lassen sich also alle Skalarprodukte sowie auch die Basisdarstellung eines beliebigen Vektors auf die einfachst denkbare Weise darstellen.

Da die Vektoren einer Orthonormalbasis oftmals Wurzelausdrücke als Komponenten enthalten, sind vor allem beim Rechnen ohne Hilfsmittel Orthogonalbasen oft übersichtlicher als Orthonormalbasen; wir wollen den Satz also der Vollständigkeit halber auch für Othogonalbasen formulieren:

Satz: $(\vec{e}_1, \dots, \vec{e}_n)$ sei eine Orthogonalbasis eines EUKLIDischen oder HERMITEschen Vektorraums V.
a) Sind $\vec{v} = a_1\vec{e}_1 + \cdots + a_n\vec{e}_n$ und $\vec{w} = b_1\vec{e}_1 + \cdots + b_n\vec{e}_n$ zwei Vektoren aus V, so ist

$$\vec{v} \cdot \vec{w} = \sum_{i=1}^{n} a_i \overline{b_i} \, |\vec{e}_i|^2 \; .$$

b) Für einen Vektor $\vec{v} \in V$ ist

$$\vec{v} = a_1\vec{e}_1 + \cdots + a_n\vec{e}_n \quad \text{mit} \quad a_i = \frac{(\vec{v} \cdot \vec{e}_i)}{|\vec{e}_i|^2} \; .$$

Die *Beweise* sind fast dieselben Einzeiler:

$$\vec{v} \cdot \vec{w} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{b_j} \; \vec{e}_i \cdot \vec{e}_j = \sum_{i=1}^{n} a_i \overline{b_i} \, |\vec{e}_i|^2 \; , \quad \text{denn} \quad \vec{e}_i \cdot \vec{e}_j = \delta_{ij} \, |\vec{e}_i|^2$$

$$\vec{v} \cdot \vec{e}_i = (a_1\vec{e}_1 + \cdots + a_n\vec{e}_n) \cdot \vec{e}_i = \sum_{j=1}^{n} a_j \vec{e}_j \cdot \vec{e}_i = a_i \, |\vec{e}_i|^2 \; .$$

■

f) Die QR-Zerlegung einer Matrix

Wir betrachten den Vektorraum $V = \mathbb{R}^n$ oder $V = \mathbb{C}^n$ mit seiner Standardbasis und mit seiner üblichen Struktur als EUKLIDischer *bzw.* HERMITEscher Vektorraum, außerdem noch irgendeine weitere Basis

$(\vec{b}_1, \dots, \vec{b}_n)$. Nach dem GRAM-SCHMIDTschen Orthogonalisierungsverfahren können wir daraus eine Orthogonalbasis $(\vec{c}_1, \dots, \vec{c}_n)$ von V konstruieren, wobei gilt

$$\vec{c}_k = \vec{b}_k + \lambda_{k,k-1}\vec{b}_{k-1} + \cdots + \lambda_{k1}\vec{c}_1 \quad \text{mit} \quad \lambda_{kj} = -\frac{\vec{b}_k \cdot \vec{c}_j}{\vec{c}_j \cdot \vec{c}_j} \; .$$

Lösen wir auf nach $\vec{b}_k$, erhalten wir

$$\vec{b}_k = \vec{c}_k - \lambda_{k,k-1}\vec{b}_{k-1} - \cdots - \lambda_{k1}\vec{c}_1 = \sum_{j=1}^{n} r_{jk}\vec{c}_j$$

$$\text{mit } r_{jk} = \begin{cases} -\lambda_{kj} = \dfrac{\vec{b}_k \cdot \vec{c}_j}{\vec{c}_j \cdot \vec{c}_j} & \text{für } j < k \\ 1 & \text{für } j = k \\ 0 & \text{für } j > k \end{cases} \; .$$

Fassen wir die Vektoren $\vec{b}_1, \dots \vec{b}_n$ auf als Spaltenvektoren einer Matrix B und die $\vec{c}_j$ als Spalten von C, so gilt also für die i-ten Komponenten

$$b_{ik} = \sum_{j=1}^{n} r_{jk} c_{ij} = \sum_{j=1}^{n} c_{ij} r_{jk} \quad \text{oder} \quad B = CR \quad \text{mit} \quad R = (r_{ij}) \, .$$

Da r_{ij} für $i > j$ verschwindet und alle $r_{ii} = 1$ sind, ist R eine obere Dreiecksmatrix mit Einsen in der Hauptdiagonalen.

Ist B eine beliebige invertierbare $n \times n$-Matrix, so bilden die Spaltenvektoren $\vec{b}_i$ von B eine Basis des $\mathbb{R}^n$ *bzw.* $\mathbb{C}^n$; wie wir gerade gesehen haben, gibt es also eine obere Dreiecksmatrix R und eine Matrix C, deren Spalten eine Orthogonalbasis bilden, so daß $B = CR$ ist.

Normieren wir die Spalten von C auf Länge eins, erhalten wir eine $n \times n$-Matrix Q, deren Spalten eine Orthonormalbasis bilden; um die Beziehung $B = QR$ zu erhalten, müssen wir dann allerdings auch noch für jedes $i = 1, \dots, n$ die i-te Zeile von R mit der Länge des i-ten Spaltenvektors $\vec{c}_i$ von C multiplizieren. Dabei bleibt R zwar eine obere Dreiecksmatrix, in der Hauptdiagonalen stehen nun aber im allgemeinen keine Einsen mehr.

Was passiert, wenn B nicht invertierbar ist? Selbst wenn wir für B eine beliebige $n \times m$-Matrix nehmen, die Anzahl m der Spaltenvektoren $\vec{b}_i$

also nicht mehr gleich der Anzahl der Zeilen ist, lassen sich die Rechenschritte des GRAM-SCHMIDTschen Orthogonalisierungsverfahrens weiterhin durchführen – mit einem wesentlichen Unterschied:

Betrachten wir eine Folge $\vec{b}_1, \ldots, \vec{b}_m$ von Vektoren aus $\mathbb{R}^n$ *bzw.* $\mathbb{C}^n$, versehen mit dem EUKLIDischen oder HERMITEschen Standardskalarprodukt. Wir konstruieren wieder im k-ten Schritt zunächst eine Orthogonalbasis des von $\vec{b}_1$ bis $\vec{b}_k$ aufgespannten Untervektorraums.

Schon im ersten Schritt kann es eine Änderung geben: Der Vektor $\vec{b}_1$ könnte der Nullvektor sein und damit definitiv nicht als erster Basisvektor in Frage kommen.

Wir können also im ersten Schritt nicht einfach den Vektor $\vec{b}_1$ als ersten Vektor der zu konstruierenden Orthogonalbasis nehmen, sondern wir müssen den ersten Vektor $\vec{b}_\ell$ nehmen, der vom Nullvektor verschieden ist.

Die eventuell davor stehenden Nullvektoren können wir allerdings nicht ganz vergessen, denn sie sind schließlich Spalten der Matrix und müssen am Ende auch im Produkt CR wieder auftauchen. Wir müssen deshalb in der Matrix R die ersten $\ell - 1$ Zeilen auf Null setzen.

Danach wird $\vec{c}_1 = \vec{b}_\ell$ gesetzt; gegebenenfalls kann man den Vektor auch gleich noch auf Länge eins normieren: Zumindest wenn man von Hand rechnet, wird das davon abhängen, wie kompliziert die Länge ist. Die ℓ-te Zeile von R beginnt mit einer Eins (oder der Länge von $\vec{b}_\ell$, falls $\vec{c}_1$ auf Länge Eins normiert wurde) und enthält ansonsten lauter Nullen.

Im $(k+1)$-ten Schritt, für $k \geq 1$, gehen wir davon aus, daß wir eine Orthogonalbasis $(\vec{c}_1, \ldots, \vec{c}_k)$ des von $\vec{b}_1$ bis $\vec{b}_\ell$ aufgespannten Untervektorraums gefunden haben für ein $\ell \leq m$. Falls $\ell < m$ ist, machen wir den Ansatz

$$\vec{c}_{k+1} = \vec{b}_{\ell+1} - r_{\ell+1,1}\vec{c}_1 - \cdots - r_{\ell+1,k}\vec{c}_k$$

und bestimmen die Koeffizienten $\vec{r}_{ij}$ so, daß $\vec{c}_{k+1}\vec{c}_j = 0$ ist für alle $j \leq k$, d.h.

$$r_{\ell+1,j} = \frac{\vec{b}_{\ell+1} \cdot \vec{c}_j}{\vec{c}_j \cdot \vec{c}_j} .$$

Falls $\vec{b}_{\ell+1}$ linear unabhängig ist von $\vec{b}_1, \ldots, \vec{b}_\ell$, ist dann

$$\vec{c}_{k+1} = \vec{b}_{\ell+1} - r_{\ell+1,1}\vec{c}_1 - \cdots - r_{\ell+1,k}\vec{c}_k$$

linear unabhängig von $\vec{c}_1, \ldots, \vec{c}_k$, denn diese Vektoren erzeugen denselben Untervektorraum wie $\vec{b}_1, \ldots, \vec{b}_\ell$.

Wenn allerdings $\vec{b}_{\ell+1}$ linear abhängig ist von $\vec{b}_1, \ldots, \vec{b}_\ell$, ist $\vec{b}_{\ell+1}$ auch linear abhängig von $\vec{c}_1, \ldots, \vec{c}_k$; da diese Vektoren eine Orthogonalbasis des von ihnen erzeugten Vektorraums sind, ist daher

$$\vec{b}_{\ell+1} = \lambda_1\vec{c}_1 + \cdots + \lambda_k\vec{c}_k \quad \text{mit} \quad \lambda_j = \frac{\vec{b}_{\ell+1} \cdot \vec{c}_j}{\vec{c}_j \cdot \vec{c}_j} .$$

Vergleicht man dies mit der obigen Formel für $r_{\ell+1,j}$, so sieht man, daß in diesem Fall $\vec{c}_{k+1}$ der Nullvektor ist, wir bekommen also kein neues Element der Orthogonalbasis.

