5a. Lineare Nebenbedingungen

Gleichungen

Optimierung SoSe 2020

Dr. Alexey Agaltsov

Plan

- Optimalitätsbedingungen
- Sensitivitätsanalyse
- Quadratische Programme mit Gleichungsnebenbedingungen

Mathematisches Programm

$$\text{Minimiere } f(x) \text{ über } x \in \mathbb{R}^n$$
$$\text{u.d.N. } Ax = b$$
$$f \in C^1(\mathbb{R}^n),\ A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

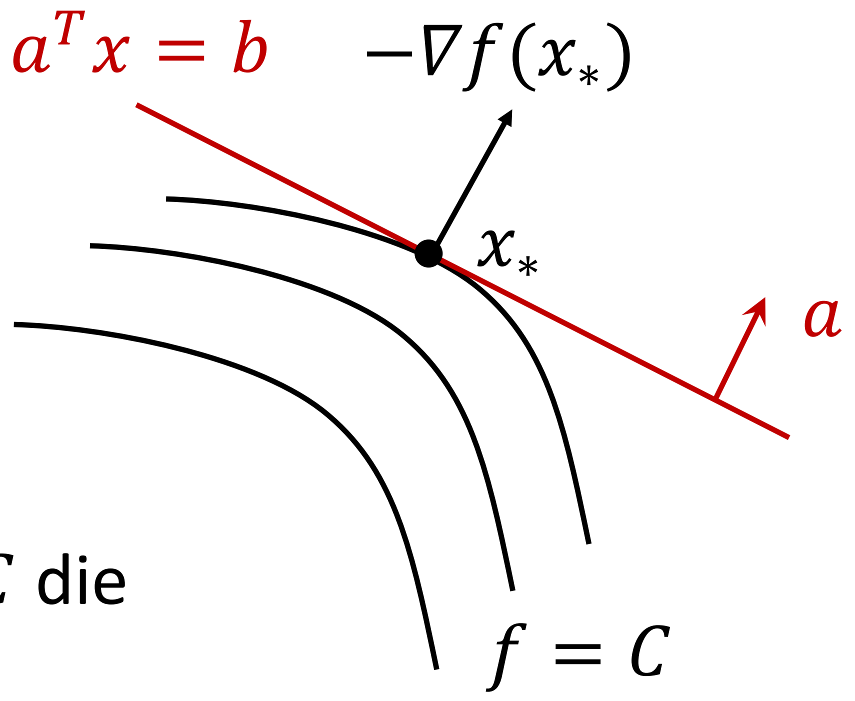
- Angenommen, $\mathrm{rang}(A) = m$ und $m < n$
- Sei x_* eine optimale Lösung
- Wie kann man x_* charakterisieren und bestimmen?

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

Geometrische Intuition

Minimiere $f(x)$

u.d.N. $a^T x = b$



- Wir starten auf der Hyperfläche $f = C$ mit großem C
- Wir verkleinern C bis die Hyperfläche $f = C$ die Hyperebene $a^T x = b$ berührt
- Berührbedingung im Punkt x_*:

$$\nabla f(x_*) = \lambda a$$

Lagrange-Multiplikator

Eliminierung

- In unrestringiertem Fall erfüllt ein optimales x_* die Bedingung:

$$\nabla f(x_*) = 0$$

- Mit der Gleichung $Ax = b$ kann man einige Entscheidungsvariablen eliminieren
- Damit kann man Optimalitätskriterien für unrestringierte Programme auf Gleichungsrestringierte Probleme übertragen

Modifikationen

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

NOB

$$\nabla f(x_*) = 0$$

$$\nabla^2 f(x_*) \succcurlyeq 0$$

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $Ax = b$

NOB

Lagrange-Multiplikatoren

$$\nabla f(x_*) \perp \ker A \; (\Leftrightarrow \nabla f(x_*) = A^T \lambda)$$

$$\nabla^2 f(x_*)|_{\ker A} \succcurlyeq 0$$

Satz 5.1. Optimalitätsbedingungen

Minimiere $f(x)$ $\quad$ $f \in C^1(\mathbb{R}^n),\ A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$

u.d.N. $Ax = b$ $\quad$ $\mathrm{rang}(A) = m < n$

Lagrange-Multiplikatoren

1. Sei x_* ein lokales Minimum , so $\exists!\, \lambda_* \in \mathbb{R}^m$ so dass $\nabla f(x_*) = A^T\lambda_*$

   Ist zusätzlich $f \in C^2(\mathbb{R}^n)$, so gilt $\nabla^2 f(x_*)|_{\ker A} \succcurlyeq 0$:

