

D-CHAB und D-BIOL: Zusammenfassung Lineare Algebra

Hansruedi Künsch, D-MATH, ETHZ

Frühlingssemester 2012

1 Vorbemerkung

Dieser Text ist kein Lehrbuch, sondern ein Versuch, die wesentlichen Begriffe und Resultate sowie deren Zusammenhang kompakt zusammenzustellen. Beispiele, die man braucht, um diesen Text zu verstehen, werden in der Vorlesung gebracht und stehen im Buch von Nipp und Stoffer, Lineare Algebra, vdf. Da sich der Aufbau etwas unterscheidet von dem bei Nipp & Stoffer, werden die Bezüge jeweils am Ende eines Kapitels kurz aufgezeigt. Hingegen habe ich mich bemüht, die Dinge, die ich aus der Vektorgeometrie der Mittelschule, bzw. der Analysis-Vorlesung voraussetze, jeweils kurz in Erinnerung zu rufen.

2 Lineare Gleichungssysteme

2.1 Zwei, bzw. drei Unbekannte

Lineare Gleichungssysteme mit 2 Gleichungen und 2 Unbekannten sind aus der Mittelschule bekannt. Man löst die erste Gleichung nach x_1 auf und setzt das Resultat in die zweite Gleichung ein, oder äquivalent dazu, man subtrahiert von der zweiten Gleichung ein Vielfaches der ersten Gleichung, so dass x_1 nicht mehr vorkommt.

Im Fall von 3 Gleichungen mit 3 Unbekannten geht man gleich vor: Zuerst löst man die erste Gleichung nach x_1 auf und setzt das Ergebnis in die zweite und dritte Gleichung ein. Dann hat man 2 Gleichungen mit den Unbekannten x_2 und x_3 , und man geht vor wie oben beschrieben. Wieder kann man den ersten Schritt ersetzen dadurch, dass man von der zweiten, bzw. dritten Gleichung ein Vielfaches der ersten Gleichung subtrahiert.

Im Folgenden verallgemeinern wir dieses Verfahren für beliebige Anzahlen von Unbekannten und Gleichungen. Wir überlegen uns auch, was dabei eventuell schief gehen kann.

2.2 Der Gauss-Algorithmus allgemein

Ein lineares Gleichungssystem mit m Gleichungen und n Unbekannten hat die Form

$$\begin{array}{cccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & & & \vdots & = & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & b_m. \end{array} \tag{1}$$

Dabei sind die Koeffizienten a_{ik} und die rechten Seiten b_i vorgegebenen und die Unbekannten x_k sind gesucht. Man beachte: Der erste Index bei a_{ik} gibt die Nummer der Gleichung an, der zweite die Nummer der Unbekannten. Das Schema der Koeffizienten a_{ik} heisst die Koeffizientenmatrix des Gleichungssystems.

Der *Gaussalgorithmus* formt das vorgegebene Gleichungssystem in ein einfacheres, aber äquivalentes System um. Äquivalent soll heissen, dass die beiden Gleichungssysteme die gleiche Lösungsmenge haben (wir wollen Fälle, wo es keine oder mehrere Lösungen gibt, nicht ausschliessen). Nach der ersten Umformung soll das äquivalente System so aussehen:

$$\left(\begin{array}{cccccc} a'_{11}x_1 & + & a'_{12}x_2 & + & \dots & + & a'_{1n}x_n & = & b'_1 \\ & & a'_{22}x_2 & + & \dots & + & a'_{2n}x_n & = & b'_2 \\ & & a'_{32}x_2 & + & \dots & + & a'_{3n}x_n & = & b'_3 \\ & & \vdots & & & & \vdots & = & \vdots \\ & & a'_{m2}x_2 & + & \dots & + & a'_{mn}x_n & = & b'_m. \end{array} \right) \quad (2)$$

Man sieht sofort, dass man dieses neue System in 2 Schritten lösen kann: Wenn man eine Lösung x_2, \dots, x_n der Gleichungen 2, \dots, m hat, dann kann man x_1 aus der ersten Gleichung bestimmen und hat dann eine Lösung des ursprünglichen Systems. Wenn wir so fortfahren, dann haben erhalten wir jedes Mal eine Gleichung und eine Unbekannte weniger. Wenn $n = m$, dann haben wir am Schluss also eine Gleichung mit einer Unbekannten. Ist $n > m$, dann bleibt eine Gleichung mit $n - m$ Unbekannten, d.h. dann erwarten wir unendlich viele Lösungen. Wenn $m > n$, dann haben wir am Schluss $m - n$ Gleichungen ohne Unbekannte. Was das heisst, werden wir noch diskutieren.

Zuerst müssen wir uns noch überlegen, wie wir von (1) zu (2) kommen. Um nachher x_1 aus der ersten Gleichung zu bestimmen, darf a'_{11} nicht Null sein. Wenn x_1 in (1) tatsächlich vorkommt, können wir mit dem Vertauschen von Gleichungen erreichen, dass $a_{11} \neq 0$. a_{11} heisst dann das erste *Pivot*. Um x_1 aus den anderen Gleichungen zu eliminieren, subtrahieren wir dann von der Gleichung k das a_{k1}/a_{11} -fache der ersten Gleichung für $k = 2, 3, \dots, n$.

Wenn x_1 in (1) nicht vorkommt, dann können wir x_1 beliebig wählen, x_1 ist dann eine sogenannte freie Unbekannte. Wenn wir mit dem Gauss-Algorithmus weiterfahren, dann sehen wir, dass es wieder darauf ankommt, ob x_2 in den Gleichungen 2, 3, \dots, m von (2) überhaupt vorkommt, d.h. ob einer der Koeffizienten $a'_{22}, a'_{32}, \dots, a'_{m2}$ von Null verschieden ist. Wenn ja, dann erreichen wir mit dem Vertauschen von Gleichungen, dass $a'_{22} \neq 0$. a'_{22} heisst dann das zweite *Pivot*. Anderenfalls ist x_2 eine freie Variable, es bleiben $m - 1$ Gleichungen für $n - 2$ Unbekannte. Wenn wir in jedem Schritt ein *Pivot* finden, dann sprechen wir vom regulären Fall, sonst vom Ausnahme- oder singulären Fall.

Zur Vereinfachung der Notation führen wir im Folgenden die Eliminationschritte durch auf der *Matrix* der Koeffizienten (a_{ik}), augmentiert durch die Werte (b_i) der rechten Seite, d.h. man lässt alle x_k , “+” und “=” weg. Ausserdem überschreiben wir die Elemente der Matrix jeweils mit den neuen Werten, d.h. die Werte von a_{ij} und b_i ändern sich im Verlaufe des Verfahrens.

2.3 Der reguläre Fall

Wenn $m = n$ und wir für jede Unbekannte ein Pivot finden, dann hat die augmentierte Matrix am Schluss die Dreiecksform

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ 0 & a_{22} & \dots & a_{2n} & b_2 \\ 0 & 0 & a_{33} & \dots & a_{3n} & b_3 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & & \dots & 0 & a_{nn} & b_n \end{pmatrix}$$

wobei die Pivots $a_{11}, a_{22}, \dots, a_{nn}$ alle $\neq 0$ sind. Dann gibt es für beliebige rechte Seiten genau eine Lösung, welche man durch *Rückwärtseinsetzen* findet: Die letzte Gleichung liefert direkt x_n , Einsetzen in die zweitletzte Gleichung liefert dann x_{n-1} und so weiter.

Wenn $m < n$ ist, dann haben wir – wieder unter der Annahme, dass wir in jedem Schritt ein Pivot finden – nach $m-1$ Eliminationsschritten als letzte Gleichung $a_{mm}x_m + \dots + a_{mn}x_n = b_m$. Im regulären Fall ist auch das m -te Pivot $a_{mm} \neq 0$. Die Unbekannten x_{m+1}, \dots, x_n können beliebig gewählt werden, sind also freie Unbekannte. Die übrigen Unbekannten x_m, \dots, x_1 sind dann durch Rückwärtseinsetzen eindeutig bestimmt. Es existieren also stets unendlich viele Lösungen, und zwar haben wir $n - m$ freie Parameter.

Wenn $m > n$, dann haben wir – immer noch unter der Annahme, dass wir in jedem Schritt ein Pivot finden – nach n Schritten folgendes Bild

$$\begin{pmatrix} a_{11} & \dots & & a_{1n} & b_1 \\ 0 & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ & \vdots & \ddots & \ddots & \\ 0 & 0 & \dots & 0 & a_{nn} & b_n \\ 0 & 0 & \dots & 0 & 0 & b_{n+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & b_m \end{pmatrix}$$

wobei die Pivots $a_{11}, a_{22}, \dots, a_{nn}$ alle $\neq 0$ sind. Jede der letzten $m - n$ Gleichungen lautet $0 = b_i$, was entweder automatisch richtig ist oder einen Widerspruch darstellt. Wenn daher einer oder mehrere der Werte b_{n+1}, \dots, b_m im Endschema $\neq 0$ sind, dann hat das Gleichungssystem keine Lösung. Falls aber $b_{n+1} = \dots = b_m = 0$, dann kann man die letzten $m - n$ Gleichungen weglassen, und es gibt genau eine Lösung, die man durch Rückwärtseinsetzen bestimmen kann.