In diesem Fall müssen wir daher mit dem nächsten Vektor $\vec{b}_{\ell+2}$ – so es einen gibt – einen neuen Ansatz

$$\vec{c}_{k+1} = \vec{b}_{\ell+2} - r_{\ell+1,1}\vec{c}_1 - \cdots - r_{\ell+2,k}\vec{c}_k$$

machen, und wieder die Koeffizienten $\vec{r}_{ij}$ so bestimmen, daß $\vec{c}_{k+1}\vec{c}_j = 0$ ist für alle $j \leq k$, d.h.

$$r_{\ell+2,j} = \frac{\vec{b}_{\ell+2} \cdot \vec{c}_j}{\vec{c}_j \cdot \vec{c}_j} ,$$

wobei der entstehende Vektor $\vec{c}_{k+1}$ wieder der Nullvektor sein kann und so weiter, bis wir entweder einen vom Nullvektor verschiedenen Vektor $\vec{c}_{k+1}$ erhalten oder aber kein neuer Vektor $\vec{b}_j$ mehr existiert.

Im ersteren Fall machen wir weiter mit dem $(k+2)$-ten Schritt, andernfalls haben wir eine Basis $(\vec{c}_1, \ldots, \vec{c}_k)$ gefunden für den von $\vec{b}_1, \ldots, \vec{b}_m$ aufgespannten Untervektorraum von $\mathbb{R}^n$ *bzw.* $\mathbb{C}^n$.

Das muß allerdings noch keine Basis des gesamten $\mathbb{R}^n$ *bzw.* $\mathbb{C}^n$ sein, denn offensichtlich ist die Anzahl der Vektoren $\vec{c}_k$, die wir so erhalten, gleich dem Rang von A, und der könnte auch kleiner als n sein – für $m < n$ muß das sogar so sein.

Wenn wir eine quadratische Matrix C suchen, müssen wir daher gegebenenfalls noch die Vektoren $\vec{c}_1, \dots, \vec{c}_k$ zu einer vollen Orthogonalbasis ergänzen; dazu können wir beispielsweise so lange GRAM-SCHMIDT auf Vektoren aus der Standardbasis anwenden, bis wir n orthogonale Vektoren $\vec{c}_i$ gefunden haben.

Falls alle Vektoren $\vec{c}_i$ gleich auf Länge eins normiert wurden, bilden sie sogar eine Orthonormalbasis. Da eine solche Normierung allerdings oft Nenner mit Wurzeln produziert, ist es oft rechnerisch angenehmer, die $\vec{c}_i$ unnormiert zu lassen. In diesem Fall müssen sie zum Schluß auf die Länge eins gebracht werden; wir setzen dazu

$$\vec{q}_i = \frac{1}{|\vec{c}_i|}\vec{c}_i$$

und definieren Q als die $n \times n$-Matrix mit Spaltenvektoren $\vec{q}_1$ bis $\vec{q}_n$.

Da die Einträge der Matrix R bislang so berechnet waren, daß $CR = A$ ist, müssen wir auch diese noch modifizieren: Um von C auf Q zu kommen, haben wir die i-te Spalte durch $|\vec{c}_i|$ dividiert; zur Kompensation müssen wir daher in R die i-te *Zeile* mit $|\vec{c}_i|$ *multiplizieren;* dann haben wir mit der so modifizierten Matrix R die Produktzerlegung $A = QR$, die sogenannte QR-Zerlegung von A.

Obwohl die Matrix R hier nicht mehr quadratisch ist, wollen wir sie als (obere) Dreiecksmatrix bezeichnen, denn auch hier verschwindet r_{ij} wann immer $i > j$ ist.

Damit haben wir gezeigt:

Satz: Zu jeder reellen oder komplexen $n \times m$-Matrix A gibt es eine $n \times n$-Matrix Q, deren Spalten eine Orthonormalbasis von $\mathbb{R}^n$ *bzw.* $\mathbb{C}^n$ bilden, sowie eine obere Dreiecksmatrix $R \in \mathbb{R}^{n\times m}$ *bzw.* $\mathbb{C}^{n\times m}$, so daß gilt: $A = QR$.
Hat A einen Rang $k < n$, so enthalten die letzten $n - k$ Zeilen von R nur Nullen; daher ist auch $A = Q'R'$, wobei Q' die $n \times k$-Matrix aus den ersten k Spalten von Q ist und die $k \times m$-Matrix R' aus den ersten k Zeilen von R besteht. ■

Als Beispiel betrachten wir die Matrix

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

ihre Spaltenvektoren bezeichnen wir mit $\vec{a}_1, \dots, \vec{a}_5$.

$\vec{a}_1$ hat bereits die Länge eins, kann also als erster Vektor $\vec{q}_1$ der Orthonormalbasis genommen werden, so daß der erste Spaltenvektors $\vec{r}_1$ der Matrix R einfach der erste Koordinateneinheitsvektor ist.

Der zweite Spaltenvektor $\vec{a}_2$ von A hat mit $\vec{q}_1$ das Skalarprodukt eins; da $\vec{q}_1$ selbst bereits Länge eins hat, setzen wir also

$$\vec{q}_2 = \vec{a}_2 - \vec{q}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{und} \quad \vec{r}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

denn $\vec{a}_2 = \vec{q}_1 + \vec{q}_2$.

Im dritten Schritt suchen wir zunächst einen Vektor der Form

$$\vec{c}_3 = \vec{a}_3 + \lambda\vec{q}_1 + \mu\vec{q}_3 ,$$

der auf $\vec{q}_1$ und $\vec{q}_2$ senkrecht steht. Da $\vec{q}_1$ und $\vec{q}_2$ die Länge eins haben, ist hier einfach $\lambda = -\vec{a}_3 \cdot \vec{q}_1 = -3$ und $\mu = -\vec{a}_3 \cdot \vec{q}_2 = -2$, der gesuchte Vektor ist also der dreifache vierte Einheitsvektor, und als zugehörigen Vektor $\vec{q}_3$ der Länge eins nehmen wir den vierten Einheitsvektor selbst. Dann ist

$$\vec{a}_3 = 3\vec{q}_1 + 2\vec{q}_2 + 3\vec{q}_3, \quad \text{also} \quad \vec{r}_3 = \begin{pmatrix} 3 \\ 2 \\ 3 \\ 0 \end{pmatrix}.$$

Beim nächsten Spaltenvektor $\vec{a}_4$ sieht man eigentlich bereits ohne Rechnung, daß er im Erzeugnis von $\vec{q}_1$ bis $\vec{q}_3$ liegt, denn offensichtlich ist $\vec{a}_4 = \vec{a}_3 + \vec{q}_2 + \vec{q}_3$. Wenn wir trotzdem stur nach Schema F losrechnen mit dem Ansatz

$$\vec{c}_4 = \vec{a}_4 + \lambda_1\vec{q}_1 + \lambda_2\vec{q}_2 + \lambda_3\vec{q}_3 \quad \text{mit} \quad \lambda_i = -\vec{a}_4 \cdot \vec{q}_i ,$$

erhalten wir erwartungsgemäß $\vec{c}_4 = \vec{0}$, also keinen neuen Vektor der Orthogonalbasis. Wir bekommen aber, da wir auch $\vec{a}_4$ als Linearkombination der $\vec{q}_i$ darstellen müssen, eine neue Spalte $\vec{r}_4$ der Matrix R; wegen

$$\vec{a}_4 = 3\vec{q}_1 + 3\vec{q}_2 + 4\vec{q}_3 \quad \text{ist} \quad \vec{r}_4 = \begin{pmatrix} 3 \\ 3 \\ 4 \\ 0 \end{pmatrix} .$$

In der letzten Spalte von A stehen lauter Nullen; also brauchen wir auch hier keinen neuen Basisvektor und sehen auch ohne jede Rechnung, daß $\vec{r}_5$ gleich dem Nullvektor ist.

Damit kennen wir die Matrix R vollständig, allerdings fehlt für eine vollständige QR-Zerlegung noch eine Spalte von Q. Diese kann offensichtlich völlig unabhängig von A gewählt werden als irgendein Vektor, der auf $\vec{q}_1$ bis $\vec{q}_3$ senkrecht steht. Die kanonische Wahl ist natürlich der dritte Einheitsvektor; die einzige Alternative dazu wäre sein negatives.

Damit ist $A = QR$ mit

$$Q = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{und} \quad R = \begin{pmatrix} 1 & 1 & 3 & 3 & 4 & 0 \\ 0 & 1 & 2 & 3 & 4 & 0 \\ 0 & 0 & 3 & 4 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} .$$

Natürlich ist auch

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 3 & 3 & 4 & 0 \\ 0 & 1 & 2 & 3 & 4 & 0 \\ 0 & 0 & 3 & 4 & 5 & 0 \end{pmatrix} ,$$

denn die vierte Spalte von Q wird ja bei der Multiplikation mit R stets mit Null gewichtet.

g) Orthogonale und unitäre Matrizen

Bei der Lektüre des letzten Abschnitts hat sich sicherlich mancher die Frage gestellt, warum wir selbst bei kleinem Rang von A an einer Zerlegung interessiert sind, bei der Q eine quadratische Matrix ist, deren Spalten eine Orthonormalbasis des gesamten $\mathbb{R}^n$ *bzw.* $\mathbb{C}^n$ bilden. Der

Grund liegt darin, daß solche Matrizen Q eine ganze Reihe interessanter Eigenschaften haben, die zur Lösung mehrerer wichtiger Probleme verwendet werden können. Dies wollen wir uns im folgenden etwas genauer ansehen.

Betrachten wir zunächst den reellen Fall.

