1

Deterministic Matrices

Matrices appear in all corners of science, from mathematics to physics, computer science, biology, economics and quantitative finance. In fact, before Schrodinger's equation, quantum mechanics was formulated by Heisenberg in terms of what he called "Matrix Mechanics". In many cases, the matrices that appear are deterministic, and their properties are encapsulated in their eigenvalues and eigenvectors. This first chapter gives several elementary results in linear algebra, in particular concerning eigenvalues. These results will be extremely useful in the rest of the book where we will deal with random matrices, and in particular the statistical properties of their eigenvalues and eigenvectors.

1.1 Matrices, Eigenvalues and Singular Values

1.1.1 Some Problems Where Matrices Appear

Let us give three examples motivating the study of matrices, and the different forms that those can take.

Dynamical System

Consider a generic *dynamical system* describing the time evolution of a certain N-dimensional vector $\mathbf{x}(t)$, for example the three-dimensional position of a point in space. Let us write the equation of motion as

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}(\mathbf{x}), \tag{1.1}$$

where $\mathbf{F}(\mathbf{x})$ is an arbitrary vector field. Equilibrium points $\mathbf{x}^*$ are such that $\mathbf{F}(\mathbf{x}^*) = 0$. Consider now small deviations from equilibrium, i.e. $\mathbf{x} = \mathbf{x}^* + \epsilon\mathbf{y}$ where $\epsilon \ll 1$. To first order in ϵ, the dynamics becomes linear, and given by

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \mathbf{A}\mathbf{y}, \tag{1.2}$$

where $\mathbf{A}$ is a matrix whose elements are given by $\mathbf{A}_{ij} = \partial_j F_i(\mathbf{x}^*)$, where i, j are indices that run from 1 to N. When $\mathbf{F}$ can itself be written as the gradient of some potential V, i.e. $F_i = -\partial_i V(\mathbf{x})$, the matrix $\mathbf{A}$ becomes symmetric, i.e. $\mathbf{A}_{ij} = \mathbf{A}_{ji} = -\partial_{ij} V$. But this is not

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

always the case; in general the linearized dynamics is described by a matrix $\mathbf{A}$ without any particular property – except that it is a square $N \times N$ array of real numbers.

Master Equation

Another standard setting is the so-called Master equation for the evolution of probabilities. Call $i = 1, \ldots, N$ the different possible states of a system and $P_i(t)$ the probability to find the system in state i at time t. When memory effects can be neglected, the dynamics is called Markovian and the evolution of $P_i(t)$ is described by the following discrete time equation:

$$P_i(t+1) = \sum_{j=1}^{N} \mathbf{A}_{ij} P_j(t), \tag{1.3}$$

meaning that the system has a probability $\mathbf{A}_{ij}$ to jump from state j to state i between t and $t+1$. Note that all elements of $\mathbf{A}$ are positive; furthermore, since all jump possibilities must be exhausted, one must have, for each j, $\sum_i \mathbf{A}_{ij} = 1$. This ensures that $\sum_i P_i(t) = 1$ at all times, since

$$\sum_{i=1}^{N} P_i(t+1) = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{A}_{ij} P_j(t) = \sum_{j=1}^{N}\sum_{i=1}^{N} \mathbf{A}_{ij} P_j(t) = \sum_{j=1}^{N} P_j(t) = 1. \tag{1.4}$$

Matrices such that all elements are positive and such that the sum over all rows is equal to unity for each column are called *stochastic matrices*. In matrix form, Eq. (1.3) reads $\mathbf{P}(t+1) = \mathbf{A}\mathbf{P}(t)$, leading to $\mathbf{P}(t) = \mathbf{A}^t\mathbf{P}(0)$, i.e. $\mathbf{A}$ raised to the t-th power applied to the initial distribution.

Covariance Matrices

As a third important example, let us consider random, N-dimensional real vectors $\mathbf{X}$, with some given multivariate distribution $P(\mathbf{X})$. The covariance matrix $\mathbf{C}$ of the $\mathbf{X}$'s is defined as

$$C_{ij} = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j], \tag{1.5}$$

where $\mathbb{E}$ means that we are averaging over the distribution $P(\mathbf{X})$. Clearly, the matrix $\mathbf{C}$ is real and symmetric. It is also positive semi-definite, in the sense that for any vector $\mathbf{x}$,

$$\mathbf{x}^T\mathbf{C}\mathbf{x} \geq 0. \tag{1.6}$$

If it were not the case, it would be possible to find a linear combination of the vectors $\mathbf{X}$ with a negative variance, which is obviously impossible.

The three examples above are all such that the corresponding matrices are $N \times N$ square matrices. Examples where matrices are rectangular also abound. For example, one could consider two sets of random real vectors: $\mathbf{X}$ of dimension N_1 and $\mathbf{Y}$ of dimension N_2. The *cross-covariance* matrix defined as

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

$$C_{ia} = \mathbb{E}[X_i Y_a] - \mathbb{E}[X_i]\mathbb{E}[Y_a]; \qquad i = 1, \ldots, N_1; \qquad a = 1, \ldots, N_2, \tag{1.7}$$

is an $N_1 \times N_2$ matrix that describes the correlations between the two sets of vectors.

