Föreläsning 7: Kvadratisk optimering

1. Kvadratisk optimering utan bivillkor
2. Positivt definita och semidefinita matriser
3. $\mathbf{LDL}^\mathsf{T}$ faktorisering
4. Kvadratisk optimering under linjära bivillkor
5. Minsta kvadratproblem

Kvadratisk optimering utan bivillkor

$$\text{minimera}\ \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{H}\mathbf{x} + \mathbf{c}^{\mathsf{T}}\mathbf{x} + c_0$$

där $\mathbf{H} = \mathbf{H}^{\mathsf{T}} \in \mathbf{R}^{n\times n}$, $\mathbf{c} \in \mathbf{R}^{n}$, $c_0 \in \mathbf{R}$.

- Vanligt förekommande i tillämpningar, t.ex.
  - linjär regression (modelanpassning)
  - minimering av fysikaliskt betingade målfunktioner som t.ex. minimering av energi, varians, e.t.c.
  - Kvadratisk approximation av olinjära optimeringsproblem

Kvadratiska termen

Låt

$$f(x) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$$

där $\mathbf{H} = \mathbf{H}^\mathsf{T} \in \mathbf{R}^{n\times n}$, $\mathbf{c} \in \mathbf{R}^n$, $c_0 \in \mathbf{R}$.

Vi kan anta att matrisen $\mathbf{H}$ är symmetrisk.

Om $\mathbf{H}$ inte är symmetrisk så gäller att

$$\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \frac{1}{2}(\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x})^\mathsf{T} = \frac{1}{2}\mathbf{x}^\mathsf{T}(\mathbf{H} + \mathbf{H}^\mathsf{T})\mathbf{x},$$

d.v.s. $\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} = \mathbf{x}^\mathsf{T}\tilde{\mathbf{H}}\mathbf{x}$ där $\tilde{\mathbf{H}} = \frac{1}{2}(\mathbf{H} + \mathbf{H}^\mathsf{T})$ är en symmetrisk matris.

Positivt definita och positivt semidefinita matriser

Definition 1 (Positivt semidefinit). *En symmetrisk matris* $\mathbf{H} = \mathbf{H}^\mathsf{T} \in \mathbf{R}^{n \times n}$ *säges vara positivt semidefinit om*

$$\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} \geq 0 \quad \textit{för alla } \mathbf{x} \in \mathbf{R}^n$$

Definition 2 (Positivt definit). *En symmetrisk matris* $\mathbf{H} = \mathbf{H}^\mathsf{T} \in \mathbf{R}^{n \times n}$ *säges vara positivt definit om*

$$\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} > 0 \quad \textit{för alla } \mathbf{x} \in \mathbf{R}^n - \{\mathbf{0}\}$$

Metoder att avgöra om en symmetrisk matris $\mathbf{H} = \mathbf{H}^{\mathsf{T}} \in \mathbf{R}^{n\times n}$ är positivt (semi) definit:

- $\mathbf{LDL}^{\mathsf{T}}$ faktorisering (lämpligt vid handräkning).
  Kapitel 26.6 i OK (Kapitel 7 i gula häftet).
- Uttnyttja matrisens spektralfaktorisering $\mathbf{H} = \mathbf{Q}\Lambda\mathbf{Q}^{\mathsf{T}}$, där $\mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{I}$ och $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ där λ_k är matrisens egenvärden. Från detta kan man se att
  - $\mathbf{H}$ är positivt semidefinit om och endast om $\lambda_k \geq 0$, $k = 1, \dots, n$
  - $\mathbf{H}$ är positivt definit om och endast om $\lambda_k > 0$, $k = 1, \dots, n$

  (Du kan läsa mer om detta i kapitel 8 i gula häftet (kursivt).)

$\mathbf{LDL}^\mathsf{T}$ faktorisering

Sats 1 (26.21 i OK). *En symmetrisk matris* $\mathbf{H} = \mathbf{H}^\mathsf{T} \in \mathbf{R}^{n\times n}$ *är positivt semidefinit om och endast om det finns en faktorisering* $\mathbf{H} = \mathbf{LDL}^\mathsf{T}$ *där*

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 \\ \vdots & & & \ddots & \\ l_{n1} & l_{n2} & l_{n3} & \dots & 1 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & d_n \end{bmatrix}$$

och $d_k \geq 0$, $k = 1, \dots, n$.

