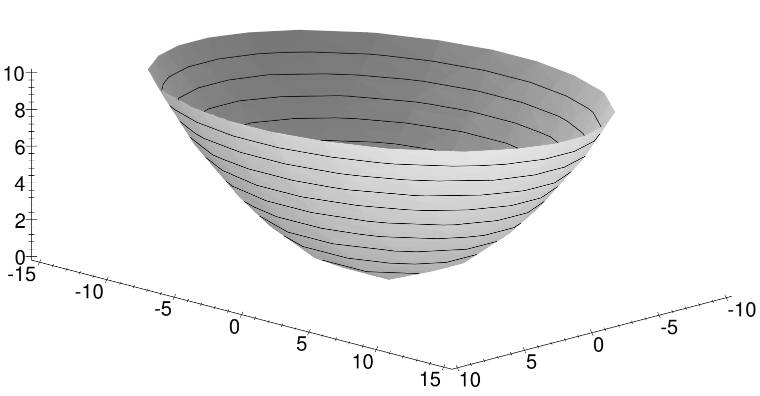

Abb. 16: Das elliptischen Paraboloid $x^2 + y^2/2 - 10z = 0$

Der Fall $s = 1$ liefert die beiden Normalformen

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} + \beta z = 0 \quad \text{und} \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1\,.$$

Letztere definiert einen Zylinder mit hyperbolischer Grundlinie, erstere hat für konstantes z Hyperbeln als Schnittkurven und für konstantes x oder y Parabeln; man spricht daher von einem *hyperbolischen Paraboloid* oder, aus Gründen, die man in Abbildung 17 erkennen kann, einer *Sattelfläche.*

Für $s = 0$ schließlich ist entsprechend nur

$$-\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

interessant, und diese Gleichung ist unlösbar.

Damit ist der Fall $r = 2$ erledigt; wir kommen zu $r = 1$. Für $s = 1$ haben wir die Normalformen

$$\frac{x^2}{a^2} + by + cz = 0 \quad \text{und} \quad \frac{x^2}{a^2} = 1\,.$$

Letztere definiert die beiden parallelen Ebenen $x = \pm a$, erstere können wir durch den Koordinatenwechsel $y \mapsto by + cz$ vereinfachen zu

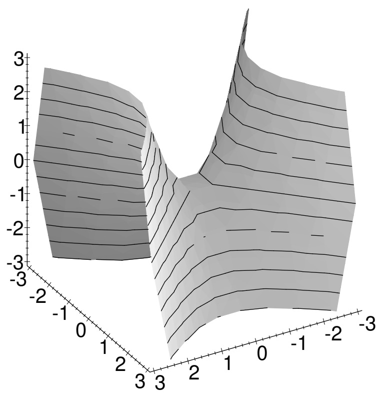

Abb. 17: Die Sattelfläche $x^2 - y^2 - z = 0$

$x^2 + a^2y = 0$, d.h. wir haben einen Zylinder mit einer Parabel als Grundlinie.

Für $s = 0$ sind die Normalformen entsprechend

$$-\frac{x^2}{a} + by + cz = 0 \quad \text{und} \quad -\frac{x^2}{a} = 1\,;$$

nun ist die zweite Gleichung unlösbar, und die erste definiert weiterhin einen parabolischen Zylinder.

Damit bleibt nur noch der Fall $r = 0$, in dem wir keine echte Quadrik haben, sondern eine Ebene, die gegebenfalls auch zur leeren Menge oder zum gesamten $\mathbb{R}^3$ degenerieren kann.

Als Beispiel für die Bestimmung einer Normalform betrachten wir die Quadrik

$$x^2 + 4xy + 4yz + 12xz + z^2 + 10x - 10z = 1\,.$$

Als erstes müssen wir eine symmetrische Matrix A finden, so daß

$$(x,\, y,\, z)A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^2 + 4xy + 4yz + 12xz + z^2$$

ist. Da y^2 nicht vorkommt und sowohl x^2 als auch z^2 den Koeffizienten eins haben, sind die Diagonaleinträge $1, 0, 1$. Weiter schreiben wir, da wir eine symmetrische Matrix wollen, $4xy = 2xy + 2yx$, so daß $a_{12} = a_{21} = 2$ ist, und entsprechend findet man auch $a_{23} = a_{32} = 2$ und $a_{13} = a_{31} = 6$. Also ist

$$A = \begin{pmatrix} 1 & 2 & 6 \\ 2 & 0 & 2 \\ 6 & 2 & 1 \end{pmatrix} \quad \text{und mit} \quad \vec{b} = \begin{pmatrix} 10 \\ 0 \\ -10 \end{pmatrix}$$

wird die Quadrik in Matrix-Vektor-Schreibweise durch die Gleichung

$${}^t\vec{x}A\vec{x} + \vec{b} \cdot \vec{x} = 1$$

beschrieben.

