Andragradsekvationen som nåldyna

Håkan Lennerstad
Blekinge Tekniska Högskola

Inledning

En av de mest allmänt kända matematiska formlerna är lösningarna till andragradsekvationen $x^2 + px + q = 0$. Dessa kan ju via kvadratkomplettering skrivas

$$x_1 = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}$$

och

$$x_2 = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q}.$$

Avsikten i denna artikel är att titta på hur lösningarna x_1 och x_2 beter sig som funktioner av de reella koefficienterna p och q. Kan man genom att titta på p och q direkt avgöra om lösningarna är reella? Heltal? Hur mycket ändrar sig värdena på x_1 och x_2 om man rubbar p och q? Vi kommer bland annat att finna en speciell geometrisk form, en viss hyperbolisk paraboloid, som representerar alla andragradsekvationer tillsammans med dess lösningar.

Vi ska alltså syna de två funktionerna

$$x_1(p, q) = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}$$

och

$$x_2(p, q) = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q}$$

närmare i sömmarna, med variablerna p och q.

Indelning av *pq*-planet

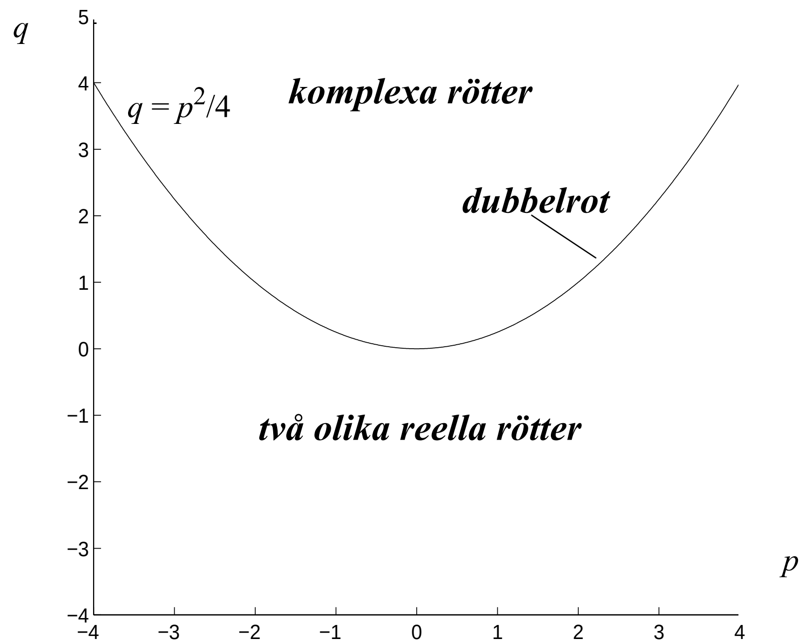
Vi kommer bara att vara intresserade av reella lösningar x_1 och x_2. Därmed får det som står under rottecknet inte vara negativt: vi har villkoret

$$\frac{p^2}{4} - q \geq 0.$$

Geometriskt betyder detta nedanför parabeln $q = p^2/4$ i pq-planet. Dubbelrot har vi då rottecknet är noll, detta inträffar således på parabeln. Kvantiteten $p^2 - 4q$ kallas ibland diskriminant. Särskilt i Storbritannien brukar man ofta undersöka diskriminantens tecken i samband med lösning av andragradsekvationer. Vi sammanfattar detta i en (mycket trivial) sats och i en figur i pq-planet.

Sats 1(*Indelning av pq-planet*): *Antag att p och q är reella tal. Då har andragradsekvationen* $x^2 + px + q = 0$ *lösningar enligt ett av följande tre fall:*

1. $q > p^2/4$: två *olika komplexa rötter*
2. $q = p^2/4$: *en dubbelrot*
3. $q < p^2/4$: två *olika reella rötter.*



Figur 1. Indelning av pq-planet i områden efter lösningarnas (rötternas) typ.

Den första frågan i inledningen kan vi genast svara på med en blick i denna figur. Om q är negativ har vi alltid två reella lösningar. Om q är positiv har vi komplexa lösningar om p är noll eller nära noll, men olika reella om p är ett stort positivt eller stort negativt tal.
Samma sak i formelspråk: vi har två reella lösningar om $q < p^2/4$. Så om q är negativ är olikheten alltid uppfylld. Om q är positiv är den uppfylld bara om p är långt från noll, nämligen om $p > 2\sqrt{q}$ eller $p < -2\sqrt{q}$.

