LINEAR ALGEBRA - TMA4115

FRANZ LUEF

1. Primer on Linear Algebra

We review some facts about spanning sets and basis in finite-dimensional vector spaces.

Definition 1.1. A *vector space* over the real numbers $\mathbb{R}$ is a set X together with the operations of addition $X \times X \to X$ and scalar multiplication $\mathbb{R} \times X \to X$ satisfying the following properties:

(1) Commutativity: $x + y = y + x$ for all $x, y \in X$ and $(\alpha\beta x) = \alpha(\beta x)$ for all $\alpha, \beta \in \mathbb{R}$;
(2) Associativity: $(x + y) + z = x + (y + z)$ for all $x, y, z \in X$;
(3) Additive identity: There exists an element $0 \in X$ such that $0 + x = x$ for all $x \in X$;
(4) Additive inverse: For every $x \in X$, there exists an element $y \in X$ such that $x + y = 0$, we denote y by $-x$;
(5) Multiplicative identity: $1x = x$ for all $x \in X$;
(6) Distributivity: $\alpha(x+y) = \alpha x+\alpha y$ and $(\alpha+\beta)x = \alpha x+\beta x$ for all $x, y \in X$ and $\alpha, \beta \in \mathbb{R}$.

The elements of a vector space are called vectors. Given $x_1, ..., x_n$ in X and scalars $\alpha_1, ..., \alpha_n \in \mathbb{R}$ we call the vector

$$x = \alpha_1 x_1 + \cdots + \alpha_n x_n$$

a *linear combination.* The set of all possible linear combinations of the vectors $x_1, ..., x_n$ in X is called the span of $\{\mathrm{x}_1, ..., x_n\}, denoted by \mathrm{span}\{\mathrm{x}_1, ..., x_n\}$. Recall that a set of vectors $\{x_1, ..., x_n\} \subset X$ is *linearly independent* if for all scalars $\alpha_1, ..., \alpha_n$ the equation $\alpha_1 x_1 + + \alpha_n x_n = 0$ has only $\alpha_1 = \cdots \alpha_n = 0$ as solution.
If there exists a non-trivial linear combination of the x_i's that give a representation of 0, then we call the $\{x_1, ..., x_n\}$ linearly dependent.

Lemma 1.1. *$\{x_1, ..., x_n\} \subset X$ is linearly dependent if and only if there exists a vector, e.g. x_j, that is a linear combination of the others, i.e.*

$$span\{x_1, ..., x_j, ..., x_n\} = span\{x_1, ..., x_{j-1}, x_{j+1}, ..., x_n\}$$

Lemma 1.2. *$\{x_1, ..., x_n\} \subset X$ is linearly independent if and only if every $x \in span\{x_1, ..., x_n\}$ can be written uniquely as a linear combination of elements of $\{x_1, ..., x_n\}$.*

Proof. ($\Rightarrow$) Assume $\{x_1, ..., x_n\}$ is linearly independent. Suppose there are two ways to express x:

$$x = \alpha_1 x_1 + \cdots + \alpha_n x_n$$
$$x = \alpha'_1 x_1 + \cdots + \alpha'_n x_n.$$

Then we have

$$0 = (\alpha_1 - \alpha'_1)x_1 + \cdots + (\alpha_n - \alpha'_n)x_n.$$

By linear independence all these scalars have to be zero, hence the representation is unique. Contradicting our assumption.

Date: Fall 2022.

($\Leftarrow$) Suppose every $x \in \text{span}\{x_1, ..., x_n\}$ can be written uniquely as a linear combination of elements of $\{x_1, ..., x_n\}$. Hence there exist unique scalars $\alpha_1, ..., \alpha_n$ for every $x \in \text{span}\{x_1, ..., x_n\}$ such that

$$x = \alpha_1 x_1 + \cdots + \alpha_n x_n.$$

In particular $x = 0$ is uniquely represented, hence the trivial decomposition $\alpha_1 = \cdots = \alpha_n = 0$ is the only way to represent the zero vector. Hence the set $\{x_1, ..., x_n\}$ is linearly independent. □

There are two central notions in the theory of vector spaces:

Definition 1.2. Let X be a vector space.

