SELECTED TOPICS IN LINEAR ALGEBRA

In this note, we discuss certain aspects of bounded linear operators between finite-dimensional normed spaces X and Y. We begin by establishing the fundamental fact that $B(X,Y) \cong \mathcal{M}_{m\times n}(\mathbb{C})$ if X and Y are complex vector spaces of dimensions n and m, respectively. We then go on to discuss spectral theory for linear operators between finite-dimensional vector spaces, and finally consider certain useful matrix decompositions.

1. Bounded linear transformations between finite-dimensional spaces

We have seen that any finite-dimensional vector space X of dimension n has a set of n linearly independent spanning vectors $\{x_1, \dots, x_n\}$. We call this set a (Hamel) basis for X, and any other basis must necessarily have the same number of spanning vectors. As a consequence, we have the following.

Theorem 1. Let X be a complex vector space with basis $\{e_1, \dots e_n\}$. Then X is isomorphic to $\mathbb{C}^n$,

$$X \cong \mathbb{C}^n.$$

Similarly, if X is a real vector space of dimension n, then $X \cong \mathbb{R}^n$.

Proof. Let X be a complex vector space. By the definition of a basis, any $x \in X$ has a unique representation

$$x = \sum_{j=1}^{n} a_j e_j, \quad a_j \in \mathbb{C}.$$

Let $T : X \to \mathbb{C}^n$ be the mapping defined by

$$Tx = (a_1, a_2, \dots, a_n).$$

T is *linear*: if $x = \sum a_j e_j$ and $y = \sum b_j e_j$, then

$$T(\alpha x+\beta y) = (\alpha a_1+\beta b_1, \dots, \alpha a_n+\beta b_n) = \alpha(a_1, \dots, a_n)+\beta(b_1, \dots, b_n) = \alpha Tx+\beta Ty.$$

T is *surjective*:

$$\text{for any } (a_1, \dots, a_n) \in \mathbb{C}^n \quad \text{there exists } x = \sum_{j=1}^{n} a_j e_j \quad \text{such that } Tx = (a_1, \dots, a_n).$$

T is *injective*:

$$Tx = Ty \quad \Leftrightarrow \quad a_j = b_j \text{ for each } j = 1, 2, \dots, n \quad \Rightarrow x = y.$$

Thus, T is a vector space isomorphism. The same proof works when X is a real vector space. □

Now let $T : X \to Y$ be a linear operator between finite-dimensional vector spaces X and Y. We make the useful observation that T *is determined by its action on any basis of* X.

Lemma 2. Let X be a finite-dimensional vector space with basis $\{e_1, \dots, e_n\}$. For any values $y_1, \dots, y_n \in Y$ there exists precisely one linear transformation $T : X \to Y$ such that
$$Te_j = y_j, \quad j = 1, \dots, n.$$

Proof. Any $x \in X$ has a unique representation $x = \sum_{j=1}^n x_j e_j$. Define T as
$$Tx = \sum_{j=1}^n x_j y_j.$$
Then $Te_j = y_j$, and T is clearly linear (since it acts as matrix multiplication with a $(1 \times n)$ matrix). Finally, T is also unique: If $S : X \to Y$ is a linear map satisfying $Se_j = y_j$, then
$$Sx = S\left(\sum_{j=1}^n x_j e_j\right) = \sum_{j=1}^n x_j Se_j = \sum_{j=1}^n x_j y_j = Tx \quad \text{for all } x \in X,$$
so $S = T$. $\square$

Example 3. Let $T : \mathbb{C}^n \to \mathbb{C}^m$ be the linear map given by matrix multiplication
$$Tx = Ax, \quad A \in \mathcal{M}_{m\times n}(\mathbb{C}).$$
Then the columns A_j of the matrix A are determined by the action on the standard basis $\{e_j\}_{j=1}^n$:
$$Ae_j = A_j, \quad j = 1, \dots, n.$$
Note that A_j plays the role of y_j in the above lemma.

Remark 1. If X and Y are both finite-dimensional normed spaces, then any linear transformation $T : X \to Y$ is automatically bounded. We therefore use $B(X, Y)$ to denote the linear transformations from X to Y when X and Y are finite-dimensional, even though we originally introduced the notation for *bounded* linear transformations.

We are now equipped to clarify the link between matrices and linear transformations. We have already seen that an $(m\times n)$ matrix A defines a linear transformation from $\mathbb{C}^n$ to $\mathbb{C}^m$ by matrix multiplication. On the other hand, any linear transformation between finite-dimensional vector spaces can be represented in matrix form once we have chosen bases for X and Y.

Theorem 4. Let X and Y be complex vector spaces of dimension n and m, respectively. Then $B(X, Y) \cong \mathcal{M}_{m\times n}(\mathbb{C})$. Similarly, if X and Y are real vector spaces, then $B(X, Y) \cong \mathcal{M}_{m\times n}(\mathbb{R})$.

Proof. Since $X \cong \mathbb{C}^n$ and $Y \cong \mathbb{C}^m$, it suffices to prove the statement for these choices of X and Y. Let $\{e_j\}_{j=1}^n$ be the standard basis for $\mathbb{C}^n$. Then
$$T : \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \to \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

is a linear transformation $\mathbb{C}^n \to \mathbb{C}^m$ satisfying

$$Te_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

According to Lemma 2, there is precisely one such $T \in B(\mathbb{C}^n, \mathbb{C}^m)$. As we can choose the columns A_j of A to be any elements $A_j \in \mathbb{C}^m$, we get all possible $T \in B(\mathbb{C}^n, \mathbb{C}^m)$. $\square$

Example 5. The differential operator $\frac{d}{dx}$ is a linear operator on $\mathcal{P}_n(\mathbb{R})$. Since $\mathcal{P}_2(\mathbb{R}) \cong \mathbb{R}^3$ via the vector space isomorphism

$$\sum_{j=0}^{2} a_j x^j \to (a_0, a_1, a_2),$$

we see that

$$\frac{d}{dx} : \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ 2a_2 \\ 0 \end{bmatrix}$$

expresses the derivation

$$\frac{d}{dx}(a_0 + a_1 x + a_2 x^2) = a_1 + 2a_2 x + 0x^2.$$

Notice that in the proof of Theorem 4, the matrix representation of T depends on the choice of basis of the space X; in particular, we use the standard basis $\{e_j\}_{j=1}^n$ for $X \cong \mathbb{C}^n$. However, we may equally well choose a different basis $\{f_j\}_{j=1}^n$. Let us see how this affects the matrix representation of T.

