**Dr. C.J. Luchsinger, christoph.luchsinger@math.uzh.ch , 2023-10-24**

**Crash-Kurs Lineare Algebra für die Statistik**

Matrizen sind aus zwei Gründen für die Statistik sehr wichtig: Sie ermöglichen uns *einerseits* eine sehr elegante und kompakte Formulierung von wichtigen Formeln. *Andererseits* können wir Dank geometrischer Überlegungen im $\mathbb{R}^n$ viele Resultate einfach herleiten, welche sonst extrem mühsam erarbeitet werden müssten.
Wir verzichten auf Beweise. Zudem wird davon ausgegangen, dass dieses Kapitel weitgehend Repetition ist. LeserInnen, welche bisher mit der linearen Algebra auf Kriegsfuss gestanden haben, können insofern beruhigt sein, als dass wir in diesem Kapitel in erster Linie einfach alle relevanten Resultate im Zusammenhang mit *Matrizen* repetieren, welche für die Statistik wichtig sind. Vielleicht wird man durch die Anwendungen die Schönheit der linearen Algebra entdecken. Vorkenntnisse in Statistik sind sicher hilfreich, aber nicht zwingend notwendig.

## 1 Grundlegende Definitionen

Eine $(m \times n)$-**Matrix** $A$ ist ein rechteckiges Zahlenschema der folgenden Art:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ & & \cdots & & \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}.$$

Diese Matrix hat $m$ Zeilen (rows) und $n$ Spalten (columns). Wenn $m = n$ sprechen wir von einer **quadratischen Matrix** (square matrix). Die Einträge $a_{ij}$ ($a_{\text{Zeile,Spalte}}$) sind bei uns immer aus $\mathbb{R}$, falls nicht explizit anders angegeben. Matrizen werden wir immer mit grossen Buchstaben bezeichnen. Wenn $m = n = 1$, so ist die Matrix (für uns) einfach eine reelle Zahl. Die dritte Spalte bezeichnen wir mit $a_{.3}$, die zweite Zeile mit $a_{2.}$. In der Statistik werden wir Datenmatrizen (Design-Matrizen) der folgenden Art haben:

$$\begin{pmatrix} 1 & 68 & 6 & \cdots & 85000 \\ 1 & 72 & 4.5 & \cdots & 66000 \\ & & \cdots & & \\ 0 & 66 & 5.5 & \cdots & 73000 \end{pmatrix}.$$

Die Daten in derselben Zeile gehören dann zum gleichen Objekt (Person) und die Daten in den Spalten sind verschiedene Merkmale von Personen: die erste Spalte kann das Geschlecht sein (0 = männlich, 1 = weiblich), die zweite Spalte ist der Jahrgang, die dritte Spalte die Note in einem Fach, die letzte Spalte das Einkommen der Eltern.

Ein **Vektor** $\vec{x}$ ist eine geordnete Menge von reellen Zahlen. Wir werden Vektoren immer als Spaltenvektoren auffassen:

$$\vec{x} := \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{pmatrix}.$$

Wir werden die Spalten von Matrizen manchmal als Vektoren auffassen und die Zeilen als (transponierte (s.u.)) Vektoren. Umgekehrt kann man Vektoren immer auch als $(n \times 1)$-Matrizen auffassen. Zudem werden wir Vektoren immer mit dem Pfeil über dem (kleinen) Buchstaben bezeichnen, um sie von der Notation her von reellen Zahlen zu unterscheiden. Ausnahmen sind die oben eingeführten Zeilen und Spalten von Matrizen ($a_{2.}$, $a_{.3}$). Die **kanonischen Einheitsvektoren** im $\mathbb{R}^n$ sind bei uns

$$\vec{e}_i := \begin{pmatrix} 0 \\ 0 \\ . \\ 1 \\ . \\ 0 \end{pmatrix}$$

mit einem Eintrag 1 in der $i$-ten Koordinate und sonst mit Einträgen $= 0$.

Spezielle Matrizen sind

* die **symmetrische Matrix** (z.B. Kovarianzmatrizen in der Statistik) wo $a_{ij} = a_{ji}$ für alle $1 \leq i, j \leq m = n$, zum Beispiel

$$\begin{pmatrix} 2 & 3 & 4 \\ 3 & 5 & 6 \\ 4 & 6 & 9 \end{pmatrix}.$$

* die **Diagonalmatrix** wo $a_{ij} = 0$ für alle $1 \leq i, j \leq m = n$ mit $i \neq j$, zum Beispiel

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

* die **Einheitsmatrix (Identitätsmatrix)**; sie ist eine Diagonalmatrix mit Diagonaleinträgen 1:

$$\mathbf{I}_3 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Wir geben mit dem Index 3 gleich die Dimension an.

