

Studiehandledning till linjär algebra

Avsnitt 4

Kapitel 6 och 9.3 i Anton/Rorres:
"Elementary Linear Algebra: Applications version" (7:e uppl.)

I avsnitt 4 inför vi en inre produkt i våra vektorrum. En inre produkt är en generalisering av den välkända skalärprodukten. På samma sätt som skalärprodukten gav oss möjlighet att tala om begrepp som norm och ortogonalitet i det euklidiska n -dimensionella rummet, så får vi med hjälp av inre produkten tillgång till dessa begrepp i allmänna s.k. inre-produkt-rum.

Geometriska begrepp som längd, avstånd och vinkel får vi "på köpet".

För att underlätta beräkningar är det bra om basen är väl vald. Oftast är det fördelaktigt att arbeta med baser som har samma egenskaper som standardbasen i R^n , d.v.s. en bas där basvektorerna är parvis ortogonala och har alla normen (längden) 1. En sådan bas kallas ortonormal.

Vi lär oss använda Gram–Schmidts ortogonaliseringsmetod för att utgående från en godtycklig bas konstruera en ortonormal dito.

När man ställs inför ett överbestämt ekvationssystem saknas i regel exakt lösning. Ofta är man intresserad av att hitta en approximativ lösning, d.v.s. en vektor som gör avvikelsen så liten som möjligt. En metod för detta är minsta-kvadrat-metoden. Här kommer begrepp som avstånd och ortogonal projektion in.

Vi gör ett litet hopp fram till kapitel 9.3 för att studera hur minsta-kvadrat-metoden kan tillämpas på situationen där man vill anpassa en rät linje (eller en annan kurvform) till ett datamaterial.

Efter denna utflykt i boken återvänder vi till kapitel 6, som avslutas med en betraktelse över hur byte av bas kan systematiseras med hjälp av matriser, s.k. övergångsmatriser. Härvid dyker en speciell typ av matriser, ortogonalmatriser, upp i det fall då båda de aktuella baserna är ortonormala.

Läsanvisningar och kommentarer till läroboken

6.1 Inre produkter

Den välkända skalärprodukten $\mathbf{u} \diamond \mathbf{v} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n$ och dess egenskaper tas som modell när vi definierar en mer allmän **inre produkt** av två vektorer. Skalärprodukten är en linjär funktion av var och en av de två vektorerna, och antar värden bland de reella talen. Om de båda vektorerna är lika så är skalärprodukten större än eller lika med noll, med likhet exakt då de båda är nollvektorn.

Ovan nämnda egenskaper (som formuleras i sats 4.1.2) tas nu som definition; ett sätt att till varje par av vektorer ordna ett reellt tal kallas en inre produkt om dessa egenskaper gäller. Den inre produkten av vektorerna $\mathbf{u}$ och $\mathbf{v}$ betecknas $\langle \mathbf{u}, \mathbf{v} \rangle$. Ett vektorrum där man har definierat en inre produkt kallas för ett **inre-produkt-rum** (*inner product space*).

Det vanligaste sättet att definiera en inre produkt i R^n är att använda skalärprodukten, som också kallas den euklidiska inre produkten. Om inget annat sägs så underförstås att det är den euklidiska inre produkten som används.

En lättare modifikation av denna är att man tillmäter komponenterna olika stor vikt. Den **viktade** (eller **vägda**) **euklidiska inre produkten** (*weighted Euclidean inner product*) av $\mathbf{u}$ och $\mathbf{v}$ är $\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \dots + w_n u_n v_n$, där $u_1, \dots, u_n$ och $v_1, \dots, v_n$ är komponenterna av $\mathbf{u}$ respektive $\mathbf{v}$, och $w_1, \dots, w_n$ är positiva reella tal. Dessa senare tal kallas **vikter** (*weights*) i den inre produkten.

Man bör vara på det klara med att det är nödvändigt att ha en inre produkt för att man ska kunna använda begrepp som norm (eller längd), avstånd, vinkel och ortogonalitet. Normen av en vektor definieras som

$$\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2},$$

och avståndet mellan vektorerna (eller punkterna) $\mathbf{u}$ och $\mathbf{v}$ är

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Observera att våra välkända definitioner från kapitel 4 (sid 170-171) helt enkelt är specialfallet att den inre produkten är euklidisk: $\langle \mathbf{u}, \mathbf{u} \rangle = \mathbf{u} \diamond \mathbf{u}$.