We focus on the case when $X = Y = \mathbb{C}^n$. For any $x \in X$, we have that

$$x = \sum_{j=1}^{n} \alpha_j e_j = \sum_{j=1}^{n} \beta_j f_j$$

for unique scalars $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$. Let us now denote by x_e the vector $\sum_{j=1}^n \alpha_j e_j = (\alpha_1, \ldots, \alpha_n)^\top$, by x_f the vector $\sum_{j=1}^n \beta_j e_j = (\beta_1, \ldots, \beta_n)^\top$, and by F the $(n \times n)$ matrix whose columns are the basis vectors $f_1, \ldots, f_n$. We then have $f_j = Fe_j$, and

$$x_e = x = \sum \beta_j f_j = \sum \beta_j F e_j = F\left(\sum \beta_j e_j\right) = F(x_f).$$

The matrix F is invertible, so we have

$$x_e = F x_f \quad \text{ and } x_f = F^{-1} x_e.$$

Now let $T : \mathbb{C}^n \to \mathbb{C}^n$, and suppose A_e is the matrix representation of T in the standard basis $\{e_j\}$. What is then its matrix representation A_f in the basis $\{f_j\}$? Defining y_e and y_f as above, we have that

$$y_e = A_e x_e \quad \Leftrightarrow \quad y_f = F^{-1} y_e = F^{-1} A_e x_e = F^{-1} A_e F x_f.$$

Thus,

$$A_f = F^{-1} A_e F$$

is the matrix representation of T in the basis $\{f_j\}$.

Recall that the kernel of a linear operator $T : X \to Y$,

$$\ker(T) = \{x \in X \,:\, Tx = 0\},$$

is a vector subspace of X, whereas the range of T,

$$\operatorname{ran}(T) = \{y \in Y \,:\, Tx = y \text{ for some } x \in X\},$$

is a vector subspace of Y. When X and Y are finite-dimensional, and T is represented by a matrix $A \in \mathcal{M}_{m\times n}(\mathbb{C})$, then these subspaces are equivalently given by the so-called *null space* and *column space* of the matrix A:

- **Null space of A:** The kernel of T represented by A is clearly equal to the null space of A. We have

$$x \in \ker(T) \quad\Leftrightarrow\quad Ax = 0 \quad\Leftrightarrow\quad \sum_{j=1}^{n} a_{ij}x_j = 0 \quad \forall i = 1,\dots m$$
$$\Leftrightarrow (x_1,\dots,x_n) \perp (\overline{a_{i1}},\dots,\overline{a_{in}}) \quad \forall i = 1,\dots m.$$

  Note that the final line above tells us that the kernel of T (or null space of A) is the space of vectors $x \in \mathbb{C}^n$ orthogonal to the conjugated row vectors of A. We call the dimension of this subspace the **nullity** of T.
- **Column space of A:** The column space of A is the range of T. Since

$$Tx = Ax = A_1x_1 + \cdots + A_nx_n,$$

  where $A_j = (a_{1j},\dots,a_{mj})^\top$ is the jth column vector of A, we have that

$$\operatorname{ran}(T) = \{Ax \,:\, x \in \mathbb{C}^n\} = \operatorname{span}\{A_1,\dots,A_n\}.$$

  This is precisely the column space of A. We call the dimension of this subspace the **rank** of T.
- **Row space of A:** The row space of A is the space spanned by the row vectors of A. Note that

$$\text{row space of } A = \text{column space of } A^\top,$$

  where $A^\top$ is the transpose of A. The following result follows almost immediately.

Proposition 6. Let $A \in \mathcal{M}_{m\times n}(\mathbb{C})$. Then

$$\ker(A) \perp \operatorname{ran}(\overline{A}^\top).$$

In words, the kernel of A is orthogonal to the range of $\overline{A}^\top$.

Proof. We have just seen that the kernel, or null space, of A is orthogonal to the row space of $\overline{A}$. This is in turn equal to the column space, or range, of $\overline{A}^\top$. $\square$

Finally let us state the *rank-nullity theorem* and see some important consequences.

Theorem 7. Let $T \in B(\mathbb{C}^n, \mathbb{C}^m)$. Then

$$\dim\ker(T) + \dim\operatorname{ran}(T) = n.$$

Proof. Pick a basis $\{e_1, \ldots, e_k\}$ for $\ker T$. If $k = n$ and $\ker(T) = \mathbb{C}^n$, we are done, since then $\operatorname{ran}(T) = \{0\}$, and

$$\dim \ker(T) + \dim \operatorname{ran}(T) = n + 0 = n.$$

Now assume $k < n$, and extend $\{e_1, \ldots, e_k\}$ to a basis $\{e_1, \ldots, e_k, f_1, \ldots, f_l\}$ for $\mathbb{C}^n$. This can be done in the following way: pick $f_1 \notin \operatorname{span}\{e_1, \ldots, e_k\}$. Then $\{e_1, \ldots, e_k, f_1\}$ is linearly independent. If this set of vectors spans all of $\mathbb{C}^n$, we stop. If not, we pick $f_2 \notin \operatorname{span}\{e_1, \ldots, e_k, f_1\}$. This process will necessarily stop when $k + l = n$ (because any linearly independent set of vectors spanning $\mathbb{C}^n$ has precisely n elements).