**1.R Statistik-Paket R/S-PLUS**

matrix(c(3,2,1,4,5,6), c(2,3)) liefert uns (spaltenweise eingelesen!) die Matrix

$$\begin{pmatrix} 3 & 1 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Wenn wir bereits die Vektoren $a = (3, 2), b = (1, 4), c = (5, 6)$ haben, so können wir mit $cbind(a, b, c)$ (**c**olumn **b**ind) obige Matrix ebenfalls erzeugen. Mit $A[2, 2]$ erhalten wir den Eintrag $a_{2,2}$ der Matrix $A$. Mit $A[, 3]$ erhalten wir die dritte Spalte, mit $A[2, ]$ erhalten wir die zweite Zeile. Mit $a[3]$ erhalten wir die dritte Koordinate des Vektors $\vec{a}$. Mit diag(c$(1, 2, 3, 4)$) erhalten wir eine Diagonalmatrix der Dimension 4 mit Einträgen $1, 2, 3, 4$ (entsprechend gibt man die Einheitsmatrix ein).

## 2 Algebraische Eigenschaften von Matrizen

Vektoren kann man auch als spezielle $(n \times 1)$-Matrizen auffassen. Deshalb gelten die Resultate unten auch für Vektoren.

### 2.1 Gleichheit von Matrizen

Die $(m \times n)$-Matrix $A$ und die $(q \times r)$-Matrix $B$ sind per Definitionem genau dann gleich, wenn $m = q, n = r$ und $a_{ij} = b_{ij}$ für alle $1 \leq i \leq m; 1 \leq j \leq n$.

### 2.2 Transponieren von Matrizen

Wir transponieren eine Matrix, indem wir die $i$-te Zeile als $i$-te Spalte schreiben: $B$ ist per Definitionem genau dann die **transponierte Matrix** von $A$ (Notation $A^t$), wenn

$$b_{ij} = a_{ji}$$

für alle $i, j$. Ist $A$ eine $(m \times n)$-Matrix, so ist $A^t$ eine $(n \times m)$-Matrix.

Eine Matrix ist genau dann symmetrisch, wenn $A^t = A$. Es gilt $(A^t)^t = A$.

### 2.3 Addition von Matrizen und skalare Multiplikation

Die **Addition (analog Subtraktion) von Matrizen** (von gleicher Dimension!) erfolgt elementweise: Sind $A$ und $B$ $(m \times n)$-Matrizen, so ist die $(m \times n)$-Matrix $C$ die Summe von $A$ und $B$ ($C = A + B$) genau dann wenn

$$c_{ij} = a_{ij} + b_{ij}$$

für alle $1 \leq i \leq m, 1 \leq j \leq n$. Die Nullmatrix $\mathbf{0}$ (alle Einträge gleich 0) liefert:

$$A + \mathbf{0} = \mathbf{0} + A = A.$$

Die Addition von Matrizen ist kommutativ und assoziativ:

$$A + B = B + A, \qquad (A + B) + C = A + (B + C)$$

und

$$(A+B)^t = A^t + B^t.$$

Die **skalare Multiplikation** (Multiplikation mit einem Skalar (einer reellen Zahl)) bei Matrizen erfolgt ebenfalls elementweise: $B := \lambda A$, genau dann wenn

$$b_{ij} = \lambda a_{ij}$$

für alle $1 \leq i \leq m, 1 \leq j \leq n$.

### 2.4 Skalarprodukt und Matrizenmultiplikation

Wenn wir zwei Vektoren $\vec{a}, \vec{b}$ von gleicher Dimension $n$ haben, so definieren wir das **Skalarprodukt (oder inneres Produkt)** als

$$(\vec{a})^t \vec{b} := \sum_{i=1}^{n} a_i b_i \in \mathbb{R}. \tag{1.1}$$

Wir haben also mit $(\vec{a})^t$ eine (liegende) Zeile und mit $\vec{b}$ eine (stehende) Spalte. Am besten wird man dieses Paar je als kleine lange Rechtecke derart darstellen, dass die obere rechte Ecke von $(\vec{a})^t$ mit der unteren linken Ecke von $\vec{b}$ zusammenfällt. Dann kann man schön Bögen schlagen (die Produkte $a_i b_i$) und aufsummieren, um die Summe aus (1.1) zu erhalten:

Dies ist praktisch, zum Beispiel wenn Vektoren (und gleich nachfolgend Matrizen) voll von Nullen sind. Man sieht dann sofort, wo relevante Ausdrücke vorkommen. Offensichtlich gilt $(\vec{b})^t \vec{a} = (\vec{a})^t \vec{b}$; siehe (1.1).

Nach diesen Vorarbeiten kann man die **Matrizenmultiplikation** einfach einführen: Sei $A$ eine $(m \times n)$-Matrix, $B$ eine $(n \times k)$-Matrix und $C$ eine $(m \times k)$-Matrix. Dann ist $C$ genau dann das Produkt $AB$, wenn für alle $c_{ij}$, $1 \leq i \leq m, 1 \leq j \leq k$ gilt

$$c_{ij} = \sum_{u=1}^{n} a_{iu} b_{uj}. \tag{1.2}$$

**Warnung:** Die Anzahl Spalten von $A$ und die Anzahl Zeilen von $B$ muss identisch sein. $C$ "erbt" die Anzahl Zeilen von $A$ und die Anzahl Spalten von $B$. Wir werden im folgenden manchmal einfach $AB$ schreiben, ohne jedesmal die notwendigen Dimensionen der involvierten Matrizen anzugeben.

