Wissenschaftliches Rechnen - Großübung 6.2

Themen: Newton-Verfahren, Lagrange-Multiplikatoren

Ugo & Gabriel

7. Februar 2023

Aufgabe 1: Newton-Verfahren

1. Wie lautet der Iterationsschritt des Newton-Verfahrens zum finden von Optima einer mehrdimensionalen Funktion $f : \mathbb{R}^n \to \mathbb{R}$?

   Lösung

   $$\mathbf{x}_{t+1} \leftarrow x_t - \mathbf{H}_f^{-1}(\mathbf{x}_t)\nabla f(\mathbf{x}_t)$$

   Lösung Ende

2. Welche geometrische Bedeutung hat ein Iterationsschritt des Newton-Verfahrens?

   Lösung

   Wir führen eine Taylorapproximation 2. Ordnung T_2 für die Funktion f am Punkt $\mathbf{x}_t$ durch. Der neue Punkt $\mathbf{x}_{t+1}$ ist der bzw. ein kritischer Punkt von T_2.

   Lösung Ende

3. Der Iterationsschritt hängt von der Hessematrix des aktuellen Zustandes ab. Wie verhält sich das Verfahren, wenn die Hessematrix
   a) positiv definit ist?
   b) negativ definit ist?
   c) indefinit, aber regulär ist?

   Lösung

   a) Falls $\mathbf{H}_f$ positiv definit ist, so ist T_2 konvex und der neue Punkt ist das Minimum von T_2.
   b) Falls $\mathbf{H}_f$ negativ definit ist, so ist T_2 konkav und der neue Punkt dessen Maximum.
   c) Falls $\mathbf{H}_f$ indefinit (und regulär) ist, so ist der Sattelpunkt von T_2 der neue Punkt.

   Lösung Ende

4. Wie lautet der Iterationsschritt des Gradientenabstiegsverfahrens (Gradient Descent)?

   Lösung

   $$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - \alpha\nabla f(\mathbf{x}_t)$$

   Lösung Ende

5. Unter welchen Umständen verhält sich das Gradientenverfahren wie das Newton-Verfahren?

Lösung

Die Verfahren verhalten sich gleich, falls f eine konstante Hessematrix mit $\mathbf{H}_f(\mathbf{x}) = \frac{1}{\alpha}\mathbf{I}$ hat. Somit ist f eine quadratische Funktion, dessen Niveaumengen Kreise sind.

Lösung Ende

6. Nun möchten wir untersuchen, wie sich das Newton-Verfahren sowie das Gradientenabstiegsverfahren bei einer affinen[1] Koordinatentransformation $T : \mathbb{R}^n \to \mathbb{R}^n$ mit $T(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ verhalten, wobei $\mathbf{A} \in \mathbb{R}^{n \times n}$ regulär ist und $\mathbf{b} \in \mathbb{R}^n$. Zeigen Sie, dass das Newton-Verfahren invariant bezüglich besagter Transformation ist, das Gradientenverfahren hingegen nicht.

   Hinweis: Sie sollen untersuchen, ob $T(\mathbf{y}_{t+1}) = \mathbf{x}_{t+1}$, falls $T(\mathbf{y}_t) = \mathbf{x}_t$.

Lösung

Zunächst schauen wir uns an, wie der Gradient und die Hessematrix sich unter affinen Transformationen verhalten. Seien $\mathbf{x}_t = \mathbf{A}\mathbf{y}_t + \mathbf{b}$ sowie $g(\mathbf{y}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$. Es gilt:

$$\nabla g(\mathbf{y}) = \mathbf{A}^\top \nabla f(\mathbf{A}\mathbf{y} + \mathbf{b}) \qquad \mathbf{H}_g(\mathbf{y}) = \mathbf{A}^\top \mathbf{H}_f(\mathbf{A}\mathbf{y} + \mathbf{b})\mathbf{A}$$

Ferner ist dann:

$$\begin{aligned}
T(\mathbf{y}_{t+1}) &= \mathbf{A}(\mathbf{y}_t - \Delta\mathbf{y}_t) + \mathbf{b} \\
&= \mathbf{A}\mathbf{y}_t - \mathbf{A}\Delta\mathbf{y}_t + \mathbf{b} \\
&= \mathbf{A}\mathbf{x}_t + \mathbf{b} - \mathbf{A}\mathbf{H}_g^{-1}(\mathbf{y}_t)\nabla g(\mathbf{y}_t) \\
&= \mathbf{A}\mathbf{y}_t + \mathbf{b} - \mathbf{A}(\mathbf{A}^\top \mathbf{H}_f(\mathbf{A}\mathbf{y}_t + \mathbf{b})\mathbf{A})^{-1}\mathbf{A}^\top \nabla f(\mathbf{A}\mathbf{y}_t + \mathbf{b}) \\
&= \mathbf{A}\mathbf{y}_t + \mathbf{b} - \mathbf{A}\mathbf{A}^{-1}\mathbf{H}_f^{-1}(\mathbf{A}\mathbf{y}_t + \mathbf{b})\mathbf{A}^{-\top}\mathbf{A}^\top \nabla f(\mathbf{A}\mathbf{y}_t + \mathbf{b}) \\
&= \mathbf{A}\mathbf{y}_t + \mathbf{b} - \mathbf{H}_f^{-1}(\mathbf{A}\mathbf{y} + \mathbf{b})\nabla f(\mathbf{A}\mathbf{y}_t + \mathbf{b}) \\
&= \mathbf{x}_t - \mathbf{H}_f^{-1}(\mathbf{x}_t)\nabla f(\mathbf{x}_t) \\
&= \mathbf{x}_{t+1}
\end{aligned}$$