2.4 Der Ausnahmefall: Zeilenstufenform

Wir betrachten jetzt also den Fall, wo wir für eine Unbekannte kein Pivot finden, und wir nehmen an, dass das zum ersten Mal für die Variable x_j , d.h. nach dem $j - 1$ -ten Eliminationsschritt auftritt. Es gilt also $a_{jj} = a_{j+1,j} = \dots = a_{n,j} = 0$, damit ist x_j eine freie Unbekannte. Dann suchen wir einfach die nächste Unbekannte, für die ein Pivot

existiert und die also nicht frei ist. Falls dies x_k ist, dann sieht das Schema wie folgt aus:

$$\left(\begin{array}{cccccccccc} a_{11} & \dots & & & & & & & a_{1n} & b_1 \\ 0 & a_{22} & & & & & & & a_{2n} & b_2 \\ \vdots & \ddots & \ddots & & & & & & \vdots & \vdots \\ 0 & \dots & & a_{j-1,j-1} & \dots & & & & a_{j-1,n} & b_{j-1} \\ 0 & \dots & & 0 & 0 & \dots & 0 & a_{jk} & \dots & a_{jn} & b_j \\ \vdots & & & \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots \\ 0 & \dots & & 0 & 0 & \dots & 0 & a_{mk} & \dots & a_{mn} & b_m \end{array} \right)$$

wobei nach Voraussetzung $a_{11}, a_{22}, \dots, a_{j-1,j-1}$ alle $\neq 0$ sind und $k > j$. Ferner ist – eventuell nach einer Vertauschung von Gleichungen – auch $a_{jk} \neq 0$. Die Gleichungen $j, j+1, \dots, m$ stellen dann ein Gleichungssystem für x_k, x_{k+1}, \dots, x_n dar, wobei x_k darin auch tatsächlich vorkommt. Wenn x_k, \dots, x_n eine Lösung davon ist, dann erhalten wir eine Lösung des ursprünglichen Systems, indem wir x_j, \dots, x_{k-1} beliebig wählen und x_1, \dots, x_{j-1} durch Rückwärtseinsetzen bestimmen.

Wir setzen also das Eliminationsverfahren auf diesem kleineren System fort. Am Schluss erhalten wir dann die sogenannte *Zeilenstufenform*: Jede Gleichung enthält weniger Unbekannte als die vorangehende. Eine Unbekannte, die als erste in einer Gleichung steht, heisst eine Pivot-Unbekannte, und der zugehörige Koeffizient ist ein Pivot. Die anderen Unbekannten sind freie Unbekannte. Die Anzahl der vorkommenden Pivots nennen wir den *Rang des Gleichungssystems*, den wir mit r bezeichnen. Offensichtlich ist $r \leq m$ und $r \leq n$. Im vorherigen Unterabschnitt hatten wir den Fall $r = \min(m, n)$ besprochen.

Aus der Zeilenstufenform erhalten wir den vollständigen Überblick darüber, wie viele Lösungen es gibt, und wie wir alle Lösungen bestimmen können. Ist $r < m$, so haben wir am Schluss $m - r$ Gleichungen, die – je nach dem Wert der b_i 's – entweder einen Widerspruch beinhalten oder weggelassen werden können. Will man, dass für beliebige b_i 's eine Lösung existiert, dann muss $r = m$ sein. Wenn $r < n$, dann gibt es freie Unbekannte, d.h. wenn es eine Lösung gibt, gibt es sogar unendlich viele. Eine eindeutige Lösung existiert für beliebige rechte Seiten genau dann, wenn $r = m = n$.

2.5 Zusammenhang mit Nipp & Stoffer

Dieses Kapitel entspricht 1.1. und 1.2 in jenem Buch.

3 Vektoren und Matrizen

3.1 Vektoren im \mathbb{R}^2 und \mathbb{R}^3

Paaren von reellen Zahlen entsprechen bekanntlich Punkte in der Ebene, bzw. Äquivalenzklassen von “Pfeilen” (gerichtete Strecken, die durch Parallelverschiebung zur Deckung gebracht werden können). Analog ist es bei Trippeln und Punkten, bzw. “Pfeilen” im Raum. Komponentenweise Multiplikation mit einem Skalar entspricht Streckung/Stauchung der Pfeile, Komponentenweise Addition der Zusammensetzung gemäss der Parallelogrammregel.

3.2 Vektoren und Matrizen allgemein

Ein n -Vektor ist ein n -Tupel von Zahlen, die untereinander geschrieben werden. Der Nullvektor besteht aus n Nullen und wird mit 0 bezeichnet. Eine $m \times n$ Matrix ist ein rechteckiges Schema von $m \cdot n$ Zahlen, angeordnet in m Zeilen und n Spalten (Kolonnen). Diese Zahlen heissen die *Elemente* der Matrix. Die Position eines Elements wird durch zwei Indizes festgelegt, wie bei der Koeffizientenmatrix bei einem linearen Gleichungssystem: Der erste Index gibt die Nummer der Zeile an und der zweite die Nummer der Spalte. Ein Vektor ist also eine $m \times 1$ Matrix. Manchmal betrachtet man auch $1 \times n$ Matrizen als Vektoren und spricht dann zur Unterscheidung der beiden Möglichkeiten von Spalten-, bzw. Zeilenvektoren. Wenn nichts anderes gesagt ist, sind Vektoren für uns immer Spaltenvektoren. Die Menge aller Spaltenvektoren mit n Elementen wird als \mathbb{R}^n bezeichnet und die Menge aller Matrizen mit m Zeilen und n Spalten als $\mathbb{R}^{m \times n}$.

Die Addition zweier Matrizen (Vektoren) und die Multiplikation einer Matrix (Vektors) mit einem Skalar sind komponentenweise definiert. Es gelten die gleichen Regeln wie beim Rechnen mit Zahlen. Eine weitere einfache Operation bei Matrizen ist das *Transponieren*: Die *Transponierte* A^T einer Matrix A entsteht durch Spiegelung an der Diagonalen, d.h. die Zeilen von A^T sind die Spalten von A : $(A^T)_{ij} = A_{ji}$. Es gilt $(A^T)^T = A$. Ist b ein Spaltenvektor, so ist b^T ein Zeilenvektor.

Das Produkt einer $m \times n$ Matrix A mit einem $n \times 1$ Spaltenvektor x ist ein $m \times 1$ Vektor mit den Komponenten

$$(Ax)_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = \sum_{j=1}^n a_{ij}x_j \quad (i = 1, 2, \dots, m).$$

Es ist wichtig, dass die Matrix A links steht und der Vektor rechts, sonst gibt es später ein Durcheinander bei der Multiplikation zweier Matrizen. Kurz gesagt ist die i -te Komponente von Ax das "Skalarprodukt" des i -ten Zeilenvektors von A mit dem Vektor x .

Ein lineares Gleichungssystem lässt sich daher in der Kurzform $Ax = b$ schreiben. Ferner ist Ax nichts anderes als die Linearkombination x_1 mal erster Spaltenvektor von A plus x_2 mal zweiter Spaltenvektor von A plus usw. plus x_n mal letzter Spaltenvektor von A :

$$Ax = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Für das Folgende sind die beiden Regeln

$$A(x + y) = Ax + Ay, \quad A(\lambda x) = \lambda \cdot Ax \quad (A \in \mathbb{R}^{m \times n}, x, y \in \mathbb{R}^n, \lambda \in \mathbb{R}) \quad (3)$$

absolut zentral. Der Beweis ist einfach, man schreibt die Ausdrücke in Komponenten hin. Wir sagen, die Multiplikation Vektor mal Matrix ist linear. Damit kann man insbesondere zeigen, dass man alle Lösungen von $Ax = b$ bekommt, wenn man zu irgendeiner Lösung von $Ax = b$ alle Lösungen von $Ax = 0$ addiert. In Worten: Die allgemeine Lösung des inhomogenen Systems (d.h. für $b \neq 0$) ist eine spezielle Lösung des inhomogenen Systems plus die allgemeine Lösung des homogenen Systems (d.h. für $b = 0$)

3.3 Geometrische Interpretation von linearen Gleichungssystemen mit $m \leq 3, n \leq 3$: Spaltensichtweise

Für $n = m = 2$ ist $Ax = x_1a^{(1)} + x_2a^{(2)}$. Lösen von $Ax = b$ ist also das Gleiche wie b darstellen als Linearkombination der beiden Spaltenvektoren. Geometrisch sieht man sofort, dass das für jedes b auf genau eine Art geht, ausser wenn die beiden Spaltenvektoren kollinear sind. Für $m = 2$ und $n = 3$ gibt eine Lösung von $Ax = b$ eine Darstellung von b mit 3 Spaltenvektoren: Ausser wenn alle 3 kollinear sind, geht das für beliebiges b auf unendlich viele Arten.

Für $n = m = 3$ entspricht ein Lösung von $Ax = b$ einer Darstellung von b als Linearkombination der 3 Spaltenvektoren. Dies geht für jedes b auf genau eine Art, sofern die drei Spaltenvektoren nicht koplanar sind. Für $m = 3$ und $n = 2$ spannen die beiden Spaltenvektoren eine Ebene durch den Ursprung auf, sofern sie nicht kollinear sind: $Ax = b$ ist dann nur lösbar, wenn b in dieser Ebene liegt.

3.4 Lineare Hülle und lineare Unabhängigkeit

Wir verallgemeinern nun die geometrische Interpretation des vorherigen Unterabschnitts für beliebige m und n . Die *lineare Hülle* U von vorgegebenen Vektoren $a^{(1)}, a^{(2)}, \dots, a^{(n)} \in \mathbb{R}^m$ besteht aus allen Linearkombinationen der Vektoren $a^{(i)}$

$$U = \{y \in \mathbb{R}^m; y = x_1a^{(1)} + x_2a^{(2)} + \dots + x_na^{(n)} \text{ mit } x_1, x_2, \dots, x_n \in \mathbb{R}\}.$$

Wir schreiben dann auch $U = \text{span}\{a^{(1)}, a^{(2)}, \dots, a^{(n)}\}$. Lineare Hüllen sind höherdimensionale Verallgemeinerungen von Geraden und Ebenen durch den Ursprung.

Mit Hilfe der Matrix A , deren Spaltenvektoren gerade die Vektoren $a^{(i)}$ sind, können wir auch schreiben

$$\text{span}\{a^{(1)}, a^{(2)}, \dots, a^{(n)}\} = \{y; y = Ax \text{ mit } x \in \mathbb{R}^n\}.$$

$Ax = b$ ist also genau dann lösbar, wenn b in der linearen Hülle der Spaltenvektoren von A , dem sogenannten Spaltenraum von A liegt.