2. Sei $f \in C^2(\mathbb{R}^n)$. Seien $x_* \in \mathbb{R}^n, \lambda_* \in \mathbb{R}^m$ so, dass:

$$Ax_* = b$$

$$\nabla f(x_*) = A^T\lambda_*,\quad \nabla^2 f(x_*)|_{\ker A} \succ 0$$

   Dann ist x_* ein lokales Minimum

Beweis: Eliminierung

- Wähle $\mathcal{B} \in \{1, \ldots n\}$ so, dass $A_{\mathcal{B}} \coloneqq A[:,\mathcal{B}]$ regulär ist.
- Setze $\mathcal{N} = \mathcal{B}^c$ und $A_{\mathcal{N}} \coloneqq A[:,\mathcal{N}]$

$$Ax = b$$

$$A_{\mathcal{B}}x[\mathcal{B}] + A_{\mathcal{N}}x[\mathcal{N}] = b$$

$$x[\mathcal{B}] = A_{\mathcal{B}}^{-1}(b - A_{\mathcal{N}}x[\mathcal{N}])$$

Beweis: Eliminierung

$$\text{Minimiere } f(x) \text{ über } x \in \mathbb{R}^n$$
$$\text{u.d.N. } Ax = b$$

Angenommen, $\mathcal{B} = \{1, \dots, m\}$

$$\text{Minimiere } f(x[\mathcal{B}], x[\mathcal{N}]) \text{ über } x[\mathcal{B}] \in \mathbb{R}^m, x[\mathcal{N}] \in \mathbb{R}^{n-m}$$
$$\text{u.d.N. } A_{\mathcal{B}} x[\mathcal{B}] + A_{\mathcal{N}} x[\mathcal{N}] = b$$

$$x[\mathcal{N}] = y$$
$$x[\mathcal{B}] = A_{\mathcal{B}}^{-1}(b - A_{\mathcal{N}} y)$$

$$\text{Minimiere } \bar{f}(y) = f(A_{\mathcal{B}}^{-1}(b - A_{\mathcal{N}} y), y) \text{ über } y \in \mathbb{R}^{n-m}$$

- x_* ist ein lokales Min. $\Leftrightarrow$ $y_* = x_*[\mathcal{N}]$ ist ein lokales Min.

Beweis: NOB erster Ordnung

Minimiere $\bar{f}(y) = f(A_{\mathcal{B}}^{-1}(b - A_{\mathcal{N}}y), y)$ über $y \in \mathbb{R}^{n-m}$

Sei $y_* = x_*[\mathcal{N}]$ ein lokales Minimum

$$\nabla_y f\left(A_{\mathcal{B}}^{-1}(b - A_{\mathcal{N}}y_*), y_*\right) = 0$$

$$-\mathrm{A}_{\mathcal{N}}^T \underbrace{A_{\mathcal{B}}^{-T}\nabla_{\mathcal{B}} f(x_*)}_{\lambda_*} + \nabla_{\mathcal{N}} f(x_*) = 0$$

$$\begin{aligned} \nabla_{\mathcal{B}} f(x_*) &= A_{\mathcal{B}}^T \lambda_* \\ \nabla_{\mathcal{N}} f(x_*) &= A_{\mathcal{N}}^T \lambda_* \end{aligned} \quad \Longleftrightarrow \quad \nabla f(x_*) = A^T \lambda_*$$

Beweis: OB zweiter Ordnung

$$\bar{f}(y) = f(A_{\mathcal{B}}^{-1}(b - A_{\mathcal{N}}y), y)$$

$$\nabla^2 \bar{f} = \begin{bmatrix} -A_{\mathcal{N}}^T A_{\mathcal{B}}^{-T} & I \end{bmatrix} \nabla^2 f \begin{bmatrix} -A_{\mathcal{B}}^{-1} A_{\mathcal{N}} \\ I \end{bmatrix}$$

$p^T(\nabla^2 \bar{f})p$

$$p^T(\nabla^2 \bar{f})p = d^T(\nabla^2 f)d, \quad d = \begin{bmatrix} -A_{\mathcal{B}}^{-1} A_{\mathcal{N}} p \\ p \end{bmatrix}$$

