Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Grundlagen der Mathematik II

Lineare Algebra

FS 2017

Dr. Marcel Dettling

Institute for Data Analysis and Process Design

Zurich University of Applied Sciences

CH-8401 Winterthur

1 Lineare Gleichungssysteme

Lineare Gleichungssysteme (LGS) treten in vielen Anwendungen im Alltag, in der Wissenschaft und in der Technik auf, siehe die einführenden Beispiele in der Vorlesung. Sie lassen sich allgemein in der folgenden Form schreiben:

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

Wir haben hier ein System mit m Gleichungen und n Unbekannten $x_1,...,x_n$, welche es zu bestimmen gilt. Die a_{ij} nennen wir Koeffizienten, sie sind in Anwendungen bekannte, skalare Grössen. Ebenso vorgegeben sind die rechten Seiten, d.h. die b_i. Um sich die Schreibarbeit zu sparen, verwendet man oft auch die Notation als sogenannte augmentierte Matrix, welche sich auf die Koeffizienten und die rechte Seite beschränkt.

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}$$

Eine weitere Möglichkeit besteht darin, ein LGS als Matrixgleichung der Form $Ax = b$ zu schreiben. Hierbei ist A die Matrix mit den Koeffizienten, x ist der Spaltenvektor mit den Unbekannten und b ist der Spaltenvektor mit der rechten Seite.

Während wir uns erst in Kapitel 2 um das Lösen von LGS kümmern werden, stellen wir hier einige geometrische Überlegungen zur Bedeutung von LGS und deren Lösbarkeit an. Die Fragestellung in einem LGS besteht nämlich darin, wie die verschiedenen Spaltenvektoren $a^{(j)}$ linear kombiniert werden müssen, um den Zielvektor b zu erzeugen. Die $x_1,...,x_n$ sind dann die Vielfachen, mit denen die Spaltenvektoren multipliziert werden müssen. Wie man sich an einem Beispiel im $\mathbb{R}^2$ leicht überlegt, kann man mit zwei Spaltenvektoren jeden beliebigen Zielvektor erzeugen, solange es sich bei den beiden Spaltenvektoren nicht um den Nullvektor handelt, oder sie auf demselben Strahl liegen. Liegen die beiden Spaltenvektoren indes auf demselben Strahl, so hängt die Lösbarkeit davon ab, ob auch der Zielvektor auf diesem liegt oder nicht. Im ersten Fall gibt es unendlich viele Kombinationen der beiden Spaltenvektoren, welche uns zum Zielvektor bringen. Im zweiten Fall ist es hingegen nicht möglich, zum Zielvektor zu gelangen, da wir quasi „auf dem Strahl gefangen sind“.

Bei einem 3×3-Gleichungssystem kann man sich ähnliche Überlegungen im $\mathbb{R}^3$ machen, die Situation ist allerdings bereits etwas komplizierter. Erneut geht es um die Frage, wie man die Spaltenvektoren linear kombinieren muss, um den Zielvektor zu erzeugen. Zeigen die Spaltenvektoren in drei echt verschiedene

Richtungen des Raumes (d.h. sie liegen weder in einer Ebene, noch auf einem Strahl durch den Nullpunkt, noch handelt es sich um den Nullvektor), so kann man wiederum jeden beliebigen Zielvektor erzeugen und es gibt dazu nur eine einzige Kombinationsmöglichkeit. Liegen die drei Vektoren hingegen in einer Ebene, so hängt die Lösbarkeit des LGS davon ab, ob der Zielvektor auch in dieser Ebene liegt (sodann es unendlich viele Kombinationsmöglichkeiten gibt) oder nicht, wobei es in letzterem Fall keine Lösung gibt, usw. Wir werden die Lösbarkeit von LGS zu einem späteren Zeitpunkt noch vertieft und systematisiert betrachten. Bereits zum jetzigen Zeitpunkt können wir aber festhalten, dass es einen fundamentalen Unterschied zwischen LGS gibt, wo die Spaltenvektoren linear unabhängig sind, und solchen wo dies nicht zutrifft.

2 Gauss-Algorithmus

Der Gauss-Algorithmus ist ein Verfahren zur Bestimmung der Lösungsmenge von LGS. Die Idee besteht darin, das LGS so umzuformen, dass die Lösungsmenge dabei identisch bleibt (man spricht dabei von *äquivalenten* LGS), jedoch einfacher abzulesen ist. Dies ist dann der Fall, wenn die i-te Gleichungs des Systems nur die Unbekannten $x_i, x_{i+1}, ..., x_n$ enthält. Im Idealfall erreicht man dabei die sogenannte *Dreiecksform*, aus welcher die Lösungsmenge durch *Rückwärtseinsetzen* abgelesen wird, siehe das Beispiel in der Vorlesung. Für das äquivalente Umformen von LGS sind zwei Operationen erlaubt. Erstens das Vertauschen von Zeilen, zweitens das Addieren eines Vielfachen einer Gleichung zu sich selbst oder einer anderen Gleichung.

An einem weiteren Beispiel mit $m = n = 3$ erkennen wir, dass beim Durchführen des Gauss-Algorithmus eine *Verträglichkeitsbedingung* entstehen kann. Dies ist der Fall, wenn in einer Zeile des LGS auf der linken Seite alle Koeffizienten eliminiert werden, d.h. nur noch Nullen stehen. Ist in diesem Fall die rechte Seite von Null verschieden, so ergibt sich ein Widerspruch. Das LGS hat keine Lösung, bzw. die Lösungsmenge ist leer. Ist die Verträglichkeitsbedingung hingegen erfüllt, so hat das LGS in diesem Beispiel unendlich viele Lösungen. Wir können beim Rückwärtseinsetzen einen Parameter (d.h. die Unbekannte x_3) frei wählen. Die anderen beiden Unbekannten x_1, x_2 ergeben sich dann aufgrund der Wahl von x_3. Man spricht in diesem Fall von einer 1-parametrigen Schar von Lösungen.

Ein drittes Beispiel bespricht die Situation $m = n = 5$. Beim Anwenden des Gauss-Algorithmus ergeben sich zwei Verträglichkeitsbedingungen. Das heisst auch, dass wir beim Eliminieren nur 3 Pivots finden konnten bzw. mussten. Falls eine der beiden Verträglichkeitsbedingungen verletzt ist, so hat das LGS keine Lösung. Andernfalls können wir zwei Unbekannte frei wählen, d.h. wir haben eine 2-parametrige Schar von Lösungen, insgesamt gibt es also unendlich viele davon. Geometrisch bedeutet dieses Beispiel, dass die 5 zur Verfügung stehenden Spaltenvektoren des $\mathbb{R}^5$ in einem echten Unterraum des $\mathbb{R}^5$ mit Dimension 3 liegen. Je nachdem, ob auch der Zielvektor in diesem Unterraum liegt, gibt es unendlich viele Lösungen zum LGS, oder gar keine. Man beachte, dass einige der hier verwendeten Begriffe (Unterraum, Dimension) erst zu einem späteren Zeitpunkt formell definiert werden.