1.1.2 Eigenvalues and Eigenvectors

One learns a great deal about matrices by studying their eigenvalues and eigenvectors. For a square matrix $\mathbf{A}$ a pair of scalar and non-zero vector $(\lambda, \mathbf{v})$ satisfying

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \tag{1.8}$$

is called an eigenvalue–eigenvector pair.

Trivially if $\mathbf{v}$ is an eigenvector $\alpha\mathbf{v}$ is also an eigenvector when α is a non-zero real number. Sometimes multiple non-collinear eigenvectors share the same eigenvalue; we say that this eigenvalue is degenerate and has multiplicity equal to the dimension of the vector space spanned by its eigenvectors.

If Eq. (1.8) is true, it implies that the equation $(\mathbf{A} - \lambda\mathbf{1})\mathbf{v} = 0$ has non-trivial solutions, which requires that $\det(\lambda\mathbf{1} - \mathbf{A}) = 0$. The eigenvalues λ are thus the roots of the so-called characteristic polynomial of the matrix $\mathbf{A}$, obtained by expanding $\det(\lambda\mathbf{1} - \mathbf{A})$. Clearly, this polynomial[1] is of order N and therefore has at most N different roots, which correspond to the (possibly complex) eigenvalues of $\mathbf{A}$. Note that the characteristic polynomial of $\mathbf{A}^T$ coincides with the characteristic polynomial of $\mathbf{A}$, so the eigenvalues of $\mathbf{A}$ and $\mathbf{A}^T$ are identical.

Now, let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the N eigenvalues of $\mathbf{A}$ with $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ the corresponding eigenvectors. We define Λ as the $N \times N$ diagonal matrix with λ_i on the diagonal, and $\mathbf{V}$ as the $N \times N$ matrix whose jth column is $\mathbf{v}_j$, i.e. $V_{ij} = (\mathbf{v}_j)_i$ is the ith component of $\mathbf{v}_j$. Then, by definition,

$$\mathbf{A}\mathbf{V} = \mathbf{V}\Lambda, \tag{1.9}$$

since once expanded, this reads

$$\sum_k \mathbf{A}_{ik} V_{kj} = V_{ij}\lambda_j, \tag{1.10}$$

or $\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}_j$. If the eigenvectors are linearly independent (which is not true for all matrices), the matrix inverse $\mathbf{V}^{-1}$ exists and one can therefore write $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}, \tag{1.11}$$

which is called the eigenvalue decomposition of the matrix $\mathbf{A}$.

Symmetric matrices (such that $\mathbf{A} = \mathbf{A}^T$) have very nice properties regarding their eigenvalues and eigenvectors.

[1] The characteristic polynomial $Q_N(\lambda) = \det(\lambda\mathbf{1} - \mathbf{A})$ always has a coefficient 1 in front of its highest power ($Q_N(\lambda) = \lambda^N + O(\lambda^{N-1})$), such polynomials are called *monic*.

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

- They have exactly N eigenvalues when counted with their multiplicity.
- All their eigenvalues and eigenvectors are real.
- Their eigenvectors are orthogonal and can be chosen to be orthonormal (i.e. $\mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$). Here we assume that for degenerate eigenvalues we pick an orthogonal set of corresponding eigenvectors.

If we choose orthonormal eigenvectors, the matrix $\mathbf{V}$ has the property $\mathbf{V}^T\mathbf{V} = \mathbf{1}$ ($\Rightarrow \mathbf{V}^T = \mathbf{V}^{-1}$). Hence it is an *orthogonal matrix* $\mathbf{V} = \mathbf{O}$ and Eq. (1.11) reads

$$\mathbf{A} = \mathbf{O}\Lambda\mathbf{O}^T, \tag{1.12}$$

where Λ is a diagonal matrix containing the eigenvalues associated with the eigenvectors in the columns of $\mathbf{O}$. A symmetric matrix can be diagonalized by an orthogonal matrix. Remark that an $N \times N$ orthogonal matrix is fully parameterized by $N(N-1)/2$ "angles", whereas Λ contains N diagonal elements. So the total number of parameters of the diagonal decomposition is $N(N-1)/2 + N$, which is identical, as it should be, to the number of different elements of a symmetric $N \times N$ matrix.

