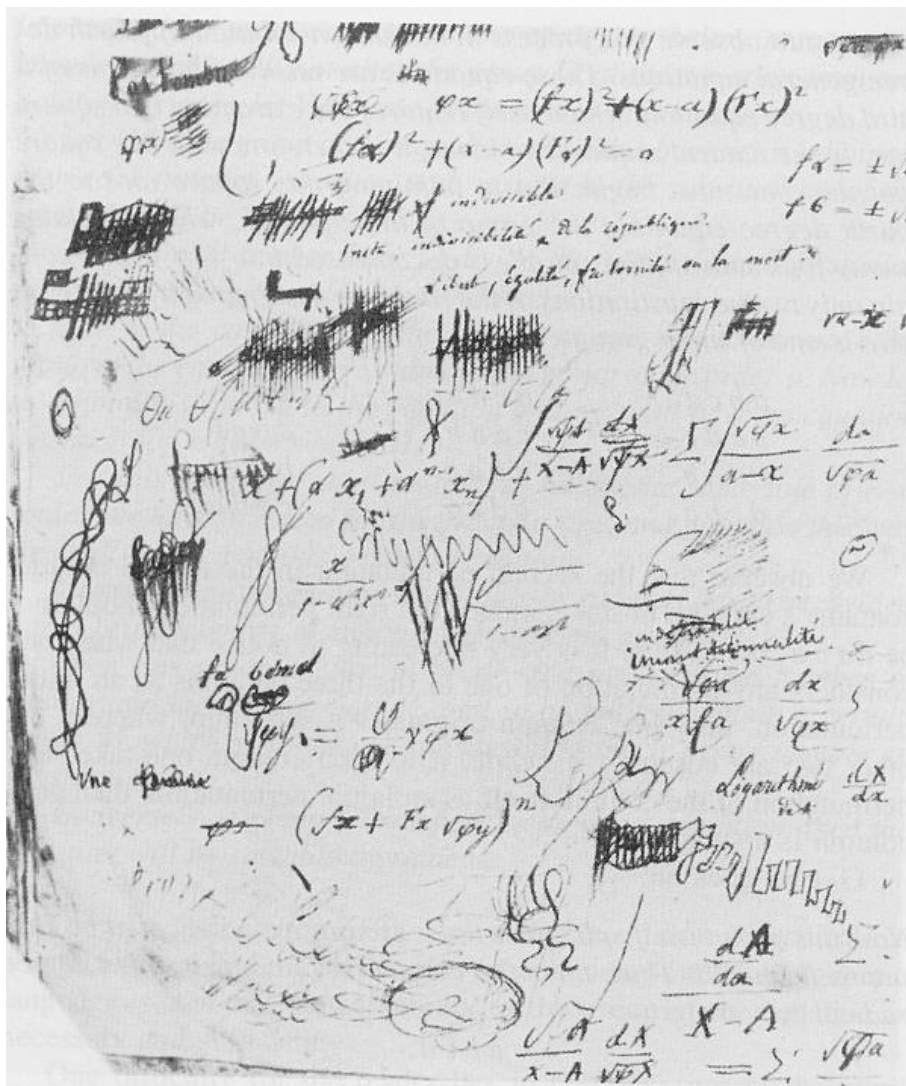


Algebrans fundamentalsats och polynomekvationer

Anton Holmquist, Axel Antonsson, Jimmy Pettersson, Andreas Argelius

15 maj 2006



Anteckningar från den betydelsefulla matematikern
Évariste Galois (1811-1832)

Innehåll

1	Inledning	3
2	Algebrans fundamentalsats	4
3	Polynomekvationer	7
3.1	Förstegradsekvationer	7
3.2	Andragradsekvationer	7
3.3	Tredjegradssekvationer	8
3.4	Fjärdegradsekvationer	10
3.5	Femtegradsekvationer och högre	11
4	Betydelsefulla personer inom området	14
4.1	Carl Friedrich Gauss	14
4.2	Niels Henrik Abel	14
4.3	Évariste Galois	15
5	Referenser	17

1 Inledning

Då polynomekvationer är en grundläggande form av ekvation har dessa studerats extra mycket. Den enklaste polynomekvationen är förstgradsekvationen som grafiskt representerar en rak linje. Andragradsekvationer försökte man lösa redan i Babylonien för 4000 år sedan, men man kunde då bara lösa vissa specialfall. Det kom att dröja till 800-talet innan man fann en fullständig lösningsformel. Tredjegradssekvationer studeras några hundra år före kristus av indiska matematiker men det kom att dröja länge innan man kunde lösa dem. Fram till och med 1500-talet kunde man bara lösa vissa specialfall. I början av 1500-talet lyckades Scipione del Ferro lösa ekvationer av formen $x^3 + mx - n = 0$. En av hans studenter utvecklade detta till en fullständig lösningsformel då han insåg att alla tredjegradssekvationer kunde skrivas på formen ovan. Lösningsformeln för tredjegradssekvationer är dock mycket avancerad och sågs som en stor upptäckt på sin tid. Samma lösningsmetod kunde sedan bara justeras lite för att lösa även fjärdegradsekvationer. När man nu hittat allmänna lösningsmetoder för både tredje- och fjärdegradsekvationer med så kort mellanrum trodde alla att femtegradsekvationerna snart också skulle vara lösbara. Det skulle dock komma att ta stopp här och under nästan 300 år var detta problem en drivfjäder för matematiken. Både Euler och Lagrange försökte sig på problemet men tvingades ge upp. Inte förrän i början på 1800-talet kom man svaret på spåren. Det var Abel och Galois som stod för upptäckten som även grundade en ny gren inom matematiken som kom att kallas *gruppteori*. Svaret var att femtegradsekvationerna inte har någon allmän lösningsformel. Detta oväntade resultat chockerade dåtidens matematiker och många vägrade acceptera detta faktum. Sökandet efter lösningsmetoder till polynomekvationerna drev matematiken framåt under många år och i försök att lösa dessa fann man metoder som visade sig användbara inom andra områden av matematiken.

2 Algebrans fundamentalsats

Algebrans fundamentalsats går uttrycka på en stort antal sätt. Nedan följer två av de vanligaste.

- Varje polynom av grad n med reella eller komplexa koefficienter

$$f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$$

har minst en lösning i det komplexa talplanet.

- Varje polynom av grad n ,

$$f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$$

kan faktoriseras till en produkt av exakt n faktorer,

$$(x - \alpha_1)(x - \alpha_2)\dots(x - \alpha_n),$$

där $\alpha_1, \alpha_2, \dots, \alpha_n$, är komplexa tal och rötterna till ekvationen $f(\alpha) = 0$.

Historia

1608 kom boken *Arithmetica Philosophica* av Peter Rothe, i vilken han skriver att ett polynom av n :e graden med reella koefficienter kan ha n stycken lösningar. 21 år senare, i boken *L'invention nouvelle en l'Algèbre*, antog Albert Girard att ett polynom av n :e graden har n stycken lösningar, men han nämnde aldrig något om att de var komplexa lösningar. Han lade till att hans antagande var sant så länge ekvationen var komplett, med vilket han menade att ingen av koefficienterna var lika med 0. Dock visade det sig att han trodde på sitt antagande för alla polynom då han visar att ekvationen $x^4 = 4x - 3$ har de fyra lösningarna $1, 1, -1 + i\sqrt{2}$, och $-1 - i\sqrt{2}$.

Ifrån fundamentalsatsen följer det att varje polynom med reella koefficienter och av en grad större än 0 kan faktoriseras till en produkt av reella polynom av grad 1 och 2. 1702 påstod Gottfried Leibniz att ett polynom som skrivs på formen $x^4 + a^4$, där a är reell och skild från 0, inte kan faktoriseras på det sätt som har beskrivits ovan. Senare gjorde Nikolaus Bernoulli samma antagande, denna gång gällande polynomet $x^4 - 4x^3 + 2x^2 + 4x + 4$. men han fick ett brev 1742 ifrån Leonhard Euler i vilket det stod att hans polynom råkade vara lika med $(x^2 - (2 + \alpha)x + 1 + \sqrt{7} + \alpha)(x^2 - (2 - \alpha)x + 1 + \sqrt{7} - \alpha)$, där α är roten ur $4 + 2\sqrt{7}$. Detta visar också att Leibniz påstående var falskt ty, $x^4 + a^4 = (x^2 + a\sqrt{2}x + a^2)(x^2 - a\sqrt{2}x + a^2)$.

Den första som gav sig på att försöka bevisa fundamentalsatsen var d'Alembert. Detta gjordes 1746 men beviset var ofullständigt. Ett av de största problemen men hans bevis var att det antog att ett annat teorem var sant, Puiseux's teorem som bevisades över hundra år senare. Likväl så antog Puiseux's teorem att fundamentalsatsen var sann så de båda vilade på varandra utan att vara bevisade.

Det skulle vara många över åren som försökte sig på att bevisa fundamentalsatsen, namnkunniga matematiker så som Euler, de Foncenex, Lagrange och Laplace försökte sig på att bevisa fundamentalsatsen men alla begick samma misstag. De antog implicit att Girards antagande var sant, att lösningarna existerade och allt som var kvar att bevisa var att de kunde skrivas på formen $a + bi$, a och b reella tal.

Gauss presenterade sitt första bevis 1799, vilket byggde på geometri men det hade ett topologiskt fel. Argand publicerade 1806 ett noggrant bevis som för första gången tog i anspråk komplexa koefficienter. Gauss publicerade sedan två bevis till 1816 och sedan en ny version av hans första bevis 1849.

Det första bevis som kom att tryckas i en bok var Cauchys - *Cours d'analyse de l'École Royale Polytechnique*. Det trycktes 1821 och innehöll Argands bevis men på något vis hade Cauchy glömt att nämna Argand överhuvudtaget.

Det första konstruktiva beviset skrevs av Hellmuth Kneser och förenklades sedan av hans son Martin Kneser.

Bakgrund

Varje algebraisk ekvation av grad n med reella, eller komplexa, koefficienter

$$f(x) = x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_1x + a_0 = 0$$

har lösningar i det komplexa talplanet. Detta fastställdes redan på 1500-talet för 3:e och 4:e gradens ekvationer. Gauss var som bekant den som först bevisade att lösningarna existerade (men det återstod att visa att man kunde uttrycka lösningarna med rationella operationer med rotuttryck.) Gauss teorem säger att för varje algebraisk ekvation av formen (ovan) där n är ett positivt heltal och varje a är vilket reellt eller även komplext tal som helst så existerar det åtminstone ett komplext tal $\alpha = c + di$ så att

$$f(\alpha) = 0.$$

Talet α kallas roten till ekvationen ovan. Om vi för stunden antar att ovanstående är sant utan att bevisa det så kan vi med hjälp av faktorsatsen bevisa fundamentalsatsen och visa att varje polynom av grad n kan faktoriseras till en produkt av precis n faktorer så att

$$f(x) = (x - \alpha_1)(x - \alpha_2)\dots(x - \alpha_n).$$

Att de α :n som man har fått fram är rötter till ekvationen $f(\alpha) = 0$ ses enklast efter faktoriseringen som vi gjort ovan, eftersom $x - \alpha_n$ är en faktor blir således hela $f(x)0$. I vissa fall blir faktorerna likadana, till exempel för polynomet

$$f(x) = x^2 - 2x + 1 = (x - 1)(x - 1)$$

Här ser vi efter faktoriseringen att polynomet bara har en rot för $x = 1$, man säger att polynomet har en rot med multiplicitet 2.

Olika bevismetoder

De bevis som finns till fundamentalsatsen grundar sig i analysen, eller mer precist på kontinuitet. Vissa bevis, särskilt de tidigare, bevisar bara att ett givet polynom med reella koefficienter har minst en komplex rot. Detta är tillräckligt eftersom ett givet polynom $q(z) = p(z)\overline{p(\bar{z})}$ bara har reella koefficienter. Om z är nollställe till $q(z)$ så måste z vara nollställe till p eller till \bar{p} . Om z är nollställe till \bar{p} så är \bar{z} nollställe till p .

Många av de icke-algebraiska bevisen använder sig av det faktum att $p(z)$ beter sig som z^n då $|z|$ är stort (kallas growth lemma). Det går ut på att det finns ett tal R så att $|z|^n/2 < |p(z)| < 3|z|^n/2$ då $|z| > R$.

Ett fullständigt bevis

Sätt

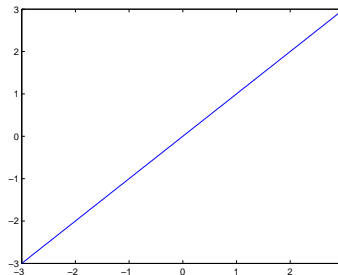
$$|p(z)| = |z|^n \left| a_n + \frac{a_{n-1}}{z} + \dots + \frac{a_0}{z^n} \right|,$$

där $a_n \neq 0$. Här ser vi att $|p(z)| \rightarrow \infty$ då $z \rightarrow \infty$. Antag att $p(z)$ saknar rötter. Då är funktionen $\frac{1}{p(z)}$ en hel analytisk funktion. Funktionen har gränsvärdet 0 då $z \rightarrow \infty$ vilket innebär att den är begränsad i hela det komplexa talplanet. $\frac{1}{p(z)}$ är då enligt Liouilles sats konstant vilket innebär att $p(z)$ skulle vara konstant. Detta är en motsägelse då $n > 0$. Härav har $p(z)$ minst ett komplext nollställe. Detta bevis som blir väldigt kort och lättläst bygger på Liouilles sats som säger att de enda hela begränsade funktionerna är de konstanta funktionerna, samt att man har kunskap om analytiska funktioner.

3 Polynomekvationer

3.1 Förstgradsekvationer

En ekvation av första ordningen, även kallad linjär ekvation, är en ekvation som endast innehåller summor och produkter av konstanter och variabler med exponenten 1. Ordet linjär kommer från att de representerar raka linjer i ett kartesiskt koordinatsystem. Ett vanligt sätt att uttrycka en förstgradsekvation är på formen $y = mx + c$. På den här formen fastställer m grafens lutning och c punkten där linjen skär y -axeln.



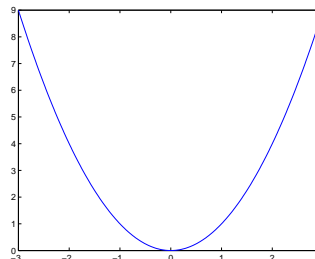
Lösning av förstgradsekvationer

Att lösa en linjär ekvation är på grund av dess enkla natur mycket triviale och har varit möjligt sedan matematikens begynnelse. (Läsaren antas redan vara bekant med dess lösningsmetoder, men skulle inte så vara fallet rekommenderas en grundläggande genomgång i högstadiematematik.)

3.2 Andragradsekvationer

Historia

I babylonien har man hittat andragradsekvationer på lertavlor som tros vara omkring 4000 år gamla. Där fanns också beskrivet enkla lösningsmetoder till vissa specialfall. 1500 år senare hade den grundläggande matematiken utvecklats en hel del och babylonierna hade genom att utnyttja kvadratkompletterna åstadkommit en metod för att hitta positiva rötter till en allmän andragradsekvation. I jakten på en fullt generell formeln testades många metoder, en av de mera lyckosamma var en mer abstrakt geometrisk metod producerad av Euklides några hundra år före kristi födelse, men inte heller denna metod var fullständig. Faktum var att det skulle dröja till 600-talet innan någon presenterade en metod för att hitta negativa rötter, detta gjordes av den indiske matematikern Brahmagupta. Två hundra år senare var problemet med andragradsekvationer löst. På 800-talet blev indierna Shridhara den första att ge ut ett verk med en fullständig lösningsmetod, hans originalverk har tyvärr kommit bort men så här blev hans lösningen citerad i senare verk:

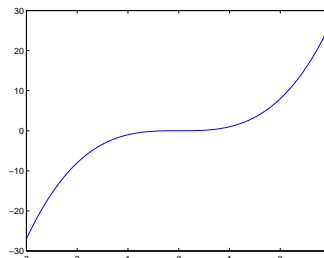


Multiplicera båda sidorna i ekvationen med ett tal fyra gånger så stort som koefficienten till kvadraten på den okända. Addera båda sidorna med kvadraten på koefficienten till den okända. Ta sedan roten ur.

3.3 Tredjegrads ekvationer

Historia

Tredjegrads ekvationer upptäcktes ett par hundra år före Kristus av indiska matematiker. På grund av komplexiteten av sådana ekvationer dröjde det länge innan en allmän lösning metod hittades. Fram till 1500-talet var enda verktyget ett par geometriska knep som kunde användas för att hitta rötter till vissa ekvationsmässiga specialfall. Det skulle dock visa sig att tredjegrads ekvationer är fullt lösbare med algebraiska metoder.



I början av 1500-talet hittade den italienska matematikern Scipione del Ferro en allmän metod för att lösa ekvationer av typen $x^3 + mx - n = 0$. På denna tid låg det stor prestige i att ha kunskapen att kunna lösa någonting som ingen annan kunde, så del Ferro höll sin metod hemlig fram till just före sin död då han lät en av sina studenter vid namn Tartaglia ta del av hans upptäckt. Tartaglia själv fortsatte att utveckla metoden och lyckades färdigställa den genom att visa att *alla* tredjegrads ekvationer kan skrivas om på ovan nämnda form. Tartaglia insåg att han nu hade tillgång till en fullständig lösningsformel och avslöjade den för vetenskapsmannen Gerolamo Cardano, som publicerade i den i sitt verk *Ars Magna* år 1545. Nedan följer en härledning av formeln.

Det är inte svårt att förstå varför tredjegrads ekvationer länge gäckade matematiker över hela världen då följande allmänna lösningsmetod är långt mer komplex än motsvarande formel för andragrads ekvationer.

del Ferro och Tartaglias lösningsmetod

En allmän tredjegrads ekvationer kan skrivas på formen

$$x^3 + ax^2 + bx + c = 0. \quad (1)$$

I avsikt att reducera bort andragradstermen gör vi substitutionen $x = t - \frac{a}{3}$, som ger oss $(t - \frac{a}{3})^3 + a(t - \frac{a}{3})^2 + b(t - \frac{a}{3}) + c = 0$. Utvecklar vi sedan uttrycket enligt binomialsatsen och stuvlar om lite så erhåller vi följande tredjegrads ekvation i t :

$$t^3 + (b - \frac{1}{3}a^2)t + \frac{1}{27}(2a^3 - 9ab) + c = 0.$$

Återigen gör vi nu en variabelsubstitution genom att sätta $p = b - \frac{1}{3}a^2$, och $q = \frac{1}{27}(2a^3 - 9ab) + c$. I och med detta har vi reducerat ursprungsproblemet till att lösa en ekvation på den *komprimerade* tredjegradsformen

$$t^3 + pt + q = 0. \quad (2)$$

Eventuella läsare som ännu inte fått nog av variabelsubstitutioner gläds nog åt vårt följande antagande att det finns tal u och v så att

$$\begin{cases} q = u^3 - v^3 \\ p = 3uv \end{cases}$$

Meningen med denna substitution är att ekvationen (2) då får lösningen $t = v - u$, vilket kan inses genom insättning och användande av binomialsatsen. Eftersom vi har att $x = t - \frac{a}{3}$, $t = v - u$, och $v = \frac{p}{3u}$ kan vi skriva de eftersökta lösningarna x till (1) som

$$x = \frac{p}{3u} - u + \frac{a}{3}. \quad (3)$$

Vårt nästa steg är att uttrycka u i variablerna p och q . Man måste till att börja med dock fråga sig om det alltid är möjligt att uttrycka p , q i u , v enligt ovanstående. Svaret på detta är ja. Om vi bryter ut v ur den undre ekvationen och sätter in i den övre får vi att $q = u^3 - \frac{p^3}{27u^3}$. Detta kan ses som en andragradsekvation i u^3 vilket kan förtydligas genom att multiplicera båda leden med u^3 så att vi får $u^6 - \frac{p^3}{27} - qu^3 = 0$. Variabelsubstitutionen $s = u^3$ ger oss ekvationen $s^2 - qs - \frac{p^3}{27} = 0$ som enligt den bekanta andragradslösningsformeln har lösningarna

$$s = u^3 = \frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \Rightarrow u_1 = \sqrt[3]{\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \quad (4)$$

Den vänstra tredjegradsekvationen kallas *binomisk* och kan lösas allmänt genom att man skriver om uttrycket på polär form. För att hitta de andra två skriver vi om båda leden på polär form.

$$\begin{cases} u = re^{i\theta} \\ \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} = \rho e^{i\varphi}. \end{cases}$$

Att lösa ekvationen (3) blir alltså ekvivalent med att lösa $r^3 e^{3i\theta} = \rho e^{i\varphi}$. Identifierar vi belopp och argument i de polära formerna i ekvationen erhåller vi följande värden på r och θ :

$$\begin{cases} r = \sqrt[3]{\rho} \\ \theta = \frac{\varphi}{3} + k\frac{2\pi}{3}, \quad k \in 0, 1, 2. \end{cases}$$

Vi får alltså tre värden på u .

$$\begin{aligned} k = 0 \text{ ger } u_1 &= \sqrt[3]{\rho} \cdot e^{i\frac{\varphi}{3}} = \sqrt[3]{\rho \cdot e^{i\varphi}}, \\ k = 1 \text{ ger } u_2 &= \sqrt[3]{\rho} \cdot e^{i(\frac{\varphi}{3} + \frac{2\pi}{3})} = \sqrt[3]{\rho} \cdot e^{i\frac{\varphi}{3}} e^{i\frac{2\pi}{3}} = \sqrt[3]{\rho \cdot e^{i\varphi}} \cdot e^{i\frac{2\pi}{3}}, \\ k = 2 \text{ ger } u_3 &= \sqrt[3]{\rho} \cdot e^{i(\frac{\varphi}{3} + \frac{4\pi}{3})} = \sqrt[3]{\rho} \cdot e^{i\frac{\varphi}{3}} e^{i\frac{4\pi}{3}} = \sqrt[3]{\rho \cdot e^{i\varphi}} \cdot e^{i\frac{4\pi}{3}}. \end{aligned}$$

Men $\sqrt[3]{\rho \cdot e^{i\varphi}}$ är ju enligt (4) lika med u_1 . Dessutom kan vi, enligt Eulers formel, $e^{iy} = \cos y + i \sin y$, göra omskrivningarna

$$\begin{aligned} e^{i\frac{2\pi}{3}} &= -\frac{1}{2} + i\frac{\sqrt{3}}{2} \text{ och} \\ e^{i2\cdot\frac{2\pi}{3}} &= -\frac{1}{2} - i\frac{\sqrt{3}}{2}. \end{aligned}$$

Tillsammans kan vi ur detta finna uttryck även för u_2 och u_3 .

$$\begin{cases} u_1 = \sqrt[3]{\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ u_2 = u_1 \cdot \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) = \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) \cdot \sqrt[3]{\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ u_3 = u_1 \cdot \left(-\frac{1}{2} - i\frac{\sqrt{3}}{2}\right) = \left(-\frac{1}{2} - i\frac{\sqrt{3}}{2}\right) \cdot \sqrt[3]{\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \end{cases}$$

Vi har nu uttryckt u i variablerna p och q som i sin tur är uttryckt i a , b och c från (1). Vi är nu i stort sett klara. Allt som återstår är att sätta in $p = b - \frac{1}{3}a^2$ och $q = \frac{1}{27}(2a^3 - 9ab) + c$ i ovanstående uttryck för u och sedan i sin tur sätta in dessa värden på u i uttrycket för x i (3). Detta kräver som ni ser en stunds räknande som vi av utrymmesskal här ej redovisar för, men nedan följer resultatet.