Sind $\vec{q}_i$ die Spaltenvektoren von Q, so ist die Tatsache, daß die $\vec{q}_i$ eine Orthonormalbasis bilden, äquivalent dazu, daß $\vec{q}_i \cdot \vec{q}_j = \delta_{ij}$ für alle i, j. Da wir $\vec{q}_j$ nicht nur als j-te Spalte von Q, sondern auch als j-te Zeilte der transponierten Matrix ${}^tQ$ auffassen können, können wir diese n^2 Gleichungen zusammenfassen zu der einen Matrixgleichung ${}^tQQ = E$.

Auch im HERMITEschen Fall ist $\vec{q}_i \cdot \vec{q}_j = \delta_{ij}$ für alle i, j, aber nun ist das HERMITEsche Skalarprodukt nicht mehr durch dieselbe Formel definiert, die auch bei der Matrixmultiplikation Anwendung findet, sondern enthält noch eine komplexe Konjugation im zweiten Faktor. Daher sind die Gleichungen $\vec{q}_i \cdot \vec{q}_j = \delta_{ij}$ hier äquivalent dazu, daß ${}^tQ\overline{Q}$ die Einheitsmatrix ist. Transponieren macht dies zu ${}^t({}^tQ\overline{Q}) = {}^t\overline{Q}Q = {}^tE = E$.

Die Schreibweise ${}^t\overline{Q}$ ist etwas umständlich; deshalb führen wir eine Abkürzung ein:

Definition: Die *adjungierte Matrix* zu einer Matrix $A \in \mathbb{C}^{n\times m}$ ist die Matrix $A^* = {}^t\overline{A} = \overline{{}^tA}$.

Damit ist also beispielsweise

$$A^* = \begin{pmatrix} 1-i & 2-i & 3-i \\ 4-i & 5-i & 6-i \end{pmatrix} \quad \text{für} \quad A = \begin{pmatrix} 1+i & 4+i \\ 2+i & 5+i \\ 3+i & 6+i \end{pmatrix} .$$

Für eine reelle Matrix A ist natürlich $A^* = {}^tA$ einfach die transponierte Matrix.

Definition: *a)* Eine Matrix $Q \in \mathbb{R}^{n\times n}$ heißt *orthogonal*, wenn ${}^tQQ = E$ ist.
b) Eine Matrix $U \in \mathbb{C}^{n\times n}$ heißt *unitär*, wenn $U^*U = E$ ist.

Die Matrix Q aus der QR-Zerlegung einer reellen Matrix ist somit orthogonal; im Falle einer komplexen Matrix ist Q unitär.

Lemma: *a)* Eine Matrix $Q \in \mathbb{R}^{n\times n}$ ist genau dann orthogonal, wenn $Q^{-1} = {}^tQ$ ist. Insbesondere ist jede orthogonale Matrix invertierbar.
b) Eine Matrix $U \in \mathbb{C}^{n\times n}$ ist genau dann unitär, wenn $U^{-1} = U^*$ ist; insbesondere ist jede unitäre Matrix invertierbar.
c) Die inverse Matrix einer orthogonalen *bzw.* unitären Matrix ist wieder orthogonal *bzw.* unitär.
d) Das Produkt zweier orthogonaler *bzw.* unitärer Matrizen ist wieder orthogonal *bzw.* unitär.

Beweis: Da Unitarität und Orthogonalität für reelle Matrizen äquivalent sind, können wir uns auf unitäre Matrizen beschränken.

a) und *b)* sind klar, denn nach Definition einer unitären Matrix U ist U^* invers zu U. Für *c)* muß somit gezeigt werden, daß U^* wieder unitär ist. Nach Definition ist U invers zu dieser Matrix, wegen $(U^*)^* = U$ folgt die Unitarität.

Für *d)* schließlich seien U_1 und U_2 zwei unitäre Matrizen; dann ist

$$(U_1U_2)^* = U_2^*U_1^* = U_2^{-1}U_1^{-1} = (U_1U_2)^{-1} ,$$

und damit ist auch U_1U_2 unitär. ■

Insbesondere die Aussagen *a)* und *b)* zeigen einen großen praktischen Vorteil orthogonaler und unitärer Matrizen: Man kann sie mit minimalem Aufwand invertieren.

Ist etwa $A\vec{x} = \vec{b}$ ein lineares Gleichungssystem mit einer $n \times m$-Koeffizientenmatrix A, und ist $A = QR$ die QR-Zerlegung von A, so ist

$$A\vec{x} = \vec{b} \iff QR\vec{x} = \vec{b} \iff R\vec{x} = Q^{-1}\vec{b} = Q^*\vec{b} .$$

Da R eine obere Dreiecksmatrix ist, hat das Produkt $R\vec{x}$ ausgeschrieben die Treppengestalt, die der GAUSSS-Algorithmus produziert, und die rechte Seite läßt sich für jede neue rechte Seite zum Preis einer einzigen Matrixmultiplikation berechnen. Im Gegensatz zur LR-Zerlegung ist also keine Matrixinversion nötig, und dazu ist diese Methode auch noch numerisch stabiler, falls man mit Gleitkommazahlen rechnet und einen guten numerischen Algorithmus zur Berechnung der QR-Zerlegung verwendet. (GRAM-SCHMIDT ist numerisch eher nicht zu empfehlen.)

Der wesentliche Grund für die guten numerischen Eigenschaften orthogonaler und unitärer Matrizen besteht darin, daß sie Längen respektieren. Um das einzusehen, beweisen wir zunächst ein allgemeines Lemma (das auch der historische Grund für die Bezeichnung „adjungierte Matrix“ ist):

Lemma: Für $k = \mathbb{R}$ oder $\mathbb{C}$ und $A \in k^{n\times m}$ ist

$$(A\vec{v}) \cdot \vec{w} = \vec{v} \cdot (A^*\vec{w}) \quad \text{für alle } \vec{v} \in k^m \text{ und } \vec{w} \in k^n ,$$

wobei links das (HERMITEsche) Standardskalarprodukt von k^n steht und rechts das von k^m.

Beweis: Es genügt, den HERMITEschen Fall zu betrachten. Dazu fassen wir einen Vektor $\vec{w} \in \mathbb{C}^n$ auf als eine $n \times 1$-Matrix $w_M \in \mathbb{C}^{n\times 1}$; für einen weiteren Vektor $\vec{u} \in \mathbb{C}^n$, aufgefaßt als Matrix $u_M \in \mathbb{C}^{n\times 1}$, ist dann das HERMITEsche Skalarprodukt $\vec{v} \cdot \vec{w}$ gleich dem Matrixprodukt ${}^tv_M \cdot \overline{w_M}$.

Ganz entsprechend ordnen wir dem Vektor $\vec{v} \in \mathbb{C}^m$ eine $m \times 1$-Matrix $v_M \in \mathbb{C}^{m\times 1}$ zu; die Matrix zum Vektor $A\vec{v}$ ist dann die Produktmatrix Av_M.

Somit ist

$$\begin{aligned}(A\vec{v}) \cdot \vec{w} &= {}^t(Av_M) \cdot \overline{w_M} = {}^tv_M \cdot {}^tA \cdot \overline{w_M} = {}^tv_M \cdot \overline{\overline{{}^tA} \cdot w_M} \\ &= {}^tv_M \cdot \overline{A^* \cdot w_M} = \vec{v} \cdot (A^*\vec{w}) .\end{aligned}$$

■

Als mehr oder weniger unmittelbare Folgerung erhalten wir:

Satz: *a)* Eine Matrix $A \in \mathbb{R}^{n\times n}$ ist genau dann orthogonal, wenn für das Standardskalarprodukt des $\mathbb{R}^n$ gilt: $(A\vec{v}) \cdot (A\vec{w}) = \vec{v} \cdot \vec{w}$ für alle $\vec{v}, \vec{w} \in \mathbb{R}^n$.
b) Eine Matrix $A \in \mathbb{C}^{n\times n}$ ist genau dann unitär, wenn für das HERMITEsche Standardprodukt des $\mathbb{C}^n$ gilt: $(A\vec{v}) \cdot (A\vec{w}) = \vec{v} \cdot \vec{w}$ für alle $\vec{v}, \vec{w} \in \mathbb{C}^n$.

Beweis: Nach dem gerade bewiesenen Lemma ist

$$(A\vec{v}) \cdot (A\vec{w}) = \vec{v} \cdot \big(A^*(A\vec{w})\big) = \vec{v} \cdot \big((A^*A)\vec{w}\big) .$$

A ist genau dann orthogonal *bzw.* unitär, wenn AA^* gleich der Einheitsmatrix ist; in diesem Fall ist $(A^*A)\vec{w} = \vec{w}$ und damit $(A\vec{v})\cdot(A\vec{w}) = \vec{v}\cdot\vec{w}$.

Falls wir umgekehrt wissen, daß $(A\vec{v})\cdot(A\vec{w}) = \vec{v}\cdot\vec{w}$ ist für alle Vektoren $\vec{v}, \vec{w}$, so ist auch $\vec{v}\cdot\big((A^*A)\vec{w}\big) = \vec{v}\cdot\vec{w}$, insbesondere also

$$\vec{e}_i \cdot \big((A^*A)\vec{e}_j\big) = \vec{e}_i \cdot \vec{e}_j = \delta_{ij}$$

für die Koordinateneinheitsvektoren.

$(A^*A)\vec{e}_j$ ist die j-te Spalte der Matrix A^*A, ihr Skalarprodukt mit $\vec{e}_i$ also der ij-Eintrag von A^*A. Da dieser gleich δ_{ij} sein muß, ist also $A^*A = E$ und A somit orthogonal *bzw.* unitär. ■

h) Orthogonale Projektionen

Ist U ein r-dimensionaler Untervektorraum eines n-dimensionalen Vektorraums V, so können wie jede Basis $\{\vec{b}_1, \ldots, \vec{b}_r\}$ von U ergänzen zu einer Basis $\{\vec{b}_1, \ldots, \vec{b}_n\}$ von V. Der von $\vec{b}_{r+1}$ bis $\vec{b}_n$ erzeugte Untervektorraum $W \le V$ hat dann die Eigenschaft, daß $U \cap W$ der Nullraum ist, während $U \cup W$ dem gesamten Vektorraum V erzeugt. Einen solchen Untervektorraum W bezeichnen wir als *Komplement* von U; es ist natürlich, genau wie seine Basisvektoren $\vec{b}_{r+1}$ bis $\vec{b}_n$, alles andere als eindeutig bestimmt.