(1) If there exists a set $S \subseteq X$ with $\text{span}(S) = X$, then we call S a *spanning set*. In case that S consists of finitely many elements $\{x_1, ..., x_n\}$, then we say that X is *finite-dimensional*. Finally, if there exists no finite spanning set for X, then we call the vector space *infinite-dimensional*.
(2) If there exists a linearly independent spanning set B for X, then we call B a basis for X.

Proposition 1.3 (Basis Reduction Theorem)**.** *If $\{x_1, ..., x_n\}$ is a spanning set for X, then either $\{x_1, ..., x_n\}$ is a basis for X or some x_j's can be removed from $\{\{x_1, ..., x_n\}\}$ to obtain a basis.*

As a consequence we get that every finite-dimensional vector space has a basis.

Proposition 1.4. *Every finite-dimensional vector space has a basis.*

An often used result is the following one:

Proposition 1.5 (Basis Extension Theorem)**.** *Let X be a finite-dimensional vector space. Then any linearly independent subset of X can be extended to a basis.*
Any two bases of a finite-dimensional vector space have the same number of elements.

Lemma 1.6. *Let X be a finite-dimensional vector space of dimension n. Then any set $\{x_1, ..., x_n\}$ of n linearly independent vectors is a basis of X. In other words, any set of vectors $\{x_1, ..., x_m\}$ with $m > n$ is linearly dependent.*

These observations motivate

Definition 1.3. Suppose X has a basis $\{x_1, ..., x_n\}$. Then we call the number of elements of this basis the *dimension* of X, denoted by $\dim(X)$. If X is infinite-dimensional, then we write $\dim(X) = \infty$.

Example 1.1. We have that $\dim(\mathbb{R}^n) = n$, the dimension of the space of all polynomials of degree at most n is $\dim(\mathcal{P}_n) = n + 1$ and the vector space of all polynomials is infinite dimensional, $\dim(\mathcal{P}) = \infty$.

Transformations that preserve linear combinations are called linear transformations, but also might referred to as linear mappings and linear operators.

Definition 1.4. Suppose X and Y are vector spaces. A mapping $T : X \to Y$ that satisfies $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ is called a linear mapping.
A bijective linear mapping $T : X \to Y$ is called an isomorphism between X and Y. We then refer to X and Y as isomorphic vector spaces.

Maybe one of the most important example of isomorphic vector spaces is that $\mathbb{R}^n$ is isomorphic to any n-dimensional real vector space.

Theorem 1.7. *Any n-dimensional vector space X is isomorphic to $\mathbb{R}^n$.*

We introduce some notation before we prove this statement.

Definition 1.5. Suppose X is a n-dimensional vector space. If $\mathcal{B} = \{b_1, ..., b_n\}$ is a basis of X, then for any $x \in X$ we denote the scalars in the unique expansion of x with respect to the basis $\mathcal{B}$ by $\alpha_1, , , ., \alpha_n$:

$$x = \alpha_1 b_1 + \cdots + \alpha_n b_n.$$

The linear mapping from X to $\mathbb{R}^n$ defined by $x \mapsto [x]_{\mathcal{B}} := (\alpha_1, ..., \alpha_n)^T$ is called the *coefficient mapping* and denoted by $C_{\mathcal{B}}$.

Proof. Let $\mathcal{B}$ be a basis for X. Then the mapping $C_{\mathcal{B}} : X \to \mathbb{R}^n$ is linear and bijective.

- Claim: $C_{\mathcal{B}}$ is linear. Let $[x]_{\mathcal{B}} = (\alpha_1, ..., \alpha_n)^T$ and $[y]_{\mathcal{B}} = (\beta_1, ..., \beta_n)^T$ be the coefficients of $x, y \in X$. Then we have $[x + \lambda y]_{\mathcal{B}} = [x]_{\mathcal{B}} + \lambda[y]_{\mathcal{B}}$ since

$$x + \lambda y = \sum_{j=1}^{n} \alpha_j b_j + \lambda \sum_{j=1}^{n} \beta_j b_j = \sum_{j=1}^{n} (\alpha_j + \lambda \beta_j) b_j.$$

- Claim: $C_{\mathcal{B}}$ is bijective.
  There are several ways to see this: (i) One is the show that $C_{\mathcal{B}}$ is injective and surjective. (ii) Or instead, find the inverse to $C_{\mathcal{B}}$ and show that it is linear. (iii) Using the fact that a linear mapping is bijective if and only if it sends a basis to a basis.