$\mathbf{H}$ *är positivt definit om och endast om* $d_k > 0$, $k = 1, \dots, n$, *i* $\mathbf{LDL}^\mathsf{T}$ *faktoriseringen.*

Bevis: *Se kapitel 26.9 i OK (kapitel 7 i gula häftet).*

Man erhåller $\mathbf{LDL}^{\mathsf{T}}$ faktoriseringen genom endera av följande metoder:

- Elementära rad och kolonnoperationer.
  Se kapitel 26.6-7 och 26.9-10 i OK
  (kapitel 7.6-7 och 7.9-10 i gula häftet).

- Kvadratkomplettering.
  Se kapitel 26.8 i OK (kapitel 7.8 i gula häftet).

Obegränsade Problem

Låt $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$.

Antag att $\mathbf{H}$ inte är positivt semidefinit. Då finns $\mathbf{d} \in \mathbf{R}^n$ sådant att $\mathbf{d}^\mathsf{T}\mathbf{H}\mathbf{d} < 0$. Det följer att

$$f(t\mathbf{d}) = \frac{1}{2}t^2\mathbf{d}^\mathsf{T}\mathbf{H}\mathbf{d} + t\mathbf{c}^\mathsf{T}\mathbf{d} + c_0 \to -\infty \text{ då } t \to \infty$$

Slutsats: Ett nödvändigt villkor för att $f(\mathbf{x})$ skall ha ett minimum är att $\mathbf{H}$ är positivt semidefinit.

Antag att $\mathbf{H}$ är positivt semidefinit.
Då kan vi använda faktoriseringen $\mathbf{H} = \mathbf{L}\mathbf{D}\mathbf{L}^\mathsf{T}$, vilket ger

$$
\begin{aligned}
&\min_{x} \quad \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0 \\
&= \min_{x} \quad \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{D}\mathbf{L}^\mathsf{T}\mathbf{x} + \mathbf{c}^\mathsf{T}(\mathbf{L}^\mathsf{T})^{-1}\mathbf{L}^\mathsf{T}\mathbf{x} + c_0 \\
&= \min_{y_1,\cdots,y_n} \quad \sum_{k=1}^{n} \frac{1}{2}d_k y_k^2 + \tilde{c}_k y_k + c_0 \\
&= c_0 + \sum_{k=1}^{n} \min_{y_k} \quad \frac{1}{2}d_k y_k^2 + \tilde{c}_k y_k
\end{aligned}
$$

där vi introducerat de nya koordinaterna $y = \mathbf{L}^\mathsf{T}\mathbf{x}$ och $\tilde{c} = \mathbf{L}^{-1}\mathbf{c}$.

Vårt optimeringsproblem har separaterat till n oberoende skalära kvadratiska optimeringsproblem vilka var för sig är enkla att lösa.

Skalära Kvadratiska optimeringsproblem utan bivillkor

Det skalära kvadratiska optimeringsproblemet

$$\text{minimera}_{x \in \mathbf{R}} \quad \frac{1}{2}hx^2 + cx + c_0$$

har en ändlig lösning om, och endast om,

- $h = 0$ och $c = 0$, eller
- $h > 0$.

I det första fallet så är optimum inte unikt.

I det andra fallet så är $\frac{1}{2}hx^2 + cx + c_0 = \frac{1}{2}h(x + c/h)^2 + c_0 - c^2/(2h)$, och optimum antas för $x = -c/h$.

Separat lösning av minimeringsproblemen ger att $\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 & \dots & \hat{y}_n \end{bmatrix}^\mathsf{T}$ är en minpunkt om

$$\begin{array}{l} (i)' \; d_k \geq 0, \;\; k = 1, \dots, n \\ (ii)' \; d_k \hat{y}_k = -\tilde{c}_k \end{array} \quad \Leftrightarrow \quad \begin{array}{l} (i) \; \mathbf{H} \text{ är positivt semidefinit} \\ (ii) \; \mathbf{D}\mathbf{L}^\mathsf{T}\hat{\mathbf{x}} = -\mathbf{L}^{-1}\mathbf{c} \end{array}$$

Villkor (ii) kan ersättas med $\mathbf{L}\mathbf{D}\mathbf{L}^\mathsf{T}\hat{\mathbf{x}} = -\mathbf{c}$ d.v.s. $\mathbf{H}\hat{\mathbf{x}} = -\mathbf{c}$.