Als nächstes muß die Matrix A diagonalisiert werden, d.h. wir brauchen ihre Eigenwerte. Das charakteristische Polynom kann (beispielsweise nach der SARRUSschen Regel) leicht berechnet werden; es ergibt sich zu

$$\det(A - \lambda E) = -\lambda^3 + 2\lambda^2 + 43\lambda + 40\,.$$

Die Nullstellen eines solchen kubischen Polynoms lassen sich im allgemeinen nur mit Mühe bestimmen; einfacher wird es, wenn man eine davon kennt. Gerade bei einer relativ symmetrisch aufgebauten Matrix wie hier kann man oft durch hinreichend scharfes Hinschauen einen Eigenwert sehen: In

$$A - \lambda E = \begin{pmatrix} 1-\lambda & 2 & 6 \\ 2 & 0 & 2 \\ 6 & 2 & 1-\lambda \end{pmatrix}$$

sind der erste und der letzte Spaltenvektor spiegelverkehrt zueinander, und für $\lambda = -5$ werden sie sogar gleich, insbesondere also linear abhängig. Somit ist -5 ein Eigenwert von A; ein Eigenvektor dazu ist etwa

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Da

$$\begin{aligned} &(\lambda_1 - \lambda)(\lambda_2 - \lambda)(\lambda_3 - \lambda) \\ = &-\lambda^3 + (\lambda_1 + \lambda_2 + \lambda_3)\lambda^2 - (\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1)\lambda + \lambda_1\lambda_2\lambda_3 \end{aligned}$$

ist, wissen wir, daß das Produkt aller drei Eigenwerte 40 und ihre Summe gleich zwei ist, das Produkt der beiden noch unbekannten Eigenwerte ist also -8 und ihre Summe ist 7. Damit kennen wir die drei Eigenwerte

$$\lambda_1 = 8, \quad \lambda_2 = -1 \quad \text{und} \quad \lambda_3 = -5\,.$$

und wissen, daß die Quadrik den Rang drei hat.

Auch die Signatur $(1,\ 2)$ scheint klar zu sein, ist es aber nicht: Durch die Umformungen, die wir auf dem Weg zur Normalform noch vornehmen werden, kann eventuell die Konstante auf der rechten Seite negativ werden, und wenn wir die Gleichung dann durch Division wieder auf rechte Seite eins bringen, ändern sich alle Vorzeichen, d.h. die Signatur wird möglicherweise zu $(2,\ 1)$.

Um die Quadrik besser zu verstehen, müssen wir zunächst die Eigenvektoren bestimmen und die Quadrikgleichung auf das neue Koordinatensystem umrechnen.

Den Eigenvektor zu -5 kennen wir bereits. Für $\lambda = 8$ ist

$$A - 8E = \begin{pmatrix} -7 & 2 & 6 \\ 2 & -8 & 2 \\ 6 & 2 & -7 \end{pmatrix},$$

und da die Summe der ersten und der dritten Spalte gleich minus der Hälte der zweiten ist, können wir als Eigenvektor

$$\begin{pmatrix} 1 \\ 1/2 \\ 1 \end{pmatrix} \quad \text{oder besser} \quad \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

nehmen. Auch bei $\lambda = -1$ und

$$A + E = \begin{pmatrix} 2 & 2 & 6 \\ 2 & 1 & 2 \\ 6 & 2 & 2 \end{pmatrix}$$

können wir einen Eigenvektor erraten, wenn wir den ersten und den dritten Spaltenvektor addieren: Wir erhalten das Vierfache der zweiten Spalte und kennen damit den Eigenvektor

$$\begin{pmatrix} 1 \\ -4 \\ 1 \end{pmatrix}.$$

Da die drei Eigenvektoren zu verschiedenen Eigenwerten gehören, ist klar, daß sie eine Orthogonalbasis des EUKLIDischen Vektorraums $\mathbb{R}^3$ bilden, was zur Bestimmung des *Typs* der Quadrik genügen würde; wenn wir aber auch Längen und Längenverhältnisse bezüglich der neuen Basis mit den üblichen Formeln korrekt berechnen wollen, brauchen wir eine Orthonormalbasis. Wir müssen daher die Eigenvektoren noch auf Länge eins normieren und arbeiten deshalb mit der Basis aus

$$\vec{b}_1 = \frac{1}{3}\begin{pmatrix}2\\1\\2\end{pmatrix}, \quad \vec{b}_2 = \frac{\sqrt{2}}{6}\begin{pmatrix}1\\-4\\1\end{pmatrix} \quad \text{und} \quad \vec{b}_3 = \frac{\sqrt{2}}{2}\begin{pmatrix}1\\0\\-1\end{pmatrix}.$$

Mit

$$\vec{x} = u\vec{b}_1 + v\vec{b}_2 + w\vec{b}_3$$

ist dann

$$\begin{aligned} {}^t\vec{x}A\vec{x} &= {}^t\vec{x}(uA\vec{b}_1 + vA\vec{b}_2 + wA\vec{b}_3) = {}^t\vec{x}(8u\vec{b}_1 - v\vec{b}_2 - 5w\vec{b}_3) \\ &= {}^t(u\vec{b}_1 + v\vec{b}_2 + w\vec{b}_3)(8u\vec{b}_1 - v\vec{b}_2 - 5w\vec{b}_3) = 8u^2 - v^2 - 5w^2, \end{aligned}$$

denn ${}^t\vec{b}_i\vec{b}_j = \delta_{ij}$.

