Matrixzerlegungen, überbestimmte Systeme, Datenkompression

5. Vorlesung

170 004 Numerische Methoden I

Alexander Steinicke

Montanuniversität Leoben

16. November 2023

Matrixzerlegungen, überbestimmte Systeme, Datenkompression

Gliederung 5. Vorlesung

1. Matrixzerlegungen
   - Links-Rechts-Zerlegung
   - Orthogonale Matrizen
   - *QR*-Zerlegung
   - Singulärwertzerlegung
2. Überbestimmte Systeme
   - Einleitung, Beispiele
   - Geometrische Interpretation
   - Rechenweg über *QR*-Zerlegung und SVD
   - Zahlenbeispiel, verschiedene Methoden
3. Inverse, Pseudoinverse, Singlärwertzerlegung
   - Inverses Problem
   - Pseudoinverse via SVD
   - SVD und Datenkompression

Die Matrix-Vektor-Multiplikation $\mathbf{y} = A \cdot \mathbf{x}$

Grundlegende Bedeutung einer $n \times m$-Matrix A

Eine Matrix definiert durch $\mathbf{y} = A \cdot \mathbf{x}$ eine *lineare Abbildung* $\mathbb{R}^m \to \mathbb{R}^n$.

Hauptberuf einer Matrix: aus einem Vektor einen anderen zu machen.

Im Allgemeinen ändert die Matrix dadurch Richtung, Betrag (und sogar Dimension) eines Vektors. Matrixzerlegungen spalten die Aktion der Matrix in leichter zu durchblickende Einzelschritte auf.

- Links-Rechts-Zerlegung $A = L \cdot R$
- QR-Zerlegung $A = Q \cdot R$
- Singulärwertzerlegung $A = U \cdot S \cdot V^T$

Die Matrix-Vektor-Multiplikation $\mathbf{y} = A \cdot \mathbf{x}$

Grundlegende Bedeutung einer $n \times m$-Matrix A

Eine Matrix definiert durch $\mathbf{y} = A \cdot \mathbf{x}$ eine *lineare Abbildung* $\mathbb{R}^m \to \mathbb{R}^n$.

Hauptberuf einer Matrix: aus einem Vektor einen anderen zu machen.

Im Allgemeinen ändert die Matrix dadurch Richtung, Betrag (und sogar Dimension) eines Vektors. Matrixzerlegungen spalten die Aktion der Matrix in leichter zu durchblickende Einzelschritte auf.

- Links-Rechts-Zerlegung $A = L \cdot R$
- QR-Zerlegung $A = Q \cdot R$
- Singulärwertzerlegung $A = U \cdot S \cdot V^T$

Die Matrix-Vektor-Multiplikation $\mathbf{y} = A \cdot \mathbf{x}$

Grundlegende Bedeutung einer $n \times m$-Matrix A

Eine Matrix definiert durch $\mathbf{y} = A \cdot \mathbf{x}$ eine *lineare Abbildung* $\mathbb{R}^m \to \mathbb{R}^n$.

Hauptberuf einer Matrix: aus einem Vektor einen anderen zu machen.

Im Allgemeinen ändert die Matrix dadurch Richtung, Betrag (und sogar Dimension) eines Vektors. Matrixzerlegungen spalten die Aktion der Matrix in leichter zu durchblickende Einzelschritte auf.

- Links-Rechts-Zerlegung $A = L \cdot R$
- QR-Zerlegung $A = Q \cdot R$
- Singulärwertzerlegung $A = U \cdot S \cdot V^T$

Wiederholung: LR-Zerlegung $A = L \cdot R$

Das Gaußsche Eliminationsverfahren ohne Pivotisierung faktorisiert (wenn es nicht abbricht) eine Matrix A in ein Produkt $A = LR$ aus einer linken unteren Dreiecksmatrix L und einer rechten oberen Dreiecksmatrix R.

$$\begin{bmatrix} 5 & 6 & 7 \\ 10 & 20 & 23 \\ 15 & 50 & 67 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 5 & 6 & 7 \\ 0 & 8 & 9 \\ 0 & 0 & 10 \end{bmatrix}$$

LR-Zerlegung allgemein

Jede quadratische oder auch rechteckige Matrix lässt sich in ein Produkt der Form

$$A = L \cdot R$$

zerlegen. Dabei ist

- L eine links-unten-Dreiecksmatrix mit 1-Diagonale (zumindest, wenn man die Zeilen von L passend umordnet)
- R eine obere Dreiecksmatrix

MATLAB: `[L U]=lu(A)` zerlegt beispielsweise

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 5 & 6 \\ 0 & \frac{3}{4} & \frac{3}{2} \end{bmatrix}$$

Anwendungen zur LR-Zerlegung

- lineare Gleichungssysteme
- Determinante
- Inverse

MATLAB-Hilfe: "*Most of the algorithms for computing LU factorization are variants of Gaussian elimination. The factorization is a key step in obtaining the inverse with inv and the determinant with det. It is also the basis for the linear equation solution obtained with* \"

Orthogonale Matrizen

Transponierte ist zugleich Inverse

Definition

Eine quadratische Matrix Q heißt orthogonal, wenn gilt

$$Q^T \cdot Q = I$$

Die Spalten von Q sind Einheitsvektoren und paarweise orthogonal. Es gilt dann auch

$$Q \cdot Q^T = I$$

Die Zeilen von Q sind ebenfalls Einheitsvektoren und paarweise orthogonal.

Orthogonale Matrizen: Beispiele

$$P = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad Q = \frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix}$$

$$\left.\begin{matrix} \text{Zwei-} \\ \text{Drei-} \end{matrix}\right\} \text{dimensionale orthogonale Matrizen mit Determinante 1}$$

$$\text{entsprechen} \left\{\begin{matrix} \text{ebenen} \\ \text{räumlichen} \end{matrix}\right\} \text{Drehungen}$$

Orthogonale Matrizen mit Determinante -1 entsprechen Spiegelungen.

Orthogonale Matrizen: Beispiele

$$P = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad Q = \frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix}$$

$$\left.\begin{array}{c}\text{Zwei-}\\ \text{Drei-}\end{array}\right\} \text{dimensionale orthogonale Matrizen mit Determinante 1}$$

$$\text{entsprechen} \left\{\begin{array}{c}\text{ebenen}\\ \text{räumlichen}\end{array}\right\} \text{Drehungen}$$

Orthogonale Matrizen mit Determinante -1 entsprechen Spiegelungen.

Fundamentale Eigenschaft

2-Norm bleibt unverändert

Multiplikation eines Vektors mit einer orthogonalen Matrix dreht (oder spiegelt) den Vektor, lässt aber seine Länge (2-Norm) *unverändert.*

$$\|\mathbf{x}\|_2 = \|\mathbf{Q} \cdot \mathbf{x}\|_2$$

QR-Zerlegung

Satz

Jede reelle $n \times m$ Matrix lässt sich in ein Produkt der Form

$$A = Q \cdot R$$

mit einer orthogonalen $n \times n$ Matrix Q und einer $n \times m$ oberen Dreiecksmatrix R zerlegen.

MATLAB: `[Q R]=qr(A)`
Anwendungen: Eigenwert-Berechnung, überbestimmte Systeme

Singulärwert-Zerlegung

Singular Value Decomposition, SVD

Satz

Jede reelle $n \times m$ Matrix lässt sich in ein Produkt der Form

$$A = U \cdot S \cdot V^T$$

mit einer orthogonalen $n \times n$ Matrix U, einer $n \times m$-Diagonalmatrix S und einer orthogonalen $m \times m$ Matrix V zerlegen.
Die Diagonalwerte von S heißen die Singulärwerte von A.

MATLAB: `[U S V]=svd(A)`

Anwendungen

Pseudoinverse, über- und unterbestimmte Gleichungssysteme, Datenanalyse, -komprimierung, -clustering, etc. etc.

Singulärwert-Zerlegung

Anschauliche Interpretation

Die Singulärwertzerlegung $A = U \cdot S \cdot V^T$ zerlegt die Multiplikation $A \cdot \mathbf{x}$ in drei Einzelschritte:

- $\mathbf{a} = V^T \cdot \mathbf{x}$ dreht den Vektor $\mathbf{x}$
- $\mathbf{b} = S \cdot \mathbf{a}$ skaliert die Komponenten von $\mathbf{a}$
- $\mathbf{y} = U \cdot \mathbf{b}$ dreht den Vektor $\mathbf{b}$

Abgesehen von den beiden Drehungen beschreibt die Diagonalmatrix S, was die die Matrix A „eigentlich“ tut.