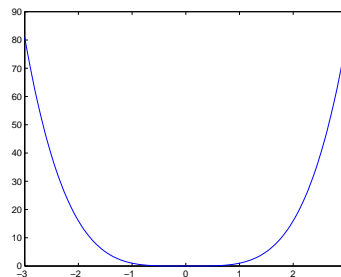
$$\begin{aligned} x_1 &= -\frac{a}{3} + \sqrt[3]{\frac{-2a^3+9ab-27c+\sqrt{(2a^3-9ab+27c)^2+4(-a^2+3b)^3}}{54}} \\ &\quad + \sqrt[3]{\frac{-2a^3+9ab-27c-\sqrt{(2a^3-9ab+27c)^2+4(-a^2+3b)^3}}{54}} \\ x_2 &= -\frac{a}{3} - \frac{1+i\sqrt{3}}{2} \sqrt[3]{\frac{-2a^3+9ab-27c+\sqrt{(2a^3-9ab+27c)^2+4(-a^2+3b)^3}}{54}} \\ &\quad + \frac{1+i\sqrt{3}}{2} \sqrt[3]{\frac{-2a^3+9ab-27c-\sqrt{(2a^3-9ab+27c)^2+4(-a^2+3b)^3}}{54}} \\ x_3 &= -\frac{a}{3} + \frac{1+i\sqrt{3}}{2} \sqrt[3]{\frac{-2a^3+9ab-27c+\sqrt{(2a^3-9ab+27c)^2+4(-a^2+3b)^3}}{54}} \\ &\quad - \frac{1+i\sqrt{3}}{2} \sqrt[3]{\frac{-2a^3+9ab-27c-\sqrt{(2a^3-9ab+27c)^2+4(-a^2+3b)^3}}{54}} \end{aligned}$$

Helt trivialt är det alltså inte att härleda uttryck för rötterna till tredjegrads-ekvationer. Som tur är kan dagens matematiker ta hjälp av datorer för att enkelt hitta lösningarna.

3.4 Fjärdegradsekvationer

Historia

Fjärdegradsekvationer upptäcktes av matematiker i det gamla Indien ungefär samtidigt som tredjegrads-ekvationerna, alltså ett par hundra år före Kristus. Sedan dess har det lagts ner mycket möda på att hitta dess rötter. Det skulle dock dröja länge innan en allmän formel hittades, men det visade sig faktiskt att en sådan faktiskt existerade. Först med bedriften blev italienern Lodovico Ferrari på 1500-talet och lösningsformeln blev publicerad, tillsammans med den för tredjegrads-ekvationer, i Gerolamo Cardano's verk *Ars magna*.



Om lösningsformeln

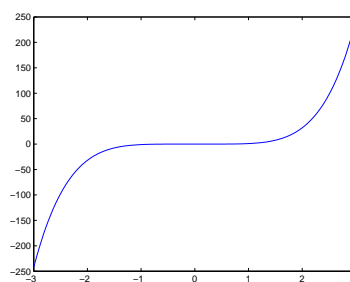
Lösningsmetoden är till uppbyggnad och idé mycket lik motsvarande formel för tredjegrads ekvationer, men dock bra mycket mer komplex. Den färdiga formeln för att hitta rötterna är ungefär fyra gånger längre än tredjegradsformeln så vi avstår från att skriva ut den.

De som är intresserade kan hitta formeln här:

<http://planetmath.org/encyclopedia/QuarticFormula.html>

3.5 Femtegradsekvationer och högre

Den logiska fortsättningen borde vara att hitta formler för att lösa polynom-ekvationer av femte graden och högre. Man kanske till och med skulle kunna hitta ett samband med vilket man kan konstruera rötterna till ett allmänt polynom av godtycklig grad. I själva verket har man till dags dato inte ens klarat av att ställa upp en lösningsformel för ekvationer av femte graden. Eftersom lösningar till tredje- och fjärdegraderna upptäcktes kort efter varandra verkar steget till att lösa femtegraderna inte vara långt. Detta måste varit väldigt förbryllande för dåtidens matematiker. Bland de som gjort seriösa ansatser att lösa problemet finner man stora namn som Leonard Euler och Joseph-Louis Lagrange.



I början av 1800-talet lyckades dock två unga matematiker bringa klarhet i problemet och dessutom bana väg för en helt ny gren av matematiken. Den norske matematikern Niels Henrik Abel och fransmannen Évariste Galois lade oberoende av varandra grunden till vad som kom att kallas grupp teori. Abel publicerade 1824 ett bevis till en sats som säger att:

Det är omöjligt att lösa en allmän polynomekvation där graden är högre än fyra med ett *ändligt* antal algebraiska operationer.

Italienaren Paolo Ruffini hade redan elva år tidigare lagt fram ett bevis för satsen baserat på tidigare arbeten av Lagrange. Hans bevis innehöll dock ett smärre fel och blev ignorerat av de flesta. Ruffini försökte desperat få medhåll på sitt bevis. Han ville få Lagrange att studera det men till sin förtvivlan fick Ruffini inga svar då han postade sitt arbete till honom. Han försökte posta det ytterligare två gånger men även då utan framgång. Orsaken är oklar men kan bero på att Lagrange helt enkelt inte ville veta av att femtegraderna inte har någon lösning. Satsen brukar populärt kallas Abel-Ruffini-satsen. □

Alla är inte olösliga

Att en ekvation inte går att lösas algebraiskt innebär att dess rötter inte går att skrivas som ett uttryck med de vanliga räknesätten (addition, subtraktion, etc.) samt rotutdragningar. Detta innebär bland annat att det inte går att ställa

upp en formel som löser alla femtegradsekvationer. Det utesluter däremot inte möjligheten att approximera rötterna med numeriska metoder. Exempelvis med Newton-Raphson-metoden. Rötterna (fem stycken enligt algebras fundamentalsats) finns men vi kan inte bestämma dem exakt. Det finns däremot vissa ekvationer som går att faktorisera (och därmed lösa algebraiskt). Som ett exempel kan polynomekvationen

$$x^5 - x^4 - x + 1 = 0 \text{ skrivs som } (x^2 + 1)(x + 1)^2(x - 1) = 0$$

och har rötterna: $(1, -1, -1, i, -i)$. Detta konstaterande är viktigt och för oss in på näste matematiker, Galois. Abel bevisade att det var omöjligt att lösa en polynomekvation av femte graden eller högre algebraiskt. Galois gick ett steg längre. Som vi ser ovan finns det vissa undantag där lösningarna faktiskt kan framställas med algebraiska operationer.

Frågan Galois ställde sig var: Finns det något sätt att avgöra huruvida en given ekvation går att lösa eller ej? Han lyckades uppfinna en metod för att göra precis det. Matematiken bakom är väldigt komplicerad men har precis som Abels bevis sin grund i gruppteorin. Galois visade att om en ekvations *Galoisgrupp* är lösbar går ekvationen att lösa exakt. För ekvationer av graderna fyra eller lägre är *Galoisgruppen* alltid lösbar. Då graden är högre finns det ekvationer som t ex

$$x^5 - x + 1 = 0$$

vars rötter inte går att finna algebraiskt. Enligt *Galoisteorin* går alltså ovanstående ekvations Galoisgrupp inte att lösa. Intressant att påpeka är även att i femtegradarens fall kan antingen *inga* eller *alla* av rötterna uttryckas som en ändligt uttryck av algebraiska operationer. Detta förstår man lättast genom att ta en rot, α , som är ett algebraiskt tal. Då kan man utföra en division med $(x - \alpha)$. Enligt faktorsatsen reduceras problemet då till ett polynom av fjärde graden. Dessa har som bekant alltid algebraiska lösningar.

Abel-Ruffini-satsen kan å ena sidan ses som ett bakslag för matematiken. Den visar att vi aldrig kommer kunna lösa alla polynomekvationer exakt utan måste förlita oss på approximativa lösningar i vissa fall. Matematiken är en exakt vetenskap där man vill undvika uppskattningar i så stor utsträckning som möjligt. Å andra sidan gav Abels och Galois arbeten upphov till gruppteorin, ett väldigt användbart hjälpmedel.

Kort om grupper och Galoisteori

Gruppteorin har trots sitt ursprung i algebran många användningsområden utöver att lösa ekvationer. Den kan tillämpas i flera grenar av matematiken. *Symmetrier* är ett grundläggande begrepp i gruppteorin som därför är ett värdefullt verktyg i geometrin. Matematiskt innebär en symmetri att ett objekt(en ekvation, en geometrisk figur eller dylikt) inte påverkas av en viss operation. Detta kan t ex vara att rotera en liksidig triangel 120° motsols.

Det kan tyckas konstigt att tala om symmetri för något så abstrakt som en ekvation men det har många fördelar. En fjärdegradsekvation kan t ex ha rötterna $(\sqrt{2}, -\sqrt{2}, \sqrt{5}, -\sqrt{5})$. Vi kallar dessa för α, β, γ respektive δ . Man ser att α och β hänger samman. Samma sak gäller för γ och δ . I själva verket är det

omöjligt att skilja paren åt med hjälp av en ekvation med rationella koefficienter. För γ och δ gäller både $\gamma^2 - 5 = 0$ och $\delta^2 - 5 = 0$. Samtidigt ser vi att $\gamma + \delta = 0$ och $\gamma\delta + 5 = 0$ gäller även om man byter plats på rötterna. Det finns oändligt många exempel. En Galoisgrupp är en grupp av alla permutationer där symmetrin upprätthålls. Man kan alltså inte byta ut α eller β mot γ eller δ . I det här fallet finns det fyra sådana permutationer.

Gruppteorin används även med framgång i andra vetenskaper. Främst i fysiken men även i kemin där man använder sig av gruppteorin för att undersöka kristallstrukturer.

4 Betydelsefulla personer inom området

4.1 Carl Friedrich Gauss

Gauss föddes 1777 i Braunschweig i Tyskland och var redan i tidig ålder intresserad av matematik. Hans far var murare och hans mor tjänsteflicka. Fadern ville att hans son också skulle bli murare men på grund av hans färdigheter inom matematik blev han det inte. Hertig Ferdinand av Brunswick hörde talas om Gauss begåvning och finansierade hans studier och blev under många år hans beskyddare. Gauss studerade mellan 1795-1798 på universitetet i Göttingen. Där utarbetade han bland annat minsta kvadratmetoden med vilkens hjälp man med stor noggrannhet kan avgöra värdet på en variabel. Gauss upptäckte 1796 att en regelbunden 17-hörning kan konstrueras med en passare och en ograderad linjal. I Gauss doktorsavhandling publicerade han sitt första bevis av algebrans fundamentalsats. Detta bevis innehöll dock geometriska argument och var inte helt rigoröst. Han publicerade sen ytterligare tre bevis. Det tredje beviset demonstrerade hans färdigheter inom komplex analys och det fjärde arbetade han med komplexa polynom. Gauss främsta intresse inom matematiken var talteori. Han införde begreppet kongruens och utarbetade sin kända funktion för primtalens fördelning. Gauss försökte hitta matematik inom allt han såg, ett exempel på detta är Gauss påskformel som används för att beräkna vilken dag påskdagen infaller ett visst år.

1801 upptäcks den första asteroiden (Ceres) då den passerar mellan jorden och solen. Asteroiden är bara synlig under en kort tid och Gauss får uppdraget att beräkna när den kommer att bli synlig nästa gång. Trots de knappa data som står till hans förfogande lyckas han beräkna dess bana med stor exakthet och asteroiden återfinns där Gauss förespätt. Gauss beräknade detta med minsta kvadratmetoden och tar inte bara hänsyn till solens gravitation utan även omgivande planeters. För detta erbjöds han en utnämning vid akademien i St. Petersburg men avböjde och stannade i Tyskland. Astronomin kom sedan att ta upp mycket av hans tid och han fann ytterligare tre asteroider (Pallas, Juno och Vestra).

Han gifte sig med Johanna Osthoff 1805 och fick med henne tre barn. Hon dog dock vid förlossningen av det tredje. Gauss gifte sig sedan med hennes bästa vän Minna Waldrick och som fick ytterligare tre barn. År 1806 i kriget mot Napoleon dog Gauss bidragsgivare men Gauss fick året därpå anställning på universitetet i Göttingen som professor och kunde fortsätta sitt arbete.

Gauss var en av de främsta matematikerna genom tiderna och hans arbete förde matematiken långt framåt. Han berörde i stort sätt alla områden inom matematiken och var även verksam inom astronomi och fysik. Det är svårt att förutsäga var matematiken skulle varit idag om det inte varit för Gauss men det står säkert att hans arbete var ett gigantiskt kliv framåt i utvecklingen.

4.2 Niels Henrik Abel

Abel föddes 5 augusti 1802 i Rogaland, Norge och dog 6 april 1829. Abel blev alltså bara 27 år gammal men det hindrade inte honom från att bli en av den mo-

derna tidens mest erkända matematiker. Abel hann under sitt korta liv bidra med mycket inom matematiken, exempelvis gjorde han stora insatser inom teorin för de elliptiska funktionerna, men det han är mest känd för är sitt bevis som visar att det är omöjligt att lösa femtegradsekvationer allmänt med algebraiska metoder. Abel har fått ge namn åt Abelpriset vilket delas ut en gång om året till den matematiker som har gjort det största bidraget till dagens matematik.

Abel var under större delen av sitt liv väldigt fattig och fick förlita sig på donationer ifrån sina medmänniskor för att kunna jobba med matematiken. 1824 lät han på egen bekostnad trycka sin avhandling i matematik i vilken han bevisade att det var omöjligt att fullständigt lösa ekvationer av femte graden med algebraiska metoder.

Efter sin avhandling fick han ett stipendie av Stortinget som gjorde det möjligt för honom att resa till Berlin där han fick tryckt ett par artiklar i *Journal für die reine und angewandte Mathematik*. Detta ledde Abel vidare både till Italien och Paris där han också presenterade sina senaste verk men de ledde inte till någon uppmärksamhet där. Tre år efter sin avhandling fick Abel tuberkulos och han kämpade mot den i två år innan sjukdomen tog hans liv. Det var under de två sista åren som han gjorde framstående arbeten inom de elliptiska funktionerna. I Berlin ville man att Abel skulle tilldelas en professur vid universitetet men precis innan han skulle tilldelas titeln så avled han.

4.3 Évariste Galois

Galois föddes i Frankrike år 1811 och var tidigt intresserad av matematik. Han kom senare att inleda den abstrakta algebran och har fått stor betydelse för matematikens utveckling. Under den korta tid han levde fick han dock inte uppleva någon uppskattning.

Hans intresse för matematik startade tidigt. Han var mycket begåvad, inte bara för sin ålder utan han behärskade under sin skoltid långt mer avancerad matematik än sina lärare. Han studerade på egen hand Abels och Lagranges texter eftersom ingen i hans omgivning förstod dem. Många av dåtidens professionella matematiker kunde inte tillgodogöra sig dessa texter. Galois sökte in till den högt ansedda skolan *Ecole Polytechnique* men vid inträdesproven ansågs han sakna begåvning. Han kom dock in på en annan skola och fick där uppskattning för sina idéer. Han skickade därpå in två texter till franska akademien som berörde lösningar till algebraiska ekvationer. Dessa dokument försvann dock. 1830 ställde Galois upp i franska akademiens stora matematiktävling men hans bidrag lästes aldrig igenom eftersom den ansvarige dog innan han hann läsa det. Den texten har inte återfunnits. Senare samma år var Julirevolutionen och Galois drogs tillsammans med de flesta andra studenter i oroligheterna. Han lämnade ytterligare texter till akademien men fick inget svar på flera månader. Senare fick han svaret att hans texter var obegripliga och han ombads förenkla dem. Galois var politiskt aktiv och revolutionär och fängslades flertalet gånger i samband med demonstrationer. Han inledde ett förhållande med en kvinna som indirekt blev hans död. Han kom att hamna i en duell med en annan man under oklara förhållanden. Han dog i sviterna av den duellen den 31 maj 1832.

Galois skrev under sitt korta liv inte lika många texter som många av de stora matematikerna men han tog ett stort steg i matematikens utveckling. Han fick inget allmänt erkännande medan han var i livet men efter hans död har hans arbete blivit mycket uppmärksammat.

5 Referenser

What is mathematics? - An elementary approach to ideas and methods

Richard Courant and Herbert Robbins (1969)

A History of Algebra

B.L. van der Waerden (1985)

Matematikens historia

Bo Göran Johansson

Wikipedia.org

http://en.wikipedia.org/wiki/Fundamental_theorem_of_algebra

http://en.wikipedia.org/wiki/Linear_equation

http://en.wikipedia.org/wiki/Quadratic_equation

http://en.wikipedia.org/wiki/Cubic_equation

http://en.wikipedia.org/wiki/Quartic_equation

http://en.wikipedia.org/wiki/Quintic_equation

PlanetMath.org

<http://planetmath.org>

Mathworld

<http://mathworld.wolfram.com/>

Ekvationssystem och Gröbnerbaser

David Nilsson, Tomas Palm,
Jonathan Svárdén, Jesper Öqvist

LUNDS TEKNISKA HÖGSKOLA

18 maj 2006

Innehåll

1	Inledning	3
2	Ideal och ekvationssystem	4
2.1	Ideal	4
2.2	Lösningssmängder	4
2.3	Ekvationssystem i en variabel	5
3	Termordning	7
3.1	Lexikografisk ordning	7
4	Gröbnerbaser	7
5	Maple-exempel	9

1 Inledning

Vid problemlösning uppkommer allt som oftast ekvationssystem innehållandes polynomekvationer, både linjära och icke-linjära. Linjära ekvationssystem löses som bekant med vanlig Gausselimination, medan det icke-linjära fallet handlar om ekvationssystem med godtyckliga polynom i flera variabler. Sådana ekvationssystem innehållandes godtyckliga polynom kan i de enklaste fallen lösas med direkt insättning. Som exempel har vi,

$$x^2 + y^2 = 1 \tag{1}$$

$$y = x \tag{2}$$

Med insättning i (1) med (2) så fås,

$$2x^2 = 1 \iff x^2 = \frac{1}{2}. \tag{3}$$

Då $x = y$ fås lösningarna

$$x = \frac{1}{\sqrt{2}} \quad y = \frac{1}{\sqrt{2}}$$

eller,

$$x = -\frac{1}{\sqrt{2}} \quad y = -\frac{1}{\sqrt{2}}.$$

Som vi ser här så får vi en ekvation (3) i en variabel som är lätt att lösa. Just detta med att lösa en ekvation i en variabel kommer att bli den brännande punkten lite senare i denna rapport. Direkt insättning gäller bara de enklaste fallen och är väldigt svårt att applicera på mer komplicerade system. Ett systematiskt sätt att behandla ekvationssystem är Gröbnerbaser. Detta skall vi försöka förklara lite närmre i denna rapport.

Men innan dess så tittar vi lite på vad som ligger bakom tänkandet. Sedan den dagen då man först kom i kontakt med matematik har man fått lära sig vad som är tillåtet och inte när man räknar med algebraiska uttryck. Det är inte förrän mycket senare som man kommer i kontakt med den omkringliggande strukturen som talar om för en varför saker och ting är som de är. Den delen kallas för *abstrakt algebra*. I denna del av matematiken spelar språket en viktig roll. Som i alla andra fackområden så används ett fackspråk för att kommunikationen skall kunna ske utan missförstånd. Som en grundpelare inom matematiken finns mängder i olika former. I vårt fall en mängd av polynom, där namnet *Ringar* dyker upp. Andra mängder är t.ex. \mathbb{Z} , mängden av alla hela tal. Operationerna på elementen i mängderna kallas för binära operationer och innefattar de vanliga räknesätten $\{+, -, *, /\}$. Beroende på vilken struktur, (*ringar, kroppar* etc), som avses så är en eller flera binära operationer tillåtna. Som vi ska se senare så har vi vår mängd av alla polynom och dess delmängder *ideal*.

Teorin bakom Gröbnerbaser introducerades år 1965 av Bruno Buchberger i hans avhandling [4], men namngavs efter hans mentor Wolfgang Gröbner. Relevansen bakom teorin var under en längre tid ifrågasatt, inte förrän datorn fick sin framfart började den matematiska allmänheten förstå hur viktig teorin bakom

Gröbnerbaser är. Detta är ett område som för närvarande är under aktiv utveckling, både genom nya teoretiska insikter, men även genom en mängd nya praktiska tillämpningar.

2 Ideal och ekvationssystem

För att kunna definiera Gröbnerbaser krävs någonting som kallas ideal. Ideal är ett slags matematiskt objekt och kommer i sig från en mer generell matematisk struktur som kallas ring. Vi behöver inte ha djupa kunskaper om ringar eller ideal för att skaffa oss en uppfattning om hur Gröbnerbaser fungerar, men vi måste förstå vissa delar om just *varför* det fungerar.

2.1 Ideal

Låt $K[x_1, \dots, x_n]$ beteckna mängden av alla polynom i n variabler. Här står K för att koefficienterna kommer från en generell matematisk struktur som kallas *kropp*, men vi kan här strunta i det och tänka på K som t.ex. de reella talen.

