

Zusammenfassung

Lineare Algebra

Dr. Marcel Dettling
Institut für Datenanalyse und Prozessdesign
Zürcher Hochschule für angewandte Wissenschaften

Frühlingssemester 2010

1 Lineare Gleichungssysteme

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

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

Dabei sind die Koeffizienten a_{ik} und die rechten Seiten b_i vorgegebenen und die x_k sind gesucht. Wenn alle $b_i = 0$ sind, heisst das System *homogen*, sonst *inhomogen*. Ein homogenes System hat immer die triviale Lösung

$$x_1 = x_2 = \dots = x_n = 0,$$

und die allgemeine Lösung eines inhomogenen Systems ist gleich einer speziellen Lösung plus der allgemeinen Lösung des zugehörigen homogenen Systems.

Der *Gauss-Algorithmus* formt dieses Gleichungssystem in ein anderes, einfacheres System um, welches die gleiche Lösungsmenge hat. Einfacher heisst hier, dass die i -te Gleichung nur die Unbekannten x_i, x_{i+1}, \dots, x_n enthält. Die Lösungen eines solchen Systems sind leicht zu bestimmen, siehe unten. Die Umformungen bestehen darin, dass man Gleichungen vertauscht und von einer Gleichung ein Vielfaches einer andern Gleichung subtrahiert. Zur Vereinfachung führt man diese Schritte durch auf der *Matrix* der Koeffizienten (a_{ik}), augmentiert durch den *Spaltenvektor* (b_i) der rechten Seite, d.h. man lässt alle x_k , “+” und “=” weg. Statt von Gleichungen spricht man dann auch von *Zeilen*.

Zu Beginn des j -ten *Eliminationsschritts* sieht die augmentierte Matrix wie folgt aus

$$\begin{pmatrix} a_{11} & \dots & & & & & a_{1n} & b_1 \\ 0 & a_{22} & \dots & & & & \vdots & \vdots \\ \vdots & 0 & \ddots & & & & & \\ & \vdots & \ddots & \ddots & & & & \\ & & & 0 & a_{jj} & \dots & & \\ & & & \vdots & \vdots & & & \\ 0 & 0 & \dots & 0 & a_{mj} & \dots & a_{mn} & b_m \end{pmatrix}.$$

Dabei sind die Zahlen a_{ik} und b_i nicht mehr alle dieselben wie am Anfang. Im Normalfall, den wir zuerst besprechen, sind die Werte $a_{11}, \dots, a_{j-1,j-1}$, die sogenannten *Pivots*, alle ungleich null. Wir nehmen an, dass eine der Zahlen a_{jj}, \dots, a_{mj} nicht null ist. Dann können wir mit einer Zeilenvertauschung erreichen, dass auch $a_{jj} \neq 0$. Dies ist unser nächstes Pivot. Der eigentliche Eliminationsschritt besteht dann darin, dass wir das a_{kj}/a_{jj} -fache der Zeile Nummer j von der Zeile Nummer k subtrahieren, und zwar für $k = j+1, \dots, m$. Dies ergibt dann lauter Nullen unterhalb von a_{jj} , d.h. wir sind einen Schritt weitergekommen. Nach m Eliminationsschritten ist das Verfahren beendet.

Für $m = n$ haben wir dann am Schluss ein Gleichungssystem in Dreiecksform, und es ist klar, dass wir die Lösung durch *Rückwärtseinsetzen* finden können. Insbesondere existiert also im Normalfall (d.h. sofern man in jedem Eliminationsschritt ein Pivot gefunden hat) für beliebige rechte Seiten immer genau eine Lösung.

Wenn $m < n$ ist, dann können wir (wieder im Normalfall) x_{m+1}, \dots, x_n beliebig wählen und danach x_m, \dots, x_1 durch Rückwärtseinsetzen bestimmen. Es existieren also stets unendlich viele Lösungen, und zwar haben wir $n - m$ freie Parameter.

Wenn $m > n$, dann haben wir (immer noch im Normalfall) am Schluss $m - n$ Gleichungen, bei denen alle Koeffizienten null sind. Wenn alle Werte b_{m+1}, \dots, b_n auf der rechten Seite (nach Beendigung des Gauss-Algorithmus) ebenfalls gleich null sind, dann kann man die letzten Gleichungen weglassen und die Lösung durch Rückwärtseinsetzen bestimmen. Wenn hingegen eine oder mehrere dieser Zahlen von null verschieden sind, dann gibt es keine Lösung.

