Krister Svanberg, april 2012

1 Minsta-kvadratproblem (MK-problem)

Betrakta ett ekvationssystem $\mathbf{A}\mathbf{x} = \mathbf{b}$, med $\mathbf{b} \in I\!R^m$, $\mathbf{x} \in I\!R^n$ och $\mathbf{A} \in I\!R^{m\times n}$. I många tillämpningar kan det vara så att $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$, varvid ekvationssystemet saknar lösning. Ofta vill man då i stället bestämma en punkt $\mathbf{x} \in I\!R^n$ som "så nära som möjligt" uppfyller $\mathbf{A}\mathbf{x} = \mathbf{b}$. Ett naturligt mått på att $\mathbf{x}$ "nästan" uppfyller $\mathbf{A}\mathbf{x} = \mathbf{b}$ är att $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2 = (\mathbf{A}\mathbf{x}-\mathbf{b})^\mathsf{T}(\mathbf{A}\mathbf{x}-\mathbf{b})$ är "liten". Därmed leds man till följande minsta-kvadratproblem (MK-problem):

$$\text{minimera } \ \tfrac{1}{2}(\mathbf{A}\mathbf{x}-\mathbf{b})^\mathsf{T}(\mathbf{A}\mathbf{x}-\mathbf{b}) \ \text{ då } \mathbf{x} \in I\!R^n. \tag{1.1}$$

Man vill alltså minimera längden i kvadrat på "felvektorn" $\mathbf{A}\mathbf{x}-\mathbf{b}$. Faktorn $\frac{1}{2}$ är instoppad bara för att förenkla vissa uttryck längre fram.

1.1 Ett modellanpassningsexempel

Antag att s är en storhet som beror av variabeln t, säg $s = g(t)$ där g är en inte helt känd funktion. Antag vidare att man baserat på givna mätdata ska skatta funktionen g. Mätdata består av m stycken givna punkter $(t_1, s_1), \ldots, (t_m, s_m)$, där s_i är ett uppmätt värde på s för $t = t_i$. Ett vanligt angreppssätt är då att man gör en ansats på formen

$$g(t) \approx \alpha_1\psi_1(t) + \alpha_2\psi_2(t) + \ldots + \alpha_n\psi_n(t) = \sum_{j=1}^{n} \alpha_j\psi_j(t), \tag{1.2}$$

där $\psi_j(t)$ är givna "basfunktioner" medan α_j är okända koefficienter. Basfunktionerna kan exempelvis vara polynom, $\psi_j(t) = t^{\,j-1}$, trigonometriska funktioner, $\psi_j(t) = sin\dfrac{2\pi jt}{T}$, etc. Idealt skulle man vilja välja koefficienterna α_j så att

$$\sum_{j=1}^{n} \alpha_j\psi_j(t_i) = s_i, \ \text{för } i = 1, \ldots, m, \tag{1.3}$$

men eftersom det typiskt gäller att
1. antalet mätpunkter är större än antalet koefficienter α_j, dvs $m > n$,
2. approximationen i (1.2) inte är exakt, och
3. mätningarna av s innehåller mätfel,
så kan man normalt inte lösa ekvationssystem.(1.3). Man söker då istället de värden på koefficienterna α_j som gör följande kvadratsumma så liten som möjligt:

$$\tfrac{1}{2}\sum_{i=1}^{m}\left(\sum_{j=1}^{n}\alpha_j\psi_j(t_i) - s_i\right)^2. \tag{1.4}$$

Genom att införa följande matris och vektorer:

$$\mathbf{A} = \begin{bmatrix} \psi_1(t_1) & \cdots & \psi_n(t_1) \\ \psi_1(t_2) & \cdots & \psi_n(t_2) \\ \vdots & & \vdots \\ \psi_1(t_m) & \cdots & \psi_n(t_m) \end{bmatrix}, \quad \mathbf{b} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \tag{1.5}$$

så kan ekvationssystemet (1.3) skrivas på formen $\mathbf{Ax} = \mathbf{b}$, medan problemet att minimera kvadratsumman (1.4) kan skrivas på formen (1.1) ovan.

