Mathematik für Informatiker III

Dr. Klaus Kriegel

Wintersemester 2005/06

Kapitel 1

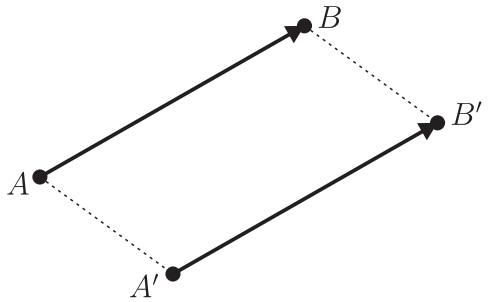
Lineare Algebra

1.1 Einführung: Anschauliche Vektorrechnung

Zur Vorbereitung auf die Beschäftigung mit abstrakten Vektorräumen ist es angebracht, sich noch einmal mit intuitiven, elementargeometrischen Grundgedanken zu dieser Theorie zu beschäftigen. Dabei geht man von Punkten in der Euklidischen Ebene bzw. im (dreidimensionalen) Euklidischen Raum aus, die durch Koordinatenpaare bzw. Koordinatentripel beschrieben sind. Zur leichteren Veranschaulichung werden wir im Folgenden vorwiegend Punkte in der Ebene betrachten, aber alle Überlegungen lassen sich sinngemäß auf den Raum übertragen.

Freie und gebundene Vektoren

Jedes geordnete Punktepaar beschreibt einen *gebundenen Vektor* $\overrightarrow{AB}$, veranschaulicht durch die gerichtete Strecke von A nach B. Auf der Menge der gebundenen Vektoren kann man eine Äquivalenzrelation einführen, unter der zwei Vektoren $\overrightarrow{AB}$ und $\overrightarrow{A'B'}$ äquivalent sind, wenn es eine Parallelverschiebung (Translation) gibt, die A in A' und B in B' überführt. Es ist klar, dass die Vektoren $\overrightarrow{AB}$ und $\overrightarrow{A'B'}$ genau dann äquivalent sind, wenn die Tupel der Koordinatendifferenzen zwischen B und A bzw. zwischen B' und A' gleich sind.



Eine Äquivalenzklasse dieser Relation nennt man einen *freien Vektor*. Anschaulich kann man also einen freien Vektor als ein Objekt beschreiben, dass eine bestimmte Richtung und eine bestimmte Länge, aber keinen festgelgten Anfangspunkt hat. Da ein freier Vektor durch das Differenzentupel der zu Grunde liegenden gebundenen Vektoren eindeutig charakterisiert wird, verwendet man dieses Tupel auch als Bezeichnung für den freien Vektor. Eine besondere Rolle unter den freien Vektoren spielt der Nullvektor $(0, 0)$, der die Äqui-

valenzklasse aller gebundenen Vektoren der Form $\overrightarrow{AA}$ ist. Für den Nullvektor wird auch oft die Kurzbezeichnung $\vec{0}$ verwendet.

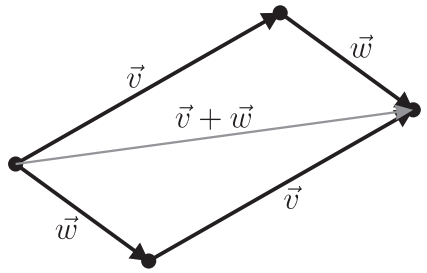
Beispiel: Für die Punkte $A = (-1, 2)$ und $B = (1, 3)$ und den gebundenen Vektor $\overrightarrow{AB}$ entsteht das Differenzentupel $(1 - (-1), 3 - 2) = (2, 1)$, das den zugehörigen freien Vektor bezeichnet.

Der Standardrepräsentant eines freien Vektors (a_1, a_2) ist der gebundene Vektor $\overrightarrow{OA}$, der vom Koordinatenursprung $O = (0, 0)$ zum Punkt $A = (a_1, a_2)$ führt. Man nennt diesen Vektor deshalb auch *Ortsvektor* des Punkts A.

Addition von Vektoren

Gebundene Vektoren sind ein nützliches Werkzeug in der Physik, z.B. kann man $\overrightarrow{AB}$ zur Beschreibung einer Kraft verwenden, die auf den Punkt A wirkt, wobei Richtung und Betrag der Kraft durch den freien Vektor beschrieben sind. Darüber hinaus sind sie sehr gut dazu geeignet, die Addition von Vektoren zu veranschaulichen: Die Summe von zwei Vektoren der Form $\overrightarrow{AB}$ und $\overrightarrow{BC}$ ist der Vektor $\overrightarrow{AC}$. Uns interessiert vor allem die Übertragung dieser Idee auf die freien Vektoren. Dazu muss man die Addition durch komponentenweise Addition der Koordinaten realisieren, d.h. $(s_1, s_2) + (t_1, t_2) = (s_1 + t_1, s_2 + t_2)$.

Zur Veranschaulichung der Addition von zwei freien Vektoren, die durch entsprechenden Ortsvektoren gegeben sind, konstruiert man ein Parallelogramm, dessen vom Koordinatenursprung abgehende Diagonale die Summe der zwei Vektoren repräsentiert.



Durch die Parallelogrammkonstruktion wird auch die Kommutativität der Vektoraddition sehr gut veranschaulicht. Zu jedem freien Vektor $\vec{v}$ kann man durch Umkehrung der Vorzeichen bei allen Komponenten den sogenannten *inversen Vektor* $-\vec{v}$ konstruieren, der sich durch die Eigenschaft $\vec{v} + (-\vec{v}) = \vec{0}$ auszeichnet. Ist $\overrightarrow{AB}$ ein Repräsentant von $\vec{v}$, dann ist $\overrightarrow{BA}$ ein Repräsentant von $-\vec{v}$.

Multiplikation mit Skalaren

Freie Vektoren können verlängert oder verkürzt (skaliert) werden indem die Länge des Vektors $\vec{v}$ mit einem bestimmten Faktor multipliziert wird, aber die Richtung gleich bleibt. Diese Faktoren - man nennt sie Skalare - können zunächst beliebige positv-reelle Zahlen sein, aber auch negative Zahlen $r \in \mathbb{R}$ kommen in Frage, wenn man den inversen Vektor $-\vec{v}$ mit dem Absolutbetrag $|r|$ skaliert. Die Skalierung eines Vektors $\vec{v} = (s_1, s_2)$ mit einem Faktor $r \in \mathbb{R}$ wird als Multiplikation $r \cdot \vec{v}$ notiert und durch die Regel $r \cdot (s_1, s_2) = (rs_1, rs_2)$ ausgeführt.

Verbindung zu linearen Gleichungssystemen und geometrischen Fragen

Durch Anwendung von Multiplikation mit Skalaren und Vektoraddition entstehen sogenannte Linearkombinationen von Vektoren, also Ausdrücke der Form $r_1 \cdot \vec{v_1} + \ldots + r_k \cdot \vec{v_k}$. Eine wichtiges Problem, mit der wir uns genauer beschäftigen werden, ist die Frage, ob ein bestimmter Vektor $\vec{v}$ als Linearkombination aus vorgegebenen Vektoren $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ erzeugt werden kann. Wir demonstrieren den Zusammenhang zwischen dieser Frage, der Lösung von linearen Gleichungsystemen und einem geometrischen Problem an einem Beispiel in dreidimensionalen Raum.

Problem 1: Kann man den Vektor $\vec{v} = (-1, 2, 5)$ als Linearkombination aus den Vektoren $\vec{v_1} = (5, 4, 3)$ und $\vec{v_2} = (3, 1, -1)$ darstellen?

Problem 2: Hat das folgende lineare Gleichungssystem eine Lösung?

$$\begin{array}{rcrcr} 5\alpha & + & 3\beta & = & -1 \\ 4\alpha & + & \beta & = & 2 \\ 3\alpha & - & \beta & = & 5 \end{array}$$

Problem 3: Liegt der Punkt $(-1, 2, 5)$ in der Ebene, die von den Punkten $(0, 0, 0)$, $(5, 4, 3)$ und $(3, 1, -1)$ aufgespannt wird?

Man kann sich leicht von der Gleichwertigkeit der drei Probleme überzeugen: Eine konkrete Lösung des Gleichungssystems würde die Koeffizienten für die Linearkombination von $\vec{v}$ liefern und zeigen, wie man den Ortsvektor des Punkts $(-1, 2, 5)$ aus den Ortsvektoren der Punkte $(5, 4, 3)$ und $(3, 1, -1)$ erzeugen könnte. Umgekehrt wären Koeffizienten einer Linearkombination von $\vec{v}$ aus $\vec{v_1}$ und $\vec{v_2}$ auch Lösungen des Gleichungssystems, usw.

Wir werden die lineare Algebra als eine Theorie kennenlernen, die es erlaubt, Probleme wie die oben genannten in ihrer allgemeinsten Form zu lösen und Zusammenhänge zu weiteren interessanten Fragestellungen herzustellen.

1.2 Gruppen und Körper

In diesem Abschnitt wollen wir uns damit beschäftigen, welche mathematischen Strukturen geeignet sind, Skalare eines Vektorraums zu beschreiben. Mit diesem strukturellen Ansatz werden die Gemeinsamkeiten bzw. Analogien zwischen den verschiedenen Zahlenbereichen, aber auch zu anderen mathematischen Objekten herausgearbeitet. Eine mathematische Struktur wird in der Regel durch eine Trägermenge, Operationen auf dieser Trägermenge und Eigenschaften dieser Operationen beschrieben.

Definition: Eine *Gruppe* $(G, *)$ besteht aus einer Trägermenge G und einer Operation $* : G \times G \longrightarrow G$ mit den folgenden drei Eigenschaften:

(G1) $\forall a, b, c \in G \quad (a * b) * c = a * (b * c)$ (Assoziativität)

(G2) $\exists e \in G \quad \forall a \in G \quad a * e = a = e * a$ (e ist neutrales Element)

(G3) $\forall a \in G \quad \exists \bar{a} \in G \quad a * \bar{a} = e = \bar{a} * a$ ($\bar{a}$ ist das zu a inverse Element)

$(G, *)$ ist *kommutative (abelsche) Gruppe*, falls zudem $\forall a, b \in G \quad a * b = b * a$ gilt.

Ist für ein Paar $(G, *)$ die Eigenschaft (G1) erfüllt, spricht man von einer *Halbgruppe* und sind die Eigenschaften (G1) und (G2) erfüllt, spricht man von einem *Monoid*.

Beispiele:

- $(\mathbb{Z}, +)$ ist eine kommutative Gruppe .
- $(\mathbb{N}, +)$ erfüllt mit $e = 0$ nur die Kriterien (G1) und (G2) und ist deshalb ein Monoid.
- $(\mathbb{N}^+, +)$ erfüllt nur das Kriterium (G1) und ist damit eine Halbgruppe.
- $(\mathbb{Q}, +)$ ist eine Gruppe.
- $(\mathbb{Q}, \cdot)$ ist ein Monoid (kein zu 0 inverses Element).
- $(\mathbb{Q} \setminus \{0\}, \cdot)$ ist eine Gruppe.
- $(S(M), \circ)$, wobei $S(M)$ die Menge der bijektiven Funktionen von M auf M und $\circ$ die Funktionskomposition darstellen, ist eine Gruppe. Dabei bildet die identische Funktion Id_M das neutrale Element und die inversen Elemente sind durch die Umkehrfunktionen f^{-1} gegeben.
- Bezeichne $\mathcal{B}_n$ die Menge aller n-stelligen Booleschen Funktionen und $\vee, \wedge, \oplus$ die Operationen Disjunktion, Konjunktion und Antivalenz.
  - $(\mathcal{B}_n, \vee)$ ist ein Monoid wobei das neutrale Element die Überall-Null-Funktion ist.
  - $(\mathcal{B}_n, \wedge)$ ist ein Monoid wobei das neutrale Element die Überall-Eins-Funktion ist.
  - $(\mathcal{B}_n, \oplus)$ ist eine Gruppe wobei das neutrale Element wieder durch die Überall-Null-Funktion gestellt wird und jede Funktion zu sich selbst invers ist.

Definition: Ein *Körper* $(K, \oplus, \odot)$ besteht aus einer Menge K und zwei Operationen $\oplus, \odot : R \times R \longrightarrow R$ mit den folgenden Eigenschaften:

(K1) $(K, \oplus)$ ist eine kommutative Gruppe mit einem neutralen Element 0.

(K2) $(K \setminus \{0\}, \odot)$ ist eine kommutative Gruppe mit einem neutralen Element 1.

(K3) $\forall a, b, c \in K \qquad a \odot (b \oplus c) = a \odot b \oplus a \odot c.$

Das zu einem $a \in K$ $\oplus$-inverse Element wird mit $-a$ bezeichnet. Ist $a \neq 0$, bezeichent man das $\odot$-inverse Element mit a^{-1} oder mit $\frac{1}{a}$.

Beispiele:

- $(\mathbb{Q}, +, \cdot)$ ist ein Körper.
- $(\mathbb{R}, +, \cdot)$ ist ein Körper.
- $(\mathbb{C}, +, \cdot)$ ist ein Körper.
- $(\mathbb{Z}, +, \cdot)$ ist **kein** Körper.

- Für jede Primzahl p bezeichne $\mathbb{Z}_p$ die Menge $\{0, 1, \ldots, p-1\}$ der möglichen Reste beim Teilen durch p. Definiert man darauf die Operationen Addition modulo p und Multiplikation modulo p entsteht ein Körper. Der Beweis, dass immer inverse Elemente bezüglich der Multiplikation existieren und Methoden zur Berechnung dieser Elemente, werden später besprochen.

Schlussfolgerungen aus den Axiomen:

Ein wesentlicher Vorteil der strukturellen Betrachtungen zeigt sich darin, dass man Sätze über Strukturen beweisen kann, die dann in jedem konkreten Exemplar der Struktur gültig sind und somit angewendet werden können. Exemplarisch werden hier ein paar einfache Schlussfolgerungen aus den Gruppen- und Körperaxiomen vorgestellt.

1) Das neutrale Element in einer Gruppe ist eindeutig.
Beweis: Angenommen zwei Elemente e und e' einer Gruppe erfüllen (G2). Dann wäre $e * e' = e'$ wegen (G2) für e und $e * e' = e$ wegen (G2) für e'. Damit muss $e = e'$ sein, d.h. das neutrale Element ist eindeutig.

2) In einer Gruppe ist für jedes Element $a \in G$ das zu a inverse Element eindeutig.
Beweis: Angenommen $\bar{a}$ und $\tilde{a}$ erfüllen beide (G3). Wir betrachten das Gruppenelement $b = (\bar{a} * a) * \tilde{a}$:

$$\begin{array}{rll} (\bar{a} * a) * \tilde{a} & = b = & \bar{a} * (a * \tilde{a}) \qquad |(G1) \\ e * \tilde{a} & = b = & \bar{a} * e \qquad |(G3) \text{ für } \bar{a} \text{ und für } \tilde{a} \\ \tilde{a} & = b = & \bar{a} \qquad |((G2)) \end{array}$$

3) In einer Gruppe hat jede Gleichung der Form $(a * x) * b = c$ eine eindeutige Lösung für die Varaible x.
Beweis: Die Gleichung wird äquivalent umgeformt, d.h. jeder Schritt der Umformung ist auch umkehrbar. Werden beide Seiten der Gleichung von rechts mit dem zu b inversen Element verknüpft, und auf der linken Seite (G3) und (G2) angewendet, enhält man

$$(a * x) * b = c \Longleftrightarrow a * x = c * \bar{b}$$

Werden analog beide Seiten der neuen Gleichung von links mit $\bar{a}$ verknüpft , so ergibt sich die eindeutige Lösung der Gleichung:

$$a * x = c * \bar{b} \Longleftrightarrow x = \bar{a} * (c * \bar{b})$$

4) In einem Körper hat jede Gleichung der Form $a \odot x \oplus b = c$ eine eindeutige Lösung, wenn $a \neq 0$ ist.
Beweis: Man verwendet das Symbol $-b$ für das bezüglich $\oplus$ zu b inverse Element und a^{-1} für das bezülich $\odot$ zu a inverse Element und erreicht mit ähnlichen Umformungen wie beim dritten Punkt die folgende Äquivalenz:

$$a \odot x \oplus b = c \Longleftrightarrow x = a^{-1} * (c \oplus -b)$$

Die nützliche Eigenschaft der eindeutigen Lösbarkeit beschränkt sich in Körpern im Allgemeinen auf lineare Gleichungen. Eine einfache quadratische Gleichung, wie $x \odot x = 2$ ist im Körper der rationalen Zahlen nicht lösbar.

1.3 Vektorräume

In allen nachfolgenden Betrachtungen wird K einen Körper mit den Operationen $+$ und $\cdot$ und den neutralen Elementen 0 und 1 bezeichnen.

Definition: Eine *Vektorraum* (abgekürzt VR) über dem Körper K besteht aus einer Menge V mit zwei Operationen $\oplus : V \times V \to V$ und $\odot : K \times V \to V$ mit den folgenden Eigenschaften:

- $(V, \oplus)$ ist eine kommutative Gruppe mit neutralem Element $\vec{0}$
  ($\ominus\vec{v}$ bezeichnet das $\oplus$-inverse Element zu $\vec{v}$)
- $\forall\, \lambda, \mu \in K \quad \forall\, \vec{v} \in V \quad \lambda \odot (\mu \odot \vec{v}) = (\lambda \cdot \mu) \odot \vec{v}$
- $\forall\, \vec{v} \in V \quad 1 \odot \vec{v} = \vec{v}$
- $\forall\, \lambda, \mu \in K \quad \forall\, \vec{v} \in V \quad (\lambda + \mu) \odot \vec{v} = (\lambda \odot \vec{v}) \oplus (\mu \odot \vec{v})$
- $\forall\, \lambda \in K \quad \forall\, \vec{v}, \vec{w} \in V \quad \lambda \odot (\vec{v} \oplus \vec{w}) = (\lambda \odot \vec{v}) \oplus (\lambda \odot \vec{w})$

Beispiele:

1. Der reelle Vektorraum $\mathbb{R}^n$ über dem Körper $\mathbb{R}$:

   $$V = \mathbb{R}^n = \{(x_1, \ldots x_n) \mid x_i \in \mathbb{R}\}$$

   Die Addition und Multiplikation mit Skalaren erfolgen komponentenweise:

   $$\begin{aligned} (x_1, \ldots, x_n) \oplus (y_1, \ldots, y_n) &= (x_1 + y_1, \ldots, x_n + y_n) \\ \lambda \odot (x_1, \ldots, x_n) &= (\lambda x_1, \ldots, \lambda x_n) \end{aligned}$$

   Es ist offensichtlich, dass $(0, \ldots, 0)$ der Nullvektor ist und dass man inverse Vektoren durch komponentenweise Umkehrung des Vorzeichens erhält:

   $$\ominus(x_1, \ldots, x_n) = (-x_1, \ldots, -x_n)$$

2. Die auf dem Intervall $[0, 1]$ definierten reellwertigen Funktionen bilden einen Vektorraum über dem Körper $\mathbb{R}$:

   $$V = \{f \mid f : [\,0, 1\,] \to \mathbb{R}\}$$

   Die Addition von zwei Funktionen $f, g \in V$ und die Multiplikation einer Funktion f mit einem Skalar λ ergeben neue Funktionen $f \oplus g$ bzw. $\lambda \odot f$, die wie folgt punktweise definiert werden:

   $$\begin{aligned} (f \oplus g)(x) &= f(x) + g(x) \\ (\lambda \odot f)(x) &= \lambda \cdot (f(x)) \end{aligned}$$

3. Die reellen Zahlen $\mathbb{R}$ sind ein Vektorraum über dem Körper $\mathbb{Q}$.

Bemerkung 1: Die Frage, ob man Vektoren des $\mathbb{R}^n$ in Zeilen- oder Spaltenform schreibt ist zunächst zweitrangig. Hier haben wir uns aus Platzgründen für die Zeilenform entschieden, aber insbesondere wenn lineare Abbildungen durch Matrizen repräsentiert werden, muss man die Spaltenform nutzen.

Bemerkung 2: Die Verwendung von verschiedenen Symbolen für die Addition von Vektoren und von Skalaren hat rein didaktischen Charakter. Ab jetzt werden wir in beiden Fällen das übliche + verwenden. Welche Operation anzuwenden ist, ergibt sich eindeutig aus dem Kontext. Gleiches gilt für die Multiplikationen und für die Subtraktion, welche eigentlich als Addition des inversen Elements zu verstehen ist, d.h. $\vec{v} - \vec{w} := \vec{v} \oplus (\ominus \vec{w})$.

Unterräume

Definition: Eine nichleere Teilmenge U eines Vektorraums V über K wird *Unterraum* oder genauer *Untervektorraum* von V (abgekürzt UR) genannt, falls

- $\forall\, \vec{v}, \vec{w} \in U \qquad \vec{v} + \vec{w} \in U$
- $\forall\, \vec{v} \in U \;\; \forall\, \lambda \in K \qquad \lambda \vec{v} \in U$

Diese Eigenschaften bedeuten, dass die Menge U abgeschlossen gegen Vektoraddition und Multiplikation mit Skalaren sein muss.

Beispiele:

1. Für den Vektorraum $V = \mathbb{R}^n = \{(x_1, \ldots x_n) \mid x_i \in \mathbb{R}\}$ ist die Teilmenge $U = \{(x_1, x_2, 0, \ldots 0) \mid x_1, x_2 \in \mathbb{R}\}$ ein Unterraum von V.
2. Für den Vektorraum $V = \{f \mid f : [0,1] \to \mathbb{R}\}$ sind die Teilmengen
   - $U = \{f \mid f : [0,1] \to \mathbb{R},\ f \text{ ist stetig}\}$ und
   - $U' = \{f \mid f : [0,1] \to \mathbb{R},\ f \text{ ist linear, d.h. } f(x) = ax + b\}$

   Unterräume von V.
3. Betrachtet man $V = \mathbb{R}$ als Vektorraum über dem Körper $\mathbb{Q}$, dann ist die Teilmenge $U = \{q_1 + q_2\sqrt{2} \mid q_1, q_2 \in \mathbb{Q}\}$ ein ein Unterraum von V.

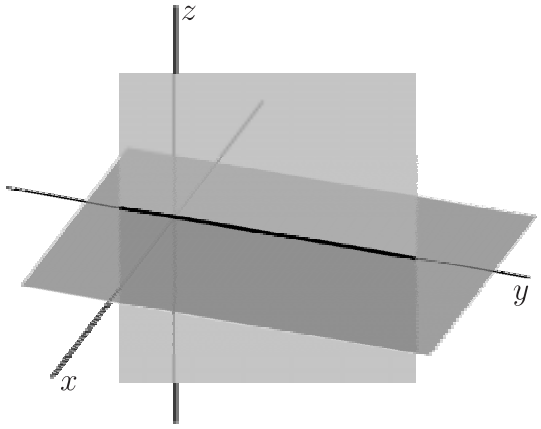
Satz: Sei V ein Vektorraum über einem Körper K und $\{U_i \mid i \in I\}$ eine Familie von Unterräumen, dann ist $\bigcap\limits_{i \in I} U_i$ auch ein Unterraum von V.

Beweis: Sei $\vec{u}, \vec{v} \in \bigcap\limits_{i \in I} U_i$ und $\lambda \in K$, dann gilt

- $\vec{u}$ und $\vec{v}$ sind Elemente von allen U_i
- $\vec{u} + \vec{v}$ und $\lambda \vec{u}$ sind Elemente von allen U_i

Daraus folgt, dass $\vec{u} + \vec{v} \in \bigcap\limits_{i \in I} U_i$ und $\lambda \vec{u} \in \bigcap\limits_{i \in I} U_i$.

Beispiel: Durchschnitt von xy-Ebene und der yz-Ebene in $\mathbb{R}^3$ ist die y-Achse.



Die folgenden zwei Beobachtungen sollten als Übung leicht zu beweisen sein. Ist U ein Unterraum eines Vektorraums V, dann gilt:

- Der Nullvektor $\vec{0}$ gehört zu U;
- Für jeden Vektor $\vec{u} \in U$ gehört auch der inverse Vektor $-\vec{u}$ zu U.

Linearkombinationen und lineare Hülle

Definition: Sind $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in V$ paarweise verschiedene Vektoren und $\lambda_1, \lambda_2, \ldots, \lambda_k \in K$ beliebige Skalare, so nennt man den Vektor

$$\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_k\vec{v}_k$$

eine *Linearkombination* (abgekürzt LK) aus den Vektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$.

Lemma: Sei $M \subseteq V$ eine Teilmenge eines Vektorraums V, dann bildet die Menge U_M aller Linearkombinationen von Vektoren aus M einen Unterraum von V.

Beweis: Man muss die Abgeschlossenheit von U_M bezüglich Vektoraddition und Multiplikation mit Skalaren nachweisen. Dazu betrachten wir ein Skalar $\alpha \in K$ zwei Vektoren aus U_M:

- $\vec{v} = \lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_k\vec{v}_k$ mit $\vec{v_i} \in M$ für $1 \leq i \leq k$
- $\vec{w} = \mu_1\vec{w}_1 + \mu_2\vec{w}_2 + \ldots + \mu_l\vec{w}_l$ mit $\vec{w_i} \in M$ für $1 \leq i \leq l$

Abgeschlossenheit bezüglich Multiplikation mit Skalaren:

$$\alpha\vec{v} = \alpha(\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_k\vec{v}_k) = (\alpha\lambda_1)\vec{v}_1 + (\alpha\lambda_2)\vec{v}_2 + \ldots + (\alpha\lambda_k)\vec{v}_k \in U_M$$

Um die Abgeschlossenheit bezüglich Addition zu zeigen, nehmen wir oBdA. an (Kommutativgesetz anwenden), dass alle Vektoren, die in den Linearkombinationen von $\vec{v}$ und $\vec{w}$ gemeinsam auftreten (die Anzahl sei j), linksbündig stehen, d.h. $\vec{v_i} = \vec{w_i}$ für alle i zwischen 1 und j und $\{\vec{v}_{j+1}, \ldots, \vec{v_k}\} \cap \{\vec{w}_{j+1}, \ldots, \vec{w}_l\} = \emptyset$. Dieser kleine technische Trick ist notwendig, um eine Linearkombination von paarweise verschiedenen Vektoren zu konstruieren:

$$\begin{aligned}\vec{v} + \vec{w} &= (\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_k\vec{v}_k) + (\mu_1\vec{w}_1 + \mu_2\vec{w}_2 + \ldots + \mu\vec{w}) \\ &= (\lambda_1 + \mu_1)\vec{v_1} + \ldots + (\lambda_j + \mu_j)\vec{v_j} + \lambda_{j+1}\vec{v}_{j+1} + \ldots + \lambda_k\vec{v_k} + \mu_{j+1}\vec{w}_{j+1} + \ldots + \mu_l\vec{w}_l\end{aligned}$$

Damit liegt $\vec{v} + \vec{w}$ als Linearkombination von Vektoren aus M auch in U_M.

Definition: Sei $M \subseteq V$ eine Menge von Vektoren, dann ist die *lineare Hülle* $\mathrm{Lin}(M)$ von M der kleinste Unterraum von V (bezüglich Inklusion), der M enthält, d.h.

$$\mathrm{Lin}(M) = \bigcap_{\substack{U \text{ ist UR von } V \\ M \subseteq U}} U$$

Satz: Die lineare Hülle einer Menge $M \subseteq V$ ist die Menge aller Linearkombinationen der Vektoren aus M, d.h.

$$\mathrm{Lin}(M) = \{\lambda_1 \vec{v}_1 + \ldots \lambda_k \vec{v}_k \mid \lambda_i \in K, \vec{v}_i \in M\}$$

Beweis: Einerseits bildet die Menge U_M aller Linearkombinationen von Vektoren aus M einen Unterraum (Lemma). Andererseits enthält jeder Unterraum U, der M enthält, auch alle Linearkombinationen von Vektoren aus M (Abgeschlossenheit von Unterräumen bezüglich der Addition und der Multiplikation mit Skalaren). Daraus folgt, dass U_M der kleinste Unterraum ist, der M enthält.

1.4 Lineare Unabhängigkeit, Basis und Dimension

Lineare Unabhängigkeit

Definition: Eine Menge $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ von k Vektoren heißt *linear abhängig* (l.a.), wenn eine Linearkombination existiert, mit

$$\lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2 + \ldots + \lambda_k \vec{v}_k = \vec{0}$$

wobei mindestens ein $\lambda_i \neq 0$ ist. Eine solche Linearkombination nennt man nichttriviale Linearkombination des Nullvektors.

Definition: Eine Menge $M \subseteq V$ ist linear unabhängig, wenn es **keine** nichttriviale Linearkombination des Nullvektors von Vektoren aus M gibt.

Folgerung: Eine Menge $M \subseteq V$ ist linear unabhängig, wenn jede endliche Teilmenge von M linear unabhängig ist.

Bemerkung 1: Man kann die Definition der linearen Unabhängigkeit auch sinngemäß auf Folgen von Vektoren anwenden. In diesem Fall reicht die Wiederholung eines Vektors in der Folge aber schon aus, lineare Abhängigkeit zu erzeugen, denn wenn $\vec{v_i} = \vec{v_j}$ ist, dann ist $1 \cdot \vec{v_i} + (-1) \cdot \vec{v_j}$ einen nichttriviale Linearkombination des Nullvektors.
Bemerkung 2: Aus $\vec{0} \in M$ folgt lineare Abhängigkeit, denn $\vec{0} = 1 \cdot \vec{0}$ ist eine nichttriviale Linearkombination des Nullvektors.

Beispiele:

1. Die Vektoren

$$\vec{v_1} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \qquad \vec{v_2} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in \mathbb{R}^3$$

sind linear unabhängig, denn für jede Linearkombination $\lambda_1\vec{v_1} + \lambda_2\vec{v_2} = \vec{0}$ gilt

$$\begin{aligned} 0 &= \lambda_2 \\ 0 &= \lambda_1 + \lambda_2 \\ 0 &= \lambda_1 + \lambda_2 \end{aligned}$$

und daraus folgt, $\lambda_1 = \lambda_2 = 0$.

2. Im Vektorraum $V = \{f \mid f : [0,1] \to \mathbb{R}\}$ sind die Funktionen f und g, die durch $f(x) = x + 1$ und $g(x) = 2$ definiert sind, linear unabhängig, denn für jede Linearkombination $\lambda f + \mu g$, die den Nullvektor, also die Funktion $h_0(x) = 0$ ergibt, wäre

$$\begin{aligned} h_0(0) = 0 &= \lambda f(0) + \mu g(0) = \lambda + 2\mu \\ h_0(1) = 0 &= \lambda f(1) + \mu g(1) = 2\lambda + 2\mu \end{aligned}$$

und daraus folgt, $\lambda = \mu = 0$.

Satz: Für jede Teilmenge M eines Vektorraums V sind die folgenden Aussagen äquivalent:

1. Die Menge M ist linear unabhängig.
2. Kein Vektor $\vec{v} \in M$ kann als Linearkombination aus den übrigen Vektoren aus M dargestellt werden.
3. Jeder Vektor $\vec{v} \in \mathrm{Lin}(M)$ hat eine *eindeutige* Darstellung als Linearkombination aus M.

Beweis: Der Satz wird über die negierten Aussagen nach folgendem Schema bewiesen:

$$\neg(1) \underset{\text{1. Schritt}}{\Rightarrow} \neg(2) \underset{\text{2. Schritt}}{\Rightarrow} \neg(3) \underset{\text{3. Schritt}}{\Rightarrow} \neg(1)$$

Zuerst formulieren wir die Negationen der drei Aussagen:
$\neg(1)$: Es gibt eine nichttriviale Linearkombination von $\vec{0}$.
$\neg(2)$: Es gibt einen Vektor $\vec{v} \in M$, der Linearkombination der übrigen Vektoren ist.
$\neg(3)$: Es gibt einen Vektor $\vec{v} \in \mathrm{Lin}(M)$ mit verschiedenen Linearkombinationen aus M.
Die drei Implikationen aus dem Schema kann man wie folgt beweisen.

- Schritt 1: Angenommen es gibt eine nichttriviale Linearkombination von $\vec{0}$:

$$\vec{0} = \lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_k\vec{v}_k$$

mit $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in M$, $\lambda_1, \lambda_2, \ldots, \lambda_k \in K$ und $\exists\, \lambda_i \neq 0$.

Ohne Beschränkung der Allgemeinheit können wir $\lambda_1 \neq 0$ annehmen. Diese Gleichung wird in zwei Schritten nach $\vec{v}_1$ umgestellt:

$$\begin{aligned} (-\lambda_1)\vec{v}_1 &= \lambda_2\vec{v}_2 + \lambda_3\vec{v}_3 + \ldots + \lambda_k\vec{v}_k \\ \vec{v}_1 &= (-\lambda_1)^{-1}\lambda_2\vec{v}_2 + (-\lambda_1)^{-1}\lambda_3\vec{v}_3 + \ldots + (-\lambda_1)^{-1}\lambda_k\vec{v}_k \end{aligned}$$

Damit ist $\vec{v}_1$ eine Linearkombination aus den übrigen Vektoren aaus M.

- Schritt 2: Angemonnen es gibt einen Vektor $\vec{v}$, der Linearkombination der übrigen Vektoren ist:

$$\vec{v} = \lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_k\vec{v}_k$$

wobei $\vec{v} \notin \{\vec{v}_1, \ldots, \vec{v}_k\}$. Damit existieren mindestens zwei verschiedene Linearkombinationen von $\vec{v_1}$:

$$\begin{aligned} \vec{v}_1 &= 1 \cdot \vec{v} + 0 \cdot \vec{v}_1 + 0 \cdot \vec{v}_2 + \ldots + 0 \cdot \vec{v}_k \\ &= 0 \cdot \vec{v} + \lambda_1 \cdot \vec{v}_1 + \lambda_2 \cdot \vec{v}_2 + \ldots + \lambda_k \cdot \vec{v}_k \end{aligned}$$

- Schritt 3: Angenommen, es existiert ein Vektor $\vec{v} \in \mathrm{Lin}(M)$ mit zwei verschiedenen Linearkombinationen aus M:

$$\begin{aligned} \vec{v} &= \lambda_1\vec{u}_1 + \lambda_2\vec{u}_2 + \ldots + \lambda_m\vec{u}_m \\ &= \mu_1\vec{w}_1 + \mu_2\vec{w}_2 + \ldots + \mu_n\vec{w}_n \end{aligned}$$

Dann betrachten wir die Vereinigung der zwei Vektormengen

$$\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\} \cup \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\} \subseteq M$$

und erweitern die beiden gegebnen Linearkombinationen zu Linearkombinationen über dieser Vereinigung indem für die Koeffizienten der jeweils fehlenden Vektoren Nullen gesetzt werden:

$$\begin{aligned} \vec{v} &= \lambda'_1\vec{v}_1 + \lambda'_2\vec{v}_2 + \ldots + \lambda'_k\vec{v}_k \\ &= \mu'_1\vec{v}_1 + \mu'_2\vec{v}_2 + \ldots + \mu'_k\vec{v}_k \end{aligned}$$

wobei

$$\lambda'_i = \begin{cases} \lambda_j & \text{falls } \vec{v}_i = \vec{u}_j \\ 0 & \text{sonst} \end{cases} \quad \text{und} \quad \mu'_i = \begin{cases} \mu_j & \text{falls } \vec{v}_i = \vec{w}_j \\ 0 & \text{sonst} \end{cases}$$

Da auch diese Linearkombinationen verschieden sind, gibt es ein i_0, so dass $\lambda'_{i_0} \neq \mu'_{i_0}$. Durch Subtraktion der beiden Gleichungen ergibt sich

$$\begin{aligned} \vec{0} = \vec{v} - \vec{v} &= (\lambda'_1\vec{v}_1 + \lambda'_2\vec{v}_2 + \ldots + \lambda'_k\vec{v}_k) - (\mu'_1\vec{v}_1 + \mu'_2\vec{v}_2 + \ldots + \mu'_k\vec{v}_k) \\ &= (\lambda'_1 - \mu'_1)\vec{v}_1 + \ldots + \underbrace{(\lambda'_{i_0} - \mu'_{i_0})}_{\neq 0}\vec{v}_{i_0} + \ldots + (\lambda'_k - \mu'_k)\vec{v}_k \end{aligned}$$

Damit wurde die Existenz einer nichttrivialen Linearkombination von $\vec{0}$ abgeleitet. □

Erzeugendensystem und Basis

Definition: Eine Teilmenge $M \subseteq V$ heißt *Erzeugendensystem* von V, wenn die lineare Hülle von M der Vektorraum V ist, d.h. wenn $\mathrm{Lin}(M) = V$.

Definition: Eine Teilmenge $M \subseteq V$ heißt *Basis* von V, wenn sie ein Erzeugendensystem von V und linear unabhängig ist.

Folgerung: Eine Teilmenge $M \subseteq V$ ist genau dann eine Basis von V, wenn jeder Vektor $\vec{v} \in V$ eine *eindeutige* Darstellung als Linearkombination aus M hat.

Beispiele:

- Die Vektoren

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots \quad \vec{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

  bilden eine Basis des Vektorraums $\mathbb{R}^n$, welche man *kanonische Basis* oder *Standardbasis* von $\mathbb{R}^n$ nennt. Zum Nachweis der Basiseigenschaften reicht die Überlegung, dass jeder Vektor aus $\mathbb{R}^n$ eindeutig als Linearkombination darstellbar ist:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1\vec{e}_1 + a_2\vec{e}_2 + \ldots + a_n\vec{e}_n$$

- Die Vektoren

$$\vec{v}_1 = \vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \vec{v}_2 = \vec{e}_1 + \vec{e}_2 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots \quad \vec{v}_n = \vec{e}_1 + \ldots + \vec{e}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

  bilden eine andere Basis des Vektorraums $\mathbb{R}^n$.

Satz: Für jede Teilmenge $M \subseteq V$ sind die folgenden Bedingungen äquivalent:

1. Die Menge M ist Basis von V.
2. Die Menge M ist ein minimales Erzeugendensystem von V.
3. Die Menge M ist eine maximale linear unabhängige Menge.

Die Begriffe „minimal“ und „maximal“ beziehen sich dabei auf die Inklusionsrelation von Mengen.

Während man die Äquivalenz der ersten beiden Bedingungen aus dem Satz über die verschiedenen Charakterisierungen der linearen Unabhängigkeit ableiten kann, hilft bei der Äquivalenz zwischen der ersten und der dritten Bedingung das folgende Lemma.

Lemma: Ist eine Teilmenge $M \subseteq V$ linear unabhängig und der Vektor $\vec{v} \in V$ nicht in der linearen Hüllen von M, dann ist die Menge $M \cup \{\vec{v}\}$ ebenfalls linear unabhängig.

Beweis (indirekt): Angenommen $M \cup \{\vec{v}\}$ wäre linear abhängig, dann existiert eine nichttriviale Linearkombination

$$\vec{0} = \lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_k\vec{v}_k + \lambda\vec{v}$$

in der $\lambda \neq 0$ sein muss, denn anderenfalls wäre das eine nichttriviale Linearkombination des Nullvektors über M. Durch Gleichungsumstellung ergibt sich

$$\vec{v} = (-\lambda)^{-1}\lambda_1\vec{v}_1 + (-\lambda)^{-1}\lambda_2\vec{v}_2 + \ldots + (-\lambda)^{-1}\lambda_k\vec{v}_k$$

Damit ist $\vec{v} \in \text{Lin}(M)$, ein Widerspruch zur Annahme. □

Basisergänzungssatz von Steinitz : Sei V ein Vektorraum über dem Körper K, $M = \{\vec{v_1}, \ldots, \vec{v_k}\}$ eine linear unabhängige Teilmenge von V und $N = \{\vec{w}_1, \ldots, \vec{w}_l\}$ eine weitere endliche Teilmenge von V, so dass die Vereinigung $M \cup N$ ein Erzeugendensystem von V ist. Dann kann man die Menge M durch eventuelle Hinzunahme von Vektoren aus der Menge N zu einer Basis des Vektorraumes V erweitern.

Beweisidee: Man beweist diesen Satz mit vollständiger Induktion nach $l = |N|$.

- Der Induktionsanfang mit $l = 0$ ist einfach, denn dann ist M nach den Voraussetzungen bereits eine Basis und muss nicht ergänzt werden.
- Für den Induktionsschritt von $l - 1$ nach l macht man eine Fallunterscheidung:
  1. Ist $\text{Lin}(M) = V$, dann ist M bereits eine Basis.
  2. Ist $\text{Lin}(M) \neq V$, dann muss es ein $\vec{w}_i \in N$ geben, das nicht zu $\text{Lin}(M)$ gehört (anderenfall wäre $\text{Lin}(M \cup N) = \text{Lin}(M) \neq V$). Nach obigen Lemma ist dann die Menge $M' = M \cup \{\vec{w}_i\}$ linear unabhängig und das Mengenpaar M' und $N' = N \setminus \{\vec{w}_i\}$ erfüllt die Induktionsvoraussetzung, weil $|N'| = l - 1$. Folglich kann man M' durch eventuelle Hinzunahme von Vektoren aus der Menge N' zu einer Basis des Vektorraumes V erweitern. □

Beispiel: Sei $M = \left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix} \right\}$ und N die Standardbasis von $\mathbb{R}^3$.

Der Vektor $\vec{e}_1$ ist bereits in M_2 enthalten und damit keine geeignete Ergänzung für M. Der Vektor $\vec{e}_2$ ist auch keine Basisergänzung, weil $M \cup \{\vec{e}_2\}$ linear abhängig wäre, doch der dritte Vektor $\vec{e}_3$ ergänzt die Menge M zu einer Basis.

Die folgenden zwei Aussagen sind unmittelbare Konsequenzen aus dem dem Basisergänzungssatz.

Austauschlemma: Sind die Mengen $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ und $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ zwei Basen eines Vektorraums V, dann gibt es für jeden Vektor $\vec{v}_i$ einen Vektor $\vec{w}_j$, so dass die Menge $(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \setminus \{\vec{v}_i\}) \cup \{\vec{w}_j\}$ ebenfalls Basis von V ist, d.h. man kann in der ersten Basis $\vec{v}_i$ gegen $\vec{w}_j$. austauschen.

Satz: Besitzt ein Vektorraum V eine endliche n-elementige Basis, dann ist jede andere Basis von V auch endlich und hat n Elemente.

Beweis: Sei eine n-elementige Basis B_1 und eine weitere Basis B_2 von V gegeben. Wendet man auf B_1 n-mal das Austauschlemma an, so entsteht einen Basis B von V, die keine Elemente B_1 sondern nur Elemente aus B_2 enthält. Da man weiss, dass kein Element aus B_2 mehrfach eingetauscht wurde (anderenfalls würde lineare Abhängigkeit entstehen), müssen es genau n Elemente sein. Damit ist $|B_1| = n \leq |B_2|$. Andererseits kann B_2 nicht mehr als n Elemente haben, denn $B \subseteq B_2$ ist bereits ein Erzeugendensystem von V und B_2 ist als Basis von V ein minimales Erzeugendensystem.