Nivåkurvor för $x_1(p, q)$ och $x_2(p, q)$

Vi kan givetvis rita upp de två ytorna $x_1(p, q) = -p/2 + \sqrt{p^2/4 - q}$ och $x_2(p, q) = -p/2 - \sqrt{p^2/4 - q}$, som alltså är definierade i punkter (p, q) söder om (och på) kurvan $q = p^2/4$. Låt oss göra det genom att rita upp nivåkurvor.
Att vi har dubbelrot på $q = p^2/4$ betyder att $x_2(p, q) = x_1(p, q)$ om p och q uppfyller denna relation. Vi vet därför i förväg att ytorna skär varandra endast i ytterkanten $q = p^2/4$ av defini-

tionsmängden. Hur ser nivåkurvorna ut, och hur sker dessa skärningar?

Vi ska visa följande:

Sats 2: *Nivåkurvorna* $x_1(p, q) = C$ *och* $x_2(p, q) = C$ *är räta linjen* $q = -Cp - C^2$, *vilken tangerar* $p^2/4 - q = 0$ *i punkten* $(-2C, C^2)$. *Delen av tangenten till höger om tangeringspunkten, för* $p \geq -2C$, *är nivåkurva till* $x_1(p, q)$, *och delen till vänster,* $p \leq -2C$, *är nivåkurva till* $x_2(p, q)$.

En nivåkurva är bestämd av $x_{1,2}(p, q) = C$, man hoppas förstås att p eller q kan lösas ut ur ekvationen. Här får vi $-p/2 \pm \sqrt{p^2/4 - q} = C$. Detta ger

$$\pm\sqrt{p^2/4 - q} = C + p/2.$$

När vi kvadrerar får vi samma ekvation för $x_1(p, q) = C$ och $x_2(p, q) = C$:

$$p^2/4 - q = (C + p/2)^2$$

som ger

$$q = -Cp - C^2.$$

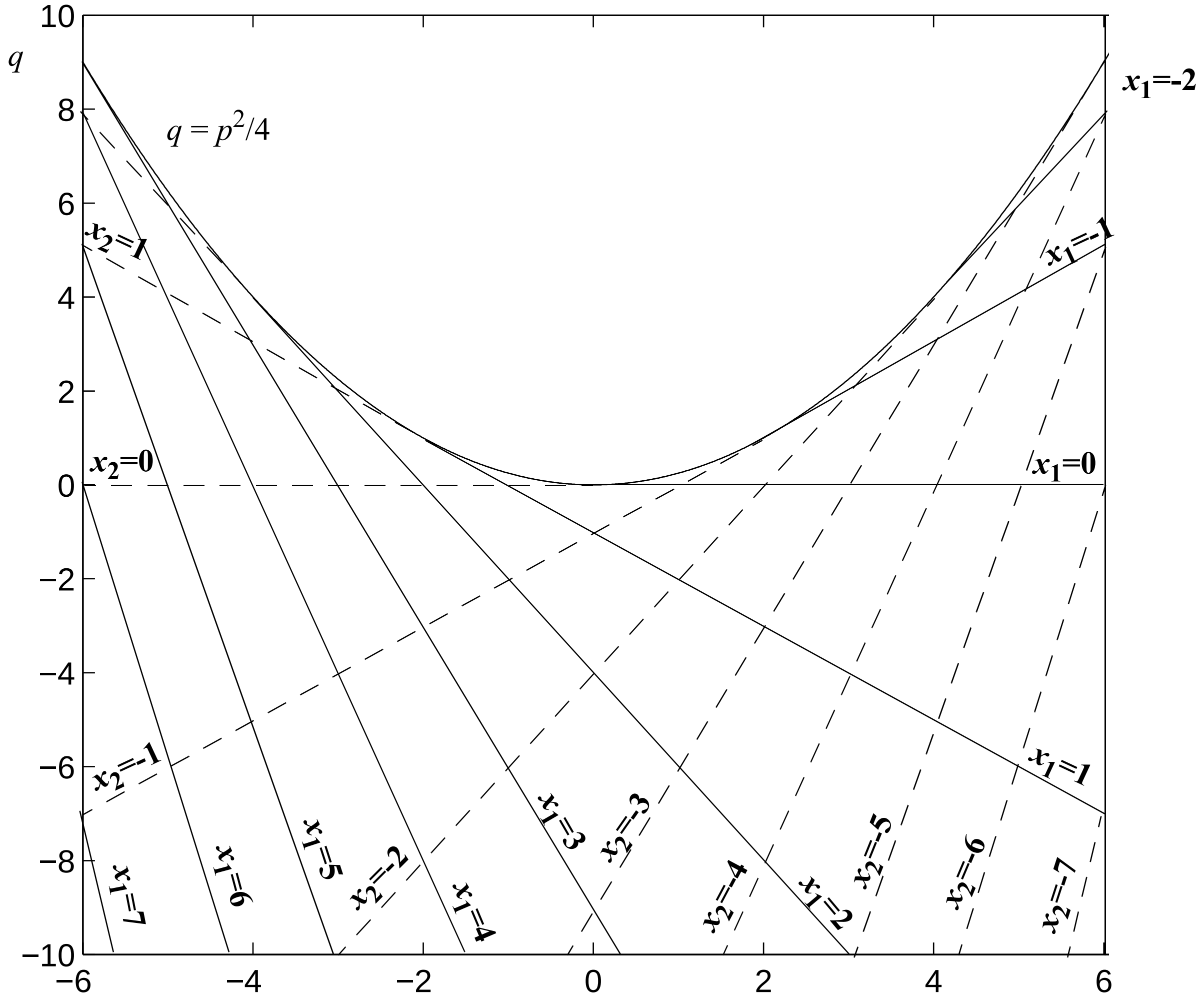
Kvadreringen kräver kontroll av att vi fick riktiga rötter. Insättning i $x_1(p, q) = C$ ger