To finish the proof, we prove that $Tf = \{Tf_1, \ldots, Tf_l\}$ is a basis for $\operatorname{ran}(T)$. We observe first that Tf is linearly independent:

$$\begin{aligned} \sum_{j=1}^{l} a_j T f_j = T\left(\sum_{j=1}^{l} a_j f_j\right) = 0 \quad &\Leftrightarrow \quad \sum_{j=1}^{l} a_j f_j \in \ker T \\ &\Leftrightarrow \quad a_j = 0 \text{ for } j = 1, 2, \ldots, l. \end{aligned}$$

The last implication follows from the fact that by construction, no nonzero linear combination of vectors f_j lies in $\ker(T)$. Now let us see that Tf spans $\operatorname{ran}(T)$. By the linearity of T we have

$$\begin{aligned} \operatorname{ran}(T) = \{Tx \,:\, x \in \mathbb{C}^n\} &= \left\{ T\Big(\sum_{j=1}^{k} a_j e_j + \sum_{j=1}^{l} b_j f_j\Big) \,:\, a_j, b_j \in \mathbb{C} \right\} \\ &= \left\{ T\Big(\sum_{j=1}^{k} a_j e_j\Big) + T\Big(\sum_{j=1}^{l} b_j f_j\Big) \,:\, a_j, b_j \in \mathbb{C} \right\} \\ &= \left\{ \sum_{j=1}^{l} b_j T f_j \,:\, b_j \in \mathbb{C} \right\}. \end{aligned}$$

Hence $\{Tf_1, \ldots, Tf_l\}$ is a basis for $\operatorname{ran}(T)$, and

$$\dim \ker(T) + \dim \operatorname{ran}(T) = k + l = n.$$

□

An immediate consequence of the rank-nullity theorem is that a linear map $T : \mathbb{C}^n \to \mathbb{C}^n$ is injective if and only if it is surjective.

Corollary 8. Let $T \in B(\mathbb{C}^n, \mathbb{C}^n)$. Then the following are equivalent.

i) T is injective ($\ker(T) = \{0\}$).

ii) T is surjective ($\operatorname{ran}(T) = \mathbb{C}^n$).

iii) T is invertible.

iv) The matrix representation A of T (in any given basis) is invertible.

v) For any $b \in \mathbb{C}^n$, the system $Ax = b$ has a unique solution.

2. Eigenvalues and eigenvectors

In the next section, we will discuss similarity transformations between matrices and establish Schur's triangulation lemma. This requires that we recall some properties of eigenvalues and eigenvectors.

Definition 2. Let $T : X \to X$ be a linear transformation (for example, $T : \mathbb{C}^n \to \mathbb{C}^n$ given by a matrix A). Then the scalar $\lambda \in \mathbb{C}$ is called an *eigenvalue* of T if there exists a nonzero vector $v \in X$ such that

$$Tv = \lambda v.$$

The vector v is called an *eigenvector* corresponding to the eigenvalue λ.

Definition 3. Let $T : X \to X$ be a linear transformation. The set $\sigma(T)$ of scalars satisfying

$$\sigma(T) = \{z \in \mathbb{C} : T - zI \text{ is not invertible}\}$$

is called the *spectrum* of T.

Proposition 9. For a linear transformation represented by $A \in \mathcal{M}_{n\times n}(\mathbb{C})$,

$$\sigma(A) = \{\lambda \in \mathbb{C} : \det(A - \lambda I) = 0\}$$

consists of the roots $(\lambda_1, \ldots, \lambda_n)$ of the *characteristic polynomial* $p_A(\lambda) = \det(A - \lambda I)$; these are precisely the eigenvalues of A.

Proof. Exercise. □

We recall the following notions related to eigenvalues of a matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$:

- The multiplicity of a root λ of $p_A(\lambda)$ is the **algebraic multiplicity** of the eigenvalue λ.
- The eigenvectors corresponding to an eigenvalue λ span a subspace of $\mathbb{C}^n$,

  $$\ker(A - \lambda I),$$

  called the **eigenspace** of λ. The dimension of this space is the **geometric multiplicity** of λ.

Definition 4. Suppose that the matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ has n linearly independent eigenvectors. If these eigenvectors are the columns of a matrix S, then $S^{-1}AS$ is a diagonal matrix Λ with the eigenvalues of A on its diagonal:

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

This is called the *diagonalization* of A.

Note that the definition above is a simple consequence of the fact that if A has eigenvectors $\lambda_1, \dots, \lambda_n$ with associated, and linearly independent, eigenvectors $v_1, \dots, v_n$, then we may rewrite the set of equations

$$Av_1 = \lambda_1 v_1$$
$$\vdots$$
$$Av_n = \lambda_n v_n$$

in matrix form $AS = S\Lambda$, where S is the matrix with column vectors $v_1, \dots v_n$. Since the vectors v_j are linearly independent, the matrix S is invertible.

Remark 5. i) If the eigenvectors $v_1, \dots, v_k$ correspond to *different* eigenvalues $\lambda_1, \dots, \lambda_k$, then they are automatically linearly independent. Therefore any $(n \times n)$ matrix with n distinct eigenvalues can be diagonalized.

ii) The diagonalization is not unique, as any eigenvector v_j can be multiplied by a constant and remains an eigenvector. Repeated eigenvalues leave even more freedom. For the trivial example $A = I$, any invertible S will do, since $S^{-1}IS = I$ is diagonal.

iii) Not all matrices possess n linearly independent eigenvectors, so not all matrices are diagonalizable. The standard example of a "defective" matrix is

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

Exercise: Show that this matrix cannot be diagonalized.

Recall that a map $T \in B(\mathbb{C}^n)$ is called

i) normal if $TT^* = T^*T$,

ii) unitary if $T^* = T^{-1}$, and

iii) self-adjoint or Hermitian if $T = T^*$.

Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ be the matrix representation of T. We have seen that T^* has matrix representation $\overline{A}^\top$. Accordingly, we let $A^* = \overline{A}^\top$, and call the matrix A

i) normal if $AA^* = A^*A$,

ii) unitary if $A^* = A^{-1}$, and

iii) Hermitian if $A = A^*$.

We make certain observations on the eigenvalues and eigenvectors of Hermitian and unitary matrices.

Proposition 10. Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a Hermitian matrix. Then all eigenvalues of A are real, and any two eigenvectors corresponding to different eigenvalues are orthogonal.

Proof. Let λ be an eigenvalue of A, and v the corresponding eigenvector. Then

$$\langle Av, v\rangle = \langle v, A^*v\rangle = \langle v, Av\rangle,$$

and since the inner product is conjugate symmetric ($\langle x, y\rangle = \overline{\langle y, x\rangle}$), it follows that $\langle Av, v\rangle$ is real-valued. On the other hand, we have

$$\langle Av, v\rangle = \langle \lambda v, v\rangle = \lambda \|v\|^2,$$

and since both $\langle Av, v\rangle$ and $\|v\|^2$ are real, the eigenvalue λ must be real-valued.