Der lineare Teil der Quadrikgleichung ist

$$10x - 10z = 10\begin{pmatrix}1\\0\\-1\end{pmatrix}\cdot\vec{x} = 10\cdot\sqrt{2}\vec{b}_3\cdot\vec{x} = 10\sqrt{2}w,$$

bezüglich der neuen Vektorraumbasis hat die Quadrik nun also die Gleichung

$$8u^2 - v^2 - 5w^2 + 10\sqrt{2}w = 1.$$

Für die Normalform müssen wir als nächstes den linearen Term eliminieren; wir ersetzen w durch $t = w - \sqrt{2}$, verschieben also den Ursprung des Koordinatensystems auf der w-Achse um $\sqrt{2}$; dann ist

$$-5t^2 = -5w^2 + 10\sqrt{2}w - 10 \quad \text{oder} \quad -5w^2 + 10\sqrt{2}w = -5t^2 + 10.$$

Im neuen Koordinatensystem hat die Quadrik also die Gleichung

$$8u^2 - v^2 - 5t^2 + 10 = 1 \quad \text{oder} \quad 8u^2 - v^2 - 5t^2 = -9.$$

Das ist schon fast eine Normalform, nur steht auf der rechten Seite anstelle einer Eins noch -9. Also müssen wir die gesamte Gleichung

durch -9 dividieren; dabei ändern sich alle Vorzeichen und die damit auch die Signatur. Die Normalform ist also

$$\frac{v^2}{9} + \frac{5t^2}{9} - \frac{8u^2}{9} = 1.$$

Somit hat die Quadrik Signatur $(2,\ 1)$ und ist ein einschaliges Hyperboloid.

3) Positiv und negativ definite Matrizen: Quadriken im $\mathbb{R}^4$ und in höherdimensionalen Räumen lassen sich zumindest im Prinzip mit denselben Methoden klassifizieren, die wir gerade für die Ebene und den $\mathbb{R}^3$ angewandt haben. Da die Geometrie von mehr als dreidimensionalen Räumen jedoch nicht sonderlich anschaulich ist, wollen wir darauf verzichten.

Stattdessen wollen wir uns zum Ende dieses Abschnitts mit der Frage beschäftigen, wann die quadratische Form ${}^t\vec{x}A\vec{x}$ nur positive oder nur negative Werte annimmt; dies wird uns im nächsten Semester bei Extremwertproblemen in mehreren Veränderlichen nützlich sein.

Definition: *a)* Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ heißt *positiv semidefinit,* wenn für alle $\vec{v} \in \mathbb{R}^n$ gilt: ${}^t\vec{v}A\vec{v} \geq 0$.
b) A heißt *negativ semidefinit,* wenn für alle $\vec{v} \in \mathbb{R}^n$ gilt: ${}^t\vec{v}A\vec{v} \leq 0$.
c) A heißt positiv *bzw.* negativ *definit,* wenn A positiv *bzw.* negativ semidefinit ist und ${}^t\vec{v}A\vec{v} = 0$ genau dann, wenn $\vec{v} = \vec{0}$ ist.
d) A heißt *indefinit,* wenn ${}^t\vec{v}A\vec{v}$ sowohl positive als auch negative Werte annehmen kann.

In diesem Sinne ist beispielsweise die Einheitsmatrix positiv definit, denn $\sum_{i=1}^n v_i^2$ ist nie negativ und verschwindet genau dann, wenn alle v_i verschwinden. Entsprechend ist die negative Einheitsmatrix negativ definit, und die Matrix

$$A = \begin{pmatrix}0&0&0\\0&1&0\\0&0&4\end{pmatrix}$$

ist positiv semidefinit, denn zwar ist

$${}^t\vec{v}A\vec{v} = (v_1,\ v_2,\ v_3)\begin{pmatrix}0&0&0\\0&1&0\\0&0&4\end{pmatrix}\begin{pmatrix}v_1\\v_2\\v_3\end{pmatrix} = v_2^2 + 4v_3^2 \geq 0$$

für alle Vektoren $\vec{v} \in \mathbb{R}^3$, aber auch für den ersten Einheitsvektor $\vec{e}_1$ ist ${}^t\vec{e}_1 A\vec{e}_1 = 0$, da das Produkt nicht von der ersten Komponente abhängt. Allgemein gilt

Lemma: Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ ist genau dann positiv (negativ) definit, wenn alle ihre Eigenwerte positiv (negativ) sind. Sie ist positiv (negativ) semidefinit, wenn alle ihre Eigenwerte nichtnegativ (nichtpositiv) sind, und sie ist indefinit, wenn es sowohl positive als auch negative Eigenwerte gibt.

Beweis: Wir gehen wieder aus von einer Orthonormalbasis $\{\vec{v}_1, \dots, \vec{v}_n\}$ aus Eigenvektoren von A; für einen Vektor $\vec{v} = c_1\vec{v}_1 + \ldots + c_n\vec{v}_n$ ist dann

$${}^t\vec{v}A\vec{v} = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2\,,$$

wobei λ_i der Eigenwert zum Eigenvektor $\vec{v}_i$ ist. Da alle $c_i^2 \geq 0$ sind und gleich null nur dann, wenn c_i verschwindet, folgt die Behauptung. ■

Im Fall $n = 2$ läßt sich dies in ein einfacheres Kriterium übersetzen: Die Eigenwerte λ_1 und λ_2 einer symmetrischen 2×2-Matrix

$$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$$

sind die Wurzeln des charakterischen Polynoms

$$\begin{vmatrix} a-\lambda & b \\ b & d-\lambda \end{vmatrix} = (a-\lambda)(d-\lambda) - b^2 = \lambda^2 - (a+d)\lambda + (ad-b^2)$$
$$= (\lambda-\lambda_1)(\lambda-\lambda_2) = \lambda^2 - (\lambda_1+\lambda_2)\lambda + \lambda_1\lambda_2\,.$$

Also ist das Produkt der beiden Eigenwerte gleich der Determinanten (was wir auch so schon wußten), und ihre Summe ist $a + d$.