$$x_1(p, Cp - C^2) = -p/2 + \sqrt{p^2/4 - (-Cp - C^2)} = -p/2 + |C + p/2|$$

som verkligen är C bara då $p \geq -2C$. På återstoden av den räta linjen får vi $x_1(p, q) = -C - p$. På analogt sätt blir $x_2(p, q) = C$ om $p \leq -2C$, annars $x_2(p, q) = -C - p$. De två funktionerna är alltså på denna räta linje styckevis linjära och inte deriverbara i punkten $(p, p^2/4)$.

Beräkning av skärningen mellan $q = -Cp - C^2$ och $q = p^2/4$ ger en ny andragradsekvation: $p^2 + 4Cp + 4C^2 = 0$. Denna har dubbelroten $q = -2C$. I denna punkt, $(-2C, C^2)$, har $q = p^2/4$ lutningen $-C$, vilket är samma lutning som hos nivåkurvan $q = -Cp - C^2$. Därmed är nivåkurvan en tangent. Sats 2 är bevisad.

Sammantaget på svenska: nivåkurvorna för x_1 och x_2 är räta linjer som tangerar dubbelrotsparabeln. Räta linjen till höger om tangeringspunkten är nivåkurva till x_1, delen till vänster är nivåkurva till x_2. Vi visste innan att nivåkurvorna måste skära varandra på dubbelrotsparabeln. Nu vet vi att det sker i form av räta halvlinjer som möts “frontalt”.



Figur 2. Nivåkurvorna för $x_1(p, q)$ (heldragna) och $x_2(p, q)$ (streckade) är räta halvlinjer som möts frontalt.

En biprodukt från kalkylen är att funktionerna $x_1(p, q)$ och $x_2(p, q)$ inte är deriverbara på parabeln. Även som komplexa funktioner i hela pq-planet kommer funktionerna således att vara kontinuerliga men inte differentierbara. Dubbelrotsparabeln är en verklig gräns för funktionernas regularitet.

Vi kan också notera att

$$\lim_{p \to \infty} x_1(p, q) = \lim_{p \to \infty} (-p/2 + \sqrt{p^2/4 - q}) = 0,$$

och på samma sätt $\lim_{p \to -\infty} x_2(p, q) = 0$, i båda fallen för godtyckligt konstant q. Funktionerna går mot plus respektive minus oändligheten åt andra hållet. Ytorna är inte symmetriska runt q-axeln - båda “lutar neråt höger”.

