EE240A - MAE 270A - Fall 2002
Review of some elements of linear algebra

Prof. Fernando Paganini

September 27, 2002

1 Linear spaces and mappings

In this section we will introduce some of the basic ideas in linear algebra. Our treatment is primarily intended as a review for the reader's convenience, with some additional focus on the geometric aspects of the subject. References are given at the end of the chapter for more details at both introductory and advanced levels.

1.1 Vector spaces

The structure introduced now will pervade our course, that of a *vector space*, also called a linear space. This is a set that has a natural addition operation defined on it, together with scalar multiplication. Because this is such an important concept, and arises in a number of different ways, it is worth defining it precisely below.

Before proceeding we set some basic notation. The real numbers will be denoted by $\mathbb{R}$, and the complex numbers by $\mathbb{C}$; we will use $j := \sqrt{-1}$ for the imaginary unit. Also, given a complex number $z = x + jy$ with $x, y \in \mathbb{R}$:

- $z^* = x - jy$ is the complex conjugate;

- $|z| = \sqrt{x^2 + y^2}$ is the complex magnitude;

- $x = \text{Re}(z)$ is the real part.

We use $\mathbb{C}^+$ to denote the open right half-plane formed from the complex numbers with positive real part; $\bar{\mathbb{C}}^+$ is the corresponding closed half-plane, and the left half-planes $\mathbb{C}^-$ and $\bar{\mathbb{C}}^-$ are analogously defined. Finally, $j\mathbb{R}$ denotes the imaginary axis.

We now define a vector space. In the definition, the field $\mathbb{F}$ can be taken here to be the real numbers $\mathbb{R}$, or the complex numbers $\mathbb{C}$. The terminology *real* vector space, or *complex* vector space is used to specify these alternatives.

Definition 1. *Suppose $\mathcal{V}$ is a nonempty set and $\mathbb{F}$ is a field, and that operations of vector addition and scalar multiplication are defined in the following way.*

(a) *For every pair u, $v \in \mathcal{V}$ a unique element $u + v \in \mathcal{V}$ is assigned called their sum;*

(b) *for each $\alpha \in \mathbb{F}$ and $v \in \mathcal{V}$, there is a unique element $\alpha v \in \mathcal{V}$ called their product.*

Then $\mathcal{V}$ is a vector space if the following properties hold for all u, v, $w \in \mathcal{V}$, and all α, $\beta \in \mathbb{F}$:

(i) *There exists a zero element in $\mathcal{V}$, denoted by 0, such that $v + 0 = v$;*

(ii) *there exists a vector $-v$ in $\mathcal{V}$, such that $v + (-v) = 0$;*

(iii) *the association $u + (v + w) = (u + v) + w$ is satisfied;*

(iv) *the commutativity relationship $u + v = v + u$ holds;*

(v) *scalar distributivity $\alpha(u + v) = \alpha u + \alpha v$ holds;*

(vi) *vector distributivity $(\alpha + \beta)v = \alpha v + \beta v$ is satisfied;*

(vii) *the associative rule $(\alpha\beta)v = \alpha(\beta v)$ for scalar multiplication holds;*

(viii) *for the unit scalar $1 \in \mathbb{F}$ the equality $1v = v$ holds.*

Formally, a vector space is an additive group together with a scalar multiplication operation defined over a field $\mathbb{F}$, which must satisfy the usual rules (v–viii) of distributivity and associativity. Notice that both $\mathcal{V}$ and $\mathbb{F}$ contain the zero element, which we will denote by "0" regardless of the instance.

Given two vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$, with the same associated scalar field, we use $\mathcal{V}_1 \times \mathcal{V}_2$ to denote the vector space formed by their Cartesian product. Thus every element of $\mathcal{V}_1 \times \mathcal{V}_2$ is of the form

$$(v_1, v_2) \quad \text{where } v_1 \in \mathcal{V}_1 \text{ and } v_2 \in \mathcal{V}_2.$$

Having defined a vector space we now consider a number of examples.

Examples:

Both $\mathbb{R}$ and $\mathbb{C}$ can be considered as real vector spaces, although $\mathbb{C}$ is more commonly regarded as a complex vector space. The most common example of a real vector space is $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$; namely, n copies of $\mathbb{R}$. We represent elements of $\mathbb{R}^n$ in a column vector notation

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n, \quad \text{where each } x_k \in \mathbb{R}.$$

2

Addition and scalar multiplication in $\mathbb{R}^n$ are defined componentwise:

$$x + y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}, \quad \alpha x = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}, \quad \text{for} \quad \alpha \in \mathbb{R}, \quad x, y \in \mathbb{R}^n.$$

Identical definitions apply to the complex space $\mathbb{C}^n$. As a further step, consider the space $\mathbb{C}^{m \times n}$ of complex $m \times n$ matrices of the form

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

Using once again componentwise addition and scalar multiplication, $\mathbb{C}^{m \times n}$ is a (real or complex) vector space.

We now define two vector spaces of matrices which will be central in our course. First, we define the *transpose* of the above matrix $A \in \mathbb{C}^{m \times n}$ by

$$A' = \begin{bmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{C}^{n \times m},$$

and its *Hermitian conjugate* or *adjoint* by

$$A^* = \begin{bmatrix} a_{11}^* & \cdots & a_{m1}^* \\ \vdots & \ddots & \vdots \\ a_{1n}^* & \cdots & a_{mn}^* \end{bmatrix} \in \mathbb{C}^{n \times m}.$$

In both cases the indices have been transposed, but in the latter we also take the complex conjugate of each element. Clearly both operations coincide if the matrix is real; we thus favor the notation A^*, which will serve to indicate both the transpose of a real matrix, and the adjoint of a complex matrix.[1]

The square matrix $A \in \mathbb{C}^{n \times n}$ is *Hermitian* or *self-adjoint* if

$$A = A^*.$$

The space of Hermitian matrices is denoted $\mathbb{H}^n$, and is a *real* vector space. If a Hermitian matrix A is in $\mathbb{R}^{n \times n}$ it is more specifically referred to as *symmetric*. The set of symmetric matrices is also a real vector space and will be written $\mathbb{S}^n$.

