Phys 6124 zipped!

World Wide Quest to Tame Math Methods

Predrag Cvitanović

September 13, 2020

Overview

> The career of a young theoretical physicist consists of treating the harmonic oscillator in ever-increasing levels of abstraction.
> — Sidney Coleman

I am leaving the course notes here, not so much for the notes themselves –they cannot be understood on their own, without the (unrecorded) live lectures– but for the hyperlinks to various source texts you might find useful later on in your research.

We change the topics covered year to year, in hope that they reflect better what a graduate student needs to know. This year's experiment is taking the course online. Let's work together to make it work for everyone in the course.

- *Course outline : An ode in 15 stanzas*
- *Course policy*
- *My teaching philosophy : Bologna*
- *How does one pronounce 'Euler'? 'Cvitanovic'?*

After a while you might notice a pattern: Every week we start with something obvious that you already know, let mathematics lead us on, and then suddenly end up someplace surprising and highly non-intuitive.

Contents

mathematical methods - week 1

Linear algebra

Georgia Tech PHYS-6124

Homework HW #1 due Tuesday, August 25, 2020

== show all your work for maximum credit,
== put labels, title, legends on any graphs
== acknowledge study group member, if collective effort
== if you are LaTeXing, here is the source code

Exercise 1.1 *Trace-log of a matrix*	4 points
Exercise 1.2 *Stability, diagonal case*	2 points
Exercise 1.3 *The matrix square root*	4 points
Exercise 1.4 *Exponential of a matrix of Jordan form*	4 bonus points

Total of 10 points = 100 % score. Bonus points accumulate, can help you later if you miss a few problems.

edited September 13, 2020

Week 1 syllabus

Tuesday, August 18, 2020

> Diagonalizing the matrix: that's the key to the whole thing.
> — Governor Arnold Schwarzenegger

Anything prefixed by AWH, like "Kronecker product AWH eq. (2.55)" refers to Arfken, Weber & Harris [1] *Mathematical Methods for Physicists: A Comprehensive Guide* (get it in GaTech Library). Light blue text in this PDF is a live hyperlink. When you encounter a web login: All copyright-protected references are on a password protected site. What password? Have your ears up in the class; the password will be posted on the Canvas for a week or so, so remember to write it down.

This week's lectures are related to AWH Chapter 2 *Determinants and matrices* (click here) and Chapter 6 *Eigenvalue problems* (click here)

- Sect. 1.2 *Matrix-valued functions*, AWH p. 113 Functions of Matrices

  - AWH Section 2.2 *Matrices*
  - *Matrices : 2 kinds*
  - *Derivative of a matrix function*
  - *Exponential, logarithm of a matrix*
  - AWH Example 2.2.6 *Exponential of a diagonal matrix*
  - *Determinant is a volume*
  - *log det = tr log* (updated Aug 18, 2020)
  - *Multi-matrix functions* (optional, for QM inclined)

- Sect. 1.3 *A linear diversion*
  There are two representations of exponential of constant matrix, the Taylor series and the compound interest (Euler product) formulas (1.10). If the matrix (for example, the Hamiltonian) changes with time, the exponential has to be time-ordered. The Taylor series generalizes to the nested integrals formula (**??**), and the Euler product to time-ordered product (1.11). The first is used in formal quantum-mechanical calculations, the second in practical, numerical calculations.

  - *Linear differential equations*
  - *Nonlinear differential equations*

- Sect. 1.4 *Eigenvalues and eigenvectors*
  Hamilton-Cayley equation, projection operators (1.21), any matrix function is evaluated by spectral decomposition (1.24). Work through example 1.3.

  - AWH Section 6.1 *Eigenvalue Equations*
  - *Eigenvalues and eigenvectors*
  - *What's the deal with Hamilton-Cayley?*

- *Spectral decomposition*
- *Spectral decomposition and completeness*
- *Right, left eigenvectors*
- *A projection operators workout*
- *Jordan form* AWH p. 324 *Defective matrices* (optional, for QM inclined)
- *Are there Jordan form matrices in physics?* (optional, for QM inclined)

1.1 Other sources

The subject of linear algebra is a vast and very alive research area, generates innumerable tomes of its own, and is way beyond what we can exhaustively cover here. A few resources that you might find helpful going forward:

Linear operators and matrices reading (optional reading for week 1, not required for this course):

- Stone and Goldbart [12], *Mathematics for Physics: A Guided Tour for Graduate Students*, Appendix A. This is an advanced summary where you will find almost everything one needs to know.
- In sect. 1.2 I make matrix functions appear easier than they really are. For an in-depth discussion, consult Golub and Van Loan [6] *Matrix Computations*, chap. 9 *Functions of Matrices* (click here). exercise 1.3
- Much more than you ever wanted to know about linear algebra: Axler [2] *Down with determinants!* (click here).
- Karan Shah: I like Grant Sanderson's 3Blue1Brown geometrical explanations of linear algebra eigenstuff (click here).

Question 1.1. Henriette Roux finds course notes confusing

Q Couldn't you use one single, definitive text for methods taught in the course?

A It's a grad school, so it is research focused - I myself am (re)learning the topics that we are going through the course, using various sources. My emphasis in this course is on understanding and meaning, not on getting all signs and 2π's right, and I find reading about the topic from several perspectives helpful. But if you really find one book more comfortable, nearly all topics are covered in Arfken, Weber & Harris [1].

1.2 Matrix-valued functions

What is a matrix?

—Werner Heisenberg (1925)

What is the matrix?

