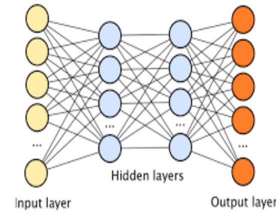
Math 4997

Math Review I: Linear Algebra



Notation: Some phrases are used so often that one uses symbols for them:

symbol	meaning	symbol	meaning
$\in$	"is an element of"	$\subseteq$	"is a subset of or equal to"
$\forall$	"for all"	$\subsetneq$	"is strictly a subset of"
$\exists$	"there exists"	$\ni$ or s.t.	"such that"
$\therefore$	"therefore"	$\square$	"End of Proof"

The number systems frequently encountered are:

$$\begin{aligned}
\mathbb{N} &= \{1,\ 2,\ 3, \dots\} \quad \text{(the natural numbers)}\\
\mathbb{N}_0 &= \{0,\ 1,\ 2,\ 3, \dots\} \quad \text{(the whole numbers)}\\
\mathbb{Z} &= \{\dots,\ -2,\ -1,\ 0,\ 1,\ 2, \dots\} \quad \text{(the integers)}\\
\mathbb{Q} &= \{p/q \ :\ p \in \mathbb{Z},\ q \in \mathbb{N}\} \quad \text{(the rational numbers)}\\
\mathbb{R} &= \text{the real numbers}\\
\mathbb{C} &= \{a + i\,b \ :\ a,\ b \in \mathbb{R}\} \quad \text{(the complex numbers), } i \text{ is a symbol satisfying } i^2 = -1\\
\mathbb{F} &= \text{Field of scalars, usually taken to be } \mathbb{R}\text{, but could be } \mathbb{Q} \text{ or } \mathbb{C} \text{ as well.}
\end{aligned}$$

For $n \in \mathbb{N}$, the n-dimensional Euclidean space $\mathbb{R}^n$ is defined as the set of all (column) vectors of length n. Specifically, we have

$$\mathbb{R}^n = \left\{ \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \;\middle|\; x_j \in \mathbb{R} \text{ for all } j = 1, \dots, n \right\} \tag{1}$$

The real number x_j appearing in (1) is called the j^{th} component or coordinate of $\mathbf{x}$. If a row vector of the form $(x_1, \dots, x_n)$ is desired, we can use the transpose operation and write $\mathbf{x}^\top = (x_1, \dots, x_n)$ where $\mathbf{x} \in \mathbb{R}^n$. It is sometimes more convenient to write $\mathbf{x}^\top$ in text because it takes up less space on the page. In the case $n = 1$, often $\mathbb{R}^1$ is written as $\mathbb{R}$, but the former notation could be used if we want to highlight the vector space structure of the real numbers instead of thinking of them only as scalars.

Here and in many places below, we could develop a more general theory by replacing the real scalar field $\mathbb{R}$ of real numbers by any field $\mathbb{F}$. For simplicity, we consider for the most part only $\mathbb{F} = \mathbb{R}$ except when discussing eigenvalues, where the complex numbers $\mathbb{C}$ *must* enter the picture.

1 Linear Algebra and vector space structure

One of the difficulties in introducing vector spaces is deciding on the appropriate level of abstraction. Although vector spaces are not usually "thought of" as functions, they can be introduced and be represented like that in the most natural way. Let X be any set; we shall use only elementary properties of the real numbers $\mathbb{R}$. The real numbers are being designated as the scalars here, hence our description of the vector space should perhaps be called a *real* vector space. The complex numbers $\mathbb{C}$ could replace $\mathbb{R}$ as the scalars, or in fact any *field* $\mathbb{F}$, in which case the space is a *complex* or $\mathbb{F}$ vector space.

Every vector space can be viewed as a set of all functions from X to $\mathbb{R}$ (for those already familiar with the concept of a vector space, X is an index set for a *basis*.) Let $\mathbb{R}^X$ denote this class of functions: $\mathbb{R}^X := \left\{ f(\cdot) \middle| f : X \to \mathbb{R} \right\}$. With $f_1(\cdot), f_2(\cdot) \in \mathbb{R}^X$, one can naturally define the sum $(f_1 + f_2)(\cdot)$ as a new element in $\mathbb{R}^X$ by $(f_1 + f_2)(x) := f_1(x) + f_2(x)\ \forall x \in X$. With $f(\cdot) \in \mathbb{R}^X$ and $r \in \mathbb{R}$, one can also define a new function $(r \cdot f)(\cdot) \in \mathbb{R}^X$ by $(r \cdot f)(x) = r \cdot f(x)\ \forall x \in X$. Note that the same symbols "+" and "·" are used in two distinct ways: the first is a definition of an operation that takes place in $\mathbb{R}^X$, and the second as the usual operation in $\mathbb{R}$. We shall encounter this multiple use of symbols many times but it should never be ambiguous or cause confusion. If X is a set with only finitely many elements, say n, then $\mathbb{R}^X$ is *finite* dimensional and can be easily identified with $\mathbb{R}^n$ as given above. Otherwise, The vector space $\mathbb{R}^X$ is said to be *infinite* dimensional. For example, with $X = [0, 1]$, $\mathbb{R}^{[0,1]}$ is the set of all real-valued functions defined on the interval $[0, 1]$. Most interesting applications of these infinite dimensional spaces involve so-called *subspaces* (a subspace will defined shortly) of the ambient space that have additional properties. For example, the continuous functions defined on $[0, 1]$ is a subspace of $\mathbb{R}^{[0,1]}$. If $Y \subseteq X$, then the vector space $\mathbb{R}^Y$ can be viewed as a subset of $\mathbb{R}^X$ by extending any $f \in \mathbb{R}^Y$ to a function defined on all of X by setting $f(x) = 0$ for any $x \in X \setminus Y$.

1.1 Algebraic operations in $\mathbb{R}^n$

Let $n \in \mathbb{N}$ and $X = \{1, 2, \dots, n\}$, but X could be any set consisting of n elements; we just label it that way for convenience. Then $\mathbb{R}^n$ defined in (1) can naturally be identified as the vector space $\mathbb{R}^X$ as described above. Each element of the set X is in essence acting like a component "identifier" in the manner that you naturally think about components.

The above discussion is perhaps too abstract, and so we now give more details in describing the vector space $\mathbb{R}^n$ which is more in line with what you have learned thus far about vector spaces. Given vectors $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^n$ and a scalar $r \in \mathbb{R}$, another vector $r\mathbf{x} + \mathbf{y}$ is formed by carrying out the usual algebraic operations defined on $\mathbb{R}$ component-wise in $\mathbb{R}^n$:

$$r\mathbf{x} + \mathbf{y} = r \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} rx_1 + y_1 \\ \vdots \\ rx_n + y_n \end{pmatrix} \tag{2}$$

It is clear that such additions and multiplications can be extended and are well-defined for any finite many vectors and scalars. The operation $r\mathbf{x}$ is called scalar multiplication (of a vector $\mathbf{x}$ by a scalar $r \in \mathbb{R}$), and $\mathbf{x} + \mathbf{y}$ is just called addition (of vectors). Notice again the double use of the addition sign "+" in (2): the first two expressions on the left use + as an addition of vectors, whereas on the right it designates addition of real numbers. Similar remarks pertain to scalar multiplication. These simple operations define the vector space structure on $\mathbb{R}^n$ A vector $\mathbf{x}$ is said to be a *linear combination* of $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ if there exist scalars $r_1, \dots, r_m$ so that $\mathbf{x} = \sum_{j=1}^{m} r_j \mathbf{v}_j$.

There is a special element $\mathbf{0_n} \in \mathbb{R}^n$, called the origin, which has all of its components equal to 0. It acts as the additive identity since it satisfies $\mathbf{x} + \mathbf{0_n} = \mathbf{x}\ \forall \mathbf{x} \in \mathbb{R}^n$. Associated with each vector $\mathbf{x}$ is a unique element $\mathbf{y}$, called the inverse of $\mathbf{x}$, that satisfies $\mathbf{x} + \mathbf{y} = \mathbf{0_n}$. Since the inverse is unique, it is written as $-\mathbf{x}$, and is the same element as $\mathbf{x}$ being multiplied by the scalar -1. The addition and scalar multiplication properties satisfy a list of axioms which are all rather obvious and will not be elaborated upon here.

Multiplication between vectors is not in general defined, at least in the sense that it provides another vector. There is however the important concept of an inner product between vectors that produces a scalar. The inner product of two vectors $\mathbf{x} = (x_1, \dots, x_n)^\top$ and $\mathbf{y} = (y_1, \dots, y_n)^\top$ is (algebraically) defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^{n} x_j y_j. \tag{3}$$

It is clear that the inner product is symmetric in the sense that $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ and has the following linear property:

$$\langle r\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y} \rangle = r\langle \mathbf{x}_1, \mathbf{y} \rangle + \langle \mathbf{x}_2, \mathbf{y} \rangle \quad \forall \mathbf{x}_1,\, \mathbf{x}_2,\, \mathbf{y} \in \mathbb{R}^n \text{ and } r \in \mathbb{R}.$$

Observe $\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{j=1}^{n} x_j^2} =: \|\mathbf{x}\|$, which is called the 2-norm of $\mathbf{x}$. We'll discuss norms in further detail below. Note in particular that if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for all $\mathbf{y} \in \mathbb{R}^n$, then $\mathbf{x} = \mathbf{0_n}$. It is not immediately obvious what the inner product does. An alternate and equivalent definition of the inner product will be given below and will clarify its geometrical meaning.

1.2 Matrices and linear maps

Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. An $m \times n$ matrix $\mathbf{A}$ consists of a rectangular array of numbers, and is indexed as follows:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{1\star} \\ \mathbf{A}_{2\star} \\ \vdots \\ \mathbf{A}_{m\star} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{\star 1} & \mathbf{A}_{\star 2} & \cdots & \mathbf{A}_{\star n} \end{pmatrix},$$

where $\mathbf{A}_{i\star}$ is the i^{th} row of $\mathbf{A}$ (and so $\mathbf{A}_{i\star}^\top \in \mathbb{R}^n$) and $\mathbf{A}_{\star j}$ is the j^{th} column of $\mathbf{A}$ (and so belongs to $\mathbb{R}^m$). The set of all $m \times n$ matrices is denoted by $\mathbb{R}^{m\times n}$. For a pair (i, j) of indices, $1 \le i \le m$ and $1 \le j \le n$, the real number a_{ij} is called the ij^{th} entry of $\mathbf{A}$. Sometimes to fix notation, we write $\mathbf{A} = (a_{ij})$.

At first thought, you might well ask what could possibly be interesting about a rectangular array of numbers? An answer lies in understanding how an $m \times n$ matrix defines a *linear mapping* from $\mathbb{R}^n$ to $\mathbb{R}^m$. The same symbol $\mathbf{A}$ is used to represent the actual matrix as well as the map $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$, which is defined by so-called matrix multiplication:

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix} = \begin{pmatrix} \langle \mathbf{A}_{1\star}^\top, \mathbf{x} \rangle \\ \langle \mathbf{A}_{2\star}^\top, \mathbf{x} \rangle \\ \vdots \\ \langle \mathbf{A}_{m\star}^\top, \mathbf{x} \rangle \end{pmatrix} = \sum_{j=1}^{n} x_j \mathbf{A}_{\star j}$$

You may now see the notational advantage of insisting an element in $\mathbb{R}^n$ be a column vector. The map $\mathbf{A}$ takes a vector $\mathbf{x} \in \mathbb{R}^n$ and produces a vector $\mathbf{A}\mathbf{x} \in \mathbb{R}^m$, where the i^{th} component $(\mathbf{A}\mathbf{x})_i$ of $\mathbf{A}\mathbf{x} \in \mathbb{R}^m$ is $(\mathbf{A}\mathbf{x})_i = \mathbf{A}_{i\star}\mathbf{x} = \sum_{j=1}^n a_{ij}x_j = \langle \mathbf{A}_{i\star}^\top, \mathbf{x} \rangle$. One always writes $\mathbf{A}\mathbf{x}$ instead of the usual function notation which would normally be $\mathbf{A}(\mathbf{x})$.

In general, a map $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *linear* provided it preserves the linear structure of each space. This means $\forall \mathbf{x}_1,\, \mathbf{x}_2 \in \mathbb{R}^n$ and $r \in \mathbb{R}$, we have $T\big(r\mathbf{x}_1 + \mathbf{x}_2\big) = rT\mathbf{x}_1 + T\mathbf{x}_2$. The set all linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ is denoted by $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. It is a simple exercise to show every $\mathbf{A} \in \mathbb{R}^{m\times n}$ belongs to $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, but in fact every linear map $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ can be represented by a unique matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$. To see this, suppose $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, and set

$\mathbf{w}_j := T(\mathbf{e}_j) \in \mathbb{R}^m$. Now let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be the matrix whose columns are the $\mathbf{w}_j$'s, and obviously $T(\mathbf{e}_j) = \mathbf{w}_j = \mathbf{A}\mathbf{e}_j$ for each $j = 1, \dots, n$. Since any $\mathbf{x} \in \mathbb{R}^n$ satisfies $\mathbf{x} = \sum_{j=1}^n x_j \mathbf{e}_j$, the linear property implies

$$T(\mathbf{x}) = T\left(\sum_{j=1}^n x_j \mathbf{e}_j\right) = \sum_{j=1}^n x_j T(\mathbf{e}_j) = \sum_{j=1}^n x_j \mathbf{w}_j = \big(\mathbf{w}_1, \dots, \mathbf{w}_n\big)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{A}\mathbf{x}.$$

This explains why the same notation $\mathbf{A}$ is used to represent the array of numbers as well as the linear mapping that it defines. In this sense, then, $\mathcal{L}\big(\mathbb{R}^n, \mathbb{R}^m\big)$ is the same as $\mathbb{R}^{m\times n}$.

1.2.1 The transpose of a matrix

The transpose of a vector was given above, and the concept can be extended to matrices. If $\mathbf{A} \in \mathbb{R}^{m\times n}$, then $\mathbf{A}^\top \in \mathbb{R}^{n\times m}$ is constructed by transforming the rows of $\mathbf{A}$ into the columns of $\mathbf{A}^\top$. Thus the ij^{th} entry a_{ij} of $\mathbf{A}$ is the ji^{th} of $\mathbf{A}^\top$. As a map, $\mathbf{A}^\top$ takes elements of $\mathbb{R}^m$ into $\mathbb{R}^n$, and one can check for any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ that

$$\langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^m (\mathbf{A}\mathbf{x})_i \, y_i = \sum_{i=1}^m \left[\sum_{j=1}^n a_{ij}\, x_j\right] y_i = \sum_{j=1}^n x_j \left[\sum_{i=1}^m a_{ij}\, y_i\right] = \sum_{j=1}^n x_j \, (\mathbf{A}^\top \mathbf{y})_j = \langle \mathbf{x}, \mathbf{A}^\top \mathbf{y}\rangle. \tag{4}$$

In fact, (4) is in essence the defining property of $\mathbf{A}^\top$ in the sense that $\mathbf{A}^\top$ is the only matrix in $\mathbb{R}^{n\times m}$ that satisfies (4) for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Notice that the first inner product in (4) is taken between vectors in $\mathbb{R}^m$, and the last between vectors in $\mathbb{R}^n$, but the same notation is used. This never causes confusion. The inner product can be stated in terms of matrix multiplication, since for $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$, we have $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^\top \mathbf{y} = \mathbf{y}^\top \mathbf{x}$.