Heltalslösningar

Om $x_1$och x_2 är lösningar till $x^2 + px + q = 0$, så har vi ju

$$x^2 + px + q = (x - x_1)(x - x_2) = 0,$$

multiplikation av parenteserna och identifikation av koefficienter ger

$$p = -x_1 - x_2$$

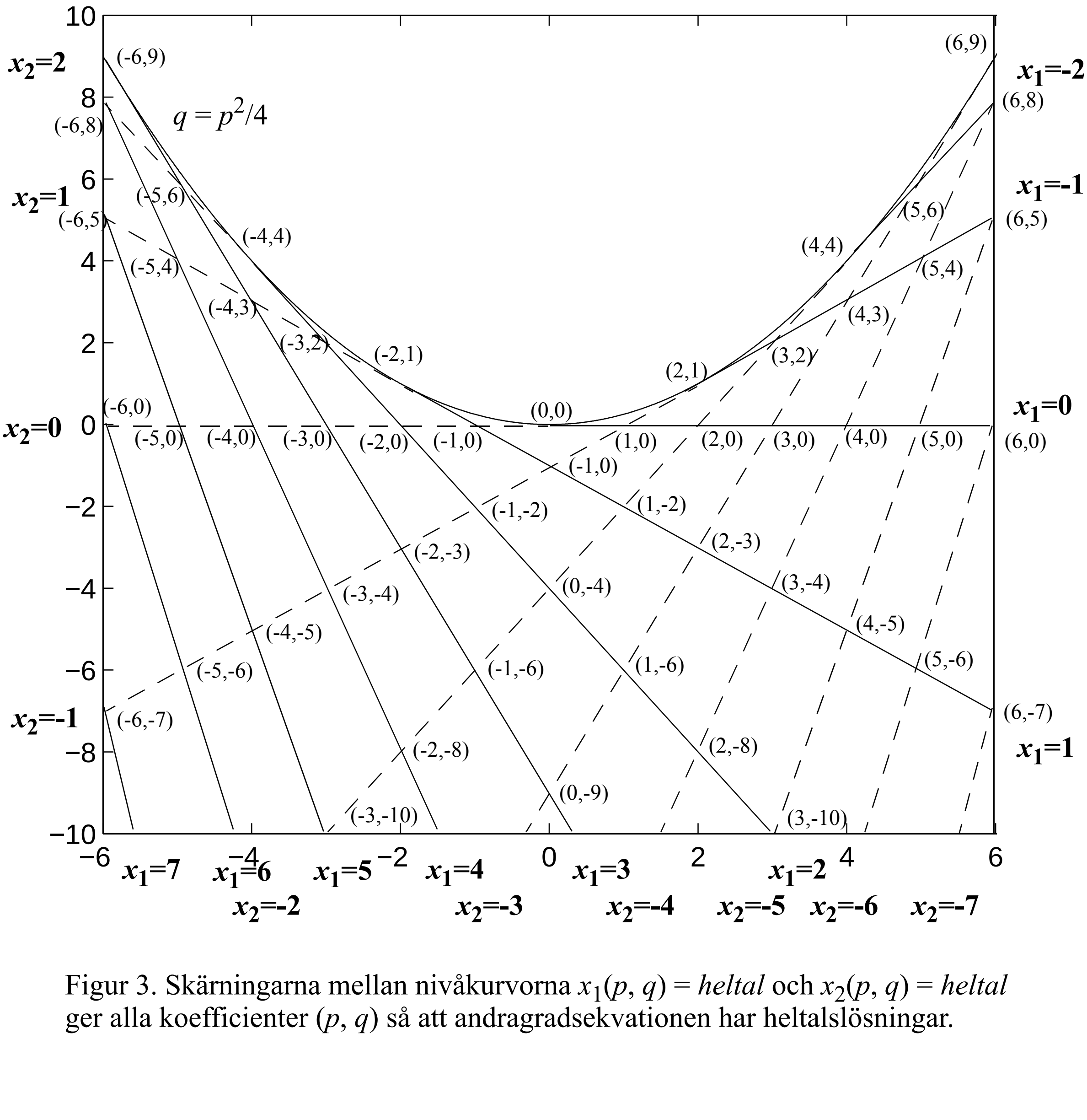
$$q = x_1 x_2 .$$

Detta är en annan utgångspunkt som kan användas för att beräkna nivåkurvorna. Vi kan också använda den för att bestämma de koefficienter som ger heltalslösningar. Skärningarna mellan $x_1(p, q) = n$ och $x_2(p, q) = m$ om n och m är heltal sker givetvis i de punkter (p, q) som ges av

