Geometrisches Rechnen
(WS 2013/14)

Martin Held

FB Computerwissenschaften
Universität Salzburg
A-5020 Salzburg, Austria
held@cosy.sbg.ac.at

August 20, 2014

UNIVERSITÄT SALZBURG
Computational Geometry and Applications Lab
Personalia

Instructor: M. Held.

Email: held@cosy.sbg.ac.at.

Base-URL: http://www.cosy.sbg.ac.at/~held.

Office: Universität Salzburg, Computerwissenschaften, Rm. 1.20, Jakob-Haringer Str. 2, 5020 Salzburg-Itzling.

Phone number (office): (0662) 8044-6304.

Phone number (secr.): (0662) 8044-6328.
Formalia

URL of course: .../teaching/geom_rechnen/geom_rechnen.html.

Lecture times:  Wednesday 11\textsuperscript{00} – 13\textsuperscript{40}.

Venue:  T02, Computerwissenschaften, Jakob-Haringer Str. 2.

Note  This lecture is graded according to continuous-assessment mode!
Electronic Slides and Online Material

In addition to these slides, you are encouraged to consult the WWW home-page of this lecture:

http://www.cosy.sbg.ac.at/~held/teaching/geom_rechnen/geom_rechnen.html.

In particular, this WWW page contains links to online manuals, slides, and code.
A Few Words of Warning

I hope that these slides will serve as a practice-minded introduction to the mathematics of geometric computing. I would like to warn you explicitly not to regard these slides as the sole source of information on the topics of my course. It may and will happen that I’ll use the lecture for talking about subtle details that need not be covered in these slides! In particular, the slides won’t contain all sample calculations, proofs of theorems, demonstrations of algorithms, or solutions to problems posed during my lecture. That is, by making these slides available to you I do not intend to encourage you to attend the lecture on an irregular basis.
Acknowledgments

These slides are a revised and extended version of notes and slides originally prepared for my graphics courses. Those graphics slides were partially based on write-ups of former students, and I would like to express my thankfulness for their help with those graphics slides. This revision and extension was carried out by myself, and I am responsible for all errors.

Salzburg, September 2013

Martin Held
Legal Fine Print and Disclaimer

To the best of our knowledge, these slides do not violate or infringe upon somebody else’s copyrights. If copyrighted material appears in these slides then it was considered to be available in a non-profit manner and as an educational tool for teaching at an academic institution, within the limits of the “fair use” policy. For copyrighted material we strive to give references to the copyright holders (if known). Of course, any trademarks mentioned in these slides are properties of their respective owners.

Please note that these slides are copyrighted. The copyright holder(s) grant you the right to download and print it for your personal use. Any other use, including non-profit instructional use and re-distribution in electronic or printed form of significant portions of it, beyond the limits of “fair use”, requires the explicit permission of the copyright holder(s). All rights reserved.

These slides are made available without warrant of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. In no event shall the copyright holder(s) and/or their respective employers be liable for any special, indirect or consequential damages or any damages whatsoever resulting from loss of use, data or profits, arising out of or in connection with the use of information provided in these slides.
# Table of Content

1. Notation
2. Algebraic Concepts
3. Basic Linear Algebra
4. Transformations
5. Geometric Objects
6. Basic Concepts of Topology
7. Floating-Point Arithmetic and Numerical Mathematics
1 Notation
Notation

- We use letters in italics to denote scalar values: \(a, b, c, t\). Greek letters, such as \(\alpha\) or \(\lambda\), typically also denote scalars.

- Points (in 2D or 3D) are denoted by capital letters written in italics: \(A\). If no misunderstanding is possible, we may also resort to using lower-case italics letters to denote points.

- We user Euler Fraktur for denoting vectors: \(\mathbf{v}\). E.g., the vector pointing from the coordinate origin to the point \(A\) is given by \(\mathbf{a}\).

- The coordinates of a vector are denoted by using indices (or numbers): e.g., \(\mathbf{a} = (a_x, a_y, a_z)\), or \(\mathbf{a} = (a_1, a_2, \ldots, a_n)\).

- The term \(\mathbf{ab}\) denotes the vector from the point \(A\) to the point \(B\). That is, \(\mathbf{ab} = \mathbf{b} - \mathbf{a}\).

- The dot product of two vectors \(\mathbf{a}\) and \(\mathbf{b}\) is denoted by \(\langle \mathbf{a}, \mathbf{b} \rangle\).

- The vector cross-product is denoted by a cross: \(\mathbf{a} \times \mathbf{b}\).
Notation

- The length of a vector $\alpha$ is denoted by $\|\alpha\|$.
- The straight-line segment between the points $A$ and $B$ is denoted by $\overline{AB}$.
- Bold capital letters, such as $\mathbf{M}$, are used for matrices.
- The set $\{1, 2, 3, \ldots\}$ of natural numbers is denoted by $\mathbb{N}$, while $\mathbb{Z}$ denotes the integers (positive and negative), $\mathbb{Q}$ stands for the rationals, and $\mathbb{R}$ for the reals. The non-negative reals are denoted by $\mathbb{R}_0^+$, and the positive reals by $\mathbb{R}^+$.
- Open or closed intervals $I \subset \mathbb{R}$ are denoted using square brackets: e.g., $l_1 = [a_1, b_1]$ or $l_2 = [a_2, b_2[$, where the right-hand “[” indicates that the value $b_2$ is not included in $l_2$.
- It goes without saying that we may find it convenient to deviate from these rules here and there...
2 Algebraic Concepts

- Algebraic Structures
- Real Numbers and Vector Space $\mathbb{R}^n$
- Complex Numbers $\mathbb{C}$
- Quaternions $\mathbb{H}$
Algebraic Structures

- An algebraic structure consists of a set together with one or more operations on the set which satisfy certain axioms.
- Well-known examples: \( \mathbb{N} \) with the standard addition “+”.
- Algebraic structures get their names based on the type of operations and axioms supported.
- Well-known structures include group, ring, field, and vector space. (Many more algebraic structures are studied in abstract algebra!)
Definition 1 (Group, Dt.: Gruppe)

A set $G$ together with an operation ("addition") $+: G \times G \rightarrow G$ defines a group if the following conditions hold:

1. **Associativity:** $\forall a, b, c \in G \quad (a + b) + c = a + (b + c)$.
2. **Neutral element:** There exists an element $0 \in G$ ("zero" element) such that $\forall a \in G \quad 0 + a = a = a + 0$.
3. **Inverse element:** For all $a \in G$ there exists an inverse $b \in G$, satisfying $a + b = 0 = b + a$. (Conventionally, such a $b$ is denoted by $-a$.)

- Needless to say, the binary operation "+" need not bear any relation to the standard addition of numbers!
- It should be noted that $a + b = b + a$ is not required. That is, commutativity need not hold! A group $(G, +)$ in which commutativity holds is called *Abelian*.
- Sample (Abelian) groups: the integers $\mathbb{Z}$ under addition, non-zero rational numbers $\mathbb{Q} \setminus \{0\}$ under multiplication.
- Not a group: The integers under multiplication.
Finite Group

- A group \((G, +)\) is finite if \(G\) has a finite number of elements.
- The number of elements of a finite group is called the \textit{order} of the group.
- A finite group is completely described by its \textit{multiplication table} (aka \textit{Cayley table}).
- Multiplication tables for groups with two and three elements:

\begin{align*}
+ & | 0 & a \\
0 & | 0 & a \\
a & | a & 0
\end{align*}

\begin{align*}
+ & | 0 & a & b \\
0 & | 0 & a & b \\
a & | a & b & 0 \\
b & | b & 0 & a
\end{align*}
Finite Group: Dihedral Group $D_4$

- The **dihedral group** $D_4$ is formed by the rotations and reflections of a square:
  - $id$, $r_1$ (rotation by $90^\circ$), $r_2$ (rotation by $180^\circ$), $r_3$ (rotation by $270^\circ$);
  - $f_v$ (vertical flip), $f_h$ (horizontal flip), $f_d$ (diagonal flip), $f_c$ (counter-diagonal flip).

[Image credit: Wikipedia.]
Finite Group: Dihedral Group $D_4$

- Multiplication table of $D_4$:

<table>
<thead>
<tr>
<th></th>
<th>$id$</th>
<th>$r_1$</th>
<th>$r_2$</th>
<th>$r_3$</th>
<th>$f_v$</th>
<th>$f_h$</th>
<th>$f_d$</th>
<th>$f_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$id$</td>
<td>$id$</td>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r_3$</td>
<td>$f_v$</td>
<td>$f_h$</td>
<td>$f_d$</td>
<td>$f_c$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r_3$</td>
<td>$id$</td>
<td>$f_c$</td>
<td>$f_d$</td>
<td>$f_v$</td>
<td>$f_h$</td>
</tr>
<tr>
<td>$r_2$</td>
<td>$r_2$</td>
<td>$r_3$</td>
<td>$id$</td>
<td>$r_1$</td>
<td>$f_h$</td>
<td>$f_v$</td>
<td>$f_c$</td>
<td>$f_d$</td>
</tr>
<tr>
<td>$r_3$</td>
<td>$r_3$</td>
<td>$id$</td>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$f_d$</td>
<td>$f_c$</td>
<td>$f_h$</td>
<td>$f_v$</td>
</tr>
<tr>
<td>$f_v$</td>
<td>$f_v$</td>
<td>$f_d$</td>
<td>$f_h$</td>
<td>$f_c$</td>
<td>$id$</td>
<td>$r_2$</td>
<td>$r_1$</td>
<td>$r_3$</td>
</tr>
<tr>
<td>$f_h$</td>
<td>$f_h$</td>
<td>$f_c$</td>
<td>$f_v$</td>
<td>$f_d$</td>
<td>$r_2$</td>
<td>$id$</td>
<td>$r_3$</td>
<td>$r_1$</td>
</tr>
<tr>
<td>$f_d$</td>
<td>$f_d$</td>
<td>$f_h$</td>
<td>$f_c$</td>
<td>$f_v$</td>
<td>$r_3$</td>
<td>$r_1$</td>
<td>$id$</td>
<td>$r_2$</td>
</tr>
<tr>
<td>$f_c$</td>
<td>$f_c$</td>
<td>$f_v$</td>
<td>$f_d$</td>
<td>$f_h$</td>
<td>$r_1$</td>
<td>$r_3$</td>
<td>$r_2$</td>
<td>$id$</td>
</tr>
</tbody>
</table>

- Note: $D_4$ is not commutative.
**Definition 2 (Ring)**

A set $R$ which possesses an addition $+ : R \times R \rightarrow R$ and a multiplication $\cdot : R \times R \rightarrow R$ defines a (unit) ring if the following conditions hold:

1. $(R, +)$ is an Abelian group:
2. **Associativity**: $\forall a, b, c \in R \quad (a \cdot b) \cdot c = a \cdot (b \cdot c)$.
3. **Distributivity**:
   - $\forall a, b, c \in R \quad a \cdot (b + c) = (a \cdot b) + (a \cdot c)$;
   - $\forall a, b, c \in R \quad (a + b) \cdot c = (a \cdot c) + (b \cdot c)$.
4. **Neutral element**: There exists an element $1 \in R$ ("one" element) such that $\forall a \in R \quad 1 \cdot a = a = a \cdot 1$.

- Note that $a \cdot b = b \cdot a$ is not required. If commutativity holds then $(R, +, \cdot)$ forms a **commutative ring**.
- Sample ring: The set of all continuous real-valued functions defined over an interval $[\alpha, \beta] \subset \mathbb{R}$, with addition and multiplication of functions as operations, forms a ring.
Definition 3 (Polynomial, Dt.: Polynom)

Consider a commutative ring \((R, +, \cdot)\) and a symbol \(x\) that is not contained in \(R\). A (univariate) polynomial over \(R\) with indeterminate \(x\) ("polynomial in \(x\)"") is a term of the form

\[
a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,
\]

with coefficients \(a_0, \ldots, a_n \in R\), for some \(n \in \mathbb{N}_0\).

The highest occurring power of \(x\) — that is, the maximum \(n\) such that \(a_n \neq 0\) — is called the degree (Dt.: Grad) of the polynomial; its coefficient \(a_n\) is called the leading coefficient.

- Two polynomials are equal if and only if the sequences of their coefficients are equal.
- Polynomials of degree
  0 are called constant polynomials,
  1 are called linear polynomials,
  2 are called quadratic polynomials,
  3 are called cubic polynomials,
  4 are called quartic polynomials,
  5 are called quintic polynomials.
Ring of Polynomials

- We define the addition of polynomials based on the pairwise addition of corresponding coefficients. The multiplication of polynomials is based on the multiplication within $R$, distributivity, and the rules:

\[ ax = xa \quad \text{and} \quad x^m x^k = x^{m+k} \]

for all $a \in R$ and $m, k \in \mathbb{N}$.

Lemma 4

The set of all polynomials in $x$ with the multiplication as defined above and with coefficients in the ring $(R, +, \cdot)$ forms itself a ring, the ring of polynomials over $R$, commonly denoted by $R[x]$.

- Multivariate polynomials are defined by a natural extension of this scheme, using several distinct indeterminates. E.g.,

\[ a_{2,3} x^2 y^3 + a_{0,1} y + a_{2,1} x^2 y + a_{1,0} x + a_{0,0} = (a_{2,3} x^2) y^3 + (a_{0,1} + a_{1,1} x^2) y + (a_{1,0} x + a_{0,0}) \]

is an element of $R[x, y] := (R[x])[y]$. 
**Definition 5 (Polynomial function, Dt.: Polynomfunktion)**

With every polynomial $p$ in $R[x]$ one can associate a polynomial function $f_p : R \rightarrow R$ as follows: For an argument $r \in R$ we obtain $f_p(r)$ by replacing all occurrences of $x$ in $p$ by $r$, and by evaluating the resulting term by applying the standard arithmetic rules for $+$ and $\cdot$ in $R$.

- As usual, two (polynomial) functions are identical if their values are identical for all arguments in $R$.
- Note: Two different polynomials may result in the same polynomial function! (E.g., over finite fields.)

**Definition 6 (Root, Dt.: Nullstelle)**

The polynomial function $f_p$ has a root (or zero) $r \in R$ if $f_p(r) = 0$. 
### Definition 7 (Field, Dt.: Körper)

A set $F$ which possesses an addition $+: F \times F \rightarrow F$ and a multiplication $\cdot: F \times F \rightarrow F$ defines a field if the following conditions hold:

1. **Associativity:** $(a + b) + c = a + (b + c)$ $\forall a, b, c \in F$.
2. **Associativity:** $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ $\forall a, b, c \in F$.
3. **Commutativity:** $a + b = b + a$ $\forall a, b \in F$.
4. **Commutativity:** $a \cdot b = b \cdot a$ $\forall a, b \in F$.
5. **Distributivity:** $a \cdot (b + c) = a \cdot b + a \cdot c$ $\forall a, b, c \in F$.
6. **Neutral element:** There exists an element $0 \in F$ such that $0 + a = a$ $\forall a \in F$.
7. **Neutral element:** There exists an element $1 \in F$ such that $1 \cdot a = a$ $\forall a \in F$.
8. For all $a \in F$ there exists an additive inverse $b \in F$, satisfying $a + b = 0$.
9. For all $a \in F \setminus \{0\}$ there exists a multiplicative inverse $b \in F$, satisfying $a \cdot b = 1$.
10. $0 \neq 1$. 

© M. Held (Univ. Salzburg)
Field

- **Note**: Every field has at least two distinct elements.
- **Sample fields**: \(\mathbb{Q}, \mathbb{R}\) and complex numbers \(\mathbb{C}\) with the standard addition and multiplication, and \(\mathbb{Z}_p\) with addition and multiplication modulo \(p\) (for prime numbers \(p\)).
- **Note** that the multiplication sign often is dropped if the meaning is clear within a specific context: \(ab\) rather than \(a \cdot b\).
Vector Space

Definition 8 (Vector space, Dt.: Vektorraum)

A set $V$ together with an addition $+: V \times V \rightarrow V$ and a scalar multiplication $\cdot: F \times V \rightarrow V$ defines a vector space over a field $F$ if the following eight conditions hold:

1. $a + (b + c) = (a + b) + c \quad \forall a, b, c \in V$.
2. $a + b = b + a \quad \forall a, b \in V$.
3. There exists an element $0 \in V$ such that $a + 0 = a \quad \forall a \in V$.
4. $\forall a \in V$ there exists $b \in V$ such that $a + b = 0$.
5. $\lambda (a + b) = \lambda \cdot a + \lambda \cdot b \quad \forall \lambda \in F, \forall a, b \in V$.
6. $(\lambda + \mu) \cdot a = \lambda \cdot a + \mu \cdot a \quad \forall \lambda, \mu \in F, \forall a \in V$.
7. $\lambda \cdot (\mu \cdot a) = (\lambda \mu) \cdot a \quad \forall \lambda, \mu \in F, \forall a \in V$.
8. $1 \cdot a = a \quad \forall a \in V$.

The multiplication sign is often dropped if the meaning is clear within a specific context: $\lambda a$ rather than $\lambda \cdot a$. 
**Definition 9 (Cartesian product, Dt.: Mengenprodukt, Kreuzprodukt)**

For a field $F$ and $n \in \mathbb{N}$, we define

$$ F^n := \underbrace{F \times F \times \cdots \times F}_{n \text{ times}} := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : x_1, \ldots, x_n \in F \right\}. $$

- Well-known sample: $\mathbb{R}^n$, i.e., $F := \mathbb{R}$. You may find it convenient to “visualize” $F^n$ as $\mathbb{R}^n$.
- It is trivial to generalize this definition to $F_1 \times F_2 \times \cdots \times F_n$.
- For $a, b \in F^n$, we define the addition $+$ as follows:

$$ a + b = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} := \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix}. $$
Vector Space $F^n$

- For $a \in F^n$, we define the multiplication by a scalar $\lambda \in F$ as follows:

$$\lambda a = \lambda \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} := \begin{pmatrix} \lambda a_1 \\ \vdots \\ \lambda a_n \end{pmatrix}.$$

- Furthermore, we use

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

as zero vector 0 of $F^n$, and

$$\begin{pmatrix} -a_1 \\ \vdots \\ -a_n \end{pmatrix}$$

as the additive inverse for

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

**Theorem 10**

Let $F$ be a field. Then $F^n$ with addition and scalar multiplication as defined above constitutes a vector space.
"Exotic" Vector Spaces: Functions, Polynomials, Sequences

**Lemma 11**
The set of all real-valued functions \( f : \mathbb{R} \rightarrow \mathbb{R} \) forms a vector space over \( \mathbb{R} \).

**Lemma 12**
The set of all polynomials \( a_n x^n + \ldots + a_1 x + a_0 \) of degree at most \( n \) (for some fixed \( n \in \mathbb{N} \)), with real coefficients \( a_0, a_1, \ldots, a_n \in \mathbb{R} \), forms a vector space over \( \mathbb{R} \).

**Lemma 13**
The set of all infinite sequences \( a = (a_1, a_2, a_3, \ldots) \) of real numbers forms a vector space over \( \mathbb{R} \). (Formally, a real sequence is a function that maps \( \mathbb{N} \) to \( \mathbb{R} \).)

- **Caveats:**
  - Subsets of functions characterized by an additional property — e.g., positive, not continuous — need not form a vector space.
  - The set of all polynomials of degree exactly \( n \) does not form a vector space!
  - Subsets of sequences characterized by an additional property — e.g., divergent sequences, monotonic sequences — need not form a vector space!
Definition 14 (Subspace, Dt.: Teilraum, Unterraum)

A subset $S$ of a vector space $V$ is called a *subspace* of $V$ if

1. the zero vector belongs to $S$; (that is, $0 \in S$);
2. $\forall a, b \in S \quad a + b \in S$ (S is said to be closed under vector addition);
3. $\forall a \in S \forall \lambda \in F \quad \lambda a \in S$ (S is said to be closed under scalar multiplication).

Lemma 15

The set of all continuous (real-valued) functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and the set of all linear functions form subspaces of the vector space of all (real-valued) functions.
**Definition 16 (Linear combination, Dt.: Linearkombination)**

Let $V$ be a vector space over $F$, and $a_1, \ldots, a_k \in V$ and $\lambda_1, \ldots, \lambda_k \in F$, for some $k \in \mathbb{N}$. The vector

$$a = \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_k a_k$$

is called a *linear combination* of the vectors $a_1, \ldots, a_k$.

**Definition 17 (Linear hull, Dt.: lineare Hülle)**

For $A \subseteq V$,

$$[A] := \{ \lambda_1 a_1 + \cdots + \lambda_k a_k : k \in \mathbb{N}, a_1, \ldots, a_k \in A, \lambda_1, \ldots, \lambda_k \in F \}$$

forms the *linear hull* of $A$.

- **Note:** Any linear combination is formed by a finite number of vectors, even if we are allowed to pick those vectors from an infinite set!