Eine lineare Hülle U von Vektoren hat die Eigenschaften

$$y \in U, \lambda \in \mathbb{R} \Rightarrow \lambda y \in U, \quad y, z \in U \Rightarrow y + z \in U.$$

Jede (nichtleere) Teilmenge des \mathbb{R}^m mit diesen Eigenschaften heisst ein *Unterraum*. Man kann zeigen, dass jeder Unterraum des \mathbb{R}^m ein Erzeugendensystem hat, d.h. es gibt Vektoren $a^{(1)}, a^{(2)}, \dots, a^{(n)}$, so dass $U = \text{span}\{a^{(1)}, a^{(2)}, \dots, a^{(n)}\}$.

Der gleiche Unterraum U hat jeweils viele unterschiedliche Erzeugendensysteme. Von Interesse sind solche, die möglichst wenige Vektoren enthalten. Wann kann man einen Vektor $a^{(j)}$ in einem Erzeugendensystem $(a^{(1)}, a^{(2)}, \dots, a^{(n)})$ weglassen, ohne dass sich die lineare Hülle verkleinert? Offensichtlich genau dann, wenn man $a^{(j)}$ als Linearkombination der übrigen $k - 1$ Vektoren darstellen kann:

$$a^{(j)} = x_1a^{(1)} + x_2a^{(2)} + \dots + x_{j-1}a^{(j-1)} + x_{j+1}a^{(j+1)} + \dots + x_na^{(n)}.$$

Nimmt man $a^{(j)}$ auf die rechte Seite, dann steht eine Darstellung des Nullvektors als Linearkombination der $a^{(i)}$'s da mit Koeffizienten, die nicht alle gleich Null sind. Man sieht leicht, dass eine solche Darstellung sogar äquivalent ist dazu, dass das Erzeugendensystem

reduziert werden kann. Wir nennen daher $a^{(1)}, a^{(2)}, \dots, a^{(n)}$ *linear abhängig*, wenn es Zahlen x_1, x_2, \dots, x_n gibt, die nicht alle gleich null sind, so dass

$$x_1 a^{(1)} + x_2 a^{(2)} + \dots + x_n a^{(n)} = 0. \quad (4)$$

Hat (4) nur die triviale Lösung $x_1 = x_2 = \dots = x_n = 0$, dann heissen $a^{(1)}, a^{(2)}, \dots, a^{(n)}$ *linear unabhängig*. 2 Vektoren sind genau dann linear abhängig, wenn sie kollinear sind. 3 Vektoren sind genau dann linear abhängig, wenn sie koplanar sind.

In einem Erzeugendensystem eines Unterraums U bestehend aus linear unabhängigen können wir also keine Vektoren weglassen. Ein solches Erzeugendensystem von U heisst eine Basis von U . Ein Unterraum hat viele verschiedene Basen, aber alle Basen bestehen aus gleich vielen Vektoren. Diese Anzahl heisst die Dimension von U . Ein eindimensionaler Unterraum besteht aus allen Vektoren in einer festen Richtung, ist also anschaulich gesprochen eine Gerade im \mathbb{R}^m durch den Ursprung. Ein zweidimensionaler Unterraum besteht aus allen Linearkombinationen von zwei Vektoren mit verschiedener Richtung, ist also anschaulich gesprochen eine Ebene im \mathbb{R}^m durch den Ursprung. Unterräume höherer Dimension nennt man dann "Hyperebenen" durch den Ursprung. Der einzige m -dimensionale Unterraum von \mathbb{R}^m ist \mathbb{R}^m selber.

Da (4) auch als $Ax = 0$ geschrieben werden kann, können wir den Gauss-Algorithmus benutzen, um lineare Abhängigkeit zu prüfen, sofern wir dies den Vektoren nicht direkt ansehen: $a^{(1)}, a^{(2)}, \dots, a^{(n)} \in \mathbb{R}^m$ sind genau dann linear unabhängig, falls $Ax = 0$ nur die triviale Lösung $x = 0$ hat, also nach früheren Resultaten falls $\text{Rang}(A) = n$. Weil $\text{Rang}(A) \leq m$, sind mehr als m Vektoren im \mathbb{R}^m also immer linear abhängig. Der Gauss-Algorithmus gibt sogar noch genauere Information: Ist $\text{Rang}(A) = r < n$ und haben die Pivot-Unbekannten die Indizes i_1, i_2, \dots, i_r , dann ist $a^{(i_1)}, a^{(i_2)}, \dots, a^{(i_r)}$ eine Basis des Spaltenraums. Die Dimension des Spaltenraums von A ist also gleich dem Rang von A .

3.5 Koordinaten, Basistransformation

Ein Unterraum hat wie gesagt viele verschiedene Basen, die aber alle aus gleich vielen Vektoren bestehen. Wenn U die Dimension k hat und $u^{(1)}, u^{(2)}, \dots, u^{(k)}$ eine Basis von U ist, dann lässt sich jedes $y \in U$ auf genau eine Art als Linearkombination der $u^{(j)}$ darstellen:

$$y = c_1 u^{(1)} + c_2 u^{(2)} + \dots + c_k u^{(k)} = Uc.$$

Die Koeffizienten c_j heissen die Koordinaten von y bezüglich der Basis $u^{(1)}, u^{(2)}, \dots, u^{(k)}$.

Die einfachste Basis des \mathbb{R}^m ist die sogenannte Standardbasis, die aus den Vektoren $e^{(j)}$ ($j = 1, 2, \dots, m$) besteht, welche an der j -ten Stelle eine 1 und sonst lauter Nullen haben. Jedes $y \in \mathbb{R}^m$ hat offensichtlich die Darstellung

$$y = y_1 e^{(1)} + y_2 e^{(2)} + \dots + y_m e^{(m)},$$

das heisst die Komponenten von y sind gerade die Koordinaten von y in der Standardbasis. Für manche Probleme ist es aber von Vorteil, eine andere Basis zu verwenden, d.h. man stellt alle Vektoren dar als Linearkombination von $u^{(1)}, u^{(2)}, \dots, u^{(m)}$ statt als Linearkombination von $e^{(1)}, e^{(2)}, \dots, e^{(m)}$. Die Koordinaten bezüglich der neuen Basis berechnen sich dann als Lösung von $Uc = y$.

3.6 Geometrische Interpretation von linearen Gleichungssystemen mit $m \leq 3, n \leq 3$: Zeilensichtweise

Eine lineare Gleichung mit 2 Unbekannten stellt bekanntlich eine Gerade in der Ebene dar, deren Normalenvektor durch die Koeffizienten gegeben ist. Die Lösungsmenge eines linearen Gleichungssystems mit m Gleichungen und 2 Unbekannten ist daher der Durchschnitt von m Geraden, und man kann sich geometrisch leicht vorstellen, wie die Geraden liegen müssen, damit die Lösungsmenge ein Punkt, eine Gerade oder die leere Menge ist. Insbesondere haben zwei Geraden genau einen Schnittpunkt, wenn die beiden Normalenvektoren nicht kollinear sind. Wenn die m Geraden alle den Ursprung enthalten, d.h. wenn die rechten Seiten des Gleichungssystems alle gleich Null sind, dann ist der Durchschnitt entweder eine Gerade oder er besteht nur aus dem Ursprung selber. Der erste Fall tritt dann ein, wenn alle Normalenvektoren kollinear sind, der zweite Fall dann, wenn es zwei linear unabhängige Normalenvektoren gibt.

Analog stellt eine lineare Gleichung mit 3 Unbekannten eine Ebene im Raum dar, deren Normalenvektor durch die Koeffizienten gegeben ist. Die Lösungsmenge eines linearen Gleichungssystems mit 3 Unbekannten und m Gleichungen ist also gleich dem Durchschnitt von m Ebenen und damit handelt es sich um eine Ebene, eine Gerade, einen Punkt oder die leere Menge. Die geometrische Anschauung hilft einem wieder zu verstehen, wann welcher Fall zutrifft. Insbesondere haben drei Ebenen genau einen Schnittpunkt, wenn die drei Normalenvektoren nicht koplanar sind. Wenn die m Ebenen alle den Ursprung enthalten, d.h. wenn die rechten Seiten des Gleichungssystems alle gleich Null sind, dann ist der Durchschnitt entweder eine Ebene oder eine Gerade oder er besteht nur aus dem Ursprung selber. Eine Ebene erhält man, wenn alle Normalenvektoren kollinear sind, eine Gerade, wenn die Normalenvektoren alle in einer Ebene liegen, und einen Punkt, wenn es drei linear unabhängige Normalenvektoren gibt.

3.7 Der Kern von A

Eine lineare Gleichung $a_1x_1 + a_2x_2 + \dots + a_nx_n = b$, bei der nicht alle $a_i = 0$ sind, stellt eine “ $(n - 1)$ -dimensionalen Hyperebene” im \mathbb{R}^n dar: Wenn zum Beispiel $a_1 \neq 0$, dann hat diese Gleichung nämlich die allgemeine Lösung

$$x = \begin{pmatrix} -b/a_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -a_2/a_1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + t_n \begin{pmatrix} -a_n/a_1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

mit beliebigen $t_i \in \mathbb{R}$. Die Lösungsmenge eines Gleichungssystems $Ax = b$ mit $A \in \mathbb{R}^{m \times n}$ kann daher als Durchschnitt von m “ $(n - 1)$ -dimensionalen Hyperebenen” im \mathbb{R}^n aufgefasst werden. Um einen Überblick zu gewinnen, wie ein solcher Durchschnitt aussieht, beschränken wir uns auf den Fall $b = 0$, d.h. wir studieren die Menge $\{x \in \mathbb{R}^n; Ax = 0\}$, den sogenannten *Kern von A* . Für ein beliebiges b ist die Lösungsmenge entweder leer, oder der Kern von A wird um einen festen Vektor verschoben.

