Google unterstützt diese Erwartung in zweierlei Weise: Zunächst gewichtet es das Vorkommen des Suchbegriffs an verschiedenen Stellen des Dokuments in verschiedener Weise: Zu Einzelheiten schweigt man sich dort natürlich aus (Eine gute Suchmaschine versucht, Webspam zu bekämpfen, nicht durch gute Tips zu unterstützen), aber es ist anzunehmen, daß beispielsweise das Vorkommen des Suchbegriffs im URL stärker berücksichtigt wird als die bloße Erwähnung gegen Ende des Dokuments, womöglich noch weiß auf weiß. Von manchen Suchmaschinen werden auch Dokumente, in denen die Suchbegriffe weit oben als Überschriften oder hervorgehoben erscheinen, stärker berücksicht als andere oder aber auch die (von Google völlig ignorierten) META-Tags.

Dies allein reicht allerdings nicht: Ein Webspammer könnte sich beispielsweise die domain `uni-mannheim.tv` sichern, um den BWL-Studenten des selbsternannten deutschen Harvards garantiert echte MBAs der richtigen Harvard Business School zu verkaufen. Eine gute Suchmaschine muß daher unterscheiden können zwischen `uni-mannheim.de` und `uni-mannheim.tv`.

Der wesentliche Unterschied zwischen den beiden ist, daß zumindest in Teilen des Subnetzes `uni-mannheim.de` wirkliche Inhalte angeboten werden und daß zumindest ein Teil dieser Seiten auch auf die Hauptseite `www.uni-mannheim.de` verweisen; außerdem gibt es externe Verweise auf zumindest einen Teil dieser Seiten. Die Spam-Domain `uni-mannheim.tv` kann zwar auch viele Seiten generieren, die auf sie verweisen, aber es ist sehr unwahrscheinlich, daß – abgesehen von anderen Spam-Domains – externe Links dorthin führen. Von den wirklich wichtigen Seiten wird wohl keine einen Link darauf haben.

Das wesentlich neue Element, durch das sich Google von seinen Vorgängern unterscheidet, ist die globale Anordnung aller erfaßten Seiten, unabhängig von irgendwelchen Suchbegriffen, nach dieser „Wichtigkeit“, die Google durch den sogenannten PageRank zu quantifizieren versucht. Das Wort PageRank geht dabei wohl nicht auf das englische Wort *page* für *Seite* zurück, sondern auf einen der beiden Gründer von Google, LAWRENCE PAGE.

LAWRENCE PAGE wurde 1973 in East Lansing, Michigan geboren; sein Vater war Informatikprofessor an der Michigan State University. Er selbst studierte an der University of Michigan in Ann Arbor technische Informatik, was er 1995 mit einem B.S.E. (Bachelor of Science in Engineering) abschloß. Zum anschließenden Promotionsstudium ging an die Stanford University, wo er SERGEY BRIN kennenlernte und gemeinsam mit ihm über Suchmaschinen arbeitete. 1998 gründeten sie, mit Hilfe der Universität, Google. PAGE ließ sich nach Erhalt seines Masters von seinem Promotionsstudium beurlauben und ist seither einer der beiden Präsidenten von Google.

SERGEY MICHAILOWITSCH BRIN (Сергей Михайлович Брин), der Mitbegründer und andere Präsident von Google, wurde 1973 in Moskau geboren als Sohn eines bei der Planungsbehörde *Gosplan* tätigen Mathematikers. Wegen der zahlreichen Repressalien gegen Juden während der Breschnew-Zeit verließ die Familie 1979 die Sowjetunion und siedelte um nach USA, wo der Vater an der University of Maryland Mathematik lehrte. Dort studierte auch der Sohn Mathematik und Informatik, bis er nach seinen Bachelor 1993 zum Promotionsstudium nach Stanford wechselte. Dort erhielt er 1995 seinen Master der Informatik. Seit Gründung von Google ist auch er von seinem Promotionsstudium beurlaubt.

Die Grundidee ist einfach: Eine Seite ist wichtig, wenn wichtige Seiten auf sie verweisen. Allerdings ist es natürlich ein Unterschied, ob sie in einem Linkverzeichnis mit Hunderten von Einträgen steht, oder ob sie einer von nur drei Links auf einer Seite ist. Eine erste naive Idee zur Definition der Wichtigkeit $w(x)$ einer Seite $x$ wäre also der folgende Ansatz: Sei $M_x$ die Menge aller Seiten $y$, die auf $x$ verweisen, und sei $n(y)$ jeweils die Anzahl der Links, die von einer Seite $y$ ausgehen. Dann ist

$$w(x) = \sum_{y \in M_x} \frac{w(y)}{n(y)} .$$

Jede Seite $y$ kann also nur den festen Betrag $w(y)$ an Wichtigkeit auf andere Seiten verteilen; je mehr solche Seiten es gibt, desto weniger bekommt jede einzelne davon ab.

Auf den ersten Blick beist sich diese Definition in den Schwanz: Um die Wichtigkeit einer Seite berechnen zu können, müssen wir zuvor die Wichtigkeit einer jeden Seite $y$ kennen, die auf $x$ verweist. Um deren Wichtigkeit zu berechnen, brauchen wir aber die Wichtigkeiten der Seiten, die auf $y$ verweisen, und darunter könnte durchaus auch $x$ sein, *usw.*

Das ist aber tatsächlich kein großes Problem: Angenommen, die Suchmaschine kennt insgesamt $N$ Seiten. (Im Augenblich ist $N$ für Google in der Größenordnung von etwa acht Milliarden, Tendenz steigend.) Dann ist das $N$-Tupel aller Wichtigkeiten $w(x)$ ein Vektor $\vec{v} \in \mathbb{R}^N$.

Dazu definieren wir eine $N \times N$-Matrix $A$, deren Eintrag an der Stelle $xy$ eine Null ist, falls die Seite $y$ keinen Link auf $x$ hat, und $1/n(x)$ sonst. Dann wird die obige Gleichung zu $\vec{v} = A\vec{v}$, d.h. $\vec{v}$ ist ein Eigenvektor von $A$ zum Eigenwert eins.

Als solcher ist er natürlich nicht eindeutig bestimmt: Selbst wenn der Eigenraum eindimensional ist, sind mit jedem Vektor $\vec{v}$ auch noch alle seine vom Nullvektor verschiedenen Vielfache Lösungen derselben Gleichung.

Solche Vielfache stellen jedoch kein großes Problem dar, denn selbstverständlich ist *Wichtigkeit* ein relativer Begriff; sobald man die Summe der Wichtigkeiten *aller* bekannter Seiten festlegt, ist klar, welches der Vielfachen als Einziges in Frage kommt. Man kann beispielsweise den Vektor so normieren, daß die Summe aller Einträge gleich Eins ist – das ist für theoretische Betrachtungen ganz nützlich und wird in der 1998 publizierten ersten Vorstellung von PageRank in

> LAWRENCE PAGE, SERGEY BRIN, RAJEEV MOTWANI, TERRY WINOGRAD: The PageRank Citation Ranking: Bringing Order to the Web, `http://dbpubs.stanford.edu/pub/1999-66`

auch so gemacht; für die Praxis hat es allerdings den Nachteil, daß dann fast alle Wichtigkeiten mit sehr vielen Nullen nach dem Komma beginnen. Praktikabler ist daher, die Summe beispielsweise festzulegen auf die Anzahl $N$ der Dokumente, die der Suchmaschine bekannt sind.



Als Problem könnte auch erscheinen, daß die Eins möglicherweise gar kein Eigenwert von $A$ ist. Falls keine Spalte von $A$ gleich dem Nullvektor ist, falls es also kein Dokument ohne Verweise gibt, kann das nicht passieren: In diesem Fall ist die Summe der Einträge einer jeden Spalte gleich eins, und für solche Matrizen gilt

**Lemma:** In der Matrix $A = (a_{ij}) \in \mathbb{R}^{N \times N}$ seien alle Einträge $a_{ij} \geq 0$ und $\sum_{i=1}^{N} a_{ij} = 1$ für alle $j$. Dann gilt:
*a)* Die Eins ist Eigenwert von $A$.
*b)* Jeder (reelle oder komplexe) Eigenwert von $A$ hat höchstens den Betrag eins.
*c)* Ist $\vec{v}$ Eigenvektor zu einem Eigenwert $\lambda \neq 1$, so ist die Summe aller Komponenten von $\vec{v}$ gleich null.

*Beweis: a)* Da die Summe der Einträge einer jeden Spalte von $A$ gleich eins ist, ist die entsprechende Summe für die Matrix $A - E$ gleich null, denn in jeder Spalte wird gegenüber $A$ noch in der Diagonale $-1$ subtrahiert. Damit ist die Summe aller Zeilenvektoren von $A - E$ gleich dem Nullvektor, d.h. die Zeilenvektoren von $A - E$ sind linear abhängig, und damit verschwindet $\det(A - E)$. Letzteres ist aber äquivalent dazu, daß die Eins Eigenwert von $A$ ist.

*b)* Ist $\vec{v}$ Eigenvektor zum Eigenwert $\lambda \in \mathbb{C}$, so ist $A\vec{v} = \lambda\vec{v}$, also

$$\sum_{j=1}^{N} a_{ij}\, v_j = \lambda v_i \,.$$

Daher ist

$$\begin{aligned} |\lambda| \sum_{i=1}^{N} |v_i| &= \sum_{i=1}^{N} |\lambda v_i| = \sum_{i=1}^{N} \left| \sum_{j=1}^{N} a_{ij}\, v_j \right| \\ &\leq \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}\, |v_j| = \sum_{j=1}^{N} \left( \sum_{i=1}^{N} a_{ij} \right) |v_j| = \sum_{j=1}^{N} |v_j| \,, \end{aligned}$$

was offensichtlich nur für $|\lambda| \leq 1$ möglich ist.

*c)* Die ist im wesentlichen die gleiche Rechnung wie bei *b),* nur ohne Betragsstriche: Mit Bezeichnungen wie eben ist

$$\lambda \sum_{i=1}^{N} v_i = \sum_{i=1}^{N} \lambda v_i = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}\, v_j = \sum_{j=1}^{N} \left( \sum_{i=1}^{N} a_{ij} \right) v_j = \sum_{j=1}^{N} v_j \,;$$

falls die Summe der Komponenten von $\vec{v}$ nicht verschwindet, können wir durch sie dividieren und erhalten $\lambda = 1$. Bei einem Eigenvektor zu einem Eigenwert ungleich eins muß daher die Summe der Komponenten verschwinden.

■

Nachdem wir nun wissen, daß die Eins stets Eigenwert von $A$ sein muß, fehlt nur noch ein zugehöriger Eigenvektor, d.h. eine Lösung des linearen Gleichungssystems $(A - E)\vec{v} = \vec{0}$. Für $N$ in der Größenordnung von mehreren Milliarden ist die Anwendung des GAUSS-Algorithmus allerdings nicht zu empfehlen, selbst wenn Google Supercomputer hätte statt seiner (etwa 10 000) billigen Linux-PCs.

Google verwendet daher ein anderes Verfahren: Zunächst erhält jede Seite die Wichtigkeit eins; man startet also mit einem Vektor $\vec{v}_0 \in \mathbb{R}^N$, dessen sämtliche Komponenten gleich eins sind. Dann berechnet man nacheinander für $i \geq 1$ die Vektoren $\vec{v}_i = A\vec{v}^{(i-1)}$, bis die Differenz zwischen $\vec{v}^{(i)}$ und $\vec{v}^{(i-1)}$ hinreichend nahe beim Nullvektor liegt. Die Multiplikation eines Vektors mit einer Matrix ist bei den Größenordnungen, mit denen wir es hier zu tun haben, zwar auch sehr aufwendig, aber da es im Internet viele Bereiche gibt, die kaum etwas miteinander zu tun haben (Mathematikseiten verweisen selten auf Internetapotheken und umgekehrt), ist in der Matrix $A$ nur ein sehr geringer Anteil der Einträge von Null verschieden. Natürlich werden nur diese Einträge gespeichert, und die Multiplikation kann auch leicht parallelisiert werden, indem man verschiedene Teilsummen auf verschiedene PCs verteilt.

