}, \mathbf{X})^* .$$

The quadratic form $\mathcal{H}(\mathbf{X}, \mathbf{X})$ must also be definite positive (Hermitian norm). The canonical scalar product relative to a given basis reads:

$$\mathcal{H}(\mathbf{X}, \mathbf{Y}) = \sum\nolimits_i x_i^* y_j \, .$$

The change from an orthonormal basis to another one yields

$$u_{ij} = \mathcal{H}(\mathbf{F}_i, \mathbf{E}_j) = \mathcal{H}(\mathbf{E}_j, \mathbf{F}_i)^* = t_{ji}^* .$$

Such transformations are said to be *unitarian*.

Let us consider the general case of a Hermitian scalar product. An operator $\mathcal{L}$ being given, its *adjoint* $\mathcal{L}^\dagger$ is defined by

$$\mathcal{H}(\mathbf{Y}, \mathcal{L}^\dagger \mathbf{X}) = \mathcal{H}(\mathcal{L}\mathbf{Y}, \mathbf{X}) = \mathcal{H}(\mathbf{X}, \mathcal{L}\mathbf{Y})^* . \tag{A.12}$$

Adjoint operators are thus represented by matrices such that: $(l^\dagger)_{ij} = l_{ji}^*$.

An operator $\mathcal{L}$ is said to be *normal* if it commutes with its adjoint, *i.e.* $\mathcal{L}^\dagger \mathcal{L} \equiv \mathcal{L}\mathcal{L}^\dagger$. It is said to be *self-adjoint* or *Hermitian* if it is identical to its adjoint: $\mathcal{L}^\dagger \equiv \mathcal{L}$.

Normality and hermiticity properties interfere with those arising from the eigenvalue decomposition. It can be shown that a normal operator $\mathcal{L}$ possesses an orthogonal eigen-basis in which it is represented by a diagonal matrix. If in addition it is Hermitian, all its eigenvalues are real.

When the operator is non-normal (*i.e.* does not commute with its adjoint), these properties are lost. Not only the operator can have complex eigenvalues, but there is no orthogonal basis in which it is represented by a diagonal matrix (cf. Jordan normal form). On the other hand, it is possible to find a double series of basis vectors $\{\mathbf{E}_j, \mathbf{F}_j, j = 1, \ldots, d\}$ such that $\{\mathbf{E}_i\}$ is an eigen-basis of $\mathcal{L}$ and $\{\mathbf{F}_j\}$ an eigen-basis of $\mathcal{L}^\dagger$ such that

$$(s_j^* - s_i)\mathcal{H}(\mathbf{F}_j, \mathbf{E}_i) = 0 \, ,$$

i.e. vectors $\mathbf{E}_i$ and $\mathbf{F}_j$ are orthogonal except when s_j and s_i are complex conjugate. The series is said to be *bi-orthogonal*. Unfortunately, this says nothing about scalar products $\mathcal{H}(\mathbf{E}_j, \mathbf{E}_i)$ and $\mathcal{H}(\mathbf{F}_j, \mathbf{F}_i)$, see Exercise A.4.

Remark: Non-normality and transient growth of the energy. Let a linear differential dynamical system be governed by

$$\dot{\mathbf{X}} = \mathcal{L}\mathbf{X} \, ,$$

one can interpret the canonical Hermitian norm

$$2K = \mathcal{H}(\mathbf{X}, \mathbf{X})$$

as an energy. This interpretation is particularly appropriate when considering systems of mechanical origin, whose dependent variables are velocities, *e.g.* fluids governed by the Navier–Stokes equations linearized around some base state for which the kinetic energy $K = \frac{1}{2}\int_{\mathcal{V}}(u'^2 + v'^2 + w'^2)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ contained in a perturbation has all the properties requested for a norm.

On general grounds, asymptotic stability properties (long time limit) of the fixed point at the origin are obtained from the spectrum of eigenvalues s of $\mathcal{L}$ (*i.e.*, stability, neutrality, or instability for $\mathcal{R}e(s) < 0$, $= 0$ or > 0, respectively). When the considered system is unstable, the energy contained in an arbitrary (but small) perturbation may decrease when the trajectory are called back the fixed point along strongly attracting eigendirections but always escape exponentially fast in the unstable subspace since the long time behavior is controlled by eigenvalues with positive real parts. The case of asymptotically stable systems is less transparent because, while they always approach the fixed point at late stage, trajectories may initially depart from it. This going away of trajectories, in the sense of the norm, can be interpreted as a transient growth of the energy for some initial conditions. One can indeed have $\dot{K} > 0$ in some sectors of the tangent phase space, which may be surprising for a stable system. This is because the quadratic form:

$$\dot{K} = \tfrac{1}{2}[\mathcal{H}(\dot{\mathbf{X}}, \mathbf{X}) + \mathcal{H}(\mathbf{X}, \dot{\mathbf{X}})] = \tfrac{1}{2}[\mathcal{H}(\mathcal{L}\mathbf{X}, \mathbf{X}) + \mathcal{H}(\mathbf{X}, \mathcal{L}\mathbf{X})] = \mathcal{R}e(\mathcal{H}(\mathbf{X}, \mathcal{L}\mathbf{X}))$$

is not necessarily definite negative when $\mathcal{L}$ is non-normal, even when all eigenvalues are negative or have negative real parts. A simple example is treated in Exercise A.4.4.

The interest of this property in hydrodynamics derives from the fact that nonlinearities conserve the kinetic energy, so that the initial growth of the energy in the perturbations around the base state is controlled by the linear part of the tangent operator whatever their amplitude.[21] In practice, transient effects are all the more marked that diagonal terms are small when compared to non-diagonal terms. If one notices that diagonal terms mainly account for viscous dissipation, one understands that the situation is highly degenerate at large Reynolds numbers, so that this growth has essentially

[21] D.S. Henningson & S.C. Reddy, "On the role of linear mechanisms in transition to turbulence," Phys. Fluids **6** (1994) 1396-1398.

the same origin as the sub-dominant algebraic evolution called 'secular' in the study of the Jordan normal form. See also the introduction of Chapter 6 and §6.3.4.

A.3.2 *Fredholm Alternative*

Let us now consider the role played by the adjoints in the resolution of inhomogeneous linear problems (as they appear for example in perturbation theory, Chapter 2, p. 55ff):

$$\mathcal{L}\mathbf{X} = \mathbf{F}\,. \tag{A.13}$$

The solution is unique and given by $\mathbf{X} = \mathcal{L}^{-1}\mathbf{F}$ for all $\mathbf{F}$, as long as $\mathcal{L}$ is invertible, that is to say as long as its kernel is trivial ($\mathcal{L}$ of maximal rank). The homogeneous problem

$$\mathcal{L}\mathbf{X} = \mathbf{0} \tag{A.14}$$

then admits only the null vector $\mathbf{X} = \mathbf{0}$ as a solution. When this is not the case, *i.e.* $\mathcal{L}$ is not of maximal rank and has a nontrivial kernel, the adjoint homogeneous problem

$$\mathcal{L}^\dagger\tilde{\mathbf{X}} = \mathbf{0} \tag{A.15}$$

also has nontrivial solutions $\tilde{\mathbf{X}} \neq \mathbf{0}$ (*Fredholm alternative*). The inhomogeneous problem (A.13) then has solutions only if the right hand side $\mathbf{F}$ is orthogonal to the kernel of the adjoint (*Fredholm theorem*). Vector $\tilde{\mathbf{X}}$ being in the kernel of $\mathcal{L}^\dagger$, this reads

$$\mathcal{H}(\tilde{\mathbf{X}}, \mathbf{F}) = 0\,, \tag{A.16}$$

see Exercise A.4.3. When this condition is fulfilled the solution exists but is not unique. This indeterminacy can be raised by imposing, *e.g.* the orthogonality of the solution to the kernel of $\mathcal{L}^\dagger$, which, in perturbations problems presents the solution to the inhomogeneous problem as a true correction (in the sense of the scalar product) to the solution to the homogeneous problem.

A.3.3 *Boundary value problems and adjoint operators.*

Up to now we have considered only finite dimensional spaces. Let us have a look as spaces whose elements are real or complex functions defined on

some interval $[a, b]$. The natural extension of the sum appearing in the definition of the scalar product is now an integral:

$$\mathcal{H}(g, f) = \int_a^b g(x)^* f(x)\, \mathrm{d}x\,.$$

The case of interest to us is when these functions fulfill a differential boundary value problem of order n, with the differential operator $\mathcal{L}$ written as:

$$\mathcal{L}f = \sum\nolimits_{m=0}^{n} a_m(x) f^{(m)}\,,$$

$f^{(m)}$ denoting the mth derivative of f with respect to x and $f^{(0)} \equiv f$. In order to formulate a well posed problem, we must add n boundary conditions:

$$\sum\nolimits_{m=0}^{n-1} \alpha_{jm} f^{(m)}(a) + \beta_{jm} f^{(m)}(b) = u_j\,, \qquad j = 1, \ldots, n\,.$$

In many practical cases, these boundary conditions apply separately at each end of the interval, that is $\beta_{jm} = 0$ when $\alpha_{jm} \neq 0$ and *vice versa*. When $u_j = 0$ for all j, boundary conditions are said to be homogeneous.

Returning to the definition of the adjoint operator (A.12), we see that the expression of $\mathcal{L}^\dagger$ has to be determined from:

$$\int_a^b g^*(x) \mathcal{L}^\dagger f(x)\, \mathrm{d}x = \int_a^b (\mathcal{L} g(x))^*\, f(x)\, \mathrm{d}x\,,$$

by removing g^* from the action of the differential operator on the right hand side of this identity. This is done by means of a series of integrations by parts:

$$\int_a^b a_m^* (g^*)^{(m)} f\, \mathrm{d}x = \left[(g^*)^{(m-1)} a_m^* f \right]_a^b - \int_a^b (g^*)^{(m-1)} (a_m^* f)'\, \mathrm{d}x\,.$$

This operation progressively lowers the order of derivatives acting on g^* and must be pursued down to zero. The formal expression of $\mathcal{L}^\dagger$ is then easily obtained as:

$$\mathcal{L}^\dagger f = \sum\nolimits_{m=0}^{n} (-1)^m (a_m^* f)^{(m)}\,.$$

However, the integrations by parts leaves us with a complicated boundary form made of integrated terms evaluated at the boundaries and involving derivatives of f, the a_m^*, and g^*. Inserting the boundary conditions of the direct problem in this linear form and imposing the cancellation of the residue gives the set of boundary conditions to be applied to the functions of the adjoint problem.

A.4 Exercises

Among the results recalled in this Appendix, several find a nontrivial illustration already in dimension two. The following, extremely elementary exercises may help one better understand the generalizations stated without derivation in the text. Vectors $\mathbf{X}$ are represented in the canonical basis by column arrays:

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and the linear operator $\mathcal{L}$ by a matrix $[\mathrm{L}]$:

$$\mathcal{L} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}, \tag{A.17}$$

The complex extension of the vector space on $\mathbb{R}$ (*i.e.*, $x_1 \in \mathbb{R}$, $x_2 \in \mathbb{R}$) is the set of vectors with complex coefficients (*i.e.*, $x_1 \in \mathbb{C}$, $x_2 \in \mathbb{C}$). When necessary, the canonical scalar product is defined as:

$$\mathcal{H}(\mathbf{Y}, \mathbf{X}) = y_1^* x_1 + y_2^* x_2 \,.$$

A.4.1 *Eigenvalues and eigenvectors, non-degenerate case*

1) Write down the eigenvalue problem $\mathcal{L}\mathbf{X} = s\mathbf{X}$ and the corresponding characteristic equation. Rewrite the latter using the sum S and the product P of the roots ($S = \mathrm{trc}(\mathcal{L}) = l_{11} + l_{22}$, $P = \det(\mathcal{L}) = l_{11}l_{22} - l_{21}l_{12}$. Discuss the number and the nature of the roots in the plane of parameters (S, P) and identify the possible bifurcations upon variation of a control parameter r, $S = S(r)$ and $P = P(r)$. [see Fig. A.3].

2) Identify the domain of parameters corresponding to two real distinct eigenvalues. Compute the eigenvalues, determine the corresponding eigenvectors and the matrix expressing the change from the canonical basis to the eigen-basis. Compute the inverse of this matrix.

3) Determine trajectories of $\dot{\mathbf{X}} = \mathcal{L}\mathbf{X}$ with $\mathbf{X}(0) = \mathbf{X}_0$ in the canonical basis by changing back and forth to the eigen-basis. (As a concrete example, take $\dot{X}_1 = 2X_1 + X_2$, $\dot{X}_2 = 4X_1 - X_2$.)