1.2.2 Matrix algebra

Two matrices of the same dimension can be added to produce a new matrix of the same dimension by adding the corresponding entries, and scalar multiplication is defined by multiplying each entry by the scalar. Matrix multiplication is more complicated, and is defined generally between elements of $\mathbb{R}^{m\times n}$ and $\mathbb{R}^{n\times p}$ as follows: Suppose $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$, say $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{jk})$ where $1 \le i \le m$, $1 \le j \le n$, and $1 \le k \le p$. Then the product $\mathbf{AB}$ is well-defined and is an element in $\mathbb{R}^{m\times p}$ given by

$$\mathbf{AB} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j} b_{j1} & \sum_{j=1}^n a_{1j} b_{j2} & \dots & \sum_{j=1}^n a_{1j} b_{jp} \\ \sum_{j=1}^n a_{2j} b_{j1} & \sum_{j=1}^n a_{2j} b_{j2} & \dots & \sum_{j=1}^n a_{2j} b_{jp} \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{j=1}^n a_{mj} b_{j1} & \sum_{j=1}^n a_{mj} b_{j2} & \dots & \sum_{j=1}^n a_{mj} b_{jp} \end{pmatrix}$$

In shorter notation, the (i,k) entry $(\mathbf{AB})_{ik}$ equals $\big\langle \mathbf{A}^\top_{i\star}, \mathbf{B}_{\star k}\big\rangle$, where the notation for the columns and rows of a matrix is as above. If $\mathbf{AB}$ is defined, then so is $\mathbf{B}^\top \mathbf{A}^\top$, and the latter equals $(\mathbf{AB})^\top$.

It is important to note that matrix multiplication is not commutative (one usually has $\mathbf{AB} \neq \mathbf{BA}$ even when $n = m = p$), and $\mathbf{AB}$ is only defined if the number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$. It does however satisfy other arithmetic properties like associativity and the distributive property (assuming the dimensions are in accord). In this way, $\mathbb{R}^{m\times n}$ is itself another vector space (of dimension mn).

The case $m = n$ deserves some special treatment. With $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, both $\mathbf{AB}$ and $\mathbf{BA}$ are defined. There are two special matrices in $\mathbb{R}^{n\times n}$. The zero matrix $\mathbf{0}_{n\times n} \in \mathbb{R}^{n\times n}$ has entries that are all 0 and is the additive identity since it satisfies $\mathbf{A} + \mathbf{0}_{n\times n} = \mathbf{A} = \mathbf{0}_{n\times n} + \mathbf{A}$ for all $\mathbf{A} \in \mathbb{R}^{n\times n}$. The (multiplicative) identity matrix $\mathbf{I_n} \in \mathbb{R}^{n\times n}$ has entries with 1 along the main diagonal and 0's elsewhere, and satisfies $\mathbf{AI_n} = \mathbf{A} = \mathbf{I_n A}$ for all $\mathbf{A} \in \mathbb{R}^{n\times n}$. If $p(\lambda) := \lambda^k + a_{k-1}\lambda^{k-1} + \cdots + a_1\lambda + a_0$ is any polynomial, then $p(\mathbf{A}) \in \mathbb{R}^{n\times n}$ is defined by inserting $\mathbf{A}$ for λ (each power $\mathbf{A}^i$ is well-defined and belongs to $\mathbb{R}^{n\times n}$).

If $\mathbf{A} = \mathbf{A}^\top$, then $\mathbf{A}$ is said to be *symmetric*. It is always the case that $\mathbf{AA}^\top \in \mathbb{R}^{m\times m}$ and $\mathbf{A}^\top\mathbf{A} \in \mathbb{R}^{n\times n}$ are symmetric for $\mathbf{A} \in \mathbb{R}^{m\times n}$. In general one may have $\mathbf{AA}^\top \neq \mathbf{A}^\top\mathbf{A}$ even when $m = n$.

A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is invertible provided there exists a matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ with $\mathbf{AB} = \mathbf{I_n}$. In this case, it can be shown there is only one such matrix and it satisfies $\mathbf{BA} = \mathbf{I_n}$ as well. Since it is unique, one writes unambiguously $\mathbf{A}^{-1}$ for this matrix, and calls it the (multiplicative) inverse of $\mathbf{A}$. Observe if $\mathbf{A}$ is invertible, then so are $\mathbf{A}^{-1}$ and $\mathbf{A}^\top$, and that $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ and $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$. If $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ are invertible, then so is $\mathbf{AB}$, and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

In general, it can be computationally expensive and difficult in practice to find the inverse or even determine if it exists. An exception is when $n = 2$, where a matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has an inverse if and only if $ad - bc \neq 0$, in which case

$$\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{5}$$

One can check this directly.

1.3 Linear independence, spanning sets, and bases

There are many conditions that characterize the property that a matrix is invertible, and some of these are given in Proposition 1.1 below.

Definition 1.1. *Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a set of vectors in $\mathbb{R}^n$.*

(a) The set $\mathcal{B}$ is linearly independent provided that if $\sum_{i=1}^{m} r_i\mathbf{v}_i = \mathbf{0_n}$ for some scalars $r_1, \ldots, r_m$, then necessarily $r_1 = r_2 = \cdots = r_m = 0$.

(b) The span of $\mathcal{B}$ is defined as

$$\mathbf{span}(\mathcal{B}) := \left\{ \mathbf{x} \in \mathbb{R}^n : \quad \exists \text{ scalars } r_1, \ldots, r_m \text{ with with } \mathbf{x} = \sum_{i=1}^{m} r_i\mathbf{v}_i \right\}, \tag{6}$$

and $\mathcal{B}$ is said to span $\mathbb{R}^n$ provided $\mathbf{span}(\mathcal{B}) = \mathbb{R}^n$.

(c) $\mathcal{B}$ is a basis of $\mathbb{R}^n$ if it is linearly independent and spans $\mathbb{R}^n$.

An example of a basis is the so-called canonical basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, where for $1 \leq i \leq n$, $\mathbf{e}_i \in \mathbb{R}^n$ is the element in $\mathbb{R}^n$ with 1 as the i^{th} component and all the other components equal to 0. The identity matrix $\mathbf{I_n}$ is the matrix whose columns are the $\mathbf{e}_i$'s. It can be shown that every basis $\mathcal{B}$ of $\mathbb{R}^n$ contains exactly n elements. The latter statement is an instance of a more general fact about vector spaces, namely, that any two bases of a vector space have the same number of elements. The number of elements in a basis is called the dimension of the vector space $\mathcal{V}$, and is denoted by $\dim(\mathcal{V})$. Another important fact is that if $\mathcal{B}_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is any independent set in $\mathbb{R}^n$ with $k < n$, then there are additional vectors $\mathcal{B}_2 = \{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ so that the aggregate collection $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for $\mathbb{R}^n$.

Suppose $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$, and let $\mathbf{x} \in \mathbb{R}^n$. Since $\mathcal{B}$ spans $\mathbb{R}^n$, there exist scalars $x_j, \dots, x_n$ with $\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{v}_j$. It is also true that these scalars are unique. This is because if $\sum_{j=1}^{n} x_j \mathbf{v}_j = \mathbf{x} = \sum_{j=1}^{n} r_j \mathbf{v}_j$ are two representations, then $\mathbf{0_n} = \mathbf{x} - \mathbf{x} = \sum_{j=1}^{n} (x_j - r_j)\mathbf{v}_j$. Since $\mathcal{B}$ is independent, we have $x_j = r_j$ for each $j = 1, \dots, n$. The x_j's are called the coordinates of $\mathbf{x}$ with respect to $\mathcal{B}$, and is designated by $\left[\mathbf{x}\right]_{\mathcal{B}}$. Its defining property is

$$\left[\mathbf{x}\right]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{B}} \quad \iff \quad \mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{v}_j. \tag{7}$$

Note the coordinates of any $\mathbf{x} \in \mathbb{R}^n$ is another element of $\mathbb{R}^n$. We used the term coordinate before in (1) when referring to elements of $\mathbb{R}^n$. In accordance with our new usage, it is more accurate to say the x_j-terms in (1) are the coordinates with respect to the canonical basis. The map that takes an element $\mathbf{x} \in \mathbb{R}^n$ written in canonical coordinates to its $\mathcal{B}$-coordinates $\left[\mathbf{x}\right]_{\mathcal{B}}$ is denoted by $\mathcal{I}_{\mathcal{B}}$, and is an example of an *isomorphism* on from $\mathbb{R}^n$ to $\mathbb{R}^n$. An isomorphism between two vector spaces is by definition a map that is one-to-one, onto, and linear.

We showed above the relationship between matrices and linear maps. Namely, associated with a linear map $T \in \mathcal{L}\left(\mathbb{R}^n, \mathbb{R}^m\right)$ is a unique matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ in which $T(\mathbf{x})$ is $\mathbf{Ax}$. We delve further into that observation by considering bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ that may not be the canonical ones. Suppose $\mathcal{B}^n = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$ and $\mathcal{B}^m = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ is a basis of $\mathbb{R}^m$. Let $T \in \mathcal{L}\left(\mathbb{R}^n, \mathbb{R}^m\right)$ and $\mathbf{x} \in \mathbb{R}^n$, and note

$$T\left(\mathbf{x}\right) = T\left(\sum_{j=1}^{n} x_j \mathbf{v}_j\right) = \sum_{j=1}^{n} x_j T\left(\mathbf{v}_j\right).$$

Hence T is completely determined by what it does on the basis $\mathcal{B}^n$. We can now find the $\mathcal{B}^m$-coordinates of each $T\left(\mathbf{v}_j\right)$, $j = 1, \dots, n$, and define the $m\times n$ matrix $\mathbf{A}_{\mathcal{B}^n,\mathcal{B}^m}$ by letting the j^{th} column be equal to $\left[T(\mathbf{v}_j)\right]_{\mathcal{B}^m}$. The situation seems a bit complicated, but it is not, really, and is easily summarized by the following commutative diagram:

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "T"] \arrow[d, "\mathcal{I}_{\mathcal{B}^n}"'] & \mathbb{R}^m \arrow[d, "\mathcal{I}_{\mathcal{B}^m}"] \\
\mathbb{R}^n \arrow[r, "\mathbf{A}_{\mathcal{B}^n,\mathcal{B}^m}"'] & \mathbb{R}^m
\end{tikzcd}$$

Commuting property: For all $\mathbf{x} \in \mathbb{R}^n$,

$$\left(\mathcal{I}_{\mathcal{B}^m} \circ T\right)(\mathbf{x}) = \left(\mathbf{A}_{\mathcal{B}^n,\mathcal{B}^m} \circ \mathcal{I}_{\mathcal{B}^n}\right)(\mathbf{x}).$$

The point here is that starting with $\mathbf{x} \in \mathbb{R}^n$ written in canonical coordinates, one can first apply T (moving from the top left of the diagram to the top right), and then find the coordinates of $T(\mathbf{x})$ w.r.t. $\mathcal{B}^m$ (landing at the bottom right). Or, one can first find the coordinates of $\mathbf{x}$ w.r.t. $\mathcal{B}^n$ (moving this time from the top left of the diagram to the bottom left), and then multiplying this vector by $\mathbf{A}_{\mathcal{B}^n,\mathcal{B}^m}$ (landing again on the bottom right). Saying the diagram commutes means the same vector is arrived at either way.

We introduced vector spaces by describing them as all functions from a given set X to $\mathbb{R}$. Here is a direct illustration of that: If $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is any basis of $\mathbb{R}^n$, let $X = \mathcal{B}$ and define a map $T : \mathbb{R}^X \to \mathbb{R}^n$ by $T\left(f(\cdot)\right) = \mathbf{x} = (x_1, \dots, x_n)^\top$ where $x_i = f(\mathbf{v}_i)$. Then $T(\cdot)$ is an isomorphism.

The following Proposition summarizes properties of a basis and its relation to invertible matrices.

Proposition 1.1. *Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a set of n vectors in $\mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$ be the square matrix whose columns are the vectors in $\mathcal{B}$. The following properties are equivalent:*

(a) $\mathcal{B}$ *is a basis of* $\mathbb{R}^n$.

(b) $\mathbf{A}$ *is invertible.*

(c) $\mathbf{A}^\top$ *is invertible.*

(d) If $\mathbf{Ax} = 0$ *for some* $\mathbf{x} \in \mathbb{R}^n$*, then* $\mathbf{x} = \mathbf{0_n}$*; that is, the columns of A are linearly independent.*

(e) $\mathbf{A}$ *is an onto map; that is,* $\mathcal{B}$ *spans* $\mathbb{R}^n$.

1.3.1 Orthonormal bases and unitary maps

The canonical basis is an example of an *orthonormal* basis, the latter being a basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ with the additional property

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij} := \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

Hence with an orthonormal basis $\mathcal{B}$, the j^{th} component x_j of any $\mathbf{x} \in \mathbb{R}^n$ is $\langle \mathbf{x}, \mathbf{v}_j \rangle$, and $\mathbf{x} = \sum_{j=1}^{n} \langle \mathbf{x}, \mathbf{v}_j \rangle \mathbf{v}_j$.

A very special type of an invertible linear map $\mathbf{U} \in \mathbb{R}^{n \times n}$ is when $\mathbf{U}^{-1} = \mathbf{U}^\top$. Such a map $\mathbf{U}$ is called *unitary*. We next show with dimension $n = 2$ that unitary maps have the form

$$\mathbf{U} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{8}$$

for some $\theta \in [0, 2\pi)$. To see this, recall the formula (5) for the inverse of an invertible 2×2 matrix $\mathbf{U} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The invertibility property (recall (5)) requires $\delta := ad - bc \neq 0$, and the unitary property says that

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \delta \mathbf{U}^{-1} = \delta \mathbf{U}^\top = \begin{pmatrix} \delta a & \delta c \\ \delta b & \delta d \end{pmatrix}.$$

Comparing entries of the matrices, we have $d = \delta a$ and $a = \delta d$, or that $a = \delta^2 a$. This implies $\delta = 1$ or $a = 0$. We also have $b = -\delta c$ and $c = -\delta b$, or that $b = \delta^2 b$. If $a = 0$, then $b \neq 0$ since $\delta \neq 0$. In any case, we see $\delta = 1$ and $a = d$, $b = -c$, and $1 = \delta = a^2 + b^2$. There exist two θ-values with $0 \leq \theta < 2\pi$ satisfying $\sin(\theta) = -b$. If $a \geq 0$, we choose value of θ with $\cos(\theta) \geq 0$, and if $a < 0$, then θ should be chosen such that $\frac{\pi}{2} < \theta < \frac{3\pi}{2}$. Thus $\mathbf{U}$ has the form of (8) as claimed.

For any $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$, the polar coordinates of $\mathbf{x}$ are $(r, \phi) \in [0, \infty) \times [0, 2\pi)$ where $r = \sqrt{x^2 + x_2^2}$ and $\phi = \arctan\left(\frac{x_2}{x_1}\right)$ (if $x_1 = 0$, then $\phi = \frac{\pi}{2}$ if $x_2 > 0$ and $\phi = \frac{3\pi}{2}$ if $x_2 < 0$). Note that $x_1 = r\cos(\phi)$ and $x_2 = r\sin(\phi)$. Consider the image of $\mathbf{x}$ under $\mathbf{U}$:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{Ux} = \begin{pmatrix} \cos(\theta)x_1 - \sin(\theta)x_2 \\ \sin(\theta)x_1 + \cos(\theta)x_2 \end{pmatrix} = r \begin{pmatrix} \cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi) \\ \sin(\theta)\cos(\phi) + \cos(\theta)\sin(\phi) \end{pmatrix} = r \begin{pmatrix} \cos(\theta + \phi) \\ \sin(\theta + \phi) \end{pmatrix}.$$

The last equality is justified by trig identities, and one sees $y_1 = r\cos(\theta+\phi)$ and $y_2 = r\sin(\theta+\phi)$. That is, the effect of a unitary map $\mathbf{U}$ on a vector $\mathbf{x} \in \mathbb{R}^2$ is to "rotate" $\mathbf{x}$ by the fixed angle θ associated to $\mathbf{U}$ as in (8).

A general characterization of a unitary map acting on $\mathbb{R}^n$ is that it maps an orthonormal basis to another orthonormal basis. It follows that $\mathbf{U}$ is unitary precisely when it preserves the inner product. That is, $\mathbf{U}$ is unitary if and only if

$$\langle \mathbf{U}\mathbf{x}, \mathbf{U}\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle \text{ for all } \mathbf{x},\ \mathbf{y} \in \mathbb{R}^n. \tag{9}$$

The proof of this follows from (4) and the defining property $\mathbf{U}^{\top} = \mathbf{U}^{-1}$ of a unitary map $\mathbf{U}$.