Normen och avståndet definierade som i kapitel 4 kallas därför den euklidiska normen respektive det euklidiska avståndet. (Ex 3)

Längd, avstånd och vinkel är ju grundläggande geometriska begrepp. Dessa är baserade på den inre produkten. Man kan alltså säga att geometrin beror på valet av inre produkt. I Ex 5 ser vi hur en figur, som vi normalt (dvs med euklidisk inre produkt) uppfattar som en ellips, med en speciell inre produkt i stället är en cirkel. (En cirkel kännetecknas av att alla punkter ligger på ett och samma avstånd från en given punkt.)

Ett sätt att konstruera en inre produkt i R^n är att utgå från en inverterbar $n \times n$ -matris A , och definiera $\langle \mathbf{u}, \mathbf{v} \rangle = A\mathbf{u} \diamond A\mathbf{v}$. Detta innebär att inre produkten av $\mathbf{u}$ och $\mathbf{v}$ får vara skalärprodukten av bilderna av $\mathbf{u}$ respektive $\mathbf{v}$ under en linjär operator, vars matris är A .¹

Inre produkten kan då med hjälp av matrismultiplikation skrivas $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T A^T A \mathbf{v}$, där $\mathbf{u}$ och $\mathbf{v}$ uppfattas som kolonnmatriser.

(Jämför ekvation (4) på sid 292. Där står $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T A^T A \mathbf{u}$ i stället. Detta är samma sak. Varför?)

Om A är en diagonalmatris fås som specialfall en viktad euklidisk inre produkt, och om A är enhetsmatrisen fås förstås den vanliga euklidiska inre produkten $\mathbf{u} \diamond \mathbf{v}$.² Se Ex 6.

Exempel 7 till 10 visar hur inre produkter i andra rum än R^n kan se ut. Som vi sett tidigare kan vi ju ha exempelvis matriser eller funktioner som vektorer.

I rutan på sid 295 sammanfattas några räkneregler för inre produkt. De är precis desamma som för vanlig skalärprodukt, vilket är naturligt eftersom den allmänna inre produkten definitionsmässigt har samma yttre egenskaper³ som den euklidiska.

Övningar: 1c, 2c, 3a, 4a, 5, 6a, 8a, 9bc, 10c, 11c, 13a, 14.

6.2

Vinkel och ortogonalitet i inre-produkt-rum

Som vi vet sedan tidigare är ju begreppen skalärprodukt, norm och vinkel relaterade genom sambandet

$$\mathbf{u} \diamond \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta ,$$

där θ är vinkeln mellan de båda (nollskilda) vektorerna $\mathbf{u}$ och $\mathbf{v}$. Detta gäller i de euklidiska två- resp. tre-dimensionella rummen. Sambandet kan tas som definition på vinkel mellan två (nollskilda) vektorer i ett godtyckligt inre-produkt-rum, om vi byter skalärprodukten mot den aktuella inre produkten:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta .$$

(För att θ ska vara entydigt bestämd så kräver vi att dess värde ligger i intervallet från 0 till π .)

¹ Tänk gärna på den geometriska tolkningen i fallet $n = 2$ som vridning, töjning o.s.v.

² Observera att matrisen A inte är entydigt bestämd av en given inre produkt, utan det finns olika A som ger samma inre produkt. Den intressanta matrisen är ju $A^T A$. Prova gärna i fallet $n = 2$ med att låta A vara matrisen för en vridning, eller en spegling.

³ Med "yttre egenskaper" menas här de egenskaper som är oberoende av vektorernas komponenter.

Att definitionen fungerar bygger på att $\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|}$ är ett tal mellan -1 och 1 . Detta garanteras av Cauchy-Schwarz olikhet:

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|.$$

Beviset för denna olikhet (sid 300) bygger på ett litet trick, men är inte svårt. Du kanske inte känner till begreppet diskriminant, som nämns i beviset. Detta är inget märkvärdigt, och krävs ej heller. Kvadratkomplettera andragradspolynomet. Villkoret innebär att konstanttermen måste vara ≥ 0 , vilket ger olikheten.