In Anlehnung an unser Schema zur Berechnung des Skalarprodukts, kann man hier ein praktisches Schema für die Matrizenmultiplikation aufstellen, welches obige Warnung überflüssig macht. (1.2) ist offensichtlich auch ein Skalarprodukt, nämlich von der $i$-ten Zeile von $A$ mit der $j$-ten Spalte von $B$. Damit wird man die Matrizenmultiplikation am besten folgendermassen durchführen: die drei Matrizen $A, B$ und $C$ sind Rechtecke. Wieder wird man $A$ und $B$ so eintragen, dass die obere rechte Ecke von $A$ mit der unteren linken Ecke von $B$ zusammenfällt. Dann erhält man rechts von $A$ und unter $B$ die Matrix $C$. Die Dimensionen stimmen automatisch und die Einträge erhält man durch Skalarprodukte mit (1.2):

Die Matrizenmultiplikation ist im Allgemeinen **nicht kommutativ**: $AB = BA$ gilt im allgemeinen nicht. Man bedenke, dass wegen obiger Warnung $AB$ zwar definiert sein kann, $BA$ aber nicht existieren muss oder eventuell nicht von gleicher Dimension sein muss wie $AB$. Selbst wenn die Dimensionen von $AB$ gleich den Dimensionen von $BA$ sind, muss die Matrix $AB$ nicht gleich der Matrix $BA$ sein.

Die Multiplikation einer Matrix mit einem Vektor (falls die Dimensionen dies zulassen) geschieht analog, da ein Vektor ja einfach eine spezielle Matrix ist. Wir gelangen durch die Multiplikation einer Matrix mit einem Vektor zum Begriff der **Linearkombination:** Wenn

$$A\vec{b} = \vec{c},$$

dann ist das gleichbedeutend mit

$$b_1 a_{.1} + b_2 a_{.2} + \ldots + b_n a_{.n} = \vec{c}.$$

Wir sagen, $\vec{c}$ ist eine Linearkombination der Spalten $a_{.i}$ der Matrix $A$. Wenn wir (aggregiert) anstelle von $\vec{b}$ eine Matrix $B$ und anstelle von $\vec{c}$ eine Matrix $C$ haben:

$$AB = C,$$

so ist analog jede Spalte von $C$ eine Linearkombination der Spalten von Matrix $A$:

$$Ab_{.j} = c_{.j}.$$

Die kanonischen Einheitsvektoren sind praktisch, wenn wir eine spezielle Spalte (oder Zeile) aus der Matrix $A$ "herausschlagen" wollen:

$$A\vec{e}_i = a_{.i}, \tag{1.3}$$

mit

$$(\vec{e}_i)^t A = a_{i.}$$

erhalten wir analog die $i$-te Zeile von $A$ (Vorsicht: Dimensionen!). Aus (1.3) folgt sofort

$$A\mathbf{I} = A,$$

analog

$$\mathbf{I}A = A.$$

Seien Sie sich bitte bewusst, dass in obigen beiden Formeln die Matrix $A$ (im Gegensatz zu $\mathbf{I}$) *nicht* quadratisch sein muss. Die Multiplikation von $A$ mit $\mathbf{I}$ muss jedoch von den Dimensionen her erlaubt sein!

Weitere Rechenregeln der Matrizenmultiplikation:

Multiplikation mit der Nullmatrix: $\mathbf{0}A = \mathbf{0}$, $A\mathbf{0} = \mathbf{0}$.

Assoziativgesetz: $(AB)C = A(BC)$

Distributivgesetz: $A(B + C) = AB + AC$ und $(A + B)C = AC + BC$

Transponieren eines Produkts (wichtig in der Statistik): $(AB)^t = B^t A^t$

Verallgemeinert (Transponieren von Produkten): $(ABC)^t = C^t B^t A^t$.

**2.R Statistik-Paket R/S-PLUS**

Mit $t(A)$ transponieren wir die Matrix. $A+B$ liefert die Summe der Matrizen, $3*B$ ist die skalare Multiplikation. Das Skalarprodukt von $a$ und $b$ ist $a\%*\%b$ (Transponieren von $a$ ist nicht notwendig) - Vorsicht: bei $a * b$ wird einfach die $i$-te Koordinate von $a$ mit der $i$-ten Koordinate von $b$ multipliziert, das Resultat ist wieder ein Vektor! Matrizenmultiplikation: $A\% * \%B$ ($A * B$ ist wieder die elementweise Multiplikation).

## 3 Elegante und kompakte Formulierung wichtiger Formeln aus der Statistik

Wir führen hier noch den Eins-Vektor **1** ein (nicht mit den Kanonischen Einheitsvektoren verwechseln):

$$\mathbf{1} := \sum_{i=1}^{n} \vec{e}_i = \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \\ 1 \end{pmatrix}.$$

Der Eins-Vektor ist offenbar ein Vektor mit lauter Einsen. Wenn notwendig geben wir mit $\mathbf{1_n}$ noch die Länge des Vektors an. Wir verzichten auf einen Pfeil über dem **1**.