Vidare gäller att $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$ är nedåt obegränsad om

$$\begin{array}{l} (i)' \;\; \text{något } d_k < 0, \text{ eller} \\ (ii)' \;\; \text{någon av ekvationerna} \\ \qquad d_k \hat{y}_k = -\tilde{c}_k \text{ saknar lösning} \\ \qquad \text{d.v.s. } d_k = 0 \text{ och } \tilde{c}_k \neq 0 \end{array} \quad \Leftrightarrow \quad \begin{array}{l} (i) \; \mathbf{H} \text{ är ej positivt semidefinit} \\ (ii) \text{ ekvationen } \mathbf{H}\hat{\mathbf{x}} = -\mathbf{c} \\ \qquad \text{saknar lösning} \end{array}$$

Ovanstående resonemang kan sammanfattas i följande sats

Sats 2. *Låt* $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$. *Då är* $\hat{\mathbf{x}}$ *en minpunkt om*

$$(i)\ \mathbf{H} \text{ är positivt semidefinit}$$

$$(ii)\ \mathbf{H}\hat{\mathbf{x}} = -\mathbf{c}$$

Om $\mathbf{H}$ *är positivt definit så är* $\hat{\mathbf{x}} = -\mathbf{H}^{-1}\mathbf{c}$.

Om $\mathbf{H}$ *ej är positivt semidefinit eller* $\mathbf{H}\mathbf{x} = -\mathbf{c}$ *saknar lösning så är* f *nedåt obegränsad.*

Kommentar 1. *Villkoret att* $\mathbf{H}\mathbf{x} = -\mathbf{c}$ *är lösbar är ekvivalent med att* $\mathbf{c} \in \mathcal{R}(\mathbf{H})$. *Således har* f *en minpunkt om och endast om* $\mathbf{H}$ *är positivt semidefinit och* $\mathbf{c} \in \mathcal{R}(\mathbf{H})$.

Kvadratisk optimering med linjära bivillkor

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{H}\mathbf{x} + \mathbf{c}^{\mathsf{T}}\mathbf{x} + c_0 \\ \text{då} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{1}$$

där $\mathbf{H} = \mathbf{H}^{\mathsf{T}} \in \mathbf{R}^{n\times n}$, $\mathbf{A} \in \mathbf{R}^{m\times n}$, $\mathbf{c} \in \mathbf{R}^{n}$, and $\mathbf{b} \in \mathbf{R}^{m}$, and $c_0 \in \mathbf{R}$.

- Vi antar att $\mathbf{b} \in \mathcal{R}(\mathbf{A})$ och $n > m$, d.v.s. $\mathcal{N}(\mathbf{A}) \neq \{\mathbf{0}\}$.
- Antag att $\mathcal{N}(A) = \text{span}\{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$ och definiera nollrumsmatrisen $\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 & \ldots & \mathbf{z}_k \end{bmatrix}$.
- Om $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$ så gäller att en godtycklig lösning till det linjära bivillkoret har formen $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Z}\mathbf{v}$, för något $\mathbf{v} \in \mathbf{R}^{k}$.
  Detta följer eftersom

$$\mathbf{A}(\bar{\mathbf{x}} + \mathbf{Z}\mathbf{v}) = \mathbf{A}\bar{\mathbf{x}} + \mathbf{A}\mathbf{Z}\mathbf{v} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

Med $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Z}\mathbf{v}$ insatt i $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$ får vi

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{v}^\mathsf{T}\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}\mathbf{v} + (\mathbf{Z}^\mathsf{T}(\mathbf{H}\bar{\mathbf{x}} + \mathbf{c}))^\mathsf{T}\mathbf{v} + f(\bar{\mathbf{x}}).$$

Minimeringsproblemet (1) är således ekvivalent med

$$\min_{\mathbf{v}} \quad \frac{1}{2}\mathbf{v}^\mathsf{T}\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}\mathbf{v} + (\mathbf{Z}^\mathsf{T}(\mathbf{H}\bar{\mathbf{x}} + \mathbf{c}))^\mathsf{T}\mathbf{v} + f(\bar{\mathbf{x}}).$$

Vi kan tillämpa Sats 2 vilket ger följande resultat.