4) Find the adjoint $\mathcal{L}^\dagger$ to $\mathcal{L}$ for the canonical scalar product (general case, $[\mathrm{L}]$ with coefficients in $\mathbb{C}$). Show that when $[\mathrm{L}]$ has real coefficients, $\mathcal{L}^\dagger$ is normal ($\mathcal{L}\mathcal{L}^\dagger \boxminus \mathcal{L}^\dagger\mathcal{L}$) only when it is symmetric ($l_{12} = l_{21}$) or such that $l_{11} = l_{22}$ and $l_{12} = -l_{21}$. Find the eigenvectors $(\mathbf{E}_1, \mathbf{E}_2)$ of $\mathcal{L}$ and $(\mathbf{F}_1, \mathbf{F}_2)$

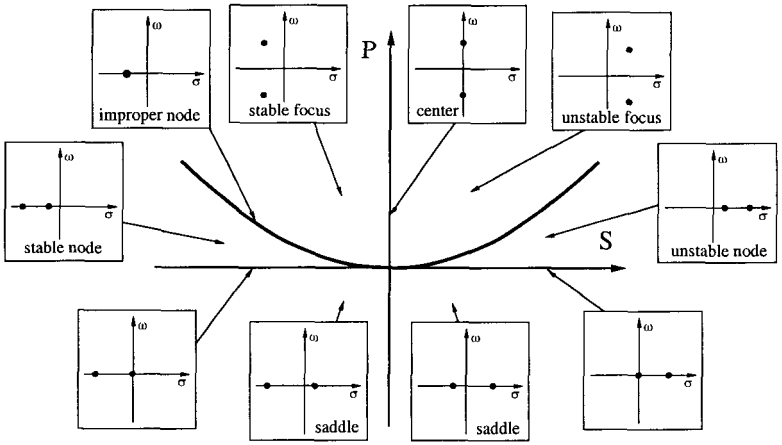

Fig. A.3 Nature of the roots of the characteristic equation of the two-dimensional problem as a function of S and P.

of $\mathcal{L}^\dagger$. Check that, in all cases, $\mathcal{H}(\mathbf{F}_1, \mathbf{E}_2) = 0 = \mathcal{H}(\mathbf{F}_2, \mathbf{E}_1)$ but that one has $\mathcal{H}(\mathbf{E}_1, \mathbf{E}_2) = 0 = \mathcal{H}(\mathbf{F}_1, \mathbf{F}_2)$ only when $\mathcal{L}$ is normal.

5) When there are two complex conjugate eigenvalues, $s = \sigma \pm i\omega$, find a basis in which the system is represented by (2.31), Chapter 2, p. 39.

A.4.2 *Degenerate case and Jordan normal form*

1) Find the condition fulfilled by the l_{ij} when the characteristic polynomial has a double root. Find eigenvectors and determine the condition for any vector of the plane to be an eigenvector.

2) In the general case, this condition is not satisfied and there is only one eigen-direction, generated by a single vector $\mathbf{E}$, the components of which will be determined. Compute the image of an arbitrary vector $\mathbf{X}$ and check that $\mathbf{Y} = (\mathcal{L} - \bar{s}\mathcal{I})\mathbf{X}$ is parallel to $\mathbf{E}$. Derive from this an illustration of the Cayley–Hamilton theorem.

3) Show that in the basis $\{\mathbf{Y} = (\mathcal{L} - \bar{s}\mathcal{I})\mathbf{X}; \mathbf{X}\}$, the matrix representing $\mathcal{L}$ is in canonical Jordan form, which justifies writing a two-dimensional dynamical system with degenerate eigenvalues as (2.35, 2.36), p. 40.

4) Unfold the singularity by setting $s_{(\pm)} = \bar{s} \pm \delta s$, where δs is a small perturbation, and consider the dynamical system:

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} X - 2\bar{s}\, X + \left(\bar{s}^2 - \delta s^2\right) X = \left(\frac{\mathrm{d}}{\mathrm{d}t} - s_{(+)}\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - s_{(-)}\right) X = 0\,,$$

integrate it explicitly by solving successively $\left(\frac{\mathrm{d}}{\mathrm{d}t} - s_{(+)}\right) Y = 0$ and next $\left(\frac{\mathrm{d}}{\mathrm{d}t} - s_{(-)}\right) X = Y$, then show how secular terms associated to linear resonance appear by taking the limit $\delta s \to 0$.

A.4.3 *Fredholm Alternative*

1) Compute the kernel of operator $\mathcal{L}$ in (A.17).

2) Solve the inhomogeneous problem $\mathcal{L}\mathbf{X} = \mathbf{F}$ taken in the form

$$l_{11}x_1 + l_{22}x_2 = f_1\,,$$
$$l_{21}x_1 + l_{22}x_2 = f_2\,,$$

when the kernel of $\mathcal{L}$ is not trivially reduced to the null vector. Observe that the problem has solutions only when the two equations are proportional and express this condition.

3) Determine the adjoint $\mathcal{L}^\dagger$ of $\mathcal{L}$ and its kernel. Then check that the existence condition obtained can be written in the form (A.16).

A.4.4 *Transient growth of the energy*

Consider $\dot{\mathbf{X}} = \mathcal{L}\mathbf{X}$ defined by:

$$\dot{X}_1 = -aX_1 + X_2\,,$$
$$\dot{X}_2 = -bX_2\,,$$

where constants a and b are positive and $b > a$ (already considered in Exercise 4.6.6, p. 172).

1) This system is non-normal, and thus has no orthogonal eigen-basis. Find the eigenvalues and the eigen vectors of $\mathcal{L}$ and its adjoint $\mathcal{L}^\dagger$ (for the canonical scalar product).

2) Consider the energy $K = \frac{1}{2}(X_1^2 + X_2^2)$ and determine the condition on a and b rendering $\dot{K}$ definite negative (monotonic stability). When this condition is not satisfied, identify the region of the (X_1, X_2) plane where $\dot{K} > 0$ and the direction $\rho_{\max} = X_2/X_1$ that maximize the initial growth of the energy. Examine in particular the case $a \ll 1$, $b \ll 1$, so that the behavior of the system is dominated by the out-of-diagonal term.

3) Compute the solution (X_1, X_2) issued from the initial condition $X_1^{(0)}, X_2^{(0)} = \rho X_1^{(0)}$, the evolution of the energy associated to this trajectory and the value of ρ that maximizes $K(t)/K^{(0)}$.

Appendix B

Numerical Approach

Present-day computers make numerical simulations of complex physical systems feasible. Their flexibility and accuracy allow us to gather quantitative information on situations of practical interest but also to obtain more qualitative hints about the nature of interactions at work. In this appendix we focus on the simplest, low-cost, approaches most particularly appropriate to the second aspect, leaving the first one to specialists. We present some classical numerical schemes used in simulations of *initial value problems* defined in terms of *ordinary* (ODEs) or *partial* (PDEs) *differential equations*. The level is kept very elementary and theoretical discussions are avoided, though numerical stability properties derive directly from the application of techniques developed in other parts of the present work, as can be seen by working out the proposed exercises.

In practice, the evolution of the system under consideration is determined at a series of regularly spaced times $t_n = n\,\Delta t$. The corresponding integration methods, appropriate to ODEs as well as PDEs, are introduced in §B.1. The physical-space dependence inherent in PDEs implies a specific coupling between degrees of freedom. This can be treated in the most naive way by estimating the values of the different fields at the nodes of a discrete lattice, $x \to x_j = j\,\Delta x$. Replacing partial derivatives by their *finite differences* yields the numerical schemes presented in §B.2.1. Another description can be achieved in terms of projection of the dynamics on a functional basis spanning the whole physical space. These so-called *spectral* methods are briefly introduced, using Fourier modes in §B.2.2.

Finite element methods and their variants specifically developed for fluid mechanics are left completely aside despite their accuracy and their versatility regarding complex geometries encountered in technical applications for which commercial softwares are fully adapted. Instead of relying on

such black-boxes, we prefer present simple algorithms that can be straightforwardly translated into programming languages such as MATLAB. To our belief, the required investment is light but rewarding in terms of the insight gained in the complexity arising from nonlinearities.

B.1 Treatment of the Time Dependence

We consider the following general initial value problem:

$$\frac{\mathrm{d}}{\mathrm{d}t}v = f(v,t) \qquad \text{with} \quad v(t_0) = v_0\,. \tag{B.1}$$

here written as a non-autonomous ODE but the case of PDEs is the same as far as time dependence is concerned.

Expanding the solution as a Taylor series in the neighborhood of t_n, we get:

$$v_{n+1} = v_n + \Delta t \frac{\mathrm{d}}{\mathrm{d}t}v\Big|_n + \tfrac{1}{2}\Delta t^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2}v\Big|_n + \tfrac{1}{6}\Delta t^3 \frac{\mathrm{d}^3}{\mathrm{d}t^3}v\Big|_n + \cdots \tag{B.2}$$

where the successive derivatives of v, total derivative with respect to t, can be evaluated at t_n recursively, using (B.1):

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}v\big|_n &= f_n\,,\\
\tfrac{\mathrm{d}^2}{\mathrm{d}t^2}v\big|_n &= \tfrac{\mathrm{d}}{\mathrm{d}t}\tfrac{\mathrm{d}}{\mathrm{d}t}v\big|_n = \tfrac{\mathrm{d}}{\mathrm{d}t}f(v,t)\big|_n\\
&= \left(\partial_v f \tfrac{\mathrm{d}}{\mathrm{d}t}v + \partial_t f\right)\big|_n = (\partial_v f_n) f_n + \partial_t f_n\\
\tfrac{\mathrm{d}^3}{\mathrm{d}t^3}\big|_n &= \cdots\,,
\end{aligned}$$

provided that we can evaluate the successive partial derivatives of $f(v,t)$ analytically, which is hardly conceivable beyond the lowest order without the help of formal algebra softwares. The idea is thus to look for numerical approximations of these derivatives.[22]

Truncating expression (B.2) beyond first order, we get the most elementary formula (Euler, cf. Fig. B.1):

$$v_{n+1} = v_n + \Delta t\, f(v_n, t_n)\,. \tag{B.3}$$

This integration scheme is said to be *first order* because it is exact up to correction of order Δt^2, and *explicit* because the knowledge of v and $f(v,t)$ at previous steps is sufficient to compute the solution at the next step.

[22] For elementary formulas consult [Abramowitz and Stegun (1972)]. A detailed presentation can be found in, *e.g.* [Lapidus and Seinfeld (1971)].

Another first order scheme is obtained by evaluating v_n using quantities computed at t_{n+1}:

$$v_n = v_{n+1} - \Delta t \left.\frac{\mathrm{d}}{\mathrm{d}t} v\right|_{n+1} + \frac{1}{2}\Delta t^2 \left.\frac{\mathrm{d}^2}{\mathrm{d}t^2} v\right|_{n+1} - \frac{1}{6}\Delta t^3 \left.\frac{\mathrm{d}^3}{\mathrm{d}t^3}\right|_{n+1} + \ldots , \quad \text{(B.4)}$$

which leads to:

$$v_{n+1} = v_n + \Delta t\, f(v_{n+1}, t_{n+1}) .$$

This scheme, still exact up to corrections $\mathcal{O}(\Delta t^2)$, is termed *implicit* since v_{n+1} is obtained from an implicit equation:

$$v_{n+1} - \Delta t\, f(v_{n+1}, t_{n+1}) = v_n . \quad \text{(B.5)}$$

Subtracting (B.4) from (B.2) and noticing that

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2} v\right|_{n+1} = \left.\frac{\mathrm{d}^2}{\mathrm{d}t^2} v\right|_{n} + \Delta t \left.\frac{\mathrm{d}^3}{\mathrm{d}t^3} v\right|_{n} + \mathcal{O}(\Delta t^2) ,$$

one can see that the term $\mathcal{O}(\Delta t^2)$ disappears, so that:

$$v_{n+1} = v_n + \frac{\Delta t}{2}\big(f(v_n, t_n) + f(v_{n+1}, t_{n+1})\big) + \mathcal{O}(\Delta t^3) . \quad \text{(B.6)}$$

This *implicit second order* formula is called the *Crank–Nicholson* scheme (cf. Fig. B.1).

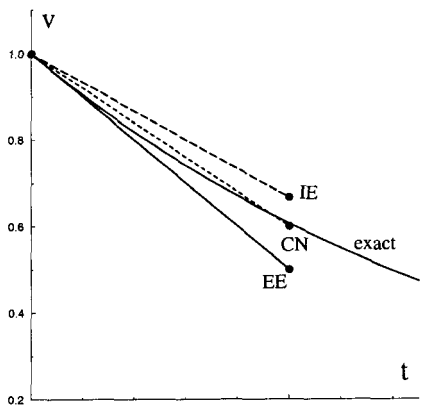

Fig. B.1 Solution of $\frac{\mathrm{d}}{\mathrm{d}t} v = -v$ at $t = \Delta t = 0.5$ with $v(0) = 1$: Exact result compared to estimates from the explicit Euler (EE), implicit Euler (IE) and Crank–Nicholson (CN) formulas, Eqs. (B.3), (B.5), and (B.6), respectively.

Let us compare the error terms using $f(v) = -v$ with $v_0 = 1$ as a simple example. We get $v = \exp(-t)$ and after a time step

$$\begin{aligned}
\text{Exact } (\exp(-\Delta t)) : &\quad 1 - \Delta t + \tfrac{1}{2}\Delta t^2 - \tfrac{1}{6}\Delta t^3 + \dots \\
\text{Explicit Euler (B.3)} : &\quad 1 - \Delta t \\
\text{Implicit Euler (B.5)} : &\quad \frac{1}{1+\Delta t} = 1 - \Delta t + \Delta t^2 - \Delta t^3 + \dots , \\
\text{Crank–Nicholson (B.6)} : &\quad \frac{1-\frac{1}{2}\Delta t}{1+\frac{1}{2}\Delta t} = 1 - \Delta t + \tfrac{1}{2}\Delta t^2 - \tfrac{1}{4}\Delta t^3 + \dots ,
\end{aligned}$$

which clearly shows the improvement brought by the Crank–Nicholson scheme (cf. exercise B.3.1). The obvious computational drawback of implicit iteration schemes, largely compensated by better stability properties as shown later, is that the solution at t_{n+1} is not expressed simply in terms of the solution at t_n. The inversion is trivial only if v is a single scalar variable and f is linear as in the example chosen. When this is not the case, specific techniques are required, see §B.2.1.