### 3.1 Summe von Zahlen (z.B. bei arithmetischem Mittel)

Mit $(\vec{a})^t := (a_1, a_2, \ldots, a_n)$:

$$\sum_{i=1}^{n} a_i = \mathbf{1}^t \vec{a} = (\vec{a})^t \mathbf{1}.$$

Das arithmetische Mittel ist demnach

$$\overline{a} = \frac{\mathbf{1}^t \vec{a}}{n}.$$

### 3.2 Summe von Quadraten (z.B. bei der Varianz)

Mit $(\vec{a})^t := (a_1, a_2, \ldots, a_n)$:

$$\sum_{i=1}^{n} a_i^2 = (\vec{a})^t \vec{a}.$$

### 3.3 Summe von Kreuzprodukten (z.B. bei der Kovarianz)

Mit $(\vec{a})^t := (a_1, a_2, \ldots, a_n), (\vec{b})^t := (b_1, b_2, \ldots, b_n)$:

$$\sum_{i=1}^{n} a_i b_i = (\vec{a})^t \vec{b}.$$

### 3.4 $A^t A$ (z.B. bei der (multiplen) linearen Regression)

Sei $A$ eine $(n \times k)$-Matrix. $a_{i.}$ ist die $i$-te Zeile von $A$. Dann gilt:

$$A^t A = \sum_{i=1}^{n} a_{i.}^t a_{i.}. \tag{1.4}$$

Die $(k \times k)$-Matrix $A^t A$ kann also mit (1.4) als Summe von $n$ Matrizen dargestellt werden, welche alle Rang 1 haben.

### 3.5 Wichtige Idempotente Matrizen

Eine quadratische Matrix $M$ mit der Eigenschaft

$$MM = M$$

nennen wir **idempotent**. Wenn $M$ sogar symmetrisch ist, dann gilt auch $M^t M = MM^t = M = M^t$.

#### 3.5.1 Beispiel Idempotente Matrizen I: Mittelwert entfernen

Häufig will man in der Statistik einen Datenvektor $\vec{x}$ zentrieren (den Mittelwert abziehen). Wir verwenden dazu den Eins-Vektor (s.o.). Kleine Vorarbeit:

$$\begin{pmatrix} \overline{x} \\ \overline{x} \\ . \\ . \\ . \\ \overline{x} \\ \overline{x} \end{pmatrix} = \mathbf{1}\overline{x} = \mathbf{1}\frac{1}{n}\mathbf{1}^t\vec{x} = \frac{1}{n}\mathbf{1}\mathbf{1}^t\vec{x}.$$

Die Matrix $\frac{1}{n}\mathbf{1}\mathbf{1}^t$ ist eine $(n \times n)$-Matrix mit Einträgen $1/n$. Die zentrierten Daten erhalten wir mit

$$\begin{pmatrix} x_1 - \overline{x} \\ x_2 - \overline{x} \\ . \\ . \\ . \\ x_{n-1} - \overline{x} \\ x_n - \overline{x} \end{pmatrix} = \vec{x} - \mathbf{1}\overline{x} = \vec{x} - \frac{1}{n}\mathbf{1}\mathbf{1}^t\vec{x} = (\mathbf{I_n} - \frac{1}{n}\mathbf{1}\mathbf{1}^t)\vec{x} =: M^z\vec{x}.$$

Wir nennen $M^z$ die **Z**entrierungsmatrix. Die Diagonaleinträge sind alle $1 - 1/n$ und die restlichen Einträge $-1/n$. $M^z$ ist insbesondere symmetrisch. Diese Matrix ist idempotent (Mittelwert nochmals abziehen bringt nichts, ist ja nach dem ersten Mal schon 0). Des weiteren gilt:

$$M^z \mathbf{1} = \vec{0},$$

woraus sofort folgt

$$\mathbf{1}^t M^z = (\vec{0})^t.$$

Damit kann man einfach (mit $M^z$) zeigen:

$$\sum_{i=1}^{n} (x_i - \overline{x}) = 0.$$

Es gelten auch

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = (\vec{x})^t M^z \vec{x}$$

und

$$\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = (\vec{x})^t M^z \vec{y}.$$