Als erstes beachten wir, dass der $\text{Kern}(A)$ ein Unterraum des \mathbb{R}^n ist, d.h. wenn x und y im Kern sind, dann auch $x + y$ und λx . Wir konstruieren eine Basis von $\text{Kern}(A)$ und bestimmen damit die Dimension des Kerns. Dazu verwenden wir wieder den Gauss-Algorithmus: Ist der Rang r von A gleich n , dann hat $Ax = 0$ nur die Lösung $x =$

0. Anderenfalls ist eine Lösung von $Ax = 0$ eindeutig bestimmt, wenn wir die Werte $x_{i_1}, x_{i_2}, \dots, x_{i_{n-r}}$ vorgeben, wobei i_1, i_2, \dots, i_{n-r} die Indizes der freien Unbekannten (das heisst derjenigen x_k , die nicht zu einem Pivot gehören) bezeichnen. Die $n - r$ Vektoren, bei denen genau eine der freien Unbekannten gleich 1 und die anderen freien Unbekannten gleich 0 ist, erzeugen daher $\text{Kern}(A)$. Da sie auch linear unabhängig sind, bilden sie eine Basis von $\text{Kern}(A)$. Die Dimension von $\text{Kern}(A)$ ist gleich der Anzahl Spalten minus Rang, bzw. gleich Anzahl Spalten minus Dimension des Spaltenraums. $\text{Kern}(A)$ ist also eine “ $(n - r)$ -dimensionale Hyperebene” durch den Ursprung. Ist $r = m$, dann ist der Durchschnitt von m “ $(n - 1)$ -dimensionalen Hyperebenen” eine “ $(n - m)$ -dimensionale Hyperebene”.

Beim Gaußalgorithmus ändert sich der Unterraum, der von den Zeilen von A aufgespannt ist, nicht, weil wir ja nur andere Linearkombinationen bilden. In der Zeilenstufenform sind die ersten r Zeilenvektoren offensichtlich linear unabhängig, und die letzten $m - r$ Zeilenvektoren sind 0 und können weggelassen werden. Ds heisst r ist nicht nur die Dimension des Spalten-, sondern auch des Zeilenraums. Da der Zeilenraum von A gleich dem Spaltenraum von A^T ist, folgt damit $\text{Rang}(A) = \text{Rang}(A^T)$.

3.8 Zusammenhang mit Nipp & Stoffer

Der in diesem Kapitel behandelte Stoff findet sich bei Nipp & Stoffer in den Kapitel 2.1, 2.2, 4.1, 4.2 und 6.1. Die wesentlichen Unterschiede sind, dass ich die Multiplikation von Matrizen erst später bei den linearen Abbildungen einführe und dass ich den Begriff eines abstrakten Vektorraums weglasse. Die wichtigsten Resultate hier findet man bei Nipp & Stoffer in Satz 2.2, Satz 2.4 (ohne iii)), und auf den Seiten 77-82, sowie 122-124.

4 Längen, Winkel und Volumen im \mathbb{R}^n

4.1 Längen und Winkel im \mathbb{R}^3

Im Raum definiert man das Skalarprodukt (x, y) zweier Vektoren x und y als $\|x\| \cdot \|y\| \cdot \cos(\phi)$, wobei $\|x\|$ die Länge $\sqrt{x_1^2 + x_2^2 + x_3^2}$ bezeichnet und ϕ den Zwischenwinkel. Man zeigt dann, dass dieses Skalarprodukt linear ist

$$(x, y + z) = (x, y) + (x, z), \quad (x, \lambda y) = \lambda(x, y)$$

und folgert daraus, dass

$$(x, y) = x_1y_1 + x_2y_2 + x_3y_3.$$

Damit kann man Winkel algebraisch berechnen, und insbesondere hat man die Bedingung $(x, y) = 0$ für die Orthogonalität von x und y .

4.2 Längen und Winkel allgemein

Da wir im \mathbb{R}^n Winkel und Längen nicht geometrisch bestimmen können, gehen wir gerade umgekehrt vor. Wir definieren das sogenannte Euklid'sche Skalarprodukt durch die Formel

$$(x, y) = \sum_{i=1}^n x_i y_i = x^T y$$

und definieren dann Winkel und Längen in Analogie zum \mathbb{R}^3 :

$$\|x\| = \sqrt{(x, x)}, \quad \cos(\phi) = \frac{(x, y)}{\|x\| \cdot \|y\|}.$$

Ist diese Definition sinnvoll? Zunächst einmal muss man sicher stellen, dass die rechte Seite bei der Definition von $\cos(\phi)$ zwischen -1 und 1 liegt. Ferner möchte man, dass einige Eigenschaften von Längen und Winkeln, die im \mathbb{R}^3 offensichtlich sind, auch in beliebigen Dimensionen gelten.

Als erstes untersuchen wir den Zusammenhang zwischen Orthogonalprojektion und Minimierung des Abstandes: Seien x und y zwei Vektoren $\neq 0$ im \mathbb{R}^n . Die orthogonale Projektion von x auf die von y aufgespannte Gerade durch den Ursprung ist derjenige Vektor λy , für den gilt $(x - \lambda y, y) = 0$. Mit der Linearität des Skalarprodukts folgt sofort $\lambda = (x, y)/(y, y)$. Andererseits ist der quadrierte Abstand von x von einem beliebigen Punkt λy dieser Geraden wieder wegen der Linearität und der Symmetrie des Skalarprodukts gleich

$$\|x - \lambda y\|^2 = (x - \lambda y, x - \lambda y) = (x, x) - 2\lambda(x, y) + \lambda^2(y, y).$$

Ableiten nach λ und Null setzen ergibt die Bedingung $\lambda = (x, y)/(y, y)$ für ein Minimum. Die Orthogonalprojektion minimiert also tatsächlich den Abstand. Wir erhalten daraus aber noch mehr Information:

$$0 \leq \min_{\lambda} \|y - \lambda x\|^2 = (x, x) - 2\frac{(x, y)^2}{(y, y)} + \frac{(x, y)^2}{(y, y)} = (x, x) - \frac{(x, y)^2}{(y, y)}.$$

Daher gilt die sogenannte Schwarz'sche Ungleichung

$$(x, y)^2 \leq (x, x)(y, y), \quad \text{bzw. } |(x, y)| \leq \|x\| \cdot \|y\|,$$

und wir haben insbesondere gezeigt, dass obige Definition des Winkels sinnvoll ist. Eine weitere Folgerung aus der Schwarz'schen Ungleichung lautet.

$$\|x + y\|^2 = (x + y, x + y) = \|x\|^2 + \|y\|^2 + 2(x, y) \leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| = (\|x\| + \|y\|)^2.$$

Nach Ziehen der Wurzel steht die sogenannte Dreiecksungleichung da $\|x + y\| \leq \|x\| + \|y\|$.

Es zeigt sich aber, dass es noch andere vernünftige Arten gibt, Längen (auch Normen genannt) und Winkel zu definieren. Die beiden wichtigsten anderen Längen sind die sogenannten L_1 - und die L_∞ -Normen:

$$\|x\|_1 = \sum_{i=1}^n |x_i|, \quad \|x\|_\infty = \max(|x_1|, |x_2|, \dots, |x_n|).$$

Sie erfüllen ebenfalls die Eigenschaften

$$\|x\| > 0 \text{ falls } x \neq 0, \quad \|\lambda x\| = |\lambda| \|x\|, \quad \|x + y\| \leq \|x\| + \|y\|,$$

welche als Axiome gefordert werden. Die Axiome für ein Skalarprodukt lauten

$$(x, y + z) = (x, y) + (x, z), \quad (x, \lambda y) = \lambda(x, y), \quad (x, y) = (y, x), \quad (x, x) > 0 \text{ falls } x \neq 0.$$

Wir verzichten darauf, andere Skalarprodukte als das oben definierte Euklid'sche Skalarprodukt anzugeben. Zu jedem Skalarprodukt gehört die Norm $\sqrt{(x, x)}$, aber nicht jede Norm

wird durch ein Skalarprodukt erzeugt: Zum Beispiel sind gibt es kein Skalarprodukt hinter der L_1 - oder der L_∞ -Norm.

Paarweise orthogonale Vektoren $u^{(1)}, \dots, u^{(k)}$ (ungleich dem Nullvektor) sind automatisch linear unabhängig. Zudem können wir in diesem Fall die Koordinaten (c_1, \dots, c_k) eines Vektors $x \in \text{span}\{u^{(1)}, \dots, u^{(k)}\}$ leicht berechnen: Multipliziert man nämlich $x = \sum_{i=1}^k c_i u^{(i)}$ skalar mit $u^{(j)}$, so folgt

$$(x, u^{(j)}) = \sum_{i=1}^k c_i (u^{(i)}, u^{(j)}) = c_j (u^{(j)}, u^{(j)}) \Rightarrow c_j = \frac{(x, u^{(j)})}{(u^{(j)}, u^{(j)})}.$$

Sind die $u^{(j)}$ auch noch auf Länge 1 normiert, so kann man kurz schreiben $c = U^T x$.

4.3 Das Vektorprodukt im \mathbb{R}^3

Das Vektorprodukt $x \times y$ zweier Vektoren im \mathbb{R}^3 ist der durch die folgenden Forderungen eindeutig bestimmte Vektor im \mathbb{R}^3 : i) $x \times y$ steht senkrecht auf x und auf y ,

ii) $\|x \times y\| = \|x\| \cdot \|y\| \sin(\phi) = \text{Fläche des Parallelogramms aufgespannt von } x \text{ und } y$ (ϕ ist der Zwischenwinkel der beiden Vektoren), iii) $x, y, x \times y$ bilden ein Rechtssystem.

Das Vektorprodukt hat die folgenden Eigenschaften

$$y \times x = -x \times y, \quad x \times (\lambda y) = \lambda x \times y, \quad x \times (y + z) = x \times y + x \times z.$$

Daraus folgt, dass man $x \times y$ wie folgt berechnen kann:

$$(x_1, x_2, x_3)^T \times (y_1, y_2, y_3)^T = (x_2 y_3 - x_3 y_2, x_3 y_1 - x_1 y_3, x_1 y_2 - x_2 y_1)^T.$$

4.4 Determinanten

Die Determinante ist eine für quadratische $n \times n$ Matrizen definierte Funktion, mit den folgenden zwei Hauptanwendungen: i) $\det(A) = 0$ genau dann, wenn die Spalten von A linear unabhängig sind, ii) die Determinante wird für Volumenberechnungen definiert.