—-Keanu Reeves (1999)

(optional, for QM inclined)

Why should a working physicist care about linear algebra? Physicists were blissfully ignorant of group theory until 1920's, but with Heisenberg's sojourn in Helgoland, everything changed. Quantum Mechanics was formulated as

$$|\phi(t)\rangle = \hat{U}^t|\phi(0)\rangle\,, \qquad \hat{U}^t = e^{-\frac{i}{\hbar}t\hat{H}}\,, \tag{1.1}$$

where $|\phi(t)\rangle$ is the quantum wave function at time t, $\hat{U}^t$ is the unitary quantum evolution operator, and $\hat{H}$ is the Hamiltonian operator. Fine, but what does this equation *mean*? In the first lecture we deconstruct it, make $\hat{U}^t$ computationally explicit as a the time-ordered product (1.12).

It would not be fair to students to expect a prior exposure to Heisenberg's matrix quantum mechanics (1.1), so if you do not 'get' the QM comments of this section, it's OK. It is not needed for what follows, and I'll do it in the class only if you request me to do it.

The matrices that have to be evaluated are very high-dimensional, in principle infinite dimensional, and the numerical challenges can quickly get out of hand. What made it possible to solve these equations analytically in 1920's for a few iconic problems, such as the hydrogen atom, are the symmetries, or in other words group theory, a subject of another course, our group theory course.

Whenever you are confused about an "operator", think "matrix". Here we recapitulate a few matrix algebra concepts that we found essential. The punch line is (1.27): Hamilton-Cayley equation $\prod(\mathbf{M} - \lambda_i\mathbf{1}) = 0$ associates with each distinct root λ_i of a matrix $\mathbf{M}$ a projection onto ith vector subspace

$$P_i = \prod_{j\neq i} \frac{\mathbf{M} - \lambda_j\mathbf{1}}{\lambda_i - \lambda_j}\,.$$

What follows - for this week - is a jumble of Predrag's notes. If you understand the examples, we are on the roll. If not, ask :)

How are we to think of the quantum operator (1.1)

$$\hat{H} = \hat{T} + \hat{V}\,, \qquad \hat{T} = \hat{p}^2/2m\,, \qquad \hat{V} = V(\hat{q})\,, \tag{1.2}$$

corresponding to a classical Hamiltonian $H = T + V$, where T is kinetic energy, and V is the potential?

Expressed in terms of basis functions, the quantum evolution operator is an infinite-dimensional matrix; if we happen to know the eigenbasis of the Hamiltonian, the problem is solved already. In real life we have to guess that some complete basis set is good starting point for solving the problem, and go from there. In practice we truncate such operator representations to finite-dimensional matrices, so it pays to recapitulate a few relevant facts about matrix algebra and some of the properties of functions of finite-dimensional matrices.

1.3 A linear diversion

(Notes based of ChaosBook.org/chapters/stability.pdf)

Linear fields are the simplest vector fields, described by linear differential equations which can be solved explicitly, with solutions that are good for all times. The state space for linear differential equations is $\mathcal{M} = \mathbb{R}^d$, and the equations of motion are written in terms of a vector x and a constant stability matrix A as

$$\dot{x} = v(x) = Ax\,. \tag{1.3}$$

Solving this equation means finding the state space trajectory

$$x(t) = (x_1(t), x_2(t), \ldots, x_d(t))$$

passing through a given initial point x_0. If $x(t)$ is a solution with $x(0) = x_0$ and $y(t)$ another solution with $y(0) = y_0$, then the linear combination $ax(t) + by(t)$ with $a, b \in \mathbb{R}$ is also a solution, but now starting at the point $ax_0 + by_0$. At any instant in time, the space of solutions is a d-dimensional vector space, spanned by a basis of d linearly independent solutions.

How do we solve the linear differential equation (1.3)? If instead of a matrix equation we have a scalar one, $\dot{x} = \lambda x$, the solution is $x(t) = e^{t\lambda}x_0$. In order to solve the d-dimensional matrix case, it is helpful to rederive this solution by studying what happens for a short time step δt. If time $t = 0$ coincides with position $x(0)$, then

$$\frac{x(\delta t) - x(0)}{\delta t} = \lambda x(0)\,, \tag{1.4}$$

which we iterate m times to obtain Euler's formula for compounding interest

$$x(t) \approx \left(1 + \frac{t}{m}\lambda\right)^m x(0) \approx e^{t\lambda}x(0)\,. \tag{1.5}$$

The term in parentheses acts on the initial condition $x(0)$ and evolves it to $x(t)$ by taking m small time steps $\delta t = t/m$. As $m \to \infty$, the term in parentheses converges to $e^{t\lambda}$. Consider now the matrix version of equation (1.4):

$$\frac{x(\delta t) - x(0)}{\delta t} = A\,x(0)\,. \tag{1.6}$$

A representative point x is now a vector in $\mathbb{R}^d$ acted on by the matrix A, as in (1.3). Denoting by $\mathbf{1}$ the identity matrix, and repeating the steps (1.4) and (1.5) we obtain Euler's formula

$$x(t) = J^t x(0)\,, \qquad J^t = e^{tA} = \lim_{m\to\infty}\left(\mathbf{1} + \frac{t}{m}A\right)^m\,. \tag{1.7}$$

We will find this definition for the exponential of a matrix helpful in the general case, where the matrix $A = A(x(t))$ varies along a trajectory.