Up to now we have used only *one-step* formulas, requiring the knowledge of the solution at a single time. A simple improvement of (B.3) is obtained by inserting:

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} v \Big|_n = \frac{1}{\Delta t} \left(\frac{\mathrm{d}}{\mathrm{d}t} v|_n - \frac{\mathrm{d}}{\mathrm{d}t} v|_{n-1} \right) + \mathcal{O}(\Delta t)$$

in (B.2), which leads to the second-order *Adams–Bashford* formula:

$$v_{n+1} = v_n + \Delta t \left(\tfrac{3}{2} f(v_n, t_n) - \tfrac{1}{2} f(v_{n-1}, t_{n-1}) \right) , \tag{B.7}$$

exact up to $\mathcal{O}(\Delta t^3)$. This formula is one of the simplest examples of *multi-step* schemes, computing v_{n+1} from the knowledge of v at several previous times, here t_n and t_{n-1}.

An equally simple possibility is to write (B.4) at times t_{n-1} and t_n instead of t_n and t_{n+1}, which yields

$$v_{n-1} = v_n - \Delta t \, \frac{\mathrm{d}}{\mathrm{d}t} v|_n + \tfrac{1}{2}\Delta t^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} v|_n - \tfrac{1}{6}\Delta t^3 \frac{\mathrm{d}^3}{\mathrm{d}t^3} v|_n + \dots ,$$

so that after subtraction of (B.2) one gets another second order formula, also exact up to $\mathcal{O}(\Delta t^3)$:

$$v_{n+1} = v_{n-1} + 2\,\Delta t\, f(v_n, t_n) , \tag{B.8}$$

called the *leap-frog* scheme.

In addition to accuracy requirements, the choice of an integration scheme is also subjected to stability requirements. As a matter of fact, the numerical scheme must not amplify round-off errors so that the solution obtained is the physical one and not some parasitic solution. The problem is important since when, for example, one uses a two-step formula to integrate a first-order differential equation, one obtains an iteration that starts with two initial conditions instead of one for the time-continuous problem. The added mode must be damped in order that the simulation keeps its meaning. Let us compare schemes (B.7) and (B.8) used to integrate $\frac{\mathrm{d}}{\mathrm{d}t}v = -v$:

- Iteration (B.7) reads

$$v_{n+1} = v_n - \Delta t\left(\tfrac{3}{2}v_n - \tfrac{1}{2}v_{n-1}\right)$$

or, better,

$$\begin{aligned} u_{n+1} &= v_n\,, \\ v_{n+1} &= \tfrac{1}{2}\Delta t\, u_n + \left(1 - \tfrac{3}{2}\Delta t\right)v_n\,, \end{aligned}$$

which, from standard linear stability analysis (Exercise B.3.2), leads to the characteristic equation:

$$s^2 - s\left(1 - \tfrac{3}{2}\Delta t\right) - \tfrac{1}{2}\Delta t = 0\,.$$

At lowest order in Δt, the two roots are $s_{(+)} = 1 - \Delta t$ and $s_{(-)} = -\frac{1}{2}\Delta t$. The first one clearly corresponds to the physical solution and the second one to the numerical mode. Recalling that the differential equation is here replaced by a discrete-time system like those introduced in Chapter 3 (see in particular Figure 4.10, p. 140) we obtain that the latter mode is damped provided that $|s_{(-)}| < 1$, which can be achieved by taking Δt sufficiently small, here $\Delta t < 2$. This does not mean that the numerical solution obtained with Δt just below the limit is a good approximation, but just that it is not polluted by the numerical mode. The Adams–Bashford scheme is said to be *conditionally stable*.

- Iteration (B.8) reads:

$$v_{n+1} = v_{n-1} - 2\,\Delta t\, v_n,$$

that is:

$$\begin{aligned} u_{n+1} &= v_n\,, \\ v_{n+1} &= u_n - 2\,\Delta t\, v_n\,, \end{aligned}$$

which leads to

$$s^2 + 2\,\Delta t\, s - 1 = 0\,,$$

so that $s_{(+)} \simeq 1 - \Delta t$ as before. But now one has $s_{(-)} \simeq -1 - \Delta t$ and since $|s_{(-)}| > 1$, the numerical mode is always unstable. The leap-frog scheme is thus always *unconditionally unstable.* Since $s_{(-)}$ is negative, successive iterates alternate from positive to negative values. This subharmonic instability is due to the fact that the scheme leaves even time steps uncoupled to odd ones. Such a behavior is frequently the signature of a *numerical instability* (see later Figure B.16). Scheme (B.8) is however little demanding and its use is not forbidden provided the growth of errors is controlled. In general, the numerical mode is tamed by averaging over two iterations every p computation, with $p \gg 1$ but still small enough, so that its amplitude remains small when compared to the truncation errors.

High order accuracy can be reached with sophisticated multi-step formulas. Their drawback stems from the need to generate numerical initial conditions from the physical ones, especially when the time step has to be changed for accuracy reasons. Explicit formulas of *Runge–Kutta* type avoid this problem. The idea is to use the basic first-order Euler formula but with a better estimate of the slope. Second order accuracy is obtained by averaging the slopes evaluated at t_n and t_{n+1} but the resulting (Crank–Nicholson) scheme is implicit. To avoid such a difficulty one can first estimate v_{n+1} with the first-order explicit formula:

$$v' = v_n + \Delta t\, f(v_n, t_n)\,, \tag{B.9}$$

then compute the slope at this point $f(v'_{n+1}, t_{n+1})$, and take the average with the slope at (v_n, t_n) as an improved guess. This yields:

$$v_{n+1} = v_n + \tfrac{1}{2}\Delta t\big(f(v_n, t_n) + f(v', t_{n+1})\big)\,. \tag{B.10}$$

The correction is sufficient to achieve second order accuracy while keeping an explicit scheme (Exercise B.3.1). This case is illustrated in Figure B.2, left.

Another way to achieve the same goal is to introduce an intermediate time $t_{n+1/2}$, to compute v' with the first-order formula

$$v' = v_n + \tfrac{1}{2}\Delta t f(v_n, t_n)\,, \tag{B.11}$$

and to use the slope at that point to extrapolate the solution at time t_{n+1},

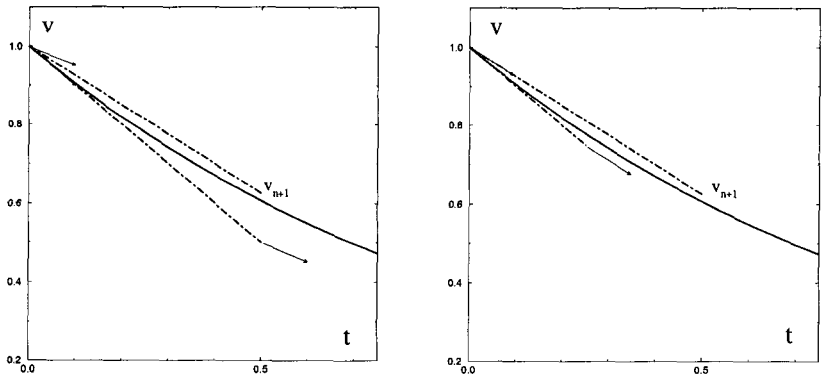

Fig. B.2 Achieving second order accuracy with Runge–Kutta formulas (B.9, B.10) and (B.11, B.12).

as shown in Figure B.2, right:

$$v_{n+1} = v_n + \Delta t\, f(v', t_{n+1/2})\,. \tag{B.12}$$

High-order formulas combine these two basic techniques to get the best estimate of the slope. The classical fourth-order scheme

$$v_{n+1} = v_n + \tfrac{1}{6}\Delta t\big(f(v_n, t_n) + 2f(v', t_{n+1/2}) + 2f\big(v'', t_{n+1/2}\big) + f(v''', t_{n+1})\big) \tag{B.13}$$

involves three intermediate estimates:

$$\begin{aligned} v' &= v_n + \tfrac{1}{2}\Delta t f(v_n, t_n)\,, \\ v'' &= v_n + \tfrac{1}{2}\Delta t f(v', t_{n+1/2})\,, \\ v''' &= v_n + \Delta t\, f\big(v'', t_{n+1/2}\big)\,. \end{aligned}$$

One full time step thus requires four evaluations of f, which may be numerically demanding, but the scheme stays explicit and starts with a single initial condition so that the time step can easily be changed.

Predictor-corrector methods are multi-step schemes that are often preferred to Runge–Kutta methods because they require a smaller number of evaluations of f to reach the same accuracy though it is more difficult to change the time step. For more information, consult [Lapidus and Seinfeld (1971)] or [Press *et al.* (1986)].

Explicit methods can easily be extended to treat differential systems coupling several variables, see the examples considered in §B.4.

B.2 Treatment of Space Dependence in PDEs

B.2.1 *Finite difference methods*

The derivation of an integration scheme is made of two main steps. First, a consistent approximation of space differential operators has to be found. Next a time-integration scheme is chosen. The resulting space-time iteration is then studied for stability. A supplementary step would be to demonstrate convergence in the limit of infinitesimal space and time steps. For *linear* partial differential problems, it can be shown that *consistence* and *stability* implies *convergence.* Here we do worry about convergence and are happy with numerical solutions that look "physical."

We consider only systems that extend in one space direction. The solution is sought for at regularly spaced nodes $\{x_j\}$ of a periodic lattice with spacing $\Delta x = \ell/N$ where ℓ is the length of the domain and $N+1$ the number of points, including end points. The numerical solution $v_{n,j}$ is thus noted with two subscripts but the first one, corresponding to time, will be dropped where it is not necessary.

B.2.1.1 *Space discretization and consistence*

Finite difference approximations of space derivatives are obtained in the same way as those of time derivatives.[23] For example, combination of two first order approximations of the first order derivatives:

$$\partial_x v|_j = (v_{j+1} - v_j)/\Delta x \qquad \text{and} \qquad \partial_x v|_j = (v_j - v_{j-1})/\Delta x \tag{B.14}$$

yields the second order centered formula:

$$\partial_x v|_j = (v_{j+1} - v_{j-1})/2\Delta x\,. \tag{B.15}$$

In the same way, the expression of the second derivative at order Δx^2 reads:

$$\partial_{xx} v|_j = (v_{j+1} - 2v_j + v_{j-1})/\Delta x^2\,, \tag{B.16}$$

Higher order accuracy requires more points, *e.g.* five at fourth order:

$$\partial_{xx} v|_j = (-v_{j+2} + 16v_{j+1} - 30v_j + 16v_{j-1} - v_{j-2})/12\Delta x^2. \tag{B.17}$$

The second-order approximation of the fourth derivative reads:

$$\partial_{xxxx} v|_j = (v_{j+2} - 4v_{j+1} + 6v_j - 4v_{j-1} + v_{j-2})/\Delta x^4\,. \tag{B.18}$$

[23] Useful formulas are found in [Abramowitz and Stegun (1972)]. A complete presentation is given by [Richtmyer and Morton (1967)]. Consult also [Acton (1970)].

The order of *consistency* of some finite difference approximation is the value of the exponent that controls the decay of the global truncation error when Δx decreases. It serves nothing to approximate different derivatives appearing in a model by formulas at different orders. In practice, the representation of differential operators by finite differences is not very good since truncation errors decay only algebraically, as a power of Δx. From this point of view spectral methods briefly introduced in Section B.2.2 behave in a much better way.

B.2.1.2 *Boundary conditions*

Consistency considerations also affect the treatment of boundary conditions when they involve derivatives (Neumann conditions). Since it is in general easy to achieve consistency at order Δx^2 for interior points —for example by preferring (B.15) to (B.14) for ∂_x— it would be a pity to spoil the quality of an approximation by a bad account of boundary conditions, even if their effects do not propagate beyond some narrow "numerical boundary layer."

Consider for example the condition $\partial_x v = 0$ at one end point of the x interval. Using (B.14) we get an expression accurate at order Δx, which simply implies $v_j = v_{j+1}$, with $j = 0$ and 1 for the boundary point and the first interior point, respectively. To achieve second order consistency, we can use:

$$\partial_x v|_j = \left(-3v_j + 4v_{j+1} - v_{j+2}\right)/2\Delta x \tag{B.19}$$

so that canceling $\partial_x v$ at the boundary gives $3v_0 - 4v_1 + v_2 = 0$. Another possibility is to add a fictitious point outside, $j = -1$ and associate a variable v_{-1} whose value is forced to that of the first interior point at all times, $v_{-1} = v_1$, as dictated by formula (B.15).

B.2.1.3 *Time discretization and stability*

For more specificity we consider the one-dimensional *diffusion equation*:

$$\partial_t v = \partial_{xx} v\,, \tag{B.20}$$

with boundary conditions that need not be given here. The temporal scheme is obtained according to one the recipes of section B.1 (here we restore subscript n). From the spatial viewpoint we assume second order consistency, *i.e.* using (B.16) to discretize $\partial_{xx} v$. The most important question is that of the numerical stability of the scheme, as determined from the growth rate of the "noise" generated by truncation and round-off

errors. Here, we only compare the case of the two simplest schemes deriving from (B.3) and (B.5). In the *explicit* case, the space derivative $\partial_{xx}v$ is evaluated at t_n, that is:

$$v_{n+1,j} = v_{n,j} + \frac{\Delta t}{\Delta x^2}\left(v_{n,j+1} - 2v_{n,j} + v_{n,j-1}\right), \tag{B.21}$$

while it is evaluated at t_{n+1} in the *implicit* case, which yields:

$$-\frac{\Delta t}{\Delta x^2}v_{n+1,j+1} + \left(1 + \frac{2\Delta t}{\Delta x^2}\right)v_{n+1,j} - \frac{\Delta t}{\Delta x^2}v_{n+1,j-1} = v_{n,i}\,. \tag{B.22}$$

The explicit method is straightforward and requires less computational work than the implicit method that requires the inversion of a linear system with a band structure and only three near-diagonal terms. As we will see, this amount of work is rewarding.