Die Determinante einer 2×2 Matrix ist definiert als

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11} a_{22} - a_{12} a_{21}.$$

Die Determinante ist also nichts anderes als die dritte Komponente von $(a_{11}, a_{21}, 0)^T \times (a_{12}, a_{22}, 0)^T$. Der Absolutbetrag der Determinante ist daher die Fläche des Parallelogramms aufgespannt von den beiden Spaltenvektoren, und das Vorzeichen ist durch die Orientierung der beiden Spaltenvektoren bestimmt.

Für $n = 3$ ist

$$\det(A) = (a^{(1)}, a^{(2)} \times a^{(3)}) = a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{21} \det \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} + a_{31} \det \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}.$$

Der Absolutbetrag der Determinante ist daher das Volumen des Körpers (genannt Spat oder Parallelepipid), der von den drei Spaltenvektoren aufgespannt wird, denn $\|a^{(1)}\|$ mal der Kosinus des Winkels zwischen $a^{(1)}$ und $a^{(2)} \times a^{(3)}$ ist gerade die Höhe dieses Körpers. Das Vorzeichen der Determinante hängt wieder von der Orientierung der drei Spaltenvektoren ab. Für eine schnelle Berechnung im Fall $n = 3$ schreibt man die ersten

beiden Spalten nochmals hinter die dritte und multipliziert jeweils drei Zahlen entlang von Geraden mit Steigung ± 1 , addiert die drei Produkte entlang Geraden mit Steigung -1 und subtrahiert die andern drei Produkte:

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

Für $n > 3$ ist die Determinante rekursiv definiert durch Entwicklung nach der ersten Spalte, analog wie in der obigen Formel für $n = 3$. Wichtiger als die Definition sind jedoch die folgenden Eigenschaften der Determinante: Man kann leicht nachprüfen, dass die Determinante von 2×2 und 3×3 Matrizen den folgenden Regeln genügt

1. Bei der Vertauschung zweier Spalten ändert die Determinante das Vorzeichen. Wenn zwei Spalten gleich sind, ist die Determinante daher null.
2. Multipliziert man eine Spalte mit einer Konstanten λ , dann multipliziert sich die Determinante ebenfalls mit λ .
3. Addiert man zu einer Spalte ein Vielfaches einer anderen Spalte, so ändert sich die Determinante nicht.
4. Die Determinante einer Dreiecksmatrix, d.h. einer Matrix, bei der die Elemente unterhalb der Diagonalen alle gleich Null sind, ist gleich dem Produkt der Diagonalelemente.
5. Es gilt $\det(A^T) = \det(A)$. Damit gelten die ersten 3 Regeln auch für Zeilen statt Spalten.
6. $\det(A) = 0$ genau dann, wenn $\text{Rang}(A) < n$.

Mit Induktion kann man zeigen, dass alle Eigenschaften auch für $n > 3$ gelten, die Beweise sind aber relativ kompliziert. Die Eigenschaften 2, 3, und 4 entsprechen auch intuitiven Anforderungen an einen Volumenbegriff. Daher sagt man, dass $|\det(A)|$ das Volumen des Körpers aufgespannt von den Spaltenvektoren von A ist.

Aus diesen Eigenschaften folgt insbesondere, dass man eine Determinante mit dem Gauss-Algorithmus berechnen kann, denn bei jedem Eliminationsschritt ändert sich höchstens das Vorzeichen, und am Ende hat man eine Dreiecksmatrix. Es gibt auch eine Formel für Determinanten, aber diese besteht aus einer Summe von $n!$ Termen, von denen jeder ein Produkt von n Matrixelementen ist. Diese Formel ist daher nicht geeignet für Berechnungen.

4.5 Zusammenhang mit Nipp & Stoffer

Der Stoff dieses Kapitels ist bei Nipp & Stoffer in den Kapiteln 4.3, 4.4, sowie 3.1, 3.2, 3.3. Die Teile über Normen und Skalarprodukte in anderen Vektorräumen als \mathbb{R}^n kommen bei mir nicht vor, die Regel für die Determinante eines Produkts von 2 Matrizen behandle ich später. Der Zusammenhang von Determinante und Volumen ist bei Nipp und Stoffer auf p. 137 ff. Die wichtigsten Resultate sind in den Sätzen 3.1, 3.3, 3.11, 4.5 und 4.6.

5 Lineare Abbildungen und Matrixmultiplikation

5.1 Der allgemeine Abbildungsbegriff

Aus der Mittelschule sind Sie vertraut mit Funktionen, bei denen sowohl das Argument als auch der Funktionswert reelle Zahlen sind: $x \in \mathbb{R} \mapsto y = f(x) \in \mathbb{R}$. Wir betrachten hier nun Funktionen F , auch Abbildungen genannt, mit beliebigem Definitionsbereich D und beliebigem Wertebereich W : Darunter versteht man eine Vorschrift, die jedem Argument $x \in D$ eindeutig einen Funktionswert $y = F(x) \in W$ zuordnet. Wir schreiben

$$F : D \longrightarrow W, \quad x \mapsto y = F(x) = y(x) \in W.$$

Speziell von Bedeutung sind Funktionen mit $D \subseteq \mathbb{R}^n$ und $W \subseteq \mathbb{R}^m$. Für $n = 1$ und $m = 2, 3$ bezeichnen wir das Argument meist mit t : $t \mapsto (y_1(t), y_2(t), y_3(t))^T$. Eine Abbildung ist dann also eine Parameterdarstellung einer Kurve in der Ebene oder im Raum. Für $n = 2$ und $m = 1$ haben wir ein skalares Feld, z.B. die Temperatur zu einer festen Zeit als Funktion des Ortes. Die Punkte $(x_1, x_2, F(x_1, x_2))$ stellen dann eine Fläche über der Ebene dar. Für $n = m = 2$ haben wir ein Vektorfeld in der Ebene, z.B. die Windgeschwindigkeit zu einer festen Zeit als Funktion des Ortes. Ein anderes Beispiel mit $n = m = 2$ ergibt sich, wenn wir ein Teilchen betrachten, das sich in der Ebene bewegt (gemäss einer Differentialgleichung), und wir jeder Anfangsposition x zur Zeit $t = 0$ die Position y zur Zeit $t = 1$ zuordnen.

Das Bild einer Abbildung $F : D \longrightarrow W$ ist die Menge aller möglicher Funktionswerte: $\text{Bild}(F) = \{y; y = F(x) \text{ für ein } x \in D\}$. Eine Abbildung $F : D \longrightarrow W$ heisst *surjektiv*, falls $\text{Bild}(F) = W$. Sie heisst *injektiv*, falls verschiedene Argumente verschiedene Funktionswerte haben: $x \neq x' \Rightarrow F(x) \neq F(x')$. Sie heisst *bijektiv*, falls sie sowohl injektiv als auch surjektiv ist. Sind D und W endlich, so kann man eine Funktion durch Pfeile darstellen, die von D nach W gehen: Von jedem $x \in D$ muss genau ein Pfeil ausgehen. Ist F surjektiv, so endet an jedem $y \in W$ mindestens ein Pfeil. Ist F injektiv, so endet an jedem $y \in W$ höchstens ein Pfeil. Ist F bijektiv, so endet an jedem $y \in W$ genau ein Pfeil. Dann müssen D und W gleich viele Elemente haben.

Zwei Abbildungen $F_1 : D_1 \longrightarrow W_1$, $F_2 : D_2 \longrightarrow W_2$ kann man zusammensetzen, wenn $W_1 \subseteq D_2$ gilt. Die Zusammensetzung ist definiert als

$$F_2 \circ F_1 : D_1 \longrightarrow W_2, \quad x \mapsto (F_2 \circ F_1)(x) = F_2(F_1(x)).$$

Bei der Zusammensetzung von zwei Abbildungen wird also die Abbildung, die rechts steht, zuerst ausgeführt. Wenn $F_2 \circ F_1$ definiert ist, dann ist die umgekehrte Zusammensetzung $F_1 \circ F_2$ nicht immer definiert, denn dafür muss $W_2 \subseteq D_1$. Selbst wenn beide Zusammensetzungen definiert sind, beschreiben sie im Allgemeinen nicht die gleiche Abbildung $F_1 \circ F_2 \neq F_2 \circ F_1$, die Zusammensetzung von Abbildungen ist nicht kommutativ (Gegenbeispiele kann man sehr leicht konstruieren).

Hingegen ist die Zusammensetzung von Abbildungen assoziativ: Wenn $F_2 \circ F_1$ und $F_3 \circ F_2$ beide definiert sind, dann ist auch die Zusammensetzung der drei Abbildungen definiert, und es gilt

$$F_3 \circ (F_2 \circ F_1) = (F_3 \circ F_2) \circ F_1.$$

In beiden Fällen wird nämlich x auf $F_3(F_2(F_1(x)))$ abgebildet. Wir lassen daher die Klammern weg und schreiben $F_3 \circ F_2 \circ F_1$.

Ist $F : D \rightarrow W$ bijektiv, dann existiert die Umkehrabbildung $F^{-1} : W \rightarrow D$: Für $y \in W$ ist $F^{-1}(y)$ das eindeutig festgelegte x mit $F(x) = y$. Dann gilt

$$F^{-1}(F(x)) = x \quad (x \in D), \quad F(F^{-1}(y)) = y \quad (y \in W).$$

$F \circ F^{-1}$ ist also die Identität auf W und $F^{-1} \circ F$ die Identität auf D .

