SUMMARY OF MATH 341

May 6, 2022

1. Set theory

1.1. **Definition ($A = B$).** Two sets A and B are equal if they have the same elements, i.e. for every $x \in A$ it is true that $x \in B$, and for every $x \in B$ it is true that $x \in A$.

1.2. **Definition ($A \subset B$).** A set A is a subset of a set B if every $x \in A$ also is an element of B.

1.3. **Definition ($A \cup B$).** The union of two sets A and B is

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

1.4. **Definition ($A \cap B$).** The intersection of two sets A and B is

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

1.5. **Definition ($A \times B$).** The product of two sets A and B is the set of all pairs (a, b) where $a \in A$ and $b \in B$:

$$A \times B = \big\{(a, b) \mid a \in A, b \in B\big\}.$$

2. Vector spaces

2.1. **Vector space axioms.** $(V, \mathbb{F}, +, \cdot)$ defines a vector space if V is a set, $\mathbb{F}$ is a number field, and if addition of elements of V and multiplication of numbers and vectors are defined so that they satisfy these axioms:

Commutativity: For all $x, y \in V$ one has $x + y = y + x$

Associativity: For all $x, y, z \in V$ one has $(x + y) + z = x + (y + z)$

Zero vector: There is a $0_V \in V$ such that for all $x \in V$ one has $x + 0_V = x$

Additive inverse: For each $x \in V$ there is a $y \in V$ such that $x + y = 0_V$

Multiplicative identity: For each $x \in V$ one has $1_{\mathbb{F}} x = x$

Associative and distributive properties: For all $a, b \in \mathbb{F}$, and all $x, y \in V$ one has $(ab)x = a(bx)$, $\quad a(x + y) = ax + ay$, $\quad (a + b)x = ax + bx$.

2.2. **Definition (linear combination).** A linear combination of vectors $v_1, v_2, \ldots, v_n \in V$ is any vector of the form $a_1 v_1 + \cdots + a_n v_n$, for any choice of numbers $a_1, \ldots, a_n \in \mathbb{F}$.

2.3. **Definition (basis).** n vectors $v_1, \ldots, v_n \in V$ are called a basis for V if for every $v \in V$ there exists a unique n-tuple of numbers $a_1, \ldots, a_n \in \mathbb{F}$ with $v = a_1 v_1 + \cdots + a_n v_n$.

2.4. **Definition (generate, span).** An n-tuple of vectors $v_1, \dots, v_n \in V$ generates V if every $v \in V$ is a linear combination of $v_1, \dots, v_n$.

2.5. **Definition (independence).** An n-tuple of vectors $v_1, \dots, v_n \in V$ is linearly independent if the only linear combination $c_1 v_1 + \cdots + c_n v_n$ with $c_1 v_1 + \cdots + c_n v_n = 0$ is the one where all coefficients are zero.

In symbols: $\forall c_1, \dots, c_n \in \mathbb{F} : c_1 v_1 + \cdots + c_n v_n = 0 \implies c_1 = \cdots = c_n = 0$

2.6. **Theorem.** Vectors $v_1, \dots, v_n \in V$ are linearly dependent if and only if one of the vectors v_i can be written as a linear combination of the others.

In other words, $v_1, \dots, v_n \in V$ are linearly dependent if and only if there is an $i \in \{1, 2, \dots, n\}$ and there are numbers $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n \in \mathbb{F}$ such that $v_i = a_1 v_1 + \cdots + a_{i-1} v_{i-1} + a_{i+1} v_{i+1} + \cdots + a_n v_n$.

2.7. **Basis Selection Theorem.** If $v_1, \dots, v_n$ spans V then there is a subset $v_{i_1}, \dots, v_{i_p}$ that is a basis for V.

2.8. **Components of a vector with respect to a basis.** Let $u_1, \dots, u_n \in V$ be given vectors. Then $u_1, \dots, u_n$ is a basis for V if and only if $u_1, \dots, u_n$ is linearly independent and spans V.

See Appendix A.1 for a proof.

The numbers $x_1, \dots, x_n$ are called the components, or coefficients, of the vector x with respect to the basis $u_1, \dots, u_n$.

3. Linear Transformations

3.1. **Definition — linear transformation.** A map $T : V \to W$ from one vector space V to another W is called linear if for all $x, y \in V$ and all $a, b \in \mathbb{F}$ one has $T(ax + by) = aT(x) + bT(y)$.

3.2. **Definition — the zero transformation.** If V and W are vector spaces then the map $O : V \to W$ defined by $O(x) = 0_W$ for all $x \in V$ is linear. O is called the zero transformation.

3.3. **Definition — the identity transformation.** If V is a vector space then the map $I : V \to V$ defined by $I(x) = x$ for all $x \in V$ is called the identity transformation.

3.4. **Theorem.** If $T, S : V \to W$ are linear and if $a \in \mathbb{F}$ then the maps $T + S : V \to W$ and $aT : V \to W$ are linear. The set $\mathscr{L}(V, W)$ of linear maps $T : V \to W$ is a vector space.

3.5. **Theorem — the matrix of $A:\mathbb{F}^n\to\mathbb{F}^m$.** Let $a_{11},\dots,a_{mn}\in\mathbb{F}$ be given numbers. Then the map $A:\mathbb{F}^n\to\mathbb{F}^m$ defined by

$$(\star)\qquad A\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}a_{11}x_1+\cdots+a_{1n}x_n\\a_{21}x_1+\cdots+a_{2n}x_n\\\vdots\\a_{m1}x_1+\cdots+a_{mn}x_n\end{pmatrix}$$

is linear.

Conversely, if $A:\mathbb{F}^n\to\mathbb{F}^m$ is a linear map then there exist numbers $a_{11},\dots,a_{mn}\in\mathbb{F}$ such that Ax is given by $(\star)$.

The matrix of A is

$$\begin{pmatrix}a_{11}&\cdots&a_{1n}\\a_{21}&\cdots&a_{2n}\\\vdots&&\vdots\\a_{m1}&\cdots&a_{mn}\end{pmatrix}$$

The equation $(\star)$ is written as a matrix product

$$Ax=\begin{pmatrix}a_{11}&\cdots&a_{1n}\\a_{21}&\cdots&a_{2n}\\\vdots&&\vdots\\a_{m1}&\cdots&a_{mn}\end{pmatrix}\begin{pmatrix}x_1\\\vdots\\x_n\end{pmatrix}$$

3.6. **Theorem — finding the matrix of $A:\mathbb{F}^n\to\mathbb{F}^m$.** If the linear map $A:\mathbb{F}^n\to\mathbb{F}^m$ is given by $(\star)$ then columns of the matrix of A are the vectors Ae_1, Ae_2, …, Ae_n, i.e.

$$Ae_1=\begin{pmatrix}a_{11}\\a_{21}\\\vdots\\a_{m1}\end{pmatrix},\quad Ae_2=\begin{pmatrix}a_{11}\\a_{22}\\\vdots\\a_{m2}\end{pmatrix},\quad Ae_3=\begin{pmatrix}a_{11}\\a_{23}\\\vdots\\a_{m3}\end{pmatrix},\quad\dots,\quad Ae_n=\begin{pmatrix}a_{11}\\a_{2n}\\\vdots\\a_{mn}\end{pmatrix}$$

Sometimes the following notation is used to express this:

$$A=\begin{bmatrix}|&|&&|\\Ae_1&Ae_2&\cdots&Ae_n\\|&|&&|\end{bmatrix}$$

The matrices of the zero and identity transformations are

$$O=\begin{pmatrix}0&0&\cdots&0\\0&0&\cdots&0\\\vdots&&&\vdots\\0&0&\cdots&0\end{pmatrix}\qquad I=\begin{pmatrix}1&0&\cdots&0\\0&1&\cdots&0\\&&\ddots&\\0&0&\cdots&1\end{pmatrix}.$$

3.7. **Definition — composition of linear transformations.** If V,W,U are vector spaces and if $T:V\to W$ and $S:W\to U$ are linear transformations, then the composition of S and T is the map $S\circ T:V\to U$, with $S\circ T(x)=S\big(T(x)\big)$

3.8. **Theorem.** The composition $S\circ T$ is linear.

