Nr 8, 4 april -05, Amelia 2

8 Minsta kvadratmetoden

8.1 Ekvationssystem med en lösning, 2×2-fallet

Ett linjärt ekvationssystem, som

$$\begin{cases} 3x + 7y = 1 \\ x - 2y = 2 \end{cases},$$

har en entydig lösning om koefficientdeterminanten, här $\begin{vmatrix} 3 & 7 \\ 1 & -2 \end{vmatrix}$, är skild från noll. I detta fall har vi $\begin{vmatrix} 3 & 7 \\ 1 & -2 \end{vmatrix} = 3(-2) - 7 = -13 \neq 0$. Determinanten är skild från noll då och endast då kolumnvektorerna är linjärt beroende. Ty om determinanten är noll, dvs $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = 0$, så kan vi lösa ut $d = bc/a$. Det ger kolonnvektorerna $\begin{pmatrix} a \\ b \end{pmatrix}$ och

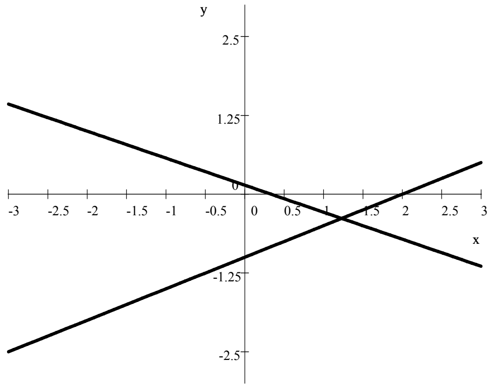
$$\begin{aligned} \begin{pmatrix} c \\ d \end{pmatrix} &= \begin{pmatrix} c \\ bc/a \end{pmatrix} = \{\text{bryt ut } c/a\} \\ &= c/a \begin{pmatrix} a \\ b \end{pmatrix}. \end{aligned}$$

Så de två kolonnvektorerna $\begin{pmatrix} a \\ b \end{pmatrix}$ och $\begin{pmatrix} c \\ d \end{pmatrix} = c/a \begin{pmatrix} a \\ b \end{pmatrix}$ är linjärt beroende – den ena är en faktor ac av den andra. Om $a = 0$ kan vi i stället lösa ut c.

En geometrisk tolkning av ekvationssystemet

$$\begin{cases} 3x + 7y = 1 \\ x - 2y = 2 \end{cases}$$

är skärningen av två räta linjer, $3x + 7y = 1$ och $x - 2y = 2$. Båda ekvationerna uppfyllda betyder att punkten (x, y) måste ligga på båda linjerna, dvs i skärningspunkten.



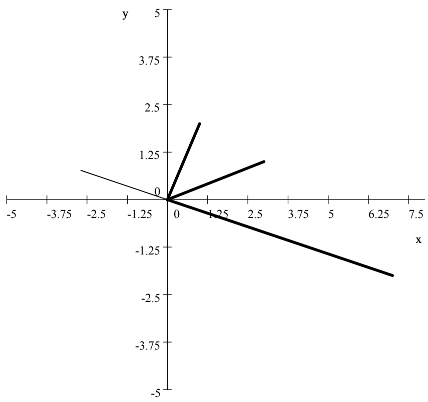
tLösningen är $x = \frac{16}{13}$ och $y = -\frac{5}{13}$.
En annan geometrisk tolkning av ekvationssystemet

$$\begin{cases} 3x + 7y = 1 \\ x - 2y = 2 \end{cases}$$

är

$$\begin{pmatrix} 3 \\ 1 \end{pmatrix} x + \begin{pmatrix} 7 \\ -2 \end{pmatrix} y = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Vi väljer här talen x och y så att vektorn $\binom{1}{2}$ kan bildas som en linjärkombination av $\binom{3}{1}$ och $\binom{7}{-2}$, som är kolonnvektorer i matrisen. Uppenbarligen kan detta ske på endast ett sätt (vi har redan löst ekvationssystemet), och då måste $x = \frac{16}{13}$ och $y = -\frac{5}{13}$. Denna senare tolkning bygger minsta kvadratmetoden på.