5.2 Multiplikation von Matrix mal Vektor als lineare Abbildung

Wenn A eine $m \times n$ Matrix ist, dann definiert $x \mapsto y = Ax$ eine Abbildung F_A von \mathbb{R}^n nach \mathbb{R}^m . Wegen (3) ist diese Abbildung linear in folgendem Sinne

$$F_A(x + x') = F_A(x) + F_A(x'), \quad F_A(\lambda x) = \lambda F_A(x).$$

Der j -te Spaltenvektor $a^{(j)}$ ist das Bild des j -ten Standardbasisvektors $e^{(j)}$, der an der j -ten Stelle eine Eins und sonst alles Nullen hat. Jede lineare Abbildung von \mathbb{R}^n nach \mathbb{R}^m ist von der Form $F(x) = Ax$, und zwar ist A gerade bestimmt durch $a^{(j)} = F(e^{(j)})$. Aus der Linearität folgt nämlich

$$F(x) = F(x_1 e^{(1)} + \dots + x_n e^{(n)}) = x_1 F(e^{(1)}) + \dots + x_n F(e^{(n)}) = x_1 a^{(1)} + \dots + x_n a^{(n)} = Ax.$$

Jede lineare Abbildung bildet 0 in $A0 = 0$ ab (0 ist also ein Fixpunkt von F_A). Weiter gehen Geraden in Geraden oder einen einzigen Punkt über:

$$A(x + \lambda u) = Ax + \lambda(Au),$$

Ebenen gehen in Ebenen oder eine Gerade oder einen Punkt über, etc..

Beispiele von linearen Abbildungen von $\mathbb{R}^2 \rightarrow \mathbb{R}^2$. Sei

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Dann ist F_A eine Spiegelung an der Winkelhalbierenden $x_1 = x_2$, F_B eine Drehung um den Winkel ϕ im Gegenuhrzeigersinn und F_C eine Streckung um den Faktor 2.

Die $n \times n$ Matrix I_n , welche in der Diagonalen alles Einsen hat und ausserhalb der Diagonalen alles Nullen, stellt die Identität dar: $I_n x = x$ für alle $x \in \mathbb{R}^n$. Wir nennen sie die *Einheitsmatrix*.

Das wichtigste Beispiel einer linearen Abbildung tritt in der Differentialrechnung mehrerer Variablen auf. Wenn F eine beliebige Abbildung $\mathbb{R}^n \rightarrow \mathbb{R}^m$ ist, dann ist F differenzierbar an der Stelle x_0 , wenn sie sich durch eine lineare Abbildung approximieren lässt

$$F(x) - F(x_0) \approx J(x_0) \cdot (x - x_0)$$

wobei die Approximation umso besser ist, je näher x bei x_0 liegt. Die Elemente der Matrix $J(x_0)$ sind dann die partiellen Ableitungen der i -ten Komponente von F nach der j -ten Komponente von x an der Stelle x_0 (die sogenannte Jacobi-Matrix).

5.3 Zusammensetzung von linearen Abbildungen und Matrixmultiplikation

Ist A eine $m \times n$ und B eine $p \times m$ Matrix, dann lassen sich F_A und F_B zusammensetzen. $F_B \circ F_A$ ist wieder linear

$$\begin{aligned}(F_B \circ F_A)(x + x') &= F_B(F_A(x + x')) = F_B(F_A(x) + F_A(x')) = F_B(F_A(x)) + F_B(F_A(x')) \\ &= F_B \circ F_A(x) + F_B \circ F_A(x').\end{aligned}$$

Also muss die Zusammensetzung wieder eine Multiplikation mit einer Matrix C sein, und zwar ist $C \in \mathbb{R}^{p \times n}$. Wir bezeichnen diese Matrix C als das Produkt der Matrizen $B \cdot A$. Die j -te Spalte von C ist gleich

$$(F_B \circ F_A)(e^{(j)}) = B(Ae^{(j)}) = Ba^{(j)} = \left(\sum_{k=1}^m B_{1k}A_{kj}, \dots, \sum_{k=1}^m B_{pk}A_{kj} \right)^T,$$

folglich

$$C_{ij} = (BA)_{ij} = \sum_{k=1}^m B_{ik}A_{kj} = B_{i1}A_{1j} + B_{i2}A_{2j} + \dots + B_{im}A_{mj}.$$

Für die Matrixmultiplikation berechnet man also alle Skalarprodukte von Zeilen des linken Faktors mit Spalten des rechten Faktors. Damit die Matrixmultiplikation definiert ist, müssen die Matrizen "zusammenpassen", d.h. die linke Matrix muss gleich viele Spalten haben wie die rechte Zeilen.

Die Eigenschaften über das Zusammensetzen von Abbildungen übersetzen sich direkt in Eigenschaften der Matrixmultiplikation: $A(BC) = (AB)C$, aber im Allgemeinen $AB \neq BA$, selbst wenn beide Produkte definiert sind. Ferner prüft man leicht nach, dass

$$A(B + C) = AB + AC, \quad (A + B)C = AC + BC, \quad (AB)^T = B^T A^T,$$

sofern die Ausdrücke auf der linken Seite definiert sind.

5.4 Die inverse Matrix

Wann ist die durch $A \in \mathbb{R}^{m \times n}$ definierte lineare Abbildung F_A injektiv? Genau dann, wenn aus $Ax = Ax'$ folgt, dass $x = x'$, oder wegen der Linearität, wenn aus $A(x - x') = 0$ folgt $x - x' = 0$, das heißt wenn $\text{Kern}(A) = \{0\}$. Vom Gaußalgorithmus her wissen wir, dass die Dimension von $\text{Kern}(A) = n - \text{Rang}(A)$ ist. Das heißt F_A ist injektiv, falls $\text{Rang}(A) = n$.

Wann ist F_A surjektiv? Genau dann, wenn das Bild $\{y; y = Ax \text{ für ein } x \in \mathbb{R}^n\}$ gleich \mathbb{R}^m ist. Weil Ax eine Linearkombination der Spaltenvektoren von A ist, ist das Bild nichts anderes als die lineare Hülle der Spalten von A , und diese hat bekanntlich die Dimension $\text{Rang}(A)$. Also ist F_A surjektiv, falls $\text{Rang}(A) = m$.

Damit ist F_A genau dann bijektiv, wenn $\text{Rang}(A) = m = n$. Solche Matrizen nennen wir regulär. Wenn F_A bijektiv ist, dann existiert die Umkehrabbildung F_A^{-1} : $F_A^{-1}y$ ist nichts anderes als die Lösung von $Ax = y$, und es lässt sich zeigen, dass F_A^{-1} wieder linear und damit von der Form Multiplikation einer Matrix mit einem Vektor ist. Diese Matrix nennen wir die Inverse von A und bezeichnen sie mit A^{-1} . Die Lösung von $Ax = y$ ist damit gleich $x = A^{-1}y$. Weil $F_A \circ F_A^{-1}$ und $F_A^{-1} \circ F_A$ beide die Identität sind, gilt auch

$$AA^{-1} = A^{-1}A = I_n.$$

Wenn A und B reguläre $n \times n$ Matrizen sind, dann kann man zeigen, dass auch A^T und AB regulär sind und dass gilt

$$(A^T)^{-1} = (A^{-1})^T, \quad (AB)^{-1} = B^{-1}A^{-1}.$$

Mit Hilfe der inversen Matrix kann man die Lösung von $Ax = y$ kurz schreiben als $x = A^{-1}y$. Will man diese Lösung aber tatsächlich ausrechnen, so ist es keine gute Idee, zuerst A^{-1} zu berechnen. Die j -te Spalte von A^{-1} ist nämlich nichts anderes als die Lösung von $Ax = e^{(j)}$, und im Allgemeinen ist dies der einzige Weg, um A^{-1} zu berechnen.

5.5 Orthogonale Matrizen

Im Allgemeinen verändert eine lineare Abbildung Längen und Winkel. Lineare Abbildungen, welche Längen und Winkel unverändert lassen, sind von speziellem Interesse. Für welche Matrizen A ist das der Fall? Wir gehen aus von $a^{(j)} = Ae^{(j)}$. Wenn F_A also Längen und Winkel nicht ändert, dann müssen die $a^{(j)}$ auch Länge eins haben und senkrecht aufeinander stehen:

$$(a^{(j)}, a^{(j)}) = 1, \quad (a^{(j)}, a^{(k)}) = 0 \quad (j \neq k).$$

Diese Bedingungen lassen sich in Matrixform kürzer schreiben als

$$A^T A = I_n$$

(Die Elemente von $A^T A$ sind die Skalarprodukte von Zeilen von A^T mit Spalten von A , und die Zeilen von A^T sind die Spalten von A). Diese Bedingung ist auch hinreichend dafür, dass F_A längen- und winkeltreu ist:

$$(Ax, Ax') = (Ax)^T (Ax') = x^T A^T Ax' = x^T I_n x' = (x, x').$$

Man kann sogar zeigen, dass aus Längentreue allein schon $A^T A = I_n$ folgt.

Die Bedingung $A^T A = I_n$ kann nur erfüllt sein, wenn $m \geq n$, weil die Spalten von A orthogonal und damit linear unabhängig sind. Quadratische Matrizen A mit $AA^T = I_n$ heissen *orthogonal*. Für diese Matrizen gilt $A^{-1} = A^T$ und damit auch $AA^T = I_n$, d.h. auch die Zeilen haben Länge 1 und stehen senkrecht aufeinander. Für orthogonale Matrizen lässt sich also die Inverse und damit auch die Lösung von $Ax = y$ ohne Gaußalgorithmus leicht berechnen.

5.6 Volumenänderung bei linearen Abbildungen

Analog zur Längen- und Winkeltreue kann man auch fragen, wie sich das Volumen bei einer linearen Abbildung ändert. Der Körper aufgespannt von den Standardbasisvektoren $e^{(j)}$ hat Volumen 1 und geht unter der linearen Abbildung F_A über in den Körper aufgespannt von den Spaltenvektoren $a^{(j)}$, dessen Volumen für $m = n$ gleich $|\det(A)|$ ist. Wegen der Linearität wird jeder Würfel im Ausgangsraum unter F_A mit dem gleichen Faktor $|\det(A)|$ vergrößert oder verkleinert. Da man nicht-würfelförmige Körper durch eine Vereinigung kleiner disjunkter Würfel approximieren kann, folgt daher, dass unter F_A das Volumen jedes Körpers mit $|\det(A)|$ verändert wird. Das Vorzeichen der Determinante sagt dann zusätzlich wie sich die Orientierung unter F_A ändert.

