Mathematical Engineering

Gordon Blower

Linear Systems

Springer

Mathematical Engineering

Series Editors

Jörg Schröder, Institute of Mechanics, University of Duisburg-Essen, Essen, Germany

Bernhard Weigand, Institute of Aerospace Thermodynamics, University of Stuttgart, Stuttgart, Germany

Jan-Philip Schmidt, Universität of Heidelberg, Heidelberg, Germany

Advisory Editors

Günter Brenn, Institut für Strömungslehre und Wärmeübertragung, TU Graz, Graz, Austria

David Katoshevski, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Jean Levine, CAS- Mathematiques et Systemes, MINES-ParsTech, Fontainebleau, France

Gabriel Wittum, Goethe-University Frankfurt am Main, Frankfurt am Main, Germany

Bassam Younis, Civil and Environmental Engineering, University of California, Davis, Davis, CA, USA

Today, the development of high-tech systems is unthinkable without mathematical modeling and analysis of system behavior. As such, many fields in the modern engineering sciences (e.g. control engineering, communications engineering, mechanical engineering, and robotics) call for sophisticated mathematical methods in order to solve the tasks at hand.

The series Mathematical Engineering presents new or heretofore little-known methods to support engineers in finding suitable answers to their questions, presenting those methods in such manner as to make them ideally comprehensible and applicable in practice.

Therefore, the primary focus is—without neglecting mathematical accuracy—on comprehensibility and real-world applicability.

To submit a proposal or request further information, please use the PDF Proposal Form or contact directly: Dr. Thomas Ditzinger (thomas.ditzinger@springer.com)

Indexed by SCOPUS, zbMATH, SCImago.

Gordon Blower

Linear Systems

Springer

Gordon Blower
Department of Mathematics and Statistics
Lancaster University
Lancaster, UK

ISSN 2192-4732 ISSN 2192-4740 (electronic)
Mathematical Engineering
ISBN 978-3-031-21239-0 ISBN 978-3-031-21240-6 (eBook)
https://doi.org/10.1007/978-3-031-21240-6

Mathematics Subject Classification: 93C05, 93B52, 15A16, 34D08

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Jian

Preface

The book is about the mathematics of linear systems, particularly continuous time autonomous linear systems with finite-dimensional state space. These cover an interesting range of applications in science and engineering and are basic to the study of more complex systems. The material is at a level suitable for a third or fourth year undergraduate student in a UK university. We assume that the reader is familiar with calculus, linear algebra and basic complex analysis and develop these ideas further within the particular context of linear systems.

Chapter 1 is about how to describe linear systems by differential equation, block diagrams and linear algebra. One of the attractions of linear systems is that they can be discussed in several different ways. Chapter 2 begins the systematic development of the theory, where methods of linear algebra are used to describe linear systems and compute transfer functions. The purpose of transfer functions is deferred until Chap. 4, where they are introduced in parallel via the Laplace transform. In Chap. 2, we use only basic linear algebra, while in chapter three more advanced techniques are introduced. In most cases, the results are proved in detail, and there are indications about how the mathematical questions can be posed in a form suitable for calculation via MATLAB. A crucial aspect of linear systems theory is its adaptability to treat systems of very high dimension, for which computers are essential. Some readers may wish to defer the final few sections of Chap. 3 until later. More generally, some sections of the book are more challenging than others, and readers can pass over some results if they find them difficult.

In Chap. 4, we give a conventional discussion of the Laplace transform for functions on $(0, \infty)$ with basic applications to differential equations. Inversion of Laplace transforms is a complicated topic, and in this book we give some special cases such as Heaviside's expansion theorem before addressing the general case in Chap. 6; this is consistent with the historical development of the subject. Methods of complex analysis are mainly deferred until Chap. 5. We also cover the Fourier cosine transform and obtain an inversion formula for the L^1 cosine transform. The Fourier transform is essential for applications to signal processing, which are pursued in Chap. 10.

In Chaps. 5, 6, and 7, we consider three approaches to the stability problem. Chapter 5 uses methods of complex analysis and geometric function theory to help us visualize the transfer function in terms of its Nyquist contour. In Chap. 6, we present algebraic approaches which are algorithmic and can be carried out in exact arithmetic without approximation. Then in Chap. 7 we use linear algebra in a manner that is especially suited to large matrices. All of these approaches are most effective when they are implemented with the aid of computers, and for large systems, computers are essential. Although these approaches are separated into distinct chapters, the difference between them should not be overstated. They are different routes towards the same goal and the same problem can be expressed in different but equivalent ways in terms of Laplace transforms, polynomials or matrices. Also, some families of transcendental functions such as the Bessel functions can be conveniently described in terms of algebraic differential rings.

In Chap. 8, we consider orthogonal polynomials. This topic is often taught alongside numerical analysis, as examples in the theory of differential equations or as an application of Hilbert space theory. In this book, we emphasize that sequences of orthogonal polynomials can be generated efficiently using discrete time linear systems, which make use of the three term recurrence relation. In applications to signal processing, it is common to use examples such as the Chebyshev polynomials to create filters, and these are particularly well suited to our approach. The chapter covers some of the other classical orthogonal polynomials, such as the Laguerre system, which we later use to prove fundamental results about Fourier integrals.

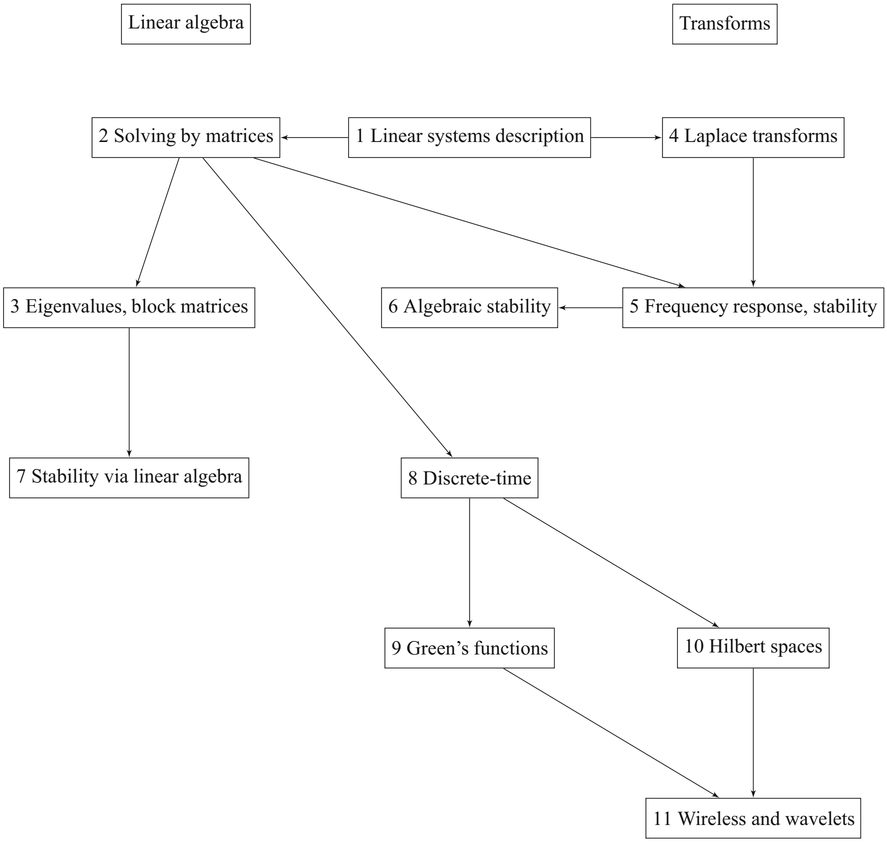
Chapter 9 is concerned with Green's functions, in the sense of Cauchy transforms of an integrable weight on a bounded real interval. This has obvious applications to orthogonal polynomials and moments, and some less obvious application to random linear processes. Some of these results describe fundamental examples in random matrix theory such as the semicircle law. There are diverse applications, such as the famous May-Wigner law in mathematical biology, which demonstrate how useful linear systems are in treating complex problems in many branches of science.

The results of the first nine chapters mainly apply to linear systems with finite dimensional state spaces and the finite matrices that operate on them. In Chap. 10, we introduce Hilbert space with a view to describing some infinite-dimensional linear systems. As soon as one attained the generality of Hilbert space, one realizes the need to have suitably adapted tools with which to carry out explicit computations. For this reason, we introduce Hardy space on the right half-plane with its orthonormal basis derived from the Laguerre functions. The intended application of these is to signal processing, and in Chap. 11 we include famous results such as Shannon's sampling theorem for band limited functions. Although the historical development had an interlude of several decades, it is now natural to follow this directly with the Shannon wavelet basis for L^2. The book concludes with a discussion of Telatar's model of wireless transmission, which is often called 'single user MIMO'. This is an important instance where random matrix theory enters into linear systems and suggests areas for further study.

Chapters 1 and 2 are essential for understanding the rest of the book. Then readers who are mainly interesting in linear algebra can progress to Chaps. 3 and 7. Chapters 4 and 5 feature Laplace transforms and do not depend on the more advanced tools from linear algebra. Chapters 8 and 9 are best read together, and these feed naturally into Chap. 10. The final Chap. 11 requires Chaps. 9 and 10. The following diagram indicates this logical dependence.

There are many aspects of control theory that we do not discuss in this book. Reliability, cost of manufacture, usability and tuning of components are all important topics that we leave to books that emphasize engineering.

This book is based on lectures for a third-year module at Lancaster University for mathematics students and several projects for students in mathematics, physics and environmental science. I am grateful to these students for helpful comments on the course materials which progressively improved the module. The module also benefited from a helpful review by my colleagues Nadia Mazza and David Towers, who persuaded me to incorporate projects into the module assessment and enabled students to work on more extended exercises. Lucinda Hadley's PhD thesis on wireless communication suggested some of the contemporary topics. I am grateful to Remi Lodh of Springer who helped guide the project. A former student Yufei Li proofread the manuscript and eradicated several errors.



After studying this book, we hope that the reader is confidently prepared to pursue the topic via $IEEE$ journals or applications to engineering problems.

Lancaster, UK Gordon Blower

Contents

Chapter 1
Linear Systems and Their Description

In linear systems, we consider a machine made up of several components, which are connected together. We take $t > 0$ to be continuous time and consider the evolution of the system through time. The machine has an input $u = u(t)$, and output $y = y(t)$ and the internal state of the machine is described by a state $x = x(t)$, and we take u, x, y to be vector-valued functions of t. The state of a system is a set of variables whose values, together with the input and the equations describing the dynamics, will describe the future state and output of the system. Generally, we want to know how $y(t)$ depends upon $u(t)$. The component parts of the machine are represented by various linear operators, which leads to the terminology 'linear system'. In this book, we consider a special class of linear systems that we can analyze by means of linear algebra. When studying a linear system, it is important to have:

- general results which enable us to classify and describe a significant class of linear systems;
- specific methods for solving these linear systems;
- results that are in a form that allows effective and explicit computation of solutions, usually using computers.

In this book, we achieve these criteria for (A, B, C, D) systems.

1.1 Linear Systems and Their Description

Let $\mathbb{C}$ be the field of complex numbers, let V and W be vector spaces over $\mathbb{C}$, so $\lambda f + \mu g \in V$ for all $f, g \in V$ and $\lambda, \mu \in \mathbb{C}$. (We use $\mathbb{C}$ since the results of basic linear algebra about square matrices work best for $\mathbb{C}$, and we also use some results from complex analysis.) Time is $t > 0$. A map $L : V \to W$ is called linear if

$$L(\lambda f + \mu g) = \lambda L f + \mu L g. \tag{1.1}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_1

Example 1.1 The following give the basic examples of linear maps and their diagrammatic representation:

$$V = W = \{\text{continuously differentiable functions} \quad f : [0, \infty) \to \mathbb{C}\}.$$

Notation (Operations)

1. [(i)] Differentiator $Lf = \frac{df}{dt}$; symbolized by $[d/dt]$;
2. [(ii)] Integrator $Lf(x) = \int_0^x f(t)\,dt$, symbolized as $[\int]$;
3. [(iii)] an amplifier acts by multiplication by $a \in \mathbb{C}$, symbolized as $[a]$;
4. [(iv)] a matrix multiplier acts by multiplication on the left by a matrix A on a column vector v as in $v \mapsto Av$, symbolized by $[A]$;
5. [(v)] Multiplication by $h \in V$, $Lf(t) = h(t)f(t)$ symbolized as $[h]$;
6. [(vi)] Evaluation at t_0, $f \mapsto f(t_0)$ symbolized as $[\delta_{t_0}]$.

Definition 1.2 (Diagrams) Let V be the space of infinitely differentiable functions $f : (0, \infty) \to \mathbb{C}$. Let $u \in V$ be the input, $y \in V$ be the output. A diagram is a graph built up from vertices u, y and others chosen from

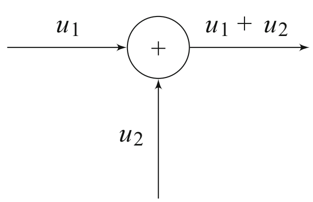
$$\Big\{u, y, [\delta_0], \oplus, \cdot, [d/dt], [\int], [a], [h]\Big\}$$

which are connected by directed edges, drawn as arrows. The following apply.

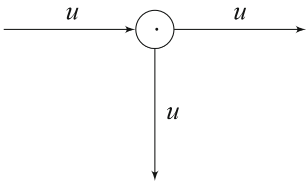
(1) u is the input and y is the output. The vertices u and y have degree one; whereas all other vertices have degree two or three with one or two arrows pointing into the vertex, and one or two pointing out.
(2) The graph is simple, so there are no multiple edges, and no vertex is directly connected to itself by some edge. All the vertices lie on some directed path consisting of consecutive arrows from u to y.
(3) If the diagram contains a circuit, then we say that there is feedback.
(4) If the diagram does not contain a circuit, then the system is 'open' or 'straight through'.

Block Diagrams

1. [(i)] Summing junction $\oplus$ has $y = u_1 + u_2$.

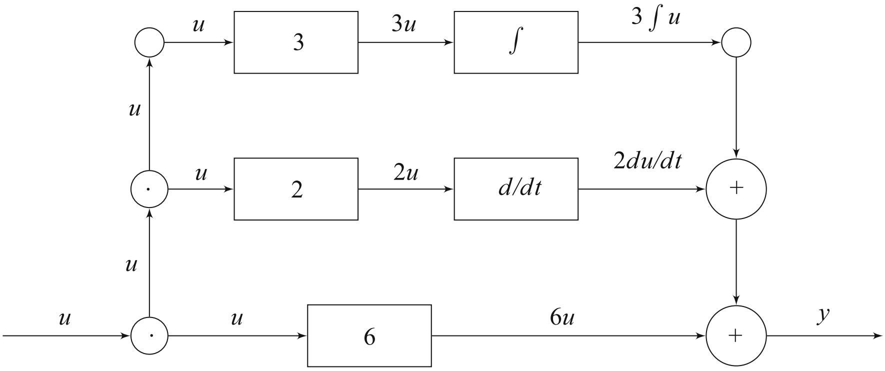


2. [(ii)] Tap • splits an input u into two copies of u (the effect of a tap is like voltage at an electrical junction, not like water flow in a plumbing).



Example 1.3 We can build up more complicated systems as in

$$y = 3\int u + 2\frac{du}{dt} + 6u. \tag{1.2}$$



1.2 Feedback

Linear systems have two basic types, namely open or closed loop.

Open Loop Here the input u is subject to linear operations and produces an output y. For example

$$y(t) = atu(t) + \frac{d^2u}{dt^2}(t). \tag{1.3}$$

Some machines are open loop, for instance rockets, or primitive turbines.

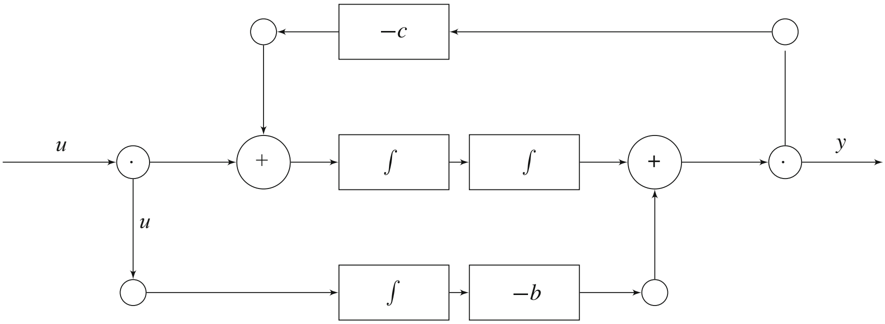
Closed Loop or Feedback Systems Here we take the input, subject it to linear operations, and also take the output, feed it back into the system, apply linear operations to the output and then add the modified input and the modified output. Most modern alliances involve some feedback or control systems. For instance, car engines, wind turbines involve feedback systems to ensure that their rate of rotation is under control. For example

$$\frac{d^2y}{dt^2} + cy = u - b\frac{u}{dt} \tag{1.4}$$

can be written as

$$y = -c\int\int y - b\int u + \int\int u. \tag{1.5}$$

Here we take the output y, integrate y twice, and add to the twice integrated input $\int\int u$ and the once integrated input multiplied by $-b$. Later, we'll use feedback to stabilize linear systems.



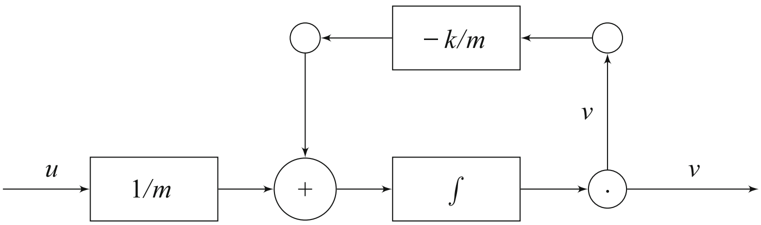
Example 1.4 (Negative Feedback) Feedback can occur by force of nature. A cyclist pedals harder, hence goes faster, but the faster the cyclist goes, the greater the air resistance. Let m be the mass of the cyclist, v the velocity, k a constant of proportionality, and u the force imparted on the pedals. Then by Newton's second law of motion,

$$m\frac{dv}{dt} = -kv + u \tag{1.6}$$

which we can express as a feedback system

$$v = \frac{-k}{m}\int v + \frac{1}{m}\int u, \tag{1.7}$$

and the sign of $-k/m$ indicates negative feedback.



Positive feedback occurs in the following examples, naturally or by design.

Example 1.5 (Bird Populations) Let x be the number of birds on an isolated island, say puffins on Ailsa Craig. The birth rate of birds is proportional to the number of birds, and birds can come and go by flying to other nesting sites at rate u, so, for some $k > 0$,

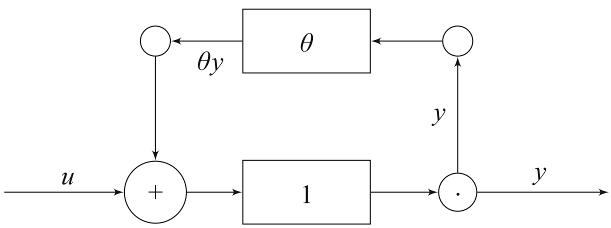
$$\frac{dx}{dt} = kx + u. \tag{1.8}$$

Example 1.6 (Turbocharger) A car engine has a turbocharger. This consists of a simple gas turbine driven by the exhaust outlet, connected by a shaft to a turbine which forces air into the inlet manifold. The faster the engine revolves, the more exhaust it produces, so the turbine forces more air into the engine, so the engine goes faster and so on.

Most practical devices of this kind also incorporate some negative feedback so that they do not damage themselves or their users.

Example 1.7 (Black's Amplifier) Let $0 < \theta < 1$. The output y is fed back into the input u, after multiplication by θ, so $y = \theta y + u$, hence $y = u/(1-\theta)$ is an amplified version of the input. This will also amplify any noise, so is not practical by itself as an amplifier.

Note that the summing junction lies to the left of the tap in this diagram. The loop is characteristic of block diagrams of feedback systems.



1.3 Linear Differential Equations

Let t be the independent variable. By combining amplifiers, summing junctions and differentiators, we can construct linear differential operators

$$Lf(t) = a_n(t)\frac{d^n f}{dt^n} + a_{n-1}(t)\frac{d^{n-1} f}{dt^{n-1}} + \cdots + a_0(t) f(t); \tag{1.9}$$

the number of derivatives n is the order of L; the $a_j(t)$ are the coefficient (functions). When the a_j are constants, we talk about a linear differential operator with constant coefficients. A linear equation of order n is

$$a_n(t)\frac{d^n f}{dt^n} + a_{n-1}(t)\frac{d^{n-1} f}{dt^{n-1}} + \cdots + a_0(t) f(t) = u(t), \tag{1.10}$$

where $a_n(t), \ldots, a_0(t), u(t)$ are given and $f(t)$ is to be found.

Proposition 1.8 *The differential equation*

$$a_n\frac{d^n y}{dt^n} + \cdots + a_0 y = b_m\frac{d^m u}{dt^m} + b_{m-1}\frac{d^{m-1} u}{dt^{m-1}} + \cdots + b_0 u(t) \tag{1.11}$$

with constant coefficients can be realized as a feedback system with input u and output y involving taps, amplifiers, summing junctions, integrators and differentiators.

Proof When $a_n \neq 0$, we integrate n times and divide by a_n to get

$$y = -\frac{a_{n-1}}{a_n}\int y - \cdots - \frac{a_0}{a_n}\int^{(n)} y + \frac{b_m}{a_n}\int^{(n)} \frac{d^m u}{dt^m} + \cdots + \frac{b_0}{a_n}\int^{(n)} u. \tag{1.12}$$

Then it is straightforward to realize the system.

The left half of the diagram is an open loop system

$$x = \frac{b_m}{a_n}\frac{d^m u}{dt^m} + \frac{b_{m-1}}{a_n}\frac{d^{m-1} y}{dt^{m-1}} + \cdots + \frac{b_0}{a_n}u;$$

whereas the right-half is a feedback system

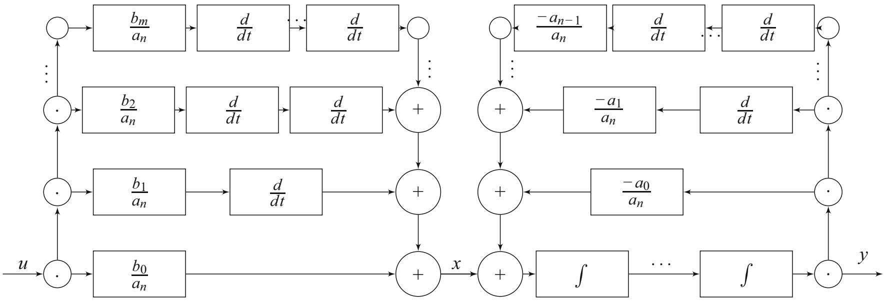
$$\frac{d^n y}{dt^n} + \frac{a_{n-1}}{a_n}\frac{d^{n-1} y}{dt^{n-1}} + \cdots + \frac{a_0}{a_n}y = x,$$

or

$$y = \int\int \cdots \int \Big(- \frac{a_{n-1}}{a_n}\frac{d^{n-1} y}{dt^{n-1}} - \cdots - \frac{a_0}{a_n}y + x\Big),$$

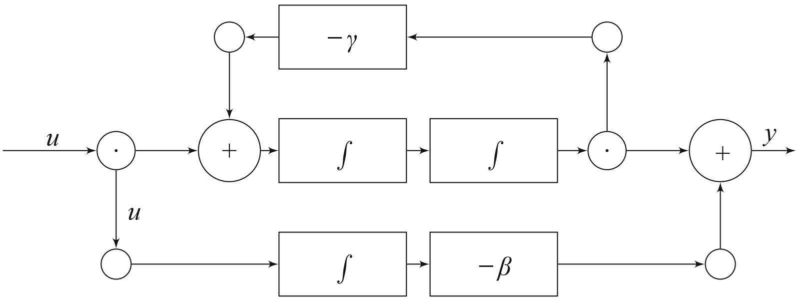
with n integrations. □

Differential Equations as Feedback Systems
The ellipses (dots) in the diagram indicate omitted terms.



1.4 Damped Harmonic Oscillator

Damped harmonic oscillators are important, because they (i) arise in various physical systems such as a mass on a spring or electrical systems such as inductor and capacitor circuits; and (ii) they exhibit the effects that describe more general linear systems.



Let t be time and y displacement from rest of a mass on a spring. The velocity is dy/dt and acceleration d^2y/dt^2, the mass is driven by an external driving force $u(t)$. The equation is

$$\frac{d^2y}{dt^2} + \beta\frac{dy}{dt} + \gamma y = u \tag{1.13}$$

where β and γ are constants; usually $\gamma > 0$ and $\beta \geq 0$. This models the suspension of a car moving along a road. Here $u(t)$ represents the force imparted on the car by the road; the suspension involves a spring which gives a restoring force $-\gamma y$, while the shock absorbers give a damping force $-\beta dy/dt$. We can write this as a feedback system, using the formula

$$y = -\beta \int y - \gamma \int \int y + \int \int u. \tag{1.14}$$

Example 1.9 (Matrix Form of the Damped Harmonic Oscillator) We introduce a new state variable v, the velocity, so $dy/dt = v$. Then we have a pair of equations $dy/dt = v$ and $dv/dt = -\beta v - \gamma y + u$. We put these together in a matrix equation

$$X = \begin{bmatrix} y \\ v \end{bmatrix}, \tag{1.15}$$

$$A = \begin{bmatrix} 0 & 1 \\ -\gamma & -\beta \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad D = 0 \tag{1.16}$$

so the matrix form of the differential equation is

$$\frac{dX}{dt} = AX + Bu \tag{1.17}$$

$$y = CX + Du \tag{1.18}$$

We often choose input $u(t) = e^{i\omega t} = \cos(\omega t) + i\,\sin(\omega t)$, with angular frequency ω to model a periodic input force.

1.5 Reduction of Order of Linear ODE

Lemma 1.10 *The linear ordinary differential equation*

$$\frac{d^n y}{dt^n} + a_{n-1}(t)\frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_0(t)y(t) = u(t), \tag{1.19}$$

may be expressed as the matrix system

$$\frac{dX}{dt} = AX + Bu \tag{1.20}$$

where X is $(n \times 1)$, A is $(n \times n)$ and B is $(n \times 1)$.

Thus we replace a n^{th} order differential equation with one independent variable with a first order differential equation with a $(n \times 1)$ vector independent variable.

Proof We express the differential equation in terms of matrices. Let

$$X = \begin{bmatrix} y \\ dy/dt \\ \vdots \\ d^{n-1}y/dt^{n-1} \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ -a_0 & -a_1 & \dots & -a_{n-1} \end{bmatrix} \tag{1.21}$$

so the differential equation is

$$\frac{dX}{dt} = AX + Bu. \tag{1.22}$$

□

Example 1.11 (Constant Coefficient Case) The advantages of Lemma 1.10 are that first order differential equations are apparently easier to solve than n^{th} order equation, and we can use linear algebra on the matrix A; see the discussion below. Suppose that the a_j are constant, or equivalently A is constant; then the system is said to have constant coefficients, and the differential equation can be expressed as the feedback system

$$X = A \int X + B \int U. \tag{1.23}$$

Definition 1.12 (Companion Matrix) Let $a_{n-1}, a_{n-2}, \dots, a_0 \in \mathbb{C}$. Then the $(n \times n)$ matrix

$$A_c = \begin{bmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \ddots \\ \vdots & \vdots & \ddots & \ddots \\ -a_0 & -a_1 & \dots & -a_{n-1} \end{bmatrix} \tag{1.24}$$

is called a companion matrix for $a_{n-1}, \dots, a_0$ (note the signs of the a_j).

1.6 Exercises

Exercise 1.1 Express the following differential and integral equations as block diagrams

$$(i) \qquad y = 4\frac{d^2u}{dt^2} + 3\int u + 6u;$$

$$(ii) \qquad y + 4\frac{d^2y}{dt^2} = 2\frac{du}{dt} + 7\int u.$$

Exercise 1.2

(i) Express the following coupled differential equations as a block diagram, where u is the input, y is the output, x is a state variable, and a, b, c and d are constants:

$$\frac{dx}{dt} = ax + bu,$$
$$\frac{dy}{dt} = cx + du.$$

(ii) Express the following coupled differential and integral equations as a block diagram, where u_1 and u_2 are the inputs, y is the output, x is a state variable, and a, c, b_1, b_2, d_1 and d_2 are constants:

$$\frac{dx}{dt} = ax + b_1u_1 + b_2u_2,$$
$$\frac{dy}{dt} = cx + d_1u_1 + d_2u_2.$$

Exercise 1.3 A simple harmonic oscillator satisfies

$$m\frac{d^2x}{dt^2} + kx = u,$$

where t is time, and k and m are positive constants. By introducing an extra state variable $v = dx/dt$, write this as a first order system of differential equations.

Chapter 2
Solving Linear Systems by Matrix Theory

2.1 Matrix Terminology

Let V and W be finite-dimensional complex vector spaces, and suppose that V has basis $\{e_j : j = 1, \dots, n\}$ and W has basis $\{f_j : j = 1, \dots, m\}$. We generally write vectors as columns, so

$$v = \sum_{j=1}^{n} v_j e_j \leftrightarrow \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = [v_1; v_2; \dots; v_n] \in \mathbb{C}^{n \times 1}.$$

A map $T : V \to W$ is defined to be a linear transformation if $T(\lambda v + \mu w) = \lambda T v + \mu T w$ for all $v, w \in V$ and $\lambda, \mu \in \mathbb{C}$. We write $T(e_k) = \sum_{j=1}^{m} T_{jk} f_j$ for $k = 1, \dots n$ where the coefficients $T_{jk} \in \mathbb{C}$ are uniquely determined, and thus we associate T with the $m \times n$ complex matrix $[T_{jk}]_{j=1,\dots,m;k=1,\dots n}$. Conversely, any such $m \times n$ complex matrix determines a unique linear transformation T with respect to the specified bases via this formula, and $Tv = \sum_{j=1}^{m} \sum_{k=1}^{n} T_{jk} v_k f_j$ in the preceding notation.

Definition 2.1

(i) The range of T is $\{w \in W : w = Tv, v \in V\}$ which is otherwise known as the image of T. This is a vector space with dimension called the rank of T, denoted by rank(T);
(ii) The null space of T is null$(T) = \{v \in V : Tv = 0\}$, which is a vector space of dimension called the nullity of T, denoted by nullity(T).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_2

Theorem 2.2 *The rank and nullity of a linear transformation $T : V \to W$ between finite-dimensional vector spaces satisfy*

$$rank(T) + nullity(T) = dimension(V). \tag{2.1}$$

Proof See [6] page 213. □

Definition 2.3

(1) The elementary row operations on a complex matrix are:

  (i) interchanging two rows;
  (ii) multiplying one row by a nonzero scalar in $\mathbb{C}$;
  (iii) adding a complex multiple of one row to another row.

(2) Matrices S and T are row equivalent when S can be transformed to T by some finite sequence of elementary row operations.
(3) A matrix is in echelon form when

$$\begin{bmatrix} 0 & a_1 & * & * & * & * \\ 0 & 0 & 0 & a_2 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.2}$$

has row leaders a_1 and a_2 that are non-zero, all the entries directly below a row leader are all zero, row leaders appear to the right of the row leaders of rows above them, and zero rows are at the bottom of the matrix. A matrix in echelon form is moreover in reduced echelon form when the row leaders are all 1, and the entries directly above row leaders are all zero, as in

$$\begin{bmatrix} 0 & 1 & * & 0 & * & * \\ 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.3}$$

Lemma 2.4 (Rank of a Matrix) *For $X_1, \ldots, X_n \in \mathbb{C}^{m\times 1}$, the following quantities are equal:*

(i) the dimension of the vector space U spanned by $X_1, \ldots, X_n$;
(ii) the number of linearly independent columns in the $m \times n$ matrix $T = [X_1 \ldots X_n]$;
(iii) the number of linearly independent rows in T;
(iv) the number of nonzero rows in any row-equivalent echelon form of T.

Proof See [6]. □

The rank can be computed in MATLAB and Scilab using $rank(T)$. Alternatively, by row-reducing the matrix $[T_{jk}]$ to echelon form by elementary row operations, one can find a basis for the range of T, and thus compute rank(T) via Lemma 2.4 (iv).

Definition 2.5 (Transpose) Let $T \in M_{n\times m}(\mathbb{C})$ be a matrix with n rows, m columns and entry $t_{j,k}$ is row j and column k. Then the transpose of T is $T^\top \in M_{m\times n}(\mathbb{C})$ is the matrix with m rows, n columns and entry $t_{k,j}$ in row j and column k.

We suppose in particular that $m = n$, and we choose the same basis $\{e_j : j = 1, \dots, n\}$ for V and W, and naturally write $V = W$. Then we prefer the notation $A : V \to V$ for the linear transformation of V, and the $n \times n$ complex matrix that represents this linear transformation with respect to $\{e_j : j = 1, \dots, n\}$. Observe that by Theorem 2.2, $\mathrm{null}(A) = \{0\}$ if and only if $\mathrm{rank}(A) = n$. Under these equivalent conditions, A has an inverse transformation A^{-1} such that $AA^{-1} = A^{-1}A = I$. This inverse matrix may be expressed in terms of determinants via Proposition 2.7.

Determinants In this book, we give only a brief sketch of the theory, mainly to establish notation, and refer the reader to [60] or [8] for a fuller account in the same spirit. We begin with the 2×2 case

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \qquad (a, b, c, d \in \mathbb{C}). \tag{2.4}$$

Then we obtain large determinants by expanding in terms of smaller ones. To obtain the determinant of

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{2.5}$$

observe the chess-board of signs

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}, \tag{2.6}$$

pick one column, say the second, then multiply out with signs from the chess-board

$$\det A = -a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{32} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \tag{2.7}$$

where the 2×2 determinants exclude the row and column of their coefficient, and can be computed by the formula (2.4).

Definition 2.6 (Adjugate)

(i) The $n \times n$ sign chessboard is the matrix that has (j, k) entry given by $(-1)^{j+k}$ for $j, k = 1, \dots, n$.

(ii) The determinant of the submatrix of A that excludes row j and column k, multiplied by $(-1)^{j+k}$, is called the cofactor A_{jk} of a_{jk}, so

$$\det A = \sum_{j=1}^{n} a_{jk} A_{jk} \tag{2.8}$$

is the expansion by column k for all $k = 1, \dots, n$.

(iii) The adjugate matrix $\mathrm{adj}(A)$ is the transpose of the matrix $[A_{jk}]$ of cofactors. (In some books, our adjugate is called the adjoint, but we use this terms for something else in Definition 2.15.)

(iv) A square matrix $[a_{j,k}]_{j,k=1}^{n}$ is said to be lower triangular if all the entries above the leading diagonal are zero, , so $a_{j,k} = 0$ for all $1 \leq j < k \leq n$.

Proposition 2.7

(i) For all square matrices A adj$(A) = (\det A)I$.

(ii) The determinant of a lower triangular matrix equals the product of the entries on the leading diagonal, so $\det[a_{j,k}] = a_{1,1}a_{2,2} \dots a_{n,n}$.

Proof

(i) A square matrix with a repeated row has zero determinant, so $\sum_{j=1}^{n} a_{jk} A_{jm} = 0$ for $k \neq m$, which gives the off-diagonal entries of $A\,\mathrm{adj}(A)$ to be zero. For the diagonal entries, we use the definition of the cofactor.

(ii) One can check this by repeatedly expanding by the first row of the determinant.

□

Definition 2.8 A square matrix A is unimodular if $\det A = 1$. (We use a more restrictive definition than some authors who permit ± 1.)

2.2 Characteristic Polynomial

Definition 2.9 A complex polynomial $p(s)$ is monic and of degree n if it has the form $p(s) = s^n + a_{n-1}s^{n-1} + \dots + a_0$ for coefficients $a_j \in \mathbb{C}$.

Definition 2.10 (Characteristic Polynomial) The characteristic polynomial of a $(n \times n)$ complex matrix A is $\chi_A(\lambda) = \det(\lambda I - A)$, where I is the $(n \times n)$ identity matrix.

Some books on linear algebra $c_A(\lambda) = \det(A - \lambda I)$. The definition used in this book is standard in control theory. Also

$$\chi_A(\lambda) = \det(\lambda I - A) = \lambda^n - \lambda^{n-1}\mathrm{trace}(A) + \dots + (-1)^n \det A, \tag{2.9}$$

so $\chi_A(\lambda)$ is a monic polynomial of degree n.

Lemma 2.11 *If* $\det(sI - A) \neq 0$*, then* $sI - A$ *is invertible and*

$$(sI - A)^{-1} = (\det(sI - A))^{-1} adj(sI - A). \tag{2.10}$$

Proof This follows directly from Proposition 2.7 applied to $sI - A$. □

This (2.10) may or may not be an appropriate formula for computing the inverse, depending on n. It does tell us that $sI - A$ is invertible, except at finitely many values of s. Often $(sI - A)^{-1}$ is called the resolvent of A.

Proposition 2.12 (Characteristic Polynomials of Companion Matrix) *The characteristic polynomial of the companion matrix* A_c *is*

$$\det(\lambda I - A_c) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0. \tag{2.11}$$

Thus any monic complex polynomial arises as the characteristic polynomial of some complex matrix.

Proof We prove this by induction on n. Let P_n be the statement that the above identity holds for some positive integer n. Then P_1 is trivially true. Assume that the identity holds for $1, \dots, n-1$ and consider P_n. We expand the determinant by the first column, and obtain

$$\det(\lambda I - A_c) = \det \begin{bmatrix} \lambda & -1 & 0 & \dots \\ 0 & \lambda & -1 & \ddots \\ \vdots & \vdots & \ddots & \ddots \\ a_0 & a_1 & \dots & \lambda + a_{n-1} \end{bmatrix}$$

$$= \lambda \det \begin{bmatrix} \lambda & -1 & 0 & \dots \\ 0 & \lambda & -1 & \ddots \\ \vdots & \vdots & \ddots & \ddots \\ a_1 & a_2 & \dots & \lambda + a_{n-1} \end{bmatrix} + (-1)^{n-1} a_0 \det \begin{bmatrix} -1 & 0 & 0 & \dots \\ \lambda & -1 & 0 & \ddots \\ 0 & \ddots & \ddots & \ddots \\ 0 & \dots & \lambda & -1 \end{bmatrix} \tag{2.12}$$

so we use the induction hypothesis to deal with the first determinant, and observe that the second is lower triangular, so by Proposition 2.7 (ii)

$$\begin{aligned} \det(\lambda I - A) &= \lambda(\lambda^{n-1} + a_{n-1}\lambda^{n-2} + \dots + a_2\lambda + a_1) + a_0 \\ &= \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0. \end{aligned} \tag{2.13}$$

Given any monic complex polynomial $p(\lambda)$ of degree n, we choose the $n \times n$ companion matrix with entries from the coefficients of $p(\lambda)$ which has characteristic polynomial $p(\lambda)$. □

Example 2.13 Proposition 2.12 is an existence result, not a uniqueness theorem. The matrices

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

both have characteristic polynomial s^2, although they are not similar.

Example 2.14 The polynomial $f(s) = s^4 + s^3 + 20s^2 + 400s + 200$ is the characteristic polynomial of

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -200 & -400 & -20 & -1 \end{bmatrix}. \tag{2.14}$$

Numerical results show that $f(s) = 0$ has two roots -6.5527 and -0.5130 in $(-\infty, 0)$ and a pair complex conjugate roots $3.0329 \mp i7.0922$ in the right half-plane $\{s \in \mathbb{C} : \Re s > 0\}$. In Sect. 2.8 onwards, we describe polynomials with all their roots in the open left half-plane as stable. Therefore, $f(s)$ is unstable, although all its coefficients are positive.

2.3 Norm of a Vector

In this section we use complex conjugates, so $z = x + iy \in \mathbb{C}$ has $\bar{z} = x - iy$ for $x, y \in \mathbb{R}$, so $|z|^2 = z\bar{z} = x^2 + y^2$. We write $\Re z = x$ and $\Im z = y$.

Definition 2.15 (Adjoint)

(i) For a column $z = \mathrm{col}[z_j]_{j=1}^n \in \mathbb{C}^{n\times 1}$, we take $z' = \mathrm{row}[\bar{z}_j]_{j=1}^n \in \mathbb{C}^{1\times n}$.
(ii) We define the adjoint of a $n \times m$ complex matrix $T = [t_{jk}]$ to be the $n \times m$ matrix $T' = [\bar{t}_{kj}]$, found by interchanging the rows and columns and taking the complex conjugate of each entry.

Remark 2.16

(i) Here we employ the MATLAB notation T' for adjoint. For real matrices T the adjoint coincides with the transpose $T^\top$, so $T' = T^\top$. In this case, our notation is consistent with [8]. For derivatives in the sense of calculus, we use df/dt.
(ii) In some books, the adjoint is called the Hermitian conjugate or conjugate transpose and the notation used is A^*. Note that in MATLAB $A * B$ is the usual product of the matrices A and B, with no transposition or conjugation involved. In physics, a common notation is $A^\dagger$, although often a different definition is used for the inner product.

Let $V = \mathbb{C}^{n\times 1}$, the complex vector space of column vectors. The standard inner product on V is defined for

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}, \qquad w = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

by

$$\langle z, w\rangle = w' z = \begin{bmatrix} \bar{w}_1 \dots \bar{w}_n \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \sum_{j=1}^{n} \bar{w}_j z_j \tag{2.15}$$

equivalently written $z = \text{column}(z_j)_{j=1}^n$ $w = \text{column}(w_j)_{j=1}^n$ by

$$\langle z, w\rangle = \bar{w}^\top z = \sum_{j=1}^{n} z_j \bar{w}_j. \tag{2.16}$$

Then $\langle z, z\rangle = \sum_{j=1}^n |z_j|^2$, so $\langle z, z\rangle \geq 0$, with $\langle z, z\rangle = 0 \Rightarrow z = 0$;

$$\langle z + u, w\rangle = \langle z, w\rangle + \langle u, w\rangle \tag{2.17}$$

$$\langle \lambda z, w\rangle = \lambda\langle z, w\rangle, \quad \overline{\langle z, w\rangle} = \langle w, z\rangle. \tag{2.18}$$

2.4 Cauchy–Schwarz Inequality

On V the standard norm is the Euclidean norm for $z = \text{column}(z_j)_{j=1}^n$

$$\|(z_j)_{j=1}^n\| = \Big(\sum_{j=1}^{n} |z_j|^2\Big)^{1/2} = (\bar{z}^\top z)^{1/2} = \langle z, z\rangle^{1/2}. \tag{2.19}$$

Proposition 2.17 (Cauchy–Schwarz Inequality)

(i) All $z, w \in V$ *satisfy*

$$|\langle z, w\rangle| \leq \|z\|\|w\|, \tag{2.20}$$

(ii) and the triangle inequality

$$\|z + w\| \leq \|z\| + \|w\|. \tag{2.21}$$

Proof

(i) Recall that a nonzero complex number ζ has a polar decomposition $\zeta = se^{i\theta}$, where $\theta \in (-\pi, \pi]$ and $s = |\zeta| > 0$; so $e^{-i\theta}\zeta = s > 0$. If $\langle z, w\rangle = 0$, then (i) is clearly true. Otherwise, there exists $u \in \mathbb{C}$ such that $u\bar{u} = 1$ and $u\langle z, w\rangle = |\langle z, w\rangle|$. Now we have a real quadratic in the real variable t which is non negative

$$\begin{aligned} 0 \le \|tw + uz\|^2 &= \langle tw + uz, tw + uz\rangle \\ &= t^2\langle w, w\rangle + t\langle w, uz\rangle + \langle uz, tw\rangle + \langle uz, uz\rangle \\ &= t^2\|w\|^2 + 2t|\langle z, w\rangle| + \|z\|^2. \end{aligned} \tag{2.22}$$

We cannot have a pair of distinct real roots, since $y = t^2\|w\|^2 + 2t|\langle z, w\rangle| + \|z\|^2$ is a parabola that does not cross the t-axis. Hence this quadratic has discriminant $b^2 - 4ac \le 0$, so

$$4|\langle z, w\rangle|^2 \le 4\|z\|^2\|w\|^2. \tag{2.23}$$

(ii) By (i) we have

$$\begin{aligned} \|z + w\|^2 &= \langle z + w, z + w\rangle \\ &= \langle z, z\rangle + \langle z, w\rangle + \langle w, z\rangle + \langle w, w\rangle \\ &= \|z\|^2 + 2\Re\langle z, w\rangle + \|w\|^2 \\ &\le \|z\|^2 + 2\|z\|\|w\| + \|w\|^2 \\ &= \big(\|z\| + \|w\|\big)^2. \end{aligned} \tag{2.24}$$

□

Definition 2.18 (Matrix Norm) Suppose that $A \in M_{n\times n}(\mathbb{C})$. Then A operates on the space $V = \mathbb{C}^{(n\times 1)}$ of column vectors so $A : V \to V : v \mapsto Av$ by multiplication on the left. Then the matrix norm of A is

$$\|A\| = \sup\{\|Av\| : v \in V; \|v\| \le 1\}. \tag{2.25}$$

The supremum in this definition can be calculated in various ways, depending upon the specific form of A, as in Proposition 2.19 and Lemma 2.21.

Proposition 2.19

(i) For a square diagonal matrix D with diagonal entries $\lambda_1, \dots, \lambda_n$, the norm is

$$\|D\| = \max\{|\lambda_j| : j = 1, \dots, n\}. \tag{2.26}$$

(ii) Let A have columns $A = [A_1, A_2, \dots, A_n]$, and let column $A_j \in \mathbb{C}^{n\times 1}$ have norm $\|A_j\|$. Then $\|A\| \le (\sum_{j=1}^n \|A_j\|^2)^{1/2}$.

Proof

(i) Let $(e_j)_{j=1}^n$ be the usual orthonormal basis of $\mathbb{C}^{n\times 1}$ and $v = \sum_{j=1}^n v_j e_j$, so $\|v\| = (\sum_{j=1}^n |v_j|^2)^{1/2}$. Then $Dv = \sum_{j=1}^n v_j De_j = \sum_{j=1}^n v_j \lambda_j e_j$, so

$$\|Dv\|^2 = \sum_{j=1}^n |\lambda_j v_j|^2 \le \max_k |\lambda_k|^2 \sum_{j=1}^n |v_j|^2 = \max_k |\lambda_k|^2 \|v\|^2. \tag{2.27}$$

We can achieve equality in this inequality by considering the largest $|\lambda_k|$, and selecting $v = e_k$.

(ii) Here we have $Av = \sum_{j=1}^n v_j Ae_j = \sum_{j=1}^n v_j A_j$, so by the triangle inequality,

$$\|Av\| \le \sum_{j=1}^n |v_j|\|A_j\| \le \Big(\sum_{j=1}^n |v_j|^2\Big)^{1/2}\Big(\sum_{j=1}^n \|A_j\|^2\Big)^{1/2}, \tag{2.28}$$

where the last step follows by Cauchy-Schwarz, so

$$\|Av\| \le \Big(\sum_{j=1}^n \|A_j\|^2\Big)^{1/2} \|v\|. \tag{2.29}$$

□

Remark 2.20 This inequality (ii) of Proposition 2.19 show that the norm of any finite matrix is finite. With j^{th} column $A_j = [a_{kj}]_{k=1}^n$ we can consider $A = [A_1 \; \dots \; A_n]$ with

$$\|A\|_{HS} = \Big(\sum_{j,k=1}^n |a_{kj}|^2\Big)^{1/2} = \Big(\sum_{j=1}^n \|A_j\|^2\Big)^{1/2} \tag{2.30}$$

which defines the Hilbert-Schmidt norm $\|A\|_{HS}$ of A. This is straightforward to calculate, but generally gives an overestimate on $\|A\|$.

Lemma 2.21 (Properties of the Norm of a Matrix) *The matrix norm satisfies, for $A, B \in M_{n\times n}(\mathbb{C})$,*

$$(i) \quad \|A + B\| \le \|A\| + \|B\|; \quad (ii) \quad \|\lambda A\| = |\lambda|\|A\| \qquad (\lambda \in \mathbb{C}); \tag{2.31}$$

$$(iii) \quad \|AB\| \le \|A\|\|B\|, \quad (iv) \quad \|A'\| = \|A\|, \quad (v) \quad \|A\|^2 = \|A'A\|. \tag{2.32}$$

Proof

(i) Let $v \in V$ have $\|v\| \leq 1$. We have

$$\|(A+B)v\| = \|Av + Bv\| \leq \|Av\| + \|Bv\| \leq \|A\| + \|B\| \tag{2.33}$$

so $\|A+B\| \leq \|A\| + \|B\|$.

(ii) We have $\|(\lambda A)v\| = |\lambda|\|Av\|$, so $\|\lambda A\| = |\lambda|\|A\|$.

(iii) We observe that $\|A\| = \inf\{t : \|Aw\| \leq t\|w\| : \forall w \in V\}$. Then with $w = Bv$, we have

$$\|(AB)v\| \leq \|A\|\|Bv\| \leq \|A\|\|B\|\|v\|, \tag{2.34}$$

so $\|AB\| \leq \|A\|\|B\|$.

(iv) First observe that

$$\begin{aligned}\|A\| &= \sup\{\Re\langle Ax, y\rangle : \|x\| = \|y\| = 1\} \\ &= \sup\{\Re\langle x, A'y\rangle : \|x\| = \|y\| = 1\} = \|A'\|.\end{aligned} \tag{2.35}$$

(v) We choose $x \neq 0$ so that $\|Ax\| = \|A\|\|x\|$, then

$$\|A\|^2\|x\|^2 = \langle Ax, Ax\rangle = \langle A'Ax, x\rangle \leq \|A'A\|\|x\|^2 \leq \|A\|^2\|x\|^2, \tag{2.36}$$

so we have equality throughout.

□

Definition 2.22 (Polynomial Functions of a Matrix) Let A be a $n \times n$ complex matrix. As in basic linear algebra we form polynomials in A. Let I be the $n \times n$ identity matrix. We can form $A, A^2, A^3, \ldots$ by matrix multiplication, and hence given $g(z) = a_m z^m + \cdots + a_0$ we build polynomials

$$g(A) = a_m A^m + a_{m-1}A^{m-1} + \cdots + a_0 I \tag{2.37}$$

for complex coefficients a_j. We regard $A^0 = I$, the identity matrix.

Definition 2.23 (Eigenvalue Equation) An eigenvector is a non zero solution v of

$$Av = \lambda v \tag{2.38}$$

where $\lambda \in \mathbb{C}$ is the corresponding eigenvalue. This (2.38) is called the eigenvalue equation.

Lemma 2.24 *The eigenvalues of $n \times n$ complex matrix A are the roots of the characteristic equation*

$$\chi_A(\lambda) = 0. \tag{2.39}$$

Proof Recall that $\det(sI - A) = \chi_A(s)$. By the Fundamental Theorem of Algebra [6], there are n complex roots, counted according to algebraic multiplicity. When $\chi_A(\lambda) = 0$, the matrix $\lambda I - A$ is not invertible, so by the rank-nullity theorem 2.2 there exists $v \in V$, with $v \neq 0$ and $Av = \lambda v$ and λ is an eigenvalue. Conversely, if there exists $v \neq 0$ and $\lambda \in \mathbb{C}$ such that $Av = \lambda v$, then $\lambda I - A$ is not invertible, so $\chi_A(\lambda) = 0$. □

Definition 2.25 (Spectrum) The spectrum $\mathrm{spec}(A)$ of an $n \times n$ complex matrix A is the set of $\lambda \in \mathbb{C}$ such that $\lambda I - A$ does not have an inverse.

Remark 2.26

(i) The square matrix $\lambda I - A$ is invertible if and only if the nullspace of $\lambda I - A$ equals $\{0\}$, by the rank-nullity theorem 2.2. Hence the spectrum is the set of all the eigenvalues of A.

(ii) By the Lemma 2.24, the spectrum of A has at least one element and at most n elements. Often one lists roots of polynomial equations according to their algebraic multiplicity, so that the roots of $(s - \lambda)^2$ are listed as λ, λ. In this sense, there are n eigenvalues, listed according to algebraic multiplicity. However, with eigenvalues, we also need to consider the eigenvalue equation (2.38) as well as the characteristic equation (2.39).

(iii) The geometric multiplicity of λ is the number of linearly independent solutions of $Av = \lambda v$. Now by a slight extension of this Lemma 2.24 , one can show that

$$1 \leq (\text{geometric multiplicity}) \leq (\text{algebraic multiplicity}).$$

Lemma 2.27 (Similarity to a Diagonal Matrix) *For a complex $n \times n$ matrix A, the following are equivalent:*

(i) There exists an invertible matrix S such that $S^{-1}AS$ is a diagonal matrix;
(ii) there exist n linearly independent eigenvectors of A;
(iii) there exists a basis of $\mathbb{C}^n$ that consists of eigenvectors of A.

Proof $(iii) \Rightarrow (i)$ Let $X_j \neq 0$ be a $n \times 1$ columns forming a basis of $\mathbb{C}^n$ such that $AX_j = \lambda_j X_j$, and let $S = [X_1\ X_2\ \dots\ X_n]$; then the X_j are linearly independent by assumption and hence S has column rank n. Hence S is invertible. Now let

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots \\ 0 & \lambda_2 & 0 & \ddots \\ \vdots & \vdots & \ddots & \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} \tag{2.40}$$

and note the chain of identities

$$\begin{aligned} AS &= A\left[X_1 \dots X_n\right] = \left[AX_1 \dots AX_n\right] \\ &= \left[\lambda_1 X_1 \dots \lambda_n X_n\right] = \left[X_1 \dots X_n\right] D = SD \end{aligned} \tag{2.41}$$

where S is invertible, so $A = SDS^{-1}$. □

Proposition 2.28 (Functions of a Matrix) *Suppose that A has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Then there exists an invertible $n \times n$ matrix S such that*

$$g(A) = S \begin{bmatrix} g(\lambda_1) & 0 & 0 & \dots \\ 0 & g(\lambda_2) & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & g(\lambda_n) \end{bmatrix} S^{-1} \tag{2.42}$$

for all complex polynomials $g(\lambda)$.

Soon we'll extend this to the functions $g(x) = \exp(tx)$ and $g(x) = 1/(s - x)$.

Proof The eigenvectors corresponding to distinct eigenvalues are linearly independent, hence form a basis of $\mathbb{C}^n$, so we can introduce S as in the Lemma 2.27. Then

$$A^2 = SDS^{-1}SDS^{-1} = SD^2S^{-1} \tag{2.43}$$

and so on so,

$$\begin{aligned} g(A) &= a_m A^m + a_{m-1}A^{m-1} + \dots + a_0 I \\ &= S(a_m D^m + a_{m-1}D^{m-1} + \dots + a_0 I)S^{-1} = Sg(D)S^{-1}, \end{aligned} \tag{2.44}$$

and we can easily check that $g(D)$ is as above. □

Theorem 2.29 (Cayley–Hamilton) *Let A be a $n \times n$ complex matrix with characteristic polynomial $\chi_A(s)$. Then*

$$\chi_A(A) = 0. \tag{2.45}$$

Proof See [6]. □

Complex Exponential

For $z \in \mathbb{C}$ write $z = \Re z + i\Im z$ where $\Re z$ is the real part and $\Im z$ is the imaginary part. We define

$$\exp(z) = e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots, \tag{2.46}$$

which converges for all $z \in \mathbb{C}$. We have $\exp(z + w) = \exp(z)\exp(w)$ and $(d/dz)\exp(z) = \exp(z)$. Also, $e^{i\theta} = \cos\theta + i \sin\theta$ has $|e^{i\theta}| = 1$. Hence e^z has modulus $|e^z| = e^{\Re z}$ and argument $\arg e^z = \Im z$. In contemporary English, argument is often used to mean a dispute; it also means legal case; in mathematics, the term applies to the angle in the polar decomposition of a complex number. The latter is denoted arg, or Arg, especially when the values is taken in $(-\pi, \pi]$. MATLAB uses angle for argument, while engineers often use phase.

In some applications, we also need

$$\cosh(x + iy) = \cosh x \cos y + i \sinh x \sin y. \tag{2.47}$$

2.5 Matrix Exponential exp(A) or expm (A)

Definition 2.30 For any $n \times n$ complex matrix A, we define the matrix exponential by

$$\exp(A) = I + A + \frac{A^2}{2!} + \cdots + \frac{A^m}{m!} + \ldots.$$

The MATLAB command $expm(A)$ gives this series for a square complex matrix A; whereas the command $exp(A)$ gives the matrix arising from the exponential function applied to the entries of A individually, which is a quite different function.

Proposition 2.31 (Wedderburn)

(i) *For any* $A \in M_{n\times n}(\mathbb{C})$, *the exponential series converges, and* $\|\exp(A)\| \le e^{\|A\|}$.
(ii) $\exp(zA)\exp(wA) = \exp((z + w)A)$ *for all* $z, w \in \mathbb{C}$;
(iii) $\exp(zA)$ *has inverse* $\exp(-zA)$ *for all* $z \in \mathbb{C}$;
(iv) *Let* λ *be an eigenvalue of* A. *Then* $e^{z\lambda}$ *is an eigenvalue of* $\exp(zA)$.
(v)

$$\frac{d}{dz}\exp(zA) = A\exp(zA). \tag{2.48}$$

Proof

(i) Note that for a matrix X the entries X_{jk} satisfy $X_{jk} = \langle Xe_k, e_j\rangle$ so $|X_{jk}| \le \|X\|$. Also, by Lemma 2.21

$$\|A^2\| \le \|A\|^2, \ldots, \|A^m\| \le \|A\|^m, \tag{2.49}$$

so

$$p_m(A) = I + A + \frac{A^2}{2!} + \cdots + \frac{A^m}{m!} \tag{2.50}$$

satisfy

$$\|p_m(A) - p_k(A)\| \leq \left\| \frac{A^{k+1}}{(k+1)!} + \cdots + \frac{A^m}{m!} \right\| \leq \sum_{j=k+1}^{m} \frac{\|A\|^j}{j!} \tag{2.51}$$

where $e^{\|A\|} = \sum_{j=0}^{\infty} \|A\|^j / j!$ converges. Hence each entry of $p_m(A)$ converges as $m \to \infty$, giving $\exp(A)$ as the limit.

(ii) We write

$$p_m(zA)p_m(wA) = \sum_{j=0}^{m} \frac{z^j A^j}{j!} \sum_{k=0}^{m} \frac{w^k A^k}{k!} \tag{2.52}$$

and compare with

$$p_{2m}((z+w)A) = \sum_{r=0}^{2m} \frac{(z+w)^r A^r}{r!} = \sum_{r=0}^{2m} \sum_{s=0}^{r} \frac{z^s w^{r-s} A^r}{s!(r-s)!} \tag{2.53}$$

where $p_m(zA) \to \exp(zA)$, $p_m(wA) \to \exp(wA)$, and $p_{2m}((z+w)A) \to \exp((z+w)A)$, so $\exp((z+w)A) = \exp(zA)\exp(wA)$.

(iii) Observe that, by (ii)

$$\exp(zA)\exp(-zA) = I = \exp(-zA)\exp(zA) \tag{2.54}$$

(iv) Let $v \in V$ satisfy $v \neq 0$ and $Av = \lambda v$. Then

$$\exp(zA)v = \sum_{j=0}^{\infty} \frac{z^j A^j v}{j!} = \sum_{j=0}^{\infty} \frac{z^j \lambda^j v}{j!} = e^{z\lambda} v. \tag{2.55}$$

(v) We consider (ii), and obtain as $h \to 0$

$$\begin{aligned} \frac{\exp((z+h)A) - \exp(zA)}{h} &= \exp(zA)\Big(\frac{\exp(hA) - I}{h}\Big) \\ &= \exp(zA)\Big(A + \frac{hA^2}{2!} + \frac{h^2A^3}{3!} + \dots\Big) \\ &\to \exp(zA)A. \end{aligned}$$

□

2.6 Exponential of a Diagonable Matrix

Lemma 2.32 *Suppose that A has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Then there exists an invertible $n \times n$ matrix S such that*

$$\exp(tA) = S \begin{bmatrix} e^{t\lambda_1} & 0 & 0 & \dots \\ 0 & e^{t\lambda_2} & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & e^{t\lambda_n} \end{bmatrix} S^{-1}. \tag{2.56}$$

Proof Introduce the matrix $S = [X_1\, X_2 \dots X_n]$ with columns given by the eigenvectors of A, hence the X_j are linearly independent and S has rank n. Hence S is an invertible $n \times n$ matrix such that $A = SDS^{-1}$ where D is diagonal with entries $\lambda_1, \dots, \lambda_n$. Hence $\exp(tA) = S \exp(tD) S^{-1}$. □

Exponentials of Diagonable Matrices

[**Proposition 2.33**] *Suppose that A has distinct eigenvalues λ_j such that $\Re\lambda_j \le \kappa$ for all $j = 1, \dots, n$, all the eigenvalues lie in the closed left half-plane $\{\lambda : \Re\lambda \le \kappa\}$ which consists of the points in the complex plane that lie on or to the left of the vertical line $\{\lambda : \Re\lambda = \kappa\}$.*

1. *[(i)] Then the general solution of $\frac{dX}{dt} = AX$ is $X = \sum_{j=1}^{n} a_j e^{\lambda_j t} X_j$, where X_j is an eigenvector corresponding to λ_j and $a_j \in \mathbb{C}$ are arbitrary.*
2. *[(ii)] There exists M such that $\|\exp(tA)\| \le Me^{\kappa t} \qquad (t \ge 0)$.*
3. *[(iii)] In particular, suppose that $\Re\lambda_j \le 0$ for all $j = 1, \dots, n$. Then there exists M such that $\|\exp(tA)\| \le M \qquad (t \ge 0)$.*

Proof

(i) Checking the solution: For arbitrary $X_0 \in \mathbb{C}^n$, we observe that $X(t) = \exp(tA)X_0$ satisfies $dX(t)/dt = AX(t)$ and $X(0) = X_0$. We can write $X_0 = a_1X_1 + \dots + a_nX_n$ for some $a_j \in \mathbb{C}$ since $\{X_1, X_2, \dots, X_n\}$ is a basis for $\mathbb{C}^n$. Also, $\exp(tA)X_j = e^{\lambda_j t}X_j$, so $X(t) = \sum_{j=1}^{n} a_j e^{\lambda_j t} X_j$ is the general solution.

(ii) Checking the bounds: Consider

$$\begin{bmatrix} e^{t\lambda_1} & 0 & \dots & \dots \\ 0 & e^{t\lambda_2} & 0 & \dots \\ 0 & \ddots & \dots & 0 \\ 0 & \dots & \dots & e^{t\lambda_n} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} e^{t\lambda_1} z_1 \\ e^{t\lambda_2} z_2 \\ \vdots \\ e^{t\lambda_n} z_n \end{bmatrix} \tag{2.57}$$

where $|e^{t\lambda_j}| = e^{t\Re\lambda_j} \le e^{t\kappa}$ for all $t \ge 0$, hence

$$\sum_{j=1}^{n} |e^{t\lambda_j} z_j|^2 \le e^{2t\kappa} \sum_{j=1}^{n} |z_j|^2 \tag{2.58}$$

so $\| \exp(tD)z\| \le e^{t\kappa}\|z\|$; hence

$$\begin{aligned} \| \exp(tA)\| &\le \|S\| \left\| \begin{bmatrix} e^{t\lambda_1} & 0 & \dots \\ 0 & \ddots & 0 \\ 0 & \dots & e^{t\lambda_n} \end{bmatrix} \right\| \|S^{-1}\| \\ &\le \|S\|\|S^{-1}\| \max_{j=1,\dots,n} |e^{t\lambda_j}| \le \|S\|\|S^{-1}\|e^{t\kappa}. \end{aligned} \tag{2.59}$$

□

2.7 Solving MIMO (A, B, C, D)

Definition 2.34 (SISO) Let A, B, C, D be constant complex matrices with shapes:

$$A \quad (n \times n); \quad B \quad (n \times 1); \quad C \quad (1 \times n); \quad D \quad (1 \times 1). \tag{2.60}$$

Then the continuous time linear system with one input, n states and one output is

$$\begin{aligned} \frac{dX}{dt} &= AX + Bu \\ y &= CX + Du \end{aligned} \tag{2.61}$$

Here t is time, u is the input, X is the state, and y is the output. We call the system single-input single-output or **SISO**.

Example 2.35

(i) An electrical fan is a SISO system. The input is electricity, and the output is moving air. The states can involve the speed of rotation, voltage, current and so on.
(ii) A wind turbine has input moving air and output electricity.

Given $B \in \mathbb{C}^{n\times 1}$ and $C \in \mathbb{C}^{1\times n}$, we can build various linear operations.

(i) CB is simply a complex number;
(ii) $b \mapsto Bb$ gives a linear map $\mathbb{C} \to \mathbb{C}^{n\times 1}$;
(iii) $Y \mapsto CY$ gives a linear map $\mathbb{C}^{n\times 1} \to \mathbb{C}$;

(iv) $Y \mapsto BCY$ gives a linear map $\mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$, determined by the matrix $BC \in M_{n\times n}(\mathbb{C})$ which has rank one if $B, C \neq 0$.

Definition 2.36 (MIMO (A,B,C,D)) Let A, B, C, D be constant complex matrices with shapes: A $(n \times n)$; B $(n \times k)$; C $(m \times n)$; D $(m \times k)$. Then the continuous-time linear system with k inputs, n states and m outputs is

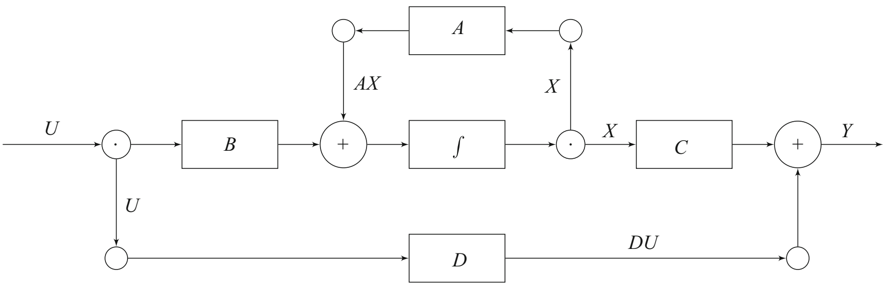
$$\frac{dX}{dt} = AX + BU,$$
$$Y = CX + DU. \tag{2.62}$$

Here $U \in \mathbb{C}^{k\times 1}$ is the input, $X \in \mathbb{C}^{n\times 1}$ is the state, and $Y \in \mathbb{C}^{m\times 1}$ is the output, which all depend upon t. The matrices are:

- A=state matrix (main transformation);
- B= input matrix (input transformation);
- C= output matrix;
- D=straight through matrix (external transformation).

This data gives the multiple-input multiple-output system (A, B, C, D), called MIMO.

The following diagram gives the standard form of MIMO, which is the main object of study in this book.



Describing MIMO (A,B,C,D)
The system (A, B, C, D) is a state model and the variables are in the time domain, in the sense that they are functions of t.

1. [(i)] If $k = 1$, then we say the system is single input; if $m = 1$, then the system is single output. If $k = m = 1$, then the system is **SISO**.
2. [(ii)] If $k > 1$, then we say the system is multi input; if $m > 1$ then the system is multi output.

3. [(iii)] If $k \geq 1$ and $m \geq 1$, then we call the system **MIMO**. (SISO is a special case of MIMO)

We write MIMO as a rectangular $(n + k) \times (n + m)$ block array with sizes

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad \begin{bmatrix} n \times n & n \times k \\ m \times n & m \times k \end{bmatrix}. \tag{2.63}$$

Example 2.37 (Domestic MIMO)

(i) A washing machine is a MIMO system. The inputs are cold water, soap powder, and electricity; whereas the outputs are hot soapy water, cold rinsing water and hot air. The state of the washing machine can be complicated, and relate to the rotation of the drum, temperature of various components, washing cycles and so on.
(ii) A domestic central heating system is a MIMO system with inputs gas, cold air, electricity and cold water; whereas the outputs are hot air and hot water.
(iii) Mobile telephone networks are MIMO systems. There are multiple transmitting antennas, and multiple receiving antennas, as we discuss in Proposition 11.14 about single user MIMO. Massive MIMO is a further example to model 5G wireless transmission. In a given district, there may be 64 receivers, and 64 transmitters; so we need a matrix A of size 64×64, and signals may be split up into many component parts. This explains the terminology 'massive'.

Example 2.38 (MIMO Transposed) Let A, B, C, D be complex matrices with shapes A $(n \times n)$; B $(n \times k)$; C $(m \times n)$; D $(m \times k)$ Recall

$$(m \times n) \times (n \times n) \times (n \times k) = (m \times k) \tag{2.64}$$

Then $(A^\top, C^\top, B^\top, D^\top)$ also gives a linear system

$$\begin{bmatrix} A^\top & C^\top \\ B^\top & D^\top \end{bmatrix} \qquad \begin{bmatrix} n \times n & n \times m \\ k \times n & k \times m \end{bmatrix}$$

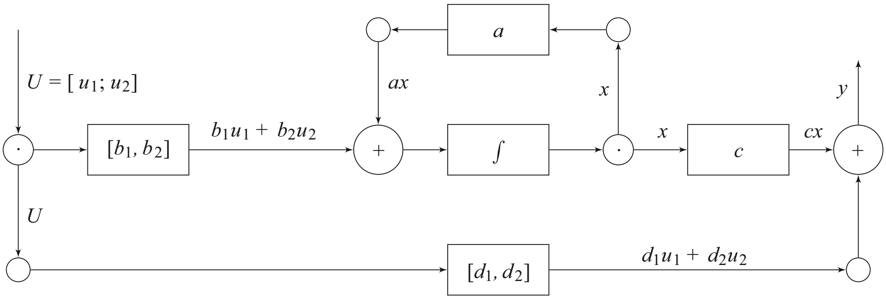
where $\det(sI - A^\top) = \det(sI - A)$. The properties of the system and its transpose are thus closely related, and sometimes one can easily obtain properties of one from the other. We will exploit this idea in our discussion of controllability and observability in section 3.12.

Example 2.39 The following system has two inputs, one state variable and one output

$$\frac{dx}{dt} = ax + b_1 u_1 + b_2 u_2$$

$$y = cx + d_1 u_1 + d_2 u_2$$

and can be represented the diagram below.



There are three issues involved in solving any differential equation:

- existence, that is, showing there is some solution;
- uniqueness, that is, showing that there is at most one solution;
- finding a useful expression for the solution.

The following theorem achieves all of these and is the fundamental result for solving (A, B, C, D) systems throughout this book.

Theorem 2.40 (Solution of Basic ODE) *Suppose that A is a constant $(n \times n)$ matrix and that $BU(t)$ is a $(n \times k)$ matrix with continuous functions $[0, \infty) \to \mathbb{C}$ as entries. Then for any constant $(n \times k)$ complex matrix X_0, the $(n \times k)$ matrix function*

$$X(t) = \exp(tA)X_0 + \int_0^t \exp((t-s)A)BU(s)\,ds \tag{2.65}$$

satisfies the matrix differential equation

$$\frac{dX}{dt} = AX + BU \tag{2.66}$$

with initial value

$$X(0) = X_0.$$

Proof Uniqueness: We suppose that a solution exists, and find a formula for the solution. Write the ODE in the standard form of a first order linear ODE

$$\frac{dX}{dt} - AX = BU \tag{2.67}$$

so the integrating factor is $\exp(-tA)$, and

$$\exp(-tA)\frac{dX}{dt} - \exp(-tA)AX = \exp(-tA)BU$$

$$\frac{d}{dt}\big(\exp(-tA)X\big) = \exp(-tA)BU,$$

so we integrate to get the unique solution of this differential equation

$$\big[\exp(-wA)X\big]_0^t = \int_0^t \exp(-wA)BU(w)dw$$

$$\exp(-tA)X(t) - \exp(0)X(0) = \int_0^t \exp(-wA)BU(w)dw$$

so we can solve for X and obtain

$$\begin{aligned} X(t) - \exp(tA)X_0 &= \int_0^t \exp((t-w)A)BU(w)dw \\ &= \exp(tA)\int_0^t \exp(-wA)BU(w)dw. \end{aligned}$$

Existence: To check the proposed solution works, let $t = 0$ to get $X(0) = X_0$. Then we apply standard results of calculus, and work on one entry of the matrix at a time. So by the fundamental theorem of calculus,

$$X(t) = \exp(tA)X_0 + \exp(tA)\int_0^t \exp(-wA)BU(w)dw \tag{2.68}$$

is a differentiable function of t, with derivative

$$\begin{aligned} \frac{d}{dt}X &= A\exp(tA)X_0 + \exp(0)BU(t) + A\exp(tA)\int_0^t \exp(-wA)BU(w)\,dw \\ &= AX + BU. \end{aligned}$$

□

Corollary 2.41 (Solution of MIMO) *For any initial condition X_0 and any continuous input U, the solution of (A,B,C,D) is*

$$Y(t) = C\exp(tA)X_0 + \int_0^t C\exp((t-v)A)BU(v)\,dv + DU(t). \tag{2.69}$$

Proof From Theorem 2.40, we take

$$X(t) = \exp(tA)X_0 + \int_0^t \exp((t-v)A)BU(v)\,dv \tag{2.70}$$

and then

$$Y(t) = CX(t) + DU(t). \tag{2.71}$$

□

Terminology Concerning Solutions
The terminology of Differential Equations reappears in linear systems. Consider the inhomogeneous differential equation

$$\frac{dX}{dt} = AX + BU(t). \tag{2.72}$$

1. [(i)] Let $T_t = \exp(tA)$, which satisfies $T_t(\lambda X_0 + \mu Y_0) = \lambda T_t X_0 + \mu T_t Y_0$, also $T_{t+v} = T_t T_v$ and

$$\frac{T_t - I}{t} \to A \qquad (t \to 0+). \tag{2.73}$$

2. [(ii)] The expression $X(t) = T_t X_0$ with X_0 arbitrary is known as the complementary function, since it satisfies the homogeneous differential equation

$$\frac{dX}{dt} = AX$$
$$X(0) = X_0.$$

3. [(iii)] The term $X_t = \int_0^t \exp((t-v)A)BU(v)\,dv$ is a particular integral of the inhomogeneous differential equation

$$\frac{dX}{dt} = AX + BU$$
$$X(0) = 0.$$

4. [(iv)] The general solution of the inhomogeneous equation is the complementary function plus a particular integral, so

$$X_t = T_t X_0 + \int_0^t \exp((t-v)A)BU(v)\,dv \tag{2.74}$$

satisfies

$$\frac{dX}{dt} = AX + BU$$
$$X(0) = X_0.$$

5. [(v)] We consider the unit impulse $U(v) = \delta_0(v_0)$ in (iii), and observe that

$$X_t = 0 \qquad (t < v_0)$$
$$= T_{t-v_0}B \qquad (t > v_0)$$

gives a solution of

$$\frac{dX}{dt} = AX + BU \qquad (t > 0)$$
$$X(0) = 0.$$

This initial value problem must be interpreted with great care, since the differential equation now involves a measure.

Example 2.42 (Damped Harmonic Oscillator) Suppose that $\beta, \gamma > 0$. Then the damped harmonic oscillator is

$$\frac{d^2x}{dt^2} + \beta\frac{dx}{dt} + \gamma x = u$$
$$x(0) = x_0, \quad \frac{dx}{dt}(0) = v_0. \tag{2.75}$$

Here we regard the variable x as the output, and we aim to find x for a given input u. We can regard this as the linear system specified by

$$A = \begin{bmatrix} 0 & 1 \\ -\gamma & -\beta \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \end{bmatrix}, D = 0. \tag{2.76}$$

Then

$$(sI - A)^{-1} = \begin{bmatrix} s & -1 \\ \gamma & s+\beta \end{bmatrix}^{-1} = \frac{1}{s^2 + \beta s + \gamma}\begin{bmatrix} s+\beta & 1 \\ -\gamma & s \end{bmatrix}. \tag{2.77}$$

In later discussion, we use the transfer matrix (2.93), which is defined by

$$T(s) = D + C(sI - A)^{-1}B = \frac{1}{s^2 + \beta s + \gamma}. \tag{2.78}$$

First suppose that $\beta^2 - 4\gamma \neq 0$. Then we have eigenvalues for A at the roots of $s^2 + \beta s + \gamma = 0$, namely

$$\lambda_{\pm} = \frac{-\beta \pm \sqrt{\beta^2 - 4\gamma}}{2}, \tag{2.79}$$

which are distinct since $\beta^2 - 4\gamma \neq 0$. The corresponding eigenvalues are

$$\lambda_+ : \qquad \begin{bmatrix} 1 \\ \lambda_+ \end{bmatrix}, \qquad \lambda_- : \qquad \begin{bmatrix} 1 \\ \lambda_- \end{bmatrix}, \tag{2.80}$$

so we introduce the invertible matrix

$$S = \begin{bmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{bmatrix}, \tag{2.81}$$

so that

$$A = S \begin{bmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{bmatrix} S^{-1}. \tag{2.82}$$

Hence we have

$$\begin{aligned}
\exp(tA) &= S \begin{bmatrix} e^{\lambda_+ t} & 0 \\ 0 & e^{\lambda_- t} \end{bmatrix} S^{-1} \\
&= \begin{bmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{bmatrix} \begin{bmatrix} e^{\lambda_+ t} & 0 \\ 0 & e^{\lambda_- t} \end{bmatrix} (\lambda_- - \lambda_+)^{-1} \begin{bmatrix} \lambda_- & -1 \\ -\lambda_+ & 1 \end{bmatrix} \\
&= (\lambda_- - \lambda_+)^{-1} \begin{bmatrix} \lambda_- e^{\lambda_+ t} - \lambda_+ e^{\lambda_- t} & e^{\lambda_- t} - e^{\lambda_+ t} \\ \lambda_- \lambda_+ (e^{\lambda_+ t} - e^{\lambda_- t}) & \lambda_- e^{\lambda_- t} - \lambda_+ e^{\lambda_+ t} \end{bmatrix}
\end{aligned}$$

so

$$\begin{aligned}
x(t) &= C \exp(tA) \begin{bmatrix} x_0 \\ v_0 \end{bmatrix} + \int_0^t C \exp\big((t-\tau)A\big) B u(\tau) d\tau \\
&= \frac{\lambda_- e^{\lambda_+ t} - \lambda_+ e^{\lambda_- t}}{\lambda_- - \lambda_+} x_0 + \frac{e^{\lambda_- t} - e^{\lambda_+ t}}{\lambda_- - \lambda_+} v_0 + \int_0^t \frac{e^{\lambda_-(t-\tau)} - e^{\lambda_+(t-\tau)}}{\lambda_- - \lambda_+} u(\tau) d\tau.
\end{aligned}$$

Now suppose that $\beta^2 = 4\gamma$. Then we have an eigenvalue $\lambda = -\beta/2$ with algebraic multiplicity two, so we introduce an eigenvector V such that $AV = \lambda V$ and W such that $AW - \lambda W = V$; then with $S = [V, W]$, we have

$$A = \begin{bmatrix} 0 & 1 \\ -\lambda^2 & 2\lambda \end{bmatrix}, S = \begin{bmatrix} 1 & -1/\lambda \\ \lambda & 0 \end{bmatrix}, S^{-1} = \begin{bmatrix} 0 & 1/\lambda \\ -\lambda & 1 \end{bmatrix}, \tag{2.83}$$

so that

$$S^{-1}AS = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \tag{2.84}$$

has the form of a Jordan block. From the exponential series, we have

$$\exp\left(t \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}\right) = e^{\lambda t} \exp\left(\begin{bmatrix} 0 & t \\ 0 & 0 \end{bmatrix}\right) = e^{\lambda t} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \tag{2.85}$$

which gives

$$\exp(tA) = S \begin{bmatrix} e^{\lambda t} & te^{\lambda t} \\ 0 & e^{\lambda t} \end{bmatrix} S^{-1} = \begin{bmatrix} (1-\lambda t)e^{\lambda t} & te^{\lambda t} \\ -\lambda^2 te^{\lambda t} & (\lambda t + 1)e^{\lambda t} \end{bmatrix}. \tag{2.86}$$

Hence we obtain

$$\begin{aligned} x(t) &= C\exp(tA)\begin{bmatrix} x_0 \\ v_0 \end{bmatrix} + \int_0^t C\exp\big((t-\tau)A\big)Bu(\tau)d\tau \\ &= (1-\lambda t)e^{\lambda t}x_0 + te^{\lambda t}v_0 + \int_0^t (t-\tau)e^{\lambda(t-\tau)}u(\tau)\,d\tau. \end{aligned}$$

2.8 Rational Functions

In this section we summarize some terminology about complex rational functions that we use repeatedly later on; see [6] page 55. Let s be an algebraic indeterminate (variable), let $\mathbb{C}[s]$ the space of complex polynomials in s. Let $g(s)$ and $h(s)$ be complex polynomials, with $h(s)$ not the zero polynomial. Then

$$f(s) = \frac{g(s)}{h(s)}$$

is said to be a rational function. The set of all complex rational functions in s is denoted $\mathbb{C}(s)$, with the usual operations of multiplication, addition, division and differentiation.

Definition 2.43

(i) If the degree of $g(s)$ is less than or equal to the degree of $h(s)$, then $f(s)$ is said to be **proper** rational. If the degree of $g(s)$ is strictly less than the degree of $h(s)$, then $f(s)$ is said to be strictly proper. We write $\mathbb{C}(s)_p$ for the proper rational functions and $\mathbb{C}(s)_0$ for the strictly proper rational functions.

(ii) For a nonzero rational function $f = g/h$, one can define

$$\deg f = \deg g - \deg h. \tag{2.87}$$

Then f is proper if and only if $\deg f \leq 0$. A rational function is strictly proper if and only if $\deg f < 0$.

(iii) A zero of g is $s_0 \in \mathbb{C}$ such that $g(s_0) = 0$; this is otherwise called a root of $g(s) = 0$.

Suppose that $g(s)$ and $h(s)$ have no common factors other than constants. Then zeros of $g(s)$ give zeros of $f(s)$; while zeros of $h(s)$ give **poles** of $f(s)$. One can feed a rational function into MATLAB by way of the coefficients. For example

$$f(s) = \frac{g(s)}{h(s)} = \frac{2s^3 - is^2 + 6s + 5}{3s^4 + 7s^2 - 4s + 3} \tag{2.88}$$

is a strictly proper rational function, which can be entered into MATLAB code as

$$>> \quad f = tf([2\ -i\ 6\ 5], [3\ 0\ 7\ -4\ 3])$$

with numerator before denominator and starting with the leading coefficients; there is no need for commas. The abbreviation tf is for transfer function, specifically a continuous-time transfer function of the type we consider in the first seven chapters. One can then find numerical values for the zeros and poles via

$$>> \quad zero(f)$$

$$>> \quad pole(f)$$

MATLAB diagrams indicate poles with crosses $\times$ and zeros with small circles $\circ$.

MATLAB operations sometimes do not work properly when there are complex coefficients.

To give zeros and poles equivalent status, it is convenient to work with the Riemann sphere $\mathbb{C} \cup \{\infty\}$. Then one can regard $g(s)$ as having zeros at $z_1, \dots, z_n$ and a pole of order n at ∞. This allows us to keep track of zeros and poles when we make invertible rational changes of variable such as $s = (z-1)/(z+1)$, which takes $\infty \mapsto 1$.

Definition 2.44

(i) A rational function is stable if it is proper and all the poles are in $LHP = \{s : \Re s < 0\}$. We write $RHP = \{s \in \mathbb{C} : \Re s > 0\}$ for the right half plane.

(ii) The notation $F(s) = O(1/s^k)$ as $s \to \infty$ means that there exist $r, M > 0$ such that $|F(s)| \leq M|s|^k$ for all $s \in \mathbb{C}$ such that $|s| \geq r$.

Example 2.45

(i) The rational function $1/(1+s)$ is stable. The importance of such functions in linear systems will be considered in Chaps. 5 and 6.
(ii) We have $(s+1)/(s-2)^3 = O(1/s^2)$ as $s \to \infty$. One can show that a rational function is proper if $F(s) = O(1)$ as $s \to \infty$, and strictly proper if $F(s) = O(1/s)$ as $s \to \infty$.

2.9 Block Matrices

Let S, U, V and W be complex vector spaces. Then we can form the direct sum

$$S \oplus V = \begin{matrix} S \\ V \end{matrix} = \left\{ \begin{bmatrix} u \\ v \end{bmatrix} : u \in S, v \in V \right\} \tag{2.89}$$

with the operations

$$\lambda \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \lambda u \\ \lambda v \end{bmatrix}, \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} + \begin{bmatrix} u_2 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + u_2 \\ v_1 + v_2 \end{bmatrix}, \tag{2.90}$$

and inner product

$$\left\langle \begin{bmatrix} u_1 \\ v_1 \end{bmatrix}, \begin{bmatrix} u_2 \\ v_2 \end{bmatrix} \right\rangle = \langle u_1, u_2 \rangle + \langle v_1, v_2 \rangle \qquad (u_1, u_2 \in S; v_1, v_2 \in V, \lambda \in \mathbb{C}). \tag{2.91}$$

We write $\mathcal{L}(U, S)$ for the space of linear transformations $B : U \to S$. Then for $A \in \mathcal{L}(S, S)$, $B \in \mathcal{L}(U, S)$, $C \in \mathcal{L}(S, W)$ and $D \in \mathcal{L}(U, W)$ we form the linear transformation

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} : \begin{matrix} S \\ U \end{matrix} \to \begin{matrix} S \\ W \end{matrix} : \qquad \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} Au + Bv \\ Cu + Dv \end{bmatrix} \tag{2.92}$$

known as a block matrix or block transformation. Block matrices of appropriate sizes matrices can be added and multiplied. The MATLAB command for the matrix in (2.92) is $[A, B; C, D]$.

In the context of linear systems, U is called the input space, S is the state space and W the output space.

2.10 The Transfer Function of (A, B, C, D)

Definition 2.46 (Transfer Function) The transfer matrix function of MIMO system (A, B, C, D) is

$$T(s) = D + C(sI - A)^{-1}B. \tag{2.93}$$

Lemma 2.47 *The transfer function may be found by exact arithmetic over* $\mathbb{C}(s)$ *by elementary row operation.*

Proof We show that if $sI - A$ is invertible, then the following block matrices are row equivalent over $\mathbb{C}(s)$:

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \cong \begin{bmatrix} I & (A - sI)^{-1}B \\ 0 & T(s) \end{bmatrix}. \tag{2.94}$$

(i) Suppose that $sI - A$ is invertible. Use multiplication on the left to show that the matrices

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} I & (A - sI)^{-1}B \\ 0 & D + C(sI - A)^{-1}B \end{bmatrix} \tag{2.95}$$

are row equivalent. We multiply on the left by $\begin{bmatrix} (A - sI)^{-1} & 0 \\ 0 & I \end{bmatrix}$, obtaining

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \cong \begin{bmatrix} I & (A - sI)^{-1}B \\ C & D \end{bmatrix}; \tag{2.96}$$

now we multiply on the left by $\begin{bmatrix} I & 0 \\ -C & I \end{bmatrix}$, obtaining

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \cong \begin{bmatrix} I & (A - sI)^{-1}B \\ 0 & D - C(A - sI)^{-1}B \end{bmatrix}; \tag{2.97}$$

hence the result. We observe that $T(s)$ can thus be computed exactly by matrix multiplication and elementary row operations, which involve rational arithmetic.

□

We follow this with some determinant calculations. Note that

$$\det \begin{bmatrix} I & 0 \\ -C & I \end{bmatrix} = 1 \tag{2.98}$$

by Proposition 2.7 since the matrix is triangular with ones on the diagonal. Also, by the row reductions in the proof of the Lemma 2.47, we have

$$\det\begin{bmatrix}(A-sI)^{-1} & 0\\ 0 & I\end{bmatrix}\det\begin{bmatrix}A-sI & B\\ C & D\end{bmatrix}=\det\begin{bmatrix}I & (A-sI)^{-1}B\\ 0 & D-C(A-sI)^{-1}B\end{bmatrix}, \tag{2.99}$$

so

$$\det(A-sI)^{-1}\det\begin{bmatrix}A-sI & B\\ C & D\end{bmatrix}=\det(D+C(sI-A)^{-1}B). \tag{2.100}$$

When D is a 1×1 matrix, as in a SISO, we can reduce this formula to a determinant formula for the transfer function

$$\det(A-sI)^{-1}\det\begin{bmatrix}A-sI & B\\ C & D\end{bmatrix}=D+C(sI-A)^{-1}B. \tag{2.101}$$

Proposition 2.48 (Transfer Function) *The transfer function $T(s)=D+C(sI-A)^{-1}B$ of a SISO system is a proper rational function, and all the poles are eigenvalues of A.*

Proof The characteristic polynomial $\det(sI-A)$ has degree n, and leading term s^n. A cofactor of $sI-A$ is the determinant of a $(n-1)\times(n-1)$ submatrix of $sI-A$ and hence is a polynomial of degree less than or equal to $n-1$. Now

$$(sI-A)^{-1}=\det(sI-A)^{-1}\mathrm{adj}(sI-A) \tag{2.102}$$

where $\mathrm{adj}(sI-A)$ is the transpose of the matrix of cofactors. Hence the entries of $(sI-A)^{-1}$ are strictly proper rational functions. The eigenvalues of A are precisely the zeros of $\det(sI-A)$, hence are the only possible poles of entries of $(sI-A)^{-1}$. Since C, B and D are constant matrices, they do not introduce any more factors involving s, so $T(s)$ is a proper rational function. It is not asserted that all eigenvalues of A lead to poles of $T(s)$, since there may be cancellation. □

Corollary 2.49 (The Transfer Function of MIMO (A,B,C,D)) *The transfer function of a MIMO is a $(m\times k)$ matrix of proper rational functions, and all the poles are eigenvalues of A.*

Proof This follows from the proof of Proposition 2.48. □

In Sect. 6.12 we use invariant factors to refine this result, and give a condition under which we can cancel out some of the poles, and reduce the denominator in the transfer function.

Remark 2.50 Let V_1 and V_2 be complex vector spaces. A map $L: V_1\to V_2$ is said to be affine if

$$L(\lambda X+(1-\lambda Y))=\lambda L(x)+(1-\lambda)L(Y) \qquad (X,Y\in V_1)$$

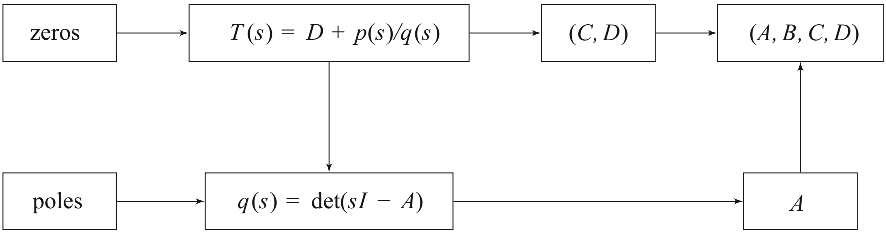
and $\lambda \in \mathbb{C}$. Let (A, B, C, D) be a linear system with transfer function $T(s) = D + C(sI - A)^{-1}B$. Then the following are affine maps

(i) $M_{m\times k}(\mathbb{C}) \to M_{m\times k}(\mathbb{C}(s)) : D \mapsto T(s)$;
(ii) $M_{n\times k}(\mathbb{C}) \to M_{m\times k}(\mathbb{C}(s)) : B \mapsto T(s)$;
(iii) $M_{m\times n}(\mathbb{C}) \to M_{m\times k}(\mathbb{C}(s)) : C \mapsto T(s)$.

We cannot make any such simple statement about $A \mapsto T(s)$, since $(sI - A)^{-1} = \operatorname{adj}(sI - A)/\det(sI - A)$ depends upon the entries of A in a complicated manner.

2.11 Realization with a SISO

Next we consider the converse of the Proposition 2.48. Realization means devising a linear system with a given transfer function; we think of this as building a gadget with desired effect. The arrows in the diagram point from data in the source box to data in the destination box, indicating that such a choice is possible.



Proposition 2.51 *The general strictly proper rational function*

$$T(s) = \frac{\sum_{j=0}^{n-1} \gamma_j s^j}{s^n + \sum_{j=0}^{n-1} \alpha_j s^j} \tag{2.103}$$

is the transfer function of the SISO (A, B, C, D)*, where*

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ -\alpha_0 & -\alpha_1 & \dots & \dots & -\alpha_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \tag{2.104}$$

$$C = \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{n-1} \end{bmatrix}, \quad D = 0. \tag{2.105}$$

Proof We require to prove $T(s) = C(sI - A)^{-1}B$. Recall that

$$(sI - A)^{-1} = \det(sI - A)^{-1}\mathrm{adj}(sI - A) \tag{2.106}$$

where the adjugate is the transpose of the matrix of cofactors. The coefficients α_j appear in the denominator, but not in the numerator. Also $\mathrm{adj}(sI - A)B$ equals the last column of $\mathrm{adj}(sI - A)$, so by transposition, $\mathrm{adj}(sI - A)B$ equals the final row of the matrix of cofactors of $sI - A$, where

$$sI - A = \begin{bmatrix} s & -1 & 0 & \dots & 0 \\ 0 & s & -1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & s & -1 \\ \alpha_0 & \alpha_1 & \dots & \alpha_{n-2} & s + \alpha_{n-1} \end{bmatrix} \tag{2.107}$$

We compute these one after another, and find that the α_j do not appear in these cofactors.

Cofactors: Recall that the determinant of an upper or lower triangular matrix equals the product of the diagonal entries. The cofactor of the entry α_0 in place $(n, 1)$ is

$$(-1)^{n-1} \det \begin{bmatrix} -1 & 0 & \dots & 0 \\ s & -1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & s & -1 \end{bmatrix} = 1$$

The cofactor of the entry α_1 in place $(n, 2)$ is

$$(-1)^{n-2} \det \begin{bmatrix} s & 0 & \dots & 0 \\ 0 & -1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -1 \end{bmatrix} = s \tag{2.108}$$

and so until the cofactor of the entry $s + \alpha_{n-1}$ in place (n, n) is

$$\det \begin{bmatrix} s & -1 & \dots & 0 \\ 0 & s & -1 & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & s \end{bmatrix} = s^{n-1} \tag{2.109}$$

so

$$C\mathrm{adj}(sI - A)B = \begin{bmatrix}\gamma_0 & \gamma_1 & \dots & \gamma_{n-1}\end{bmatrix} \begin{bmatrix} 1 \\ s \\ \vdots \\ s^{n-2} \\ s^{n-1} \end{bmatrix}$$

$$= \gamma_0 + \gamma_1 s + \cdots + \gamma_{n-1} s^{n-1}. \tag{2.110}$$

Note that A is a companion matrix, so by Proposition 2.12,

$$\det(sI - A) = s^n + \alpha_{n-1} s^{n-1} + \cdots + \alpha_1 s + \alpha_0. \tag{2.111}$$

Hence

$$T(s) = \frac{C\mathrm{adj}(sI - A)B}{\det(sI - A)}$$

$$= \frac{\gamma_0 + \gamma_1 s + \cdots + \gamma_{n-1} s^{n-1}}{s^n + \alpha_{n-1} s^{n-1} + \cdots + \alpha_0}. \tag{2.112}$$

□

Proposition 2.51 is an existence theorem, not a uniqueness theorem about the choice of $(A, B, C, 0)$. In the next chapter we give a slight extension of this result, Proposition 3.15, which applies to stable rational functions and includes a determinant formula for $T(s)$.

Example 2.52 To realize

$$T(s) = \frac{2s^2 + 3s + 1}{s^3 + 6s^2 + 8s - 2} \tag{2.113}$$

as the transfer function of a SISO.

MATLAB calls this a continuous-time transfer function, and one can introduce this example as

$$>> \qquad T = tf([2\ 3\ 1], [1\ 6\ 8\ {-2}])$$

As in Proposition 2.51, we introduce

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -8 & -6 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}, D = 0. \tag{2.114}$$

which can be realized as a feedback linear system without differentiators. In MATLAB, or similar, one can check that $T(s) = C(sI - A)^{-1}B$.
Checking the solution: Here

$$sI - A = \begin{bmatrix} s & -1 & 0 \\ 0 & s & -1 \\ -2 & 8 & s+6 \end{bmatrix}; \tag{2.115}$$

computing only the relevant entries, we have

$$\operatorname{adj}(sI - A)^{\top} = \begin{bmatrix} * & * & * \\ * & * & * \\ 1 & s & s^2 \end{bmatrix} \tag{2.116}$$

$$\begin{aligned} C\operatorname{adj}(sI - A)B &= \begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} * & * & 1 \\ * & * & s \\ * & * & s^2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= 1 + 3s + 2s^2. \end{aligned} \tag{2.117}$$

Since A is a companion matrix, we have

$$\det(sI - A) = s^3 + 6s^2 + 8s - 2 \tag{2.118}$$

so

$$T(s) = \frac{C\operatorname{adj}(sI - A)B}{\det(sI - A)} = \frac{1 + 3s + 2s^2}{s^3 + 6s^2 + 8s - 2}, \tag{2.119}$$

as required.

MIMO as a Feedback System, Without Differentiators
Differentiators are sometimes regarded as bad components to have in a linear system since they can introduce noise. Integrators are preferable, since they depend upon the long term history of system. So it is advantageous to produce linear systems without differentiators.

Proposition 2.53 *The MIMO system* (A, B, C, D) *can be realized as a feedback system involving taps, matrix amplifiers, summing junctions and integrators, but no differentiators.*

Proof We write

$$\frac{d}{dt}X = AX + BU$$

in the form

$$X = A \int X + B \int U$$

and realize this as a feedback with $\int$ and the matrix amplifiers A and B.

Then we take U and X as inputs into the system

$$Y = CX + DU. \tag{2.120}$$

□

Remark 2.54 If in MIMO we replace the integrator $\int$ by multiplication by $1/s$ we obtain

$$\begin{aligned} \hat{X} &= (1/s)A\hat{X} + (1/s)B\hat{U} \\ \hat{Y} &= C\hat{X} + D\hat{U} \end{aligned} \tag{2.121}$$

with solution

$$\hat{Y} = (D + C(sI - A)^{-1}B)\hat{U}. \tag{2.122}$$

The theoretical justification of this is the Laplace transform, as in the proof of Theorem 4.21.

Example 2.55 Consider the second-order differential equation

$$\frac{d^2X}{dt^2} = AX + U \tag{2.123}$$

where $X, U \in \mathbb{C}^{n\times 1}$ and $A \in M_{n\times n}(\mathbb{C})$. This is equivalent to the first order differential equation

$$\frac{d}{dt}\begin{bmatrix} X \\ V \end{bmatrix} = \begin{bmatrix} 0 & I \\ A & 0 \end{bmatrix}\begin{bmatrix} X \\ V \end{bmatrix} + \begin{bmatrix} 0 \\ U \end{bmatrix}, \tag{2.124}$$

so we consider the MIMO linear system

$$\left(\begin{bmatrix} 0 & I \\ A & 0 \end{bmatrix}, \begin{bmatrix} 0 \\ I \end{bmatrix}, \begin{bmatrix} I & 0 \end{bmatrix}, 0\right), \tag{2.125}$$

where

$$\begin{bmatrix} sI & -I \\ -A & sI \end{bmatrix}^{-1} = \begin{bmatrix} s(s^2I - A)^{-1} & (s^2I - A)^{-1} \\ A(s^2I - A)^{-1} & s(s^2I - A)^{-1} \end{bmatrix},$$

so the transfer function is

$$T(s) = (s^2 I - A)^{-1}. \tag{2.126}$$

2.12 Exercises

Exercise 2.1 Let

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 6 & 2 & 1 \\ 1 & 7 & 8 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.127}$$

$$B = \begin{bmatrix} 1 & 2 \\ 7 & 4 \\ 5 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 2 & 7 \\ 1 & 4 & 3 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 5 \\ 6 & 7 \end{bmatrix}. \tag{2.128}$$

Compute the matrix transfer function

$$T(s) = C(sI - A)^{-1}B + D \tag{2.129}$$

either by hand or using suitable computer software. Here s is an algebraic variable (indeterminate).

Exercise 2.2 Show from the definition 2.30 that

$$\exp\left(t \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} \qquad (t \in \mathbb{R}). \tag{2.130}$$

This example show that the exponential of a real matrix can have negative entries. The conditions on A that ensure $\exp(tA)$ has nonnegative entries are discussed in section 9.9 and [48].

Exercise 2.3 Let A be a $n \times n$ complex matrix and let $T_t = \exp(tA)$ for $t \in \mathbb{R}$.

(i) Show that $T_t = T_t'$ for all $t \in \mathbb{R}$, if and only if $A = A'$.
(ii) Show that T_t is unitary, so $T_t' = T_t^{-1}$ if and only if $A' = -A$, so A is skew.

Exercise 2.4 Let A be a complex $(n \times n)$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Show that there exists an invertible matrix S such that

$$(sI - A)^{-1} = S \begin{bmatrix} \frac{1}{s-\lambda_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{s-\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \frac{1}{s-\lambda_n} \end{bmatrix} S^{-1} \tag{2.131}$$

for all $s \neq \lambda_1, \ldots, \lambda_n$.

Exercise 2.5

(i) Find a SISO system that has transfer function

$$T(s) = \frac{2s^2 - 3s + 4}{s^3 + 5s^2 + 6s + 7}. \tag{2.132}$$

(ii) Find approximate numerical values for the eigenvalues of A.

Exercise 2.6 Let A be the matrix

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a & b & 0 & 0 \\ c & d & 0 & 0 \end{bmatrix}. \tag{2.133}$$

(i) Find $\det A$.
(ii) Use reduction of determinants to find $\det(sI - A)$.

Exercise 2.7 Find a SISO system (A, B, C, D) that has transfer function

$$T(s) = \frac{2s^3 + s^2 - 5s + 1}{s^4 - 6s^3 + 5s^2 + 4s + 2}. \tag{2.134}$$

Exercise 2.8 Let (A, B, C, D) be a SISO with transfer function T. Show that $(A^\top, C^\top, B^\top, D)$ is also a SISO with transfer function T, where here $A^\top$ denotes the transpose of A.

Exercise 2.9 Let $a, b, d \in \mathbb{R}$ and consider

$$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}.$$

(i) Show that $\langle A\xi, \xi \rangle \geq 0$ for all $\xi \in \mathbb{R}^2$, if and only if $a, d, ad - b^2 \geq 0$.
(ii) Show that if the conditions of (i) hold, then

$$(\det A)^{1/2} \leq 2^{-1}\mathrm{trace}(A) \leq \|A\| \leq \mathrm{trace}(A).$$

Exercise 2.10 Find a SISO system (A, B, C, D) that has transfer function

$$T(s) = \frac{5s^4 + 7s^3 - 6s^2 + s + 2}{s^4 - 3s^3 + 4s^2 + 7s + 6}, \tag{2.135}$$

and find numerical values for the eigenvalues of A. Start by dividing numerator by denominator.

Exercise 2.11 Let A be a real (3×3) matrix.

(i) Show that $\det(sI - A)$ has either (a) three real zeros, or (b) one real root and a pair of complex conjugate zeros.
(ii) Show that, in both cases (a) and (b), A has a real eigenvector.

Exercise 2.12 Let

$$A = \begin{bmatrix} 1 & 4 & 10 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}. \tag{2.136}$$

Find $(sI - A)^{-1}$, where s is an algebraic variable.

Exercise 2.13 Let

$$A = \begin{bmatrix} 1 & -1 \\ 3 & 5 \end{bmatrix}; \tag{2.137}$$

find an invertible matrix S and a diagonal matrix D such that

$$A = SDS^{-1}. \tag{2.138}$$

Hence or otherwise find $\exp(tA)$, where t is a real variable.

Exercise 2.14 Find a matrix A such that $s^4 + 2s^3 + s^2 + 4s + 2$ is the characteristic polynomial of A. Then find the eigenvalues of A numerically.

Exercise 2.15

(i) Let A be the 3×3 complex matrix

$$A = \begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}. \tag{2.139}$$

Find A^2 and A^3, and deduce that the matrix $\exp(tA)$ has entries which are quadratic in t.

(ii) Let B be a strictly upper triangular 3×3 complex matrix

$$B = \begin{bmatrix} 0 & b_{1,2} & \dots & b_{1,n} \\ 0 & 0 & b_{2,3} & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \tag{2.140}$$

Show that $B^n = 0$, and deduce that $\exp(tB)$ has entries which are polynomials in t of degree $\leq n - 1$.

Exercise 2.16 (i) Find the eigenvalues and eigenvectors V of

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}. \tag{2.141}$$

(ii) By considering expressions of the form $Z(t) = e^{tz}V$, find the general solution to

$$\frac{dZ}{dt} = AZ. \tag{2.142}$$

(iii) By considering expressions of the form $Y(t) = e^{tw}V$, find the general solution to

$$\frac{d^2Y}{dt^2} = AY. \tag{2.143}$$

This is a model for three identical particles on a common circular track, connected by elastic springs.

(iv) State how many independent constants your solutions to (ii) and (iii) involve, and explain why this is the correct number in each case.

Exercise 2.17 (Cross Product) Some books on mechanics use the cross product $X \times U$ if $X, U \in \mathbb{R}^3$, as in vector calculus. This exercises expresses the cross product in matrix terms. Let

$$A = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad U = \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \quad X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$$

so that

$$A \times X = \begin{bmatrix} bz - cy \\ cx - az \\ ay - bx \end{bmatrix}.$$

We write

$$L_A = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \tag{2.144}$$

and similarly for the other vectors.

(i) Show that $L_A X = A \times X$ and $L_A L_U - L_U L_A = L_{A\times U}$.
(ii) Show that L_A has eigenvalues $0, \pm i\omega$ where $\omega^2 = a^2 + b^2 + c^2$.
(iii) Show that $\exp(L_A) = p(L_A)$ for some quadratic polynomial p.
(iv) By considering the eigenvalues, deduce that for $A \neq 0$,

$$\exp(L_A) = I_3 + \frac{\sin\omega}{\omega} L_A + \frac{1 - \cos\omega}{\omega^2} L_A^2. \tag{2.145}$$

Exercise 2.18 Let A be a $n \times n$ complex matrix with some of the following properties: (i) upper triangular; (ii) diagonal; (iii) real entries; (iv) nonnegative entries; (v) positive entries. In each case, (i)–(v), show that $\exp(A)$ also has this property.

Exercise 2.19 Let $B \in \mathbb{C}^{n\times 1}$ and $C \in \mathbb{C}^{1\times n}$, so that $BC \in M_{n\times n}(\mathbb{C})$ has rank one. Find $\det(I_n + \alpha BC)$, and find the inverse of $I_n + \alpha BC$ when it exists for $\alpha \in \mathbb{C}$.

Exercise 2.20 Let $A_1, A_2 \in M_{n\times n}(\mathbb{C})$. Show that

$$(sI_n - A_2)^{-1} = (sI_n - A_1)^{-1} + (sI_n - A_1)^{-1}(A_2 - A_1)(sI_n - A_2)^{-1}$$

for typical $s \in \mathbb{C}$ and deduce that

$$\operatorname{rank}\big((sI_n - A_2)^{-1} - (sI_n - A_1)^{-1}\big) = \operatorname{rank}(A_2 - A_1).$$

Exercise 2.21 (Spectral Factorization) Let $P(s)$ and $Q(s)$ be polynomials with real coefficients which are even so that $P(s) = P(-s)$ and $Q(s) = Q(-s)$ and suppose that P and Q have no zeros on $i\mathbb{R}$ and that the degree of $P(s)$ is less than or equal to the degree of $Q(s)$.

(i) Show that $F(s) = P(s)/Q(s)$ is even, so $F(s) = F(-s)$, $F(i\omega)$ is real for all $\omega \in \mathbb{R}$ and that $F(s)$ is proper. Let $a_1, \dots, a_n$ be the zeros of P that in the left half-plane, and $b_1, \dots, b_m$ be the zeros of Q that are in the left half-plane; then let

$$G(s) = \frac{\prod_{j=1}^{n}(s + a_j)}{\prod_{j=1}^{m}(s + b_j)}.$$

Show that $G(s)$ is stable and free from zeros in RHP, and

$$F(s) = CG(s)G(-s)$$

for some constant C.

Chapter 3
Eigenvalues and Block Decompositions of Matrices

In Chap. 2, we introduced the fundamental MIMO system (A, B, C, D) and solved it by the matrix exponential functions. For matrices A that are similar to diagonal matrices, we computed $\exp(tA)$. However, this does not address the typical case, and in this chapter we introduce Jordan decompositions to deal with multiple eigenvalues by splitting matrices into smaller blocks. The norm of $\exp(tA)$ is related to the position of the eigenvalues of A even when A is not similar to a diagonal matrix, as the results of this chapter show. We also consider positive definite matrices, which will turn out to be important in later chapters as an alternative method for controlling the size of matrix exponentials. We also look at ways of decomposing the state space. Chapters 4 and 5 can be read independently of Chap. 3, so readers mainly interested in differential equations can proceed to there.

3.1 The Transfer Function of Similar SISOs (A, B, C, D)

Lemma 3.1 *Let $\Sigma_1 = (A, B, C, D)$ be a linear system with transfer function $T(s)$, where A is a $n \times n$ complex matrix. Then for any invertible $n \times n$ complex matrix S, the linear system $\Sigma_2 = (S^{-1}AS, S^{-1}B, CS, D)$ also has transfer function $T(s)$.*

Proof We simply compute the new transfer function

$$D + CS(sI - S^{-1}AS)^{-1}S^{-1}B = D + C(sI - A)^{-1}B = T(s). \tag{3.1}$$

□

This result suggests that one can simplify the original linear system by reducing the main transformation A to $S^{-1}AS$ with a specific form. The similarity S is otherwise described by choosing a basis for $\mathbb{C}^{n\times 1}$ other than the standard basis $\{e_1, \dots, e_n\}$, so that the new basis is adapted to the matrix A. In particular, in the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_3

next section we consider Jordan decomposition of A, then in Sect. 3.3 we consider the implications for the resolvent $(sI - A)^{-1}$.

3.2 Jordan Blocks

A $k \times k$ Jordan block with eigenvalue $\lambda \in \mathbb{C}$ is the matrix

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & 0 & \dots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \ddots & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{bmatrix} \tag{3.2}$$

$$J_1(\lambda) = [\lambda], \quad J_2(\lambda) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \tag{3.3}$$

$$J_3(\lambda) = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} \tag{3.4}$$

Jordan canonical form (Jordan normal form)

(i) Let λ be an entry on the leading diagonal of a square matrix A such that all the other entries in the same column as λ are 0. Observe if λ is in row j, then the column vector with 1 in row j and 0 elsewhere is an eigenvector corresponding to eigenvalue λ. Hence λ is an eigenvalue of A.
(ii) In a Jordan block, all the entries down the leading diagonal are equal to some λ, all the entries in the diagonal above the leading diagonal are 1, and all other entries are zero. By (i), λ is an eigenvalue of the Jordan block.
(iii) We put Jordan blocks of various sizes into a block matrix, so that the blocks on the block diagonal are Jordan blocks. Then all the entries below the leading diagonal are zero; all the entries in the diagonal directly above the leading diagonal are 0 or 1, and all the entries above this diagonal are all zero.

Definition 3.2 (Eigenvalue Terminology)

(i) Each eigenvalue λ_j has algebraic multiplicity n_j, where n_j is the largest power of $(z - \lambda_j)$ that divides the characteristic polynomial of A.
(ii) For each eigenvalue λ_j, there is an eigenvector v_j. Let $E(\lambda_j) = \{v : Av = \lambda_j v\}$ be the eigenspace. The geometric multiplicity of λ_j is the dimension of $E(\lambda_j)$.

(iii) For each λ_j, the geometric multiplicity is the number of Jordan blocks that involve λ_j, so the geometric multiplicity is less than or equal to the algebraic multiplicity.
(iv) When a Jordan block has shape $k \times k$, where $k > 1$, it has both eigenvectors and generalized eigenvectors. A generalized eigenvector is $v \neq 0$ such that $(\lambda_j I - A)^m v = 0$ for some $m = 2, \dots, k$.

Example 3.3 Consider the matrix A that is given by

$$A = \begin{bmatrix} 2^* & 1 & & \dots & \dots & & & & 0 \\ 0 & 2 & & \dots & & & & & \vdots \\ & \ddots & 2^* & 1 & 0 \dots & & & & \vdots \\ \vdots & & 0 & 2 & 1 & & \dots & & \\ & \dots & 0 & 0 & 2 & & \dots & & \\ \vdots & & & & & 2^* & & & \\ & & & & & & 2^* & & \\ \vdots & & & & & & & 3^* & \\ 0 & \dots & & & & & & & 3^* \end{bmatrix} \tag{3.5}$$

in which zeros outside the Jordan blocks on the diagonal are mostly omitted. Here the eigenvalues are marked 2^* and 3^*; these are the only entries in their own column; the eigenspaces are $E(2)$ of dimension 4 since there are four blocks involving eigenvalue 2, and $E(3)$ of dimension 2 since there are two blocks involving eigenvalue 3. The eigenvalues are 2, 2, 2, 2, 3, 3 and the corresponding eigenvectors are $e_1, e_3, e_6, e_7, e_8, e_9$.The matrix A has Jordan blocks

$$J_2(2) \oplus J_3(2) \oplus J_1(2) \oplus J_1(2) \oplus J_1(3) \oplus J_1(3) \tag{3.6}$$

and A has shape 9×9 since $2+3+1+1+1+1 = 9$; the characteristic polynomial is

$$\chi_A(s) = (s-2)^7 (s-3)^2 \tag{3.7}$$

which is given by the product of the diagonal terms of $sI - A$ and has roots $2, 2, 2, 2, 2, 2, 2, 3, 3$; the minimal polynomial is

$$m(s) = (s-2)^3 (s-3) \tag{3.8}$$

since the largest block involving eigenvalue 2 is $J_3(2)$ and the largest with eigenvalue 3 is $J_1(3)$.

Remark 3.4 Computing eigenvalues

(i) One way to compute eigenvalues is to solve the characteristic equation (2.39), either algebraically or numerically. Computers can employ some special algorithms to find approximate values for eigenvalues.
(ii) For complex matrices up to and including size 4×4, it is possible to compute the Jordan canonical form by using algebraic results as in Sect. 6.3 on stable cubics. However, computing JCF by hand is a bore.
(iii) In any method, multiple eigenvalues are a technical challenge. MATLAB has a command $jordan$ for finding the JCF.
(iv) However, computers find it easy to check whether matrices are positive definite as in Theorem 3.23, so researchers have found clever ways to use linear matrix inequalities instead of computing eigenvalues.
(v) Given $n \times n$ complex matrices A_1 and A_2, we can compute a Jordan canonical form for A_1 and a Jordan canonical form for A_2. However, the bases and similarity matrices that we choose for A_1 and A_2 might not be related to one another in any simple way. The topic of simultaneous reduction of matrices is complicated, and various results discussed in [28]. In some applications to linear systems, it is possible to avoid this problem by using results such as Proposition 7.10.

