Appendix A

Complements on Matrix Algebra

The set of $n \times m$ matrices with entries in the field $\mathbb{F}$ is denoted as $\mathcal{M}_{n\times m}(\mathbb{F})$. A square matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is called **non-singular** if A^{-1} exists, or equivalently, if $\det(A) \neq 0$. The **identity matrix** of $\mathcal{M}_{n\times n}$ will be denoted as $\mathbf{I}$. Its dimension will normally be clear from the context.

The **transposition** operation consists in inverting rows and columns of a matrix. It is denoted with the superscript "^{T}". If $A \in \mathcal{M}_{n\times m}$, then $A^T \in \mathcal{M}_{m\times n}$.

We shall use two sort of "vectors": the standard **column vectors** in $\mathcal{M}_{n\times 1}$ and the **row vectors** in $\mathcal{M}_{1\times n}$. The transpose of a row vector is a column vector, and vice versa. The default for a vector is to be column, but the good practice is to specify each time which sort of vector we are considering.

A row vector which is formed of positive numbers which add up to 1 is called a **probability vector**. In algebraic notation, the equation $\boldsymbol{\pi}\mathbf{1} = 1$ expresses the fact that the sum of the components of vector $\boldsymbol{\pi}$ is 1.

A.1 Eigenvalues, spectrum and Jordan decomposition

Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a square matrix with complex entries. The complex number λ is an **eigenvalue** of A if there exists a vector $x \neq 0$ such that $Ax = \lambda x$. Such a vector is called a right-**eigenvector** associated to eigenvalue λ. For each eigenvalue, there are also left-eigenvectors.

The set of all eigenvalues is finite and called the **spectrum** of the matrix. We shall denote it as $\mathrm{sp}(A)$. The **spectral radius** of A is defined as:

$$\rho(A) \;=\; \max\{|\lambda| \mid \lambda \in \mathrm{sp}(A)\ \} \ .$$

For several results, it is useful do define the value:

$$\rho_2(A) \;=\; \max\{|\lambda| \mid \lambda \in \mathrm{sp}(A) \text{ and } |\lambda| \neq \rho(A)\ \} \ .$$

With some abuse of terminology, this is called the **second largest eigenvalue** (although $\rho_2(A)$ is not an eigenvalue of A in general).

The eigenvalues of A are the roots of the **characteristic polynomial**:

$$\chi_A(x) \;=\; \det(A - xI) \;=\; \prod_{i=1}^{p} (\lambda_i - x)^{\alpha_i} \ . \tag{A.1}$$

Accordingly, the different eigenvalues are denoted as: $\lambda_1, \ldots, \lambda_p$, with algebraic multiplicities $\alpha_1, \ldots, \alpha_p$. Since the characteristic polynomial if of degree n, we have: $\sum \alpha_i = n$.

If $\alpha_i = 1$, the eigenvalue λ_i is called *simple*.

Every square matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ admits a **Jordan decomposition**: there exists a non-singular matrix S such that:

$$A = S\,J\,S^{-1} \tag{A.2}$$

where

$$J = \begin{pmatrix} J_{\nu_1}(\mu_1) & & & \\ & J_{\nu_2}(\mu_2) & & \\ & & \ddots & \\ & & & J_{\nu_k}(\mu_k) \end{pmatrix}.$$

Here, the square matrices $J_m(\mu)$, are called **Jordan blocks** and are of the form:

$$J_m(\mu) = \begin{pmatrix} \mu & 1 & & \\ & \mu & \ddots & \\ & & \ddots & 1 \\ & & & \mu \end{pmatrix}$$

and the μ_m are eigenvalues of A (they are one of the λ_i). The Jordan matrix J in the decomposition is unique up to permutation of the Jordan blocks. The transformation matrix S is not unique.

There may be several Jordan blocks for one particular eigenvalue λ_i: this is the case for the identity matrix for instance.

When all Jordan blocks are of size $m = 1$, the matrix is called **diagonalizable**. In that case, the i-th row of matrix $\mathbf{S}$ is a left-eigenvector for the i-th eigenvalue, whereas the i-th column of matrix $\mathbf{S}^{-1}$ is a right-eigenvector for that eigenvalue.

A.2 Matrices and graphs

Any matrix can be associated to a directed graph, and properties of this graph are useful to understand that of the matrix.