  (i) $C_{\mathcal{B}}$ is injective: Suppose $C_{\mathcal{B}}(x) \neq C_{\mathcal{B}}(y)$ for $x, y \in X$. Then by the uniqueness of the coefficients in the expansion wrt to the basis $\mathcal{B}$, we have $x \neq y$.
  $C_{\mathcal{B}}$ is surjective: Suppose $(\alpha_1, ..., \alpha_n)^T$ in $\mathbb{R}^n$. Then the vector $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$ satisfies $[x]_{\mathcal{B}}$ equal to the vector we started with.
  (ii) The inverse of $C_{\mathcal{B}}^{-1}$ is a mapping from $\mathbb{R}^n$ to X and given by $((\alpha_1, ..., \alpha_n)^T) \mapsto x = \alpha_1 x_1 + \cdots + \alpha_n x_n$. Show that it is linear and that $C_{\mathcal{B}} \circ C_{\mathcal{B}}^{-1} = \mathrm{id} = C_{\mathcal{B}}^{-1} \circ C_{\mathcal{B}}$.
  (iii) Show that $C_{\mathcal{B}}(b_j) = e_j$, where e_j denotes the $j - th$ standard vector in $\mathbb{R}^n$.

□

Next we discuss the link between matrices and linear transformations. On the one hand an $m \times n$ matrix A with real entries defines a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ by $Tx = Ax$.

On the other hand **any** linear transformation T between finite-dimensional vector spaces X and Y can by represented as a matrix-vector transformation after picking a basis for X and Y, respectively.

Let $\mathcal{B} = \{b_1, ..., b_n\}$ be a basis of X and $\mathcal{C} = \{c_1, ..., c_m\}$ be a basis of Y. Suppose T is a linear transformation $T : X \to Y$ Then

$$x = \sum_{j=1}^{n} \alpha_j b_j \qquad \mapsto \qquad T(x) = \sum_{j=1}^{n} \alpha_j T(b_j).$$

Thus we have

$$[T(x)]_{\mathcal{C}} = \sum_{j=1}^{n} \alpha_i [T(b_j)]_{\mathcal{C}}.$$

We define a $m \times n$ matrix A which has as its j-th column $[T(b_j)]_{\mathcal{C}}$. Then we have

$$[Tx]_{\mathcal{C}} = A[x]_{\mathcal{B}}.$$

The matrix A represents T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$. Sometimes, we denote this A sometimes by $[T]_{\mathcal{B}}^{\mathcal{C}}$.

We address now the relation between the matrix representation of T depending on the change of bases of X and Y, respectively.
Suppose we have two bases $\mathcal{B} = \{x_1, ..., x_n\}$ and $\mathcal{R} = \{y_1, ..., y_n\}$ for X. Let $x = \sum_{j=1}^{n} \alpha_i x_i$. Then

$$[x]_{\mathcal{R}} = \sum_{j=1}^{n} \alpha_i \vec{x_i}_{\mathcal{R}}.$$

Define the $n \times n$ matrix P with j-th column $\vec{x_j}_{\mathcal{R}}$, and we call P the *change of bases matrix*:

$$[x]_{\mathcal{R}} = P[x]_{\mathcal{B}}$$

and by the invertibility of P we also have

$$[x]_{\mathcal{B}} = P^{-1}[x]_{\mathcal{R}}.$$

Let now $\mathcal{C}$ and $\mathcal{S}$ be two bases for Y. Then a linear transformation $T : X \to Y$ has two matrix representations:

$$A = [T]_{\mathcal{B}}^{\mathcal{C}} \text{ and } B = [T]_{\mathcal{R}}^{\mathcal{S}}.$$

