Ludwig-Maximilians-Universität München
Department für Computerlinguistik
WS 2010/11 – Hauptseminar "Matrixmethoden in Textmining"
Dozent: Prof. Dr. Klaus Schulz
Referentin: Sarah Söhlemann

22.11.2010

Lineare Gleichungssysteme und die Methode der kleinsten Quadrate

<u>Inhalt der Stunde:</u>

1. **Permutationsmatrizen und Eliminationsmatrizen**
2. **LR-Zerlegung** (vgl. Elden, 3.1)
3. **Sonderfall: symmetrische, positiv definite Matrizen** (vgl. Elden, 3.2)
4. **Störungstheorie und Konditionszahlen** (vgl. Elden, 3.3)
5. **Rundungsfehler bei Gaußscher Elimination** (vgl. Elden, 3.4)
6. **Die Methode der kleinsten Quadrate** (vgl. Elden, 3.6)

---

1. <u>Permutationsmatrix, Eliminationsmatrix</u>

Die <u>Permutations</u>- oder Zeilenvertauschungsmatrix $P_{i,j}$ ist eine Einheitsmatrix, deren Zeilen *i,j* vertauscht sind. Multipliziert man eine Matrix A mit der Permutationsmatrix, vertauscht man dadurch auch die jeweiligen Zeilen von A. Bei einer Pivotisierung im Zuge der Gaußelimination wird dies auf die erweiterte Koeffizientenmatrix A' angewendet. Beispiel:

$$P_{12} \cdot A' = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{bmatrix} = \begin{bmatrix} 4 & 9 & -3 & 8 \\ 2 & 4 & -2 & 2 \\ -2 & -3 & 7 & 10 \end{bmatrix}$$ (vgl. Strang)

Mithilfe der <u>Eliminationsmatrix</u> $E_{i,j}$ kann man das *l*-fache der *j*-ten Zeile von Zeile *i* subtrahieren. Sie ist eine Einheitsmatrix, hat aber an der Stelle (*i*, *j*) statt der Null den Eintrag –*l*. Die Multiplikation wird auf ganze Zeilen des LGS angewendet, d.h. wir verwenden die erweiterte Koeffizientenmatrix A'. Beispiel:

$$EA' = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 10 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 & 2 \\ 0 & 1 & 1 & 4 \\ -2 & -3 & 7 & 10 \end{bmatrix}$$

(vgl. Strang, S. 57)

2. <u>Die LR-Zerlegung (LU Decomposition)</u>

Jede reguläre (n × n)-Matrix A kann zerlegt werden in

$$PA = LU$$

wobei *P* eine Permutationsmatrix ist, *L* eine linksuntere (***l**ower*) Dreiecksmatrix mit Einsen in der Diagonalen und Nullen oberhalb der Diagonalen und *U* eine rechtsobere (***u**pper*) Dreiecksmatrix mit Nullen unterhalb der Diagonalen.

P dient der Teilpivotisierung von *A* (d.h. die Vertauschung der Zeilen der erweiterten Koeffizientenmatrix), so dass in der zu zerlegenden Matrix *A* auf der Hauptdiagonalen keine Nullen stehen und das größte Element der ersten Spalte in die Position (1,1) gerückt wird.
Erster Schritt: $P_1A = L_1A^{(1)}$. L_1 ist die vorläufige linksuntere Dreiecksmatrix nach dem ersten Schritt (die Elemente unterhalb der Diagonalen stellen die Multiplikatoren $m_1 \ldots m_n$ aus der Gaußschen Elimination dar). *I* ist die Einheitsmatrix.

$$\text{mit } L_1 = \begin{pmatrix} 1 & 0 \\ m_1 & I \end{pmatrix}, \qquad m_1 = \begin{pmatrix} m_{21} \\ m_{31} \\ \vdots \\ m_{n1} \end{pmatrix}$$

Teilergebnis:

$$A^{(1)} = \begin{pmatrix} a'_{11} & a'_{12} & \ldots & a'_{1n} \\ 0 & a^{(1)}_{22} & \ldots & a^{(1)}_{2n} \\ \vdots & & & \\ 0 & a^{(1)}_{n2} & \ldots & a^{(1)}_{nn} \end{pmatrix}$$

(Teilmatrix B)

(vgl. Elden, 24)

Diese Schritte werden mit den übrigen Teilmatrizen wiederholt, solange bis man nur noch Nullen unterhalb der Diagonalen (= U) stehen hat.

$$B = \begin{pmatrix} a^{(1)}_{22} & \ldots & a^{(1)}_{2n} \\ \vdots & & \\ a^{(1)}_{n2} & \ldots & a^{(1)}_{nn} \end{pmatrix}.$$