1.4 Subspaces

If $\mathcal{V} \subseteq \mathbb{R}^n$ is any subset that satisfies

$$\mathbf{x} \in \mathcal{V},\ \mathbf{y} \in \mathcal{V},\ r \in \mathbb{R} \Rightarrow (r\mathbf{x} + \mathbf{y}) \in \mathcal{V},$$

then $\mathcal{V}$ is called a *subspace* of $\mathbb{R}^n$. Subspaces are subsets of $\mathbb{R}^n$ that are vector spaces in their own right where the algebraic operations are inherited from the ambient space $\mathbb{R}^n$. If $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is any collection of vectors in $\mathbb{R}^n$, then $\mathcal{V} := \mathbf{span}(\mathcal{B})$ is an example of a subspace. This also suggests a way to "split" $\mathbb{R}^n$ into pieces, something called a direct sum. As mentioned earlier, if $\mathcal{B}_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is an independent set, then there exist vectors $\mathcal{B}_2 = \{\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$ so that $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for $\mathbb{R}^n$. Let $\mathcal{W}_1 = \mathbf{span}(\mathcal{B}_1)$ and $\mathcal{W}_2 = \mathbf{span}(\mathcal{B}_2)$, which of course are two subspaces. Any element $\mathbf{x} \in \mathbb{R}^n$ can be written uniquely as

$$\mathbf{x} = \sum_{j=1}^{m} r_j \mathbf{v}_j + \sum_{j=m+1}^{n} r_j \mathbf{v}_j =: \mathbf{w}_1 + \mathbf{w}_2 \in \mathcal{W}_1 + \mathcal{W}_2, \tag{10}$$

where the last sum has the obvious definition $\mathcal{W}_1 + \mathcal{W}_2 := \left\{\mathbf{w}_1 + \mathbf{w}_2 \middle| \mathbf{w}_1 \in \mathcal{W},\ \mathbf{w}_2 \in \mathcal{W}_2\right\}$.

1.4.1 Direct sums

The decomposition of elements $\mathbf{x} \in \mathbb{R}^n$ as described in (10) is often called an *internal* direct sum because the subspaces $\mathcal{W}_1$, $\mathcal{W}_2$ lie in the ambient space $\mathbb{R}^n$. A related but slightly different construction is the *external* direct sum of two vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$, which is defined as

$$\mathcal{V}_1 \oplus \mathcal{V}_2 = \left\{(\mathbf{v}_1, \mathbf{v}_2) : \mathbf{v}_1 \in \mathcal{V}_1,\ \mathbf{v}_2 \in \mathcal{V}_2\right\}. \tag{11}$$

The addition and scalar multiplication is defined component-wise, by which is meant $(\mathbf{v}_1, \mathbf{v}_2) + (\mathbf{v}_3, \mathbf{v}_4) = (\mathbf{v}_1 + \mathbf{v}_3, \mathbf{v}_2 + \mathbf{v}_4)$ and $r(\mathbf{v}_1, \mathbf{v}_2) = (r\mathbf{v}_1, r\mathbf{v}_2)$. With either internal or external direct sums, the dimension of the direct sum is the sum of the dimensions of the two components.

If the vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ are subspaces of $\mathbb{R}^n$ with $\mathcal{V}_1 \cap \mathcal{V}_2 = \{\mathbf{0_n}\}$, then one often identifies the ordered pair $(\mathbf{v}_1, \mathbf{v}_2)$ appearing in (11) with the sum $\mathbf{v}_1 + \mathbf{v}_2$ of the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, and by so doing is merging the external and internal forms. The special symbol "$\oplus$" in (11) emphasizes the coordinate feature of how vectors are "decomposed" into smaller pieces. Two subspaces $\mathcal{V}_1$ and $\mathcal{V}_2$ of $\mathbb{R}^n$ are called *complementary* if $\mathbb{R}^n = \mathcal{V}_1 \oplus \mathcal{V}_2$, which means $\mathcal{V}_1 \cap \mathcal{V}_2 = \{\mathbf{0_n}\}$ and $\mathbb{R}^n = \mathcal{V}_1 + \mathcal{V}_2$. Given a nontrivial (i.e. not $\emptyset$ or all of $\mathbb{R}^n$) subspace $\mathcal{V}_1$, there are many complementary subspaces $\mathcal{V}_2$ associated with $\mathcal{V}_1$, and the *projection* map $\mathbf{P} : \mathbb{R}^n \to \mathbb{R}^n$ given by $\mathbf{P}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{v}_1$ depends on which complementary subspace $\mathcal{V}_2$ is used.

Two sets $\mathcal{V}_1$ and $\mathcal{V}_2$ of $\mathbb{R}^n$ are orthogonal to each other (written as $\mathcal{V}_1 \perp \mathcal{V}_2$) provided $\langle \mathbf{v}_1, \mathbf{v}_2\rangle = 0\ \forall \mathbf{v}_1 \in \mathcal{V}_1,\ \mathbf{v}_2 \in \mathcal{V}_2$. If $\mathcal{V}_1$ and $\mathcal{V}_2$ are in addition subspaces, then it is clear $\mathcal{V}_1 \perp \mathcal{V}_2$ implies $\mathcal{V}_1 \cap \mathcal{V}_2 = \{\mathbf{0_n}\}$. So in this case the direct sum $\mathcal{V}_1 \oplus \mathcal{V}_2$ is well-defined. To emphasize the additional orthogonality property, one can write $\mathcal{V}_1 \obot \mathcal{V}_2$ in place of $\mathcal{V}_1 \oplus \mathcal{V}_2$.

For any subspace $\mathcal{V} \subseteq \mathbb{R}^n$, the inner product can be used to find a particular subspace complementary to it. The *orthogonal complement* $\mathcal{V}^\perp$ of $\mathcal{V}$ is defined by

$$\mathcal{V}^\perp := \left\{ \mathbf{v} \in \mathbb{R}^n \;\middle|\; \langle \mathbf{v}, \mathbf{w} \rangle = 0 \quad \forall \mathbf{w} \in \mathcal{V} \right\}.$$

One has $\mathbb{R}^n = \mathcal{V} \oplus \mathcal{V}^\perp$ and $\dim \mathcal{V}^\perp = n - \dim \mathcal{V}$.

1.4.2 The Gram-Schmidt process

The Gram-Schmidt process is an algorithmic technique that transforms an independent set $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of $\mathbb{R}^n$ into another independent set $\tilde{\mathcal{B}} = \{\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_m\}$ of $\mathbb{R}^n$ that is orthonormal and has the same span as $\mathcal{B}$. It works like this:

$\tilde{\mathbf{v}}_1 = \mathbf{v}_1/\|\mathbf{v}_1\|$	$\mathbf{u}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \tilde{\mathbf{v}}_1 \rangle \tilde{\mathbf{v}}_1$
$\tilde{\mathbf{v}}_2 = \mathbf{u}_2/\|\mathbf{u}_2\|$	$\mathbf{u}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \tilde{\mathbf{v}}_1 \rangle \tilde{\mathbf{v}}_1 - \langle \mathbf{v}_3, \tilde{\mathbf{v}}_2 \rangle \tilde{\mathbf{v}}_2$
$\tilde{\mathbf{v}}_3 = \mathbf{u}_3/\|\mathbf{u}_3\|$	$\cdots$
$\vdots$	$\vdots$
$\tilde{\mathbf{v}}_{m-1} = \mathbf{u}_{m-1}/\|\mathbf{u}_{m-1}\|$	$\mathbf{u}_m = \mathbf{v}_m - \sum_{j=1}^{m-1} \langle \mathbf{v}_m, \tilde{\mathbf{v}}_j \rangle \tilde{\mathbf{v}}_j$
$\tilde{\mathbf{v}}_m = \mathbf{u}_m/\|\mathbf{u}_m\|$	

One can check $\tilde{B}$ is independent and spans the span of $\mathcal{B}$; in fact, $\mathbf{span}\big(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}\big) = \mathbf{span}\big(\{\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_k\}\big)$ for each $1 \le k \le m$. The process is not as complicated as it may first appear, for it has a simple geometrical interpretation that we'll see later after we interpret the inner product geometrically.

Proposition 1.2. *For any $\mathbf{0_n} \neq \mathbf{y} \in \mathbb{R}^n$, there exists an orthonormal basis whose first element is $\mathbf{y}/\|\mathbf{y}\|$.*

Proof. One can find a basis of $\mathbb{R}^n$ consisting of $\{\mathbf{y}, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. Applying the Gram -Scmidt process produces the desired orthonormal basis. □

There is a generalization of the characterization of unitary maps that was described in (8) for dimension two. A unit vector $\mathbf{u} \in \mathbb{R}^n$ is one with $\langle \mathbf{u}, \mathbf{u} \rangle = 1$, and the set of all unit vectors is denoted by S^1. It can be shown that $\mathbf{U} \in \mathbb{R}^{n \times n}$ is unitary if and only if the columns $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ all belong to S^1 and satisfy $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ for each $i \neq j$. This is because the ij entry of $\mathbf{U}\mathbf{U}^\top$ is $\langle \mathbf{u}_i, \mathbf{u}_j \rangle$.

1.4.3 The null and range spaces

The existence of subspaces and properties of matrices are inextricably intertwined. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. The *null space* of $\mathbf{A}$ is the set

$$\mathcal{N}(\mathbf{A}) := \left\{ \mathbf{x} \in \mathbb{R}^m \middle| \mathbf{A}\mathbf{x} = \mathbf{0_n} \right\}$$

and is a subspace of $\mathbb{R}^n$. The *range space* of $\mathbf{A}$ is the set

$$\mathcal{R}(\mathbf{A}) = \left\{ \mathbf{y} \in \mathbb{R}^m \middle| \; \exists \mathbf{x} \in \mathbb{R}^n \text{ with } \mathbf{A}\mathbf{x} = \mathbf{y} \right\},$$

and is a subspace of $\mathbb{R}^m$. One should note that $\mathcal{R}(\mathbf{A})$ is just the span of the columns of $\mathbf{A}$ and $\dim\big(\mathcal{R}(\mathbf{A})\big)$ is called the *rank* of $\mathbf{A}$. Conversely, suppose $\mathcal{W} \subseteq \mathbb{R}^n$ is some subspace of dimension m with $0 \le m \le n$. Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ be a basis of $\mathcal{W}$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$ the matrix whose rows consist of the $\mathbf{v}_i^\top$'s. Then $\mathcal{R}(\mathbf{A}) = \mathcal{W}$, and illustrates that a subset $\mathcal{W} \subseteq \mathbb{R}^m$ is a subspace if and only if it is the range of some $m \times n$ matrix.

Suppose $\mathbf{A} \in \mathbb{R}^{m\times n}$. It can be easily shown from (4) that $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^\top)^\perp$. Thus $\mathbb{R}^n = \mathcal{N}(\mathbf{A}) \oplus \mathcal{R}(\mathbf{A}^\top)$, from whence we conclude

$$n = \dim\big(\mathcal{N}(\mathbf{A})\big) + \dim\big(\mathcal{R}(\mathbf{A})\big). \tag{12}$$

The *rank* of $\mathbf{A}$ by definition is the dimension of $\mathcal{R}(\mathbf{A})$, and it equals the rank of $\mathbf{A}^\top$.

Proposition 1.1(d) says (with $m = n$) $\mathbf{A}$ is invertible if and only if $\mathcal{N}(\mathbf{A}) = \{\mathbf{0_n}\}$, and part(e) of that same proposition says $\mathbf{A}$ is invertible if and only if $\mathcal{R}(\mathbf{A}) = \mathbb{R}^n$.

1.5 Eigenvalues and eigenvectors

Throughout this section we consider only square matrices ($m = n$). Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$. A scalar $\lambda \in \mathbb{R}$ is an *eigenvalue* of $\mathbf{A}$ provided there exists a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ (called an eigenvector associated to λ) so that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. That is, $\mathbf{A}$ acts in a very simple and special way on an eigenvector $\mathbf{v}$ by just multiplying $\mathbf{v}$ by λ. If $\mathbf{v}$ is an eigenvector, then so is any nonzero multiple of $\mathbf{v}$ also an eigenvector of the same eigenvalue. It is sometimes convenient to choose the eigenvector $\mathbf{v}$ with $\|\mathbf{v}\| = 1$.

We shall see how to find eigenvalues and eigenvectors shortly, but let us first reverse the process and build a linear mapping by specifying its eigen-structure. Suppose $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is any basis of $\mathbb{R}^n$, and let $\lambda_1, \dots, \lambda_n$ be any collection of n scalars. Every $\mathbf{x} \in \mathbb{R}^n$ can be written uniquely as $\mathbf{x} = \sum_{j=1}^{n} \alpha_j \mathbf{v}_j$. We now *define* $T : \mathbb{R}^n \to \mathbb{R}^n$ by

$$T(\mathbf{x}) = T\left(\sum_{j=1}^{n} \alpha_j \mathbf{v}_j\right) = \sum_{j=1}^{n} \lambda_j\, \alpha_j\, \mathbf{v}_j. \tag{13}$$

Any linear map is determined by what it does on a basis. Here, we started with a basis $\mathcal{B}$ and specified that the map T take an element $\mathbf{v}_j \in \mathcal{B}$ to $\lambda_j \mathbf{v}_j$. Requiring that T be linear determines the behavior of T on all of $\mathbb{R}^n$. We have constructed a linear map T on $\mathbb{R}^n$ whose associated matrix representation $\mathbf{A}$ has eigenvalues $\lambda_1, \dots, \lambda_n$ with associated eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$.

This is how a matrix $\mathbf{A}$ can be constructed where the eigenvalues and eigenvectors are prescribed, and is a desirable form to know because it is so easy to understand. The mapping $\mathbf{A}$ just distorts the direction of each eigenvector by its eigenvalue. The next obvious question is whether every such mapping can be "decomposed" in this manner, and unfortunately the answer is "no ". But sometimes it can, which means a basis can be found where each vector in the basis is an eigenvector. If such a basis exists, it is called an *eigenbasis* and $\mathbf{A}$ is said to be *diagonalizable*.

We next see how to find eigenvalues. Recall an eigenvalue λ of $\mathbf{A} \in \mathbb{R}^{n\times n}$ satisfies $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ for some $\mathbf{v} \neq \mathbf{0_n}$, which means the same thing as saying the null space $\mathcal{N}(\lambda\mathbf{I_n} - \mathbf{A})$ is not $\{\mathbf{0_n}\}$. By Proposition 1.1(d), we can conclude that λ is an eigenvalue if and only if the inverse $(\lambda\mathbf{I_n} - \mathbf{A})^{-1}$ does not exist. We need a (relatively) simple manner to determine that, and this leads naturally to determinants.

1.6 Determinants

Associated with a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a real number called its *determinant*. There are several ways to define the determinant algebraically, but before doing that, we first explore its geometric meaning. The best situation is when there exists an orthonormal eigenbasis, in which case the absolute value of the determinant $\det(\mathbf{A})$ is the volume of $\mathbf{A}\mathbb{B}_1^+ := \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{B}_1^+\}$, where $\mathbb{B}_1^+ := \{\mathbf{x} \in \mathbb{R}^n \big| 0 \le x_j \le 1\}$ is the unit cube. This means the determinant is the amount that $\mathbf{A}$ distorts a unit of space, (and is negative if an odd number of directions is "flipped"). The next best situation is when $\mathbf{A}$ is diagonalizable, in which the determinant is the product of the distortions in each eigen-direction. With $n = 1$, a matrix has only one row and column, and so has $\mathbf{A} = (\mathbf{a})$. The determinant in this case is just $\mathbf{a}$. Here $\mathcal{B}_1 = [0,1]$, and $\mathbf{A}\mathcal{B}_1 = [0,a]$ if $a \geq 0$ and $[a,0]$ if $a < 0$. Consider dimension two, $\mathbf{A} \in \mathbb{R}^{2\times 2}$. Suppose there exists an orthonormal eigenbasis $\{\mathbf{v}_1, \mathbf{v}_2\}$ with eigenvalues λ_1, λ_2, and let $\mathbf{U} = \begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{pmatrix}$ be the 2×2 matrix whose columns are $\mathbf{v}_1$ and $\mathbf{v}_2$. Then $\mathbf{U}$ is a unitary matrix and preserves areas, since it has the form (8) which is a rotation. Since $\mathbf{A}\mathbf{U}\mathbb{B}_1^+ = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2$. The "signed" area of the rectangle with sides $\lambda_1\mathbf{v}_1$ and $\lambda_2\mathbf{v}_2$ is $\lambda_1\lambda_2$. If $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an eigenbasis but not orthogonal, then this interpretation of the determinant as a distortion of area does not hold exactly. Nonetheless, the determinant is the product of the eigenvalues, and thus is still a measurement of how much distortion is done in each eigen-direction. It gets a little more complicated yet if there is no eigenbasis, but this only happens when an eigenvalue is repeated and the determinant in that case is still the product of the eigenvalues but with the repeated eigenvalues raised to the power of their number of repetitions. What is important from this geometrical interpretation is that 0 is an eigenvalue if and only if the distortion gives a zero volume, which is the case if and only if $\mathcal{R}(\mathbf{A})$ is not all of $\mathbb{R}^n$, and which is true if and only if $\mathbf{A}$ is not invertible.