Dimension

Definition: Besitzt ein Vektorraum V eine endliche Basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, dann nennt man V einen *endlich-dimensionalen* und konkreter einen *n-dimensionalen* Vektorraum . Die Dimension des Raums wird mit $\dim V = n$ bezeichnet.
Ein Vektorraum, der keine endliche Basis besitzt, wird *unendlich-dimensional* genannt und man verwendet dafür die formale Schreibweise $\dim V = \infty$.

Satz: Ist $M = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ eine Teilmenge eines Vektorraums V mit $k > \dim V$, so ist M linear abhängig.

Beweis: Wäre M linear unabhängig, könnte man durch k-fache Anwendung des Basisergänzungssatzes die Vektoren aus M gegen Vektoren einer Basis von V austauschen. Da dabei zwangsläufig mindestens ein Basisvektor mehrfach in M eingefügt werden müsste, entstünde lineare Abhängigkeit - ein Wiederspruch.

Satz: Jeder Vektorraum besitzt eine Basis.

Dieser Satz ist von fundamentaler Bedeutung, aber leider übersteigt sein Beweis unsere bisher zur Verfügung stehenden mathematischen Mittel. Deshalb kann er im Rahmen dieser Vorlesung nur genannt, aber nicht bewiesen werden.
Im Gegensatz dazu, sind die folgenden Aussagen einfache Konsequenzen aus dem Basisergänzungssatz und dem Austauschlemma.

Satz: Ist die Dimension eines Vektorraumes V endlich und U ein Unterraum von V, dann gilt:
i) $\dim U \leq \dim V$
ii) $\dim U < \dim V \quad \Leftrightarrow \quad U \neq V$

Definition: Sind U_1 und U_2 Unterräume von V, so nennt man die Menge

$$U_1 + U_2 = \{\vec{x} + \vec{y} \mid \vec{x} \in U_1, \vec{y} \in U_2\}$$

die Summe von U_1 und U_2.

Beispiel: Sei $V = \mathbb{R}^4$ mit den Unterräumen $U_1 = \text{Lin}(\{\vec{e}_1, \vec{e}_2, \vec{e}_4\})$ und $U_2 = \text{Lin}(\{\vec{e}_1, \vec{e}_3, \vec{e}_4\})$ sowie $U_3 = \text{Lin}(\{\vec{e}_1 + \vec{e}_2, \vec{e}_1 + \vec{e}_4\}$, dann ist

$$U_1 + U_2 = \mathbb{R}^4 \qquad \text{und} \qquad U_1 + U_3 = U_1$$

Satz: Die Summe von zwei Unterräumen ist ein Unterraum. Für zwei endlich-dimensionale Unterräume U_1 und U_2 gilt:

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

Beweisidee:

- Die Abgeschlossenheit von $U = U_1 + U_2$ bezüglich Addition und Multiplikation mit Skalaren ist trivial.
- Sei $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}$ eine Basis von $U_1 \cap U_2$.
- Wir ergänzen B zu einer Basis B_1 von U_1:

$$B_1 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{u}_1, \vec{u}_2, \ldots, \vec{u}_s\}$$

- Wir ergänzen B zu einer Basis B_2 von U_2:

$$B_2 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{u}_1, \vec{u}_2, \ldots, \vec{u}_s\}$$

- Man weist nach, dass $B_1 \cup B_2 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{u}_1, \vec{u}_2, \ldots, \vec{u}_s, \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_t\}$ eine Basis von $U = U_1 + U_2$ ist

- Damit gilt für die Dimenstionen:

$$\dim(U_1 + U_2) = r + s + t = r + s + r + t - r = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

Beispiele:

- Sei $V = \mathbb{R}^3$, der Unterraum U_1 eine Ebene durch den Koordinatenursprung und der Unterraum U_2 eine Gerade durch den Koordinatenursprung, die aber nicht in U_1 liegt. Dann ist $\dim U_1 = 2$, $\dim U_2 = 1$ und $\dim(U_1 \cap U_2) = 0$ (wegen $U_1 \cap U_2 = \{\vec{0}\}$). Aus dem Satz folgt dann:

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_1 - \dim(U_1 \cap U_2) = 2 + 1 - 0 = 3$$

  Damit ist $U_1 + U_2 = \mathbb{R}^3$.

- Sind U_1 und U_2 zwei verschiedene Unterräume des $\mathbb{R}^n$ mit $\dim U_1 = \dim U_2 = n - 1$, dann ist $U_1 + U_2 = \mathbb{R}^n$ und folglich $\dim (U_1 \cap U_2) = n - 2$.

1.5 Lineare Abbildungen

Definition: Seien V und W zwei Vektorräume über einem Körper K. Eine Abbildung $f : V \to W$ heißt *linear* (oder *Vektorraumhomomorphismus*), wenn für alle $\vec{v}, \vec{w} \in V$ und für alle $\lambda \in K$ gilt:

$$\begin{aligned} f(\vec{v} + \vec{w}) &= f(\vec{v}) + f(\vec{w}) \\ f(\lambda \cdot \vec{v}) &= \lambda \cdot f(\vec{v}) \end{aligned}$$

$\mathrm{Hom}(V, W)$ bezeichnet die Menge aller linearer Abbildungen $f : V \to W$.

Beobachtungen:

- Sei $f \in \mathrm{Hom}(V, W)$, dann gilt:

$$f(\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_k\vec{v}_k) = \lambda_1 \cdot f(\vec{v}_1) + \lambda_2 \cdot f(\vec{v}_2) + \ldots + \lambda_k \cdot f(\vec{v}_k)$$

  für alle $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in V$ und alle $\lambda_1, \lambda_2, \ldots, \lambda_k \in K$.

- Die Verknüpfung von linearen Abbildungen $f : V \to W$ und $g : W \to Y$ ist eine lineare Abbildung $gf : V \to Y$, wobei die Verknüpfung wie folgt operiert:

$$gf(\vec{v}) = g(f(\vec{v}))$$

- Die Menge aller linearen Abbildungen $\mathrm{Hom}(V, W)$ ist selbst ein Vektorraum mit den Operationen:
  - $(f + g)(\vec{v}) = f(\vec{v}) + g(\vec{v})$
  - $(\lambda \cdot f)(\vec{v}) = \lambda \cdot f(\vec{v})$

  Dazu muss man nur nachprüfen, dass für alle $f, g \in \mathrm{Hom}(V, W)$, und $\lambda \in K$ die Abbildungen $f + g$ und $\lambda \cdot f$ auch linear sind. Das kann man aus den Definitionen der Operationen in $\mathrm{Hom}(V, W)$, den Eigenschaften von linearen Abbildungen und den Vektorraumeigenschaften ableiten:

$$\begin{aligned}(f+g)(\vec{u}+\vec{v}) &= f(\vec{u}+\vec{v}) + g(\vec{u}+\vec{v}) \\ &= f(\vec{u}) + f(\vec{v}) + g(\vec{u}) + g(\vec{v}) \\ &= f(\vec{u}) + g(\vec{u}) + f(\vec{v}) + g(\vec{v}) \\ &= (f+g)(\vec{u}) + (f+g)(\vec{v})\end{aligned}$$

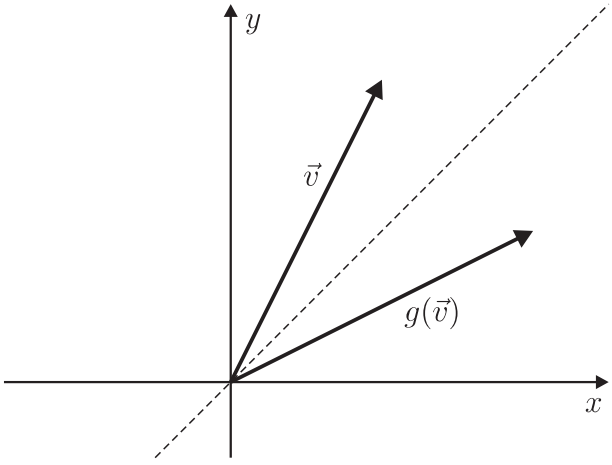
  Analog erfolgt auch die Ableitungen der drei anderen Bedingungen:

$$\begin{aligned}(\lambda \cdot f)(\vec{u}+\vec{v}) &= (\lambda \cdot f)(\vec{u}) + (\lambda \cdot f)(\vec{v}) \\ (f+g)(\mu\vec{u}) &= \mu(f+g)(\vec{u}) \\ (\lambda \cdot f)(\mu\vec{u}) &= \mu(\lambda \cdot f)(\vec{u})\end{aligned}$$

Eine Reihe häufig verwendeter geometrischer Transformationen, wie Drehungen um den Koordinatenursprung, Spiegelungen an Geraden bzw. Ebenen, die durch den Koordinatenursprung verlaufen, sowie Projektionen auf solche Geraden und Ebenen sind lineare Abbildungen. Die folgenden Beispiele geben eine Idee, warum das so ist:

a) Die Spiegelung an der x-Achse in $\mathbb{R}^2$ erfolgt durch die Abbildung $f\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}x\\-y\end{pmatrix}$

b) Die Spiegelung an der Geraden $y = x$ erfolgt durch die Abbildung $g\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}y\\x\end{pmatrix}$



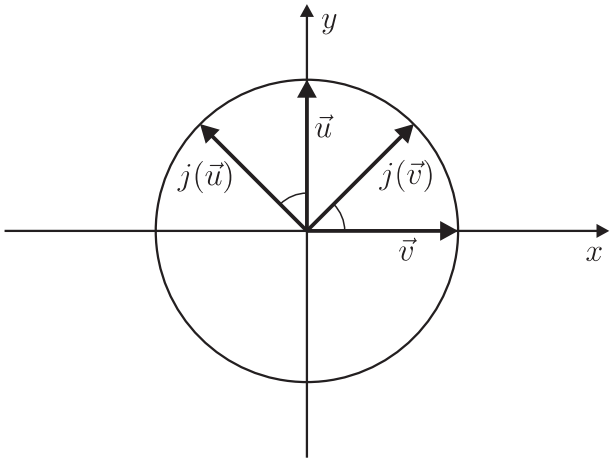
c) Die Projektion auf die y-Achse erfolgt durch die Abbildung $h\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0\\y\end{pmatrix}$

d) Die Drehung um 45° in $\mathbb{R}^2$ erfolgt durch eine Abbildung j, die man zuerst auf den Basisvektoren beschreibt:

$$j\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix} \qquad j\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}$$

Das erweitert man zu einer linearen Abbildung:

$$\begin{aligned} j\begin{pmatrix}x\\y\end{pmatrix} = j\left(x\cdot\begin{pmatrix}1\\0\end{pmatrix} + y\cdot\begin{pmatrix}0\\1\end{pmatrix}\right) &= x\cdot j\begin{pmatrix}1\\0\end{pmatrix} + y\cdot j\begin{pmatrix}0\\1\end{pmatrix} \\ &= \begin{pmatrix}\frac{1}{\sqrt{2}}\,x - \frac{1}{\sqrt{2}}\,y\\ \frac{1}{\sqrt{2}}\,x + \frac{1}{\sqrt{2}}\,y\end{pmatrix}\end{aligned}$$



Man beachte, dass Tranlationen (also Verschiebungen) keine linearen Abbildungen sind, da eine lineare Abbildung immer den Nullvektor auf den Nullvektor abbilden muss. Man kann Translationen erst durch einen Trick, nämlich die Einführung homogener Koordinatensysteme, als lineare Abbildung darstellen.

Kern und Bild von linearen Abbildungen

Definition: Der *Kern* Ker f und das *Bild* Im f einer linearen Abbildung $f \in \mathrm{Hom}(V, W)$ sind wie folgt definiert:

$$\begin{aligned} \mathrm{Ker}\ f &= \{\vec{v} \in V \mid f(\vec{v}) = \vec{0}\} \\ \mathrm{Im}\ f &= \{\vec{w} \in W \mid \exists\, \vec{v}\ \ f(\vec{v}) = \vec{w}\} \end{aligned}$$

Beispiele: Wir bestimmen die Kerne und Bilder der oben eingeführten geometrischen Transformationen.

a) Ker $f = \{\vec{0}\}$, denn

$$f\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}x\\-y\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \quad\Leftrightarrow\quad x = 0 \quad\wedge\quad y = 0$$

Im $f = \mathbb{R}^2$, denn

$$\begin{pmatrix} x \\ y \end{pmatrix} = f\begin{pmatrix} x \\ -y \end{pmatrix} \quad \text{für alle} \quad \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$$

Damit ist F eine bijektive Abbildung von $\mathbb{R}^2$ nach $\mathbb{R}^2$ mit der Umkehrabbildung $f^{-1} = f$.

b) Ker $g = \{\vec{0}\}$ und Im $g = \mathbb{R}^2$. Damit ist auch g ist eine bijektive Abbildung von $\mathbb{R}^2$ nach $\mathbb{R}^2$ mit der Umkehrabbildung $g^{-1} = g$.

c) Ker $h = \text{Lin}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}$, denn

$$h\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad y = 0$$

Im $h = \text{Lin}\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}$, denn die x-Komponente aller Elemente aus dem Bild ist 0.

d) Ker $j = \{\vec{0}\}$, denn

$$j\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}\, x - \frac{1}{\sqrt{2}}\, y \\ \frac{1}{\sqrt{2}}\, x + \frac{1}{\sqrt{2}}\, y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad x = 0 \quad \wedge \quad y = 0$$

Die Abbildung j ist eine bijektive Abbildung von $\mathbb{R}^2$ nach $\mathbb{R}^2$, weil mit

$$j^{-1}\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} \frac{1}{\sqrt{2}}\, x + \frac{1}{\sqrt{2}}\, y \\ -\frac{1}{\sqrt{2}}\, x + \frac{1}{\sqrt{2}}\, y \end{pmatrix}$$

eine Umkehrabbildung existiert. Daraus folgt Im $f = \mathbb{R}^2$.

Lemma: Der Kern einer linearen Abbildung $f \in \text{Hom}(V, W)$ ist ein Unterraum von V und das Bild von f ist ein Unterraum von W.

Beweis (Kern): Seien $\vec{u}, \vec{v} \in$ Ker f und $\lambda \in K$ (Körper zu V).

- Ker f ist nicht leer, denn:

$$f(\vec{0}) = f(\vec{0} - \vec{0}) = f(\vec{0}) - f(\vec{0}) = \vec{0}$$

Damit ist $\vec{0} \in$ Ker f.

- Abgeschlossenheit bezüglich der Addition:

$$\begin{array}{lrcl} \text{Für alle } \vec{u}, \vec{v} \in \text{Ker } f\text{:} & f(\vec{u} + \vec{v}) & = & f(\vec{u}) + f(\vec{v}) \\ & & = & \vec{0} + \vec{0} \\ & & = & \vec{0} \end{array}$$

Damit ist auch $\vec{u} + \vec{v} \in$ Ker f.

- Abgeschlossenheit bezüglich der Multiplikation mit Skalaren:

$$\text{Für alle } \vec{u} \in \text{Ker } f \text{ und für alle } \lambda \in K: \qquad \begin{aligned} f(\lambda\vec{u}) &= \lambda \cdot f(\vec{u}) \\ &= \lambda \cdot \vec{0} \quad (\text{da } \vec{u} \in \text{Ker } f) \\ &= \vec{0} \end{aligned}$$

  Damit ist auch $\lambda\vec{u} \in \text{Ker } f$. □

Beweis (Bild): Seien $\vec{u}, \vec{v} \in \text{Im } f$ und $\lambda \in K$ (Körper zu W).

- Im f ist nicht leer, denn $f(\vec{0}) = \vec{0} \in \text{Im } f$:
- Abgeschlossenheit bezüglich der Addition:

$$\text{Für alle } \vec{w} = f(\vec{v}), \vec{w'} = f(\vec{v'}) \in \text{Im } f \qquad \begin{aligned} \vec{w} + \vec{w'} &= f(\vec{v}) + f(\vec{v'}) \\ &= f(\vec{v} + \vec{v'}) \in \text{Im } f \end{aligned}$$

- Abgeschlossenheit bezüglich der Multiplikation mit Skalaren:

$$\text{Für alle } \vec{w} = f(\vec{v}) \in \text{Im } f \text{ und alle } \lambda \in K \qquad \begin{aligned} \lambda\vec{w} &= \lambda \cdot f(\vec{v}) \\ &= f(\lambda\vec{p}) \in \text{Im } f \end{aligned}$$

Lemma: Eine lineare Abbildung $f \in \text{Hom}(V, W)$ ist genau dann injektiv, wenn ihr Kern nur aus dem Nullvektor besteht, d.h.

$$f \text{ ist injektiv} \qquad \Longleftrightarrow \qquad \text{Ker } f = \{\vec{0}\}$$

Beweis: Die Richtung $\Rightarrow$ ist offensichtlich, denn da $f(\vec{0}) = \vec{0}$ und f injektiv ist, kann kein anderer Vektor auf $\vec{0}$ abgebildet werden.
Für die Richtung $\Leftarrow$ erfolgt der Beweis durch Widerspruch:
Angenommen Ker $f = \{\vec{0}\}$ und f ist *nicht* injektiv, dann existieren zwei Vektoren $\vec{u}, \vec{v} \in V$, so dass $\vec{u} \neq \vec{v}$ aber $f(\vec{u}) = f(\vec{v})$. Daraus folgt

$$f(\vec{u} - \vec{v}) = f(\vec{u}) - f(\vec{w}) = \vec{0}$$

und damit liegt $\vec{u} - \vec{v} \neq \vec{0}$ in Ker f. ein Widerspruch zur Annahme. □

Spezielle Homomorphismen

Definitionen: Einen Homomorphismus $f \in \text{Hom}(V, W)$ nennt man einen

- *Monomorphismus*, wenn f injektiv ist,
- *Epimorphismus*, wenn f surjektiv ist,
- *Isomorphismus*, wenn f bijektiv ist,
- *Endomorphismus*, wenn $V = W$,
- *Automorphismus*, wenn $V = W$ und f bijektiv ist.

Der folgende Satz ist eine einfache Konsequenz aus den bekannten Fakten, dass die Komposition von bijektiven Abbildungen auch bijektiv und die Komposition von linearen Abbildungen auch linear ist.

Satz: Die Komposition (Verkettung) von zwei Isomorphismen ist auch wieder ein Isomorphismus.

Satz: Ist $f \in \mathrm{Hom}(V,W)$ ein Isomorphismus, dann ist auch $f^{-1} \in \mathrm{Hom}(W,V)$ ein Isomorphismus.

Beweis: Da ein Isomorphismus bijektiv ist, gibt es eine eindeutige Umkehrfunktion $f^{-1} : W \longrightarrow V$, die durch

$$f(\vec{v}) = \vec{w} \Longleftrightarrow f^{-1}(\vec{w}) = \vec{v}$$

charakterisiert ist. Man muss nur noch die Abbildung f^{-1} auf Linearität überprüfen: Seien $\vec{w} = f(\vec{v})$ und $\vec{w'} = f(\vec{v'})$ gegeben. Da f linear ist, gilt $f(\vec{v}+\vec{v'}) = f(\vec{v})+f(\vec{v'}) = \vec{w}+\vec{w'}$ und $f(\lambda\vec{v}) = \lambda f(\vec{v}) = \lambda\vec{w}$. Jetzt ergibt sich die Linearität von f^{-1} durch Anwendung der oben beschriebene Äquivalenz auf die zwei Gleichungen:

$$\begin{aligned} f^{-1}(\vec{w}+\vec{w'}) &= \vec{v}+\vec{v'} = f^{-1}(\vec{w}) + f^{-1}(\vec{w'}) \\ f^{-1}(\lambda\vec{w}) &= \lambda\vec{v} = \lambda f^{-1}(\vec{w}) \end{aligned}$$

Satz: Seien V und W zwei Vektorräume über K, $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subseteq V$ eine Basis von V und $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n \in W$ eine beliebige Folge von Vektoren aus W, dann gibt es eine *eindeutige* lineare Abbildung $f \in \mathrm{Hom}(V,W)$ definiert durch

$$f(\vec{v}_i) = \vec{w}_i \quad \text{für} \quad i = 1, 2, \ldots n$$

Beweis: Jeder Vektor $\vec{v} \in V$ hat eine eindeutige Darstellung als Linearkombination aus den Basisvektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$:

$$\vec{v} = \lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_n\vec{v}_n$$

Wenn eine lineare Abbildung mit der Eigenschaft $f(\vec{v}_i) = \vec{w}_i$ für $i = 1, 2, \ldots n$ existiert, dann muss

$$f(\vec{v}) = \lambda_1\vec{w}_1 + \lambda_2\vec{w}_2 + \ldots + \lambda_n\vec{w}_n = \lambda_1 \cdot f(\vec{v}_1) + \lambda_2 \cdot f(\vec{v}_2) + \ldots + \lambda_n \cdot f(\vec{v}_n)$$

gelten. Es bleibt also nur noch zu zeigen, dass diese eindeutige Zuordnungsregel eine lineare Abbildung beschreibt. Dazu muss die Veträglichkeit mit der Addition und mit der Multiplikation mit Skalaren überprüft werden. Dazu seien $\lambda \in K$ sowie

$$\begin{aligned} \vec{v} &= \lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_n\vec{v}_n \\ \vec{u} &= \mu_1\vec{v}_1 + \mu_2\vec{v}_2 + \ldots + \mu_n\vec{v}_n \qquad \text{gegeben, woraus} \\ \vec{v}+\vec{u} &= (\lambda_1+\mu_1)\vec{v_1} + (\lambda_2+\mu_2)\vec{v_2} + (\lambda_n+\mu_n)\vec{v_n} \qquad \text{folgt.} \end{aligned}$$

Nach der Zuordnungsregel ist

$$\begin{aligned} f(\vec{v}+\vec{u}) &= (\lambda_1+\mu_1)f(\vec{v_1}) + (\lambda_2+\mu_2)f(\vec{v_2}) + (\lambda_n+\mu_n)f(vecv_n) \\ &= \lambda_1 f(\vec{v_1}) + \lambda_2 f(\vec{v_2}) + \lambda_n f(vecv_n) + \mu_1 f(\vec{v_1}) + \mu_2 f(\vec{v_2}) + \mu_n f(vecv_n) \\ &= f(\vec{v}) + f(\vec{u}) \end{aligned}$$

Analog kann man die zweite Eigenschaft $f(\lambda\vec{v}) = \lambda f(\vec{v})$ nachrechnen.

Folgerung: Zu zwei n-dimensionalen Vektorräumen existiert mindestens ein Isomorphismus, der den einen Vektorraum in den anderen überführt.

Rang einer linearen Abbildung

Definition: Der *Rang* einer linearen Abbildung $f \in \mathrm{Hom}(V, W)$ ist die Dimension des Bildes von f:

$$\mathrm{rg}\ f = \dim(\mathrm{Im}\ f)$$

Satz (Dimensionsformel für lineare Abbildungen): Für jede lineare Abbildung $f \in \mathrm{Hom}(V, W)$ auf einem endlichdimensionalen Vektorraum V gilt:

$$\dim V = \dim(\mathrm{Ker}\ f) + \dim(\mathrm{Im}\ f) = \dim(\mathrm{Ker}\ f) + \mathrm{rg}\ f$$

Beweis: Zuerst betrachten wir eine Basis $B_1 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ des Kerns $\mathrm{Ker}\ f \subseteq V$. Man kann B_1 zu einer Basis B von V erweitern. Sei $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, d.h. $\vec{v_{k+1}}, \ldots \vec{v_n}$ sind die Ergänzungsvektoren.
Wir betrachten die Bilder der Ergänzungsvektoren und werden zeigen, dass sie eine Basis $B_2 = \{f(\vec{v_{k+1}}), \ldots f(\vec{v_n})\}$ des Unterraums Im f bilden.

- **Erzeugendensystem:** Da man jeden Vektor $\vec{w}$ aus Im f als Linearkombination $\vec{w} = \lambda_1 f(\vec{v_1}) + \ldots + \lambda_n f(\vec{v_n})$ darstellen kann und außerdem $f(\vec{v_1}) = \ldots = f(\vec{v_k}) = \vec{0}$ gilt, reichen die Vektoren aus B_2 aus, um Im f zu erzeugen.

- **Lineare Unabhängigkeit:** Wir betrachten eine Linearkombination des Nullvektors aus B_2

  $$\vec{0} = \lambda_{k+1} f(\vec{v_{k+1}}) + \ldots + \lambda_n f(\vec{v_n})$$

  Damit ist

  $$f(\lambda_{k+1}\vec{v_{k+1}} + \ldots + \lambda_n\vec{v_n}) = \vec{0} = \lambda_{k+1} f(\vec{v_{k+1}}) + \ldots + \lambda_n f(\vec{v_n}) = \vec{0}$$

  und folglich ist $\vec{v} = \lambda_{k+1}\vec{v_{k+1}} + \ldots + \lambda_n\vec{v_n}$ Element des Kerns von f. Als solches muss $\vec{v}$ aber auch als Linearkombination aus B_1 darstellbar sein, d.h. $\vec{v} = \mu_1\vec{v}_1 + \mu_2\vec{v}_2 + \ldots + \mu_k\vec{v}_k$. Da jeder Vektor aus V, also insbesondere auch $\vec{v}$, eine **eindeutige** Basisdarstellung aus V hat, kann es nur eine Konsequenz geben: $\vec{v} = \vec{0}$ und $\mu_1 = \ldots = \mu_k = \lambda_{k+1} = \ldots = \lambda_n = 0$, womit gezeigt ist, dass die eingangs betrachtete Linearkombination des Nullvektors aus B_2 trivial sein muss.

Da wir nach den oben gewählten Bezeichnungen von $\dim(\mathrm{Ker}\ f) = k$, $\dim V = n$ und $\dim(\mathrm{Im}\ f) = n - k$ ausgehen können, ergibt sich die Dimensionsformel durch einfache Zusammenfassung

$$\dim(\mathrm{Ker}\ f) + \dim(\mathrm{Im}\ f) = k + (n - k) = n = \dim V$$

1.6 Matrizen

Definition: Eine $m \times n$-*Matrix* über einem Körper K ist eine Anordnung von $m \times n$ Elementen aus K nach dem folgenden Schema:

$$A = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \cdots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \cdots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \cdots & a_{m\,n} \end{pmatrix}$$

Alternativ kann man auch die Schreibweise $A = (a_{i\,j})_{(i,j)\in m\times n}$ verwenden. Die horizontalen n-Tupel werden Zeilen und die Vertikalen m-Tupel werden Spalten der Matrix genannt. Die Skalare $a_{i\,j}$ nennt man die Koeffizieten (oder Einträge) der Matrix.
Die Menge aller $m \times n$-Matrizen über K wird mit $M(m \times n, K)$ bezeichnet.

Beobachtung: Die Menge $M(m\times n, K)$ ist ein Vektorraum mit den folgenden Operationen ($\lambda \in K$):

$$\begin{pmatrix} a_{1\,1} & \dots & a_{1\,n} \\ \vdots & & \vdots \\ a_{m\,1} & \dots & a_{m\,n} \end{pmatrix} + \begin{pmatrix} b_{1\,1} & \dots & b_{1\,n} \\ \vdots & & \vdots \\ b_{m\,1} & \dots & b_{m\,n} \end{pmatrix} = \begin{pmatrix} a_{1\,1}+b_{1\,1} & \dots & a_{1\,n}+b_{1\,n} \\ \vdots & & \vdots \\ a_{m\,1}+b_{m\,1} & \dots & a_{m\,n}+b_{m\,n} \end{pmatrix}$$

$$\lambda \cdot \begin{pmatrix} a_{1\,1} & \dots & a_{1\,n} \\ \vdots & & \vdots \\ a_{m\,1} & \dots & a_{m\,n} \end{pmatrix} = \begin{pmatrix} \lambda a_{1\,1} & \dots & \lambda a_{1\,n} \\ \vdots & & \vdots \\ \lambda a_{m\,1} & \dots & \lambda a_{m\,n} \end{pmatrix}$$

Multiplikation von Matrizen

Definition: Ist $A \in M(m \times n, K)$ und $B \in M(n \times r, K)$ (wichtig ist der gemeinsame Parameter n), dann kann man das Produkt dieser Matrizen als eine Matrix $C = AB \in M(m \times r, K)$ definieren, deren Koeffizienten $c_{i\,j}$ die folgende Form haben:

$$\begin{aligned} c_{i\,j} &= a_{i\,1} \cdot b_{1\,j} + a_{i\,2} \cdot b_{2\,j} + \ldots + a_{i\,n} \cdot b_{n\,j} \\ &= \sum_{k=1}^{n} a_{i\,k} \cdot b_{k\,j} \end{aligned}$$

Man kann sich diese Regel so einprägen, dass man um $c_{i\,j}$ zu erhalten, die i-te Zeile von A mit der j-ten Spalte von B „multipliziert“, wobei multiplizieren hier bedeutet, die Produkte der sich entsprechenden Koeffizientenpaare aufzuaddieren.

Beispiel:

$$\begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 0+0-3 & 2+0-1 \\ 0+1+0 & 4-1+0 \end{pmatrix} = \begin{pmatrix} -3 & 1 \\ 1 & 3 \end{pmatrix}$$

Eine spezielle Ausprägung bekommt diese Regel bei der Multiplikation einer $m \times n$-Matrix mit einem Spaltenvektor aus K^n, der gleichzeitig eine $n \times 1$-Matrix ist. Das Ergebnis ist eine $m \times 1$-Matrix, also ein Spaltenvektor aus K^m. Auf diese Weise kann die Matrix A als Abbildung von K^n nach K^m interpretiert werden.

Satz: Die Multiplikation von Matrizen ist assoziativ, d.h. für alle $A \in M(m \times n, K)$, $B \in M(n \times r, K)$ und $C \in M(r \times s, K)$ gilt

$$(A \cdot B) \cdot C = A \cdot (B \cdot C)$$

Da der Beweis bis auf das Jonglieren mit komplizierten Summenformeln keine herausragenden Überraschungsmomente enthält, werden wir uns Zeit und Aufwand dafür sparen.

Achtung: Wie das folgende Beispiel zeigt, ist die Multiplikation von Matrizen im Allgemeinen *nicht* kommutativ:

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

Matrixdarstellung von lineare Abbildungen

Wir haben bereits gesehen, wie man eine Matrix als lineare Abbildung interpretieren kann. Jetzt geht es um den umgekehrten Weg bei dem eine lineare Abbildung $f \in \mathrm{Hom}(V, W)$ gegeben ist, die als Matrix dargestellt werden soll. Das ist aber nur möglich, wenn man vorher eine Basis von V und eine Basis von W festlegt.

Definition: Sei $f \in \mathrm{Hom}(V, W)$ eine lineare Abbildung, $B_1 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ Basis von V und $B_2 = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ Basis von W, dann wird der Abbildung f eine Matrix $A \in M(m \times n, K)$ zugeordnet, deren Koeffizienten $a_{i\,j}$ sich aus der Darstellung der Bilder der Basisvektoren $f(\vec{v}_i)$ in der Basis $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ wie folgt ergeben:

$$\begin{aligned} f(\vec{v}_1) &= a_{1\,1}\,\vec{w}_1 + a_{2\,1}\,\vec{w}_2 + \ldots + a_{m\,1}\,\vec{w}_m \\ f(\vec{v}_2) &= a_{1\,2}\,\vec{w}_1 + a_{2\,2}\,\vec{w}_2 + \ldots + a_{m\,2}\,\vec{w}_m \\ &\vdots \\ f(\vec{v}_n) &= a_{1\,n}\,\vec{w}_1 + a_{2\,n}\,\vec{w}_2 + \ldots + a_{m\,n}\,\vec{w}_m \end{aligned}$$

Umgekehrt bestimmt jede Matrix $A \in M(m \times n, K)$ durch die oberen Formeln eine lineare Abbildung $f \in \mathrm{Hom}(V, W)$, denn wir wissen, dass eine lineare Abbildung bereits durch die Bilder der Basisvektoren eindeutig beschrieben ist.

Häufig trifft man auf die besondere Situation, dass B_1 die Standardbasis von $V = \mathbb{R}^n$ und B_2 die Standardbasis von $W = \mathbb{R}^m$ ist. Für diesen Fall kann man sich die folgende Regel einprägen:

die j-te Spalte der Matrix A ist das Bild des j-ten Basisvektors von V, also $f(\vec{e}_j)$

Folgerung 1: Die Vektorräume der linearen Abbildungen $\mathrm{Hom}(V, W)$ und der Matrizen $M(m \times n, K)$ sind isomorph. Der Isomorphismus wird (nach Festlegung von zwei Basen für V und W) durch die oben beschriebene Zuordnung zwischen linearen Abbildungen und Matrizen realisiert.

Folgerung 2: Seien für $V = K^n$ und $W = K^m$ die jeweiligen Standardbasen festgelegt und sei $A \in M(m \times n, K)$ die zu einer Abbildung $f \in \mathrm{Hom}(V, W)$ gehörige Matrix, dann erhält man das Bild $f(\vec{v})$ eines beliebigen Spaltenvektors $\vec{v} \in V$ durch Multiplikation der Matrix A mit $\vec{v}$, d.h.

$$A \cdot \vec{v} = f(\vec{v})$$

Man kann die zweite Folgerung durch einfaches Nachrechnen überprüfen:

$$\begin{aligned}
A \cdot \vec{v} &= \begin{pmatrix} a_{1\,1} & \cdots & a_{1\,n} \\ \vdots & & \vdots \\ a_{m\,1} & \cdots & a_{m\,n} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= \begin{pmatrix} a_{1\,1} \cdot x_1 + \ldots + a_{1\,n} \cdot x_n \\ \vdots \\ a_{m\,1} \cdot x_1 + \ldots + a_{m\,n} \cdot x_n \end{pmatrix} \\
f(\vec{v}) = f \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &= f\left(x_1 \cdot \vec{e}_1^{\,(n)} + \ldots + x_n \cdot \vec{e}_n^{\,(n)}\right) \\
&= x_1 \cdot f\left(\vec{e}_1^{\,(n)}\right) + \ldots + x_n \cdot f\left(\vec{e}_n^{\,(n)}\right) \\
&= x_1 \cdot \begin{pmatrix} a_{1\,1} \\ \vdots \\ a_{m\,1} \end{pmatrix} + \ldots + x_n \cdot \begin{pmatrix} a_{1\,n} \\ \vdots \\ a_{m\,n} \end{pmatrix} \\
&= \begin{pmatrix} a_{1\,1} \cdot x_1 + \ldots + a_{1\,n} \cdot x_n \\ \vdots \\ a_{m\,1} \cdot x_1 + \ldots + a_{m\,n} \cdot x_n \end{pmatrix} \\
\Rightarrow \quad A \cdot \vec{v} &= f(\vec{v})
\end{aligned}$$

Satz: Sind $f \in \mathrm{Hom}(K^q, K^p)$ sowie $g \in \mathrm{Hom}(K^r, K^q)$ lineare Abbildungen und $A \in M(p \times q, K)$ bzw. $B \in M(q \times r, K)$ die zu f bzw. g gehörigen Matrizen bezüglich der Standardbasen von K^p, K^q und K^r dann entspricht das Produkt der Matrizen $A \cdot B$ der Abbildungskomposition fg, vereinfacht geschrieben:

$$C = A \cdot B \in M(p \times r, K) \quad \longleftrightarrow \quad fg \in \mathrm{Hom}(K^r, K^p)$$

Auf den Beweis dieses Satzes wird verzichtet, weil er auch in die Kategorie der rechnerisch aufwändigen, aber nicht sehr originellen Beweise gehört.

Beispiele:

a) Die Skalierung des Raumes $\mathbb{R}^n$ um einen Faktor $c \in \mathbb{R}$ wird durch die folgende Matrix realisiert:

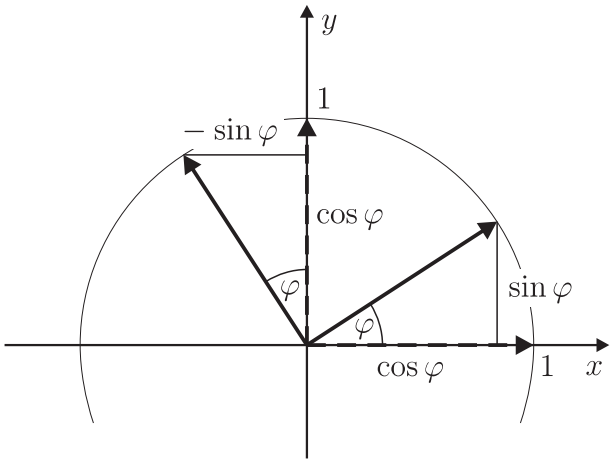
$$A = \begin{pmatrix} c & 0 & \cdots & 0 \\ 0 & c & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c \end{pmatrix}$$

b) Die Projektion dea Raums $\mathbb{R}^3$ auf die xy-Ebene im gleichen Raum wird durch die folgende Matrix realisiert:

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

c) Die Drehung der Ebene $\mathbb{R}^2$ mit dem Winkel φ um den Koordinatenursprung wird durch die folgende Matrix realisiert:

$$C = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$



Die Drehung des Punkts $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ mit dem Winkel $\frac{\pi}{3}$ um den Koordinatenursprung kann man dann wie folgt berechnen:

$$\begin{pmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 - 2\sqrt{3} \\ \sqrt{3} + 2 \end{pmatrix}$$

1.7 Der Rang einer Matrix

Der Rang einer Matrix kann auf drei verschiedene Arten definiert werden. Um diese beser unterscheiden zu können führen wir zuerst drei Begriffe ein, von denen später gezeigt wird, dass sie gleich sind.

Definition: Sei $A \in M(m \times n, K)$ eine Matrix und $f \in \mathrm{Hom}(K^n, K^m)$ die zugehörige lineare Abbildung (bezüglich der Standardbasen).

- Der *Rang* von A ist definiert durch $\quad \mathrm{rg}\ A := \mathrm{rg}\ f = \dim\ (\mathrm{Im}\ f)$
- Der *Zeilenrang* von A ist die maximale Anzahl von linear unabhängigen Zeilenvektoren aus A.
- Der *Spaltenrang* von A ist die maximale Anzahl von linear unabhängigen Spaltenvektoren aus A.

Satz: Der Rang und der Spaltenrang einer Matrix A sind gleich.

Beweis: Die Spalten von A sind die Bilder der Basisvektoren. Folglich bilden die Spaltenvektoren der Matrix ein Erzeugendensystem von Im f. Damit ist jede maximale linear

unabhängige Teilmenge der Spaltenvektoren eine Basis von Im f, und daraus folgt, dass der Spaltenrang von A gleich dim (Im f) = rg f = rg A ist.

Lemma: Ist $\vec{w_k}$ ein Spaltenvektor von $A \in M(m \times n, K)$, der sich als Linearkombination der übrigen Spalten darstellen lässt und ist A' die Matrix A ohne Spalte $\vec{w_k}$, dann gilt:

$$\text{Spaltenrang } A' = \text{Spaltenrang } A \quad \textbf{und} \quad \text{Zeilenrang } A' = \text{Zeilenrang } A$$

Die gleiche Aussage gilt auch, wenn man aus der Matrix einen Zeilenvektor streicht, der sich als Linearkombination aus den anderen Zeilen darstellen lässt.

Beweis: Die maximale Anzahl von linear unabhängigen Spaltenvektoren ist gleichzeitig die Dimension der linearen Hülle der Menge aller Spaltenvektoren $\{\vec{w_1}, \ldots, \vec{w_n}\}$ von A. Ist der Vektor

$$\vec{w_k} = \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda j \, \vec{w_j}$$

eine Linearkombination der anderen Spaltenvektoren, dann bleibt die lineare Hülle nach seiner Streichung unverändert und deshalb bleibt der Spaltenrang gleich.
Für die Betrachtung des Zeilenrangs sei $I \subseteq \{1, 2, \ldots, m\}$ eine maximale Menge, so dass die Zeilenvektoren $\{\vec{u_i} \mid i \in I\}$ linear unabhängig sind, und sei $\{\vec{u_i}' \mid i \in I\}$ die entsprechende Menge von Zeilenvektoren aus A', in denen also jeweils die k-te Stelle gestrichen ist. Um die Gleichheit des Zeilenrangs von A und A' zu zeigen, genügt es, die lineare Unabhängigkeit von $\{\vec{u_i}' \mid i \in I\}$ nachzuweisen. Sei

$$\sum_{i \in I} \mu_i \, \vec{u_i}' = \vec{0}$$

eine Linearkombination des Nullvektors in K^{n-1}. Wir werden zeigen, dass dann auch die Linearkombination $\sum_{i \in I} \mu_i \vec{u_i}$ den Nullvektor in K^n erzeugt. Damit müssen alle Skalare μ_i gleich 0 sein und folglich ist $\{\vec{u_i}' \mid i \in I\}$ linear unabhängig. In der Linearkombination $\sum_{i \in I} \mu_i \vec{u_i}$ sind bereits alle Stellen bis auf die k-te gleich Null. Bleibt also $\sum_{i \in I} \mu_i a_{i\,k} = 0$ zu zeigen. Nach Voraussetzung über den Spaltenvektor $\vec{w_k}$ wissen wir für alle $i \in I$

$$\begin{aligned}
a_{i\,k} &= \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_j a_{i\,j} \qquad\qquad \text{und folglich} \\
\sum_{i \in I} \mu_i a_{i\,k} &= \sum_{i \in I} \Big(\mu_i \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_j \, a_{i\,j} \Big) \\
&= \sum_{\substack{j=1 \\ j \neq k}}^{n} \Big(\lambda_j \underbrace{\sum_{i \in I} \mu_i \, a_{i\,j}}_{= 0 \text{ da } j \neq k} \Big) \\
&= \sum_{\substack{j=1 \\ j \neq k}}^{n} 0 = 0 \qquad\qquad \square
\end{aligned}$$

Satz: Der Spaltenrang und der Zeilenrang einer Matrix A sind gleich und damit gilt:

$$\text{rg } A = \text{Spaltenrang } A = \text{Zeilenrang } A$$

Beweis: Wir streichen aus A Zeilen bzw. Spalten, die jeweils Linearkombinationen der übrigen Zeilen bzw. Spalten sind, solange das möglich ist.

$$A \mapsto A' \mapsto A'' \mapsto \ldots \mapsto A^{(\text{end})}$$

Dann ist in der Matrix $A^{(\text{end})}$ die Menge der Zeilenvektoren linear unabhängig und auch die Menge der Spaltenvektoren ist linear unabhängig. Sei$A^{(\text{end})}$ eine $m \times n$–Matrix, dann gilt mach dem Lemma:

$$\begin{aligned} \text{Spaltenrang } A &= \text{Spaltenrang } A^{(\text{end})} = n \\ \text{Zeilenrang } A &= \text{Zeilenrang } A^{(\text{end})} = m \end{aligned}$$

Es gibt m linear unabhängige Zeilenvektoren in $A^{(\text{end})}$, aber das sind Vektoren aus K^n und deshalb muss $m \leq n$ sein. Andererseits gibt es n linear unabhängige Spaltenvektoren in $A^{(\text{end})}$, aber das sind Vektoren aus K^m und deshalb muss $n \leq m$ sein.
Folglich ist Spaltenrang $A = n = m =$ Zeilenrang A. □

Definition: Sei $A = (a_{i\,j}) \in M(m \times n, K)$ eine Matrix, dann ist transponierte Matrix von A definiert durch

$$A^t = (a^t_{i\,j}) \in M(n \times m, K) \quad \text{mit} \quad a^t_{i\,j} = a_{j\,i}$$

Beispiel:

$$\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 4 & 0 \end{pmatrix}^t = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 0 \end{pmatrix}$$

Folgerung: Der Rang einer Matrix A und der transponierten Matrix A^t ist gleich.

$$\text{rg } A = \text{rg } A^t$$

Elementare Umformungen

Definition: Die folgenden Operationen auf einer Matrix werden elementaren Umformungen genannt:

- Typ 1: Vertauschung von zwei Zeilen bzw. von zwei Spalten.
- Typ 2: Multiplikation einer Zeile bzw. Spalte mit einem Skalar $\lambda \neq 0$.
- Typ 3: Addition des λ-fachen einer Zeile bzw. Spalte zu einer anderen Zeile bzw. Spalte.