$p \in \mathbb{R}^{n-m} \Leftrightarrow d \in \ker A$

$$\nabla^2 \bar{f} \succcurlyeq 0 \; \Leftrightarrow \; \nabla^2 f|_{\ker A} \succcurlyeq 0$$

Satz 2.19

bzw. $\nabla^2 \bar{f} \succ 0 \; \Leftrightarrow \; \nabla^2 f|_{\ker A} \succ 0$

Satz 2.19

x_* lokales Min. $\Rightarrow \nabla^2 f(x_*)|_{\ker A} \succcurlyeq 0$

$\nabla f(x_*) = A^T \lambda^*, \nabla^2 f(x_*)|_{\ker A} \succ 0 \; \Rightarrow x_*$ lokales Min. $\square$

Aufgabe 5.2. Konvexe Zielfunktion

$$\text{Minimiere } f(x) \text{ über } x \in \mathbb{R}^n$$
$$\text{u.d.N. } Ax = b$$

- Sei $f \in C^1(\mathbb{R}^n)$ konvex. Seien $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\operatorname{rang}(A) = m < n$
- $x_* \in \mathbb{R}^n$ mit $Ax_* = b$ ist eine optimale Lösung $\Leftrightarrow \exists \lambda_* \in \mathbb{R}^m$:

$$\nabla f(x_*) = A^T \lambda_*$$

Tipp: Benutzen Sie Eliminierung und verweisen Sie auf Satz 2.20

Lagrange-Funktion

Minimiere $f(x)$ über $x \in \mathbb{R}^n$ $\qquad f \in C^1(\mathbb{R}^n)$

u.d.N. $Ax = b$ $\qquad A \in \mathbb{R}^{m \times n}, \mathrm{rang}(A) = m < n$

- Die Lagrange-Funktion ist gegeben durch:

$$\mathcal{L}(x,\lambda) = f(x) + \lambda^T(b - Ax), \quad x \in \mathbb{R}^n, \ \lambda \in \mathbb{R}^m$$

$$\nabla_x \mathcal{L}(x,\lambda) = \nabla f(x) - A^T\lambda$$
$$\nabla_\lambda \mathcal{L}(x,\lambda) = b - Ax$$

- $x_* \in \mathbb{R}^n$ ist ein lokales Min. $\Rightarrow \exists \lambda_* \in \mathbb{R}^m$ sodass

$$\nabla_x \mathcal{L}(x_*,\lambda_*) = 0, \ \nabla_\lambda \mathcal{L}(x_*,\lambda_*) = 0$$

Beispiel

$$\text{Minimiere} \quad f(x) = \|x\|_2^2 \;\; \text{über } x \in \mathbb{R}^n$$

$$\text{u.d.N.} \quad Ax = b$$

$$A \in \mathbb{R}^{m\times n}, \mathrm{rang}(A) = m < n$$

Optimalitätsbedingung

$$\mathcal{L}(x,\lambda) = x^T x + \lambda^T (b - Ax)$$

$$\begin{aligned} \nabla_x \mathcal{L}(x_*,\lambda_*) &= 2x_* - A^T\lambda_* = 0 \\ \nabla_\lambda \mathcal{L}(x_*,\lambda_*) &= Ax_* - b = 0 \end{aligned} \quad \Rightarrow \quad \begin{aligned} x_* &= A^T\lambda_*/2 \\ \lambda_* &= 2b(AA^T)^{-1} \end{aligned}$$

$$x_* = \underbrace{A^T(AA^T)^{-1}}_{A^+}b$$

Pseudo-Inverse

f ist konvex $\Rightarrow$ optimale Lösung *(Aufgabe 5.2)*

Plan

- Optimalitätsbedingungen
- Sensitivitätsanalyse
- Quadratische Programme mit Gleichungsnebenbedingungen

Sensitivitätsanalyse

- Häufig wollen wir wissen, wie sich der Optimalwert ändert, wenn sich ein oder mehrere Parameterwerte ändern

- Die Lagrange-Multiplikatoren messen die Sensitivät des Optimalwertes auf Änderungen in den rechten Seiten der Nebenbedingungen

Intuition: Störungen in den rechten Seiten

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $a_i^T x = b_i, \quad i = 1, \dots, m$