Definition Idealet I i $K[x_1, \dots, x_n]$ som genereras av polynomen f_1, \dots, f_s är mängden av alla polynom som kan skrivas $u_1 f_1 + \dots + u_s f_s$ där u_i är polynom i $K[x_1, \dots, x_n]$. Idealet som genereras av f_1, \dots, f_s skrivs

$$I = \langle f_1, \dots, f_s \rangle$$

EXEMPEL Låt $I = \langle f, g \rangle$. Det är då självklart att f och g var för sig finns i I ty alla element på formen $u_1 f + u_2 g$ finns i idealet och om man väljer t.ex. $u_1 = 1$ och $u_2 = 0$ så får man f som ett enskilt element i I .

EXEMPEL Låt I genereras av $\{x^2 y^3 - xy, xy^2 - y\} = \{p_1, p_2\}$. Att exempelvis polynomen $t = x^2 y^4 - xy^2$ och $p = xy^3 - y^2$ ligger i I ses lätt ty

$$t = yp_1 + 0 \cdot p_2 = y(x^2 y^3 - xy) = x^2 y^4 - xy^2$$

$$p = 0 \cdot p_1 + yp_2 = y(xy^2 - y) = xy^3 - y^2$$

2.2 Lösningssmängder

Vi skall nu studera lösningarna till ekvationssystem där högerleden är noll och vi har polynom i vänsterleden. Om vi t.ex. tittar på systemet

$$\begin{cases} f_1 = 0 \\ \dots \\ f_s = 0 \end{cases}$$

så gör vi kopplingen till idealet I som genereras av polynomen f_1, \dots, f_s :

$$I = \langle f_1, \dots, f_s \rangle.$$

Vi ska nu se att alla lösningar till det ovanstående ekvationssystemet är de samma som lösningarna till alla polynom i det ideal som genereras av de i ekvationssystemet ingående polynomen.

Sats Om $Z(f_1, \dots, f_s)$ betecknar alla lösningar $X = (x_1, \dots, x_n)$ som satisfierar $f_i(X) = 0$ för alla polynom f_1, \dots, f_s och om $Z(\langle f_1, \dots, f_s \rangle)$ är alla X som satisfierar $u_1 f_1 + \dots + u_s f_s = 0$ för alla polynom u_i så är

$$Z(f_1, \dots, f_s) = Z(\langle f_1, \dots, f_s \rangle).$$

Bevis. Om $f_1(X) = \dots = f_n(X) = 0$ för något X så måste $u_1 f_1(X) + \dots + u_n f_n(X) = 0$ för godtyckliga polynom u_i ty $u_i \cdot 0 = 0$. Omvänt så måste $u_1 f_1(X) + \dots + u_n f_n(X) = 0$ medföra att $f_1(X) = \dots = f_n(X) = 0$ ty det är bara att välja $u_i = 1$, och alla andra till noll så att $u_1 f_1(X) + \dots + u_n f_n(X) = f_i(X)$. Alla lösningar $Z(f_1, \dots, f_n)$ är alltså även lösningar i $Z(\langle f_1, \dots, f_n \rangle)$ och vice versa. \square

Antag nu att två olika uppsättningar polynom f_1, \dots, f_s och g_1, \dots, g_t genererar samma ideal – då måste motsvarande ekvationssystem enligt föregående sats ha samma lösningsmängd ty

$$Z(f_1, \dots, f_s) = Z(\langle f_1, \dots, f_s \rangle) = Z(\langle g_1, \dots, g_t \rangle) = Z(g_1, \dots, g_t).$$

2.3 Ekvationssystem i en variabel

För två heltal går det som bekant att beräkna den så kallade största gemensamma delaren (*greatest common divisor* på engelska). Den största gemensamma delaren för heltal beräknas med hjälp av *Euklides algoritm* (Tabell 1) och den så kallade *divisionsalgoritmen*, eller det faktum att varje helt tal a kan skrivas som en produkt av ett annat heltal b , en kvot q och en rest r :

$$a = qb + r.$$

Om resten är noll så går divisionen jämnt upp och $a/b = q$.

```

INPUT a, b hela tal
OUTPUT sgd, största gemensamma delaren av a och b
WHILE (b är skild från noll)
{
    r := resten av a dividerat med b
    a := b
    b := r
}
sgd := a

```

Tabell 1: Euklides algoritm

För polynom finns det som bekant polynomdivision som fungerar analogt med heltalsdivision. Det går även att beräkna en största gemensamma delare för polynom. Den största gemensamma delaren för två polynom f_1 och f_2 betecknar vi $sgd(f_1, f_2)$ och kan beräknas med Euklides algoritim precis som med heltal – den enda skillnaden är att alla variabler är polynom och det beräknas polynoma kvoter och rester.

Det är värt att nämna att det också går att beräkna största gemensamma delare för ett godtyckligt antal polynom induktivt enligt

$$sgd(f_1, \dots, f_n) = sgd(f_1, sgd(f_2, \dots, f_n)).$$

Nyttan utav att kunna beräkna största gemensamma delare för polynom kommer av att idealet genererat av en mängd godtyckliga polynom f_1, \dots, f_n i en variabel är det samma som idealet genererat av deras största gemensamma delare, det behövs alltså bara en ekvation för att lösa ett system med flera polynom i en variabel:

$$sgd(f_1, \dots, f_n) = 0 \iff \begin{cases} f_1 = 0 \\ \dots \\ f_s = 0 \end{cases}$$

Detta beror som sagt på följande sats där vi hoppar över beviset.

Sats *Givet polynom f_1, \dots, f_s så gäller det att $\langle sgd(f_1, \dots, f_s) \rangle = \langle f_1, \dots, f_s \rangle$.*

EXEMPEL Låt

$$\begin{cases} f_1 = x^3 - 3x + 2 = 0 \\ f_2 = x^2 - 1 = 0 \end{cases}$$

vara ett ekvationssystem som vi vill lösa. Vi kan då med polynomdivision och Euklides algoritim beräkna största gemensamma delare. Vi får $sgd(f_1, f_2) = x - 1$ och $x - 1 = 0$ ger $x = 1$. Lösningen till

$$\begin{cases} x^3 - 3x + 2 = 0 \\ x^2 - 1 = 0 \end{cases}$$

är alltså $x = 1$.

3 Termordning

Då man utför polynomdivision i endast en variabel är det naturligt att ordna termerna fallande efter gradtal. I flervariabelfallet finns inget direkt motsvarande sätt att ordna termerna, därför måste man i förväg bestämma sig för termernas inbördes ordning. Tag t.ex. xy^2 och x^2y , de har samma grad men är uppenbart olika. Termordningar visar sig vara helt avgörande för beräkningen av Gröbnerbaser. Det finns en mängd olika sätt att ordna termer på, t.ex. total gradordning eller lexikografisk ordning. Vi ska här beskriva lexikografisk ordning som vi behöver senare.

3.1 Lexikografisk ordning

När det gäller lexikografisk ordning så måste man bestämma hur variablerna ska ordnas, t.ex. $x_1 > \dots > x_n$.

Att lexikografiskt sortera termerna i ett polynom i n variabler är en process i n steg. Dvs. (1) sortera fallande genom att jämföra största variabelns exponent, om likhet, (2) sortera fallande efter andra variabelns exponent, ..., (n) sortera fallande efter n :te variabelns exponent. Om vi har ett polynom p så definierar vi den ledande termen $LT(p)$ som den största termen i polynomet.

EXEMPEL Polynomet $y^3 + 3y^2x^2 + 2x^3$ ska sorteras i lexikografisk ordning. Om $x > y$ så sorteras polynomet till $2x^3 + 3x^2y^2 + y^3$ och $LT(p) = 2x^3$. Om $y > x$ så sorteras polynomet till $y^3 + 3y^2x^2 + 2x^3$ och $LT(p) = y^3$.

4 Gröbnerbaser

En Gröbnerbas för ett givet ideal är en mängd polynom $\{g_1, \dots, g_t\}$ som har vissa egenskaper. Det finns flera olika definitioner för Gröbnerbaser vilka alla såklart är ekvivalenta. Vi skall använda nedanstående.

Definition Låt I vara ett ideal i $K[x_1, \dots, x_n]$. En ändlig delmängd $\{g_1, \dots, g_t\} \subset I$ kallas en Gröbnerbas för I om varje polynom $f \in I$ går att skriva

$$f = u_1g_1 + \dots + u_tg_t, \text{ där } LT(f) \geq LT(u_i g_i), 1 \leq i \leq t \quad (4)$$

Vi ser att en Gröbnerbas för I genererar I . Det ytterligare villkoret är påståendet om de ledande termerna. Således är Gröbnerbaser beroende av den termordning man väljer. Givet en mängd generatorer för ett ideal så finns en algoritm kallad *Buchbergers algoritm* för att beräkna en Gröbnerbas. Metoden som används för att bestämma en Gröbnerbas kan simplificerat beskrivas som en kombination av Gausseliminering och en generalisering av Euklides algoritm. Motsvarigheten till polynomdivision i flera variabler kallas reduktion. För en utförlig behandling av Gröbnerbaser refererar vi till [1], [2], [3], [5], [6].

Vi skall nu studera egenskapen som visar varför Gröbnerbaser är lämpliga för att lösa ekvationssystem. Denna kallas ibland eliminationsegenskapen.

För ett ideal $I \in K[x_1, \dots, x_n]$ låt $I_k = I \cap K[x_{k+1}, \dots, x_n]$, dvs. polynomen i I som ej innehåller någon av variablerna x_1, \dots, x_k . Det går att visa att I_k är ett ideal i $K[x_{k+1}, \dots, x_n]$.

Sats Om $G = \{g_1, \dots, g_t\}$ är en Gröbnerbas för ett ideal I i $K[x_1, \dots, x_n]$, med lexikografisk termordning med $x_1 > x_2 > \dots > x_n$, då är $G_k = G \cap K[x_{k+1}, \dots, x_n]$ en Gröbnerbas för I_k .

Bevis. Om ett polynom $f \in I_k \subset I$ framställs på formen (4) så måste varje term i högerledet innehållandes x_1, \dots, x_k vara större än $LT(f)$ eftersom $f \in I_k$, och I_k endast innehåller x_{k+1}, \dots, x_n . I högerledet finns således bara polynom i x_{k+1}, \dots, x_n , speciellt är alla inblandade (där $u_i \neq 0$) $g_i \in G_k$. Basen G_k är alltså en Gröbnerbas i I_k . \square

Vi antar nu för enkelhetens skull att vi har ett ekvationssystem

$$\begin{cases} f_1 = 0 \\ \dots \\ f_s = 0 \end{cases}$$

där $f_i \in K[x_1, \dots, x_n]$ med ett ändligt antal lösningar (principen gäller även för andra system). Man kan då visa att idealet $\langle f_1, \dots, f_s \rangle$ innehåller, för varje x_i , ett polynom i bara x_i . Låt $I = \langle f_1, \dots, f_s \rangle$ och G en Gröbnerbas för I med samma lexikografiska ordning som ovan. Om vi använder oss av ovanstående slutsats för x_n -variabeln ser vi att I_{n-1} och således G_{n-1} är icke-tomma. Med samma argument innehåller I_{n-2} polynom i både x_{n-1} och x_n så att $G_{n-2} \subset G$ måste ha polynom innehållandes x_{n-1} och x_n , men inga andra variabler. På samma sätt ser vi att G måste vara på "triangulär" form, där G_{n-1} endast innehåller x_n , G_{n-2} innehåller x_{n-1} och x_n , G_{n-3} innehåller x_{n-2}, x_{n-1}, x_n och så vidare.

Nu löser vi envariabelsystemet $\{f(x_n) = 0 \mid f \in G_{n-1}\}$, sedan återsubstituerar vi dessa lösningar i de polynom som bara innehåller x_n och x_{n-1} från G_{n-2} och så vidare, tills hela ursprungssystemet är löst. Till slut har vi fått fram alla lösningar till polynomen i G och således till vårt ursprungliga system. Slutsatsen blir att om man använder sig av Gröbnerbaser kan man reducera problemet från att lösa ett system av godtyckliga polynom i $K[x_1, \dots, x_n]$ till att lösa envariabelsystem.

5 Maple-exempel

För att referera till tidigare exempel där direkt insättning var ett sätt att lösa ett system av polynomekvationer så visar vi hur samma system förenklas med Gröbnerbaser. Därefter löser vi ett inte fullt så enkelt exempel. Systemen löses i programet Maple som har inbyggt stöd för Gröbnerbaser.

Termordningen i Maple definieras med kommandot “plex” och står för “pure lexicographic order” där ordningen helt och hållet är lexikografisk. Om vi använder exemplet från inledningen

$$\begin{cases} x^2 + y^2 - 1 = 0 \\ y - x = 0 \end{cases}$$

i Maple med termordningen $y > x$, så får vi:

```
with(Groebner):
```

```
f1:=x^2+y^2-1;
```

$$f1 := x^2 + y^2 - 1$$

```
f2:=y-x;
```

$$f2 : y - x$$

```
gbasis({f1,f2},plex(y,x));
```

$$[2x^2 - 1, -x + y]$$

När vi anger termordningen $y > x$ så elimineras variabeln y och vi får en ekvation $2x^2 - 1 = 0$ i bara x . Samma exempel från inledningen av rapporten med termordningen $x > y$ i Maple visar,

```
f1:=x^2+y^2-1;
```

$$f1 := x^2 + y^2 - 1$$

```
f2:=y-x;
```

$$f2 : y - x$$

```
gbasis({f1,f2},plex(x,y));
```

$$[2y^2 - 1, x - y]$$

När vi anger termordningen $x > y$ så elimineras variabeln x och vi får en ekvation i bara y .

Om vi då tittar på ett exempel som inte är fullt så enkelt

$$\begin{cases} x^2y + x - xy - 1 = 0 \\ x^4y^2 + yx^3 - x + 1 = 0 \end{cases}$$

och väljer $y > x$ och gör en körning i Maple så får vi,

```
f1:=x^2*y+x-x*y-1;
```

$$f1 := x^2y + x - xy - 1$$

```
f2:=x^4*y^2+y*x^3-x+1;
```

$$f2 := x^4y^2 + yx^3 - x + 1$$

```
gbasis({f1,f2},plex(y,x));
```

$$[x^2 - 2x + 1, x + xy - 1 - y, y^2 + y]$$

Om vi tittar på Gröbnerbasen $[x^2 - 2x + 1, x + xy - 1 - y, y^2 + y]$ så ser vi att två av polynomen är i en variabel. När vi erhållit lösningarna till

$$\begin{cases} x^2 - 2x + 1 = 0 \\ x + xy - 1 - y = 0 \\ y^2 + y = 0 \end{cases}$$

så är ursprungsproblemet löst. Om vi vänder på termordningen och väljer $x > y$ så får vi

```
f1:=x^2*y+x-x*y-1;
```

$$f1 := x^2y + x - xy - 1$$

```
f2:=x^4*y^2+y*x^3-x+1;
```

$$f2 := x^4y^2 + yx^3 - x + 1$$

```
gbasis({f1,f2},plex(x,y));
```

$$[y^2 + y, x + xy - 1 - y, x^2 - 2x + 1]$$

I detta fallet så får vi ingen skillnad mer än att y skrivs före x i svaret. Dock är det fortfarande på en enklare form som är lätt att lösa. Här hade det varit svårt att gissa sig till lösningarna och metoden med direkt insättning hade inte gett oss någon tillfredställande lösning. Därav Gröbnerbasernas stora styrka och användbarhet.

Referenser

- [1] William W. Adams and Philippe Lounstaunau. *An introduction to Gröbner bases*. American Mathematical Society, Providence, RI, 1994.
- [2] Per Andersson. Gröbnerbaser - ett verktyg för att konstruera och räkna med matematiska objekt. Technical report, Luleå Tekniska Universitet, 1997.
- [3] Thomas Becker and Volker Weispfenning. *Gröbner bases*. Springer-Verlag, New York, 1993. A computational approach to commutative algebra, In cooperation with Heinz Kredel.
- [4] Bruno Buchberger. *On finding a vector space basis of the residue class ring modulo a zero dimensional polynomial ideal (German)*. PhD thesis, University of Innsbruck, Austria, 1965.
- [5] David Cox, John Little, and Donal O'Shea. *Ideals, varieties, and algorithms*. Springer-Verlag, New York, second edition, 1997. An introduction to computational algebraic geometry and commutative algebra.
- [6] Ralf Fröberg. *An introduction to Gröbner bases*. John Wiley & Sons Ltd., Chichester, 1997.

Gruppteori

Hussein al-Hadad, Tobias Karlsson, Martin Nordholts, Johan Richter

16 maj 2006

Introduktion till gruppteori

Gruppteorin utvecklades i början av 1800-talet som ett matematiskt verktyg för att bevisa att vissa ekvationer inte har någon lösningsformel. Gruppteorin är en del av den abstrakta algebran, där man inte tar hänsyn till vilka objekt man hanterar, utan bara relationerna mellan dem.

Idag har gruppteorin andra tillämpningar, såväl inom den rena matematiken som andra vetenskaper. Inom matematiken används gruppteorin inom bl a talteorin och geometrin. Även inom fysiken spelar gruppteorin en viktig roll för att beskriva olika typer av symmetrier. T ex så utgör byten av koordinatsystem i mekanik en grupp. Inom partikelfysiken spelar gruppteorin en stor roll för att beskriva elementarpartiklarna.

Grundläggande gruppteori

Vad är en grupp?

En grupp består av en mängd och en operation. För att en mängd ska vara en grupp för en operation krävs 4 egenskaper.

- $a, b \in G \Rightarrow a \star b \in G$.
- $a \star (b \star c) = (a \star b) \star c$ för alla $a, b, c \in G$.
- Det finns ett element $e \in G$ (kallad identitets-elementet) sådant att $a \star e = e \star a = a$ för alla $a \in G$.
- För varje $a \in G$ finns det ett element $a^{-1} \in G$ (kallad inversen till a) sådant att $a \star a^{-1} = a^{-1} \star a = e$

Man benämner en grupp som abelisk om elementen även uppfyller den kommutativa lagen

- $a \star b = b \star a$ för alla $a, b \in G$.

Exempel på grupper

Heltal

Mängden \mathbb{Z} av alla heltal med operationen addition utgör en grupp. Den associativa lagen gäller: $(a+b)+c = a+(b+c)$, det finns ett neutralt element: $0+a = a+0 = a$, det finns en invers till varje tal: $a + (-a) = (-a) + a = 0$ och denna grupp är även abelisk: $a + b = b + a$. Om man istället väljer mängden alla icke-negativa heltal och operationen addition är det inte en grupp eftersom det då inte finns någon invers.

Reella tal

Mängden \mathbb{R} med operationen addition bildar en grupp, varvid 0 är ett neutralt element och a har inversen $-a$. Även de nollskilda reella talen \mathbb{R}^* med operationen multiplikation är en grupp, varvid 1 är ett neutralt element och a^{-1} är invers till a . Notera att noll inte har någon invers. Båda grupperna är abeliska eftersom addition och multiplikation är kommutativa operationer.

Icke-singulära $n \times n$ matriser

Mängden av alla icke-singulära $n \times n$ matriser (icke-singulär betyder att matrisen har en invers) utgör en grupp med operationen matrismultiplikation. Att axiomen ovan gäller för denna grupp har bevisats i den linjära algebran. Dessa grupper är ej abeliska när $n \geq 2$.

Permutationer

Mängden permutationer som avbildar t ex ABC på ABC, ACB, BAC, CAB, BCA, CBA är en grupp med operationen sammansättning. Elementen av denna grupp är:

$$e = \begin{pmatrix} A & B & C \\ A & B & C \end{pmatrix} \quad a = \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix} \quad b = \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix} \\ c = \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} \quad d = \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix} \quad g = \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix}$$

Till exempel så anger $a = \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}$

att A avbildas på B och B avbildas på C o.s.v. Operationen ab menas: först gör permutation b och sedan gör permutation a på förgående resultat. Denna grupp är ej abelisk eftersom $ac \neq ca$.

Exempel på satser

Utifrån de givna axiomen kan man härleda andra räknelagar och satser. Vi ger här några exempel.

SATS 1. Man kan förkorta bort element i en grupp, dvs om

$$ab = ac$$

följer att

$$b = c$$

och om

$$ba = ca$$

följer att

$$b = c.$$

BEVIS. Låt $ab = ac$. Vi vet att a har minst en invers. Låt d vara en invers till a och multiplicera med d från vänster.

$$dab = dac \Leftrightarrow (da)b = (da)c \Leftrightarrow ea = ec \Leftrightarrow a = c.$$

Den andra hälften av satsen bevisas på precis samma sätt. □

SATS 2. Varje element har en och endast en invers.

BEVIS. Att varje element har minst en invers är ett av axiomen för en grupp. Vi ska således visa att varje element har högst en invers. Anta att a har två inverser d och c . Då gäller att

$$ca = e$$

och

$$da = e.$$

Men det betyder att

$$ca = da \Leftrightarrow c = d.$$

I sista steget har vi använt föregående sats. □

Den unika inversen till a skrivs a^{-1} .

Vi kan notera att beviset för föregående sats utesluter möjligheten till en "vänsterinvers", dvs ett element $d \neq a^{-1}$ sådant att $a \star d = e$. På samma sätt kan man utesluta existensen av en högerinvers.

SATS 3. Låt a och b vara element i en grupp. Då är $(a \star b)^{-1} = b^{-1} \star a^{-1}$.

BEVIS.

$$(b^{-1} \star a^{-1}) \star (a \star b) = b^{-1} \star a^{-1} \star a \star b = b^{-1} \star (a^{-1} \star a) \star b = b^{-1} \star e \star b = b^{-1} \star b = e.$$

Eftersom inversen är unik måste $(a \star b)^{-1}$ vara lika med $b^{-1} \star a^{-1}$. □

SATS 4.

$$(a^{-1})^{-1} = a.$$

BEVIS. Vi vet att $(a^{-1})(a^{-1})^{-1} = e$ och att $a^{-1}a = e$. Men detta innebär att $(a^{-1})(a^{-1})^{-1} = a^{-1}a$ och genom att förkorta bort a^{-1} får vi att

$$(a^{-1})^{-1} = a.$$

□

Permutationsgrupper

En särskild viktig typ av grupper är permutationsgrupper. Betrakta en mängd M med n element. Numrera dem $1, 2, 3, \dots, n$. Betrakta sedan alla permutationer av M , dvs alla bijektiva funktioner från M till sig själv. Mängden av alla dessa permutationer, med den binära operationen funktionsammansättning, sägs utgöra permutationsgruppen av ordning n . Denna grupp betecknas ofta S_n . Det är viktigt att observera att elementen i S_n är bijektionsfunktioner av M och inte elementen i M .