Sats 3. $\hat{\mathbf{x}}$ *är en optimal lösning till (1) om*

(i) $\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}$ *är positivt semidefinit.*

(ii) $\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{Z}\hat{\mathbf{v}}$ *där* $\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}\hat{\mathbf{v}} = -\mathbf{Z}^\mathsf{T}(\mathbf{H}\bar{\mathbf{x}} + \mathbf{c})$

Kommentar 2. *Det andra villkoret i satsen är ekvivalent med att det finns* $\hat{\mathbf{u}} \in \mathbf{R}^m$ *så att*

$$\begin{bmatrix} \mathbf{H} & -\mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} -\mathbf{c} \\ \mathbf{b} \end{bmatrix}$$

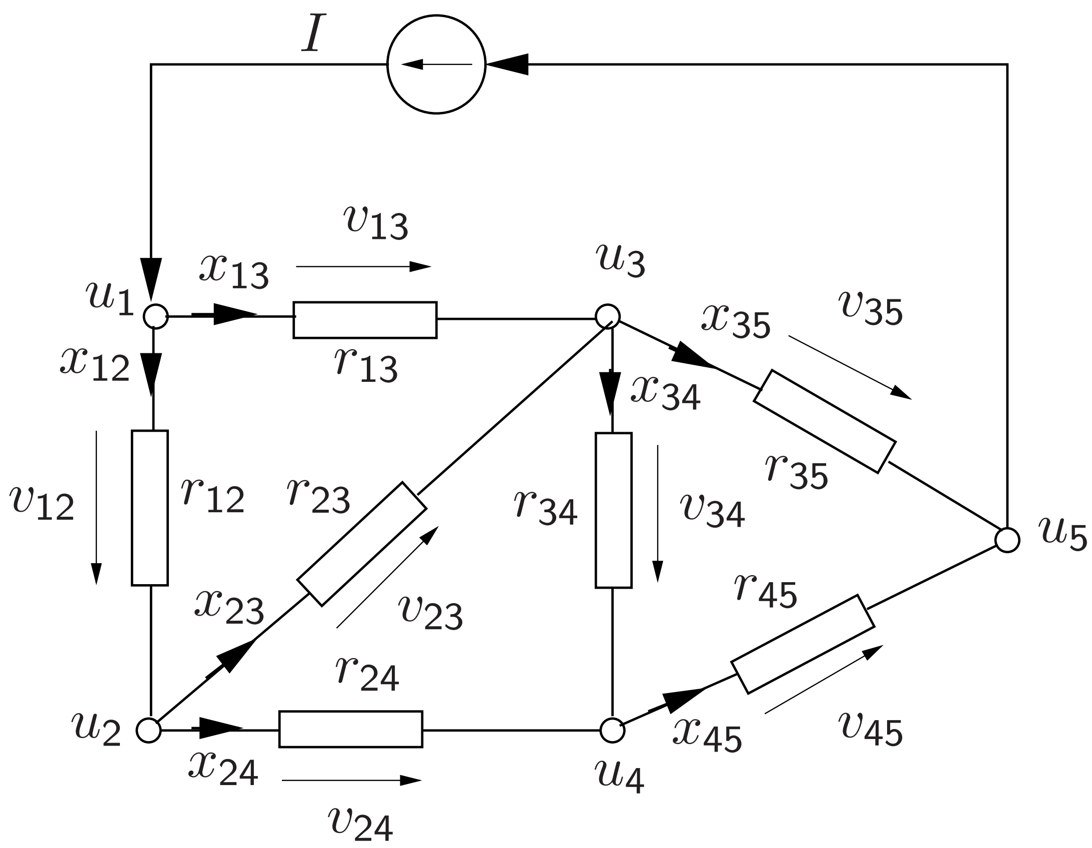
Bevis: *Det andra villkoret i satsen kan skrivas*

$$\mathbf{Z}^\top(\mathbf{H}(\bar{\mathbf{x}} + \mathbf{Z}\hat{\mathbf{v}}) + \mathbf{c}) = \mathbf{0}$$