Die Weiterführung der obigen Beispiele bringt uns zur Zeilenstufenform, dem allgemein formulierten Endschema des Gauss-Algorithmus.

$$\begin{pmatrix} \otimes & * & \cdots & * & * & \cdots & * & * & \cdots & * & c_1 \\ 0 & 0 & \cdots & \otimes & * & \cdots & * & * & \cdots & * & c_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \cdots & 0 & \otimes & \cdots & * & c_r \\ 0 & 0 & \cdots & 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 & c_{r+1} \\ 0 & 0 & \cdots & 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 & c_m \end{pmatrix}$$

Die Dreiecksform ist ein Spezialfall der Zeilenstufenform, nämlich jener wo $m = n = r$. Aus dem Endschema, d.h. der Ausprägung der Zeilenstufenform lassen sich nun allgemeine Aussagen zu Existenz und Eindeutigkeit der Lösung eines LGS treffen. Man beachte, dass dies aufgrund vom Ausgangsschema im Allgemeinen nicht möglich ist. Ein wichtiger Begriff ist der *Rang* r des LGS. Es handelt sich dabei um die Anzahl Nicht-Nullzeilen im Endschema, welche auch der Anzahl Pivots entspricht. Sie entspricht auch der Dimension des Unterraums, welche die Spaltenvektoren der Koeffizienten aufspannen. Es gilt ganz allgemein, dass:

$r \geq 0$, sowie auch $r \leq m$ und $r \leq n$.

Falls $r < m$, d.h. der Rang ist kleiner wie die Anzahl Gleichungen, so existieren in jedem Fall Nullzeilen und damit Verträglichkeitsbedingungen. Falls letztere verletzt sind, so gibt es keine Lösung zum LGS. Gilt hingegen $r = m$, so gibt es keine Nullzeilen und Verträglichkeitsbedingungen und das LGS hat auf jeden Fall eine Lösung, allenfalls unendlich viele. Die folgende Tabelle fasst alle Möglichkeiten nochmals zusammen, in der Vorlesung werden für alle Situationen darüber hinaus auch noch einfache Beispiele präsentiert.

Dimension	# Lösungen	r=m	r=n	r<m	r<n
m=n	0,1,∞	1		0,∞	
m<n	0,∞	∞	-	0,∞	-
m>n	0,1,∞	-	0,1	-	0,∞

Als nächstes wird der Begriff von homogenen und inhomogenen LGS eingeführt. Ein LGS heisst homogen, falls die rechte Seite überall gleich null ist. Andernfalls nennt man es inhomogen. Führt man bei einem homogenen LGS den Gauss-Algorithmus durch, so bleiben die Nullen auf der rechten Seite erhalten. Allfällige Verträglichkeitsbedingungen im Endschema sind also stets erfüllt, somit tritt der Fall „keine Lösung“ bei homogenen LGS nicht auf, sie besitzen also immer (mindestens) eine Lösung.

Wenn man sich nun an einem Beispiel (siehe Vorlesung) überlegt, so hat ein homogenes LGS welches man durch den Gauss-Algorithmus auf Dreiecksform bringt, nur die Lösung $x_1 = x_2 = ... = x_n = 0$. Dies nennt man die triviale Lösung eines homogenen LGS, sie existiert in jedem Fall. Geometrisch entspricht ein homogenes LGS dem Problem, wie man die Spaltenvektoren linear kombinieren muss, um zum Nullvektor zurückzukehren. Falls die Vektoren linear unabhängig sind und somit den gesamten $\mathbb{R}^n$ aufspannen, so ist nur die triviale Lösung möglich (d.h., wer einmal vom Ursprung weggeht, kehrt nie mehr zurück). Sind die Spaltenvektoren hingegen linear abhängig (d.h. sie spannen einen echten Unterraum des $\mathbb{R}^n$ auf), so gibt es unendlich viele Kombinationsmöglichkeiten, um wieder zum Ursprung zurückzukehren.

3 Matrizen

Matrizen kommen bei vielen verschiedenen Anwendungen in der Mathematik zum Einsatz. Wir verwenden sie für die vereinfachte Schreibweise von linearen Gleichungssystemen, zur Beschreibung von linearen Abbildungen und zum Lösen von linearen Differentialgleichungssystemen

3.1 Definition und spezielle Matrizen

Eine $m \times n$-Matrix A ist ein Schema von $m \cdot n$ Zahlen, angeordnet in m Zeilen und n Spalten. Diese $m \cdot n$ Zahlen nennt man *Elemente* der Matrix.

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

Matrizen werden grundsätzlich mit Grossbuchstaben notiert. Das Element der i-ten Zeile und j-ten Spalte wird mit a_{ij} oder $(A)_{ij}$ bezeichnet. Von einer *quadratischen Matrix* spricht man, falls $m = n$. Es folgen einige weitere Definitionen:

a) Zwei Matrizen A und B sind gleich, falls sie die gleiche Anzahl Zeilen und Spalten haben, und alle ihre Elemente identisch sind. Das heisst, es ist $A = B \Leftrightarrow a_{ij} = b_{ij}$ für alle i, j

b) Eine $m \times n$-Matrix heisst *Nullmatrix*, falls jedes Element gleich null ist.

c) Weiter gibt es den Begriff der *oberen Dreiecksmatrix* oder *Rechtsdreiecksmatrix*. Dies bezeichnet quadratische Matrizen, in denen alle Elemente unterhalb der Diagonale gleich null sind, $r_{ij} = 0$ für alle $i > j$.

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

d) Ebenso kann man *untere Dreiecksmatrizen* bzw. *Linksdreiecksmatrizen* definieren. Es handelt sich dabei um quadratische Matrizen, bei welchen alle Einträge oberhalb der Diagonale gleich null sind: $l_{ij} = 0$ für alle $i < j$.

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

e) Eine quadratische Matrix heisst *Diagonalmatrix*, falls höchstens die Elemente auf der Diagonale von null verschieden sind.

$$D = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = diag(5,2,3)$$

f) Die quadratische $n \times n$-Matrix $I_n = diag(1,1,...,1)$ heisst *Einheitsmatrix* oder *Identität*.

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

g) Ein *Spaltenvektor* ist eine $n \times 1$-Matrix. Seine Elemente werden üblicherweise Komponenten genannt. Spaltenvektoren schreibt man per Definition per Kleinbuchstaben.

$$b = \begin{pmatrix} 2 \\ -4 \\ 7 \\ 0 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}$$

Ebenso gibt es *Zeilenvektoren*, das sind $1 \times n$-Matrizen. Auch hier spricht man von Komponenten und verwendet Kleinbuchstaben zur Notation. Ganz wichtig: ein Zeilen- und ein Spaltenvektor sind nicht identisch, auch wenn alle Elemente übereinstimmen. Sie haben nämlich nicht die gleiche Dimension.