### 3.5.2 Beispiel Idempotente Matrizen II: Projektionen in der multiplen Regression

Sei $A$ eine $(n \times k)$-Matrix, $n \geq k$, von Rang $k$. Wir definieren:

$$H := A(A^t A)^{-1} A^t$$

und

$$M := \mathbf{I} - H.$$

Die Ähnlichkeit mit 3.5.1 ist *nicht* zufällig (wähle einfach $A = \mathbf{1}$)! Es gilt: $H$ und $M$ sind symmetrisch und idempotent, $HA = A$ und $H$ und $M$ sind orthogonal zueinander, das heisst, es gilt:

$$HM = 0.$$

Diese Resultate kann man in der linearen Regression zeitsparend einsetzen. Nur so viel sei vorausgeschickt (kein Problem, wer hier nur Bahnhof versteht): Die *einfache* Regression hat (neben dem $\alpha$) nur eine erklärende Variable (mit Parameter $\beta$):

$$y_i = \alpha + \beta x_i + \epsilon_i. \tag{1.5}$$

In der Regression werden wir die *multiple* Regression besprechen. Das Pendant zu (1.5) ist in der multiplen Regression

$$\vec{Y} = A\vec{\beta} + \vec{\epsilon}. \tag{1.6}$$

Dort haben wir (wie bisher) $n$ Datenpunkte. Jetzt wollen wir diese aber durch z.B. $k$ (sog. erklärende) Variablen erklären. Es geht dann darum, die Daten (im $\mathbb{R}^n$) auf eine (höchstens) $k$-dimensionale Hyperebene ($A\mathbb{R}^k$) zu projizieren (als $H\vec{Y}$).

# 4 Geometrische Sicht

## 4.1 Linearkombinationen und Basis, linear (un)abhängig, Unterraum

Unter einem **Vektorraum** verstehen wir eine Menge von Vektoren, welche bezüglich Addition und skalarer Multiplikation abgeschlossen ist. Umgangssprachlich: man fliegt nie aus dieser Menge heraus, wenn man Vektoren aneinanderhängt und/oder streckt. In vielen Vorlesungen wird es immer der $\mathbb{R}^n$ sein - es darf aber auch eine beliebige Menge von abstrakten Objekten sein. Dem Begriff der **Linearkombination** sind wir bereits in 2.4 begegnet. Wenn wir an den $\mathbb{R}^2$ denken, so ist klar, dass wir mit zwei beliebigen Vektoren aus dem $\mathbb{R}^2$ durch Linearkombinationen alle anderen Vektoren im $\mathbb{R}^2$ erreichen, vorausgesetzt, der eine Vektor ist nicht ein skalares Vielfaches des anderen. Wir können auch mehr als 2 Vektoren nehmen. Zentral ist jedoch in diesem Zusammenhang der Begriff der **Basis**: Eine Teilmenge eines Vektorraums ist eine Basis des Vektorraums, genau dann wenn jeder Vektor dieses Vektorraums auf genau eine Art als Linearkombination von Elementen der Basis dargestellt werden kann. Die Anzahl Vektoren der Basis ist die **Dimension** des Vektorraums. Die Auswahl von Elementen, welche man zu einer Basis zusammenfügen will, ist nicht eindeutig. Für den $\mathbb{R}^n$ drängen sich die $n$ Einheitsvektoren auf. Dort ist auch sofort ersichtlich, dass es eine Basis ist: jeder Vektor kann mit den Einheitsvektoren dargestellt werden und die Darstellung ist eindeutig. Wenn wir mehr als $n$ Vektoren im $\mathbb{R}^n$ nehmen, so sind diese **linear abhängig**: man kann jeden beliebigen Vektor als Linearkombination der anderen Vektoren darstellen. Wir nennen eine Menge von Vektoren **linear unabhängig** genau dann wenn gilt:

$$\alpha_1\vec{a}_1 + \alpha_2\vec{a}_2 + \ldots + \alpha_n\vec{a}_n = 0 \Rightarrow \alpha_i = 0 \;\forall\; 1 \leq i \leq n.$$

Die Elemente einer Basis sind linear unabhängig. Die Forderung, dass $n$ Vektoren eine Basis sein sollen, ist stark. Viel weniger anspruchsvoll ist es, einfach einen **Unterraum** zu erzeugen: Die Menge von Linearkombinationen von Vektoren $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ nennen wir den (linearen) Unterraum, welcher von $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ erzeugt wird. Dabei müssen diese $k$ Vektoren nicht linear unabhängig sein.

### 4.2 Rang einer Matrix

Wir können Matrizen als Menge von (Spalten)-Vektoren auffassen. Dann definieren wir: Der **Rang einer Matrix** $A$ ist die Dimension des Vektorraums, welcher von den Spalten-Vektoren aufgespannt wird. Dies ist übrigens gleich der Dimension des Vektorraums, welcher von den Zeilen-Vektoren aufgespannt wird (obschon z.B. bei einer $(n \times k)$-Matrix die $k$ Spalten-Vektoren aus dem $\mathbb{R}^n$ sind und die $n$ Zeilen-Vektoren aus dem $\mathbb{R}^k$). Man sagt kurz: Zeilenrang = Spaltenrang; nur deshalb sprechen wir vom Rang einer Matrix. Es gilt logischerweise für eine $(n \times k)$-Matrix $A$:

$$\mathrm{rang}(A) \leq \min(n, k).$$

Wir werden in der Statistik (Regression) üblicherweise $(n \times k)$-Matrizen $A$ haben, wo $n \geq k$ (hoffentlich sogar $n \gg k$) und die Matrix wird Rang $k$ haben. Wir sagen dann, die Matrix habe **vollen Rang**. Es gelten noch:

$$\mathrm{rang}(AB) \leq \min(\mathrm{rang}(A), \mathrm{rang}(B)) \tag{1.7}$$

und (vgl. 3.5.2)

$$\mathrm{rang}(A) = \mathrm{rang}(A^t A) = \mathrm{rang}(AA^t). \tag{1.8}$$

### 4.3 Determinanten

Für *quadratische* Matrizen lässt sich eine sogenannte Determinante definieren. Im Fall von $(2 \times 2)$-Matrizen

$$A := \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

ist dies bekanntlich $a_{11}a_{22} - a_{21}a_{12}$. Bei $(3 \times 3)$-Matrizen benutzt man zur Berechnung am Besten das Schema von Cramer:

Bei höheren Dimensionen wird es kompliziert (nicht schwierig!). Meist braucht man aber nur die relevanten Rechenregeln (z.B. Produktregel) und die Berechnung von Determinanten in Spezialfällen (z.B. Diagonalmatrix) zu kennen. Ohnehin hat man heute gute Computer, welche die Arbeit abnehmen. Die allgemeine Rechenregel geht wie folgt: die Determinante einer $(n \times n)$-Matrix $A$ ist

$$\det(A) = \sum_{i=1}^{n} a_{ij}(-1)^{i+j} \det(A^{(i,j)}). \tag{1.9}$$

Dabei ist $A^{(i,j)}$ diejenige Matrix, welche man erhält, wenn man in $A$ die $i$-te Zeile und $j$-te Spalte einfach komplett herausstreicht. Man kann ein beliebiges $j$ wählen oder auch ein beliebiges $i$ und über die $j$ summieren (tönt wunderbar, ist auch wunderbar: man nehme eine Zeile oder Spalte mit möglichst vielen Nullen, dann haben wir wenig Summanden). Die geometrische Interpretation der Determinanten ist die, dass es das Volumen des Körpers ist, der von den Spaltenvektoren der Matrix aufspannt wird. Wenn wir einen Körper mit Hilfe einer Matrix abbilden, so gibt die Determinante die Volumenänderung an (negativ wenn Anzahl Spiegelungen ungerade).

Wichtige Rechenregeln für Determinanten:

Sei $A$ eine $(n \times n)$-Matrix. Dann gilt: $\det(A) \neq 0 \Leftrightarrow \text{Rang}(A) = n$

Determinante einer Diagonalmatrix ist das Produkt der Diagonalelemente. Also ist z.B. die Determinante der Identitätsmatrix gleich 1 (geometrische Interpretation: das Volumen ändert sich ja wirklich nicht).

$\det(aA) = a^n \det(A)$, wenn $a$ eine reelle Zahl ist und $A$ eine $(n \times n)$-Matrix. Geometrische Interpretation: Wenn wir einen Körper im $\mathbb{R}^3$ um den Faktor 2 in allen Dimensionen strecken (mit $2\mathbf{I}_3$ abbilden), dann wächst das Volumen um den Faktor $2^3 = 8$.

$\det(A) = \det(A^t)$

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

## 5 Inverse von (quadratischen) Matrizen

Eine $(n \times n)$-Matrix $A$ mit Rang $n$ ist invertierbar, das heisst, es existiert eine Matrix $B$, sodass

$$AB = BA = \mathbf{I}_n.$$

Diese Matrix $B$ ist eindeutig und wird mit $A^{-1}$ bezeichnet. Der Begriff der Invertierbarkeit kommt vor allem von linearen Gleichungssystemen. Dort hat man ein Problem der Art $A\vec{x} = \vec{b}$. Man möchte jetzt $\vec{x}$ berechnen. Wenn wir eine Matrix $B$ wie oben haben, so erhalten wir $\vec{x} = B\vec{b} = A^{-1}\vec{b}$.

Wichtige Rechenregeln für Inverse:

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

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

$\det(A^{-1}) = 1/\det(A)$ (geometrische Interpretation: Wenn die Abbildung eines Körpers mit $A$ das Volumen um den Faktor $\det(A)$ vergrössert, dann wird die Umkehrabbildung $A^{-1}$ das Volumen um genau diesen Faktor verkleinern.)

$A$ symmetrisch, dann $A^{-1}$ symmetrisch

Falls involvierte Inverse alle existieren:

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}.$$

Bei obiger Formel sollte man vor allem an Abbildungen denken: erst am Schluss wird man bei der linken Seite mit $A$ abgebildet - wenn man jetzt das Ganze umkehrt (rechte Seite), wird man zuerst mit $A^{-1}$ abgebildet.

Sei $A$ eine $(n \times k)$-Matrix mit $n \geq k$ und $\text{Rang}(A) = k$. Dann haben die $(k \times k)$-Matrix $(A^tA)$ und die $(n \times n)$-Matrix $(AA^t)$ beide Rang $k$ - insbesondere ist $(A^tA)$ invertierbar.