The set $\mathcal{F}(\mathbb{R}^m, \mathbb{R}^n)$ of functions mapping m real variables to $\mathbb{R}^n$ is a vector space. Addition between two functions f_1 and f_2 is defined by

$$(f_1 + f_2)(x_1, \ldots, x_m) = f_1(x_1, \ldots, x_m) + f_2(x_1, \ldots, x_m)$$

---

[1] The transpose, without conjugation, of a complex matrix A will be denoted by A'; however, it is seldom required.

for any variables $x_1, \ldots, x_m$; this is called pointwise addition. Scalar multiplication by a real number α is defined by

$$(\alpha f)(x_1, \ldots, x_m) = \alpha f(x_1, \ldots, x_m).$$

An example of a less standard vector space is given by the set composed of multinomials in m variables, that have homogeneous order n. We denote this set by $P_m^{[n]}$. To illustrate the elements of this set consider

$$p_1(x_1, x_2, x_3) = x_1^2 x_2 x_3, \quad p_2(x_1, x_2, x_3) = x_1^3 x_2, \quad p_3(x_1, x_2, x_3) = x_1 x_2 x_3.$$

Each of these is a multinomial in three variables; however, p_1 and p_2 have order four, whereas the order of p_3 is three. Thus only p_1 and p_2 are in $P_3^{[4]}$. Similarly of

$$p_4(x_1, x_2, x_3) = x_1^4 + x_2 x_3^3 \quad \text{and} \quad p_5(x_1, x_2, x_3) = x_1^2 x_2 x_3 + x_1$$

only p_4 is in $P_3^{[4]}$, whereas p_5 is not in any $P_3^{[n]}$ space since its terms are not homogeneous in order. Some thought will convince you that $P_m^{[n]}$ is a vector space under pointwise addition. □

1.2 Subspaces

A *subspace* of a vector space $\mathcal{V}$ is a subset of $\mathcal{V}$ which is also a vector space with respect to the same field and operations; equivalently, it is a subset which is closed under the operations on $\mathcal{V}$.

Examples:

A vector space can have many subspaces, and the simplest of these is the zero subspace, denoted by $\{0\}$. This is a subspace of any vector space and contains only the zero element. Excepting the zero subspace and the entire space, the simplest type of subspace in $\mathcal{V}$ is of the form

$$\mathcal{S}_v = \{s \in \mathcal{V} : s = \alpha v, \text{ for some } \alpha \in \mathbb{R}\},$$

given v in $\mathcal{V}$. That is, each element in $\mathcal{V}$ generates a subspace by multiplying it by all possible scalars. In $\mathbb{R}^2$ or $\mathbb{R}^3$, such subspaces correspond to lines going through the origin.

Going back to our earlier examples of vector spaces we see that the multinomials $P_m^{[n]}$ are subspaces of $\mathcal{F}(\mathbb{R}^m, \mathbb{R})$, for any n.

Now $\mathbb{R}^n$ has many subspaces and an important set is those associated with the natural insertion of $\mathbb{R}^m$ into $\mathbb{R}^n$, when $m < n$. Elements of these subspaces are of the form

$$x = \begin{bmatrix} \bar{x} \\ 0 \end{bmatrix},$$

where $\bar{x} \in \mathbb{R}^m$ and $0 \in \mathbb{R}^{n-m}$. □

Given two subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ we can define the addition

$$\mathcal{S}_1 + \mathcal{S}_2 = \{ s \in \mathcal{V} : s = s_1 + s_2 \text{ for some } s_1 \in \mathcal{S}_1 \text{ and } s_2 \in \mathcal{S}_2 \}$$

which is easily verified to be a subspace.

1.3 Bases, spans, and linear independence

We now define some key vector space concepts. Given elements $v_1, \ldots, v_m$ in a vector space we denote their *span* by

$$\text{span}\{v_1, \ldots, v_m\},$$

which is the set of all vectors v that can be written as

$$v = \alpha_1 v_1 + \cdots + \alpha_m v_m$$

for some scalars $\alpha_k \in \mathbb{F}$; the above expression is called a *linear combination* of the vectors $v_1, \ldots, v_m$. It is straightforward to verify that the span always defines a subspace. If for some vectors we have

$$\text{span}\{v_1, \ldots, v_m\} = \mathcal{V},$$

we say that the vector space $\mathcal{V}$ is finite dimensional. If no such finite set of vectors exists we say the vector space is infinite dimensional. Our focus for the remainder of the chapter is *exclusively* finite dimensional vector spaces. We will pursue the study of some infinite dimensional spaces in Chapter 3.

If a vector space $\mathcal{V}$ is finite dimensional we define its *dimension*, denoted $\dim(\mathcal{V})$, to be the smallest number n such that there exist vectors $v_1, \ldots, v_n$ satisfying

$$\text{span}\{v_1, \ldots, v_n\} = \mathcal{V}.$$

In that case we say that the set

$$\{v_1, \ldots, v_n\} \text{ is a } basis \text{ for } \mathcal{V}.$$

Notice that a basis will automatically satisfy the *linear independence* property, which means that the only solution to the equation

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$$

is $\alpha_1 = \cdots = \alpha_n = 0$. Otherwise, one of the elements v_i could be expressed as a linear combination of the others and $\mathcal{V}$ would be spanned by fewer than n vectors. Given this observation, it follows easily that for a given $v \in \mathcal{V}$, the scalars $(\alpha_1, \ldots, \alpha_n)$ satisfying

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = v$$

are unique; they are termed the *coordinates* of v in the basis $\{v_1, \ldots, v_n\}$.

Linear independence is defined analogously for any set of vectors $\{v_1, \ldots, v_m\}$; it is equivalent to saying the vectors are a basis for their span. The maximal number of linearly independent vectors is n, the dimension of the space; in fact any linearly independent set can be extended with additional vectors to form a basis.

5

Examples:

From our examples so far $\mathbb{R}^n, \mathbb{C}^{m \times n}$, and $P_m^{[n]}$ are all finite dimensional vector spaces; however, $\mathcal{F}(\mathbb{R}^m, \mathbb{R}^n)$ is infinite dimensional. The real vector space $\mathbb{R}^n$ and complex vector space $\mathbb{C}^{m \times n}$ are n and mn dimensional, respectively. The dimension of $P_m^{[n]}$ is more challenging to compute and its determination is an exercise at the end of the chapter.