3.9. **Theorem — matrix of the composition.** If $T : \mathbb{F}^n \to \mathbb{F}^m$ and $S : \mathbb{F}^m \to \mathbb{F}^\ell$ are linear, and if their matrices are given by

$$T = \begin{pmatrix} t_{11} & \dots & t_{1n} \\ \vdots & & \vdots \\ t_{m1} & \dots & t_{mn} \end{pmatrix} \qquad S = \begin{pmatrix} s_{11} & \dots & s_{1m} \\ \vdots & & \vdots \\ s_{\ell 1} & \dots & s_{\ell m} \end{pmatrix}$$

then the matrix of the composition $S \circ T$ is given by the matrix product

$$ST = \begin{pmatrix} s_{11} & \dots & s_{1m} \\ \vdots & & \vdots \\ s_{\ell 1} & \dots & s_{\ell m} \end{pmatrix} \begin{pmatrix} t_{11} & \dots & t_{1n} \\ \vdots & & \vdots \\ t_{m1} & \dots & t_{mn} \end{pmatrix}$$

The ij entry of ST is computed by taking the "dot product" of the i^{th} row of S with the j^{th} column of T, e.g.:

$$\overset{[ST]}{} \qquad \overset{[S]}{} \qquad \overset{[T]}{}$$

$$i \begin{bmatrix} \cdot & \cdot & \cdot & | & \cdot \\ — & — & — & z & — \\ \cdot & \cdot & \cdot & | & \cdot \end{bmatrix}^{j} = i \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ a & b & c & d \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot & \cdot & k & \cdot \\ \cdot & \cdot & \cdot & l & \cdot \\ \cdot & \cdot & \cdot & m & \cdot \\ \cdot & \cdot & \cdot & n & \cdot \end{bmatrix}^{j}$$

$$z = ak + bl + cm + dn.$$

3.10. **Definition — injective, surjective, bijective.** A linear transformation $T : V \to W$ is

- injective (one-to-one, 1-1) if for all $x, y \in V$ it is true that $T(x) = T(y)$ implies $x = y$
- surjective (onto) if for every $w \in W$ there is a $v \in V$ with$Tv = w$
- bijective (invertible) if T is both injective and surjective

3.11. **Definition — the inverse.** If the linear map $T : V \to W$ is bijective then its inverse $T^{-1} : W \to V$ is defined by

$$\forall x \in W, y \in V : \quad T^{-1}(x) = y \iff x = T(y).$$

3.12. **Theorem.** If $T : V \to W$ is bijective then the inverse $T^{-1} : W \to V$ is linear.

3.13. **Theorem.** If $T : V \to W$ is bijective, then $T^{-1} \circ T = I_V$ and $T \circ T^{-1} = I_W$.

3.14. **Theorem.** If $T : V \to W$ is a linear map, then consider the equation $Tx = y$, where $y \in W$ is given and $x \in V$ is unknown.

- If T is injective then the equation $Tx = y$ has at most one solution.
- If T is surjective then the equation $Tx = y$ has at least one solution.
- If T is bijective then the equation $Tx = y$ has exactly one solution.

3.15. **Definition — powers.** If $T : V \to V$ is linear, then T^k, the k^{th} power of T is defined by $T^k = \overbrace{T \cdot T \cdot T \cdots T}^{k \text{ factors}}$ if k is a positive integer. If T is invertible, then one also defines $T^{-k} = \left(T^{-1}\right)^k$.

3.16. **Theorem — power law.** $T^{k+l} = T^k T^l$ for all $k, l \in \mathbb{N}$ and all $T \in \mathscr{L}(V)$.

4. Linear Subspaces

4.1. **Definition.** Suppose V is a vector space. A subset $W \subset V$ is a linear subspace if it satisfies

- for all $x, y \in W$ one has $x + y \in W$
- for all $x \in W$ and $a \in \mathbb{F}$ one has $ax \in W$
- $0 \in W$

4.2. **Theorem.** If V is a vector space then any linear subspace $W \subset V$ is also a vector space.

4.3. **Examples — smallest and largest subspaces.** For any vector space V

- V is a subspace of V
- the set $\{0_V\}$ is a subspace of V

4.4. **Definition of Null Space and Range.** If $T : V \to W$ is a linear map then the null space of T is

$$N(T) = \{x \in V \mid Tx = 0_W\}$$

and the range of T is

$$R(T) = \{Tx \mid x \in V\}.$$

The null space is sometimes also called the kernel of T and one writes $N(T) = \ker(T)$.

4.5. **Theorem.** If $T : V \to W$ is linear then $N(T)$ is a linear subspace of V and $R(T)$ is a linear subspace of W.

4.6. **Definition.** If V is a vector space and if $v_1, \ldots, v_n \in V$ are given then the span of $v_1, \ldots, v_n$ is the set of all linear combinations of $v_1, \ldots, v_n$. In symbols:

$$\text{span}\{v_1, \ldots, v_n\} = \left\{a_1 v_1 + \cdots + a_n v_n \mid a_1, \ldots, a_n \in \mathbb{F}\right\}$$

4.7. **Theorem.** $\text{span}\{v_1, \ldots, v_n\}$ is a linear subspace of V.

4.8. **Injectivity Theorem.** A linear map $T : V \to W$ is injective *iff* $N(T) = \{0\}$.

"*iff*" is a common abbreviation for "if and only if"

4.9. **Surjectivity Theorem.** A linear map $T : V \to W$ is surjective *iff* $R(T) = W$.

5. Linear independence, Bases, and Dimension

5.1. **Definition of independence.** A set of vectors $\{u_1, \ldots, u_n\} \subset V$ is linearly independent if for any $a_1, \ldots, a_n \in \mathbb{F}$ one has $a_1 u_1 + \cdots + a_n u_n = 0 \implies a_1 = a_2 = \cdots = a_n = 0$.

5.2. **Definition of basis.** A set of vectors $\{u_1, \ldots, u_n\} \subset V$ is a basis for V if

- $\{u_1, \ldots, u_n\}$ is linearly independent, and
- $\{u_1, \ldots, u_n\}$ spans V.

5.3. **Extension Theorem for Independent Sets.** If $u_1, \dots, u_n \in V$ are linearly independent, and $v \in V$, then

$$v \in \text{span}(u_1, \dots, u_n) \iff u_1, \dots, u_n, v \text{ are dependent}$$

The proof is in Appendix A.2.

5.4. **The Basis Selection Theorem.** If $u_1, \dots, u_n \in V$ span the vector space V, then there exist $i_1, \dots, i_k \in \{1, \dots, n\}$ such that $\{u_{i_1}, \dots, u_{i_k}\}$ is a basis for V.

A proof is in Appendix A.3.

5.5. **The Dimension Theorem.** If $v_1, \dots, v_m \in \text{span}(u_1, \dots, u_n)$ and if $v_1, \dots, v_m$ are linearly independent then $m \le n$.

Appendix A.4 has the proof of this theorem.

5.6. **First consequence of the Dimension Theorem.** If $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_m\}$ both are bases of a vector space V, then $m = n$.

Proof. This is true because $\{v_1, \dots, v_m\} \subset \text{span}(u_1, \dots, u_n)$ and $\{v_1, \dots, v_m\}$ is linearly independent, so the Dimension Theorem implies $m \le n$.

On the other hand, $u_1, \dots, u_n \in \text{span}(v_1, \dots, v_n)$ and $\{u_1, \dots, u_n\}$ is linearly independent, so the Dimension Theorem implies $n \le m$.

Since $n \le m$ and $m \le n$ we conclude $n = m$. ////

5.7. **Definition of dimension.** If a vector space V has a basis $\{u_1, \dots, u_n\}$ with n elements, then n is the dimension of V.

If a vector space V has a basis with finitely many vectors then V is called finite dimensional.

The Dimension Theorem and its corollary imply that the dimension as defined above does not depend on which basis of V you consider.

5.8. **Finite dimensional subspace theorem.** If $L \subset V$ is a linear subspace and V is finite dimensional then $\dim L \le \dim V$. If $\dim L = \dim V$ then $L = V$.

5.9. **Extending a basis of a subspace to a basis of the whole space.** If $L \subset V$ is a linear subspace and V is finite dimensional, and if $\{v_1, \dots, v_k\} \subset L$ is a basis for L, then there exist vectors $v_{k+1}, \dots, v_n \in V$ such that $\{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$ is a basis for V.

5.10. **Definition of the Rank and Nullity of a linear transformation.** Let $T : V \to W$ be a linear transformation.

The rank of $T : V \to W$ is the dimension of the range of T.

The nullity of $T : V \to W$ is the dimension of the null space of T.

5.11. **Rank+Nullity Theorem.** If $T : V \to W$ is linear, and if V is finite dimensional, then

$$\dim N(T) + \dim R(T) = \dim V.$$

See Appendix A.4 for the proof.