$c = \begin{pmatrix} 2 & -4 & 7 & 0 \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & c_3 & c_4 \end{pmatrix}$ und es gilt $b \neq c$.

3.2 Rechnen mit Matrizen

Wir diskutieren an dieser Stelle Addition, skalare Multiplikation und Matrixmultiplikation. Während die ersten zwei Operationen trivial sind, erfordert letztere besondere Aufmerksamkeit. Addieren lassen sich zwei Matrizen genau dann, wenn beide dieselbe Dimension aufweisen, d.h. $m \times n$-Matrizen sind. Auch

die Summe ist dann eine $m \times n$-Matrix, man erhält sie durch elementweise Addition, d.h.:

$$A + B \text{ mit } (A+B)_{ij} = (A)_{ij} + (B)_{ij} .$$

Matrizen mit unterschiedlicher Dimension lassen sich hingegen nicht addieren. Bei der skalaren Multiplikation wird eine $m \times n$-Matrix A mit einer Zahl α multipliziert. Das bedeutet, dass jedes Element der Matrix mit α multipliziert wird:

$$\alpha A \text{ mit } (\alpha A)_{ij} = \alpha \cdot (A)_{ij} .$$

Die Matrixmultiplikation $A \cdot B$ ist hingegen nicht elementweise definiert. Damit sie möglich ist, muss A eine $m \times n$-Matrix und B eine $n \times p$-Matrix sein, d.h. die Spaltendimension der ersten Matrix muss mit der Zeilendimension der zweiten Matrix übereinstimmen. Das Produkt ist eine $m \times p$-Matrix, es übernimmt also die Anzahl Zeilen der ersten und die Anzahl Spalten der zweiten Matrix. Die Elemente sind wie folgt definiert:

$$(AB)_{ij} = \sum_{k=1}^{n} (A)_{ik} (B)_{kj} .$$

Obwohl wir den Begriff noch nicht formal definiert haben, halten wir an dieser Stelle fest, dass $(AB)_{ij}$ sich als Skalarprodukt der i-ten Zeile von A und der j-ten Spalte von B ergibt. Beispiele zur Matrixmultiplikation werden in der Vorlesung vorgerechnet, dabei wird auch das *Schema von Falk* vorgeführt, welches bei der Berechnung hilfreich sein kann. Es verbleibt noch die Frage, weshalb die Matrixmultiplikation auf diese, relativ komplizierte Art und Weise definiert ist. Die Antwort liegt in der Geometrie, bzw. in der Theorie zu linearen Abbildungen. Jede $m \times n$-Matrix definiert eine lineare Abbildung vom $\mathbb{R}^n$ in den $\mathbb{R}^m$, d.h. jeder Vektor aus dem $\mathbb{R}^n$ wird durch die Matrix auf einen Vektor im $\mathbb{R}^m$ abgebildet. Mit der obigen Definition der Matrixmultiplikation können lineare Abbildungen hintereinandergeschaltet werden.

Zentral sind die Rechenregeln für Matrizen, welche im Buch von Nipp/Stoffer als Satz 2.1 angegeben sind. Sie lauten wie folgt:

1) Für $m \times n$-Matrizen A und B gilt das Kommutativgesetz bzgl. der Addition:

$$A + B = B + A$$

2) Für $m \times n$-Matrizen A, B, C gilt das Assoziativgesetz bzgl. der Addition:

$$(A+B)+C = A+(B+C)$$

3) Für jede $m \times n$-Matrix A, $n \times p$-Matrix B und $p \times q$-Matrix C gilt das Assoziativgesetz bzgl. der Multiplikation:

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

4) Für $m \times n$-Matrizen A, B und $n \times p$-Matrizen C, D gelten die Distributivgesetze:

$$\begin{aligned} (A+B)C &= AC + BC \\ A(C+D) &= AC + AD \end{aligned}$$

Hingegen gilt das Kommutativgesetz bzgl. der Multiplikation nicht, d.h. man kann nicht darauf zählen, dass $AB = BA$ gilt. Je nach Dimension der Matrizen existiert das Produkt BA gar nicht, wenn AB existiert. Doch selbst für quadratische Matrizen, wo es sowohl AB wie auch BA gibt, kann $AB \neq BA$ sein. Es gibt aber auch Fälle, wo $AB = BA$ zutrifft. Letzterer Fall bedeutet, dass es keine Rolle spielt, ob zuerst die Abbildung A oder die Abbildung B ausgeführt wird. Dies ist aber nicht in jedem Fall so.

3.3 Lineare Abbildungen

Für eine $m \times n$-Matrix A betrachten wir die Abbildung $F : \mathbb{R}^n \to \mathbb{R}^m$, welche jeden beliebigen Spaltenvektor $x \in \mathbb{R}^n$ überführt in $x' = F(x) = Ax \in \mathbb{R}^m$. Diese Abbildung ist linear, weil für beliebige $x, y \in \mathbb{R}^n$ und beliebiges $\alpha \in \mathbb{R}$ gilt:

$$F(x+y) = F(x) + F(y),\ F(\alpha x) = \alpha F(x).$$

Umgekehrt ist jede lineare Abbildung F vom $\mathbb{R}^n$ in dem $\mathbb{R}^m$ durch eine Matrix A definiert, d.h. es gilt stets $F(x) = Ax$. Beispiele von linearen Abbildungen sind Geraden- und Ebenenspiegelungen, Drehungen, Parallelprojektionen, Streckungen. Zentral ist dabei, dass bei der Abbildung der Ursprung festgehalten wird, andernfalls ist sie nicht linear.

Zwei Abbildungen $F : \mathbb{R}^n \to \mathbb{R}^m$ und $G : \mathbb{R}^m \to \mathbb{R}^p$ können zusammengesetzt werden als $H = G \circ F$, welche von $\mathbb{R}^n \to \mathbb{R}^p$ abbildet. Es gilt dann $H(x) = G(F(x))$, bzw. wenn $F(x) = Ax$ und $G(y) = By$, so ist $H(x) = (BA)x$, d.h. die Zusammensetzung von linearen Abbildungen ist wieder linear und man muss einfach die entsprechenden Abbildungsmatrizen miteinander multiplizieren. Einige anschauliche Abbildungen vom $\mathbb{R}^2$ in den $\mathbb{R}^2$ und die zugehörigen Abbildungsmatrizen werden in der Vorlesung präsentiert.