Det finns flera alternativa sätt att beteckna elementen i S_n . Ett sätt är med hjälp av en sorts tabell med två rader där den översta raden helt enkelt består av siffrorna 1 till n . I den undre raden på plats j står vilket element som hamnar på plats j .

Till exempel så anger

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$$

att 1 avbildas på 2, 2 avbildas på 3 och så vidare.

Ett annat sätt att skriva elementen, som kommer användas i resten av denna artikel, är den så kallade cykliska notationen. Där anger $(2, 1, 3)$ att element på plats 2 går till plats 1, elementet på plats 1 går till plats 3 och element på plats 3 går till plats 2. En sådan permutation kallas cyklisk. För att kunna beskriva alla permutationer på detta sätt måste produkter av cykliska permutationer tillåtas. De element som inte ska flyttas anges överhuvudtaget inte. Exempel $(1, 2, 3)(9, 8, 4)$ beskriver en permutation cykliskt permuterar 1, 2 och 3 med varandra och 9, 8 och 4 med varandra. Övriga element flyttas inte.

En anledning till att permutationsgrupper är så viktiga är att alla ändliga grupper kan beskrivas som en undergrupp till något S_n . Denna sats, Cayleys sats, bevisas inte här.

Undergrupper

Definition 1. En delmängd H av en grupp G är en undergrupp till G om H själv är en grupp under operationen i G .

Alla grupper G har två triviala undergrupper; G själv och e . Alla andra grupper kallas äkta undergrupper.

Eftersom associativitetsaxiomet automatiskt ärvt till alla undergrupper räcker det att verifiera två axiomer för en undergrupp.

SATS 5. En icke-tom delmängd H av en grupp G är en undergrupp av G om

- Om $a, b \in H$, så gäller att $a \star b \in H$
- Om $a \in H$ så gäller att $a^{-1} \in H$.

BEVIS. Enligt ovan gäller associativitet för alla element i H , så alla axiom för en grupp är uppfyllda om vi kan visa att e tillhör H . Eftersom gruppen är icke-tom finns det ett element c . Då finns enligt andra punkten även c^{-1} . Enligt första punkten finns då också $c \star c^{-1} = e$.

□

Om H är än ändlig grupp, räcker det med att ett axiom uppfylls.

SATS 6. Låt H vara en icke-tom ändlig delmängd av en grupp G . Om H är begränsad under operationen i G , d.v.s. om

- Om $a, b \in H$, så gäller att $a \star b \in H$

så är H en undergrupp till G .

BEVIS. Enligt sats 5 behöver vi bara visa att inversen av varje element i H också är i H . Om $a \in H$, så kommer $a^k \in H$ för varje positivt heltal k . Eftersom H är ändlig, kan inte alla a^k vara olika. Alltså finns två heltal $m > n$ så att $a^m = a^n$. Detta kan förenklas till $a^{m-n} = e$. Vi betraktar nu två fall, det ena då $m - n = 1$, då är $a = e$ och därmed inversen till sig själv. Annars är $m - n > 1$, då kan vi skriva om uttrycket som $a^{m-n-1}a = e$, och eftersom inversen är unik, är $a^{-1} = a^{m-n-1}$, och därmed ligger inversen i mängden.

□

Med denna sats kan man generera undergrupper med så kallade *generatorer*. Denna sats är användbar då man vill generera permutationsgrupper för till exempel Rubiks kub. Man fastställer då ett antal permutationer som man genom att kombinera kan skapa alla godtyckliga manipulationer av klumpen. Man låter sedan lämpligen en dator applicera ovanstående sats på dessa permutationer och får då fram den permutationsgrupp som motsvarar Rubiks kub.

Cykliska grupper och generatorer

Låt G vara en grupp. Välj ut ett element a och bilda mängden $F = \{a^n | n \in \mathbb{Z}\}$. F består alltså av alla potenser av a . F utgör en undergrupp vilket lätt inses. Produkten av två potenser av a är också en potens av a och inversen till ett element a^k är a^{-k} som tillhör F . Det räcker att kontrollera dessa egenskaper för att vara säker på att det är en undergrupp.

Elementet a sägs vara en generator till gruppen F . F kan givetvis vara lika med G . En grupp som genereras av ett element sägs vara cyklisk. Alla cykliska grupper är abeliska eftersom $a^k \star a^n = a^n \star a^k$.

Man kan generalisera det ovanstående genom att ta fler än ett element som får fungera som generator. Låt således G vara en grupp. Välj ut en mängd M av element i G . Låt H vara undergruppen bestående av elementen som kan skrivas som produkter av potenser av elementen i M . Att H verkligen är en undergrupp kontrolleras på samma sätt som vi gjorde när vi bara hade en generator. Vi har dock ingen garanti att gruppen är abelisk.

Elementen i M sägs utgöra generatorerna för H . Vi kommer att använda generatorer i vårt arbete för att på ett smidigt sätt kunna beskriva grupper.

Om antalet element i en ändlig grupp är ett primtal så är gruppen cyklisk.

En tillämpning av gruppteorin

Med hjälp av gruppteori kan vi lösa pussel av samma sort som Rubiks kub. Vi väljer att numrera de olika rörliga delarna av den s.k. klumpen på ett lämpligt sätt enligt figur 1. Målet är att hitta en kombination av de olika rotationer man kan utföra som får alla delar att hamna på rätt plats. Med rätt plats menar vi att alla delar hamnar i nummerordning och med de vita på en sida och de svarta på andra sidan. Vi betraktar alla möjliga sätt placera delarna på klumpen som en grupp. Fördelen med att göra detta är att de vridningar vi kan utföra på klumpen också blir gruppens generatorer. Vi har använt oss av sex generatorer, som i praktiken är att vrida framsidan 90° medurs, baksidan 90° medurs, vrida de sex delarna på ovansidan ett halvt varv, och på samma sätt vrida delarna på högerkanten, vänsterkanten och undersidan. Med hjälp av numreringen av klumpen kan dessa vridningar skrivas som rena permutationer. Dessa är

$$W = (1, 3, 9, 7)(2, 6, 8, 4)$$

$$B = (11, 13, 19, 17)(12, 16, 18, 14)$$

$$U = (1, 11)(2, 12)(3, 13)$$

$$L = (1, 19)(4, 16)(7, 13)$$

$$D = (7, 17)(8, 18)(9, 19)$$

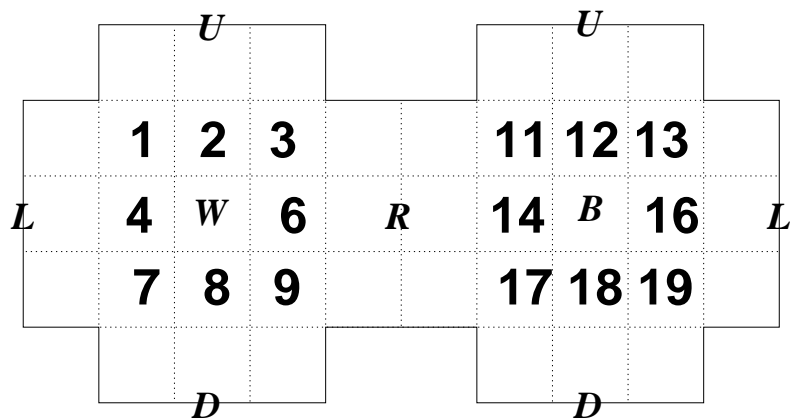
$$R = (3, 17)(6, 14)(9, 11)$$

Notera också att vår grupp blir en undergrupp av S_{16} .

Se figur 1 för en bild av de beteckningar vi använder för klumpen sidor.

GAP

GAP står för Groups, Algorithms and Programming. Det är gratis för nedladdning och användbart för beräkningar med grupper. Vi visar här hur vi



Figur 1: Beteckningarna för klumpens sidor. Figuren skall ses som en "utvikt" klump. Då klumpen är löst ligger den vita ettan på sida 1, den vita tvåan på sida 2, den vita fyran på sida 4, den svarta ettan på sida 11, den svarta tvåan på sida 12, den svarta fyran på sida 14, etc.

gick tillväga föra att lösa vårt problem, och på samma sätt kan liknande pussel lösas.

Vi beskriver de grundläggande rotationerna som permutationer.

```
gap> W:=(1,3,9,7)(2,6,8,4);
gap> B:=(11,13,19,17)(12,16,18,14);
gap> U:=(1,11)(2,12)(3,13);
gap> L:=(1,19)(4,16)(7,13);
gap> D:=(7,17)(8,18)(9,19);
gap> R:=(3,17)(6,14)(9,11);
```

Vi bildar en grupp med dessa som generatorer.

```
gap> G:=Group(W,B,U,L,D,R);
```

Vi kan räkna ut storleken på gruppen. Det som står utan 'gap>' är utskriften från GAP.

```
gap> Size(G);
1625702400
```

Vi får namnge vår grupps generatorer så som vi vill att de ska skrivas i lösningen. Här är det viktigt att namnen står i samma ordning som när man skapade gruppen.

```
gap> hom:=EpimorphismFromFreeGroup(
  G:names:=["w","b","u","l","d","r"]);
```

Och slutligen matade vi in den permutation som sätter alla siffror på rätt plats.

```
gap> PreImagesRepresentative(
  hom, (2,8,6,16,4,14,12,18)(3,7,9,13));
```

Utskriften är de generatorer man ska utföra för att uppnå den inskrivna permutationen, och således lösa klumpen. I utskriften använder GAP även inverser och potenser av generatorerna.

```
w^-1*l^-1*w^-2*l^-1*w^-2*l^-1*^-1*w^-1*l*u*
w*d*w^-1*u*w*d*w*^-1*l^-1*w*l*l^-1*w*l^-1*
u^-1*^-1*u^-1*w*l^-1*w^-2*r^-1*d^-1*r^-1*w*
d^-1*w*l^-1*b
```

Man kan förkorta lösningen genom att använda andra gruppteoretiska tekniker, som Gröbner-baser.

Referenser

Vi har främst använt följande bok:

Hungerford, T.W., *Abstract Algebra: An Introduction*, ISBN 0-03-030558-6, Saunders College Publishing, 1990.

GAP kan hämtas från:

<http://www-gap.mcs.st-and.ac.uk/>
2006-04-21

Webbplats med gruppteori:

<http://members.tripod.com/dogschool/>
2006-04-20

Insiration för lösningen av klumpen har vi hämtat från:

Radu, Silviu, *A New Upper Bound on Rubik's Cube*, Lund Institute of Technology, 2006.

Hyperreella tal

Per Ganestam, Martin Ericsson, Daniel Knös och Patrik Meijer

18 maj 2006

Innehåll

1	Världens största tal	3
2	Historia	4
3	Euler använder oändligheter	5
3.1	Bevis	5
4	Räkneregler för hyperreella tal	6
5	Ett hyperreellt talsystem	7
6	Appendix	11
6.1	Filter och ultrafilter	11

1 Världens största tal

Vilket är världens största tal? Kan det finnas ett tal som är större än alla andra tal ω . Finns det i så fall ett tal $\frac{1}{\omega}$ som är mindre än alla tal men ändå är större än noll? Vad ska man med ett tal till, kan man använda det i beräkningar och ungefär hur stort är det. Kan man inte lägga till ett och få ett ännu större tal. För att få svar på frågorna tar vi en tillbakablick och studerar hur vår matematik har blivit vad den är idag.

Vi är vana att se följande sekvens.

$$N \subseteq Z \subseteq Q \subseteq R \subseteq C$$

Det är den sekvensen som beskriver alla de olika tal som vi använder idag och deras relation till varandra. Symbolerna är i tur och ordning

- Naturliga tal N (1, 2, 3...).
- Heltal Z (... - 2, -1, 0, 1, 2...).
- Rationella tal Q ($\frac{p}{q}$, $p, q \in Z$).
- Reella tal R . Alla rationella tal men även tal som inte kan skrivas som en kvot av två heltal t.ex. (π , $\sqrt{2}$, e).
- Komplexa tal C ($a + bi$, $a, b \in R$, $i = \sqrt{-1}$).

Vad man kanske inte tänker på när man ser den här sekvensen är att den inte alltid har sett ut såhär. Innan man insåg de reella talens existens beskrevs våra tal istället av:

$$N \subseteq Z \subseteq Q,$$

och långt innan det bestod den bara av de naturliga talen och beskrevs då av ett ensamt N .

Allt eftersom människan har utvecklats har också hennes krav på levnadsstandard ökat. När behovet byts från att räkna stenar till att bygga flygplan måste nya verktyg till. Ett av dessa verktyg är ett fungerande matematiskt tänkande. När det enda problemet människan hade var att skilja ett antal från ett annat räckte det med de naturliga talen. I takt med utvecklingen behöver matematiken ändras och detta har lett oss till dagens definierade tal som sträcker sig från de naturliga talen till de komplexa, via heltalen, de rationella talen och de reella talen.

Men om nu matematiken genom alla dessa år ständigt har förändrats och utökats till att omfatta fler typer av tal, skulle då inte en ny definition av

tal kunna införas?

Det är här oändligheten, kommer in. Matematiker har ofta varit försiktiga med att använda oändligheter i sina beräkningar. Man har använt sig av gränsvärden som fått närma sig oändligheten men aldrig riktigt nått fram. Kan det vara så att det finns ett sätt att använda sig av oändligheten när man räknar? Idén har funnits länge och talen har sammanfattats i den mängd som kallas de hyperreella talen *R . De hyperreella talen är en utvidgning av de reella talen, dvs

$$R \subseteq {}^*R.$$

2 Historia

Oändligheten har alltid förbryllat människan. Hur beskriver man en oändlighet? Finns det olika grader av oändligheter? Ni har säkert hört om sköldpaddan och kaninen som aldrig kommer i mål. Detta filosofiskt- matematiska problem leker med oändligheten på ett vis som ger helt absurda konsekvenser. Archimedes använde sig av oändligt små bitar då han uppskattade en cirkels area och följaktligen kunde bestämma ett närmevärde till pi. Archimedes vågade dock inte att betrakta det som matematiskt gediget och utvecklade inte sina tankar. Archimedes verkade redan 300 f.kr. men det var inte förrän långt senare som infinitesimalen myntades som begrepp och fick de definitioner som idag gäller. Gottfried Wilhelm Leibnitz lade under 1600-talet grunderna till de beteckningar som nu används vid derivering och integralkalkyl t.ex. $\frac{d}{dx}F(x)$ samt $\int f(x)dx$ men vem som ska ha äran för metoderna är lite oklart. Det verkar nämligen som att Isaac Newton och Leibnitz ovetande om varandras arbete gjorde detta på var sitt håll. Leibniz, som publicerade sitt arbete först, fick till en början allmänhetens erkännande men Newton kunde bevisa att han arbetat med det två decennier tidigare vilket ledde till att Leibniz blev anklagad för att ha plagierat Newton. Långt innan varken Newton eller Leibniz hade publicerat något inom ämnet så hade Newton diskuterat det med personer som stod i kontakt med Leibniz, vilket förstås stödde teorin om plagiat.

Hur det egentligen ligger till är det ingen som vet men det är säkert att både Leibniz och Newton har bidragit oerhört mycket till differentialekalkylen. Både derivering och integralkalkylen bygger lite förenklat på att man delar upp något i oändligt små delar, s.k. infinitesimaler. Infinitesimalen definieras som det tal som är mindre än alla reella tal men större än noll och den betecknas oftast med lilla epsilon, ϵ .

Efter både Newton och Leibniz död yttrade sig den matematiskt kunnige biskop George Berkeley kritiskt om derivationsreglerna. Berkeley menade att

teorierna inte var matematiskt korrekta eftersom beräkningarna medför att man först betraktar infinitesimalen som ett reellt tal om än oändligt litet, för att sedan betrakta infinitesimalen som noll. Han menade att inkonsekvensen gör det origoröst. Denna kritik ledde till att man lämnade infinitesimalen och började använda gränsvärden istället. Infinitesimalen återupptogs inte förrän på 1960-talet i och med att Abraham Robinson gjorde verklighet av icke-standardanalysen, och det var inte förrän då som den fick just beteckningen infinitesimal och de regler som används nu.

3 Euler använder oändligheter

En matematiker som med stor framgång använde oändligheter i sina beräkningar var schweizaren Leonhard Euler (1707-1783). Bland annat lyckades han med hjälp av infinitesimaler bevisa formeln för Taylorutvecklingen av sinus och cosinus. Nedan följer cosinusbeviset.

3.1 Bevis

Enligt Eulers formel kan $\cos x$ skrivas $\frac{e^{ix} + e^{-ix}}{2}$. Om man här ersätter x med $n\theta$ får man istället

$$\cos x = \cos n\theta = \frac{e^{i\theta n} + e^{-i\theta n}}{2} = \frac{(e^{i\theta})^n + (e^{-i\theta})^n}{2}.$$

Man kan utveckla $e^{i\theta}$ till $\cos \theta + i \sin \theta$ och genom att ersätta termerna i täljaren med det uttrycket får vi ett nytt uttryck som kan skrivas:

$$\frac{(e^{i\theta})^n + (e^{-i\theta})^n}{2} = \frac{(\cos \theta + i \sin \theta)^n + (\cos \theta - i \sin \theta)^n}{2}.$$

Med hjälp av binomialsatsen kan täljaren förenklas något.

$$\begin{aligned} & (\cos \theta + i \sin \theta)^n + (\cos \theta - i \sin \theta)^n = \\ & = \sum_{k=0}^n \binom{n}{k} \cos^{n-k} \theta \cdot i^k \sin^k \theta + \sum_{k=0}^n \binom{n}{k} \cos^{n-k} \theta \cdot (-i)^k \sin^k \theta \end{aligned}$$

Det man ser nu är att alla termer där k är udda, dvs de med en komplex del kommer att ta ut varandra och försvinna. Alla andra termer dubblas men skall sedan delas med två och resultatet blir därför:

$$\cos^n \theta - \frac{n(n-1) \cos^{n-2} \theta \sin^2 \theta}{2!} + \frac{n(n-1)(n-2)(n-3) \cos^{n-4} \theta \sin^4 \theta}{4!} \dots$$

Om vi nu låter n vara oändligt stort så blir $\theta = \frac{x}{n}$ infinitesimalt litet. Då blir $\cos \theta = 1$ och $\sin \theta = \theta = \frac{x}{n}$. Vidare så kan man, eftersom att n är

oändligt stort, sätta $n - 1 = n$. Detsamma kan göras med $n - 2$, $n - 3$ och så vidare. Då kan vi nu ändra om i vårt uttryck och får istället

$$\cos x = \cos n\theta = 1^n - \frac{n^2 1^{n-2} (\frac{x}{n})^2}{2!} + \frac{n^4 1^{n-4} (\frac{x}{n})^4}{4!} \dots = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} \dots$$

vilket är Taylorutvecklingen av $\cos x$, som vi nu har lyckats bevisa genom att räkna med oändligheter. På liknande sätt kan Taylorutvecklingen för $\sin x$ visas.

4 Räknerregler för hyperreella tal

En infinitesimal definieras som ett tal vars absolutbelopp är mindre än alla reella tal men större än noll. Med hjälp av den definitionen kan man fastställa olika räknerregler för att möjliggöra en användning av talen.

Till att börja med, om ϵ, δ båda är infinitesimala tal så är deras summa och produkt också infinitesimalt liten.

Bevis Låt r vara ett reellt tal större än noll. Då är $|\epsilon| < \frac{r}{2}$ och $|\delta| < \frac{r}{2}$. Det ger att $|\epsilon + \delta| < r$, och summabeviset är färdigt eftersom r kan väljas hur litet som helst. För produkten kan man istället skriva $|\epsilon| < \sqrt{r}$, och $|\delta| < \sqrt{r}$, vilket ger att $|\epsilon\delta| < r$.

Produkten av ett reellt tal och en infinitesimal är också infinitesimal. Låt ϵ vara infinitesimal och b ett ändligt reellt tal. Då är $|b| < s$ för något reellt $s > 0$. Dessutom är $|\epsilon| < \frac{r}{s}$ för något $r > 0$. Då följer att $|\epsilon b| < r$ och därför är produkten infinitesimal.

Om $(x + \epsilon) \in {}^*R$ och ϵ är infinitesimalt litet och x är ett godtyckligt reellt tal så är $st(x + \epsilon) = x$ där st är en förkortning för standarddel. Om x och y är hyperreella tal så följer:

$$(1) \quad st(x \pm y) = st(x) \pm st(y),$$

$$(2) \quad st(xy) = st(x)st(y),$$

$$(3) \quad st\left(\frac{x}{y}\right) = \frac{st(x)}{st(y)},$$

Bevis av (1) Om $x = x' + \epsilon$, och $y = y' + \delta$ så är $x + y = (x' + y') + (\epsilon + \delta)$. Summan av två infinitesimaler är också infinitesimal, av detta följer första påståendet.

Om x och y är två hyperreella tal och $x - y$ är en infinitesimal så är x och y infinitesimalt nära, vilket skrivs $x \simeq y$. Om $x \simeq x'$ och $y \simeq y'$ så följer:

$$(1) \quad x \pm y \simeq x' \pm y',$$

$$(2) \quad xy \simeq x'y',$$

$$(3) \quad \frac{x}{y} \simeq \frac{x'}{y'}, \quad y \neq 0.$$

5 Ett hyperreellt talsystem

För att konstruera ett hyperreellt talsystem, *R , vill vi att följande egenskaper ska vara uppfyllda för det.

(1) Det innehåller de reella talen.

(2) Det innehåller en infinitesimal ϵ .

(3) Varje mening i språket S som är sann i *R är sann i R .

För att förklara vad ett språk innebär är det lättast att titta på exempel av språk. Låt oss kalla vårt språk S . Språket S gäller för de reella talen och innehåller konstanter, variabler, funktioner och relationer.