Let us come back to our dynamical system example, Eq. (1.2). One basic question is to know whether the perturbation $\mathbf{y}$ will grow with time, or decay with time. The answer to this question is readily given by the eigenvalues of $\mathbf{A}$. For simplicity, we assume $\mathbf{F}$ to be a gradient such that $\mathbf{A}$ is symmetric. Since the eigenvectors of $\mathbf{A}$ are orthonormal, one can decompose $\mathbf{y}$ in term of the $\mathbf{v}$'s as

$$\mathbf{y}(t) = \sum_{i=1}^{N} c_i(t)\mathbf{v}_i. \tag{1.13}$$

Taking the dot product of Eq. (1.2) with $\mathbf{v}_i$ then shows that the dynamics of the coefficients $c_i(t)$ are decoupled and given by

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = \lambda_i c_i, \tag{1.14}$$

where λ_i is the eigenvalue associated with $\mathbf{v}_i$. Therefore, any component of the initial perturbation $\mathbf{y}(t=0)$ that is along an eigenvector with positive eigenvalue will grow exponentially with time, until the linearized approximation leading to Eq. (1.2) breaks down. Conversely, components along directions with negative eigenvalues decrease exponentially with time. An equilibrium $\mathbf{x}^*$ is called stable provided all eigenvalues are negative, and marginally stable if some eigenvalues are zero while all others are negative.

The important message carried by the example above is that diagonalizing a matrix amounts to finding a way to decouple the different degrees of freedom, and convert a matrix equation into a set of N scalar equations, as Eqs. (1.14). We will see later that the same idea holds for covariance matrices as well: their diagonalization allows one to find a set of uncorrelated vectors. This is usually called Principal Component Analysis (PCA).

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

Exercise 1.1.1 Instability of eigenvalues of non-symmetric matrices

Consider the $N \times N$ square band diagonal matrix $\mathbf{M}_0$ defined by $[\mathbf{M}_0]_{ij} = 2\delta_{i,j-1}$:

$$\mathbf{M}_0 = \begin{pmatrix} 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 2 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{1.15}$$

(a) Show that $\mathbf{M}_0^N = 0$ and so all the eigenvalues of $\mathbf{M}_0$ must be zero. Use a numerical eigenvalue solver for non-symmetric matrices and confirm numerically that this is the case for $N = 100$.

(b) If $\mathbf{O}$ is an orthogonal matrix ($\mathbf{O}\mathbf{O}^T = \mathbf{1}$), $\mathbf{O}\mathbf{M}_0\mathbf{O}^T$ has the same eigenvalues as $\mathbf{M}_0$. Following Exercise 1.2.4, generate a random orthogonal matrix $\mathbf{O}$. Numerically find the eigenvalues of $\mathbf{O}\mathbf{M}_0\mathbf{O}^T$. Do you get the same answer as in (a)?

(c) Consider $\mathbf{M}_1$ whose elements are all equal to those of $\mathbf{M}_0$ except for one element in the lower left corner $[\mathbf{M}_1]_{N,1} = (1/2)^{N-1}$. Show that $\mathbf{M}_1^N = \mathbf{1}$; more precisely, show that the characteristic polynomial of $\mathbf{M}_1$ is given by $\det(\mathbf{M}_1 - \lambda\mathbf{1}) = \lambda^N - 1$, therefore $\mathbf{M}_1$ has N distinct eigenvalues equal to the N complex roots of unity $\lambda_k = \mathrm{e}^{2\pi \mathrm{i} k/N}$.

(d) For N greater than about 60, $\mathbf{O}\mathbf{M}_0\mathbf{O}^T$ and $\mathbf{O}\mathbf{M}_1\mathbf{O}^T$ are indistinguishable to machine precision. Compare numerically the eigenvalues of these two rotated matrices.

1.1.3 Singular Values

A non-symmetric, square matrix cannot in general be decomposed as $\mathbf{A} = \mathbf{O}\Lambda\mathbf{O}^T$, where Λ is a diagonal matrix and $\mathbf{O}$ an orthogonal matrix. One can however find a very useful alternative decomposition as

$$\mathbf{A} = \mathbf{V}\mathbf{S}\mathbf{U}^T, \tag{1.16}$$

where $\mathbf{S}$ is a non-negative diagonal matrix, whose elements are called the *singular values* of $\mathbf{A}$, and $\mathbf{U}, \mathbf{V}$ are two real, orthogonal matrices. Whenever $\mathbf{A}$ is symmetric positive semi-definite, one has $\mathbf{S} = \Lambda$ and $\mathbf{U} = \mathbf{V}$.

Equation (1.16) also holds for rectangular $N \times T$ matrices, where $\mathbf{V}$ is $N \times N$ orthogonal, $\mathbf{U}$ is $T \times T$ orthogonal and $\mathbf{S}$ is $N \times T$ diagonal as defined below. To construct the *singular value decomposition* (SVD) of $\mathbf{A}$, we first introduce two matrices $\mathbf{B}$ and $\widehat{\mathbf{B}}$, defined as $\mathbf{B} := \mathbf{A}\mathbf{A}^T$ and $\widehat{\mathbf{B}} = \mathbf{A}^T\mathbf{A}$. It is plain to see that these matrices are symmetric, since

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

$\mathbf{B}^T = (\mathbf{A}\mathbf{A}^T)^T = \mathbf{A}^{TT}\mathbf{A}^T = \mathbf{B}$ (and similarly for $\widehat{\mathbf{B}}$). They are also positive semi-definite as for any vector $\mathbf{x}$ we have $\mathbf{x}^T\mathbf{B}\mathbf{x} = ||\mathbf{A}^T\mathbf{x}||^2 \geq 0$.