The stability analysis of finite-difference systems is performed in the same way as that of physical systems (see Chap. 2, §3.1.3 and Chap. 3, §4.2). The evolution of a perturbation $\delta v_{n,j}$ to a given numerical solution $v_{j,n}$ is studied by means of a discrete Fourier transform:

$$\delta v_{n,j} = \delta\tilde{v}_{n,k}\exp(ikj)\,.$$

Inserting this expression into equations (B.21) and (B.22), we obtain:

$$\text{Explicit:} \qquad \delta\tilde{v}_{n+1,k} = \left(1 - \frac{2\,\Delta t}{\Delta x^2}\big(1 - \cos(k)\big)\right)\delta\tilde{v}_{n,k}\,, \tag{B.23}$$

$$\text{Implicit:} \qquad \left(1 + \frac{2\Delta t}{\Delta x^2}\big(1 - \cos(k)\big)\right)\delta\tilde{v}_{n+1,k} = \delta\tilde{v}_{n,k}\,, \tag{B.24}$$

both of the general form:

$$\delta\tilde{v}_{n+1,k} = \xi(k)\,\delta\tilde{v}_{n,k}\,,$$

and, as we know, the perturbation grows when $|\xi(k)| > 1$ and decays when $|\xi(k)| < 1$.

For the explicit scheme, this gives:

$$1 - \frac{4\,\Delta t}{\Delta x^2} \le \xi_{\text{exp}} = 1 - \frac{2\,\Delta t}{\Delta x^2}\big(1 - \cos(k)\big) \le 1\,,$$

The upper bound condition brings nothing, whereas the lower bound implies $-1 < \xi_{\text{exp}}(\pi)$ for stability. This leads to:

$$-1 < 1 - \frac{4\,\Delta t}{\Delta x^2} \quad \Rightarrow \quad \frac{4\,\Delta t}{\Delta x^2} < 2 \quad \Rightarrow \quad \Delta t < \frac{\Delta x^2}{2}\,,$$

which means that the time step must be small enough at given spacing: the explicit scheme is thus *conditionally stable*.

By comparison, the growth rate of perturbations with the implicit scheme:

$$\xi_{\text{imp}} = \left(1 + \frac{2\,\Delta t}{\Delta x^2}\big(1 - \cos(k)\big)\right)^{-1}$$

is such that $0 < \xi_{\text{imp}} \leq 1$ for all k since $\big(1 - \cos(k)\big) \geq 0$. This scheme is therefore *unconditionally stable*, so that Δt can be chosen for temporal accuracy requirement only, whatever the value of Δx.

For the explicit scheme the most unstable perturbations are those with $k = \pi$. Going back to physical space, one finds $\delta v_{n,j} = \delta \bar{v}_n\,(-1)^j$: the behavior of the solution at even nodes is the opposite of that at odd nodes, which is typical of numerical instabilities. At this stage, the instability can be tamed by decreasing the time step until the stability condition is fulfilled but it is better to turn to an implicit scheme, especially when higher derivatives are involved (Exercise B.3.3). Higher order time discretization would be studied in the same way.

B.2.1.4 *Efficient treatment of implicit schemes*

The supplementary work needed to find the solution at time step t_{n+1} by the implicit scheme (B.22) is easily achieved by a simple numerical "double sweep" method since the matrix to be inverted has only a few terms close to the diagonal. This is in fact a special case of the classical "*LU*" decomposition used to solve linear systems.

Let the initial system be $\mathcal{M}\mathbf{V} = \mathbf{F}$, the operator $\mathcal{M}$ is then written as the product $\mathcal{L}\mathcal{U}$ of two operators, $\mathcal{L}$ represented by a lower triangular matrix $[\,\mathrm{L}\,]$ with elements $l_{ij} = 0$ when $i < j$, and an operator $\mathcal{U}$ represented by an upper triangular matrix $[\,\mathrm{U}\,]$ such that $u_{ij} = 0$ when $i > j$. The decomposition is unique upon requiring that the diagonal elements of $[\,\mathrm{U}\,]$ are scaled to unity, $u_{ii} = 1$. The starting matrix $[\,\mathrm{M}\,]$ is three-diagonal, *i.e.* $m_{ij} \neq 0$ for $j = i$ and $j = i \pm 1$, $[\,\mathrm{L}\,]$ and $[\,\mathrm{U}\,]$ have a band structure with two non-zero diagonals only: $l_{ij} \neq 0$ for $j = i$ and $j = i - 1$, $u_{jj} = 1$, and $u_{ij} \neq 0$ for $j = i + 1$.

The problem which reads:

$$(\mathcal{L}\mathcal{U})\,\mathbf{V} = \mathcal{L}\,(\mathcal{U}\,\mathbf{V}) = \mathcal{L}\,\mathbf{W} = \mathbf{F}$$

is then replaced by two equations

$$\mathcal{L}\,\mathbf{W} = \mathbf{F} \qquad \text{and} \qquad \mathcal{U}\,\mathbf{V} = \mathbf{W}\,,$$

that are solved through two first-order recursion relations, first a *forward* iteration to obtain $\mathbf{W}$:

$$\begin{aligned} l_{11} w_1 &= f_1\,, \\ l_{i\,i-1} w_{i-1} + l_{ii} w_i &= f_i\,, \qquad i = 2, \ldots, N\,, \end{aligned}$$

and second a *backward* iteration to get $\mathbf{V}$ from $\mathbf{W}$:

$$\begin{aligned} v_N &= w_N\,, \\ v_i + u_{i\,i+1} v_{i+1} &= w_i\,, \qquad i = N-1, \ldots, 1\,. \end{aligned}$$

where the coefficients l_{ij} and u_{ij} are computed once for all using two simple first-order recursion relations. The general term of the product reads $\sum_k l_{ik} u_{kj} = m_{ij}$, with two special cases for $i = 1$ and $j = N$. One readily obtains

$$\begin{aligned} l_{11} &= m_{11}\,, \\ l_{11} u_{12} &= m_{12}\,, \end{aligned}$$

then for $i = 2$ to $N - 1$

$$\begin{aligned} l_{ii-1} &= m_{ii-1}\,, \\ l_{ii-1} u_{i-1\,i} + l_{ii} &= m_{ii}\,, \\ l_{ii} u_{ii+1} &= m_{ii+1}\,, \end{aligned}$$

and finally for $i = N$

$$\begin{aligned} l_{NN-1} &= m_{NN-1}\,, \\ l_{NN-1} u_{N-1N} + l_{NN} &= m_{NN}\,. \end{aligned}$$

These relations are easily implemented numerically and can be generalized to the case with any finite and small number of elements close to the diagonal. This happens for example when the model contains fourth order derivatives, hence five terms from (B.18), which leads to similar but second-order recursion relations, see exercise B.3.3 and B.4.4.2 focusing on the Swift–Hohenberg model.

For problems that are two-dimensional in space, implicit schemes lead to linear systems where the matrix [M] is sparse with a block-diagonal structure. Instead of generalizing the previous algorithms to block matrices,

which is possible but requires a lot of storage, one usually turns to iterative methods. The description of corresponding algorithms, over-relaxation, conjugate-gradients, *etc.*, goes beyond our present purpose.[24]

B.2.1.5 *Treatment of nonlinear terms*

Up to now we have considered linear problems. In most cases of interest, nonlinearities are present, so let us write formally $\partial_t v = \mathcal{L}v + \mathcal{N}(v)$. The linear part is often the most dangerous part from the point of view of numerical stability since it usually contains partial derivatives of higher order than the nonlinear part, the dangerous character of which comes from other sources as we know. In order to preserve second order accuracy while avoiding a fully implicit treatment of nonlinearities, one can develop *quasi-linearized* schemes of the Crank–Nicholson type by replacing v_{n+1} by $v_n + (v_{n+1} - v_n)$ in the nonlinear term, expanding it to first order in $(v_{n+1} - v_n)$ assumed to be small enough:

$$\mathcal{N}(v)\big|_{n+1} \approx \mathcal{N}(v)\big|_n + \mathcal{Q}_n\big(v_{n+1} - v_n\big)\,,$$

where $\mathcal{Q}_n = \delta\mathcal{N}/\delta v\big|_n$, the functional derivative of $\mathcal{N}$ with respect to v is a coefficient evaluated at time t_n. When nonlinearities are not numerically dangerous, another possibility is to treat the linear and nonlinear terms separately, using an implicit Crank–Nicholson scheme for the linear part and an explicit Adams–Bashford scheme for the nonlinear part

$$v_{n+1} = v_n + \tfrac{1}{2}\Delta t\left[\big(\mathcal{L}v\big|_{n+1} + \mathcal{L}v\big|_n\big) + \big(3\mathcal{N}(v)\big|_n - \mathcal{N}(v)\big|_{n-1}\big)\right].$$

B.2.2 *Spectral methods*

Spectral methods rest on a conversion of the partial differential problem into an ordinary differential problem for the amplitudes of modes obtained by projection onto a complete functional basis. For simplicity, we consider here only the case of periodic boundary conditions at the ends of an interval of length ℓ, which allows one to use Fourier modes:

$$v(x) = \sum_{m\in\mathbb{Z}} \hat{v}_m \exp(2\pi i m x/\ell)\,, \tag{B.25}$$

[24]The interested reader is invited to consult specialized references, *e.g.* J.H. Wilkinson, "Solution of linear algebraic equations and matrix problems by direct methods," Chapter 2.2 in *Digital computer user's handbook*, M.K. Klerer and G.A. Korn, Eds. (McGraw-Hill, 1967).

so that, when $v \in \mathbb{R}$, one must have $\hat{v}_{-m} = \hat{v}_m^*$. The coefficients of the Fourier series are given by:

$$\hat{v}_m = \frac{1}{\ell} \int_0^\ell v(x) \exp(-2\pi i m\, x/\ell)\, \mathrm{d}x\,, \quad m \in \mathbb{Z}\,. \tag{B.26}$$

We recall that:

$$\frac{1}{\ell} \int_0^\ell \exp\left(2\pi i(m' - m)x/\ell\right) \mathrm{d}x = \delta_{mm'}\,, \tag{B.27}$$

where $\delta_{mm'}$ is the Kroneker symbol (= 1 when $m = m'$ and zero otherwise), and that, reciprocally,

$$\sum_{m=-\infty}^{+\infty} \exp\left(2\pi i m(x - x')/\ell\right) = \delta_{\mathrm{D}}(x - x')\,, \tag{B.28}$$

where $\delta_{\mathrm{D}}(x)$ is the Dirac distribution (generalized function) such that

$$\int_{-\infty}^{+\infty} f(x')\delta_{\mathrm{D}}(x - x')\, \mathrm{d}x = f(x)\,.$$

For the diffusion equation (B.20), the evolution of the solution is easily obtained by integrating the differential equations for the Fourier coefficients $\hat{v}_m = \hat{v}_m(t)$:

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{v}_m = -k_m^2 \hat{v}_m \qquad \text{with} \qquad k_m = 2\pi m/\ell\,,$$

which yields

$$\hat{v}_m(t) = \hat{v}_m(0) \exp(-k_m^2 t)\,, \quad m \in \mathbb{Z}.$$

Let us notice that, here, the differential operator is diagonal in Fourier space. This may not be the case with other boundary conditions requiring other basis functions, *e.g.* Chebyshev polynomials, but, in general, an exact representation of the differential operator in spectral space can still be found.

In practice, the series is truncated beyond some maximal value $N/2$, so that $\sum_{m=-\infty}^{+\infty}$ is replaced by $\sum_{-N/2+1}^{+N/2}$, see Fig. B.3(a, b). This truncation introduces some approximation forbidding us to resolve the dynamics of structures with wavelengths shorter than $\lambda_{\min} = 2\pi/k_{\max} = 2\ell/N$, which can be understood as the result of sampling the solution on a regular lattice with spacing $\Delta x = \ell/N$. One may say that the method is of *infinite order* since the error term decreases exponentially with N, as $\exp\left(-(N\pi/\ell)^2 t\right)$ at given t, *i.e.* faster than any power of Δx.

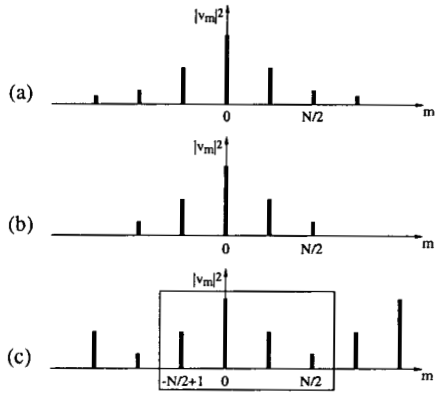

Fig. B.3 (a) Spectrum $|\hat{v}_m|^2$ of a periodic function. (b) Same spectrum, truncated beyond $N/2$. (c) Discrete Fourier transform of the function that has the same value as the solution at the collocation points.

In fact we have not yet said how integral (B.26) is computed. In practice the solution is only represented at the nodes of the lattice mentioned above and called the *collocation points*. This allows us to pass from continuous Fourier transforms to discrete ones.[25] Let $\{x_j = j\Delta x; j = 0, \ldots, N-1\}$ be the set of these points, (B.26) is then modified to read

$$\hat{v}_m^{(\mathrm{d})} = \frac{1}{N} \sum_{j=0}^{N-1} v_j \exp\left(-2\pi i \frac{mj}{N}\right), \tag{B.29}$$

while the solution (B.25) is written as

$$v_j = \sum_{m=-N/2+1}^{+N/2} \hat{v}_m^{(\mathrm{d})} \exp\left(2\pi i \frac{mj}{N}\right). \tag{B.30}$$

Notice that, depending on the computer program, the $1/N$ factor may be placed in (B.30) or in (B.29), or one can find a factor $1/\sqrt{N}$ in each.