Now let λ_1 and λ_2 be two distinct eigenvalues of A, with corresponding eigenvectors x and y:

$$Ax = \lambda_1 x \quad \text{and} \quad Ay = \lambda_2 y.$$

Then

$$\lambda_1 \langle x, y\rangle = \langle Ax, y\rangle = \langle x, A^* y\rangle = \langle x, Ay\rangle = \lambda_2 \langle x, y\rangle,$$

and it follows that we must have $\langle x, y\rangle = 0$, meaning $x \perp y$. □

Proposition 11. Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a unitary matrix. Then every eigenvalue of A has absolute value $|\lambda| = 1$. Moreover, eigenvectors corresponding to different eigenvalues are orthogonal.

Proof. Let λ be an eigenvalue of A and v the corresponding eigenvector. Then

$$\langle Av, Av\rangle = \left\langle v, A^{-1}Av\right\rangle = \langle v, v\rangle = \|v\|^2.$$

On the other hand

$$\langle Av, Av\rangle = \langle \lambda v, \lambda v\rangle = |\lambda|^2 \|v\|^2,$$

and it follows that $|\lambda| = 1$.

Now let λ_1 and λ_2 be two distinct eigenvalues of A, with corresponding eigenvectors x and y:

$$Ax = \lambda_1 x \quad \text{and} \quad Ay = \lambda_2 y.$$

Then

$$\langle x, y\rangle = \langle Ax, Ay\rangle = \lambda_1 \overline{\lambda_2} \langle x, y\rangle,$$

which implies that either $\lambda_1\overline{\lambda_2} = 1$ or $\langle x, y\rangle = 0$. Multiplying both sides of the first equality by λ_2, we get

$$\lambda_1 |\lambda_2|^2 = \lambda_1 = \lambda_2.$$

This is a contradiction, as the two eigenvalues λ_1 and λ_2 are distinct. Thus the condition $\lambda_1\overline{\lambda_2} = 1$ cannot hold, and we conclude that $\langle x, y\rangle = 0$. □

3. Similarity transformations and Schur's lemma

We saw in the previous section that if a matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ has n linearly independent eigenvectors, then it has a diagonalization $\Lambda = S^{-1}AS$, where the matrix S has the eigenvectors of A as its columns. Let us now look at *all* combinations $M^{-1}AM$ formed with an invertible matrix M on the right and its inverse on the left.

Definition 6. We say that the matrices A and B in $\mathcal{M}_{n\times n}(\mathbb{C})$ are *similar* if there exists an invertible matrix M such that

$$B = M^{-1}AM.$$

The matrix M provides a *similarity transformation* from A to B. If M can be chosen unitary, then we say that A and B are *unitarily equivalent*.

At first glance it might not be obvious why we would be interested in similarity transforms, but the general idea is that a matrix B similar to A shares many properties with A, yet B might have a much more useful form than A.

Example 12. Similarity transformations arise in systems of differential equations, when a "change of variables" $u = Mv$ introduces the new unknown v:

$$\frac{du}{dt} = Au \quad \text{becomes} \quad M\frac{dv}{dt} = AMv, \quad \text{or} \quad \frac{dv}{dt} = M^{-1}AMv.$$

The new matrix in the equation is $M^{-1}AM$. In the special case that M is the eigenvector matrix S, the system becomes completely uncoupled, because $\Lambda = S^{-1}AS$ is diagonal. This is a maximal simplification, but other M's can also be useful. We try to make $M^{-1}AM$ easier to work with than A.

Note also that the similar matrix $B = M^{-1}AM$ is closely connected to A if we go back to linear transformations. Recall the key idea: Every linear transformation is represented by a matrix. However, this matrix depends on the choice of basis. If we recall our observations on page 91, we see that if we change the basis from $e = \{e_1, \dots, e_n\}$ to Me, then we change the matrix from A to B.

We will try to shed light on the following two questions:

(1) What do similar matrices $M^{-1}AM$ have in common?

(2) By picking M in a clever way, can we ensure that $M^{-1}AM$ has a special form?

Our first observation is that similar matrices have the same eigenvalues.

Lemma 13. If $B = M^{-1}AM$, then A and B have the same eigenvalues.

Proof. We consider the characteristic polynomial of B:

$$\begin{aligned} p_B(z) &= \det(M^{-1}AM - zI) = \det(M^{-1}AM - M^{-1}Mz) \\ &= \det(M^{-1})\det(AM - zM) = \det(M^{-1})\det(A - zI)\det(M) = p_A(z) \end{aligned}$$

It follows that A and B must have the same eigenvalues. □

Let us now focus on question (2) above. We restrict our attention to the case where $M = U$ is unitary (meaning $U^* = U^{-1}$, which necessarily implies that U has orthonormal columns). Unless the eigenvectors of A are orthogonal, it is impossible for $U^{-1}AU$ to be diagonal. However, Schur's lemma states the very useful fact that $U^{-1}AU$ can always achieve a triangular form.

Theorem 14 (Schur's triangulation lemma)**.** For any $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ there exists a unitary matrix U such that

$$U^{-1}AU = U^*AU = T,$$

where T is an upper triangular matrix, and where the eigenvalues of A appear (with multiplicity) along the diagonal of T.

We recall that an upper triangular matrix is one with only zeros below its diagonal:

$$T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$$

Proof of Theorem 14. We proceed by induction on $n \geq 1$. For $n = 1$ there is nothing to do. Suppose now that the result is true for matrices up to size $n-1$ ($n \geq 2$). Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ with eigenvalues $\lambda_1, \dots, \lambda_n$ (counting multiplicities). Consider an eigenvector v_1 associated to λ_1, and assume that $\|v_1\| = 1$. We use it to form an orthonormal basis $(v_1, \dots, v_n)$, and we let V be the unitary matrix with v_j as its columns. The matrix A is equivalent to the matrix of the linear map $x \to Ax$ relative to the basis V, i.e.

$$A = V \left[\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \tilde{A} & \\ 0 & & & \end{array}\right] V^{-1} =: V\tilde{T}V^{-1}, \tag{1}$$

The matrices A and $\tilde{T}$ are similar, so they have the same eigenvalues. We see that $p_A(z) = (\lambda_1 - z)p_{\tilde{A}}(z)$, so the eigenvalues of the matrix $\tilde{A}$ must be $\lambda_2, \dots, \lambda_n$. By the induction hypothesis there exists an $(n-1)\times(n-1)$ unitary matrix $\tilde{W}$ such that

$$\tilde{A} = \tilde{W} \begin{bmatrix} \lambda_2 & * & \cdots & * \\ 0 & \ddots & & \vdots \\ \vdots & & & * \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \tilde{W}^{-1}.$$

By a tedious calculation it is not difficult to check that if we let

$$W := \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \tilde{W} & \\ 0 & & & \end{array}\right],$$

then

$$W^{-1}\tilde{T}W = \begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & \ddots & & \vdots \\ \vdots & & & * \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} =: T.$$

It follows that $\tilde{T} = WTW^{-1}$, and inserting this in equation (1), we get

$$A = VWTW^{-1}V^{-1} = (VW)T(VW)^{-1}.$$

Finally, we observe that W and V are both unitary, so VW is also unitary, and the matrix T is of the desired form. $\square$

As an immediate consequence of Schur's lemma, we have the following.