Theorem 3.5 (Jordan Canonical Form) *Let A be an $(n \times n)$ complex matrix. Then there exist S an invertible $n \times n$ complex matrix, a partition of n into $n = k_1 + k_2 + \cdots + k_r$ and $k_j \times k_j$ Jordan blocks $J_{k_j}(\lambda_j)$ where λ_j is some eigenvalue of A, such that A is similar to the sum of Jordan blocks*

$$A = S \begin{bmatrix} J_{k_1}(\lambda_1) & 0 & \dots & 0 \\ 0 & J_{k_2}(\lambda_2) & 0 & \dots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & J_{k_r}(\lambda_r) \end{bmatrix} S^{-1}. \tag{3.9}$$

Proof See [20] page 183. □

3.3 Exponentials and Eigenvalues of Complex Matrices

Lemma 3.6 *Let A be a $(n \times n)$ complex matrix with eigenvalues λ_j, where* $\max_j \Re\lambda_j < \beta$ *for some real β.*

(i) Then the entries of $\exp(tA)$ *are complex linear combinations of $t^k e^{t\lambda_j}$ for integers $k = 0, 1, \dots, n-1$.*
(ii) There exists M such that

$$\|\exp(tA)\| \le Me^{\beta t} \qquad (t > 0). \tag{3.10}$$

The condition $\Re\lambda_j < \beta$ means that the point λ_j lies strictly to the left of the vertical line in the complex plane through β on the real axis.

Proof Reducing to Jordan blocks: From the Jordan canonical form 3.5, we have

$$\exp(tA) = S \begin{bmatrix} \exp(tJ_{k_1}(\lambda_1)) & 0 & \dots & 0 \\ 0 & \exp(tJ_{k_2}(\lambda_2)) & 0 & \dots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \exp(tJ_{k_r}(\lambda_r)) \end{bmatrix} S^{-1}, \tag{3.11}$$

so we consider a typical block $J_k(\lambda)$. Now

$$J_k(\lambda) = \begin{bmatrix} \lambda & 0 & \dots\dots & 0 \\ 0 & \lambda & 0 & \dots \\ \vdots & 0 & \ddots & \ddots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 0 \end{bmatrix} \tag{3.12}$$

which we write as

$$J_k(\lambda) = \lambda I_k + N_k \tag{3.13}$$

where N_k is strictly upper triangular, and I_k and N_k commute, so the exponential of a Jordan block is

$$\exp(tJ_k(\lambda)) = \exp(t\lambda I_k)\exp(tN_k). \tag{3.14}$$

Now $N_k^k = 0$, so we have a polynomial of degree $k - 1 < n$

$$\exp(tN_k) = I + tN_k + \dots + t^{k-1}N_k^{k-1}/(k-1)! \tag{3.15}$$

and $|\exp(t\lambda)| = e^{t\Re\lambda}$, hence we obtain the bound

$$\|\exp(tJ_k(\lambda))\| \le e^{t\Re\lambda}\Big(1 + t\|N_k\| + \dots + \frac{t^{k-1}\|N_k\|^{k-1}}{(k-1)!}\Big). \tag{3.16}$$

Hence

$$\begin{aligned} \|\exp(tA)\| &\le \|S\|\|S^{-1}\| \sum_{j=1}^{r} \|\exp(tJ_{k_j}(\lambda_j))\| \\ &\le \|S\|\|S^{-1}\| \sum_{j=1}^{r} e^{t\Re\lambda_j} \sum_{\ell=0}^{k_j-1} \frac{t^\ell \|N_{k_j}^\ell\|}{\ell!}. \end{aligned}$$

Now choose $\varepsilon > 0$ such that $\Re\lambda_j + \varepsilon < \beta$.

For all $t \geq 0$, we have

$$e^{\varepsilon t} = 1 + \varepsilon t + \frac{\varepsilon^2 t^2}{2!} + \cdots + \frac{\varepsilon^\ell t^\ell}{\ell!} + \cdots \geq \frac{\varepsilon^\ell t^\ell}{\ell!},$$

so $t^\ell \leq \ell! e^{\varepsilon t}/\varepsilon^\ell$. Observe that $t^\ell e^{-\varepsilon t}$ is bounded for all $t > 0$, also $e^{t\Re\lambda_j} \leq e^{\beta t} e^{-\varepsilon t}$, so

$$M = \sup_{t>0} \Big(e^{-\varepsilon t} \|S\| \|S^{-1}\| \sum_{j=1}^{r} \sum_{\ell=0}^{k_j-1} \frac{t^\ell \|N_{k_j}^\ell\|}{\ell!} \Big) \tag{3.17}$$

is finite. Hence

$$\|\exp(tA)\| \leq M e^{\beta t} \qquad (t > 0). \tag{3.18}$$

□

The following result gives conditions under which solutions grow.

Proposition 3.7 (Growth of Solutions) *Let A be a $n \times n$ complex matrix and consider*

$$\frac{dX}{dt} = AX; \qquad X(0) = X_0 \tag{3.19}$$

where the initial value $X_0 \in \mathbb{C}^{n\times 1}$ is to be chosen.

(i) Suppose that A has an eigenvalue λ such that $\Re\lambda > 0$. Then there exists a solution that grows at exponential rate $\Re\lambda$, so $\|X(t)\| \leq M e^{t\Re\lambda}$ for all $t > 0$ for some $M > 0$.

(ii) Suppose that A has an eigenvalue λ of geometric multiplicity $k > 1$ such that $\Re\lambda = 0$. Then there exists a solution that grows at polynomial rate $k-1$, so $\|X(t)\| \leq M(1 + t^{k-1})$ for all $t > 0$ for some $M > 0$.

Proof

(i) Let V be an eigenvector corresponding to eigenvalue λ, and with $X_0 = V$ introduce the solution $X(t) = \exp(tA)V$. Then $X(t) = e^{\lambda t}V$ satisfies $\|X\| = e^{t\Re\lambda}\|V\|$, hence grows exponentially at rate $\Re\lambda > 0$.

(ii) Here A is similar to a Jordan block matrix that contains a Jordan block $J_k(\lambda) = \lambda I_k + N_k$, where

$$N_k^{k-1} = \begin{bmatrix} 0 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \tag{3.20}$$

and $N_k^k = 0$. Then

$$\exp(tJ_k(\lambda)) = \exp(t\lambda I_k)\exp(tN_k) = e^{t\lambda}\Big(I_k + tN_k + \cdots + \frac{t^{k-1}N_k^{k-1}}{k!}\Big), \tag{3.21}$$

so the exists an invertible matrix S such that

$$\|S\|\|S^{-1}\|\|\exp(tA)\| \geq \|S^{-1}\exp(tA)S\| \geq \|\exp(tJ_k(\lambda))\| \geq \frac{t^{k-1}}{k!}. \tag{3.22}$$

Hence we can choose X_0 to produce a growing solution.

This result does not assert that every solution grows, only that some initial conditions produce growing solutions. the result does not apply to the situation in which A has eigenvalues of geometric multiplicity one on the imaginary axis $\Re\lambda = 0$. To address this subtle case, we introduce the notion of resonance in Chap. 5. □

3.4 Exponentials and the Resolvent

Definition 3.8 (Resolvent) Let A be $n \times n$ complex matrix with set of eigenvalues $\sigma = \{\lambda_j;\ j = 1, \dots, n\}$. Here σ is called the spectrum, $\mathbb{C} \setminus \sigma$ is the resolvent set, and the matrix function $R(s) = (sI - A)^{-1}$ on $\mathbb{C} \setminus \sigma$ is called the resolvent.

Proposition 3.9

(i) (Resolvent identity) The resolvent $R(s) = (sI - A)^{-1}$ *satisfies*

$$R(s) - R(\lambda) = (\lambda - s)R(s)R(\lambda). \tag{3.23}$$

(ii) Also $R(s)$ *is a differentiable function such that*

$$\frac{d}{ds}R(s) = -R(s)^2$$

when s *is not an eigenvalue of* A.

Proof

(i) We have

$$\lambda I - A = (sI - A) + (\lambda - s)I$$

which we multiply on the left by $R(\lambda) = (\lambda I - A)^{-1}$ and multiply on the right by $R(s) = (sI - A)^{-1}$ to give

$$R(s) = R(\lambda) + (\lambda - s)R(\lambda)R(s).$$

(ii) When s is not an eigenvalue, $R(\lambda)$ is a continuous function of λ on an open neighbourhood of s. We deduce from (i) that

$$\begin{aligned}\frac{R(s) - R(\lambda)}{s - \lambda} &= -R(s)R(\lambda)\\ &\to -R(s)^2 \qquad (\lambda \to s),\end{aligned}$$

so $R(s)$ is complex differentiable.

□

There are various formulas for the resolvent, including the following one which links the exponential to the resolvent.

Proposition 3.10 (Resolvent Formula) *Let A be an $(n \times n)$ complex matrix such that $\|\exp(tA)\| \leq Me^{\beta t}$ for all $t > 0$ Then for $\Re s > \beta$, the matrix $sI - A$ is invertible with inverse*

$$(sI - A)^{-1} = \int_0^\infty e^{-st}\exp(tA)\,dt. \tag{3.24}$$

Proof We have

$$(sI - A)\exp(t(A - sI)) = -\frac{d}{dt}\exp(t(A - sI)) \tag{3.25}$$

so

$$\int_0^R (sI - A)\exp(t(A - sI))\,dt = \int_0^R -\frac{d}{dt}\exp(t(A - sI))\,dt \tag{3.26}$$

so by Fundamental Theorem of Calculus

$$(sI - A)\int_0^R \exp(t(A - sI))\,dt = \big[-\exp(t(A - sI))\big]_0^R = I - \exp R(A - sI); \tag{3.27}$$

where

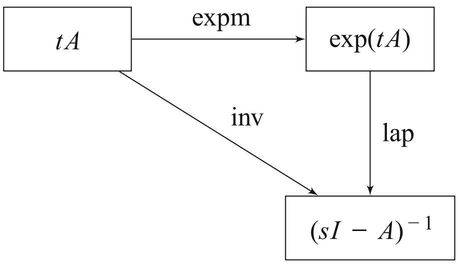
$$\|\exp R(A - sI)\| = \|\exp RA\||e^{-sR}| \leq Me^{\beta R - R\Re s} \tag{3.28}$$

where $\beta - \Re s < 0$, so by the assumption on s, we can take the limit as $R \to \infty$ to obtain

$$(sI - A)\int_0^\infty \exp(t(A - sI))\,dt = I. \tag{3.29}$$

Hence $sI - A$ is invertible. □

The following diagram gives the basic MATLAB commands for computing these.



Proposition 3.11 (Spectral Radius Formula) *Let* $\rho = \limsup_{n\to\infty} \|A^n\|^{1/n}$.

(i) Suppose that $s \in \mathbb{C}$ *has* $|s| > \rho$. *Then*

$$R(s) = \sum_{n=0}^\infty \frac{A^n}{|s|^{n+1}} \tag{3.30}$$

converges and satisfies $R(s)(sI - A) = (sI - A)R(s) = I$.

(ii) The spectrum of A *is contained in the closed disc* $\{\lambda \in \mathbb{C} : |\lambda| \le \rho\}$.

Proof

(i) The n^{th} term in the series has norm $\|A^n\|/|s|^{n+1}$ where

$$\limsup_{n\to\infty} \frac{\|A^n\|^{1/n}}{|s|^{(n+1)/n}} = \frac{\rho}{|s|} < 1 \tag{3.31}$$

hence the series converges by the n^{th} root test. Let

$$R_n(s) = \frac{1}{s}I + \frac{1}{s^2}A + \frac{1}{s^3}A^2 + \frac{1}{s^4}A^3 + \cdots + \frac{1}{s^n}A^{n-1}, \tag{3.32}$$

so

$$sR_n(s) - AR_n(s) = I - \frac{1}{s^n}A^n, \tag{3.33}$$

so $R_n(s) \to R(s)$ as $n \to \infty$, where $(sI - A)R(s) = I$. Likewise $R(s)(sI - A) = I$.

(ii) The series for $R(s)$ converges whenever $|s| > \rho$, and gives an inverse for $sI - A$. Hence $\det(sI - A) \neq 0$, and s cannot be an eigenvalue of A. Conversely if v is an eigenvector corresponding to eigenvalue λ, then $\|A^n v\| = |\lambda|^n \|v\|$, so $\limsup_{n\to\infty} \|A^n\|^{1/n} \geq |\lambda|$, so $\rho \geq |\lambda|$.

□

3.5 Schur Complements

Definition 3.12 (Schur Complements) Given an invertible $n \times n$ matrix A, the Schur complement of A in the block matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{3.34}$$

is $D - CA^{-1}B$. For D square and invertible, the Schur complement of D is $A - BD^{-1}C$.

Example 3.13 The Schur complement of $A - sI$ in

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \tag{3.35}$$

is the transfer function

$$T(s) = D - C(A - sI)^{-1}B = D + C(sI - A)^{-1}B. \tag{3.36}$$

For the next two results, we specialize to the shape

$$U = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad \begin{bmatrix} n \times n & n \times k \\ k \times n & k \times k \end{bmatrix}. \tag{3.37}$$

Lemma 3.14 *Suppose that A is invertible. Then*

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det(D - CA^{-1}B). \tag{3.38}$$

Proof One can show that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}, \tag{3.39}$$

and one deduces the required result form taking the determinant of these matrix products. The final matrix is upper triangular with 1 on the leading diagonal, hence has determinant 1; the first matrix on right-hand side is lower triangular with 1 on leading diagonal, hence has determinant 1. □

Proposition 3.15 (A Determinant Formula for Realization by SISO) *Let $T(s) = D + p(s)/q(s)$ be a proper rational function, where $q(s)$ is a monic polynomial and $p(s)$ is a complex polynomial of degree less than $q(s)$. Then $T(s)$ can be realized as the transfer function of a SISO (A, B, C, D), so $T(s) = D + C(sI - A)^{-1}B$, where*

- *A is a companion matrix with final row given by the coefficients of $q(s)$ after the leading coefficient, reversed in order and with minus signs;*
- *B is the column $[0; \dots; 0; 1]$;*
- *C is a row vector given by the coefficients of $p(s)$, reversed in order;*
- *D is given by $T(s) \to D$ as $s \to \infty$.*

In particular, with A, B, C as above and $D \in \mathbb{C}$,

$$T(s) = \frac{\det \begin{bmatrix} A - sI & B \\ C & D \end{bmatrix}}{\det[A - sI]}.$$

Proof This follows by combining the Example 3.13 and the Lemma 3.14. For a SISO, the entry D is a scalar. □

Proposition 3.16 *Suppose that A and $D - CA^{-1}B$ are invertible. Then U is invertible with inverse*

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}. \tag{3.40}$$

Proof To motivate this formula, we carry out elementary row operations on

$$V = \begin{bmatrix} A & B & | & I & 0 \\ C & D & | & 0 & I \end{bmatrix}, \tag{3.41}$$

and thus find the inverse of U. An elementary row operation on V amounts to multiplying V on the left by an invertible matrix E to form EV.

Since A is invertible, we have row equivalences

$$\begin{bmatrix} A & B & | & I & 0 \\ C & D & | & 0 & I \end{bmatrix} \sim \begin{bmatrix} I & A^{-1}B & | & A^{-1} & 0 \\ C & D & | & 0 & I \end{bmatrix}$$
$$\sim \begin{bmatrix} I & A^{-1}B & | & A^{-1} & 0 \\ 0 & D - CA^{-1}B & | & -CA^{-1} & I \end{bmatrix} \tag{3.42}$$

It is now clear that the 2×2 block matrix is invertible if and only if $D - CA^{-1}B$ is invertible, in which case we have

$$V \sim \begin{bmatrix} I & A^{-1}B & | & A^{-1} & 0 \\ 0 & I & | & -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$
$$\sim \begin{bmatrix} I & 0 & | & A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ 0 & I & | & -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}, \tag{3.43}$$

so the inverse is the right block, namely

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}. \tag{3.44}$$

□

3.6 Self-adjoint Matrices

We write $z = (z_j)_{j=1}^n$ and $w = (w_j)_{j=1}^n$, and introduce the inner product $\langle z, w \rangle = \sum_{j=1}^n z_j \bar{w}_j$. With linear operators $T : \mathbb{C}^m \to \mathbb{C}^n$ the adjoint has $T' : \mathbb{C}^n \to \mathbb{C}^m$ and is characterized by the identity

$$\langle Tv, w \rangle = \langle v, T'w \rangle \qquad (v \in \mathbb{C}^m, w \in \mathbb{C}^n) \tag{3.45}$$

for the standard inner product. For many purposes, (3.45) is the most helpful way to think about the adjoint, instead of the matricial definition 2.15. In particular, one sees that for $n \times n$ matrices A and B, the adjoint reverses the order in matrix products, so $(AB)' = B'A'$.

Lemma 3.17 *Let* $A \in M_{n \times n}(\mathbb{C})$.

(i) If $\langle AX, X \rangle = 0$ *for all* $X \in \mathbb{C}^n$, *then* $A = 0$.
(ii) $A = A'$ *if and only if* $\langle AX, X \rangle$ *is real for all* $X \in \mathbb{C}^n$.

Proof

(i) We need to consider complex vectors X, as evidenced by the example in $M_{2\times 2}(\mathbb{R})$ given by

$$\left\langle \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x \\ y \end{bmatrix} \right\rangle = 0 \qquad (x, y \in \mathbb{R}). \tag{3.46}$$

We take $X, Y \in \mathbb{C}^n$ and $s \in \mathbb{C}$, and write

$$\begin{aligned} 0 &= \langle A(X + sY), X + sY \rangle \\ &= \langle AX, X \rangle + s\langle AY, X \rangle + \bar{s}\langle AX, Y \rangle + |s|^2 \langle AY, Y \rangle \\ &= s\langle AY, X \rangle + \bar{s}\langle AX, Y \rangle, \end{aligned} \tag{3.47}$$

so by considering $s = t$ and $s = it$ for $t \in \mathbb{R}$, we deduce that $\langle AX, Y \rangle = 0$, so $A = 0$.

(ii) We have

$$\begin{aligned} \langle AX, X \rangle \in \mathbb{R} \Leftrightarrow \langle AX, X \rangle = \overline{\langle AX, X \rangle} \Leftrightarrow \langle AX, X \rangle \\ = \overline{\langle X, A'X \rangle} \Leftrightarrow \langle AX, X \rangle = \langle A'X, X \rangle \end{aligned}$$

and by (i), this is equivalent to $A = A'$.

□

Definition 3.18 (Self-adjoint)

(i) We say that $A \in M_{n\times n}(\mathbb{C})$ is self-adjoint if $A = A'$.
(ii) We say that $S \in M_{n\times n}(\mathbb{C})$ is skew self-adjoint if $S' = -S$. Often one says that S is skew. Equivalently $S = iA$ where A is self-adjoint.

Example 3.19 For any $A \in M_{n\times n}(\mathbb{C})$, the operators $A + A'$, AA' and $A'A$ are self-adjoint, whereas $A - A'$ is skew self-adjoint.

Theorem 3.20 (Spectral Theorem for Self-adjoint Matrices) *Suppose that $A \in M_{n\times n}(\mathbb{C})$ is self-adjoint, so $A = A'$.*

(i) Then the eigenvalues of A are all real,
(ii) eigenvectors corresponding to distinct eigenvalues are orthogonal, and
(iii) there exists a unitary matrix U such that $UU' = U'U = I$ and $A = UDU'$ where D is a real diagonal matrix.

Proof

(i) For an eigenvalue λ with corresponding eigenvector X, we have $\lambda X = AX$, so

$$\lambda\langle X, X \rangle = \langle AX, X \rangle = \langle X, A'X \rangle = \langle X, AX \rangle = \bar{\lambda}\langle X, X \rangle, \tag{3.48}$$

and $\langle X, X \rangle = \|X\|^2 > 0$ since $X \neq 0$, so $\lambda = \bar{\lambda}$.

(ii) For an eigenvalue λ with corresponding eigenvector X, and an eigenvalue μ with $\mu \neq \lambda$ with corresponding eigenvector Y, we have

$$\lambda\langle X, Y\rangle = \langle AX, Y\rangle = \langle X, A'Y\rangle = \langle X, AY\rangle = \mu\langle X, Y\rangle, \tag{3.49}$$

so $\langle X, Y\rangle = 0$.

(iii) We find the largest eigenvalue λ_1. The set $K_1 = \{X \in \mathbb{C}^n : \langle X, X\rangle = 1\}$ is closed and bounded, so the function $K_1 \to \mathbb{R} : X \mapsto \langle AX, X\rangle$ is bounded and attains its supremum. We write λ_1 for this supremum, and choose $X_1 \in K_1$ such that $\langle AX_1, X_1\rangle = \lambda_1\langle X_1, X_1\rangle$. Then for any fixed $Y \in \mathbb{C}^n$, the real function

$$f(t) = \langle A(X_1+tY), X_1+tY\rangle - \lambda_1\langle X_1+tY, X_1+tY\rangle \qquad (t \in \mathbb{R}) \tag{3.50}$$

has $f(t) \leq 0$ for all t and $f(0) = 0$, so by calculus

$$0 = \frac{df}{dt}(0) = \langle AX_1, Y\rangle + \langle AY, X_1\rangle - \lambda_1\langle X_1, Y\rangle - \lambda_1\langle Y, X_1\rangle \tag{3.51}$$

so

$$\langle AX_1 - \lambda_1 X_1, Y\rangle + \langle Y, AX_1 - \lambda_1 X\rangle = 0; \tag{3.52}$$

but Y was arbitrary, so we can choose $Y = AX_1 - \lambda_1 X_1$ to deduce that

$$\langle AX_1 - \lambda_1 X_1, AX_1 - \lambda_1 X_1\rangle = 0;$$

hence $AX_1 - \lambda_1 X_1 = 0$ and we have an eigenvector X_1 with corresponding eigenvalue λ_1.

Observe that

$$\langle X, X_1\rangle = 0 \Rightarrow \langle X, \lambda_1 X_1\rangle = 0 \Rightarrow \langle X, AX_1\rangle = 0 \Rightarrow \langle AX, X_1\rangle = 0. \tag{3.53}$$

Hence we can repeat the argument with $K_2 = \{X \in \mathbb{C}^n : \langle X, X\rangle = 1, \langle X, X_1\rangle = 0\}$ in place of K_1 to find another eigenvalue λ_2. Thus by an induction argument we can find eigenvectors $X_1, \dots, X_n$ and corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. By (ii), the eigenvectors X_j of A are orthogonal, so we can choose them so that $\langle X_j, X_k\rangle = 0$ for $j \neq k$ and $\langle X_j, X_j\rangle = 1$. Then $U = [X_1 \dots X_n]$ satisfies $U'U = I$. Also $U'AU$ is the diagonal matrix $D = \mathrm{diag}[\lambda_1, \dots, \lambda_n]$. □

Exercise

(i) Suppose that $A = A'$. Show that the eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ of A determine the norm of A via $\|A\| = \max\{|\lambda_j| : j = 1, \dots, n\}$.

(ii) Let B be a $m \times n$ matrix. Show that $B'B$ and BB' are square matrices and their largest eigenvalues are equal.

Exercise Suppose that $K = K'$ has all its eigenvalues positive. Using the spectral theorem, show that

(i) K is invertible, and the inverse K^{-1} also has all its eigenvalues positive;
(ii) there exists a L such that $L' = L$, all eigenvalues of L are positive and $L^2 = K$. This L is unique, and is known as the positive square root of K.

The following result is a variant of the rank-nullity theorem 2.2.

Proposition 3.21 *Let $A \in M_{n\times n}(\mathbb{C})$ and let $V = \{x \in \mathbb{C}^n : A'x = 0\}$ be the nullspace of A'.*

(i) Then V is a linear subspace of $\mathbb{C}^n$, and its orthogonal complement $V^\perp = \{y \in \mathbb{C}^n : \langle y, v\rangle = 0, \forall v \in V\}$ is equal to the range $\{Ay : y \in \mathbb{C}^n\}$ of A.
(ii) Also A' maps V into V, and A maps $V^\perp$ into $V^\perp$.
(iii) rank$(A) = \dim(V^\perp) = r$ and nullity$(A) = \dim V = k$, where $n = k + r$.
(iv) There exist a unitary $U \in M_{n\times n}(\mathbb{C})$, $A_{1,1} \in M_{r\times r}(\mathbb{C})$ and $A_{1,2} \in M_{r\times k}(\mathbb{C})$ such that

$$U'AU = \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & 0 \end{bmatrix}, \quad U'A'U = \begin{bmatrix} A'_{1,1} & 0 \\ A'_{1,2} & 0 \end{bmatrix}.$$

Proof

(i) For $x, z \in V$ and $\lambda, \mu \in \mathbb{C}$, we have $A'(\lambda x + \mu z) = \lambda A'x + \mu A'z = 0$, so $\lambda x + \mu z \in V$. Hence V is a linear subspace of $\mathbb{C}^n$. To identify its orthogonal complement, we observe that $x \in V$ if and only if $\langle y, A'x\rangle = 0$ for all $y \in \mathbb{C}^n$ so $\langle Ay, x\rangle = 0$ for all y; so x is perpendicular to the range of A.
(ii) This follows from the definition of V and (i).
(iii) We have an orthogonal direct sum $\mathbb{C}^n = V \oplus V^\perp$, so we can add the dimensions $n = \dim(V) + \dim(V^\perp)$. Note that the row rank of A is equal to the column rank of A, so rank$(A) = $ rank(A').
(iv) Using the Gram-Schmidt process or otherwise [**?**], we choose an orthonormal basis $\{e_1, \dots, e_r\}$ of $V^\perp$ and an orthonormal basis $\{e_{r+1}, \dots, e_n\}$ of V and let U be the unitary that takes $\{e_1, \dots, e_n\}$ to the standard basis of $\mathbb{C}^n$. The matrix decomposition then follows from (ii).

□

3.7 Positive Definite Matrices

Definition 3.22 (Positive Definite) An $n \times n$ complex matrix K is said to be positive definite if $K = K'$ and $\langle KZ, Z\rangle > 0$ for all $Z \in \mathbb{C}^n$ such that $Z \neq 0$; we write $K \succ 0$. We say that L is negative definite if $K = -L$ is positive definite; we write $L \prec 0$.

Beware that the product of positive definite matrices is generally not positive definite. Exercise 3.7 gives basic properties of positive definite matrices. The fundamental characterization is the following theorem. A principal leading minor is the determinant of one of the top left corner blocks of a matrix.

Theorem 3.23 (Sylvester) *Let K be a $(n \times n)$ complex matrix such that $K = K'$. Then the following are equivalent:*

(i) $\langle KZ, Z\rangle > 0$ for all $Z \in \mathbb{C}^n$ such that $Z \neq 0$;
(ii) the eigenvalues κ_j of K are all real and $\kappa_j > 0$ for all j;
(iii) the leading principal minors Δ_j of K are all positive, so $\Delta_j > 0$ for all j.

Proof $(i) \Rightarrow (ii)$ For an eigenvector X with corresponding eigenvalue κ, we have $\kappa X = KX$ so $\kappa\langle X, X\rangle = \langle KX, X\rangle > 0$, hence $\kappa > 0$.

$(ii) \Rightarrow (i)$ By the spectral theorem, we have $K = UDU'$ where D is the diagonal matrix with entries the eigenvalues of K, which are all positive, so (i) follows.

$(i) \Rightarrow (iii)$ Let K_j be the $j \times j$ submatrix of K in the top left corner. Then $\langle K_j X, X\rangle > 0$ by (i). Then the eigenvalues of K_j are all positive since $(i) \Rightarrow (ii)$; hence $\Delta_j = \det K_j > 0$.

$(iii) \Rightarrow (i)$ The proof is by induction on the number n of rows of the matrix. The basis of induction is the case $n = 1$, which is evident. Suppose that $(iii) \Rightarrow (i)$ holds for matrices with n rows, and consider $n + 1$; that is, consider a self-adjoint matrix K of the shape

$$K = \begin{bmatrix} A & B \\ B' & D \end{bmatrix} \qquad \begin{bmatrix} n \times n & n \times 1 \\ 1 \times n & 1 \times 1 \end{bmatrix} \tag{3.54}$$

and suppose that all the leading minors have $\Delta_j > 0$. Then by the induction hypothesis, A is positive definite, and in particular is invertible; also A has a positive definite square root $A^{1/2}$ with inverse $A^{-1/2}$ by the exercises. Then

$$\Delta_{n+1} = \det \begin{bmatrix} A & B \\ B' & D \end{bmatrix} = (\det A)(D - B'A^{-1}B) = \Delta_n(D - B'A^{-1}B) \tag{3.55}$$

so the final factor is positive. We can then write

$$\begin{aligned} \left\langle \begin{bmatrix} A & B \\ B' & D \end{bmatrix} \begin{bmatrix} X \\ \xi \end{bmatrix}, \begin{bmatrix} X \\ \xi \end{bmatrix} \right\rangle &= \langle AX, X\rangle + \xi\langle B, X\rangle + \bar{\xi}\langle X, B\rangle + D|\xi|^2 \\ &= \|A^{1/2}X + \xi A^{-1/2}B\|^2 + (D - \langle A^{-1/2}B, A^{-1/2}B\rangle)|\xi|^2, \end{aligned} \tag{3.56}$$

which is a sum of nonnegative terms. If $\xi \neq 0$, then the final term is positive; whereas if $\xi = 0$, then we are left with $\langle AX, X\rangle$ which is positive unless X satisfies $X = 0$. Hence the matrix K is positive definite. □

For small matrices, (iii) is easy to check, and computers can carry out (iii) in exact arithmetic for medium sized matrices. For large sized matrices, computers can handle (i). The condition (ii) involves the eigenvalues, which are useful to find for other purposes, but can be difficult to locate exactly.

Example 3.24 Part (iii) of the Theorem 3.23 is about positive leading minors; whereas nonnegative leading minors themselves do not carry much information. The leading minors of

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \tag{3.57}$$

are $\Delta_1 = 0 \geq 0$ and $\Delta_2 = 0 \geq 0$, but A has a negative eigenvalue -1 and is not positive definite.

Exercise Let $B \in M_{m\times n}(\mathbb{C})$. Show that the following are equivalent :

(i) $\|B\| \leq 1$;
(ii) $B'B$ has all its eigenvalues less than or equal to 1;
(iii) $\langle (I_n - B'B)X, X\rangle \geq 0$ for all $X \in \mathbb{C}^n$.

Definition 3.25 Let K be a $(n \times n)$ complex matrix such that $K = K'$, and the eigenvalues κ_j of K are all real and $\kappa_j \geq 0$ for all j. Then K is said to be positive semidefinite.

3.8 Linear Fractional Transformations

An important idea is to compare the scalar function $1/(s - \lambda)$ with the matrix function $(sI - A)^{-1}$, especially when λ is chosen to be an eigenvalue of A. To do this systematically, we broaden the scope of the scalar valued functions slightly, and consider linear fractional transformations.

Definition 3.26 (Linear Fractional Transformations) Given an invertible matrix

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad (ad - bc \neq 0) \tag{3.58}$$

we introduce the Möbius or linear fractional transformation

$$\varphi_M(s) = \frac{as + b}{cs + d} \qquad (cs + d \neq 0). \tag{3.59}$$

Example 3.27 The following are linear fractional transformations:
$T_\alpha(s) = s + \alpha$, translation through $\alpha \in \mathbb{C}$;
$D_r(s) = rs$, dilation through scale factor $r > 0$;

$R_\theta(s) = e^{i\theta}s$, rotation about $\theta \in [0, 2\pi)$;

$J(s) = 1/s$, the reciprocal map. This is not to be confused with $s \mapsto s/|s|^2$, which gives inversion in the unit circle.

Conversely, any linear fractional transformation is a composition of some combination of these basic transformations.

To see this, we consider φ_M as above; there are two cases:

suppose $c = 0$; then $a, d \neq 0$ and we have $\varphi_M(s) = (a/d)s + b/d$. so we make a polar decomposition $a/d = re^{i\theta}$ and write $\varphi_M(s) = re^{i\theta}s + b/d$, so φ_M is the composition of R_θ followed by D_r, followed by T_α, with $\alpha = b/d$.

Now suppose $c \neq 0$, and write

$$\varphi_M(s) = \frac{a}{c} - \frac{(ad - bc)/c}{cs + d} \qquad (cs + d \neq 0), \tag{3.60}$$

which we can express as a composition of the basic transformations, including J.

Let $\mathcal{C}$ be the set of circles and straight lines in $\mathbb{C}$. A typical circle has centre α and radius $r > 0$, so has the formulas $|s - \alpha| = r$, so $s\bar{s} - \alpha\bar{s} - \bar{\alpha}s + |\alpha|^2 = r^2$; a straight line has the form $y = mx + c$ with $m, c \in \mathbb{R}$, so $(s - \bar{s})/(2i) = m(s + \bar{s})/2 + c$; or $x = d$ with $d \in \mathbb{R}$, so $s + \bar{s} = 2d$. We suppose that straight lines pass through ∞.

Proposition 3.28 *Linear fractional transformations map $\mathcal{C}$ to itself.*

Proof By the example, it suffices to show that T_α, D_r, R_θ and J all map $\mathcal{C}$ to itself. The most challenging case is J, so we consider the line $s + \bar{s} = 2d$, which J maps to $s + \bar{s} = 2ds\bar{s}$, which is a circle with centre $1/(2d)$ and radius $1/|2d|$ for $d \neq 0$. For $s + \bar{s} = 0$, we have the imaginary axis, which is mapped by J to itself. Other cases are proved likewise. □

Example 3.29 The following linear fractional transformations are particularly important.

(i) $\varphi(s) = \frac{1}{s+1}$ is used in changes of variable $\lambda = \frac{1}{s+1}$ with inverse $s = \frac{1-\lambda}{\lambda}$.
(ii) $\varphi(s) = \frac{s-1}{s+1}$ takes $RHP = \{s \in \mathbb{C} : \Re s > 0\}$ onto the unit disc $\mathbb{D}(0, 1) = \{s \in \mathbb{C} : |s| < 1\}$.
(iii) Suppose that $a, d, c, d \in \mathbb{R}$ with $ad - bc > 0$. Then $\varphi_M(s)$ takes the upper half plane $\{s : \Im s > 0\}$ onto itself. The inverse function is

$$\varphi_M^{-1}(s) = \varphi_{M^{-1}}(s) = \frac{ds - b}{-cs + a} \qquad (-cs + a \neq 0),$$

which has a similar form.

Example 3.30 Let $T(s) = d + c(s - a)^{-1}b$, which is the transfer function of a linear system with one-dimensional state space. Then $\{T(i\omega) : -\infty \leq \omega \leq \infty\}$ is a circle or straight line in $\mathbb{C}$, passing through d.

3.9 Stable Matrices

Definition 3.31 A complex matrix A is said to be stable if all of its eigenvalues lie in $LHP = \{\lambda \in \mathbb{C} : \Re\lambda < 0\}$.

Proposition 3.32 *Let A be stable, and let $\alpha \in (0, \infty)$ and $\beta \in [0, \infty)$. Then A', $\alpha A - \beta I$ and $-(I - \alpha A)^{-1}$ are also stable.*

Proof First observe that $\det(\lambda I - A) = 0$ if and only if $\det(\bar{\lambda} - A') = 0$, so λ is an eigenvalue of A if and only if $\bar{\lambda}$ is an eigenvalue of A'. Also $\Re\lambda = \Re\bar{\lambda}$, which shows A' is stable if and only if A is stable.

The equations $(\alpha A - \beta I)w = \mu w$ and $Aw = (\mu + \beta)w/\alpha$ are equivalent. In particular, choosing v to be an eigenvector of A corresponding to eigenvalue λ, we have $(\alpha A - \beta I)v = (\alpha\lambda - \beta)v$, where $\Re(\alpha\lambda - \beta) < 0$. The eigenvalues of $\alpha A - I$ are $\alpha\lambda - 1$, where λ is an eigenvalue of A, so $\alpha\lambda - 1 \neq 0$ and $\alpha A - I$ is invertible. Also, $-(I - \alpha A)^{-1}$ has eigenvalues $-1/(1 - \alpha\lambda)$, where

$$\Re\frac{-1}{1 - \alpha\lambda} = \Re\frac{\alpha\lambda - 1}{|1 - \alpha\lambda|^2} < 0. \tag{3.61}$$

□

The notion of stability is fundamentally important in linear systems. Later we see how stability of A relates to other notions of stability, such as stability of polynomials and stability of transfer functions. The definition involves locating all the eigenvalues, which can be computationally difficult for large matrices. Hence we introduce a stricter notion called strict dissipativity, which can be checked without finding eigenvalues, and is a route towards proving stability. See [28] for more details.

3.10 Dissipative Matrices

Definition 3.33

(i) A $n \times n$ complex matrix is strictly dissipative if $\Re\langle Av, v\rangle < 0$ for all $v \in \mathbb{C}^{n\times 1} \setminus \{0\}$. Let $\mathcal{D}_n$ be the set of $n \times n$ strictly dissipative matrices. See [18].
(ii) A $n \times n$ complex matrix A is contractive if $\|A\| \leq 1$.

Exercise

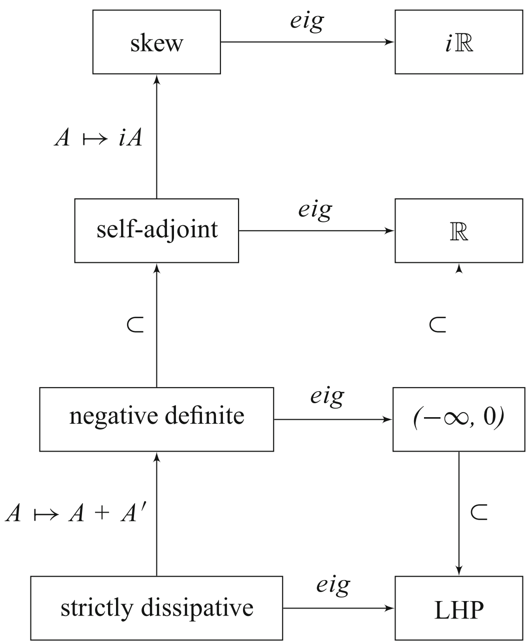
(i) Show that

$$2\Re\langle Av, v\rangle = \langle Av, v\rangle + \overline{\langle Av, v\rangle} = \langle (A + A')v, v\rangle \tag{3.62}$$

where $A + A'$ is self-adjoint.

(ii) Suppose that $A = [a_{jk}]_{j,k=1}^{n}$ is stricly dissipative. Show that the diagonal entries satisfy $\Re a_{jj} < 0$, and $\operatorname{trace}(A) < 0$.
(iii) Suppose that $A = A'$. Show that $A \in \mathcal{D}_n$ if and only if all eigenvalues of A lie in $(-\infty, 0)$. (Hence A is negative definite.)

The following diagram describes the location of the eigenvalues of various types of matrices.



Proposition 3.34

(i) *If* $A \in \mathcal{D}_n$, *then* A *is stable.*
(ii) $\alpha A - \beta I \in \mathcal{D}_n$ *for all* $A \in \mathcal{D}_n$, $\alpha \in (0, \infty)$ *and* $\beta \in [0, \infty)$.
(iii) $A \in \mathcal{D}_n$ *if and only if* $A' \in \mathcal{D}_n$,
(iv) $A \in \mathcal{D}_n$ *if and only if* $A + A' \in \mathcal{D}_n$; *that is* $(A + A')$ *is negative definite;*
(v) *If* $A_1, A_2 \in \mathcal{D}_n$, *then* $A_1 + A_2 \in \mathcal{D}_n$.
(vi) $I - \alpha A \in \mathcal{D}_n$ *for all* $A \in \mathcal{D}_n$ *and* $\alpha \in [0, \infty)$, *and* $\|(I - \alpha A)^{-1}\| \leq 1$.
(vii) *The Cayley transform matrix* $Z = (I + A)(I - A)^{-1}$ *satisfies* $\|Z\| \leq 1$.
(viii) *For* $A \in \mathcal{D}_n$, *there exists* $\kappa > 0$ *such that*

$$\|\exp(tA)\| \leq e^{-\kappa t} \qquad (t > 0). \tag{3.63}$$

(ix) *Conversely, if* A *is an* $n \times n$ *complex matrix that satisfies the inequality of (viii) for some* $\kappa > 0$, *then* $A \in \mathcal{D}_n$.

Proof

(i) The eigenvalue equation gives $v \in \mathbb{C}^{n\times 1} \setminus \{0\}$ such that $Av = \lambda v$, so

$$\Re\langle Av, v\rangle = \Re\lambda\langle v, v\rangle \tag{3.64}$$

where $\Re\langle Av, v\rangle < 0$ and $\langle v, v\rangle > 0$, so $\Re\lambda < 0$, and A is stable.

(ii) For all $v \in \mathbb{C}^{n\times 1} \setminus \{0\}$, we have

$$\Re\langle(\alpha A - \beta I)v, v\rangle = \alpha\Re\langle Av, v\rangle - \beta\langle v, v\rangle < 0. \tag{3.65}$$

(iii) We observe that

$$\langle Av, v\rangle = \langle v, A'v\rangle = \overline{\langle A'v, v\rangle}, \tag{3.66}$$

so $\Re\langle Av, v\rangle = \Re\langle A'v, v\rangle$, hence (iii).

(iv) The proof of (iii) also gives (iv). Observe that $A+A'$ is self-adjoint and strictly dissipative, *ie* equivalent to $K = -(A + A')$ being positive definite.

(v) Likewise, we have

$$\Re\langle(A_1 + A_2)v, v\rangle = \Re\langle A_1 v, v\rangle + \Re\langle A_2 v, v\rangle < 0. \tag{3.67}$$

(vi) By the Cauchy-Schwarz inequality, we write

$$\|(I-\alpha A)v\|\|v\| \geq \Re\langle(I-\alpha A)v, v\rangle = \langle v, v\rangle - \alpha\Re\langle Av, v\rangle \geq \|v\|^2 \tag{3.68}$$

for all $v \in \mathbb{C}^{n\times 1}$. From this we deduce that $\|(I - \alpha A)v\| \geq \|v\|$, so $I - \alpha A$ has zero nullspace and hence is invertible. For all $w \in \mathbb{C}^{n\times 1}$, there exists $v \in \mathbb{C}^{n\times 1}$ such that $w = (I - \alpha A)v$ and $\|w\| \geq \|(I - \alpha A)^{-1}w\|$.

(vii) We have

$$\begin{aligned}\|(I + A)v\|^2 &= \langle v, v\rangle + \langle(A + A')v, v\rangle + \langle Av, Av\rangle \\ &\leq \langle v, v\rangle - \langle(A + A')v, v\rangle + \langle Av, Av\rangle \\ &= \|(I - A)v\|^2,\end{aligned}$$

and since $I - A$ is invertible, we can replace v by $v = (I - A)^{-1}w$ to give $\|(I + A)(I - A)^{-1}w\| \leq \|w\|$.

(viii) First we show that the exponentials satisfy $\|\exp(tA)\| \leq 1$ for all $t > 0$. For $v \in \mathbb{C}^{n\times 1}$, the function $V(t) = \langle \exp(tA)v, \exp(tA)v\rangle$ is non-negative and

$V(0) = \|v\|^2$. Also $V(t)$ is differentiable, with

$$\begin{aligned}\frac{dV}{dt} &= \langle A\exp(tA)v, \exp(tA)v\rangle + \langle \exp(tA)v, A\exp(tA)v\rangle \\ &= \langle (A + A')\exp(tA)v, \exp(tA)v\rangle \le 0, \end{aligned} \tag{3.69}$$

since A is dissipative, so $V(t)$ is decreasing. Hence

$$\|v\|^2 = V(0) \ge \langle \exp(tA)v, \exp(tA)v\rangle = \|\exp(tA)v\|^2 \qquad (t > 0). \tag{3.70}$$

We refine this estimate as follows. Consider the unit sphere $S^{n-1} = \{v \in \mathbb{C}^n : \|v\| = 1\}$, and the continuous map $S^{n-1} \to \mathbb{R}\ v \mapsto \Re\langle Av, v\rangle$, which attains its supremum at v_0, say. Then $\Re\langle Av, v\rangle \le \Re\langle Av_0, v_0\rangle = -\kappa_0$, where $\kappa_0 > 0$ since A is strictly dissipative. Then $A + (\kappa_0/2)I$ is also strictly dissipative since

$$\Re\langle (A + (\kappa_0/2))v, v\rangle = \Re\langle Av, v\rangle + \kappa_0/2 \le -\kappa_0/2 < 0, \tag{3.71}$$

for all $v \in S^n$, so $\Re\langle (A + (\kappa_0/2))v, v\rangle \le -\kappa_0\|v\|^2/2$ for all $v \in \mathbb{C}^n$.

Then we write $\exp(tA) = e^{-\kappa_0 t/2}\exp(t(A+(\kappa_0/2)I)t)$, so by the previous estimate, we have

$$\|\exp(tA)\| \le e^{-\kappa_0 t/2}\|\exp(t(A + (\kappa_0/2)I)t)\| \le e^{-\kappa_0 t/2} \qquad (t > 0). \tag{3.72}$$

(ix) Let $W(t) = \langle \exp(t(A + \kappa I))v, \exp(t(A + \kappa I))v\rangle$, which satisfies $W(t) \le W(0)$ by hypothesis, so $(W(t) - W(0))/t \le 0$ for all $t > 0$, hence taking $t \to 0+$, we deduce that

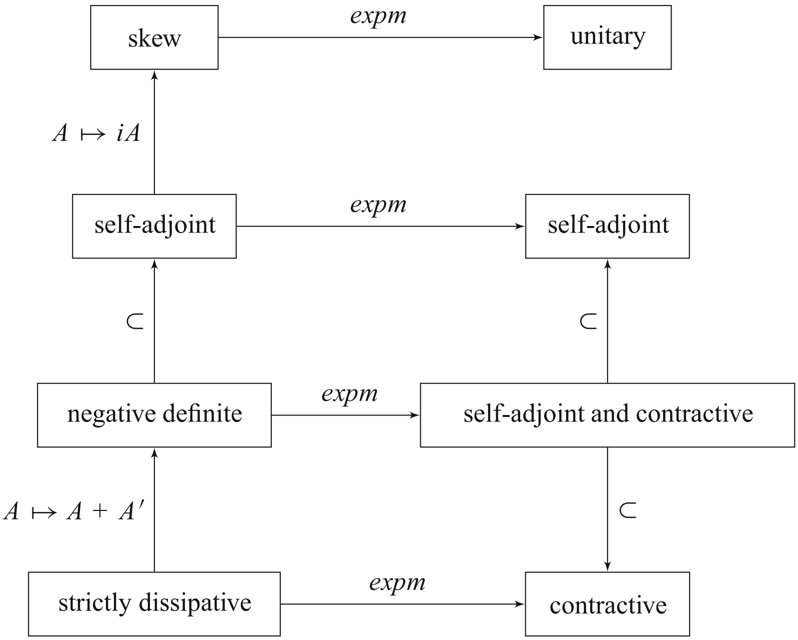
$$\langle (A + \kappa I)v, v\rangle + \langle v, (A + \kappa I)v\rangle = W'(0) \le 0, \tag{3.73}$$

so A is strictly stable.

□

Remark 3.35 Part (iv) of the Proposition 3.34 can be checked in many different ways, as discussed in Theorem 3.23. Part (v) is a simple result, but can be used even when A_1 and A_2 do not commute. For these reasons, it is a good idea to check whether A is strictly dissipative before embarking on an eigenvalue hunt to see whether A is stable.

The following diagram summarizes the effect of the matrix exponential function, denoted *expm* in MATLAB, on some spaces of matrices.



Proposition 3.36 *For all* $A \in \mathcal{D}_n$,

$$\Big(I - \frac{t}{m}A\Big)^{-m} \to \exp(tA) \qquad (m \to \infty), \qquad (t > 0). \tag{3.74}$$

Proof See [18]. The relevance of the following calculation will become clear at the end of the proof. Let X be a Poisson random variable with parameter m, so that $\mathbb{P}[X = k] = e^{-m}m^k/k!$ for $k = 0, 1, \dots$. Then X has expectation

$$\mathbb{E}X = \sum_{k=1}^{\infty} k\mathbb{P}[X = k] = \sum_{k=0}^{\infty} e^{-m}\frac{km^k}{k!} = me^{-m}\sum_{\ell=0}^{\infty}\frac{m^\ell}{\ell!} = m \tag{3.75}$$

and X^2 has expectation

$$\begin{aligned}\mathbb{E}X^2 &= \sum_{k=1}^{\infty} k^2\mathbb{P}[X = k] \\ &= \sum_{k=0}^{\infty} e^{-m}\frac{k(k-1)m^k}{k!} + \sum_{k=0}^{\infty} e^{-m}\frac{km^k}{k!}\end{aligned}$$

$$= m^2 e^{-m} \sum_{\ell=0}^{\infty} \frac{m^\ell}{\ell!} + m e^{-m} \sum_{\ell=0}^{\infty} \frac{m^\ell}{\ell!}$$
$$= m^2 + m.$$

Then the variance of X is $\mathbb{E}X^2 - (\mathbb{E}X)^2 = m$, so the standard deviation is $\sqrt{m}$. Also

$$\mathbb{E}|X - \mathbb{E}X| \le \big(\mathbb{E}|X - \mathbb{E}X|^2\big)^{1/2} = \sqrt{m} \tag{3.76}$$

so

$$\mathbb{E}|X - \mathbb{E}X| = \sum_{k=0}^{\infty} |k-m| \mathbb{P}[X = k] = e^{-m} \sum_{k=0}^{\infty} \frac{|k-m| m^k}{k!}. \tag{3.77}$$

is bounded above by $\sqrt{m}$.

We choose $\alpha = t/m > 0$ and observe that by (vi),

$$\left\| \Big(I - \frac{t}{m} A\Big)^{-m} v \right\| \le \|v\|. \tag{3.78}$$

To prove the limit formula, we introduce $L = (I - tA/m)^{-1}$, so that $\|L\| \le 1$ by (vi), and $L - I = (tA/m)(I - tA/m)^{-1}$, so

$$m(L - I) = tA(I - tA/m)^{-1} \to tA \tag{3.79}$$

as $m \to \infty$, hence one can check that $\exp(m(L - I)) \to \exp(tA)$ as $m \to \infty$.

We have

$$\begin{aligned} \exp\big(m(L-I)\big) - L^m &= e^{-m}\big(\exp(mL) - e^m L^m\big) \\ &= e^{-m}\Big(\sum_{k=0}^{\infty} \frac{m^k L^k}{k!} - \sum_{k=0}^{\infty} \frac{L^m m^k}{k!}\Big) \\ &= e^{-m} \sum_{k=0}^{\infty} \frac{m^k (L^k - L^m)}{k!}. \end{aligned}$$

Now for $k > m$, we have

$$L^k - L^m = (L^{k-m} - I)L^m = (L - I)(L^{k-m-1} + L^{k-m-2} + \cdots + I)L^m$$

so

$$\|L^k - L^m\| \le \|L - I\|\big(\|L^{k-m-1}\| + \cdots + \|I\|\big)\|L^m\| \le (k-m)\|\|L - I\| \tag{3.80}$$

and likewise for $k < m$, we have

$$\|L^k - L^m\| \le (m-k)\|L - I\|. \tag{3.81}$$

Hence we have

$$\Big\| \exp\big(m(L-I)\big) - L^m \Big\| \le e^{-m} \sum_{k=0}^{\infty} \frac{m^k \|L^m - L^k\|}{k!} \le \|L - I\| e^{-m} \sum_{k=0}^{\infty} \frac{m^k |m-k|}{k!} \tag{3.82}$$

which by the example of Poisson random variables is

$$\begin{aligned} \Big\| \exp\big(m(L-I)\big) - L^m \Big\| &\le \sqrt{m}\|L - I\| \\ &= \sqrt{m}\Big\| \frac{tA}{m}\Big(I - \frac{tA}{m}\Big)^{-1} \Big\| \\ &\le \frac{t\|A\|}{\sqrt{m}} \end{aligned}$$

Hence

$$\Big(I - \frac{t}{m}A\Big)^{-m} - \exp\Big(tA\Big(I - \frac{tA}{m}\Big)\Big) \to 0 \qquad (m \to \infty). \tag{3.83}$$

□

Proposition 3.37 *(i) Suppose that $a, d, c, d \in \mathbb{R}$ with $ad - bc > 0$.*

(i) Then

$$\varphi(s) = \frac{as - ib}{ics + d} \qquad (ics + d \ne 0) \tag{3.84}$$

takes RHP onto RHP.

(ii) If A has all its eigenvalues in LHP, then $\varphi(A)$ also has all its eigenvalues in LHP.

(iii) If A is strictly dissipative, then $\varphi(A)$ is also strictly dissipative.

Proof

(i) Consider $\psi(s) = \frac{as+b}{cs+d}$, which maps the upper half plane onto itself. Then $\varphi(s) = -i\psi(is)$, and this has the effect of rotating the left half-plane through $\pi/2$ to the upper half plane, transforming the upper half plane by ψ, then rotating the upper half plane back to the left half-plane. The inverse function is

$$\varphi^{-1}(s) = \frac{ds + ib}{-ics + a} \qquad (-ics + a \ne 0). \tag{3.85}$$

(ii) Let λ be an eigenvalue of A in LHP with corresponding eigenvector v, so $Av = \lambda v$. This equation leads to $(aA - ibI)v = (a\lambda - ib)v$ and $(icA + dI)v = (ic\lambda + d)v$, so $\varphi(A)v = \varphi(\lambda)v$, so $\varphi(A)$ has eigenvalues in the left half-plane.

(iii) We consider

$$\begin{aligned}
&\varphi(A) + \varphi(A)' \\
&= (icA + dI)^{-1}(aA - ibI) + (aA' + ibI)(-icA' + dI)^{-1} \\
&= (icA + dI)^{-1}\Big((aA - ibI)(-icA' + dI) \\
&\quad + (icA + dI)(aA' + ibI)\Big)(-icA' + dI)^{-1} \\
&= (icA + dI)^{-1}(ad - bc)(A + A')(-icA' + dI)^{-1}
\end{aligned}$$

in which $ad - bc > 0, icA + dI$ is invertible and $-(A + A')$ is positive definite, so $-(\varphi(A) + \varphi(A)')$ is also positive definite.

□

3.11 A Determinant Formula

Lemma 3.38 *Let A and B be $(n \times n)$ complex matrices.*

(i) Then the characteristic polynomials of AB and BA are equal

$$\det(sI - AB) = \det(sI - BA). \tag{3.86}$$

and the eigenvalues of AB are equal to the eigenvalues of BA.

(ii) For $s \neq 0$ such that $\det(sI - AB) \neq 0$, *the inverses satisfy*

$$(sI - AB)^{-1} = s^{-1}\big(I + A(sI - BA)^{-1}B\big). \tag{3.87}$$

Proof

(i) We consider the $(2n \times 2n)$ matrices

$$X = \begin{bmatrix} I & A \\ 0 & I \end{bmatrix}, Y = \begin{bmatrix} sI & -A \\ -B & I \end{bmatrix}, Z = \begin{bmatrix} I & A/s \\ 0 & I \end{bmatrix}, \tag{3.88}$$

with products

$$XY = \begin{bmatrix} I & A \\ 0 & I \end{bmatrix}\begin{bmatrix} sI & -A \\ -B & I \end{bmatrix} = \begin{bmatrix} sI - AB & 0 \\ -B & I \end{bmatrix} \tag{3.89}$$

so

$$\det\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \det\begin{bmatrix} sI & -A \\ -B & I \end{bmatrix} = \det\begin{bmatrix} sI - AB & 0 \\ -B & I \end{bmatrix} \tag{3.90}$$

hence

$$\det\begin{bmatrix} sI & -A \\ -B & I \end{bmatrix} = \det(sI - AB); \tag{3.91}$$

also

$$YZ = \begin{bmatrix} sI & -A \\ -B & I \end{bmatrix} \begin{bmatrix} I & A/s \\ 0 & I \end{bmatrix} = \begin{bmatrix} sI & 0 \\ -B & I - BA/s \end{bmatrix}, \tag{3.92}$$

$$\det\begin{bmatrix} sI & -A \\ -B & I \end{bmatrix} \det\begin{bmatrix} I & A/s \\ 0 & I \end{bmatrix} = \det\begin{bmatrix} sI & 0 \\ -B & I - BA/s \end{bmatrix} \tag{3.93}$$

$$\det\begin{bmatrix} sI & -A \\ -B & I \end{bmatrix} = s^n \det(I - BA/s) = \det(sI - BA); \tag{3.94}$$

hence by combining (3.91) and (3.94), we obtain

$$\det(sI - AB) = \det(sI - BA). \tag{3.95}$$

Hence λ is an eigenvalue of AB if and only if both of these are zero as $s = \lambda$, or equivalently λ is an eigenvalue of BA.

(ii) We have

$$\begin{aligned}
(sI - AB)s^{-1}\big(I + A(sI - BA)^{-1}B\big) \\
&= s^{-1}\big(sI - AB + (sI - AB)A(sI - BA)^{-1}B\big) \\
&= s^{-1}\big(sI - AB + A(sI - BA)(sI - BA)^{-1}B\big) \\
&= s^{-1}\big(sI - AB + AB\big) = I,
\end{aligned} \tag{3.96}$$

and similarly

$$s^{-1}\big(I + A(sI - BA)^{-1}B\big)(sI - AB) = I. \tag{3.97}$$

We can also swap A and B in these formulas.

□

3.12 Observability and Controllability

See [25]. Let (A, B, C, D) be a MIMO where $A \in M_{n\times n}(\mathbb{C})$, with transfer function $T(s) = D + C(sI - A)^{-1}B$, which we expand as

$$T(s) = D + \sum_{k=0}^{\infty} \frac{CA^k B}{s^{k+1}} \qquad (|s| > \|A\|). \tag{3.98}$$

This suggests that the coefficients CA^kB should contain useful information about the linear system. In this section we study the vector spaces $\mathrm{span}\{A^jB : j = 0, \dots, n-1\}$ and $\mathrm{span}\{CA^j : j = 0, \dots, n-1\}$, and use them to obtain decompositions of the state space.

Let (A, B, C, D) be a SISO, and suppose that A is $n \times n$ and C is $1 \times n$. Then we introduce the $n \times 1$ complex matrix

$$L = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}. \tag{3.99}$$

Proposition 3.39 (Observability) *The following conditions are equivalent:*

(i) *span*$\{C, CA, \dots, CA^{n-1}\} = \mathbb{C}^{1\times n}$*;*
(ii) *If* $CA^jv = 0$ *for* $j = 0, \dots, n-1$*, then* $v = 0$*;*
(iii) *rank*$(L) = n$*;*
(iv) *the observability Gramian* $L'L$ *is positive definite.*

Proof $(ii) \Leftrightarrow (iv)$ We observe that

$$L'L = \begin{bmatrix} C' & A'C' & \dots (A')^{(n-1)}C' \end{bmatrix} \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix} = \sum_{j=0}^{n-1} (A')^j C' C A^j \tag{3.100}$$

so that $L'L$ is positive definite if and only if $\sum_{j=0}^{n-1} \|CA^jv\|^2 > 0$ for all $v \neq 0$, which is equivalent to (i).

$(iii) \Leftrightarrow (iv)$ We observe that $\|Lv\|^2 = \langle L'Lv, v\rangle$, so the null space of $L'L$ is equal to the null space of L, hence (iv) is equivalent to the nullity of L being zero. But $\mathrm{rank}(L) + \mathrm{nullity}(L) = n$, so (iv) is equivalent to the $\mathrm{rank}(L) = n$.

$(i) \Leftrightarrow (iii)$. We observe that (i) is equivalent to the statement that span$\{C, CA, \dots, CA^{n-1}\}$ has dimension equal to n, which is equivalent to (iii). □

Definition 3.40 (Observability) A linear system (A, B, C, D) that satisfies the equivalent conditions of Proposition 3.39 is called observable.

Remark 3.41 The terminology observable refers to ability to observe an initial state of the system via the output. See [11]. Clearly only A and C are involved in the conditions. By the Cayley–Hamilton Theorem 2.29, all the vectors CA^k belong to span$\{C, CA, \dots, CA^{n-1}\}$, so we only need to consider the first n such expressions. Condition (iii) is convenient for computer calculation, as one can find the echelon form of L. Condition (iv) brings us to the familiar criteria for positive definiteness of a finite matrix, as in Theorem 3.23.

Now let $K : \mathbb{C}^n \to \mathbb{C}^n$ be the linear transformation with matrix

$$K = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix}. \tag{3.101}$$

Proposition 3.42 *For a SISO (A, B, C, D) consider range(K) for K as above.*

(i) Then range$(K) = \{0\}$ if and only if $B = 0$.
(ii) range(K) has dimension one, if and only if B is an eigenvector of A.
(iii) range(K) has dimension n, if and only if K has rank n.

Proof We observe that

$$\text{range}(K) = \Big\{ \sum_{j=1}^{n} a_j A^{j-1} B;\ (a_j)_{j=1}^n \in \mathbb{C}^n \Big\} \tag{3.102}$$

$$= \text{span}\{A^j B;\ j = 0, \dots\} \tag{3.103}$$

is the column space of K. Then one can consider the various cases.

(i) If $B = 0$, then $K = 0$, The converse is clear.
(ii) If B is an eigenvector, then $AB = \lambda B$ for λ the eigenvalue, where $B \neq 0$ by definition of eigenvector, hence $A^j B = \lambda^j B$ for $j = 1, \dots, n$, so the column space of K has basis (B). Conversely, if the column space of K has dimension one, then $B \neq 0$ by (i) and K is spanned by B. Hence AB is a multiple of B, so B is an eigenvector of A.
(iii) The rank of K is the dimension of the column space of K, hence result.

□

Proposition 3.43 (Controllability) *The following conditions are equivalent:*

(i) span$\{B, AB, \dots, A^{(n-1)}B\} = \mathbb{C}^{n\times 1}$;
(ii) span$\{B', B'A', \dots, B'(A')^{(n-1)}\} = \mathbb{C}^{1\times n}$;
(iii) If $B'(A')^j v = 0$ for $j = 0, \dots, n-1$, then $v = 0$;
(iv) rank$(K') = n$;
(v) the controllability Gramian KK' is positive definite.

Proof $(i) \Leftrightarrow (ii)$ follows by taking the adjoint, which clearly does not change the dimension of a vector space.

The remainder of the proof follows by replacing (A, C) by (A', B'), and L by K'. □

Corollary 3.44 *Suppose that A has characteristic polynomial $s^n + \sum_{j=0}^{n-1} a_j s^j$, and that $V = span\{B, AB, \dots, A^{n-1}B\}$ has dimension n. Then V has (non-orthogonal) basis $(e_j)_{j=1}^n$ where $e_j = A^{j-1}B$ for $j = 1, \dots, n$, and*

$$Ae_j = e_{j+1}, \quad (j = 1, \dots, n-1); \quad Ae_n = -\sum_{j=1}^{n-1} a_{j-1} e_j, \tag{3.104}$$

so A gives a linear transformation of V such that A is similar to

$$\begin{bmatrix} 0 & 0 & \dots & -a_0 \\ 1 & 0 & \dots & \vdots \\ 0 & \ddots & \ddots & -a_{n-2} \\ 0 & \dots & 1 & -a_{n-1} \end{bmatrix}, \tag{3.105}$$

which is the transpose of a companion matrix.

Definition 3.45 (Controllable)

(i) A linear system (A, B, C, D) that satisfies the equivalent conditions of Proposition 3.43 is called controllable.
(ii) The controllability space of (A, B) is $\text{span}\{A^{n-1}B, \dots, AB, B\}$.

Remark 3.46

(1) The terminology controllable refers to ability to attain any state of the system from the input, and thereby control the states of the system. Here only A and B are involved in the conditions.
(2) The discussion shows that (A, B, C, D) is observable if and only if (A', C', B', D') is controllable.
(3) We also observe that

$$LK = \begin{bmatrix} CB & CAB & CA^2B & \dots & CA^{n-1}B \\ CAB & CA^2B & \dots & \dots & CA^nB \\ CA^2B & CA^3B & \dots & & \\ \vdots & \dots & \dots & \dots & \\ CA^{n-1}B & \dots & \dots & \dots & CA^{2n-2}B \end{bmatrix} \tag{3.106}$$

is a matrix which is constant on cross diagonals. This is called a finite Hankel matrix, and the entries are the coefficients of the power series $T(s) = D + \sum_{k=0}^{\infty} CA^k B/s^{k+1}$.

Theorem 3.47 (Popov–Belevitch–Hautus Test for Controllability) *Let A be a $n \times n$ complex matrix and B be a $n \times 1$ column matrix; let*

$$V = \Big\{ \sum_{j=0}^{n-1} a_j A^j B; a_j \in \mathbb{C}; j = 0, \dots, n-1 \Big\}. \tag{3.107}$$

Then the following conditions are equivalent:

(i) $V = \mathbb{C}^{n\times 1}$;
(ii) the nullspace of $B^\top$ *contains no eigenvectors of* $A^\top$;
(iii) the rank of $[A - \lambda I \quad B]$ *equals* n *for all* $\lambda \in \mathbb{C}$;
(iv) the rank of Q *equals* n, *where* $Q = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix}$.

Proof (not (iii) implies not (ii)) Suppose that there exists $\lambda \in \mathbb{C}$ such that $[A - \lambda I \; B]$ has rank k where $k < n$. Then by the rank-nullity theorem 2.2, the nullspace of

$$\begin{bmatrix} A^\top - \lambda I \\ B^\top \end{bmatrix} \tag{3.108}$$

has dimension $n - k > 0$, so there exists on non-zero $y \in \mathbb{C}^{n\times 1}$ in the nullspace; hence

$$(A^\top - \lambda I)y = 0, \quad B^\top y = 0; \tag{3.109}$$

thus the nullspace of $B^\top$ contains an eigenvector of $A^\top$, namely y.

(not (ii) implies not (i)) Suppose that $Y \in \mathbb{C}^{n\times 1}$ satisfies

$$A^\top y = \lambda y, \quad B^\top y = 0, \quad y \neq 0. \tag{3.110}$$

Then

$$\sum_{j=0}^{n-1} a_j B^\top (A^\top)^j y = \sum_{j=0}^{n-1} a_j \lambda^j B^\top y = 0 \tag{3.111}$$

for all $a_j \in \mathbb{C}$. Hence y is perpendicular to V, so V is a strictly proper subspace of $\mathbb{C}^{n\times 1}$.

(not (i) implies not (iii)) Suppose that V is a strictly proper subspace of $\mathbf{C}^{n\times 1}$, so there exists a nonzero $y \in \mathbb{C}^{n\times 1}$ such that

$$\sum_{j=0}^{n-1} a_j B^\top (A^\top)^j y = 0. \tag{3.112}$$

Then the subspace $W = \{\sum_{j=0}^{n-1} a_j (A^\top)^j y\}$ is non zero and mapped to itself by left multiplication by $A^\top$, hence W contains an eigenvector z of $A^\top$. Now $A^\top z = \lambda z$ for some λ and $B^\top z = 0$. Hence z is in the nullspace of $B^\top$ and is an eigenvector of $A^\top$. So

$$\begin{bmatrix} A^\top - \lambda I \\ B^\top \end{bmatrix} z = 0 \tag{3.113}$$

so by the rank-nullity theorem 2.2, the rank of $[A - \lambda I \; B]$ is less than n.

((i) is equivalent to (iv)) We have

$$\begin{aligned} V &= \Big\{ \sum_{j=0}^{n-1} a_j A^j B; a_j \in \mathbb{C}; j = 0, \dots, n-1 \Big\} \\ &= \Big\{ \big[B \; AB \; A^2B \; \dots \; A^{n-1}B\big] \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} : a_j \in \mathbb{C}; j = 0, \dots, n-1 \Big\} \\ &= \Big\{ Q \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} : a_j \in \mathbb{C}; j = 0, \dots, n-1 \Big\} \end{aligned} \tag{3.114}$$

so V equals the range of Q; hence $\dim V = \operatorname{rank}(Q)$. We deduce that $V = \mathbb{C}^{n\times 1}$ if and only if $\dim V = \operatorname{rank}(Q) = n$.

In condition (iii) we consider the vector space spanned by the rows of $[A - \lambda I \;\; B]$, which depends upon $\lambda \in \mathbb{C}$, and its dimension could possibly depend upon λ. Clearly $\operatorname{rank}[A - \lambda I \;\; B] = n$ for large $|\lambda|$, and $\operatorname{rank}[A - \lambda I] < n$ when λ is an eigenvalue of A. The condition states that, nevertheless, $\operatorname{rank}[A - \lambda I \;\; B] = n$ is all cases. □

3.13 Kalman's Decomposition

Before considering the general result, we look at a special case.

Proposition 3.48 *Let* (A, B, C, D) *be a SISO.*

(i) Then the state space $\mathbb{C}^n$ *has an orthogonal decomposition*

$$\mathbb{C}^{n\times 1} = \mathit{span}\{A^j B : j = 0, 1, \dots\} \oplus \{X \in \mathbb{C}^{n\times 1} : B'(A')^j X = 0; j = 0, 1, \dots\}. \tag{3.115}$$

(ii) Then A is similar to a block matrix with respect to this decomposition

$$\begin{bmatrix} A_1 & A_2 \\ 0 & A_4 \end{bmatrix}. \tag{3.116}$$

Proof

(i) With the $n \times n$ matrix

$$K = \left[B \ AB \ A^2B \ \ldots \ A^{n-1}B\right], \tag{3.117}$$

we have an orthogonal decomposition

$$\mathbb{C}^{n\times 1} = \text{range}(K) \oplus \text{null}(K'), \tag{3.118}$$

where

$$\begin{aligned} \text{range}(K) &= \{K(a_j)_{j=1}^n : (a_j)_{j=1}^n \in \mathbb{C}^{n\times 1}\} \\ &= \left\{ \sum_{j=1}^n a_j A^{j-1} B : (a_j)_{j=1}^n \in \mathbb{C}^{n\times 1} \right\} \\ &= \text{span}\{A^j B : j = 0, \ldots, n-1\} \\ &= \text{span}\{A^j B : j = 0, 1 \ldots, \} \end{aligned} \tag{3.119}$$

where the final step follows from the Cayley-Hamilton theorem 2.29.
Likewise, we have

$$\begin{aligned} \text{null}(K') &= \left\{X \in \mathbb{C}^{n\times 1} : K'X = 0\right\} \qquad (3.120) \\ &= \left\{X \in \mathbb{C}^{n\times 1} : \langle X, K(a_j)_{j=1}^n \rangle = 0 : (a_j)_{j=1}^n \in \mathbb{C}^{n\times 1}\right\} \\ &= \left\{X \in \mathbb{C}^{n\times 1} : \sum_{j=1}^n a_j \langle X, A^{j-1}B \rangle = 0 : (a_j)_{j=1}^n \in \mathbb{C}^{n\times 1}\right\} \\ &= \left\{X \in \mathbb{C}^{n\times 1} : \langle X, A^{j-1}B \rangle = 0 : j = 1, \ldots, n\right\} \\ &= \left\{X \in \mathbb{C}^{n\times 1} : \langle X, A^{j-1}B \rangle = 0 : j = 1, 2, \ldots\right\} \qquad (3.121) \end{aligned}$$

where the final step follows from the Cayley-Hamilton theorem 2.29.

(ii) The subspace range(K) is evidently invariant under A, so we can choose bases of range(K) and null(K') so that A can be expressed as a matrix of the stated block form.

Consider a linear system represented by (A, B, C, D) where A is $n \times n$ and B is $n \times k$. The basic idea is to introduce a basis for $\mathbb{C}^{n\times 1}$ which is especially adapted to (A, B, C) such that the linear systems has a special block form. To deal with (A, B), we let V be the linear span of the columns in $\begin{bmatrix} A^{n-1}B \ A^{n-2}B \ \ldots AB \ B \end{bmatrix}$, so that V is a linear subspace of $\mathbb{C}^{n\times 1}$ which is invariant under left multiplication by A. Suppose that V has dimension $\ell < n$; then we choose a basis $\{v_1, \ldots, v_\ell\}$ of V, and extend to a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{C}^{n\times 1}$, thereby introducing a complementary space W spanned by $\{v_{\ell+1}, \ldots, v_n\}$. We can find these bases by carrying out elementary column operations to find echelon forms. There is an invertible linear transformation S on $\mathbb{C}^{n\times 1}$ determined by $Sv_j = e_j$ where $\{e_1, \ldots, e_n\}$ is the standard basis. Then in term of $\{v_1, \ldots, v_n\}$, we have an upper triangular block form

$$S^{-1}AS = \begin{bmatrix} A_1 & A_2 \\ 0 & A_4 \end{bmatrix}, S^{-1}B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, CS = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \tag{3.122}$$

and as similarity does not change the transfer matrix, we have

$$\begin{aligned} T(s) &= D + C(sI - A)^{-1}B \\ &= D + \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} sI - A_1 & -A_2 \\ 0 & sI - A_4 \end{bmatrix}^{-1} \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \\ &= D + C_1(sI - A_1)^{-1}B_1. \end{aligned} \tag{3.123}$$

We now consider $A_1 : V \to V$ and its transpose $A_1^\top : V \to V$, and let V_1 be the subspace of V that is spanned by the columns of

$$\begin{bmatrix} A_1^{\top,\ell-1}C_1^\top \ A_1^{\top,\ell-2}C_1^\top \ \ldots A_1^\top C_1^\top \ C_1^\top \end{bmatrix}. \tag{3.124}$$

Then V_1 is a subspace of V that is invariant under left multiplication by $A_1^\top$, so as before, V_1 has a complementary subspace W_1 in V, and we can introduce an invertible transformation $S_1 : V \to V$ such that the upper triangular block form

$$S_1^{-1}A_1^\top S_1 = \begin{bmatrix} A_{1,1}^\top & A_{2,1}^\top \\ 0 & A_{2,2}^\top \end{bmatrix}, S_1^{-1}C_1^\top = \begin{bmatrix} C_{1,1}^\top \\ 0 \end{bmatrix}, B_1^\top S_1 = \begin{bmatrix} B_{1,1}^T & B_{1,2}^\top \end{bmatrix}, \tag{3.125}$$

which transposes to a lower triangular block form

$$S_1^\top A_1 S_1^{-\top} = \begin{bmatrix} A_{1,1} & 0 \\ A_{2,1} & A_{2,2} \end{bmatrix}, S_1^\top B_1 = \begin{bmatrix} B_{1,1} \\ B_{1,2} \end{bmatrix}, C_1 S_1^{-\top} = \begin{bmatrix} C_{1,1} & 0 \end{bmatrix}, \tag{3.126}$$

Hence we can further reduce the transfer function

$$\begin{aligned} T(s) &= D + C_1(sI - A_1)^{-1}B_1 \\ &= D + [C_{1,1}\ 0]\begin{bmatrix} sI - A_{1,1} & 0 \\ -A_{2,1} & sI - A_{2,2} \end{bmatrix}^{-1}\begin{bmatrix} B_{1,1} \\ B_{1,2} \end{bmatrix} \\ &= D + C_{1,1}(sI - A_{1,1})^{-1}B_{1,1}. \end{aligned}$$

The full Kalman decomposition of the linear system is

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & A_{13} & 0 \\ A_{21} & A_{22} & A_{23} & A_{24} \\ 0 & 0 & A_{33} & 0 \\ 0 & 0 & A_{43} & A_{44} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ 0 \\ 0 \end{bmatrix} \tag{3.127}$$

$$y = \begin{bmatrix} C_1 & 0 & C_3 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + Du \tag{3.128}$$

for a suitable basis of $\mathbb{C}^{n\times 1}$. The basis can be found by elementary row and column operations, as above. □

3.14 Kronecker Product of Matrices

Let $(e_{jk})_{j=1,k=1}^{r,s}$ of $M_{r,s}(\mathbb{C})$ such that

$$\sum_{j=1}^{r}\sum_{k=1}^{s} a_{j,k}e_{j,k} = \begin{bmatrix} a_{1,1} & \dots & a_{1,s} \\ \vdots & \ddots & \vdots \\ a_{r,1} & \dots & a_{r,s} \end{bmatrix} \tag{3.129}$$

where $a_{j,k} \in \mathbb{C}$. Then for $A_{j,k} \in M_{p,q}(\mathbb{C})$, we form the block matrix

$$\sum_{j=1}^{r}\sum_{k=1}^{s} A_{j,k} \otimes e_{j,k} = \begin{bmatrix} A_{1,1} & \dots & A_{1,s} \\ \vdots & \ddots & \vdots \\ A_{r,1} & \dots & A_{r,s} \end{bmatrix} \tag{3.130}$$

with r block rows and s block columns, so the block matrix belongs to $M_{pr,qs}(\mathbb{C})$. This defines the Kronecker product $M_{p,q}(\mathbb{C}) \otimes M_{r,s}(\mathbb{C}) = M_{pr,qs}(\mathbb{C})$, with multiplication of blocks. We also take scalars across the tensor $\otimes$ symbol, so

$$\sum_{j=1}^{r}\sum_{k=1}^{s} A_{j,k} \otimes x_{j,k}e_{j,k} = \sum_{j=1}^{r}\sum_{k=1}^{s} x_{j,k}A_{j,k} \otimes e_{j,k} = \begin{bmatrix} x_{1,1}A_{1,1} & \dots & x_{1,s}A_{1,s} \\ \vdots & \ddots & \vdots \\ x_{r,1}A_{r,1} & \dots & x_{r,s}A_{r,s} \end{bmatrix}. \tag{3.131}$$

Given a linear map $\phi : M_{p,q}(\mathbb{C}) \to (\mathbb{C})$ there exists a unique linear map

$$\Phi : M_{p,q}(\mathbb{C}) \otimes M_{r,s}(\mathbb{C}) \to M_{r,s}(\mathbb{C}) : \begin{bmatrix} A_{1,1} & \dots & A_{1,s} \\ \vdots & \ddots & \vdots \\ A_{r,1} & \dots & A_{r,s} \end{bmatrix} \mapsto \begin{bmatrix} \phi(A_{1,1}) & \dots & \phi(A_{1,s}) \\ \vdots & \ddots & \vdots \\ \phi(A_{r,1}) & \dots & \phi(A_{r,s}) \end{bmatrix} \tag{3.132}$$

obtained by applying ϕ to the blocks in the matrix.

3.15 Exercises

Exercise 3.1 (Hadamard Matrices) The Hadamard matrices have applications in signal processing. This exercise gives the construction for matrices of size $2^n \times 2^n$.

(i) Let H be an $n \times n$ matrix. Show that

$$H \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} H & H \\ H & -H \end{bmatrix}. \tag{3.133}$$

(ii) Let

$$H_0 = 1, \quad H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad H_2 = H_1 \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \dots, H_{n+1} = H_n \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{3.134}$$

(iii) Show that H_n has size $2^n \times 2^n$, that $H_n^T = H_n$ and $H_n H_n^T = 2^n I_{2^n}$.
(iv) Show that all the entries of H_n are in $\{\mp 1\}$.

Exercise 3.2 A complex square matrix is stable if all the eigenvalues λ have $\Re\lambda < 0$, where $\Re\lambda$ is the real part of λ. For each of the following matrices, find the eigenvalues numerically using computer software to test whether $\pm A$, $\pm B$, $\pm C$

are stable:

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 7 & 5 \\ 1 & 8 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 7 \\ 9 & 8 & 4 \\ 2 & 2 & 9 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 7 & 9 & 4 \\ 8 & 1 & 7 & \imath \\ 2 & 2\imath & 2 & 4 \end{bmatrix}. \tag{3.135}$$

Exercise 3.3 Let

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & -2 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}; \tag{3.136}$$

let Q be the (4×8) matrix written as (4×2) blocks

$$Q = \begin{bmatrix} B & AB & A^2B & A^3B \end{bmatrix}. \tag{3.137}$$

Find the rank of Q.

Exercise 3.4 Let A be a $(n \times n)$ complex matrix and C be a $(1 \times n)$ complex matrix; then let

$$Q = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}, \quad R = \begin{bmatrix} C^\top & A^\top C^\top & \dots & (A^\top)^{n-1} C^\top \end{bmatrix}, \tag{3.138}$$

which are $n \times n$ complex matrices, and $C^\top$ is the transpose of C.

(i) Show that $Q^\top = R$ and $\mathrm{rank}(Q) = \mathrm{rank}(R)$.
(ii) Suppose that $C^\top$ is an eigenvector of $A^\top$. Find R, and compute $\mathrm{rank}(R)$.
(iii) Find the rank of Q when

$$A = \begin{bmatrix} 1 & 3 & 5 & 0 \\ 0 & 1 & 9 & 6 \\ 1 & 1 & 4 & -7 \\ 2 & 2 & 1 & 8 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -5 & -1/2 & 3 \end{bmatrix}. \tag{3.139}$$

Exercise 3.5

(i) Let D be a $(n \times n)$ diagonal matrix with positive diagonal entries $\kappa_1 \ge \kappa_2 \ge \dots \ge \kappa_n$. Show that

$$\kappa_1 \|X\|^2 \ge \langle DX, X \rangle \ge \kappa_n \|X\|^2 \qquad (X \in \mathbb{R}^{n \times 1}). \tag{3.140}$$

(ii) Let K be a $(n \times n)$ real symmetric matrix with positive eigenvalues $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_n$. Show that

$$\kappa_1 \|X\|^2 \geq \langle KX, X\rangle \geq \kappa_n \|X\|^2 \qquad (X \in \mathbb{R}^{n\times 1}). \tag{3.141}$$

Exercise 3.6 The Lorenz system is

$$\begin{aligned} \frac{dx}{dt} &= -\sigma x + \sigma y \\ \frac{dy}{dt} &= \rho x - y - xz \\ \frac{dz}{dt} &= -\beta z + xy. \end{aligned}$$

where ρ, β and σ are real constants. A linear version of this system is

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\sigma & \sigma & 0 \\ \rho & -1 & 0 \\ 0 & 0 & -\beta \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{3.142}$$

Find the eigenvalues of this matrix

$$A = \begin{bmatrix} -\sigma & \sigma & 0 \\ \rho & -1 & 0 \\ 0 & 0 & -\beta \end{bmatrix}, \tag{3.143}$$

and state conditions on ρ, β and σ for all the eigenvalues λ to lie in $\{\lambda \in \mathbb{C} : \Re\lambda < 0\}$.

Exercise 3.7 Consider $(n \times n)$ matrices A, S, K, L. Let K be a positive definite matrix.

(i) Show that if λ is an eigenvalue of K, then $\lambda > 0$.
(ii) Deduce that $\det K > 0$ and $\operatorname{trace}(K) > 0$.
(iii) Let S an invertible matrix. Show that $S'KS$ is also positive definite.
(iv) Deduce that $\exp(A')K\exp(A)$ is also positive definite.
(v) Suppose that L is positive definite. Show that $K + L$ is also positive definite.

Exercise 3.8 Let U be a nonzero proper subspace of a finite-dimensional vector space V, and $T : V \to V$ a linear transformation.

(i) Show that T maps U into itself, if and only if T has the block form

$$T = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \qquad \begin{matrix} U \\ W \end{matrix} \tag{3.144}$$

with respect to a suitable basis of $V = U \oplus W$, where W is a complementary subspace of U.

(ii) Show that such a T is invertible, if and only if A and D are both invertible, in which case

$$T^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}. \tag{3.145}$$

Exercise 3.9

(i) Let B be $n \times m$ complex matrix, such that $\|B\| \leq 1$. Show that this condition is equivalent to the condition that $I - B'B$ is a self-adjoint $m \times m$ matrix with nonnegative eigenvalues.

(ii) By considering the binomial series

$$C = \sum_{k=0}^{\infty} (-1)^k \binom{1/2}{k} (B'B)^k \tag{3.146}$$

show that there exists a self-adjoint $m \times m$ matrix C with nonnegative eigenvalues such that $C^2 = I - B'B$.

(iii) Show that $\|B'\| \leq 1$. Deduce that there exists a self-adjoint $n \times n$ matrix D with nonnegative eigenvalues such that $D^2 = I - BB'$.

(iv) Deduce that

$$U = \begin{bmatrix} B & D \\ -C & B' \end{bmatrix} \tag{3.147}$$

satisfies $UU' = I$.

This exercise shows that a sub-block of a unitary matrix is equivalent to a matrix B of norm less than or equal to one.

Exercise 3.10 Let $A_1, A_2 \in \mathcal{D}_n$. Without assuming that A_1 and A_2 commute, show that

$$\left(\left(I - \frac{t}{m}A_1\right)\left(I - \frac{t}{m}A_2\right)\right)^{-m} \to \exp(t(A_1 + A_2)) \qquad (m \to \infty), \qquad (t > 0). \tag{3.148}$$

Exercise 3.11 Consider $\mathbb{C} \cup \{\infty\}$ with the interpretation that $z \to \infty$ means $|z| \to \infty$.

(i) Show that for distinct $z_1, z_2, z_3 \in \mathbb{C} \cup \{\infty\}$, there exists a $L \in \mathcal{C}$ that passes through z_1, z_2 and z_3.

(ii) Show there exists a linear fractional transformation φ such that

$$\varphi(z_1) = 0, \quad \varphi(z_2) = \infty, \quad \varphi(z_3) = 1.$$

(iii) Deduce that given $L_1, L_2 \in \mathcal{C}$, there exist a linear fractional transformation ϕ such that $\phi(L_1) = L_2$.
(iv) Deduce that the group of linear fractional transformations acts transitively on $\mathcal{C}$.

Exercise 3.12 Let

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{2\times 2}(\mathbb{C}) \tag{3.149}$$

have $ad - bc \neq 0$, and let φ_M be the corresponding linear fractional transformation.

(i) Show that for $\lambda \in \mathbb{C} \setminus \{0\}$, the matrices M and λM give the same linear fractional transformation.
(ii) Show that there exists $\lambda \in \mathbb{C} \setminus \{0\}$ such that $\det(\lambda M) = 1$.
(iii) Suppose that $M \in M_{2\times 2}(\mathbb{R})$ has $\det M \neq 0$. Show that there exists $\lambda \in \mathbb{R} \setminus \{0\}$ such that $\det(\lambda M) \in \{1, -1\}$.

Exercise 3.13 (Controllability and Block Matrices) Let $A_1 \in M_{k\times k}(\mathbb{C})$, $A_2 \in M_{m\times m}(\mathbb{C})$ and $B_1 \in \mathbb{C}^{k\times 1}$, $B_2 \in \mathbb{C}^{m\times 1}$, then form the block matrices

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}. \tag{3.150}$$

(i) Find $L \in M_{k\times(m+k)}(\mathbb{C})$ and $M \in M_{m\times(m+k)}(\mathbb{C})$ such that

$$K = \begin{bmatrix} A^{n-1}B \ \dots \ AB \ B \end{bmatrix} = \begin{bmatrix} L \\ M \end{bmatrix}. \tag{3.151}$$

(ii) By considering $K^\top$, show that

$$\operatorname{rank}(K) = \operatorname{rank}(L) + \operatorname{rank}(M) - \dim(\operatorname{range}(L^\top) \cap \operatorname{range}(M^\top)). \tag{3.152}$$

(iii) Deduce a formula relating the dimension of the controllability space of (A, B) to the dimensions of the controllability spaces of (A_1, B_1) and (A_2, B_2).
(iv) Show that the controllability space of (A, B) has dimension $m + k$ if and only if

$$KK' = \begin{bmatrix} LL' & LM' \\ ML' & MM' \end{bmatrix} \tag{3.153}$$

is positive definite.

Exercise 3.14

(i) Suppose that $T(s) = D + C(sI - A)^{-1}B$ is invertible. Show that

$$T(s)^{-1} = \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} sI - A & B \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ I \end{bmatrix}. \tag{3.154}$$

(ii) (Higman's trick) Discuss the validity of the formula

$$\big[[1\ 0]\big] \begin{bmatrix} f & a \\ -b & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{f + ab}. \tag{3.155}$$

Exercise 3.15 Let $A \in M_{n\times n}(\mathbb{C})$ have eigenvalue λ and $B \in M_{m\times m}(\mathbb{C})$ have eigenvalue μ.

(i) Show that $A \otimes B$ has eigenvalue $\lambda\mu$.
(ii) Show that $A \otimes I_m + I_n \otimes B$ has eigenvalue $\lambda + \mu$.
(iii) Show that

$$\exp\big(t(A \otimes I_m + I_n \otimes B)\big) = \exp(tA) \otimes \exp(tB) \qquad (t \in \mathbb{R}). \tag{3.156}$$

Exercise 3.16 (i) (Second Resolvent Identity) Let A and $A^\times$ be $n \times n$ complex matrices. Show that their resolvents satisfy

$$(sI - A)^{-1} - (sI - A^\times)^{-1} = (sI - A)^{-1}(A - A^\times)(sI - A^\times)^{-1}$$

when s is in the resolvent set of A and the resolvent set of $A^\times$.

(ii) (Inverse of a transfer function) Suppose that $\Sigma = (A, B, C, D)$ has D invertible and let $T(s) = D + C(sI - A)^{-1}B$ be the transfer function; then let

$$\Sigma^\times = \begin{bmatrix} A^\times & B^\times \\ C^\times & D^\times \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C & BD^{-1} \\ -D^{-1}C & D^{-1} \end{bmatrix}$$

have transfer function $T^\times(s) = D^\times + C^\times(sI - A^\times)^{-1}B^\times$. Show that

$$T^\times(s)T(s) = I.$$

Exercise 3.17

(i) Let $V \in M_{n\times 1}(\mathbb{C})$ and $C = M_{1\times n}(\mathbb{C})$, so VC is of rank one. By considering the Jordan form of VC or otherwise, show that

$$\det(I - sVC) = 1 - s\,\mathrm{trace}\, VC = 1 - sCV.$$

(ii) Deduce that

$$\det(sI - A - VC) = \det(sI - A) - C\operatorname{adj}(sI - A)V.$$

(iii) Suppose that V is an eigenvector of A corresponding to eigenvalue λ. Deduce that there is a factorization of polynomials

$$\det(sI - A - VC) = \frac{\det(sI - A)}{s - \lambda}\big(s - \lambda - CV\big).$$

Exercise 3.18 Let $X \in M_{2\times 2}(\mathbb{C})$ satisfy $\operatorname{trace}(X) = 0$.

(i) Use the Cayley-Hamilton theorem 2.29 to show that $X^2 = -\delta^2 I_2$ for some $\delta \in \mathbb{C}$.
(ii) Deduce that, the terminology of Definition 4.43,

$$\exp(X) = \cos(\delta)I_2 + \operatorname{sinc}(\delta)X.$$

(iii) Deduce that the equation

$$\exp(X) = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

has no such solution X. [In [52], p111, this example is credited to Engel.]

Exercise 3.19 (Matrix Logarithm) Show that for $X \in M_{n\times n}(\mathbb{C})$ with $\|X\| < 1$, the integral

$$L(X) = \int_0^\infty \Big((1+t)^{-1}I_n - (tI + I + X)^{-1}\Big)dt \tag{3.157}$$

is convergent.

(ii) By considering Taylor's series of $L(\lambda X)$ or otherwise, obtain the formula

$$L(X) = X - \frac{X^2}{2} + \frac{X^3}{3} - \frac{X^4}{4} + \dots,$$

as in $L(X) = \log(I + X)$.

(iii) Using the integral from (i), show how to define a positive definite $L(X)$ for X positive definite.

Exercise 3.20 Suppose that S is an invertible matrix. From the SISO system (A, B, C, D), we introduce another SISO by $(\hat{A}, \hat{B}, \hat{C}, \hat{D}) = (S^{-1}AS, S^{-1}B, CS, D)$.

(i) Show that

$$\det \begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} = \det \begin{bmatrix} \hat{A} - sI & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix}. \tag{3.158}$$

(ii) Using Lemma 3.14, or otherwise, deduce that the transfer functions of these linear systems are equal.

Exercise 3.21 (Variant of the Schur Complement Formula) Let (A, B, C, D) have D invertible.

(i) Derive the formula

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -D^{-1}C \end{bmatrix} (A - BD^{-1}C)^{-1} \begin{bmatrix} I & -BD^{-1} \end{bmatrix}.$$

(ii) Replace A by $A - sI$ and compute the right-hand side.

Chapter 4
Laplace Transforms

The Laplace transform is a fundamental tool for solving differential equations with constant coefficients. The merit of the Laplace transform is that solutions of linear systems such as constant coefficient ordinary differential equations have Laplace transforms which are well-behaved functions, such as holomorphic on a half plane. Holomorphic means analytic, or differentiable as a function of a complex variable. In this chapter, we present several of the fundamental results about the Laplace transform and obtain famous results such as Heaviside's expansion theorem which was important in the historical development of linear systems. In this book, we have introduced the theory in terms of state space models with a differential equation in time variable t for a state vector X satisfying a linear differential equation with constant matrix coefficients. Here we consider how the MIMO system (A, B, C, D) can be transformed via the Laplace transform, and we discover the meaning of the transfer function $T(s)$ which previously was defined by a largely unmotivated formula. The Laplace transform replaces d/dt by multiplication by a variable s, which leads to a description of linear systems in terms of algebra in which $T(s)$ is central to the discussion. In Chap. 5, we will also interpret transfer functions geometrically in terms of plots involving s.

4.1 Laplace Transforms

Definition 4.1 (Laplace Transform)

(i) A function $f : (0, \infty) \to \mathbb{C}$ is said to be piecewise continuous if there exists an increasing sequence $(a_j)_{j=1}^{\infty}$ with $a_j \to \infty$ as $j \to \infty$ such that the restricted function $f|(a_j, a_{j+1})$ is continuous.

(ii) Suppose that $f : (0, \infty) \to \mathbb{C}$ is a piecewise continuous function such that

$$(E) \qquad |f(x)| \leq Me^{\beta x} \qquad (x > 0) \tag{4.1}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_4

for some $M > 0$ and $\beta \in \mathbb{R}$. Here β is called the exponential type or growth rate. Then we say that f is of exponential type, or satisfies (E).

(iii) We then define the Laplace transform by

$$\mathcal{L}(f)(s) = \int_0^\infty f(x)e^{-sx}\,dx \qquad (\Re s > \beta). \tag{4.2}$$

Sometimes $\mathcal{L}(f)(s)$ is written as $\hat{f}(s)$. Here x, t are time variables; whereas s is the transform variable. Writers often contrast the time domain with s-space, to emphasize the difference in interpretation. The term s-space is not an abbreviation for state space, since the latter relates to the time domain.

Example 4.2 (Laplace Transform Table) Let a be a non-zero real number, $b > 0$ and $\alpha > -1$.

$f(t)$	$\mathcal{L}(f)(s)$
1	$1/s$
t^n	$\dfrac{n!}{s^{n+1}}$
e^{at}	$\dfrac{1}{s-a} \quad (s > a)$
t^α	$\dfrac{\Gamma(\alpha+1)}{s^{\alpha+1}}$
$\sin at$	$\dfrac{a}{s^2+a^2}$
$\cos at$	$\dfrac{s}{s^2+a^2}$
$\delta_b(dt)$	e^{-bs}
$H(t-b)$	$\dfrac{e^{-bs}}{s}$
$f(t/b)$	$b\hat{f}(bs)$
$H(t-b)f(t-b)$	$e^{-bs}\hat{f}(s)$
$e^{at}f(t)$	$\hat{f}(s-a)$

Example 4.3 Calculating Some Laplace Transforms

(i) For all $\varepsilon, t \geq 0$ and $\ell \in \mathbb{N}$, we have

$$e^{\varepsilon t} = 1 + \varepsilon t + \frac{\varepsilon^2 t^2}{2!} + \cdots + \frac{\varepsilon^\ell t^\ell}{\ell!} + \cdots \geq \frac{\varepsilon^\ell t^\ell}{\ell!},$$

so $t^\ell \leq \ell! e^{\varepsilon t}/\varepsilon^\ell$, so t^ℓ satisfies (E).

(ii) Now consider $f(t) = t^2$, with

$$\mathcal{L}(t^2; s) = \frac{2}{s^3} \qquad (s > 0). \tag{4.3}$$

To see this, consider $R > 0$ and integrate by parts

$$\begin{aligned}
\int_0^R t^2 e^{-st}\,dt &= \left[\frac{t^2 e^{-st}}{-s}\right]_0^R + \frac{2}{s}\int_0^R t e^{-st}\,dt \\
&= \left[\frac{t^2 e^{-st}}{-s}\right]_0^R + \left[\frac{2te^{-st}}{-s^2}\right]_0^R + \frac{2}{s^2}\int_0^R e^{-st}\,dt \\
&= \left[\frac{t^2 e^{-st}}{-s}\right]_0^R + \left[\frac{2te^{-st}}{-s^2}\right]_0^R + \left[\frac{2e^{-st}}{-s^3}\right]_0^R \\
&= -\frac{R^2 e^{-Rs}}{s} - \frac{2Re^{-Rs}}{s^2} - \frac{2e^{-Rs}}{s^3} + \frac{2}{s^3} \qquad (s > 0) \\
&\to \frac{2}{s^3} \qquad \text{as} \qquad R \to \infty.
\end{aligned}$$

Remark 4.4 Comments Concerning Some Functions

(i) In the above table t^α satisfies (E) for $\alpha \geq 0$; for $0 > \alpha > -1$, t^α diverges as $x \to 0+$, but the Laplace transform integral exists as an improper Riemann integral.

(ii) The Dirac delta function δ_b is not actually a function, instead δ_b is the measure that assigns unit mass to the point $b \geq 0$ on the line. So $\int f(t)\delta_b(dt) = f(b)$ for all continuous real functions f. The measure δ_0 is often called the unit impulse function; as an input, it gives the system a kick start.

(iii) The Heaviside function

$$\begin{aligned}
H(t) &= 1, \quad t \geq 0; \\
H(t) &= 0, \quad t < 0;
\end{aligned} \tag{4.4}$$

is a step function with a jump at $x = 0$, so $H(x - b)$ is a step function with a jump at $x = b$. Hence $H(x - b) = \int_{(-\infty, x]} \delta_b(dt)$. While H has a jump, it is piecewise continuous and bounded, so the Laplace transform is defined using the same formula as above. The Heaviside function has

$$\begin{aligned}
\mathcal{L}(H(t - b); s) &= \int_0^\infty H(t - b)e^{-st}\,dt \\
&= \int_b^\infty e^{-st}\,dt \\
&= \frac{e^{-sb}}{s} \qquad (s > 0).
\end{aligned}$$

Exercise Suppose that $(a_n)_{n=0}^{\infty}$ is a complex sequence such that $|a_n| \le Mr^n$ for all $n = 0, 1, \ldots$ for some $M, r > 0$. Show that $f(t) = \sum_{n=0}^{\infty} a_n t^n/n!$ converges for all $t \in \mathbb{C}$ and defines a continuous function of type (E) on $[0, \infty)$. Show how $f(t) = \sin(2t)$ and $f(t) = \cos(4t)$ arise in this way.

Proposition 4.5 (Properties of the Laplace Transform) *Here (E) refers to some $M > 0$ and $\beta \in \mathbb{R}$, and s is real.*

(i) The Laplace transform exists for all $s > \beta$, and $|\mathcal{L}(f)(s)| \le \frac{M}{s-\beta}$ for all $s > \beta$.
(ii) The Laplace transform is linear so, that if f, g satisfy (E), then for all $\lambda, \mu \in \mathbb{C}$ the function $\lambda f + \mu g$ also satisfies (E) and

$$\mathcal{L}(\lambda f + \mu g)(s) = \lambda \mathcal{L}(f)(s) + \mu \mathcal{L}(g)(s). \tag{4.5}$$

(iii) $tf(t)$ also satisfies (E) and $\mathcal{L}(f)(s)$ is differentiable with

$$\mathcal{L}(tf(t))(s) = -\frac{d}{ds}\mathcal{L}(f)(s). \tag{4.6}$$

(iv) If f is continuously differentiable and df/dt satisfies (E), then f also satisfies (E) and $\mathcal{L}(df/dt)(s) = s\mathcal{L}(f)(s) - f(0)$.

Proof

(i) Let $|f(x)| \le Me^{\beta x}$. Then for $0 < W < R$

$$\begin{aligned}\Big|\int_W^R e^{-sx} f(x)\,dx\Big| &\le M\int_W^R e^{\beta x}e^{-sx}\,dx\\ &= \Big[\frac{M}{\beta - s}e^{(\beta-s)x}\Big]_W^R\\ &= \frac{M}{\beta-s}e^{(\beta-s)R} - \frac{M}{\beta-s}e^{(\beta-s)W} \to 0\end{aligned}$$

as $W \to \infty$. Also, we can let $R \to \infty$ and $W \to 0+$ to get

$$\Big|\int_0^{\infty} e^{-sx} f(x)\,dx\Big| \le \frac{M}{s-\beta}. \tag{4.7}$$

(ii) Suppose that $|f(x)| \le Pe^{ax}$ and $|g(x)| \le Re^{bx}$ for all $x > 0$. Then with $\beta = \max\{a, b\}$ and $M = |\lambda|P + |\mu|R$, we have

$$|\lambda f(x) + \mu g(x)| \le Me^{\beta x}, \tag{4.8}$$

so we can integrate

$$\int_0^\infty e^{-sx}(\lambda f(x) + \mu g(x))\,dx = \lambda \int_0^\infty e^{-sx} f(x)\,dx + \mu \int_0^\infty e^{-sx} g(x)\,dx. \tag{4.9}$$

(iii) Suppose that $|df/dx| \leq Me^{\beta x}$ for all $x > 0$ and some $\beta > 0$, so df/dx belongs to (E); we verify that f also belongs to (E). By fundamental theorem of calculus,

$$f(x) = f(0) + \int_0^x \frac{df}{dt}(t)\,dt,$$

so

$$\begin{aligned} |f(x)| &\leq |f(0)| + \int_0^x Me^{\beta t}\,dt \\ &= |f(0)| + \Big[\frac{Me^{\beta t}}{\beta}\Big]_0^x \\ &= |f(0)| + \frac{Me^{\beta x}}{\beta} - \frac{M}{\beta}, \end{aligned}$$

so f satisfies (E). Now for $s > \beta$, we integrate by parts to get

$$\begin{aligned} \int_0^R e^{-sx}\frac{df}{dx}\,dx &= \Big[e^{-sx} f(x)\Big]_0^R + s\int_0^R e^{-sx} f(x)\,dx \\ &= e^{-sR} f(R) - f(0) + s\int_0^R e^{-sx} f(x)\,dx \end{aligned}$$

so we let $R \to \infty$ to get

$$\int_0^\infty e^{-sx}\frac{df}{dx}\,dx = -f(0) + s\int_0^\infty e^{-sx} f(x)\,dx.$$

(iv) Differentiating Laplace transforms: Let $s > \beta + \delta$ for some $\delta > 0$ and consider $-\delta < h < \delta$. Note that $e^{\delta x}e^{-sx} f(x)$ is integrable, and $x \leq e^{\delta x}/\delta$, so $xf(x)$ also satisfies (E). Also

$$\frac{e^{-(s+h)x} - e^{-sx}}{h} = e^{-sx}\Big(\frac{e^{-hx} - 1}{h}\Big) \to -xe^{-sx} \tag{4.10}$$

as $h \to 0$. Hence

$$\begin{aligned}\frac{\mathcal{L}(f)(s+h)-\mathcal{L}(f)(s)}{h} &= \int_0^\infty \frac{e^{-(s+h)x}-e^{-sx}}{h} f(x)\,dx\\ &= \int_0^\infty \frac{e^{-xh}-1}{h} e^{-sx} f(x)\,dx\\ &\to -\int_0^\infty x e^{-sx} f(x)\,dx.\end{aligned}$$

To make this precise, we consider

$$\frac{e^{-xh}-1}{h}+x = h\Big(\frac{e^{-xh}-1+hx}{h^2}\Big) \tag{4.11}$$

where by comparing the coefficients in the power series

$$\begin{aligned}\Big|\frac{e^{-xh}-1+hx}{h^2}\Big| &= \Big|\frac{x^2}{2!}-\frac{hx^3}{3!}+\frac{h^2x^4}{4!}-\dots\Big|\\ &\le \frac{x^2}{2!}+\frac{\delta x^3}{3!}+\frac{\delta^2x^4}{4!}+\dots\\ &= \frac{e^{\delta x}-1-\delta x}{\delta^2} \le \delta^{-2}e^{\delta x}\end{aligned}$$

we have $e^{\delta x}e^{-sx}f(x)$ is integrable, and

$$\Big|\int_0^\infty \frac{e^{-xh}-1+hx}{h} e^{-sx} f(x)\,dx\Big| \le |h|\int_0^\infty \delta^{-2}e^{\delta x}e^{-sx} f(x)\,dx. \tag{4.12}$$

□

Proposition 4.6 (Holomorphic Laplace Transform) *Suppose that f satisfies (E). Then*

(i) $\mathcal{L}(f)(s)$ defines a holomorphic (complex differentiable) function of s on the open left half-plane $\{s \in \mathbb{C} : \Re s > \beta\}$;
(ii) $\mathcal{L}(f)(s) \to 0$ as $s \to \infty$ along $(0, \infty)$.
(iii) Let $\bar{f}$ be the complex conjugate of f. Then $\mathcal{L}(\bar{f})(s) = \mathcal{L}(f)(\bar{s})$.

Proof

(i) Similar to Proposition 4.5 (iii)
(ii) This is similar to Proposition 4.5 (i).
(iii) We have

$$\mathcal{L}(\bar{f})(s) = \int_0^\infty \bar{f}(t)e^{-st}dt \tag{4.13}$$

is the complex conjugate of

$$\mathcal{L}(f)(\bar{s}) = \int_0^\infty f(t)e^{-\bar{s}t}dt. \tag{4.14}$$

□

Definition 4.7 Let Euler's Gamma function be $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$ for $\alpha > 0$.

Proposition 4.8 (Laplace Transform of Exponentials and Powers) *Let $\nu_j > -1$ and let $p_j \in \mathbb{C}$ for $j = 1, \dots, N$. Then*

$$f(t) = \sum_{j=1}^N a_j t^{\nu_j} e^{p_j t} \tag{4.15}$$

satisfies (E) for $\beta > \max\{\Re p_j\}$ and $\mathcal{L}(f)(s)$ is holomorphic for $\Re s > \beta$.

Proof By direct calculation, we have

$$\mathcal{L}(f)(s) = \sum_{j=1}^N a_j \frac{\Gamma(\nu_j + 1)}{(s - p_j)^{\nu_j+1}}. \tag{4.16}$$

The Laplace transform is holomorphic on the half plane $\{s \in \mathbb{C} : \Re s > \beta\}$, so we stay to the right of the singularities at the p_j. If the ν_j are all integers, then $\mathcal{L}(f)(s)$ has a pole of order $\nu_j + 1$ at p_j, and $\mathcal{L}(f)(s)$ is a rational function which is holomorphic on $\mathbb{C} \setminus \{p_1, \dots, p_N\}$. □

4.2 Laplace Convolution

Definition 4.9 (Convolution) Suppose that f and g both satisfy (E). Then their Laplace convolution is

$$f * g(x) = \int_0^x f(x - y)g(y)\,dy. \tag{4.17}$$

Observe that the variable y moves along the range of integration $[0, x]$, and we have y and $x-y$ in the integrand. We use the phrase Laplace convolution to avoid possible confusion with convolution on $\mathbb{R}$, where in the latter case the range of integration is $\mathbb{R}$. We can take our functions f and g to live on $[0, \infty)$, so there is no ambiguity.

Proposition 4.10 *The Laplace convolution is:*

(i) commutative, so $f * g = g * f$;
(ii) linear, so $(\lambda f + \mu g) * h = \lambda f * h + \mu g * h$;

(iii) multiplicative with respect to the Laplace transform, so $f * g$ satisfies (E) and

$$\mathcal{L}(f * g)(s) = \mathcal{L}(f)(s)\mathcal{L}(g)(s) \qquad (s > s_0); \tag{4.18}$$

(iv) associative, so $f * (g * h) = (f * g) * h$.

Proof

(i) Change variable to $u = y - x$.
(ii) is easy.
(iii) Bounds on convolution formula: We choose M, β such that

$$|f(x)| \leq Me^{\beta x}, \quad |g(x)| \leq Me^{\beta x} \qquad (x > 0); \tag{4.19}$$

then

$$|f(x-y)g(y)| \leq Me^{\beta(x-y)}Me^{\beta y} = M^2e^{\beta x} \tag{4.20}$$

so

$$\left|\int_0^x f(x-y)g(y)\,dy\right| \leq xM^2e^{\beta x} \tag{4.21}$$

so $f * g$ satisfies (E).

Proof of convolution formula: Also, when we change order of integration, then let $u = x - y$,

$$\begin{aligned}\mathcal{L}(f * g)(s) &= \int_0^\infty e^{-sx} f * g(x)\,dx \\ &= \int_0^\infty e^{-sx}\int_0^x f(x-y)g(y)\,dy\,dx \\ &= \int_0^\infty \Big(\int_y^\infty e^{-s(x-y)}f(x-y)\,dx\Big)e^{-sy}g(y)\,dy \\ &= \int_0^\infty e^{-su}f(u)\,du\int_0^\infty e^{-sy}g(y)\,dy \\ &= \mathcal{L}(f)(s)\mathcal{L}(g)(s).\end{aligned}$$

Thus the Laplace transform converts convolution to multiplication.

(iv) Associativity: This can be proved in a similar way to (i). Alternatively, one uses (iii) to compute

$$\begin{aligned}\mathcal{L}((f * g) * h)(s) &= \mathcal{L}(f * g)(s)\mathcal{L}(h)(s) = \mathcal{L}(f)(s)\mathcal{L}(g)(s)\mathcal{L}(h)(s) \\ &= \mathcal{L}(f * (g * h))(s)\end{aligned} \tag{4.22}$$

and then use uniqueness, as discussed below.

□

4.3 Laplace Uniqueness Theorem

Theorem 4.11 *Suppose that f and g satisfy (E) and that there exists s_0 such that*

$$\mathcal{L}(f)(s) = \mathcal{L}(g)(s) \qquad (s > s_0). \tag{4.23}$$

Then $f(x) = g(x)$ for all $x > 0$ at which $f - g$ is continuous.

Proof We defer the proof of this theorem until Sect. 4.10. In Corollary 9.5 we obtain a stronger version due to Lerch. □

Meanwhile, if one knows that $F(s)$ occurs as $\mathcal{L}(f)(s)$ for some f, then the best way to find f is by comparing F with known Laplace transforms in tables, then invoking the uniqueness theorem. It would also help to describe the functions $F(s)$ that arise as Laplace transforms, and have an effective formula that produces an explicit $f(t)$ from $F(s)$. There is an inversion formula, credited to Bromwich, which takes a suitable $F(s)$ and produces this function $f(t)$ via a contour integral as in [53] page 177. The following result covers some cases of interest.

Proposition 4.12 (Holomorphic at Infinity) *Suppose that $F(s)$ is holomorphic near ∞ with $F(\infty) = 0$ so that $F(s)$ has a convergent Laurent series $F(s) = \sum_{n=0}^{\infty} a_n s^{-n-1}$ on $\{s : |s| > \sigma\}$ for some $\sigma > 0$. Then $f(t) = \sum_{n=0}^{\infty} a_n t^n/n!$ is of type (E) and $F(s)$ is the Laplace transform of $f(t)$.*

Proof A function F is holomorphic at infinity if $F(1/s)$ has a removable singularity at 0, so we can write $F(1/s) = a_{-1} + \sum_{n=0}^{\infty} a_n s^{n+1}$ where $F(1/s) \to a_{-1}$ as $s \to 0$, and we can interpret a_{-1} as $F(\infty)$. In particular, F vanishes at infinity when $F(\infty) = 0$. See page 123 [**?**] and Exercise 4.13. Note that $F(s)$ is holomorphic on the half plane $\{s : \Re s > \sigma\}$ with $F(s) \to 0$ as $s \to \infty$, as in Sect. 4.1. □

Remark 4.13

(i) A strictly proper rational function $F(s)$ is holomorphic near ∞ with $F(\infty) = 0$, and we obtain an explicit form for $f(t)$ in terms of partial fractions and residues in Proposition 6.55. In linear systems, strictly proper stable rational functions with simple poles occur frequently, so we deal with this special case in Sect. 4.7 with Heaviside's expansion.

(ii) The algebraic function $1/\sqrt{1+s^2}$ is holomorphic at infinity, and this occurs in the theory of Bessel functions as in Exercise 4.12.

(iii) Whereas $\sqrt{\pi}/\sqrt{s}$ is holomorphic on $\mathbb{C} \setminus (-\infty, 0]$ it is not holomorphic at ∞. This function arises as the Laplace transform of $1/\sqrt{t}$, but the inversion is much more complicated and uses a specially chosen contour.

(iv) In Sect. 10.7, we prove the Paley-Wiener theorem 10.36 which gives the definitive description of the functions that arise as Laplace transforms of square integrable functions, along with a general inversion theorem.

Example 4.14 We consider some examples relating to exponential and trigonometric functions, which are interesting in applications, and we use methods which also work for Bessel functions in subsequent sections.

(i) The hyperbolic function $\sinh at$ has a series $\sinh at = \sum_{n=0}^{\infty}(at)^{2n+1}/(2n+1)!$ that gives a function satisfying (E), and for $s > a > 0$, this has Laplace transform

$$\begin{aligned}\mathcal{L}(\sinh at)(s) &= \int_0^\infty \sum_{n=0}^\infty \frac{a^{2n+1}t^{2n+1}}{(2n+1)!}e^{-st}\,dt \\ &= \sum_{n=0}^\infty \int_0^\infty \frac{a^{2n+1}t^{2n+1}}{(2n+1)!}e^{-st}\,dt \\ &= \sum_{n=0}^\infty \frac{a^{2n+1}}{s^{2n+2}} \\ &= \frac{a}{s^2-a^2},\end{aligned}$$

where the change in order of integration and summation is justified by uniform convergence or the monotone convergence theorem. The Laplace transform is rational and holomorphic at infinity. One can otherwise obtain this integral from $\sinh at = (e^{at} - e^{-at})/2$.

(ii) The Laplace transform of $f(t) = t^{-1/2}\cos(at^{1/2})$ is $\sqrt{(\pi/s)}e^{-a^2/(4s)}$ for $a \in \mathbb{R}$. To check this, we compute

$$\begin{aligned}\mathcal{L}f(s) &= \int_0^\infty e^{-st}\sum_{j=0}^\infty \frac{(-1)^j a^{2j}t^{j-1/2}}{(2j)!}dt \\ &= \sum_{j=0}^\infty \int_0^\infty \frac{(-1)^j a^{2j}e^{-st}t^{j-1/2}}{(2j)!}dt \\ &= \sum_{j=0}^\infty \frac{(-1)^j a^{2j}\Gamma(j+1/2)}{(2j)!s^{j+1/2}},\end{aligned} \tag{4.24}$$

and we can simplify this by multiplying the numerator and denominator by $2^j j!$ to obtain

$$\begin{aligned}\mathcal{L}f(s) &= \frac{\sqrt{\pi}}{\sqrt{s}}\sum_{j=0}^\infty \frac{(-1)^j a^{2j}}{2^{2j}j!s^j} \\ &= \frac{\sqrt{\pi}}{\sqrt{s}}e^{-a^2/(4s)}.\end{aligned}$$

(iii) We introduce the error function by $\operatorname{erf}(x) = 2\int_0^x e^{-t^2} dt/\sqrt{\pi}$. This defines an entire function, with Taylor expansion

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j+1}}{(2j+1)j!}.$$

Let $g(t) = t^{-1}\sin(at^{1/2})$; then by a similar calculation to the preceding one (4.24), the Laplace transform satisfies

$$\mathcal{L}g(s) = \frac{\sqrt{\pi}}{\sqrt{s}} \sum_{j=0}^{\infty} \frac{(-1)^j a^{2j}}{2^j(2j+1)j!s^j},$$

and we deduce that

$$\mathcal{L}g(s) = \pi \operatorname{erf}\left(\frac{a}{2\sqrt{s}}\right),$$

which is holomorphic at infinity.