**Lemma 18**

For $A \subseteq V$, the linear hull $[A]$ forms a subspace of $V$. 
Linear Independence

Definition 19 (Linear independence, Dt.: lineare Unabhängigkeit)

The vectors $a_1, a_2, \ldots, a_k$ of a vector space $V$ over $F$ are *linearly dependent* if there exist scalars $\lambda_1, \ldots, \lambda_k \in F$, not all zero, such that

$$\lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_k a_k = 0.$$ 

Otherwise, the vectors $a_1, a_2, \ldots, a_k$ are *linearly independent*.

Lemma 20

If the vectors $a_1, a_2, \ldots, a_k$ of a vector space $V$ are linearly independent then

$$\lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_k a_k = 0 \Rightarrow \lambda_1 = \lambda_2 = \cdots = \lambda_k = 0.$$ 

Lemma 21

The vectors $a_1, a_2, \ldots, a_k$ of a vector space $V$ are linearly independent if and only if none of them can be expressed as a linear combination of the other vectors.
**Basis of a Vector Space**

**Definition 22 (Basis)**

The vectors $a_1, a_2, \ldots, a_n \in V$ form a *basis* of the vector space $V$ over $F$ if

1. $a_1, \ldots, a_n$ are linearly independent;
2. $\{a_1, \ldots, a_n\} = V$.

**Theorem 23**

Every basis of a finite vector space has the same number of basis vectors.

- The number $n$ of vectors of a basis is called the *dimension* of the vector space.
- $V$ is said to have *finite dimension* if the number of basis vectors is finite.
- Note: Infinitely-dimensional vector spaces tend to be significantly more complicated than finite vector spaces!

**Theorem 24**

If $a_1, \ldots, a_n$ form a basis for $V$ over $F$ then for all $a \in V$ exist uniquely determined $\lambda_1, \ldots, \lambda_n \in F$ such that $a = \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_n a_n$. 
A point is a location in a (vector) space. From a mathematical point of view it does not have any size or any other property besides its location.

A vector has a direction and a length as its main properties.

The position vector (Dt.: Ortsvektor) of a point is the vector that points from the origin of the space to the point.

It is common not to make a clean distinction between a point and its position vector.

We use the column matrix $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ to denote the $n$-dimensional vector $v$.

Note that vectors can be regarded both as column matrices and as row matrices.

While it does not matter for most applications whether or not to specify a vector as a column or row matrix, there exist a few applications for which it does matter! (E.g., multiplication of a matrix and a vector.)

Thus, pay close attention to how vectors are treated when studying a textbook or using a graphics package.
We define $\mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $\mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$, $\ldots$, $\mathbf{e}_n := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$.

$\mathbf{e}_1, \ldots, \mathbf{e}_n$ are linearly independent since $\lambda_1 \mathbf{e}_1 + \cdots + \lambda_n \mathbf{e}_n = 0$ implies

$$
\begin{pmatrix} 
\lambda_1 \\
\vdots \\
\lambda_n
\end{pmatrix} =
\begin{pmatrix} 
0 \\
\vdots \\
0
\end{pmatrix},
$$

i.e., $\lambda_1 = 0, \ldots, \lambda_n = 0$.

Let $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + a_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$.

Hence,

$$
\mathbf{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_n \mathbf{e}_n.
$$
Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

- For a 2D vector $\mathbf{v}$, we may find it more convenient to use the symbols $x$ and $y$ for its components, rather than $v_x$ and $v_y$:

$$\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{or} \quad \mathbf{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}.$$ 

- Similarly for a 3D vector:

$$\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{or} \quad \mathbf{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}.$$
Vector Algebra in $\mathbb{R}^2$ and $\mathbb{R}^3$

- Adding two 2D vectors $a$ and $b$:

$$a + b = \begin{pmatrix} a_x \\ a_y \end{pmatrix} + \begin{pmatrix} b_x \\ b_y \end{pmatrix} := \begin{pmatrix} a_x + b_x \\ a_y + b_y \end{pmatrix}.$$ 

- Similarly for higher-dimensional vectors. E.g., for 3D vectors $a$, $b$:

$$a + b = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} := \begin{pmatrix} a_x + b_x \\ a_y + b_y \\ a_z + b_z \end{pmatrix}.$$
Standard Coordinate Systems in $\mathbb{R}^2$ and $\mathbb{R}^3$

- Cartesian coordinates: $(a, b, c)$.
- Polar coordinates: $(\rho, \theta)$, with $\theta \in [0, 2\pi]$.
- Spherical coordinates: $(r, \theta, \phi)$, with $\theta \in [0, 2\pi]$ and $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.
- Cylindrical coordinates: $(\rho, \theta, c)$, with $\theta \in [0, 2\pi]$.
The z-axis of the coordinate system is aligned with the spin axis of the Earth, with the coordinate origin at the earth’s center.

The equator is defined as the intersection of the $xy$-plane (“fundamental plane”) of this coordinate system with the Earth.

Two angles are measured from the center of the Earth: latitude (Dt. “Breite”) measures the angle between any point and the equator. The other angle, longitude (Dt. “Länge”), measures the angle along the equator from an arbitrary point on the earth. (Greenwich, England, is the generally accepted zero-longitude point (Dt. “Nullmeridian”).)

Lines of constant latitude are called parallels, with the equator having latitude 0.

Lines of constant longitude are great circles that intersect at the poles and are called meridians.

A position is specified as $\alpha$ degrees East or West, and $\beta$ degrees North or South. Thus, $\alpha \in [0, 180]$, and $\beta \in [0, 90]$. 
Affine and Convex Combinations

Definition 25 (Affine combination, Dt.: Affinkombination)

Let \( P_1, P_2, \ldots, P_k \) be \( k \) points in \( \mathbb{R}^n \). An affine combination of the points \( P_1, \ldots, P_k \) is given by

\[
\sum_{i=1}^{k} \lambda_i p_i \quad \text{with} \quad \sum_{i=1}^{k} \lambda_i = 1,
\]

where \( \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R} \) are scalars.

Definition 26 (Convex combination, Dt.: Konvexkombination)

Let \( P_1, P_2, \ldots, P_k \) be \( k \) points in \( \mathbb{R}^n \). A convex combination of the points \( P_1, \ldots, P_k \) is defined as

\[
\sum_{i=1}^{k} \lambda_i p_i \quad \text{with} \quad \sum_{i=1}^{k} \lambda_i = 1 \quad \text{and} \quad \forall (1 \leq i \leq k) \quad \lambda_i \geq 0,
\]

where \( \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R} \) are scalars.
Definition 27 (Affine hull, Dt.: affine Hülle)

Let $P_1, P_2, \ldots, P_k$ be $k$ points in $\mathbb{R}^n$. The **affine hull** of the points $P_1, \ldots, P_k$ is the set

$$\left\{ \sum_{i=1}^{k} \lambda_i p_i : \lambda_1, \ldots, \lambda_k \in \mathbb{R} \text{ and } \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$ 

For a set $A \subseteq \mathbb{R}^n$ (with possibly infinitely many points), the **affine hull** of $A$ is the set

$$\left\{ \sum_{i=1}^{k} \lambda_i p_i : k \in \mathbb{N} \text{ and } P_1, P_2, \ldots, P_k \in A \text{ and } \lambda_1, \ldots, \lambda_k \in \mathbb{R} \text{ and } \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$
Convex Hull

Definition 28 (Convex hull, Dt.: konvexe Hülle)

Let $P_1, P_2, \ldots, P_k$ be $k$ points in $\mathbb{R}^n$. The convex hull of the points $P_1, \ldots, P_k$ is the set

$$\left\{ \sum_{i=1}^{k} \lambda_i p_i : \lambda_1, \ldots, \lambda_k \in \mathbb{R}_0^+ \text{ and } \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$ 

For a set $A \subseteq \mathbb{R}^n$ (with possibly infinitely many points), the convex hull of $A$ is the set

$$\left\{ \sum_{i=1}^{k} \lambda_i p_i : k \in \mathbb{N} \text{ and } P_1, P_2, \ldots, P_k \in A \text{ and } \lambda_1, \ldots, \lambda_k \in \mathbb{R}_0^+ \text{ and } \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$ 

The convex hull of $A$ is commonly denoted by $CH(A)$. 
## Convexity

### Definition 29 (Convex set, Dt.: konvexe Menge)

A set $A \subseteq \mathbb{R}^n$ is **convex** if $PQ \subseteq A$ for all points $P, Q \in A$.

### Lemma 30

For $A \subseteq \mathbb{R}^n$, the convex hull $CH(A)$ of $A$ is a convex set.

### Lemma 31

For $A \subseteq \mathbb{R}^n$, the following definitions are equivalent to Def. 28:

- $CH(A)$ is the smallest convex super-set of $A$.
- $CH(A)$ is the intersection of all convex super-sets of $A$.

- The definition of a convex hull (and of convexity) is readily extended from $\mathbb{R}^n$ to other vector spaces over $\mathbb{R}$. 

© M. Held (Univ. Salzburg)
Complex Numbers

Definition 32 (Complex numbers, Dt.: komplexe Zahlen)

The complex numbers, \( \mathbb{C} \), is formed by the set of ordered pairs of real numbers together with the operations \( + : \mathbb{C} \times \mathbb{C} \to \mathbb{C} \) and \( \cdot : \mathbb{C} \times \mathbb{C} \to \mathbb{C} \) defined as follows:

- \((a, b) + (c, d) := (a + c, b + d)\quad \forall a, b, c, d \in \mathbb{R},\)
- \((a, b) \cdot (c, d) := (ac - bd, bc + ad)\quad \forall a, b, c, d \in \mathbb{R},\)

The addition and multiplication of real numbers follows standard rules of \( \mathbb{R} \).

- Alternate view: a complex number \((a, b)\) is regarded as the sum of a real and an imaginary part: \(a + ib\), with \(i^2 := -1\). Complex numbers can be embedded into \(\mathbb{R}^2\) by regarding the complex number \((a, 0)\) as the real number \(a\), and treating the complex number \((0, 1)\) as the imaginary unit \(i\).
- Applying standard rules of algebra used when multiplying real numbers is consistent with the definitions above.

Lemma 33

Commutativity, associativity and distributivity hold for \((\mathbb{C}, +, \cdot)\).
Complex Numbers

**Definition 34 (Absolute value)**

The *absolute value* $|z|$ (or modulus or magnitude) of a complex number $z = a + ib \in \mathbb{C}$ is given by

$$|z| := \sqrt{a^2 + b^2}.$$ 

**Definition 35 (Complex conjugate, Dt.: konjugiert-komplexe Zahl)**

The complex *conjugate* $\overline{z}$ of the complex number $z = a + ib \in \mathbb{C}$ is given by

$$\overline{z} := a - ib.$$ 

**Definition 36 (Multiplicative inverse)**

The *multiplicative inverse* for $z \in \mathbb{C}$, with $z \neq 0$ is defined as

$$z^{-1} := \overline{z} |z|^{-2} = \frac{\overline{z}}{|z|^2}.$$
Complex Numbers

Lemma 37
Easy to check for all $z_1, z_2 \in \mathbb{C}$:

$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$
$\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$
$\overline{\overline{z_1}} = z_1$

$|z_1| = |\overline{z_1}|$
$z_1 \cdot z_1^{-1} = 1$
$|z_1|^2 = z_1 \cdot \overline{z_1}$

Theorem 38
The complex numbers $(\mathbb{C}, +, \cdot)$ form a field.
Complex Numbers and Euler-Moivre Formula

- A complex number $z = a + ib$ can also be written as
  
  $z = a + ib = r(\cos \phi + i \sin \phi),$
  
  with $r := |a + ib|$ and $\phi$ such that $a = r \cos \phi$ and $b = r \sin \phi$.

- By applying standard trigonometric identities, we get
  
  $z_1 \cdot z_2 = r_1 r_2[\cos(\phi_1 + \phi_2) + i \sin(\phi_1 + \phi_2)],$
  
  $z_1/z_2 = r_1/r_2[\cos(\phi_1 - \phi_2) + i \sin(\phi_1 - \phi_2)].$

- Thus, the multiplication of one complex number with another complex number can be seen as a simultaneous rotation and stretching.

Theorem 39 (Euler-Moivre)

For any $\phi \in \mathbb{R}$,

$e^{i\phi} = \cos \phi + i \sin \phi.$

- Thus, $e^{i\phi}$ traces out the unit circle as $\phi$ runs from 0 to $2\pi$.
- Important application: Modeling (electric) signals that vary periodically over time.
Mandelbrot Set and Julia Set

- A **Mandelbrot set** is the locus of points $c$ of the complex plane for which the series $(z_0, z_1, z_2, \ldots)$, with $z_{n+1} := z_n \cdot z_n + c$ and $z_0 := (0, 0)$, never leaves a circle of radius 2 centered at $(0, 0)$.
- A **Julia set** is the locus of points $z$ of the complex plane for which the series $(z_0, z_1, z_2, \ldots)$, with $z_{n+1} := z_n \cdot z_n + c$ and $z_0 := z$ and an arbitrary seed $c$, never leaves a circle of radius 2 centered at $(0, 0)$.
**Definition 40**

The set of quaternions, \( \mathbb{H} \), is given by quadrupels \((s, v)\) of real numbers, with \( s \in \mathbb{R} \) and \( v \in \mathbb{R}^3 \), together with the operations \( + : \mathbb{H} \times \mathbb{H} \to \mathbb{H} \) and \( \cdot : \mathbb{H} \times \mathbb{H} \to \mathbb{H} \) defined as follows for all \((s_1, v_1), (s_2, v_2) \in \mathbb{H} \):

\[
(s_1, v_1) + (s_2, v_2) := (s_1 + s_2, v_1 + v_2),
\]

\[
(s_1, v_1) \cdot (s_2, v_2) := (s_1s_2 - v_1v_2, s_1v_2 + s_2v_1 + v_1 \times v_2).
\]

- Discovered by William R. Hamilton in 1843. In honor of Hamilton, the quaternions are denoted by \( \mathbb{H} \).
- Extension of complex numbers: A quaternion \( P \) can also be regarded as

\[
P := s + ia + jb + kc, \quad \text{with} \quad s, a, b, c \in \mathbb{R},
\]

where the multiplication of the imaginary elements \( i, j, \) and \( k \) is defined by

\[
i^2 = j^2 = k^2 := -1 \quad \text{and} \quad ij = -ji := k.
\]

- This definition implies

\[
jk = -kj = i \quad \text{and} \quad ki = -ik = j.
\]
**Quaternions**

**Definition 41**
The conjugate of a quaternion $\mathcal{P} = (s, v) = (s, a, b, c) \in \mathbb{H}$ is defined as

$$\overline{\mathcal{P}} := (s, -v) = s - ia - jb - kc.$$ 

**Definition 42 (Unit quaternion, Dt.: Einheitsquaternion)**
The norm of a quaternion $\mathcal{P} = (s, v) = (s, a, b, c) \in \mathbb{H}$ is defined as

$$\|\mathcal{P}\| := \sqrt{s^2 + \langle v, v \rangle} = \sqrt{s^2 + a^2 + b^2 + c^2}.$$ 

A unit quaternion is a quaternion whose norm is 1.

**Definition 43**
The multiplicative inverse $\mathcal{P}^{-1}$ of a quaternion $\mathcal{P} = (s, v) \in \mathbb{H}$ is defined as

$$\mathcal{P}^{-1} := \frac{\overline{\mathcal{P}}}{\|\mathcal{P}\|^2} = \frac{1}{\|\mathcal{P}\|^2} (s, -v).$$
Quatfernion Algebra

**Lemma 44**
For all $P, Q \in \mathbb{H}$, we have
\[ \overline{P} = P \quad \text{and} \quad P \cdot Q = Q \cdot P. \]

**Lemma 45**
For all $P, Q \in \mathbb{H}$, we have
\[ (P^{-1})^{-1} = P \quad \text{and} \quad (P \cdot Q)^{-1} = Q^{-1} \cdot P^{-1}. \]

**Lemma 46**
The inverse of a unit quaternion and the product of unit quaternions are themselves unit quaternions.

- Note: The multiplication of quaternions is associative but not commutative!
- A unit quaternion can be represented by $(\cos \phi, u \sin \phi)$, where $u \in \mathbb{R}^3$ with $\|u\| = 1$.
- Important application in graphics: Modeling and interpolating spatial rotations.
Basic Linear Algebra

- Matrices
- Linear Equations
- Determinants
- Eigenvalues and Eigenvectors
- Dot Product
- Vector Cross-Product
Matrices

Definition 47 (Matrix)

An $m \times n$ matrix is a scheme of $m \cdot n$ numbers $a_{ij}$ from a field $F$, with $1 \leq i \leq m, 1 \leq j \leq n$, arranged as follows:

\[
\begin{pmatrix}
  a_{11} & \cdots & a_{1n} \\
  \vdots & \ddots & \vdots \\
  a_{m1} & \cdots & a_{mn}
\end{pmatrix}.
\]

The numbers $a_{ij}$ are called the coefficients of the matrix. The horizontal $n$-tuples $(a_{i1} \cdots a_{in})$ are called the rows of the matrix, while the vertical $m$-tuples $(a_{ij} \cdots a_{mj})$ are called columns of the matrix.

- For virtually all our applications, $a_{ij}$ will be real numbers, i.e., $F = \mathbb{R}$.
- Short-hand notation:

  \[
  A = [a_{ij}]_{i=1,j=1}^{m,n} \text{ or simply } A = [a_{ij}].
  \]

- The collection of all $m \times n$ matrices over $F$ is denoted by $M_{m \times n}(F)$, or simply by $M_{m \times n}$ if the field is irrelevant. (Again, you may think of $F$ as $\mathbb{R}$.)
Matrices

Definition 48 (Size)
The numbers $m$ and $n$ describe the size of a matrix. A matrix is square if $m = n$.

Definition 49 (Zero matrix, Dt.: Nullmatrix)
For $m, n \in \mathbb{N}$, the matrix in $M_{m \times n}(F)$ with all elements equal to 0 is called the zero matrix (of size $m \times n$), and is denoted by the symbol $0$.

Definition 50 (Identity matrix, Dt.: Einheitsmatrix)
For $n \in \mathbb{N}$, the $n \times n$ matrix $I := [\delta_{ij}]$, defined by $\delta_{ij} := 1$ if $i = j$ and $\delta_{ij} := 0$ otherwise, is called the $n \times n$ identity matrix.

E.g., for $n = 4$ we have

$0 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and $I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. 
Matrices

**Definition 51 (Matrix identity)**

Two matrices $A$ and $B$ over the same field $F$ are said to be equal if $A$ and $B$ have the same size and if corresponding elements are equal; that is, $A, B \in M_{m\times n}(F)$ and $A = [a_{ij}], B = [b_{ij}]$, with $a_{ij} = b_{ij}$ for $1 \leq i \leq m, 1 \leq j \leq n$.

**Definition 52 (Rank, Dt.: Rang)**

The *rank* of a matrix $A$ is the number of linearly independent columns of $A$.

**Definition 53 (Sparse, Dt.: dünn besetzt)**

For $m, n \in \mathbb{N}$, the $m \times n$ matrix $A$ is called *sparse* if $k \ll m \cdot n$ holds for the number $k$ of non-zero coefficients of $A$.

- Note: Storing an $n \times n$ matrix consumes $O(n^2)$ space, unless special precautions are taken (e.g., in the case of sparse matrices)!

**Theorem 54**

$M_{m\times n}(F)$, with addition and scalar multiplication as defined in Def. 55+56 forms a vector space over $F$ for all $m, n \in \mathbb{N}$. 
**Definition 55 (Matrix addition)**

Let $A$, $B$ be matrices of the same size. Then $A + B$ is the matrix obtained by adding corresponding elements of $A$ and $B$; that is,

$$A + B = [a_{ij}] + [b_{ij}] := \begin{pmatrix}
    a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\
    a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn}
\end{pmatrix}.$$

**Definition 56 (Scalar multiplication)**

Consider a matrix $A$ and $\lambda \in F$. (Thus, $\lambda$ is a scalar.) Then $\lambda A$ is the matrix obtained by multiplying all elements of $A$ by $\lambda$; that is,

$$\lambda A = \lambda[a_{ij}] := [\lambda a_{ij}] = \begin{pmatrix}
    \lambda a_{11} & \cdots & \lambda a_{1n} \\
    \lambda a_{21} & \cdots & \lambda a_{2n} \\
    \vdots & \ddots & \vdots \\
    \lambda a_{m1} & \cdots & \lambda a_{mn}
\end{pmatrix}.$$
Lemma 57

The matrix operations of addition, scalar multiplication, additive inverse and subtraction satisfy the usual laws of arithmetic. (In what follows, $A$, $B$, $C$ are matrices of the same size over the same field $F$, and $\lambda$, $\mu$ are scalars out of $F$.)

1. $(A + B) + C = A + (B + C)$;
2. $A + B = B + A$;
3. $0 + A = A$;
4. $A + (-A) = 0$;
5. $(\lambda + \mu)A = \lambda A + \mu A$;
6. $\lambda(A + B) = \lambda A + \lambda B$;
7. $\lambda(\mu A) = (\lambda \mu)A$;
8. $1A = A$;
9. $0A = 0$;
10. $(-1)A = -A$;
11. $\lambda A = 0 \Rightarrow \lambda = 0$ or $A = 0$. 