3.4 Transponierte Matrizen

Die Transponierte A^T einer beliebigen Matrix A erhält man, indem man die Zeilen von A in die Spalten von A^T einfüllt. Ist A eine $m \times n$-Matrix, so ist A^T eine $n \times m$-Matrix.

$$A = \begin{pmatrix} 7 & 8 & 2 \\ 0 & 1 & 3 \end{pmatrix} \quad \Leftrightarrow \quad A^T = \begin{pmatrix} 7 & 0 \\ 8 & 1 \\ 2 & 3 \end{pmatrix}$$

Eine quadratische $n \times n$-Matrix B nennt man symmetrisch, falls $B = B^T$. Wir führen noch einige wichtige Rechenregeln fürs Transponieren auf:

1) $(A^T)^T = A$

2) $(A+B)^T = A^T + B^T$

3) $(AB)^T = B^T A^T$

3.5 Die Inverse einer Matrix

Die Inverse ist nur für quadratische Matrizen definiert, und selbst für quadratische Matrizen existiert sie nicht immer. Geometrisch handelt es sich um die Umkehrabbildung. Es ist leicht nachvollziehbar, dass Drehungen stets rückgängig gemacht werden können, d.h. die inverse Abbildung existiert. Bei Projektionen ist dies hingegen nicht möglich.

Die $n \times n$-Matrix X heisst Inverse der $n \times n$-Matrix A, falls $AX = I_n$ gilt. Existiert zur $n \times n$-Matrix A die Inverse, so nennt man sie *regulär* oder auch *invertierbar*. Andernfalls spricht man von einer *singulären* Matrix. Wir schreiben $X = A^{-1}$. Bestimmen kann man die Inverse, indem man die Matrixgleichung $AX = I_n$ spaltenweise löst. Dies läuft auf das Lösen eines einzigen LGS (mit Koeffizientenmatrix A) für n verschiedene rechte Seiten, nämlich die Einheitsvektoren $e^{(1)}, e^{(2)}, ..., e^{(n)}$ hinaus. Falls eines dieser n LGS keine Lösung hat, so existiert die Inverse nicht. Weiter kann es nur genau eine einzige Inverse geben. Hat nämlich A im Endschema Nullzeilen, so ist sicher in einem dieser LGS eine Verträglichkeitsbedingung verletzt. Sind A, B zwei reguläre $n \times n$-Matrizen, so gelten die folgenden Rechenregeln für die Matrixinversion:

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

2) A^{-1} ist regulär und $(A^{-1})^{-1} = A$

3) I_n ist regulär und $I_n^{-1} = I_n$

4) AB ist regulär und $(AB)^{-1} = B^{-1}A^{-1}$

5) A^T ist regulär und $(A^T)^{-1} = (A^{-1})^T$

Absolut zentral ist auch der Zusammenhang zwischen dem Lösen von LGS $Ax = b$ und der Existenz der Inversen von A. Die folgenden Aussagen sind äquivalent, d.h. wenn eine gilt, so gelten alle anderen auch:

1) A ist invertierbar bzw. regulär.

2) A hat Rang n.

3) $Ax = b$ ist für jede rechte Seite b lösbar.

4) $Ax = 0$ hat nur die triviale Lösung.

Durch Matrixinversion lässt sich auch die Lösung eines LGS $Ax = b$ sehr bequem angeben, nämlich als $x = A^{-1}b$. Allerdings ist dies in der Praxis ein ineffizienter

Weg. Zur Bestimmung der Inversen muss ja auch der Gauss-Algorithmus auf die Matrix A angewandt werden, dies darüber hinaus für n anstatt nur für eine einzige rechte Seite. Danach ist noch die Matrixmultiplikation $A^{-1}b$ nötig.

3.6 Orthogonale Matrizen

Orthogonale Matrizen beschreiben längentreue Abbildungen vom $\mathbb{R}^n$ in den $\mathbb{R}^n$, also beispielsweise Drehungen oder Spiegelungen. Sie sind wichtig für das Eigenwertproblem, die Ausgleichsrechnung, d.h. spätere Anwendungen der Matrizenrechnung. Eine quadratische $n \times n$-Matrix A ist orthogonal, falls:

$$A^T A = I_n .$$

Die Elemente von $A^T A$ sind gerade die Skalarprodukte der Spalten von A. Daraus ergibt sich, dass die Spaltenvektoren von A Länge 1 haben und senkrecht (d.h. orthogonal) zueinander sind. Folgende Rechenregeln gelten, falls A, B zwei orthogonale $n \times n$-Matrizen sind:

1) A ist invertierbar und $A^{-1} = A^T$.

2) A^{-1} ist orthogonal.

3) AB ist orthogonal.

4) I_n ist orthogonal.

4 Determinanten

Die Determinante ist eine für quadratische $n \times n$-Matrizen definierte Funktion, welche jeder Matrix eine Zahl $\in \mathbb{R}$ zuordnet. Diese Zahl charakterisiert die Eigenschaften der Matrix in Bezug auf die Lösbarkeit von LGS, der Existenz der Inversen und der Eigenwerte. Wir schreiben:

$$\det(A) = |A| .$$

Zur Interpretation der Determinante lässt sich sagen, dass sie das Volumen des Spats definiert, welcher von den Spaltenvektoren von A im $\mathbb{R}^n$ aufgespannt werden. Wenn dieses Volumen (und damit die Determinante) gleich null sind, so liegen die Spaltenvektoren in einem Unterraum des $\mathbb{R}^n$ (z.B. auf einer Gerade im $\mathbb{R}^2$, einer Gerade oder Ebene im $\mathbb{R}^3$, …). Dies hat entsprechende, bereits bekannte Auswirkungen auf die Lösbarkeit des durch die Matrix A definierten LGS. Es folgt, dass die Determinante auch die Existenz der Inversen bestimmt. Für reguläre Matrizen A gilt $\det(A) \neq 0$.

Zur Berechnung der Determinante beachte man die Beispiele in der Vorlesung. Im trivialen Fall, d.h. für eine 1×1-Matrix ist die Determinante die entsprechende Zahl selber. Bei einer 2×2-Matrix ist sie das Produkt der Hauptdiagonal minus das Produkt der Nebendiagonale:

$$\det(A) = \det\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = a_1 b_2 - a_2 b_1 .$$

Determinanten von 3×3-Matrizen lassen sich mit der *Regel von Sarrus* (siehe Vorlesung) noch direkt rechnen. Ein allgemeineres Verfahren ist jedoch durch die Entwicklung nach der ersten Spalte und der nachfolgenden Berechnung von Subdeterminanten bestimmt. Es lässt sich auch auf grössere Matrizen übertragen. Wie man leicht bemerkt, wird der Aufwand für eine Handrechnung jedoch rasch sehr gross. Mit den folgenden Aussagen legen wir die Grundlage für eine effiziente Berechnung:

1) Wenn man in einer Matrix 2 Zeilen vertauscht, so ändert die Determinante das Vorzeichen.

2) Addieren eines Vielfachen einer Zeile zu einer anderen Zeile lässt die Determinante unverändert.

3) Multiplizieren einer Zeile mit einer Zahl α erhöht die Determinante um den Faktor α.