1.2 Existens av optimala lösningar till MK-problemet

Kalla målfunktionen till MK-problemet (1.1) för f, dvs

$$f(\mathbf{x}) = \tfrac{1}{2}(\mathbf{Ax} - \mathbf{b})^{\mathsf{T}}(\mathbf{Ax} - \mathbf{b}) = \tfrac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{Ax} - \mathbf{b}^{\mathsf{T}}\mathbf{Ax} + \tfrac{1}{2}\mathbf{b}^{\mathsf{T}}\mathbf{b}. \tag{1.6}$$

Vi ser att f är en kvadratisk funktion med $\mathbf{H} = \mathbf{A}^{\mathsf{T}}\mathbf{A}$, $\mathbf{c} = -\mathbf{A}^{\mathsf{T}}\mathbf{b}$ och $c_0 = \frac{1}{2}\mathbf{b}^{\mathsf{T}}\mathbf{b}$.
Ett trevligt faktum är att det alltid finns minst en minimalpunkt $\hat{\mathbf{x}}$ till f. Enligt ett tidigare avsnitt har den kvadratiska funktionen f minst en minimalpunkt om $\mathbf{H}$ är positivt semidefinit och $\mathbf{c} \in \mathcal{R}(\mathbf{H})$, vilket är uppfyllt här eftersom $\mathbf{H} = \mathbf{A}^{\mathsf{T}}\mathbf{A}$ alltid är positivt semidefinit och $\mathbf{c} = -\mathbf{A}^{\mathsf{T}}\mathbf{b} = \mathbf{A}^{\mathsf{T}}(-\mathbf{b}) \in \mathcal{R}(\mathbf{A}^{\mathsf{T}}) = \mathcal{R}(\mathbf{A}^{\mathsf{T}}\mathbf{A}) = \mathcal{R}(\mathbf{H})$.

1.3 Normalekvationerna till MK-problemet

Varje minimalpunkt $\mathbf{x}$ till en kvadratisk funktion erhålls enligt ett tidigare avsnitt som lösning till ekvationssystemet $\mathbf{Hx} = -\mathbf{c}$, som här får utseendet

$$\mathbf{A}^{\mathsf{T}}\mathbf{Ax} = \mathbf{A}^{\mathsf{T}}\mathbf{b}. \tag{1.7}$$

Detta ekvationssystem brukar kallas *normalekvationerna* för MK-problemet (1.1).

Eftersom $\mathbf{x}$ är en lösning till ekvationssystemet (1.7) om och endast om $\mathbf{x}$ är en optimal lösning till MK-problemet (1.1), och vi ovan konstaterade att MK-problemet alltid har minst en optimal lösning, så betyder det att systemet (1.7) alltid har minst en lösning! Alternativt följer detta av att högerledsvektorn i systemet (1.7) faktiskt tillhör $\mathcal{R}(\mathbf{A}^{\mathsf{T}}\mathbf{A})$, eftersom $\mathcal{R}(\mathbf{A}^{\mathsf{T}}) = \mathcal{R}(\mathbf{A}^{\mathsf{T}}\mathbf{A})$.

Även om det i vissa fall finns oändligt många olika lösningar $\mathbf{x}$ till normalekvationerna så gäller att vektorn $\mathbf{Ax}$ är densamma för samtliga dessa lösningar. Om nämligen både $\mathbf{A}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\mathbf{b}$ och $\mathbf{A}^{\mathsf{T}}\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\mathbf{b}$ så är $\mathbf{A}^{\mathsf{T}}\mathbf{A}(\hat{\mathbf{x}} - \bar{\mathbf{x}}) = \mathbf{0}$, dvs $\hat{\mathbf{x}} - \bar{\mathbf{x}} \in \mathcal{N}(\mathbf{A}^{\mathsf{T}}\mathbf{A})$.
Men eftersom $\mathcal{N}(\mathbf{A}^{\mathsf{T}}\mathbf{A}) = \mathcal{N}(\mathbf{A})$ så erhålls att $\mathbf{A}(\hat{\mathbf{x}} - \bar{\mathbf{x}}) = \mathbf{0}$, dvs $\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}\bar{\mathbf{x}}$.