Angenommen:
eindeutige Lösung $\forall b$

- Seien x_*, λ_* eine optimale Lösung und der Vektor der zugehörigen Lagrange-Multiplikatoren
- Ersetze b_i durch $b_i + \Delta b_i$ für $i = 1, \dots, m$ und bezeichne die neue optimale Lösung als $x_* + \Delta x$

$$
\begin{aligned}
f(x_* + \Delta x) &\approx f(x_*) + \nabla f(x_*)^T \Delta x \\
&= f(x_*) + \textstyle\sum_{i=1}^m \lambda_{*,i} a_i^T \Delta x \\
&= f(x_*) + \textstyle\sum_{i=1}^m \lambda_{*,i} \Delta b_i
\end{aligned}
$$

$\nabla f(x_*) = [a_1, \dots, a_m]\lambda_*$

$a_i^T \Delta x = \Delta b_i$

Satz 5.3. Umhüllungssatz

$$g(a) = \min\{f(x,a):\ x \in \mathbb{R}^n\}\ ,\ a \in \mathbb{R}^k$$

$$f \in C^1(\mathbb{R}^n \times \mathbb{R}^k)$$

- Angenommen, $x_*(a)$ ist ein eindeutiges Minimum $\forall a \in \mathbb{R}^k$ und $x_*(a) \in C^1(\mathbb{R}^k)$. Dann gilt:

$$\nabla g(a) = \nabla_a f(x_*(a), a)$$

Ist $f \in C^2(\mathbb{R}^n)$ und $\nabla^2_{xx} f(x_*(a), a) \succ 0$, so folgt es aus dem Satz von der impliziten Funktion

Beweis

$$g(a) = f(x_*(a), a)$$

$$\frac{\partial g}{\partial a_j}(a) = \sum_{i=1}^{n} \underbrace{\frac{\partial f}{\partial x_i}(x_*(a), a)}_{=0} \frac{\partial x_{*,i}(a)}{\partial a_j} + \frac{\partial f}{\partial a_j}(x_*(a), a)$$

Optimalitätsbedingung

$$\frac{\partial g}{\partial a_j}(a) = \frac{\partial f}{\partial a_j}(x_*(a), a)$$

$\square$

Satz 5.4. Interpretation der Lagrange-Multiplikatoren

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

u.d.N. $Ax = b$

$f \in C^1(\mathbb{R}^n)$

$A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$

$\mathrm{rang}(A) = m < n$

- Sei $x_*(b)$ ein eindeutiges Minimum und $\lambda_*(b) \in \mathbb{R}^m$ der zugehörige Lagrange-Multiplikator $\forall b \in \mathbb{R}^m$. Angenommen, $x_*(b) \in C^1(\mathbb{R}^m)$
- Dann gilt:

$$\nabla_b f(x_*(b)) = \lambda_*(b)$$

Ist $f \in C^2(\mathbb{R}^n), \nabla^2 f|_{\ker A} \succ 0$, so folgt es aus dem Satz von der impliziten Funktion

Beweis

Angenommen (o.B.d.A.), $A[:,\mathcal{B}]$ ist regulär mit $\mathcal{B} = \{1, \dots, m\}$

$$\xrightarrow{\textit{Eliminierung}}$$

$$f(x_*(b)) = \min\{f\left(A_\mathcal{B}^{-1}(b - A_\mathcal{N} y), y\right) \text{ über } y \in \mathbb{R}^{n-m}\}$$

$$\xrightarrow{y_*(b) = x_*[\mathcal{B}^c]}$$

$$= f\left(A_\mathcal{B}^{-1}\left(b - A_\mathcal{N} y_*(b)\right), y_*(b)\right)$$

$$\xrightarrow{\text{Umhüllungssatz 5.3}}$$

$$\nabla_b f(x_*(b)) = \nabla_b f\left(A_\mathcal{B}^{-1}(b - A_\mathcal{N} y), y\right)\Big|_{y=y_*(b)}$$

$$= \underbrace{A_\mathcal{B}^{-T} \nabla_\mathcal{B} f(x_*(b))}_{\lambda_*(b)}$$

$\square$

Plan

- Optimalitätsbedingungen
- Sensitivitätsanalyse
- Quadratische Programme mit Gleichungsnebenbedingungen

Quadratisches Programm mit GNB

Minimiere $f(x) = \frac{1}{2}x^T Qx - c^T x + r$ über $x \in \mathbb{R}^n$

u.d.N. $Ax = b$

$Q \in \mathbb{S}^n_{\succcurlyeq}$, $A \in \mathbb{R}^{m \times n}, \operatorname{rang}(A) = m < n$

$$\nabla f(x) = Qx - c$$

$$\nabla^2 f(x) = Q \succcurlyeq 0$$

$\Longrightarrow$ f ist konvex

Optimalitätsbedingung

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^T Q x - c^T x + r + \lambda^T (Ax - b)$$

$$\begin{aligned} \nabla_x \mathcal{L}(x,\lambda) &= Qx - c + A^T\lambda = 0 \\ \nabla_\lambda \mathcal{L}(x,\lambda) &= b - Ax = 0 \end{aligned} \quad \Rightarrow \quad \begin{aligned} Qx_* + A^T\lambda_* &= c \\ Ax_* &= b \end{aligned}$$