Daraus lässt sich folgern, dass durch den Gauss-Algorithmus nur das Vorzeichen, jedoch nicht der Betrag der Determinante verändert wird. Des Weiteren ergibt sich die Determinante einer Dreiecksmatrix als das Produkt ihrer Diagonalelemente. Dies ist darin begründet, dass bei der Entwicklung nach der jeweils ersten Zeile und der Berechnung von Subdeterminanten jeweils viele Terme wegen Multiplikation mit null wegfallen. Somit ist der einfachste Weg zum Berechnen der Determinante von grossen Matrizen die Anwendung des Gauss-Algorithmus, um die Matrix auf Dreiecksform zu bringen. Für 2×2- und 3×3-Matrizen ist eine direkte Berechnung jedoch meist effizienter. Vor diesem Hintergrund ist es wenig erstaunlich, dass die Determinante einer Matrix mit zwei gleichen Zeilen null ist, ebenso natürlich die Determinante einer Matrix mit einer Nullzeile. Es ist offensichtlich, dass in diesem Fällen mindestens eine Null auf der Diagonal im Gauss-Endschema steht. Mehr Einsatz erfordert der Beweis, dass:

$$\det(A) = \det(A^T) .$$

Es braucht dazu die Technik der Elementarmatrizen, welche im Rahmen dieser Vorlesung nicht behandelt wird. Es folgt aber direkt, dass auch Matrizen mit zwei gleichen Spalten oder einer Nullspalte eine Determinante von null haben. Darüber hinaus ist die Determinante der Produktmatrix zweier quadratischer Matrizen gleich dem Produkt der Determinanten. Seien $A, B \in \mathbb{R}^{n\times n}$. Dann gilt:

$$\det(AB) = \det(A)\cdot\det(B) .$$

Daraus folgt auch direkt, dass die Determinante der Inversen von A dem Kehrwert von $\det(A)$ entspricht:

$$\det(A^{-1}) = 1/\det(A) .$$

Singuläre Matrizen (d.h. solche mit $\det(A)=0$) sind nicht invertierbar. Wir fassen alle Ergebnisse zu diesem Thema in einem abschliessenden Satz zusammen. Es sei A eine $n\times n$-Matrix. Dann sind folgende Aussagen äquivalent:

- A ist regulär bzw. invertierbar.
- $\det(A)\neq 0$.
- Im Gauss-Endschema ist $r=n$, wir haben echte Dreiecksform.
- Das LGS $Ax=b$ ist für jede rechte Seite b lösbar.
- Das homogene LGS $Ax=0$ hat nur die triviale Lösung.

5 Vektorräume

Ein Vektorraum ist eine Menge von Elementen, für welche gewisse Rechenregeln bzgl. Addition und Multiplikation gelten. Die prominentesten Beispiel sind die Vektorräume $\mathbb{R}^n$, auf welche wir hauptsächlich fokussieren werden. Es gibt jedoch auch abstraktere Vektorräume, welche wir jedoch nur am Rande erwähnen.

5.1 Definition und Eigenschaften

Damit eine Menge von Elementen V ein Vektorraum ist, sind folgende Rechenvorschriften gefordert:

A1) $a+b=b+a$ für alle $a,b\in V$
A2) $(a+b)+c=a+(b+c)$ für alle $a,b,c\in V$
A3) $a+0=a$ für alle $a\in V$
A4) $a+(-a)=0$ für alle $a\in V$
M1) $\alpha(\beta a)=(\alpha\beta)a$ für alle $\alpha,\beta\in\mathbb{R}$ und $a\in V$
M2) $(\alpha+\beta)a=\alpha a+\beta a$ für alle $\alpha,\beta\in\mathbb{R}$ und $a\in V$
M3) $1\cdot a=a$ für alle $a\in V$

Wie man leicht verifizieren kann, sind alle diese Vorschriften in jedem $\mathbb{R}^n$ erfüllt. Zentral im Zusammenhang mit den Vektorräumen ist der Begriff eines Unterraums.

5.2 Unterräume

Ein Unterraum ist eine nichtleere Teilmenge eines Vektorraums (bzw. im Speziellen des $\mathbb{R}^n$), welcher gegenüber Linearkombinationen seiner Elemente abgeschlossen ist. Die formale Definition verlangt, dass für alle $x,y\in U$ und $\alpha,\beta\in\mathbb{R}$ gilt:

1) $x+y\in U$
2) $\alpha x\in U,\ \beta y\in U$
3) $\alpha x+\beta y\in U$

So bildet zum Beispiel jede Ebene durch den Nullpunkt im $\mathbb{R}^3$ einen Unterraum. Diese lässt sich entweder durch eine Parameterdarstellung $\{sa^{(1)} + ta^{(2)} \mid s,t \in \mathbb{R}\}$ oder als Lösungsmenge einer Gleichung $ax_1 + bx_2 + cx_3 = 0$ angeben. Dies lässt sich sinngemäss auch auf andere Dimensionen übertragen. Es gibt auch komplexer definierte Unterräume des $\mathbb{R}^n$, jedoch ist längst nicht jede Teilmenge des $\mathbb{R}^n$ ein Unterraum, siehe die Beispiele in der Vorlesung. Jeder Unterraum ist jedoch auch ein Vektorraum, und jeder Vektorraum V hat zwei triviale Unterräume, nämlich V und $\{0\}$.

Wir behandeln nun einige spezielle Unterräume welche im Zusammenhang mit dem Lösen von LGS auftreten. So bilden sämtliche Lösungen eines homogenen LGS $Ax = 0$ einen Unterraum. Dieser umfasst alle Vektoren, welche durch die Matrix A auf den Nullvektor abgebildet werden. Man nennt ihn *Kern* der Matrix (bzw. der linearen Abbildung) A:

$$Kern\ A = \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

Um zu sehen, dass dies ein Unterraum ist, nehme man zwei Lösungen a, b zum homogenen LGS, d.h. $Aa = 0$ und $Ab = 0$. Weiter seien $\alpha, \beta \in \mathbb{R}$. Dann gilt stets, dass $\alpha a + \beta b$ auch eine Lösung zum homogenen LGS ist, weil $A(\alpha a + \beta b) = \alpha(Aa) + \beta(Ab) = 0$. Ebenso bildet die Menge aller rechten Seiten b, für welche ein LGS $Ax = b$ eine Lösung hat, einen Unterraum des $\mathbb{R}^n$. Man nennt ihn das *Bild* der Matrix (bzw. linearen Abbildung) A:

$$Bild\ A = \{y \in \mathbb{R}^m \mid Es\ gibt\ ein\ x \in \mathbb{R}^n,\ so\ dass\ y = Ax\}.$$

Auch hier ist es einfach zu erkennen, dass die Definition von einem Unterraum erfüllt ist. Es seinen y, z zwei Vektoren aus dem Bild von A. Dann existieren u, v, so dass $y = Au$ und $z = Av$. Dann ist $y + z = Au + Av = A(u + v)$, somit gehört auch $y + z$ zum Bild von A. Zudem ist $\alpha y = \alpha Au = A(\alpha u)$, somit ist auch αy ein Element aus dem Bild von A.