An important computational concept in vector space analysis is associating a general k dimensional vector space $\mathcal{V}$ with the vector space $\mathbb{F}^k$. This is done by taking a basis $\{v_1, \ldots, v_k\}$ for $\mathcal{V}$, and associating each vector v in $\mathcal{V}$ with the vector of coordinates in the given basis,

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} \in \mathbb{F}^k.$$

Equivalently, each vector v_i in the basis is associated with the vector

$$e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{F}^k.$$

That is, e_i is the vector with zeros everywhere excepts its ith entry, which is equal to one. Thus we are identifying the basis $\{v_1, \ldots, v_k\}$ in $\mathcal{V}$ with the set $\{e_1, \ldots, e_k\}$ which is in fact a basis of $\mathbb{F}^k$, called the *canonical* basis.

To see how this type of identification is made, suppose we are dealing with $\mathbb{R}^{n \times m}$, which has dimension $k = nm$. Then a basis for this vector space is

$$E_{ir} = \begin{bmatrix} 0 & \cdots & & 0 \\ \vdots & \ddots & 1 & \vdots \\ 0 & \cdots & & 0 \end{bmatrix},$$

which are the matrices that are zero everywhere but their (i, r)th-entry, which is one. Then we identify each of these with the vector $e_{n(r-1)+i} \in \mathbb{R}^k$. Thus addition or scalar multiplication on $\mathbb{R}^{n \times m}$ can be translated to equivalent operations on $\mathbb{R}^k$. □

1.4 Mappings and matrix representations

We now introduce the concept of a *linear* mapping between vector spaces. The mapping $A : \mathcal{V} \to \mathcal{W}$ is linear if

$$A(\alpha v_1 + \beta v_2) = \alpha A v_1 + \beta A v_2$$

for all v_1, v_2 in $\mathcal{V}$, and all scalars α_1 and α_2. Here $\mathcal{V}$ and $\mathcal{W}$ are vector spaces with the same associated field $\mathbb{F}$. The space $\mathcal{V}$ is called the *domain* of the mapping, and $\mathcal{W}$ its *codomain*.

Given bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ for $\mathcal{V}$ and $\mathcal{W}$, respectively, we associate scalars a_{ik} with the mapping A, defining them such that they satisfy

$$Av_k = a_{1k}w_1 + a_{2k}w_2 + \cdots + a_{mk}w_m,$$

for each $1 \leq k \leq n$. Namely, given any basis vector v_k, the coefficients a_{ik} are the coordinates of Av_k in the chosen basis for $\mathcal{W}$. It turns out that these mn numbers a_{ik} completely specify the linear mapping A. To see this is true consider any vector $v \in \mathcal{V}$, and let $w = Av$. We can express both vectors in their respective bases as $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $w = \beta_1 w_1 + \cdots + \beta_m w_m$. Now we have

$$\begin{aligned}
w = Av &= A(\alpha_1 v_1 + \cdots + \alpha_n v_n) \\
&= \alpha_1 Av_1 + \cdots + \alpha_n Av_n \\
&= \sum_{k=1}^{n} \sum_{i=1}^{m} \alpha_k a_{ik} w_i = \sum_{i=1}^{m} \left(\sum_{k=1}^{n} \alpha_k a_{ik} \right) w_i,
\end{aligned}$$

and therefore by uniqueness of the coordinates we must have

$$\beta_i = \sum_{k=1}^{n} \alpha_k a_{ik}, \quad i = 1, \ldots, m.$$

To express this relationship in a more convenient form, we can write the set of numbers a_{ik} as the $m \times n$ matrix

$$[A] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

Then via the standard matrix product we have

$$\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

In summary any linear mapping A between vector spaces can be regarded as a matrix $[A]$ mapping $\mathbb{F}^n$ to $\mathbb{F}^m$ via matrix multiplication.

Notice that the numbers a_{ik} depend intimately on the bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$. Frequently we use only one basis for $\mathcal{V}$ and one for $\mathcal{W}$ and thus there is no need to distinguish between the map A and the basis dependent matrix $[A]$. Therefore after this section we will simply write A to denote either the map or the matrix, making which is meant context dependent.

We now give two examples to illustrate the above discussion more clearly.

Examples:

Given matrices $B \in \mathbb{C}^{k \times k}$ and $D \in \mathbb{C}^{l \times l}$ we define the map $\Pi : \mathbb{C}^{k \times l} \to \mathbb{C}^{k \times l}$ by

$$\Pi(X) = BX - XD,$$

where the right-hand side is in terms of matrix addition and multiplication. Clearly Π is a linear mapping since

$$\begin{aligned}
\Pi(\alpha X_1 + \beta X_2) &= B(\alpha X_1 + \beta X_2) - (\alpha X_1 + \beta X_2)D \\
&= \alpha(BX_1 - X_1 D) + \beta(BX_2 - X_2 D) \\
&= \alpha \Pi(X_1) + \beta \Pi(X_2).
\end{aligned}$$

If we now consider the identification between the matrix space $\mathbb{C}^{k \times l}$ and the product space $\mathbb{C}^{kl}$, then Π can be thought of as a map from $\mathbb{C}^{kl}$ to $\mathbb{C}^{kl}$, and can accordingly be represented by a complex matrix, which is $kl \times kl$. We now do an explicit 2×2 example for illustration. Suppose $k = l = 2$ and that

$$B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } D = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}.$$

We would like to find a matrix representation for Π. Since the domain and codomain of Π are equal, we will use the standard basis for $\mathbb{C}^{2 \times 2}$ for each. This basis is given by the matrices E_{ir} defined earlier. We have

$$\Pi(E_{11}) = \begin{bmatrix} -4 & 0 \\ 3 & 0 \end{bmatrix} = -4E_{11} + 3E_{21};$$

$$\Pi(E_{12}) = \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix} = E_{12} + 3E_{22};$$

$$\Pi(E_{21}) = \begin{bmatrix} 2 & 0 \\ -1 & 0 \end{bmatrix} = 2E_{11} - E_{21};$$

$$\Pi(E_{22}) = \begin{bmatrix} 0 & 2 \\ 0 & 4 \end{bmatrix} = 2E_{12} + 4E_{22}.$$

Now we identify the basis $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ with the standard basis for $\mathbb{C}^4$ given by $\{e_1, e_2, e_3, e_4\}$. Therefore we get that

$$[\Pi] = \begin{bmatrix} -4 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & -1 & 0 \\ 0 & 3 & 0 & 4 \end{bmatrix}$$

in this basis.