Im Ausnahmefall kann man während einem Eliminationsschritt kein Pivot finden. Das heisst einfach, dass die entsprechende Unbekannte bereits eliminiert ist und man den entsprechenden Schritt überspringen kann. Die Pivots stehen dann nicht mehr alle in der Diagonalen, und man erhält am Schluss eine sogenannte *Zeilenstufenform*. Jede Stufe wird von einem Pivot, das definitionsgemäss nicht null ist, angeführt. Die Anzahl Pivots heisst auch der *Rang* des Gleichungssystems und wird mit r bezeichnet. Wenn $r < m$, dann kommt es auf die rechten Seiten an, ob eine Lösung existiert. Für $r = m$ existiert immer

mindestens eine Lösung. Wenn eine Lösung existiert, dann ist sie eindeutig falls $r = n$ ist, während man für $r < n$ alle x_k , die nicht zu einem Pivot gehören, beliebig wählen kann.

Damit haben wir die Frage nach der Existenz und der Anzahl Lösungen eines linearen Gleichungssystems vollständig beantwortet, und wir können auch alle Lösungen berechnen. Die einzige Schwierigkeit ist die, dass man einem Gleichungssystem im Allgemeinen nicht ansieht, ob man im Normal- oder im Ausnahmefall ist. Das ergibt sich erst im Verlaufe des Eliminationsverfahrens. Zum Schluss noch ein paar kleinere Bemerkungen:

- Wenn man ein lineares Gleichungssystem lösen will für verschiedene rechte Seiten, schreibt man alle rechten Seiten nebeneinander. Die augmentierte Matrix hat dann also mehr als $n + 1$ Spalten. Die Eliminationsschritte kann man so simultan durchführen. Das Rückwärtseinsetzen muss man jedoch für jede rechte Seite separat ausführen.
- Der Aufwand für das Lösen eines linearen $n \times n$ Gleichungssystems benötigt ungefähr $n^3/3$ Divisionen bzw. Multiplikationen. Heutzutage führen viele numerische Probleme am Schluss auf riesige lineare Gleichungssysteme. In diesen Situationen ist ein Aufwand der Ordnung $n^3/3$ beachtlich, und man entwickelt auch heute noch effizientere Algorithmen für Gleichungssysteme mit speziellen Eigenschaften.
- Bei allen Berechnungen auf dem Computer muss man sich Gedanken über Rundungsfehler machen. Beim Gauss-Algorithmus kann man Rundungsfehler klein halten durch eine geschickte Wahl der Pivots.

2 Matrizenrechnung

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: Der erste gibt die Nummer der Zeile an und der zweite die Spalte an. Eine $m \times 1$ Matrix ist ein *Spaltenvektor* und eine $1 \times n$ Matrix ein *Zeilenvektor*. Elemente von Vektoren benötigen nur einen Index.

Die Multiplikation einer Matrix mit einer Zahl (einem sogenannten *Skalar*) und die Addition zweier $m \times n$ Matrizen erfolgt Element für Element. Die Multiplikation von Matrizen hingegen ist anders definiert: Wenn A eine $m \times n$ Matrix ist mit Elementen a_{ij} und B eine $n \times p$ Matrix mit Elementen b_{ij} , dann ist AB eine $m \times p$ Matrix mit den folgenden Elementen

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}.$$

Man bildet also das "Skalarprodukt" der i -ten Zeile von A mit der j -ten Spalte von B . Wenn die Anzahl Spalten von A verschieden ist von der Anzahl Zeilen von B , existiert das Produkt AB nicht. Der Grund für diese Definition wird klar werden, wenn wir lineare Abbildungen betrachten.

Es gelten die üblichen Rechenregeln, die auch für Zahlen gelten, z.B. $(AB)C = A(BC)$, $A(B + C) = AB + AC$ etc. (vgl. die vollständige Liste in Satz 2.1 von Nipp und Stoffer), mit einer wichtigen Ausnahme: Die *Matrixmultiplikation ist nicht kommutativ*. Im Allgemeinen existieren nicht beide Produkte AB und BA , und wenn beide existieren, dann ist im Allgemeinen $AB \neq BA$.

Das Produkt $m \times n$ Matrix A mal $n \times 1$ Spaltenvektor x ist ein $m \times 1$ Vektor, und zwar gerade 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 . Ein lineares Gleichungssystem lässt sich daher mit der Matrixmultiplikation in der Kurzform $Ax = b$ schreiben. Ferner gilt: Wenn $b^{(1)}, \dots, b^{(p)}$ die Spalten von B sind, dann sind die Spalten von AB gerade $Ab^{(1)}, \dots, Ab^{(p)}$.

Die $n \times n$ Matrix mit lauter 1 in der Diagonalen und 0 ausserhalb heisst die *Einheitsmatrix* I_n . Es gilt $AI_n = I_n A = A$ für jede $m \times n$ Matrix A . 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 $(A^T)^T = A$ und $(AB)^T = B^T A^T$.