Example 4.15 To solve the integral equation

$$y(t) = 2e^{at} + \int_0^t e^{b(t-u)} y(u)\, du \tag{4.25}$$

where a, b are constants with $a \neq b + 1$, and y satisfies (E).
Solution Note that the the integral is a convolution of y with e^{bt}, so by Proposition 4.10(iii), we have

$$\hat{y}(s) = \frac{2}{s-a} + \frac{\hat{y}(s)}{s-b}; \tag{4.26}$$

after a little reduction we obtain

$$\hat{y}(s) = \frac{2(s-b)}{(s-a)(s-b-1)}; \tag{4.27}$$

which we write as partial fractions

$$\hat{y}(s) = \frac{A}{s-a} + \frac{B}{s-b-1}; \tag{4.28}$$

then

$$A = \frac{2(a-b)}{a-b-1}; \quad B = \frac{-2}{a-b-1}; \tag{4.29}$$

by the uniqueness of Laplace transforms, we have a unique solution

$$y(t) = \frac{2(a-b)e^{at} - 2e^{(b+1)t}}{a-b-1}. \tag{4.30}$$

Example 4.16 (Unique Solutions of a Population Equation) Let x be the size of a population at time $t > 0$. The birth rate and death rate depend upon the age profile of the population, as represented by a function g, and there can be emigration and immigration, represented by an input u, so the rate of change of population is given by

$$\frac{dx}{dt} = \int_0^t g(t-\tau)x(\tau)d\tau + u(t). \tag{4.31}$$

This is equivalent to the integral equation

$$x(t) = x_0 + \int_0^t \int_0^{t-\tau} g(v)dv x(\tau)d\tau + \int_0^t u(\tau)d\tau \qquad (t > 0). \tag{4.32}$$

Assuming g and u are bounded and piecewise continuous, one use a version of Gronwall's inequality to deduce that x satisfies (E); see [26] page 371. Then we deduce that the equation has Laplace transform

$$sX(s) - x(0) = G(s)X(s) + U(s) \tag{4.33}$$

so

$$X(s) = \frac{x(0)}{s - G(s)} + \frac{1}{s - G(s)}U(s). \tag{4.34}$$

If we can invert the Laplace transform on the right-hand side, then this leads to an explicit solution. Otherwise, we can regard this as a uniqueness result pertaining to the solution.

4.4 Laplace Transform of a Differential Equation

Proposition 4.17 *Suppose that* $y(0) = p_0$, $\frac{dy}{dt}(0) = p_1, \dots, \frac{d^{n-1}y}{dt^{n-1}}(0) = p_{n-1}$ *and*

$$\frac{d^n y}{dt^n}(t) + a_{n-1}\frac{d^{n-1}y}{dt^{n-1}}(t) + \cdots + a_0 y = u(t) \tag{4.35}$$

where the coefficients are complex constants and u satisfies (E). Then there exists a complex polynomial $q_{n-1}(s)$ of degree $\leq n-1$, that depends only upon the a_j and p_k, such that

$$(s^n + a_{n-1}s^{n-1} + \cdots + a_0)\mathcal{L}(y)(s) + q_{n-1}(s) = \mathcal{L}(u)(s) \tag{4.36}$$

where $s^n + a_{n-1}s^{n-1} + \cdots + a_0$ is the characteristic polynomial, as in Lemma 1.10 and Definition 2.10.

Proof By Theorem 2.40, we know that this initial value problem has a solution y, and by Sect. 4.5 y is of type (E). By repeatedly applying Proposition 4.5, we have

$$\begin{aligned}
\mathcal{L}\Big(\frac{dy}{dt}\Big)(s) &= s\mathcal{L}(y)(s) - y(0) \\
\mathcal{L}\Big(\frac{d^2y}{dt^2}\Big)(s) &= s\mathcal{L}\Big(\frac{dy}{dt}\Big)(s) - \Big(\frac{dy}{dt}\Big)(0) \\
\mathcal{L}\Big(\frac{d^ny}{dt^n}\Big)(s) &= s\mathcal{L}\Big(\frac{d^{n-1}y}{dt^{n-1}}\Big)(s) - \Big(\frac{d^{n-1}y}{dt^{n-1}}\Big)(0),
\end{aligned}$$

so we can substitute backwards and get

$$\begin{aligned}
\mathcal{L}\Big(\frac{d^2y}{dt^2}\Big)(s) &= s^2\mathcal{L}(y)(s) - sy(0) - \Big(\frac{dy}{dt}\Big)(0) \\
\mathcal{L}\Big(\frac{d^3y}{dt^3}\Big)(s) &= s^3\mathcal{L}(y)(s) - s^2y(0) - s\Big(\frac{dy}{dt}\Big)(0) - \Big(\frac{dy^2}{dt^2}\Big)(0)
\end{aligned}$$

and thus obtain $q_{n-1}(s)$ with coefficients $p_j = \frac{d^jy}{dt^j}(0)$ as in the initial conditions $y(0), \ldots, \frac{d^{n-1}y}{dt^{n-1}}(0)$. The characteristic polynomial here is the same as we get from Lemma 1.10 and Definition 2.10. □

Proposition 4.17 takes us from the data in the differential equation to an algebraic relation between their Laplace transforms. This leads directly to some interesting information, as we see in Chap. 5, but to make full use of the result, we need a systematic method for deriving the solution in the time variable. In Theorem 4.27 we obtain such an inversion process that works for stable characteristic polynomials.

Example 4.18 (Laplace Transform of Differential Equation) To find the Laplace transform of

$$\begin{aligned}
\frac{d^3y}{dt^3} + 2\frac{d^2y}{dt^2} - \frac{dy}{dt} + y &= u \\
\frac{d^2y}{dt^2}(0) = 5, \frac{dy}{dt}(0) = -1, y(0) &= 7
\end{aligned}$$

where $u \in (E)$.

Solution The Laplace transforms are found recursively, with

$$\begin{aligned}\mathcal{L}(y)(s) &= \hat{y}(s)\\ \mathcal{L}\Big(\frac{dy}{dt}\Big)(s) &= s\hat{y}(s) - 7\\ \mathcal{L}\Big(\frac{d^2y}{dt^2}\Big)(s) &= s^2\hat{y}(s) - 7s + 1\\ \mathcal{L}\Big(\frac{d^3y}{dt^3}\Big)(s) &= s^3\hat{y}(s) - 7s^2 + s - 5\end{aligned}$$

and substituting this into

$$\mathcal{L}\Big(\frac{d^3y}{dt^3}\Big) + 2\mathcal{L}\Big(\frac{d^2y}{dt^2}\Big) - \mathcal{L}\Big(\frac{dy}{dt}\Big) + \mathcal{L}(y) = \mathcal{L}(u) \tag{4.37}$$

gives

$$(s^3 + 2s^2 - s + 1)\hat{y}(s) - 7s^2 - 13s + 4 = \hat{u}(s), \tag{4.38}$$

which we write as

$$\hat{y}(s) = \frac{7s^2 + 13s - 4}{s^3 + 2s^2 - s + 1} + \frac{1}{s^3 + 2s^2 - s + 1}\hat{u}(s), \tag{4.39}$$

where the first term on the right-hand side is the Laplace transform of the complementary function with constants chosen four the boundary values and the final term is the Laplace transform of the particular integral.To make further progress, one needs to find the roots of $s^3 + 2s^2 + 13s - 4 = 0$, which are approximately -2.8312 and $0.4156 \pm 0.4248i$.

Remark 4.19

(i) Alternatively, we can represent this as an (A, B, C, D) SISO system and use the theorem of the next section.
(ii) We often take s such that $\Re s \geq 0$, so s is in the left half-plane, and denote the points on the imaginary axis by $s = i\omega$, where $\omega \in \mathbb{R}$ is the angular frequency.

4.5 Solving MIMO by Laplace Transforms

Definition 4.20 (Transfer Function) Consider a linear system $Y = LU$ where L is a linear operator, and such that all the entries of the $(k \times 1)$ input U and $(m \times 1)$ output Y satisfy (E), and let the initial conditions be zero. Suppose that $T(s)$ is a

$(m \times k)$ matrix of functions such that

$$\hat{Y}(s) = T(s)\hat{U}(s) \qquad (s > \beta). \tag{4.40}$$

Then $T(s)$ is called the transfer function of the linear system.

Theorem 4.21 *Let A, B, C, D be constant matrices, and suppose that the input function satisfies (E). Then the output Y of the linear system*

$$\begin{aligned} \frac{dX}{dt} &= AX + BU \\ Y &= CX + DU \end{aligned} \tag{4.41}$$

with initial condition $X(0) = 0$ in (E) is uniquely determined, and the Laplace transform satisfies

$$\hat{Y}(s) = T(s)\hat{U}(s) \tag{4.42}$$

where the transfer function is $T(s) = D + C(sI - A)^{-1}B$.

Proof By Theorem 2.40, the solution is determined by the state

$$X(t) = \int_0^t \exp(((t - v)A)BU(v)\, dv \tag{4.43}$$

which is a convolution type integral of functions in (E), since

$$\|\exp(tA)\| \le M_1 e^{\beta_1 t}, \quad \|U(t)\| \le M_2 e^{\beta_2 t} \tag{4.44}$$

so with $M = M_1 M_2 \|B\|$ and $\beta = \max\{\beta_1, \beta_2\}$, we have

$$\|\exp(((t - v)A)BU(v)\| \le Me^{\beta t} \tag{4.45}$$

so

$$\|X(t)\| \le tMe^{\beta t} \le Me^{(\beta+1)t} \tag{4.46}$$

and so X satisfies (E) and has a Laplace transform. From the differential equation, dX/dt also satisfies (E) and has a Laplace transform, and likewise Y satisfies (E).

The Laplace transform of

$$\begin{aligned} \frac{dX}{dt} &= AX + BU \\ Y &= CX + DU \end{aligned} \tag{4.47}$$

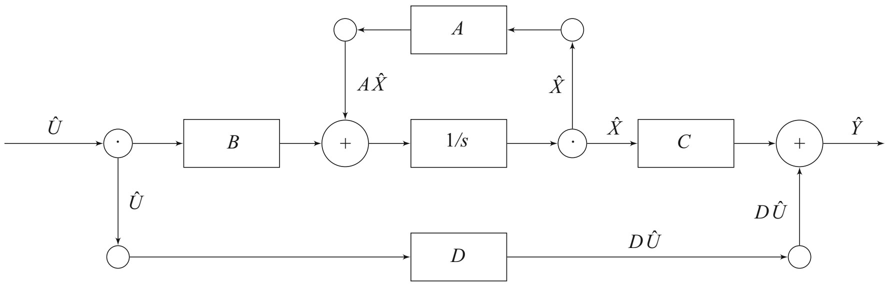
is

$$s\hat{X}(s) - X(0) = A\hat{X}(s) + B\hat{U}(s)$$
$$\hat{Y}(s) = C\hat{X}(s) + D\hat{U}(s). \tag{4.48}$$

Hence when s is not an eigenvalue of A,

$$\hat{X}(s) = (sI - A)^{-1}B\hat{U}(s) + (sI - A)^{-1}X(0)$$
$$\hat{Y}(s) = C(sI - A)^{-1}B\hat{U}(s) + D\hat{U}(s) + C(sI - A)^{-1}X(0). \tag{4.49}$$

When $X(0) = 0$ we get $\hat{Y}(s) = T(s)\hat{U}(s)$. □



Block diagram for the Laplace transform of the MIMO system

4.6 Partial Fractions

In the previous Sects. 4.4 and 4.5, we have obtained solution of differential equations such that the Laplace transforms are rational functions. In this section we give an informal discussion of how to express these rational functions, which we will make more systematic in Chap. 6. See also [6], page 79.

Proposition 4.22

(i) Let $f(s)$ be a complex rational function. Then there exists a complex polynomial $q(s)$, integers $m_j > 0$ and poles $\lambda_j \in \mathbb{C}$ and $a_{j,k} \in \mathbb{C}$, all uniquely determined, such that

$$f(s) = q(s) + \sum_{j=1}^{N}\sum_{k=1}^{m_j} \frac{a_{j,k}}{(s - \lambda_j)^k}. \tag{4.50}$$

(ii) Suppose that $f(s) = r(s)/h(s)$ where the degree N of $h(s)$ is greater than the degree of $r(s)$ and $h(s)$ has only simple zeros at $z_1, \dots, z_N$. Then the partial fractions decomposition of $f(s)$ is

$$f(s) = \sum_{j=1}^{N} \frac{r(z_j)}{\frac{dh}{ds}(z_j)(s - z_j)}. \tag{4.51}$$

Proof

(i) Outline of the proof of existence. Recall the process of long division for polynomials; see [6], page 64. Starting with $f(s) = g(s)/h(s)$, we use the Euclidean algorithm to write

$$g(s) = q(s)h(s) + r(s) \tag{4.52}$$

where $q(s)$ and $r(s)$ are polynomials, and either $r(s) = 0$ or the degree of $r(s)$ is strictly less than the degree of $h(s)$; hence

$$f(s) = q(s) + \frac{r(s)}{h(s)} \tag{4.53}$$

where $r(s)/h(s)$ is strictly proper. Now by the fundamental theorem of algebra [6] page 101,

$$h(s) = b \prod_{j=1}^{N} (s - \lambda_j)^{m_j} \tag{4.54}$$

where the $\lambda_j \in \mathbb{C}$ are distinct. One can derive from this a partial fractions decomposition by repeatedly using the division algorithm for polynomials, as we discuss in Proposition 6.24. By such a process, we obtain coefficients $a_{j,k}$ such that

$$\frac{r(s)}{h(s)} = \sum_{j=1}^{N} \sum_{k=1}^{m_j} \frac{a_{j,k}}{(s - \lambda_j)^k}, \tag{4.55}$$

with integers m_j that give the multiplicity of the poles λ_j. The poles λ_j and coefficients $a_{j,k}$, are unique, as one can show by considering the Cauchy integral formula to $(z - \lambda_j)^p f(z)$ to suitably chosen contour integrals about the λ_j.

(ii) The function

$$g(s) = f(s) - \sum_{j=1}^{N} \frac{r(z_j)}{\frac{dh}{ds}(z_j)(s - z_j)}$$

is proper and rational, with $h(s) \to 0$ as $s \to \infty$. The only possible poles are simple poles at the z_j, but we find that $(s - z_j)g(s) \to 0$ as $s \to z_j$, so there are no such poles. Hence $g(s)$ is holomorphic and bounded on $\mathbb{C}$, so $g(s)$ is constant by Liouville's theorem. But the constant must be zero, since $g(s) \to 0$ as $s \to \infty$.

□

This result is very useful, so long as one can locate the poles λ_j. Using Corollary 6.27, one can check simplicity of the poles, as in the hypothesis (ii) of Proposition 4.22, without locating the poles λ_j.

Corollary 4.23 (Laplace Transforms Which Are Strictly Proper Rationals) *Let $y(t)$ be a function of the form*

$$y(t) = \sum_{j=1}^{n} a_{j,n_j} t^{n_j} e^{\lambda_j t}. \tag{4.56}$$

Then the Laplace transform Y of y is a strictly proper rational function with partial fraction decomposition

$$Y(s) = \sum_{j=1}^{n} \frac{n_j! a_{j,n_j}}{(s - \lambda_j)^{n_j+1}}. \tag{4.57}$$

Conversely, all strictly proper rational functions arise thus.

Proof Let $\Re\lambda < \beta$ and $\beta < s$. We substitute $z = (s - \lambda)t$ and find

$$\begin{aligned}\int_0^\infty t^n e^{\lambda t} e^{-st}\, dt &= \int_0^\infty t^n e^{-(s-\lambda)t}\, dt \\ &= \frac{1}{(s-\lambda)^{n+1}} \int_0^\infty z^n e^{-z} dz;\end{aligned}$$

this can be justified by Cauchy's theorem from complex analysis. Integrating by parts, we obtain

$$\begin{aligned}\int_0^\infty t^n e^{\lambda t} e^{-st}\, dt &= \frac{1}{(s-\lambda)^{n+1}} \Big[-z^n e^{-z} \Big]_0^\infty + \frac{n}{(s-\lambda)^{n+1}} \int_0^\infty z^{n-1} e^{-z}\, dz \\ &= 0 + \frac{n}{(s-\lambda)^{n+1}} \Big[-z^{n-1} e^{-z} \Big]_0^\infty + \frac{n(n-1)}{(s-\lambda)^{n+1}} \int_0^\infty z^{n-2} e^{-z}\, dz\end{aligned}$$

and so until

$$\int_0^\infty t^n e^{\lambda t} e^{-st}\,dt = \frac{n!}{(s-\lambda)^{n+1}}. \tag{4.58}$$

□

In the next section, we give an inversion formula via a contour integral when the poles are simple.

4.7 Dirichlet's Integral and Heaviside's Expansions

Inverting the Laplace transform involves the following crucial calculation. We write $sgn(t) = 1$ for $t > 0$ and $sgn(t) = -1$ for $t < 0$.

Lemma 4.24 (Dirichlet's Integral)

$$\lim_{R\to\infty} \int_0^R \frac{\sin t\omega}{\omega} d\omega = \frac{\pi}{2} sgn(t). \tag{4.59}$$

Proof By a simple scaling argument, we can replace $t\omega$ by ω, taking account of the change in sign of the resulting integral when $t < 0$. The function $f(s) = e^{-s}/s$ is holomorphic except for a simple pole at $s = 0$, so we use the contour

$$\Gamma = [-Ri, -\delta i] \oplus S_\delta \oplus [\delta i, Ri] \oplus (-S_R), \tag{4.60}$$

where $0 < \delta < R$ and the indentation around $s = 0$ ensures that 0 lies to the left of Γ as in Fig. 4.1. By Cauchy's Theorem,

$$\int_\Gamma \frac{e^{-s}}{s} ds = 0.$$

We can express this integral as the sum of the four parts corresponding to the arcs in (4.60), taking first the two segments on the imaginary axis with $s = i\omega$

$$\begin{aligned}\int_{[-Ri,-\delta i]} + \int_{[\delta i, Ri]} \frac{e^{-s}}{s} ds &= \int_{-R}^{-\delta} + \int_\delta^R \frac{e^{-i\omega}}{\omega} d\omega \\ &= \int_\delta^R \frac{e^{-i\omega} - e^{i\omega}}{\omega} d\omega \\ &= -2i \int_\delta^R \frac{\sin\omega}{\omega} d\omega;\end{aligned}$$

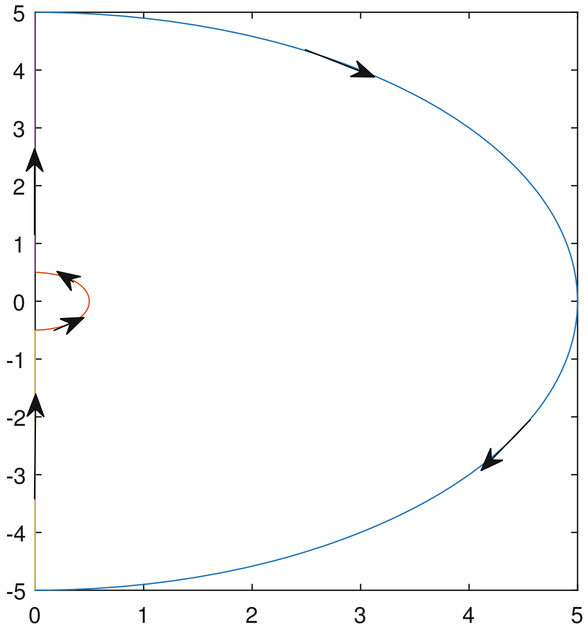


Fig. 4.1 Semicircular contour in the left half-plane with indentation

then taking the integral round the indentation, with $s = \delta e^{i\theta}$ for $-\pi/2 \leq \theta \leq \pi/2$,

$$\int_{S_\delta} \frac{e^{-s}}{s} ds = i \int_{-\pi/2}^{\pi/2} e^{-\delta e^{i\theta}} d\theta$$
$$\to \pi i \qquad (\delta \to 0+);$$

and finally taking the integral round the large semicircle with $s = Re^{i\theta}$ for $-\pi/2 \leq \theta \leq \pi/2$

$$\int_{S_R} \frac{e^{-s}}{s} ds = i \int_{-\pi/2}^{\pi/2} e^{-Re^{i\theta}} d\theta$$
$$= i \int_{-\pi/2}^{\pi/2} e^{-R\cos\theta - iR\sin\theta} d\theta;$$

where the final integral is bounded by

$$2 \int_0^{\pi/2} e^{-R\cos\theta} d\theta = 2 \int_0^{\pi/2} e^{-R\sin\phi} d\phi$$
$$\leq 2 \int_0^{\pi/2} e^{-2R\phi/\pi} d\phi$$
$$\leq \frac{\pi}{R}.$$

Combining these identities, we deduce that

$$\int_0^R \frac{\sin\omega}{\omega}d\omega = \frac{\pi}{2} + O\Big(\frac{1}{R}\Big) \qquad (R \to \infty). \tag{4.61}$$

□

Definition 4.25 (Hilbert Transform) The Hilbert transform is defined by the Cauchy principal value integral

$$Hg(y) = \lim_{\varepsilon\to 0+}\Big\{\int_{-\infty}^{y-\varepsilon} + \int_{y+\varepsilon}^{\infty}\Big\}\frac{g(\omega)}{y-\omega}\frac{d\omega}{\pi} \tag{4.62}$$

for $g \in L^2(\mathbb{R})$.

Example 4.26 It follows from Dirichlet's integral 4.24 with the change of variable $\omega \mapsto \omega - y$ that

$$\lim_{\varepsilon\to 0+}\Big\{\int_{-\infty}^{y-\varepsilon} + \int_{y+\varepsilon}^{\infty}\Big\}\frac{e^{it\omega}}{y-\omega}\frac{d\omega}{\pi} = -i\,\mathrm{sgn}(t)e^{ity}. \tag{4.63}$$

The following is a useful inversion formula for special Laplace transforms, which we extend in Proposition 6.55.

Proposition 4.27 (Heaviside's Expansion Theorem) *Let $F(s) = p(s)/q(s)$ be a strictly proper rational function with simple poles at $z_j \in LHP$ for $j = 1,\dots,n$, and let*

$$f(t) = \lim_{R\to\infty}\frac{1}{2\pi i}\int_{-iR}^{iR} e^{st}F(s)\,ds \qquad (t > 0). \tag{4.64}$$

Then $F(s)$ is the Laplace transform of $f(t)$, and

$$f(t) = \sum_{j=1}^{n}\frac{p(z_j)}{\frac{dq}{ds}(z_j)}e^{tz_j} \qquad (t > 0). \tag{4.65}$$

Proof The rational function has a partial fractions decomposition

$$F(s) = \sum_{j=1}^{n}\frac{p(z_j)}{\frac{dq}{ds}(z_j)(s-z_j)} \tag{4.66}$$

where we have computed the residues of $p(s)/q(s)$ at the simple poles z_j by the formula of Proposition 4.22. Note that e^{st} appears instead of e^{-st}, and for $t > 0$ the function e^{st} is bounded for s in the left half-plane. We integrate $e^{st}F(s)$ round a semicircular contour of large radius $R > 0$ in the LHP, which winds round all the poles. As in the calculation of Lemma 4.24 Dirichlet's integral , the contribution of the semicircular arc tends to 0 as $R \to \infty$. The poles are all simple, and residues of $e^{st}F(s)$ are

$$\mathrm{Res}\big\{e^{st}F(s); z_j\big\} = \lim_{s\to z_j} \frac{(s-z_j)p(s)e^{st}}{q(s)} = \frac{p(z_j)}{\frac{dq}{ds}(z_j)} e^{tz_j}, \tag{4.67}$$

so the formula for $f(t)$ follows from Cauchy's Residue Theorem. Then $F(s)$ coincides with the Laplace transform of $f(t)$ by a simple case of Corollary 4.23. If $F(s) = O(1/s^2)$ as $s \to \infty$, then the integral (4.64) is absolutely convergent. □

Remark 4.28

(i) If $F(s)$ has multiple poles, then the formula (4.64) is still valid, but the expansion formula needs amending with more complicated formulas for the residues at the multiple poles. See Proposition 6.55 for details.
(ii) The reader will find it instructive to extend to the case in which the poles are possibly in RHP; it only takes a translation in the variable s.

Heaviside's expansion gives a succinct solution of some differential equations. For polynomial $q(s) = a_m s^m + \cdots + a_0$ we write

$$q\Big(\frac{d}{dt}\Big) = a_m \frac{d^m}{dt^m} + \cdots + a_1 \frac{d}{dt} + a_0,$$

as in the style of Proposition 4.17.

Corollary 4.29 (Heaviside's Solution) *Suppose that q is a complex polynomial of degree m with all its zeros simple and in LHP, and suppose that p is a complex polynomial of degree n where $n < m$ and let $f(t)$ be as in (4.65). Then for any n-times continuously differentiable input u of the type (E), the unique solution of the initial value problem*

$$q\Big(\frac{d}{dt}\Big)y = p\Big(\frac{d}{dt}\Big)u \tag{4.68}$$

$$y(0) = \frac{dy}{dt}(0) = \cdots = \frac{d^{m-1}y}{dt^{m-1}}(0) = 0 = u(0) = \cdots = \frac{d^{n-1}u}{dt^{n-1}}(0) \tag{4.69}$$

is

$$y(t) = \int_0^t f(t-\tau)u(\tau)\,d\tau \qquad (t > 0). \tag{4.70}$$

Proof By Theorem 2.40, there exists a unique solution, which belongs to (E) by section 4.5. The Laplace transform of the differential equation is $q(s)Y(s) =$

$p(s)U(s)$ so $Y(s) = F(s)U(s)$, so the solution is expressed as a convolution with f as in (4.65). □

4.8 Final Value Theorem

The results of this section are useful for finding or checking constants in solutions of differential equations. For a continuous and bounded function $f : (0, \infty) \to \mathbb{C}$, we can interpret

$$sF(s) = s \int_0^\infty e^{-st} f(t)dt \qquad (s > 0) \tag{4.71}$$

as a weighted average of f, since $s \int_0^\infty e^{-st} dt = 1$. This suggests that the values of $sF(s)$ should be strongly related to the values of f as $s \to 0$ or $s \to \infty$. In the literature there are two types of results about limits of Laplace transforms:

(i) Abelian theorems, which show that $f(t)$ has certain limits as $t \to 0+$ or $t \to \infty$;
(ii) Tauberian theorems, which have the hypothesis that $f(t)$ has certain limits as $t \to 0+$ or $t \to \infty$, and conclusions that $F(s)$ has certain limits as $s \to \infty$ or $s \to 0+$.

It is important not to confuse the hypotheses and conclusions. See [56].

There following two results are Tauberian theorems for the Laplace transform, and may be applied with due care about the hypotheses. See [53].

Proposition 4.30 (Final Value Theorem) *Suppose that $f : (0, \infty) \to \mathbb{C}$ is a continuous and bounded function such that $f(t) \to L$ as $t \to \infty$ for some $L \in \mathbb{C}$. Then the Laplace transform $F(s)$ of f satisfies* $\lim_{s \to 0+} sF(s) = L$.

Proof Take M such that $|f(t)| \le M$ for all $t > 0$, and let $\varepsilon > 0$. Then we split the integral

$$sF(s) - L = \int_0^\infty (f(t) - L)se^{-st}\, dt \tag{4.72}$$

into $\int_0^R + \int_R^\infty$; where $R > 0$ is to be chosen. We take R such that $|f(t) - L| < \varepsilon$ for all $t > R$, so

$$\Big| \int_R^\infty (f(t) - L)se^{-st}\, dt \Big| \le \varepsilon \int_R^\infty se^{-st}\, dt \le \varepsilon; \tag{4.73}$$

when we have $|f(t) - L| \le M + |L| \le 2M$, so

$$\Big| \int_0^R (f(t) - L)se^{-st}\, dt \Big| \le 2M \int_0^R se^{-st}\, dt = 2M(1 - e^{-sR}); \tag{4.74}$$

there exists $s_0 > 0$ such that $2M(1 - e^{-sR}) \leq \varepsilon$ for all $0 < s < s_0$; hence the result. □

Proposition 4.31 (Initial Value Theorem) *Let f be continuous on $[0, \infty)$ and suppose that f satisfies (E). Then the Laplace transform $F(s)$ satisfies*

$$f(0) = \lim_{s \to \infty} sF(s). \tag{4.75}$$

Proof By a simple scaling, one can show that

$$sF(s) - f(0) = \int_0^\infty \big(f(x/s) - f(0)\big)e^{-x}\,dx.$$

Given $\varepsilon > 0$, and $M, \alpha > 0$ such that

$$|f(t)| \leq Me^{\alpha t} \qquad (t \geq 0), \tag{4.76}$$

consider $s > 2\alpha$ and $R > 0$. Then

$$\left| \int_R^\infty \big(f(x/s) - f(0)\big)e^{-x}\,dx \right| \leq 4Me^{-R/2} \qquad (s > 2\alpha). \tag{4.77}$$

We now choose and fix R so large that $4Me^{-R/2} < \varepsilon$. By continuity of f at 0, $f(x/s) - f(0) \to 0$ as $s \to \infty$, so there exists s_0 such that

$$\left| \int_0^R \big(f(x/s) - f(0)\big)e^{-x}\,dx \right| \leq \varepsilon \tag{4.78}$$

for all $s > s_0$.

From the preceding estimates, we deduce that for all $s > \max\{s_0, 2\alpha\}$,

$$|sF(s) - f(0)| \leq \left| \int_0^\infty \big(f(x/s) - f(0)\big)e^{-x}\,dx \right| \leq 2\varepsilon. \tag{4.79}$$

□

Example 4.32 In the context of Proposition 4.27, one can check that $f(0) = \lim_{s\to\infty} sF(s)$ and $\lim_{t\to\infty} f(t) = 0 = \lim_{s\to 0} sF(s)$ as in the initial and final value theorems.

Remark 4.33

(i) If $f : [0, \infty) \to \mathbb{C}$ is continuous and $\lim_{t\to\infty} f(t) = L$ exists, then f is bounded so that there exists $M > 0$ such that $|f(t)| \leq M$ for all $t \geq 0$. In this situation, f satisfies the hypotheses of both the initial value theorem and the final value theorem.

(ii) The hypotheses of the theorems are nevertheless different. The function $\sin t$ is continuous and bounded on $[0, \infty)$, but does not have a limit as $t \to \infty$. The function $\sin(1/t)$ is bounded and continuous on $(0, \infty)$, but does not have a limit as $t \to 0+$.

(iii) In the final value theorem, we assume that $\lim_{t\to\infty} f(t)$ exists, and identify this limit in terms of $sF(s)$ as $s \to \infty$; in the initial value theorem, we assume that $f(0) = \lim_{t\to 0+} f(t)$ exists, and express this limit in terms of $sF(s)$ as $s \to 0$. The results do not say that limits for $sF(s)$ imply existence of limits for $f(t)$.

(iv) The initial value theorem can be extended to a more subtle versions known as Watson's lemma; see [53].

4.9 Laplace Transforms of Periodic Functions

In many applications, especially to signal processing, one works with periodic functions, which have Laplace transforms with a special form. We consider a basic result and two significant examples, namely sine waves and square waves.

Definition 4.34 (Periodic Function)

(i) A piecewise continuous and nonconstant function $f : \mathbb{R} \to \mathbb{C}$ is said to be periodic with period $p > 0$ if $f(t + p) = f(t)$ for all $t \in \mathbb{R}$, and no such identity holds when p is replaced by $0 < q < p$.

(ii) A complex function $F(s)$ is said to be meromorphic if F is holomorphic apart from some poles. All rational functions are meromorphic.

Proposition 4.35 *Let f be periodic. Then f is bounded and has a Laplace transform which is a meromorphic function that satisfies*

$$\mathcal{L}f(s) = \frac{\int_0^p e^{-us} f(u)du}{1 - e^{-ps}}. \tag{4.80}$$

Proof Since f is piecewise continuous, it is bounded on $[0, p]$ so is evidently bounded on $\mathbb{R}$ as the graph repeats itself when we translate it to the right through steps of p. Then f when restricted to $(0, \infty)$, has a Laplace transform, which we compute by splitting the range of integration into intervals $[np, (n + 1)p)$, on which we change variables to $t = np + u$. We have

$$\begin{aligned}\mathcal{L}f(s) &= \int_0^\infty e^{-st} f(t)dt \\ &= \sum_{n=0}^\infty \int_{np}^{(n+1)p} e^{-st} f(t)dt \\ &= \sum_{n=1}^\infty e^{-nps} \int_0^p e^{-su} f(np + u)du,\end{aligned}$$

where $f(np+u) = f(u)$ by periodicity, so all the integral are equal and we can sum the geometric series to obtain

$$\mathcal{L}f(s) = \frac{\int_0^p e^{-su} f(u)du}{1-e^{-ps}}$$

where the numerator is an entire function of s, and the denominator is an entire function of s with zeros at $e^{sp} = 1$; that is $s = 2\pi ni/p$ where $n \in \mathbb{Z}$. Therefore the Laplace transform is a meromorphic function with possible poles on the imaginary axis, equally spaced with gaps $2\pi/p$ between them. However, the possible poles may be canceled by zeros on the numerator. □

Example 4.36 (Sine Waves) The function of periodic functions include $\sin(2\pi t/p)$ is periodic with period $p > 0$, and we have

$$\mathcal{L}(\sin(2\pi t/p))(s) = \frac{2\pi/p}{(s-2\pi i/p)(s+2\pi i/p)} \tag{4.81}$$

which has only two poles, at $\pm 2\pi i/p$, so in this case all but two of the possible poles are canceled out.

Example 4.37 (Square Waves) Consider the initial value problem for $k > 0$

$$\frac{d^2y}{dt^2} + k^2 y(t) = u(t)$$
$$y(0) = y'(0) = 0$$

for a bounded and piecewise continuous input u. Then the solution is

$$y(t) = \int_0^t \frac{\sin k(t-\tau)}{k} u(\tau)d\tau \qquad (t > 0). \tag{4.82}$$

One can verify that this works by differentiating twice. To derive the formula, we take Laplace transforms, and obtain

$$s^2Y(s) + k^2Y(s) = U(s), \tag{4.83}$$

so

$$Y(s) = \frac{1}{k}\frac{k}{s^2+k^2}U(s), \tag{4.84}$$

where $k/(s^2+k^2)$ is the Laplace transform of $\sin kt$, so we obtain the solution as the convolution of $(\sin kt)/k$ with $u(t)$.

In particular, we can take the square wave input

$$u(t) = 1 \qquad (t \in [0,1) \cup [2,3) \cup [4,5) \cup \dots)$$
$$-1 \qquad (t \in [1,2) \cup [3,4) \cup [5,6) \cup \dots$$

which is known as the square wave on account of its graph, which resembles the top of the curtain wall of a medieval castle. One can write

$$u(t) = \sum_{k=0}^{\infty} \big(H(t-2k) - 2H(t-1-2k) + H(t-2-2k) \big) \qquad (t > 0) \tag{4.85}$$

which is a finite sum for each $t > 0$ since $H(t-n) = 0$ for all $n > t$.

The Laplace transform is

$$\begin{aligned} U(s) &= \sum_{n=0}^{\infty} \int_{2n}^{2n+1} e^{-st} dt - \sum_{n=1}^{\infty} \int_{2n-1}^{2n} e^{-st} dt \\ &= \frac{1}{s} \sum_{n=0}^{\infty} (e^{-2ns} - e^{-(2n+1)s}) - \frac{1}{s} \sum_{n=1}^{\infty} (e^{-(2n-1)s} - e^{-2ns}) \\ &= \frac{1 - 2e^{-s} + e^{-2s}}{s(1 - e^{-2s})} \\ &= \frac{1}{s} \tanh \frac{s}{2}, \end{aligned}$$

where we used geometric series to make the summation. This calculation is easily justified by uniform convergence since the partial sums of the series for $u(t)$ are uniformly bounded. Hence $sU(s) = \tanh(s/2)$, and since $u(t)$ is right-continuous at $t = 0+$, we can use the initial value theorem to confirm that $u(0+) = \lim_{s \to \infty} sU(s) = 1$. Whereas $sU(s) \to 0$ as $s \to 0+$, the square wave does not have a limit as $t \to \infty$, and we cannot apply the final value theorem.

For large $t > 0$, we choose N to be the largest integer such that $2N + 2 \leq t$, and we write the solution as

$$y(t) = \int_0^{2N+2} \frac{\sin k(t-\tau)}{k} u(\tau) d\tau + \int_{2N+2}^{t} \frac{\sin k(t-\tau)}{k} u(\tau) d\tau \tag{4.86}$$

where the final integral is bounded independent of t, and the other integral is evaluated by splitting $[0, 2N+2]$ into subintervals of length 2. A typical subinterval contributes

$$\begin{aligned}\int_{2n}^{2n+2}\frac{\sin k(t-\tau)}{k}u(\tau)d\tau &= \int_{2n}^{2n+1}\frac{\sin k(t-\tau)}{k}d\tau - \int_{2n+1}^{2n+2}\frac{\sin k(t-\tau)}{k}d\tau\\ &= \frac{1}{k^2}\Big(2\cos k(2n+1-t)-\cos k(2n-t)\\ &\quad -\cos k(2n+2-t)\Big)\\ &= \frac{1}{k^2}\Big(-2\sin(k/2)\sin k(2n+1/2-t)\\ &\quad +2\sin(k/2)\sin k(2n+3/2-t)\Big)\\ &= \frac{4}{k^2}\sin^2(k/2)\cos k(2n+1-t).\end{aligned}$$

- When $k \neq (2m+1)\pi$ for $m = 0, 1, \ldots,$ we have $\cos(k/2) \neq 0$, and we continue with

$$\begin{aligned}\int_{2n}^{2n+2}\frac{\sin k(t-\tau)}{k}u(\tau)d\tau &= 2\frac{\sin^2(k/2)}{k^2\sin k}2\sin k\cos k(2n+1-t)\\ &= \frac{1}{k^2}\tan(k/2)\big(\sin k(2n+2-t)-\sin k(2n-t)\big),\end{aligned}$$

  and we deduce that

$$\begin{aligned}\int_{0}^{2N+2}\frac{\sin k(t-\tau)}{k}u(\tau)d\tau &= \sum_{n=0}^{N}\int_{2n}^{2n+2}\frac{\sin k(t-\tau)}{k}u(\tau)d\tau\\ &= \sum_{n=0}^{N}\frac{1}{k^2}\tan(k/2)\big(\sin k(2n+2-t)\\ &\quad -\sin k(2n-t)\big)\\ &= \frac{1}{k^2}\tan(k/2)\big(\sin k(2N+2-t)+\sin kt\big),\end{aligned}$$

  which is bounded independent of t, so the solution $y(t)$ is bounded.
- When $k = (2m+1)\pi$ for some $m = 0, 1, 2, \ldots,$ we have $\cos(k/2) = 0$, and

$$\begin{aligned}\int_{2n}^{2n+2}\frac{\sin k(t-\tau)}{k}u(\tau)d\tau &= \frac{4}{k^2}\Big(\cos(2m+1)(2n+1)\pi\cos(2m+1)t\pi\\ &\quad +\sin(2m+1)(2n+1)\pi\sin(2m+1)t\pi\Big)\\ &= -\frac{4}{k^2}\cos(2m+1)t\pi,\end{aligned}$$

and we deduce that

$$\int_0^{2N+2} \frac{\sin k(t-\tau)}{k} u(\tau)d\tau = \sum_{n=0}^{N} -\frac{4}{k^2}\cos(2m+1)t\pi$$
$$= -\frac{4(N+1)}{k^2}\cos(2m+1)t\pi,$$

so $y(t)$ oscillates unboundedly as $t \to \infty$, and we have a resonance effect for such k.

Remark 4.38 Given a piecewise continuous function $f : [0, p] \to \mathbb{C}$, there is a natural extension of f to a periodic function $f : \mathbb{R} \to \mathbb{C}$. Given $t \in \mathbb{R}$, there exist a unique $n \in \mathbb{Z}$ and $u \in [0, p)$ such that $t = np + u$. For $u \in (0, p)$, we define $f(t) = f(u)$. For $u = 0$, we can define either:

(i) $f(np) = \lim_{v\to 0+} f(v)$, if one wants a cadlag function (continuous from the right with limits from the left); or
(ii) $f(np) = \lim_{v\to 0+}(1/2)(f(v) + f(p-v))$, which is useful in the context of Fourier series.

4.10 Fourier Cosine Transform

The Fourier transform is fundamentally important in signal processing and theory of linear differential equations. In this section we give some fundamental results, including an inversion theorem. The Laplace transform and Fourier transform are different, but they are related; in particular, we obtain the uniqueness theorem for Laplace transforms via the Fourier inversion formula. Throughout this section, we suppose that $f : [0, \infty) \to \mathbb{C}$ is a piecewise continuous function such that $\int_0^\infty |f(t)|dt$ converges, so that f is integrable. We regard $t > 0$ as time and introduce $\omega \in \mathbb{R}$ as the angular frequency. The function $\cos(\omega t)$ is a periodic function of t with period $2\pi/\omega$ for $\omega > 0$. Models described in terms of ω referred to as frequency domain models. The Fourier transform takes us from time domain models to frequency domain models.

Definition 4.39 We define the Fourier cosine transform of f by

$$\phi(\omega) = \int_0^\infty \cos(\omega t) f(t)\,dt \qquad (\omega \in \mathbb{R}). \tag{4.87}$$

Proposition 4.40 *The Fourier cosine transform is a continuous and bounded function.*

Proof By the triangle inequality, we have a bound

$$|\phi(\omega)| \leq \int_0^\infty |\cos(\omega t)||f(t)|dt \leq \int_0^\infty |f(t)|dt \qquad (\omega \in \mathbb{R}), \tag{4.88}$$

so the integral is absolutely convergent and uniformly bounded in ω.

By continuity of cosine, the partial integrals

$$\phi_n(\omega) = \int_0^n \cos(\omega t) f(t)dt \qquad (\omega \in \mathbb{R}) \tag{4.89}$$

are all continuous and

$$|\phi(\omega) - \phi_n(\omega)| \leq \int_n^\infty |\cos(\omega t)||f(t)|dt \leq \int_n^\infty |f(t)|dt \qquad (\omega \in \mathbb{R}) \tag{4.90}$$

so $\phi_n \to \phi$ uniformly on $[0, \infty)$ as $n \to \infty$. Hence ϕ is also continuous. □

Remark 4.41

(i) Suppose that f has Laplace transform F and cosine transform ϕ. Then from $\cos(\omega t) = 2^{-1}(e^{i\omega t} + e^{-i\omega t})$, we deduce that

$$\phi(\omega) = 2^{-1}\big(F(i\omega) + F(-i\omega)\big) \qquad (\omega \in \mathbb{R}). \tag{4.91}$$

(ii) Suppose that f is integrable and real-valued on $(0, \infty)$, so has Laplace transform is $F(s)$; then with $s = i\omega$ and $\omega \in \mathbb{R}$, we have

$$\Re F(s) = \int_0^\infty f(t)\cos(\omega t)\,dt = \phi(\omega), \tag{4.92}$$

namely the Fourier cosine transform.

Hence we can convert Laplace transform formulas into Fourier cosine formulas.

Example 4.42

(i) By integrating twice by parts, one can show that $f(t) = e^{-t}$ has Fourier cosine transform $\phi(\omega) = 1/(1 + \omega^2)$.
(ii) Let $\mathbb{I}_{(0,a)}$ be the indicator function of $(0, a)$, so $\mathbb{I}_{(0,a)}(t) = 1$ for $t \in (0, a)$ and $\mathbb{I}_{(0,a)}(t) = 0$ for $t \in \mathbb{R} \setminus (0, a)$. Then the Fourier cosine transform is $\sin(a\omega)/\omega$. Note that $\sin(a\omega)/\omega \to a$ as $\omega \to 0$, so we have continuity.

Definition 4.43 The unnormalized sinc function is

$$\operatorname{sinc}(t) = \frac{\sin t}{t} \qquad (t \in \mathbb{R}). \tag{4.93}$$

This is sometimes called the cardinal sine function.

Theorem 4.44 (Integrated Inversion Formula)

$$\int_0^x f(t)\,dt = \lim_{R\to\infty} \frac{2}{\pi} \int_0^R \frac{\sin(x\omega)}{\omega}\phi(\omega)d\omega. \tag{4.94}$$

Proof By Lemma 4.24, we have

$$\begin{aligned}\lim_{R\to\infty} \int_0^R \frac{\sin(u\omega)}{\omega} d\omega &= \pi/2 \qquad (u > 0)\\ &= -\pi/2 \qquad (u < 0).\end{aligned}$$

Then we have

$$\begin{aligned}\int_0^R \frac{\sin(x\omega)}{\omega}\phi(\omega)d\omega &= \int_0^R \frac{\sin(x\omega)}{\omega} \int_0^\infty \cos(\omega t) f(t)dtd\omega\\ &= \int_0^\infty \Big(\int_0^R \frac{\sin(x\omega)\cos(\omega t)}{\omega} d\omega\Big) f(t)dt\\ &= \int_0^\infty \Big(\int_0^R \frac{\sin((x-t)\omega) + \sin((x+t)\omega)}{2\omega} d\omega\Big) f(t)dt\end{aligned}$$

where we have changed the order of integration and used a trigonometric addition rule. The inside integral has limit

$$\begin{aligned}\int_0^R \frac{\sin((x-t)\omega) + \sin((x+t)\omega)}{2\omega} d\omega &\to \frac{\pi}{2} \frac{\text{sgn}(x-t) + \text{sgn}(x+t)}{2}\\ &= \frac{\pi}{2}\mathbb{I}_{(-x,x)}(t)\end{aligned}$$

as $R \to \infty$. From integration theory, we deduce that

$$\lim_{R\to\infty} \int_0^R \frac{\sin(x\omega)}{\omega}\phi(\omega)d\omega = \frac{\pi}{2}\int_0^\infty \mathbb{I}_{(-x,x)}(t) f(t)dt = \frac{\pi}{2}\int_0^x f(t)dt. \tag{4.95}$$

□

Corollary 4.45 (Laplace Transform Uniqueness) *Suppose that f has Laplace transform F, where $F(s) = 0$ for all $s \in (s_0, \infty)$ for some $s_0 > 0$. Then $f(t) = 0$ at all points of continuity of f.*

Proof By Propositions 4.6 and 4.40, F is holomorphic on $\{s : \Re s > 0\}$ and continuous on the closed left half-plane $\{s : \Re s \geq 0\}$. By the principle of isolated zeros, we deduce that $F(s) = 0$ on $\{s : \Re s \geq 0\}$. In particular, $\phi(\omega) = 2^{-1}(F(i\omega) + F(-i\omega)) = 0$ when $s = i\omega$ is on the imaginary axis, so from the integrated inversion formula (4.94), we have

$$\int_0^x f(t)dt = 0 \qquad (x \geq 0). \tag{4.96}$$

By the Fundamental Theorem of Calculus, we have $f(t) = 0$ at all points t at which f is continuous. See Exercise 4.23 for an inversion formula.

This proves the Laplace uniqueness Theorem 4.11 on Sect. 4.4, and in Corollary 9.5, we prove a stronger version of this result due to Lerch. □

4.11 Impulse Response

Proposition 4.46 *For a stable system* (A, B, C, D) *as in Sect.* 5.6 *let* $\phi(t) = D\delta_0(t) + C\exp(tA)B$. *Then*

$$\mathcal{L}(\phi)(s) = T(s). \tag{4.97}$$

Proof We have

$$\begin{aligned}\int_0^\infty e^{-st}\phi(t)dt &= D + C\int_0^\infty e^{-st}\exp(tA)dt B\\ &= D + C(sI - A)^{-1}B = T(s),\end{aligned}$$

by the Proposition 3.10. □

This ϕ frequently appears in the literature, without having a ubiquitous name. One can call ϕ a scattering function, by analogy with similar functions which appear in physics; alternatively the impulse response function as it is the signal that arises from an input of δ_0.

We consider some standard inputs u for the SISO system $(A, B, C, 0)$, where the initial condition of the state is $X(0) = 0$. Let $\phi(t) = C\exp(tA)B$.

(1) Let $u_1(t) = H(t)$, so

$$y_1(t) = \int_0^t \phi(t - v)dv = \int_0^t \phi(v)dv. \tag{4.98}$$

(2) Let $u_2(t) = \delta_0(dt)$, which is the unit impulse. Then $y_2(t) = \phi(t)$, which is the derivative of the output in case (1).

(3) Let $u_3(t) = e^{i\nu t}$ for real angular frequency ν. In this case, it is more helpful to consider the Laplace transforms Y_3 of the output y_3, $T(s)$ of ϕ and U_3 of u_3, so that

$$Y_3(s) = T(s)U_3(s) = C(sI - A)^{-1}B\frac{1}{s - i\nu}. \tag{4.99}$$

Evidently, $Y_3(s)$ has a possible pole at $s = i\nu$, as we discuss in the next chapter.

Example 4.47 Suppose that we use an (A, B, C, D) model for a pension fund, in which the employee contributes an input $u(t)$ from the start of employment at time t_0 until retirement at time $t = 0$, and then draws a pension $y(t)$ for $t > 0$. After retirement the contributions cease, so $U(t) = 0$ so amount of money in the pension fund is the state variable $X(t)$, which satisfies

$$X(t) = \exp((t - t_0)A)X(t_0) + \int_{t_0}^{\min\{t,0\}} \exp((t - \tau)A)BU(\tau)d\tau, \tag{4.100}$$

and the output, namely the pension is

$$y(t) = CX(t) = C\exp((t - t_0)A)X(t_0) + \int_{t_0}^{\min\{0,t\}} C\exp((t - \tau)A)BU(\tau)d\tau. \tag{4.101}$$

If we assume that t_0 is in the remote past, and A is stable, then $\exp((t - t_0)A) \to 0$ as $t_0 \to -\infty$, so we are therefore led to consider

$$y(t) = \int_0^\infty C\exp((t + v)A)BU(-v)dv \tag{4.102}$$

where we have substituted $v = -\tau$. With $\phi(t) = C\exp(tA)B$ and $f(v) = U(-v)$, we have

$$y(t) = \int_0^\infty \phi(t + v)f(v)dv. \tag{4.103}$$

In this formula, $t + v$ is the total time elapsed between payment of a pension contribution and a receipt of the pension.

4.12 Transmitting Signals

(i) Morse. Suppose that we have a radio transmitter that is able to transmit radio waves at angular frequency ω_c. We use this to send out short pulses called dots of duration d seconds and longer pulses called dashes of duration D seconds.

The letters of the alphabet can be represented by specific combinations of dots and dashes, in Morse code. A signal consists of dots and dashes emitted at times t_j for $t_1 < \cdots < t_N$, so the signal is

$$a \sum_{j=1}^{n} \mathbb{I}_{[t_j, t_j + d_j]}(t) \sin \omega_c t \tag{4.104}$$

for some $a > 0$ where $d_j \in \{d, D\}$. The signal is obtained from the carrier $\sin \omega_c t$ by multiplying by an on-off switch, known as Morse key. The receiver records the transmission and communicates this by a loudspeaker to a human receiver, who identifies the pattern of dots and dashes in the signal as letters, and thus reconstructs the message text. This system was used in radio communication in the first half of the 20th century, particularly in maritime and military contexts. The advantage is that only very simple transmitters and receivers are required, and the message can be interpreted when the signal is rather faint. The disadvantage is that one can only communicate text, and the rate of communication is slow.

(ii) Amplitude modulation (AM). Suppose that we wish to communicate sound waves at angular frequency ω_m , such as the middle C note of a piano has 264 Hz so $\omega_m = 2\pi(264)$ and the wavelength is $1.25m$. We transmit a carrier signal $\sin \omega_c$ as above, but we modulate the amplitude of the signal at the angular frequency ω, so that the combined signal is

$$(A + a \sin \omega_m t) \sin \omega_c t. \tag{4.105}$$

For instance, Radio 4 uses long wave $1514m$ at frequency $198kHz$, so the angular frequency of the carrier wave is much larger than the angular frequency of the signal. The input into the transmitter derives from electrical signals from microphones, and the receiver reverses the process by broadcasting the received signal via a loudspeaker. This system is effective for transmitting the spoken word, and requires relatively simple equipment.

(iii) Frequency modulation (FM). Let $x(t)$ be a signal with polar decomposition $x(t) = A(t)e^{i\theta(t)}$; then we define the instantaneous angular frequency to be $\frac{d\theta}{dt}$. Suppose in particular that we have a carrier wave $e^{i\omega_c t}$ which we modulate by adding a phase $\phi(t)$ so that $\theta(t) = \omega_c t + \phi(t)$ with instantaneous angular frequency $\frac{d\theta}{dt} = \omega_c + \frac{d\phi}{dt}$. Given a bounded and continuous function $m : [0, \infty) \to \mathbb{R}$, we can choose $\phi(t) = \int_0^t m(u)du$ so that $\theta(t) = \omega_c t + \phi(t)$ has instantaneous angular frequency

$$\frac{d\theta}{dt} = \omega_c + m(t); \tag{4.106}$$

the carrier frequency ω_c is thus modulated by the signal $m(t)$. If we choose ω_c so that $|m(t)| < \omega_c$, then θ is a strictly increasing and continuously differentiable function. This is the basic principle underlying frequency modulation, which is used for radio transmission, especially for broadcasting music with high fidelity. Radio 3 uses very high frequency transmission of 90 MHz so the wavelength of the carrier signal is about $3.33m$. For comparison, the highest note on a piano has frequency of about $4185Hz$, so the frequency of the carrier signal is much larger than the modulating frequency.

4.13 Exercises

Exercise 4.1

(i) Calculate the Laplace transforms of $\cos 2\omega t$ from the definition, and
(ii) deduce the Laplace transform of $\sin^2 \omega t$ where $\omega > 0$ is a constant.

Exercise 4.2 Solve the initial value problem

$$\frac{dy}{dt} - 7y = \sin 2t,$$
$$y(0) = 0;$$

by taking Laplace transforms. *Use partial fractions at the final step of the calculation.*

Exercise 4.3 Solve the integral equation

$$y(t) = e^{-2t} + \int_0^t e^{u-t} y(u)\, du,$$

where y has property (E), by using Laplace transforms.

Exercise 4.4

(i) Show that

$$\delta_b * f(t) = f(t-b)H(t-b) \qquad (t, b > 0).$$

(ii) Let $h(t) = H(t-b)$ for some $b > 0$. Show that

$$h * f(t) = H(t-b) \int_0^{t-b} f(u)du.$$

Exercise 4.5 Solve the differential equation

$$\frac{d^3y}{dt^3} + 3\frac{d^2y}{dt^2} + 6\frac{dy}{dt} + 4y = 21\frac{d^2u}{dt^2} + 39\frac{du}{dt} - 12u$$

$$\frac{d^2y}{dt^2}(0) = \frac{dy}{dt}(0) = y(0) = 0 = \frac{du}{dt}(0) = u(0)$$

by Laplace transforms.

Exercise 4.6 (Poles at -1) Let $\mathcal{R}$ be the set of functions of the form

$$f(s) = \sum_{j=1}^{n} \frac{a_j}{(1+s)^j}$$

where $n \geq 0$ and $a_j \in \mathbb{C}$.

(i) Show that $f(s)$ is differentiable, and $df/ds \in \mathcal{R}$.
(ii) Show that, for all $f(s), g(s) \in \mathcal{R}$, the sum $f(s) + g(s)$ and the product $f(s)g(s)$ also belong to $\mathcal{R}$.
(iii) Show that $f(s)$ is the Laplace transform of

$$y(t) = \sum_{j=1}^{n} \frac{a_j t^{j-1} e^{-t}}{(j-1)!} \qquad (t > 0).$$

Exercise 4.7 Let $y(t) = (1/2)\delta_0(t) + \sum_{j=1}^{\infty}(-1)^j \delta_j(t)$ be an alternating sum of Dirac point masses on the nonnegative integers.

(i) Calculate the Laplace transform $Y(s)$ of y.
(ii) Show that $Y(s)$ has zeros at even integer multiples of πi, and poles at odd integer multiples of πi.

Exercise 4.8 Compute the Laplace transform $F(s)$ of

$$f(x) = \sum_{j=1}^{n} (a_j \cos(b_j x) + c_j \sin(d_j x))$$

and consider the values $\lim_{s\to\infty} sF(s)$ and $\lim_{s\to 0+} sF(s)$ in relation to f.

Exercise 4.9 Show that

$$\int_0^x e^{-t} dt = \frac{2}{\pi} \int_0^\infty \frac{\sin(\omega x)}{\omega} \frac{d\omega}{1+\omega^2} \qquad (x > 0)$$

and deduce that

$$e^{-x} = \frac{2}{\pi}\int_0^\infty \frac{\cos(\omega x)}{1+\omega^2}d\omega \qquad (x > 0).$$

By taking $x \to 0+$, confirm that the constants are correct.

Exercise 4.10 (Carleman's Integral) See [47]. Let $y(t)$ be a bounded, continuous and integrable function that has Laplace transform $Y(u)$. By taking the Laplace transform of Y, derive the formula (the Laplace transform of the Laplace transform)

$$\mathcal{L}^2(y)(s) = \int_0^\infty \frac{1}{s+t}y(t)dt.$$

The right-hand side was studied by Carleman, and in operator theory by Power [47] and others. It leads to a fundamentally important example of a Hankel operator. In books of standard integrals, it is sometimes known as the Stieltjes transform of y; see Titchmarsh [57] page 317.

For λ such that $\Re\lambda > 0$, let

$$\Gamma f(x) = \int_0^\infty e^{-\lambda(x+y)} f(y)dy \qquad (f \in L^2(0,\infty)).$$

Then the range of Γ is $\{ce^{-\lambda x}; c \in \mathbb{C}\}$, so Γ has rank one.

Exercise 4.11 Suppose that g is piecewise continuous on $(0,\infty)$ of type (E). Suppose also that the Laplace transform $G(s)$ of g satisfies $G(s) = 0$ for all $s \in (s_0,\infty)$ for some $s_0 > 0$.

(i) Show that there exists $\kappa > 0$ such that $f(t) = e^{-\kappa t}g(t)$ is piecewise continuous and $\int_0^\infty |f(t)|dt$ converges.
(ii) Show that f has Laplace transform $F(s) = G(s+\kappa)$, and deduce that $F(s) = 0$ on $\{s : \Re s \geq 0\}$.
(iii) Using Corollary 4.45, deduce that $g(t) = 0$ at all points of continuity of h.

Exercise 4.12 (The Series and Laplace Transform of J_0) See [53]. Let Bessel's function of the first kind of order zero be defined by

$$J_0(t) = \frac{1}{2\pi}\int_0^{2\pi} \cos(t\cos\theta)d\theta.$$

(i) Show that $J_0(0) = 1$ and J_0 satisfies Bessel's equation

$$t^2\frac{d^2J_0}{dt^2} + t\frac{dJ_0}{dt} + t^2J_0(t) = 0.$$

(ii) By expanding the outer cosine function as a series, obtain the power series expansion

$$J_0(t) = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{2^{2n}(n!)^2},$$

and verify that it converges for all $t \in \mathbb{C}$.

(iii) Show that the Laplace transform satisfies

$$\mathcal{L}(J_0)(s) = \int_0^{2\pi} \frac{s}{s^2 + \cos^2\theta} \frac{d\theta}{2\pi},$$

and by calculus of residues or otherwise, deduce that

$$\mathcal{L}(J_0)(s) = \frac{1}{\sqrt{1+s^2}}.$$

(iv) Obtain this Laplace transform from the differential equation and the initial value theorem.

(v) Expand the Laplace transform as a power series in $1/s$ for $|s| > 1$ by the binomial theorem to obtain

$$\frac{1}{\sqrt{1+s^2}} = \sum_{n=0}^{\infty} \frac{(-1)^n (2n)!}{2^{2n}(n!)^2 s^{2n+1}},$$

and compare with the series that you obtain by taking the Laplace transform of the power series in (ii) term by term. This step is justified by Exercise 4.13.

Exercise 4.13 (Bessel Functions) Suppose that $\sum_{n=0}^{\infty} a_n z^n$ is a complex power series with radius of convergence $r > 0$, so the series converges for all $z \in \mathbb{C}$ such that $|z| < r$.

(i) Show that $f(t) = \sum_{n=0}^{\infty} a_n t^n/n!$ converges for all $t \in \mathbb{C}$, and that $f(t)$ for $t > 0$ determines a function of type (E) with Laplace transform

$$F(s) = \int_0^{\infty} f(t)e^{-st}dt = \sum_{n=0}^{\infty} a_n s^{-n-1} \qquad (\Re s > 1/r). \tag{4.107}$$

(ii) Show also that $\sum_{n=0}^{\infty} a_n s^{-n-1}$ converges uniformly on $\{s \in \mathbb{C} : |s| > \sigma\}$ for all $\sigma > 1/r$.

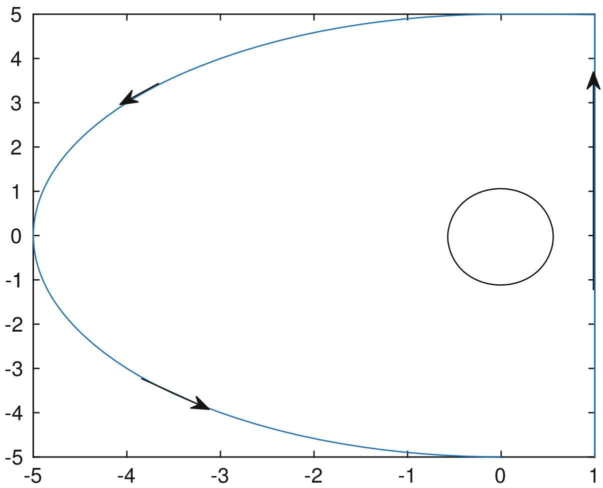


Fig. 4.2 Bromwich contour for Laplace inversion

(iii) Let $\Gamma_R(\sigma)$ be the Bromwich semicircular contour of Fig. 4.2 where $R > \sigma$, so $\Gamma_R(\sigma)$ lies outside of the circle $C(0, 1/r)$. Show that

$$\int_{\Gamma_R(\sigma)} \sum_{n=0}^{\infty} a_n s^{-n-1} e^{ts} \frac{ds}{2\pi i} = \sum_{n=0}^{\infty} \frac{a_n t^n}{n!} \qquad (t \in \mathbb{C}). \tag{4.108}$$

(iv) Show that for $t > 0$,

$$\lim_{R\to\infty} \int_{\sigma-iR}^{\sigma+iR} \sum_{n=0}^{\infty} a_n s^{-n-1} e^{ts} \frac{ds}{2\pi i} = \sum_{n=0}^{\infty} \frac{a_n t^n}{n!} \qquad (t > 0). \tag{4.109}$$

so that there is an inverse Laplace transform formula

$$f(t) = \lim_{R\to\infty} \int_{\sigma-iR}^{\sigma+iR} F(s) e^{ts} \frac{ds}{2\pi i} \qquad (t > 0). \tag{4.110}$$

This exercise is related to Borel summability as in [56] Exercise 4.21 and applies to examples of $f(t)$ such as Bessel's function $J_0(t)$ in Exercise 4.12.

(v) In the case of the Bessel function of order zero, it shows that

$$J_0(t) = \lim_{R\to\infty} \int_{1-iR}^{1+iR} \frac{e^{st}}{\sqrt{1+s^2}} \frac{ds}{2\pi i}.$$

Here the function $\sqrt{1+s^2}$ is holomorphic on $\mathbb{C} \setminus [-i, i]$ and takes opposite signs on either side of the cut $[-i, i]$; see (6.120). To reconcile this formula

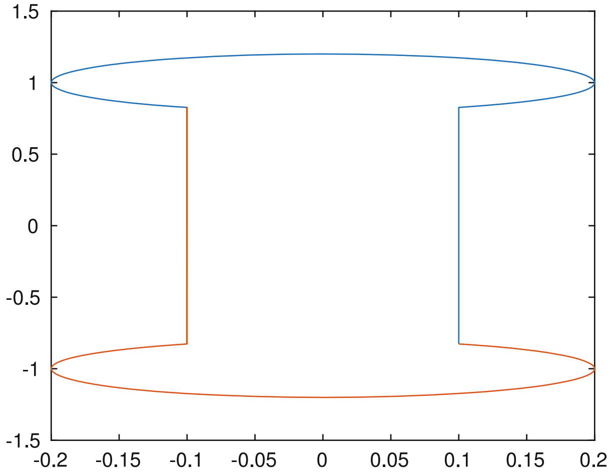


Fig. 4.3 A dog-bone contour

with the definition as given in Exercise 4.12, show that

$$\lim_{R\to\infty}\int_{1-iR}^{1+iR}\frac{e^{st}}{\sqrt{1+s^2}}\frac{ds}{2\pi i}=\int_B\frac{e^{st}}{\sqrt{1+s^2}}\frac{ds}{2\pi i}$$

where B is the dog-bone contour as in Fig. 4.3 that goes from $-i+\delta$ to $i+\delta$, goes round i on an arc of a circle, then goes down from $i-\delta$ to $-i-\delta$, then goes round $-i$ on a semicircular arc back to $-i+\delta$. Evaluate this integral by letting $\delta\to 0+$ and substituting $s=i\cos\theta$ for $-\pi\leq\theta\leq 0$, thereby recovering $J_0(t)$.

Exercise 4.14 Suppose that $(A, B, C, 0)$ is a stable linear system where A is similar to a diagonal matrix with eigenvalues $\lambda_1,\dots,\lambda_n$.

(i) Show that $\phi(t)=C\exp(tA)B$ satisfies

$$\phi(t)=\sum_{j=1}^{n}a_j e^{\lambda_j t}\qquad (t>0)$$

for some $a_j\in\mathbb{C}$.

(ii) Let f be a bounded and continuous function and let

$$y(t)=\int_0^\infty \phi(t+v)f(v)dv.$$

Show that

$$y(t) = \sum_{j=1}^{n} b_j e^{\lambda_j t} \qquad (t > 0)$$

for some $b_j \in \mathbb{C}$.

Exercise 4.15 Let $f, g : [0, \infty) \to \mathbb{R}$ be continuously differentiable functions of class (E) such that $f(0) = g(0)$. An approximate form of the telegraph equation gives rise to the initial value problem

$$\begin{aligned} \frac{\partial^2 u}{\partial x \partial y} + u = 0, \qquad & (x, y > 0), \\ u(x, 0) = f(x), \qquad & (x > 0) \\ u(0, y) = g(y), \qquad & (y > 0). \end{aligned} \tag{4.111}$$

(i) By integrating by parts and changing order of integration, show that

$$\begin{aligned} (1 + pq) \int_0^\infty \int_0^\infty e^{-px-qy} u(x, y) dx dy = f(0) + \int_0^\infty p e^{-px} f(x) dx - f(0) \\ + \int_0^\infty q e^{-qy} g(y) dy - g(0) \end{aligned} \tag{4.112}$$

(ii) Using the power series in Exercise 4.12 or otherwise, show that

$$\int_0^\infty \int_0^\infty e^{-px-qy} J_0\big(2\sqrt{xy}\big) dx dy = \frac{1}{1 + pq} \qquad (p, q > 0).$$

(iii) Find the Laplace transform of

$$\int_0^x \frac{df}{dt} J_0\big(2\sqrt{y(x - t)}\big) dt,$$

and deduce that

$$\begin{aligned} \int_0^\infty \int_0^\infty e^{-px-qy} \int_0^x \frac{df}{dt} J_0\big(2\sqrt{y(x - t)}\big) dt dx dy \\ = \frac{-f(0)}{1 + pq} + \frac{p}{1 + pq} \int_0^\infty e^{-px} f(x) dx. \end{aligned}$$

(iv) Deduce that

$$u(x,y) = f(0)J_0\big(2\sqrt{xy}\big) + \int_0^x \frac{df}{dt} J_0\big(2\sqrt{y(x-t)}\big)dt + \int_0^y \frac{dg}{dt} J_0\big(2\sqrt{x(y-t)}\big)dt$$

gives a solution of the initial value problem.

(v) Use the change of variables $x + y = \xi$ and $x - y = \eta$ to solve the inital value problem

$$\begin{aligned} \frac{\partial^2 U}{\partial \xi^2} - \frac{\partial^2 U}{\partial \eta^2} + U = 0, \qquad & (\xi \pm \eta > 0), \\ U(\xi,\xi) = f(\xi), \qquad & (\xi = \eta) \\ U(\xi,-\xi) = g(\xi), \qquad & (\xi = -\eta). \end{aligned} \tag{4.113}$$

Exercise 4.16

(i) Given the Laplace integral formula

$$\int_0^\infty \exp\Big(-av^2 - \frac{b}{v^2}\Big)dv = \frac{1}{2}\sqrt{\frac{\pi}{a}}e^{-2\sqrt{ab}} \qquad (a,b > 0),$$

deduce that for $\kappa, x > 0$

$$\int_0^\infty \frac{x\exp(-x^2/(4\kappa t))}{\sqrt{4\pi\kappa t^3}}e^{-st}dt = \exp\big(-x\sqrt{s/\kappa}\big).$$

(ii) Let $u(x,t)$ be a solution of the telegraph equation

$$\begin{aligned} \frac{\partial u}{\partial t} = \kappa\frac{\partial^2 u}{\partial x^2}, \qquad & (x, t > 0), \\ u(0,t) = f(t), \qquad & (t > 0) \\ u(x,t) \to 0, \qquad & (x \to \infty, t > 0). \end{aligned} \tag{4.114}$$

Take the Laplace transform $U(x,s) = \int_0^\infty e^{-st}u(x,t)dt$ in the t variable and show that it satisfies the ordinary differential equation in the x variable

$$\kappa\frac{\partial^2 U(x,s)}{\partial x^2} = sU(x,s)$$

where $U(x,s) \to 0$ as $x \to \infty$. By solving this, show that

$$\int_0^\infty e^{-st}u(x,t)dt = \exp\big(-x\sqrt{s/\kappa}\big)\int_0^\infty e^{-s\tau}f(\tau)d\tau.$$

(iii) Deduce that

$$u(x, t) = \int_0^t \frac{x \exp(-x^2/(4\kappa\tau))}{\sqrt{4\pi\kappa\tau^3}} f(t - \tau) d\tau.$$

Exercise 4.17 (Tent Function) For $a > 0$, let $f(t)$ be the tent function

$$\begin{aligned} f(t) = a - t \qquad & (0 < t < a); \\ a + t \qquad & (-a < t \leq 0); \\ 0 \qquad & \text{else}. \end{aligned}$$

Show that the Fourier transform of f is

$$\int_{-\infty}^{\infty} e^{-ixt} f(t) dt = \frac{4 \sin^2(ax/2)}{x^2},$$

and that

$$\int_{-\infty}^{\infty} \frac{4 \sin^2(ax/2)}{x^2} dx = 2\pi a.$$

Exercise 4.18 Show that sinc is log-concave, in the sense that

$$\frac{d^2}{dt^2} \log \operatorname{sinc}(t) \leq 0 \qquad (-\pi < t < \pi).$$

Exercise 4.19 (Bounded Convolution)

(i) Say that $f : (0, \infty) \to \mathbb{C}$ belongs to $L^1(0, \infty)$ if f is integrable and $\int_0^\infty |f(x)| dx$ is finite. Say that $u : (0, \infty) \to \mathbb{C}$ is bounded if there exists M such that $|u(t)| \leq M$ for all $t > 0$. Show that if $f \in L^1(0, \infty)$ and u is bounded and continuous, then $f * u$ is bounded.

(ii) In the context of the differential equation (4.68) suppose that the input u is bounded for $t \in [0, \infty)$. Show that the output y is also bounded.

Exercise 4.20 (Saw Tooth) The saw-tooth wave is periodic with period 2 and $u(t) = t - 1$ for $0 < t < 2$. Show that the Laplace transform of u is

$$U(s) = \frac{1}{s^2} - \frac{\coth s}{s}.$$

Using the logarithmic series or otherwise, show that

$$t - 1 = \sum_{n=-\infty; n\neq 0}^{\infty} \frac{i e^{i\pi n t}}{\pi n} \qquad (0 < t < 2).$$

Exercise 4.21 Suppose that $f(z)$ is entire and there exist $\beta, M > 0$ such that $|f(z)| \leq M e^{\beta|z|}$ for all $z \in \mathbb{C}$. By considering

$$\frac{d^n f}{dz^n}(0) = \frac{n!}{2\pi i} \int_{C(0,n)} \frac{f(z)}{z^{n+1}} dz$$

show that

$$\left|\frac{d^n f}{dz^n}(0)\right| \leq \frac{n! M e^{n\beta}}{n^n} \qquad (n = 1, 2, \dots)$$

so the series $g(w) = \sum_{n=0}^{\infty} \frac{d^n f}{dz^n}(0) w^n$ has radius of convergence r where $r \geq e^{1-\beta}$. Calculate the Laplace transform of $f(t) = \sum_{n=0}^{\infty} \frac{d^n f}{dz^n}(0) t^n / n!$ for $t > 0$ and compare with g.

Exercise 4.22 (Error Function)

(i) The error function is $\operatorname{erf}(t) = 2\int_0^t e^{-x^2} dx / \sqrt{\pi}$. Express $\operatorname{erf}(t)$ as a power series, show that $\operatorname{erf}(1/s)$ is holomorphic near ∞ and find the inverse Laplace transform $g(t)$ of $\operatorname{erf}(1/s)$.
(ii) Find the Laplace transform of $g(\sqrt{t})$, and compare this with $(1/\sqrt{s}) \sin(1/\sqrt{s})$.

Exercise 4.23 (Fourier cosine inversion formula) Let $f : (0, \infty) \to \mathbb{C}$ be a continuous function such that $\int_0^{\infty} |f(t)| dt$ converges. Let ϕ be the Fourier cosine transform of f, and suppose that $\int_0^{\infty} |\phi(\omega)| \, d\omega$ converges. Show that

$$f(t) = (2/\pi) \int_0^{\infty} \cos(\omega t) \, \phi(\omega) \, d\omega \qquad (t > 0).$$

Chapter 5
Transfer Functions, Frequency Response, Realization and Stability

This chapter considers the Laplace transforms of linear systems, particularly $SISOs$ that have rational transfer functions. The aim is to reinterpret the properties of solutions $y(t)$ in terms of the transfer function $T(s)$. The centrally important idea is stability, and we focus attention on BIBO stability, which means that bounded inputs always lead to bounded outputs. This chapter contains the crucial theorem that BIBO stability of a linear system (A, B, C, D) is equivalent to stability of its transfer function as a rational function. Results of complex analysis are crucial to the theory, and we begin by considering some contours and winding numbers. Nyquist and Bode observed that much of the essential information about a linear system (A, B, C, D) is captured by the frequency response function $T(i\omega)$, which can be plotted in a diagram known as a Nyquist plot. With computers it is straightforward to plot Nyquist diagrams and when suitably interpreted they encapsulate much information about the linear system. We consider these plots geometrically and relate them to properties of the transfer function such as gain and phase. The plots lead to criteria for various linear systems to be BIBO stable. Using these tools from geometric function theory, we are able to solve stability problems as posed by Maxwell.

5.1 Winding Numbers

Let $\gamma : [a, b] \to \mathbb{C}$ be a continuously differentiable function. Then we say that γ is an arc with initial point $\gamma(a)$ and final point $\gamma(b)$, and that $d\gamma/dt$ gives the tangent to γ at $\gamma(t)$. A curve γ is a continuous function that is made up of consecutive arcs such that the final point of one arc is the initial point of the next arc. If $\gamma(a) = \gamma(b)$ then we call γ a contour.

Definition 5.1 (Winding Number) Let γ be a contour. If $z = \gamma(t)$ for some $t \in [a, b]$, then we say that z lies on γ, or that γ passes through z. Otherwise, for an arc

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_5

γ we define

$$n(\gamma, z) = \frac{1}{2\pi i}\int_a^b \frac{\frac{d\gamma}{dt}dt}{\gamma(t) - z} = \frac{1}{2\pi i}\int_\gamma \frac{ds}{s - z} \tag{5.1}$$

to be the winding number of γ about z. The definition extends to contours made up of several arcs by splitting the integral into integrals over arcs.

One can show that $n(\gamma, z)$ is an integer, and that for all z not on γ there exists $\delta > 0$ such that $n(\gamma, z) = n(\gamma, w)$ for all $w \in \mathbb{C}$ such that $|z - w| < \delta$.

(i) In particular, if $n(\gamma, z) = 1$, then we say that γ winds round z once in the positive sense.
(ii) By Cauchy's theorem $n(\gamma, z) = 0$ for all z such that $|z|$ is sufficiently large. If $\{z \in \mathbb{C} : n(\gamma, z) = 0\}$ consists of a connected open set, then its elements are said to lie outside of γ.

Example 5.2 For $r > 0$ and $a \in \mathbb{C}$, the circle $\gamma = C(a, r)$ with centre a and radius r is given by $s = a + re^{i\theta}$ for $0 \le \theta \le 2\pi$. Then $n(\gamma, z) = 1$ for $|z - a| < r$, namely the points in the open disc of centre a and radius r; whereas $n(\gamma, z) = 0$ for $|z - a| > r$, namely the points outside the closed disc of centre a and radius r.

Suppose that f is a rational function such that the poles of f are not on γ. Then $\Gamma = f \circ \gamma : [a, b] \to \mathbb{C}$ is an arc. If γ is a contour, then Γ is also a contour. For $z \in \mathbb{C}$, we consider whether Γ winds round or passes through z, and introduce

$$n(f \circ \gamma, z) = \frac{1}{2\pi i}\int_a^b \frac{f(\gamma(t))(d\gamma/dt)dt}{f(\gamma(t)) - z} = \frac{1}{2\pi i}\int_\gamma \frac{(df/ds)ds}{f(s) - z}. \tag{5.2}$$

In other words, we use $n(f \circ \gamma, z)$ to determine whether $f(s)$ winds round or passes through z as s describes γ.

In complex analysis it is usual to use continuous curves that are made up of consecutive arcs. The previous observations apply likewise to this case (Fig. 5.1).

Definition 5.3 (Semicircular Contours) Let $R > 0$. In complex analysis one considers the semicircular contour in the left half-plane $\Gamma = S_R \oplus [Ri, -Ri]$, which is given by the semicircular arc $S_R : z(\theta) = Re^{i\theta}$ for $-\pi/2 \le \theta \le \pi/2$ with centre 0 from $-iR$ to iR in the left half-plane, followed by the line segment $[iR, -iR]$ $z(\omega) = -i\omega$ for $-R \le \omega \le R$ from iR down the imaginary axis to $-iR$. Evidently Γ is continuous, and starts and finishes at $-iR$, hence defines a contour. We say that Γ is described in the positive sense, namely anti-clockwise.

In some engineering books, a different convention is followed, and one considers the reverse of Γ, namely $(-\Gamma) = [-Ri, Ri] \oplus (-S_R)$. Here we take the line segment $[-Ri, Ri]$ $z(\omega) = i\omega$ for $-R \le \omega \le R$ from $-iR$ up the imaginary axis to iR, then the semicircular arc $(-S_R)$ $z(\theta) = Re^{-i\theta}$ for $-\pi/2 \le \theta \le \pi/2$ with centre 0 from iR to $-iR$ in the left half-plane. The contour $(-\Gamma)$ is taken in the negative sense,

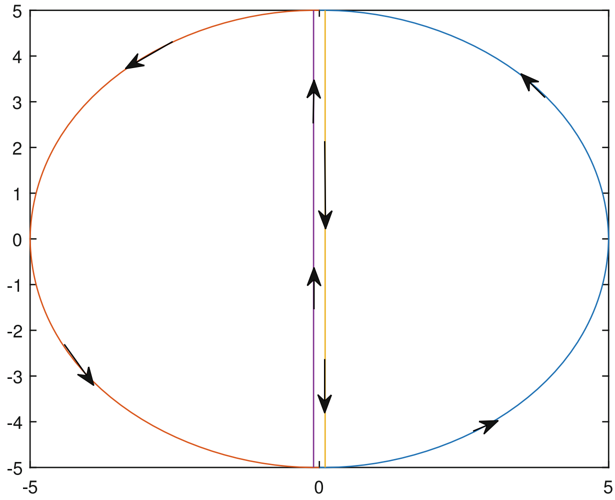


Fig. 5.1 Semicircular contours in left and left half-planes

namely clockwise. Hence we need to interpret the formulas of complex analysis carefully, reversing the signs as necessary.

Now consider a proper rational function $T(s)$; note that $T(s)$ has no poles on the imaginary axis. By choosing $R > 0$ sufficiently large, we can ensure that there are no poles on S_R. Since $T(s)$ is proper, there exists $c \in \mathbb{C}$ such that $T(s) \to c$ as $|s| \to \infty$, so in particular, $T(s) \to c$ as $R \to \infty$ for all s on S_R. Pictorially, the image $\{T(s) : s \in S_R\}$ reduces to a curve joining the points $T(iR)$ to $T(-iR)$ where $T(iR) \to c$ and $T(-iR) \to c$. For this reason, one often replaces the full contour $[-Ri, Ri] \oplus (-S_R)$ by the line segment $[-Ri, Ri]$, and fills in the gap between $T(iR)$ to $T(-iR)$. In this context, we can regard $[-Ri, Ri]$ for large $R > 0$ as a contour that goes round points in the open left half-plane once in the negative sense. For s on $[-iR, iR]$, we use the natural parametrization $s = i\omega$ where $-R \leq \omega \leq R$ is the range of natural frequencies, and consider $T(i\omega)$, the frequency response function.

If $T(s)$ is also stable, there are no poles inside or on $(-\Gamma)$ for sufficiently large $R > 0$; the poles of $T(s)$ are either in LHP outside $(-\Gamma)$, or in the open RHP inside $(-\Gamma)$.

Proposition 5.4 *Let $f(s)$ be a rational function that has no zeros or poles on the imaginary axis, and let*

$$\begin{aligned} Z_R &= \sharp\{\text{zeros of } f(s) \text{ in RHP}\} \\ Z_L &= \sharp\{\text{zeros of } f(s) \text{ in LHP}\} \\ P_R &= \sharp\{\text{poles of } f(s) \text{ in RHP}\} \\ P_L &= \sharp\{\text{poles of } f(s) \text{ in LHP}\}. \end{aligned} \tag{5.3}$$

Then $Z_L + Z_R$ equals the degree of the numerator, $P_L + P_R$ equals the degree of the denominator, and for all sufficiently large $r > 0$

$$\frac{1}{2\pi i}\int_{C(0,r)} \frac{df/ds}{f(s)} ds = Z_R + Z_L - P_R - P_L \tag{5.4}$$

and

$$\lim_{r\to\infty} \frac{1}{\pi}\int_{-r}^{r} \frac{(df/ds)(i\omega)}{f(i\omega)} d\omega = Z_L - Z_R + P_R - P_L. \tag{5.5}$$

Proof We use the contours of Fig. 5.1. We choose $r > 0$ so large that all the zeros and poles lie inside $C(0, r)$. Then we introduce the semicircular contour $S_r \oplus [ir, -ir]$ and apply Cauchy's Residue Theorem to obtain

$$\int_{S_r} + \int_{[ir,-ir]} \frac{df/ds}{f(s)} ds = 2\pi i(Z_R - P_R). \tag{5.6}$$

Likewise, when we take the semicircular contour $[-ir, ir] \oplus T_r$ which is taken anti-clockwise the left half-plane, we obtain

$$\int_{T_r} + \int_{[-ir,ir]} \frac{df/ds}{f(s)} ds = 2\pi i(Z_L - P_L). \tag{5.7}$$

The sum of these gives

$$\int_{C(0,r)} \frac{df/ds}{f(s)} ds = \int_{S_r} \frac{df/ds}{f(s)} ds + \int_{T_r} \frac{df/ds}{f(s)} ds = 2\pi i(Z_R + Z_L - P_R - P_L), \tag{5.8}$$

since the contribution from $[-ir, ir]$ cancels the contribution from $[ir, -ir]$. Also

$$\int_{S_r} \frac{df/ds}{f(s)} ds - \int_{T_r} \frac{df/ds}{f(s)} ds = O\Big(\frac{1}{r}\Big) \tag{5.9}$$

as $r \to \infty$. The reason is that

$$\frac{df/ds}{f(s)} = \frac{Z_R + Z_L - P_L - P_R}{s} + O\Big(\frac{1}{s^2}\Big) \tag{5.10}$$

and we can compute these with the substitution $s = re^{i\theta}$. Then by taking (5.7)-(5.6)+(5.9), we obtain

$$2\int_{[-ir,ir]} \frac{df/ds}{f(s)} ds = 2\pi i(Z_L - Z_R + P_R - P_L) + O\Big(\frac{1}{r}\Big), \tag{5.11}$$

and finally we take $s = i\omega$ to parametrize the integral.

Note that the first of the integrals gives the degree of the numerator minus the degree of the denominator, while the second integral gives us extra information. When a function has poles on the imaginary axis, we need to modify the contours, as in Dirichlet's integral. □

Proposition 5.5 (Argument Principle) *Let f be a rational function with no zeros or poles on a simple contour γ. Let P be the number of poles inside γ, counted according to multiplicity, and Z be the number of zeros of f, counted according to multiplicity. Then*

$$\frac{1}{2\pi i}\int_{\gamma}\frac{df/ds}{f(s)}ds = Z - P. \tag{5.12}$$

5.2 Realization

We now apply the results of the previous section to transfer functions.

Definition 5.6 (Transfer Function) Consider a linear system $Y = LU$ where L is a linear operator, and such that all the entries of the $(k \times 1)$ input U and $(m \times 1)$ output Y satisfy (E) of Sect. 4.1, and let the initial conditions be zero. Suppose that $T(s)$ is a $(m \times k)$ matrix of functions such that

$$\hat{Y}(s) = T(s)\hat{U}(s) \qquad (s > \beta). \tag{5.13}$$

Then $T(s)$ is called the transfer function of the linear system.

Consider a SISO linear system $Y = LU$ where L is a linear operator, and such that all the input U and output Y satisfy (E). Let $\hat{Y}$ and $\hat{U}$ be the Laplace transforms of Y and U. Suppose that $T(s)$ is a complex function such that $\hat{Y}(s) = T(s)\hat{U}(s)$ for $s > \beta$ so $T(s)$ is the transfer function of the linear system. Conversely, we have a realization theorem.

Theorem 5.7 (Realizing a SISO by a Rational Function) *Let T be a complex rational function. Then there exists a SISO linear system Σ, possibly with feedback, composed of taps, amplifiers, summing junctions, integrators, and differentiators, such that the transfer function of Σ is T.*

Proof Let the transfer function be $T(s) = p(s)/q(s)$ where $p(s) = \sum_{j=0}^{n} a_j s^j$ and $q(s) = \sum_{k=0}^{m} b_k s^k$ are polynomials with $b_m = 1$. Consider the differential equation

$$y = \sum_{j=0}^{n} a_j \frac{d^j u}{dt^j} \tag{5.14}$$

which has Laplace transform

$$Y(s) = p(s)U(s). \tag{5.15}$$

Also, we can realize the proper rational function $1/q(s)$ as the transfer function of a SISO system, as in Proposition 2.51. By combining these in series, we realize a system with transfer function $T(s) = p(s)/q(s)$. □

Corollary 5.8 *Let T be a matrix of complex rational functions. Then there exists a MIMO linear system Σ, possibly with feedback, composed of taps, matrix amplifiers, summing junctions, differentiators and integrators such that the transfer function of Σ is T.*

5.3 Frequency Response

Suppose that we have a SISO with Laplace transform $\hat{Y}(s) = T(s)\hat{U}(s)$. We change variable to $s = i\omega$ so $\hat{Y}(i\omega) = T(i\omega)\hat{U}(i\omega)$. (Consider input $e^{i\omega t}$, with $i\omega$ on imaginary axis in s plane.)

Definition 5.9

(i) Let $T(s)$ be a (proper) rational function. Then the frequency response function is $T(i\omega)$ where $\omega \in (-\infty, \infty)$.
(ii) The Nyquist plot of T is the curve $\{T(i\omega) : -\infty \leq \omega \leq \infty\}$.

Note that $\omega \mapsto e^{ia\omega}$ for $a > 0$ is periodic with period $2\pi/a$. We interpret ω as an angular frequency. Nyquist introduced a plot of the frequency response function $T(i\omega) = \Gamma(\omega)e^{i\phi(\omega)}$. The Nyquist plot is easy to produce on computer, and one can glean a great deal of useful information about the linear system from the shape of Nyquist plot. Here we focus attention on Nyquist's criterion for stability Theorem 5.30, which is the starting point for the other application. In examples it is helpful to produce Nyquist plots of all the frequency response functions in use.

Remark 5.10 (Geometrical Interpretation of Phase and Gain) In the Nyquist plot, the gain and phase can be found from the polar form of points on the Nyquist contour:

- $\Gamma(\omega) = |T(i\omega)|$ is the gain, namely the distance of $T(i\omega)$ to 0;
- $\phi(\omega) = \arg T(i\omega)$ is the phase, namely the angle between $T(i\omega)$ and the positive real axis.

In complex analysis, a contour is a continuous curve $\gamma : [a, b] \to \mathbb{C}$ such that $\gamma(a) = \gamma(b)$, and γ is piecewise continuously differentiable. The following phases are noteworthy

$\phi(\omega) = 0, 2\pi$ when the Nyquist contour crosses the positive real axis $(0, \infty)$;

$\phi(\omega) = \pi/2$ when the Nyquist contour crosses the positive imaginary axis $(0, i\infty)$;

$\phi(\omega) = \pi, -\pi$ when the Nyquist contour crosses the negative real axis $(-\infty, 0)$;

$\phi(\omega) = -\pi/2, 3\pi/2$ when the Nyquist contour crosses the negative imaginary axis $(-i\infty, 0)$.

5.4 Nyquist's Locus

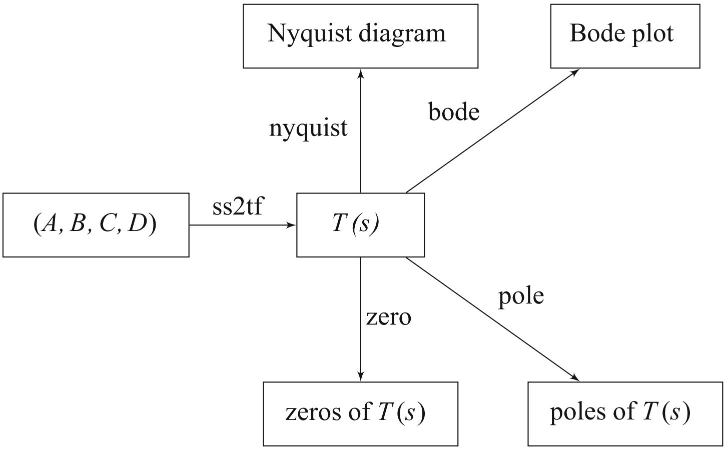


Diagram to show the information that may be derived from $T(s)$, graphically (Fig. 5.2).

Proposition 5.11 (Nyquist's Locus) *Let R be a proper rational function with all its poles in LHP. Then $R(i\omega)$ for $-\infty \leq \omega \leq \infty$ gives a contour in $\mathbb{C}$ that starts and ends at some $c \in \mathbb{C}$ where $R(s) \to c$ as $s \to \infty$.*

Proof Write $R(s) = c + p(s)/q(s)$ where degree of $p(s)$ is strictly less than the degree of $q(s)$, where $R(s) \to c$ as $s \to \infty$. There are finitely many poles, at λ such that $q(\lambda) = 0$, and there exists $\delta > 0$ such that $\Re\lambda < -\delta$ for all poles λ. Hence for $-\infty < \omega < \infty$, the function $R(i\omega)$ is continuously differentiable and $R(i\omega) \to c$ as $\omega \to \pm\infty$. We can write $\omega = \tan t$ where $t \in (-\pi/2, \pi/2)$, and $R(i \tan t) \to c$ as $t \to (-\pi/2)+$ and $t \to \pi/2-$, so $\gamma(t) = R(i \tan t)$ is a contour in the sense of complex analysis. Since R is proper with no poles on the imaginary axis, there exists M such that $|\frac{dR}{ds}(i\omega)| \leq M/(1 + \omega^2)$ for all real ω , hence $\int_{-\infty}^{\infty} |\frac{dR}{ds}(i\omega)|d\omega$ converges and the length of the contour is finite. The contour starts and ends at c, since $p(s)/q(s) \to 0$ as $s = i\omega \to \pm i\infty$. □

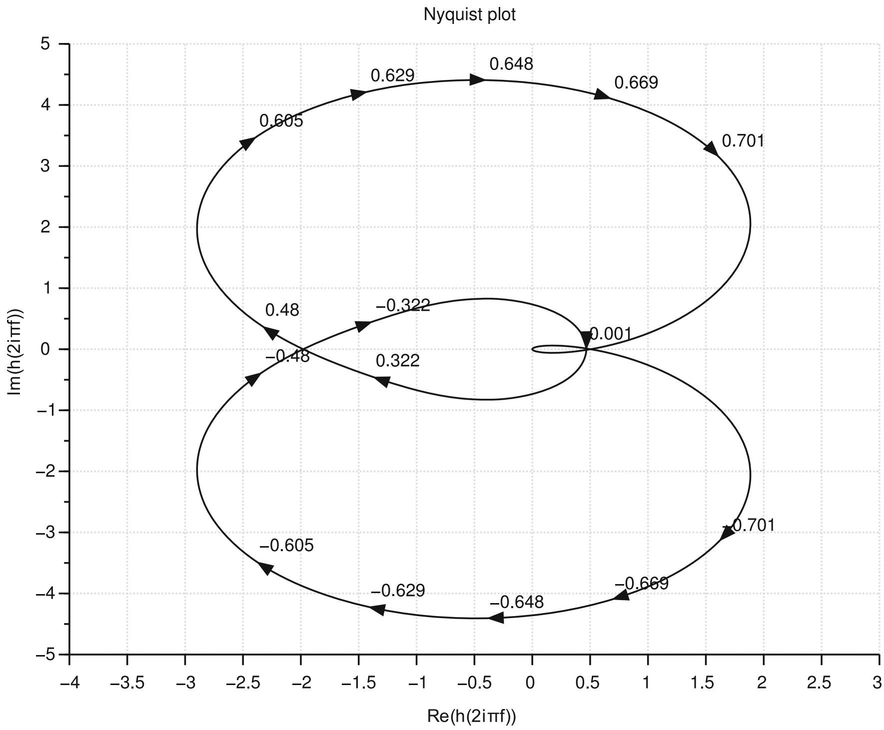


Fig. 5.2 Nyquist plot for the transfer function $(s^2 - 20s + 7)/(s^3 + 2s^2 + (70/4)s + 15)$. Note that a Nyquist plot can cross itself repeatedly, and the arrows indicate the direction of travel as $i\omega$ runs up the imaginary axis in the s plane

5.5 Gain and Phase

The polar decomposition of the frequency response function gives the gain and phase (Fig. 5.3).

- gain measures the factor by which the device multiplies the amplitude of a signal.
- phase describes the relative position of peaks in the input and output.

Definition 5.12 (Gain and Phase) Define the frequency response to be $T(i\omega)$, and make a polar decomposition $T(i\omega) = \Gamma(\omega)e^{i\phi(\omega)}$. Then define the gain (or amplitude gain) of the system to be $\Gamma(\omega) = |T(i\omega)|$ at angular frequency $\omega \in \mathbb{R}$; define the phase (shift) to be $\phi(\omega) = \arg T(i\omega)$. Equivalently,

$$\log T(i\omega) = \log \Gamma(\omega) + i\phi(\omega).$$

The phase (or phase shift) $\phi(\omega)$ is the change in phase of the signal. When $\phi(\omega) > 0$, one talks of a phase gain, so the output is running ahead of the input.

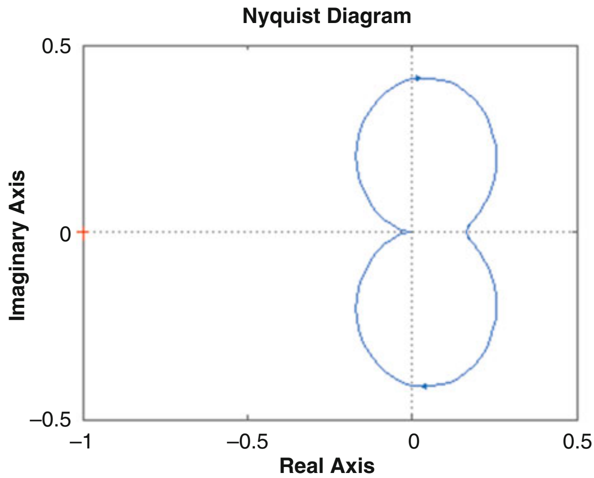
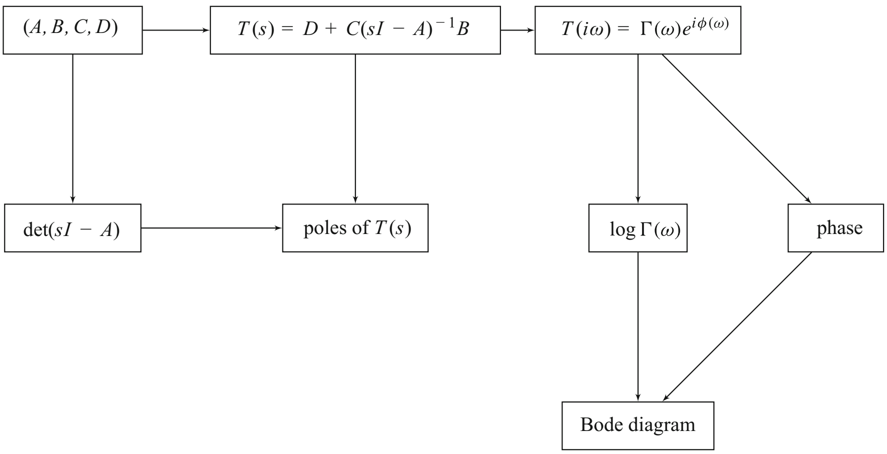


Fig. 5.3 Nyquist plot for the transfer function $1/(s^2 + s + 6)$

When $\phi(\omega) < 0$, there is a phase lag. In engineering, the frequency response is relatively easy to measure. The Bode plot consists of the graphs of $\log \Gamma(\omega)$ and $\phi(\omega)$ against ω, usually plotted on the same diagram; there are various options as to whether one uses natural logarithms, logarithms to base 10 for $\log \Gamma(\omega)$, and whether ϕ is in radians or degrees. MATLAB can give the logarithmic gain as expressed in decibels (dB), as in $20 \log_{10} \Gamma(\omega)$. For instance, $\Gamma(\omega) = 100$ gives $40dB$, while $\Gamma(\omega) = 0.1$ gives $-20dB$. The factor of $20 = 2 \times 10$ involves 10 to convert bels to decibels, while the 2 accounts for Γ^2, which is gain in the power of the transmitted signal. The bel is an inconveniently large unit, so decibels are more popular.



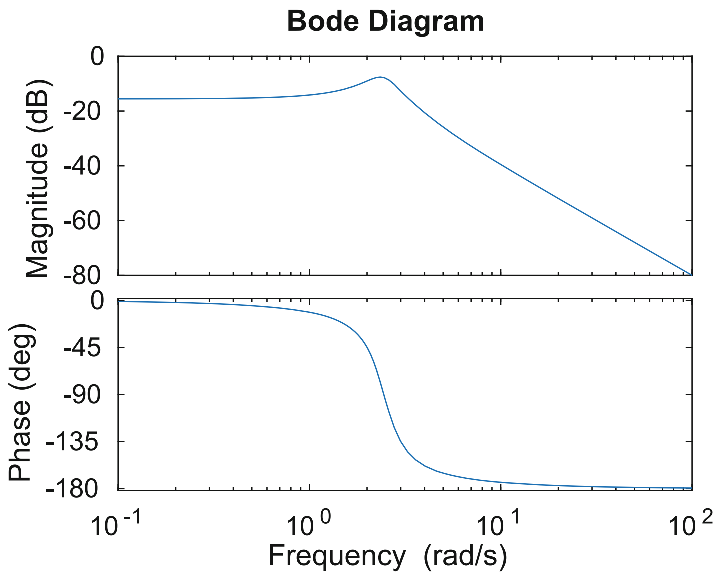


Fig. 5.4 Bode plot and phase for the transfer function $1/(s^2 + s + 6)$

Data derived from a MIMO (A, B, C, D) (Fig. 5.4)

Example 5.13 If the transfer function has poles on the imaginary axis, then the frequency response function and phase need to be interpreted carefully. In this example we write $\check{s} = -\bar{s}$, so $\check{s}$ is the reflection of s in the imaginary axis $\Re s = 0$; in particular, $s = \check{s}$ if and only if $\Re s = 0$. Suppose that (A, B, C, D) is a SISO with D real, $A' = -A$ and $C = iB'$. Then $T(\check{s}) = \overline{T(s)}$, so $T(s)$ is real for all $\Re s = 0$; to see this, write

$$T(s) = D + iB'(sI - A)^{-1}B \tag{5.16}$$

so

$$\overline{T(s)} = D' - iB'(\bar{s}I - A')^{-1}B = D + iB(\check{s}I - A)^{-1}B = T(\check{s}). \tag{5.17}$$

In particular, for

$$A = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, C = \begin{bmatrix} i & i \end{bmatrix}, D = 1 \tag{5.18}$$

we have transfer function

$$T(s) = 1 + \frac{2is}{s^2 + 1} \tag{5.19}$$

with poles at $s = \pm i$ on the imaginary axis, so the frequency response function is

$$T(i\omega) = \frac{\omega^2 + 2\omega - 1}{\omega^2 - 1} = \frac{(\omega + 1 + \sqrt{2})(\omega + 1 - \sqrt{2})}{(\omega + 1)(\omega - 1)}, \tag{5.20}$$

so $T(i\omega)$ is real for $\omega \in \mathbb{R} \setminus \{\pm 1\}$ and has sign

$$T(i\omega) > 0 \qquad (\omega \in (-\infty, -1 - \sqrt{2}) \cup (-1, \sqrt{2} - 1) \cup (1, \infty)), \tag{5.21}$$

$$T(i\omega) < 0 \qquad (\omega \in (-1 - \sqrt{2}, -1) \cup (\sqrt{2} - 1, 1)), \tag{5.22}$$

so the phase changes abruptly between 0 and π at the endpoints of these intervals.

Example 5.14

(i) For $a > 0$ and $\theta, b \in \mathbb{R}$, we introduce $\alpha = a + ib$ and the transfer function

$$T(s) = e^{i\theta} \frac{s - \alpha}{s + \bar{\alpha}} \tag{5.23}$$

which has a simple zero at $\alpha \in RHP$ and a simple pole at $-\bar{\alpha} \in LHP$. On the imaginary axis, we write $s = i\omega$ where $\omega = b + a\cot(\phi/2)$ so the frequency response function is

$$T(i\omega) = e^{i\theta} \frac{i\omega - ib - a}{i\omega - ib + a} = e^{i\theta} \frac{i\cot(\phi/2) - 1}{i\cot(\phi/2) + 1} = e^{i\theta} \frac{\cos(\phi/2) + i\sin(\phi/2)}{\cos(\phi/2) - i\sin(\phi/2)} = e^{i(\theta+\phi)}, \tag{5.24}$$

so that the gain is constant with $\Gamma = |T(i\omega)| = 1$, and the phase is $\theta + \phi$. This calculation is a variant on the $\tan t/2$ substitution which is commonly used in integral calculus.

(ii) We now take $a_j, c_k > 0$ and $b_j, d_k \in \mathbb{R}$ and introduce $\alpha_j = a_j + ib_j \in RHP$ and $\beta_k = c_k + id_k \in RHP$; then let

$$T(s) = e^{i\theta} \prod_{j=1}^{n} \frac{s - \alpha_j}{s + \bar{\alpha}_j} \prod_{k=1}^{m} \frac{s + \bar{\beta}_k}{s - \beta_k}, \tag{5.25}$$

which has zeros at $\alpha_j \in RHP$ and at $-\bar{\beta}_k \in LHP$, and poles at $-\alpha_j \in LHP$ and at $\beta_k \in RHP$. As in (i), the gain of the transfer function is constant $\Gamma = 1$. To find the phase ϕ, we introduce new variables ϕ_j and ψ_k depending upon ω by

$$\omega = b_j + a_j \cot(\phi_j/2), \quad \omega = d_k + c_k \cot(\psi_k/2) \tag{5.26}$$

and as in (i), obtain phase

$$\phi = \theta + \sum_{j=1}^{n} \phi_j - \sum_{k=1}^{m} \psi_k. \tag{5.27}$$

Example 5.15 (Gain and Phase of Damped Harmonic Oscillator) For $a, b > 0$, u_0 a constant and $y(0) = \frac{dy}{dt}(0) = 0$, we find the gain and phase of

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + y = u_0 \cos \omega t. \tag{5.28}$$

Then the Laplace transform is

$$(as^2 + bs + 1)Y(s) = u_0 \frac{s}{\omega^2 + s^2}, \tag{5.29}$$

so that taking $s = i\omega$ in the formula

$$T(s) = \frac{1}{as^2 + bs + 1}, \tag{5.30}$$

$$T(i\omega) = \frac{1}{1 - a\omega^2 + ib\omega} = \frac{1 - a\omega^2 - ib\omega}{(1 - a\omega^2)^2 + b^2\omega^2}, \tag{5.31}$$

so the gain is

$$\Gamma(\omega) = |T(i\omega)| = \frac{1}{\sqrt{(1 - a\omega^2)^2 + b^2\omega^2}}; \tag{5.32}$$

while the phase $\phi(\omega)$ satisfies

$$\tan \phi = \frac{b\omega}{\omega^2 a - 1}, \quad \phi = \tan^{-1} \frac{b\omega}{\omega^2 a - 1} \tag{5.33}$$

which has sign depending on the value of ω. Note that

$$\phi(\omega) = \arg(1 - a\omega^2 - ib\omega) \tag{5.34}$$

and so $T(0) = 1$, hence $\phi(0) = 0$; while $T(i\omega)$ is in the third quadrant as $\omega \to \infty$, whereas $T(i\omega)$ is in the second quadrant as $\omega \to -\infty$; at $1 - a\omega^2 = 0$, $T(i\omega)$ is on the imaginary axis so $\phi(\pm 1/\sqrt{a}) = \mp\pi/2$.

5.6 BIBO Stability

Definition 5.16 (BIBO) Let (A, B, C, D) be a linear system

$$\frac{dX}{dt} = AX + Bu$$
$$Y = CX + Du \tag{5.35}$$

such that for all bounded inputs $u(t)$ for $t \in (0, \infty)$, all outputs y are bounded for $t \in (0, \infty)$. Then we say that the system is bounded-input bounded-output stable, or BIBO stable.

Bounded Exponentials of Matrices

[**Lemma 5.17**]*Suppose that A has (not necessarily distinct) eigenvalues such that $\Re\lambda_j < 0$ for all $j = 1, \ldots, n$. Then there exists $M, \delta > 0$ such that*

$$\|\exp(tA)\| \leq Me^{-\delta t} \qquad (t \geq 0). \tag{5.36}$$

[***Proof***]This follows from Lemma 3.6. □

The difference between $\Re\lambda \leq 0$ in Proposition 2.33 (iii) and $\Re\lambda < 0$ in the Lemma 5.17 is subtle, and historically important in the theory. Maxwell realized that the stronger hypothesis of the Lemma 5.17, requiring strict inequality, is needed to cover the case of multiple eigenvalues, and deal with resonance.

Remark 5.18 (Stability Cases) Consider $dX/dt = AX$ with $X(0) = X_0$. This has solution $X(t) = \exp(tA)X_0$, and we distinguish the following cases.

(i) **Exponentially stable**: there exist $M, \delta > 0$ such that $\|X(t)\| \leq Me^{-\delta t}$ for all $t > 0$ and all X_0. This occurs when $\Re\lambda_j < 0$ for all eigenvalues λ_j. In Theorem 5.21 we find this to be BIBO stable.

(ii) **Marginally stable**: $X(t)$ is bounded for $t > 0$ for all X_0, which occurs when $\Re\lambda_j < 0$, or $\Re\lambda_j = 0$ and the corresponding Jordan blocks are all of size 1×1. Later we will resolve this marginal case as BIBO unstable. Whereas the complementary function $X(t)$ is bounded, a bounded input can give an unbounded particular integral. This effect occurs via resonance, which we discuss in the context of the harmonic oscillator.

(iii) **Unstable**: $X(t)$ $(t > 0)$ is unbounded for some X_0, which occurs when either $\Re\lambda_j > 0$ for some eigenvalue λ_j of A, or $\Re\lambda_j = 0$ for some λ_j that has a Jordan block of size ≥ 2. This is also found to be BIBO unstable in general.

These cases will be considered with reference to a crucial example, the damped harmonic oscillator.

Example 5.19 (Damped Harmonic Oscillator) Matrix form of the damped harmonic oscillator is

$$\frac{dX}{dt} = AX + BU, \tag{5.37}$$

where $\gamma > 0$ and β real in

$$A = \begin{bmatrix} 0 & 1 \\ -\gamma & -\beta \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{5.38}$$

The characteristic equation of A is

$$\det \begin{bmatrix} \lambda & -1 \\ \gamma & \lambda + \beta \end{bmatrix} = \lambda^2 + \lambda\beta + \gamma = 0, \tag{5.39}$$

so eigenvalues are

$$\lambda_\pm = 2^{-1}(-\beta \pm \sqrt{\beta^2 - 4\gamma}), \tag{5.40}$$

with corresponding eigenvectors

$$\begin{bmatrix} 1 \\ \lambda_+ \end{bmatrix}, \qquad \begin{bmatrix} 1 \\ \lambda_- \end{bmatrix} \tag{5.41}$$

so when $\lambda_+ \neq \lambda_-$, we introduce

$$S = \begin{bmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{bmatrix} \tag{5.42}$$

so that S is invertible and

$$A = S \begin{bmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{bmatrix} S^{-1} \tag{5.43}$$

and

$$\exp(tA) = S \begin{bmatrix} e^{t\lambda_+} & 0 \\ 0 & e^{t\lambda_-} \end{bmatrix} S^{-1} \tag{5.44}$$

Cases of the damped harmonic oscillator

Consider $1/(s^2 + \beta s + \gamma)$ with $\gamma > 0$; poles at $\lambda_\pm = (1/2)(-\beta \pm \sqrt{\Delta})$ where $\Delta = \beta^2 - 4\gamma$. The results are summarized in the following table.

solutions	$\Delta < 0$	$\Delta = 0$	$\Delta > 0$
$\beta > 0$	decaying oscillations	critically damped	exp decay
$\beta = 0$	periodic	constant	hyperbolic
$\beta < 0$	unbounded oscillations	exponential growth	exp growth

(5.45)

The damped oscillator is exponentially stable if and only if $\beta > 0$ and $\gamma > 0$. When $\beta = 0$ and $\gamma > 0$, the oscillator is marginally stable. For $\beta < 0$, the oscillator is unstable.

Poles of the Transfer Function of the Damped Harmonic Oscillator
Consider $1/(s^2+\beta s+\gamma)$ with $\gamma > 0$ and β real with poles at $\lambda_\pm = (1/2)(-\beta \pm \sqrt{\Delta})$ where $\Delta = \beta^2 - 4\gamma$. Then

poles	β	$\Delta < 0$	$\Delta = 0$	$\Delta > 0$
		$\lambda_+ = \bar{\lambda}_-$	$\lambda_+ = \lambda_-$	distinct real roots
unstable	$\beta < 0$	$\Re\lambda_\pm > 0$	$\lambda_\pm > 0$	$0 < \lambda_- < \lambda_+$
marginal	$\beta = 0$	$\Re\lambda_\pm = 0$	$\lambda_\pm = 0$	$\lambda_- < 0 < \lambda_+$
stable	$\beta > 0$	$\Re\lambda_\pm < 0$	$\lambda_\pm < 0$	$\lambda_- < \lambda_+ < 0$

(5.46)

For a damped harmonic oscillator, we have $\beta, \gamma > 0$, so only the last row matters. The last row gives the stable cases.

5.7 Undamped Harmonic Oscillator: Marginal Stability and Resonance

Example 5.20 The undamped harmonic oscillator

$$\frac{d^2x}{dt^2} + \nu^2 x = U_0 \cos \omega t \tag{5.47}$$

with U_0 real and $\omega, \nu > 0$ is marginally stable, but not BIBO stable.

We introduce

$$X = \begin{bmatrix} x \\ v \end{bmatrix}, A = \begin{bmatrix} 0 & 1 \\ -\nu^2 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{5.48}$$

$$\frac{d}{dt}X = AX + BU_0 \cos \omega t. \tag{5.49}$$

Note that A has eigenvalues $\pm i\nu$ on the imaginary axis.

Marginal Stability

The general solution is given by the complementary function plus a particular integral. The complementary function arises when the input is zero. For $U = 0$ the system oscillates at its natural angular frequency ν and the general solution of $\frac{d}{dt}X = AX$ is

$$X = c_1 \begin{bmatrix} \cos \nu t \\ -\nu \sin \nu t \end{bmatrix} + c_2 \begin{bmatrix} \sin \nu t \\ \nu \cos \nu t \end{bmatrix}, \tag{5.50}$$

for constants c_1, c_2. In particular, all these solutions are bounded, so we have marginal stability.

- For $U_0 \neq 0$ and $\omega \neq \nu$, the input has angular frequency different from the natural angular frequency, and the solution is the complementary function plus a particular integral

$$X = c_1 \begin{bmatrix} \cos \nu t \\ -\nu \sin \nu t \end{bmatrix} + c_2 \begin{bmatrix} \sin \nu t \\ \nu \cos \nu t \end{bmatrix} + \frac{U_0}{\nu^2 - \omega^2} \begin{bmatrix} \cos \omega t \\ -\omega \sin \omega t \end{bmatrix}; \tag{5.51}$$

  here the complementary function oscillates at natural angular frequency ν; whereas the particular integral oscillates at the input angular frequency ω. These solutions are all bounded. One can obtain these particular integrals by W3.2, or by guesswork.
  Resonance
- Let $U_0 \neq 0$ and $\nu = \omega$, so that the input angular frequency equals the natural angular frequency. Then the general solution is

$$X = c_1 \begin{bmatrix} \cos \nu t \\ -\sin \nu t \end{bmatrix} + c_2 \begin{bmatrix} \sin \nu t \\ \nu \cos \nu t \end{bmatrix} + \frac{U_0}{2\nu} \begin{bmatrix} t \sin \nu t \\ \sin \nu t + t\nu \cos \nu t \end{bmatrix} \tag{5.52}$$

  where the solution oscillates unboundedly; this effect is called resonance. The input is bounded whereas the output is unbounded, so the system is not BIBO stable. A system is prone to resonance when the transfer function has a pole on the imaginary axis. The term marginal stability is used to describe the situation in which the complementary function is bounded, whereas the particular integral is unbounded for suitably chosen bounded inputs; this means that the system is not BIBO stable.

