2006–05–11

Optimering, fortsättning

$$(P) \quad \begin{cases} \min f(x) \\ \text{då } \begin{cases} g(x) = 0 \\ h(x) \leqslant 0 \end{cases} \end{cases}$$

I optimering möts två matematiska världar: analkys och linjär algebra.

Exempel. Minimera $f(\boldsymbol{x})$ då $A\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geqslant \boldsymbol{0}, \boldsymbol{x} \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, m < n$.

Tag en punkt $\boldsymbol{x}_0$ i planet. ⟨fig20⟩. $\boldsymbol{d}_0$ är ortogonala projektionen av $-\nabla f(\boldsymbol{x}_0)$ på planet $\{\boldsymbol{x}: A\boldsymbol{x} = \boldsymbol{b}\}$. Om $\boldsymbol{d}_0 \in \mathrm{Nul}(A)$ så blir $\boldsymbol{x}_1$ ok. Visa att om $\boldsymbol{x}_0$ är tillåten och $\boldsymbol{d}_0 \in \mathrm{Nul}(A)$ så är $\boldsymbol{x}_1$ tillåten.

Bevis. $A\boldsymbol{x}_0 = \boldsymbol{b} \Longrightarrow A(\boldsymbol{x}_0 + \boldsymbol{d}_0) = A\boldsymbol{x}_0 + A\boldsymbol{d}_0 = A\boldsymbol{x}_0 = \boldsymbol{b}$, dvs $\boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{d}_0$ tillåten ty $A\boldsymbol{x}_1 = \boldsymbol{b}$. □

Matematiska grundbegrepp

Riktad derivata av ordning k i punkten $\boldsymbol{x}$ i riktning $\boldsymbol{y}$.

$$\left.\frac{\mathrm{d}^k}{\mathrm{d}\tau^2} f(\boldsymbol{x} + \tau\boldsymbol{y})\right|_{\tau=0}$$

Speciellt om $k = 1$ får vi

$$\frac{\mathrm{d}}{\mathrm{d}\tau} f(\boldsymbol{x} + \tau\boldsymbol{y}) = \nabla f(\boldsymbol{x})^T \boldsymbol{y}$$

där

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{pmatrix}$$

är **gradienten** av f i punkten $\boldsymbol{x}$.

Hessianen, $H(x)$, är matrisen med element $h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x})$. Hessianen är alltså Jakobianen av gradenten. H är symmetrisk på grund av symmetri.

Exempel.

$$f(\boldsymbol{x}) = x_1^2 x_3 + x_2^2 x_3^2 + x_2^3$$

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 2x_1 x_3 \\ 2x_2 x_3^2 + 3x_2^2 \\ x_1^2 + 2x_2^2 x_3 \end{pmatrix}$$

$$H(\boldsymbol{x}) = J(\nabla f(\boldsymbol{x})) = \begin{pmatrix} 2x_3 & 0 & 2x_1 \\ 0 & 2x_3^2 + 6x_2 & 4x_2 x_3 \\ 2x_1 & 4x_2 x_3 & 2x_2^2 \end{pmatrix}$$

Speciella typer av optimeringsproblem Linjär programering (LP).

$$\min \boldsymbol{c}^T\boldsymbol{x} \quad \Big| \quad \boldsymbol{c}^T\boldsymbol{x} = c_1 x_1 + \cdots + c_n x_n$$

då $A\,\boldsymbol{x} \leqslant \boldsymbol{b}$, $\boldsymbol{x} \geqslant \boldsymbol{0}$. I ett LP problem har vi alltså en linjär objektfunktion och linjära bivillkor, positivitetskrav.

Matlab: `linprog`.

Exempel (Matnyttigt.) Du behöver 4 enheter protein per dag. Du äter jordnötsmör och/eller biffstek.

Jordnötssmör innehåller 1 enhet protein och kostar 2$ per pund.

Biffstek innehåller 2 enheter protein och kostar 3$ per pund. Hur ska du äta till minimal kostnad?

Lösning: Låt x_1 vara antal pund jordnötssmör man ska äta, x_2 antal pund biffstek.

$$(P) \quad \begin{cases} \min 2x_1 + 3\,x_2 \\ \text{då } x_1 + 2\,x_2 \geqslant 4, \quad x_1 \geqslant 0, x_2 \geqslant 0 \end{cases}$$

Grafisk lösning. ⟨fig21⟩.

Observationer kring (LP)-problemet:

- nivåkurvorna är räta linjer.
- tillåtet område är ett polygonområde (simplex)
- lösning i någon av hörnpunkterna.
- en metod som systematiskt söker upp optimal hörnpunkt är simplexmetoden.