Omvänt gäller att om $\bar{\mathbf{x}}$ är en lösning till (1.7) och $\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}\bar{\mathbf{x}}$ så gäller att även $\hat{\mathbf{x}}$ är en lösning till (1.7) eftersom $\mathbf{A}^{\mathsf{T}}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}(\mathbf{A}\hat{\mathbf{x}}) = \mathbf{A}^{\mathsf{T}}(\mathbf{A}\bar{\mathbf{x}}) = \mathbf{A}^{\mathsf{T}}\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\mathbf{b}$.

1.4 Geometrisk tolkning av MK-problemet och normalekvationerna

Det finns en naturlig geometrisk tolkning av MK-problemet (1.1) och tillhörande normalekvationer (1.7). Problemet (1.1) kan tolkas som att man vill bestämma den punkt i underrummet

$\mathcal{R}(\mathbf{A})$ som ligger närmast punkten $\mathbf{b} \in I\!R^m$, ty problemet kan ekvivalent skrivas

$$\begin{array}{rl} \text{minimera} & \frac{1}{2}\|\,\mathbf{y}-\mathbf{b}\,\|^2 \\ \text{då} & \mathbf{y} \in \mathcal{R}(\mathbf{A}). \end{array} \tag{1.8}$$

Normalekvationerna (1.7) kan skrivas $\mathbf{A}^{\mathsf{T}}(\mathbf{b}-\mathbf{A}\mathbf{x}) = \mathbf{0}$, vilket är ekvivalent med att $\mathbf{b}-\mathbf{A}\mathbf{x} \in \mathcal{N}(\mathbf{A}^{\mathsf{T}})$, vilket i sin tur är ekvivalent med att $\mathbf{b}-\mathbf{A}\mathbf{x} \in \mathcal{R}(\mathbf{A})^{\perp}$. Normalekvationerna säger alltså att "felvektorn" $\mathbf{b}-\mathbf{A}\mathbf{x}$ ska vara *ortogonal* mot $\mathcal{R}(\mathbf{A})$.

Den optimala punkten $\hat{\mathbf{y}}$ $(=\mathbf{A}\hat{\mathbf{x}})$ till problemet (1.8) bestäms alltså av att $\hat{\mathbf{y}} \in \mathcal{R}(\mathbf{A})$ och $\mathbf{b}-\hat{\mathbf{y}} \in \mathcal{R}(\mathbf{A})^{\perp}$, dvs av att $\mathbf{b}-\hat{\mathbf{y}}$ är ortogonal mot $\mathcal{R}(\mathbf{A})$. Som vi såg ovan är punkten $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}$ unik, även om inte $\hat{\mathbf{x}}$ är det.

1.5 Specialfallet att kolonnerna i A är linjärt oberoende

Om kolonnerna i matrisen $\mathbf{A}$ är linjärt oberoende så är matrisen $\mathbf{A}^{\mathsf{T}}\mathbf{A}$ positivt definit och därmed icke-singulär. Då har normalekvationerna (1.7) en *unik* lösning $\hat{\mathbf{x}}$.

1.6 Minsta-normlösningen vid linjärt beroende kolonner i A

Om $\mathbf{A}$ har linjärt beroende kolonner så är $\mathbf{A}^{\mathsf{T}}\mathbf{A}$ positivt semidefinit men *inte* positivt definit, och då finns det oändligt många lösningar $\mathbf{x}$ till normalekvationerna (1.7). Ett vanligt sätt att välja ut *en* av alla dessa lösningar är att ta den "kortaste", dvs den med "minst norm".

Låt $\bar{\mathbf{x}} \in I\!R^n$ vara *en* lösning till (1.7). Enligt avsnitt 1.3 ges då samtliga lösningar till normalekvationerna av lösningarna $\mathbf{x}$ till ekvationssystemet $\mathbf{A}\mathbf{x} = \mathbf{A}\bar{\mathbf{x}}$. Därmed kan problemet att bestämma minsta-normlösningen till normalekvationerna skrivas på formen

$$\begin{array}{rl} \text{minimera} & \frac{1}{2}\|\,\mathbf{x}\,\|^2 \\ \text{då} & \mathbf{A}\mathbf{x} = \mathbf{A}\bar{\mathbf{x}}. \end{array} \tag{1.9}$$