Satz: Elementare Umformungen ändern den Rang einer Matrix nicht.
Beweis: Man muss sich nur davon überzeugen, dass sich die lineare Hülle der Zeilen- bzw. Spaltenvektoren durch die Umformungen nicht ändert. Für Umformungen vom Typ 1 und Typ 2 ist das offensichtlich.
Für eine Umformungen vom Typ 3 seien $\vec{v}_i, \vec{v}_k$ zwei Zeilenvektoren und $\lambda \in K$ ein Skalar. Nach der Umformung hat man an Stelle von $\vec{v}_i$ den Vektor

$$\vec{v}^*_i = \vec{v}_i + \lambda \vec{v}_k$$

Es ist eine leichte Übung, in Linearkombinationen $\vec{v}_i$ gegen $\vec{v}_i^*$ auszutauschen:

$$\begin{aligned}\mu_1\vec{v}_1+\ldots+\mu_i\vec{v}_i+\ldots+\mu_k\vec{v}_k+\ldots+\mu_n\vec{v}_n &= \mu_1\vec{v}_1+\ldots+\mu_i\vec{v}_i^*+\ldots+\underline{(\mu_k-\lambda\mu_i)}\vec{v}_k+\ldots\\ &\quad\ldots+\mu_n\vec{v}_n\\ \mu_1\vec{v}_1+\ldots+\mu_i\vec{v}_i^*+\ldots+\mu_k\vec{v}_k+\ldots+\mu_n\vec{v}_n &= \mu_1\vec{v}_1+\ldots+\mu_i\vec{v}_i+\ldots+\underline{(\mu_k+\lambda\mu_i)}\vec{v}_k+\ldots\\ &\quad\ldots+\mu_n\vec{v}_n\end{aligned}$$

Obere Dreiecksform

Definition: Die Matrix A ist in oberer Dreiecksform, wenn die Matrix die folgende Form hat, wobei das Symbol $*$ für beliebige Inhalte steht:

$$A=\left(\begin{array}{ccccc|ccc} a_{1\,1} & * & * & \cdots & * & * & \cdots & * \\ 0 & a_{2\,2} & * & \cdots & * & \vdots & \ddots & \vdots \\ 0 & 0 & a_{3\,3} & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & * & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{r\,r} & * & \cdots & * \\ \hline 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 \end{array}\right)$$

und die Werte $a_{1\,1}, a_{2\,2}, a_{3\,3}, \ldots a_{r\,r}$ ungleich Null sind, d.h.

$$a_{1\,1}\cdot a_{2\,2}\cdot a_{3\,3}\cdot\ldots\cdot a_{r\,r}\neq 0$$

Beobachtung: Der Rang einer solchen Matrix ist r.

Algorithmus zur Umwandlung in eine obere Dreiecksform:
Mit dem folgenden Verfahren kan man eine beliebige Matrix $A\in M(m\times n,K)$ mit elementaren Umformungen in eine obere Dreiecksform überführen und damit auch ihren Rang bestimmen. Das Verfahren arbeitet in $r=\operatorname{rg}A$ Stufen.
Zustandsinvariante: Nach jeder Stufe $k\geq 1$ wird eine Matrix A_k der folgenden Form erreicht:

$$A_k=\left(\begin{array}{ccccc|ccc} a_{1\,1} & * & * & \cdots & * & * & \cdots & * \\ 0 & a_{2\,2} & * & \cdots & * & \vdots & \ddots & \vdots \\ 0 & 0 & a_{3\,3} & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & * & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{k\,k} & * & \cdots & * \\ \hline 0 & \cdots & \cdots & \cdots & 0 & b_{k+1\,k+1} & \cdots & b_{k+1\,n} \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & b_{m\,k+1} & \cdots & b_{m\,n} \end{array}\right)$$

Dabei können die Koeffzienten $b_{i\,j}$ (mit $i=k+1,k+2,\ldots m$ und $j=k+1,k+2,\ldots n$) beliebig sein, aber gefordert ist $a_{1\,1}\cdot a_{2\,2}\cdot a_{3\,3}\cdot\ldots\cdot a_{k\,k}\neq 0$.

Sei $B_k = \begin{pmatrix} b_{k+1\,k+1} & \cdots & b_{k+1\,n} \\ \vdots & \ddots & \vdots \\ b_{m\,k+1} & \cdots & b_{m\,n} \end{pmatrix}$ die rechte untere Teilmatrix von A_k.

Initialisierung: Wir setzen $A_0 = A$. Offensichtlich gibt es für A_0 keine Einschränkungen, denn $B_0 = A_0$.

Abbruchkriterium: Das Verfahren ist beendet, wenn B_k die Nullmatrix ist, denn dann ist A_k in oberer Dreiecksform.

Umwandlung von A_k in A_{k+1}: Wenn das Abbruchkriterium für A_k noch nicht erfüllt ist, gibt es in B_k einen Koeffizienten $b_{i\,j} \neq 0$.

- Vertausche Zeilen und/oder Spalten, die durch B gehen, um den Koeffizieten $b_{i,j}$ an die Stelle von $b_{k+1\,k+1}$ zu bringen. Die so entstandene Matrix A'_k folgende Gestalt:

$$A'_k = \left(\begin{array}{cccc|cccc} a_{1\,1} & * & \cdots & * & * & * & \cdots & * \\ 0 & a_{2\,2} & \ddots & * & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & a_{k\,k} & * & * & \cdots & * \\ \hline 0 & \cdots & \cdots & 0 & b'_{k+1\,k+1} & \cdots & \cdots & b'_{k+1\,n} \\ \vdots & \ddots & \ddots & \vdots & b'_{k+2\,k+1} & \cdots & \cdots & b'_{k+2\,n} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & b'_{m\,k+1} & \cdots & \cdots & b'_{m\,n} \end{array}\right)$$

  wobei nun $b'_{k+1\,k+1} = b_{i,j} \neq 0$ ist.

- Für die Stellen $b'_{k+2\,k+1}, b'_{k+3\,k+1}, \ldots b'_{m\,k+1}$ werden durch Typ-3-Umformungen Nullen erzeugt. Man verwendet für die Zeilen $i = k{+}2, k{+}3, \ldots m$ die folgenden Umformungen:

$$\text{Zeile}_i := \text{Zeile}_i - \left(b'_{i\,k+1} \cdot (b'_{i\,k+1})^{-1}\right) \cdot \ \text{Zeile}_{k+1}$$

  Dadurch entsteht die neue Matix $A''_k = A_{k+1}$:

$$A''_k = \left(\begin{array}{cccc|cccc} a_{1\,1} & * & \cdots & * & * & * & \cdots & * \\ 0 & a_{2\,2} & \ddots & * & \vdots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & a_{k\,k} & * & * & \cdots & * \\ \hline 0 & \cdots & \cdots & 0 & b'_{k+1\,k+1} & \cdots & \cdots & b'_{k+1\,n} \\ \vdots & \ddots & \ddots & \vdots & 0 & b''_{k+2\,k+2} & \cdots & b''_{k+2\,n} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & b''_{m\,k+2} & \cdots & b''_{m\,n} \end{array}\right)$$

  wobei für alle $i = k+2, k+3, \ldots m$ und $j = k+1, k+2, \ldots n$ die Koeffizienten in B_{k+1} die folgenden Werte haben:

$$b''_{i\,j} := b'_{i\,j} - b'_{i\,k+1} \cdot \frac{b'_{k+1\,j}}{b'_{k+1\,k+1}}$$

Beispiel: Die folgende Matrix A soll in eine obere Dreiecksform überführt werden:

$$A = \begin{pmatrix} 0 & -2 & 4 \\ 2 & 1 & 0 \\ 1 & 0 & 2 \\ 2 & 0 & 3 \end{pmatrix}$$

- Vertausche die erste und die dritte Zeile, so dass an der Stelle $a_{1\,1}$ ein Koeffizient $\neq 0$ steht:

$$\begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & -2 & 4 \\ 2 & 0 & 3 \end{pmatrix}$$

- Erzeuge an den Stellen $a_{2\,1}$ und $a_{4\,1}$ Nullen durch Typ-3-Umformungen mit der ersten Zeile:

$$\begin{pmatrix} 1 & 0 & 2 \\ 2-2\cdot 1 & 1-2\cdot 0 & 0-2\cdot 2 \\ 0 & -2 & 4 \\ 2-2\cdot 1 & 0-2\cdot 0 & 3-2\cdot 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -4 \\ 0 & -2 & 4 \\ 0 & 0 & -1 \end{pmatrix}$$

- An der Stelle $a_{2\,2}$ befindet sich ein Koeffizient $\neq 0$. Damit muss nur noch an der Stelle $a_{2\,3}$ eine Null erzeugt werden:

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -4 \\ 0 & -2-(-2)\cdot 1 & 4-(-2)\cdot(-4) \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -4 \\ 0 & 0 & -4 \\ 0 & 0 & -1 \end{pmatrix}$$

- An der Stelle $a_{3\,3}$ muss eine Null erzeugt werden:

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -4 \\ 0 & 0 & -4 \\ 0 & 0 & -1-\frac{1}{4}\cdot(-4) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -4 \\ 0 & 0 & -4 \\ 0 & 0 & 0 \end{pmatrix}$$

Elementarmatrizen

Elementare Matrixumformungen kann man auch durch Multiplikation der umzuformenden Matrix mit einer sogenannten Elementarmatrix beschreiben. Für jeden Umformungstyp gibt es eine Elementarmatrix–Standardkonstruktion, so dass die Multiplikation mit der Elementarmatrix auf der linken Seite die entsprechene Zeilenumformung und die Multiplikation mit der Elementarmatrix auf der rechten Seite die entsprechene Spaltenumformung realisiert.

- **Typ 1:** Für die Vertauschung der i-ten und der j-ten Zeile (Spalte) in einer Matrix A wird eine Matrix $T_{i\,j}$ konstruiert, so dass die Multiplikation $A' = T_{i\,j} \cdot A$ die Vertauschung der i-ten und der j-ten Zeile von A bewirkt und die Multiplikation $A' = A \cdot T_{i\,j}$ die Vertauschung der i-ten und der j-ten Spalte von A bewirkt.

Dazu muss T_{ij} die folgende Form haben:

$$T_{ij} = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & 0 & \\ & & 1 & & & & & & \\ & & & 0 & \cdots & \to & \cdots & 1 & & & \longleftarrow & i\text{-te Zeile} \\ & & & \vdots & 1 & & & \vdots & & & & \\ & & & \downarrow & & \ddots & & \uparrow & & & & \\ & & & \vdots & & & 1 & \vdots & & & & \\ & & & 1 & \cdots & \leftarrow & \cdots & 0 & & & \longleftarrow & j\text{-te Zeile} \\ & & & & & & & & 1 & & & \\ & 0 & & & & & & & & \ddots & & \\ & & & & & & & & & & 1 & \end{pmatrix}$$

- **Typ 2:** Für die Multiplikation einer Zeile (Spalte) mit dem Faktor λ in einer Matrix A wird eine Matrix $S_{i\,\lambda}$ konstruiert, so dass die Multiplikation $A' = S_{i\,\lambda} \cdot A$ die Multiplikation der i-ten Zeile von A mit λ bewirkt und die Multiplikation $A' = A \cdot S_{i\,\lambda}$ die Multiplikation der j-ten Spalte von A mit λ bewirkt.

$$S_{i\,\lambda} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & 0 & \\ & & 1 & & & & \\ & & & \lambda & & & \\ & & & & 1 & & \\ & 0 & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \longleftarrow i\text{-te Zeile}$$

- **Typ 3:** Man konstruiert eine Matrix $K_{i\,j\,\lambda}$, so dass die Multiplikation $A' = K_{i\,j\,\lambda} \cdot A$ die Addition des λ-fachen der j-ten Zeile zur i-ten Zeile von A bewirkt und die Multiplikation $A' = A \cdot K_{i\,j\,\lambda}$ die Addition des λ-fachen der i-ten Spalte zur j-ten Spalte von A bewirkt.

$$K_{i\,j\,\lambda} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & 0 & \\ & & \ddots & & & & \\ & & & \ddots & \lambda & & \\ & & & & \ddots & & \\ & 0 & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \longleftarrow i\text{-te Zeile}$$

$$\uparrow$$

j-te Spalte

1.8 Lineare Gleichungssysteme

Definition: Ein *lineares Gleichungssystem* (LGS) mit Koeffizienten in einem Körper K, mit m Gleichungen und n Unbekannten wird durch eine Matrix $A = (a_{ij})_{(i,j)\in m\times n} \in M(m\times n, K)$ und einem Spaltenvektor $\vec{b} \in K^m$ repräsentiert und wie folgt als Gleichungssystem $(*)$ interpretiert:

$$\begin{array}{ccccccccc} a_{11}\cdot x_1 & + & a_{12}\cdot x_2 & + & \ldots & + & a_{1n}\cdot x_n & = & b_1 \\ a_{21}\cdot x_1 & + & a_{22}\cdot x_2 & + & \ldots & + & a_{2n}\cdot x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}\cdot x_1 & + & a_{m2}\cdot x_2 & + & \ldots & + & a_{mn}\cdot x_n & = & b_m \end{array} \qquad (*)$$

Die Matrixrepräsentation von linearen Gleichungssystemen hat den Vorteil, dass man das Gleichungsystem (*) mit Hilfe des Matrixprodukts auch als eine Vektor–Gleichung beschreiben kann:

$$(*) \qquad \Longleftrightarrow \qquad A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \vec{b}$$

Ein Tupel $(x_1, x_2, \ldots x_n) \in K^n$ ist also genau dann eine Lösung des linearen Gleichungssystems, wenn $A\cdot\vec{x} = \vec{b}$ gilt, wobei $\vec{x}$ die Spaltenvektor–Darstellung des Tupels bezeichnet. Man bezeichnet mit $(A \,|\, \vec{b})$ die Erweiterung der Matrix A mit der zusätzlichen $(n+1)$-ten Spalte $\vec{b}$:

$$(A \,|\, b) = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right)$$

Satz: Das lineare Gleichungssystem $A\cdot\vec{x} = \vec{b}$ ist genau dann lösbar, wenn

$$\text{rg } A = \text{rg}(A \,|\, b)$$

Beweis:

$$\begin{aligned} & \text{rg } A = \text{rg}(A \mid b) \\ \Leftrightarrow\ & \text{Spaltenrang}(A) = \text{Spaltenrang}(A \mid b) \\ \Leftrightarrow\ & \vec{b} \in \text{Lin}\left(\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \ldots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\}\right) \\ \Leftrightarrow\ & \exists x_1, \ldots x_n \in K \qquad \vec{b} = x_1 \cdot \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \ldots + x_n \cdot \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \\ \Leftrightarrow\ & \exists x_1, \ldots x_n \in K \qquad A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \end{aligned}$$

Definition: Ein lineares Gleichungssystem $A \cdot \vec{x} = \vec{b}$ wird *homogenes Gleichungssystem* genannt, wenn der Vektor $\vec{b}$ der Nullvektor ist.

Jedes lineare Gleichungssystem hat ein assoziiertes homogenes Gleichungssystem, welches durch die Ersetzung des Vektors $\vec{b}$ durch den Nullvektor entsteht.

Definition: Die Lösungsmenge eines linaren Gleichungssystems $A \cdot \vec{x} = \vec{b}$ ist die Vektormenge

$$\text{Lös}(A, \vec{b}) = \{\vec{x} \mid A \cdot \vec{x} = \vec{b}\}$$

Satz: Sei $A \in M(m \times n, K)$ die Matrix einer linearen Abbildung $f : K^n \to K^m$ bezüglich der Standardbasis und $\vec{b} \in K^m$, dann ist

a) Die Lösungsmenge $\text{Lös}(A, \vec{0})$ ist gleich Ker f. Damit ist $\text{Lös}(A, \vec{0})$ ein Unterraum von K^n.

b) Sei $\vec{x}, \vec{y} \in \text{Lös}(A, \vec{b})$, dann ist $\vec{x} - \vec{y} \in \text{Lös}(A, \vec{0})$

c) Sei $\vec{x} \in \text{Lös}(A, \vec{b})$ und $\vec{z} \in \text{Lös}(A, \vec{0})$, dann ist $\vec{x} + \vec{z} \in \text{Lös}(A, \vec{b})$

Beweis:

a) Es genügt, die entsprechenden Definition anzuwenden:

$$\begin{aligned} \text{Lös}(A, \vec{0}) &= \{\vec{x} \mid A \cdot \vec{x} = \vec{0}\} \\ &= \{\vec{x} \mid f(\vec{x}) = \vec{0}\} \\ &= \text{Ker } f \end{aligned}$$

b) Seien $\vec{x}, \vec{y} \in \text{Lös}(A, \vec{b})$, dann gilt:

$$f(\vec{x}) = \vec{b} \quad \text{und} \quad f(\vec{y}) = \vec{b}$$

Daraus folgt:

$$f(\vec{x} - \vec{y}) = f(\vec{x}) - f(\vec{y}) = \vec{b} - \vec{b} = \vec{0}$$

Das heißt:

$$\vec{x} - \vec{y} \in \text{Lös}(A, \vec{0})$$

c) Seien $\vec{x} \in \text{Lös}(A, \vec{b})$ und $\vec{z} \in \text{Lös}(A, \vec{0}^{(m)})$, dann gilt:

$$f(\vec{x}) = \vec{b} \quad \text{und} \quad f(\vec{z}) = \vec{0}$$

Daraus folgt:

$$f(\vec{x} + \vec{z}) = f(\vec{x}) - f(\vec{z}) = \vec{b} - \vec{0} = \vec{b}$$

Das heißt:

$$\vec{x} + \vec{z} \in \text{Lös}(A, \vec{b})$$

Beobachtung: Sei $A \in M(m \times n, K)$ die Matrix einer linearen Abbildung $f : K^n \to K^m$ bezüglich der Standardbasis, dann ist die Lösungsmenge $\text{Lös}(A, b)$ genau dann *nicht* leer, wenn $\vec{b} \in \text{Im } f$.

Gauß'scher Algorithmus

Das folgende Lösungsverfahren für lineare Gleichungssysteme geht auf Carl Friedrich Gauß zurück. Es ist auch unter dem Namen Gauß–Elimination bekannt. Die Lösung des Gleichungssystems erfolgt dabei in drei Stufen. In der ersten Stufe wird durch Entwicklung einer oberen Dreiecksform und Rangbetrachtungen festgestellt, ob das System überhaupt eine Lösung hat. Wenn die Antwort positiv ist, wird in der zweiten Stufe eine spezielle Lösung bestimmt. In der dritten Stufe werden alle Lösungen des assoziierten homogenen Systems bestimmt und daraus die komplette Lösungsmenge generiert.
Sei ein Gleichungsystem der Form $A \cdot \vec{x} = \vec{b}$ mit wobei $a \in M(m \times n, K)$ und $\vec{b} \in K^m$ gegeben.

a) Zur Überprüfung, ob das Gleichungssystem eine Lösung hat, wird die Matrix A in obere Dreiecksform gebracht, aber dabei alle Zeilenumformungen auf die erweiterte Matrix $(A \mid b)$ angewendet.

   Achtung: Spaltenvertauschungen in A bedeuten Variablenvertauschung im linearen Gleichungssystem.

   Sei das Ergebnis dieses ersten Schritts das System:

$$A' = \left(\begin{array}{ccccc|ccc|c} a'_{1\,1} & * & * & \cdots & * & * & \cdots & * & b'_1 \\ 0 & a'_{2\,2} & * & \cdots & * & \vdots & \ddots & \vdots & b'_2 \\ 0 & 0 & a'_{3\,3} & \ddots & \vdots & \vdots & \ddots & \vdots & b'_3 \\ \vdots & \vdots & \ddots & \ddots & * & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a'_{r\,r} & * & \cdots & * & b'_r \\ \hline 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 & b'_{r+1} \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 & b'_m \end{array}\right)$$

   - Fall 1: Falls mindestens einer der Werte $b'_{r+1}, b'_{r+2}, \ldots, b'_m$ ungleich Null ist, dann ist rg $(A) <$ rg$(A \mid b)$, und das System hat **keine** Lösung. Das Verfahren wird abgebrochen.
   - Fall 2: Falls $b'_{r+1} = b'_{r+2} = \ldots = b'_m = 0$, dann ist rg $(A) =$ rg$(A \mid b)$, und folglich hat das System hat eine Lösung.

b) Zur Bestimmung einer speziellen Lösung werden die Zeilen $(r+1)$ bis m gestrichen und die Koeffizientenmatrix zwischen der r-ten und $(r+1)$-ten Spalte in zwei Teilmatrizen T und S getrennt:

$$(T \mid S \mid b') = \left(\begin{array}{ccccc|ccc|c} a'_{1\,1} & * & * & \cdots & * & * & \cdots & * & b'_1 \\ 0 & a'_{2\,2} & * & \cdots & * & \vdots & \ddots & \vdots & b'_2 \\ 0 & 0 & a'_{3\,3} & \ddots & \vdots & \vdots & \ddots & \vdots & b'_3 \\ \vdots & \vdots & \ddots & \ddots & * & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a'_{r\,r} & * & \cdots & * & b'_r \end{array}\right)$$

Das Gleichungssystem nimmt dadurch die folgende Gestalt an:

$$(T \mid S) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix} = T \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} + S \cdot \begin{pmatrix} x_{r+1} \\ x_{r+2} \\ \vdots \\ x_n \end{pmatrix} = \vec{b}'$$

Man setzt $x_{r+1} = x_{r+2} = \ldots = x_n = 0$ und reduziert das System dadurch auf:

$$T \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} = \vec{b}' \quad \Rightarrow \quad \begin{pmatrix} a'_{1\,1} & a'_{1\,2} & \cdots & a'_{1\,r} \\ 0 & a'_{2\,2} & \cdots & a'_{2\,r} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a'_{r\,r} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} = \begin{pmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_r \end{pmatrix}$$

Die Werte von $x_1, x_2, \ldots x_r$ können nun direkt bestimmt werden:

$$\begin{array}{rclcrcl} b'_r & = & a'_{r\,r} \cdot x_r & \Rightarrow & x_r & = & \frac{b'_r}{a'_{r\,r}} \\ b'_{r-1} & = & a'_{r-1\,r-1} \cdot x_{r-1} + a'_{r-1\,r} \cdot x_r & \Rightarrow & x_{r-1} & = & \frac{b'_{r-1} - a'_{r-1\,r} \cdot x_r}{a'_{r-1\,r-1}} \\ & & & \vdots & & & \\ b'_1 & = & a'_{1\,1} \cdot x_1 + \ldots + a'_{1\,r} \cdot x_r & \Rightarrow & x_1 & = & \frac{b'_1 - a'_{1\,r} \cdot x_r - \ldots - a'_{1\,2} \cdot x_2}{a'_{1\,1}} \end{array}$$

Somit wurde eine spezielle Lösung des linearen Gleichungssystem berechnet, die im Weitern mit $\vec{v}$ bezeichnet wird:

$$\vec{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

c) Auf Grund der Vorüberlegungen wissen wir, dass die Lösungsmenge des homogenen Gleichungssystems ein Vektorraum der Dimension $n - r$ ist. Man erhält den j-ten Basisvektor dieses Raums ($1 \leq j \leq n - r$) indem die Variablen $x_{r+1}, \ldots, x_n$ jeweils mit den folgenden Werten belegt werden:

$$x_{r+j} = 1 \quad \text{und} \quad x_{r+1} = \ldots = x_{r+j-1} = x_{r+j+1} = \ldots = x_n = 0$$

Bezeichnet man die Koeffizienten der rechten Teilmatrix S durch

$$S = (s_{i\,j})_{(i,j) \in r \times (n-r)}$$

so nimmt das Gleichungsytem die folgende Form an

$$(T \mid S)\cdot\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a'_{1\,1} & a'_{1\,2} & \cdots & a'_{1\,r} \\ 0 & a'_{2\,2} & \cdots & a'_{2\,r} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a'_{r\,r} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} + \begin{pmatrix} s_{1\,1} & \cdots & s_{1\,n-r} \\ \vdots & \ddots & \vdots \\ s_{r\,1} & \cdots & s_{r\,r} \end{pmatrix} \cdot \begin{pmatrix} x_{r+1} \\ x_{r+2} \\ \vdots \\ x_n \end{pmatrix} = \vec{0}$$

Berücksichtig man die speziellen Werte $x_{r+j} = 1$ und $x_{r+1} = \ldots = x_{r+j-1} = x_{r+j+1} = \ldots = x_n = 0$, ergibt sich das folgende lineare Gleichungssystem

$$\begin{pmatrix} a'_{1\,1} & a'_{1\,2} & \cdots & a'_{1\,r} \\ 0 & a'_{2\,2} & \cdots & a'_{2\,r} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a'_{r\,r} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} = \begin{pmatrix} -s_{1\,j} \\ -s_{2\,j} \\ \vdots \\ -s_{r\,j} \end{pmatrix}$$

Die Werte von $x_1, x_2, \ldots x_r$ können nun wie bei der speziellen Lösung bestimmt werden.

Das Verfahren muss für alle $n-r$ Spalten von S durchgeführt werden. Sei $\vec{u}_j$ der dabei berechnete j-te Basisvektor des Lösungsraums des homogenen Gleichungssystems.

Bleibt nur noch, die Lösungsmenge des linearen Gleichungssystem aus der speziellen Lösung und der Lösungsmenge des homogenen Systems zusammenzusetzen:

$$\text{Lös}(A \mid b) = \text{Lös}(T \mid S \mid b') = \left\{ \vec{v} + \sum_{j=1}^{n-r} \lambda_j \cdot \vec{u}_j \mid \lambda_1, \lambda_2, \ldots \lambda_{n-r} \in \mathbb{R} \right\}$$

Beispiel: Gegeben sei das folgende Gleichungssystem:

$$\begin{array}{rcrcrcrcr} & & 2x_2 & + & x_3 & - & x_4 & = & 6 \\ x_1 & - & x_2 & + & 2x_3 & & & = & -1 \\ 2x_1 & & & + & 5x_3 & & & = & 3 \\ - & x_1 & - & x_2 & - & 3x_3 & + & 2x_4 & = & -6 \end{array}$$

a) Die dazugehörige Matrix

$$(A \mid \vec{b}) = \left(\begin{array}{rrrr|r} 0 & 2 & 1 & -1 & 6 \\ 1 & -1 & 2 & 0 & -1 \\ 2 & 0 & 5 & 0 & 3 \\ -1 & -1 & -3 & 2 & -6 \end{array}\right)$$

Diese Matrix muss zuerst in obere Dreiecksform überführt werden.

Erste und zweite Zeile vertauschen:

$$\left(\begin{array}{rrrr|r} 1 & -1 & 2 & 0 & -1 \\ 0 & 2 & 1 & -1 & 6 \\ 2 & 0 & 5 & 0 & 3 \\ -1 & -1 & -3 & 2 & -6 \end{array}\right)$$

In der ersten Spalte unter $a_{1\,1}$ Nullen erzeugen:

$$\left(\begin{array}{rrrr|r} 1 & -1 & 2 & 0 & -1 \\ 0 & 2 & 1 & -1 & 6 \\ 0 & 2 & 1 & 0 & 5 \\ 0 & -2 & -1 & 2 & -7 \end{array}\right)$$

In der zweiten Spalte unter $a_{2\,2}$ Nullen erzeugen:

$$\left(\begin{array}{rrrr|r} 1 & -1 & 2 & 0 & -1 \\ 0 & 2 & 1 & -1 & 6 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

Dritte und vierte Spalte tauschen, um an der Stelle $a_{3\,3}$ einen Wert $\neq 0$ zu erzeugen (Achtung: $x_3 \leftrightarrow x_4$):

$$\left(\begin{array}{rrrr|r} 1 & -1 & 0 & 2 & -1 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \end{array}\right)$$

In der dritten Spalte unter $a_{3\,3}$ Nullen erzeugen:

$$\left(\begin{array}{rrr|r|r} 1 & -1 & 0 & 2 & -1 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 1 & 0 & -1 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Die Matrix hat nun obere Dreiecksform.

Es existiert eine Lösung, da der untere Teil von $\vec{b}$ aus einer Null besteht. Die Matrix kann nun folgendermaßen reduziert werden:

$$\left(\begin{array}{rrr|r|r} 1 & -1 & 0 & 2 & -1 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 1 & 0 & -1 \end{array}\right)$$

Das lineare Gleichungssystem wird geteilt:

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_4 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \cdot (x_3) = \begin{pmatrix} -1 \\ 6 \\ -1 \end{pmatrix}$$

b) Bestimmung der speziellen Lösung:

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1 \\ 6 \\ -1 \end{pmatrix}$$

Wähle $x_3 = 0$ und bestimme die übrigen Variablen:

$$\begin{array}{rcrcrcrcl} & & & & 1 \cdot x_4 & = & -1 & \Rightarrow & x_4 = -1 \\ & & 2 \cdot x_2 & + & (-1) \cdot x_4 & = & 6 & \Rightarrow & x_2 = 2,5 \\ 1 \cdot x_1 & + & (-1) \cdot x_2 & + & 0 \cdot x_4 & = & -1 & \Rightarrow & x_1 = 1,5 \end{array}$$

c) Bestimmung des ersten (und einzigen) Basisvektors von $(A \mid \vec{0}^{(4)})$:

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ 0 \end{pmatrix}$$

Wähle $x_3 = 1$ und bestimme die übrigen Variablen:

$$\begin{array}{rcrcrcrcr} & & & & 1 \cdot x_4 & = & 0 & \Rightarrow & x_4 = 0 \\ & & 2 \cdot x_2 & + & (-1) \cdot x_4 & = & -1 & \Rightarrow & x_2 = -0,5 \\ 1 \cdot x_1 & + & (-1) \cdot x_2 & + & 0 \cdot x_4 & = & -2 & \Rightarrow & x_1 = -2,5 \end{array}$$

d) Lösungsmenge:

$$\text{Lös}(A \mid b) = \left\{ \begin{pmatrix} 1,5 \\ 2,5 \\ 0 \\ -1 \end{pmatrix} + \lambda \cdot \begin{pmatrix} -2,5 \\ -0,5 \\ 1 \\ 0 \end{pmatrix} \;\middle|\; \lambda \in \mathbb{R} \right\}$$

Quotientenräume

Wir haben gesehen, dass die Lösungsmenge eines linearen Gleichungssystems in der Regel ist kein Unterraum, sondern eine „Verschiebung“ eines Unterraums ist. Durch die folgende Begriffsbildung kann man eine etwas allgemeinere Sicht auf diese Konstruktion gewinnen.

Definition: Sei V ein Vektorraum, U ein Unterraum von V und $\vec{v}$ ein Vektor aus V, dann nennt man

$$\vec{v} + U = \{\vec{v} + \vec{u} \mid \vec{u} \in U\}$$

die *Nebenklasse* von $\vec{v}$ bezüglich U.

Satz: Sei V ein Vektorraum, U ein Unterraum von V und $\vec{w}, \vec{w}$ zwei Vektoren aus V, dann gilt:

$$\vec{v} + U = \vec{w} + U \quad \Leftrightarrow \quad \vec{v} - \vec{w} \in U \quad \Leftrightarrow \quad \vec{w} \in \vec{v} + U$$

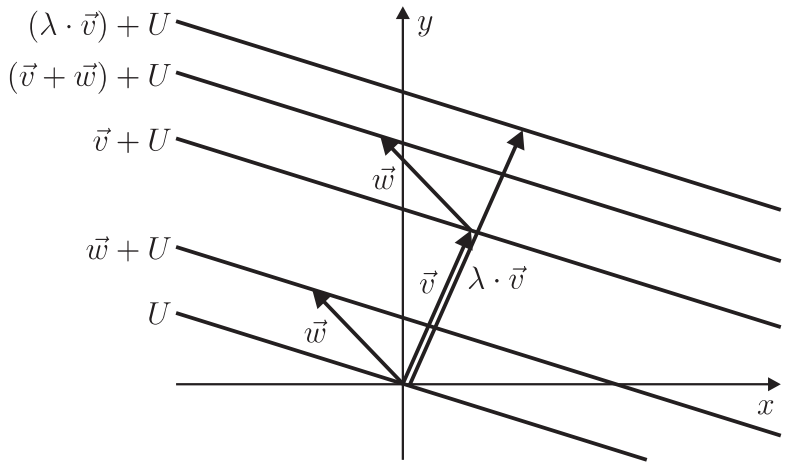
Definition: Sei V ein Vektorraum, U ein Unterraum von V und K der Körper von V, dann bezeichnet man die Menge

$$V_{/U} = \{\vec{v} + U \mid \vec{v} \in V\}$$

als *Quotientenraum* von V nach U.

Beobachtung: Der Quotientenraum $V_{/U}$ ist ein Vektorraum mit den Operationen

$$\begin{array}{rcl} (\vec{v} + U) + (\vec{w} + U) & = & (\vec{v} + \vec{w}) + U \\ \lambda \cdot (\vec{v} + U) & = & (\lambda \cdot \vec{v}) + U \\ \text{und dem neutralen Element} & & \vec{0} + U = U. \end{array}$$



Satz: Sei V ein Vektorraum, U ein Unterraum von V, dann ist

$$\dim V_{/U} = \dim V - \dim U$$

Beweis: Man definiert eine Abbildung $\varphi : V \to V_{/U}$ durch

$$\vec{v} \mapsto \vec{v} + U$$

Die Abbidlung φ ist linear und surjektiv (d.h. Im $\varphi = V_{/U}$) und Ker $\varphi = U$, denn

$$\vec{v} + U = U \iff \vec{v} \in U$$

Dann folgt aus der Dimensionsformel:

$$\dim V = \dim(\text{Ker } \varphi) + \dim(\text{Im } \varphi) = \dim U + \dim V_{/U}$$

Ist $A \in M(m \times n, K)$ eine Matrix und $\vec{b} \in K^n$ ein Spaltenvektor, dann ist die Lösungsmenge $U = \text{Lös}(A, \vec{0})$ des homogenen Gleichungssystems ist ein Unterraum von $V = K^n$.
Die Lösungsmenge Lös$(A, \vec{b})$ des linearen Gleichungssystems $A \cdot \vec{x} = \vec{b}$ ist eine Nebenklasse von U. Die Menge der Lösungsmengen bildet den Quotientenraum $V_{/U}$ in dem die Operationen die folgenden zusätzlichen Eigenschaften haben:

- Addition:

  Ist Lös$(A, \vec{b}) = \vec{x} + U$ und Lös$(A, \vec{c}) = \vec{y} + U$, dann ist $(\vec{x} + \vec{y}) + U = \text{Lös}(A, \vec{b} + \vec{c})$

- Multiplikation mit Skalaren:

  $$\text{Ist Lös}(A \,|\, \vec{b}) = \vec{x} + U \text{ und } \lambda \in K, \text{ dann ist } \lambda\,\vec{x} + U = \text{Lös}(A \,|\, \lambda\,\vec{b})$$

1.9 Inverse Matrizen

Der Ring $M(n \times n, K)$

Die $(n \times n)$–Matrizen über einen Körper K bilden einen Vektorraum und damit eine kommutative Gruppe bezüglich der Addition. Wie wir bereits wissen, ist die Multiplikation von Matrizen aus $M(n \times n, K)$ assoziativ. Darüber hinaus gibt es mit der Einheitsmatrix E_n ein neutrales Element:

$$E_n = \begin{pmatrix} \mathbf{1} & 0 & \cdots & 0 \\ 0 & \mathbf{1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \mathbf{1} \end{pmatrix} \in M(n \times n, K),$$

denn für jede Matrix $A \in M(n \times n, K))$ gilt $E_n \cdot A = A = A \cdot E_n$. Da außerdem das Assoziativitätsgesetz gilt, bilden die $(n \times n)$–Matrizen einen sogenannten Ring. Diese Struktur ist ähnlich zu der eines Körpers, aber die Multiplikation ist nicht notwendigerweise kommutativ und es muss nicht für jedes nicht-Null Element ein inverses Element bezüglich der Multiplikation geben, welche wie folgt definiert sind.

Definition: Sei $A \in M(n \times n, K)$, dann ist A^{-1} die zu A inverse Matrix, wenn

$$A \cdot A^{-1} = E_n = A^{-1} \cdot A$$

Man nennt eine Matrix A invertierbar, wenn eine zu A inverse Matrix A^{-1} existiert.

Satz: Eine Matrix $A \in M(n \times n, K)$ ist genau dann invertierbar, wenn rg $A = n$.

Beweis: Man beweist diesen Satz, indem man die Bijektion zwischen Matrizen und linearen Abbildungen verwendet und dann die Dimensioinsformel für lineare Abbildungen anwendet. Ist $f \in \mathrm{Hom}(K^n, K^n)$ die zur Matrix $A \in M(n \times n, K)$ assoziierte lineare Abbildung, dann gilt die folgende Kette von Äquivalenzen:

$$\begin{aligned}
\text{rg } A = n &\iff \dim(\text{Im } f) = n \\
&\iff \dim(\text{Im } f) = n \text{ und } \dim(\text{Ker } f) = 0 \\
&\iff f \text{ ist surjektiv und injektiv} \iff f \text{ ist bijektiv} \\
&\iff \text{Es gibt für } f \text{ eine Umkehrfunktion } g \text{ mit } fg = Id_{K^n} \text{ und } gf = Id_{K^n} \\
&\iff \text{Für die zu } g \text{ assoziierte Matrix } B \text{ gilt } AB = E_n \text{ und } BA = E_n \\
&\iff A \text{ ist invertierbar}
\end{aligned}$$

Da man nun mit der Umwandlung in obere Dreiecksform ein einfaches Mittel zur Verfügung hat, das Entscheidungsproblem der Invertierbarkeit einer Matrix zu beantworten, bleibt noch die Frage offen, wie man für eine invertierbare Matrix A die inverse Matrix berechnen kann.
Wir werden dazu mehrere Verfahren kennen lernen. Zuerst überlegen wir uns, wie man dieses Problem als lineares Gleichungssystem formulieren kann, danach werden wir den Algorithmus zur Entwicklung der oberen Dreiecksform zu einem Verfahren zur Matrixinvertierung modifizieren

Bestimmung der inversen Matrix mit einem LGS

Eine einfache Überlegung zu der mit A assoziierten linearen Abbildung f führt zu einer ersten, naiven Methode zum Invertieren einer Matrix. Wenn eine inverse Matrix existiert, korrespondiert Sie zur inversen linearen Abbildung f^{-1}. Damit muss in der j-ten Spalte von A^{-1} das Bild des j-ten Basisvektors unter f^{-1} stehen, also der Spaltenvektor $f^{-1}(e_j)$. Dieser Vektor ist aber gerade die (eindeutige!) Lösung des linearen Gleichungssystems $A \cdot \vec{x} = \vec{e_j}$. Man kann also die inverse Matriv durch Lösung von n Gleichungssystemen bestimmen.

Bestimmung der inversen Matrix mit elementaren Umformungen

Dieses zweite Verfahren basiert auf einer Reihe von einfachen Beobachtungen:

a) Seien $A, B, C \in M(n \times n, K)$ Matrixen und $A \cdot B = C$. Überführt man mit den gleichen elementaren Zeilenumformungen A in A' und C in C' (B bleibt unverändert), so gilt $A' \cdot B = C'$.
Begründung: Jede Zeilenumformungen kann durch Multiplikation von links mit einer entsprechenden Elementarmatrix ausgeführt werden:

$$A' \cdot B = \underbrace{D_k \cdot \ldots \cdot D_2 \cdot D_1}_{\text{Elementarmatrizen}} \cdot A \cdot B = \underbrace{D_k \cdot \ldots \cdot D_2 \cdot D_1}_{\text{Elementarmatrizen}} \cdot C = C'$$

b) Ist eine Matrix $A \in M(n \times n, K)$ invertierbar, so kann man A mit elementaren Zeilenumformungen in E_n überführen.
Begründung: Da A den vollen Rang n hat, sind nach Überführung in eine obere Dreiecksform alle Diagonalelemente ungleich Null und man kann schrittweise mit Unformungen vom Typ 3 alle Elemente über der Diagonale in Nullen verwandeln und letzlich mit Unformungen vom Typ 2 alle Diagonalelemente in Einsen verwandeln.

c) Überführt man die Matrix A durch Zeilenumformungen in E_n und wendet die gleichen Umformungen auf E_n an, so erhält man A^{-1}.