Zu einer $n \times n$ Matrix A gibt es höchstens eine $n \times n$ Matrix X mit $AX = I_n$, und wenn ein solches X existiert, dann gilt auch $XA = I_n$. Diese Matrix X heisst die *Inverse* von A und wird mit A^{-1} bezeichnet. Wenn die Inverse existiert, heisst A invertierbar, oder *regulär*, andernfalls heisst A *singulär*. Es gilt: A ist genau dann regulär, wenn das Gleichungssystem $Ax = b$ für beliebige rechte Seiten b lösbar ist, d.h. genau dann, wenn der Rang von A gleich n ist.

Wenn A und B reguläre $n \times n$ Matrizen sind, dann sind auch AB und A^T regulär, und es gilt $(AB)^{-1} = B^{-1}A^{-1}$ und $(A^T)^{-1} = (A^{-1})^T$.

Eine $n \times n$ Matrix heisst *orthogonal*, falls $A^T A = I_n$ gilt. Weil die Elemente von $A^T A$ gerade die Skalarprodukte der Spalten von A sind, haben die Spalten einer orthogonalen Matrix "Länge" 1 und stehen "senkrecht" aufeinander. Wenn A orthogonal ist, dann ist A regulär mit $A^{-1} = A^T$ und es gilt auch $AA^T = I_n$.

3 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 A singulär ist (d.h. die Determinante ist eine Testgrösse für die Singularität einer Matrix), ii) $|\det(A)|$ ist das Volumen des Spats aufgespannt von den Spaltenvektoren von A .

Für $n = 2$ gilt

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

Für $n = 3$ ist

$$\det(A) = a^{(1)T} (a^{(2)} \times a^{(3)}),$$

wobei $a^{(j)}$ der j -te Spaltenvektor von A ist und \times das Vektorprodukt bezeichnet:

$$a \times b = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)^T.$$

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:

$$\begin{array}{cccccc} a & b & c & a & b & \\ d & e & f & d & e & \\ g & h & k & g & h & \end{array}$$

Für $n > 3$ ist die Determinante rekursiv definiert durch Entwicklung nach der ersten Spalte, siehe Nipp und Stoffer, p. 52 oben.

Wichtiger als die Definition sind die folgenden Eigenschaften (ohne Beweis):

- Bei der Vertauschung zweier Spalten ändert die Determinante das Vorzeichen. Wenn zwei Spalten gleich sind, ist die Determinante daher null.
- Die Determinante ist linear in jeder Spalte. d.h.

$$\begin{aligned} \det(a^{(1)}, \dots, a^{(j-1)}, \lambda a^{(j)} + \mu b^{(j)}, a^{(j+1)}, \dots, a^{(n)}) = \\ \lambda \det(a^{(1)}, \dots, a^{(j-1)}, a^{(j)}, a^{(j+1)}, \dots, a^{(n)}) + \\ \mu \det(a^{(1)}, \dots, a^{(j-1)}, b^{(j)}, a^{(j+1)}, \dots, a^{(n)}). \end{aligned}$$

- Wenn A eine obere Dreiecksmatrix ist, d.h. $a_{ij} = 0$ für $i > j$, dann ist $\det(A) = a_{11} \cdot a_{22} \cdots a_{nn}$.
- $\det(A^T) = \det(A)$. Deshalb gelten die ersten beiden Regeln auch für Zeilen anstelle von Spalten.
- $\det(AB) = \det(A) \det(B)$.

Daraus folgt, dass man eine Determinante mit dem Gauss-Algorithmus berechnen kann, denn bei jedem Eliminationsschritt ändert sich höchstens das Vorzeichen.

4 Vektorräume

4.1 Der Vektorraum \mathbb{R}^n

Spaltenvektoren mit n Elementen kann man in den Fällen $n = 2$ und 3 geometrisch als "Orts-" oder "Richtungs-Vektoren" in der Ebene, bzw. im Raum

auffassen. Multiplikation mit einer Zahl entspricht dann dem Strecken, bzw. Stauchen, und Addition dem Zusammensetzen gemäss der bekannten Parallelogrammregel. Wir werden im Folgenden diese geometrische Anschauung auch für $n > 3$ verwenden, d.h. wir interpretieren Spaltenvektoren der Länge n als “Orts-” oder “Richtungs-Vektoren” in einem “ n -dimensionalen” Raum. Dies führt unter anderem zu einem geometrischen Verständnis von linearen Gleichungssystemen.