We can show that $\mathbf{B}$ and $\widehat{\mathbf{B}}$ have the same non-zero eigenvalues. In fact, let $\lambda > 0$ be an eigenvalue of $\mathbf{B}$ and $\mathbf{v} \neq 0$ is the corresponding eigenvector. Then we have, by definition,

$$\mathbf{A}\mathbf{A}^T\mathbf{v} = \lambda\mathbf{v}. \tag{1.17}$$

Let $\mathbf{u} = \mathbf{A}^T\mathbf{v}$, then we can get from the above equation that

$$\mathbf{A}^T\mathbf{A}\mathbf{A}^T\mathbf{v} = \lambda\mathbf{A}^T\mathbf{v} \Rightarrow \widehat{\mathbf{B}}\mathbf{u} = \lambda\mathbf{u}. \tag{1.18}$$

Moreover,

$$\|\mathbf{u}\|^2 = \mathbf{v}^T\mathbf{A}\mathbf{A}^T\mathbf{v} = \mathbf{v}^T\mathbf{B}\mathbf{v} \neq 0 \Rightarrow \mathbf{u} \neq 0. \tag{1.19}$$

Hence λ is also an eigenvalue of $\widehat{\mathbf{B}}$. Note that for degenerate eigenvalues λ of $\mathbf{B}$, an orthogonal set of corresponding eigenvectors $\{\mathbf{v}_\ell\}$ gives rise to an orthogonal set $\{\mathbf{A}^T\mathbf{v}_\ell\}$ of eigenvectors of $\widehat{\mathbf{B}}$. Hence the multiplicity of λ in $\widehat{\mathbf{B}}$ is at least that of $\mathbf{B}$. Similarly, we can show that any non-zero eigenvalue of $\widehat{\mathbf{B}}$ is also an eigenvalue of $\mathbf{B}$. This finishes the proof of the claim.

Note that $\mathbf{B}$ has at most N non-zero eigenvalues and $\widehat{\mathbf{B}}$ has at most T non-zero eigenvalues. Thus by the above claim, if $T > N$, $\widehat{\mathbf{B}}$ has at least $T - N$ zero eigenvalues, and if $T < N$, $\mathbf{B}$ has at least $N - T$ zero eigenvalues. We denote the other $\min\{N, T\}$ eigenvalues of $\mathbf{B}$ and $\widehat{\mathbf{B}}$ by $\{\lambda_k\}_{1\leq k\leq \min\{N,T\}}$. Then the SVD of $\mathbf{A}$ is expressed as Eq. (1.16), where $\mathbf{V}$ is the $N \times N$ orthogonal matrix consisting of the N normalized eigenvectors of $\mathbf{B}$, $\mathbf{U}$ is the $T \times T$ orthogonal matrix consisting of the T normalized eigenvectors of $\widehat{\mathbf{B}}$, and $\mathbf{S}$ is an $N \times T$ rectangular diagonal matrix with $S_{kk} = \sqrt{\lambda_k} \geq 0, 1 \leq k \leq \min\{N, T\}$ and all other entries equal to zero.

For instance, if $N < T$, we have

$$\mathbf{S} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 & \cdots & 0 \\ & & & & & \\ 0 & 0 & \ddots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sqrt{\lambda_N} & \cdots & 0 \end{pmatrix}. \tag{1.20}$$

Although (non-degenerate) normalized eigenvectors are unique up to a sign, the choice of the positive sign for the square-root $\sqrt{\lambda_k}$ imposes a condition on the combined sign for the left and right singular vectors $\mathbf{v}_k$ and $\mathbf{u}_k$. In other words, simultaneously changing both $\mathbf{v}_k$ and $\mathbf{u}_k$ to $-\mathbf{v}_k$ and $-\mathbf{u}_k$ leaves the matrix $\mathbf{A}$ invariant, but for non-zero singular values one cannot individually change the sign of either $\mathbf{v}_k$ or $\mathbf{u}_k$.