Corollary 15 (Spectral theorem for Hermitian matrices)**.** Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ be Hermitian. Then A can be diagonalized, meaning there exists a diagonal matrix Λ (with the eigenvalues of A on the diagonal) and a unitary matrix U such that

$$A = U\Lambda U^{-1} = U\Lambda U^*.$$

Proof. By Schur's lemma there exists a unitary matrix U and a triangular matrix T such that

$$A = UTU^*.$$

If A is Hermitian, then $A = A^*$, and it follows that

$$A = A^* = (UTU^*)^* = UT^*U^*.$$

This means T must also be Hermitian in addition to triangular, which forces T to be diagonal. □

The corollary above is known as the spectral theorem for Hermitian matrices. However, we will see in the following section that this result can be extended to all normal matrices.

4. The spectral theorem

We have seen that a Hermitian matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ can be diagonalized. This is a sufficient, but not a necessary, condition for diagonalization. The following theorem, known as the Spectral Theorem, tells us precisely which matrices can be diagonalized.

Theorem 16 (Spectral Theorem)**.** Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$. Then A is diagonalizable, meaning there exists a diagonal matrix Λ (with the eigenvalues of A on the diagonal) and a unitary matrix U such that

$$A = U\Lambda U^{-1} = U\Lambda U^*,$$

if and only if A is normal (meaning $AA^* = A^*A$).

Before proving Theorem 16, we establish the following preliminary result.

Lemma 17. An upper triangular matrix is normal if and only if it is diagonal.

Proof. $(\Rightarrow)$: Suppose T is an upper triangular matrix. Then the (n,n)-th entry of TT^* is $|t_{nn}|^2$, while the (n,n)-th entry of T^*T is $|t_{nn}|^2 + \sum_{i=1}^{n-1} |t_{in}|^2$. If T is normal, then these two entries have to be the same. Hence $t_{in} = 0$ for $i = 1, ..., n-1$. Repeating this argument for the entries $(n-1, n-1), ...(2,2), (1,1)$ gives that T is diagonal.
$(\Leftarrow)$: If T is diagonal, then T is certainly normal. □

Proof of Theorem 16. By Schur's lemma, there exists a unitary matrix U and an upper triangular matrix T such that

$$U^*AU = U^{-1}AU = T.$$

We observe that the matrix T is normal if A is normal, since

$$\begin{aligned}TT^* = (U^*AU)(U^*AU)^* = U^*AUU^*A^*U &= U^*AA^*U \\ &= U^*A^*AU = U^*A^*UU^*AU = T^*T,\end{aligned}$$

and similarly A is normal if T is normal. Finally, by Lemma 17, T is normal if and only if it is diagonal. We know from Schur's lemma that we must have

$$T = \Lambda,$$

where Λ is the matrix with the eigenvalues of A on its diagonal. Finally, we observe that it follows from

$$AU = U\Lambda$$

that the columns of U must be the (orthonormal) eigenvectors of A. □

5. Singular value decomposition and applications

Let $A \in \mathcal{M}_{m\times n}(\mathbb{C})$. If $m \neq n$, it no longer makes sense to ask if A can be diagonalized. However, one can raise the question of whether there exist two *different* unitary matrices U and V such that

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

and where Σ is a diagonal (but rectangular) matrix. It turns out that the answer to this question is yes, and that the specific factorization, known as the *singular value decomposition*, is closely related to the diagonalization of the normal matrix AA^* (or similarly A^*A). Before we state the singular value decomposition in detail and prove its existence, let us briefly discuss positive definite matrices.

Definition 7. A self-adjoint matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is said to be *positive definite* if

$$\langle Ax, x\rangle > 0, \qquad \text{for all nonzero } x \in \mathbb{C}^n.$$

Similarly, if A satisfies the weaker condition

$$\langle Ax, x\rangle \geq 0, \qquad \text{for all nonzero } x \in \mathbb{C}^n,$$

the A is said to be *positive semi-definite*.

A useful test for positive definiteness (or semi-definiteness) is to consider the eigenvalues of the matrix in question.

Proposition 18. A self-adjoint matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is positive definite if and only if all its eigenvalues are positive. Similarly, A is positive semi-definite if and only if all its eigenvalues are non-negative.

Proof. ($\Leftarrow$): Suppose A is positive definite. Then

$$\langle Ax, x\rangle > 0 \qquad \text{for all nonzero } x \in \mathbb{C}^n.$$

In particular, this holds for any eigenvector of A. Let x be an eigenvector associated to the eigenvalue λ. We have

$$\langle Ax, x\rangle = \langle \lambda x, x\rangle = \lambda\|x\|^2 > 0,$$

and it follows that $\lambda > 0$.

($\Rightarrow$): By the Spectral Theorem, there exists a unitary matrix U such that

$$A = U^* \Lambda U,$$

and where Λ is a diagonal matrix with the positive eigenvalues of A on its diagonal. It follows that

$$\langle Ax, x \rangle = \langle U^* \Lambda U x, x \rangle = \langle \Lambda U x, U x \rangle .$$

Now let $y := Ux \in \mathbb{C}^n$. We then have

$$\langle Ax, x \rangle = \langle \Lambda y, y \rangle = \lambda_1 |y_1|^2 + \cdots \lambda_n |y_n|^2,$$

which is greater than zero for all nonzero $y \in \mathbb{C}^n$. Finally note that $y = 0$ if and only if $x = 0$. □

An important pair of self-adjoint, positive semi-definite matrices is AA^* and A^*A for any given $A \in \mathcal{M}_{m \times n}(\mathbb{C})$. The following result follows almost immediately from the proposition above.

