Hoofdstuk 3

Vectorruimten met inproduct

3.1 Inleiding

In $\mathbb{R}^2$ en $\mathbb{R}^3$ hebben we behalve de optelling en scalairvermenigvuldiging nog meer "structuur"; bij een vector kun je spreken over zijn lengte en bij twee vectoren hebben we het zgn. inproduct dat een verband heeft met de hoek tussen de vectoren. Precieser, als $v = (x_1, x_2, x_3)^t \in \mathbb{R}^3$ dan wordt volgens de stelling van Pythagoras de lengte van v, genoteerd $\|v\|$ gegeven door $\|v\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$. Het inproduct $\langle v, w\rangle$ van v en w wordt gegeven door

$$\langle v, w\rangle = \|v\| \, \|w\| \cos\alpha \tag{3.1}$$

waarbij α de hoek tussen de vectoren v en w is. Gebruik makend van de cosinusregel kan men nagaan (opgave 3.1.3) dat geldt

$$\langle v, w\rangle = x_1y_1 + x_2y_2 + x_3y_3.$$

Deze formule kunnen we gebruiken om een inproduct op $\mathbb{R}^n$ te definiëren:

Definitie 3.1.1 Zij $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ en $w = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ in $\mathbb{R}^n$. Met de voorschrijft

$$\langle v, w\rangle := x_1y_1 + \cdots + x_ny_n$$

wordt aan ieder tweetal vectoren v en w uit $\mathbb{R}^n$ een reëel getal $\langle v, w\rangle$ toegevoegd m.a.w. we hebben een afbeelding $\langle\ ,\ \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ gedefinieerd.

Opgave 3.1.2 Bewijs dat de in 3.1.1 gedefinieerde afbeelding aan de volgende eigenschappen voldoet.

i) $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$, voor alle v_1, v_2, w in $\mathbb{R}^n$.
ii) $\langle cv, w\rangle = c\langle v, w\rangle$, voor alle v, w in $\mathbb{R}^n$ en alle c in $\mathbb{R}$.
iii) $\langle v, w\rangle = \langle w, v\rangle$, voor alle v, w in $\mathbb{R}^n$.
iv) $\langle v, v\rangle > 0$ als $v \neq 0$.

Opgave 3.1.3 Zij $v = (x_1, x_2, x_3)^t$ en $w = (y_1, y_2, y_3)^t$ in $\mathbb{R}^3$ en $\langle v, w\rangle := \|v\| \, \|w\| \cos\alpha$, waarbij α de hoek is tussen v en w. Bewijs dat $\langle v, w\rangle = x_1y_1 + x_2y_2 + x_3y_3$.

[Aanw.: Bekijk de driehoek met als zijden de vector v, de vector w en de vector die loopt van het eindpunt van w naar het eindpunt van v. Deze zijden hebben lengte resp. $\|v\|$, $\|w\|$ en $\|v-w\|$. Volgens de cosinusregel (3.3.2 uit LA1) geldt dan

$$\|v-w\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\|\,\|w\|\cos\alpha.$$

Gebruik nu dat $\|v\|^2 = x_1^2 + x_2^2 + x_3^2$ enz.]

3.2 Abstracte inproductruimten

De eigenschappen van het inproduct op $\mathbb{R}^n$ zoals in 3.1.2 beschreven leiden ons tot de volgende definitie.

Definitie 3.2.1 Zij V een vectorruimte. Een *inproduct* op V is een afbeelding $\langle\ ,\ \rangle : V\times V \to \mathbb{R}$ die voldoet aan:
i) $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$, voor alle v_1, v_2, w in V.
ii) $\langle cv, w\rangle = c\langle v, w\rangle$, voor alle v, w in V en alle c in $\mathbb{R}$.
iii) $\langle v, w\rangle = \langle w, v\rangle$, voor alle v, w in V.
iv) $\langle v, v\rangle > 0$ als $v \neq 0$.
Een vectorruimte V met daarop een inproduct heet een *inproductruimte* of ook *ruimte met inproduct* of *prehilbertruimte*.
De *lengte* van een vector v in een inproductruimte definiëren we als $\|v\| := \sqrt{\langle v, v\rangle}$.

Omdat er bij twee vectoren nauwelijks een misverstand met andere vermenigvuldigingen kan bestaan, worden inproducten vaak gewoon met $v \cdot w$ i.p.v. $\langle v, w\rangle$ genoteerd,

Opgave 3.2.2 Ga na dat uit i)-iv) in 3.2.1 volgt dat $\langle v, 0\rangle = 0$ en $\langle 0, w\rangle = 0$ en dat $\langle v, -w\rangle = -\langle v, w\rangle$ voor alle $v, w \in V$.

Voorbeeld 3.2.3 $\mathbb{R}^n$ met $\langle\ ,\ \rangle$ zoals in 3.1.1 is volgens 3.1.2 een inproductruimte. We noemen het in 3.1.1 gedefinieerde inproduct het *standaard inproduct* op $\mathbb{R}^n$.

Voorbeeld 3.2.4 Zij $V := C([0,1])$ de vectorruimte der continue functies van $[0,1]$ naar $\mathbb{R}$. Definieer

$$\langle f, g\rangle := \int_0^1 f(x)g(x)\,dx, \text{ voor alle } f, g \in V.$$

Laat zien dat $\langle\ ,\ \rangle$ een inproduct op V is.

Oplossing. i)-iii): $\langle f_1 + f_2, g\rangle = \int_0^1 (f_1 + f_2)(x)g(x)\,dx = \int_0^1 (f_1(x)g(x) + f_2(x)g(x))\,dx = \langle f_1, g\rangle + \langle f_2, g\rangle$. Ook ziet men eenvoudig dat $\langle cf, g\rangle = c\langle f, g\rangle$, voor alle $c \in \mathbb{R}$ en dat $\langle f, g\rangle = \langle g, f\rangle$.

iv): Omdat $f(x)^2 \geq 0$ is het duidelijk dat $\langle f, f\rangle := \int_0^1 f(x)^2\,dx \geq 0$. We moeten nog inzien: als $\langle f, f\rangle = 0$, dan $f = 0$. Neem dus aan dat $\langle f, f\rangle = 0$ m.a.w. $\int_0^1 f(x)^2\,dx = 0$ en stel dat $f(x_0) \neq 0$ voor zekere $x_0 \in (0,1)$. Volgens de continuïteit van $f(x)$ bestaat dan $\delta > 0$ en $\varepsilon > 0$ zdd $|f(x)| \geq \varepsilon$ voor alle $x \in [x_0 - \delta, x_0 + \delta]$. Dan

$$0 = \int_0^1 f(x)^2\,dx \geq \int_{x_0-\delta}^{x_0+\delta} f(x)^2\,dx \geq \int_{x_0-\delta}^{x_0+\delta} \varepsilon^2\,dx = 2\delta\varepsilon^2 > 0,$$

tegenspraak. Dus blijkbaar $f(x_0) = 0$ voor alle $x_0 \in (0,1)$ en dus $f = 0$ op $[0,1]$ vanwege de continuïteit van f.

Opgave 3.2.5 Zij $V := \mathrm{Pol}(n)$ (zie voorbeeld 1.2.7) en $p = a_0 + a_1x + \cdots + a_nx^n$ en $q = b_0 + b_1x + \cdots + b_nx^n$ in V. Laat zien dat

$$\langle p, q\rangle := a_0b_0 + a_1b_1 + \cdots + a_nb_n$$

een inproduct op V definieert.

Opgave 3.2.6 Zij A een $n \times n$ matrix. Het *spoor* van A, genoteerd $\mathrm{tr}\, A$ (voor *trace* A), is per definitie de som der diagonaal elementen van A. Laat zien dat voor $n \times n$ matrices geldt

i) $\mathrm{tr}\,(A + B) = \mathrm{tr}\, A + \mathrm{tr}\, B$.
ii) $\mathrm{tr}\,(cA) = c.\mathrm{tr}\, A$, voor alle c in $\mathbb{R}$.
iii) $\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)$.

Opgave 3.2.7 Zij $V := M_n(\mathbb{R})$. Definieer

$$\langle A, B\rangle := \mathrm{tr}\,(A^tB), \text{ voor alle } A, B \in V.$$

Laat zien dat $\langle\ ,\ \rangle$ een inproduct op V definieert.

Terugkerend naar formule (3.1) zien we dat $|\langle v, w\rangle| \leq \|v\|\,\|w\|$ voor alle v, w in $\mathbb{R}^3$ (want $|\cos\alpha| \leq 1$). Het verrassende is dat zo'n ongelijkheid geldt voor *ieder inproduct*!

Stelling 3.2.8 (Ongelijkheid van Cauchy-Bunyakovsky-Schwarz.)
Zij V een vectorruimte met inproduct $\langle\ ,\ \rangle$. Dan geldt

$$|\langle v, w\rangle| \leq \|v\|\,\|w\| \text{ of te wel } \langle v, w\rangle^2 \leq \langle v, v\rangle \cdot \langle w, w\rangle \text{ voor alle } v, w \in V.$$

Gelijkheid treedt op d.e.s.d.a. v en w afhankelijk zijn.

Bewijs. i) Voor iedere $t \in \mathbb{R}$ geldt $\langle v + tw, v + tw\rangle \geq 0$ en dus

$$\langle v, v\rangle + 2t\langle v, w\rangle + t^2\langle w, w\rangle \geq 0 \text{ voor alle } t \in \mathbb{R}. \tag{3.2}$$

Als $\langle w, w\rangle = 0$, dan $w = 0$ en dus $\langle v, w\rangle = 0$ en dus geldt de ongelijkheid. We mogen dus aannemen dat $\langle w, w\rangle > 0$. Kies dan voor t de waarde $t_0 := -\frac{\langle v,w\rangle}{\langle w,w\rangle}$. Invullen in (3.2) geeft

$$\langle v + t_0w, v + t_0w\rangle = \langle v, v\rangle - 2\frac{\langle v, w\rangle}{\langle w, w\rangle}\langle v, w\rangle + \frac{\langle v, w\rangle^2}{\langle w, w\rangle^2}\langle w, w\rangle = \langle v, v\rangle - \frac{\langle v, w\rangle^2}{\langle w, w\rangle} \geq 0. \tag{3.3}$$

Dus $\langle v, w\rangle^2 \leq \langle v, v\rangle \cdot \langle w, w\rangle$. Trek dan wortels.
ii) Uit (3.3) volgt dat we gelijkheid hebben d.e.s.d.a. $\langle v+t_0w, v+t_0w\rangle = 0$ d.e.s.d.a. $v+t_0w = 0$. Hieruit volgt de stelling. $\square$

Het mooie van 3.2.8 is dat we hem kunnen toepassen op allerlei inproductruimten. Bijvoorbeeld op die van 3.2.4. Dit levert:

Gevolg 3.2.9 Zij $f, g \in C([0,1])$. Dan geldt de volgende vaak gebruikte ongelijkheid voor integralen:

$$\left|\int_0^1 f(x)g(x)dx\right|^2 \leq \int_0^1 f(x)^2dx \cdot \int_0^1 g(x)^2dx.$$

Een ander gevolg van 3.2.8 is de zgn. *driehoeksongelijkheid*.

Stelling 3.2.10 Zij V een inproductruimte en v, w in V. Dan geldt:

$$\|v + w\| \leq \|v\| + \|w\|.$$

Bewijs. Volgens 3.2.8 is $\langle v, w\rangle \leq |\langle v, w\rangle| \leq \|v\| \, \|w\|$. Dus

$$\begin{aligned}\|v + w\|^2 &= \langle v + w, v + w\rangle = \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle \\ &\leq \|v\|^2 + 2\|v\| \, \|w\| + \|w\|^2 = (\|v\| + \|w\|)^2.\end{aligned}$$

Trek dan wortels. □

Ook kunnen we 3.2.8 gebruiken om de hoek tussen twee vectoren ($\neq 0$) in een abstracte inproductruimte te definiëren. Daartoe merk op dat als v en w in V beiden niet nul zijn dan (volgens 3.2.8)

$$-1 \leq \frac{\langle v, w\rangle}{\|v\| \, \|w\|} \leq 1.$$

Dus bestaat er precies één θ met $0 \leq \theta \leq \pi$ zdd

$$\cos\theta = \frac{\langle v, w\rangle}{\|v\| \, \|w\|}.$$

Deze hoek θ heet *de hoek tussen* v *en* w.

Definitie 3.2.11 We noemen twee willekeurige vectoren v en w in een inproductruimte V *orthogonaal*, notatie $v \perp w$, als $\langle v, w\rangle = 0$. We zeggen dan ook dat v en w *loodrecht op elkaar staan*.

Definitie 3.2.12 Zij $v_1, \ldots, v_n$ in V. Dan heet $(v_1, \ldots, v_n)$ een *orthogonaal* stelsel in V als iedere $v_i \neq 0$ en $v_i \perp v_j$ voor alle $i \neq j$. Als bovendien iedere v_i lengte 1 heeft m.a.w. $\|v_i\| = 1$ voor alle i, dan heet $(v_1, \ldots, v_n)$ een *orthonormaal* stelsel in V.

Voorbeeld 3.2.13 Laat zien dat in $C([0, 1])$ de vectoren $v := x$ en $w := 2 - 3x$ loodrecht op elkaar staan.

Oplossing. $\langle v, w\rangle = \int_0^1 x(2 - 3x)\, dx = \int_0^1 2x\, dx - \int_0^1 3x^2\, dx = x^2\big|_0^1 - x^3\big|_0^1 = 1 - 1 = 0.$

Voorbeeld 3.2.14 In $\mathbb{R}^n$ is $(e_1, \ldots, e_n)$ een orthonormaal stelsel t.o.v. het standaard inproduct.

Een van de redenen waarom een orthogonaal stelsel handig is, is dat we de coördinaten van een willekeurige vector m.b.t. een orthogonaal stelsel rechtstreeks m.b.v. het inproduct kunnen bepalen.

Stelling 3.2.15 Zij $(v_1, \ldots, v_n)$ een orthogonaal stelsel in V. Voor $v = c_1 v_1 + \cdots + c_n v_n \in V$ geldt dan $c_i = \dfrac{\langle v, v_i\rangle}{\langle v_i, v_i}\rangle$, voor alle i.

Bewijs. $\langle v, v_i\rangle = \langle c_1 v_1 + \cdots + c_n v_n, v_i\rangle = c_1\langle v_1, v_i\rangle + \cdots + c_i\langle v_i, v_i\rangle + \cdots + c_n\langle v_n, v_i\rangle = c_i\langle v_i, v_i\rangle$. Omdat $v_i \neq 0$ is ook $\langle v_i, v_i\rangle \neq 0$ en dus $c_i = \frac{\langle v, v_i\rangle}{\langle v_i, v_i\rangle}$. □

Gevolg 3.2.16 Als $(v_1, \ldots, v_n)$ een orthogonaal stelsel is, is het onafhankelijk.

Bewijs. Pas 3.2.15 toe met $v = 0$. $\square$

Gevolg 3.2.17 Zij $(v_1, \ldots, v_n)$ een orthogonaal stelsel in V.
i) Als $\dim V = n$, is het een basis van V.
ii) Als $(v_1, \ldots, v_n)$ volledig is in V, is het een basis van V.

Bewijs. ii) volgt uit 3.2.16 en i) uit 3.2.16 en 1.4.4. $\square$

Voorbeeld 3.2.18 In $\mathbb{R}^3$ vormen $v_1 := \begin{pmatrix}1\\1\\0\end{pmatrix}$, $v_2 := \begin{pmatrix}1\\-1\\0\end{pmatrix}$ en $v_3 := \begin{pmatrix}0\\0\\1\end{pmatrix}$ een orthogonaal stelsel t.o.v. het standaard inproduct (ga na!). Dus is volgens 3.2.17 i) (v_1, v_2, v_3) een basis van $\mathbb{R}^3$.
Zij $v := \begin{pmatrix}2\\3\\7\end{pmatrix}$. Bepaal $c_1, c_2, c_3 \in \mathbb{R}$ met $v = c_1v_1 + v_2c_2 + c_3v_3$.

Oplossing. Volgens 3.2.15 is $c_1 = \frac{\langle v,v_1\rangle}{\langle v_1,v_1\rangle} = \frac{2.1+3.1+7.0}{1.1+1.1+0.0} = \frac{5}{2}$. Net zo vinden we $c_2 = -\frac{1}{2}$ en $c_3 = 7$.

We geven nu een algoritme om uit een gegeven basis $(v_1, \ldots, v_n)$ van V een orthogonale basis te maken:

Stelling 3.2.19 (Gram-Schmidt orthogonalisatie algoritme.)
Zij V een inproductruimte met basis $(v_1, \ldots, v_n)$. Definiëer inductief $e_1, \ldots, e_n$ in V d.m.v.

$$e_1 := v_1, \quad e_2 := v_2 - \frac{\langle v_2, e_1\rangle}{\langle e_1, e_1\rangle} e_1, \quad \ldots$$
$$e_k := v_k - \frac{\langle v_k, e_1\rangle}{\langle e_1, e_1\rangle} e_1 - \ldots - \frac{\langle v_k, e_{k-1}\rangle}{\langle e_{k-1}, e_{k-1}\rangle} e_{k-1} = v_k - \sum_{j=1}^{k-1} \frac{\langle v_k, e_j\rangle}{\langle e_j, e_j\rangle} e_j.$$

Dan is voor iedere $1 \leq k \leq n$ het stelsel $(e_1, \ldots, e_k)$ orthogonaal. I.h.b. is $(e_1, \ldots, e_n)$ een orthogonale basis van V.

Bewijs. Met inductie naar k. Duidelijk als $k = 1$. Neem dus aan $k \geq 2$ en dat de stelling al voor $k - 1$ bewezen is, m.a.w. $(e_1, \ldots, e_{k-1})$ is orthogonaal. We moeten dan nog inzien dat $\langle e_k, e_i\rangle = 0$ voor alle $1 \leq i \leq k - 1$. Er geldt

$$\langle e_k, e_i\rangle = \langle v_k - \sum_{j=1}^{k-1} \frac{\langle v_k, e_j\rangle}{\langle e_j, e_j\rangle} e_j, e_i\rangle = \langle v_k, e_i\rangle - \sum_{j=1}^{k-1} \frac{\langle v_k, e_j\rangle}{\langle e_j, e_j\rangle}\langle e_j, e_i\rangle.$$

Omdat $\langle e_j, e_i\rangle = 0$ als $1 \leq i \neq j \leq k - 1$ volgt

$$\langle e_k, e_i\rangle = \langle v_k, e_i\rangle - \frac{\langle v_k, e_i\rangle}{\langle e_i, e_i\rangle}\langle e_i, e_i\rangle = 0.$$

Dus is $(e_1, \ldots, e_k)$ orthogonaal voor alle $1 \leq k \leq n$. I.h.b. is $(e_1, \ldots, e_n)$ orthogonaal en dus een basis vanwege 3.2.17 i). $\square$

Opmerking 3.2.20 Het meetkundige idee achter de Gram-Schmidt orthogonalisatie is vrij simpel. Voor een vector v en een orthogonaal stelsel $(e_1, \ldots, e_{k-1})$ is $v' = \sum_{j=1}^{k-1} \frac{\langle v, e_j \rangle}{\langle e_j, e_j \rangle} e_j$ de orthogonale projectie van v in de deelruimte U opgespannen door $e_1, \ldots, e_{k-1}$ (zie 3.2.27 hieronder). De verschilvector $e_k = v - v'$ staat dan loodrecht op alle vectoren in de deelruimte U, dus i.h.b. op $e_1, \ldots, e_{k-1}$. Omdat $v = e_k + v'$ en v' in U ligt, is $\langle e_1, \ldots, e_{k-1}, v \rangle = \langle e_1, \ldots, e_{k-1}, e_k \rangle$.

Voorbeeld 3.2.21 Zij $P_n[-1, 1]$ de verzameling der veeltermfuncties van graad $\leq n$ op het interval $[-1, 1]$. Dit is een lineaire deelruimte van $C([-1, 1])$, de vectorruimte der continue reëelwaardige functies op $[-1, 1]$. Op $P_n[-1, 1]$ nemen we het inproduct $\langle f, g \rangle := \int_{-1}^{1} f(x)g(x)\,dx$. Bepaal een orthogonale basis van $P_2[-1, 1]$.

Oplossing. $(1, x, x^2)$ is een basis van $P_2[-1, 1]$. Pas nu het Gram-Schmidt proces toe: $e_1 := 1$, $e_2 := x - \frac{\langle x, 1 \rangle}{\langle 1, 1 \rangle} 1$. Omdat $\langle x, 1 \rangle = \int_{-1}^{1} x\,dx = 0$ volgt $e_2 = x$. Tenslotte

$$e_3 = x^2 - \frac{\langle x^2, 1 \rangle}{\langle 1, 1 \rangle} 1 - \frac{\langle x^2, x \rangle}{\langle x, x \rangle} x.$$

Omdat $\langle x^2, x \rangle = \int_{-1}^{1} x^3\,dx = 0$ en $\langle x^2, 1 \rangle = \frac{2}{3}$ en $\langle 1, 1 \rangle = 2$ volgt dat $e_3 = x^2 - \frac{1}{3}$. Dus $(1, x, x^2 - \frac{1}{3})$ is een orthogonale basis van $P_2[-1, 1]$.

Deze polynomen heten de *Legendre polynomen*. Eenzelfde verhaal voor $n = 5$ geeft de eerste zes Legendre polynomen

$$1,\ x,\ x^2 - \frac{1}{3},\ x^3 - \frac{3}{5}x,\ x^4 - \frac{6}{7}x^2 + \frac{3}{35},\ x^5 - \frac{10}{9}x^3 + \frac{5}{21}x.$$

Definitie 3.2.22 Zij U een lineaire deelruimte van een inproductruimte V. Het *orthogonale complement van* U (in V), notatie $U^\perp$, is gedefinieerd d.m.v.

$$U^\perp := \{v \in V \mid \langle v, u \rangle = 0 \text{ voor alle } u \in U\}.$$

Opgave 3.2.23 i) Laat zien dat $U^\perp$ een lineaire deelruimte van V is.
ii) Stel dat $(u_1, \ldots, u_m)$ een volledig stelsel voor V is. Laat zien dat

$$U^\perp = \{v \in V \mid \langle v, u_1 \rangle = 0, \ldots, \langle v, u_m \rangle = 0\}.$$

Voorbeeld 3.2.24 Zij $U := \left\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle \subset \mathbb{R}^2$. Bereken $U^\perp$.

Oplossing. Volgens 3.2.23 geldt:

$$U^\perp = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \mid \langle \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \rangle = 0 \right\} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 0 \right\}$$

m.a.w. $U^\perp$ is het vlak $x_1 + x_2 + x_3 = 0$.

Opgave 3.2.25 Zij A een $m \times n$ matrix. Laat zien dat in $\mathbb{R}^m$ het volgende verband geldt tussen de kolommenruimte $\mathrm{Col}(A)$ (zie sectie 2.3) en de nulruimte $\mathrm{Nul}(A^t)$ (zie 1.2.14) van de getransponeerde matrix: $\mathrm{Col}(A)^\perp = \mathrm{Nul}(A^t)$.

De volgende stelling is een zeer nuttige generalisatie van het feit dat iedere vector uit $\mathbb{R}^2$ eenduidig te schrijven is als een vector uit $U := \langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle$ en een vector uit $U^\perp = \langle \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle$.

Stelling 3.2.26 Zij U een eindig dimensionale lineaire deelruimte van een inproductruimte V. Dan is iedere $v \in V$ ééndúidig te schrijven als $v = v_1 + v_2$ met $v_1 \in U$ en $v_2 \in U^\perp$.

Bewijs. i) Kies m.b.v. 3.2.19 een orthogonale basis $(e_1, \ldots, e_m)$ van U. Zij $v \in V$ en definiëer

$$v_1 := \frac{\langle v, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 + \cdots + \frac{\langle v, e_m \rangle}{\langle e_m, e_m \rangle} e_m \quad \text{en} \quad v_2 := v - v_1.$$

Dan is duidelijk dat v_1 in U en $v = v_1 + v_2$. Om in te zien dat $v_2 \in U^\perp$ is het volgens 3.2.23 voldoende te bewijzen dat $\langle v_2, e_i \rangle = 0$ voor alle $1 \leq i \leq m$. Zij daartoe $1 \leq i \leq m$. Dan $\langle v_2, e_i \rangle = \langle v, e_i \rangle - \langle v_1, e_i \rangle$. Omdat $\langle e_j, e_i \rangle = 0$ als $j \neq i$ volgt

$$\langle v_1, e_i \rangle = \frac{\langle v, e_i \rangle}{\langle e_i, e_i \rangle} \langle e_i, e_i \rangle = \langle v, e_i \rangle.$$

Dus

$$\langle v_2, e_i \rangle = \langle v, e_i \rangle - \langle v_1, e_i \rangle = 0.$$

ii) Rest ons nog de eenduidigheid te bewijzen. Stel daarom dat ook $v = v_1' + v_2'$ met $v_1' \in U$ en $v_2' \in U^\perp$. Dan $w := v_1 - v_1' = v_2' - v_2$. Omdat $v_1, v_1' \in U$ volgt $w \in U$. Net zo $w \in U^\perp$, omdat $w_2, w_2' \in U^\perp$. Dus $\langle w, u \rangle = 0$ voor alle $u \in U$. I.h.b. voor $u = w$, m.a.w. $\langle w, w \rangle = 0$. Dus $w = 0$ en dus $v_1 = v_1'$ en $v_2 = v_2'$. $\square$

Definitie 3.2.27 Zij U een eindig dimensionale lineaire deelruimte van een inproductruimte V en zij $v = v_1 + v_2 \in V$ met $v_1 \in U$ en $v_2 \in U^\perp$. Dan heet de (eenduidig bepaalde) vector $v_1 \in U$ de *projectie van v op U*, notatie $\mathrm{proj}_U(v)$. De vector $v_2 \in U^\perp$ heet *de component van v loodrecht op U*.

Uit bovenstaand bewijs zien we: als $(e_1, \ldots, e_m)$ een orthogonale basis van U is dan is

$$\mathrm{proj}_U(v) = \frac{\langle v, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 + \ldots + \frac{\langle v, e_m \rangle}{\langle e_m, e_m \rangle} e_m.$$

en $v_2 = v - \mathrm{proj}_U(v)$. Het belang van de projectie van v op U komt voort uit de volgende (zeer nuttige) stelling:

Stelling 3.2.28 Zij U een eindig dimensionale lineaire deelruimte van een inproductruimte V en zij $v \in V$. Dan is $\mathrm{proj}_U(v)$ die vector in U die de kleinste afstand tot v heeft, d.w.z.

$$\|v - \mathrm{proj}_U(v)\| \leq \|v - u\|, \text{ voor alle } u \in U.$$

Bewijs. Zij $u \in U$ en $v_1 := \mathrm{proj}_U(v)$. Dan $v - v_1 \in U^\perp$ en $v_1 - u \in U$. Schrijf nu $v - u = (v - v_1) + (v_1 - u)$. Dan volgt wegens $\langle v - v_1, v_1 - u \rangle = 0$:

$$\begin{aligned} \|v - u\|^2 &= \langle (v - v_1) + (v_1 - u), (v - v_1) + (v_1 - u) \rangle = \langle v - v_1, v - v_1 \rangle + \langle v_1 - u, v_1 - u \rangle \\ &= \|v - v_1\|^2 + \|v_1 - u\|^2. \end{aligned}$$

Dus i.h.b. $\|v - u\|^2 \geq \|v - v_1\|^2$, waarmee de stelling bewezen is. $\square$

Voorbeeld 3.2.29 Zij U de lineaire deelruimte van $\mathbb{R}^4$ gegeven door de vergelijkingen

$$x_1 + x_2 + x_3 = 0 \quad \text{en} \quad 2x_2 + x_3 + x_4 = 0.$$

Bepaal de projectie van $v = \begin{pmatrix} 5 \\ 6 \\ 7 \\ 2 \end{pmatrix}$ op U en de kleinste afstand van v tot U.

Oplossing. Het stelsel vergelijkingen oplossend vinden we $x_2 = -\frac{1}{2}x_3 - \frac{1}{2}x_4$ en $x_1 = -x_2 - x_3 = -\frac{1}{2}x_3 + \frac{1}{2}x_4$. Dit geeft

$$U = \left\{ \begin{pmatrix} -\frac{1}{2}x_3 + \frac{1}{2}x_4 \\ -\frac{1}{2}x_3 - \frac{1}{2}x_4 \\ x_3 \\ x_4 \end{pmatrix} \mid x_3, x_4 \in \mathbb{R} \right\} = \left\langle e_1 := \begin{pmatrix} -1 \\ -1 \\ 2 \\ 0 \end{pmatrix}, \; e_2 := \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix} \right\rangle.$$

Merk op dat e_1 en e_2 orthogonaal zijn. Dus

$$\begin{aligned} \operatorname{proj}_U(v) &= \frac{\langle v, e_1 \rangle}{\langle e_1, e_1 \rangle} e_1 + \frac{\langle v, e_2 \rangle}{\langle e_2, e_2 \rangle} e_2 \\ &= \frac{5.(-1) + 6.(-1) + 7.2 + 2.0}{(-1).(-1) + (-1).(-1) + 2.2 + 0.0} e_1 + \frac{5.1 + 6.(-1) + 7.0 + 2.2}{1.1 + (-1).(-1) + 0.0 + 2.2} e_2 \\ &= \frac{1}{2}e_1 + \frac{1}{2}e_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 1 \end{pmatrix}. \end{aligned}$$

De kortste afstand van v tot U is volgens 3.2.28:

$$\|v - \operatorname{proj}_U(v)\| = \left\| \begin{pmatrix} 5 \\ 6 \\ 7 \\ 2 \end{pmatrix} - \begin{pmatrix} 0 \\ -1 \\ 1 \\ 1 \end{pmatrix} \right\| = \sqrt{25 + 49 + 36 + 1} = \sqrt{111}.$$

3.3 Toepassingen

a) De methode der kleinste kwadraten

In veel fysische situaties is het gebruikelijk dat een stelsel $Ax = b$ om theoretische redenen een oplossing moet hebben, maar dat het in "werkelijkheid" geen oplossing heeft vanwege meetfouten die de matrix elementen van A en b verstoren. In zulke situaties zoeken we naar een x die een zo goed mogelijke oplossing is in de zin dat $\|Ax - b\|$ minimaal is ($\| \;\; \|$ is de gewone Euclidische lengte).

Probleem der klassieke kwadraten.
Gegeven een lineair systeem $Ax = b$ van m vergelijkingen met n onbekenden. Vind een vector $x \in \mathbb{R}^n$ zó dat $\|Ax - b\|$ minimaal is. Zo'n x heet een *kleinste kwadraten oplossing van* $Ax = b$.

Opmerking 3.3.1 De naam " kleinste kwadraten" zien we als volgt: zij $\varepsilon := Ax - b$ de "error" functie, zeg $\varepsilon = (\varepsilon_1 \, \ldots \, \varepsilon_m)^t$. Dan $\|\varepsilon\| = \sqrt{\varepsilon_1^2 + \cdots + \varepsilon_m^2}$ is minimaal d.e.s.d.a. $\|\varepsilon\|^2 = \varepsilon_1^2 + \cdots + \varepsilon_m^2$ minimaal. Dit verklaart de naam!

Om het kleinste kwadraten probleem op te lossen definiëren we

$$U := \{Ax \mid x \in \mathbb{R}^n\} \subset \mathbb{R}^m.$$

We zien dus dat U de lineaire deelruimte $\operatorname{Im} f_A = \operatorname{Col}(A)$ is (zie sectie 2.3).
We zoeken een $x \in \mathbb{R}^n$ zodat Ax dié vector uit U is die zo dicht mogelijk bij b ligt. Volgens 3.2.28 moet dan $Ax = \operatorname{proj}_U b$ zijn!. M.a.w. we moeten x oplossen uit

$$Ax = \operatorname{proj}_U b.$$

Lemma 3.3.2 Er geldt $Ax = \operatorname{proj}_U b$ d.e.s.d.a. $A^t Ax = A^t b$.

Bewijs. Merk op dat $U^\perp = \operatorname{Col}(A)^\perp = \operatorname{Nul}(A^t)$ (3.2.25).
$\Rightarrow$: Stel $Ax = \operatorname{proj}_U b$. Dan $b - Ax = b - \operatorname{proj}_U b \in U^\perp = \operatorname{Nul}(A^t)$, dus $A^t(b - Ax) = 0$ m.a.w. $A^t Ax = A^t b$.
$\Leftarrow$: Omgekeerd, stel $A^t(Ax - b) = 0$. Dan $Ax - b \in \operatorname{Nul}(A^t) = U^\perp$. Ook $b - \operatorname{proj}_U b \in U^\perp$. Dus $Ax - \operatorname{proj}_U b \in U^\perp$. Per definitie zitten Ax en $\operatorname{proj}_U b$ in U, dus $Ax - \operatorname{proj}_U b \in U$. Totaal $Ax - \operatorname{proj}_U b \in U \cap U^\perp = \{0\}$. Dus $Ax = \operatorname{proj}_U b$. □

Gevolg 3.3.3 Als de kolommen van A lineair onafhankelijk zijn, dan is een kleinste kwadraten oplossing uniek.

Bewijs. Uit 3.3.2 volgt dat x een kleinste kwadraten oplossing is d.e.s.d.a. $A^t Ax = A^t b$. We laten nu zien dat $A^t A$ inverteerbaar is (waaruit dan de bewering volgt). Volgens 2.3.7 is het voldoende te bewijzen dat uit $A^t Ax = 0$ volgt dat $x = 0$. Neem dus aan dat $A^t Ax = 0$, m.a.w. $A^t(Ax) = 0$. Dus $Ax \in \operatorname{Nul}(A^t) = U^\perp$. Ook $Ax \in U$. Dus $Ax \in U \cap U^\perp = \{0\}$, m.a.w. $x_1 A_1 + \cdots + x_n A_n = 0$, waarbij A_i de i-de kolom van A is. Uit het gegeven volgt dan dat $x_1 = \ldots = x_n = 0$ m.a.w. $x = 0$. □

Voorbeeld 3.3.4 Gegeven het stelsel $Ax = b$ d.m.v.

$$\begin{aligned} x_1 - x_2 &= 4 \\ 3x_1 + 2x_2 &= 1 \\ -2x_1 + 4x_2 &= 3 \end{aligned}$$

Laat zien dat er precies een kleinste kwadraten oplossing bestaat en vind die.

Oplossing. Hier $A = \begin{pmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{pmatrix}$ en $b = \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix}$.
Men ziet meteen dat de kolommen van A onafhankelijk zijn. Dus is er precies één kleinste kwadraten oplossing van $Ax = b$ (3.3.3). Om deze te bepalen moeten we volgens 3.3.2 x oplossen uit $A^t Ax = A^t b$. Uitrekenen van $A^t A$ en $A^t b$ levert het stelsel

$$\begin{pmatrix} -14 & -3 \\ -3 & 21 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 10 \end{pmatrix}.$$

Hieruit volgt dan eenvoudig: $(x_1, x_2) = \left(\frac{17}{95}, \frac{143}{285}\right)$.

b) Hoofdcomponenten analyse

Bij statistische evaluaties kijkt men vaak naar het gemiddelde van de metingen en naar de spreiding van de metingen rond het gemiddelde. Als de metingen $x_1, \ldots, x_m$ zijn, dan is het gemiddelde (in de kansrekening *verwachtingswaarde* geheten) $\mu = \frac{1}{m}\sum_{i=1}^m x_i$ en de variantie is $\sigma^2 = \frac{1}{m}\sum_{i=1}^m (x_i - \mu)^2$, d.w.z. het gemiddelde van de kwadratische afwijkingen van de verwachtingswaarde.

Als we nu een experiment hebben met verschillende parameters, dan zijn de metingen vectoren v_i met een zeker aantal componenten. Bij drie parameters is een meting bijvoorbeeld een vector van de vorm $v_i = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}$. Het gemiddelde vinden we weer door de metingen bij elkaar op te tellen en door het aantal metingen te delen, dit geeft bij drie parameters de vector

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix} = \frac{1}{m}\sum_{i=1}^m v_i.$$

Het is nu niet meer vanzelfsprekend dat de grootste spreiding van de metingen in één van de richtingen der parameters ligt, want misschien is er een combinatie van parameters, die het grootste effect op het experiment heeft. We willen dus graag de richting bepalen waarin de spreiding maximaal wordt.

Dit legt het verband met de orthogonale projecties want als we het over de spreiding in een bepaalde richting hebben, praten we over de spreiding van de orthogonale projecties op een lijn.

Stel we willen de spreiding in de richting van een vector u met $\|u\| = 1$ bepalen. Volgens 3.2.26 is de orthogonale projectie van een vector v op de lijn langs u gegeven door $\mathrm{proj}_u(v) = \langle v, u\rangle u$ en wegens $\|u\| = 1$ is het kwadraat van de lengte van deze projectie $\langle v, u\rangle^2$.

Als we nu metingen $v_1, \ldots, v_m$ met een gemiddelde μ hebben, is de spreiding in de richting van u gegeven door

$$\sigma_u^2 = \frac{1}{m}\sum_{i=1}^m \langle v_i - \mu, u\rangle^2$$

Omdat we in dit geval altijd met het standaard inproduct op $\mathbb{R}^n$ werken, kunnen we $\langle v, u\rangle^2$ nog iets anders schrijven, namelijk

$$\langle v, u\rangle^2 = u^t \cdot v \cdot v^t \cdot u$$

Dit toegepast op σ_u^2 geeft

$$\sigma_u^2 = u^t \left(\frac{1}{m}\sum_{i=1}^m (v_i - \mu)\cdot(v_i - \mu)^t \right) u.$$

Opmerking 3.3.5 De uitdrukking

$$V := \frac{1}{m}\sum_{i=1}^m (v_i - \mu)\cdot(v_i - \mu)^t$$

is een $n \times n$ matrix die alleen maar van de metingen v_i afhangt. Deze matrix heet de *covariantiematrix* van de metingen $v_1, \ldots, v_n$. M.b.v. de covariantiematrix laat zich σ_u^2 makkelijk schrijven als $\sigma_u^2 = u^t V u$.

Gezocht is nu de vector u met $\|u\| = 1$ waarvoor $u^t V u$ maximaal is. Nu laat zich voor covariantiematrices bewijzen, dat ze altijd diagonaliseerbaar zijn, met alleen maar positieve reële eigenwaarden. Met behulp hiervan gaat men na dat de vector u waarvoor $u^t V u$ maximaal is de eigenvector voor de grootste eigenwaarde van V is.

Voorbeeld 3.3.6 Gegeven zijn de tien metingen

$$(x_1, y_1) = (-5, -2); \quad (x_2, y_2) = (-4, -3); \quad (x_3, y_3) = (-3, 0); \quad (x_4, y_4) = (-2, -1);$$
$$(x_5, y_5) = (-1, -1); \quad (x_6, y_6) = (1, 0); \quad (x_7, y_7) = (2, 3); \quad (x_8, y_8) = (3, 2);$$
$$(x_9, y_9) = (4, 0); \quad (x_{10}, y_{10}) = (5, 2).$$

Bepaal de richting van de grootste spreiding van deze metingen.

Oplossing. De punten zijn zo gekozen dat het gemiddelde $\mu = \frac{1}{10} \begin{pmatrix} \sum_{i=1}^{10} x_i \\ \sum_{i=1}^{10} y_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is. We moeten dus alleen maar de covariantiematrix bepalen, deze is

$$V = \frac{1}{10} \begin{pmatrix} \sum_{i=1}^{10} x_i^2 & \sum_{i=1}^{10} x_i y_i \\ \sum_{i=1}^{10} x_i y_i & \sum_{i=1}^{10} y_i^2 \end{pmatrix} = \begin{pmatrix} 11 & 3.2 \\ 3.2 & 4.7 \end{pmatrix}$$

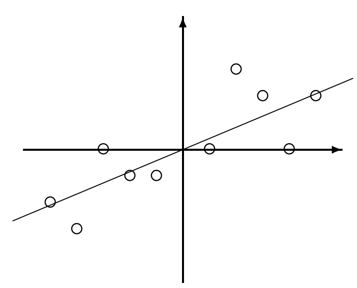
De eigenwaarden van V zijn

$$\lambda_1 = \frac{1}{20}(157 + \sqrt{8065}) \approx 12.34, \quad \lambda_2 = \frac{1}{20}(157 - \sqrt{8065}) \approx 3.36$$

en de eigenvector voor de grootste eigenwaarde λ_1 is

$$u = \begin{pmatrix} 64 \\ \sqrt{8065} - 63 \end{pmatrix} \approx \begin{pmatrix} 64 \\ 26.8 \end{pmatrix}$$

Het plaatje hieronder geeft de metingen en de lijn in de richting van de grootste spreiding aan. Merk op dat de lijn vanwege $\mu = 0$ door de oorsprong loopt.



Opmerking 3.3.7 In het algemeen wordt deze methode gebruikt om uit een groot aantal parameters de belangrijkste combinaties eruit te vissen. Deze combinaties heten de *hoofdcomponenten*, vandaar de naam *hoofdcomponenten analyse*. Men neemt dan bijvoorbeeld in een experiment met honderd parameters de eigenvectoren die bij de grootste tien eigenwaarden horen en heeft hiermee vaak de belangrijkste eigenschappen voor het experiment beschreven.

Historische opmerkingen

Een techniek die veel lijkt op de methode der kleinste kwadraten was ontwikkeld door Roger Cotes (1682-1716) in een werk (rond 1715) waarin hij afwijkingen in astronomische waarnemingen behandelde. Het complete principe werd voor het eerst door Gauss op 16-jarige leeftijd (1793) geformuleerd terwijl hij bezig was de verdeling der priemgetallen te bestuderen. Gauss zei later dat hij deze methode heel veel gebruikt had gedurende vele jaren bijvoorbeeld bij het berekenen van banen van astroïden. Hij publiceerde zijn methode in 1809 en gaf 14 jaar later een nog uitgebreidere uiteenzetting. Echter Adrien-Marie Legendre (1752-1833) was de eerste die de methode der kleinste kwadraten publiceerde in een werk in 1806 over de bepaling van komeetbanen. Legendre was niet echt blij met het feit dat Gauss in 1809 beweerde dat hij de methode als eerste had bedacht: tot 1827 bleef Legendre Gauss daarover schrijven.