Sats 6.2.2 och 6.2.3 säger att längd och avstånd i allmänna inre-produkt-rum har samma yttre egenskaper som i det euklidiska n -dimensionella rummet (kap 4.1). Detta hänger ju ihop med motsvarande samband för inre produkten. Satserna skulle ha hört till slutet av kap 6.1, om det inte vore för att beviset för triangelolikheten bygger på Cauchy-Schwarz olikhet.

Det är svårt nog att geometriskt tänka sig begrepp som längd och avstånd i högre dimensioner, eller i allmänna inre-produkt-rum. Skall vi nu också behöva räkna med vinklar i dessa? Nej, i praktiken är vi i dessa rum mest intresserade av en enda vinkel, nämligen den räta. **Ortogonalitet** är ett centralt begrepp i allmänna inre-produkt-rum. Vektorerna $\mathbf{u}$ och $\mathbf{v}$ sägs vara ortogonala om $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Pythagoras sats gäller generellt i alla inre-produkt-rum:

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2,$$

om $\mathbf{u}$ och $\mathbf{v}$ är ortogonala vektorer.

Resten av kap 6.2 kan hoppas över.

Övningar: 1e, 3c, 4a, 5, 6a, 7bd, 10a, 11a.

6.3 Ortonormala baser; Gram-Schmidt-ortogonalisering; QR-faktorisering

Som vi såg i kapitel 5.4 så har man stor frihet när det gäller att välja bas till ett vektorrum. Det finns oändligt många sätt att göra det på. Men alla är inte lika praktiska att ha att göra med. Det är en stor fördel om basvektorerna är ortogonala mot varandra.⁴ Dessutom är det en fördel om de är **normerade**,

⁴ Basvektorer är ju alltid linjärt oberoende. Ortogonalitet är ett strängare krav. Geometriskt är det så att två linjärt oberoende vektorer är icke-parallella, medan två ortogonala vektorer är vinkelräta. Vi förstår att ortogonalitet medför linjärt oberoende, men inte tvärtom.

Observera att egenskapen linjärt oberoende kan definieras utan referens till inre produkt, medan egenskapen ortogonalitet är beroende av en sådan.

d.v.s. har normen 1. En bas som har dessa egenskaper kallas **ortonormal**, eller förkortat **ON**.⁵

I så fall är det exempelvis lätt att uttrycka en godtycklig vektor i basen. Koordinaterna för en vektor $\mathbf{u}$ relativt basen är helt enkelt inre produkterna av vektorn med respektive basvektor $\mathbf{v}_i$ (sats 6.3.1):

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \dots + \langle \mathbf{u}, \mathbf{v}_n \rangle \mathbf{v}_n$$

Det algebraiska beviset på sid 313 för denna sats är inte svårt. Det bygger på egenskaperna för inre produkt och på att basvektorerna är normerade och ortogonala. Det kan i alla fall vara bra att också förankra tankebyggnaden i det välkända exemplet med vår euklidiska tredimensionella rymd.

I tredimensionella rymden med euklidisk inre produkt och standard-basvektorerna $\mathbf{i} = (1,0,0)$, $\mathbf{j} = (0,1,0)$ och $\mathbf{k} = (0,0,1)$ har vi för en godtycklig vektor $\mathbf{u} = (a,b,c)$ att $\langle \mathbf{u}, \mathbf{i} \rangle = \mathbf{u} \cdot \mathbf{i} = (a,b,c) \cdot (1,0,0) = a$. På samma sätt har vi att $\langle \mathbf{u}, \mathbf{j} \rangle = b$ och $\langle \mathbf{u}, \mathbf{k} \rangle = c$. Satsen säger alltså i detta fall att $(a,b,c) = a(1,0,0) + b(0,1,0) + c(0,0,1)$, vilket uppenbart är sant.