In geeignet gedrehten Koordinaten wird aus A eine Diagonalmatrix S. Der richtige Dreh vereinfacht viele Aufgaben.

Singulärwert-Zerlegung

Anschauliche Interpretation

Die Singulärwertzerlegung $A = U \cdot S \cdot V^T$ zerlegt die Multiplikation $A \cdot \mathbf{x}$ in drei Einzelschritte:

- $\mathbf{a} = V^T \cdot \mathbf{x}$ dreht den Vektor $\mathbf{x}$
- $\mathbf{b} = S \cdot \mathbf{a}$ skaliert die Komponenten von $\mathbf{a}$
- $\mathbf{y} = U \cdot \mathbf{b}$ dreht den Vektor $\mathbf{b}$

Abgesehen von den beiden Drehungen beschreibt die Diagonalmatrix S, was die die Matrix A „eigentlich“ tut.

In geeignet gedrehten Koordinaten wird aus A eine Diagonalmatrix S. Der richtige Dreh vereinfacht viele Aufgaben.

Gliederung 5. Vorlesung

Überbestimmte Systeme

Ein lineares Gleichungssystem mit mehr Gleichungen als Unbekannten heißt überbestimmt

Siehe die einführenden Folien am Ende der 4. Vorlesung!

$$A\mathbf{x} = \mathbf{b} \quad \text{mit einer } m \times n\text{-Matrix } A, \text{ wobei } m > n$$

- In der Regel hat ein solches System keine exakte Lösung.
- Kompromiss-Lösung sucht möglichst kleinen Residuenvektor

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}$$

- Die Näherungs mit $\|\mathbf{r}\|_2 \to \min!$ heißt kleinste-Quadrate-Lösung.
- Klassisches Rechenverfahren: Normalengleichungen

$$A^T A\mathbf{x} = A^T \mathbf{b}$$

- Moderne, oft bessere Verfahren: QR-, Singulärwert-Zerlegung.

Übersicht der Verfahren

für überbestimmte Systeme $A \cdot \mathbf{x} = \mathbf{b}$

Normalengleichungen $A^T \cdot A = A^T \mathbf{b}$ Klassischer Lösungsweg. Anfällig für Daten- und Rundungsfehler (schlechte Konditionszahl).

QR-Zerlegung Standardverfahren zur numerischen Lösung. Algebraisch äquivalent zu Norm'gleichungen, aber numerisch weniger fehlerempfindlich.

MATLAB `x = A\b` verwendet bei überbestimmten Systemen automatisch QR-Zerlegung

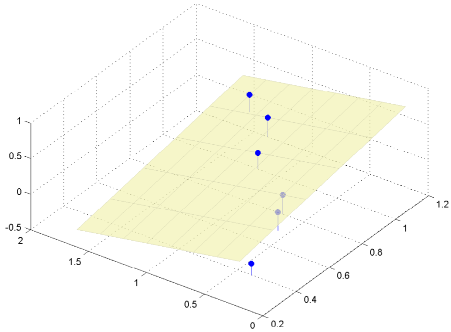
Sonderfälle (rang $A \leq n$) Überbestimmte Systeme, die trotzdem exakt lösbar sind oder eine Schar von (ex. oder kl. Quadr.) Lösungen haben. Normalengleichungen können versagen. Methode: Singulärwert-Zerlegung.

Lineares Modell in zwei Variablen

Anpassen einer Ausgleichs-Ebene: Beispiel aus der Matlab-Hilfe

Angenommen, eine Größe y hängt von zwei Parametern x_1 und x_2 ab. Folgende Messwerte liegen vor:

x_1 :	0.2	0.5	0.6	0.8	1.0	1.1
x_2 :	0.1	0.3	0.4	0.9	1.1	1.4
y :	0.17	0.26	0.28	0.23	0.27	0.24



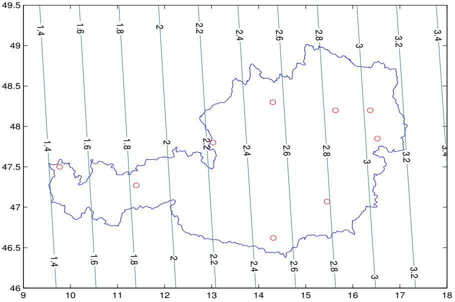
Wir nehmen ein lineares Modell $y = a_0 + a_1x_1 + a_2x_2$ an und setzen die gegebenen Datentripel ein

$\longrightarrow$ führt auf ein System von 6 linearen Gleichungen in den 3 unbekannten Koeffizienten a_0, a_1, a_2.

$\longrightarrow$ Kleinste-Quadrate-Lösung liefert Ebene mit „bestmöglicher“ Anpassung an Daten

Lineares Modell in zwei Variablen

Beispiel: Magnetische Deklinationswerte 2008.5 in Österreich



Wien	3°	00’
Eisenstadt	3°	02’
St.Pölten	2°	54’
Graz	2°	47’
Linz	2°	33’
Klagenfurt	2°	30’
Salzburg	2°	15’
Innsbruck	1°	53’
Bregenz	1°	24’

Daten: ZAMG

Kleinste-Quadrate-Anpassung

liefert Modell:

$$\delta = -2.0987 + 0.2365\lambda + 0.0261\phi$$

$A\mathbf{x} = \mathbf{b}$ Zwei mögliche geometrische Interpretationen

Im Raum $\mathbb{R}^n$ der Lösungsvektoren

- Gleichungen entsprechen Geraden (Ebenen, Hyperebenen)
- Optimal-Lösung im Schnittbereich (Fehlerdreieck, -Polyeder, ...)
- Nachteil: „Mitte des Fehlerdreiecks“ nicht exakt definiert.

Im Raum $\mathbb{R}^m$ der rechten-Seite-Vektoren

- Matrix mal Vektor ergibt Linearkombination der Spaltenvektoren
- Bei „zu wenig“ Spaltenvektoren lässt sich nicht jede rechte Seite als Linearkombination erreichen.
- Der geringste Abstand zwischen Linearkombination und rechter Seite ist Normalabstand.

$A\mathbf{x} = \mathbf{b}$ Zwei mögliche geometrische Interpretationen

Im Raum $\mathbb{R}^n$ der Lösungsvektoren

- Gleichungen entsprechen Geraden (Ebenen, Hyperebenen)
- Optimal-Lösung im Schnittbereich (Fehlerdreieck, -Polyeder, ...)
- Nachteil: „Mitte des Fehlerdreiecks“ nicht exakt definiert.

Im Raum $\mathbb{R}^m$ der rechten-Seite-Vektoren

- Matrix mal Vektor ergibt Linearkombination der Spaltenvektoren
- Bei „zu wenig“ Spaltenvektoren lässt sich nicht jede rechte Seite als Linearkombination erreichen.
- Der geringste Abstand zwischen Linearkombination und rechter Seite ist Normalabstand.

$A\mathbf{x} = \mathbf{b}$ Zwei mögliche geometrische Interpretationen

Im Raum $\mathbb{R}^n$ der Lösungsvektoren

- Gleichungen entsprechen Geraden (Ebenen, Hyperebenen)
- Optimal-Lösung im Schnittbereich (Fehlerdreieck, -Polyeder, ...)
- Nachteil: „Mitte des Fehlerdreiecks“ nicht exakt definiert.

Im Raum $\mathbb{R}^m$ der rechten-Seite-Vektoren

- Matrix mal Vektor ergibt Linearkombination der Spaltenvektoren
- Bei „zu wenig“ Spaltenvektoren lässt sich nicht jede rechte Seite als Linearkombination erreichen.
- Der geringste Abstand zwischen Linearkombination und rechter Seite ist Normalabstand.

$A\mathbf{x} = \mathbf{b}$ Zwei mögliche geometrische Interpretationen

Im Raum $\mathbb{R}^n$ der Lösungsvektoren

- Gleichungen entsprechen Geraden (Ebenen, Hyperebenen)
- Optimal-Lösung im Schnittbereich (Fehlerdreieck, -Polyeder, ...)
- Nachteil: „Mitte des Fehlerdreiecks“ nicht exakt definiert.

Im Raum $\mathbb{R}^m$ der rechten-Seite-Vektoren

- Matrix mal Vektor ergibt Linearkombination der Spaltenvektoren
- Bei „zu wenig“ Spaltenvektoren lässt sich nicht jede rechte Seite als Linearkombination erreichen.
- Der geringste Abstand zwischen Linearkombination und rechter Seite ist Normalabstand.