In other words we have

$$[Tx]_{\mathcal{C}} = A[x]_{\mathcal{B}} \quad , [Tx]_{\mathcal{S}} = B[x]_{\mathcal{R}}$$

for any $x \in X$. Let P be the change of bases matrix of size $n \times n$ such that $[x]_{\mathcal{R}} = P[x]_{\mathcal{B}}$ for any $x \in X$ and let Q be the invertible $m \times m$ matrix such that $[y]_{\mathcal{S}} = Q[y]_{\mathcal{C}}$.
Hence we get that

$$[Tx]_{\mathcal{S}} = BP[x]_{\mathcal{B}}$$

and

$$[y]_{\mathcal{S}} = [Tx]_{\mathcal{S}} = Q[Tx]_{\mathcal{C}} = QA[x]_{\mathcal{B}}$$

for any $x \in X$. Hence we get that

$$B = QAP^{-1} \text{ and } A = Q^{-1}BP.$$

In the case $X = Y$ we have $P = Q$ and we set $S = Q^{-1}$ to get $B = S^{-1}AS$. Then the matrices A and B represent the same linear transformation T on X with respect to different bases.
These observations motivate the definition of matrices representing the same linear transformation.

Definition 1.6. Two $m \times n$ matrices A and B are called *equivalent* if there exists an invertible matrix S such that $B = QAP^{-1}$. Furthermore, Two $n \times n$ matrices A and B are called *similar* if there exists an invertible matrix S such that $B = S^{-1}AS$.

Note that two similar matrices describe the same linear transformation on X with respect to different bases of X.

2. Invariant subspaces and matrix decomposition

Invariance of a class of objects under some structures is an integral part of mathematics. In the case of linear transformations between vector spaces the invariance of a subspace under a linear transformation is one of the crucial notions. Since it allows one to address the main problem of linear algebra: Show that given a linear transformation on a vector space X. There exists a basis of X with respect to which T has a reasonable simple matrix representation.

In order to achieve this goal we have to break up our linear transformation on X into "smaller" ones, by decomposing X into subspaces that allow us to restrict the linear transformation onto these subspaces.

Suppose $T : X \to Y$ is a linear transformation. Then the *kernel* of T, $\ker(T)$, is a subspace of X consisting of all $x \in X$ for which $Tx = 0$, and the *image* or *range* of T, denoted by $\text{im}(T)$ or $\text{ran}(T)$, is the subspace of all $y \in Y$ that are of the form $y = Tx$ for some $x \in X$.
If one represents $T : X \to Y$ as a matrix-vector transformation $y = Ax$, then instead of the kernel we refer to it as the *nullity* of A and the range of T becomes the *column space* of A. Recall that the dimension of the column space is called the *rank* of A.

In order to make sense of the restriction of a linear mapping $T : X \to X$ to a subspace M, it needs to satisfy that $T(M) \subseteq M$.

Definition 2.1. Suppose T is a linear transformation on a vector space. A subspace M of X is called *invariant* under T if $x \in M$ implies $Tx \in M$. We will also refer to M as T-invariant subspace.

Here are some examples of invariant subspaces. Let T be a linear transformation on a vector space X.

(1) $\{0\}$ and X;
(2) The kernel and the range of T.
(3) Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear mapping defined by $T(x_1, x_2, x_3) = (x_1, x_2, 0)$. Then the subspace M spanned by $(1, 0, 0)$ and $(0, 1, 0)$ is T-invariant. Note that T is the orthogonal projection of $\mathbb{R}^3$ onto M.

A question of interest is if a linear operator on a vector space has an invariant subspace. We will later demonstrate that any linear transformation on a complex vector space has an invariant subspace. This is not the case for linear mappings on real vector spaces, e.g. take the rotation R_α by the angle α in $\mathbb{R}^2$.

Lemma 2.1. *Suppose $T : X \to X$ is a linear mapping and M a subspace of X. Then M is T-invariant if and only if $T(b_j) \in M$, $j = 1, ..., k$ for any basis $\{b_1, ..., b_k\}$ of M.*

Proof. This follows from the observation that a linear mapping is uniquely determined by its values on a basis. □

Suppose M is a subspace of X. Then we know from other courses that the orthogonal complement $M^\perp$ of M allows us to decompose $x \in X$ in the part x_M in M and its part $x_{M^\perp}$ in $M^\perp$ $x = x_M + x_{M^\perp}$ where x_M and $x_{M^\perp}$ are unique. The reason underlying the uniqueness of the decomposition is that $M \cap M^\perp = \{0\}$.
Recall that the sum $M + N$ of two subspaces of X is defined to be the set $M + N = \{m + n : m \in M, n \in N\}$, which is also a subspace of X. There is a relation between the dimensions of subspaces of a finite-dimensional vector space X and the dimensions of their intersection and sum:

$$\dim(M + N) + \dim(M \cap N) = \dim(M) + \dim(N).$$

Let us focus on the case when the subspaces M and N have trivial intersection, i.e. $M \cap N = \{0\}$. Sums of subspaces that satisfy this additional condition are called **direct sums**.