Für eine definite Matrix müssen λ_1 und λ_2 entweder beide positiv sein oder beide negativ; dies ist genau dann der Fall, wenn ihr Produkt, die Determinante also, positiv ist. Alsdann haben beide dasselbe Vorzeichen wie $a+d$; die Matrix ist also genau dann positiv definit, wenn $a+d$ positiv ist, und negativ definit, wenn es negativ ist.

Bei einer nur semidefiniten Matrix kann die Determinante als Produkt der beiden Eigenwerte auch verschwinden, sie muß also nur größer

oder gleich null sein, und dasselbe gilt im positiv semidefiniten Fall für ihre Summe, denn im Extremfall könnte die Matrix auch gleich der Nullmatrix sein. Entsprechend muß $a + d$ im negativ semidefiniten Fall nur kleiner oder gleich null sein.

Für definite Matrizen können wir noch etwas weiter vereinfachen: Wenn die Determinante $ad - b^2$ positiv ist, ist erst recht auch $ad > 0$, d.h. a und d haben dasselbe Vorzeichen, das dann natürlich gleich dem Vorzeichen ihrer Summe $a + d$ sein muß. Damit ergibt sich folgendes Kriterium:

Lemma: *a)* Die 2×2-Matrix $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ ist indefinit, falls $\det A$ negativ ist.
b) Ist $\det A > 0$, so ist A positiv definit für positives a und negativ definit für negatives a.
c) Ist $\det A \geq 0$, so ist A positiv semidefinit für $a + d \geq 0$ und negativ semidefinit für $a + d \leq 0$.
■

Für höhere Dimensionen gibt es ähnliche Kriterien, die allerdings deutlich weniger attraktiv sind, da man außer der Determinante der Matrix selbst auch noch die von mehreren Untermatrizen berechnen muß.

§8: Komprimierung von Bild- und Audiodaten

Nach der recht langen und trockenen Beschäftigung mit Eigenwerten und Eigenvektoren im letzten Paragraphen wollen wir wenigstens kurz eine praktische Anwendung von Eigenvektoren und Matrizen kennenlernen, die Komprimierung von Bild- und Audiodaten.

Hinter den Verfahren, die hierfür typischerweise angewandt werden, stehen auch mathematische Theorien, die wir bislang noch nicht kennen und die hier nur kurz informell eingeführt werden können; der wesentliche Aspekt, der am Ende zur Komprimierung führt, sind aber die mittlerweise zu Genüge bekannten Eigenvektoren symmetrischer Matrizen, und daher sollten, wenn auch vielleicht nicht alle Einzelheiten ganz klar werden, zumindest die Grundgedanken dieses Paragraphen verständlich sein.

a) Datenkompression

Ziel der Datenkompression ist es, eine Datei für Zwecke der Speicherung oder Übertragung möglichst stark zu verkleinern, das aber in einer solchen Weise, daß sich die ursprüngliche Datei aus der verkleinerten wieder exakt rekonstruieren läßt.

Es ist klar, daß es keinen universellen Algorithmus zur Datenkompression geben kann: Gäbe es nämlich ein Verfahren, das für beliebige Dateien einen Kompressionsfaktor $\alpha < 1$ garantieren würde, so könnte man dieses Verfahren iterativ anwenden und nach n Anwendungen eine Kompressionsrate von α^n erreichen. Wenn man n nur hinreichend groß wählt, könnte man daher jede Datei auf weniger als ein Bit komprimieren, was natürlich absurd ist.

Ein Kompressionsverfahren kann also nur auf Dateien mit spezieller Struktur erfolgreich angewandt werden und muß die spezielle Redundanz in diesen Dateien ausnutzen. In Textdateien beispielsweise ist dies die Redundanz der Sprache, die schon bei bloßer Beachtung der höchst unterschiedlichen Buchstabenhäufigkeiten Kompressionen von rund 50% gestattet.

Bild- und Audiodaten werden typischerweise als Matrizen ganzer Zahlen zwischen 0 und 255 *bzw.* Vektoren ganzer Zahlen zwischen 0 und 65535 ($= 2^{16} - 1$) digitalisiert. (Der Unterschied zwischen den Wertenbereichen liegt darin begründet, daß unser Auge selbst bei gedruckten Bildern mit nur 64 Graustufen praktisch keine Artefakte mehr erkennen kann, wohingegen unser Gehör noch auf sehr feine Unterschiede reagiert.)

Bei einer Musik-CD etwa wird das Signal 44 100-mal pro Sekunde abgetastet (dies bedeutet nach dem Abtasttheorem von NYQUIST, daß ein auf den Bereich von 0 bis 22,05kHz bandbegrenztes Signal fehlerfrei rekonstruiert werden kann), und das Ergebnis wird dann so skaliert und quantisiert (d.h. gerundet), daß eine Zahl zwischen 0 und 65535 entsteht. Bei Bilddaten werden beispielsweise 256×256 oder 512×512 Bildpunkte abgetastet (je nach Seitenverhältnis des Bilds und den Beschränkungen des Scanners kommen auch fast beliebige andere Bildgrößen vor), bei Schwarzweißbilder nur nach Helligkeit, bei Farbbildern nach insgesamt

drei physikalischen Größen, die vom jeweiligen Farbmodell abhängen. Das Ergebnis dieser Abtastungen wird dann entsprechend skaliert und quantisiert.