The recipe to find the SVD, Eq. (1.16), is thus to diagonalize both $\mathbf{A}\mathbf{A}^T$ (to obtain $\mathbf{V}$ and $\mathbf{S}^2$) and $\mathbf{A}^T\mathbf{A}$ (to obtain $\mathbf{U}$ and again $\mathbf{S}^2$). It is insightful to again count the number of parameters involved in this decomposition. Consider a general $N \times T$ matrix with $T \geq N$ (the case $N \geq T$ follows similarly). The N eigenvectors of $\mathbf{A}\mathbf{A}^T$ are generically unique up to a sign, while for $T - N > 0$ the matrix $\mathbf{A}^T\mathbf{A}$ will have a degenerate eigenspace associated with the eigenvalue 0 of size $T - N$, hence its eigenvectors are only unique up

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

to an arbitrary rotation in $T-N$ dimension. So generically the SVD decomposition amounts to writing the NT elements of $\mathbf{A}$ as

$$NT \equiv \frac{1}{2}N(N-1) + N + \frac{1}{2}T(T-1) - \frac{1}{2}(T-N)(T-N-1). \tag{1.21}$$

The interpretation of Eq. (1.16) for $N \times N$ matrices is that one can always find an orthonormal basis of vectors $\{\mathbf{u}\}$ such that the application of a matrix A amounts to a rotation (or an improper rotation) of $\{\mathbf{u}\}$ into another orthonormal set $\{\mathbf{v}\}$, followed by a dilation of each $\mathbf{v}_k$ by a positive factor $\sqrt{\lambda_k}$.

Normal matrices are such that $\mathbf{U} = \mathbf{V}$. In other words, $\mathbf{A}$ is normal whenever $\mathbf{A}$ commutes with its transpose: $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}$. Symmetric, skew-symmetric and orthogonal matrices are normal, but other cases are possible. For example a 3×3 matrix such that each row and each column has exactly two elements equal to 1 and one element equal to 0 is normal.

1.2 Some Useful Theorems and Identities

In this section, we state without proof very useful theorems on eigenvalues and matrices.

1.2.1 Gershgorin Circle Theorem

Let $\mathbf{A}$ be a real matrix, with elements $\mathbf{A}_{ij}$. Define R_i as $R_i = \sum_{j\neq i} |\mathbf{A}_{ij}|$, and $\mathcal{D}_i$ a disk in the complex plane centered on $\mathbf{A}_{ii}$ and of radius R_i. Then every eigenvalue of $\mathbf{A}$ lies within at least one disk $\mathcal{D}_i$. For example, for the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -0.2 & 0.2 \\ -0.3 & 2 & -0.2 \\ 0 & 1.1 & 3 \end{pmatrix}, \tag{1.22}$$

the three circles are located on the real axis at $x = 1, 2$ and 3 with radii 0.4, 0.5 and 1.1 respectively (see Fig. 1.1).

In particular, eigenvalues corresponding to eigenvectors with a maximum amplitude on i lie within the disk $\mathcal{D}_i$.

1.2.2 The Perron–Frobenius Theorem

Let $\mathbf{A}$ be a real matrix, with all its elements positive $\mathbf{A}_{ij} > 0$. Then the top eigenvalue $\lambda_{\max}$ is unique and real (all other eigenvalues have a smaller real part). The corresponding top eigenvector $\mathbf{v}^*$ has all its elements positive:

$$\mathbf{A}\mathbf{v}^* = \lambda_{\max}\mathbf{v}^*; \qquad \mathbf{v}^*_k > 0, \forall k. \tag{1.23}$$

The top eigenvalue satisfies the following inequalities:

$$\min_i \sum_j \mathbf{A}_{ij} \leq \lambda_{\max} \leq \max_i \sum_j \mathbf{A}_{ij}. \tag{1.24}$$

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

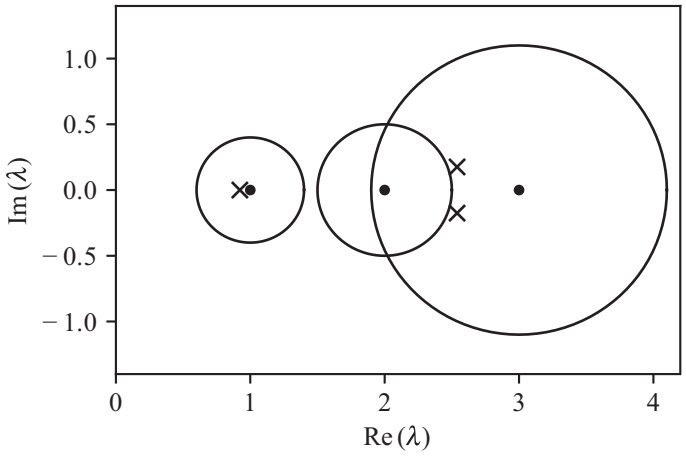

Figure 1.1 The three complex eigenvalues of the matrix (1.22) (crosses) and its three Gershgorin circles. The first eigenvalue $\lambda_1 \approx 0.92$ falls in the first circle while the other two $\lambda_{2,3} \approx 2.54 \pm 0.18\mathrm{i}$ fall in the third one.

