

Restringierte Optimierung – Die Lagrange-Funktion

DIPL.-MATH. MARTIN GUBISCH

WIEDERHOLUNG.

Wir betrachten das restringierte Optimierungsproblem

$$\min_{x \in \mathbb{R}^n} J(x) \quad \text{u.d.N.B.} \quad e(x) = 0 \quad (1)$$

mit **Kostenfunktional** $J \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R})$, **zulässiger Menge** $F := \{x \in \mathbb{R}^n \mid e(x) = 0\}$ und **Gleichungsnebenbedingungen** $e \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$, $m \leq n$.

SATZ. (Notwendige Optimalitätsbedingungen erster Ordnung)

Sei x^* eine reguläre, lokale Lösung von (1), d.h. es gelten

- (a) Es gibt $U \subseteq \mathbb{R}^n$ offen mit $x^* \in U$ und $J(x^*) \leq J(x)$ für alle $x \in U \cap F$,
- (b) $\{\nabla e_1(x^*), \dots, \nabla e_m(x^*)\} \subseteq \mathbb{R}^n$ ist linear unabhängig.

Dann existiert genau ein Vektor $\lambda^* \in \mathbb{R}^m$ von **Lagrange-Multiplikatoren** mit

$$\nabla J(x^*) + \sum_{i=1}^m \lambda_i^* \nabla e_i(x^*) = \nabla J(x^*) + (\lambda^*)^t \nabla e(x^*) = 0.$$

Insbesondere ist $\nabla J(x^*) \in \text{span}(\nabla e_1(x^*), \dots, \nabla e_m(x^*))$.

BEWEIS.

Eindeutigkeit: Seien λ^* und μ^* solche Multiplikatoren, dann gilt

$$\sum_{i=1}^m (\lambda_i^* - \mu_i^*) \nabla e_i(x^*) = 0 \quad \implies \quad \text{alle Koeffizienten } \lambda_i^* - \mu_i^* = 0,$$

d.h. $\lambda^* = \mu^*$.

Existenz: Der Fall $m = n$ ist trivial, da $\nabla J(x^*) \in \mathbb{R}^n = \text{span}(\nabla e_1(x^*), \dots, \nabla e_m(x^*))$.

Gelte also $m < n$. Trick: Wir lösen die m Nebenbedingungen nach gewissen m Variablen auf und ersetzen diese Variablen in J dann durch die Nebenbedingungen. Auf diese Weise erhalten wir ein $(n - m)$ -dimensionales, unrestringiertes Optimierungsproblem.

Wegen $\dim \text{Zeilenraum} \nabla e(x^*) = m$ lassen sich die Variablen x_1, \dots, x_n so umnummerieren, dass $x = (x_B, x_R) \in \mathbb{R}^m \times \mathbb{R}^{n-m}$ und $\nabla_B e(x^*) \in \mathbb{R}^{m \times m}$ invertierbar ist. Wegen $e(x_B^*, x_R^*) = 0$ kann die Variable x_B nach dem Satz über implizite Funktionen dann lokal bei x^* eliminiert werden, d.h. es gibt ein $r > 0$ und ein $\Phi : B_r(x_R^*) \rightarrow \mathbb{R}^m$ mit

$$\Phi(x_R^*) = x_B^* \quad \& \quad \forall x_R \in B_r(x_R^*) : e(\Phi(x_R), x_R) = 0.$$

Nach der Regel zum impliziten Differenzieren gilt weiter

$$\forall x_R \in B_r(x_R^*) : \nabla \Phi(x_R) = -(\nabla_B e(\Phi(x_R), x_R))^{-1} \nabla_R e(\Phi(x_R), x_R). \quad (2)$$

