KTH Matematik

Lösningar till SF1852 Optimeringslära för E, 25/10–08

1. (a) Flödesbalans i noderna ger upphov till följande optimeringsproblem

$$\begin{aligned}
\text{maximera} \quad & f \\
\text{då} \quad & x_{12} + x_{13} = f \\
& -x_{12} + x_{23} + x_{24} = 0 \\
& -x_{13} - x_{23} + x_{34} = 0 \\
& -x_{24} - x_{34} + x_{45} = 0 \\
& -x_{45} = -f \\
& 0 \leq x_{ij} \leq c_{ij}
\end{aligned}$$

För att överföra problemet på standardform inför vi slackvariablerna $z_{ij} \geq 0$ sådana att $x_{ij} + z_{ij} = c_{ij}$. Om vi låter

$$\mathbf{x} = \begin{bmatrix} x_{12} & x_{13} & x_{23} & x_{24} & x_{34} & x_{45} & z_{12} & z_{13} & z_{23} & z_{24} & z_{34} & z_{45} & f \end{bmatrix}^{\mathsf{T}}$$

$$\mathbf{c} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}^{\mathsf{T}}$$

$$\mathbf{A}_{11} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{A}_{13} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

$$\mathbf{A}_{21} = \mathbf{A}_{22} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} c_{12} \\ c_{13} \\ c_{23} \\ c_{24} \\ c_{34} \\ c_{45} \end{bmatrix}$$

och

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{b}_1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{b}_2 \end{bmatrix}$$

Vårt optimeringsproblem är då ekvivalent med följande standardproblem

$$\begin{aligned}
\text{minimera} \quad & \mathbf{c}^{\mathsf{T}}\mathbf{x} \\
\text{då} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{x} \geq \mathbf{0}
\end{aligned}$$

(b) Vi använder Gauss-Jordans metod på den givna matrisen. Elementära radoperationer ger upphov till följande sekvens av ekvivalenta matriser

$$\mathbf{A}=\begin{bmatrix}1&1&0&0&0\\-1&0&1&1&0\\0&-1&-1&0&1\\0&0&0&-1&-1\end{bmatrix}\sim\begin{bmatrix}1&1&0&0&0\\0&1&1&1&0\\0&-1&-1&0&1\\0&0&0&-1&-1\end{bmatrix}\sim\begin{bmatrix}1&1&0&0&0\\0&1&1&1&0\\0&0&0&1&1\\0&0&0&-1&-1\end{bmatrix}$$
$$\sim\begin{bmatrix}1&1&0&0&0\\0&1&1&1&0\\0&0&0&1&1\\0&0&0&0&0\end{bmatrix}\sim\begin{bmatrix}1&0&-1&-1&0\\0&1&1&1&0\\0&0&0&1&1\\0&0&0&0&0\end{bmatrix}\sim\begin{bmatrix}1&0&-1&0&1\\0&1&1&0&-1\\0&0&0&1&1\\0&0&0&0&0\end{bmatrix}=\begin{bmatrix}\mathbf{T}\\\mathbf{0}\end{bmatrix}$$

Nu är $\mathbf{A}$ överförd till trappstegsform med tre trappstegsettor.

En bas till $\mathcal{R}(\mathbf{A})$ erhålls genom att som basvektorer välja de kolonner i $\mathbf{A}$ som svarar mot trappstegsettor i $\mathbf{T}$, dvs kolonnerna nr 1, 2 och 4 i $\mathbf{A}$. Vi får alltså

$$\mathcal{R}(\mathbf{A})=\operatorname{span}\left\{\begin{bmatrix}1\\-1\\0\\0\end{bmatrix},\begin{bmatrix}1\\0\\-1\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\\-1\end{bmatrix}\right\}$$

En bas till $\mathcal{N}(\mathbf{A})$ kan bestämmas enligt följande. Sätt $x_3=1$ och $x_5=0$ och bestäm sedan x_1, x_2, och x_4 (variablerna svarande mot trappstegsettor) så att $\mathbf{Tx}=\mathbf{0}$. Detta ger en nollrums vektor. En andra nollrumsvektor erhålls om vi går tillväga på samma sätt med $x_3=0$ och $x_5=1$. Detta ger att

$$\mathcal{N}(\mathbf{A})=\operatorname{span}\left\{\begin{bmatrix}1\\-1\\1\\0\\0\end{bmatrix},\begin{bmatrix}-1\\1\\0\\-1\\1\end{bmatrix}\right\}.$$