Eftersom $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{Z})$ *så är detta ekvivalent med*

$$\begin{aligned} & \mathbf{H}(\underbrace{\bar{\mathbf{x}} + \mathbf{Z}\hat{\mathbf{v}}}_{\hat{\mathbf{x}}}) + \mathbf{c} \in \mathcal{N}(\mathbf{A})^\perp = \mathcal{R}(\mathbf{A}^\top) \\ \Leftrightarrow \quad & \mathbf{H}\hat{\mathbf{x}} + \mathbf{c} = \mathbf{A}^\top\hat{\mathbf{u}}, \text{ för något } \hat{\mathbf{u}} \in \mathbf{R}^m \end{aligned}$$

Exempel: Analys av ett resistansnätverk



- Vi skickar en ström I genom nätverket från nod 1 till nod 5.
- Vad blir lösningen till kretsekvationen?

Strömmen x_{ij} går från i till nod j om $x_{ij} > 0$. Kirschoff's strömlag ger

$$\begin{aligned} x_{12} + x_{13} &= I \\ x_{12} - x_{23} - x_{24} &= 0 \\ x_{13} + x_{23} - x_{34} - x_{35} &= 0 \\ x_{24} + x_{34} - x_{45} &= 0 \\ x_{35} + x_{45} &= -I \end{aligned}$$

Detta kan skrivas $\mathbf{A}\mathbf{x} = \mathbf{b}$ där $\mathbf{x} = \begin{bmatrix} x_{12} & x_{13} & x_{23} & x_{24} & x_{34} & x_{35} & x_{45} \end{bmatrix}^{\mathsf{T}}$,

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \\ -I \end{bmatrix}$$

- Nodpotentialerna betecknas u_j, $j = 1, \ldots, 5$
- Spänningsfallet över motståndet r_{ij} betecknas v_{ij} och ges av ekvationerna

$$
\begin{aligned}
v_{ij} &= u_i - u_j \\
v_{ij} &= r_{ij} x_{ij} \quad \text{(Ohms lag)}
\end{aligned}
$$

Detta kan också skrivas

$$
\begin{aligned}
\mathbf{v} &= \mathbf{A}^\mathsf{T} \mathbf{u} \\
\mathbf{v} &= \mathbf{D} \mathbf{x}
\end{aligned}
$$

där

$$\mathbf{u} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 \end{bmatrix}^\mathsf{T} \quad \mathbf{v} = \begin{bmatrix} v_{12} & v_{13} & v_{23} & v_{24} & v_{34} & v_{35} & v_{45} \end{bmatrix}^\mathsf{T}$$

$$\mathbf{D} = \mathrm{diag}(r_{12}, r_{13}, r_{23}, r_{24}, r_{34}, r_{35}, r_{45})$$

Elkretsens ekvationer:

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{v} &= \mathbf{A}^\mathsf{T}\mathbf{u} \\ \mathbf{v} &= \mathbf{D}\mathbf{x} \end{aligned} \quad \Leftrightarrow \quad \begin{bmatrix} \mathbf{D} & -\mathbf{A}^\mathsf{T} \\ \mathbf{A} & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix}$$

Eftersom $\mathbf{D}$ är positivt definit $(r_{ij} > 0)$ så är detta enligt kommentar 2 ekvivalent med optimalitetsvillkoren för följande optimeringsproblem

$$\begin{aligned} \text{minimera} \quad & \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{D}\mathbf{x} \\ \text{då} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