<u>Beweis</u>: Wir nehmen an, dass sich B in $P_BB = L_BU_B$ zerlegen lässt (wie A), woraus sich per Induktionsbeweis PA = LU schließen lässt:

$$U = \begin{pmatrix} a'_{11} & a_2^T \\ 0 & U_B \end{pmatrix}, \qquad L = \begin{pmatrix} 1 & 0 \\ P_Bm_1 & L_B \end{pmatrix}, \qquad P = \begin{pmatrix} 1 & 0 \\ 0 & P_B \end{pmatrix} P_1,$$

$$\text{mit } a_2^T = (a'_{12}\, a'_{13} \ldots a'_{1n}). \qquad \square$$

Beispiel (hier ohne Permutation):

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 3 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -2 \end{pmatrix} = L \cdot R$$

(vgl.Wikipedia)

Mit der LU-Form der Matrix lässt sich schnell der Lösungsvektor *x* errechnen:
$Ax = b$ und $A = LU \rightarrow L\underline{Ux} = b$
Wir ersetzen $Ux = y$ und bekommen: $Ly = b$, aus dem wir durch Vorwärtssubstitution leicht *y* bekommen. Damit lösen wir nun durch Rückwärtssubstitution $Ux = b$ und bekommen *x*.

Der Vorteil gegenüber dem Gaußverfahren liegt darin, dass die Zerlegung von $A = LU$ für jedes beliebige *b* gilt, denn hat man *L* und *U* für *A* einmal gefunden, lässt sich bei wechselndem *b* schnell ein neuer Lösungsvektor *x* berechnen, ohne von vorne anzufangen.

3. <u>Symmetrische, positiv definite Matrizen</u>

Für die LR-Zerlegung von symmetrischen, positiv definiten Matrizen ist keine Pivotisierung nötig. Durch ihre Symmetrie sind auch die zerlegten Dreiecksmatrizen symmetrisch. Daher können sie etwa doppelt so schnell berechnet werden.
Eine beliebige quadratische (n × n)-Matrix *A* ist positiv definit, falls gilt: $x^TAx > 0$, $x \neq 0$.

4.1 Störungstheorie und Konditionszahl

In der Realität gibt es viele Arten von Fehlern (Störungen), die ein errechnetes Ergebnis verfälschen (Modellfehler, Datenfehler, Verfahrensfehler, Rundungsfehler). Störungstheorie (Vorwärtsfehleranalyse): Man ändert die Matrix *A* und den Vektor *b* eines eindeutig lösbaren LGS so ab, dass deren (nun geänderte) Lösung *y* sich an die Lösung des eigentlichen, vorher nicht lösbaren Systems annähert.

Unter dem Konditionsbegriff versteht man die Auswirkung kleinster Fehler in den Eingangsdaten auf die Genauigkeit der Lösung. Ein LGS wird schlecht konditioniert genannt, wenn es auf kleinste Fehler mit starken Abweichungen in der Lösung reagiert, es ist also sehr störungsempfindlich.

Mit der Konditionszahl $\kappa(A)$ einer Matrix *A* gibt man an, wie störempfindlich das LGS ist – je kleiner κ, desto störungsunempfindlicher, desto besser konditioniert ist das LGS. Definition: $\kappa(A) = ||A|| \cdot ||A^{-1}||$ [das Produkt der Norm der Matrix *A* mit der Norm ihrer Inversen A^{-1}]. Wir wenden die euklidische Norm (2-Norm) an.
Bei der Konditionsanalyse betrachtet man den Einfluss von Fehlern in der Eingabe auf das Ergebnis der numerischen Aufgabe. Hier bleiben die Rundungsfehler der numerischen Berechnung unberücksichtigt.

4.2 Rundungsfehleranalyse

Rückwärtsfehleranalyse: Man ändert den Lösungsvektor *x* so ab, dass er zur genauen Lösung eines gestörten LGS wird. Konkret stellt man die Frage: Welche Eingabe hätte mit der exakten Berechnung das Ergebnis des numerischen Verfahrens geliefert?

Bei der Rundungsfehleranalyse betrachtet man den Einfluss von Rundungsfehlern, die während der numerischen Berechnung auftreten, auf das Endergebnis. Fehler in der Eingabe bleiben unberücksichtigt.
Wichtig: kleine Fehler können an den falschen Stellen zu großen Abweichungen im Ergebnis führen.

5. Die Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate (auch: kleinste Fehlerquadrate; engl. *method of least squares*) benutzt man, um ein überbestimmtes lineares Gleichungssystem mittels Normalgleichungen zu lösen.
Konkretes Beispiel: Bestimmung der Elastizitätseigenschaften einer Feder durch verschiedene Gewichte und das Messen der Federlänge nach der Formel: $e + \kappa F = l$.