© M. Held (Univ. Salzburg)
Matrix Algebra

**Definition 58 (Matrix multiplication)**

Let $A$ be a matrix of size $m \times n$ and $B$ be a matrix of size $n \times p$; that is, the number of columns of $A$ equals the number of rows of $B$. Then $A \cdot B$ is the $m \times p$ matrix $C = [c_{ik}]$ whose $(i, k)$-th element is defined by the formula

$$c_{ik} := \sum_{j=1}^{n} a_{ij} b_{jk} = a_{i1} b_{1k} + \cdots + a_{in} b_{nk}.$$

**Lemma 59**

Matrix multiplication obeys most of the standard laws of arithmetic except for the commutative law:

1. $(AB)C = A(BC)$ if $A$, $B$, $C$ are $m \times n$, $n \times p$, $p \times q$, respectively;
2. $\lambda(AB) = (\lambda A)B = A(\lambda B)$;
3. $A(-B) = (-A)B = -(AB)$;
4. $(A + B)C = AC + BC$ if $A$, $B$ are $m \times n$ and $C$ is $n \times p$;
5. $D(A + B) = DA + DB$ if $A$, $B$ are $m \times n$ and $D$ is $p \times m$.

Note: $AB \neq BA$ even if $A$, $B$ are square. Also, $AB = 0 \iff [A = 0$ or $B = 0]$. 

© M. Held (Univ. Salzburg) Geometrisches Rechnen (WS 2013/14)
Inversion of a Matrix

Definition 60 (Invertible, Dt.: invertierbar)

An \( n \times n \) matrix \( A \) is invertible (or non-singular) if there exists an \( n \times n \) matrix \( B \) such that

\[
AB = BA = I.
\]

If \( A \) is invertible then the inverse matrix is denoted by \( A^{-1} \).

Theorem 61

If \( A \) has inverse matrices \( B, C \) then \( B = C \).

- Note that \( A^{-1} \) can be obtained (if it exists) by solving \( A \xi_i = e_i \) for \( 1 \leq i \leq n \); the vectors \( \xi_i \) form the columns of \( A^{-1} \).

Theorem 62

If \( A, B \) are non-singular matrices of the same size then \( AB \) is not singular, and

\[
(AB)^{-1} = B^{-1}A^{-1},
\]

i.e., the inverse of the product equals the product of the inverses in the reverse order.
Definition 63 (Transpose, Dt.: transponiert)

Consider an \( m \times n \) matrix \( A \). The \textit{transpose} of \( A \) is the matrix \( A^t \) obtained by interchanging the rows and columns of \( A \).

Consequently, \( A^t \) is an \( n \times m \) matrix.

Lemma 64

The transpose operation has the following properties for all matrices \( A \):

1. \((A^t)^t = A\);
2. \((A + B)^t = A^t + B^t\);
3. \((\lambda A)^t = \lambda A^t\) for a scalar \( \lambda \).
4. \((AB)^t = B^t A^t\);
5. If \( A \) is non-singular then \( A^t \) is also non-singular and we have \((A^t)^{-1} = (A^{-1})^t\).
**Definition 65 (Symmetric, Dt.: symmetrisch)**
A matrix $A$ is called *symmetric* if $A^t = A$.

**Definition 66 (Diagonal matrix, Dt.: Diagonalmatrix)**
A square matrix $A$ is called *diagonal* if $a_{ij} = 0$ for $i \neq j$.

**Definition 67 (Upper-triangular, Dt.: obere Dreiecksmatrix)**
A square matrix $A$ is called *upper-triangular* if $a_{ij} = 0$ for $i > j$.

**Definition 68 (Orthogonal, Dt.: orthogonal)**
A square matrix $A$ is called *orthogonal* if $A \cdot A^t = I = A^t \cdot A$.

**Lemma 69**
If a square matrix $A$ is *orthogonal* then $A^{-1} = A^t$. 
Fast Matrix Multiplication

- Standard multiplication of two $n \times n$ matrices results in $O(n^3)$ many multiplications.

**Lemma 70 (Strassen 1969)**

Seven multiplications suffice to compute the multiplication of two $2 \times 2$ matrices. In general, $O(n^{\log_2 7}) \equiv O(n^{2.807\ldots})$ multiplications suffice for $n \times n$ matrices.

- Strassen’s algorithm is more complex and less stable than the standard naïve algorithm. But it is considerably more efficient for $n > 100$, and it is very useful for large matrices over finite fields.

**Lemma 71 (Coppersmith&Winograd 1990)**

$O(n^{2.376\ldots})$ multiplications suffice for $n \times n$ matrices.

**Lemma 72 (Stothers 2010)**

$O(n^{2.374\ldots})$ multiplications suffice for $n \times n$ matrices. (No peer-reviewed paper yet!)

**Lemma 73 (Williams 2011, 2012)**

$O(n^{2.373\ldots})$ multiplications suffice for $n \times n$ matrices.
Fast Matrix Multiplication

- **Strassen**: For $A, B \in M_{2 \times 2}$, we compute $C = A \cdot B$ via

  \[
  p_1 := (a_{1,2} - a_{2,2})(b_{2,1} + b_{2,2})
  \]
  \[
  p_2 := (a_{1,1} + a_{2,2})(b_{1,1} + b_{2,2})
  \]
  \[
  p_3 := (a_{1,1} - a_{2,1})(b_{1,1} + b_{1,2})
  \]
  \[
  p_4 := (a_{1,1} + a_{1,2})b_{2,2}
  \]
  \[
  p_5 := a_{1,1}(b_{1,2} - b_{2,2})
  \]
  \[
  p_6 := a_{2,2}(b_{2,1} - b_{1,1})
  \]
  \[
  p_7 := (a_{2,1} + a_{2,2})b_{1,1}
  \]

  and set

  \[
  c_{1,1} := a_{1,1}b_{1,1} + a_{1,2}b_{2,1} = p_1 + p_2 - p_4 + p_6
  \]
  \[
  c_{1,2} := a_{1,1}b_{1,2} + a_{1,2}b_{2,2} = p_4 + p_5
  \]
  \[
  c_{2,1} := a_{2,1}b_{1,1} + a_{2,2}b_{2,1} = p_6 + p_7
  \]
  \[
  c_{2,2} := a_{2,1}b_{1,2} + a_{2,2}b_{2,2} = p_2 - p_3 + p_5 - p_7.
  \]

- Use block matrices to apply this concept for $n > 2$. 
Definition 74 (Linear equation, Dt.: lineare Gleichung)

A linear equation in $n$ unknowns $x_1, x_2, \ldots, x_n$ is an equation of the form

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b,$$

where $a_1, \ldots, a_n, b$ are given (real) numbers.

Definition 75 (System of linear equations, Dt.: lineares Gleichungssystem)

A system of $m$ linear equations in $n$ unknowns $x_1, x_2, \ldots, x_n$ is a family of linear equations

$$a_{11} x_1 + \cdots + a_{1n} x_n = b_1$$
$$\vdots$$
$$a_{m1} x_1 + \cdots + a_{mn} x_n = b_m,$$

where $a_{11}, \ldots, a_{mn}, b_1, \ldots, b_m$ are given (real) numbers. The system is called homogeneous if $b_1 = b_2 = \cdots = b_m = 0.$
Matrices and Linear Equations

Consider an \( m \times n \) matrix \( A = [a_{ij}]_{i=1,j=1}^{m,n} \) and a vector \( b \in \mathbb{R}^m \). For \( x \in \mathbb{R}^n \), the term

\[
A x = b
\]
describes the system of linear equations

\[
\begin{pmatrix}
  b_1 \\
  b_2 \\
  \vdots \\
  b_m
\end{pmatrix}
= b =
A x =
\begin{pmatrix}
  a_{11} & \cdots & a_{1n} \\
  a_{21} & \cdots & a_{2n} \\
  \vdots & \ddots & \vdots \\
  a_{m1} & \cdots & a_{mn}
\end{pmatrix}
\begin{pmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_n
\end{pmatrix}
\]

\[
= \begin{pmatrix}
  a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n \\
  a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n \\
  \vdots \\
  a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n
\end{pmatrix}.
\]
Matrices and Linear Equations

The matrix
\[
\begin{pmatrix}
  a_{11} & \cdots & a_{1n} \\
  a_{21} & \cdots & a_{2n} \\
  \vdots & \ddots & \vdots \\
  a_{m1} & \cdots & a_{mn}
\end{pmatrix}
\]
is called the coefficient matrix of the system.

The matrix
\[
\begin{pmatrix}
  a_{11} & \cdots & a_{1n} & b_1 \\
  a_{21} & \cdots & a_{2n} & b_2 \\
  \vdots & \ddots & \vdots \\
  a_{m1} & \cdots & a_{mn} & b_m
\end{pmatrix}
\]
is called the augmented matrix of the system.
Solutions of Linear Equations

Definition 76
A system of linear equations in $n$ unknowns is called *consistent* if it has a solution, i.e., if there exist (real) numbers $x_1, x_2, \ldots, x_n$ that satisfy all equations simultaneously.

- A homogeneous system is always consistent, since $x_1 = x_2 = \cdots = x_n = 0$ always is a solution, which is called *trivial* solution. Any other solution of a homogeneous system is called a *non-trivial* solution.

Theorem 77
A homogeneous system of $m$ linear equations in $n$ unknowns always has a non-trivial solution if $m < n$. 

© M. Held (Univ. Salzburg) 

Geometrisches Rechnen (WS 2013/14)

71
Theorem 78

The system $A \mathbf{x} = \mathbf{b}$ is consistent if and only if the number of linearly independent columns of the coefficient matrix equals the number of linearly independent columns of the augmented matrix, i.e., if and only if the rank of the coefficient matrix equals the rank of the augmented matrix.

Theorem 79

Assume that the system $A \mathbf{x} = \mathbf{b}$ is consistent. This system has a unique solution if and only if the number of linearly independent columns of the coefficient matrix equals the number of unknowns.
Lemma 80

The following three types of elementary row operations may be performed on matrices without changing the rank:

1. Interchanging two rows;
2. Multiplying a row by a nonzero scalar;
3. Adding a multiple of one row to another row.

Definition 81

A matrix $A$ is row-equivalent to a matrix $B$ if $B$ is obtained from $A$ by a sequence of elementary row operations.

Theorem 82

If $A$ and $B$ are row-equivalent augmented matrices of two systems of linear equations, then the two systems have the same solution sets.
Definition 83 (Reduced row-echelon form, Dt.: Treppennormalform)

A matrix is in reduced row-echelon form if

1. all zero rows (if any) are at the bottom of the matrix;
2. if two successive rows are nonzero then the second row starts with more zeros than the first (moving from left to right and top to bottom);
3. the leading (leftmost nonzero) entry in each nonzero row is 1;
4. all other elements of the column in which the leading entry 1 occurs are zeros.

Sample matrix in reduced row-echelon form:

\[
\begin{pmatrix}
0 & 1 & * & 0 & 0 & * & * & 0 & * \\
0 & 0 & 0 & 1 & 0 & * & * & 0 & * \\
0 & 0 & 0 & 0 & 1 & * & * & 0 & * \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]
Gauss-Jordan Algorithm

- The following algorithm transforms an augmented matrix $A$ into a matrix $A'$ that is in reduced row-echelon form, using elementary row operations:
  
  Initially, $k := 1$.
  
  If the rows $k, \ldots, m$ all are zero then the matrix is in reduced row-echelon form.
  
  Otherwise, suppose that the first column which has a non-zero element in the rows below the first $k - 1$ rows is column $c_k$. By interchanging the rows below the first $k - 1$ rows, if necessary, we ensure that the element $a_{k,c_k}$ is nonzero. Convert $a_{k,c_k}$ to 1. By adding suitable multiples of row $k$ to the remaining rows, where necessary, we ensure that all remaining elements in column $c_k$ are zero.
  
  If $k < m$, repeat this process for $k := k + 1$.
  
  This process will eventually stop after $r$ steps, either because we run out of rows (if $k = m$), or because we run out of non-zero columns.
  
  In general, the final matrix $A'$ will be in reduced row-echelon form and will have $r$ non-zero rows, with leading entries 1 in columns $c_1, \ldots, c_r$, respectively.
  
  By swapping columns (and updating the solution vector $\mathbf{x}$ accordingly) we can guarantee that the $r$ non-zero rows have their leading 1’s in columns $1, \ldots, r$. 
Gauss-Jordan Algorithm

Thus, the Gauss-Jordan algorithm transforms an augmented matrix $A$ into a matrix $A'$ of the following form:

$$
\begin{pmatrix}
1 & 0 & a'_{1,r+1} & \cdots & a'_{1n} & b'_1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 1 & a'_{r,r+1} & \cdots & a'_{rn} & b'_r \\
\hline
0 & \cdots & \cdots & \cdots & \cdots & b'_{r+1} \\
0 & \cdots & \cdots & \cdots & \cdots & b'_m
\end{pmatrix}
$$

If $r = n + 1$ then the system is inconsistent. (The last row reads $0 \cdot x'_1 + 0 \cdot x'_2 + \cdots + 0 \cdot x'_n = 1$, which has no solutions.)

If $r \leq n$ then the system is inconsistent unless $b'_{r+1} = b'_{r+2} = \cdots = b'_m = 0$.

If $r = n$ and $b'_{r+1} = b'_{r+2} = \cdots = b'_m = 0$, then there exists a unique solution $x'_1 = b'_1, x'_2 = b'_2, \ldots, x'_n = b'_n$. 
Gauss-Jordan Algorithm

- The Gauss-Jordan algorithm transforms an augmented matrix $A$ into a matrix $A'$ of the following form:

\[
\begin{pmatrix}
1 & 0 & a'_{1,r+1} & \cdots & a'_{1n} & b'_1 \\
\vdots & \vdots & \vdots & & \vdots & \\
0 & 1 & a'_{r,r+1} & \cdots & a'_{rn} & b'_r \\
\hline
0 & & & & & b'_{r+1} \\
\end{pmatrix}
\]

- If $r < n$ and $b'_{r+1} = b'_{r+2} = \cdots = b'_m = 0$, then there are infinitely many solutions:

\[
x'_1 = b'_1 - a'_{1,r+1}x'_{r+1} - a'_{1,r+2}x'_{r+2} - \cdots - a'_{1n}x'_n,
\]

\[
\vdots
\]

\[
x'_r = b'_r - a'_{r,r+1}x'_{r+1} - a'_{r,r+2}x'_{r+2} - \cdots - a'_{rn}x'_n.
\]

The independent unknowns $x'_{r+1}, \ldots, x'_n$ may take on arbitrary values.
Sample Linear System

\[
\begin{align*}
\begin{cases}
  x_1 + x_2 + 2x_3 + 3x_4 &= 4 \\
  2x_1 + 2x_2 + 3x_3 + 4x_4 &= 5
\end{cases}
\end{align*}
\]

\[
(A, b) = \begin{pmatrix}
  1 & 1 & 2 & 3 & 4 \\
  2 & 2 & 3 & 4 & 5
\end{pmatrix} + I \cdot (-2) \sim \begin{pmatrix}
  1 & 1 & 2 & 3 & 4 \\
  0 & 0 & -1 & -2 & -3
\end{pmatrix} \cdot (-1)
\]

\[
\sim \begin{pmatrix}
  1 & 1 & 2 & 3 & 4 \\
  0 & 0 & 1 & 2 & 3
\end{pmatrix} x_2 \leftrightarrow x_3 \sim \begin{pmatrix}
  1 & 2 & 1 & 3 & 4 \\
  0 & 1 & 0 & 2 & 3
\end{pmatrix} + II \cdot (-2)
\]

\[
\sim \begin{pmatrix}
  1 & 0 & 1 & -1 & -2 \\
  0 & 1 & 0 & 2 & 3
\end{pmatrix}
\]

\[\Rightarrow \text{Solution} \equiv \left\{ \begin{pmatrix}
  -2 \\
  0 \\
  3 \\
  0
\end{pmatrix} + \lambda_1 \begin{pmatrix}
  -1 \\
  1 \\
  0 \\
  0
\end{pmatrix} + \lambda_2 \begin{pmatrix}
  1 \\
  0 \\
  -2 \\
  1
\end{pmatrix} : \lambda_1, \lambda_2 \in \mathbb{R} \right\}
Determinants

**Definition 84 (Submatrix, Dt.: Untermatrix)**

Let $A \in M_{n \times n}$, with $n \geq 2$. Let $A_{ij}(A)$, or simply $A_{ij}$ if there is no ambiguity, denote the $(n-1) \times (n-1)$ submatrix of $A$ formed by deleting the $i$-th row and $j$-th column of $A$.

- Example:

  \[
  A = \begin{pmatrix}
  1 & 0 & 1 \\
  2 & 1 & 2 \\
  0 & 4 & 4 \\
  \end{pmatrix}, \quad A_{12} = \begin{pmatrix}
  2 & 2 \\
  0 & 4 \\
  \end{pmatrix}, \quad A_{33} = \begin{pmatrix}
  1 & 0 \\
  2 & 1 \\
  \end{pmatrix}
  \]

**Definition 85 (Determinant)**

The determinant, $\det(A)$, of an $n \times n$ matrix $A$ is defined recursively by the so-called first-row Laplace expansion:

\[
\det(A) := \begin{cases}
  a_{11} & \text{if } n = 1, \\
  \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \cdot \det(A_{1j}) & \text{if } n > 1.
\end{cases}
\]

- Note that the term $|A|$ is also commonly used for denoting the determinant of $A$. 
Laplace Expansion

- One can prove (albeit the proof is not entirely straightforward) that a determinant can be obtained by using any row or column for expansion if the following chess-board pattern is used for determining the signs of the summands:

\[
\begin{bmatrix}
+ & - & + & \cdots \\
- & + & - & \cdots \\
+ & - & + & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

- E.g.,

\[
\det(A) = \sum_{j=1}^{n} (-1)^j a_{2j} \cdot \det(A_{2j}) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \cdot \det(A_{1j})
\]

and

\[
\det(A) = \sum_{i=1}^{n} (-1)^i a_{i2} \cdot \det(A_{i2}) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \cdot \det(A_{1j}).
\]
2 × 2 and 3 × 3 Determinants

Lemma 86

Determinant of a 2 × 2 matrix: For all $a, b, c, d \in \mathbb{R}$,

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Determinant of a 3 × 3 matrix: For all $a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33} \in \mathbb{R}$,

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$= a_{11} \cdot \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{21} \cdot \det \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} + a_{31} \cdot \det \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{13}a_{22})$$

$$= a_{11}a_{22}a_{33} + a_{21}a_{13}a_{32} + a_{31}a_{12}a_{23} - a_{11}a_{23}a_{32} - a_{21}a_{12}a_{33} - a_{31}a_{13}a_{22}.$$
Mnemonic for Computing $3 \times 3$ Determinants (Sarrus)

$$\det \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix} = x_1 y_2 z_3 + y_1 z_2 x_3 + z_1 x_2 y_3 - x_3 y_2 z_1 - y_3 z_2 x_1 - z_3 x_2 y_1.$$
Properties of Determinants

Lemma 87
If two columns or rows of a matrix are interchanged, then the determinant changes sign (but its absolute value does not change).

Lemma 88
If a row (or column) of a matrix is zero, then its determinant is zero.

Lemma 89
The determinant is a linear function of each row and column.

Lemma 90
If a multiple of a row is added to another row, then the value of the determinant remains unchanged. Similarly for columns.

Lemma 91
If two rows or columns of a matrix are equal then the determinant is zero.
Properties of Determinants

Lemma 92
The determinant of the product of two (square) matrices is the product of the determinants of the matrices:

\[ \det(AB) = \det(A) \det(B) \]

for all \( A, B \in M_{n \times n} \).

Lemma 93
A matrix and its transpose have equal determinants, i.e., for all (square) \( A \),

\[ \det A^t = \det A. \]

Lemma 94
The determinant of an orthogonal matrix is \( \pm 1 \).