5.12. **Bijectivity Theorem.** If V is a finite dimensional vector space, and if $T : V \to V$ is a linear transformation then the following are equivalent:

(1) T is injective (one-to-one)
(2) $N(T) = \{0\}$
(3) $\operatorname{rank} T = \dim V$
(4) T is surjective (onto)

6. Determinants

6.1. **Permutations.** A permutation of $(1, 2, \ldots, n)$ is a sequence of n integers $i_1, \ldots, i_n \in \{1, \ldots, n\}$ such that $i_k \neq i_l$ for all $k \neq l$.

For example, all possible permutations of $(1, 2, 3)$ are

$$(1,2,3), \quad (1,3,2), \quad (2,1,3), \quad (2,3,1), \quad (3,1,2), \quad (3,2,2).$$

There are $n!$ permutations of $(1, 2, \ldots, n)$.

6.2. **Sign of a permutation.** By definition, the number of **inversions** of a permutation $(i_1, \ldots, i_n)$ is

$$\delta(i_1, \ldots, i_n) \stackrel{\text{def}}{=} \#\{k < l \mid i_k > i_l\}$$

A permutation $(i_1, i_2, \ldots, i_n)$ is called **even** if $\delta(i_1, \ldots, i_n)$ is even.

A permutation $(i_1, i_2, \ldots, i_n)$ is called **odd** if $\delta(i_1, \ldots, i_n)$ is odd.

The sign of the permutation $(i_1, \ldots, i_n)$ is defined to be

$$\epsilon_{i_1 i_2 \cdots i_n} = (-1)^{\delta(i_1, i_2, \ldots, i_n)} = \begin{cases} +1 & \text{if } \delta(i_1, \ldots, i_n) \text{ is even} \\ -1 & \text{if } \delta(i_1, \ldots, i_n) \text{ is odd} \end{cases}$$

The quantity $\epsilon_{i_1 i_2 \cdots i_n}$ is called the Levi-Civita symbol.

6.3. **Definition of the determinant.**

$$\det A = \sum_{i_1, \ldots, i_n} \epsilon_{i_1 i_2 \cdots i_n} a_{1 i_1} a_{2 i_2} \cdots a_{n i_n}$$

6.4. **Properties of the sign** $\epsilon_{i_1 i_2 \cdots i_n}$**.**

- $\epsilon_{i_1 \cdots i_k \cdots i_\ell \cdots i_n} = -\epsilon_{i_1 \cdots i_\ell \cdots i_k \cdots i_n}$
- if $i_1, \ldots, i_k < i_{k+1}, \ldots, i_n$ then $\epsilon_{i_1 i_2 \cdots i_n} = \epsilon_{i_1 \cdots i_k} \epsilon_{i_{k+1} \cdots i_n}$

6.5. **What is the sign of** $a_{14} a_{22} a_{35} a_{43} a_{51}$ **when you expand a** 5×5 **determinant?**

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{vmatrix}$$

The question can be rephrased as: compute ϵ_{42531}

6.6. **Special case — 2×2 determinants.**

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

6.7. **Special case — 3×3 determinants.**

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1b_2c_3 - a_1b_3c_2 - a_2b_1c_3 + a_2b_3c_1 + a_3b_1c_2 - a_3b_2c_1$$

6.8. **Determinant of upper triangular matrices.** If all entries of a matrix below its diagonal are zero, then the determinant of the matrix is the product of its diagonal entries:

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots a_{nn}$$

6.9. **Determinant of the identity matrix.**

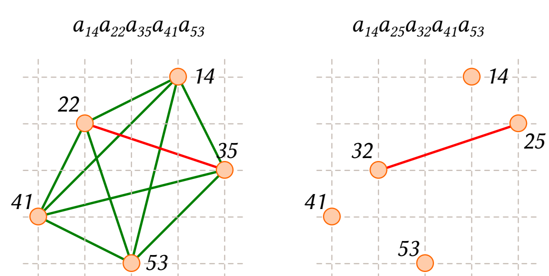
$$\det I = 1.$$

6.10. **Determinant of block triangular matrices.** If A is a $k \times k$ matrix, B a $k \times l$ matrix, and C an $l \times l$ matrix then

$$\begin{vmatrix} A & B \\ 0 & C \end{vmatrix} = (\det A)(\det C)$$

6.11. **Determinant of the transpose.** $\det A^\top = \det A$

6.12. **Swapping rows or columns changes the sign.** If B is the matrix you get by swapping two rows, or by swapping two columns in the matrix A, then $\det A = -\det B$



6.13. **The determinant as a function of its rows.** If we have n row vectors $a_1,\dots,a_n \in \mathbb{F}^n$, given by

$$\begin{aligned} a_1 &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}, \\ a_2 &= \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}, \\ &\vdots \\ a_n &= \begin{pmatrix} a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \end{aligned}$$

then we define

$$\det(a_1,a_2,\dots,a_n) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

One has for all $a,b,a_1,a_2,\dots,a_n \in \mathbb{F}^n$ and $t \in \mathbb{F}$:

$$\begin{aligned} \det(a+b,a_2,a_3,\dots,a_n) &= \det(a,a_2,a_3,\dots,a_n) + \det(b,a_2,a_3,\dots,a_n) \\ \det(ta,a_2,a_3,\dots,a_n) &= t\det(a,a_2,a_3,\dots,a_n) \\ \det(a_1,\dots,a_i,\dots,a_j,\dots a_n) &= -\det(a_1,\dots,a_j,\dots,a_i,\dots a_n) \\ \det(a_1,\dots,a_i,\dots,a_i,\dots a_n) &= 0 \\ \det(a_1,\dots,a_i,\dots,a_j,\dots a_n) &= \det(a_1,\dots,a_i+ta_j,\dots,a_j,\dots a_n) \end{aligned}$$

6.14. **Cofactor expansion.** The ij-minor of an $n\times n$ matrix A is the $(n-1)\times(n-1)$ matrix obtained by deleting the i^{th} row and j^{th} column from A. Let us write $\tilde{A}_{ij}$ for the ij-minor of A.

The ij-cofactor of the matrix A is the number

$$c_{ij} = (-1)^{i+j}\det\tilde{A}_{ij}$$

6.15. **Cofactor Expansion Theorem.** If $A = (a_{ij})$ is an $n\times n$ matrix, then one has

$$\det A = a_{i1}c_{i1} + a_{i2}c_{i2} + \cdots + a_{in}c_{in}$$

for any $i \in \{1,2,\dots,n\}$.

A **consequence** of the cofactor expansion theorem is that if $i \neq j$ then

$$a_{j1}c_{i1} + a_{j2}c_{i2} + \cdots + a_{jn}c_{in} = 0.$$

6.16. **Example.** Expanding a 3×3 determinant along its middle row:

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ -2 & 4 & 3 \end{vmatrix} &= a_{21}c_{21} + a_{22}c_{22} + a_{23}c_{23} \\ &= -3\cdot\begin{vmatrix} 2 & 3 \\ 4 & 3 \end{vmatrix} + 2\cdot\begin{vmatrix} 1 & 3 \\ -2 & 3 \end{vmatrix} - 1\cdot\begin{vmatrix} 1 & 2 \\ -2 & 4 \end{vmatrix} = \cdots \end{aligned}$$

6.17. **A formula to invert a matrix.** For any $n \times n$ matrix A one has

$$AC^\top = C^\top A = (\det A)I,$$

where C is the cofactor matrix of A. If $\det A \neq 0$ then A is invertible, and the inverse matrix is given by

$$A^{-1} = \frac{1}{\det A} C^\top.$$

Proof: see Appendix A.6.

6.18. **Cramer's rule.** For any $y \in \mathbb{F}^n$ the solution of $Ax = y$ is given by

$$x_1 = \frac{\begin{vmatrix} y_1 & a_{12} & \dots & a_{1n} \\ y_2 & a_{22} & \dots & a_{2n} \\ & & \dots & \\ y_n & a_{n2} & \dots & a_{nn} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}}, \quad x_2 = \frac{\begin{vmatrix} a_{11} & y_1 & \dots & a_{1n} \\ a_{21} & y_2 & \dots & a_{2n} \\ & & \dots & \\ a_{n1} & y_n & \dots & a_{nn} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}}, \quad \dots, \quad x_n = \frac{\begin{vmatrix} a_{11} & a_{12} & \dots & y_1 \\ a_{21} & a_{22} & \dots & y_2 \\ & & \dots & \\ a_{n1} & a_{n2} & \dots & y_n \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}}$$

A proof is in Appendix A.7.

6.19. **Example — the inverse of a 2×2 matrix.**

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

6.20. **Determinant of a matrix product.** For any two $n \times n$ matrices A and B one has

$$\det(AB) = \det(A)\det(B)$$

See Appendix A.8 for a proof.