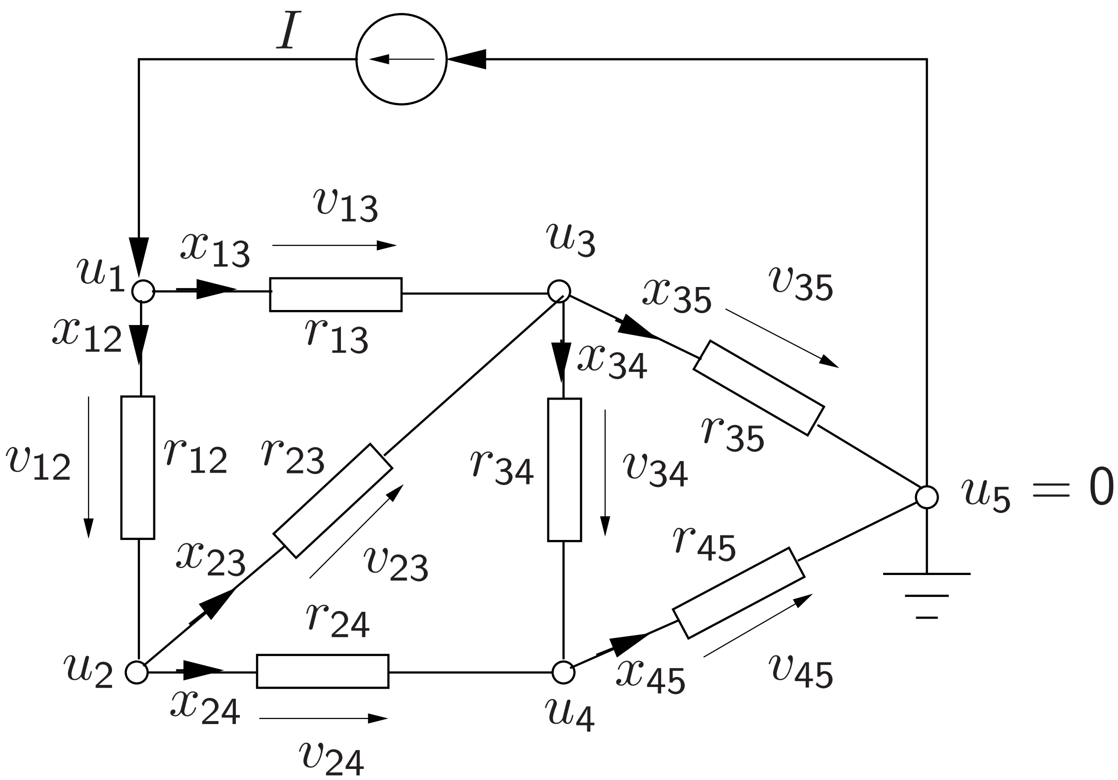
Kommentar

I avsnitten om nätverksoptimering såg vi att matrisen $\mathcal{N}(\mathbf{A}) \neq \mathbf{0}$ (då kallades matrisen $\tilde{\mathbf{A}}$) och att nollrummet svarade mot slingor i elkretsen.

Elektricitetsteorin (Ohms lag + Kirchoffs lag) väljer ut den ström som svarar mot minst effektförlust. Detta följer eftersom målfunktionen kan skrivas

$$\frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{D}\mathbf{x} = \frac{1}{2}\sum_{ij} r_{ij}x_{ij}^2$$

Kommentar 3. *Det är i regel en fördel att jorda en av noderna. Låt oss till exempel jorda nod 5 varvid* $u_5 = 0$.



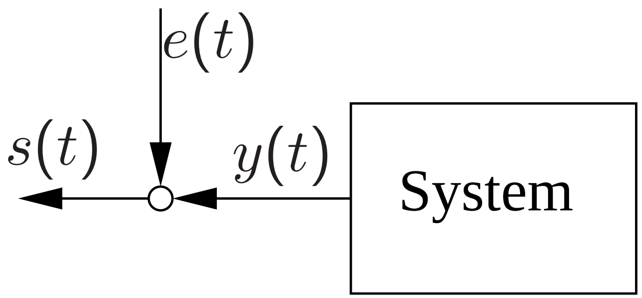
Med potentialen i nod 5 fixerad till $u_5 = 0$ utgår sista kolonnen i spänningsekvationen $\mathbf{v} = \mathbf{A}^\mathsf{T}\mathbf{u}$. Spänningsekvationen och strömbalansen kan ersättas med $\mathbf{v} = \bar{\mathbf{A}}^\mathsf{T}\bar{\mathbf{u}}$ och $\bar{\mathbf{A}}\mathbf{x} = \bar{\mathbf{b}}$ där

$$\bar{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \end{bmatrix}, \quad \bar{\mathbf{b}} = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \bar{\mathbf{u}} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}$$

Fördelen med detta är att raderna i $\bar{\mathbf{A}}$ är linjärt oberoende. Detta underlättar lösningen av optimeringsproblemet.

Minsta kvadratproblem

Tillämpning: Linjär regression (modellanpassing)



Problem: Anpassa en linjär regressionsmodell till mätdata.

$$\text{Regressionsmodellen :} \quad y(t) = \sum_{j=1}^{n} \alpha_j \psi_j(t)$$

- $\psi_k(t)$, $k = 1, \ldots, n$ är regressorerna (kända funktioner)
- $\alpha_k = k = 1, \ldots, n$ är modellparametrarna (skall bestämmas)
- $e(t)$ mätbrus (ej känd)
- $s(t)$ observationer.