Now that we have some feeling for the qualitative behavior of a linear flow, we are ready to return to the nonlinear case. Consider an infinitesimal perturbation of the initial state, $x_0 + \delta x(x_0, 0)$. How do we compute the exponential (1.7) that describes linearized perturbation $\delta x(x_0, t)$?

$$x(t) = f^t(x_0)\,, \quad \delta x(x_0, t) \;=\; J^t(x_0)\,\delta x(x_0, 0)\,. \tag{1.8}$$

The equations are linear, so we should be able to integrate them–but in order to make sense of the answer, we derive this integration step by step. The Jacobian matrix is computed by integrating the equations of variations

$$\dot{x}_i = v_i(x)\,, \quad \dot{\delta x}_i = \sum_j A_{ij}(x)\delta x_j \tag{1.9}$$

Consider the case of a general, non-stationary trajectory $x(t)$. The exponential of a constant matrix can be defined either by its Taylor series expansion or in terms of the Euler limit (1.7):

$$e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = \lim_{m\to\infty} \left(\mathbf{1} + \frac{t}{m} A\right)^m . \tag{1.10}$$

Taylor expanding is fine if A is a constant matrix. However, only the second, tax-accountant's discrete step definition of an exponential is appropriate for the task at hand. For dynamical systems, the local rate of neighborhood distortion $A(x)$ depends on where we are along the trajectory. The linearized neighborhood is deformed along the flow, and the m discrete time-step approximation to J^t is therefore given by a generalization of the Euler product (1.10):

$$\begin{aligned} J^t &= \lim_{m\to\infty} \prod_{n=m}^{1} (\mathbf{1} + \delta t A(x_n)) = \lim_{m\to\infty} \prod_{n=m}^{1} e^{\delta t\, A(x_n)} \\ &= \lim_{m\to\infty} e^{\delta t\, A(x_m)} e^{\delta t\, A(x_{m-1})} \cdots e^{\delta t\, A(x_2)} e^{\delta t\, A(x_1)} , \end{aligned} \tag{1.11}$$

where $\delta t = (t - t_0)/m$, and $x_n = x(t_0 + n\delta t)$. Indexing of the products indicates that the successive infinitesimal deformation are applied by multiplying from the left. The $m \to \infty$ limit of this procedure is the formal integral

$$J^t_{ij}(x_0) = \left[\mathbf{T} e^{\int_0^t d\tau A(x(\tau))}\right]_{ij} , \tag{1.12}$$

where $\mathbf{T}$ stands for time-ordered integration, *defined* as the continuum limit of the successive multiplications (1.11). This integral formula for J^t is the finite time companion of the differential definition

$$\dot{J}(t) = A(t)J(t), \tag{1.13}$$

with the initial condition $J(0) = 1$. The definition makes evident important properties of Jacobian matrices, such as their being multiplicative along the flow,

$$J^{t+t'}(x) = J^{t'}(x')\, J^t(x), \qquad \text{where } x' = f^t(x_0)\,, \tag{1.14}$$

which is an immediate consequence of the time-ordered product structure of (1.11). However, in practice J is evaluated by integrating differential equation (1.13) along with the ODEs (3.6) that define a particular flow.

1.4 Eigenvalues and eigenvectors

10. Try to leave out the part that readers tend to skip.
— Elmore Leonard's Ten Rules of Writing.

Eigenvalues of a $[d\times d]$ matrix $\mathbf{M}$ are the roots of its characteristic polynomial

$$\det(\mathbf{M}-\lambda\mathbf{1}) = \prod(\lambda_i-\lambda) = 0\,. \tag{1.15}$$

Given a nonsingular matrix $\mathbf{M}$, $\det\mathbf{M}\neq 0$, with all $\lambda_i \neq 0$, acting on d-dimensional vectors $\mathbf{x}$, we would like to determine *eigenvectors* $\mathbf{e}^{(i)}$ of $\mathbf{M}$ on which $\mathbf{M}$ acts by scalar multiplication by eigenvalue λ_i

$$\mathbf{M}\,\mathbf{e}^{(i)} = \lambda_i\mathbf{e}^{(i)}\,. \tag{1.16}$$

If $\lambda_i \neq \lambda_j$, $\mathbf{e}^{(i)}$ and $\mathbf{e}^{(j)}$ are linearly independent. There are at most d distinct eigenvalues, which we assume have been computed by some method, and ordered by their real parts, $\mathrm{Re}\,\lambda_i \geq \mathrm{Re}\,\lambda_{i+1}$.

If all eigenvalues are distinct $\mathbf{e}^{(j)}$ are d linearly independent vectors which can be used as a (non-orthogonal) basis for any d-dimensional vector $\mathbf{x}\in\mathbb{R}^d$

$$\mathbf{x} = x_1\,\mathbf{e}^{(1)} + x_2\,\mathbf{e}^{(2)} + \cdots + x_d\,\mathbf{e}^{(d)}\,. \tag{1.17}$$

From (1.16) it follows that

$$(\mathbf{M}-\lambda_i\mathbf{1})\,\mathbf{e}^{(j)} = (\lambda_j-\lambda_i)\,\mathbf{e}^{(j)}\,,$$

matrix $(\mathbf{M}-\lambda_j\mathbf{1})$ annihilates $\mathbf{e}^{(j)}$, the product of all such factors annihilates any vector, and the matrix $\mathbf{M}$ satisfies its characteristic equation

$$\prod_{i=1}^{d}(\mathbf{M}-\lambda_i\mathbf{1}) = 0\,. \tag{1.18}$$

This humble fact has a name: the Hamilton-Cayley theorem. If we delete one term from this product, we find that the remainder projects $\mathbf{x}$ from (1.17) onto the corresponding eigenspace:

$$\prod_{j\neq i}(\mathbf{M}-\lambda_j\mathbf{1})\,\mathbf{x} = \prod_{j\neq i}(\lambda_i-\lambda_j)\,x_i\mathbf{e}^{(i)}\,.$$