Detta är ett QP-problem med likhetsbivillkor, med $\mathbf{H} = \mathbf{I}$, $\mathbf{c} = \mathbf{0}$, $\mathbf{A} = \mathbf{A}$, $\mathbf{b} = \mathbf{A}\bar{\mathbf{x}}$. Enligt ett tidigare avsnitt är $\hat{\mathbf{x}} \in I\!R^n$ en optimal lösning till detta problem (1.9) om och endast om $\hat{\mathbf{x}}$ utgör "$\mathbf{x}$-delen" av en lösning till följande linjära ekvationssystem i $\mathbf{x}$ och $\mathbf{u}$:

$$\begin{array}{ccccc} \mathbf{I}\,\mathbf{x} & - & \mathbf{A}^{\mathsf{T}}\mathbf{u} & = & \mathbf{0} \\ \mathbf{A}\mathbf{x} & & & = & \mathbf{A}\bar{\mathbf{x}} \end{array} \tag{1.10}$$

De övre ekvationerna ger att $\mathbf{x} = \mathbf{A}^{\mathsf{T}}\mathbf{u}$, som insatt i de undre ekvationerna ger att

$$\mathbf{A}\mathbf{A}^{\mathsf{T}}\mathbf{u} = \mathbf{A}\bar{\mathbf{x}}. \tag{1.11}$$

Eftersom $\mathbf{A}\bar{\mathbf{x}} \in \mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{A}\mathbf{A}^{\mathsf{T}})$, så har systemet (1.11) alltid minst en lösning $\hat{\mathbf{u}}$. Motsvarande $\hat{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\hat{\mathbf{u}}$ är då en optimal lösning till problemet (1.9).

Även om det i vissa fall finns oändligt många olika lösningar $\mathbf{u}$ till systemet (1.11) så gäller att vektorn $\mathbf{A}^{\mathsf{T}}\mathbf{u}$ är densamma för samtliga dessa lösningar. Om nämligen både $\mathbf{A}\mathbf{A}^{\mathsf{T}}\hat{\mathbf{u}} = \mathbf{A}\bar{\mathbf{x}}$ och $\mathbf{A}\mathbf{A}^{\mathsf{T}}\bar{\mathbf{u}} = \mathbf{A}\bar{\mathbf{x}}$ så är $\mathbf{A}\mathbf{A}^{\mathsf{T}}(\hat{\mathbf{u}}-\bar{\mathbf{u}}) = \mathbf{0}$, dvs $\hat{\mathbf{u}}-\bar{\mathbf{u}} \in \mathcal{N}(\mathbf{A}\mathbf{A}^{\mathsf{T}})$. Men eftersom $\mathcal{N}(\mathbf{A}\mathbf{A}^{\mathsf{T}}) = \mathcal{N}(\mathbf{A}^{\mathsf{T}})$ så erhålls att $\mathbf{A}^{\mathsf{T}}(\hat{\mathbf{u}}-\bar{\mathbf{u}}) = \mathbf{0}$, dvs $\mathbf{A}^{\mathsf{T}}\hat{\mathbf{u}} = \mathbf{A}^{\mathsf{T}}\bar{\mathbf{u}}$.

Det betyder att $\hat{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\hat{\mathbf{u}}$ är den *unika* optimala lösningen till problemet (1.9), även om lösningen $\hat{\mathbf{u}}$ till systemet (1.11) inte är unik!

1.7 Pseudoinversen till en matris

Antag att man ska bestämma MN-lösningen till ett MK-problem enligt avsnitt 1.6 ovan, och att matrisen $\mathbf{A}$ har *singulärvärdesfaktoriserats* (se Gula häftet), dvs faktoriserats på formen

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{S}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{\mathsf{T}} \\ \mathbf{V}_2^{\mathsf{T}} \end{bmatrix} = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^{\mathsf{T}}, \tag{1.12}$$

där såväl $n{\times}r$–matrisen $\mathbf{U}_1$ som $m{\times}r$–matrisen $\mathbf{V}_1$ har ortonormala kolonner, dvs $\mathbf{U}_1^{\mathsf{T}}\mathbf{U}_1 = \mathbf{I}$ och $\mathbf{V}_1^{\mathsf{T}}\mathbf{V}_1 = \mathbf{I}$, medan $r \times r$–matrisen $\mathbf{S}_1$ är diagonal med strikt positiva diagonalelement.