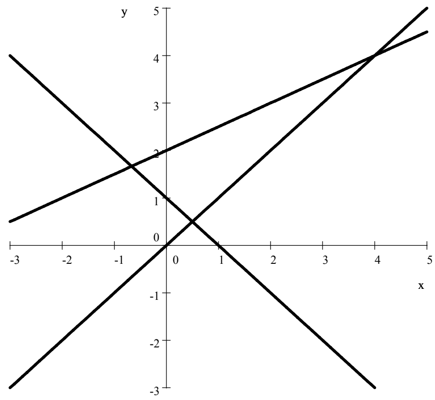
Vektorerna $(3, 1)$, $(7, -2)$, $(1, 2)$ och $-\frac{5}{13}(7, -2)$ (streckad).

8.2 Ekvationssystem utan lösning

Ekvationssystemet

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

har inte någon lösning, för det representerar tre räta linjer som inte har någon gemensam punkt.



Men vad är den minst dåliga lösningen, den som är "ganska nära" alla tre räta linjerna? Problemet kan skrivas $Ax = b$. Det finns alltså inget x så att $Ax - b = 0$ i detta fall. I brist på exakt lösning söker vi då istället ett x så att vektorn $Ax - b$ är så kort som möjligt (om vi har en lösning så är längden av denna vektor noll). Det är grundidén i minsta kvadratmetoden.

Vilka x har vi att välja på? Vi tolkar uttrycket Ax som en linjärkombination av A:s kolonnvektorer, som vi kan beteckna med A_1 och A_2:

$$\begin{aligned} Ax &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \\ a_{31}x_1 + a_{32}x_2 \end{pmatrix} = \{\text{matrislagar}\} \\ &= \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} x_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} x_2 \\ &= A_1x_1 + A_2x_2. \end{aligned}$$

Här är x_1 och x_2 två reella tal. Vi söker bland alla tänkbara linjärkombinationer $A_1x_1 + A_2x_2$, för olika x_1 och x_2. De bildar ett plan i rummet som går genom origo (ty med $x_1 = x_2 = 0$ får vi $A_1 \cdot 0 + A_2 \cdot 0 = \mathbf{0}$). Vektorn b ligger inte i detta plan, ty om b gjorde det så skulle vi kunna välja x_1 och x_2 så att vi har $A_1x_1 + A_2x_2 = b$ alltså $Ax = b$.

Vektorn $Ax - b$ går mellan b och en punkt Ax i planet. Denna vektor är kortast då $Ax - b$ är ortogonal mot planet. Att den är ortogonal betyder att den är ortogonal mot alla vektorer i planet, bl.a. mot A_1 och A_2, dvs att skalärprodukten är noll. Skalärprodukt av kolonnvektorerna u och v kan skrivas i matrisnotation som $u^T v$.

Att skalärprodukten mellan A_1 och $Ax-b$ är noll kan skrivas $A_1^T(Ax-b) = 0$. Vi har också $A_2^T(Ax-b) = 0$. Dessa två ekvationer kan skrivas som en ekvation:

$$\begin{aligned} A^T(Ax-b) &= 0, \text{ dvs} \\ A^T Ax &= A^T b. \end{aligned}$$

Definition 1 *Ekvationen $A^T Ax = A^T b$ kallas **normalekvationen**.*

Observera att argumentet ovan gäller i vilken dimension som helst (om vi byter "plan genom origo" mot "linjärt rum"), dvs vi kan anta att A har dimension $n \times m$. Vi har visat att

Theorem 2 *En lösning x till $A^T Ax = A^T b$ minimerar $|(Ax-b)|$.*

Här är $|\ |$ vektornorm: $|x| = \sqrt{x_1^2 + ... + x_n^2}$. Observera att koefficientmatrisen $A^T A$ till detta ekvationssystem alltid är kvadratisk och symmetrisk. Kvadratisk ty om A har dimension $n \times m$ så har A^T dimension $m \times n$ och $A^T A$ har dimension $m \times m$. Symmetrisk ty

$$\begin{aligned} (A^T A)^T &= \{\text{för matriser gäller ju } (AB)^T = B^T A^T\} \\ &= A^T (A^T)^T = A^T A. \end{aligned}$$

Så matrisen $A^T A$ är oförändrad vid transponering, således symmetrisk.