6.21. **Theorem.** A is invertible $\iff \det A \neq 0$

The proof is short: If A is invertible then the matrix $B = A^{-1}$ satisfies $AB = I$. This implies $\det(AB) = 1$ and hence $(\det A)(\det B) = 1$. Therefore $\det A \neq 0$.

Conversely, if $\det A \neq 0$ then A is invertible because we have a formula for its inverse, $A^{-1} = \frac{C^\top}{\det A}$.

6.22. **The determinant independence test.** If $a_1, \dots, a_n \in \mathbb{F}^n$ then

$$\{a_1, \dots, a_n\} \text{ is linearly independent } \iff \det(a_1, \dots, a_n) \neq 0$$

The proof is in Appendix A.9.

7. EIGENVALUES, SPECTRAL THEORY

7.1. **Definition.** Let V be a vector space and $A : V \to V$ a linear transformation. If $v \in V$ and $\lambda \in \mathbb{F}$ satisfy

$$Av = \lambda v, \quad v \neq 0$$

then v is called an eigenvector of A and λ is called an eigenvalue of A.

The set

$$\operatorname{Spec}(A) \stackrel{\text{def}}{=} \{\lambda \mid \lambda \text{ is an eigenvalue of } A\}$$

is called the spectrum of A.

7.2. **Theorem.** If $A : V \to V$ is a linear transformation then

- ▶ $v \in V$ is an eigenvector of A with eigenvalue $\lambda \iff v \neq 0$ and $v \in N(\lambda I - A)$
- ▶ $\lambda \in \mathbb{F}$ is an eigenvalue of A if and only if $N(\lambda I - A) \neq \{0\}$
- ▶ $\lambda \in \mathbb{F}$ is an eigenvalue of A if and only if $\det(\lambda I - A) = 0$

This theorem gives a method of finding all eigenvalues and vectors, namely:

- Compute the characteristic polynomial $\det(\lambda I - A)$
- Solve $\det(\lambda I - A) = 0$ for λ. Let $\lambda_1, \ldots, \lambda_k$ be all the solutions you find
- For each $i = 1, \ldots, k$ find all the vectors in $N(\lambda_i I - A)$, i.e. solve $\lambda_i v - Av = 0$ for v.

7.3. **Finding eigenvalues/vectors for $A : \mathbb{F}^n \to \mathbb{F}^n$.** The characteristic polynomial of $A : \mathbb{F}^n \to \mathbb{F}^n$ is, by definition,

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ & & \ddots & \\ -a_{11} & -a_{12} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

After expanding, this turns out to be a polynomial in λ of degree n

$$\det(\lambda I - A) = \lambda^n + c_1 \lambda^{n-1} + c_2 \lambda^{n-2} + \cdots + c_{n-1}\lambda + c_n.$$

Here

$$-c_1 = a_{11} + a_{22} + \cdots + a_{nn}$$

is called the trace of the matrix A, and

$$(-1)^n c_n = \det A$$

is its determinant.

7.4. **Independence of Eigenvectors Theorem.** If $v_1, \ldots, v_m \in V$ are eigenvectors of $A : V \to V$ with eigenvalues $\lambda_1, \ldots, \lambda_m$, and if the eigenvalues are distinct (i.e. $\lambda_i \neq \lambda_j$ for all $i \neq j$) then $\{v_1, \ldots, v_m\}$ is linearly independent.

7.5. **The use of eigenvalues and vectors.** If a vector $v \in V$ is known as a linear combination of eigenvectors of A then it is easy to compute Av, A^2v, etc. If

$$v = a_1 v_1 + \cdots + a_k v_k$$

where $Av_i = \lambda_i v_i$, then

$$\begin{aligned} Av &= \lambda_1 a_1 v_1 + \cdots + \lambda_k a_k v_k \\ A^2 v &= \lambda_1^2 a_1 v_1 + \cdots + \lambda_k^2 a_k v_k \\ &\vdots \\ A^\ell v &= \lambda_1^\ell a_1 v_1 + \cdots + \lambda_k^\ell a_k v_k \end{aligned}$$

for any $\ell \in \mathbb{N}$. If none of the eigenvalues λ_i vanishes then one solution of $Aw = v$ is given by

$$w = \frac{a_1}{\lambda_1} v_1 + \cdots + \frac{a_k}{\lambda_k} v_k.$$

For this to be useful we need to find as many eigenvectors as we can.

7.6. **Solving a system of linear differential equations.**

$$\begin{aligned} \frac{dx_1}{dt} &= a_{11}x_1 + \cdots + a_{1n}x_n \\ \frac{dx_2}{dt} &= a_{21}x_1 + \cdots + a_{2n}x_n \\ &\vdots \\ \frac{dx_n}{dt} &= a_{n1}x_1 + \cdots + a_{nn}x_n \end{aligned} \iff \frac{dx}{dt} = Ax, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

If $v_1, \dots, v_n$ are a basis of eigenvectors of A with eigenvalues $\lambda_1, \dots, \lambda_n$, then a solution is a vector function of the form

$$x(t) = c_1(t)v_1 + \cdots + c_n(t)v_n.$$

$$x'(t) = c_1'(t)v_1 + \cdots + c_n'(t)v_n \text{ and } Ax(t) = c_1(t)\lambda_1 v_1 + \cdots + c_n(t)\lambda_n v_n$$

The diffeq $x' = Ax$ is then equivalent with

$$c_1'(t) = \lambda_1 c_1(t), \quad \dots \quad c_n'(t) = \lambda_n c_n(t)$$

whose solutions are $c_1(t) = e^{\lambda_1 t}c_1(0)$, ..., $c_n(t) = e^{\lambda_n t}c_n(0)$.

The solution $x(t)$ is therefore

$$x(t) = e^{\lambda_1 t}c_1(0)v_1 + \cdots + e^{\lambda_n t}c_n(0)v_n$$

Diagonalization

Since the characteristic polynomial $\det(\lambda I - A)$ is a polynomial of degree n, it has at most n zeroes. If it has n distinct zeroes, λ_1, ..., λ_n then we choose an eigenvector v_i for each eigenvalue λ_i. The vectors $v_1, \dots, v_n$ are linearly independent in $\mathbb{F}^n$ and therefore they form a basis for $\mathbb{F}^n$.

Diagonalization Theorem (version 1). If a linear transformation $A : V \to V$ has a basis $v_1, \dots, v_n \in V$ of eigenvectors, with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, then the matrix of A with respect to this basis is

$$[A]_{v_1,\dots,v_n} = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

This follows directly from the fact that $Av_i = \lambda_i v_i$ for $i = 1, 2, \dots, n$.

Diagonalization Theorem (version 2). If an $n \times n$ matrix A has a basis $v_1, \dots, v_n \in \mathbb{F}^n$ of eigenvectors, with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, then we have

$$S^{-1}AS = D,$$

where

$$S = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$$

is the matrix whose columns are the eigenvectors, and D is the diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

7.7. **Definition – Similarity of linear transformations.** Two linear transformations $A : V \to V$ and $B : W \to W$ are called similar if there is an invertible linear transformation $S : V \to W$ such that

$$SA = BS, \text{ or } A = S^{-1}BS, \text{ or } SAS^{-1} = B.$$

The three conditions are equivalent.

8. Inner product spaces, Symmetric Matrices, and their Eigenvalues

In the theory of inner product spaces we assume that the number field $\mathbb{F}$ is either the real numbers or the complex numbers:

$$\mathbb{F} = \mathbb{R} \text{ or } \mathbb{F} = \mathbb{C}$$

The theories for real and complex inner products are very similar.

8.1. **Definition.** An real inner product on a real vector space V is a real valued function on $V \times V$, usually written as (x, y) or $\langle x, y \rangle$ that satisfies the following properties

- $\langle x, y \rangle = \langle y, x \rangle$
- $\langle ax, y \rangle = a\langle x, y \rangle$
- $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- $\langle x, x \rangle > 0$ if $x \neq 0$

for all $x, y, z \in V$ and $a \in \mathbb{R}$.

8.2. **Definition.** A complex inner product on a complex vector space V is a complex valued function on $V \times V$, usually written as (x, y) or $\langle x, y \rangle$ that satisfies the following properties

- $\langle x, y \rangle = \overline{\langle y, x \rangle}$ where $\overline{z}$ is the complex conjugate of $z \in \mathbb{C}$
- $\langle ax, y \rangle = a\langle x, y \rangle$
- $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- $\langle x, x \rangle > 0$ if $x \neq 0$

for all $x, y, z \in V$ and $a \in \mathbb{C}$.

