MATHS CRASH COURSE

German Research Center for Artificial Intelligence

LINEAR ALGEBRA

Julian Liersch
AG Robotik, Universität Bremen

Prof. Dr. Dr. h.c. Frank Kirchner
AG Robotik, Universität Bremen
`https://robotik.dfki-bremen.de/`
`robotik@dfki.de`

October 13, 2023 – Bremen, Deutschland

University of Bremen

German Research Center for Artificial Intelligence

General information

- Our goal is to have a common language and understanding of the relevant content.
- The crash course shall provide you an opportunity to identify knowledge gaps.
- If something is too fast, feel free to pause the video.

University of Bremen

Table of contents

German
Research Center
for Artificial
Intelligence

University
of Bremen

Vector Spaces

German
Research Center
for Artificial
Intelligence

Content

- Vector spaces and subspaces
- Linear (in-)dependence
- Bases and dimensions
- Inner products
- Orthogonality and -normality

German Research Center for Artificial Intelligence

Definition: Vector space

A triple $(V, +, \cdot)$ is called a *vector space over* $\mathbb{F}$ or a *real vector space* for $\mathbb{F} = \mathbb{R}$, if $V \neq \emptyset$, $(\mathbb{F}, +_{\mathbb{F}}, \cdot_{\mathbb{F}})$ is a field and the mappings $+ : V \times V \to V$ as well as $\cdot : \mathbb{F} \times V \to V$ are such that the following properties regarding *vector addition* and *scalar multiplication* hold true:

Vector addition

$\forall u, v, w \in V : (u + v) + w = u + (v + w)$	associativity
$\forall u, v \in V : u + v = v + u$	commutativity
$\exists \mathbf{0}_V \in V : \forall v \in V : \mathbf{0}_V + v = v$	identity element
$\forall v \in V : \exists w \in V : v + w = \mathbf{0}_V$	inverse elements

University of Bremen

German Research Center for Artificial Intelligence

Definition: Vector space

A triple $(V, +, \cdot)$ is called a *vector space over* $\mathbb{F}$ or a *real vector space* for $\mathbb{F} = \mathbb{R}$, if $V \neq \emptyset$, $(\mathbb{F}, +_{\mathbb{F}}, \cdot_{\mathbb{F}})$ is a field and the mappings $+ : V \times V \to V$ as well as $\cdot : \mathbb{F} \times V \to V$ are such that the following properties regarding *vector addition* and *scalar multiplication* hold true:

Scalar multiplication

$\forall \lambda \in \mathbb{F}, u, v \in V : \lambda \cdot (u + v) = \lambda \cdot u + \lambda \cdot v$	distributivity 1
$\forall \lambda, \mu \in \mathbb{F}, v \in V : (\lambda +_{\mathbb{F}} \mu) \cdot v = \lambda \cdot v + \mu \cdot v$	distributivity 2
$\forall v \in V : 1 \cdot v = v$	identity element
$\forall \lambda, \mu \in \mathbb{F}, v \in V : \lambda \cdot_{\mathbb{F}} \mu \cdot v = \lambda \cdot (\mu \cdot v)$	multiplicative compatibility

University of Bremen

Convention

The set V by itself may be called a vector space if $+$ and $\cdot$ are irrelevant or trivial.

Examples of real vector spaces

- Set of n-dimensional vectors $\mathbb{R}^n$ with $+$ and $\cdot$ defined element-wise
- Set of square matrices $\mathbb{R}^{n \times n}$ with $+$ and $\cdot$ defined element-wise
- Set of real continuous functions on closed interval $I \subseteq \mathbb{R}$:
  $C(I) := \{f : I \to \mathbb{R} | f \text{ continuous}\}$ with $+$ and $\cdot$ defined element-wise

Definition: Subspaces

Let $(V, +, \cdot)$ be a vector space over $\mathbb{F}$ and $U \subseteq V$. If $(U, +_{|U \times U}, \cdot_{|\mathbb{F} \times U})$ is also a vector space, it is called a *subspace of* V. This is denoted as $U \leq V$.