Typische Komprimierungsverfahren arbeiten daher mit Vektoren oder Matrizen aus Zahlen zwischen 0 und einer geeigneten Zahl M, die aus praktischen Gründen meist von der Form $2^{8r} - 1$ ist, wobei die Zahl r der Empfindlichkeit unserer Sinne angepaßt gleich eins oder zwei ist.

Da man zur eindeutigen Festlegung von N beliebigen Zahlen zwischen 0 und 2^{8r} nicht mit weniger als den $8Nr$ Bit auskommen kann, die man zum Hinschreiben der Zahlen braucht, sehen wir auch hier wieder, daß kein Verfahren *alle* solchen Vektoren komprimieren kann; wir müssen also eine Teilmenge auszeichnen.

Die ideale solche Teilmenge wäre hier natürlich die Menge aller möglicher Bilder (oder Audiosequenzen), aber diese Menge dürfte mathematisch wohl kaum definierbar sein: Schließlich hängt es sehr vom Betrachter ab, welches Pixelmuster er noch als „Bild“ gelten läßt und welches nicht. Sinnvoll läßt sich eine solche Menge daher höchstens definieren, wenn von vornherein feststeht, welche Bilder berücksichtigt werden sollen – und dann ist wohl ein Verfahren, das statt vom Bildinhalt von einer Bildnummer ausgeht, unschlagbar.

Alle Verfahren, die beliebige, aber realistische Bilder komprimieren möchten, gehen aus von einem *statistischen Modell,* das zwar auch viele Matrizen produziert, die niemand als „Bilder“ anerkennen würde, das aber dennoch genügend viele Eigenschaften realer Bilder reproduziert, um eine große Anzahl von „Nichtbildern“ auszuschließen.

Ausgangspunkt ist die Beobachtung, daß es in einem Bild oder Musikstück nur wenige aprupte Übergänge gibt. Zwar gibt es natürlich immer wieder ein plötzliches *fortissimo,* das auf eine leise Stelle folgt, aber da das Signal 44 100-mal pro Sekunde abgetastet wird und solche übergänge selbst bei der schrägsten Musik deutlich seltener als im Sekundenrythmus erfolgen, sind diese Sprünge innerhalb des zu behandelnden Datenstroms in der Tat sehr seltene Ereignisse. Wir können daher davon ausgehen, daß sich die unmittelbaren Nachbarn eines Datums *im Mittel* nur wenig vom gegebenen Datum unterscheiden.

Dasselbe gilt auch für Bilddaten: Falls das Bild digital hinreichend fein dargestellt wird, so daß keine Rastereffekte erkennbar sind, kommen große Sprünge in den Helligkeitswerten nur selten vor.

Unser nächstes Ziel ist es daher, diesen engen Zusammenhang zwischen benachbarten Daten mathematisch zu fassen. Das ist am anschaulichsten, wenn wir keine Bilder oder Musikstücke betrachten, sondern einen einfachen und überschaubaren Datensatz – auch wenn dieser nicht das geringste mit Bild- oder Audiodaten zu tun hat:

b) Kovarianzen und Korrelationen bei der Europawahl 1999

Wir betrachten dazu das Ergebnis der Europawahl vom 13. Juni 1999 in den verschiedenen Mannheimer Stadtteilen. Laut Mannheimer Morgen vom 15. Juni gab es folgende Prozentzahlen:

Bezirk	Bet.	CDU	SPD	Grüne	Rep.	FDP	PDS	Sonst.
(A)	37,7	40,7	31,8	13,0	2,4	5,0	3,4	4,2
(B)	30,9	33,1	39,8	12,9	3,9	2,1	4,1	4,1
(D)	36,1	39,5	36,1	10,8	2,8	3,1	3,7	4,0
(D)	44,9	46,0	26,6	12,8	1,9	6,6	2,5	3,6
(E)	46,0	44,2	30,0	12,0	2,4	5,6	2,1	3,7
(F)	48,6	51,1	23,7	10,8	2,6	6,3	2,5	3,0
(G)	36,8	43,3	42,0	4,6	3,1	2,2	1,3	3,5
(H)	28,3	39,2	46,0	4,0	4,1	1,6	1,5	3,6
(I)	37,9	40,1	42,7	5,7	3,7	2,8	1,7	3,3
(J)	37,9	44,9	37,3	6,4	3,3	2,6	1,8	3,7
(K)	40,9	44,8	40,9	3,8	3,2	2,5	1,6	3,2
(L)	49,1	46,3	35,9	7,7	2,2	3,7	1,2	3,0
(M)	47,4	47,0	30,2	11,9	1,6	4,2	1,8	3,3
(N)	42,5	48,8	32,7	9,1	1,9	3,5	1,4	2,6
(O)	40,9	44,8	38,1	7,4	1,8	3,2	1,8	2,9
(P)	42,5	47,8	30,3	10,2	2,2	3,8	2,1	3,5
(Q)	38,4	47,6	34,9	6,0	3,4	2,9	1,2	4,0
MA	**39,8**	**44,2**	**34,6**	**9,0**	**2,7**	**3,7**	**2,1**	**3,7**

Die Stadtteile sind, damit alles in eine Zeile paßt, durch Buchstaben bezeichnet; für Interessenten seien auch die richtigen Namen angegeben:

(A) steht für Innenstadt/Jungbusch, (B) für Neckarstadt West, (C) für Neckarstadt Ost/Wohlgelegen, (D) ist die Oststadt/Schwetzingerstadt, (E) der Lindenhof, (F) Neuostheim/Neuhermsheim, (G) Sandhofen, (H) Schönau, (I) Waldhof, (J) Käfertal, (K) Vogelstang, (L) Wallstadt, (M) Feudenheim, (N) Seckenheim, (O) Friedrichsfeld, (P) Neckarau, (Q) Rheinau und MA schließlich Mannheim insgesamt.