8.3. **Example — $\mathbb{R}^n$ and the dot product.** On $\mathbb{R}^n$ we have the dot-product from vector calculus, i.e.

$$\langle x, y \rangle = x \cdot y \stackrel{\text{def}}{=} x_1 y_1 + \cdots + x_n y_n$$

for any two vectors $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$.

A variation on this example is the weighted dot product given by

$$\langle x, y \rangle_w \stackrel{\text{def}}{=} w_1 x_1 y_1 + \cdots + w_n x_n y_n$$

for all $x, y \in \mathbb{R}^n$, and where $w_1, \ldots, w_n > 0$ are given constants, called the weights in the inner product.

8.4. **Example — $\mathbb{C}^n$ and the complex dot product.** For any two vectors $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{C}^n$ the complex dot product of x and y is defined by:

$$\langle x, y \rangle = x \cdot y \stackrel{\text{def}}{=} x_1 \overline{y_1} + \cdots + x_n \overline{y_n}$$

8.5. **Example — Inner product on a Function Space.** Let V be the space of continuous functions $f : [0, 1] \to \mathbb{R}$. Then

$$\langle f, g \rangle \stackrel{\text{def}}{=} \int_0^1 f(t) g(t) \, dt$$

defines an inner product on V.

This kind of inner product plays a large role in Quantum Mechanics and in the theory of Fourier series.

8.6. **Inner Product when you have a Basis.** If $\{v_1, \dots, v_n\}$ is a basis for a complex vector space, and if $x, y \in V$ satisfy

$$x = x_1 v_1 + \dots + x_n v_n, \quad y = y_1 v_1 + \dots + y_n v_n,$$

then

$$\langle x, y \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} x_i \overline{y_j} \quad \text{where} \quad g_{ij} \stackrel{\text{def}}{=} \langle v_i, v_j \rangle.$$

Norms, Distances, Angles, and Inequalities

8.7. **Definition.** Let V be a real or complex inner product space with inner product $\langle x, y \rangle$. Then the norm (or length) of a vector $x \in V$ is

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Note that $\langle x, x \rangle$ is never negative, so the square root is always defined.

8.8. **Theorem — properties of the length.**

- $\|x\| > 0$ for all $x \in V$ with $x \neq 0$
- $\|ax\| = |a| \, \|x\|$ for all $x \in V$ and $a \in \mathbb{F}$
- $|\langle x, y \rangle| \leq \|x\| \, \|y\|$ (the Cauchy-Schwarz inequality)
- $\|x + y\| \leq \|x\| + \|y\|$ (the triangle inequality)

8.9. **Definition.** The distance between two vectors $x, y \in V$ is $d(x, y) \stackrel{\text{def}}{=} \|x - y\|$.

8.10. **Definition.** Two vectors $x, y \in V$ are said to be orthogonal if $\langle x, y \rangle = 0$.

Notation: $x \perp y$ means x, y are orthogonal.

In particular, the zero vector is orthogonal to every other vector, because $\langle x, 0 \rangle = 0$ for all $x \in V$.

More generally, if V is a real inner product space, then the angle between two non-zero vectors $x, y \in V$ is defined to be

$$\angle(x, y) \stackrel{\text{def}}{=} \arccos \frac{\langle x, y \rangle}{\|x\| \, \|y\|}$$

The Cauchy-Schwarz inequality implies $-1 \leq \frac{\langle x, y \rangle}{\|x\| \, \|y\|} \leq 1$ so the inverse cosine is always defined.

8.11. **Example in $\mathbb{R}^4$.** Find the lengths and the angle between the vectors $u = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 3 \end{pmatrix}$ and $v = \begin{pmatrix} -1 \\ 2 \\ 1 \\ 3 \end{pmatrix}$ with respect to to the standard inner product on $\mathbb{R}^4$. Also find the distance between u and v.

Solution: We have

$$\|u\| = \sqrt{1^2 + 0^2 + 2^2 + 3^2} = \sqrt{14}$$

$$\|v\| = \sqrt{(-1)^2 + 2^2 + 1^2 + 3^2} = \sqrt{15}$$

$$\langle u, v \rangle = 1 \cdot (-1) + 0 \cdot 2 + 2 \cdot 1 + 3 \cdot 3 = 10.$$

Therefore

$$\angle(u,v) = \arccos\frac{10}{\sqrt{14}\sqrt{15}} = \arccos\frac{10}{\sqrt{210}} \approx 46.3647\ldots^\circ$$

Finally, the distance between u and v is

$$\|u-v\| = \sqrt{(1-(-1))^2+(0-2)^2+(2-1)^2+(3-3)^2} = \sqrt{10}.$$

8.12. **Example in a function space.** Find the lengths and angle between the functions $f(t)=1$ and $f(t)=t^2$ in the real function space $V=C([0,1])$ with inner product $\langle f,g\rangle = \int_0^1 f(t)g(t)dt$.

Solution: By definition we have

$$\begin{aligned}
\|f\| &= \sqrt{\int_0^1 1^2 dt} = \sqrt{1} = 1\\
\|g\| &= \sqrt{\int_0^1 \left(t^2\right)^2 dt} = \sqrt{\int_0^1 t^4 dt} = \sqrt{\frac{1}{5}} = \frac{1}{5}\sqrt{5}\\
\langle f,g\rangle &= \int_0^1 1\cdot t^2\, dt = \frac{1}{3}\\
\cos\angle(f,g) &= \frac{1/3}{1\cdot\frac{1}{5}\sqrt{5}} = \frac{1}{3}\sqrt{5} \implies \angle(f,g) = \arccos\frac{\sqrt{5}}{3} \approx 41.81\ldots^\circ
\end{aligned}$$

Orthogonal sets of vectors

Let V be a real or complex inner product space

8.13. **Definition.** A set of vectors $\{v_1,\ldots,v_k\}\subset V$ is orthogonal if

- $v_i \neq 0$ for all i
- $v_i \perp v_j$ for all $i \neq j$

An orthogonal set $\{v_1,\ldots,v_k\}\subset V$ is called orthonormal if $\|v_i\| = 1$ for $i=1,\ldots,k$.

8.14. **Theorem.** If $\{v_1,\ldots,v_n\}\subset V$ is orthogonal, then $\{v_1,\ldots,v_n\}$ is linearly independent.

If, in addition, $n=\dim V$, then $\{v_1,\ldots,v_n\}$ is a basis for V. In this case $\{v_1,\ldots,v_n\}$ is called an orthogonal basis.

For vectors expressed in terms of an orthogonal basis one has the following formulas for the inner product and norm: if $x = x_1v_1+\cdots+x_nv_n$, $y = y_1v_1+\cdots+y_nv_n$, then

$$\langle x,y\rangle = w_1x_1y_1+\cdots+w_nx_ny_n, \qquad \|x\|^2 = w_1x_1^2+\cdots+w_nx_n^2$$

where the weights w_i are given by $w_i = \|v_i\|^2$.

If the basis $\{v_1,\ldots,v_n\}$ is orthonormal, then $w_i = \|v_i\|^2 = 1$ for all i, and thus

$$\langle x,y\rangle = x_1y_1+\cdots+x_ny_n, \qquad \|x\|^2 = x_1^2+\cdots+x_n^2$$

8.15. **Theorem.** Every finite dimensional inner product space has an orthonormal basis.

8.16. **Proof using the Gram–Schmidt procedure.** Let $\{v_1, \dots, v_m\}$ be a basis of V. Define

$$w_1 = v_1 \qquad u_1 = \frac{v_1}{\|v_1\|}$$
$$w_2 = v_2 - \langle v_2, u_1\rangle u_1 \qquad u_2 = \frac{w_2}{\|w_2\|}$$
$$w_3 = v_3 - \langle v_3, u_1\rangle u_1 - \langle v_3, u_2\rangle u_2 \qquad u_3 = \frac{w_3}{\|w_3\|}$$

and in general,

$$w_j = v_j - \langle v_j, u_1\rangle u_1 - \dots - \langle v_j, u_{j-1}\rangle u_{j-1},$$
$$u_j = \frac{w_j}{\|w_j\|}$$

Then $\{u_1, \dots, u_n\}$ is an orthonormal basis of V. More precisely, by induction on j one shows that

- $w_j \perp \{v_1, \dots, v_{j-1}\}$ and hence $u_j \perp \{v_1, \dots, v_{j-1}\}$
- $\text{span}\{v_1, \dots, v_j\} = \text{span}\{u_1, \dots, u_j\}$

The above procedure that created the orthonormal basis $\{u_1, \dots, u_n\}$ from the given basis $\{v_1, \dots, v_n\}$ is called Gram–Schmidt orthognoalization.