German Research Center for Artificial Intelligence

Definition: Span and linear combination

Let $U \subseteq V$ be a subset of a vector space V over $\mathbb{F}$. Then the *span* of U is defined as

$$\mathrm{span}(U) := \left\{ \sum_{i=1}^{k} \lambda_i v_i \,\middle|\, k \in \mathbb{N}, \forall i \in \{1, \dots, k\} : \lambda_i \in \mathbb{F}, v_i \in V \right\}$$

A combination of elements of U as $\sum_{i=1}^{k} \lambda_i v_i$ is called a *linear combination*.

Proposition: A span is a vector space

Let V be a vector space and $U \subseteq V$. Then $\mathrm{span}(U) \leq V$.

University of Bremen

German Research Center for Artificial Intelligence

Definition: Linear (in-)dependence

Let V be a vector space. A finite set of vectors $\{v_1, \ldots, v_k\} \subseteq V$ is called *linear dependent*, if there is a set of scalars $\{\lambda_1, \ldots, \lambda_k\} \subseteq \mathbb{F}$ with $\lambda_i \neq 0$ for some $i \in \{1, \ldots k\}$ and

$$\sum_{i=1}^{k} \lambda_i v_i = 0$$

If there is no such set of scalars, the set of vectors is called *linear independent*.

Examples

- The set $\{\binom{1}{0}, \binom{0}{1}, \binom{1}{1}\} \subseteq \mathbb{R}^2$ is linear dependent as $\binom{1}{0} + \binom{0}{1} - \binom{1}{1} = \binom{0}{0} = 0_{\mathbb{R}^2}$
- The set $\{\binom{1}{0}, \binom{0}{1}\} \subseteq \mathbb{R}^2$ is linear independent

University of Bremen

German Research Center for Artificial Intelligence

Definition: Basis

A linear independent set $B \subseteq V$ with $\mathrm{span}(B) = V$ is called a *basis for V*.

Definition: Unit basis

Let $V = \mathbb{R}^n$ for some $n \in \mathbb{N}$, then the set $B := \{(\delta^i_j)^n_{i=1}\}^n_{j=1}$ yields a basis for V.It is called *the unit basis*.Its elements are usually denoted as $e_j := (\delta^i_j)^n_{i=1} \in \mathbb{R}^n$, where δ^i_j is the Kronecker delta, equal to 1 if $i = j$ and else equal to 0.

Definition: Vector space dimension

Let V be a vector space with basis B. We call the cardinality of B the *dimension of V* and denote it as $\dim(V) := |B|$.

University of Bremen

German Research Center for Artificial Intelligence

Definition: Inner product

Let $(V, +, \cdot)$ be a vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, then a mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ is called an *inner product* or *scalar product*, if the following conditions hold:

Requirements for inner products

$\forall u, v \in V : \langle u, v \rangle = \overline{\langle v, u \rangle}$	(conjugate) symmetry
$\forall \lambda, \mu \in \mathbb{F}, u, v, w \in V : \langle \lambda u + \mu v, w \rangle = \lambda \langle u, w \rangle + \mu \langle v, w \rangle$	linearity in first argument
$\forall v \in V \setminus \{0_V\} : \langle v, v \rangle > 0$	positive definiteness

German Research Center for Artificial Intelligence

Definition: Inner product space

A vector space having a inner product is called an *inner product space*.

Example

The dot product defined as $\langle u, v\rangle := \sum_{i=1}^{n} u_i v_i$ is an inner product for $\mathbb{R}^n$ yielding $(\mathbb{R}^n, +, \cdot, \langle\cdot,\cdot\rangle)$ an inner product space.

University of Bremen

German Research Center for Artificial Intelligence

Definition: Orthogonal basis

Let B be a basis for an inner product space V, then B is called *orthogonal* if for all $u \neq v \in B$ the equality $\langle u, v \rangle = 0$ holds true.