Lemma 2.2. *Let M and N be subspaces of a finite-dimensional vector space X. Then $M \cap N = \{0\}$ if and only if for every $z \in M + N$ there exist unique elements $m \in M$ and $n \in N$ such that $z = m + n$.*

Proof. ($\Rightarrow$) Suppose we have $M \cap N = \{0\}$. Let $z \in M + N$ have two decompositions $z = m_1 + n_1 = m_2 + n_2$ for $m_i \in M$ and $n_i \in N$, $i = 1, 2$. Then we have $0 = m_1 - m_2 + (n_1 - n_2)$. We set $m := m_1 - m_2$ and $n := n_1 - n_2$ and note that $m \in M$ and $n \in N$. Hence we have $m = -n$,

which implies that $m \in N$ and $n \in M$. Consequently m and n are in $M \cap N$. By assumption $M \cap N = \{0\}$ implies that $m = n = 0$, which yields that $m_1 = m_2$ and $n_1 = n_2$. This shows the desired uniqueness of the decomposition.
($\Leftarrow$) Suppose that every element $z \in M + N$ can be uniquely written as $z = m + n$ for $m \in M$ and $n \in N$. Assume that $b \in M \cap N$, i.e. $b \in M$ and $b \in N$.Since N is a subspace, we have also $-b \in N$. Hence we have $0 = b + (-b)$ where $b \in M$ and $-b \in N$. On the other hand 0 has also the decomposition $0 = 0 + 0$. The uniqueness condition yields that $b = 0$. Since b was arbitrary, we have $M \cap N = \{0\}$. $\square$

A result of utmost importance is the existence of complements for a subspace of a finite-dimensional vector space.

Proposition 2.3. *Let X be a finite-dimensional vector space and let M be any subspace of X. Then there exists a subspace N of X such that $M \otimes N = X$.*

We call the subspace $N \subseteq X$ a *complement* of M.

Proof. Let $\{x_1, ..., x_k\}$ be a basis of M. Then there exist vectors $y_1, .., y_l$ in X such that $\{x_1, ..., x_k, y_1, ..., y_l\}$ is a basis of X. We define N to be the span of $\{y_1, ..., y_l\}$ and note that this set is also a basis of N. By construction we have $M + N = X$.
Let us show that $M \cap N = \{0\}$. Suppose $z \in M \cap N$. Then $z = \alpha_1 x_1 + \cdots + \alpha_k x_k$ since it is an element of M and $z = \beta_1 y_1 + \cdots + \beta_l y_l$. Consequently, $0 = \alpha_1 x_1 + \cdots + \alpha_k x_k - \beta_1 y_1 - \cdots - \beta_l y_l$ which yields that $\alpha_1 = \cdots = \alpha_k = \beta_1 = \cdots = \beta_l = 0$. Hence $z = 0$ and since z was arbitrary, we have $M \cap N = \{0\}$. $\square$

We explore the implications of invariant subspaces and direct sums for matrix representations of linear mappings.

Proposition 2.4. *Let $T : X \to X$ be a linear mapping and M a T-invariant subspace of X. Suppose $\mathcal{B}_M = \{b_1, ..., b_k\}$ is a basis of M and $\mathcal{B} = \{b_1, ..., b_k, b_{k+1}, ..., b_n\}$ be a basis of X. Then the matrix representation of T wrt $\mathcal{B}$ is of the form*

$$[T]_{\mathcal{B}} = \begin{bmatrix} [T]_{\mathcal{B}_M} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

where $[T]_{\mathcal{B}_M}$ is the matrix representation of T wrt to $\mathcal{B}_M$, A_{12} is an $k \times (n-k)$ matrix and A_{22} is an $(n-k) \times (n-k)$ matrix.