5.3 Linearkombinationen und Erzeugendensysteme

Eine *Linearkombination* von einer beliebigen Anzahl k Spaltenvektoren $a^{(1)}, ..., a^{(k)}$ aus dem $\mathbb{R}^n$ erhält man, indem man sie mit beliebigen Vielfachen $x_1, ..., x_k \in \mathbb{R}$ multipliziert und hernach die Summe bildet:

$$x_1 a^{(1)} + x_2 a^{(2)} + ... + x_k a^{(k)} \in \mathbb{R}^n$$

Wir hatten den Begriff schon früher im Zusammenhang mit LGS hin und wieder verwendet. Ein wichtiger Typ von Unterraum ist jener, der alle Vektoren enthält, welche sich als Linearkombination von $a^{(1)}, ..., a^{(k)}$ schreiben lassen.

$$U = \{x_1 a^{(1)} + ... + x_k a^{(k)} \mid x_1, ..., x_k \in \mathbb{R}\} = span\left\{a^{(1)}, ..., a^{(k)}\right\}$$

Man nennt U den von $a^{(1)}, ..., a^{(k)}$ aufgespannten oder erzeugten Unterraum, bzw. U ist die lineare Hülle von $a^{(1)}, ..., a^{(k)}$. Dass es sich dabei um einen Unterraum handelt, ist leicht zu sehen (siehe Vorlesung).

Wir kümmern uns als nächstes um die wichtige Frage, ob ein Vektor $c \in \mathbb{R}^n$ zu einem definierten Unterraum gehört. Handelt es sich beim Unterraum um den Kern einer Matrix, so können wir einfach einsetzen und prüfen ob $Ac = 0$ ergibt. Spannender ist der Fall, wo der Unterraum durch $U = span\left\{a^{(1)},...,a^{(k)}\right\}$ gegeben ist. Hier müssen wir prüfen, ob es eine Linearkombination gibt, so dass der Vektor c erzeugt wird. Dies entspricht gerade dem Lösen des LGS $Ax = c$, wobei A die Matrix mit den Spaltenvektoren $a^{(1)},...,a^{(k)}$ ist. Eine Lösung existiert genau dann, wenn $c \in U$ und jede Lösung gibt eine Möglichkeit an, wie man c als Linearkombination der Spaltenvektoren ausdrücken kann. Es gibt genau dann mehrere Lösungen, falls das homogene System $Ax = 0$ nichttriviale Lösungen hat. Oder mit anderen Worten, wenn es Zahlen $x_1,...,x_k$ gibt, welche nicht alle gleich null sind, so dass:

$$x_1 a^{(1)} + ... + x_k a^{(k)} = 0 .$$

In diesem Fall heissen die Spaltenvektoren $a^{(1)},...,a^{(k)}$ *linear abhängig*. Dies bedeutet, dass mindestens einer (allenfalls auch mehrere) Spaltenvektoren „überflüssig sind“. Konkret heisst dies, dass wir den Unterraum U mit einem verkleinerten Set an Vektoren immer noch aufspannen können, also gibt es ein $\ell < k$ mit $U = span\left\{a^{(1)},...,a^{(\ell)}\right\}$. In diesem Fall gibt es auch mindestens einen Vektor $a^{(i)}$, welcher sich als Linearkombination der übrigen Vektoren $a^{(1)},...,a^{(i-1)},a^{(i+1)},...,a^{(k)}$ schreiben lässt.

Der andere Fall ist jener, wo es zum LGS $Ax = c$ genau eine Lösung gibt. Das zugehörige homogene System $Ax = 0$ hat dann nur die triviale Lösung und die Spaltenvektoren $a^{(1)},...,a^{(k)}$ heissen *linear unabhängig*. Ein solches Erzeugendensystem kann nicht verkleinert werden, ohne dass sich die lineare Hülle verändert. Es ist in diesem Sinne minimal und heisst *Basis* des Unterraums U. Ein Unterraum $U \neq \{0\}$ hat viele Basen, aber alle bestehen aus gleich vielen Vektoren. Diese Zahl k nennt man die *Dimension* von U, d.h. $k = \dim(U)$. Nimmt man $\ell > k$ Vektoren aus U, so sind diese stets linear abhängig. Andererseits lässt sich jeder Vektor aus U eindeutig als Linearkombination der Basisvektoren darstellen, dies via die Gleichung $Ax = c$. Die dabei auftretenden Vielfachen (d.h. die Lösung x von $Ax = c$) heissen die *Koordinaten* des Vektors bzgl. der Basis $a^{(1)},...,a^{(k)}$.

5.4 Zusammenhang mit dem Lösen von LGS

Das Bild von A ist die lineare Hülle der Spaltenvektoren der Matrix A. Die Dimension dieses Unterraums des $\mathbb{R}^m$ ist gleich r, dem Rang von A. Ist also $r < m$, d.h. im Endschema des Gauss-Algorithmus existieren Nullzeilen, so ist das Bild ein echter Unterraum mit Dimension $r < m$. Wir können also nicht für jede rechte Seite $b \in \mathbb{R}^m$ eine Lösung finden. Ob das LGS keine oder unendlich viele Lösungen hat, hängt davon ab, ob b in der linearen Hülle der Spaltenvektoren liegt oder nicht. Für alle $m \times n$-Matrizen gilt:

1) $\dim(Kern\ A) + \dim(Bild\ A) = n$
2) $\dim(Bild\ A) = \dim(Bild\ A^T)$, d.h. Zeilenrang=Spaltenrang

5.5 Normen

Eine Norm ordnet jedem Vektor a aus dem Vektorraum V eine Zahl zu, und zwar dessen Länge. Wir schreiben $\|a\|$ für die Norm. Es gibt in der Regel verschiedene Normen für einen Vektorraum V. Folgende Gesetze müssen aber für jede Norm gelten:

1) $\|a\| \geq 0$
2) falls $\|a\| = 0$, so ist $a = 0$.
3) $\|\alpha \cdot a\| = |\alpha| \cdot \|a\|$ für alle $\alpha \in \mathbb{R}$, $a \in V$
4) $\|a + b\| \leq \|a\| + \|b\|$ für alle $a, b \in V$

Beispiele umfassen die Euklid'sche Norm, welche den „normale" (bzw. kanonischen) Längenbegriff im $\mathbb{R}^n$ vermittelt. Man sagt auch, dies sei die $\|\cdot\|_2$-Norm oder L_2-Norm. Die Definition ist wie folgt:

$$\|x\|_2 = \sqrt{x_1^2 + ... + x_n^2}\ .$$

Es gibt jedoch auf dem $\mathbb{R}^n$ auch weitere Normen, wie z.B. die L_1-Norm, auch als *Manhattan Block Distance* bekannt. Die Definition lautet $\|x\|_1 = |x_1| + ... + |x_n|$. Wie der Name schon suggeriert, kann diese Distanzfunktion im $\mathbb{R}^2$ durchaus Sinn machen, wenn man sich in einer Stadt mit einem rechtwinkligen Strassennetz befindet, weil die Distanz zwischen zwei Punkten dann tatsächlich auf diese Art und Weise berechnet werden muss.