Kvadratisk programmering (QP)

$$\begin{cases} \min \frac{1}{2}\boldsymbol{x}^T H\,\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \\ \text{då } \begin{cases} A\boldsymbol{x} = \boldsymbol{b} \\ \boldsymbol{x} \geqslant \boldsymbol{0} \end{cases}, \text{ linjära bivillkor} \end{cases}$$

Matlab: `quadprog`

Exempel: $\min x_1^2 + 2\,x_2^2 - 2\,x_1$ då $x_1 + 2x_2 \geqslant 4, x_1 \geqslant 0, x_2 \geqslant 0$. Här är:

$$H = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad \boldsymbol{c} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \quad A = [1 \quad -2], \quad b = -4$$

Lösning utan bivillkor: $H\,\boldsymbol{x}^* = -\boldsymbol{c}$, dvs $\boldsymbol{x}^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Med bivillkor: Nivåkurvorna är ellipser:

$$x_1^2 + 2\,x_2^2 - 2\,x_1 = \alpha \Longleftrightarrow (x_1 - 1)^2 + 2\,x_2^2 = \alpha + 1$$

⟨fig22⟩. Bestäm skärningen mellan ellipsen och bivillkor:

$$(3 - 2\,x_2)^2 + 2\,x_2^2 = \alpha + 1 \Longleftrightarrow (x_2 - 1)^2 + \frac{1}{2} = \frac{\alpha + 1}{6}$$

Dubbelrot om $\frac{\alpha - 1}{6} = \frac{1}{2} \Longrightarrow \alpha = 2$ och då är $x_2 = 1$ och $x_1 = 2$.

Några definitioner

$$\begin{cases} \min f(x) \\ \text{då } x \in X, \text{tillåtet område} \end{cases}$$

$x^* \in X$ är **globalt** minimum om $f(x^*) \leqslant f(x), \forall x \in X$.

$x^* \in X$ är **lokalt** minimum om $f(x^*) \leqslant f(x), \forall x \in X$ i en omgivning av x^*.

Strikt olikhet $\Rightarrow$ strikt minimum.

En riktning $\boldsymbol{s}$ i $\boldsymbol{x}$ är **tillåten** om $\boldsymbol{x} + \alpha \boldsymbol{s}$ är tillåten för $0 < \alpha < \delta_1$, något δ_1.

En riktning $\boldsymbol{s}$ i $\boldsymbol{s}$ är en **descentriktning** om $f(\boldsymbol{x} + \alpha\, \boldsymbol{s}) < f(\boldsymbol{x})$ för $0 < \alpha < \delta_2$.

Intressanta riktningar vid minimering: tillåtna descentriktningar.

Ett par resultat

- $\boldsymbol{s}$ är en descentriktning om $\nabla f(\boldsymbol{x})^T \boldsymbol{s} < 0$
- $\boldsymbol{x}^*$ lokalt min $\Rightarrow$ det finns ingen tillåten descentriktning. Omvändningen gäller inte.
- $\boldsymbol{x}^*$ ligger i det inre av X och lokalt min $\Rightarrow \nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

Endimensionell optimering

Problem:

$$\min_{x \in \mathbb{R}} f(x), \qquad f: \mathbb{R} \rightarrow \mathbb{R}$$

Några metoder:

1. Gyllne snittet: kräver endast f-beräkningar, inga derivator.

   Förutsättning: f är unimodal på (för enkelhets skull) [0,1], dvs $\exists! x^*$ i intervallet så att $f(\boldsymbol{x}^*)$ är min och f är strikt avtagande $x \leqslant x^*$ och strikt växande $x \geqslant x^*$.

   Idé: Att med så få f-beräkningar som möjligt stänga in x^* i mindre intervall. ⟨fig23⟩

   Test: Om $f(x_1) < f(x_2)$ så $x^* \in [0, x_2]$. Om $f(x_1) > f(x_2)$ så $x^* \in [x_1, 1]$.

   Som fortsättning ska delintervallen vara lika långa. Anta $x^* \in [x_1, 1]$; vi vill återanvända $f(x_2)$ för att bara göra *en* f-beräkning i x_3. Vi ser att x_2 ska ligga i intervalet $[x_1, x_3]$ ligger intervallet $[x_1, 1]$. Villkoret blir $\frac{1-\tau}{\tau} = \frac{\tau}{1} \Longrightarrow \tau^2 + \tau - 1 = 0$.