Prof. Marc Alexa, Ugo Finnendahl,
Dimitrios Bogiokas, Gabriel Dogadov

Wissenschaftliches Rechnen - Übung 6.2

Optimierung

05.02.2024 bis 09.02.2024

Aufgabe 1: Iterative Optimierung

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

Lösung

Der Iterationsschritt lautet wie folgt:

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

Lösung Ende

2. Erklären Sie, warum der Iterationsschritt aus numerischen Gründen ungeeignet ist. Schreiben Sie den Iterationsschritt um, sodass dieser numerisch stabiler wird.

Lösung

Die Hesse-Matrix muss invertiert werden, was aus numerischen Gründen vermieden werden sollte. Stattdessen können wir folgendes lineare Gleichungssystem lösen:

$$\begin{aligned}\mathbf{x}_{t+1} = \mathbf{x}_t - \mathbf{H}_f^{-1}(\mathbf{x}_t)\nabla f(\mathbf{x}_t) &\Leftrightarrow \mathbf{x}_t - \mathbf{x}_{t+1} = \mathbf{H}_f^{-1}(\mathbf{x}_t)\nabla f(\mathbf{x}_t)\\ &\Leftrightarrow \mathbf{H}_f(\mathbf{x}_t)(\mathbf{x}_{t+1} - \mathbf{x}_t) = -\nabla f(\mathbf{x}_t)\\ &\Leftrightarrow \mathbf{H}_f(\mathbf{x}_t)\Delta\mathbf{x} = -\nabla f(\mathbf{x}_t)\end{aligned}$$

wobei $\Delta\mathbf{x} = \mathbf{x}_{t+1} - \mathbf{x}_t$ die Differenz zwischen zwei konsekutiven Schritten ist. Das lineare Gleichungssystem kann zum Beispiel mit der Gauß-Elimination oder einer geeigneten Matrixzerlegung gelöst werden (da die Hesse-Matrix symmetrisch ist, falls f zweimal stetig differenzierbar ist, kann eine LDL-Zerlegung bzw. eine Cholesky-Zerlegung verwendet werden, falls $\mathbf{H}_f$ zusätzlich PD ist).

Lösung Ende

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

Lösung

Man berechnet implizit das Taylorpolynom 2. Ordnung von f am Punkt $\mathbf{x}_t$, approximiert f also durch das quadratische Polynom

$$T_2(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_t)^\mathsf{T}\mathbf{H}_f(\mathbf{x}_t)(\mathbf{x} - \mathbf{x}_t) + (\mathbf{x} - \mathbf{x}_t)^\mathsf{T}\nabla f(\mathbf{x}_t) + f(\mathbf{x}_t),$$

und der neue Punkt $\mathbf{x}_{t+1}$ ist das Extremum bzw. der Sattelpunkt von T_2. Dieser ist die Nullstelle des Gradienten von T_2:

$$\begin{aligned}\nabla T_2(\mathbf{x}_{t+1}) = \mathbf{H}_f(\mathbf{x}_t)(\mathbf{x}_{t+1} - \mathbf{x}_t) + \nabla f(\mathbf{x}_t) &\overset{!}{=} \mathbf{0}\\ \Leftrightarrow \mathbf{H}_f(\mathbf{x}_t)(\mathbf{x}_{t+1} - \mathbf{x}_t) &= -\nabla f(\mathbf{x}_t)\\ \Leftrightarrow \mathbf{x}_{t+1} - \mathbf{x}_t &= -\mathbf{H}_f^{-1}(\mathbf{x}_t)\nabla f(\mathbf{x}_t)\\ \Leftrightarrow \mathbf{x}_{t+1} &= \mathbf{x}_t - \mathbf{H}_f^{-1}(\mathbf{x}_t)\nabla f(\mathbf{x}_t).\end{aligned}$$

Das Taylorpolynom 2. Ordnung hat genau dann einen und nur einen kritischen Punkt, falls die Hesse-Matrix regulär ist.