Zur Begründung wendet man die Beobachtung a) an. Wir betrachten die Folge von Zeilenumformungen, die A in E_n überführen. Sei X die Matrix, die man durch Anwendung der gleichen Folge von Umformungen auf E_n erhält. Wir wollen zeigen, dass $X = A^{-1}$ ist. Das folgt aber aus der Anwendung von a) in der folgenden Situation:

$$A = A \qquad B = A^{-1} \qquad \text{und} \qquad C = E_n$$

Offensichtlich ist mit $AB = C$ die Voraussetzung erfüllt und wir haben

$$\begin{aligned} A &\rightsquigarrow A' = E_n \\ C = E_n &\rightsquigarrow C' = X \end{aligned}$$

Nach a) ist dann $A'B = C'$, damit $E_n \cdot A^{-1} = X$ und letztlich $A^{-1} = X$

Wir demonstrieren das beprochene Verfahren an einem Beispiel:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \\ 0 & -1 & 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_n$$

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & -1 & 2 \\ 0 & -1 & 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} -1 & 2 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 4 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} -1 & 2 & 0 \\ -2 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} -3 & 4 & -2 \\ -2 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

$$E_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -3 & 4 & -2 \\ 2 & -2 & 1 \\ 0,5 & -0,5 & 0,5 \end{pmatrix} = A^{-1}$$

Probe:

$$\begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \\ 0 & -1 & 4 \end{pmatrix} \cdot \begin{pmatrix} -3 & 4 & -2 \\ 2 & -2 & 1 \\ 0,5 & -0,5 & 0,5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

1.10 Determinanten

Begriffseinführung für den Anwender

Definition: Die *Determinante* $\det A$ ist eine Kenngröße einer quadratischen Matrix $A \in M(n \times n, K)$, die man rekursiv bestimmen kann:

- Fall 1: $n = 1$

$$\det a_{1\,1} = a_{1\,1}$$

- Fall 2: $n > 1$

  Entwicklung nach der ersten Spalte:

$$\det A = \sum_{i=1}^{n} (-1)^{i+1} a_{i\,1} \cdot \det A_{i\,1}$$

  Dabei ist $A_{i\,1}$ die Matrix, die man aus A durch Streichen der i-ten Zeile und der ersten Spalte enthält.

An Stelle der Schreibweise $\det A$ kann man die Matrix auch mir Betragsstrichen begrenzen:

$$\det \begin{pmatrix} a_{1\,1} & \cdots & a_{1\,n} \\ \vdots & \ddots & \vdots \\ a_{n\,1} & \cdots & a_{n\,n} \end{pmatrix} = \begin{vmatrix} a_{1\,1} & \cdots & a_{1\,n} \\ \vdots & \ddots & \vdots \\ a_{n\,1} & \cdots & a_{n\,n} \end{vmatrix}$$

Beispiel:

$$\begin{aligned} \begin{vmatrix} \mathbf{0} & 1 & 2 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{4} & \mathbf{6} \\ \mathbf{0} & 1 & 5 & 1 \\ \mathbf{0} & 0 & 2 & 0 \end{vmatrix} &= (-1)^{2+1} \cdot 1 \cdot \begin{vmatrix} 1 & 2 & 0 \\ 1 & 5 & 1 \\ 0 & 2 & 0 \end{vmatrix} \\ &= -\left(1 \cdot \begin{vmatrix} 5 & 1 \\ 2 & 0 \end{vmatrix} - 1 \cdot \begin{vmatrix} 2 & 0 \\ 2 & 0 \end{vmatrix} \right) \\ &= -\left((5 \cdot 0 - 2 \cdot 1) - (2 \cdot 0 - 2 \cdot 0) \right) \\ &= 2 \end{aligned}$$

Beobachtung: Für die Spezialfälle $n = 2$ und $n = 3$ ergibt sich aus der angegebenen Formel ein einfaches Schema zur Bestimmung der Determinanten, die sogenannte Regel von Sarrus:

$$\begin{vmatrix} a_{1\,1} & a_{1\,2} \\ a_{2\,1} & a_{2\,2} \end{vmatrix} = a_{1\,1} \cdot a_{2\,2} - a_{1\,2} \cdot a_{2\,1}$$

$$\begin{aligned} \begin{vmatrix} a_{1\,1} & a_{1\,2} & a_{1\,3} \\ a_{2\,1} & a_{2\,2} & a_{2\,3} \\ a_{3\,1} & a_{3\,2} & a_{3\,3} \end{vmatrix} &= (a_{1\,1} \cdot a_{2\,2} \cdot a_{3\,3} + a_{1\,2} \cdot a_{2\,3} \cdot a_{3\,1} + a_{1\,3} \cdot a_{2\,1} \cdot a_{3\,2}) \\ &\quad - (a_{1\,3} \cdot a_{2\,2} \cdot a_{3\,1} + a_{1\,1} \cdot a_{2\,3} \cdot a_{3\,2} + a_{1\,2} \cdot a_{2\,1} \cdot a_{3\,3}) \end{aligned}$$

Das Schema für (3×3)–Matrizen erhält man, indem die erste und die zweite Spalte noch einmal auf der rechten Seiten der Matrix angehängt werden:

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

Alle Summanden mit positiven Vorzeichen ergeben sich dann als Produkte der Werte auf den Diagonalen von links oben nach rechts unten und alle Summanden mit negativen Vorzeichen ergeben sich dann als Produkte der Werte auf den Diagonalen von rechts oben nach links unten.

Achtung: Ab $n = 4$ funktioniert dieses Schema nicht mehr!

Begriffseinführung für den Mathematiker

Üblicherweise werden Determinanten durch den folgenden Satz eingeführt, dessen Beweis aber technisch sehr aufwändig ist.

Satz: Es gibt genau eine Abbildung $\det : M(n\times n, K) \to K$ mit den folgenden Eigenschaften:

1. Die Abbildung det ist linear in jeder Zeile.
2. Wenn $\operatorname{rg} A < n$, dann gilt $\det A = 0$.
3. Für die Einheitsmatrix E_n gilt: $\det E_n = 1$.

Diese Abbildung lässt sich durch die am Anfang angegebene Entwicklungsformel bestimmen.

Dabei bezieht sich der Begriff linear jeder Zeile zu sein auf zwei Matrizen A und A' die sich nur in einer (der i-ten) Zeile unterscheiden und an allen anderen Stellen identisch sind:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \qquad \textit{und} \qquad A' = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a'_{i1} & \cdots & a'_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

Die Determinante muss dann die folgenden Bedingungen erfüllen:

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} + a'_{i1} & \cdots & a_{in} + a'_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \det A + \det A' \quad \text{und} \quad \det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ \lambda a_{i1} & \cdots & \lambda a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \lambda \cdot \det A$$

Wir verzichten auf einen vollständigen Beweis und werden nur einige Ideen ableiten, die sich unmittelbar aus den drei Eigenschaften ergeben. Zuerst untersuchen wir das Verhalten bei elementaren Zeilenumformungen.

- Bei Typ 1 Umformungen, d.h. Multiplikation einer Zeile mit einem Skalar λ, folgt schon aus der ersten Eigenschaft, dass die Determinante der alten Matrix auch mit λ multipliziert werden muss.

- Jede Abbildung det mit den Eigenschaften 1) bis 3) ist invariant bei Zeilenumformungen vom Typ 3, d.h. die Determinante bleibt bei solchen Umformungen gleich:

$$\det \begin{pmatrix} a_{1\,1} & \cdots & a_{1\,n} \\ \vdots & & \vdots \\ a_{j\,1} & \cdots & a_{j\,n} \\ \vdots & & \vdots \\ a_{i\,1} + \lambda a_{j\,1} & \cdots & a_{i\,n} + \lambda a_{j\,n} \\ \vdots & & \vdots \\ a_{n\,1} & \cdots & a_{n\,n} \end{pmatrix} = \det \begin{pmatrix} a_{1\,1} & \cdots & a_{1\,n} \\ \vdots & & \vdots \\ a_{j\,1} & \cdots & a_{j\,n} \\ \vdots & & \vdots \\ a_{i\,1} & \cdots & a_{i\,n} \\ \vdots & & \vdots \\ a_{n\,1} & \cdots & a_{n\,n} \end{pmatrix} + \det \begin{pmatrix} a_{1\,1} & \cdots & a_{1\,n} \\ \vdots & & \vdots \\ a_{j\,1} & \cdots & a_{j\,n} \\ \vdots & & \vdots \\ \lambda a_{j\,1} & & \lambda a_{j\,n} \\ \vdots & & \vdots \\ a_{n\,1} & \cdots & a_{n\,n} \end{pmatrix}$$

$$= \det A + 0$$

Dabei ist die zweite Determinante gleich Null, weil in der Matrix die j-te und die i-te Zeile linear abhängig sind und folglich der Rang kleiner als n ist.

- Jede elementare Zeilenumformung vom Typ 1 (Vertauschung von zwei Zeilen) bewirkt die Änderung des Vorzeichens der Determinate bei gleichbleibendem Betrag. Man kann eine solche Umformung durch drei Umformungen vom Typ 3 und eine Umformung vom Typ 1 simulieren:

$A^{(0)}$		Beginn mit Matrix A
$A^{(1)}$	Zeile i	:= Zeile i+ Zeile j
$A^{(2)}$	Zeile j	:= Zeile j− Zeile i = − Zeile i von A
$A^{(3)}$	Zeile i	:= Zeile i+ Zeile j = Zeile j von A
$A^{(4)}$	Zeile j	:= $(-1)\cdot$ Zeile j = Zeile i von A

Die Umformungen (1) bis (3) sind vom Typ 3 und ändern die Determinate nicht, die letzte Umformung ändert das Vorzeichen.

Folgerung: Die Determinante kann als Produkt der Diagonalelemente einer Matrix in oberer Dreiecksform berechnet werden, wobei man die Überführung in diese Form nur durch Zeilenumformungen realisieren und für jeden Zeilentausch zusätzlich mit (-1) multiplizieren muss.

Beispiel:

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 2 & 5 & 3 \end{pmatrix} \mapsto \underbrace{\begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 3 & -1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 2 \\ 0 & 3 & -1 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{Zeilentausch}}$$

und folglich ist $\det A = (-1)\cdot 1\cdot 3\cdot 1 = -3$.

Man beachte, dass wir bisher die Kernaussage des Satzes, nämlich dass eine solche Abbildung det überhaupt existiert, noch nicht bewiesen haben, sondern nur unter der Annahme, dass die Abbildung existiert, einige nützliche Eigenschaften nachgewiesen haben. Der Kernbeweis erfordert einigen technischen Aufwand, den wir hier vermeiden werden. Als Nebenprodukt

ergibt sich dabei auch die Tatsache, dass die Determinante durch Entwicklung nach einer beliebigen Spalte oder nach einer beliebigen Zeile berechent werden kann:

$$\begin{aligned} \det A &= \sum_{i=1}^{n} (-1)^{i+k} a_{i\,k} \det A_{i\,k} && \text{Entwicklung nach Spalte } k \\ &= \sum_{j=1}^{n} (-1)^{l+j} a_{l\,j} \det A_{l\,j} && \text{Entwicklung nach Zeile } l \end{aligned}$$

Die Regel, welches Vorzeichen für welchen Summanden verwendet werden muss, kann man sich als Schachbrettmuster einprägen:

$$\begin{pmatrix} + & - & + & \dots \\ - & + & - & \dots \\ + & - & + & \dots \\ \vdots & \vdots & \vdots & \ddots \\ & & & \end{pmatrix}$$

Wie das folgende Beispiel zeigt, kann man diese Eigenschaften sehr gut zur Vereinfachung der Determinantenberechnung nutzen, indem man vorrangig nach Zeilen bzw. Spalten mit vielen Nullen entwickelt:

$$\begin{aligned} \begin{vmatrix} 7 & 3 & \mathbf{0} & -1 \\ 2 & 4 & \mathbf{0} & 5 \\ 8 & -5 & \mathbf{2} & 4 \\ 2 & 1 & \mathbf{0} & 0 \end{vmatrix} &= (-1)^{3+3} \cdot 2 \cdot \begin{vmatrix} 7 & 3 & -1 \\ 2 & 4 & 5 \\ \mathbf{2} & \mathbf{1} & \mathbf{0} \end{vmatrix} \\ &= 2 \cdot \left(2 \cdot \begin{vmatrix} 3 & -1 \\ 4 & 5 \end{vmatrix} - 1 \cdot \begin{vmatrix} 7 & -1 \\ 2 & 5 \end{vmatrix} \right) \\ &= 2 \cdot (2 \cdot 19 - 1 \cdot 37) = 2 \end{aligned}$$

Letztlich kann man aus den oben genannten Fakten auch die folgende Beobachtung ableiten.
Folgerung: Die Determinanten einer Matrix A und der zu A transponierten Matrix A^t sind gleich:

$$\det A = \det A^t$$

Anwendungen von Determinanten

Determinanten haben sich als ein äußert nützliches Werkzeug für vielfältige Anwendungen (in der Linearen Algebra und darüber hinaus) erwiesen. Wir werden uns hier mit drei Anwendungsfeldern genauer beschäftigen:

1. Lösung von (speziellen) linearen Gleichungssystemen (Cramersche Regel)
2. Geometrische Anwendungen
3. Invertierung von Matrizen

Cramersche Regel

Sei $A \in M(n \times n, K)$ eine Matrix mit rg $A = n$ und $\vec{b} \in K^n$ ein Vektor. Dann hat dat das lineare Gleichungssystem $A \cdot \vec{x} = \vec{b}$ eine eindeutige Lösung, die man wie folgt bestimmen kann:

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{mit} \quad x_i = \frac{\det A_i}{\det A} \quad \text{für alle } 1 \leq i \leq n$$

Dabei ist A_i die Matrix, die man erhält, wenn man die i-te Spalte von A durch $\vec{b}$ ersetzt.

Beispiel: Es ist Lösung des folgenden linearen Gleichungssystem zu bestimmen:

$$\begin{array}{rcrcl} 2x_1 & + & 3x_2 & = & 5 \\ x_1 & - & 2x_2 & = & 2 \end{array} \quad \leftrightarrow \quad \begin{pmatrix} 2 & 3 \\ 1 & -2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$$

Anwendung der Cramerschen Regel:

$$x_1 = \frac{\begin{vmatrix} 5 & 3 \\ 2 & -2 \end{vmatrix}}{\begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix}} = \frac{-10-6}{-4-3} = \frac{16}{7}$$

$$x_2 = \frac{\begin{vmatrix} 2 & 5 \\ 1 & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix}} = \frac{4-5}{-4-3} = \frac{1}{7}$$

Geometrische Anwendungen I: Baryzentrische Koordinaten

Im Folgenden werden Punkte in der Ebene $\mathbb{R}^2$ und im Raum $\mathbb{R}^3$ betrachtet. Wir werden die Punkte mit Großbuchstaben und die zugehörigen Ortsvektoren mit den entsprechenden Kleinbuchstaben bezeichnen.

Seien $P, Q, R \in \mathbb{R}^2$ drei Punkte in der Ebenene, die nicht auf einer Geraden liegen, dann kann man den Ortsvektor $\vec{t}$ eines beliebigen Punkts $T \in \mathbb{R}^2$ eindeutig als Linearkombination

$$\vec{t} = a \cdot \vec{p} + b \cdot \vec{q} + c \cdot \vec{r} \quad \text{mit} \quad a + b + c = 1$$

darstellen. Die Koeffizienten a, b und c nennt man die *baryzentrischen Koordinaten* oder auch *Schwerpunktskoordinaten* von T bezüglich P, Q und R. Man kann diese Koordinaten als Lösung eines linearen Gleichungssystems berechnen:

$$\begin{array}{rrrl} p_x \cdot a + & q_x \cdot b + & r_x \cdot c & = t_x \\ p_y \cdot a + & q_y \cdot b + & r_y \cdot c & = t_y \\ a + & b + & c & = 1 \end{array}$$

Die Voraussetzung, dass P, Q und R nicht auf einer Geraden liegen, sorgt dafür, dass dieses System eine eindeutige Lösung hat, die man mit der Cramerschen Regel finden kann. Wir

verwenden dazu eine spezielle Funktion, die wir mit pdet (abgekürzt für Punktdeterminante) bezeichnen:

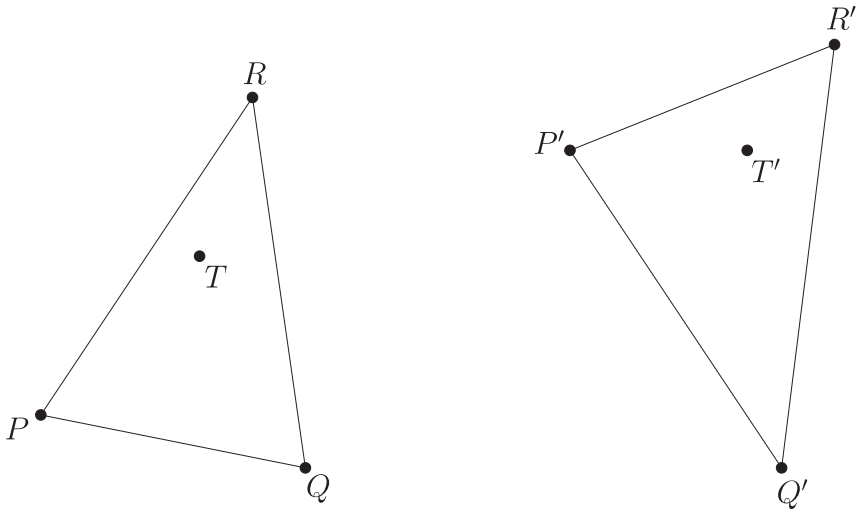
$$\begin{aligned} \text{pdet} \;:\; \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \;&\to\; \mathbb{R} \\ \text{pdet}\,(\vec{p},\,\vec{q},\,\vec{r}) \;&:=\; det \begin{pmatrix} p_x & q_x & r_x \\ p_y & q_y & r_y \\ 1 & 1 & 1 \end{pmatrix} \end{aligned}$$

Aus der Cramerschen Regel ergeben sich die baryzentrischen Koordinaten wie folgt:

$$a = \frac{\text{pdet}(\vec{t},\vec{q},\vec{r})}{\text{pdet}(\vec{p},\vec{q},\vec{r})} \qquad b = \frac{\text{pdet}(\vec{p},\vec{t},\vec{r})}{\text{pdet}(\vec{p},\vec{q},\vec{r})} \qquad c = \frac{\text{pdet}(\vec{p},\vec{q},\vec{t})}{\text{pdet}(\vec{p},\vec{q},\vec{r})}$$

Eine wichtige Eigenschaft der baryzentrischen Koordinaten besteht darin, dass der Punkt T genau dann in dem von P, Q und R aufgespannten Dreieck liegt, wenn alle baryzentrischen Koordinaten von T im Intervall $[\,0\,,\,1\,]$ liegen. Der Punkt T liegt genau dann auf dem Rand des Dreiecks, wenn eine baryzentrischen Koordinate gleich Null ist und die anderen beiden in $[\,0\,,\,1\,]$ liegen.

Mit Hilfe der baryzentrischen Koordinaten kann man auch Bildverzerrungen wie das in der Computergrafik verwendete *Warping* realisieren. Dabei wird vorausgesetzt, dass über ein gegebenes Bild ein Dreiecksgitter gelegt ist und die verzerrte Abbildung der Gitterpunkte bekannt ist. Zu bestimmen ist, wohin die inneren Punkte der Dreiecke abzubilden sind. Dazu berechnet man die baryzentrischen Koordinaten des abzubildenden Punktes T in einem Dreieck $\Delta(P, Q, R)$ und definiert den Bildpunkt T' als Punkt mit den gleichen baryzentrischen Koordinaten im verzerrten Dreieck $\Delta(P', Q', R')$.



Geometrische Anwendungen II: Dreiecksfläche und Volumen eines Simplexes
Seien $\vec{p}, \vec{q}, \vec{r} \in \mathbb{R}^2$ die Ortsvektoren der Punkte P, Q und R, dann kann man aus der oben eingeführten Punktdeterminate die folgenden geometrischen Eigenschaften ablesen:

- Die Punkte P, Q und R liegen genau dann auf einer Linie, wenn

$$\text{pdet}(\vec{p},\vec{q},\vec{r}) = 0$$

- Der Punkt R liegt genau dann links von der gerichteten Geraden $\overrightarrow{PQ}$, wenn

$$\mathrm{pdet}(\vec{p}, \vec{q}, \vec{r}) > 0$$

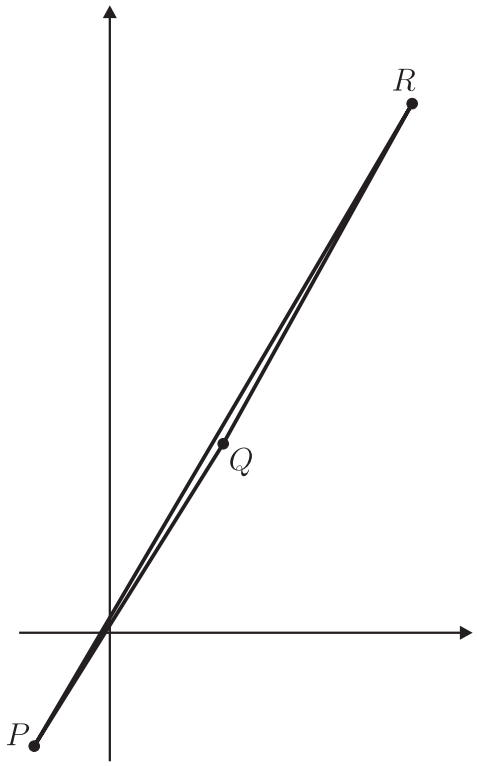
- Der Punkt R liegt genau dann rechts von der gerichteten Geraden $\overrightarrow{PQ}$, wenn

$$\mathrm{pdet}(\vec{p}, \vec{q}, \vec{r}) < 0$$

- Die Fläche des von P, Q und R aufgespannten Dreiecks beträgt:

$$\left| \frac{\mathrm{pdet}(\vec{p}, \vec{q}, \vec{r})}{2} \right|$$

Beispiel: $P = (-2, -3)$, $Q = (3, 5)$ und $R = (8, 14)$ in $\mathbb{R}^2$:



$$\mathrm{pdet}(\vec{p}, \vec{q}, \vec{r}) = \begin{vmatrix} -2 & 3 & 8 \\ -3 & 5 & 14 \\ 1 & 1 & 1 \end{vmatrix} = -10 + 42 - 24 - 40 + 9 + 28 = 5 > 0$$

Daraus folgt, dass $(8, 14)$ links von der gerichteten Geraden $\overrightarrow{(-2, -3)(3, 5)}$ liegt und dass die Fläche des von den drei Punkten aufgespannten Dreiecks $2, 5$ beträgt.

Diese Eigenschaften lassen sich auch auf höhere Dimensionen übertragen. Seien P, Q, R und S Punkte in $\mathbb{R}^3$, dann ist die Größe

$$\left| \frac{1}{6} \cdot \det \begin{pmatrix} p_x & q_x & r_x & s_x \\ p_y & q_y & r_y & s_y \\ p_z & q_z & r_z & s_z \\ 1 & 1 & 1 & 1 \end{pmatrix} \right|$$

das Volumen des von den vier Punkten aufgespannten Simplexes. Falls die vier Punkte auf einer Ebene liegen, ist der Wert 0.

Komplementärmatrix

Definition: Die zu einer Matrix $A \in M(n\times n, K)$ *komplementäre Matrix* $\tilde{A} = (\tilde{a}_{i\,j})_{(i,j)\in n\times n}$ ist wie folgt definiert:

$$\tilde{a}_{i\,j} = (-1)^{i+j} \cdot \det\, A_{j\,i}$$

Dabei ist $A_{j\,i}$ die Matrix, die man aus A durch Streichen der j-ten Zeile und der i-ten Spalte enthält.

Satz: $A \cdot \tilde{A} = (\det\, A) \cdot E_n$

Beweis: Auf der Diagonalen von $C := A \cdot \tilde{A}$ steht immer det A, denn der Koeffizient $c_{i\,i}$ kann leicht in die Entwicklung von det A nach der i–ten Zeile umgewandelt werden:

$$\begin{aligned} c_{i\,i} &= \sum_{k=1}^{n} a_{i\,k} \cdot \tilde{a}_{k\,i} \\ &= \sum_{k=1}^{n} a_{i\,k} \cdot (-1)^{k+i} \cdot \det\, A_{i\,k} \\ &= det(A) \end{aligned}$$

Bei der Berechnung eines Koeffizienten $c_{i\,j}$, der nicht auf der Diagonalen liegt, beginnt man mit den gleichen Ansatz, stellt dann aber fest, dass die Werte der Terme det $A_{j\,k}$ überhaupt nicht vom Inhalt der j-ten Zeile von A abhängen, d.h. wenn man aus A eine Matrix A' bildet, bei der die j–te Zeile durch die i–te Zeile ersetzt ist, erhält man die gleiche Formel. In A' gilt wegen der Identität der i–ten und j–ten Zeile $a'_{j\,k} = a'_{i\,k} = a_{i\,k}$. Auf diesem Umweg kommt man zur Entwicklung der Determinate von A' nach der j–ten Zeile. Da aber A' zwei identische Zeilen hat, ist det $A' = 0$.

$$\begin{aligned} c_{i\,j} &= \sum_{k=1}^{n} a_{i\,k} \cdot \tilde{a}_{k\,j} \\ &= \sum_{k=1}^{n} a_{i\,k} \cdot (-1)^{k+j} \cdot \det\, A_{j\,k} \\ &= \sum_{k=1}^{n} (-1)^{k+j} \cdot a'_{j\,k} \cdot \det\, A'_{j\,k} \\ &= det(A') = 0 \end{aligned}$$

Folgerung: Ist det $A \neq 0$, dann ist A invertierbar und $A^{-1} = \dfrac{\tilde{A}}{\det\, A}$.

Spezialfall: Für $A \in M(2 \times 2, K)$ gilt:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad \text{für} \quad ad - bc \neq 0$$

Die Determinate eines Endomorphismus'

Satz: Für alle $A, B \in M(n \times n, K)$ gilt:

$$\det(A \cdot B) = \det\, A \cdot \det\, B$$

Wir verzichten auf einen Beweis diese Satzes, wollen aber dafür eine wichtige Schlussfolgerung ziehen.

Satz: Sei $f : K^n \to K^n$ ein Endomorphismus, A die zu f gehörende Matrix bezüglich der Standardbasis $\mathcal{B}_1$ und B eine zu f gehörende Matrix bezüglich einer anderen Basis $\mathcal{B}_2$, dann ist $\det A = \det B$.

Beweis: Sie C die Matrix, deren Spalten die Vektoren aus $\mathcal{B}_2$ sind. Damit beschreibt C die Abbildung des Basiswechsel von $\mathcal{B}_1$ nach $\mathcal{B}_2$ bezüglich der Standardbasis. Die Matrix C ist invertierbar und die inverse Matrix C^{-1} beschreibt den umgekehrten Basiswechsel. Offensichtlich gilt nun:

$$C^{-1} \cdot A \cdot C = B.$$

Durch Anwendung des Satzes über die Determinante des Matrixprodukts erhalten wir:

$$\det B = \det\,(C^{-1}) \cdot \det A \cdot \det C = \det A \cdot \det\,(C^{-1} \cdot C) = \det A \cdot \det E_n = \det A.$$

Folgerung: Man kann jedem Endomorphismus $f : K^n \to K^n$ eindeutig seine Determinante $\det f = \det A$ zuordnen, wobei A die Abbildung f bezüglich einer beliebigen Basis repräsentiert.

Beispiel: Wir wollen die im Beweis beschriebene Methode zum Basiswechsel an einem einfachen Beispiel nachvollziehen. Dazu betrachten wir als Endomorphismus f die Spiegelung des Raums $\mathbb{R}^2$ an der x–Achse. Wir wählen

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \text{ die Standardbasis und } \mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \text{ eine andere Basis}$$

Das hat den Vorteil, dass wir nicht nur die Matrix A kennen, sondern auch schon B, denn durch die Spiegelung an der x–Achse wird der erste Vektor von $\mathcal{B}_2$ auf den zweiten und der zweite Vektor von $\mathcal{B}_2$ auf den ersten abgebildet, d.h.

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{und} \qquad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Damit kennen wir schon das Ergebnis, das eigentlich noch berechnet werden soll. Wir wissen, dass $C = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ ist und können C^{-1} durch die Komplementärmatrix bestimmen:

$$C^{-1} = \frac{1}{\det C} \cdot \tilde{C} = \frac{1}{-2} \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}$$

Durch Ausführung der Matrixmultiplikation $C^{-1} \cdot A \cdot C$ erhält man tatsächlich die oben dargestellte Matrix B.

1.11 Euklidische Vektorräume

Definition: Sei V ein reeller Vektorraum. Ein Skalarprodukt über V ist eine Abbildung $< , >: V \times V \to \mathbb{R}$ mit folgenden drei Eigenschaften:

1. Biliniarität, d.h. für jedes $\vec{v} \in V$ sind die Abbildungen
   - $< , \vec{v} >: V \to \mathbb{R}$ mit $\vec{w} \mapsto < \vec{w}, \vec{v} >$ oder
   - $< \vec{v}, >: V \to \mathbb{R}$ mit $\vec{w} \mapsto < \vec{v}, \vec{w} >$

   sind linear.
2. Symmetrie, d.h. $< \vec{v}, \vec{w} >=< \vec{w}, \vec{v} >$ für alle $\vec{v}, \vec{w} \in V$.
3. Positive Definitheit, d.h. $< \vec{v}, \vec{v} >> 0$ für alle $\vec{v} \neq \vec{0}$.

Ein ein reeller Vektorraum mit einem Skalarprodukt wird ein Euklidischer Vektorraum genannt.

Beispiele:

1. $V = \mathbb{R}^n$, Standardskalarprodukt:
$$\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$
$$\langle (1, 5, 0, 3), (3, 0, 7, -4) \rangle = 3 + 0 + 0 + (-12) = -9$$
2. $V = \{f : [-1, 1] \to \mathbb{R} \mid \text{stetige Funktion}\}$
$$\langle f, g \rangle := \int_{-1}^{1} f(x) g(x) dx$$

Definition: Die Norm eines Vektors $\vec{v}$ in einem Euklidischen Raum ist definiert durch
$$\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$$

Beispiel:
$$\|(1, 2, 0, 2, 4)\| = \sqrt{1 + 4 + 0 + 4 + 16} = \sqrt{25} = 5$$
Diese Norm beschreibt den Euklidischen Abstand zwischen $(0, 0, \ldots 0)$ und $(1, 2, 0, 2, 4)$ in $\mathbb{R}^5$.

Satz (Ungleichung von Cauchy-Schwarz): In jedem Euklidschen Vektorraum gilt für alle $\vec{u}, \vec{v} \in V$
$$|\langle \vec{u}, \vec{v} \rangle| \leq \|\vec{u}\| \cdot \|\vec{v}\|$$
Spezielle Form für das Standardskalarprodukt in $\mathbb{R}^n$ mit $\vec{u} = (a_1, a_2, \ldots, a_n)$ und $\vec{v} = (b_1, b_2, \ldots, b_n)$:
$$|a_1 b_1 + a_2 b_2 + \ldots + a_n b_b| \leq \sqrt{a_1^2, a_2^2, \ldots, a_n^2} \cdot \sqrt{b_1^2, b_2^2, \ldots, b_n^2}$$
Spezielle Form für $V = \{f : [-1, 1] \to \mathbb{R} \mid \text{stetige Funktion}\}$:
$$\left| \int_{-1}^{1} f(x) g(x) dx \right| \leq \sqrt{\int_{-1}^{1} (f(x))^2 dx} \cdot \sqrt{\int_{-1}^{1} (g(x))^2 dx}$$

Beweis:

1. Fall 1: Für $\vec{v} = \vec{0}$, dann ergibt die Cauchy-Schwarz-Ungleichung $0 = 0$, ist also korrekt.

2. Fall 2: Für $\vec{v} \neq \vec{0}$ setzt man
$$\lambda := \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} = \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{v}\|^2}$$

Daraus ergibt sich:
$$\begin{aligned} 0 &\leq \langle \vec{u} - \lambda\vec{v}, \vec{u} - \lambda\vec{v} \rangle \\ &= \langle \vec{u}, \vec{u} - \lambda\vec{v} \rangle - \lambda\langle \vec{v}, \vec{u} - \lambda\vec{v} \rangle \\ &= \langle \vec{u}, \vec{u} \rangle - \lambda\langle \vec{u}, \vec{v} \rangle - \lambda\langle \vec{v}, \vec{u} \rangle + \lambda^2\langle \vec{v}, \vec{v} \rangle \\ &= \|\vec{u}\|^2 - 2 \cdot \frac{\langle \vec{u}, \vec{v} \rangle^2}{\|\vec{v}\|} + \frac{\langle \vec{u}, \vec{v} \rangle^2 \cdot \|\vec{v}\|^2}{\|\vec{v}\|^4} \\ &= \|\vec{u}\|^2 - \frac{\langle \vec{u}, \vec{v} \rangle^2}{\|\vec{v}\|^2} \end{aligned}$$

Also:
$$\langle \vec{u}, \vec{v} \rangle^2 \leq \|\vec{u}\|^2 \cdot \|\vec{v}\|^2 \quad \Rightarrow \quad |\langle \vec{u}, \vec{v} \rangle| \leq \|\vec{u}\| \cdot \|\vec{v}\|$$

Satz: Die Norm in einem Euklidischen Vektorraum hat die folgenden Eigenschaften:

1. $\|\vec{v}\| \geq 0$ für alle $\vec{v} \in V$
2. $\|\vec{v}\| = 0 \quad \Leftrightarrow \quad \vec{v} = \vec{0}$
3. $\|\lambda\vec{v}\| = |\lambda|\|\vec{v}\|$
4. $\|\vec{u} + \vec{v}\| \leq \|\vec{v}\| + \|\vec{u}\|$ (Dreiecksungleichung)

Beweis: Die Eigenschaften 1 und 2 folgen unmittelbar aus der positiven Definitheit. Es genügt deshalb, die Eigenschaften 3 und 4 nachzuweisen.

$$\begin{aligned} \|\lambda\vec{v}\| &= \sqrt{\langle \lambda\vec{v}, \lambda\vec{v} \rangle} \\ &= \sqrt{\lambda\langle \vec{v}, \lambda\vec{v} \rangle} \\ &= \sqrt{\lambda^2\langle \vec{v}, \vec{v} \rangle} \\ &= \sqrt{\lambda^2} \cdot \sqrt{\langle \vec{v}, \vec{v} \rangle} \\ &= |\lambda|\|\vec{v}\| \end{aligned}$$

Die Dreiecksungleichung beweist man in der quadrierten Form und beginnt dabei von der rechten Seite:

$$\begin{aligned} (\|\vec{u}\| + \|\vec{v}\|)^2 &= \|\vec{u}\|^2 + 2\|\vec{u}\|\|\vec{v}\| + \|\vec{v}\|^2 \\ &\geq \|\vec{u}\|^2 + 2\langle \vec{u}, \vec{v} \rangle + \|\vec{v}\|^2 \qquad \text{(Cauchy-Schwarz)} \\ &= \langle \vec{u}, \vec{u} \rangle + 2\langle \vec{u}, \vec{v} \rangle + \langle \vec{v}, \vec{v} \rangle \\ &= \langle \vec{u} + \vec{v}, \vec{u} + \vec{v} \rangle \\ &= \|\vec{u} + \vec{v}\|^2 \end{aligned}$$

Ein Spezialfall dieser allgemeinen Dreiecksungleichung ist die herkömmliche Dreiecksungleichung aus der Geometrie, die aussagt, dass für jedes Dreieck mit den Seitenlängen a, b, c die Ungleichung $c \leq a + b$ gilt. Die Beziehungen zur Geometrie gehen aber noch wesentlich weiter. Das Skalarprodukt und die Norm können auch verwendet werden, um den Öffnungswinkel zwischen zwei Vektoren zu beschreiben.

Definition: Der Öffnungswinkel zwischen zwei Vektoren $\vec{v}, \vec{w} \in V$ in einem Euklidischen Vektorraum ist wie folgt definiert:

$$\sphericalangle(\vec{u}, \vec{v}) = \arccos \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \|\vec{v}\|}$$

Man kann wieder beobachten, dass diese Definition für das Standardskalarprodukt mit dem herkömmlichen Winkel übereinstimmt. Wir betrachten dazu das Dreieck, dass durch zwei Ortsvektoren $\vec{u}$ und $\vec{v}$ aufgespannt wird und bezeichnen die Seitenlängen mit $a = \|\vec{u}\|$, $b = \|\vec{v}\|$ und $c = \|\vec{u} - \vec{v}\|$. Der von $\vec{u}$ und $\vec{v}$ eingeschlossene Winkel wird in der Dreiecksgeometrie mit γ bezeichnet und nach Cosinussatz gilt:

$$c^2 = a^2 + b^2 - 2ab\cos\gamma = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\,\|\vec{v}\|\cos\gamma.$$

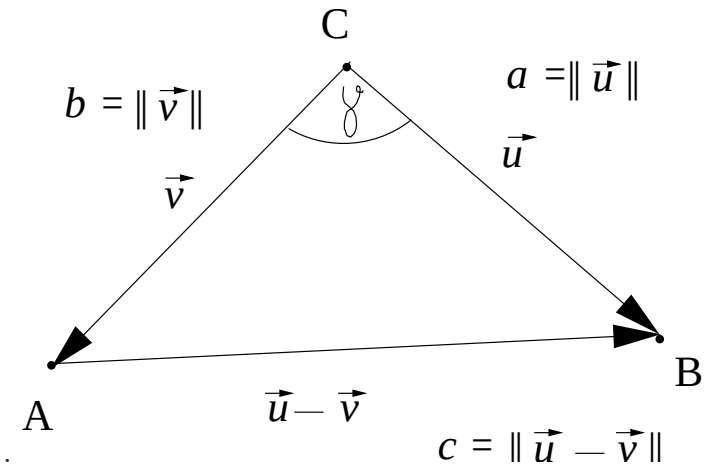
Wir stellen c^2 nun noch einmal durch die Norm dar:

$$\begin{aligned} c^2 = \|\vec{u} - \vec{v}\|^2 &= \langle \vec{u} - \vec{v}\,,\, \vec{u} - \vec{v} \rangle \\ &= \langle \vec{u}\,,\, \vec{u} \rangle - 2\langle \vec{u}\,,\, \vec{v} \rangle + \langle \vec{v}\,,\, \vec{v} \rangle \\ &= \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\langle \vec{u}\,,\, \vec{v} \rangle. \end{aligned}$$

Aus der Kombination beider Gleichungen ergibt sich die Identität

$$\|\vec{u}\|\,\|\vec{v}\|\cos\gamma = \langle \vec{u}\,,\, \vec{v} \rangle\,,$$

welche man leicht in die oben gegebene Winkeldefinition umformen kann.



Beispiel: Der von den Vektoren $\vec{u} = (-4, 3, 0)$ und $\vec{v} = (2, -4, \sqrt{44})$ aufgespannte Winkel wird wie folgt berechnet:

$$\begin{aligned}
\cos(\sphericalangle(\vec{u}, \vec{v})) &= \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \cdot \|\vec{v}\|} = \frac{\langle (-4, 3, 0), (2, -4, \sqrt{44}) \rangle}{\|(-4, 3, 0)\| \cdot \|(2, -4, \sqrt{44})\|} \\
&= \frac{-8 - 12 + 0}{\sqrt{16 + 9} \cdot \sqrt{4 + 16 + 44}} = \frac{-20}{5 \cdot 8} = -\frac{1}{2} \\
\sphericalangle(\vec{u}, \vec{v}) &= arcos\left(-\frac{1}{2}\right) = \frac{2\pi}{3}
\end{aligned}$$

Definition: Zwei Vektoren $\vec{u}$ und $\vec{v}$ in einem Euklidischen Vektorraum $(V, \langle\ ,\ \rangle)$ sind orthogonal (senkrecht) zueinander, wenn das Skalarprodukt $\langle \vec{u}, \vec{v} \rangle$ gleich Null ist. Eine Teilmenge $M \subseteq V$ ist orthogonal zu $\vec{u}$, wenn das Skalarprodukt $\langle \vec{v}, \vec{u} \rangle$ für alle $\vec{v} \in M$ gleich Null ist.
Für die Orthogonalität verwenden wir die Notationen $\vec{u} \perp \vec{v}$ und $\vec{u} \in M$.

Man beachte, dass der hier eingeführte Begriff mit der weiter obenen gegebenen Winkeldefinition im Einklag steht:

$$\sphericalangle(\vec{u}, \vec{v}) = \frac{\pi}{2} \quad \Leftrightarrow \quad \langle \vec{u}, \vec{v} \rangle = 0 \quad \Leftrightarrow \quad \vec{u} \text{ steht senkrecht auf } \vec{v}$$

Definition: Das orthogonale Komplement $M^\perp$ einer Menge $M \subseteq V$ ist die Menge aller Vektoren, die orthogonal zu M sind, d.h.

$$M^\perp = \{\vec{u} \in V \mid M \perp \vec{u}\} = \{\vec{u} \in V \mid \forall\, \vec{v} \in M \qquad \langle \vec{u}, \vec{v} \rangle = 0\}$$

Satz: Die Menge $M^\perp$ ist ein Untervektorraum.
Beweis:

- Der Nullvektoren gehört zu $M^\perp$, d.h. $M^\perp$ ist nicht leer.
- Addition: Sind $\vec{u}, \vec{u'} \in M^\perp$ dann ist $\langle \vec{u}, \vec{v} \rangle = \langle \vec{u'}, \vec{v} \rangle = 0$ für alle $\vec{v} \in M$.
  Aus der Bilinearität des Skalarprodukts folgt $\langle \vec{u} + \vec{u'}, \vec{v} \rangle = \langle \vec{u}, \vec{v} \rangle + \langle \vec{u'}, \vec{v} \rangle = 0$ und damit ist $\vec{u} + \vec{u'} \in M^\perp$.
- Multiplikation: Ist $\vec{u} \in M^\perp$ und $\lambda \in \mathbb{R}$, dann ist $\langle \vec{u}, \vec{v} \rangle = 0$ für alle $\vec{v} \in M$.
  Aus der Bilinearität des Skalarprodukts folgt $\langle \lambda \vec{u}, \vec{v} \rangle = \lambda \langle \vec{u}, \vec{v} \rangle = 0$ und damit ist $\lambda \vec{u} \in M^\perp$.

Definition: Eine Menge von Vektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r$ wird Orthonormalsystem genannt, falls $\|\vec{v}_i\| = 1$ für alle $i = 1, 2, \ldots r$ und $\langle \vec{v}_i, \vec{v}_j \rangle = 0$ für alle $i \neq j$, in verkürzter Schreibweise

$$\langle \vec{v}_i, \vec{v}_j \rangle = \delta_{i\,j} = \begin{cases} 1 & \text{falls} \quad i = j \\ 0 & \text{sonst} \end{cases}$$

Beispiel: Die Standardbasen der Vektorräume $\mathbb{R}^n$ sind Orthonormalsysteme.

Die folgenden drei Lemmata bilden die Grundlage eines Orthogonalisierungsverfahrens, mit dem man aus einer beliebigen Basis eine Orthonormalbasis erzeugen kann.

Lemma 1: Die Vektoren eines Orthogonalsystem sind linear unabhängig.