Proof. Let $\mathcal{B}_M$ be a basis of M. Then the condition $T(b_j) \in M$ for $j = 1, ..., k$ implies that $[T]_{\mathcal{B}_M}$ is an $k \times k$ matrix since the columns of $[T]_{\mathcal{B}_M}$ are linear combinations of the elements of $\mathcal{B}_M$. Hence this yields the zeros in the first k columns of $[T]_{\mathcal{B}}$. $\square$

Proposition 2.5. *Suppose $T : X \to X$ is a linear mapping and let M, N be T-invariant subspaces sucht that $X = M \oplus N$. If $\mathcal{B} = \mathcal{B}_M \cup \mathcal{B}_N$ is a basis of X where $\mathcal{B}_M$ and $\mathcal{B}_N$ are bases of M and N, then the matrix representation of T wrt $\mathcal{B}$ is of the form*

$$[T]_{\mathcal{B}} = \begin{bmatrix} [T]_{\mathcal{B}_M} & 0 \\ 0 & [T]_{\mathcal{B}_N} \end{bmatrix}.$$

Proof. Let $\mathcal{B}_M = \{b_1, ..., b_m\}$ be a basis of M and let $\mathcal{B}_N = \{\tilde{b_1}, ..., \tilde{b_n}\}$ be a basis of N. Since $T(b_i)$ is in M for $i = 1, ..., m$ and $T(\tilde{b_j})$ is in N for $j = 1, ..., n$, we have

$$\begin{aligned}
T(b_1) &= a_{11}b_1 + \cdots a_{1m}b_m + 0 \cdots \tilde{b}_1 + \cdots + 0 \cdots \tilde{b}_n \\
T(b_2) &= a_{21}b_1 + \cdots a_{2m}b_m + 0 \cdots \tilde{b}_1 + \cdots + 0 \cdots \tilde{b}_n \\
&\vdots \\
T(b_m) &= a_{m1}b_1 + \cdots a_{mm}b_m + 0 \cdots \tilde{b}_1 + \cdots + 0 \cdots \tilde{b}_n \\
T(\tilde{b}_1) &= 0 \cdots b_1 + \cdots + 0 \cdots b_m + b_{11}\tilde{b}_1 + \cdots + b_{1n}\tilde{b}_n \\
T(\tilde{b}_2) &= 0 \cdots b_1 + \cdots + 0 \cdots b_m + b_{21}\tilde{b}_1 + \cdots + b_{2n}\tilde{b}_n \\
&\vdots \\
T(\tilde{b}_n) &= 0 \cdots b_1 + \cdots + 0 \cdots b_m + b_{n1}\tilde{b}_1 + \cdots + b_{nn}\tilde{b}_n
\end{aligned}$$

i.e. $[T]_\mathcal{B} = \begin{bmatrix} [T]_{\mathcal{B}_M} & 0 \\ 0 & [T]_{\mathcal{B}_N} \end{bmatrix}$ where $[T]_{\mathcal{B}_M} = (a_{ij})_{i,j=1}^m$ and $[T]_{\mathcal{B}_N} = (b_{ij})_{i,j=1}^n$. □

3. Eigenspaces and Generalized Eigenspaces

Let us investigate one-dimensional invariant subspaces.

Proposition 3.1. *A linear transformation on a finite-dimensional vector space has a one-dimensional invariant subspace if and only if T has an eigenvector.*

Proof. (•) Suppose M is invariant under T, then $Tx \in M$ and hence there is a scalar $\lambda \in \mathbb{F}$ such that $Tx = \lambda x$.

(•) If $Tx = \lambda x$ for some $\lambda' in \mathbb{F}$ and some non-zero $x \in X$, then the span(x) is a one-dimensional subspace. This subspace is invariant under T. □

We restrict our discussion to complex vector spaces, i.e. the scalars in our linear combinations are complex numbers.

Definition 3.1. A scalar λ is called an *eigenvalue* of a linear transformation $T : X \to X$ if there exists a non-zero $x \in X$ such that $Tx = \lambda x$. The set $\sigma(T)$ of $\mathbb{C}$

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

is known as the spectrum of T.

In other words, x is an eigenvector of T if and only if $x \in \ker T - \lambda I$. For finite-dimensional vector spaces $\sigma(T)$ is the set of all eigenvalues counting multiplicities of T.