Überbestimmte Systeme: geometrische Interpretation 2

$$\begin{array}{ccccc} x & & & = & 1 \\ & & y & = & 2 \\ x & + & y & = & 4 \end{array} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

Matrix-Vektor-Multiplikation ergibt Linearkombination der Spaltenvektoren

Hier: kann x mal erster plus y mal zweiter Spaltenvektor rechte Seite ergeben?

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot x + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \cdot y = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

Die Linearkombinationen von zwei Vektoren im Raum spannen eine Ebene auf. Nur wenn die rechte Seite ein Punkt dieser Ebene ist, gibt es eine Lösung für x und y.

Überbestimmte Systeme: geometrische Interpretation 2

$$\begin{array}{ccccc} x & & & = & 1 \\ & & y & = & 2 \\ x & + & y & = & 4 \end{array} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

Matrix-Vektor-Multiplikation ergibt Linearkombination der Spaltenvektoren

Hier: kann x mal erster plus y mal zweiter Spaltenvektor rechte Seite ergeben?

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot x + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \cdot y = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

Die Linearkombinationen von zwei Vektoren im Raum spannen eine Ebene auf. Nur wenn die rechte Seite ein Punkt dieser Ebene ist, gibt es eine Lösung für x und y.

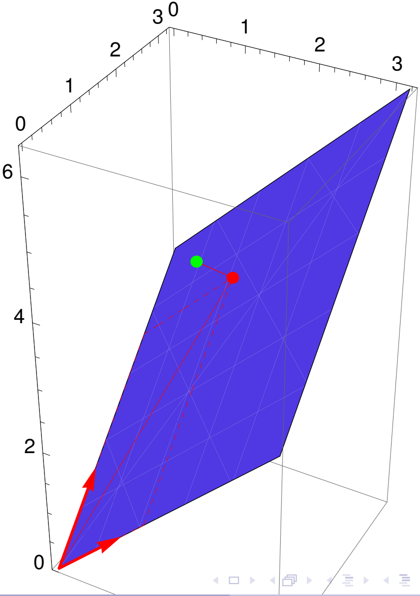
Orthogonalitätsbedingung im Bildraum

Der Ausdruck $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$ entspricht der Parameterdarstellung eine Ebene im $\mathbb{R}^3$.

Der Punkt $(1; 2; 4)$ liegt nicht auf dieser Ebene. Der Residuenvektor

$$\mathbf{r} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

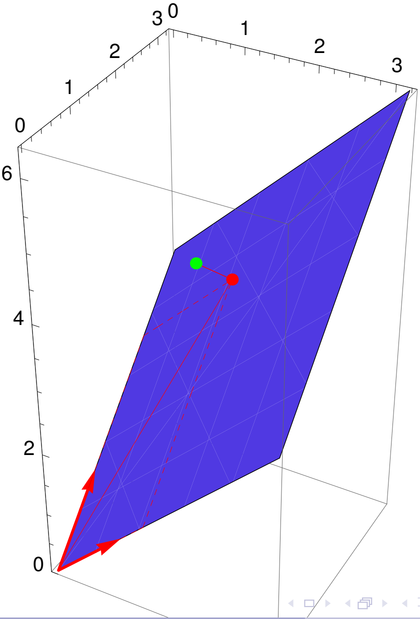
hat minimale Länge, wenn er auf die Ebene normal steht.



Drehe den Bildraum, dann geht's leichter!

Wenn die zwei die Ebene aufspannenden Vektoren irgendwo hin in den $\mathbb{R}^3$ zeigen, ist der Punkt mit minimalem Abstand von $\mathbf{b}$ nicht so leicht zu finden...

Drehen wir vorsichtig das Koordinatensystem... bis die zwei Vektoren in der xy-Ebene liegen. Die Lösung liegt nun genau senkrecht unter $\mathbf{b}$

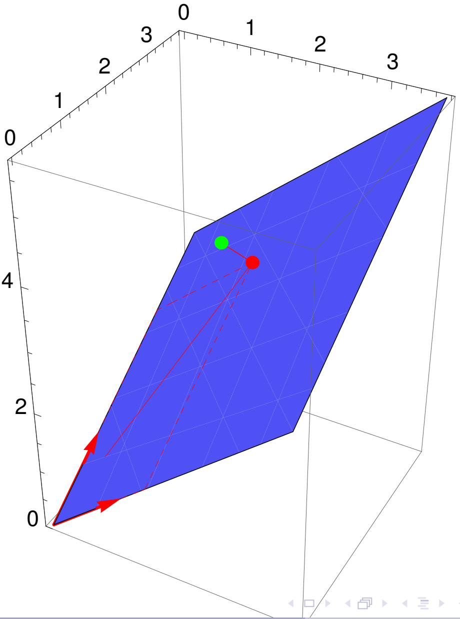


Drehe den Bildraum, dann geht's leichter!

Wenn die zwei die Ebene aufspannenden Vektoren irgendwo hin in den $\mathbb{R}^3$ zeigen, ist der Punkt mit minimalem Abstand von **b** nicht so leicht zu finden...

Drehen wir vorsichtig das Koordinatensystem...

bis die zwei Vektoren in der *xy*-Ebene liegen. Die Lösung liegt nun genau senkrecht unter **b**

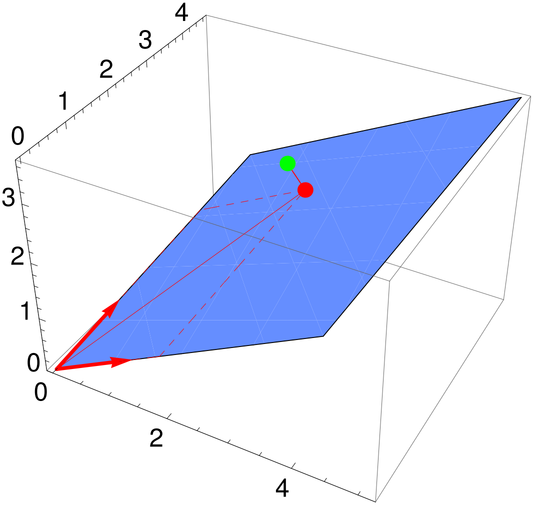


Drehe den Bildraum, dann geht's leichter!

Wenn die zwei die Ebene aufspannenden Vektoren irgendwo hin in den $\mathbb{R}^3$ zeigen, ist der Punkt mit minimalem Abstand von $\mathbf{b}$ nicht so leicht zu finden...

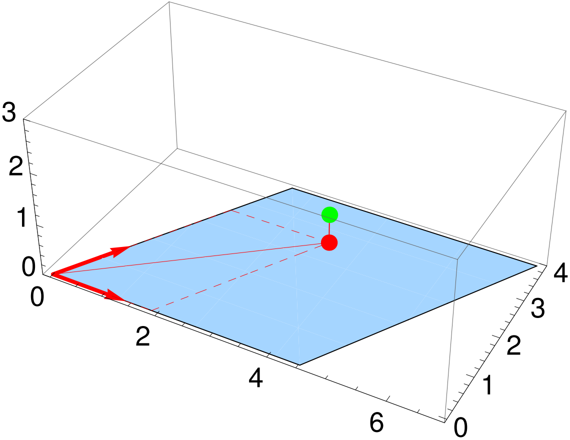
Drehen wir vorsichtig das Koordinatensystem...

bis die zwei Vektoren in der xy-Ebene liegen. Die Lösung liegt nun genau senkrecht unter $\mathbf{b}$



Drehe den Bildraum, dann geht's leichter!

Wenn die zwei die Ebene aufspannenden Vektoren irgendwo hin in den $\mathbb{R}^3$ zeigen, ist der Punkt mit minimalem Abstand von **b** nicht so leicht zu finden. . .



Drehen wir vorsichtig das Koordinatensystem. . . bis die zwei Vektoren in der xy-Ebene liegen. Die Lösung liegt nun genau senkrecht unter **b**

Suche eine orthogonale Matrix Q, welche das Gleichungssystem in einen „einfacheren" Bildraum dreht!

Beispiel von vorhin

Die QR-Zerlegung von A liefert die gewünschte Drehmatrix!