Examples

8.17. **Example — in the plane.** The vectors $\binom{1}{1} \in \mathbb{R}^2$ and $\binom{1}{-1} \in \mathbb{R}^2$ are orthogonal. Therefore they are independent, and, since there are two of them, they form a basis of $\mathbb{R}^2$. Since, they are orthogonal, $\binom{1}{1}, \binom{1}{-1}$ is an orthogonal basis for $\mathbb{R}^2$.

On the other hand $\|\binom{1}{1}\| = \|\binom{1}{-1}\| = \sqrt{2} \neq 1$ so $\binom{1}{1}, \binom{1}{-1}$ is not an orthonormal basis for $\mathbb{R}^2$.

8.18. **Example — in $\mathbb{R}^n$ and $\mathbb{C}^n$.** The standard basis $\{e_1, \dots, e_n\}$ is orthonormal and hence is an orthonormal basis for $\mathbb{R}^n$ and also for $\mathbb{C}^n$.

8.19. **Example — in a function space.** Take $\mathbb{F} = \mathbb{R}$.

We define V to be the set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ that are 2π-periodic, i.e. all continuous functions $f : \mathbb{R} \to \mathbb{R}$ that satisfy

$$\forall x \in \mathbb{R} : f(x + 2\pi) = f(x).$$

Then V is a real vector space, and one defines an inner product on V by setting $\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)\,dt$.

The set of functions

$$\cos t, \cos 2t, \cos 3t, \dots$$

is orthogonal. To prove this, compute the integrals

$$\int_0^{2\pi} \cos nt \cos mt\, dt = \begin{cases} \pi & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$$

We can add more functions to this set and still have an orthogonal set: the set of functions

$$\beta : \qquad 1, \cos t, \sin t, \cos 2t, \sin 2t, \cos 3t, \sin 3t, \dots$$

is also an orthogonal set in V.

The theory of Fourier series says that β is a "basis" for V, in the sense that every function $f \in V$ can be written as

$$f(t) = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t + a_3 \cos 3t + b_3 \sin 3t + \cdots$$

for suitable $a_0, a_1, b_1, \cdots \in \mathbb{R}$. This statement does not quite fit in the linear algebra from this course because the sum above contains infinitely many terms.

THE ADJOINT OF A MATRIX

8.20. **Definition.** Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. If $A : \mathbb{F}^n \to \mathbb{F}^n$ has matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix},$$

then the adjoint of A is the complex conjugate of the transpose of A, i.e.

$$A^* = \overline{A^\top} = \begin{pmatrix} \overline{a_{11}} & \cdots & \overline{a_{n1}} \\ \vdots & & \vdots \\ \overline{a_{1n}} & \cdots & \overline{a_{nn}} \end{pmatrix}.$$

In the real case there is no need to take the complex conjugate, and for real matrices one has $A^* = A^\top$.

8.21. **Definition.**

- A real $n \times n$ matrix is called symmetric if $A = A^\top$.
- A complex $n \times n$ matrix is called Hermitian if $A = \overline{A^\top}$.

In both cases the matrix, and the corresponding linear operator $A : \mathbb{F}^n \to \mathbb{F}^n$ are called self-adjoint.

8.22. **Theorem.** Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. For any $x, y \in \mathbb{F}^n$ and any matrix A one has

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

8.23. **Example.** The matrices

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

are Hermitian.

8.24. **Theorem.** All eigenvalues of a Hermitian matrix are real

Proof. If $Av = \lambda v$, then

$$\lambda \|v\|^2 = \langle Av, v \rangle = \langle v, A^* v \rangle = \langle v, Av \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \langle v, v \rangle = \overline{\lambda} \|v\|^2.$$

Since $\|v\| \neq 0$, this implies $\lambda = \overline{\lambda}$, i.e. λ is real.

8.25. **Theorem.** If $A : \mathbb{F}^n \to \mathbb{F}^n$ is Hermitian and if v, w are eigenvectors corresponding to different eigenvalues $\lambda \neq \mu$, then $v \perp w$.

Proof. Since A is self adjoint, all its eigenvalues are real, and thus $\lambda, \mu \in \mathbb{R}$.

We have $Av = \lambda v$ and $Aw = \mu w$, and therefore

$$\lambda \langle v, w \rangle = \langle Av, w \rangle = \langle v, Aw \rangle = \left\langle v, \mu w \right\rangle = \overline{\mu} \langle v, w \rangle = \mu \langle v, w \rangle$$

because $\mu \in \mathbb{R}$. We find

$$(\lambda - \mu) \langle v, w \rangle = 0.$$

The eigenvalues λ and μ are different, so $\lambda - \mu \neq 0$. Therefore $\langle v, w \rangle = 0$.

8.26. **The Spectral Theorem.** Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and let $A : \mathbb{F}^n \to \mathbb{F}^n$ be Hermitian. Then $\mathbb{F}^n$ has an orthonormal basis consisting of eigenvectors of A.

For a proof see appendix A.11.

A.1. **Proof of the Components from a Basis Theorem.** Suppose $u_1,\dots,u_n$ is a basis. By definition, every vector is a linear combination of $u_1,\dots,u_n$, so $u_1,\dots,u_n$ spans V.

If $c_1u_1+\cdots+c_nu_n=0$, then $c_1u_1+\cdots+c_nu_n=0=0\cdot u_1+\cdots+0\cdot u_n=0$. Since $u_1,;u_n$ is a basis we the coefficients in the expansion are unique, i.e. $c_1=0$, $\dots$, $c_n=0$. Therefore the vectors are linearly independent.

Next, we show the converse: suppose $u_1,\dots,u_n$ spans V and is linearly independent. Then, for any vector $x\in V$ there exist numbers $x_1,\dots,x_n\in\mathbb{F}$ such that

$$x=x_1u_1+\cdots+x_nu_n.$$

The coefficients $x_1,\dots,x_n$ are uniquely determined by the vectors x and $u_1,\dots,u_n$. Namely, if $x_1'u_1+\cdots+x_n'u_n=x_1u_1+\cdots+x_nu_n$ then subtract x_1u_1, $\dots$, x_nu_n from both sides. We end up with $(x_1'-x_1)u_1+\cdots+(x_n'-x_n)u_n=0$. Since $u_1,\dots,u_n$ are independent, this implies $x_1=x_1'$, $\dots$, $x_n=x_n'$.

A.2. **Proof of the Extension Theorem for Independent Sets.** The theorem claims "if and only if" so we have to prove two implications.

First we prove: $v\in\operatorname{span}(u_1,\dots,u_n)\implies\{u_1,\dots,u_n,v\}$ is linearly dependent.

If $v\in\operatorname{span}(u_1,\dots,u_n)$ then there are $a_1,\dots,a_n\in\mathbb{F}$ with $v=a_1u_1+\cdots+a_nu_n$. Therefore $-a_1u_1-\cdots-a_nu_n+v=0$, so we have a nontrivial linear combination of $\{u_1,\dots,u_n,v\}$ that adds up to zero. Hence $\{u_1,\dots,u_n,v\}$ is linearly dependent.

Next, we show: $\{u_1,\dots,u_n,v\}$ is linearly dependent $\implies v\in\operatorname{span}(u_1,\dots,u_n)$.

Suppose $\{u_1,\dots,u_n,v\}$ is linearly dependent. Then there exist $a_1,\dots,a_n,b\in\mathbb{F}$ such that $a_1u_1+\cdots+a_nu_n+bv=0$, and such that at least one of the coefficients $a_1,\dots,a_n,b$ is nonzero.

If $b=0$ then we have $a_1u_1+\cdots+a_nu_n=0$. Since $\{u_1,\dots,u_n\}$ is independent, this implies $a_1=\cdots=a_n=0$, which is impossible because at least one of $a_1,\dots,a_n,b$ does not vanish.

Therefore $b\neq 0$. This implies that $v=-\frac{a_1}{b}u_1-\cdots-\frac{a_n}{b}u_n\in\operatorname{span}(u_1,\dots,u_n)$.

A.3. **Proof of the Bases Selection Theorem.** For any subset $\{i_1,\dots,i_k\}\subset\{1,\dots,n\}$ the vectors $u_{i_1},\dots,u_{i_k}$ either are independent or they are dependent.

Of all possible choices $\{i_1,\dots,i_k\}\subset\{1,\dots,n\}$for which the vectors $u_{i_1},\dots,u_{i_k}$ are independent, choose one with the largest possible k. For any $i\in\{1,\dots,n\}$ with $i\neq i_1,\dots,i\neq i_k$ the vectors $u_{i_1},\dots,u_{i_k},u_i$ must be dependent. The Extension Theorem 5.3 implies that $u_i\in\operatorname{span}(u_{i_1},\dots,u_{i_k})$.