Eftersom $\mathbf{A}^{\mathsf{T}} = \mathbf{V}_1\mathbf{S}_1\mathbf{U}_1^{\mathsf{T}}$, så blir

$$\mathbf{A}^{\mathsf{T}}\mathbf{A} = \mathbf{V}_1\mathbf{S}_1^2\,\mathbf{V}_1^{\mathsf{T}} \quad \text{och} \quad \mathbf{A}\mathbf{A}^{\mathsf{T}} = \mathbf{U}_1\mathbf{S}_1^2\,\mathbf{U}_1^{\mathsf{T}}. \tag{1.13}$$

Normalekvationerna (1.7) får då utseendet

$$\mathbf{V}_1\mathbf{S}_1^2\,\mathbf{V}_1^{\mathsf{T}}\mathbf{x} = \mathbf{V}_1\mathbf{S}_1\mathbf{U}_1^{\mathsf{T}}\mathbf{b}, \tag{1.14}$$

som är ekvivalent med ekvationssystemet

$$\mathbf{V}_1^{\mathsf{T}}\mathbf{x} = \mathbf{S}_1^{-1}\mathbf{U}_1^{\mathsf{T}}\mathbf{b}. \tag{1.15}$$

Om $\bar{\mathbf{x}}$ är *en* lösning till (1.15) så blir ekvationssystemet (1.11) nu

$$\mathbf{U}_1\mathbf{S}_1^2\,\mathbf{U}_1^{\mathsf{T}}\mathbf{u} = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^{\mathsf{T}}\bar{\mathbf{x}}, \tag{1.16}$$

som är ekvivalent med ekvationssystemet

$$\mathbf{S}_1\mathbf{U}_1^{\mathsf{T}}\mathbf{u} = \mathbf{V}_1^{\mathsf{T}}\bar{\mathbf{x}}. \tag{1.17}$$

Om $\hat{\mathbf{u}}$ är *en* lösning till (1.17) så ges MN-lösningen till MK-problemet således av

$$\hat{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\hat{\mathbf{u}} = \mathbf{V}_1\mathbf{S}_1\mathbf{U}_1^{\mathsf{T}}\hat{\mathbf{u}} = \mathbf{V}_1\mathbf{V}_1^{\mathsf{T}}\bar{\mathbf{x}} = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^{\mathsf{T}}\mathbf{b} = \mathbf{A}^{+}\mathbf{b}, \tag{1.18}$$

där matrisen $\mathbf{A}^{+} = \mathbf{V}_1\mathbf{S}_1^{-1}\mathbf{U}_1^{\mathsf{T}}$ kallas för *pseudoinversen* till matrisen $\mathbf{A} = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^{\mathsf{T}}$.

1.8 Sammanfattning av resultaten i detta kapitel

MK-problemet består i att minimera $\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \frac{1}{2}(\mathbf{A}\mathbf{x} - \mathbf{b})^{\mathsf{T}}(\mathbf{A}\mathbf{x} - \mathbf{b})$.
Vektorn $\hat{\mathbf{x}}$ är en optimal lösning till MK-problemet om och endast om $\hat{\mathbf{x}}$ är en lösning till *normalekvationerna* $\mathbf{A}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathsf{T}}\mathbf{b}$ som alltid har minst en lösning.
Om kolonnerna i $\mathbf{A}$ är linjärt oberoende så har normalekvationerna en *unik* lösning $\hat{\mathbf{x}}$.
Om kolonnerna i $\mathbf{A}$ är linjärt beroende så har normalekvationerna oändligt många lösningar.
Bland alla dessa lösningar finns det en unik vektor $\hat{\mathbf{x}}$ med minst norm (dvs kortast längd).
Denna minsta-normlösning ges av $\hat{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\hat{\mathbf{u}}$, där $\hat{\mathbf{u}}$ är en lösning till systemet $\mathbf{A}\mathbf{A}^{\mathsf{T}}\mathbf{u} = \mathbf{A}\bar{\mathbf{x}}$.
Här är $\bar{\mathbf{x}}$ en godtycklig lösning till normalekvationerna.