Definition: Induced norm

Let V be an inner product space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. The mapping $||\cdot|| : V \to \mathbb{R}_0^+$ defined as $||v|| := \sqrt{\langle v, v \rangle}$ is called the *induced norm for* V.

Definition: Orthonormal basis

Let B be a orthogonal basis for an inner product space V. If for all $v \in B$ the equality $||v|| = 1$ holds, B is called an *orthonormal* basis.

University of Bremen

Linear mappings

Content

- Definition of linear mappings
- Range and kernel
- Rank-nullity theorem

Definition: Linear mapping

Let V, W be vector spaces over a field $\mathbb{F}$ and $L : V \to W$ be a mapping. L is called a *linear mapping* or *linear*, if

$$\forall \lambda \in \mathbb{F}, v, w \in V : L(\lambda \cdot v + w) = \lambda \cdot L(v) + L(w)$$

Proposition: Linear mappings are uniquely defined by their definition on a basis

Let V, W be vector spaces and B be a basis for V. If $L_1, L_2 : V \to W$ are linear with $L_1(v) = L_2(v)$ for all $v \in B$, then $L_1(v) = L_2(v)$ for all $v \in V$.

German Research Center for Artificial Intelligence

Definition: Range

Let $L : V \to W$ be a linear mapping between vector spaces V, W. Then the *image* or *range of* L is defined as

$$\mathrm{Im}(L) := L[V] = \{w \in W | \exists v \in V : L(v) = w\}$$

Definition: Kernel

Let $L : V \to W$ be as above. Then the *kernel* or *null space of* L is defined as

$$\ker(L) := L^{-1}[\{\mathbf{0}_W\}] = \{v \in V | L(v) = \mathbf{0}_W\}$$

Proposition: The kernel and range of linear mappings are subspaces

For vector spaces V, W and $L : V \to W$ linear it is $\mathrm{Im}(L) \leq W$ and $\ker L \leq V$.

German Research Center for Artificial Intelligence

Reminder

A mapping $f : X \to Y$ is called *injective*, if for each $y \in f[X] = \{y \in Y | \exists x \in X : f(x) = y\}$ there is **exactly one** $x \in X$ with $f(x) = y$. Short:

$$f(x_1) = f(x_2) \Rightarrow x_1 = x_2$$

Proposition: Relation to kernel

A linear mapping $L : V \to W$ is injective iff (short for if and only if) its kernel is trivial, meaning $\ker(L) = \{\mathbf{0}_V\}$.

German Research Center for Artificial Intelligence

Theorem: Rank-nullity theorem

Let $L : V \to W$ be a linear mapping between two vector spaces V and W. Then the following identity holds:

$$\dim(\mathrm{Im}(L)) + \dim(\ker(L)) = \dim(V)$$

Definition: Rank and nullity

The dimension of the image of L is also called *rank* and the dimension of the kernel *nullity*.

Corollary

Injective linear mappings $L : V \to W$ are surjective (and thus bijective), if $\dim(V) = \dim(W)$.

Matrices

German Research Center for Artificial Intelligence

Content

- Identification of linear mappings with matrices
- Matrix operations
- Determinant, inverse matrices and the general linear group
- Eigenvector decomposition

University of Bremen

German Research Center for Artificial Intelligence

Observation

Let V, W be finite dimensional real vector spaces with basis $B_V := \{v_1, \dots v_n\}$, $B_W := \{w_1, \dots, w_m\}$ respectively and $L : V \to W$ be linear.

- For all $v_i \in B_V$ there are $\mu_{ji} \in \mathbb{F}$ such that $L(v_i) = \sum_{j=1}^{m} \mu_{ji} w_j$.