Das reduzierte Problem

$$\min_{x_R \in \mathbb{R}^{n-m}} \hat{J}(x_R) \quad \text{u.d.N.B.} \quad x_R \in B_r(x_R^*) \quad (3)$$

mit reduziertem Kostenfunktional $\hat{J} : \mathbb{R}^{n-m} \rightarrow \mathbb{R}$, $\hat{J}(x_R) = J(\Phi(x_R), x_R)$ ist dann unrestringiert und wird lokal von x_R^* gelöst. Es gilt also die notwendige Optimalitätsbedingung:

$$\nabla \hat{J}(x_R^*) = 0. \quad (4)$$

Berechnung von $\nabla \hat{J}(x_R^*)$:

$$\begin{aligned} & \nabla \hat{J}(x_R) \\ \stackrel{\text{Kettenregel}}{=} & \nabla_B J(\Phi(x_R), x_R) \nabla \Phi(x_R) + \nabla_R J(\Phi(x_R), x_R) \\ \stackrel{(2)}{=} & -\nabla_B J(\Phi(x_R), x_R) (\nabla_B e(\Phi(x_R), x_R))^{-1} \nabla_R e(\Phi(x_R), x_R) + \nabla_R J(\Phi(x_R), x_R) \\ \stackrel{x_R := x_R^*}{=} & -\nabla_B J(x^*) (\nabla_B e(x^*))^{-1} \nabla_R e(x^*) + \nabla_R J(x^*) \\ = & (\lambda^*)^t \nabla_R e(x^*) + \nabla_R J(x^*) \\ \stackrel{(4)}{=} & 0 \end{aligned} \quad (5)$$

mit

$$\lambda^* := -(\nabla_B J(x^*) (\nabla_B e(x^*))^{-1})^t \in \mathbb{R}^m. \quad (6)$$

Es gelten also

$$\begin{cases} \nabla_R J(x^*) + (\lambda^*)^t \nabla_R e(x^*) \stackrel{(5)}{=} 0 \\ \nabla_B J(x^*) + (\lambda^*)^t \nabla_B e(x^*) \stackrel{(6)}{=} 0, \end{cases} \quad (7)$$

zusammen folgt die Behauptung

$$\nabla J(x^*) + (\lambda^*)^t \nabla e(x^*) = 0 \quad \square$$

BEMERKUNG. (Lagrange-Funktion)

Das Optimalitätssystem

$$\begin{cases} \nabla J(x^*) + (\lambda^*)^t \nabla e(x^*) = 0 \\ e(x^*) = 0 \end{cases} \quad (8)$$

ist ein nichtlineares Gleichungssystem in den Unbestimmten $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$.

Mit Hilfe der **Lagrange-Funktion**

$$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}, \quad \mathcal{L}(x, \lambda) := J(x) + \lambda^t e(x) \quad (9)$$

lässt sich (8) schreiben als

$$\nabla \mathcal{L}(x^*, \lambda^*) = 0. \quad (10)$$

BEMERKUNG. (Numerische Lösungsstrategien)

(1) **Newton-Verfahren.**

Unter den Voraussetzungen

(a) $J \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$ und $e \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^m)$ mit Lipschitz-stetigen zweiten Ableitungen und

(b) $\nabla_x^2 \mathcal{L}(x^k, \lambda^k)$ ist positiv definit auf $\text{Kern}(\nabla e(x^k))$

ist die Newton-Iterationsvorschrift

$$\begin{cases} \begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} + \begin{pmatrix} \Delta x^k \\ \Delta \lambda^k \end{pmatrix} \\ \nabla^2 \mathcal{L}(x^k, \lambda^k) \begin{pmatrix} \Delta x^k \\ \Delta \lambda^k \end{pmatrix} = -\nabla \mathcal{L}(x^k, \lambda^k) \end{cases}$$

wohldefiniert und die generierte Iterationsfolge $(x^k)_{k \in \mathbb{N}}$ konvergiert gegen x^* .

Beachte: Wegen

$$\nabla^2 \mathcal{L}(x, \lambda) = \begin{pmatrix} \nabla_x^2 \mathcal{L}(x, \lambda) & \nabla e(x)^t \\ \nabla e(x) & 0 \end{pmatrix}$$

wird in jedem Schritt der Iteration zuerst $\Delta x^k = \Delta x^k(x^k, \lambda^k)$ bestimmt und danach $\Delta \lambda^k = \Delta \lambda^k(\Delta x^k, x^k, \lambda^k)$.