Sind $A$ und $B$ invertierbare Matrizen, so ist $\text{Rang}(ACB) = \text{Rang}(C)$.

Gilt mit *quadratischen!* Matrizen $AB = \mathbf{I}$, so ist $B = A^{-1}$.

## 6 Eigenwerte und Eigenvektoren

Vorbereitung: Zwei Vektoren $\vec{a}$ und $\vec{b}$ sind **orthogonal** zueinander ($\vec{a} \perp \vec{b}$), genau dann wenn

$$(\vec{a})^t \vec{b} = (\vec{b})^t \vec{a} = 0.$$

Dies stimmt mit unserer Vorstellung von "senkrecht" im $\mathbb{R}^n, n \in \{2, 3\}$, überein.

Sei $A$ eine quadratische Matrix, $\lambda$ eine reelle Zahl und $\vec{c} \neq \vec{0}$ ein Vektor. Falls hierfür gilt

$$A\vec{c} = \lambda\vec{c}, \tag{1.10}$$

so nennen wir $\lambda$ einen **Eigenwert** und $\vec{c}$ einen **Eigenvektor** von $A$. Offenbar ist jedes Vielfaches $a\vec{c}$ mit $a$ einer reellen Zahl wieder ein Eigenvektor. Um diese Unbestimmtheit zu eliminieren, fordern wir gerne, dass der Eigenvektor $\vec{c}$ auf 1 normiert sein soll:

$$(\vec{c})^t \vec{c} = 1.$$

Kovarianzmatrizen aus der Statistik sind symmetrisch - und Gott sei Dank vereinfacht dies viele Resultate aus der allgemeinen Theorie. Wir fassen diese Resultate zusammen:

<u>Wichtige Rechenregeln für symmetrische ($n \times n$)-Matrizen (Eigenwerte und -vektoren)</u>:

Die Eigenwerte sind alle reell.

Die Eigenwerte müssen *nicht* alle verschieden sein (z.B. $\mathbf{I_n}$ mit Eigenwert 1).

Die Eigenwerte von $A^t A$ und $AA^t$ sind gleich.

Die Eigenvektoren zu verschiedenen Eigenwerten stehen orthogonal aufeinander.

Es gibt $n$ zueinander orthogonale Eigenvektoren der Länge 1 (z.B. die $n$ kanonischen Einheitsvektoren für $\mathbf{I_n}$).

Die letzte Eigenschaft erlaubt uns folgende Vorgehensweise: Wir ordnen die Eigenwerte zuerst der Grösse nach (in der entsprechenden Reihenfolge ordnen wir auch die dazugehörigen Eigenvektoren). Wir sammeln die Eigenvektoren $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ als Spalten und fassen diese zur $(n \times n)$-Matrix $C$ zusammen: $C := (\vec{c}_1\, \vec{c}_2\, \cdots\, \vec{c}_n)$. Die dazugehörigen Eigenwerte $\lambda_i$ fassen wir in einer Diagonalmatrix $\Lambda$ zusammen, wo der $i$-te Diagonaleintrag $\lambda_i$ ist. Da von (1.10) für jedes $1 \leq i \leq n$ gilt:

$$A\vec{c}_i = \lambda_i \vec{c}_i,$$

haben wir aggregiert auch folgende Gleichung:

$$AC = C\Lambda. \tag{1.11}$$

Da die Eigenvektoren senkrecht aufeinander stehen und auf 1 normiert sind (ortho-normal), haben wir $C^tC = \mathbf{I}$. Damit ist aber $C^t = C^{-1}$. Wir haben damit $C^tC = CC^t = \mathbf{I}$. Offenbar stehen auch die Zeilen von $C$ senkrecht aufeinander (und sind auch auf 1 normiert). Dies hat weitreichende Konsequenzen:

### 6.1 Diagonalisierung

Wir können (1.11) von links mit $C^t$ multiplizieren und erhalten:

$$C^tAC = C^tC\Lambda = \mathbf{I}\Lambda = \Lambda. \tag{1.12}$$

Wir haben $A$ *diagonalisiert.* Die geometrische Interpretation dreht sich um "Basiswechsel" und "Hauptachsentransformation". *Wir* machen dies vor allem, um uns die Rechnungen zu erleichtern. Von 6.5 haben wir, weil $C$ invertierbar ist, dass $A$ und $\Lambda$ gleichen Rang haben. Der Rang von $\Lambda$ ist denkbar einfach zu eruieren: es ist einfach die Anzahl der Diagonalelemente ungleich 0. Damit ist auch der Rang einer symmetrischen Matrix einfach zu eruieren: es ist die Anzahl der Eigenwerte, welche von 0 verschieden sind (mit Mehrfachzählungen: Eigenwerte können mehrmals auftreten (vgl. **I**)).

### 6.2 Spektralzerlegung einer Matrix

Wir können (1.11) von rechts mit $C^t$ multiplizieren und erhalten:

$$A = ACC^t = C\Lambda C^t = \sum_{i=1}^{n} \lambda_i \vec{c}_i (\vec{c}_i)^t. \tag{1.13}$$

Dies ist die Spektralzerlegung einer Matrix $A$. Wir haben $A$ als Summe von $n$ Matrizen dargestellt, welche alle Rang 1 haben.