$$p = -n - m$$

$$q = nm .$$

Detta är nog enklaste sättet att generera alla koefficienter p och q som ger heltalslösningar.



Figur 3. Skärningarna mellan nivåkurvorna $x_1(p, q)$ = *heltal* och $x_2(p, q)$ = *heltal* ger alla koefficienter (p, q) så att andragradsekvationen har heltalslösningar.

Lösningarna som variabeltransformation

Vi kan också betrakta

$$x_1 = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}$$

och

$$x_2 = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q}$$

som en variabeltranformation, vars inverstransformation är:

$$p = -x_1 - x_2$$

$$q = x_1 x_2,$$

och med Jacobianen (lokala ytförstoringen)

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial}{\partial x_1}p & \frac{\partial}{\partial x_2}p \\ \frac{\partial}{\partial x_1}q & \frac{\partial}{\partial x_2}q \end{vmatrix} = \begin{vmatrix} -1 & -1 \\ x_2 & x_1 \end{vmatrix} = x_2 - x_1 ..$$

Jacobianen är noll (urartar) endast på linjen $x_1 = x_2$ (dubbelrot!). Så som vi valt x_1 och x_2 är det klart att $x_1 – x_2 = 2\sqrt{p^2/4 - q} \geq 0$, således: $x_1 \geq x_2$. Området under dubbelrotsparabeln i *pq*-planet avbildas därför till området under linjen $x_1 = x_2$. Jacobianen är här negativ. En sluten kurva som ligger i detta område kommer att efter transformation (punkt för punkt) till *pq*-planet fortfarande att vara en sluten kurva (på grund av avbildningens kontinuitet), men som genomlöpes i motsatt riktning (på grund av att Jacobianen är negativ).
Nivåytorna $p(x_1, x_2) = C$ och $q(x_1, x_2) = C$ i detta område blir tydligen räta linjer respektive hyperbler.

Lösningarna som *en* yta

Vi har två funktionsytor som möter varandra i ytterkanten - de två ytorna måste utgöra en sammanhängande yta. Likt de två ytorna $z_1(x, y) = \sqrt{1 - x^2 - y^2}$ och $z_2(x, y) = -\sqrt{1 - x^2 - y^2}$ som är två halvklot, men som givetvis tillsammans bildar en hel sfär. En sfär utan skarv. Uppdelningen är bara bokföringsteknisk, den svarar inte mot någon geometrisk egenskap.
Vi får inte någon skarv i vårt fall heller om vi ändrar vår bokföring till att använda *x* och *p* som variabler. Funktionen $q(x, p)$ blir då

$$q(x, p) = -x^2 - xp ..$$

alltså ett andragradspolynom i variablerna *x* och *p*. För varje fixt *p* har vi en nedåtvänd parabel i *x*, för varje fixt *x* en rät linje i *p*. Om vi tittar på ytan/ytorna från *q*-axeln, norrifrån i Figur 2, kommer vi därför att få en enda funktionsyta. Liksom alla polynom uppvisar den inte några

synliga skarvar: det finns hur många derivator som helst i alla punkter.
Denna geometriska form, vilken representerar alla andragradsekvationer och deras lösningar, är en hyperbolisk paraboloid. Som vi sett ligger den inte helt symmetriskt i *p,q*-planet. Vi ska härnäst beräkna exakt hur den är orienterad.