Das zentrale Konzept ist das des *Unterraums*. Darunter versteht man eine nicht-leere Teilmenge U von \mathbb{R}^n mit den Eigenschaften, dass mit $x \in U$ und $y \in U$ auch $\alpha x + \beta y \in U$ für beliebige reelle Zahlen α und β . Ein Unterraum ist also abgeschlossen bezüglich Addition und Multiplikation mit einer Zahl. Für $n = 2, 3$ ist ein Unterraum eine Gerade oder eine Ebene durch den Ursprung. Im \mathbb{R}^3 kann man eine Ebene durch den Ursprung entweder durch eine Parameterdarstellung $\{sa^{(1)} + ta^{(2)} \mid s, t \in \mathbb{R}\}$ angeben oder als Lösungsmenge einer Gleichung $ax_1 + bx_2 + cx_3 = 0$. Dies überträgt sich auf andere Dimensionen.

Die Verallgemeinerung der Parameterdarstellung ist die Darstellung eines Unterraums als *lineare Hülle* von vorgegebenen Vektoren $a^{(1)}, a^{(2)}, \dots, a^{(k)} \in \mathbb{R}^n$, einem sogenannten *Erzeugendensystem*. Das heisst

$$U = \text{span}\{a^{(1)}, a^{(2)}, \dots, a^{(k)}\} = \{x_1 a^{(1)} + x_2 a^{(2)} + \dots + x_k a^{(k)} \mid x_1, x_2, \dots, x_k \in \mathbb{R}\}.$$

Die zweite mögliche Darstellung eines Unterraums ist die als Menge der Lösungen eines homogenen linearen Gleichungssystems: $U = \{x \mid Bx = 0\}$. Dabei ist B eine vorgegebene $m \times n$ Matrix.

Eine Grundaufgabe besteht darin festzustellen, ob ein Vektor c in einem Unterraum U liegt. Wenn U als Lösungsmenge eines homogenen linearen Gleichungssystems gegeben ist, setzt man einfach ein. Wenn U durch ein Erzeugendensystem gegeben ist, betrachtet man das Gleichungssystem $Ax = c$, wobei A die Matrix mit den Spaltenvektoren $a^{(1)}, a^{(2)}, \dots, a^{(k)}$ ist. Eine Lösung existiert genau dann, wenn c in diesem Unterraum liegt, und jede Lösung gibt eine Möglichkeit an, wie man c als Linearkombination des Erzeugendensystems ausdrücken kann. Es gibt genau dann mehrere Lösungen, wenn das homogene System $Ax = 0$ nichttriviale Lösungen hat, oder in andern Worten, wenn es Zahlen x_1, x_2, \dots, x_k gibt, welche nicht alle gleich null sind, so dass

$$x_1 a^{(1)} + x_2 a^{(2)} + \dots + x_k a^{(k)} = 0.$$

Dann heissen $a^{(1)}, a^{(2)}, \dots, a^{(k)}$ *linear abhängig*. Ist hingegen obige Gleichung nur erfüllt für $x_1 = x_2 = \dots = x_k = 0$, so heissen die Vektoren *linear unabhängig*.

Wenn ein Erzeugendensystem linear abhängig ist, dann sind einer (oder mehrere) der Vektoren im Erzeugendensystem überflüssig: Man kann mindestens einen Vektor weglassen, ohne dass sich die lineare Hülle verkleinert. Ein linear unabhängiges Erzeugendensystem hingegen kann nicht weiter reduziert werden. Ein solches minimales Erzeugendensystem heisst eine *Basis* des Unterraums. Ein Unterraum $U \neq \{0\}$ hat viele Basen, aber alle Basen bestehen aus gleich

vielen Vektoren. Diese Anzahl heisst die *Dimension* von U . Jeder Vektor aus U lässt sich eindeutig als Linearkombination von Basisvektoren darstellen. Die dabei auftretenden Zahlen heissen die *Koordinaten* des Vektors.

Eine weitere Grundaufgabe besteht darin, eine Basis für einen Unterraum U zu finden. Dazu verwendet man den Eliminationsalgorithmus von Gauss. Wenn U durch ein Erzeugendensystem gegeben ist, so schreibt man die erzeugenden Vektoren in die Spalten einer Matrix A : Die Spaltenvektoren von A , die zu den Pivotvariablen gehören, bilden dann eine Basis von U , die andern Spaltenvektoren sind überflüssig. Die Dimension ist also gleich dem Rang von A . Wenn U gegeben ist als die Menge der Lösungen eines homogenen linearen Gleichungssystems $Bx = 0$, dann erhält man den i -ten Basisvektor, indem man die i -te Nicht-Pivotvariable gleich 1 und die andern Nicht-Pivotvariablen gleich 0 setzt und die Pivotvariablen durch Rückwärtseinsetzen bestimmt. Die Dimension dieses Unterraums ist daher die Anzahl Spalten minus der Rang von B .