We have shown that every u_i either is one of the $u_{i_1},\dots,u_{i_k}$, or else is a linear combination of $u_{i_1},\dots,u_{i_k}$.

Since every $v\in V$ is a linear combination of $u_1,\dots,u_n$ it follows that v also is a linear combination of $u_{i_1},\dots,u_{i_k}$.

Hence $u_{i_1},\dots,u_{i_k}$ span V. Since they are also independent we have shown that $u_{i_1},\dots,u_{i_k}$ is a basis for V.

A.4. **Proof of the dimension theorem.** Statement of the theorem: If $v_1,\dots,v_m \in \operatorname{span}(u_1,\dots,u_n)$ and if $v_1,\dots,v_m$ are linearly independent then $m \le n$.

We will show that if $m > n$ and $v_1,\dots,v_m$ are linear combinations of $u_1,\dots,u_n$, then $\{v_1,\dots,v_m\}$ is linearly dependent.

To do this we use mathematical induction on n.

We begin with the case $n = 1$. There is only one vector u_1, and each v_1 and v_2 are linear combinations of u_1, i.e. multiples of u_1. Thus for certain numbers $a_1, a_2 \in \mathbb{F}$ we have $v_1 = a_1u_1, \quad v_2 = a_2u_1$. If $a_1 = 0$ then $v_1 = 0$ and $\{v_1, v_2\}$ is dependent.

If $a_1 \neq 0$ then we have $-a_2v_1 + a_1v_2 + 0\cdot v_3 + \cdots + 0\cdot v_m = 0$. Since $a_1 \neq 0$ this is a nontrivial linear combination of $v_1,\dots,v_m$ that adds up to zero. Hence $\{v_1,\dots,v_m\}$ is dependent.

Next we consider the general case $n > 1$, and we assume that the case $n-1$ has already been proven.

In this case each v_i is a linear combination of the vectors $u_1,\dots,u_n$. So we have

$$\begin{aligned} v_1 &= a_{11}u_1 + \cdots + a_{1n}u_n \\ v_2 &= a_{21}u_1 + \cdots + a_{2n}u_n \\ &\vdots \\ v_m &= a_{m1}u_1 + \cdots + a_{mn}u_n \end{aligned}$$

for certain numbers $a_{11},\dots,a_{mn} \in \mathbb{F}$.

If all the coefficients $a_{1n}, a_{2n},\dots,a_{mn} = 0$ then the v_i are linear combinations of $u_1,\dots,u_{n-1}$. Since $m > n$ we have $m > n-1$ and therefore the induction hypothesis tells us that $v_1,\dots,v_m$ are linearly dependent.

We are left with the case in which one of the coefficients $a_{1n},\dots,a_{mn}$ does not vanish. Assume that $a_{1n} \neq 0$. Then we consider the vectors

$$w_2 = v_2 - \frac{a_{2n}}{a_{1n}}v_1, \quad \dots \quad w_m = v_m - \frac{a_{mn}}{a_{1n}}v_1.$$

The vectors $w_2,\dots,w_m$ are linear combinations of $u_1,\dots,u_{n-1}$. Since $m-1 > n-1$ the induction hypothesis applies. We therefore know that $w_2,\dots,w_m$ are linearly dependent, i.e. there exist $c_2,\dots,c_m \in \mathbb{F}$ such that

$$c_2w_2 + \cdots + c_mw_m = 0,$$

and such that at least one of $c_2,\dots,c_m$ does not vanish.

By substituting the definition of the w_i in this linear combination, we find

$$c_2\big(v_2 - \frac{a_{2n}}{a_{1n}}v_1\big) + \cdots + c_m\big(v_m - \frac{a_{mn}}{a_{1n}}v_1\big) = 0$$

This implies

$$-\left(\frac{a_{2n}}{a_{1n}}c_2 + \cdots + \frac{a_{mn}}{a_{1n}}c_m\right)v_1 + c_2v_2 + \cdots + c_mv_m = 0.$$

Thus $v_1,\dots,v_m$ is linearly dependent.

A.5. **Proof of the Rank+Nullity Theorem.** Choose a basis $\{v_1,\dots,v_r\}$ of the null space $N(T)$. Then choose vectors $v_{r+1},\dots,v_n \in V$ so that $\{v_1,\dots,v_r,v_{r+1},\dots,v_n\}$ is a basis for V. We will show that $\{Tv_{r+1},\dots,Tv_n\}$ is a basis for $R(T)$. The rank+nullity formula then follows because we will have shown that $\dim V = n$, $\dim N(T) = r$, and $\dim R(T) = n - r$.

$Tv_{r+1},\dots,Tv_n$ spans $R(T)$:
if $y \in R(T)$ then there is an $x \in V$ with $y = Tx$. We can write $x = x_1v_1 + \cdots + x_nv_n$. Since $Tv_1 = \cdots = Tv_r = 0$ we have

$$\begin{aligned} y = Tx &= T(x_1v_1 + \cdots + x_nv_n) \\ &= x_{r+1}Tv_{r+1} + \cdots + x_nTv_n \\ &= x_{r+1}w_{r+1} + \cdots + x_nw_n \\ &\in \operatorname{span}(w_{r+1},\dots,w_n). \end{aligned}$$

$Tv_{r+1},\dots,Tv_n$ is linearly independent:
Suppose $c_{r+1}w_{r+1} + \cdots + c_nw_n = 0$ for certain $c_{r+1},\dots,c_n \in \mathbb{F}$. Then

$$T\big(c_{r+1}v_{r+1} + \cdots + c_nv_n\big) = 0,$$

which implies $c_{r+1}v_{r+1} + \cdots + c_nv_n \in N(T)$. It follows that there are numbers $c_1,\dots,c_r$ such that

$$c_{r+1}v_{r+1} + \cdots + c_nv_n = c_1v_1 + \cdots + c_rv_r.$$

Since $\{v_1,\dots,v_n\}$ is a basis for V we conclude that $c_1 = \cdots = c_n = 0$. Hence $Tv_{r+1},\dots,Tv_n$ is linearly independent.

A.6. **Proof of the matrix inverse formula (Theorem 6.17).** Let $A = (a_{ij})$ and $C = (c_{ij})$ where $c_{ij} = (-1)^{i+j}\det\tilde{A}_{ij}$. Consider $B = AC^\top$. By definition of matrix multiplication you have

$$B_{ij} = a_{i1}c_{j1} + \cdots + a_{in}c_{jn}$$

The expression on the right is what we get if we replace the j^{th} row of the matrix A with $(a_{i1} \cdots a_{in})$, i.e. with the i^{th} row of A.

If $i \neq j$ then rows i and j in the resulting determinant are equal, so the determinant vanishes: $B_{ij} = 0$ for all $i \neq j$.

If $i = j$ then replacing row j with row i does nothing, and we just get $\det A$: you get $B_{ii} = \det A$ for all i.

The result is

$$AC^\top = \begin{pmatrix} \det A & 0 & 0 & \cdots & 0 \\ 0 & \det A & 0 & \cdots & 0 \\ 0 & 0 & \det A & \cdots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & & \det A \end{pmatrix} = (\det A)I.$$

A.7. **Proof of Cramer's rule.** Compute $C^\top y$:

$$C^\top y = \begin{pmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ \vdots & & & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} y_1 c_{11} + \cdots + y_n c_{n1} \\ \vdots \\ y_1 c_{1n} + \cdots + y_n c_{nn} \end{pmatrix}$$

Our formula for the inverse of A implies that the solution $x = A^{-1}y$ is given by

$$x = \frac{1}{\det A} \begin{pmatrix} y_1 c_{11} + \cdots + y_n c_{n1} \\ \vdots \\ y_1 c_{1n} + \cdots + y_n c_{nn} \end{pmatrix}$$

Thus the i^{th} component x_1 of the solution is given by

$$x_i = \frac{y_1 c_{1i} + \cdots + y_n c_{ni}}{\det A}.$$

The numerator in this fraction is the determinant that we get by replacing the i^{th} column in $\det A$ with the column vector y.