Vi har fått en kvadratisk form

$$q(x,p) = -x^2 - xp = \begin{pmatrix} x & p \end{pmatrix} \begin{pmatrix} -1 & -\frac{1}{2} \\ -\frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} x & p \end{pmatrix} A \begin{pmatrix} x \\ p \end{pmatrix},$$

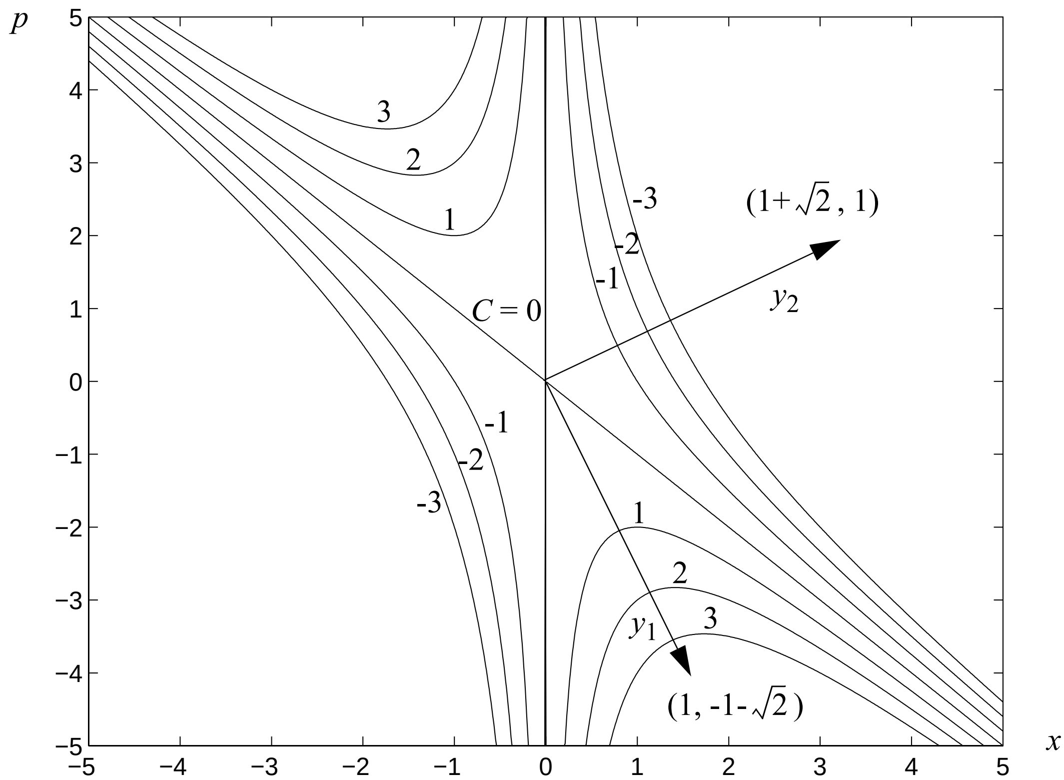
där matrisen A har egenvärdena

$$\lambda = -\frac{1}{2} \pm \frac{1}{\sqrt{2}}.$$

Det finns då en bas y_1, y_2 där den kvadratiska formen kan skrivas i diagonaliserad form:

$$q(y_1, y_2) = \left(\frac{1}{\sqrt{2}} - \frac{1}{2}\right) y_1^2 - \left(\frac{1}{2} + \frac{1}{\sqrt{2}}\right) y_2^2 \ldots$$

Koefficienten framför y_1 är positiv, medan koefficienten framför y_2 är negativ. I skärningen med $y_2 = 0$ har vi alltså en uppåtvänd parabel, i skärningen med $y_1 = 0$ har vi en nedåtvänd.



Figur 3. Några nivåkurvor för $q(x, p)$, med karaktäristiska riktningar.

En hyperbolisk paraboloid har alltid en sadelpunkt i origo. Andragradsekvationens paraboloid är förskjuten något och sammanpressad jämfört med den “kanoniska” hyperboliska paraboloiden (med koefficienter 1 och -1): $q = y_1^2 - y_2^2$. Egenvektorer till A visar sig vara $(1, -1-\sqrt{2})$ och $(1+\sqrt{2}, 1)$. Den relativa sammanpressningen kan mätas som kvoten mellan det till beloppet största och minsta egenvärdet, som är $(-1/2-1/\sqrt{2})/(-1/2+1/\sqrt{2}) = -3 - 2\sqrt{2} \approx -5.83$.