Corollary 19. Let $A \in \mathcal{M}_{m \times n}(\mathbb{C})$. Then the $(n \times n)$ matrix A^*A and the $(m \times m)$ matrix AA^* are self-adjoint with non-negative eigenvalues, and the positive eigenvalues of the two matrices coincide.

For the proof of Corollary 19 we need the following lemma, which we state without proof.

Lemma 20. For any $A \in \mathcal{M}_{m \times n}(\mathbb{C})$ and $B \in \mathcal{M}_{n \times m}(\mathbb{C})$, the matrices AB and BA have the same non-zero eigenvalues.

Proof of Corollary 19. It is clear that AA^* and A^*A are both self-adjoint. Moreover, we have that

$$\|Ax\|^2 = \langle Ax, Ax \rangle = \langle A^*Ax, x \rangle \geq 0,$$

so A^*A is clearly positive semi-definite. Running the same argument with $\|A^*x\|$ shows that also AA^* is positive semi-definite. By Proposition 18, the eigenvalues of both matrices are non-negative, and by the preceeding lemma it finally follows that the positive eigenvalues of the two matrices coincide. □

Let us now return to the so-called singular value decomposition of a matrix.

Definition 8. Let $A \in \mathcal{M}_{m \times n}(\mathbb{C})$ have rank r. Let $\sigma_1^2 \geq \cdots \geq \sigma_r^2$ be the positive eigenvalues of A^*A. The scalars $\sigma_1, \ldots, \sigma_r$ are called the *positive singular values* of A.

Since the matrix A^*A is of size $n \times n$, it has n eigenvalues. Those that are not positive are necessarily equal to zero, and accordingly the matrix A has $n - r$ singular values $\sigma_j = 0$, $j = r+1, \ldots, n$. As we have just established that AA^* and A^*A have the same nonzero eigenvalues, one may choose either one for determining the positive singular values of A.

Theorem 21 (Singular Value Decomposition)**.** Suppose $A \in \mathcal{M}_{m \times n}(\mathbb{C})$ is of rank r, and let $\sigma_1 \geq \cdots \geq \sigma_r$ be the positive singular values of A. Let Σ be the $(m \times n)$ matrix defined by

$$\Sigma_{ij} = \begin{cases} \sigma_i & \text{if } i = j \leq r \\ 0 & \text{otherwise.} \end{cases}$$

Then there exists an $(m \times m)$ unitary matrix U and an $(n \times n)$ unitary matrix V such that

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

Through the proof of Theorem 21 below, we will see that the columns of V are the (orthonormal) eigenvectors of A^*A.

Proof. The matrix A^*A is self-adjoint with positive eigenvalues $\sigma_1^2 \geq \cdots \geq \sigma_r^2$ and $(n-r)$ eigenvalues equal to zero. Thus, by the Spectral Theorem, there exists an $(n \times n)$ unitary matrix V such that

$$V^*A^*AV = (AV)^*(AV) = D, \tag{2}$$

where $D = \Sigma^*\Sigma$ is the $(n \times n)$ diagonal matrix with

$$D_{ii} = \sigma_i^2, \quad i = 1, \ldots, r,$$

and zeros elsewhere. It is clear from (2) that the (i,j)th entry of V^*A^*AV is the inner product of columns i and j in AV. Thus, the columns $(AV)_j$ of AV are pairwise orthogonal. Moreover, for $1 \leq j \leq r$, the length of $(AV)_j$ is σ_j. Let U_r denote the $(m \times r)$ matrix with $(AV)_j/\sigma_j$ as its jth column. Complete U_r to an $(m \times m)$ unitary matrix U by finding an orthonormal basis for the orthogonal complement of (the column space of) U_r, and using these basis vectors as the last $(m-r)$ columns in U. We then have

$$AV = U\Sigma \quad \Leftrightarrow \quad A = U\Sigma V^*.$$

□

Remark 9. Since only the first r diagonal entries of Σ are nonzero, we see that the last $(m-r)$ columns of U, and likewise the last $(n-r)$ columns of V, are superfluous. As a consequence, we have that a given matrix A has an SVD where the diagonal matrix Σ is uniquely determined, but the unitary matrices U and V are *not*.

Example 22. Let us determine the singular value decomposition of

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}.$$

The procedure for finding the SVD is as follows: We begin by determining the positive eigenvalues of A^*A (or similarly AA^*). We have

$$A^*A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$$

The positive eigenvalues of this matrix are $\sigma_1^2 = 25$ and $\sigma_2^2 = 9$. The last eigenvalue is $\sigma_3^2 = 0$. Since A^*A is self-adjoint (or Hermitian), the eigenvectors corresponding to σ_1^2, σ_2^2 and σ_3^2 are necessarily orthogonal. We find these eigenvectors, and choose them to have length 1:

$\underline{\sigma_1^2 = 25:}$

$$A^*A - 25I = \begin{bmatrix} 13-25 & 12 & 2 \\ 12 & 13-25 & -2 \\ 2 & -2 & 8-25 \end{bmatrix} \sim \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & -\frac{17}{2} \end{bmatrix},$$

and solving for $A^*A - 25I = 0$, we find that $v_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}$ is a normalized eigenvector.

$\underline{\sigma_2^2 = 9}$:

$$A^*A - 9I = \begin{bmatrix} 13-9 & 12 & 2 \\ 12 & 13-9 & -2 \\ 2 & -2 & 8-9 \end{bmatrix} \sim \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & \frac{1}{4} \\ 1 & 0 & -\frac{1}{4} \end{bmatrix},$$

and solving for $A^*A - 9I = 0$, we find that $v_2 = \begin{pmatrix} \frac{\sqrt{2}}{6} \\ -\frac{\sqrt{2}}{6} \\ \frac{2\sqrt{2}}{3} \end{pmatrix}$ is a normalized eigenvector.

$\underline{\sigma_3^2 = 0}$:

$$A^*A = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix} \sim \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -2 \\ 1 & 0 & 2 \end{bmatrix},$$

and solving for $A^*A = 0$, we find that $v_3 = \begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$ is a normalized eigenvector.