5.6 Skalaprodukt

Das Skalarprodukt ordnet jedem Paar von Vektoren x, y aus einem Vektorraum V eine Zahl zu. Wir schreiben $\langle x, y \rangle \in \mathbb{R}$ für das Skalarprodukt. Es charakterisiert, wie die beiden Vektoren zueinander in Beziehung stehen, im kanonischen Fall entspricht dies der Winkelmessung. Es gibt jedoch auf einem Vektorraum V im Allgemeinen verschiedene Skalarprodukte. Folgende Gesetze müssen gelten, damit eine Funktion, welche einem Paar von Vektoren eine Zahl zuordnet, als Skalarprodukt gilt:

1) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ für alle $x, y, z \in V$
2) $\langle \alpha x, y \rangle = \alpha \cdot \langle x, y \rangle$ für alle $\alpha \in \mathbb{R}$; $x, y \in V$
3) $\langle x, y \rangle = \langle y, x \rangle$ für alle $x, y \in V$
4) $\langle x, x \rangle \geq 0$ für alle $x \in V$
5) Falls $\langle x, x \rangle = 0$, so folgt $x = 0$

Das kanonische Skalarprodukt im $\mathbb{R}^n$ ist wie folgt definiert:

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

Wie man leicht überprüfen kann, sind die zuvor aufgestellten Bedingungen für ein Skalarprodukt erfüllt. Dies ist jedoch nicht die einzige Möglichkeit im $\mathbb{R}^n$: Jede Funktion der Art $x^T D y$, wo $D = diag(d_1, ..., d_n)$ eine Diagonalmatrix mit lauter positiven Elementen $d_i > 0$ ist, bildet ebenfalls ein Skalarprodukt. Folgende Bemerkungen sind im Zusammenhang mit dem Skalarprodukt ebenfalls zentral:

$$\|x\|_2 = \sqrt{\langle x, x \rangle}$$

Die Euklid'sche Norm und damit die kanonische Länge eines Vektors im $\mathbb{R}^n$ kann via das kanonische Skalarprodukt im $\mathbb{R}^n$ definiert werden. Wie bereits erwähnt wurde, dient es auch der Winkelmessung:

$$\cos\varphi = \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} = \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle \langle y, y \rangle}}$$

Daraus lässt sich ableiten, dass zwei Vektoren genau dann orthogonal zueinander sind (d.h. rechtwinklig zueinander liegen), falls $\langle x, y \rangle = 0$, d.h. ihr Skalarprodukt gleich null ist.

5.7 Orthgonale Projektion

Eine wichtige Anwendung der Theorie in den letzten beiden Kapiteln ist die *orthogonale Projektion*, siehe auch die Skizzen dazu in der Vorlesung. Mathematisch geht es darum, ein $\lambda \in \mathbb{R}$ zu bestimmen, so dass:

$$\langle z, y \rangle = \langle x - \lambda y, y \rangle = 0 .$$

Die Lösung lässt sich mit den Rechengesetzen für das Skalarprodukt leicht finden:

$$\lambda = \frac{\langle x, y \rangle}{\langle y, y \rangle}$$

Die im Buch von Nipp/Stoffer als Satz 4.5 gekennzeichnete Passage macht einige aus der der Geometrie bereits wohlbekannte Aussagen:

1) $\langle x, y \rangle^2 \le \langle x, x \rangle \langle y, y \rangle$ für alle $x, y \in \mathbb{R}^n$ (Schwarz'sche Ungleichung)
2) $\|x\| = \sqrt{\langle x, x \rangle}$ ist stets eine Norm.
3) falls $\langle x, y \rangle = 0$, so ist $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ (Satz von Pythagoras)

Ebenfalls wird postuliert, dass paarweise orthogonale Einheitsvektoren im $\mathbb{R}^n$ linear unabhängig sind und damit eine (orthonormale) Basis bilden.

6 Ausgleichsrechnung

Mit der Ausgleichsrechnung besprechen wir ein Verfahren zum Umgang mit LGS $Ax = b$, welche keine Lösung haben. Wie wir bereits erkannt haben, ist dies genau dann der Fall, wenn y nicht in der linearen Hülle der Spaltenvektoren $U = span\left\{a^{(1)},...,a^{(k)}\right\}$ liegt. In vielen praktischen Anwendungen ist es dann sinnvoll, eine approximative Lösung zu suchen. Sprich, wir suchen dasjenige x, so dass der Fehlervektor

$$r = y - Ax$$

eine möglichst kleine Länge hat. Wir minimieren also $\|y - Ax\|$ (oder $\|y - Ax\|^2$) bezüglich $x \in \mathbb{R}^k$. Dies nennt man das *Ausgleichsproblem* oder *Ausgleichsrechnung*. Die Lösung hängt offensichtlich von der verwendeten Norm ab. Besonders einfach wird es, wenn die Euklid'sche Norm zum Einsatz kommt. Wir minimieren in diesem Fall:

$$\|r\|_2^2 = \|y - Ax\|_2^2 = \sum_{i=1}^{n} (y_i - a_{i1}x_1 - ... - a_{ik}x_k)^2$$

Man nennt dies auch die *Methode der kleinsten Quadrate*. Dieses Verfahren wurde unabhängig voneinander durch Gauss und Legendre zu Beginn des 19. Jahrhunderts entdeckt. Wie kommt man zur Lösung? Man kann einerseits die Analysis bedienen, den obigen Ausdruck partiell nach den unbekannten $x_1,...,x_k$ ableiten und nullsetzen. Anschaulicher ist die geometrische Überlegung, dass die Länge des Fehlervektors r genau dann minimal ist, falls Ax gerade die orthogonale Projektion von y auf den Unterraum $U = span\left\{a^{(1)},...,a^{(k)}\right\}$ ist. Diese Projektion, d.h. die approximative Lösung, schreiben wir als $\hat{y}$ (siehe Skizzen in der Vorlesung). Weil $\hat{y}$ in U liegt, lässt es sich als Linearkombination der Spaltenvektoren von A schreiben:

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

Wir setzen ins Skalarprodukt ein und verwenden dessen Linearität:

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

Diesen Ausdruck kann man auch einfacher als Matrixgleichung schreiben, er lautet in diesem Fall $(A^T A)\hat{x} = A^T y$. Dieses lineare Gleichungssystem, welches die Lösung zum Ausgleichsproblem liefert, nennt man *Normalengleichungen*. Aber hat es denn auch eine eindeutige Lösung? Die Anschauung besagt uns sofort, dass die Lösung nur dann eindeutig sein kann, falls die Spaltenvektoren $a^{(1)},...,a^{(k)}$ linear unabhängig sind. Ansonsten besteht nämlich Redundanz in der lineare Hülle, und die gesuchte Projektion $\hat{y}$ liesse sich auf verschiedene Art und Weise schreiben. Ein Beispiel in der Vorlesung, wo ein überbestimmtes LGS behandelt wird, welches wegen Messfehlern ohne Lösung bleibt, illustriert das Vorgehen.