Application: Suppose $\mathbf{A}$ is a stochastic matrix, such that all its elements are positive and satisfy $\sum_i \mathbf{A}_{ij} = 1$, $\forall j$. Then clearly the vector $\vec{1}$ is an eigenvector of $\mathbf{A}^T$, with eigenvalue $\lambda = 1$. But since the Perron–Frobenius can be applied to $\mathbf{A}^T$, the inequalities (1.24) ensure that λ *is* the top eigenvalue of $\mathbf{A}^T$, and thus also of $\mathbf{A}$. All the elements of the corresponding eigenvector $\mathbf{v}^*$ are positive, and describe the stationary state of the associated Master equation, i.e.

$$P_i^* = \sum_j \mathbf{A}_{ij} P_j^* \longrightarrow P_i^* = \frac{\mathbf{v}_i^*}{\sum_k \mathbf{v}_k^*}. \tag{1.25}$$

Exercise 1.2.1 Gershgorin and Perron–Frobenius
Show that the upper bound in Eq. (1.24) is a simple consequence of the Gershgorin theorem.

1.2.3 The Eigenvalue Interlacing Theorem

Let $\mathbf{A}$ be an $N \times N$ symmetric matrix (or more generally Hermitian matrix) with eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$. Consider the $N-1 \times N-1$ submatrix $\mathbf{A}_{\setminus i}$ obtained by removing the ith row and ith columns of $\mathbf{A}$. Its eigenvalues are $\mu_1^{(i)} \geq \mu_2^{(i)} \cdots \geq \mu_{N-1}^{(i)}$. Then the following *interlacing* inequalities hold:

$$\lambda_1 \geq \mu_1^{(i)} \geq \lambda_2 \cdots \geq \mu_{N-1}^{(i)} \geq \lambda_N. \tag{1.26}$$

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

Very recently, a formula relating eigenvectors to eigenvalues was (re-)discovered. Calling $\mathbf{v}_i$ the eigenvector of $\mathbf{A}$ associated with λ_i, one has[2]

$$\left|(\mathbf{v}_i)_j\right|^2 = \frac{\prod_{k=1}^{N-1} \lambda_i - \mu_k^{(j)}}{\prod_{\ell=1,\,\ell\neq i}^{N} \lambda_i - \lambda_\ell}. \tag{1.27}$$

1.2.4 Sherman–Morrison Formula

The Sherman–Morrison formula gives the inverse of a matrix $\mathbf{A}$ perturbed by a rank-1 perturbation:

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}, \tag{1.28}$$

valid for any invertible matrix $\mathbf{A}$ and vectors $\mathbf{u}$ and $\mathbf{v}$ such that the denominator does not vanish. This is a special case of the Woodbury identity, which reads

$$\left(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V}^T\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}^T\mathbf{A}^{-1}, \tag{1.29}$$

where $\mathbf{U}, \mathbf{V}$ are $N \times K$ matrices and $\mathbf{C}$ is a $K \times K$ matrix. Equation (1.28) corresponds to the case $K = 1$.

The associated Sherman–Morrison determinant lemma reads

$$\det(\mathbf{A} + \mathbf{v}\mathbf{u}^T) = \det\mathbf{A} \cdot \left(1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{v}\right) \tag{1.30}$$

for invertible $\mathbf{A}$.

Exercise 1.2.2 Sherman–Morrison

Show that Eq. (1.28) is correct by multiplying both sides by $(\mathbf{A} + \mathbf{u}\mathbf{v}^T)$.

1.2.5 Schur Complement Formula

The Schur complement, also called inversion by partitioning, relates the blocks of the inverse of a matrix to the inverse of blocks of the original matrix. Let $\mathbf{M}$ be an invertible matrix which we divide in four blocks as

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix} \text{ and } \mathbf{M}^{-1} = \mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}, \tag{1.31}$$

where $[\mathbf{M}_{11}] = n \times n$, $[\mathbf{M}_{12}] = n \times (N-n)$, $[\mathbf{M}_{21}] = (N-n) \times n$, $[\mathbf{M}_{22}] = (N-n) \times (N-n)$, and $\mathbf{M}_{22}$ is invertible. The integer n can take any values from 1 to $N-1$.

[2] See: P. Denton, S. Parke, T. Tao, X. Zhang, *Eigenvalues from Eigenvectors: a survey of a basic identity in linear algebra, arXiv:1908.03795.*

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

Then the upper left $n \times n$ block of $\mathbf{Q}$ is given by

$$\mathbf{Q}_{11}^{-1} = \mathbf{M}_{11} - \mathbf{M}_{12}(\mathbf{M}_{22})^{-1}\mathbf{M}_{21}, \tag{1.32}$$

where the right hand side is called the Schur complement of the block $\mathbf{M}_{22}$ of the matrix $\mathbf{M}$.