We can now "build" all the matrices that enter into the SVD of the matrix A. We get

$$V = \begin{bmatrix} v_1 | v_2 | v_3 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & \frac{2}{3} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{6} & -\frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & -\frac{1}{3} \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}.$$

Finally, we find that

$$U = \begin{bmatrix} U_1 | U_2 \end{bmatrix} = \begin{bmatrix} \frac{Av_1}{\|Av_1\|} | \frac{Av_2}{\|Av_2\|} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

With these choices of U, Σ and V, we have that $A = U\Sigma V^*$, or explicitly written out:

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{6} & -\frac{\sqrt{2}}{6} & \frac{2\sqrt{2}}{3} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{bmatrix}.$$

Let us now discuss some consequences and applications of the SVD Theorem.

Proposition 23. Let $A \in \mathcal{M}_{m\times n}(\mathbb{C})$ have positive singular values $\sigma_1 \geq \cdots \geq \sigma_r$. Then the operator norm of A (that is, the norm of the bounded linear operator associated with A) is

$$\|A\| = \sigma_1.$$

Proof. Let $A = U\Sigma V^*$ be the singular value decomposition of A, and let v_1 be the first column vector of V. The vector v_1 has length 1, and from the equation $AV = U\Sigma$ it is clear that $\|Av_1\| = \sigma_1$. It follows that

$$\|A\| = \sup_{\|x\|=1} \|Ax\| \geq \sigma_1.$$

Now let $x \in \mathbb{C}^n$ be any vector of length 1, and consider the equation $Ax = U\Sigma V^*x$. Since V^* is unitary, it represents an isometry, and it follows that $\|V^*x\| = 1$. Let us denote this vector by $y := V^*x$. Moreover, we note that Σy is the vector where the jth component of y is multiplied by σ_j. Thus, we have $\|\Sigma y\| \leq \sigma_1\|y\|$. Finally, since U is also unitary, we have

$$\|Ax\| = \|U\Sigma y\| = \|\Sigma y\| \leq \sigma_1\|y\| = \sigma_1,$$

and it follows that $\|A\| \leq \sigma_1$. We thus conclude that $\|A\| = \sigma_1$. □

Let us now see that the SVD of a matrix can be used to obtain so-called *polar decompositions*. A polar decomposition factors a square matrix in a manner analogous to the factoring of a complex number as the product of a complex number of length 1 and a nonnegative number ($z = |z|e^{2\pi i\varphi}$). In the case of matrices, the complex number of length 1 is replaced by a unitary matrix, and the nonnegative number is replaced by a positive semi-definite matrix.

Theorem 24 (Polar decomposition)**.** For any square matrix A, there exists a unitary matrix W and a positive semi-definite matrix P such that

$$A = WP.$$

Proof. By the singular value decomposition theorem, there exist unitary matrices U and V and a diagonal matrix Σ with nonnegative diagonal entries such that $A = U\Sigma V^*$. It follows that

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

where $W = UV^*$ and $P = V\Sigma V^*$. Since W is the product of unitary matrices, W is unitary. Moreover, since Σ is positive semi-definite, so is the matrix P. □

Example 25. To find the polar decomposition of

$$A = \begin{bmatrix} 11 & -5 \\ -2 & 10 \end{bmatrix},$$

we begin by finding the SVD of $A = U\Sigma V^*$. It can be shown that

$$v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

are orthonormal eigenvectors of A^*A with corresponding eigenvalues $\sigma_1^2 = 200$ and $\sigma_2^2 = 50$. Thus, we have

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix}.$$

Next, we find the columns of U:

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{5}\begin{pmatrix} 4 \\ -3 \end{pmatrix} \quad \text{and} \quad u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{5}\begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

Thus,

$$U = \begin{bmatrix} \frac{4}{5} & \frac{3}{5} \\ \frac{-3}{5} & \frac{4}{5} \end{bmatrix}.$$

Therefore, in the notation of the polar decomposition theorem, we have

$$W = UV^* = \begin{bmatrix} \frac{4}{5} & \frac{3}{5} \\ \frac{-3}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{5\sqrt{2}} \begin{bmatrix} 7 & -1 \\ 1 & 7 \end{bmatrix},$$

and

$$P = V\Sigma V^* = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{5}{\sqrt{2}} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}.$$

Finally, let us illustrate one possible application of SVD's to image processing.

Example 26. Suppose a satellite takes a picture, and wants to send it to earth. The picture may contain 1000×1000 pixels - a million little squares each with a definite color. We can code the colors, and send back 1000000 numbers. However, it is more convenient if we can find the *essential* information, and send only this.

Suppose we know the SVD, and specifically the matrix of singular values Σ. Typically, some of the σ's are significant, whereas others are extremely small. If we keep, say, 20 singular values, and discard the remaining 980, then we need only send the corresponding 20 columns of U and V. Thus, if only 20 singular values are kept, we send 20×2000 numbers rather than a million (and this is a 25 to 1 compression).

There is, of course, the additional cost of computing the SVD. This has become quite efficient, but is still expensive for big matrices.

6. The pseudoinverse

Let V and W be finite-dimensional inner product spaces over the same field $\mathbb{F}$, and let $T : V \to W$ be a linear transformation. It is desirable to have a linear transformation from W to V which captures some of the essence of an inverse of T even if T is not invertible. A simple (but fruitful) approach to this problem is to focus on the "part" of T that is invertible, namely the restriction of T to $\ker(T)^\perp$. Let $L : \ker(T)^\perp \to \operatorname{ran}(T)$ be the linear transformation defined by $L(x) = T(x)$ for all $x \in \ker(T)^\perp$. Then L is invertible, and we can use the inverse of L to construct a linear transformation from W to V which restores some of the benefits of an inverse of T.

Definition 10. Let V and W be finite-dimensional inner product spaces over the same field, and let $T : V \to W$ be a linear transformation. Let L :

$\ker(T)^\perp \to \operatorname{ran}(T)$ be the linear transformation defined by $L(x) = T(x)$ for all $x \in \ker(T)^\perp$. The *pseudoinverse* of T, denoted T^+, is defined as the unique linear transformation from W to V such that

$$T^+(y) = \begin{cases} L^{-1}(y) & \text{for } y \in \operatorname{ran}(T) \\ 0 & \text{for } y \in \operatorname{ran}(T)^\perp \end{cases}.$$

The pseudoinverse of a linear transformation T on a finite-dimensional inner product space exists even if T is not invertible. Furthermore, if T *is* invertible, then $T^+ = T^{-1}$, because $\ker(T)^\perp = V$ and L coincides with T.