4.2 Abstrakte Vektorräume

Ein abstrakter Vektorraum ist eine Menge von mathematischen Objekten, welche man addieren und mit Zahlen multiplizieren kann gemäss den üblichen Regeln. Die wichtigsten Beispiele neben den geometrischen Vektoren und dem \mathbb{R}^n sind die Menge der Matrizen von fester Form und die Menge von Funktionen mit einem festen Definitionsbereich. Die Begriffe Unterraum, Erzeugendensystem, lineare (Un-)Abhängigkeit und Basis sind analog definiert wie im \mathbb{R}^n . Der entscheidende Unterschied ist jedoch, dass es im Allgemeinen Unterräume gibt, die kein (endliches) Erzeugendensystem haben, und damit auch keine Basis. Ihre Dimension bezeichnet man daher als unendlich. Im Falle endlicher Dimension kann man jeden Unterraum mit \mathbb{R}^n identifizieren, indem man die *Koordinaten* bezüglich einer Basis betrachtet.

4.3 Längen und Winkel in Vektorräumen

Das Standard- (oder Euklid'sche) Skalarprodukt im \mathbb{R}^n ist definiert als

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

Daraus ergibt sich die Definition der Länge von x , $\|x\| = \sqrt{(x, x)}$, und des Winkels zwischen x und y : $\cos(\varphi) = (x, y)/(\|x\| \|y\|)$. Wir ändern also die Sichtweise gegenüber dem Vorgehen in der Mittelschule, wo man für $n = 3$ von der Definition $(x, y) = \|x\| \|y\| \cos(\varphi)$ ausgeht und dann die Formel $(x, y) = x_1 y_1 + x_2 y_2 + x_3 y_3$ herleitet. In unserem Zugang muss man daher die sogenannte *Schwarz'sche Ungleichung* beweisen, welche besagt, dass

$$|(x, y)| \leq \|x\| \|y\|.$$

Dies folgt durch Minimierung von

$$0 \leq \|y - \lambda x\|^2 = \|y\|^2 - 2\lambda(x, y) + \lambda^2\|x\|^2$$

bezüglich λ . Die Lösung ist $\lambda = (x, y)/\|x\|^2$, und Einsetzen ergibt die gewünschte Ungleichung.

Es zeigt sich aber, dass es noch andere vernünftige Arten gibt, eine Länge (auch Norm genannt) und Winkel zu definieren. Das Einzige, was man für eine Länge braucht, sind die Axiome (N1) – (N3) (p. 84 im Buch). Dies ist insbesondere von Bedeutung in abstrakten Vektorräumen, aber selbst im \mathbb{R}^n gibt es viele Normen. Die beiden wichtigsten sind die L_1 - und die L_∞ -Norm:

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

Analog kann man ein Skalarprodukt (und damit Winkel) mit Hilfe der Axiome (S1) – (S3) (p. 93 im Buch) definieren. Überraschenderweise gibt es jedoch kein Skalarprodukt, das die L_1 - oder die L_∞ -Norm erzeugt.

Bei jedem Skalarprodukt sind paarweise orthogonale Vektoren $b^{(1)}, \dots, b^{(k)}$ (ungleich dem Nullvektor) automatisch linear unabhängig. Wenn $x = \sum_{i=1}^k c_i b^{(i)}$, d.h. x liegt in $U = \text{span}\{b^{(1)}, \dots, b^{(k)}\}$, dann folgt durch skalare Multiplikation beider Seiten mit $b^{(j)}$, dass $c_j = (x, b^{(j)})/(b^{(j)}, b^{(j)})$. Die Koordinaten von x können in diesem Fall also ganz einfach berechnet werden.

4.4 Ausgleichsrechnung

Wenn y nicht im Unterraum $U = \text{span}\{a^{(1)}, \dots, a^{(k)}\}$ liegt, hat das Gleichungssystem $Ax = y$ keine Lösung. In diesem Fall begnügen wir uns mit einer approximativen Lösung: Wir suchen dasjenige x , für das der Fehler $r = y - Ax$ eine möglichst kleine Länge hat. Wir minimieren also $\|y - Ax\|$ (oder äquivalent $\|y - Ax\|^2$) bezüglich $x \in \mathbb{R}^k$. 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

$$\sum_{i=1}^n (y_i - a_{i1}x_1 - \dots - a_{ik}x_k)^2$$

minimieren, spricht man von der *Methode der kleinsten Quadrate*. Die Lösung kann man analytisch finden durch Nullsetzen der partiellen Ableitungen. Wir bevorzugen jedoch die geometrische Lösung. Aus geometrischer Anschauung ist es offensichtlich, dass $\|y - Ax\|$ minimal wird, falls Ax gleich der orthogonalen Projektion \hat{y} von y auf den Unterraum U ist. Das heisst, es muss $(y - \hat{y}, a^{(i)}) =$