A **valued directed graph** is a triple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ such that $\mathcal{E} \subset \mathcal{V}\times\mathcal{V}$ and $\mathcal{W} : \mathcal{E} \to \mathbb{C}\setminus\{0\}$.

The set $\mathcal{V}$ is any discrete set, and the elements are called vertices. $\mathcal{E}$ is the set of edges connecting the vertices. $\mathcal{W}$ is a function from the set of edges to $\mathbb{C}\setminus\{0\}$: to each edge is associated a value, called "weight" in the following.

To each valued graph, it is possible to associate a square matrix with entries in $\mathbb{C}$ and dimension $N = |\mathcal{V}|$. Given an ordering of the vertices in $\mathcal{V}$: $\mathcal{V} = \{v(1), \ldots, v(N)\}$, let $A_{i,j} = \mathcal{W}(v(i), v(j))$ if $(i,j) \in \mathcal{E}$ and $A_{i,j} = 0$ otherwise.

Conversely, to any matrix $A \in \mathcal{M}_{N\times N}(\mathbb{C})$, one associates naturally a valued directed graph $\mathcal{G}[A]$:

- The vertices in $\mathcal{V}$ are $\{1, 2, \ldots, N\}$,

- The edge (i, j) is in $\mathcal{E}$ if and only if $A_{i,j} \neq 0$.
- This edge has then the weight $\mathcal{W}(i, j) = A_{i,j}$.

A **path** in a graph $\mathcal{G}$ is a nonempty sequence of vertices $\gamma = (i_1, \ldots, i_p)$ such that $(i_j, i_{j+1}) \in \mathcal{E}$ for $j \in \{1, \ldots, p-1\}$. The first edge i_1 is the origin of the path, the last one is the end of the path. The convention here is that the *length* of the path, noted $|\gamma|$, is the number of edges that it contains (not the number of vertices), that is, $p - 1$. A path with 0 length contains only one vertex.

For a valued graph, we associate to each path γ a (multiplicative) weight $w(\gamma)$ as follows:

$$\gamma = (i_1, \ldots, i_n) \quad \Longrightarrow \quad w(\gamma) \;=\; \prod_{j=1}^{n-1} \mathcal{W}(i_j, i_{j+1}) \; . \tag{A.3}$$

By convention, $w(\gamma) = 1$ if $|\gamma| = 0$. By definition of the weighted graph, the weight of a path is never 0. The definition and the convention are consistent with the operation of *appending* paths: if $\gamma_1 = (i_1, \ldots, i_p)$ and $\gamma_2 = (j_1, \ldots, j_q)$ are two paths such that $i_p = j_1$, then one defines the path $\gamma = (i_1, \ldots, i_p, j_2, \ldots, j_q)$. Then $|\gamma| = |\gamma_1| + |\gamma_2|$ and $w(\gamma) = w(\gamma_1)w(\gamma_2)$.

Some important structural properties of the matrix can be defined from properties of this graph. We recall or define first the graph properties. These graphs do not depend on the weight of the edges.

Definition A.1. *A (strongly)* ***connected component*** *of a graph $\mathcal{G}$ is a set $\mathcal{C}$ of vertices such that there exists a path in the graph between any two pairs of vertices.*

A graph with a single connected component is called ***strongly connected****.*

The ***periodicity*** *of a vertex i is the number $p(i)$ equal to the l.c.d. (least common divisor) of the lengths of all cycles in $\mathcal{G}[\mathbf{A}]$ which go through i. All vertices in a connected component have the same periodicity.*

When a graph is strongly connected, and the periodicity of any of its vertices is p, the graph is called ***periodic*** *with period p.*

We can now state structural properties of matrices. Again, these properties do not depend on the values of the weights, just on the edges that are present.

Definition A.2. *Let $\mathbf{A} \in \mathcal{M}_{n\times n}$ be a square matrix, and $\mathcal{G}[\mathbf{A}]$ be its associated weighted graph. The matrix is said to be:*

irreducible *if the graph $\mathcal{G}$ is strongly connected;*

reducible *if it is not irreducible*

periodic *with period d if the graph is periodic with period d.*

aperiodic *if the graph is periodic with period $d = 1$.*

A.3 Powers of matrices

This section reviews results and methods for computing powers of matrices. An important theoretical result is that when the powers are large, they depend essentially on the *spectral radius* of the matrix.

A.3.1 Powers and paths

The first important observation is that powers of matrices and paths in graphs are closely related.