Lösung Ende

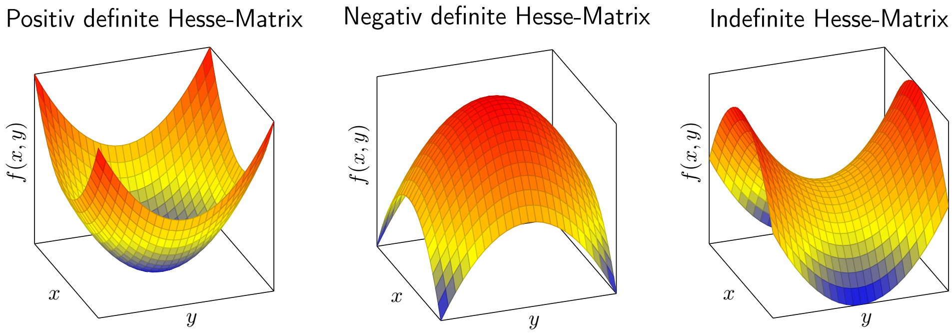
4. Wie wirkt sich die Definitheit der Hesse-Matrix auf den Iterationsschritt des Verfahrens aus?

Lösung

Zunächst stellen wir fest, dass die Hesse-Matrix der zu optimierenden Funktion f am Punkt $\mathbf{x}_t$ mit der Hesse-Matrix des Taylorpolynoms 2. Ordnung an jenem Entwicklungspunkt übereinstimmt, wobei das Taylorpolynom (an einem festen Entwicklungspunkt), anders als f, notwendigerweise eine konstante Hesse-Matrix hat. Die Hesse-Matrix gibt Auskunft über das Krümmungsverhalten der Funktion und die Art ihrer kritischen Punkte:

- Positiv definite Hesse-Matrix: Die Taylorentwicklung 2. Ordnung ist (strikt) konvex und der neue Punkt $\mathbf{x}_{t+1}$ ist dessen (eindeutige) Minimum.
- Negativ definite Hesse-Matrix: Die Taylorentwicklung 2. Ordnung ist (strikt) konkav und der neue Punkt $\mathbf{x}_{t+1}$ ist dessen (eindeutige) Maximum.
- Indefinite Hesse-Matrix: Die Taylorentwicklung 2. Ordnung ist weder konvex noch konkav und der neue Punkt $\mathbf{x}_{t+1}$ ist dessen (eindeutiger) Sattelpunkt.

Im Folgenden sind drei quadratische Polynome $\mathbb{R}^2 \to \mathbb{R}$ abgebildet, deren Hesse-Matrix unterschiedliche Definitheit hat.



Lösung Ende

5. Gegeben sei die Funktion $f(x, y) = -3x^2 + 2xy + y^2 + 3x - y + 1$. Berechnen Sie alle kritischen Punkte von f und klassifizieren Sie diese (Minimum, Maximum, Sattelpunkt). Wie viele Iterationsschritte benötigt das Newton-Verfahren mit Startpunkt $(13, 3)^\mathsf{T}$, um zu einem kritischen Punkt zu konvergieren?

Lösung

Der Gradient von f ist $\nabla f(x, y) = (-6x + 2y + 3, 2x + 2y - 1)^\mathsf{T}$. Dessen einzige Nullstelle ist $(x^*, y^*) = (\frac{1}{2}, 0)$. Die Hesse-Matrix von f ist die konstante Matrix

$$\mathbf{H}_f(x, y) = \begin{bmatrix} -6 & 2 \\ 2 & 2 \end{bmatrix}$$

und man sieht an ihrer Diagonale, dass sie indefinit ist. Somit ist $(\frac{1}{2}, 0)^\mathsf{T}$ ein Sattelpunkt. Da f eine quadratische Funktion ist (was man u.a. an der konstanten Hesse-Matrix sieht), ist sie ihre eigene Taylorentwicklung 2. Ordnung (egal, an welchem Entwicklungspunkt sie gebildet wird) und das Newton-Verfahren findet den Sattelpunkt nach nur einer einzigen Iteration. Dies gilt für alle quadratischen Funktionen: Das Newton-Verfahren (zum Optimieren) approximiert die Funktion (lokal) durch eine quadratische Funktion, und falls f bereits ein quadratisches Polynom ist, wird ihr Optimum nach einer einzigen Iteration gefunden.

Lösung Ende

6. Wie lautet der Iterationsschritt des Gradientenabstiegsverfahrens (Gradient Descent)? Worin unterscheidet es sich vom Newton-Verfahren? Welche wichtige Funktion erfüllt die Schrittweite?

Lösung

Der Iterationsschritt lautet wie folgt:

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

In jeder Iteration wird der Gradient am aktuellen Punkt mit einer (womöglich adaptiven) Schrittweite $\alpha_t \in \mathbb{R}^+$ skaliert vom aktuellen Punkt abgezogen (daher der Name Gradientenabstieg).