Man kan konstruera de hyperreella talen genom att betrakta sekvenser av reella tal så som

$$\langle 4, 2, 1, 1/2, 1/3, 1/4, \dots \rangle$$

$$\langle 13, -45, 13, -45, 13, -45, \dots \rangle$$

Tar vi det reella talet $\sqrt{2}$, skulle dess sekvens bli

$$\langle \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \dots \rangle$$

Två olika sekvenser kommer att kunna representera samma hyperreella tal. Ta sekvenserna

$$p = \langle 3, 2, 3, 19, 43, 6, 6, 6, 6, 6, \dots \rangle$$

$$j = \langle 7, 1, 2, 6, 6, 6, 6, 6, 6, 6, \dots \rangle$$

Vi säger att sekvenserna p och j representerar samma hyperreella tal eftersom de är lika på nästan alla ställen. Vi säger att de är lika då mängden av de ställen där de inte är lika är ändlig. Tyvärr är inte detta tillräckligt starkt för att bygga upp vårt hyperreella talsystem. Därför behöver vi skapa ett ultrafilter. Detta ultrafilter innehåller bl.a. just de mängder vars komplement är ändliga. En utförligare diskussion kring ultrafilter ges i appendix.

Som tidigare nämnt representerades de hyperreella talen av sekvenser, om vi bildar sekvenserna

$\langle a_1, a_2, a_3, \dots \rangle$

och

$\langle b_1, b_2, b_3, \dots \rangle$

kommer dessa att representera samma hyperreella tal omm

$$\{n \in N; a_n = b_n\} \in U.$$

För att sedan gå vidare och se hur relationer och funktioner fungerar på de hyperreella talen, börjar vi med att välja en allmän funktion f som gäller i vårt språk S . Vi tar sedan två hyperreella tal a och b , dvs sekvenserna $\langle a_1, a_2, a_3, \dots \rangle$ och $\langle b_1, b_2, b_3, \dots \rangle$ och bildar $f(a)$ och $f(b)$, som vi definierar till

$$\langle f(a_1), f(a_2), f(a_3), \dots \rangle$$

respektive

$$\langle f(b_1), f(b_2), f(b_3), \dots \rangle$$

Om vi låter a och b representera samma hyperreella tal, vill vi att $f(a)$ och $f(b)$ också gör det. För att kontrollera att de gör det bildar vi mängderna

$$A = \{n \in N; a_n = b_n\}$$

och

$$B = \{n \in N; f(a_n) = f(b_n)\}.$$

Vi vet sedan tidigare att $A \in U$ och vi inser snabbt att $A \subseteq B$. Detta ger att $B \in U$ och därmed har vi visat att $f(a)$ och $f(b)$ representerar samma tal.

För att sedan gå vidare och definiera vad vi menar med relationer för de hyperreella talen, börjar vi med en relation som vi skriver som $R()$. För att sedan avgöra om $R(a)$ är sann definierar vi helt enkelt att $R(a)$ är sann då $\{n \in N; R(a_n) \text{ är sant}\} \in U$. För att avgöra om vår definition är okej låter vi som tidigare a och b vara samma hyperreella tal, då måste det förhålla sig så att $R(a)$ är sant omm $R(b)$ är sant. För att visa att så också är fallet bildar vi mängderna

$$A = \{n \in N; R(a_n) \text{ är sant}\}$$

$$B = \{n \in N; R(b_n) \text{ är sant}\}$$

och

$$C = \{n \in N; a_n = b_n\}.$$

Om vi antar att $R(a)$ är sant. Då gäller att $A \in \mathcal{U}$ och vi vet sen tidigare att $C \in \mathcal{U}$. Enligt definitionen av filter och ultrafilter gäller då även att $A \cap C \in \mathcal{U}$ och eftersom $A \cap C \subseteq B$, så gäller att $B \in \mathcal{U}$, se appendix.

Nu när vi definierat vad vi menar med ett hyperreellt tal vill vi visa att de verkligen uppfyller de tre egenskaperna vi proklamerade i början av detta avsnitt.

Egenskap 1 visar vi enkelt genom att låta ett reellt tal r , representeras med sekvensen $\langle r, r, r, r, r, r, \dots \rangle$. Som exempel kan vi ta det reella talet 4, som då kommer att representeras av sekvensen $\langle 4, 4, 4, 4, 4, 4, \dots \rangle$.

Den andra egenskapen var att vårt system skulle innehålla en infinitesimal ϵ , där $0 < \epsilon < r$ för alla reella tal r . Bildar vi sekvensen

$$\epsilon = \langle 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \dots \rangle$$

ser vi att $\epsilon > 0$ eftersom om vi låter 0 representeras av sekvensen $p = \langle 0, 0, 0, 0, 0, 0, \dots \rangle$, ser vi att $\{n \in N; \epsilon_n < p_n\}$ blir alla naturliga tal och således tillhör mängden ultrafiltret. Vidare ser vi att olikheten $\epsilon < r$ även gäller för alla reella tal r , eftersom om vi låter r representeras av sekvensen $q = \langle r, r, r, r, r, r, \dots \rangle$ och bildar mängden

$$A = \{n \in N; \epsilon_n < q_n\}$$

Oavsett hur litet vi väljer vårt reella tal r kommer vi alltid att kunna hitta ett heltal m så att olikheten $\frac{1}{m} < r$ är uppfylld. Då gäller även att

$$\langle \frac{1}{m} < r, \frac{1}{m+1} < r, \frac{1}{m+2} < r, \dots \rangle$$

Bildar vi sedan mängden

$$B = \{m, m+1, m+2, m+3, \dots\}$$

ser vi att $B \in \mathcal{U}$ och även att $B \subseteq A$, därför gäller $A \in \mathcal{U}$ och därmed är vi klara.

Tredje och sista egenskapen är den svåraste att visa, tittar vi på egenskaperna hos språket S ser vi att det innehåller följande:

- (1) En konstantsymbol r för alla reella tal.
- (2) En funktionssymbol f för alla reella funktioner.
- (3) En relationssymbol R för alla reella relationer.

Vi vill sedan bygga ut detta språk genom att lägga till en konstant symbol för alla hyperreella tal. Låt oss kalla det nya språket för $*S$. Om vi i detta språk låter de hyperreella talen representeras av symbolerna $j_1, j_2, j_3, \dots, j_k$, där varje j_i egentligen bara är sekvenser av reella tal. Så innebär detta att $j_{1_n}, j_{2_n}, j_{3_n}, \dots, j_{k_n}$ bara innehåller reella tal. Nu kan vi göra följande definition.

Om $G = G(j_1, j_2, j_3, \dots, j_k)$ är en formel i $*S$, (exempel på en formel är $E = mc^2$, sekvensen inom parantesen visar bara vilka symboler formeln gäller för) så gäller att $G_n = G(j_{1_n}, j_{2_n}, j_{3_n}, \dots, j_{k_n})$ är motsvarande formel i språket S .

Det slutgiltiga satsen som ger oss vårt hyperreella talsystem lyder:

Om G är en formel i $*S$, så är G sann på $*R$ om och endast om $\{n \in N; G_n \text{ är sant i } R\}$ tillhör ultrafiltret.

I denna text kommer inte något bevis för satsen att ges, men om vi antar att den är sann, vilket den är, så har vi visat egenskap 3. Därmed är vårt talsystem klart.

6 Appendix

6.1 Filter och ultrafilter

För att bygga upp och hantera de hyperreella talen krävs ett ultrafilter. Först efter att man valt sitt ultrafilter kan man avgöra huruvida två talföljder $\langle x_n \rangle$ och $\langle y_n \rangle$ är ekvivalenta. Det är de nämligen omm mängden $\{n \in N; x_n = y_n\}$ tillhör ultrafiltret.

Definition En icke-tom mängd \mathcal{F} av delmängder till I är ett filter på I om,

- (1) $\emptyset \notin \mathcal{F}$
- (2) $A, B \in \mathcal{F}$ medför att $A \cap B \in \mathcal{F}$
- (3) $A \in \mathcal{F}$ och $B \supseteq A$ medför att $B \in \mathcal{F}$

Familjen av delmängder till N med ändligt komplement uppfyller villkoren (1), (2) och (3) och kan kallas för ett filter. Just detta filtret är känt som Frechetfiltret och är det används för vad som kallas den svaga ekvivalensen. Dessvärre räcker det inte med ett filter av denna uppbyggnad i alla situationer. Vad som behövs då är ett "maximalt" filter, ett ultrafilter. En avgörande egenskap hos ultrafiltret är,

- (4) Ett filter \mathcal{U} är ett ultrafilter omm för varje delmängd A av I gäller antingen $A \in \mathcal{U}$ eller komplementet $A' = I - A \in \mathcal{U}$.

Om man ser på den svaga ekvivalensen, utan ultrafilter, är $[\langle x_n \rangle] = [\langle y_n \rangle]$ omm $\#\{n \in N; x_n \neq y_n\} < +\infty$. Det går även att garantera en multiplikativ invers för varje hyperreellt tal endast med den svaga ekvivalensen. För en totalt ordnad kropp där man alltid kan avgöra vilket av två tal som är störst krävs den starkare formen av ekvivalens. Då använder man sig av ett ultrafilter. Den svaga ekvivalensen kan t.ex. inte avgöra vilken av sekvenserna $\langle x_n \rangle = \langle 1, 0, 1, 0, \dots \rangle$ och $\langle y_n \rangle = \langle 0, 1, 0, 1, \dots \rangle$ som är störst och vilken som är ekvivalent med vilket tal. Enligt (4) gäller antingen $\{n \in N; x_n = 0\} \in \mathcal{U}$ eller $\{n \in N; x_n = 1\} \in \mathcal{U}$, men aldrig båda. Av första alternativet ges $[\langle x_n \rangle]$ ekvivalent med nollsekvensen och därav $[\langle y_n \rangle]$ ekvivalent med ettsekvensen, den andra möjligheten ger naturligtvis omvänt resultat. Vilket av fallen som gäller beror på hur ultrafiltret ser ut. Det går inte att avgöra vilket ultrafilter som används men det är inte heller relevant. Det intressanta är vetenskapen om att det ena eller det andra gäller, men inte båda.

Ultrafilteraxiomet Om \mathcal{F} är ett filter av delmängder till N , då finns det ett ultrafilter \mathcal{U} vilket innefattar \mathcal{F} . Ultrafiltret är då ett maximalt filter.

Ultrafilteraxiomet går att härleda med hjälp av Zorns lemma eller med det ekvivalenta urvalsaxiomet.

Zorns lemma Låt X vara en partiellt ordnad mängd, om varje kedja (linjärt ordnade delmängder) är uppåt begränsad, så har X åtminstone ett maximalt element.

Urvalsaxiomet För varje mängd A av icke-tomma mängder finns det en funktion f sådan att $f(x) \in x$ för varje $x \in A$. Funktionen f kallas en valfunktion.

Man kan säga att valfunktionen väljer ett element ur varje delmängd av en godtycklig mängd.

Det är mestadels Zorns lemma och urvalsaxiomet som anses vara kontroversiellt inom icke-standardanalysen. En stor kontrovers med urvalsaxiomet är huruvida det handlar om att det "existerar" en valfunktion eller om att "finna" en valfunktion. För att godta axiomet får man intuitivt tänka sig att det existerar en valfunktion men acceptera att man inte kommer hitta den. För delmängder av N kan man tänka sig en valfunktion f där $f(S)$ är det minsta elementet i S . Det går att fundera ut mer avancerade valfunktioner, även på den reella tallinjen. Problemet blir när man vill skapa en valfunktion som gäller för alla icke-tomma delmängder av R . Ingen har ännu hittat en sådan funktion och det finns teorier som menar att den aldrig kommer hittas. Detta utesluter dock inte existensen av funktionen.

En annan märklig uppkomst efter urvalsaxiomet är Banach-Tarskis paradox. Matematiskt är det egentligen ingen paradox men den mänskliga intuitionen har svårt att acceptera den. Enligt "paradoxen" kan man skapa ett klot, dela upp det i ändligt många bitar för att sedan pussla bitarna till två nya klot båda identiska med ursprungsklotet. Detta beror på att man kan skära sitt klot så att bitarna blir omätbara. När man sätter ihop sina omätbara bitar kan det göras så man får två nya klot identiska med ursprunget.

Matematisk kommunikation

Konvex optimering och igenkänning

Jacob Gumpert, Gustaf Pettersson,
Carl-Johan Rosenvinge, Sara Tapper
Teknisk matematik Vt-05, Lunds tekniska högskola

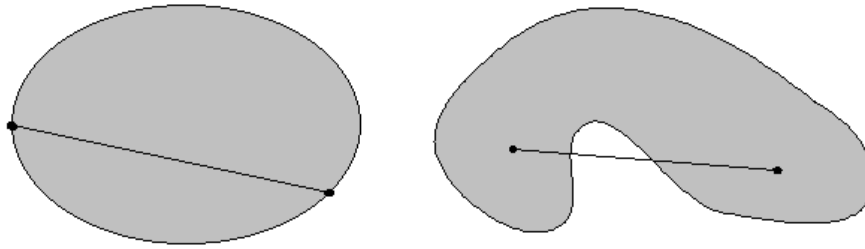
18 maj 2006

Sammanfattning

Detta arbete syftar till att ge läsaren en grundläggande förståelse för vad konvex optimering är, vilka teorier den bygger på samt vad den kan användas till. I detta fallet har användningsområdet igenkänning valts. Inledningsvis presenteras teori kring konvexa mängder och konvexa funktioner. Därefter följer ett avsnitt om konvex optimering och slutligen ett avsnitt om igenkänning. För att illustrera tillämpning av konvex optimering visas först ett syntetiskt exempel. Därefter beskrivs hur konvex optimering kan användas vid ansiktsdetektion.

Innehåll

1	Inledning	3
2	Konvexa mängder	3
2.1	Hyperplan	4
2.2	Separation av konvexa mängder	5
3	Konvexa Funktioner	6
3.1	Konvexa funktioner av en variabel	6
3.2	Konvexa funktioner av flera variabler	7
3.3	Kontinuitet	8
4	Konvex Optimering	8
4.1	Mönsterigenkänning	9
4.2	Klassificering	9
5	Syntetiskt exempel	10
6	Ansiktsdetektion	10
7	Slutsatser	13



Figur 1: Till vänster: En konvex mängd. Till höger: En ickekonvex mängd.

1 Inledning

Att lösa optimeringsproblem är detsamma som att hitta ett minsta eller största värde till en funktion. Optimeringen av ett problem ger den lösning som är bäst att använda för en viss situation. Det är dock långt ifrån alltid möjligt att genomföra en optimering på en problemuppställning, och dessutom finns väldigt många olika metoder.

I det här arbetet behandlas en viss klass av optimeringsproblem som kallas för konvex optimering, se [2, 4, 1]. Ett konvext optimeringsproblem består i att minimera en konvex funktion över en konvex mängd. Dessa begrepp beskrivs inledningsvis i arbetet, och därefter introdueras läsaren till användningen av konvex optimering. Konvexitet är en egenskap hos mängder och funktioner som avsevärt förenklar arbetet. Huvudsakligen underlättas optimeringen av att konvexitet medför att ett lokalt minimum är ett globalt minimum, vilket visas i arbetet.

Det är uppenbart att denna egenskap visar sig mycket användbar i skrivandet av olika algoritmer med konvex optimering. Konvex optimering kan bland annat användas i stödvektormaskiner till igenkänning, se [3], vilket kommer beskrivas utförligare senare. Avslutningsvis visas ett exempel på hur konvex optimering kan användas till ansiktsgigenkänning, se [5].

2 Konvexa mängder

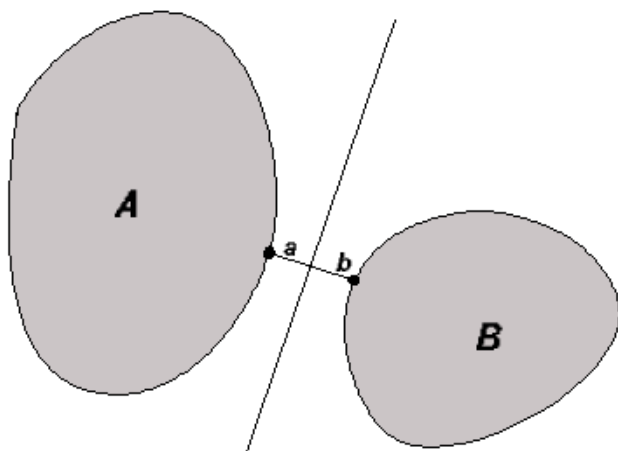
Ett fundamentalt begrepp som behöver definieras för att förstå konvex optimering är konvex mängd. En enkel definition på en konvex mängd är:

Definition 1. *En delmängd av \mathbb{R}^n kallas konvex om sträckan mellan två godtyckliga punkter i mängden helt ligger i mängden.*

Denna definition förstås lättast genom att se vad detta betyder geometriskt. Som vi kan se i figur 1 betyder det att mängden är konvex om den räta linje som dras mellan de två punkterna helt ligger i mängden. För att kunna utveckla detta ytterligare behövs en mer matematisk definition av konvex mängd.

Definition 2. *En mängd $S \subseteq \mathbb{R}^n$ är konvex om*

$$\mathbf{a}, \mathbf{b} \in S \implies \lambda \mathbf{a} + (1 - \lambda) \mathbf{b} \in S \text{ för alla } \lambda \in [0, 1].$$



Figur 2: Ett hyperplan separerar två konvexa mängder.

Detta innebär att en konvex mängd innehåller alla tänkbara linjärkombinationer av punkter i mängden, förutsatt att alla koefficienter ≥ 0 , och att summan av koefficienterna = 1. Exempel på konvex mängd i det tredimensionella fallet är klot, och hyperplan.

2.1 Hyperplan

Ett hyperplan i \mathbb{R}^n är av dimension \mathbb{R}^{n-1} , och delar därmed en mängd i \mathbb{R}^n i två delar. Hyperplanet i två dimensioner är en linje, och hyperplanet i tre dimensioner är ett plan. Man kan på följande sätt bestämma ekvationen till ett hyperplan. Om planet normalvektor betecknas \mathbf{p} och en punkt som finns i planet ges beteckningen $\tilde{\mathbf{x}}$ så kan hyperplanet skrivas

$$\mathbf{p}^T(\mathbf{x} - \tilde{\mathbf{x}}) = 0,$$

där \mathbf{p}^T är transponatet till \mathbf{p} . Det är nödvändigt att transponera en av två vektorer som ska multipliceras. På det sättet bli den ena en radvektor och den andra en kolonnvektor, och multiplikationen resulterar i en skalär. Genom att ersätta $\mathbf{p}^T\tilde{\mathbf{x}}$ med α i ekvationen ovan fås

$$\mathbf{p}^T\mathbf{x} = \alpha.$$

I fallet då $\mathbf{p}^T\mathbf{x} = 0$ så delas mängden i två på varsin sida av origo. Detta kan nu beskrivas formellt.

Definition 3. Låt A och B vara delmängder av \mathbb{R}^n . Hyperplanet $\mathbf{p}^T\mathbf{x} = \alpha$ delar dessa om

$$\mathbf{p}^T\mathbf{x} > \alpha \text{ då } x \in A$$

$$\mathbf{p}^T\mathbf{x} < \alpha \text{ då } x \in B$$

och tvärt om.

2.2 Separation av konvexa mängder

Med hjälp av de nu introducerade begreppen är det möjligt att följa ett fortsatt resonemang om de egenskaper som finns för konvexa mängder. Satsen som nu följer visar hur en konvex mängd förhåller sig till en punkt som finns utanför mängden.

Sats 1. *Låt M vara en sluten, konvex, samt icke tom mängd i \mathbb{R}^n . Antag att det finns en punkt $\mathbf{a} \notin M$. Då finns en entydigt bestämd punkt $\tilde{\mathbf{b}} \in M$ så att*

$$|\mathbf{a} - \tilde{\mathbf{b}}| = \min_{\mathbf{x} \in \mathbb{R}^n} |\mathbf{a} - \mathbf{x}|,$$

för alla $\mathbf{x} \in M$.

Bevis. De exakta detaljerna i beviset anses inte relevanta för fortsatt förståelse, men beviset är indelat i tre steg. Först visas att ett minsta avstånd existerar, sedan visas att det avståndet finns just för punkten $\tilde{\mathbf{b}}$, och till sist visas att det inte finns någon annan punkt $\mathbf{x} \in M$ med ett kortare avstånd. \square

Det är nu sagt att hyperplan kan separera mängder, samt att konvexa mängder har ett entydigt bestämt minsta avstånd till punkter som ligger utanför mängden. Det ska snart visas att ett hyperplan används med fördel för att separera den konvexa mängden från punkter utanför mängden, vilka kan betecknas \mathbf{a} . Om hyperplanet väljs mittemellan punkterna \mathbf{a} och $\tilde{\mathbf{b}}$ vinkelrät mot den entydigt kortaste sträckan mellan mängden och \mathbf{a} så hamnar givetvis alla andra punkter i mängden på ett längre avstånd från hyperplanet.

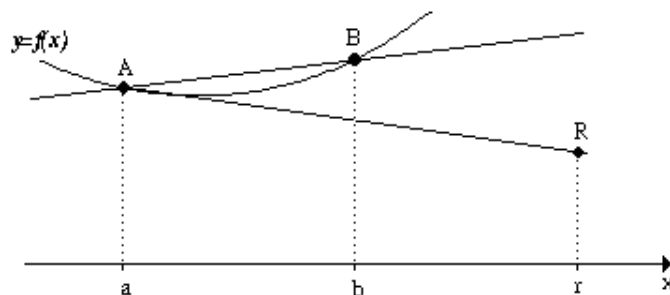
Sats 2. *Låt M vara en sluten, konvex, samt icke tom mängd i \mathbb{R}^n . Antag att det finns en punkt $\mathbf{a} \notin M$. Då har punkten \mathbf{a} och mängden M ett separerande hyperplan.*

Bevis. Det ska visas att $\mathbf{a} > \mathbf{p}^T \mathbf{x} - \alpha$, samt att $M < \mathbf{p}^T \mathbf{x} - \alpha$, eftersom planet ska dela mängden och punkten enligt definition 3. Som tidigare konstaterats skrivs hyperplanets ekvation $\mathbf{p}^T (\mathbf{x} - \tilde{\mathbf{x}}) = 0$. För enkelhetens skull väljs planets placering precis på mitten av det kortaste avståndet mellan M och \mathbf{a} , alltså på halva sträckan $(\mathbf{a} - \tilde{\mathbf{b}})$. Då kan $\frac{1}{2}\mathbf{p}^T(\mathbf{a} + \tilde{\mathbf{b}})$ ersättas med α . Det kan visas att $\mathbf{p}^T \mathbf{a} > \alpha$, och med hjälp av sats 1 kan det visas att om $\mathbf{x} \in M$ så är $\mathbf{p}^T \mathbf{x} < \alpha$. Detaljerna för beviset utelämnas, men det är då enligt definition 3 klart att hyperplanet separerar punkten \mathbf{a} och mängden M . \square

Det är nu uppenbart att om \mathbf{a} finns i mängden $Y = \{\mathbf{a}\}$ så separerar hyperplanet de två mängderna Y och M vilket visas i figur 2.