$A^T A$ är emellertid inte alltid inverterbar. Om t.ex. A är nollmatrisen, så kommer även $A^T A$ att vara nollmatrisen (men med annorlunda dimension), som inte är inverterbar. Vi har inte ännu undersökt om det alltid finns någon lösning till normalekvationen, eller under vilka villkor det finns flera lösningar. Vi räknar först ett exempel – det ekvationssystem som är omnämnt ovan.

Exempel 3 *Beräkna en minsta kvadratlösning till*

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

och medelfelet.

Lösning: Vi har evationssystemet

$$\begin{pmatrix} 1 & 1 \\ -1 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix},$$

så vi får

$$
\begin{aligned}
A &= \begin{pmatrix} 1 & 1 \\ -1 & 2 \\ 1 & -1 \end{pmatrix} \text{ som ger} \\
A^T A &= \begin{pmatrix} 1 & -1 & 1 \\ 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 2 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 6 \end{pmatrix}.
\end{aligned}
$$

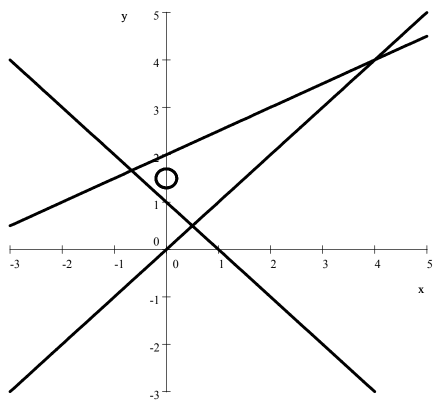
Högerledet blir

$$
\begin{aligned}
b &= \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \text{ som ger} \\
A^T b &= \begin{pmatrix} 1 & -1 & 1 \\ 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ 9 \end{pmatrix}.
\end{aligned}
$$

Normalekvationen $A^T Ax = A^T b$ är i detta fall

$$
\begin{pmatrix} 3 & -2 \\ -2 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -3 \\ 9 \end{pmatrix}.
$$

Detta ger lösingen $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{3}{2} \end{pmatrix}$. Detta är minsta kvadratlösningen. Det är inte nödvändigtvis en mittpunkt i den triangel som bildas av tre linjer som inte är parallella och inte skär varandra i en gemensam punkt. Detta syns i nedanstående figur.



Svar: $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{3}{2} \end{pmatrix}$.

8.3 Normalekvationens lösbarhet

Naturligtvis har inte alla ekvationssystem med kvadratiska koefficientmatriser lösningar, inte ens om matrisen är symmetrisk (tag exempelvis A som nollmatrisen och ett högerled som inte är endast nollor). Men normalekvationen är inte vilket sådant ekvationssystem som helst, ty det har alltid lösningar. Orsaken är att vi har inte vilket högerled som helst, det är transformerat med A^T.

Theorem 4 *Normalekvationen $A^TAx = A^Tb$ har alltid minst en lösning.*

Satsen följer från härledningen av normalekvationen. Minimum av funktionen $|Ax - b|$ existerar eftersom det är en nedåt begränsad funktion (≥ 0) och mängden tillåtna värden x är icke-tom och en sluten mängd. Ett sådant minimum realiseras av en lösning till $A^TAx = A^Tb$, ty om x_0 är en lösning till normalekvationen och x en punkt vilken som helst, så gäller att (Pythagoras sats!)

$$\begin{aligned}
|Ax - b|^2 &= |A(x - x_0 + x_0) - b|^2 \\
&= |Ax_0 - b + A(x - x_0)|^2 \\
&= |Ax_0 - b|^2 + |A(x - x_0)|^2 + 2(x - x_0)^T A^T (Ax_0 - b) \\
\{A^T(Ax_0 - b) \text{ är noll!}\} &= |Ax_0 - b|^2 + |A(x - x_0)|^2.
\end{aligned}$$

Hur kan vi nu välja x så att

$$|Ax - b|^2 = |Ax_0 - b|^2 + |A(x - x_0)|^2$$

är minimalt? Jo genom att välja $x = x_0$,ty då försvinner den andra termen:

$$|Ax - b|^2 = |Ax_0 - b|^2 + 0.$$

Alltså har normalekvationen alltid någon lösning, ty den sammanfaller med ett minimum som existerar.