Lemma 95
The (square) matrix \( A \) is not singular if and only if \( \det(A) \neq 0 \).
Properties of Determinants

Lemma 96

The determinant of an upper-triangular matrix

\[
\begin{pmatrix}
    a_{11} & * & \cdots & \cdots & * \\
    0 & a_{22} & \cdots & \cdots & \vdots \\
    \vdots & \vdots & \ddots & \cdots & \vdots \\
    \vdots & \vdots & \ddots & * & \vdots \\
    0 & \cdots & \cdots & 0 & a_{nn}
\end{pmatrix}
\]

is given by the product of its diagonal elements: \( \det(A) = \prod_{i=1}^{n} a_{ii} \).
Calculating Determinants Manually

Make sure to make good use of the lemmas stated on previous slides!

\[
\begin{vmatrix}
1 & 2 & -1 & 3 \\
0 & 1 & 4 & 2 \\
0 & 1 & 0 & 4 \\
1 & 0 & 2 & 1
\end{vmatrix}
\]

\[
\begin{vmatrix}
0 & 2 & -3 & 2 \\
0 & 1 & 4 & 2 \\
0 & 1 & 0 & 4 \\
1 & 0 & 2 & 1
\end{vmatrix}
\]

Expansion by first column

\[
\begin{vmatrix}
2 & -3 & 2 \\
1 & 4 & 2 \\
1 & 0 & 4
\end{vmatrix}
\]

\[
\begin{vmatrix}
0 & -3 & -6 \\
0 & 4 & -2 \\
1 & 0 & 4
\end{vmatrix}
\]

\[
\begin{vmatrix}
-3 & -6 \\
4 & -2
\end{vmatrix}
\]

\[
\begin{vmatrix}
-3 \cdot (-2) - (-6 \cdot 4)
\end{vmatrix}
\]

\[
-((-3 \cdot (-2)) - (-6 \cdot 4)) = -30.
\]
Implementing Determinant Calculations

- The recursive formula results in a horrendous algorithmic complexity: If $T(n)$ denotes the number of multiplications needed for computing the determinant of an $n \times n$ matrix, with $T(2) := 2$, then $T(n) = n + n \cdot T(n-1)$ and, thus, $T(n) > n!$.
- Hence, the recursive formula is not suitable for anything but small matrices.
- Standard alternative: Apply Gaussian elimination in order to transform the input matrix into an upper-triangular matrix, at a cost of $\Theta(n^3)$ operations.
- Unfortunately, this transformation introduces divisions.
- Bird (IPL 111(21–22), 2011) presents a simple method that requires $O(nM(n))$ additions and multiplications for an $n \times n$ matrix, where $M(n)$ is the number of such operations needed for matrix multiplication.
- If naïve matrix multiplication is used then we get $\Theta(n^4)$; no $\Theta(n^3)$ division-free determinant calculation is known.
Lemma 97

The linear system $A\mathbf{x} = b$, with $A \in M_{n \times n}$, has a unique solution if and only if $\det(A) \neq 0$.

Lemma 98 (Cramer’s Rule)

If $\det(A) \neq 0$, for $A \in M_{n \times n}(\mathbb{R})$, then the solution of $A\mathbf{x} = b$ is given by

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \ldots, \quad x_n = \frac{\det(A_n)}{\det(A)},$$

where $A_i$ is the matrix formed by replacing the $i$-th column of the coefficient matrix $A$ by the right-hand side $b$. 
Geometric Interpretation of Determinants

Lemma 99

For points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ in $\mathbb{R}^2$,

$$\det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}$$

is positive if the triangle formed by the origin $O = (0, 0)$ and the points $P_1$ and $P_2$ has counter-clockwise (CCW) orientation. It is negative for a clockwise (CW) orientation. This determinant is zero if $P_1$, $P_2$ and $O$ are collinear.

Lemma 100

For points $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$ and $P_3 = (x_3, y_3)$ in $\mathbb{R}^2$,

$$\det \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix}$$

is positive if the triangle $\Delta(P_1, P_2, P_3)$ formed by $P_1$, $P_2$, $P_3$ has counter-clockwise (CCW) orientation. It is negative for a clockwise (CW) orientation. This determinant is zero if $P_1$, $P_2$ and $P_3$ are collinear.
Geometric Interpretation of Determinants

**Lemma 101**
Apart from the sign,

\[
\frac{1}{2} \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}
\]
corresponds to the area of the triangle \( \Delta(O, P_1, P_2) \).

**Lemma 102**
Apart from the sign,

\[
\frac{1}{2} \det \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix}
\]
corresponds to the area of the triangle \( \Delta(P_1, P_2, P_3) \).
Definition 103 (Eigenvalue, Dt.: Eigenwert)

Consider a square matrix $A$. A scalar $\lambda$ is called eigenvalue of $A$ if a vector $v \neq 0$ exists such that

$$Av = \lambda v.$$ 

Such a vector $v$ is called eigenvector of $A$.

- A scalar $\lambda$ is an eigenvalue of matrix $A$ if and only if the homogeneous linear system of equations $(A - \lambda I)v = 0$ has a non-zero solution. This is the case if and only if $(A - \lambda I)$ is singular, that is, if $\det(A - \lambda I) = 0$.
- Thus, the eigenvalues of a matrix $A$ are the zeros of the characteristic polynomial

$$p_A(\lambda) = \det(A - \lambda I).$$

- While this approach works for any $n \times n$ matrix, it becomes tedious for $n > 4$.
- Sample application of eigenvalues and eigenvectors: Principal Components Analysis.
Principal Components Analysis (PCA)

- Sample problem: Suppose that you are given a cloud of points in $\mathbb{R}^3$. Somebody tells you that all those points lie inside of an (unknown) ellipsoid. How would you rotate and translate those points such that the main axes of the ellipsoid coincide with the coordinate axes?

- Roughly, Principal Components Analysis (PCA) is a statistical method for finding “structure” in such a point cloud.

- PCA starts with subtracting the mean of all points from every point. This is equivalent to translating the point cloud such that its centroid coincides with the origin.

- Then, PCA chooses the first PCA axis as that line which goes through the centroid of the point cloud, but also minimizes the square of the distance of each point to that line. Thus, the line is as close to all of the points as possible. Equivalently, the line goes through the maximum variation in the point cloud.

- The second PCA axis also goes through the centroid, and also goes through the maximum variation in the points in a direction that is orthogonal to the first axes. Similarly for the third axes.
Principal Components Analysis

- Consider $n$ points $P_i = (x_i, y, z_i) \in \mathbb{R}^3$.
- For our sample application, the PCA axes can be computed by finding the eigenvalues and eigenvectors of the covariance matrix $\text{Cov}$ of the coordinates of the $n$ points:

$$\text{Cov}(x, y, z) := \begin{pmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{pmatrix},$$

where

$$\text{cov}(x, y) := \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

and $\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\bar{y} := \frac{1}{n} \sum_{i=1}^{n} y_i$.

Similarly for the other entries of the covariance matrix.
- The origin of the PCA axes is given by the point $(\bar{x}, \bar{y}, \bar{z})$. 
Dot Product

Definition 104 (Dot product, Dt.: Skalarprodukt, inneres Produkt)

Consider a vector space $V$ over a field $F$, where $F$ is either $\mathbb{R}$ or $\mathbb{C}$. A mapping

$$
\cdot : V \times V \rightarrow F
$$

$$(a, b) \mapsto \langle a, b \rangle$$

is called a \textit{dot product} (or \textit{inner product}) on $V$ if $\forall a, b, c \in V$

1. $\langle \lambda_1 a + \lambda_2 b, c \rangle = \lambda_1 \langle a, c \rangle + \lambda_2 \langle b, c \rangle \ \forall \lambda_1, \lambda_2 \in F$;
2. $\langle a, b \rangle = \langle b, a \rangle$;
3. $\langle a, a \rangle \geq 0$;
4. $\langle a, a \rangle = 0 \Rightarrow a = 0$.

- Note that Condition 2 ensures that $\langle a, a \rangle \in \mathbb{R}$.
- If $F$ is $\mathbb{R}$ then commutativity holds. (In the sequel we will assume $F$ to be $\mathbb{R}$.)
- Be warned that the notation is not uniform: $a \cdot b$ and $(a \ | \ b)$ are two other common notations for denoting the dot product of $a$ and $b$.
- Note the difference between $a \cdot b$ for $a, b \in V$, and $\lambda \cdot a$ for $\lambda \in F$ and $a \in V$.!
Norm and Triangle Inequality

**Definition 105 (Length)**

Based on a dot product on $V$ (over $\mathbb{R}$), we can define the *length* (or *norm*) of a vector $a \in V$ as the following mapping from $V$ to $\mathbb{R}$:

$$\|a\| := \sqrt{\langle a, a \rangle}.$$ 

**Lemma 106**

We get the following standard properties of a norm for $\| . \|$ for $\forall a, b \in V$:

1. $\|a\| \geq 0$;
2. $\|a\| = 0 \implies a = 0$;
3. $\|\lambda a\| = |\lambda| \cdot \|a\| \quad \forall \lambda \in \mathbb{R}$;
4. Triangle Inequality (Dt.: Dreiecksungleichung):
   $$\|a + b\| \leq \|a\| + \|b\|.$$
Lemma 107 (Cauchy-Schwarz Inequality)

\[ |\langle a, b \rangle| \leq \|a\| \cdot \|b\| \quad \forall a, b \in V. \]

- Note that, for \( a, b \neq 0 \), the Cauchy-Schwarz inequality implies

\[-1 \leq \frac{\langle a, b \rangle}{\|a\| \cdot \|b\|} \leq 1.\]

We will make use of this fact when defining angles between vectors.
Standard Dot Product and Standard Norm on $\mathbb{R}^n$

- For $V = \mathbb{R}^n$, it is easy to prove that the definition

$$\langle a, b \rangle := a_1 b_1 + \ldots + a_n b_n \quad \text{for} \quad a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

does indeed yield a dot product on $V$.

- In the sequel, we will always use this dot product when referring to “the dot product” on $\mathbb{R}^n$.

- Note that this definition of a dot product and its corresponding norm on $\mathbb{R}^n$ matches our intuitive notion of the distance, $d(a, b)$, of two points $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ in $\mathbb{R}^n$: their distance is given by the length of the vector from $a$ to $b$, i.e.,

$$d(a, b) := \|a - b\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}.$$

Definition 108 (Unit vector, Dt.: Einheitsvektor)

A vector $a$ is said to be a unit vector if $\|a\| = 1$. 

© M. Held (Univ. Salzburg) Geometrisches Rechnen (WS 2013/14) 100
Other Widely Used Norms on $\mathbb{R}^n$

- The norm
  \[ \|a - b\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2} \]
  is also called $L_2$-norm and then denoted by $\|a - b\|_2$, in order to distinguish it from other well-known norms on $\mathbb{R}^n$, such as the $L_1$-norm (Manhattan metric)
  \[ \|a - b\|_1 := |a_1 - b_1| + |a_2 - b_2| + \cdots + |a_n - b_n|, \]
  or the $L_\infty$-norm (maximum norm)
  \[ \|a - b\|_\infty := \max_{1 \leq i \leq n} |a_i - b_i|. \]

- Note that the shape of a “unit circle” changes relative to different norms.
Definition 109 (Angle between vectors)

The \textit{angle}, \( \alpha \), \textit{between non-zero vectors} \( \mathbf{a}, \mathbf{b} \in \mathbb{R}^n \) is given by

\[
\cos \alpha := \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\| \mathbf{a} \| \cdot \| \mathbf{b} \|}.
\]

Definition 110 (Perpendicular, Dt.: senkrecht)

The vectors \( \mathbf{a}, \mathbf{b} \in \mathbb{R}^n \) are said to be \textit{perpendicular} (or \textit{orthogonal}), denoted by \( \mathbf{a} \perp \mathbf{b} \), if

\[
\langle \mathbf{a}, \mathbf{b} \rangle = 0.
\]

Lemma 111 (Pythagoras)

For \( \mathbf{a}, \mathbf{b} \in \mathbb{R}^n \),

\[
\langle \mathbf{a}, \mathbf{b} \rangle = 0 \implies \| \mathbf{a} + \mathbf{b} \|^2 = \| \mathbf{a} \|^2 + \| \mathbf{b} \|^2.
\]
Orthogonality, Angle and Projections

**Definition 112 (Parallel)**

The non-zero vectors $a, b \in \mathbb{R}^n$ are said to be *parallel*, denoted by $a \parallel b$, if there exists $\lambda \in \mathbb{R}$ such that

$$a = \lambda b.$$

**Lemma 113**

The length of the orthogonal projection of a vector $a$ onto a non-zero vector $b$ is given by

$$\frac{\langle a, b \rangle}{\|b\|}.$$
Orthonormal Basis of a Vector Space

**Definition 114 (Orthogonal basis)**

The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ form an *orthogonal basis* of a vector space $V$ (over $\mathbb{R}$, such as $\mathbb{R}^n$) if

1. the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ form a basis of $V$;
2. $\forall (1 \leq i, j \leq n), \ [i \neq j \Rightarrow \mathbf{a}_i \cdot \mathbf{a}_j = 0]$.

**Definition 115 (Orthonormal basis)**

The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ form an *orthonormal basis* of a vector space $V$ (over $\mathbb{R}$, such as $\mathbb{R}^n$) if

1. the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ form a basis of $V$;
2. $\forall (1 \leq i, j \leq n), \ \mathbf{a}_i \cdot \mathbf{a}_j = \delta_{ij}$.

- The algorithm by Gram-Schmidt can be used to transform an arbitrary basis into an orthonormal basis.

**Lemma 116**

An $n \times n$ matrix $\mathbf{A} \in M_{n\times n}(\mathbb{R})$ is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$. 
Vector Cross-Product in $\mathbb{R}^3$

**Definition 117 (Cross-product, Dt.: Kreuzprodukt)**

Let $a = (a_x, a_y, a_z), b = (b_x, b_y, b_z) \in \mathbb{R}^3$. The (vector) cross-product of $a$ and $b$ is given by

$$a \times b := \begin{vmatrix}
    a_y & b_y \\
    a_z & b_z \\
    a_x & b_x \\
    a_y & b_y \\
\end{vmatrix}.$$

Note: The cross-product is only defined in $\mathbb{R}^3$!
Properties of the Cross-Product

Lemma 118

The following properties of the vector cross-product follow from the properties of $2 \times 2$ and $3 \times 3$ determinants:

1. $e_1 \times e_2 = e_3$, $e_2 \times e_3 = e_1$, $e_3 \times e_1 = e_2$;
2. $a \times a = 0$;
3. $a \times b = -(b \times a) = -b \times a$;
4. $a \times (b + c) = a \times b + a \times c$;
5. $(\lambda a) \times (\mu b) = \lambda \mu (a \times b)$;
6. $\langle a, b \times c \rangle = \begin{vmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{vmatrix} = \langle a \times b, c \rangle$;
7. $\langle a, a \times b \rangle = 0 = \langle b, a \times b \rangle$;
8. $\|a \times b\| = \sqrt{\|a\|^2 \|b\|^2 - (\langle a, b \rangle)^2}$;
9. For non-zero vectors $a, b$, if $\alpha$ is the angle between $a$ and $b$, then
   \[
   \sin \alpha = \frac{\|a \times b\|}{\|a\| \cdot \|b\|}.
   \]
Properties of the Cross-Product

- In particular, $a \times b$ is perpendicular on both $a$ and $b$.

**Lemma 119**

If $A, B, C$ are distinct non-collinear points in $\mathbb{R}^3$, then the area of the triangle $\Delta(A, B, C)$ equals

$$\frac{1}{2} \| a b \times a c \| .$$

**Lemma 120**

If $A, B, C$ are distinct non-collinear points in $\mathbb{R}^3$, then the distance $d$ from $C$ to the line through $A$ and $B$ is given by

$$d = \frac{|| a b \times a c ||}{|| a b ||}.$$
Transformations

- Linear Maps
- Coordinate Transformations in $\mathbb{R}^2$
- Coordinate Transformations in $\mathbb{R}^3$
- Classification of Transformations
- Transformation of Coordinate Systems
- Applications of Coordinate (System) Transformations
- Projections
Linear Maps

Definition 121 (Linear map, Dt.: lineare Abbildung)

Let $V, W$ be vector spaces over $\mathbb{R}$, e.g., $V := \mathbb{R}^n$ and $W := \mathbb{R}^m$. A map $g : V \rightarrow W$ is called a linear map if

1. $g(v_1 + v_2) = g(v_1) + g(v_2)$ $\forall v_1, v_2 \in V,$
2. $g(\lambda v) = \lambda g(v)$ $\forall v \in V, \forall \lambda \in \mathbb{R}.$

Lemma 122

Let $e_1, \ldots, e_n$ be a basis of $V$, and $e'_1, \ldots, e'_m$ be a basis of $W$. A linear map $g : V \rightarrow W$ is uniquely determined by the images of the basis vectors $e_j$ relative to $e'_i$: For $v := \sum_{j=1}^{n} v_j e_j$ and $w := \sum_{i=1}^{m} w_i e'_i$, with $w = g(v)$, we get

$$w = g(v) = g\left(\sum_{j=1}^{n} v_j e_j\right) = \sum_{j=1}^{n} v_j g(e_j) = \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{m} a_{ij} e'_i\right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} v_j\right) e'_i = A v,$$

where $A := [a_{ij}]_{i=1,j=1}^{m,n}$.

- Every linear map has a corresponding matrix transformation, and it maps a linear combination of vectors to the linear combination of the images of the vectors.
Linear Maps

- Sample linear maps in \( \mathbb{R}^2 \): rotation about origin, stretching, reflection (about coordinate axis or origin), shear transformation.
- Note: Translation is not linear!

**Lemma 123**

If a linear map has an inverse map then the inverse map is also linear.

**Lemma 124**

If a linear map \( g \) has an inverse linear map then the matrix which corresponds to \( g \) is invertible.
Composition of Linear Maps

Definition 125 (Composition, Dt.: Zusammensetzung)
Consider two linear maps $g : U \rightarrow V$ and $h : V \rightarrow W$. The composition $h \circ g$ is a map from $U$ to $W$ such that every $u \in U$ is mapped to $h(g(u)) \in W$.

Warning
There is absolutely no consensus in the literature on whether $(h \circ g)(x)$ shall mean $h(g(x))$ or $g(h(x))$!

Lemma 126
The composition of two linear maps is a linear map.
Combining Matrix Transformations

- Suppose that $P'$ is obtained by applying the matrix transformation $T_1$ to $P$, and $P''$ is obtained from $P'$ via $T_2$, and so on till $P^{(n)}$:

  \[
  \begin{pmatrix}
  x' \\
  y'
  \end{pmatrix} = T_1 \cdot \begin{pmatrix}
  x \\
  y
  \end{pmatrix} \wedge \begin{pmatrix}
  x'' \\
  y''
  \end{pmatrix} = T_2 \cdot \begin{pmatrix}
  x' \\
  y'
  \end{pmatrix} \wedge \ldots \wedge \begin{pmatrix}
  x^{(n)} \\
  y^{(n)}
  \end{pmatrix} = T_n \cdot \begin{pmatrix}
  x^{(n-1)} \\
  y^{(n-1)}
  \end{pmatrix}.
  \]

  Then the dependence of $P^{(n)}$ on $P$ can be expressed as

  \[
  \begin{pmatrix}
  x^{(n)} \\
  y^{(n)}
  \end{pmatrix} = T_n \cdot \left( T_{n-1} \cdot \ldots \left( T_2 \cdot \left( T_1 \cdot \begin{pmatrix}
  x \\
  y
  \end{pmatrix} \right) \right) \right) =
  \]

  \[
  = (T_n \cdot T_{n-1} \cdot \ldots \cdot T_2 \cdot T_1) \cdot \begin{pmatrix}
  x \\
  y
  \end{pmatrix} =
  \]

  \[
  = T \cdot \begin{pmatrix}
  x \\
  y
  \end{pmatrix},
  \]

  where $T := T_n \cdot T_{n-1} \cdot \ldots \cdot T_2 \cdot T_1$.

- Note the order of the matrix multiplications!
- Recall that matrix multiplication is associative but not commutative!
Order of Transformations Matters

- $T$: Translate by $(5, 0)$;
- $R$: Rotate about origin by $\pi/4$.  

\[
\begin{align*}
T &: \text{Translate by } (5, 0) \\
R &: \text{Rotate about origin by } \pi/4.
\end{align*}
\]
Rotation in $\mathbb{R}^2$

Rotation of $P$ by $\theta$ about the origin yields $P'$.

Polar coordinates: $x = r \cos \varphi, \quad y = r \sin \varphi$.

\[
x' &= r \cos(\theta + \varphi) \\
    &= r \cos \theta \cos \varphi - r \sin \theta \sin \varphi \\
    &= x \cos \theta - y \sin \theta.
\]

\[
y' &= r \sin(\theta + \varphi) \\
    &= r \sin \theta \cos \varphi + r \cos \theta \sin \varphi \\
    &= x \sin \theta + y \cos \theta.
\]
Rotation as a Matrix Transformation

- We have

\[
\begin{pmatrix}
  x' \\
y'
\end{pmatrix} = \begin{pmatrix}
x \cos \theta - y \sin \theta \\
x \sin \theta + y \cos \theta
\end{pmatrix}
\]

for a rotation about the origin by the angle \( \theta \).