Exercise 1.2.3 Combining Schur and Sherman–Morrison

In the notation of Eq. (1.31) for $n = 1$ and any $N > 1$, combine the Schur complement of the lower right block with the Sherman–Morrison formula to show that

$$\mathbf{Q}_{22} = (\mathbf{M}_{22})^{-1} + \frac{(\mathbf{M}_{22})^{-1}\mathbf{M}_{21}\mathbf{M}_{12}(\mathbf{M}_{22})^{-1}}{\mathbf{M}_{11} - \mathbf{M}_{12}(\mathbf{M}_{22})^{-1}\mathbf{M}_{21}}. \tag{1.33}$$

1.2.6 Function of a Matrix and Matrix Derivative

In our study of random matrices, we will need to extend real or complex scalar functions to take a symmetric matrix $\mathbf{M}$ as its argument. The simplest way to extend such a function is to apply it to each eigenvalue of the matrix $\mathbf{M} = \mathbf{O}\Lambda\mathbf{O}^T$:

$$F(\mathbf{M}) = \mathbf{O}F(\Lambda)\mathbf{O}^T, \tag{1.34}$$

where $F(\Lambda)$ is the diagonal matrix where we have applied the function F to each (diagonal) entry of Λ. The function $F(\mathbf{M})$ is now a matrix valued function of a matrix. Scalar polynomial functions can obviously be extended directly as

$$F(x) = \sum_{k=0}^{K} a_k x^k \Rightarrow F(\mathbf{M}) = \sum_{k=0}^{K} a_k \mathbf{M}^k, \tag{1.35}$$

but this is equivalent to applying the polynomial to the eigenvalues of $\mathbf{M}$. By extension, when the Taylor series of the function $F(x)$ converges for every eigenvalue of $\mathbf{M}$ the matrix Taylor series coincides with our definition.

Taking the trace of $F(\mathbf{M})$ will yield a matrix function that returns a scalar. This construction is rotationally invariant in the following sense:

$$\operatorname{Tr} F(\mathbf{U}\mathbf{M}\mathbf{U}^T) = \operatorname{Tr} F(\mathbf{M}) \text{ for any } \mathbf{U}\mathbf{U}^T = \mathbf{1}. \tag{1.36}$$

We can take the derivative of a scalar-valued function $\operatorname{Tr} F(\mathbf{M})$ with respect to each element of the matrix $\mathbf{M}$:

$$\frac{\mathrm{d}}{\mathrm{d}[\mathbf{M}]_{ij}} \operatorname{Tr}(F(\mathbf{M})) = [F'(\mathbf{M})]_{ij} \Rightarrow \frac{\mathrm{d}}{\mathrm{d}\mathbf{M}} \operatorname{Tr}(F(\mathbf{M})) = F'(\mathbf{M}). \tag{1.37}$$

Equation (1.37) is easy to derive when $F(x)$ is a monomial $a_k x^k$ and by linearity for polynomial or Taylor series $F(x)$.

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

1.2.7 Jacobian of Simple Matrix Transformations

Suppose one transforms an $N \times N$ matrix $\mathbf{A}$ into another $N \times N$ matrix $\mathbf{B}$ through some function of the matrix elements. The Jacobian of the transformation is defined as the determinant of the partial derivatives:

$$\mathbb{G}_{ij,k\ell} = \frac{\partial \mathbf{B}_{k\ell}}{\partial \mathbf{A}_{ij}}. \tag{1.38}$$

The simplest case is just multiplication by a scalar: $\mathbf{B} = \alpha\mathbf{A}$, leading to $\mathbb{G}_{ij,k\ell} = \alpha\delta_{ik}\delta_{j\ell}$. $\mathbb{G}$ is therefore the tensor product of $\alpha\mathbf{1}$ with $\mathbf{1}$, and its determinant is thus equal to α^N. Not much more difficult is the case of an orthogonal transformation $\mathbf{B} = \mathbf{OAO}^T$, for which $\mathbb{G}_{ij,k\ell} = \mathbf{O}_{ik}\mathbf{O}_{j\ell}$. $\mathbb{G}$ is now the tensor product $\mathbb{G} = \mathbf{O} \otimes \mathbf{O}$ and therefore its determinant is unity.

Slightly more complicated is the case where $\mathbf{B} = \mathbf{A}^{-1}$. Using simple algebra, one readily obtains, for symmetric matrices,

$$\mathbb{G}_{ij,k\ell} = \frac{1}{2}[\mathbf{A}^{-1}]_{ik}[\mathbf{A}^{-1}]_{j\ell} + \frac{1}{2}[\mathbf{A}^{-1}]_{i\ell}[\mathbf{A}^{-1}]_{jk}. \tag{1.39}$$

Let us now assume that $\mathbf{A}$ has eigenvalues λ_α and eigenvectors $\mathbf{v}_\alpha$. One can easily diagonalize $\mathbb{G}_{ij,k\ell}$ within the symmetric sector, since

$$\sum_{k\ell} \mathbb{G}_{ij,k\ell}\left[\mathbf{v}_{\alpha,k}\mathbf{v}_{\beta,\ell} + \mathbf{v}_{\alpha,\ell}\mathbf{v}_{\beta,k}\right] = \frac{1}{\lambda_\alpha\lambda_\beta}\left[\mathbf{v}_{\alpha,i}\mathbf{v}_{\beta,j} + \mathbf{v}_{\alpha,j}\mathbf{v}_{\beta,i}\right]. \tag{1.40}$$

So the determinant of $\mathbb{G}$ is simply $\prod_{\alpha,\beta\geq\alpha}(\lambda_\alpha\lambda_\beta)^{-1}$. Taking the logarithm of this product helps avoiding counting mistakes, and finally leads to the result

$$\det \mathbb{G} = (\det \mathbf{A})^{-N-1}. \tag{1.41}$$

Exercise 1.2.4 Random Matrices

We conclude this chapter on deterministic matrices with a numerical exercise on random matrices. Most of the results of this exercise will be explored theoretically in the following chapters.