Matrix representation of linear mappings

As L is fully described by its operation on B_V and each $L(v_i)$ can be written as a unique [!] linear combination of B_W, L is (for given B_V and B_W) uniquely defined by the μ_{ji} as above.These are usually arranged in a matrix $M_{B_W}^{B_V}(L) \in \mathbb{R}^{m \times n}$ which can be identified with the linear mapping:

$$L \simeq M_{B_W}^{B_V}(L) := \begin{pmatrix} \mu_{11} & \cdots & \mu_{1n} \\ \vdots & \ddots & \vdots \\ \mu_{m1} & \cdots & \mu_{mn} \end{pmatrix}$$

University of Bremen

German Research Center for Artificial Intelligence

Convention for further slides

For simplicity, we will focus on B_V and B_W being unit bases and drop explicit mentioning of the used bases.

Definition: Matrix-vector multiplication

Let V, W be finite dimensional real vector spaces and $L : V \to W$ be linear with matrix representation $M(L) \in \mathbb{R}^{m \times n}$. The *matrix-vector multiplication* $\cdot : \mathbb{R}^{m \times n} \times \mathbb{R}^n \to \mathbb{R}^m$ is defined via the evaluation of L at v as

$$M(L) \cdot v := L(v) = \left(\sum_{i=1}^{n} \mu_{ji} v_i \right)_{j=1}^{m}$$

German Research Center for Artificial Intelligence

Definition: Matrix-matrix multiplication

Let V_1, V_2, V_3 be finite dimensional real vector spaces with dimensions $n_1, n_2, n_3 \in \mathbb{N}$ respectively and $L_1 : V_1 \to V_2, L_2 : V_2 \to V_3$ be linear mappings. For $M(L_1) \in \mathbb{R}^{n_2 \times n_1}, M(L_2) \in \mathbb{R}^{n_3 \times n_2}$ we can define the *matrix-matrix multiplication* $\cdot : \mathbb{R}^{n_3 \times n_2} \times \mathbb{R}^{n_2 \times n_1} \to \mathbb{R}^{n_3 \times n_1}$ via the composition of L_2 and L_1:

$$M(L_2) \cdot M(L_1) := M(L_2 \circ L_1)$$

Matrix-matrix multiplication mnemonic

The product $A \cdot B$ can be calculated as dot products of the rows of A and columns of B:

$$\mathbb{R}^{n_2 \times n_3} \ni B = \begin{pmatrix} b_{11} & \cdots & b_{1n_3} \\ \vdots & \ddots & \vdots \\ b_{n_2 1} & \cdots & b_{n_2 n_3} \end{pmatrix}$$

$$\mathbb{R}^{n_1 \times n_2} \ni A = \begin{pmatrix} a_{11} & \cdots & a_{1n_2} \\ \vdots & \ddots & \vdots \\ a_{n_1 1} & \cdots & a_{n_1 n_2} \end{pmatrix} \quad \begin{pmatrix} \sum_{i=1}^{n_2} a_{1i} b_{i1} & \dots & \sum_{i=1}^{n_2} a_{1i} b_{in_3} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n_2} a_{n_1 i} b_{i1} & \dots & \sum_{i=1}^{n_2} a_{n_1 i} b_{in_3} \end{pmatrix} = A \cdot B \in \mathbb{R}^{n_1 \times n_3}$$

Definition: Transpose

Let $A := (a_{ij})_{i,j=1}^{i=n,j=m} \in \mathbb{R}^{n\times m}$ be a matrix, then $A := (a_{ji})_{i,j=1}^{i=n,j=m} \in \mathbb{R}^{m\times n}$ is called the *transpose of* A.

Definition: Symmetric matrices

If for $A \in \mathbb{R}^{n\times n}$ the equality $A^T = A$ holds, A is called *symmetric*.

Definition: Orthogonal matrices

Let V be an n-dimensional inner product space and $A \in \mathbb{R}^{n\times n}$. Then A is called *orthogonal* if $A^{-1} = A^T$.

German Research Center for Artificial Intelligence

Definition: Square matrix

A matrix $A \in \mathbb{R}^{n \times m}$ is called *square matrix* if $n = m$.