Theorem A.1. *Let* $\mathbf{A}$ *be a square matrix,* $\mathcal{G}$ *the associated weighted graph, and* $\Gamma^n(i,j)$ *be the set of paths of length* n *in* $\mathcal{G}$ *which go from* i *to* j*. The following identity holds:*

$$(\mathbf{A}^n)_{i,j} \;=\; \sum_{\gamma\in\Gamma^n(i,j)} w(\gamma) \;=\; \sum_{(i_1,\dots,i_n)\in\Gamma^n(i,j)} \mathcal{W}(i_1,i_2)\mathcal{W}(i_2,i_3)\dots\mathcal{W}(i_{n-1},i_n)\ . \tag{A.4}$$

Proof. Expanding the formula for the power of matrices, one obtains:

$$(\mathbf{A}^n)_{i,j} \;=\; \sum_{i_1,i_2,\dots,i_n} A_{i_1,i_2}A_{i_2,i_3}\dots A_{i_{n-1},i_n}\ .$$

But the only terms that are not 0 in this sum correspond precisely to sequences $(i_1,\dots,i_n)$ which are paths of $\mathcal{G}$. The result follows from the definition of the weight of paths (A.3). □

Observe again the consistency of the conventions made on the length of paths and their weights. If $n=0$ for instance, the $\Gamma^0(i,j)$ of paths with length 0 going from i to j is: a/ empty if $i\neq j$, so that $\mathbf{A}^0_{i,j}=0$; b/ reduced to (i,i) if $i=j$, so that $\mathbf{A}^0_{i,i}=1$. We find the expected identity: $\mathbf{A}^0=\mathbf{I}$. The operation of appending paths works well with the matrix identity: $\mathbf{A}^{p+q}=\mathbf{A}^p\mathbf{A}^q$.

A.3.2 Powers and Jordan decomposition

From the Jordan decomposition (A.2), one has:

$$A^n \;=\; (SJS^{-1})^n \;=\; SJ^nS^{-1}\ . \tag{A.5}$$

Given the form of J, this is:

$$A^n \;=\; S\begin{pmatrix} J^n_{\nu_1} & & & \\ & J^n_{\nu_2} & & \\ & & \ddots & \\ & & & J^n_{\nu_k}\end{pmatrix} S^{-1}\ . \tag{A.6}$$

Computing the power of a matrix is reduced to computing the power of a Jordan block. The latter are of the form:

$$((J_m(\mu)^k))_{i,j} \;=\; \binom{k}{j-i}\ \mu^{k-(j-i)}\ \mathbf{1}_{\{j\geq i\}}\ ,$$

that is:

$$J_m(\mu)^k \;=\; \begin{pmatrix} \mu^k & k\mu^{k-1} & & \binom{k}{n-1}\mu^{k-n+1} \\ 0 & \mu^k & \ddots & \\ & \ddots & \ddots & k\mu^{k-1} \\ & & 0 & \mu^k \end{pmatrix}\ . \tag{A.7}$$

The next results provides the exact value of $\mathbf{A}^n$ in function of n and the eigenvalues of $\mathbf{A}$. For this reason, this type of result is called a *spectral expansion*.

Theorem A.2 (Spectral Expansion). *i/ Let $\mathbf{A}$ be a matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_p$, with respective multiplicities $\alpha_1, \ldots, \alpha_p$. Then there exist p matrices $\mathbf{B}_i(n, \lambda_i)$, $1 \le i \le p$, such that each entry in $\mathbf{B}_i$ is a polynomial in λ_i with degree at most n, which coefficients are monomials in n of degree at most $\alpha_i - 1$, and such that:*

$$\mathbf{A}^n \;=\; \sum_{i=1}^{p} \mathbf{B}_i(n, \lambda_i)\;. \tag{A.8}$$

ii/ In the case where λ_i is a simple eigenvalue, then $\mathbf{B}_i(n, \lambda_i) = \lambda_i^n \mathbf{B}_i$, where $\mathbf{B}_i$ is constant and of rank 1: there exist a row vector $\mathbf{v}_i$ and a column vector $\mathbf{w}_i$ such that $\mathbf{v}_i.\mathbf{w}_i = 1$ and $\mathbf{B}_i = \mathbf{w}_i.\mathbf{v}_i$.