Zerlegung $A = QR$: $$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & -1 & -\sqrt{2} \\ 0 & 2 & -\sqrt{2} \\ \sqrt{3} & 1 & \sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{2} & \sqrt{2}/2 \\ 0 & \sqrt{3/2} \\ 0 & 0 \end{bmatrix}$$

Transformiertes System $R\mathbf{x} = Q^T\mathbf{b}$: $$\begin{bmatrix} \sqrt{2} & \sqrt{2}/2 \\ 0 & \sqrt{3/2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5/\sqrt{2} \\ 7/\sqrt{6} \\ 1/\sqrt{3} \end{bmatrix}$$

Drehung mit Q anschaulich interpretiert

QR-Zerlegung transformiert überbestimmtes System in lösbaren Anteil und unlösbaren Rest

Bestmögliche Kompromiss-Lösung: unlösbare Gleichungen weglassen, die anderen exakt lösen.

QR-Zerlegung „löst“ Gleichungssysteme

Löst nicht exakt, aber nähert bestmöglich in 2-Norm

Zur bestmöglichen Näherungs-Lösung von $A\mathbf{x} = \mathbf{b}$ führt man QR-Zerlegung durch und „löst“ das Dreieckssystem $R\mathbf{x} = Q^T \cdot \mathbf{b}$.

Begründung

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ Q \cdot R\mathbf{x} &= \mathbf{b} && |Q^T \cdot \\ Q^T \cdot Q \cdot R\mathbf{x} &= Q^T \cdot \mathbf{b} \\ R\mathbf{x} &= Q^T \cdot \mathbf{b} \end{aligned}$$

Bei $n \times n$-Systemen setzt man diese Methode normalerweise nicht ein, weil sie rechenaufwändiger als die LR-Zerlegung ist.
Bei $n \times m$-Systemen ($n > m$) liefern die ersten m Zeilen von R die kleinste-Quadrate-Lösung.

Der Rechenweg über die Normalengleichungen ist oft anfälliger für Rundungsfehler als der über QR-Zerlegung.

Überbestimmte Systeme $A \cdot \mathbf{x} = \mathbf{b}$: QR-Zerlegung

Zusammenfassende mathematische Begründung

Aufgabe

Suche $\mathbf{x}$ so, dass Residuums-2-Norm $\|\mathbf{r}\|_2$ minimal wird

$$\|\mathbf{r}\|_2 = \|\mathbf{b} - A \cdot \mathbf{x}\|_2 \to \text{min!}$$

Vorgangsweise

Zerlege $A = Q \cdot R$
Multiplikation von $\mathbf{r}$ mit Q^T lässt 2-Norm unverändert. Wandle um:

$$\|\mathbf{r}\|_2 = \|Q^T \cdot \mathbf{r}\|_2 = \|Q^T(\mathbf{b} - A \cdot \mathbf{x})\|_2 = \|Q^T\mathbf{b} - R \cdot \mathbf{x}\|_2$$

Im transformierten System ist die Minimallösung direkt ablesbar...

Angenommen, das überbestimmte System besteht aus m Unbekannten und $n = m + k$ Gleichungen ($k > 0$). Dann ist R eine $n \times m$-Matrix, deren letzte k Zeilen lauter Nullen enthalten.

Die letzten k Zeilen des Residuumsvektors hängen nicht von $\mathbf{x}$ ab.

Löse die ersten m Gleichungen des Systems $R \cdot \mathbf{x} = Q^T\mathbf{b}$ exakt (wenn möglich): Diese Gleichungen liefern keinen Beitrag zu Residuum.

Die restlichen k Gleichungen hängen von $\mathbf{x}$ nicht ab. Keine Wahl von $\mathbf{x}$ kann den Beitrag dieser Gleichungen zum Residuum ändern.

Daher ist die gewählte Lösung optimal für $R \cdot \mathbf{x} = Q^T\mathbf{b}$

Weil die Q^T-Transformation die Norm nicht beeinflusst, ist diese Lösung auch optimale Lösung von $A \cdot \mathbf{x} = \mathbf{b}$

Sie könnten dieselbe Argumentation auch mit LR-Zerlegung versuchen. Auch dabei erhalten sie eine R-Matrix mit lauter Nullen in den letzten k Zeilen. Aber die Transformation mit der L-Matrix ändert die Norm des Residuumsvektors. Eine optimale Lösung des transformierten Systems ist dann keine Optimallösung des Originalsystems.

Angenommen, das überbestimmte System besteht aus m Unbekannten und $n = m + k$ Gleichungen ($k > 0$). Dann ist R eine $n \times m$-Matrix, deren letzte k Zeilen lauter Nullen enthalten.
Die letzten k Zeilen des Residuumsvektors hängen nicht von $\mathbf{x}$ ab.
Löse die ersten m Gleichungen des Systems $R \cdot \mathbf{x} = Q^T\mathbf{b}$ exakt (wenn möglich): Diese Gleichungen liefern keinen Beitrag zu Residuum.
Die restlichen k Gleichungen hängen von $\mathbf{x}$ nicht ab. Keine Wahl von $\mathbf{x}$ kann den Beitrag dieser Gleichungen zum Residuum ändern.
Daher ist die gewählte Lösung optimal für $R \cdot \mathbf{x} = Q^T\mathbf{b}$
Weil die Q^T-Transformation die Norm nicht beeinflusst, ist diese Lösung auch optimale Lösung von $A \cdot \mathbf{x} = \mathbf{b}$

Sie könnten dieselbe Argumentation auch mit LR-Zerlegung versuchen. Auch dabei erhalten sie eine R-Matrix mit lauter Nullen in den letzten k Zeilen. Aber die Transformation mit der L-Matrix ändert die Norm des Residuumsvektors. Eine optimale Lösung des transformierten Systems ist dann keine Optimallösung des Originalsystems.

Überbestimmte Systeme $A \cdot \mathbf{x} = \mathbf{b}$: *SVD*-Zerlegung

Aufgabe

Suche $\mathbf{x}$ so, dass Residuums-2-Norm $\|\mathbf{r}\|_2$ minimal wird

$$\|\mathbf{r}\| = \|\mathbf{b} - A \cdot \mathbf{x}\| \to \text{min!}$$

Vorgangsweise

Zerlege $A = U \cdot S \cdot V^T$, substituiere $\mathbf{y} = V^T \cdot \mathbf{x}$.
Multiplikation von $\mathbf{r}$ mit U^T lässt 2-Norm unverändert. Wandle um:

$$\|\mathbf{r}\| = \|U^T \cdot \mathbf{r}\| = \|U^T\mathbf{b} - S \cdot V^T \cdot \mathbf{x}\| = \|Q^T\mathbf{b} - S \cdot \mathbf{y}\|$$

Für das gedrehte System ist die Minimallösung völlig offensichtlich...

Zahlenbeispiel, verschiedene Methoden

Wir rechnen ein Zahlenbeispiel mit 3 Gleichungen für 2 Unbekannte auf drei Arten durch:

- Normalengleichungen
- *QR*-Zerlegung
- Singulärwertzerlegung

Für das Zahlenbeispiel gibt es MATLAB-Zusatzmaterial:
`UebbGlsysBeisp.m`

Illustration: Farbmischung

Additive Farbmischung↔Linerakombination von Farbvektoren

Wie gut lässt sich der Farbton *SlateGray* (RGB 112 128 144) aus *Turquoise* (RGB 64 224 208) und *DeepPink* (RGB 255 20 147) zusammenmischen?

$$\begin{pmatrix} 64 \\ 224 \\ 208 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 255 \\ 20 \\ 147 \end{pmatrix} \cdot x_2 =? \begin{pmatrix} 112 \\ 128 \\ 144 \end{pmatrix}$$

Bestmögliche Mischfarbe

Die Linearkombination mit $x_1 = 0,520 \quad x_2 = 0,294$ ergibt die bestmögliche Mischfarbe (RGB 108 122 151) - keine andere kommt (in der 2-Norm) näher an RGB 112 128 144 heran.

Illustration: Farbmischung

Additive Farbmischung↔Linerakombination von Farbvektoren

Wie gut lässt sich der Farbton *SlateGray* (RGB 112 128 144) aus *Turquoise* (RGB 64 224 208) und *DeepPink* (RGB 255 20 147) zusammenmischen?

$$\begin{matrix} 64 \\ 224 \\ 208 \end{matrix} \cdot 0,520 + \begin{matrix} 255 \\ 20 \\ 147 \end{matrix} \cdot 0,294 = \begin{matrix} 108 \\ 122 \\ 151 \end{matrix}$$

Bestmögliche Mischfarbe

Die Linearkombination mit $x_1 = 0,520 \quad x_2 = 0,294$ ergibt die bestmögliche Mischfarbe (RGB 108 122 151) - keine andere kommt (in der 2-Norm) näher an RGB 112 128 144 heran.