Resonance is desirable or undesirable depending upon the application. The process of tuning involves inputting a signal with a single oscillating frequency such as a sine wave, and then identifying the frequency that produces a large output. Musical instruments are tuned so that they resonate at particular frequencies in the process of tuning. However, in automotive engineering, one avoids having structural components that resonate at the frequency of the engine's rotation, as this would produce noisy vibrations. In the example of square waves in Sect. 4.9, we identified a system that could be adjusted to have several resonant frequencies, which in the context of music are known as harmonics.

5.8 BIBO Stability in Terms of Eigenvalues of A

Theorem 5.21 *Suppose that all eigenvalues λ_j of A satisfy $\Re\lambda_j < 0$, and that U is bounded on $(0, \infty)$. Then all solutions to*

$$\frac{dX}{dt} = AX + BU$$
$$Y = CX + DU \tag{5.53}$$

are bounded on $(0, \infty)$. Hence (A, B, C, D) is BIBO stable.

Proof By the Theorem 2.40, the general solution to the differential equation is

$$X(t) = \exp(tA)X_0 + \int_0^t \exp((t-s)A)BU(s)\,ds \tag{5.54}$$

where by hypotheses there exists $K > 0$ such that

$$\|B\|\|U(s)\| \leq K \qquad (s > 0) \tag{5.55}$$

and by the Lemma 5.17

$$\|\exp(tA)\| \leq Me^{-t\delta} \qquad (t > 0), \tag{5.56}$$

so

$$\|X(t)\| \leq Me^{-t\delta}\|X_0\| + \int_0^t \|\exp((t-s)A)\|\|B\|\|U(s)\|ds \tag{5.57}$$

Hence

$$\begin{aligned}\|X(t)\| &\leq M\|X_0\| + KM\int_0^t e^{-\delta(t-s)}\,ds \\ &= M\|X_0\| + \frac{KM}{\delta}\Big[-e^{-\delta(t-s)}\Big]_0^t \\ &= M\|X_0\| + \frac{KM}{\delta}\big(1 - e^{-\delta t}\big) \\ &\leq M\|X_0\| + \frac{KM}{\delta} \qquad (t > 0).\end{aligned}$$

Hence $X(t)$ is bounded, so the output is also bounded, since

$$Y(t) = CX(t) + Du(t) \tag{5.58}$$

is the sum of two bounded functions. □

Transfer functions and stability criteria: Next we combine ideas about transfer functions with the notion of stability, so as to obtain criteria for stability of a system solely in terms of properties of transfer functions. The idea is to describe the properties of solutions of the differential equation, without having to solve the differential equations explicitly. Thus we go from differential equations to algebra via the Laplace transform. Instead of working with functions of time t in the state space or time domain, we work with functions of s in s-space, where s is a complex variable.

When building devices out of components, the main operations on the transfer functions are:

- amplification $\lambda f(s)$
- addition $f(s) + g(s)$
- multiplication $f(s)g(s)$.

We investigate these complex functions, starting with polynomials, and progressing to rational functions. In the rest of this chapter we use geometrical tools, and in the following chapter we introduce more sophisticated methods from algebra.

5.9 Maxwell's Stability Problem

Definition 5.22 (Stable Polynomials) A polynomial $h(s)$ is said to be stable if all of its zeros are in the open left half-plane $LHP = \{s \in \mathbb{C} : \Re s < 0\}$.

Problem (Maxwell's Problem) Find necessary and sufficient conditions on the coefficients of a monic complex polynomial for the polynomial to be stable.

Finding the zeros exactly can be very difficult, especially when the polynomial has large degree and there are multiple zeros near to the imaginary axis. Practical modern method: given a monic complex polynomial $p(s)$, there exists a complex matrix A such that $\det(sI - A) = p(s)$. Then one can find the eigenvalues of A numerically. If all the eigenvalues are comfortably in the open left half-plane, then $p(s)$ is stable.

Proposition 5.23 (Necessary Condition for Stability) *Suppose that $h(s)$ is a monic real polynomial that is stable. Then all the coefficients of $h(s)$ are positive.*

Proof Here $h(s)$ has real coefficients, so $h(\lambda) = 0$ if and only if $h(\bar{\lambda}) = 0$. Hence the roots of $h(s)$ are either real $\mu_j < 0$; or pairs of conjugate complex roots λ_k and $\bar{\lambda}_k$ with $\Re\lambda_k < 0$, which combine to give real quadratic factors $(s - \lambda_k)(s - \bar{\lambda}_k) =$

$s^2 - 2s\Re\lambda_k + |\lambda_k|^2$. Hence $h(s)$ factorizes as

$$h(s) = \prod_{j=1}^{n}(s - \mu_j)\prod_{k=1}^{m}(s^2 - 2s\Re\lambda_k + |\lambda_k|^2), \tag{5.59}$$

where $-\mu_j > 0$, $-2\Re\lambda_k > 0$ and $|\lambda_k|^2 > 0$; hence all the coefficients that we obtain on multiplying out are positive. □

This necessary condition for stability is easy to check, but it not sufficient. For example

$$\frac{s^5 - 1}{s - 1} = s^4 + s^3 + s^2 + s + 1 = \Big(s^2 + \frac{1+\sqrt{5}}{2}s + 1\Big)\Big(s^2 + \frac{1-\sqrt{5}}{2}s + 1\Big) \tag{5.60}$$

has roots at the complex fifth roots of unit, namely two roots in LHP and two roots in RHP, hence is unstable. In Proposition 6.7 we characterize stable real cubics. Routh and Hurwitz [30] extended this to a sufficient condition for general real polynomials, as we present in Theorem 6.12.

5.10 Stable Rational Transfer Functions

Definition 5.24 (Stable Rational Functions) Let $LHP = \{s \in \mathbb{C} : \Re s < 0\}$ be the open left half-plane. A complex rational function $f(s)$ is said to be stable if

(i) $f(s)$ is proper, and
(ii) all the poles of $f(s)$ are in the open left half-plane.
The space of stable rational functions is denoted $\mathcal{S}$.

Equivalently, $f(s) = g(s)/h(s)$ is stable if

(i) degree$(g(s)) \leq$ degree$(h(s))$, so $f(s)$ is proper, and
(ii) all the zeros of $h(s)$ have $\Re s < 0$, so $h(s)$ is stable.

(So a polynomial $h(s)$ is stable, if and only if $1/h(s)$ is a stable rational function.)

For a linear system such as (A, B, C, D), we have two notions of stability, one is BIBO stability, relating to the solutions of the associated differential equation; the other is stability of the transfer function as a rational function. The following result resolves these two interpretations. The merit of the result is that one can often determine whether transfer functions are stable by basic algebra.

Theorem 5.25 (Stability for Systems and Transfer Functions) *Let* $\Sigma = (A, B, C, D)$ *be a linear system with rational transfer function* T. *Then* Σ *is BIBO stable if and only if* T *is stable.*

Proof T not stable implies Σ not BIBO stable: Suppose that the system is BIBO, and that T is not stable. Recall $\hat{Y}(s) = T(s)\hat{U}(s)$. Then we can choose a bounded input $U = 1$ such that $\hat{U}(s) = 1/s$. But Σ is BIBO stable, so Y is bounded, so $|Y(t)| \leq M$ for some M and all $t > 0$, so

$$\begin{aligned} |\hat{Y}(s)| &= \Big| \int_0^\infty e^{-st} Y(t)\, dt \Big| \\ &\leq \Big| \int_0^\infty e^{-t\Re s} M\, dt \Big| \leq \frac{M}{\Re s}. \end{aligned}$$

Hence $\hat{Y}(s)$ is holomorphic on $\{s : \Re s > 0\}$ and $\hat{Y}(s) \to 0$ as $s \to \infty$ along $(0, \infty)$. So $T(s) = s\hat{Y}(s)$ must be proper rational.

Suppose that T has a pole at λ. If $\Re\lambda > 0$, then $T(s)\hat{U}(s) = T(s)/s$ also has a pole at λ. But $\hat{Y}(s)$ cannot have a pole at $s = \lambda$ by Prop.

Now suppose that $\Re\lambda = 0$, so $\lambda = i\nu$ for some real ν. The idea is to cause resonance, so we let $U(t) = \cos \nu t$, which is bounded, and

$$\hat{U}(s) = \frac{s}{s^2 + \nu^2} = \frac{1/2}{s - i\nu} + \frac{1/2}{s + i\nu} \tag{5.61}$$

has a pole at $i\nu$, and hence $\hat{Y}(s) = T(s)\hat{U}(s)$ has a double (or triple, ...) pole at $i\nu$.

Now consider s with $\Re s > 0$ and $s \to i\nu$. Now $T(s)\hat{U}(s)$ diverges like $1/(s - i\nu)^2$ or $1/(s - i\nu)^3$ etc.; whereas $\hat{Y}(s)$ can only diverge like $M/\Re s$ at worst. This contradicts the identity $\hat{Y}(s) = T(s)\hat{U}(s)$.

We deduce that Y has at most a simple pole on the imaginary axis, so T has no poles in the imaginary axis. Hence $T(s)$ has all its poles in LHP. Hence T is stable.
T stable implies BIBO stable:

Conversely, suppose that T is stable. Then by Proposition 2.51, there exists a SISO (A, B, C, D) such that the transfer function is T and the eigenvalues λ of A are the poles of T, hence satisfy $\Re\lambda < 0$. Then by Theorem 5.21, (A, B, C, D) is BIBO stable. □

We state two results which summarize results from elsewhere in the book.

Theorem 5.26 (Realization)

(i) Every monic complex polynomial is the characteristic polynomial of some complex matrix.
(ii) Every proper complex rational function is the transfer function $T(s)$ of some SISO system (A, B, C, D).
(iii) Every stable complex rational function is the transfer function of some BIBO stable system (A, B, C, D).

Realization suggests building a gadget with desired properties.

Proposition 5.27 (Stable Matrices) *For a $n \times n$ complex matrix A, the following conditions are equivalent.*

(i) *All the eigenvalues of A are in the open left half-plane.*
(ii) *There exists a positive definite $n \times n$ complex matrix K such that $-AK - KA' = I$.*
(iii) *The characteristic polynomial of A is stable.*
(iv) *For all (B, C, D) complex matrices of shape $(n \times 1, 1 \times n, 1 \times 1)$, the transfer function of (A, B, C, D)*

$$T(s) = D + C(sI - A)^{-1}B \tag{5.62}$$

is a stable rational function.
(v) *All solutions of $\frac{d}{dt}X = AX$ decay exponentially to 0 as $t \to \infty$.*
(vi) *For all (B, C, D) complex matrices of shape $(n \times 1, 1 \times n, 1 \times 1)$, the linear system (A, B, C, D) is BIBO stable.*

Example 5.28 (Three Rational Filters) In signal processing, the term filter is often used for a type of transfer function. Rational filters are easy to construct and analyze, and the following three examples have specific properties for their phase and gain.

(i) For $x > 0$ and $y \in \mathbb{R}$, let $z = x + iy \in RHP$, and $-\bar{z} = -x + iy \in LHP$ be its reflection in the imaginary axis. Then

$$B(s) = \frac{s - z}{s + \bar{z}} \tag{5.63}$$

is a stable rational function with a zero at $z \in RHP$ and a pole at $-\bar{z} \in LHP$. With $s = i\omega$ on the imaginary axis, we have the frequency response function

$$B(i\omega) = \frac{i\omega - iy - x}{i\omega - iy + x} = \frac{(\omega - y)^2 - x^2 + 2ix(\omega - y)}{(\omega - y)^2 + x^2} \tag{5.64}$$

so $B(i\omega)$ has constant gain $\Gamma = 1$ and phase $\phi(\omega)$ where

$$\tan\phi(\omega) = \frac{2x(\omega - y)}{(\omega - y)^2 - x^2}; \tag{5.65}$$

hence

$$\phi(\omega) \to 0 \quad (\omega \to -\infty); \quad \phi(\omega) \to -\pi/2 \quad (\omega \to (y - x)-) \quad \phi(y) = \pi; \tag{5.66}$$

$$\phi(\omega) \to \pi/2 \quad (\omega \to (y + x)+); \quad \phi(\omega) \to 0 \quad (\omega \to \infty); \tag{5.67}$$

so $B(i\omega)$ loops once round 0 in the clockwise (negative) sense. The Nyquist contour of B is the circle of centre 0 and radius 1, taken clockwise and starting and ending at 1.

(ii) Let

$$P_z(s) = \frac{-1}{s-z} + \frac{1}{s+\bar{z}} \tag{5.68}$$

so $P_z(s)$ is proper with $P_z(s) = O(1/s^2)$ as $s \to \infty$ with poles at $z \in RHP$ and $-\bar{z} \in LHP$, so $P_z(s)$ is not stable. Then the corresponding frequency response function is

$$P_z(i\omega) = \frac{2x}{(\omega - y)^2 + ax^2} > 0 \tag{5.69}$$

so phase $\phi(\omega) = 0$. Since $P_z(i\omega) \to 0$ as $\omega \to \pm\infty$, this filter reduces high frequency signals. By choosing $x > 0$ small, we can make $P_z(i\omega)$ be sharply peaked near to $\omega = y$, where $P_z(iy) = 2/x$ is the maximum of $P_z(i\omega)$. For $f(s)$ holomorphic and bounded on $\{s : \Re s > -\delta\}$ for some $\delta > 0$, we have an absolutely convergent integral

$$\int_{-i\infty}^{i\infty} f(s)\Big(\frac{-1}{s-z} + \frac{1}{s+\bar{z}}\Big)\frac{ds}{2\pi i} = f(z) \tag{5.70}$$

by Cauchy's integral formula; see Sect. 5.1 for discussion of the relevant semicircular contour. This is known as Poisson's integral formula for RHP.

If $f(z)$ is a stable rational transfer function, then f is determined by its frequency response function via this absolutely convergent integral.

(iii) Let

$$Q_z(s) = \frac{1}{s-z} + \frac{1}{s+\bar{z}} \tag{5.71}$$

so $Q_z(s)$ is proper with $Q(s) = O(1/s)$ as with poles at $z \in RHP$ and $-\bar{z} \in LHP$, so $Q_z(s)$ is not stable. Then the corresponding frequency response function is

$$Q_z(i\omega) = \frac{-2i(\omega - y)}{(\omega - y)^2 + x^2} \tag{5.72}$$

is purely imaginary, so phase $\phi(\omega) = \pi/2$ for $\omega < y$ and $\phi(\omega) = -\pi/2$ for $\omega > y$; thus the phase is discontinuous with a jump at y of size π. We also have

$$\begin{aligned} \int_{-i\infty}^{i\infty} e^{-ts}\Big(\frac{1}{s-z} + \frac{1}{s+\bar{z}}\Big)\frac{ds}{2\pi i} &= -e^{-tz} \qquad (t > 0) \\ &= e^{t\bar{z}} \qquad (t < 0) \end{aligned} \tag{5.73}$$

by Cauchy's integral formula. Suppose that f is holomorphic and bounded on $\{s : \Re s > -\delta\}$ for some $\delta > 0$. Then by Cauchy's Theorem we have

$$\int_{-i\infty}^{i\infty} f(s)Q_z(s)\frac{ds}{2\pi i} \to \lim_{\varepsilon\to 0+}\int_{-\infty}^{y-\varepsilon} + \int_{y+\varepsilon}^{\infty} \frac{f(i\omega)}{\omega - y}\frac{d\omega}{i\pi} \qquad (z \to iy). \tag{5.74}$$

The right hand side is i times the Hilbert transform of $f(iy)$; see (4.62).

Proposition 5.29 (Factorization of Stable Rational Functions) *Let $T \in \mathcal{S}$. Then $T(s) = S(s)B(s)$ for:*

(i) $S \in \mathcal{S}$ that has no zeros in RHP and $|T(iy)| = |S(iy)|$ for all $y \in \mathbb{R}$;
(ii) $B \in \mathcal{S}$ such that $|B(iy)| = 1$ for all $y \in \mathbb{R}$; and the factors are uniquely determined up to multiplication by a unimodular complex constant factor.

Proof Let the zeros of T in the open RHP be $b_1, \dots, b_m$; let the other zeros of T be $c_1, \dots, c_n$; let the poles of T be $a_1, \dots, a_p$, all listed according to multiplicity. Since T is stable, we have $p \geq m + n$, and $\Re a_j < 0$ for all j. We introduce

$$B(z) = \prod_{j=1}^{m} \frac{z - b_j}{z + \bar{b}_j} \tag{5.75}$$

which has zeros at $b_1, \dots, b_m \in RHP$ and poles at $-\bar{b}_1, \dots -\bar{b}_m \in LHP$, hence B is stable. Observe that $-\bar{b}_j$ is the reflection of b_j in the imaginary axis, so $|iy - b_j| = |iy + \bar{b}_j|$ for all $y \in \mathbb{R}$, so $|B(iy)| = 1$ for all $y \in \mathbb{R}$.

Now let

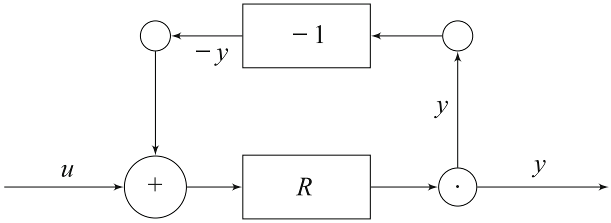
$$S(z) = \lambda \frac{\prod_{j=1}^{m}(z + \bar{b}_j)\prod_{k=1}^{n}(z - c_k)}{\prod_{\ell=1}^{p}(z - a_\ell)}, \tag{5.76}$$

where $\lambda \neq 0$ is to be chosen. Then S has poles at $a_1, \dots, a_p \in LHP$, zeros at $-\bar{b}_1, \dots - \bar{b}_m \in LHP$ and zeros at $c_1, \dots, c_n$ where $\Re c_j \leq 0$; hence S is stable. Also by cancellation, $S(s)B(s)/T(s)$ is a rational function with no zeros or poles, hence by Liouville's theorem is a constant, and by adjusting λ we can ensure that $T(s) = S(s)B(s)$. Hence $|T(iy)| = |B(iy)||S(iy)| = |S(iy)|$ for all $y \in \mathbb{R}$. A similar argument establishes uniqueness.

The factor $B(s)$ is called a finite Blaschke product, the inner factor of T or an all pass filter. The $S(s)$ is called an outer factor or minimum phase factor. In Chap. 8, we show how to introduce all pass filters by means of linear systems specified by matrices. □

5.11 Nyquist's Criterion for Stability of *T*

Consider the feedback loop with constant feedback -1, so the transfer function is $T = R/(1+R)$.



Theorem 5.30 (Nyquist's Criterion) *Let R be the transfer function of a plant such that R is stable. Suppose that the contour $R(i\omega)$ $(-\infty \leq \omega \leq \infty)$ does not pass through or wind around -1. Then $T = R/(1+R)$ is also stable, so the feedback system with constant feedback -1 is also stable.*

Proof First we give a proof that depends upon the Argument Principle, then in the next section we give a more detailed proof that uses contour integration. We let $c = \lim_{s\to\infty} R(s)$ where $c \neq -1$ by assumption. Hence we can write $R(s) = c + p(s)/q(s)$ where $p(s)$ and $q(s)$ are polynomials, and the degree of $p(s)$ is less than the degree of $q(s)$. Then

$$T(s) = \frac{R(s)}{1+R(s)} = \frac{c + p(s)/q(s)}{c+1+p(s)/q(s)} = \frac{cq(s)+p(s)}{(1+c)q(s)+p(s)} \tag{5.77}$$

and the degree of $(1+c)q(s)+p(s)$ equals the degree of $q(s)$, hence $T(s)$ is proper. Note that poles of R give finite values of T. So the possible poles of $T(s)$ are the zeros of $1 + R(s)$, and these are not canceled by the zeros of $R(s)$. Let

- N be the number of times that the Nyquist contour of R winds around -1, clockwise;
- Z be the number of zeros of $R(s) + 1$ in the left half-plane;
- P be the number of poles of $R(s) + 1$ in the left half-plane;

Then, by the Argument Principle of complex analysis applied to a semicircular contour in the left half-plane,

$$N = Z - P.$$

Here $N = P = 0$ by hypothesis, so $Z = 0$. Hence T has no poles in the left half-plane, hence T is stable. □

5.12 Nyquist's Criterion Proof

Proposition 5.31 (Nyquist's Criterion) *Let $R(s) = p(s)/q(s)$ where $p(s)$ and $q(s)$ are complex polynomials with degree of $p(s)$ less than the degree of $q(s)$, and suppose that $q(s)$ has all its zeros in LHP. Suppose that the Nyquist contour $R(\imath\omega)$ for $-\infty \leq \omega \leq \infty$ does not pass through or wind around -1. Then $R/(1+R)$ is a strictly proper and stable rational function.*

Proof

(i) First, we show that there exists M_1 such that

$$|R(s)| \leq \frac{M_1}{1+|s|} \tag{5.78}$$

and that there exists $M_2 > 0$ such that

$$\left|\frac{dR}{ds}\right| \leq \frac{M_2}{1+|s|^2} \tag{5.79}$$

for all $s \in RHP$. Since the degree of $q(s)$ is greater than the degree of $p(s)$, we have

$$sR(s) = \frac{sp(s)}{q(s)} \to c \qquad (|s| \to \infty); \tag{5.80}$$

for some $c \in \mathbb{C}$. Also, $q(s)$ has only finitely many zeros, so we can choose r_0 to be the largest modulus of any zero of p; then we can choose M_1 such that

$$|R(s)| \leq \frac{M_1}{1+|s|} \tag{5.81}$$

for all $|s| \geq r_0 + 1$. By hypothesis, R is stable, so R is holomorphic and hence bounded on $\{z : \Re z \geq 0; |z| \leq r_0 + 1\}$. So by changing M_1 if necessary, we obtain the stated upper bound for all $s \in RHP$.

Likewise

$$\frac{dR}{ds} = \frac{\frac{dp}{ds}q(s) - p(s)\frac{dq}{ds}}{q(s)^2} \tag{5.82}$$

where the degree of $q(s)^2$ exceeds the degree of $\frac{dp}{ds}q(s) - p(s)\frac{dq}{ds}$ by two. Also, $q(s)$ has only finitely many zeros, so we can choose r_0 to be the largest modulus of any zero of q; then we can choose M such that

$$\left|\frac{dR}{ds}\right| = \left|\frac{\frac{dp}{ds}q(s) - p(s)\frac{dq}{ds}}{q(s)^2}\right| \leq \frac{M}{1+|s|^2} \tag{5.83}$$

for all s such that $|s| > r_0+1$. By hypothesis, R is stable, so R' is holomorphic and hence bounded on $\{z : \Re z \geq 0; |z| \leq r_0 + 1\}$. So by changing M if necessary, we obtain the stated upper bound for all $s \in RHP$.

(ii) Let S_r be the semicircle in the left half-plane $S_r : z = re^{\imath\theta}$ for $-\pi/2 \leq \theta \leq \pi/2$. For $s = re^{\imath\theta}$ on S_r and $r > M_1$, we have

$$\left|\frac{\frac{dR}{ds}}{1 + R(s)}\right| \leq \frac{M_2/(1 + r^2)}{1 - M_1/(1 + r)} \tag{5.84}$$

so

$$\left|\int_{S_r} \frac{\frac{dR}{ds}ds}{1 + R(s)}\right| \leq \frac{2\pi M_2 r(1 + r)}{(1 + r^2)(1 + r - M_1)}, \tag{5.85}$$

hence

$$\int_{S_r} \frac{\frac{dR}{ds}ds}{1 + R(s)} \to 0 \tag{5.86}$$

as $r \to \infty$.

(iii) Let $\gamma_r = S_r \oplus [\imath r, -\imath r]$ be the contour made of joining the ends of the semicircle S_r with part of the imaginary axis; then let $\Gamma_r = R(z)$ for z on γ_r be the image of γ_t under R. We show that for all sufficiently large r, the contour Γ_r does not pass through or wind around -1. Note that the image of the contour γ_r under the holomorphic map R is again a contour. The image of the interval $[\imath r, -\imath r]$ is $\{T(\imath\omega) : -r \leq \omega \leq r\} \subset \{T(\imath\omega) : -\infty \leq \omega \leq \infty\}$, which does not pass through -1. Also,

$$|R(re^{\imath\theta})| \leq \frac{M_1}{1 + r} < 1 \tag{5.87}$$

for all $r > M_1$, so $R(S_r)$ does not pass through -1. Indeed, Γ_r does not pass through or wind around -1 for all sufficiently large r.

(iv) Let

$$J_r = \frac{1}{2\pi i}\int_{\gamma_r} \frac{\frac{dR}{ds}ds}{1 + R(s)}; \tag{5.88}$$

we aim to prove that $J_r = 0$ for all sufficiently large r. The function $1 + R$ is holomorphic on RHP, so $R'/(1 + R)$ is holomorphic, except where $1 + R$ has zeros. Suppose that $1 + R$ has a zero of order m at s_0. Then there is a holomorphic function $g(s)$ such that $g(s_0) \neq 0$ and $1 + R(s) = (s - s_0)^m g(s)$ on some neighbourhood of s_0, so

$$\frac{\frac{dR}{ds}}{1 + R(s)} = \frac{m}{s - s_0} + \frac{\frac{dg}{ds}}{g(s)}, \tag{5.89}$$

where $(dg/ds)/g(s)$ is holomorphic on some neighbourhood of s_0, so

$$\mathrm{Res}\Big(\frac{\frac{dR}{ds}}{1+R}; s_0\Big) = m. \tag{5.90}$$

By Cauchy's Residue Theorem,

$$J_r = \sum_{j=1}^{n_r} \mathrm{Res}\Big(\frac{\frac{dR}{ds}}{1+R}; s_j\Big) = \sum_{j=1}^{n_r} m_j, \tag{5.91}$$

where the sum is over all the orders of all zeros s_j inside γ_r. Hence J_r is a non negative integer, and increases with increasing r. Now

$$J_r = \frac{1}{2\pi\imath}\int_{S_r} \frac{\frac{dR}{ds}ds}{1+R(s)} + \frac{1}{2\pi\imath}\int_{[\imath r,-\imath r]} \frac{\frac{dR}{ds}ds}{1+R(s)}, \tag{5.92}$$

and by (ii) we deduce that

$$J_r \to \frac{1}{2\pi\imath}\int_{[\imath\infty,-\imath\infty]} \frac{\frac{dR}{ds}ds}{1+R(s)} \tag{5.93}$$

as $r \to \infty$. The final integral converges, by the estimates from (i). An increasing function which takes integer values and is bounded must ultimately be constant, so the left-hand side satisfies

$$J_r = \frac{1}{2\pi\imath}\int_{[\imath\infty,-\imath\infty]} \frac{\frac{dR}{ds}ds}{1+R(s)} \tag{5.94}$$

for all r sufficiently large. Now the value of the constant is 0, since γ_r does not pass through or wind around -1. Hence $J_r = 0$ for all $r > 0$, since the left-hand side increases with increasing r. We deduce that $1 + R$ has no zeros inside γ_r for all $r > 0$, hence has no zeros in the left half-plane.

(v) Finally, we deduce that $1+R(s)$ has all its zeros in LHP, and hence that $R/(1+R)$ is a strictly proper and stable rational function. By (iv), we deduce that $1+R$ has all its zeros in LHP, and by hypothesis R has all its poles in LHP. Hence $R/(1+R)$ has all poles in LHP and is strictly proper.

□

Remark 5.32 (Root Locus) The Nyquist Criterion Theorem 5.30 appears to emphasize the controller $K = -1$ unduly; however, this is to simplify the statement of the result. For a rational function G, we can let $\kappa > 0$ be a positive parameter and consider $K = -1/\kappa$ which corresponds to the transfer function $\kappa G/(1 + \kappa G)$, so that the zeros of

$$1 + \kappa G(s) = 0 \tag{5.95}$$

give rise to poles of the transfer function. When viewed as functions of $\kappa \in [0, \infty)$, the zeros give the root locus, and by Weierstrass's preparation theorem of complex analysis the root locus is made up of continuous curves; see [27] p 267. The root locus plot shows in particular if any roots lie in the RHP, and hence give unstable poles of the transfer function. MATLAB has a convenient function *rlocus* for plotting the root locus.

The region $\mathbb{C} \setminus (-\infty, -1]$ consists of the complex plane with part of the negative real axis removed, and is starlike with star centre in the sense that for all $\zeta \in \mathbb{C} \setminus (-\infty - 1]$ and $\kappa \in [0, \infty)$, the point $\kappa\zeta \in \mathbb{C} \setminus (-\infty - 1]$. If $G(i\omega) \in \mathbb{C} \setminus (-\infty - 1]$ for all $-\infty \leq \omega \leq \infty$, then the Nyquist contour of G does not pass through or wind around -1. Hence $1 + \kappa G(i\omega) \in \mathbb{C} \setminus (-\infty, 0]$ for all $-\infty \leq \omega \leq \infty$, so the Nyquist contour of $1 + \kappa G$ does not pass through or wind around -1, and $1 + \kappa G(i\omega)$ is nonzero for all $-\infty \leq \omega \leq \infty$. This helps to describe the effect of scaling some transfer functions. There are commands in MATLAB that describe the ways in which a Nyquist contour can cross the axis.

If the Nyquist contour crosses $(-1, 0)$ (but possibly not $(-\infty, -1]$), then the gain margin is the smallest $\kappa > 1$ such that $1 + \kappa G(i\omega) = 0$ for some $-\infty \leq \omega \leq \infty$. The gain margin measures how much we need to scale up the Nyquist diagram of G for marginal instability.

If the Nyquist contour of G crosses the unit circle $C(0, 1)$ but does pass through -1, then the phase margin is the smallest $|\phi|$ such that $e^{i\phi}G(i\omega) + 1 = 0$ for some $-\infty \leq \omega \leq \infty$. This measures how much we need to rotate the Nyquist diagram of G, or lag the phase, for marginal instability.

5.13 *M* and *N* Circles

We introduce a geometrical device which will enable us to visualize both R and T by a single Nyquist plot. We consider the Argand diagram, namely the Euclidean plane with complex coordinates. Let φ be the Möbius transformation

$$\varphi(z) = \frac{az + b}{cz + d}; \tag{5.96}$$

by general theory φ maps circles and straight lines to circles and straight lines. For example, the map $z \mapsto z/(1 + z)$ takes the imaginary axis to the circle $\{s \in \mathbb{C} : |s - 1/2| = 1/2\}$. In particular, we consider the relations

$$T = \frac{R}{R + 1}, \qquad R = \frac{T}{1 - T}. \tag{5.97}$$

In the T plane, an M circle is determined by $|T| = M$, and the M circles give a concentric family of circles with radius M and centre 0 such that every point in $\mathbb{C} \setminus \{0\}$ lies on precisely one M circle. In the T plane, an N circle is the straight

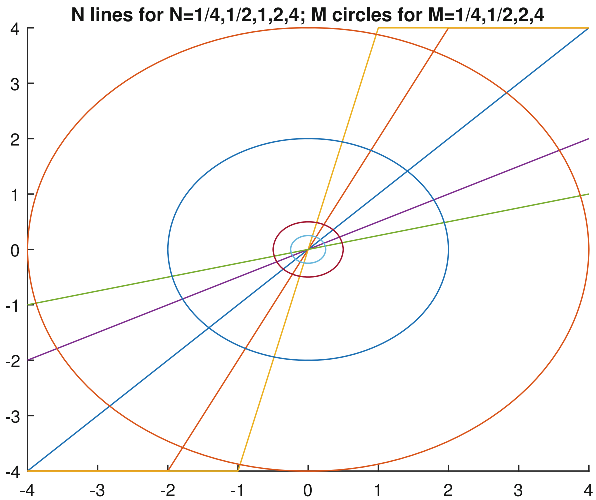


Fig. 5.5 M circles and N lines

line through 0 with gradient N where $-\infty < N \leq \infty$, and we take $N = \tan\phi$ for $-\pi/2 < \phi \leq \pi/2$. Every point in $\mathbb{C} \setminus \{0\}$ lies on precisely one N circle. The M circles and N circles intersect at right angles (Fig. 5.5).

We map these back to the R plane, retaining the names M and N circles. Then $R = u + iv$ is on an M circle if

$$M^2 = |T|^2 = \left|\frac{u+iv}{u+1+iv}\right|^2 \tag{5.98}$$

or

$$(M-1)u^2 + 2M^2u + (M^2-1)v^2 + M^2 = 0 \tag{5.99}$$

so

$$\left(u + \frac{M^2}{M^2-1}\right)^2 + v^2 = \frac{M^2}{(M^2-1)^2}, \tag{5.100}$$

so that an M circle in the R plane has centre $-M^2/(M^2-1) + i0$ and radius $M/|M^2-1|$ (Fig. 5.6).

Let $N = \tan\phi$ and consider $R = u + iv$ on an N circle; then

$$e^{i\phi}(u+1+iv) = u+iv \tag{5.101}$$

so

$$\phi = \tan^{-1}(v/u) - \tan^{-1}(v/(u+1)), \tag{5.102}$$

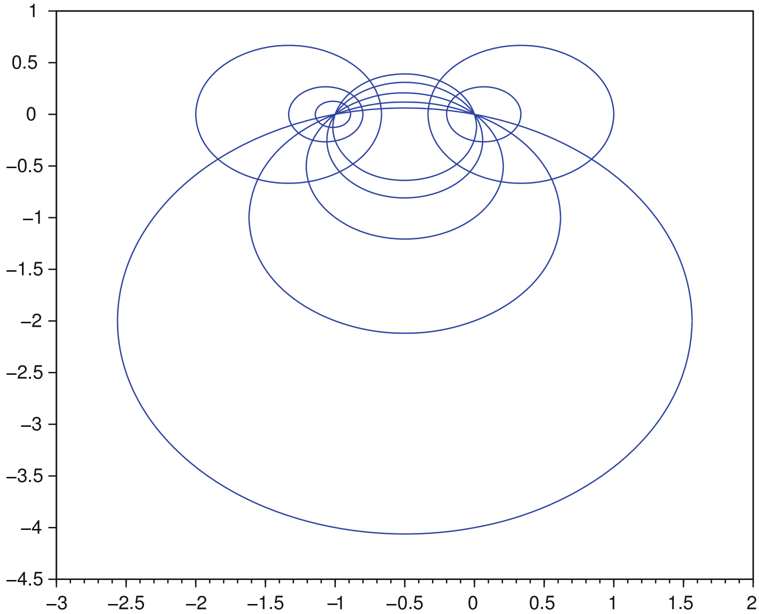


Fig. 5.6 M circles with M=1/4,1/2,2,4,8; N circles with N=1/4,1/2,1,2,4

and by the tangent addition formula

$$N = \tan\phi = \frac{v/u - v/(u+1)}{1 + v^2/(u(u+1))} = \frac{v}{u^2 + u + v^2}, \tag{5.103}$$

so

$$\left(u + \frac{1}{2}\right)^2 + \left(v - \frac{1}{2N}\right)^2 = \frac{N^2+1}{4N^2}, \tag{5.104}$$

so an N circle in the R plane has centre $-1/2 + i/(2N)$ and radius $\sqrt{N^2+1}/(2N)$. Since Möbius transformations are conformal and bijective, the M and N circles intersect at right angles, and every nonzero point in the R plane lies on exactly one M circle and one N circle.

The R plane is plotted with a background of M and N circles, with the following interpretation: in polar coordinates, a typical point is $R = \Gamma e^{i\theta}$ where Γ is the gain and θ is the phase; also R lies on an M circle and an N circle, where $T = R/(1+R)$ has polar decomposition $T = Me^{i\phi}$ where $\tan\phi = N$ (Fig. 5.7).

Example 5.33 The linear system is given by

$$A = \left[1, 2; 3, 4\right], \quad B = \left[0; 1\right], C = \left[3, 5\right], D = 1.$$

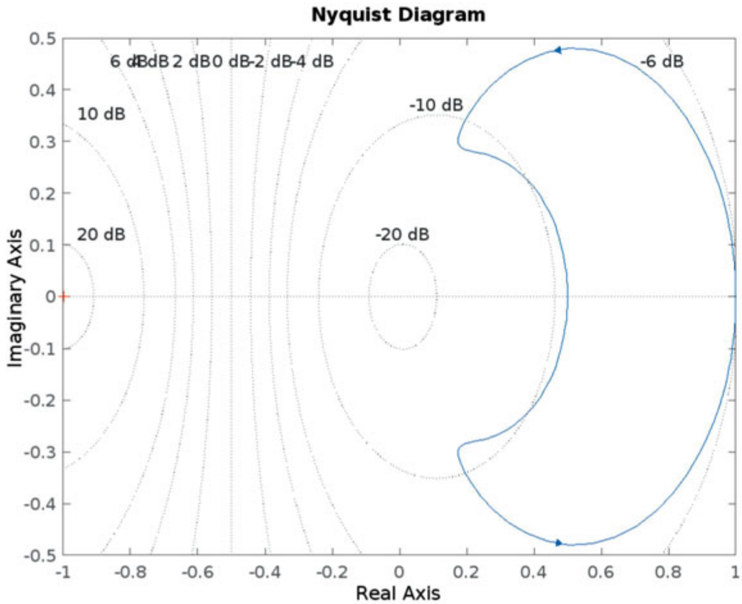


Fig. 5.7 Nyquist plot with grid of M and N circles with transfer function $(s^2 - 1)/(s^2 - 5s - 2)$. The scale on the background plot refers to the gain in units of decibels

To produce the Nyquist plot in MATLAB, one can use the commands

$$>> [b, a] = ss2tf(A, B, C, D)$$

to find the transfer function as a quotient of polynomials, Here the transfer function is

$$R(s) = \frac{s^2 - 1}{s^2 - 5s - 2}. \tag{5.105}$$

then to obtain the Nyquist plot, enter

$$>> R = tf([1, 0, -1], [1, -5, -2])$$

$$>> nyquist(R)$$

Then include the grid by

$$>> grid \quad on$$

5.14 Exercises

Exercise 5.1 Calculate the transfer function T associated with the linear differential equation

$$\frac{d^2y}{dt^2} + 6\frac{dy}{dt} + y = -3\frac{du}{dt} + u. \tag{5.106}$$

with $y(0) = (dy/dt)(0) = 0 = u(0)$; here y is the output and u is the input.

(ii) Find the gain Γ and phase ϕ of the frequency response function $T(i\omega)$.

Exercise 5.2

(i) Let A be a $n \times n$ complex matrix with eigenvalues $\lambda_1, \dots, \lambda_n$. Find the eigenvalues of $A - kI$ for any $k \in \mathbb{C}$.
(ii) Deduce that given any MIMO (A, B, C, D) there exists $k \in \mathbb{C}$ such that $(A - kI, B, C, D)$ is BIBO stable.

Exercise 5.3 (Nyquist and Bode Plots) Recall $s = i\omega$ and let

$$T(s) = \frac{8s + 8\imath + 4}{(s + 1)(s + 2 + \imath)}. \tag{5.107}$$

Take care with complex numbers when carrying out the following plots.

(i) Plot the Nyquist locus of T; that is, plot $\{T(i\omega) : -\infty \leq \omega \leq \infty\}$.
(ii) Let $\Gamma(\omega)$ be the gain and let $\phi(\omega)$ be the phase of T. The Bode plot consists of the graphs of $\log \Gamma(\omega)$ and $\phi(\omega)$ against ω. Produce the Bode plot for $-100 < \omega < 100$.

Exercise 5.4 (More Bode Plots)

(i) Let $T(s) = p(s)/q(s)$, where $p(s)$ and $q(s)$ are polynomials with real coefficients; then $T(s)$ is said to be a real rational function. Show that the gain Γ and phase ϕ of T satisfy

$$\Gamma(\omega) = \Gamma(-\omega), \quad \phi(-\omega) = -\phi(\omega) \qquad (\omega \in \mathbf{R}). \tag{5.108}$$

(ii) For $T(s) = 1/(1 + s)$ and $s = \imath\omega$, plot $\log \Gamma(\omega)$ and $\phi(\omega)$ against ω for $-100 < \omega < 100$.
(iii) When $T(s)$ is a transfer function as in (i), we can plot $\log \Gamma$ and ϕ against $\log \omega$ for $0 < \omega < \infty$. Do this for $T(s) = 1/(1 + s)$.

Exercise 5.5 Let $p(s)$ be a complex polynomial with leading term s^n, and let $r(s) = p(s) - (s + 1)^n$.

(i) Show that

$$R(s) = \frac{r(s)}{(s+1)^n} \tag{5.109}$$

is stable.
(ii) Show that $p(s) = 0$ has no roots in the left half-plane if and only if the Nyquist contour of $R(s)$ does not pass through or wind around -1.

(ii) Hence show that

$$p(s) = s^4 + 3s^3 + 2s^2 + s + 1 \tag{5.110}$$

has zeros in the left half-plane.

Exercise 5.6 Let $p(s) = s^2 - 2s + 7$ and $q(s) = s^3 + 2s^2 + (69/4)s + 65/4$.

(i) Verify that $R(s) = p(s)/q(s)$ is stable.
(ii) By considering the Nyquist locus of R, discuss whether $T = R/(1+R)$ is also stable. Supply graphs to justify your results.
(iii) Replace $p(s)$ by $r(s) = s^2 - 2s - 20$, and repeat (i) and (ii).

Exercise 5.7 Let θ be an improper rational function such that $\theta(s) + 1$ has no zeros in RHP.

(i) Show that $R(s) = -1/(1 + \theta(s))$ is stable.
(ii) If the Nyquist locus of R does not pass through or wind around -1, show that $\theta(s)$ has no zeros in RHP.

Exercise 5.8 At frequency ω, the Nyquist contour of R points in the direction $iR'(i\omega)$. Show that

$$\frac{i\frac{dR}{ds}(i\omega)}{R(i\omega)} = \frac{\frac{d\Gamma}{d\omega}}{\Gamma(\omega)} + i\frac{d\phi}{d\omega} \tag{5.111}$$

is a decomposition into real and imaginary parts.

Exercise 5.9 (Cumulants) Suppose that a piecewise continuous function $y : (0, \infty) \to \mathbb{C}$ satisfies the conditions:

(1) there exist $\kappa, M > 0$ such that $|y(t)| \leq Me^{-\kappa t}$ for all $t > 0$;
(2) $\int_0^\infty y(t)dt \neq 0$.

(i) Show that $Y(s) = \int_0^\infty e^{-st} y(t)\, dt$ converges for all s such that $\Re s > -\kappa$, and defines a holomorphic function of such s.
(ii) Show that $\varphi(s) = \log Y(s)$ defines a holomorphic function on $\{s : |s| < \delta\}$ for some $0 < \delta \leq \kappa$, hence has a convergent power series

$$\varphi(s) = \sum_{j=0}^{\infty} \frac{c_j s^j}{j!} \qquad (|s| < \delta).$$

In probability theory, one considers the case where $y \geq 0$ and $\int_0^\infty y(t)dt = 1$, and the terms $(-1)^j c_j$ are known as cumulants.

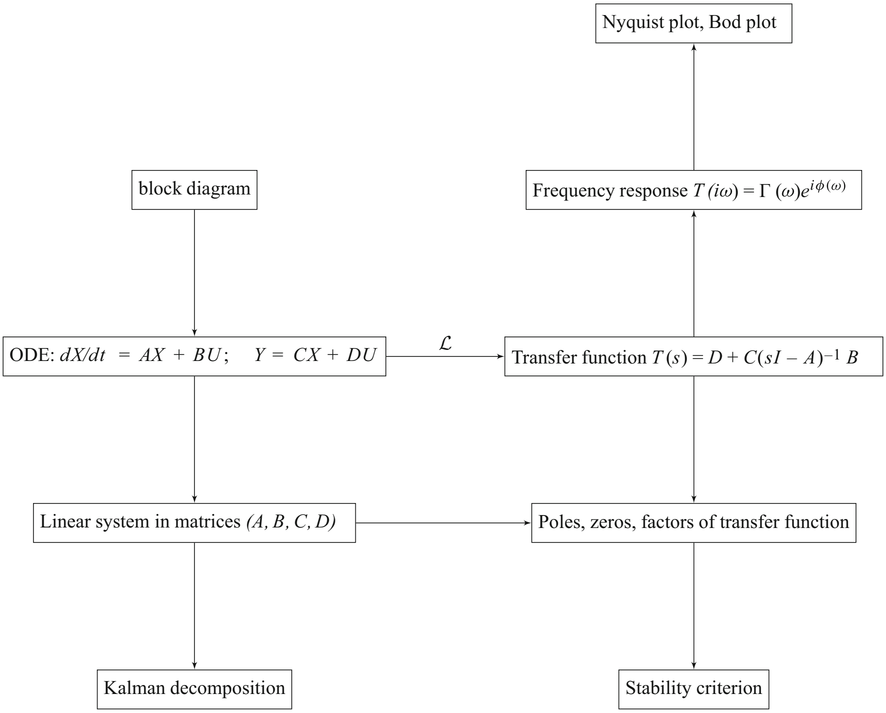
(iii) Let y_1 and y_2 satisfy (1) and (2) for some constants. Show that $y = y_1 * y_2$ also satisfies (1) and (2), for some constants.

(iv) Show that the φ corresponding to y and the φ_1 and φ_2 corresponding to y_1 and y_2 satisfy

$$\varphi(s) = \varphi_1(s) + \varphi_2(s),$$

and the c_j are likewise additive.

(v) For $y(t) = e^{-\kappa t} \sin at$ where $\kappa > 0$ and $a > 0$, obtain an expression for $\varphi(s)$ and the c_j.



Chapter 6
Algebraic Characterizations of Stability

The results of the previous chapter provide a geometrical and analytical approach to the problem of stability. The tools they use are effective when implemented by modern computers. However, the solutions they provide are often only approximate, as they depend upon solving algebraic or transcendental equations which often admit only of numerical rather then exact solutions. In this chapter, we take rational transfer functions, and consider algebraic approaches to stability.

- We use algorithms which can be carried out in exact arithmetic without approximation, such as:
- polynomial long division which gives a Euclidean algorithm for polynomials;
- elementary row operations for matrices with polynomial entries.
- We move between conditions on coefficients of polynomials and entries of matrices.
- Basic computer algebra makes determinant calculations easy, so we present Hurwitz's Theorem 6.12 solving Maxwell's stability problem.

This discussion involves functions $F(s)$ where s is the variable that arises in the Laplace transform. At the end of the chapter, we use the inverse Laplace transform to take us back to state space models in terms of $f(t)$.

6.1 Feedback Control

Vehicles are usually under human control.

Example 6.1

(i) An airliner starts at rest on the runway and then is accelerated until it achieves lift-off. To achieve this, the engines are run almost flat out. Once the aircraft

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_6

reaches cruising altitude, the pilot will reduce the amount of fuel going into the engines, and fly at a steady speed.

(ii) A car on the motor way is not run flat out. Instead, the driver regulates the speed by means of the accelerator, so that when the car is going too fast, say over 70 miles per hour, the driver can reduce the amount of fuel going into the engine, and thus slow down the car. Likewise the driver can speed up the car by allowing more fuel to go into the engine. Regulating speed thus involves the continual attention of the driver; so can we automate this?

Remark 6.2 (Feedback Controllers) The engineer Watt produced various controllers (governors) for steam engines and developed the principle of feedback control. The output is fed through a machine back into the input. A plant is some sort of machine, described by a rational matrix transfer function. Consider a plant given by a linear system G $(m \times k)$ and a controller represented by a linear system $-K$ $(k \times m)$. The output from G is fed back into the input. The minus sign indicates that we want negative feedback (if the engine goes too fast, the controller will slow it down).

- Controllers generally use negative feedback.
- Usually, plants are described by their Laplace transforms.
- Rational transfer functions and controllers are simplest to deal with.
- Proportional-integral-differentiator controllers PID are widely used, as discussed below in Sect. 6.2.

Definition 6.3 (Simple Feedback Loop)

(i) Suppose that G has k inputs and m outputs. We choose K to have m inputs and k outputs. Then the simple feedback loop (SFL) has transfer function

$$T = (I + GK)^{-1}G. \tag{6.1}$$

(ii) Suppose that G and K are rational matrix functions. The poles of G are called open loop poles; the poles of T are called closed loop poles.

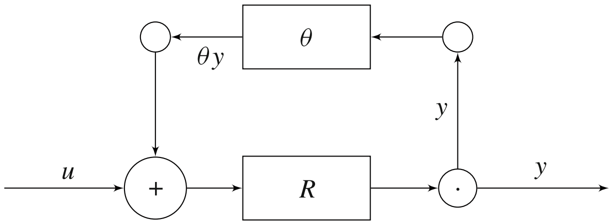
6.2 PID Controllers

Definition 6.4 PID (proportional-integral-derivative) controllers K have the form

$$K(s) = a + \frac{b}{s} + cs, \tag{6.2}$$

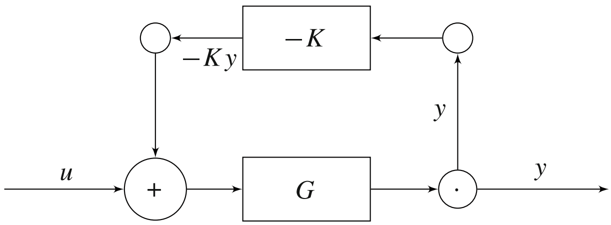
where a, b, c are constants, usually real. The differentiator is expressed in s-space as cs.

In 1866, Robert Whitehead, a naval engineer from Bolton in England, invented PD controllers for torpedoes, in which the differentiator moves the errant torpedo abruptly back on track. In more sensitive applications, PI controllers are preferable, as they have a milder effect on the system. A PID controller has three parameters a, b, c, and by choosing these carefully, engineers can often select a controller that ensures stability while having the right degree of responsiveness in the resulting controlled system. For instance, the steering of a car should be stable, so that the car does not drift off the intended route, but the driver should still be able to change direction of the vehicle. Note that $K(s)$ is unstable unless $b = c = 0$. It may seem paradoxical to stabilize an unstable plant by adding an unstable controller, but this choice is sometimes made. Inevitably, this poses potential difficulties which we address later in our discussion of internal stability.



Problem 1 Given a stable plant $R(s)$, we consider a feedback loop with proportional feedback θ, so that the new transfer function is $T = R/(1 - \theta R)$. When is this stable?

Certainly T is stable for $\theta = 0$, so the question is how much latitude we have in the choice of θ while retaining a stable system. Such feedback might arise deliberately, as in Black's amplifier Example 1.7, or inadvertently; this issue is whether the feedback can be accommodated. By a simple scaling argument, replacing $-\theta R$ by R, we note that the case of $T = R/(1 + R)$ is sufficiently general as to give results in the general case. Nyquist's Criterion Theorem 5.30 is therefore formulated for a stable plant and a simple feedback look with proportional controller $K = -1$, giving $T = R/(1 + R)$.



Problem 2 Given a SISO system with rational transfer function G, can we find a rational controller $-K$ such that

$$T = (I + GK)^{-1}G \tag{6.3}$$

is stable rational?

This is a more general question in which $G(s)$ is not necessarily stable, but we allow K to be a rational controller, possibly unstable.

Example 6.5 (Proportional Feedback) Consider a SISO system with transfer function $G(s) = p(s)/q(s)$ with p, q complex polynomials. The zeros of p are called open loop zeros, while the zeros of q are called open loop poles. Then the controller $-K = -k$ with constant $k > 0$ gives a proportional negative feedback loop with

$$T = (I + KG)^{-1}G = \frac{p}{kp + q}. \tag{6.4}$$

In particular, let $p(s) = s^2$ and $q(s) = bs + c$ where $b, c > 0$. Then $G(s)$ is not proper; whereas

$$T(s) = \frac{s^2}{ks^2 + bs + c} \tag{6.5}$$

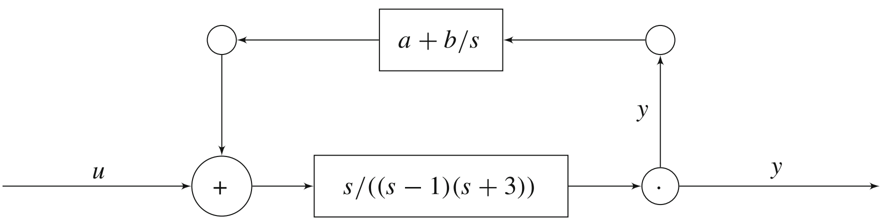
is proper and has poles in the open left half-plane at

$$s = \frac{-b \pm \sqrt{b^2 - 4ck}}{2k}, \tag{6.6}$$

so the controlled system with T is stable.

Example 6.6 (Proportional-Integral Controller) Find a PI feedback controller of the form $-a - b/s$ that stabilizes the plant with transfer function

$$G(s) = s/(s - 1)(s + 3).$$



The plant has poles at $s = 1, -3$, hence is unstable. We consider

$$\frac{G}{1 + GK} = \frac{1}{(1/G) + K} = \frac{s}{s^2 + (2 - a)s - b - 3}. \tag{6.7}$$

For stability we require positive coefficients, so $2 - a > 0$ and $-b - 3 > 0$. In particular, we choose $b = -4$ and $a = 0$, and get

$$\frac{G}{1 + GK} = \frac{s}{(1 + s)^2}, \tag{6.8}$$

which is stable.

6.3 Stable Cubics

Proposition 6.7 *Let a, b, c, A, B, C be real constants. Then*

(i) $s + a$ *is stable, if and only if* $a > 0$;
(ii) $s^2 + bs + c$ *is stable, if and only if* $b, c > 0$;
(iii) $s^3 + As^2 + Bs + C$ *is stable, if and only if* $A, B, C > 0$ *and* $AB - C > 0$.

Proof

(i) Obvious.
(ii) The roots of the quadratic are

$$z_{\pm} = \frac{-b \pm \sqrt{b^2 - 4c}}{2}, \tag{6.9}$$

where $c = z_+ z_-$ and $b = -(z_+ + z_-)$. If $c < 0$, then $b^2 - 4c > 0$ and $s^2 + bs + c$ has real zeros of opposite sign, hence is unstable. If $c = 0$, then $s = 0$ is a zero, so the quadratic is unstable. Hence we need $c > 0$; also $-b$ as the sum of the roots, must be negative, so $b > 0$. Conversely, if $b, c > 0$, then $b^2 - 4c < b^2$ and either there are two negative roots, or a pair of complex conjugate roots in LHP.
(iii) Note that $s^3 + As^2 + Bs + C \to \infty$ as $s \to \infty$ and $s^3 + As^2 + Bs + C \to -\infty$ as $s \to -\infty$. By the intermediate value theorem there exists a real root $s = -a$. Hence we have a factorization

$$s^3 + As^2 + Bs + C = (s + a)(s^2 + bs + c) \tag{6.10}$$

where $A = a + b$, $B = ab + c$ and $C = ac$. Now the cubic is stable if and only if $s + a$ and $s^2 + bs + c$ are stable; that is, if and only if $a, b, c > 0$ by (i) and (ii). Hence $A, B, C > 0$, and

$$AB - C = a^2 b + bab + bc = b(a(a + b) + c) > 0. \tag{6.11}$$

Conversely, if $A, B, C, AB - C > 0$, then $a, b, c > 0$. First note that $ac > 0$ so a and c have the same sign, and if $a, c < 0$, then we cannot have $A, B > 0$. So $a, c > 0$, and $AB - C = b(Aa + c) > 0$, so $b > 0$ also.

□

Example 6.8 (Governors) In the nineteenth century, the term governor was used for what we would now call a proportional-integral controller; see [32] for an historical account. Maxwell [40] considered Jenkin's governor, and found that the nature of solutions depended upon the roots of a cubic equation

$$MBn^3 + (MY + FB)n^2 + FYn + FG = 0 \tag{6.12}$$

where M, B, Y, F, G are various real constants. He solved this cubic by the trigonometric method described in the following exercise.

Example 6.9 (Discriminant of the Cubic) Suppose that the depressed cubic equation $z^3 + pz + q = 0$ has roots z_1, z_2 and z_3. From the factorization

$$z^3 + pz + q = (z - z_1)(z - z_2)(z - z_3)$$

we deduce by comparing coefficients that

$$0 = z_1 + z_2 + z_3, \quad p = z_1 z_2 + z_2 z_3 + z_1 z_3, \quad q = -z_1 z_2 z_3;$$

hence one can use the identities $z_j^3 + pz_j + q = 0$ for $j = 1, 2, 3$ to show

$$0 = (z_1 + z_2 + z_3)^2 = z_1^2 + z_2^2 + z_3^2 + 2p;$$

$$0 = z_1^3 + z_2^3 + z_3^3 + 3q = 0;$$

$$0 = z_1^4 + z_2^4 + z_3^4 - 2p^2.$$

Let

$$\delta = \begin{vmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \\ z_1^2 & z_2^2 & z_3^2, \end{vmatrix}$$

so that by the multiplicative property of determinants

$$\begin{aligned} \delta^2 &= \begin{vmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \\ z_1^2 & z_2^2 & z_3^2 \end{vmatrix} \begin{vmatrix} 1 & z_1 & z_1^2 \\ 1 & z_2 & z_2^2 \\ 1 & z_3 & z_3^2 \end{vmatrix} \\ &= \begin{vmatrix} 3 & 0 & -2p \\ 0 & -2p & -3q \\ -2p & -3q & 2p^2 \end{vmatrix} \\ &= -4p^3 - 27q^2, \end{aligned}$$

which is given by the coefficients of the cubic. Note that δ^2 is real if p and q are real.

Exercise (Trigonometric Solutions of the Cubic) It is possible to obtain algebraic surd expressions for the roots of cubics and biquadratics by a carefully chosen sequence of substitutions. These results were known in the 16th century and were published by Cardano. Unfortunately, the surd expressions are rather complicated, and do not make it easy to see where the roots are located in the complex plane. For this reason, the trigonometric solution is sometimes preferable, and gives accessible conditions for the roots of the real cubic to be real. The following exercise gives the method, which was know to Viete.

Solving the depressed cubic by trigonometry

(i) Show that the substitution $s = z - A/3$ reduces the real cubic equation

$$s^3 + As^2 + Bs + C = 0 \tag{6.13}$$

to the depressed cubic in the style of Scipione del Ferro

$$z^3 + pz + q = 0 \tag{6.14}$$

where the new real coefficients are

$$p = B - A^2/3, \qquad q = C - AB/3 + 2A^3/27. \tag{6.15}$$

(ii) Show that $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$ for all $\theta \in \mathbb{C}$; this is the crucial identity.

(iii) Let $\gamma^2 = (-4p/3)$ so that γ is real for $p \leq 0$ and purely imaginary for $p > 0$. For $p \neq 0$, and $z = \gamma\cos\theta$, show that the depressed cubic reduces to

$$\cos 3\theta = \frac{3q}{\gamma p}. \tag{6.16}$$

(iv) Suppose that $p < 0$. Show that for $-1 < 3q/(\gamma p) < 1$, there are three real roots for the depressed cubic equation in (i), given by

$$z = \gamma\cos\theta, \quad \gamma\cos(\theta + 2\pi/3) \quad \gamma\cos(\theta - 2\pi/3) \tag{6.17}$$

where $\theta \in \mathbb{R}$ satisfies the identity in (iii).

(v) Suppose that $p < 0$. Show likewise that for $3q/(\gamma p) \in (1, \infty)$ and for $3q/(\gamma p) \in (-\infty, -1)$ and for there is a real root and a pair of complex conjugate roots. In the last case, it helps to consider $3\theta = \pi + 3i\phi$ with $\phi \in \mathbb{R}$, so $\cos 3\theta = -\cosh 3\phi$.

(vi) Suppose that $p > 0$. Show that there is a real root, and a pair of complex conjugate roots. Here it helps to consider $3\theta = \pi/2 + 3i\phi$.

(vii) Let $\Delta = -(4p^3 + 27q^2)$ be the discriminant of the depressed cubic. Deduce that (i) has three real roots if and only if $\Delta > 0$.

Example 6.10 Suppose that $G(s) = p(s)/q(s)$ is a plant where $p(s), q(s) \in \mathbb{C}[s]$, and we wish to stabilize G in a simple feedback loop involving the PID controller $K(s) = a + b/s + cs$. Then we have

$$\frac{1}{1+GK}\begin{bmatrix} 1 & G \\ K & GK \end{bmatrix} = \frac{1}{sq + p(as+b+cs^2)}\begin{bmatrix} sq & sp \\ (as+b+cs^2)q & p(as+b+cs^2) \end{bmatrix}. \tag{6.18}$$

The common denominator is

$$\Delta(s) = sq + p(as + b + cs^2), \tag{6.19}$$

and we require all the zeros of $\Delta(s)$ to lie in LHP. A necessary condition for stability is that all the coefficients of $\Delta(s)$ are of the same sign; this gives a system of linear inequalities in a, b, c, one inequality for each coefficient of $\Delta(s)$. By linear programming, we either have no solution, or a feasible region in which all the linear inequalities are satisfied. We can then determine whether some points in this feasible region do indeed give roots in LHP.

Example 6.11 Consider a plant $G(s) = s^2 + \beta s + \gamma$, where $\gamma > 0$ and $\beta \in \mathbb{R}$, and form the simple feedback loop with a PI controller $K(s) = a + b/s$. Then the transfer function is

$$\frac{G}{1+KG} = \frac{s^3 + \beta s^2 + \gamma s}{s + (s^2 + \beta s + \gamma)(as + b)} = \frac{s^3 + \beta s^2 + \gamma s}{as^3 + (b + \beta a)s^2 + (1 + \beta b + \gamma a)s + \gamma b}, \tag{6.20}$$

so by the Proposition 6.7 we have stability when the coefficients of the cubic on the denominator satisfy

$$A = b/a + \beta > 0, \quad B = 1/a + \beta b/a + \gamma > 0, \quad C = b\gamma/a > 0, \tag{6.21}$$

$$AB - C = (b/a + \beta)(1/a + \beta b/a) + \beta\gamma > 0. \tag{6.22}$$

For example, when $\gamma > 0$ and $1 + (2 + \gamma)\beta > 0$, we can choose $a = b = 1$ for a stable system.

Ferrari solved the biquadratic (quartic) equation of degree four by radicals, thus obtaining the roots in terms of algebraic surds. For quintic equations, Hermite demonstrated a solution in terms of elliptic functions. These approaches become increasingly complicated, and for polynomials of higher degree, one has to resort to numerical or graphical approaches for locating the roots. Maxwell's stability question is solved by the Routh–Hurwitz criterion Theorem 6.12 in the next section.

6.4 Hurwitz's Stability Criterion

We consider the equation

$$a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n = 0 \tag{6.23}$$

where $a_0 > 0, a_1, \dots, a_{n-1} \in \mathbb{R}$, and for notational convenience we take $a_j = 0$ for $j < 0$ and for $j > n$. Then we build the $n \times n$ matrix

$$H = \begin{bmatrix} a_1 & a_3 & a_5 & \dots & a_{2n-1} \\ a_0 & a_2 & a_4 & \dots & a_{2n-2} \\ 0 & a_1 & a_3 & \dots & a_{2n-3} \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & * & \dots & a_n \end{bmatrix} \tag{6.24}$$

in which the indices increase in steps of two as we move from left to right along each row, and decrease in steps of one as we move down each column. The leading diagonal of the Hurwitz matrix gives the coefficients of the polynomial in order $a_1, a_2, \dots, a_n$, omitting the leading coefficient a_0.

Theorem 6.12 (Hurwitz) *All the roots of the polynomial equation have negative real parts, if and only if all the leading minors of H are positive, so*

$$\Delta_1 = a_1 > 0, \quad \Delta_2 > 0, \quad \dots, \quad \Delta_n = \det(H) > 0. \tag{6.25}$$

Proof The proof in [30] involves a complicated application of the argument principle, and is omitted. □

In calculations, we can assume that all the coefficients are positive, since this has already been established as a necessary condition for stability. With the aid of computers, Hurwitz's condition 6.12 becomes a feasible calculation for medium sized matrices, and can be carried out in exact arithmetic without root finding. Another advantage is that one can compute the minors when they involve additional parameters, which often happens in control problems.

Example 6.13 For the cubic equation $s^3 + As^2 + Bs + C = 0$, we have

$$H = \begin{bmatrix} A & C & 0 \\ 1 & B & 0 \\ 0 & A & C \end{bmatrix}, \tag{6.26}$$

with leading minors

$$A, \quad \begin{vmatrix} A & C \\ 1 & B \end{vmatrix} = AB - C, \quad \begin{vmatrix} A & C & 0 \\ 1 & B & 0 \\ 0 & A & C \end{vmatrix} = (AB - C)C, \tag{6.27}$$

and Hurwitz condition is that all of these are positive. Considering the last two minors, we see that $C > 0$, so the condition is $A, C, AB - C > 0$, which is equivalent to Proposition 6.7.

Hurwitz's criterion 6.12 turns out to be equivalent to the solution achieved by Routh. In this book we present Hurwitz's version since the former can be expressed in terms of determinants rather then Routh's special arrays.

6.5 Units and Factors

We can solve the stability problem using some commutative algebra. Ultimately we will describe and solve the problem using the polynomial ring $\mathbb{C}[s]$, and it is helpful to introduce some related rings of rational functions.

Definition 6.14 Let R be a commutative ring with 1.

(i) Say that $u \in R$ is a unit if there exists $v \in R$ such that $uv = 1$.
(ii) We say that $f \in R$ divides $g \in R$ if there exists $h \in R$ such that $g = fh$. Such an f is called a factor or divisor of g, denoted $f \mid g$.
(iii) Given nonzero $g_1, g_2 \in R$, an element d is called a greatest common divisor (or highest common factor) if d divides both g_1 and g_2, and all common divisors of g_1 and g_2 divide d.

Remark 6.15

(i) In $\mathbb{Z}$, the units are $\{\pm 1\}$.
(ii) In $\mathbb{Z}$, we let $(f) = \{af : a \in \mathbb{Z}\}$ be the integers that are divisible by f. Then $f|g \Leftrightarrow (f) \supseteq (g)$ for $f, g \in \mathbb{Z} \setminus \{0\}$.
(iii) Note that if d is a greatest common divisor of g_1 and g_2, then ud is also a greatest common divisor for any unit $u \in R$.

6.6 Euclidean Algorithm and Principal Ideal Domains

For a general introduction to ideals and factorization, see [6] chapter XIII.

Definition 6.16 (Ideals)

(i) Let R be a commutative ring with 1. An ideal J is a subset of R such that $0 \in J$, $-a \in J$ for all $a \in J$, $a + b \in J$ for all $a, b \in J$ and $ra \in J$ for all $r \in R$ and $a \in J$. In particular, $\{0\}$ and R are ideals of R.
(ii) An ideal J is called principal if there exists $a \in J$ such that $J = \{ra : a \in R\}$, and such an ideal is denoted (a). Observe that $a|b \Leftrightarrow (a) \supseteq (b)$.
(iii) An integral domain R is a commutative ring with 1 in which $fg = 0$ implies $f = 0$ or $g = 0$. An integral domain R in which all ideals are principal, is called a principal ideal domain. In algebra, the abbreviation PID is commonly used, although this conflicts with the abbreviation we use for proportional-integral-differentiator controllers.

Example 6.17

(i) The prototype of a principal ideal domain is $\mathbb{Z}$ with the usual multiplication and addition. Note that (2) is the ideal of even integers, $(2) \cap (3) = (6)$ and $(2) + (3) = \mathbb{Z}$. Ideals in $\mathbb{Z}$ are studied using Euclid's algorithm.

(ii) In the field of rational functions $\mathbb{C}(x)$, the only ideals are $\{0\}$ and $\mathbb{C}(x)$.

(iii) In $\mathbb{C}[x]$ the ideal $J = \{f \in \mathbb{C}[x] : f(1) = 0 = (df/ds)(1) = f(2)\}$ can be expressed as $J = ((x-1)^2(x-2))$. There is a Euclidean algorithm for polynomials, based upon polynomial long division, which we use to study ideals, and in Proposition 6.18 we show all ideals in this ring are principal.

(iv) The space $\mathbb{C}(s)_p$ of proper rational functions may be regarded as a ring of functions near ∞ since each $F \in \mathbb{C}(s)_p$ has a well defined limit $\lim_{s\to\infty} F(s)$. Hence we can introduce the strictly proper rational functions $\mathbb{C}(s)_0 = \{F(s) \in \mathbb{C}(s)_p : F(\infty) = 0\}$ which is the principal ideal generated by $1/s$. To see this, note that $F(s) = (1/s)(sF(s))$ where $1/s \in \mathbb{C}(s)_0$ and $sF(s) \in \mathbb{C}(s)_p$ for all $F \in \mathbb{C}(s)_0$.

(v) It is not possible to extend the discussion of (iv) to $\mathbb{C}(s)$. The identity $1 = (1+s)(1+s)^{-1}$ shows that one cannot regard elements of $\mathbb{C}(s)$ as a ring of functions from $\mathbb{C} \cup \{\infty\}$ to $\mathbb{C}$ since there can be poles at ∞.

(vi) The principal ideal domains that are most useful in control theory are $\mathbb{C}[s]$ and $\mathbb{C}[1/(1+s)]$.

Proposition 6.18 *The ring* $\mathbb{C}[x]$ *is a principal ideal domain with units* $\mathbb{C} \setminus \{0\}$.

Proof Let f, g be nonzero polynomials with degrees n and m. Then fg is a polynomial of degree $n+m$. So the identity $fg = 1$ occurs only if f and g are nonzero constants.

Let J be an ideal in $\mathbb{C}[x]$. If $J = \{0\}$, then $J = (0)$. If J contains a nonzero constant polynomial λ, then J also contains $\lambda^{-1}\lambda = 1$, so $J = (1) = \mathbb{C}[x]$. Otherwise, we consider the set $\{\text{degree}(f) : f \in J, f \neq 0\}$, which is a non-empty subset of $\mathbb{N}$, hence has a smallest member m. Now $m = \text{degree}(g)$ for some $g \in J$, and if $f \in J$ with $f \neq 0$, then by the Euclidean algorithm for polynomials $f = qg + r$ where $q, r \in \mathbb{C}[x]$ and either $r = 0$ or $\text{degree}(r) < m$. Now $r = f - qg \in J$, so by the minimality of m, we must have $r = 0$. Hence $f = qg$, and we deduce that $J = (g)$. Multiplying g by the reciprocal of its leading coefficient, if necessary, we can suppose that g is monic, and this choice is then unique. □

Corollary 6.19 (Minimal Polynomial) *Let* $A \in M_{n\times n}(\mathbb{C})$. *Then there exists a unique* $m(s) \in \mathbb{C}[s]$ *that is monic and of degree less than or equal to n such that* $p(A) = 0$ *for* $p(s) \in \mathbb{C}[s]$ *if and only if* $m(s)$ *is a factor of* $p(s)$ *in* $\mathbb{C}[s]$.

Proof Consider $J = \{p(s) \in \mathbb{C}[s] : p(A) = 0\}$ and observe that J is an ideal in $\mathbb{C}[s]$. Further, by the Cayley-Hamilton Theorem 2.29, we have $\chi_A(s) \in J$. By Proposition 6.18, there exists a unique monic $m(s)$ such that $J = (m(s))$, and the degree of $m(s)$ is less than or equal to n by (2.9). □

Lemma 6.20 *Let f and g be nonzero in a principal ideal domain R, and $(f, g) = \{af + bg : a, b \in R\}$. Then (f, g) is an ideal in R, and $(f, g) = (h)$ where h is a greatest common divisor of f and g. This h is unique up to multiplication by a unit.*

Proof One easily checks that (f, g) is an ideal, hence $(f, g) = (h)$ for some nonzero $h \in R$. Also, $f = 1f + 0g \in J$, so $f = kh$ for some $k \in R$. Likewise $g = 0f + 1g \in J$, so $g = ph$ for some $p \in R$, hence h is a common divisor of f and g. Conversely, if r is a common divisor of f and g, then $f = ur$ and $g = wr$ for some R, so $h = af + bg = (au + bw)r$, and r is a divisor of h. If $h = rs$ and $r = hk$, then $h = skh$, so $h(1 - sk) = 0$, hence $1 = sk$ since $h \neq 0$, hence s is a unit. □

The formula $h = af + bg$ expressing the greatest common divisor of R as a combination with $a, b \in R$ is known as Bezout's identity. One can write

$$h = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix}. \tag{6.28}$$

We say that f and g are coprime if there exist $a, b \in R$ such that $1 = af + gb$. This is equivalent to $R = \{af + bg : a, b \in R\}$. If R is a Euclidean domain such as $\mathbb{C}[s]$ or $\mathbb{Z}$, one can use Euclid's algorithm to determine whether f and g are coprime by finding a highest common factor h and a, b such that $h = af + bg$.

Definition 6.21 (Euclidean Domain) Let R be an integral domain with unit 1. Say that R is a Euclidean domain if there exists $\delta : R \setminus \{0\} \to \{0, 1, \dots\}$ such that for all $x, y \in R \setminus \{0\}$ there exists $q, r \in R$ such that $x = qy + r$ and either $r = 0$, or $\delta(r) < \delta(y)$.

Example 6.22

(i) The integers $\mathbb{Z}$ give a Euclidean domain for $\delta(x) = |x|$.
(ii) The polynomial ring $\mathbb{C}[s]$ is Euclidean for $\delta(f)$ the degree of the polynomial f. Likewise, $\mathbb{K}[s]$ is a Euclidean domain for any field $\mathbb{K}$.
(iii) The ring $\mathbb{Z}[s]$ is not Euclidean.

Proposition 6.23 *Any Euclidean domain is a principal ideal domain.*

Proof This follows exactly as in Proposition 6.18. □

Algorithm *The iterated Euclidean algorithm applies to a Euclidean domain R with $\delta : R \setminus \{0\} \to \{0, 1, \dots\}$ the Euclidean function.*

Given $x_0, x_1 \in R$ the algorithm determines $a, b, x \in R$ such that $(x) = (x_0, x_1)$ and $x = ax_0 + bx_1$.

Step 0. Let $x_0, x_1 \in R$ and suppose $\delta(x_0) \geq \delta(x_1)$.

Step 1. Introduce $q_1, x_2 \in R$ such that $x_0 = q_1x_1 + x_2$ and either $x_2 = 0$, in which case stop; or $\delta(x_1) > \delta(x_2)$, in which case continue.

Step 2. Introduce $q_2, x_3 \in R$ such that $x_1 = q_2x_2 + x_3$ and either $x_3 = 0$, in which case stop; or $\delta(x_2) > \delta(x_3)$, in which case continue.

Step n. Introduce $q_n \in R$ such that $x_{n-1} = q_nx_n$ in which case stop.

The algorithm terminates since $\delta(x_1) > \delta(x_2) > \delta(x_3) > \ldots$ is a strictly decreasing sequence of nonnegative integers, so must reach 0 in at most $\delta(x_1)$ steps. Also

$$(x_0, x_1) = (x_1, x_2) = (x_2, x_3) = \cdots = (x_{n-1}, x_n) = (x_n).$$

We reverse the steps and make the remainder the subject of the formulas

$$\begin{aligned} x_n &= x_{n-2} - q_{n-1}x_{n-1} \\ x_{n-1} &= x_{n-3} - q_{n-2}x_{n-2} \\ \vdots &\qquad \vdots \\ x_2 &= x_0 - q_1 x_1, \end{aligned}$$

so we recover $a_n, b_n \in R$ such that $x_n = a_n x_0 + b_n x_1$ by substituting back. Equivalently, we can write

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & -q_1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \\ \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & -q_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ \vdots &\quad \vdots \\ \begin{bmatrix} x_{n-1} \\ x_n \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & -q_{n-1} \end{bmatrix} \begin{bmatrix} x_{n-2} \\ x_{n-1} \end{bmatrix} \end{aligned}$$

and we can recover $x_n = a_n x_0 + b_n x_1$ by matrix multiplication.

6.7 Ideals in the Complex Polynomials

Proposition 6.24 *Let F be a nonempty finite set of complex polynomials, and $J(F)$ the ideal in $\mathbb{C}[s]$ generated by F; that is, the intersection of all the ideals in $\mathbb{C}[s]$ that contain F. Then by finitely many applications of the division algorithm, one determines $p \in \mathbb{C}[s]$ such that $(p) = J(F)$.*

Proof We give an algorithm for finding p. For $F = \{p_1, \ldots, p_m\}$ we can describe $J(F)$ explicitly as

$$J(F) = \{h_1 p_1 + \cdots + h_m p_m : h_1, \ldots, h_m \in \mathbb{C}[s]\}. \tag{6.29}$$

We assume that the elements of F are nonzero and let $d(F) = \min\{\deg(p); p \in F\}$.

Start with $F_0 = F$ and choose $p_0 \in F_0$ such that $\deg(p_0) = d(F_0)$. For each $p \in F_0$, we use the Euclidean algorithm for polynomials to write $p = qp_0 + r$ where $q, r \in \mathbb{C}[s]$ and either $r = 0$ or $\deg(r) < \deg(p_0)$. Note that $r = p - qp_0 \in J(F_0)$. If remainder $r = 0$ for all $p \in F_0$, then p_0 divides all $p \in F_0$ and we have $J(F_0) = (p_0)$. Otherwise, there exist nonzero remainders, and we introduce the set of nonzero remainders $F_1 = \{r = p - qp_0; r \neq 0; p \in F_0\}$ where $J(F_1) = J(F_0)$ and $d(F_1) < d(F_0)$. Then we repeat with F_1 instead of F_0.

A strictly decreasing sequence of positive integers is finite, so after at most $d(F)$ steps, we obtain a nonempty finite set of nonzero complex polynomials F_j such that $J(F_j) = J(F)$ and $J(F_j) = (p_j)$ for some $p_j \in F_j$. □

Example 6.25 (Lowest Common Denominator) Let F be a nonempty and finite set of complex rational functions, so $F \subset \mathbb{C}(s)$, and let

$$J = \{p(s) \in \mathbb{C}[s] : p(s)f(s) \in \mathbb{C}[s], \quad \forall f(s) \in F\}. \tag{6.30}$$

Then J is an ideal in $\mathbb{C}[s]$. To see this, note that $0 \in J$ since $0 \in \mathbb{C}[s]$ and for $p(s), q(s) \in J$ we have $(p(s) + q(s))f(s) \in \mathbb{C}[s]$ for all $f(s) \in F$; likewise $p(s)f(s)g(s) \in \mathbb{C}[s]$ for all $p(s) \in J$, $f(s) \in F$ and $g(s) \in \mathbb{C}[s]$.

Now $J = \mathbb{C}[s]$ if and only if $1 \in J$, or equivalently, $F \subset \mathbb{C}[s]$. This happens in particular if $F = \{0\}$. Otherwise, we let $f_j(s) = p_j(s)/q_j(s)$ with $p_j(s), q_j(s) \in \mathbb{C}[s]$ for $j = 1, \dots, n$ be the nonzero elements of F, and observe that $q(s) = q_1(s) \dots q_n(s)$ satisfies $q(s)f_j(s) \in \mathbb{C}[s]$ for all $j = 1, \dots, n$, so $q(s)$ is a nonzero element of J. Then by the Proposition 6.18, there exists $p(s) \in \mathbb{C}[s]$ such that $J = (p(s))$, so we can clear the denominators of the elements of F by multiplying by $p(s)$. This $p(s)$ is often called the lowest common denominator, and we can find it explicitly by the Euclidean algorithm. More precisely, $p(s)$ is the polynomial of lowest degree in $J \setminus \{0\}$, and is unique up to multiplication by a nonzero complex number. This result has a significant application to finding partial factions, and enables us to calculate inverse Laplace transforms, as in Proposition 6.55

6.8 Highest Common Factor and Common Zeros

By Proposition 6.18 and Lemma 6.20, any two nonzero polynomials in $\mathbb{C}[s]$ have a greatest common divisor. In $\mathbb{C}[s]$, the units are precisely the nonzero constant polynomials, so we can replace a greatest common divisor by a monic greatest common divisor. The following result characterizes the case in which the greatest common divisor is 1. The criterion can be assessed via the Euclidean algorithm, which one can carry out in exact arithmetic, with computer assistance as required.

Proposition 6.26 *Nonzero complex polynomials P and Q have no common complex zeros if and only if there exist complex polynomials M and N such that*

$$PM + QN = 1. \tag{6.31}$$

Proof If P and Q have a common zero λ, then

$$P(\lambda)M(\lambda) + Q(\lambda)N(\lambda) = 0,$$

contrary to the equality.

Conversely, suppose that P and Q have no common zeros and carry out the Euclidean algorithm for P and Q to obtain complex polynomials M, N, R such that R is the highest common factor of P and Q and $PM + QN = R$. If $R = r$ is a nonzero constant, then we can multiply through by r^{-1} to obtain $PM/r + QN/r = 1$, as required. Otherwise, R is a complex polynomial of positive degree, and hence by the fundamental theorem of algebra [6] has a complex zero λ. Now R is a factor of both P and of Q, so $s - \lambda$ is a factor of both P and Q, so λ is a common zero of P and Q, contrary to assumption. □

Corollary 6.27 *A nonconstant polynomial P has only simple zeros if and only if the highest common factor of P and P' is* 1.

Proof Now P has multiple zeros if and only if $(s - \lambda)^2$ is a factor of P for some $\lambda \in \mathbb{C}$, or equivalently $P(\lambda) = (dP/ds)(\lambda) = 0$. Otherwise, P and dP/ds have no common zeros and we have the situation of the Proposition 6.26. See also [6], page 403. □

Remark 6.28

(i) MATLAB can find the greatest common divisor (highest common factor) for polynomials via command
>> [g,M,N]=gcd(P,Q)

(ii) Let G be a nonconstant complex rational function. Then we can write $G = P/Q$ for complex polynomials P and Q, and find their highest common factor H by the Euclidean algorithm. First, if $H = 1$, then P and Q have no common complex zeros and the algorithm gives complex polynomials X and Y such that $PX + QY = 1$. The X and Y can can be found using the Euclidean algorithm, which can be carried out in exact arithmetic–no need to find roots of polynomial equations. If H is a polynomial of positive degree, then we can write $P = Hp$ and $Q = Hq$ where p, q are complex polynomials such that $p/q = P/Q = G$, and p and q have highest common factor 1, and we are back in the first case. Thus we can reduce G to $G = p/q$ where p, q are complex polynomials with no common complex zeros.

(iii) The algebra $\mathbb{C}[x, y]$ of complex polynomials in two variables is not a principal ideal domain, but does have unique factorization. See [6], pages 76 and 349.

(iv) The Euclidean algorithm for polynomials [6] page 64 works for $\mathbb{K}[s]$ for any field $\mathbb{K}$ such as the reals $\mathbb{R}$ or the rationals $\mathbb{Q}$, so Proposition 6.18 holds in this context. However, $1+s^2$ and $(1+s^2)^2$ have no common zeros in $\mathbb{R}$, but clearly they have common factor $1+s^2$. Proposition 6.26 makes essential use of the fundamental theorem of algebra, which establishes algebraic completeness of $\mathbb{C}$.

6.9 Rings of Fractions

Definition 6.29 (Multiplicative Sets) Let R be an integral domain, and $S \subseteq R$ such that $1 \in S$ and $s, t \in S$ implies $st \in S$; then S is said to be multiplicative. A multiplicative set S is said to be saturated if $ab \in S$ for $a, b \in R$ implies $a, b \in S$.

Example 6.30 The following sets are multiplicative for the usual multiplication operation:

(i) in an integral domain R, the set $R^\sharp = \{r \in R : r \neq 0\}$;
(ii) the set of units, namely the set of $u \in R$ such that $uv = 1$ for some $v \in R$;
(iii) in $\mathbb{C}[s]$, the set of monic stable polynomials;
(iv) in $\mathbb{C}[s]$, the set $\{(1+s)^n : n = 0, 1, \dots\}$ of powers of $(1+s)$;
(v) in $\mathbb{C}[s]$, the set of even polynomials $f(s)$ so that $f(s) = f(-s)$;
(vi) $\mathcal{S}^\sharp$ the set of nonzero stable rational functions.

The examples (i), (ii), (iii) and (iv) are saturated; whereas (v) is not, since s^2 is an even polynomial which is the product of the odd polynomials s and s. (vi) This is saturated as a subset of the ring of stable rational functions, which forms an integral domain. However $(1+s)/(1+s) = 1$ is stable, whereas $(1+s)$ is not stable, so (vi) is unsaturated in $\mathbb{C}(s)$.

Definition 6.31 (Ring of Fractions)

(i) For a multiplicative subset S of an integral domain R, we introduce $S^{-1}R = \{a/b : a \in R, b \in S\}$, the set of fractions with numerator in R and denominator in S. We identify a/b with c/d when $ac = bd$, and identify $a/1$ with $a \in R$. Then $S^{-1}R$ becomes a commutative ring for the obvious operations $a/b + f/g = (ag+bf)/(bg)$ and $(a/b)(f/g) = (af)/(bg)$, and we can regard R as a subring of $S^{-1}R$.
(ii) In particular $Q(R) = \{P/Q : P \in R, Q \in R^\sharp\}$ gives the field of fractions of R.

Example 6.32

(i) We can choose $S = \{g \in \mathbb{C}[s] : g \neq 0\}$ and $R = \mathbb{C}[s]$ so that $S^{-1}R = \{f/g : f, g \in \mathbb{C}[s] : g \neq 0\} = \mathbb{C}(s)$.

(ii) The ring $\mathcal{S}$ consists of proper rational functions $p(s)/q(s)$ with denominator $q(s)$ a monic stable polynomial. In Proposition 6.58, we summarize the properties of $\mathcal{S}$, and write out the formulas in detail there.
(iii) We can choose $\mathcal{S}^\sharp$ in $\mathcal{S}$, to form $\{P/Q : P \in \mathcal{S}, Q \in \mathcal{S}^\sharp\}$ which we show in Proposition 6.36 out to be all of $\mathbb{C}(s)$.
(iv) The advantage of a saturated set is that we can take a fraction $p/q \in S^{-1}R$ and any factorization $p = p_1 p_2$ and $q = q_1 q_2$ in R with give $p/q = (p_1/q_1)(p_2/q_2)$ with factors p_1/q_1 and p_2/q_2 in $S^{-1}R$.

The following result introduces a ring which is surprisingly important in the theory. We have already encountered this in Exercise 4.6.

Lemma 6.33 *Let $\mathcal{R}$ be the space of proper complex rational functions with poles only at -1. Then $\mathcal{R}$ is a principal ideal domain, and a subring of $\mathcal{S}$.*

Proof Observe that $\{g(s)/(1+s)^n : f(s) \in \mathbb{C}[s], n = 0, 1, \dots\}$ gives a ring of rational functions with the only possible poles at -1 and ∞. When $g(s)/(1+s)^n$ is proper, then the only possible pole is at -1.

For $f(s) \in \mathcal{R}$, there is Laurent expansion about -1 given by

$$f(s) = p(s) + \sum_{k=1}^{n} \frac{a_k}{(s+1)^k}, \tag{6.32}$$

where the principal part $p(s)$ is a polynomial, which reduces to a constant since f is proper. Hence the map $\mathbb{C}[\lambda] \to \mathcal{R}$

$$\sum_{k=0}^{n} a_k \lambda^k \mapsto \sum_{k=0}^{n} \frac{a_k}{(s+1)^k} \tag{6.33}$$

is an isomorphism of algebras. Hence $\mathcal{R}$ is a principal ideal domain. □

In Proposition 6.36, we show that $\{P/Q : P \in \mathcal{R}, Q \in \mathcal{R}^\sharp\}$ is all of $\mathbb{C}(s)$.

Example 6.34 (Changes of Variable in the Rational Functions) Consider $\mathbb{C}(s)$ and let $g(s) \in \mathbb{C}(s)$. Then the map $\lambda \mapsto g(s)$ and $1 \mapsto 1$ determines a homomorphism of fields $\mathbb{C}(\lambda) \to \mathbb{C}(s)$ via $f(\lambda) \mapsto f(g(s))$. Consider $a, b, c, b \in \mathbb{C}$ such that $ad - bc \neq 0$, and write

$$\lambda = \frac{as+b}{cs+d}, \quad s = \frac{d\lambda - b}{-c\lambda + d}. \tag{6.34}$$

There is an isomorphism of fields $\mathbb{C}(\lambda) \to \mathbb{C}(s) : f(\lambda) \mapsto f(\frac{as+b}{cs+d})$ with inverse $f(s) \mapsto f(\frac{d\lambda-b}{-c\lambda+d})$. In particular, we can take

$$\lambda = \frac{1}{s+1}, \quad s = \frac{\lambda - 1}{-\lambda}. \tag{6.35}$$

Let $P(\lambda) \in \mathbb{C}[\lambda]$. Then $P(1/(1+s))$ gives a stable rational function in s. We have a ring $\mathcal{R} = \mathbb{C}[1/(1+s)]$ which is a principal ideal domain and a subring of $\mathcal{S}$.

In the next few sections we consider how the stable rational functions can be used to solve control problems. The implications for Laplace transforms are discussed in the final two sections of this chapter.

6.10 Coprime Factorization in the Stable Rational Functions

Definition 6.35 (Coprime) Let $P(s), Q(s) \in \mathcal{S}$ be non zero. We say that P and Q are coprime if there exist $X(s)$ and $Y(s)$ in $\mathcal{S}$ such that

$$P(s)X(s) + Q(s)Y(s) = 1. \tag{6.36}$$

Proposition 6.36 (Coprime Factorization into Stable Rationals) *Let $G(s)$ be a complex rational function. Then there exist $P(s)$ and $Q(s)$ in $\mathcal{S}$ such that*

$$G(s) = \frac{P(s)}{Q(s)} \tag{6.37}$$

and $P(s)$ and $Q(s)$ are coprime in $\mathcal{S}$. In particular, the field of fractions of $\mathcal{S}$ is $\mathbb{C}[s]$.

Proof

(1) We have shown that $\mathbb{C}(s)$ and $\mathbb{C}(1/(s+1))$ are isomorphic. The issue is to show that we can choose $P(s)$ and $Q(s)$ coprime in $\mathcal{S}$. Since $G(s)$ is rational, we can write $G(s) = M(s)/N(s)$ where complex polynomials M, N have no common zeros.

(2) We introduce a new variable $\lambda = 1/(1+s)$ and write

$$\tilde{M}(\lambda) = \lambda^m M\Big(\frac{1-\lambda}{\lambda}\Big)$$
$$\tilde{N}(\lambda) = \lambda^m N\Big(\frac{1-\lambda}{\lambda}\Big) \tag{6.38}$$

where m is the maximum of the degrees of M and N, so that $\tilde{M}(\lambda)$ and $\tilde{N}(\lambda)$ are polynomials. Now $\tilde{M}(\lambda)$ and $\tilde{N}(\lambda)$ have no common zeros. The problematic case is $\lambda = 0$, but we note that $\tilde{M}(0)$ is the m^{th} coefficient of M, and $\tilde{N}(0)$ is the m^{th} coefficient of N; so either $\tilde{M}(0)$ or $\tilde{N}(0)$ is not zero by the choice of m.

(3) By Proposition 6.26 there exist complex polynomials $\tilde{X}(\lambda)$ and $\tilde{Y}(\lambda)$ such that

$$\tilde{M}(\lambda)\tilde{X}(\lambda) + \tilde{N}(\lambda)\tilde{Y}(\lambda) = 1. \tag{6.39}$$

(4) Finally we convert back to the original variable $s = (1-\lambda)/\lambda$ and introduce

$$P(s) = \tilde{M}\Big(\frac{1}{1+s}\Big); \quad Q(s) = \tilde{N}\Big(\frac{1}{1+s}\Big); \tag{6.40}$$

$$X(s) = \tilde{X}\Big(\frac{1}{1+s}\Big); \; Y(s) = \tilde{Y}\Big(\frac{1}{1+s}\Big); \tag{6.41}$$

so that $P(s), Q(s), X(s), Y(s)$ belong to $\mathcal{S}$. Indeed, they are all proper and the only poles are at $s = -1$. Furthermore, $P(s)$ and $Q(s)$ satisfy

$$P(s)X(s) + Q(s)Y(s) = 1 \tag{6.42}$$

and

$$G(s) = \frac{M(s)}{N(s)} = \frac{\tilde{M}(1/(1+s))}{\tilde{N}(1/(1+s))} = \frac{P(s)}{Q(s)}. \tag{6.43}$$

We have shown that $\mathbb{C}(s)$ is the field generated by $\mathcal{S}$.

□

Algorithm (Coprime Factorization Algorithm for Stable Rationals)

(1) Write $G(s)$ as a quotient of polynomials in s with no common zeros.
(2) Introduce λ via $s = (1-\lambda)/\lambda$ and clear the denominators.
(3) Apply the Euclidean algorithm for polynomials in λ.
(4) Convert back to the original variable s by $\lambda = 1/(1+s)$.

Example 6.37 (Coprime Factorization in Stable Rationals) Let

$$G(s) = \frac{s^2+5}{s^2-s+1}. \tag{6.44}$$

Then with $s = (1-\lambda)/\lambda$, we have

$$G(s) = \frac{(1-\lambda)^2 + 5\lambda^2}{(1-\lambda)^2 - \lambda(1-\lambda) + \lambda^2} = \frac{6\lambda^2 - 2\lambda + 1}{3\lambda^2 - 3\lambda + 1}. \tag{6.45}$$

Then by the Euclidean algorithm

$$\Big(\frac{24\lambda - 2}{7}\Big)\Big(3\lambda^2 - 3\lambda + 1\Big) - \Big(\frac{12\lambda - 9}{7}\Big)\Big(6\lambda^2 - 2\lambda + 1\Big) = 1 \tag{6.46}$$

so letting $\lambda = 1/(1+s)$, we have $PX + QY = 1$ with $P, Q, X, Y \in \mathcal{S}$

$$\Big(\frac{22-2s}{7(1+s)}\Big)\Big(\frac{3}{(1+s)^2} - \frac{3}{1+s} + 1\Big) - \Big(\frac{3-9s}{7(1+s)}\Big)\Big(\frac{6}{(1+s)^2} - \frac{2}{1+s} + 1\Big) = 1. \tag{6.47}$$

Realizing the quotient via linear systems. Suppose that $G(s) = P(s)/Q(s)$ where $P(s)$ and $Q(s)$ are coprime in $\mathcal{S}$, where $Q(s)$ is proper but not strictly proper. Then there exists a BIBO stable linear system $\Sigma_1 = (A_1, B_1, C_1, D_1)$ with transfer function $P(s)$, and a BIBO stable linear system $\Sigma_2 = (A_2, B_2, C_2, D_2)$ with transfer function $Q(s)$. Then there exists a linear system $\Sigma_2^\times = (A_2^\times, B_2^\times, C_2^\times, D_2^\times)$ with transfer function $1/Q(s)$, which is not necessarily stable. Then $G(s)$ is the transfer function that arises from multiplying $1/Q(s)$ and $P(s)$, namely running the linear systems Σ_1 and $\Sigma_2^\times$ in series; see Proposition 7.14.

Definition 6.38 (Unstable Poles) For rational transfer function $G(s)$, the unstable poles of G are the poles in the closed left half-plane $\{s : \Re s \geq 0\} \cup \{\infty\}$.

Corollary 6.39 *The poles of $G(s)$ in $\{s : \Re s \geq 0\} \cup \{\infty\}$ are given by the zeros of $Q(s)$ in $\{s : \Re s \geq 0\} \cup \{\infty\}$.*

Proof Suppose that $s_0 \in \{s : \Re s \geq 0\} \cup \{\infty\}$ has $Q(s_0) = 0$. Then s_0 is not a pole of $X(s)$ or of $Y(s)$, so from the equation $P(s)X(s) + Q(s)Y(s) = 1$, we deduce that $P(s_0)X(s_0) = 1$, so $P(s_0) \neq 0$ and s_0 is a pole of $G(s)$. Conversely, the poles of $G(s) = P(s)/Q(s)$ are either poles of $P(s)$, which are all in LHP, or zeros of $Q(s)$. □

6.11 Controlling Rational Systems

Proposition 6.40 *Let $G = P/Q$ be a complex rational function, as in Proposition 6.36. Then the rational $K = X/Y$ is such that*

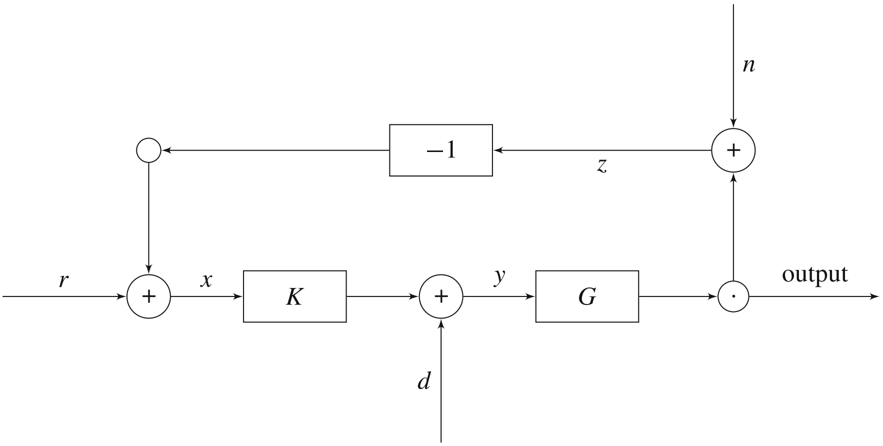
$$\frac{G}{1+GK}, \quad \frac{K}{1+GK}, \quad \frac{1}{1+GK}, \quad \frac{GK}{1+GK} \tag{6.48}$$

are all stable rational functions.

Proof We recall that $PX + QY = 1$, we we have

$$\begin{aligned} \frac{G}{1+GK} &= \frac{P/Q}{1+PX/QY} = PY; \\ \frac{K}{1+GK} &= \frac{X/Y}{1+PX/QY} = XQ; \\ \frac{1}{1+GK} &= \frac{1}{1+PX/QY} = QY; \\ \frac{GK}{1+GK} &= \frac{PX/QY}{1+PX/QY} = PX. \end{aligned} \tag{6.49}$$

Given any rational SISO, we have produced an algorithm for finding a controller to stabilize it. □



Wellposedness of the SFL diagram

The standard form of the rational simple feedback loop linear system has inputs r, d and n and states x, y and z so that

$$\begin{aligned} x &= r - z \\ y &= d + Kx \\ z &= n + Gy \end{aligned} \tag{6.50}$$

for some rational functions G and K, hence

$$\begin{bmatrix} 1 & 0 & 1 \\ -K & 1 & 0 \\ 0 & -G & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r \\ d \\ n \end{bmatrix}. \tag{6.51}$$

Definition 6.41 (Well Posed) The system is said to be well posed if the inputs (r, d, n) uniquely determine the states (x, y, z).

Lemma 6.42 *The SFL system is well posed if and only if* $1 + GK \neq 0$*, and in this case, the inputs determine the states by*

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{1 + GK} \begin{bmatrix} 1 & -G & -1 \\ K & 1 & -K \\ GK & G & 1 \end{bmatrix} \begin{bmatrix} r \\ d \\ n \end{bmatrix}. \tag{6.52}$$

We abbreviate this to $X = \Phi U$.

Proof The proof consists of calculating the inverse of the matrix above. □

Definition 6.43 (Internal Stability of SFL) The SFL system is internally stable if all the transfer functions in the matrix Φ are stable, so

$$\frac{1}{1+GK}, \frac{G}{1+GK}, \frac{GK}{1+GK}, \frac{K}{1+GK}, \tag{6.53}$$

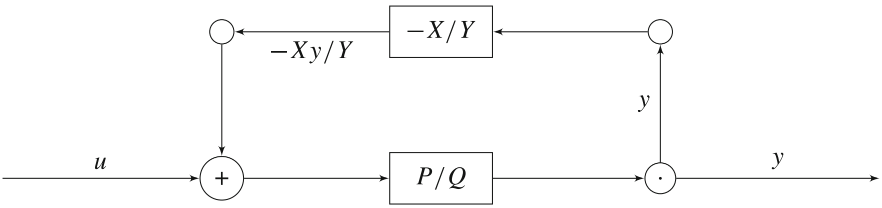
all belong to $\mathcal{S}$.

This is equivalent to having the all entries of the matrix

$$\frac{1}{1+GK}\begin{bmatrix} GK & K \\ G & 1 \end{bmatrix} \tag{6.54}$$

in $\mathcal{S}$. This matrix does not have any particular physical interpretation, but its entries are all stable if and only if the entries of Φ are all stable, where Φ is physically meaningful. The idea is that we want all the junctions in the diagram of the SFL to be stable. (Think of the x, y, z as representing the temperature of the components of a tumble drier; we need to ensure that these do not overheat.)

Consider a SISO system (A, B, C, D) with transfer function $G(s)$, and consider the corresponding simple feedback loop with rational controller K. We now allow $G(s)$ to be an arbitrary rational function, and write it in the form $G = P/Q$ with coprime $P, Q \in \mathcal{S}$.



By previous results, there exists another SISO such that K is the corresponding transfer function, so in this sense we have shown that any SISO with rational transfer function can be stabilized by a rational (possibly unstable) controller. In applications, one may have other criteria in mind when selecting a controller. In applications, one can tune controllers so that they ensure stability, but allow the system to be responsive.

Theorem 6.44 (Youla's Parametrization of Stabilizing Controllers) *Suppose that $G = P/Q$ has a coprime factorization $PX + QY = 1$ where $P, Q, X, Y \in \mathcal{S}$.*

(i) Then $K = X/Y$ internally stabilizes SFL.
(ii) The set of all rational controllers K that internally stabilize SFL is

$$\left\{K = \frac{X+QR}{Y-PR} : R \in \mathcal{S}; Y - PR \neq 0\right\}. \tag{6.55}$$

Proof (i) Suppose that $K = X/Y$. Then,

$$I + GK = 1 + \frac{X}{Y}\frac{P}{Q} = \frac{YQ + XP}{YQ} = \frac{1}{YQ}; \tag{6.56}$$

so $\frac{1}{1+GK} = YQ$ and hence Φ has all entries in $\mathcal{S}$, since

$$\Phi = YQ \begin{bmatrix} 1 & -P/Q & -1 \\ X/Y & 1 & -X/Y \\ PX/YQ & P/Q & 1 \end{bmatrix} = \begin{bmatrix} YQ & -YP & -YQ \\ XQ & YQ & -XQ \\ XP & YP & YQ \end{bmatrix}. \tag{6.57}$$

Suppose that $G = P/Q$ and $K = X/Y$ where $PX + QY = 1$. Then

$$\frac{1}{1+GK}\begin{bmatrix} GK & K \\ G & 1 \end{bmatrix} = \frac{1}{PX + QY}\begin{bmatrix} PX & QX \\ PY & QY \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix}\begin{bmatrix} P & Q \end{bmatrix} \tag{6.58}$$

Hence we have found a controller that internally stabilizes the simple feedback loop system.
(ii) Suppose that $K = X/Y$. Then, as in Lemma 6.42

$$I + GK = 1 + \frac{X}{Y}\frac{P}{Q} = \frac{YQ + XP}{YQ} = \frac{1}{YQ}; \tag{6.59}$$

so $\frac{1}{1+GK} = YQ$ and hence Φ has all entries in $\mathcal{S}$, since

$$\Phi = YQ \begin{bmatrix} 1 & -P/Q & -1 \\ X/Y & 1 & -X/Y \\ PX/YQ & P/Q & 1 \end{bmatrix} = \begin{bmatrix} YQ & -YP & -YQ \\ XQ & YQ & -XQ \\ XP & YP & YQ \end{bmatrix}. \tag{6.60}$$

Hence we have found a controller that internally stabilizes the simple feedback loop system. Suppose that

$$K = \frac{X + QR}{Y - PR} \tag{6.61}$$

Then

$$I + GK = 1 + \frac{X + QR}{Y - PR}\frac{P}{Q} = \frac{YQ + XP}{(Y - PR)Q} = \frac{1}{(Y - PR)Q}; \tag{6.62}$$

so

$$\frac{1}{1 + GK} = (Y - PR)Q \tag{6.63}$$

and hence

$$\Phi = \begin{bmatrix} (Y-PR)Q & -(Y-PR)P & -(Y-PR)Q \\ (X+QR)Q & (Y-PR)Q & -(X+QR)Q \\ (X+QR)P & (Y-PR)P & (Y-PR)Q \end{bmatrix} \tag{6.64}$$

has all its entries in $\mathcal{S}$. Conversely, let K be a rational controller that stabilizes SFL; then

$$\frac{1}{1+KG}\begin{bmatrix} GK & G \\ K & 1 \end{bmatrix} \tag{6.65}$$

is a submatrix of Φ, hence has all its entries in $\mathcal{S}$. Hence

$$R = \begin{bmatrix} X & Y \end{bmatrix} \frac{1}{1+KG}\begin{bmatrix} GK & G \\ K & 1 \end{bmatrix}\begin{bmatrix} Y \\ -X \end{bmatrix} \tag{6.66}$$

belongs to $\mathcal{S}$, and we write this as

$$\begin{aligned} R &= \begin{bmatrix} X & Y \end{bmatrix}\begin{bmatrix} G \\ 1 \end{bmatrix}(1+KG)^{-1}\begin{bmatrix} K & 1 \end{bmatrix}\begin{bmatrix} Y \\ -X \end{bmatrix} \\ &= \frac{(XG+Y)(KY-X)}{1+KG} \\ &= \frac{(XP/Q+Y)(KY-X)}{1+KP/Q} \\ &= \frac{(PX+QY)(KY-X)}{Q+KP} = \frac{KY-X}{KP+Q}. \end{aligned}$$

$$Y-PR = Y - \frac{PKY-PX}{KP+Q} = \frac{QY+PX}{KP+Q} = \frac{1}{KP+Q} \neq 0; \tag{6.67}$$

and we solve for K from

$$(KP+Q)R = KY - X \tag{6.68}$$

to get

$$K = \frac{X+QR}{Y-PR}. \tag{6.69}$$

□

6.12 Invariant Factors

The theory of linear algebra is often formulated for matrices with entries in a field. Much of theory extends to matrices with entries in a principal ideal domain $\mathcal{R}$, so we are able to introduce the machinery necessary to describe MIMO systems in terms of transfer functions involving $\mathbb{C}[s]$ and $\mathbb{C}[1/(1+s)]$.

Definition 6.45 The elementary row operations on a matrix over $\mathcal{R}$ are:

(i) interchanging two rows;
(ii) multiplying one row by a unit in $\mathcal{R}$;
(iii) adding a multiple of one row to another.

For later use, we prove the following basic results.

Lemma 6.46 *Let $\mathcal{R}$ be a principal ideal domain.*

(i) An $n \times n$ matrix X with entries in a principal ideal domain $\mathcal{R}$ is invertible in $M_{n\times n}(\mathcal{R})$ if and only if $\det X$ *is a unit in $\mathcal{R}$.*
(ii) For all $B \in \mathcal{R}^{n\times 1}$, there exists $X \in M_{n\times n}(\mathcal{R})$ with $\det X$ *a unit in $\mathcal{R}$ such that $XB = col[r, 0, \dots, 0]$, where (r) is the ideal generated by the entries of B.*

Proof

(i) Existence follows from

$$X\mathrm{adj}(X) = (\det X)I_n \tag{6.70}$$

and the forward implication follows from identity $\det X \det(X^{-1}) = \det I_n = 1$.

(ii) Suppose that $B = \mathrm{col}[b_1, \dots, b_n]$. First observe that we can permute the entries of B by interchanging pairs of neighbouring entries, as in

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & \vdots \\ 0 & \dots & I \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ B' \end{bmatrix} = \begin{bmatrix} b_2 \\ b_1 \\ B' \end{bmatrix} \tag{6.71}$$

where $B' = \mathrm{col}[b_3, \dots, b_n]$ and the matrix here has $(\det X)^2 = \det X^2 = 1$, so $\det X$ is a unit. Now we prove the statement by induction on n, starting with the crucial case $n = 2$. Let $x, y \in \mathcal{R}$ with $y \neq 0$. Then

$$\begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} p & q \\ -b & a \end{bmatrix} \begin{bmatrix} r \\ 0 \end{bmatrix} \tag{6.72}$$

where the matrix is unimodular in $M_{2\times 2}(\mathcal{R})$ and r is the generator of the ideal $(x, y) = \{wx+zy : w, z \in \mathcal{R}\}$. To see this, we introduce $(r) = (x, y)$, so $r \neq 0$, and we write $y = ar$ and $x = br$ for some $a, b \in \mathcal{R}$. We have $r = py + qx$ for some $p, q \in \mathcal{R}$; then $r = par + qbr$ and $1 = ap + bq$ by cancellation; then

$$\begin{bmatrix} p & q \\ -b & a \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}, \tag{6.73}$$

so we can left multiply by the inverse matrix to solve.

Suppose we have the result for n, and consider $n + 1$. Then $B_{n+1} = \mathrm{col}[b_1, \dots, b_n, b_{n+1}]$ can be written as $B_{n+1} = [B_n; b_{n+1}]$ where $B_n = \mathrm{col}[b_1, \dots, b_n]$, so we have $X_n \in M_{n\times n}(\mathcal{R})$ with $\det X_n$ a unit in $\mathcal{R}$ such that

$$\begin{bmatrix} X_n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} B_n \\ b_{n+1} \end{bmatrix} = \begin{bmatrix} r_n \\ 0 \\ b_{n+1} \end{bmatrix}, \tag{6.74}$$

where (r_n) is the ideal generated by the entries of B_n. We now permute the entries of the final column vector, and apply the case of $n = 2$ to find $r_{n+1} = (r_n, b_{n+1})$.

□

In particular, when $\mathcal{R}$ is a Euclidean domain, we can find the entries of X by following through the steps in this induction proof and using the Euclidean algorithm in the case $n = 2$.

Proposition 6.47 (Unimodular-Upper Triangular Decomposition) *Let $\mathcal{R}$ be a principal ideal domain, and $A \in M_{n\times n}(\mathcal{R})$. Then there exists a upper triangular $T \in M_{n\times n}(\mathcal{R})$ and a unimodular $U \in M_{n\times n}(\mathcal{R})$ such that $A = UT$.*

Proof By the Lemma 6.46, there exists $X_0 \in M_{n\times n}(\mathcal{R})$ with $\det X_0$ a unit in $\mathcal{R}$ such that

$$X_0 A = \begin{bmatrix} r_1 & s_1 \\ 0 & A_1 \end{bmatrix}, \tag{6.75}$$

so we can introduce $X_1 \in M_{(n-1)\times(n-1)}(\mathcal{R})$ with $\det X_1$ a unit in $\mathcal{R}$ such that

$$\begin{bmatrix} 1 & 0 \\ 0 & X_1 \end{bmatrix} XA = \begin{bmatrix} r_1 & \dots & \dots \\ 0 & r_2 & \dots \\ 0 & 0 & A_2 \end{bmatrix} \tag{6.76}$$

and repeat until we have an upper triangular matrix T_0 and $X \in M_{n\times n}(\mathcal{R})$ with $\det X_0$ a unit in $\mathcal{R}$ such that $XA = T_0$; then we let $U_0 = X^{-1}$ so $A = U_0 T_0$. Finally, we multiply the first column of U_0 by $1/\det U_0$ to get a unimodular U, and the first row of T_0 by $\det U_0$ to get an upper triangular T with $A = UT$. □

Let $B = [b_{jk}]$ be an $n \times m$ matrix with entries in $\mathcal{R}$. For $S \subseteq \{1, \dots, n\}$ and $T \subseteq \{1, \dots, m\}$ with $\sharp S = \sharp T = \ell$, let $\det[b_{jk}]_{j\in S, k\in T}$ be the determinant formed from the submatrix of B, known as an ℓ- minor. Then let

$$J_\ell = \left(\det[b_{jk}]_{j\in S,k\in T} : \sharp S = \sharp T = \ell\right) \tag{6.77}$$

be the ideal generated by all of these ℓ-minors. By considering the determinant expansion of each minor in terms of smaller minors, we observe that

$$J_n \subseteq J_{n-1} \subseteq \cdots \subseteq J_1, \tag{6.78}$$

where J_1 is simply the ideal generated by the individual entries of B. Since $\mathcal{R}$ is a principal ideal domain, we have $J_\ell = (\Delta_\ell)$, with $\Delta_1 \mid \Delta_2 \mid \cdots \mid \Delta_n$. Recall that Δ_ℓ is uniquely determined up to multiplication by a unit in $\mathcal{R}$. With $\Delta_n = d_n\Delta_{n-1}$, the additive quotient group J_{n-1}/J_n is naturally isomorphic to the additive quotient group $R/(d_n)$, namely the group with addition modulo (d_n). To see this, use the group homomorphism $\mathcal{R} \to J_{n-1} : p \mapsto p\Delta_{n-1}$ followed by the quotient map $J_{n-1} \to J_{n-1}/J_n$ which is a group homomorphism. Note that $\Delta_n \mid \Delta_{n-1}p$ if and only if $d_n \mid p$.

The divisibility conditions can be made more precise by a theorem which was proved by H.J.S. Smith for $\mathcal{R} = \mathbb{Z}$.

Theorem 6.48 (Invariant Factors) *Let $\mathcal{R}$ be a principal ideal domain and $B \in M_{n\times m}(\mathcal{R})$. Then there exist unimodular matrices $X \in M_{n\times n}(\mathcal{R})$ and $Y \in M_{m\times m}(\mathcal{R})$ and a $n \times m$ matrix D such that $B = XDY$, where*

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & \ddots & d_r & \vdots \\ 0 & \dots & \dots & 0 \end{bmatrix}$$

with $r \leq \min\{m, n\}$ and $d_1 \mid d_2 \mid \cdots \mid d_r$, and $\Delta_u = d_1, d_2 \dots d_u$. The sequence $(d_1, \dots, d_r)$ is called an invariant factor sequence and D is an invariant factor matrix. The factors are unique up to multiplication by units.

Proof The proof is a considerable refinement of the method used to prove unimodular-upper triangular factorization, and is given in detail in [20]. By calculating the Δ_n, one can find the d_k via $\Delta_k = d_1, \dots d_k$, and hence find D. Uniqueness can be proved from determinants identities, as in Exercise 6.18. For small matrices with entries in $\mathbb{C}[s]$ we can use the algorithm of Proposition 6.24 to compute the Δ_k. □

Example 6.49 For $G, K \in R$, the matrix

$$\begin{bmatrix} 1 & -G & -1 \\ K & 1 & -K \\ GK & G & 1 \end{bmatrix} \tag{6.79}$$

has ideals

$$J_3 = ((1+GK)^2), \quad J_2 = (1+GK), \quad J_1 = (1), \tag{6.80}$$

as one easily shows by calculating the minors.

We now revert to the notation used previously.

Proposition 6.50 *Let (A, B, C, D) be a MIMO with $A \in M_{n\times n}(\mathbb{C})$.*

(i) Then the transfer function is

$$T(s) = \frac{P(s)}{\chi_A(s)} \tag{6.81}$$

where $\chi_A(s)$ is the characteristic polynomial of A and the entries of $P(s)$ belongs to the ideal J_{n-1} in $\mathbb{C}[s]$ that is generated by the $n-1$-minors of $sI - A$.