Idé för modellanpassning: Minimera kvadratsumman av prediktionsfelen

$$\begin{aligned}
&\text{minimera}\ \frac{1}{2}\sum_{i=1}^{m}\Big(\sum_{j=1}^{n}\alpha_j\psi_j(t_i)-s(t_i)\Big)^2\\
=&\text{minimera}\ \frac{1}{2}\sum_{i=1}^{m}\big(\psi(t_i)^{\mathsf{T}}\mathbf{x}-s(t_i)\big)^2\\
=&\text{minimera}\ \frac{1}{2}(\mathbf{A}\mathbf{x}-\mathbf{b})^{\mathsf{T}}(\mathbf{A}\mathbf{x}-\mathbf{b})
\end{aligned} \tag{2}$$

$$\psi(t)=\begin{bmatrix}\psi_1(t)\\ \vdots\\ \psi_n(t)\end{bmatrix},\quad \mathbf{x}=\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{bmatrix},\quad \mathbf{A}=\begin{bmatrix}\psi(t_1)^{\mathsf{T}}\\ \vdots\\ \psi(t_n)^{\mathsf{T}}\end{bmatrix},\quad \mathbf{b}=\begin{bmatrix}s(t_1)\\ \vdots\\ s(t_n)\end{bmatrix}$$

Minsta kvadratproblemets (MK) lösning

Minsta kvadratproblemet (2) kan ekvivalent skrivas

$$\text{minimera}\ \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$$

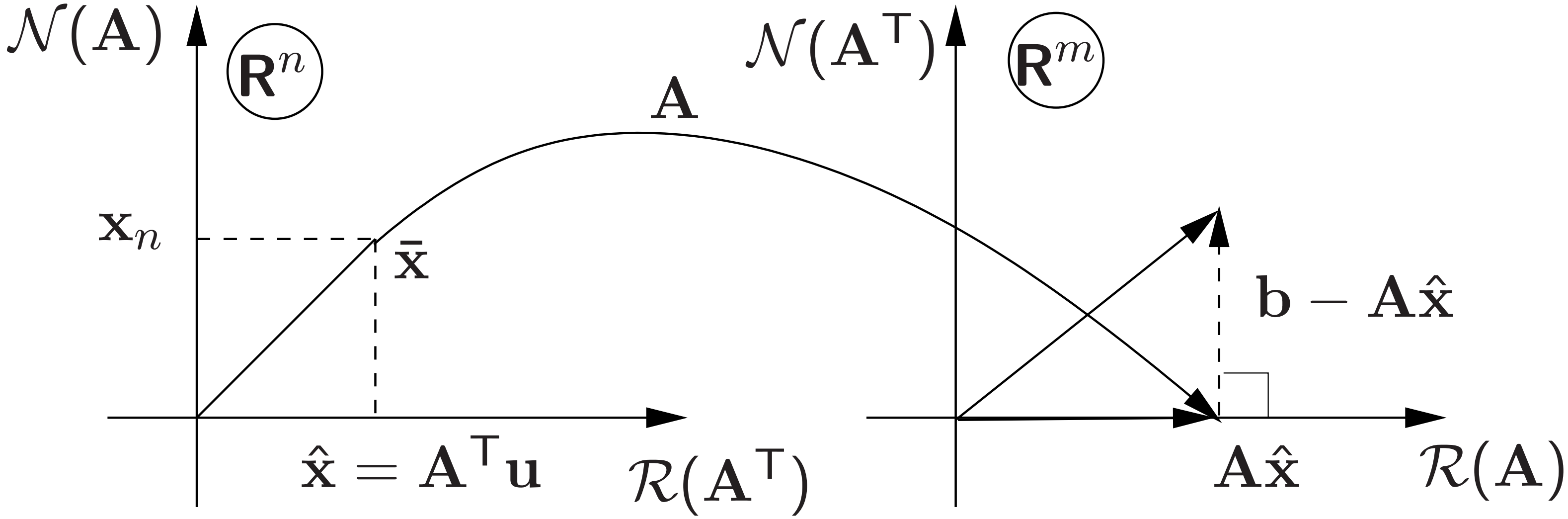
där $\mathbf{H} = \mathbf{A}^\mathsf{T}\mathbf{A}$, $\mathbf{c} = -\mathbf{A}^\mathsf{T}\mathbf{b}$, $c_0 = \frac{1}{2}\mathbf{b}^\mathsf{T}\mathbf{b}$. Vi noterar att