Illustration: Farbmischung

Additive Farbmischung↔Linerakombination von Farbvektoren

Wie gut lässt sich der Farbton *SlateGray* (RGB 112 128 144) aus *Turquoise* (RGB 64 224 208) und *DeepPink* (RGB 255 20 147) zusammenmischen?

$$\begin{pmatrix} 64 \\ 224 \\ 208 \end{pmatrix} \cdot 0,520 + \begin{pmatrix} 255 \\ 20 \\ 147 \end{pmatrix} \cdot 0,294 = \begin{pmatrix} 108 \\ 122 \\ 151 \end{pmatrix} \approx \begin{pmatrix} 112 \\ 128 \\ 144 \end{pmatrix}$$

Bestmögliche Mischfarbe

Die Linearkombination mit $x_1 = 0,520 \quad x_2 = 0,294$ ergibt die bestmögliche Mischfarbe (RGB 108 122 151) - keine andere kommt (in der 2-Norm) näher an RGB 112 128 144 heran.

Rechenweg über Normalengleichungen

Die Standard-Lehrbuch-Lösung

Gleichungssystem $A \cdot \mathbf{x} = \mathbf{b}$
$$\begin{bmatrix} 64 & 255 \\ 224 & 20 \\ 208 & 147 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 112 \\ 128 \\ 144 \end{bmatrix}$$

Multipliziere Matrix und rechte Seite mit der transponierten Matrix

$$A^T \cdot A = \begin{bmatrix} 64 & 224 & 208 \\ 255 & 20 & 147 \end{bmatrix} \cdot \begin{bmatrix} 64 & 255 \\ 224 & 20 \\ 208 & 147 \end{bmatrix} = \begin{bmatrix} 97536 & 51376 \\ 51376 & 87034 \end{bmatrix}$$

$$A^T \cdot \mathbf{b} = \begin{bmatrix} 64 & 224 & 208 \\ 255 & 20 & 147 \end{bmatrix} \cdot \begin{bmatrix} 112 \\ 128 \\ 144 \end{bmatrix} = \begin{bmatrix} 65792 \\ 52288 \end{bmatrix}$$

A und A^T sind hier nur der Deutlichkeit halber farblich hinterlegt; die Farben haben sonst keine tiefere Bedeutung

Rechenweg über Normalengleichungen

Die Standard-Lehrbuch-Lösung

Gleichungssystem $A \cdot \mathbf{x} = \mathbf{b}$ $\qquad \begin{bmatrix} 64 & 255 \\ 224 & 20 \\ 208 & 147 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 112 \\ 128 \\ 144 \end{bmatrix}$

Multipliziere Matrix und rechte Seite mit der transponierten Matrix

$$A^T \cdot A = \begin{bmatrix} 64 & 224 & 208 \\ 255 & 20 & 147 \end{bmatrix} \cdot \begin{bmatrix} 64 & 255 \\ 224 & 20 \\ 208 & 147 \end{bmatrix} = \begin{bmatrix} 97536 & 51376 \\ 51376 & 87034 \end{bmatrix}$$

$$A^T \cdot \mathbf{b} = \begin{bmatrix} 64 & 224 & 208 \\ 255 & 20 & 147 \end{bmatrix} \cdot \begin{bmatrix} 112 \\ 128 \\ 144 \end{bmatrix} = \begin{bmatrix} 65792 \\ 52288 \end{bmatrix}$$

A und A^T sind hier nur der Deutlichkeit halber farblich hinterlegt; die Farben haben sonst keine tiefere Bedeutung

Rechenweg über Normalengleichungen

Die Standard-Lehrbuch-Lösung

Gleichungssystem $A \cdot \mathbf{x} = \mathbf{b}$ $$\begin{bmatrix} 64 & 255 \\ 224 & 20 \\ 208 & 147 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 112 \\ 128 \\ 144 \end{bmatrix}$$

Multipliziere Matrix und rechte Seite mit der transponierten Matrix

$$A^T \cdot A = \begin{bmatrix} 64 & 224 & 208 \\ 255 & 20 & 147 \end{bmatrix} \cdot \begin{bmatrix} 64 & 255 \\ 224 & 20 \\ 208 & 147 \end{bmatrix} = \begin{bmatrix} 97536 & 51376 \\ 51376 & 87034 \end{bmatrix}$$

$$A^T \cdot \mathbf{b} = \begin{bmatrix} 64 & 224 & 208 \\ 255 & 20 & 147 \end{bmatrix} \cdot \begin{bmatrix} 112 \\ 128 \\ 144 \end{bmatrix} = \begin{bmatrix} 65792 \\ 52288 \end{bmatrix}$$

A und A^T sind hier nur der Deutlichkeit halber farblich hinterlegt; die Farben haben sonst keine tiefere Bedeutung

Rechenweg über Normalengleichungen (Forts.)

System der Normalengleichungen

$$(A^T \cdot A) \cdot \mathbf{x} = A^T \cdot \mathbf{b} \qquad \begin{bmatrix} 97536 & 51376 \\ 51376 & 87034 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 65792 \\ 52288 \end{bmatrix}$$

- Matrix $A^T \cdot A$ ist symmetrisch
- Größenordnung der Zahlenwerte in $A^T \cdot A$ ist Quadrat der Zahlenwerte in der Originalmatrix
- Konditionszahl der Normalengleichungen ist Quadrat der Original-Konditionszahl. Das vergrößert Rundungsfehler !

  (Das ist bei kleinen Beispielen, so wie hier, kein Thema – erst bei „wirklichen“, großen Systemen)
- Lösung der Normalengleichungen $\mathbf{x} = [0.520; 0.294]$ gibt die „am wenigsten falsche Lösung“ des überbestimmten Systems.

Rechenweg über Normalengleichungen (Forts.)

System der Normalengleichungen

$$(A^T \cdot A) \cdot \mathbf{x} = A^T \cdot \mathbf{b} \qquad \begin{bmatrix} 97536 & 51376 \\ 51376 & 87034 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 65792 \\ 52288 \end{bmatrix}$$

- Matrix $A^T \cdot A$ ist symmetrisch
- Größenordnung der Zahlenwerte in $A^T \cdot A$ ist Quadrat der Zahlenwerte in der Originalmatrix
- Konditionszahl der Normalengleichungen ist Quadrat der Original-Konditionszahl. Das vergrößert Rundungsfehler !

  (Das ist bei kleinen Beispielen, so wie hier, kein Thema – erst bei „wirklichen“, großen Systemen)
- Lösung der Normalengleichungen $\mathbf{x} = [0.520; 0.294]$ gibt die „am wenigsten falsche Lösung“ des überbestimmten Systems.

Rechenweg mit QR-Zerlegung, anschaulich

Original-System in Spaltenvektor-Schreibung

$$\begin{pmatrix} 64 \\ 224 \\ 208 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 255 \\ 20 \\ 147 \end{pmatrix} \cdot x_2 = \begin{pmatrix} 112 \\ 128 \\ 144 \end{pmatrix} \qquad A\mathbf{x} = \mathbf{b}$$

System in gedrehten Koordinaten

Die Matrix Q^T aus der QR-Zerlegung dreht alle Vektoren in ein einfacheres Koordinatensystem. Die Matrix R enthält die gedrehten Spalten von A

$$\begin{pmatrix} 312 \\ 0 \\ 0 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 165 \\ -245 \\ 0 \end{pmatrix} \cdot x_2 = \begin{pmatrix} 211 \\ -72 \\ 10 \end{pmatrix} \qquad R\mathbf{x} = Q^T\mathbf{b}$$

Die Geometrie (Längen, Winkel) der Spaltenvektoren ist dieselbe, nur der Blickwinkel ist anders!

Rechenweg mit QR-Zerlegung, anschaulich

Original-System in Spaltenvektor-Schreibung

$$\begin{pmatrix} 64 \\ 224 \\ 208 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 255 \\ 20 \\ 147 \end{pmatrix} \cdot x_2 = \begin{pmatrix} 112 \\ 128 \\ 144 \end{pmatrix} \qquad A\mathbf{x} = \mathbf{b}$$

System in gedrehten Koordinaten

Die Matrix Q^T aus der QR-Zerlegung dreht alle Vektoren in ein einfacheres Koordinatensystem. Die Matrix R enthält die gedrehten Spalten von A

$$\begin{pmatrix} 312 \\ 0 \\ 0 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 165 \\ -245 \\ 0 \end{pmatrix} \cdot x_2 = \begin{pmatrix} 211 \\ -72 \\ 10 \end{pmatrix} \qquad R\mathbf{x} = Q^T\mathbf{b}$$

Die Geometrie (Längen, Winkel) der Spaltenvektoren ist dieselbe, nur der Blickwinkel ist anders!