Dividing through by the $(\lambda_i-\lambda_j)$ factors yields the *projection operators*

$$P_i = \prod_{j\neq i}\frac{\mathbf{M}-\lambda_j\mathbf{1}}{\lambda_i-\lambda_j}\,, \tag{1.19}$$

which are *orthogonal* and *complete*:

$$P_i P_j = \delta_{ij} P_j \,, \quad (\text{no sum on } j)\,, \qquad \sum_{i=1}^{r} P_i = \mathbf{1}\,, \tag{1.20}$$

with the dimension of the ith subspace given by $d_i = \text{tr}\, P_i$. For each distinct eigenvalue λ_i of $\mathbf{M}$,

$$(\mathbf{M} - \lambda_j \mathbf{1}) P_j = P_j (\mathbf{M} - \lambda_j \mathbf{1}) = 0\,, \tag{1.21}$$

the colums/rows of P_i are the right/left eigenvectors $\mathbf{e}^{(k)}$, $\mathbf{e}_{(k)}$ of $\mathbf{M}$ which (provided $\mathbf{M}$ is not of Jordan type, see example 1.1) span the corresponding linearized subspace.

The main take-home is that once the distinct non-zero eigenvalues $\{\lambda_i\}$ are computed, projection operators are polynomials in $\mathbf{M}$ which need no further diagonalizations or orthogonalizations. It follows from the characteristic equation (1.21) that λ_i is the eigenvalue of $\mathbf{M}$ on P_i subspace:

$$\mathbf{M}\, P_i = \lambda_i P_i \qquad (\text{no sum on } i)\,. \tag{1.22}$$

Using $\mathbf{M} = \mathbf{M}\,\mathbf{1}$ and completeness relation (1.20) we can rewrite $\mathbf{M}$ as

$$\mathbf{M} = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_d P_d\,. \tag{1.23}$$

Any matrix function $f(\mathbf{M})$ takes the scalar value $f(\lambda_i)$ on the P_i subspace, $f(\mathbf{M})\, P_i = f(\lambda_i)\, P_i$, and is thus easily evaluated through its *spectral decomposition* (see AWH Exercise 3.5.34)

remark 1.1

$$f(\mathbf{M}) = \sum_i f(\lambda_i) P_i\,. \tag{1.24}$$

This, of course, is the reason why anyone but a fool works with irreducible reps: they reduce matrix (AKA "operator") evaluations to manipulations with numbers.

By (1.21) every column of P_i is proportional to a right eigenvector $\mathbf{e}^{(i)}$, and its every row to a left eigenvector $\mathbf{e}_{(i)}$. In general, neither set is orthogonal, but by the idempotence condition (1.20), they are mutually orthogonal,

$$\mathbf{e}_{(i)} \cdot \mathbf{e}^{(j)} = c\, \delta_i^j\,. \tag{1.25}$$

The non-zero constant c is convention dependent and not worth fixing, unless you feel nostalgic about Clebsch-Gordan coefficients. We shall set $c = 1$. Then it is convenient to collect all left and right eigenvectors into a single matrix.

Example 1.1. ***Degenerate eigenvalues.*** *While for a matrix with generic real elements all eigenvalues are distinct with probability 1, that is not true in presence of symmetries, or spacial parameter values (bifurcation points). What can one say about situation where d_α eigenvalues are degenerate, $\lambda_\alpha = \lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+d_\alpha-1}$? Hamilton-Cayley (1.18) now takes form*

$$\prod_{\alpha=1}^{r} (\mathbf{M} - \lambda_\alpha \mathbf{1})^{d_\alpha} = 0\,, \qquad \sum_\alpha d_\alpha = d\,. \tag{1.26}$$

We distinguish two cases:

M can be brought to diagonal form. *The characteristic equation (1.26) can be replaced by the minimal polynomial,*

$$\prod_{\alpha=1}^{r}(\mathbf{M}-\lambda_\alpha\mathbf{1})=0\,, \tag{1.27}$$

where the product includes each distinct eigenvalue only once. Matrix $\mathbf{M}$ acts multiplicatively

$$\mathbf{M}\,\mathbf{e}^{(\alpha,k)}=\lambda_i\mathbf{e}^{(\alpha,k)}\,, \tag{1.28}$$

on a d_α-dimensional subspace spanned by a linearly independent set of basis eigenvectors $\{\mathbf{e}^{(\alpha,1)},\mathbf{e}^{(\alpha,2)},\cdots,\mathbf{e}^{(\alpha,d_\alpha)}\}$. This is the easy case. Luckily, if the degeneracy is due to a finite or compact symmetry group, relevant $\mathbf{M}$ matrices can always be brought to such Hermitian, diagonalizable form.

M can only be brought to upper-triangular, Jordan form. *This is the messy case, so we only illustrate the key idea in example 1.2.* *(optional, for QM inclined)*

Example 1.2. ***Decomposition of 2-dimensional vector spaces:*** *Enumeration of every possible kind of linear algebra eigenvalue / eigenvector combination is beyond what we can reasonably undertake here. However, enumerating solutions for the simplest case, a general $[2\times2]$ non-singular matrix*

$$\mathbf{M}=\left[\begin{array}{cc} M_{11} & M_{12}\\ M_{21} & M_{22}\end{array}\right].$$

takes us a long way toward developing intuition about arbitrary finite-dimensional matrices. The eigenvalues

$$\lambda_{1,2}=\frac{1}{2}\mathrm{tr}\,\mathbf{M}\pm\frac{1}{2}\sqrt{(\mathrm{tr}\,\mathbf{M})^2-4\det\mathbf{M}} \tag{1.29}$$

are the roots of the characteristic (secular) equation (1.15):

$$\begin{aligned}\det(\mathbf{M}-\lambda\,\mathbf{1}) &= (\lambda_1-\lambda)(\lambda_2-\lambda)\\ &= \lambda^2-\mathrm{tr}\,\mathbf{M}\,\lambda+\det\mathbf{M}\;=\;0\,.\end{aligned}$$