Beweis: Wir betrachten eine Linearkombination $\lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_r\vec{v}_r = \vec{0}$ und müssen $\lambda_1 = \lambda_2 = \ldots = \lambda_r = 0$ zeigen.
Um $\lambda_i = 0$ nachzuweisen, bildet man das Skalarprodukt aus der Linearkombination und dem vektor $\vec{v}_i$:

$$\begin{aligned} 0 &= \langle \vec{0}, \vec{v}_i \rangle \\ &= \langle \lambda_1\vec{v}_1 + \lambda_2\vec{v}_2 + \ldots + \lambda_r\vec{v}_r, \vec{v}_i \rangle \\ &= \lambda_1 \underbrace{\langle \vec{v}_1, \vec{v}_i \rangle}_{=0} + \ldots + \lambda_i \underbrace{\langle \vec{v}_i, \vec{v}_i \rangle}_{=1} + \ldots + \lambda_r \underbrace{\langle \vec{v}_r, \vec{v}_i \rangle}_{=0} \\ &= \lambda_i \end{aligned}$$

Lemma 2: Ist $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ eine orthonormale Basis von V, so gilt für jedes $\vec{v} \in V$ die folgende *Entwicklungsformel*:

$$\vec{v} = \sum_{i=1}^{n} \langle \vec{v}, \vec{v}_i \rangle \cdot \vec{v}_i$$

Beweis: Ähnlich wie im Beweis von Lemma 1 reicht einfaches Nachrechnen aus. Geht man von der eindeutigen Darstellung $\vec{v} = \sum_{i=1}^{n} \lambda_i \vec{v}_i$ in der gegebenen Basis aus, muss nur noch $\lambda_j = \langle \vec{v}, \vec{v}_j \rangle$ für alle $1 \leq j \leq n$ nachgewiesen werden:

$$\langle \vec{v}, \vec{v}_j \rangle = \left\langle \sum_{i=1}^{n} \lambda_i \vec{v}_i \, , \, \vec{v}_j \right\rangle = \sum_{i=1}^{n} \lambda_i \, \langle \vec{v}_i, \vec{v}_j \rangle = \lambda_j$$

Beispiel: Für die Standardbasis im $\mathbb{R}^3$, ist die Gültigkeit der Entwicklungsformel offensichtlich.

Sei $\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\vec{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ und $\vec{v} = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$, dann ergibt sich

$$\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = (2+0+0) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (0+3+0) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + (0+0+0) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 2 \cdot \vec{v}_1 + 3 \cdot \vec{v}_2 + 0 \cdot \vec{v}_3$$

Lemma 3: Ist $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r$ ein Orthonormalsystem in V und $U = \mathrm{Lin}(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\})$, so hat jedes $\vec{v} \in V$ eine eindeutige Darstellung

$$\vec{v} = \vec{u} + \vec{w} \text{ mit } \vec{u} \in U \text{ und } \vec{w} \in U^{\perp}$$

Beweis: Durch Anwendung der Entwicklungsformel kann man die Zerlegung konstruktiv angeben:

$$\vec{u} = \sum_{i=0}^{r} \langle \vec{v}, \vec{v}_i \rangle \cdot \vec{v}_i \qquad \text{und} \qquad \vec{w} = \vec{v} - \vec{u}$$

Es ist klar, dass der so konstruierte Vektor $\vec{u}$ in U liegt. Zum Nachweis, dass der Vektor $\vec{w}$ in $U^\perp$ liegt, genügt es zu zeigen, dass das Skalarprodukt aus $\vec{w}$ und einem beliebigen Basisvektor $\vec{v}_j$ von U gleich Null ist:

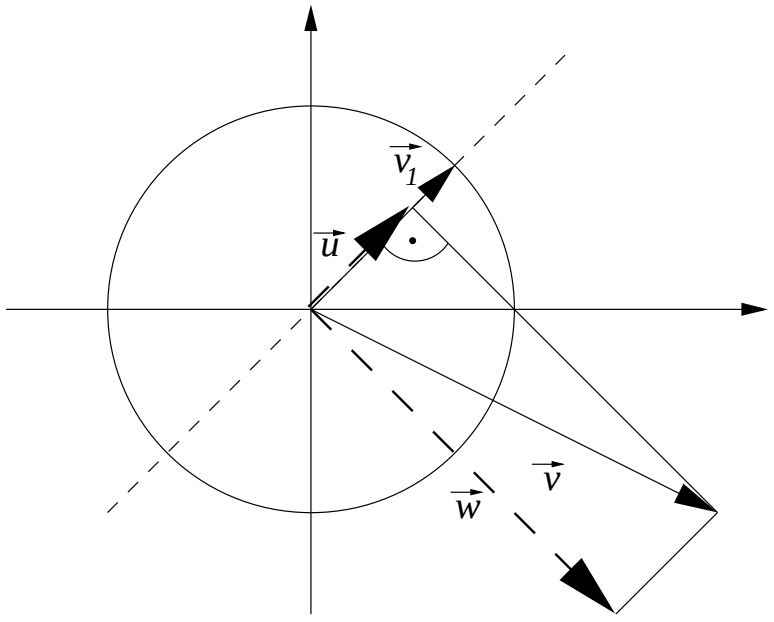
$$\begin{aligned}\langle \vec{w}\,,\,\vec{v}_j\rangle = \langle \vec{v}-\vec{u}\,,\,\vec{v}_j\rangle &= \langle \vec{v} - \sum_{i=0}^{r}\langle \vec{v},\vec{v}_i\rangle\cdot\vec{v}_i\,,\,\vec{v}_j\rangle \\ &= \langle \vec{v},\vec{v}_j\rangle - (0+\ldots+0+\langle \vec{v},\vec{v}_j\rangle\cdot\langle \vec{v}_j,\vec{v}_j\rangle+0\ldots+0) \\ &= \langle \vec{v},\vec{v}_j\rangle - \langle \vec{v},\vec{v}_j\rangle\cdot 1 = 0\end{aligned}$$

Zum Beweis der Eindeutigkeit betrachten wir zwei Darstellungen

$$\vec{v} = \vec{u}+\vec{w} = \vec{u'}+\vec{w'} \quad \text{mit} \quad \vec{u},\vec{u'}\in U \quad \text{und} \quad \vec{w},\vec{w'}\in U^\perp$$

Durch Umstellung der Gleichung entsteht ein Vektor $\vec{x} = \vec{u}-\vec{u'} = \vec{w'}-\vec{w}$, der gleichzeitig in U und in $U^\perp$ liegt. Aus der Definition des orthogonalen Komplements folgt $\langle \vec{x},\vec{x}\rangle = 0$ und da das Skalarprodukt positiv definit ist, ergibt sich $\vec{x} = \vec{0}$ und letztlich $\vec{u} = \vec{u'}$ sowie $\vec{w} = \vec{w'}$.

Beispiel: Wir betrachten im Raum $V = \mathbb{R}^2$ ein einelementiges Orthonormalsystem ($r = 1$) mit dem Vektor $\vec{v}_1 = \begin{pmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{pmatrix}$. Es sei $\vec{v} = \begin{pmatrix}2\\-1\end{pmatrix}$ der zu zerlegende Vektor. Wie man an der Abbildung nachvollziehen kann, führt der Konstruktion aus dem Beweis von Lemma 3 dazu, dass der Vektor $\vec{u}$ die orthogonale Projektion (d.h. Fällen des Lots) von $\vec{v}$ auf den Unterraum U ist und folglich der Differenzvektor $\vec{w} = \vec{v}-\vec{u}$ senkrecht auf U steht.



Die detaillierte Berechnung:

$$\begin{aligned}\vec{u} &= \langle \vec{v},\vec{v}_1\rangle\vec{v}_1 = \left\langle \begin{pmatrix}2\\-1\end{pmatrix}\,,\,\begin{pmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{pmatrix}\right\rangle \begin{pmatrix}\sqrt{2}/2\\ \sqrt{2}/2\end{pmatrix} = \frac{\sqrt{2}}{2}\begin{pmatrix}\sqrt{2}2\\ \sqrt{2}/2\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\ \frac{1}{2}\end{pmatrix} \\ \vec{w} &= \begin{pmatrix}2\\-1\end{pmatrix} - \begin{pmatrix}1/2\\1/2\end{pmatrix} = \begin{pmatrix}3/2\\-3/2\end{pmatrix}\end{aligned}$$

Satz (Erhard Schmidt'sches Orthonormalisierungsverfahren): Seien die Vektoren $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r$ linear unabhängig, dann bilden die wie folgt bestimmten Vektoren $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r$ ein Orthonormalsystem

$$\begin{array}{lcllcll} \tilde{u}_1 &=& \vec{v}_1 & \text{und} & \vec{u}_1 &=& \dfrac{\tilde{u}_1}{\|\tilde{u}_1\|} \\ \tilde{u}_2 &=& \vec{v}_2 - \langle \vec{v}_2, \vec{u}_1 \rangle \cdot \vec{u}_1 & \text{und} & \vec{u_2} &=& \dfrac{\tilde{u}_2}{\|\tilde{u}_2\|} \quad \text{und allgemein} \\ \tilde{u}_k &=& \vec{v}_k - \sum_{i=1}^{k-1} \langle \vec{v}_k, \vec{u}_i \rangle \cdot \vec{u}_i & \text{und} & \vec{u}_k &=& \dfrac{\tilde{u}_k}{\|\tilde{u}_k\|} \quad \text{für } k = 2, 3, \ldots r. \end{array}$$

Darüber hinaus hat das neue Orthonormalsystem die Eigenschaft, dass

$$\mathrm{Lin}(\{\vec{v}_1, \vec{v}_2, \ldots \vec{v}_i\}) = \mathrm{Lin}(\{\tilde{v}_1, \tilde{v}_2, \ldots \tilde{v}_i\}) \text{ für alle } i = 1, 2, \ldots r.$$

Alle Aussagen dieses Satzes ergeben sich durch direkte Anwendung der vorangestellten Lemmata.

Beispiel: Gegeben ist eine Basis $\vec{v}_1 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}$, $\vec{v}_2 = \begin{pmatrix} -3 \\ -1 \\ 0 \end{pmatrix}$, $\vec{v}_3 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

$$\tilde{u}_1 = \vec{v}_1 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} \qquad \vec{u}_1 = \tfrac{1}{\sqrt{8}} \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}$$

$$\tilde{u}_2 = \vec{v}_2 - \langle \vec{v}_2, \vec{u}_1 \rangle \vec{u}_1 = \begin{pmatrix} -3 \\ -1 \\ 0 \end{pmatrix} - (-2\sqrt{2}) \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} -3 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \qquad \vec{u}_2 = \tfrac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{-\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}$$

$$\tilde{u}_3 = \vec{v}_3 - \langle \vec{v}_3, \vec{u}_1 \rangle \vec{u}_1 - \langle \vec{v}_3, \vec{u}_2 \rangle \vec{u}_2$$

$$= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \tfrac{3\sqrt{2}}{2} \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix} - \tfrac{\sqrt{2}}{2} \begin{pmatrix} \frac{-\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} \frac{3}{2} \\ \frac{3}{2} \\ 0 \end{pmatrix} - \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} \qquad \vec{u}_3 = \tfrac{1}{3} \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Wie wir bereits im Beispiel zu Lemma 3 gesehen haben, entsteht die Zerlegung eines Vektors $\vec{v}$ in einen Vektor $\vec{u} \in U$ und einen Vektor $\vec{w} \in U^\perp$ anschaulich durch Projektion von $\vec{v}$ auf den Unterraum U. Im Folgenden werden wir auf den Begriff der Orthogonalprojektion noch einmal genauer eingehen.

Definition: Ist U ein Unterraum eines Euklidischen Vektorraums V, dann nennt die Abbildung $P_U : V \to U$ deren Einschränkung auf U die identische Abbildung ist und deren Kern $U^\perp$ ist, die Orthogonalprojektion von V auf U.

Zur konkreten Beschreibung der Orthogonalprojektion P_U konstruiert man eine Orthonormalbasis $\vec{u}_1, \ldots, \vec{u}_k$ von U und definiert

$$P_U(\vec{v}) := \langle \vec{v}, \vec{u}_1 \rangle \cdot \vec{u}_1 + \ldots + \langle \vec{v}, \vec{u}_k \rangle \cdot \vec{u}_k \in U.$$

Durch die Definition der Orthogonalprojektion P_U als Abbildung von V nach U ist bereits gesichert, dass P_U surjektiv ist. Das hat aber den Nachteil, dass man bei der Matrixdarstellung von P_U im Allgemeinen nur auf der linken Seite die Standardbasis verwenden kann. Deshalb wird für die Matrixdarstellung in der Regel P_U als ein Endomorphismus auf V (also als Abbildung von V nach V) betrachtet.

Beispiel: Sei $U \subseteq \mathbb{R}^2$ die lineare Hülle des Vektors $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Zu bestimmen ist die Matrix der Orthogonalprojektion $P_U : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$.

1. Orthonormalbasis für U

$$\vec{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \mapsto \quad \vec{u}_1 = \frac{1}{\left\| \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\|} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$$

2. Projektionsformel: $P_U(\vec{v}) = \langle \vec{v}, \tilde{v}_1 \rangle \cdot \tilde{v}_1$

   Bestimmung der Bilder der Basisvektoren $\vec{e}_1$ und $\vec{e}_2$:

$$\begin{aligned} P_U\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) &= \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \right\rangle \cdot \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{4}{5} \\ \frac{2}{5} \end{pmatrix} \\ P_U\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) &= \left\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \right\rangle \cdot \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{2}{5} \\ \frac{1}{5} \end{pmatrix} \end{aligned}$$

3. Matrixdarstellung:

$$A = \begin{pmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix} = \begin{pmatrix} 0,8 & 0,4 \\ 0,4 & 0,2 \end{pmatrix}$$

Isometrische Abbildungen

Definition: Ein Homomorphismus $f \in \mathrm{Hom}(V, W)$ von einem Euklidischen Vektorraum V in einen Euklidischen Vektorraum W wird *isometrische Abbildung* (auch *Isometrie* oder *orthogonale Abbildung*) genannt, wenn für alle Vektoren $\vec{v}, \vec{w}$ die folgende Identität erfüllt ist: $\langle \vec{v}, \vec{w} \rangle = \langle f(\vec{v}), f(\vec{w}) \rangle$

Bemerkung 1: Diese Identität impliziert, dass eine Isometrie die Norm von Vektoren nicht verändert, geometrisch interpretiert, dass die Abbildung abstandserhaltend ist. Damit ist jede Isometrie ein injektive Abbildung.

Bemerkung 2: Orthogonalprojektionen P_U haben (mit Ausnahme der identischen Abbildung, die eine Projektion auf ganz V ist) den nichttrivialen Kern $U^\perp$. Sie sind also nicht

injektiv und damit im Sinne der Definition keine orthogonalen Abbildungen. Um diesen Begriffswirrwarr zu vermeiden, bevorzugen wir den Begriff der isometrischen Abbildung.

Lemma: Ist $f \in \mathrm{Hom}(V,W)$ und $\vec{v}_1, \ldots, \vec{v}_n$ eine Orthonormalbasis von V, dann ist f genau dann eine Isometrie, wenn die Vektoren $f(\vec{v}_1), \ldots, f(\vec{v}_n)$ ein Orthonormalsystem in W bilden.

Beweis: Die Richtung ($\Longrightarrow$) folgt aus der Definition von isometrischen Abbildungen. Für die Rückrichtung setzen wir voraus, dass $f(\vec{v}_1), \ldots, f(\vec{v}_n)$ ein Orthonormalsystem ist und betrachten zwei Vektoren $\vec{u}, \vec{v}$ aus V in ihrer Basisdarstellung:

$$\vec{u} = \sum_{i=1}^{n} \lambda_i \vec{v}_i \qquad\qquad \vec{v} = \sum_{j=1}^{n} \mu_j \vec{v}_j$$

$$\begin{aligned} \langle \vec{u}, \vec{v} \rangle &= \left\langle \sum_{i=1}^{n} \lambda_i \vec{v}_i \,,\, \sum_{j=1}^{n} \mu_j \vec{v}_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mu_j \cdot \langle \vec{v}_i, \vec{v}_j \rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mu_j \cdot \delta_{ij} = \sum_{i=1}^{n} \lambda_i \mu_i \end{aligned}$$

Da aber auch $\langle f(\vec{v}_i), f(\vec{v}_j) \rangle = \delta_{ij}$ gilt und f ein lineare Abbildung ist, kann man auf die gleiche Weise auch

$$\langle f(\vec{u}), f(\vec{v}) \rangle = \sum_{i=1}^{n} \lambda_i \mu_i$$

ableiten und hat damit $\langle \vec{u}, \vec{v} \rangle = \langle f(\vec{u}), f(\vec{v}) \rangle$ gezeigt.

Ist $f \in \mathrm{Hom}(V,V)$ eine Isometrie und ein Endomorphismus, dann nennt man die Matrix von f bezüglich der Standarbasis eine *orthogonale Matrix*. Nutzt man das Wissen über transponierte Matrizen, gelangt man leicht zu der folgenden Charakterisierung von orthogonalen Matrizen.

Satz: Für einen Matrix $A \in M(n \times n, \mathbb{R})$ sind die folgenden Bedingungen äquivalent:

1. A ist eine orthogonale Matrix
2. Die Spalten von A sind ein Orthonormalsystem
3. $A^t \cdot A = E_n$
4. A ist invertierbar und $A^{-1} = A^t$
5. $A \cdot A^t = E_n$
6. Die Zeilen von A sind ein Orthonormalsystem

Die einzige Lücke, die zum Beweis noch geschlossen werden muss, besteht zwischen der zweiten und der dritten Aussage. Man kann sie aber leicht durch die folgende Beobachtung schließen: Der Koeffizient in der i-ten Zeile und j-ten Spalte von $A^t \cdot A$ ist das Skalarprodukt der i-ten und j-ten Spalte von A und damit ist $A^t \cdot A = E_n$ nur eine Umschreibung der Tatsache, dass die Spalten von A ein Orthonormalsystem sind.

1.12 Eigenwerte und Eigenvektoren

Diagonalisierbarkeit

Wir haben bereits im Zusammenhang mit inversen Matrizen die Auswirkungen eines Basiswechsel für die Matrixdarstellung eines Endomorphismus kennengelernt, sind dabei aber noch nicht auf die Frage eingegangen, ob und wann ein solcher Basiswechsel eigentlich sinnvoll ist. Der Nutzen eines Basiswechsels liegt sicherlich immer dann klar auf der Hand, wenn durch diesen Wechsel eine sehr einfache Matrix entsteht. Das trifft insbesondere auf den folgenden Fall zu.

Definition: Ein Endomorphismus $f \in \text{Hom}\,(V,V)$ wird *diagonalisierbar* genannt, wenn eine Basis von V existiert, für die die Matrixdarstellung A von f Diagonalgestalt hat, d.h. wenn A außerhalb der Diagonalen nur Nullen hat. Eine solche Basis nennt man *Diagonalbasis* von f.

Zu den wesentlichen Vorteilen einer Diagaonalmatrix $A \in M(n\times n, K)$ gehören die folgenden Aspekte:

- Das Bild eines beliebigen Vektors $\vec{v} \in V$ kann mit n Multiplikationen und $n-1$ Additionen bestimmt werden.
- Man kann die Determinante von A mit $n-1$ Multiplikationen bestimmen und damit auch entscheiden, ob A invertierbar ist.
- Falls A invertierbar ist, kann man A^{-1} mit n Divisionen berechnen.

Der Fakt, dass jeder Vektor $\vec{v}$ einer Diagonalbasis die Eigenschaft $f(\vec{v}) = \lambda\vec{v}$ hat, führt zur folgenden Begriffsbildung.

Definition: Ist $f \in \text{Hom}\,(V,V)$ ein Endomorphismus, $\vec{v} \neq \vec{0}$ ein Vektor aus V und $\lambda \in K$, so dass

$$f(\vec{v}) = \lambda\vec{v}$$

gilt, so wird $\vec{v}$ *Eigenvektor* von f zum *Eigenwert* λ genannt.

Lemma: Ist $M = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}$ eine Menge von Eigenvektoren eines Endomorphismus f zu paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_r$, so ist M linear unabhängig.

Beweis durch Induktion nach r:
Der Induktionsanfang für $r = 1$ ist trivial, denn ein Eigenvektor ist nach Definition nicht der Nullvektor und damit ist $\{\vec{v}_1\}$ linear unabhängig.
Sei die Aussage wahr für alle Mengen mit r Vektoren und sei $M' = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{r+1}\}$ eine Menge von Eigenvektoren eines Endomorphismus f zu paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_{r+1}$. Zu zeigen ist, dass jede Linearkombination

$$\vec{0} = \mu_1\vec{v}_1 + \ldots + \mu_r\vec{v}_r + \mu_{r+1}\vec{v}_{r+1}$$

trivial ist, d.h. dass $\mu_1 = \ldots = \mu_r = \mu_{r+1} = 0$ ist. Dazu betrachtet man die Abbildung dieser Linearkombination unter f:

$$\begin{aligned}
\vec{0} = f(\vec{0}) &= f(\mu_1\vec{v}_1 + \ldots + \mu_r\vec{v}_r + \mu_{r+1}\vec{v}_{r+1}) \\
&= \mu_1 f(\vec{v}_1) + \ldots + \mu_r f(\vec{v}_r) + \mu_{r+1} f(\vec{v}_{r+1}) \\
&= \mu_1\lambda_1\vec{v}_1 + \ldots + \mu_r\lambda_r\vec{v}_r + \mu_{r+1}\lambda_{r+1}\vec{v}_{r+1}
\end{aligned}$$

Durch subtraktive Verknüpfung dieser Gleichung mit der Gleichung

$$\begin{aligned}\vec{0} = \lambda_{r+1}\vec{0} &= \lambda_{r+1}\left(\mu_1\vec{v}_1 + \ldots + \mu_r\vec{v}_r + \mu_{r+1}\vec{v}_{r+1}\right) \\ &= \mu_1\lambda_{r+1}\vec{v}_1 + \ldots + \mu_r\lambda_{r+1}\vec{v}_r + \mu_{r+1}\lambda_{r+1}\vec{v}_{r+1}\end{aligned}$$

erhält man

$$\vec{0} = \mu_1(\lambda_1 - \lambda_{r+1})\vec{v}_1 + \ldots + \mu_r(\lambda_r - \lambda_{r+1})\vec{v}_r + \mu_{r+1}()\underbrace{(\lambda_{r+1} - \lambda_{r+1})}_{=0}\vec{v}_{r+1}.$$

Da $\vec{v}_{r+1}$ in dieser Linearkombination verschwindet und nach Induktionsvoraussetzung die Menge $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}$ linear unabhängig ist, muss die Linearkombination trivial sein. Unter Berücksichtigung der Verschiedenheit der Eigenwerte ergibt sich

$$\mu_1\underbrace{(\lambda_1 - \lambda_{r+1})}_{\neq 0} = \ldots = \mu_r\underbrace{(\lambda_r - \lambda_{r+1})}_{\neq 0} = 0 \quad \Longrightarrow \quad \mu_1 = \ldots = \mu_r = 0$$

und daraus folgt auch $\mu_{r+1} = 0$, womit die Induktionsbehauptung bewiesen ist. □

Folgerung: Ist $\dim V = n$ und der Endomorphismus $f \in \mathrm{Hom}\,(V, V)$ hat n Eigenwerte, dann ist f diagonalsierbar.

Auf Grund des Lemmas bildet jede Menge von Eigenvektoren zu den n Eigenwerten eine Basis von V. Offensichtlich ist das eine Diagonalbasis, denn die zugehörige Matrixdarstellung von f hat auf der Diagonalen die Eigenwerte und sonst nur Nullen. □

Wie wir sehen werden, ist die Existenz von n Eigenwerten aber keine notwendige Voraussetzung für die Diagonaliesierbarkeit.

Eigenräume

Sind $\vec{v}$ und $\vec{w}$ zwei Eigenvektoren von $f \in \mathrm{Hom}\,(V, V)$ zum selben Eigenwert λ, dann ist auch jede (von $\vec{0}$ verschiedene) Linearkombination aus $\vec{v}$ und $\vec{w}$ ein Eigenvektor zum Eigenwert λ, mit anderen Worten bilden die Eigenvektoren zu λ zusammen mit dem Nullvektor einen Unterraum von V. Da $f(\vec{v}) = \lambda\vec{v}$ äquivalent ist zu $(f - \lambda\,\mathrm{Id}_V)(\vec{v}) = \vec{0}$, kann man diesen Unterraum auch als Kern des Endomorphismus $f - \lambda\,\mathrm{Id}_V$ charkterisieren.
Definition: Ist λ ein Eigenwert von $f \in \mathrm{Hom}\,(V, V)$, dann nennt man den Unterraum $E_\lambda := \mathrm{Ker}\,(f - \lambda\,\mathrm{Id}_V)$ den *Eigenraum* von λ. Die Dimension von E_λ wird *geometrische Vielfachheit* von λ genannt.

Satz: Ein Endomorphismus $f \in \mathrm{Hom}\,(V, V)$ mit den Eigenwerten $\lambda_1, \ldots, \lambda_r$ ist genau dann diagonalisierbar, wenn

$$\sum_{k=1}^{r} \dim E_{\lambda_k} = \dim V.$$

Beweis: Man verwendet das Lemma über die lineare Unabhängigkeit von Eigenvektoren zu verschiedenen Eigenwerten um zu zeigen, dass die Vereinigung der Basen der Eigenräume $E_{\lambda_1}, \ldots, E_{\lambda_r}$ auch eine linear unabhängige Menge ist. Damit ist diese Vereinigung genau dann eine Basis von V, wenn sie $n = \dim V$ Elemente hat.

Wenn man alle Eigenwerte von f kennt, ist es nicht schwer, die Diagonalisierbarkeit von f überprüfen, denn die Eigenräume zu den Eigenwerten λ_i sind die Kerne der Endomorphismen $f - \lambda_i \operatorname{Id}_V$. Es bleibt also zu klären, wie man die Menge der Eigenwerte von f bestimmen kann.

Charakteristisches Polynom

Definition: Sei V ein n–dimensionaler Vektorraum und $f \in \operatorname{Hom}(V, V)$, dann nennt man die Determinante $\det(f - \lambda \cdot \operatorname{Id}_V)$ das *charakteristisches Polynom* von f. Da wir wissen, dass die Determinante eines Epimorhismus auf V nicht von der Wahl der Basis von V abhängig ist, kann man das charakteristische Polynom von f auch als $\det(A - \lambda \cdot E_n)$ definieren, wobei A eine Matrixdarstellung von f für eine beliebig gewählte Basis von V ist. Das charakteristische Polynom ist ein Polynom vom Grad n mit der Unbekannten λ und wird mit $P_f(\lambda)$ bezeichnet.

Satz: Sei V ein n–dimensionaler Vektorraum und $f \in \operatorname{Hom}(V, V)$. Ein Element α des Körpers ist genau dann Eigenwert von f, wenn es Nullstelle des charakteristischen Polynoms $P_f(\lambda)$ ist.

Beweis: Wie wir bereits wissen, ist ein α genau dann ein Eigenwert von f, wenn es einen (von $\vec{0}$ verschiedenen!) Eigenvektor zu α gibt, d.h. wenn der Eigenraum $E_\alpha = \operatorname{Ker}(f - \alpha \cdot \operatorname{Id}_V)$ mindestens Dimension 1 hat. Nach Dimensionsformel gilt dann:

$$\dim\ (\operatorname{Im}\ (f - \alpha \cdot \operatorname{Id}_V)) = \dim V - \dim\ (\operatorname{Ker}\ (f - \alpha \cdot \operatorname{Id}_V)) \leq n - 1$$

Da ein Endomorphismus genau dann vollen Rang hat, wenn seine Determinante ungleich 0 ist, kann man obige Bedingung äquivalent in $\det(f - \alpha \cdot \operatorname{Id}_V) = 0$ umwandeln und damit ist α Nullstelle des charakteristischen Polynoms.

Beispiel:

Für den durch die Matrix $A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & -1 & 2 \\ 2 & 0 & 1 \end{pmatrix}$ gegebenen Endomorphismus f sollen alle Eigenwerte und Basen der jeweiligen Eigenräume bestimmt werden.

- Charakteristisches Polynom:

$$P_f(\lambda) = \det\ (A - \lambda \cdot E_n) = \det \begin{pmatrix} 1-\lambda & 0 & 2 \\ 0 & -1-\lambda & 2 \\ 2 & 0 & 1-\lambda \end{pmatrix} = -\lambda^3 + \lambda^2 + 5\lambda + 3$$

- Nullstellen des charakteristischen Polynoms:
  $\lambda_1 = -1$ (eine doppelte Nullstelle) und $\lambda_2 = 3$ (einfache Nullstelle)

- Basis des Eigenraums E_{-1}:
  Dazu betrachtet man den Kern der Matrix $(A - (-1)E_n) = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 0 & 2 \\ 2 & 0 & 2 \end{pmatrix}$.
  Da der Rang dieser Matrix 2 ist, hat der Kern nur Dimension 1 und offensichtlich ist der Vektor $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ ein Basisvektor des Kerns.

- Basis des Eigenraums E_3:
  Dazu betrachtet man den Kern der Matrix $(A - 3E_n) = \begin{pmatrix} -2 & 0 & 2 \\ 0 & -4 & 2 \\ 2 & 0 & -2 \end{pmatrix}$.
  Da der Rang dieser Matrix 2 ist, hat der Kern nur Dimension 1 und offensichtlich ist der Vektor $\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$ ein Basisvektor des Kerns.
- Schlussfolgerung: Da die Summe der Eigenraumdimensionen kleiner als 3 ist, ist die Abbildung f nicht diagonalisierbar.

Hauptachsentransformation

Wir werden in diesem Abschnitt nachweisen, dass ein Endomorphismus auf einem reellen Vektorraum mit einer symmetrischen Matrix bezüglich einer Orthonormalbasis in jedem Fall diagonalisierbar ist. Die Methode zur Bestimmung einer Diagonalbasis nennt man Hauptachsentransformation. Zuerst beschäftigen wir uns mit einer Charakterisierung solcher Endomorphismen.

Selbstadjungierte Endomorphismen

Definition: Ein Endomorphismus $f \in \text{Hom}\,(V, V)$ auf einem Euklidischen Vektorraum V ist *selbstadjuingiert*, wenn für alle $\vec{v}, \vec{w} \in V$ die folgende Gleichung erfüllt ist:

$$\langle\, f(\vec{v})\,,\, \vec{w}\,\rangle \;=\; \langle\, \vec{v}\,,\, f(\vec{w})\,\rangle.$$

Satz: Ein Endomorphismus $f \in \text{Hom}\,(V, V)$ ist genau dann selbstadjungiert, wenn seine Matrix A bezüglich einer Orthonormalbasis $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ symmetrisch ist.

Beweis: Bekanntlich repräsentiert die j–te Spalte von A das Bild des Basisvektors $\vec{v}_j$ durch

$$f(\vec{v}_j) \;=\; \sum_{i=1}^{n} a_{i\,,\,j}\, \vec{v}_i$$

Da B eine Orthonormalbasis ist, ergibt sich $\langle\, f(\vec{v}_j)\,,\, \vec{v}_i\,\rangle = a_{i\,j}$ und $\langle\, f(\vec{v}_i)\,,\, \vec{v}_j\,\rangle = a_{j\,i}$. Setzt man voraus, dass f selbstadjungiert ist, folgt unter Berücksichtigung der Symmetrie des Skalarprodukts

$$a_{i\,j} = \langle\, f(\vec{v}_j)\,,\, \vec{v}_i\,\rangle = \langle\, \vec{v}_j\,,\, f(\vec{v}_i)\,\rangle = \langle\, f(\vec{v}_i)\,,\, \vec{v}_j\,\rangle = a_{j\,i}.$$

Damit ist gezeigt, dass die Matrix A symmetrisch ist.
Für die Gegenrichtung setzen wir die Symmetrie von A voraus. Aus der Umkehrung der bisherigen Betrachtungen folgt, dass dann die definierende Eigenschaft eines selbstadjungierten Endomorphismus zumindest für die Basisvektoren erfüllt ist, d.h.

$$\langle\, f(\vec{v}_i)\,,\, \vec{v}_j\,\rangle \;=\; \langle\, \vec{v}_i\,,\, f(\vec{v}_j)\,\rangle.$$

für alle $\vec{v}_i, \vec{v}_j \in B$. Um diese Eigenschaft auf beliebige Vektoren $\vec{v}, \vec{w} \in V$ zu übertragen, reicht es aus, die Definition von linearen Abbildungen und die Bilinearität des Skalarprodukts für zwei Vektoren

$$\vec{v} = \sum_{i=1}^{n} \lambda_i \vec{v}_i \qquad \text{und} \qquad \vec{w} = \sum_{j=1}^{n} \mu_j \vec{v}_j$$

anzuwenden:

$$\begin{aligned}
\langle\, f(\vec{v})\,,\, \vec{w}\,\rangle &= \left\langle\, f(\sum_{i=1}^{n}\lambda_i\vec{v}_i)\;,\;\sum_{j=1}^{n}\mu_j\vec{v}_j\,\right\rangle \\
&= \left\langle\,\sum_{i=1}^{n}\lambda_i f(\vec{v}_i)\;,\;\sum_{j=1}^{n}\mu_j\vec{v}_j\,\right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i n\,\mu_j\,\langle\, f(\vec{v}_i)\,,\,\vec{v}_j\,\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\,\mu_j\,\langle\,\vec{v}_i\,,\,f(\vec{v}_j)\,\rangle \\
&= \left\langle\,\sum_{i=1}^{n}\lambda_i\vec{v}_i\;,\;\sum_{j=1}^{n}\mu_j f(\vec{v}_j)\,\right\rangle \\
&= \langle\,\vec{v}\,,\,f(\vec{w})\,\rangle \qquad\qquad \square
\end{aligned}$$

Beobachtung 1: Ist $f \in \mathrm{Hom}\,(V,V)$ ein selbstadjungierter Endomorphismus und sind $\vec{v}$ und $\vec{w}$ zwei Eigenvektoren von f zu verschiedenen Eigenwerten λ und μ, dann sind $\vec{v}$ und $\vec{w}$ orthogonal zueinander.

Beweis: Aus den Voraussetzungen und der Bilinearität des Skalarprodukts folgt

$$\begin{aligned}
\langle\, f(\vec{v})\,,\,\vec{w}\,\rangle &= \langle\,\lambda\,\vec{v},\,,\,\vec{w}\,\rangle = \lambda\,\langle\,\vec{v}),\,,\,\vec{w}\,\rangle \\
\langle\,\vec{v}\,,\,f(\vec{w})\,\rangle &= \langle\,\vec{v}\,,\,\mu\,\vec{w}\,\rangle\; = \mu\,\langle\,\vec{v}\,,\,\vec{w}\,\rangle
\end{aligned}$$

Da f selbstadjungiert ist, sind beide Terme gleich und daraus ergibt sich

$$\underbrace{(\lambda-\mu)}_{\neq 0}\,\langle\,\vec{v}\,,\,\vec{w}\,\rangle = 0 \qquad\Longrightarrow\qquad \langle\,\vec{v}\,,\,\vec{w}\,\rangle = 0 \qquad\qquad \square$$

Beobachtung 2: Ist $f \in \mathrm{Hom}\,(V,V)$ ein selbstadjungierter Endomorphismus und $\vec{v}$ ein Eigenvektor (zum Eigenwert λ) von f, dann ist die Beschränkung von f auf den Unterraum $U = \{\vec{v}\}^{\perp}$ wieder selbstadjungierter Endomorphismus.

Beweis: Es genügt nachzuweisen, dass f alle Vektoren $\vec{u} \in U$ nach U abbildet. Da U orthogonal zu $\vec{v}$ ist, gilt $\langle\,\vec{u}\,,\,\vec{v}\,\rangle = 0$. Um zu zeigen, dass auch $f(\vec{u})$ orthogonal zu $\vec{v}$ ist, nutzt man die Eigenschaften von selbstadjungierten Endomorphismen und von Eigenvektoren:

$$\langle\, f(\vec{u})\,,\,\vec{v}\,\rangle = \langle\,\vec{u}\,,\,f(\vec{v})\,\rangle = \langle\,\vec{u}\,,\,\lambda\,\vec{v}\,\rangle = \lambda\,\langle\,\vec{u}\,,\,\vec{v}\,\rangle = 0 \qquad\qquad \square$$

Satz: Jeder selbstadjungierte Endomorphismus $f \in \mathrm{Hom}\,(V,V)$ auf einen n–dimensionalen, Euklidischen Vektorraum V ist diagonalisierbar und besitzt eine orthonormale Diagonalbasis.

Beweisidee: Man führt einen Induktionsbeweis nach n. Der Induktionsanfang für $n = 1$ ist trivial, da jede (1×1)–Matrix bereits in Diagonalform ist.
Für den Induktionsschritt setzt man voraus, dass der Satz schon für alle selbstadjungierten Endomorphismen auf $(n-1)$–dimensionalen Räumen bewiesen ist und betrachtet einen selbstadjungierten Endomorphismus f mit einer symmetrischen $(n \times n)$–Matrix A bezüglich

einer Orthonormalbasis von V. Nach dem Fundamentalsatz der Algebra hat das charakteristische Polynom $P_f(\lambda) = \det(A - \lambda E_n)$ mindestens eine komplexe Nullstelle. Interpretiert man A als Matrix eines Endomorphismus auf $\mathbb{C}^n$, dann hat diese Erweiterung mindestens einen komplexen Eigenvektor $\vec{v}$, den man in seinen Real– und Imaginärteil aufspalten kann. Durch Nutzung der Tatsache, dass f selbstadjungiert ist, kann man (relativ einfach) nachrechnen, dass der Imaginärteil dieses Eigenvektors Null ist und folglich der zugehörige Eigenwert reell ist. Natürlich ist dann auch der normierte Vektor $\tilde{v} = \frac{\vec{v}}{\|\vec{v}\|}$ ein Eigenvektor von f.
Da nach Beobachtung 2 und nach Induktionsvoraussetzung die Einschränkung von f auf den Unterraum $U = \{\vec{v}\}^\perp$ diagonalisierbar ist, muss man zur orthogonalen Diagonalbasis dieser Einschränkung nur noch den Vektot $\tilde{v}$ hinzufügen und hat damit die Induktionsbehauptung nachgewiesen.

Algorithmus zur Hauptachsentransformation

- Gegeben ist ein selbstadjungierter Endomorphismus auf $\mathbb{R}^n$ durch seine symmetrische Matrix A bezüglich der Standardbasis.

- Bestimmung des charakteristischen Polynoms $P_f(\lambda) = \det(A - \lambda E_n)$ und aller Nullstellen von $P_f(\lambda)$:
  Anmerkung: Obwohl gesichert ist, dass alle Nullstellen reell sind, kann ihre Bestimmung im Allgemeinen nur durch numerische Näherungsverfahren gelöst werden.

- Sei $\{\lambda_1, \ldots, \lambda_k\}$ die Menge aller Nullstellen von $P_f(\lambda)$. Man bestimmt für die Eigenräume $E_{\lambda_1}, \ldots, E_{\lambda_k}$ jeweils eine Basis durch Lösung der linearen Gleichungsysteme $(A - \lambda_i E_n) \cdot \vec{x} = \vec{0}$ für alle $i = 1, \ldots, k$.

- Die Basen der Eigenräume werden durch das Schmidtsche Orthonormalisierungsverfahren in Orthogonalbasen umgewandelt.

- Die orthonormale Diagonalbasis von f entsteht als Vereinigung der Orthonormalbasen aller Eigenräume.

Kapitel 2

Endliche Körper und Lineare Codes

2.1 Endliche Körper

Teilbarkeit und Teilen mit Rest

In diesem Abschnitt werden wir uns mit Teilbarkeitsrelation auf der Menge der ganzen Zahlen, dem Teilen mit Rest und dem Rechnen mit Resten genauer vertraut machen. Wir bezeichen mit $\mathbb{Z}$ die Menge der ganzen Zahlen und mit $\mathbb{Z}^+$ die Menge der positiven ganzen Zahlen (ohne Null). In den folgenden Definitionen und Sätzen werden noch einmal die wichtigsten Grundbegriffe zusammengefasst.

Definition: Die hier auftretenden Variablen a, b, c, d und p bezeichnen immer ganze Zahlen.

1. $b \neq 0$ ist ein *Teiler* von a, wenn es ein c gibt, so dass $a = b \cdot c$ gilt. Man sagt dann auch, dass a durch b *teilbar* ist und drückt das symbolisch durch $b \,|\, a$ (gesprochen b teilt a) aus. Nach dieser Definition ist 0 durch jede ganze Zahl $b \neq 0$ teilbar.

2. $p > 1$ heißt *irreduzibel*, wenn p nur die Teiler $-p, -1, 1$ und p hat.

3. $p > 1$ heißt *Primzahl*, wenn für alle $a, b \in \mathbb{Z}$ die folgende Implikation gilt:

$$p \,|\, ab \implies p \,|\, a \vee p \,|\, b$$

4. Die Zahl $c > 0$ ist der größte gemeinsamer Teiler von a und b (geschrieben $c = \mathrm{ggT}(a, b)$), wenn c Teiler von a und von b ist und jeder andere gemeinsame Teiler von a und b auch ein Teiler von c ist, d.h. wenn

$$c \,|\, a \ \wedge \ c \,|\, b \ \wedge \ \forall d \ (\, c \,|\, a \ \wedge \ c \,|\, b \ \longrightarrow \ d \,|\, c\,)$$

5. a und b heißen *teilerfremd* oder *koprim*, wenn $\mathrm{ggT}(a, b) = 1$.

Bemerkung: Für die ganzen Zahlen sind die Begriffe irreduzibel und Primzahl äquivalent und deshalb werden in der Schulmathematik Primzahlen wie in Punkt 2 definiert. Der Beweis dieser Äquivalenz ist nicht ganz offensichtlich. Wir werden sie hier als gegeben voraussetzen und begnügen uns mit dem Hinweis, dass man den Beweis auf die in diesem Abschnitt besprochene Umkehrung des Euklidischen Algorithmus zurückführen kann. Letztlich folgt aus dieser Äquivalenz auch der Fakt, dass jede positive ganze Zahl eine eindeutige Primzahlzerlegung hat.

Satz über die ganzzahlige Division: Für beliebige $a \in \mathbb{Z}$ und $d \in \mathbb{Z}^+$ existieren eindeutig bestimmte ganze Zahlen q und r mit $0 \leq r < d$, so daß $a = qd + r$.