Wie man sieht, haben die einzelnen Parteien in den verschiedenen Stadtteilen recht unterschiedlich abgeschnitten; bei der FDP etwa ergibt sich mehr als ein Faktor vier zwischen den 1,6% in der Schönau und den 6,6% in Neuostheim/Neuhermsheim. Es ist auch klar, daß die CDU dort unterdurchschnittlich viele Stimmen hat, wo die SPD überdurchschnittlich abgeschnitten hat; anders geht das bei den Stimmenzahlen dieser beiden Parteien gar nicht. Wie steht es aber mit den kleinen Parteien?

Eine Partei hat in einem Stadtteil dann überdurchschnittlich gut abgeschnitten, wenn sie dort einen höheren Prozentsatz erreicht hat als für Mannheim insgesamt, wenn also die Differenz zwischen dem Prozentsatz dort und dem für ganz Mannheim positiv ist. Bei negativer Differenz hat sie unterdurchschnittlich abgeschnitten.

Nehmen wir nun an, zwei Parteien A und B hätten ihre Hochburgen und ihre Schwächegebiete jeweils in denselben Stadtteilen; dann ist also die Differenz für die eine Partei genau dann positiv *bzw.* negativ, wenn sie das auch für die andere Partei ist. Das Produkt der beiden Differenzen ist daher immer positiv und auch die Summe dieser Produkte ist positiv.

Falls dagegen die Hochburgen von A die Schwächegebiete von B sind und umgekehrt, sind alle Produkte und damit auch deren Summe negativ, denn für jedem Stadtteil wird eine positive Zahl mit einer negativen Zahl multipliziert.

Wenn es schließlich keinen signifikaten Zusammenhang zwischen den Hochburgen und Schwächegebieten der beiden Parteien gibt, wird es unter den Hochburgen von A welche geben, in denen auch B überdurchschnittlich abschneidet, es wird aber auch welche geben, wo B unterdurchschnittlich viele Stimmen erhielt, so daß wir sowohl positive als auch negative Produkte addieren und die Summe daher betragsmäßig relativ klein sein sollte.

Das mittlere Produkte der Abweichungen vom Mittelwert bezeichnen wir als *Kovarianz* zwischen den beiden Parteien A und B:

Definition: Die Kovarianz zwischen zwei Datenvektoren $\vec{x}, \vec{y} \in \mathbb{R}^N$ ist

$$\operatorname{cov}(\vec{x}, \vec{y}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y}),$$

wobei

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \text{und} \quad \overline{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

die Mittelwerte der beiden Datensätze sind.

In diesem Beispiel ist natürlich $N = 17$ die Anzahl der Stadtteile.

(Tatsächlich ist die Kovarianz aus obiger Definition nicht genau die, mit der wir hier arbeiten: Da wir als Mittelwerte das Abschneiden in ganz Mannheim nehmen und die verschiedenen Stadtteile nicht alle die gleiche Größe haben, haben wir hier einen etwas aussagekräftigeren gewichteten Mittelwert anstelle von $\overline{x}$ und $\overline{y}$.)

Wenn wir, wie in diesem Beispiel, nicht nur zwei, sondern m Datensätze $\vec{x}_1, \dots \vec{x}_m \in \mathbb{R}^N$ haben, können wir die Kovarianzen $\operatorname{cov}(\vec{x}_i, \vec{x}_j)$ zu einer Matrix zusammenfassen, der *Kovarianzmatrix.* Sie ist offensichtlich eine symmetrische Matrix.

Wenn wir die Wahlbeteiligung formal wie eine Partei namens W behandeln, erhalten wir hier die folgende Kovarianzmatrix:

	W	CDU	SPD	Grüne	Rep.	FDP	PDS	Sonst.
W	32,4	18,8	−24,3	5,66	−3,27	5,79	−1,32	−1,46
CDU	18,8	18,2	−15,7	−0,58	−1,96	3,00	−2,07	−1,08
SPD	−24,3	−15,7	34,9	−13,8	3,11	−7,66	−1,07	0,10
Grüne	5,66	−0,58	−13,8	10,2	−1,21	3,13	1,93	0,44
Rep.	−3,27	−1,96	3,11	−1,21	0,57	−0,69	0,05	0,14
FDP	5,79	3,00	−7,66	3,13	−0,69	2,04	0,26	−0,04
PDS	−1,32	−2,07	−1,07	1,93	0,05	0,26	0,73	0,22
Sonst.	−1,46	−1,08	0,10	0,44	0,14	−0,04	0,23	0,24

Diese Matrix ist noch etwas unübersichtlich: Wann immer eine große Partei mit im Spiel ist, erhalten wir große Werte, ansonsten eher kleine. Auch die Diagonaleinträge

$$\operatorname{cov}(\vec{x}, \vec{x}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2 .$$

der Matrix sind noch interpretationsbedürftig: Sie geben offenbar an, wie stark die x_i um ihren Mittelwert $\overline{x}$ schwanken.