- $\mathbf{H} = \mathbf{A}^\mathsf{T}\mathbf{A}$ är positivt semidefinit eftersom $\mathbf{x}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 \geq 0$.
- $\mathbf{c} \in \mathcal{R}(\mathbf{H})$ eftersom $\mathbf{c} = -\mathbf{A}^\mathsf{T}\mathbf{b} \in \mathcal{R}(\mathbf{A}^\mathsf{T}) = \mathcal{R}(\mathbf{A}^\mathsf{T}\mathbf{A}) = \mathcal{R}(\mathbf{H})$.

Villkoren i Sats 2 och Kommentar (1) är uppfyllda och det följer att MK-skattningen ges av

$$\mathbf{H}\hat{\mathbf{x}} = -\mathbf{c} \quad \Leftrightarrow \quad \mathbf{A}^\mathsf{T}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^\mathsf{T}\mathbf{b} \quad \text{(Normal ekvationen)}$$

Linjär algebratolkning



- Normalekvationen kan skrivas

$$\mathbf{A}^\mathsf{T}(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{b} - \mathbf{A}\hat{\mathbf{x}} \in \mathcal{N}(\mathbf{A}^\mathsf{T})$$

  Denna ortogonalitetsegenskap har en geometrisk tolkning villken illustreras i figuren ovan.

- Linjära algebrans fundamentalsats förklarar MK lösningen geometriskt.

MK skattningens unikhet

<u>Fall 1</u> Om kolonnerna i $\mathbf{A}$ är linjärt oberoende, d.v.s. $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$, så gäller att $\mathbf{H} = \mathbf{A}^\mathsf{T}\mathbf{A}$ är positivt definit och därmed inverterbar. Normalekvationen har då den unika lösningen

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

<u>Fall 2</u> Om $\mathbf{A}$ har linjärt beroende kolonner, d.v.s. $\mathcal{N}(\mathbf{A}) \neq \{\mathbf{0}\}$ så är det naturligt att välja den lösning med kortast längd. Från figuren på föregående sida ser vi att vi då skall låta $\hat{\mathbf{x}} = \mathbf{A}^\mathsf{T}\hat{\mathbf{u}}$, där $\mathbf{A}\mathbf{A}^\mathsf{T}\hat{\mathbf{u}} = \mathbf{A}\bar{\mathbf{x}}$, där $\bar{\mathbf{x}}$ är en godtycklig lösning till MK problemet.

Vårt val av $\hat{\mathbf{x}}$ i fall 2 kan ekvivalent tolkas som lösningen till det kvadratiska optimeringsproblemet

$$\begin{array}{rl} \text{minimera} & \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{x} \\ \text{då} & \mathbf{A}\mathbf{x} = \mathbf{A}\bar{\mathbf{x}} \end{array}$$

Enligt Kommentar 2 ges lösningen till detta optimeringsproblem som lösningen till ekvationssystemet:

$$\begin{bmatrix} I & -\mathbf{A}^{\mathsf{T}} \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{A}\bar{\mathbf{x}} \end{bmatrix}$$

d.v.s. $\hat{\mathbf{x}} = \mathbf{A}^{\mathsf{T}}\hat{\mathbf{u}}$, där $\mathbf{A}\mathbf{A}^{\mathsf{T}}\hat{\mathbf{u}} = \mathbf{A}\bar{\mathbf{x}}$

Läsanvisningar

- Kvadratisk optimering utan bivillkor: Kapitel 9 och 26 i OK (Blå häftet sidan 2-10. Gula häftet sidan 36-49)
- Kvadratisk optimering under bivillkor: Kapitel 10 i OK. (Blå häftet sidan 11-20)
- Minsta-kvadratmetoden: Kapitel 11 i OK (11.7 och 11.9 kursivt) (Blå häftet sidan 21-23 (sidan 24 läses kursivt)).