A.8. **Proof that** $\det AB = \det A \cdot \det B$**.** The short version of the proof uses block matrix notation and goes like this:

$$\begin{aligned} \det(A)\det(B) &= \begin{vmatrix} A & 0 \\ -I & B \end{vmatrix} \qquad \begin{pmatrix} \text{add the first } n \text{ columns } B \text{ times} \\ \text{to the second } n \text{ columns} \end{pmatrix} \\ &= \begin{vmatrix} A & AB \\ -I & 0 \end{vmatrix} \\ &= (-1)^n \begin{vmatrix} -I & A \\ 0 & AB \end{vmatrix} \\ &= (-1)^n \det(-I)\det(AB) \\ &= \det(AB) \end{aligned}$$

If we avoid block matrix notation then we get the following more detailed version of the computation. By definition,

$$\begin{vmatrix} A & 0 \\ -I & B \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & 0 & \cdots & 0 \\ & & \vdots & & & \ddots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 & b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & -1 & & 0 & b_{21} & b_{22} & \cdots & b_{2n} \\ & & \vdots & \vdots & \vdots & & \vdots & \\ 0 & 0 & \cdots & -1 & b_{n1} & b_{n2} & \cdots & b_{nn} \end{vmatrix}.$$

Add the first column b_{11} times to column $(n+1)$, the second column b_{12} times to column $(n+2)$, etc, clearing out the top row in the b section. The result is:

$$\begin{vmatrix} A & 0 \\ -I & B \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} & a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} & a_{21}b_{11} & a_{21}b_{12} & \cdots & a_{21}b_{1n} \\ & & \vdots & & & \ddots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} & a_{n1}b_{11} & a_{n1}b_{12} & \cdots & a_{n1}b_{1n} \\ -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & & 0 & b_{21} & b_{22} & \cdots & b_{2n} \\ & & \vdots & \vdots & \vdots & & \vdots & \\ 0 & 0 & \cdots & -1 & b_{n1} & b_{n2} & \cdots & b_{nn} \end{vmatrix}$$

Next clear out the second row in the b-section. We get

$$\begin{vmatrix} A & 0 \\ -I & B \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} & a_{11}b_{11}+a_{12}b_{21} & \cdots & a_{11}b_{1n}+a_{12}b_{21} \\ a_{21} & a_{22} & \cdots & a_{2n} & a_{21}b_{11}+a_{22}b_{21} & \cdots & a_{21}b_{1n}+a_{22}b_{21} \\ & & \vdots & & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} & a_{n1}b_{11}+a_{n2}b_{21} & \cdots & a_{n1}b_{1n}+a_{n2}b_{21} \\ -1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & -1 & & 0 & 0 & \cdots & 0 \\ & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 & b_{n1} & \cdots & b_{nn} \end{vmatrix}$$

Repeating this for the remaining rows in the b-section of the determinant, we end up with

$$\begin{vmatrix} A & 0 \\ -I & B \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} & a_{11}b_{11}+\cdots+a_{1n}b_{n1} & \cdots & a_{11}b_{1n}+\cdots+a_{1n}b_{nn} \\ a_{21} & a_{22} & \cdots & a_{2n} & a_{21}b_{11}+\cdots+a_{2n}b_{n1} & \cdots & a_{21}b_{1n}+\cdots+a_{2n}b_{nn} \\ & & \vdots & & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} & a_{n1}b_{11}+\cdots+a_{nn}b_{n1} & \cdots & a_{n1}b_{1n}+\cdots+a_{nn}b_{nn} \\ -1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & -1 & & 0 & 0 & \cdots & 0 \\ & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 & 0 & \cdots & 0 \end{vmatrix}$$

i.e. we have

$$\begin{vmatrix} A & 0 \\ -I & B \end{vmatrix} = \begin{vmatrix} A & AB \\ -I & 0 \end{vmatrix}.$$

A.9. **Proof of Theorem 6.22, the determinant-independence test.** Consider the matrix A whose columns are $a_1, \dots, a_n$. Then for any $c_1, \dots, c_n \in \mathbb{F}$ we have

$$c_1 a_1 + \cdots + c_n a_n = A \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

$\det A \neq 0 \implies \{a_1, \dots, a_n\}$ independent. If $\det A \neq 0$ then A is invertible. This implies that $N(A) = \{0\}$. Suppose that there are $c_1, \dots, c_n \in \mathbb{F}$ with $c_1 a_1 + \cdots + c_n a_n = 0$. Then

$c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ satisfies $Ac = 0$, so $c \in N(A)$. Therefore $c = 0$, i.e. $c_1 = \cdots = c_n = 0$. This implies that $\{a_1, \ldots, a_n\}$ is independent.

$\{a_1, \ldots, a_n\}$ independent implies $\det A \neq 0$. Any $c \in N(A)$ satisfies $c_1 a_1 + \cdots + c_n a_n = 0$. Since $\{a_1, \ldots, a_n\}$ is independent this implies $c_1 = \cdots = c_n = 0$, i.e. $c = 0$. We conclude that $N(A) = \{0\}$. It follows that A is injective, and hence also that A is bijective (because of the Bijectivity Theorem). Bijective means the same as invertible, so A is invertible, and thus $\det A \neq 0$.

A.10. **Proof of the independence of eigenvectors.** The proof goes by induction on m.

A.10.1. *The case $m = 1$:* There is only one vector v_1. Since v_1 is an eigenvector we have $v_1 \neq 0$. Therefore v_1 is independent.

A.10.2. *The induction step:* Assume we have already proved the theorem for $m - 1$.

Let $v_1, \ldots, v_m$ be eigenvectors with different eigenvalues, and assume

$$c_1 v_1 + c_2 v_2 + \cdots c_m v_m = 0.$$

Then we will show that $c_1 = \cdots = c_m = 0$.

Multiply the equation with A to get

$$A(c_1 v_1 + c_2 v_2 + \cdots c_m v_m) = 0$$
$$\implies c_1 A v_1 + c_2 A v_2 + \cdots c_m A v_m = 0$$
$$\implies c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \cdots c_m \lambda_m v_m = 0$$

We can also multiply $c_1 v_1 + c_2 v_2 + \cdots c_m v_m = 0$ with λ_m to get

$$c_1 \lambda_m v_1 + c_2 \lambda_m v_2 + \cdots c_m \lambda_m v_m = 0$$

Subtract the last two equations:

$$c_1(\lambda_1 - \lambda_m) v_1 + \cdots + c_{m-1}(\lambda_{m-1} - \lambda_m) v_{m-1} = 0.$$

The vectors $v_1, \ldots, v_{m-1}$ are eigenvectors for A with distinct eigenvalues. The induction hypothesis implies that they are independent, and therefore we find

$$c_1(\lambda_1 - \lambda_m) = 0, \quad c_2(\lambda_2 - \lambda_m) = 0, \quad \ldots \quad c_{m-1}(\lambda_{m-1} - \lambda_m) = 0.$$

The eigenvalues are distinct, so $\lambda_1 - \lambda_m \neq 0$, $\ldots$, $\lambda_{m-1} - \lambda_m \neq 0$. It follows that $c_1 = \cdots = c_{m-1} = 0$.

We still have to show that $c_m = 0$. Since $c_1 = \cdots = c_{m-1} = 0$ we have

$$c_1 v_1 + \cdots + c_m v_m = c_m v_m = 0.$$

The vector v_m is an eigenvector so $v_m \neq 0$. Therefore $c_m v_m = 0$ implies $c_m = 0$.

A.11. **Proof of the spectral theorem.**

Lemma. Let v be an eigenvector of a self-adjoint operator $A : V \to V$ and consider the set $L = \{x \in V \mid x \perp v\}$. Then

- L is a linear subspace of V
- if $\dim V < \infty$ then $\dim L = \dim V - 1$
- L is invariant under A, i.e. for all $x \in L$ one has $Ax \in L$

Proof of the Lemma: an exercise.

To prove the Spectral theorem we use induction on $n = \dim V$.

A is self-adjoint so all its eigenvalues are real. Let v be an eigenvector with eigenvalue $\lambda \in \mathbb{R}$: $Av = \lambda v$ and $v \neq 0$.

We may assume $\|v\| = 1$.

Define $L = \{x \in V \mid x \perp v\}$. The Lemma implies that L is invariant under A. Then $A : L \to L$ is also self-adjoint and, since $\dim L < \dim V$, there is an orthonormal basis $\{v_1, \ldots, v_{n-1}\}$ of L consisting of eigenvectors of A.

Since $v \perp \{v_1, \ldots, v_{n-1}\}$ the set $\{v_1, \ldots, v_{n-1}, v\}$ is an orthonormal, and therefore linearly independent set of vectors in V. Moreover, the set $\{v_1, \ldots, v_{n-1}, v\}$ contains exactly $\dim V$ vectors, so that $\{v_1, \ldots, v_{n-1}, v\}$ is a basis of V.