Definition: Sind $a, q, r \in \mathbb{Z}$, $d \in \mathbb{Z}^+$ mit $0 \leq r < d$ und $a = qd + r$, dann wird q der ganzzzahlige Quotient aus a und d genannt und r als *Rest von a bezüglich (modulo) d* bezeichnet. Als Notation verwenden wir

$$q = \left\lfloor \frac{a}{d} \right\rfloor \quad \text{und} \quad r = a \bmod d.$$

Zwei ganze Zahlen a und b, die den gleichen Rest bezüglich d haben, werden *kongruent* bezüglich (modulo) d genannt, wofür die folgende Schreibweise vereinbart wird:

$$a \equiv b \pmod{d}$$

Satz: Die Relation $\equiv \pmod{d}$ ist eine Äquivalenzrelation und die Zahlen $\{0, 1, \ldots, d-1\}$ bilden ein Repräsentantensystem für die Äquivalenzklassen. Darüber hinaus ist die Relation verträglich mit der Addition, Subtraktion und Multiplikation, d.h.:

ist	a	$\equiv$	a'	$(\bmod d)$
und	b	$\equiv$	b'	$(\bmod d)$
dann ist auch	$(a+b)$	$\equiv$	$(a'+b')$	$(\bmod d)$
und	$(a-b)$	$\equiv$	$(a'-b')$	$(\bmod d)$
und	ab	$\equiv$	$a'b'$	$(\bmod d)$

Euklidischer Algorithmus

Lemma: Seien $a, b \in \mathbb{Z}^+$ mit $a > b$ und sei $r = a \bmod b$.
1) Ist $r = 0$, dann gilt $b|a$ und $\mathrm{ggT}(a, b) = b$.
2) Ist $r \neq 0$, dann gilt $ggT(a, b) = \mathrm{ggT}(b, r)$.

Der *Euklidische Algorithmus* zur rekursiven Berechnung des größten gemeinsamen Teilers ist eine direkte Anwendung dieses Lemmas, wobei der erste Punkt die Abbruchbedingung und der zweite Punkt einen Rekursionsschritt beschreibt:

procedure $\mathrm{ggT}(a, b : \mathit{aus}\ \mathbb{Z}^+)$
$x := a$
$y := b$
while $y \neq 0$
  $r := x \bmod y$
  $x := y$
  $y := r$
return x

Beispiel: Berechnung von $\mathrm{ggT}(252, 198)$

$$\begin{aligned} 252 &= 1 \cdot 198 + 54 \\ 198 &= 3 \cdot 54 + 36 \\ 54 &= 1 \cdot 36 + 18 \\ 36 &= 2 \cdot 18 + 0 \quad \rightsquigarrow \quad \mathrm{ggT}(252, 198) = 18 \end{aligned}$$

Satz: Sind $a, b \in \mathbb{Z}^+$, dann kann man $\mathrm{ggT}(a, b)$ als Linearkombination von a und b mit ganzzahligen Koeffizienten darstellen, d.h. es existieren $r, s \in \mathbb{Z}$, so daß

$$\mathrm{ggT}(a, b) = sa + tb.$$

Beweis: Die Idee ist sehr einfach – man muss den Euklidischen Algorithmus nur umkehren. Zur Illustration zeigen wir das am obigen Beispiel. Zuerst werden alle Gleichungen nach dem Rest umgestellt:

$$\begin{aligned} 18 &= 54 - 1 \cdot 36 \\ 36 &= 198 - 3 \cdot 54 \\ 54 &= 252 - 1 \cdot 198 \end{aligned}$$

Durch schrittweise Substitution erhalten wir:

$$\begin{aligned} 18 = \; & 54 - 1 \cdot 36 = 54 - 1 \cdot (198 - 3 \cdot 54) = 4 \cdot 54 - 1 \cdot 198 = \\ & = 4 \cdot (252 - 1 \cdot 198) - 1 \cdot 198 = 4 \cdot 252 - 5 \cdot 198 \end{aligned}$$

Der formale Beweis erfolgt durch vollständige Induktion nach der Anzahl n der Durchläufe der while–Schleife, die der Euklidische Algorithmus auf der Eingabe (a, b) ausführt.
Induktionsanfang: Im Fall $n = 1$ ist der Rest $r = a \bmod b$ gleich Null und $\mathrm{ggT}(a, b) = b$. Wie man leicht sieht, leistet die Linearkombination $b = 0 \cdot a + 1 \cdot b$ das Gewünschte.
Induktionsschritt: Wenn die while–Schleife mehr als einmal durchlaufen wird, ist der Rest $r = a \bmod b$ größer als Null und $\mathrm{ggT}(a, b) = \mathrm{ggT}(b, r)$. Da der Euklidische Algorithmus die while–Schleife für die Eingabe (b, r) einmal weniger durchläuft als für (a, b) kann man die Induktionsvoraussetzung auf (b, r) anwenden:

$$\exists s, t \in \mathbb{Z} \; \mathrm{ggT}(b, r) \; = \; sb \; + \; tr$$

Wir stellen die Gleichung $a = qb + r$ nach r um, setzen das Ergebnis in obige Linearkombination ein

$$\mathrm{ggT}(a, b) = \mathrm{ggT}(b, r) \; = \; sb \; + \; tr \; = \; sb \; + \; t(a - qb) = ta \; + \; (s - tq)b$$

und haben damit die Induktionsbehauptung nachgewiesen. □

Satz: Seien m und a zwei positive, teilerfremde Zahlen (d.h. $\mathrm{ggT}(a, m) = 1$), dann gibt es genau ein $b \in \{0, 1, \ldots, m - 1\}$, so daß $ab \equiv 1 \bmod m$.
Beweis: Es existieren $s, t \in \mathbb{Z}$ mit $sa + tm = 1$. Setzt man $b := s \bmod m$, so erhält man

$$\begin{array}{lcll} b & \equiv & s & (\mathrm{mod}\ m) \qquad (1) \\ a & \equiv & a & (\mathrm{mod}\ m) \qquad (2) \\ 0 & \equiv & tm & (\mathrm{mod}\ m) \qquad (3) \\ ba & \equiv & sa & (\mathrm{mod}\ m) \qquad (4) \end{array}$$

wobei sich (4) durch die multiplikative Verknüpfung von (1) und (2) ergibt. Durch additive Verknüpfung von (3) und (4) erhalten wir

$$ba + 0 \quad \equiv \quad sa + tm \quad \equiv \quad 1 \pmod{m}.$$

Zum Beweis der Eindeutigkeit sei $ba \equiv ca \bmod m$ und $0 \leq b \leq c \leq m - 1$. Dann ist $(c - b)a \equiv 0 \bmod m$, und da m und a teilerfremd sind, muß m ein Teiler von $c - b$ sein. Wegen $0 \leq c - b < m$ folgt dann $c = b$ (die Eindeutigkeit). □

Folgerung: Sei p eine Primzahl und $a \in \{1, 2, \ldots, p - 1\}$, dann hat a ein eindeutig bestimmtes Inverses modulo p, d.h. es gibt genau ein $b \in \{1, 2, \ldots, p - 1\}$, so daß $ab \equiv 1 \bmod p$.

Beweis: Setzt für das m aus dem vorhergehenden Satz die Primzahl p, so ist $m = p$ teilerfremd zu jedem $a \in \{1, 2, \ldots, p-1\}$ und damit ergibt sich die Existenz und die Eindeutigkeit von b aus dem Satz. □

Endliche Körper

Definition: Für jede Primzahl p bildet die Menge $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ der Reste beim Teilen durch p mit den Operationen Addition modulo p und Multiplikation modulo p einen endlichen Körper, der auch mit $GF(p)$ bezeichnet wird.
Will man die Operationen in $GF(p)$ deutlich von der Addition und Multiplikation in $\mathbb{Z}$ abgrenzen, kann man neue Symbole wie $+_p$ und $\cdot_p$ einführen, aber allgemein werden nach einer Festlegung, dass in $GF(p)$ gerechnet wird, die üblichen Operationszeichen $+$ und $\cdot$ verwendet. Das zu einem $a \in \mathbb{Z}_p$ bezüglich Addition inverse Element wird in diesem Sinne mit $-a$ bezeichnet. Für das bezüglich Multiplikation inverse Element verwendet man die Bezeichnung a^{-1}.

Beispiel: Der Körper $GF(5)$ hat 5 Elemente: $\mathbb{Z}_5 = \{1, 2, 3, 4, 5\}$.
In diesem Körper ist $3 +_5 4 = (3 + 4) \bmod 5 = 7 \bmod 5 = 2$ und $3 \cdot_5 4 = (3 \cdot 4) \bmod 5 = 12 \bmod 5 = 2$. Da wir uns schon auf $GF(5)$ festgelegt haben, kann man auch kurz $3+4 = 2$ und $3 \cdot 4 = 2$ schreiben.
Das bezüglich Addition inverse Element zu 2 ist die 3, denn $2 + 3 = 0$.

Allgemein gilt in $GF(p)$ die Regel $-0 = 0$ und $-a := p - a$ für alle $a \in \{1, 2, \ldots, p-1\}$.

Die bezüglich Multiplikation inversen Elemente kann man für kleine Primzahlen durch systematische Probieren bestimmen: In $GF(5)$ ist $4 \cdot 2 = 3 \neq 1$ und $4 \cdot 3 = 2 \neq 1$, aber $4 \cdot 4 = 1$, also ist $4^{-1} = 4$.

Als allgemeine Methode zur Bestimmung von a^{-1} in $GF(p)$ bleibt nur der in den vorangegangenen Sätzen dargestellte Weg:

1. Berechnung von $\mathrm{ggT}(a, p)$ mit dem Euklidischen Algorithmus.
2. Darstellung von $1 = \mathrm{ggT}(a, p)$ als Linearkombination $1 = s \cdot a + t \cdot p$ durch Umkehrung des Euklidischen Algorithmus.
3. Verwendung der Formel $a^{-1} := s \bmod p$.

Beispiel: In $GF(19)$ benötigt man zur Bestimmung von 8^{-1} die Zwischenschritte $19 = 2 \cdot 8 + 3$, $8 = 2 \cdot 3 + 2$ und $3 = 1 \cdot 2 + 1$ aus dem Euklidischen Algorithmus. Durch Umkehrung (d.h. Umstellung nach den Resten und schrittweise Substitution) ergibt sich:

$$1 = 3 - 2 = 3 - (8 - 2 \cdot 3) = 3 \cdot 3 - 8 = 3 \cdot (19 - 2 \cdot 8) - 8 = 3 \cdot 19 - 7 \cdot 8.$$

Folglich ist $8^{-1} = -7 \bmod 19 = 12$, was man durch $8 \cdot 12 = 96$ (in $\mathbb{Z}$) und $96 \bmod 19 = 1$ leicht überprüfen kann.

Die Konstruktion von weiteren (von $GF(p)$ verschiedenen) endlichen Körpern ist wesentlich schwieriger. Wir begnügen uns hier mit dem folgenden Faktenwissen ohne Beweis.

Satz: Für jeden endlichen Körper K gibt es eine Primzahl p und einen Exponenten $k \in \mathbb{Z}^+$, so dass K genau p^k Elemente hat und einen zu $GF(p)$ isomorphen Unterkörper enthält. Umgekehrt, gibt es zu jeder Primzahlpotenz p^k einen Körper mit p^k Elementen, der $GF(p)$ als Unterkörper enthält. Man verwendet für einen solchen Körper die Bezeichnung $GF(p^k)$.

Es ist wichtig zu betonen, dass die Operationen in $GF(p^k)$ für alle $k > 1$ **nicht** als Addition bzw. Multiplikation modulo p^k ausgeführt werden.

Chinesischer Restesatz

Abschließend wollen wir einen weiteren Satz zum Rechen mit Resten behandeln. Die Bedeutung dieses Satzes für endliche Körper ist zwar eher marginal, aber er hat dafür wichtige Anwendunegn in der Kryptographie.

Chinesischer Restesatz: Seien $m_1, m_2, \ldots, m_n$ paarweise teilerfremde, positive Zahlen und $m = m_1 \cdot m_2 \cdot \ldots \cdot m_n$. Dann gibt es für beliebig gewählte $a_1, \ldots, a_n$ genau eine Zahl x mit $0 \leq x < m$, die die folgenden Kongruenzen erfüllt:

$$\begin{aligned} x &\equiv a_1 \mod m_1, \\ x &\equiv a_2 \mod m_2, \\ &\vdots \\ x &\equiv a_n \mod m_n. \end{aligned}$$

Beweis: Definiert man $M_k = m/m_k$ für $k = 1, 2, \ldots, n$, dann ist $ggT(M_k, m_k) = 1$, und folglich gibt es Zahlen y_k mit

$$M_k y_k \equiv 1 \mod m_k.$$

Wir setzen $x = a_1 M_1 y_1 + a_2 M_2 y_2 + \ldots + a_n M_n y_n \bmod m$ und überprüfen, ob es die obigen Kongruenzen erfüllt. Zuerst stellen wir fest, daß für alle $k \neq j$ m_k ein Teiler von M_j ist, d.h.

$$\begin{aligned} M_j &\equiv 0 \mod m_k. \qquad \text{Daraus folgt für alle } k \\ x &\equiv 0 + \ldots + 0 + a_k M_k y_k + 0 + \ldots + 0 \equiv a_k \cdot 1 \equiv a_k \mod m_k \end{aligned}$$

Zum Beweis der Eindeutigkeit nimmt man an, daß es zwei Lösungen $0 \leq x \leq y < m$ gibt. Dann sind alle m_k Teiler von $y - x$. Nach Voraussetzung (paarweise teilerfremd) ist auch m Teiler von $y - x$, und folglich ist $y - x = 0$, d.h. $y = x$. □

Anwendung 1: Exakte Arithmetik

In vielen Algorithmen ist es erforderlich, auch mit sehr großen Zahlen (die weit über den `Int` der Rechner liegen) exakt zu rechnen, insbesondere dann, wenn die Fließkomma-Arithmetik zu ungenau wird. Die üblicherweise verwendete Lösung besteht darin, große Zahlen durch Folgen (Listen) von `Int`-Werten darzustellen. Die Multiplikation zweier aus jeweils n `Int`-Werten zusammengesetzter Zahlen erfordert dann quadratischen Aufwand.

Alternativ kann man aber auch n Primzahlen $p_1, \ldots, p_n$ auswählen, die fast so groß wie der maximale `Int`-Werte sind und die Operanden durch die n–Tupel ihrer Reste modulo der Primzahlen repräsentieren. Bei Nutzung des Chinesischen Restesatzes ist der Aufwand für eine Multiplikation von zwei durch Restetupel repräsentierte Zahlen nur linear, denn man muss nur die jeweiligen Reste modulo p_k multipizieren. Dafür ist es aufwendig, aus einer Resterepräsentation das Ergebnis explizit zu berechnen (siehe Beweis des Restklassensatzes). Deshalb ist diese alternative Methode vor allem dann geeignet, wenn ein Algorithmus viele Zwischenergebnisse produziert, deren explizite Darstellung nicht benötigt wird, zum Beispiel bei der Polynomauswertung mit dem Horner–Schema oder bei Matrixmultiplikationen.

Anwendung 2: RSA-Codes

In der Kryptographie untersucht man Methoden zur Verschlüsselung von geheimen Nachrichten (Quelltext) in eine Binärfolge (Kanalcode), die über einen Übertragungskanal zum Empfänger geschickt wird. Der Empfänger soll in der Lage sein, den Kanalcode zu entschlüsseln, d.h. ihn wieder in den Quelltext zurückzuverwandeln, eventuelle Lauscher sollen dazu aber nicht in der Lage sein.
Lange ging man davon aus, daß zu diesem Zweck sowohl die Verschlüsselungs- als auch Entschlüsselungsmethode geheim sein müssen, und es war schon eine Pionierleistung, einfach diese Denkblockade zu durchbrechen und nach *Public–Key–Systemen* zu suchen. Solche Systeme zeichnen sich dadurch aus, dass die Verschlüsselungsmethode nicht mehr geheim gehalten werden muß. 1976 entwickelten R. **R**ivest, A. **S**hamir und L. **A**dleman ein solches Public–Key–System, die nach ihnen benannten RSA-Codes.

Eine wichtige Voraussetzung dieses Ansatzes besteht darin, dass der Quelltext bereits als Binärstring vorliegt und durch Unterteilung in Blöcke fester Länge als Folge von Zahlen (beschränkter Größe) interpretiert werden kann. Damit reduziert sich die Verschlüsselung und Entschlüsselung auf die Codierung und Decodierung von positiven ganzen Zahlen. Zur Installierung und Nutzung eines RSA-Codes geht man nach folgendem Protokoll vor:

- Der Nachrichtenempfänger **Bob** generiert zwei große Primzahlen p und q sowie eine Zahl $e \in \mathbb{Z}$, die teilerfremd zu $(p-1)(q-1)$ ist. Er berechnet das Produkt $n = p \cdot q$ und die (eindeutige) Zahl $d \in \{1, 2, \ldots, (p-1)(q-1)-1\}$ für welche die Bedingung

  $$c \cdot d \;\equiv\; 1 \mod (p-1)(q-1)$$

  erfüllt ist. Bob gibt n und e bekannt, die Zahlen p, q und d bleiben sein Geheimnis.

- Die Nachrichtensenderin **Alice** zerlegt den zu übetragenden Quelltext in Blöcke der Länge $K = \lfloor \log_2 n \rfloor$ und interpretiert jeden Block als eine Zahl $m < n$. Jedes m wird durch die folgende Formel chiffriert:

  $$E(m) := (m^e) \bmod n$$

  Alice sendet die Folge der $E(m)$ an Bob.

- Bob empfängt die Folge der chiffrierten Blöcke. Jeder Block wird wieder als eine Zahl $M < n$ interpretiert und wie folgt dechiffriert:

  $$D(M) := (M^d) \bmod n$$

  Danach wird jedes $D(M)$ durch führende Nullen auf die Länge K erweitert und der Quelltext durch Konkatenation dieser Blöcke zusammengesetzt.

Der zentrale Punkt dieser Methode ist die Tatsache, dass die Funktion D wirklich die Umkehrung der Funktion E ist. Zum Beweis verwendet man einen alten Satz aus der Zahlentheorie, den wir hier nicht beweisen werden.

Kleiner Satz von Fermat: Ist p eine Primzahl, dann gilt für jede nicht durch p teilbare Zahl a

$$a^{p-1} \;\equiv\; 1 \mod p$$

Satz: Für jede ganze Zahl m zwischen 0 und $n-1$ gilt $D(E(m)) \equiv m$.

Beweis: Nach Definition der Funktionen E und D ist klar, dass ihr Bild in $\{0, 1, \ldots, n-1\}$ liegt. Deshalb ist es hinreichend, $D(E(m)) \equiv m \mod n$ nachzuweisen, und da $n = p \cdot q$ das Produkt von zwei teilerfremden Zahlen ist, reicht es nach chinesischem Restesatz aus,

$$D(E(m)) \equiv m \mod p \qquad \text{und} \qquad D(E(m)) \equiv m \mod q$$

zu zeigen. Da man die Rollen von p und q vertauschen kann, konzentrieren wir uns auf die Kongruenz modulo p und unterscheiden die folgenden zwei Fälle:

1. $m \equiv 0 \mod p$:
   In diesem Fall ist p ein Teiler von m, damit ein Teiler von m^e und wegen $n = p \cdot q$ ist p auch ein Teiler $(m^e \bmod n)$. Folglich ist $E(m) \equiv 0 \mod p$ und mit der gleichen Argumentation ergibt sich $D(E(m)) \equiv 0 \mod p$, also $D(E(m)) \equiv m \mod p$.

2. $m \not\equiv 0 \mod p$:
   In diesem Fall kommt der kleine Satz von Fermat zum Einsatz und ergibt die Kongruenz $m^{p-1} \equiv 1 \mod p$.
   Nach Voraussetzung ist $d \cdot e \equiv 1 \mod ((p-1)(q-1))$. Folglich gibt es ein $k \in \mathbb{Z}$, so dass $d \cdot e = 1 + k(p-1)(q-1)$.
   Schließlich ist $D(E(m)) \equiv m^{d \cdot e} \mod n$ und da p ein Teiler von n ist sind beide Seiten auch kongruent modulo p. Nach den Rechenregeln mit Kongruenzen ergibt die Nutzung der oben zusammengestellten Fakten:

$$\begin{aligned} D(E(m)) &\equiv m^{d \cdot e} && \mod p \\ &\equiv m^{1+k(p-1)(q-1)} && \mod p \\ &\equiv m \cdot \left(m^{(p-1)}\right)^{k(q-1)} && \mod p \\ &\equiv m \cdot 1^{k(q-1)} && \mod p \\ &\equiv m && \mod p \end{aligned}$$ □

Die Sicherheit dieses Systems beruht allein auf der Geheimhaltung der Zahl d. Wie wir gesehen haben, kann Bob (und damit auch jeder potentielle Gegenspieler) d ausrechnen, wenn er p und q kennt. Bob muss also sehr große Primzahlen p und q wählen, damit die Zerlegung von n in die Faktoren p und q für den Gegenspieler zu einer praktisch unlösbaren Aufgabe wird – theoretisch ist das für den Gegenspieler natürlich einfach, aber wenn er dafür hundert Jahre braucht, ist dass für Bob ein zumutbares Risiko.

Abschließend sei bemerkt, dass die Chiffrierungs– und Dechiffrierungsfunktionen von RSA–Systemen leicht zu implementieren sind und schnelle Laufzeiten haben. Insbesondere kann man die Berechnung der Potenzen $(m^e \bmod n)$ durch wiederholtes Quadrieren und Nutzung der Binärdarstellung von e mit $O(\log_2 e)$ Multiplikationen berechnen, im Gegensatz zu $e-1$ Multiplikationen, die man mit dem naiven Ansatz benötigt.

2.2 Fehlererkennung und Fehlerkorrektur in Codes

Blockcodes

Untersuchungen zu Codierungen von Informationen, die über einen Nachrichtenkanal übertragen werden sollen, konzentrieren sich in der Regel auf einen der folgenden Aspekte:

1. Kompaktheit: Die codierte Information sollte möglichst kurz sein.
2. Geheimhaltung: Die im Kanal übertragene Information sollte ohne Kenntnis eines Schlüssels nicht (oder nur sehr schwer) zu decodieren sein.
3. Fehlererkennung und Fehlerkorrektur: Treten bei der Übertragung vereinzelte Fehler auf sollte das erkannt werden und gegebenenfalls korrigiert werden können.

Es ist klar, dass man nicht alle Aspekte in einer Codierung berücksichtigen kann, denn die Aspekte 1 und 3 stehen sich diametral gegenüber: Um eine Information kompakt zu repräsentieren, muss man auf Redundanz verzichten, aber andererseits ist Redundanz notwendig, um Übertragungsfehler erkennen.

Wichtige Themen im Zusammenhang mit dem ersten Aspekt sind statistische Datenanalyse, Stringmatching–Algorithmen und Huffman–Codes.
Der zweite Aspekt wird nicht der Codierungstheorie zugeordnet, sondern begründet ein eigenständiges Forschungsgebiet, die Kryptographie. Mit den RSA-Codes haben wir schon ein Thema diese Gebiets kennengelernt.
Im Folgenden werden wir uns nur noch mit dem dritten Aspekt beschäftigen und werden dabei Blockcodes im Allgemeinen und speziell lineare Codes kennenlernen.

Definition: Sei A eine endliche Menge von zu codierenden Symbolen (Informationen) und Q ein q–elementiges Alphabet, dass sogenannte Kanalalphabet . Eine injektive Funktion $\varphi : A \longrightarrow Q^*$ wird *Codierung* von A genannt. Das Bild der Funktion φ ist der zugehörige *Code* $C = \mathrm{Im}(\varphi)$, die Elemente von C werden *Codewörter* genannt. Der Code C ist ein *Blockcode* der Blocklänge n, wenn alle Codewörter aus C die Länge n haben.

Jeder Blockcode C der Blocklänge n ist damit eine Teilmenge von Q^n, seine Elemente sind n–Tupel. Dabei wird Q in der Regel ein Körper sein. Damit wird Q^n zu einem Vektorraum und die Codewörter zu Vektoren. Zur besseren Lesbarkeit vereinbaren wir die folgenden Bezeichnungen. Mit $\vec{u}, \vec{v}, \vec{w}$ werden beliebige Tupel (Vektoren) aus Q^n benannt. Wenn betont werden soll, dass es sich bei dem Tupel um ein Codewort handelt, verwenden wir die Bezeichnungen $\vec{c}, \vec{c'}$ oder Ähnliches.

Hamming–Abstand

Definition: Der Hamming–Abstand zwischen zwei n–Tupeln $\vec{v} = (v_1, \ldots, v_n)$ und $\vec{w} = (w_1, \ldots, w_n)$ wird mit $d_H(\vec{v}, \vec{w})$ oder kurz mit $d(\vec{v}, \vec{w})$ bezeichnet. Er ist bestimmt durch die Anzahl der Stellen an denen sich die zwei Tupel unterscheiden, d.h.

$$d_H(\vec{v}, \vec{w}) = d(\vec{v}, \vec{w}) = |\,\{i \,|\, 1 \leq i \leq n \wedge v_i \neq w_i\}\,|$$

Zur Fehlerekennung und Fehlerkorrektur muss einschränkend gesagt werden, dass hier nur Fehler der Art betrachtet werden, dass einzelne Einträge eines Tupels, also Symbole aus Q, bei der Übertragung verändert werden. Der Fall von fehlenden oder fälschlicherweise

eingeschobenen Einträgen wird nicht berücksichtigt. Damit ist der Hamming–Abstand ein geeignetes Beschreibungsmittel für die Fehleranzahl:

Wenn bei der Übertragung eines Codeworts $\vec{c}$ genau k Fehler auftreten, dann ist k der Hamming–Abstand zwischen $\vec{c}$ und dem empfangenen Tupel $\vec{v}$. Die folgende Definition dient zur Beschreibung aller Tupel mit höchstens k Fehlern.

Definition: Die Kugel mit Radius $k \in \mathbb{N}$ in Q^n um einen Punkt $\vec{u}$ besteht aus allen Punkten $\vec{v}$ deren Hamming–Abstand zu $\vec{u}$ kleiner oder gleich k ist:

$$B_k(\vec{u}) = \{\, \vec{v} \in Q^n \mid d(\vec{v}, \vec{u}) \leq k \,\}$$

Lemma: Der Hamming–Abstand ist eine Metrik auf Q^n, d.h. für beliebige $\vec{u}, \vec{v}, \vec{w} \in Q^n$ sind die folgenden drei Eigenschaften erfüllt:

1. $d(\vec{u}, \vec{v}) \geq 0$ und $d(\vec{u}, \vec{v}) = 0 \iff \vec{u} = \vec{v}$
2. $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$ (Symmetrie)
3. $d(\vec{u}, \vec{v}) + d(\vec{v}, \vec{w}) \geq d(\vec{u}, \vec{w})$ (Dreiecksungleichung)

Definition: Der Minimalabstand eines Codes $C \subseteq Q^n$ wird mit $d(C)$ bezeichnet und ist der kleinste Abstand zwischen zwei Codewörtern aus C, d.h.

$$d(C) = \min \{\, d(\vec{c}, \vec{c'}) \mid \vec{c}, \vec{c'} \in C \text{ und } \vec{c} \neq \vec{c'} \}$$

Wie wir sehen werden, ist der Minimalabstand ein entscheidender Schlüssel zur Charakterisierung von Fehlererkennungs– und Fehlerkorrektur–Eigenschaften eines Codes.

Definition: Ein Code $C \subseteq Q^n$ ist k–fehlererkennend, wenn für jedes Codewort $\vec{c} \in C$ jedes Tupel $\vec{v} \in B_k(\vec{c}) \setminus \{\vec{c}\}$ (das sich also von $\vec{c}$ an mindestens einer und höchstens k Stellen unterscheidet) nicht in C liegt und damit als fehlerhaft erkannt wird.
Der Code C ist k–fehlerkorrigierend, wenn für jedes Codewort $\vec{c} \in C$ und für jedes Tupel $\vec{v} \in B_k(\vec{c})$ (das sich also von $\vec{c}$ an höchstens k Stellen unterscheidet) $\vec{c}$ das eindeutig näheste Codewort zu $\vec{v}$ ist und damit die $\leq k$ Fehler in $\vec{v}$ durch Suche nach dem nähesten Codewort korrigiert werden können.

Da in dieser Definition der Zusammenhang zwischen Fehlern und Hamming–Abstand schon sehr deutlich hervorgehoben wurde, ist die folgende Charakterisierung von fehlererkennenden und fehlerkorrigierenden Codes offensichtlich.

Satz: Für einen Code $C \subseteq Q^n$ gelten die folgenden Äquivalenzen:

- C ist k–fehlererkennend $\iff \forall\, \vec{c} \in C \quad B_k(\vec{c}) \cap C = \{\vec{c}\}$
  $\iff d(C) \geq k + 1$
- C ist k–fehlerkorrigierend $\iff \forall\, \vec{c}, \vec{c'} \in C \quad \left(\vec{c} \neq \vec{c'} \Rightarrow B_k(\vec{c}) \cap B_k(\vec{c'}) = \emptyset\right)$
  $\iff d(C) \geq 2k + 1$

Beispiele: Die folgenden binären Codierungen (d.h. $Q = \{0, 1\}$) beziehen sich auf die Situation, dass die zu codierende Menge A bereits die Form Q^m hat, also aus binären m–Tupeln besteht, die bei der Codierung durch Hinzufügung redundanter Informationen in n–Tupel umgewandelt werden. Die zu codierenden Tupel haben demnach immer die Form $\vec{v} = (v_1, \ldots, v_m)$ bzw. $\vec{w} = (w_1, \ldots, w_m)$.

1. Doppelcodierung: $\varphi_2 : Q^m \longrightarrow Q^{2m}$ $(n = 2m)$, wobei
$$\varphi_2(v_1, \ldots, v_m) = (v_1, \ldots, v_m, v_1, \ldots, v_m)$$
Der zugehörige Code $C_2 = \mathrm{Im}(\varphi_2)$ hat den Minimalabstand 2, denn $d(\varphi_2(\vec{v}), \varphi_2(\vec{w})) = 2 \cdot d(\vec{v}, \vec{w})$. Damit ist C_2 1–fehlererkennend.

2. Codierung mit Paritätsbit: $\varphi_{par} : Q^m \longrightarrow Q^{m+1}$ $(n = m + 1)$, wobei
$$\varphi_{par}(v_1, \ldots, v_m) = (v_1, \ldots, v_m, p) \quad \text{mit} \quad p = (v_1 + \ldots + v_m) \bmod 2$$
Auch hier hat der zugehörige Code $C_{par} = \mathrm{Im}(\varphi_{par})$ hat den Minimalabstand 2, denn im Fall $d(\vec{v}, \vec{w}) \geq 2$ überträgt sich die Ungleichung automatisch auf die Codewörter und im Fall $d(\vec{v}, \vec{w}) = 1$ müssen die Paritätsbits für $\vec{v}$ und $\vec{w}$ verscheiden sein, woraus sich $d(\varphi_p(\vec{v}), \varphi_p(\vec{w})) = 2$ ergibt. Damit ist C_{par} 1–fehlererkennend.

3. Um den Minimalabstand zu vergrößern kann man von der Doppelcodierung zur dreifachen oder k–fachen Codierung übergehen und erreicht damit $d(C_k) = k$.

4. Eine etwas bessere Alternative zur Dreifachcodierung ist die Doppelcodierung mit Paritätsbit $\varphi_{2,par} : Q^m \longrightarrow Q^{2m+1}$, wobei
$$\varphi_{2,par}(v_1, \ldots, v_m) = (v_1, \ldots, v_m, v_1, \ldots, v_m, p) \quad \text{mit} \quad p = (v_1 + \ldots + v_m) \bmod 2$$
Der Minimalabstand $d(C_{2,par}) = 3$ ergibt sich durch eine einfache Fallunterscheidung.

5. Leider erreicht man durch mehrfache Paritätsbits über der gleichen Grundmenge keine Verbesserung des Minimalabstands. Sinnvoller ist es, mehrere Paritätsbits über verschiedenen (geeignet gewählten!) Teilmengen der Eingabebits zu bilden. Ein Beispiel dafür sind die sogenannten Kreuzsicherungscodes, bei denen m eine Quadratzahl oder zumindest Produkt aus zwei annähernd gleichen ganzen Zahlen sein sollte.
Sei $m = k^2$ dann verwendet man für den Kreuzsicherungscode $\varphi_{kr} : Q^m \longrightarrow Q^{m+2k}$ insgesamt $2k = 2\sqrt{m}$ Paritätsbits. Dazu trägt man die k^2 Bits des Eingabetupels $\vec{v}$ Zeile für Zeile in eine $k \times k$ Matrix ein und bildet alle Zeilenparitätsbits $p_1, \ldots p_k$ und alle Spaltenparitätsbits $p'_1, \ldots p'_k$ und definiert:
$$\varphi_{kr}(v_1, \ldots, v_m) = (v_1, \ldots, v_m, p_1, \ldots, p_k, p'_1, \ldots p'_k)$$
Der Minimalabstand $d(C_{kr}) = 3$ ergibt sich aus der folgenden Fallunterscheidung:

$d(\vec{v}, \vec{w}) \geq 3$		$d\left(\varphi_{kr}(\vec{v}), \varphi_{kr}(\vec{w})\right) \geq 3$
$d(\vec{v}, \vec{w}) = 2$	Unterschiedliche Stellen in einer Zeile	$d\left(\varphi_{kr}(\vec{v}), \varphi_{kr}(\vec{w})\right) = 2 + 2 = 4$ (2 + 2 Spaltenparitätsbits)
	Unterschiedliche Stellen in einer Spalte	$d\left(\varphi_{kr}(\vec{v}), \varphi_{kr}(\vec{w})\right) = 2 + 2 = 4$ (2 + 2 Zeilenparitätsbits)
	sonst	$d\left(\varphi_{kr}(\vec{v}), \varphi_{kr}(\vec{w})\right) = 2 + 2 + 2 = 6$ (2 + 2 Spaltenpar.-bits + 2 Zeilenenpar.-bits)
$d(\vec{v}, \vec{w}) = 1$		$d\left(\varphi_{kr}(\vec{v}), \varphi_{kr}(\vec{w})\right) = 1 + 1 + 1 = 3$ (1 + 1 Spaltenpar.-bit + 1 Zeilenenpar.-bit)

Hamming–Code

Der Kreuzsicherungscode benötigt zur Codierung von $A = \{0,1\}^4$ vier zusätzliche Paritätsbits, um den Minimalabstand 3 zu realisieren. Mit etwas Knobelei kommt man darauf, dass auch schon drei zusätzliche Paritätsbits ausreichen, wie der folgende Code zeigt, der nach Hamming benannt ist:

$$\begin{aligned} \text{Sei} \quad \vec{v} &= (v_1, v_2, v_3, v_4). \\ p_1 &:= (v_2 + v_3 + v_4) \bmod 2 \\ p_2 &:= (v_1 + v_3 + v_4) \bmod 2 \\ p_3 &:= (v_1 + v_2 + v_4) \bmod 2 \\ \varphi_{ham}(\vec{v}) &:= (v_1, v_2, v_3, v_4, p_1, p_2, p_3) \end{aligned}$$

Die Analyse, dass der Minimalabstand dieses Hamming–Codes 3 ist, erfolgt wieder durch eine Fallunterscheidung:
Ist $d(\vec{v}, \vec{w}) \geq 3$, dann überträgt sich diese Ungleichung auch auf die Codewörter.
Im Fall $d(\vec{v}, \vec{w}) = 2$ zeigt sich, dass entweder ein Paritätsbit verschieden ist (wenn sich $\vec{v}$ und $\vec{w}$ an der vierten Stelle unterscheiden) oder sogar zwei (wenn $\vec{v}$ und $\vec{w}$ an der vierten Stelle gleich sind)
Im Fall $d(\vec{v}, \vec{w}) = 1$ sind entweder zwei Paritätsbits verschieden (wenn $\vec{v}$ und $\vec{w}$ an der vierten Stelle gleich sind) oder alle drei (wenn sich $\vec{v}$ und $\vec{w}$ an der vierten Stelle unterscheiden).

Aus diesem Beispiel ergeben sich zwei wichtige Fragen:

- Ist diese Lösung optimal, d.h. muss man zur Codiereung von $A = \{0,1\}^4$ mit Minimalabstand 3 mindestens drei zusätzliche Bits verwenden oder geht es mit weniger?
- Kann man diese Art der Codierung auch auf andere $\{0,1\}^m$ übertragen oder handelt es sich eher um eine zufällige Auflösung eines Puzzles?

Wir werden beide Fragen in den nächsten Vorlesungen beantworten, wollen aber an dieser Stelle schon eine Idee für die Beantwortung der zweiten Frage vorwegnehmen. Die oben beschriebene Codierung ist bei genauerer Betrachtung eine lineare Abbildung von $GF(2)^4$ nach $GF(2)^7$, die man durch die folgende Matrix beschreiben kann:

$$\varphi_{ham}(\vec{v}) = G \cdot \vec{v} \qquad \text{wobei} \qquad G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

Wir betrachten außerdem eine lineare Abbildung von $GF(2)^7$ nach $GF(2)^3$, welche durch die folgende Matrix H repräsentiert ist:

$$H = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Mit etwas Mühe kann man Folgendes nachrechnen:

- Der Code C_{ham} ist der Kern der von H repräsentierten Abbildung, d.h. für alle $\vec{w} \in GF(2)^7$ ist $H \cdot \vec{w}$ genau dann der Nullvektor, wenn $\vec{w}$ ein Codewort aus C_{ham} ist.
- Für alle $\vec{w} \in GF(2)^7$ ist $H \cdot \vec{w} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ genau dann, wenn sich $\vec{w}$ nur an der ersten Stelle von einem Codewort $\vec{c} \in C_{ham}$ unterscheidet.
- Für alle $\vec{w} \in GF(2)^7$ ist $H \cdot \vec{w} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ genau dann, wenn sich $\vec{w}$ nur an der zweiten Stelle von einem Codewort $\vec{c} \in C_{ham}$ unterscheidet.
- Allgemein gibt $H \cdot \vec{w}$ als Binärzahl gelesen (kleinste Bits stehen oben) die Stelle an, an der sich $\vec{w}$ von einem Codewort $\vec{c} \in C_{ham}$ unterscheidet, d.h. die Auswertung von $H \cdot \vec{w}$ ist ausreichend, um die Fehlerkorrektur vornehmen zu können.

Mit der Verallgemeinerung dieses Ansatzes werden wir uns in der übernächsten Vorlesung beschäftigen. Zuvor wird gezeigt, dass der oben präsentierte Hamming–Code hinsichtlich der Anzahl der redundanten Bits nicht mehr verbessert werden kann.

2.3 Allgemeine Schranken für die Informationsrate

Informationsrate

Definition: Die Informationsrate eines Codes $C \subseteq \{0,1\}^n$ ist der Quotient $\frac{\log_2 |C|}{n}$ oder allgemein für $C \subseteq Q^n$ der Quotient $\frac{\log_q |C|}{n}$, wobei $q = |Q|$. Damit beschreibt die Informationsrate das Längenverhältnis der Informationswörter zu den Codewörtern.

Für die besprochenen Beispiele ergeben sich die folgenden Informationsraten:

- Die Informationsrate der Doppelcodierung ist $\frac{1}{2}$ und die Informationsrate der Dreifachcodierung ist $\frac{1}{3}$.
- Die Informationsrate der Doppelcodierung mit Paritätsbit ist $\frac{m}{2m+1} < \frac{1}{2}$.
- Die Informationsrate des Kreuzsicherungscodes für $A = \{0,1\}^4$ ist $\frac{1}{2}$ und der besprochene Hamming–Code hat eine bessere Informationsrate von $\frac{4}{7} > \frac{1}{2}$.
- Die Informationsrate des Kreuzsicherungscodes für $A = \{0,1\}^{k^2}$ ist $\frac{k^2}{k^2+2k} = 1 - \frac{2}{k}$, d.h. mit ausreichend großen Werten von k man kann sich beliebig dicht an die 1 annähern.

Der letzte Punkt muss aber kritisch betrachtet werden. Es ist zwar möglich, mit großem k einen 1-fehlerkorrigierenden Code mit sehr guter Informationrate zu konstruieren, aber dann darf auf k^2 Bits auch nur ein Fehler auftreten. Um also zu einer weniger oberflächlichen Bewertung zu kommen, muss man die Informationsrate mit der Fehlerwahrscheinlichkeit in Zusammenhang bringen. Grundlage dafür ist das Modell eines binären, symmetrischen Kanals, das von Shannon eingefürt wurde.

Satz von Shannon

Unter einem binären, symmetrischen Kanal mit Fehlerwahrscheinlichkeit $p < \frac{1}{2}$ verstehen wir ein Modell eines Nachrichtenübertragungskanals mit den folgenden Eigenschaften:

- Das Kanalalphabet ist $Q = \{0, 1\}$, d.h. es werden Bitfolgen übertragen.
- Die Wahrscheinlichkeit, dass das i–te Bit fehlerhaft übertragen wird, ist geich $p < \frac{1}{2}$. Sie ist unabhängig davon, ob das i–te Bit 0 oder 1 ist (Symmetrie).
- Die Ereignissse, dass das erste, bzw. das zweite, das dritte, ... Bit fehlerhaft übertragen werden, sind unabhängig voneinander.

Insbesondere die dritte Bedingung sorgt dafür, dass dieses Modell nicht sehr realistisch ist und man muss in der Praxis einige Tricks anwenden, um Aussagen nutzen zu können, die auf diesem Modell beruhen.

Definition: Die Kapazität $H(p)$ eines binären, symmetrischen Kanals mit Fehlerwahrscheinlichkeit p ist durch die folgende Formel definiert:

$$H(p) = 1 + p \cdot \log_2 p + (1 - p) \cdot \log_2 (1 - p)$$

Man kann leicht nachrechnen (erste Ableitung nach p), dass sich die Funktion $H(p)$ im Bereich $0 < p \leq \frac{1}{2}$ streng monoton fallend von 1 (Grenzwert für $p \to 0$) nach $0 = H(\frac{1}{2})$ bewegt. Eine geringe Fehlerwahrscheinlichkeit bewirkt also eine hohe Kapazität (nahe 1) und hohe Fehlerwahrscheinlichkeiten (nahe $0,5$) lassen die Kapazität gegen Null gehen. Ohne Beweis wollen wir hier die zentrale Aussage zu diesem Modell zitieren.

Satz (Shannon): Sei ein binärer, symmetrischer Kanal mit Fehlerwahrscheinlichkeit $p < \frac{1}{2}$ und ein $\varepsilon > 0$ gegeben, dann charakterisieren die folgenden Aussagen die Zusammenhänge zwischen der Imformationrate eines Codes C und der Wahrscheinlichkeit $\Pr(C \text{ fails})$, dass bei der Decodierung eines empfangenen und mit zufälligen Fehlern behaftenen Codeworts die Fehlerkorrektur durch Suche nach dem nähesten korrekten Codewort eine falsche Antwort liefert.

1. Für jedes $0 < R < H(p)$ gibt es einen Code C mit Informationsrate $\geq R$, so dass $\Pr(C$
   $mbox fails) < \varepsilon$.
2. Für jedes $R > H(p)$ gibt es eine Konstante $K_R > 0$, so dass für jeden Code C mit Informationsrate $\geq R$ die Wahrscheinlichkeit $\Pr(C \text{ fails})$ größer als K_R ist (also nicht beliebig klein werden kann).