Theorem 5 *Normalekvationen $A^TAx = A^Tb$ har minst två lösningar om och endast om kolonnerna i A är linjärt beroende.*

Observera att kravet är att kolonnerna i A är linjärt beroende, inte kolonnerna i A^TA, vilket vore självklart. En kvadratisk matris är ju inverterbar om och endast om dess kolonner är linjärt oberoende om och endast om motsvarande ekvationssystem har exakt en lösning.

Man kan alltså genom att studera A avgöra om normalekvationen har en eller flera lösningar. Vi utelämnar detta bevis.

8.3.1 Kurvpassning

Minsta kvadratproblem kan också uppstå ur ett motsatt problem: kurvpassing. Här anpassar man en kurva till ett antal punkter, så att summan av kvadratfelen är så litet som möjligt.

Exempel 6 *Antag att vi vill finna en rät linje som går genom punkterna* $(1, -1)$, $(2, 2)$, $(3, 2)$ *och* $(4, 6)$. *Dessa fyra punkter ligger inte på rät linje, så någon exakt lösning finns inte. Vi söker en minstakvadratlösning, och medelfelet.*

Lösning: Insättning av punkterna i $y = kx + l$ ger de fyra kraven

$$\begin{aligned} -1 &= k + l \\ 2 &= 2k + l \\ 2 &= 3k + l \\ 6 &= 4k + l. \end{aligned}$$

Men detta är ett linjärt ekvationssystem:

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} k \\ l \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 2 \\ 6 \end{pmatrix}.$$

Här får vi

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}$$

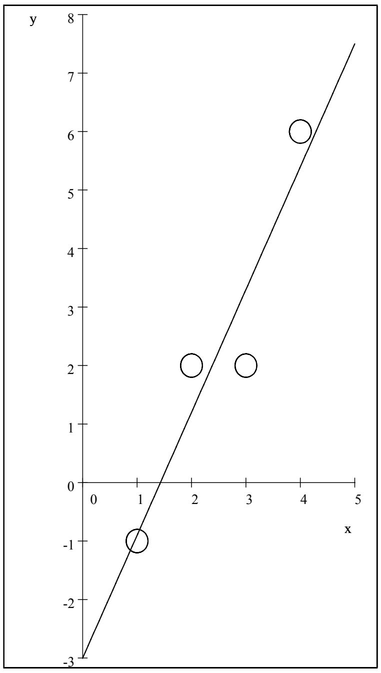
och

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 33 \\ 9 \end{pmatrix},$$

så normalekvationen är här

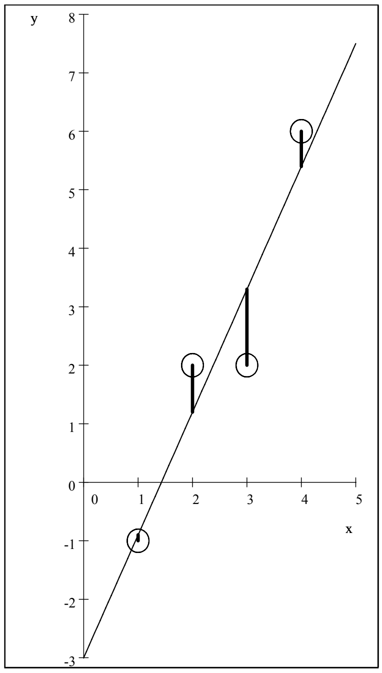
$$\begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix} \begin{pmatrix} k \\ l \end{pmatrix} = \begin{pmatrix} 33 \\ 9 \end{pmatrix}$$

med lösning $\binom{k}{l} = \begin{pmatrix} \frac{21}{10} \\ -3 \end{pmatrix}$.



Räta linjen $\frac{21}{10}x - 3$.

Summan av kvadraterna på följande sträckor minimeras av lösningen $y = 2.1x - 3$.



Sträckor vars kvadratsumma minimeras.