Distinct eigenvalues case *has already been described in full generality. The left/right eigenvectors are the rows/columns of projection operators*

$$P_1=\frac{\mathbf{M}-\lambda_2\mathbf{1}}{\lambda_1-\lambda_2}\,,\qquad P_2=\frac{\mathbf{M}-\lambda_1\mathbf{1}}{\lambda_2-\lambda_1}\,,\qquad \lambda_1\neq\lambda_2\,. \tag{1.30}$$

Degenerate eigenvalues. *If $\lambda_1=\lambda_2=\lambda$, we distinguish two cases: (a) $\mathbf{M}$ can be brought to diagonal form. This is the easy case. (b) $\mathbf{M}$ can be brought to* Jordan form, *with zeros everywhere except for the diagonal, and some 1's directly above it; for a $[2\times2]$ matrix the Jordan form is* *(optional, for QM inclined)*

$$\mathbf{M}=\left[\begin{array}{cc}\lambda & 1\\ 0 & \lambda\end{array}\right],\qquad \mathbf{e}^{(1)}=\begin{bmatrix}1\\0\end{bmatrix},\;\;\mathbf{v}^{(2)}=\begin{bmatrix}0\\1\end{bmatrix}.$$

$\mathbf{v}^{(2)}$ helps span the 2-dimensional space, $(\mathbf{M}-\lambda)^2\mathbf{v}^{(2)}=0$, but is not an eigenvector, as $\mathbf{M}\mathbf{v}^{(2)} = \lambda\mathbf{v}^{(2)} + \mathbf{e}^{(1)}$. For every such Jordan $[d_\alpha \times d_\alpha]$ block there is only one eigenvector per block. Noting that

$$\mathbf{M}^m = \left[\begin{array}{cc} \lambda^m & m\lambda^{m-1} \\ 0 & \lambda^m \end{array} \right] ,$$

we see that instead of acting multiplicatively on $\mathbb{R}^2$, Jacobian matrix $J^t = \exp(t\mathbf{M})$

$$e^{t\mathbf{M}} \begin{pmatrix} u \\ v \end{pmatrix} = e^{t\lambda} \begin{pmatrix} u+tv \\ v \end{pmatrix} \tag{1.31}$$

picks up a power-low correction. That spells trouble (logarithmic term $\ln t$ if we bring the extra term into the exponent).

Example 1.3. ***Projection operator decomposition in 2 dimensions:*** *Let's illustrate how the distinct eigenvalues case works with the $[2\times 2]$ matrix*

$$\mathbf{M} = \left[\begin{array}{cc} 4 & 1 \\ 3 & 2 \end{array} \right] .$$

Its eigenvalues $\{\lambda_1, \lambda_2\} = \{5, 1\}$ are the roots of (1.29):

$$\det(\mathbf{M} - \lambda \mathbf{1}) = \lambda^2 - 6\,\lambda + 5 = (5-\lambda)(1-\lambda) = 0\,.$$

That $\mathbf{M}$ satisfies its secular equation (Hamilton-Cayley theorem) can be verified by explicit calculation:

$$\left[\begin{array}{cc} 4 & 1 \\ 3 & 2 \end{array} \right]^2 - 6 \left[\begin{array}{cc} 4 & 1 \\ 3 & 2 \end{array} \right] + 5 \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right] = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} \right] .$$

Associated with each root λ_i is the projection operator *(1.30)*

$$P_1 = \frac{1}{4}(\mathbf{M} - \mathbf{1}) = \frac{1}{4} \left[\begin{array}{cc} 3 & 1 \\ 3 & 1 \end{array} \right] \tag{1.32}$$

$$P_2 = \frac{1}{4}(\mathbf{M} - 5\cdot\mathbf{1}) = \frac{1}{4} \left[\begin{array}{cc} 1 & -1 \\ -3 & 3 \end{array} \right] . \tag{1.33}$$

Matrices P_i are orthonormal and complete, The dimension of the ith subspace is given by $d_i = \mathrm{tr}\, P_i$; in case at hand both subspaces are 1-dimensional. From the characteristic equation it follows that P_i satisfies the eigenvalue equation $\mathbf{M}\,P_i = \lambda_i P_i$. Two consequences are immediate. First, we can easily evaluate any function of $\mathbf{M}$ by spectral decomposition, for example

$$\mathbf{M}^7 - 3\cdot\mathbf{1} = (5^7-3)P_1 + (1-3)P_2 = \left[\begin{array}{cc} 58591 & 19531 \\ 58593 & 19529 \end{array} \right] .$$

Second, as P_i satisfies the eigenvalue equation, its every column is a right eigenvector, and every row a left eigenvector. Picking first row/column we get the eigenvectors:

$$\{\mathbf{e}^{(1)}, \mathbf{e}^{(2)}\} = \{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} , \begin{bmatrix} 1 \\ -3 \end{bmatrix} \}$$

$$\{\mathbf{e}_{(1)}, \mathbf{e}_{(2)}\} = \{ \begin{bmatrix} 3 \\ 1 \end{bmatrix} , \begin{bmatrix} 1 \\ -1 \end{bmatrix} \} ,$$

with overall scale arbitrary. The matrix is not symmetric, so $\{\mathbf{e}^{(j)}\}$ do not form an orthogonal basis. The left-right eigenvector dot products $\mathbf{e}_{(j)} \cdot \mathbf{e}^{(k)}$, however, are orthogonal as in (1.25), by inspection. (Continued in example ??.)