Setzt man zwei Abbildungen F_A und F_B zusammen, dann ist die gesamte Volumenänderung gleich dem Produkt der Volumenänderungen unter F_A , bzw. F_B , d.h. $|\det(AB)| =$

$|\det(A)||\det(B)|$. Diese Produktregel gilt auch für das Vorzeichen der Determinante, d.h. es gilt

$$\det(AB) = \det(A) \det(B) = \det(B) \det(A) = \det(BA) \quad (\text{auch wenn } AB \neq BA).$$

Da $\det(I_n) = 1$, folgt also insbesondere $\det(A^{-1}) = 1/\det(A)$. Wenn A orthogonal ist dann ist $\det(A) = \det(A^T) = \det(A^{-1}) = 1/\det(A)$, also $\det(A) = \pm 1$: Wenn Längen und Winkel gleich bleiben, dann auch das Volumen. Die Umkehrung gilt im Allgemeinen nicht.

5.7 Zusammenhang mit Nipp & Stoffer

Der Inhalt dieses Kapitels findet sich bei Nipp & Stoffer in den Kapiteln 2.2, 2.3 und 6. Satz 3.6 enthält die Regel $\det(AB) = \det(A) \det(B)$. Matrixnormen und 6.2 habe ich weggelassen.

6 Eigenwertprobleme

6.1 Eigenwerte und Eigenvektoren

Wenn A eine Matrix ist, in der die meisten Elemente ungleich Null sind, ist es schwierig zu sehen, was die lineare Abbildung $x \mapsto Ax$ bewirkt. Einfacher ist es im Fall einer Diagonalmatrix $A = \text{diag}(a_1, \dots, a_n)$. Dann wird der i -te Standardbasisvektor $e^{(i)}$ einfach mit dem Faktor a_i gestreckt oder gestaucht, aber er zeigt in die gleiche oder in die entgegengesetzte Richtung. Für eine beliebige Matrix A suchen wir nun analog Vektoren x , für die die Abbildung nur die Länge, aber nicht die Richtung ändert: Eine Zahl λ heisst *Eigenwert* einer $n \times n$ Matrix A falls ein Vektor $x \neq 0$ existiert mit $Ax = \lambda x$. Jeder solche Vektor x heisst ein Eigenvektor von A zum Eigenwert λ .

Eigenwerte und Eigenvektoren spielen auch eine wichtige Rolle bei homogenen linearen Differentialgleichungssystemen mit konstanten Koeffizienten

$$\dot{y}(t) = \frac{d}{dt}y(t) = Ay(t),$$

oder ausgeschrieben

$$\dot{y}_i(t) = a_{i1}y_1(t) + a_{i2}y_2(t) + \dots + a_{in}y_n(t) \quad (i = 1, 2, \dots, n).$$

Die Lösungen sind Kurven im \mathbb{R}^n , deren Tangentenvektor im Punkt $y(t)$ gleich dem "Strömungsfeld" $Ay(t)$ ist. Wenn x ein Eigenvektor von A zum Eigenwert λ ist, dann hat das Strömungsfeld die gleiche Richtung wie der Ortsvektor, d.h. wenn man zur Zeit 0 mit $y(0) = x$ startet, dann bleibt $y(t)$ für alle Zeiten t ein Vielfaches von x . Der Eigenwert gibt dabei an, wie wie der multiplizierende Skalar als Funktion von t aussieht:

$$y(t) = e^{\lambda t}x.$$

Eigenvektoren führen also zu besonders einfachen Lösungen eines solchen Gleichungssystems.

Wie können wir Eigenwerte und Eigenvektoren finden? Die Eigenwertgleichung $Ax = \lambda x$ lässt sich auch schreiben als $(A - \lambda I_n)x = 0$. Dies ist ein homogenes lineares Gleichungssystem und hat Lösungen $x \neq 0$ genau dann, wenn der Rang von $A - \lambda I_n$ kleiner als n ist,

oder anders gesagt, wenn $A - \lambda I_n$ singularär ist. Da man die Singularität mit Hilfe der Determinante testen kann, ist λ genau dann ein Eigenwert von A falls $\det(A - \lambda I_n) = 0$.

Mit Induktion kann man zeigen, dass die Funktion $\lambda \mapsto p_A(\lambda) = \det(A - \lambda I_n)$ ein Polynom vom Grad n in λ ist, in dem der Term λ^n den Koeffizienten $(-1)^n$ hat. Wir nennen p_A das charakteristische Polynom der Matrix A . Aus der Theorie der Polynome folgt, dass $p_A(\lambda)$ im Komplexen n eventuell zusammenfallende Nullstellen hat und sich wie folgt faktorisieren lässt:

$$p_A(\lambda) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n), \quad (\lambda_i \in \mathbb{C}).$$

Dass es Matrizen gibt, die keine reellen Eigenwerte haben, sieht man bereits am Fall von 2×2 Drehmatrizen, wo Ax die Richtung aller reeller Vektoren ändert.

Um alle Eigenvektoren zu einem Eigenwert λ_i zu finden, müssen wir wie gesagt alle nicht-trivialen Lösungen von $(A - \lambda_i I_n)x = 0$ bestimmen. Das kann mit Hilfe des Gauss-Algorithmus geschehen. Die Menge aller Lösungen von $(A - \lambda_i I_n)x = 0$ ist – wie wir früher gelernt haben – ein Unterraum der Dimension $n - \text{Rang}(A - \lambda_i I_n)$. Diese Dimension muss mindestens eins sein (sonst ist λ_i kein Eigenwert) und man kann zeigen, dass diese Dimension höchstens gleich der Vielfachheit von λ_i als Nullstelle von $p_A(\lambda)$ ist.

6.2 Diagonalisierbarkeit von Matrizen

Eigenwerte und Eigenvektoren sind insbesondere dann nützlich, wenn es n linear unabhängige Eigenvektoren $u^{(1)}, \dots, u^{(n)}$ gibt. Diese Eigenvektoren können wir dann als Basis des \mathbb{R}^n benutzen, das heißt wir schreiben ein beliebiges x als

$$x = c_1 u^{(1)} + \dots + c_n u^{(n)} = Uc.$$

Daraus folgt

$$Ax = c_1 Au^{(1)} + \dots + c_n Au^{(n)} = c_1 \lambda_1 u^{(1)} + \dots + c_n \lambda_n u^{(n)} = UDUc$$

wobei die Spalten von U aus den Eigenvektoren $u^{(i)}$ bestehen und D die Diagonalmatrix mit den Eigenwerten λ_i in der Diagonalen bezeichnet. Kombination dieser beiden Gleichungen ergibt $Ax = UDU^{-1}x$, und weil dies für alle x gilt, folgt

$$A = UDU^{-1} \Leftrightarrow U^{-1}AU = D.$$

Wir haben also gezeigt, dass sich eine Matrix “diagonalisieren” lässt falls n linear unabhängige Eigenvektoren existieren. Es lässt sich leicht zeigen, dass die Umkehrung auch gilt: Wenn es eine reguläre Matrix U und eine Diagonalmatrix D gibt mit $A = UDU^{-1}$, dann sind die Spalten von U Eigenvektoren von A und die Diagonalelemente von D die zugehörigen Eigenwerte.

Wann existieren nun n linear unabhängige Eigenvektoren? Ein positives Resultat besagt, dass k Eigenvektoren zu k verschiedenen Eigenwerten automatisch linear unabhängig sind. Die Diagonalisierbarkeit einer quadratischen Matrix mit reellem U und D ist also dann in Frage gestellt, wenn das charakteristische Polynom komplexe oder zusammenfallende Nullstellen hat. Komplexe Nullstellen werden wir unten diskutieren. Bei zusammenfallenden Nullstellen kann es (aber es muss nicht) geschehen, dass weniger als n linear unabhängige Eigenvektoren existieren.

6.3 Der Spezialfall von symmetrischen Matrizen

Besonders erfreulich ist die Situation bei symmetrischen Matrizen A , d.h. wenn $A^T = A$. Dann sind die Eigenwerte immer reell und es gibt auch bei zusammenfallenden Nullstellen stets n linear unabhängige Eigenvektoren. Symmetrische Matrizen lassen sich also stets im Reellen diagonalisieren. Es gilt sogar noch mehr: Eigenvektoren zu verschiedenen Eigenwerten stehen stets senkrecht aufeinander, und man kann die Matrix U stets als orthogonale Matrix wählen

$$A = UDU^T \Leftrightarrow U^T AU = D.$$

Falls keine Eigenwerte zusammenfallen, muss man dafür die Eigenvektoren einfach auf Länge 1 normieren.

6.4 Matrixpotenzen und lineare Differentialgleichungssysteme

Ist A diagonalisierbar, dann folgt sofort dass

$$A^k = A = UD^kU^{-1} \quad (k = 1, 2, \dots,$$

d.h. wenn man die Eigenwerte und Eigenvektoren kennt, kann man Matrixpotenzen leicht berechnen. Daraus gewinnt man Information, wie sich die Matrixpotenzen asymptotisch verhalten für $k \rightarrow \infty$. Wenn z.B. der Absolutbetrag aller Eigenwerte kleiner als 1 ist, dann konvergieren alle Elemente von A^k gegen null.

Wenn A diagonalisierbar ist, dann ist $y(t) = e^{\lambda_i t} u^{(i)}$ eine Lösung des Differentialgleichungssystems $\dot{y}(t) = Ay(t)$ für jedes $i = 1, 2, \dots, n$ eine Lösung. Wegen der Linearität ist dann auch

$$y(t) = c_1 e^{\lambda_1 t} u^{(1)} + c_2 e^{\lambda_2 t} u^{(2)} + \dots + c_n e^{\lambda_n t} u^{(n)}$$

eine Lösung für beliebige Koeffizienten c_i . Diese Lösung hat die Anfangsbedingung $y(0) = c_1 u^{(1)} + \dots + c_n u^{(n)} = Uc$, d.h. man kann jede Anfangsbedingung durch geeignete Wahl der c_i erfüllen. Da die Lösung des Differentialgleichungssystems für gegebene Anfangsbedingung eindeutig ist, haben wir damit die allgemeine Lösung gefunden.