2. **(a)** Vi använder nollrumsmetoden. Gauss-Jordans metod på bivillkoret ger att

$$\begin{bmatrix}1&1&1\\1&1&0\end{bmatrix}\mathbf{x}=\begin{bmatrix}2\\1\end{bmatrix}\quad\sim\quad\begin{bmatrix}1&1&1\\0&0&-1\end{bmatrix}\mathbf{x}=\begin{bmatrix}2\\-1\end{bmatrix}$$

vilket ger följande lösning och nollrumsmatris

$$\bar{\mathbf{x}}=\begin{bmatrix}1\\0\\1\end{bmatrix},\qquad\mathbf{Z}=\begin{bmatrix}1\\-1\\0\end{bmatrix}$$

Således uppfyller $\mathbf{x}=\bar{\mathbf{x}}+\mathbf{Z}v$ bivillkoret. Variabelbytet $\mathbf{x}=\bar{\mathbf{x}}+\mathbf{Z}v$ ger upphov till ett ekvivalent problem

$$\begin{array}{rl}\text{minimera} & \frac{1}{2}(x_1-x_2-x_3-1)^2 \quad = \quad \text{minimera}\,(2v-1)^2\\ \text{då} & x_1+x_2+x_3=2\\ & x_1+x_2=1\end{array}$$

Vi får optimalvärdet $\hat{v} = \frac{1}{2}$ varvid

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{Z}\hat{v} = \frac{1}{2}\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$$

(**b**) Vi ser att

$$\bar{\mathbf{x}} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{Z} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

är lösning till bivilkoret respektive nollrumsmatris för $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$. Således uppfyller $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Z}\mathbf{v}$ bivillkoret. Variabelbytet $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Z}v$ ger upphov till ett ekvivalent problem

$$\begin{array}{rlcl} \text{minimera} & \frac{1}{2}(x_1 - x_2 - x_3 - 1)^2 & = & \text{minimera}\,(2v_1 - 1)^2 \\ \text{då} & x_1 + x_2 + x_3 = 2 & & \end{array}$$

Vi får optimalvärdet $\hat{v}_1 = \frac{1}{2}$ samt $\hat{v}_2 \in R$ godtycklig. Detta ger lösningen

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + Z\hat{\mathbf{v}} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \frac{1}{2} \\ \hat{v}_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ -\frac{1}{2} + \hat{v}_2 \\ 1 - \hat{v}_2 \end{bmatrix}$$

3. (**a**) Gradienten är

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 4(x_1^2 - 1)x_1 + 2x_2 & 2x_1 + x_2 \end{bmatrix}$$

Vi ser att $\nabla f(\mathbf{x}) = \mathbf{0}^\mathsf{T}$ då $(x_1, x_2) = (0, 0)$, $(x_1, x_2) = (\sqrt{2}, -2\sqrt{2})$ och $(x_1, x_2) = (-\sqrt{2}, 2\sqrt{2})$. Hessianen för f är

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 12x_1^2 - 4 & 2 \\ 2 & 1 \end{bmatrix}.$$

Vi får

$$\begin{aligned} \nabla^2 f((0,0)) &= \begin{bmatrix} -4 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{indefinit} \\ \nabla^2 f((\sqrt{2}, -2\sqrt{2})) &= \begin{bmatrix} 20 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{positivt definit} \\ \nabla^2 f((-\sqrt{2}, 2\sqrt{2})) &= \begin{bmatrix} 20 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{positivt definit} \end{aligned}$$

Således är punkterna $(x_1, x_2) = (\sqrt{2}, -2\sqrt{2})$ och $(x_1, x_2) = (-\sqrt{2}, 2\sqrt{2})$ minpunkter till f.