Definition: Unit matrix

For $n \in \mathbb{N}$, the matrix $\mathbb{I}_n := \left(\delta_i^j\right)_{i,j=1}^{n}$ is called the *identity matrix of rank n*.

University of Bremen

German Research Center for Artificial Intelligence

Definition: Inverse matrix

Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. If there is a matrix $B \in \mathbb{R}^{n \times n}$ with $A \cdot B = B \cdot A = \mathbb{I}_n$, A is called *invertible* and $B =: A^{-1}$ is called the *inverse of A*.

Proposition: Existence of inverse

A square matrix $A \in \mathbb{R}^{n \times n}$ is invertible if and only if and only if it has full rank, meaning

$$\dim(\mathrm{Im}(v \mapsto A \cdot v)) = n$$

Equivalently it is invertible if its nullity is 0, meaning

$$\dim(\ker(v \mapsto A \cdot v)) = 0$$

This also means the mapping $v \mapsto A \cdot v$ is bijective for invertible matrices.

German Research Center for Artificial Intelligence

Definition: Determinant

For a square matrix $A \in \mathbb{R}^{n \times n}$, the *determinant of A* is defined via the Leibniz-formula as

$$\det(A) := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i}$$

Here S_n is the group of permutations of $\{1, \ldots, n\}$ and $\operatorname{sgn}(\sigma)$ is the sign of σ, which equals $+1$ if the number of inversions the permutation composes of is even and -1 if it is odd.

Proposition: Link to invertibility

A square matrix $A \in \mathbb{R}^{n \times n}$ is invertible if and only if its determinant is non-zero. Equivalently, if $\det(A) = 0$, then $\dim(\ker(A)) > 0$.

University of Bremen

German Research Center for Artificial Intelligence

Definition: Group

Let $G \neq \emptyset$ and $* : G \times G \to G$. Then $(G, *)$ is called a *group*, if:

$$
\begin{aligned}
&\forall a, b, c : (a * b) * c = a * (b * c) && \text{associativity} \\
&\exists e \in G : \forall a \in G : e * a = a * e = a && \text{existence of neutral element} \\
&\forall a \in G : \exists b \in G : a * b = e \wedge b * a = e && \text{existence of inverse elements}
\end{aligned}
$$

Proposition and definition: General linear group

The set of invertible matrices $GL_n := \{A \in \mathbb{R}^{n \times n} | \det(A) \neq 0\}$ together with the matrix-matrix multiplication forms a group. It is called the *general linear group* and its neutral element is $\mathbb{I}_n$.

University of Bremen

German Research Center for Artificial Intelligence

Definition: Diagonalizable

A square matrix $A \in \mathbb{R}^{n \times n}$ is called *diagonalizable*, if there is an invertible matrix $P \in GL_n$ such that $D := P^{-1}AP$ is a diagonal matrix. This means the elements of D are only non-zero on the main diagonal as such:

$$D = \begin{pmatrix} d_{11} & 0 & \cdots & 0 & 0 \\ 0 & d_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d_{n-1,n-1} & 0 \\ 0 & 0 & \cdots & 0 & d_{nn} \end{pmatrix}$$

University of Bremen

German Research Center for Artificial Intelligence

Definition: Eigenvectors and -values

Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. If $A \cdot v = \lambda v$ for $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^n \setminus \{0\}$ holds, λ is called an *eigenvalue for A* and v is called an *eigenvector with eigenvalue λ*.

Definition: Eigenvalue-problem

To find eigenvectors and values for a square matrix, one can find non-trivial solutions ($v \neq 0$) of the following linear equation system, the *eigenvalue problem*:

$$(A - \lambda \mathbb{I}_n) \cdot v = 0$$

Non-trivial solutions exist for λ such that $\dim(\ker(A - \lambda \mathbb{I}_n)) > 0$,which is the case iff $p_A(\lambda) := \det(A - \lambda \mathbb{I}_n) = 0$. The mapping $p_A : \mathbb{R} \to \mathbb{R}$ is called the *characteristic polynomial of A*.