Für EUKLIDische und HERMITEsche Vektorräume können wir allerdings jedem Untervektorraum ein wohlbestimmtes ausgezeichnetes Komplement zuordnen, das *orthogonale Komplement.*

Definition: V sei ein EUKLIDischer oder HERMITEscher Vektorraum und $U \le V$ sei ein Untervektorraum von V. Das orthogonale Komplement von U ist der Untervektorraum

$$U^\perp \underset{\text{def}}{=} \big\{\vec{v} \in V \;\big|\; \vec{u}\cdot\vec{v} = 0 \;\text{ für alle } \vec{u} \in U\big\}\,.$$

Wegen der Linearität des EUKLIDischen wie auch HERMITEschen Skalarprodukts im ersten Argument ist klar, daß $U^\perp$ ein Untervektorraum von V ist. Außerdem ist klar, daß es reicht die Bedingung $\vec{v}\cdot\vec{u} = 0$ für die Vektoren $\vec{u}$ aus einer Basis von U nachzurechnen, denn wenn alle

diese Produkte verschwinden, verschwindet auch jedes Produkt mit einer Linearkombination solcher Vektoren. (Es stört dabei nicht, daß wir im HERMITEschen Fall keine Linearität im zweiten Argument haben, sondern die Koeffizienten komplex konjugieren müssen.)

Lemma: U sei ein Untervektorraum des n-dimensionalen EUKLIDischen oder HERMITEschen Vektorraums V, und $(\vec{b}_1, \ldots, \vec{b}_r)$ sei eine Orthogonalbasis von U. Ergänzt man diese zu einer Orthogonalbasis $(\vec{b}_1, \ldots, \vec{b}_n)$ von V, so ist $(\vec{b}_{r+1}, \ldots, \vec{b}_n)$ eine Orthogonalbasis von $U^\perp$. Insbesondere hat also das orthogonale Komplement eines r-dimensionalen Untervektorraum die Dimension $n-r$ und $U \cap U^\perp = \{\vec{0}\}$.

Beweis: Nach dem gerade Gesagten liegt ein Vektor $\vec{v} = \lambda_1\vec{b}_1 + \cdots + \lambda_n\vec{b}_n$ aus V genau dann in $U^\perp$, wenn für alle $i \le r$ gilt:

$$\vec{v}\cdot\vec{b}_i = \left(\sum_{j=1}^{n} \lambda_j\vec{b}_j\right)\cdot\vec{b}_i = \sum_{j=1}^{n} \lambda_j\vec{b}_j\cdot\vec{b}_i = \lambda_i\vec{b}_i\cdot\vec{b}_i = 0\,.$$

Da $\vec{b}_i$ als Basisvektor nicht der Nullvektor sein kann, ist $\vec{b}_i\cdot\vec{b}_i \ne 0$; daher ist dies äquivalent zum Verschwinden aller λ_i mit $i \le r$, also zur Darstellbarkeit von $\vec{v}$ als Linearkombination der Vektoren $\vec{b}_{r+1}, \ldots, \vec{b}_n$. Als Teil einer Basis sind diese linear unabhängig, also Basis ihres Erzeugnisses $U^\perp$. ■

Korollar: *a)* V sei ein endlichdimensionaler EUKLIDischer oder HERMITEscher Vektorraum, und $U \le V$ sei ein Untervektorraum. Dann läßt sich jedes Element $\vec{v} \in V$ eindeutig schreiben als $\vec{v} = \vec{u} + \vec{w}$ mit $\vec{u} \in U$ und $\vec{w} \in U^\perp$.
b) ${U^\perp}^\perp = U$

Beweis: a) Wir wählen eine Orthogonalbasis $(\vec{b}_1, \ldots, \vec{b}_r)$ von U und ergänzen sie zu einer Orthogonalbasis $(\vec{b}_1, \ldots, \vec{b}_n)$ von V; nach dem Lemma ist dann $(\vec{b}_{r+1}, \ldots, \vec{b}_n)$ eine Orthogonalbasis von $U^\perp$. Schreiben wir $\vec{v} = v_1\vec{b}_1 + \cdots + v_n\vec{b}_n$, so ist also

$$\vec{u} \underset{\text{def}}{=} v_1\vec{b}_1 + \cdots + v_r\vec{b}_r \in U, \quad \vec{w} \underset{\text{def}}{=} v_{r+1}\vec{b}_{r+1} + \cdots + v_n\vec{b}_n \in U^\perp$$

und $\vec v = \vec u + \vec w$.

Ist $\vec v = \vec x + \vec y$ irgendeine Darstellung von $\vec v$ als Summe zweier Vektoren $\vec x \in U$ und $\vec y \in V$, so ist

$$\vec u + \vec w = \vec x + \vec y \Longrightarrow \vec u - \vec x = \vec y - \vec w\,.$$

In der letzteren Gleichung steht links der Vektor $\vec u - \vec x \in U$ und rechts $\vec y - \vec w \in U^\perp$; wegen $U \cap U^\perp = \{\vec 0\}$ ist also $\vec u = \vec x$ und $\vec w = \vec y$, was die Eindeutigkeit dieser Zerlegung zeigt.·

b) Für $\vec u \in U$ und $\vec w \in U^\perp$ verschwindet nach Definition von $U^\perp$ das Produkt $\vec w \cdot \vec u$, also wegen dessen (HERMITEscher) Symmetrie auch $\vec u \cdot \vec w$. Damit ist

$$\vec u \in {U^\perp}^\perp = \left\{\vec v \in V \;\middle|\; \vec v \cdot \vec w = 0 \;\text{ für alle } \vec w \in U^\perp\right. ,$$

also liegt U in ${U^\perp}^\perp$. Nach dem obigen Lemma ist

$$\dim {U^\perp}^\perp = \dim V - \dim U^\perp = \dim V - (\dim V - \dim U) = \dim U\,,$$

also muß $U = {U^\perp}^\perp$ sein ■

Bemerkung: Tatsächlich gilt dieses Korollar auch für unendlichdimensionale Vektorräume; da die Existenz von wie auch der Umgang mit Basen im Unendlichdimensionalen etwas problematisch ist, soll aber hier, wie bereits mehrfach in diesem Skriptum, der endlichdimensionale Fall genügen.

Definition: V sei ein endlichdimensionaler EUKLIDischer oder HERMITEscher Vektorraum, und $U \le V$ sei ein Untervektorraum. Die Abbildung $\pi_U\colon V \to U$, die jedem Vektor $\vec v = \vec u + \vec w \in V$ mit $\vec u \in U$ und $\vec w \in U^\perp$ den Vektor $\vec u$ zuordnet, heißt *orthogonale Projektion* von V nach U.

Wegen der eindeutigen Zerlegbarkeit eines Vektors in eine Komponente aus U und eine aus $U^\perp$ ist π_U offensichtlich wohldefiniert und linear; der Kern von π_U ist $U^\perp$.

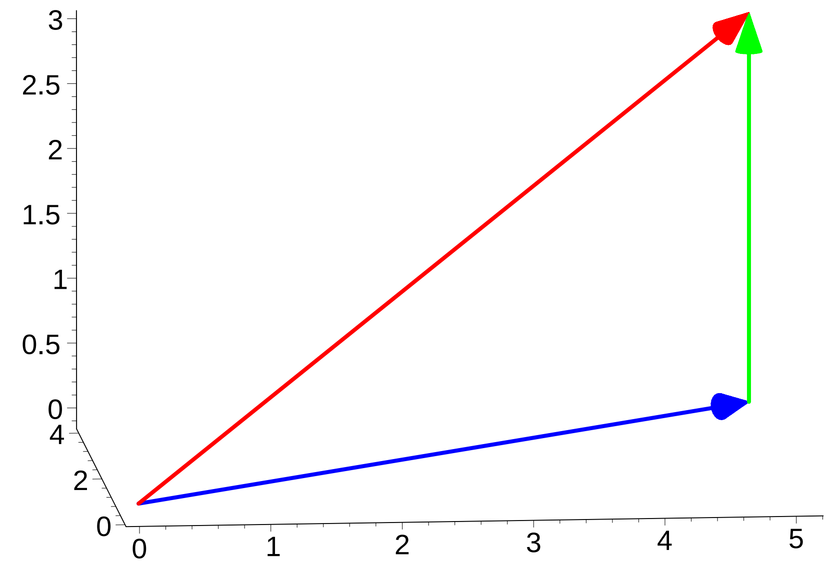

Abb. 15: Orthogonale Projektion eines Vektors

Orthogonale Projektionen sind aus der Geometrie bekannt, beispielsweise als Grundriß, Aufriß und Kreuzriß eines dreidimensionalen Körpers; uns interessiert hier vor allem ihre folgende Eigenschaft:

Lemma: V sei ein endlichdimensionaler EUKLIDischer oder HERMITEscher Vektorraum, $U \le V$ ein Untervektorraum und $\vec v \in V$. Dann gilt für jeden Vektor $\vec u \in U$ die Ungleichung $|\vec v - \vec u| \le |\vec v - \pi_U(\vec v)|$, d.h. $\pi_U(\vec v)$ ist derjenige Vektor aus U, dessen Differenz mit $\vec v$ am kürzesten ist.