Definition: $\sigma_{\vec{x}}^2 \underset{\text{def}}{=} \operatorname{cov}(\vec{x}, \vec{x}) = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} (x_i - \overline{x})^2$ heißt *Varianz* des Datensatzes $\vec{x}$; die Quadratwurzel $\sigma_{\vec{x}} = \sqrt{\sigma_{\vec{x}}^2}$ heißt *Standardabweichung.*

Nach der CAUCHY-SCHWARZschen Ungleichung ist für je zwei Vektoren $\vec{v}, \vec{w} \in \mathbb{R}^n$ stets $|\vec{v} \cdot \vec{w}| \le |\vec{v}| \cdot |\vec{w}|$, also ist auch stets

$$|\operatorname{cov}(\vec{x}, \vec{y})| \le \sigma_{\vec{x}} \cdot \sigma_{\vec{y}}, \quad \text{d.h.} \quad -1 \le \frac{\operatorname{cov}(\vec{x}, \vec{y})}{\sigma_{\vec{x}} \cdot \sigma_{\vec{y}}} \le 1 .$$

Definition: $\rho(\vec{x}, \vec{y}) = \dfrac{\operatorname{cov}(\vec{x}, \vec{y})}{\sigma_{\vec{x}} \cdot \sigma_{\vec{y}}}$ heißt *Korrelation* zwischen den Datensätzen $\vec{x}$ und $\vec{y}$.

Analog zur Kovarianzmatrix haben wir dann auch die *Korrelationsmatrix,* deren sämtliche Einträge zwischen -1 und 1 liegen; im vorliegenden Fall ist sie

	W	CDU	SPD	Grüne	Rep.	FDP	PDS	Sonst.
W	1,00	0,77	−0,72	0,31	−0,76	0,71	−0,27	−0,52
CDU	0,77	1,00	−0,62	−0,04	−0,61	0,49	−0,57	−0,52
SPD	−0,72	−0,62	1,00	−0,73	0,70	−0,91	−0,21	0,03
Grüne	0,31	−0,04	−0,73	1,00	−0,50	0,69	0,71	0,28
Rep.	−0,76	−0,61	0,70	−0,50	1,00	−0,64	0,07	0,37
FDP	0,71	0,49	−0,91	0,69	−0,64	1,00	0,21	−0,05
PDS	−0,27	−0,56	−0,21	0,71	0,07	0,21	1,00	0,53
Sonst.	−0,52	−0,52	0,03	0,28	0,37	−0,05	0,53	1,00

Dieser Matrix können wir nun aller gewünschten Informationen entnehmen. Die erste Zeile beispielsweise sagt uns, daß die CDU, FDP und in geringerem Ausmaß auch die Grünen dort überdurchschnittlich gut abgeschnitten haben, wo auch die Wahlbeteilung überdurchschnittlich war, diese drei Parteien konnten also anscheinend ihre Anhänger am besten mobilisieren.

Wo die CDU überdurchschnittlich gut abschnitt, waren alle anderen Parteien mit Ausnahme der FDP unterdurchschnittlich; wo die SPD überdurchschnitt gut abschnitt, waren ebenfalls fast alle anderen unterdurchschnittlich; die Ausnahme hier sind die Repuplikaner. Zwischen SPD und FDP haben wir mit -0,91 den betragsgrößten Korrelationskoeffizienten; die Hochburgen dieser beiden Parteien sind also sehr unterschiedlich. Die Grünen schließlich sind positiv korreliert mit FDP, PDS und Sonstigen.

c) Zufallsvariablen und ihre statistischen Kenngrößen

Die Europawahl vom 13. Juni 1999 ist vorbei und ausgezählt, wir konnten daher im letzten Abschnitt mit wohlbekannten Daten rechnen. Beim Umgang mit Bild- und Tondaten gilt das nur in eingeschränktem Maße: Zwar können Daten erst dann verarbeitet werden, wenn sie da sind, aber eine Digitalkamera beispielsweise muß doch schon heute auf die Bilder vorbereitet sein, die ihr Besitzer erst in ein paar Jahren aufnehmen wird.

Für den Grundalgorithmus zur Datenkompression müssen wir daher Daten zulassen, über die wir noch nichts konkretes wissen – abgesehen von gewissen vagen Gesetzmäßigkeiten, durch die sich echte Bilddaten von beliebigen Matrizen unterscheiden. Als Hilfmittel dazu dient der Begriff der *Zufallsvariablen.*

Definition: Eine (diskrete) Zufallsvariablen ist ein Prozeß, der zufällig einen Wert aus einer vorgegebenen endlichen Menge

$$\{x_0, \dots, x_m\}$$

liefert.

Dieser „Zufall“ muß natürlich, falls er mathematisch faßbar sein soll, irgendwelchen Regeln genügen, und hier kommt der zweite fundamentale Begriff zum Einsatz: Wir nehmen an, daß für jeden der möglichen Werte x_i feststeht, mit welcher *Wahrscheinlichkeit* p_i er angenommen wird. Diese „Wahrscheinlichkeit“ definieren wir informell so, daß bei einer großen Anzahl m von Versuchen *ungefähr* $p_i m$-mal der Wert x_i geliefert wird. Das „Gesetz der großen Zahlen“, das wir im nächsten Semester kennenlernen werden, sagt, daß diese Definition sinnvoll ist und man die Wahrscheinlichkeiten p_i damit in wohldefinierter Weise mit beliebiger Genauigkeit bestimmen kann.

Zwei formale Konsequenzen der Definition sind offensichtlich:

$$0 \le p_i \le 1 \quad \text{für alle } i \text{ und} \quad \sum_{i=0}^{m} p_i = 1 \,.$$

Sollen beispielsweise alle x_i mit gleicher Wahrscheinlichkeit angenommen werden (was für Bilddaten nicht unbedingt ein sehr realistisches Modell ist), so sind alle $p_i = 1/(m+1)$.