Let $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n\times n}$ for some $n \in \mathbb{N}$. The formal definition of the determinant $\det(\mathbf{A})$ is given inductively on n. When $n = 1$, then $\det(\mathbf{A}) = a_{11}$. Suppose for some $n \in \mathbb{N}$, $\det(\mathbf{A})$ has defined for any $\mathbf{A} \in \mathbb{R}^{(n-1)\times(n-1)}$, and consider $\mathbf{A} \in \mathbb{R}^{n\times n}$. For each $1 \leq i,\ j \leq n$, let $\mathbf{A}(i|j) \in \mathbb{R}^{(n-1)\times(n-1)}$ be defined by deleting the i^{th} row and the j^{th} column from $\mathbf{A}$. The induction hypothesis implies $\det(\mathbf{A}(i|j))$ is defined for all i, j, and $\det(\mathbf{A})$ is defined by setting

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(\mathbf{A}(1|j)).$$

This definition gives a special status to the first row $i_0 = 1$, but in fact, expanding along any row i_0 or column j_0 results in the same quantity. We have

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{i_0+j} a_{i_0 j} \det(\mathbf{A}(i_0|j)) = \sum_{i=1}^{n} (-1)^{i+j_0} a_{i j_0} \det(\mathbf{A}(i|j_0)) \qquad \forall i_0,\ j_0 \in \{1,\dots,n\}. \tag{14}$$

These are all the same because they all equal

$$\det(\mathbf{A}) = \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)}\, a_{2,\sigma(2)}\, \cdots\, a_{n,\sigma(n)}, \tag{15}$$

where Σ_n is the collection of all permutations of $\{1,\dots,n\}$. Hence $\sigma \in \Sigma_n$ is nothing but a one-to-one onto map from $\{1,\dots,n\}$ to itself. Its "sign" $sgn(\sigma)$ is -1 if the number of single interchanges needed to produce σ is odd, and $+1$ if an even number is needed. For example, with $n = 3$, the permutations $(1\ 2\ 3) \xrightarrow{\sigma} (2\ 1\ 3)$ and $(1\ 2\ 3) \xrightarrow{\sigma} (3\ 2\ 1)$ are both odd, and the permutations $(1\ 2\ 3) \xrightarrow{\sigma} (2\ 3\ 1)$ and $(1\ 2\ 3) \xrightarrow{\sigma} (3\ 1\ 2)$ are both even.

Proposition 1.3. *The determinant has the following properties (where $n \in \mathbb{N}$, $\mathbf{A}$, $\mathbf{B} \in \mathbb{R}^{n\times n}$).*

(a) $\det(\mathbf{A}) = \det(\mathbf{A}^\top)$.

(b) $\mathbf{A}$ *is invertible if and only if* $\det(\mathbf{A}) \neq 0$. *If* $\mathbf{A}$ *is invertible, then* $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$.

(c) $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$.

(d) *If* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by switching two rows or columns of* $\mathbf{A}$, *then* $\det(\mathbf{B}) = -\det(\mathbf{A})$.

(e) *If* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by multiplying one row or column of* $\mathbf{A}$ *by a scalar* $r \in \mathbb{R}$, *then* $\det(\mathbf{B}) = r\det(\mathbf{A})$.

(f) *Suppose* $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{R}^{n\times n}$ *have the form (for some* $i_0 \in \{1,\dots,n\}$*)*

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{i_0 1} & \cdots & a_{i_0 n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}; \quad \mathbf{B} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ b_1 & \cdots & b_n \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}; \quad \mathbf{C} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ c_1 & \cdots & c_n \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix},$$

where $a_{i_0 j} = b_j + c_j$ *for* $j = 1,\dots,n$. *Then* $\det\big(\mathbf{A}\big) = \det\big(\mathbf{B}\big) + \det\big(\mathbf{C}\big)$

(g) *Suppose* $\mathbf{A}$ *is upper (resp. lower) diagonal. Then* $\det(\mathbf{A})$ *is the product of its diagonal entries. That is, if*

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1(n-1)} & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2(n-1)} & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3(n-1)} & a_{3n} \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix} \text{ or, resp.,} \mathbf{A} = \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 & 0 \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ a_{(n-1)1} & a_{(n-1)2} & a_{(n-1)3} & \cdots & a_{(n-1)(n-1)} & 0 \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{(n-1)(n-1)} & a_{nn} \end{pmatrix}$$

then $\det(\mathbf{A}) = a_{11} \cdot a_{22} \cdot a_{33} \cdot \cdots \cdot a_{nn}$.

(h) *If* $\mathbf{A}$ *has the* block *structure*

$$\mathbf{A} = \left(\begin{array}{c|c|c|c|c} \mathbf{A}_1 & \mathbf{0}_{n_1 n_2} & \cdots & \mathbf{0}_{n_1 n_{k-1}} & \mathbf{0}_{n_1 n_k} \\ \hline \mathbf{0}_{n_2 n_1} & \mathbf{A}_2 & \cdots & \mathbf{0}_{n_2 n_{k-1}} & \mathbf{0}_{n_2 n_k} \\ \hline \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline \mathbf{0}_{n_{k-1} n_1} & \mathbf{0}_{n_{k-1} n_2} & \cdots & \mathbf{A}_{k-1} & \mathbf{0}_{n_{k-1} n_k} \\ \hline \mathbf{0}_{n_k n_1} & \mathbf{0}_{n_k n_2} & \cdots & \mathbf{0}_{n_k n_{k-1}} & \mathbf{A}_k \end{array}\right),$$

where each $\mathbf{A}_i \in \mathbb{R}^{n_i \times n_i}$ *with* $n_1 + \cdots + n_k = n$ *and* $\mathbf{0}_{n_i n_j}$ *denotes the zero matrix in* $\mathbb{R}^{n_i \times n_j}$*, then*

$$\det\big(\mathbf{A}\big) = \prod_{i=1}^{k} \det(\mathbf{A}_i). \tag{16}$$

[i) $\det(\mathbf{A})$ *is the product of all the eigenvalues (including multiplicity).*

The *characteristic polynomial* of $\mathbf{A} \in \mathbb{R}^{n\times n}$ is defined as $p(\lambda) := \det\big(\lambda \mathbf{I_n} - \mathbf{A}\big)$, and is a polynomial in λ of degree n. The roots of $p(\cdot)$ are precisely the eigenvalues of $\mathbf{A}$, and thus provides the means to find eigenvalues - just find the roots of $p(\cdot)$. Of course, polynomials may have complex roots, so it is not immediately clear what this means as regards to eigenvalues. If $\mathbf{A}$ has real entries (as we are assuming throughout these notes), and $\lambda = a + ib \in \mathbb{C}$ is a complex root of $p(\cdot)$, then the Complex conjugate $\overline{\lambda} = a - ib$ is also a root. This is because $p(\cdot)$ has real coefficients and so $\overline{p(\lambda)} = p\left(\overline{\lambda}\right)$. Here we used the simple facts that for any complex numbers λ_1 and λ_2, one has $\overline{\lambda_1 \lambda_2} = \overline{\lambda_1}\,\overline{\lambda_2}$ and $\overline{\lambda_1 + \lambda_2} = \overline{\lambda_1} + \overline{\lambda_2}$. One should note that $|\lambda| := \sqrt{\alpha^2 + \beta^2} = \sqrt{\overline{\lambda}\lambda} = |\overline{\lambda}|$.

One way to interpret these complex eigenvalues is to think of $\mathbf{A}$ as a linear operator on $\mathbb{C}^n$ rather than on $\mathbb{R}^n$, where $\mathbb{C}^n := \left\{\mathbf{z} = \mathbf{x} + i\mathbf{y} \;\middle|\; \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n\right\}$. Then λ is an eigenvalue if and only if there exists an associated eigenvector $\mathbf{z} = \mathbf{x} + i\mathbf{y} \in \mathbb{C}^n$ satisfying $\mathbf{z} \neq \mathbf{0_n} + i\mathbf{0_n}$ and $\mathbf{A}\mathbf{z} = \lambda\mathbf{z}$. As mentioned above, the assumption that $\mathbf{A}$ has real entries assures that any pure complex eigenvalue can occur only when its complex conjugate is also an eigenvalue. With $\overline{\mathbf{z}} := \mathbf{x} - i\mathbf{y}$, we have $\mathbf{A}\overline{\mathbf{z}} = \overline{\mathbf{A}\mathbf{z}} = \overline{\lambda\mathbf{z}} = \overline{\lambda}\overline{\mathbf{z}}$, and so $\overline{\mathbf{z}}$ is an associated eigenvector associated to $\overline{\lambda}$.

If $\mathbf{A}$ has entries that are complex numbers, extending the definitions of the eigen-concepts to complex numbers and vectors is straightforward. It may then be the case that eigenvalues do not occur as conjugate pairs.

The set of all eigenvalues (not including repeated values) is denoted by $\sigma(\mathbf{A})$, and is also called the *spectrum* of $\mathbf{A}$. The spectrum provides important information about how the matrix operates, but of course it does not characterize the operator. The same polynomial can be the characteristic polynomial of several essentially different matrices. An important endeavor is to precisely define what is meant by "essentially different". Two matrices $\mathbf{A}_1$, $\mathbf{A}_2 \in \mathbb{R}^{n\times n}$ are said to be *similar* if there exists an invertible matrix $\mathbf{U} \in \mathbb{R}^{n\times n}$ so that $\mathbf{A}_1 = \mathbf{U}\mathbf{A}_2\mathbf{U}^{-1}$, or equivalently, $\mathbf{A}_1\mathbf{U} = \mathbf{U}\mathbf{A}_2$. In this case, one writes $\mathbf{A}_1 \sim \mathbf{A}_2$, and this property defines an equivalence relation on $\mathbb{R}^{n\times n}$. We shall discuss the similarity relation in greater detail below in Section 1.8

Proposition 1.4. *Suppose* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *has n distinct real eigenvalues. Then* $\mathbf{A}$ *is diagonalizable.*

Proof. Each eigenvalue λ_i has an associated eigenvector $\mathbf{v}_i$, and we need to show the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is independent. Suppose $\alpha_1, \ldots, \alpha_n$ are scalars satisfying $\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \mathbf{0_n}$. We first show $\alpha_1 = 0$. Let $p_1(\lambda) := (\lambda - \lambda_2)(\lambda - \lambda_3)\ldots(\lambda - \lambda_n)$. Then $p_1(\mathbf{A})\mathbf{v}_i = \mathbf{0_n}$ for each $i = 2, \ldots, n$ since $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Also since the λ_i's are distinct, we have $p_1(\mathbf{A})\mathbf{v}_1 = p_1(\lambda_1)\mathbf{v}_1 \neq \mathbf{0_n}$. We therefore have

$$\mathbf{0_n} = p_1(\mathbf{A}) \left[\sum_{i=1}^{n} \alpha_i \mathbf{v}_i \right] = \sum_{i=1}^{n} \alpha_i p_1(\mathbf{A})\, \mathbf{v}_i = \alpha_1 p_1(\lambda_1)\mathbf{v}_1.$$

We conclude $\alpha_1 = 0$. We can do a similar analysis to every index and conclude $\alpha_j = 0$ for all $j = 1, \ldots, n$. □

A second proof of Proposition 1.4 will begin at the end of Section 1.7.4 below.

1.7 Polynomials

Suppose $p(\lambda) = \lambda^k + a_{k-1}\lambda^{k-1} + \cdots + a_1\lambda + a_0$ is a polynomial, where each $a_i \in \mathbb{R}$. The variable λ can be thought of as just a placeholder, and in particular it makes sense to insert a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ for λ ($\mathbf{A}$ needs to be square for all the powers $\mathbf{A}^i$ to be defined). In this way, we obtain a new matrix $p(\mathbf{A})$. Factoring, multiplying, and other algebraic operations on polynomials carries over to matrices. For example, if $p(\lambda) = p_1(\lambda)p_2(\lambda)$, then $p(\mathbf{A}) = p_1(\mathbf{A})p_2(\mathbf{A})$.

Factoring polynomials in essence extends to "factoring" matrices: If $\sigma(\mathbf{A})$ has distinct real eigenvalues $\sigma(\mathbf{A}) = \{\lambda_1, \ldots, \lambda_k\}$ and $p(\lambda) := \det(\lambda\mathbf{I_n} - \mathbf{A})$, then

$$p(\mathbf{A}) = (\mathbf{A} - \lambda_1\mathbf{I_n})^{n_1}(\mathbf{A} - \lambda_2\mathbf{I_n})^{n_2} \ldots (\mathbf{A} - \lambda_k\mathbf{I_n})^{n_k}$$

where $n_1 + n_2 + \cdots + n_k = n$. This is because $p(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \ldots (\lambda - \lambda_k)^{n_k}$. The exponent n_i of the factor $(\lambda - \lambda_i)$ is called the *algebraic multiplicity* of the eigenvalue λ_i. The *geometrical multiplicity* is defined as the number of independent eigenvectors of λ_i, and is possibly strictly smaller.

1.7.1 The spectral mapping theorem

The spectral mapping theorem says that

$$\sigma\big(p(\mathbf{A})\big) = p\big(\sigma(\mathbf{A})\big) := \big\{p(\lambda) : \lambda \in \sigma(\mathbf{A})\big\}.$$

This is easy to see. Indeed, $\lambda \in \sigma(\mathbf{A})$ if and only if there exists $\mathbf{v} \neq \mathbf{0_n}$ with $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. In this case, one also has $\mathbf{A}^i\mathbf{v} = \lambda^i\mathbf{v}$ for every $i \in \mathbb{N}$. It is immediate then that $p(\mathbf{A})\mathbf{v} = p(\lambda)\mathbf{v}$, or that $p(\lambda) \in \sigma\big(p(\mathbf{A})\big)$.

1.7.2 The Cayley-Hamilton theorem

Theorem 1.1. *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *and* $p(\lambda) := \det(\lambda\mathbf{I_n} - \mathbf{A})$. *Then one has* $p(\mathbf{A}) = \mathbf{0}_{n\times n}$.

1.7.3 Minimal polynomials

Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$. One might well ask if there is a polynomial $m(\cdot)$ of degree less than n with $m(\mathbf{A}) = \mathbf{0}_{n\times n}$. The Cayley-Hamilton Theorem says the degree of $m(\mathbf{A})$ is less than or equal to n. For example, consider

$$\mathbf{A}_1 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},\ \mathbf{A}_2 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},\ \mathbf{A}_3 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix},\ \mathbf{A}_4 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},\ \mathbf{A}_5 = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

Then $(\lambda-2)^4$ is the characteristic of each of these. The minimal polynomial for $\mathbf{A}_1$ is $(\lambda-2)$, for $\mathbf{A}_2$ and $\mathbf{A}_3$ it is $(\lambda-2)^2$, for $\mathbf{A}_4$ it is $(\lambda-2)^3$, and for $\mathbf{A}_5$ is $(\lambda-2)^4$. The algebraic multiplicity for each of the $\mathbf{A}_i$'s is 4, and the respective geometric multiplicities are 4, 3, 2, 2, 1.

The minimal polynomial $m(\lambda)$ must divide the characteristic polynomial since the roots of the two polynomials must coincide. The multiplicity of a root of $m(\cdot)$ is less than or equal to the multiplicity associated with the characteristic polynomial, and may be strictly less as the above examples show.