Now let $A \in \mathcal{M}_{m\times n}(\mathbb{C})$ be the matrix representation of the linear map T. Then there exists a unique $(n \times m)$ matrix B which represents the pseudoinverse T^+. We call B the *pseudoinverse* of A and denote it by $B = A^+$. It turns out that the pseudoinverse A^+ can be computed with the aid of the singular value decomposition of A.

Theorem 27. Let $A \in \mathcal{M}_{m\times n}(\mathbb{C})$ have rank r and singular value decomposition $A = U\Sigma V^*$, where $\sigma_1 \geq \cdots \geq \sigma_r$ are the positive singular values of A. Let Σ^+ be the $(n \times m)$ matrix

$$\Sigma^+_{ij} = \begin{cases} \frac{1}{\sigma_i} & \text{if } i = j \leq r \\ 0 & \text{otherwise} \end{cases}.$$

Then $A^+ = V\Sigma^+U^*$.

We state this result without proof, and focus on its applications.

Let $b \in \mathbb{C}^m$, and consider the system of linear equations

$$Ax = b.$$

We know that this system has either no solution, a unique solution, or infinitely many solutions. It has a unique solution for every $b \in \mathbb{C}^m$ if and only if A is invertible, in which case the solution is given by $A^{-1}b$. Moreover, if A is invertible, then $A^{-1} = A^+$, so we could have written the solution as $x = A^+b$. If, on the other hand, the system $Ax = b$ is underdetermined or inconsistent, then A^+b still exists. This raises the question: How is the vector A^+b related to the system of linear equations $Ax = b$? In order to answer this question, we need the following lemma.

Lemma 28. Let V and W be finite-dimensional inner product spaces, and let $T : V \to W$ be linear. Then

i) T^+T is the orthogonal projection of V on $\ker(T)^\perp$.

ii) TT^+ is the orthogonal projection of W on $\operatorname{ran}(T)$.

Proof. As above, we define $L : \ker(T)^\perp \to \operatorname{ran}(T)$ by $L(x) = T(x)$ for $x \in \ker(T)^\perp$. If $x \in \ker(T)^\perp$, then

$$T^+T(x) = L^{-1}L(x) = x,$$

and if $x \in \ker(T)$, then

$$T^+T(x) = T^+(0) = 0.$$

Consequently, T^+T is the orthogonal projection of V on $\ker(T)^\perp$. This proves part i). Part ii) is proved similarly. $\square$

Theorem 29. Consider the system of linear equations $Ax = b$, where $A \in \mathcal{M}_{m\times n}(\mathbb{C})$ and $b \in \mathbb{C}^m$. If $z = A^+b$, then z has the following properties.

i) If $Ax = b$ is consistent, then z is the unique solution to the system having minimum norm. That is, z is a solution to the system, and if y is any other solution to the system, then $\|y\| > \|z\|$.

ii) If $Ax = b$ is inconsistent, then z is the unique best approximation to a solution having minimum norm. That is

$$\|Az - b\| \leq \|Ay - b\| \quad \text{for any } y \in \mathbb{C}^n,$$

with equality if and only if $Ay = Az$. Moreover, if $Ay = Az$, then $\|z\| \leq \|y\|$ with equality if and only if $z = y$.

Proof. Let T be the linear map associated to the matrix A

i) Suppose that $Ax = b$ is consistent, and let $z = A^+b$. Observe that $b \in \operatorname{ran}(T)$, and therefore

$$Az = AA^+b = TT^+b = b,$$

by Lemma 28ii). Thus, z is a solution to the system $Ax = b$. Now let y be any solution to the system. Then

$$T^+Ty = A^+Ay = A^+b = z.$$

Thus, z is the orthogonal projection of y on $\ker(T)^\perp$. By the projection theorem, we have $y = z + v$ with $v \in \ker(T)$, and $\|y\|^2 = \|z\|^2 + \|v\|^2$. It follows that $\|y\| > \|z\|$ unless $v = 0$ and $y = z$.

ii) Suppose that $Ax = b$ is inconsistent. By Lemma 28ii), we have that

$$Az = AA^+b = TT^+b$$

is the orthogonal projection of b on $\operatorname{ran}(T)$. Therefore, by the projection theorem, Az is the vector in $\operatorname{ran}(T)$ nearest b. If Ay is any other vector in $\operatorname{ran}(T)$, then necessarily

$$\|Az - b\| \leq \|Ay - b\|,$$

with equality if and only if $Az = Ay$. Finally, suppose that y is any vector in $\mathbb{C}^n$ such that $Az = Ay = c$. Then

$$A^+c = A^+Az = A^+AA^+b = A^+b = z,$$

where we have used that $A^+AA^+ = A^+$ (this is easily checked by writing out the SVD of A). Hence, we may apply part i) of this theorem to the system $Ax = c$ to conclude that $\|y\| \geq \|z\|$ with equality if and only if $y = z$.

$\square$

Example 30. Let us find the minimal norm solution of

$$-x_1 + 2x_2 + 2x_3 = b, \quad \text{for } b \in \mathbb{R}.$$

According to Theorem 29i), this is given by

$$z = A^+ b,$$

where A^+ is the pseudoinverse of the (1×3) matrix $A = \begin{bmatrix} -1 & 2 & 2 \end{bmatrix}$. The SVD of A is $A = U\Sigma V^*$, where

$$U = \begin{bmatrix} 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} -\frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{\sqrt{5}}{3} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & -\frac{4}{3\sqrt{5}} \end{bmatrix}.$$

The pseudoinverse of A is thus given by

$$A^+ = V\Sigma^+ U^* = \begin{bmatrix} -\frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{\sqrt{5}}{3} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & -\frac{4}{3\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9}, \end{bmatrix}$$

and it follows that the minimal norm solution of $Ax = b$ is

$$z = A^+ b = \begin{bmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9}, \end{bmatrix} b.$$

Any other solution of the system $Ax = b$ is necessarily of the form

$$y = A^+ b + v, \quad v \in \ker(A).$$