(**b**) Optimeringsproblemet kan skrivas

$$\begin{array}{rl} \text{minimera} & f(\mathbf{x}) \\ \text{då} & g_1(\mathbf{x}) \leq 0 \\ & g_2(\mathbf{x}) \leq 0 \end{array}$$

där

$$f(\mathbf{x}) = x_1^3 - x_1^2 x_2, \quad g_1(\mathbf{x}) = x_1^4 + 2x_2^4 - 4, \quad g_2(\mathbf{x}) = 5x_1^4 + x_2^4 - 6$$

Gradienterna är

$$\begin{aligned} \nabla f(\mathbf{x}) &= \begin{bmatrix} 3x_1^2 - 2x_1 x_2 & -x_1^2 \end{bmatrix} \\ \nabla g_1(\mathbf{x}) &= \begin{bmatrix} 4x_1^3 & 8x_2^3 \end{bmatrix} \\ \nabla g_2(\mathbf{x}) &= \begin{bmatrix} 20x_1^3 & 4x_2^3 \end{bmatrix} \end{aligned}$$

Vi testar härnäst hurvida de föreslagna punkterna uppfyller KKT-villkoren

(i) Då $(x_1, x_2) = (1, -1)$ får vi

$$\begin{aligned} \nabla f((1,-1)) &= \begin{bmatrix} 5 & -1 \end{bmatrix} \\ \nabla g_1((1,-1)) &= \begin{bmatrix} 4 & -8 \end{bmatrix} \\ \nabla g_2((1,-1)) &= \begin{bmatrix} 20 & -4 \end{bmatrix} \end{aligned}$$

Eftersom $\nabla g_1((1,-1))$ och $\nabla g_2((1,-1))$ är linjärt oberoende och nollskilda så är regularitetskravet uppfyllt och KKT ger nödvändiga optimalitetsvillkor (notera att f inte är konvex). KKT-villkoren är

$$\begin{array}{ll} \text{KKT1} & \begin{bmatrix} 5 & -1 \end{bmatrix} + \begin{bmatrix} 4 & -8 \end{bmatrix} y_1 + \begin{bmatrix} 20 & -4 \end{bmatrix} y_2 = 0 \\ \text{KKT2} & \begin{array}{l} g_1((1,-1)) = 1 + 2 - 4 < 0 \quad \text{ok} \\ g_2((1,-1)) = 5 + 1 - 6 = 0 \quad \text{ok} \end{array} \\ \text{KKT3} & \begin{array}{l} y_1 \geq 0 \\ y_2 \geq 0 \end{array} \\ \text{KKT4} & \begin{array}{l} y_1(1 + 2 - 4) = 0 \\ y_2(5 + 1 - 6) = 0 \end{array} \end{array}$$

Från KKT4 ser vi att $y_1 = 0$. Om vi sedan löser ekvationssystemet i KKT1 med avseende på y_2 får vi $(y_1, y_2) = (0, -\frac{1}{4})$. Detta bryter mot KKT3 så punkten $(1, -1)$ uppfyller inte KKT villkoren.

(ii) Då $(x_1, x_2) = (-1, 1)$ får vi

$$\begin{aligned} \nabla f((-1,1)) &= \begin{bmatrix} 5 & -1 \end{bmatrix} \\ \nabla g_1((-1,1)) &= \begin{bmatrix} -4 & 8 \end{bmatrix} \\ \nabla g_2((-1,1)) &= \begin{bmatrix} -20 & 4 \end{bmatrix} \end{aligned}$$

Eftersom $\nabla g_1((-1,1))$ och $\nabla g_2((-1,1))$ är linjärt oberoende och nollskilda så är regularitetskravet uppfyllt och KKT ger nödvändiga optimalitetsvillkor (notera att f inte är konvex). KKT-villkoren är

KKT1	$\begin{bmatrix}5 & -1\end{bmatrix} + \begin{bmatrix}-4 & 8\end{bmatrix} y_1 + \begin{bmatrix}-20 & 4\end{bmatrix} y_2 = 0$
KKT2	$g_1((-1,1)) = 1+2-4 < 0$ ok $g_2((-1,1)) = 5+1-6 = 0$ ok
KKT3	$y_1 \geq 0$ $y_2 \geq 0$
KKT4	$y_1(1+2-4) = 0$ $y_2(5+1-6) = 0$

Från KKT4 ser vi att $y_1 = 0$. Om vi sedan löser ekvationssystemet i KKT1 med avseende på y_2 får vi $(y_1, y_2) = (0, \frac{1}{4})$. Eftersom KKKT2-KKT4 är uppfyllda så har vi visat att $(x_1, x_2) = (1, -1)$ är en KKT punkt.