1.7.4 The companion matrix

Given a polynomial $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$ with $a_i \in \mathbb{R}$, the companion matrix associated to $p(\cdot)$ is defined as

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ & \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-2} & -a_{n-1} \end{pmatrix}$$

It turns out that $p(\cdot)$ is both the characteristic and minimal polynomial of $\mathbf{A}$. The characteristic polynomial can be calculated by using (14) with $i_0 = n$:

$$p(\lambda) = \det\big(\lambda\mathbf{I_n} - \mathbf{A}\big) = \det \begin{pmatrix} \lambda & -1 & \cdots & 0 & 0 \\ 0 & \lambda & \ldots & 0 & 0 \\ & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda & -1 \\ a_0 & a_1 & \cdots & a_{n-2} & \lambda + a_{n-1} \end{pmatrix}$$

$$= (-1)^{n+1}(a_0)\det\begin{pmatrix} -1 & \cdots & 0 & 0 \\ \lambda & \ldots & 0 & 0 \\ & \vdots & & \vdots \\ 0 & \cdots & \lambda & -1 \end{pmatrix} + (-1)^{n+2}(a_1)\det\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & -1 & \ldots & 0 \\ & \vdots & & \vdots \\ 0 & \cdots & \lambda & -1 \end{pmatrix} + \cdots + (-1)^{n+n}(\lambda + a_{n-1})\det\begin{pmatrix} \lambda & -1 & \cdots & 0 \\ 0 & \lambda & \ldots & 0 \\ & \vdots & & \vdots \\ 0 & & \cdots & \lambda \end{pmatrix}$$

$$= (-1)^{n+1}a_0(-1)^{n-1} + (-1)^{n+2}a_1(-1)^{n-2}\lambda + \cdots + (-1)^{2n}\big(\lambda + a_{n-1}\big)\lambda^{n-1}$$

$$= a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n$$

It is, as is typical for this issue, somewhat harder to show $p(\cdot)$ is also the minimal polynomial. We'll illustrate this in only a very special case. Consider $n = 4$ and $p(\lambda) = (\lambda-2)^4 = \lambda^4 - 8\lambda^3 + 24\lambda^2 - 32\lambda + 16 = (\lambda-1)^4$. It is clear that the companion matrix $\mathbf{A} \in \mathbb{R}^{4\times 4}$ has only one eigenvalue $\lambda_1 = 2$, and we show next that $\mathbf{A}$ has exactly one (independent) eigenvector. Indeed, we show $\mathbf{x}_1 := (x_1,\ x_2, \dots, x_n)^\top$ is an eigenvector if and only if

$$2\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = 2\mathbf{x}_1 = \mathbf{A}\mathbf{x}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -16 & 32 & -24 & 8 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_3 \\ x_4 \\ -16x_1 + 32x_2 - 24x_3 + 8x_4 \end{pmatrix}$$

$$\Leftrightarrow \left\{ \begin{array}{c} x_2 = 2x_1 \\ x_3 = 2x_2 = 4x_1 \\ x_4 = 2x_3 = 8x_1 \\ -16x_1 + 32x_2 - 24x_3 + 8x_4 = 2x_4 \end{array} \right\} \Leftrightarrow \quad (-16 + 64 - 96 + 64)x_1 = 16x_1 \quad \Leftrightarrow \quad \begin{array}{c} x_1 \in \mathbb{R} \text{ is} \\ \text{arbitrary} \end{array}.$$

Therefore, (taking $x_1 = r \neq 0$), one has that $\mathbf{x}_1$ is an eigenvector, and every eigenvector has the form $\mathbf{v} = r(1,\ 2,\ 4,\ 8)^\top$. For definiteness, we take $r = 1$. We next claim there exists a vector $\mathbf{x}_2 := (x_1, x_2, x_3, x_4)^\top \neq \mathbf{0_n}$ for which $(2\mathbf{I}_4 - \mathbf{A})\mathbf{x}_2 = \mathbf{x}_1$, in which case it follows that $\mathcal{N}(2\mathbf{I}_4 - \mathbf{A})^2 \neq \mathbb{R}^4$ since $\mathbf{x}_2 \notin \mathcal{N}(2\mathbf{I}_4 - \mathbf{A})^2$. We just check:

$$\begin{pmatrix} 1 \\ 2 \\ 4 \\ 8 \end{pmatrix} = \mathbf{x}_1 = (2\mathbf{I}_4 - \mathbf{A})\mathbf{x}_2 = \begin{pmatrix} 2 & -1 & 0 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 2 & -1 \\ 16 & -32 & 24 & -6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ 2x_2 - x_3 \\ 2x_3 - x_4 \\ 16x_1 - 32x_2 + 24x_3 - 6x_4 \end{pmatrix}$$

$$\Leftrightarrow \left\{ \begin{array}{c} x_2 = -1 + 2x_1 \\ x_3 = -2 + 2x_2 = -4 + 4x_1 \\ x_4 = -4 + 2x_3 = -12 + 8x_1 \\ (16 - 64 + 96 - 48)x_1 = 8 - 32 + 96 - 72 \end{array} \right\} \Leftrightarrow \quad \begin{array}{c} x_1 \in \mathbb{R} \text{ is} \\ \text{arbitrary} \end{array}.$$

Again we take $x_1 = 1$ and so $\mathbf{x}_2 = (1, 1, 0, 4)^\top$. Continuing this same line of reasoning, we seek $\mathbf{x}_3 := (x_1, x_2, x_3, x_4)^\top \neq \mathbf{0_n}$ for which $(2\mathbf{I}_4 - \mathbf{A})\mathbf{x}_3 = \mathbf{x}_2$, in which case it follows that $\mathcal{N}(2\mathbf{I}_4 - \mathbf{A})^3 \neq \mathbb{R}^4$ since $\mathbf{x}_3 \notin \mathcal{N}(2\mathbf{I}_4 - \mathbf{A})^2$. Again we check:

$$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 4 \end{pmatrix} = \mathbf{x}_2 = (2\mathbf{I}_4 - \mathbf{A})\mathbf{x}_3 = \begin{pmatrix} 2 & -1 & 0 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 2 & -1 \\ 16 & -32 & 24 & -6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ 2x_2 - x_3 \\ 2x_3 - x_4 \\ 16x_1 - 32x_2 + 24x_3 - 6x_4 \end{pmatrix}$$

$$\Leftrightarrow \left\{ \begin{array}{c} x_2 = -1 + 2x_1 \\ x_3 = -1 + 2x_2 = -3 + 4x_1 \\ x_4 = -2x_3 = 6 - 8x_1 \\ (16 - 64 + 96 + 48)x_1 = 4 + 32 + 24 + 36 \end{array} \right\} \Leftrightarrow \quad \begin{array}{c} x_1 \in \mathbb{R} \text{ is} \\ \text{arbitrary} \end{array}.$$

So the minimal polynomial coincides with the characteristic polynomial in this example. Note also that there is only one eigenvector $\mathbf{x}_1$, and hence the geometric multiplicity is one. If $\mathbf{U} \in \mathbb{R}^{4\times 4}$ has its columns equal to $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$, then one can check that $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{A}_5$, where $\mathbf{A}_5$ is as defined above. One need not compute $\mathbf{U}^{-1}$ to see this, but rather just observe $\mathbf{A}\mathbf{x}_1 = 2\mathbf{x}_1$, $\mathbf{A}\mathbf{x}_2 = 2\mathbf{x}_2 - \mathbf{x}_1$, $\mathbf{A}\mathbf{x}_3 = 2\mathbf{x}_3 - \mathbf{x}_2$, and $\mathbf{A}\mathbf{x}_4 = 2\mathbf{x}_4 - \mathbf{x}_3$. These vector multiplications can be condensed into one matrix multiplication $\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{A}_5$, which is equivalent to $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{A}_5$.

1.7.5 Vandermonde matrices

Given a finite sequence $\lambda_1, \ldots, \lambda_n$ of real numbers, a Vandermonde matrix has the form

$$\mathbf{V} = \begin{pmatrix} 1 & \ldots & 1 \\ \lambda_1 & \ldots & \lambda_n \\ \lambda_1^2 & \ldots & \lambda_n^2 \\ \vdots & \vdots & \vdots \\ \lambda_1^{m-1} & \ldots & \lambda_n^{m-1} \end{pmatrix}.$$

When $m = n$, the determinant of $\mathbf{V}$ turns out to satisfy

$$\det(\mathbf{V}) = \prod_{1\le i<j\le n} (\lambda_j - \lambda_i) \tag{17}$$

Obviously then a square Vandermonde matrix is invertible precisely when the λ_i's are distinct. Moreover, in this case, if $p(\lambda) := \prod_{i=1}^{n}(\lambda - \lambda_i)$ is the polynomial with (distinct) roots λ_i and $\mathbf{A}$ is the companion matrix associated with $p(\cdot)$, then $\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{D}$, where $\mathbf{D}$ is the diagonal matrix with $\lambda_1, \ldots, \lambda_n$ along the diagonal. The order of the λ_i's aligns with the ordering of the λ_i's that appear in $\mathbf{V}$. Again, one need not compute $\mathbf{V}^{-1}$ to show this, but the rather easy verification

$$\mathbf{AV} = \mathbf{VD} = \begin{pmatrix} \lambda_1 & \ldots & \lambda_n \\ \lambda_1^2 & \ldots & \lambda_n^2 \\ \lambda_1^3 & \ldots & \lambda_n^3 \\ \vdots & \vdots & \vdots \\ \lambda_1^n & \ldots & \lambda_n^n \end{pmatrix}.$$

does the same thing.

1.8 Similar matrices

As mentioned earlier, two $n \times n$ matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are *similar* if there exists an invertible map $\mathbf{U} \in \mathbb{R}^{n\times n}$ so that $\mathbf{A}_2 = \mathbf{U}\,\mathbf{A}_1\,\mathbf{U}^{-1}$ or equivalently, $\mathbf{A}_2\mathbf{U} = \mathbf{U}\mathbf{A}_1$. We write $\mathbf{A}_1 \sim \mathbf{A}_2$ if they are similar, and it turns out the concept of being similar is an equivalence relation: one has

a. $\mathbf{A} \sim \mathbf{A} \quad \forall \mathbf{A} \in \mathbb{R}^{n\times n}$,

b. $\mathbf{A}_1 \sim \mathbf{A}_2 \;\Rightarrow\; \mathbf{A}_2 \sim \mathbf{A}_1 \quad \forall \mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{n\times n}$, and

c. $\mathbf{A}_1 \sim \mathbf{A}_2$ and $\mathbf{A}_2 \sim \mathbf{A}_3 \;\Rightarrow\; \mathbf{A}_1 \sim \mathbf{A}_3 \quad \forall \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \in \mathbb{R}^{n\times n}$.

The property $\mathbf{A}_1 \sim \mathbf{A}_2$ says that $\mathbf{A}_1$ and $\mathbf{A}_2$ operate as linear maps in the same manner up to a change in basis. The situation can be summarized by the following equivalent commutative diagrams

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "\mathbf{A}_1"] & \mathbb{R}^n \arrow[d, "\mathbf{U}"] \\ \mathbb{R}^n \arrow[u, "\mathbf{U}^{-1}"] \arrow[r, "\mathbf{A}_2"'] & \mathbb{R}^n \end{tikzcd} \qquad \begin{tikzcd} \mathbb{R}^n \arrow[r, "\mathbf{A}_1"] \arrow[d, "\mathbf{U}"'] & \mathbb{R}^n \arrow[d, "\mathbf{U}"] \\ \mathbb{R}^n \arrow[r, "\mathbf{A}_2"'] & \mathbb{R}^n \end{tikzcd}$$

Let $\mathcal{B}$ be the n vectors that are the columns of $\mathbf{U}^{-1}$. Proposition1.1 says $\mathcal{B}$ is a basis if and only if $\mathbf{U}$ is invertible. Starting with $\mathbf{x} \in \mathbb{R}^n$ in the bottom left corner, we can think of $\mathbf{x}$ as the coordinate vector with respect to the canonical basis. Calculating $\mathbf{U}^{-1}\mathbf{x}$ produces a new vector in $\mathbb{R}^n$ whose coordinates with respect to $\mathcal{B}$ is $\mathbf{x}$. Next, multiplying by $\mathbf{A}_1$ is the result of the linear map acting on these coordinates producing $\mathbf{A}_1\mathbf{U}^{-1}\mathbf{x}$. Finally, applying $\mathbf{U}$ to the result transforms the $\mathcal{B}$-coordinates back to the canonical coordinates. The similarity property $\mathbf{A}_2 = \mathbf{U}\mathbf{A}_1\mathbf{U}^{-1}$ is thus what we said above: the maps $\mathbf{A}_1$ and $\mathbf{A}_2$ are the "same" modulo a change of coordinates.

Two similar matrices have the same characteristic and minimal polynomials, and each eigenvalue has the same algebraic and geometric multiplicity. But these properties are not by themselves sufficient to characterize the property of being similar. The main result can be summarized by saying that two matrices are similar if and only if they can be transformed into the same Jordan canonical form (modulo the order of the blocks). Essentially this means that each equivalence class contains a special type of matrix that can be described in "block" form, which we explain next.

1.8.1 The Jordan form of a square matrix

The *nilpotent* matrix $\mathbf{N}_n$ (of size $n \in \mathbb{N}$) has the form

$$\mathbf{N}_n = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

That is, $\mathbf{N}_n \in \mathbb{R}^{n\times n}$ has zeroes everywhere except along the "superdiagonal" where it has all ones. One can check that $\mathbf{N}_n^{n-1}$ has only one non-zero element that is a one in the top right corner, and $\mathbf{N}_n^n = \mathbf{0}_{n\times n}$.

Let $\lambda \in \mathbb{R}$ and $n \in \mathbb{N}$. An *elementary Jordan block* $\mathbf{J}_\lambda$ of size n with eigenvalue λ has the form

$$\mathbf{J}_{\lambda,n} = \begin{pmatrix} \lambda & 1 & 0 & 0 & \dots & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 \\ 0 & 0 & \lambda & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & 0 & 0 & \lambda \end{pmatrix} = \lambda\mathbf{I_n} + \mathbf{N}_n,$$

Note that $\mathbf{J}_{\lambda,n}$ consists of the $n \times n$ matrix which has all λ's along the diagonal and 1's along the super diagonal.

A *Jordan block* is a matrix $\mathbf{J} \in \mathbb{R}^{n\times n}$ with $\lambda \in \mathbb{R}$ along the diagonal , only 0's and 1's along the superdiagonal, and 0's elsewhere. If the geometric multiplicity of λ is $k \leq n$, then there exists k independent eigenvectors which gives rise to k elementary Jordan blocks $\mathbf{J}_{\lambda,n_i}$, where $n_i \in \mathbb{N}$ with $n_1 + \dots + n_k = n$. That is, a Jordan block has the form

$$\mathbf{J}_\lambda = \begin{pmatrix} \mathbf{J}_{\lambda,n_1} & \mathbf{0}_{n_1,n_2} & \mathbf{0}_{n_1,n_3} & \cdots & \mathbf{0}_{n_1,n_k} \\ \mathbf{0}_{n_2,n_1} & \mathbf{J}_{\lambda,n_2} & \mathbf{0}_{n_2,n_3} & \cdots & \mathbf{0}_{n_2,n_k} \\ \mathbf{0}_{n_3,n_1} & \mathbf{0}_{n_3,n_2} & \mathbf{J}_{\lambda,n_3} & \cdots & \mathbf{0}_{n_3,n_k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_k,n_1} & \mathbf{0}_{n_k,n_2} & \mathbf{0}_{n_k,n_3} & \cdots & \mathbf{J}_{\lambda,n_k} \end{pmatrix}$$

where each $\mathbf{J}_{\lambda,n_i}$ is an elementary Jordan block. A simpler and less messy way to display this is

$$\mathbf{J}_\lambda = \begin{pmatrix} \boxed{\mathbf{J}_{\lambda,n_1}} & & & & \\ & \boxed{\mathbf{J}_{\lambda,n_2}} & & \mathbf{0} & \\ & & \ddots & & \\ & \mathbf{0} & & \boxed{\mathbf{J}_{\lambda,n_{k-1}}} & \\ & & & & \boxed{\mathbf{J}_{\lambda,n_k}} \end{pmatrix}$$

where it is understood that zeroes are in all the positions that are not within the boxes. One should note the minimal polynomial is $p(\alpha) = (\alpha - \lambda)^{\bar{n}}$, where $\bar{n} = \max\{n_1, \ldots, n_k\}$.