### 6.3 Spur einer (quadratischen) Matrix

Sei $A$ eine $(n \times n)$-Matrix. Wir definieren die **Spur** [engl. Trace] als die Summe der Diagonalelemente:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Wichtige Rechenregeln für die Spur:

$\mathrm{tr}(cA) = c\,\mathrm{tr}(A)$, wenn $c$ eine reelle Zahl ist

$\mathrm{tr}(A) = \mathrm{tr}(A^t)$; $\quad \mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$; $\quad \mathrm{tr}(\mathbf{I_n}) = n$

$\mathrm{tr}(AB) = \mathrm{tr}(BA)$, wichtige Spezialfälle sind: $\vec{a}^t\vec{a} = \mathrm{tr}(\vec{a}^t\vec{a}) = \mathrm{tr}(\vec{a}\vec{a}^t)$ und im Fall einer $(n \times k)$-Matrix $A$ von (1.4) her (beachten Sie: $a_{i.}$ ist eine (liegende) Zeile)

$$\mathrm{tr}(A^tA) = \mathrm{tr}(\sum_{i=1}^{n} a_{i.}^t a_{i.}) = \sum_{i=1}^{n} \mathrm{tr}(a_{i.}^t a_{i.}) = \sum_{i=1}^{n} \mathrm{tr}(a_{i.} a_{i.}^t) = \sum_{i=1}^{n} (a_{i.} a_{i.}^t) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_{ij}^2.$$

Warnung: es gilt im Allgemeinen nicht: $\mathrm{tr}(ABC) = \mathrm{tr}(CBA)$ und Verwandte, hingegen darf man die Reihenfolge zyklisch vertauschen, z.B.:

$$\mathrm{tr}(ABCD) = \mathrm{tr}(DABC) = \mathrm{tr}(CDAB) = \mathrm{tr}(BCDA).$$

Damit gilt für symmetrische Matrizen wegen (1.13)

$$\mathrm{tr}(\Lambda) = \mathrm{tr}(\Lambda C^t C) = \mathrm{tr}(C\Lambda C^t) = \mathrm{tr}(A).$$

Damit gilt: bei einer symmetrischen Matrix ist die Spur gleich der Summe der Eigenwerte.

### 6.4 Determinanten revisited

Für symmetrische Matrizen können wir ähnlich wie oben, jetzt mit (1.12), auch Determinanten einfach berechnen:

$$\det(\Lambda) = \det(C^t AC) = \det(C^t)\det(A)\det(C) = \det(A).$$

Damit gilt: bei einer symmetrischen Matrix ist die Determinante gleich dem Produkt der Eigenwerte.

### 6.5 $A^{1'000}$

Die weiteren Resultate erhalten wir durch einfache Rechnungen. Wenn wir im Fall einer symmetrischen Matrix $A$ das "Quadrat" $A^2 := AA$ berechnen wollen, so können wir folgendes Verfahren anwenden:

$$A^2 = AA = (C\Lambda C^t)(C\Lambda C^t) = C\Lambda C^t C\Lambda C^t = C\Lambda\Lambda C^t = C\Lambda^2 C^t. \tag{1.14}$$

Da $\Lambda$ eine Diagonalmatrix ist, können wir sehr einfach $\Lambda^2$ berechnen: wir müssen nur die Diagonalelemente quadrieren. Dieses Verfahren kann man auf beliebige positive, ganzzahlige Potenzen anwenden. Sogar auf negative, ganzzahlige Potenzen (mit invertierbarem $A$):

$$A^{-1} = (C\Lambda C^t)^{-1} = (C^t)^{-1}\Lambda^{-1}(C)^{-1} = C\Lambda^{-1}C^t.$$

Damit erhalten wir zwei wichtige Resultate für invertierbare, symmetrische Matrizen $A$ mit Eigenwerten $(\lambda_i)_{i=1}^n$:

1. $A^k = C\Lambda^k C^t$, $k \in \mathbb{Z}$.

2. Die Eigenwerte von $A^k$ sind $(\lambda_i^k)_{i=1}^n$, $k \in \mathbb{Z}$, und die Eigenvektoren bleiben gleich.

### 6.6 Die "Wurzel" aus einer Matrix

Wir haben bisher nur ganzzahlige Potenzen von $A$ betrachtet. In der Statistik kommt es aber vor, dass wir für $A$ ein $B$ suchen, sodass $B^2 = A$ oder $B^2 = A^{-1}$. Dieses Problem ist mit den Resultaten von 6.5 einfach lösbar. Eine Lösung im Fall von symmetrischen Matrizen mit nichtnegativen Eigenwerten lautet

$$A^{1/2} = C\Lambda^{1/2}C^t,$$

wie leicht nachzurechnen ist. Ebenso gilt

$$A^{-1/2} = C\Lambda^{-1/2}C^t.$$