Google verwendet allerdings nicht genau die Definition von $\vec{v}$, die wir bislang betrachtet haben, sondern eine Modifikation davon. Um einzusehen warum, überlegen wir uns zunächst, warum oder besser *wann* das angegebene Verfahren konvergiert.



Nehmen wir zunächst an, die Matrix $A$ sei (zumindest über $\mathbb{C}$) diagonalisierbar. Dann können wir sie bezüglich einer geeigneten Basis als Diagonalmatrix $D$ schreiben, wobei die Diagonaleinträge die Eigenwerte von $A$ sind. Das ist zwar mit heutiger Technologie nicht praktisch möglich, aber es ist ein Wesensmerkmal der Mathematik, daß sie oft auch aus Konstruktionen, die nur theoretisch möglich sind, praktisch relevante Folgerungen ziehen kann. Wir stellen auch die Vektoren $\vec{v}_0$ in der neuen Basis dar; Multiplikation mit $A$ bedeutet dann einfach, daß der $i$-te Eintrag mit dem $i$-ten Diagonaleintrag von $D$ multipliziert wird.

Mindestens einer dieser Diagonaleinträge ist die Eins; von den anderen wissen wir, daß ihre Beträge kleiner oder gleich eins sind. *Falls* alle diese Beträge echt kleiner als eins sind, gehen ihre Potenzen gegen null, so daß im Limes nur Komponenten von $\vec{v}$ übrigbleiben, die zu Eigenvektoren zum Eigenwert eins gehören; der Limes existiert also und ist ein Eigenvektor zum Eigenwert eins – es sei denn, in der Basisdarstellung des Ausgangsvektors $\vec{v}_0$ kommt kein Eigenvektor zum Eigenwert eins vor, was für einen zufällig gewählten Vektor $\vec{v}_0$ extrem unwahrscheinlich ist.

Zurückübersetzt in die Ausgangsbasis heißt dies, daß die Folge der Vektoren $A^n\vec{v}$ für jeden Vektor $\vec{v} \in \mathbb{R}^n$, der nicht im Erzeugnis der restlichen Eigenvektoren liegt, gegen einen Eigenvektor zum Eigenwert eins konvergiert.

Zur Herleitung dieses Ergebnisses haben wir angenommen, daß dies Matrix $A$ diagonalisiert werden kann; mit den Methoden, die wir im nächsten Semester kennenlernen werden, kann man aber zeigen, daß die Folge der $\vec{v}_i$ auch für nicht diagonalisierbare Matrizen $A$ gegen einen Eigenvektor zum Eigenwert eins konvergiert – falls alle anderen Eigenwerte einen Betrag echt kleiner eins haben.

Diese Annahme ist leider im allgemeinen nicht erfüllt: Das obige Lemma sagt uns nur, daß ihr Betrag kleiner oder gleich eins ist. Wie das folgende Beispiel zeigt, kommen Eigenwerte vom Betrag eins, aber ungleich eins, durchaus vor:

Wir betrachten vier Webseiten $u, v, w, x$, derart, daß $u$ ausschließlich auf $v$ verweist und $v$ ausschließlich auf $w$; auf $u$ verweise sowohl $w$ als

auch $x$, und auf $x$ schließlich keine Seite. Dann haben wir für diesen Teil des Webs die Matrix

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

Wie man leicht nachrechnet, ist

$$\det(A - \lambda E) = \lambda^4 - \lambda = \lambda(\lambda - 1)(\lambda^2 + \lambda + 1) ,$$

die Eigenwerte sind also 0, 1 und $-\frac{1}{2} \pm \frac{i}{2}\sqrt{3}$. Abgesehen von der Null haben sie alle den Betrag eins.

Berechnet man hier die Vektoren $\vec{v}_i = A\vec{v}_0$ mit dem Vektor $\vec{v}_0$ aus lauter Einsen, so gehen die $\vec{v}_i$ zyklisch hin und her zwischen den drei Vektoren

$$\vec{v}_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \end{pmatrix} , \quad \vec{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad \vec{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 0 \end{pmatrix} ;$$

es gibt also keinen Grenzwert.

Aus diesem Grund, und auch weil es Webseiten gibt, die keine Verweise auf andere Seiten haben (z.B. die Postskript- und pdf-Seiten dieses Skriptums), modifizierten PAGE und BRIN bereits in der oben zitierten Arbeit den naiven Ansatz zur Definition der Wichtigkeit einer Seite:

In der zitierten Arbeit gehen sie zunächst aus von Anfangswichtigkeiten $E(x)$, die zu den oben definierten Wichtigkeiten addiert werden:

$$w(x) = c \left( \sum_{y \in M_x} \frac{w(y)}{n(y)} + E(x) \right) ,$$

wobei die Konstante $c$ so gewählt wird, daß die Summe aller Wichtigkeiten konstant bleibt. In ihrer zweiten Arbeit

> SERGEY BRIN, LAWRENCE PAGE: The Anatomy of a Large-Scale Hypertextual Web Search Engine,
> http://www-db.stanford.edu/pub/papers/google.pdf



setzen sie $E(x)$ auf eine von $x$ unabhängige Konstante:

$$w(x) = d \sum_{y \in M_x} \frac{w(y)}{n(y)} + (1 - d) ,$$

wobei sie für die Konstante $d$ die Zahl 0,85 als geeignet erwähnen. Wie Google heute tatsächlich arbeitet, ist natürlich nicht bekannt. Hier wird also ein Vektor $\vec{v}$ gesucht, für dessen Komponenten $v_i$ gilt

$$v_i = d \sum_{j=1}^{N} a_{ij}\, v_j + (1 - d) \quad \text{oder} \quad \vec{v} = dA\vec{v} + (1 - d)\vec{1} ,$$

wobei $A = (a_{ij})$ die oben definierte Matrix ist und $\vec{1} \in \mathbb{R}^N$ der Vektor mit lauter Einsen als Komponenten.

Auch diese Modifikation läßt wieder als Eigenwertproblem formulieren, allerdings nur, wenn man sich auf Vektoren beschränkt, für die die Summe der Komponenten gleich einer festen Zahl ist, z.B. gleich $N$: Bezeichnet $\mathbb{E}$ die $N \times N$-Matrix mit lauter Einsen als Einträgen, so muß gelten

$$\vec{v} = \left( dA + \frac{1 - d}{N} \mathbb{E} \right) \vec{v} ,$$

denn $\mathbb{E}\vec{v}$ ist die Summe aller Einträge von $\vec{v}$, also $N$. Somit ist $\vec{v}$ ein Eigenvektor zum Eigenwert eins von $M = dA + \frac{1-d}{N}\mathbb{E}$.

Falls in der Matrix $A$ alle Spaltensummen gleich eins sind, gilt dasselbe auch für $M$, wir wissen also nach obigem Lemma, daß es mindestens einen Eigenvektor $\vec{v}$ zum Eigenwert eins gibt. Falls dieser Vektor Komponentensumme $N$ hat, löst er auch die Gleichung $\vec{v} = dA\vec{v} + (1 - d)\vec{1}$ und ist somit ein Kandidat für einen PageRank.

Falls die Summe der Komponenten einen anderen von null verschiedenen Wert hat, liefert Multiplikation mit einer geeigneten Konstanten einen Eigenvektor mit Summe $N$; Probleme gibt es also nur, falls die Summe der Komponenten von $\vec{v}$ verschwindet. Dann ist aber $\mathbb{E}\vec{v} = 0$, also $\vec{v} = M\vec{v} = dA\vec{v}$ oder $A\vec{v} = \frac{1}{d}\vec{v}$. Somit ist $\vec{v}$ ein Eigenvektor von $A$ zum Eigenwert $1/d$. Für $d < 1$ ist dies aber nicht möglich, denn alle Eigenwerte von $A$ haben Betrag kleiner oder gleich eins.

Somit gibt es auch für das modifizierte Problem stets mindestens eine Lösung. Tatsächlich gilt für $d < 1$ sogar:

**Lemma:** *a)* Für $0 < d < 1$ gibt es genau einen Vektor $\vec{v} \in \mathbb{R}^N$, für den gilt

$$\vec{v} = dA\vec{v} + (1-d)\vec{1}\,.$$

*b)* Ist $\vec{v}^{(0)}$ irgendein Vektor aus $\mathbb{R}^N$, so konvergiert die für $i \geq 1$ durch

$$\vec{v}^{(i)} = dA\vec{v}^{(i-1)} + (1-d)\vec{1}$$

definierte Folge gegen $\vec{v}$.

Zum *Beweis* definieren wir zunächst für jeden Vektor $\vec{v} \in \mathbb{R}^n$ seine sogenannte $\mathrm{L}^1$-Norm

$$\|\vec{v}\|_1 \underset{\text{def}}{=} \sum_{i=1}^{N} |v_i|$$

als Summe der Beträge seiner Komponenten. Sie ist offensichtlich genau dann gleich null, wenn $\vec{v}$ der Nullvektor ist.

Für zwei Vektoren $\vec{v}, \vec{w} \in \mathbb{R}^N$ seien nun

$$\vec{v}^* = dA\vec{v} + (1-d)\vec{1} \quad \text{und} \quad \vec{w}^* = dA\vec{w} + (1-d)\vec{1}\,.$$

Dann ist

$$\begin{aligned}
\left\|\vec{v}^* - \vec{w}^*\right\|_1 &= \sum_{i=1}^{N} \left|v_i^* - w_i^*\right| = \sum_{i=1}^{N} \left| d \sum_{j=1}^{N} a_{ij}(v_j - w_j) \right| \\
&\leq \sum_{i=1}^{N} d \sum_{j=1}^{N} a_{ij} \left|v_j - w_j\right| = d \sum_{j=1}^{N} \left( \sum_{i=1}^{N} a_{ij} \right) \left|v_j - w_j\right| \\
&= d \sum_{j=1}^{N} \left|v_j - w_j\right| = d \left\|\vec{v} - \vec{w}\right\|_1 \,.
\end{aligned}$$

Mit dieser Formel können wir nun zunächst *a)* beweisen: Wir wissen bereits, daß es *mindestens* einen Vektor $\vec{v} \in \mathbb{R}^N$ gibt mit der Eigenschaft $\vec{v} = dA\vec{v} + (1-d)\vec{1}$; was noch fehlt, ist, daß es nicht mehr als einen



geben kann. Sind $\vec{v}$ und $\vec{w}$ zwei solche Vektoren, so ist mit obigen Bezeichnungen $\vec{v}^* = \vec{v}$ und $\vec{w}^* = \vec{w}$, also

$$\|\vec{v} - \vec{w}\|_1 = \left\|\vec{v}^* - \vec{w}^*\right\|_1 \leq d\,\|\vec{v} - \vec{w}\|_1 \,.$$

Für $0 \leq d < 1$ ist dies offensichtlich nur dann möglich, wenn $\|\vec{v} - \vec{w}\|_1 = 0$, also $\vec{v} = \vec{w}$ ist.

Zum Beweis von *b)* setzen wir in der obigen Formel $\vec{v}$ gleich dem eindeutig bestimmten Lösungsvektor und $\vec{w} = \vec{v}_i$ für irgendein $i \geq 0$. Dann ist $\vec{v}^* = \vec{v}$ und $\vec{w}^* = \vec{v}_{i+1}$, also $\|\vec{v}_{i+1} - \vec{v}\|_1 \leq d\,\|\vec{v}_i - \vec{v}\|_1$. Da $d < 1$ vorausgesetzt war, zeigt dies, daß die Folge der $\vec{v}_i$ unabhängig vom Startwert gegen $\vec{v}$ konvergiert.

■

Damit ist ein praktikables Verfahren gefunden, um die „Wichtigkeit" einer Webseite zu definieren: Man stellt zunächst die Matrix $A$ der Verweise auf, wobei in diesem Stadium alle Seiten ohne ausgehende Links unberücksichtigt bleiben. Dann beginnt man mit irgendeinem Vektor $\vec{v}^{(0)}$ und berechnet so lange $\vec{v}^{(i+1)} = dA\vec{v}^{(i)} + (1-d)$, bis das Ergebnis einigermaßen stabil ist. Das Ergebnis ist der Vektor der Wichtigkeiten, der nun noch dazu benutzt werden kann, um auch die Ränge der Seiten ohne ausgehende Linkszu bestimmen.

Google publiziert wie üblich keine Einzelheiten, aber in unabhängigen Schätzungen ist davon die Rede, daß etwa zwanzig bis dreißig Iterationen nötig sind und daß Google mehrere Tage braucht, um diese durchzuführen.

Weitere Informationen über die Mathematik hinter Google findet man beispielsweise bei

> AMY N. LANGVILLE, CARL D. MEYER: Google's PageRank and Beyond: The Science of Search Engine Rankings, *Princeton University Press,* 2006.

Die Werte für die Wichtigkeiten können beträchtlich schwanken: der minimale Wert ist offensichtlich gleich $1-d$, also für $d = 0{,}85$ gleich $0{,}15$; die theoretische Obergrenze liegt bei $dN$, also im Milliardenbereich. Google zerteilt diesen Bereich in elf Teilintervalle, denen die PageRanks null bis zehn zugeordnet werden. Über die Definition dieser

Intervalle ist nichts bekannt, allerdings wird vermutet, daß die Intervallängen ungefähr in einer geometrischen Progression ansteigen, so daß der PageRank ungefähr gleich einem Logarithmus der gerade bestimmten Wichtigkeit ist, dessen Basis in der Gegend von sechs oder sieben liegen dürfte ($6^{10} = 60\,466\,176$ und $7^{10} = 282\,475\,249$).

Die *home page* dieser Vorlesung, beispielsweise, hat natürlich PageRank null, da außer meiner *home page* nichts darauf verweist. Auf meine *home page* jedoch verweisen fast alle meine Seiten, so daß diese auf einen PageRank von vier kommt. Das Institut für Mathematik, auf das wohl alle hiesigen Mathematiker verweisen, hat den nächsthöheren Rang fünf, die Fakultät für Mathematik und Informatik sechs, die Wirtschaftshochschule Mannheim sieben. Universitäten wie Karlsruhe, Heidelberg, Stuttgart oder Bielefeld kommen auf acht, Eliteuniversitäten wie Harvard oder das MIT auf neun oder gar zehn wie Stanford, die Heimatuniversität der beiden Google-Gründer.

Auch Google selbst hat PageRank zehn, `amazon.com` und Microsoft immerhin noch neun, `amazon.de` und Der Spiegel acht.

## k) Die Cramersche Regel

Determinanten können auch angewandt werden, um die Lösungen eines linearen Gleichungssystems vom Rang $n$ aus $n$ Gleichungen in $n$ Unbekannten in geschlossener Form als Funktion der Koeffizienten darzustellen. Dazu schreiben wir das Gleichungssystem $A\vec{x} = \vec{b}$ in der Form $x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \vec{b}$, wobei die $\vec{a}_i$ die Spaltenvektoren der Matrix $A$ seien und $(x_1, \ldots, x_n)$ eine Lösung des linearen Gleichungssystems.

Nun ersetzen wir in $\det A = \det\bigl(\vec{a}_1, \ldots, \vec{a}_n\bigr)$ rechts den Vektor $\vec{a}_i$ durch die rechte Seite $\vec{b}$ des Gleichungssystems. (Es gibt eigentlich keinen vernünftigen Grund, warum wir das tun sollten; auch dieser Trick wird, wie so viele, erst nachträglich durch das Ergebnis gerechtfertigt.) Die so



entstehende Determinante ist

$$\begin{aligned}
&\det\bigl(\vec{a}_1, \ldots, \vec{b}, \ldots, \vec{a}_n\bigr) \\
&= \det\Bigl(\vec{a}_1, \ldots, \sum_{j=1}^{n} x_j\vec{a}_j, \ldots, \vec{a}_n\Bigr) \\
&= \sum_{j=1}^{n} x_j \det\bigl(\vec{a}_1, \ldots, \vec{a}_j, \ldots, \vec{a}_n\bigr) \\
&= x_i \det\bigl(\vec{a}_1, \ldots, \vec{a}_i, \ldots, \vec{a}_n\bigr) = x_i \det A\,,
\end{aligned}$$

da für jeden Index $j \neq i$ der Vektor $\vec{a}_j$ zweimal als Argument der Determinante auftritt, so daß alle Summanden bis auf den $i$-ten verschwinden.

Falls $\det A = 0$ ist, nützt uns diese Formel überhaupt nichts; ist allerdings $\det A \neq 0$, wissen wir bereits, daß das Gleichungssystem eindeutig lösbar ist, und wir können die Komponenten dieser eindeutig bestimmten Lösung explizit ausdrücken durch die

CRAMERsche Regel: $$x_i = \frac{\det\bigl(\vec{a}_1, \ldots, \vec{b}, \ldots, \vec{a}_n\bigr)}{\det A}$$

Der Schweizer Mathematiker GABRIEL CRAMER (1704–1752) lehrte an der Universität Genf. Bekannt wurde er vor allem durch seine Arbeiten über Determinanten, er beschäftigte sich aber auch viel mit Analysis und Geometrie, insbesondere ist er Autor eines Buchs über algebraische Kurven. Weitere Arbeitsgebiete sind mathematische Methoden der Physik sowie die Geschichte der Mathematik. Die CRAMERsche Regel im Spezialfall $n = 2$ ist bereits 1545 in der *Ars magna* des italienischen Mathematikers GIROLAMO CARDANO (1501–1576) zu finden.

Die CRAMERsche Regel ist sicherlich kein Verfahren, das man oft anwendet zur Lösung eines einzelnen Gleichungssystems: Abgesehen von einigen Fällen mit sehr spärlich besetzter Matrix ist der Aufwand für die Berechnung von $n + 1$ Determinanten weit größer als eine Lösung nach dem GAUSS-Algorithmus. Falls man allerdings eine ganze Familie ähnlicher Gleichungssysteme hat, in der sich nur wenige Parameter

ändern und bei denen die Determinanten auf Grund einer speziellen Form des Gleichungssystems gut berechenbar sind, kann es sich lohnen, mit Hilfe der CRAMERschen Regel eine (von den Parametern abhängige) Lösungsformel zu berechnen und dann diese anzuwenden.

### l) Geschichte und Anwendungen von Determinanten

Determinanten erblickten das Licht der Welt im Jahr 1683, und zwar gleich zweimal: Der japanische Mathematiker SEKI benutzte sie, ohne Ihnen einen Namen zu geben, zur Lösung von Gleichungen höheren Grades und zeigte anhand von Beispielen, wie man sie für $2 \times 2$- bis $5 \times 5$-Matrizen berechnet. Im gleichen Jahr schrieb auch LEIBNIZ einen Brief an DE L'HÔPITAL, in dem er erwähnte, daß ein gewisses homogenes lineares Gleichungssystem in drei Variablen nichttrivial lösbar sei, da (in heutiger Terminologie) seine Determinante verschwinde.

TAKAKAZU SEKI KOWA (1642–1708) war Sohn eines Samurai, wurde aber schon sehr jung von einem Adeligen namens SEKI GOROZAYEMON adoptiert. Einer von dessen Dienern weckte das Interesse des neunjährigen SEKI an der Mathematik, woraufhin dieser eine große Bibliothek japanischer und chinesischer Mathematikbücher anschaffte, anhand derer er sich selbst in das Gebiet einarbeitete. Als Staatsbeamter und ab 1704 Zeremonienmeister des Shogun befaßte er sich weiterhin viel mit Mathematik und entdeckte außer Determinanten beispielsweise auch das NEWTON-Verfahren und (vor JAKOB BERNOULLI) die BERNOULLI-Zahlen.

LEIBNIZ sprach noch nicht von Determinanten, sondern von *Resultanten;* ein Begriff, den unabhängig davon 1772 auch LAPLACE benutzte, als er damit die Bahnen der inneren Planeten berechnete. Das *Wort* Determinante erschien erstmal 1801 in den *Disquisitiones arithmeticae* von GAUSS, der damit die Eigenschaften quadratischer Formen untersuchte. Auch CAUCHY, der 1812 den Multiplikationssatz bewies, sprach von Determinanten.

Heute bezeichnet man als *Resultanten* spezielle Determinaten, die zur Lösung nichtlinearer Gleichungssysteme benutzt werden und auf den englischen Mathematiker JAMES JOSEPH SYLVESTER (1814–1897) zurückgehen. Mit Hilfe solcher Resultanten gelang es beispielsweise 1985 zwei Mathematikern bei *General Motors* das inverse kinematische Problem für einen Roboterarm mit sechs Freiheitsgraden zu lösen, d.h. also ein Verfahren zu entwickeln, mit dem der Manipulator am Ende des Arms *automatisch* auf eine vorgegebene Position und Ausrichtung gebracht werden kann.



Weitere wichtige Anwendungen haben Determinanten auch in der Numerik, wo sie sowohl *Konditionszahlen* definieren, die etwas über die Stabilität und Robustheit eines Verfahrens aussagen, als auch beispielsweise (wie die VANDERMONDEsche Determinante) bei der Approximation von Funktionen oder Datenpunkten durch Polynomfunktionen verwendet werden. Ebenfalls wichtig sind sie für Volumenberechnungen; dieser Aspekt wird uns im nächsten Kapitel begegnen, wenn wir Mehrfachintegrale von einem Koordinatensystem in ein anderes transformieren.

## § 5: Euklidische und Hermitesche Vektorräume

In §2*a)* hatten wir, zur Definition von Vektoren im $\mathbb{R}^3$, Pfeile betrachtet und vereinbart, daß zwei Pfeile genau dann den gleichen Vektor darstellen sollen, wenn sie dieselbe Länge und (so diese Länge von Null verschieden ist) dieselbe Richtung haben. In der anschließenden Definition des Vektorraums allerdings kamen Längen und Richtungen nicht mehr vor – aus gutem Grund, denn es fällt in der Tat schwer, sich etwas vorzustellen unter der Länge eines Vektors über dem Körper $\mathbb{F}_{256}$. Über den reellen (und, mit Modifikationen) den komplexen Zahlen aber führen Längen und Richtungen zu interessanten Strukturen, deren Bedeutung weit über die Geometrie hinausgeht: Beispielsweise lassen sich Energie oder Leistung eines Signals oft interpretieren als eine Art Länge in einem unendlichdimensionalen Vektorraum; außerdem spielen solche verallgemeinerte Längen und Richtungen eine wichtige Rolle in der Fehler- und Ausgleichsrechnung sowie in der Statistik. Seit einiger Zeit werden werden Winkel auch bei der Informationssuche angewandt: Einige Suchmaschine etwa berechnen bei der Suche nach den besten Dokumenten zu einer Anfrage unter anderem Winkel zwischen einem

Dokumentenvektor und einem Anfragevektor, wobei beide in einem reellen Vektorraum liegen, dessen Dimension einige Millionen oder gar Milliarden betragen kann.

## a) Längen und Winkel in $\mathbb{R}^2$ und $\mathbb{R}^3$

Beginnen wir mit leichter vorstellbaren Längen und Winkeln.

Die Länge eines Vektors $\vec{v}$ im $\mathbb{R}^2$ läßt sich leicht nach dem Satz des PYTHAGORAS berechnen: Wir nehmen die beiden Koordinateneinheitsvektoren als Basis und schreiben bezüglich dieser Basis

$$\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ b \end{pmatrix} .$$

Wenn wir den Koordinateneinheitsvektoren, ihrem Namen entsprechend, die Länge eins zuordnen, haben die Vektoren auf der rechten Seite offensichtlich die Längen $a$ und $b$. Außerdem bilden diese Vektoren zusammen mit dem Vektor $\vec{v}$ ein rechtwinkliges Dreieck; nach PYTHAGORAS erfüllt die Länge $c$ von $\vec{v}$ daher die Gleichung

$$c^2 = a^2 + b^2 \quad \text{oder} \quad c = \sqrt{a^2 + b^2} .$$

(Hier sieht man einen der Gründe, warum wir über beliebigen Körpern nicht von Längen sprechen: In Körpern wie $\mathbb{Q}$ existieren Quadratwurzeln nur ausnahmsweise, und in Körpern wie $\mathbb{C}$, in denen sie immer existieren, sind sie nicht eindeutig, da mit $c$ stets auch $-c$ eine Wurzel aus $c^2$ ist. (Lediglich in Körpern, in denen (wie in $\mathbb{F}_{2^n}$) jedes Element gleich seinem Negativen ist, ist die Wurzel eindeutig.) In $\mathbb{R}$ gibt es zwar auch zwei Quadratwurzeln, aber da $\mathbb{R}$ im Gegensatz etwa zu $\mathbb{C}$ ein angeordneter Körper ist, können wir in konsistenter Weise eine der beiden auszeichnen, nämlich die nichtnegative.)

Für Vektoren im $\mathbb{R}^3$ müssen wir den Satz des PYTHAGORAS zweimal anwenden: Wir schreiben

$$\vec{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ b \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix} = \vec{u} + \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}$$



mit

$$\vec{u} = \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ b \\ 0 \end{pmatrix} .$$

Dann bildet $\vec{u}$ zusammen mit den ersten beiden Vielfachen von Einheitsvektoren ein rechtwinkliges Dreieck, seine Länge $d$ erfüllt also die Gleichung $d^2 = a^2 + b^2$. Da der Vektor $\vec{u}$ in der $x, y$-Ebene liegt, auf der der Vektor mit Komponenten $0, 0, c$ senkrecht steht, bilden auch $\vec{u}, \vec{v}$ und dieser Vektor ein rechtwinkliges Dreieck, so daß für die Länge $e$ von $\vec{v}$ gilt

$$e^2 = d^2 + c^2 = a^2 + b^2 + c^2 \quad \text{oder} \quad e = \sqrt{a^2 + b^2 + c^2} .$$

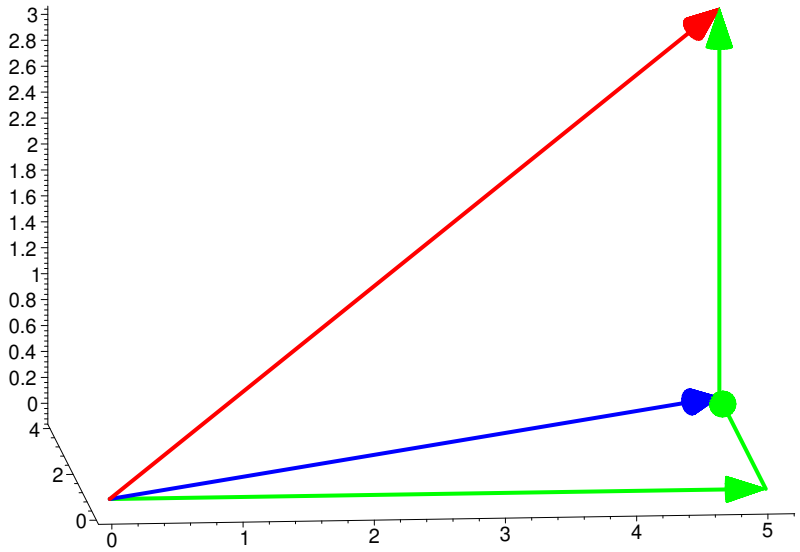



Abb. 13: Längenberechnung im $\mathbb{R}^3$

Längen sind also problemlos berechenbar.

Zur Berechnung von Winkeln verwenden wir das (den meisten wohl bereits aus der Schule bekannte) Skalarprodukt. Es hat seinen Namen

daher, daß es zwei Vektoren zu einem *Skalar* verknüpft; dieser wird mit $\vec{v} \cdot \vec{w}$ oder auch kurz $\vec{v}\vec{w}$ bezeichnet. Nach Definition ist

$$\vec{v} \cdot \vec{w} \underset{\text{def}}{=} |\vec{v}|\,|\vec{w}| \cos \measuredangle(\vec{v}, \vec{w}) \,,$$

wobei Betragsstriche die Länge eines Vektors bezeichnen sollen und $\measuredangle(\vec{v}, \vec{w})$ für den Winkel zwischen $\vec{v}$ und $\vec{w}$ steht. Da

$$\cos(2\pi - \varphi) = \cos(-\varphi) = \cos\varphi$$

ist, folgt sofort das *Kommutativgesetz*

$$\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$$

für das Skalarprodukt; außerdem folgt wegen $\cos 0 = 1$, daß

$$\vec{v} \cdot \vec{v} = |\vec{v}|^2 \quad \text{oder} \quad |\vec{v}| = \sqrt{\vec{v} \cdot \vec{v}}$$

ist. Das Skalarprodukt erlaubt also auch die Berechnung von Längen.

Weiter ist $\cos(\pi/2) = \cos(3\pi/2) = 0$ und damit $\vec{v} \cdot \vec{w} = 0$, falls $\vec{v}$ und $\vec{w}$ aufeinander senkrecht stehen; falls keiner der beiden Vektoren der Nullvektor ist, stehen $\vec{v}$ und $\vec{w}$ *genau dann* senkrecht aufeinander, wenn $\vec{v} \cdot \vec{w} = 0$ ist.

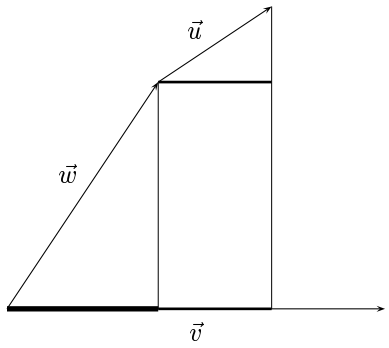



Abb. 14: Geometrische Interpretation des Skalarprodukts



Das Skalarprodukt kann auch geometrisch interpretiert werden: Da der Cosinus eines Winkels gleich Ankathete durch Hypothenuse ist, zeigt Abbildung 14, daß $|\vec{w}| \cos \measuredangle(\vec{v}, \vec{w})$ gerade die Länge des senkrecht auf die von $\vec{v}$ erzeugte Gerade projizierten Vektors $\vec{w}$ ist; in der Zeichnung ist dies die fett eingezeichnete Strecke.

Addieren wir zu $\vec{w}$ einen weiteren Vektor $\vec{u}$, so ist auch hier wieder $|\vec{u}| \cos \measuredangle(\vec{v}, \vec{u})$ die (in Abbildung 14 halbfett eingezeichnete) Länge des auf die von $\vec{v}$ erzeugte Gerade projizierten Vektors $\vec{u}$ und entsprechendes gilt für $\vec{w} + \vec{u}$; wir erhalten somit die Regel

$$\vec{v} \cdot (\vec{w} + \vec{u}) = \vec{v} \cdot \vec{w} + \vec{v} \cdot \vec{u} \,,$$

und damit wegen des Kommutativgesetzes auch

$$(\vec{v} + \vec{w}) \cdot \vec{u} = \vec{v} \cdot \vec{u} + \vec{w} \cdot \vec{u} \,.$$

Tatsächlich ist das Skalarprodukt sogar linear in beiden Argumenten, denn für positive Werte von $\lambda$ ist natürlich $(\lambda\vec{v}) \cdot \vec{w} = \lambda(\vec{v} \cdot \vec{w})$, da sich am Winkel nichts ändert und die Länge von $\vec{v}$ mit $\lambda$ multipliziert wurde. Bei negativem $\lambda$ wird der Winkel durch seinen Komplementärwinkel ersetzt, wobei der Kosinus sein Vorzeichen wechselt, und die Länge von $\vec{v}$ wird mit $|\lambda| = -\lambda$ multipliziert, so daß insgesamt wieder ein Faktor $\lambda$ vor dem Skalarprodukt steht.

Zum Berechnen des Skalarprodukts ist die bisherige Definition für viele Fälle recht unhandlich, da Vektoren oft in einer solchen Weise gegeben sind, daß man den Winkel zwischen ihnen *nicht* ohne weiteres kennt. Um gut rechnen zu können, wählen wir eine Basis aus drei paarweise aufeinander senkrecht stehenden Vektoren $\vec{e}_1, \vec{e}_2, \vec{e}_3$ der Länge eins. Für diese Vektoren läßt sich das Skalarprodukt leicht ausrechnen: Da zwei verschiedene stets aufeinander senkrecht stehen und jeder einzelne die Länge eins hat, ist

$$\vec{e}_i \cdot \vec{e}_j = \delta_{ij} = \begin{cases} 0 & \text{falls } i \neq j \\ 1 & \text{falls } i = j \end{cases} \,.$$

($\delta_{ij}$ ist das aus §2 bekannte KRONECKER-$\delta$)

Für zwei Vektoren

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad \text{und} \quad \vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

ist dann

$$\vec{v} \cdot \vec{w} = (v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3) \cdot (w_1\vec{e}_1 + w_2\vec{e}_2 + w_3\vec{e}_3)\,,$$

und nach obigen Rechenregeln läßt sich dies berechnen als

$$\vec{v} \cdot \vec{w} = \sum_{i=1}^{3}\sum_{j=1}^{3} v_i w_j\, \vec{e}_i\vec{e}_j = \sum_{i=1}^{3}\sum_{j=1}^{3} v_i w_j \delta_{ij} = \sum_{i=1}^{3} v_i w_i\,.$$

In dieser Form werden Skalarprodukte meistens ausgerechnet, und wir können aus dem Ergebnis rückwärts den Winkel zwischen den beiden Faktoren bestimmen, denn nach der ursprünglichen Definition des Skalarprodukt ist

$$\cos \measuredangle(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| \cdot |\vec{w}|} = \frac{\vec{v} \cdot \vec{w}}{\sqrt{(\vec{v} \cdot \vec{v})(\vec{w} \cdot \vec{w})}}\,.$$

Da wir nur den Cosinus des Winkels kennen, ist dieser nur bis auf Vielfache von $180^\circ$ im Gradmaß *bzw.* $\pi$ im Bogenmaß bestimmt, aber mehr ist im Dreidimensionalen ohnehin nicht sinnvoll: Für Vektoren in der Ebene unterscheiden wir zwei entgegengesetzt gleiche Winkel $\alpha$ und $-\alpha$ dadurch, daß wir den Gegenuhrzeigersinn als mathematisch positiv auszeichnen. Im $\mathbb{R}^3$ aber können wir die Uhr auf zwei Arten in die von zwei Vektoren aufgespannte Ebene legen: mit dem Ziffernblatt nach „oben“ oder nach „unten“, wobei wir diese Begriffe nicht konsistent unterscheiden können. Damit entfällt die Unterscheidung zwischen $\alpha$ und $-\alpha$.

## b) Euklidische Vektorräume

Im letzten Abschnitt haben wir gesehen, daß das Skalarprodukt im $\mathbb{R}^3$ eng mit Längen und Winkeln zusammenhängt; da dies zwei der Grundbegriffe der EUKLIDischen Geometrie sind, werden wir reelle Vektorräume mit Skalarprodukt allgemein als EUKLIDische Vektorräume bezeichnen.

Wir beginnen mit einem etwas schwächeren Begriff als dem des Skalarprodukts, der sich im Gegensatz zu letzterem noch für Vektorräume über beliebigen Körpern definieren läßt:



**Definition:** Eine *symmetrische Bilinearform* auf dem $k$-Vektorraum $V$ ist eine Abbildung $\cdot\colon V \times V \to k$ mit folgenden Eigenschaften:
*a)* $\cdot$ ist *bilinear,* d.h.

$$(\lambda\vec{v}_1 + \mu\vec{v}_2) \cdot \vec{w} = \lambda(\vec{v}_1 \cdot \vec{w}) + \mu(\vec{v}_2 \cdot \vec{w}) \qquad \text{und}$$
$$\vec{v} \cdot (\lambda\vec{w}_1 + \mu\vec{w}_2) = \lambda(\vec{v} \cdot \vec{w}_1) + \mu(\vec{v} \cdot \vec{w}_2)$$

für alle $\vec{v}_1, \vec{v}_2, \vec{w}, \vec{w}_1, \vec{w}_2 \in V$ und $\lambda, \mu \in k$.
*b)* $\cdot$ ist *symmetrisch,* d.h. $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$ für alle $\vec{v}, \vec{w} \in V$.

Die Skalarprodukte in $\mathbb{R}^2$ und $\mathbb{R}^3$ sind natürlich symmetrische Bilinearformen im Sinne dieser Definition; allgemeiner wird für jeden $\mathbb{R}^n$ durch

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v_1 w_1 + \cdots + v_n w_n$$

eine Bilinearform definiert.

Noch allgemeiner erklärt diese Formel auch für die Vektorräume $k^n$ über einem beliebigen Körper $k$ eine symmetrische Bilinearform, die auch für (nach unserem derzeitigen Kenntnisstand) eher exotische Körper durchaus nützliche Anwendungen haben kann: Ist etwa $\mathbb{F}_2 = \{0, 1\}$ der Körper mit zwei Elementen (in dem Addition und Multiplikation modulo zwei definiert sind), so ist das Produkt eines Vektors $\vec{v} \in \mathbb{F}_2^n$ mit dem Vektor, dessen sämtliche Komponenten gleich eins sind, genau dann gleich null, wenn die Anzahl der Einsen im Vektor $\vec{v}$ gerade ist; ansonsten ist es null. Auf diese Weise läßt sich also eine Paritätsprüfung einfach und kompakt formulieren. Wenn man außer dem Vektor mit lauter Einsen als Komponenten noch weitere geeignete Vektoren wählt, lassen sich nicht nur Paritätsfehler, sondern auch noch andere Fehler erkennen und eventuell sogar korrigieren.

Verglichen mit dem Skalarprodukt, wie wir es aus $\mathbb{R}^2$ und $\mathbb{R}^3$ gewohnt sind, fehlt diesen Bilinearformen jedoch eine wesentliche Eigenschaft: In $\mathbb{R}^2$ und $\mathbb{R}^3$ ist das Skalarprodukt eines Vektors mit sich selbst gleich dem Quadrat der Länge und verschwindet somit genau dann, wenn der Vektor gleich dem Nullvektor ist. Dies gilt auch für die gerade definierte

Bilinearform auf $\mathbb{R}^n$, denn

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = v_1^2 + \cdots + v_n^2$$

verschwindet als Summe von Quadraten genau dann, wenn jedes einzelne $v_i$ verschwindet.

Über dem Körper $\mathbb{F}_2$ dagegen ist beispielsweise

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} = 1^2 + 0^2 + 1^2 + 0^2 = 1 + 1 = 0\,,$$

und auch über den komplexen Zahlen ist etwa in $\mathbb{C}^2$

$$\begin{pmatrix} 1 \\ i \end{pmatrix} \cdot \begin{pmatrix} 1 \\ i \end{pmatrix} = 1^2 + i^2 = 1 - 1 = 0\,.$$

Auch über den reellen Zahlen lassen sich Bilinearformen finden, für die das Produkt eines Vektors mit sich selbst verschwinden kann, ohne daß der Vektor gleich dem Nullvektor sein müßte: Für die Bilinearform

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \star \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \underset{\text{def}}{=} v_1 w_1 - v_2 w_2$$

etwa ist

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \star \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1^2 - 1^2 = 0\,.$$

Da wir im $\mathbb{R}^n$ (und später auch in allgemeineren Räumen) Längen und Abstände definieren wollen, um so auch dort analytische Grundbegriffe wie Konvergenz und Stetigkeit einführen zu können, müssen wir solche Bilinearformen ausschließen: Abstände dürfen nicht negativ sein, und der Abstand zwischen zwei Punkten soll nur dann verschwinden, wenn beide Punkte gleich sind.

Da man in beliebigen Körpern nicht von Positivität und Negativität reden kann, beschränken wir uns ab jetzt auf den Fall $k = \mathbb{R}$ und definieren:



**Definition:** *a)* Eine symmetrische Bilinearform $V \times V \to \mathbb{R}$ auf einem reellen Vektorraum $V$ heißt *positiv semidefinit,* wenn

$$\vec{v} \cdot \vec{v} \geq 0 \quad \text{für alle } \vec{v} \in V\,;$$

sie heißt *positiv definit* oder *Skalarprodukt,* wenn zusätzlich gilt

$$\vec{v} \cdot \vec{v} = 0 \Rightarrow \vec{v} = \vec{0}\,.$$

*b)* Ein EUKLIDischer Vektorraum ist ein Paar $(V, \cdot)$ bestehend aus einem $\mathbb{R}$-Vektorraum $V$ und einem Skalarprodukt $\cdot : V \times V \to \mathbb{R}$.

Wie bei Produkten üblich, werden wir den Malpunkt oft weglassen. Gelegentlich, wenn Verwechslungen mit anderen Produkten zu befürchten sind, werden wir anstelle von $\vec{v} \cdot \vec{w}$ auch $(\vec{v},\, \vec{w})$ schreiben; in der Literatur findet man manchmal auch die Schreibweise $\langle \vec{v},\, \vec{w} \rangle$.

Typisches Beispiel eines EUKLIDischen Vektorraums ist natürlich der $\mathbb{R}^n$ mit seinem Standardskalarprodukt

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v_1 w_1 + \cdots + v_n w_n\,.$$

Ein anderes Beispiel ist der Raum $C^0([0,\,1],\,\mathbb{R})$ aller stetiger Funktionen vom Einheitsintervall $[0,\,1]$ nach $\mathbb{R}$ mit dem Skalarprodukt

$$(f,\,g) \underset{\text{def}}{=} \int_0^1 f(t) g(t)\, dt\,.$$

(Dies ist ein typischer Fall, wo die Klammerschreibweise vorzuziehen ist, denn $f \cdot g$ ist bereits vergeben für die Funktion, die jedem $t \in [0,\,1]$ den Wert $f(t) \cdot g(t)$ zuordnet.)

Die Bilinearität des so definierten Produkts ist klar; außerdem ist es offensichtlich positiv semidefinit:

$$(f,\,f) = \int_0^1 f(t)^2\, dt \geq 0\,,$$

da der Integrand nirgends negativ wird. Falls $f$ nicht identisch verschwindet, gibt es einen Punkt $t_0 \in (0,\ 1)$ mit $f(t_0) = h \neq 0$. Wegen der Stetigkeit von $f$ gibt es dazu eine Umgebung $(a,\ b)$, so daß $|f(t)| > h/2$ für $t \in (a,\ b)$. Damit ist

$$(f,\ f) = \int_0^1 f(t)^2\,dt \geq \int_a^b f(t)^2\,dt > \int_a^b \frac{h^2}{4}\,dt = \frac{h^2}{4}\,(b-a) > 0\,,$$

$(f,\ f) = 0$ ist also nur möglich, wenn $f$ überall verschwindet.

Man beachte, daß hier die Stetigkeit von $f$ eine wesentliche Rolle spielt: Für die Funktion $f$, die überall verschwindet außer im Punkt $1/2$, wo sie den Wert 1 annimmt, ist das Integral über $f(t)^2$ gleich null, obwohl die Funktion nicht die Nullfunktion ist.

Da wir EUKLIDische Vektorräume eingeführt haben, um dort so etwas wie EUKLIDische Geometrie zu betreiben, sollten wir als nächstes deren Grundbegriffe definieren:

**Definition:** *a)* Die Länge eines Vektors $\vec{v}$ aus einem EUKLIDischen Vektorraum $V$ ist $|\vec{v}| = \sqrt{\vec{v}\cdot\vec{v}}$.
*b)* Der Winkel zwischen zwei Vektoren $\vec{v}, \vec{w} \in V \setminus \{\vec{0}\}$ ist

$$\measuredangle(\vec{v}, \vec{w}) = \arccos\left(\frac{\vec{v}\cdot\vec{w}}{\sqrt{(\vec{v}\cdot\vec{v})(\vec{w}\cdot\vec{w})}}\right)\,.$$

Da der Arkuscosinus nur Werte zwischen null und $\pi$ annimmt, im Gradmaß also zwischen $0^\circ$ und $180^\circ$, ist der so definierte Winkel *unorientiert.*

Wir sollten uns allerdings die Frage stellen, ob hiermit *überhaupt* ein Winkel erklärt ist: Da der Cosinus nur Werte zwischen $-1$ und 1 annimmt, ist das offensichtlich nur dann der Fall, wenn das Argument des Arkuskosinus in obiger Definition höchstens den Betrag eins hat.

Für das Standardskalarprodukt im $\mathbb{R}^2$ oder $\mathbb{R}^3$ ist das kein Problem: Das Skalarprodukt im $\mathbb{R}^3$ hatten wir im vorigen Abschnitt *definiert* als Produkt aus den Längen der beiden Vektoren und dem Cosinus des eingeschlossenen Winkels; wir hatten dann nachgerechnet, daß dies mit



der Koordinatendefinition des hier definierten Skalarprodukts übereinstimmt.

Da zwei Vektoren stets eine Ebene aufspannen, sollten wir erwarten, daß Entsprechendes auch für beliebige reelle Vektorräume gilt, was auch in der Tat der Fall ist. Mit dem Beweis wollen wir allerdings noch warten bis zum übernächsten Abschnitt, wo wir ohne nennenswerten zusätzlichen Aufwand gleich auch noch eine ähnliche Formel für komplexe Vektorräume beweisen können.

Für EUKLIDische Vektorräume wie $\mathcal{C}^0\big([0,\ 1], \mathbb{R}\big)$ ist die „Länge“ eines „Vektors“ natürlich nichts mehr, was man sich geometrisch anschaulich vorstellen könnte; trotzdem ist sie oft eine nützliche Größe. Eine stetige Funktion $I\colon [0,\ 1] \to \mathbb{R}$ könnte beispielsweise einen zeitabhängigen Strom beschreiben, der durch einen Widerstand $R$ fließt; nach dem OHMschen Gesetz fällt dort eine Spannung $U(t) = R \cdot I(t)$ ab, so daß die elektrische Leistung gleich $U(t)\cdot I(t) = R \cdot I(t)^2$ ist. Die elektrische Arbeit, die während des Zeitraums $[0,\ 1]$ verrichtet wird, ist daher gleich

$$\int_0^1 U(t)\cdot I(t)\,dt = \int_0^1 R\cdot I(t)^2\,dt = R\cdot\int_0^1 I(t)^2\,dt = R\cdot|I|\,,$$

d.h. die „Länge“ von $I$ ist bis auf einen konstanten Faktor gerade gleich der Energie des Signals $I$.

Ähnliche physikalische Interpretationen gibt es bei fast allen Anwendungen von Vektorräumen, deren Elemente Funktionen sind.

## c) Hermitesche Vektorräume

Die unmittelbare Verallgemeinerung des Skalarprodukts des $\mathbb{R}^n$ auf $\mathbb{C}^n$ ist nicht sonderlich nützlich: Dann hätten wir nämlich zum Beispiel

$$\begin{pmatrix}1\\ i\end{pmatrix}\cdot\begin{pmatrix}1\\ i\end{pmatrix} = 1\cdot 1 + i\cdot i = 0\,.$$

Solche Vektoren mit Quadrat Null sind zwar in einigen Anwendungen (wie etwa der speziellen Relativitätstheorie) durchaus sinnvoll und nützlich, meist möchte man aber das Skalarprodukt eines Vektors mit sich

selbst als Quadrat seiner Länge interpretieren, und die sollte für alle Vektoren außer dem Nullvektor positiv sein.

Wir haben bereits jeder komplexen Zahl ungleich Null eine positive reelle Zahl zugeordnet, nämlich ihren Betrag

$$|z| = \sqrt{z\overline{z}}\,.$$

Entsprechend können wir zwei komplexen Vektoren

$$\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \quad \text{und} \quad \vec{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$

aus $\mathbb{C}^n$ das Produkt

$$\vec{v} \cdot \vec{w} = \sum_{\ell=1}^{n} v_\ell \overline{w_\ell}$$

zuordnen. Auch hier ist offensichtlich $\vec{v} \cdot \vec{v}$ für jeden Vektor außer dem Nullvektor eine positive reelle Zahl, denn wir addieren ja die Betragsquadrate der Komponenten des Vektors.

Dieses Produkt erfüllt nun allerdings nicht mehr die Forderungen aus der Definition eines Skalarprodukts: Es ist zwar noch linear im ersten Argument, nicht mehr aber im zweiten, denn schon bei Multiplikation des zweiten Vektors mit einer komplexen Zahl $\lambda \notin \mathbb{R}$ ist

$$\vec{v} \cdot (\lambda\vec{w}) = \sum_{\ell=1}^{n} v_\ell \cdot \overline{\lambda w_\ell} = \overline{\lambda} \cdot \sum_{\ell=1}^{n} v_\ell \overline{w_\ell} = \overline{\lambda} \cdot (\vec{v} \cdot \vec{w}) \neq \lambda \cdot (\vec{v} \cdot \vec{w})\,.$$

Bezüglich der Addition gibt es keine Probleme; somit wird die Linearitätsregel für das zweite Argument zu

$$\vec{u} \cdot (\lambda\vec{v} + \mu\vec{w}) = \overline{\lambda} \cdot (\vec{u} \cdot \vec{v}) + \overline{\mu} \cdot (\vec{u} \cdot \vec{w})\,.$$

Die Symmetriebedingung ist ebenfalls verletzt, denn schon für zwei komplexe Zahlen $z$ und $w$ ist das Produkt $z\overline{w} = \overline{w}z = \overline{w\overline{z}}$ im allgemeinen verschieden von $w\overline{z}$, und entsprechend ist auch für zwei Vektoren

$$\vec{v} \cdot \vec{w} = \overline{\vec{w} \cdot \vec{v}}\,.$$

Der Begriff des HERMITEschen Vektorraums formalisiert diese Eigenschaften:



**Definition:** Ein HERMITEscher Vektorraum ist ein Paar $(V, \cdot)$ bestehend aus einem $\mathbb{C}$-Vektorraum $V$ und einer Abbildung

$$\cdot : V \times V \to \mathbb{C}; \quad (\vec{v}, \vec{w}) \mapsto \vec{v} \cdot \vec{w}$$

mit folgenden Eigenschaften:

*a)* $\cdot$ ist linear im ersten Argument, d.h.

$$(\lambda\vec{u} + \mu\vec{v}) \cdot \vec{w} = \lambda(\vec{u} \cdot \vec{w}) + \mu(\vec{v} \cdot \vec{w})$$

für alle $\vec{u}, \vec{v}, \vec{w} \in V$ und $\lambda, \mu \in \mathbb{C}$.

*b)* $\cdot$ ist HERMITEsch symmetrisch, d.h.

$$\vec{v} \cdot \vec{w} = \overline{\vec{w} \cdot \vec{v}}$$

für alle $\vec{v}, \vec{w} \in V$.

*c)* $\cdot$ ist positiv definit, d.h.

$$\vec{v} \cdot \vec{v} > 0$$

ist eine positive *reelle* Zahl für alle Vektoren $\vec{v} \neq \vec{0}$ aus $V$.

Die bilineare Abbildung $\cdot$ heißt HERMITE*sches Skalarprodukt;* auch hier werden wir den Malpunkt nicht immer hinschreiben und bei Verwechslungsgefahr mit anderen Produkten auch gelegentlich $(\vec{v}, \vec{w})$ anstelle von $\vec{v} \cdot \vec{w}$ schreiben.

CHARLES HERMITE (1822–1901) war einer der bedeutendsten Mathematiker des neunzehnten Jahrhunderts. Zu seinen Resultaten zählen eine Vereinfachung des ABELschen Beweises, daß Gleichungen fünften Grades im allgemeinen nicht durch Wurzelausdrücke gelöst werden können, die explizite Lösung solcher Gleichungen durch elliptische Funktionen, der Nachweis, daß $e$ eine transzendente Zahl ist, also keiner algebraischen Gleichung über $\mathbb{Q}$ genügt, eine Interpolationsformel und vieles mehr. HERMITE galt als ein sehr guter akademischer Lehrer; er unterrichtete an der École Polytechnique, dem Collège de France, der École Normale Supérieure und der Sorbonne.

Durch Kombination der Eigenschaften *a)* und *b)* kann man leicht ausrechnen, was anstelle der Linearität für das zweite Argument gilt: Genau wie im obigen Beispiel ist

$$\begin{aligned}\vec{u}\cdot(\lambda\vec{v}+\mu\vec{w}) &= \overline{(\lambda\vec{v}+\mu\vec{w})\cdot\vec{u}} = \overline{\lambda(\vec{v}\cdot\vec{u})+\mu(\vec{w}\cdot\vec{u})}\\ &= \overline{\lambda}\,\overline{(\vec{v}\cdot\vec{u})}+\overline{\mu}\,\overline{(\vec{w}\cdot\vec{u})} = \overline{\lambda}\,(\vec{u}\cdot\vec{v})+\overline{\mu}\,(\vec{u}\cdot\vec{w})\,.\end{aligned}$$

Diese Eigenschaft, zusammen mit der Linearität im ersten Argument, bezeichnet man gelegentlich als *Sesquilinearität,* d.h. „anderthalbfache Linearität", im Gegensatz zur echten Linearität in zwei Argumenten, der *Bilinearität.*

Typisches Beispiel eines HERMITEschen Vektorraums ist der $\mathbb{C}^n$ mit dem eingangs definierten Produkt, jedoch wird sich im nächsten Semester zeigen, daß gerade HERMITEsche Vektorräume von komplexwertigen Funktionen sehr interessante Anwendungen haben.

Hier sei nur als Beispiel für das Rechnen mit HERMITEschen Skalarprodukten in $\mathbb{C}^n$ gezeigt, mit dem wir auch nochmals das Rechnen mit Matrizen wiederholen können, und zwar geht es um ein Verfahren von G. PHILIPPE, veröffentlicht in *Quadrature,* Oct.-Déc. 2000, S. 23–34, wie man aus komplexen Vektoren reelle quadratische Matrizen $A, B$ definieren kann, für die $AB = -BA$ ist, die also *antikommutieren.*

Dazu betrachten wir einen Vektor $\vec{v} \in \mathbb{C}^n$, wobei $\mathbb{C}^n$ sein Standard-HERMITEsches Produkt habe. Zu den Komponenten $v_i$ von $\vec{v}$ betrachten wir die $n \times n$-Matrix $W$ mit Einträgen $w_{ij} = v_i\overline{v_j}$ und die Matrix

$$M = a\cdot E - 2W \quad \text{mit} \quad a = \vec{v}\cdot\vec{v}\,.$$

Da $aE$ als Diagonalmatrix mit $W$ kommutiert, ist

$$M^2 = a^2E - 4aW + 4W^2\,.$$

Der Eintrag an der Stelle $ik$ von $W^2$ ist

$$\sum_{j=1}^{n} w_{ij}w_{jk} = \sum_{j=1}^{n} v_i\overline{v_j}\cdot v_j\overline{v_k} = v_i\overline{v}_k\sum_{j=1}^{n} v_j\overline{v_j} = w_{ik}(\vec{v}\cdot\vec{v}) = aw_{ik}\,,$$

d.h. $M^2 = a^2E$.



Bezeichen $A$ und $B$ die reellen $n \times n$-Matrizen aus den Real- und Imaginärteilen von $M$, ist also $M = A + iB$, so ist demnach

$$(A+iB)^2 = A^2 - B^2 + i(AB+BA) = a^2E$$

eine reelle Matrix, der Imaginärteil $AB + BA$ muß also verschwinden. Das ist aber gleichbedeutend damit, daß $AB = -BA$ ist.

Als Beispiel betrachten wir den Vektor $\binom{1}{i}$ aus $\mathbb{C}^2$. Hier ist $a = 2$ und

$$M = \begin{pmatrix} 1 & -i \\ i & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + i\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\,,$$

und in der Tat ist

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = -E \quad \text{und} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = E\,.$$

Kompliziertere Vektoren $\vec{v}$ führen zu interessanteren Beispielen.

Auch in HERMITEschen Vektorräumen werden wir $|\vec{v}| = \sqrt{\vec{v}\cdot\vec{v}}$ gelegentlich als *Länge* des Vektors $\vec{v}$ bezeichnen; auf die Definition von Winkeln hingegen wollen wir verzichten, da die Übernahme der entsprechenden Definition für EUKLIDische Vektorräume hier auf geometrisch nicht sinnvolle komplexe Winkel führen würde.

Ansonsten können und werden wir EUKLIDische und HERMITEsche Vektorräume im folgenden gleichzeitig behandeln: Ersetzt man in der Definition eines HERMITEschen Vektorraums überall $\mathbb{C}$ durch $\mathbb{R}$, so erhält man einen EUKLIDischen Vektorraum. Zwar stehen noch an vielen Stellen die Querstriche für komplexe Konjugation, aber da sich eine reelle Zahl unter komplexer Konjugation nicht ändert, sind die Konjugationsstriche nur überflüssig, nicht schädlich.

## d) Die Cauchy-Schwarzsche Ungleichung

Als erstes Beispiel wollen wir das noch offene Problem aus Abschnitt *b)* lösen und zeigen, daß die Zahl, die wir dort als Cosinus eines Winkels definiert hatten, tatsächlich zwischen null und eins liegt. Natürlich zeigen wir dies gleich etwas allgemeiner so, daß wir auch eine Aussage für HERMITEsche Vektorräume bekommen, und wir verallgemeinern

auch gleich noch etwas weiter in Hinblick auf die Tatsache, daß wir in Vektorräumen, die auch unstetige Funktionen enthalten, im allgemeinen kein wirkliches Skalar- *bzw.* HERMITEsches Produkt haben. Trotzdem werden Produkte in solchen Vektorräumen im nächsten Semester für die FOURIER-Theorie sehr nützlich sein.

**Cauchy-Schwarzsche Ungleichung:** Für $k = \mathbb{R}$ oder $\mathbb{C}$ sei auf dem $k$-Vektorraum $V$ eine Abbildung $\cdot\colon V \times V \to k;\ (f, g) \mapsto f \cdot g$ gegeben mit den Eigenschaften
*a)* $(\lambda\vec{u} + \mu\vec{v}) \cdot \vec{w} = \lambda(\vec{u} \cdot \vec{w}) + \mu(\vec{v} \cdot \vec{w})$
*b)* $\vec{w} \cdot \vec{v} = \overline{\vec{v} \cdot \vec{w}}$
*c)* $|\vec{v}|^2 = \vec{v} \cdot \vec{v} \geq 0$ für alle $\vec{v} \in V$.
Dann ist für alle Vektoren $\vec{v}, \vec{w} \in V$

$$|\vec{v} \cdot \vec{w}| \leq |\vec{v}| \cdot |\vec{w}| \ .$$

*(Man beachte, daß hier nicht gefordert wird, daß $\vec{v} \cdot \vec{v} > 0$ für $\vec{v} \neq \vec{0}$.)*

Der *Beweis* beruht auf folgendem Trick: Wegen *c)* ist für $\lambda, \mu \in \mathbb{C}$

$$\begin{aligned}(\lambda\vec{v} + \mu\vec{w}) \cdot (\lambda\vec{v} + \mu\vec{w}) &= \lambda\overline{\lambda}\,\vec{v} \cdot \vec{v} + \lambda\overline{\mu}\,\vec{v} \cdot \vec{w} + \mu\overline{\lambda}\,\vec{w} \cdot \vec{v} + \mu\overline{\mu}\,\vec{w} \cdot \vec{w} \\ &= \lambda\overline{\lambda}\,\vec{v} \cdot \vec{v} + \lambda\overline{\mu}\,\vec{v} \cdot \vec{w} + \overline{\lambda}\mu\,\overline{\vec{v} \cdot \vec{w}} + \mu\overline{\mu}\,\vec{w} \cdot \vec{w}\end{aligned}$$

stets nichtnegativ. Speziell für $\lambda = (\vec{w} \cdot \vec{w}) \in \mathbb{R}$ und $\mu = -(\vec{v} \cdot \vec{w})$ erhalten wir

$$(\vec{w} \cdot \vec{w})^2(\vec{v} \cdot \vec{v}) - 2(\vec{w} \cdot \vec{w})(\overline{\vec{v} \cdot \vec{w}})(\vec{v} \cdot \vec{w}) + (\vec{v} \cdot \vec{w})(\overline{\vec{v} \cdot \vec{w}})(\vec{w} \cdot \vec{w}) \geq 0\,,$$

also

$$(\vec{w} \cdot \vec{w})\big((\vec{w} \cdot \vec{w})(\vec{v} \cdot \vec{v}) - |\vec{v} \cdot \vec{w}|^2\big) \geq 0\,.$$

Ist hier $\vec{w} \cdot \vec{w} \neq 0$, können wir durch diese Zahl dividieren und die Behauptung ist bewiesen. Andernfalls können wir, falls wenigstens $\vec{v} \cdot \vec{v}$ nicht null ist, im obigen Argument die Rollen von $\vec{v}$ und $\vec{w}$ vertauschen und damit die Behauptung beweisen.

Wenn schließlich $\vec{v} \cdot \vec{v}$ und $\vec{w} \cdot \vec{w}$ beide null sind, folgt aus

$$(\vec{v} \pm \vec{w}) \cdot (\vec{v} \pm \vec{w}) = \pm(\vec{v} \cdot \vec{w} + \overline{\vec{v} \cdot \vec{w}}) = \pm 2\,\mathfrak{Re}(\vec{v}\vec{w}) \geq 0\,,$$



daß der Realteil von $\vec{v} \cdot \vec{w}$ verschwindet, und genauso verschwindet auch der Imaginärteil, da $(i\vec{v} \pm \vec{w}) \cdot (i\vec{v} \pm \vec{w}) = \pm 2\,\mathfrak{Im}(\vec{v} \cdot \vec{w}) \geq 0$ ist. Somit ist $\vec{v} \cdot \vec{w} = 0$, und die Ungleichung gilt auch in diesem Fall. ■

Baron AUGUSTIN LOUIS CAUCHY (1789–1857) stellte als erster durch die exakte Definition von Begriffen wie *Konvergenz* und *Stetigkeit* die Analysis auf ein sicheres Fundament. In insgesamt 789 Arbeiten beschäftigte er sich u.a. auch mit komplexer Analysis, Variationsrechnung, Differentialgleichungen, FOURIER-Analysis, Permutationsgruppen, der Diagonalisierung von Matrizen und der theoretischen Mechanik. Als überzeugter Royalist hatte er häufig Schwierigkeiten mit den damaligen Regierungen; er lebte daher mehrere Jahre im Exil in Turin und später in Prag, wo er (mit sehr mäßigem Erfolg) den französischen Thronfolger unterrichtete.

Der deutsche Mathematiker KARL HERMAN AMANDUS SCHWARZ (1843–1921) beschäftigte sich hauptsächlich mit konformen Abbildungen und mit sogenannten Minimalflächen, d.h. Flächen mit vorgegebenen Eigenschaften, deren Flächeninhalt minimal ist. Im Rahmen einer entsprechenden Arbeit für die WEIERSTRASS-Festschrift von 1885 (im Falle eines durch Doppelintegrale definierten Skalarprodukts) bewies er die obige Ungleichung; CAUCHY hatte sie bereits in seinem Analysislehrbuch von 1821 für endlichdimensionale Vektoren bewiesen. SCHWARZ lehrte nacheinander in Halle, Zürich, Göttingen und Berlin.

## e) Orthonormalbasen

Genau wie im Falle des $\mathbb{R}^3$ wollen wir auch für beliebige EUKLIDische oder HERMITEsche Vektorräume zwei Vektoren als *orthogonal* bezeichnen, wenn ihr (HERMITEsches) Skalarprodukt verschwindet. Viele Rechnungen vereinfachen sich, wenn man von einer Basis ausgeht, deren Vektoren paarweise orthogonal sind und möglicherweise zusätzlich noch die Länge eins habe; um diese Art von Basen soll es hier gehen:

**Definition:** *a)* Eine Basis $\mathcal{B}$ eines EUKLIDischen oder HERMITEschen Vektorraums heißt *Orthogonalbasis,* wenn jedes Element von $\mathcal{B}$ orthogonal zu allen übrigen ist.

*b)* Eine Orthogonalbasis $\mathcal{B}$ heißt *Orthonormalbasis,* wenn zusätzlich für jeden Vektor $\vec{b} \in \mathcal{B}$ gilt: $\vec{b} \cdot \vec{b} = 1$.

Standardbeispiel sind die Einheitsvektoren

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \vec{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

die sowohl für $\mathbb{R}^n$ als auch für $\mathbb{C}^n$ bezüglich des jeweiligen Standardskalar- oder HERMITEschen Produkts eine Orthonormalbasis bilden.

Tatsächlich hat sogar *jeder* EUKLIDische oder HERMITEsche Vektorraum eine Orthonormalbasis; für den Beweis werden wir uns allerdings, wie immer, wenn von Basen die Rede ist, zur Vermeidung logischer Schwierigkeiten auf den endlichdimensionalen Fall beschränken:

**Satz:** Jeder endlichdimensionale EUKLIDische oder HERMITEsche Vektorraum $V$ hat eine Orthonormalbasis.

*Beweis:* Zunächst wissen wir, daß $V$ überhaupt eine Basis $(\vec{b}_1, \ldots, \vec{b}_n)$ hat. Daraus konstruieren wir schrittweise eine *Orthogonalbasis* nach dem sogenannten GRAM-SCHMIDTschen Orthogonalisierungsverfahren. Dieses liefert in seinem $r$-ten Schritt eine Orthogonalbasis $(\vec{c}_1, \ldots, \vec{c}_r)$ des von dern ersten $r$ Basisvektoren aufgespannten Untervektorraums $[\vec{b}_1, \ldots, \vec{b}_r]$.

Der erste Schritt ist der einfachste: Da es noch keine Orthogonalitätsbedingung für $\vec{c}_1$ gibt, können wir einfach $\vec{c}_1 = \vec{b}_1$ setzen.

Nachdem wir $r \geq 1$ Schritte durchgeführt haben, haben wir $r$ linear unabhängige Vektoren $\vec{c}_1, \ldots, \vec{c}_r$ mit $\vec{c}_i \cdot \vec{c}_j = 0$ für $i \neq j$ aus dem von $\vec{b}_1, \ldots, \vec{b}_r$ aufgespannten Untervektorraum. Ist $r = n$, haben wir eine Orthogonalbasis; andernfalls muß ein auf den bisher konstruierten $\vec{c}_i$ senkrecht stehender Vektor $\vec{c}_{r+1}$ gefunden werden, der zusammen mit diesen den von $\vec{b}_1$ bis $\vec{b}_{r+1}$ erzeugten Untervektorraum erzeugt.

Da $\vec{c}_1, \ldots, \vec{c}_r$ und $\vec{b}_1, \ldots, \vec{b}_r$ denselben Untervektorraum erzeugen, gilt dasselbe für $\vec{c}_1, \ldots, \vec{c}_r, \vec{b}_{r+1}$ und $\vec{b}_1, \ldots, \vec{b}_{r+1}$; das Problem ist, daß $\vec{b}_{r+1}$



im allgemeinen nicht orthogonal zu den $\vec{c}_i$ sein wird. Wir dürfen $\vec{b}_{r+1}$ aber abändern um einen beliebigen Vektor aus dem von $\vec{c}_1, \ldots, \vec{c}_r$ aufgespannten Untervektorraum; also setzen wir

$$\vec{c}_{r+1} = \vec{b}_{r+1} + \lambda_1 \vec{c}_1 + \cdots + \lambda_r \vec{c}_r$$

und versuchen, die $\lambda_i$ so zu bestimmen, daß dieser Vektor orthogonal zu $\vec{c}_1, \ldots, \vec{c}_r$ wird.

Wegen der Orthogonalität der $\vec{c}_i$ ist

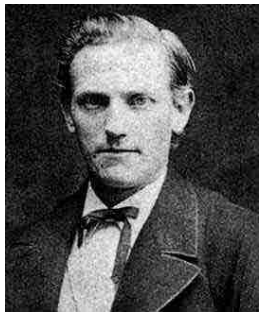
$$\vec{c}_{r+1} \cdot \vec{c}_i = \vec{b}_{r+1} \cdot \vec{c}_i + \sum_{j=1} \lambda_j (\vec{c}_j \cdot \vec{c}_i) = \vec{b}_{r+1} \cdot \vec{c}_i + \lambda_i (\vec{c}_i \cdot \vec{c}_i)\,;$$

setzen wir daher $\lambda_i = -\dfrac{\vec{b}_{r+1} \cdot \vec{c}_i}{\vec{c}_i \cdot \vec{c}_i}$, so ist $\vec{v} \cdot \vec{c}_i = 0$ für alle $i = 1, \ldots, r$.

Nach dem $n$-ten Schritt haben wir eine Orthogonalbasis $(\vec{c}_1, \ldots, \vec{c}_n)$ von $V$ konstruiert. Daraus wird die gewünschte *Orthonormalbasis* $(\vec{e}_1, \ldots, \vec{e}_n)$, wenn wir jeden Vektor durch seine Länge dividieren, d.h.

$$\vec{e}_i = \frac{\vec{c}_i}{|\vec{c}_i|}\,.$$

■



Der dänische Mathematiker JØRGAN PEDERSEN GRAM (1850–1916) lehrte an der Universität Kopenhagen, war aber gleichzeitig auch noch geschäftsführender Direktor einer Versicherungsgesellschaft und Präsident des Verbands der dänischen Versicherungsunternehmen. Er publizierte anscheinend nur eine einzige mathematische Arbeit *Sur quelque théorèmes fondamentaux de l'algèbre moderne,* die 1874 erschien. Das GRAM-SCHMIDTsche Orthogonalisierungsverfahren, durch das er heute hauptsächlich bekannt ist, stammt wohl von LAPLACE (1749–1827) und wurde auch schon 1836 von CAUCHY verwendet.

ERHARD SCHMIDT (1876–1959) wurde in Estland geboren; er studierte in Berlin bei SCHWARZ und promovierte in Göttingen bei HILBERT.

ERHARD SCHMIDT ist einer der Begründer der modernen Funktionalanalysis; insbesondere geht die Verallgemeinerung EUKLIDischer und HERMITEscher Vektorräume zu sogenannten HILBERT-Räumen, mit der wir uns im nächsten Semester im Zusammenhang mit der FOURIER-Analysis und der Theorie der Differentialgleichungen beschäftigen werden, auf ihn zurück. Er war Professor in Zürich, Erlangen, Breslau und Berlin.

Um auch den Umgang mit HERMITEschen Skalarprodukten und das Rechnen mit komplexen Zahlen zu üben, betrachten wir als Beispiel den von

$$\vec{b}_1 = \begin{pmatrix} 1 \\ i \\ -i \\ -1 \end{pmatrix}, \quad \vec{b}_2 = \begin{pmatrix} 1+2i \\ -2+i \\ -i \\ -1 \end{pmatrix} \quad \text{und} \quad \vec{b}_3 = \begin{pmatrix} 1+2i \\ -i \\ -4-i \\ 1+2i \end{pmatrix}$$

aufgespannten Untervektorraum von $\mathbb{C}^4$.

Wie oben im Beweis können wir bei der Anwendung des GRAM-SCHMIDTschen Orthogonalisierungsverfahrens $\vec{c}_1 = \vec{b}_1$ setzen; im zweiten Schritt suchen wir einen Vektor $\vec{c}_2 = \vec{b}_2 + \lambda_1 \vec{c}_1$ für den gilt:

$$\vec{c}_2 \cdot \vec{c}_1 = (\vec{b}_2 + \lambda_1 \vec{c}_1) \cdot \vec{c}_1 = \vec{b}_2 \cdot \vec{c}_1 + \lambda_1 (\vec{c}_1 \cdot \vec{c}_1) = 0$$

(Wir könnten natürlich auch fordern, daß

$$\vec{c}_1 \cdot \vec{c}_2 = \vec{c}_1 \cdot (\vec{b}_2 + \lambda_1 \vec{c}_1) = \vec{c}_1 \cdot \vec{b}_2 + \overline{\lambda}_1 (\vec{c}_1 \cdot \vec{c}_1)$$

verschwindet, aber dann erhalten wir zunächst eine Formel für $\overline{\lambda}_1$ und müssen noch komplex konjugieren. Daher ist zumindest im komplexen Fall die Forderung $\vec{c}_2 \cdot \vec{c}_1 = 0$ rechnerisch etwas bequemer – auch wenn natürlich beides auf dasselbe Ergebnis führt.)

Wir brauchen also die HERMITEschen Skalarprodukte

$$\vec{b}_2 \cdot \vec{c}_1 = \begin{pmatrix} 1+2i \\ -2+i \\ -i \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ i \\ -i \\ -1 \end{pmatrix} \text{ und } \vec{c}_1 \cdot \vec{c}_1 = \begin{pmatrix} 1 \\ i \\ -i \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ i \\ -i \\ -1 \end{pmatrix} .$$



Bei deren Berechnung darf man auf keinen Fall vergessen, die Einträge des jeweils zweiten Vektors komplex zu konjugieren, d.h.

$$\begin{aligned} \vec{b}_2 \cdot \vec{c}_1 &= (1+2i) \cdot \overline{1} + (-2+i) \cdot \overline{i} + (-i) \cdot \overline{(-i)} + (-1) \cdot \overline{(-1)} \\ &= (1+2i) \cdot 1 + (-2+i) \cdot (-i) + (-i) \cdot i + (-1) \cdot (-1) \ . \\ &= (1+2i) + (1+2i) + 1 + 1 = 4 + 4i \end{aligned}$$

Entsprechend ist

$$\begin{aligned} \vec{c}_1 \cdot \vec{c}_1 &= 1 \cdot \overline{1} + i \cdot \overline{i} + (-i) \cdot \overline{(-i)} + (-1) \cdot \overline{(-1)} \\ &= 1 \cdot 1 + i \cdot (-i) + (-i) \cdot i + (-1) \cdot (-1) = 1 + 1 + 1 + 1 = 4 \, , \end{aligned}$$

wobei das allerdings auch mit weniger Mühe einzusehen ist: Da für einen Vektor $\vec{v} \in \mathbb{C}^n$ mit Komponenten $v_1, \ldots, v_n$

$$\vec{v} \cdot \vec{v} = \sum_{i=1}^{n} v_i \cdot \overline{v_i} = \sum_{i=1}^{n} |v_i|^2$$

ist, hätte es genügt, einfach die Betragsquadrate der vier Einträge aufzusummieren, wobei im vorliegenden Fall trivial ist, daß diese allesamt gleich eins sind.

Einsetzen der beiden Skalarprodukte in die Gleichung für $\lambda_1$ zeigt, daß

$$(4+4i) + 4\lambda_1 = 0 \quad \text{und damit} \quad \lambda_1 = \frac{-4-4i}{4} = -1 - i$$

sein muß. Somit ist

$$\vec{c}_2 = \vec{b}_2 + \lambda_1 \vec{c}_1 = \begin{pmatrix} 1+2i \\ -2+i \\ -i \\ -1 \end{pmatrix} - \begin{pmatrix} 1+i \\ -1+i \\ 1-i \\ -1-i \end{pmatrix} = \begin{pmatrix} i \\ -1 \\ -1 \\ i \end{pmatrix} .$$

Im dritten Schritt muß ein Vektor

$$\vec{c}_3 = \vec{b}_3 + \mu_1 \vec{c}_1 + \mu_2 \vec{c}_2$$

berechnet werden mit

$$\vec{c}_3 \cdot \vec{c}_1 = \vec{c}_3 \cdot \vec{c}_2 = 0 \, .$$

Hier ist

$$\vec{c}_3 \cdot \vec{c}_1 = \vec{b}_3 \cdot \vec{c}_1 + \mu_1 \vec{c}_1 \cdot \vec{c}_1 + \mu_2 \vec{c}_2 \cdot \vec{c}_1 = \vec{b}_3 \cdot \vec{c}_1 + \mu_1 \vec{c}_1 \cdot \vec{c}_1 \, ,$$

da der zweite Schritt sicherstellte, daß $\vec{c}_2 \cdot \vec{c}_1 = 0$ ist. Das HERMITEsche Skalarprodukt $\vec{c}_1 \cdot \vec{c}_1 = 4$ kennen wir bereits, und eine kurze Rechnung zeigt, daß

$$\vec{b}_3 \cdot \vec{c}_1 = (1+2i) + (-1) + (1-4i) + (-1-2i) = -4i$$

ist. Also folgt

$$\mu_1 = -\frac{\vec{b}_3 \cdot \vec{c}_1}{\vec{c}_1 \cdot \vec{c}_1} = -\frac{-4i}{4} = i$$

und entsprechend folgt auch

$$\mu_2 = -\frac{\vec{b}_3 \cdot \vec{c}_2}{\vec{c}_2 \cdot \vec{c}_2} = -\frac{8}{4} = -2\,,$$

wobei diesesmal die genaue Berechnung der beiden HERMITEschen Skalarprodukte dem Leser als Übungsaufgabe überlassen sei. Der dritte Basisvektor $\vec{c}_3 = \vec{b}_3 + i\vec{c}_1 - 2\vec{c}_2$ ist daher gleich

$$\begin{pmatrix} 1+2i \\ -i \\ -4-i \\ 1+2i \end{pmatrix} + i \begin{pmatrix} 1 \\ i \\ -i \\ -1 \end{pmatrix} - 2 \begin{pmatrix} i \\ -1 \\ -1 \\ i \end{pmatrix} = \begin{pmatrix} 1+i \\ 1-i \\ -1-i \\ 1-i \end{pmatrix}.$$

Damit ist eine *Orthogonalbasis* gefunden; wenn wir eine *Orthonormalbasis* wollen, müssen wir die Vektoren noch durch ihre Längen dividieren: Wir wissen bereits, daß $\vec{c}_1 \cdot \vec{c}_1 = \vec{c}_2 \cdot \vec{c}_2 = 4$ ist, so daß die ersten beiden Vektoren die Länge zwei haben; $\vec{c}_3$ hat vier Komponenten vom Betrag zwei, also ist

$$\vec{c}_3 \cdot \vec{c}_3 = 4 \cdot 2 = 8 \quad \text{und} \quad |\vec{c}_2| = \sqrt{8} = 2\sqrt{2}\,.$$

Die gesuchte Orthonormalbasis besteht daher aus den Vektoren

$$\frac{1}{2}\begin{pmatrix} 1 \\ i \\ -i \\ 1 \end{pmatrix}, \quad \frac{1}{2}\begin{pmatrix} i \\ -1 \\ -1 \\ i \end{pmatrix} \quad \text{und} \quad \frac{\sqrt{2}}{4}\begin{pmatrix} 1+i \\ 1-i \\ -1-i \\ 1-i \end{pmatrix}.$$

Die Nützlichkeit von Orthonormal- und Orthogonalbasen ergibt sich vor allem aus den beiden folgenden Sätzen:



**Satz:** $(\vec{e}_1, \ldots, \vec{e}_n)$ sei eine Orthonormalbasis 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

$$\vec{v} \cdot \vec{w} = \sum_{i=1}^{n} a_i \overline{b_i}\,.$$

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

$$\vec{v} = a_1\vec{e}_1 + \cdots + a_n\vec{e}_n \quad \text{mit} \quad a_i = \vec{v} \cdot \vec{e}_i\,.$$

Der *Beweis* ist in beiden Fällen ein Einzeiler:

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

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

■

Bezüglich einer Orthonormalbasis sieht also jedes (gewöhnliche oder HERMITEsche) Skalarprodukt aus wie das entsprechende Standardskalarprodukt – und das unabhängig von der gewählten Orthonormalbasis. Sobald man Vektoren bezüglich einer Orthonormalbasis dargestellt hat, 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:** $(\vec{e}_1, \ldots, \vec{e}_n)$ 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 $(\vec{b}_1,\ldots,\vec{b}_n)$. Nach dem GRAM-SCHMIDTschen Orthogonalisierungsverfahren können wir daraus eine Orthogonalbasis $(\vec{c}_1,\ldots,\vec{c}_n)$ von $V$ konstruieren, wobei gilt

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

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

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

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

Fassen wir die Vektoren $\vec{b}_1,\ldots\vec{b}_n$ 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 $i = 1,\ldots,n$ 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 $\ell$-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 $\vec{c}_1, \ldots, \vec{c}_k$ 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 $\vec{a}_1, \ldots, \vec{a}_5$.

$\vec{a}_1$ 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.

Insgesamt ist also $A = QR$ mit

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

Natürlich ist auch

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

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 Determinante einer orthogonalen *bzw.* unitären Matrix hat den Betrag eins.
*d)* Die inverse Matrix einer orthogonalen *bzw.* unitären Matrix ist wieder orthogonal *bzw.* unitär.
*e)* 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)* sei $U$ eine orthogonale oder unitäre Matrix. Dann ist $\det {}^tU = \det U$, also $\det U^* = \overline{\det U}$ und $\det U \cdot \det U^* = \det U \cdot \overline{\det U} = |\det U|^2$. Andererseits ist $\det U \cdot \det U^* = \det(UU^*) = \det E = 1$; somit hat $\det U$ den Betrag eins.

Für *d)* muß wegen *a)* und *b)* gezeigt werden, daß $U^*$ wieder unitär ist. Nach Definition ist $U$ invers zu dieser Matrix, wegen $(U^*)^* = U$ folgt die Unitarität.

Für *e)* 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

$$(A\vec{v}) \cdot (A\vec{w}) = \vec{v} \cdot \left(A^*(A\vec{w})\right) = \vec{v} \cdot \left((A^*A)\vec{w}\right)\,.$$

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