Another linear operator involves the multinomial function $P_m^{[n]}$ defined earlier in this section. Given an element $a \in P_m^{[k]}$ we can define the mapping $\Omega : P_m^{[n]} \to P_m^{[n+k]}$ by function multiplication

$$\Omega(p)(x_1, x_2, \ldots, x_m) := a(x_1, x_2, \ldots, x_m)p(x_1, x_2, \ldots, x_m).$$

Again Ω can be regarded as a matrix, which maps $\mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, where d_1 and d_2 are the dimensions of $P_m^{[n]}$ and $P_m^{[n+k]}$, respectively. $\qquad\square$

Associated with any linear map $A : \mathcal{V} \to \mathcal{W}$ is its *image* space, which is defined by

$$\mathrm{Im}A = \{w \in \mathcal{W} : \text{ there exists } v \in \mathcal{V} \text{ satisfying } Av = w\}.$$

This set contains all the elements of $\mathcal{W}$ which are the image of some point in $\mathcal{V}$. Clearly if $\{v_1, \dots, v_n\}$ is a basis for $\mathcal{V}$, then

$$\mathrm{Im}A = \mathrm{span}\{Av_1, \dots, Av_n\}$$

and is thus a subspace. The map A is called *surjective* when $\mathrm{Im}A = \mathcal{W}$.

The dimension of the image space is called the *rank* of the linear mapping A, and the concept is applied as well to the associated matrix $[A]$. Namely,

$$\mathrm{rank}[A] = \dim(\mathrm{Im}A).$$

If $\mathcal{S}$ is a subspace of $\mathcal{V}$, then the image of $\mathcal{S}$ under the mapping A is denoted $A\mathcal{S}$. That is,

$$A\mathcal{S} = \{w \in \mathcal{W} : \text{ there exists } s \in \mathcal{S} \text{ satisfying } As = w\}.$$

In particular, this means that $A\mathcal{V} = \mathrm{Im}A$.

Another important set related to A is its *kernel*, or null space, defined by

$$\mathrm{Ker}A = \{v \in \mathcal{V} : \quad Av = 0\}.$$

In words, $\mathrm{Ker}A$ is the set of vectors in $\mathcal{V}$ which get mapped by A to the zero element in $\mathcal{W}$, and is easily verified to be a subspace of $\mathcal{V}$.

If we consider the equation $Av = w$, suppose v_a and v_b are both solutions; then

$$A(v_a - v_b) = 0.$$

Plainly, the difference between any two solutions is in the kernel of A. Thus given any solution v_a to the equation, all solutions are parametrized by

$$v_a + v_0,$$

where v_0 is any element in $\mathrm{Ker}A$.

In particular, when $\mathrm{Ker}A$ is the zero subspace, there is at most a unique solution to the equation $Av = w$. This means $Av_a = Av_b$ only when $v_a = v_b$; a mapping with this property is called *injective*.

In summary, a solution to the equation $Av = w$ will exist if and only if $w \in \mathrm{Im}A$; it will be unique only when $\mathrm{Ker}A$ is the zero subspace.

The dimensions of the image and kernel of A are linked by the relationship

$$\dim(\mathcal{V}) = \dim(\mathrm{Im}A) + \dim(\mathrm{Ker}A),$$

9

proved in the exercises at the end of the chapter.

A mapping is called *bijective* when it is both injective and surjective; that is, for every $w \in \mathcal{W}$ there exists a unique v satisfying $Av = w$. In this case there is a well-defined inverse mapping $A^{-1} : \mathcal{W} \to \mathcal{V}$, such that

$$A^{-1}A = I_{\mathcal{V}}, \qquad AA^{-1} = I_{\mathcal{W}}.$$

In the above, I denotes the identity mapping in each space, that is the map that leaves elements unchanged. For instance, $I_{\mathcal{V}} : v \mapsto v$ for every $v \in \mathcal{V}$.

From the above property on dimensions we see that if there exists a bijective linear mapping between two spaces $\mathcal{V}$ and $\mathcal{W}$, then the spaces must have the same dimension. Also, if a mapping A is from $\mathcal{V}$ back to itself, namely, $A : \mathcal{V} \to \mathcal{V}$, then one of the two properties (injectivity or surjectivity) suffices to guarantee the other.

We will also use the terms *nonsingular* or *invertible* to describe bijective mappings, and apply these terms as well to their associated matrices. Notice that invertibility of the mapping A is equivalent to invertibility of $[A]$ in terms of the standard matrix product; this holds true regardless of the chosen bases.

Examples:

To illustrate these notions let us return to the mappings Π and Ω defined above. For the 2×2 numerical example given, Π maps $\mathbb{C}^{2 \times 2}$ back to itself. It is easily checked that it is invertible by showing either

$$\operatorname{Im} \Pi = \mathbb{C}^{2 \times 2}, \quad \text{or equivalently} \quad \operatorname{Ker} \Pi = 0.$$

In contrast Ω is not a map on the same space, instead taking $P_m^{[n]}$ to the larger space $P_m^{[n+k]}$. And we see that the dimension of the image of Ω is at most n, and the dimension of its kernel at least k. Thus assuming $k > 0$, there are at least some elements $w \in P_m^{[n+k]}$ for which

$$\Omega v = w$$

cannot be solved. These are exactly the values of w that are not in $\operatorname{Im}\Omega$.

$\square$

1.5   Change of basis and invariance

We have already discussed the idea of choosing a basis $\{v_1, \dots, v_n\}$ for the vector space $\mathcal{V}$, and then associating every vector x in $\mathcal{V}$ with its coordinates

$$x_v = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{F}^n,$$

which are the unique scalars satisfying $x = \alpha_1 v_1 + \cdots + \alpha_n v_n$. This raises the question, suppose we choose another basis $u_1, \ldots, u_n$ for $\mathcal{V}$, how can we effectively move between these basis representations? That is, given $x \in \mathcal{V}$, how are the coordinate vectors $x_v, x_u \in \mathbb{F}^n$ related?

The answer is as follows. Suppose that each basis vector u_k is expressed by

$$u_k = t_{1k} v_1 + \cdots + t_{nk} v_n,$$

in the basis $\{v_1, \ldots, v_n\}$. Then the coefficients t_{ik} define the matrix

$$T = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix}.$$

Notice that such a matrix is nonsingular, since it represents the identity mapping $I_{\mathcal{V}}$ in the bases $\{v_1, \ldots, v_n\}$ and $\{u_1, \ldots, u_n\}$. Then the relationship between the two coordinate vectors is

$$T x_u = x_v.$$

Now suppose $A : \mathcal{V} \to \mathcal{V}$ and that $A_v : \mathbb{F}^n \to \mathbb{F}^n$ is the representation of A on the basis $v_1, \ldots, v_n$, and A_u is the representation of A using the basis $u_1, \ldots, u_n$. How is A_u related to A_v?