Således är planet $y_1 = 0$, med normalen $(x, p, q) = (1+\sqrt{2}, 1, 0)$, ett symmetriplan för paraboloiden (liksom $y_2 = 0$), detta syns bland annat i figuren. Spegling i detta plan betyder teckenbyte på y_2, som enligt vår diagonalisering inte ändrar värdet på q. Detta är planet

$$(1+\sqrt{2})x + p = 0$$

som ger

$$x = -\frac{1}{1+\sqrt{2}}p.$$

Å andra sidan måste väl $x = -p/2$ ligga mittemellan, eftersom

$$x_{1,2}(p, q) = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}..$$

Skärningskurvan $(-p/2, p, p^2/4)$ mellan de två funktionerna $x_1(p, q)$ och $x_2(p, q)$ ligger ju i detta plan.
Det är klart att vi har “mittemellan” i två olika betydelser - vi speglar i olika riktningar. Ytan förändras inte vid ortogonal spegling i planet $x = -p/(1+\sqrt{2})$. Speglar vi $x_1(p, q)$ i planet $x = -p/2$ ortogonalt **mot** pq**-planet** får vi $x_2(p, q)$ - även här är ytan oförändrad. För övrigt är inte bara ortogonala speglingar linjära.

Vi kan sammanfatta dessa observationer som en sats om andragradsekvationerna som en yta:

Sats 3: *Ytan* $q(x, p) = -x^2 - xp$ *är en hyperbolisk paraboloid. Det är också en kvadratisk form med matrisen*

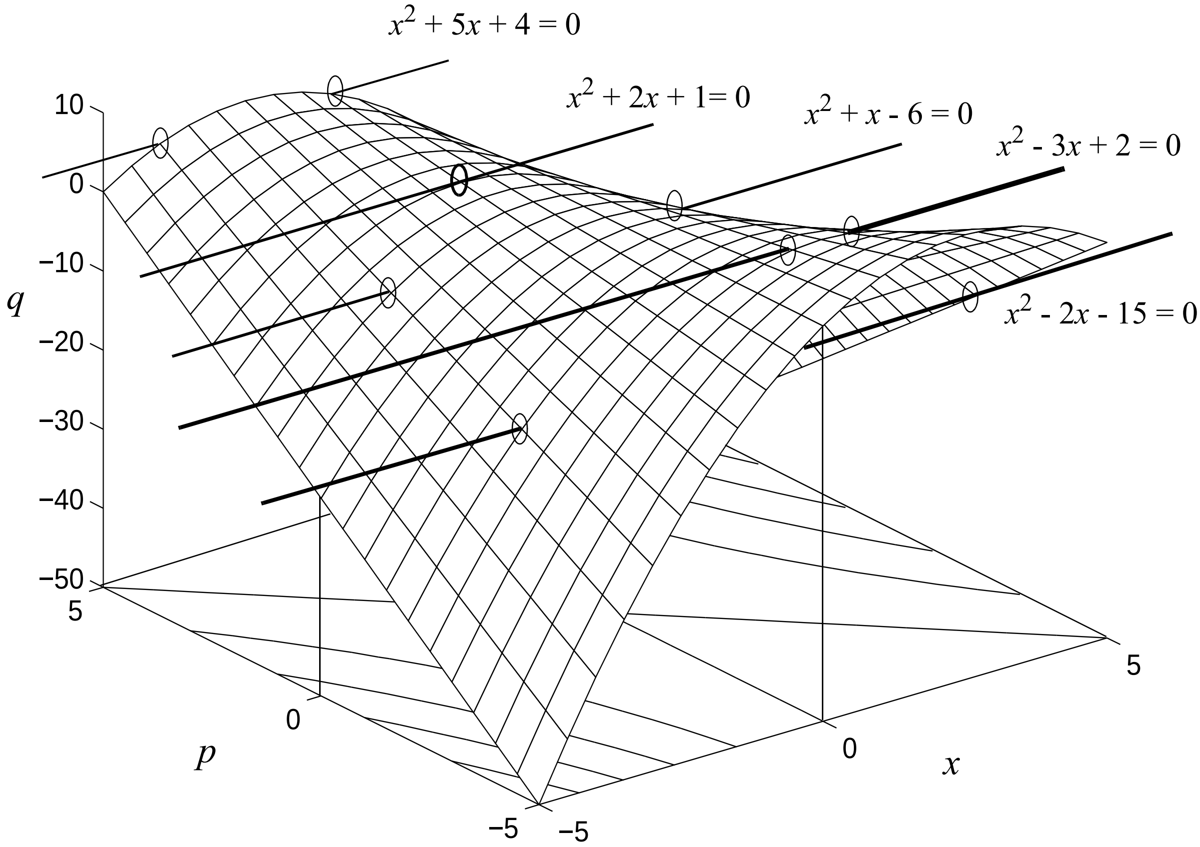
$$A = \begin{pmatrix} -1 & -\frac{1}{2} \\ -\frac{1}{2} & 0 \end{pmatrix},$$