(ii) Suppose that $J_{n-1} = (\Delta_{n-1}(s))$ where the degree of $\Delta_{n-1}(s)$ is positive. Then $\chi_A(s) = d_n(s)\Delta_{n-1}(s)$ and

$$T(s) = D + \frac{R(s)}{d_n(s)} \tag{6.82}$$

where the fraction is a matrix of strictly proper rational functions.

Proof

(i) In the principal ideal domain $\mathbb{C}[s]$ we can write $J_n = (\chi_A(s))$, and J_{n-1} for the ideal generated by the $n-1$-minors of $sI - A$, or equivalently by the entries of $\operatorname{adj}(sI - A)$. Then the result follows from the formula

$$T(s) = \chi_A(s)^{-1}\big(\chi_A(s)D + C\operatorname{adj}(sI - A)B\big). \tag{6.83}$$

(ii) If $J_{n-1} = (\Delta_{n-1}(s))$ where the degree of $\Delta_{n-1}(s)$ is positive, then we can write $\chi_A(s) = d_n(s)\Delta_{n-1}(s)$ and $d_n(s) \in \mathbb{C}[s]$ is a new common denominator for the entries of $T(s)$ of lower degree less than n. By the Euclidean algorithm, a typical $p(s) \in J_{n-1}$ has the form $p(s) = \chi_A(q)q(s) + \Delta_{n-1}(s)r(s)$ where $q(s) \in \mathbb{C}[s]$ and $r(s) \in \mathbb{C}[s]$ has either $r(s) = 0$ or the degree of $r(s)$ is less than the degree of $d_n(s)$. We apply this to the entries $p(s)$ of $P(s)$ in (i).

This applies in particular to cases in which A has Jordan block decomposition as in

$$A = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}, \tag{6.84}$$

and $J_2 = (s-\lambda)$ with $J_3 = ((s-\lambda)^3)$, so $d_3(s) = (s-\lambda)^2$. □

6.13 Matrix Factorizations to Stabilize MIMO

Let $\mathcal{R}$ be an integral domain. The matrix identity

$$\begin{bmatrix} x & -y \\ -q & p \end{bmatrix} \begin{bmatrix} p & y \\ q & x \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.85}$$

is equivalent to $px - qy = 1$ for $x, y, p, q \in \mathcal{R}$. So given $p, q \in \mathcal{R}$ we can look for $x, y \in \mathcal{R}$ to make the matrices on the left-hand side invertible. Equivalently, given $p, q \in \mathcal{R}$ we can look for $x, y \in \mathcal{R}$ to make the determinants of matrices on the left-hand side be units in $\mathcal{R}$. If $p \neq 0$, we can then consider the element q/p as a fraction in its lower terms in the quotient field over $\mathcal{R}$. Note that this involves an additional assumption on p, since the matrix identity

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.86}$$

does not lead to such a fraction.

We now consider coprime factorization in $M_{n\times n}(\mathcal{R})$ for $n > 1$ which is not commutative and has zero divisors. We need to respect left and right factors.

Definition 6.51

(i) Given $P, Q \in M_{n\times n}(\mathcal{R})$, we say that $[P; Q]$ are right coprime if there exist $X, Y \in M_{n\times n}(\mathcal{R})$ such that $XP - YQ = I_n$; so that

$$\begin{bmatrix} X & -Y \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} = I_n. \tag{6.87}$$

(ii) Given $W, Z \in M_{n\times n}(\mathcal{R})$, we say that $[W, Z]$ are left coprime if there exist $R, S \in M_{n\times n}(\mathcal{R})$ such that $-WR + ZS = I_n$; so that

$$\begin{bmatrix} -W & Z \end{bmatrix} \begin{bmatrix} R \\ S \end{bmatrix} = I_n. \tag{6.88}$$

Note that $[P; Q]$ are right coprime if and only if $[P^\top, Q^\top]$ are left coprime, as we see by taking the transpose of $XP - YQ = I_n$.

Lemma 6.52 *The following data are equivalent in $M_{n\times n}(\mathcal{R})$:*

(i) a right coprime $[P; Q]$, and a left coprime $[-W, Z]$ such that $-WP + ZQ = 0$;
(ii) given $P, Q, W, Z \in M_{n\times n}(\mathcal{R})$ such that the block matrix identity

$$\begin{bmatrix} X & -Y \\ -W & Z \end{bmatrix} \begin{bmatrix} P & R \\ Q & S \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix} \tag{6.89}$$

holds for some $X, Y, R, S \in M_{n\times n}(\mathcal{R})$.

Proof Given (ii), we immediately have (i) by considering the block diagonal and bottom left entries. Conversely, given (i) we have $X, Y, R, S \in M_{n\times n}(\mathcal{R})$ such that

$$\begin{bmatrix} X & -Y \\ -W & Z \end{bmatrix} \begin{bmatrix} P & R \\ Q & S \end{bmatrix} = \begin{bmatrix} I_n & E \\ 0 & I_n \end{bmatrix} \tag{6.90}$$

where $E = XR - YS$. To remove this term, we postmultiply by $\begin{bmatrix} I_n & -E \\ 0 & I_n \end{bmatrix}$ to obtain

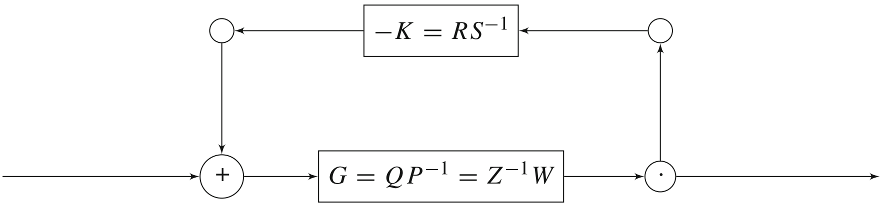
$$\begin{bmatrix} X & -Y \\ -W & Z \end{bmatrix} \begin{bmatrix} P & R - PE \\ Q & S - QE \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix} \tag{6.91}$$

which gives a matrix factorization with $P, Q, -W, Z$ in the required positions.

Suppose that we have a factorization as in the Lemma 6.52 for $\mathcal{R} = \mathbb{C}[1/(1+s)]$, and suppose further that P and Z are invertible; then

$$G = QP^{-1} = Z^{-1}W \tag{6.92}$$

is an $n \times n$ matrix with entries in the quotient field over $\mathcal{R}$. □



Proposition 6.53 (Stabilizing MIMOs) *Given the data as in the Lemma 6.52, suppose further that $X, P \in M_{n\times n}(\mathcal{R})$ are invertible in $M_{n\times n}(\mathbb{C}(s))$. Then*

(i) $G = QP^{-1} = Z^{-1}W$ has a doubly coprime factorization;
(ii) The simple feedback loop with plant G is stabilized with controller $K = -RS^{-1}$, so that

$$\begin{aligned}
(I + GK)^{-1}G &= SW \in M_{n\times n}(\mathcal{S}),\\
(I + GK)^{-1} &= SZ \in M_{n\times n}(\mathcal{S}),\\
(I + KG)^{-1}K &= -PY \in M_{n\times n}(\mathcal{S}),\\
(I + KG)^{-1} &= PX \in M_{n\times n}(\mathcal{S}).
\end{aligned}$$

Proof Proposition 3.16 Schur complements, $S - RP^{-1}Q \in M_{n\times n}(\mathbb{C}(s))$ is invertible with inverse $Z \in M_{n\times n}(\mathcal{S})$; likewise $Z - WX^{-1}Y$ is invertible in $M_{n\times n}(\mathbb{C}(s))$ with inverse $S \in M_{n\times n}(\mathcal{S})$. In the simple feedback loop, suppose that

$G = Z^{-1}W$ and $K = -RS^{-1}$. Then we have a sequence of identities:

$$\begin{aligned}(I + GK)y &= Gu \\ (I - Z^{-1}WRS^{-1})y &= Z^{-1}Wu \\ Z^{-1}(ZS - WR)S^{-1}y &= Z^{-1}Wu \\ y &= SWu;\end{aligned}$$

the case of $(I + GK)^{-1}$ is similar. Likewise we use $G = QP^{-1}$ and $K = -X^{-1}Y$, so

$$\begin{aligned}(I + KG)y &= Ku \\ (I - X^{-1}YQP^{-1})y &= -X^{-1}Yu \\ (XP - YQ)P^{-1}y &= -Yu \\ y &= -PYu,\end{aligned}$$

and $(I + KG)^{-1}$ is found similarly. We can carry out all the calculations within the domain $\mathcal{R} \subset \mathcal{S}$, and express the hypothesis as $\det X \neq 0$, and $\det P \neq 0$.

The results also apply to

$$\begin{bmatrix} X & -Y \\ -W & Z \end{bmatrix}, \begin{bmatrix} P & R \\ Q & S \end{bmatrix} \qquad \begin{bmatrix} k \times k & k \times m \\ m \times k & m \times m \end{bmatrix} \tag{6.93}$$

where $G \in M_{m\times k}(\mathbb{C}(s))$ and $K \in M_{k\times m}(\mathbb{C}(s))$ so that GK and KG are defined.

□

We can use these results in cases of interest due to the following theorem.

Theorem 6.54 (Coprime Factorization) *Let $\mathcal{R}$ be a principal ideal domain with field of fractions $Q\mathcal{R}$. Then for all $G \in M_{m\times k}(Q\mathcal{R})$, there exist a left coprime factorization and a right coprime factorization.*

Proof Let $G_{j,\ell} = q_{j,\ell}/p_{j,\ell}$ be a nonzero entry of G, written as a fraction where $p_{j,\ell}$ and $q_{j,\ell}$ are coprime; then let y be the least common multiple of all the $p_{j,\ell}$. Then $P = yI_k \in M_{k\times k}(\mathcal{R})$ and $Q = yG \in M_{m\times k}(\mathcal{R})$ have $G = QP^{-1}$. Then by the unimodular-triangular decomposition result, there exists a unimodular $U \in M_{(k+m)\times(k+m)}(\mathcal{R})$ with block form

$$U = \begin{bmatrix} U_{1,1} & U_{1,2} \\ U_{2,1} & U_{2,2} \end{bmatrix} \tag{6.94}$$

such that

$$\begin{bmatrix} U_{1,1} & U_{1,2} \\ U_{2,1} & U_{2,2} \end{bmatrix} \begin{bmatrix} yI_k \\ Q \end{bmatrix} = \begin{bmatrix} T \\ 0 \end{bmatrix} \tag{6.95}$$

where T is upper triangular. Then we introduce the inverse of U in block form

$$\begin{bmatrix} U_{1,1} & U_{1,2} \\ U_{2,1} & U_{2,2} \end{bmatrix} \begin{bmatrix} S_{1,1} & S_{1,2} \\ S_{2,1} & S_{2,2} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \tag{6.96}$$

where $yI_k = S_{1,1}T$ and $Q = S_{2,1}T$, so $S_{1,1}$ and T are invertible, and $G = Q(yI_k)^{-1} = S_{2,1}TT^{-1}S_{1,1}^{-1} = S_{2,1}S_{1,1}^{-1}$. Also $U_{1,1}S_{1,1} + U_{1,2}S_{2,1} = I_k$, so we have a right coprime factorization of G. □

6.14 Inverse Laplace Transforms of Strictly Proper Rational Functions

In the final three sections of this chapter, we reinterpret the results in terms of the state space functions via the Laplace transform and its inverse. We introduce a complex vector space V of functions $f(t)$ satisfying (E) and a complex vector space of R of functions $F(s)$ that are holomorphic near to ∞. The Laplace transform takes $\mathcal{L} : V \to R$, and we also introduce a inverse Laplace transform via a contour integral such that $J : \mathcal{R} \to V$, such that $\mathcal{L}J = I : R \to R$, as one can verify by computing the formulas. Further, for all $f \in V$, we introduce $g = J\mathcal{L}f - f$ which has $\mathcal{L}g = 0$, so by the Laplace uniqueness theorem 4.11, $g = 0$, hence $J\mathcal{L} = I : V \to V$, so $\mathcal{L}$ has inverse J.

Proposition 6.55 (Laplace Inversion for Rationals) *Let $F(s)$ be a strictly proper rational function with poles at distinct λ_j of order $d_j + 1$ $(j = 1, \ldots, m)$ and let $\sigma > \Re\lambda_j$ for all $j = 1, \ldots, m$. Then $F(s)$ is the Laplace transform of*

$$f(t) = \lim_{R\to\infty} \frac{1}{2\pi i} \int_{\sigma-iR}^{\sigma+iR} e^{st} F(s)ds \qquad (t > 0) \tag{6.97}$$

where

$$f(t) = \sum_{j=1}^{m} Res\{e^{st}F(s); s = \lambda_j\} \tag{6.98}$$

is a complex linear combination of $t^\ell e^{\lambda_j t}$ where $\ell = 0, \ldots, d_j$ for $j = 1, \ldots, m$.

Proof This result is more general but less explicit than Proposition 4.27. To calculate the integral, we need to know the poles and then we can compute the necessary partial fractions by the Euclidean algorithm of Proposition 6.24 . We write $F(s) = p(s)/q(s)$ where the degree of $p(s)$ is less than the degree of $q(s)$, and $q(s) = \prod_{j=1}^{m}(s-\lambda_j)^{d_j+1}$ where λ_j are the poles of multiplicity $d_j + 1$. Then $q_j(s) = \prod_{k=1;\, j\neq k}^{m}(s-\lambda_k)^{d_k+1}$ give an ideal

$$J = \big(q_1(s), \dots, q_m(s)\big) \tag{6.99}$$

in $\mathbb{C}[s]$, so $J = (p)$ for some $p \in \mathbb{C}[s]$ such that p divides q_j for all $j = 1, \dots, m$; since the λ_j are distinct, p is a unit, so $J = (1)$. Hence the algorithm of Proposition 6.24 gives $h_j(s) \in \mathbb{C}[s]$ such that $1 = \sum_{j=1}^{m} h_j(s)q_j(s)$. Now we apply the division algorithm and obtain $g_j(s), r_j(s) \in \mathbb{C}[s]$ such that $p(s)h_j(s) = g_j(s)(s-\lambda_j)^{d_j+1} + r_j(s)$ and degree of $r_j(s)$ is less than $d_j + 1$; hence

$$\begin{aligned} F(s) = \frac{p(s)}{q(s)} &= \sum_{j=1}^{m} \frac{p(s)h_j(s)q_j(s)}{q(s)} \\ &= \sum_{j=1}^{m} g_j(s) + \sum_{j=1}^{n} \frac{r_j(s)}{(s-\lambda_j)^{d_j+1}}. \end{aligned} \tag{6.100}$$

We note that $r_j(s)$ is nonzero since $F(s)$ does have a pole at λ_j, but $F(s) \to 0$ as $s \to \infty$ so $\sum_{j=1}^{m} g_j(s) = 0$. We can introduce constants $a_{j,\ell}$ such that $F(s)$ has partial fractions decomposition

$$F(s) = \sum_{j=1}^{m}\sum_{\ell=0}^{d_j} \frac{a_{j,\ell}}{(s-\lambda_j)^{\ell+1}}. \tag{6.101}$$

We proceed to calculate the contour integral round a semicircular contour Γ_R with centre σ and large radius $R > 0$ that goes into the left half-plane and encircles all the poles; see Fig. 4.2 in Exercise 4.13 for this Bromwich contour. We have

$$\begin{aligned} \frac{1}{2\pi i}\int_{\Gamma_R} e^{st}F(s)ds &= \sum_{j=1}^{m} \mathrm{Res}\big\{e^{st}F(s) : s = \lambda_j\big\} \\ &= \sum_{j=1}^{m}\sum_{\ell=0}^{d_j} \frac{a_{j,\ell}}{\ell!}\frac{d^\ell}{ds^\ell}e^{st}\Big|_{s=\lambda_j} \\ &= \sum_{j=1}^{m}\sum_{\ell=0}^{d_j} \frac{a_{j,\ell}}{\ell!}t^\ell e^{\lambda_j t} \qquad (t > 0) \end{aligned} \tag{6.102}$$

by Cauchy's integral formula. Without changing the value of the integral, we can replace the integral round Γ_R by an (improper) integral up the line from $\sigma - i\infty$

to $\sigma + i\infty$ since the integral round the semicircular arc S_R parametrized by $s = \sigma + Re^{i\theta}$ for $\pi/2 \leq \theta \leq 3\pi/2$ and a typical summand contributes

$$\int_{S_R} \frac{e^{ts}}{(s-\lambda_j)^{\ell+1}} \frac{ds}{2\pi i} = \frac{e^{t\sigma}}{2\pi} \int_{\pi/2}^{3\pi/2} \frac{\exp(tRe^{i\theta}) Re^{i\theta} d\theta}{(\sigma + Re^{i\theta} - \lambda_j)^{\ell+1}} \tag{6.103}$$

which converges to 0 as $R \to \infty$. As in Corollary 4.23, the Laplace transform of $f(t)$ is $F(s)$. If $F(s) = O(1/s^2)$ as $s \to \infty$, then the integral is absolutely convergent. □

Corollary 6.56 (Laplace Transforms and Stable Rational Functions) *The Laplace transform gives a bijection between*

$$span\{t^n e^{\lambda t} : n = 0, 1, \dots; \Re\lambda < 0\} \tag{6.104}$$

and the set of strictly proper complex rational functions that have all their poles in the open left half-plane $\{\lambda : \Re\lambda < 0\}$. *All stable strictly proper rational functions arise as transfer functions of SISOs* $(A, B, C, 0)$ *where* A *has all its eigenvalues* λ_j *in LHP.*

Proof The Laplace transform of $t^n e^{\lambda t}$ is $n!/(s-\lambda)^{n+1}$, which is a strictly proper rational function that has a pole of order $n+1$ at λ in the open left half-plane. Conversely, any strictly proper rational function has a partial fractions decomposition as in the preceding proof. With $\sigma = 0$, Proposition 6.55 gives an inverse formula for the Laplace transform, hence the Laplace transform is bijective between these sets of functions. The final statement is immediate from the realization result in Sect. 2.11. □

Example 6.57 Let (A, B, C, D) be a SISO with stable transfer function $T(s)$. Suppose that $u(t) = \sum_{j=1}^{N} a_j e^{i\omega_j t}$ with $a_j \in \mathbb{C}$ and distinct $\omega_j \in \mathbb{R}$ is chosen as the input so that the output $y(t)$ has Laplace transform

$$Y(s) = T(s) \sum_{j=1}^{N} \frac{a_j}{s - i\omega_j}.$$

By a slight extension of Proposition 4.27, one can show that

$$y(t) = \sum_{j=1}^{N} T(i\omega_j) a_j e^{i\omega_j t} + z(t) \tag{6.105}$$

where $z(t) \to 0$ as $t \to \infty$. Suppose that we wish to pick out the part of the signal that has angular frequency ω_1. Then we choose $T(s)$ so that $T(i\omega_1) = 1$ and $T(i\omega_j) = 0$ for $j = 2, \dots, N$, so $y(t) = a_1 e^{i\omega_1 t} + z(t)$.

6.15 Differential Rings

Proposition 6.58 *The set $\mathcal{S}$ of stable rational functions forms a differential ring, so*

(i) *$\mathcal{S}$ is a ring under pointwise multiplication and addition, so that $F(s), G(s) \in \mathcal{S}$ and $\lambda, \mu \in \mathbb{C}$ imply $F(s)G(s) \in \mathcal{S}$ and $\lambda F(s) + \mu G(s) \in \mathcal{S}$;*

(ii) *multiplication is commutative $F(s)G(s) = G(s)F(s)$, and there is $1 \in \mathcal{S}$;*

(iii) *$F(s)G(s) = 0$ for all $s \in RHP$ implies $F(s) = 0$ or $G(s) = 0$ on RHP, so $\mathcal{S}$ is an integral domain;*

(iv) *the units in $\mathcal{S}$ are P/Q where P and Q are nonzero stable polynomials of equal degree; $\mathcal{S}$ is not a field.*

(v) *Matrix $X \in M_{n\times n}(\mathcal{S})$ has inverse $Y \in M_{n\times n}(\mathcal{S})$ if and only if $\det X$ is a unit in $\mathcal{S}$.*

(vi) *For all $F(s) \in \mathcal{S}$ the derivative dF/ds also belongs to $\mathcal{S}$ and is strictly proper;*

(vii) *for all $a \in RHP$, the shifted function $F(s+a)$ also belongs to $\mathcal{S}$;*

(viii) *Let $G(s) \in \mathcal{S}$ be strictly proper with inverse Laplace transform $g(t)$; then dG/ds is the Laplace transform of $-tg(t)$, and $G(s+a)$ is the Laplace transform of $e^{-at}g(t)$.*

Proof

(i) Multiplication and addition: Given $F_1(s) = G_1(s)/H_1(s)$ and $F_2(s) = G_2(s)/H_2(s)$ with degree$(G_1(s)) \leq$ degree$(H_1(s))$ and degree$(G_2(s)) \leq$ degree$(H_2(s))$ we have

$$F_1(s)F_2(s) = \frac{G_1(s)G_2(s)}{H_1(s)H_2(s)} \tag{6.106}$$

where degree$(G_1(s)G_2(s)) \leq$ degree$(H_1(s)H_2(s))$. Also, the zeros of $H_1(s)H_2(s)$ are either zeros of $H_1(s)$ or zeros of $H_2(s)$, hence are in LHP. By partial fractions, we can write $F \in \mathcal{S}$ as

$$F(s) = Q + \sum_{j=1}^{N} a_j(s-\lambda_j)^{-n_j}, \tag{6.107}$$

where here $Q \in \mathbb{C}$ is a constant since $F \in \mathcal{S}$ is proper, and all the λ_j have $\Re\lambda_j < 0$. So we can take linear combinations of such F, and stay in $\mathcal{S}$. Also

$$F_1(s) + F_2(s) = \frac{G_1(s)H_2(s) + H_1(s)G_2(s)}{H_1(s)H_2(s)} \in \mathcal{S}. \tag{6.108}$$

(ii) Commutativity of multiplication follows from the corresponding property for polynomials;

(iii) likewise.

(iv) Whereas $1/(s+1)$ belongs to $\mathcal{S}$, the inverse $s+1$ is not proper, hence not in $\mathcal{S}$. Note that $P/Q \in \mathcal{S}$ if and only if Q is a nonzero stable polynomial of degree greater than or equal to the degree of P. Hence P/Q and Q/P both belong to $\mathcal{S}$ if and only if P and Q are both nonzero, stable and of equal degree.

(v) For $X, Y \in M_n(\mathcal{S})$ the equation $XY = I_n$ implies $\det X \det Y = 1$, so $\det X$ is a unit in $\mathcal{S}$ and X has inverse $X^{-1} = (\det X)^{-1}\mathrm{adj}(X) = Y$. The converse also holds.

(vi) We can differentiate

$$\frac{dF}{ds} = \sum_{j=1}^{N} \frac{-n_j a_j}{(s-\lambda_j)^{n_j+1}}, \tag{6.109}$$

which is strictly proper and the poles are at λ_j in open left half-plane.

(vii) Note that $\lambda \in LHP$ is a pole of $F(s)$ if and only if $\lambda - a \in LHP$ is a pole of $F(s+a)$. In terms of linear systems, this amounts to replacing (A, B, C, D) by $(A + aI, B, C, D)$.

(viii) The inverse Laplace transform is given by Propositions 6.55, 4.6 and Corollary 6.56.

□

Theorem 6.59 *With the usual multiplication and differentiation, let:*

- *$\mathcal{H}$ be the set of complex functions that are holomorphic near ∞;*
- *$\mathbb{C}(s)_p$ be the set of proper rational functions;*
- *$\mathcal{S}$ be the set of stable rational functions;*
- *$\mathcal{R}$ be the set of proper rational functions with poles only at -1; so*

$$\mathcal{R} \subset \mathcal{S} \subset \mathbb{C}(s)_p \subset \mathcal{H}. \tag{6.110}$$

(i) Let R be one of these. Then R is a differential ring of holomorphic functions and $R_0 = \{F(s) \in R : F(\infty) = 0\}$ is an ideal.

(ii) Under the inverse Laplace transform

$$f(t) = \lim_{r\to\infty} \frac{1}{2\pi i} \int_{\sigma - ir}^{\sigma + ir} e^{st} F(s) ds \qquad (t > 0) \tag{6.111}$$

with large $\sigma > 0$, this R_0 corresponds to a complex vector space V of functions f satisfying the exponential growth condition (E) such that $f, g \in V$ implies $tf(t) \in V$ and $f * g \in V$.

Proof This is similar to Proposition 6.58. When F is holomorphic near infinity and $F(s) \to 0$ as $s \to \infty$, we have $F(s) = O(1/s)$ and the contour integral for the inverse Laplace transform is well defined. Given $f, g \in \mathcal{H}$, there exist neighbourhoods $\{s : |s| > r_1\}$ on which f is holomorphic and $\{s : |s| > r_2\}$ on which g is holomorphic, so both f and g are holomorphic on $\{s : |s| > \max\{r_1, r_2\}\}$.

The particular spaces V are specified in Propositions 4.12, 6.55, Corollary 6.56 and Exercise 4.6. By the remarks preceding Proposition 6.55, the Laplace transform gives a bijection $\mathcal{L} : V \to R$. □

In the case of $\mathcal{H}$, we are still dealing with transcendental objects such as infinite power series. However, in special cases, we can reduce to algebraic functions, as in the following section.

6.16 Bessel Functions of Integral Order

The definitive account the theory of Bessel functions remains [59], a copy of which was chained to a table at the University of Chicago during the construction of the first atomic pile [39]. Here we are concerned with Bessel functions of the first kind of integral order. These have the remarkable property that their Laplace transforms are algebraic functions, which leads to some significant applications, and greatly simplifies the analysis. In this section, we introduce Bessel functions of integral order by one convenient definition, then discuss their properties, and conclude with an application to signal transmission.

Definition 6.60 The Bessel function of the first kind of integer order n may be defined as in [61] page 362 by

$$J_n(x) = \int_{-\pi}^{\pi} e^{ix\sin\theta - in\theta} \frac{d\theta}{2\pi}. \tag{6.112}$$

(i) Note that $J_n(x)$ is bounded and real for all real x. Then the function $f_x(\theta) = e^{ix\sin\theta}$ is continuously differentiable and 2π periodic with n^{the} Fourier coefficient $J_n(x)$, so

$$e^{ix\sin\theta} = \sum_{n=-\infty}^{\infty} J_n(x)e^{in\theta}. \tag{6.113}$$

We have with $\theta = \pi - \phi$ the identities

$$\begin{aligned} J_n(x) &= \int_0^{\pi} \cos(x\sin\theta - n\theta)\frac{d\theta}{\pi} \\ &= \int_0^{\pi} \cos(x\sin\phi + n\phi - n\pi)\frac{d\phi}{\pi} \\ &= (-1)^n \int_0^{\pi} \cos(x\sin\phi + n\phi)\frac{d\phi}{\pi} \\ &= (-1)^n J_{-n}(x). \end{aligned}$$

(ii) We have the identities

$$J_{n+1}(x) - J_{n-1}(x) = -2i \int_{-\pi}^{\pi} e^{ix\sin\theta} \sin\theta e^{-in\theta} \frac{d\theta}{2\pi} = -2\frac{dJ_n}{dx} \tag{6.114}$$

and by integration by parts

$$\begin{aligned} ix\big(J_{n+1}(x) + J_{n-1}(x)\big) &= 2ix \int_{-\pi}^{\pi} \cos\theta e^{ix\sin\theta} e^{-in\theta} \frac{d\theta}{2\pi} \\ &= \Big[\frac{1}{\pi} e^{ix\sin\theta} e^{-in\theta}\Big]_{-\pi}^{\pi} + 2in \int_{-\pi}^{\pi} e^{ix\sin\theta} e^{-in\theta} \frac{d\theta}{2\pi} \\ &= 2inJ_n(x) \end{aligned}$$

between the Bessel functions of various orders.

(iii) By expanding the exponential as a series, we have for $n \geq 0$

$$\begin{aligned} J_n(x) &= \int_{-\pi}^{\pi} \sum_{k=0}^{\infty} \frac{(ix)^k}{k!} \sin^k\theta\, e^{-in\theta} \frac{d\theta}{2\pi} \\ &= \int_{-\pi}^{\pi} \sum_{k=0}^{\infty} \frac{(ix)^k}{k!(2i)^k} (e^{i\theta} - e^{-i\theta})^k e^{-in\theta} \frac{d\theta}{2\pi} \end{aligned}$$

and we can use the binomial theorem to express this as

$$\begin{aligned} J_n(x) &= \int_{-\pi}^{\pi} \sum_{k=0}^{\infty} \frac{(ix)^k}{k!(2i)^k} \sum_{\ell=0}^{k} \binom{k}{\ell} e^{i(k-2\ell-n)\theta} \frac{d\theta}{2\pi} \\ &= \sum_{\ell=0}^{\infty} \frac{(ix)^{n+2\ell}}{(n+2\ell)!(2i)^{n+2\ell}} (-1)^{\ell} \binom{n+2\ell}{\ell} \\ &= \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell} x^{n+2\ell}}{2^{n+2\ell}(n+\ell)!\ell!} \end{aligned} \tag{6.115}$$

It is easy to justify the change in order of the integration and summation, since the series are uniformly convergent.

(iv) By differentiating the power series and comparing coefficients, one checks that $y(x) = J_n(x)$ satisfies Bessel's differential equation

$$x^2 \frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - n^2)y = 0. \tag{6.116}$$

Since $J_n(x) = (-1)^n J_{-n}(x)$, we have found only one independent solution; the other solution is a Bessel function of the second kind, which is unbounded at $t = 0$; see [61] p 370.

(v) The Laplace transform of $J_n(x)$ is

$$
\begin{aligned}
Y_n(s) &= \int_0^\infty e^{-sx} \int_{-\pi}^{\pi} e^{ix\sin\theta - in\theta} \frac{d\theta}{2\pi} dx \\
&= \int_{-\pi}^{\pi} \int_0^\infty e^{ix\sin\theta - in\theta} e^{-sx} dx \frac{d\theta}{2\pi} \\
&= \int_{-\pi}^{\pi} \frac{e^{-in\theta}}{s - i\sin\theta} \frac{d\theta}{2\pi}; \qquad (6.117)
\end{aligned}
$$

so with the substitution $z = e^{i\theta}$ we obtain a contour integral round $C(0, 1)$ with

$$
\begin{aligned}
Y_n(s) &= \frac{-1}{\pi i} \int_{C(0,1)} \frac{dz}{(z^2 - 2sz - 1)z^n} \\
&= \frac{-1}{\pi i} \int_{C(0,1)} \Big(\frac{1}{z - z_-} - \frac{1}{z - z_+} \Big) \frac{dz}{(z_- - z_+)z^n}
\end{aligned}
$$

where $z_\pm = s \pm \sqrt{s^2 + 1}$ are the quadratic roots. For $n \le 0$, there is only a simple pole at $z = z_-$ inside $C(0, 1)$, so by Cauchy's theorem

$$
Y_n(s) = \frac{-1}{\pi i} \frac{2\pi i}{z_-^n} \frac{1}{-2\sqrt{s^2 + 1}} = \frac{1}{(s - \sqrt{s^2 + 1})^n \sqrt{s^2 + 1}}; \qquad (6.118)
$$

whereas for $n > 0$, we have an n^{th} order pole at $z = 0$, so we write

$$
\begin{aligned}
Y_n(s) &= \frac{-1}{\pi i} \int_{C(0,1)} \Big(-\sum_{k=0}^{\infty} \frac{z^k}{z_-^{k+1}} + \sum_{k=0}^{\infty} \frac{z^k}{z_+^{k+1}} \Big) \frac{dz}{(z_- - z_+)z^n} \\
&= \frac{1}{(s + \sqrt{s^2 + 1})^n \sqrt{s^2 + 1}} \\
&= \frac{(\sqrt{s^2 + 1} - s)^n}{\sqrt{s^2 + 1}} \qquad (6.119)
\end{aligned}
$$

by Cauchy's residue theorem. This $Y_n(s)$ is holomorphic near to ∞, on account of (6.117), so in (vi) we clarify the interpretation of the square root.

(vi) Using the binomial expansion, we can define the appropriate square root function by

$$\sqrt{1+s^2} - s = \sum_{k=1}^{\infty} \binom{1/2}{k} \frac{1}{s^{2k-1}} \qquad (|s| > 1) \tag{6.120}$$

where the series converges and determines a holomorphic function near ∞ which vanishes at ∞. Then we can extend the definition of $\sqrt{1+s^2} - s$ to a holomorphic function on $\mathbb{C} \setminus [-i, i]$ so that $\sqrt{1+s^2}$ takes values $\pm\sqrt{1-y^2}$ for $s = iy \pm 0$ on either side of the cut $[-i, i]$. In the Laplace inversion formula, the integrand $Y_n(s)e^{st}$ is holomorphic on $\mathbb{C} \setminus [-i, i]$, so we can replace the the Bromwich contour integral 4.2 by

$$J_n(t) = \int_B \frac{e^{st}(\sqrt{s^2+1} - s)^n}{\sqrt{1+s^2}} \frac{ds}{2\pi i}$$

where B is the dog-bone contour that goes from $-i+\delta$ to $i+\delta$, goes round i on an arc of a circle, then goes down from $i-\delta$ to $-i-\delta$, then goes round $-i$ on a semicircular arc back to $-i+\delta$; see Exercise 4.13 and [61] page 365. By substituting $s = i \sin\theta$, one can check consistency with the above definition of $J_n(t)$.

(vii) We have $J_0(0) = 1$, $J_1(t) = -dJ_0/dt$ and $J_n(0) = 0$ for all $n = 1, 2, \dots,$ so from (ii) there is a recursion formula for the Laplace transforms

$$\begin{bmatrix} Y_{n+1}(s) \\ Y_n(s) \end{bmatrix} = \begin{bmatrix} -2s & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Y_n(s) \\ Y_{n-1}(s) \end{bmatrix}, \qquad \begin{bmatrix} Y_1(s) \\ Y_0(s) \end{bmatrix} = \frac{1}{\sqrt{1+s^2}} \begin{bmatrix} \sqrt{1+s^2} - s \\ 1 \end{bmatrix} \tag{6.121}$$

for $n = 1, 2, \dots,$ which is slightly different from a recursion formula that we will encounter for the Chebyshev polynomials (8.30).

(viii) The Bessel functions $J_{n+1/2}(t)$ are also of interest in signal processing. These may be written as $J_{n+1/2}(t) = P_n(1/\sqrt{t}) \sin t + Q_n(1/\sqrt{t}) \cos t$ for polynomials P_n and Q_n.

Proposition 6.61 (Differential Ring for Bessel Functions Transforms) *Under the usual pointwise operations,*

$$\mathcal{B} = \{f(s) + g(s)\sqrt{1+s^2};\ f(s), g(s) \in \mathbb{C}(s)\} \tag{6.122}$$

gives a differential field of meromorphic functions on $\mathbb{C} \setminus [-i, i]$.

Proof We have $Y_0(s) = \sqrt{1+s^2}/(1+s^2)$ and related formulas for $Y_n(s)$. The polynomial equation $Z^2 = 1 + s^2$ may be viewed as an irreducible equation in $\mathbb{C}(s)[Z]$ for indeterminate Z with coefficients in the field $\mathbb{C}(s)$, and we can create a new field $\mathcal{B} = \mathbb{C}(s)[Z]/(Z^2 - 1 - s^2)$ to solve it. By (6.120) of the Example, the elements of $\mathcal{B}$ are holomorphic functions on $\mathbb{C} \setminus [-i, i]$ apart from possible poles from the rational functions.

(i) Addition is the rule

$$\lambda(f_1(s) + g_1(s)\sqrt{1+s^2}) + \mu(f_2(s) + g_2(s)\sqrt{1+s^2})$$
$$= \lambda f_1(s) + \mu f_2(s) + (\lambda g_1(s) + \mu g_2(s))\sqrt{1+s^2};$$

(ii) multiplication works as

$$(f_1(s) + g_1(s)\sqrt{1+s^2})(f_2(s) + g_2(s)\sqrt{1+s^2})$$
$$= f_1(s) f_2 + g_1(s)g_2(s)(1+s^2) + (f_1(s)g_2(s) + f_2(s)g_1(s))\sqrt{1+s^2};$$

(iii) differentiation is

$$\frac{d}{ds}\big(f(s) + g(s)\sqrt{1+s^2}\big) = \frac{df}{ds} + \frac{dg}{ds}\sqrt{1+s^2} + \frac{sg(s)}{1+s^2}\sqrt{1+s^2},$$

(iv) and since $\sqrt{1+s^2}$ is not a rational function, we can take reciprocals

$$\frac{1}{f(s) + g(s)\sqrt{1+s^2}} = \frac{f(s) - g(s)\sqrt{1+s^2}}{f(s)^2 - g(s)^2(1+s^2)}, \tag{6.123}$$

where the denominator is a nonzero rational function.

□

Example 6.62 (Periodic Signals) Returning to the time domain, we can use Bessel functions to express periodic signals. Suppose that $g : \mathbb{R} \to \mathbb{R}$ is periodic with period p and $\int_0^p g(t)dt = 0$, so $\phi(t) = \int_0^t g(u)du$ is also periodic with period p. Then $f(t) = e^{i\phi(t)}$ is periodic with period p and has a Fourier representation $f(t) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{2\pi int/p}$ where $\hat{f}(n) = \int_0^p e^{i\phi(t)-2\pi int/p}dt/p$.

In particular, the single tone $g(t) = x\cos t$ gives $\phi(t) = x \sin t$ which is periodic in t with period 2π, so the Fourier coefficients are

$$\hat{f}(n) = \int_{-\pi}^{\pi} e^{ix\sin t - int}\frac{dt}{2\pi} = J_n(x), \tag{6.124}$$

where we recognize the Bessel function of integral order n as the n^{th} Fourier coefficient 6.112 , so as in [61] page 358

$$e^{ix\sin t} = \sum_{n=-\infty}^{\infty} J_n(x)e^{int} = J_0(x) + 2\sum_{n=1}^{\infty} J_{2n}(x)\cos 2nt + 2i\sum_{n=1}^{\infty} J_{2n-1}(x)\sin(2n-1)t. \tag{6.125}$$

Note that $\cos 2nt$ and $\sin(2n-1)t$ involve higher harmonics than were present in the original $\sin t$. For x with $|x|$ small, we can approximate $J_n(x)$ by the first few terms in the Maclaurin expansion $J_n(x) = (x/2)^n (n!)^{-1} + \dots$, so

$$e^{ix\sin t} = 1 - \frac{x^2}{4} + \frac{x^2}{4}\cos 2t + i\Big(x - \frac{x^3}{8}\Big)\sin t + \frac{ix^3}{24}\sin 3t + O(x^4) \qquad (6.126)$$

which gives an approximate formula involving the first few harmonics. The main term in (6.126) is in the polynomial ring $\mathbb{C}[\sin t, \cos t, x]$, so is well suited for calculation for reasons discussed in (8.30). One can express the trigonometric functions in terms of $\tau = \tan(t/2)$, so that

$$\cos t = \frac{1-\tau^2}{1+\tau^2}, \quad \sin t = \frac{2\tau}{1+\tau^2}.$$

This is a familiar, though unpopular, device from elementary calculus for expressing the circle as a rational curve. See also [3] page 68.

6.17 Exercises

Exercise 6.1 Let K be the matrix

$$K = \begin{bmatrix} m & n \\ -x & y \end{bmatrix}, \qquad (6.127)$$

where m, n, x, y are all integers.

(i) Write down a formula for the inverse matrix K^{-1}, assuming it exists. By considering $\det K^{-1}$, show that K has an inverse K^{-1} with integer entries, if and only if $my + xn = \pm 1$.
(ii) Show the condition of (i) is equivalent to m and n having highest common factor 1.
(iii) Show conversely that if m and n have highest common factor 1, then one can choose integers x and y such that K as above is invertible and K^{-1} has integer entries.

Exercise 6.2 Let K be the matrix

$$K = \begin{bmatrix} P(s) & Q(s) \\ -X(s) & Y(s) \end{bmatrix}, \qquad (6.128)$$

where $P(s), Q(s), X(s), Y(s)$ are all complex polynomials.

(i) Write down an expression for the inverse matrix K^{-1}. By considering $\det K^{-1}$, show that K has an inverse with polynomial entries, if and only if $P(s)Y(s) + X(s)Q(s) = \kappa$ for some $\kappa \neq 0$ a constant.
(ii) Show that given $P(s)$ and $Q(s)$, there exist $X(s)$ and $Y(s)$ such that $P(s)Y(s) + X(s)Q(s) = \kappa$ for some $\kappa \neq 0$, if and only if $P(s)$ and $Q(s)$ have highest common factor 1.
(iii) Show conversely that if $P(s)$ and $Q(s)$ have no common zeros, then one can choose polynomials $X(s)$ and $Y(s)$ as entries of K such that K is invertible and K^{-1} has polynomial entries.
(iv) Given $P(s) = s^2 + 3s + 2$ and $Q(s) = s^2 + 2s - 3$, find a K as in (i).

Exercise 6.3 An amplifier and its controller have transfer functions

$$G(s) = \frac{\alpha}{1 + \beta s}, \qquad K(s) = b + \frac{c}{s}, \tag{6.129}$$

where α, β, b, c are real constants with $\alpha, \beta \neq 0$.

(i) State conditions under which $G(s)$ is stable.
(ii) Compute the entries of

$$\Psi = \frac{1}{1 + GK} \begin{bmatrix} 1 & G \\ K & GK \end{bmatrix}, \tag{6.130}$$

and state conditions for all the entries to be stable.
(iii) Deduce that for all G there exists K such that Ψ is stable.

Exercise 6.4

(i) Find the zeros of the polynomial

$$p(s) = s^3 + 10s^2 + 16s + 160. \tag{6.131}$$

(ii) Obtain numerical approximations to the zeros of

$$q(s) = s^3 + 11s^2 + 16s + 160, \tag{6.132}$$

$$r(s) = s^3 + 9s^2 + 16s + 160. \tag{6.133}$$

(iii) Discuss which of these polynomials p, q, r is stable.

Exercise 6.5 (Descartes's Rule of Signs) *Let σ be the number of changes in sign in the real sequence $a_0, \dots, a_n$, ignoring 0. Let r be the number of positive roots of*

$$a_0 + a_1 x + \cdots + a_n x^n = 0. \tag{6.134}$$

Then $r \leq \sigma$, and $\sigma - r$ is even.

Deduce the possible value of r for the polynomial equations:

(i) $-2 + 3x + 5x^2 + x^3 = 0$;
(ii) $2 + 3x - 4x^2 + (1/2)x^3 + x^4 - x^5 + 6x^2 - x^7 = 0$.
(iii) Find the roots of

$$-2 + 3x + 5x^2 + x^3 = 0$$

numerically; hence find r.
(iv) Likewise, find the roots of

$$2 + 3x - 4x^2 + (1/2)x^3 + x^4 - x^5 + 6x^6 - x^7 = 0 \tag{6.135}$$

numerically; hence find r.

Exercise 6.6 Show that R is a commutative ring with 1, where

$$R = \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} : a, b \in \mathbb{C} \right\}, \tag{6.136}$$

and find the units in R.

Exercise 6.7

(i) Let $f, g \in \mathbb{C}[s] \setminus \{0\}$. Show that

$$(fg) \subseteq (f) \cap (g) \subseteq (f, g) \subseteq (1), \tag{6.137}$$

and interpret the ideals in terms of the zeros of f and g.
(ii) Show that f has simple zeros if and only if $(f, df/ds) = (1)$.

Exercise 6.8 Express the rational function

$$G(s) = \frac{s^2 + s + 1}{s^2 - 2} \tag{6.138}$$

as the quotient $G = P/Q$ of stable rational functions P and Q that are coprime in $\mathcal{S}$.

Exercise 6.9 Let A be a nonzero $n \times n$ matrix with entries in $\mathbb{C}[s]$, and $d(A)$ the degree of the nonzero polynomial of minimal degree in A.

(i) Show that one can find the greatest common divisor of the entries of A in at most $d(A)(n^2 - 1)$ applications of the Euclidean algorithm.
(ii) Let J_ℓ be the ideal in $\mathbb{C}[s]$ that is generated by the $k \times k$ minors of A. Estimate the number of applications of the Euclidean algorithm needed to find p_k such that $J_k = (p_k)$ for $k = 1, 2, \ldots, n$.

Exercise 6.10

(i) Show that the following are principal ideal domains: (1) $\mathbb{C}$, (2) $\mathbb{C}(s)$, (3) $\mathbb{C}(s)[\lambda]$ the polynomials in λ with coefficients that are rational functions in s.
(ii) Let $T(s)$ be the transfer function of a a MIMO with input space $\mathbb{C}^n$ and output space $\mathbb{C}^n$. Show that $\det(\lambda I_n - T(s))$ belongs to $\mathbb{C}(s)[\lambda]$.

Exercise 6.11 Use Hurwitz's criterion Theorem 6.12 to show that the real polynomial

$$s^4 + As^3 + Bs^2 + Cs + D$$

is stable if and only if

$$A, B, C, D > 0,$$

$$AB - D^2 > 0,$$

$$ABC - A^2 D - C^2 > 0.$$

Exercise 6.12 Maxwell sought necessary and sufficient conditions for a real quintic to be stable. Using Hurwitz's criterion Theorem 6.12, find conditions on the coefficients for all the roots of

$$s^5 + As^4 + Bs^3 + Cs^2 + Ds + E = 0 \tag{6.139}$$

to have negative real parts. Consider the leading minors of

$$\begin{bmatrix} A & C & E & 0 & 0 \\ 1 & B & D & 0 & 0 \\ 0 & A & C & E & 0 \\ 0 & 1 & B & D & 0 \\ 0 & 0 & A & C & E \end{bmatrix}.$$

Exercise 6.13 Consider the matrices

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 4 & 1 & 1 \\ 5 & -2 & 4 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 3 & 0 \\ -1 & 1 & 2 \end{bmatrix} \tag{6.140}$$

and the matrices $sI - A$ and $sI - F$ over $\mathbb{C}[s]$.

(i) Show that the ideals generated by the k-minors of $sI - A$ are $J_1 = (1)$, $J_2 = (1)$ and $J_3 = ((s-2)^3)$.
(ii) Compare these with the ideals generated by the k-minors of $sI - F$.

Exercise 6.14 Let $\mathcal{R}$ be a principal ideal domain , and $A \in M_{n\times m}(\mathcal{R})$, and $B \in M_{m\times n}(\mathcal{R})$ where $m > n$. The Cauchy–Binet formula states that

$$\det(AB) = \sum_S \det A|_{[n]\times S} \det B|_{S\times[n]}, \tag{6.141}$$

where $[n] = \{1, 2, \dots, n\}$, S ranges over all the subsets of $\{1, \dots, m\}$ that have n elements, and $A|_{[n]\times S}$ is the submatrix of $A = [a_{k,\ell}]$ with $(k, \ell) \in [n] \times S$.

Let $X \in M_{n\times n}(\mathcal{R})$ and $Y \in M_{m\times m}(\mathcal{R})$ be unimodular matrices such that $B = XAY$.

(i) Show that the ideal generated by the k-minors of A and AY satisfy $J_k(A) \subseteq J_k(AY)$ for $k = 1, \dots, n$ and $J_k(AY) \subseteq J_k(A)$.
(ii) Deduce that the ideal generated by the k-minors of A and B satisfy $J_k(A) = J_k(B)$ for $k = 1, \dots, n$.
(iii) Show that the invariant factors of A are unique up multiplication by units of $\mathcal{R}$.

Exercise 6.15 Consider Laguerre's differential equation

$$t\frac{d^2L_n(t)}{dt^2} + (1-t)\frac{dL_n(t)}{dt} + nL_n(t) = 0.$$

(i) Show that the Laplace transform of this equation is

$$(s - s^2)\frac{d\hat{L}_n(s)}{ds} + (n + 1 - s)\hat{L}_n(s) = 0,$$

and that

$$\hat{L}_n(s) = \frac{(s-1)^n}{s^{n+1}}$$

gives a strictly proper rational solution of this equation.
(ii) Deduce that

$$L_n(t) = \lim_{R\to\infty} \int_{1-iR}^{1+iR} e^{st}\frac{(s-1)^n}{s^{n+1}}\frac{ds}{2\pi i} = \frac{1}{n!}\frac{d^n}{ds^n}\Big((s-1)^n e^{st}\Big)\Big|_{s=0}$$

is a polynomial of degree n with $L_n(0) = 1$ that satisfies Laguerre's equation.
(iii) By multiplying Laguerre's equation by $e^{-t}L_m(t)$ and integrating by parts, show that

$$\int_0^\infty e^{-t}L_n(t)L_m(t)dt = 0 \qquad (n \neq m).$$

(iv) Find the Laplace transform of $e^{-t}L_n(2t)$.

Exercise 6.16 Let $\mathcal{SP}$ be the space of monic complex stable polynomials in $\mathbb{C}[s]$, and introduce the ring of fractions $\mathcal{S}_\infty = \{f/p : f \in \mathbb{C}[s], p \in \mathcal{SP}\}$.

(i) Show that $\mathcal{S} \subset \mathcal{S}_\infty \subset \mathbb{C}(s)$, so $\mathcal{S}_\infty$ in an integral domain.

(ii) Show that for $s_r \in RHP$, there is a well-defined homomorphism $\mathcal{S}_\infty \to \mathbb{C}$ given by $f/p \mapsto f(s_r)/p(s_r)$ for $p \in \mathcal{SP}$ and $f \in \mathbb{C}[s]$. Deduce that for a finite subset $S = \{s_1, \dots, s_n\}$ with distinct points $s_j \in RHP$, we can introduce $J = \{f \in \mathcal{S}_\infty : f(s_j) = 0, j = 1, \dots, n\}$. This is an ideal, and $J = ((s - s_1)\dots(s - s_n))$.

(iii) Let $\iota : \mathbb{C}[s] \to \mathcal{S}_\infty$ be the natural homomorphism $f \mapsto f/1$, and J be a nonzero ideal in $\mathcal{S}_\infty$. Show that $\iota^{-1}(J) = \{f \in \mathbb{C}[s] : f/1 \in J\}$ is an ideal in $\mathbb{C}[s]$, which is a principal ideal domain, so there exists $f_J \in \mathbb{C}[s]$ such that $(f_J) = \iota^{-1}(J)$ and that $J = (\iota(f_J))$, so J is generated by (the image of) a polynomial. This point is discussed in [29] page 146.

(iv) Deduce that $\mathcal{S}_\infty$ is a principal ideal domain.

(v) Let $P, Q \in \mathcal{S}_\infty$ be nonzero. By considering the ideal (P, Q) in $\mathcal{S}_\infty$, show that either

(1) there exist $X, Y \in \mathcal{S}_\infty$ such that $PX + QY = 1$; or
(2) there exists $s_0 \in RHP$ such that $P(s_0) = Q(s_0) = 0$.

Exercise 6.17 (Finite-Rank Hankel Operators) For $\lambda_j \in LHP$ and $d_j \in \{0, 1, \dots\}$ for $j = 1, \dots, m$, let

$$V = \mathrm{span}\left\{t^n e^{\lambda_j t} : n = 0, 1, \dots, d_j; j = 1, \dots, m\right\} \tag{6.142}$$

which is a subspace of the space that appears in Corollary 6.56. For $\phi \in V$, let

$$\Gamma_\phi f(t) = \int_0^\infty \phi(t + u) f(u) du$$

for bounded continuous functions f. Show that $\Gamma_\phi f \in V$. This Γ_ϕ gives a Hankel integral operator in the time domain.

Chapter 7
Stability and Transfer Functions via Linear Algebra

This chapter considers stability criteria for linear systems that involve linear algebra for a MIMO system (A, B, C, D). As in Chaps. 5 and 6, we are concerned with stability of transfer functions, but this time focus attention on the matrix formulation, especially the main transformation A.

- The aim is to have criteria that are computationally effective for large matrices, and apply to MIMO systems.
- The new tools are linear matrix inequalities Riccati's matrix inequality and Lyapunov's equation.
- We also consider how transfer functions can be added and multiplied by combining linear systems (A, B, C, D).
- Continuing the theme of matrix algebra, the chapter also includes some periodic linear systems and the discrete Fourier transform.

7.1 Lyapunov's Criterion

Theorem 7.1 *Suppose that A is a complex $n \times n$ matrix and that there exists a positive definite matrix K such that $Q = -(A'K + KA)$ is also positive definite. Then all the solutions of*

$$\frac{dX}{dt} = AX \tag{7.1}$$

are bounded on $(0, \infty)$.

Note that a positive definite K satisfies $K = K'$ and $(A'K + KA)') = (KA + A'K)$, so the issue is whether Q satisfies the equivalent conditions (i)–(iii) of Theorem 3.23. One can either (i) try many positive definite K, and test whether

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_7

Q is also positive definite, or (ii) choose a positive definite Q, and try to find K such that $KA + A'K + Q = 0$.

Proof To show that $\langle X(t), X(t)\rangle$ is bounded, the trick is to consider $V(t) = \langle KX(t), X(t)\rangle$. Observe that $V(t) \geq 0$ for all $t \geq 0$, and use the differential equation to find

$$\begin{aligned}\frac{dV}{dt} &= \Big\langle K\frac{dX}{dt}, X(t)\Big\rangle + \Big\langle KX, \frac{dX}{dt}\Big\rangle\\ &= \big\langle KAX(t), X(t)\big\rangle + \big\langle KX(t), AX(t)\big\rangle\\ &= \big\langle KAX(t), X(t)\big\rangle + \big\langle A'KX(t), X(t)\big\rangle\\ &= \big\langle (A'K + KA)X(t), X(t)\big\rangle \leq 0.\end{aligned}$$

Hence $V(t)$ is decreasing on $(0, \infty)$. Since K is positive definite, the eigenvalues of K are $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_n$, where $\kappa_n > 0$; so

$$0 \leq \kappa_n\langle X(t), X(t)\rangle \leq \langle KX(t), X(t)\rangle \leq \langle KX(0), X(0)\rangle,$$

and so $\|X(t)\| \leq (\langle KX_0, X_0\rangle/\kappa_n)^{1/2}$ for all $t \geq 0$. □

7.2 Sylvester's Equation $AY + YB + C = 0$

Given $n \times n$ matrices A, B and C, the problem is to find a $n \times n$ matrix Y such that

$$AY + YB = -C; \tag{7.2}$$

this is called Sylvester's equation see [8] .

Proposition 7.2 *There are three possibilities for Sylvester's equation: either*

(i) there exists a unique solution;
(ii) there exist infinitely many solutions;
(iii) there does not exist any solution.

Proof In terms of the matrix entries, we can write $A = [a_{jk}]$, $B = [b_{jk}]$ and $C = [c_{jk}]$, and then the unknown $Y = [y_{jk}]$ is given by the system

$$\sum_{\ell=1}^{n} a_{j\ell}y_{\ell k} + \sum_{\ell=1}^{n} y_{j\ell}b_{\ell k} = -c_{jk}, \tag{7.3}$$

which is a linear system of n^2 equations in the n^2 unknowns $[y_{jk}]$. So the above possibilities arise from the general theory of linear equations. Gauss–Jordan elimination reduces this system to reduced echelon form and gives the solutions in cases (i) and (ii). In case (iii), the system is inconsistent, so there is no solution. □

Proof *(Another of the Same)* There is an equivalent way of expressing the preceding proof in terms of linear transformations. Fix A and B, regard C as a matrix of parameters and Y as a variable. Then the transformation $T : M_{n\times n} \to M_{n\times n} : Y \mapsto AY + YB$ is linear on the n^2-dimensional vector space $M_{n\times n}$, so either:

(i) T is invertible, and for all C there exists a unique Y such that $T(Y) = -C$; or
(ii) T is not invertible and hence does not have full rank, so $T(Y) = -C$ has no solution for some C. More precisely, the rank of T is r where $0 \le r < n^2$, so there is an r-dimensional subspace $R = \{T(Y) : Y \in M_{n\times n}\}$ such that $T(Y) = -C$ has a solution if and only if $C \in R$. The nullspace $K = \{Z : T(Z) = 0\}$ is subspace of dimension $n^2 - r$, so for $C \in R$, the general solution has the form $Z + Y$ where $Z \in K$ and Y is some solution of $T(Y) = -C$.
(iii) When C is not an element of R, there is no solution.

□

Proposition 7.3 (Sylvester's Criterion) *Given A and B, Sylvester's equation $AY + YB = -C$ has a unique solution Y for all C if and only if A and $-B$ have no common eigenvalues.*

Proof

(i) First suppose that A and $-B$ have no common eigenvalues, so that their characteristic polynomials $\chi_A(\lambda)$ and $\chi_{-B}(\lambda)$ have highest common factor 1; so by Proposition 6.26, there exist polynomials $p(\lambda)$ and $q(\lambda)$ such that

$$p(\lambda)\chi_A(\lambda) + q(\lambda)\chi_{-B}(\lambda) = 1. \tag{7.4}$$

By the Cayley–Hamilton theorem 2.29, $\chi_{-B}(-B) = 0$ and $\chi_A(A) = 0$, so $\chi_{-B}(A)q(A) = I$. Let Y be any solution of $T(Y) = 0$, so $AY = -YB$. Hence $\chi_{-B}(A)Y = Y\chi_{-B}(-B) = 0$, so

$$Y = q(A)\chi_{-B}(A)Y = q(A)Y\chi_{-B}(-B) = 0; \tag{7.5}$$

hence T is one-to-one. By the rank-nullity theorem 2.2, T is also invertible, so for all C, there exists a unique Y such that $T(Y) = -C$, hence $AY+YB = -C$.
(ii) Conversely, suppose that μ is a common eigenvalue of A and $-B$; then

$$0 = \det(\mu I - A) = \det(\mu I - A^\top), \tag{7.6}$$

so μ is also an eigenvalue of $A^\top$; hence there exist nonzero vectors v and w such that $A^\top w = \mu w$ and $Bv = -\mu v$. Choose C to satisfy $-Cv = \bar{w}$, and suppose with a view to obtaining a contradiction that Y satisfies $AY + YB = -C$; then with the bilinear pairing $\langle v, w \rangle = \sum_{j=1}^{n} v_j w_j$, we have

$$\begin{aligned} \big\langle (AY + YB)v, w \big\rangle &= \big\langle -Cv, w \big\rangle \\ \langle Yv, A^\top w \rangle + \langle -\mu Yv, w \rangle &= \langle \bar{w}, w \rangle \\ \langle Yv, \mu w \rangle + \langle -\mu Yv, w \rangle &= \|w\|^2 > 0, \end{aligned} \tag{7.7}$$

a contradiction, since $\langle Yv, \mu w \rangle + \langle -\mu Yv, w \rangle = 0$.

□

Proposition 7.4 (An Integral Solution of Sylvester's Equation) *Suppose that all the eigenvalues of A and B are in the open left half-plane* $\{\lambda \in \mathbb{C} : \Re\lambda < 0\}$. *Then*

$$Y = \int_0^\infty \exp(tA) C \exp(tB)\, dt \tag{7.8}$$

gives the unique solution to

$$AY + YB = -C. \tag{7.9}$$

The formula in this Proposition is often not the most practical way of finding Y; instead one can use the systems of linear equations in (7.3). The reference [5] gives alternative expressions for the solution. One can also use computer packages.

MATLAB takes the standard form of Sylvester's equation to be $AY + YB + C = 0$, consistent with this book, and gives the solution $Y = lyap(A, B, C)$.

Proof First observe that by Lemma 3.6 there exist $M_1, M_2, \delta_1, \delta_2 > 0$ such that $\|\exp(tA)\| \le M_1 e^{-\delta_1 t}$ and $\|\exp(tB)\| \le M_2 e^{-\delta_2 t}$, so the integral is convergent. Also

$$\begin{aligned} AY + YB &= \int_0^\infty \Big(A\exp(tA)C\exp(tB) + \exp(tA)C\exp(tB)B \Big)\, dt \\ &= \int_0^\infty \frac{d}{dt}\Big(\exp(tA)C\exp(tB) \Big) dt \\ &= \Big[\exp(tA)C\exp(tB) \Big]_0^\infty \\ &= -C. \end{aligned}$$

This shows that the map $Y \mapsto AY + YB$ from $M_{n\times n} \to M_{n\times n}$ is surjective, so by the rank-nullity theorem 2.2, the map is also injective. Hence the solution exists and is unique. □

7.3 A Solution of Lyapunov's Equation $AL + LA' + P = 0$

Corollary 7.5 *Suppose that A is a $n \times n$ complex matrix such all its eigenvalues are in the open left half-plane $\{\lambda \in \mathbb{C} : \Re\lambda < 0\}$. Then for all positive definite P, there exists a unique positive definite L such that*

$$AL + LA' = -P. \tag{7.10}$$

Proof From the characteristic equation, it follows that λ is an eigenvalue of A, if and only if $\bar{\lambda}$ is an eigenvalue of A', so we can introduce $M, \delta > 0$ such that $\| \exp(tA)\| \le Me^{-\delta t}$ and $\| \exp(tA')\| \le Me^{-\delta t}$, so the integral

$$L = \int_0^\infty \exp(tA)P \exp(tA')dt \tag{7.11}$$

converges and gives a solution of (7.10) for any matrix P. Hence the map $Y \mapsto AP + PA'$ is surjective, and by the rank plus nullity theorem is also injective, so the solution is unique. In particular, let P be positive definite. Then $\exp(tA)P \exp(tA')$ is positive definite by exercise since

$$\langle \exp(tA)P \exp(tA')Y, Y\rangle = \langle P \exp(tA')Y, \exp(tA')Y\rangle \tag{7.12}$$

which is positive and continuous for $t > 0$ and $Y \neq 0$. Hence L is a positive definite solution, whenever P is positive definite, and is the unique solution, as observed.

□

For a more advanced discussion of this topic, see [42]. We have chosen a slightly different form for Lyapunov's equation in this Corollary than in the proof of Theorem 7.1, so as to conform with the MATLAB convention. MATLAB takes the standard form of Lyapunov's equation to be $AL + LA' + P = 0$ where $A = A'$ for positive definite P so $L = lyap(A, P)$ or equivalently $L = lyap(A, A', P)$. In Theorem 7.1, we used $KA + A'K + Q = 0$, so $K = lyap(A', Q)$.

For K positive definite, let $\langle\langle v, w\rangle\rangle = \langle Kv, w\rangle$, which defines an inner product of $\mathbb{C}^{n\times 1}$. For $A \in M_{n\times n}(\mathbb{C})$ let $A^\diamond = K^{-1}A'K$, so

$$\mathrm{spec}(A^\diamond) = \mathrm{spec}(A') = \{\bar{\lambda} : \lambda \in \mathrm{spec}(A)\}, \tag{7.13}$$

and

$$\langle\langle Av, w\rangle\rangle = \langle KAv, w\rangle = \langle v, A'Kw\rangle = \langle v, KA^\diamond w\rangle = \langle\langle v, A^\diamond w\rangle\rangle, \tag{7.14}$$

so $A^\diamond$ is the adjoint of A with respect to $\langle\langle\cdot, \cdot\rangle\rangle$, and

$$\langle\langle (A + A^\diamond)v, w\rangle\rangle = \langle (KA + KA^\diamond)v, w\rangle = \langle (KA + A'K)v, w\rangle. \tag{7.15}$$

When K and A are as in Corollary 7.5, A is strictly dissipative with respect to $\langle\langle\cdot, \cdot\rangle\rangle$.

7.4 Stable and Dissipative Linear Systems

For the standard norm on $\mathbb{C}^{n\times 1}$, the following sets

$$\{A \text{ negative definite}\} \subset \{A \text{ strictly dissipative}\} \subset \{A \text{ stable}\}. \tag{7.16}$$

have strict containments, but for each A in the largest set, we can change the norm to move into the middle set.

Theorem 7.6 *The following conditions are equivalent for a $n \times n$ complex matrix A.*

(i) A is stable, so all eigenvalues of A are in LHP;
(ii) for all positive definite P, there exists a positive definite K such that $KA + A'K = -P$;
(iii) there exist $\kappa, M > 0$ such that $\|\exp(tA)\| \leq Me^{-\kappa t}$ for all $t > 0$.

Proof $(i) \Rightarrow (iii)$ Use the Jordan decomposition of A, as in Theorem 3.5.

$(iii) \Rightarrow (i)$ Use resolvent formula for exponentials (3.10).

$(ii) \Rightarrow (iii)$ Apply Lyapunov's criterion 7.1 to dissipative $A + \kappa I$ for some $\kappa > 0$.

$(iii) \Rightarrow (ii)$ Corollary 7.5, giving solution of Lyapunov's criterion 7.1.

$(i) \Rightarrow (ii)$ From (i) and Sylvester's criterion 7.3, there exists K such $KA + A'K = -P$, and by uniqueness $K = K'$. Unfortunately, it is not evident that K is positive definite, so we need to proceed by the route $(i) \Rightarrow (iii) \Rightarrow (ii)$.
We remark that Proposition 3.34 (viii) and (ix) show that we can take $M = 1$ if and only if A is dissipative, which is the case in which $K = I$ gives a positive definite $P = (-A - A')$. □

7.5 Almost Stable Linear Systems

In some cases, it is possible to stabilize a system that has a single pole in RHP, by perturbing the main transformation as follows.

Lemma 7.7 *Let (A, B, C, D) be a SISO such that A has an eigenvalue λ_0 with algebraic multiplicity one and corresponding eigenvector V. Then either:*

(i) $CA^kV = 0$ for $k = 0, 1, \ldots$; or
(ii) there exists $\alpha \in \mathbb{C}$ such that the transfer function of $(A - \alpha VC, B, C, D)$ does not have a pole at λ_0.

Proof

(i) If $CV = 0$, then $CA^kV = \lambda_0^k CV = 0$. For $C \in \mathbb{C}^{1\times n}$, we also observe that this condition implies

$$\operatorname{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} < n, \tag{7.17}$$

so the system (A, B, C, D) is not observable.

(ii) Otherwise, we choose α such that $\alpha CV = \lambda_0 + \nu$ for some $\nu \in (0, \infty)$; then VC is a rank-one matrix such that

$$(sI - A)^{-1}(sI - A + \alpha VC) = I + \alpha(sI - A)^{-1}VC = I + \alpha(s - \lambda_0)^{-1}VC, \tag{7.18}$$

so as in Exercise 3.17 the determinants satisfy

$$\begin{aligned}\det(sI - A)^{-1}\det(sI - A + \alpha VC) &= \det\big(I + \alpha(s - \lambda_0)^{-1}VC\big)\\ &= 1 + \alpha(s - \lambda_0)^{-1}\operatorname{trace} VC\\ &= 1 + \alpha(s - \lambda_0)^{-1}\operatorname{trace} CV\\ &= 1 + (s - \lambda_0)^{-1}(\lambda_0 + \nu)\end{aligned}$$

so that

$$\det(sI - A + \alpha VC) = \frac{s + \nu}{s - \lambda_0}\det(sI - A) \tag{7.19}$$

and the simple zero of $\det(sI - A)$ at λ_0 is canceled out.

□

Theorem 7.8 *Suppose that (A, B, C, D) is an observable SISO, where A has distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Let V be an eigenvector of A for eigenvalue λ_1. Then there exists α such that $\Sigma_\alpha = (A - \alpha VC, B, C, D)$ is an observable SISO, with distinct eigenvalues $\hat\lambda_1, \lambda_2, \dots, \lambda_n$, where $\Re\hat\lambda_1 < 0$. The transfer function of Σ_α is*

$$T_\alpha(s) = \frac{(s - \lambda_1)T_0(s) - \alpha DCV}{s - \lambda_1 - \alpha CV}. \tag{7.20}$$

Proof Suppose that (A, B, C, D) is observable. We observe that

$$\begin{aligned}C(A-\alpha VC)W &= CAW - \alpha CVCW\\ C(A-\alpha VC)^2W &= CA^2W - \alpha CVCAW - \alpha CACVW + \alpha^2 CVCVCW\\ &\vdots \quad \vdots\\ C(A-\alpha VC)^{n-1}W &= CA^{n-1}W + \cdots + (-\alpha)^{n-1}(CV)^{n-1}CW.\end{aligned}$$

Suppose that $C(A-\alpha VC)^kW = 0$ for $k = 0, \dots, n-1$. Then $CA^kW = 0$ for $k = 0, \dots, n-1$. To see this, we proceed by recursion from one line to the next. Given that $CA^kW = 0$ for $k = 0, \dots, j-1$, then at line j, we see that $0 = C(A-\alpha VC)^jW = CA^jW$ since the other summands involve factors $CA^\ell W = 0$ for $\ell < j$.

Since (A, B, C, D) is observable, the only solution to $CA^kW = 0$ for $k = 0, \dots, n-1$ is $W = 0$. Hence $(A-\alpha VC, B, C, D)$ is also observable. We have freedom to choose α and $\nu \in (0, \infty)$ as in the Lemma 7.7 so that $\Re\hat{\lambda}_1 < 0$ and the eigenvalues are all distinct.

There are various formulas relating the transfer function T_α of the new system $(A-\alpha VC, B, C, D)$ with the transfer function $T_0(s)$ of the old one (A, B, C, D). For instance, we start with

$$(sI - A + \alpha VC) = (sI - A) + \alpha VC \tag{7.21}$$

and premultiply by $(sI-A)^{-1}$ and postmultiply by $(sI - A + \alpha VC)^{-1}$; this gives

$$(sI-A)^{-1} = (sI-A+\alpha VC)^{-1} + \alpha(sI-A)^{-1}VC(sI-A+\alpha VC)^{-1}, \tag{7.22}$$

hence

$$(sI - A + \alpha VC)^{-1} = \big(I - \alpha(sI-A)^{-1}VC\big)^{-1}(sI-A)^{-1}. \tag{7.23}$$

The matrix $(sI-A)^{-1}VC$ has rank one, so we can compute the middle inverse matrix by a special argument. With $\beta = \alpha(1 - \alpha C(sI-A)^{-1}V)^{-1}$, we find that

$$\big(I + \beta(sI-A)^{-1}VC\big)\big(I - \alpha(sI-A)^{-1}VC\big) = I, \tag{7.24}$$

so

$$(sI - A + \alpha VC)^{-1} = \Big(I + \frac{\alpha(sI-A)^{-1}VC}{1 - \alpha C(sI-A)^{-1}V}\Big)(sI-A)^{-1}. \tag{7.25}$$

hence

$$C(sI - A + \alpha VC)^{-1}B = \Big(1 + \frac{\alpha C(sI - A)^{-1}V}{1 - \alpha C(sI - A)^{-1}V}\Big)C(sI - A)^{-1}B, \qquad (7.26)$$

and the transfer function satisfy

$$T_\alpha(s) = \frac{T_0(s)}{1 - \alpha C(sI - A)^{-1}V} - \frac{\alpha C(sI - A)^{-1}VD}{1 - \alpha C(sI - A)^{-1}V}; \qquad (7.27)$$

this discussion applies to a typical V, and since in our case $AV = \lambda_1 V$ we can simplify the expression to obtain (7.20). □

This result can be applied repeatedly, to remove troublesome eigenvalues one by one, but the main transformation, hence the eigenvectors, change at each step. In Corollary 7.15, we consider an alternative approach based upon linear systems, and in Sect. 9.12 we look again at the determinants.

Example 7.9 Let

$$(A, B, C, D) = \left(\begin{bmatrix} 1 & -3 & -4 \\ 0 & -1 & -1 \\ 0 & 0 & -5 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}, 0\right) \qquad (7.28)$$

eigenvalues $1, -1, -5$ and transfer function

$$T(s) = \frac{15}{8(s+1)} - \frac{13}{4(s-1)} - \frac{5}{8(s+5)}. \qquad (7.29)$$

To remove the unstable pole at 1, we introduce

$$V = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad (7.30)$$

and consider

$$(A - \alpha VC, B, C, D) = \left(\begin{bmatrix} 1-\alpha & -3 & \alpha - 4 \\ 0 & -1 & -1 \\ 0 & 0 & -5 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}, 0\right) \qquad (7.31)$$

with transfer function

$$\frac{15}{4(\alpha - 6)(s+5)} - \frac{15}{4(\alpha - 2)(s+1)} - \frac{2\alpha^2 - 16\alpha + 19}{(s - 1 + \alpha)(\alpha^2 - 8\alpha + 12)}. \qquad (7.32)$$

By computing

$$\det \begin{bmatrix} C \\ C(A - \alpha VC) \\ C(A - \alpha VC)^2 \end{bmatrix} = 15, \tag{7.33}$$

we deduce that these linear systems are observable for all α.

7.6 Simultaneous Diagonalization

Say that self-adjoint matrices L_0 and L_1 are congruent if there exists an invertible S such that $L_1 = S'L_0S$. If we can choose S to be unitary, so that $S'S = I$ and $L_1 = S'L_0S$, then we say that L_1 and L_0 are unitarily equivalent. Given a pair of self-adjoint $n \times n$ matrices K and L, we can reduce K to a diagonal matrix by unitary conjugation, and L to a diagonal matrix by unitary conjugation. If K and L commute, then we can introduce a unitary W such that $W'KW$ and $W'LW$ are both diagonal matrices. This is possible only if K and L commute. The following result is a partial substitute for simultaneous diagonalization.

Proposition 7.10 *Suppose that K is a positive definite $n \times n$ matrix and L is a self-adjoint $n \times n$ matrix. Then there exist real diagonal matrices D_K and D_L such that D_K is unitarily equivalent to K , D_L is congruent to L and*

$$\det(\lambda K - L) = \det(\lambda D_K - D_L) \qquad (\lambda \in \mathbb{C}). \tag{7.34}$$

Proof We introduce a unitary matrix U such that $U'KU = D_K$, where $D_K = \mathrm{diag}(\lambda_j)$ is a diagonal matrix with positive entries λ_j on the leading diagonal, which are given by the eigenvalues of K. Then we introduce $D_K^{1/2} = \mathrm{diag}(\lambda_j^{1/2})$, which satisfies $D_K^{1/2}D_K^{1/2} = D_K$ and has an inverse $D_K^{-1/2} = \mathrm{diag}(\lambda_j^{-1/2})$. Then $D_K^{-1/2}U'LUD_K^{-1/2}$ is self-adjoint, so there exists a unitary matrix V such that $V'D_K^{-1/2}U'LUD_K^{-1/2}V = D_2$ is a real diagonal matrix; we then define $D_L = D_KD_2$. Since diagonal matrices commute, we can write

$$\begin{aligned} L &= UD_K^{1/2}VD_K^{-1}D_LV'D_K^{1/2}U', \\ K &= UD_KU' = UD_K^{1/2}VV'D_K^{1/2}U' = UD_K^{1/2}VD_K^{-1}D_KV'D_K^{1/2}U', \end{aligned}$$

so that

$$\lambda K - L = UD_K^{1/2}VD_K^{-1/2}(\lambda D_K - D_L)D_K^{-1/2}V'D_K^{1/2}U'. \tag{7.35}$$

We observe that $\det U D_K^{1/2} V D_K^{-1/2} = \det U \det V$, and the multiplicative property of determinants also gives

$$\det(\lambda K - L) = \det U \det V \det(\lambda D_K - D_L) \det V' \det U' = \det(\lambda D_K - D_L). \tag{7.36}$$

Observe that if we can choose $V = I$, then L and K commute and L is unitarily equivalent to D_L and we achieve simultaneous diagonalization. □

7.7 A Linear Matrix Inequality

See [23] and [24]. Consider real matrices $(A, B, C, 0)$, and P positive definite. Previously, we considered the equality $PA + A'P + C'C = 0$ for P positive definite. Here we consider the condition

$$PA + A'P + C'C + PBB'P \prec 0 \tag{7.37}$$

in the sense that the matrix on the left-hand side is negative definite. We introduce

$$L = \begin{bmatrix} PA + A'P + C'C & PB \\ B'P & -I \end{bmatrix}. \tag{7.38}$$

and observe that

$$L = \begin{bmatrix} C'C & 0 \\ 0 & -I \end{bmatrix} + \begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} A' & 0 \\ B' & 0 \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix} \tag{7.39}$$

which is an affine linear expression in P.

Proposition 7.11 (Riccati Matrix Inequality) *The matrix L is negative definite if and only if the Schur complement of $-I$ in L is negative definite, that is*

$$L \prec 0 \Leftrightarrow C'C + PA + A'P + PBB'P \prec 0. \tag{7.40}$$

Proof Then by completing the squares, we obtain

$$\begin{aligned} &\left\langle \begin{bmatrix} PA + A'P + C'C & PB \\ B'P & -I \end{bmatrix}\begin{bmatrix} x \\ u \end{bmatrix}, \begin{bmatrix} x \\ u \end{bmatrix} \right\rangle \\ &= \langle (C'C + PA + A'P)x, x\rangle + \langle x, PBu\rangle + \langle B'Px, u\rangle - \langle u, u\rangle \\ &= \langle (C'C + PA + A'P + PBB'P)x, x\rangle - \langle u - B'Px, u - B'Px\rangle. \end{aligned}$$

We maximize this expression by choosing $u = B'Px$, removing the final term and leaving

$$\big\langle (C'C + PA + A'P + PBB'P)x, x\big\rangle. \tag{7.41}$$

Suppose that L is negative definite; then choosing $u = 0$, we deduce that

$$\langle (C'C + PA + A'P + PBB'P)x, x\rangle < 0 \qquad (x \neq 0) \tag{7.42}$$

so $C'C + PA + A'P + PBB'P$ is negative definite. Suppose conversely, that $C'C + PA + A'P + PBB'P$ is negative definite. If $x = 0$ and $u \neq 0$, then $\langle L \begin{bmatrix} x \\ u \end{bmatrix}; \begin{bmatrix} x \\ u \end{bmatrix}\rangle < 0$. Otherwise $x \neq 0$, so $\langle L \begin{bmatrix} x \\ u \end{bmatrix}; \begin{bmatrix} x \\ u \end{bmatrix}\rangle < 0$ by the preceding calculations.

In this case, $(A, B, C, 0)$ has

$$\begin{aligned}\frac{d}{dt}\langle Px, x\rangle + \langle Cx, Cx\rangle - \langle u, u\rangle \\ &= \langle P(Ax + Bu), x\rangle + \langle Px, Ax + Bu\rangle + \langle Cx, Cx\rangle - \langle u, u\rangle \\ &= \big\langle (PA + A'P + C'C)x, x\big\rangle + \langle PBu, x\rangle + \langle B'Px, u\rangle - \langle u, u\rangle \\ &< 0\end{aligned}$$

for all nonzero inputs u. □

7.8 Differential Equations Relating to Sylvester's Equation

There are various matrix differential equations relating to Sylvester's equation. In the following result, we use an R_t that satisfies the differential equation

$$\frac{dR_t}{dt} = AR_t + R_t A \tag{7.43}$$

with initial condition $dR_0/dt = -BC$, so R_0 gives a solution of Sylvester's equation $AR_0 + R_0 A = -BC$.

Proposition 7.12 *Let* $(A, B, C, 0)$ *be a SISO with* A *stable. Let* $\phi(t) = C\exp(tA)B$ *and*

$$R_t = \int_t^\infty \exp(uA)BC\exp(uA)du, \tag{7.44}$$

and let

$$T(t,u) = -C\exp(tA)(I+R_t)^{-1}\exp(uA)B. \tag{7.45}$$

Then there exists $t_0 > 0$ *such that* $T(t,u)$ *satisfies*

$$\phi(t+u) + T(t,u) + \int_t^\infty T(t,v)\phi(v+u)dv = 0 \qquad (t_0 < t < u) \tag{7.46}$$

and

$$T(t,t) = \frac{d}{dt}\log\det(I+R_t) \qquad (t_0 < t). \tag{7.47}$$

Proof There exist $\kappa, M > 0$ such that $\|\exp(tA)\| \le Me^{-\kappa t}$ for all $t > 0$. Hence the integral for R_t converges, and there exists $t_0 > 0$ such that

$$\begin{aligned}\|R_t\| &\le \int_0^\infty \|\exp(uA)\|\|BC\|\|\exp(uA)\|du \le \int_t^\infty M^2\|BC\|e^{-2\kappa u}du\\ &= \frac{M^2\|BC\|e^{-2\kappa t}}{2\kappa},\end{aligned}$$

so there exists t_0 such that $\|R_t\| < 1$ for all $t > t_0$. Then $I + R_t$ has an inverse, and $T(t,u)$ is defined. Then

$$\begin{aligned}\phi(t+u) &+ T(t,u) + \int_t^\infty T(t,v)\phi(v+u)dv\\ &= C\exp(tA)\exp(uA)B - C\exp(tA)(I+R_t)^{-1}\exp(uA)B\\ &\quad - C\exp(tA)(I+R_t)^{-1}\int_t^\infty \exp(vA)BC\exp(vA)\exp(uA)B\,dv\\ &= C\exp(tA)\Big(I - (I+R_t)^{-1} - R_t(I+R_t)^{-1}\Big)\exp(uA)B = 0.\end{aligned}$$

Observe that

$$\frac{dR_t}{dt} = \frac{d}{dt}\int_t^\infty \exp(uA)BC\exp(uA)du = -\exp(tA)BC\exp(tA),$$

by the fundamental theorem of calculus. We write

$$\begin{aligned}T(t,t) &= -\mathrm{trace}\big(C\exp(tA)(I+R_t)^{-1}\exp(tA)B\big)\\ &= -\mathrm{trace}\big(\exp(tA)BC\exp(tA)(I+R_t)^{-1}\big)\end{aligned}$$

$$= \operatorname{trace}\Big(\frac{dR_t}{dt}(I + R_t)^{-1}\Big)$$
$$= \frac{d}{dt}\operatorname{trace}\log(I + R_t)$$
$$= \frac{d}{dt}\log\det(I + R_t).$$

□

Example 7.13 (Duhamel's Formula) Suppose that A and D are stable matrices and consider a block matrix

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

Then by Schur complements (3.34), we have

$$\begin{bmatrix} sI - A & -B \\ 0 & sI - D \end{bmatrix}^{-1} = \begin{bmatrix} (sI - A)^{-1} & (sI - A)^{-1}B(sI - D)^{-1} \\ 0 & (sI - D)^{-1} \end{bmatrix}. \tag{7.48}$$

From this, we deduce that

$$\exp\left(t\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}\right) = \begin{bmatrix} \exp(tA) & \int_0^t \exp((t-\tau)A)B\exp(\tau D)d\tau \\ 0 & \exp(tD) \end{bmatrix}. \tag{7.49}$$

The Laplace transforms of both sides of this equation are equal to the right-hand side of the previous formula (7.48).

Mainly in this book we are interested in systems that are autonomous, so that the coefficients of the differential equations do not depend upon time. However, some of the formulas can be adapted to deal with systems that have specific types of time dependence. Given a block matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

with A and D square, and a bounded input u, the system

$$\frac{dx}{dt} = Ax + B\exp(tD)u$$
$$y = Cx$$
$$x(0) = x_0, \quad y(0) = Cx_0$$

has solution

$$y(t) = C\exp(tA)x_0 + \int_0^t C\exp((t-\tau)A)B\exp(\tau D)u(\tau)d\tau.$$

If A and D are stable, then the output y is bounded.

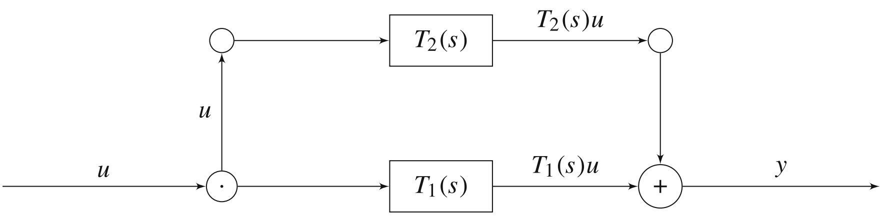
7.9 Transfer Functions tf

Rational linear systems can be described in terms of matrices (A, B, C, D) or transfer functions $T(s)$. The transfer functions have the advantage that there are natural operations of multiplication and addition, by which rational functions form an algebra. The matrix description also has advantages in terms of ease of computations. In this section, we consider various ways of building new transfer functions from old, in terms of (A, B, C, D). The advantage of the following formulas is that they can be carried out in exact arithmetic, where possible. We do not need to solve eigenvalue equations or compute partial fractions, which can involve solving polynomial equations.

Suppose $\Sigma = (A, B, C, D)$ has tf $T(s) = D + C(sI - A)^{-1}B$, and $\phi(t) = D\delta_0 + C\exp(tA)B$. Note that $T(s)$ is the Laplace transform of $\phi(t)$ since

$$\begin{aligned}\mathcal{L}(\phi)(s) &= \int_0^\infty e^{-st}\big(D\delta_0 + C\exp(tA)B\big)dt \\ &= D + C\int_0^\infty \exp(t(A - sI))dtB \\ &= D + C(sI - A)^{-1}B = T(s).\end{aligned}$$

(i) *Adding transfer functions*



The idea is to have devices represented by linear systems Σ_1 and Σ_2 in parallel, combined into a single linear system. Suppose that $\Sigma_1 = (A_1, B_1, C_1, D_1)$ has tf

$$T_1(s) = D_1 + C_1(sI - A_1)^{-1}B_1,$$

and $\phi_1(t) = C_1 \exp(tA_1)B_1$, and $\Sigma_2 = (A_2, B_2, C_2, D_2)$ has tf $T_2(s) = D_2 + C_2(sI - A_2)^{-1}B_2$, and $\phi_2(t) = C_2 \exp(tA_2)B_2$. Then

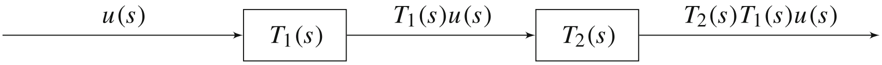
$$\left(\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \begin{bmatrix} C_1 & C_2 \end{bmatrix}, D_1 + D_2 \right) \tag{7.50}$$

has transfer function

$$T(s) = T_1(s) + T_2(s) \tag{7.51}$$

and $\phi(t) = \phi_1(t) + \phi_2(t)$.

(ii) *Multiplying transfer functions*



The idea is to have devices represented by linear systems Σ_1 and Σ_2 in series, combined into a single linear system. In the notation of (i), we write the differential equation for Σ_1 as

$$\begin{aligned} \frac{dx}{dt} &= A_1 x + B_1 u \\ v &= C_1 x + D_1 u \end{aligned}$$

and for input u and state variable x, and use the output v of Σ_1 as the input for Σ_2, which has differential equation

$$\begin{aligned} \frac{d\xi}{dt} &= A_2 \xi + B_2 v \\ y &= C_2 \xi + D_2 v \end{aligned}$$

with state variable ξ and output y. We eliminate v, and use x and ξ for the state variables in the combined differential equation

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} x \\ \xi \end{bmatrix} &= \begin{bmatrix} A_1 & 0 \\ B_2 C_1 & A_2 \end{bmatrix} \begin{bmatrix} x \\ \xi \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 D_1 \end{bmatrix} u \\ y &= \begin{bmatrix} D_2 C_1 & C_2 \end{bmatrix} \begin{bmatrix} x \\ \xi \end{bmatrix} + D_2 D_1 u. \end{aligned}$$

From the differential equations, or by direct verification, we deduce that the linear system

$$\left(\begin{bmatrix} A_1 & 0 \\ B_2C_1 & A_2 \end{bmatrix}, \begin{bmatrix} B_1 \\ B_2D_1 \end{bmatrix}, \begin{bmatrix} D_2C_1 & C_2 \end{bmatrix}, \begin{bmatrix} D_2D_1 \end{bmatrix}\right) \tag{7.52}$$

has transfer function $T_2(s)T_1(s)$.

(iii) *Multiplying transfer functions: an alternative*

This approach uses Sylvester's equation to produce a type of partial fraction decomposition for products of transfer functions. Suppose that we have SISO systems (A_1, B_1, C_1, D_1) has tf

$$T_1(s) = D_1 + C_1(sI - A_1)^{-1}B_1,$$

and (A_2, B_2, C_2, D_2) has tf

$$T_2(s) = D_2 + C_2(sI - A_2)^{-1}B_2,$$

where A_1 and A_2 are $n \times n$ matrices such that $\mathrm{spec}(A_1) \cap \mathrm{spec}(A_2) = \emptyset$. Then there exists X such that $B_1C_2 = A_1X - XA_2$, so $B_1C_2 = -(sI - A_1)X + X(sI - A_2)$; then

$$(sI - A_1)^{-1}B_1C_2(sI - A_2)^{-1} = -X(sI - A_2)^{-1} + (sI - A_1)^{-1}X;$$

hence

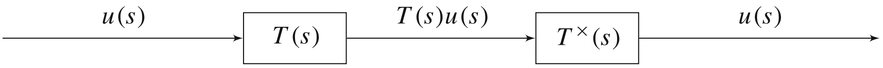
$$\begin{aligned} T_1(s)T_2(s) &= D_1D_2 + C_1(sI - A_1)^{-1}B_1D_2 + D_1C_2(sI - A_2)^{-1}B_2 \\ &\quad + C_1(sI - A_1)^{-1}B_1C_2(sI - A_2)^{-1}B_2 \\ &= D_1D_2 + C_1(sI - A_1)^{-1}B_1D_2 + D_1C_2(sI - A_2)^{-1}B_2 - C_1X(sI - A_2)^{-1}B_2 \\ &\quad + C_1(sI - A_1)^{-1}XB_2 \end{aligned}$$

which is the transfer function of

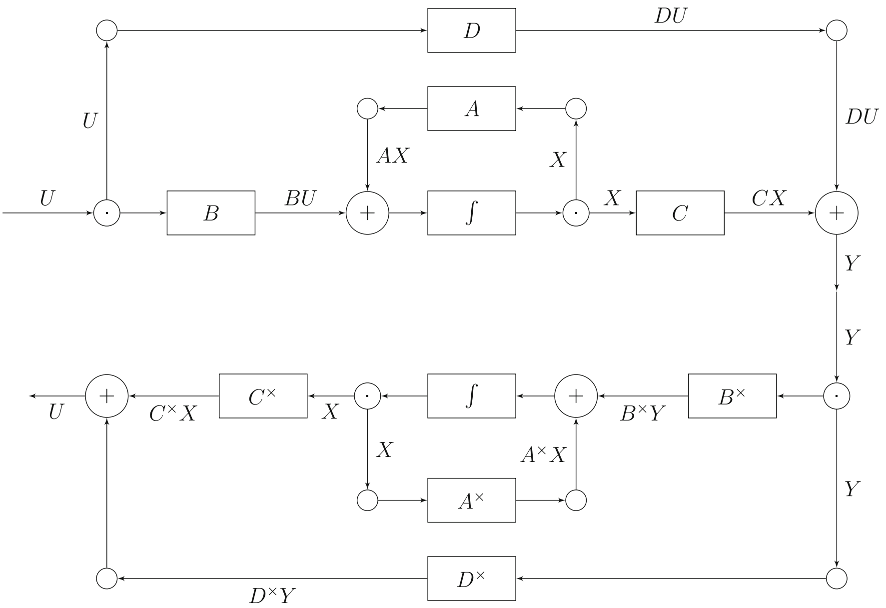
$$\left(\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} B_1D_2 + XB_2 \\ B_2 \end{bmatrix}, \begin{bmatrix} C_1 & D_1C_2 - C_1X \end{bmatrix}, D_1D_2\right). \tag{7.53}$$

More generally, we can use similar formulas to multiply transfer functions of MIMOs of appropriate shapes whenever A_1, B_1 and B_1C_2 are $n \times n$ matrices such that $B_1C_2 = A_1X - XA_2$ has a solution X.

(iv) *Inverting transfer functions*



A radio station takes a message U and uses a rational transfer function $T(s)$ to convert the message into a signal Y which it broadcasts. The receiver wishes to take the signal Y and recover U. Presumably, the receiver should use a transfer function such as $T(s)^{-1}$. Here is how to realize this inverse as a linear system.



Proposition 7.14 (Inverse System) *Suppose that* $\Sigma = (A, B, C, D)$ *has* D *invertible. Then*

$$\Sigma^\times = \begin{bmatrix} A^\times & B^\times \\ C^\times & D^\times \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C & BD^{-1} \\ -D^{-1}C & D^{-1} \end{bmatrix}$$

has transfer function $T^\times(s) = D^\times + C^\times(sI - A^\times)^{-1}B^\times$ *such that*

$$T(s)T^\times(s) = T^\times(s)T(s) = I.$$

Proof To see this, we take the usual differential equation

$$\begin{aligned} \frac{dX}{dt} &= AX + BU \\ Y &= CX + DU \end{aligned} \tag{7.54}$$

with input U and output Y, and then solve for U so that the input becomes U and the output becomes Y. This gives

$$U = -D^{-1}CX + D^{-1}Y \tag{7.55}$$

$$\frac{dX}{dt} = AX + B\big(-D^{-1}CX + D^{-1}Y\big), \tag{7.56}$$

which we can rearrange to give

$$\frac{dX}{dt} = (A - BD^{-1}C)X + BD^{-1}Y \tag{7.57}$$

$$U = -D^{-1}CX + D^{-1}Y, \tag{7.58}$$

so we obtain $\Sigma^{\times}$.

We can also use the Schur complement formula (3.34) to show that for a SISO system (A, B, C, D) with transfer function $T(s)$, the inverse is

$$T(s)^{-1} = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} A - sI & B \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{7.59}$$

□

The purpose of the following result is to replace the main transformation A by $A + BF$ in cases in which $A + BF$ is stable; see Lemma 7.7 and Theorem 7.8. Note that the choice $F = 0$ gives $T_2(s) = I$ and $T(s) = T_1(s)$.

Corollary 7.15 *Suppose that*

$$T(s) = tf \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{7.60}$$

$$T_1(s) = tf \begin{bmatrix} A + BF & B \\ C + DF & D \end{bmatrix}; \quad T_2(s) = tf \begin{bmatrix} A + BF & B \\ F & I \end{bmatrix}. \tag{7.61}$$

Then

$$T(s) = T_1(s)T_2(s)^{-1}. \tag{7.62}$$

Proof In particular, we have

$$\begin{bmatrix} A + BF & B \\ F & I \end{bmatrix}^{\times} = \begin{bmatrix} A & B \\ -F & I \end{bmatrix} \tag{7.63}$$

so

$$T_2(s)T_2(s)^{-1} = \big(I + F(sI - A - BF)^{-1}B\big)\big(I - F(sI - A)^{-1}B\big) = I. \quad (7.64)$$

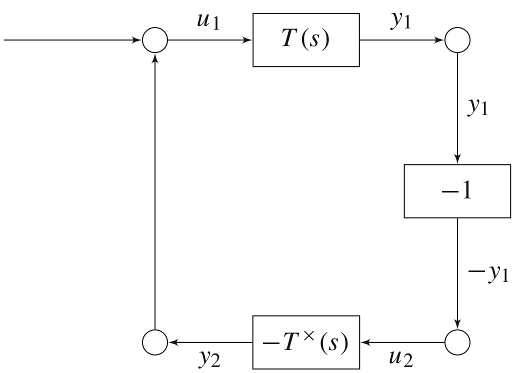
Hence

$$\begin{aligned} T_1(s)T_2(s)^{-1} &= \big(D + (C + DF)(sI - A - BF)^{-1}B\big)\big(I - F(sI - A)^{-1}B\big) \\ &= D + C\big((sI - A - BF)^{-1} - (sI - A - BF)^{-1}BF(sI - A)^{-1}\big)B \\ &\quad + DF\big(- (sI - A)^{-1} + (sI - A - BF)^{-1} \\ &\quad - (sI - A - BF)^{-1}BF(sI - A)^{-1}\big)B \\ &= D + C(sI - A)^{-1}B \\ &= T(s). \end{aligned}$$

□

Example 7.16 (An Input-Output Closed System) Suppose that SISO (A, B, C, D) has transfer function $T(s)$ and $(A^\times, B^\times, C^\times, D^\times)$ has transfer function $T^\times(s)$, where $T(s)T(s)^\times = 1$. Find a SISO that has transfer function $-T^\times(s)$.

Consider the diagram

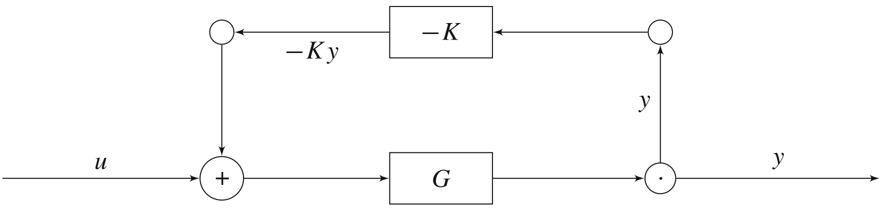


Show that

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

It is instructive to consider the cases in which $T(s)$ and $T^\times(s)$ individually are stable.

(v) *Transfer functions for controllers*



Suppose that we have a plant $G = (A, B, C, D)$ with transfer function

$$G(s) = D + C(sI - A)^{-1}B \tag{7.65}$$

which is to be controlled by another plant $K = (a, b, c, d)$ with transfer function

$$K(s) = d + c(sI - a)^{-1}b \tag{7.66}$$

is a simple feedback loop, so that the combined system has transfer function

$$H(s) = (1 + G(s)K(s))^{-1}G(s). \tag{7.67}$$

We wish to represent this as the transfer function of a single linear system. The differential equation for G is

$$\begin{aligned}\frac{dx}{dt} &= Ax + Bv\\ y &= Cx + Dv\end{aligned}$$

while the differential equation for K is

$$\begin{aligned}\frac{d\xi}{dt} &= a\xi + by\\ w &= c\xi + dy\end{aligned}$$

The combined system is to have input u and output y, and we use x and ξ as the new state variables. Suppose for the moment that $d = 0$. Then the output of K is w, which is multiplied by -1, then added to the input u of the whole system, so that $v = u - w$ is the input into G. We eliminate v and w by writing $w = c\xi$ and the input for the x differential equation becomes

$$v = u - w = u - c\xi, \tag{7.68}$$

and we have an input for the ξ differential equation

$$y = Cx + Dv = Cx + D(u - c\xi),$$

then we have our new differential equations

$$\frac{d}{dt}\begin{bmatrix} x \\ \xi \end{bmatrix} = \begin{bmatrix} A & -Bc \\ bC & a - bDc \end{bmatrix}\begin{bmatrix} x \\ \xi \end{bmatrix} + \begin{bmatrix} B \\ bD \end{bmatrix} u$$
$$y = \begin{bmatrix} C & -Dc \end{bmatrix}\begin{bmatrix} x \\ \xi \end{bmatrix} + Du$$

representing the feedback loop system as in the following block matrix.

$$\left[\begin{array}{cc|c} A & -Bc & B \\ bC & a - bDc & bD \\ -- & -- & -- \\ C & -Dc & D \end{array}\right] \tag{7.69}$$

If $d \neq 0$ and $1 + dD$ is invertible, we use

$$w = c\xi + dy$$

to eliminate w, then substitute this into the equations for v

$$v = u - w = u - c\xi - dy = u - c\xi - d(Cx + Dv)$$

and solve this

$$v = (1 + dD)^{-1}\big(u - c\xi - dCx\big)$$

With $q = (1 + dD)^{-1}$, the differential equations become

$$\frac{d}{dt}\begin{bmatrix} x \\ \xi \end{bmatrix} = \begin{bmatrix} A - qBdC & -qBc \\ bC - qbDdC & a - qbDc \end{bmatrix}\begin{bmatrix} x \\ \xi \end{bmatrix} + \begin{bmatrix} qB \\ qbD \end{bmatrix} u$$
$$y = \begin{bmatrix} C - qDdC & -qDc \end{bmatrix}\begin{bmatrix} x \\ \xi \end{bmatrix} + qDu.$$

Note that matrices denoted with lower case a, b, c, d multiply matrices with upper case A, B, C, D.

(vii) *Conjugating transfer functions*
We consider

$$\Sigma \leftrightarrow \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \Sigma' \leftrightarrow \begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix}$$

and the corresponding transfer functions.

Suppose $\Sigma = (A, B, C, D)$ has tf $T(s) = D + C(sI - A)^{-1}B$, and $\phi(t) = C\exp(tA)B$.
Then $\Sigma' = (A', C', B', D')$ has tf $\tilde{T}(s) = D' + B'(sI - A')^{-1}C'$ and $\tilde{\phi}(t) = B'\exp(tA')C'$,
so $\tilde{T}(s) = T(\bar{s})'$ and $\tilde{\phi}(t) = \phi(t)'$.

Also

$$\Sigma^* = \left(\begin{bmatrix} A & 0 \\ 0 & A' \end{bmatrix}, \begin{bmatrix} 0 & B \\ C' & 0 \end{bmatrix}, \begin{bmatrix} C & 0 \\ 0 & B' \end{bmatrix}, \begin{bmatrix} 0 & D \\ D' & 0 \end{bmatrix} \right) \tag{7.70}$$

has transfer function

$$F(s) = \begin{bmatrix} 0 & T(s) \\ \tilde{T}(s) & 0 \end{bmatrix} \tag{7.71}$$

and

$$\Phi(t) = \begin{bmatrix} C & 0 \\ 0 & B' \end{bmatrix} \exp\left(t \begin{bmatrix} A & 0 \\ 0 & A' \end{bmatrix} \right) \begin{bmatrix} 0 & B \\ C' & 0 \end{bmatrix} = \begin{bmatrix} 0 & \phi(t) \\ \phi(t)' & 0 \end{bmatrix}. \tag{7.72}$$

where $F(\bar{s})' = F(s)$ and $\Phi(t) = \Phi(t)'$ for $t \in (0, \infty)$.

This produces a symmetrical looking transfer function, but the matrices in Σ^* are not themselves self-adjoint.

(vi) *Conjugating transfer functions for square matrices*

There is a variant for I, A, B, C, D all $n \times n$ matrices. Let

$$J = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \tag{7.73}$$

$$\hat{\Sigma} = (\hat{A}, \hat{B}, \hat{C}, \hat{D}) = \left(\begin{bmatrix} 0 & A' \\ A & 0 \end{bmatrix}, \begin{bmatrix} 0 & B' \\ B & 0 \end{bmatrix}, \begin{bmatrix} 0 & C' \\ C & 0 \end{bmatrix}, \begin{bmatrix} 0 & D \\ D' & 0 \end{bmatrix} \right) \tag{7.74}$$

which are all self-adjoint, and consider

$$\begin{aligned} J\frac{d}{dt}X &= \hat{A}X + \hat{B}U \\ Y &= \hat{C}X + \hat{D}U. \end{aligned} \tag{7.75}$$

Then the corresponding transfer function is

$$\hat{T}(s) = \hat{D} + \hat{C}(sJ - \hat{A})^{-1}\hat{B}, \tag{7.76}$$

which is expressed in terms of self-adjoint matrices, and reduces to

$$\hat{T}(s) = \begin{bmatrix} 0 & D + C(sI - A)^{-1}B \\ D' + C'(sI - A')^{-1}B' & 0 \end{bmatrix}; \tag{7.77}$$

which is not self-adjoint.

7.10 Small Groups of Matrices

Example 7.17

(i) The identity matrix and reflection in the origin give a group D_2 with two elements

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{7.78}$$

(ii) There is also a cyclic group C_4 of order four, given by rotations about the origin through $0, \pi/2, \pi, 3\pi/2$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{7.79}$$

all of which have determinant one.

(iii) By adding elements, we can introduce the quaternion group Q_8 with elements

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}, \tag{7.80}$$

which is a subgroup of $SU(2)$.

(iv) Alternatively, we can introduce the dihedral group of order 8, given by the symmetries of the square from the products of elements of D_2 and C_4.

We can choose

$$J = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}, \quad \begin{bmatrix} iI & 0 \\ 0 & -iI \end{bmatrix}, \quad \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & iI \\ iI & 0 \end{bmatrix}, \tag{7.81}$$

and consider the differential equation and transfer function as in (7.75).

7.11 How to Convert Complex Matrices into Real Matrices

In some cases, it is easier to work with matrices with entries that real numbers rather than complex numbers. The following result shows one way of converting real into larger complex matrices. We observe that

$$a + ib \leftrightarrow \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad (a, b \in \mathbb{R})$$

gives a bijective correspondence between the complex number $a + ib$ and the matrix $aI + bJ$, where $J^2 = -I$, so $\{I, J, -I, -J\}$ gives a group isomorphic to $\{1, i, -1, -i\}$, or the group C_4 of rotations through multiples of right angles. We extend this idea as follows. Let $\Re A = (A + \bar{A})/2$ and $\Im A = (A - \bar{A})/(2i)$ be the matrices given by the real and imaginary parts of the entries of a matrix A, so $A = \Re A + i\Im A$.

Lemma 7.18 *There is a homomorphism $M_{n\times n}(\mathbb{C}) \to M_{2n\times 2n}(\mathbb{R})$*

$$A \mapsto \hat{A} = \begin{bmatrix} \Re A & -\Im A \\ \Im A & \Re A \end{bmatrix}$$

which is

(i) *real linear, so $\lambda A \mapsto \lambda\hat{A}$ for all $\lambda \in \mathbb{R}$, and $A \in M_{n\times n}(\mathbb{C})$;*
(ii) *additive $A + B \mapsto \hat{A} + \hat{B}$ for all $A, B \in M_{n\times n}(\mathbb{C})$;*
(iii) *multiplicative, so $AB \mapsto \hat{A}\hat{B}$ for all $A, B \in M_{n\times n}(\mathbb{C})$;*
(iv) *unital, so $I_n \mapsto I_{2n}$;*
(v) *Hermitian matrices $A = A'$ are mapped to real symmetric matrices, so $\hat{A} = \hat{A}^\top$;*
(vi) $\det \hat{A} = |\det A|^2$ *for all $A \in M_n(\mathbb{C})$, so an invertible A is mapped to an invertible $\hat{A}$.*

Proof (i), (ii) and (iv) are straightforward.

(iii) We have

$$AB = (\Re A + i\Im A)(\Re B + i\Im B) = \Re A\Re B - \Im A\Im B + i(\Re A\Im B + \Im A\Re B) \tag{7.82}$$

while for comparison

$$\begin{bmatrix} \Re A & -\Im A \\ \Im A & \Re A \end{bmatrix}\begin{bmatrix} \Re B & -\Im B \\ \Im B & \Re B \end{bmatrix} = \begin{bmatrix} \Re A\Re B - \Im A\Im B & -\Re A\Im B - \Im A\Re B \\ \Re A\Im B + \Im A\Re B & \Re A\Re B - \Im A\Im B \end{bmatrix}. \tag{7.83}$$

(v) For Hermitian A, we have $A = A'$, so $(\Re A + i\Im A) = (\Re A + i\Im A)'$, so $\Re A = (\Re A)^\top$ and $\Im A = -(\Im A)^\top$. Hence $\hat{A}$ is symmetric.

(vi) Working in $M_{2n}(\mathbb{C})$, we have a similarity of matrices

$$\det \hat{A} = \det\left(\begin{bmatrix} I & iI \\ 0 & I\end{bmatrix}\begin{bmatrix} \Re A & -\Im A \\ \Im A & \Re A\end{bmatrix}\begin{bmatrix} I & -iI \\ 0 & I\end{bmatrix}\right) = \det\begin{bmatrix} A & 0 \\ \Im A & A'\end{bmatrix} \tag{7.84}$$

so $\det \hat{A} = \det A \det A' = |\det A|^2$. Also, for an invertible matrix A, we have $\det A \neq 0$, so $\det \hat{A} \neq 0$, and $\hat{A}$ is also invertible. □

Let M be a $n \times n$ complex matrix. Then

$$\hat{M} = \begin{bmatrix} 0 & M \\ M' & 0\end{bmatrix} \tag{7.85}$$

is a self-adjoint matrix, and for $\lambda \in \mathbb{C} \setminus \mathbb{R}$, there is an inverse

$$(\lambda I_{2n} - \hat{M})^{-1} = \begin{bmatrix} \lambda I & -M \\ -M' & \lambda I\end{bmatrix}^{-1} = \begin{bmatrix} \lambda(\lambda^2 I - MM')^{-1} & M(\lambda^2 I - M'M)^{-1} \\ M'(\lambda^2 I - MM')^{-1} & \lambda(\lambda^2 I - M'M)^{-1}\end{bmatrix}. \tag{7.86}$$

Note that $\lambda^2 I - M'M$ and $\lambda^2 I - MM'$ are invertible, since $\lambda^2 > 0$ implies $\lambda \in \mathbb{R}$.

Definition 7.19 Let M be a $n \times n$ complex matrix. Say that $\sigma \geq 0$ is a singular number of M if there exists $v \in \mathbb{C}^{n\times 1}$ such that $v \neq 0$ and $M'Mv = \sigma^2 v$. We can list them according to multiplicity as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

Usually singular numbers of M are defined to be the eigenvalues of $(M'M)^{1/2}$, but it is possible to avoid the complication of square roots, on account of the following lemma.

Lemma 7.20 *Let $\sigma > 0$. Then σ is a singular number of M if and only if there exist eigenvectors of $\hat{M}$ of the form*

$$\begin{bmatrix} w_- \\ v\end{bmatrix} = \begin{bmatrix} -Mv/\sigma \\ v\end{bmatrix}, \quad \begin{bmatrix} w_+ \\ v\end{bmatrix} = \begin{bmatrix} Mv/\sigma \\ v\end{bmatrix}$$

corresponding to $-\sigma$ and σ respectively. Conversely, all the eigenvectors corresponding to nonzero eigenvalue of $\hat{M}$ arise in pairs of this form.

Proof Let $v \neq 0$ satisfy $M'Mv = \sigma^2 v$; then

$$\begin{bmatrix} 0 & M \\ M' & 0\end{bmatrix}\begin{bmatrix} -Mv/\sigma \\ v\end{bmatrix} = -\sigma\begin{bmatrix} -Mv/\sigma \\ v\end{bmatrix}, \tag{7.87}$$

$$\begin{bmatrix} 0 & M \\ M' & 0\end{bmatrix}\begin{bmatrix} Mv/\sigma \\ v\end{bmatrix} = \sigma\begin{bmatrix} Mv/\sigma \\ v\end{bmatrix}, \tag{7.88}$$

and we have an eigenvalue pair for $\hat{M}$ for eigenvalues $\pm\sigma$. Conversely, suppose that we have an eigenvector equation

$$\begin{bmatrix} 0 & M \\ M' & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix} = \lambda \begin{bmatrix} w \\ v \end{bmatrix} \tag{7.89}$$

with $\lambda \neq 0$, so λ is real, and $\lambda^2 > 0$; then $Mv = \lambda w$ and $M'w = \lambda v$; hence $M'Mv = \lambda^2 v$. We take $\sigma = |\lambda| > 0$ and a pair of non zero vector $w_- = -Mv/\sigma$ and $w_+ = Mv/\sigma$.These are distinct, $w = \pm w_\pm$, according to whether $\lambda < 0$ or $\lambda > 0$. This we obtain a pair of eigenvectors corresponding to $\pm\lambda$, where $\sigma = |\lambda|$ is a singular number of M. □

7.12 Periods

For $\omega_j \in \mathbb{R}$ and $a_j \in \mathbb{C}$, the sum

$$f(t) = \sum_{j=1}^{n} a_j e^{i\omega_j t} \tag{7.90}$$

represents a signal with periodic summands $e^{i\omega_j t}$ of various periods. To describe the behaviour of the sum it is helpful to determine the relationship between the periods, as we do here by some algebra. The following results are special cases of the main theorem of [20], which provide an algorithm for computing all the quantities we mention here. Let

$$A = \{k_1\omega_1 + k_2\omega_2 + \cdots + k_n\omega_n : k_1, \ldots, k_n \in \mathbb{Z}\} \tag{7.91}$$

be the additive group that is generated by the ω_j. We note that $\varphi : A \to \{s : |s| = 1\}$

$$\varphi(k_1\omega_1 + k_2\omega_2 + \cdots + k_n\omega_n) = \exp\big(i(k_1\omega_1 + k_2\omega_2 + \cdots + k_n\omega_n)\big) \tag{7.92}$$

is a group homomorphism to the circle, and $\varphi(k_1\omega_1 + k_2\omega_2 + \cdots + k_n\omega_n) = 1$ if and only if $k_1\omega_1 + k_2\omega_2 + \cdots + k_n\omega_n = 2\pi k$ for some $k \in \mathbb{Z}$. We therefore introduce the subgroup $A \cap 2\pi\mathbb{Z}$ and the quotient group $M = A/(A \cap 2\pi\mathbb{Z})$. We can interpret the elements of M as sums $2\pi k + k_1\omega_1 + k_2\omega_2 + \cdots + k_n\omega_n$ modulo $2\pi\mathbb{Z}$.

Then M is a finitely generated Abelian group, and by general theory has a decomposition as a direct sum of nonzero subgroups

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_r \tag{7.93}$$

where $M_j = \{k\sigma_j : k \in \mathbb{Z}\}$ is generated by a single element σ_j and $\theta(M_j) = \{k \in \mathbb{Z} : k\sigma_j = 0\}$ is a proper subgroup of $\mathbb{Z}$, so $\theta(M_j) = (d_j) = d_j\mathbb{Z}$, and M_j is isomorphic as a group to $\mathbb{Z}/(d_j)$.

If $d_j = 0$, then M_j is isomorphic as a group to $\mathbb{Z}$, and we have an infinite group. If $d_j \neq 0$, then we can assume that $d_j > 1$, and M_j is isomorphic as a group to $\mathbb{Z}/(d_j)$, the cyclic group of order d_j.

By the theory, we can take $1 \leq s \leq r$, and arrange the indices so that the positive d_j appear first, with $d_1|d_2|\dots|d_s$, followed by $d_{s+1} = \cdots = d_r = 0$. This gives two possibilities:

Proposition 7.21

(i) Either $r = s$, and M is a finite group such that $d_r m = 0$ for all $m \in M$; this is equivalent to

$$\exp\big(i(k_1\omega_1 + k_2\omega_2 + \cdots + k_n\omega_n)d_r\big) = 1 \qquad (k_1, \dots, k_n \in \mathbb{Z}). \tag{7.94}$$

In particular, $\omega_j = 2\pi q_j/d_r$ for some $q_j \in \mathbb{Z}$ and all $j = 1, \dots, r$ and $2\pi d_r$ is the period of sums such as $f(t)$.

(ii) Alternatively, $s < r$ and M contains an infinite subgroup isomorphic to $\mathbb{Z}^{r-s}$.

Proof See [20] for a general discussion of finitely generated Abelian groups. □

In case (ii), we cannot describe the frequencies simply in terms of fractions with a single common denominator. The sum $f(t)$ is an almost periodic function, as described in Bohr's theory. We refer the reader to [33] and [9] for the general theory and to [45] for application to linear systems. In the next section return to case (i) and consider further the notion of sums of terms $e^{i\omega_j t}$ with a common period.