Funktionen $y = 2.1x - 3$. tar i punkterna 1, 2, 3, och 4 värdena -0.9, 1.2, 3.3 respektive 5.4, medan de önskade värdena är -1, 2, 2, respektive 6. Felen är alltså 0.1, 0.8, 1.3 respektive -0.6. Det linjära medelfelet är då medelvärdet av beloppen ($\frac{1}{n}(|e_1| + ... + |e_n|)$)

$$\frac{1}{4}(0.1 + 0.8 + 1.3 + 0.6) = 0.7.$$

Det kvadratiska medelfelet ($\frac{1}{n}\sqrt{e_1^2 + ... + e_n^2}$) är här

$$\frac{1}{4}\sqrt{0.01 + 0.64 + 1.69 + 0.36} = 0.41079.$$

Det kvadratiska medelfelet minimeras när man använder minstakvadratmetoden.

Svar: Rät linje $y = 2.1x - 3$ med kvadratiskt medelfel är 0.41 079.

Exempel 7 *Bestäm den andragradskurva som bäst anpassar sig till punkterna* $(-1, 1)$, $(0, 0)$, $(1, 1)$, $(2, 8)$ *i minsta kvadratmening och beräkna medelfelet.*

Lösning: Med $y = ax^2 + bx + c$ får vi här följande fyra villkor.

$$\begin{aligned} a - b + c &= 1 \\ c &= 0 \\ a + b + c &= 1 \\ 4a + 2b + c &= 8. \end{aligned}$$

Vi får det linjära ekvationssystemet

$$\begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 8 \end{pmatrix}.$$

Vi får

$$A^T A = \begin{pmatrix} 1 & 0 & 1 & 4 \\ -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 18 & 8 & 6 \\ 8 & 6 & 2 \\ 6 & 2 & 4 \end{pmatrix}$$

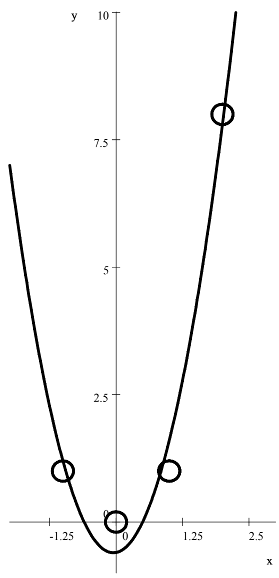
och

$$A^T b = \begin{pmatrix} 1 & 0 & 1 & 4 \\ -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 8 \end{pmatrix} = \begin{pmatrix} 34 \\ 16 \\ 10 \end{pmatrix}.$$

Normalekvaitonen $A^T Ax = A^T b$ är ekvationssystemet

$$\begin{pmatrix} 18 & 8 & 6 \\ 8 & 6 & 2 \\ 6 & 2 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 34 \\ 16 \\ 10 \end{pmatrix}.$$

Detta ekvationssystem kan lösas med exempelvis Gausselimination. Det ger $a = 2$, $b = 1/5$ och $c = -3/5$. Minstakvadratlösningen är alltså $2x^2 + \frac{1}{5}x - \frac{3}{5}$.



Eftersom punktena är tagna osymmetriskt runt y-axeln, kommer lösningen att vara osymmetrisk (pga att termen $\frac{1}{5}x$ finns med). Funktionen $2x^2 + \frac{1}{5}x - \frac{3}{5}$ tar i punkterna $x = -1, 0, 1, 2$ värdena $1.2, -0.6, 1.6, 7.8$ jämfört med de önskade $1, 0, 1, 8$. Vi har alltså det kvadratiska medelfelet

$$\frac{1}{4}\sqrt{0.04 + 0.36 + 0.36 + 0.04} = 0.223\,61.$$

Svar: Minstakvadratlösningen är $2x^2 + \frac{1}{5}x - \frac{3}{5}$ med kvadratiska medelfelet 0.22361.

Exempel 8 *Bestäm den andragradskurva som bäst anpassar sig till punkterna* $(-2, 8)$, $(-1, 1)$, $(0, 0)$, $(1, 1)$, $(2, 8)$ *i minsta kvadratmening, och beräkna medelfelet.*

Till skillnad från förra problemet är dessa punkter symmetriskt placerad runt y-axeln, och vi får en symmetrisk minstakvadratlösning – vi får $b = 0$.