som har egenvektorer (1, -1-$\sqrt{2}$*) och (1+*$\sqrt{2}$*, 1) svarande mot egenvärdena* $-1/2-1/\sqrt{2}$ *respektive* $-1/2+1/\sqrt{2}$. *Ytan är invariant vid ortogonal spegling i planet .*

$$x = -(1/(1+\sqrt{2}))p..$$

Andragradsekvationen som en nåldyna

Man säga att varje rät linje i *xpq*-rummet som är parallel med *x*-axeln svarar mot en andragradsekvation på det viset att varje par (p, q) svarar mot ekvationen $x^2 + px + q = 0$. Några sådana räta linjer tillsammans med $q(x, p)$ visas i följande figur. Här är varje ekvatiion som en nål som sitter i en nåldyna - skärningen med nåldynan är ekvationens lösningar.



Figur 4. Grafen för $q(x, p)$, med några andragradsekvationer och dess lösningar (skärningar med paraboloiden). Figurens bottenplan innehåller några nivåkurvor för $q(x, p)$, vilka bättre kan studeras i Figur 3.

Om en linje som svarar mot $x^2 + px + q = 0$ inte skär den hyperboliska paraboloiden så har ekvationen ingen reell lösning. Om det finns skärningar så ges lösningarna av skärningspunkternas *x*-koordinater. Dubbelrot innebär i denna geometriska tolkning att linjen tangerar paraboloiden - det finns bara en skärningspunkt.

Generaliseringar

På analogt sätt definierar givetvis tredjegradsekvationen $x^3 + px^2 + qx + r = 0$ en särskild geometrisk mängd i ett fyrdimensionellt rum ($\mathbf{R}^4$). Denna kan man betrakta visuellt som en "rörlig" tredimensionell figur - en tredimensionell skärning som med konstant hastighet parallelförflyttar sig i $\mathbf{R}^4$. Vi byter då ut den fjärde rumsdimensionen mot tiden. I detta rum är vi som blinda som måste treva oss fram. Dock kan ju blinda inte sällan skaffa sig en ganska bra uppfattning om rummet...

Detta att betrakta mängden av lösningar som ett geometriskt objekt är en generell matematisk metod. Enklaste exemplet på detta är kanske lösningarna till ett linjärt ekvationssystem. Finns det en lösning kan man se den som en punkt i rummet - ett geometriskt objekt. Finns det många lösningar så bildar de alltid en rät linje eller plan (eller mängd av högre dimension). Ett annat exempel är lösningarna till linjära homogena differentialekvationer, där lösningarna utgör ett linjärt rum. Här byggs rummet inte upp av vektorer som det vanliga rummet, utan av funktioner. Sådana rum kan man inte visualisera på något enkelt sätt. Många sådana abstrakta rum har oändligt antal dimensioner. Ändå fungerar en geometrisk terminologi - detta gör att dessa rum är hanterbara fast man inte kan föreställa sig dem visuellt.
Ett linjärt rum är i sin tur ett specialfall av ett begrepp kallat mångfald. Detta är en differentiabel mångfald om “ytan” lokalt är differentiabel (saknar “skarpa veck”). Detta gäller för den hyperboliska paraboloiden som vi fick från andragradsekvationen - det är en mångfald. Det gäller också för många andra mindre elementära fall, bland annat för lösningsmängderna till de flesta differentialekvationer som är centrala i fysik, kemi och biologi.