7 Das Eigenwertproblem

Die Diskussion von Eigenwerten und Eigenvektoren beschränkt sich auf quadratische Matrizen. Es sei also A eine $n \times n$-Matrix. Man stellt sich beim Eigenwertproblem die Frage, ob es bei der zu A zugehörigen linearen Abbildung Vektoren gibt, welche nur gestreckt, aber nicht gedreht werden. Dies bedeutet, dass sie durch A auf ein Vielfaches ihrer selbst abgebildet werden. So lautet also die formale Definition:

Eine Zahl $\lambda \in \mathbb{C}$ heisst *Eigenwert* zur $n \times n$-Matrix A, falls es einen Vektor $x \in \mathbb{C}^n$ gibt, so dass $Ax = \lambda x$ gilt. Andererseits heisst jeder Vektor $x \neq 0$, für welchen $Ax = \lambda x$ gilt, *Eigenvektor* von A zum Eigenwert λ.

Man beachte, dass Eigenwerte und Eigenvektoren selbst für reelle Matrizen komplex sein können. Nur für symmetrische Matrizen spielt sich das Eigenwertproblem garantiert in den reellen Zahlen ab. Während man in einigen einfachen Fällen die Eigenwerte und Eigenvektoren einfach ablesen kann, so erfolgt die Bestimmung im Allgemeinen mit dem charakteristischen Polynom.

7.1 Berechnung von Eigenwerten und Eigenvektoren

Für die Berechnung der Eigenwerte benützen wir die Eigenschaft, dass x genau dann ein Eigenvektor zum Eigenwert λ ist, falls $Ax - \lambda x = Ax - \lambda I_n x = (A - \lambda I_n)x = 0$ Somit ist λ genau dann ein Eigenwert von A, falls das homogene LGS $(A - \lambda I_n)x = 0$ nichttriviale Lösungen hat. Dies ist der Fall, wenn die Koeffizientenmatrix singulär ist, d.h. $\det(A - \lambda I_n) = 0$. Diesen letzten Ausdruck nennt man das charakteristische Polynom, seine Nullstellen ergeben die Eigenwerte.

Zur Bestimmung der Eigenvektoren setzt man die gefundenen λ ein und löst die entsprechenden, homogenen LGS. Per Definition haben diese mehr als eine Lösung. Oder mit anderen Worten: alle Vektoren, welche in eine bestimmte Richtung zeigen, sind Eigenvektoren zu einer Matrix, unabhängig von ihrer Länge. Es hat sich die Konvention durchgesetzt, dass man als Eigenvektor jenen wählt, dessen Euklid'sche Länge gleich 1 ist.

7.2 Diagonalisierung

Die Diagonalisierung beruht auf der Idee, eine Transformationsmatrix T zu finden, so dass eine beliebige Matrix A auf Diagonalform gebracht werden kann:

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

Dies ist für zahlreichen Anwendungen sehr nützlich, so z.B. für das Potenzieren von Matrizen, sowie auch für das Lösen von linearen Differentialgleichungs-systemen. Es stellt sich die Frage, wie man die Transformationsmatrix T finden kann. Dazu betrachten wir die obige, definierende Gleichung spaltenweise. Somit ergibt sich:

$AT = T \cdot diag(\lambda_1, ..., \lambda_n)$ wird zu $At^{(j)} = \lambda_j t^{(j)}$ für alle $j = 1, ..., n$.

Somit ist direkt ersichtlich, dass die Transformationsmatrix T aus den Eigenvektoren zu A besteht, welche in die Spalten von A eingefüllt werden müssen. Die Diagonalmatrix, welche sich schliesslich ergibt, besteht aus den Eigenwerten zu A. In der Basis der Eigenvektoren lässt sich also jede lineare Abbildung besonders einfach charakterisieren.

7.3 Struktur des Eigenraums

Wie definieren zuerst den Begriff der *algebraischen Vielfachheit (aV)*. Dieser besagt, wie oft ein bestimmter Eigenwert λ_0 in der Linearfaktorzerlegung des charakteristischen Polynoms vorhanden ist. Bzw. mehr umgangssprachlich, mit einer wievielfachen Nullstelle wir es zu tun haben. Es gilt $1 \leq aV \leq n$. Zählt man die Eigenwerte mit ihrer aV, so hat jede $n \times n$-Matrix genau n Eigenwerte. Die *geometrische Vielfachheit (gV)* ist die Dimension des sogenannten Eigenraums, d.h. die Dimension der Lösungsmenge von $Ax = \lambda_0 x$, bzw. $(A - \lambda_o I_n)x = 0$. Es gilt $1 \leq gV \leq aV$.

Wenn wir mit $\lambda_1, ..., \lambda_k$ paarweise verschiedene Eigenwerte der Matrix A betrachten und $u^{(1)}, ..., u^{(k)}$ die zugehörigen Eigenvektoren sind, so sind diese linear unabhängig und bilden die *Eigenbasis* zu A. Sie existiert genau dann, wenn die Summe der geometrischen Vielfachheiten der Dimension n der Matrix entspricht. Matrizen, welche eine Eigenbasis besitzen, sind für die Anwendung wichtig, so dass man für diese Eigenschaft einen eigenen Namen kreiert hat. Eine Matrix A heisst *einfach*, wenn für alle ihre Eigenwerte $aV = gV = 1$ gilt. Ist zumindest noch die Bedingung $\sum aV = \sum gV = n$ erfüllt, so nennt man die Matrix *halbeinfach*. (Halb)einfache Matrizen sind auch stets diagonalisierbar.

Wir schliessen dieses Kapitel mit einigen Bemerkungen zu reellen, symmetrischen Matrizen. Wie bereits früher erwähnt, weisen diese reelle Eigenwerte auf. Weiter sind die Eigenvektoren zu verschiedenen Eigenwerten orthogonal. Solche Matrizen sind zumindest halbeinfach und damit auch diagonalisierbar. Weil die Eigenvektoren orthogonal sind, fällt die Diagonalisierung (bzw. das Bestimmen der Inverse der Transformationsmatrix T) besonders einfach aus: $D = T^T AT$.