7.13 Discrete Fourier Transform

Consider a time interval $[0, 2\pi]$ and split this into N equal parts by introducing the times $t_j = 2\pi j/N$ for $j = 0, \dots, N-1$. Given a function $f : [0, 2\pi] \to \mathbb{C}$, we can introduce samples $f(t_j)$. The set of indices $\{0, 1, \dots, N-1\}$ can be regarded as a group under addition modulo N, namely the additive group $\mathbb{Z}/(N)$. Equivalently, we can introduce the multiplicative group

$$G_N = \{1, e^{2\pi i/N}, e^{4\pi i/N}, \dots, e^{2(N-1)\pi i/N}\} \tag{7.95}$$

in which $e^{2\pi ji/N}e^{2\pi ki/N} = e^{2\pi(j+k)i/N}$ and $e^{2\pi ji/N} = e^{2\pi ki/N}$ if and only if $j \cong k$ modulo N. Then $(G_N, \cdot)$ and $(\mathbb{Z}/(N), +)$ give the cyclic group with N elements. It might appear strange to have a cyclical structure for time; however, cyclic patterns are common in music.

Let V be the complex vector space of functions $F : \{0, 1, \dots, N-1\} \to \mathbb{C}$, with the usual pointwise addition and scalar multiplication. We also introduce the scalar product

$$\langle F, G\rangle = \frac{1}{N}\sum_{j=0}^{N-1} F(j)\overline{G(j)} \qquad (F, G \in V). \tag{7.96}$$

Lemma 7.22 *Let $E_k(j) = e^{2\pi jki/N}$ for $j, k \in \{0, 1, \dots, N-1\}$. Then $(E_j)_{j=0}^{N-1}$ gives a complete orthonormal basis for V.*

Proof We have

$$\langle E_k, E_k\rangle = \frac{1}{N}\sum_{j=0}^{N-1} E_k(j)\overline{E_k(j)} = \frac{1}{N}\sum_{j=0}^{N-1} 1 = 1;$$

whereas for $k \neq \ell$, we have $-N < j - \ell < N$, so $e^{2\pi(k-\ell)i/N} \neq 1$ and we can use the geometric sum formula

$$\begin{aligned}\langle E_k, E_\ell\rangle &= \frac{1}{N}\sum_{j=0}^{N-1} e^{2\pi(k-\ell)ji/N}\\ &= \frac{1}{N}\sum_{j=0}^{N-1}\left(e^{2\pi(k-\ell)i/N}\right)^j\\ &= \frac{1}{N}\frac{1-e^{2\pi(k-\ell)i}}{1-e^{2\pi(k-\ell)i/N}}\\ &= 0.\end{aligned}$$

The space V evidently has dimension N since we can specify any $F \in V$ by its values at N points; so we have a complete orthonormal basis. □

Proposition 7.23 *For all $F \in V$, there is an orthogonal expansion*

$$F = \sum_{k=0}^{N-1} a_k E_k \tag{7.97}$$

where $a_k = \langle F, E_k\rangle$ and

$$\langle F, F\rangle = \sum_{k=0}^{N-1} |a_k|^2. \tag{7.98}$$

Proof This is an immediate consequence of the Lemma 7.22 and basic facts about orthogonal bases. □

The a_k are known as discrete Fourier coefficients and the sequence $(a_k)_{k=0}^{N-1}$ as the discrete Fourier transform. Note also that $E_k : \{0, 1, \dots, N-1\} \to \mathbb{C}$ satisfies $E_k(m+n) = E_k(m)E_k(n)$ and $|E_k(n)| = 1$, hence E_k is a multiplicative character on $\{0, 1, \dots, N-1\} = \mathbb{Z}/(N)$. We also have $E_{j+\ell}(n) = E_j(n)E_\ell(n)$.

Using $\{0, 1, \dots, N-1\}$ to index the rows and columns, we introduce the matrix

$$U = \frac{1}{\sqrt{N}}\Big[e^{2\pi jki/N}\Big]_{j,k=0}^{N-1}. \tag{7.99}$$

Corollary 7.24 *The Fourier expansion of $F \in V$ is*

$$F = \sqrt{N}\,U \begin{bmatrix} a_0 \\ \vdots \\ a_{N-1} \end{bmatrix} \tag{7.100}$$

where U has the properties:

(i) $U'U = I$;
(ii) U^2 *is a permutation matrix on the basis* $\{E_j : j = 0, 1, \dots, N-1\}$*;*
(iii) $U^4 = I$.

Proof

(i) We observe that $E_k(j) = e^{2\pi jki/N}$, so writing the E_k as columns, we have

$$U = \frac{1}{\sqrt{N}}\big[E_0\ E_1\ \dots\ E_{N-1}\big], \tag{7.101}$$

where the columns are orthonormal in V by the Lemma 7.22; equivalently, $U'U = [\langle E_j, E_k\rangle]$ is the identity matrix. By the Proposition 7.23, $F = \sum_{k=0}^{N-1} a_k E_k$.

(ii) The (j, ℓ) entry of U^2 is

$$[U^2]_{j,\ell} = \frac{1}{N}\sum_{k=0}^{N-1} e^{2\pi i(j+\ell)k/N} = \langle E_{j+\ell}, E_0\rangle, \tag{7.102}$$

so we apply the Lemma 7.22 . In the case $k = \ell = 0$, we have $\langle E_{j+\ell}, E_0\rangle = 1$; in the case $j = N - \ell$, we have $j + \ell \cong 0$ modulo (N), and we have $\langle E_{j+N-j}, E_0\rangle = \langle E_0, E_0\rangle = 1$; in all other cases, $j + \ell$ is not congruent to 0 modulo (N), so $\langle E_{j+\ell}, E_0\rangle = 0$. Hence U^2 has the form

$$U^2 = \begin{bmatrix} 1 & 0 & \ldots\ldots & 0 \\ 0 & \ldots & 0 & 1 \\ \vdots & & 1 & 0 \\ \vdots & . & & \vdots \\ 0 & 1 & \ldots & 0 \end{bmatrix}. \tag{7.103}$$

The effect of U^2 is to fix E_0, and take E_j to E_{N-j} for $j = 1, \ldots, N-1$.

(iii) Given the shape of E^2, it is clear that $U^4 = I$.

□

7.14 Exercises

Exercise 7.1 For

$$A = B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{7.104}$$

let $T : M_{2\times2}(\mathbb{C}) \to M_{2\times2}(\mathbb{C})$ be the operator

$$T(X) = AX + XB \qquad (X \in M_{2\times2}(\mathbb{C})). \tag{7.105}$$

(i) Show that T is linear.
(ii) Find $\mathrm{null}(T) = \{X : T(X) = 0\}$ and $\mathrm{range}(T) = \{T(X) : X \in M_{2\times2}(\mathbb{C})\}$.

Exercise 7.2 Consider the matrix

$$A = -\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 1 & 2 & 7 \end{bmatrix}. \tag{7.106}$$

(i) Show that $-A - A'$ is not positive definite, by considering the determinant or otherwise.
(ii) Show that there exists a positive definite K such that

$$-AK - KA' = I \tag{7.107}$$

has a solution, and find K numerically. (*Use appropriate computer programs.*)

Exercise 7.3 Let $S : \mathbb{C} \to M_{2\times 2}(\mathbb{C})$ be a matrix function such that

$$S(k)S(-k) = I_2, \quad S(k)^\top = S(k), \quad S(-\bar{k}) = \overline{S(k)} \qquad (k \in \mathbb{C}),$$

where the last matrix has entries that are the complex conjugates of the entries of $S(k)$.

(i) Show that for k real, $S(k)$ is unitary.
(ii) For

$$S(k) = \begin{bmatrix} r(k) & t(k) \\ t(r) & r(k) \end{bmatrix}, \quad \text{let} \quad \Phi(k) = \frac{1}{t(k)t(-k)} \begin{bmatrix} -r(k)t(-k) & t(-k) \\ t(k) & -r(-k)t(k) \end{bmatrix}.$$

Show that trace $\Phi(k) = 0$ and det $\Phi(k) = -1$ for all $k \in \mathbb{C}$.
(iii) For

$$S(k) = e^{i\theta(k)} \begin{bmatrix} \cos\psi(k) & i\sin\psi(k) \\ i\sin\psi(k) & \cos\psi(k) \end{bmatrix},$$

find conditions on $\theta, \psi : \mathbb{C} \to \mathbb{C}$ that ensure that S satisfies the stipulated conditions, and compute $\Phi(k)$.

In scattering theory, $S(k)$ is known as the scattering matrix, while $\Phi(k)$ is the transfer matrix.

Exercise 7.4 (Gramians in continuous time)

(i) Suppose that the controllability Gramian

$$K_C = \int_0^\infty \exp(tA)BB' \exp(tA')dt$$

converges. Show that

$$G(t) = \int_t^\infty \exp(uA)BB' \exp(uA')du$$

gives a solution of

$$\frac{dG}{dt} = AG + GA', \quad G(0) = K_C.$$

(ii) Suppose that the observability Gramian in continuous time

$$K_O = \int_0^\infty \exp(tA')C'C \exp(tA)dt$$

converges. Show that

$$H(t) = \int_t^\infty \exp(uA')C'C\exp(uA)du$$

gives a solution of

$$\frac{dH}{dt} = A'H + HA, \quad H(0) = K_O.$$

Exercise 7.5 By considering the trace, show that the equation

$$AX - XA = I_n$$

has no solution with $A, X \in M_{n\times n}(\mathbb{C})$.

Exercise 7.6 (Pauli Matrices)

(i) Show that

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \tag{7.108}$$

give a linear basis for $M_{2\times 2}(\mathbb{C})$.

(ii) Let $[A, X] = AX - XA$. Show that

$$[\sigma_0, \sigma_j] = 0; [\sigma_1, \sigma_2] = 2\sigma_3, [\sigma_1, \sigma_3] = -2\sigma_2, [\sigma_2, \sigma_3] = -2\sigma_1.$$

(iii) For $A = \sum_{j=0}^3 a_j\sigma_j$, $C = \sum_{j=0}^3 c_j\sigma_j$ and $X = \sum_{j=0}^3 x_j\sigma_j$, deduce that the equation $AX - XA + C = 0$ has a solution if and only if $c_0 = 0$ and $-a_1c_1 + a_2c_2 + a_3c_3 = 0$. Find this solution.

Chapter 8
Discrete Time Systems

- This chapter considers linear systems in discrete time which are specified by a difference equation. Initially, the results are similar to those achieved in previous chapters for continuous time linear systems, and involve tools such as the z-transform which is analogous to the Laplace transform of previous chapters. There is a corresponding notion of transfer function.
- There is a particularly important difference equation called the three term recurrence relation for orthogonal polynomials. This provides us with a route into the classical theory of orthogonal polynomials on bounded intervals of the real line. Orthogonal polynomials are important in signal processing as they can be used to construct filters. The exercises cover examples such as Bessel filters.
- We consider some classical examples of orthogonal polynomials such as the Chebyshev polynomials of the first kind, the Laguerre polynomials and the Hermite polynomials. In Chap. 9, we will use the Chebyshev polynomials and variants to solve some random linear systems and models from physics. In Chap. 10, we use the Laguerre polynomials and their Laplace transforms to study signals in wireless communication.

8.1 Discrete-Time Linear Systems

In this chapter we consider time as a variable which takes values $0, 1, 2, \dots,$ as if viewing the system at unit time intervals. By rescaling, one can adjust the time interval to be $h > 0$. The system under consideration has inputs $u_0, u_1, \dots, \in \mathbb{C}$, and outputs $y_0, y_1, \dots \in \mathbb{C}$, and the corresponding states are $x_0, x_1, \dots \in \mathbb{C}^{N\times 1}$. Given constant matrices

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad \begin{bmatrix} N\times N & N\times 1 \\ 1\times N & 1\times 1 \end{bmatrix}, \tag{8.1}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_8

the corresponding linear system is the system of difference equations

$$\begin{aligned} x_{n+1} &= Ax_n + Bu_n \\ y_n &= Cx_n + Du_n \qquad (n = 0, 1, 2, \ldots). \end{aligned}$$

A convenient way to represent the inputs is by way of a power series

$$(u_n)_{n=0}^{\infty} \leftrightarrow u(z) = \sum_{n=0}^{\infty} u_n z^n; \tag{8.2}$$

if the power series has radius of convergence $R > 0$, then $u(z)$ represents the Taylor series of a holomorphic function on $\mathbb{D}(0, R) = \{z \in \mathbb{C} : |z| < R\}$. Likewise we introduce formal power series

$$(x_n)_{n=0}^{\infty} \leftrightarrow X(z) = \sum_{n=0}^{\infty} x_n z^n, \qquad (y_n)_{n=0}^{\infty} \leftrightarrow Y(z) = \sum_{n=0}^{\infty} y_n z^n \tag{8.3}$$

to represent the state and the output, and interpret them as holomorphic functions when the series converge.

Definition 8.1 (Z-Transform) The function $X(1/z) = \sum_{n=0}^{\infty} x_n z^{-n}$ is known as the unilateral Z-transform of $(x_n)_{n=0}^{\infty}$, and may be regarded as a discrete-time Laplace transform. This is a Laurent series in negative powers of z, so the natural domain of convergence is $\{z \in \mathbb{C} : |z| > r\}$ for some $r > 0$.

8.2 Transfer Function for a Discrete Time Linear System

Definition 8.2 The transfer function of a discrete-time linear system is

$$T(z) = D + zC(I - zA)^{-1}B. \tag{8.4}$$

Proposition 8.3

(i) Then $T(z)$ defines a holomorphic function on $\mathbb{D}(0, 1/\|A\|)$.
(ii) Let $r = \min\{R, 1/\|A\|\}$. Then for $x_0 = 0$, there exists a unique solution to the linear system, which is determined by the coefficients in the power series, where

$$Y(z) = T(z)X(z) \qquad (z \in \mathbb{D}(0, r)). \tag{8.5}$$

Proof

(i) As in Proposition 2.48, $(I - zA)^{-1}$ has rational entries with possible poles at the zeros of $\det(I - zA)$. Hence by Proposition 3.11, $I - zA$ is invertible for $|z|\|A\| < 1$, and $(I - zA)^{-1}$ generates a convergent power series $\sum_{j=0}^{\infty} z^j A^j$, so we have an unambiguous interpretation of $T(z)$ as a holomorphic function via the convergent power series

$$T(z) = D + \sum_{n=0}^{\infty} CA^n Bz^{n+1}. \tag{8.6}$$

(ii) We multiply the state difference equation by z^{n+1} and sum over n to obtain

$$\sum_{n=0}^{\infty} x_{n+1}z^{n+1} = z\sum_{n=0}^{\infty} Ax_n z^n + zB\sum_{n=0}^{\infty} u_n z^n \tag{8.7}$$

so

$$X(z) - x_0 = zAX(z) + zBu(z), \tag{8.8}$$

where by assumption $x_0 = 0$. Now $I - zA$ is invertible for $|z|\|A\| < 1$, so we obtain

$$X(z) = z(I - zA)^{-1}Bu(z), \tag{8.9}$$

where the right-hand side is holomorphic on $\mathbb{D}(0, r)$ since $u(z)$ and $(I - zD)^{-1}$ may be expressed as convergent power series. Multiplying the output equation by z^n and summing over n, we obtain

$$Y(z) = CX(z) + Du(z), \tag{8.10}$$

hence we obtain the solution

$$Y(z) = zC(I - zA)^{-1}Bu(z) + Du(z) \qquad (z \in \mathbb{D}(0, r)), \tag{8.11}$$

which immediately gives $Y(z) = T(z)u(z)$.

Given the convergent power series $X(z)$ and $Y(z)$ we can recover the coefficients from Taylor's formula

$$x_n = \frac{1}{n!}\frac{d^n X}{dz^n}(0), \quad y_n = \frac{1}{n!}\frac{d^n Y}{dz^n}(0) \qquad (n = 0, 1, 2, \dots) \tag{8.12}$$

so we have a unique solution for the system of difference equations. □

Example 8.4 Suppose that $A \in M_{2\times 2}(\mathbb{C})$. Then the characteristic equation of A is the quadratic

$$\det(sI - A) = s^2 - \operatorname{trace} As + \det A$$

where the eigenvalues λ and μ of A satisfy $\lambda + \mu = \operatorname{trace} A$ and $\lambda\mu = \det A$; so either:

(i) $A = \lambda I_2$ for some $\lambda \in \mathbb{C}$;

(ii) A is similar to

$$\begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix} \tag{8.13}$$

for some distinct $\lambda, \mu \in \mathbb{C}$; or

(iii) A is similar to the Jordan block

$$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \tag{8.14}$$

for some $\lambda \in \mathbb{C}$.

Based upon this, one can easily compute A^n and hence the coefficients of the power series $T(z)$. For instance, in case (iii), we have

$$T(z) = D + \frac{z\hat{C}}{(1-\lambda z)^2}\begin{bmatrix} 1-\lambda z & z \\ 0 & 1-\lambda z \end{bmatrix}\hat{B} \tag{8.15}$$

for vectors $\hat{C}$ and $\hat{B}$, so $T(z)$ has a possible double pole.

This Example arises in applications such as (8.30), (6.121) and Proposition 8.27.

Remark 8.5

(i) Proposition 8.3 has a converse Proposition 10.29 which realizes a holomorphic function on the disc as the transfer function of a discrete-time linear system.

(ii) The results in this chapter have been formulated so far for SISO systems. The extension to MIMO systems is straightforward and only involves allowing matrices (A, B, C, D) with suitable shapes. We carry this out in the remainder of this section. The reader can check that Proposition 8.3 extends as required.

8.3 Correspondence Between Continuous- and Discrete-Time Systems

First we recapitulate some previous concepts. We consider the matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad \begin{bmatrix} n \times n & n \times k \\ m \times n & m \times k \end{bmatrix}. \tag{8.16}$$

and observe that the sequence $(CA^jB)_{j=0}^{\infty}$ arises in the following situations:

(i) The continuous time linear system (A, B, C, D) has a transfer function $T(s)$ with a Laurent series

$$T(s) = D + C(sI - A)^{-1}B = D + \sum_{j=0}^{\infty} \frac{CA^jB}{s^{j+1}}; \tag{8.17}$$

(ii) the scattering function of (A, B, C, D) has a Taylor series

$$\phi(t) = D + C\exp(tA)B = D + \sum_{j=0}^{\infty} \frac{CA^jBt^j}{j!}; \tag{8.18}$$

(iii) the discrete time linear system (A, B, C, D) has a transfer function with Taylor series

$$T_d(z) = D + Cz(I - zA)^{-1}B = D + \sum_{j=0}^{\infty} z^{j+1}CA^jB; \tag{8.19}$$

(iv) the operators L and K of Sect. 3.11 have product

$$LK = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \left[B, AB, A^2B, \dots\right] = \begin{bmatrix} CB & CAB & CA^2B & \dots \\ CAB & CA^2B & CA^3B & \dots \\ CA^2B & CA^3B & CA^4B & \dots \\ \vdots & \ddots & \ddots & \dots \end{bmatrix}. \tag{8.20}$$

(v) If A is similar to a diagonal matrix with eigenvalues λ_ℓ, so $A = SDS^{-1}$, then $CA^jB = CSD^jS^{-1}B$ involves the powers λ_ℓ^j.

These statements indicate that the continuous time and discrete time systems are related via the sequence $(CA^jB)_{j=0}^{\infty}$. In the remainder of this section, we consider a more profound connection between the transfer functions.

Example 8.6 (Matrix Möbius Transforms) Suppose that A_d, B_d, C_d and D_d be complex matrices such that

$$\begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix} \qquad \begin{bmatrix} n \times n & n \times k \\ k \times n & k \times k \end{bmatrix} \tag{8.21}$$

is unitary. This relates to Exercise 3.9. Then

$$\Sigma = \begin{bmatrix} -C_d & D_d \\ -A_d & B_d \end{bmatrix} \qquad \begin{bmatrix} k \times n & k \times k \\ n \times n & n \times k \end{bmatrix} \tag{8.22}$$

is also unitary. This matrix is associated with a map

$$\Psi_\Sigma : M_{n\times n}(\mathbb{C}) \to M_{k\times k}(\mathbb{C}) : Z \mapsto D_d + C_d Z(I - A_d Z)^{-1} B_d \tag{8.23}$$

known as the (matrix) Möbius transform, which is discussed in [65, page 146]. By calculation, one shows that

$$I_n - \Psi_\Sigma(Z)'\Psi_\Sigma(Z) = B_d'(I_n - Z'A_d')^{-1}(I_n - Z'Z)(I_n - A_d Z)^{-1} B_d, \tag{8.24}$$

which has the consequence that

$$I_n - Z'Z \succ 0 \Rightarrow I_n - \Psi_\Sigma(Z)'\Psi_\Sigma(Z) \geq 0. \tag{8.25}$$

In particular, we can take $Z = zI_n$ and recover the transfer function

$$\Psi_\Sigma(zI_n) = D_d + C_d z(I - zA_d)^{-1} B_d = T_d(z). \tag{8.26}$$

Proposition 8.7 *Suppose that Σ is unitary. Then $T_d(z)$ is holomorphic on $\mathbb{D}$, with*

(i) $\|T_d(z)\| \leq 1$ for all $z \in \mathbb{D}$;
(ii) $T_d(z)$ is a unitary $k \times k$ matrix for all z such that $|z| = 1$.

Proof

(i) This follows since

$$\begin{aligned} I - T_d(z)'T_d(z) &= I - (D_d' + \bar{z}B_d'(I - \bar{z}A_d')^{-1}C_d')(D_d + zC_d(I - zA_d)^{-1}B_d) \\ &= I - D_d'D_d - \bar{z}B_d'(I - \bar{z}A_d')^{-1}C_d'D_d - zD_d'C_d(I - zA_d)^{-1}B_d \\ &\quad - \bar{z}zB_d'(I - \bar{z}A_d')^{-1}C_d'C_d(I - zA_d)^{-1}B_d \\ &= B_d'B_d + \bar{z}B_d'(I - \bar{z}A_d')^{-1}A_d'B + zB_d'A_d(I - zA_d)^{-1}B_d \\ &\quad - \bar{z}zB_d'(I - \bar{z}A_d')^{-1}(I - A_d'A_d)(I - zA_d)^{-1}B_d \\ &= B_d'(I - \bar{z}A_d')^{-1}\Big((I - \bar{z}A_d')(I - zA_d) + \bar{z}A_d'(I - zA_d) + z(I - \bar{z}A_d')A_d \end{aligned}$$

$$- |z|^2(I - A_d'A_d)\Big)(I - zA_d)^{-1}B_d$$

$$= (1 - |z|^2)B_d'(I - \bar{z}A_d')^{-1}(I - zA_d)^{-1}B_d$$

$$\geq 0.$$

(ii) Note that $T(z)$ is a $k \times k$ matrix and when $|z| = 1$, we have $T(z)'T(z) = I$, so $T(z)$ is unitary.

□

This Proposition 8.7 gives a way of constructing transfer functions that are bounded and holomorphic on the unit disc. We can deduce a similar result for continuous time linear systems.

Theorem 8.8 (Discrete-Time and Continuous-Time Transfer Functions)

(i) Suppose that Σ has the block form

$$\Sigma = \begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix} \qquad \begin{bmatrix} n \times n & n \times k \\ k \times n & k \times k \end{bmatrix} \tag{8.27}$$

where 1 is not an eigenvalue of A_d. Then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} (A_d + I)(A_d - I)^{-1} & 2(A_d - I)^{-1}B_d \\ C_d(I - A_d)^{-1} & D_d + C_d(I - A_d)^{-1}B_d \end{bmatrix} \tag{8.28}$$

gives a continuous-time linear system with transfer function $T(s) = D + C(sI - A)^{-1}B$.

(ii) If A_d has all its eigenvalues in $\mathbb{D}$, then A has all its eigenvalues in LHP so $T(s)$ is stable.

(iii) If Σ is unitary, then $T(s)$ is holomorphic for $s \in RHP$ and

$$\|T(s)\| \leq 1 \qquad (s \in RHP), \tag{8.29}$$

and $T(s)$ is unitary for all $s = i\omega$ with $\omega \in \mathbb{R}$.

Proof

(i) The transfer function $T_d(z)$ for the discrete-time system Σ is defined as in Sect. 8.2, while the transfer function $T(s)$ for (A, B, C, D) is defined as in Sect. 2.10, and we need to show that these match up. For $s \in RHP$ we write $z = (s - 1)(s + 1)^{-1}$, so $z \in \mathbb{D}$, and we calculate

$$T_d(z) = D_d + zC_d(I - zA_d)^{-1}B_d$$

$$= D_d + (s - 1)C_d\big((s + 1)I - (s - 1)A_d\big)^{-1}B_d$$

$$= D_d + (s-1)C_d\big(s(I-A_d) + I + A_d\big)^{-1}B_d$$
$$= D_d + (s-1)C_d(I-A_d)^{-1}\big(sI - A\big)^{-1}B_d$$

then we use $(s-1)I = sI - A + A - I$ and $A - I = 2(A_d - I)^{-1}$ to write

$$\begin{aligned} T_d(z) &= D_d + C_d(I - A_d)^{-1}B_d + 2C_d(I - A_d)^{-1}(sI - A)^{-1}(A_d - I)^{-1}B_d \\ &= D + C(sI - A)^{-1}B \\ &= T(s). \end{aligned}$$

(ii) We observe that if $\lambda \in \mathbb{D}$, then $(\lambda - 1)/(\lambda + 1) \in RHP$ so $(\lambda + 1)/(\lambda - 1) \in LHP$. This is relevant for the eigenvalues λ of A_d and A.

(iii) Since Σ is unitary, we have $A_d' A_d + C_d' C_d = I$, so $\|A_d\| \leq 1$ and all the eigenvalues of A_d are in $\overline{\mathbb{D}}$. Hence $T_d(z)$ is homomorphic on $\mathbb{D}$, and $T(s) = T_d(z)$ is holomorphic for $s \in RHP$. By the Proposition 8.7, we have $\|T(s)\| = \|T_d(z)\| \leq 1$ for all $s \in RHP$. Also with $e^{i\theta} = (i\omega - 1)/(i\omega + 1)$ we have $T(i\omega)'T(i\omega) = T_d(e^{i\theta})'T_d(e^{i\theta}) = I$.

□

8.4 Chebyshev Polynomials and Filters

Example 8.9 (Chebyshev Polynomials) Consider the linear system

$$A = \begin{bmatrix} 2s & -1 \\ 1 & 0 \end{bmatrix}, \quad B = 0, \quad C = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad D = 0 \tag{8.30}$$

where s is here regarded as a complex parameter, so

$$\begin{aligned} x_{n+1} &= Ax_n \\ y_n &= Cx_n \end{aligned} \tag{8.31}$$

and we choose $x_0 = \begin{bmatrix} s \\ 1 \end{bmatrix}$. Then $y_n = CA^n x_0$ gives the solution, which is well adapted for computer algebra. For instance, one can compute

$$[y_0, y_1, y_2, y_3] = C * [x_0, A * x_0, A * A * x_0, A * A * A * x_0] \tag{8.32}$$

to find $C_0, \dots, C_3$.

(i) Observe that A has characteristic equation $\lambda^2 - 2s\lambda + 1 = 0$ with roots $s \mp \sqrt{s^2 - 1}$ so $s = \cos\theta$ gives eigenvalues $e^{\pm i\theta}$ for A; whereas $s = \cosh\theta$ gives eigenvalues $e^{\pm\theta}$. This suggests we use the substitution $s = \cos\theta$.

(ii) Then we define C_n as the Chebyshev polynomial of the first kind of degree n by the output $y_n = C_n(s)$. These polynomials are characterized by the property that $C_n(\cos\theta) = \cos(n\theta)$, since one can show by induction that, with $s = \cos\theta$,

$$x_n = \begin{bmatrix} \cos(n+1)\theta \\ \cos n\theta \end{bmatrix} \qquad (n = 0, 1, \dots). \tag{8.33}$$

Of course, the induction step is the trigonometric addition rule

$$\cos(n+1)\theta + \cos(n-1)\theta = 2\cos\theta\cos n\theta. \tag{8.34}$$

The zeros of C_n are given by $s = \cos\theta$ such that $\cos n\theta = 0$, so there are n zeros in $[-1, 1]$ from the equally spaced angles

$$\theta = \frac{\pi}{2n}, \frac{3\pi}{2n}, \dots, \frac{(2n-1)\pi}{2n}. \tag{8.35}$$

(iii) We also have the rule

$$\sin(n+1)\theta + \sin(n-1)\theta = 2\cos\theta\sin n\theta, \tag{8.36}$$

which suggests the definition of the Chebyshev polynomials $(U_n)_{n=0}^{\infty}$ such that $U_n(\cos\theta) = \sin(n+1)\theta/\sin\theta$. These are generated by the same recurrence relation (8.30), but the initial condition is

$$\begin{bmatrix} U_1(s) \\ U_0(s) \end{bmatrix} = \begin{bmatrix} 2s \\ 1 \end{bmatrix}. \tag{8.37}$$

(iv) Chebyshev Filters (Fig. 8.1)

Suppose that we require a filter that cuts off signals like the indicator function $\mathbb{I}_{(-1,1)}(x)$, which has a rectangular graph. We cannot find a meromorphic function T in the RHP such that $|T(i\omega)| = \mathbb{I}_{(-1,1)}(\omega)$, since this conflicts with results about the boundary values of holomorphic functions. However, we can approximate the indicator function by using the Chebyshev polynomials. Let $\varepsilon > 0$, and introduce a transfer function $T_n(s) = 1/(1 + \varepsilon i C_n(s/i))$, so the frequency response function is

$$T_n(i\omega) = \frac{1}{1 + \varepsilon i C_n(\omega)} \qquad (\omega \in \mathbb{R}). \tag{8.38}$$

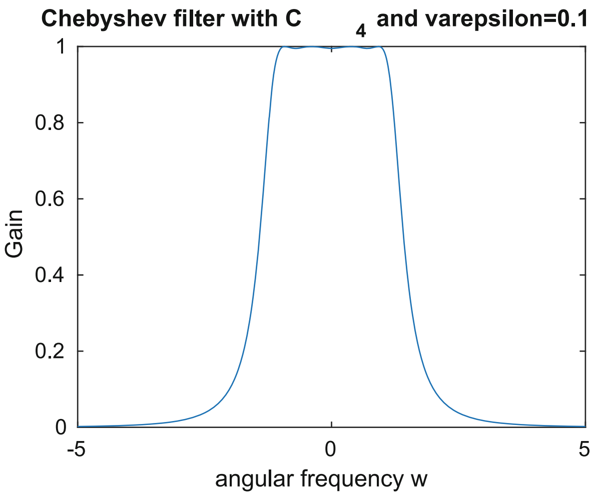


Fig. 8.1 Chebyshev filter

Then for $-1 \leq \omega \leq 1$, we can write $\omega = \cos\theta$ for some $\theta \in \mathbb{R}$, so $C_n(\cos\theta) = \cos n\theta$ and $|C_n(\omega)| \leq 1$; hence the gain is

$$1 \geq |T(i\omega)| = \frac{1}{\sqrt{1 + \varepsilon^2 C_n(\omega)^2}} \geq \frac{1}{\sqrt{1 + \varepsilon^2}} \qquad (\omega \in [-1, 1]), \tag{8.39}$$

while the phase has $\tan\phi = -\varepsilon C_n(\omega)$, so $|\phi| \leq \varepsilon$.

For $\omega > 1$ we write $\omega = \cosh\theta$ for some $\theta > 0$, so $C_n(\cosh\theta) = \cosh n\theta$ and by induction we have $C_n(\cosh\theta) \geq \cosh^n\theta$, so the gain is

$$|T(i\omega)| = \frac{1}{\sqrt{1 + \varepsilon^2 C_n(\omega)^2}} \leq \frac{1}{\sqrt{1 + \varepsilon^2 \omega^{2n}}} \qquad (\omega \in (1, \infty)). \tag{8.40}$$

For $\omega < -1$ we write $\omega = -\cosh\theta = \cosh(\theta + i\pi)$ for some $\theta > 0$, so $C_n(\cosh(\theta + i\pi)) = \cosh n(\theta + i\pi) = (-1)^n \cosh n\theta$ and we argue as in the previous case.

For large n, the gain has a graph which resembles the graph of the indicator function. The Chebyshev filter is easy to compute as the iteration scheme in (8.30) is simple. On the interval $[-1, 1]$ the graph of the gain exhibits a slight ripple effect.

This example is the simplest instance of a rather general result that applies to orthogonal polynomials with respect to weights, which we discuss next.

8.5 Hankel Matrices and Moments

Let $w : \mathbb{R} \to [0, \infty)$ be an integrable function such that

$$0 < \int_{-\infty}^{\infty} |x|^k w(x)dx < \infty \tag{8.41}$$

for all $k = 0, 1, 2, \dots$. Then w is called a weight, and we can generate a sequence of moments

$$\mu_k = \int_{-\infty}^{\infty} x^k w(x)dx \qquad (k = 0, 1, \dots) \tag{8.42}$$

such that $\mu_k \in \mathbb{R}$. Clearly the even moments satisfy $\mu_{2k} \geq 0$; whereas there is no reason to suppose that the odd moments μ_{2k+1} are nonnegative. We introduce the Hankel matrix

$$\Gamma = [\mu_{j+k}]_{j,k=0}^{\infty} = \begin{bmatrix} \mu_0 & \mu_1 & \mu_2 & \dots \\ \mu_1 & \mu_2 & \mu_3 & \dots \\ \mu_2 & \mu_3 & \mu_4 & \dots \\ \mu_3 & \mu_4 & \mu_5 & \dots \\ \vdots & \dots & \dots & \dots \end{bmatrix} \tag{8.43}$$

where the top row gives the sequence of moments. Then Γ is real and symmetric, and has the characteristic property of Hankel matrices that the cross diagonals are constant. In this section we use the subscript N to refer to $(N + 1) \times (N + 1)$ complex matrices, and entries will be indexed with indices j starting from $j = 0$.

Lemma 8.10 *The top left block $\Gamma_N = [\mu_{j+k}]_{j,k=0}^{N}$ is a positive definite matrix.*

Proof Let $f(t) = \sum_{j=0}^{N} a_j t^j$ and $g(t) = \sum_{j=0}^{N} b_j t^j$ be complex polynomials of degree N. Then

$$\begin{aligned} \left\langle \Gamma(a_j)_{j=0}^{N}, (b_j)_{j=0}^{N} \right\rangle &= \sum_{j,k=0}^{N} a_j \bar{b}_k \mu_{j+k} \\ &= \sum_{j,k=0}^{N} a_j \bar{b}_k \int_{-\infty}^{\infty} t^{j+k} w(t)dt \\ &= \int_{-\infty}^{\infty} f(t)\bar{g}(t)w(t)dt, \end{aligned}$$

so in particular

$$\left\langle \Gamma(a_j)_{j=0}^N, (a_j)_{j=0}^N \right\rangle = \int_{-\infty}^{\infty} |f(t)|^2 w(t)dt. \tag{8.44}$$

If $a_j \neq 0$ for some j, then $|f(t)|^2$ is a continuous function that is zero at only finitely many points hence the integral is strictly positive, hence Γ_N is positive definite. □

8.6 Orthogonal Polynomials

We introduce an inner product

$$\langle f, g \rangle_w = \int_{-\infty}^{\infty} f(t)\bar{g}(t)w(t)dt \tag{8.45}$$

on the complex polynomials f, g.

Proposition 8.11 *There exists a sequence of real orthogonal polynomials*

$$f_n(t) = \sum_{j=0}^{n} a_j^{(n)} t^j \tag{8.46}$$

such that (i) $\langle f_j, f_k \rangle = 0$ *for all* $j \neq k$*;*

(ii) the matrix $U_N = [a_j^{(k)}]_{j,k=0}^N$ *is upper triangular with positive diagonal entries* $a_n^{(n)} > 0$*,*

(iii) $U_N' \Gamma_N U_N$ *is diagonal with positive entries on the diagonal.*

Proof

(i) We can apply the Gram-Schmidt process [51, page 258] to $(t^j)_{j=0}^N$ to produce the required $f_j(t)$.

(ii) The matrix of coefficients has the form

$$U_N = \begin{bmatrix} a_0^{(0} & a_0^{(1)} & \dots & a_0^{(N)} \\ 0 & a_1^{(1)} & \dots & a_1^{(N)} \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & \dots & a_N^{(N)} \end{bmatrix}, \tag{8.47}$$

and the jth diagonal entry is the leading coefficients of $f_j(t)$.

(iii) We take an arbitrary pair of complex polynomials of degree $\leq N$, and write $f(t) = \sum_{j=0}^{N} \xi_j f_j(t)$ and $g(t) = \sum_{j=0}^{N} \eta_k f_k(t)$, so that

$$\langle f, g\rangle_w = \sum_{j,k=0}^{N} \xi_j \bar{\eta}_k \langle f_j, f_k\rangle_w = \sum_{j=0}^{N} \xi_j \bar{\eta}_j h_j \tag{8.48}$$

where we have introduced

$$h_j = \int_{-\infty}^{\infty} |f_j(t)|^2 w(t) dt = \langle f_j, f_j\rangle_w. \tag{8.49}$$

We can also write $(e_j)_{j=0}^{N}$ for the standard orthonormal basis for $\mathbb{C}^{(N+1)\times 1}$, and then

$$U_N = \left[U(e_0)\ U(e_1)\ \ldots\ U(e_N)\right], \tag{8.50}$$

so

$$\sum_{j=0}^{N} \xi_j \bar{\eta}_j h_j = \langle f, g\rangle_w = \langle \Gamma_N U(\xi_j)_{j=0}^{N}, U(\eta_j)_{j=0}^{N}\rangle \tag{8.51}$$

as in the Lemma 8.10. Hence we have a diagonal matrix

$$U_N' \Gamma_N U_N = \begin{bmatrix} h_0 & 0 & \ldots & 0 \\ 0 & h_1 & \ddots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & h_N \end{bmatrix}. \tag{8.52}$$

□

There are two common choices for the diagonal constants as in the following Corollary, which correspond to two commonly used normalizations for orthogonal polynomials. For some classical orthogonal polynomials, there are other special normalizations, as in the example of Chebyshev polynomials of the first kind or the Laguerre polynomials. So it is always worth checking which normalization an author is using.

8.7 Hankel Determinants

Corollary 8.12 (Hankel Determinants) *The normalizing constants of the orthogonal polynomials are determined by the sequence of Hankel determinants* $(\Delta_N)_{N=0}^{\infty}$ *in the following distinct cases.*

(i) Let $(P_j)_{j=0}^{N}$ *be the unique sequence of monic orthogonal polynomials with* $h_j = \langle P_j, P_j\rangle_w$. *Then*

$$\Delta_N = \det \Gamma_N = \prod_{j=0}^{N} h_j. \tag{8.53}$$

(ii) Let $(f_j)_{j=0}^{N}$ *be the orthonormal sequence with* $\langle f_j, f_j\rangle_w = 1$ *where* f_j *has leading coefficient* $k_j > 0$. *Then*

$$\Delta_N = \det \Gamma_N = \prod_{j=0}^{N} k_j^{-2}. \tag{8.54}$$

Proof

(i) We can select $a_j^{(j)} = 1$ for $j = 0, \dots, N$, so that P_j is a monic polynomial of degree j and U_N has ones on the leading diagonal. Then matrix U_N is upper triangular, with ones on its leading diagonal, so $\det U_N = 1 = \det U_N'$. Hence from the formula (8.52), we have

$$\det \Gamma_N = \det U_N' \det \Gamma_N \det U_N = \prod_{j=0}^{N} h_j. \tag{8.55}$$

Of course, we can recover h_n from $h_0 = \Delta_0$ and $h_N = \Delta_N/\Delta_{N-1}$ for $N \geq 1$.

(ii) We can choose $h_j = 1$ for all $j = 0, \dots, N$, so that $(f_j)_{j=0}^{N}$ is an orthonormal sequence, and then we write $k_j = a_j^{(j)}$ for the leading coefficients and $U_N' \Gamma_N U_N = I$. In this case U_N is upper triangular with entries k_j on the leading diagonal. Then $U_N' \Gamma_N U_N = I_N$, so taking determinants, we see that

$$1 = \det U_N' \det \Gamma_N \det U_N = \det \Gamma_N \prod_{j=0}^{N} k_j^2. \tag{8.56}$$

Of course, we can recover k_n from $k_0^{-2} = \Delta_0$ and $k_N^{-2} = \Delta_N/\Delta_{N-1}$ for $N \geq 1$.

□

Exercise Show that there exists an upper triangular $(N+1)\times(N+1)$ matrix G such that $\Gamma_N = G'G$.

Definition 8.13 (Completeness) Let $w : [a,b] \to [0,\infty)$ be a weight, and $(P_n(t))_{n=0}^{\infty}$ the corresponding sequence of monic orthogonal polynomials for w, and suppose that $\int_a^b |f(t)|^2 w(t)dt$ converges. We say that $(P_n(t))_{n=0}^{\infty}$ is complete if

$$\int_a^b f(t)P_n(t)dt = 0 \qquad (n = 0, 1, \dots)$$

implies that $f(t) = 0$ on $[a,b]$.

8.8 Laguerre Polynomials

Example 8.14 (Laguerre Polynomials) We introduce the weight $w(t) = e^{-t}$ for $t > 0$, so that the moments are the factorials

$$\mu_k = \int_0^\infty t^k e^{-t}dt = k! \qquad (k = 0, 1, \dots). \tag{8.57}$$

The corresponding sequence of monic orthogonal polynomials begins with

$$f_0(t) = 1, \quad f_1(t) = t - 1, \quad f_2(t) = t^2 - 4t + 2, \tag{8.58}$$

with the corresponding

$$h_0 = 1, h_1 = 1, h_2 = 4; \tag{8.59}$$

hence

$$U_2 = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Gamma_2 = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 6 \\ 2 & 6 & 24 \end{bmatrix}, \quad U_2'\Gamma_2 U_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}. \tag{8.60}$$

The standard Laguerre polynomials $L_n(t)$ are defined as in Exercise 6.15 by

$$L_n(t) = \frac{e^t}{n!}\frac{d^n}{dt^n}\Big(e^{-t}t^n\Big), \tag{8.61}$$

so that $f_n(t) = (-1)^n n! L_n(t)$ is a monic polynomial of degree n. The standard Laguerre polynomials are normalized so that $L_n(0) = 1$, and the leading coefficients are negative for odd n. One can show that

$$\begin{bmatrix} f_{n+1}(t) \\ f_n(t) \end{bmatrix} = \begin{bmatrix} t-1-2n & -n^2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_n(t) \\ f_{n-1}(t) \end{bmatrix}. \tag{8.62}$$

Let the nth Laguerre function be $y_n(t) = e^{-t/2}L_n(t)$, which has Laplace transform

$$Y_n(s) = \frac{1}{n!}\int_0^\infty e^{-(s-1/2)t}\frac{d^n}{dt^n}\Big(e^{-t}t^n\Big)dt \tag{8.63}$$

so we integrate by parts n times over to get

$$Y_n(s) = \frac{(s-1/2)^n}{n!}\int_0^\infty t^n e^{-(s+1/2)t}dt \tag{8.64}$$

so with the substitution $u = (s+1/2)t$, we get

$$Y_n(s) = \frac{(s-1/2)^n}{(s+1/2)^{n+1}}, \tag{8.65}$$

which is a stable rational function with zero of order n at $1/2$ and a pole of order $n+1$ at $-1/2$ in LHP. For this reason, the scaled Laguerre functions $(\sqrt{2}y_n(2t))_{n=0}^\infty$ are a particularly convenient orthonormal basis for $L^2(0,\infty)$. They have Laplace transforms

$$\mathcal{L}(\sqrt{2}e^{-t}L_n(2t))(s) = \sqrt{2}\frac{(s-1)^n}{(s+1)^{n+1}}, \tag{8.66}$$

which match with the functions from Exercise 6.15 as follows. From Lemma 6.33, we recall the algebra $\mathcal{R}$ of proper rational functions with only poles at -1. We observe that

$$\mathrm{span}\Big\{\frac{1}{(1+s)^{k+1}} : k = 0, \dots, n\Big\} = \mathrm{span}\Big\{\sqrt{2}\frac{(s-1)^k}{(s+1)^{k+1}} : k = 0, \dots, n\Big\}$$

for $n = 0, 1, \dots$. We prove this by induction on n. The case $n = 0$ is trivially true, so we suppose the identity has been established for all cases up to $n-1$, and consider the case n. The new function in the left-hand space can be written as

$$\frac{1^n}{(s+1)^{n+1}} = \frac{((s+1)-(s-1))^n}{2^n(s+1)^n} = \sum_{k=0}^n \frac{(-1)^k}{2^n}\binom{n}{k}\frac{(s+1)^{n-k}(s-1)^k}{(s+1)^{n+1}}$$
$$= \sum_{k=0}^n \frac{(-1)^k}{2^n}\binom{n}{k}\frac{(s-1)^k}{(s+1)^{k+1}},$$

hence belongs to the space on the right-hand side, and the spaces are of equal dimension $n + 1$.

The advantage of this basis is that the latest function is the Laplace transform of the following sum of orthogonal functions in $L^2(0, \infty)$, namely

$$\sum_{k=0}^{n} \frac{(-1)^k}{2^n\sqrt{2}} \binom{n}{k} \sqrt{2}e^{-t} L_k(2t). \tag{8.67}$$

In Chap. 10, we will show that the Laguerre orthogonal polynomials $(L_n)_{n=0}^{\infty}$ are complete for the weight e^{-t} on $(0, \infty)$, or equivalently that Laguerre functions $(e^{-t/2}L_n(t))_{n=0}^{\infty}$ give a complete orthonormal basis of $L^2(0, \infty)$. This is the key step in our proof of the Paley–Wiener Theorem 10.36. Our proof also uses the uniqueness of Fourier transforms of L^1-functions. In [50, p. 350] there is an alternative approach, which uses Vitali's completeness theorem from [50, p. 25] as in Exercise 8.12. This uses some special identities for the Laguerre polynomials and has the advantage of being more elementary, but more specialized. In Exercise 8.4, we consider another approach based upon Green's functions which is suitable for Legendre and Chebyshev polynomials which live on bounded intervals.

8.9 Three-Term Recurrence Relation

Proposition 8.15 (Three-Term Recurrence Relation) *Suppose that (f_n) is as in Corollary 8.12 and let*

$$A_n = \frac{k_n}{k_{n-1}}, \quad B_n = -A_n\langle tf_{n-1}, f_{n-1}\rangle_w, \quad C_n = \frac{A_n}{A_{n-1}} = \frac{k_n k_{n-2}}{k_{n-1}^2} \qquad (n = 2, 3, \dots). \tag{8.68}$$

Then the (f_n) satisfy the recurrence relation

$$\begin{bmatrix} f_{n+1}(t) \\ f_n(t) \end{bmatrix} = \begin{bmatrix} A_{n+1}t + B_{n+1} & -C_{n+1} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_n(t) \\ f_{n-1}(t) \end{bmatrix} \qquad (n = 1, 2, \dots). \tag{8.69}$$

We can regard the three-term recurrence relation as generating a discrete time process in which the matrix depends upon the discrete time n, and has t as a parameter.

Proof Szegö [54] gives the three-term recurrence relation in the style

$$f_n(t) = (A_n t + B_n) f_{n-1}(t) - C_n f_{n-2}(t), \tag{8.70}$$

which is equivalent to the matrix version here. To establish this relation, we observe that the given A_n makes $f_n(t) - A_n t f_{n-1}(t)$ a polynomial of degree $\leq n-1$, since we have canceled out the leading coefficients. Hence there exist constants ξ_j such that

$$f_n(t) - A_n t f_{n-1}(t) = \sum_{j=0}^{n-1} \xi_j f_j(t). \tag{8.71}$$

Now for $j = 0, \ldots, n-3$ we have $\langle f_n, f_j \rangle_w = 0$ and $\langle t f_{n-1}(t), f_j(t) \rangle_w = 0$, since $t f_j(t)$ is of degree less than $n-1$. We take the inner product of both sides of (8.71) with f_j and we find $\xi_j = 0$ for $j = 0, \ldots, n-3$. This leaves us with

$$f_n(t) - A_n t f_{n-1}(t) = \xi_{n-1} f_{n-1}(t) + \xi_{n-2} f_{n-2}(t). \tag{8.72}$$

By taking the inner product with f_{n-1} we obtain

$$-A_n \langle t f_{n-1}(t), f_{n-1}(t) \rangle_w = \xi_{n-1} = B_n. \tag{8.73}$$

We then take the inner product with f_{n-2} and obtain

$$\begin{aligned} -C_n &= \xi_{n-2} = -A_n \langle t f_{n-1}, f_{n-2} \rangle_w \\ &= -A_{n-1} \frac{k_{n-2}}{k_{n-1}} \Big\langle f_{n-1}, k_{n-1} t^{n-1} + \text{terms of lower degree} \Big\rangle_w \\ &= -A_{n-1} \frac{k_{n-2}}{k_{n-1}} \langle f_{n-1}, f_{n-1} \rangle_w = -\frac{A_n}{A_{n-1}}. \end{aligned}$$

□

Corollary 8.16 *With respect to the space spanned by the basis $(f_n)_{n=0}^{\infty}$, the operation of multiplication by t is represented by a real symmetric tridiagonal matrix such that the entries on the diagonal below the leading diagonal are positive and the entries on the diagonal are real.*

Proof Here $1/A_{j-1} = C_j/A_j > 0$. Hence the matrix J is real symmetric and tridiagonal with entries on the diagonal below the leading diagonal that are positive; such a matrix is called a Jacobi matrix. □

Definition 8.17 A $(n+1) \times (n+1)$ Jacobi matrix looks like

$$\begin{bmatrix} b_0 & a_0 & \dots & \\ a_0 & b_1 & a_1 & \ddots \\ \vdots & & \ddots & a_{n-1} \\ 0 & \dots & a_{n-1} & b_n \end{bmatrix} \tag{8.74}$$

with $b_j \in \mathbb{R}$ and $a_j > 0$.

Such a Jacobi matrix has real eigenvalues. Given $(k_n)_{n=1}^{\infty}$, we can generate the sequences $(A_n)_{n=1}^{\infty}$ and $(C_n)_{n=1}^{\infty}$ by the recursion formula. Then given (B_n) we can compute the entire sequence $(f_n(t))_{n=0}^{\infty}$ from this recurrence relation. An equivalent form of the three-term recurrence relation is

$$tf_{n-1}(t) = \frac{1}{A_n} f_n(t) - \frac{B_n}{A_n} f_{n-1}(t) + \frac{C_n}{A_n} f_{n-2}(t) \tag{8.75}$$

which expresses the operation of multiplication by t as a tridiagonal matrix with respect to the orthonormal sequence $(f_n)_{n=0}^{\infty}$.

$$\begin{bmatrix} -B_1/A_1 & C_2/A_2 & 0 & \dots \\ 1/A_1 & -B_2/A_2 & C_3/A_3 & \ddots \\ 0 & 1/A_2 & -B_3/A_3 & \ddots \\ \vdots & 0 & 1/A_3 & \ddots \\ \vdots & & \ddots & \ddots \end{bmatrix} \tag{8.76}$$

Corollary 8.18 *Suppose that w is even, so $w(t) = w(-t)$.*

(i) Then the odd moments vanish $\mu_{2j-1} = 0$ and $B_j = 0$ for all $j = 1, 2, \dots$.
(ii) The sequence of monic orthogonal polynomials (f_j) is determined by the recurrence relation (8.70) in which the coefficients A_n and C_n are determined by the Hankel determinants $(\Delta_N)_{N=0}^{\infty}$.

Proof We have $\mu_{2j-1} = \int_{-\infty}^{\infty} t^{2j-1} w(t)dt = 0$. The even indexed polynomials $(f_{2j}(t))_{j=0}^{\infty}$ are all even functions involving only even powers of t, whereas the odd indexed polynomials $(f_{2j-1})_{j=1}^{\infty}$ are all odd functions involving only odd powers of t, so $f_{2j-1}(-t) = -f_{2j-1}(t)$. One can check these facts from the Gram-Schmidt construction [51, page 258], and make a formal proof by induction on the degree. In either case we have $f_{n-1}(t)^2$ even, so

$$B_n = -A_n \int_{-\infty}^{\infty} tf_{n-1}(t)^2 w(t)dt = 0. \tag{8.77}$$

The other coefficients in the recurrence relation are

$$A_n = \frac{k_n}{k_{n-1}} = \left(\frac{\Delta_{n-1}^2}{\Delta_n \Delta_{n-2}}\right)^{1/2}, \quad C_n = \frac{A_n}{A_{n-1}}. \tag{8.78}$$

□

Powers of Infinite Jacobi Matrices

A matrix is fundamentally a table of data. For instance, the outputs of a linear system can be listed in an infinite column vector $[y_j]_{j=0}^{\infty}$. When all the y_j belong to a vector space V we can carry out addition and scalar multiplication on the entries coordinatewise, so that

$$\lambda[y_j]_{j=0}^{\infty} + \mu[z_j]_{j=0}^{\infty} = [\lambda y_j + \mu z_j]_{j=0}^{\infty}. \tag{8.79}$$

Likewise, given a doubly-indexed collection of scalars $a_{j,k}$, we can form an infinite matrix $A = [a_{j,k}]_{j,k=0}^{\infty}$ and carry out scalar addition and multiplication on the entries of such matrices. However, forming the product AB of infinite matrices $A = [a_{j,k}]_{j,k=0}^{\infty}$ and $B = [b_{j,k}]_{j,k=0}^{\infty}$ involves the series $\sum_{k=0}^{\infty} a_{j,k}b_{k,\ell}$ for $j, \ell = 0, 1, \dots$, and we need to ensure that these converge. In the special case in which most of the entries are zero, then this problem is less significant. The identity matrix I with 1 on the leading diagonal and 0 in all the off-diagonal entries satisfies $AI = IA = A$. If A and B have $AB = BA = I$, then A is said to be invertible with inverse B.

In the remainder of this section we consider matrices A that are tridiagonal, so that the entries are all zero, apart from those on the leading diagonal or directly beside the leading diagonal. It is easy to multiply matrices of this special shape, and most importantly they arise in the theory of orthogonal polynomials, as follows.

Let w be a weight on a bounded interval $[a, b]$ such that $\int_a^b w(t)dt = 1$, and let $(f_j)_{j=0}^{\infty}$ be the system of orthonormal polynomials where f_j has degree j and positive leading coefficient. There is a three term recurrence relation

$$tf_j(t) = a_j f_{j+1}(t) + b_j f_j(t) + a_{j-1} f_{j-1}(t) \tag{8.80}$$

where $f_{-1} = 0$ and the coefficients are

$$A = \begin{bmatrix} b_0 & a_0 & \dots & \\ a_0 & b_1 & a_1 & \ddots \\ 0 & a_2 & b_2 & \ddots \\ \vdots & \dots & \ddots & \ddots \end{bmatrix} \tag{8.81}$$

where $a_j > 0$.

Proposition 8.19

(i) The operation of left multiplication by A^n on the column vector $[f_j(t)]_{j=0}^{\infty}$ represents multiplication by t^n, so the (j, k)th entry of A^n is given by

$$[A^n]_{j,\ell} = \int_a^b t^n f_j(t) f_\ell(t) w(t)dt \qquad (n, j, \ell = 0, 1, \dots). \tag{8.82}$$

(ii) Suppose that $(sI - A)$ is invertible for some $s \in \mathbb{C}$ with $|s| > |a|, |b|$. Then the (j, k)th entry of $(sI - A)^{-1}$ is given by

$$\big[(sI - A)^{-1}\big]_{j,k} = \int_a^b \frac{f_j(t) f_k(t) w(t) dt}{s - t} \qquad (j, k = 0, \dots). \tag{8.83}$$

Proof

(i) We write the entries of A as $[A]_{j,k}$ for $j, k = 0, 1, \dots,$ so the three-term recurrence relation becomes

$$t f_j(t) = \sum_{k=0}^{\infty} [A]_{j,k} f_k(t), \tag{8.84}$$

where only finitely many of the terms in this sum are nonzero since A is tridiagonal. Hence we have

$$t^2 f_j(t) = \sum_{k=0}^{\infty} [A]_{jk} t f_k(t) = \sum_{k,\ell=0}^{\infty} [A]_{j,k}[A]_{k,\ell} f_\ell(t) = \sum_{\ell=0}^{\infty} [A^2]_{j,\ell} f_\ell(t), \tag{8.85}$$

and generally by induction on n we have

$$t^n f_j(t) = \sum_{k=0}^{\infty} [A^n]_{j,k} f_k(t) \qquad (j = 0, 1, \dots). \tag{8.86}$$

where all but finitely many terms in the series are zero. From (8.86) we have the formula

$$\int_a^b t^n f_j(t) f_\ell(t) w(t) dt = \sum_{k=0}^{\infty} [A^n]_{jk} \int_a^b f_k(t) f_\ell(t) w(t) dt = [A^n]_{j,\ell} \tag{8.87}$$

which follows by orthogonality.

(ii) We can express the infinite identity matrix as $I = [\int_a^b f_j(t) f_k(t) w(t) dt]_{j,k=0}^{\infty}$. Then by geometric series we have

$$\begin{aligned} \int_a^b \frac{f_j(t) f_k(t) w(t) dt}{s - t} &= \sum_{n=0}^{\infty} \frac{1}{s^{n+1}} \int_a^b t^n f_j(t) f_k(t) w(t) dt \\ &= \sum_{n=0}^{\infty} \frac{1}{s^{n+1}} [A^n]_{j,k} \end{aligned}$$

$$= \Big[\sum_{n=0}^{\infty} s^{-n-1} A^n\Big]_{j,k}$$
$$= \big[(sI - A)^{-1}\big]_{j,k}. \tag{8.88}$$

To check the final identity, one can estimate the entries of $(sI - A)A^n/s^{n+1}$ as $n \to \infty$ and prove they converge to zero.

□

Remark 8.20 Classical orthogonal polynomials

In this book, we mention classical orthogonal polynomials associated with Chebyshev, Legendre, Hermite and Laguerre. Classical orthogonal polynomials have some additional special features, and can be introduced in various ways.

(i) They are orthogonal with respect to weights $w(x)$ defined in terms of elementary functions. In many examples, $(dw/dx)/w(x) = 2V(x)/W(x)$ where V and W are polynomials with W not zero.
(ii) The classical orthogonal polynomials they satisfy differential equations with polynomial coefficients. This is important in applications to physics, and many classical orthogonal polynomials were discovered as solutions of differential equations in various geometrical coordinates. Also, one can classify second-order differential equations with polynomial coefficients according to the singular points where the coefficients are zero; see [61]. There are various results dating back to Laguerre regarding the weights and the differential equations.
(iii) Bochner considered sequences of polynomials $(P_n(x))_{n=0}^{\infty}$ where P_n has degree n that satisfy the differential equation

$$p_0(x)\frac{d^2P_n}{dx^2} + p_1(x)\frac{dP_n}{dx} + p_2(x)P_n(x) + \lambda_n P_n(x) = 0$$

for all x in some common real interval where $p_0(x)$, $p_1(x)$ and $p_2(x)$ are real polynomials and $\lambda_n \in \mathbb{R}$. After scaling transformations, he concluded that the only cases are essentially:

(1) Legendre polynomials, as in (8.90), and related examples of hypergeometric functions;
(2) Laguerre polynomials, as in (8.61), which we use several times in this book;
(3) Hermite polynomials as in (8.111), which are important in the quantum harmonic oscillator;
(4) Bessel type polynomials, as in (8.127) and (8.128) which are related to the Bessel functions $J_{n+1/2}$ and used in linear filters.

(iv) We emphasize the recurrence relations, since these provide an efficient way to calculate orthogonal polynomials. For the above classical polynomials, the recurrence coefficients are rational, so the polynomials can be computed in exact arithmetic. Generally, the coefficients of the three-term recurrence relation (8.70) also determine the properties of the orthogonal polynomials themselves. In particular, the distribution of the zeros of orthogonal polynomials of high degree and the asymptotic form of the polynomials are described in Szegö's theory. In the next chapter we consider $G(s) = \int w(x)dx/(s-x)$, which is the Cauchy transform of the weight w and the moment generating function of (μ_n). We remark that in classical examples

$$W(s)\frac{dG}{ds} = 2V(s)G(s) + U(s) \tag{8.89}$$

for polynomials U, V and W with W nonzero. For a modern discussion of (i),(ii) and (iv), see [37].

Example 8.21 Legendre polynomials

We introduce the Legendre polynomials by

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\Big((x^2-1)^n\Big). \tag{8.90}$$

From the binomial expansion, we deduce that

$$\frac{1}{2^n n!}\sum_{k=0;2k\geq n}^{n}\binom{n}{k}\frac{(2k)!}{(2k-n)!}(-1)^{n-k}x^{2k-n}, \tag{8.91}$$

so in particular P_n has degree n. From this, we deduce that P_n satisfies Legendre's differential equation

$$(1-x^2)\frac{d^2P_n}{dx^2} - 2x\frac{dP_n}{dx} + n(n+1)P_n(x) = 0. \tag{8.92}$$

By integrating by parts for $m > n$, we see that

$$\int_{-1}^{1} P_m(x)P_n(x)dx = \int_{-1}^{1}\frac{1}{2^m m!}\frac{d^m}{dx^m}\Big((x^2-1)^m\Big)\frac{1}{2^n n!}\frac{d^n}{dx^n}\Big((x^2-1)^n\Big)dx = 0. \tag{8.93}$$

We also have

$$\begin{aligned}\int_{-1}^{1} P_n(x)P_n(x)dx &= \frac{(-1)^n}{2^{2n}(n!)^2}\int_{-1}^{1}(x^2-1)^n\frac{d^{2n}}{dx^{2n}}\Big((x^2-1)^n\Big)dx\\ &= \frac{(-1)^n(2n)!}{2^{2n}(n!)^2}\int_{-1}^{1}(x^2-1)^n dx\end{aligned}$$

$$= \frac{(2n)!}{2^{2n}(n!)^2} \int_0^{\pi} \sin^{2n+1}\theta \, d\theta$$
$$= \frac{2}{2n+1},$$

where we have used the substitution $x = \cos\theta$. For small values of n, one can compute the Legendre polynomials by applying the Gram–Schmidt process to the polynomials $(x^n)_{n=1}^{\infty}$ in $L^2[-1, 1]$ for the weight $w(x) = 1/2$. More generally, one can consider the differential equation

$$(1 - x^2)\frac{d^2 S_n}{dx^2} - 2x\frac{dS_n}{dx} - t^2 x^2 S_n(x) - \mu_n S_n = 0 \tag{8.94}$$

which has solutions given by the prolate spheroidal wave functions; see page 295 of [7], page 99 of [43] and page 213 of [36]. These functions have applications to signal processing. See also [35].

8.10 Moments via Discrete Time Linear Systems

Example 8.22 (Moments from a Discrete Time Linear System) Suppose that w is a weight on $[-1, 1]$, and introduce an inner produce by $\langle f, g\rangle = \int_{-1}^{1} f(t)\bar{g}(t)dt$. This inner product is associated with the space $L^2[-1, 1]$ and does not involve w. Then we introduce

$$\begin{aligned} A :& L^2[-1, 1] \to L^2[-1, 1] : f(t) \mapsto tf(t) \qquad (f \in L^2[-1, 1]),\\ B :& \mathbb{C} \to L^2[-1, 1] : \qquad b \mapsto \sqrt{w(t)}b \qquad (b \in \mathbb{C}),\\ C :& L^2[-1, 1] \to \mathbb{C} : \qquad f \mapsto \int_{-1}^{1} f(t)\sqrt{w(t)}dt \qquad (f \in L^2[-1, 1]),\\ D :& \mathbb{C} \to \mathbb{C} : \qquad c \mapsto 0 \qquad (c \in \mathbb{C}). \end{aligned}$$

Then we have

$$CA^n B = \int_{-1}^{1} t^n w(t)dt = \mu_n \tag{8.95}$$

and the transfer function is

$$T(z) = D + \sum_{n=0}^{\infty} z^{n+1} C A^n B = \int_{-1}^{1} \sum_{n=0}^{\infty} z^{n+1} t^n w(t) dt$$
$$= \int_{-1}^{1} \frac{z}{1 - zt} w(t) dt.$$

This transfer function is commonly studied in a slightly different form, since

$$T(1/z) = \int_{-1}^{1} \frac{1}{z - t} w(t) dt \tag{8.96}$$

is the Cauchy transform of w. This example is a realization of a transfer function via a discrete time system, which is different from the situation of Proposition 8.3 since the state space $L^2[-1, 1]$ is infinite-dimensional. This extra flexibility allows us to consider a wider range of examples, and we will pursue this idea in Proposition 10.29.

Example 8.23 (Chebyshev Polynomials) For example, let

$$w(t) = \frac{1}{\sqrt{1 - t^2}} \qquad (-1 < t < 1) \tag{8.97}$$

be the Chebyshev weight on $(-1, 1)$. The corresponding transfer function is

$$T(z) = \int_{-1}^{1} \frac{z}{1 - zt} \frac{1}{\sqrt{1 - t^2}} dt \tag{8.98}$$

which reduces with the substitution $t = \sin\theta$ to

$$T(z) = \int_{-\pi/2}^{\pi/2} \frac{z}{1 - z\sin\theta} d\theta = \frac{\pi z}{\sqrt{1 - z^2}}. \tag{8.99}$$

The final step is given by contour integration, or a $\tan\theta/2$ substitution.

The Chebyshev polynomials of the first kind are the orthogonal polynomials with respect to this weight, with the normalization $C_n(\cos\theta) = \cos(n\theta)$. Then

$$\mu_n = \int_{-1}^{1} \frac{t^n}{\sqrt{1 - t^2}} dt \tag{8.100}$$

can be computed using the substitution $t = \sin\theta$ or otherwise to give

$$\mu_{2k} = \int_{-\pi/2}^{\pi/2} \sin^{2k}\theta d\theta = \frac{(2k - 1)(2k - 3)\dots 1}{(2k)(2k - 2)\dots 2}\pi \tag{8.101}$$

and $\mu_{2k-1} = 0$. Then the Hankel matrix has the characteristic banded pattern arising from an even weight

$$\Gamma_5 = \pi \begin{bmatrix} 1 & 0 & 1/2 & 0 & 3/8 & 0 \\ 0 & 1/2 & 0 & 3/8 & 0 & 5/16 \\ 1/2 & 0 & 3/8 & 0 & 5/16 & 0 \\ 0 & 3/8 & 0 & 5/16 & 0 & 35/128 \\ 3/8 & 0 & 5/16 & 0 & 35/128 & 0 \\ 0 & 5/16 & 0 & 35/128 & 0 & 63/256 \end{bmatrix}. \tag{8.102}$$

The coefficients of the Chebyshev polynomials are given by the columns of

$$U_5 = \begin{bmatrix} 1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -3 & 0 & 5 \\ 0 & 0 & 2 & 0 & -8 & 0 \\ 0 & 0 & 0 & 4 & 0 & -20 \\ 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 16 \end{bmatrix} \tag{8.103}$$

as in $T_2(u) = 2u^2 - 1$, so that

$$U_5' \Gamma_5 U_5 = \pi \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 \end{bmatrix}, \tag{8.104}$$

where the diagonal entries arise from the special choice of normalization.

(ii) Let ξ be a random variable such that $\mathbb{P}[\xi \leq x] = \int_{-1}^{x} w(t)dt/\pi$. Then ξ is called an arc sine random variable, and $-\xi$ is distributed as ξ.

Example 8.24 (Semicircle Moments) The transfer function of the moment sequence of $S(0, 2)$ is given by expanding the geometric series

$$T(z) = \frac{1}{2\pi} \int_{-2}^{2} \frac{z}{1 - xz} \sqrt{4 - x^2} dx = \frac{1}{2\pi} \int_{-2}^{2} \sum_{k=0}^{\infty} z^{k+1} x^k \sqrt{4 - x^2} dx \tag{8.105}$$

and substituting $x = 2 \sin\theta$ to give

$$T(z) = \frac{2}{\pi} \int_{-\pi/2}^{\pi/2} \sum_{k=0}^{\infty} z^{k+1} 2^k \sin^k\theta (1 - \sin^2\theta) d\theta \tag{8.106}$$

in which only the even powers contribute, so we obtain

$$T(z) = \sum_{k=0}^{\infty} (2z)^{2k+1} \frac{1}{2k+1} \frac{(2k+1)(2k-1)\dots 3\cdot 1}{(2k+2)(2k)\dots 2} = \sum_{k=0}^{\infty} (2z)^{2k+1} (-1)^k \binom{1/2}{k+1} \tag{8.107}$$

so by the binomial theorem we conclude that

$$T(z) = \frac{1-\sqrt{1-4z^2}}{2z} \qquad (|z| < 1/2). \tag{8.108}$$

Example 8.25 (Gaussian Weight) The weight

$$\gamma(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \tag{8.109}$$

is even and the even moments are

$$\mu_{2n} = (2n-1)(2n-3)\dots 1 = \frac{(2n)!}{2^n n!}. \tag{8.110}$$

This weight gives rise to an orthogonal sequence of monic polynomials called the Hermite polynomials He_n. These satisfy the recurrence relation

$$\begin{bmatrix} He_{n+1}(x) \\ He_n(x) \end{bmatrix} = \begin{bmatrix} x & -n \\ 1 & 0 \end{bmatrix} \begin{bmatrix} He_n(x) \\ He_{n-1}(x) \end{bmatrix}. \tag{8.111}$$

This γ is the probability density function for a Gaussian or normal $N(0, 1)$ random variable X with mean 0 and variance 1.

8.11 Floquet Multipliers

Let $q : \mathbb{R} \to \mathbb{R}$ be a continuous and periodic function with period 2π; then Hill's equation is

$$-\frac{d^2x}{dt^2} + q(t)x(t) = \lambda x(t) \qquad (t \in \mathbb{R}), \tag{8.112}$$

where $\lambda \in \mathbb{C}$ is a complex parameter and $x : (0, \infty) \to \mathbb{C}$ is a solution. There may or may not be a periodic solution; so given a solution, we can consider how $x(t)$ relates to $x(2\pi)$. To do this systematically, the basic idea is to link the continuous time differential equation with a discrete time process determined by a 2×2 matrix

A, which is variously called the transit matrix or the monodromy matrix. The matrix A depends upon λ, but we often suppress this in the notation; likewise, we consider the discrete-time process $(X(0), X(2\pi), X(4\pi)\dots,)$, where the vectors $X(2\pi j) \in \mathbb{C}^{2\times 1}$ also depend upon λ. We write the differential equation as

$$\frac{dX}{dt} = \begin{bmatrix} 0 & 1 \\ q(t) - \lambda & 0 \end{bmatrix} X, \qquad X = \begin{bmatrix} x \\ \frac{dx}{dt} \end{bmatrix} \tag{8.113}$$

so that the matrix is periodic in t and has zeros on the leading diagonal. We can build 2×2 matrix solutions by taking vectors X_1 and X_2 that satisfy the differential equation, and forming $F = [X_1, X_2]$.

Lemma 8.26

(i) There exists a unique 2×2 *matrix* F *that satisfies this differential equation and the initial condition* $F(0) = I_2$*.*
(ii) Let $A = F(2\pi)$*, so*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{8.114}$$

Then the characteristic equation of A *is*

$$s^2 - (a + d)s + 1 = 0, \tag{8.115}$$

with Hill's discriminant $\Delta_\lambda = a + b$*.*

Proof By basic theory of differential equations [26], there exists a unique 2×2 matrix F that satisfies this differential equation and the initial condition $F(0) = I_2$. Since the matrix in the differential equation has zero trace, the Wronskian

$$\det F(t) = \det \begin{bmatrix} x_1(t) & x_2(t) \\ \frac{dx_1}{dt} & \frac{dx_2}{dt} \end{bmatrix} \tag{8.116}$$

is a constant, and we can prove directly since the differential equation gives

$$\begin{aligned} \frac{d}{dt}\Big(x_1(t)\frac{dx_2}{dt} - \frac{dx_1}{dt}x_2(t)\Big) &= x_1(t)\frac{d^2x_2}{dt^2} - \frac{d^2x_1}{dt^2}x_2(t) \\ &= x_1(t)x_2(t)(q(t) - \lambda - q(t) + \lambda) = 0. \end{aligned}$$

The initial condition gives $\det F(0) = 1$. The matrix $F(t + 2\pi)$ also satisfies the differential equation, so by uniqueness, we have $F(2\pi + t) = F(t)A$ for some matrix A that is independent of t. Hence $F(2\pi) = F(0)A = A$. Also

$$1 = \det F(2\pi + 0) = \det F(0) \det A = \det A. \tag{8.117}$$

Hence we have where $\det A = ad - bc = 1$ and $\text{trace}(A) = a + d$. See [38] for a discussion of Floquet theory. □

Proposition 8.27 *Suppose that λ is real. Then A is also real, and there are four cases for the roots of the characteristic equation.*

(i) If $(a+d)^2 < 4$, then A has a pair of complex conjugate eigenvalues on the circle $\{s : |s| = 1\}$.
(ii) If $(a+d)^2 > 4$, then A has a pair of real eigenvalues of the same sign, one inside $\{s : |s| = 1\}$, the other outside.
(iii) If $a + d = -2$, then A has an eigenvalue $s = 1$, and (8.113) has a periodic solution.
(iv) If $a + d = 2$, then A has an eigenvalue $s = -1$, and (8.113) has an anti periodic solution such that $x(t + 2\pi) = -x(t)$.

Proof

(i) Here

$$\lambda_{\pm} = \frac{(a+d) \pm \sqrt{(a+d)^2 - 4}}{2} \tag{8.118}$$

and the eigenvalues have product 1 so are a pair of complex conjugate roots on the unit circle.
(ii) Here the eigenvalues are real and have product 1, hence are of the same sign, and exactly one of them is inside the unit circle.
(iii) Let V_+ be a nonzero vector such that $AV_+ = V_+$; then $X(t) = F(t)V_+$ is a solution of the differential equation that satisfies $X(2\pi) = AV_+ = V_+ = X(0)$, so $X(t)$ gives a periodic solution.
(iv) We let V_+ be a nonzero vector such that $AV_+ = V_+$ and choose $X(t) = F(t)V_-$ so $X(2\pi) = AV_- = -V_- = -X(0)$.

The fundamental solution satisfies $F(2\pi n) = A^n$, so we expect to have bounded solutions in case (i) and unbounded solutions in case (ii). Hill's discriminant determines the eigenvalues via (8.118), hence describes the nature of the solutions. □

8.12 Exercises

Exercise 8.1 (Fibonacci Sequence) For the recurrence relation

$$X_{n+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} X_n, \quad X_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{8.119}$$

compute the first few terms, and the eigenvalues of the matrix.

Exercise 8.2 (Legendre Polynomials) For the weight $w(t) = 1$ for $t \in (0, 1)$ and $w(t) = 0$ otherwise, calculate the first few monic orthogonal polynomials.

In the notation of the section, introduce the coefficients of the polynomials and the moments by

$$U_2 = \begin{bmatrix} 1 & -1/2 & 1/6 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \Gamma_2 = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \tag{8.120}$$

such that the normalizing coefficients satisfy

$$U_2' G_2 U_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/12 & 0 \\ 0 & 0 & 1/180 \end{bmatrix}. \tag{8.121}$$

Exercise 8.3 See [47]. Let $w(t) = 1/2$ for $t \in (-1, 1)$ and $w(t) = 0$ otherwise.

(i) Show that the moment sequence of w is $(1, 0, 1/3, 0, 1/5, \dots)$.
(ii) Show that the Cauchy transform $G(s) = \int_{-1}^{1} w(x)dx/(s - x)$ satisfies

$$G(s) = \frac{1}{2} \log \left(\frac{s+1}{s-1} \right) \qquad (s \in \mathbb{C} \setminus [-1, 1]). \tag{8.122}$$

Exercise 8.4 See [47]. Let w be a weight on $[-1, 1]$, and

$$T(z) = D + \int_{-1}^{1} \frac{zw(x)dx}{1 - xz}. \tag{8.123}$$

Show that the change of variables $z = (s - 1)/(s + 1)$ and $x = (t - 1)/(t + 1)$ transforms this to

$$Z(s) = D + \frac{s-1}{2} \int_0^\infty \frac{W(t)dt}{s+t} \tag{8.124}$$

where

$$W(t) = \frac{2}{(1+t)^2} w\left(\frac{t-1}{t+1} \right) \qquad (t > 0), \tag{8.125}$$

and the new integral involves a Carleman integral.

Exercise 8.5 (Chebyshev Filter)

(i) Use the recursion formula to compute the Chebyshev polynomial $C_6(s)$.
(ii) Plot the gain of the frequency response function $T_6(i\omega) = 1/(1 + i\varepsilon C_6(\omega))$ where $\varepsilon = 0.1$.

Exercise 8.6 (Bessel Polynomials) The Bessel polynomials may be defined by the recurrence relation

$$
\begin{aligned}
y_0(x) &= 1, \quad y_1(x) = 1 + x \\
y_{n+1}(x) &= (2n+1)xy_n(x) + y_{n-1}(x).
\end{aligned}
\tag{8.126}
$$

(i) Compare this with the recurrence relation for Bessel functions of integral order, and show that

$$y_n(x) = \sum_{k=0}^{n} \frac{(n+k)!}{k!(n-k)!}\left(\frac{x}{2}\right)^k \tag{8.127}$$

satisfies this relation.

(ii) Compute the Laplace transform $Y_n(s)$ of $y_n(x)$.

Exercise 8.7 (Reverse Bessel Polynomials) Let $\theta_n(x) = x^n y_n(1/x)$ be the reverse Bessel polynomial where y_n is as in Exercise 8.6.

(i) Show that

$$\theta_n(x) = \sum_{k=0}^{n} \frac{(n+k)!}{k!(n-k)!}\frac{x^{n-k}}{2^k} \tag{8.128}$$

and find the Laplace transform of $\theta_n(x)$.

(ii) Show that the reverse Bessel polynomials may be defined by

$$\begin{bmatrix}\theta_1(s)\\ \theta_0(s)\end{bmatrix} = \begin{bmatrix}s+1\\ 1\end{bmatrix} \tag{8.129}$$

and the recursion formula

$$\begin{bmatrix}\theta_{n+1}(s)\\ \theta_n(s)\end{bmatrix} = \begin{bmatrix}2n+1 & s^2\\ 1 & 0\end{bmatrix}\begin{bmatrix}\theta_n(s)\\ \theta_{n-1}(s)\end{bmatrix} \qquad (n = 1, 2, \ldots). \tag{8.130}$$

(iii) Show that θ_n is a monic polynomial of degree n with positive coefficients, and find an expression for $\theta_n(0)$.

The Bessel filter has transfer function

$$\Theta_n(s) = \frac{\theta_n(0)}{\theta_n(s/\omega_0)}, \tag{8.131}$$

where $\omega_0 > 0$ is a scaling parameter for the frequency.

Exercise 8.8 (Laguerre Polynomials) The Laguerre polynomial of order α and degree n for $\alpha, n = 0, 1, \dots$ is $L_n^{(\alpha)}(x)$ which satisfies the differential equation

$$x\frac{d^2}{dx^2}L_n^{(\alpha)}(x) + (\alpha + 1 - x)\frac{d}{dx}L_n^{(\alpha)}(x) + nL_n^{(\alpha)}(x) = 0. \tag{8.132}$$

(i) Let $h_n^{(\alpha)}(x) = x^\alpha e^{-x} L_n^{(\alpha)}(x)$, and show that

$$x\frac{d^2}{dx^2}h_n^{(\alpha)}(x) + (1-\alpha)\frac{d}{dx}h_n^{(\alpha)}(x) + (2n+1+\alpha-x)h_n^{(\alpha)}(x) = 0. \tag{8.133}$$

(ii) Show that the Laplace transform $H_n^{(\alpha)}(s)$ of $h_n^{(\alpha)}(x)$ satisfies

$$H_n^{(\alpha)}(s) = C_{n,\alpha}\frac{(s-1)^n}{(s+1)^{n+1+\alpha}}, \tag{8.134}$$

for some constant $C_{n,\alpha}$.

Exercise 8.9 (Toda's Equation) Suppose that orthogonal polynomials $(P_n)_{n=0}^{\infty}$ make vectors

$$X_n(s) = \begin{bmatrix} P_{n+1}(s) \\ P_n(s) \end{bmatrix} \tag{8.135}$$

that satisfies the system of equations

$$\begin{aligned} X_{n+1}(s) &= A_n(s)X_n(s), \\ \frac{d}{ds}X_n(s) &= \Omega_n(s)X_n(s) \qquad (n = 0, 1, \dots), \end{aligned} \tag{8.136}$$

where $A_n(s)$ and $\Omega_n(s)$ are 2×2 matrices with rational function entries. Show that these are consistent, provided that

$$\frac{d}{ds}A_n(s) = \Omega_{n+1}(s)A_n(s) - A_n(s)\Omega_n(s). \tag{8.137}$$

Exercise 8.10 (Uniqueness of Moment Sequences) Suppose that w is a weight on $\mathbb{R}$ such that

$$\int_{-\infty}^{\infty} \cosh(\omega_0 t)w(t)\,dt$$

converges for some $\omega_0 > 0$.

(i) Show that

$$f(x+iy) = \int_{-\infty}^{\infty} e^{-(x+iy)t} w(t)dt$$

defines a holomorphic function on the vertical strip $\{x+iy : -\omega_0 < x < \omega_0\}$ with

$$\frac{d^k f}{dz^k}(0) = (-i)^k \int_{-\infty}^{\infty} t^k w(t)dt \qquad (k = 0, 1, \dots).$$

(ii) Show that $d^k f/dz^k)(0))_{k=0}^{\infty}$ uniquely determines f, and hence determines w.

Exercise 8.11 (Prolate Spheroidal Wave Functions) For $\lambda \in \mathbb{R}$, let K be the differential operator

$$Kf(x) = (1-x^2)\frac{d^2 f}{dx^2} - 2x\frac{df}{dx} - \lambda^2 x^2 f(x), \tag{8.138}$$

and let

$$Uf(x) = \int_{-1}^{1} e^{i\lambda xy} f(y)dy. \tag{8.139}$$

Show that

$$KUf(x) = \int_{-1}^{1} \big(\lambda^2 x^2 y^2 - \lambda^2 y^2 - \lambda^2 x^2 - 2i\lambda xy\big)e^{i\lambda xy} f(y)dy \tag{8.140}$$

and show by integration by parts that $UKf(x) = KUf(x)$ for all $f \in C^2([-1,1];\mathbb{C})$. This calculation can be used to show that the eigenfunctions of $U'U$ are eigenfunctions of K, satisfying

$$(1-x^2)\frac{d^2 f}{dx^2} - 2x\frac{df}{dx} - \lambda^2 x^2 f(x) = \mu f(x), \tag{8.141}$$

for some $\mu = \mu(\lambda)$. The integral operator $U'U$ is computed in Exercise 10.11.

Exercise 8.12 (Vitali's Completeness Theorem; See [50, p. 25]) Let w be a weight on $[a,b]$ and let $(f_n(t))_{n=0}^{\infty}$ be the sequence of orthonormal polynomials for w. Show that

$$\int_a^x w(t)dt \geq \sum_{n=0}^{\infty} \Big(\int_a^x f_n(t)w(t)dt\Big)^2 \qquad (x \in [a,b]).$$

Vitali's theorem states that $(f_n(t))_{n=0}^{\infty}$ is a complete sequence of orthogonal polynomials, if and only if equality holds for all $x \in [a, b]$.

Exercise 8.13 Prove the identity (8.24). The method is similar to Exercise 3.16 regarding the second resolvent identity.

Exercise 8.14 Theorem 8.8 gives a map $\Sigma \mapsto (A, B, C, D)$ from discrete to continuous-time linear systems. Find the inverse map, and obtain converse statements for Theorem 8.8 (i) and (ii). The starting point is to show that A_d is the Cayley transform of A.

Exercise 8.15

(i) Find the eigenvalues and eigenvectors for the matrix A in (8.30).
(ii) Do likewise for the matrix in (6.121).

Chapter 9
Random Linear Systems and Green's Functions

In this chapter we consider some applications of discrete time linear systems to various models. We consider a case in which either the input to the linear systems is random, known as the ARMA process. Then we consider the Cauchy transform of Green's function associated with a distribution. This enables us to use results of complex analysis and we can use ideas from the preceding chapter regarding orthogonal polynomials. We consider models in which the main transformation is a random matrix, and achieve results in specific cases where we can carry out calculations explicitly. These include results on the semicircle distribution, which is an important topic in modern wireless communication. Another application is to population dynamics, namely the May–Wigner model.

9.1 ARMA Process

Auto-regressive moving average models are commonly used in economics. The input is taken to be random, to reflect changing economic circumstances. The output involves outputs from the recent past and inputs from the recent past; for instance; current process can be affected by prices from the recent past. The relationship between these quantities is expressed in a linear equation with constant coefficient. More specifically, let $(\varepsilon_k)_{k=0}^{\infty}$ be a sequence of mutually independent random variables with identical distribution, with mean $\mathbb{E}\varepsilon_k = 0$ and variance $\mathbb{E}\varepsilon_k^2 = 1$. We take constants $a_1, \dots, a_n$ and $c_1, \dots, c_m$, and suppose that the outputs and inputs from the recent past are related by the linear equation

$$y_k + a_1 y_{k-1} + \dots + a_n y_{k-n} = \varepsilon_k + c_1 \varepsilon_{k-1} + \dots + c_m \varepsilon_{k-m} \qquad (k = 1, 2, \dots). \tag{9.1}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_9

In Chap. 2, we showed how an nth order differential equation could be transformed into a first-order matrix differential equation. In a similar way, we can transform a difference equation in several variables into a matrix difference equation. We let the state be

$$X_k = \begin{bmatrix} y_{k-1} \\ y_{k-2} \\ \vdots \\ y_{k-n} \end{bmatrix}, \tag{9.2}$$

and the random input be

$$U_k = \begin{bmatrix} \varepsilon_k \\ \varepsilon_{k-1} \\ \vdots \\ \varepsilon_{k-m} \end{bmatrix}, \tag{9.3}$$

which we make into a discrete time system via

$$\begin{aligned} A &= \begin{bmatrix} -a_1 & -a_2 & \dots\dots & -a_n \\ 1 & 0 & \dots\dots & 0 \\ 0 & 1 & \dots\dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots\ 1 & 0 \end{bmatrix} \in M_{n\times n}(\mathbb{C}) \\ B &= \begin{bmatrix} 1\ c_1 & \dots & c_m \\ & 0_{(n-1)\times m} & \end{bmatrix} \in M_{n\times(m+1)}(\mathbb{C}) \\ C &= \begin{bmatrix} 1\ 0 \dots 0 \end{bmatrix} \in M_{1\times n}(\mathbb{C}) \\ D &= 0 \end{aligned} \tag{9.4}$$

Then (9.1) is equivalent to

$$\begin{aligned} X_{k+1} &= AX_k + BU_k \\ y_{k-1} &= CX_k. \end{aligned} \tag{9.5}$$

We have

$$\det(sI - A) = s^n + a_1 s^{n-1} + \cdots + a_n. \tag{9.6}$$

Suppose that all the eigenvalues λ of A satisfy $|\lambda| < 1$. Then $(I - zA)$ is invertible for all $z \in \mathbb{D}(0, 1)$.

9.2 Distributions on a Bounded Interval

Proposition 9.1 *The following data are equivalent, and define the notion of an distribution on* $[-M, M]$*:*

(1) An increasing function $F : [-M, M] \to [0, 1]$ *such that* $F(-M) = 0$ *and* $F(M) = 1$*, where* F *is right-continuous so that* $\lim_{y \to x+} F(y) = F(x)$ *for all* $x \in [-M, M)$*;*

(2) A positive linear functional $\phi : C([-M, M]; \mathbb{R}) \to \mathbb{R}$ *such that*

$$\phi(\lambda f + \mu g) = \lambda\phi(f) + \mu\phi(g) \qquad (\lambda, \mu \in \mathbb{R};\ f, g \in C([-M, M]; \mathbb{R})), \tag{9.7}$$

such that $\phi(1) = 1$ *and* $\phi(h) \geq 0$ *for all* $h \in C([-M, M]; \mathbb{R})$ *such that* $h(x) \geq 0$ *for all* $x \in [-M, M]$*;*

(3) A probability measure ν *on* $[-M, M]$*;*

(4) The cumulative distribution function F *of a bounded random variable* $\xi : \Omega \to [-M, M]$ *on a probability space* $(\Omega, \mathbb{P})$ *such that* $F(x) = \mathbb{P}[\xi \leq x]$*.*

Proof The details of this equivalence are discussed in books on measure theory, so we give only a brief indication of how the quantities relate to one another. Given (1) illustrated by Fig. 9.1, we can construct a Stieltjes integral, which defines ϕ via

$$\phi(g) = \int_{[-M,M]} g(x)dF(x) \qquad (g \in C([-M, M]; \mathbb{R})); \tag{9.8}$$

and one easily show that ϕ satisfies the conditions of (2). Conversely, F. Riesz showed that all ϕ from (2) arise from an integral in this way; this is the representation theorem for linear functionals. Here we make essential use of the assumption that $[-M, M]$ is closed and bounded.

(3) The Stieltjes integral can equivalently be defined in terms of a probability measure ν such that $\nu(a, b] = F(b) - F(a)$.

Suppose as in (4) that ξ is a bounded random variable on a probability space with probability measure $\mathbb{P}$. Then the distribution of ξ is specified by the cumulative distribution function F on $[-M, M]$ so that $\mathbb{P}[\xi \leq x] = F(x)$, and we can write the expectation of the random variable $g \circ \xi$ as

$$\mathbb{E}g(\xi) = \int_{[-M,M]} g(x)dF(x) \qquad (g \in C([-M, M]; \mathbb{R})). \tag{9.9}$$

Observe that $\phi(g) = \mathbb{E}g(\xi)$ has the properties of (2).

In particular, ξ has expectation or mean given by the first moment, so $\mu_1 = \mathbb{E}\xi = \int t dF(t)$. The second moment is $\mu_2 = \mathbb{E}\xi^2 = \int t^2 dF(t)$, and the variance is $\sigma^2 = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$. Generally, the nth moment $\mathbb{E}\xi^n$ arises from $g(t) = t^n$. □

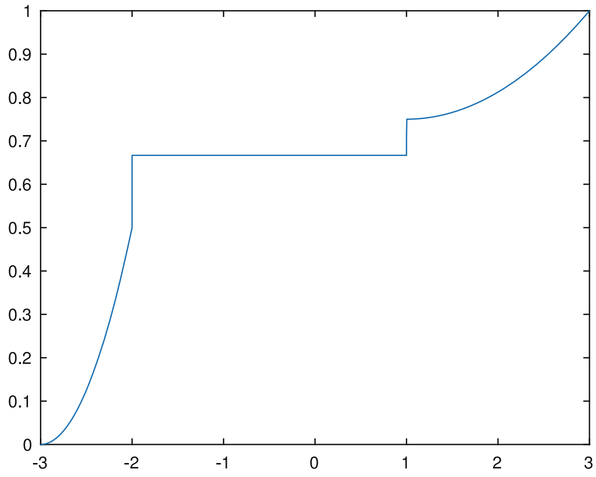


Fig. 9.1 Graph of a cumulative distribution function with jumps at -2 and 1

9.3 Cauchy Transforms

In terms of the previous section, the Cauchy transform is equivalently defined by

$$(1)\quad G(s) = \int \frac{dF(x)}{s-x}, \qquad (2)\quad G(s) = \phi\Big(\frac{1}{s-x}\Big),$$
$$(3)\quad G(s) = \int \frac{\nu(dx)}{s-x}, \qquad (4)\quad G(s) = \mathbb{E}\Big(\frac{1}{s-\xi}\Big),$$

where we take $g(x) = 1/(s-x)$ for $x \in [-M, M]$ and $s \in \mathbb{C} \setminus [-M, M]$. The set $\mathbb{C} \setminus [-M, M]$ is known as the one-cut plane, and is a connected open set.

In this section, we focus on (1), and in the following Lemma prove properties (i)–(iv) of $G(s)$ that reflect the properties of F. The cumulative distribution function F can be discontinuous. For instance, there can exist a sequence of $x_j \in [-M, M]$ such that $\mathbb{P}[\xi = x_j] = F(x_j) - F(x_j-) > 0$, so F jumps up at each x_j. Case (iii) of the following result can be used in this case. Another circumstance is when F is continuously differentiable on (a, b) so that $F(x) = F(a) + \int_a^x f(t)dt$ for $a < x < b$, so $F'(x) = f(x)$. Case (iv) can be used when ξ is a continuous random variable with continuous probability density function dF/dx.

The Cauchy transform proves us with a generating function for moments, with the following properties.

Lemma 9.2 (Cauchy Transforms) *Suppose that* $F : [-M, M] \to [0, 1]$ *is increasing with* $F(-M) = 0$ *and* $F(M) = 1$*, and right-continuous, so that*

$\lim_{y\to x+} F(y) = F(x)$ for all $x \in [-M, M)$. Then the Cauchy transform of F is

$$G(s) = \int_{-M}^{M} \frac{dF(x)}{s-x}. \tag{9.10}$$

(i) Then $G(s)$ holomorphic on $\mathbb{C} \setminus [-M, M]$ with $G(\bar{s}) = \overline{G(s)}$ and $\Im G(s) < 0$ for all s such that $\Im s > 0$, so $-G(s)$ and $G(-s)$ take the upper half plane $\{s : \Im s > 0\}$ to the upper half plane.

(ii) There is a convergent power series expansion

$$G(s) = \frac{1}{s} + \sum_{n=1}^{\infty} \frac{\mu_n}{s^{n+1}} \tag{9.11}$$

for $|s| > M$, which is determined by the moments $\mu_n = \int_{-M}^{M} x^n dF(x)$, so $G(s)$ is holomorphic near ∞.

(iii) Suppose that F jumps at x_0. Then the height of the jump is

$$F(x_0) - F(x_0-) = \lim_{h\to 0+} \frac{h}{2i}\big(G(x_0 - ih) - G(x_0 + ih)\big). \tag{9.12}$$

(iv) Suppose that F is differentiable on $(x_0 - \varepsilon, x_0 + \varepsilon)$ for some $\varepsilon > 0$ and that the derivative dF/dx is continuous there. Then

$$\frac{dF}{dx}(x_0) = \lim_{h\to 0+} \frac{1}{2\pi i}\big(G(x_0 - ih) - G(x_0 + ih)\big). \tag{9.13}$$

Proof

(i) The function $1/(s-x)$ is differentiable with respect to the complex variable s for $x \in [-M, M]$ and $s \in \mathbb{C} \setminus [-M, M]$. We check that

$$\frac{dG}{ds} = -\int_{-M}^{M} \frac{dF(x)}{(s-x)^2} \qquad (\mathbb{C} \setminus [-M, M]) \tag{9.14}$$

and

$$G(t + i\sigma) = \int_{-M}^{M} \frac{dF(x)}{t + i\sigma - x} = \int_{-M}^{M} \frac{(t - x - i\sigma)dF(x)}{(x-t)^2 + \sigma^2}, \tag{9.15}$$

so

$$\Im G(t + i\sigma) = \int_{-M}^{M} \frac{-\sigma dF(x)}{(x-t)^2 + \sigma^2}, \tag{9.16}$$

so $\Im G(t+i\sigma)$ takes the same sign as $-\sigma$.

(ii) For $|s|>M$, the geometric series $1/(s-x)=\sum_{n=0}^{\infty}x^n/s^{n+1}$ is absolutely and uniformly convergent for $x\in[-M,M]$, hence can be integrated term by term against $dF(x)$. The resulting Laurent series converges for all $|s|>M$, so the coefficients $(\mu_n)_{n=1}^{\infty}$ determine $G(s)$ for all $s\in\mathbb{C}\setminus[-M,M]$ and conversely. This is precisely what is meant by $G(s)$ being holomorphic near infinity.

(iii) In this case we have

$$\begin{aligned}\frac{h}{2i}\big(G(x_0-ih)-G(x_0+ih)\big)&=\frac{h}{2i}\int_{-M}^{M}\Big(\frac{1}{x_0-ih-x}-\frac{1}{x_0+ih-x}\Big)dF(x)\\&=\int_{-M}^{M}\frac{h^2dF(x)}{(x-x_0)^2+h^2}\end{aligned}$$

and we can take the limit as $h\to 0+$. We have

$$\begin{aligned}F(x_0)-F(x_0-)&\leq\int_{x_0-\delta}^{x_0+\delta}\frac{h^2dF(x)}{(x-x_0)^2+h^2}\\&\leq\int_{x_0-\delta}^{x_0+\delta}dF(x)\\&=F(x_0+\delta)-F(x_0-\delta).\end{aligned}$$

We can make the left-hand side and right-hand side as close as we please by taking $\delta>0$ sufficiently small, since F is right continuous. Having fixed $\delta>0$, we then take

$$\int_{-M}^{x_0-\delta}+\int_{x_0+\delta}^{M}\frac{h^2dF(x)}{(x-x_0)^2+h^2}\leq\frac{h^2}{\delta^2}\int_{-M}^{M}dF(x)=\frac{h^2}{\delta^2}$$

small by letting $h\to 0+$.

(iv) The method is similar to (iii), except the constants are different. Here

$$\begin{aligned}\frac{1}{2\pi i}\big(G(x_0-ih)-G(x_0+ih)\big)&=\frac{1}{2\pi i}\int_{-M}^{M}\Big(\frac{1}{x_0-ih-x}-\frac{1}{x_0+ih-x}\Big)dF(x)\\&=\frac{1}{\pi}\int_{-M}^{M}\frac{h\,dF(x)}{(x-x_0)^2+h^2}\qquad(h>0),\end{aligned}$$

which is the Poisson integral of dF, where $dF(x)=(dF/dx)dx$ on $(x_0-\varepsilon,x_0+\varepsilon)$. See [56] and [34]. Taking the limit as $h\to 0+$, we recover $F'(x_0)$, as follows. Given $\eta>0$, there exists $\varepsilon>\delta>0$ such that $(dF/dx)(x_0)-\eta<$

$(dF/dx)(x) < (dF/dx)(x_0) + \eta$ for all $x \in [x_0 - \delta, x_0 + \delta]$, so

$$\frac{((dF/dx)(x_0) - \eta)}{\pi} \int_{x_0-\delta}^{x_0+\delta} \frac{hdx}{(x-x_0)^2 + h^2} \leq \frac{1}{\pi} \int_{x_0-\delta}^{x_0+\delta} \frac{h(dF/dx)dx}{(x-x_0)^2 + h^2}$$
$$\leq \frac{((dF/dx)(x_0) + \eta)}{\pi} \int_{x_0-\delta}^{x_0+\delta} \frac{hdx}{(x-x_0)^2 + h^2}.$$

Now we fix this $\delta > 0$, and split the integral

$$1 = \int_{-\infty}^{\infty} \frac{h}{(x-x_0)^2 + h^2} \frac{dx}{\pi} = \int_{x_0-\delta}^{x_0+\delta} + \int_{-\infty}^{x_0-\delta} + \int_{x_0+\delta}^{\infty} \frac{hdx}{(x-x_0)^2 + h^2} \frac{dx}{\pi}, \tag{9.17}$$

in which the final two summands are equal and satisfy

$$\frac{1}{\pi} \int_{-\infty}^{x_0-\delta} \frac{hdx}{(x-x_0)^2 + h^2} = \frac{1}{\pi} \int_{x_0+\delta}^{\infty} \frac{hdx}{(x-x_0)^2 + h^2} \leq \frac{h}{\pi} \int_{x_0+\delta}^{\infty} \frac{dx}{(x-x_0)^2} = \frac{h}{\pi\delta}, \tag{9.18}$$

and as in (iii)

$$\frac{1}{\pi} \int_{-M}^{x_0-\delta} + \int_{x_0+\delta}^{M} \frac{hdF(x)}{(x-x_0)^2 + h^2} \leq \frac{h}{\delta^2\pi} \int_{-M}^{M} dF(x) = \frac{h}{\delta^2\pi}.$$

We let $h \to 0+$ and deduce that

$$\frac{1}{\pi} \int_{x_0-\delta}^{x_0+\delta} \frac{h(dF/dx)dx}{(x-x_0)^2 + h^2} \to \frac{dF}{dx}(x_0). \tag{9.19}$$

□

Example 9.3 Let $A \in M_{N\times N}(\mathbb{C})$ satisfy $A = A'$ and let $\mathbb{C}^N$ have an orthonormal basis of eigenvectors $(e_j)_{j=1}^N$ so that $Ae_j = \lambda_j e_j$ where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. Then a typical unit vector $X \in \mathbb{C}^N$ satisfies $X = \sum_{j=1}^N a_j e_j$ where $a_j = \langle X, e_j\rangle$, so $(sI - A)^{-1}X = \sum_{j=1}^N (s - \lambda_j)^{-1} a_j e_j$ and

$$G(s) = \big\langle (sI - A)^{-1}X, X\big\rangle = \sum_{j=1}^{N} (s - \lambda_j)^{-1} |a_j|^2 \tag{9.20}$$

is the Cauchy transform of the probability measure

$$\mu = \sum_{j=1}^{N} |a_j|^2 \delta_{\lambda_j}. \tag{9.21}$$

with cumulative distribution function $F(t) = \sum_{\{j:\lambda_j \le t\}} |a_j|^2$ Now $-G$ maps the upper half plane to itself since

$$-2i\Im G(s) = -G(s) + \overline{G(s)} = \sum_{j=1}^{N} \frac{s - \bar{s}}{|\lambda_j - s|^2} |a_j|^2 \tag{9.22}$$

is positive for $\Im s > 0$, and the jumps in F occur at the eigenvalues λ_j. The heights of the jumps depend upon X.

In particular, we can choose $X = \sum_{j=1}^{N} N^{-1/2} e_j$ so $a_j = 1/\sqrt{N}$ and

$$G(s) = \big\langle (sI - A)^{-1} X, X \big\rangle = \frac{1}{N} \sum_{j=1}^{N} (s - \lambda_j)^{-1} \tag{9.23}$$

is the Cauchy transform of the probability measure

$$\mu = \frac{1}{N} \sum_{j=1}^{n} \delta_{\lambda_j}. \tag{9.24}$$

Corollary 9.4 *Let $w : [-M, M] \to [0, \infty)$ be a continuous weight such that $\int_{-M}^{M} w(t)dt = 1$. Then the sequence $(\mu_n)_{n=0}^{\infty}$ of moments of w determines w.*

Proof By (i) and (ii) applied to $F(x) = \int_{-M}^{x} w(t)dt$, there is a Cauchy transform $G(s)$ determined by $(\mu_n)_{n=0}^{\infty}$, and we can apply (iv) of the Lemma 9.2 to $dF/dx = w(x)$ to recover w from $G(s)$. □

Corollary 9.5 (Lerch) *Suppose that f is a function of class (E) for which there exist $s_0, \ell > 0$ such that the Laplace transform F satisfies $F(s_0 + n\ell) = 0$ for $n = 0, 1, 2, \ldots$, so F is zero on some infinite real arithmetic progression. Then $f(t) = 0$ for all $t > 0$.*

Proof We have

$$\int_0^\infty e^{-n\ell t} e^{-s_0 t} f(t)dt = 0 \qquad (n = 0, 1, 2, \ldots) \tag{9.25}$$

so with the new variable $x = e^{-\ell t}$, we have

$$\int_0^1 x^n x^{s_0/\ell} f(-\ell^{-1} \log x) \ell^{-1} \frac{dx}{x} = 0 \qquad (n = 0, 1, \ldots) \tag{9.26}$$

where

$$\int_0^1 x^{s_0/\ell}|f(-\ell^{-1}\log x)|\ell^{-1}\frac{dx}{x} = \int_0^\infty e^{-s_0 t}|f(t)|dt < \infty. \tag{9.27}$$

We deduce that

$$G(s) = \int_0^1 \frac{\ell x^{-1+s_0/\ell} f(-\ell^{-1}\log x)}{s-x}dx \tag{9.28}$$

is a holomorphic function for $s \in \mathbb{C} \setminus [0,1]$ such that the Laurent series for $|s| > 0$, has coefficient that are all zero. Hence $G(s) = 0$ for all $s \in \mathbb{C} \setminus [0,1]$, and by considering $G(x+ih) - G(x-ih)$ as $h \to 0+$, we deduce that $x^{-1+s_0/\ell} f(-\ell^{-1}\log x) = 0$ for all $x \in (0,1)$, so $f(t) = 0$ for all $t > 0$. □

Example 9.6 The sine function $\sin z$ is not the Laplace transform of a bounded function. Note that $\sin z = 0$ for $z = n\pi$ for all $n \in \mathbb{Z}$.

Remark 9.7

(i) Cases (iii) and (iv) are useful in applications, but do not cover all eventualities of cumulative distribution functions. The details of the convergence in other cases are discussed in detail in [34], which presents a theorem of Fatou on the integral (9.10).

(ii) The connection between distribution functions F on $[0,1]$ and moment sequences $(\mu_n)_{n=0}^\infty$ is discussed in the Hausdorff moment problem; see [54]. In some applications, one can change $[-M, M]$ to $[0,1]$ by a simple linear scaling. Moment problems for distribution functions on $[0,\infty)$ or $(-\infty,\infty)$ are much more difficult than for $[0,1]$, and Corollary 9.4 is not always valid for weights on $[0,\infty)$.

(iii) We now have several tools for studying moment sequences.

$$\begin{array}{ccccc} \text{cdf} & F & \longrightarrow & (\mu_n)_{n=0}^\infty & \text{moments} \\ & \downarrow & \swarrow & & \downarrow \\ \text{Cauchy transform} & G & & \Gamma & \text{Hankel matrix} \end{array} \tag{9.29}$$

(iv) In Sect. 4.3 and Proposition 6.55 we considered a function that is holomorphic near ∞ and the contour integral

$$f(t) = \lim_{R\to\infty} \int_{\sigma - iR}^{\sigma + iR} e^{st} G(s)\frac{ds}{2\pi i} \qquad (t > 0). \tag{9.30}$$

In the context of the Lemma 9.2, we take G to be the Cauchy transform of w and obtain $f(t) = \int_{-M}^{M} e^{tx} w(x)\,dx$, which is the moment generating function of w. This appears in basic probability theory.

Given a nonempty open subset Ω of $\mathbb{C}$, we can consider holomorphic functions $\varphi_1, \varphi_2 : \Omega \to \Omega_1$ and then form their composition $\varphi = \varphi_1 \circ \varphi_2$, so that $\varphi : \Omega \to \Omega$ is also holomorphic. The simplest examples to consider are $\Omega = \mathbb{D}$, or the upper half plane $\{s : \Re s > 0\}$. The example we have in mind is $\Omega = \mathbb{C} \setminus [-M, M]$, the plane with the interval $[-M, M]$ cut out.

9.4 Herglotz Functions

Definition 9.8 Holomorphic functions that take the upper half plane to itself are known as Nevanlinna or Herglotz functions.

Exercise

(i) Show that the linear fractional transformations

$$\varphi(s) = \frac{as + b}{cs + d}$$

with $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$ are holomorphic on the upper half plane $\{s : \Im s > 0\}$ and take $\{s : \Im s > 0\}$ to itself. They also satisfy $\varphi(\bar{s}) = \overline{\varphi(s)}$. Find the inverse transformation of φ, and the composition $\varphi \circ \psi$ of two such transformations.

(ii) The sum of Herglotz functions is also Herglotz. Let $\alpha, \gamma \geq 0$, and $x, \beta \in \mathbb{R}$ and

$$\varphi(s) = \alpha s + \beta - \frac{\gamma}{s - x}.$$

Show that $\varphi(\bar{s}) = \overline{\varphi(s)}$; $\Im\varphi(s) \geq 0$ for all $\Im s > 0$, and $\varphi(s)$ is holomorphic except at x.

(iii) If $\varphi(s)$ is a Herglotz function, then $s \mapsto \varphi(is)/i$ takes $RHP \to RHP$; also $s \mapsto i\varphi(s/i)$ takes $LHP \to LHP$.

(iv) The logarithm function $\log s = \log|s| + i \arg s$ is a Herglotz function, which also follows from

$$\log s = \int_0^\infty \Big(\frac{1}{1 + t} - \frac{1}{s + t} \Big) dt \qquad (\Im s > 0), \tag{9.31}$$

which we encountered in (3.157).

(v) Another Herglotz function is $i\sqrt{z}$; see Exercise 9.6 and (9.121). By contrast, z^2 is not Herglotz.

(vi) Let $G(s)$ be a Green's function as in (9.10). It was shown there that the functions $-G(s)$ and $G(-s)$ are Herglotz functions. (Some authors define the Cauchy transform with $1/(x - s)$ to obtain Herglotz functions.) There is a converse to the Lemma 9.2 on Cauchy transforms, which shows that a large collection of Herglotz functions can be built out of examples (i) and (ii).

Proposition 9.9 (Evans) *Suppose that for some $M > 0$ the function φ satisfies*

(i) *φ is holomorphic on $\mathbb{C} \setminus [-M, M]$;*
(ii) *$\varphi(\bar{s}) = \overline{\varphi(s)}$;*
(iii) *$\Im\varphi(s) \geq 0$ for all $\Im s > 0$.*
Then there exist unique $\alpha \geq 0$, $\beta \in \mathbb{R}$ and $\gamma \geq 0$ and a cumulative distribution function F on $[-M, M]$ such that

$$\varphi(s) = \alpha s + \beta - \gamma \int_{-M}^{M} \frac{dF(x)}{s - x}. \tag{9.32}$$

(iv) *Also $\alpha > 0$ if and only if $\Re\varphi(s) \to \pm\infty$ as $\Re s \to \pm\infty$.*

Proof This is given in [34]. □

There are numerous variants and refinements of this result. Given functions φ_1, φ_2 satisfying (i)–(iv), we can form the composition $\varphi(s) = \varphi_1 \circ \varphi_2$, which also satisfies (i)–(iv), for some possibly different $M > 0$. The original $(\alpha_1, \beta_1, \gamma_1, F_1)$ and $(\alpha_2, \beta_2, \gamma_2, F_2)$ are composed to produce new data $(\alpha, \beta, \gamma, F)$. The functions have Laurent series beginning with

$$\begin{aligned} \varphi_1(s) &= \alpha_1 s + \beta_1 - \frac{\gamma_1}{s} + O\Big(\frac{1}{s^2}\Big) \qquad (s \to \infty), \\ \varphi_2(s) &= \alpha_2 s + \beta_2 - \frac{\gamma_2}{s} + O\Big(\frac{1}{s^2}\Big) \qquad (s \to \infty), \end{aligned} \tag{9.33}$$

Then by substitution we obtain

$$\varphi(s) = \alpha_1\alpha_2 s + \alpha_1\beta_2 + \beta_1 - \frac{\alpha_1\gamma_2 + \gamma_1/\alpha_2}{s} + O\Big(\frac{1}{s^2}\Big) \qquad (s \to \infty). \tag{9.34}$$

This determines (α, β, γ), and in favourable cases one can also find F via (iii) and (iv) of the Lemma 9.2. We can associate $(\alpha_1, \beta_1, \gamma_1)$ with the matrix

$$\begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\alpha_1 \end{bmatrix}, \tag{9.35}$$

and use the usual matrix multiplication.

9.5 Green's Functions

In the physics literature, the term Green's function can mean the average value of the resolvent $(sI - A)^{-1}$ with respect to sums of entries or some underlying probability measure. Here we consider some examples. (The term Green's function can also refer to the integral kernel of $(sI - A)^{-1}$, particularly when A is a differential operator. Also, different authors have diverse sign conventions.)

(i) Let A be an $N \times N$ complex matrix, with eigenvalues $\lambda_1, \dots, \lambda_N$ listed according to algebraic multiplicity. Then we define

$$G(s) = \frac{1}{N}\mathrm{trace}\big((sI - A)^{-1}\big). \tag{9.36}$$

We have the important formula

$$G(s) = \sum_{n=0}^{\infty} \frac{1}{s^{n+1}} \frac{1}{N}\mathrm{trace}\big(A^n\big) \qquad (|s| > \|A\|) \tag{9.37}$$

which shows that the Green's function is determined by the moments $\mathrm{trace}(A^n)/N$ and conversely.

The Green's function may be expressed as

$$G(s) = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{s - \lambda_j} = \frac{1}{N}\sum_{j=1}^{N} e_j^\top (sI - A)^{-1} e_j \tag{9.38}$$

where $(e_j)_{j=1}^N$ is the standard basis for $\mathbb{C}^{n\times 1}$. Then $G(s)$ is known as the Green's function or the Cauchy transform of the eigenvalue distribution of A; compare (9.23). Using the final formula, we can realize this as the transfer function of the linear system

$$\left(\begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & A \end{bmatrix}, \frac{1}{\sqrt{N}}\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}, \frac{1}{\sqrt{N}}\left[e_1^\top \; e_2^\top \; \dots \; e_N^\top\right], 0\right). \tag{9.39}$$

Here B and C are vectors of norm one with $C = B'$ so $CB = 1$.

Proposition 9.10 *The Green's function and the characteristic polynomial are related by*

$$\frac{1}{N}\mathit{trace}\big((sI - A)^{-1}\big) = \frac{1}{N}\frac{d}{ds}\log\det(sI - A). \tag{9.40}$$

Proof

(i) We have

$$\log\det(sI - A) = \operatorname{trace}\log(sI - A) \tag{9.41}$$

and we can differentiate both sides of this formula. The Green's function captures similar information to the characteristic polynomial, and is sometimes easier to work with.

(ii) In particular, let A be a $N \times N$ self-adjoint matrix, with real eigenvalues $\lambda_1, \dots, \lambda_N$ listed according to algebraic multiplicity. Then the normalized eigenvalue counting function is

$$F_N(x) = \frac{1}{N}\sharp\{j : \lambda_j \le x\} \tag{9.42}$$

which has a graph which resembles a staircase of total height one, which increases from left to right by steps that are of height some positive integer multiple of $1/N$. Then

$$\frac{1}{N}\operatorname{trace}\big((sI - A)^{-1}\big) = \int_{-\infty}^{\infty}\frac{dF_N(\lambda)}{s - \lambda} \qquad (\Im s > 0). \tag{9.43}$$

□

When the A come from a common family, we can consider convergence of this expression as $N \to \infty$.

Theorem 9.11 (Helly) *Let $(F_N)_{N=1}^{\infty}$ be a sequence of cumulative distribution functions on a bounded interval $[a, b]$. Then there exists a subsequence (F_{N_k}) and a cumulative distribution function F on $[a, b]$ such that:*

(i) $F_{N_k}(x) \to F(x)$ as $N_k \to \infty$ for all $x \in [a, b]$;

(ii) $\int_a^b g(x)dF_{N_k}(x) \to \int_a^b g(x)dF(x)$ as $N_k \to \infty$ for all continuous functions $g : [a, b] \to \mathbb{C}$;

(iii) the corresponding Cauchy integrals converge, so that

$$\int_a^b \frac{dF_{N_k}(x)}{s - x} \to \int_a^b \frac{dF(x)}{s - x} \qquad s \in \mathbb{C} \setminus [a, b]$$

uniformly for s in closed and bounded subsets of $\mathbb{C} \setminus [a, b]$ as $N_k \to \infty$.

Proof

(i) This is Helly's choice theorem, as in page 56 of [16].

(ii) This if Helly's convergence theorem, as in page 56 of [16]. It is important for this application that the F_N live on a common bounded interval $[a, b]$.