To study this, take any $x \in \mathcal{V}$ and let x_v, x_u be its coordinates in the respective bases, and z_v, z_u be the coordinates of Ax. Then we have

$$z_u = T^{-1} z_v = T^{-1} A_v x_v = T^{-1} A_v T x_u.$$

Since the above identity and

$$z_u = A_u x_u$$

both hold for every x_u, we conclude that

$$A_u = T^{-1} A_v T.$$

The above relationship is called a *similarity transformation*. This discussion can be summarized in the following commutative diagram. Let $E : \mathcal{V} \to \mathbb{F}^n$ be the map that takes elements of $\mathcal{V}$ to their representation in $\mathbb{F}^n$ with respect to the basis $\{v_1, \ldots, v_n\}$. Then
Next we examine mappings when viewed with respect to a subspace. Suppose that $\mathcal{S} \subset \mathcal{V}$ is a k-dimensional subspace of $\mathcal{V}$, and that $v_1, \ldots, v_n$ is a basis for $\mathcal{V}$ with

$$\operatorname{span}\{v_1, \ldots, v_k\} = \mathcal{S}.$$

That is the first k vectors of this basis forms a basis for $\mathcal{S}$. If $E : \mathcal{V} \to \mathbb{F}^n$ is the associated map which maps the basis vectors in $\mathcal{V}$ to the standard basis on $\mathbb{F}^n$, then

$$E\mathcal{S} = \mathbb{F}^k \times \{0\} \subset \mathbb{F}^n.$$

$$\begin{array}{ccccc}
\mathcal{V} & \xrightarrow{\;E\;} & \mathbb{F}^n & \xrightarrow{\;T^{-1}\;} & \mathbb{F}^n \\
\Big\downarrow{\scriptstyle A} & & \Big\downarrow{\scriptstyle A_v} & & \Big\downarrow{\scriptstyle A_u = T^{-1}A_v T.} \\
\mathcal{V} & \xrightarrow{\;E\;} & \mathbb{F}^n & \xrightarrow{\;T^{-1}\;} & \mathbb{F}^n
\end{array}$$

Thus in $\mathbb{F}^n$ we can view $\mathcal{S}$ as the elements of the form

$$\begin{bmatrix} x \\ 0 \end{bmatrix} \text{ where } x \in \mathbb{F}^k.$$

From the point of view of a linear mapping $A : \mathcal{V} \to \mathcal{V}$ this partitioning of $\mathbb{F}^n$ gives a useful decomposition of the corresponding matrix $[A]$. Namely, we can regard $[A]$ as

$$[A] = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix},$$

where $A_1 : \mathbb{F}^k \to \mathbb{F}^k$, $A_2 : \mathbb{F}^{n-k} \to \mathbb{F}^k$, $A_3 : \mathbb{F}^k \to \mathbb{F}^{n-k}$, and $A_4 : \mathbb{F}^{n-k} \to \mathbb{F}^{n-k}$. We have that

$$EA\mathcal{S} = \mathrm{Im} \begin{bmatrix} A_1 \\ A_3 \end{bmatrix}.$$

Finally to end this section we have the notion of *invariance* of a subspace to a mapping. We say that a subspace $\mathcal{S} \subset \mathcal{V}$ is A-invariant if $A : \mathcal{V} \to \mathcal{V}$ and

$$A\mathcal{S} \subset \mathcal{S}.$$

Clearly every map has at least two invariant subspaces, the zero subspace and entire domain $\mathcal{V}$. For subspaces $\mathcal{S}$ of intermediate dimension, the invariance property is expressed most clearly by saying the associated matrix has the form

$$[A] = \begin{bmatrix} A_1 & A_2 \\ 0 & A_4 \end{bmatrix}.$$

Here we are assuming, as above, that our basis for $\mathcal{V}$ is obtained by extending a basis for $\mathcal{S}$. Similarly if a matrix has this form the subspace $\mathbb{F}^k \times \{0\}$ is $[A]$-invariant.

We will revisit the question of finding non-trivial invariant subspaces later in the chapter, when studying eigenvectors and the Jordan decomposition.

2 Matrix theory

The material of this section is aimed directly at both analysis and computation. Our goals will be to review some basic facts about matrices, and present some additional results for later reference, including two matrix decompositions which have tremendous application, the *Jordan* form and *singular value* decomposition. Both are extremely useful for analytical purposes, and the singular value decomposition is also very important in computations. We will also present some results about self-adjoint and positive definite matrices.

2.1 Eigenvalues and Jordan form

In this section we are concerned exclusively with complex square matrices. We begin with a definition: if $A \in \mathbb{C}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if

$$Ax = \lambda x \tag{1}$$

can be satisfied for some nonzero vector x in $\mathbb{C}^n$. Such a vector x is called an *eigenvector*. Equivalently this means that $\mathrm{Ker}\,(\lambda I - A) \neq 0$ or $\lambda I - A$ is singular. A matrix is singular exactly when its determinant is zero, and therefore we have that λ is an eigenvalue if and only if

$$\det(\lambda I - A) = 0,$$

where $\det(\cdot)$ denotes determinant. Regarding λ as a variable we call the polynomial

$$\det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$$

the *characteristic polynomial* of A. If A is a real matrix then the coefficients a_k will be real as well. The characteristic polynomial can be factored as

$$\det(\lambda I - A) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n).$$

The n complex roots λ_k, which need not be distinct, are the eigenvalues of A, and are collectively denoted by $\mathrm{eig}(A)$. Furthermore if A is a real matrix, then any nonreal eigenvalues must appear in conjugate pairs. Also, a matrix has the eigenvalue zero if and only if it is singular.

Associated with every eigenvalue λ_k is the subspace

$$\mathcal{E}_k = \mathrm{Ker}\,(\lambda_k I - A);$$

every nonzero element in $\mathcal{E}_k$ is an eigenvector corresponding to the eigenvalue λ_k. Now suppose that a set of eigenvectors satisfies

$$\mathrm{span}\{x_1, \ldots, x_n\} = \mathbb{C}^n.$$

Then we can define the invertible matrix $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$, and from the matrix product we find

$$AX = \begin{bmatrix} Ax_1 & \cdots & Ax_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \cdots & \lambda_n x_n \end{bmatrix} = X\Lambda$$

13

where Λ is the diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

Thus in this case we have a similarity transformation X such that $X^{-1}AX = \Lambda$ is diagonal, and we say that the matrix A is *diagonalizable*.