Sats 6.3.2 innebär att när vi beräknar inre produkt, norm eller avstånd i ett allmänt inre-produkt-rym, så kan vi räkna på vanligt euklidiskt sätt (som i kap 4.1) på koordinatvektorerna, förutsatt att basen är ortonormal.

Om basen man arbetar med är ortogonal, men basvektorerna inte är normerade, så måste formeln i sats 6.3.1 modifieras för att ta hänsyn till detta. Resultatet är (formel (1) sid 315):

$$\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 + \dots + \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2} \mathbf{v}_n.$$

(Som vi ser får vi uttrycket i sats 6.3.1 i specialfallet att samtliga basvektorer har normen 1.) Vi kommer strax att ha användning för detta sätt att uttrycka en godtycklig vektor, när vi beskriver Gram–Schmidt-metoden för att givet en godtycklig bas konstruera en ortogonal bas för ett inre-produkt-rym.

Sats 6.3.3 säger att om n vektorer är ortogonala så är de också linjärt oberoende. (Se också fotnot 4 i dessa blad.) Jämför beviset för denna sats och beviset för sats 6.3.1.

Ett mycket viktigt resultat är projektionssatsen (6.3.4). Varje vektor kan på exakt ett sätt skrivas som en summa av två komponenter, sådana att en ligger i ett delrum och den andra är ortogonal mot delrummet.

Detta illustreras väl av figur 2 på sid 317. Man har ofta glädje av en enkel geometrisk åskådning som denna, för att lättare kunna ta till sig abstrakta

⁵ Begreppen ortogonal respektive ortonormal används även om godtyckliga mängder av vektorer, inte bara baser. En mängd är ortogonal om varje par av (olika) vektorer i mängden är sinsemellan ortogonala. Om dessutom varje vektor i mängden är normerad är mängden ortonormal.

algebraiska sammanhang. Vi säger att vi projicerar vektorn $\mathbf{u}$ ortogonalt på delrummet W . Figuren visar egentligen bara situationen där vektorrummet (inre-produkt-rummet) är R^3 och delrummet är ett plan, men satsen säger att motsvarande gäller allmänt. Därför är detta en bra bild av ortogonal projektion på ett delrum till ett allmänt inre-produkt-rum.

Beviset för satsen är inte svårt men lite långt. Det ges sist i underkapitlet men kan hoppas över.

Sats 6.3.5 anger hur den ortogonala projektionen av $\mathbf{u}$ på ett delrum W uttrycks relativt en lämplig bas för W . Med lämplig menas här en ON-bas (a), eller åtminstone en ortogonal bas (b). I fallet ON-bas, med basvektorena $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$, ges projektionen $\text{proj}_W \mathbf{u}$ av

$$\text{proj}_W \mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \dots + \langle \mathbf{u}, \mathbf{v}_r \rangle \mathbf{v}_r$$

Detta kan lämpligen förstås på följande sätt:

Delrummet W är också ett inre-produkt-rum med samma inre produkt som V . Nu är $\text{proj}_W \mathbf{u}$ en vektor i detta rum, och kan därför enligt sats 6.3.1 skrivas

$$\text{proj}_W \mathbf{u} = \langle \text{proj}_W \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \text{proj}_W \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \dots + \langle \text{proj}_W \mathbf{u}, \mathbf{v}_r \rangle \mathbf{v}_r.$$

Detta ser ut som resultatet vi vill komma fram till, förutom att det står $\text{proj}_W \mathbf{u}$ i stället för $\mathbf{u}$ i de inre produkterna. Det är i själva verket exakt samma sak. Vi har till exempel i den första termen att

$$\langle \text{proj}_W \mathbf{u}, \mathbf{v}_1 \rangle = \langle \mathbf{u} - \text{proj}_{W^\perp} \mathbf{u}, \mathbf{v}_1 \rangle = \langle \mathbf{u}, \mathbf{v}_1 \rangle - \langle \text{proj}_{W^\perp} \mathbf{u}, \mathbf{v}_1 \rangle.$$

Men $\langle \text{proj}_{W^\perp} \mathbf{u}, \mathbf{v}_1 \rangle = 0$, eftersom $\text{proj}_{W^\perp} \mathbf{u}$ definitionsmässigt är ortogonal mot alla vektorer i W , däribland $\mathbf{v}_1$. Alltså är $\langle \text{proj}_W \mathbf{u}, \mathbf{v}_1 \rangle = \langle \mathbf{u}, \mathbf{v}_1 \rangle$. På samma sätt fås att $\langle \text{proj}_W \mathbf{u}, \mathbf{v}_2 \rangle = \langle \mathbf{u}, \mathbf{v}_2 \rangle$ o.s.v.