Proof. Consider the Jordan decomposition (A.2) and the representation of powers (A.6). The result is obtained defining:

$$\mathbf{B}_i(n, \lambda_i) \;=\; \sum_{\mu_m = \lambda_i} S \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & J_{\nu_m}(\mu_m)^n & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix} S^{-1}\;.$$

The matrices $\mathbf{B}_i$ have the required properties since S and S^{-1} are constant, and given the expression (A.7) for the powers of a Jordan block. This proves *i/*.

When λ_i is simple, the corresponding Jordan block is of size 1. Then *ii/* follows by defining for $\mathbf{v}_i$ the corresponding column of S, and $\mathbf{w}_i$ the corresponding row of S^{-1}. The scalar product of these two vectors is 1 because of the identity $S^{-1}S = \mathbf{I}$. This proves *ii/*. □

A.3.3 Asymptotic behavior

This section contains results about the growth of the coefficients of the matrix $\mathbf{A}^n$ when $n \to \infty$.

To express quantitatively the "size" of a matrix, we use the norm: for $\mathbf{A} \in \mathcal{M}_{N\times N}$,

$$\mu(\mathbf{A}) \;=\; \max\{|A_{ij}|,\; 1 \le i, j \le N\}\;.$$

Accordingly, if we write that $\mu(\mathbf{A}^n) = O(f(n))$, we express the fact that *all* coefficients of the matrix $\mathbf{A}^n$ are bounded by some constant times $f(n)$ as $n \to \infty$.

Lemma A.3. *Let $A \in \mathcal{M}_{N\times N}(\mathbb{C})$. For all $n \in \mathbb{N}$ and $\varepsilon > 0$, we have:*

$$\mu(\mathbf{A}^n) \;=\; \mathcal{O}((\rho(\mathbf{A}) + \varepsilon)^n)\;, \tag{A.9}$$

where $\rho(\mathbf{A})$ is the spectral radius of $\mathbf{A}$.

Introducing more properties of the matrix $\mathbf{A}$, we have the stronger result.

Theorem A.4. *Let $\mathbf{A} \in \mathcal{M}_{N\times N}(\mathbb{R}_+)$ be a real, positive and irreducible matrix. Then:*

i/ $\mu(\mathbf{A}^n) \;=\; \mathcal{O}(\rho(\mathbf{A})^n)$.

ii/ *There exist a row vector* $\mathbf{v}$ *and a column vector* $\mathbf{w}$ *which are strictly positive, such that* $\mathbf{v}.\mathbf{w} = 1$ *and:*

$$\mu\left(\frac{\mathbf{A}^n}{\rho(\mathbf{A})^n} - \mathbf{wv}\right) = \mathcal{O}(\varphi^n)$$

for all φ *such that* $\rho_2(\mathbf{A})/\rho(\mathbf{A}) < \varphi < 1$.

A.4 The exponential of matrices

Let $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a square matrix with complex coefficients. The exponential of $\mathbf{A}$ is defined as:

$$e^{\mathbf{A}} \;=\; \sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n \;. \tag{A.10}$$

This series converges for any matrix: this is a consequence of Lemma A.3.

The following properties are used in the course:

- If $\mathbf{D} = diag(d_1, \ldots, d_n)$ is a diagonal matrix, then $e^{\mathbf{D}} = diag(e^{d_1}, \ldots, e^{d_n})$.
- In particular, for any $x \in \mathbb{C}$,
$$e^{x\mathbf{I}} \;=\; e^x \;\mathbf{I}\;.$$
- In particular again: $e^{\mathbf{0}} = \mathbf{I}$.
- The matrices $\mathbf{A}$ and $e^{\mathbf{A}}$ commute: $\mathbf{A}e^{\mathbf{A}} = e^{\mathbf{A}}\mathbf{A}$. So do $p(\mathbf{A})$ and $e^{\mathbf{A}}$ for every polynomial $p(\cdot)$.
- If $\mathbf{A}$ and $\mathbf{B}$ are square matrices such that $\mathbf{AB} = \mathbf{BA}$ (commuting matrices), then:
$$e^{\mathbf{A}+\mathbf{B}} \;=\; e^{\mathbf{A}}\;e^{\mathbf{B}} \;=\; e^{\mathbf{B}}\;e^{\mathbf{A}}\;.$$

A.5 Positive matrices and the Perron-Frobenius theorem

Positive matrices enjoy a very particular property: their spectral radius is in fact one of the eigenvalues. This result and related ones are grouped in the following theorem.