Vid tillämpning av konvexa mängder i optimeringsproblem måste ofta den konvexa mängden begränsas av olika bivillkor. Definitionsmängden blir då snittet av olika konvexa mängder. Eftersom definitionsmängden i ett konvext optimeringsproblem förutsätts vara en konvex mängd är det ett krav att snittet är en konvex mängd, vilket visas i följande sats.

Sats 3. *Snittet av ett godtyckligt antal konvexa mängder i \mathbb{R}^n är en konvex mängd.*



Figur 3: Om funktionen är konvex kan det inte existera något R som ligger under den räta linjen genom A och B

Bevis. Låt A och B vara två konvexa mängder. Antag att $A \cap B$ inte är konvex. Alltså finns det två punkter \mathbf{a} och \mathbf{b} i $A \cap B$ som inte uppfyller villkoret

$$\lambda \mathbf{a} + (1 - \lambda) \mathbf{b} \in A \cap B, \quad \lambda \in [0, 1]. \quad (1)$$

Eftersom hela $A \cap B$ ligger i både A och B , vilka båda är konvexa mängder, måste dock (1) uppfyllas och vi får en motsägelse. Alltså måste snittet $A \cap B$ vara konvext. Beviset kan upprepas analogt för ett godtyckligt antal mängder. \square

3 Konvexa Funktioner

Funktionen som minimeras i ett konvext optimeringsproblem definieras på följande sätt.

Definition 4. Låt f vara en funktion av en variabel definierad på ett intervall I . Funktionen f är konvex om

$$\begin{cases} a, b \in I \\ 0 < \lambda < 1 \end{cases} \Rightarrow f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda)f(b)$$

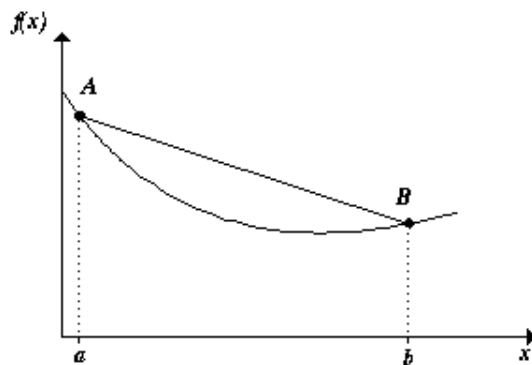
och strängt konvex om

$$\begin{cases} a, b \in I \\ 0 < \lambda < 1 \end{cases} \Rightarrow f(\lambda a + (1 - \lambda)b) < \lambda f(a) + (1 - \lambda)f(b).$$

3.1 Konvexa funktioner av en variabel

Att en funktion är konvex är ekvivalent med att en rät linje mellan punkterna $(a, f(a))$ och $(b, f(b))$ ligger över funktionskurvan mellan $(a, f(a))$ och $(b, f(b))$. Det kan även ses som att epigrafen, alltså den mängd som ligger över kurvan, skall vara en konvex mängd.

Lemma 1. Låt f vara konvex på intervallet I och låt $a, b \in I$. Då följer att om $x < a$ eller $x > b$ att kurvan $y = f(x)$ ligger över den räta linjen L som går genom $(a, f(a))$ och $(b, f(b))$.



Figur 4: Om funktionen är konvex och $f(a) < f(b)$ kan inte funktionen ha ett lokalt minimum i $f(a)$.

Bevis. Betrakta intervallet $x > b$. Antag att det finns en punkt $(r, f(r))$ på kurvan där $r > b$ och ligger strikt under linjen L . Definitionen av konvexitet kräver att kurvan mellan a och r ligger under linjen L , men punkten $(b, f(b))$ motsäger det. \square

Sats 4. Om en konvex funktion f har en lokal minimipunkt i punkten a på intervallet I , följer att a även är global minimipunkt.

Bevis. Låt a vara en inre punkt på intervallet. Genom att anta att det finns en omgivning $a - \delta < x < a + \delta$ så att $f(x) \geq f(a)$ för alla x i denna omgivning. Betrakta linjen som skär punkterna på kurvan med x-koordinater a och $a + \frac{\delta}{2}$. Givet är $f(a + \frac{\delta}{2}) \geq f(a)$, alltså att linjens riktningskoefficient är större än eller lika med noll. Av lemmat följer att kurvan är över linjen för $x \geq a + \frac{\delta}{2}$. Det ger att

$$f(x) \geq f(a) \text{ för alla } x \in I \text{ med } x \geq a.$$

\square

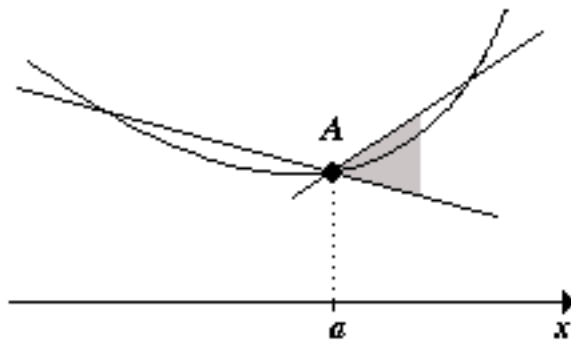
Fallet då $x \leq a$ följer analogt. Detta bevis är även giltigt då a är en ändpunkt på intervallet.

3.2 Konvexa funktioner av flera variabler

Sats 5. Om en konvex funktion f har en lokalminimipunkt i punkten a i den konvexa mängden M , följer att a även är global minimipunkt.

Bevis. Låt a vara en lokal minimipunkt i M och låt f vara en konvex funktion. Antag att det finns en punkt $b \in M$ där $f(b) < f(a)$. Det finns ett linjestycke mellan punkterna a och b som beskrivs av $x = (1 - \lambda)a + \lambda b$ där $\lambda \in [0, 1]$, då mängden är konvex $x \in M$ för alla λ .

Betrakta intervallet $\lambda \in [0, 1]$. Från definitionen av konvexa funktioner följer att funktionskurvan ligger under kordan mellan $f(a)$ och $f(b)$ alltså är för alla $t > 0$ $f(a + t) < f(a)$ och a kan då inte vara en lokal minimipunkt, vilket är en



Figur 5: Funktionen måste ligga i det markerade området.

motsägelse. Om a är en lokal minimipunkt existerar det inget b där $f(b) < f(a)$, följaktligen är a en global minimipunkt. \square

Ovanstående bevis inses enkelt med hjälp av figur 5. Vad som sagts är att om det finns en minimipunkt $f(b)$ som är mindre än minimipunkten $f(a)$ så ska närliggande punkter till $f(a)$ uppfylla två villkor. För det första ska de vara större än $f(a)$, och för det andra ska de ligga under en linje som sammanbinder $f(a)$ och $f(b)$. Det framgår i figuren att inga punkter kan uppfylla dessa två villkor samtidigt.

3.3 Kontinuitet

Som kuriosas visas nu en intressant egenskap angående konvexa funktioner och kontinuitet, detta avsnitt är ej av avgörande betydelse för fortsatt förståelse.

Sats 6. Om funktionen f är konvex på intervallet I är den kontinuerlig på det öppna intervallet I .

Bevis. Låt a vara en inre punkt i I och låt $A = (a, f(a))$. Eftersom A är en inre punkt kan två kordor dras genom A , en till vänster och en till höger. Ur definitionen av konvexitet och lemma 1 följer att kurvan $f(x)$ ligger under en korda mellan A och annan godtycklig punkt på kurvan, till höger om A . Eftersom kordorna skär varandra i A gäller att kurvan $f(x)$ ligger i ett område till höger om A mellan de två kordorna, vilket innebär att

$$\lim_{x \rightarrow a^+} f(x) = f(a).$$

Detta kan visas analogt för ett område till vänster om A , alltså $\lim_{x \rightarrow a^-} f(x) = f(a)$. Därmed är f kontinuerlig i a . \square

4 Konvex Optimering

Optimeringsproblem som behandlar konvexa funktioner med konvexa mängder som definitionsområde är relativt lätta att lösa. En användbar metod för

praktiska problem är därför att formulera problemet som ett konvext optimeringsproblem. De exakta algoritmer och dataprogram som utför optimeringen beskrivs inte i det här arbetet, utan istället beskrivs ett användningsområde för konvex optimering.

4.1 Mönsterigenkänning

På samma sätt som ett barn kan lära sig att känna igen bokstäverna i alfabetet kan en dator lära sig avgöra om en bild föreställer ett ansikte eller inte. För att få en dator att lära sig själv känna igen samband och dra slutsatser av dessa kan man använda sig av mönsterigenkänning. Mönsterigenkänning innebär inlärning genom observationer. Inlärningsproblem med binära utfall, alltså ja eller nej, kallas *binära klassificeringsproblem*. Ett problem med ett ändligt antal olika utfall kallas *multippelklass klassificering*, däremot ett problem med reellvärda utfall kallas *regression*.

4.2 Klassificering

I *binära klassificeringsproblem* är målet att separera två mängder så tydligt som möjligt. Detta maximerar chansen att placera en ny okänd observation i rätt kategori. Genom att separera mängderna på ett sätt som gör att avståndet mellan mängderna blir så stort som möjligt så blir skillnaden mellan dem tydligast, vilket underlättar klassifikationen av nya observationer. Detta sker med en metod som kallas för största marginalmetoden. Det finns ett entydigt hyperplan som på det sättet separerar de konvexa mängderna, om syftet är att maximera avståndet mellan mängderna. Som tidigare visats har varje mängd en viss punkt som klart utgör minsta avståndet mellan sig själv och en annan mängd. Genom att lägga planet vinkelrätt mot detta kortaste avstånd, samt mittemellan mängderna, så separeras mängderna maximalt. Givet ett linjärt separabelt exempel

$$M = \{(\mathbf{x}_1, \beta_1), \dots, (\mathbf{x}_i, \beta_i)\},$$

där de två olika mängderna ges av identifikationsvariabeln $\beta = 1$ för ena mängden, respektive $\beta = -1$ den andra mängden. Hyperplanet som separerar de två mängderna kan skrivas $\mathbf{p}^T \mathbf{x} + \alpha = 0$. För varje punkt är avståndet till hyperplanet

$$\frac{|\mathbf{p}^T \mathbf{x} + \alpha|}{\sqrt{\mathbf{p}^T \mathbf{p}}}$$

vilket ska maximeras. Genom att sätta upp bivillkoren

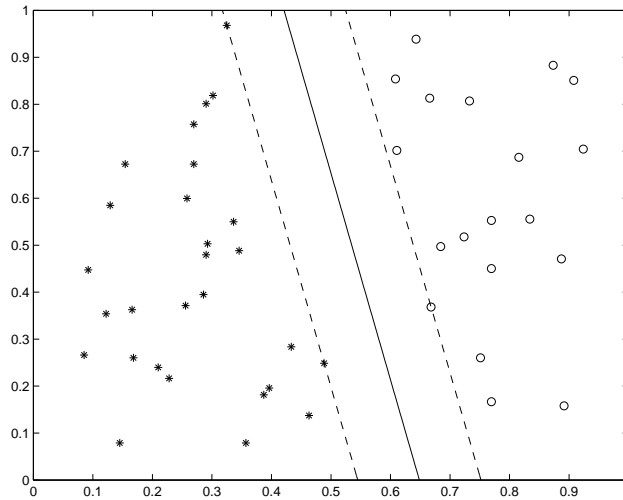
$$\beta_j(\mathbf{p}^T \mathbf{x} + \alpha) \geq 1, \quad j = 1, \dots, i,$$

är ekvivalent med att maximera

$$\frac{1}{\sqrt{\mathbf{p}^T \mathbf{p}}}$$

vilket är detsamma som att minimera $\mathbf{p}^T \mathbf{p}$. Funktionen $\mathbf{p}^T \mathbf{p}$ är en kvadratisk funktion och således konvex. Definitionsmängden beskrivs av hyperplanen

$$\beta_j(\mathbf{p}^T \mathbf{x} + \alpha) \geq 1, \quad j = 1, \dots, i.$$



Figur 6: Ett syntetiskt exempel med två mängder som representeras av cirklar och stjärnor.

Definitionsmängden är snittet av hyperplanen, och eftersom ett hyperplan är en konvex mängd är också snittet en konvex mängd, enligt sats 3. Följdaktligen är maximering av största marginalen ett konvext optimeringsproblem.

5 Syntetiskt exempel

Ett enkelt syntetiskt exempel är det tvådimensionella fallet, se figur 6

$$M = \{(x_1, y_1, \beta_1), \dots, (x_i, y_i, \beta_i)\},$$

där de två olika mängderna ges av identifikationsvariabeln $\beta = 1$ för mängden som representeras av cirklar respektive $\beta = -1$ den mängd som representeras av stjärnor.

Hyperplanet är en linje $ax + by + c = 0$ där a, b och c ges av att minimera $(a, b) \cdot (a, b) = a^2 + b^2$ med bivillkoren

$$\beta_j(ax + by + c) \geq 1, \quad j = 1, \dots, i,$$

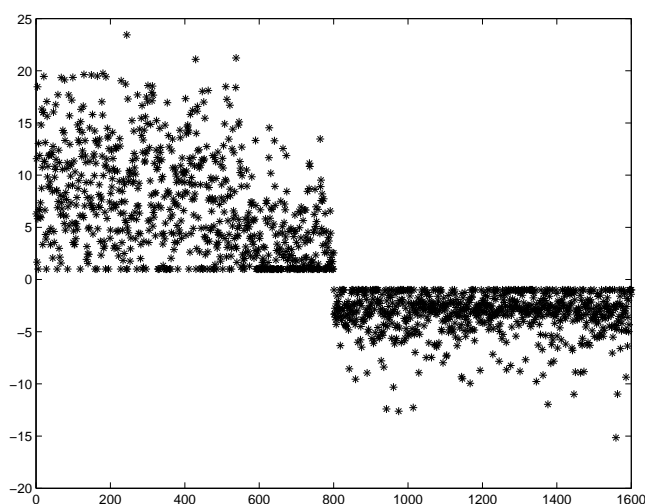
som kan tolkas som att alla stjärnor ska ligga till vänster om linjen och alla cirklar till höger. När marginalen optimeras kommer några av punkterna att hamna på lika långt avstånd från linjen. Dessa kallas för stödvektorer, då punkter även kan ses som vektorer, eftersom de stödjer "skivan".

6 Ansiktsdetektion

För att åskådliggöra användbarheten med konvex optimering demonstreras ett experiment med förenklad igenkänning. Experimentet går ut på att låta ett datorprogram avgöra om ett digitalt foto, med storleken 19x19 pixlar, föreställer



Figur 7: Exempel på inlästa bilder.



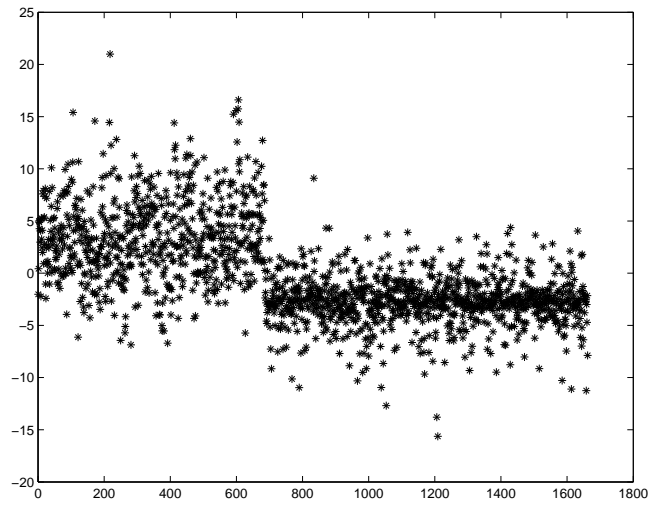
Figur 8: Skiss över träningsbildernas avstånd till det separerande hyperplanet.

ett ansikte eller ej. Precis som i det syntetiska exemplet låter man programmet skapa två separabla mängder och separera dessa med ett hyperplan. En av mängderna skapas utifrån foton som föreställer ansikten och den andra utifrån foton som inte föreställer ansikten. I detta fallet 800 av vardera sort. Att bilderna består av 19×19 pixlar medför att problemets dimensionstal blir precis 19^2 , det vill säga 361.

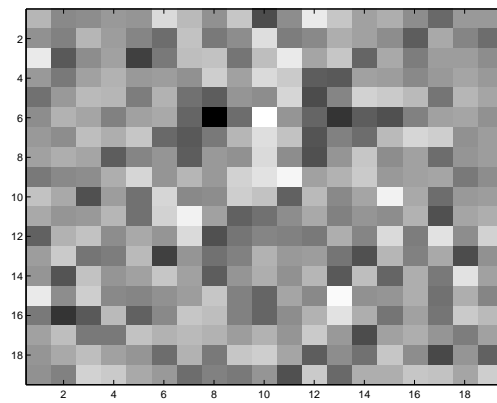
Låt varje bild representeras av en punkt i \mathbb{R}^{361} . Färgvärdet i en pixel betyder då punktens koordinat i den aktuella dimensionen. I figur 9 åskådliggörs punkternas avstånd till hyperplanet.

Programmet testas sedan genom att en ny bild läses och representeras av en punkt. Beroende på vilken sida om hyperplanen punkten ligger klassificeras den som ett ansikte eller ej. I figur 10 åskådliggörs hur nya okända foton klassificerats av programmet. De första 700 foton föreställer ansikten och resterande gör det inte. Man kan tydligt se att majoriteten av ansiktsbilderna ligger ovanför hyperplanet och majoriteten av de andra bilderna ligger under hyperplanet.

På samma sätt som en bild kan representeras av en punkt kan en punkt ge upphov till en bild. I figur 11 åskådliggörs den bild hyperplanetets normalvektor ger upphov till. Man kan med viss svårighet urskilja vissa drag i bilden som påminner om ett ansikte.



Figur 9: Skiss över testbildernas avstånd till det separerande hyperplanet.



Figur 10: Hyperplanets normalvektor återskapad som en bild. Läsaren uppmanas kisa för att urskilja ett ansikte.

7 Slutsatser

Rapporten innehåller en redogörelse för konvexa mängder, konvexa funktioner samt konvexa optimeringsproblem. Den visar dessutom hur konvex optimering kan användas i igenkänningsproblem. När ett problem kan omformuleras till ett konvext optimeringsproblem underlättas lösningen väsentligt. Detta beror på de egenskaper som visas hos konvexa mängder och funktioner, huvudsakligen i sats 4 och sats 5. I och med att varje lokalt minimum i ett konvext optimeringsproblem är ett globalt minimum sparas otroligt mycket tid i sökandet efter minsta funktionsvärdet.

Vi vill tacka vår handledare Karl Åström, som varit till stor hjälp under arbetets gång.

Referenser

- [1] Stephen P. Boyd. *Convex Optimization*. Cambridge University Press, 2004.
- [2] Lars-Christer Böiers. *Lectures on Optimization*. KFS AB, 2004.
- [3] N Cristianini and J.S. Taylor. *An introduction to support vector machines*. Cambridge, 2000.
- [4] Eike Petermann. *Konvexitet och Optimering*. Reprocentralen Stockholms Universitet, 1974.
- [5] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

Kvinnliga matematiker

Lisa Gustafsson, Sara Kiani, Marianne Lindkvist, Malin Sallhammar

5 maj 2006

Sammanfattning

Tre av världens mest kända kvinnliga matematiker; Hypatia, Sofya Kovalovsky och Emmy Noether har trots deras kön lyckats bli erkända inom den mansdominerande akademiska världen. Nykomlingen Salla Franzén, doktorand inom matematik vid Uppsala Universitet, jobbar under betydligt bättre förutsättningar än sina företrädare. I detta arbete har vi analyserat kvinnornas situation då och nu inom matematiken.

Innehåll

1	Inledning	3
2	Hypatia - den första kvinnliga matematikern	4
2.1	Verk	4
3	Sofya Kovalevsky - ryskan som härjade i Sverige	5
3.1	Utbildning	5
3.2	Livsverk	6
4	Emmy Noether - algebrans moder	7
4.1	Emmy Noethers liv	7
4.2	Livsverk	8
5	Salla Franzén - den moderna kvinnliga matematikern	9
6	Kvinnornas förutsättningar	10
7	Slutord	10
8	Källförteckning	11

1 Inledning

Hur kommer det sig att de flesta framstående vetenskapsmännen och matematikerna är män? Vilka är de mest kända kvinnliga matematikerna och vad har de åstadkommit? Detta är några frågor som vi kommer att ta upp i vårt arbete och förhoppningsvis svara på. Kvinnans roll i den akademiska världen har aldrig varit central, men inte obetydlig. Kan man helt säkert säga att Einstein var hela hjärnan bakom verken, eller bidrog hans fru till mer än vad man tror?

2 Hypatia - den första kvinnliga matematikern

Hypatia anses vara den första kvinnliga matematikern. Hon var inte bara matematiker, utan även filosof och astronom, som de flesta vetenskapsmännen på den tiden var. Hon växte upp under 300-talet i en akademisk miljö i staden Alexandria i Egypten, och var dotter till Theon, en av de mest bildade männen i staden som var professor i matematik på universitetet i Alexandria. Under denna tid var det relativt ovanligt att kvinnor överhuvudtaget fick någon intellektuell fostran, men Hypatia studerade tillsammans med sin far och med andra vetenskapsmän. Med sin far som lärare, följde Hypatia hans fotspår och blev lärare i matematik och filosofi vid samma institut som honom. Hon lockade till sig studenter från hela världen tack vare sin populära undervisning och även sin skönhet.