Lösning: Med $y = ax^2 + bx + c$ får vi från de fem punkterna fem villkor.

$$\begin{aligned}
4a - 2b + c &= 8 \\
a - b + c &= 1 \\
c &= 0 \\
a + b + c &= 1 \\
4a + 2b + c &= 8.
\end{aligned}$$

På liknande sätt får vi det linjära ekvationssystemet

$$\begin{pmatrix} 4 & -2 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 8 \\ 1 \\ 0 \\ 1 \\ 8 \end{pmatrix}.$$

Så

$$A^T A = \begin{pmatrix} 4 & 1 & 0 & 1 & 4 \\ -2 & -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 4 & -2 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 34 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 5 \end{pmatrix},$$

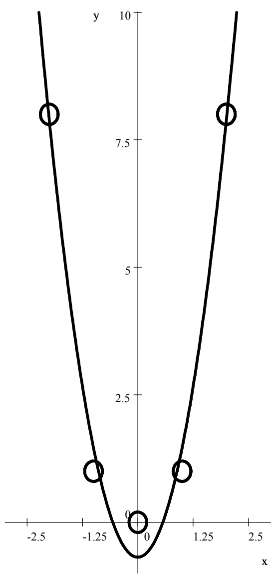
och :

$$A^T b = \begin{pmatrix} 4 & 1 & 0 & 1 & 4 \\ -2 & -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 8 \\ 1 \\ 0 \\ 1 \\ 8 \end{pmatrix} = \begin{pmatrix} 66 \\ 0 \\ 18 \end{pmatrix}.$$

Normalekvationen blir nu

$$\begin{pmatrix} 34 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 5 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 66 \\ 0 \\ 18 \end{pmatrix},$$

som har lösning $\begin{pmatrix} \frac{15}{7} \\ 0 \\ -\frac{24}{35} \end{pmatrix}$. Nu får vi alltså andragradskurvan $\frac{15}{7}x^2 - \frac{24}{35}$, som är symmetrisk, till skillnad från i det förra fallet.



Vi har $\frac{15}{7}2^2 - \frac{24}{35} = \frac{276}{35}$ och $\frac{15}{7} - \frac{24}{35} = \frac{51}{35}$, så får här det kvadratiska medelfelet

$$\frac{1}{5}\sqrt{(\frac{276}{35} - 8)^2 + (\frac{51}{35} - 1)^2 + (\frac{24}{35} - 0)^2 + (\frac{51}{35} - 1)^2 + (\frac{276}{35} - 8)^2} = 0.191\,24.$$

Medelfelet blev något mindre med en punkt till. Man får inte nödvändigtvis bättre överensstämmelse med flera punkter, ty överensstämmelsen beror på vilken del av funktionen (x^3) vi arbetar med – hur bra den stämmer överens med approximerande funktion i området.

Svar: $y = \frac{15}{7}x^2 - \frac{24}{35}$, kvadratiskt medelfel $0.191\,24$.

8.4 Kvasiinvers

Som bekant är en delmängd av de kvadratiska matriserna inverterbara. Om A är inverterbar så finns det en annan kvadratisk matris med samma dimension, inversen A^{-1}, så att $AA^{-1} = E$ och $A^{-1}A = E$.

För en icke inverterbar matris A, som inte ens behöver vara kvadratisk, kan man emellertid definera en **kvasiinvers**, betecknat med A^+. Om A är inverterbar så är inversen kvasiinversen ($A^+ = A^{-1}$). Om man vill lösa ekvationssystemet $Ax = b$ för en generell matris A, så kommer $x = A^+b$ att vara just en minsta kvadratlösning om ekvationssystemet saknar exakta lösningar. Dessutom, om det finns många minstakvadratlösningar, så kommer $x = A^+b$ att vara den minstakvadratlösning som är närmast origo. Vi går här inte in på explicit definition och beräkningsalgoritmer för kvasiinversen.

Symboliska programvaror som MATLAB ger ofta kvasiinversen till en matris om den inte är inverterbar, utan någon särskild kommentar.