Finally, a matrix $\mathbf{J}$ is in *Jordan form* if it looks like

$$\mathbf{J} = \begin{pmatrix} \boxed{\mathbf{J}_{\lambda_1}} & & & & \\ & \boxed{\mathbf{J}_{\lambda_2}} & & \mathbf{0} & \\ & & \ddots & & \\ & \mathbf{0} & & \boxed{\mathbf{J}_{\lambda_{k-1}}} & \\ & & & & \boxed{\mathbf{J}_{\lambda_k}} \end{pmatrix}$$

where each $\mathbf{J}_{\lambda_i}$ is a Jordan block and $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues.

Some theorems are relatively easy to understand because they are easy to prove. Others are easy to understand because the statement is simple, but the proof seems inaccessible. The following is one of the latter, and is an amazing and beautiful result. The proof of the general result is indeed complicated and we shall not do it here.

Theorem 1.2. *Every square matrix with real eigenvalues is similar to a matrix in Jordan form. Moreover, the Jordan form is unique up to the order that the elementary Jordan blocks appear.*

We can finally characterize the similarity property: Two matrices (with only real eigenvalues) are similar if they have the same Jordan form.

1.8.2 Complex eigenvalues

There is a Jordan form where complex eigenvalues appear as well. Essentially everything done above for $\mathbb{R}^n$ can be done for $\mathbb{C}^n$, or in fact for $\mathbf{F}^n$ where $\mathbf{F}$ is any field. Such generality assumes $\mathbf{F}^n$ is a vector space over the field $\mathbf{F}$ and every eigenvalue belongs to $\mathbf{F}$. With $\mathbf{F} = \mathbb{C}$, the latter property is automatically true, thanks to the celebrated Fundamental Theorem of Algebra that says every n^{th} degree polynomial with $\mathbb{C}$ coefficients has precisely n roots including multiplicity. Thus the above progression to the Jordan form for $\mathbb{C}^n$ covers all cases and needs no further comment. However, if the original matrix has only $\mathbb{R}$ values, it is generally desirable to have a Jordan form with only real entries. Recall that if $\mathbf{A}$ has only real entries and $\lambda = a + ib$ is a truly complex eigenvalue (the adjective *truly* refers to $b \neq 0$), then $\overline{\lambda} = a - ib$ is also an eigenvalue, and thus both Jordan blocks J_λ and $J_{\overline{\lambda}}$ appear in the complex version of the Jordan form. There is redundancy, but this redundancy of a two complex dimensional subspace (and hence lying in a four dimensional real subspace) can be replaced by a two dimensional real subspace that is embedded in a four dimensional real subspace.

We try to explain this further, but first consider the simple example $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The eigenvalues are $\pm i$, and the (Complex) Jordan form is

$$\mathbf{J} := \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \text{and } \mathbf{A} = \mathbf{U}\mathbf{J}\mathbf{U}^{-1} \text{ where } \mathbf{U} = \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}.$$

There are two complex Jordan blocks of size one, and of course the eigenvalues are conjugates since $\mathbf{A}$ has only real entries. The invertible matrix $\mathbf{U}$ has to be truly complex since its columns are the eigenvectors associated with the complex eigenvalues. The issue is to give a Jordan form with only real matrix data. In this example, the matrix $\mathbf{A}$ itself is actually in (real) Jordan form associated to the complex eigenvalues $\pm i$.

We can think of $\mathbb{C}^n$ as a vector space over $\mathbb{C}$ (with dimension n) or over $\mathbb{R}$ (with dimension $2n$), and it is the latter we are now interested in since our matrix $\mathbf{A}$ has only real entries. If there exists a truly complex eigenvalue $\lambda = a+ib$ with associated eigenvector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ ($\mathbf{x} \in \mathbb{R}^n$, $\mathbf{0} \neq \mathbf{y} \in \mathbb{R}^n$), then its conjugate $\overline{\lambda} = a - ib$ is also an eigenvalue with eigenvector $\overline{\mathbf{z}} = \mathbf{x} - i\mathbf{y}$. We think of $\mathbb{C}^n$ (as a vector space over $\mathbb{R}$) first in the form $\mathbb{R}^n + i\mathbb{R}^n \simeq \mathbb{R}^n \bigoplus \mathbb{R}^n \simeq \mathbb{R}^{2n} \simeq (\mathbb{R}^2)^n$, but since it is over $\mathbb{R}$, we are really only interested in a certain n dimensional (real) subspace of $\mathbb{C}^n$.

In the simplest case of a single complex eigenvalue λ, the complex space containing it must be large enough to contain both λ and $\overline{\lambda}$. We identify an element $\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} = \begin{pmatrix} a_1 + ib_1 \\ a_2 + ib_2 \end{pmatrix} \in \mathbb{C}^2$ with the 2×2 matrix $\begin{pmatrix} a_1 & b_2 \\ b_1 & a_2 \end{pmatrix}$. More precisely, define the map $T : \mathbb{C}^2 \to \mathbb{R}^{2\times 2}$ by

$$T \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} a_1 & b_2 \\ b_1 & a_2 \end{pmatrix},$$

which is an isomorphism of the two real vector spaces $\mathbb{C}^2$ and $\mathbb{R}^{2\times 2}$. Notice if $z = a + ib$, then

$$T \begin{pmatrix} z \\ \overline{z} \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

and 2×2 matrices in the latter form are capturing all the information in $\mathbb{R}^{2\times 2}$ that $(z, \overline{z})$ provide in $\mathbb{C}^2$. The Jordan form with a complex eigenvalue looks exactly the same as in the real case, except that the real eigenvalue is replace by a 2×2 matrix that represents both the complex eigenvalue and its conjugate. An elementary Jordan block with truly complex eigenvalue $\lambda = a + ib$ of size n is the $2n \times 2n$ matrix of the form

$$\mathbf{J}_{\lambda,n} = \begin{pmatrix} a & -b & 1 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ b & a & 0 & 1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & a & -b & 1 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & b & a & 0 & 1 & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \dots & \dots & a & -b & 1 & 0 \\ 0 & 0 & \dots & \dots & \dots & \dots & b & a & 0 & 1 \\ 0 & 0 & \dots & \dots & \dots & \dots & 0 & 0 & a & -b \\ 0 & 0 & \dots & \dots & \dots & \dots & 0 & 0 & b & a \end{pmatrix}$$

Notice the diagonal blocks correspond to λ, and the dashed superdiagonal blocks are 2×2 identity matrices.

Jordan blocks and the Jordan form are defined in the complex case as they were for the real case. Notice, however, that with a truly complex eigenvalue λ of a matrix $\mathbf{A}$ with only real entries, there are two types of Jordan forms. The complex one uses both $n \times n$ blocks $\mathbf{J}_\lambda$ and $\mathbf{J}_{\overline{\lambda}}$, while the real one uses one of these (it doesn't matter which one) in the real form as above (which is of dimension $2n$).

1.9 Symmetric matrices

A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric if $\mathbf{A} = \mathbf{A}^\top$. Symmetric matrices occur more naturally than might at first be realized. For example, the Hessian matrix of a C^2 function is symmetric. Also, if $\mathbf{A} \in \mathbb{R}^{m\times n}$, then $\mathbf{A}\mathbf{A}^\top \in \mathbb{R}^{m\times m}$ and $\mathbf{A}^\top\mathbf{A} \in \mathbb{R}^{n\times n}$ are always both symmetric.

Proposition 1.5. *Suppose* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *is symmetric. Then*

(i) $\mathbf{A}$ *has only real eigenvalues;*

(ii) Any two eigenvectors associated with different eigenvalues are orthogonal; and

(iii) $\mathbf{A}$ *is diagonalizable, and there exists an orthonormal eigenbasis. This is equivalent to saying there is a unitary matrix* $\mathbf{V}$ *and a diagonal matrix* $\mathbf{D}$ *so that* $\mathbf{A} = \mathbf{V}\ \mathbf{D}\ \mathbf{V}^\top$. *The diagonal entries of* $\mathbf{D}$ *are the eigenvalues of* $\mathbf{A}$ *whose associated eigenvectors are the columns of* $\mathbf{V}$.

Proof. (i) Suppose $\lambda = \alpha + i\beta$ is a (possibly complex) eigenvalue with associated eigenvector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$. On the one hand, we have

$$\overline{\mathbf{z}}^\top \mathbf{A}\,\mathbf{z} = \overline{\mathbf{z}}^\top \lambda\,\mathbf{z} = \lambda\,\overline{\mathbf{z}}^\top \mathbf{z}.$$

We have noted above that $\overline{\mathbf{z}} = \mathbf{x} - i\mathbf{y}$ is an eigenvector associated with $\overline{\lambda} = \alpha - i\beta$, so on the other hand (and since $\mathbf{A} = \mathbf{A}^\top$), we have

$$\overline{\mathbf{z}}^\top \mathbf{A}\,\mathbf{z} = \left(\mathbf{A}^\top \overline{\mathbf{z}}\right)^\top \mathbf{z} = \left(\mathbf{A}\,\overline{\mathbf{z}}\right)^\top \mathbf{z} = \overline{\lambda}\,\overline{\mathbf{z}}^\top \mathbf{z} = \overline{\lambda}\,\overline{\mathbf{z}}^\top \mathbf{z}.$$

Since these two are the same, we have $0 = (\lambda - \overline{\lambda})\overline{\mathbf{z}}^\top \mathbf{z} = i2\,\beta\,\overline{\mathbf{z}}^\top \mathbf{z}$. Since $\overline{\mathbf{z}}^\top \mathbf{z} > 0$, we conclude $\beta = 0$, or that λ is real.
(ii) Suppose $\lambda_1 \neq \lambda_2$ be eigenvalues with associated eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$. We have

$$\lambda_1\langle \mathbf{v}_1, \mathbf{v}_2\rangle = \langle \lambda_1\mathbf{v}_1, \mathbf{v}_2\rangle = \langle \mathbf{A}\mathbf{v}_1, \mathbf{v}_2\rangle = \langle \mathbf{v}_1, \mathbf{A}\mathbf{v}_2\rangle = \langle \mathbf{v}_1, \lambda_2\mathbf{v}_2\rangle = \lambda_2\langle \mathbf{v}_1, \mathbf{v}_2\rangle.$$

Since $\lambda_1 \neq \lambda_2$, we must have $\langle \mathbf{v}_1, \mathbf{v}_2\rangle = 0$, and so $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.
(iii) This is closely related to Theorem 1.2, but has additional information regarding the type of similarity. We prove the result by induction on n. Let λ be an eigenvalue of $\mathbf{A}$ with an associated unit eigenvector $\mathbf{v}_1$. We can choose vectors $\mathbf{v}_2, \dots, \mathbf{v}_n$ so that $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is an orthonormal basis, and the matrix $\mathbf{V}$ whose i^{th} column is $\mathbf{v}_i$ is a unitary matrix (i.e. satisfies $\mathbf{V}^{-1} = \mathbf{V}^\top$). We have

$$\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \left(\begin{array}{c|ccc} \lambda & \alpha_2 & \dots & \alpha_n \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_1 & \\ 0 & & & \end{array}\right) \tag{18}$$

where $\alpha_2, \dots, \alpha_n$ are some scalars and $\mathbf{A}_1 \in \mathbb{R}^{(n-1)\times(n-1)}$. Recall $\mathbf{A} = \mathbf{A}^\top$ and $\mathbf{V}^\top = \mathbf{V}^{-1}$, and therefore $(\mathbf{V}^{-1}\mathbf{A}\mathbf{V})^\top = \mathbf{V}^\top\mathbf{A}^\top(\mathbf{V}^{-1})^\top = \mathbf{V}^{-1}\mathbf{A}\mathbf{V}$. Thus the matrix in (18) is symmetric, from whence it follows $\alpha_2 = \cdots = \alpha_n = 0$ and $\mathbf{A}_1$ is also symmetric. Applying the induction hypothesis to $\mathbf{A}_1$ completes the proof. $\square$

A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is called positive semi-definite if $\langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, and strictly positive definite if the inequality is strict for all $\mathbf{x} \neq \mathbf{0_n}$.

Proposition 1.6. *Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric. Then $\mathbf{A}$ is positive semi-definite (resp. strictly positive definite) if and only if all its eigenvalues are nonnegative (resp. strictly positive).*

2 Topological structure in $\mathbb{R}^n$

The vector space structure underpins the representation and manipulation of elements in R^n. The topological structure provides the framework to talk about convergence. The basic topological concepts reviewed here have more general formulations, but we are only interested in how they pertain in conjunction with the vector space structure.

2.1 Geometric form of the inner product and norms

The inner product $\langle \mathbf{x}, \mathbf{y}\rangle$ of $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$ was defined earlier in (3), which we call the algebraic definition since it relies on the explicit component representation of the vectors $\mathbf{x}$ and $\mathbf{y}$. The *norm* of a vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} = \sqrt{\sum_{j=1}^{n} x_j^2}, \tag{19}$$

and has a geometric character since it measures "how far" the vector $\mathbf{x}$ is from the origin $\mathbf{0_n}$. The inner product also equals

$$\langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{x}\|\,\|\mathbf{y}\|\,\cos(\theta), \tag{20}$$

where θ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. This is called the geometric form of the inner product. We next justify the equivalence of the algebraic form (3) and geometric form (20). If $\mathbf{y} = \mathbf{e}_1$, then $x_1 = \sum_{i=1}^{n} x_i\, y_i$. In this

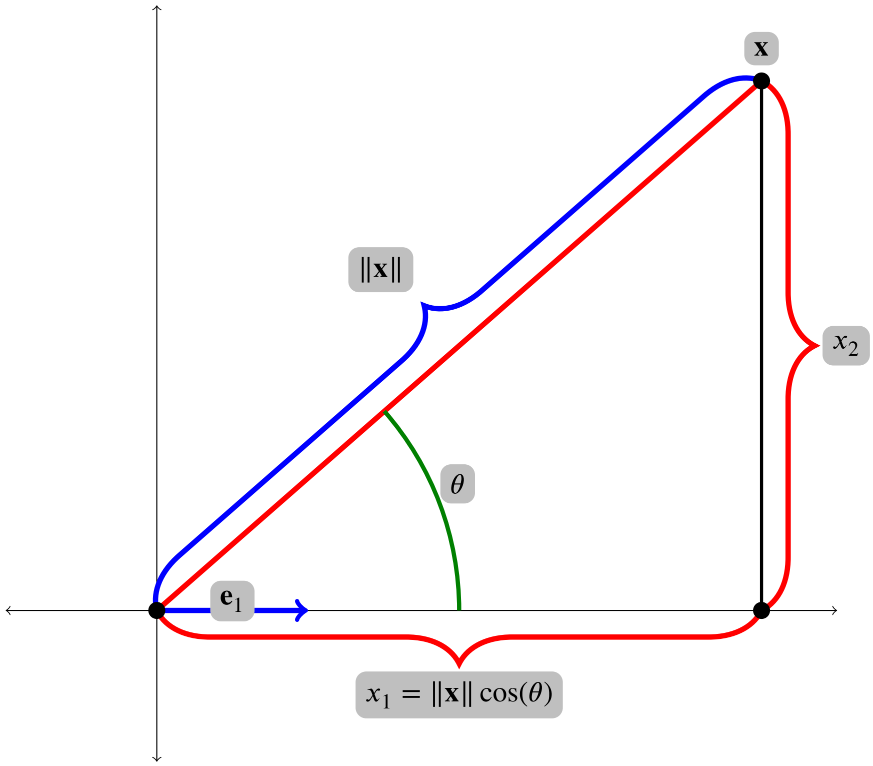


Figure 1: Geometric form of inner product

case, it is clear (20) holds, as is demonstrated in Figure 1 in dimension two: For general $\mathbf{y} \in S^1$, one can find a unitary map $\mathbf{U} : \mathbb{R}^n \to \mathbb{R}^n$ that takes $\mathbf{y}$ to $\mathbf{e}_1$. To see this, choose any orthonormal basis with $\mathbf{y}$ as its first element (see Proposition 1.2), and let $\mathbf{U}$ be the matrix whose columns consist of this basis. Then $\mathbf{U}\mathbf{y} = \mathbf{e}_1$. We have seen that unitary maps preserve the inner product (see (8)), and so the general situation essentially reduces to the case $n = 2$.