Beweis: Wir schreiben $\vec v = \vec p + \vec w$ mit $\vec p = \pi_U(\vec v) \in U$ und $\vec w \in U^\perp$. Für jeden Vektor $\vec u \in U$ ist dann

$$\begin{aligned}|\vec v - \vec u|^2 = |\vec p + \vec w - \vec u|^2 &= |(\vec p - \vec u) + \vec w|^2\\ &= (\vec p - \vec u)\cdot(\vec p - \vec u) + \vec w \cdot (\vec p - \vec u) + (\vec p - \vec u)\cdot \vec w + \vec w \cdot \vec w\\ &= |\vec p - \vec u|^2 + |\vec w|^2\,,\end{aligned}$$

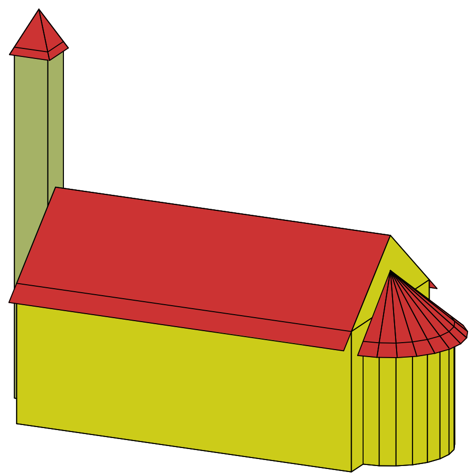

Abb. 16: Ein dreidimensionales Objekt

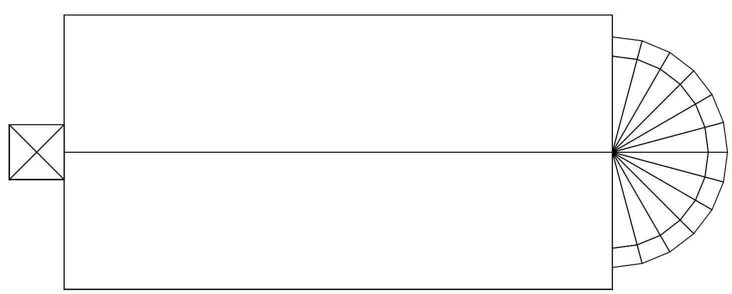

Abb. 17: Der Grundriß des Objekts aus Abbildung 16

denn $\vec{p}-\vec{u}$ liegt in U und $\vec{w}$ in $U^{\perp}$. Also ist $|\vec{v}-\vec{u}|$ nie kleiner als $|\vec{v}-\vec{p}|$, und die beiden Vektoren sind genau dann gleich lang, wenn $\vec{u}-\vec{p}$ der Nullvektor ist, also $\vec{u}=\pi_U(\vec{v})$. ■

Betrachten wir orthogonale Projektionen statt in Vektorräumen in den dazugehörigen affinen Räumen, entsprich die orthogonale Projektion auf einen Unterraum geometrisch also einfach der Konstruktion des Lotfußpunkts in diesem Unterraum.

i) Die Methode der kleinsten Quadrate

Oftmals ist zu gegebenen Beobachtungsdaten grundsätzlich bekannt, welcher *Art* von Gesetz sie genügen sollten; das Problem besteht „nur“ noch darin, die in diesem Gesetz vorkommenden *Parameter* zu bestimmen. Im einfachsten Fall könnte man etwa an einen Widerstand denken, der dadurch gemessen wird, daß man verschiedene Spannungen U_i anlegt und die zugehörigen Stromstärken I_i mißt. Nach dem Ohmschen Gesetz ist dann $U_i = R\cdot I_i$, aber aufgrund der unvermeidlichen Meßfehler werden die verschiedenen Quotienten U_i/I_i natürlich nicht alle gleich sein. Die Lösung dieses Problems ist klar: Man nimmt den Mittelwert der Quotienten. Schwieriger wird es, wenn mehrere Parameter ins Spiel kommen, wenn die Meßreihe als mehr als nur einen Parameter bestimmen soll.

Solche Fälle treten nicht nur auf in Naturwissenschaft und Technik, sondern auch in den Wirtschafts- und Sozialwissenschaften, wo es zwar selten *exakte* Gesetze gibt, man den Zusammenhang zwischen verschiedenen Größen aber trotzdem zumindest näherungsweise durch eine mathematische Formel beschreiben will – auch wenn diese in konkreten Einzelfällen gelegentlich ziemlich falsch sein kann.

Als Beispiel dieser Art können wir den Zusammenhang zwischen Korruption und Wohlstand in verschiedenen Staaten betrachten: edes Jahr veröffentlicht die Organisation *Transparancy International* ihren *corruption perceptions index (CPI)*, in dem jedem Land eine Zahl zwischen null und zehn zugeordnet wird, je nachdem, wie stark Geschäftsleute, Risikospezialisten und die Bevölkerung die Korruption im betreffenden Land einschätzen: Ein Index von zehn bedeutet, daß es praktische keine Korruption gibt, während bei null nichts läuft ohne Bimbes. Die neuesten Daten stammen vom 7. Oktober 2003 und sind unter

`http://www.transparency.org/cpi/`

zu finden. Die Zahlen werden als Mittelwerte über die letzten drei Jahren berechnet, so daß singuläre Ereignisse eines Jahres nicht zu sehr ins Gewicht fallen.

Wir vergleichen diese Zahlen mit dem Bruttoinlandsprodukt pro Einwohner, wie es die Weltbank für 2001 festgestellt hat. Dies sind die neuesten verfügbaren Zahlen; sie sind beispielsweise auf dem Server des Statistischen Bundesamtes unter

`http://www.destatis.de/ausl_prog/suche_ausland.htm`

zu finden, indem man unter „Indikatoren“ das Feld „BIP je Einwohner (real)“ auswählt. In der folgenden Tabelle sind Staaten aufgelistet, für die sowohl das Bruttosozialprodukt pro Einwohner als auch der CPI für 2003 vorliegt; das Bruttosozialprodukt pro Einwohner in US-$ ist kursiv gedruckt, der Korruptionsindex fett. Der Index ist jeweils ein Mittelwert aus verschiedenen Quellen; die angegebenen Fehlerschranken sind die Standardabweichungen der betreffenden Daten und geben an, wie groß die Unterschiede zwischen den einzelnen Quellen sind.

Ägypten	*1229*	**3,3** ± 1,3
Albanien	*1032*	**2,5** ± 0,6
Algerien	*1616*	**2,6** ± 0,5
Angola	*525*	**1,8** ± 0,3
Argentinien	*7468*	**2,5** ± 0,5
Armenien	*1068*	**3,0** ± 0,8
Aserbaidschan	*460*	**1,8** ± 0,3
Äthiopien	*121*	**2,5** ± 0,8
Australien	*24204*	**8,8** ± 0,9
Bahrain	*10609*	**6,1** ± 1,1
Bangladesch	*386*	**1,3** ± 0,7
Belgien	*31186*	**7,6** ± 0,9
Belize	*3189*	**4,5** ± 0,9
Bolivien	*944*	**2,3** ± 0,4
Bosnien und Herzegowina	*1584*	**3,3** ± 0,7
Botsuana	*4130*	**5,7** ± 0,9
Brasilien	*4633*	**3,9** ± 0,5
Bulgarien	*1670*	**3,9** ± 0,9
Chile	*5385*	**7,4** ± 0,9
China	*878*	**3,4** ± 1,0
Costa Rica	*3900*	**4,3** ± 0,7
Côte d’Ivoire	*715*	**2,1** ± 0,5
Dänemark	*39000*	**9,5** ± 0,4

Deutschland	*32855*	**7,7** ± 1,2
Dominikanische Republik	*2077*	**3,3** ± 0,4
Ecuador	*1478*	**2,2** ± 0,3
El Salvador	*1757*	**3,7** ± 1,5
Estland	*3457*	**5,5** ± 0,6
Finnland	*31853*	**9,7** ± 0,3
Frankreich	*30374*	**6,9** ± 1,1
Gambia	*382*	**2,5** ± 0,9
Georgien	*499*	**1,8** ± 0,7
Ghana	*421*	**3,3** ± 0,9
Griechenland	*13709*	**4,3** ± 0,8
Guatemala	*1554*	**2,4** ± 0,6
Haiti	*354*	**1,5** ± 0,6
Honduras	*711*	**2,3** ± 0,6
Indien	*477*	**2,8** ± 0,4
Indonesien	*1034*	**1,9** ± 0,5
Iran	*1714*	**3,0** ± 1,0
Irland	*29352*	**7,5** ± 0,7
Island	*32161*	**9,6** ± 0,3
Israel	*16576*	**7,0** ± 1,2
Italien	*21363*	**5,3** ± 1,1
Jamaika	*2171*	**3,8** ± 0,4
Japan	*44353*	**7,0** ± 1,1
Jemen	*316*	**2,6** ± 0,7
Jordanien	*1639*	**4,6** ± 1,1
Kamerun	*696*	**1,8** ± 0,2
Kanada	*23138*	**8,7** ± 0,9
Kasachstan	*1712*	**2,4** ± 0,9
Kenia	*325*	**1,9** ± 0,3
Kirgisistan	*417*	**2,1** ± 0,4
Kolumbien	*2277*	**3,7** ± 0,5
Kongo	*792*	**2,2** ± 0,5
Korea, Republik	*13581*	**4,3** ± 1,0
Kroatien	*5355*	**3,7** ± 0,6
Kuba	*1480*	**4,6** ± 1,0
Kuwait	*13345*	**5,3** ± 1,7