Rechenweg mit QR-Zerlegung, anschaulich

Original-System in Spaltenvektor-Schreibung

$$\begin{pmatrix} 64 \\ 224 \\ 208 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 255 \\ 20 \\ 147 \end{pmatrix} \cdot x_2 = \begin{pmatrix} 112 \\ 128 \\ 144 \end{pmatrix} \qquad A\mathbf{x} = \mathbf{b}$$

System in gedrehten Koordinaten

Die Matrix Q^T aus der QR-Zerlegung dreht alle Vektoren in ein einfacheres Koordinatensystem. Die Matrix R enthält die gedrehten Spalten von A

$$\begin{pmatrix} 312 \\ 0 \\ 0 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 165 \\ -245 \\ 0 \end{pmatrix} \cdot x_2 = \begin{pmatrix} 211 \\ -72 \\ 10 \end{pmatrix} \qquad R\mathbf{x} = Q^T\mathbf{b}$$

Die Geometrie (Längen, Winkel) der Spaltenvektoren ist dieselbe, nur der Blickwinkel ist anders!

Singulärwert-Zerlegung, anschaulich

System in gedrehten Koordinaten

Die Matrix U^T aus der Singulärwert-Zerlegung $A = U \cdot S \cdot V^T$ dreht die Spalten in ein neues Koordinatensystem

$$\begin{pmatrix} 282 \\ 135 \\ 0 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 254 \\ -150 \\ 0 \end{pmatrix} \cdot x_2 = \begin{pmatrix} 221 \\ 26 \\ 10 \end{pmatrix} \qquad S \cdot V^T \mathbf{x} = U^T \mathbf{b}$$

Lösungsvektor auch noch gedreht

Die Matrix V^T aus der Singulärwert-Zerlegung $A = U \cdot S \cdot V^T$ dreht den Lösungsvektor: $\mathbf{y} = V^T \cdot \mathbf{x}$. Die Gleichungen für $\mathbf{y}$ werden ganz einfach

$$\begin{pmatrix} 379 \\ 0 \\ 0 \end{pmatrix} \cdot y_1 + \begin{pmatrix} 0 \\ 202 \\ 0 \end{pmatrix} \cdot y_2 = \begin{pmatrix} 221 \\ 26 \\ 10 \end{pmatrix} \qquad S\mathbf{y} = U^T \mathbf{b}$$

Singulärwert-Zerlegung, anschaulich

System in gedrehten Koordinaten

Die Matrix U^T aus der Singulärwert-Zerlegung $A = U \cdot S \cdot V^T$ dreht die Spalten in ein neues Koordinatensystem

$$\begin{pmatrix} 282 \\ 135 \\ 0 \end{pmatrix} \cdot x_1 + \begin{pmatrix} 254 \\ -150 \\ 0 \end{pmatrix} \cdot x_2 = \begin{pmatrix} 221 \\ 26 \\ 10 \end{pmatrix} \qquad S \cdot V^T \mathbf{x} = U^T \mathbf{b}$$

Lösungsvektor auch noch gedreht

Die Matrix V^T aus der Singulärwert-Zerlegung $A = U \cdot S \cdot V^T$ dreht den Lösungsvektor: $\mathbf{y} = V^T \cdot \mathbf{x}$. Die Gleichungen für $\mathbf{y}$ werden ganz einfach

$$\begin{pmatrix} 379 \\ 0 \\ 0 \end{pmatrix} \cdot y_1 + \begin{pmatrix} 0 \\ 202 \\ 0 \end{pmatrix} \cdot y_2 = \begin{pmatrix} 221 \\ 26 \\ 10 \end{pmatrix} \qquad S\mathbf{y} = U^T \mathbf{b}$$

Überbestimmte Systeme, Zusammenfassung

Normalengleichungen: Löse das System

$$(A^T \cdot A) \cdot \mathbf{x} = A^T \cdot b$$

QR-Zerlegung: Löse das System

$$R \cdot \mathbf{x} = Q^T \cdot b$$

Singulärwert-Zerlegung: Löse die Systeme

$$S \cdot \mathbf{y} = U^T \cdot b$$
$$\mathbf{x} = V \cdot \mathbf{y}$$

Pseudoinverse (kommt noch)

$$\mathbf{x} = A^+ \cdot \mathbf{b}$$

Überbestimmte Systeme, Zusammenfassung

Normalengleichungen: Löse das System

$$(A^T \cdot A) \cdot \mathbf{x} = A^T \cdot b$$

***QR*-Zerlegung: Löse das System**

$$R \cdot \mathbf{x} = Q^T \cdot b$$

Singulärwert-Zerlegung: Löse die Systeme

$$S \cdot \mathbf{y} = U^T \cdot b$$
$$\mathbf{x} = V \cdot \mathbf{y}$$

Pseudoinverse (kommt noch)

$$\mathbf{x} = A^+ \cdot \mathbf{b}$$

Gliederung 5. Vorlesung

Ein einfaches inverses Problem

Ein Zaubertrick?

Gegeben ist die Matrix

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

- Denken Sie sich einen Vektor
- multiplizieren Sie ihn mit A
- sagen Sie mir das Ergebnis; ich kann den Vektor finden, an den Sie gedacht haben!

Hier erfahren Sie, wie dieser Trick funktioniert, und was passiert, wenn Sie den Trick mit

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{oder} \quad A = \begin{bmatrix} 1 & 2{,}1 \\ 2 & 4 \end{bmatrix}$$

wiederholen möchten.

Dazu Zusatzmaterial `InversesProblem.m`

Abbildung und inverse Abbildung

Eine Matrix definiert durch $\mathbf{y} = A \cdot \mathbf{x}$ eine lineare Abbildung.

Die Matrix ist dazu gedacht,
dass sie aus einem Vektor einen anderen macht.

Die inverse Matrix macht diese Abbildung rückgängig: $\mathbf{x} = A^{-1} \cdot \mathbf{y}$

Was eine Matrix tut,
macht die Inverse wieder gut.

Aber das ist nicht immer möglich: Eine lineare Abbildung auf den Nullvektor lässt sich nicht umkehren.

Doch wenn ein Vektor ganz verschwindet,
gibt 's keine Matrix, die ihn wiederfindet.

Die pseudoinverse Matrix macht rückgängig, so gut es eben geht.

Abbildung und inverse Abbildung

Eine Matrix definiert durch $\mathbf{y} = A \cdot \mathbf{x}$ eine lineare Abbildung.

Die Matrix ist dazu gedacht,
dass sie aus einem Vektor einen anderen macht.

Die inverse Matrix macht diese Abbildung rückgängig: $\mathbf{x} = A^{-1} \cdot \mathbf{y}$

Was eine Matrix tut,
macht die Inverse wieder gut.

Aber das ist nicht immer möglich: Eine lineare Abbildung auf den Nullvektor lässt sich nicht umkehren.

Doch wenn ein Vektor ganz verschwindet,
gibt 's keine Matrix, die ihn wiederfindet.

Die pseudoinverse Matrix macht rückgängig, so gut es eben geht.

Abbildung und inverse Abbildung

Eine Matrix definiert durch $\mathbf{y} = A \cdot \mathbf{x}$ eine lineare Abbildung.

Die Matrix ist dazu gedacht,
dass sie aus einem Vektor einen anderen macht.

Die inverse Matrix macht diese Abbildung rückgängig: $\mathbf{x} = A^{-1} \cdot \mathbf{y}$

Was eine Matrix tut,
macht die Inverse wieder gut.

Aber das ist nicht immer möglich: Eine lineare Abbildung auf den Nullvektor lässt sich nicht umkehren.

Doch wenn ein Vektor ganz verschwindet,
gibt 's keine Matrix, die ihn wiederfindet.

Die pseudoinverse Matrix macht rückgängig, so gut es eben geht.

Abbildung und inverse Abbildung

Eine Matrix definiert durch $\mathbf{y} = A \cdot \mathbf{x}$ eine lineare Abbildung.

Die Matrix ist dazu gedacht,
dass sie aus einem Vektor einen anderen macht.

Die inverse Matrix macht diese Abbildung rückgängig: $\mathbf{x} = A^{-1} \cdot \mathbf{y}$

Was eine Matrix tut,
macht die Inverse wieder gut.

Aber das ist nicht immer möglich: Eine lineare Abbildung auf den Nullvektor lässt sich nicht umkehren.

Doch wenn ein Vektor ganz verschwindet,
gibt 's keine Matrix, die ihn wiederfindet.