Another way to verify (20) is to use the Law of Cosines, which says $\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\,\|\mathbf{y}\|\,\cos(\theta)$. One also has $\|\mathbf{x}-\mathbf{y}\|^2 = \langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\langle \mathbf{x}, \mathbf{y}\rangle$. Equating these two ways of writing $\|\mathbf{x}-\mathbf{y}\|^2$ and rearranging terms gives (20).

2.1.1 Other inner products

Observe for all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in \mathbb{R}^n$ and $r \in \mathbb{R}$, that $r\langle \mathbf{x}, \mathbf{y}\rangle = \langle r\,\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, r\,\mathbf{y}\rangle = r\langle \mathbf{y}, \mathbf{x}\rangle$ and $\langle \mathbf{x}+\mathbf{z}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{z}, \mathbf{y}\rangle$. These are the defining properties of an inner product, and there are in fact many different inner products that can be defined on $\mathbb{R}^n$. For example, if $\mathbf{A} \in \mathbb{R}^n$ is any strictly positive definite matrix, then $\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbf{A}} := \langle \mathbf{x}, \mathbf{A}\mathbf{y}\rangle$ defines another inner product; Any inner product gives rise to another norm. For simplicity we will only refer to the usual usual inner product (which is the case with $\mathbf{A} = \mathbf{I_n}$).

2.1.2 Other norms

The norm $\|\cdot\|$ satisfies (i) $\|\mathbf{x}\| \geq 0\ \forall \mathbf{x} \in \mathbb{R}^n$, and equals 0 if and only if $\mathbf{x} = \mathbf{0_n}$; (ii) $\|r\,\mathbf{x}\| = |r|\,\|\mathbf{x}\|\ \forall \mathbf{x} \in \mathbb{R}^n$, $\forall r \in \mathbb{R}$, and (iii) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|\ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. These properties are usually presented as the axioms of a norm (and not just for the particular norm we are using). If we need to be more specific, we denote the norm defined in (19) by

$\|\mathbf{x}\|_2$. For any $1 \le p < p$, the p-norm is given by

$$\|\mathbf{x}\|_p := \sqrt[p]{\sum_{j=1}^{n} x_j^p}, \tag{21}$$

and when $p = \infty$, one defines the ∞-norm by $\|\mathbf{x}\|_\infty = \sup\left\{|x_j| \,:\, 1 \le j \le n\right\}$. Although each such p gives a different norm, the only such norm that agrees with the norm defined through some inner product is the case $p = 2$. Unless explicitly stated or denoted otherwise, we shall use the 2-norm.

Since $|\cos(\theta)| \le 1$ for any angle θ, (20) implies the *Cauchy-Schwarz* inequality holds:

$$\langle \mathbf{x}, \mathbf{y} \rangle \le \|\mathbf{x}\|_2 \, \|\mathbf{y}\|_2 \quad \forall \mathbf{x},\, \mathbf{y} \in \mathbb{R}^n, \tag{22}$$

and equality holds in this inequality precisely when $\theta = 0$, or that there exists a constant $\rho \ge 0$ with $\mathbf{x} = \rho\mathbf{y}$. It is also clear from the geometric form (20) that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ if and only if the vectors $\mathbf{x}$ and $\mathbf{y}$ are perpendicular to each other.

Actually (22) is a special case of what is called Hölder's inequality. This states that if $1 \le p,\, q \le \infty$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\langle \mathbf{x}, \mathbf{y} \rangle \le \|\mathbf{x}\|_p \, \|\mathbf{y}\|_q \quad \forall \mathbf{x},\, \mathbf{y} \in \mathbb{R}^n,$$

but we won't use that much here.

With $\mathbf{u} \in S^1$ and $\mathbf{x} \in \mathbb{R}^n$, the projection (or component) of $\mathbf{x}$ in direction $\mathbf{u}$ is $\|\mathbf{x}\| \cos(\theta)\mathbf{u} = \langle \mathbf{x}, \mathbf{u} \rangle \mathbf{u}$. The Gram-Schmidt procedure can now be explained geometrically. By starting with a basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, one first normalizes $\mathbf{v}_1$ by setting $\tilde{\mathbf{v}}_1 := \mathbf{v}_1/\|\mathbf{v}_1\|$. Then consider $\mathbf{v}_2$, and let $\mathbf{u}_2$ be $\mathbf{v}_2$ minus its $\tilde{\mathbf{v}}_1$-component, and obtain $\tilde{\mathbf{v}}_2$ by normalizing $\mathbf{u}_2$. Each step brings in a new vector $\mathbf{v}_k$ independent of the previous ones, and one subtracts all of its components in the previous directions to get $\mathbf{u}_k$, and finally, $\tilde{\mathbf{v}}_k$ is the normalization of $\mathbf{u}_k$.

2.1.3 The Operator Norm

Suppose $\mathbf{A} = \left(a_{ij}\right) \in \mathbb{R}^{m\times n}$. Then $\mathbf{A}$ as a map from $\mathbb{R}^n$ to $\mathbb{R}^m$ is an example of a *bounded* operator because there exists a constant $M \ge 0$ so that

$$\left\|\mathbf{A}\mathbf{x}\right\| \le M \, \|\mathbf{x}\| \quad \forall \mathbf{x} \in \mathbb{R}^n. \tag{23}$$

For example, if $M = \sup\left\{|a_{ij}| \,:\, 1 \le i \le m,\ 1 \le j \le n\right\}$ (called the sup-norm of $\mathbf{A}$), or if $M = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$ (also called the Fröbenius norm of $\mathbf{A}$), then (23) holds. These two bounds can be verified by considering $\mathbf{A}$ as an element in $\mathbb{R}^{mn}$, and as such do not reflect the operator-character of $\mathbf{A}$ as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. The smallest M that satisfies (23) is called the *operator norm* of $\mathbf{A}$, and is denoted by $\left\|\mathbf{A}\right\|$. There are only so many symbols to use, and recycling symbols keeps notation from getting out of hand. Another way to write this is

$$\left\|\mathbf{A}\right\| = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\left\|\mathbf{A}\mathbf{x}\right\|}{\|\mathbf{x}\|} = \sup_{\|\mathbf{x}\|=1} \left\|\mathbf{A}\mathbf{x}\right\|.$$

Here we have the same notation $\|\cdot\|$ being used in three different ways in the same line! But, in each instance it is a norm. The norms being used in $\mathbb{R}^n$ and/or $\mathbb{R}^m$ affect the operator norm, and when this is of import, one writes (where $1 \le p,\, q \le \infty$)

$$\left\|\mathbf{A}\right\|_{p,q} = \inf_{\|x\|_p=1} \left\|\mathbf{A}\mathbf{x}\right\|_q.$$

We generally use only the 2-norm in both $\mathbb{R}^n$ and $\mathbb{R}^m$, and keep the notation simple by writing $\left\|\mathbf{A}\right\|$ for $\left\|\mathbf{A}\right\|_{2,2}$.

Proposition 2.1. *The operator norm has the following properties:*

(a) $\left\|\mathbf{A}\right\| \geq 0 \quad \forall \mathbf{A} \in \mathbb{R}^{m\times n}$*, and* $= 0$ *if and only if* $\mathbf{A} = \mathbf{0}_{m\times n}$*;*

(b) $\left\|r\mathbf{A}\right\| = |r| \left\|\mathbf{A}\right\| \quad \forall \mathbf{A} \in \mathbb{R}^{m\times n},\ r \in \mathbb{R}$*;*

(c) $\left\|\mathbf{A}_1 + \mathbf{A}_2\right\| \leq \left\|\mathbf{A}_1\right\| + \left\|\mathbf{A}_2\right\| \quad \forall \mathbf{A}_1,\ \mathbf{A}_2 \in \mathbb{R}^{m\times n}$*;*

(d) $\left\|\mathbf{A}\,\mathbf{B}\right\| \leq \left\|\mathbf{A}\right\| \left\|\mathbf{B}\right\| \quad \forall \mathbf{A} \in \mathbb{R}^{m\times n},\ \mathbf{B} \in \mathbb{R}_{n\times p}$*;*

(e) $\left\|\mathbf{A}\right\| = \left\|\mathbf{A}^\top\right\| \quad \forall \mathbf{A} \in \mathbb{R}^{m\times n}$*;*

(f) $\left\|\mathbf{A}\right\| = \max\left\{ \sqrt{\lambda} \,:\, \lambda \in \sigma\left(\mathbf{A}^\top\mathbf{A}\right) \right\} \quad \forall \mathbf{A} \in \mathbb{R}^{m\times n}$*;*

(g) If $\mathbf{A} \in \mathbb{R}^{n\times n}$ *is diagonalizable, then* $\left\|\mathbf{A}\right\| = \max\left\{|\lambda| \,:\, \lambda \in \sigma(\mathbf{A})\right\}$.

Proof. (f) Suppose $\mathbf{A} \in \mathbb{R}^{m\times n}$, and note that $\mathbf{A}^\top\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric and positive semidefinite. So $\mathbf{A}^\top\mathbf{A}$ is diagonalizable by Proposition 1.5, has nonnegative eigenvalues by Proposition 1.6, and there is an orthonormal eigenbasis $\mathcal{B} = \left\{\mathbf{v}_1, \ldots, \mathbf{v}_n\right\}$ with associated eigenvalues $\sigma(\mathbf{A}^\top\mathbf{A}) = \{\lambda_1, \ldots, \lambda_n\}$. If we let $\lambda^* := \max\left\{\sqrt{\lambda} \,:\, \lambda \in \sigma(\mathbf{A}^\top\mathbf{A})\right\}$, then we need to show $\left\|\mathbf{A}\right\|$ equals $\sqrt{\lambda^*}$.

For any $\mathbf{x} \in \mathbb{R}^n$, we can write $\mathbf{x} = \displaystyle\sum_{j=1}^{n} \alpha_j \mathbf{v}_j$ where the α's are the coordinates of $\mathbf{x}$ with respect to $\mathcal{B}$. Since $\mathcal{B}$ is orthogonal, one has $\|\mathbf{x}\|^2 = \sum_{i=1}^n \alpha_i^2$. Then

$$\left\|\mathbf{A}\mathbf{x}\right\|^2 = \left\langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}\right\rangle = \left\langle \mathbf{x}, \mathbf{A}^\top\mathbf{A}\mathbf{x}\right\rangle = \left\langle \sum_{i=1}^{n} \alpha_i \mathbf{v}_i, \sum_{i=1}^{n} \alpha_i \mathbf{A}^\top\mathbf{A}\mathbf{v}_i \right\rangle = \left\langle \sum_{i=1}^{n} \alpha_i \mathbf{v}_i, \sum_{i=1}^{n} \alpha_i \lambda_i \mathbf{v}_i \right\rangle = \sum_{i=1}^{n} \lambda_i \left|\alpha_i\right|^2 \leq \lambda^* \left\|\mathbf{x}\right\|^2.$$

This implies $\left\|\mathbf{A}\right\| \leq \sqrt{\lambda^*}$. On the other hand, there exists an index i_0 with $\lambda_{i_0} = \lambda^*$. In a calculation even simpler than the previous one, one has $\left\|\mathbf{A}\mathbf{v}_{i_0}\right\|^2 = \lambda^* \|\mathbf{v}_{i_0}\|^2$, and hence $\sqrt{\lambda^*} \leq \left\|\mathbf{A}\right\|$. □

2.1.4 The exponential of a square matrix

We have seen that for a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ and a polynomial $p(\cdot)$, one can define a new matrix $p(\mathbf{A})$. The next (and natural) step is to inquire what it would mean to let the degree of the polynomial become very large, and whether another matrix would accrue by passing to the limit. Perhaps the most important (because of its application to differential equations) illustration of this general idea is to consider the exponential function e^x. Of course one *must* remember

$$e^x = \sum_{i=0}^{\infty} \frac{1}{i!} x^i = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \cdots = \lim_{k\to\infty} \left\{ \sum_{i=0}^{k} \frac{1}{i!} x^i \right\} = \lim_{k\to\infty} \left\{ 1 + x + \frac{1}{2}x^2 + \ldots \frac{1}{k!}x^k \right\}$$

Each of the terms $\mathbf{A}^i$ is well-defined and we now have a mechanism of passing to a limit of matrices. Hence the exponential of a matrix can be defined exactly as it is for numbers:

$$\mathbf{exp}\big(\mathbf{A}\big) := \sum_{i=0}^{\infty} \frac{1}{i!} \mathbf{A}^i = 1 + \mathbf{A} + \frac{1}{2}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \cdots = \lim_{k\to\infty} \left\{ \sum_{i=0}^{k} \frac{1}{i!} \mathbf{A}^i \right\} = \lim_{k\to\infty} \left\{ 1 + \mathbf{A} + \frac{1}{2}\mathbf{A}^2 + \ldots \frac{1}{k!}\mathbf{A}^k \right\} \tag{24}$$

One can obviously go further. Any power series $s(x) = \sum_{i=1}^{\infty} \alpha_i x^i$ that absolutely converges for all $x \in \mathbb{R}^n$, one can replace x by any matrix $\mathbf{A}$ and get a new matrix $\mathbf{s}(\mathbf{A}) := \sum_{i=1}^{\infty} \alpha_i \mathbf{A}^i$.

2.1.5 Calculating the exponential matrix

It is not easy to calculate an infinite series of numbers let alone that of a matrix. This is true even for the series defining an exponential, however one can do so easily provided the Jordan form can be found. We illustrate this in a number of steps:

Step 1: The exponential of $\lambda \mathbf{I_n}$ is $\mathbf{exp}(\lambda \mathbf{I_n}) = e^{\lambda}\mathbf{I_n}$.

Step 2: For a nilpotent matrix $\mathbf{N}_n$, we have

$$\mathbf{N}_n = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad \mathbf{N}_n^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad \dots, \quad \mathbf{N}_n^{n-1} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix},$$

and $\mathbf{N}_n^n = \mathbf{0}_{n\times n}$. Therefore the sum in (24) with $\mathbf{A} = \mathbf{N}_n$ is actually a finite sum, and a direct calculation shows

$$\mathbf{exp}(\mathbf{N}_n) = \mathbf{I_n} + \mathbf{N}_n + \frac{1}{2}\mathbf{N}_n^2 + \dots + \frac{1}{(n-1)!}\mathbf{N}_n^{n-1} = \begin{pmatrix} 1 & 1 & \frac{1}{2} & \frac{1}{3!} & \dots & \frac{1}{(n-2)!} & \frac{1}{(n-1)}! \\ 0 & 1 & 1 & \frac{1}{2} & \dots & \frac{1}{(n-3)!} & \frac{1}{(n-2)!} \\ 0 & 0 & 1 & 1 & \dots & \frac{1}{(n-4)!} & \frac{1}{(n-3)!} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & \dots & 0 & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Step 3: Suppose two matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ commute; that is, $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1$. Then

$$\mathbf{exp}(\mathbf{A}_1)\mathbf{exp}(\mathbf{A}_2) = \mathbf{exp}(\mathbf{A}_2)\mathbf{exp}(\mathbf{A}_1) = \mathbf{exp}(\mathbf{A}_1 + \mathbf{A}_2).$$

This is perhaps surprisingly hard to prove since it is such a familiar property of numbers (i.e. when $n = 1$).

Step 4: If $\mathbf{A}$ is an elementary Jordan block, $\mathbf{A} = \lambda\mathbf{I_n} + \mathbf{N}_n$, then $\mathbf{exp}(\mathbf{A}) = \mathbf{exp}(\lambda)\mathbf{exp}(\mathbf{N}_n)$. This is clear because $\lambda\mathbf{I_n}\mathbf{N}_n = \mathbf{N}_n\lambda\mathbf{I_n}$ from inspection, and the result follows from Step 3.