Det viktigaste resultatet i kap 6.3 är att varje ändligt-dimensionellt inre-produkt-rum har en ON-bas.⁶ Beviset bygger på en konkret metod att konstruera en sådan bas, utgående från en godtycklig bas för rummet. Denna metod kallas Gram–Schmidts ortogonaliseringsmetod. Metoden bygger på successiva ortogonala projektioner, och beskrivs i de färgade plattorna på sid 318-319. Lär dig denna metod ordentligt!

Avslutningen av kapitel 6.3, om QR-faktorisering, är intressant men ingår inte i denna kurs. Du kanske kommer att stöta på detta i någon kurs i numerisk analys.

Övningar: 1ac, 2ac, 3c, 4c, 5b, 6a, 7b, 9b, 10b, 11a, 12, 16a, 17a.

⁶ Undantag är det triviala rummet som bara består av nollvektorn. I själva verket finns det alltid oändligt många alternativa ON-baser (utom för ett en-dimensionellt rum, som bara har två).

6.4

Minsta-kvadrat-approximation

Ett sätt att se på situationen med ortogonal projektion på delrum (figur 1, som är samma som figur 2 i kap 6.3) är följande. Antag att vi är tvungna att approximera en vektor $\mathbf{u}$ med en vektor i ett delrum W . Då är $\text{proj}_W \mathbf{u}$ den bästa (närmaste) approximationen. $\text{proj}_W \mathbf{u}$ är nämligen den vektor i W vars avstånd till $\mathbf{u}$ är minimalt. (Sats 6.4.1). Vår geometriska bild av projektionen indikerar detta (fig 1), och det kan lätt visas formellt med hjälp av Pythagoras sats.

Vi ska nu se hur dessa tankegångar kan appliceras på linjära ekvationssystem, som saknar exakt lösning. Man är då ofta intresserad av att hitta en approximativ lösning.

Ett linjärt ekvationssystem kan skrivas $A\mathbf{x} = \mathbf{b}$, eller ekvivalent $A\mathbf{x} - \mathbf{b} = \mathbf{0}$. Om lösning saknas, vilket är det normala om ekvationssystemet är överbestämt (d.v.s. då antalet ekvationer m överstiger antalet obekanta n), söker vi i stället efter den vektor $\mathbf{x}$ som gör $A\mathbf{x} - \mathbf{b}$ så nära $\mathbf{0}$ som möjligt. Vi vill alltså minimera avståndet $d(A\mathbf{x} - \mathbf{b}, \mathbf{0}) = \|A\mathbf{x} - \mathbf{b} - \mathbf{0}\| = \|A\mathbf{x} - \mathbf{b}\|$.

Norm och avstånd beror ju på valet av inre produkt. Om vi använder den euklidiska inre produkten, d.v.s. skalärprodukten, så säger vi att den vektor $\mathbf{x}$ som minimerar $\|A\mathbf{x} - \mathbf{b}\|$ är **lösningen i minsta-kvadratmetodens mening** (*least squares solution*) till ekvationssystemet $A\mathbf{x} = \mathbf{b}$.