Messreihe

F	1	2	3	4	5
l	7.97	10.2	14.2	16.0	21.2

überbestimmtes lineares Gleichungssystem

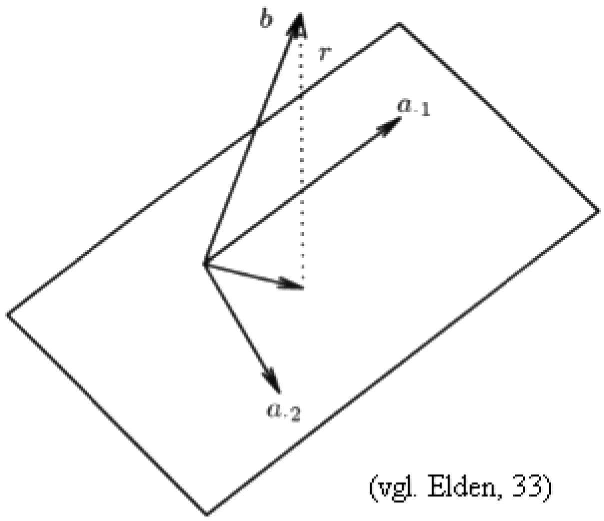
$$e + \kappa 1 = 7.97,$$
$$e + \kappa 2 = 10.2,$$
$$e + \kappa 3 = 14.2,$$
$$e + \kappa 4 = 16.0,$$
$$e + \kappa 5 = 21.2,$$

$$\xrightarrow{\text{Matrixform}} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} e \\ \kappa \end{pmatrix} = \begin{pmatrix} 7.97 \\ 10.2 \\ 14.2 \\ 16.0 \\ 21.2 \end{pmatrix}$$

(vgl. Elden, 31)

Annäherung: wir suchen eine Linearkombination von Vektoren, so dass gilt: $x_1a_1 + x_2a_2 = b$.
Problem: Überbestimmte LGS haben keine Lösung.
Grafische Darstellung: Die Spaltenvektoren $a_{.1}$ und $a_{.2}$ spannen eine Ebene auf. Läge der Vektor *b* in der Ebene, gäbe es eine Linearkombination (s.o.), also eine eindeutige Lösung.

<u>Lösungsversuch</u>: Der Abstand (= Residualvektor *r*) zwischen der Vektorspitze von *b* und einer Linearkombination der Spaltenvektoren (mathematisch: $r = b - Ax$) soll möglichst klein sein. Er ist dann am kleinsten, wenn er orthogonal zu den Spaltenvektoren ist (s. Abb.).



(vgl. Elden, 33)

Orthogonal sind zwei Vektoren dann, wenn ihr inneres Produkt (Skalarprodukt) null ist, in diesem Fall:

$$r^T a_{.j} = 0 \text{ oder } r^T A = 0.$$

Außerdem: $(r^T A)^T = A^T r$ ist ebenfalls 0.

Um eine Normalgleichungen zu bekommen:

$r = b - Ax \rightarrow Ax = b - r$	(umgestellt)
$A^TAx = A^T(b - r)$	(auf beiden Seiten mit A^T multipliziert)
$= A^Tb - A^Tr$	(s.o.: $A^Tr = 0$)
$A^TAx = A^Tb$	(Normalgleichung)

<u>These</u>: sind die Spaltenvektoren von A linear unabhängig, ist die Normalgleichung nichtsingulär/invertierbar und besitzt eine eindeutige Lösung.

<u>Beweis</u>: Zeige, dass A^TA positiv definit ist. Eine Matrix ist dann positiv definit, wenn gilt: $x^TAx > 0$, $x \neq 0$. Außerdem gilt: $Ax = y$ (Multiplikation einer Matrix mit Vektor gibt Vektor).

Also: $x^T\underline{A^TA}x = \underline{A^Tx^T}\,\underline{Ax}$ (umgestellt)
$= y^T y =$ [Vektor mal transponierter Vektor gibt Summe der quadrierten Vektorelemente*] $= \sum_{i=1}^{n} y_i^2 > 0$.

*Beispiel: $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot (x_1 \quad x_2 \quad x_3) = x_1^2 + x_2^2 + x_3^2$ mit i = 3.

Da bei linearer Unabhängigkeit gilt: $Ax \neq 0$ und mit $Ax = y$ folgt: $y \neq 0$, ist die Summe der Quadrate der Vektorelemente von y natürlich > 0. Q.e.d.

Wir wissen nun also, dass A^TA positiv definit ist. Nun müssen wir noch beweisen, dass das $\dot{x}$ aus der Normalengleichung $A^TA\,\dot{x} = A^Tb$ unser Lösungsvektor ist. Dazu beweisen wir, dass $||\dot{r}||_2 \leq ||r||_2$ für alle $r = b - Ax$ (der gesuchte Residualvektor $\dot{r}$ im euklidischen Raum ist kleiner als alle anderen möglichen r).