Die pseudoinverse Matrix macht rückgängig, so gut es eben geht.

Pseudoinverse

tritt auch bei überbestimmten Systemen auf

Für überbestimmte Systeme $A\mathbf{x} = \mathbf{b}$ lässt sich die kleinste-Quadrate-Lösung aus den Normalengleichungen bestimmen

$$
\begin{aligned}
A^T A\mathbf{x} &= A^T\mathbf{b} && \mid \quad \text{multipliziere } (A^T A)^{-1}\cdot \\
\mathbf{x} &= (A^T A)^{-1}A^T\,\mathbf{b} && \mid \quad \text{substituiere } (A^T A)^{-1}A^T \to A^+ \\
\mathbf{x} &= A^+\mathbf{b}
\end{aligned}
$$

(abgesehen von numerischen Problemen und dem Sonderfall, dass A nicht vollen Spaltenrang hat)

Die Matrix A^+ wirkt also ähnlich wie eine Inverse bei der Lösung eines „gewöhnlichen“ Gleichungssystems mit nichtsingulärer quadratischer Matrix.

Pseudoinverse

tritt auch bei überbestimmten Systemen auf

Für überbestimmte Systeme $A\mathbf{x} = \mathbf{b}$ lässt sich die kleinste-Quadrate-Lösung aus den Normalengleichungen bestimmen

$$
\begin{aligned}
A^T A\mathbf{x} &= A^T\mathbf{b} && \Big| \quad \text{multipliziere } (A^TA)^{-1}\cdot \\
\mathbf{x} &= (A^TA)^{-1}A^T\,\mathbf{b} && \Big| \quad \text{substituiere } (A^TA)^{-1}A^T \to A^+ \\
\mathbf{x} &= A^+\mathbf{b}
\end{aligned}
$$

(abgesehen von numerischen Problemen und dem Sonderfall, dass A nicht vollen Spaltenrang hat)

Die Matrix A^+ wirkt also ähnlich wie eine Inverse bei der Lösung eines „gewöhnlichen“ Gleichungssystems mit nichtsingulärer quadratischer Matrix.

Pseudoinverse

tritt auch bei überbestimmten Systemen auf

Für überbestimmte Systeme $A\mathbf{x} = \mathbf{b}$ lässt sich die kleinste-Quadrate-Lösung aus den Normalengleichungen bestimmen

$$
\begin{aligned}
A^T A\mathbf{x} &= A^T\mathbf{b} && \Big| \quad \text{multipliziere } (A^T A)^{-1}\cdot \\
\mathbf{x} &= (A^T A)^{-1}A^T\,\mathbf{b} && \Big| \quad \text{substituiere } (A^T A)^{-1}A^T \rightarrow A^+ \\
\mathbf{x} &= A^+\mathbf{b}
\end{aligned}
$$

(abgesehen von numerischen Problemen und dem Sonderfall, dass A nicht vollen Spaltenrang hat)

Die Matrix A^+ wirkt also ähnlich wie eine Inverse bei der Lösung eines „gewöhnlichen“ Gleichungssystems mit nichtsingulärer quadratischer Matrix.

Pseudoinverse

Die Definition $A^+ = (A^T A)^{-1} A^T$ ist nicht immer gültig

Problem

Die Definition $A^+ = (A^T A)^{-1} A^T$ ist nicht möglich, wenn $(A^T A)$ singulär ist.

Trotzdem lässt sich eine Matrix A^+ angeben, die eine optimale Lösung des überbestimmten Systems findet.

Existenz und Eigenschaften der Pseudoinversen

Zu jeder reellen $m \times n$-Matrix A gibt es eine eindeutig bestimmte reelle $n \times m$-Matrix A^+, die Moore-Penrose Inverse, mit den Eigenschaften

$$A \cdot A^+ \cdot A = A \qquad (A \cdot A^+)^T = A \cdot A^+$$
$$A^+ \cdot A \cdot A^+ = A^+ \qquad (A^+ \cdot A)^T = A^+ \cdot A$$

Falls $A^+ = (A^T A)^{-1} A^T$ existiert, erfüllt diese Matrix alle vier Bedingungen.

Pseudoinverse

Die Definition $A^+ = (A^T A)^{-1} A^T$ ist nicht immer gültig

Problem

Die Definition $A^+ = (A^T A)^{-1} A^T$ ist nicht möglich, wenn $(A^T A)$ singulär ist.

Trotzdem lässt sich eine Matrix A^+ angeben, die eine optimale Lösung des überbestimmten Systems findet.

Existenz und Eigenschaften der Pseudoinversen

Zu jeder reellen $m \times n$-Matrix A gibt es eine eindeutig bestimmte reelle $n \times m$-Matrix A^+, die Moore-Penrose Inverse, mit den Eigenschaften

$$A \cdot A^+ \cdot A = A \qquad (A \cdot A^+)^T = A \cdot A^+$$

$$A^+ \cdot A \cdot A^+ = A^+ \qquad (A^+ \cdot A)^T = A^+ \cdot A$$

Falls $A^+ = (A^T A)^{-1} A^T$ existiert, erfüllt diese Matrix alle vier Bedingungen.

Inverse und Pseudoinverse von Diagonalmatrizen

Für quadratische Diagonalmatrizen ist die Definition recht einfach...

Inverse einer Diagonalmatrix (falls alle $s_i \neq 0$)

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & s_n \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} \frac{1}{s_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{s_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{s_n} \end{bmatrix}$$

Pseudoinverse einer Diagonalmatrix

$$S^+ = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & r_n \end{bmatrix} \qquad \text{mit } r_i = \begin{cases} \frac{1}{s_i} \\ 0 \end{cases} \text{falls } \begin{cases} s_i \neq 0 \\ s_i = 0 \end{cases}$$

Inverse und Pseudoinverse von Diagonalmatrizen

Für quadratische Diagonalmatrizen ist die Definition recht einfach...

Inverse einer Diagonalmatrix (falls alle $s_i \neq 0$)

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & s_n \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} \frac{1}{s_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{s_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{s_n} \end{bmatrix}$$

Pseudoinverse einer Diagonalmatrix

$$S^+ = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & r_n \end{bmatrix} \qquad \text{mit } r_i = \begin{cases} \frac{1}{s_i} \\ 0 \end{cases} \text{falls } \begin{cases} s_i \neq 0 \\ s_i = 0 \end{cases}$$

Pseudoinverse von rechteckigen Diagonalmatrizen

Ist $S \in \mathbb{R}^m \times \mathbb{R}^n$, dann ist $S^+ \in \mathbb{R}^n \times \mathbb{R}^m$

Definition der r_i und s_i bleibt gleich wie vorhin, es gibt nur zusätzliche Nullzeilen oder -spalten.

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & s_n \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \qquad S^+ = \begin{bmatrix} r_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & r_2 & & 0 & 0 & \vdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & r_n & 0 & \cdots & 0 \end{bmatrix}$$

Pseudoinverse allgemein

Verwende Singulärwertzerlegung $A = U \cdot S \cdot V^T$

Bei der Multiplikation $\mathbf{y} = A \cdot \mathbf{x} = U \cdot S \cdot V^T \cdot \mathbf{x}$ spürt der Vektor $\mathbf{x}$

- zuerst V^T
- dann S
- zuletzt U

Um diese drei Multiplikationen rückgängig zu machn, muss man bei der letzten beginnen:

- U rückgängig machen: mit U^T multiplizieren
- S rückgängig machen: hier braucht man S^+
- V^T rückgängig machen: mit V multiplizieren

Pseudoinverse

$$A = U \cdot S \cdot V^T \qquad A^+ = V \cdot S^+ \cdot U^T$$

Pseudoinverse allgemein

Verwende Singulärwertzerlegung $A = U \cdot S \cdot V^T$

Bei der Multiplikation $\mathbf{y} = A \cdot \mathbf{x} = U \cdot S \cdot V^T \cdot \mathbf{x}$ spürt der Vektor $\mathbf{x}$

- zuerst V^T
- dann S
- zuletzt U

Um diese drei Multiplikationen rückgängig zu machn, muss man bei der letzten beginnen:

- U rückgängig machen: mit U^T multiplizieren
- S rückgängig machen: hier braucht man S^+
- V^T rückgängig machen: mit V multiplizieren

Pseudoinverse

$$A = U \cdot S \cdot V^T \qquad A^+ = V \cdot S^+ \cdot U^T$$

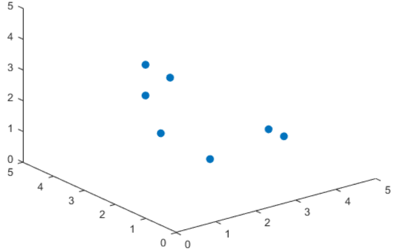
Anwendung der Singulärwert-Zerlegung: Datenkompression

Film-Bewertungen	mein Goldfisch	meine Tante	mein Freund
Amelie	2	1.5	1
Pretty Woman	1	1.5	4
Beauty and the Beast	5	3	0
Lion King	5	3	0
Finding Nemo	5	3.5	0
Star Wars: Episode VI - Return of the Jedi	0	1	5
Matrix	0	1	5
Terminator	0	1	5
Gladiator	0	1	4
Bourne Identity	0	1	5
Indiana Jones and the Raiders of the Lost Ark	0	0.5	3

Stellen Sie sich diese Tabelle riesig groß vor: 10K Filme, 100K Fans. . .