$(r, a^{(i)}) = 0$ für $i = 1, \dots, k$ gelten. Weil \hat{y} in U liegt, können wir \hat{y} mit den $a^{(j)}$'s ausdrücken: $\hat{y} = \sum_{j=1}^k \hat{x}_j a^{(j)}$. Mit der Linearität des Skalarprodukts erhalten wir

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

bzw. in Matrixform

$$(A^T A) \hat{x} = A^T y.$$

Die Lösung dieser sogenannten *Normalgleichungen* ist in jenem Fall eindeutig, wo die $a^{(1)}, \dots, a^{(k)}$ linear unabhängig sind. Zur numerischen Berechnung soll man allerdings nicht dieses Gleichungssystem lösen, sondern die sogenannte *QR-Zerlegung* verwenden, die wir hier nicht besprechen.

5 Lineare Abbildungen

Für eine gegebene $m \times n$ Matrix A können wir die Abbildung $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ betrachten, welche $x \in \mathbb{R}^n$ überführt in $x' = \mathcal{F}(x) = Ax \in \mathbb{R}^m$. Diese Abbildung ist *linear* im Sinne, dass für beliebige $x, y \in \mathbb{R}^n$ und beliebiges $\alpha \in \mathbb{R}$

$$\mathcal{F}(x + y) = \mathcal{F}(x) + \mathcal{F}(y), \quad \mathcal{F}(\alpha x) = \alpha \mathcal{F}(x)$$

gilt. Umgekehrt ist jede lineare Abbildung \mathcal{F} von \mathbb{R}^n nach \mathbb{R}^m von der Form $\mathcal{F}(x) = Ax$. Die j -te Spalte der Abbildungsmatrix A ist das Bild $\mathcal{F}(e^{(j)})$ des j -ten Basisvektors $e^{(j)}$ ($e^{(j)}$ hat an der j -ten Stelle eine Eins und sonst lauter Nullen).

Beispiele von linearen Abbildungen sind erstens einmal die Mehrzahl der üblichen Abbildungen in der Geometrie, welche den Ursprung festhalten, wie Geraden- und Ebenenspiegelungen, Drehungen, Parallelprojektionen, Streckungen etc. Nichtlineare differenzierbare Abbildungen von \mathbb{R}^n nach \mathbb{R}^m lassen sich in der Umgebung eines festen Punktes approximieren durch eine lineare Abbildung, mit der Jacobi-Matrix als Abbildungsmatrix.

Zwei Abbildungen $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $\mathcal{G} : \mathbb{R}^m \rightarrow \mathbb{R}^p$ können zusammengesetzt werden: $\mathcal{H} = \mathcal{G} \circ \mathcal{F}$ bildet \mathbb{R}^n nach \mathbb{R}^p ab, und zwar ist $\mathcal{H}(x) = \mathcal{G}(\mathcal{F}(x))$. Wenn $\mathcal{F}(x) = Ax$ und $\mathcal{G}(y) = By$, dann ist $\mathcal{H}(x) = (BA)x$, d.h. die *Zusammensetzung linearer Abbildungen* ist wieder linear, und es werden einfach die entsprechenden Abbildungsmatrizen multipliziert.

Lineare Abbildungen treten auch auf bei einem *Basiswechsel*, d.h. wenn wir statt der Standardbasis $e^{(1)}, \dots, e^{(n)}$ eine andere Basis $t^{(1)}, \dots, t^{(n)}$ wählen. Ein Punkt $x = (x_1, \dots, x_n)^T$ kann dann auch als Linearkombination der $t^{(j)}$ dargestellt werden:

$$x = \sum_{j=1}^n y_j t^{(j)} = Ty.$$

Der Zusammenhang zwischen den Koordinaten (x_1, \dots, x_n) bezüglich der alten Basis und den Koordinaten (y_1, \dots, y_n) bezüglich der neuen Basis ist also linear. Weil die Basisvektoren $t^{(j)}$ linear unabhängig sind, ist T invertierbar und damit gilt auch $y = T^{-1}x$.

Wenn $x' = Ax$ eine lineare Abbildung von \mathbb{R}^n nach \mathbb{R}^n und $x = Ty$ eine Koordinatentransformation im \mathbb{R}^n ist, dann gilt $y' = T^{-1}x' = T^{-1}Ax = T^{-1}ATy$, d.h. in der neuen Basis ist die Abbildungsmatrix gleich $T^{-1}AT$. Durch eine geschickte Wahl einer Basis kann man oft eine einfache Form der Abbildungsmatrix erreichen.