Definition 3.2. The subspace $E_\lambda = \ker(T - \lambda I)$ is called the *eigenspace* of T for the eigenvalue λ. The dimension of E_λ is called the *geometric multiplicity* of λ.

Note that E_λ consists of the eigenvectors of T and the zero vector 0.

Theorem 3.2. *Suppose T is a linear transformation on a finite-dimensional complex vector space. Then there exists an eigenvalue $\lambda \in \mathbb{C}$ for an eigenvector x of T.*

Proof. We assume that $\dim(X) = n$ and choose any non-zero vector x in X. Consider the following set of $n+1$ vectors in X:

$$\{x, Tx, T^2x, ..., T^nx\}.$$

Since $n+1$ vectors in an n-dimensional vector space X are linearly independent, there exists a non-trivial linear combination:

$$a_0x + a_1Tx + \cdots + a_nT^nx = (a_0I + a_1T + \cdots + a_nT^n)x = 0.$$

Note that not all $a_1, ..., a_n$ are zero. If they were all zero, then $a_0x = 0$ which would imply that $a_0 = 0$. Hence that the linear combination is trivial.

Let us denote by $p(z) = a_0 + a_1z + \cdots + a_nz^n$ the polynomial associated to the linear transformation T. Powers of numbers correspond to powers of T by the corresponding iterates of T and $T^0 = I$.

Then the non-trivial linear combination among the vectors turns into a polynomial equation in T:

$$p(T) = 0.$$

By the Fundamental Theorem of Algebra any polynomial can be written as a product of linear factors:

$$p(t) = c(t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n), \ \lambda_i \in \mathbb{C}, c \neq 0.$$

Hence $p(T)$ has a factorization of the form:

$$p(T) = c(T - \lambda_1I)(T - \lambda_2I) \cdots (T - \lambda_mI).$$

Hence $p(T)$ is a product of linear mappings $T - \lambda_jI$ for $j = 1, ..., m$. We know that $p(T)x = 0$ for a non-zero $x \neq 0$, which implies that at least one of these linear mappings is not invertible. Thus it has to have a non-trivial kernel, let's say $y \in \ker(T - \lambda_iI)$, which yields that y is an eigenvector for the eigenvalue λ_i. Consequently, we have shown the desired assertion. □

The assumptions of the above statement are crucial: (i) Since there are linear transformations on a real vector space, do not need to have eigenvalues. For example, the rotation by 90 degrees in the plane $\mathbb{R}^2$.

Definition 3.3. A $n \times n$ matrix A is called diagonalizable if it has n linearly independent eigenvectors.

Note that the set of eigenvectors of a diagonalizable matrix is consequently a basis for $\mathbb{C}^n$. By definition a diagonalizable $n \times n$ matrix A has eigenvalues $\lambda_1, ..., \lambda_n$ and associated eigenvectors $u_1, ..., u_n$ satisfying:

$$\begin{aligned} Au_1 &= \lambda u_1 \\ &\vdots \\ Au_n &= \lambda u_n. \end{aligned}$$

Collect the eigenvectors of A into one matrix: $U = (u_1|u_2|\cdots|u_n)$; and the eigenvalues of A into the diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & 0 \cdots & \cdots & 0 & \\ \vdots & \lambda_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \lambda_n \end{pmatrix}.$$

Then the eigenvalue equations turn into a matrix equation:

$$AU = UD.$$

Since A is diagonalizable, the eigenvectors are a basis for $\mathbb{C}^n$. Hence U is invertible and we have

$$A = UDU^{-1}.$$

Sometimes U is an unitary matrix, i.e. the eigenvectors yield an orthonormal basis for $\mathbb{C}^n$. Then we have $A = UDU^*$.

A well-known criterion for the non-invertiblity of a matrix is the vanishing of its determinant. Hence eigenvalues are the zeros of the polynomial $p_A(z) = \det(zI - A)$, known as the *characteristic polynomial*.