Relations (B.27, B.28) then read:

$$\frac{1}{N} \sum_{j=0}^{N-1} \exp\left(2\pi i \frac{(m'-m)j}{N}\right) = \delta_{mm'}$$

[25] A brief but nice introduction to discrete Fourier transforms is given by R.J. Higgins, "Fast Fourier transform: an introduction with some mini computer experiments," *Am. J. Physics* **44** (1976) 766–773. For more information, of course consult [Gottlieb and Orszag (1977)] or [Canuto *et al.* (1988)].

and

$$\sum_{m=-N/2+1}^{N/2} \exp\left(2\pi i \frac{m(j'-j)}{N}\right) = N\delta_{jj'} .$$

It can be noticed that the series of Fourier coefficients $\hat{v}_m^{(\mathrm{d})}$ is periodic with period N since:

$$\hat{v}_{m+N}^{(\mathrm{d})} = \sum_{j=0}^{N-1} v_j \exp[-2\pi i(m+N)j/N] \boxminus \sum_{j=0}^{N-1} v_j \exp(-2\pi i\, mj/N) = \hat{v}_m^{(\mathrm{d})}, \tag{B.31}$$

whereas $\hat{v}_m$ is not periodic, as illustrated in Figure B.3. When v is a real quantity, it is also observed that not only $\hat{v}_0^{(\mathrm{d})}$ is real (like $\hat{v}_0$), but that $\hat{v}_{N/2}^{(\mathrm{d})}$ is also real since one has $\left(\hat{v}_{N/2}^{(\mathrm{d})}\right)^* = \hat{v}_{-N/2}^{(\mathrm{d})} = \hat{v}_{-N/2+N}^{(\mathrm{d})} = \hat{v}_{N/2}^{(\mathrm{d})}$. The Fourier transform thus changes the N real quantities v_j, $j = 0, \ldots, N-1$, into a set of 2 real quantities, $\hat{v}_0^{(\mathrm{d})}$ and $\hat{v}_{-N/2}^{(\mathrm{d})}$, and $N/2-1$ complex quantities $\hat{v}_m^{(\mathrm{d})}$, $m = 1, \ldots, N/2-1$, so that the total number of degrees of freedom describing the solution is preserved.

The periodicity of the discrete spectrum induces what is known as the *aliasing phenomenon*: two Fourier modes distant by N in the spectrum are aliases of each other. This is another way of saying that, having no information on modes with $m > N/2$, one cannot resolve structures with scales smaller than the lattice spacing.

The important point is that we know how to go from v_j to $\hat{v}_m^{(\mathrm{d})}$ and *vice versa* by means of "fast" algorithms with operation counts of order $N \log N$ and not N^2. The *Fast Fourier Transform* (FFT) takes advantage of the standard trigonometric relations to organize the information flow so as to minimize the number of arithmetic operations. Commercial softwares such as MATLAB are effective in computing FFTs when N is a power of two but also when N can be decomposed as $N = 2^{n_2}3^{n_3}5^{n_5}\ldots$ with prime factors that are not too large. This possibility of a fast passage from physical to spectral space and back allows one to treat nonlinearities in a clever way by means of so-called *pseudo-spectral* methods, provided that aliasing is appropriately dealt with.

The difficulty with nonlinearities is that products in physical space are transformed into convolution sums in spectral space, as exemplified here using quadratic nonlinearities, *i.e.* a product vw. For the exact solution

we have:

$$\widehat{vw}_m = \sum_{m'=-\infty}^{+\infty} \hat{v}_{m-m'} \hat{w}_{m'} ,$$

which derives from the Fourier analysis of

$$\begin{aligned} v(x)w(x) &= \left(\sum_{m'=-\infty}^{+\infty} \hat{v}_{m'} \exp(2\pi i\, m' x/\ell) \right) \left(\sum_{m''=-\infty}^{+\infty} \hat{w}_{m''} \exp(2\pi i\, m'' x/\ell) \right) \\ &= \sum_{m'=-\infty}^{+\infty} \sum_{m''=-\infty}^{+\infty} \hat{v}_{m'} \hat{w}_{m''} \exp[2\pi i\, (m'+m'')x/\ell] , \end{aligned}$$

see Figure B.4(a).

In the perspective of a truncation beyond $\pm N/2$, one should only keep the combinations such that m', m'', and $m = m' + m''$ belong to the considered interval (Fig. B.4(b), gray-shaded region). The computation of the Fourier transform of a formally quadratic term thus involves N sums with N^2 products each. A clever way to get around the rapid growth of the operation count with N is to take advantage of fast transforms to compute the product in physical space instead staying in the spectral space, with a reduction factor of order $\log(N)/N$.

Things are however not simple since one must take care of aliasing. If nothing special is done, the Fourier evaluation of a product computed in physical space contains spurious terms (Fig. B.4c). Let us indeed start with:

$$v_j = \sum_{m=-N/2+1}^{N/2} \hat{v}_m^{(\mathrm{d})} \exp\left(2\pi i \frac{mj}{N}\right), \qquad w_j = \sum_{m'=-N/2+1}^{N/2} \hat{w}_{m'}^{(\mathrm{d})} \exp\left(2\pi i \frac{m'j}{N}\right),$$

next define :

$$\widehat{vw}_m^{(\mathrm{d})} = \frac{1}{N} \sum_{j=0}^{N-1} (v_j w_j) \exp\left(-2\pi i \frac{mj}{N}\right)$$

and expand the right hand side. We get:

$$\widehat{vw}_m^{(\mathrm{d})} = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{m'=-N/2+1}^{N/2} \sum_{m''=-N/2+1}^{N/2} \hat{v}_{m'}^{(\mathrm{d})} \hat{w}_m^{(\mathrm{d})} \exp\left(2\pi i \frac{(m'+m''-m)j}{N}\right)$$

with $(m', m'') \in [-N/2+1, N/2]$ and $m = -N+2, \ldots, N$. Terms with $m = m' + m''$ outside the band $[-N/2+1, N/2]$ are sent back inside the

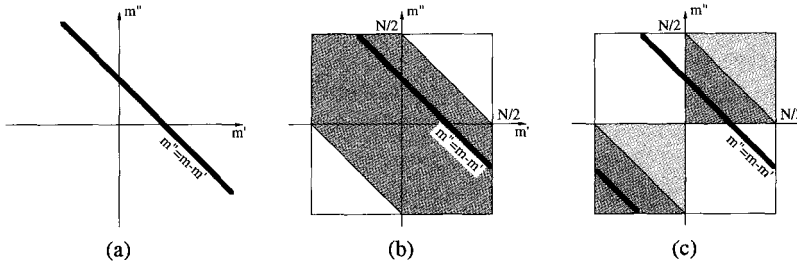

Fig. B.4 (a) The computation of the convolution involves terms such that $m = m' + m''$. (b) For the truncated system only modes with $|m'| < N/2$, $|m''| < N/2$, $|m| < N/2$ should be kept (the gray-shaded region). (c) For the discrete transformation, aliasing send each term in the triangular gray-shaded corners in the corresponding region close to the center, which perturbs the result.

band by the periodicity property (B.31) so that we get

$$\widehat{vw}_m^{(\mathrm{d})} = \sum_{m'+m''=m} \hat{v}_{m'}^{(\mathrm{d})} \hat{w}_{m''}^{(\mathrm{d})} + \sum_{m'+m''=m\pm N} \hat{v}_{m'}^{(\mathrm{d})} \hat{w}_{m''}^{(\mathrm{d})}$$

with $m = -N/2+1, \ldots, 0, \ldots, N/2$.

When the Fourier coefficients $\hat{v}_m^{(\mathrm{d})}$ and $\hat{w}_m^{(\mathrm{d})}$ decrease sufficiently fast as $|m|$ increases, the correction due to the spurious terms may be negligible but it is preferable to get rid of them ("unaliased" scheme). A popular way to achieve this purpose is illustrated in Figure B.5. Starting with the spectrum of v and w with $-N/2+1 \leq m \leq N/2$ one turns to a spectrum with $-M/2+1 \leq m \leq M/2$ and $M > N$ by adding zeroes for $-M/2+1 \leq m \leq -N/2$ and $N/2+1 \leq m \leq M/2$ (in the figure, all products in the band outside the interior square with side N cancel). Going back to the physical space one obtains an evaluation of v and w on a lattice with M regularly spaced points, finer than the original lattice. The product vw is then computed at the nodes of this lattice and next Fourier transformed. It remains to drop the Fourier coefficients outside interval $[-N/2+1, N/2]$ to get the unaliased spectrum of the product. The most dangerous non-zero term is with $m' = m'' = N/2$. Its alias, $m' + m'' - M$, gets off the band if $m' + m'' - M < -N/2 + 1$, thus if $3N/2 - 1 < M$, so that one can take $M = 3N/2$ ("3/2 rule"). Despite manipulations implied by the size increase from N to M, the evaluation of convolutions involved in the treatment of nonlinear terms is freed from a systematic error.

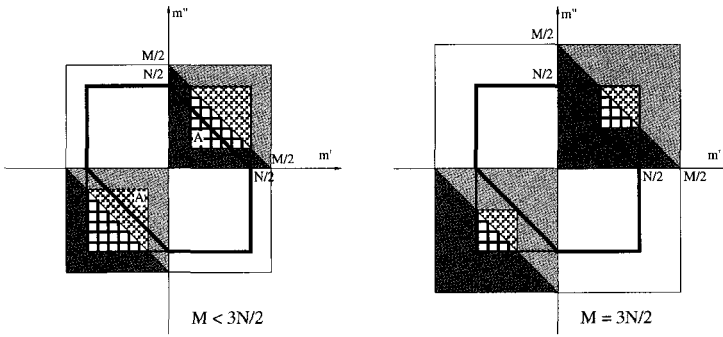

Fig. B.5 As in Figure B.4c, but now with M instead of N, the aliased terms belong to the gray-shaded triangular regions. In the regions with uniform shading, they arise from terms that are zero by construction ($|m|$ outside the $[-N/2+1, N/2]$ interval). The non-trivial aliased terms belong to hatched regions. Some still belong to the domain bounded by the thick line when $M < 3N/2$ (left), which is no longer the case for $M \geq 3N/2$ (right).

B.3 Exercises

B.3.1 *Truncation and round-off errors*

Consider the differential equation $\frac{\mathrm{d}}{\mathrm{d}t}v = -v$ with initial condition $v(0) = 1$ and exact solution $v(t) = \exp(-t)$ and point out the role of the order of a numerical scheme on the precision of the numerical integration by comparing the exact result at $t = 1$, $v = 1/e$, to approximations obtained with first-order iterations (B.3) and (B.5), second-order schemes (B.6), (B.7), (B.8), (B.9, B.10) and (B.11, B.12). In particular, show formally that the RK schemes are of order Δt^2.

Write down all the corresponding programs and determine the distance $|v_N - 1/e|$ between the exact solution and the numerical solution v_N obtained performing N successive iterations of each scheme with a time-interval $\Delta t = 1/N$. Use logarithmic scales to plot this distance as a function of Δt and to identify the regime dominated by truncation errors from that dominated by the accumulation of round-off errors. Test also the fourth-order Runge–Kutta scheme (B.13).

B.3.2 *Stability of multi-step schemes*

Consider the integration of $\frac{\mathrm{d}}{\mathrm{d}t}v = f(v)$ by means of second-order schemes (B.7) and (B.8) and assume that a solution $\{v_n, n = 0, 1, \ldots\}$ is known.

Neighboring solutions $\{\tilde{v}_n, n = 0, 1, \dots\}$ can be written as series of perturbations $\{u_n, n = 0, 1, \dots\}$, with $u_n = \tilde{v}_n - v_n$ governed by the linearized map $g_n = \mathrm{d}f/\mathrm{d}v|_n$. Show that the two numerical schemes respectively lead to:

$$u_{n+1} = u_n + \tfrac{1}{2}\Delta t\,(3g_n u_n - g_{n-1}u_{n-1})\,, \tag{B.32}$$

$$u_{n+1} = u_{n-1} + 2\Delta t\, g_n u_n\,. \tag{B.33}$$

Then write (B.32) and (B.33) as two-dimensional maps for the two variables u_n and $v_n = u_{n-1}$ and compute their eigenvalues. Determine which is the physical eigenvalue λ_{phys} controlling the solution to the linearized problem $\frac{\mathrm{d}}{\mathrm{d}t}u = g(v)u$, where v is the solution to the primitive problem $\frac{\mathrm{d}}{\mathrm{d}t}v = f(v)$, and which is the eigenvalue λ_{num} corresponding to the spurious numerical mode. The instability takes place when $|\lambda_{\mathrm{num}}| > 1$. Assuming that $|g|$ is bounded by some $g_{\max}$ for the values of v of interest, discuss the stability properties of the two schemes.

B.3.3 *Finite difference schemes for the SH model*

Derive second-order consistent finite difference approximations to the linear part of the Swift–Hohenberg (SH) model:

$$\partial_t v = rv - (\partial_{xx} + 1)^2 v \equiv (r-1)v - \partial_{xxxx}v - 2\partial_{xx}v\,. \tag{B.34}$$

Consider explicit–Euler, implicit–Euler and Crank–Nicholson temporal schemes and use (B.18) to discretize the fourth-order derivative. Study their stability by introducing discrete Fourier normal modes analogous to those leading to conditions (B.23) and (B.24).