Lettland	*2560*	**3,8** ± 0,4
Libanon	*2890*	**3,0** ± 0,8
Litauen	*2195*	**4,7** ± 1,6
Luxemburg	*57720*	**8,7** ± 0,4
Madagaskar	*253*	**2,6** ± 1,8
Malawi	*163*	**2,8** ± 1,2
Malaysia	*4708*	**5,2** ± 1,1
Mali	*292*	**3,0** ± 1,8
Marokko	*1436*	**3,3** ± 1,3
Mauritius	*4352*	**4,4** ± 0,7
Mazedonien	*2417*	**2,3** ± 0,3
Mexiko	*3705*	**3,6** ± 0,6
Moldau, Republik	*678*	**2,4** ± 0,8
Mosambik	*213*	**2,7** ± 0,7
Myanmar	*2875*	**1,6** ± 0,3
Namibia	*2383*	**4,7** ± 1,3
Neuseeland	*18684*	**9,5** ± 0,2
Niederlande	*31577*	**8,9** ± 0,3
Nigeria	*257*	**1,4** ± 0,4
Norwegen	*39749*	**8,8** ± 0,5
Oman	*7110*	**6,3** ± 0,9
Österreich	*33492*	**8,0** ± 0,7
Pakistan	*517*	**2,5** ± 0,9
Panama	*3243*	**3,4** ± 0,8
Papua-Neuguinea	*897*	**2,1** ± 0,6
Paraguay	*1703*	**1,6** ± 0,3
Peru	*2311*	**3,7** ± 0,6
Philippinen	*1165*	**2,5** ± 0,5
Polen	*4758*	**3,6** ± 1,1
Portugal	*13104*	**6,6** ± 1,2
Rumänien	*1573*	**2,8** ± 1,0
Russische Föderation	*2609*	**2,7** ± 0,8
Sambia	*405*	**2,5** ± 0,6
Saudi-Arabien	*6614*	**4,5** ± 2,0
Schweden	*33242*	**9,3** ± 0,2
Schweiz	*47273*	**8,8** ± 0,8

Senegal	*629*	**3,2** ± 1,2
Serbien und Montenegro	*1240*	**2,3** ± 0,5
Sierra Leone	*158*	**2,2** ± 0,5
Simbabwe	*559*	**2,3** ± 0,3
Singapur	*27118*	**9,4** ± 0,1
Slowakei	*4406*	**3,7** ± 0,7
Slowenien	*12010*	**5,9** ± 1,2
Spanien	*18342*	**6,9** ± 0,8
Sri Lanka	*876*	**3,4** ± 0,7
Südafrika	*4068*	**4,4** ± 0,6
Sudan	*328*	**2,3** ± 0,3
Syrien, Arabische Republik	*796*	**3,4** ± 1,3
Tadschikistan	*420*	**1,8** ± 0,3
Taiwan (China)	*15283*	**5,7** ± 1,0
Tansania, Vereinigte Republik	*197*	**2,5** ± 0,6
Thailand	*2853*	**3,3** ± 0,9
Trinidad und Tobago	*5553*	**4,6** ± 1,3
Tschechische Republik	*5555*	**3,9** ± 0,9
Tunesien	*2562*	**4,9** ± 0,7
Türkei	*2810*	**3,1** ± 0,9
Uganda	*355*	**2,2** ± 0,7
Ukraine	*986*	**2,3** ± 0,6
Ungarn	*5694*	**4,8** ± 0,6
Uruguay	*5870*	**5,5** ± 1,1
Usbekistan	*512*	**2,4** ± 0,5
Venezuela	*3326*	**2,4** ± 0,5
Vereinigte Arabische Emirate	*18110*	**5,2** ± 0,5
Vereinigte Staaten	*31399*	**7,5** ± 1,2
Vereinigtes Königreich	*22112*	**8,7** ± 0,5
Vietnam	*390*	**2,4** ± 0,8
Weißrussland	*1493*	**4,2** ± 1,8
Zypern	*14592*	**6,1** ± 1,6

Abbildung 18 zeigt die 127 Datenpunkte zu dieser Liste graphisch, wobei der Punkt für Deutschland etwas heller eingezeichnet ist.

Der erste Augenschein zeigt, daß korruptionsärmere Länder oftmals

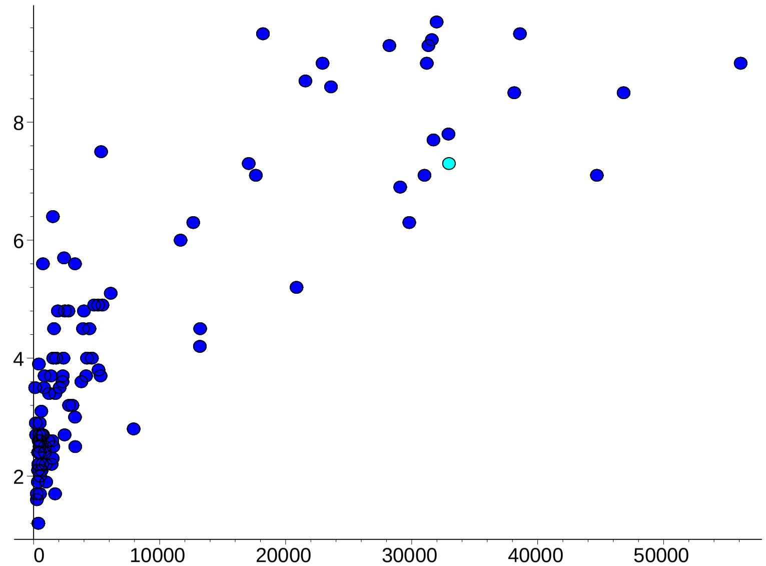

Abb. 18: Zusammenhang zwischen Korruption und Bruttoinlandsprodukt je Einwohner

reicher sind: Das weitgehend korruptionsfreie Dänemark hat ein Bruttosozialprodukt von 39 000 \$ pro Einwohner, das deutlich korruptere Deutschland nur 32 855 \$ und ein stark korruptes Land wie Bangladesch nur 386 \$. Allerdings gibt es auch Ausnahmen, denn Chile hat, obwohl kaum korrupter als Deutschland, nur ein Bruttosozialprodukt von 5 385 \$ pro Einwohner. Es gibt also sicherlich keinen deterministischen Zusammenhang zwischen Korruption und Wohlstand, aber doch eine Tendenz.

Falls wir nun versuchen, beispielsweise einen linearen Zusammenhang der Form

$$CPI = a + b \cdot BIP$$

zu finden, so haben wir 127 Gleichungen für die beiden unbekannten Koeffizienten a und b, und ein kurzer Blick auf Abbildung 18 zeigt, daß dieses lineare Gleichungssystem keine Lösung haben kann.

Wir suchen also keine Lösung, sondern zwei Zahlen a und b derart, daß

die 127 Gleichungen „möglichst gut“ gelten. Was das bedeuten soll läßt sich mathematisch auf verschiedene, nicht äquivalente Weisen definieren; da wir uns im Augenblick mit Skalarprodukten beschäftigen, bietet sich an, die 127 Bruttoinlandsprodukte pro Einwohner und die 127 Korruptionsindizes zu zwei Vektoren $\vec{x}, \vec{y} \in \mathbb{R}^{127}$ zusammenzufassen, und nach Zahlen a, b zu suchen, so daß die Länge des Differenzvektors $\vec{y} - a\vec{x} - b$ möglichst klein wird. Ausgeschrieben bedeutet dies, wenn wir die Komponenten von $\vec{x}$ mit x_i und die von $\vec{y}$ mit y_i bezeichnen, daß die Summe

$$\sum_{i=1}^{127} (y_i - ax_i - b)^2$$

der Abweichungsquadrate möglichst klein sein soll – von daher der Name „Methode der kleinsten Quadrate“ für diesen Ansatz, mit dessen Hilfe sein Schöpfer GAUSS sowohl die Position des Planetoiden Ceres vorhersagte als auch die Vermessung und Kartierung des Königreichs Hannover durchführte.

Derselbe Ansatz läßt sich natürlich auf jedes lineare Gleichungssystem über den reellen oder komplexen Zahlen anwenden: Wir haben ein möglicherweise unlösbares lineares Gleichungssystem $A\vec{x} = \vec{b}$ und wollen einen Vektor $\vec{x}$ so bestimmen, daß der Vektor $A\vec{x} - \vec{b}$ minimale Länge hat.

Falls das lineare Gleichungssystem lösbar ist, gibt es damit kein Problem: Wir bestimmen irgendeine Lösung $\vec{x}$ und haben damit einen Vektor gefunden, für den $A\vec{x} - \vec{b}$ die Länge null hat – kürzer geht es nicht.

Im allgemeinen ist aber für den gesuchten Vektor $\vec{x}$ das Produkt $A\vec{x}$ von $\vec{b}$ verschieden; es sei etwa gleich $\vec{c}$. Dann ist $\vec{c}$ ein Vektor, der sich in der Form $A\vec{x}$ darstellen läßt, und unter allen solchen Vektoren ist es derjenige, für den die Länge des Differenzvektors zu $\vec{b}$ minimal ist. Dies erinnert an die orthogonalen Projektionen aus dem vorigen Abschnitt, und in der Tat läßt sich das Problem damit lösen:

Nehmen wir an, wir haben n Gleichungen in m Unbekannten mit Koeffizienten aus $k = \mathbb{R}$ oder $k = \mathbb{C}$. Dann definiert die Matrix $A \in k^{n \times m}$

des Gleichungssystems eine lineare Abbildung

$$\varphi\colon k^m \to k^n; \quad \vec{v} \mapsto A\vec{v};$$

deren Bildraum sei U. Falls die rechte Seite $\vec{b}$ in U liegt, ist das Gleichungssystem lösbar; andernfalls suchen wir einen Vektor $\vec{x} \in k^m$, für den die Länge des Vektors $A\vec{x} - \vec{b}$ minimal wird. Da die Vektoren, die sich in der Form $A\vec{x}$ darstellen lassen, genau die Vektoren aus U sind, ist somit $A\vec{x} = \pi_U(\vec{b})$ die orthogonale Projektion von $\vec{b}$ nach U. Diese könnten wir *im Prinzip* bestimmen, indem wir die QR-Zerlegung von A berechnen, denn dann sind die ersten Spalten von Q eine Basis von U, die durch die weiteren Spalten zu einer Basis von ganz k^n ergänzt wird; danach haben wir ein lösbares lineares Gleichungssystem.

Wir wollen uns überlegen, wie wir $\vec{x}$ auch ohne die rechnerisch aufwendige QR-Zerlegung bestimmen können.