Studera figur 2 sid 330. Observera att för alla vektorer $\mathbf{u}$ så gäller att $A\mathbf{u}$ ligger i W , som här symboliseras av ett plan. För ett val av $\mathbf{u}$, denna vektor kallas $\mathbf{x}$, får vi den ortogonala projektionen av $\mathbf{b}$; den bästa approximationen. Vektorn $\mathbf{b} - A\mathbf{x}$ är alltså ortogonal mot alla vektorer $A\mathbf{u}$ i W . Detta villkor ger oss ett lösbart ekvationssystem för att bestämma $\mathbf{x}$. Ortogonaliteten ger

$$0 = A\mathbf{u} \cdot (A\mathbf{x} - \mathbf{b}) = (A\mathbf{u})^T (A\mathbf{x} - \mathbf{b}) = \mathbf{u}^T A^T (A\mathbf{x} - \mathbf{b}) = \mathbf{u}^T (A^T A\mathbf{x} - A^T \mathbf{b}).$$

Alltså har vi $\mathbf{u} \cdot (A^T A\mathbf{x} - A^T \mathbf{b}) = 0$. Men detta ska gälla för alla $\mathbf{u}$! Enda möjligheten är då att $A^T A\mathbf{x} - A^T \mathbf{b} = \mathbf{0}$. Vi har kommit fram till

$$A^T A\mathbf{x} = A^T \mathbf{b}$$

Detta system består av de så kallade **normalekvationerna** (*normal equations*) för lösningen i minsta-kvadrat-metodens mening till systemet $A\mathbf{x} = \mathbf{b}$. Det är ett kvadratisk ekvationssystem (n ekvationer och n obekanta). Systemet har minst en lösning. Om matrisen $A^T A$ är inverterbar så kan vi multiplicera ledvis med inversen och får således en unik lösning:

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}.$$

Detta är fallet precis då A har linjärt oberoende kolonnvektorer. (Annars får vi oändligt många likvärdiga lösningar.) I fallet med unik lösning får vi på köpet ett uttryck för den ortogonala projektionen av $\mathbf{b}$:

$$\text{proj}_W \mathbf{b} = A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}$$

Studera Ex 1 och 2.

De linjära operatorer som tidigare i kursen har studerats i R^2 och R^3 kan också generaliseras till allmänna inre-produkt- rum. I kap 4.2 behandlades ortogonal projektion på koordinataxlar resp. koordinatplan. Vi kan nu enkelt beskriva ortogonal projektion på delrum med hjälp av ovanstående resultat för den ortogonala projektionen av en vektor $\mathbf{b}$ (som ju för detta ändamål kan uppfattas som en godtycklig vektor). Matrisen för ortogonal projektion på W är alltså $A(A^T A)^{-1} A^T$, där A är en matris sådan att W är dess kolonnrum.⁷

Övningar: 1a, 3a, 5a, 7, 8a, 11a.

9.3

Minsta-kvadrat-anpassning till data

I detta avsnitt ser vi hur liknande ideer som i kap 6.4 kan användas för att anpassa en kurva (till exempel en rät linje) till observerade data. Vad vi är ute efter är att använda observationer för att få en uppskattning av värdena för parametrar i en matematisk modell. Antalet observationer är större än antalet parametrar (antalet ekvationer är större än antalet obekanta), så vi har ett överbestämt ekvationssystem. På grund av mätfel⁸ kommer inte alla ekvationer att kunna satisfieras, utan vi får leta efter en approximativ lösning — i minsta-kvadrat-metodens mening.

Om vi har n observationer⁹ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ så behöver vi ett sätt att mäta avståndet i R^n mellan de observerade y -värdena $\mathbf{y}$ och modellens förutsägelse $M\mathbf{v}$. Vektorn $\mathbf{v}$ innehåller de okända parametrarna och matrisen M de observerade x -värdena. Om man använder det vanliga euklidiska avståndet i R^n så svarar det mot (roten ur) summan av kvadraterna av de lodräta avstånden mellan punkterna och kurvan (fig 2 sid 462).

På samma sätt som i kap 6.4 får vi minsta-kvadrat-metodens lösning genom att lösa normalekvationerna $M^T M\mathbf{v} = M^T \mathbf{y}$

Läs igenom teorin och exemplen både för anpassning till rät linje och till polynom.

Övningar: 1, 3, 8.

(Fel i facit till övning 1. Skall vara $y = -\frac{1}{2} + \frac{7}{2}x$. Rätt i Lösningsmanualen.)