- Let $\mathbf{M}$ be a random real symmetric orthogonal matrix, that is an $N \times N$ matrix satisfying $\mathbf{M} = \mathbf{M}^T = \mathbf{M}^{-1}$. Show that all the eigenvalues of $\mathbf{M}$ are ± 1.
- Let $\mathbf{X}$ be a Wigner matrix, i.e. an $N \times N$ real symmetric matrix whose diagonal and upper triangular entries are IID Gaussian random numbers with zero mean and variance σ^2/N. You can use $\mathbf{X} = \sigma(\mathbf{H} + \mathbf{H}^T)/\sqrt{2N}$ where $\mathbf{H}$ is a non-symmetric $N \times N$ matrix with IID standard Gaussians.
- The matrix $\mathbf{P}_+$ is defined as $\mathbf{P}_+ = \frac{1}{2}(\mathbf{M} + \mathbf{1}_N)$. Convince yourself that $\mathbf{P}_+$ is the projector onto the eigenspace of $\mathbf{M}$ with eigenvalue $+1$. Explain the effect of the matrix $\mathbf{P}_+$ on eigenvectors of $\mathbf{M}$.

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press

- An easy way to generate a random matrix $\mathbf{M}$ is to generate a Wigner matrix (independent of $\mathbf{X}$), diagonalize it, replace every eigenvalue by its sign and reconstruct the matrix. The procedure does not depend on the σ used for the Wigner.
- We consider a matrix $\mathbf{E}$ of the form $\mathbf{E} = \mathbf{M}+\mathbf{X}$. To wit, $\mathbf{E}$ is a noisy version of $\mathbf{M}$. The goal of the following is to understand numerically how the matrix $\mathbf{E}$ is corrupted by the Wigner noise. Using the computer language of your choice, for a large value of N (as large as possible while keeping computing times below one minute), for three interesting values of σ of your choice, do the following numerical analysis.

(a) Plot a histogram of the eigenvalues of $\mathbf{E}$, for a single sample first, and then for many samples (say 100).

(b) From your numerical analysis, in the large N limit, for what values of σ do you expect a non-zero density of eigenvalues near zero.

(c) For every normalized eigenvector $\mathbf{v}_i$ of $\mathbf{E}$, compute the norm of the vector $\mathbf{P}_+\mathbf{v}_i$. For a single sample, do a scatter plot of $|\mathbf{P}_+\mathbf{v}_i|^2$ vs λ_i (its eigenvalue). Turn your scatter plot into an approximate conditional expectation value (using a histogram) including data from many samples.

(d) Build an estimator $\Xi(\mathbf{E})$ of $\mathbf{M}$ using only data from $\mathbf{E}$. We want to minimize the error $\mathcal{E} = \frac{1}{N}||(\Xi(\mathbf{E}) - \mathbf{M})||_{\mathrm{F}}^2$ where $||A||_{\mathrm{F}}^2 = \mathrm{Tr}\mathbf{A}\mathbf{A}^T$. Consider first $\Xi_1(\mathbf{E}) = \mathbf{E}$ and then $\Xi_0(\mathbf{E}) = 0$. What is the error $\mathcal{E}$ of these two estimators? Try to build an ad-hoc estimator $\Xi(\mathbf{E})$ that has a lower error $\mathcal{E}$ than these two.

(e) Show numerically that the eigenvalues of $\mathbf{E}$ are not IID. For each sample $\mathbf{E}$ rank its eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_N$. Consider the eigenvalue spacing $s_k = \lambda_k - \lambda_{k-1}$ for eigenvalues in the bulk ($.2N < k < .3N$ and $.7N < k < .8N$). Make a histogram of $\{s_k\}$ including data from 100 samples. Make 100 pseudo-IID samples: mix eigenvalues for 100 different samples and randomly choose N from the $100N$ possibilities, do not choose the same eigenvalue twice for a given pseudo-IID sample. For each pseudo-IID sample, compute s_k in the bulk and make a histogram of the values using data from all 100 pseudo-IID samples. (Bonus) Try to fit an exponential distribution to these two histograms. The IID case should be well fitted by the exponential but not the original data (not IID).

https://doi.org/10.1017/9781108768900.003 Published online by Cambridge University Press