Write down quasi-linearized Crank–Nicholson schemes for the original SH model, *i.e.* (B.34) with nonlinear term $\mathcal{N}(v) = -v^3$ added to its r.h.s. by expanding the nonlinear terms $\mathcal{N}(v)$ as:

$$\mathcal{N}(v)|_{n+1} = \mathcal{N}(v)|_n + \delta\mathcal{N}/\delta v|_n (v_{n+1} - v_n)\,.$$

Consider also the modified SH model completed by a term $\mathcal{N}(v) = -v\partial_x v$ on the r.h.s. (*i.e.* an advection term $v\partial_x v$ on its l.h.s).

B.4 Case studies

The working sessions are organized around two themes: ODEs and PDEs. To deal with ODEs, we choose the simplest fourth-order Runge–Kutta scheme (B.13) and apply it to dynamical systems with few degrees of freedom, the periodically forced pendulum (two-dimensional, non-autonomous, §B.4.1) and the Lorenz model (three-dimensional, autonomous, §B.4.2). Other examples are illustrated but not studied in detail, the Rössler system, §B.4.3.1, and the Chua circuit, §B.4.3.2. All this can serve to materialize concepts introduced in Chapters 2 and 4.

Problems arising in EDPs are illustrated using two variants of the SH model introduced in some exercises of Chapters 3 and 4, and considered in exercise B.3.3 above. They may help to better understand pattern formation studied in Chapter 5. The approach can be extended to illustrate the generation of dissipative waves. Boundary conditions of Neumann–Dirichlet type are better adapted to finite difference methods and will mainly serve us to illustrate numerical stability problems. Periodic boundary conditions are ideal for a first approach of spectral methods used to illustrate space-time chaotic regimes. Computational Fluid Dynamics (CFD) will be left to specialists.

B.4.1 *ODEs 1: Forced pendulum*

We consider a pendulum submitted to a sinusoidal external force:

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\theta + \eta\frac{\mathrm{d}}{\mathrm{d}t}\theta + \sin(\theta) = f\cos(\omega t) \tag{B.35}$$

where θ is the angle that the pendulum makes with the vertical, η is the damping coefficient, f is the intensity of the forcing and ω its angular frequency.

Write down a second-order Runge–Kutta program (B.11, B.12) for (B.35) written as a system of two first-order ODEs for θ and $\phi = \dot{\theta}$, taking care of the explicit time dependence introduced by the forcing. Then turn to a fourth-order scheme (B.13).

1) Use that routine to draw the projection of the trajectories in the plane (θ, ϕ). Choose for example $\omega = \frac{2}{3}$, $\eta = 0.5$ and f variable in the range $[0.5, 3]$. Consider in particular cases $f = 1.07$ and 1.15. See Figure B.6.

2) Perform a stroboscopic analysis (Poincaré section) of the system at the forcing period $T = 2\pi/\omega$ and observe the attractor in particular for the previous parameter values mentioned. See Figure B.7, left.

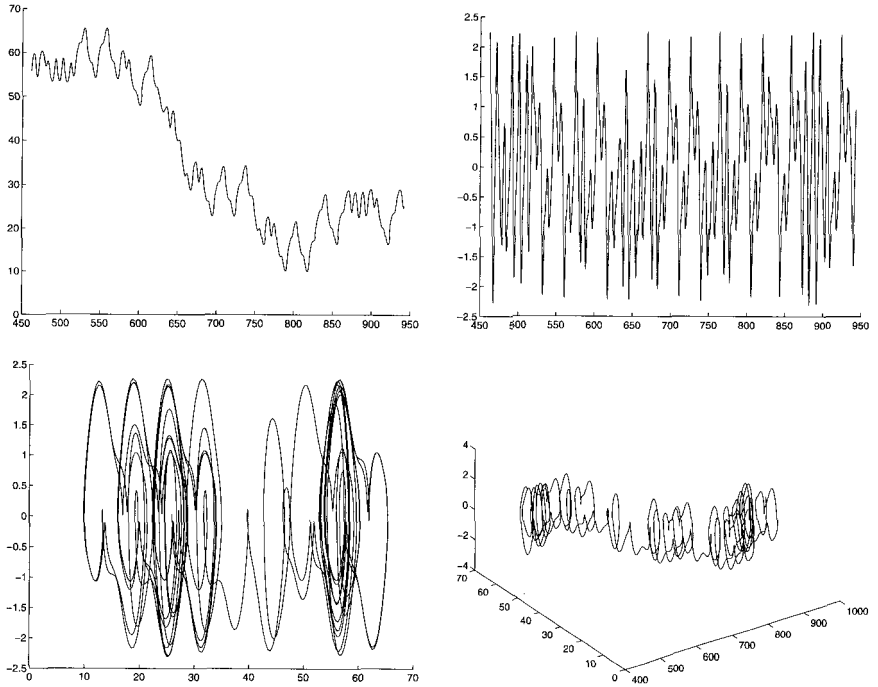

Fig. B.6 Forced pendulum for $\omega = 2/3$, $f = 1.15$. Top-left: Time series of θ. Top-right: Time series of $\dot{\theta} \equiv \phi$. Bottom-left: ϕ as a function of θ. Bottom-right: Three-dimensional representation in the space (t, θ, ϕ), ϕ-axis vertical, θ-axis to the left, t-axis to the right.

3) When the intensity of the forcing f increases, one finds parameter ranges where the attractor is periodic and other ranges where it is chaotic. Draw the bifurcation diagram obtained by varying f by recording the values reached by variables ϕ on the surface of section (for every $\Delta t = T$), at given $\omega \in [\omega_1, \omega_2]$, for a sufficiently long trajectory and after having eliminated points corresponding to the transient. See Figure B.7, right.

For all these simulations, it will be interesting to choose the time step Δt such that $\Delta t = T/N$ so that a full oscillation of the forcing corresponds exactly to an integer number of time steps. N will be sufficiently large, so that one has $\Delta t \ll T$ and $\Delta t \ll T_0 = 2\pi$, the natural period of the pendulum in the vicinity of its equilibrium position. A detailed study would show that the value of the bifurcation points is sensitive to the numerical resolution (compare results with the 2nd-order and 4th-order schemes).

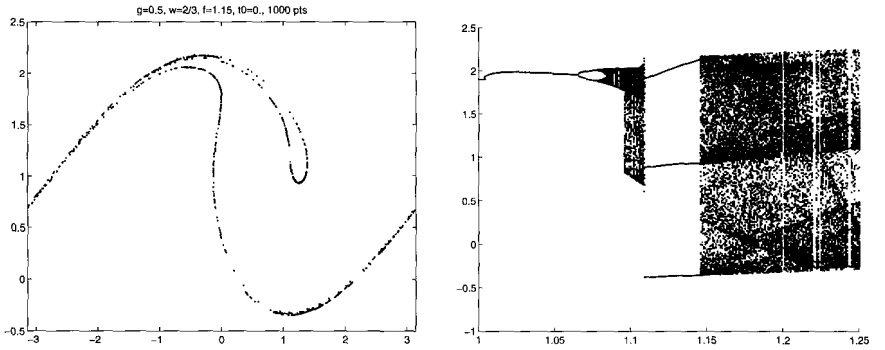

Fig. B.7 Left: Poincaré section of the forced pendulum for $f = 1.15$. Right: Bifurcation diagram for $f \in [1, 1.25]$ obtained when increasing f (the result would be slightly different when decreasing f because some bifurcations are sub-critical, different attractors may coexist in some limited ranges of the parameter f, and hysteresis takes place correspondingly).

B.4.2 *ODEs 2: Lorenz model*

In Chapter 3, §3.2.1 we introduced a simplified model of nonlinear convection initially proposed by E.N. Lorenz, cf. Note 5, p. 88 rewritten here for convenience using standard notations:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \sigma(Y - X)\,, \tag{B.36}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = -XZ + rX - Y\,, \tag{B.37}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Z = XY - bZ\,. \tag{B.38}$$

1) Display the solution in the three-dimensional space (X, Y, Z), draw the time series of Z and study the attractor obtained for the values initially chosen by Lorenz: $\sigma = 10$, $b = 8/3$ and $r = 28$ (Figure B.8).

Then vary r on the interval $r \in [145, 170]$, a range where the chaotic attractor decays into a periodic cycle through a subharmonic cascade and returns to chaos by intermittency (Figures B.9, and B.10).

2) Determine the Lorenz map defined as $Z_{k+1} = f(Z_k)$ where Z_k is defined by the condition that $Z(t)$ goes through a maximum, *i.e.* $\frac{\mathrm{d}}{\mathrm{d}t}Z = 0$ with $\frac{\mathrm{d}^2}{\mathrm{d}t^2}Z < 0$. To do so, at each iteration check the value of $D_n = X_nY_n - bZ_n$, identify intervals $[t_n, t_{n+1} = t_n + \Delta t]$ where D_n changes sign from positive to negative. By linear interpolation, find the time τ_n where the change of sign takes place. Restarting from the result at t_n, perform

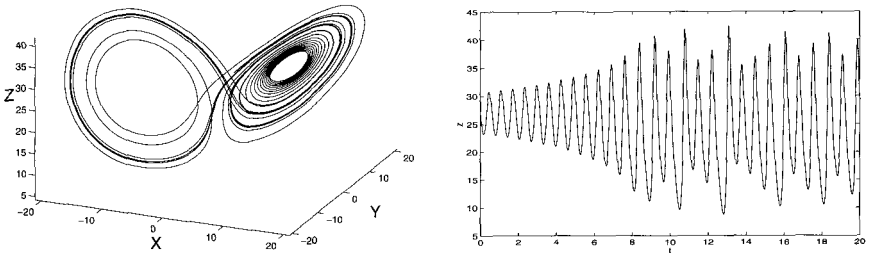

Fig. B.8 Left: The Lorenz attractor for $r = 28$ with $\Delta t = 0.01$. Initial conditions $X_0 = 6 = Y_0$, $Z_0 = 27$. Right: Time series $Z(t)$ in the same conditions.

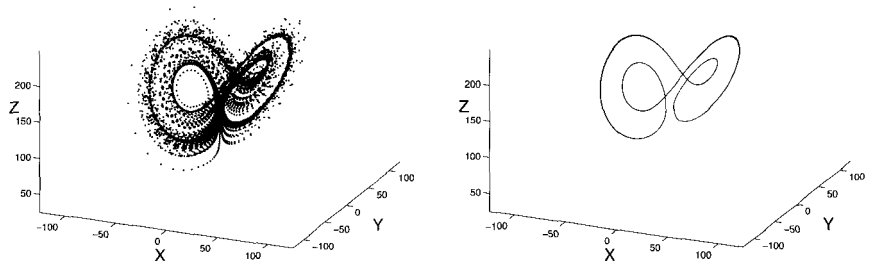

Fig. B.9 Left: "Intermittent" Lorenz attractor for $r = 166.5$ with $\Delta t = 0.01$. Initial conditions: $X_0 = 12 = Y_0$, $Z_0 = 165.5$. Lines joining successive points are not drawn to better point out the existence of a "ghost" trajectory recurrently visited by the system. Right: Limit cycle for $r = 166$ with $\Delta t = 0.01$. Initial conditions: $X_0 = 12 = Y_0$, $Z_0 = 165$. The cycle is reached after a transient that is not represented here.

a single Runge–Kutta step with time interval $\tau_n - t_n$ to obtain the set of coordinates of the corresponding point. Continue the simulation from t_{n+1} determined previously. The Lorenz map for $r = 28$ is displayed in Figure B.11. This method to obtain Poincaré maps with essentially the same accuracy as the integration itself was proposed by Hénon.[26] The same strategy can thus be used for Poincaré sections other than that defined by the condition $D = 0$. A popular one in early studies of the Lorenz model was the surface defined by $Z = r - 1$ and crossed from above (for example).

3) Construct a bifurcation diagram for the Lorenz model by plotting $Z_k - r + 1$ as a function of r (Figure B.12). Sets of isolated points at given

[26] M. Hénon, "On the numerical computation of Poincaré maps," Physica D **5** (1982) 412–414.

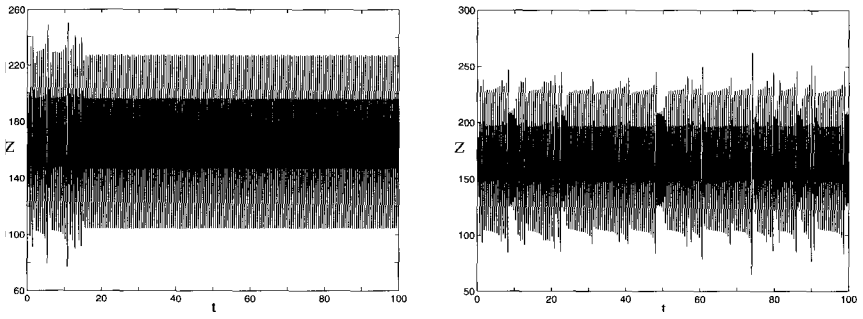

Fig. B.10 **Left:** $Z(t)$ **for** $r = 166$, periodic regime after a brief transient. Right: $Z(t)$ for $r = 166.5$, intermittent regime.

r correspond to periodic behavior, continuous rows of inhomogeneously distributed points to chaotic behavior. The bifurcation around $r = 320$ is not a period doubling but a symmetry breaking bifurcation as shown by drawing trajectories in three dimensions before and after the bifurcation.