Example 1.4. ***Computing matrix exponentials.*** *If A is diagonal (the system is uncoupled), then e^{tA} is given by*

$$\exp\begin{pmatrix} \lambda_1 t & & & \\ & \lambda_2 t & & \\ & & \ddots & \\ & & & \lambda_d t \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{\lambda_d t} \end{pmatrix}.$$

If A is diagonalizable, $A = FDF^{-1}$, where D is the diagonal matrix of the eigenvalues of A and F is the matrix of corresponding eigenvectors, the result is simple: $A^n = (FDF^{-1})(FDF^{-1})\ldots(FDF^{-1}) = FD^nF^{-1}$. Inserting this into the Taylor series for e^x gives $e^{At} = Fe^{Dt}F^{-1}$.

But A may not have d linearly independant eigenvectors, making F singular and forcing us to take a different route. To illustrate this, consider $[2\times 2]$ matrices. For any linear system in $\mathbb{R}^2$, there is a similarity transformation

$$B = U^{-1}AU\,,$$

where the columns of U consist of the generalized eigenvectors of A such that B has one of the following forms:

$$B = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}, \qquad B = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \qquad B = \begin{bmatrix} \mu & -\omega \\ \omega & \mu \end{bmatrix}.$$

These three cases, called normal forms, correspond to A having (1) distinct real eigenvalues, (2) degenerate real eigenvalues, or (3) a complex pair of eigenvalues. It follows that

$$e^{Bt} = \begin{bmatrix} e^{\lambda t} & 0 \\ 0 & e^{\mu t} \end{bmatrix}, \qquad e^{Bt} = e^{\lambda t}\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, \qquad e^{Bt} = e^{at}\begin{bmatrix} \cos bt & -\sin bt \\ \sin bt & \cos bt \end{bmatrix},$$

and $e^{At} = Ue^{Bt}U^{-1}$. What we have done is classify all $[2\times 2]$ matrices as belonging to one of three classes of geometrical transformations. The first case is scaling, the second is a shear, and the third is a combination of rotation and scaling. The generalization of these normal forms to $\mathbb{R}^d$ is called the Jordan normal form. *(J. Halcrow)*

Example 1.5. ***Determinants and traces.***

The usual textbook expression for a determinant is the sum of products of all permutations

$$\det M = \sum_{\{\pi\}} (-1)^\pi M_{1,\pi_1} M_{2,\pi_2} \cdots M_{m,\pi_m} \tag{1.34}$$

where M is a $[m\times m]$ matrix, $\{\pi\}$ denotes the set of permutations of m symbols, π_k is the permutation π applied to k, and $(-1)^\pi = \pm 1$ is the parity of permutation π. For example, for a $[2\times 2]$ matrix, the permutations are $\{\pi_m\} = \{(1)(2), (12)\}$, so

$$\det M = M_{11}M_{22} - M_{12}M_{21}\,, \tag{1.35}$$

for a $[3\times 3]$ *matrix*

$$M = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}$$

there are $6 = 3!$ *permutations,*

$$\begin{aligned} \det M &= M_{11}M_{22}M_{33} - M_{11}M_{23}M_{32} - M_{12}M_{21}M_{33} + M_{12}M_{23}M_{31} \\ &\quad + M_{13}M_{21}M_{32} - M_{13}M_{22}M_{31}\,, \end{aligned} \tag{1.36}$$

and so on. Not very illuminating.

But if $M = T - \lambda\mathbf{1}$*, evaluation of the [*$2\times 2$*] case,*

$$\det(T - \lambda\mathbf{1}) = (T_{11} - \lambda)(T_{22} - \lambda) - M_{12}M_{21} = \lambda^2 + (\mathrm{tr}\,T)\lambda + \det T\,, \tag{1.37}$$

used in (1.29), offers a hint of better things to come. This way of computing determinants is generalized to any $[m\times m]$ *matrix in ref. [5], sect. 6.4* Determinants *(click here).*

The $\ln\det M = \mathrm{tr}\,\ln M$ *relation, valid for any square matrix* M *(even the infinite dimensional 'trace class' operators* M*, as long as all* $|\mathrm{tr}\,M^k|$ *are bounded) offers a powerful alternative, universally used, for evaluating determinants.*

First, observe that both the determinant and the trace are invariant under similarity transformations $\hat{M} = S^{-1}M\,S\,,\ \det S \neq 0$*:*

$$\begin{aligned} \det \hat{M} &= \det(S^{-1}M\,S) = (\det S^{-1})\,(\det M)\,(\det S) = \det M \\ \mathrm{tr}\,\hat{M}^k &= \mathrm{tr}\,S^{-1}M\,S\cdots S^{-1}M\,S = \mathrm{tr}\,M\,S\cdots S^{-1}M\,S\,S^{-1} = \mathrm{tr}\,M^k\,, \end{aligned} \tag{1.38}$$

so any quantity, in particular the eigenvalues of M*, expressed in terms of its traces and its determinant is also invariant under all linear coordinate changes.*

Next, consider the characteristic polynomial (1.15) of $[m\times m]$ *matrix* T*, and change the variable to* $z = 1/\lambda$ *in* $\det(T - \lambda\mathbf{1})$*. The zeros* $z_j = 1/\lambda_j$ *of*

$$\det(\mathbf{1} - zT) = 0 \tag{1.39}$$

exercise 1.1

now yield the non-zero eigenvalues λ_j *of* T*. That* $\lambda_j = 0$ *eigenvalues are gone is a blessing; nobody liked them anyway. By the determinant–trace relation* $\ln\det M = \mathrm{tr}\,\ln M$*, the determinant of* $M = 1 - zT$ *is* always *expressible as*

$$\det(1 - zT) = \exp(\mathrm{tr}\,\ln(1 - zT)) = e^{-\sum_{n=1}\frac{z^n}{n}\mathrm{tr}\,T^n}\,. \tag{1.40}$$