(2) **SQP-Verfahren.**

Die Optimalitätsbedingungen der Lagrange-Funktion $\tilde{\mathcal{L}}(y^k, \mu^k) := \tilde{J}(y^k) + \mu^k \tilde{e}(y^k)$ des restringierten quadratischen Optimierungsproblems

$$\begin{cases} \min_{y^k \in \mathbb{R}^n} \frac{1}{2} \langle y^k, \nabla_x^2 \mathcal{L}(x^k, \lambda^k) y^k + \nabla J(x^k) y^k \rangle \\ \text{u.d.N.B. } \nabla e(x^k) y^k + e(x^k) = 0 \end{cases}$$

lauten – in kompakter Matrix-Schreibweise mit Lagrange-Multiplikator $\mu^k \in \mathbb{R}^m$ –

$$\begin{pmatrix} \nabla_x^2 \mathcal{L}(x^k, \lambda^k) & \nabla e(x^k)^t \\ \nabla e(x^k) & 0 \end{pmatrix} \begin{pmatrix} y^k \\ \mu^k \end{pmatrix} = \begin{pmatrix} -\nabla J(x^k)^t \\ -e(x^k) \end{pmatrix} = -\begin{pmatrix} \nabla_x \mathcal{L}(x^k, \lambda^k) \\ \nabla_\lambda \mathcal{L}(x^k, \lambda^k) \end{pmatrix} + \begin{pmatrix} \lambda^k \nabla e(x^k) \\ 0 \end{pmatrix},$$

also ist die Lösung des quadratischen Problems gegeben als

$$\begin{pmatrix} y^k \\ \mu^k \end{pmatrix} = \begin{pmatrix} \Delta x^k \\ \lambda^{k+1} \end{pmatrix} \implies \begin{pmatrix} \Delta x^k \\ \Delta \lambda^k \end{pmatrix} = \begin{pmatrix} y^k \\ \mu^k - \lambda^k \end{pmatrix}.$$

Die Lösung x^* des restringierten Optimierungsproblems (8) lässt also approximativ ermitteln durch das Lösen einer Folge quadratischer restringierter Optimierungsprobleme – daher der Name „SQP“ (**sequential quadratic programming**).

Die einzelnen quadratischen Probleme lassen sich dann beispielsweise mit dem Newton-Verfahren behandeln.

BEMERKUNG.

Betrachte das Optimierungsproblem

$$\min_{x \in \mathbb{R}^2} J(x) = x_1 + x_2 \quad \text{u.d.N.B.} \quad e(x) := \begin{pmatrix} (x_1 - 1)^2 + x_2^2 - 1 \\ (x_1 - 2)^2 + x_2^2 - 4 \end{pmatrix} = 0.$$

Geometrisch: Ist $e_{r,p}(x) = (x_1 - p_1)^2 + (x_2 - p_2)^2 - r^2$ mit $r > 0$ und $p \in \mathbb{R}^2$, dann beschreibt $K_{r,p} = \{x \in \mathbb{R}^2 \mid e_{r,p}(x) = 0\}$ den Kreis mit Radius r um den Mittelpunkt p .

Wegen $F = \{(0, 0)\}$ ist $x^* = (0, 0)$ die eindeutige Lösung des Problems und es gelten

$$\nabla J(x^*) = (1, 1), \quad \nabla e_1(x^*) = (-2, 0), \quad \nabla e_2(x^*) = (-4, 0),$$

d.h. $\nabla J(x^*) \notin \text{span}(\nabla e_1(x^*), \nabla e_2(x^*))$ und somit existiert auch kein Lagrange-Multiplikator $\lambda^2 \in \mathbb{R}^2$ zu x^* .

Dies stellt keinen Widerspruch zum Satz über die notwendigen Optimalitätsbedingungen erster Ordnung dar, da x^* kein regulärer Punkt ist.