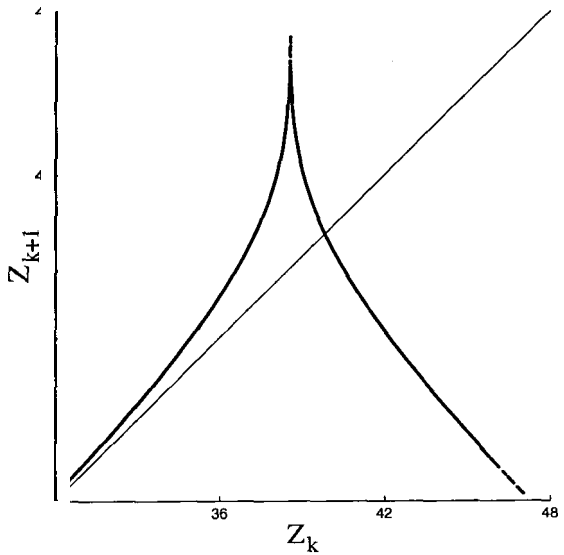

Fig. B.11 Lorenz map $Z_{k+1} = f(Z_k)$ for $r = 28$.

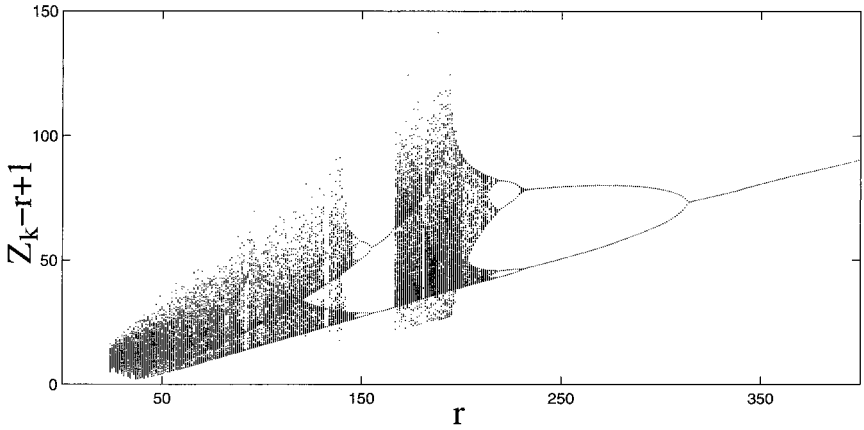

Fig. B.12 Bifurcation diagram of the Lorenz model.

B.4.3 *ODEs 3: Rössler and Chua models*

B.4.3.1 *Rössler model*

Another simple three-dimensional autonomous dynamical system with a strange attractor is the Rössler model.[27]

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}X &= -Y - Z\,,\\
\tfrac{\mathrm{d}}{\mathrm{d}t}Y &= X + aY\,,\\
\tfrac{\mathrm{d}}{\mathrm{d}t}Z &= b + (X - c)Z\,.
\end{aligned}$$

Notice that its definition is even simpler than that of the Lorenz model since it has a single nonlinear quadratic term and no obvious symmetry. The dynamics has two ingredients: approximately periodic behavior at some distance from an unstable spiral fixed point and fast excursions away from the (X, Y)-plane. It displays periodic and chaotic regimes. Typical trajectories are shown in Figure B.13. Use the same tools as for the Lorenz model to study its bifurcation diagram as a function of $c \in [1, 20]$.

[27] O.E. Rössler, "An equation for continuous chaos," Physics Lett. A **57** (1976) 397–398. Also, "Continuous chaos - Four prototype equations," Ann. NY Acad. Sc. **316** (1979) 376–392. See also: J.C. Sprott, "Simplest dissipative chaotic flows," Physics Lett. A **228** 271–274.

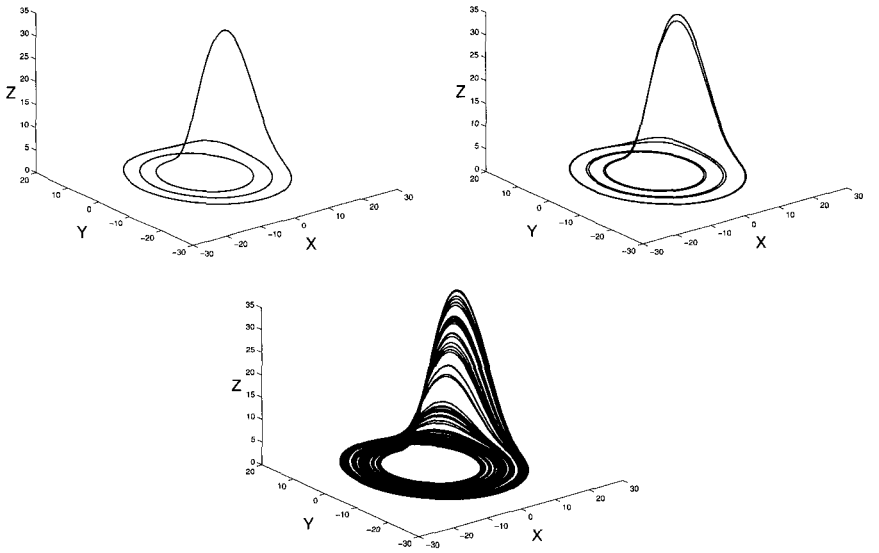

Fig. B.13 The Rössler attractor for $a = 0.1$, $b = 0.1$. Top-left: Limit cycle for $c = 12$. Top-right: Limit cycle after a period doubling for $c = 12.7$. Bottom: Chaotic attractor beyond the accumulation point of the subharmonic cascade for $c = 13.6$.

B.4.3.2 *Chua circuit*

The Chua circuit is the last example proposed here. It is defined as a prototype of analogical systems displaying chaos that can be implemented as electronic circuits with good accuracy. It is also three-dimensional and autonomous. Its main properties are reviewed in a collective work edited by R.N. Madan.[28] It reads

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}X &= \alpha(Y - X - F(X))\,, \\
\tfrac{\mathrm{d}}{\mathrm{d}t}Y &= X - Y + Z\,, \\
\tfrac{\mathrm{d}}{\mathrm{d}t}Z &= -\beta Y\,.
\end{aligned}$$

where $F(X)$ is an odd function of X that is linear by part: $F(X) = m_1 X + \frac{1}{2}(m_0 - m_1)(|X + 1| - |X - 1|)$. The "double scroll" attractor is displayed in Figure B.14. Establish the bifurcation diagram of the Chua system from a Lorenz map of variable Z, with $\beta \in [1, 25]$ and other parameters fixed as in the caption of Figure B.14. Display other typical attractors.

[28] R.N. Madan, Ed., *Chua's circuit: a paradigm for chaos* (World Scientific, 1993).

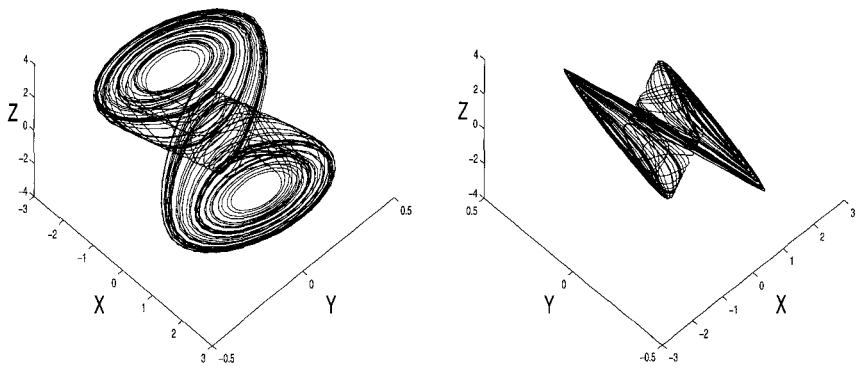

Fig. B.14 Two views of the Chua "double scroll" attractor for $\alpha = 9$, $\beta = 100/7$, $m_0 = -8/7$, $m_1 = -5/7$.

B.4.4 *PDEs 1: SH model, finite differences*

Finite difference methods are illustrated here using the original variant of the Swift–Hohenberg model with a cubic nonlinear term in one space dimension:

$$\partial_t v = rv - (\partial_{xx} + 1)^2 v - v^3 \tag{B.39}$$

with boundary conditions

$$v(0) = v(\ell) = \partial_x v(0) = \partial_x v(\ell) = 0 \tag{B.40}$$

The two parameters are r and ℓ. The model derives from a potential and is a good tool to study the formation of cellular structures in simple cases.

B.4.4.1 *Explicit scheme*

Develop the simplest possible, first order in time, second order in space, explicit numerical scheme. To maintain second order consistency at the boundaries, add fictitious exterior points enslaved to the first inner points as explained in §B.2.1.2. Observe how the numerical instability develops when the condition on Δt obtained in Exercise B.3.3, namely $\Delta t < \Delta x^4/8$ for Δx small, is violated.

The effect of the space resolution is shown in Figure B.15. In order to reach the finest grid, the time step should be considerably reduced in order to avoid the numerical instability illustrated in Figure B.16. Notice that in certain cases depending on initial conditions, the nonlinear term may be

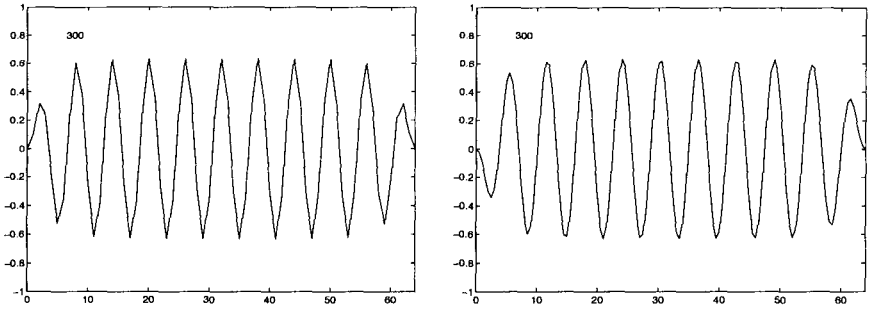

Fig. B.15 Explicit-Euler numerical simulation of the Swift–Hohenberg model with $r = 0.3$; random initial conditions with amplitude 0.01; solution at $t_f = 300$. Left: $\ell = 64$, $n = 64$, $\Delta x = 1.0$, $\Delta t = 0.1$. Right: $\ell = 64$, $n = 128$, $\Delta x = 0.5$, $\Delta t = 0.008$, numerically stable.

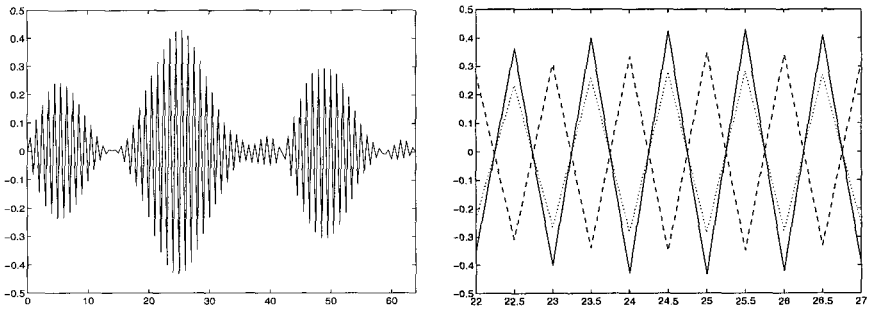

Fig. B.16 Numerical instability of the Swift–Hohenberg model with $\ell = 64$, $n = 128$, $\Delta x = 0.5$, $\Delta t = 0.01$, $r = 0.3$, random initial conditions with amplitude 0.01. Left: State at iteration #20. Right: The same numerical instability at three successive time steps, iterations #18 (dots), #19 (dashes), and #20 (solid line).

sufficiently strong to prevent the numerical divergence but the so-obtained solution is not physical in that it has a space period very different from the expected one $\simeq \lambda_c = 2\pi$ (in practice much shorter).

B.4.4.2 *Implicit scheme*

Develop an implicit Euler scheme, first order in time, second order in space, by solving the linear problem as explained in §B.2.1.4. Take $\ell = 64$ ($\simeq 10\lambda_c$) in order to deal with a sufficiently extended system. The number of grid points N will be varied to check the effect of the numerical resolution

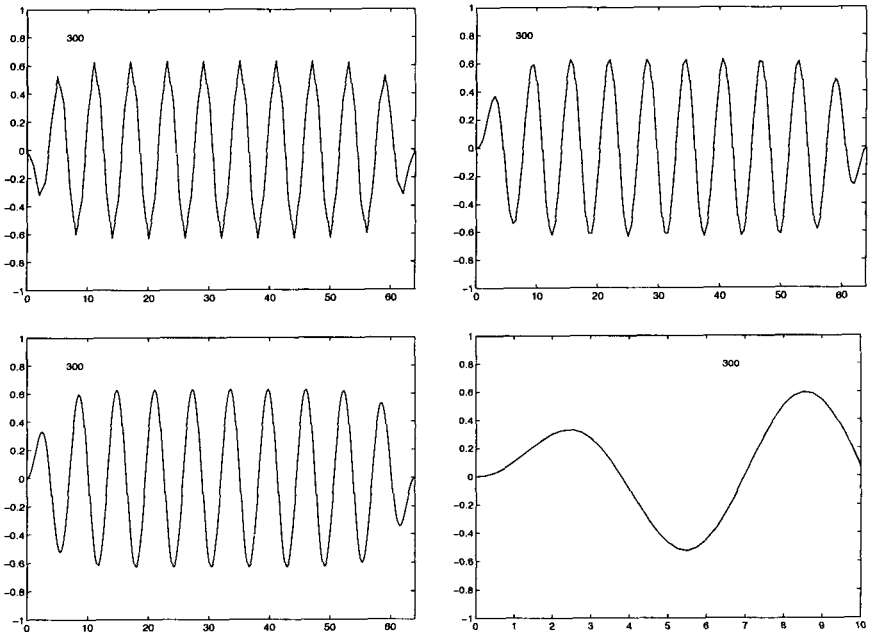

Fig. B.17 Swift–Hohenberg model with $\ell = 64$, $\Delta t = 0.1$, $r = 0.3$, random initial conditions with amplitude 0.01; solution at $t_f = 300$. (a) $n = 64$, $\Delta x = 1.0$. (b) $n = 128$, $\Delta x = 0.5$. (c) $n = 256$, $\Delta x = 0.25$. (d) zoom on the left boundary showing the accuracy of the numerical account of boundary condition $\partial_x w = 0$.