Summarizing we have the following result.

Proposition 2. *A matrix A is diagonalizable if and only if*

$$\mathcal{E}_1 + \mathcal{E}_2 + \cdots + \mathcal{E}_n = \mathbb{C}^n \ \ holds.$$

The following example shows that not all matrices can be diagonalized. Consider the 2×2 matrix

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

It has a repeated eigenvalue at zero, but only one linearly independent eigenvector. Thus it cannot be diagonalized. Matrices of this form have a special role in the decomposition we are about to introduce: define the $n \times n$ matrix N by

$$N = \begin{bmatrix} 0 & 1 & & 0 \\ & & \ddots & \\ & & & 1 \\ 0 & & & 0 \end{bmatrix},$$

where $N = 0$ if the dimension $n = 1$. Such matrices are called nilpotent because $N^n = 0$. Using these we define a matrix to be a *Jordan block* if it is of the form

$$J = \lambda I + N = \begin{bmatrix} \lambda & 1 & & 0 \\ & & \ddots & \\ & & & 1 \\ 0 & & & \lambda \end{bmatrix}.$$

Notice all scalars are 1×1 Jordan blocks. A Jordan block has one eigenvalue λ of multiplicity n. However, it has only one linearly independent eigenvector. A key feature of a Jordan block is that it has precisely n subspaces which are J-invariant. They are given by

$$\mathbb{C}^k \times \{0\},$$

for $1 \le k \le n$. When $k = 1$ this corresponds exactly to the subspace associated with its eigenvector. We can now state the Jordan decomposition theorem.

Theorem 3. *Suppose $A \in \mathbb{C}^{n \times n}$. Then there exists a nonsingular matrix $T \in \mathbb{C}^{n \times n}$, and an integer $1 \leq p \leq n$, such that*

$$T^{-1}AT = J = \begin{bmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_p \end{bmatrix},$$

where the matrices J_k are Jordan blocks.

This theorem states that a matrix can be transformed to one that is block-diagonal, where each of the diagonal matrices is a Jordan block. Clearly if a matrix is diagonalizable each Jordan block J_k will simply be a scalar equal to an eigenvalue of A. In general each block J_k has a single eigenvalue of A in all its diagonal entries; however, a given eigenvalue of A may occur in several blocks.

The relevance of the Jordan decomposition is that it provides a canonical form to characterize matrix similarity; namely, two matrices are similar if and only if they share the same Jordan form. Another related feature is that the Jordan form exhibits the structure of *invariant subspaces* of a given matrix. This is best seen by writing the above equation as

$$AT = TJ.$$

Now suppose we denote by T_1 the submatrix of T formed by its first n_1 columns, where n_1 is the dimension of the block J_1. Then the first n_1 columns of the preceding equation give

$$AT_1 = T_1 J_1,$$

which implies that $\mathcal{S}_1 = \mathrm{Im} T_1$ is invariant under A. Furthermore, we can use this formula to study the linear mapping on $\mathcal{S}_1$ obtained by restriction of A. In fact we find that in the basis defined by the columns of T_1, this linear mapping has the associated matrix J_1; in particular, the only eigenvalue of A restricted to $\mathcal{S}_1$ is λ_1.

The preceding idea can be extended by selecting T_1 to contain the columns corresponding to more than one Jordan block. The resulting invariant subspace will be such that the restriction of A to it has only the eigenvalues of the chosen blocks. Even more generally, we can pick any invariant subspace of J and generate from it invariant subspaces of A. Indeed there are exactly n_k invariant subspaces of A associated with the $n_k \times n_k$ Jordan block J_k, and all invariant subspaces of A can be constructed from this collection.

We will not explicitly require a constructive method for transforming a matrix to Jordan form, and will use this result solely for analysis.

2.2 Self-adjoint, unitary, and positive definite matrices

We have already introduced the adjoint A^* of a complex matrix A; in this section we study in more detail the structure given to the space of matrices by

this operation. A first observation, which will be used extensively below, is that

$$(AB)^* = B^* A^*$$

for matrices A and B of compatible dimensions; this follows directly by definition.

Another basic concept closely related to the adjoint is the Euclidean length of a vector $x \in \mathbb{C}^n$, defined by

$$|x| = \sqrt{x^* x}$$

This extends the usual definition of magnitude of a complex number, so our notation will not cause any ambiguity. In particular,

$$|x|^2 = x^* x = \sum_{i=1}^{n} |x_i|^2.$$

Clearly $|x|$ is never negative, and is zero only when the vector $x = 0$. Later in the course we will discuss generalizations of this concept in more general vector spaces.

We have already encountered the notion of a Hermitian matrix, characterized by the self-adjoint property $Q^* = Q$. Recall the notation $\mathbb{H}^n$ for the real vector space of complex Hermitian matrices. We now collect some properties and introduce some new definitions, for later use. Everything we will state will apply as well to the set $\mathbb{S}^n$ of real, symmetric matrices.

Our first result about self-adjoint matrices is that their eigenvalues are always *real*. Suppose $Ax = \lambda x$ for nonzero x. Then we have

$$\lambda x^* x = x^* A x = (Ax)^* x = \lambda^* x^* x.$$

Since $x^* x > 0$ we conclude that $\lambda = \lambda^*$.

We say that two vectors $x, y \in \mathbb{C}^n$ are *orthogonal* if

$$y^* x = 0.$$

Given a set of vectors $\{v_1, \dots, v_k\}$ in $\mathbb{C}^n$ we say the vectors are *orthonormal* if

$$v_i^* v_r = \begin{cases} 1, & \text{if } i = r; \\ 0, & \text{if } i \neq r. \end{cases}$$

The vectors are orthonormal if each has unit length and is orthogonal to all the others. It is easy to show that orthonormal vectors are linearly independent, so such a set can have at most n members. If $k < n$, then it is always possible to find a vector v_{k+1} such that $\{v_1, \dots, v_{k+1}\}$ is an orthonormal set. To see this, form the $k \times n$ matrix

$$V_k^* = \begin{bmatrix} v_1^* \\ \vdots \\ v_k^* \end{bmatrix}.$$

The kernel of V_k^* has the nonzero dimension $n - k$, and therefore any element of the kernel is orthogonal to the vectors $\{v_1, \ldots, v_k\}$. We conclude that any element of unit length in $\mathrm{Ker}\,V_k^*$ is a suitable candidate for v_{k+1}. Applying this procedure repeatedly we can generate an orthonormal basis $\{v_1, \ldots, v_n\}$ for $\mathbb{C}^n$.