Lemma 3.3. *Similar matrices have the same characteristic equation.*

Proof. Let A and B be similar matrices. Thus there exists an invertible matrix S such that $B = S^{-1}AS$.

$$p_B(z) = \det(zI - S^{-1}AS) = \det(zS^{-1}S - S^{-1}AS) = \det(S^{-1}(zI - A)S) = p_A(z).$$

□

As an important consequence of the existence of an eigenvector for linear mappings between complex finite-dimensional vector spaces we prove Schur's triangularization theorem, our first classification theorem. Before we introduce a refined version of similarity. Namely, if the matrix S in the definition of similar matrices may be chosen as a chosen as a unitary matrix, then we call the matrices A and B *unitarily equivalent*.

Theorem 3.4 (Triangularization Theorem)**.** *Given a $n \times n$ matrix with eigenvalues $\lambda_1, ..., \lambda_n$, counting multiplicities. There exists a unitary $n \times n$ matrix U such that*

$$A = UTU^*$$

for an upper triangular matrix T with the eigenvalues on the diagonal. Hence any matrix is similar to an upper triangular matrix.

We refer to the decomposition of the theorem as *Schur form.*

Proof. We proceed by induction on n. For $n = 1$, there is nothing to show. Suppose that the result is true up to matrices of size $n - 1$.
Let A be a $n \times n$ matrix with eigenvalues $\lambda_1, ..., \lambda_n$ counting multiplicities. Choose a normalized eigenvector u_1 for the eigenvalue λ_1. Then we extend u_1 to a basis $\{u_1, ..., u_n\}$ of $\mathbb{C}^n$ and we choose this basis to be orthonormal. Relative to this basis the matrix is of the form

$$A = U \begin{pmatrix} \lambda_1 & x & \cdots & x \\ 0 & & & \\ \vdots & A_{n-1} & & \\ 0 & & & \end{pmatrix} U^{-1},$$

where U is the matrix of the system $\{u_1, ..., u_n\}$ relative to the canonical basis. Since this is a unitary matrix, the similarity, is actually a unitary equivalence. By the induction hypothesis there exists a $(n-1) \times (n-1)$-matrix V such that VAV^* is upper triangular. Set $\tilde{V}$ to be the $n \times n$ matrix where $v_11 = 1$ and the other entries of the first column and row are zero. Then $\tilde{V}$ is a unitary matrix and $U\tilde{V}$ is the desired unitary matrix. □

Example 3.1. Find the Schur form of $A = \begin{pmatrix} 5 & 7 \\ -2 & -4 \end{pmatrix}$.
First step: Find an eigenvalue of A and associated eigenvector. The characteristic polynomial is $\lambda^2 - \lambda - 6 = 0$ and so $\lambda_1 = -2$ and $\lambda_2 = 3$. An eigenvector for $\lambda_1 = -2$ is $x_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.
The second step is to complete it to a basis of $\mathbb{C}^2$. In our case we take the eigenvector to the second eigenvalue and note that the corresponding set of vectors is linearly independent: $x_2 = \begin{pmatrix} 7 \\ -2 \end{pmatrix}$.
Third step: Use a orthonormalization procedure, e.g. Gram-Schmidt, to turn the system $\{x_1, x_2\}$ into a basis $\{u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}\}$.
Final step: Form the matrix $U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Computation of $U^*AU = \begin{pmatrix} 2 & 9 \\ 0 & 3 \end{pmatrix}$, which has the eigenvalues of A on its diagonal and is upper triangular.

Schur's triangularization theorem has a number of important consequences.

Theorem 3.5 (Cayley-Hamilton)**.** *Given a $n \times n$ matrix. Then*

$$p_A(A) = 0,$$

where $p_A(A)$ is the characteristic polynomial of A.

We state a refined version of Schur's triangularization theorem

Theorem 3.6 (Schur normal form)**.** *Given a $n \times n$ matrix A with distinct eigenvalues $\lambda_1, ..., \lambda_k$ with $k \leq n$. Then A is unitarily equivalent to*

$$\begin{pmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & \ddots & 0 \\ \vdots & \ddots & \vdots & \\ 0 & \dots & 0 & T_k \end{pmatrix}$$

where T_i has the form

$$\begin{pmatrix} \lambda_i & x & \cdots & x \\ 0 & \lambda_i & \ddots & x \\ \vdots & \ddots & \ddots & x \\ 0 & \dots & 0 & \lambda_i \end{pmatrix}$$