Für den gesuchten Vektor $\vec{x}$ (oder für die gesuchten Vektoren $\vec{x}$) ist $A\vec{x} = \varphi_U(\vec{b})$. Da $A\vec{x}$ bereits in U liegt, ist $\pi_U(A\vec{x}) = A\vec{x}$, also ist die Gleichung $A\vec{x} = \pi_U(\vec{b})$ äquivalent zu

$$\pi_U(A\vec{x}) = \pi_U(\vec{b}) \quad \text{oder} \quad A\vec{x} - \vec{b} \in \operatorname{Kern} \pi_U = U^\perp .$$

Das orthogonale Komplement $U^\perp$ von U besteht aus allen Vektoren $\vec{y} \in k^n$, die senkrecht stehen auf U, für die also gilt

$$(A\vec{x}) \cdot \vec{y} = 0 \quad \text{für alle } \vec{x} \in k^m .$$

Wie wir im vorletzten Abschnitt gesehen haben, ist

$$(A\vec{x}) \cdot \vec{y} = \vec{x} \cdot A^*\vec{y} \quad \text{für alle } \vec{x} \in k^m, \vec{y} \in k^n ,$$

$\vec{y}$ liegt also genau dann in $U^\perp$, wenn $A^*\vec{y}$ senkrecht steht auf allen Vektoren $\vec{x} \in k^m$. Ein solcher Vektor aus k^m ist insbesondere $A^*\vec{y}$ selbst; wegen der positiven Definitheit des (HERMITEschen) Skalarprodukts ist also $A^*\vec{y} = \vec{0}$ und somit

$$U^\perp = \left\{ \vec{y} \in k^n \;\middle|\; A^*\vec{y} = \vec{0} \right\} .$$

$A\vec{x} - \vec{b}$ liegt also genau dann im Kern von π_U, wenn $A^*(A\vec{x} - \vec{b}) = \vec{0}$ ist oder, anders ausgedrückt, wenn $\vec{x}$ eine Lösung des linearen Gleichungssystems

$$(A^*A)\,\vec{x} = A^*\vec{b}$$

ist. Da die adjungierte Matrix A^* einfach die transponierte Matrix zur komplex konjugierten Matrix zu A ist, wobei die komplexe Konjugation über $\mathbb{R}$ natürlich entfällt, läßt sich dieses Gleichungssystem schnell aufstellen und dann nach GAUSS lösen.

Betrachten wir dies konkret im eingangs diskutierten Fall eines linearen Zusammenhangs $y = ax + b$ zu N Wertepaaren $(x_i, y_i) \in \mathbb{R}^2$, wobei N sinnvollerweise größer als zwei sein sollte. Wir haben dann N Gleichungen

$$y_i = ax_i + b \quad \text{oder} \quad x_i a + b = y_i ,$$

wobei hier im Gegensatz zu unserer sonstigen Gewohnheit die Parameter a und b unbekannt sind, während die x_i und die y_i bekannt sind. Wir haben also ein lineares Geichungssystem von N Gleichungen in den beiden Variablen a und b.

Fassen wir die Werte x_i zusammen zu einem Vektor $\vec{x} \in \mathbb{R}^N$ und die y_i zu einem Vektor $\vec{y} \in \mathbb{R}^n$, so läßt sich dieses Gleichungssystem kurz schreiben als

$$\vec{x} \cdot a + \vec{1} \cdot b = \vec{y} ,$$

wobei $\vec{1} \in \mathbb{R}^N$ jenen Vektor bezeichnen soll, dessen sämtliche Komponenten eins sind.

Die Matrix des Gleichungssystems ist somit die $N \times 2$-Matrix A mit Spalten $\vec{x}$ und $\vec{1}$. Da wir mit reellen Zahlen rechnen, ist A^* einfach die transponierte Matrix dazu, also die $2 \times N$-Matrix, in deren erster Zeile die x_i stehen, während in der zweiten lauter Einser stehen. Somit ist

$${}^tAA = \begin{pmatrix} \vec{x} \cdot \vec{x} & \vec{x} \cdot \vec{1} \\ \vec{x} \cdot \vec{1} & \vec{1} \cdot \vec{1} \end{pmatrix} \quad \text{und} \quad {}^tA\vec{b} = \begin{pmatrix} \vec{x} \cdot \vec{y} \\ \vec{1} \cdot \vec{y} \end{pmatrix} ,$$

das Gleichungssystem wird also zu

$$(\vec{x} \cdot \vec{x})a + (\vec{x} \cdot \vec{1})b = \vec{x} \cdot \vec{y} \quad \text{und} \quad (\vec{x} \cdot \vec{1}) + Nb = \vec{y} \cdot \vec{1} .$$

Seine Matrix ist genau dann singulär, wenn die beiden Spalten proportional zueinander sind, wenn also $N(\vec{x} \cdot \vec{x}) = (\vec{x} \cdot \vec{1})^2$ ist. Nach der CAUCHY-SCHWARZschen Ungleichung ist

$$\left|\vec{1} \cdot \vec{x}\right| \le \left|\vec{1}\right| \cdot |\vec{x}| = \sqrt{N}\,|\vec{x}| , \quad \text{also} \quad \left|\vec{1} \cdot \vec{x}\right|^2 \le N(\vec{x} \cdot \vec{x})$$

mit Gleichheit nur dann, wenn die Vektoren $\vec{x}$ und $\vec{1}$ linear abhängig sind, wenn also alle x_i gleich sind. In diesem Fall ist die erste Gleichung ein Vielfaches der zweiten, es gibt also unendlich viele Lösungen.

Andernfalls ist die Matrix invertierbar, die Lösung also eindeutig.

Führen wir die (in der Ausgleichsrechnung ziemlich verbreiteten) Abkürzungen

$$[x^r] = \sum_{i=1}^{N} x_i^r, \quad [y^r] = \sum_{i=1}^{N} x_i^r \quad \text{und} \quad [x^r y^s] = \sum_{i=1}^{N} x_i^r y_i^s$$

ein, so erhält das Gleichungssystem die übersichtlichere Gestalt

$$[x^2]a + [x]b = [xy] \quad \text{und} \quad [x]a + Nn = [y]\,.$$

Subtraktion von $[x]/[x^2]$ mal der ersten Gleichung von der zweiten führt auf

$$\left(N - \frac{[x]^2}{[x^2]}\right) b = [y] - \frac{[x]}{[x^2]}[xy]$$

oder $(N[x^2] - [x]^2)b = [y][x^2] - [x][xy]$, d.h.

$$b = \frac{[y][x^2] - [x][xy]}{N[x^2] - [x]^2}\,.$$

(Man beachte, daß im Falle der eindeutigen Lösbarkeit sowohl $[x^2] > 0$ als auch $N[x^2] - [x]^2 > 0$ ist.)

Einsetzen von b in die erste Gleichung ergibt dann auch

$$a = \frac{[xy] - [x]b}{[x^2]}\,.$$

Im Falle des Zusammenhangs zwischen Korruptionsindex CPI und Bruttoinlandsprodukt pro Einwohner BIP erhalten wir nach diesen Formeln die Ausgleichsgerade

$$CPI = 2{,}917 + 0{,}000163\, BIP\,,$$

die Steigung ist also erwartungsgemäß positiv. Der relativ große konstante Term zeigt, daß *im Mittel* Korruption selbst bei sehr armen Ländern

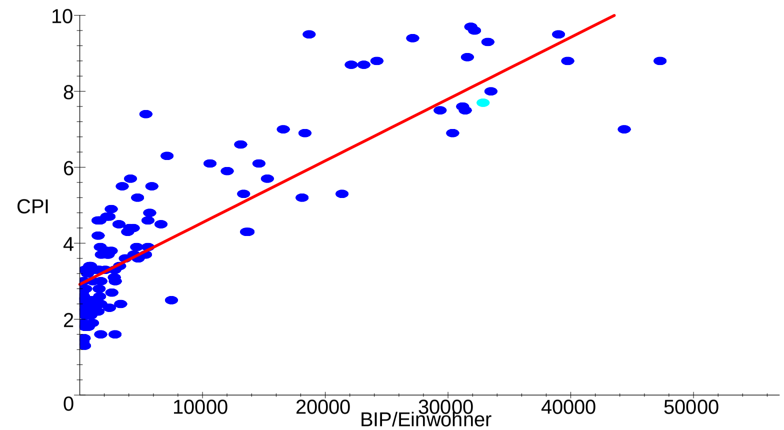

Abb. 19: Ausgleichsgerade zu Abbildung 18

deutlich über dem unteren Ende der Skala liegt. Abbildung 19 zeigt die Ausgleichsgerade zusammen mit den Daten.

Natürlich sind die Datenpunkte relativ breit gestreut um die Ausgleichsgerade; der Zusammenhang zwischen Korruption und Wohlstand ist schließlich zum Glück kein unausweichliches deterministisches Gesetz, sondern nur eine empirische Beobachtung.

Auch bei Messungen physikalischer Größen, wo die verschiedenen Meßgrößen meist durch wohlbekannte Naturgesetze miteinander verbunden sind, gibt es praktisch immer eine Streuung der Daten um die theoretisch richtige Meßkurve; absolut fehlerfreie Messungen sind, trotz aller Mühe der Experimentatoren, fast nie möglich, da es praktisch immer ein Grundrauschen der Meßgeräte und/oder nicht in ihrer Gesamtheit erfaßbare Umgebungseinflüsse *usw.* gibt. Vor allem bei Messungen, mit denen Konstanten für Naturgesetze ermittelt werden sollen oder gar ein Experiment zwischen zwei oder mehr Hypothesen entscheiden soll, ist es daher wichtig zu wissen, wie gut die Übereinstimmung zwischen den Daten und der berechneten Kurve (oder Fläche *usw.*) wirklich ist.