A square matrix $U \in \mathbb{C}^{n \times n}$ is called *unitary* if it satisfies

$$U^*U = I.$$

From this definition we see that the columns of any unitary matrix forms an orthonormal basis for $\mathbb{C}^n$. Further, since U is square it must be that $U^* = U^{-1}$ and therefore $UU^* = I$. So the columns of U^* also form an orthonormal basis. A key property of unitary matrices is that if $y = Ux$, for some $x \in \mathbb{C}^n$, then the length of y is equal to that of x:

$$|y| = \sqrt{y^*y} = \sqrt{(Ux)^*(Ux)} = \sqrt{x^*U^*Ux} = |x|.$$

Unitary matrices are the only matrices that leave the length of every vector unchanged. We are now ready to state the *spectral theorem* for Hermitian matrices.

Theorem 4. *Suppose H is a matrix in $\mathbb{H}^n$. Then there exist a unitary matrix U and a real diagonal matrix Λ such that*

$$H = U\Lambda U^*.$$

Notice that since $U^* = U^{-1}$ for a unitary U, the above expression is a similarity transformation. Therefore the theorem says that a self-adjoint matrix can be diagonalized by a unitary similarity transformation. Thus the columns of U are all eigenvectors of H. Since the proof of this result assembles a number of concepts from this chapter we provide it below.

Proof. We will use an induction argument. Clearly the result is true if H is simply a scalar, and it is therefore sufficient to show that if the result holds for matrices in $\mathbb{H}^{n-1}$ then it holds for $H \in \mathbb{H}^n$. We proceed with the assumption that the decomposition result holds for $(n-1) \times (n-1)$ Hermitian matrices.

The matrix H has at least one eigenvalue λ_1, and λ_1 is real since H is Hermitian. Let x_1 be an eigenvector associated with this eigenvalue, and without loss of generality we assume it to have length one. Define X to be any unitary matrix with x_1 as its first column, namely,

$$X = [x_1 \cdots x_n].$$

Now consider the product X^*HX. Its first column is given by $X^*Hx_1 = \lambda_1 X^*x_1 = \lambda_1 e_1$, where e_1 is the first element of the canonical basis. Its first row is described by x_1^*HX, which is equal to $\lambda_1 x_1^*X = \lambda_1 e_1^*$, since $x_1^*H = \lambda_1 x_1^*$ because H is self-adjoint. Thus we have

$$X^*HX = \begin{bmatrix} \lambda_1 & 0 \\ 0 & H_2 \end{bmatrix},$$

where H_2 a Hermitian matrix in $\mathbb{H}^{n-1}$. By the inductive hypothesis there exists a unitary matrix X_2 in $\mathbb{C}^{(n-1)\times(n-1)}$ such that $H_2 = X_2\Lambda_2 X_2^*$, where Λ_2 is both diagonal and real. We conclude that

$$H = \left(X \begin{bmatrix} I & 0 \\ 0 & X_2 \end{bmatrix} \right) \begin{bmatrix} \lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \left(\begin{bmatrix} I & 0 \\ 0 & X_2^* \end{bmatrix} X^* \right).$$

The right-hand side gives the desired decomposition.

$\blacksquare$

We remark, additionally, that the eigenvalues of H can be arranged in decreasing order in the diagonal of Λ. This follows directly from the above induction argument: just take λ_1 to be the largest eigenvalue.

We now focus on the case where these eigenvalues have a definite sign. Given $Q \in \mathbb{H}^n$, we say it is *positive definite*, denoted $Q > 0$, if

$$x^* Q x > 0,$$

for all nonzero $x \in \mathbb{C}^n$. Similarly Q is positive *semidefinite*, denoted $Q \geq 0$, if the inequality is nonstrict; and negative definite and negative semidefinite are similarly defined. If a matrix is not positive or negative semidefinite, then it is *indefinite*.

The following properties of positive matrices follow directly from the definition, and are left as exercises:

- If $Q > 0$ and $A \in \mathbb{C}^{n \times m}$, then $A^* Q A \geq 0$. If $\mathrm{Ker}(A) = \{0\}$, then $A^* Q A > 0$.

- If $Q_1 > 0$, $Q_2 > 0$, then $\mu_1 Q_1 + \mu_2 Q_2 > 0$ whenever $\mu_1 > 0$, $\mu_2 \geq 0$. In particular, the set of positive definite matrices is a *convex cone* in $\mathbb{H}^n$, as defined in the previous section.

At this point we may well ask, how can we check whether a matrix is positive definite? The following answer is derived from Theorem 4:

If $Q \in \mathbb{H}^n$, then $Q > 0$ if and only if the eigenvalues of Q are all positive.

Notice in particular that a positive definite matrix is always invertible, and its inverse is also positive definite. Also a matrix is positive semidefinite exactly when none of its eigenvalues are negative; in that case the number of strictly positive eigenvalues is equal to the rank of the matrix.

An additional useful property for positive matrices is the existence of a square root. Let $Q = U\Lambda U^* \geq 0$, in other words the diagonal elements of Λ are non-negative. Then we can define $\Lambda^{\frac{1}{2}}$ to be the matrix with diagonal elements $\lambda_k^{\frac{1}{2}}$, and

$$Q^{\frac{1}{2}} := U\Lambda^{\frac{1}{2}}U^*.$$

Then $Q^{\frac{1}{2}} \geq 0$ (also $Q^{\frac{1}{2}} > 0$ when $Q > 0$) and it is easily verified that $Q^{\frac{1}{2}}Q^{\frac{1}{2}} = Q$.

Having defined a notion of positivity, our next aim is to generalize the idea of ordering to matrices: namely, what does it mean for a matrix to be larger than another matrix? We write

$$Q > S$$

for matrices $Q, S \in \mathbb{H}^n$ to denote that $Q - S > 0$. We refer to such expressions generally as matrix inequalities. Note that for matrices that it may be that neither $Q \leq S$ nor $Q \geq S$ holds; that is, not all matrices are comparable.

We conclude our discussion by establishing a very useful result, known as the Schur complement formula.