2.1 Verk

Hypatia blev känd mest för sina matematiska idéer om koniska sektioner, som tidigare hade introducerats av Apollonius, och skrev en kommentar till Apollonius verk *Om koniska snitt*, vilket handlar om koner som skärs av ett plan och ger upphov till s.k. koniska sektioner, beroende på vart planet skär. Hon skrev flera andra läroböcker, bl a en analys av sin faders utgåva av Euklides Elementa, som handlade om geometri. Dessutom uppfann hon flera instrument, däribland ett astrolabium för astronomiska observationer, en destillationsapparat och ett vattenpass, men varken dessa eller skrifter av dem finns bevarade.

Historiker tror att hon även skrev om bland annat planeternas rörelser och talteori. Hennes verk förvarades i biblioteket i Alexandria, men förstördes då biblioteket sattes i brand av araberna under deras erövningar. Det lilla man vet om Hypatia och hennes verk har man fått redan på från brev skrivna till henne från hennes samtida, andras verk och de som har citerat henne. Hypatia mördades tragiskt av en fanatisk kristen grupp 415 e.Kr. eftersom hon var gudsfönekare.

Efter Hypatia dröjde det mer än tusen år innan det på nytt trädde fram kvinnliga matematiker. Det var först på 1500- och 1600-talet som mycket begåvade kvinnor från samhällets högre klass ägnade sig åt vetenskaplig verksamhet, däribland matematik. Dessa kvinnor fick emellertid inte göra detta självständigt, utan i samband med översättning eller som medarbetare till någon lärd man.

3 Sofya Kovalevsky - ryskan som härjade i Sverige

Sofya föddes år 1850 och växte upp på ett lantställe i Vitryssland med mamma, pappa, en storsyster och en lillebror. Barnen gick inte i vanlig skola utan de undervisades av en privatlärare som bodde hos dem.

När Sofya var tretton år lyckades hon komma över en bok om algebra som hon studerade i hemlighet. Ett år senare kom en professor, en granne till dem, som skrivit en bok i fysik. Sofya läste boken men läsningen hindrades av att hon inte kunde trigonometri. Hon frågade sin brors privatlärare som dock inte ansåg sig kunna tillräckligt. Därför började Sofya själv att analysera formlerna. Vid nästa besök av professorn ställde hon frågor till honom men han kunde inte tro att hon hade förstått någonting. Sofya förklarade hur hon hade tänkt och professorn blev mycket imponerad och berättade om hennes begåvning för pappan. Professorn menade att Sofya skulle få ordentlig undervisning i matematik och detta fick hon några år senare när familjen tillbringade en tid i Petersburg.

3.1 Utbildning

På den tiden (mitten av 1800-talet) hade inte flickor tillträde till universitet i Ryssland. Deras enda möjlighet var att komma till ett universitet som tillät kvinnor att studera i Schweiz eller Tyskland. En ung dam kunde inte åka ensam till främmande land utan att ha manligt stöd. Alltså var Sofya tvungen att finna en äkta man. Detta fann hon i Vladimir Onufrievich Kovalévsky som hon ingick ett skenäktenskap med. Det krävdes dock en del övertalning för att hennes far skulle godkänna giftermålet. Paret bodde i St Petersburg och där började Sofya studera fysiologi och anatomi.

Dock långtade båda ut i världen och 1869 åkte de tillsammans med Sofyas syster Aniuta till Tyskland. Men när Sofya kom till Heidelberg insåg hon att kvinnor inte hade tillstånd att studera vid universitetet där. Efter mycket krångel lyckades dock Sofya att få lov att närvara vid föreläsningar, inofficiellt. Sofya gjorde stort intryck på lärarna med sin matematiska begåvning.

1870 hade Sofya kommit underfund med att det var matematik som intresserade henne allra mest. Därför bestämde hon sig för att bli elev till professor Karl Weierstrass i Berlin. Sofya åkte dit och hade med sig rekommendationsbrev av professorer i Heidelberg. Weierstrass kunde inte ta emot henne som student vid universitetet men han blev så imponerad av hennes kunskaper att han gick med på att undervisa henne privat. Under några år arbetade Sofya intensivt med sina studier och hon unnade sig inget uteliv. Arbetet resulterade i tre originalverk som vart och ett i sig skulle kunna ha utgjort en doktorsavhandling. Men enligt Weierstrass kunde de inte läggas fram i Berlin. Han förordade universitetet i Göttingen eftersom det tidigare hade godkänt doktorsavhandlingar för utlänningar utan att ha studerat där. Så i juli 1874 fick Sofya en doktorsgrad i matematik. Hennes man Vladimir hade fått sin doktorsgrad i paleontologi två år tidigare i Jena. Deras äktenskap var inte särskilt bra. Det hade blivit sämre på grund av att de hade lite pengar, inte hade bött tillsammans på många år och framför allt att ingen av dem ville vara sämre än den andre. För att förbättra sitt förhållande åkte de tillbaka till Ryssland 1874.

Första vintern i St Petersburg kom det fram nya sidor av Sofyas personlighet.

Efter att tidigare ha isolerat sig från omvärlden började hon nu intressera sig för vänner och underhållning. Ett problem var dock att hitta ett arbete. Det bästa hon kunde hitta var att undervisa matematik i en flickskola, något hon inte fann särskilt stimulerande då det var långt under hennes nivå. Inte heller hennes man fick det jobb han hoppats på trots sin höga examen. Istället försökte Vladimir tjäna pengar på att köpa och sälja hus men misslyckades fullständigt och det slutade i stora skulder. Samtidigt gick Sofya vidare och under en period på två år arbetade hon med något hon tyckte om. Hon var teaterrecensent och vetenskapsskribent och skrev bland annat om telefonens, skrivmaskinens samt flygplanets utveckling.

3.2 Livsverk

I oktober 1878 födde Sofya en dotter som fick smeknamnet Fufa. Under de första åren lade mamman stor energi på att ta hand om sin dotter men i det långa loppet var inte ens ett barn tillräckligt stimulerande för henne. Då Vladimirs affärer bara gick sämre ökade klyftan mellan dem ytterligare. Sofya hittar åter tillbaka till matematiken och tar fram ett sex år gammalt arbete om abelianska integraler som imponerar stort på de ryska matematikerna. I ett enda slag var hon återigen en matematiker att räkna med. En man vid namnet Gösta Mittag-Leffler försökte ordna ett jobb vid universitetet i Helsingfors åt Sofya men det misslyckades då fakulteten vägrade att ta emot en rysk kvinna. Sofya lämnar återigen Ryssland och börjar forska om ljusets brytning i kristaller på Weierstrass inrådan. Hemma i Ryssland har Vladimir det svårt och begår självmord på våren 1883. Sofya tar på sig skulden och försöker svälta sig själv till döds men misslyckas. Hon återhämtar sig snart och fortsätter med sina studier. Mittag-Leffler lyckas ordna en tjänst till henne vid Stockholms universitet med det villkoret att hennes första år skulle vara ett prövoår utan lön. I juli 1884, sex månader efter hennes första föreläsning, utnämndes hon till professor i matematik. I Stockholm föreläste hon tre gånger i veckan och gav även privatundervisning. Senare blev hon redaktör för *Acta Mathematica* i vilken hon publicerade sina översättningar av två avhandlingar av den ryske matematikern P L Chebyshev. År 1885 blev hon professor i mekanik och matematik vid universitetet i Stockholm och år 1889 fick hon professur på livstid. Det största Sofya gjorde var nog ändå att vinna Prix Bordin (den franska vetenskapsakademins högsta pris) genom hennes avhandling i en stel kropps rotation runt en fast punkt. Det var fjärde gången detta ämne blev föremål för tävlan efter att Euler, Lagrange och Poisson misslyckats med att lösa problemet. Efter att anonymt lämnat in bidraget blev det slutligen utvalt ibland 15 andra tävlande. Det var så elegant gjort att prissumman höjdes från 3000 francs till 5000 francs.

Efter ett besök i Paris fick hon en hjärtattack. Hon återvände till Stockholm i februari 1891, inom några dagar var hon död. Hon blev 41 år gammal.

4 Emmy Noether - algebrans moder

Emmy Noether anses vara en av de största kvinnliga matematikerna genom tiderna. Trots alla motgångar och motsättningar på grund av sitt kön, lyckades hon ändå bli erkänd inom matematiken och har bidragit med mycket inom algebran. Hon influerade även andra matematiker på den tiden, såväl kvinnliga som manliga. Eftersom hon var grundaren till den abstrakta algebran så går hon även under beteckningen "algebrans moder".

4.1 Emmy Noethers liv

Amalie Noether, mer känd som Emmy Noether, föddes den 23 mars 1882 i Tyskland. Hennes pappa Max var en framgångsrik matematiker, vilket även hennes broder så småningom blev. Under skolåren studerade hon språk eftersom matematiken inte direkt var riktad till kvinnor, men vid 18 års ålder valde Noether att börja läsa matematik på universitetet där hennes pappa var professor. Noether var en av de två kvinnliga studenter på universitetet, som för övrigt bestod av 984 manliga studenter.

År 1907 då Emmy var 25 år gammal klarade hon av sin muntliga doktorsexamen och ett år senare blev hennes doktorsavhandling godkänd och registrerad. Hon var då den andra kvinnan efter Kovalevsky som någonsin hade doktorerat i matematik. Dock hade denna avhandling ingenting att göra med det som senare skulle göra henne världsberömd.

Följande 10 åren "arbetade" Noether vid den matematiska institutionen i hemstaden. Universitetet var motvilliga till att anställa henne så därför jobbade hon för sin pappa på universitetet, främst med forskning men hon hade även en del föreläsningar. Allt hon gjorde var dock i hennes pappas namn. Även efter första världskrigets slut då Tyskland blev republik och kvinnor fick rösträtt så fick hon fortfarande inte någon lön för sitt arbete. Så småningom började hon publicera artiklar och skriva avhandlingar om sin forskning i sitt eget namn och detta ledde till att två manliga matematiker som jobbade vidare på Einsteins teorier tog kontakt med henne. De bjöd in Noether till att arbeta med dem och detta gjorde Noether mer känd inom matematiska kretsar och hon fick så småningom ett obetalt jobb som lektor. Emmy Noether gladdes av detta eftersom det var första gången hon arbetade i sitt eget namn, och efter tre år som lektor fick hon även en mindre lön.

Åren 1920-1927 studerade Emmy Noether kommutativa algebror och några år efter det studerade hon icke kommutativa algebror. Hon skrev avhandlingar om detta som kom att bli riktigt stora. Innan detta sysslade hon med invariantteori. Det var dessa tre områden som gjorde henne så framgångsrik och hennes upptäckter spelade stor roll inom den abstrakta algebran och därav kom hon att kallas algebrans moder.

1933 flyttade Noether till USA där hon arbetade på ett amerikanskt universitet för kvinnor. Hon avled 2 år senare, endast 53 år gammal i samband med en operation. Det var först efter hennes död som hon verkligen blev en erkänd matematiker. Många av hennes idéer utvecklades senare av hennes elever eller medarbetare.

Att vara kvinnlig matematiker i Tyskland under Emmy Noethers tid var alltså inte speciellt lätt. Under hennes universitetstid fick kvinnliga studenter inte registrera sig på vanligt sätt utan fick komma som åhörare, då med ett speciellt tillstånd av undervisaren. De hade rätt att neka kvinnor tillträde till föreläsningarna. Flera av de högst uppsatta männen inom universitetet tyckte att "Närvaron av kvinnor på universiteten förstör det grundläggande syftet med akademiskt liv", som en talesman för traditionalisterna i Tyskland, som Heinrich von Treitschke uttryckte det. Att Noether inte heller fick betalt för sitt arete även då hon flera gånger visat sin matematiska begåvning, är ännu ett bevis på påfrestningen för kvinnorna i matematikens värld.

4.2 Livsverk

Emmy ägnade sig först åt invariantteorin. Detta hade en stor betydelse för den kommutativa algebrans utveckling. Detta ledde henne också in på en mer fysikalisk bana. Hon lyckades formulera en sats, så kallad Noethers sats, som fortfarande är mycket viktig inom matematisk fysik. Detta var dock bara ett sidospår från hennes matematiska karriär och inget som hon ytterligare fördjupade sig inom.

1921 publicerar Emmy Noether sin artikel om kedjevillkor för ringar, "Idealtheorie in Ringbereichen". Hon klassificerade en typ av ringar som kom att heta noethersta ringar. Denna artikel blev väldigt inflytelserik och det som gjorde att folk inom den matematiska världen hörde talas om hennes namn. Med denna artikel öppnade hon nya matematiska vägar och ett helt nytt algebraiskt tankesätt. Noether generaliserade det som andra matematiker som Lasker och Macualay redan visat. Hon lyckades bevisa sina teorier på ett sätt som var enkelt och smidigt att förstå jämfört med tidigare liknande bevis. En annan av Noethers stora avhandlingar som bör nämnas är "Abstrakter Aufbau der Idealtheorie om algebraischen Zahl- und Funktionenkörpern", som hon skrev år 1927. Denna avhandling har sina rötter i framförallt algebraisk talteori. Dessa två avhandlingar lade tillsammans grunden för kommutativ algebra eftersom hon introducerade nya begrepp samt utvecklade en bra teknik och bevisade viktiga resultat.

Emmy Noether sysslade även med icke kommutativ algebra och representations-teori. Mest uppmärksammas blev hennes avhandling "hyperkomplexe Größen und Darstellungstheorie", som kom år 1929. Denna avhandling handlade om grupprepresentations-teori i termer av strukturteori för hyperkomplexa system.

Några exempel på begrepp namngivna efter Emmy Noether är noetherska ringar, noetherska moduler, noetherska topologiska rum och noetherska faktorsystem.

5 Salla Franzén - den moderna kvinnliga matematikern

När vi letade efter en modern kvinnlig matematiker hittade vi av en slump nätverket Lilla Mys" hemsida. Därigenom fick vi kontakt med Salla Franzén.

Salla doktorerar inom komplex analys i Uppsala och på frågan varför hon blev matematiker svarar hon Min dröm har alltid varit att ha ett yrke som är krävande och som jag kan vara stolt över. När jag förstod att man kan bli matematiker tänkte jag genast att det var något jag ville prova på".

Hon började med att läsa fristående kurser inom matematiken i Uppsala, bl.a. en kurs i komplex analys. Det var så hon först kom i kontakt med den kvinnliga professor som kom att bli hennes handledare för ex-jobbet. Denna professor var för några år sedan den enda kvinnliga professorn inom matematik i Sverige!

Varför det är så få kvinnor inom matematiken har Salla inte något definitivt svar på, men hon säger att bilden av att matematiker är män antagligen blir starkare därför att många matematiklärare i grundskolan är just män. Hon är själv ganska trött på att det ses som ett problem, att det inte finns fler kvinnliga matematiker i Sverige, utan att någon gör något konkret åt det.

En av de saker som hon märkt av inom matematiken är att man som tjej inte är en riktig matematiker förrän man har godkänts av en duktig man. Det har hon själv märkt och hon har också hört flera andra tycka detsamma. Dock vill hon påpeka att det inte är så i alla sällskap, hon själv undviker att umgås med människor som har en dålig inställning.

Salla anser att man som kvinnlig doktorand måste hitta en hållbar inställning till att mannen är normen. Hon säger att man måste kunna ta sig igenom de situationer där man känner sig som ett undantag. Trots alla tankar och reflektioner hon har kring jämställdheten säger hon att hon inte försöker företräda hela det kvinnliga släktet, det skulle bli för svårt. Hon tror att om man vill bli matematiker för att bevisa att kvinnor kan så kan det bli väldigt svårt, men om man forskar för att det är kul går det bättre!

Just nu anser Salla att det är en väldigt trevlig situation inom hennes forskningsområde i Sverige. Det är totalt 6 tjejer som doktorerar inom komplex analys i flera variabler.

6 Kvinnornas förutsättningar

Kvinnorna hade det i allmänhet svårt att slå igenom i den vetenskapliga världen på grund av fördomar, hinder orsakat av männens dominans och eftersom det ansågs vara opassande. Ett exempel är fransyskan Sophie Germain som under slutet av 1700-talet fick kämpa för att nå en position som motsvarade hennes kompetens. Trots konkurrensen och svårigheterna med att studera på universitet och överhuvudtaget ta en akademisk examen i matematik, lyckades italienskan Maria Gaetana Agnesi, som levde under 1700-talet, bli den första kvinnliga matematikern som blev professor vid ett universitet.

Före andra världskriget avrättades eller hindrades kvinnor i många utbildningsmässigt mer utvecklade länder från att studera matematik vid universitet. Deras situation kan jämföras med hur de hade det exempelvis under Pythagoras tid (ca 570-500 f.Kr.) då kvinnor till och med kunde vara lärare i matematik. Detta kom att upphöra. Med väldigt få undantag har kvinnor varit utestängda från dessa studier, tills nyligen.

7 Slutord

Vad dessa kvinnor har gemensamt, förutom passionen till matematiken, är svårigheterna med att utöva den i praktiken. Eftersom många inte ens hade tillträde till universiteten eller arbeten inom ämnet, nådde deras idéer inte fram. De som väl studerade gjorde det under manliga pseudonymer men fick därför ofta ingen "credit" för sitt arbete. Dagens kvinnliga matematiker är däremot mer attraktiva än förut, just för att de är kvinnor och för att gynna jämställdheten inom arbetsmarknaden. Vi i gruppen har inte upplevt några hinder på grund av att vi är kvinnor, men fortfarande är det inte lika jämställt på dagens mansdominerade företag som det är på högskolorna.

8 Källförteckning

<http://www.agnesscott.edu/lriddle/women/hypatia.htm> 060404

<http://womenshistory.about.com/od/hypati1/a/hypatia.htm> 060404

Matematikens kulturhistoria, McLeish, John, 1991, Forum AB, ISBN 91-37-10979-0

Stora matematiker: från Fibonacci till Wiles, Picutti, Edwards, Schwartz m.fl., 1996, ISBN 91-44-01034-6

Den matematiska människan, Butterworth, Brian, 1999, ISBN 91-46-17406-0

A Russian Childhood, Kovalevskay, 1978, ISBN 0-387-90348-8

Arvet efter Hypatia, Margaret Alic, 2002, ISBN 91-501-0204 4

<http://www.agnesscott.edu/lriddle/women/noether.htm> 060404

<http://epubl.ltu.se/1402-1781/2001/06/index.html> 060404

Mailkontakt med Salla Franzén.

Kan man höra formen på en trumma?

Patrick Fogwall, Caroline Malmberg, Julia Nilsson, Albin Olsson
 $\pi 05$

17 maj 2006

Vad handlar problemet om?

Matematiker har alltid varit intresserade av vad som händer i kända situationer när man tar problemet till högre dimensioner. Detta är en naturlig frågeställning som leder matematiken och vetenskapen framåt. Man kan lätt tänka sig att givet egenfrekvenser för en gitarrsträng kan man ta reda på dess längd, men hur är det för ett trumskinn. Kan man ta reda på formen av en trumma genom att veta vilka egenfrekvenser den ger upphov till?

Första gången denna fråga officiellt ställdes var 1966 av den polske matematikern Marc Kac (1914-1984) i artikeln "Can one hear the shape of a drum?" i tidskriften *The American Mathematical*

Monthly. Denna artikel skapade stor uppmärksamhet och 1968 tilldelades Kac *The Chauvenet Prize* från *The Mathematical Association of America* med motiveringen

"for the most outstanding expository article on a mathematical topic by a member of the Association."

Varför var då denna artikel så intressant? Fysiker och matematiker visste sedan länge att formen av den rand som innesluter ett membran spelar en avgörande roll för membranets frekvensspektrum. Med sin artikel satte Kac fokus på motsatt problem. Om man vet ett membrans frekvensspektrum, kan man då avgöra dess geometriska form? Problemet går alltså ut på huruvida man från de utsända frekvenserna kan avgöra formen på trumman som skickat ut dem. En del förenklingar och antaganden gjordes, bl.a. att trumman egentligen var en tamburin dvs. ett membran fastsatt i sin ram, att lyssnaren hade absolut gehör, på så vis att denne kunde avgöra exakt vilka frekvenser den uppfattade tonen var uppbyggd av, samt att isometriska membran räknas som samma, dvs. att det inte görs skillnad på om trumman roteras eller om denna vänds upp och ner, ej heller spelar trummans placering in.

Sedan 1800-talet är det känt att ett membran som är fastsatt längst sin rand uppfyller vågekvationen då det sätts i rörelse (se nästa sida). Membranet svänger då på ett sätt som är resultanten av ett antal enkla svängningar. Dessa svängningar är rena egenfrekvenser vilka kan vara grundfrekvensen eller någon av de oändligt många övertonerna. Dessa frekvenser kan sedan användas för att få fram egenvärdena genom att frekvensen är proportionell mot kvadraten ur egenvärdet. Med egenvärdenas hjälp är det möjligt att komma fram till en del detaljer om membranets form, t.ex. dess area och perimeter dvs utformning av randen. I sin artikel utgick således Kac från dessa kända fakta om att ett membrans area och omkrets lämnar tydliga spår i dess frekvensspektrum. Kac inspirerades även av ett problem han hört talas om tio år tidigare. Förenklat lät problemet ungefär såhär:

Låt Ω_1 och Ω_2 vara plana ytor respektive begränsade av kurvorna Γ_1 och Γ_2 , där egenvärdena definieras av Laplace-operatorn, anta att varje egenvärde i re-

spektive följd överensstämmer med de andra. Är då ytorna kongruenta?

I sin artikel resonerade Kac vidare kring ett membrans svängningsrörelse men lämnade frågan om huruvida man kunde höra dess form öppen. Han anade dock att det inte gick att lyssna sig till en trummas form. Följande citat från artikeln visar på hans osäkerhet i frågan;

“I believe that one cannot “hear” the shape of a tambourine, but I may well be wrong and I am not prepared to bet large sums either way.”