Solche Maße stellt die Statistik zur Verfügung; für ihr Verständnis sind daher meist zumindest Grundlagenkenntnisse der Statistik notwendig, wie wir sie (wenn auch nur kurz) im nächsten Semester behandeln werden. Im einfachsten und zugleich wichtigsten Fall eines linearen Zusammenhangs zwischen zwei Größen allerdings reicht die lineare Algebra, um das sowohl in der Theorie wie auch den Anwendungen wichtigste Qualitätsmaß zu definieren, den Korrelationskoeffizienten.

Angenommen, wir haben N Datenpaare (x_i, y_i), zwischen denen ein perfekter linearer Zusammenhang besteht, d.h.

$$y_i = ax_i + b \quad \text{für alle } i = 1, \ldots, N\,.$$

Wir wollen den Datenvektoren $\vec{x} \in \mathbb{R}^n$ mit Komponenten x_i und $\vec{y} \in \mathbb{R}^N$ mit Komponenten y_i Vektoren zuordnen, die nicht nur in einem linearen Zusammenhang stehen, sondern sogar gleich sind; mit anderen Worten, wir wollen die Parameter a und b aus obiger Gleichung eliminieren.

Dazu betrachten wir als erstes die Mittelwerte

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

Da $y_i = ax_i + b$ ist für alle i, folgt sofort, daß auch

$$\overline{y} = a\overline{x} + b$$

ist, und damit

$$(y_i - \overline{y}) = a(x_i - \overline{x}) \quad \text{für alle } i = 1, \ldots, N\,.$$

Damit haben wir den Parameter b eliminiert. Bezeichnen wir wieder mit $\vec{1} \in \mathbb{R}^N$ den Vektor, dessen sämtliche Komponenten Einser sind, ist nun also $(\vec{y} - \overline{y}\vec{1}) = a(\vec{x} - \overline{x}\vec{1})$. Aus dieser Gleichung können wir nun leicht a bis auf sein Vorzeichen eliminieren, indem wir die beiden Vektoren durch ihre Länge dividieren. Dies ist natürlich nur möglich, wenn keiner der beiden Vektoren gleich dem Nullvektor ist, wenn also nicht alle $x_i = \overline{x}$ oder alle $y_i = \overline{y}$ sind. Bei nicht getürkten Messungen ist dies allerdings *praktisch* nie der Fall, so daß die Nützlichkeit der folgenden Diskussion und Definition nicht darunter leidet, daß wir diesen Fall ausschließen müssen.

Falls also weder $\vec{y} - \overline{y}\vec{1}$ noch $\vec{x} - \overline{x}\vec{1}$ der Nullvektor ist, betrachten wir die beiden auf Länge eins normierten Vektoren

$$\frac{\vec{y} - \overline{y}\vec{1}}{\left|\vec{y} - \overline{y}\vec{1}\right|} \quad \text{und} \quad \frac{\vec{x} - \overline{x}\vec{1}}{\left|\vec{x} - \overline{x}\vec{1}\right|}\,.$$

Diese sind nun offensichtlich entweder gleich (für $a > 0$) oder entgegengesetzt gleich (für $a < 0$).

Wenn (wie in der Realität meist der Fall) *kein* perfekter linearer Zusammenhang zwischen den x_i und den y_i besteht, können wir trotzdem – falls weder $\vec{y} - \overline{y}\vec{1}$ noch $\vec{x} - \overline{x}\vec{1}$ der Nullvektor ist – die beiden Vektoren

$$\frac{\vec{y} - \overline{y}\vec{1}}{\left|\vec{y} - \overline{y}\vec{1}\right|} \quad \text{und} \quad \frac{\vec{x} - \overline{x}\vec{1}}{\left|\vec{x} - \overline{x}\vec{1}\right|}$$

betrachten. Da beides Einheitsvektoren sind, unterscheiden sie sich nur in der Richtung; als Maß für ihren Unterschied bietet sich daher den Winkel zwischen $\vec{x}$ und $\vec{y}$ an. Rechnerisch einfacher ist der Cosinus dieses Winkels, denn der ist bei Einheitsvektoren einfach gleich dem Skalarprodukt.

Definition: Der Korrelationskoeffizient zwischen zwei Datenvektoren $\vec{x}$ und $\vec{y} \in \mathbb{R}^n$, die keine Vielfachen des Vektors $\vec{1} \in \mathbb{R}^N$ sind, ist

$$\rho = \frac{(\vec{x} - \overline{x}\cdot\vec{1})\cdot(\vec{y} - \overline{y}\cdot\vec{1})}{\left|\vec{x} - \overline{x}\cdot\vec{1}\right|\cdot\left|\vec{y} - \overline{y}\cdot\vec{1}\right|}\,.$$

Damit ist also $\rho = \pm 1$ genau dann, wenn es einen perfekten linearen Zusammenhang $y_i = ax_i + b$ zwischen den beiden Größen gibt, mit $\rho = 1$ für $a > 0$ und $\rho = -1$ für $a < 0$. Ansonsten ist der Zusammenhang umso besser, je größer der Betrag von ρ ist. Für $\rho = 0$ stehen die beiden Vektoren $\vec{x} - \overline{x}\vec{1}$ und $\vec{y} - \overline{y}\vec{1}$ senkrecht aufeinander, d.h. wenn x_i größer ist als der Mittelwert $\overline{x}$, kann y_i im Mittel genauso gut größer wie auch kleiner als der Mittelwert $\overline{y}$ sein. (in der Statistik ist dies die *Definition* für die Unabhängigkeit von Daten.)

Definition: Zwei Größen x und y heißen $\left\{\begin{matrix}\text{positiv}\\ \text{negativ}\end{matrix}\right\}$ korreliert, wenn $\rho \left\{\begin{matrix}>\\<\end{matrix}\right\} 0$ ist. Sie heißen unkorreliert oder voneinander unabhängig, wenn $\rho = 0$ ist.

Im Beispiel der Korruption erhalten wir einen Korrelationskoeffizienten von

$$\rho \approx 0{,}8627227\,;$$

dies entspricht einem Winkel von etwa 30,376° zwischen den oben definierten Vektoren.

Um ein Gefühl für Korrelationskoeffizienten zu bekommen, wollen wir zwei Beispiele betrachten, die sich zumindest visuell sehr unterscheiden:

Der CPI für Deutschland hatte in den letzten Jahren folgende Werte:

Jahr:	1980–1985	1988–1992	1995	1996	1997	1998
CPI:	8,14	8,13	8,14	8,27	8,23	7,9
Jahr:	1999	2000	2001	2002	2003	
CPI:	8,0	7,6	7,4	7,3	7,7	

Wie Abbildung 20 zeigt, sieht der Zusammenhang zwischen Jahr und CPI nicht sonderlich linear aus: Der Bimbesknick ist unverkennbar, jedoch scheint die Talsohle inzwischen durchschritten. Der Korrelationskoeffizient $\kappa \approx -0{,}634$ sieht auf den ersten Blick recht gut aus; bedenkt man aber, daß dies der Kosinus eines Winkels von rund 130° ist, sieht man schon diesem Wert an, daß der Zusammenhang nicht sonderlich gut durch eine lineare Funktion beschrieben werden kann. In der Tat ist die eingezeichnete Ausgleichsgerade nicht einmal eine grobe qualitative Beschreibung des Zusammenhangs.

Vergleichen wir dagegen die Ergebnisse von Europawahl und Gemeinderatswahl am 13. Juni 2004 miteinander, so sind die Stimmenanteile der vier Parteien, die an beiden teilgenommen haben, zwar teilweise deutlich verschieden, der Zusammenhang ist aber, wie Abbildung 21 zeigt, doch in recht guter Näherung linear. In der Tat erhalten wir hier $\kappa \approx 0{,}9900614$, was einem Winkel von etwa acht Grad entspricht.

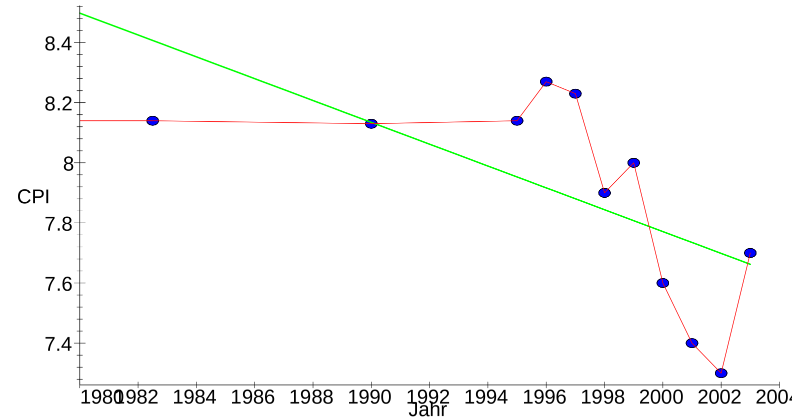

Abb. 20: Zeitabhängigkeit des CPI für Deutschland

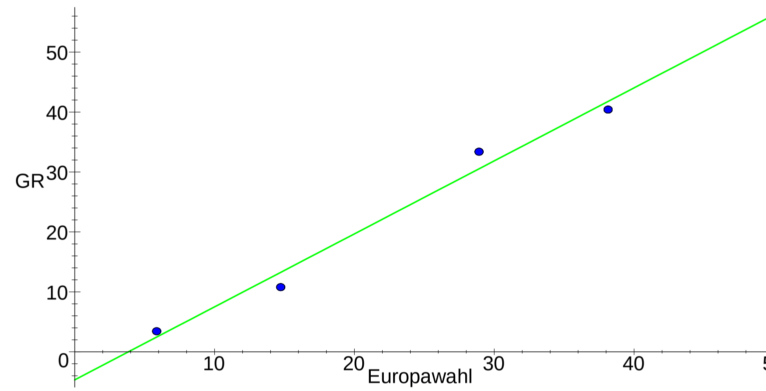

Abb. 21: Zusammenhang Europawahl/Gemeinderatswahl

	Europawahl	*Gemeinderatswahl*
CDU	38,14	40,41
SPD	28,91	33,38
Grüne	14,72	10,19
FDP	5,86	3,43