- This relation can also be expressed by means of a rotation matrix \( \text{Rot}(\theta) \):

\[
\begin{pmatrix}
  x' \\
y'
\end{pmatrix} = \begin{pmatrix}
  \cos \theta & -\sin \theta \\
  \sin \theta & \cos \theta
\end{pmatrix} \cdot \begin{pmatrix}
x \\
y
\end{pmatrix} = \text{Rot}(\theta) \cdot \begin{pmatrix}
x \\
y
\end{pmatrix};
\]

that is,

\[
\text{Rot}(\theta) := \begin{pmatrix}
  \cos \theta & -\sin \theta \\
  \sin \theta & \cos \theta
\end{pmatrix}.
\]

**Lemma 127**

Rotation matrices are orthogonal: \( \text{Rot}(\theta)^{-1} = \text{Rot}(\theta)^t \).
General Rotation in $\mathbb{R}^2$

- Rotation of $P$ by $\theta$ about $A$, with $a := \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, yields $P'$.

\[
\begin{align*}
 x &= \alpha + r \cos \varphi \quad \Rightarrow \quad r \cos \varphi = x - \alpha \\
 y &= \beta + r \sin \varphi \quad \Rightarrow \quad r \sin \varphi = y - \beta.
\end{align*}
\]

\[
\begin{align*}
 x' &= \alpha + r \cos(\theta + \varphi) \\
 &= \alpha + r \cos \theta \cos \varphi - r \sin \theta \sin \varphi \\
 &= \alpha + (x - \alpha) \cos \theta - (y - \beta) \sin \theta.
\end{align*}
\]

\[
\begin{align*}
 y' &= \beta + (x - \alpha) \sin \theta + (y - \beta) \cos \theta.
\end{align*}
\]
Stretching in $\mathbb{R}^2$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \underbrace{\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}}_{\mathbf{s}(\lambda_1, \lambda_2)} \cdot \begin{pmatrix} x \\ y \end{pmatrix}.$$ 

- If $\lambda_1 = \lambda_2$: (uniform) scaling;
- If $\lambda_1 \neq \lambda_2$: non-uniform scaling or stretching.

If $\lambda_1 = 2$ and $\lambda_2 = 2$:

- Uniform scaling.

If $\lambda_1 = 2$ and $\lambda_2 = 1$:

- Non-uniform scaling or stretching.
Shear Transformation in $\mathbb{R}^2$

- Suppose that we want to map a point $P$ to a point $P'$ such that

$$x' = x + ay \quad \text{and} \quad y' = y.$$ 

Hence, a horizontal segment at height $y$ is shifted in the $x$-direction by $ay$.

- The corresponding transformation matrix is given by

$$\text{SH}_x(a) = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}.$$
Reflection in $\mathbb{R}^2$

- Reflection about $x$-axis:
  \[
  \begin{pmatrix}
  x' \\
  y'
  \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}.
  \]

- Reflection about $y$-axis:
  \[
  \begin{pmatrix}
  x' \\
  y'
  \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}.
  \]

- Reflection about origin:
  \[
  \begin{pmatrix}
  x' \\
  y'
  \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix};
  \]
  that is, a reflection about the origin is identical to a rotation about the origin by $\pi$. 
Translation in $\mathbb{R}^2$

- Translation: Move a point $P$ along a vector $t$ from its original location $p$ to its new location $p'$.

$$ p = \begin{pmatrix} x \\ y \end{pmatrix}, \quad t = \begin{pmatrix} t_x \\ t_y \end{pmatrix}, \quad p' = \begin{pmatrix} x' \\ y' \end{pmatrix}. $$

$$ x' = x + t_x, \quad y' = y + t_y, \quad p' = p + t. $$

$$ \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}. $$
Translating a Rigid Body in $\mathbb{R}^2$

- Translate every point of $\Delta$ by $t$:
  \[
  \Delta' = \{ p + t : p \in \Delta \}. 
  \]
- For polygons and polytopes it suffices to translate the vertices.
Translation as a Matrix Transformation

Question
What is the matrix of a translation?

Answer
No $n \times n$ matrix is the matrix of a (non-trivial) translation in $\mathbb{R}^n$!

Why? Since the fixed point set of every matrix transformation includes the origin, but the origin is not invariant under a translation.

We will resort to homogeneous coordinates, which is a concept borrowed from projective geometry.
Homogeneous Coordinates: Motivation

- A rational number $\frac{x}{y}$ is an equivalence class of appropriate pairs $(x', y')$.

- $2 \simeq (2, 1), (4, 2), \ldots$

- $1/3 \simeq (1/3, 1), (1, 3), (2, 6), \ldots$

- Not a unique representation: All points on a line through the origin represent the same rational number.

- $(x_1, y_1) \cdot (x_2, y_2) = (x_1x_2, y_1y_2)$  
  $(x_1, y_1)/(x_2, y_2) = (x_1y_2, y_1x_2)$.

- Infinity does not need to be treated separately:  
  $\infty \simeq (1, 0), (2, 0), \ldots$
Homogeneous Coordinates in $\mathbb{R}^2$

Definition 128 (Homogeneous coordinates, Dt.: homogene Koordinaten)

The *homogeneous coordinates* of \( \begin{pmatrix} x \\ y \end{pmatrix} \) are given by \( \begin{pmatrix} w \cdot x \\ w \cdot y \\ w \end{pmatrix} \), for \( w \neq 0 \), while the *inhomogeneous coordinates* of \( \begin{pmatrix} x \\ y \\ w \end{pmatrix} \) are given by \( \begin{pmatrix} x/w \\ y/w \end{pmatrix} \).

- Thus, \( \begin{pmatrix} u \\ v \\ w \end{pmatrix} \) are the homogeneous coordinates of \( \begin{pmatrix} x \\ y \end{pmatrix} \), and \( \begin{pmatrix} x \\ y \end{pmatrix} \) are the inhomogeneous coordinates of \( \begin{pmatrix} u \\ v \\ w \end{pmatrix} \).

\[ \iff x = \frac{u}{w} \text{ and } y = \frac{v}{w}. \]

- We will find it convenient to assume \( w = 1 \).
Homogeneous Coordinates in $\mathbb{R}^2$

- We identify the point $\left( \begin{array}{c} x \\ y \end{array} \right) \in \mathbb{R}^2$ with $\left( \begin{array}{c} x \\ y \\ 1 \end{array} \right) \in \mathbb{R}^3$.

- Thus, $\mathbb{R}^2$ is embedded into $\mathbb{R}^3$ by identifying it with the plane $z = 1$. 
Homogeneous Coordinates in $\mathbb{R}^2$

- All points on a line through the origin in $\mathbb{R}^3$ represent the same point in $\mathbb{R}^2$.
- $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ can be regarded as the point at infinity on the line through $\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$.
- Homogeneous coordinates allow us to express translation, rotation and scaling in $\mathbb{R}^2$ by means of one $3 \times 3$ transformation matrix.
- Homogeneous coordinates support scaling in a natural way, and build the basis of projective geometry.
- In particular, note that the plane $z = 1$ of $\mathbb{R}^3$ is invariant under matrix transformations of the form

$$
\begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
0 & 0 & 1
\end{pmatrix}.
$$
Transformation Matrices Based on Homogeneous Coordinates for $\mathbb{R}^2$

Translation:

\[
\begin{pmatrix}
    x' \\
    y' \\
    w'
\end{pmatrix} =
\begin{pmatrix}
    1 & 0 & t_x \\
    0 & 1 & t_y \\
    0 & 0 & 1
\end{pmatrix}
\cdot
\begin{pmatrix}
    x \\
    y \\
    w
\end{pmatrix}.
\]

We get $\text{Trans}(t_x, t_y)^{-1} = \text{Trans}(-t_x, -t_y)$.

Stretching:

\[
\begin{pmatrix}
    x' \\
    y' \\
    w'
\end{pmatrix} =
\begin{pmatrix}
    \lambda_1 & 0 & 0 \\
    0 & \lambda_2 & 0 \\
    0 & 0 & 1
\end{pmatrix}
\cdot
\begin{pmatrix}
    x \\
    y \\
    w
\end{pmatrix}.
\]

We get $\text{S}(\lambda_1, \lambda_2)^{-1} = \text{S}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}\right)$. 
Rotation:

\[
\begin{pmatrix}
x' \\ y' \\ w'
\end{pmatrix} = \begin{pmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{pmatrix} \cdot \begin{pmatrix}
x \\ y \\ w
\end{pmatrix}.
\]

We get \( \text{Rot}(\theta)^{-1} = \text{Rot}(-\theta) = \text{Rot}(\theta)^t \).

- Rotation involves either trigonometric functions or square roots.
- Power series may be used to approximate the terms of a rotation matrix for small values of \( \theta \).
Homogeneous Coordinates and Transformations in $\mathbb{R}^3$

- Homogeneous coordinates in $\mathbb{R}^3$:

  \[(x, y, z, w) \simeq \left( \frac{x}{w}, \frac{y}{w}, \frac{z}{w} \right).\]

- For a right-hand coordinate system the positive (CCW) rotation about a coordinate axis is defined as follows:
  - Look along axis towards origin from $+\infty$;
  - Counter-clockwise rotation about axis by angle $\pi/2$ transforms one axis to another, obeying the cyclic order $x \rightarrow y \rightarrow z \rightarrow x$. 

![Diagram of coordinate system with rotations]

© M. Held (Univ. Salzburg)  Geometrisches Rechnen  (WS 2013/14)
Rotation about $z$-Axis

A rotation about the $z$-axis can be regarded as a rotation in $\mathbb{R}^2$ about the origin that is extended to $\mathbb{R}^3$. That is,

\[
\begin{align*}
x' &= x \cos \theta - y \sin \theta, \\
y' &= x \sin \theta + y \cos \theta, \\
z' &= z.
\end{align*}
\]
Rotation about $x$-Axis

Rotation about the $x$-axis: Substitute $x \to y, y \to z, z \to x$ in the equations for the rotation about $z$.

\[
y' = y \cos \theta - z \sin \theta,
\]
\[
z' = y \sin \theta + z \cos \theta,
\]
\[
x' = x.
\]
Rotation about $y$-Axis

Similarly for a rotation about the $y$-axis: Substitute $x \rightarrow y$, $y \rightarrow z$, $z \rightarrow x$ in the previous equations.

\[
z' = z \cos \theta - x \sin \theta,
\]
\[
x' = z \sin \theta + x \cos \theta,
\]
\[
y' = y.
\]
**Transformation Matrices for $\mathbb{R}^3$**

- **Rotation (about x-Axis):**
  
  $\begin{pmatrix} x' \\ y' \\ z' \\ w' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi & 0 \\ 0 & \sin \phi & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$

- **Rotation (about y-Axis):**
  
  $\begin{pmatrix} x' \\ y' \\ z' \\ w' \end{pmatrix} = \begin{pmatrix} \cos \phi & 0 & \sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$

- **Rotation (about z-Axis):**
  
  $\begin{pmatrix} x' \\ y' \\ z' \\ w' \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi & 0 & 0 \\ \sin \phi & \cos \phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$
Transformation Matrices for $\mathbb{R}^3$

Translation:

$$
\begin{pmatrix}
    x' \\
    y' \\
    z' \\
    w'
\end{pmatrix} = \begin{pmatrix}
    1 & 0 & 0 & t_x \\
    0 & 1 & 0 & t_y \\
    0 & 0 & 1 & t_z \\
    0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
    x \\
    y \\
    z \\
    w
\end{pmatrix}
$$

Stretching/Scaling:

$$
\begin{pmatrix}
    x' \\
    y' \\
    z' \\
    w'
\end{pmatrix} = \begin{pmatrix}
    \lambda_1 & 0 & 0 & 0 \\
    0 & \lambda_2 & 0 & 0 \\
    0 & 0 & \lambda_3 & 0 \\
    0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
    x \\
    y \\
    z \\
    w
\end{pmatrix}
$$
Quaternions and Rotation

- A unit quaternion \( Q := (\cos \phi, u \sin \phi) \) represents the rotation of a point, \( P \), in \( \mathbb{R}^3 \) by the angle \( 2\phi \) about the axis \( u \) through the origin, with \( \|u\| = 1 \):
  - Let \( P := (0, p) \). We get \( P' := QPQ^{-1} \), with \( P' = (0, p') \).
  - Let \( s := \cos \phi \), \( v := u \sin \phi \). Then

\[
p' = s^2 p + \langle p, v \rangle v + 2s (v \times p) + v \times (v \times p).
\]

- Thus, there is a one-to-one correspondence between unit quaternions and rotations in \( \mathbb{R}^3 \).

- Suppose that we are given two unit quaternions \( Q_0, Q_1 \) and would like to interpolate the rotations specified by these quaternions linearly.

- Recall that a unit quaternion can be regarded as a point on the unit sphere in \( \mathbb{R}^4 \).

- Hence, a natural approach to a linear interpolation of two quaternions is a spherical interpolation along the great circle between \( Q_0 \) and \( Q_1 \).
Consider a transformation matrix $T$, a plane $\varepsilon$: $ax + by + cz + d = 0$, and let $\xi := (x, y, z, 1)^t$ and $A := (a, b, c, d)$.

- Thus, the plane equation is given by $\varepsilon$: $A \cdot \xi = 0$.
- The equation of the transformed plane is $\varepsilon': (A \cdot T^{-1}) \cdot \xi' = 0$, with $\xi' = T \cdot \xi$, since

$$0 = A \cdot \xi = A \cdot (T^{-1} \cdot T) \cdot \xi = (A \cdot T^{-1}) \cdot (T \cdot \xi) = (A \cdot T^{-1}) \cdot \xi'.$$

- A line in $\mathbb{R}^3$ can be regarded as the intersection of two planes, and, thus, be transformed by transforming those two planes.
Transformation of Quadrics in Equational Form

- Consider a transformation matrix $\mathbf{T}$, and a quadric $\mathbf{x}^t \cdot \mathbf{Q} \cdot \mathbf{x} = 0$.
- We get

$$0 = \mathbf{x}^t \mathbf{Q} \mathbf{x} = (\mathbf{T}^{-1} \cdot \mathbf{T} \cdot \mathbf{x})^t \cdot \mathbf{Q} \cdot \mathbf{T}^{-1} \cdot \mathbf{T} \cdot \mathbf{x}$$

$$= \mathbf{x}^t \mathbf{T}^t \cdot (\mathbf{T}^t)^{-1} \cdot \mathbf{Q} \cdot \mathbf{T}^{-1} \cdot \mathbf{T} \cdot \mathbf{x}$$

$$= \mathbf{x}^t \cdot ((\mathbf{T}^t)^{-1} \cdot \mathbf{Q} \cdot \mathbf{T}^{-1}) \cdot \mathbf{x}' .$$

- Thus, the matrix of the transformed quadric is $(\mathbf{T}^t)^{-1} \cdot \mathbf{Q} \cdot \mathbf{T}^{-1}$, where $(\mathbf{T}^t)^{-1} = \mathbf{T}$ if $\mathbf{T}$ is orthonormal.
- In order to transform more general curves and surfaces, it is common to compute a linear approximation and transform its vertices.
Classification of Transformations

- Consider a mapping $g : \mathbb{R}^n \to \mathbb{R}^n$ and a distance metric $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.
- E.g., take $n = 2$ and the standard Euclidean distance

$$d(P, Q) := \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}.$$  

**Definition 129 (Isometry, Dt.: Isometrie)**

A mapping $g$ is called an *isometry* (or *rigid motion*, Dt.: *Bewegung*) if it maps pairs of points to points the same distance apart. That is,

$$\forall (P, Q \in \mathbb{R}^n) \quad d(g(P), g(Q)) = d(P, Q).$$

- Another widely-used term for characterizing an isometry is *distance-preserving transformation*.
- In planar Euclidean geometry such a mapping is also called a *congruence*.
- E.g., two triangles which are congruent have corresponding sides of equal length.
Classification of Transformations

**Definition 130 (Similarity mapping, Dt.: Ähnlichkeitsabbildung)**

A mapping $g$ is called a *similarity mapping* if it preserves angles.

- E.g., two triangles which are similar have identical angles, and their sides are "in proportion".

**Lemma 131**

A distance-preserving transformation is a similarity mapping, i.e., it preserves angles.

**Definition 132 (Orthogonal transformation, Dt.: orthogonale Transformation)**

A linear mapping that preserves distance is called *orthogonal transformation*. (And the class of all such transformations on $\mathbb{R}^n$ forms the *orthogonal group* of $\mathbb{R}^n$.)

**Lemma 133**

The group of all isometries on $\mathbb{R}^n$ is given by composites of a translation and an orthogonal transformation.
Classification of Transformations

**Lemma 134**

With respect to an orthonormal basis of $\mathbb{R}^n$, an orthogonal transformation has a corresponding *orthogonal matrix*, i.e., a matrix whose columns and rows are orthonormal vectors.

**Corollary 135**

An orthogonal transformation is invertible: If its matrix is $A$ then the inverse transformation has matrix $A^t$. Furthermore, $\det A = \pm 1$.

**Lemma 136**

A $2 \times 2$ orthogonal matrix $A$ is the matrix of a rotation about the origin if and only if $\det A = 1$. If $\det A = -1$ then it is the matrix of a reflection.

**Lemma 137**

A $3 \times 3$ orthogonal matrix $A$ is the matrix of a rotation about a straight line through the origin if and only if $\det A = 1$. 
Classification of Transformations

**Definition 138 (Affine transformation, Dt.: affine Abbildung)**

A mapping $g$ is called *affine transformation* (or *affinity*) if it is a composite of a translation and a linear transformation.

- Affine transformations need not preserve distance, angle, area or volume.

**Lemma 139**

If $g$ is an affine transformation and $P, Q, R$ are collinear, then $g(P), g(Q), g(R)$ are collinear. That is, affine transformations preserve collinearity.

**Corollary 140**

An affine transformation maps parallel lines to parallel lines.

**Lemma 141**

An affine transformation preserves ratios of lengths of intervals on any line.
Group Hierarchy of Transformations

- **Orthogonal**
- **Isometry**
- **Similarity**
- **Affine**
- **Projective**

These transformations are nested, with each category being a subset of the one above it.
Transformation of Coordinate Systems

- What are the coordinates \( P_{C'} = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \) of a point \( P_C = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \) relative to a new coordinate system \( C' \)?

- We assume that the mapping from \( P_C \) to \( P_{C'} \) is an isometry, and consider the matrix

\[
T_C := \begin{pmatrix} e'_1 & e'_2 & e'_3 & \delta \\ 0 & 0 & 0 & 1 \end{pmatrix},
\]

where \( e'_1 \) represents the unit vector of the \( x' \)-axis of \( C' \) in terms of \( C \). Analogously for \( e'_2, e'_3 \). Hence, \([e'_1, e'_2, e'_3]\) is an orthogonal matrix if \( e_1, e_2, e_3 \) are orthonormal.
Transformation of Coordinate Systems

- We have
\[
\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \xrightarrow{T_C} \begin{pmatrix} e_1' + \delta \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \xrightarrow{T_C} \begin{pmatrix} e_2' + \delta \\ 1 \end{pmatrix},
\]
\[
\begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \xrightarrow{T_C} \begin{pmatrix} e_3' + \delta \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \xrightarrow{T_C} \begin{pmatrix} \delta \\ 1 \end{pmatrix},
\]

that is
\[
\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} \xrightarrow{T_C} \begin{pmatrix} x' e_1' + y' e_2' + z' e_3' + \delta \\ 1 \end{pmatrix} =: \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}_C.
\]

- We understand that the coordinates of a point specified relative to \(C'\) are converted by \(T_C\) to coordinates relative to \(C\):

**Theorem 142**

With \(T_C\) as defined above, we get
\[
P_C = T_C \cdot P_{C'} \quad \text{and} \quad P_{C'} = T_C^{-1} \cdot P_C.
\]
Inverse Transformation

If $T$ is the matrix of an isometry then, by Lem 133,

$$T = \begin{pmatrix}
1 & 0 & 0 & \varphi \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \cdot \begin{pmatrix}
\begin{array}{c}
R \\
0 \\
0 \\
0
\end{array} & 0 \\
0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix},$$

where $R$ is an orthogonal matrix, and $\varphi$ describes the translation.

Since $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$, we get

$$T^{-1} = \begin{pmatrix}
\begin{array}{c}
R^{-1} \\
0 \\
0 \\
0
\end{array} & 0 \\
0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \cdot \begin{pmatrix}
\begin{array}{c}
1 \\
0 \\
0 \\
0
\end{array} & 0 \\
0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}^{-1}.$$

Since $R$ is orthogonal, we have $R^{-1} = R^t$ and get

$$T^{-1} = \begin{pmatrix}
\begin{array}{c}
R^t \\
0 \\
0 \\
0
\end{array} & 0 \\
0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \cdot \begin{pmatrix}
\begin{array}{c}
1 \\
0 \\
0 \\
0
\end{array} & 0 \\
0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}.$$

Inverse Transformation

Theorem 143

If \([n, o, a]\) is orthogonal then we get

\[
T^{-1} = \begin{pmatrix}
    n_x & n_y & n_z & -\langle d, n \rangle \\
    o_x & o_y & o_z & -\langle d, o \rangle \\
    a_x & a_y & a_z & -\langle d, a \rangle \\
    0 & 0 & 0 & 1
\end{pmatrix}
\]

for

\[
T := \begin{pmatrix}
    n_x & o_x & a_x & d_x \\
    n_y & o_y & a_y & d_y \\
    n_z & o_z & a_z & d_z \\
    0 & 0 & 0 & 1
\end{pmatrix}.
\]

- Recall that the matrix of a general affine transformation is not orthogonal!
Rotating About a General Axis

- What is the matrix of the rotation about a general axis $u$ (through the origin) by an angle $\phi$?