Interpretation des Ergebnisses: Das Newton-Verfahren kann durch einen Koordinatenwechsel nicht verbessert werden. Hier ein Diagramm dazu:

$$\begin{tikzcd}
\mathbf{y}_t \arrow[r, "T"] \arrow[d, "\text{Newton mit } g"'] & \mathbf{x}_t \arrow[d, "\text{Newton mit } f"] \\
\mathbf{y}_{t+1} \arrow[r, "T"] & \mathbf{x}_{t+1}
\end{tikzcd}$$

[1]Eine affine Transformation entspricht einer linearen Transformation gefolgt von einer Translation.

Das Gradientenverfahren ist nicht invariant bzgl. affinen Transformationen.

$$\begin{aligned}
T(\mathbf{y}_{t+1}) &= \mathbf{A}(\mathbf{y}_t - \alpha \nabla g(\mathbf{y}_t)) + \mathbf{b} \\
&= \mathbf{A}\mathbf{y}_t + \mathbf{b} - \alpha \mathbf{A} \nabla g(\mathbf{y}_t) \\
&= \mathbf{A}\mathbf{y}_t + \mathbf{b} - \alpha \mathbf{A}\mathbf{A}^\mathsf{T} \nabla f(\mathbf{A}\mathbf{y}_t + \mathbf{b}) \\
&= \mathbf{x}_t - \alpha \mathbf{A}\mathbf{A}^\mathsf{T} \nabla f(\mathbf{x}_t) \\
&\neq \mathbf{x}_{t+1}
\end{aligned}$$

Lösung Ende

Aufgabe 2: Lagrange-Multiplikatoren

1. Wir haben einen 100 m langen Metalldraht, welchen wir zu einem rechteckigen Zaun mit den Seitenlängen x und y und maximaler Fläche spannen wollen. Das kann man als folgendes Optimierungsproblem schreiben:

$$\max \quad xy \quad \text{s.t.} \quad 2x + 2y = 100$$

Die Funktion $f(x, y) = xy$ beschreibt die Fläche und die Funktion $g(x, y) = 2x + 2x$ beschreibt den Umfang.

a) Lösen Sie das Problem ohne Verwendung von Lagrange-Multiplikatoren.

b) Lösen Sie das Problem mithilfe von Lagrange-Multiplikatoren.

Lösung

a) Wir können y als Funktion von x darstellen und dies in die Zielfunktion einsetzen. Wir erhalten $y = 50 - x$ und die neue (eindimensionale) Zielfunktion $f(x) = x(50 - x)$. Diese leiten wir ab: $f'(x) = 50 - 2x$. Der einzige Kandidat ist $x = 25$. Wegen der zweiten Ableitung $f''(x) = -2$ wissen wir, dass es sich dabei um ein Maximum handelt. Aus $x = 25$ können wir y rekonstruieren, nämlich $y = 50 - x = 50 - 25 = 25$. Somit sollte der Zaun quadratisch sein.

b) Wir definieren die Lagrange-Funktion:

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - c) = xy - \lambda(2x + 2y - 100)$$

Diese leiten wir partiell ab und erhalten nach Gleichsetzen mit Null folgendes Gleichungssystem:

$$\frac{\partial \mathcal{L}}{\partial x} = y - 2\lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial x} = x - 2\lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial x} = 2x + 2y - 100 = 0$$

Aus den ersten zwei Bedingungen erhalten wir $x = y$. Setzen wir das in die Nebenbedingung (also die dritte Gleichung) ein, erhalten wir $4x = 100$, also $x = 25$ und damit auch $y = 25$.

Das Folgende geht über den Stoff dieses Moduls hinaus:

Zu überprüfen, ob der kritische Punkt (x^*, y^*) ein Minimum, Maximum oder Sattelpunkt unter der Nebenbedingung ist, ist mithilfe des üblichen Kriteriums (Definitheit der Hessematrix) nicht möglich. Dazu gibt es ein anderes Kriterium, mit der sogenannten Bordered Hessian:

$$\mathbf{H}_{\mathcal{L}}(\mathbf{x}, \lambda) = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \lambda^2} & \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \mathbf{x}} \\ \left(\frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \mathbf{x}}\right)^\mathsf{T} & \frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}^2} \end{bmatrix} = \begin{bmatrix} 0 & \frac{\partial g}{\partial \mathbf{x}} \\ \left(\frac{\partial g}{\partial \mathbf{x}}\right)^\mathsf{T} & \frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}^2} \end{bmatrix}$$

Aufgrund der Null bzw. des Nullblocks bei mehreren Bedingungen ist die Matrix niemals positiv definit bzw. negativ definit. Stattdessen lautet das Kriterium:

- Falls ihre Determinante positiv ist, so ist der kritische Punkt ein lokales Maximum.