In unserem Modell soll ein „Bild“ dann produziert werden durch eine Matrix (X_{ij}) von geeigneten Zufallsvariablen; eine „Audiosequenz“ entsprechend durch einen Vektor $(X_1, \dots, X_N)$.

Welche Zufallsvariablen sind geeignet? In unserem einfachen Modell wollen wir uns nicht darauf festlegen, wie die p_i gewählt werden sollen, sondern stattdessen mit ihrer Hilfe Analoga zu den Kenngrößen aus dem vorigen Abschnitt definieren.

Definition: Der *Erwartungswert* $E(X)$ einer Zufallsvariablen X ist

$$E(X) = \sum_{i=0}^{m} p_i x_i \,.$$

Falls alle x_i mit gleicher Wahrscheinlichkeit angenommen werden, ist der Erwartungswert also einfach das arithmetische Mittel

$$\frac{1}{m+1} \sum_{i=0}^{m} x_i$$

der möglichen Werte; ansonsten ist er ein sogenanntes gewichtetes arithmetisches Mittel. Für einen Würfel etwa, der die Augenzahlen von $x_0 = 1$ bis $x_5 = 6$ mit jeweils gleicher Wahrscheinlichkeit produziert, ist

$$E(X) = \frac{1+2+3+4+5+6}{6} = 3\frac{1}{2}\,.$$

Als nächste wichtige Kenngröße hatten wir imletzten Abschnitt die mittlere quadratiscche Abweichung vom Mittelwert, die Varianz, betrachtet; in völliger Analogie zu dort definieren wir

Definition: Die Varianz einer Zufallsvariablen X mit Erwartungswert $E(X)$ ist

$$\sigma_X^2 = E\big((X - E(X))^2\big) = \sum_{i=0}^{m} p_i\big(x_i - E(X)\big)^2\,;$$

ihre Standardabweichung ist $\sigma_X = \sqrt{\sigma_X^2}$.

Beim Würfel wäre also

$$\sigma_X^2 = \frac{\left(-2\frac12\right)^2 + \left(-1\frac12\right)^2 + \left(-\frac12\right)^2 + \left(\frac12\right)^2 + \left(1\frac12\right)^2 + \left(2\frac12\right)^2}{6} = \frac{35}{12}$$

und

$$\sigma_X = \sqrt{\frac{35}{12}} \approx 1{,}7078\,.$$

Um die Bedeutung dieser Kenngrößen in der Bildverarbeitung zu veranschaulichen, sind auf der nächsten Doppelseit sechs beliebte Testbilder zusammen mit den Werten dieser Kenngrößen abgedruckt. Die Werte sind entnommen aus

P.M. FARELLE: Recursive Block Coding for Image Data Compression, *Springer,* 1990 ;

sie beziehen sich natürlich auf die Originalbilder und nicht auf das, was der Druckvorgang hier im Skriptum daraus gemacht hat. Trotzdem sollte der Vergleich von Bildern und Daten einen einigermaßen korrekten

Eindruck zumindest der relativen Situation vermitteln, da hoffentlich alle hier abgedruckten Bilder in derselben Weise verunstaltet sind.

Die mittlere Helligkeit eines Bildes, dessen (viele) Pixel durch je eine Zufallsvariable mit Erwartungswert μ produziert werden, sollte ziemlich nahe bei μ liegen; der beste Schätzwert für den gemeinsamen Erwartungswert der Zufallsvariablen ist also die mittlere Helligkeit des Bildes. Typischerweise werden Helligkeiten durch Zahlen zwischen 0 und 255 kodiert, wobei schwarz der Zahl Null entspricht und weiß der 255. Dies sieht man gut an den Beispielbildern, wo das mit Abstand hellste Bild „Tiffany“ auch den mit Abstand größten Mittelwert μ hat; den kleinsten Wert hat das auch visuell dunkelste Bild „Lenna“.

Die nächste wichtige Kenngröße ist die Varianz, welche angibt, wie stark eine Zufallsvariable um ihren Erwartungswert streut. Auch hier wollen wir wieder davon ausgehen, daß alle Zufallsvariablen zu einem gegebenen Bild *bzw.* einer gegebenen Audiosequenz aus N dieselbe Varianz haben. Wir schätzen diese gemeinsame Varianz aufgrund der vorliegenden Daten als

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2\,,$$

wobei $y_1, \ldots, y_N$ die Helligkeits- *bzw.* Lautstärkewerte sind. (Es ist kein Druckfehler, daß vor dieser Summe mit N Summanden nur $N-1$ im Nenner steht; es liegt im wesentlichen daran, daß wir den Erwartungswert der Zufallsvariablen aufgrund des Bildes nur schätzen können. Im nächsten Semester, bei der Fehler- und Ausgleichsrechnung, werden wir solche Fragen genauer betrachten und dabei auch diese Formel herleiten.)

Was die Varianz und die Standardabweichung bedeuten, sieht man wieder deutlich an den Beispielbildern: Bilder mit geringem Kontrast wie „Tiffany“ oder „Lenna“ haben deutlich geringere Werte als die kontrastreicheren Bilder „Peppers“ und „Sailboat“.

Zur Beschreibung der gegenseitigen Abhängigkeit zweier Datensätze hatten wir im letzten Abschnitt Kovarianz und Korrelation eingeführt;