4. (**a**) Dualen är

$$\begin{aligned} \text{maximera} \quad & 8y_1 + 4y_2 \\ \text{då} \quad & y_1 + y_2 \leq -3 \\ & y_1 - y_2 \leq 4 \\ & -y_1 + y_2 \leq -2 \\ & -y_1 - y_2 \leq 5 \end{aligned} \tag{0.1}$$

Motsvarande tillåtna område och isokostline finns illustrerat i Figur 1.

(**b**) Enligt figuren i uppgiften ser vi att de två första bivillkoren är bindande i optimalpunkten. Från detta kan man lösa ut att $\hat{y} = \begin{bmatrix}0.5 & -3.5\end{bmatrix}^T$ med motsvarande optimalvärde $\hat{z} = 8 \cdot 0.5 - 4 \cdot 3.5 = -10$. Detta framgår förståss också ur figuren.

(**c**) Genom uppdelning av variablerna i basvariabler respektive icke-basvariabler erhåller vi uttrycket

$$\mathbf{c}^\mathsf{T}\mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{y} = (\mathbf{c}_\beta - \mathbf{A}_\beta^\mathsf{T}\mathbf{y})^\mathsf{T}\mathbf{x}_\beta + (\mathbf{c}_\delta - \mathbf{A}_\delta^\mathsf{T}\mathbf{y})^\mathsf{T}\mathbf{x}_\delta = 0$$

Eftersom alla faktorer är positiva gäller följande välkända komplementaritetsegenskap

$$\begin{aligned} &\mathbf{A}_\delta^\mathsf{T}\mathbf{y} < \mathbf{c}_\delta \;\; \Rightarrow \mathbf{x}_\delta = 0 \\ &\mathbf{A}_\beta^\mathsf{T}\mathbf{y} = \mathbf{c}_\beta \;\; \Rightarrow \mathbf{x}_\beta \geq 0 \quad (\mathbf{x}_\beta > 0 \text{ såvida inte baslösningen är degenererad}). \end{aligned}$$

I vårt fall har vi att

$$\begin{aligned} \hat{y}_1 + \hat{y}_2 &= -3 \\ \hat{y}_1 - \hat{y}_2 &= 4 \\ -\hat{y}_1 + \hat{y}_2 &= -4 < -2 \\ -\hat{y}_1 - \hat{y}_2 &= 3 < 5 \end{aligned}$$

varvid följer att $\beta = (1,2)$ och $\delta = (3,4)$. Motsvarande baslösning ges av

$$\begin{bmatrix}x_1 \\ x_2\end{bmatrix} = \mathbf{A}_\beta^{-1}\begin{bmatrix}8 \\ 4\end{bmatrix} = \begin{bmatrix}1 & 1 \\ 1 & -1\end{bmatrix}^{-1}\begin{bmatrix}8 \\ 4\end{bmatrix} = \begin{bmatrix}6 \\ 2\end{bmatrix}$$

vilket ger målfunktionesvärdet $\hat{z} = 6 \cdot (-3) + 4 \cdot 2 = -10$. Primala och duala målfunktionsvärdena är lika vilket bekräftar att $\hat{x} = \begin{bmatrix} 6 & 2 & 0 & 0 \end{bmatrix}^{\mathsf{T}}$ är den primala optimallösningen.