(iii) Let K be any closed and bounded subset of $\mathbb{C} \setminus [a, b]$. We can apply (ii) to the function $g(x) = 1/(s - x)$ and deduce that the Green's functions converge pointwise on K. It is easy to show that the Green's functions are uniformly bounded on K. Since the Green's functions are also holomorphic, we can apply Vitali's convergence theorem 5.21 from [56] to obtain uniform convergence on K.

Suppose that the limiting cumulative distribution function $F(x)$ is continuously differentiable with derivative $f(x) = dF/dx$. Then in physical applications, f is called the density of states. In the following example, we use convergence of the Green's functions to identify the density of states. □

Example 9.12 (Green's Function for a One-Dimensional Periodic Lattice) We consider N points arranged in a ring, so that each point interacts with its immediate neighbours, and no others. The interaction is described by the matrix $\Delta^{(1)} : \mathbb{C}^N \to \mathbb{C}^N$ defined for $X = (x_n)_{n=1}^N$ by the formula

$$\Delta^{(1)}(x_n) = (2x_n - x_{n+1} - x_{n-1})_{n=1}^N \tag{9.44}$$

with the convention that $x_0 = x_N$. For $N = 4$, we have

$$\Delta^{(1)} = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix} \tag{9.45}$$

which is tri-diagonal, apart from -1 in the top right and bottom left entries.

The periodicity of the ring is expressed via an arithmetic condition. A sequence $(\exp(in\theta))_{n=1}^N$ has $1 = \exp(iN\theta)$ if $N\theta = 2\pi k$ for some $k = 1, 2, \dots, N$, so we introduce

$$X_k = \Big(\exp \frac{2\pi i k n}{N}\Big)_{n=1}^N; \tag{9.46}$$

then

$$2\exp\frac{2\pi ikn}{N} - \exp\frac{2\pi ik(n+1)}{N} - \exp\frac{2\pi ik(n-1)}{N} = \Big(2 - 2\cos\frac{2\pi k}{N}\Big)\exp\frac{2\pi ikn}{N}, \tag{9.47}$$

so

$$\Delta^{(1)} X_k = \Big(2 - 2\cos\frac{2\pi k}{N}\Big) X_k. \tag{9.48}$$

We deduce that X_k are orthogonal eigenvectors that correspond to eigenvalues $\lambda_k = 2 - 2\cos(2\pi k/N) \in [0, 4]$ of a real symmetric matrix, and since there are N of them, we have an orthogonal basis $(X_k)_{k=1}^N$. They also give an orthogonal basis for the matrix $(\Delta^{(1)} - I)/2$, so we consider the corresponding Green's function

$$G_N^{(1)}(s) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{s-\cos(2\pi k/N)}. \tag{9.49}$$

We can interpret this as a Riemann sum for the integral of $1/(s-\cos\theta)$. In the limit as $N \to \infty$, we have

$$G^{(1)}(s) = \frac{1}{2\pi}\int_0^{2\pi}\frac{d\theta}{s-\cos\theta} \tag{9.50}$$

with uniform convergence on compact subsets of $\mathbb{C} \setminus [-1, 1]$. To evaluate this integral, one can use geometric series or contour integration, to obtain

$$G^{(1)}(s) = \frac{1}{\sqrt{s^2-1}}, \tag{9.51}$$

with the square root so chosen that $sG^{(1)}(s) \to 1$ as $|s| \to \infty$. Note that

$$G^{(1)}(s) = \frac{1}{\pi}\int_{-1}^{1}\frac{dx}{(s-x)\sqrt{1-x^2}}, \tag{9.52}$$

which is the Cauchy transform of a Chebyshev (or arcsine) random variable.

Example 9.13 (Green's Function for the Square Lattice) Now consider a square lattice made of N^2 points in the style of a chess board, with the interpretation that opposite edges of the square are identified to produce a torus. Each point interacts with its nearest neighbours and no others on the board, where the interaction is described by the matrix $\Delta^{(2)} : \mathbb{C}^{N^2} \to \mathbb{C}^{N^2}$

$$\Delta^{(2)}(x_{n,m})_{n,m=1}^N = (4x_{n,m} - x_{n+1,m} - x_{n-1,m} - x_{n,m+1} - x_{n,m-1})_{n,m=1}^N \tag{9.53}$$

where $x_{n,0} = x_{n,N}$ and $x_{0,m} = x_{N,m}$ for all $n, m = 1, \dots, N$. Then

$$X^{j,k} = \left(x^{j,k}\right)_{n,m=1}^N = \left(\exp\frac{2\pi ijn}{N}\exp\frac{2\pi ikn}{N}\right)_{n,m=1}^N \tag{9.54}$$

satisfies the periodicity condition for $j, k = 1, \dots, N$. We also have

$$\begin{aligned}\left\langle X^{j,k}, X^{\ell,p}\right\rangle &= N^2, \quad (j=\ell, k=p);\\ &= 0 \quad \text{else};\end{aligned}$$

and

$$\Delta^{(2)} X^{j,k} = \Big(4 - 2\cos\frac{2\pi j}{N} - 2\cos\frac{2\pi k}{N}\Big) X^{j,k} \qquad (j,k = 1,\ldots,N). \tag{9.55}$$

Note that the eigenvalues of $\Delta^{(1)}$ are λ_k for $k = 1,\ldots,N$, and the eigenvalues of $\Delta^{(2)}$ are the pairwise sums $\lambda_j + \lambda_k$ for $j,k = 1,\ldots,N$. This can also be seen from the identity

$$\Delta^{(2)} = \Delta^{(1)} \otimes I_N + I_N \otimes \Delta^{(1)} \tag{9.56}$$

and exercise Exercise 3.15.

The Green's function for $(\Delta^{(2)} - 4I)/2$ is

$$G_N^{(2)}(s) = \frac{1}{N^2} \sum_{j,k=1}^{N} \frac{1}{s - \cos(2\pi j/N) - \cos(2\pi k/N)}. \tag{9.57}$$

Taking the limit as $N \to \infty$, we obtain

$$G^{(2)}(s) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} \frac{d\theta d\phi}{s - \cos\theta - \cos\phi} \tag{9.58}$$

with uniform convergence for s in compact subsets of $\mathbb{C} \setminus [-2, 2]$.

To evaluate this, we need Jacobi's complete elliptic integral [41]

$$K(z) = \int_0^{\pi/2} \frac{d\psi}{\sqrt{1 - z^2 \sin^2 \psi}}. \tag{9.59}$$

We start the calculation as in the preceding example

$$\begin{aligned}
G^{(2)}(s) &= \frac{1}{4\pi^2} \int_0^{2\pi} \frac{1}{s - \cos\theta} \int_0^{2\pi} \frac{d\phi}{1 - (s - \cos\theta)^{-1} \cos\phi} d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} \frac{d\theta}{s - \cos\theta} \frac{1}{\sqrt{1 - (s - \cos\theta)^{-2}}} \\
&= \frac{1}{2\pi} \int_0^{2\pi} \frac{d\theta}{\sqrt{(s - \cos\theta)^2 - 1}} \\
&= \frac{1}{\pi} \int_0^{\pi} \frac{d\theta}{\sqrt{(s - \cos\theta - 1)(s - \cos\theta + 1)}}
\end{aligned}$$

we substitute $t = \tan(\theta/2)$ and $\cos\theta = (1-t^2)/(1+t^2)$ so

$$G^{(2)}(s) = \frac{2}{\pi}\int_0^\infty \frac{dt}{\sqrt{(s-2+st^2)(s+(s+2)t^2)}} \tag{9.60}$$

in which we substitute $t = u\sqrt{(s-2)/s}$ to get

$$G^{(2)}(s) = \frac{2}{\pi s}\int_0^\infty \frac{du}{\sqrt{(1+u^2)(1+(s^2-4)u^2/s^2)}} \tag{9.61}$$

in which we substitute $u = \tan\psi$ to get

$$G^{(2)}(s) = \frac{2}{\pi s}\int_0^{\pi/2} \frac{\sec^2\psi\, d\psi}{\sqrt{(1+\tan^2\psi)(1+(s^2-4)\tan^2\psi/s^2)}} \tag{9.62}$$

and we multiply numerator and denominator by $\cos^2\psi$ to obtain

$$\begin{aligned} G^{(2)}(s) &= \frac{2}{\pi s}\int_0^{\pi/2} \frac{d\psi}{\sqrt{(\cos^2\psi + (s^2-4)\sin^2\psi/s^2)}} \\ &= \frac{2}{\pi s}\int_0^{\pi/2} \frac{d\psi}{\sqrt{1-(2/s)^2\sin^2\psi}} \\ &= \frac{2}{\pi s}K\Big(\frac{2}{s}\Big). \end{aligned}$$

9.6 Random Diagonal Transformations

The following result is a version of the weak law of large numbers.

We consider

$$A = \begin{bmatrix} \xi_1 & 0 & \dots & 0 \\ 0 & \xi_2 & \ddots & \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \xi_N \end{bmatrix}, \quad (sI-A)^{-1} = \begin{bmatrix} (s-\xi_1)^{-1} & 0 & \dots & 0 \\ 0 & (s-\xi_2)^{-1} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & (s-\xi_N)^{-1} \end{bmatrix} \tag{9.63}$$

Proposition 9.14 *Suppose that ξ is a bounded random variable with distribution function F on $[-M, M]$. Suppose that A is a $N \times N$ real diagonal matrix with diagonal entries $\xi_1, \dots, \xi_N$ which are mutually independent random variables* *distributed as* ξ*. Then*

$$\frac{1}{N}trace\big((sI-A)^{-1}\big) \to \int_{-M}^{M} \frac{dF(t)}{s-t} \qquad (s \in \mathbb{C}\setminus[-M,M]) \tag{9.64}$$

in mean square and in probability as $N \to \infty$.

Proof The term

$$\frac{1}{N}\mathrm{trace}((sI-A)^{-1}) = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{s-\xi_j} \tag{9.65}$$

is a complex random variable for $s \in \mathbb{C}\setminus[-M,M]$. We aim to show that as $N \to \infty$, these random variables converge to a non-random quantity. The probability and distribution function are related by $\mathbb{P}[\xi \leq x] = F(x)$, where F is increasing from $F(-M) = 0$ to $F(M) = 1$ and F is right-continuous. Here $1/(s-\xi)$ has mean

$$G(s) = \mathbb{E}\frac{1}{s-\xi} = \int_{-M}^{M}\frac{dF(t)}{s-t} \qquad (s \in \mathbb{C}\setminus[-M,M]) \tag{9.66}$$

and

$$\mathbb{E}\frac{1}{|s-\xi|^2} = \int_{-M}^{M}\frac{dF(t)}{|s-t|^2} \qquad (s \in \mathbb{C}\setminus[-M,M]), \tag{9.67}$$

so the variance is

$$\begin{aligned}
\sigma^2 &= \mathbb{E}\frac{1}{|s-\xi|^2} - \Big|\mathbb{E}\frac{1}{s-\xi}\Big|^2 \\
&= \int_{-M}^{M}\int_{-M}^{M}\Big(\frac{1}{|s-t|^2} - \frac{1}{(s-t)(\bar{s}-u)}\Big)dF(t)dF(u) \\
&= \frac{1}{2}\int_{-M}^{M}\int_{-M}^{M}\Big|\frac{1}{s-t} - \frac{1}{s-u}\Big|^2 dF(t)dF(u) \qquad (s \in \mathbb{C}\setminus[-M,M]).
\end{aligned}$$

By independence, we deduce that

$$\frac{1}{N}\sum_{j=1}^{N}\frac{1}{s-\xi_j} \tag{9.68}$$

has mean μ and variance σ^2/N, so by Chebyshev's inequality

$$\frac{\sigma^2 t^2}{N}\mathbb{P}\Big[\Big|\frac{1}{N}\sum_{j=1}^{N}\frac{1}{s-\xi_j}-\mu\Big|>\frac{t\sigma}{\sqrt{N}}\Big]\leq \mathbb{E}\Big|\frac{1}{N}\sum_{j=1}^{N}\frac{1}{s-\xi_j}-\mu\Big|^2=\frac{\sigma^2}{N}\quad (t>0). \tag{9.69}$$

Hence

$$t^2\mathbb{P}\Big[\Big|\frac{1}{N}\mathrm{trace}\big((sI-A)^{-1}\big)-G(s)\Big|>\frac{t\sigma}{\sqrt{N}}\Big]\leq 1 \qquad (t>0). \tag{9.70}$$

The stated result follows from these estimates. □

9.7 Wigner Matrices

Definition 9.15 A Wigner matrix is a random matrix W that satisfies the following conditions.

(i) W is real and symmetric, with $W \in M_{N\times N}(\mathbb{R})$;
(ii) The entries $w_{j,k}$ for $j \geq k$ that lie on or above the leading diagonal are mutually independent random variables, with $w_{j,k} = w_{k,j}$ by (i);
(iii) $\mathbb{E}w_{j,k} = 0$ and $\mathbb{E}w_{j,k}^2 = 1$, so the mean is zero and the variance is one;
(iv) either there exists $M > 0$ such that $|w_{j,k}| \leq M$ for all j, k; or $w_{j,k}$ has a standard $N(0, 1)$ normal distribution with probability density function γ, for all j, k, in which case we call W a Gaussian Wigner matrix.

On account of condition (i), the eigenvalues of W are real numbers.

By condition (iii) we have $\mathbb{E}W = 0$, and by (iii) and (ii) we have $\mathbb{E}W^2 = NI$ since the (j, j)th diagonal entry is

$$\mathbb{E}\sum_{\ell=1}^{N} w_{j,\ell}w_{\ell,j} = \mathbb{E}\sum_{\ell=1}^{N} w_{j,\ell}^2 = N \tag{9.71}$$

whereas for $j \neq k$, the (j, k)th off diagonal entry is

$$\mathbb{E}\sum_{\ell=1}^{N} w_{j,\ell}w_{\ell,k} = \mathbb{E}\sum_{\ell=1}^{N} w_{j,\ell}w_{k,\ell} = \sum_{\ell=1}^{N}\mathbb{E}w_{j,\ell}\mathbb{E}w_{k,\ell} = 0. \tag{9.72}$$

We often use the random matrix $\frac{v}{\sqrt{N}}W$, which satisfies

$$\mathbb{E}\frac{v}{\sqrt{N}}W = 0, \quad \mathbb{E}\frac{v^2}{N}W^2 = v^2 I, \tag{9.73}$$

where the entries on the right-hand side do not depend upon N.

Exercise Recall that if ξ_1 and ξ_2 are independent normal random variables, where ξ_j has mean μ_j and variance σ_j^2, then $\xi_1 + \xi_2$ is also a normal random variable with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Let W_1 and W_2 be independent $N \times N$ Wigner matrices with normal entries.

(i) Show that $\sigma_1 W_1 + \sigma_2 W_2$ has the same distribution as $\sqrt{\sigma_1^2 + \sigma_2^2} W_1$.
(ii) Let U be an $N \times N$ orthogonal matrix. Show that $UW_1U^\top$ has the same distribution as W_1.

Normal random variable are widely used in statistical applications. The central limit theorem of probability theory gives conditions under which a sum of statistically independent random variables converges in distribution to a normal random variable. In the theory of random matrices, the semicircle distribution is a counterpart of the normal distribution, and has some analogous properties. The semicircle law was used by [Wigner, 1958] to model the distribution of energy levels of nucleons in an atomic nucleus with large atomic number. Recently, the semicircle law has been used in models of wireless communication, where one considers a large number of transmitting aerials which broadcast to a large number of receivers. Both the normal and semicircle laws satisfy special replication properties such as the addition rules which we establish in this book. See [58] for modern developments of the theory.

Example 9.16 (Semicircle Law) The semicircle law $S(a, r)$ with centre $a \in \mathbb{R}$ and radius $r > 0$ is the probability density function

$$w(x) = \frac{2}{\pi r^2}\sqrt{r^2 - (x-a)^2} \qquad (-r < x < r). \tag{9.74}$$

The Cauchy transform is

$$G(s) = \frac{2}{\pi r^2}\int_{a-r}^{a+r} \frac{\sqrt{r^2 - (x-a)^2}}{s - x} dx \tag{9.75}$$

in which we substitute $x = a + r\sin\theta$

$$G(s) = \frac{2}{\pi r^2}\int_{-\pi/2}^{\pi/2} \frac{r^2\cos^2\theta d\theta}{s - a - r\sin\theta} \tag{9.76}$$

and expand as a geometric series

$$G(s) = \frac{2}{\pi}\int_{-\pi/2}^{\pi/2}\sum_{k=0}^{\infty}\frac{r^k}{(s-a)^{k+1}}\sin^k\theta(1-\sin^2\theta)d\theta \tag{9.77}$$

in which only the even powers contribute, so we obtain

$$\begin{aligned} G(s) &= \sum_{k=0}^{\infty} \frac{2r^{2k}}{(s-a)^{2k+1}} \frac{1}{2k+1} \frac{(2k+1)(2k-1)\dots 3\cdot 1}{(2k+2)(2k)\dots 2} \\ &= \frac{2(s-a)}{r^2} \sum_{k=0}^{\infty} \frac{r^{2k+2}}{(s-a)^{2k+2}} (-1)^k \binom{1/2}{k+1} \\ &= \frac{2(s-a)}{r^2}\Big(1-\sqrt{1-\frac{r^2}{(s-a)^2}}\Big) \\ &= \frac{2}{r^2}\Big((s-a)-\sqrt{(s-a)^2-r^2}\Big). \end{aligned}$$

From this calculation, or otherwise, one checks that the mean is $\mu_1 = a$, and $\mu_2 = a^2 + r^2/4$. Writing $r = 2v$, we deduce that a random variable with distribution $S(a, 2v)$ therefore has mean a and variance v^2.

Theorem 9.17 (Wigner) *Let W be an $N \times N$ Wigner matrix as above. Then*

$$\mathbb{E}\frac{1}{N}trace\Big(\Big(sI - \frac{W}{\sqrt{N}}\Big)^{-1}\Big) \to \frac{1}{2\pi}\int_{-2}^{2} \frac{\sqrt{4-x^2}}{s-x}dx \qquad (s \in \mathbb{C} \setminus [-2, 2]) \tag{9.78}$$

as $N \to \infty$.

Proof See [63]. The proof involves a detailed analysis of $\mu_n = \mathbb{E}\frac{1}{N}\mathrm{trace}(W/\sqrt{N})^n$ as $N \to \infty$. We computed the cases $n = 1$ and $n = 2$; the higher odd powers $n = 2k + 1$ give $\mu_{2k+1} = 0$, whereas the higher even powers $n = 2k$ involve increasingly complicated sums of products of random variables. □

9.8 Pastur's Theorem

The following theorem combines the result about diagonal matrices in Proposition 9.14 with Wigner's semicircle law.

Theorem 9.18 (Pastur)

(i) Suppose that ξ is a bounded random variable with distribution function F on $[-M, M]$, and let G_F be the Cauchy transform of F.
(ii) Suppose that A is a $N \times N$ real diagonal matrix with diagonal entries $\xi_1, \dots, \xi_N$ which are mutually independent random variables distributed as ξ.
(iii) Let W be an $N \times N$ Wigner matrix as above, independent of A.

Then there exists a cumulative distribution function on some bounded interval $[-M_1, M_1]$ *with Cauchy transform G, such that*

$$\mathbb{E}\frac{1}{N}trace\Big(\Big(sI - A - \frac{vW}{\sqrt{N}}\Big)^{-1}\Big) \to G(s), \tag{9.79}$$

as $N \to \infty$ *and* $G(s)$ *satisfies the fixed point equation*

$$G(s) = G_F(s - v^2 G(s)). \tag{9.80}$$

Proof See Pastur [46]. □

We present two cases in which we can solve the fixed point equation, and a further case in which we solve a matrix variant of the fixed point equation.

(i) Let $\xi = \lambda_0$ be a constant, so $dF = \delta_{\lambda_0}$ and $G_F(s) = 1/(s - \lambda_0)$

Then the equation

$$G_v(s) = \frac{1}{s - v^2 G_v(s) - \lambda_0} \tag{9.81}$$

gives the quadratic

$$v^2 G(s)^2 + (\lambda_0 - s)G(s) + 1 = 0 \tag{9.82}$$

with solution

$$G_v(s) = \frac{s - \lambda_0 - \sqrt{(s - \lambda_0)^2 - 4v^2}}{2v^2}. \tag{9.83}$$

This is the Cauchy transform of the $S(\lambda_0, 2v)$ distribution by Example 9.16.

9.9 May–Wigner Model

Consider N species of animals living on an isolated island. The population x_j of species j gives the jth entry of a state vector $X \in \mathbb{R}^{N \times 1}$ at time $t > 0$. The environmental conditions on the island encourage proportional growth or decay of all the populations through time at a common constant rate λ_0. Additionally, there are symmetrical interactions between species j and k which may be mutually disadvantageous, such as red and grey squirrels competing for the same food supply, or mutually advantageous, as in sheep and sheep tick. The rate of interaction between species j and k is represented by a random variable w_{jk} with $w_{kj} = w_{jk}$ where the random variables $w_{j,k}$ for $j < k$ are mutually independent and identically distributed. In particular, we can consider the interaction matrix to be a Wigner matrix as above, which gives the May–Wigner model. See [21]

For a $N \times N$ real diagonal matrix $\lambda_0 I$ and a constant $v > 0$, we consider the differential equation

$$\frac{d}{dt}X = \Big(\lambda_0 I + \frac{v}{\sqrt{N}}W\Big)X. \tag{9.84}$$

The associated Green's function is

$$G_N(s) = \mathbb{E}\bigg(\text{trace}\frac{1}{N}\Big(sI - \lambda_0 I - \frac{v}{\sqrt{N}}W\Big)^{-1}\bigg). \tag{9.85}$$

By Wigner's theorem and Pastur's theorem [46], the distribution of eigenvalues converges as $N \to \infty$ to the $S(\lambda_0, 2v)$ distribution, which is supported on $[\lambda_0 - 2v, \lambda_0 + 2v]$. There are therefore two cases concerning the stability of this system for large $N > 0$, depending on whether this interval intersects with $(0, \infty)$.

(i) If $\lambda_0 + 2v < 0$, then most solutions of the differential equation are bounded.
(ii) Whereas if $\lambda_0 + 2v > 0$, then there are unbounded solutions with positive probability.

The conclusion is that in case (ii), one population grows unboundedly large.

This model can be refined to address more realistic assumptions regarding the interaction of species. For instance, one can consider predator-prey relationships in which w_{jk} and w_{kj} have opposite sign, or models in which each species interacts with only a bounded number of other species. There is also the question of whether the solution to the differential equation produces credible values for a population model, as in the next exercise.

Exercise Let $C = \big[1 \dots 1\big] \in \mathbb{R}^{1\times N}$, let $A = [a_{jk}]_{j,k=1}^N \in M_{N\times N}(\mathbb{R})$ and for $X_0 \in \mathbb{R}^{N\times 1}$ consider the initial value problem

$$\frac{dX}{dt} = AX, \quad X(0) = X_0, \tag{9.86}$$

where X represents populations from various species as in the May-Wigner model.

(i) Show that CX gives the total population of all species at time t.
(ii) Show that CX is constant with respect to time for all X_0, if and only if $CA = 0$. Express this as a condition on the entries $[a_{jk}]$.
(iii) Suppose that X has nonnegative entries for all X_0 that have nonnegative entries and all $t > 0$. Show that $a_{jk} \geq 0$ for all $j \neq k$.

(iv) Let $f_{jk}(t) = \langle \exp(tA)e_k, e_j \rangle$. Show that

$$\frac{df_{jk}}{dt} = a_{jj} f_{jk} + \sum_{\ell:\ell \neq j} a_{j\ell} f_{\ell,k} \qquad (j,k = 1,\dots,N) \tag{9.87}$$

and deduce the integral equation

$$f_{jk}(t) = e^{a_{jj}t}\delta_{jk} + \int_0^t e^{a_{jj}(t-u)} \sum_{\ell:\ell\neq j} a_{j\ell} f_{\ell,k}(u)du. \tag{9.88}$$

See [48] for discussion of this integral equation and [18] for the associated semigroups. Semigroups that respect positivity conditions arise in probability theory and were studied by Feller, Kolmogorov, Markov and others.

9.10 Semicircle Addition Law

With F the semicircle $S(0, 2\alpha)$ law, the Cauchy transform of F is

$$G_{\alpha^2}(z) = \frac{z - \sqrt{z^2 - 4\alpha^2}}{2\alpha^2} \tag{9.89}$$

so that with $z = s - v^2 G$, the fixed point equation $G = G_{\alpha^2}(s - v^2 G)$ gives

$$G = \frac{s - v^2 G - \sqrt{(s - v^2 G)^2 - 4\alpha^2}}{2\alpha^2}, \tag{9.90}$$

which gives, when we isolate and square the root,

$$\big((2\alpha^2 + v^2)G - s\big)^2 = (s - v^2 G)^2 - 4\alpha^2 \tag{9.91}$$

which reduces to a quadratic equation

$$(\alpha^2 + v^2)G^2 - Gs + 1 = 0 \tag{9.92}$$

with solution

$$G = G_{\alpha^2 + v^2}(s) = \frac{s - \sqrt{s^2 - 4(\alpha^2 + v^2)}}{2(\alpha^2 + v^2)} \tag{9.93}$$

Pastur's theorem shows that if a random diagonal matrix H_0 has entries that satisfy the $S(0, 2\alpha)$ law is perturbed by an independent Wigner matrix $vW/\sqrt{N}$, then $H_0 + vW/\sqrt{N}$ has eigenvalues that satisfy a $S(0, 2\sqrt{\alpha^2 + v^2})$ law. The addition rule for semicircular distributions is an instance of the composition law for Herglotz functions from Sect. 9.4.

We observe that

$$\frac{s + \sqrt{s^2 - 4v^2}}{2} + v^2 \frac{s - \sqrt{s^2 - 4v^2}}{2v^2} = s \tag{9.94}$$

so

$$\frac{1}{G_{v^2}(s)} = s - v^2 G_{v^2}(s) \tag{9.95}$$

is a map of the above form, corresponding to $(1, 0, v^2, S(0, 2v))$. With $K_t(s) = s + t/s$, we have $K_{v^2}(1/G_{v^2}(s)) = s$ and

$$K_t(K_u(s)) = s + \frac{u}{s} + \frac{t}{s}\Big(1 + \sum_{n=1}^{\infty} \frac{(-u)^n}{s^{2n}}\Big). \tag{9.96}$$

For this reason, we add the variances α^2 and v^2 to produce $\alpha^2 + v^2$.

9.11 Matrix Version of Pastur's Fixed Point Equation

See [31]. Let $\phi : M_{2\times 2}(\mathbb{C}) \to \mathbb{C} : A \mapsto \text{trace}(A)$ and extend to $\Phi : M_{2\times 2}(\mathbb{C}) \otimes M_{2\times 2}(\mathbb{C}) \to M_{2\times 2}(\mathbb{C}) : [A_{j,k}] \mapsto [\phi(A_{j,k})]$. Then we introduce

$$sI - D - \Sigma = \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} m & 0 \\ 0 & -m \end{bmatrix} - \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tag{9.97}$$

with inverse

$$\Big(sI - D - \Sigma\Big)^{-1} = \begin{bmatrix} (s - m - \Sigma_1)^{-1} & 0 \\ 0 & (s + m - \Sigma_2)^{-1} \end{bmatrix}; \tag{9.98}$$

then with

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \in M_{2\times 2} \otimes M_{2\times 2}, \tag{9.99}$$

we consider the fixed point equation

$$\Phi\Big(\big((sI - D - \Sigma)^{-1} \otimes I_2\big)E\Big) = \Sigma \tag{9.100}$$

so the diagonal entries are

$$\frac{1}{s+m-\Sigma_2} = \Sigma_1$$
$$\frac{1}{s-m-\Sigma_1} = \Sigma_2 \tag{9.101}$$

giving a pair of quadratic equations

$$1 = \Sigma_1(s+m) - \Sigma_1\Sigma_2$$
$$1 = \Sigma_2(s-m) - \Sigma_1\Sigma_2. \tag{9.102}$$

We find that

$$\begin{aligned} G(s) &= \frac{1}{2}\Big(\Sigma_1 + \Sigma_2\Big) \\ &= \frac{s}{2}\left(1 - \frac{\sqrt{s^2-m^2-4}}{\sqrt{s^2-m^2}}\right). \end{aligned}$$

Proposition 9.19 *This $G(s)$ is the Cauchy transform of the weight function*

$$w(x) = \frac{1}{2\pi}\frac{|x|\sqrt{x^2-m^2}}{\sqrt{4+m^2-x^2}} \qquad (x \in (-\sqrt{4+m^2}, -m) \cup (m, \sqrt{m^2+4})).$$

Proof To prove this, we observe that the Cauchy transform is

$$\begin{aligned} G(s) &= \int_{-\sqrt{4+m^2}}^{-m} + \int_m^{\sqrt{m^2+4}} \frac{1}{s-x}\frac{|x|\sqrt{x^2-m^2}}{\sqrt{4+m^2-x^2}}\frac{dx}{2\pi} \\ &= \int_m^{\sqrt{m^2+4}} \frac{2s}{s^2-x^2}\frac{|x|\sqrt{x^2-m^2}}{\sqrt{4+m^2-x^2}}\frac{dx}{2\pi} \\ &= \int_{m^2}^{m^2+4} \frac{s}{s^2-u}\frac{\sqrt{u-m^2}}{\sqrt{4+m^2-u}}\frac{du}{2\pi} \end{aligned}$$

so with $u = m^2 + 4\sin^2\theta$, we have

$$\begin{aligned} G(s) &= \frac{4s}{\pi}\int_0^{\pi/2} \frac{\sin^2\theta}{s^2-m^2-4\sin^2\theta}d\theta \\ &= \frac{4s}{\pi}\sum_{k=0}^{\infty}\frac{4^k}{(s^2-m^2)^{k+1}}\int_0^{\pi/2}\sin^{2k+2}\theta\, d\theta \end{aligned}$$

$$= \frac{s}{2}\sum_{k=0}^{\infty}\frac{(2k+1)(2k-1)\dots 3.1}{(2k+2)2k\dots 4.2}\frac{4^{k+1}}{(s^2-m^2)^{k+1}}$$

$$= \frac{s}{2}\sum_{k=0}^{\infty}\binom{1/2}{k+1}(-1)^k\frac{4^{k+1}}{(s^2-m^2)^{k+1}}$$

$$= \frac{s}{2}\Big(1-\Big(1-\frac{4}{(s^2-m^2)}\Big)^{1/2}\Big)$$

$$= \frac{s}{2}\Big(1-\sqrt{\frac{s^2-m^2-4}{s^2-m^2}}\Big).$$

□

Let $G(s)$ be a Green's function, as above. Then the Dyson–Schwinger equation

$$G(z) = \frac{1}{z-\Sigma(z)} \tag{9.103}$$

introduces the self-energy $\Sigma(z) = z - 1/G(z)$. Voiculescu [58] considered the R-transform

$$R(z) = G^{-1}(z) - 1/z, \tag{9.104}$$

in which G^{-1} is the functional inverse so $G \circ G^{-1}(z) = z$. Substituting $z = G(s)$, we obtain

$$R(G(s)) = s - 1/G(s) = \Sigma(s) \tag{9.105}$$

and obtain an alternative formula for the self-energy.

9.12 Rank One Perturbations on Green's Functions

In Sect. 7.5, we considered the effect of adding a rank one operator to A with a view to making an almost stable system become stable. In this section, we consider a related question concerning the Green's functions. The following is based upon Appendix 1 from [17]. Let $A_0 = A_0' \in M_{n\times n}(\mathbb{C})$, and let $B = C' \in \mathbb{C}^{n\times 1}$. We introduce

$$A_t = A_0 + tBC \tag{9.106}$$

so that (A_t) gives a one-parameter family of self-adjoint matrices, obtained by adding scalar multiples of a rank one operator BC to A_0.

Proposition 9.20 *Let* $f_t(s) = C(sI - A_t)^{-1}B$. *Then*

$$trace\Big((sI - A_0)^{-1} - (sI - A_t)^{-1}\Big) = \frac{d}{ds}\log(1 - tf_0(s)). \tag{9.107}$$

Proof We start with the formula

$$sI - A_t = sI - A_0 - tBC, \tag{9.108}$$

and multiply by $(sI - A_0)^{-1}$ on the left and $(sI - A_t)^{-1}$ on the right. This gives

$$(sI - A_0)^{-1} = (sI - A_t)^{-1} - t(sI - A_0)^{-1}BC(sI - A_t)^{-1}; \tag{9.109}$$

then we multiply by C and the left and B on the right to get

$$f_0(s) = f_t(s) - tf_0(s)f_t(s), \tag{9.110}$$

which we rearrange to

$$f_t(s) = \frac{f_0(s)}{1 - tf_0(s)}. \tag{9.111}$$

We return to (9.109), and multiply by B on the right to

$$(sI - A_0)^{-1}B = (sI - A_t)^{-1}B - tf_t(s)(sI - A_0)^{-1}B; \tag{9.112}$$

changing the subject of the formula gives

$$(sI - A_t)^{-1}B = (sI - A_0)^{-1}B + tf_t(s)(sI - A_0)^{-1}B \tag{9.113}$$

and we use the previous formula for $f_t(s)$ to give

$$(sI - A_t)^{-1}B = (1 - tf_0(s))^{-1}(sI - A_0)^{-1}B. \tag{9.114}$$

We recall that $C = B'$, and observe that $\overline{f_0(\bar{s})} = f_0(s)$, so when we take adjoints of (9.114) and replace s by $\bar{s}$, we have

$$C(sI - A_t)^{-1} = (1 - tf_0(s))^{-1}C(sI - A_0)^{-1}. \tag{9.115}$$

This gives

$$(sI - A_0)^{-1} = (sI - A_t)^{-1} - t(1 - tf_0(s))^{-1}(sI - A_0)^{-1}BC(sI - A_0)^{-1}, \tag{9.116}$$

and when we take the trace, we obtain

$$\begin{aligned}\mathrm{trace}\Big((sI - A_0)^{-1} - (sI - A_t)^{-1}\Big) &= -t(1 - tf_0(s))^{-1}\mathrm{trace}(sI - A_0)^{-1}BC(sI - A_0)^{-1}\\ &= t(1 - tf_0(s))^{-1}\mathrm{trace}C(sI - A_0)^{-2}B\\ &= \frac{-t(df_0/ds)}{1 - tf_0(s)} = \frac{d}{ds}\log(1 - tf_0(s)).\end{aligned}$$

□

9.13 Exercises

Exercise 9.1 For a distribution F on $[-M, M]$, the corresponding logarithmic potential is defined by

$$L(z) = \int_{[-M,M]} \log|z - t|\, dF(t) \qquad (z \in \mathbb{C} \setminus [-M, M]). \tag{9.117}$$

(i) Show that

$$\Big(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\Big)L(x+iy) = G(x+iy) \qquad (x+iy \in \mathbb{C}\setminus[-M, M]). \tag{9.118}$$

(ii) Deduce that

$$\Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\Big)L(x+iy) = \Big(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\Big)G(x+iy) = 0 \qquad (x+iy \in \mathbb{C}\setminus[-M, M]). \tag{9.119}$$

Exercise 9.2 By expanding the exponential as a power series in ixt and using a trigonometric substitution, show that

$$\frac{2}{\pi}\int_{-1}^{1} e^{ixt}\sqrt{1 - x^2}dx = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{2^{2n}(n+1)!n!} \qquad (t \in \mathbb{R}). \tag{9.120}$$

Feller identified the sum on the right-hand side with $2J_1(t)/t$. See Wigner [63] and (6.115). Alternatively, one can use the dog-bone contour of Exercise 4.13 to invert the Laplace transform.

Exercise 9.3 Let $J_0(s) = \int_0^{\pi} \cos(s\cos\theta)d\theta/\pi$ and

$$\phi(t) = \int_{-a}^{a} e^{itx}\sqrt{a^2 - x^2}\frac{dx}{\pi a^2}.$$

Show that

$$\phi(t) = -\Big(J_0(at) + \frac{d^2 J_0}{ds^2}(at)\Big).$$

Exercise 9.4 Let $w : [a, b] \to [0, \infty)$ be a continuous weight let $(P_n(t))_{n=0}^{\infty}$ be the sequence of monic orthogonal polynomials for w.

(i) Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function such that

$$\int_a^b f(t)P_n(t)w(t)dt = 0 \qquad (n = 0, 1, 2, \dots).$$

(ii) Show that

$$\int_a^b t^n f(t)w(t)dt = 0 \qquad (n = 0, 1, 2, \dots),$$

and deduce that

$$\int_a^b \frac{f(t)w(t)}{s-t}dt = 0 \qquad (s \in \mathbb{C} \setminus [a, b]).$$

(iii) By considering the proof of the Lemma 9.2, show that $f(t)w(t) = 0$ for all $t \in [a, b]$.

Exercise 9.5

(i) Let $A \in M_{1\times n}(\mathbb{C})$, $B \in M_{n\times 1}(\mathbb{C})$, $C \in M_{1\times n}(\mathbb{C})$. Show that

$$\det(sI - A - tBC) = \det(sI - A) - tC\operatorname{adj}(sI - A)B,$$

so the characteristic equation for $A + tBC$ is a polynomial equation of degree n in s with coefficients that are of degree at most one as functions of t.

(ii) Show how to solve this in the case of $n = 2$, and compare with (6.121) and (6.119).

Exercise 9.6

(i) Show that $i\sqrt{z}$ defines a Herglotz function.

(ii) By substituting $E = k^2$ and using contour integration, show that

$$\frac{1}{\pi}\int_0^\infty \Big(\frac{1}{E-z} - \frac{E}{1+E^2}\Big)\sqrt{E}dE = i\sqrt{z} + \frac{1}{\sqrt{2}}. \tag{9.121}$$

This formula can be used to define $i\sqrt{A}$ when A is a matrix with eigenvalues in the upper half plane.

Chapter 10
Hilbert Spaces

In previous chapters, we have used the state space $\mathbb{C}^N$ where N is finite but possibly large. The next step in the development of the theory is to take the state space to be infinite-dimensional. Amongst many possible options, the most suitable type of space to use is Hilbert space. The essential feature of Hilbert space is that it comes equipped with an inner product that replicates the properties of the scalar product on Euclidean space. Any complex Hilbert space has a complete orthonormal basis and one can use this to introduce a system of coordinates for the Hilbert space. In this chapter, we look at the basic models of Hilbert space and operators on them. Methods of Hilbert space theory fit well with complex analysis, and allow us to use spaces of holomorphic functions as models for linear systems. The main models for Hilbert space are Hilbert sequence space ℓ^2 of square summable complex sequences as in Sect. 10.1, Hardy space of square summable power series as in Sect. 10.2, and Hardy space on the left half-plane, as in Sect. 10.5. The crucially important operator is the shift, which we discuss in Sects. 10.3 and 10.4 along with its interpretation for discrete time linear systems. We use the Laguerre polynomials from Chap. 8 and the Laplace transform from Chap. 4 to study the Hardy space on the left half-plane, and refine previous results about the Laplace transform and its inverse. The main result is the Paley–Wiener theorem, which has to significant applications to signal processing. We consider sampling of band limited functions.

10.1 Hilbert Sequence Space

Definition 10.1 (Hilbert Sequence Space) Let

$$\ell^2 = \{(u_n)_{n=0}^{\infty} : u_n \in \mathbb{C}; \sum_{n=0}^{\infty} |u_n|^2 \text{ converges}\}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_10

be the space of square summable complex sequences. Then ℓ^2 forms a vector space under the coordinatewise operations

$$\lambda(u_n)_{n=0}^{\infty} = (\lambda u_n)_{n=0}^{\infty} \tag{10.1}$$

$$(u_n)_{n=0}^{\infty} + (v_n)_{n=0}^{\infty} = (u_n + v_n)_{n=0}^{\infty} \tag{10.2}$$

since $|u_n + v_n|^2 \leq 2|u_n|^2 + 2|v_n|^2$ makes the latter series square summable. We can define an inner product by

$$\left\langle (u_n)_{n=0}^{\infty}, (v_n)_{n=0}^{\infty} \right\rangle = \sum_{n=0}^{\infty} u_n \bar{v}_n. \tag{10.3}$$

Then $\langle u, u\rangle = \sum_{n=0}^{\infty} |u_n|^2$, so $\langle u, u\rangle \geq 0$, with $\langle u, u\rangle = 0 \Rightarrow u = 0$;

$$\langle v + u, w\rangle = \langle v, w\rangle + \langle u, w\rangle \tag{10.4}$$

$$\langle \lambda v, w\rangle = \lambda\langle v, w\rangle, \quad \overline{\langle v, w\rangle} = \langle w, v\rangle. \tag{10.5}$$

Then we define a norm by $\|u\| = \langle u, u\rangle^{1/2}$. The Cauchy–Schwarz inequality as in (2.20) gives

$$\left|\langle x, y\rangle\right| \leq \|x\|\|y\| \qquad (x, y \in H) \tag{10.6}$$

with equality if and only if x and y are parallel.

Definition 10.2 (Inner Product) An complex inner product space H is a complex vector space with vector addition and scalar multiplication satisfying the usual rules such as

$$\lambda(x + y) = \lambda x + \lambda y, (\lambda(\mu x)) = (\lambda\mu)x, \qquad 1x = x \tag{10.7}$$

There is an inner product $\langle \cdot, \cdot\rangle : H \times H \to \mathbb{C}$ such that

$$\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle, \quad \lambda\langle x, z\rangle = \langle \lambda x, z\rangle; \tag{10.8}$$

$$\overline{\langle x, z\rangle} = \langle z, x\rangle \qquad (x, y, z \in H; \lambda \in \mathbb{C}); \tag{10.9}$$

$$\langle x, x\rangle > 0 \qquad (x \in H, x \neq 0). \tag{10.10}$$

We introduce the norm by $\|x\| = \langle x, x\rangle^{1/2}$.

The Cauchy–Schwarz inequality (2.20) and triangle inequality hold for H just as for $\mathbb{C}^n$ with the same proof. We also have the parallelogram law

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \qquad (x, y \in H), \tag{10.11}$$

as the reader can check by multiplying out the terms on the left-hand side.

Definition 10.3 (Hilbert Space)

(i) A complex inner product space H is said to be complete if for all $(s_n)_{n=1}^{\infty}$ in H such that $\|s_n - s_m\| \to 0$ as $n, m \to \infty$, there exists $s \in H$ such that $\|s_n - s\| \to 0$ as $n \to \infty$.
(ii) A Hilbert space is a complete inner product space.
(iii) We further suppose that H is separable, so there exists a countable subset $(x_n)_{n=1}^{\infty}$ of H such that for all $x \in H$ and $\varepsilon > 0$, there exists n such that $\|x - x_n\| < \varepsilon$.

Example 10.4 Hilbert sequence space ℓ^2 is a Hilbert space for the above inner product (10.3).

Example 10.5 (Notions of Convergence in Hilbert Space)

(i) If $(x_n)_{n=1}^{\infty}$ is a sequence in H and $x \in H$ is such that $\|x_n - x\| \to 0$ as $n \to \infty$, then we say that x_n converges to x in norm and write $x_n \to x$ as $n \to \infty$. One checks that $\langle x_n, z\rangle \to \langle x, z\rangle$ for all $z \in H$.
(ii) If $(x_n)_{n=1}^{\infty}$ is a sequence in H and $x \in H$ is such that $\langle x_n, z\rangle \to \langle x, z\rangle$ for all $z \in H$, then we say that x_n converges weakly to x. By (i), convergence in norm implies weak convergence. The converse is false, but there is a remarkable connection between the notions of convergence.
(iii) Suppose that $x_n, z \in H$ have $\|x_n\| = \|z\| = 1$ for $n = 1, 2, \dots,$ and $\langle x_n, z\rangle \to 1$ as $n \to \infty$. Then $\|x_n - z\| \to 0$ as $n \to \infty$.

To see this, we use the triangle inequality to show $\|x_n + x_m\| \leq 2$, and we observe that $\langle x_n + x_m, z\rangle \to 2$ as $n, m \to \infty$. From the Cauchy–Schwarz inequality, it follows that $\|x_n + x_m\| \to 2$ as $n \to \infty$; then the parallelogram law gives

$$\|x_n - x_m\|^2 = 2\|x_n\|^2 + 2\|x_m\|^2 - \|x_n + x_m\|^2 \to 0 \qquad (n, m \to \infty). \tag{10.12}$$

We deduce that there exists $x \in H$ such that $\|x_n - x\| \to 0$ as $n \to \infty$; hence $\langle x, z\rangle = \lim_{n\to\infty}\langle x_n, z\rangle = 1$, and $\|x\| = \lim_{n\to\infty}\|x_n\| = 1$. Since $\|z\| = 1$ we deduce that $x = z$, so $\|x_n - z\| \to 0$ as $n \to \infty$.

Definition 10.6 (Orthonormal Sequence)

(i) Let $(e_n)_{n=1}^{\infty}$ be a sequence in H such that $\langle e_n, e_m\rangle = \delta_{n,m}$. Then $(e_n)_{n=1}^{\infty}$ is said to be an orthonormal sequence.
(ii) Suppose further that (a_n) is a complex sequence. Then we call $\sum_{n=1}^{\infty} a_n e_n$ an orthonormal series, and $s_n = \sum_{k=1}^{n} a_k e_k$ the nth partial sum.
(iii) Say that $(e_j)_{j=1}^{\infty}$ is a complete orthonormal basis of H if $\langle e_j, e_k\rangle = \delta_{j,k}$ and for every $x \in H$ there exists $(a_j)_{j=1}^{\infty} \in \ell^2$ such that $x = \sum_{j=1}^{\infty} a_j e_j$.

Using the Gram-Schmidt process [51, page 258], one can easily construct orthonormal sequences in a Hilbert space. Also, one can show that any separable Hilbert space has a complete orthonormal basis see [51, page 255]. In particular, we have already seen that orthogonal polynomials can be used to form orthonormal bases in $L^2(w)$, for suitable weights, and the notion of completeness and the notion of completeness for the orthogonal polynomials coincides with the notion of completeness of the orthonormal sequence in $L^2(w)$. Proving that a given orthonormal sequence is complete can be difficult and in this book we require classical results such as the Fourier uniqueness theorem as in (4.94) to prove the Laguerre system is complete.

Example 10.7 (Standard ONB) The prototype is the standard orthonormal basis (e_n) of ℓ^2 where e_n is the standard unit vector with 1 in place n and zeros elsewhere. We have $(a_n)_{n=1}^{\infty} = \sum_{n=1}^{\infty} a_n e_n$.

Proposition 10.8 (Riesz–Fischer) *Let $\sum_{n=1}^{\infty} a_n e_n$ be an orthonormal series.*

(i) If $\sum_{n=1}^{\infty} |a_n|^2$ converges then the series $s = \sum_{n=1}^{\infty} a_n e_n$ converges in the sense that there exists $s \in H$ such that $\|s - s_n\| \to 0$ as $n \to \infty$ and $\|s\|^2 = \sum_{n=1}^{\infty} |a_n|^2$ where $s_n = \sum_{j=1}^{n} a_j e_j$.
(ii) In the case (i), the sum s is the same whenever the terms are reordered or regrouped.
(iii) If $\sum_{n=1}^{\infty} |a_n|^2$ diverges, then $\|s_n\| \to \infty$ as $n \to \infty$.

Proof See [65]. □

Definition 10.9

(i) A map $T : \ell^2 \to \ell^2$ is linear if

$$T(\lambda x + \mu y) = \lambda T(x) + \mu T(y) \qquad (x, y \in \ell^2; \lambda, \mu \in \mathbb{C}). \tag{10.13}$$

(ii) A linear map is bounded if there exists $M > 0$ such that $\|Tx\| \le M\|x\|$ for all $x \in \ell^2$. This is equivalent to the notion of continuity for a linear operator, as discussed in [51, page 219] and [65, page 60].
(iii) The operator norm of a bounded linear operator T is $\|T\| = \sup\{\|Tx\| : x \in \ell^2 : \|x\| \le 1\}$. This definition is consistent with the definition of the norm of a matrix in Definition 2.18.
(iv) A linear map $V : \ell^2 \to \ell^2$ is an isometry if $\|Vx\| = \|x\|$ for all $x \in H$.

10.2 Hardy Space on the Disc

The space of power series on the unit disc with square summable Taylor coefficients gives a Hilbert space with important applications. In this section, we show how this Hilbert space can be described in terms of the sequence of coefficients and equivalently in terms of holomorphic functions on the unit disc, when it is usually known as Hardy space. We let H^2 be the space of holomorphic functions u on the unit disc $\mathbb{D}(0, 1) = \{z \in \mathbb{C} : |z| < 1\}$ such that

$$\|u\|_{H^2}^2 = \sup_{0<r<1} \int_0^{2\pi} |u(re^{i\theta})|^2 \frac{d\theta}{2\pi} < \infty. \tag{10.14}$$

For $f, g \in H^2$, we write

$$\langle f, g\rangle = \int_{C(0,1)} f(z)\overline{g(z)}\frac{dz}{2\pi iz} = \int_0^{2\pi} f(e^{i\theta})\overline{g(e^{i\theta})}\frac{d\theta}{2\pi}, \tag{10.15}$$

which gives the inner product on H^2. The precise meaning of the final integral in (10.15) will become clear via Lemma 10.10.

Lemma 10.10 *For $w \in \mathbb{D}$ let $k_w(z) = 1/(1 - \bar{w}z)$. Then $k_w \in H^2(\mathbb{D})$ and*

$$f(w) = \langle f, k_w\rangle \qquad (f \in H^2(\mathbb{D})) \tag{10.16}$$

so that $f \mapsto f(w)$ gives a continuous linear functional $H^2(\mathbb{D}) \to \mathbb{C}$.

Proof The map $f \mapsto f(w)$ is clearly linear. For $w \in \mathbb{D}$ let $k_w(z) = 1/(1 - \bar{w}z)$, which is a rational function of z with pole at $1/\bar{w}$ outside $\overline{\mathbb{D}}$. Also, by Cauchy's formula for the circle $C(0, 1)$ we have

$$\langle f, k_w\rangle = \int_{C(0,1)} \frac{f(z)}{1 - \bar{z}w}\frac{dz}{2\pi iz} = \int_{C(0,1)} \frac{f(z)}{z - w}\frac{dz}{2\pi i} = f(w). \tag{10.17}$$

Taking $f = k_w$, we have

$$\langle k_w, k_w\rangle = k_w(w) = \frac{1}{1 - |w|^2}. \tag{10.18}$$

Hence $|f(w)| \leq \|f\|\|k_w\| = \|f\|/\sqrt{1 - |w|^2}$. □

Lemma 10.11 *The map $(u_n)_{n=0}^\infty \mapsto u(z) = \sum_{n=1}^\infty u_n z^n$ gives a linear isometric isomorphism between the space ℓ^2 of square summable complex sequences and the Hardy space H^2.*

Proof We identify each $(u_n)_{n=0}^\infty$ with the corresponding power series $u(z) = \sum_{n=1}^\infty u_n z^n$. Note that $(z^n)_{n=0}^\infty$ gives an orthonormal sequence in H^2. Given that $\sum_{n=0}^\infty |u_n|^2$ converges, we have $u_n z^n \to 0$ as $n \to \infty$ for all $z \in \mathbb{D}(0, 1)$. Hence $u(z)$ has radius of convergence ≥ 1, and $u(z) = \sum_{n=1}^\infty u_n z^n$ converges and determines a holomorphic function on $\mathbb{D}(0, 1)$. Now we take $0 \leq r < 1$, and write $z = re^{i\theta}$ so $u(re^{i\theta}) = \sum_{n=1}^\infty u_n r^n e^{in\theta}$ gives an absolutely and uniformly convergent series of functions for $\theta \in [0, 2\pi]$. We can therefore write

$$\begin{aligned}
\int_0^{2\pi} |u(re^{i\theta})|^2\frac{d\theta}{2\pi} &= \int_0^{2\pi} \Big(\sum_{n=0}^\infty u_n r^n e^{in\theta}\Big)\Big(\sum_{n=0}^\infty \bar{u}_n r^n e^{-in\theta}\Big)\frac{d\theta}{2\pi}\\
&= \sum_{n=0}^\infty\sum_{m=0}^\infty u_n\bar{u}_m r^{n+m}\int_0^{2\pi} e^{i(n-m)\theta}\frac{d\theta}{2\pi}\\
&= \sum_{n=0}^\infty |u_n|^2 r^{2n}.
\end{aligned}$$

Letting $r \to 1-$, we deduce that

$$\lim_{r\to 1-}\int_0^{2\pi} |u(re^{i\theta})|^2\frac{d\theta}{2\pi} = \sum_{n=0}^{\infty} |u_n|^2.$$

This shows that $u \in H^2$. The map $(u_n)_{n=0}^{\infty} \mapsto u(z)$ is linear, when we interpret $u(z) + v(z)$ as the usual pointwise sum of holomorphic functions on $\mathbb{D}(0, 1)$.

Conversely, every $u \in H^2$ is holomorphic on $\mathbb{D}(0, 1)$ and determines a Taylor series $u(z) = \sum_{n=0}^{\infty} u_n z^n$, which by the preceding calculation satisfies

$$\sum_{n=0}^{\infty} |u_n|^2 = \sup_{0<r<1}\int_0^{2\pi} |u(re^{i\theta})|^2\frac{d\theta}{2\pi} < \infty.$$

Hence $(u_n)_{n=0}^{\infty} \in \ell^2$. Thus every $u \in H^2$ arises from a uniquely determined $(u_n)_{n=0}^{\infty} \in \ell^2$. □

10.3 Subspaces and Blocks

Let H and K be Hilbert spaces and form their direct sum $L = H \oplus K = \{(\xi; \eta) : \xi \in K, \eta \in K\}$ with coordinatewise addition and inner product

$$\langle(\xi_1; \eta_1), (\xi_2; \eta_2)\rangle_L = \langle\xi_1, \xi_2\rangle_H + \langle\eta_1, \eta_2\rangle_K \qquad (\xi_1, \xi_2 \in H; \eta_1, L : \xi\eta_2 \in K).$$

Then there is a natural isometric linear embedding $\iota : H \to L\ \xi \mapsto (\xi; 0)$ and a linear projection $P : L \to H : (\xi; \eta) \to \xi$ such that $P\iota = I : H \to H$. Often ι is suppressed so that H is regarded as a closed linear subspace of L. We can write $K = L \ominus H$ to indicate that K is the complementary subspace to H within L.

We can also form direct sums of subspaces from inside a given Hilbert space, as follows.

Definition 10.12 (Orthogonal Complement) Let K be a linear subspace of H which is closed in the sense that if $k_n \in K$ and $k \in H$ has the property that $\|k - k_n\| \to 0$ as $n \to \infty$ then $k \in K$. The orthogonal complement of K in H is defined to be

$$K^{\perp} = H \ominus K = \{h \in H : \langle k, h\rangle = 0; \forall k \in K\}. \tag{10.19}$$

One easily shows that $K^{\perp}$ is a closed linear subspace of H and $H = K \oplus K^{\perp}$ in the sense that for all $h \in H$ there exist unique $k \in K$ and $\ell \in K^{\perp}$ such that $h = k + \ell$. Given this, one can show that $(K^{\perp})^{\perp} = K$.

Proposition 10.13 (Orthogonal Projection) *Let K be a nonzero closed linear subspace of a Hilbert space H, and suppose that $(e_j)_{j=1}^{\infty}$ is an orthonormal basis for K. Let $Px = \sum_{j=1}^{\infty} \langle x, e_j \rangle e_j$. Then P is the unique operator with the following properties:*

(i) $P : H \to K$ *is a bounded linear operator with* $\|P\| = 1$;
(ii) $\langle Px, y \rangle = \langle x, Py \rangle = \langle Px, Py \rangle$ *for all* $x, y \in H$;
(iii) $Pk = k$ *for all* $k \in K$, *and* $\langle x - Px, k \rangle = 0$ *for all* $x \in H$ *and* $k \in K$;
(iv) $\|x - Px\| = \inf\{\|x - k\| : k \in K\}$, *and the infimum is uniquely attained at* $k = Px$.

Proof

(i) Linearity follows from linearity of the inner product. Also $Px \in K$ and $x = Px + (x - Px)$ where $x - Px$ is orthogonal to all of the e_j, so Px is orthogonal to $x - Px$ and $\|x\|^2 = \|Px\|^2 + \|x - Px\|^2$ where

$$\|Px\|^2 = \sum_{j=1}^{\infty} |\langle x, e_j \rangle|^2 \leq \|x\|^2,$$

and $Pe_1 = 1_1$, so $\|P\| = 1$.

(ii) The three expressions in (ii) all equal $\langle Px, y \rangle = \sum_{j=1}^{\infty} \langle x, e_j \rangle \overline{\langle y, e_j \rangle}$.

(iii) For $k \in K$, we have $k = \sum_{j=1}^{\infty} \langle k, e_j \rangle e_j$, so $Pk = k$ and

$$\langle Px, k \rangle = \sum_{j=1}^{\infty} \langle x, e_j \rangle \overline{\langle k, e_j \rangle} = \langle x, k \rangle$$

so $\langle x - Px, k \rangle = 0$.

(iv) We have an orthogonal decomposition $x - k = (x - Px) + (Px - k)$ so

$$\|x - k\|^2 = \|x - Px\|^2 + \|Px - k\|^2,$$

so we minimize the right hand side by taking $k = Px$, and this choice is unique. By (iii) and (iv), P is unique.

□

Definition 10.14 (Orthogonal Projection) The P in Proposition 10.13 is called the orthogonal projection onto K.

Corollary 10.15 (F. Reisz-Fréchet) *Given a linear map $\phi : H \to \mathbb{C}$ such that $|\phi(x)| \leq C\|x\|$ for all $x \in H$ for some $C > 0$, there exists a unique $e \in H$ such that $\phi(x) = \langle x, e \rangle$.*

Proof First observe that any $e \in H$ gives rise to such a linear functional via $\phi(x) = \langle x, e \rangle$, so we need to show that all functionals arise thus. Let $K = \{x \in H : \phi(x) = 0\}$ which is a linear subspace; also K is closed since $x_n \to x$ with $x_n \in K$ implies

$|\phi(x)| = |\phi(x_n - x)| \leq C\|x_n - x\| \to 0$, so $\phi(x) = 0$ and $x \in K$. Let P be the orthogonal projection onto K. If $\phi = 0$, then we can choose $e = 0$; otherwise, we choose $f \in H$ such that $\phi(f) \neq 0$, and introduce $u = (f - Pf)/\|f - Pf\|$ so that u has $\|u\| = 1$ and $u \in K^\perp$; also $\phi(Pf) = 0$ so $\phi(u) \neq 0$. Now for any $x \in H$, we have

$$x = \left(x - \frac{\phi(x)}{\phi(u)}u\right) + \frac{\phi(x)}{\phi(u)}u$$

where the term in parenthesis belongs to K, hence is perpendicular to u. Finally we take the inner product with $e = \overline{\phi(u)}u$ to get $\langle x, e\rangle = \phi(x)$, and from the Cauchy–Schwarz inequality we also obtain $\|\phi\| = \|e\|$ is the best possible choice of C. □

Lemma 10.16 (Adjoint) *Let $T : H \to H$ be a bounded linear operator. Then there exists a unique bounded linear operator $T' : H \to H$ such that $\langle Tx, y\rangle = \langle x, T'y\rangle$ for all $x, y \in H$.*

Proof For $y \in H$, the linear map $\phi : H \to \mathbb{C}$ given by $\phi(x) = \langle Tx, y\rangle$ gives $T'y \in H$ such that $\langle Tx, y\rangle = \langle x, T'y\rangle$ by Corollary 10.15. One can check that $y \mapsto T'y$ is a linear map. Also T' is bounded since

$$\begin{aligned}\|T'\| &= \sup\{|\langle x, T'y\rangle; x, y \in H; \|x\| = \|y\| = 1\}\\ &= \sup\{|\langle Tx, y\rangle; x, y \in H; \|x\| = \|y\| = 1\}\\ &= \|T\|. \end{aligned}\tag{10.20}$$

□

Exercise

(i) Verify that this definition is consistent with (3.45) and Definition 2.15 for a finite-dimensional Hilbert space such as $\mathbb{C}^{n\times 1}$ with the standard inner product.
(ii) Show that Lemma 3.17 extends to this context. Show also that the adjoint T' is uniquely determined by its defining equation.
(iii) Deduce that a linear operator $V : H \to B$ is an isometry if and only if $V'V = I$.
(iv) Let $(e_n)_{n=1}^\infty$ be an orthonormal sequence in H and $V : H \to H$ a linear isometry. Show that $(Ve_n)_{n=1}^\infty$ is also an orthonormal sequence in H.

We consider block matrices of the form

$$T = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}, T' = \begin{bmatrix} A' & 0 \\ B' & 0 \end{bmatrix} \quad \begin{matrix} H \ominus K \\ K \end{matrix}\tag{10.21}$$

Note that matrices of the form of T arise when one considers elementary row operations to produce zero rows at the bottom of the array.

Proposition 10.17 *Let $T : H \to H$ be a bounded linear operator and let $K = \{x \in H : T'x = 0\}$ be the nullspace of T'.*

(i) Then K is a closed linear subspace of H, and its orthogonal complement $H \ominus K$ is the closure of the range $\{Ty : y \in H\}$ of T.
(ii) Also T' maps K into K, and T maps $H \ominus K$ into $H \ominus K$.

Proof

(i) For $x, z \in K$ and $\lambda, \mu \in \mathbb{C}$, we have $T'(\lambda x + \mu z) = \lambda T'x + \mu T'z = 0$, so $\lambda x + \mu z \in K$. Also, if $x_n \in K$ and $x_n \to x$ as $n \to \infty$, then by continuity $T'x_n \to T'x$ as $n \to \infty$, so $T'x = 0$, hence $x \in K$. Hence K is a closed linear subspace of H. To identify its orthogonal complement, we observe that $x \in K$ if and only if $\langle y, T'x\rangle = 0$ for all $y \in H$ so $\langle Ty, x\rangle = 0$ for all y; so x is perpendicular to the range of T.
(ii) This follows from the definition of K and (i).

□

Definition 10.18 We can form the following block matrices of bounded linear operators. See [15]

(i) We say that $A : H \to H$ has dilation $T : L \to L$ if $PT\iota = A$; equivalently we say that A is the compression of T to H, when we have a block matrix and a commuting diagram

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad \begin{array}{ccc} L & \longrightarrow & L \\ \iota \uparrow & & \downarrow P. \\ H & \longrightarrow & H \end{array} \tag{10.22}$$

(ii) In particular, $A : H \to H$ has lifting $T : L \to L$ if $PT = A$ so that we have a block matrix and a commuting diagram

$$T = \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \qquad \begin{array}{ccc} L & \longrightarrow & L \\ P \downarrow & & \downarrow P. \\ H & \longrightarrow & H \end{array} \tag{10.23}$$

(iii) Also $A : H \to H$ has an extension $T : L \to L$ if $TP = A$, so we have a block matrix and a commuting diagram

$$T = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \qquad \begin{array}{ccc} L & \longrightarrow & L \\ \iota \uparrow & & \uparrow \iota. \\ H & \longrightarrow & H \end{array} \tag{10.24}$$

Hence A has extension T if and only if $A' : H \to H$ has lifting $T' : L \to L$ so $PT' = A'$.

When $L = H \oplus K$, we can write

$$0 \longrightarrow H \longrightarrow L \longrightarrow K \longrightarrow 0$$

to indicate that H is isometrically included in L and that H is the nullspace of the orthogonal projection from L onto K. Then in situation (iii), we can say that T is an extension of A and a lifting of D, as in the commuting diagram

$$\begin{array}{ccccccccc} 0 \longrightarrow & H & \longrightarrow & L & \longrightarrow & K & \longrightarrow 0 \\ & A \downarrow & & T \downarrow & & \downarrow D & \\ 0 \longrightarrow & H & \longrightarrow & L & \longrightarrow & K & \longrightarrow 0 \end{array} \quad . \tag{10.25}$$

which leaves B undetermined.

10.4 Shifts and Multiplication Operators

The shift operator S and its adjoint S' are fundamental to the theory, and arises naturally when one considers ℓ^2 and H^2.

Definition 10.19 (Shifts) Let $u(z) = \sum_{n=0}^{\infty} u_n z^n$ be a convergent power series on some open disc. Then the (forward) shift operator on ℓ^2 and H^2 is

$$Su(z) = zu(z) \qquad S : (u_0, u_1, \dots) \mapsto (0, u_0, u_1, \dots); \tag{10.26}$$

the backward shift operator on ℓ^2 and H^2 is

$$Au(z) = \frac{u(z) - u(0)}{z} \qquad A : (u_0, u_1, \dots) \mapsto (u_1, u_2, \dots). \tag{10.27}$$

We also introduce a linear projection on ℓ^2 and H^2 by

$$Pu(z) = u(0) \qquad P : (u_0, u_1, \dots) \mapsto (u_0, 0, \dots), \tag{10.28}$$

Exercise For $\lambda \in \mathbb{D}(0, 1)$, let

$$\varphi_\lambda(z) = \frac{1}{1 - \bar{\lambda} z}. \tag{10.29}$$

(i) Use power series to show that $A\varphi_\lambda = \bar{\lambda}\varphi_\lambda$, so φ_λ is an eigenvector corresponding to eigenvalue $\bar{\lambda}$ of A.
(ii) Likewise show that $\langle f, \varphi_\lambda \rangle = f(\lambda)$ for all $f \in H^2$.

Proposition 10.20 (Shifts)

(i) The shift S is an isometric linear transformation on H^2, so $\|Su\| = \|u\|$ for all $u \in H^2$.
(ii) The backward shift A is a bounded linear operator on H^2, so $\|A\| = 1$, and $\|A^n u\| \to 0$ as $n \to \infty$ for all $u \in H^2$.
(iii) $AS = I$ and $SA = I - P$, where $P = P'$ is of rank one and $P = P^2$;
(iv) $A = S'$ and $S = A'$.

Proof

(i) Clearly S is linear, and $\|Su\| = \|u\|$.
(ii) Also A is linear and $u(z) = \sum_{n=0}^{\infty} u_n z^n$ satisfies $\|Au\|^2 = \sum_{k=1}^{\infty} |u_k|^2 \leq \|u\|^2$; hence $\|A\| \leq 1$, and by choosing $u(0) = 0$, we can achieve equality. We have

$$\|A^n u\|^2 = \sum_{k=n}^{\infty} |u_k|^2 \to 0 \tag{10.30}$$

as $n \to \infty$.
(iii) Note that $Su(0) = 0$, so $ASu(z) = u(z)$, hence $AS = I$; also $SAu(z) = u(z) - u(0)$, so $SA = I - P$. The operator P is the orthogonal projection onto the constant functions.
(iv) With $v(z) = \sum_{n=0}^{\infty} v_n z^n$, we have

$$\langle Su, v\rangle = \sum_{k=0}^{\infty} u_k \bar{v}_{k+1} = \langle u, Av\rangle. \tag{10.31}$$

Here A is coisometric in the sense that $A' = S$ is isometric.

□

Exercise Let $C : H^2 \to \mathbb{C} : f(z) \mapsto f(0)$.

(i) Find $C' : \mathbb{C} \to H^2$ and deduce that $C'C = P : H^2 \to H^2$ and $CC' : \mathbb{C} \to \mathbb{C}$. Note the distinction between constants in $\mathbb{C}$ and the constant functions in H^2.
(ii) Deduce that

$$\begin{bmatrix} A \\ C \end{bmatrix} : H^2 \to \begin{matrix} H^2 \\ \mathbb{C} \end{matrix}, \quad \begin{bmatrix} S \ C' \end{bmatrix} : \begin{matrix} H^2 \\ \mathbb{C} \end{matrix} \to H^2$$

are adjoints of one another, and

$$\begin{bmatrix} S \ C' \end{bmatrix} \begin{bmatrix} A \\ C \end{bmatrix} = I : H^2 \to H^2, \quad \begin{bmatrix} A \\ C \end{bmatrix} \begin{bmatrix} S \ C' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} : \begin{matrix} H^2 \\ \mathbb{C} \end{matrix} \to \begin{matrix} H^2 \\ \mathbb{C} \end{matrix}. \tag{10.32}$$

Remark 10.21

(i) To have a shift operator that satisfies the Proposition 10.20, it is essential that H is infinite-dimensional. A linear isometry on a finite-dimensional Hilbert space is unitary, since for $U \in M_{n\times n}(\mathbb{C})$ the condition $U'U = I_n$ implies $UU' = I_n$.

(ii) Beurling [4] characterized all the closed linear subspaces K of H^2 such that $SK \subseteq K$; these are the shift-invariant subspaces.

(iii) He also considered the closed linear subspaces K of H^2 such that $AK \subset K$. An example of such is $\mathrm{span}\{\varphi_\lambda : \lambda \in E\}$ for any finite subset E of $\mathbb{D}$. This follows from the identity $A\varphi_\lambda = \bar{\lambda}\varphi_\lambda$.

(iv) The shift operator is an essential tool in the study of stationary stochastic processes. Wiener and Masani [62] use Hardy spaces of holomorphic functions on the disc as a model space and then extend some results to matrix valued holomorphic functions. In this way, questions about stochastic processes are converted into questions about operators on Hilbert space, with the shift operator being the crucial example.

Let $\varphi(s) = (s-1)/(s+1)$. Then φ is a rational function with a pole at $s = -1$ which maps RHP onto $\mathbb{D}$ and has inverse $\psi(z) = (z+1)/(-z+1)$, where ψ maps $\mathbb{D}$ onto RHP. Now consider $T(s) \in \mathbb{C}(s)$, and write $W(z) = T \circ \psi(z)$. Then $T \mapsto T \circ \psi$ gives an algebra isomorphism $\mathbb{C}(s) \to \mathbb{C}(z)$ with inverse $W \mapsto W \circ \varphi$. Observe that T has all its poles in LHP if and only if W has none of its poles in $\overline{\mathbb{D}}$. Further, T has poles on $i\mathbb{R} \cup \{\infty\}$ if and only if W has poles on the unit circle $C(0, 1)$. This proves the following result.

Proposition 10.22 *The space $\mathcal{S}$ of stable rational functions corresponds to the space $\mathcal{S}_{\mathbb{D}}$ of rational functions that have no poles in $\overline{\mathbb{D}}$ under the map $T \mapsto W = T \circ \psi$.*

Example 10.23 Let $W(z) = z$ and consider $Sf(z) = zf(z)$, so that $S : H^2 \to H^2$ is the shift operator. The following result gives a version of Proposition 10.20 for the operator of multiplication by a typical $W \in \mathcal{S}_{\mathbb{D}}$. One can consider $f(z) = \sum_{n=0}^{\infty} a_n z^n$ as power series in z corresponding to a signal $(a_n)_{n=1}^{\infty}$ with $\sum_{n=0}^{\infty} |a_n|^2$ convergent, and $W(z)$ as a transfer function from a discrete-time linear system as in (8.4).

Lemma 10.24 *A function $W : \mathbb{D} \to \mathbb{C}$ defines a bounded linear operator $M_W : H^2 \to H^2 : \quad f(z) \mapsto W(z)f(z)$ if and only if W is bounded and holomorphic on $\mathbb{D}$.*

Proof ($\Rightarrow$) We have $W(z) = M_W 1 \in H^2(\mathbb{D})$, so W is holomorphic on $\mathbb{D}$. Also

$$\langle Wk_z, k_z\rangle = W(z)k_z(z) = W(z)\|k_z\|^2,$$

while

$$|\langle Wk_z, k_z\rangle| = |\langle M_W k_z, k_z\rangle| \le \|M_W\|\|k_z\|^2,$$

so $|W(z)| \leq \|M_W\|$, for all $z \in \mathbb{C}$, hence W is bounded.

($\Leftarrow$) Suppose that $|W(z)| \leq M$ for all $z \in \mathbb{D}$. Then $W(z)f(z)$ is holomorphic on $\mathbb{D}$ for all $f \in H^2$, and

$$\int_0^{2\pi} |W(re^{i\theta})f(re^{i\theta})|^2 \frac{d\theta}{2\pi} \leq M^2 \int_0^{2\pi} |f(re^{i\theta})|^2 \frac{d\theta}{2\pi} \tag{10.33}$$

so $\|Wf\|_{H^2} \leq M\|f\|_{H^2}$, hence $\|M_W\| \leq M$. The operator M_W is evidently linear. □

Proposition 10.25 *Let $W \in \mathcal{S}_{\mathbb{D}}$ be a rational function with no poles on $\overline{\mathbb{D}}$, and let*

$$M_W : H^2 \to H^2 : \quad f(z) \mapsto W(z)f(z). \tag{10.34}$$

(i) Then M_W gives a bounded linear operator on H^2.
(ii) The adjoint M_W' has eigenvector k_w with eigenvalue $\overline{W(w)}$ for all $w \in \mathbb{D}$.
(iii) Suppose that W is non zero and has zeros $w_j \in \mathbb{D}$ for $j = 1, \dots, m$. Then the null space of M_W is $\{0\}$, and the orthogonal complement of the range of M_W contains span$\{k_{w_j}; j = 1, \dots, m\}$.

Proof

(i) The function W is continuous on the closed and bounded set $\overline{\mathbb{D}}$, hence it is bounded there with $|W(z)| \leq M$ for all $z \in \overline{\mathbb{D}}$ for some $M \geq 0$. By the Lemma, M_W gives a bounded linear operator on $H^2(\mathbb{D})$.

(ii) The operator M_W has an adjoint M_W' which is a bounded linear operator on H^2, characterized by $\langle M_W f, g\rangle = \langle f, M_W' g\rangle$ for all $f, g \in H^2$. Taking $g = k_w$, we use the formula (10.16) to show that

$$\langle f, M_W' k_w\rangle = \langle M_W f, k_w\rangle = \langle Wf, k_w\rangle = W(w)f(w), \tag{10.35}$$

so that

$$\langle f, M_W' k_w\rangle = \langle f, \overline{W(w)}k_w\rangle, \tag{10.36}$$

hence $M_W' k_w = \overline{W(w)}k_w$. [Suppose W is not a constant. Then operator M_W' is not multiplication by $\bar{W}$, since $\bar{W}$ is not holomorphic.]

(iii) Suppose that $W \in \mathcal{S}_{\mathbb{D}}$ is non constant. Then the null space of M_W is $\{f \in H^2 : W(z)f(z) = 0, \forall z \in \mathbb{D}\}$ is $\{0\}$. Equivalently, the range of M_W' is dense in H^2, since f is orthogonal to the range of M_W' if and only if $\langle f, M_W' g\rangle = 0$ for all $g \in H^2$, so that $\langle Wf, g\rangle = 0$ for all $g \in H^2$, so $Wf = 0$.

The range of M_W is often denoted WH^2, and is $\{Wf : f \in H^2\}$. Suppose that W has zeros at $w_1, \dots, w_m \in \mathbb{D}$; then $\langle Wf, k_{w_j}\rangle = W(w_j)f(w_j) = 0$ for all $f \in H^2$, so k_{w_j} is orthogonal to WH^2. Hence the orthogonal complement of WH^2 is $H \ominus WH^2$, which contains span$(k_{w_j}; j = 1, \dots, m\}$. We also observe that $H^2 \ominus WH^2 = \text{null}(M_W')$. □

Exercise Suppose that W has simple zeros at $w_1, \dots, w_n \in \mathbb{D}$, and has no other zeros in $\overline{\mathbb{D}}$. For example, let

$$W(z) = \prod_{j=1}^{n} \frac{z - w_j}{1 - \bar{w}_j z} \tag{10.37}$$

which satisfies $|W(z)| < 1$ for all $z \in \mathbb{D}$.

(i) Show that

$$\{h \in H^2 : h(w_j) = 0; j = 1, \dots, n\} = \mathrm{span}\{k_{w_j} : j = 1, \dots, n\}^{\perp}. \tag{10.38}$$

(ii) Show that if $h \in H^2$ has $h(w_j) = 0$ for $j = 1, \dots, n$, then $h/W \in H^2$, and deduce that

$$WH^2 = \mathrm{span}\{k_{w_j} : j = 1, \dots, n\}^{\perp}, \tag{10.39}$$

$$H^2 = WH^2 \oplus \mathrm{span}\{k_{w_j} : j = 1, \dots, n\}. \tag{10.40}$$

Definition 10.26 Let W be a rational function that maps $\mathbb{D}$ into itself, and for distinct points $z_1, \dots, z_n \in \mathbb{D}$ let P be the matrix

$$P = \Big[\frac{1 - \overline{W(z_j)}W(z_\ell)}{1 - \bar{z}_j z_\ell}\Big]_{j,\ell=1}^{n}. \tag{10.41}$$

We call W a Pick function and P the Pick matrix for the points $w_1, \dots, w_n$.

Proposition 10.27 *Then the Pick matrix P is positive semidefinite.*

Proof The matrix is hermitian symmetric, so $P = P'$. Let $\alpha = (a_j)_{j=1}^{n} \in \mathbb{C}^{n\times 1}$ and consider

$$\begin{aligned}\langle P\alpha, \alpha\rangle &= \sum_{j,\ell=1}^{n} \frac{1 - \overline{W(z_j)}W(z_\ell)}{1 - \bar{z}_j z_\ell} a_j \bar{a}_\ell \\ &= \sum_{j,\ell=1}^{n} \frac{a_j \bar{a}_\ell}{1 - \bar{z}_j z_\ell} - \sum_{j,\ell=1}^{n} \frac{\overline{W(z_j)} a_j W(z_\ell) \bar{a}_\ell}{1 - \bar{z}_j z_\ell}.\end{aligned} \tag{10.42}$$

The first double sum in the last line is

$$\sum_{j,\ell=1}^{n} a_j \bar{a}_\ell \langle k_{z_j} k_{z_\ell}\rangle = \Big\langle \sum_{j=1}^{n} a_j k_{z_j}, \sum_{\ell=1}^{n} a_\ell k_{z_\ell} \Big\rangle = \langle f, f\rangle,$$

where we have introduced $f(z) = \sum_{j=1}^{n} a_j k_{z_j}$. We now consider

$$M'_W f = \sum_{j=1}^{n} a_j M'_W k_{z_j} = \sum_{j=1}^{n} a_j \overline{W(z_j)} k_{z_j},$$

so

$$\begin{aligned}\big\langle M'_W f, M'_W f\big\rangle &= \Big\langle \sum_{j=1}^{n} a_j M'_W k_{z_j}, \sum_{\ell=1}^{n} a_\ell M'_W k_{z_\ell}\Big\rangle \\ &= \sum_{j,\ell=1}^{n} \overline{W(z_j)} W(z_\ell) a_j \bar{a}_\ell \langle k_{z_j}, k_{z_\ell}\rangle \\ &= \sum_{j,\ell=1}^{n} \frac{\overline{W(z_j)} W(z_\ell)}{1-\bar{z}_j z_\ell} a_j \bar{a}_\ell\end{aligned}$$

which we recognize as the final summand in (10.42). Hence we have

$$\langle P\alpha, \alpha\rangle = \langle f, f\rangle - \langle M'_W f, M'_W f\rangle = \|f\|^2 - \|M'_W f\|^2. \tag{10.43}$$

Since $|W(z)| < 1$ for all $z \in \mathbb{D}$, we have $\|M'_W\| = \|M_W\| \le 1$, so $\langle P\alpha, \alpha\rangle \ge 0$. □

Example 10.28 We can take $W(z) = z$, which gives a Pick matrix $P = [1]_{j,\ell=1}^{n}$. This has rank one for all $n = 1, 2, \dots$, and is not positive definite for $n \ge 2$.

We now consider some changes of variable. To respect the structure of Hardy spaces, we use the change of variables $z = (s-1)/(s+1)$ so that $dz/(2\pi i z) = ds/(\pi i(s^2-1))$, and with $z = e^{i\theta}$ and $s = i\omega$, we have $d\theta/(2\pi) = -d\omega/(\pi(1+\omega^2))$. This helps us to transform from the disc to the half-plane, as in Sect. 10.6. The correspondence between the transfer function of a discrete-time linear system $T_d(z) = D_d + zC_d(I - zA_d)^{-1}B_d$ and the transfer function of a continuous time linear system $T(s) = D + C(sI - A)^{-1}B$ is given in Theorem 8.8, which involves a similar idea.

10.5 Canonical Model

In this section we consider how the specific examples (8.4) and (10.34) relate to the backward shift operator. See [4] and [22]. In [12–15] there is a systematic discussion of dilations and extensions of operators. This is a realization theorem, which shows that a Hilbert-space valued holomorphic function on the unit disc may be realized as the transfer function of a discrete time linear system. Let H_0 and H_1 be Hilbert spaces, so H_1 is the state space and H_0 is the input space and output space. Let (A, B, C, D) be continuous linear operators

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{matrix} H_1 \\ H_0 \end{matrix} \to \begin{matrix} H_1 \\ H_0 \end{matrix};$$

- $A : H_1 \to H_1$ is the main operator,
- $B : H_0 \to H_1$ is the input operator,
- $C : H_1 \to H_0$ is the output operator,
- $D : H_0 \to H_0$ is the external (or straight-through) operator.

Given $W(z) = \sum_{n=0}^{\infty} W_n z^n$ and $M > 0$ with $W_n : H_0 \to H_0$ a bounded linear operator with $\|W_n\| \leq M$ for all n, and $W_0 = D$, we seek (A, B, C) such that $W_{n+1} = CA^n B$. We introduce the space of power series with coefficients in H_0

$$H^2(H_0) = \Big\{ \sum_{n=0}^{\infty} a_n z^n; a_n \in H_0; \sum_{n=0}^{\infty} \|a_n\|^2 < \infty \Big\} \tag{10.44}$$

which forms a Hilbert space H_1 with inner product

$$\Big\langle \sum_{n=0}^{\infty} a_n z^n, \sum_{n=0}^{\infty} b_n z^n \Big\rangle = \sum_{n=0}^{\infty} \langle a_n, b_n \rangle_{H_0}. \tag{10.45}$$

We introduce the linear operators:

$$\begin{aligned} A :& H^2(H_0) \to H^2(H_0) : f(z) \mapsto \frac{f(z) - f(0)}{z}, \\ B :& H_0 \to H^2(H_0) : \qquad b \mapsto \frac{W(z) - W_0}{z} b, \\ C :& H^2(H_0) \to H_0 : \qquad f(z) \mapsto f(0), \\ D :& H_0 \to H_0 : \qquad b \mapsto W_0 b. \end{aligned}$$

Proposition 10.29 *Let $W(z)$ be as above. Then (A, B, C, D) are bounded linear operators that determine a linear system with transfer function*

$$W(z) = D + zC(I - zA)^{-1} B \tag{10.46}$$

where $W(z) = \sum_{n=0}^{\infty} W_n z^n$ is holomorphic on $\mathbb{D}$, with Taylor coefficients $W_0 = D$ and $W_{n+1} = CA^n B$ for $n = 0, 1, \ldots$.

Proof We have $Bb = \sum_{k=1}^{\infty} W_k z^{k-1} b$ and $A^n : \sum_{k=0}^{\infty} a_k z^k \mapsto \sum_{k=n}^{\infty} a_k z^{k-n}$, so for $n = 1, 2, \ldots$, we have

$$A^n Bb = \sum_{k=n+1}^{\infty} z^{k-n-1} W_k b \tag{10.47}$$

so $CA^nBb = W_{n+1}b$; hence

$$\begin{aligned} Db + zC(I - zA)^{-1}Bb &= Db + zCBb + \sum_{n=1}^{\infty} z^{n+1}CA^nBb \\ &= W_0b + zW_1b + \sum_{n=1}^{\infty} z^{n+1}W_{n+1}b \\ &= W(z)b \qquad (b \in H_0). \end{aligned}$$

□

Exercise Find formulas for A' and C', and show that

$$\begin{bmatrix} A \\ C \end{bmatrix} : \begin{matrix} H^2(H_0) \\ H_0 \end{matrix} \to H^2(H_0), \quad \begin{bmatrix} A' & C' \end{bmatrix} : H^2 \to \begin{matrix} H^2(H_0) \\ H_0 \end{matrix} \tag{10.48}$$

are adjoints of one another, and are inverses in the sense that

$$\begin{bmatrix} A' & C' \end{bmatrix} \begin{bmatrix} A \\ C \end{bmatrix} = I : H^2(H_0) \to H^2(H_0), \tag{10.49}$$

$$\begin{bmatrix} A \\ C \end{bmatrix} \begin{bmatrix} A' & C' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} : \begin{matrix} H^2(H_0) \\ H_0 \end{matrix} \to \begin{matrix} H^2(H_0) \\ H_0 \end{matrix}. \tag{10.50}$$

This result has a converse to the effect that a bounded transfer function can be realized from a linear system (A, B, C, D) of the above form. Furthermore, one can often realize the linear system explicitly. See [22].

10.6 Hardy Space on the Right Half-Plane

Hardy space on the unit disc is a suitable function space for describing linear systems in discrete time and their power series transforms. To describe linear systems in continuous time and their Laplace transforms, we introduce the Hardy space of functions on the right-half plane. Some of the properties are inherited from H^2 of the disc by change of variables, and we will introduce an appropriate orthonormal basis in this way. The basis is related to the Laguerre system or orthogonal polynomials via the Laplace transform, which turns out to be the crucial step in the theory. In the next section we consider how the continuous time signals can be introduced into the Hardy space via the Laplace transform.

Let $H^2(RHP)$ be the space of holomorphic functions on $\{s : \Re s > 0\}$ such that

$$\sup_{x>0} \int_{-\infty}^{\infty} |f(x+iy)|^2 \frac{dy}{2\pi} < \infty. \tag{10.51}$$

The inner product is given by

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(iy)\bar{g}(iy)\frac{dy}{2\pi}. \tag{10.52}$$

Lemma 10.30

(i) All strictly proper and stable rational functions belong to $H^2(RHP)$.
(ii) For $\Re z > 0$, let $k_z(s) = 1/(s+\bar{z})$. Then $k_z \in H^2(RHP)$ and

$$f(z) = \langle f, k_z\rangle \qquad (f \in H^2(RHP)) \tag{10.53}$$

so that $f \mapsto f(z)$ gives a bounded linear functional $H^2(RHP) \to \mathbb{C}$.

Proof

(i) For $f \in \mathcal{S}$, the poles of f are in LHP so f is bounded and holomorphic on the RHP; when f is strictly proper, $f(s) = O(1/s)$ as $s \to \infty$; hence the integral of $|f(s)|^2$ converges.
(ii) The function $k_z(s)$ has a pole at $-\bar{z}$ in the left half-plane, so one can easily check that $k_z \in H^2(RHP)$. By applying the Cauchy integral formula to the semicircular contour in the left half-plane, we have the formula

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{f(s)}{s-z}ds = -f(z),$$

where the factor of -1 arises since we go up the imaginary axis and hence describe the contour in the negative sense. By parametrizing the imaginary axis by $s = iy$, we can express the integral as an inner product in $H^2(RHP)$. This integration formula is closely related to the Poisson integral formula (5.70), but here we require $k_z \in H^2(RHP)$.

□

Example 10.31 The reader will easily check that the following functions belong to $H^2(RHP)$:

(i) $1/(1+s)$;
(ii) $P(s)$, a strictly proper stable rational function;
(iii) $(\log p(s))/(1+s)$, where $p(s)$ is a stable polynomial;
(iv) $(\log s)/(1+s)$;
(v) whereas $1/s$ does not, since $\int_{-\infty}^{\infty} d\omega/(x^2+\omega^2) \to \infty$ as $x \to 0+$. A rational function in H^2 cannot have poles on the imaginary axis.

Lemma 10.32 *The functions*

$$f_n(s) = \sqrt{2}\frac{(s-1)^n}{(s+1)^{n+1}} \quad (n = 0, 1, \dots). \tag{10.54}$$

give a complete orthonormal basis $(f_n)_{n=0}^{\infty}$ *for* $H^2(RHP)$.

Proof We observe that $f_n(s)$ is holomorphic except for poles at -1, which is in LHP. Also

$$\begin{aligned}\langle f_n, f_n\rangle &= \frac{2}{2\pi}\int_{-\infty}^{\infty}\frac{(i\omega-1)^n}{(i\omega+1)^{n+1}}\frac{(-i\omega-1)^n}{(-i\omega+1)^{n+1}}d\omega\\ &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{(\omega^2+1)^n}{(\omega^2+1)^{n+1}}d\omega = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{d\omega}{1+\omega^2} = 1.\end{aligned}$$

Given $n > k$, we write $n = k + m$ where $m \geq 1$, so

$$\begin{aligned}\langle f_n, f_k\rangle &= \frac{2}{2\pi}\int_{-\infty}^{\infty}\frac{(i\omega-1)^n}{(i\omega+1)^{n+1}}\frac{(-i\omega-1)^k}{(-i\omega+1)^{k+1}}d\omega\\ &= \frac{-1}{\pi}\int_{-\infty}^{\infty}\frac{(i\omega-1)^{m-1}}{(i\omega+1)^{m+1}}d\omega = \frac{-1}{\pi i}\int_{-i\infty}^{i\infty}\frac{(s-1)^{m-1}}{(s+1)^{m+1}}ds = 0\end{aligned}$$

by Cauchy's Theorem. More explicitly, we can apply Cauchy's Theorem to the function $f(s) = (s-1)^{m-1}/(s+1)^{m+1}$, which is holomorphic inside and on the semicircular contour $[-iR, iR] \oplus S_R$ in the left half-plane.

Now observe that

$$\begin{aligned}\langle f, f_k\rangle &= \frac{\sqrt{2}}{2\pi}\int_{-\infty}^{\infty} f(i\omega)\frac{(-i\omega-1)^k}{(-i\omega+1)^{k+1}}d\omega\\ &= \frac{-\sqrt{2}}{2\pi i}\int_{-i\infty}^{i\infty} f(s)\frac{(s+1)^k}{(s-1)^{k+1}}ds,\end{aligned}$$

which by the Cauchy integral formula applied to $[-iR, iR] \oplus S_R$ gives

$$\begin{aligned}\langle f, f_k\rangle &= \frac{\sqrt{2}}{k!}\Big(\frac{d^k}{ds^k}\Big)_{s=1}\big(f(s)(s+1)^k\big) = \sum_{j=0}^{k}\frac{\sqrt{2}}{k!}\binom{k}{j}\Big(\frac{d^{k-j}f}{ds^{k-j}}\Big)_{s=1}\Big(\frac{d^j}{ds^j}\Big)_{s=1}(s+1)^k\\ &= \sum_{j=0}^{k}\frac{\sqrt{2}}{k!}\binom{k}{j}\Big(\frac{d^{k-j}f}{ds^{k-j}}\Big)_{s=1}\frac{2^{k-j}k!}{(k-j)!},\end{aligned}$$

in which all the numerical coefficients are positive. Suppose that $\langle f, f_k\rangle = 0$ for all $k = 0, 1, \dots$. From $\langle f, f_0\rangle = 0$, we deduce that $f(1) = 0$; then $\langle f, f_1\rangle = 0$

gives $f'(1) = 0$, and so on until $f^{(k)}(1) = 0$ for all k. By the identity theorem for holomorphic functions, we deduce that $f = 0$ identically. □

Corollary 10.33 *There is an isomorphism of Hilbert spaces $H^2(\mathbb{D}) \to H^2(RHP)$ given by*

$$f(z) \mapsto \frac{\sqrt{2}}{1+s} f\Big(\frac{s-1}{s+1}\Big) \qquad (\Re s > 0). \tag{10.55}$$

Proof The linear fractional transformation $s \mapsto (s-1)/(s+1)$ is holomorphic on RHP and gives a bijection between RHP and $\mathbb{D}$ with holomorphic inverse with the map of orthonormal bases $(z^n)_{n=0}^{\infty} \mapsto (f_n(s))_{n=0}^{\infty}$. By the Lemma 10.32, the map

$$\sum_{n=0}^{\infty} a_n z^n \mapsto \sum_{n=0}^{\infty} a_n \sqrt{2} \frac{(s-1)^n}{(s+1)^{n+1}} \qquad ((a_n) \in \ell^2) \tag{10.56}$$

gives an isomorphism of Hilbert spaces, which is equivalently expressed by the formula (10.55). There is a change of variables $s = i\omega$ and $\omega = \tan(\theta/2)$ so

$$\frac{s-1}{s+1} = \frac{i\omega - 1}{i\omega + 1} = \frac{i\tan(\theta/2) - 1}{i\tan(\theta/2) + 1} = \frac{-1 + \tan^2(\theta/2) + 2i\tan(\theta/2)}{1 + \tan^2(\theta/2)} = -e^{-i\theta}. \tag{10.57}$$

In addition to the change of variables, (10.55) involves a multiplicative factor. Note that $-e^{-i\theta}$ describes the unit circle once in the negative sense for $0 \leq \theta \leq 2\pi$. This relates to our earlier comments about winding numbers for semicircular contours in Sect. 5.1. □

10.7 Paley–Wiener Theorem

Definition 10.34 Let $L^2(0, \infty)$ be the space of Lebesgue measurable functions $f : (0, \infty) \to \mathbb{C}$ such that $\int_0^\infty |f(t)|^2 dt$ converges. (Continuous functions are Lebesgue measurable, as are piecewise continuous functions and pointwise limits of sequences of continuous functions.) The inner product is

$$\langle f, g \rangle = \int_0^\infty f(t)\overline{g(t)}\, dt. \tag{10.58}$$

In this case, the Cauchy–Schwarz inequality

$$\Big| \int_0^\infty f(t)\overline{g(t)} dt \Big| \leq \Big(\int_0^\infty |f(t)|^2 dt \Big)^{1/2} \Big(\int_0^\infty |g(t)|^2 dt \Big)^{1/2} \tag{10.59}$$

follows from

$$\int_0^\infty \int_0^\infty \left| f(s)g(t) - f(t)g(s) \right|^2 ds dt \geq 0. \tag{10.60}$$

A subset E of $\mathbb{R}$ is said to have Lebesgue measure zero if for all $\varepsilon > 0$ there exists a sequence of bounded intervals (a_j, b_j) such that $E \subseteq \cup_{j=1}^\infty (a_j, b_j)$ and $\sum_{j=1}^\infty (b_j - a_j) < \varepsilon$. In this context, we identify functions f_1 and f_2 if $f_1(t) = f_2(t)$ except on a set of Lebesgue measure zero. If f_1 and f_2 are both continuous and $f_1(t) = f_2(t)$ except on a set of Lebesgue measure zero, then $f_1(t) = f_2(t)$ for all $t > 0$. With this convention, there is no difficulty in interpreting the elements of $L^2((0,\infty);\mathbb{C})$. From results of measure theory, $L^2((0,\infty);\mathbb{C})$ is complete.

Example 10.35 The function $f(t) = t^{-1/4}e^{-t}$ is in $L^2((0,\infty);\mathbb{C})$, although it is unbounded at $0+$. The function $h(t) = H(t) - H(t-2)$ also belongs to $L^2((0,\infty);\mathbb{C})$, although it is discontinuous at $t = 2$. The Laplace transforms of these functions can be computed explicitly

$$\hat{f}(s) = \frac{\Gamma(3/4)}{(s+1)^{3/4}}, \quad \hat{h}(s) = \frac{1-e^{-2s}}{s} \qquad (\Re s > 0). \tag{10.61}$$

The Laplace transform can be defined for $f \in L^2((0,\infty);\mathbb{C})$, and leads to an isomorphism with the Hardy space on the left half-plane. This is expressed in the following result which includes the Paley–Wiener theorem [49] and inversion for the Laplace transform.

Theorem 10.36 (Paley–Wiener) *Let $f \in L^2(0,\infty)$ have Laplace transform $\hat{f}(s) = \int_0^\infty e^{-st} f(t)dt$.*

(i) Then $\hat{f}(s)$ defines a holomorphic function on the RHP;
(ii) if $\langle f, \sqrt{2}e^{-t}L_n(2t)\rangle = 0$ for all n, then $f = 0$;
(iii) the Laplace transform is an isometry $L^2((0,\infty);dt) \to L^2(\mathbb{R}; d\omega/(2\pi))$, so

$$\int_0^\infty |f(t)|^2 dt = \frac{1}{2\pi}\int_{-\infty}^\infty |\hat{f}(i\omega)|^2 d\omega; \tag{10.62}$$

(iv) $\hat{f} \in H^2$, and every $g \in H^2$ arises as the Laplace transform $\hat{h}$ some $h \in L^2(0,\infty)$.

Proof

(i) With $s = x + iy$ for $x > 0$, we use the Cauchy–Schwarz inequality to show

$$|\hat{f}(s)| \leq \int_0^\infty |f(t)e^{-st}|\, dt \leq \Big(\int_0^\infty |f(t)|^2 dt\Big)^{1/2}\Big(\int_0^\infty e^{-2xt} dt\Big)^{1/2}$$
$$\leq \frac{1}{\sqrt{2x}}\Big(\int_0^\infty |f(t)|^2 dt\Big)^{1/2},$$

hence the defining integral for the Laplace transform converges. By a similar proof to Proposition 4.5, one shows that $\hat{f}(s)$ is holomorphic with

$$\frac{d\hat{f}}{ds} = -\int_0^\infty te^{-st} f(t)dt \qquad (\Re s > 0); \tag{10.63}$$

we do not assert that $tf(t)$ is square integrable, but the integral here is still convergent for $\Re s > 0$ since e^{-st} is of exponential decay.

(ii) We observe that

$$\frac{d^n}{ds^n}\hat{f}(s) = (-1)^n \int_0^\infty f(t)t^n e^{-st} dt, \tag{10.64}$$

so

$$\frac{d^n}{ds^n}\hat{f}(1) = (-1)^n \int_0^\infty f(t)t^n e^{-t} dt. \tag{10.65}$$

Suppose that $\langle f, \sqrt{2}e^{-t}L_n(2t)\rangle = 0$ for all $n = 0, 1, \dots$; then $\langle f, t^n e^{-t}\rangle = 0$ for all n, so $\hat{f}^{(n)}(1) = 0$ for all n; hence $\hat{f}(s) = 0$ for all s by the identity theorem. We deduce that

$$0 = \hat{f}(1 + iy) = \int_0^\infty f(t)e^{-t}e^{-iyt}dt, \tag{10.66}$$

so $f(t)e^{-t}$ is an integrable function on $(0, \infty)$ with zero Fourier transform, hence is zero by the Fourier uniqueness theorem (4.94). For an alternative approach, based upon Vitali's completeness theorem, see [50, p 350].

(iii) By (i) and (ii), we can express an arbitrary $f \in L^2(0, \infty)$ as an orthogonal series

$$f(t) = \sum_{n=0}^\infty a_n \sqrt{2}e^{-t}L_n(2t), \tag{10.67}$$

where $a_n = \langle f, \sqrt{2}e^{-t}L_n(2t)\rangle$ and

$$\int_0^\infty |f(t)|^2 dt = \sum_{n=0}^\infty |a_n|^2. \tag{10.68}$$

Taking the Laplace transform, we have

$$\hat{f}(s) = \sum_{n=0}^\infty a_n \sqrt{2}\frac{(s-1)^n}{(s+1)^{n+1}} \tag{10.69}$$

where

$$\int_{-\infty}^{\infty} |\hat{f}(i\omega)|^2 \frac{d\omega}{2\pi} = \sum_{n=0}^{\infty} |a_n|^2. \tag{10.70}$$

(iv) By (iii) we see that $\hat{f} \in H^2$. Conversely, given $g \in H^2$, we introduce

$$b_n = \left\langle g_n, \sqrt{2}\frac{(s-1)^n}{(s+1)^{n+1}} \right\rangle \tag{10.71}$$

so

$$g(s) = \sum_{n=0}^{\infty} b_n \sqrt{2} \frac{(s-1)^n}{(s+1)^{n+1}} \tag{10.72}$$

where $\sum_{n=0}^{\infty} |b_n|^2$ converges. Then $g = \hat{h}$, where

$$h(t) = \sum_{n=0}^{\infty} b_n \sqrt{2} e^{-t} L_n(2t), \tag{10.73}$$

gives an element of L^2 and

$$\int_0^{\infty} |h(t)|^2 dt = \sum_{n=0}^{\infty} |b_n|^2 = \int_{-\infty}^{\infty} |g(i\omega)|^2 \frac{d\omega}{2\pi} \tag{10.74}$$

as in (iii).

□

Example 10.37 Suppose that $(A, 0, C, 0)$ is a stable SISO with initial state x_0. Then the output is $y(t) = C\exp(tA)x_0$ with Laplace transform $Y(s) = C(sI - A)^{-1}x_0$. Then by Theorem 10.36,

$$\int_0^{\infty} \left|C\exp(tA)x_0\right|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left|C(i\omega I - A)^{-1}x_0\right|^2 d\omega.$$

This system has in effect zero input.

The discussion in the rest of this section is about nonzero inputs that are of finite energy in the following sense.

Definition 10.38 (Energy) A signal u is said to have finite energy if $u \in L^2((0,\infty);\mathbb{C})$.

Note that decaying signals such as $(\sin t)/t$ are of finite energy, whereas periodic signals such as $\sin \omega t$ are not of finite energy. The signal represented by unit impulse

δ_0 is not of finite energy. To see this, consider $f_n(t) = n\mathbb{I}_{[0,1/n]}(t)$ with rectangular graph of height n on $(0, 1/n)$ which has the property that

$$\int_0^\infty f_n(t)g(t)dt = n\int_0^{1/n} g(t)dt \to g(0) \qquad (n \to \infty) \tag{10.75}$$

for all continuous functions g. In particular, we can take $g(t) = e^{itx}$ with $g(0) = 1$. Note also that $\int_0^\infty f_n(t)^2dt = n \to \infty$ as $n \to \infty$. We will see that this shows δ_0 does not have finite energy.

The finite energy condition on signals is different from boundedness as in BIBO stability. Nevertheless, BIBO stable systems satisfy the following property for finite energy signals.

Theorem 10.39 *Let (A, B, C, D) be a stable SISO, and suppose that the initial state is zero. If the input is of finite energy, then the output is also of finite energy.*

Proof The transfer function $T(s) = D + C(sI - A)^{-1}B$ is a rational function that is stable, so $T(s)$ has all its poles in LHP and $T(s) \in \mathcal{S}$. Hence $T(s) - D \to 0$ as $s \to \infty$ and there exists M such that $|T(s)| \leq M$ for all s such that $\Re s \geq 0$. Now let $f(s) = \int_0^\infty e^{-st}u(t)dt$, where u is an input of finite energy. Then $f(s) \in H^2$ by the Paley–Wiener theorem. Also, $g(s) = T(s)f(s)$ is holomorphic on the RHP, and

$$\int_{-\infty}^\infty |g(x+i\omega)|^2\frac{d\omega}{\pi} = \int_{-\infty}^\infty |T(x+i\omega)f(x+i\omega)|^2\frac{d\omega}{\pi} \leq M^2\int_{-\infty}^\infty |f(x+i\omega)|^2\frac{d\omega}{\pi} \tag{10.76}$$

for all $x > 0$, so $g \in H^2$. Then $g(s)$ is the Laplace transform of the output y, so by the converse direction (iv) of the Paley–Wiener theorem, y is also of finite energy and

$$\int_0^\infty |y(t)|^2\,dt \leq M^2\int_0^\infty |u(t)|^2\,dt. \tag{10.77}$$

□

A stable rational SISO takes bounded inputs to bounded outputs, and finite energy inputs to finite energy outputs.

10.8 Rational Filters

Suppose that we have a linear system with input $u \in L^2(0, \infty)$ and output $y \in L^2(0, \infty)$. Then we introduce the Laplace transforms

$$U(s) = \int_0^\infty e^{-st}u(t)dt \in H^2$$
$$Y(s) = \int_0^\infty e^{-st}y(t)dt \in H^2 \tag{10.78}$$

and we suppose that they are linked by a multiplication formula $Y(s) = T(s)U(s)$. Suppose that T is holomorphic for $s \in RHP$ and T is bounded, so there exists $M > 0$ such that $|T(s)| \le M$ for all $s \in RHP$. The space of such functions is called H^∞. Then $T(s)U(s)$ belongs to H^2 for all $U \in H^2$ The space H^∞ forms an algebra under pointwise multiplication of functions, and H^∞ contains $\mathcal{S}$. So we can seek to develop control theory using H^∞ instead of $\mathcal{S}$.

There is a factorization theory for H^∞ functions that is based upon Beurling's notion of inner and outer functions [4]. The special feature of an outer function $R \in H^2$ is that the closed linear span of $\{e^{-\gamma s}R(s) : \gamma > 0\}$ is all of H^2. Observe that for $f \in L^2(0, \infty)$, the image of $\{f(t - \gamma) : \gamma > 0\}$ under the Laplace transform is $\{e^{-s\gamma}\hat{f}(s) : \gamma > 0\}$. Wiener studied the properties of $\{\sum_{j=1}^n a_j f(t-\gamma_j) : a_j \in \mathbb{C}; \gamma_j > 0\}$ in various problems in harmonic analysis. We present a simplified discussion that covers only the case of stable rational functions, and leave the interested reader to consult books such as [34] for a complete account of the theory. Rational transfer functions are important since they are relatively easy to calculate.

By Proposition 6.36, any rational function $G(s)$ can be expressed as $G(s) = P(s)/Q(s)$ where $P(s)$ and $Q(s)$ are stable rational functions. Our factorization theorem applies to $P(s)$.

Proposition 10.40 *Let $P(s)$ be a stable rational function with zeros $z_1, \dots, z_n$ in RHP. Then*

(i) $P(s)$ belongs to H^∞;
(ii) $P(s) = B(s)R(s)$ where $B(s)$ and $R(s)$ are stable rational functions, with

$$B(s) = \prod_{j=1}^n \frac{s - z_j}{s + \bar{z}_j}; \tag{10.79}$$

(iii) $|B(s)| \le 1$ for all $s \in RHP$, and $|B(i\omega)| = 1$ for all $\omega \in \mathbb{R}$;
(iv) $R(s)$ has no zeros in RHP, R is bounded with $|R(s)| \le \sup_{\omega\in\mathbb{R}} |R(i\omega)|$ for all $s \in RHP$, where $|R(i\omega)| = |P(i\omega)|$, and

$$\frac{\log R(s)}{1 + s} \in H^2. \tag{10.80}$$

(v) *If* $P \in H^2$ *then* $R \in H^2$ *and the linear span of* $\{e^{-\gamma s}R(s) : \gamma > 0\}$ *is a dense linear subspace of* H^2.

Proof (i) We have $P(s) = c + p(s)/q(s)$ for some $c \in \mathbb{C}$ where $p(s)$ and $q(s)$ are polynomials with $\deg p < \deg q$, and $q(s)$ is stable. Hence $P(s) \to c$ as $s \to \infty$. By the maximum modulus principle [56], we deduce that $|P(s)|$ is bounded on RHP and attains its supremum on $\mathbb{R} \cup \{\infty\}$, so $\sup\{|P(s)| : \Re s > 0\} = \sup\{|P(i\omega)| : \omega \in \mathbb{R}\}$.

(ii), (iii) We define $B(s)$ as above and observe that $-\bar{z}_j$ is the reflection of z_j in the imaginary axis, so $|s - z_j| \leq |s + \bar{z}_j|$ for all $s \in RHP$, with equality for $s = i\omega$. Hence $B(s)$ is holomorphic on RHP with zeros at $z_1, \dots, z_n$, and $B(s)$ is bounded there with $|B(s)| \leq 1$, where $|B(i\omega)| = 1$. Evidently $B(s)$ is stable rational.

(iv) The function $R(s) = P(s)/B(s)$ is also stable rational, since the zeros of $B(s)$ and $R(s)$ cancel one another; hence $R(s)$ has no zeros in RHP. (We note that $R(s)$ can have zeros on the imaginary axis, but that is not a problem.) As in (i), we can apply the maximum modulus principle to $R(s)$ to deduce that $|R(s)|$ is bounded on RHP and attains its supremum on $\mathbb{R} \cup \{\infty\}$, so $R \in H^\infty$. By (iii), we have $|P(i\omega)| = |R(i\omega)||B(i\omega)| = |R(i\omega)|$.

We can factorize

$$R(s) = \frac{a \prod_{j=1}^{p}(s - iy_j) \prod_{j=1}^{q}(s - \alpha_j)}{\prod_{j=1}^{r}(s - \beta_j)} \tag{10.81}$$

where $a \in \mathbb{C}$, $p + q \leq r$, the zeros iy_j are on the imaginary axis and $\alpha_j, \beta_j \in LHP$. Hence

$$\frac{\log R(s)}{1+s} = \frac{\log a}{1+s} + \sum_{j=1}^{p} \frac{\log(s - iy_j)}{1+s} + \sum_{j=1}^{q} \frac{\log(s - \alpha_j)}{1+s} - \sum_{j=1}^{r} \frac{\log(s - \beta_j)}{1+s}, \tag{10.82}$$

and one can easily check that each summand gives a function in H^2.

(v) Suppose that $F \in H^2$ is a nonzero function that is orthogonal to all the functions $e^{-\gamma s}R(s)$ in H^2, so

$$\int_{-\infty}^{\infty} F(i\omega)e^{i\gamma\omega}\overline{R(i\omega)}\frac{d\omega}{2\pi} = 0 \qquad (\gamma > 0); \tag{10.83}$$

then by multiplying by $e^{-\gamma z}$ and integrating with respect to $\gamma \in (0, \infty)$, we deduce that

$$\int_{-\infty}^{\infty} \frac{F(i\omega)e^{i\gamma\omega}\overline{R(i\omega)}}{i\omega - z}\frac{d\omega}{2\pi} = 0 \qquad (\Re z > 0). \tag{10.84}$$

We express $R(s) = \sum_{j=1}^{r} a_j/(s - \beta_j)$ in partial fractions where $\Re\beta_j < 0$, so that

$$\int_{-\infty}^{\infty} \frac{F(i\omega)e^{i\gamma\omega}}{i\omega - z} \sum_{j=1}^{r} \frac{\bar{a}_j}{-i\omega - \bar{\beta}_j} \frac{d\omega}{2\pi} = 0 \qquad (\Re z > 0); \tag{10.85}$$

then by the calculus of residues

$$F(z) \sum_{j=1}^{r} \frac{\bar{a}_j}{z + \bar{\beta}_j} - \sum_{j=1}^{r} \frac{F(-\bar{\beta}_j)\bar{a}_j}{z + \bar{\beta}_j} = 0, \tag{10.86}$$

so

$$F(z) = \frac{\sum_{j=1}^{r} F(-\bar{\beta}_j)\bar{a}_j/(z + \bar{\beta}_j)}{\sum_{j=1}^{r} \bar{a}_j/(z + \bar{\beta}_j)}. \tag{10.87}$$

The numerator has poles at $z = -\bar{\beta}_j \in RHP$, but these are canceled by poles on the denominator. There are also poles arising from the zeros of the denominator, namely $0 = \sum_{j=1}^{r} \bar{a}_j/(z + \bar{\beta}_j)$, so $R(-\bar{z}) = 0$; all the zeros of R are in the closure of the LHP, so z is in the closure of RHP, contrary to the assumption that $F \in H^2$. □

This $B(s)$ is a finite Blaschke product, and gives an inner function. There are sometimes known as all pass functions in the engineering literature, since the gain is one at all frequencies. The function $R(s)$ is an outer function, otherwise known as minimum phase; see [13].

10.9 Shifts on L^2

In this section, we consider transfer functions that are not rational.

Example 10.41 For $\tau > 0$, the function $T(s) = e^{-\tau s}$ belongs to H^∞, and on the imaginary axis it reduces to $T(i\omega) = e^{-i\tau\omega}$, where the gain is $|e^{-i\tau\omega}| = 1$. This transfer function represents a phase shift of $-\tau\omega$.

More generally, we can consider $\tau_j \geq 0$ and $a_j \in \mathbb{C}$, and a transfer function such as

$$T(s) = \sum_{j=1}^{n} a_j e^{-\tau_j s} \tag{10.88}$$

which belongs to H^∞.

Definition 10.42

(i) As a notational convenience for $f \in L^2(0,\infty)$ we write $f(t) = 0$ for $t < 0$. Given $\tau > 0$, the (forward or right) shift operator S_τ on $L^2(0,\infty)$ is defined by $S_\tau f(t) = f(t-\tau)$ for $f \in L^2(0,\infty)$.
(ii) The backward shift operator A_τ on $L^2(0,\infty)$ is $A_\tau f(t) = f(t+\tau)$.

At $t > 0$, it is plausible that one knows the values of $f(s)$ for $0 < s < t$, so that $S_\tau f(t) = f(t-\tau)$ involves known data. However, $A_\tau f(t) = f(t+\tau)$ involves future values of the signal, so may be inaccessible. In this interpretation, the forward and backward shifts relate to different situations. Observe that $e^{-s(t+\tau)} = e^{-s\tau}e^{-st}$. With $\Re s > 0$, $f(t) = e^{-st}$ belongs to $L^2(0,\infty)$ and $A_\tau f(t) = e^{-\tau s} f(t)$, so $f(t)$ is an eigenvector that corresponds to eigenvalue $e^{-s\tau}$.

Proposition 10.43

(i) The shift is an isometric linear transformation of $L^2(0,\infty)$ so $\|S_\tau f\| = \|f\|$ for all $f \in L^2(0,\infty)$.
(ii) Under the Laplace transform, S_τ on $L^2(0,\infty)$ corresponds to multiplication on H^2 by $e^{-\tau s}$.
(iii) The backward shift is a bounded linear transformation of $L^2(0,\infty)$ so $\|A_\tau\| = 1$, and $\|A_\tau f\| \to 0$ as $\tau \to \infty$ for all $f \in L^2(0,\infty)$.
(iv) $A_\tau S_\tau = I$, and $S_\tau A_\tau = I - P_{(0,\tau)}$, where $P_{(0,\tau)} f(t) = \mathbb{I}_{(0,\tau)}(t) f(t)$;
(v) $A_\tau' = S_\tau$ and $S_\tau' = A_\tau$.

Proof

(i) The operator S_τ is evidently linear, and the effect of S_τ on f is to shift the graph of f to the right by τ, thus opening up a gap $[0,\tau)$ on which $f(t-\tau) = 0$.
(ii) Under the Laplace transform

$$\mathcal{L}(S_\tau f)(s) = \int_0^\infty e^{-ts} f(t-\tau)dt = e^{-s\tau}\int_0^\infty e^{-us} f(u)du = e^{-s\tau}\mathcal{L}f(s). \tag{10.89}$$

(iii) The effect of A_τ is to move $[0,\infty)$ to $[\tau,\infty)$ while leaving the graph of f fixed; thus the portion of the graph of f above $[0,\tau)$ is discarded. We have

$$\|A_\tau f\|^2 = \int_0^\infty |f(t+\tau)|^2 dt = \int_\tau^\infty |f(u)|^2 du. \tag{10.90}$$

We deduce that $\|A_\tau f\| \le \|f\|$, and when $f = \mathbb{I}_{(\tau,2\tau)}$ we have equality. Also, taking the limit as $\tau \to \infty$, we see that $\|A_\tau f\| \to 0$.
(iv) This follows from carefully applying the defining formulas, noting that $f(t-\tau) = 0$ for $t < \tau$. In the product $A_\tau S_\tau$ we first shift the graph of f by τ to the right, then move the axis to catch up, and the overall effect is to preserve the graph. In the product $S_\tau A_\tau$, we first shift the axis, then move the graph. We

have $S_\tau A_\tau f(t) = \mathbb{I}_{[\tau,\infty)}(t) f(t)$, since the part of the graph of f above $[0, \tau)$ is irretrievably lost.