![Diagram of rotation about a general axis]

- Assume that $||u|| = 1$. We set up a new frame $C' = (e'_1, e'_2, e'_3)$ such that
  - $O = O'$,
  - $e'_3 = u$,
  - $\langle e'_2, e'_3 \rangle = 0 \land ||e'_2|| = 1$,
  - $e'_1 := e'_2 \times e'_3$,

and consider the transformation matrix $T := \begin{pmatrix} e'_1 & e'_2 & e'_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.
Rotating About a General Axis

We know:

\[
\begin{pmatrix}
  x' \\
  y' \\
  z' \\
  1
\end{pmatrix}
= T^{-1} \cdot 
\begin{pmatrix}
  x \\
  y \\
  z \\
  1
\end{pmatrix}.
\]

Thus, we get the following decomposition for \( \text{Rot}(u, \phi) \):

\[
\text{Rot}(u, \phi) = T \cdot 
\begin{pmatrix}
  \cos \phi & -\sin \phi & 0 & 0 \\
  \sin \phi & \cos \phi & 0 & 0 \\
  0 & 0 & 1 & 0 \\
  0 & 0 & 0 & 1
\end{pmatrix} \cdot T^{-1}.
\]

Simple algebraic operations yield:

\[
\text{Rot}(u, \phi) := 
\begin{pmatrix}
  u_x u_x \, \text{vers} \, \phi + \cos \phi & u_y u_x \, \text{vers} \, \phi - u_z \sin \phi & u_z u_x \, \text{vers} \, \phi + u_y \sin \phi & 0 \\
  u_x u_y \, \text{vers} \, \phi + u_z \sin \phi & u_y u_y \, \text{vers} \, \phi + \cos \phi & u_z u_y \, \text{vers} \, \phi - u_x \sin \phi & 0 \\
  u_x u_z \, \text{vers} \, \phi - u_y \sin \phi & u_y u_z \, \text{vers} \, \phi + u_x \sin \phi & u_z u_z \, \text{vers} \, \phi + \cos \phi & 0 \\
  0 & 0 & 0 & 1
\end{pmatrix}
\]

where \( \text{vers} \, \phi := 1 - \cos \phi \).
Rotating About a General Axis

- Given an (orthogonal) rotation matrix $T$, how can we find an axis $u$ through the origin and an angle $\phi$ such that $\text{Rot}(u, \phi) = T$?

  $$\text{Rot}(u, \phi) \overset{?}{=} T := \begin{pmatrix} n_x & o_x & a_x & 0 \\ n_y & o_y & a_y & 0 \\ n_z & o_z & a_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$ 

- Some calculations yield

  $$\tan \phi = \frac{\sqrt{(o_z - a_y)^2 + (a_x - n_z)^2 + (n_y - o_x)^2}}{n_x + o_y + a_z - 1},$$

  which defines $\phi$ within $[0, \pi]$. 

Rotating About a General Axis

Furthermore,

\[ u_x = \text{sign}(o_z - a_y) \sqrt{\frac{n_x - \cos \phi}{1 - \cos \phi}}, \]

\[ u_y = \text{sign}(a_x - n_z) \sqrt{\frac{o_y - \cos \phi}{1 - \cos \phi}}, \]

\[ u_z = \text{sign}(n_y - o_x) \sqrt{\frac{a_z - \cos \phi}{1 - \cos \phi}}. \]
Local Coordinate Systems

- Typically, objects are not modeled in world coordinates. Rather, *local coordinate systems* are used.
- In order to transform the object it suffices to fix the position and orientation of the local coordinate system relative to the world coordinate system, or relative to some other system.
Kinematics

- We consider an articulated mechanism that consists of rigid links connected by joints.
- Every joint connects exactly two links, and describes the motion of one link relative to the other link.
- The most important joints are prismatic and rotatory joints.
A mechanism can be represented as a graph, a so-called *kinematic chain*, where
- the links form the nodes, and
- the joints from the edges.

A mechanism is called an *open kinematic chain* if this graph has no cycles; *closed kinematic chain*, otherwise.

Depending on how detailed a human is modeled, a human skeleton represents either an open or a closed kinematic chain.
Local Coordinate Frames

- It is common to assign two local coordinate frames $F_{i1}$ and $F_{i2}$ to link $i$ such that
  - the $z$-axis coincides with the joint axis,
  - the $x$-axis coincides with the link axis, and
  - the $y$-axis is chosen appropriately to form a right-handed frame.
Denavit-Hartenberg Parameters

- Find a transformation matrix $i^{-1}A$ to express a point of $F_{i,2}$ in terms of $F_{i-1,2}$.

**A-Matrix**:

$$ i^{-1}A := \text{Rot}(z, \theta) \cdot \text{Trans}(0, 0, d) \cdot \text{Trans}(a, 0, 0) \cdot \text{Rot}(x, \alpha) $$

$$ = \begin{pmatrix}
\cos \theta & -\sin \theta \cos \alpha & \sin \theta \sin \alpha & a \cos \theta \\
\sin \theta & \cos \theta \cos \alpha & -\cos \theta \sin \alpha & a \sin \theta \\
0 & \sin \alpha & \cos \alpha & d \\
0 & 0 & 0 & 1
\end{pmatrix}, $$

where $a \ldots$ link length,
$\alpha \ldots$ link twist,
$d \ldots$ link offset,
$\theta \ldots$ link angle,

Denavit-Hartenberg Parameters.
Forward and Inverse Kinematics

Forward Kinematics:
- Given: joint vector.
- Compute: Frame $\mathbf{T}$ of the end-effector relative to the base frame.
- Solution:

$$\mathbf{T} = \mathbf{0A} \cdot \mathbf{1A} \cdot \ldots \cdot \mathbf{n-1A}.$$ 

Inverse Kinematics:
- Given: Frame $\mathbf{T}$ of the end-effector relative to the base frame.
- Compute: all admissible joint vectors.
- Solution: not trivial, requires solving a set of non-linear equations!
  Symbolic solution preferred over numerical solution.
Inverse Kinematics

- Truly all admissible joint vectors have to be computed!
Projections

- Virtually all output devices are two-dimensional.
- To draw a 3D scene, the scene has to be projected onto a 2D viewing plane.
Projections: History

- Plan from Mesopotamia, $\approx 2000$BC.
- Early Greeks: Agatharchus ($\approx 500$BC), Apollonius of Perga ($\approx 262$BC till $\approx 190$BC) studied projections of quadrics.
- Romans: Vitruvius wrote *De Architectura*, published specifications of plan and elevation drawings, and perspective.
- Early Renaissance period: Emphasis on point of view, interpretation of world.
  - Dürer
  - Giotto,
  - Mossacio,
  - Raphael,
  - Vinci.

- **Leon Battista Alberti** wrote the first treatise on perspective, “Della Pittura”, in 1435.
  
  “A painting is the intersection of a visual pyramid at a given distance, with a fixed center and a definite position of light, represented by art with lines and colors on a given surface.”
Camera Set-Up

- Given a scene, we have to specify how we want to view the scene:
  - **Camera Position**: Point $E$.
  - **Object Point at Center of Window**: Point $R$. Also called “to”-point.
  - **Center of View Plane**: Point $O$.
  - **“Up” direction**: Vector $V$; must not be parallel to the line through $E$ and $R$.

We may also need to specify the viewing angle, and other camera-related values.
Geometric Projections

- **Projection plane**: Plane $\Pi$.
- **Projectors**: Rays emanating from the center of projection and passing through points of the object.
- **Projection**: Intersection of projectors with plane $\Pi$.
- Non-geometric projections used in cartography.
Different Types of Geometric Projections

- **Perspective:**
  - Center of projection is at a finite distance from \( \Pi \).
  - *Perspective foreshortening* occurs.

- **Parallel:**
  - Center of Projection is at \( \infty \).
  - Defined by the direction \((x, y, z, 0)\).
  - Directions correspond to points at infinity.
Different Types of Geometric Projections

Planar geometric projection

Parallel
- Orthographic
  - Top (plan)
  - Front elevation
  - Side elevation
- Oblique
  - Cabinet
  - Axonometric
    - Isometric
- Other

Perspective
- One-point
- Two-point
- Three-point

Other elevation
- Side elevation
- Other
When formulating the mathematics of projections it is customary to place the viewpoint at \((0, 0, -d)\), in the case of a perspective projection, and to assume that the projection plane \(\Pi\) is the \(xy\)-plane.
Perspective Projection

- Perspective foreshortening gives a realistic view of 3D objects.
- Used for advertising, fine art, architecture.
- Foreshortening is not uniform.
- Parallel edges do not remain parallel; angles, scales and other geometric properties are not preserved.
- A *vanishing point* (Dt.: Fluchtpunkt) is a point in the image plane where the projections of lines parallel to one of the three principal axes of the object converge: one-point perspective, two-point perspective, or three-point perspective.
One Vanishing Point

- Π parallel to two principal axes of the cube: one vanishing point.
Two Vanishing Points

- Π is parallel to only one principal axis of the cube: two vanishing points.
Three Vanishing Points

- $\Pi$ is not parallel to any principal axis of the cube: three vanishing points.
Due to the similarity of the triangles $\triangle(Z, O, P'_{xz})$ and $\triangle(Z, P_z, P_{xz})$ we get

$$x' : d = x : (z + d), \quad \text{i.e.,} \quad x' = \frac{d \cdot x}{z + d}.$$

Analogously,

$$y' = \frac{d \cdot y}{z + d}.$$
Matrix of a Perspective Projection

- Let $P := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{d} & 1 \end{pmatrix}$.

- We get
  \[
  P \cdot p = \begin{pmatrix} p_x \\ p_y \\ 0 \\ \frac{p_z + d}{d} \end{pmatrix} \cong \begin{pmatrix} \frac{d \cdot p_x}{p_z + d} \\ \frac{d \cdot p_y}{p_z + d} \\ 0 \\ 1 \end{pmatrix} =: \begin{pmatrix} p'_x \\ p'_y \\ 0 \\ 1 \end{pmatrix}.
  \]

- Use transformation of coordinate system if the projection plane differs from $z = 0$, or if the eye point is not at $(0, 0, -d')$. 
Parallel Projection: Orthographic

- **Orthographic**: Projectors are perpendicular to the projection plane.

\[
\rightarrow P_{xy} := \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}.
\]

- **Front, top, side views**: Projectors parallel to one of the principal axes.
**Parallel Projection: Oblique**

- **Oblique:** Projectors not perpendicular to the projection plane.
- **Isometric:** Projectors make equal angle with each axis.

**Dimetric:** Equal angle with two axes.

**Trimetric:** Distinct angle with all three axes.
Mathematics of Oblique Projection

- With $\tan \beta = \frac{1}{d}$ we get
  \[
  x' = x + z \cdot d \cos \alpha, \\
  y' = y + z \cdot d \sin \alpha, \\
  z' = 0.
  \]
- Thus,
  \[
  P := \begin{pmatrix}
  1 & 0 & d \cos \alpha & 0 \\
  0 & 1 & d \sin \alpha & 0 \\
  0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 1
  \end{pmatrix}.
  \]
Special Oblique Projections

- **Cavalier projection:**
  - Angle $\beta$ between projectors and projection plane is $45^\circ$; i.e., $d = 1$.
  - The length of a segment normal to the projection plane equals the length of the projection of the segment.

- **Cabinet projection:**
  - Angle $\beta$ between projectors and projection plane is $\tan^{-1} 2 \approx 63.4^\circ$; i.e., $d = \frac{1}{2}$.
  - The length of a segment normal to the projection plane equals twice the length of the projection of the segment.
Projecting Curved Objects

- It does not suffice to project the vertices and edges of an object if the object is bounded by curved surfaces.

- Rather, we also have to project the silhouette curves of the object.
- A silhouette curve consists of all those points of the object such that the line through the point and the center of projection is tangential to the object.
- Note that the silhouette curves need not lie in one plane!
Perspective Normalization

- For computing silhouette curves, hidden-surface elimination, ray tracing, and many other algorithms, it is convenient to transform the view pyramid into a view box, while maintaining the depth ordering.

- Consider \( \mathbf{N} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{d} & 1 \end{pmatrix} \).

- We get

\[
\mathbf{N} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}; \quad \mathbf{N} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}; \quad \mathbf{N} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix};
\]

that is, the \( xy \)-plane is invariant under \( \mathbf{N} \).

- The center of projection is mapped to the point at infinity on the negative \( z \)-axis:

\[
\mathbf{N} \begin{pmatrix} 0 \\ 0 \\ \frac{-d}{1} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \frac{-d}{0} \\ 0 \end{pmatrix}.
\]
Perspective Normalization

- Summarizing, we get

\[ \mathbf{O} \cdot \mathbf{N} = \mathbf{P}, \]

where \( \mathbf{O} \) is the matrix of an orthogonal projection, and \( \mathbf{P} \) is the matrix of the corresponding perspective projection.

- \( \mathbf{N} \) maps
  - cylinder, cone \( \rightarrow \) cylinder or cone (possibly with non-circular cross-section),
  - line \( \rightarrow \) line,
  - plane \( \rightarrow \) plane,
  - sphere \( \rightarrow \) ellipsoid, elliptical paraboloid, two-sheet hyperboloid,
  - quadric \( \rightarrow \) quadric.

- We can modify \( \mathbf{N} \) such that all z-coordinates are scaled to lie between 0 and 1.
Geometric Objects

- Lines and Planes
- Circles and Spheres
- Curves and Surfaces
- Polygons and Polyhedra
Definition 144 (Straight line, Dt.: Gerade)

For two distinct points $P, Q \in \mathbb{R}^n$, the straight line defined by $P, Q$ is the set

$$\ell(P, Q) := \{p + \lambda pq : \lambda \in \mathbb{R}\}.$$

Lemma 145

For every pair of distinct points $P, Q \in \mathbb{R}^2$, there exist $n \in \mathbb{R}^2$ and $c \in \mathbb{R}$ such that

$$\ell(P, Q) = \{u \in \mathbb{R}^2 : \langle u, n \rangle = c\}.$$ 

- Note that $\langle n, pq \rangle = 0$ holds for every such $n$. That is, the vector $n$ is a normal vector of $\ell(P, Q)$. We have

$$n = \nu \begin{pmatrix} -pq_y \\ pq_x \end{pmatrix}$$

for some non-zero scalar $\nu \in \mathbb{R}$. 
Lines and Straight-Line Segments

**Definition 146 (Ray, Dt.: Strahl, Halbgerade)**
For two distinct points $P, Q \in \mathbb{R}^n$, the ray starting at $P$ through $Q$ is the set

$$\{p + \lambda pq : \lambda \in \mathbb{R}_0^+\}.$$

**Definition 147 (Straight-line segment, Dt.: Geradensegment, Strecke)**
For two distinct points $P, Q \in \mathbb{R}^n$, the (closed) straight-line segment defined by $P, Q$ is the set

$$\overline{PQ} := \{p + \lambda pq : \lambda \in [0, 1]\}.$$

**Definition 148 (Open straight-line segment)**
For two distinct points $P, Q \in \mathbb{R}^n$, the open straight-line segment defined by $P, Q$ is the set

$$\{p + \lambda pq : \lambda \in ]0, 1[\}.$$
**Lines in \( \mathbb{R}^2 \)**

**Definition 149 (Hessian normal form, Dt.: Hessische Normalform)**

A line equation is said to be in *Hessian normal form* if \( n \) (as specified in Lem. 145) is a unit vector.

**Lemma 150**

The (signed) minimum distance \( d \) of a point \( A \in \mathbb{R}^2 \) from \( \ell(P, Q) \), with \( \ell(P, Q) = \{ u \in \mathbb{R}^2 : \langle u, n \rangle = c \} \), is given by

\[
d = \frac{\langle a, n \rangle - c}{\|n\|}.
\]
Planes in $\mathbb{R}^3$

**Definition 151 (Plane, Dt.: Ebene)**

For three distinct and non-collinear points $P, Q, R \in \mathbb{R}^3$, the *plane* defined by $P, Q, R$ is the set

$$\mathcal{E}(P, Q, R) := \{ p + \lambda pq + \mu pr : \lambda, \mu \in \mathbb{R} \}.$$ 

**Lemma 152**

For every triple of distinct and non-collinear points $P, Q, R \in \mathbb{R}^3$, there exist $n \in \mathbb{R}^3$ and $c \in \mathbb{R}$ such that

$$\mathcal{E}(P, Q, R) = \{ u \in \mathbb{R}^3 : \langle u, n \rangle = c \}.$$ 

- Note that $\langle n, pq \rangle = \langle n, pr \rangle = 0$ holds for every such $n$. That is, the vector $n$ is a normal vector of $\mathcal{E}(P, Q, R)$. We have

$$n = \nu(pq \times pr) \quad \text{for some} \quad \nu \in \mathbb{R} \setminus \{0\}.$$
Planes in $\mathbb{R}^3$

**Definition 153 (Hessian normal form, Dt.: Hessische Normalform)**

A plane equation is said to be in *Hessian normal form* if $\mathbf{n}$ (as specified in Lem. 152) is a unit vector.

**Lemma 154**

The (signed) minimum distance $d$ of a point $A \in \mathbb{R}^3$ from $\varepsilon(P, Q, R)$, with $\varepsilon(P, Q, R) = \{u \in \mathbb{R}^3 : \langle u, \mathbf{n} \rangle = c\}$, is given by

$$d = \frac{\langle a, \mathbf{n} \rangle - c}{\|\mathbf{n}\|}.$$
Determinants and Lines/Planes

Lemma 155

The equation of the line through two distinct points $P$ and $Q$ in $\mathbb{R}^2$ is given by

$$\det \begin{pmatrix} x & y & 1 \\ p_x & p_y & 1 \\ q_x & q_y & 1 \end{pmatrix} = 0.$$ 

Lemma 156

The equation of the plane through three distinct and non-collinear points $P$, $Q$, $R$ in $\mathbb{R}^3$ is given by

$$\det \begin{pmatrix} x & y & z & 1 \\ p_x & p_y & p_z & 1 \\ q_x & q_y & q_z & 1 \\ r_x & r_y & r_z & 1 \end{pmatrix} = 0.$$
Half-Plane and Half-Space

- The line \( \ell(P,Q) = \{u \in \mathbb{R}^2 : \langle u, n \rangle = c \} \) partitions \( \mathbb{R}^2 \) into three disjoint sets: the actual line and the two (open) half-planes \( \{u \in \mathbb{R}^2 : \langle u, n \rangle - c < 0 \} \) and \( \{u \in \mathbb{R}^2 : \langle u, n \rangle - c > 0 \} \).

- Similarly for a plane in \( \mathbb{R}^3 \) and half-spaces.
Circles in $\mathbb{R}^2$ and Spheres in $\mathbb{R}^3$

**Definition 157 (Sphere, Dt.: Sphäre, Kugeloberfläche)**

A (hyper)-sphere in $\mathbb{R}^n$ centered at point $C$ with radius $r$ is the set

$$S(C, r) := \{ u \in \mathbb{R}^n : \|u - c\| = r \}.$$  

Conventionally, a hyper-sphere is called a *circle* in $\mathbb{R}^2$ and a *sphere* in $\mathbb{R}^3$.