Matrix von Filmbewertungen

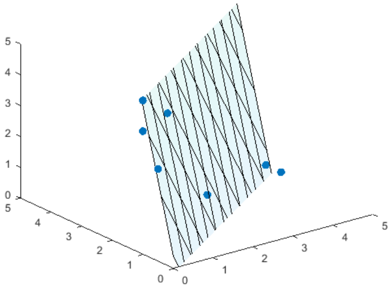
$$A = \begin{bmatrix} 2 & 1.5 & 1 \\ 1 & 1.5 & 4 \\ 5 & 3 & 0 \\ 5 & 3 & 0 \\ 5 & 3.5 & 0 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 4 \\ 0 & 1 & 5 \\ 0 & 0.5 & 3 \end{bmatrix}$$



Die Zeilen lassen sich als Punkte im $\mathbb{R}^3$ zeichnen.

Matrix von Filmbewertungen

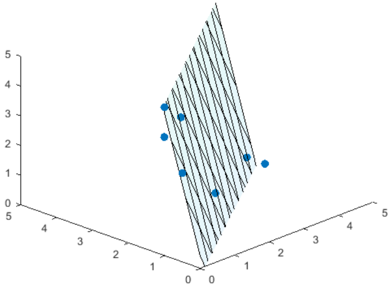
$$A = \begin{bmatrix} 2 & 1.5 & 1 \\ 1 & 1.5 & 4 \\ 5 & 3 & 0 \\ 5 & 3 & 0 \\ 5 & 3.5 & 0 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 4 \\ 0 & 1 & 5 \\ 0 & 0.5 & 3 \end{bmatrix}$$



Aber sie liegen fast in einem 2-dimensionalen Unterraum

Matrix von Filmbewertungen

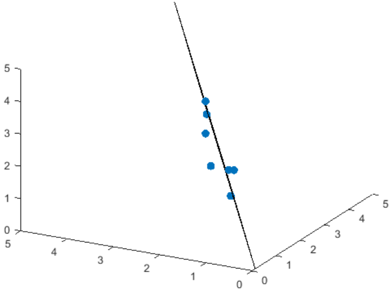
$$A = \begin{bmatrix} 2 & 1.5 & 1 \\ 1 & 1.5 & 4 \\ 5 & 3 & 0 \\ 5 & 3 & 0 \\ 5 & 3.5 & 0 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 4 \\ 0 & 1 & 5 \\ 0 & 0.5 & 3 \end{bmatrix}$$



Aber sie liegen fast in einem 2-dimensionalen Unterraum

Matrix von Filmbewertungen

$$A = \begin{bmatrix} 2 & 1.5 & 1 \\ 1 & 1.5 & 4 \\ 5 & 3 & 0 \\ 5 & 3 & 0 \\ 5 & 3.5 & 0 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 4 \\ 0 & 1 & 5 \\ 0 & 0.5 & 3 \end{bmatrix}$$



Aber sie liegen fast in einem 2-dimensionalen Unterraum

Matrix von Filmbewertungen

$$A = \begin{bmatrix} 2 & 1.5 & 1 \\ 1 & 1.5 & 4 \\ 5 & 3 & 0 \\ 5 & 3 & 0 \\ 5 & 3.5 & 0 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \\ 0 & 1 & 4 \\ 0 & 1 & 5 \\ 0 & 0.5 & 3 \end{bmatrix}$$

$$A = U \cdot S \cdot V^T$$

- Singulärwerte $\sigma_1, \sigma_2, \sigma_3$ sind [10.32.10.39].
- $\sigma_3 \ll \sigma_1, \sigma_2$
- $\sigma_3 = 0$ setzen.
- komprimierte Matrix $A_k = U_k \cdot S_k \cdot V_k^T$
- U_k und V_k enthalten nur die Spalten zu Nichtnull-Werten in S_k

A lässt sich so durch eine Matrix mit kleinerem Rang approximieren. Praktisch wichtig erst bei vielen 1000 Zeilen und Spalten: Behalte nur wenige Spalten in U und V, entsprechend den größten Singulärwerten.

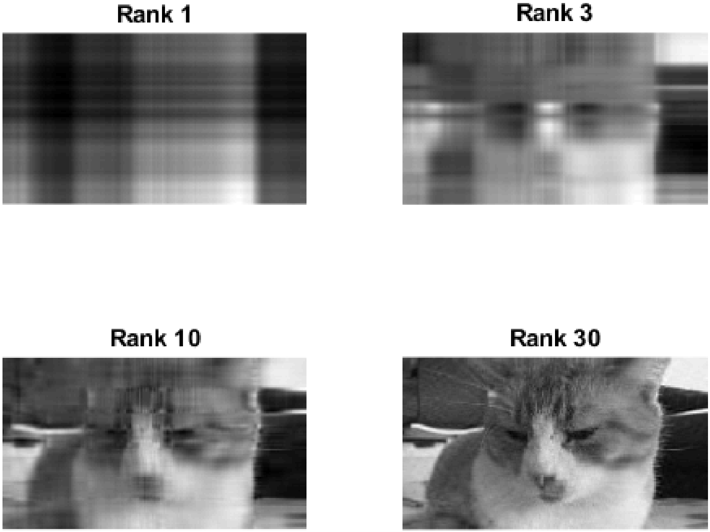
Daten-Komprimierung durch SVD

Kurzfassung, grob vereinfacht

- Daten in Matrix-Form liegen oft in niedrig-dimensionalen Unterräumen. Im vorigen Beispiel: 11 Datenpunkte im $\mathbb{R}^3$ liegen eigentlich fast im $\mathbb{R}^2$.
- Singulärwert-Zerlegung erkennt den Daten-Unterraum, der wesentlich zur Information beiträgt.
- Aus den wenigen Spaltenvektoren in U und V, die zu den größten Singulärwerten gehören, lässt sich die wichtigste Information rekonstruieren.
- Den Rest kann man weglassen.
- Gleich kommt ein Beispiel, wo sich Datenpunkte einer 2322×4128-Matrix schon durch drei Singulärvektoren erkennbar rekonstruieren lassen.

Beispiel: Bildkomprimierung durch SVD

Bildpunkte: 2322×4128-Matrix; Approximation durch 1 bis 30 Singulärvektoren



Beispiel: Gesichtserkennung mit *eigenfaces*

Hier sind einige zufällig ausgewählte Gesichter aus einem Datensatz (Yale Face Database)

The Yale Face Database contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.

Beispiel: Gesichtserkennung mit *eigenfaces*

Das ist das gemittelte Gesicht und die ersten 15 „Eigengesichter“. Sie enthalten bereits etwa 80% der gesamten Information der Daten. Jedes einzelne Gesicht lässt sich aus dem „Mittelgesicht“ und den „Eigengesichtern“ $F_1, F_2, F_3, \dots$ rekonstruieren.

Die Datei `eigenfaces.m` berechnet die eigenfaces

Beispiel: Gesichtserkennung mit *eigenfaces*

Hier ist ein Gesicht aus dem Original-Datenzatz.
Die Zutaten-Liste lautet: „Mittelgesicht" + . . .

$$-1694.2F_1 - 2810.8F_2 + 382.4F_3 + 599.7F_4 \ldots$$
$$-765.83F_5 + 358.55F_6 - 286.64F_7 \ldots$$

Darunter die Gesichter rekonstruiert aus den ersten 15 und den ersten 30 Eigengesichtern.

Es ist viel billiger, statt aller Bild-Pixel nur 15 oder 30 Zahlenwerte pro Bild zu speichern.

Auch für ein unbekanntes Gesicht ist es einfacher, die 15 oder 30 Eigen-Zahlen zu berechnen und diese mit den Werten in der Datenbank zu vergleichen.