Es gibt enge Beziehungen zwischen dem Lösen von linearen Gleichungen und linearen Abbildungen. Lösen von $Ax = b$ heisst, dass man diejenigen Vektoren $x \in \mathbb{R}^n$ sucht, welche unter \mathcal{F} nach b abgebildet werden. Lösungen gibt es genau dann, wenn b zum Unterraum gehört, der von den Spalten von A aufgespannt wird. Dieser Unterraum heisst daher auch der *Bildraum* von A . Wenn die Lösung von $Ax = b$ eindeutig ist, dann bedeutet das in der Sprache der Abbildungen, dass \mathcal{F} *injektiv* ist: $\mathcal{F}(x) \neq \mathcal{F}(z)$ für $x \neq z$. Dies ist genau dann der Fall, wenn der $\text{Kern}(A) = \{x | Ax = 0\} \subset \mathbb{R}^n$ nur aus dem Nullvektor besteht.

Aus früheren Resultaten folgt, dass die Dimension von $\text{Bild}(A)$ gleich dem Rang von A ist und die Dimension von $\text{Kern}(A)$ gleich n minus Rang von A . Zusätzlich kann man zeigen, dass der Rang von A^T gleich dem Rang von A ist und dass die Unterräume $\text{Bild}(A)$ und $\text{Kern}(A^T)$ senkrecht aufeinanderstehen und zusammen den \mathbb{R}^m aufspannen.

6 Eigenwerte und Eigenvektoren

Viele Probleme vereinfachen sich, wenn eine $n \times n$ Matrix A durch eine Koordinatentransformation in *Diagonalform* gebracht werden kann:

$$T^{-1}AT = \text{diag}(\lambda_1, \dots, \lambda_n), \text{ bzw. } A = T \text{diag}(\lambda_1, \dots, \lambda_n) T^{-1}.$$

Man sieht dann besser, was die lineare Abbildung $x \rightarrow Ax$ macht und man kann beliebige Potenzen von A leicht berechnen, indem man

$$A^k = T \text{diag}(\lambda_1^k, \dots, \lambda_n^k) T^{-1}$$

benutzt. Ebenso kann man *lineare Differentialgleichungssysteme* mit konstanten Koeffizienten lösen, siehe den nächsten Unterabschnitt.

Um eine Matrix zu diagonalisieren, d.h. um T zu finden, betrachten wir zuerst die Gleichung $AT = T \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Gemäss der Definition der Matrixmultiplikation ist dies gleichbedeutend damit, dass $At^{(j)} = \lambda_j t^{(j)}$ für $j = 1, 2, \dots, n$ ist, wobei $t^{(j)}$ der j -te Spaltenvektor von T ist. Das heisst, um A zu diagonalisieren, müssen wir Vektoren t suchen, die bei der Multiplikation mit A nur die Länge, aber nicht die Richtung ändern, sogenannte *Eigenvektoren*. Man nennt eine Zahl λ einen *Eigenwert* von A wenn ein Vektor $x \neq 0$ existiert, so dass $Ax = \lambda x$ gilt. Jedes solche x heisst dann ein Eigenvektor von A zum Eigenwert λ . Da $Ax = \lambda x$ auch geschrieben werden kann als $(A - \lambda I_n)x = 0$, ist λ genau dann ein Eigenwert von A , wenn $\det(A - \lambda I_n) = 0$ ist. Diese Determinante ist ein Polynom vom Grad n in λ und heisst das *charakteristische Polynom* von A . Im Komplexen hat dieses Polynom genau n (eventuell zusammenfallende) Nullstellen. Zur Diagonalisierung von Matrizen muss man daher im Allgemeinen mit komplexen Zahlen arbeiten.

Damit A diagonalisiert werden kann, muss T invertierbar sein, d.h. man braucht n linear unabhängige Eigenvektoren, welche die Spaltenvektoren von T ergeben. Ein wichtiger Satz besagt, dass Eigenvektoren zu verschiedenen Eigenwerten immer linear unabhängig sind. Matrizen mit n verschiedenen Eigenwerten werden als *einfach* bezeichnet. Somit ist jede einfache Matrix diagonalisierbar, zumindest wenn man mit komplexen Zahlen rechnet. Wenn ein Eigenwert λ eine k -fache Nullstelle des charakteristischen Polynoms ist, dann gibt es unter Umständen weniger als k linear unabhängige Eigenvektoren zu λ , und in diesem Ausnahmefall ist die Matrix A auch im Komplexen nicht diagonalisierbar.