- The equational form of a hyper-sphere (in the $L_2$-norm) can be re-written as

  $$(u_1 - c_1)^2 + (u_2 - c_2)^2 + \cdots + (u_n - c_n)^2 = r^2.$$  

**Definition 158 (Disk, Dt.: Kreisscheibe)**

A (closed) disk in $\mathbb{R}^2$ (with center $C$ and radius $r$) is the set

$$\{ u \in \mathbb{R}^2 : \|u - c\| \leq r \}.$$  

**Definition 159 (Open disk)**

An open disk in $\mathbb{R}^2$ (with center $C$ and radius $r$) is the set

$$\{ u \in \mathbb{R}^2 : \|u - c\| < r \}.$$
Circles in $\mathbb{R}^2$ and Spheres in $\mathbb{R}^3$

**Definition 160 (Ball, Dt.: Kugel)**

A (closed) ball in $\mathbb{R}^3$ (with center $C$ and radius $r$) is the set

$$B(C, r) := \{u \in \mathbb{R}^3 : \|u - c\| \leq r\}.$$  

**Definition 161 (Open ball)**

An open ball in $\mathbb{R}^3$ (with center $C$ and radius $r$) is the set

$$\{u \in \mathbb{R}^3 : \|u - c\| < r\}.$$  

- Of course, these definitions can be generalized to norms other than the $L_2$-norm.
- Warning: In mathematics, a terminological distinction is made between a sphere, which is a two-dimensional closed surface embedded in $\mathbb{R}^3$, and a ball, which is a shape ("solid") in $\mathbb{R}^3$ that includes the interior of its associated sphere. It is common to blur this distinction between a sphere and its enclosed set of points in application fields (e.g., graphics).
Curves

- Intuitively, a curve in $\mathbb{R}^2$ is generated by a continuous motion of a pencil on a sheet of paper.
- A formal mathematical definition is not entirely straightforward, and the term “curve” is associated with two closely related notions: kinematic and geometric.
- In the kinematic setting, a (parameterized) curve is a function of one real variable.
- In the geometric setting, a curve, also called an arc, is a 1-dimensional subset of space.
- Both notions are related: the image of a parameterized curve describes an arc. Conversely, an arc admits a parameterization.
- Since the kinematic setting is easier to introduce, we resort to a kinematic definition of “curve”.
- Note that fairly counter-intuitive curves exist: e.g., space-filling curves like the Sierpinski curve.
Sierpinski Curves

- Sierpinski curves are a sequence of recursively defined continuous and closed curves in $\mathbb{R}^2$.
- Sierpinski curve of order 1:

Their limit curve, *the Sierpinski curve*, is a space-filling curve: It fills the unit square completely!
Sierpinski Curves

- Sierpinski curves are a sequence of recursively defined continuous and closed curves in $\mathbb{R}^2$.
- Sierpinski curve of order 2:

![Sierpinski Curve of Order 2](image_url)

- Their limit curve, the Sierpinski curve, is a space-filling curve: It fills the unit square completely!
Sierpinski Curves

- Sierpinski curves are a sequence of recursively defined continuous and closed curves in $\mathbb{R}^2$.
- Sierpinski curve of order 3:

  Their limit curve, the Sierpinski curve, is a space-filling curve: It fills the unit square completely!
Sierpinski Curves

- Sierpinski curves are a sequence of recursively defined continuous and closed curves in $\mathbb{R}^2$.
- Sierpinski curve of order 4:

Their limit curve, *the Sierpinski curve*, is a space-filling curve: It fills the unit square completely!
Curves in $\mathbb{R}^n$

**Definition 162 (Curve, Dt.: Kurve)**

Let $I \subseteq \mathbb{R}$ be an interval of the real line. A continuous (vector-valued) mapping $\gamma: I \rightarrow \mathbb{R}^n$ is called a *parameterization* of $\gamma(I)$, and $\gamma(I)$ is called the (parametric) *curve* parameterized by $\gamma$.

- Well-known examples of parameterized curves include a straight-line segment, a circular arc, and a helix.
- E.g., $\gamma: [0, 1] \rightarrow \mathbb{R}^3$ with

  $$\gamma(t) := \begin{pmatrix} p_x + t \cdot (q_x - p_x) \\ p_y + t \cdot (q_y - p_y) \\ p_z + t \cdot (q_z - p_z) \end{pmatrix}$$

  maps $[0, 1]$ to a straight-line segment from point $P$ to $Q$.

- Sometimes the interval $I$ is called the *domain* of $\gamma$, and $\gamma(I)$ is called *image*.

- In daily math, the standard meaning of a “curve” is the image of the equivalence class of all paths under a certain equivalence relation. (Roughly, two paths are equivalent if they are identical up to re-parameterization.)

- For simplicity, we will not distinguish between a curve $C$ and one of its parameterizations $\gamma$ if the meaning is clear.
Curves in $\mathbb{R}^n$

**Definition 163 (Start and end point)**
If $I$ is a closed interval $[a, b]$, for some $a, b \in \mathbb{R}$, then we call $\gamma(a)$ the start point and $\gamma(b)$ the end point of the curve $\gamma: I \rightarrow \mathbb{R}^n$.

**Definition 164 (Closed, Dt.: geschlossen)**
A parameterization $\gamma: I \rightarrow \mathbb{R}^n$ is said to be closed (or a loop) if $I$ is a closed interval $[a, b]$, for some $a, b \in \mathbb{R}$, and $\gamma(a) = \gamma(b)$.

**Definition 165 (Simple, Dt.: einfach)**
A parameterization $\gamma: I \rightarrow \mathbb{R}^n$ is said to be simple if $\gamma(t_1) = \gamma(t_2)$ for $t_1 \neq t_2 \in I$ implies \{t_1, t_2\} = \{a, b\} and $I = [a, b]$, for some $a, b \in \mathbb{R}$.

- Hence, if $\gamma: I \rightarrow \mathbb{R}^n$ is simple then it is injective on int($I$).
Curves in $\mathbb{R}^n$

**Definition 166 (Tangent vector, Dt.: Tangentenvektor)**

Consider a differentiable parameterization $\gamma : I \to \mathbb{R}^n$ of a curve $C$. For $t \in I$, a tangent vector at $\gamma(t)$ with respect to $\gamma$ is given by $\gamma'(t)$.

- Note that $\gamma'(t)$ is a vector-valued function!
- It is straightforward to extend the definition of a tangent vector to parameterizations that are piecewise differentiable.
Curves in $\mathbb{R}^2$

Definition 167 (Jordan curve, Dt.: Jordankurve)

A set $C \subset \mathbb{R}^2$ (which is not a single point) is called a Jordan curve if there exists a simple and closed parameterization $\gamma : I \rightarrow \mathbb{R}^2$ that parameterizes $C$.

Theorem 168 (Jordan 1887)

Every Jordan curve $C$ partitions $\mathbb{R}^2 \setminus C$ into two disjoint open regions, a (bounded) “interior” region and an (unbounded) “exterior” region, with $C$ as the (topological) boundary of both of them.

- Although this theorem – the so-called Jordan Curve Theorem (Dt.: Jordanscher Kurvensatz) – seems obvious, a proof is not entirely trivial unless we restrict the theorem to special classes of curves, such as polygons.
Surfaces in $\mathbb{R}^3$

**Definition 169 (Parametric surface)**

Let $\Omega \subseteq \mathbb{R}^2$. A continuous mapping $\alpha : \Omega \rightarrow \mathbb{R}^3$ is called a *parameterization* of $\alpha(\Omega)$, and $\alpha(\Omega)$ is called the (parametric) *surface* parameterized by $\alpha$.

- For instance, every point on the surface of Earth can be described by the geographic coordinates longitude and latitude.

- Note that parameterizations of a surface (regarded as a set $S \subset \mathbb{R}^3$) need not be unique: two different parameterizations $\alpha$ and $\beta$ may exist such that $S = \alpha(\Omega_1) = \beta(\Omega_2)$.

- For simplicity, we will not distinguish between a surface and one of its parameterizations if the meaning is clear.
**Lemma 170**

Consider a differentiable parameterization $\alpha : \Omega \rightarrow \mathbb{R}^3$ of a surface $S$. For $(s, t) \in \Omega$, tangent vectors at $\alpha(s, t)$ with respect to $\alpha$ are given by $\frac{\partial \alpha}{\partial s}(s, t)$ and $\frac{\partial \alpha}{\partial t}(s, t)$.

**Definition 171 (Normal vector, Dt.: Normalvektor)**

Consider a differentiable parameterization $\alpha : \Omega \rightarrow \mathbb{R}^3$ of a surface $S$. A normal vector $n_{\alpha}(s, t)$ at $\alpha(s, t)$ with respect to $\alpha$ is given by

$$n_{\alpha}(s, t) := \frac{\partial \alpha}{\partial s}(s, t) \times \frac{\partial \alpha}{\partial t}(s, t).$$

- The vector $n_{\alpha}(s, t)$ is indeed a normal vector of the tangential plane at $\alpha(s, t)$. 
Standard Parameterization of Circles and Spheres

- A circle in $\mathbb{R}^2$ (with center $C$ and radius $r$) can be parameterized as follows:
  $$ (c_x + r \cos \phi, c_y + r \sin \phi) \quad \text{with} \quad \phi \in [0, 2\pi[. $$

- A sphere in $\mathbb{R}^3$ (with center $C$ and radius $r$) can be parameterized as follows:
  $$ (c_x + r \cos \delta \cos \phi, c_y + r \cos \delta \sin \phi, c_z + r \sin \delta) \quad \text{with} \quad \phi \in [0, 2\pi[, \delta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. $$
Polygons

Definition 172 (Polygonal curve, Dt.: Polygonzug)

Consider the sequence of points $V_0, V_1, V_2, \ldots, V_n \in \mathbb{R}^d$, for some $n \in \mathbb{N}$. The polygonal curve (or polygonal chain, polygonal profile) specified by these points (“vertices”) is given by

$$
\bigcup_{i=0}^{n-1} V_iV_{i+1}.
$$

Hence, a polygonal curve is a sequence of finitely many vertices connected by line segments such that each segment (except for the first) starts at the end of the previous segment.

Unless specified otherwise, we will always assume that all vertices of a polygonal curve are co-planar.

A polygonal curve with vertices $V_0, V_1, V_2, \ldots, V_n \in \mathbb{R}^d$ is commonly denoted by $(V_0, V_1, \ldots, V_{n-1}, V_n)$. 
Definition 173 (Polygon)

A polygon is a polygonal curve with vertices $V_0, V_1, V_2, \ldots, V_n \in \mathbb{R}^d$ such that $V_0 = V_n$.

- A polygon with vertices $V_0, V_1, V_2, \ldots, V_n \in \mathbb{R}^d$, with $V_0 = V_n$, is commonly called an $n$-gon.
- Note that a polygon need not form a simple curve!
Theorem 174 (Meister (1769), Gauß (1795))

Consider a simple planar polygon $\mathcal{P} := (V_0, V_1, V_2, \ldots, V_n)$, and pick a point $T$ in the plane. Then the (signed) area of $\mathcal{P}$ is given by the sum of the signed areas of the individual triangles $\Delta(T, V_i, V_{i+1})$. That is, the (signed) area of $\mathcal{P}$ equals

$$\sum_{i=0}^{n-1} A_\Delta(T, V_i, V_{i+1}) = \frac{1}{2} \cdot [(x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + \cdots + (x_ny_1 - x_1y_n)],$$

where $v_i := \begin{pmatrix} x_i \\ y_i \end{pmatrix}$. The signed area of $\mathcal{P}$ is positive if and only if $\mathcal{P}$ is oriented counter-clockwise (CCW).
Polyhedra

- In solid modeling, a \textit{solid} describes a closed object that could exist in $\mathbb{R}^3$ and, at least theoretically, be manufactured. Well-known examples of solids include balls, cubes, cylinders, and cones.

- A \textit{polyhedron} describes a solid that is bounded by planar faces formed by polygons, joined at their edges, such that every edge is shared by an even number of faces.

- In order to guarantee that (the surface of) the polyhedron is water-tight and forms a 2-manifold it is common to demand that
  1. every edge of every face belongs to exactly one other face, and
  2. the faces that share a vertex form a cyclic chain of polygons in which every pair of consecutive polygons shares an edge.

  The first requirement precludes T-junctions, while the second requirement precludes the case of two pyramids touching at a vertex.

- Note that the word “polyhedron” has slightly different meanings in solid modeling and graphics, on one hand, and combinatorial geometry and algebraic topology, on the other hand.
Basic Concepts of Topology
- Convexity and Connectedness
- Metric Space
- Topological Properties of Sets
- Topological Properties of Surfaces and Solids
- Triangulations
Convexity

Definition 175 (Convex)
A set \( S \subseteq \mathbb{R}^n \) is called \emph{convex} if for all \( P, Q \in S \)
\[ \overline{PQ} \subseteq S. \]

Definition 176 (Convex superset)
A set \( T \subseteq \mathbb{R}^n \) is called a \emph{convex superset} of a set \( S \subseteq \mathbb{R}^n \) if
\[ S \subseteq T \quad \text{and} \quad T \text{ is convex}. \]

Definition 177 (Convex hull)
A set \( T \) is called a \emph{convex hull} of a set \( S \) if
\[ T \text{ is the smallest convex superset of } S. \]
The convex hull of \( S \) is often denoted by \( CH(S) \).
Convexity

Lemma 178
For $S \subseteq \mathbb{R}^n$, the convex hull $CH(S)$ is formed by the intersection of all convex supersets of $S$.

Lemma 179
For a discrete set $S \subseteq \mathbb{R}^n$ of points, $CH(S)$ is given by all convex combinations of tuples of points of $S$. 
Connectedness

**Definition 180 (Path-connected, Dt.: Weg zusammenhängend)**

A set \( S \subset \mathbb{R}^n \) is *path-connected* if for every pair of points \( P, Q \in S \) there exists a curve that is completely contained in \( S \) and that links \( P \) and \( Q \).

**Definition 181 (Simply-connected and multiply-connected)**

A path-connected set \( S \subset \mathbb{R}^2 \) is *simply-connected* if every simple closed curve entirely contained within \( S \) encloses only points of \( S \). Otherwise, \( S \) is called *multiply-connected* (or not *simply-connected*).
Metric Space

Definition 182 (Metric space, Dt.: metrischer Raum)

A *metric space* is a set of points $\mathcal{X}$ with an associated distance function (aka *metric*) $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that the following conditions hold for all $x, y, z \in \mathcal{X}$:

1. $d(x, y) \geq 0$.
2. Identity of indiscernibles: $d(x, y) = 0 \Rightarrow x = y$.
3. Reflexivity: $d(x, x) = 0$.
4. Symmetry: $d(x, y) = d(y, x)$.
5. Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

- Easy to check: $\mathbb{R}^n$ with the standard Euclidean distance is a metric space.
- Easy to check: every normed vector space is a metric space by defining $d(x, y) := \|x - y\|$.
Definition 183 (Open ball, Dt.: offene Kugel)

Consider a metric space $\mathcal{X}$ with metric $d$. For $x \in \mathcal{X}$ and $r \in \mathbb{R}_0^+$ we define the (generalized) open ball (relative to the metric $d$) with radius $r$ centered at $x$ as

$$B(x, r) := \{ y \in \mathcal{X} : d(x, y) < r \}.$$
Consider a space $\mathcal{X}$ that has a metric, and a set $S \subseteq \mathcal{X}$. (E.g., $\mathbb{R}^d$ and the Euclidean metric, and any subset $S$ of $\mathbb{R}^d$.)

**Definition 184 (Interior point, Dt.: innerer Punkt)**

A point $x \in \mathcal{X}$ is an *interior point* of $S$ if there exists a radius $r > 0$ such that the open ball with center $x$ and radius $r$ is completely contained in $S$, i.e., $B(x, r) \subseteq S$.

**Definition 185 (Interior, Dt.: Inneres)**

The set of all interior points of $S$ is the *interior* of $S$, often denoted by $\text{int}(S)$ or $S^\circ$.

**Lemma 186**

We have $\text{int}(S) \subseteq S$ for all $S \subseteq \mathcal{X}$.

**Lemma 187**

For all $x \in \mathcal{X}$, the interior of an open ball $B(x, r) \subseteq \mathcal{X}$ is the open ball itself.
Interior, Exterior and Closure

**Definition 188 (Exterior point, Dt.: äußerer Punkt)**

A point \( y \in \mathcal{X} \) is an *exterior point* of \( S \) if there exists a radius \( r > 0 \) such that the open ball with center \( y \) and radius \( r \) is completely contained in the complement of \( S \) (with respect to \( \mathcal{X} \)), i.e., \( B(y, r) \subseteq (\mathcal{X} \setminus S) \).

**Definition 189 (Exterior, Dt.: Äußeres)**

The set of all exterior points of \( S \) is the *exterior* of \( S \), denoted by \( \text{ext}(S) \).

**Definition 190 (Boundary, Dt.: Rand)**

All points of \( \mathcal{X} \) that are neither in the interior nor in the exterior of \( S \) form the *boundary*, \( \partial S \), of \( S \).
In the figure (relative to the standard Euclidean distance in $\mathbb{R}^2$), $A$ is an interior point, $B$ is on the boundary, and $C$ is an exterior point.

Lemma 191

For all $S \subseteq \mathcal{X}$, the union of the interior, the exterior and the boundary of $S$ constitutes the whole space $\mathcal{X}$. 
Interior, Exterior and Closure

**Definition 192 (Closure, Dt.: Abschluß)**

The *closure* $\overline{S}$ of a set $S$ is the union of the interior and the boundary of $S$.

**Lemma 193**

The closure $\overline{S}$ of a set $S$ is given by all points that are not in the exterior of $S$.

**Definition 194 (Open, Dt.: offen)**

A set $S \subseteq \mathcal{X}$ is called *open* if $\text{int}(S) = S$.

**Definition 195 (Closed, Dt.: abgeschlossen)**

A set $S \subseteq \mathcal{X}$ is called *closed* if the complement of $S$ is open.

- Note that there exist $\mathcal{X}$ and subsets $S \subseteq \mathcal{X}$ such that the interior or the exterior or the boundary of $S$ are empty.
- Warning: Intuition may easily misguide one’s judgement once general spaces or metrics are studied!
Interior, Exterior and Closure

- Consider a ball in $\mathbb{E}^3$ with radius $r$ centered at the origin:
  \[
  \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq r^2 \}.
  \]
- The interior of the ball is
  \[
  \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 < r^2 \}.
  \]
- The closure of the ball is
  \[
  \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq r^2 \}.
  \]
- The exterior of the ball is
  \[
  \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 > r^2 \}.
  \]
- The boundary of the ball is
  \[
  \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = r^2 \}.
  \]
Manifolds

- Informally speaking, 2-manifolds are surfaces in 3D that are locally two-dimensional, i.e., that locally (at each point of the manifold) resemble a “bent copy of a rubber plane”.

**Definition 196 (Manifold, Dt.: Mannigfaltigkeit)**

A set $S \subseteq \mathbb{R}^3$ is a 2-*manifold* (or simply a “manifold”) if for every point $x \in S$ there exists an open subset of $S$ that contains $x$ which is homeomorphic to an open disk.

- A homeomorphism is a bijective function between two spaces that is continuous and that also has a continuous inverse; it establishes a “topological equivalence” between the spaces.
**Orientable Surface**

**Definition 197 (Orientable, Dt.: orientierbar)**

A 2-manifold is *orientable* if a unit normal vector can be defined for every point on the surface such that it varies continuously over the surface.

- Gluing the ends of a strip of paper together after a twist yields a one-sided surface called a *Möbius band*, which is not *orientable*. 

![Diagram of a Möbius band]
The topologically simplest connected 2-manifold in 3D is the surface of a sphere. By adding a “handle” to the sphere we get a torus. It is well-known that every manifold surface can be obtained by adding a certain number of handles to the sphere.

**Definition 198 (Genus, Dt.: Geschlecht)**

A connected orientable manifold surface is said to have *genus* $k$ if it can be cut along $k$ non-intersecting closed simple curves without causing the resultant manifold to become disconnected.

- Equivalently, a manifold of genus $k$ can be obtained by adding $k$ handles to the sphere.
- Note that a general surface can also be obtained by “punching holes” through a sphere.
- However, it is not difficult to see that, topologically, adding a handle is equivalent to opening a hole on a surface.
**Definition 199 (Triangulation)**

Let $S = \{P_1, P_2, \ldots, P_k\}$ be a set of $k$ points in $\mathbb{R}^2$. A structure $T$ is called a *triangulation* of $S$ if $T$ is (the embedding of) a connected planar graph such that:

- $S$ is (the embedding of) the vertex set of $T$,
- all bounded faces of $T$ are triangles,
- the union of the bounded triangular faces forms the convex hull of $S$. 

![Diagram of a triangulation with points and lines connecting them to form triangles]
Constrained Triangulation

**Definition 200 (Constrained triangulation)**

Let $S = \{P_1, P_2, \ldots, P_k\}$ be a set of $k$ points in $\mathbb{R}^2$, and $E$ be a set of line segments that link points of $S$ and that do not intersect pairwise except at common end points. A structure $T$ is called a *constrained triangulation* of $S$ if $T$ is (the embedding of) a connected planar graph such that

- $S$ is (the embedding of) the vertex set of $T$,
- all segments of $E$ are edges of $T$,
- all bounded faces of $T$ are triangles,
- the union of the bounded triangular faces forms the convex hull of $S$. 

![Diagram of constrained triangulation](image)
Floating-Point Arithmetic and Numerical Mathematics

- Floating-Point Computations
- Iterative Algorithms for Solving Non-Linear Equations
- Iterative Algorithms for Solving Linear Equations
- Numerical Integration
Floating-Point Computations

- Virtually all modern computers employ floating-point (fp) arithmetic to perform real arithmetic. Typically, the IEEE-754 standard is used, which demands 32-bit ‘single’ precision and 64-bit ‘double’ precision representations.
- No matter how many bits are used, fp-arithmetic represents a number by a fixed-length binary mantissa and an exponent of fixed size.
- Thus, only a finite number of values within a finite sub-interval of \( \mathbb{R} \) can be represented accurately; all other values have to be rounded to the closest number that is representable.
- It is well-known that \( \frac{1}{3} \) cannot be represented by a finite sum of powers of 10. Similarly, 0.1 cannot be represented by a finite sum of powers of 2!
- More generally, there are two sources of error for fp-computations: input error and round-off error.

**Input error:** It is well-known that \( \frac{1}{3} \) cannot be represented by a finite sum of powers of 10. Similarly, 0.1 cannot be represented by a finite sum of powers of 2!