Es bestehen folgende Unterschiede zum Newton-Verfahren:

- Das Gradientenverfahren nutzt die Eigenschaft des Gradienten, dass er in die Richtung des stärksten Anstiegs zeigt. Indem das Verfahren in die entgegengesetzte Richtung geht, zielt es darauf ab, die Funktion zu minimieren. Das Newton-Verfahren, hingegen, sucht lediglich eine Nullstelle des Gradienten. Das Gradientenverfahren muss aber nicht unbedingt in einem Minimum oder überhaupt konvergieren.
- Das Gradientenverfahren skaliert den Gradienten mit einer Schrittweite. Die Bestimmung dessen ist eine zentrale Herausforderung der Methode. Das Newton-Verfahren dagegen transformiert den Gradienten mit der Inversen der Hesse-Matrix, sodass f lokal durch eine quadratische Funktion approximiert wird. Das Gradientenverfahren approximiert f dagegen durch eine lineare Funktion.

Die Schrittweite bestimmt im Wesentlichen das Konvergenzverhalten des Verfahrens. Den Gradienten ohne Skalierungsfaktor zu verwenden ist meist problematisch, da der Gradient lediglich in die Richtung des stärksten Anstiegs zeigt, dessen entgegengesetzte Richtung nicht die Richtung eines Minimums sein muss. Des Weiteren entspricht die Länge des Gradienten nicht dem Abstand zum Minimum.

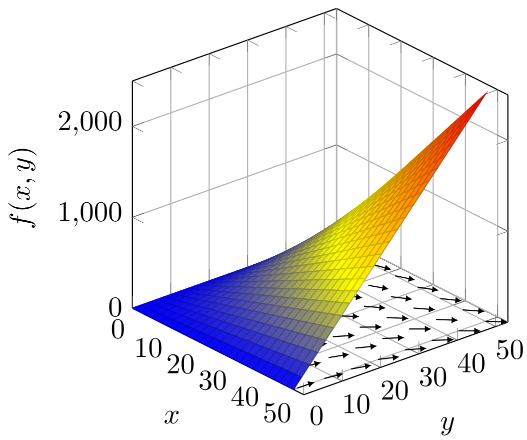
Lösung Ende

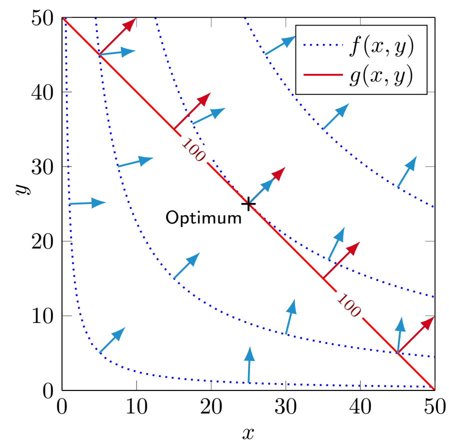
Aufgabe 2: Optimierung mit Nebenbedingungen

1. Bauer Herrmann hat einen 100 m langen Metalldraht, welchen er zu einem rechteckigen Zaun, mit Seitenlängen x und y, spannen möchte. Da er ein BIO-Bauer ist, möchte er seinen 300 Rindern den größtmöglichen Platz zur Verfügung stellen, sodass der Flächeninhalt des Zauns maximiert werden soll. Das kann man als folgendes Optimierungsproblem schreiben:

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

Die Funktion $f(x, y) = xy$ beschreibt die Fläche und die Funktion $g(x, y) = 2x + 2x$ beschreibt den Umfang des rechteckigen Stallzauns. Da wir negative Seitenlängen ausschließen, also $x, y \geq 0$ annehmen, lässt sich das Optimierungsproblem wie folgt visualisieren.





a) Lösen Sie das Problem ohne die Verwendung von Lagrange-Multiplikatoren. Stellen Sie dafür die Nebenbedingung nach y um und setzen Sie das in die Zielfunktion ein.

b) Lösen Sie das Problem mithilfe von Lagrange-Multiplikatoren. Stellen Sie dazu eine mögliche Lagrange-Funktion $\mathcal{L}(x, y, \lambda)$ auf und leiten Sie die notwendige Bedingung für einen kritischen Punkt unter der Nebenbedingung her. Berechnen Sie dann ihre kritischen Punkte. Wodurch kann man den gesuchten Punkt in der Visualisierung erkennen?

Lösung

a) Aus der Nebenbedingung erhalten wir $y = 50 - x$. