Gram–Schmidt's algoritm för en allmän vinkel

Tord Sjödin

Institutionen för matematik och matematisk statistik
Umeå universitet,
tord.sjodin@math.umu.se

Inledning.

Låt V vara ett reellt vektorrum med inre produkt $x \cdot y$ och norm $||x|| = (x \cdot x)^{1/2}$, där x, y är vektorer i V. Vi säger att två vektorer x och y i V är ortogonala om $x \cdot y = 0$. En ändlig summa $\sum_{j=1}^{k} c_j \cdot x_j$, där c_j är reella tal och x_j är vektorer i V, $1 \leq j \leq k$, kallar vi en linjär kombination av vektorerna $\{x_1, x_2, \ldots, x_k\}$. En delmängd E av V säges vara linjärt oberoende om för varje ändlig delmängd $\{x_1, x_2, \ldots, x_k\}$ av E och godtyckliga reella tal $\{c_1, c_2, \ldots, c_k\}$, villkoret $\sum_{j=1}^{k} c_j \cdot x_j = 0$ medför att $c_j = 0$, för alla $1 \leq j \leq k$. Vi säger att mängden E är en parvis ortogonal mängd om $x \cdot y = 0$, för varje par x och y av olika vektorer i V. En parvis ortogonal mängd av enhetsvektorer, vektorer med norm ett, kallas en ortonormal mängd eller helt enkelt en $ON-$mängd. Det är välkänt, och lätt att bevisa, att varje parvis ortogonal mängd också är linjärt oberoende. Omvändningen till detta påstående är naturligtvis inte sant. Det som däremot är sant är att givet en ändlig eller numrerbart oändlig och linjärt oberoende mängd E, så kan man konstruera en $ON-$mängd E' som har samma linjära hölje som E. Detta är i huvudsak innehållet i Gram-Schmidt's algoritm och fullt tillräckligt för de flesta tillämpningar. Att två vektorer x och y är ortogonala kan i vissa fall tolkas geometriskt som att vinkeln mellan x och y är 90 grader. Vi ska ge en mera allmänn algoritm, som vi kallar Gram–Schmidt's $p-$algoritm, och som konstruerar en mängd vektorer som parvis bildar samma vinkel med varandra i samtliga fall där detta är möjligt.

Det linjära höljet till en mängd $E \subset V$ är mängden av alla linjärkombinationer av vektorer i E och betecknas med $span(E)$. Det är lätt att se att $span(E)$ är ett linjärt delrum av V, det vill säga att $span(E)$ är ett vektorrum med samma operationer som V.

Vi definierar vinkeln θ mellan två vektorer x and y i V, som inte är nollvektorer, genom

$$\cos\theta = \frac{x \cdot y}{||x|| \cdot ||y||}, \quad 0 \leq \theta \leq \pi.$$

Om $x = 0$ eller $y = 0$ definierar vi vinkeln θ som noll. Två vektorer x och y är alltså ortogonala om och endast om vinkeln mellan vektorerna är 90 grader. Det är nu vanligt, och naturligt, att man identifierar vinkeln θ med dess *cosinusvärde* p, det vill säga att man definierar $p = \cos\theta$, där $-1 \leq p \leq 1$.

Gram – Schmidt's algoritm konstruerar en ortogonal mängd av vektorer från en linjärt oberoende mängd. Vi formulerar detta resultat på följande sätt, se exempelvis [1], Ch. 6, [4], Theorem 31, [2] eller [3].

Sats 1. *Låt $E = \{x_i\}$ vara en ändlig, eller numrerbart oändlig, och linjärt oberoende delmängd av V. Då finns en mängd $E' = \{y_i\}$ av vektorer i V med följande egenskaper:*

$$span(\{y_1, y_2, \ldots, y_i\}) = span(\{x_1, x_2, \ldots, x_i\}), \; \textit{för } i = 1, 2, \ldots, \tag{1}$$

$$span(E') = span(E), \tag{2}$$

$$\textit{mängden } E' \textit{ är en } ON\textit{-mängd.} \tag{3}$$

Följande sats och dess korollarium är vårt huvudresultat.

Sats 2. *Låt $\{x_i\}_{i=1}^n$ vara en linjärt oberoende delmängd av V. Då finns en mängd vektorer $\{z_i\}_{n=1}^n$ i V så att*

$$span(\{z_1, z_2, \ldots, z_i\}) = span(\{x_1, x_2, \ldots, x_i\}), \quad 1 \leq i \leq n, \tag{4}$$

och

$$z_i \cdot z_j = \begin{cases} 1, & i = j, \\ p, & i \neq j, \end{cases} \tag{5}$$

för alla $1 \leq i, j \leq n$, om och endast om $-\frac{1}{n-1} < p < 1$.

Korollarium 1. *Låt $\{x_i\}_{n=1}^\infty$ vara en numrerbart oändlig och linjärt oberoende mängd vektorer i V. Då finns en mängd vektorer $\{z_i\}_{n=1}^\infty$ i V så att (4) och (5) gäller för alla $i, j = 1, 2, \ldots,$ om och endast om $0 \leq p < 1$.*

Anmärkning. Sats 2 and Korollarium 1 kan också formuleras i termer av längden av vektorerna istället för vinklarna mellan vektorerna, ty om $||x|| = ||y|| = 1$ så gäller $||x - y||^2 = 2(1 - x \cdot y)$. Sats 2 får då följande ekvivalenta formulering:

Låt $\{x_i\}_{i=1}^n$ vara en linjärt oberoende mängd i V. Då finns vektorer $\{z_i\}_{i=1}^n$ i V som uppfyller (4) och

$$||z_i|| = 1 \quad \textit{and} \quad ||z_i - z_j|| = d$$

för $i \neq j$, $1 \leq i, j \leq n$, om och endast om $0 < d < \sqrt{\frac{2n}{n-1}}$.

Korollarium 1 har en liknande omformulering med $0 < d \leq \sqrt{2}$, vilken vi överlämnar åt läsaren.

Anmärkning. Om vi väljer $d = 1$ ovan får vi en mängd $\{0, z_1, z_2, \ldots, z_n\}$ med $n + 1$ punkter som har parvis samma avstånd. Sådana mängder kallas ekvilaterala mängder och studeras för allmänna vektorrum i [5].

Bevis

Denna sektion innehåller bevisen för Sats 2 and Korollarium 1. Vi börjar med beviset av Sats 2.

Bevis av Sats 2. Låt $\{x_i\}_{i=1}^n$ vara som i satsens formulering, låt $-\frac{1}{n-1} < p < 1$ och utför Gram–Schmidt's algoritm på vektorerna $\{x_i\}_{i=1}^n$. Detta ger vektorer $\{y_i\}_{i=1}^n$ som uppfyller (1) och (3). Vi skall nu induktivt definiera vektorerna $\{z_i\}_{i=1}^n$ och vi börjar med $z_1 = y_1$. Låt sedan $w = y_2 + a \cdot z_1$. Då gäller $||w||^2 = 1 + a^2$ och $w \cdot z_1 = a$. Vi väljer nu a så att $w \cdot z_1 = p \cdot ||w||$, vilket ger $a = p/\sqrt{1-p^2}$. Slutligen definierar vi $z_2 = w/||w||$.

Antag nu att vi har definierat $z_1, z_2, \ldots, z_k$ så att $||z_i|| = 1$ och $z_i \cdot z_j = p$, $i \neq j$, $i, j = 1, 2, \ldots k$. Låt $w = y_{k+1} - \sum_{j=1}^{k} c_j \cdot z_j$. Då gäller

$$w \cdot z_i = -c_i - p \cdot \sum_{j \neq i} c_j, \quad 1 \leq i \leq k,$$

och

$$||w||^2 = 1 + ||\sum_{j=1}^{k} c_j \cdot z_j||^2,$$

eftersom y_{k+1} är ortogonal mot $span\{z_1, z_2, \ldots, z_k\} \subset span\{y_1, y_2, \ldots, y_k\}$. Detta ger

$$w \cdot z_i = -c_i - p \cdot \sum_{j \neq i} c_j = p \cdot ||w||,$$

for $1 \leq i \leq k$. Vi subtraherar nu de sista ekvationerna parvis och får att $c_1 = c_2 = \cdots = c_k = d$ och

$$d(1 + p(k-1)) = -p \cdot ||w||. \tag{6}$$

Vidare har vi

$$||w||^2 = 1 + d^2 \cdot ||\sum_{i=1}^{k} z_i||^2 = 1 + d^2(k + p(k^2 - k)). \tag{7}$$

Genom att kombinera (6) och (7) får vi

$$d^2 \cdot (1 + p(k-1)) \cdot (1-p) \cdot (1 + pk) = p^2.$$

Denna ekvation kan lösas med avseende på d, eftersom $p > -\frac{1}{n-1} \geq -\frac{1}{k}$. Slutligen definierar vi $z_{k+1} = w/||w||$. Genom att fortsätta på detta sätt får vi vektorer $\{z_i\}_{i=1}^n$ som uppfyller (5). Det är också lätt att se att $\{z_i\}_{i=1}^n$ är linjärt oberoende, vilket bevisar (4).

För att bevisa omvändningen av satsen antar vi att $\{z_1, z_2, \ldots, z_n\}$ är vektorer i V som uppfyller (4) och (5). Då är $\{z_1, z_2, \ldots, z_n\}$ linjärt oberoende och formeln

$0 < ||z_1 + z_2 + \cdots + z_n||^2 = n(1 + p(n-1))$ bevisar att $p > -1/(n-1)$. △

Bevis av Korollarium 1. Låt $\{x_i\}_{i=1}^{\infty}$ vara som i korollariet och låt $0 \le p < 1$. Beviset av Sats 2 ger oss nu vektorer $\{z_i\}_{i=1}^{\infty}$ med de önskade egenskaperna.

För att bevisa omvändningen av korollariet antar vi att $\{z_i\}_{i=1}^{\infty}$ har egenskaperna (4) och (5). Då är $\{z_i\}_{i=1}^{\infty}$ linjärt oberoende och $0 < ||z_1 + z_2 + \cdots z_n||^2 = n(1 + p(n-1))$, $n \ge 1$. Låter vi nu $n \to \infty$ så följer att $p \ge 0$. △

Gram–Schmidt's p-algoritm

I den här sektionen skall vi beskriva den algoritm som konstruerar vektorerna $\{z_i\}_{i=1}^{n}$ i Sats 2. Låt $\{y_i\}_{i=1}^{n}$ vara en $ON-$mängd i V och låt $-\frac{1}{n-1} < p < 1$. Vi definierar induktivt vektorer $\{z_i\}_{i=1}^{n}$ med metoden i beviset för Sats 2. Sätt $z_1 = y_1$ och antag att $z_2, z_2, \ldots z_k$ har definierats. Låt

$$w = \alpha \cdot y_{k+1} + \beta \cdot (z_1 + z_2 + \cdots + z_k),$$

där α och β uppfyller

$$||w||^2 = \alpha^2 + \beta^2(k + k(k-1)p) = 1$$

och

$$w \cdot z_i = \beta \cdot (1 + (k-1)p) = p \cdot ||w||,\ 1 \le i \le k.$$

En enkel uträkning ger att

$$\alpha^2 = \frac{(1-p)(1+pk)}{1+p(k-1)} \quad \text{och} \quad \beta = \frac{p}{1+p(k-1)}, \tag{8}$$

och vi får följande algoritm.

Gram–Schmidt's p-algoritm (GS_p).
Input är en linjärt oberoende mängd $\{x_1, x_2, \ldots, x_n\}$ i ett reellt vektorrum V med inre produkt och ett reellt tal p som uppfyller $-\frac{1}{n-1} < p < 1$. Output är en mängd $\{z_1, z_2, \ldots, z_n\}$ av vektorer i V med egenskaperna (4) and (5) i Sats 2.

Steg 0. Genomför Gram–Schmidt's algoritm enligt Sats 1 och få följden $\{y_1, y_2, \ldots, y_n\}$.

Steg 1. Definiera $z_1 = y_1$.

Steg 2. För $1 \le k \le n-1$, låt

$$z_{k+1} = \sqrt{\frac{(1-p)(1+pk)}{1+p(k-1)}} \cdot y_{k+1} + \frac{p}{1+p(k-1)} \cdot (z_1 + z_2 + \cdots + z_k). \tag{9}$$

Steg 3. Den mängd $\{z_1, z_2, \ldots, z_n\}$ som har konstruerats efter det att *Steg 2* har upprepats $(n-1)$ gånger har alla de önskade egenskaperna (4) och (5).

Vi ska nu låta V vara det Euklidiska vektorerummet R^n, vars element är $n-$tupler $x = (x_1, x_2, \dots, x_n)$ av reella tal. Skalärprodukten mellan två vektorer x och w definieras av $x \cdot w = x_1 w_1 + x_2 w_2 + \cdots + x_n w_n$ och normen av vektorn x ges av $||x|| = \sqrt{x \cdot x}$. Vi låter $\{y_k\}_1^n$ vara standardbasen i R^n, det vill säga $y_k = (0, \dots, 0, 1, 0, \dots, 0)$ och ettan är på den k:e platsen. Vi ska nu härleda explicita uttryck för vektorerna z_k. Det visar sig då att de $(k-1)$ första koordinaterna i z_k inte ändras när vi går från z_k till z_{k+1}. Vi bevisar det i följande lemma.

Lemma 1. *Låt $\{y_k\}_1^n$ var som ovan och definiera $\{z_k\}_1^n$ genom (9). Då gäller att $z_{k+1}(i) = z_k(i)$, för $1 \le i \le k-1$.*

Bevis. Om $k = 1$ finns inget att bevisa. Anta att lemmat gäller för något visst k, $2 \le k \le n-1$. Då gäller

$$z_{k+1}(i) = \frac{1}{1+p(k-1)} \cdot (z_1(i) + \cdots + z_{k-1}(i) + z_k(i))$$

och

$$z_k(i) = \frac{1}{1+p(k-2)} \cdot (z_1(i) + \cdots + z_{k-1}(i)).$$

En enkel uträkning ger oss slutsatsen i Lemma 1. △

Vi kan nu härleda explicita formler för z_k, $1 \le k \le n$ och börjar med att definiera följden

$$a_m(p) = \frac{p}{1+p(m-1)} \cdot \sqrt{\frac{(1-p)(1+p(m-1))}{1+p(m-2)}}, \tag{10}$$

m=1,2,.... Med hjälp av följden $\{a_m(p)\}_1^n$ får vi följande explicita formler för vektorerna $\{z_k\}_1^n$.

Sats 3. *Låt $\{y_k\}_1^n$ vara standardbasen i R^n och låt $\{z_k\}_1^n$ vara definierad av (GS_p). Då gäller $z_1 = (1, 0, \dots, 0)$ och z_k ges av*

$$z_k = (a_1(p), a_2(p), \dots, a_{k-1}(p), \frac{1+p(k-1)}{p} \cdot a_k(p), 0, \dots, 0),$$

för $2 \le k \le n$.

Bevis. Man ser lätt att satsen gäller för $k = 2$. Antag att satsen gäller ett visst k, $2 \le k \le n-1$ och betrakta z_{k+1}. Vi observerar att $z_{k+1}(i)$ har rätt värde för $1 \le i \le k-1$ och $i = k+1$, på grund av Lemma 1 och (9). Vidare får vi

$$z_{k+1}(k) = \frac{p}{1+p(k-1)} \cdot z_k(k) = a_k(p),$$

på grund av (9) och (10). △

En asymptotisk formel.

Återstoden av denna artikel skall vi ägna åt att undersöka det asymptotiska uppförandet hos z_k för stora värden på k. Vi ska använda oss av vektorrummet l^2, som består av alla oändliga följder $\{x_k\}_1^\infty$ sådana att summan $\sum_1^\infty x_k^2$ är ändlig. Skalärprodukten mellan två vektorer x och w i l^2 och normen av x definieras av $x \cdot w = \sum_1^\infty x_k w_k$ och $||x|| = \sqrt{x \cdot x}$, respektive. Vi antar att $0 < p < 1$ och identifierar R^n med ett delrum av l^2, genom att vektorn $x = (x_1, x_2, \dots, x_n)$ i R^n identifieras med vektorn $x = (x_1, x_2, \dots, x_n, 0, \dots)$ i l^2. Slutligen definierar vi följden

$$z_0 = (a_1(p), a_2(p), \dots, a_n(p), \dots).$$

Man bevisar lätt att z_0 tillhör l^2 och $||z_0|| = \sqrt{p}$. Vi kan nu använda z_0 till att bevisa en asymptotisk formel för z_k.

Sats 4. *Låt $\{z_k\}_1^\infty$ vara som ovan, då gäller $z_k \approx z_0 + \sqrt{1-p} \cdot y_k$, as $k \to \infty$, i den meningen att*

$$\overline{\lim_{k\to\infty}} \sqrt{k} \cdot ||z_k - z_0 - \sqrt{1-p} \cdot y_k|| \le \sqrt{p(1-p)} \le 1/2.$$

Bevis. Genom triangelolikheten får vi att $||z_k - z_0 - \sqrt{1-p} \cdot y_k||$ är högst

$$\left(\sum_k^\infty |a_m(p)|^2\right)^{1/2} + \left|\sqrt{\frac{(1-p)(1+p(k-1))}{1+p(k-2)}} - \sqrt{1-p}\right| = A + B.$$

För den första termen får vi formeln

$$A^2 = \sum_k^\infty |a_m(p)|^2 = \frac{p^2(1-p)}{1+p(k-2)}$$

med hjälp av en teleskopserie och en enkel uträkning ger uppskattningen

$$B \le \frac{1}{2} \cdot \frac{p\sqrt{1-p}}{1+p(k-2)}.$$

Sats 4 följer nu genom att kombinera uppskattningarna för A och B och ta övre limes då $k \to \infty$. △

Exempel. I fallet då $p = 1/2$, vilket svarar mot vinkeln $\pi/3$ radianer mellan vektorerna, blir formlerna särskilt enkla. En uträkning för vektorn $z_0 \in l^2$ i Sats 4 ger då att

$$z_0 = \frac{1}{\sqrt{2}} \cdot \left(\frac{1}{\sqrt{1\cdot 2}}, \frac{1}{\sqrt{2\cdot 3}}, \frac{1}{\sqrt{3\cdot 4}}, \frac{1}{\sqrt{4\cdot 5}}, \dots, \frac{1}{\sqrt{n\cdot(n+1)}}, \dots\right)$$

$$= \frac{1}{2} \cdot \left(\frac{1}{\sqrt{\binom{2}{2}}}, \frac{1}{\sqrt{\binom{3}{2}}}, \frac{1}{\sqrt{\binom{4}{2}}}, \frac{1}{\sqrt{\binom{5}{2}}}, \dots, \frac{1}{\sqrt{\binom{n}{2}}}, \dots\right).$$

Vi överlämnar åt läsaren att på samma sätt beräkna vektorerna $\{z_k\}_1^n$ i Sats 3.

Anmärkning. Jørgen Pedersen Gram, 1850 – 1916, var en dansk matematiker som arbetade med både ren och tillämpad matematik. Hans bidrag till Gram–Schmidt's metod kom från hans avhandling om serieutvecklingar med användning av minstakvadratmetoden. Den tyske matematikern Erhard Schmidt, 1876 – 1959, avlade sin doktorsexamen i Göttingen under ledning av David Hilbert, en av den tidens ledande matematiker. Schmidt bidrog till att utforma teorin för Hilbertrum och betraktas som en av grundarna av den moderna funktionalanalysen. Gram–Schmidt's algorithm upptäcktes emellertid långt före både Gram och Schmidt av den inom många områden berömde franske matematikern Pierre – Simon Laplace, 1749 – 1827.

Referenser

[1] H. Anton, C. Rorres, Elementary linear algebra. Applications version, John Wiley & Sons. Inc., Ninth edition, 2005

[2] D. L. Kreider, R. G. Kuller, D. R. Ostberg, F. W. Perkins, An introduction to linear analysis, Addison–Wesley, Massachusetts, 1966

[3] M. Reed, B. Simon, Methods of Modern Mathematical Physics, Vol. 1: Functional Analysis. Academic Press, New York, 1972

[4] G. E. Shilov, An introduction to the theory of linear spaces, Prentice Hall, New Jersey, 1961

[5] K. J. Swanepoel, R Villa, *A lower bound for the equilateral number of normed spaces,* Proc. Amer. Math. Soc. 136(2008)1, pp. 127 – 131