Theorem 5. *Suppose that Q, M, and R are matrices and that M and Q are self-adjoint. Then the following are equivalent:*

(a) *The matrix inequalities $Q > 0$ and*

$$M - RQ^{-1}R^* > 0 \text{ both hold.}$$

(b) *The matrix inequality*

$$\begin{bmatrix} M & R \\ R^* & Q \end{bmatrix} > 0 \text{ is satisfied.}$$

Proof. The two inequalities listed in (a) are equivalent to the single block inequality.

$$\begin{bmatrix} M - RQ^{-1}R^* & 0 \\ 0 & Q \end{bmatrix} > 0 \ .$$

Now left- and right-multiply this inequality by the nonsingular matrix

$$\begin{bmatrix} I & RQ^{-1} \\ 0 & I \end{bmatrix}$$

and its adjoint, respectively, to get

$$\begin{bmatrix} M & R \\ R^* & Q \end{bmatrix} = \begin{bmatrix} I & RQ^{-1} \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} M - RQ^{-1}R^* & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} I & 0 \\ Q^{-1}R^* & I \end{bmatrix} > 0.$$

Therefore inequality (b) holds if and only if (a) holds. ∎

We remark that an identical result holds in the negative definite case, replacing all ">" by "<".

Having assembled some facts about self-adjoint matrices, we move on to our final matrix theory topic.

2.3 Singular value decomposition

Here we introduce the singular value decomposition of a rectangular matrix, which will have many applications in our analysis, and is of very significant computational value. The term *singular value decomposition*, or SVD, refers to the product $U\Sigma V^*$ in the statement of the theorem below.

Theorem 6. *Suppose $A \in \mathbb{C}^{m \times n}$ and that $p = \min\{m, n\}$. Then there exist unitary matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ such that*

$$A = U\Sigma V^*,$$

where $\Sigma \in \mathbb{R}^{m \times n}$ and its scalar entries satisfy

(a) *the condition $\sigma_{ir} = 0$, for $i \neq r$;*

(b) *the ordering $\sigma_{11} \geq \sigma_{22} \geq \cdots \geq \sigma_{pp} \geq 0$.*

Proof. Since the result holds for A if and only if it holds for A^*, we assume without loss of generality that $n \geq m$. To start let r be the rank of A^*A, which is Hermitian and therefore by Theorem 4 we have

$$A^*A = V \begin{bmatrix} \bar{\Sigma}^2 & 0 \\ 0 & 0 \end{bmatrix} V^*, \text{ where } \bar{\Sigma} = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix} > 0 \text{ and } V \text{ is unitary.}$$

We also assume that the nonstrict ordering $\sigma_1 \geq \cdots \geq \sigma_r$ holds. Now define

$$J = \begin{bmatrix} \bar{\Sigma} & 0 \\ 0 & I \end{bmatrix}$$

and we have

$$J^{-1}V^*A^*AVJ^{-1} = (AVJ^{-1})^*(AVJ^{-1}) = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

where I_r denotes the $r \times r$ identity matrix. From the right-hand side we see that the first r columns of AVJ^{-1} form an orthonormal set, and the remaining columns must be zero. Thus

$$AVJ^{-1} = \begin{bmatrix} U_1 & 0 \end{bmatrix},$$

where $U_1 \in \mathbb{C}^{m \times r}$. This leads to

$$A = \begin{bmatrix} U_1 & 0 \end{bmatrix} \begin{bmatrix} \bar{\Sigma} & 0 \\ 0 & I \end{bmatrix} V^* = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \bar{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} V^*,$$

where the right-hand side is valid for any $U_2 \in \mathbb{C}^{m \times (m-r)}$. So choose U_2 such that $\begin{bmatrix} U_1 & U_2 \end{bmatrix}$ is unitary. ∎

When $n = m$ the matrix Σ in the SVD is diagonal. When these dimensions are not equal Σ has the form of either

$$\begin{bmatrix} \sigma_{11} & & & 0 \\ & \ddots & & \\ 0 & & \sigma_{mm} & 0 \end{bmatrix} \text{ when } n > m, \text{ or } \begin{bmatrix} \sigma_{11} & & 0 \\ & \ddots & \\ & & \sigma_{nn} \\ 0 & & 0 \end{bmatrix} \text{ when } n < m.$$

The first p non-negative scalars σ_{kk} are called the *singular values* of the matrix A, and are denoted by the ordered set $\sigma_1, \ldots \sigma_p$, where $\sigma_k = \sigma_{kk}$. As we already saw in the proof, the decomposition of the theorem immediately gives us that

$$A^*A = V(\Sigma^*\Sigma)V^* \text{ and } AA^* = U(\Sigma\Sigma^*)U^*,$$

which are singular value decompositions of A^*A and AA^*. But since $V^* = V^{-1}$ and $U^* = U^{-1}$ it follows that these are also the diagonalizations of the matrices. Thus

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_p^2 \geq 0$$

are exactly the p largest eigenvalues of A^*A and AA^*; the remaining eigenvalues of either matrix are all necessarily equal to zero. This observation provides a straightforward method to obtain the singular value decomposition of any matrix A, by diagonalizing the Hermitian matrices A^*A and AA^*.

The SVD of a matrix has many useful properties. We use $\bar{\sigma}(A)$ to denote the largest singular value σ_1, which from the SVD has the following property.

$$\bar{\sigma}(A) = \max\{|Av| : \ v \in \mathbb{C}^n \text{ and } |v| = 1\}.$$

Namely, it gives the maximum magnification of length a vector v can experience when acted upon by A.

Finally, partition $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}$ and $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and suppose that A has r nonzero singular values. Then

$$\text{Im}A = \text{Im}\begin{bmatrix} u_1 & \cdots u_r \end{bmatrix} \text{ and } \text{Ker}A = \text{Im}\begin{bmatrix} v_{r+1} & \cdots & v_n \end{bmatrix}.$$

That is, the SVD provides an orthonormal basis for both the image and kernel of A. Furthermore notice that the rank of A is equal to r, precisely the number of nonzero singular values.

Notes and references

Given its ubiquitous presence in analytical subjects, introductory linear algebra is the subject of many excellent books; one choice is [5]. For an advanced treatment from a geometric perspective the reader is referred to [2].

Two excellent sources for matrix theory are [3] and the companion work [4]. For information and algorithms for computing with matrices see [1].

References

[1] G.H. Golub and C.F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 1996.

[2] W.H. Greub. *Linear Algebra.* Springer, 1981.

[3] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1991.

[4] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1995.

[5] G. Strang. *Linear Algebra and its Applications.* Academic Press, 1980.