Med sin artikel öppnade Kac vägen för studier inom spektalteori och som vid många tillfällen då en intressant fråga lämnas obesvarad tog andra matematiker vid för att söka svar. Det dröjde dock till 1992 tills man kunde besvara given fråga med ett nej. Carolyn Gordon, David Webb och Scott Wolpert hade då lyckats visa att två membran kan ha samma egenvärden men olika form. Detta publicerades i *Bulletin Of The American Mathematical Society* (vol 27, no 1, 1992). Gordon, Webb och Wolperts lyckades även visa att givet att membranet är cirkulär eller triangelformad kan man avgöra formen dvs. bestämma radie respektive kantlängder. Deras resultat väckte även frågan om det räcker att veta att trummembranet är konvext för att alla givna egenvärden entydigt skall bestämma dess form. Vem/vilka skickliga matematiker som ska ge svaret på denna fråga kan enbart framtiden utvisa.

Trumman uppfyller vågekvationen

Ekvationen som styr hur trumskinnets rör sig är vågekvationen

$$\frac{\partial^2 F}{\partial t^2} = c^2 \nabla^2 F$$

där c^2 är en konstant och Laplace-operatorn $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

Här utgör $F(x, y; t)$ trumskinnets avvikelse från jämviktsläget 0 i punkten $\vec{p} = (x, y)$ vid tiden t , då membranet sätts i svängning. Avvikelsen är vinkelrät mot membranets plan. Denna ekvation beskriver egentligen inte en verklig trumma eftersom den inte tar hänsyn till friktion och trumman svänger lika mycket i all oändlighet. Vågekvationen har massor av lösningar men för att få den intressanta får man precisera förutsättningarna.

Låt trumskinnets vara ett område Ω i planet med randen Γ , där randen är trummans kant. Eftersom trumskinnets sitter fast i kanten har vi randvillkoret

$$F(\vec{p}, t) = 0, \quad \text{då } \vec{p} \in \Gamma.$$

Det finns fortfarande många lösningar till vågekvationen med detta randvillkor. För att få ner antalet lösningar måste vi också ange hur lösningen ser ut när vi startar, dvs vid $t = 0$. Detta kallas begynnelsevärden och brukar se ut som

$$F(\vec{p}; 0) = f(\vec{p})$$

$$\frac{\partial F}{\partial t}(\vec{p}; 0) = g(\vec{p})$$

där f och g är kända funktioner. Vanligt är att låta $f(\vec{p}) = 0$ vilket svarar mot att trumman ligger i jämviktsläget vid tiden $t = 0$ och där $g(\vec{p})$ beskriver hur trumpinnen sätter fart på trumskinnets vilket i praktiken innebär att även $g(\vec{p}) = 0$ för de flesta \vec{p} utom där trumpinnen träffar trumman. Man kan visa att vågekvationen tillsammans med randvillkoret och begynnelsevärdena har en lösning F som beskriver hur trumskinnets rör sig i varje punkt och vid varje tidpunkt. Svårigheten är att beräkna lösningen som beror av tre variabler x , y och t . Det fina med beteckningarna är att vågekvationen ser precis likadan ut i 1, 2 eller 3 rumsdimensioner. Skillnaderna är inbyggda i beteckningarna \vec{p} och ∇^2 . När vi nu vill lösa dessa partiella differentialekvationer börjar vi först med vågekvationen och randvärdet för att senare ta hand om begynnelsevärdena i efterhand. Det är lätt att se att vågekvationen har lösningen $F = 0$ när trumskinnets är i vila. Denna lösning kallas den triviala lösningen och den vill vi undvika i fortsättningen. Vi börjar med att söka efter lösningar på formen

$$F(\vec{p}; t) = U(\vec{p})e^{i\omega t}.$$

Detta svarar mot att om man känner trumskinnets läge $U(\vec{p})$ vid en tidpunkt så känner man det för alla tidpunkter eftersom det kommer att svänga mellan utslaget $U(\vec{p})$ och $-U(\vec{p})$ med vinkelfrekvensen ω .

Dessa speciella lösningar till avvikelserna F är harmoniska över tiden och utgör egensvängningarna till membranet kring jämviktsläget. Varje rörelse hos trumskinnets kan sedan uttryckas som en superposition av egensvängningar. I en egensvängning har trumskinnets frekvensen ω . Trumskinnets sätter luftmolekylerna i svängning och en ljudvåg fortplantar sig till våra öron (denna våg beskrivs av en ny vågekvation, denna gång i tre rumsdimensioner, men den struntar vi i) och som örat uppfattar som en ton med frekvensen $f = \omega/(2\pi)$. Tidsfunktionen $e^{i\omega t}$ är uttryckt på komplex form men kan även skrivas som $A \sin(\omega t) + B \cos(\omega t)$. Varför vi har valt den komplexa formen är för att det ger snyggast räkningar. För att hitta egensvängningarna sätter man in lösningen (se ovan) i vågekvationen

$$\frac{\partial^2}{\partial t^2}(U(\vec{p})e^{i\omega t}) = c^2 \nabla^2 U(\vec{p})e^{i\omega t},$$

$$(i\omega)^2 U(\vec{p})e^{i\omega t} = c^2 \nabla^2 U(\vec{p})e^{i\omega t},$$

$$e^{i\omega t}(\omega^2 U(\vec{p}) + c^2 \nabla^2 U(\vec{p})) = 0.$$

Eftersom exponentialfunktionen är skild från noll, måste den andra faktorn vara lika med noll. Vi får då

$$\nabla^2 U = -\frac{\omega^2}{c^2} U,$$

$$U(\vec{p}) = 0 \quad \text{då } \vec{p} \in \Gamma,$$

$$U \neq 0.$$

Här är U en funktion och ω en konstant som ska bestämmas. Den tredje ekvationen säger att det är de icke-triviala lösningarna vi vill ha. Denna ekvation är enklare än den ursprungliga vågekvationen eftersom den innehåller en variabel mindre. Med $\lambda = -\omega/c^2$ så ser den väldigt lik ut en egenvärdesekvation $AF = \lambda F$, $F \neq 0$ för matriser A . Randvärdena gör att differentialekvationen bara har lösningar för vissa ω . Denna ovanstående differentialekvation är förutom i några helt symmetriska undantagsfall omöjlig att analytiskt lösa i 2 dimensioner. Däremot kan vi numeriskt beräkna fram lösningarna. För att få en bättre uppfattning om problemsituationen kan vi istället för ett trumskinn modellera en gitarrsträng vilket ger oss ett problem i en dimension.

Endimensionell lösning

Vi studerar en gitarrsträng som sitter fast i punkterna $x = 0$ och $x = L$. Då får vi istället

$$U'' = -\frac{\omega^2}{c^2}U,$$

$$U(0) = 0, \quad U(L) = 0.$$

Nu kan vi lösa denna differentialekvation och får då

$$U(x) = \begin{cases} A \sin \omega x/c + B \cos \omega x/c, & -\lambda = (\omega/c)^2 > 0, \\ Ax + B, & \omega = 0, \\ Ae^{\sqrt{\lambda}x} + Be^{-\sqrt{\lambda}x}, & \lambda > 0. \end{cases}$$

Om man försöker bestämma konstanterna A och B genom att sätta in randvärdena $U(0) = 0$ och $U(L) = 0$ fås nästan enbart triviala lösningar, dvs $A = B = 0$. I det första uttrycket kan vi emellertid undvika att $A = 0$ genom att ställa upp villkoret

$$\frac{\omega L}{c} = k\pi, \quad k = 1, 2, 3, \dots$$

Den icke-triviala lösningen blir då $U(x) = A \sin \omega x/c$ där $\omega = k\pi c/L$ för positiva heltal k .

Tillbaka till trumman

Nu ska vi tänka oss ta fram motsvarande lösningar för ett trumskinn, vilket är mycket avancerat och ligger på en nivå som är högre än vad vi själva förstår. Vi har satt oss in delar av teorin som berör trumskinnets frekvensspektrum. Då vi själva upplever teorin som svår kommer vi inte att fokusera på detaljer för att förklara teorin. Vi ska betrakta det inversa problemet till vad vi precis har löst för en gitarrsträng, fast nu i två dimensioner. Vi ska nu förutsätta att vi

känner trummans möjliga egenfrekvenser/ljud som kan uppstå när man slår på den med en trumpinne. Vi har alltså ett trumskinn Ω , precis som tidigare och en serie kända egenvärden $\lambda_1 \leq \lambda_2 \leq \lambda_3 \dots$. Varje sådant egenvärde har en motsvarande egenfunktion som också beror av en operator (här Laplace-operatorn). Egenfunktionerna är det lägesberoende vi har för att beskriva utslaget som trumskinnets vibrerar med. En egenfunktion innehåller information om utslaget och dess riktning för varje punkt på trumskinnets i förhållande till jämviktsplanet. Egenfunktionerna är lösningar till de egensvängningar som trumskinnets kan utföra, precis som lösningen för strängen i 1-dimensionsfallet.

Våra egenfunktioner $\Psi_n(\vec{p})$ för alla n ska uppfylla

$$c^2 \nabla^2 \Psi_n + \lambda_n \Psi_n = 0$$

(följer från vågekvationen) där utslaget $\Psi_n(\vec{p}) \rightarrow 0$ då vi närmar oss en punkt på randen Γ , eftersom trumskinnets sitter fast i en kant.

Vi väljer $c^2 = \frac{1}{2}$ som den konstant vilken anger de fysikaliska egenskaperna, för att förenkla räkningarna.

Det är vanligt att man normerar egenfunktionerna, Ψ_n , så att

$$\iint_{\Omega} \Psi_n^2(\vec{p}) d\vec{p} = 1,$$

Observera här att $\vec{p} = (x, y)$. Denna normering kommer också att förenkla räkningarna lite senare.

Härledning av arean

När Kac skrev sin artikel visste han att man kunde beräkna arean och omkretsen av trumskinnets när man kände dess egenfrekvenser. Dessa kända fakta ledde till en allmän hypotes om att det vore möjligt att även få fram trummembranets form.

För att ta fram ett uttryck för arean kan man angripa problemet på flera sätt. Vi kan exempelvis använda schrödingerekvationen för en ideal gas eller teori för diffusion. Utan att gå miste om förståelse för innehållet gör vi denna härledning utifrån användbara uttryck från diffusion utan att ge en fullständig förklaring till den bakomliggande teorin. Diffusion är ett begrepp som innebär spontan utjämnning av koncentration i en vätska. Detta gör att partiklar kan spridas genom tunna membran, exempelvis joner som diffunderar genom cellmembran. När vi nu studerar diffusion så lämnar vi trumskinnets och tänker det istället endast som ett område (membran) som partiklar kan diffundera genom. Detta område Ω avgränsas av randen Γ , liksom tidigare när vi studerade trumskinnets. Materia sprids från punkten \vec{p} till \vec{r} . Här är \vec{p} nu en fix punkt $\vec{p}_0 = (x_0, y_0)$. Koncentrationen av materia vid $\vec{r} = (x, y)$ vid tiden t betecknar vi $P_{\Omega}(\vec{p}|\vec{r}; t)$. Denna koncentration uppfyller differentialekvationen för diffusion,

$$\frac{\partial P_{\Omega}}{\partial t} = \frac{1}{2} \nabla^2 P_{\Omega}.$$

Idéen i härledningen är att vi låter koncentrationen approximativt närma sig sitt startvärde i \vec{p} . Då tillåts vi ta fram ett uttryck från diffusionsekvationen och ett annat från egenvärdesproblemet. Dessa två uttryck är approximativt lika för ett mycket litet tidssteg. Sedan kan vi förenkla detta sammansatta uttryck till ett uttryck där arean ingår.

Vi låter approximativt

$$P_{\Omega}(\vec{p}|\vec{r}; t) \sim P_0(\vec{p}|\vec{r}; t),$$

då $t \rightarrow 0$. Här betyder \sim , "asymptotiskt lika".

Vi har också P_0 som fortfarande uppfyller diffusionsekvationen $\frac{\partial P_0}{\partial t} = \frac{1}{2}\nabla^2 P_0$. Enligt diffusionsekvationen kan koncentrationen uttryckas i form av egenfunktioner till vågekvationen som vi såg tidigare. Man kan visa att koncentrationen givet av denna representation blir

$$P_{\Omega}(\vec{p}|\vec{r}; t) = \sum_{n=1}^{\infty} e^{-\lambda_n t} \Psi_n(\vec{p}) \Psi_n(\vec{r}).$$

Nu, för små t tänker vi oss att partiklarna inte känner av en gräns till Ω .

Där P_0 uppfyller diffusionsekvationen får vi enligt teorin för diffusion att koncentrationen är känd som

$$P_0(\vec{p}|\vec{r}; t) = \frac{1}{2\pi t} \exp\left[-\frac{|\vec{r} - \vec{p}|^2}{2t}\right].$$

Då kan vi skriva det approximativa påståendet som

$$P_{\Omega}(\vec{p}|\vec{r}; t) = \sum_{n=1}^{\infty} e^{-\lambda_n t} \Psi_n(\vec{p}) \Psi_n(\vec{r}) \sim \frac{1}{2\pi t} \exp\left[-\frac{|\vec{r} - \vec{p}|^2}{2t}\right] = P_0(\vec{p}|\vec{r}; t).$$

Om vi nu låter punkten \vec{r} som partiklar sprids till vara samma som startpunkten \vec{p} , ($\vec{r} = \vec{p}$). Då kan vi förenkla uttrycket till

$$\sum_{n=1}^{\infty} e^{-\lambda_n t} \Psi_n^2(\vec{r}) \sim \frac{1}{2\pi t}.$$

Integrerar vi detta resultat och använder normeringen $\int_{\Omega} \Psi_n^2(\vec{r}) d\vec{r} = 1$ (observera här $\vec{r} = (x, y)$), får vi

$$\begin{aligned} \int_{\Omega} \sum_{n=1}^{\infty} e^{-\lambda_n t} \Psi_n^2(\vec{r}) d\vec{r} &\sim \frac{1}{2\pi t} \int_{\Omega} d\vec{r} \\ \sum_{n=1}^{\infty} e^{-\lambda_n t} \underbrace{\int_{\Omega} \Psi_n^2(\vec{r}) d\vec{r}}_{=1} &\sim \frac{|\Omega|}{2\pi t}, \end{aligned}$$

och erhåller arean

$$|\Omega| = \lim_{t \rightarrow 0} 2\pi t \sum_{n=1}^{\infty} e^{-\lambda_n t}.$$

Denna formel innebär att om vi känner vårt frekvensspektrum, vår serie av egenvärden, så kan vi i teorin beräkna arean av trumskinet. I praktiken kan vi inte göra exakta beräkningar då vi inte har en oändlig serie av egenvärden att tillgå. Det går att göra simuleringar av detta slag med dator där man får fram ett närmevärde till arean, beräknat utifrån en känd, begränsad uppsättning egenvärden till trumskinet. Men det finns också enkla fall av symmetri då vi med säkerhet kan bestämma vilken form trumman har, vi ska titta på den cirkulära trumman.

Cirkulär trumma

Det finns ett fall då vi vet den exakta formen av trumman och problemet är lösbart, detta är fallet då randen är en cirkel. Cirkeln är den geometriska figur som har minst omkrets i förhållande till sin area. Detta kommer från lösningen av $L^2 \geq 4\pi|\Omega|$ för någon given omkrets L , där $|\Omega|$ är arean av membranet. Likhet i detta påstående inträffar bara för en cirkel. Detta kan kontrolleras genom att sätta in $L = 2\pi r$ och $|\Omega| = \pi r^2$ i uttrycket.

På liknande sätt som när vi härledde arean kan man visa att omkretsen L uppfyller

$$\sum_{n=1}^{\infty} e^{-\lambda_n t} \sim \frac{|\Omega|}{2\pi t} - \frac{L}{4} \frac{1}{\sqrt{2\pi t}} \quad \text{då } t \rightarrow 0.$$

Om vi använder uttrycket med arean och omkretsen med relationen för en cirkel får vi specialfallet

$$\sum_{n=1}^{\infty} e^{-\lambda_n t} \sim \frac{|\Omega|}{2\pi t} - \sqrt{\frac{|\Omega|}{8t}} \quad \text{då } t \rightarrow 0.$$

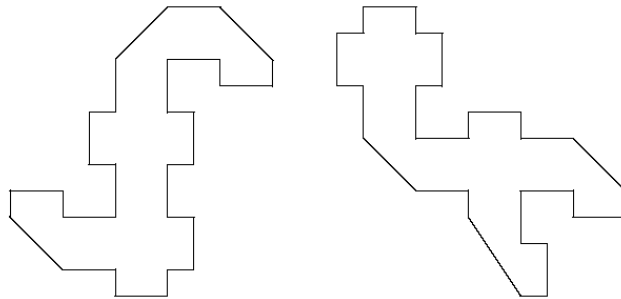
Arean för en cirkulär yta uppfyller detta uttryck om vi känner alla egenvärden.

Motbevis

Det är lätt att urskilja ljudet från olika typer av trummor, det beror på att trumskinet vibrerar med unika frekvenser, vilka beror på storleken, formen och hur hårt trumskinet är spänt. Detta spektrum av egenfrekvenser (rena toner) ger trumman sitt specifika ljud.

Vi har sett att man kan beräkna arean och omkrets av ett svängande membran/trumskinn utifrån att vi känner alla dess egenvärden. Den stora frågan som ska besvaras är om vi kan bestämma formen på trumman utifrån dessa egenvärden. Carolyn S. Gordon, David L. Webb och Scott Wolpert skapade två trummor som utgör ett motbevis.

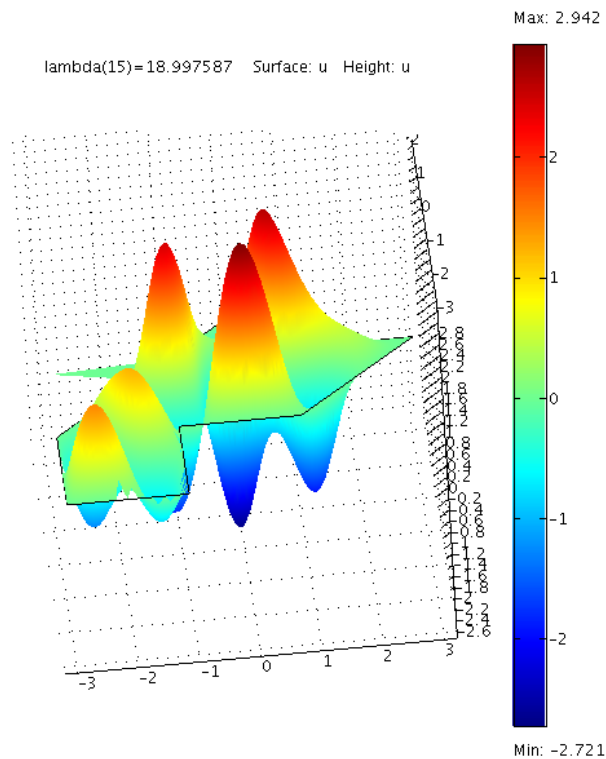
Dessa två figurer har exakt samma area och omkrets, men som kan ses nedan har de ej samma geometriska form. De har dock samma uppsättning egenvär-



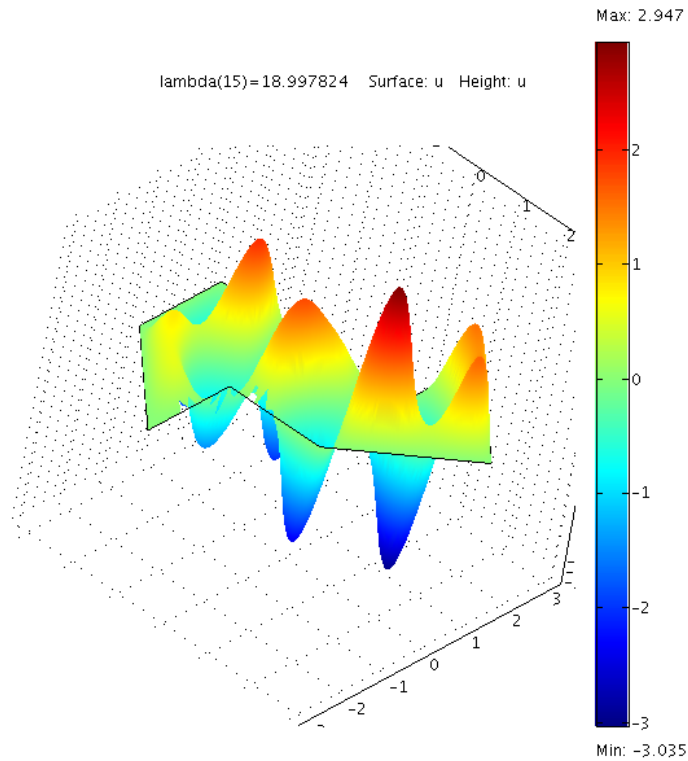
Figur 1: Två isospektrala polygoner med samma uppsättning egenvärden

den, och det är omöjligt att urskilja vilket av dessa membran som givit upphov till en svängning, något som definitivt sätter punkt för Kacs stora frågeställning.

Även vi har studerat två figurer som utgör de enklaste kända geometriska figurer vilka uppfyller motbeviset i programmet FEMLAB, resultatet kan beskådas nedan.



Figur 2: Polygon med dess 18:e egenfunktion



Figur 3: Ytterligare en polygon med dess 18:e egenfunktion

Slutord

När Kac publicerade sin artikel blev detta början på ett problem som gäckat matematiker under senare delen av 1900-talet. Många är dem som fascinerats av frågeställningen, “Kan man höra formen på en trumma?”. Så även vi. Efter att ha gått igenom teorin efter vårt bästa förstånd har vi sammanfattat den komplexa matematiken för att kunna kommunicera detta till vår omgivning. Slutsatsen är att svaret på givna frågeställning är nej.

Avslutningsvis vill vi tacka vår handledare Pelle Pettersson för hans engagemang och hjälpsamhet under vår arbetsprocess.

Källförteckning

M. Kac, Can one hear the shape of a drum?, The American Mathematical Monthly, vol 73, no 4, 1966

C. Gordon, When you can't hear the shape of a manifold, The Mathematical Intelligencer, vol 11, no 3, 1989

C. Gordon, D. Webb, S Wolpert, One cannot hear the shape of a drum, Bulletin Of The American Mathematical Society, vol 27, no 1, 1992

P. Lindqvist, Vad kan man höra? - en aktuell fråga om egenvärden, Normat, no 2, 1994 G. Sparr, A. Sparr, Kontinuerliga system, Studentlitteratur 2000, s. (17-23, 74-81)

Kontinuerliga system, Datorövning 3

<http://www.bea.st/text/drum/drum.html> 2006-04-24