(Figure B.17). Second order schemes can also be developed, either of the quasi-linearized Crank–Nicholson type (Exercise B.3.3) or according to an Adams–Bashford scheme.

Since numerical stability problems are dismissed, focus on the physics of the formation of the cellular structure. Identify the initial phase of exponential growth of perturbations selected by the linear dynamics, then the saturation phase where nonlinearities come in the foreground, and the final stage where most effects compensate so that a slow residual motion (phase diffusion) towards the final time-independent state is observed.

B.4.5 *PDEs 2: SH model, pseudo-spectral method*

Consider now the modified Swift–Hohenberg model:

$$\partial_t v + v\partial_x v = \left[r - (\partial_{xx} + 1)^2\right] v\,, \tag{B.41}$$

with periodic boundary conditions at the ends of an interval of length ℓ. This model, which does not derive from a potential and thus can have unsteady behavior, is particularly appropriate to study the growth of space-time chaos in cellular structures.

Derive a Fourier pseudo-spectral algorithm implementing the model taking advantage of the fact that $v\,\partial_x v = \frac{1}{2}\,\partial_x\left(v^2\right)$, according to the scheme:

$$\hat{v}_m \begin{cases} \mapsto \widehat{\mathcal{L}v}_m = s_m \hat{v}_m = \left(r - (k_m^2 - 1)^2\right) \hat{v}_m \\ \mapsto v(x) \mapsto \frac{1}{2} v(x)^2 \mapsto \widehat{\frac{1}{2}v^2}_m \mapsto \widehat{v\partial_x v}_m \equiv \frac{1}{2} i k_m \widehat{v^2}_m = \widehat{\mathcal{N}}_m \end{cases}$$

where $\widehat{\mathcal{N}}_m$ must be computed as discussed earlier in order to avoid aliasing errors. Next choose a time evolution scheme (§B.1). Here the linear part can be integrated exactly since it is diagonal in Fourier space. For the nonlinear part a second order Adams–Bashford scheme can be chosen: The final result is

$$\hat{v}_{n+1,m} = \exp(s_m \Delta t)\hat{v}_{n,m} + \Delta t \left[\tfrac{3}{2}\widehat{\mathcal{N}}_{n,m} - \tfrac{1}{2}\widehat{\mathcal{N}}_{n-1,m}\right] .$$

When $r = 1$ the Swift–Hohenberg model happens to be a variant of the Kuramoto–Sivashinsky equation. In a slightly different form:

$$\partial_t v + \tfrac{1}{2}(\partial_x v)^2 + \partial_{xx} v + \partial_{xxxx} v = 0\,, \tag{B.42}$$

this equation appears in the treatment of turbulent interface propagation, *e.g.* flame fronts, as shown in Figure B.18.

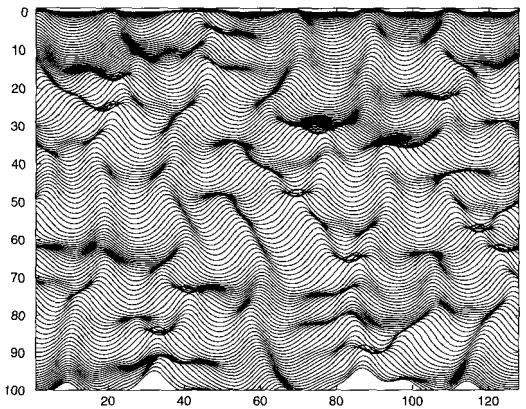

Fig. B.18 Kuramoto–Sivashinsky equation for $\ell = 64$, $n = 128$, $\Delta x = 0.5$, with a time interval $\Delta T = 1$ between two successive snapshots.

This page intentionally left blank

Bibliography

Dynamical systems, bifurcations and catastrophes, chaos

Abarbanel, H.D.I. (1996). *Analysis of Observed Chaotic Data*, (Springer).

Abraham, R.H. and Shaw, C.D. (1992). *Dynamics: the Geometry of Behavior* (Addison-Wesley).

Alligood, K.T., Sauer, T.D., and Yorke, J.A. (1996). *Chaos, an Introduction to Dynamical Systems* (Springer).

Baker, G.L. and Gollub, J.P. (1996). *Chaotic Dynamics, an Introduction*, 2nd Edition (Cambridge Univ. Press).

Beck, Ch. and Schlögl, F. (1995) *Thermodynamics of Chaotic Systems* (Cambridge Univ. Press).

Bergé, P., Pomeau, Y., and Vidal, Ch. (1986). *Order within Chaos, towards a Determinsitic Approach to Turbulence* (Wiley).

Cladis, P.E. and Palffy-Muhoray, P., Editors (1995) *Spatio-Temporal Patterns in Nonequilibrium Complex Systems* (Addison-Wesley).

Cross, M.C. and Hohenberg, P.C. (1993). Pattern formation outside equilibrium, *Rev.Mod.Phys.* **65** (1993) 851–1112.

Cvitanović, P., Editor (1989). *Universality in Chaos* (Adam Hilger) [reprint collection].

Devaney, R.L. (1989) *An Introduction to Chaotic Dynamical Systems*, 2nd edition (Addison-Wesley).

Drazin, P.G. (1992). *Nonlinear Systems* (Cambridge Univ. Press).

Eckmann, J.P. and Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57** (1985) 617–656.

Frøyland, J. (1992). *Introduction to Chaos and Coherence*, (IoP Publishing).

Glansdorff, P. and Prigogine, I. (1971). *Thermodynamics of Structures, Stability and Fluctuations* (Wiley).

Guckenheimer, J. and Holmes, Ph. (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer).

Haken, H. (1983). *Synergetics*, 3rd edition (Springer).

Hao, B.-l., Ed. (1987-...). *Directions in Chaos* (World Scientific) [several volumes].

Hao B.-l., Ed. (1990). *Chaos II* (World Scientific) [reprint collection].

Hirsch, M.W. and Smale, S. (1974). *Differential Equations, Dynamical Systems and Linear Algebra* (Academic Press).
Kaneko, K., Editor (1993). *Theory and Application of Coupled Map Lattices* (Wiley).
Kanz, H. and Schreiber, Th. (1997). *Nonlinear Time Series Analysis*, Cambridge Nonlinear Science Series 7 (Cambridge Univ. Press).
Lefschetz, S. (1977). *Differential Equations: Geometric Theory* (Dover).
Mandelbrot, B.B. (1982). *The Fractal Geometry of Nature* (Freeman).
Moon, F.C. (1992). *Chaotic and Fractal Dynamics, an Introduction for Applied Scientists and Engineers* (Wiley).
Nayfeh, A.H. and Mook, D.T. (1979). *Nonlinear Oscillations* (Wiley).
Ott, E. (1993). *Chaos in Dynamical Systems* (Cambridge Univ. Press).
Poston, T. and Stewart, I. (1978). *Catastrophe Theory and its Applications* (Pitman).
Rabinovich, M.I., Ezersky, A.B., and Weidman, P.D. (2000). *The Dynamics of Patterns* (World Scientific).
Schuster, H.G. (1988). *Deterministic Chaos, an Introduction* 2nd edition (VCH).
Schuster, H.G., ed. (1999). *Handbook of Chaos Control* (Wiley).
Steeb, W.-H. (1991). *A Handbook of Terms used in Chaos and Quantum Chaos* (BI, Mannheim).
Strogatz, S.H. (1994). *Nonlinear Dynamics and Chaos, with Applications to Physics, Chemistry and Engineering* (Addison-Wesley).
Weigend, A.S. and Gershenfeld, N.A., Editors (1993). *Time Series Prediction: Forecasting the Future and Understanding the Past* (Addison-Wesley).
Wolfram, S., Editor (1986). *Theory and Application of Cellular Automata* (World Scientific).

Hydrodynamics, instabilities, and turbulence

Batchelor, G.K. (1967). *An Introduction to Fluid Dynamics* (Cambridge Univ. Press).
Batchelor, G.K., Moffatt, H.K., and Worster, M.G., Editors (2000). *Perspectives in Fluid Dynamics. A Collective Introduction to Current Research* (Cambridge Univ. Press).
Betchov, R. and Criminale, W.O. (1967). *Stability of Parallel Flows* (Academic Press).
Chandrasekhar. S. (1961). *Hydrodynamic and Hydromagnetic Stability* (Clarendon Press).
Drazin, P.G. (2002). *Introduction to Hydrodynamic stability* (Cambridge University Press).
Drazin, P.G. and Reid, W.H. (1981). *Hydrodynamic Stability* (Cambridge Univ. Press).
Frisch, U. (2001). *Turbulence. The legacy of A.N. Kolmogorov*, 2nd edition (Cambridge Univ. Press).
Godrèche, Cl. and Manneville, P., Editors (1998). *Hydrodynamics and Nonlinear*

Instabilities (Cambridge Univ. Press).
Landau, L.D. and Lifshitz, E.M. (1987). *Fluid Mechanics - Course of Theoretical Physics, Vol. 6*, 2nd edition (Butterworth–Heinemann).
Lesieur, M. (1997). *Turbulence in Fluids* (Kluwer).
Lesieur, M., Yaglom, A., and David, F., Editors (2001) *New Trends in Turbulence.* Les Houches, session LXXIV (Springer).
Mollo-Christensen, E.L. (1972). *Flow Instabilities*, film produced by [NCFMF (1972)].
NCFMF (1972). Series of films produced by the National Commitee for Fluid Mechanics, distributed by Encyclopedia Britannica Educational Corp.; accompanying book: *Illustrated Experiments in Fluid Mechanics* (MIT Press).
Pope, S.B. (2000). *Turbulent flows* (Cambridge Univ. Press).
Rothman, D.H. and Zaleski, S. (1997). *Lattice-gas Cellular Automata. Simple Models of Complex Hydrodynamics* (Cambridge Univ. Press, 1997).
Schlichting, H. (1979). *Boundary Layer Theory* (McGraw-Hill).
Schmid, P.J. and Henningson, D.S. (2001). *Stability and Transition in Shear Flows* (Springer-Verlag).
Stewart, R.W. (1972). *Turbulence*, film produced by [NCFMF (1972)].
Swinney, H.L. and Gollub, J.P., Editors (1985). *Hydrodynamic Instabilities and the Transition to Turbulence* (Springer).
Tabeling, P. and Cardoso, O., Editors (1994) *Turbulence. A Tentative Dictionary* (Plenum Press).
Tennekes, H. and Lumley, J.L. (1972). *A First Course in Turbulence* (MIT Press).
Tritton, D.J. (1990) *Physical Fluid Dynamics* (Cambridge Univ. Press).
van Dyke, M. (1982). *An Album of Fluid Motion* (Parabolic Press).
Wilcox, D.C. (2000). *Turbulence Modeling for CFD* (DWC Industries).

Numerical modeling

Abramowitz, M. and Stegun, I.A., Editors (1972). *Handbook of Mathematical Functions* (National Bureau of Standards) [Chapter 22: Numerical interpolation, differentiation, and integration].
Acton, F.S. (1970). *Numerical Methods that Work* (Harper and Row).
Canuto, C., Hussaini, M.Y., Quarteroni, A., and Zang, T.A. (1988). *Spectral Methods in Fluid Dynamics* (Springer).
Finlayson, B.A. (1972). *The Method of Weighted Residuals and Variational Principles, with Application in Fluid Mechanics, Heat and Mass Transfer* (Academic Press).
Gottlieb, D. and Orszag, S.A. (1977) *Numerical Analysis of Spectral Methods: Theory and Applications* (SIAM Publications).
Lapidus, L. and Seinfeld, J.H. (1971). *Numerical Solution of Ordinary Differential Equations* (Academic Press).
Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1986). *Numerical Recipes, the Art of Scientific Computing* (Cambridge Univ. Press) [program sources available].

Richtmyer, R.D. and Morton, K.W. (1967) *Difference Methods for Initial Value Problems* (Interscience).

Miscellaneous

Anderson, P.W., Arrow, K.J., and Pines, D., Editors (1988). *The Economy as an Evolving Complex System*, SFI Studies in the Sciences of Complexity (Addison-Wesley).

Hall, N., ed. (1992). *The NewScientist Guide to Chaos* (Penguin Books).

Henderson-Sellers, A. and McGuffie, K. (1987). *A Climate Modelling Primer* (Wiley).

Peixoto, J.P. and Oort, A.H. (1992). *Physics of Climate* (AIP Publishing).

Philander, S.G. (2000) *Is The Temperature Rising? The Uncertain Science of Global Warming* (Princeton Univ. Press).

Stanley, H.E. (1988). *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press). [reprint original edition, Pergamon Press, 1971]

Winfree, A.T. (1990). *The Geometry of Biological Time* (Springer).

Index