> Eine Norm ist eine Verallgemeinerung des Begriffs der Länge eines Vektors, also ein reeller Wert, der den Abstand eines Punktes vom Nullelement des jeweiligen Raumes darstellt. Die Euklidische Norm $||x||_2$ (auch 2-Norm genannt) eines Vektors x ist $\sqrt{\sum_{i=1}^{n} |x_i|^2}$ (die Wurzel aus der Summe der quadrierten Beträge der Vektorelemente).
> Die quadrierte Zweinorm: $||x||_2^2 = x^Tx$ (Wurzel fällt weg).

Wir schreiben:

$r = \underline{b - A\dot{x}} + A(\dot{x} - x) = \underline{\dot{r}} + A(\dot{x} - x)$ jeder beliebige Vektor r kann geschrieben werden als der kleinste Vektor $\dot{r}$ plus die Differenz $(\dot{x} - x)$

$$\begin{aligned} ||r||_2^2 &= r^T r && \text{| (s. eukl. Norm)} \\ &= (\dot{r} + A(\dot{x} - x))^T \; (\dot{r} + A(\dot{x} - x)) && \text{| mit obiger Formel} \\ &= \dot{r}^T\dot{r} + \underline{\underline{\dot{r}^T A}}(\dot{x} - x) + (\dot{x} - x)^T \, \underline{\underline{A^T\dot{r}}} + (\dot{x} - x)^T A^T A(\dot{x} - x) && \text{| ausformuliert} \end{aligned}$$

Die doppelt unterstrichenen Terme $\dot{r}^T A$ und $A^T\dot{r}$ sollen Null sein, damit die Orthogonalität gegeben ist. Daher fallen die beiden mittleren Terme weg, übrig bleibt:

$$\begin{aligned} ||r||_2^2 &= \dot{r}^T\dot{r} + (\dot{x} - x)^T A^T A(\dot{x} - x) \text{ oder auch } = \dot{r}^T\dot{r} + \underline{A^T (\dot{x} - x)^T} \, \underline{A(\dot{x} - x)} \\ &= ||\dot{r}||_2^2 + ||A(\dot{x} - x)||_2^2 \qquad \geq ||\dot{r}||_2^2 \qquad \text{q.e.d} \end{aligned}$$

Nun haben wir bewiesen, dass $\dot{r}$ der kleinste Residualvektor für die Normalgleichung $\dot{x}$ = und folglich auch, dass $\dot{x}$ die Lösung unseres Problems des überbestimmten Gleichungssystems ist. Nun müssen wir also nur noch die Normalgleichung nach $\dot{x}$ auflösen und lösen:

$\dot{x} = \frac{A^T b}{A^T A}$ diese „Matrixdivision“ kann in MATLAB mit dem Slash-Operator \ gelöst werden, der eine Gauß-Elimination durchführt. Die MATLAB-Formel sieht dann so aus: >> x = A^TA \ A^Tb (Achtung! Das zuerst genannte steht „unterm Slash“, ist also der Nenner)

Bei der Lösung des Problems der kleinsten Quadrate mittels Normalgleichungen treten zwei Probleme auf:

1. Kann die Multiplikation einer Matrix A mit ihrer inversen A^T zu Informationsverlust führen, wenn A Elemente enthält, die sehr klein, bzw. sich in ihrer Größenordnung stark von den restlichen Elementen unterscheiden (s. Rundungsfehlerproblem in Octave)
2. Ist die Konditionszahl von A^TA per Definition das Quadrat der Konditionszahl von A. A^TA ist also wesentlich schlechter konditioniert → kleinste Fehler führen zu starken Abweichungen in der Lösung.

Quellen:

Lars Elden: Matrix Methods in Data Mining and Pattern Recognition. Society for Industrial and Applied Mathematics (Siam). Philadelphia, 2007.

Gilbert Strang: Lineare Algebra. Springer Verlag Berlin, 2003.

Doreen Seider und Enrico Tappert. Störungstheorie linearer Gleichungssysteme. Berufsakademie Mannheim, Fachrichtung Informationstechnik. <http://ba.db-nico.de/ba_mannheim_informationstechnik_2002/4_semester/seminar/gleichungssysteme.pdf>

„Gaußsches Eliminationsverfahren“. Wikipedia 2010. <http://de.wikipedia.org/wiki/Gau%C3%9Fverfahren> (zuletzt besucht am 21.11.2010)

„Methode der kleinsten Quadrate“. Mathepedia 2010. <http://www.mathepedia.de/Methode_der_kleinsten_Quadrate.aspx> (zuletzt besucht am 21.11.2010)