6.5 Komplexe Eigenwerte und Eigenvektoren

Wir hatten bisher nur lineare Gleichungssysteme mit reellen Koeffizienten betrachtet. Der Gauss-Algorithmus funktioniert aber genau gleich mit komplexen Koeffizienten; man erhält dann einfach komplexe Lösungen. Auch die Begriffe wie Rang, Kern, lineare (Un)abhängigkeit lassen sich ohne Schwierigkeiten übertragen. Das heisst zu einer komplexen Nullstelle von p_A gibt es ganz analog komplexe Eigenvektoren, und man kann daher auch Diagonalisierung einer Matrix A betrachten mit Hilfe von Matrizen U und D , deren Elemente komplexe Zahlen sind.

Bei einem linearen Differentialgleichungssystem $\dot{y}(t) = Ay(t)$ ist man jedoch an reellen Lösungen interessiert, wenn die Matrix A reell ist. Weil komplexe Nullstellen von reellen Polynomen als komplex konjugierten Paaren auftreten, erhält man solche Lösungen ohne Probleme. Wenn z.B. $\lambda_1 = \alpha + i\beta$ ein komplexer Eigenwert von A ist mit komplexem Eigenvektor $u^{(1)} = v + iw$, dann ist auch $\lambda_2 = \overline{\lambda_1} = \alpha - i\beta$ ein Eigenwert von A , und zwar mit Eigenvektor $u^{(2)} = \overline{u^{(1)}} = v - iw$. Dann ist die Linearkombination $c_1 e^{\lambda_1 t} u^{(1)} + c_2 e^{\lambda_2 t} u^{(2)}$, die in der allgemeinen Lösung oben auftritt, genau dann reell, wenn $c_2 = \overline{c_1}$.

Setzt man $c_1 = a + ib$ und $c_2 = a - ib$, so ergibt sich mit den Regeln für die komplexe Exponentialfunktion

$$c_1 e^{\lambda_1 t} u^{(1)} + c_2 e^{\lambda_2 t} u^{(2)} = 2e^{\alpha t} (a \cos(\beta t) - b \sin(\beta t))v - 2e^{\alpha t} (a \sin(\beta t) + b \cos(\beta t))w.$$

Die Werte von a und b werden dann mit Hilfe der Anfangsbedingungen festgelegt.

Aus der Formel

$$y(t) = \sum_{j=1}^n c_j e^{\lambda_j t} u^{(j)}$$

für die allgemeine Lösung lässt sich insbesondere das Verhalten für grosse Zeiten t ablesen. Wenn $y(0) = 0$, dann ist $y(t) \equiv 0$, d.h. 0 ist ein Fixpunkt des Systems. Wenn λ_j einen negativen Realteil hat, dann konvergiert $e^{\lambda_j t}$ gegen Null für $t \rightarrow \infty$. Wenn die Realteile aller Eigenwerte negativ sind, dann konvergiert also $y(t)$ für $t \rightarrow \infty$ bei beliebiger Anfangsbedingung gegen 0 , d.h. der Fixpunkt 0 ist stabil. Wenn der Realteil von einem λ_j hingegen positiv ist, dann geht $\|y(t)\|$ gegen unendlich, ausser wenn wir die Anfangsbedingung so wählen, dass das zugehörige c_j gleich Null ist. Daher ist in diesem Fall der Fixpunkt 0 instabil.

6.6 Zusammenhang mit Nipp & Stoffer

Der Stoff dieses Kapitels findet sich bei Nipp & Stoffer in den Kapiteln 7 und 8.1. Von 7.4 behandle ich nur die Berechnung der Matrixpotenzen, alles ab p. 163 nicht mehr.

7 Ausgleichsrechnung

7.1 Anpassung von Geraden an Messpunkte (x_i, y_i)

In manchen Anwendungen hat man eine genäherte lineare Beziehung zwischen zwei Variablen, d.h. für n Messpunkte (x_i, y_i) gilt

$$y_i \approx \alpha + \beta x_i.$$

Wie soll man die best-passendste Gerade bestimmen? Die gebräuchlichste Methode ist die Methode der Kleinsten Quadrate, welche α und β so bestimmt, dass die Summe der quadrierten Abweichungen in y -Richtung

$$\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

minimal wird. Durch partielles Ableiten und Null setzen erhält man die beiden Gleichungen

$$\sum_{i=1}^n (y_i - \alpha - \beta x_i) = 0, \quad \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i = 0$$

für α und β , deren Lösung man leicht berechnen kann. Wir können dies auch mit der Theorie der Vektoren im \mathbb{R}^n geometrisch deuten: Wenn wir den Vektor $y = (y_1, y_2, \dots, y_n)^T$, $\mathbf{1} = (1, 1, \dots, 1)^T$ und $x = (x_1, x_2, \dots, x_n)^T$ einführen, dann suchen wir eine genäherte Lösung des Gleichungssystems

$$y \approx \alpha \mathbf{1} + \beta x.$$

Die Methode der Kleinsten Quadrate besagt dann, dass wir die Unbekannten α und β so wählen, dass der Residuen- (oder Fehler-)Vektor $r = y - \alpha \mathbf{1} - \beta x$ minimale Länge hat. Geometrisch ist es dann sofort einleuchtend, dass dies dann der Fall ist, wenn der Fehlervektor orthogonal steht auf dem Unterraum aufgespannt von $\mathbf{1}$ und x :

$$(y - \alpha \mathbf{1} - \beta x, \mathbf{1}) = 0, \quad (y - \alpha \mathbf{1} - \beta x, x) = 0.$$

Schreiben wir die Skalarprodukte in Komponenten aus, dann haben wir dieselben Gleichungen wie oben beim Nullsetzen der partiellen Ableitungen.

7.2 Ausgleichsrechnung allgemein

Die Methode der Kleinsten Quadrate lässt sich immer anwenden, wenn wir ein überbestimmtes lineares Gleichungssystem (mehr Gleichungen als Unbekannte) haben. Für den Übergang zur Statistik ändern wir im Folgenden die Notation: Die rechte Seite nennen wir y statt b , Die Koeffizientenmatrix X statt A , und die Unbekannten β statt x (Wir bleiben damit nahe beim einführenden Beispiel aus dem vorangehenden Unterabschnitt). Wir betrachten also das Gleichungssystem

$$X\beta = y$$

mit $X \in \mathbb{R}^{n \times k}$, $y \in \mathbb{R}^n$ und $\beta \in \mathbb{R}^k$. Geometrisch heisst das, dass wir y als Linearkombination der Spaltenvektoren $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ darstellen wollen. Wenn $k < n$, dann ist der Spaltenraum ein echter Unterraum von \mathbb{R}^n und wir haben ausser in Spezialfällen keine Lösung. In diesem Fall muss man sich mit einer approximativen Lösung begnügen: Wir suchen dasjenige x , für das der Residuenvektor $r = y - Ax$ eine möglichst kleine Länge hat. Wir minimieren also $\|y - X\beta\|$ (oder äquivalent dazu $\|y - X\beta\|^2$) bezüglich $\beta \in \mathbb{R}^k$. Dies heisst auch ein *Ausgleichsproblem*.

Die Lösung hängt im Allgemeinen von der gewählten Norm ab. Besonders einfach wird es, wenn wir die übliche Euklid'sche Norm verwenden. Da wir dann den Ausdruck

$$\|y - X\beta\|^2 = \sum_{i=1}^n (y_i - x_{i1}\beta_1 - \dots - x_{ik}\beta_k)^2$$

minimieren, spricht man von der *Methode der kleinsten Quadrate*. Wir appellieren wieder an die geometrische Anschauung, dass $\|y - X\beta\|$ minimal wird, falls $X\beta$ gleich der orthogonalen Projektion von y auf den Spaltenraum von X ist. Das heisst, der Residuenvektor muss senkrecht auf den Vektoren stehen, die den Spaltenraum aufspannen.

$$(y - X\beta, x^{(i)}) = (r, x^{(i)}) = 0 \quad (i = 1, \dots, k).$$

Mit der Linearität des Skalarprodukts erhalten wir

$$0 = (y - \sum_{j=1}^k \beta_j x^{(j)}, x^{(i)}) = (y, x^{(i)}) - \sum_{j=1}^k \beta_j (x^{(j)}, x^{(i)}),$$

bzw. in Matrixform

$$(X^T X)\beta = X^T y.$$

Die Lösung dieser sogenannten *Normalgleichungen* ist eindeutig, wenn die Spaltenvektoren $x^{(1)}, \dots, x^{(k)}$ linear unabhängig sind. Um ein überbestimmtes lineares Gleichungssystem $X\beta = y$ im Sinne der kleinsten Quadrate zu lösen, multipliziert man also einfach beide Seiten mit X^T .

In der Statistik werden wir einen Schritt weitergehen bei der Betrachtung der Ausgleichsrechnung. Wir nehmen an, dass in einer idealen Welt eine Beziehung $y = X\beta_0$ gelten würde (d.h. y liegt im Spaltenraum von X), aber das wahre β_0 ist unbekannt. Auf Grund unvermeidlicher Fehler in den Messungen liegt das gemessene y ausserhalb des Spaltenraums, und mit der Kleinste-Quadrate-Approximation erhalten wir eine *Schätzung* von β_0 . Wenn wir noch annehmen, dass die Fehler $y - X\beta_0$ zufällig sind und deren Verteilung gewissen Annahmen genügt, dann können wir etwas aussagen über die Genauigkeit der Kleinste-Quadrate-Schätzung.

7.3 Zusammenhang mit Nipp & Stoffer

Der Stoff dieses Kapitels steht bei Nipp & Stoffer in 5.1.