⁷ Här har vi förutsatt att $A^T A$ är inverterbar, d.v.s. att kolonnerna är linjärt oberoende. Skulle de inte vara det kan man ju alltid plocka bort "onödiga" kolonner tills de blir det — det är ju kolonnrummet W som är det intressanta, och inte matrisen A i sig. Detta innebär i praktiken att vi kräver att kolonnerna för matrisen A utgör en bas för W .

⁸ Det kan väl inte vara något fel på den matematiska modellen...

⁹ Observera att n här står för antalet ekvationer, inte som i kapitel 6.4 för antalet obekanta.

Ortogonala matriser; basbyte

Det finns en speciell klass av kvadratiska matriser som är särskilt betydelsefulla, **ortogonala matriser**, även kallade **ON-matriser**. En anledning till deras betydelse är att de förmedlar byte mellan två ortonormala baser, på ett sätt som vi skall se i detta kapitel.

Definitionen säger att en matris är ortogonal om dess invers är lika med dess transponat: $A^{-1} = A^T$. Villkoret kan ekvivalent formuleras $A^T A = I$ eller $AA^T = I$.

Ortogonalmatriser har en rad egenskaper, som framgår av satserna 6.5.1 till och med 6.5.3. Sats 6.5.1 säger att ON-matriser känns igen på att kolonnvektorererna är normerade och parvis ortogonala. Detsamma gäller radvektorererna.¹⁰ Se till att du förstår detta med hjälp av resonemang om elementen i $A^T A$ resp. AA^T som skalärprodukter av kolonnvektorer resp. radvektorer i A (som i beviset (a) € (b) på sid 339).

En linjär operator vars matris är ortogonal kallas en ortogonal operator. Sats 6.5.3 innebär att ortogonala operatorer är sådana som inte påverkar norm eller skalärprodukt. Om vi ser det geometriskt kan vi säga att längder och vinklar är invarianta¹¹ under ortogonala operatorer.¹² Tänk på vridningar och speglingar!

När vi ska börja studera basbyten är det praktiskt att införa det som kallas för **koordinatmatriser**. En sådan är helt enkelt en koordinatvektor, men arrangerad i form av en kolonnmatris. Avsikten är att förbereda den för matrismultiplikation.

Det är viktigt att ha klart för sig att en vektor är något som "står över" konventioner som vilken bas man för tillfället använder.¹³ När man byter bas kommer koordinatvektorn att se annorlunda ut, men det är samma vektor den beskriver. En och samma vektor i R^n beskrivs av en uppsättning av n tal, koordinaterna, relativt en bas, och en annan uppsättning av n koordinater relativt en annan bas. Sambandet mellan koordinatvektorer (eller koordinatmatriserna) ges genom multiplikation med en $n \times n$ -matris, **övergångsmatrisen** (*transition matrix*).

Studera noga det allmänna basbytesproblemet i R^2 och dess lösning på sid 342-343. Motsvarande gäller i R^n , och lösningen sammanfattas i den inramade rutan på sid 343. Lär dig detta! Läs Ex 4, och sedan Ex 5 där vi vänder på steken och studerar basbytet åt andra hållet.

¹⁰ Man kan tycka att de här matriserna skulle kallas ortonormala, och inte bara ortogonala. Av historiska skäl gör de dock inte det. Matrisernas natur framgår klarare av det förkortade namnet ON-matriser.

¹¹ Invariant betyder oförändrad.

¹² Sådana geometriska avbildningar brukar kallas **isometriska**.

¹³ Tänk gärna på det konkreta exemplet med geometriska vektorer. En pil från Stockholm till Umeå är en pil från Stockholm till Umeå oavsett hur Kalle Pettersson definierar sitt koordinatsystem.

Det ligger i sakens natur att basbyten skall kunna ske åt vilket håll som helst. Det innebär att övergångsmatriser alltid är inverterbara (sats 6.5.4). Inversen är lika med övergångsmatrisen för basbytet åt andra hållet.

Avslutningsvis konstateras att övergångsmatrisen mellan två ON-baser är en ON-matris. Ex 6 och 7 visar situationen då de nya basvektorena fås genom att vrida de gamla en vinkel θ i xy -planet. Jämför gärna med kap 4.2.

Övningar: 3bcf, 5b, 7b, 8, 11, 14, 16.