**Round-off error:** It arises from rounding results of fp-computations during an algorithm. E.g., \( \sqrt{2} \) cannot be represented exactly since \( \sqrt{2} \) is an irrational number.
Real-World Example of Round-Off Error

- During the first Gulf war, an Iraqi Scud got through the Patriot anti-missile system (AMS) and hit a barracks, killing 28 people.

- To track the Scud, the system had to determine the interval between tracking times by subtracting two values of a timer. The times in tenths of a second were stored in integer registers; a stored value of 35 would be equivalent to 3.5 seconds.

- To compute the interval, the values in the registers were converted to fp-representation by multiplying them by 0.1.

- As stated previously, 0.1 has a non-terminating binary expansion. Consequently, the time interval was computed with error.

- The larger the value in the timer, the larger the error.

- At the time of the incident, the AMS had been operating for over 100 hours, resulting in an error of 0.34 seconds in the timer, causing the system to look in the wrong place for the incoming Scud.
Machine Precision

- The round-off error is bounded in terms of the *machine precision*, $\varepsilon$, which is the smallest value satisfying

$$|fp(a \circ b) - (a \circ b)| \leq \varepsilon |a \circ b|$$

for all all fp-numbers $a, b$ and any of the four operations $+, -, \cdot, /$ instead of $\circ$, for which $a \circ b$ does not cause an underflow or an overflow.

- On IEEE-754 machines, $\varepsilon = 2^{-23} \approx 1.19 \cdot 10^{-7}$ for floats, and $\varepsilon = 2^{-52} \approx 2.22 \cdot 10^{-16}$ for doubles.

- On some exotic platform, $\varepsilon$ can be determined approximately by finding the smallest positive value $x$ such that $1 + x \neq 1$.

- Note: Some compilers promote floats to doubles!

- Note: Some platforms employ extended representations, or use registers longer than standard words for intermediate results! The sad truth is that hardware vendors still prefer to stick to their own standards . . .
Floating-Point Comparisons and Precision Thresholds

- Comparing two fp-numbers $x$ and $y$ by means of $x = y$ will hardly ever yield true.
- Rather, one has to check whether $|x - y| \leq \varepsilon$ for some precision threshold $\varepsilon$, denoted by $x =_{\varepsilon} y$.
- Note that there is a difference between checking for “$\leq$” and checking for “$<$”.
- Caveat: $=_{\varepsilon}$ is no longer transitive!
- Note: fp-numbers are “denser” close to zero than far away from zero.
- Note: $|x - y| \leq \varepsilon$ need not imply $|\alpha \cdot x - \alpha \cdot y| \leq \varepsilon$.
- Thus, use relative errors or scale the data appropriately.
- Obvious disadvantage of scaling: Unless only shifts by two are performed, new errors may be introduced.
Common Manifestations of Floating-Point Errors

- Cancellation: Subtracting two numbers of almost equal magnitudes may cause a drastic loss in the number of significant digits.

- Accumulation: Adding 0.001 for 1000000 times need not yield exactly 1000. (DP-arithmetic on my PC: 999.999999832650701.)

- Adding/subtracting small and large numbers: the un-normalizing required to line up the decimal point may cause truncation. E.g., adding $2^{40} = 1099511627776$ and $2^{-14} = 0.0000610352$ yields 1099511627776 (DP-arithmetic).

- Overflow: Occurs if (absolute) value of a number is too large to be represented. The evaluation of an expression that results in an overflow will raise an error flag; the actual value of the expression is positive or negative “Inf”.

- Underflow: Occurs if (absolute) value of a number is too small to be represented. An expression that results in an underflow evaluates to zero, without returning an error.
Special Floating-Point “Numbers”

- Not a Number: 0/0 or $\sqrt{-1}$ will generate a special value, “NaN”.
- Infinity: Divisions by zero will generate positive or negative “Inf”.
- Those special numbers propagate through subsequent calculations.
Ill-Conditioned Equations

- The quartic equation $x^4 + 4x^3 + 6x^2 + 4x + 1 = 0$ has the quadruple root $x = -1$.
- Changing the coefficient of $x$ to 4.00000001 drastically affects the solution: Now we get $x = -1.01002496875 \ldots$ and $x = -0.99002503124 \ldots$ as the only real roots of the equation.
- Similarly, the linear system
  
  \[
  \begin{align*}
  x + 2y & = 3 \\
  0.48x + 0.99y & = 1.47
  \end{align*}
  \]

  has the exact solution $x = 1, y = 1$, while the system
  
  \[
  \begin{align*}
  x + 2y & = 3 \\
  0.49x + 0.99y & = 1.47
  \end{align*}
  \]

  has the exact solution $x = 3, y = 0$. Note, however, that the old solution, $x = 1, y = 1$, also “nearly” fulfills this equation.
Ill-Conditioned Equations

- Thus, a small change in the coefficients can dramatically affect the solutions of an equation!
- An equation or a system of equations is said to be *ill-conditioned* if the solutions are very sensitive to small changes — and therefore also to errors! — in the coefficients and constants. Aka “ill-posed” problem.
- An $n \times n$ system of linear equations is ill-conditioned if and only if the determinant of the coefficient matrix is close to zero.
- Note: If an equation (or a system of equations) is ill-conditioned, then the usual procedure of checking a numerical solution by calculation of the residuals may not be valid.
Ill-Conditioned Equations and Residual

- If an equation (or a system of equations) is ill-conditioned, then the usual procedure of checking a numerical solution by calculation of the residuals may not be valid.

- Consider the $2 \times 2$ linear system

  
  \[
  \begin{align*}
  1.2969x + 0.8648y &= 0.8642 \\
  0.2161x + 0.1441y &= 0.1440
  \end{align*}
  \]

  that is, $A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$.

- The exact solution is $x = 2$ and $y = -2$.

- But we get close-to-zero residuals also for other pairs of $x$ and $y$:

  \[
  \begin{align*}
  x_1 &= 0.9911 \\
  y_1 &= -0.4870 \\
  A \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} &\approx 10^{-8}
  \end{align*}
  \]

  \[
  \begin{align*}
  x_2 &= 2.001557851 \\
  y_2 &= -2.002336236 \\
  A \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} &\approx 10^{-10}
  \end{align*}
  \]

  \[
  \begin{align*}
  x_3 &= -0.000004626 \\
  y_3 &= 0.999312976 \\
  A \begin{pmatrix} x_3 \\ y_3 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} &\approx 10^{-9}
  \end{align*}
  \]
Improving the Reliability of fp-Calculations

- As far as possible, perform numerical computations relative to the original input data.
- Make sure that different calls of the same function with the “same” input will yield exactly the same output. E.g., guaranteeing
  \[ \det(P, Q, R) = \det(Q, R, P) = -\det(Q, P, R) \]
  is a must (for points \( P, Q, R \in \mathbb{R}^2 \)).
- Do not resort to multiple precision thresholds.
- At most two thresholds: One to avoid divisions by zero, and possibly another threshold to catch “nearly zero” numbers.
- Take a close look at your calculations: Different terms might be arithmetically identical, but their numerical behavior may be substantially different, and one term may be far better than the other!
Different terms might be arithmetically identical, but their numerical behavior may be substantially different, and one term may be far better than the other!

Mathematics tells us that the solutions of the quadratic equation $ax^2 + bx + c = 0$ are given by

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$ 

Unfortunately, using this formula means begging for troubles if $|a \cdot c|$ is small compared to $|b|$, since the subtraction of $\sqrt{b^2 - 4ac}$ from $b$ may cause serious cancellation.

Better: Let

$$\Delta := -\frac{1}{2}(b + \text{sign}(b)\sqrt{b^2 - 4ac}).$$

Then the roots are obtained more reliably as

$$x_1 = \frac{\Delta}{a} \quad \text{and} \quad x_2 = \frac{c}{\Delta}. \quad \text{(This is a consequence of Viète’s formulas.)}$$
Iterative Algorithms for Solving Non-Linear Equations

- We are interested in solving the equation $f(x) = 0$, for a function $f : \mathbb{R} \to \mathbb{R}$. This means finding all $x \in \mathbb{R}$ for which $f(x) = 0$.
- Explicit (algebraic) root-finding is possible for polynomial equations of degree less than five.
- For other types of non-linear equations, dozens of iterative methods have been proposed.
- Two basic schemes:
  - Bracketing: e.g., bisection, regula falsi;
  - Polishing: e.g., Newton-Raphson method, secant method.
- Extensions to vector-valued functions are possible.
Basics of Iterative Root Finding

- We attempt to compute a sequence \((x_k)_{k=0}^\infty\), depending on some initial value(s) \(x_0\) resp. \(x_0, x_1\) and on \(f\) and its derivatives.
- Ideally, \(\lim_{k \to \infty} x_k = x\).
- Question: How shall we find a suitable initial value \(x_0\)?
  Answer: Study the function \(f\).
- Question: What is a suitable initial value \(x_0\)?
  Answer: Whether or not \(x_0\) is suitable depends on \(f\) and on the iteration method used.
- Question: Is the iteration guaranteed to converge?
  Answer: Unfortunately, no – unless specific criteria are fulfilled.
- Question: Is the iteration guaranteed to find all roots?
  Answer: At best, an iteration method will find one root at a time.
- Question: How quickly will the iteration converge?
  Answer: This depends on the convergence rate of the iteration method, see later.

General advice
Do not use iteration methods on a function you do not know much about. In particular, \textit{do not} use an iteration method to test whether a root exists in the neighborhood of some initial value.
Basics of Iterative Root Finding

- How can we state how rapidly a sequence \((x_k)_{k=0}^{\infty}\) converges to the root \(\bar{x}\)?

**Definition 201 (Convergence rate, Dt.: Konvergenzrate)**

Let \((x_k)_{k=0}^{\infty}\) be a sequence that is used to approximate a root \(\bar{x}\), and let \(e_k := \bar{x} - x_k\) be the error of the \(k\)-th approximation \(x_k\) of \(\bar{x}\). The *convergence rate* of an iteration method is the largest exponent \(p\) such that

\[
\lim_{k \to \infty} \frac{|e_k|}{|e_{k-1}|^p} = c,
\]

for a suitable asymptotic error constant \(c \in \mathbb{R}^+\).

- If \(p = 1\) then the convergence is called *linear*.
- If \(p = 2\) then the convergence is called *quadratic*.
- If \(1 < p < 2\) then the convergence is called *super-linear*.

- Linear convergence means that the error is reduced by a constant factor per iteration, i.e., that that the number of correct digits increases by one after a constant number of iterations.
- Quadratic convergence means that the number of correct digits roughly doubles with each iteration.
Bisection

- Consider a continuous function $f : \mathbb{R} \to \mathbb{R}$, and assume that for $a, b \in \mathbb{R}$ you know $\text{sign}(f(a)) = -\text{sign}(f(b))$, with $a < b$ and $f(a) \cdot f(b) \neq 0$.

- Intermediate Value Theorem: Since we have opposite signs for $f$ at $a, b$, and $f$ is continuous, we conclude that $f$ has at least one root $\bar{x}$ in the interval $[a, b]$.

- By checking the sign of $f\left(\frac{a+b}{2}\right)$ and appropriately replacing $a$ or $b$ by $\frac{a+b}{2}$, this interval is halved at each step of the iteration:

\[
\begin{align*}
\text{if } \text{sign}(f\left(\frac{a+b}{2}\right)) &= 0 \quad \text{then } \bar{x} := \frac{a+b}{2}, \text{ stop;} \\
\text{or } \text{sign}(f\left(\frac{a+b}{2}\right)) &= \text{sign}(f(a)) \quad \text{then } a := \frac{a+b}{2}; \\
\text{or } \text{sign}(f\left(\frac{a+b}{2}\right)) &= \text{sign}(f(b)) \quad \text{then } b := \frac{a+b}{2}.
\end{align*}
\]

- Since bisection traps a root, it is guaranteed to converge. However, it needs at least three iterations to achieve one additional significant digit of the root!

- Caveat: Although several roots might exist within the interval $[a, b]$, only one root will be found.

- Caveat: Root-bracketing is not feasible for finding even-multiplicity roots.
Regula Falsi

- Aka “false position method” in some English literature.
- Rather than blindly testing \( c := \frac{a+b}{2} \), one could also compute the \( x \)-intercept of the secant through \((a, f(a))\) and \((b, f(b))\):

\[
c := b - \frac{f(b)(b - a)}{f(b) - f(a)}.
\]

- Now evaluate \( \text{sign}(f(c)) \), and keep either \( a \) or \( b \), just as with bisection.
- The regula falsi method shares with bisection the advantage of trapping a root and, thus, of always converging.
- However, it tends to converge faster than the bisection method if \( a \) and \( b \) are close together.
- This basic scheme can be improved further to achieve super-linear convergence; e.g., Brent-Dekker method or Illinois method.
Newton-Raphson Method

- Suppose that $f$ and $f'$ are continuous near a root $\bar{x}$ of $f$, and that $x_0$ is close to $\bar{x}$.
- The Newton-Raphson method is based on the approximation of a function $f$ by the straight-line tangent at $x = x_k$:
  \[ y = f(x_k) + f'(x_k)(x - x_k). \]
  An estimate $x_{k+1}$ for the root is obtained by setting $y := 0$ at $x = x_{k+1}$:
  \[ x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}. \]

- If $f'(x)$ is non-zero and $x_0$ sufficiently close to the actual root $\bar{x}$ then the Newton-Raphson method exhibits a quadratic convergence rate.
- That is, near a root the number of significant digits approximately doubles with each iteration.
- If the root is multiple then the rate of convergence may decrease to linear.
- Caveat: The Newton-Raphson method may be unstable near a horizontal asymptote or a local minimum, and might even diverge.
- Note: Global convergence is not guaranteed even for “nice” functions!
Secant Method

- If the derivative $f'(x_k)$ is too difficult to compute then the tangent may be replaced by the secant through two points $(x_{k-1}, f(x_{k-1}))$ and $(x_k, f(x_k))$:

$$x_{k+1} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})}.$$ 

- This yields a simplification of the Newton-Raphson method which is known as Secant method.

- The rate of convergence is super-linear, and, thus, slower than for the Newton-Raphson method.

- Note that two initial values $x_0, x_1$ are needed.
Iterative Algorithms for Solving Linear Equations

- Recall that finding the exact solution $x$ of the system of linear equations $A x = b$ requires $O(n^3)$ time for an $n \times n$ matrix $A$.
- A direct (and exact) solution turns out to be a waste of time if $n$ goes into the thousands or millions and if $A$ is sparse. In that case, iterative methods may be much faster than direct methods.
- Suppose that we know the exact solution: $x$.
- If we write $x$ as $x = x' + \Delta x$ then we get
  \[ A \Delta x = A x - A x' = b - A x'. \]
- Interpreting this equation as basis for an iterative formula $x^{(k+1)} = x^{(k)} + \Delta x$ yields
  \[ A(x^{(k+1)} - x^{(k)}) = b - A x^{(k)}. \]
- So far, we would have gained little, as we would still have to solve for $x^{(k+1)}$ . . .
- Idea: Replace $A$ on the left-hand side of this equation by an easily invertible matrix $B$ that is “close to” $A$. 
Iterative Algorithms for Solving Linear Equations

- We get

\[ B(x^{(k+1)} - x^{(k)}) = b - Ax^{(k)}, \]

or

\[ Bx^{(k+1)} = b - (B - A)x^{(k)}. \]

- One can formulate conditions under which the solution obtained by this iterative scheme is guaranteed to converge to the exact solution of \( Ax = b. \)

- Typical application in graphics: Iterative solution of a radiosity equation.
Jacobi Iteration

- Assume that all diagonal elements of $A$ are non-zero, and let $B$ be the diagonal matrix that contains all diagonal elements of $A$.

- Applying the iteration
  
  $$Bx^{(k+1)} = b - (B - A)x^{(k)}.$$  

  is equivalent to

  $$a_{ii} x_i^{(k+1)} = b_i - \sum_{j=1, j \neq i}^{n} a_{ij} x_j^{(k)}.$$  

- If $|a_{ii}| > \sum_{j=1, j \neq i}^{n} |a_{ij}|$, i.e., if $A$ is strictly diagonally dominant then this so-called Jacobi iteration is guaranteed to converge. (Different and less stringent conditions do also suffice.)
Gauss-Seidel Iteration

- Gauss-Seidel iteration is a modification of Jacobi iteration that can converge faster in some cases.
- Basic idea: Use the most up-to-date information available.
- If $x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_{i-1}^{(k+1)}$ are already known, then these new values can be used for the computation of $x_i^{(k+1)}$:

  \[
  a_{ii}x_i^{(k+1)} = b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}.
  \]

- Again, convergence is guaranteed if $A$ is strictly diagonally dominant.
- Tends to converge faster than Jacobi iteration, but is more difficult to parallelize.
Numerical Integration

- Suppose we want to compute an integral
  \[ I = \int_{a}^{b} f(x) \, dx. \]

- The best way to compute this integral would be to solve it analytically, and get
  \[ I = \int_{a}^{b} f(x) \, dx = F(b) - F(a), \quad \text{with } F'(x) = f(x). \]

- However, there are many functions that cannot be integrated analytically. Thus, methods for approximating the integral through *quadrature rules* of the form
  \[ \hat{I} = \sum_{i=1}^{n} \omega_i f(x_i) \]
  have been devised, which is essentially a weighted sum of samples of the function \( f \) at various points \( x_i \) using weights \( \omega_i \).

- The many different quadrature rules can be distinguished by their sampling patterns and weights.
Midpoint Rule for Numerical Integration

- We divide the interval \([a, b]\) into a fixed number \(n\) of subintervals, each of size \(h = (b - a)/n\).
- We then choose one sample point at the midpoint of each subinterval:
  \[
  \hat{I} = h \sum_{i=1}^{n} f(a + (i - 1/2)h)
  = h \left[ f(a + \frac{h}{2}) + f(a + \frac{3h}{2}) + \cdots + f(b - \frac{h}{2}) \right].
  \]

- The Midpoint Rule is exact for constant or linear functions. Otherwise, its error is bounded by \(O(n^{-2})\), provided that \(f\) has at least two continuous derivatives on \([a, b]\).
**Trapezoidal Rule for Numerical Integration**

- The trapezoidal rule is similar to the midpoint rule, except that we sample the function at the ends of each subinterval, and compute the area of a trapezoid for each subinterval.

\[
\hat{I} = \sum_{i=1}^{n} \frac{h}{2} [f(a + (i - 1)h) + f(a + ih)]
\]

\[
= h \left[ \frac{1}{2} f(a) + f(a + h) + f(a + 2h) + \cdots + f(b - h) + \frac{1}{2} f(b) \right].
\]

- For the trapezoid rule, the error is also bounded by \(O(n^{-2})\).
Simpson’s Rule for Numerical Integration

- Simpson’s rule is similar to the trapezoid rule, except that we compute the area under a quadratic polynomial approximation (instead of a linear approximation for the trapezoid). The equation is:

\[
\hat{I} = h \left[ \frac{1}{3} f(a) + \frac{4}{3} f(a + h) + \frac{2}{3} f(a + 2h) + \frac{4}{3} f(a + 3h) + \frac{2}{3} f(a + 4h) + \ldots + \frac{4}{3} f(b - h) + \frac{1}{3} f(b) \right].
\]

- Simpson’s rule is exact for polynomial functions up to cubics. The error can be bounded by the fourth derivative. i.e., \(O(n^{-4})\).

- It converges very quickly, assuming that \(f\) has a continuous fourth derivative.

- There are higher-order rules that can achieve even faster convergence, but require the function to be even smoother — a very rare event in computer graphics!
Multi-Dimensional Integration

- A common way to extend a 1D quadrature rule to higher dimensions is to use a tensor product rule. These rules have the form

\[ \hat{I} = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_s=1}^{n} \omega_{i_1} \omega_{i_2} \cdots \omega_{i_s} f(x_{i_1}, x_{i_2}, \ldots, x_{i_s}), \]

where \( s \) is the dimension, and the \( \omega_{i_k} \) and \( x_{i_k} \) are weights and sample locations for a given one-dimensional quadrature rule.

- Thus, if we start with an \( n \)-point quadrature rule in 1D, we need \( N = n^d \) sample points for a \( d \)-dimensional integral.

- In terms of the total number of samples the convergence is only \( O(N^{-r/d}) \) if the 1D rule has a convergence rate of \( O(n^{-r}) \).

- If we throw in a discontinuity in \( f \), things get even worse!
The basic Monte Carlo method is

\[ \int_{a}^{b} f(x) \, dx \approx \frac{b - a}{n} \sum_{i=1}^{n} f(X_i) \]

where the points \( X_i \) are chosen independently and uniformly at random within the interval \([a, b]\).

This method has a convergence rate of \( O(n^{-1/2}) \), regardless of the smoothness of the function \( f \).

Note that the convergence rate does not deteriorate in higher dimensions, and the number of samples needed does not grow astronomically.

This is particularly useful in graphics, where we often need to calculate multi-dimensional integrals of discontinuous functions, for which Newton-Cotes rules do not work well. (E.g., in distributed ray tracing.)
The End!

I hope that you enjoyed this course, and I wish you all the best for your future studies.