We evaluate such formulas in two steps. First, expand $\exp(f(z))$ *as Taylor series in* $f(z)$

$$\det(1 - zT) = \sum_{k=0}^{\infty}\frac{1}{k!}\left(-\sum_{n=1}\frac{z^n}{n}\mathrm{tr}\,T^n\right)^k$$

Then expand $(\cdots)^k$ *as series in* z^n *and combine terms of order* z^n*. The result is central to much statistical physics and field theory, where it is known as the* cumulant expansion*:*

$$\begin{aligned} \det(1 - zT) &= 1 - z\,\mathrm{tr}\,T - \frac{z^2}{2}\left((\mathrm{tr}\,T)^2 - \mathrm{tr}\,T^2\right) \\ &\quad - \frac{z^3}{3!}\left((\mathrm{tr}\,T)^3 - 3\,(\mathrm{tr}\,T)\,\mathrm{tr}\,T^2 + 2\,\mathrm{tr}\,T^3\right) \\ &\quad - \frac{z^4}{4!}\left((\mathrm{tr}\,T)^4 - 3\,(2\,(\mathrm{tr}\,T)^2 - \mathrm{tr}\,T^2)\,\mathrm{tr}\,T^2 + 8\,\mathrm{tr}\,T\,\mathrm{tr}\,T^3 - 6\,\mathrm{tr}\,T^4\right) - \dots \end{aligned} \tag{1.41}$$

If T is an $[m\times m]$ matrix, the characteristic polynomial is at most of order m, so the infinity of coefficients of z^n must vanish exactly *for $n > m$! For example, for a $[2\times 2]$ matrix, the z^2 coefficient in (1.41) is a traces expansion for the determinant (1.35),*

$$\det(T) = \frac{1}{2}\left((\mathrm{tr}\,T)^2 - \mathrm{tr}\,T^2\right) , \tag{1.42}$$

and for a $[3\times 3]$ matrix, the z^3 coefficient in (1.41) is a traces expansion for the determinant (1.36)

$$\begin{aligned}\det(T) &= \frac{1}{3!}\left((\mathrm{tr}\,T)^3 - 3\,(\mathrm{tr}\,T)\,\mathrm{tr}\,T^2 + 2\,\mathrm{tr}\,T^3\right) \\ &= M_{11}M_{22}M_{33} - M_{11}M_{23}M_{32} - M_{12}M_{21}M_{33} + M_{12}M_{23}M_{31} \\ &\quad + M_{13}M_{21}M_{32} - M_{13}M_{22}M_{31} ,\end{aligned} \tag{1.43}$$

as you can verify by hand, if you do not believe me (you should never believe anything anyone over 30 says). If you still do not believe me, verify that the z^4 coefficient vanishes

$$0 = \frac{1}{4!}\left(-6(\mathrm{tr}\,T)^2\mathrm{tr}\,T^2 + 8(\mathrm{tr}\,T)\mathrm{tr}\,T^3 + 3(\mathrm{tr}\,T^2)^2 - 6\mathrm{tr}\,T^4 + (\mathrm{tr}\,T)^4\right)$$

for $m = 1, 2, 3$, but is a traces expansion for the determinant of a $[4\times 4]$ matrix. If you need to know more, these relations were noted by Albert Girard (1629), so they are called Newton's (1666) identities.

Note also that derivative of (1.40) relates the determinant to the resolvent*,*

$$\begin{aligned}-z\frac{d}{dz}\ln\det(1 - zT) &= -\mathrm{tr}\left(z\frac{d}{dz}\ln(1 - zT)\right) \\ &= \mathrm{tr}\,\frac{zT}{1 - zT} = \sum_{k=1}^{\infty} z^n\mathrm{tr}\,(T^n) ,\end{aligned} \tag{1.44}$$

a simple but very useful relation expressing a determinant in terms of traces.

What are all these relationships? Have a fresh look at the Hamilton-Cayley theorem (1.18) that states that the matrix $\mathbf{M}$ satisfies its characteristic equation, and to be specific, look at the $m = 3$ case. The Hamilton-Cayley characteristic equation *expanded in terms of traces is*

$$0 \;=\; T^3 - (\mathrm{tr}\,T)\,T^2 + \frac{1}{2}\left((\mathrm{tr}\,T)^2 - \mathrm{tr}\,T^2\right)T - (\det T)\,\mathbf{1}\,. \tag{1.45}$$

This is the first 3 terms of the cumulant expansion *(1.41), with λ restored by $z \to 1/\lambda$, i.e., the characteristic equation for A $[3\times 3]$ matrix, and the λ replaced by T. The Hamilton-Cayley formula says that whenever you see $[m\times m]$ matrix T^m you can express it in terms of T^{m-1}, T^{m-1}, $\cdots$, T.*

To be very specific and pedestrian, consider the $[3\times 3]$ matrix

$$\begin{aligned}T &= \begin{pmatrix} 2 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 2 \end{pmatrix} , \qquad T^2 = \begin{pmatrix} 12 & 20 & 12 \\ 20 & 44 & 20 \\ 12 & 20 & 12 \end{pmatrix} \\ \mathrm{tr}\,T &= 10 , \qquad \mathrm{tr}\,T^2 = 68 \,.\end{aligned} \tag{1.46}$$

From the shape of T clearly $\det T = 0$, so the characteristic equation is

$$\begin{aligned} 0 &= \left(T^2 - (\mathrm{tr}\,T)\,T + \frac{1}{2}\left((\mathrm{tr}\,T)^2 - \mathrm{tr}\,T^2\right)\mathbf{1}\right)T\,. \\ &= \left(T^2 - 10\,T + \frac{1}{2}\,(100-68)\,\mathbf{1}\right)T \\ &= \left(T^2 - 10\,T + 16\,\mathbf{1}\right)T = (T-8\,\mathbf{1})(T-2\,\mathbf{1})\,T\,, \end{aligned} \tag{1.47}$$

with eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\} = \{8, 2, 0\}$.