(v) We have

$$\langle A_\tau f, g\rangle = \int_0^\infty f(t+\tau)\bar{g}(t)dt = \int_0^\infty f(u)\bar{g}(u-\tau)du = \langle f, S_\tau g\rangle. \tag{10.91}$$

□

Part (ii) has an important consequence. We have

$$\mathcal{L}(h(t-\tau))(s) = \int_0^\infty h(x-\tau)\mathbb{I}_{(0,\infty)}(t-\tau)e^{-st}dt = e^{-s\tau}\mathcal{L}(h)(s) \tag{10.92}$$

so that

$$\frac{\mathcal{L}(h(t-\tau))(s) - \mathcal{L}(h(t))(s)}{\tau} = \frac{e^{-s\tau}-1}{\tau}\mathcal{L}(h)(s). \tag{10.93}$$

The obvious move is to let $\tau \to 0$, as in differentiation, but we need to make sure that the functions that emerge belong to the correct spaces. So we consider $\mathcal{D} = \{f \in H^2 : sf(s) \in H^2\}$, which is a linear subspace of H^2. Also note that $|f(i\omega)| \leq 1/(1+\omega^2) + (1+\omega^2)|f(i\omega)|^2$, so $\int_{-\infty}^{\infty} |f(i\omega)|d\omega$ converges for all $f \in \mathcal{D}$. Given $W > 0$, we also note that

$$\frac{e^{-i\omega\tau}-1}{\tau} + i\omega = \frac{-i\omega}{\tau}\int_0^\tau (e^{-i\omega u}-1)du \tag{10.94}$$

converges to 0 as $\tau \to 0+$ uniformly for $\omega \in [-W, W]$; also by estimating this integral, we see that

$$\left|\frac{e^{-i\omega\tau}-1}{\tau} + i\omega\right| \leq 2|\omega|; \tag{10.95}$$

so by either the Dominated Convergence Theorem, or uniform convergence, we deduce that

$$\int_{-\infty}^{\infty} \left|\frac{e^{-i\omega\tau}-1}{\tau} + i\omega\right|^2 |f(i\omega)|^2 d\omega \to 0 \tag{10.96}$$

as $\tau \to 0+$. Hence

$$\frac{e^{-i\omega\tau}-1}{\tau} f(i\omega) \to -i\omega f(i\omega) \tag{10.97}$$

in $L^2(i\mathbb{R})$, so $(e^{-\tau s} - 1)f(s)/\tau \to -sf(s)$ in H^2 as $\tau \to 0+$. If $h \in L^2(0,\infty)$ has $f = \hat{h} \in \mathcal{D}$, then $h \in L^2(0,\infty)$ and

$$\frac{h(t-\tau) - h(t)}{\tau} \to -h'(t) \tag{10.98}$$

in $L^2(0,\infty)$ as $\tau \to 0+$.

Example 10.44 Delay differential equation

The standard (A, B, C, D) is a realistic model for processes that take place almost instantaneously, such as electrical current and radio communications. Other processes take place with some delay: medicines take a while to have effect, customers pay bills slowly, all in good time. In such examples, we can consider a delay-differential equation

$$\begin{aligned} \frac{dX}{dt} &= \sum_{j=1}^{n} A_j X(t-\tau_j) + \sum_{j=1}^{n} B_j U(t-\tau_j) \\ Y &= \sum_{j=1}^{n} C_j X(t-\tau_j) + \sum_{j=1}^{n} D_j U(t-\tau_j) \end{aligned}$$

where we have introduced delay times $\tau_j \geq 0$ and constant matrices A_j, B_j, C_j and D_j. The input U and state X are extended as functions to that $U(t) = 0$ and $X(t) = 0$ for all $t < 0$. We introduce

$$\begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} A_j e^{-\tau_j s} & \sum_{j=1}^{n} B_j e^{-\tau_j s} \\ \sum_{j=1}^{n} C_j e^{-\tau_j s} & \sum_{j=1}^{n} D_j e^{-\tau_j s} \end{bmatrix}$$

Then the Laplace transform of the delay-differential-equation is

$$\begin{aligned} s\hat{X}(s) &= A(s)\hat{X}(s) + B(s)\hat{U}(s) \\ \hat{Y}(s) &= C(s)\hat{X}(s) + D(s)\hat{U}(s), \end{aligned}$$

so that

$$\hat{Y}(s) = \big(D(s) + C(s)(sI - A(s))^{-1}B(s)\big)\hat{U}(s).$$

The entries of $A(s), B(s), C(s)$ and $D(s)$. all belong to H^∞. There is a delicate question as to when the entries of $(sI - A(s))^{-1}$ also belong to H^∞.

10.10 The Telegraph Equation as a Linear System

Consider a long wire with position $x > 0$ along the wire, and let $u(x, t)$ be the electrical current at x at time $t > 0$. In particular, let $f(t)$ be the signal at time $t > 0$ at the emitter, where $x = 0$. Let $\kappa > 0$ be a constant relating to the capacitance and electrical resistance of the wire. Then the telegraph equation is the partial differential equation

$$\frac{\partial u}{\partial t} = \kappa \frac{\partial^2 u}{\partial x^2} \tag{10.99}$$

with the initial condition $u(x, 0) = 0$, the boundary condition $u(0, t) = f(t)$ and the boundary condition at infinity $u(x, t) \to 0$ as $x \to \infty$. Then the Laplace transform in the time variable satisfies

$$\int_0^\infty e^{-st} \frac{\partial u}{\partial t} dt = \kappa \frac{\partial^2}{\partial x^2} \int_0^\infty e^{-st} u(x, t) dt, \tag{10.100}$$

so by integrating by parts and invoking the initial condition, we obtain

$$s \int_0^\infty e^{-st} u(x, t) dt = \kappa \frac{\partial^2}{\partial x^2} \int_0^\infty e^{-st} u(x, t) dt; \tag{10.101}$$

then

$$\int_0^\infty e^{-st} u(x, t) dt = A e^{-x\sqrt{s/\kappa}} + B e^{x\sqrt{s/\kappa}}, \tag{10.102}$$

and since $u(x, t) \to 0$ as $x \to \infty$, we need $B = 0$, and $A = \int_0^\infty e^{-st} f(t) dt$ from the boundary condition. Laplace calculated the integral [10, page 171]

$$\int_0^\infty \exp\Big(- \frac{x^2}{4\kappa t} - st \Big) \frac{dt}{2t^{3/2}} = \frac{\sqrt{\kappa\pi}}{x} e^{-sx/\kappa}, \tag{10.103}$$

so

$$\int_0^\infty e^{-st} u(x, t) dt = A \int_0^\infty \frac{x}{2\sqrt{\kappa\pi}} e^{-x^2/(4\kappa t) - st} \frac{dt}{t^{3/2}} \tag{10.104}$$

hence

$$u(x, t) = \int_0^t \frac{x e^{-x^2/(4\kappa\tau)}}{2\sqrt{\kappa\pi\tau^3}} f(t - \tau) d\tau. \tag{10.105}$$

The telegraph (or heat) equation is discussed in the context of semi-group theory in [2]. Here we introduce various Hilbert spaces with their inner products to describe how the various operators may be defined. First, let

$$H_0 = L^2((0,\infty);\mathbb{C}) : \langle f, g\rangle = \int_0^\infty f(x)\overline{g(x)}dx \tag{10.106}$$

or equivalently

$$H_0 = L^2((0,\infty);\mathbb{C}) : \langle f, g\rangle = \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(i\omega)\overline{\hat{g}(i\omega)}d\omega. \tag{10.107}$$

Next we introduce the space of $f \in H$ with $df/ds \in H$, namely

$$H_1 = \{f \in L^2((0,\infty);\mathbb{C}) : df/ds \in L^2((0,\infty);\mathbb{C})\}$$
$$\langle f, g\rangle_1 = \int_0^\infty f(x)\overline{g(x)}dx + \int_0^\infty \frac{df}{dx}\overline{\frac{dg}{dx}}dx \tag{10.108}$$

or equivalently $H_1 = \{f \in L^2((0,\infty);\mathbb{C}) : \omega\hat{f}(i\omega) \in L^2(\mathbb{R};\mathbb{C})\}$

$$\langle f, g\rangle_1 = \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(i\omega)\overline{\hat{g}(i\omega)}d\omega + \frac{1}{2\pi}\int_{-\infty}^\infty \omega^2\hat{f}(i\omega)\overline{\hat{g}(i\omega)}d\omega. \tag{10.109}$$

Clearly H_1 is a linear subspace of H_0, but H_1 is not a closed linear subspace of H_0. The purpose of introducing H_1 is to ensure that some useful operators are bounded.

Lemma 10.45 *The linear map* $C : H_1 \to \mathbb{C} : f \mapsto f(0)$ *is bounded.*

Proof As in Proposition 4.27, we have

$$Cf = f(0) = \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(i\omega)d\omega \tag{10.110}$$

where by the Cauchy–Schwarz inequality

$$\frac{1}{2\pi}\int_{-\infty}^\infty |\hat{f}(i\omega)|d\omega \le \Big(\frac{1}{2\pi}\int_{-\infty}^\infty (1+\omega^2)|\hat{f}(i\omega)|^2 d\omega\Big)^{1/2}\Big(\frac{1}{2\pi}\int_{-\infty}^\infty \frac{1}{1+\omega^2}d\omega\Big)^{1/2} \tag{10.111}$$

so $|Cf|^2 \le \langle f, f\rangle_1/2$.

Next we introduce

$$H_2 = \{f \in L^2((0,\infty);\mathbb{C}) : df/dx, d^2f/dx^2 \in L^2((0,\infty):\mathbb{C})\}$$

$$\langle f, g\rangle_2 = \int_0^\infty f(x)\overline{g(x)}dx + \int_0^\infty \frac{df}{dx}\overline{\frac{dg}{dx}}dx + \int_0^\infty \frac{d^2f}{dx^2}\overline{\frac{d^2g}{dx^2}}dx \tag{10.112}$$

or equivalently $H_2 = \{f \in L^2((0,\infty);\mathbb{C}) : \omega^2\hat{f}(i\omega) \in L^2(\mathbb{R}:\mathbb{C})\}$

$$\langle f, g\rangle_2 = \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(i\omega)\overline{\hat{g}(i\omega)}d\omega + \frac{1}{2\pi}\int_{-\infty}^\infty \omega^2\hat{f}(i\omega)\overline{\hat{g}(i\omega)}d\omega + \frac{1}{2\pi}\int_{-\infty}^\infty \omega^4\hat{f}(i\omega)\overline{\hat{g}(i\omega)}d\omega. \tag{10.113}$$

here H_2 is a linear subspace of H_1, and the differential operator $H_2 \to H_0$ $g \mapsto d^2g/ds^2$ is bounded.

Now we consider $\kappa > 0$ and $Ag = \kappa\frac{d^2g}{ds^2}$. There is a linear system

$$\frac{dX}{dt} = AX + Bu$$
$$y = CX$$

subject to the initial condition $X_0 = X(\cdot, 0) = 0$, where the input is an impulse at $t = 0$ so $Bu(t) = 0$ for all $t > 0$ and we require $y(t) = f(t)$, where $f : [0,\infty) \to \mathbb{C}$ is a bounded and continuous function. This problem has a solution

$$X(x,t) = \int_0^t \frac{x\exp(-x^2/(4\kappa\tau))}{\sqrt{4\pi\kappa\tau^3}} f(t-\tau)d\tau. \tag{10.114}$$

Observe that

$$\frac{x\exp(-x^2/(4\kappa\tau))}{\sqrt{4\pi\kappa\tau^3}} = -\kappa\frac{\partial}{\partial x}\frac{\exp(-x^2/(4\kappa\tau))}{\sqrt{\pi\kappa\tau}} \tag{10.115}$$

belongs to H_2 since the exponential factor decays rapidly as $x \to \infty$. Hence we can interpret this linear system as a system with state space H_0, with main transformation $A : H_2 \to H_0$ and output transformation $C : H_2 \to \mathbb{C}$. The initial condition is ill-defined since

$$\int_0^\infty \frac{x^2\exp(-x^2/(2\kappa\tau))}{4\pi\kappa\tau^3}dx \to \infty \qquad (\tau \to 0+). \tag{10.116}$$

□

10.11 Exercises

Exercise 10.1 (Laguerre Shift)

(i) Let

$$Sf(x) = f(x) - 2\int_0^x e^{-(x-t)} f(t)dt \qquad (f \in L^2(0,\infty)).$$

Show that the Laplace transform satisfies

$$\mathcal{L}Sf(s) = \frac{s-1}{s+1}\mathcal{L}f(s) \qquad (\Re s > 0).$$

(ii) Recall the Laguerre polynomials from Example 8.14, and let $h_n(t) = \sqrt{2}e^{-t}L_n(2t)$. Deduce that

$$Sh_n(t) = h_{n+1}(t).$$

Exercise 10.2 Let the Laguerre polynomials of index 1 be

$$L_n^{(1)}(x) = \frac{e^x}{n!x}\frac{d^n}{dx^n}\left(x^{n+1}e^{-x}\right) \qquad (n = 0, 1, \dots). \tag{10.117}$$

(i) Show that $L_n^{(1)}(x)$ is a polynomial of degree n. Show also that the Laplace transform of $h_n^{(1)}(x) = xe^{-x}L_n^{(1)}(2x)$ is

$$\mathcal{L}h_n^{(1)}(s) = (n+1)\frac{(s-1)^n}{(s+1)^{n+2}}. \tag{10.118}$$

(ii) Deduce that

$$\frac{d}{dt}h_n^{(1)}(t) + h_n^{(1)}(t) = (n+1)e^{-t}L_n(2t). \tag{10.119}$$

(iii) Calculate

$$\int_{-\infty}^{\infty} f(i\omega)\frac{(-i\omega-1)^n}{(-i\omega+1)^{n+2}}\frac{d\omega}{2\pi} \tag{10.120}$$

for $f \in H^2$.

Exercise 10.3 Recall the Laguerre polynomials from Example 8.14. Let $h_n(t) = \sqrt{2}e^{-t}L_n(2t)$ and $\phi(t) = e^{-st}$ for $n = 0, 1, \dots, t > 0$ and $\Re s > 0$. Show that

$$\int_0^\infty \phi(t+u)h_n(u)du = \sqrt{2}e^{-st}\frac{(s-1)^n}{(s+1)^{n+1}} \qquad (t > 0) \tag{10.121}$$

and deduce that

$$\int_0^\infty\int_0^\infty \phi(s+u)h_n(u)h_m(t)dudt = 2\frac{(s-1)^{n+m}}{(s+1)^{n+m+2}} \tag{10.122}$$

where the right-hand side gives a Hankel matrix

$$\Big[2\frac{(s-1)^{n+m}}{(s+1)^{n+m+2}}\Big]_{n,m=0}^\infty = \Big[\frac{2}{n+m+1}\int_0^\infty \phi(t)h^{(1)}_{n+m}(t)dt\Big]_{n,m=0}^\infty, \tag{10.123}$$

where $h^{(1)}(t) = te^{-t}L_n^{(1)}(2t)$.

Exercise 10.4 By [19, (8.977)], the Laguerre polynomials satisfy the addition rule

$$L_n^{(1)}(x+y) = \sum_{j=0}^n L_j(x)L_{n-j}(y) \qquad (x, y > 0; n = 0, 1, \dots). \tag{10.124}$$

Obtain a simple expression for the integral

$$\int_0^\infty e^{-x-y}L_n^{(1)}(2x+2y)f(y)dy \tag{10.125}$$

for $f \in L^2(0, \infty)$.

Exercise 10.5 As in the Proposition 10.43, let S_τ be the multiplication operator on H^2 that represents the shift on $L^2(0, \infty)$, so $S_\tau f(s) = e^{-s\tau}f(s)$ for $f \in H^2$. Calculate the Laplace transform in the $\tau > 0$

$$\mathcal{L}(S_\tau f)(z) = \int_0^\infty e^{-\tau z}S_\tau f(s)d\tau \qquad (\Re z > 0), \tag{10.126}$$

and interpret the result.

Exercise 10.6 (Shift on Hardy Space) For $\tau > 0$ and $f \in H^2$, let $S_\tau f(s) = f(s+\tau)$.

(i) Derive the formula

$$f(s+\tau) = \int_0^\infty e^{-st-\tau t}h(t)dt. \tag{10.127}$$

(ii) Deduce that S_τ is a bounded linear operator on H^2 such that

$$0 \leq \langle S_\tau f, f\rangle \leq \langle f, f\rangle \qquad (f \in H^2).$$

Exercise 10.7 (Invertible Rational Filters) Let $P(s) = B(s)R(s)$ be a factorization as in the Proposition 10.40. Show that $1/R(s)$ is also stable rational, if and only if:

(i) $B(s) = 1$;
(ii) $R(s) \to a$ as $s \to \infty$ for some $a \in \mathbb{C} \setminus \{0\}$;
(iii) $R(s)$ has no zeros in the imaginary axis $\{i\omega : \omega \in \mathbb{R}\}$.

This is the point where things become complicated for general H^∞ filters; the invertible ones are difficult to describe.

Exercise 10.8 (EVAD) Dorf and Bishop [12] propose a model for an EVAD device for managing cardiovascular illness which has a plant $G(s) = e^{-s}$ and a controller $K(s) = a/(s(s+b))$, where $a, b > 0$ are constants with indicative values $a = 5, b = 10$. For internal stability, we require

$$F = \frac{1}{1 + KG}\begin{bmatrix} 1 & K \\ G & KG \end{bmatrix} \tag{10.128}$$

to have entries in H^∞.

(i) Let $f(s) = s(s+b)+ae^{-s}$, and consider the image of the semicircular contour $[-5i, 5i] \oplus S_5$ under f. By applying the argument principle, determine a region on which $1 + KG$ has no zeros.
(ii) Find a and b such that the entries of F are in H^∞ the space of bounded and holomorphic functions on RHP.
(iii) Produce plots of the entries of F in the style of Nyquist contours.

Exercise 10.9 Find the Laplace transform of the backward shift $A_\tau h$, and show that under suitable conditions

$$\frac{A_\tau h(t) - h(t)}{\tau} \to \frac{dh}{dt}$$

in $L^2(0, \infty)$ as $\tau \to 0$.

Exercise 10.10 (A Finite-Rank Hankel Operator) Let $w_1, \dots, w_n$ be distinct points in $\mathbb{D}$ and let $a_1, \dots, a_n \in \mathbb{C}$. Define $\Gamma : H^2 \to H^2$ by

$$\Gamma f(z) = \sum_{j=1}^{n} a_j f(w_j) k_{\bar{w}_j}(z) \qquad (f \in H^2). \tag{10.129}$$

(i) Show that $\Gamma f = 0$ if and only if $f(w_j) = 0$ for $j = 1, \dots, n$.
(ii) Show that the shift operator satisfies $S'\Gamma = \Gamma S$.

(iii) Show that

$$\langle \Gamma(z^\ell), z^m \rangle = \sum_{j=1}^{n} a_j w_j^{\ell+m}. \tag{10.130}$$

This is a finite-rank Hankel operator in the frequency domain; compare Ex 6.17.

Exercise 10.11 Let $\lambda_j \in RHP$ and $a_j \in \mathbb{C}$ for $j = 1, \dots, n$. For $f \in L^2(0, \infty)$, let

$$\Gamma f(x) = \sum_{j=1}^{n} a_j \int_0^\infty e^{-\lambda_j(x+y)} f(y)\, dy, \qquad (x > 0). \tag{10.131}$$

If $f(y) = \int_{-\infty}^{\infty} g(i\omega)e^{i\omega y} d\omega/(2\pi)$, show that Γf has Laplace transform

$$\mathcal{L}(\Gamma f)(s) = \sum_{j=1}^{n} \frac{a_j g(\lambda_j)}{\lambda_j + s}. \tag{10.132}$$

Exercise 10.12

(i) Show that $\psi(t) = e^{-|t|} \sin t$ is integrable for $t \in \mathbb{R}$, that ψ is once continuously differentiable with bounded derivative and that $\psi(n\pi) = 0$ for all $n \in \mathbb{Z}$.

(ii) Let ϕ be as in Theorem 11.4 and sinc as in (11.6) Show that

$$\phi(t) = \frac{a}{\pi} \int_{-\infty}^{\infty} \operatorname{sinc}(a(t-u))\phi(u)\, du. \tag{10.133}$$

(iii) Let

$$Tf(x) = \int_{-a}^{a} e^{ixy} f(y) \frac{dy}{2a}, \qquad T'g(x) = \int_{-a}^{a} e^{-ixy} g(y) \frac{dy}{2a};$$

show that

$$T'Tf(x) = \int_{-a}^{a} \operatorname{sinc}(a(x-u))\, f(u) \frac{du}{2a} \qquad (f \in L^2([-a, a]; dx/(2a)).$$

(iv) Show that $S(t) = \operatorname{sinc}(at)$ satisfies the differential equation

$$t\frac{d^2S}{dt^2} + 2\frac{dS}{dt} + a^2 t S(t) = 0,$$

which is one of the Bessel family.

Exercise 10.13 Define Bessel's function of the first kind of integral order n by

$$J_n(x) = \int_{-\pi}^{\pi} \exp(ix \sin\omega - in\omega)\frac{d\omega}{2\pi}. \tag{10.134}$$

Let $f(\omega) = e^{-ix\sin\omega}$ for $x \in \mathbb{R}$ and $\omega \in [-\pi, \pi]$, and deduce a formula via the sampling theorem 11.4 for the corresponding signal ϕ.

Exercise 10.14 Let $B : H \to H$ be a linear operator such that $\|B\| \leq 1$. By considering Exercise 3.9, show that there exist linear operators C and D on H such that

$$U = \begin{bmatrix} B & D \\ -C & B' \end{bmatrix} \qquad \begin{matrix} H \\ H \end{matrix}$$

such that $U'U = UU' = I$ on $H \oplus H$.

Chapter 11
Wireless Transmission and Wavelets

The final chapter considers two of the most important topics in modern signal processing, namely wavelets and wireless transmissions. The origin of wavelets lies in the work of Haar and Paley on orthogonal series of functions, and the Haar wavelet was introduced as an orthonormal basis for $L^2[0, 1]$ with properties that Paley realized were remarkable. The Haar system was interesting in its own right, and was studied as a model of an orthonormal system which was apparently simpler than Fourier series. The work of Haar and Paley started a course of study that led to martingales and the sought-after results about Fourier series in the 1960s and 1970s. It was in the 1980s that the study of wavelets really sprung to life, and completely transformed signal processing. In this chapter we look at one wavelet, associated with the $sinc$ function, which is known as Shannon's wavelet.

Basic models of radio communication involve a single transmitter broadcasting to a single receiver. In modern wireless communication for mobile telephone networks, there are many transmitters and many receivers, so a more complex model of transmission is essential. We discuss the model due to Telatar [55], which has been highly influential. The results of this chapter draw on ideas from previous sections of the book, and convey the main points of the models in question. One can extend the analysis by introducing more advanced mathematics and more sophisticated computational tools, such as are discussed in the research literature.

11.1 Frequency Band Limited Functions and Sampling

It is often desirable or technically essential to consider signals such that the angular frequencies are constrained to lie in a bounded interval. For $a > 0$, we introduce the Hilbert space $L^2[-a, a]$ of square integrable complex functions $f : [-a, a] \to \mathbb{C}$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_11

with the inner product

$$\langle f, g \rangle_{L^2[-a,a]} = \int_{-a}^{a} f(\omega)\overline{g(\omega)} \frac{d\omega}{2a}, \tag{11.1}$$

where we regard $f(\omega)$ as representing a signal with angular frequency $\omega \in [-a, a]$. Then $L^2[-a, a]$ has complete orthonormal basis $(e^{i\pi n\omega/a})_{n=-\infty}^{\infty}$, so that

$$f(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{i\pi n\omega/a} \tag{11.2}$$

where the Fourier coefficients are

$$c_n = \int_{-a}^{a} f(\omega)e^{-i\pi n\omega/a} \frac{d\omega}{2a} \qquad (n \in \mathbb{Z}), \tag{11.3}$$

and they satisfy

$$\sum_{n=-\infty}^{\infty} |c_n|^2 = \int_{-a}^{a} |f(\omega)|^2 \frac{d\omega}{2a}. \tag{11.4}$$

Definition 11.1 (Unnormalized Sinc Function) Consider $f(\omega) = \mathbb{I}_{[-a,a]}(\omega)$, which has

$$\operatorname{sinc}(at) = \int_{-a}^{a} e^{it\omega} \frac{d\omega}{2a} = \frac{\sin at}{at}. \tag{11.5}$$

With $a = 1$, we have the unnormalized sinc function, which in this book we simply call sinc. With $a = \pi$ we obtain $\sin(\pi t)/(\pi t)$, which is the normalized sinc function; in signal processing and MATLAB, this is called sinc (Fig. 11.1).

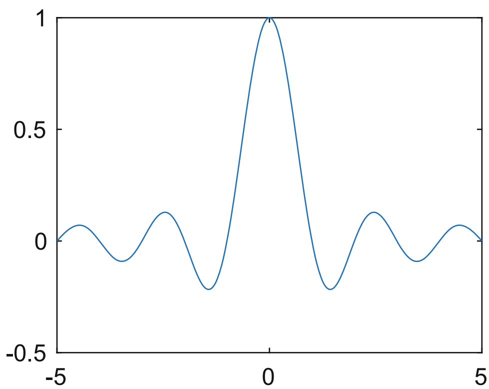


Fig. 11.1 Normalized sinc function

Lemma 11.2 *The unnormalized sinc function has the following properties:*

(i) sinc (t) = sinc $(-t)$ *for all* $t \in \mathbb{R}$*, so* sinc *is even;*
(ii) sinc $(0) = 1$ *and* $|\text{sinc}\,(t)| \leq 1$ *for all* $t \in \mathbb{R}$*;*
(iii) sinc (t) *is decreasing for* $0 \leq t \leq \pi$*;*
(iv) sinc$(t) \to 0$ *as* $t \to \pm\infty$*, and*

$$\lim_{T\to\infty} \int_{-T}^{T} \text{sinc}\,(t)dt = \pi, \tag{11.6}$$

(v) sinc *is not (absolutely) integrable over* $\mathbb{R}$*;*
(vi) sinc (z) *defines an entire function, with zeros at* $z = n\pi$ *for* $n \in \mathbb{Z} \setminus \{0\}$.

Proof

(i) Both t and $\sin t$ are odd functions, so their quotient is even.
(ii) The integrand satisfies $|e^{it\omega}| \leq 1$, which gives the bound on sinc.
(iii) Consider

$$\frac{d}{dt}\frac{\sin t}{t} = \frac{t\cos t - \sin t}{t^2}$$

which is clearly negative for $\pi/2 \leq t < \pi$; while for $0 < t < \pi/2$ it is also negative since $t < \tan t$.
(iv) We have $|\text{sinc}(t)| \leq 1/|t|$, so $|\text{sinc}(t)| \to 0$ as $|t| \to \infty$. The improper integral was found by complex analysis in Lemma 4.24.
(v) We show that

$$\int_0^T \frac{|\sin t|}{t} dt \to \infty \qquad (T \to \infty). \tag{11.7}$$

We can split the integral into integrals over intervals $[n\pi, (n+1)\pi]$, where a typical odd integral contributes

$$\int_{2n\pi}^{(2n+1)\pi} \frac{|\sin t|}{t} dt \geq \frac{1}{(2n+1)\pi} \int_{2n\pi}^{(2n+1)\pi} \sin t\, dt = \frac{2}{(2n+1)\pi}, \tag{11.8}$$

where $\sum_{n=1}^{\infty} 2/(2n+1)\pi$ diverges by comparison with the standard divergent series $\sum_{n=1}^{\infty} 1/n$. The results (iv) and (v) show that sinc(t) converges to 0 slowly as $t \to \infty$. This property also holds for some related functions in this section.
(vi) The formula $\text{sinc}(z) = (\sin z)/z$ gives an entire function with convergent power series $\sum_{n=0}^{\infty}(-1)^n z^{2n}/(2n+1)!$. The zeros of $\sin z = (e^{iz} - e^{-iz})/(2i)$ occur where $e^{2iz} = 1$, namely at $2iz = 2n\pi i$ for $n \in \mathbb{Z}$. At $z = 0$, the zeros on the numerator and denominator of $(\sin z)/z$ cancel.

□

Lemma 11.3 (Shift and Translation) *Let* $f \in L^2[-a,a]$ *give signal* $\phi(t) = \int_{-a}^{a} e^{i\omega t} f(\omega)d\omega/(2a)$. *Then* $e^{i\omega s} f(\omega)$ *gives a function in* $L^2[-a,a]$ *with corresponding signal* $\phi(s+t)$ *for* $s, t \in \mathbb{R}$.

Proof We have

$$\phi(s+t) = \int_{-a}^{a} e^{i\omega s + i\omega t} f(\omega)\frac{d\omega}{2a}. \tag{11.9}$$

□

The following result describes the corresponding signal in the time domain.

Theorem 11.4 (Sampling Theorem) *For* $f \in L^2[-a,a]$, *let*

$$\phi(z) = \int_{-a}^{a} f(\omega)e^{i\omega z}\frac{d\omega}{2a} \qquad (z \in \mathbb{C}). \tag{11.10}$$

(i) Then $\phi(z)$ *is of exponential growth of growth rate at most* a, *so that*

$$|\phi(z)|^2 \leq \frac{\sinh 2ay}{2ay}\|f\|^2_{L^2[-a,a]} \qquad (z = x + iy \in \mathbb{C}); \tag{11.11}$$

(ii) $\phi(z)$ *is an entire function;*
(iii) the energy of the signal ϕ *is finite, so*

$$\int_{-\infty}^{\infty} |\phi(t)|^2 dt = \frac{\pi}{2a^2}\int_{-a}^{a} |f(\omega)|^2 d\omega; \tag{11.12}$$

(iv) ϕ *has an orthogonal expansion in terms of* sinc *functions, so*

$$\phi(t) = \sum_{n=-\infty}^{\infty} \phi(\pi n/a)\mathrm{sinc}\,(at - n\pi); \tag{11.13}$$

thus the sequence $(\phi(\pi n/a))_{n=-\infty}^{\infty}$ *determines all the values of* $\phi(t)$ *for* $t \in \mathbb{R}$.

Proof

(i) With $x, y \in \mathbb{R}$, we have $e^{i\omega z} = e^{i\omega x}e^{-\omega y}$ where the first factor is unimodular, so by Cauchy–Schwarz inequality, we have

$$\begin{aligned} |\phi(x+iy)|^2 &= \Big| \int_{-a}^{a} e^{i\omega x}e^{-\omega y} f(\omega)\frac{d\omega}{2a}\Big|^2 \\ &\leq \int_{-a}^{a} e^{-2\omega y}\frac{d\omega}{2a}\int_{-a}^{a} |f(\omega)|^2\frac{d\omega}{2a} \\ &= \frac{\sinh 2ay}{2ay}\|f\|^2_{L^2[-a,a]}. \end{aligned}$$

Observe that $(1/2)\log\sinh 2ay \sim ay$ as $y \to \infty$, so a is the maximum growth rate of $\phi(x+iy)$.

(ii) We can differentiate through the integral sign to obtain

$$\frac{d\phi}{dz} = \int_{-a}^{a} i\omega e^{i\omega z} f(\omega)\frac{d\omega}{2a}, \tag{11.14}$$

which is justified by estimates such as in (i).

(iii) The function $g(x) = f(x-a)$ for $0 < x < 2a$ and $g(x) = 0$ for $x > 2a$ has $g \in L^2(0,\infty)$ with Laplace transform $\hat{g}(\omega) = 2ae^{-i\omega a}\phi(-\omega)$, so we can apply the Paley–Wiener theorem to g.

(iv) We multiply the series (11.2) by $e^{i\omega t}$ and integrate, and the term with index $-n$ involves

$$\int_{-a}^{a} e^{i\omega t}e^{-i\pi n\omega/a}\frac{d\omega}{2a} = \Big[\frac{e^{it\omega - i\pi n\omega/a}}{2a(it - i\pi n/a)}\Big]_{-a}^{a} = \frac{\sin(at-n\pi)}{at-n\pi}. \tag{11.15}$$

This produces the series

$$\phi(t) = \sum_{n=-\infty}^{\infty} c_{-n}\frac{\sin(at-\pi n)}{at-\pi n}, \tag{11.16}$$

stated above. The Paley–Wiener theorem shows that the Fourier transform operates as a linear isometry on $L^2[-a,a]$, so the orthonormal sequence $(e^{in\omega})_{n=-\infty}^{\infty}$ is mapped to an orthonormal sequence, and the series converges since (c_n) is square summable. The function ϕ is entire, so the value of ϕ at a point such as $\pi n/a$ is unambiguous; one checks that $c_{-n} = \phi(\pi n/a)$, and

$$\sum_{n=-\infty}^{\infty} |\phi(\pi n/a)|^2 = \int_{-a}^{a} |f(\omega)|^2\frac{d\omega}{2a}. \tag{11.17}$$

Note that the sampling sequence $(\pi n/a)_{n=-\infty}^{\infty}$ depends upon a.

□

Remark 11.5

(i) By definition, a band limited function f lives on $[-a,a]$; such a function could be continued to become a $2a$-periodic function on $\mathbb{R}$, but we choose not to make this extension; instead we cut off the function outside $[-a,a]$. The signal $\phi(t)$ is defined for $t \in \mathbb{R}$ and will not be periodic. Nevertheless, we use periodic functions and the sum (11.2) to investigate the properties of $\phi(t)$. The results of this section are related to those of Sect. 4.10, which specifically involved periodic functions.

(ii) Condition (iv) is important in applications to music. Suppose that we know in advance that the signal $\phi(z)$ is generated by an $f(\omega)$ with angular frequency

$\omega \in [-a, a]$; this $[-a, a]$ is known as the range of frequencies. Then by finding $\phi(\pi n/a)$ for $n \in \mathbb{Z}$, we can determine the complete function $\phi(z)$. We do not lose any information by sampling only at these values, and we do not gain any more information by sampling more frequently. Then $a/(2\pi)$ is known as the Nyquist frequency.

Example 11.6 (Tent Function) We take

$$\begin{aligned} f(x) = a - x \qquad & 0 < x < a; \\ a + x \qquad & -a < x < 0 \end{aligned}$$

so that by integrating by parts we obtain

$$\phi(t) = \int_{-a}^{a} e^{i\omega t} f(\omega)\frac{d\omega}{2a} = \int_0^a (a - \omega)\cos\omega t\frac{d\omega}{a} = \frac{2\sin^2 at/2}{at^2};$$

and

$$\int_{-a}^{a} f(\omega)^2\frac{d\omega}{2a} = \int_0^a (a-\omega)^2\frac{d\omega}{a} = \frac{a^2}{3}.$$

The identity from the sampling formula is

$$\frac{a^2}{3} = \int_{-a}^{a} f(\omega)^2\frac{d\omega}{2a} = \sum_{n=-\infty}^{\infty}\phi(n\pi/a)^2 = \sum_{n=-\infty}^{\infty}\frac{4\sin^4 n\pi/2}{\pi^4 n^4/a^2}$$

which involves different contributions from $n = 0$, n odd and n a non-zero even integer, so we obtain

$$\frac{\pi^4}{12} = \frac{\pi^4}{16} + 2\sum_{n=1}^{\infty}\frac{1}{(2n-1)^4},$$

which is equivalent to $\pi^4/90 = \sum_{n=1}^{\infty} 1/n^4$.

Shannon's Approximation by Finitely Many Samples

The sampling formula (11.13) does not itself give any quantitative estimate on how well the sampling formula converges, so we describe a complement due to Shannon; see [14]. We write $\phi_h(t) = \phi(t+h)$ so that $\phi_h(t) = \int_{-a}^{a} e^{i(t+h)\omega} f(\omega)d\omega/(2a)$. For $h > 0$, the signal $\phi_h(t)$ runs ahead of $\phi(t)$; whereas for $h < 0$, the signal $\phi_h(t)$ lags behind $\phi(t)$. We deduce that

$$PW_a = \left\{\phi : \phi(t) = \int_{-a}^{a} e^{it\omega} f(\omega)\frac{d\omega}{2a}; \quad f \in L^2[-a, a]\right\} \tag{11.18}$$

is a linear subspace of $L^2(\mathbb{R};\mathbb{C})$ such that $\phi \in PW_a \Rightarrow \phi_h \in PW_a$ for all $h \in \mathbb{R}$. The samples $(\phi_h(n\pi/a))_{n=-\infty}^{\infty}$ occur at points which are translates of the original sampling points, and we can eliminate h by integrating the ℓ^2 sum in the sampling formula over an interval

$$\int_0^{\pi/a} \sum_{n=-\infty}^{\infty} \left|\phi_h(n\pi/a)\right|^2 dh = \int_{-\infty}^{\infty} |\phi(t)|^2 dt.$$

From the identity

$$\phi_h(t) = \sum_{n=-\infty}^{\infty} \phi_h(\pi n/a)\mathrm{sinc}\,(at - n\pi); \tag{11.19}$$

we have

$$\int_{-\infty}^{\infty} \left|\phi_h(t) - \sum_{n=-N}^{N} \phi_h(\pi n/a)\mathrm{sinc}\,(at - n\pi)\right|^2 dt = \frac{\pi}{a} \sum_{n=-\infty}^{-N-1} + \sum_{n=N+1}^{\infty} \left|\phi_h(\pi n/a)\right|^2 \tag{11.20}$$

so

$$\frac{a}{\pi}\int_0^{\pi/a} \int_{-\infty}^{\infty} \left|\phi_h(t) - \sum_{n=-N}^{N} \phi_h(\pi n/a)\mathrm{sinc}\,(at - n\pi)\right|^2 dt dh = \int_{(N+1)\pi/a}^{\infty} + \int_{-\infty}^{-N\pi/a} |\phi(t)|^2 dt \tag{11.21}$$

where the right-hand side converges to zero as $N \to \infty$. We infer that for large N and small $|h|$, the sum

$$\sum_{n=-N}^{N} \phi_h(\pi n/a)\mathrm{sinc}\,(at - n\pi) \tag{11.22}$$

gives a useful approximation for $\phi(t)$. One can obtain more quantitative version of this statement as in page 130 of [14].

Remark 11.7

(i) The series (11.13) is known as a cardinal series after Whittaker, or Shannon's interpolation formula. Our proof uses the special property of ϕ that it is the inverse Fourier transform of $f \in L^2[-a, a]$. The converse of this Theorem is also true, but we omit the details which can be extracted from [34]. Our formulation is intended to avoid Poisson summation, which can be difficult to apply rigorously.

(ii) The width of the band needs to be interpreted carefully. The formulas $2\cos\omega t = e^{i\omega t} + e^{-i\omega t}$ and $2i\sin\omega t = e^{i\omega t} - e^{-i\omega t}$ show how one can produce real waves with angular frequencies $0 \leq \omega \leq a$, so in this respect the bandwidth is a.

Remark 11.8 (Digitizing Sound) Suppose for the sake of simplicity that we have a musical instrument capable of producing a sound at a single pitch, as represented by $A\sin\omega t$ where the (angular) frequency is $0 < \omega < \infty$. It is perceived that the sounds at frequencies ω and 2ω are similar, and we say that they are an octave apart. We then choose $\omega_0 > 0$, and call the interval $[\omega_0, 2\omega_0]$ an octave. Music is an analogue phenomenon, in the sense that one can take ω to be a continuous variable; however, to build practical instruments and simplify musical notation it is convenient to restrict the choice of frequencies we allow in the octave. Musicians therefore divide the octave into 12 subintervals, and refer to the notes as a system of semitones, for instance by using successive frequencies in the ratio $2^{1/12}$ to give the equally tempered scale. The choice of ω_0, the choice of 12 and the precise mode of dividing the octave are historical and cultural choices, as discussed in [7]. Once we have selected these, we can convert music to a digital phenomenon, which is easier to communicate. In the next section, we proceed to show how all the signals can be described in terms of a single function under scaling and translation.

11.2 The Shannon Wavelet

The modern theory of wavelets fully exploits the scaling properties of families of functions in signal processing. In [44], there is an accessible introduction to the general theory, and here we focus upon a specific example relating to band limited functions. In this section we consider $\Delta_j = [-2^{j+1}\pi, -2^j\pi) \cup (2^j\pi, 2^{j+1}\pi]$ which we regard as the range of angular frequencies for one octave; as j varies through $\mathbb{Z}$ we have pairwise disjoint sets Δ_j which correspond to all possible octaves. For each Δ_j we introduce the corresponding space W_j of finite energy signals that have frequencies in Δ_j. We show that each W_j has a natural orthonormal basis, and that the bases for different W_j are related by a scaling formula.

Exercise Show that the function

$$f(\omega) = e^{i\omega}\mathbb{I}_{[-1,-1/2]}(\omega) + e^{i\omega}\mathbb{I}_{[1/2,1]}(\omega) \tag{11.23}$$

gives signal, with graph as in Fig. 11.2,

$$\psi(t) = \mathrm{sinc}(t-1) - (1/2)\mathrm{sinc}((t-1)/2). \tag{11.24}$$

To begin the construction, we consider the intervals $[-2^j\pi, 2^j\pi]$ for $j \in \mathbb{Z}$, which are commonly used in harmonic analysis. One can think of $[-\pi, \pi]$ as the

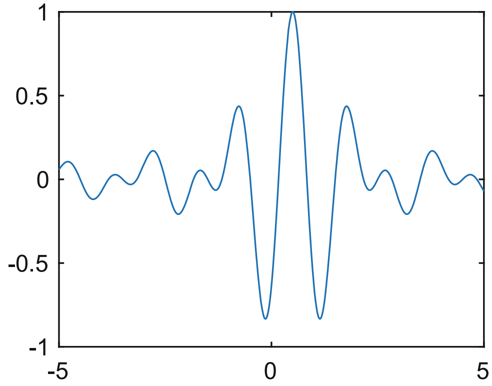


Fig. 11.2 Normalized Shannon mother wavelet $2\text{sinc}\pi(2t-1) - \text{sinc}\pi(t-1/2)$

basic interval, and then obtain all the other intervals by repeatedly doubling or halving the interval by dilating about the centre 0 by powers of 2. The intervals evidently satisfy

(i) they give an increasing sequence, so

$$\cdots \subset [-2^{-2}\pi, 2^{-2}\pi] \subset [-2^{-1}\pi, 2^{-1}\pi] \subset [-2^{0}\pi, 2^{0}\pi] \subset [-2^{1}\pi, 2^{1}\pi] \subset \ldots \mathbb{R}; \tag{11.25}$$

(ii) $\cap_{j=-\infty}^{\infty}[-2^j\pi, 2^j\pi] = \{0\}$, and $\cup_{j=-\infty}^{\infty}[-2^j\pi, 2^j\pi] = \mathbb{R}$;

(iii) $\omega \in [-2^j\pi, 2^j\pi]$ if and only if $2\omega \in [-2^{j+1}\pi, 2^{j+1}\pi]$; this is a scaling property.

These properties of the $[-2^j\pi, 2^j\pi]$ are reflected in the properties of the spaces $L^2[2^j\pi, 2^j\pi]$, which give rise to band-limited functions where the frequency range in $[-2^j\pi, 2^j\pi]$ is changed by factors of 2. Under the Fourier transform, we have spaces V_j, which we define by

$$V_j = \left\{\phi(t) = \int_{-2^j\pi}^{2^j\pi} e^{i\omega t} f(\omega)\frac{d\omega}{2^{j+1}\pi} : f \in L^2[-2^j\pi, 2^j\pi]\right\}. \tag{11.26}$$

Proposition 11.9 *The subspaces V_j satisfy:*

(i) $\cdots \subset V_2 \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset L^2(\mathbb{R})$;

(ii) $\cap_{j=-\infty}^{\infty} V_j = \{0\}$ *and* $\cup_{j=-\infty}^{\infty} V_j$ *is a dense linear subspace of* $L^2(\mathbb{R})$;

(iii) $\phi(t) \in V_j$ *if and only if* $\sqrt{2}\phi(2t) \in V_{j+1}$, *and the map* $\phi(t) \mapsto \sqrt{2}\phi(2t)$ *is an isometry;*

(iv) $\phi(t) \in V_0$ *if and only if* $\phi(t-k) \in V_0$ *for all* $k \in \mathbb{Z}$, *known as* $\mathbb{Z}$*-translation invariance;*

(v) V_0 *has an orthonormal basis* $(\text{sinc}(\pi(t-k)))_{k=-\infty}^{\infty}$.

The effect of translation as in (iv) is to move the graph of $\phi(t)$ through steps of length one to the left or right.

Proof

(i) We can map $L^2[-2^j\pi, 2^j\pi]$ isometrically into $L^2[-2^{j+1}\pi, 2^{j+1}\pi]$ by

$$f(\omega) \mapsto \sqrt{2} f(\omega)\mathbb{I}_{[-2^j\pi, 2^j\pi]}(\omega). \tag{11.27}$$

Then (i) follows from the sampling theorem 11.4.

(ii) This can be deduced from the Paley–Wiener Theorem 10.36. The key point is that for all $f \in L^2(\mathbb{R})$ we can introduce $f_j(\omega) = \mathbb{I}_{[-2^j, \phi 2^j]}(\omega) f(\omega)$ such that $f_j \in L^2(\mathbb{R})$ and

$$\int_{-\infty}^{\infty} |f(\omega) - f_j(\omega)|^2 d\omega \to 0 \qquad (j \to \infty). \tag{11.28}$$

(iii) From the definitions, we have

$$\sqrt{2}\phi(2t) = \sqrt{2} \int_{-2^j\pi}^{2^j\pi} e^{2i\omega t} f(\omega) \frac{d\omega}{2^{j+1}\pi} = \sqrt{2} \int_{-2^{j+1}\pi}^{2^{j+1}\pi} e^{i\omega t} f(\omega/2) \frac{d\omega}{2^{j+2}\pi}, \tag{11.29}$$

where $2\int_{-\infty}^{\infty} |\phi(2t)|^2 dt = \int_{-\infty}^{\infty} |\phi(t)|^2 dt$.

(iv) This follows from the Lemma (11.9).

(v) This follows from the sampling theorem 11.4.

□

We have $[2^j\pi, 2^j\pi] \subset [2^{j+1}\pi, 2^{j+1}\pi]$ and we introduce Δ_j as the difference between these sets. Let $\Delta_j = [-2^{j+1}\pi, -2^j\pi) \cup (2^j\pi, 2^{j+1}\pi]$ so that $\Delta_j \cap [-2^j\pi, 2^j\pi] = \emptyset$ and $\Delta_j \cup [-2^j\pi, 2^j\pi] = [-2^{j+1}\pi, 2^{j+1}\pi]$. One can think of Δ_j as representing the frequencies of sound in one musical octave. Then we can form a disjoint union of sets

$$\{0\} \cup \bigcup_{j=-\infty}^{\infty} \Delta_j = \mathbb{R}. \tag{11.30}$$

Remark 11.10 Proposition 11.9 shows that the subspaces V_j give a multiresolution of $L^2(\mathbb{R})$, known as an MRA. For some alternative choices of MRA, suitable for other applications, see [44] and [1].

Proposition 11.11 (Shannon's Wavelet) *Let the basic function be*

$$\psi(t) = 2sinc(2\pi t - \pi) - sinc(\pi t - \pi/2). \tag{11.31}$$

Then

$$\left(2^{j/2}\psi(2^j t - k)\right)_{j,k=-\infty}^{\infty} \tag{11.32}$$

gives a complete orthonormal basis for $L^2(\mathbb{R})$.

Proof We introduce the subspace W_j of $L^2(\mathbb{R})$ by

$$W_j = \left\{\phi(t) = \int_{\Delta_j} e^{i\omega t} f(\omega)\frac{d\omega}{2^{j+1}\pi} : f \in L^2(\Delta_j)\right\} \qquad (j \in \mathbb{Z}). \tag{11.33}$$

First we consider W_0. We can embed V_0 isometrically in V_1 by $\phi(t) \mapsto \sqrt{2}\phi(2t)$ with range $\tilde{V}_0$ and introduce the orthogonal complement W_0 of $\tilde{V}_0$ so that $V_1 = \tilde{V}_0 \oplus W_0$. Then

$$\psi_n(t) = 2\text{sinc}((2t - 2n - 1)\pi) - \text{sinc}((2t - 2n - 1)\pi)/2) \tag{11.34}$$

gives an orthogonal basis $(\psi_n)_{n=-\infty}^{\infty}$ for W_0.

We observe that $(2^{-1/2}\mathbb{I}_{[-\pi,\pi]}(\omega)e^{in\omega})_{n=-\infty}^{\infty}$ is orthonormal in $L^2(\mathbb{R})$, and

$$(2^{-1}\mathbb{I}_{[-2\pi,2\pi]}(\omega)e^{in\omega/2})_{n=-\infty}^{\infty}$$

is orthonormal in $L^2(\mathbb{R})$. Further, one checks by calculation that $(2^{-1/2}\mathbb{I}_{\Delta_0}(\omega)e^{i(2n+1)\omega/2})_{n=-\infty}^{\infty}$ is orthonormal in $L^2(\mathbb{R})$; note that the indices $2n+1$ here are all odd. This suggests that we can embed spaces of functions by using odd and even Fourier exponents. We observe that $L^2[-\pi, \pi]$ may be embedded isometrically as a subspace of $L^2[-2\pi, 2\pi]$ via by taking $g(\omega) = \sum_{k=-\infty}^{\infty} b_k e^{ik\omega} \in L^2[-\pi, \pi]$ and mapping this to $\sum_{k=-\infty}^{\infty} b_k e^{i2k\omega/2}$ with $\omega \in [-2\pi, 2\pi]$, where the index $2k$ is even, so that the orthogonal complement of the range in $L^2[-2\pi, 2\pi]$ is the space of $h(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{i(2k+1)\omega/2}$ with $\omega \in [-2\pi, 2\pi]$ and $\sum_k |c_k|^2$ convergent, giving the space with odd Fourier exponents.

Now we consider Δ_0 instead of $[-2\pi, 2\pi]$, and look at the odd indexed Fourier components. We observe for $f(\omega) = \sum_{k=-\infty}^{\infty} a_k e^{i(2k+1)\omega/2}$ in $L^2(\Delta_0)$ we have

$$a_k = \int_{\Delta_0} f(\omega)e^{-i(2k+1)\omega/2}\frac{d\omega}{2\pi} \tag{11.35}$$

hence the corresponding signal is

$$\begin{aligned}\phi(t) &= \int_{\Delta_0} e^{i\omega t} f(\omega)\frac{d\omega}{2\pi} \\ &= \sum_{k=-\infty}^{\infty} a_k \int_{-2\pi}^{-\pi} + \int_{\pi}^{2\pi} e^{i(2k+1)\omega/2 + i\omega t}\frac{d\omega}{2\pi} \\ &= \sum_{k=-\infty}^{\infty} a_k\Big(2\text{sinc}(\pi(2k+1+2t)) - \text{sinc}\frac{\pi(2k+1+2t)}{2}\Big).\end{aligned} \tag{11.36}$$

The subspaces $L^2(\Delta_j)$ are orthogonal subspaces of $L^2(\mathbb{R})$ since the Δ_j are pairwise disjoint, so by the Paley–Wiener theorem 10.36, the spaces W_j are orthogonal in $L^2(\mathbb{R})$. We have

$$2^{j/2}\psi(2^j t-k) = 2^{j/2}\int_{\Delta_0} e^{i(2^j t-k)\omega - i\omega/2}\frac{d\omega}{2\pi} = 2^{j/2}\int_{\Delta_j} e^{it\nu - i(2k+1)2^{-j}\nu}\frac{d\nu}{2^{j+1}\pi} \tag{11.37}$$

which shows that $2^{j/2}\psi(2^j t-k)$ are orthonormal for $j,k\in\mathbb{Z}$. One can prove that this system is complete in $L^2(\mathbb{R})$. □

Remark 11.12 The function ψ is known as Shannon's mother wavelet, such that the dyadic scalings and integer translations of ψ give an orthonormal basis of $L^2(\mathbb{R})$. This choice of wavelet is particularly well suited to digitization of sound, since

(i) the function ψ of (11.31) has a relatively simple formula;
(ii) the derivatives of ψ exist and are continuous;
(iii) the Fourier transform of ψ is of compact support.

The functions ψ belongs to $L^2(\mathbb{R})$; unfortunately, ψ is not integrable, since sinc is not integrable, as we noted in Lemma 11.2.

11.3 Telatar's Model of Wireless Communication

Consider a main line railway station, filled with travelers equipped with mobile telephones. There are many transmitters on the various platforms, and numerous receivers, and a background of unwanted radio signals from neighbouring buildings. When a traveler seeks to call home to report on the forthcoming journey, the call may be picked up by several receivers and degraded by the noise. We seek to model this complicated situation to understand what is transmitted and received.

Suppose that there are t transmitting antennas and r receiving antennas, and let

- $Y\in\mathbb{C}^{r\times 1}$ be the received signal;
- $X\in\mathbb{C}^{t\times 1}$ be the transmitted signal;
- $N\in\mathbb{C}^{r\times 1}$ be the noise
- $H\in M_{r\times t}(\mathbb{C})$ be the transmission matrix,

and suppose

$$Y = HX + N. \tag{11.38}$$

Let $H=[h_{jk}]$ and observe that the component h_{jk} measures how much the kth transmitter sends to the jth receiver. This transmission is degraded by jth component of the noise. More specifically, we assume that N is a Gaussian vector of the form $N=((\gamma_j+i\gamma_{-j})/\sqrt{2})_{j=1}^r$, where $(\gamma_j)_{j=-r}^r$ are mutually independent

$N(0, 1)$ Gaussian random variables. From the early days of probability theory, Gaussian random variables have been used as standard models for errors and noise. Note that UN has the same distribution as N for all unitary $U \in M_{r\times r}(\mathbb{C})$, so the noise has no preferred coordinate direction. In the simplest case $r = t$ and $H = I_{r\times r}$ so that the kth transmitter communicates only with the kth receiver; otherwise, there are nonzero entries of H for $j \neq k$ describing cross-talk between transmitters and receivers with different indices. The latter situation is what we describe in the following Lemma. Using Lemma 7.18, it is possible to replace all the complex matrices and vectors with larger real matrices and vectors, but we will persevere with the complex versions since the formulas are more compact.

Shannon observed that the logarithmic determinant in (11.39) is a crucial quantity in deciding how much information can be transmitted through the network and it is interpreted as a logarithmic capacity. The capacity of a communication link measures the mutual information between transmitters and receivers.

Lemma 11.13 *Suppose that H is constant and that the entries of X are random variables with finite second moments that are independent of the entries of N.*

(i) Then the matrices $Q = \mathbb{E}XX'$ and $R = \mathbb{E}YY'$ are positive semi definite and satisfy

$$R = I + HQH'.$$

(ii) The function $Q \mapsto \log\det(I + HQH')$ is increasing on the set of positive semi definite $t \times t$ matrices.

(iii) For $Q = \tau I$, we have

$$\log\det R = \sum_{j=1}^{\min\{r,t\}} \log(1 + \tau\sigma_j^2) \qquad (\tau \geq 0) \tag{11.39}$$

where σ_j are the singular numbers of H as in Definition 7.19.

Proof

(i) First, we find that the noise has mean and variance

$$\mathbb{E}N = 0, \quad \mathbb{E}NN' = I, \tag{11.40}$$

by the independence assumptions. Next we observe that

$$\langle Q\xi, \xi\rangle = \mathbb{E}\langle XX'\xi, \xi\rangle = \mathbb{E}\langle X'\xi, X'\xi\rangle = \mathbb{E}\|X'\xi\|^2 \geq 0$$

for all $\xi \in \mathbb{C}^{t\times 1}$, so Q is self-adjoint with all eigenvalues nonnegative.

Then we compute

$$\begin{aligned} R &= \mathbb{E}\big((HX+N)(X'H'+N')\big) \\ &= H\mathbb{E}XX'H' + \mathbb{E}(NX')H' + H\mathbb{E}(XN') + \mathbb{E}(NN') \\ &= HQH' + I \end{aligned}$$

where we have used the independence of X and N. One can check that R is positive semidefinite as with Q.

(ii) We have

$$\log\det R = \log\det(I + HQH') = \operatorname{trace}\log(I + HQH').$$

We can express this as an integral as in Exercise 3.19

$$\log\det R = \operatorname{trace}\int_0^\infty \Big((I+\tau I)^{-1} - (I + \tau I + HQH')^{-1}\Big)d\tau.$$

Now for $0 \le Q_1 \le Q_2$, we have $0 \le HQ_1H' \le HQ_2H'$, so

$$-(I+\tau I)^{-1} \le -(I+\tau I + HQ_1H')^{-1} \le -(I + \tau I + HQ_2H')^{-1},$$

and from the integral we deduce that

$$0 \le \log\det(I + HQ_1H') \le \log\det(I + HQ_2H').$$

This shows that $Q \mapsto \log\det(I + HQH')$ is increasing.

(iii) We have $\operatorname{rank}(HH') = \operatorname{rank}(H) \le \min\{r, t\}$ by the rank-nullity theorem 2.2, so there are at most $\min\{r, t\}$ nonzero singular numbers σ_j. With the specific choice of $Q = \tau I$, we have

$$\log\det R = \log\det(I + \tau HH') = \log\prod_{j=1}^{\min\{r,t\}}(1+\tau\sigma_j^2).$$

□

To make the model more realistic, it is necessary to widen the scope of assumptions about H. In practical situations, it is difficult to know in detail how much the kth transmitter can send to the jth receiver, so we assume this h_{jk} to be a random variable. In this way, H becomes a random matrix, which we suppose independent of the random entries of X and N. The computation that produced (11.39) remains valid, except that we now regard H as random, and seek the expected value. When working with $r \times r$ matrices, it is often helpful to rescale the trace by dividing by r, so that the scaled trace of the identity matrix gives $(1/r)\operatorname{trace} I_{r\times r} = 1$. Although the following result deals with an asymptotic distribution as $R \to \infty$, the conclusions can be used for matrices of size about 64×64, such as are used in applications. See page 93 of [43].

Proposition 11.14 *Suppose that $r = t$ and $H = W/\sqrt{r}$ where W is a $r \times r$ Gaussian Wigner matrix as in Definition 9.15. Then*

$$\mathbb{E}\frac{1}{r}\log\det(I + sW^2/r) \to \int_0^4 \log(1+sy)\sqrt{\frac{4-y}{y}}\frac{dy}{2\pi} \qquad (r \to \infty) \tag{11.41}$$

where the right-hand side is a holomorphic function of $s \in \mathbb{C} \setminus (-\infty, 0]$ such that

$$\frac{d}{ds}\int_0^4 \log(1+sy)\sqrt{\frac{4-y}{y}}\frac{dy}{2\pi} = \frac{2}{2s+1-\sqrt{4s+1}} \qquad (s > 0). \tag{11.42}$$

Proof For all continuous functions $f : [-2, 2] \to \mathbb{C}$, we have

$$\mathbb{E}\frac{1}{r}\mathrm{trace} f\Big(\frac{W}{\sqrt{r}}\Big) \to \int_{-2}^2 f(x)\sqrt{4-x^2}\frac{dx}{2\pi} \qquad (r \to \infty)$$

by Theorems 9.17 and 9.11. The right-hand side is the semicircle law, so we can consider σ to be a random eigenvalue in $[-2, 2]$ subject to the semicircle law, then we consider the law of σ^2, and the change of density that arises from $y = x^2$. We find that the limiting distribution of eigenvalues of W^2/r satisfies

$$\mathbb{P}[\sigma^2 \le y] = \frac{1}{2\pi}\int_{-\sqrt{y}}^{\sqrt{y}} \sqrt{4-\tau^2}d\tau$$

so the probability density function of σ^2 is

$$\frac{d}{dy}\mathbb{P}[\sigma^2 \le y] = \frac{\sqrt{4-y}}{2\pi\sqrt{y}} \qquad (0 < y < 4). \tag{11.43}$$

Hence for all continuous functions $g : [0, 4] \to \mathbb{C}$, we have

$$\mathbb{E}\frac{1}{r}\mathrm{trace} g\Big(\frac{W^2}{r}\Big) \to \int_0^4 g(y)\frac{\sqrt{4-y}}{2\pi\sqrt{y}}dy \qquad (r \to \infty).$$

In particular, we can take $g(y) = \log(1 + sy)$ for $y \in [0, 4]$ and $s \in \mathbb{C} \setminus (-\infty, 0]$ and obtain (11.41).

We can compute the derivative of this expression, as follows. Starting from the formula

$$\frac{2}{\pi}\int_{-1}^{1}\frac{\sqrt{1-x^2}}{\zeta-x}dx=\frac{2}{\zeta+\sqrt{\zeta^2-1}}\qquad (\zeta\in\mathbb{C}\setminus[-1,1]),$$

which we derived in Example 9.16, we substitute $x = y/2-1$ and $\zeta = -1-1/(2s)$ to obtain

$$\frac{1}{2\pi}\int_{0}^{4}\frac{\sqrt{4y-y^2}}{1+sy}dy=\frac{2}{2s+1-\sqrt{4s+1}}\qquad (s>0),$$

hence

$$\frac{d}{ds}\int_{0}^{4}\log(1+sy)\sqrt{\frac{4-y}{y}}\frac{dy}{2\pi}=\frac{2}{2s+1-\sqrt{4s+1}}\qquad (s>0). \qquad (11.44)$$

□

This result can be extended in several ways. We can replace the Wigner matrix W by a rectangular matrix H with Gaussian entries, and drop the assumption that H is symmetric. The Wishart matrix then arises from HH', and is one of the fundamental examples in random matrix theory. The probability density function in (11.43) is one of the family of Pastur-Marchenko distributions which arise in the context of the Wishart distribution. We refer the reader to [7], [43] or other books on random matrix theory for further discussion of this topic. Orthogonal polynomials are a useful tool for studying limit distributions of random matrices.

Example 11.15 One can easily rescale the probability density (11.43) so that it becomes

$$w(x)=(2\pi)^{-1}(x+1)^{-1/2}(1+x)^{1/2}\qquad (x\in(-1,1)).$$

The standard Jacobi polynomials $P_n^{(1/2,-1/2)}(x)$ are orthogonal with respect to this weight, and are normalized so that

$$P_n^{(1/2,-1/2)}(1)=\binom{1/2+n}{n}.$$

Then $y = P_n^{(1/2,-1/2)}(x)$ gives the only polynomial solution of the differential equation

$$(1-x^2)\frac{d^2y}{dx^2}-(1+2x)\frac{dy}{dx}+n(n+1)y=0.$$

The three-term recurrence relation is

$$4n^2(n-2)P_n^{(1/2,-1/2)}(x) = 2n(2n-1)(2n-2)xP_{n-1}^{(1/2,-1/2)}(x) - n(2n-1)(2n-3)P_{n-2}^{(1/2,-1/2)}(x)$$

as in page 71 of [54].

11.4 Exercises

Exercise 11.1 Show by substitution that

$$\int_0^4 \sqrt{\frac{4-y}{y}}\,\frac{dy}{2\pi} = \frac{(2n)!}{(n+1)(n!)^2} \qquad (n = 0, 1, \dots).$$

Exercise 11.2

(i) Let A and B be positive definite $n \times n$ matrices. Using Lemma 3.38 and Theorem 3.20, show that there exists a positive definite $\sqrt{A}$ such that $(\sqrt{A})^2 = A$ and

$$\det(I + sAB) = \det(I + s\sqrt{A}\,B\,\sqrt{A}).$$

(ii) Deduce that the eigenvalues of AB are positive. (It is not asserted that AB is positive definite).

Chapter 12
Solutions to Selected Exercises

Exercise 1.3 We write

$$\frac{dx}{dt} = v$$
$$\frac{dv}{dt} = -\frac{k}{m}x + \frac{u}{m}.$$

Then we introduce the state vector

$$X = \begin{bmatrix} x \\ v \end{bmatrix}$$

and introduce the coefficient matrices

$$A = \begin{bmatrix} 0 & 1 \\ -k/m & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1/m \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad D = 0.$$

Then

$$\frac{dX}{dt} = AX + Bu$$
$$x = CX + Du.$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6_12

Exercise 2.5

(i) As in (2.51), let

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -7 & -6 & -5 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 4 & -3 & 2 \end{bmatrix}, \quad D = 0.$$

(ii) The eigenvalues of A are

$$\text{eig}(A) = -3.9259, -0.5391 \pm \imath 1.2225;$$

all of these lie in the open left half plane, so the system is stable.

Exercise 2.12 We consider the augmented matrix for $[sI - A \quad | \quad I]$

$$\begin{bmatrix} s-1 & -4 & -10 & | & 1 & 0 & 0 \\ 0 & s-2 & 0 & | & 0 & 1 & 0 \\ 0 & 0 & s-3 & | & 0 & 0 & 1 \end{bmatrix}$$

so the row operations $r_1 \mapsto r_1/(s-1)$, $r_2 \mapsto r_2/(s-2)$ and $r_3 \mapsto r_3/(s-3)$ give

$$\begin{bmatrix} 1 & -4/(s-1) & -10/(s-1) & | & 1/(s-1) & 0 & 0 \\ 0 & 1 & 0 & | & 0 & 1/(s-2) & 0 \\ 0 & 0 & 1 & | & 0 & 0 & 1/(s-3) \end{bmatrix}$$

then $r_1 \mapsto r_1 + 4r_2/(s-1) + 10r_3/(s-1)$, gives

$$\begin{bmatrix} 1 & 0 & 0 & | & 1/(s-1) & 4/(s-1)(s-2) & 10/(s-1)(s-3) \\ 0 & 1 & 0 & | & 0 & 1/(s-2) & 0 \\ 0 & 0 & 1 & | & 0 & 0 & 1/(s-3) \end{bmatrix}$$

hence

$$(sI - A)^{-1} = \begin{bmatrix} 1/(s-1) & 4/(s-1)(s-2) & 10/(s-1)(s-3) \\ 0 & 1/(s-2) & 0 \\ 0 & 0 & 1/(s-3) \end{bmatrix}$$

Exercise 2.10 By polynomial long division we obtain

$$T(s) = 5 + \frac{22s^3 - 26s^2 - 34s - 28}{s^4 - 3s^3 + 4s^2 + 7s + 6}.$$

Hence we choose

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -6 & -7 & -4 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} -28 & -34 & -26 & 22 \end{bmatrix}, \quad D = 5.$$

The numerical values for the eigenvalues of A are found to be

$$\text{eig}(A) = 2.1014 \pm \imath 1.9797, \; -0.6014 \pm \imath 0.5985.$$

Exercise 2.13 The solution follows the method of proof of Lemma 2.32. Consider

$$\begin{aligned} \det(\lambda I - A) &= \det \begin{bmatrix} \lambda - 1 & 1 \\ -3 & \lambda - 5 \end{bmatrix} \\ &= (\lambda - 1)(\lambda - 5) + 3 \\ &= \lambda^2 - 6\lambda + 8 \\ &= (\lambda - 4)(\lambda - 2). \end{aligned}$$

For $\lambda = 2$,

$$2I - A = \begin{bmatrix} 1 & 1 \\ -3 & -3 \end{bmatrix}$$

so we choose eigenvector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$;

for $\lambda = 4$

$$4I - A = \begin{bmatrix} 3 & 1 \\ -3 & -1 \end{bmatrix}$$

so we choose eigenvector $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

Let

$$S = \begin{bmatrix} 1 & 1 \\ -1 & -3 \end{bmatrix}, D = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix};$$

then $AS = SD$, so $A = SDS^{-1}$, where

$$S^{-1} = \frac{-1}{2}\begin{bmatrix} -3 & -1 \\ 1 & 1 \end{bmatrix};$$

hence

$$\begin{aligned} \exp(tA) &= S\exp(tD)S^{-1} \\ &= \frac{-1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -3 \end{bmatrix}\begin{bmatrix} e^{2t} & 0 \\ 0 & e^{4t} \end{bmatrix}\begin{bmatrix} -3 & -1 \\ 1 & 1 \end{bmatrix} = \frac{-1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -3 \end{bmatrix}\begin{bmatrix} -3e^{2t} & -e^{2t} \\ e^{4t} & e^{4t} \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 3e^{2t} - e^{4t} & e^{2t} - e^{4t} \\ -3e^{2t} + 3e^{4t} & -e^{2t} + 3e^{4t} \end{bmatrix} \end{aligned}$$

The eigenvectors of A are only unique up to non zero constant multiples, so there are other valid choices available for S. Of course, the final answer $\exp(tA) = S\exp(tD)S^{-1}$ is unique.

Exercise 2.14 As in Proposition 2.12, we introduce the companion matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -2 & -4 & -1 & -2 \end{bmatrix}$$

with numerical eigenvalues

$$\mathrm{eig}(A) = -2.1877, 0.3516 \pm \imath 1.2843, -0.5156.$$

Given the matrix A, the final step can be carried out in MATLAB using $eig(A)$.

Exercise 2.16 The solution follows the method of Proposition 2.33. The idea is to build solutions out of each eigenvector of A.

(i) By applying column and row operations, we have

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 2 & 1 & 1 \\ 1 & \lambda - 2 & 1 \\ 1 & 1 & \lambda - 2 \end{vmatrix} \qquad (c_2 \mapsto c_2 - c_3)$$

$$= \begin{vmatrix} \lambda - 2 & 0 & 1 \\ 1 & \lambda - 3 & 1 \\ 1 & -\lambda + 3 & \lambda - 2 \end{vmatrix} \qquad (r_2 \mapsto r_2 - r_1)$$

$$= \begin{vmatrix} \lambda - 2 & 0 & 1 \\ 3 - \lambda & \lambda - 3 & 0 \\ 1 & -\lambda + 3 & \lambda - 2 \end{vmatrix} \qquad (r_3 \mapsto r_3 + r_2)$$

$$= \begin{vmatrix} \lambda - 2 & 0 & 1 \\ 3 - \lambda & \lambda - 3 & 0 \\ 4 - \lambda & 0 & \lambda - 2 \end{vmatrix}$$

$$= (\lambda - 3) \begin{vmatrix} \lambda - 2 & 1 \\ 4 - \lambda & \lambda - 2 \end{vmatrix}$$

$$= (\lambda - 3)(\lambda^2 - 3\lambda)$$

$$= \lambda(\lambda - 3)^2.$$

so the eigenvalues are $\lambda = 0, 3, 3$, listed according to algebraic multiplicity. We choose eigenvectors for the corresponding eigenvalues

$$\lambda = 0 : V_0 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \lambda = 3 : V_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}; \lambda = 3 : V_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

The choice of V_0 is unique up to constant multiples; whereas we can choose $\{V_1, V_2\}$ to be any convenient basis for the eigenspace $\{V : AV = 3V\}$. The above choice uses orthogonal vectors, but this is not an essential aspect of the solution.

(ii) Note that when V is an eigenvector corresponding to eigenvalue λ of A, the function $Z(t) = e^{zt}V$ satisfies $AZ = \lambda Z$ and $dZ/dt = zZ$; so we choose $z = \lambda$ to get $dZ/dt = AZ$. The differential equation $\frac{dx}{dt} = 3x$ has general solution $x = ae^{3t}$ while the differential equation $\frac{dx}{dt} = 0$ has general solution $x = b$. So we have

$$Z = aV_0 + b_1 e^{3t} V_1 + b_2 e^{3t} V_2.$$

(iii) Note that when V is an eigenvector corresponding to eigenvalue λ of A, the function $Y(t) = e^{wt}V$ satisfies $AY = \lambda Y$ and $d^2Y/dt^2 = w^2 Y$; so we choose $w^2 = \lambda$ to get $d^2Y/dt^2 = AY$, so $w = \pm\sqrt{\lambda}$. The differential equation $\frac{d^2x}{dt^2} = 3x$ has general solution

$$x = c_1 e^{t\sqrt{3}} + c_2 e^{-t\sqrt{3}}$$

while the differential equation $\frac{d^2x}{dt^2} = 0$ has general solution $x = at + b$. Hence the required general solution is

$$W = \frac{at+b}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{c_1 e^{t\sqrt{3}} + c_2 e^{-t\sqrt{3}}}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{d_1 e^{t\sqrt{3}} + d_2 e^{-t\sqrt{3}}}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

for real constants a, b, c_1, c_2, d_1, d_2.

(iv) In (ii) there are three constants, required to specify $Z(0)$.

In (iii) There are six constants; three specify $W(0)$ and three specify $(dW/dt)(0)$. Equivalently, we can write the system as

$$\frac{d}{dt} \begin{bmatrix} W \\ U \end{bmatrix} = \begin{bmatrix} 0 & I \\ A & 0 \end{bmatrix} \begin{bmatrix} W \\ U \end{bmatrix}$$

where the column vector is 6×1, so we need 6 independent constants.

Exercise 3.2 We compute the eigenvalues by MATLAB; specify the matrices by

$$>> A = [1, 1, 3; 2, 7, 5; 1, 8, 2]$$

$$>> B = [1, 1, 7; 9, 8, 4; 2, 2, 9]$$

$$>> C = [1, 1, 1, 1; 2, 7, 9, 4; 8, 1, 7, i; 2, 2i, 2, 4]$$

eig$(A) = 11.8679, -0.0951, -1.7729$;

eig$(B) = 13.8334, 0.0886, 4.0780$;

eig$(C) = 11.6936 + 0.8093\imath, 0.8875 - 1.7730\imath, 1.9024 + 0.8578\imath, 4.5164 + 0.1059\imath$

and the eigenvalues of $-A$ are the negatives of the eigenvalues of A, and so on. Hence $\pm A$, B and C have eigenvalues λ in the right half plane with $\Re\lambda > 0$, so are unstable. However, $-B$ and $-C$ have all their eigenvalues in the open left half plane with $\lambda < 0$, so are stable.

Exercise 3.5

(i) For $X = \text{column}(x_j)_{j=1}^n$, We have

$$\langle DX, X \rangle = \sum_{j=1}^{n} \kappa_j x_j^2,$$

and since $\kappa_1 \geq \kappa_j \geq \kappa_n$, we deduce that

$$\kappa_1 \sum_{j=1}^{n} x_j^2 \geq \sum_{j=1}^{n} \kappa_j x_j^2 \geq \kappa_n \sum_{j=1}^{n} x_j^2,$$

so

$$\kappa_1 \|X\|^2 \geq \langle DX, X\rangle \geq \kappa_n \|X\|^2.$$

(ii) By the spectral theorem 3.20, we can introduce a real orthogonal matrix S, and a diagonal matrix D as in (i) such that $K = SDS^{-1}$, where $S^{-1} = S^\top$. Then

$$\kappa_1\langle X, X\rangle \geq \langle DX, X\rangle \geq \kappa_n\langle X, X\rangle \qquad (X \in \mathbb{R}^{n\times 1}),$$

and, with $X = S^\top Y$ for $Y = SX$, we have

$$\kappa_1\langle S^\top Y, S^\top Y\rangle \geq \langle SDS^\top Y, Y\rangle \geq \kappa_n\langle S^\top Y, S^\top Y\rangle \qquad (Y \in \mathbb{R}^{n\times 1}),$$

so

$$\kappa_1\langle Y, Y\rangle \geq \langle KY, Y\rangle \geq \kappa_n\langle Y, Y\rangle \qquad (Y \in \mathbb{R}^{n\times 1}),$$

hence the result.

Exercise 3.7

(i) Let X be an eigenvector corresponding to eigenvalue κ. Then the eigenvalue equation $\kappa X = KX$ gives $\kappa\langle X, X\rangle = \langle KX, X\rangle > 0$ since K is positive definite. Also $\|X\|^2 = \langle X, X\rangle > 0$, so $\kappa > 0$.

(ii) Let the eigenvalues be $\kappa_1, \dots, \kappa_n$. By the spectral theorem 3.20 for real symmetric matrices, there exists an orthogonal matrix S such that $K = SDS'$, where D is the diagonal matrix

$$D = \begin{bmatrix} \kappa_1 & 0 & \dots & 0 \\ 0 & \kappa_2 & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \kappa_n \end{bmatrix}.$$

Then

$$\det K = \det D = \kappa_1\kappa_2\dots\kappa_n > 0,$$

$$\mathrm{trace}(K) = \mathrm{trace} D = \kappa_1 + \kappa_2 + \dots + \kappa_n > 0.$$

Alternatively, note that the usual basis vectors e_j $(j = 1, \dots, n)$ for $\mathbb{C}^n$ give

$$\operatorname{trace}(K) = \sum_{j=1}^{n} \langle Ke_j, e_j \rangle > 0$$

by the definition of trace and the assumption that K is positive definite.

(iii) For all $X \neq 0$, we have $Y = SX \neq 0$ since S is invertible, so

$$\langle S'KSX, \rangle = \langle KSX, SX \rangle = \langle KY, Y \rangle > 0,$$

hence $S'KS$ is positive definite.

(iv) Let $S = \exp(A)$. Then $\exp(-A)\exp(A) = I$, so S is invertible, and

$$\begin{aligned} S' &= \Big(I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots\Big)' \\ &= \Big(I + A' + \frac{(A')^2}{2!} + \frac{(A')^3}{3!} + \dots\Big) \\ &= \exp(A'), \end{aligned}$$

hence $\exp(A')K\exp(A) = S'KS$ is positive definite by (i).

(v) Let $Y \neq 0$ be a vector. Then

$$\langle (K + L)Y, Y \rangle = \langle KY, Y \rangle + \langle LY, Y \rangle > 0$$

so $K + L$ is positive definite.

[Beware that the eigenvalues of $K + L$ are related to the eigenvalues of K and L in a complicated way; we cannot just add eigenvalues of K to eigenvalues of L and get eigenvalues of $K + L$. Also the minors of $K + L$ are related to those of K and L is a complicated way.]

Exercise 3.21

(i) Note that V is a subset of $\mathbb{C}^{n\times 1}$ and for $\lambda, a_j, b_j \in \mathbb{C}$ we have

$$\lambda \sum_{j=0}^{n-1} a_j A^j B = \sum_{j=0}^{n-1} \lambda a_j A^j B;$$

and

$$\sum_{j=0}^{n-1} a_j A^j B + \sum_{j=0}^{n-1} b_j A^j B = \sum_{j=0}^{n-1} (a_j + b_j) A^j B.$$

Hence V is a linear subspace of $\mathbb{C}^{n\times 1}$.

(ii) First note that $V = \{0\}$, if and only if $B = 0$. If V has dimension one, then $V = \{a_0 B : a_0 \in \mathbb{C}\}$ and $AB \in V$ so $AB = a_0 B$ for some $a_0 \in \mathbb{C}$; hence B is an eigenvector of A. Conversely, if B is an eigenvector, so that $AB = \lambda B$, then $A^j B = \lambda^j B$, so

$$\sum_{j=0}^{n-1} a_j A^j B = \sum_{j=0}^{n-1} a_j \lambda^j B$$

and V evidently has dimension one.

(iii) We need to check that L_A maps V to itself, and the main problem is with $L_A A^n$. By the Cayley–Hamilton theorem, $\chi_A(A) = 0$, so

$$A^n = \mathrm{trace}(A)A^{n-1} - \cdots + (-1)^{n+1}(\det A)I;$$

so a typical $X \in V$ has the form

$$X = \sum_{j=0}^{n-1} a_j A^j B$$

has

$$\begin{aligned} L_A X &= AX \\ &= A\sum_{j=0}^{n-1} a_j A^j B \\ &= \sum_{j=0}^{n-2} a_j A^{j+1} B + a_{n-1} A^n B \\ &= \sum_{j=0}^{n-2} a_j A^{j+1} B + a_{n-1}\big(\mathrm{trace}(A)A^{n-1}B - \cdots + (-1)^{n+1}(\det A)B\big), \end{aligned}$$

so the first sum involves A^{j+1} with $j + 1 \leq n - 1$ and the other powers of A are A^k with $k \leq n - 1$; hence $AX \in V$ for all $X \in V$, so L_A maps V to itself.

(iv) MATLAB gives

$$Q = \begin{bmatrix} 1 & -16.5 & -126 & -1673 \\ -5 & 8.5 & -141.5 & -1676 \\ -0.5 & -27 & -224.5 & -1732.5 \\ 3 & 15.5 & 81 & -111.5 \end{bmatrix}$$

which has rank 4.

$$>> A = [1, 3, 5, 0; 0, 1, 9, 6; 1, 1, 4, -7; 2, 2, 1, 8]$$

$$>> B = [1; -5; -0.5; 3]$$

$$>> C = A * B$$

$$>> D = A * C$$

$$>> E = A * D$$

$$>> Q = [B, C, D, E]$$

$$>> rank(Q)$$

Alternatively, one can find

$$>> det(Q) = -1.2664e + 07$$

so $\det Q \neq 0$, hence Q is invertible and has full rank 4.

Exercise 4.1

(i) The Laplace transform can be found by integration by parts, as in

$$\begin{aligned}\int_0^R e^{-st}\cos 2\omega t\,dt &= \left[\frac{e^{-st}\cos 2\omega t}{-s}\right]_0^R - \frac{2\omega}{s}\int_0^R e^{-st}\sin 2\omega t\,dt\\ &= \frac{1}{s} - \frac{e^{-sR}\cos 2\omega R}{s} + \left[\frac{2\omega e^{-st}\sin 2\omega t}{s^2}\right]_0^R\\ &\quad - \frac{4\omega^2}{s^2}\int_0^R e^{-st}\cos 2\omega t\,dt,\end{aligned}$$

so that

$$\left(1 + \frac{4\omega^2}{s^2}\right)\int_0^R e^{-st}\cos 2\omega t\,dt = \frac{1}{s} - \frac{e^{-sR}\cos 2\omega R}{s} + \frac{2\omega e^{-sR}\sin 2\omega R}{s^2},$$

so letting $R \to \infty$, we have $e^{-sR}\cos 2\omega R \to 0$ and $e^{-sR}\sin 2\omega R \to 0$ for all $s > 0$, so

$$\int_0^\infty e^{-st}\cos 2\omega t\,dt = \frac{s}{s^2 + 4\omega^2}.$$

(ii) We have $\cos 2\omega t = 1 - 2\sin^2 \omega t$, so

$$
\begin{aligned}
Y(s) &= \int_0^\infty s^{-st} \sin^2 \omega t \, dt \\
&= \frac{1}{2} \int_0^\infty e^{-st} (1 - \cos 2\omega t) \, dt \\
&= \frac{1}{2} \int_0^\infty e^{-st} \, dt - \frac{1}{2} \int_0^\infty e^{-st} \cos 2\omega t \, dt \\
&= \frac{1}{2s} - \frac{s}{2(s^2 + 4\omega^2)} \\
&= \frac{2\omega^2}{s(s^2 + 4\omega^2)} \qquad (s > 0).
\end{aligned}
$$

Exercise 4.2 The Laplace transform is

$$sY - 7Y = \frac{2}{4 + s^2},$$

so that

$$Y(s) = \frac{2}{(s-7)(s^2+4)}.$$

The partial fractions have the form

$$Y(s) = \frac{As + B}{s^2 + 4} + \frac{C}{s - 7},$$

and we compute the undetermined coefficients by using

$$2 = (As + B)(s - 6) + C(s^2 + 4),$$

so

$$
\begin{aligned}
s^2 :& \quad 0 = A + C \\
s :& \quad 0 = -7A + B \\
1 :& \quad 2 = -7B + 4C
\end{aligned}
$$

so that $A = -2/53$, $C = 2/53$ and $B = -14/53$. Hence

$$Y(s) = \frac{-2}{53}\frac{s}{(s^2+4)} + \frac{-7}{53}\frac{2}{(s^2+4)} + \frac{2}{53}\frac{1}{(s-7)},$$

so by uniqueness of Laplace transforms

$$y(t) = \frac{-2}{53}\cos 2t - \frac{7}{53}\sin 2t + \frac{2}{53}e^{7t}.$$

Exercise 4.19 By the triangle inequality

$$\begin{aligned}|f * h(t)| &= \Big|\int_0^t h(t-s)f(s)\,ds\Big| \\ &\le \int_0^t |h(t-s)||f(s)|ds \\ &\le M\int_0^t |f(s)|\,ds \\ &\le M\int_0^\infty |f(s)|\,ds \qquad (t>0);\end{aligned}$$

hence $f * h$ is bounded.

Exercise 4.21 In $L^2[0,2]$ we have an orthonormal basis $(e^{\pi int}/\sqrt{2})_{n=-\infty}^{\infty}$, and the Fourier coefficients of $t-1$ are

$$a_0 = \int_0^2 (t-1)\frac{dt}{\sqrt{2}} = 0;$$

$$\begin{aligned}a_n &= \int_0^2 (t-1)\frac{e^{-\pi int}dt}{\sqrt{2}} \\ &= \Big[\frac{(t-1)e^{-\pi int}}{-\pi in\sqrt{2}}\Big]_0^2 + \int_0^2 \frac{e^{-\pi int}}{-\pi in\sqrt{2}}dt \\ &= \frac{e^{-2\pi in}}{-\pi in\sqrt{2}} - \frac{1}{\pi in\sqrt{2}} \\ &= \frac{-\sqrt{2}}{\pi in} \qquad (n\in\mathbb{Z}\setminus\{0\};\end{aligned}$$

while the log series gives

$$\sum_{n=-\infty}^{0} \frac{-e^{\pi int}}{\pi in} + \sum_{n=0}^{\infty} \frac{-e^{\pi int}}{\pi in} = \frac{1}{\pi i} \log \frac{1}{1-e^{-\pi it}} - \frac{1}{\pi i} \log \frac{1}{1-e^{\pi it}}$$
$$= \frac{1}{\pi i} \log \frac{1-e^{\pi it}}{1-e^{-\pi it}}$$
$$= \frac{1}{\pi i} \log e^{\pi i(t-1)}$$
$$= t-1$$

hence we have an orthogonal series

$$t-1 = \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{i}{\pi n} e^{\pi int} \qquad (t \in (0,2)).$$

Exercise 4.22 The error function is

$$\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-x^2} dx$$
$$= \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n+1}}{(2n+1)n!}.$$

The inverse Laplace transform of $\operatorname{erf}(1/s)$ is

$$g(t) = \mathcal{L}^{-1}\big(\operatorname{erf}(1/s); t\big)$$
$$= \frac{2}{\sqrt{\pi}} \int_C \sum_{n=0}^{\infty} \frac{(-1)^n s^{-(2n+1)} e^{st}}{(2n+1)n!} \frac{ds}{2\pi i}$$
$$= \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \int_C \frac{(-1)^n s^{-(2n+1)} e^{st}}{(2n+1)n!} \frac{ds}{2\pi i}$$
$$= \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{n!(2n+1)!}.$$

Now we take the Laplace transform of $g(\sqrt{t})$, obtaining

$$\begin{aligned}\int_0^\infty g(\sqrt{t})e^{-st}dt &= \frac{2}{\sqrt{\pi}}\int_0^\infty \sum_{n=0}^\infty \frac{(-1)^n t^n e^{-st}}{n!(2n+1)!}dt\\ &= \frac{2}{\sqrt{\pi}}\sum_{n=0}^\infty \int_0^\infty \frac{(-1)^n t^n e^{-st}}{n!(2n+1)!}dt\\ &= \frac{2}{\sqrt{\pi}}\sum_{n=0}^\infty \frac{(-1)^n}{(2n+1)!s^{n+1}}\\ &= \frac{2}{\sqrt{\pi}}\frac{1}{\sqrt{s}}\sin\frac{1}{\sqrt{s}}.\end{aligned}$$

Exercise 5.1

(i) The Laplace transform of the differential equation is

$$s^2Y(s) + 6sY(s) + Y(s) = -3sU(s) + U(s),$$

so

$$Y(s) = \frac{-3s+1}{s^2+6s+1}U(s),$$

so the transfer function is

$$T(s) = \frac{-3s+1}{s^2+6s+1}.$$

Hence the frequency response function is

$$\begin{aligned}T(i\omega) &= \frac{-3i\omega+1}{1-\omega^2+6i\omega}\frac{1-\omega^2-6i\omega}{1-\omega^2-6i\omega}\\ &= \frac{1-19\omega^2-9i\omega+3i\omega^3}{(1-\omega^2)^2+36\omega^2}.\end{aligned}$$

(ii) Hence the gain is

$$\Gamma(\omega) = |T(\imath\omega)| = \frac{\sqrt{1+9\omega^2}}{\sqrt{(1-\omega^2)^2+36\omega^2}},$$

while the phase shift ϕ satisfies

$$\tan\phi = \frac{3\omega^3 - 9\omega}{1 - 19\omega^2}.$$

so

$$\phi = \tan^{-1}\frac{3\omega^3 - 9\omega}{1 - 19\omega^2}.$$

Exercise 6.2

(i) Note that K is invertible if and only if $\det K = PX + YQ \neq 0$, and then

$$\begin{bmatrix} P & Q \\ -Y & X \end{bmatrix}^{-1} = \frac{1}{PY + QX}\begin{bmatrix} X & -Q \\ Y & P \end{bmatrix};$$

but we still need to consider whether the entries of the right-hand side are actually polynomials. We have $I = KK^{-1}$ so $1 = \det K \det K^{-1}$.

If K^{-1} has polynomial entries, then $\det K$ and $\det K^{-1}$ are both polynomials, so

$$\text{degree}(\det K) + \text{degree}(\det K^{-1}) = 0$$

so $\text{degree}(\det K) = 0$ and $\det K = \kappa$, for some $\kappa \neq 0$, that is $PY + XQ = \kappa$.

Conversely, if $PX + QY = \kappa$, then K is invertible, and the entries are polynomials.

(ii) Recall from the Euclidean algorithm that P and Q have highest common factor 1 if and only if $PX + QY = 1$ for some polynomials X, Y, or equivalently $PX + QY = \kappa$ for some $\kappa \neq 0$ with $\kappa \in \mathbb{C}$.

(iii) When $P(s)$ and $Q(s)$ have no common complex zero, then $P(s)$ and $Q(s)$ have highest common factor 1, so there exist polynomials X, Y such that $PX + QY = 1$, and this gives the required K.

(iv) By the Euclidean algorithm, we have

$$1 = \frac{s-2}{12}\big(s^2 + 2s - 3\big) + \frac{-s+3}{12}\big(s^2 + 3s + 2\big),$$

so we have

$$K = \begin{bmatrix} s^2 + 2s - 3 & s^2 + 3s + 2 \\ s - 3 & s - 2 \end{bmatrix}$$

with $\det K = 12$, so K^{-1} has polynomial entries.

The choice of K is not unique; indeed, one can add polynomial multiples of the first row to the second without changing the determinant. One can do the Euclidean algorithm by hand. Alternatively, use the MATLAB instructions:

$$>> syms \quad s$$

$$>> P = s^2 + 2 * s - 3$$

$$>> Q = s^2 + 3 * s + 2$$

$$>> [g, X, Y] = gcd(P, Q)$$

Exercise 6.5 Descartes's Rule of Signs

(i) Here $\sigma = 1$, so $r = 1$.
(ii) Here $\sigma = 5$, so $r = 1, 3$ or 5.
(iii) The roots are $-4.1642, 0.3914, -1.2271$; which confirms that $r = 1$.
A ponderous solution is to introduce

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -3 & -5 \end{bmatrix}$$

and then compute $eig(C)$.
(iv) The roots are $5.8580, 0.8029 \pm \imath 0.4265, -0.0829 \pm \imath 1.0466, -0.8644, -0.4336$; so $r = 1$.

Exercise 6.8 We introduce $s = (1 - \lambda)/\lambda$ and write

$$\begin{aligned} G(s) &= \frac{s^2 + s + 1}{s^2 - 2} \\ &= \frac{(1-\lambda)^2/\lambda^2 + (1-\lambda)/\lambda + 1}{(1-\lambda)^2/\lambda^2 - 2} \\ &= \frac{(1-\lambda)^2 + \lambda(1-\lambda) + \lambda^2}{(1-\lambda)^2 - 2\lambda^2} \\ &= \frac{\lambda^2 - \lambda + 1}{-\lambda^2 - 2\lambda + 1}, \end{aligned}$$

so we write

$$-\lambda^2 - 2\lambda + 1 = -(\lambda^2 - \lambda + 1) - 3\lambda + 2$$

where

$$\lambda^2 - \lambda + 1 = (-\lambda/3 + 1/9)(-3\lambda + 2) + 7/9$$

so that

$$\begin{aligned} 7/9 &= \lambda^2 - \lambda + 1 - (-\lambda/3 + 1/9)(-3\lambda + 2) \\ &= \lambda^2 - \lambda + 1 - (-\lambda/3 + 1/9)((-\lambda^2 - 2\lambda + 1) + (\lambda^2 - \lambda + 1)) \\ &= (8/9 + \lambda/3)(\lambda^2 - \lambda + 1) + (\lambda/3 - \lambda/9)(-\lambda^2 - 2\lambda + 1) \end{aligned}$$

so that

$$1 = \Big(\frac{3\lambda + 8}{7}\Big)\Big(\lambda^2 - \lambda + 1\Big) + \Big(\frac{3\lambda - 1}{7}\Big)\Big(- \lambda^2 - 2\lambda + 1\Big)$$

and substituting $\lambda = 1/(s + 1)$, we obtain

$$1 = \Big(\frac{3 + 8(1 + s)}{7(1 + s)}\Big)\Big(\frac{1}{(1 + s)^2} - \frac{1}{1 + s} + 1\Big) + \Big(\frac{3 - (1 + s)}{7(1 + s)}\Big)\Big(\frac{-1}{(1 + s)^2} - \frac{2}{1 + s} + 1\Big)$$

which shows that G is the quotient of coprime functions in $\mathcal{S}$, as in

$$G(s) = \frac{\frac{1}{(1+s)^2} - \frac{1}{1+s} + 1}{\frac{-1}{(1+s)^2} - \frac{2}{1+s} + 1}.$$

One can do this by hand. Alternatively, use MATLAB.

$$>> syms \quad x$$

$$>> P = x^2 - x + 1$$

$$>> Q = x^2 + 2 * x - 1$$

$$>> [g, M, N] = gcd(P, Q)$$

This gives polynomials M and N such that $1 = PM + QN$.

Exercise 7.2

(i) We have

$$-A - A' = \begin{bmatrix} 2 & 4 & 4 \\ 4 & 10 & 3 \\ 4 & 3 & 14 \end{bmatrix}$$

which is real symmetric, but $\det(-A - A') = -26$, hence $-A - A'$ is not positive definite.

Alternatively, one can find

$$\mathrm{eig}(-A - A') = 17.5295, 8.6421, -0.1716.$$

(ii) We have

$$\mathrm{eig}(A) = -0.1093, -8.6706, -4.2201$$

$$\mathrm{eig}(A') = -0.1093, -8.6706, -4.2201$$

which are all in the open left half plane, so there exists a solution to $AK + KA' = -I$, with K positive definite, by Corollary 7.5. MATLAB gives

$$K = \begin{bmatrix} 4.5158 & -1.7607 & -0.1648 \\ -1.7607 & 0.7974 & 0.0342 \\ -0.1648 & 0.0342 & 0.0852 \end{bmatrix}.$$

The required MATLAB command is

$$>> K = lyap(A, I)$$

or equivalently

$$>> K = lyap(A, A', I)$$

The solution in terms of factions is given by inputting

$>>$ format rational

To solve this by the Sylvester's equation, we observe that $K = K'$, so we can write

$$-AK - KA' = I$$

as

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 1 & 2 & 7 \end{bmatrix} \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} + \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 3 & 1 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so we have six linear equations for the six unknowns a, b, c, d, e, f; considering the terms on or above the leading diagonal we write these equations in the matrix format

$$\begin{bmatrix} 2 & 4 & 6 & 0 & 0 & 0 \\ 2 & 6 & 1 & 2 & 3 & 0 \\ 1 & 2 & 8 & 0 & 2 & 3 \\ 0 & 4 & 0 & 10 & 2 & 0 \\ 0 & 1 & 2 & 2 & 12 & 1 \\ 0 & 0 & 2 & 0 & 4 & 14 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

which one can solve by linear algebra to obtain a unique solution for the six unknowns a, b, c, d, e, f; this gives

$$K = \frac{1}{15680} \begin{bmatrix} 70808 & -27608 & -2584 \\ -27608 & 12504 & 536 \\ -2584 & 536 & 1336 \end{bmatrix};$$

finally, one checks that K is positive definite. Either one can invoke Corollary 7.5, or note that K has eigenvalues 5.222, 0.1183 and 00576, all positive; or one can compute the principal minors of K as

$$70808/15680,\ 123181568/15680^2,\ 240/6743;$$

so K is positive definite. All this can be carried out in exact arithmetic by hand; however, the calculation is tedious.

Exercise 8.12 Prolate Spheroidal Wave Functions Differentiating through the integral sign, we have

$$\Big((1-x^2)\frac{d^2}{dx^2} - 2x\frac{d}{dx} - \lambda^2 x^2\Big) \int_{-1}^{1} e^{i\lambda xy} f(y)dy$$
$$= \int_{-1}^{1} \Big(-(1-x^2)\lambda^2 y^2 - 2i\lambda xy - \lambda^2 x^2) e^{i\lambda x} f(y)dy$$
$$= \int_{-1}^{1} (\lambda^2 x^2 y^2 - \lambda^2 y^2 - \lambda^2 x^2 - 2i\lambda xy) e^{i\lambda xy} f(y)dy$$

which we can compare with the following identities, which occur by integration by parts

$$\int_{-1}^{1} e^{i\lambda x}\Big((1-y^2)\frac{d^2}{dy^2} - 2y\frac{d}{dy} - \lambda^2 y^2\Big) f(y)dy$$
$$= \int_{-1}^{1} e^{i\lambda xy}\Big(\frac{d}{dy}\big((1-y^2)\frac{df}{dy}\big) - \lambda^2 y^2 f(y)dy$$

$$= \int_{-1}^{1} e^{i\lambda xy}\Big(-(1-y^2)i\lambda x\frac{df}{dy} - \lambda^2 y^2 f(y)\Big)dy + \Big[e^{i\lambda xy}(1-y^2)\frac{df}{dy}\Big]_{-1}^{1}$$
$$= \int_{-1}^{1} \Big(\frac{d}{dy}\Big((1-y^2)e^{i\lambda xy}\Big)i\lambda x f(y) - \lambda^2 y^2 e^{i\lambda xy} f(y)\Big)dy$$
$$= \int_{-1}^{1} \big((-2i\lambda xy - \lambda^2 x^2(1-y^2) - \lambda^2 y^2\big)e^{i\lambda xy} f(y)dy$$
$$= \int_{-1}^{1} \big(-2i\lambda xy - \lambda^2 x^2 + \lambda^2 x^2 y^2 - \lambda^2 y^2\big)e^{i\lambda xy} f(y)dy,$$

so we find that $KUf = UKf$.

Exercise 10.12 (ii) We have

$$\int_{-\infty}^{\infty} e^{i\omega z}\operatorname{sinc} a(u-z)\frac{dz}{2\pi} = \frac{e^{i\omega u}}{2a}\mathbb{I}_{(-a,a)}(\omega),$$

so with ϕ as in Theorem 11.4,

$$\begin{aligned}\frac{a}{\pi}\int_{-\infty}^{\infty}\phi(z)\operatorname{sinc} a(u-z)\,dz &= \frac{a}{\pi}\int_{-\infty}^{\infty}\int_{-a}^{a} f(\omega)e^{i\omega z}\frac{d\omega}{2a}\operatorname{sinc} a(u-z)\,dz\\ &= \frac{a}{\pi}\int_{-a}^{a} f(\omega)\int_{-\infty}^{\infty} e^{i\omega z}\operatorname{sinc} a(u-z)\,dz\frac{d\omega}{2a}\\ &= \int_{-a}^{a} f(\omega)\frac{e^{i\omega u}}{2a}\mathbb{I}_{(-a,a)}(\omega)d\omega\\ &= \phi(u).\end{aligned}$$

(iii) Continuing with the notation of Exercise 8.12, we write

$$Ug(x) = \int_{-1}^{1} e^{i\lambda xy} g(y)dy, \qquad U'g(x) = \int_{-1}^{1} e^{-i\lambda xy} g(y)dy$$

for $\lambda \in \mathbb{R}$. Then by substitution, we have

$$\begin{aligned}U'Ug(x) &= \int_{-1}^{1} e^{-i\lambda xz} Ug(z)dz\\ &= \int_{-1}^{1} e^{-i\lambda xz}\int_{-1}^{1} e^{i\lambda yz} g(y)dy\end{aligned}$$

so regrouping the terms and changing the order of integration, we have

$$\begin{aligned}
U'Ug(x) &= \int_{-1}^{1}\int_{-1}^{1} e^{i\lambda z(x-y)}dz g(y)dy \\
&= \int_{-1}^{1}\Big[\frac{e^{i\lambda z(x-y)}}{i\lambda(x-y)}\Big]_{-1}^{1} g(y)dy \\
&= \int_{-1}^{1}\frac{2\sin\lambda(x-y)}{\lambda(x-y)}g(y)dy \\
&= 2\int_{-1}^{1}\operatorname{sinc}(\lambda(x-y))\, g(y)dy.
\end{aligned}$$

Now let $\lambda = a$, and $f(az) = g(z)$; then

$$Tf(x) = \int_{-a}^{a} e^{ixy} f(y)\frac{dy}{2a} = \int_{-1}^{1} e^{ixaz} f(az)\frac{dz}{2} = \frac{1}{2}Ug(x).$$

Exercise 10.1

(i) We note that the integral is a convolution of f with e^{-t}, so we have

$$\mathcal{L}Sf(s) = \mathcal{L}f(s) - \frac{2}{1+s}\mathcal{L}f(s) = \frac{s-1}{s+1}\mathcal{L}f(s).$$

(ii) We can choose $f_0(s) = \sqrt{2}e^{-t}$ with $\mathcal{L}f_0(s) = \sqrt{2}/(1+s)$, and generate the sequence (f_n) by the recurrence relation $f_{n+1}(s) = Sf_n(t)$, so that

$$\mathcal{L}f_n(s) = \frac{\sqrt{2}(s-1)^n}{(s+1)^{n+1}},$$

which we recognize as the Laplace transforms of the given functions $h_n(t)$. The result follows by uniqueness of Laplace transforms.

Glossary of Linear Systems Terminology

(A,B,C,D) the standard continuous-time linear system determined by constant matrices of matching size
amplitude height of the crest above the average level, regarded as a wave;
BIBO bounded input and bounded output system;
Bode plot plot of log gain and phase against angular frequency ω;
Closed loop system with feedback loop;
differentiator operator of differentiation with respect to time t;
frequency domain linear system in terms of ω, where $s = i\omega$;
frequency response function transfer function $T(s)$ when $s = i\omega$ and $\omega \in \mathbb{R}$;
gain (or amplitude gain) modulus of the transfer function;
integrator operation of integration with respect to time t, from $t = 0$;
JCF Jordan canonical form of square matrix;
$\hat{\mathbf{L}}$ Laplace transform of L;
LHP Open left half-plane $\{s \in \mathbb{C} : \Re s < 0\}$;
MIMO multiple input and multiple output linear system;
Nyquist plot graph in complex plane of $T(i\omega)$ for $-\infty < \omega < \infty$;
Open loop system without feedback;
phase (or phase shift) argument of the transfer function;
resolvent of square matrix A is $(sI - A)^{-1}$;
RHP open right half-plane $\{s \in \mathbb{C} : \Re s > 0\}$;
s-domain linear system in terms of Laplace transform variable s;
SFL simple feedback loop system;
SISO single input and single output linear system;
state-space model linear system in terms of functions of time t;
summing junction operator for adding signals;
t time, with $t \in (0, \infty)$;
tap operator for making a signal go along two routes;

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6

transfer function $T(s)$ multiplies the Laplace transform of the input to get the Laplace transform of the output;

unit impulse function δ_0 the unit point mass at time $t = 0$, also known as Dirac delta function;

$\boldsymbol{\omega}$ angular frequency, often abbreviated to 'frequency'.

Appendix A
MATLAB Commands for Matrices

>> x=3, t=2 [this assigns values $x = 3$ and $t = 2$.
>>x*t [multiply x and t]
>>x+t [add x and t]
>>x/t [divide x by t]
>> x^(-1.5) [raises x to the power -1.5]
>>2*((x+t)^3) [computes $2(x+t)^3$]
>> pi [π area of disc of unit radius]
>>j [i complex number]
>>exp (3); [creates e^3]
>>A=[5,7; 9,-2], B=[1,2,3;4,5,6] builds the matrices

$$A = \begin{bmatrix} 5 & 7 \\ 9 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

>>inv(A) [computes the inverse matrix of A]
>>det(A) [computes the determinant of A]
>> trace(A) [computes the trace of A]
>>B' [computes the adjoint (conjugate transpose) B' of B]
>>B.' [computes the transpose $B^\top$ of B]
>> rref(A) [finds the reduced echelon form of A]
>> rank(A) [finds the rank of A]
>>poly(A) [finds the coefficients of the characteristic polynomial of square matrix A]
>>eig(A) [finds the eigenvalues of square matrix A, in a list]
>> jordan(A) [finds the Jordan canonical form of A]
>>[W,D]=eig(A) [gives a matrix W with columns that are eigenvectors of A and diagonal matrix D]
>> A^2 [computes the matrix product A^2]
>>A.^2 [creates matrix by squaring each entry of A; note dot]

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6

>>A*B [calculates the matrix product AB]
>>expm(A) [calculates the matrix exponential of A]
>>exp(A) [calculates the matrix formed by taking the exponential of each entry of A]
>>syms x [introduces algebraic variable x]
>>Q=lyap(A,P) [given positive definite P and square A, solves $AQ + QA' = -P$]
>> X=lyap(A,B,C) [solves $AX + XB = -C$]
>> t=−100:0.1:100 [introduces the vector [−100, −100+0.1,. . . ,100]]
>> y=(i*t+1).^(−1) [creates a vector with entries y=1/(it+1); the dot indicates that the operations are applies to each entry at a time]
>> plot(y) [plots the imaginary part of y versus the real part of y]
>> [r,q]=polynomialReduce(P,Q) [polynomial long division to find remainder and quotient in $P = Qq + r$]
>> angle (z) [computes the argument of the complex number z]
>> abs (z) [computes the modulus (absolute value) of complex number z]
>> laplace((t^2)*exp(2*t)) [computes the Laplace transform of t^2e^{2t}, and gives values in a variable s]
>> nyquist(T) [plots the Nyquist locus of a given real rational function]
>> bode(T) [created the Bode pole of a real rational function] Beware: nyquist and bode have difficulties with complex coefficients.
>> subs(G,t,s) [substitutes expression s for t in the expression G, for algebraic variables]
>> [g,c,d]=gcd(a,b) [computes the greatest common divisor of a and b and expresses the gcd as $ac + bd$]

Appendix B
SciLab Matrix Operations

SciLab is a simplified version of MATLAB, with a similar syntax; it cannot do much symbolic manipulation. Some commands are:

$--$ > coff(A) [for a square matrix A, computes matrix of cofactors adj$(sI - A)$]
$--$ > [N,d]=coff(A) [for a square matrix A, computes $N = \text{adj}(sI - A)$ and $d = \det(sI - A)$, so $N/d = (sI - A)^{-1}$]
$--$ > i=complex(0,1) [defines the complex number i]
$--$ > det(A) [computes the determinant of a square matrix A]
$--$ > eye(3,3) [gives the 3×3 identity matrix I_3]
$--$ >expm(A) [calculates the matrix exponential of a square A]
$--$ > [gcd, U]=bezout(p,q) [gcd gives greatest common divisor of polynomials p and q, and first column of U gives polynomials a, b such that gcd=ap+bq]
$--$ > inv(A) [computes the inverse of a square matrix A]
$--$ > [X]=lyap(A,C,'c') [computes X satisfying $X' * A + A * X = C$ for symmetric C; note that SciLab uses a different sign convention from MATLAB]
$--$ > plot(real(P),imag(P)) [used for Nyquist plots]
$--$ > s=poly(0,'s') [gives an algebraic variable]
$--$ > [r,q]=pdiv(P,Q) [gives quotient q and remainder r for P=Qq+r in polynomials]
$--$ > [radius angle]=polar(z) [gives $[r\ \theta]$ where $z = re^{i\theta}$]
$--$ > rank(A) [computes the rank of a matrix A]
$--$ > rref (A) [computes the row reduced echelon form of a matrix A]
$--$ > spec (A) [computes the spectrum (eigenvalues) of a square matrix A]
$--$ > [R,D]=spec (A) [for a square matrix A, computes the eigenvectors in R and eigenvalues in diagonal matrix D]
$--$ > trace(A) [computes the trace of a square matrix A]

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6

References

1. Aboufadel E, Schlicker S (1999) Discovering wavelets. Wiley, New York
2. Arendt W, Batty CJK, Heiber M, Neubrander F (2001) Vector-valued Laplace transforms and cauchy problems. Birkhauser
3. Benson DJ (2007) Music: a mathematical offering. Cambridge University Press
4. Beurling A (1948) On two problems concerning linear transformations on Hilbert space. Acta Math 81:239–255
5. Bhatia R, Rosenthal P (1997) How and why to solve the operator equation $AX - XB = Y$. Bull Lond Math Soc 29:1–21
6. Birkhoff G, MacLane S (1965) A survey of modern algebra, 3rd edn. MacMillan
7. Blower G (2009) Random matrices: high dimensional phenomena. Cambridge University Press
8. Blyth TS, Robertson EF (2002) Basic linear algebra, 2nd edn. Springer
9. Bohr H (1925) Zur Theorie der Fastperiodischen Funtionen I, eine Verallgemeinerung der Theorie der Fourierreihe. Acta Math 45:29–127
10. Boros G, Moll VH (2004) Irresistable integrals. Cambridge University Press
11. Chen CT (2012) Linear system theory and design, 4th edn. Oxford University Press
12. Dorf RC, Bishop RH (2011) Modern control systems, 12th edn. Pearson
13. Doyle JC, Francis BA, Tannenbaum AR (2009) Feedback control theory. Dover
14. Dym H, McKean HP (1985) Fourier series and integrals. Academic Press
15. Foias C, Frazho AE (1990) The commutant lifting approach to interpolation problems. Birkhauser
16. Freud G (1971) Orthogonal polynomials. Pergamon Press
17. Gesztesy F, Simon B (1996) The xi function. Acta Math 176:49–71
18. Goldstein JA (1985) Semigroups of linear operators and applications. Oxford University Press
19. Gradsteyn IS, Ryzhik IM (1965) Table of integrals, series and products. Academic Press
20. Hartley B, Hawkes TO (1970) Rings, modules and linear algebra. Chapman and Hall
21. Hastings HM (1982) May–Wigner stability theorem for connected matrices. Bull Amer Math Soc 7:387–388
22. Helton JW (1974) Discrete time systems, operator models and scattering theory. J Funct Anal 16:15–38
23. Helton JW, Putinar M (2007) Positive polynomials in scalar and matrix variables, the spectral theorem and optimization. Theta Ser Adv Math 7 Theta
24. Helton JW, Mai T, Speicher R (2018) Applications of realizations (aka linearizations) to free probability. J Funct Anal 274:1–79
25. Hespanha JP (2009) Linear systems theory. Princeton and Oxford
26. Hille E (1972) Methods in classical and functional analysis. Addison-Wesley

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6

27. Hille E (2012) Analytic function theory, vol 1. AMS Chelsea Publishing
28. Horn RA, Johnson CA (1985) Matrix analysis. Cambridge University Press
29. Hungerford TW (1974) Algebra. Springer
30. Hurwitz A (1895) Über die Bedingen, unter welchen eine Gelichung nur Wurzeln mit negativen reellen Theilen besitzen. Math Ann (2) 46:273–284
31. Janik RA, Nowak MA, Zahed I (1997) Non-Hermitian random matrix models. Nuclear Phys B 501:603–642
32. Kang CG (2016) Origin of stability analysis "On governors" by J.C. Maxwell. IEEE Mag 36:77
33. Katznelson Y (2002) An introduction to harmonic analysis. Cambridge University Press
34. Koosis P (1980) Introduction to H_p spaces. Cambridge University Press
35. Lebedev NN (1965) Special functions and their applications. Prentice-Hall
36. Littlewood JE (1925) On inequalities in the theory of functions. Proc Lond Math Soc (2) 23:481–519
37. Magnus A (1995) Painlevé-type differential equations for the recurrence coefficients of semiclassical orthogonal polynomials. J Comput Appl Math 57:215–237
38. Magnus W, Winkler S (1979) Hill's equation. Dover
39. Martin PA (2022) London Mathematical Society Newsletter, March 2022
40. Maxwell JC (1868) On governors. Proc R Soc Lond 16:270–283
41. McKean HP, Moll V (1999) Elliptic curves. Cambridge University Press
42. Megretskii AV, Peller VV, Treil SR (1995) The inverse spectral problem for self-adjoint Hankel operators. Acta Math 174:241–309
43. Mehta ML (1991) Random matrices, 2nd edn. Academic Press
44. Mohlenkamp MJ, Pereyra MC (2008) Wavelets, their friends and what they can do for you. European Mathematical Society
45. Partington JR (2004) Linear operators and linear systems. Cambridge University Press
46. Pastur LA (1972) The spectrum of random matrices. Teoret Mat Fiz 10:102–112
47. Power SC (1982) Hankel operators on Hilbert space. Pitman Research Notes
48. Reuter GEH (1957) Denumerable Markov processes and the associated contraction semigroups on ℓ. Acta Math 97:1–46
49. Rudin W (1974) Real and complex analysis, 2nd edn. McGraw-Hill
50. Sansone G (1959) Orthogonal functions. Interscience
51. Simmons GF (1963) Introduction to topology and modern analysis. McGraw-Hill
52. Sitwell J (2008) Naive lie theory. Springer
53. Sneddon IN (1972) The use of integral transforms. McGraw-Hill
54. Szegö G (1938) Orthogonal polynomials. American Mathematical Society
55. Telatar E (1999) Capacity of multi-antenna Gaussian channels. Eur Trans Telecommun 10:585–595
56. Titchmarsh EC (1937) The theory of functions. Oxford
57. Titchmarsh EC (1937) Introduction to the theory of Fourier integrals. Oxford
58. Voiculescu D (1986) Addition of certain noncommuting random variables. J Funct Anal 66:323–346
59. Watson GN (1922) Treatise on the theory of Bessel functions. Cambridge University Press
60. Whitelaw TA (1983) An introduction to linear algebra. Blackie, Glasgow
61. Whittaker ET, Watson GN (2021) A course of modern analysis: an introduction to the general theory of infinite processes and of analytic functions; with an account of the principal transcendental functions. Cambridge University Press
62. Wiener N, Masani P (1957) The prediction theory of multivariate stochastic processes. Acta Math 98:111–150
63. Wigner EP (1955) Characteristic vectors of bordered matrices with infinite dimensions. Ann Math (2) 62:548–564
64. Wigner EP (1958) On the distribution of the roots of certain symmetric matrices. Ann Math (2) 67:325–327
65. Young N (1988) An introduction to Hilbert space. Cambridge University Press

Index

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

G. Blower, *Linear Systems*, Mathematical Engineering,
https://doi.org/10.1007/978-3-031-21240-6