5. (**a**) Låt

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}$$

samt

$$f(\mathbf{x}) = -\sum_{k=1}^{N} U_k(x_k) = -\sum_{k=1}^{N} \ln(x_k).$$

Optimeringsproblemet kan då skrivas

$$\begin{array}{rl} -\text{minimera} & f(\mathbf{x}) \\ \text{då} & \mathbf{A}\mathbf{x} \leq c \\ & \mathbf{x} \geq \mathbf{0} \end{array} \tag{0.2}$$

Eftersom $-\ln(x_k) \to \infty$ då $x_k \to 0$ så följer att bivillkoren $x_k \geq 0$ inte kan vara bindande. Dessa kan därför hållas implicita. Optimeringsproblemet är konvext eftersom det tillåtna området $\mathcal{F} = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq c; \mathbf{x} > \mathbf{0}\}$ är konvext och målfunktionens hessian

$$\mathbf{F}(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{1}{x_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{x_2^2} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \frac{1}{x_N^2} \end{bmatrix}$$

är positivt definit och därmed konvex på $\mathcal{F}$. KKT-villkoren är då nödvändiga och tillräckliga villkor för optimalitet. Lagrange funktionen blir då

$$L(\mathbf{x}, y) = \sum_{k=1}^{N} -\ln(x_k) + y\left(\sum_{k=1}^{N} x_k - c\right) = \sum_{k=1}^{N} (-\ln(x_k) + yx_k) - c$$

varvid KKT villkoren blir

$$\begin{array}{rll} \frac{1}{x_k} + y = 0, \; k = 1, \dots .N & \frac{\partial L}{\partial x_k} = 0 & KKT1 \\ \sum_{k=1}^{N} x_k \leq c & \text{primal tillåtenhet} & KKT2 \\ y \geq 0 & \text{dual tillåtenhet} & KKT3 \\ y\left(\sum_{k=1}^{N} x_k - c\right) = 0 & \text{komplementaritet} & KKT4 \end{array}$$

KKT1 leder till att $x_k = \frac{1}{y}$, $k = 1, \dots N$. Därmed följer att $y > 0$ och från komplementaritetsvillkoret i KKT4 ser vi att bivillkoret måste vara bindande, vilket ger att $y = \frac{N}{c}$ varvid det följer att optimal lösningen är $\hat{x}_k = \frac{c}{N}$, $k = 1, \dots, N$.

(**b**) Målfunktionen

$$f(\mathbf{x}) = \sum_{k=1}^{N} -w_k \frac{x_k^{1-\alpha_k}}{1-\alpha_k}$$

är konvex på det tillåtna området $\mathcal{F} = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq c; \mathbf{x} > \mathbf{0}\}$ (precis som i (**a**) följer att det räcker det att studera problemet då $x_k > 0$, $k = 1, \ldots, N$). Detta följer eftersom hessianen

$$\mathbf{F}(\mathbf{x}) = \nabla f(\mathbf{x}) = \begin{bmatrix} w_1 \frac{\alpha_1}{x_1^{1+\alpha_1}} & 0 & \ldots & 0 \\ 0 & w_2 \frac{\alpha_2}{x_2^{1+\alpha_2}} & \ldots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \ldots & w_N \frac{\alpha_N}{x_N^{1+\alpha_N}} \end{bmatrix}$$

vilken är positivt definit då $\mathbf{x} > 0$. Eftersom det tillåtna området är konvext följer att optimeringsproblemet är konvext.

Låt oss nu använda ledningen. Olikheten

$$f(\hat{\mathbf{x}}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})$$

kan skrivas om som

$$-\nabla f(\mathbf{x})(\hat{\mathbf{x}} - \mathbf{x}) \geq f(\mathbf{x}) - f(\hat{\mathbf{x}}) \geq 0$$

där vi i andra olikheten använder att $\hat{\mathbf{x}}$ är en global minpunkt. Med

$$-\nabla f(\mathbf{x}) = \begin{bmatrix} w_1 \frac{1}{x_1^{\alpha_k}} & w_2 \frac{1}{x_2^{\alpha_k}} & \ldots & w_N \frac{1}{x_N^{\alpha_k}} \end{bmatrix}$$

får vi då att

$$-\nabla f(\mathbf{x})(\hat{\mathbf{x}} - \mathbf{x}) = \sum_{k=1}^{N} w_k \frac{\hat{x}_k - x_k}{x_k^{\alpha_k}} \geq 0$$

vilket skulle visas.