For the associated projection operators, see (2.30).

Commentary

Remark 1.1. Projection operators. The construction of projection operators given in sect. 1.4 is taken from refs. [3, 4]. Sylvester [13] wrote down the spectral decomposition (1.24) in 1883 in the form we use, but lineage certainly goes all the way back to 1795 Lagrange polynomials [11], and Euler 1783. Often projection operators get drowned in sea of algebraic details. Halmos [7] is a good early reference - but we like Harter's exposition [8–10] best, for its multitude of specific examples and physical illustrations. In particular, by the time we get to (1.21) we have tacitly assumed full diagonalizability of matrix $\mathbf{M}$. That is the case for the compact groups we will study here (they are all subgroups of $\mathrm{U}(n)$) but not necessarily in other applications. A bit of what happens then (nilpotent blocks) is touched upon in example 1.2. Harter in his lecture Harter's lecture 5 (starts about min. 31 into the lecture) explains this in great detail - its well worth your time.

References

[1] G. B. Arfken, H. J. Weber, and F. E. Harris, *Mathematical Methods for Physicists: A Comprehensive Guide*, 7th ed. (Academic, New York, 2013).

[2] S. Axler, "Down with determinants!", Amer. Math. Monthly **102**, 139–154 (1995).

[3] P. Cvitanović, "Group theory for Feynman diagrams in non-Abelian gauge theories", Phys. Rev. D **14**, 1536–1553 (1976).

[4] P. Cvitanović, Classical and exceptional Lie algebras as invariance algebras, Oxford Univ. preprint 40/77, unpublished., 1977.

[5] P. Cvitanović, *Group Theory: Birdtracks, Lie's and Exceptional Groups* (Princeton Univ. Press, Princeton NJ, 2004).

[6] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. (J. Hopkins Univ. Press, Baltimore, MD, 2013).

[7] P. R. Halmos, *Finite-Dimensional Vector Spaces* (Princeton Univ. Press, Princeton NJ, 1948).

[8] W. G. Harter, "Algebraic theory of ray representations of finite groups", J. Math. Phys. **10**, 739–752 (1969).

[9] W. G. Harter, *Principles of Symmetry, Dynamics, and Spectroscopy* (Wiley, New York, 1993).

[10] W. G. Harter and N. dos Santos, "Double-group theory on the half-shell and the two-level system. I. Rotation and half-integral spin states", Amer. J. Phys. **46**, 251–263 (1978).

[11] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd ed. (Prentice-Hall, Englewood Cliffs NJ, 1971).

[12] M. Stone and P. Goldbart, *Mathematics for Physics: A Guided Tour for Graduate Students* (Cambridge Univ. Press, Cambridge UK, 2009).

[13] J. J. Sylvester, "XXXIX. On the equation to the secular inequalities in the planetary theory", Philos. Mag. **16**, 267–269 (1883).

Exercises

1.1. **Trace-log of a matrix.** Prove that

$$\det M = e^{\mathrm{tr}\,\ln M} .$$

for an arbitrary finite dimensional square matrix M, $\det M \neq 0$. (If you are not getting it, see AWH(3.171).)

1.2. **Stability, diagonal case.** Verify that for a diagonalizable matrix A the exponential is also diagonalizable

$$J^t = e^{tA} = \mathbf{U}^{-1} e^{tA_D} \mathbf{U} , \quad A_D = \mathbf{U} \mathbf{A} \mathbf{U}^{-1} . \tag{1.48}$$

1.3. **The matrix square root.** Consider matrix

$$A = \begin{bmatrix} 4 & 10 \\ 0 & 9 \end{bmatrix} .$$

Generalize the square root function $f(x) = x^{1/2}$ to a square root $f(A) = A^{1/2}$ of a matrix A.

a) Which one(s) of these are the square root of A

$$\begin{bmatrix} 2 & 2 \\ 0 & 3 \end{bmatrix} , \begin{bmatrix} -2 & 10 \\ 0 & 3 \end{bmatrix} , \begin{bmatrix} -2 & -2 \\ 0 & -3 \end{bmatrix} , \begin{bmatrix} 2 & -10 \\ 0 & -3 \end{bmatrix} ?$$

b) Assume that the eigenvalues of a $[d \times d]$ matrix are all distinct. How many square root matrices does such matrix have?

c) Given a $[2{\times}2]$ matrix A with a distinct pair of eigenvalues $\{\lambda_1, \lambda_2\}$, write down a formula that generates all square root matrices $A^{1/2}$. Hint: one can do this using the 2 projection operators associates with the matrix A. 2 points

1.4. **Exponential of a matrix of Jordan form.** A matrix B with all eigenvalues degenerate that cannot be diagonalized can always be brought to upper triangular Jordan form $B = \lambda \mathbf{1} + E$, where E is its strictly upper bidiagonal part. As an example, consider [4×4] matrix B, with

$$E = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} . \qquad (1.49)$$

a) Write down E, E^2, E^2, E^3, $\dots$

b) Write down explicitly the exponential [4×4] matrix function $\exp(tE)$.

c) Bonus points, some assembly required: Work out the kth term in the Taylor expansion of a $[d \times d]$ matrix function $f(B)$, $B = \lambda \mathbf{1} + E$ a $[d \times d]$ matrix,

$$f(B) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (B - x_0 \mathbf{1})^k . \qquad (1.50)$$

.

A side remark to the masters of QM: E is a ‘raising operator’.