- Falls ihre Determinante negativ ist, so ist der kritische Punkt ein lokales Minimum. Achtung: Hier ist es, anders als beim üblichen Kriterium, nicht umgekehrt.

Die Bordered Hessian für unser Problem ist folgende konstante Matrix:

$$\mathbf{H}_{\mathcal{L}} = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$$

Ihre Determinante ist 8, also ist der Punkt ein lokales Maximum.

Lösung Ende

2. Wie kann man das folgende Optimierungsproblem mithilfe von Lagrange-Multiplikatoren lösen, obwohl die Nebenbedingung keine Gleichung ist?

$$\min \quad x^2 - xy + y \quad \text{s.t.} \quad x^2 + y^2 \leq 9$$

Lösung

Die Nebenbedingung ist eine Ungleichung und beschreibt eine Kreisscheibe mit Mittelpunkt $(0,0)$ und Radius 3 in der Ebene. Dieses Problem lässt sich in zwei separate Probleme umwandeln:

a) Wir finden alle kritischen Punkte ohne Nebenbedingung durch $\nabla f(x,y) = (0,0)$, wobei $f(x,y) = x^2 - xy + y$. Wir überprüfen für jeden normalen kritischen Punkt, ob er die Nebenbedingung erfüllt, also im Inneren des Kreises ist.

b) Wir finden alle kritischen Punkte unter der Nebenbedingung $x^2 + y^2 = 9$ mithilfe von Lagrange-Multiplikatoren.

Dann setzen wir alle Punkte in die Zielfunktion f ein. Das Minimum unter ihnen ist der gesuchte Punkt.

Hinweis: In diesem Fall beschreibt die Bedingung eine kompakte Menge, sodass unter den gefundenen Punkten tatsächlich das globale Minimum (und sogar das globale Maximum) vorhanden ist (vgl. Satz vom Minimum/Maximum). Das muss im Allgemeinen nicht der Fall sein.

Lösung Ende

3. Wir betrachten folgendes Optimierungsproblem:

$$\max \quad \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \quad \text{s.t.} \quad \|\mathbf{x}\| = 1$$

Dabei ist $\mathbf{A} \in \mathbb{R}^{n\times n}$ eine symmetrische, positiv semidefinite Matrix und $\|\cdot\|$ die ℓ^2-Norm.

a) Um welches Ihnen wohlbekannte Problem handelt es sich hierbei?

b) Schreiben Sie die Nebenbedingung um, sodass sie keine Wurzel mehr enthält. Warum ist dies nützlich?

c) Geben Sie eine Lagrange-Funktion zum Problem an.

d) Geben Sie die notwendige Bedingung für einen kritischen Punkt unter der Nebenbedingung an.

e) Warum befindet sich das globale Minimum sowie das globale Maximum garantiert unter den kritischen Punkten?

f) Welcher Vektor $\mathbf{x} \in \mathbb{R}^n$ löst das Optimierungsproblem? Was ist der zugehörige Funktionswert?

Lösung

a) Hauptkomponentenanalyse (PCA)

b) $\|\mathbf{x}\|^2 = \mathbf{x}^\mathsf{T}\mathbf{x} = 1$

c) $\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} + \lambda(1 - \mathbf{x}^\mathsf{T}\mathbf{x})$

d) Es gilt:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{x}} &= 2\mathbf{A}\mathbf{x} - 2\lambda\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow 2\mathbf{A}\mathbf{x} = 2\lambda\mathbf{x} \\ &\Leftrightarrow \mathbf{A}\mathbf{x} = \lambda\mathbf{x} \end{aligned}$$

Jeder kritische Punkt unter der Nebenbedingung ist ein Eigenvektor von $\mathbf{A}$.

e) Die Nebenbedingung beschreibt eine Sphäre (Rand einer Kugel) im $\mathbb{R}^n$, welche eine kompakte Menge ist. Stetige Funktionen auf kompakten Mengen nehmen ihr globales Minimum sowie Maximum an.

f) Sei $\mathbf{x}$ ein Eigenvektor von $\mathbf{A}$. Es gilt:

$$\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} = \mathbf{x}^\mathsf{T}\lambda\mathbf{x} = \lambda\mathbf{x}^\mathsf{T}\mathbf{x} = \lambda$$

Da wir unter den kritischen Punkten den Eigenwert maximieren, ist die Lösung zu unserem Problem der Eigenvektor $\mathbf{v}_1$ zum größten Eigenwert von $\mathbf{A}$.

Lösung Ende