Step 5: If $\mathbf{A}$ has a block structure

$$\mathbf{A} = \begin{pmatrix} \boxed{\mathbf{A}_1} & & \mathbf{0} \\ & \boxed{\ddots} & \\ \mathbf{0} & & \boxed{\mathbf{A}_k} \end{pmatrix} \quad \text{then} \quad \mathbf{exp}(\mathbf{A}) = \begin{pmatrix} \boxed{\mathbf{exp}(\mathbf{A}_1)} & & \mathbf{0} \\ & \boxed{\ddots} & \\ \mathbf{0} & & \boxed{\mathbf{exp}(\mathbf{A}_k)} \end{pmatrix}$$

Step 6: If $\mathbf{A}_1 = \mathbf{V}^{-1}\mathbf{A}_2\mathbf{V}$, then $\mathbf{exp}(\mathbf{A}_1) = \mathbf{V}^{-1}\mathbf{exp}(\mathbf{A}_2)\mathbf{V}$.

2.2 The Singular Value Decomposition (SVD)

Theorem 2.1. *Suppose* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *has rank* r*. Then there exists unitary matrices* $\mathbf{U} \in \mathbb{R}^{m\times m}$ *and* $\mathbf{V} \in \mathbb{R}^{n\times n}$ *so that*

$$\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^{\top}, \tag{25}$$

where $\Sigma \in \mathbb{R}^{m\times n}$ *has the block form*

$$\Sigma = \left(\begin{array}{c|c} \mathbf{\Lambda} & \mathbf{0}_{r\times(n-r)} \\ \hline \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{array}\right) \tag{26}$$

and $\mathbf{\Lambda} \in \mathbb{R}^{r\times r}$ *is a diagonal matrix whose diagonal entries are the* singular *values*

$$\sigma_1 \geq \cdots \geq \sigma_r > 0, \quad \sigma_i = \sqrt{\lambda_i}, \quad \lambda_i \in \sigma(\mathbf{A}^{\top}\mathbf{A}).$$

Proof. The $n\times n$ matrix $\mathbf{A}^{\top}\mathbf{A}$ is symmetric, positive semidefinite, and of rank r. Thus by Proposition 1.5(iii), there exists an orthonormal eigenbasis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Moreover, by Proposition 1.6, the associated eigenvalues $\sigma(\mathbf{A}^{\top}\mathbf{A}) = \{\lambda_1, \ldots, \lambda_n\}$ are nonnegative and we can assume the ordering is such that $\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = 0 = \cdots = \lambda_n$. Let $\mathbf{V}$ be the matrix whose j^{th} column is $\mathbf{v}_j$, and it satisfies

$$\mathbf{A}^{\top}\mathbf{A} = \mathbf{V}\left(\begin{array}{c|c} \mathbf{\Lambda^2} & \mathbf{0}_{r\times(n-r)} \\ \hline \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(n-r)} \end{array}\right)\mathbf{V}^{\top}, \quad \text{where} \quad \mathbf{\Lambda^2} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_r \end{pmatrix}$$

We now let $\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i \in \mathbb{R}^m$ for $i = \mathbf{1}, \ldots, r$, and let $\mathbf{u}_{i+1}, \ldots, \mathbf{u}_m$ be chosen as any orthonormal basis for $\left[\mathbf{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}\right]^{\perp}$. Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ is an orthonormal basis for $\mathbb{R}^m$, and let $\mathbf{U} \in \mathbb{R}^{m\times m}$ be the unitary matrix whose columns are the $\mathbf{u}_j$'s. We will show (25) holds.

Since $\mathbf{U}^{-1} = \mathbf{U}^{\top}$ and $\mathbf{V}^{-1} = \mathbf{V}^{\top}$, (25) is equivalent to $\mathbf{U}^{\top}\ \mathbf{A}\ \mathbf{V} = \Sigma$. The latter holds since the $(i,j)^{\text{th}}$ entry is computed as

$$\left(\mathbf{U}^{\top}\ \mathbf{A}\ \mathbf{V}\right)_{ij} = (\mathbf{u}_i)^{\top}\ \mathbf{A}\ \mathbf{v}_j = \begin{cases} \frac{1}{\sigma_i}\mathbf{v}_i^{\top}\ \mathbf{A}^{\top}\ \mathbf{A}\ \mathbf{v}_j & \text{if } 1 \leq i \leq r \\ 0 & \text{if } r+1 \leq i \leq m \end{cases} = \begin{cases} \frac{\lambda_j}{\sigma_i}\delta_{ij} & \text{if } 1 \leq i \leq r \\ 0 & \text{if } r+1 \leq i \leq m \end{cases} = \begin{cases} \sigma_i\delta_{ij} & \text{if } 1 \leq i \leq r \\ 0 & \text{if } r+1 \leq i \leq m \end{cases}$$

This agrees with (26), and hence (25) holds. □

3 Exercises

(#1) **(Vector spaces as function spaces)** If $Y \subseteq X$, show $\mathbb{R}^Y$ can be viewed as a subspace of $\mathbb{R}^X$.

(#2) Let $\mathbf{A} = \left(\mathbf{A}_{\star 1} \ldots \mathbf{A}_{\star n}\right) = \left(\mathbf{A}_{1\star}^{\top} \ldots \mathbf{A}_{m\star}^{\top}\right)^{\top} \in \mathbb{R}^{m\times n}$, where $\mathbf{A}_{\star j}$ is the j^{th} column and $\mathbf{A}_{i\star}$ is the i^{th} row. Let $\mathbf{D}_n \in \mathbb{R}^{n\times n}$ (resp. $\mathbf{D}_m \in \mathbb{R}^{m\times m}$) be the diagonal matrix with $\lambda_1, \ldots, \lambda_n$ (resp. $\mu_1, \ldots, \mu_m$) along the diagonal. Show that $\mathbf{A}\mathbf{D_n} = \left(\lambda_1\mathbf{A}_{\star 1} \ldots \lambda_n\mathbf{A}_{\star n}\right)$ and $\mathbf{D}_m\mathbf{A} = \left(\mu_1\mathbf{A}_{1\star}^{\top} \ldots \mu_m\mathbf{A}_{m\star}^{\top}\right)^{\top}$.

(#3) **(Similarity of matrices)**

  (i) Show that the relation that two matrices are similar is indeed an equivalence relation.

(ii) Two $n \times n$ matrices $\mathbf{A}_1$, $\mathbf{A}_2$ are *unitarily similar* provided there exists a unitary matrix $\mathbf{U} \in \mathbb{R}^{n\times n}$ with $\mathbf{A}_1 = \mathbf{U}\mathbf{A}_2\mathbf{U}^{-1}$. Show that the relation of two matrices being unitarily similar is another equivalent relation.

(iii) Provide an example to show similar and unitarily similar are not the same relation.

(iv)

(#4) (Description of rank r matrices)

(i) Suppose $\mathbf{0_n} \neq \mathbf{v}_1 \in \mathbb{R}^n$ and $\mathbf{0}_m \neq \mathbf{w}_1 \in \mathbb{R}^m$, and let $\mathbf{A}_1 \in \mathbb{R}^{m\times n}$ be given by $\mathbf{A}_1 = \mathbf{w}_1\mathbf{v}_1^\top$. That is, if $\mathbf{v}_1 = (v_{11}, \dots, v_{1m})^\top$ and $\mathbf{w}_1 = (w_{11}, \dots, w_{1n})^\top$, then $(\mathbf{A}_1)_{ij} = v_{1i}w_{1j}$ for $1 \leq i \leq m$, $1 \leq j \leq n$. Show that $\mathbf{A}_1$ has rank 1. **Proof:** Suppose $\mathbf{v}_1 \neq \mathbf{0_n}, \mathbf{w}_1 \neq \mathbf{0}_m$ and note that $\mathbf{Ax} = (\mathbf{w}_1\mathbf{v}_1^\top)\mathbf{x} = \mathbf{w}_1(\langle \mathbf{v}_1, \mathbf{x}\rangle)$. Therefore $\mathbf{x} \in \mathcal{N}(\mathbf{A})$ if and only if $\mathbf{x} \in (\mathbf{span}\{\mathbf{v}_1\})^\perp$. Since the latter has dimension $n-1$, the rank of $\mathbf{A}$ is one (see Section 1.4.3).

(ii) Conversely, show that every matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ of rank 1 is of this form. **Proof:** The rank one property implies that the null space $\mathcal{N}(\mathbf{A})$ has dimension $(n-1)$ and we let $\mathbf{v}_1$ be with $\|\mathbf{v}_1\| = 1$ and belongs to $\mathcal{N}(\mathbf{A})^\perp$. By Proposition 1.2, there exists an orthonormal basis whose first element is $\mathbf{v}_1 = (v_{11}, \dots, v_{n1})^\top$ (or, which is the same thing, we choose an orthonormal basis for $\mathcal{N}(\mathbf{A})$), and we let $\mathbf{V}$ be the unitary matrix whose columns are the elements of this basis. The ij^{th} of $\mathbf{V}$ is denoted by v_{ij}. By setting $\mathbf{w}_1 = \mathbf{A}\mathbf{v}_1 = (w_{11}, \dots, w_{1m})^\top$ and letting $\mathbf{W} \in \mathbb{R}^{m\times n}$ be the matrix whose first column is $\mathbf{w}_1$ and has zeros elsewhere, we have

$$\mathbf{AV} = \mathbf{W} \quad \Rightarrow \quad \mathbf{A} = \mathbf{W}\mathbf{V}^{-1} = \mathbf{W}\mathbf{V}^\top.$$

A direct calculation now shows

$$\mathbf{A} = \begin{pmatrix} w_{11} & 0 & \dots & 0 \\ w_{12} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ w_{1m} & 0 & \dots & 0 \end{pmatrix}\begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ v_{1n} & v_{2n} & \dots & v_{nn} \end{pmatrix} = \begin{pmatrix} w_{11}v_{11} & w_{11}v_{21} & \dots & w_{11}v_{n1} \\ w_{12}v_{11} & w_{12}v_{21} & \dots & w_{12}v_{n1} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1m}v_{11} & w_{1m}v_{12} & \dots & w_{1m}v_{1n} \end{pmatrix} = \mathbf{w}_1\mathbf{v}_1^\top$$

(iii) Show that a rank two matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is characterized by the existence of two sets of independent vectors $\{\mathbf{v}_1, \mathbf{v}_2\} \subset \mathbb{R}^n$ and $\{\mathbf{w}_1, \mathbf{w}_2\} \subset \mathbb{R}^m$ so that $\mathbf{A} = \mathbf{w}_1\mathbf{v}_1^\top + \mathbf{w}_2\mathbf{v}_2^\top$. **Proof:** Suppose $\{\mathbf{v}_1, \mathbf{v}_2\} \subset \mathbb{R}^n$ and $\{\mathbf{w}_1, \mathbf{w}_2\} \subset \mathbb{R}^m$ are both independent sets and $\mathbf{A} = \mathbf{w}_1\mathbf{v}_1^\top + \mathbf{w}_2\mathbf{v}_2^\top$. We claim $\mathbf{A}$ has rank 2. Indeed, $\mathbf{span}\{\mathbf{v}_1, \mathbf{v}_2\}^\perp$ has dimension $n-2$, and is contained in the null space $\mathcal{N}(\mathbf{A})$ of $\mathbf{A}$. But $\mathbf{span}\{\mathbf{v}_1, \mathbf{v}_2\}^\perp$ actually equals $\mathcal{N}(\mathbf{A})$ since if $\mathbf{v} \in \mathcal{N}(\mathbf{A})$, then $\mathbf{0}_m = \mathbf{Av} = \mathbf{w}_1\mathbf{v}_1^\top\mathbf{v} + \mathbf{w}_2\mathbf{v}_2^\top\mathbf{v} = (\langle \mathbf{v}_1, \mathbf{v}\rangle)\mathbf{w}_1 + (\langle \mathbf{v}_2, \mathbf{v}\rangle)\mathbf{w}_2$. Now since $\mathbf{w}_1$ and $\mathbf{w}_2$ are independent, we have $0 = \langle \mathbf{v}_1, \mathbf{v}\rangle = \langle \mathbf{v}_1, \mathbf{v}\rangle$, which means $\mathbf{v} \in \mathbf{span}\{\mathbf{v}_1, \mathbf{v}_2\}^\perp$. Hence $\mathcal{N}(\mathbf{A})$ has dimension $n-2$, and $\mathbf{A}$ has rank 2 by (12). Conversely, if $\mathbf{A} \in \mathbb{R}^{m\times n}$ has rank 2, then there exists $\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2 \in \mathbb{R}^m$ for which $\{\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2\}$ is a basis of the range $\mathcal{R}(\mathbf{A})$ of $\mathbf{A}$. Let $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2 \in \mathbb{R}^n$ satisfy $\mathbf{A}\tilde{\mathbf{v}}_1 = \tilde{\mathbf{w}}_1$ and $\mathbf{A}\tilde{\mathbf{v}}_2 = \tilde{\mathbf{w}}_2$, which can easily be shown to be independent. By using the Gram-Schmidt process, we can produce orthomormal vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ that span the same 2 dimensional subspace spanned by $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$, and let $\mathbf{w}_1 = \mathbf{A}\mathbf{v}_1$ and $\mathbf{w}_2 = \mathbf{A}\mathbf{v}_2$. Similar to what was done in part (ii), we extend $\{\mathbf{v}_1, \mathbf{v}_2\}$ to an orthonormal basis on $\mathbb{R}^n$ and let $\mathbf{V}$ be the unitary matrix whose columns are $\mathbf{v}_i$'s. Let $\mathbf{W}$ be the matrix whose first two columns consist of $\mathbf{w}_1$ and $\mathbf{w}_2$ and has zeros elsewhere. With similar notation as in part (ii), we have $\mathbf{AV} = \mathbf{W}$ or that $\mathbf{A} = \mathbf{W}\mathbf{V}^{-1} = \mathbf{W}\mathbf{V}^\top$. Again we calculate

$$\mathbf{A} = \begin{pmatrix} w_{11} & w_{21} & 0 & \dots & 0 \\ w_{12} & w_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{1m} & w_{2m} & 0 & \dots & 0 \end{pmatrix}\begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ v_{1n} & v_{2n} & \dots & v_{nn} \end{pmatrix} = \begin{pmatrix} w_{11}v_{11} + w_{21}v_{12} & w_{11}v_{21} + w_{21}v_{22} & \dots & w_{11}v_{n1} + w_{21}v_{n2} \\ w_{12}v_{11} + w_{22}v_{12} & w_{12}v_{21} + w_{22}v_{22} & \dots & w_{12}v_{n1} + w_{22}v_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1m}v_{11} + w_{2m}v_{12} & w_{1m}v_{21} + w_{2m}v_{22} & \dots & w_{1m}v_{n1} + w_{2m}v_{n2} \end{pmatrix}$$

$$= \begin{pmatrix} w_{11}v_{11} & w_{11}v_{21} & \dots & w_{11}v_{n1} \\ w_{12}v_{11} & w_{12}v_{21} & \dots & w_{12}v_{n1} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1m}v_{11} & w_{1m}v_{21} & \dots & w_{1m}v_{n1} \end{pmatrix} + \begin{pmatrix} w_{21}v_{12} & w_{21}v_{22} & \dots & w_{21}v_{n2} \\ w_{22}v_{12} & w_{22}v_{22} & \dots & w_{22}v_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ w_{2m}v_{12} & w_{2m}v_{22} & \dots & w_{2m}v_{n2} \end{pmatrix} = \mathbf{w}_1\mathbf{v}_1^\top + \mathbf{w}_2\mathbf{v}_2^\top$$

(iv) Formulate and prove a result analogous to (iii) that characterizes the property that $\mathbf{A} \in \mathbb{R}^{m\times n}$ has rank r, where $1 \leq r \leq \min\{m, n\}$.

(#5) Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$.

(i) Show that if $\left\|\mathbf{A}\right\| < 1$, then $(\mathbf{I_n} - \mathbf{A})^{-1}$ exists and equals $\sum_{i=0}^{\infty} \mathbf{A}^i$.

(ii) Suppose $r > \left\|\mathbf{A}\right\|$. Show that $(r\mathbf{I_n} - \mathbf{A})^{-1}$ exists, and find the infinite series that it equals.