Der letzte Sachverhalt lässt sich noch etwas präziser formulieren: Die *algebraische Vielfachheit* eines Eigenwertes λ ist die Vielfachheit als Nullstelle des charakteristischen Polynoms, und die *geometrische Vielfachheit* ist die Dimension des Unterraums $\{x \mid Ax = \lambda x\}$. Die geometrische Vielfachheit ist stets mindestens eins und man kann zeigen, dass sie höchstens gleich der algebraischen Vielfachheit ist. A ist also genau dann diagonalisierbar, wenn für jeden Eigenwert die geometrische und die algebraische Vielfachheit gleich sind (man sagt auch, die Matrix sei *halbeinfach*).

Besonders einfach ist die Situation bei symmetrischen Matrizen A (d.h. $A^T = A$): Dann sind nämlich alle Eigenwerte von A reell, die geometrische Vielfachheit ist stets gleich der algebraischen, und Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal. Daher kann man n orthonormierte Eigenvektoren wählen. Mit dieser Wahl wird die Matrix T orthogonal, und damit gilt $A = T \operatorname{diag}(\lambda_1, \dots, \lambda_n) T^T$.

6.1 Lineare Differentialgleichungssysteme

Ein lineares Differentialgleichungssystem erster Ordnung mit konstanten Koeffizienten sieht in Matrixform so aus:

$$\dot{y}(t) = Ay(t).$$

Wenn A diagonalisierbar ist, dann kann man alle Lösungen wie folgt bestimmen: Man berechnet die Eigenwerte λ_j und Eigenvektoren $u^{(j)}$ von A und hat damit die Darstellung $T^{-1}AT = D$ mit $T = (u^{(1)} \dots u^{(n)})$ und $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. Die transformierte Funktion $x(t) = T^{-1}y(t)$ erfüllt dann $\dot{x}(t) = Dx(t)$, bzw. in Komponenten ausgeschrieben $\dot{x}_j(t) = \lambda_j x_j(t)$. Die allgemeine Lösung ist daher $x_j(t) = c_j e^{\lambda_j t}$ mit beliebigen Koeffizienten c_j . Daraus erhalten wir die allgemeine Lösung des ursprünglichen Systems als

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

Für eine vorgegebene Anfangsbedingung $y(0)$ bestimmt man die Koeffizienten c_j als Lösung von $Tc = y(0)$.

Wenn die Matrix A reell ist, aber komplexe Eigenwerte hat, dann kann man mit folgender Überlegung reelle Lösungen bekommen. Wenn $\lambda = \alpha + i\beta$ ein komplexer Eigenwert von A ist mit komplexem Eigenvektor $u = v + iw$, dann ist auch die zu λ komplex konjugierte Zahl $\bar{\lambda} = \alpha - i\beta$ ein Eigenwert von A . Zudem ist $\bar{u} = v - iw$ ein Eigenvektor zu $\bar{\lambda}$. Gemäss der Euler Formel ist

$$e^{(\alpha \pm i\beta)t} = e^{\alpha t} (\cos(\beta t) \pm i \sin(\beta t)),$$

und die Lösungen

$$c_1 e^{\lambda t} u + c_2 e^{\bar{\lambda} t} \bar{u}$$

sind reell, wenn auch die Koeffizienten komplex konjugiert sind: $c_2 = \bar{c}_1$. Wenn wir c_1 in der Form $a + ib$ schreiben, dann erhält man mit einer kleinen Rechnung die Lösungen

$$e^{\alpha t} (a \cos(\beta t) - b \sin(\beta t))v - e^{\alpha t} (a \sin(\beta t) + b \cos(\beta t))w.$$

(Dahinter steckt einfach die Tatsache, dass die komplex Konjugierte eines Produkts das Gleiche ist wie das Produkt der komplex Konjugierten aller Faktoren).

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 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 den Nullvektor. Das heisst, der Fixpunkt $y(t) \equiv 0$ ist stabil. Wenn der Realteil von

einem λ_j positiv ist, dann geht $\|y(t)\|$ gegen unendlich, ausser wenn wir die Anfangsbedingung so wählen, dass das zugehörige c_j gleich Null ist.

Bei Systemen zweiter Ordnung $\ddot{y}(t) = Ay(t)$ geht man analog vor. Die transformierten Variablen $x(t) = T^{-1}y(t)$ erfüllen dann $\ddot{x}(t) = Dx(t)$, und diese Gleichungen kann man explizit lösen. Man kann auch ein System zweiter Ordnung auf ein System erster Ordnung zurückführen, indem man $y(t)$ und $\dot{y}(t)$ zu einem Vektor $z(t)$ der doppelten Länge zusammenfügt. Dies ist zwar äquivalent und vom Verständnis her meist einfacher, hingegen werden die Berechnung eher komplizierter.