German Research Center for Artificial Intelligence

Theorem: Eigenvector decomposition

If for a matrix A a set of linearly independent eigenvectors $\{v_1, \ldots, v_n\} \subseteq \mathbb{R}^n$ with corresponding eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ exists, then the matrix is diagonalizable using $P := \left(v_1 \cdots v_n \right) \in \mathbb{R}^{n \times n}$ and $\Lambda := \left(\delta_i^j \lambda_i\right)_{i,j=1}^n \in \mathbb{R}^{n \times n}$ as

$$P^{-1}AP = \Lambda$$

Notes

- To prove the identity, observe that $Av_i = \lambda_i v_i$ holds for all $i \in \{1, \ldots n\}$ yielding $AP = P\Lambda$.
- Outlook: For non-square matrices the *singular value decomposition* (short *SVD*) yields a similar identity.

German Research Center for Artificial Intelligence

Definition: Positive (semi-)definite

Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then A is called *positive definite* if

$$\forall v \in \mathbb{R}^n \setminus \{\mathbf{0}_{\mathbb{R}^n}\} : \quad v^T A v > 0$$

If only $v^T A v \geq 0$ holds, A is called *positive semi-definite* instead.

Remarks

- If A is positive semi-definite, it is diagonalizable and all its eigenvalues are non-negative.
- If A is positive definite, it is diagonalizable and all its eigenvalues are positive.

University of Bremen

Matrix identities

German Research Center for Artificial Intelligence

Content

- Transpose matrices
- Inverse matrices
- Determinant of matrices

University of Bremen

German Research Center for Artificial Intelligence

Identities regarding transpose

Let $A, B \in \mathbb{R}^{n \times m}$ and $c \in \mathbb{R}$, then

$$
\begin{aligned}
(A+B)^T &= A^T + B^T \\
(AB)^T &= B^T A^T \\
(cA)^T &= cA^T \\
p_{A^T}(\lambda) &= p_A(\lambda)
\end{aligned}
$$

If A is additionally invertible, we have

$$(A^{-1})^T = (A^T)^{-1}$$

thus conventionally we write $A^{-T} := (A^{-1})^T$.

University of Bremen

Identities regarding inverse

Let $A, B \in GL_n(\mathbb{R})$ and $c \in \mathbb{R} \setminus \{0\}$, then

$$
\begin{aligned}
A + B &\text{ is not necessarily invertible! (e.g. } B = -A) \\
(AB)^{-1} &= B^{-1}A^{-1} \\
(cA)^{-1} &= c^{-1}A^{-1} \\
\lambda \neq 0 \text{ eigenvalue of } A &\Rightarrow \lambda^{-1} \text{ eigenvalue of } A^{-1}
\end{aligned}
$$

German Research Center for Artificial Intelligence

Identities regarding determinant

Let $A, B \in \mathbb{R}^{n\times n}$ and $c \in \mathbb{R} \setminus \{0\}$, as well as A invertible if needed, then

$$
\begin{aligned}
\det(cA) &= c^n \det(A) \\
\det(AB) &= \det(A)\det(B) \\
\det(A^T) &= \det(A) \\
\det(A^{-1}) &= \det(A)^{-1} \\
\det(A) &= \prod_{i=1}^{n} \lambda_i
\end{aligned}
$$

where λ_i are the eigenvalues of A, if needed repeated to occur as often as their algebraic multiplicity.

German Research Center for Artificial Intelligence

We recapped

- definitions regarding vector spaces,
- definitions regarding linear mappings,
  - as well as the rank-nullity theorem
- and content on matrices as representations of linear mappings:
  - matrix operations,
  - the general linear group
  - and eigenvector decomposition.
- Additionally, some identities regarding to matrices.

University of Bremen

Thank You!

Feel free to ask questions in the forums!