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} c \\ b \end{bmatrix}$$ *Sattelpunktsgleichung*

Satz 5.5. Optimalitätsbedingungen

$$\text{Minimiere } f(x) = \tfrac{1}{2}x^T Qx - c^T x + r,$$
$$\text{u.d.N. } Ax = b$$
$$Q \in \mathbb{S}^n_{\succcurlyeq}, A \in \mathbb{R}^{m\times n}, \mathrm{rang}(A) = m < n$$

- Die folgenden Aussagen sind äquivalent:
  1. x^* ist eine optimale Lösung
  2. $\exists \lambda^* \in \mathbb{R}^m$, sodass (x^*, λ^*) erfüllt $\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} c \\ b \end{bmatrix}$ $\Leftarrow$ Aufgabe 5.2
- Besitzt die Sattelpunktgleichung keine Lösung, so ist das Problem nicht nach unten beschränkt

Beweis: Keine Lösung ⇒ Unbeschränktheit

$K\begin{bmatrix} x_* \\ \lambda_* \end{bmatrix} = \begin{bmatrix} c \\ b \end{bmatrix}$ besitzt keine Lösung, $K = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$

$$\begin{bmatrix} c \\ b \end{bmatrix} \notin \mathrm{im}(K) = \ker(K)^\perp$$

$$\exists \underbrace{[v,w]^T \in \ker K}_{Qv + A^T w = 0,\ Av = 0}: \quad v^T c + w^T b > 0$$

Sei x so, dass $Ax = b$

$x + tv$ ist zulässig $\forall t$

$$f(x+tv) = f(x) + t(v^T Qx - c^T v) + \frac{1}{2}t^2 v^T Qv$$

$$= f(x) - t(w^T Ax + c^T v) - \frac{1}{2}t^2 w^T Av$$

$$= \underbrace{f(x) - t(w^T b + c^T v)}_{\to -\infty \text{ für } t \to \infty}$$

$v^T Q = -w^T A$

$Ax = b, Av = 0$

$\square$

Lösung der Sattelpunktsgleichung

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x_* \\ \lambda_* \end{bmatrix} = \begin{bmatrix} c \\ b \end{bmatrix}$$

Angenommen, $Q \in \mathbb{S}^n_{\succ}$
$A \in \mathbb{R}^{m \times n}$, $\mathrm{rang}(A) = m < n$

$$Qx + A^T\lambda_* = c$$

Multipliziere mit Q^{-1}

$$x_* + Q^{-1}A^T\lambda_* = Q^{-1}c$$

Multipliziere mit A

$$Ax_* + AQ^{-1}A^T\lambda_* = AQ^{-1}c$$

$$Ax_* = b$$

$$AQ^{-1}A^T\lambda_* = AQ^{-1}c - b$$

$AQ^{-1}A^T$ ist regulär

$$\lambda_* = (AQ^{-1}A^T)^{-1}(AQ^{-1}c - b)$$

$$x_* = Q^{-1}c - Q^{-1}A^T\lambda_*$$

Aufgabe 5.6: Lösbarkeit des Sattelpunktsgleichung

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} c \\ b \end{bmatrix}$$

$$Q \in \mathbb{S}^n_{\succcurlyeq}$$
$$A \in \mathbb{R}^{m \times n}, \operatorname{rang}(A) = m < n$$

Zeigen Sie, dass die folgenden Aussagen äquivalent sind:

1. Die Sattelpunktsgleichung ist eindeutig lösbar für alle c, b
2. $\ker Q \cap \ker A = \{0\}$
3. $Q + A^T A \succ 0$
4. $Z^T Q Z \succ 0$ wobei $Z \in \mathbb{R}^{n \times n-m}$ mit im $Z = \ker A$

Zusammenfassung

- Optimalitätsbedingungen
- Sensitivitätsanalyse
- Quadratische Programme mit Gleichungsnebenbedingungen

Nächstes Video

- 5b. Lineare Nebenbedingungen: Ungleichungen