Leider wirft die positive Aussage aus dem ersten Punkt keinen praktischen Nutzen ab:

- Die Aussage ist nicht konstruktiv, d.h. es ist nicht klar, wie man einen solchen Code C konstruieren kann. Im Beweis des Sates von Shannon wird der Code C durch eine Zufallskonstruktion erzeugt.
- Da C nicht systematisch aufgebaut ist, hat man keine effizienten Algorithmen zur Decodierung mit Fehlerkorrektur.

- Um ein kleines ε zu erreichen, muss die Blocklänge von C sehr groß werden, was die Decodierung zusätzlich erschwert.

Aus diesen Gründen bevorzugt man in der Praxis sogenannte lineare Codes, die systematisch aufgebaut und relativ leicht zu decodieren sind. Wir beschäftigen uns zuerst mit Aussagen über allgemeine Schranken solcher Codierungen.

Zusammenhänge zwischen Minimalabstand und Codegröße

Im folgenden sei Q ein q–elementiges Kanalalphabet und $C \subseteq Q^n$ ein Blockcode mit Blocklänge n.

Satz: Ist $C \subseteq Q^n$ ein Code mit Minimalabstand $2k+1$ dann gilt

$$|C| \leq \frac{q^n}{\sum_{i=0}^{k} \binom{n}{i}(q-1)^i}$$

Beweis: Für jedes $\vec{v} \in Q^n$ enthält die Kugel um $\vec{v}$ mit Radius k genau $\sum_{i=0}^{k} \binom{n}{i}(q-1)^i$ Tupel:

- $i = 0, 1, \ldots, k$ sind die möglichen Abstände von $\vec{v}$ zu einem Tupel $\vec{w} \in B_k(\vec{v})$.
- Für jedes i aus diesem Bereich ist $\binom{n}{i}$ die Anzahl der möglichen Auswahlen von i Stellen an denen sich $\vec{v}$ und $\vec{w}$ unterscheiden.
- Für jede ausgewählte Stelle $1j$ gibt es $q-1$ Möglichkeiten v_j durch ein anderes Symbol aus Q zu ersetzen.

Da für einen Code mit Minimalabstand $2k+1$ die k–Kugeln um die Codewörter disjunkt sein müssen, gilt

$$\sum_{\vec{c} \in C} |B_k(\vec{v})| = |C| \cdot \sum_{i=0}^{k} \binom{n}{i}(q-1)^i \leq q^n \qquad \square$$

Definition: Ein Code $C \subseteq Q^n$ mit Minimalabstand $2k+1$ ist *perfekt* ($k-perfekt$), wenn $|C| \cdot \sum_{i=0}^{k} \binom{n}{i}(q-1)^i = q^n$.

Beispiel: Der im letzten Abschnitt beschriebene Hamming–Code ist 1–perfekt. Wir fassen noch einmal zusammen, was wir über diesen Code wissen:

$q = 2 \qquad n = 7 \qquad |C| = 2^4 = 16 \quad \text{und} \quad d(C) = 3.$

Daraus ergibt sich

$$|C| \cdot \sum_{i=0}^{k} \binom{n}{i}(q-1)^i = 2^4 \cdot \left(\binom{7}{0} \cdot 1 + \binom{7}{1} \cdot 1 \right) = 2^4 \cdot (1+7) = 2^7$$

und damit ist der Code C 1–perfekt.

Aus der Ungleichung $|C| \cdot \sum_{i=0}^{k} \binom{n}{i}(q-1)^i \leq q^n$ kann man ableiten, das binäre k–fehlerkorrigierende Codes C der Länge n unter der Voraussetzung $k << n$ in der Größenordnung $k \cdot \log_2 n$ redundante Bits enthalten. Die Voraussetzung ist keine echte Einschränkung, weil sie nur bei sehr kleinen Codes verletzt wird.

Die Anzahl der redundanten Bits ist:

$$n - \log_2 |C| \geq n - \log_2 \left(\frac{2^n}{\sum_{i=0}^{k} \binom{n}{i}(2-1)^i} \right) = n - \left(\log_2 2^n - \log_2 \sum_{i=0}^{k} \binom{n}{i} \right) = \log_2 \sum_{i=0}^{k} \binom{n}{i}$$

Da wir $k << n$ voraussetzen, ist die Summe $\sum_{i=0}^{k} \binom{n}{i}$ durch den letzten Summanden $\binom{n}{k}$ dominiert und folglich

$$\log_2 \sum_{i=0}^{k} \binom{n}{i} \approx \log_2 \binom{n}{k} \leq \log_2 \frac{n^k}{k!} = k \cdot \log_2 n - \log_2 (k!) \leq k \cdot \log_2 n$$

Satz: Für jeden Code $C \subseteq Q^n$ mit $|Q| = q$ gilt für den Minimalabstand

$$d(C) \leq n - (\log_q |C|) + 1$$

Beweis: Sei $d(C) = d$. Wir betrachten die Projektion π von Q^n auf Q^{n-d+1}, die aus jedem n–Tupel die letzen $d-1$ Stellen streicht und die ersten $n-d+1$ Stellen identisch abbildet. Offensichtlich ist die Einschränkung von π auf C eine injektive Abbildung, denn wenn $\vec{c}, \vec{c'}$ verschieden sind, müssen sie sich an mindestens d Stellen unterscheiden und folglich ist auch $\pi(\vec{c}) \neq \pi(\vec{c'})$. Daraus folgt $|C| \leq q^{n-d+1}$ und durch Logarithmieren

$$\log_q |C| \leq n - d + 1 \qquad \Longrightarrow \qquad d \leq n - (\log_q |C|) + 1 \qquad \square$$

Satz: Ist $s \leq n$ und g eine Zahl, die $g \cdot \sum_{i=0}^{k-1} \binom{n}{i}(q-1)^i \leq q^n$ erfüllt, dann gibt es einem Code $C \subseteq Q^n$ mit $d(C) \geq s$ und $|C| = g$.

Beweis: Der Code C wird konstruiert, indem man mit $C = \emptyset$ beginnend in g Schritten jeweils ein neues Tupel aus Q^n in C aufnimmt. Damit der entstehende Code C die Minimalabstandsbedingung $d(C) \geq s$ erfüllt, werden mit jedem in C aufgenommenen Tupel $\vec{v}$ alle Tupel aus $B_{d-1}(\vec{v})$ als blockiert (also nicht mehr wählbar) markiert. Es ist klar, dass nach j Schritten höchstens $j \cdot \sum_{i=0}^{k-1} \binom{n}{i}(q-1)^i$ Tupel blockiert sind. Solange also die Bedingung $j \cdot \sum_{i=0}^{k-1} \binom{n}{i}(q-1)^i < q^n$ erfüllt ist, kann immer noch ein $(j+1)$–tes Tupel in C aufgenommen werden. $\square$.

2.4 Lineare Codes

Hammingraum und lineare Codes

Im gesamten Abschnitt ist $Q = GF(q)$ ein endlicher Körper mit q Elementen, wobei $q = p^i$ eine Primzahl oder Primzahlpotenz ist. Diese Körper enthalten $GF(p)$ als Unterkörper, d.h. die p–fache Summe der 1 ist gleich Null. Man beschreibt diese Eigenschaft des Köpers $GF(q)$ auch durch die Formulierung, dass er die *Charakteristik* p. hat

Definition: Der Vektorrraum $(GF(q))^n$ wird n–dimensionaler Hammingraum über $GF(q)$ genannt und alternativ mit $H(n,q)$ bezeichnet.
Eine Teilmenge $C \subseteq H(n,q)$ heißt *linearer Code*, wenn C ein Untervektorraum von $H(n,q)$ ist.

Beobachtung: Ist $C \subseteq H(n,q)$ ein linearer Code und $\dim C = m$, dann hat C genau q^m Elemente (Codewörter) und folglich die Informationsrate $\frac{\log_q |C|}{n} = \frac{m}{n}$.

Beispiel: Der bereits besprochene Hamming–Code ist ein Unterraum von $H(7,2)$ und hat die Dimension 4.

Minimalgewicht

Definition: Das *Gewicht* eines Vektors $\vec{v} \in H(n,q)$ ist die Anzahl der Komponenten des Vektors, die ungleich 0 sind. Es wird mit $w(\vec{v})$.
bezeichnet. Das *Minimalgewicht* $w(C)$ eines Codes $C \subseteq H(n,q)$ ist das minimale Gewicht eines von $\vec{0}$ verschiedenen Codeworts:

$$w(C) = \min\{\, w(\vec{c}) \mid \vec{c} \in C \ \wedge \ \vec{c} \neq \vec{0}\,\}$$

Satz: Für jeden linearen Code $C \subseteq H(n,q)$ sind Minimalgewicht und Minimalabstand gleich.

Beweis: Es sind zwei Ungleichungen nachzuweisen. Dazu nutzen wir die Identität $w(\vec{v}) = d(\vec{v},\vec{0})$ für beliebige Vektoren $\vec{v} \in H(n,q)$.

- $d(C) \leq w(C)$: Sie $\vec{c} \in C$ ein (von $\vec{0}$ verschiedenes) Codewort mit minimalem Gewicht $w(\vec{c}) = w(C)$. Wegen $\vec{0} \in C$ (C ist ein Unterraum) ist $d(C) \leq d(\vec{0},\vec{c}) = w(\vec{c}) = w(C)$.
- $w(C) \leq d(C)$: Sei $d(C)$ realisiert durch $\vec{c},\vec{c'} \in C$, d.h. $d(C) = d(\vec{c},\vec{c'})$. Dann ist auch der Vektor $\vec{u} = \vec{c} - \vec{c'}$ ein Codewort und er unterscheidet sich genau an den Stellen von $\vec{0}$ an denen sich $\vec{c}$ und $\vec{c'}$ voneinander unterscheiden. Folglich ist $w(C) \leq w(\vec{v}) = d(\vec{c},\vec{c'}) = d(C)$. □

Generatormatrix und Prüfmatrix

Definition: Ein linearer Code $C \subseteq H(n,q)$ der Dimension m wird als ein (n,m)–*Code* bezeichnet. Ist $C \subseteq H(n,q)$ ein (n,m)–Code und $G \in M(n \times m, GF(q))$ eine Matrix, deren Spaltenvektoren eine Basis von C bilden, dann nennt man G eine *Generatormatrix* von C.

Eine Generatormatrix G beschreibt eine zum Code C korrespondierende Codierung $\varphi : (GF(q))^m \longrightarrow (GF(q))^n$, d.h. $C = \operatorname{Im}\varphi$.

Definition: Eine Matrix $H \in M((n-m)\times n, GF(q))$ wird *Prüfmatrix* (oder auch *Checkmatrix*) des (n,m)–Codes C genannt, wenn C der Kern der von H repräsentierten linearen Abbildung $h : (GF(q))^n \longrightarrow (GF(q))^{n-m}$ ist.

Beobachtung: Zur Vereinfachung der Notation schreiben wir $C = \operatorname{Im} G$ an Stelle von $C = \operatorname{Im}\varphi$ und $C = \operatorname{Ker} H$ an Stelle von $C = \operatorname{Ker} h$. Nach Dimensionsformel gilt:

$$n = \dim(\operatorname{Ker} H) + \dim(\operatorname{Im} H) = \dim C + \operatorname{rg} H = m + \operatorname{rg} H$$

Folglich ist $\operatorname{rg} H = n - m$, d.h. die Zeilenvektoren einer Prüfmatrix sind immer linear unabhängig.

Satz: Zwei Matrizen $G \in M(n\times m, GF(q))$ und $H \in M((n-m)\times n, GF(q))$ mit $\operatorname{rg} G = m$ und $\operatorname{rg} H = n-m$ sind genau dann Generator– und Prüfmatrix ein und desselben linearen Codes C, wenn das Matrixprodukt $H \cdot G$ die Nullmatrix ist.

Beweis: Die erste Implikation folgt aus den Definitionen. Für die Gegenrichtung setzt man $C = \operatorname{Im} G$. Aus der Bedingung $H \cdot G = (0)$ folgt $C = \operatorname{Im} G \subseteq \operatorname{Ker} H$ und die Gleichheit folgt aus den oben verwendeten Dimensionsargumenten.

Definition: Eine Generatormatrix G eines (n, m)–Codes C ist in *Standardform*, wenn sie die folgende Gestalt hat

$$G = \begin{pmatrix} E_m \\ A \end{pmatrix} \qquad \text{wobei} \qquad A \in M((n-m) \times m, GF(q)) \quad \text{beliebig}$$

Folgerung: Ist G eine Generatormatrix von C in Standardform, so ist die Matrix $H = (-A\ E_{n-m})$ eine passende Prüfmatrix von C.

Beweis: Die Rangbedingungen $\operatorname{rg} G = m$ und $\operatorname{rg} H = n - m$ aus dem obigen Satz sind offensichtlich erfüllt und $H \cdot G = (0)$ ist leicht nachzurechnen, denn der Eintrag in der i–ten Zeile und j–ten Spalte von $H \cdot G$ hat den Wert $-a_{ij} + a_{ij} = 0$. □

Satz: Sei C ein (n, m)–Code mit Prüfmatrix H, dann ist der Minimalabstand $d(C) \geq d$ genau dann, wenn jede Menge von $d-1$ Spaltenvektoren von H linear unabhängig ist.

Beweis: Eine linear **abhängige** Menge von i Spaltenvektoren von H korrespondiert eineindeutig zu einer nichtrivialen Linearkombination des Nullvektors aus i Spaltenvektoren von H und das korrespondiert eineindeutig zu einen Vektor aus $\operatorname{Ker} H$ mit $\leq i$ Stellen ungleich 0, was äquivalent zu $w(C) = d(C) \leq i$ ist.

Folgerung: Ist in einer Prüfmatrix H keine Spalte ein Vielfaches einer anderen Spalte, so ist jede Menge von zwei Spaltenvektoren linear unabhängig und folglich ist $d(C) \geq 3$, d.h. C ist 1-fehlererkennend.

Unser nächstes Ziel ist die Verallgemeinerung der Idee, die dem bereits besprochenen Hamming–Code zu Grunde liegt. Dieser Code wird auch mit $\mathrm{Ham}_2(3)$ bezeichnet, wobei der Index 2 auf den Körper $GF(2)$ verweist und der Parameter 3 anzeigt, dass der Code 3 Redundanzbits hat, d.h. $n - m = 3$. Man kann sich leicht davon überzeugen, dass die dort verwendete Prüfmatrix H nicht in Standardform ist. Das liegt daran, dass die Auswertung von $H \cdot \vec{v}$ gelesen als Binärzahl das zu korrigierende Bit angeben sollte. Die Prüfmatrix in Standardform hat diese (schöne) Eigenschaft nicht, ist aber dafür leichter zu finden. Wie sich zeigen wird, sollte man bei der Konstruktion eines Codes nicht mit der Generatormatrix beginnen (denn es fehlt die Idee, wie man vorgehen soll), sondern mit einer Prüfmatrix in Standardform.

Beispiel 1: Wir wollen einen 1–perfekten, binären Code mit vier Redundanzbits entwerfen, den Code $\mathrm{Ham}_2(4)$. Zunächst ist dabei noch nicht klar, welche Werte wir für n und m erhalten werden:

1. Wir bilden eine Prüfmatrix in Standardform mit vier Zeilen, d.h. am rechten Ende steht E_4. Um die Informationsrate so groß wie möglich zu halten, soll die Matrix möglichst viele Spalten haben, aber mit der Einschränkung, dass je zwei Spalten linear unabhängig sind. Man beachte, dass es über $GF(2)$ schon ausreichend ist, dass die Spalten ungleich der Nullspalte und paarweise verschieden sind. Da es $2^4 - 1 = 15$ verschiedene Spaltenvektoren gibt, die ungleich der Nullspalte sind, hat unsere Prüfmatrix die Gestalt

   $$(A\, E_4) \quad \text{wobei } A \in M(4 \times 11, GF(2)).$$

   Dabei ist die Reihenfolge der von den vier Standardbasenvektoren und dem Nullvektor verschiedenen Spaltenvektoren in A egal.

2. Wir kennen jetzt $n = 15$ und $m = n-4 = 11$ und können die passende Generatormatrix als $G = \begin{pmatrix} E_{11} \\ -A \end{pmatrix} = \begin{pmatrix} E_{11} \\ A \end{pmatrix}$ zusammensetzen.
   Aus der Konstruktion von H folgt, dass der durch G definierte $(15, 11)$–Code 1–fehlerkorrigierend ist.

3. Der durch G definierte $(15, 11)$–Code C ist auch 1–perfekt, denn

$$|C| \cdot \sum_{i=0}^{1} \binom{15}{i}(q-1)^i = |C| \cdot \left(\binom{15}{0} + \binom{15}{1}\right) = 2^{11} \cdot (1+15) = 2^{11} \cdot 2^4 = 2^{15}.$$

Beispiel 2: Wir wollen einen 1–perfekten Code über $GF(3)$ mit zwei Redundanzbits entwerfen, den Code $\mathrm{Ham}_3(2)$.

1. Wir bilden eine Prüfmatrix H in Standardform mit zwei Zeilen, d.h. am rechten Ende steht E_2. Man will wieder eine maximale Menge von Spaltenenvektoren finden, die paarweise linear unabhängig sind, d.h. es muss vermieden werden, dass zwei ausgewählte Vektoren in einem gemeinsamen 1–dimensionalen Unterraum liegen. Da jeder 1–dimensionale Unterraum von $(GF(3))^2$ aus drei Vektoren besteht (der Nullvektor und zwei andere) kann man von den acht Nicht–Nullvektoren vier als Spalten von H auswählen, z.B.

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} = (A\, E_2) \quad \text{wobei } A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \in M(2 \times 2, GF(3))$$

2. Wir kennen jetzt $n = 4$ und $m = 4-2 = 2$ und können die passende Generatormatrix zusammensetzen (Achtung: In $GF(3)$ gilt $-1 = 2$ und $-2 = 1$.):

$$G = \begin{pmatrix} E_2 \\ -A \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 2 \\ 1 & 2 \end{pmatrix}$$

   Aus der Konstruktion von H folgt, dass der durch G definierte $(4, 2)$–Code 1–fehlerkorrigierend ist.

3. Der durch G definierte Code C ist auch 1–perfekt, denn

$$|C| \cdot \sum_{i=0}^{1} \binom{4}{i}(q-1)^i = |C| \cdot \left(\binom{4}{0} \cdot 1 + \binom{4}{1} \cdot 2\right) = 3^2 \cdot (1+8) = 3^4.$$

Fehlerkorrektur durch Syndrom–Decodierung

Wie wir bereits wissen, kann man für einen Code mit $d(C) \geq 2k+1$ beim Auftreten von $\leq k$ Fehlern die Fehlerkorrektur durch Suche nach dem nähesten Codewort realisieren. Aus algorithmischer Sicht ist das einfach, aber nicht sehr effizient, denn man müsste für ein zu korrigierendes Tupel $\vec{v} \in H(n, q)$ alle Tupel aus der Kugel $B_k(\vec{v})$ untersuchen (mit der Prüfmatrix multipliziern), d.h. im schlechtesten Fall wären $\sum_{i=0}^{k} \binom{n}{i}(q-1)^i$ Anwendungen der Prüfmatrix notwendig.

Ein großer Vorteil von linearen Codes liegt darin, dass es wesentlich effizientere Verfahren zur Fehlerkorrektur gibt, nämlich mit nur einer Anwendung der Prüfmatrix (nach einer Vorverarbeitungsphase).

Satz: Ist C ein (n, m)–Code mit Prüfmatrix H und Minimalabstand $d(C) \geq 2k + 1$, dann wird die Kugel $B_k(\vec{0})$ von der zu H korrespondierenden Abbildung injektiv abgebildet.

Beweis: Wir führen einen indirekten Beweis durch Widerspruch. Angenommen es gibt zwei verschiedene Vektoren $\vec{u}, \vec{v} \in B_k(\vec{0})$, so dass $H \cdot \vec{u} = H \cdot \vec{v}$, dann ist $H \cdot (\vec{u} - \vec{v}) = \vec{0}$. Folglich liegt $\vec{u} - \vec{v}$ in $\operatorname{Ker} H = C$. Da $\vec{u}$ und $\vec{v}$ voneinander verschieden sind, ist $\vec{u} - \vec{v} \neq \vec{0}$. Wegen $\vec{0} \in C$ und $d(C) \geq 2k + 1$ ergibt sich (durch Nutzung der Dreiecksungleichung) der folgende Widerspruch:

$$2k + 1 \leq d(\vec{u} - \vec{v}, \vec{0}) = d(\vec{u}, \vec{v}) \leq d(\vec{u}, \vec{0}) + d(\vec{0}, \vec{v}) \leq k + k = 2k \qquad \square$$

Die Fehlerkorrektur eines linearen Codes mit Prüfmatrix H beruht auf der folgenden Beobachtung: Ist $\vec{c} \in C$ ein Codewort und $\vec{w} \in B_k(\vec{c})$ ein Tupel, das sich an mindestens einer und höchstens k Stellen von $\vec{w}$ unterscheidet, dann ist $\vec{u} := \vec{w} - \vec{c} \in B_k(\vec{0})$ und es gilt $H \cdot \vec{u} = H \cdot \vec{w}$ und $\vec{c} = \vec{v} - \vec{u}$. Wenn man die Fehler in $\vec{w}$ korrigieren will, ist $\vec{c}$ natürlich unbekannt, denn $\vec{c}$ ist ja gerade das gesuchte näheste Codewort. Man betrachtet $\vec{c}$ als eine Variable und $\vec{u}$ als eine davon abhängige Größe. Da man aber $H \cdot \vec{u} = H \cdot \vec{w}$ kennt und H auf der Kugel $B_k(\vec{0})$ injektiv abbildet, kann $\vec{u}$ eindeutig aus seinem Bild identifiziert werden und letztlich durch $\vec{c} = \vec{w} - \vec{u}$ die Fehlerkorrektur vorgenommen werden.

Algorithmus:

- **Vorverarbeitung:** Gegeben sei ein (n, m)–Code mit Prüfmatrix H und Minimalabstand $d(C) \geq 2k + 1$.
  Für alle $\vec{u} \in B_k(\vec{0})$ wird $H \cdot \vec{u}$ berechnet und in einer Tabelle gespeichert. Man nennt den Vektor $H \cdot \vec{u}$ das Syndrom für den Fehlervektor $\vec{u}$.

- **Fehlerkorrektur:** Gegeben sei ein Vektor $\vec{w}$ von dem bekant ist, dass der in einer Kugel von Radius k um ein zu bestimmendes Codewort $\vec{c}$ liegt.
  Man berechnet $H \cdot \vec{w}$ und sucht das Ergebnis in der Syndromtabelle. Sei $\vec{u}$ der Fehlervektor, für den $H \cdot \vec{u} = H \cdot \vec{w}$ gilt, dann ist $\vec{c} := \vec{w} - \vec{u}$ das gesuchte Codewort.

Kapitel 3

Stochastik

3.1 Wahrscheinlichkeitsräume

Zufällige Prozesse und Wahrscheinlichkeitsräume

Wahrscheinlichkeitsräume dienen zur Beschreibung von idealisierten Modellen für die Ergebnisse eines zufälligen Prozesses. Wie die folgenden Beispiele zeigen, geht es dabei häufig geht um reale Prozesse, aber auch gedankliche Experimente können zum Entwurf eines Wahrscheinlichkeitsraums führen.

1. Würfelexperiment;
2. Münzwurfexperiment;
3. Folge von 5 Münzwürfen oder allgemein von n Münzwürfen, wobei n eine fest gewählte natürliche Zahl ist;
4. Folge von Münzwürfen, die dann beendet wird, wenn zum ersten Mal die Zahl fällt;
5. Folge von gewürfelten Zahlen, die dann beendet wird, wenn zum ersten Mal eine 6 fällt;
6. Zufällig gewählte natürliche Zahl, wobei alle Zahlen gleich wahrscheinlich sein sollen (Gleichverteilung);
7. Zufällig gewählte reelle Zahl aus dem Intervall $[0, 1]$, wobei alle Zahlen gleich wahrscheinlich sein sollen (Gleichverteilung);

Bei der genaueren Betrachtung fallen die folgenden Gemeinsamkeiten bzw. Unterschiede auf:

- Die Ergebnismengen in den Beispielen (1) bis (3) sind endlich, alle anderen Ergebnismengen sind unendlich.
- In den Beispielen (4) und (5) ist die Ergebnismenge abzählbar unendlich und man kann man jedem Ergebnis eine positive Wahrscheinlichkeit zuordnen, so dass sich die Gesamtwahrscheinlichkeit 1 ergibt.

- Auch in Beispiel (6) ist die Ergebnismenge abzählbar unendlich, aber man kann den einzelnen Ergebnissen keine Wahrscheinlichkeit zuordnen (ein positiver Wert würde zur Gesamtwahrscheinlichkeit ∞ und der Wert 0 würde zur Gesamtwahrscheinlichkeit 0 führen).

- In Beispiel (7) ist die Ergebnismenge überabzählbar. Wie wir später sehen werden, kann man hier den einzelnen Ergebnissen die Wahrscheinlichkeit 0 zuordnen, denn die Gesamtmenge ist keine abzählbare Vereinigung der Einzelergebnisse.

Definition: Ein diskreter Wahrscheinlichkeitsraum $(\Omega, \Pr)$ besteht aus einer abzählbaren Menge Ω von elementaren Ereignissen (Ergebnissen) und einer Verteilungsfunktion

$$\Pr : \Omega \to [0,1] \qquad \text{mit der Eigenschaft} \qquad \sum_{a\in\Omega} \Pr(a) = 1$$

Jede Teilmenge $A \subseteq \Omega$ wird ein Ereignis genannt. Die Verteilungsfunktion kann zu einem Wahrscheinlichkeitsmaß über der Menge $\mathcal{P}(\Omega)$ aller Ereignisse erweitert werden:

$$\Pr : \mathcal{P}(\Omega) \to [0,1] \quad \text{mit} \quad \Pr(A) = \sum_{a\in A} \Pr(a)$$

Die Verwendung der gleichen Bezeichnung für die Verteilung und das Wahrscheinlichkeitsmaß ist eine kleine technische Unsauberkeit, die aber in der Literatur weit verbreitet ist. Die Potenzmenge wird häufig auch mit dem Symbol 2^Ω bezeichnet.

Bemerkung: Diese Menge Ω kann endlich oder abzählbar unendlich sein. Im letzten Fall treten auch unendliche Summen auf, aber die Konvergenz ist in jedem Fall gesichert, da die Gesamtsumme über alle Ergebnisse aus Ω gleich 1 ist.

Wahrscheinlichkeitsräume für die Beispiele (1) bis (5)

1) Würfel: $\Omega = \{1,2,3,4,5,6\}$ und $\Pr(1) = \Pr(2) = \ldots = \Pr(6) = \frac{1}{6}$.
Das Ereignis A, eine gerade Zahl zu würfeln, setzt sich aus den Ergebnissen $2, 4$ und 6 zusammen und folglich ist $\Pr(A) = \Pr(\{2,4,6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$.

2) Münzwurf: $\Omega = \{0,1\}$, wobei 0 für Kopf und 1 für Zahl steht und $\Pr(0) = \Pr(1) = \frac{1}{2}$ gesetzt wird.

3) Bei einer Folge von 5 Münzwürfen setzt man $\Omega = \{0,1\}^5$ und $\Pr(a) = \frac{1}{2^5}$ für jedes 5-Tupel $a \in \Omega$.
Bezeichnet A das Ereignis, dass bei 5 Würfen genau dreimal Kopf fällt, so ist $\Pr(A) = \binom{5}{3}\frac{1}{2^5}$, da es genau $\binom{5}{3}$ elementare Ereignisse dieser Art gibt.

4) Die Ergebnisse bei einem Münzwurfexperiment, das so lange widerholt wird, bis zum ersten Mal die Zahl fällt, lassen sich durch einen String von Nullen (Kopf) mit einer abschließenden Eins (Zahl) darstellen: $\Omega = \{1, 01, 001, 0001, \ldots\}$.
Offensichtlich ist $\Pr(1) = \frac{1}{2}, \Pr(01) = \frac{1}{4}, \Pr(001) = \frac{1}{8}, \ldots$.

5) Die Ergebnisse beim Würfeln bis zum ersten Mal eine 6 fällt lassen sich durch einen String von Zahlen aus $\{1,2,3,4,5\}$ mit einer abschließenden 6 darstellen:
$\Omega = \{6, 16, 26, 36, 46, 56, 116, \ldots\}$.
Offensichtlich ist $\Pr(6) = \frac{1}{6}, \Pr(16) = \frac{1}{6\cdot 6}$ und $\Pr(116) = \frac{1}{6\cdot 6\cdot 6}$.
Allgemein ist die Wahrscheinlichkeit für das Ereignis A, dass ein Spiel über genau k Runden geht, durch $\Pr(A) = \frac{5^{k-1}}{6^k}$ beschrieben.

Verallgemeinerte Definition eines Wahrscheinlichkeitsraums

Wie bereits ausgeführt, lassen sich für die Beispiele (6) und (7) keine derartigen diskreten Wahrscheinlichkeitsräume konstruieren. Insbesondere kann man in diesen Fällen kein Wahrscheinlichkeitsmaß definieren, das jeder Teilmenge A von Ω ihre Wahrscheinlichkeit zuordnet. Der Ausweg besteht darin, nur eine Auswahl von Teilmengen als messbar zu deklarieren, so dass bis auf diese Einschränkung die wichtigsten Eigenschaften des bisherigen Ansatzes gerettet werden. Dazu zählt die Eigenschaft, dass sich die Wahrscheinlichkeiten eines Ereignisses und des Komplementärereignisses zu 1 ergänzen, und die Additivität bei disjunkten Vereinigungen:

$$\begin{array}{ll} \Pr\big(\overline{A}\big) = 1 - \Pr(A) & \text{für alle Ereignisse } A \text{ und } \overline{A} = \Omega \setminus A \\ \Pr(A \cup B) = \Pr(A) + \Pr(B) & \text{für alle disjunkten Ereignisse } A \text{ und } B \end{array}$$

Um das zu gewährleisten, muss die Menge der messbaren Ereignisse bezüglich Komplementbildung und disjunkter Vereinigung abgeschlossen sein. Im Folgenden wird sogar noch etwas mehr gefordert, nämlich die Abgeschlossenheit bezüglich beliebiger abzählbaren Vereinigungen.

Definition: Eine Mengenfamilie $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ ist eine σ-Algebra über Ω, wenn die folgenden zwei Bedingungen erfüllt sind:

$$\begin{aligned} A \in \mathcal{F} \;&\Rightarrow\; \overline{A} \in \mathcal{F} \\ A_1, A_2, \ldots \in \mathcal{F} \;&\Rightarrow\; \bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \end{aligned}$$

Die Elemente der Mengenfamilie $\mathcal{F}$ werden Ereignisse genannt. Der Begriff Ereignis steht damit für messbare Teilmengen von Ω.

Anmerkung 1: Für jede nichtleere σ-Algebra $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ ist $\Omega \in \mathcal{F}$ und $\emptyset \in \mathcal{F}$.
Da mit einem $A \in \mathcal{F}$ auch das Komplement $\overline{A}$ in $\mathcal{F}$ liegt, müssen auch $\Omega = A \cup \overline{A}$ und $\emptyset = \overline{\Omega}$ in $\mathcal{F}$ liegen.

Anmerkung 2: Die zweite Bedingung der Definition bezieht sich auf abzählbar unendliche Vereinigungen. Das schließt alle endlichen Vereigungen ein, denn man kann jede endliche Vereinigung durch Hinzunahme von abzählbar unendlich vielen Exemplaren von $\emptyset$ in eine abzählbar unendliche Vereinigung umformen.

Anmerkung 3: Schwächt man die zweite Bedingung auf endliche Vereinigungen ab, wird $\mathcal{F}$ eine *Algebra* genannt.

Definition: Ein *Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \Pr)$ besteht aus einer Ergebnismenge Ω, einer σ-Algebra $\mathcal{F}$ über Ω und einem Wahrscheinlichkeitsmaß $\Pr : \mathcal{F} \to [0,1]$ mit:

1. für alle $A \in \mathcal{F}$ ist $\Pr\big(\overline{A}\big) = 1 - \Pr(A)$
2. für jede Folge $A_1, A_2, \ldots$ von paarweise disjunkten Ereignissen $A_i \in \mathcal{F}$ ist

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$$

Eigenschaften von Wahrscheinlichkeitsräumen

Wir beginnen mit einigen unmittelbaren Schlussfolgerungen aus der Definition des Wahrscheinlichkeitsraums :

- $\mathcal{F}$ ist abgeschlossen gegen endliche und abzählbare Durchschnitte, denn

$$\begin{aligned} A, B \in \mathcal{F} &\Rightarrow A \cap B = \overline{\overline{A} \cup \overline{B}} \in \mathcal{F} \\ A_1, A_2, \ldots \in \mathcal{F} &\Rightarrow \bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \overline{A_i}} \in \mathcal{F} \end{aligned}$$

- $\mathcal{F}$ ist abgeschlossen gegen Mengendifferenzen, denn $A \backslash B = A \cap \bar{B}$
- Pr ist monoton, d.h. $A \subseteq B \Rightarrow \Pr(A) \leq \Pr(B)$.

  Aus $A \subseteq B$ kann man $B = A \cup (B \backslash A)$ ableiten. Da das eine disjunkte Vereinigung ist, folgt daraus:

$$\Pr(B) = \Pr(A) + \Pr(B \backslash A) \geq \Pr(A) + 0 = \Pr(A).$$

Satz: Ist $A_1 \subseteq A_2 \subseteq \ldots$ eine aufsteigende Folge von Ereignissen und ist A die Vereinigung dieser Ereignisse dann ist

$$\Pr(A) = \Pr(\textstyle\bigcup_{i=1}^{\infty} A_i) = \lim_{n \to \infty} \Pr(A_n).$$

Beweis: Die Folge $\Pr(A_1), \Pr(A_2), \ldots$ ist monoton wachsend und beschränkt. Damit ist sie auch konvergent. Aus $A_1 \subseteq A_2 \subseteq \ldots$ folgt

$$A = A_1 \cup \underbrace{(A_2 \backslash A_1)}_{B_2} \cup \underbrace{(A_3 \backslash A_2)}_{B_3} \cup \ldots$$

Diese Vereinigung ist disjunkt und damit ergibt sich:

$$\begin{aligned} \Pr(A) &= \Pr(A_1) + \sum_{i=2}^{\infty} \Pr(B_i) \\ &= \Pr(A_1) + \lim_{n \to \infty} \sum_{i=2}^{n} \Pr(B_i) \\ &= \lim_{n \to \infty} (\Pr(A_1) + (\Pr(A_2) - \Pr(A_1)) + (\Pr(A_3) - \Pr(A_2)) + \ldots \\ &\quad \ldots + (\Pr(A_n) - \Pr(A_{n-1}))) \\ &= \lim_{n \to \infty} \Pr(A_n) \end{aligned}$$

Folgerung: Analog gilt für jede absteigende Folge von Ereignissen $C_1 \supseteq C_2 \supseteq \ldots$:

$$\Pr\left(\bigcap_{i=1}^{\infty} C_i\right) = \lim_{n \to \infty} \Pr(C_n)$$

Gleichverteilung über einem reellen Intervall

Die folgende Konstruktion für die Gleichverteilung der reellen Zahlen im Intervall $[0, 1]$ kann man leicht auf beliebige reelle Intervalle und mit etwas mehr Aufwand auch auf kartesische Produkte von Intervallen in höher dimensionalen Räumen übertragen. Die aus dieser Konstruktion resultierenden Elemente σ-Algebra nennt man Borelsche Mengen.

- Für alle $0 \leq a < b \leq 1$ ist das Intervall $(a, b]$ ein Ereignis mit der Wahrscheinlichkeit $\Pr((a, b]) = b - a$. Offensichtlich beschreibt diese Differenz die Größe des Intervalls $(a, b]$.
  Wird an Stelle von $[0, 1]$ ein anderes Basisintervall $[s, t]$ zu Grunde gelegt, muss das Wahrscheinlichkeitsmaß mittels Division durch $t-s$ normiert werden, d.h. $\Pr((a, b]) = \frac{b-a}{t-s}$.

- Aus der σ-Algebra-Eigenschaft folgt, dass dann auch alle offenen und abgeschlossenen Intervalle in $\mathcal{F}$ sein müssen.

  - abgeschlossene Intervalle: $[a, b] = \bigcap_{i=1}^{\infty} \left(a - \frac{a}{i}, b\right]$
  - offene Intervalle: $(a, b) = (a, 1] \backslash [b, 1]$
  - insbesondere gilt für jede reelle Zahl $x \in [0, 1]$: $\{x\} \in \mathcal{F}$

Definiert man $\mathcal{F}$ als Menge aller abzählbaren disjunkten Vereinigungen von Intervallen (egal ob offen, abgeschlossen, halboffen oder Punkt), kann man die Eigenschaften einer σ-Algebra leicht nachweisen.
Die Wahrscheinlichkeit eines solchen Ereignisses $A = \bigcup_{i=1}^{\infty} \langle a_i, b_i \rangle \in \mathcal{F}$ (wobei $b_i \geq a_i$ und $\langle \in \{[, (\}$ und $\rangle \in \{),]\}$ und $b_i \leq a_{i+1}$) muss nach den oben abgeleiteten Regeln den folgenden Wert annnehmen:

$$\Pr\left(\bigcup_{i=1}^{\infty} \langle a_i, b_i \rangle\right) = \sum_{i=1}^{\infty} (b_i - a_i)$$

Ananlog verwendet man für die Gleichverteilung zufälliger Punkte aus dem Einheitsquadrat $[0, 1] \times [0, 1]$ zuerst achsenparalle Rechtecke $(a, b] \times (c, d]$, deren Wahrscheinlichkeit auf $(b - a)(d - c)$ gesetzt wird, und baut daraus die kleinstmögliche σ-Algebra $\mathcal{F}$ auf.

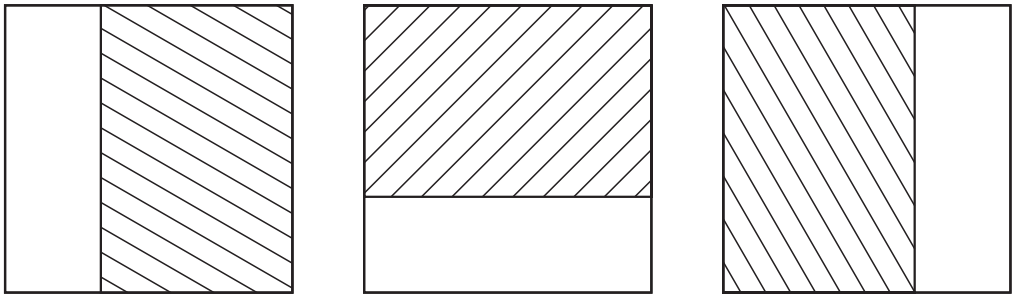
3.2 Bedingte Wahrscheinlichkeit und Unabhängigkeit

Definition: Sei $(\Omega, \mathcal{F}, \Pr)$ ein Wahrscheinlichkeitsraum. Für zwei Ereignisse $A, B \in \mathcal{F}$ mit $\Pr(B) > 0$ definiert man die *bedingte Wahrscheinlichkeit* von Ereignis A unter B durch

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Beispiel: Sei $\Omega = [0, 1] \times [0, 1]$ die Menge der Punkte im Einheitsquadrat mit Gleichverteilung und

- $A = \{(x, y) \mid x \geq \frac{1}{3}\}$ $\quad \Pr(A) = \frac{2}{3}$
- $B = \{(x, y) \mid y \geq \frac{1}{3}\}$ $\quad \Pr(B) = \frac{2}{3}$
- $C = \{(x, y) \mid x \leq \frac{2}{3}\}$ $\quad \Pr(C) = \frac{2}{3}$



Wie man leicht sieht ist $A \cap B = [\frac{1}{3}, 1] \times [\frac{1}{3}, 1]$ und $A \cap C = [\frac{1}{3}, \frac{2}{3}] \times [0, 1]$. Daraus folgt

$$\begin{aligned} \Pr(A \,|\, B) &= \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\frac{4}{9}}{\frac{2}{3}} = \frac{2}{3} \\ \Pr(A \,|\, C) &= \frac{\Pr(A \cap C)}{\Pr(C)} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2} \end{aligned}$$

Definition: Zwei Ereignisse A und B nennt man *unabhängig*, wenn

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B).$$

Im obigen Beispiel sind die Ereignisse A und B unabhängig, die Ereignisse A und C sind aber nicht unabhängig. Es gibt einen engen Zusammenhang zwischen den Begriffen der bedingten Wahrscheinlichkeit und der Unabhängigkeit:
Sind A und B unabhängige Ereignisse und $p(B) > 0$, dann hat die bedingte Wahrscheinlichkeit von A unter B den gleichen Wert wie die Wahrscheinlichkeit von A, also

$$\Pr(A \,|\, B) = \Pr(A)$$

Definition: Eine Familie $\{A_i \mid i \in I\}$ von Ereignissen ist unabhängig, wenn für jede endliche Teilmenge $J \subseteq I$

$$\Pr\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \Pr(A_i)$$

Achtung: Es gibt Familien von paarweise unabhängigen Ereignissen, die nicht unabhängig sind.

Beispiel: Sei $\Omega = \{a, b, c, d\}$ eine Menge der Elementarereignisse mit den Wahrscheinlichkeiten $\Pr(a) = \Pr(b) = \Pr(c) = \Pr(d) = \frac{1}{4}$. Wir betrachten die Ereignisse:

- $A = \{a, d\}$ $\quad \Pr(A) = \frac{1}{2}$
- $B = \{b, d\}$ $\quad \Pr(B) = \frac{1}{2}$
- $C = \{c, d\}$ $\quad \Pr(C) = \frac{1}{2}$

Offensichtlich sind diese Ereignisse paarweise unabhängig, denn

$$\begin{array}{rcl} \Pr(A \cap B) & = & \frac{1}{4} = \Pr(A) \cdot \Pr(B) \\ \Pr(A \cap C) & = & \frac{1}{4} = \Pr(A) \cdot \Pr(C) \\ \Pr(B \cap C) & = & \frac{1}{4} = \Pr(B) \cdot \Pr(C) \end{array}$$

Die Familie mit den Ereignissen $\{A, B, C\}$ ist aber **nicht** unabhängig, denn:

$$\Pr(A \cap B \cap C) = \frac{1}{4} \quad \neq \quad \frac{1}{8} = \Pr(A) \cdot \Pr(B) \cdot \Pr(C)$$

Partitionstheorem

Zum Abschluss des Themas stellen wir einen wichtigen Satz für die Verwendung von bedingten Wahrscheinlichkeiten vor und demonstrieren die Anwendung an einem Beispiel.