Theorem A.5 (Perron-Frobenius)**.** *Let* $\mathbf{A} \in \mathcal{M}_{N\times N}(\mathbb{R}_+)$ *be a square matrix, positive and irreducible. Then there exists one eigenvalue of* $\mathbf{A}$*, say* r*, with the following properties:*

a/ $r > 0$*;*

b/ there are a left-eigenvector and a right-eigenvector for r *which are* strictly positive*;*

c/ for all eigenvalue λ *of* $\mathbf{A}$*,* $|\lambda| \leq r$*;*

d/ The eigenspace associated with r *is of dimension 1;*

e/ *for all matrix* $\mathbf{B}$ *such that* $0 \leq \mathbf{B} \leq \mathbf{A}$*, and all eigenvalue* β *of* $\mathbf{B}$*,* $|\beta| \leq r$*, and the equality* $|\beta| = r$ *implies that* $\mathbf{B} = \mathbf{A}$*;*

f/ r *is a simple root of* $\chi_{\mathbf{A}}$*, the characteristic polynomial of* $\mathbf{A}$*.*

g/ *if* $\mathbf{A}$ *is periodic of period* d*, then the eigenvalues of* $\mathbf{A}$ *with modulus* r *are exactly the* ω^j*,* $0 \leq j < d$ *with* $\omega = e^{2i\pi/d}$*.*

Another way of presenting $d/$ and $f/$ is to say that the eigenvalue r is simple, and its Jordan block is of dimension 1.

The eigenvalue r is called the *Perron-Frobenius eigenvalue* of $\mathbf{A}$, or the *principal eigenvalue.*

A.6 Vector norms and Matrix norms

Let V be a vector space. A *vector norm* is a function $\|\cdot\| : V \to \mathbb{R}_+$ such that: a) $\|x\| = 0$ if and only if $x = 0$; b) $\|cx\| = |c|.\|x\|$ for every scalar c; c) $\|x + y\| \leq \|x\| + \|y\|$ (the triangular inequality).

Norms and distances are closely related: if $\|\cdot\|$ is a norm, then $d(x, y) = \|x - y\|$ is a distance for elements of V.

Some distances on $V = \mathbb{R}^m$ are more often used in practice:

$$\begin{array}{rcll} \|x\|_1 & = & \sum_i |x_i| & \text{the “sum norm” or “}\ell_1\text{ norm”;} \\ \|x\|_2 & = & \left(\sum_i (x_i)^2\right)^{1/2} & \text{the Euclidian norm;} \\ \|x\|_\infty & = & \max_i |x_i| & \text{the “sup norm” or “max norm” or “}\ell_\infty\text{ norm”.} \end{array}$$

A *matrix norm*[1] is a function $\|\cdot\|$: $\mathcal{M}_{n\times m} \to \mathbb{R}_+$ such that a) $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = 0$; b) $\|c\mathbf{A}\| = |c|.\|\mathbf{A}\|$ for every scalar c; c) $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ (the triangular inequality); d) $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\,\|\mathbf{B}\|$. This last requirement makes it different than a norm on $\mathcal{M}_{n\times m}$, considered as a vector space. In addition, a matrix norm is said to be compatible with a vector norm if: $\|\mathbf{A}x\| \leq \|\mathbf{A}\|\,\|x\|$. Note that compatibility with row products $y\mathbf{A}$ and compatibility with column products $\mathbf{A}x$ is not the same.

Usual examples of matrix norms are:

$$\begin{array}{rcll} \|A\|_1 & = & \sum_{i,j} |A_{ij}| & \text{the “sum norm” or “}\ell_1\text{ norm”;} \\ |||A|||_1 & = & \max_j \sum_i |A_{ij}| & \text{the “maximum column sum” norm, compatible with the vector norm} \\ & & & \|\cdot\|_1 \text{ for row product, and the vector norm } \|\cdot\|_\infty \text{ for column products} \\ |||A|||_\infty & = & \max_i \sum_j |A_{ij}| & \text{the “maximum row sum” norm, compatible with the vector norm} \\ & & & \|\cdot\|_\infty \text{ for row product, and the vector norm } \|\cdot\|_1 \text{ for column products} \end{array}$$

[1]The terminology “matrix norm” is taken from[10]. In the book [9], the term “consistent matrix norm” is used.