Satz: Sei $\{B_1, B_2, \ldots\}$ eine abzählbare Partition von Ω in Ereignisse $B_i \in \mathcal{F}$ mit $\Pr(B_i) > 0$. Dann ist

$$\Pr(A) = \sum_i \Pr(A \,|\, B_i) \cdot \Pr(B_i) \quad \text{für alle } A \in \mathcal{F}$$

Beweis:

$$\begin{array}{rcl} \Pr(A) & = & \Pr(A \cap \Omega) = \Pr\left(A \cap \left(\bigcup_i B_i\right)\right) \\ & = & \Pr\left(\bigcup_i (A \cap B_i)\right) \quad \text{(disjkunte Vereinigung)} \\ & = & \sum_i \Pr(A \cap B_i) = \sum_i \frac{\Pr(A \cap B_i)}{\Pr(B_i)} \cdot \Pr(B_i) \\ & = & \sum_i \Pr(A \,|\, B_i) \cdot \Pr(B_i) \end{array}$$

Beispiel: Wir betrachten die folgenden zufälligen Ereignisse, die eine Partition von Ω darstellen sollen: Morgen früh regnet es (R) oder schneit (S) oder es ist trocken (T). Für das zufällige Ereignis B einer Busverspätung seien die folgenden bedigten Wahrscheinlichkeiten bekannt:

- Bei Regen ist die Wahrscheinlichkeit für eine Busverspätung $\frac{1}{3}$.
- Bei Schnee ist die Wahrscheinlichkeit für eine Busverspätung $\frac{2}{3}$.
- Bei Trockenheit ist die Wahrscheinlichkeit für eine Busverspätung $\frac{1}{6}$.

Wenn die Wettervorhersage die Wahrscheinlichkeiten $\Pr(R) = \frac{1}{5}$ und $\Pr(S) = \Pr(T) = \frac{2}{5}$ ankündigt, kan man die Wahrscheinlichkeit einer Busverspätung mit dem Partitionstheoren bestimmen:

$$\Pr(B) = \frac{1}{3} \cdot \frac{1}{5} + \frac{2}{3} \cdot \frac{2}{5} + \frac{1}{6} \cdot \frac{2}{5} = \frac{2+8+2}{30} = \frac{12}{30} = \frac{2}{5}$$

3.3 Zufallsvariable

Diskrete Zufallsvariablen

Für einige Zufallsprozesse ist das Ergebnis bereits eine Zahl (z.B. Würfel) oder es liegt nahe, die Ergebnisse als Zahl zu interpretieren (z.B. Münzwurf). Aber auch bei Prozessen deren Ergebnisse keine Zahlen sind, kann man den konkreten Ergebnissen oft Zahlenwerte zuordnen. Zum Beispiel kann man einer Folge von n Münzwürfen die Anzahl der geworfenen Köpfe oder einem zufälligen Punkt in einer Kreisscheibe den Abstand zum Mittelpunkt zuordnen. Solche Funktionen von Ω nach $\mathbb{R}$ nennt man Zufallsvariable oder Zufallsgrößen. Da man für diskrete Wahrscheinlichkeitsräume jede Funktion verewenden kann, aber für allgemeine Wahrscheinlichkeitsräume einige Einschränkungen machen muss, werden wir für die verschiedenen Situationen mehrere Definitionen von Zufallsvariablen einführen.

Definition 1: Ist $(\Omega, \Pr)$ ein diskreter Wahrscheinlichkeitsraum , so wird jede Funktion $X : \Omega \to \mathbb{R}$ ist eine *diskrete Zufallsvariable* auf $(\Omega, \Pr)$ genannt.

Definition 2: Sei $(\Omega, \mathcal{F}, \Pr)$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \to \mathbb{R}$ ist eine *diskrete Zufallsvariable* auf $(\Omega, \mathcal{F}, \Pr)$, falls

- das Bild $\mathrm{Im}(X) = \{x \in \mathbb{R} \mid \exists\, a \in \Omega \quad x = X(a)\}$ abzählbar ist und
- für alle $x \in \mathbb{R}$ das Urbild $X^{-1}(x) = \{a \mid X(a) = x\}$ in der σ-Algebra $\mathcal{F}$ liegt.

Als Konsequenz aus dieser Definition liegt auch für alle Teilmengen $T \subseteq \mathbb{R}$ das Urbild $X^{-1}(T) \in \mathcal{F}$ in $\mathcal{F}$, denn

$$X^{-1}(T) = \bigcup_{x \in (T \cap \mathrm{Im}\ X)} X^{-1}(x)$$

Wie man leicht sieht ist die erste Definition ein Spezialfall der zweiten, denn für die Verallgemeinerung muss nur noch $\mathcal{F} = \mathcal{P}(\Omega)$ gesetzt werden.

Definition: Eine diskrete Zufallsvariable X auf $(\Omega, \mathcal{F}, p)$ induziert eine *diskrete Verteilungsfunktion* (oder *Gewichtsfunktion*) $\Pr_X : \mathbb{R} \to [0, 1]$, die wie folgt definiert ist:

$$\Pr_X(x) = \Pr(X^{-1}(x))$$

Oft verwendet man für $\Pr_X(x)$ auch die intuitiv besser erfassbare Schreibweise $\Pr(X = x)$, die man als "Wahrscheinlichkeit, dass die Variable X den Wert x annimmt" lesen kann.

Da die Urbilder $X^{-1}(x)$ über alle $x \in \mathrm{Im}(X)$ eine abzählbare Partition von Ω bilden, erhält man:

$$\begin{aligned} \sum_{x \in \mathrm{Im}(X)} \Pr_X(x) &= \sum_{x \in \mathrm{Im}(X)} \Pr(\{a \mid X(a) = x\}) = \Pr\left(\bigcup_{x \in \mathrm{Im}(X)} \{\omega \mid X(a) = x\} \right) \\ &= \Pr(\Omega) = 1 \end{aligned}$$

Diese Eigenschaft zeigt, dass eine diskrete Verteilungsfunktion $\Pr_X$ selbst ein Wahrscheinlichkeitsmaß ist, nämlich auf der Menge $\Omega' = \mathrm{Im}(X)$ und der σ-Algebra $\mathcal{P}(\Omega')$. Mit anderen Worten kann man zu jeder diskreten Verteilungsfunktion einen Wahrscheinlichkeitsraum

und eine Zufallsvariable mit der vorgegebenen Verteilungsfunktion konstruieren:
Ist eine diskrete Verteilungsfunktion durch eine abzählbare Teilmenge $S \subseteq \mathbb{R}$ Werte $\pi_s \in [\,0,1\,]$ für jedes $s \in S$ und mit $\sum_{s\in S} \pi_s = 1$ gegeben, setzt man

- $\Omega = S$
- $\mathcal{F} = \mathcal{P}(S)$
- $\Pr(A) = \sum\limits_{s\in A} \pi_s$ für jedes $A \subseteq S$
- $X : \Omega \to \mathbb{R}$ mit $X(s) = s$ für alle $s \in S$

Beispiele: In den folgenden Standardverteilungen ist p immer eine Zahl aus dem Bereich $(\,0,1\,)$ und $q = 1 - p$

1. *Zufallsvariable mit Bernoulli-Verteilung* (Parameter p):

   - $\mathrm{Im}(X) = \{0, 1\}$
   - $\Pr_X(0) = q$ und $\Pr_X(1) = p$

   Konkrete Beispiele findet man beim Münzwurf ($p = 0.5$) oder beim Ziehen einer Karte aus einem Kartenspiel, wobei die Variable dem Wert 1 annehmen soll, wenn eine Herzkarte gezogen wurde und 0 sonst ($p = 0.25$).

2. *Zufallsvariable mit Binominialverteilung* (Parameter n und p)

   - $\mathrm{Im}(X) = \{0, 1, \ldots, n\}$ und
   - $\Pr_X(k) = \binom{n}{k}\, p^k\, q^{n-k}$ für alle $k \in \{0, 1, \ldots, n\}$

   Binominialverteilungen treten bei n-facher Wiederholung eines Bernoulli-Experiments auf, wenn man davon ausgehen kann, das die einzelnen Experimente unabhängig voneinander sind.
   So ist die Wahrscheinlichkeit, dass bei n Münzwürfen genau k-mal Kopf fällt gleich $\binom{n}{k}(\frac{1}{2})^n$ und die Wahrscheinlichkeit, dass beim n-fachen Würfeln genau k Sechsen fallen gleich $\binom{n}{k}(\frac{1}{6})^k(\frac{5}{6})^{n-k}$.

3. *Zufallsvariable mit geometrischer Verteilung* (Parameter p):

   - $\mathrm{Im}(X) = \mathbb{N}^+ = \{1, 2, 3 \ldots\}$ und
   - $\Pr_X(k) = p\, q^{k-1}$ für alle $k \in \mathbb{N}^+$

   Geometrischen Verteilungen begegnet man immer dann, wenn ein Bernoulli-Experiment so lange wiederholt wird bis eine 1 auftritt und die Zufallsvariable die Anzahl der Versuche zählt. Würfelt man beispielsweise so lange bis eine Sechs fällt, ist die Wahrscheinlichkeit dafür, dass man genau n Versuche benötigt gleich $\frac{1}{6}(\frac{5}{6})^{n-1}$.

4. *Zufallsvariable mit Poisson-Verteilung* (Parameter λ):

   - $\mathrm{Im}(X) = \mathbb{N}$ (einschließlich der Null)
   - $\Pr_X(k) = \frac{1}{k!}\, \lambda^k\, e^{-\lambda}$

Es ist nicht offensichtlich, dass die Summe gleich 1 ist, aber man kann das wie folgt überprüfen:

$$\sum_{k=0}^{\infty} \Pr_X(k) = \sum_{k=0}^{\infty} \tfrac{1}{k!}\, \lambda^k\, e^{-\lambda} = e^{-\lambda} \cdot \textstyle\sum_{k=0}^{\infty} \tfrac{1}{k!}\, \lambda^k = e^{-\lambda} \cdot e^{\lambda} = 1$$

Die Poisson-Verteilung kann man in einigen Fällen als gute Näherung der Binomialverteilung verwenden, insbesondere dann, wenn n sehr groß, p sehr klein, und k wesentlich kleiner als n ist. In diesem Fall setzt man $\lambda = n \cdot p$ und erhält die Näherung:

$$\begin{aligned} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k (1-p)^{n-k} \\ &\approx \frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1-\frac{\lambda}{n}\right)^n \cdot \left(1-\frac{\lambda}{n}\right)^{-k} \\ &\approx \frac{\lambda^k}{k!} \cdot e^{-\lambda} \end{aligned}$$

Stetige Zufallsvariable

Definition 3: Algemein ist eine *Zufallsvariable* auf $(\Omega, \mathcal{F}, \Pr)$ eine Abbildung $X : \Omega \to \mathbb{R}$, so dass für alle $x \in \mathbb{R}$ gilt:

$$\{a \in \Omega \mid X(a) \leq x\} \in \mathcal{F}$$

Auch hier kann man sich leicht davon überzeugen, dass diese Definition eine Verallgemeinerung von Definition 2 ist: Bezeichnet X eine diskrete Zufallsvariable (nach Definition 2) so ist

$$\{a \in \Omega \mid X(a) \leq x\} = \bigcup_{t \leq x,\ t \in \mathrm{Im}(X)} \{a \in \Omega \mid X(a) = t\} \in \mathcal{F}.$$

Ein ähnlicher Zusammenhang besteht zwischen der nachfolgend definierten Verteilung einer Zufallsvariable und der vorher definierten diskreten Verteilung.

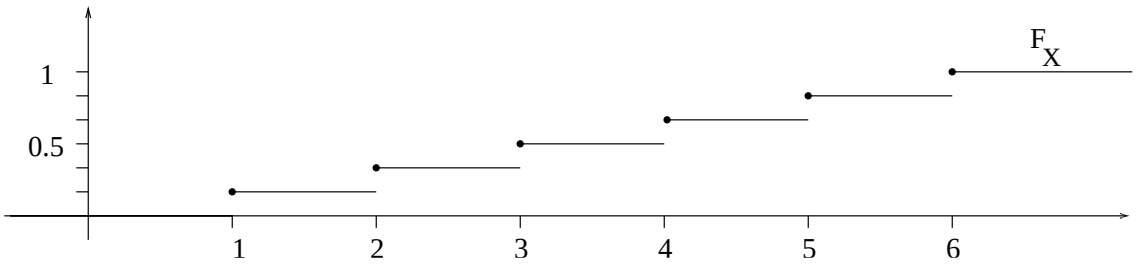
Definition: Die *Verteilung* einer (allgemeinen) Zufallsvariable X ist eine mit F_X bezeichnete Funktion von $\mathbb{R}$ nach $[0, 1]$, die wie folgt definiert ist:

$$F_X(x) = \Pr(\{a \in \Omega \mid X(a) \leq x\})$$

Für eine diskrete Variable ergibt sich mit

$$F_X(x) = \sum_{t \leq x,\ t \in \mathrm{Im}(X)} \Pr_X(t)$$

eine Darstellung als abzählbare Summe. In diesem Fall ist die Funktion F_X eine Treppenfunktion. In der Abbildung ist die Verteilung einer Zufallsvariablen X für das Würfelexperiment dargestellt. Dieses Beispiel zeigt, dass die Verteilung einer Zufallsvariablen unstetig sein kann. Es lässt sich aber auch beobachten, dass diese Verteilungsfunktion stetig von rechts ist, d.h. in jedem Punkt existiert der rechtsseitige Limes und stimmt mit dem Funktionswert überein.



Lemma: Für jede Verteilungsfunktion $F = F_X$ einer Variablen $X : \Omega \to \mathbb{R}$ gilt:

a) $x \leq y \quad \Rightarrow \quad F(x) \leq F(y)$ (Monotonie)

b) $\lim\limits_{x\to-\infty} F(x) = 0$ und $\lim\limits_{x\to\infty} F(x) = 1$

c) $\forall\, x \in \mathbb{R} \quad \lim\limits_{h\to 0+} F(x+h) = F(x)$ (Stetigkeit von Rechts)

Beweis: Sei $A_x = \{a \in \Omega \mid X(a) \leq x\} \in \mathcal{F}$ und $F(x) = \Pr(A_x)$

a) $x \leq y \quad \Rightarrow \quad A_x \subseteq A_y \quad \Rightarrow \quad \underbrace{\Pr(A_x)}_{F(x)} \leq \underbrace{\Pr(A_y)}_{F(y)}$

b) $\emptyset = \bigcap\limits_{x=1}^{\infty} A_{-x} \quad \Rightarrow \quad 0 = \Pr(\emptyset) = \lim\limits_{x\to\infty} \Pr(A_{-x}) = \lim\limits_{x\to-\infty} F(x)$

$\Omega = \bigcup\limits_{x=1}^{\infty} A_x \quad \Rightarrow \quad 1 = \Pr(\Omega) = \lim\limits_{x\to\infty} \Pr(A_n) = \lim\limits_{x\to\infty} F(x)$

c) Sei $(x_n)_{n\in\mathbb{N}}$ eine momoton fallende Folge, die (von rechts) gegen x geht, dann ist die Mengenfamilie $(A_{x_n})_{n\in\mathbb{N}}$ monoton absteigend der Durchschnitt dieser Familie ist die Menge A_x. Daraus folgt:

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \Pr(A_{x_n}) = \Pr\big(\textstyle\bigcap_{n\in\mathbb{N}} A_{x_n}\big) = \Pr(A_x) = F(x)$$

Definition 4: Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ ist *stetig*, wenn eine Funktion $f : \mathbb{R} \to \mathbb{R}^+$ existiert, so dass

$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt$$

Die Funktion $f = f_X$ wird *Dichte* der Verteilung genannt.

Beispiel: Sei X eine Zufallsvariable für die Gleichverteilung in einem reellen Intervall I, hier $I = [\,1, 3\,]$

Zur Bestimmung der Verteilungsfunktion F_X beobachtet man zuerst, dass $F_X(x)$ für alle $x < 1$ den Wert 0 annimmt (denn das Ereignis $X^{-1}(-\infty, x]$ ist leer) und für alle $x > 3$ den Wert 1. Im Bereich $x \in [\,1, 3\,]$ ist die Wahrscheinlichkeit dafür, dass ein zufälliger Punkt aus $[\,1, 3\,]$ kleiner oder gleich x ist, durch die relative Größe des Intervalls $[\,1, x\,]$ im Gesamtintervall $[\,1, 3\,]$ beschrieben, d.h.

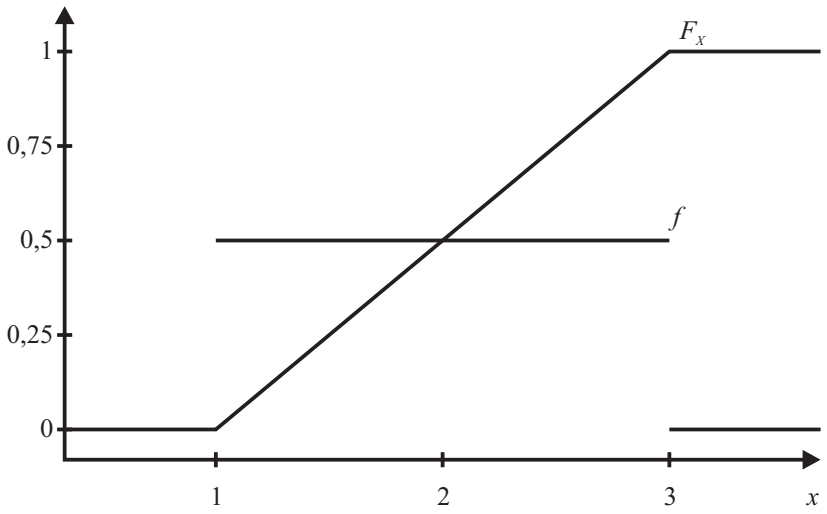
$$F_X(x) = \Pr(X \leq x) = \frac{x-1}{3-1} = \frac{x-1}{2}$$

. Da der Zusammenhang zwischen Dichtefunktion f_X und Verteilung F_X wie im (Hauptsatz der Differential- und Integralrechnung durch

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

gegeben ist, muss man die Funktion F_X ableiten, um auf f_X zu kommen:

$$f(x) = \begin{cases} \frac{1}{2} & \text{falls} \quad x \in [\,1,3\,] \\ \\ 0 & \text{sonst} \end{cases}$$



Ein Vergleich zwischen diskreten und stetigen Zufallsvariablen zeigt starke Parallelen zum Übergang von einer Riemann-Summe zum bestimmten Integral auf:

diskret	**stetig**
diskrete Verteilung $\Pr_X$ durch punktweise Wahrscheinlichkeiten $\Pr(X = x)$	stetige Verteilung F_X durch Dichtefunktion f_X
Addition der Einzelwahrscheinlichkeiten	Integration über der Dichtefunktion

3.4 Erwartungswerte

Für eine diskrete, gleichverteilte Zufallsvariable ist ihr erwarteter Wert der Durchschnitt (Mittelwert) aller möglichen Werte der Variable. Viele Zufallsgrößen sind aber nicht gleichverteilt. In einem solchen Fall gibt die Wahrscheinlichkeit $\Pr(X = x)$ ein Gewicht dafür an, dass die Variable den Wert x annimmt und man muss ein gewichtetes Mittel bilden. Für den stetigen Fall benötigt man wieder eine zweite Definition, in der die Summe durch ein Integral ersetzt wird.

Definition: Ist $X : \Omega \to \mathbb{R}$ eine diskrete Zufallsvariable, so ist der der *Erwartungswert* von X definiert durch

$$E(X) = \sum_{x \in \mathrm{Im}X} x \cdot \Pr_X(x) = \sum_{x \in \mathrm{Im}X} x \cdot \Pr(\{a \in \Omega \mid X(a) = x\})$$

falls diese Reihe absolut konvergiert.

Definition: Ist $X : \Omega \to \mathbb{R}$ eine stetige Zufallsvariable mit der Dichtefunktion f_X, so ist

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

falls beide uneigentliche Integrale existieren (Grenzwerte bei denen die untere Integrationsschranke gegen $-\infty$ und die obere Integrationsschranke gegen ∞ gehen).

Satz (Linearität der Erwartungswerte): Sind X und Y Zufallsvariablen über $(\Omega, \mathcal{F}, \Pr)$ mit den Erwartungswerten $E(X)$ und $E(Y)$ und definiert man neue Zufallsvaraible $X+Y$ und $\alpha \cdot X$, die durch $(X+Y)(a) = X(a) + Y(a)$ und $(\alpha \cdot X)(a) = \alpha \cdot X(a)$ und $\alpha \in \mathbb{R}$, wobei α eine beliebige reelle Zahl ist, dann gilt:

- $E(X+Y) = E(X) + E(Y)$ und
- $E(\alpha \cdot X) = \alpha \cdot E(X)$

Der Beweis ist trivial für diskrete Zufallsvariable (endliche Summation oder Grenzwertsätze). Im stetigen Fall ist der Beweis für die Summe technisch etwas aufwendiger, denn um auf die Defintion zurück zu kommen, muss man die Dichtefunktion von $X+Y$ bestimmen. Wir verzichten hier auf die Details.

Für die bereits eingeführten Standardbeispiele ergeben sich folgende Erwartungswerte:

1. Bernoulli-Verteilung mit Parameter p:

   Eine Variable X mit dieser Verteilung hat das Bild $\mathrm{Im}(X) = \{0, 1\}$ und die diskrete Veretilung $\Pr_X(1) = p$ und $\Pr_X(0) = 1-p$. Der Erwartungswert wird nach Definition bestimmt:
   $$E(X) = 1 \cdot p + 0 \cdot (1-p) = p$$

2. Binominialverteilung mit den Parametern n und p:

   Eine Variable X mit dieser Verteilung kann offensichtlich als Summe von n Variablen mit Bernoulli-Verteilung dargestellt werden: $X = X_1 + X_2 + X_3 + \ldots X_n$. Da man die Erwartungswerte $E(X_i) = p$ der einzelnen Summanden kennt, kann der Satz über die Linearität der Erwartungswerte angewendet werden:

   $$E(X) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n) = n \cdot p$$

3. Geometrische Verteilung mit Parameter p:

   Eine Variable X mit dieser Verteilung hat das Bild $\mathrm{Im}(X) = \mathbb{N}^+$ und die diskrete Verteilung $\Pr_X(k) = (1-p)^{k-1} \cdot p = q^{k-1} \cdot p$.
   Zur Bestimmung des Erwartungswertes werden zuerst die Summanden $k \cdot q^{k-1} \cdot p$ in k Summanden $q^{k-1} \cdot p$ zerlegt und diese neu in Summen zusammengefasst. Danach muss

man nur noch mehrfach die Summenformel für die geometrische Reihe anwenden:

$$\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k \cdot q^{k-1} \cdot p \\
&= p \cdot \left(\sum_{k=1}^{\infty} q^{k-1} + \sum_{k=2}^{\infty} q^{k-1} + \sum_{k=3}^{\infty} q^{k-1} + \ldots \right) \\
&= p \cdot \left(\sum_{k=0}^{\infty} q^{k} + q \cdot \sum_{k=0}^{\infty} q^{k} + q^2 \cdot \sum_{k=0}^{\infty} q^{k} + \ldots \right) \\
&= p \cdot \left(\frac{1}{1-q} + \frac{q}{1-q} + \frac{q^2}{1-q} + \ldots \right) \\
&= \frac{p}{1-q} \cdot \left(1 + q + q^2 + \ldots\right) \\
&= 1 \cdot \frac{1}{1-q} = \frac{1}{p}
\end{aligned}$$

4. Poisson-Verteilung mit Paramter λ:

   Eine Variable X mit dieser Verteilung hat das Bild $\mathrm{Im}(X) = \mathbb{N}$ und die diskrete Verteilung $\Pr_X(k) = \frac{1}{k!} \cdot \lambda^k \cdot e^{-\lambda}$.
   Zur Bestimmung des Erwartungswertes reichen ein paar elementare Umformungen und die Anwendung der Potenzreihendarstellung der Exponentialfunktion aus:

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k \cdot \frac{1}{k!} \cdot \lambda^k \cdot e^{-\lambda} \\
&= \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \cdot e^{-\lambda} \\
&= \lambda \cdot \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!} \cdot e^{-\lambda} \\
&= \lambda \cdot e^{\lambda} \cdot e^{-\lambda} = \lambda
\end{aligned}$$

5. Gleichverteilung über einem Intervall $[\,a, b\,]$:

   Für eine stetige Variable X, die auf einem Intervall gleichverteilt ist, muss man zuerst die die Dichtefunktion f_X kennen. Wie bereits besprochen, hat f_X den Wert $\frac{1}{b-a}$ auf den Intervall $[\,a, b\,]$ und ist sonst 0. Deshalb kann man bei der Integration die untere Grenze von $-\infty$ auf a und die obere Grenze von ∞ auf b verschieben:

$$\begin{aligned}
E(X) &= \int_{\infty}^{\infty} x \cdot f(x) dx = \int_{a}^{b} x \cdot \frac{1}{b-a}\, dx \\
&= \frac{1}{b-a} \cdot \int_{a}^{b} x\, dx = \frac{1}{b-a} \cdot \frac{1}{2}\, x^2 \Big|_{a}^{b} \\
&= \frac{1}{b-a} \cdot \left(\frac{1}{2}\, b^2 - \frac{1}{2}\, a^2 \right) = \frac{b^2 - a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} \\
&= \frac{a+b}{2}
\end{aligned}$$

3.5 Abweichungen vom Erwartungswert

Eine wichtige Frage, die sich nach Einführung des Erwartungswerts stellt, ist die nach der Stärke der Abweichung der Werte einer Variablen von ihrem Erwartungswert. Eine präzisere Formulierung dieser Fragestellung lässt sich aus dem folgenden klassichen Problem ableiten: Wird ein Würfel n mal geworfen und zählt X die Summe der geworfenen Punkte, so ist $E(X) = \frac{7n}{2}$. Durch Experimente fand man heraus, dass die Wahrscheinlichkeit dafür, dass der Wert von X um mehr als einen konstanten Faktor $\lambda > 1$ von $E(X)$ abweicht, sehr klein ist, d.h. für große n gegen Null geht. Bernoulli konnte einen mathematischen Beweis für diese empirische Tatsache geben. Wesentliche Vereinfachungen und gleichzeitg Verallgemeinerungen dieses ursprünglich sehr komplizierten Beweises gehen auf den russischen Mathematiker Tschebyscheff zurück. Zum Einstieg stellen wir ein einfach zu beweisendes Resultat vor, mit dem die obere Abweichung vom Erwartungswert abgeschätzt werden kann.

Satz (Markow-Ungleichung): Sei $X : \Omega \to \mathbb{R}^{\geq 0}$ eine Zufallsvariable mit dem Erwartungswert $E(X)$ und $t > 0$, dann gilt:

$$\Pr(X \geq t) \leq \frac{E(X)}{t}$$

Beweis: Die Summe (bzw. das Integral), die den Erwartungswert beschreiben wird so aufgetrennt, dass in einer Teilsumme alle Werte $x < t$ und in der anderen alle Werte $x > t$ zusammengefasst sind. Da nach Voraussetzung X keine negativen Werte annehmwn darf, kann man die erste Teilsumme von unten mit 0 abschätzen. Im diskreten Fall führt das zu der folgenden Ungleichung:

$$\begin{aligned}
E(X) &= \sum_{x \in \mathrm{Im}X} x \cdot \Pr(X = x) \\
&= \sum_{\substack{x \in \mathrm{Im}X \\ x < t}} x \cdot \Pr(X = x) + \sum_{\substack{x \in \mathrm{Im}X \\ x \geq t}} x \cdot \Pr(X = x) \\
&\geq \sum_{\substack{x \in \mathrm{Im}X \\ x \geq t}} x \cdot \Pr(X = x) \\
&\geq t \cdot \sum_{\substack{x \in \mathrm{Im}X \\ x \geq t}} \Pr(X = x) \\
&= t \cdot \Pr(X \geq t) \\
\\
\frac{E(X)}{t} &\geq \Pr(X \geq t)
\end{aligned}$$

Im stetigen Fall wird die gleiche Idee auf Integrale übertragen. Nach Voraussetzung kann man als untere Integrationsschranke 0 ansetzen:

$$E(X) = \int_0^\infty x \cdot f_X(x)\, dx$$

$$
\begin{aligned}
E(X) &= \int_0^t x \cdot f_X(x)\,dx + \int_t^\infty x \cdot f_X(x)\,dx \\
&\geq \int_t^\infty x \cdot f_X(x)\,dx \\
&\geq t \cdot \int_t^\infty f_X(x)\,dx \\
&= t \cdot \Pr(X \geq t)
\end{aligned}
$$

$$
\frac{E(X)}{t} \geq \Pr(X \geq t)
$$

Die Anwendung der Markow-Ungleichung liefert in vielen Fällen nur eine sehr grobe obere Abschätzung der Wahrscheinlichkeit $\Pr(X \geq t)$. Ohne zusätzliche Informationen kann man aber keine bessere Abschätzung angeben, denn wenn X nur die Werte 0 und t annimmt, gilt sogar die Gleichheit $t \cdot \Pr(X \geq t) = E(X)$.
In der folgenden typischen Anwendungssituation wird gezeigt, wie man das Abschätzungsergebnis (eventuell) verbessern kann, wenn durch eine Verschiebung des Wertebereichs der Variable dafür gesorgt wird, dass sie auch den Wert 0 oder Werte nahe der 0 annimmt.

Beispiel: In einer Gruppe von Personen mit der Durchschnittsgröße von $1,75\,m$ ist der Anteil derjenigen, die $2\,m$ oder größer sind, höchstens $87,5\%$.
Zur Begründung betrachtet man die Körpergröße als Zufallsvariable X mit $E(X) = 1.75$ und setzt $t = 2.0$. Nach Anwendung der Markow-Ungleichung ist

$$
\Pr(X \geq t) \leq \frac{E(X)}{t} = \frac{1.75}{2} = 0.875
$$

Offensichtlich liegt dieses aus praktischer Sicht unzureichenede Ergebnis darin begründet, dass man bei Verwendung der Ungleichung von der schlechtesten Annahme ausgeht, nämlich, dass alle Personen, die kleiner als $2\,m$ sind, die Größe 0 haben. Mit einer Zusatzinformation kann man zur folgenden Verbesserung kommen:
In einer Gruppe von Personen mit der Durchschnittsgröße von $1,75\,m$, in der niemand kleiner als $1.5\,m$ ist, beträgt der Anteil derjenigen, die $2\,m$ oder größer sind, höchstens 50%.
Zur Begründung betrachtet man die Zufallsvariable $Y = X - 1.5$. Nach Voraussetzung hat Y keine negativen Werte und $E(Y) = E(X) - 1.5 = 0.25$. Wegen $X(a) \geq 2.0 \Longleftrightarrow Y(a) \geq 0.5$ kann man jetzt die Markow-Ungleichung für Y anwenden:

$$
\Pr(X \geq 2.0) = \Pr(Y \geq 0.5) \leq \frac{E(Y)}{0.5} = \frac{0.25}{0.5} = 0.5
$$

Die Nutzung der Markow-Ungleichung wird durch zwei weitere Probleme eingeschränkt: Sie ist nur für Varaible ohne negative Werte anwendbar und es können nur die oberen Abweichungen vom Erwartungswert abgeschätzt werden. Um auch die unteren Abweichungen zu berücksichtigen, wäre es besser, statt $\Pr(X \geq \lambda E(X)) = \Pr(X - E(X) \geq (\lambda - 1)E(X))$ die Wahrscheinlichkeit $\Pr(|X - E(X)| \geq (\lambda - 1)E(X))$ zu untersuchen. Da die Betragsfunktion insbesondere bei der Intergration zu neuen Problemen führen kann, ersetzt man $|X - E(X)|$ durch $(X - E(X))^2$. Ein erster Schritt zu verbesserten Abschätzungen der Abweichung vom Erwartungswert führt über den Erwartungswert der Zufallsvariablen $Y = (X - E(X))^2$.

Definition: Die *Varianz* $\mathrm{Var}(X)$ einer Zufallsvariable X mit $E(X) = \mu$ ist der Erwartungswert $E((X-\mu)^2)$. Die Größe $\sigma = \sqrt{\mathrm{Var}(X)}$ wird *Standardabweichung* von X genannt.

Die Varianz von kann man auch mit dem sogenannten zweiten Moment von X berechnen (das i-te *Moment* von X ist der Erwartungswert $E(X^i)$):

$$\begin{aligned}\mathrm{Var}(X) &= E((X-\mu)^2) = E((X-E(X))^2)\\ &= E\left(X^2 - 2\cdot E(X)\cdot X + (E(X))^2\right)\\ &= E(X^2) - 2\cdot E(X)\cdot E(X) + (E(X))^2\\ &= \underbrace{E(X^2)}_{\text{2. Moment}} - \underbrace{(E(X))^2}_{\text{(1. Moment)}^2}\end{aligned}$$

Um den Erwartungswert $E(X^2)$ nach Definition zu berechen, müuste man im diskreten Fall die diskrete Verteilung von X^2 und im stetigen Fall die Dichtefunktion von X^2 bestimmen. Das ist aber nicht notwendig. Bei Anwendung der folgenden Sätze mit der Funktion $g(x) = x^2$ reicht es aus, die diskrete Verteilung von X bzw. die Dichtefunktion von X zu kennen.

Satz: Ist X eine diskrete Zufallsvariable auf $(\Omega, \mathcal{F}, \Pr)$ mit der diskreten Verteilung $\Pr_X$ und ist $g : \mathbb{R} \longrightarrow \mathbb{R}$ eine beliebige Funktion, dann ist auch die durch $Y(a) = g(X(a))$ definierte Abbildung eine diskrete Zufallsvariable auf $(\Omega, \mathcal{F}, \Pr)$ und falls der Erwartungswert von Y existiert, gilt:

$$E(Y) = \sum_{x \in Im(X)} g(x)\Pr_X(x)\, dx.$$

Satz: Ist X eine stetige Zufallsvariable auf $(\Omega, \mathcal{F}, \Pr)$ mit der Dichtefunktion f_X und ist $g : \mathbb{R} \longrightarrow \mathbb{R}$ eine stetige Funktion, dann ist auch die durch $Y(a) = g(X(a))$ definierte Abbildung eine stetige Zufallsvariable auf $(\Omega, \mathcal{F}, \Pr)$ und falls der Erwartungswert von Y existiert, gilt:

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx.$$

Wir verzichten hier sowohl auf die Beweise dieser Sätze als auch des folgenden Satzes, der bei der Bestimmung der Varianz der Binomialverteilung gute Dienste leistet.

Satz: Sind X und Y zwei unabhängige Zufalsvariablen (d.h. für beleibige $s, t \in \mathbb{R}$ sind die Ereignisse $\{a \in \Omega \,|\, X(a) \leq s\}$ und $\{a \in \Omega \,|\, Y(a) \leq t\}$) unabhängig), dann ist $E(X \cdot Y) = E(X) \cdot E(Y)$ und $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Beispiele:

1. Für eine Bernoulli-verteilte Zufallsvariable X mit Parameter p gilt:

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = (1^2 \cdot p + 0^2 \cdot (1-p)) - (1\cdot p + 0\cdot(1-p))^2 = p - p^2$$

2. Eine binomialverteilte Zufallsvariable X mit den Parametern n und p ist Summe von n unabhängigen Bernoulli-verteilten Zufallsvariable X_i und folglich:

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = n \cdot (p - p^2)$$

3. Für eine geometrisch verteilte Zufallsvariable mit Parameter p berechnet man zuerst $E(X^2)$ und verwendet dabei ähnliche Tricks wie beim Erwartungswert (Teleskopsummen, Ausklammern und Indexverschiebung, Formel für geometrische Summe):

$$\begin{aligned}
E(X^2) &= \sum_{k=1}^{\infty} k^2 \cdot q^{k-1} \cdot p \\
&= \sum_{k=1}^{\infty} 1^2 \cdot q^{k-1} \cdot p + \sum_{k=2}^{\infty} \underbrace{(2^2 - 1^2)}_{(2+1)(2-1)} \cdot q^{k-1} \cdot p + \sum_{k=3}^{\infty} \underbrace{(3^2 - 2^2)}_{(3+2)(3-2)} \cdot q^{k-1} \cdot p + \ldots \\
&= 1 + (1+2) \cdot q \cdot \underbrace{\sum_{k=1}^{\infty} p \cdot q^{k-1}}_{1} + (1+4) \cdot q^2 \cdot \underbrace{\sum_{k=1}^{\infty} p \cdot q^{k-1}}_{1} + \ldots \\
&= 1 + q + q^2 + q^3 + \ldots \quad + 2 \cdot q + 4 \cdot q^2 + 6 \cdot q^3 + \ldots \\
&= \frac{1}{1-q} + 2q \cdot \sum_{k=0}^{\infty} k \cdot q^k = \frac{1}{p} + \frac{2q}{p} \sum_{k=0}^{\infty} k \cdot q^k \cdot p \\
&= \frac{1}{p} + \frac{2q}{p} \cdot E(X) = \frac{1}{p} + \frac{2q}{p^2} \\
&= \frac{1+q}{p^2} = \frac{2-p}{p^2}
\end{aligned}$$

Daraus ergibt sich

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2} = \frac{1}{p^2} - \frac{1}{p}$$

4. Für eine Poisson-verteilte Zufallsvariable mit Parameter λ kann man mit relativ einfachen Umformungen die folgenden Werte ausrechnen:

$$E(X^2) = \lambda^2 + \lambda \qquad \text{und} \qquad \mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

5. Für eine gleichverteilte Zufallsvariable X auf dem reellen Intervall $[\,a, b\,]$ bestimmt man durch Integration

$$E(X^2) = \int_a^b \frac{x^2}{b-a}\, dx = \frac{b^2 + ab + a^2}{3} \quad \text{und erhält} \quad Var(X) = \frac{(b-a)^2}{12}$$

Satz (Tschebyscheff-Ungleichung): Sei X eine Zufallsvariable mit dem Erwartungswert $E(X) = \mu$ und der Varianz $Var(X) = \sigma^2$, dann gilt für alle $c > 0$:

$$\Pr(|X - \mu| \geq c) \quad \leq \quad \frac{\sigma^2}{c^2}$$

Spezialfall für $E(X) = \mu = 0$:

$$\Pr(|X| \geq c) \quad \leq \quad \frac{E(X^2)}{c^2}$$

Eine der wichtigsten Anwendungen dieser Ungleichung ist die am Anfang erwähnte Konzentration der Binomialverteilung um den Erwartungswert. Dazu betrachten wir eine binomialverteilte Zufallsvariable X mit den Parametern n und $p = \frac{1}{2}$. Wir wir wissen, ist $E(X) = n \cdot \frac{1}{2} = \frac{n}{2}$ und $\mathrm{Var}(X) = n \cdot \left(\frac{1}{2} - \left(\frac{1}{2}\right)^2\right) = \frac{n}{4}$ Wählt man für das c in der Tschebyscheff-Ungleichung den Wert $\frac{n}{K}$ wobei K eine feste, aber beliebig große Zahl ist, so ergibt sich,

$$\Pr\left(\left|X - \tfrac{n}{2}\right| \geq \tfrac{n}{K}\right) \quad \leq \quad \frac{\frac{n}{4}}{\left(\frac{n}{K}\right)^2} = \frac{K^2}{4n}$$

$$\lim_{n\to\infty} \frac{K^2}{4n} \quad = \quad 0$$

Das heißt, dass die Wahrscheinlichkeit dafür, dass bei n Würfen weniger als $\frac{n}{2} - \frac{n}{K}$ oder mehr als $\frac{n}{2} + \frac{n}{K}$ der Ergebnisse Köpfe sind, geht für große n gegen 0.
Dagegen liefert die Markow-Ungleichung selbst bei einer größeren Abweichung von $\frac{n}{4}$ nur eine sehr grobe Abschätzung:

$$\Pr\left(X \geq \tfrac{3n}{4}\right) \quad \leq \quad \frac{\frac{n}{2}}{\frac{3n}{4}} = \frac{2}{3}$$

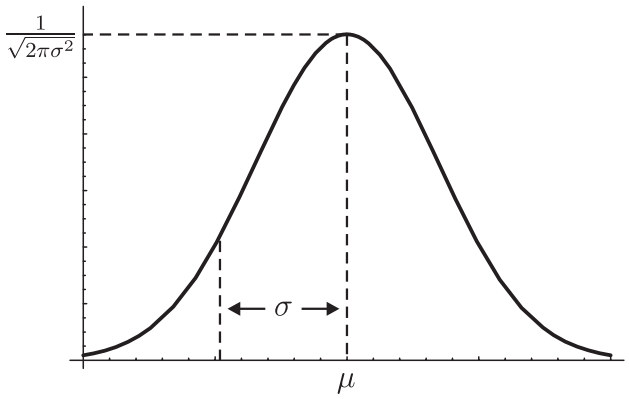
Normalverteilung und Exponentialverteilung

Abschließend führen wir zwei weitere Verteilungen ein, die in vielen Anwendungen eine wichtige Rolle spielen.

Definition: Für die Beschreibung einer Normalverteilung kann man den Erwartungswert μ und die Standardabweichung σ als Parameter vorgegeben und eine Dichtefunktion der folgenden Form definieren:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}\cdot(x-\mu)^2}$$

Der Graph der Dichtefunktion hat die Form einer Glocke. Die daraus resultierende Verteilungsfunktion wird Gaußsche Normalverteilung genannt und mit $N(\mu, \sigma^2)$ bezeichnet.



Bemerkung: Aus der Definition der Dichtefunktion $f(x)$ kann man leicht die Symmetrie bezüglich der Gerade $x = \mu$ ablesen und daraus ergibt sich μ als Erwartungswert. Dagegen ist der Nachweis, dass die Fläche unter der Kurve gleich 1 ist und die Standardabweichung gleich σ ist, wesentlich anspruchsvoller. Man kann außerdem zeigen, dass σ gleichzeitig der Abstand von μ zu den Wendestellen der Funktion $f(x)$ ist.

Definition: Eine Zufallsvariable X ist exponentialverteilt (bezüglich eines vorgegebenen Parameters $\lambda > 0$), wenn die Verteilungsfunktion F_X für alle $x \leq 0$ gleich 0 ist und für jedes positve $x \in \mathbb{R}$ den Wert $1 - e^{-\lambda x}$ annimmt. Daraus kann man die folgende Dichtefunktion ableiten:

$$f_X(x) = \begin{cases} 0 & \text{falls } x \leq 0 \\ \lambda e^{-\lambda x} & \text{sonst} \end{cases}$$

Durch partielle Integration kann man den Erwartungswert und die Varianz einer Exponentialverteilung relativ einfach berechnen:

$$E(X) = \frac{1}{\lambda} \qquad E(X^2) = \frac{2}{\lambda^2} \qquad \text{und} \qquad \mathrm{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$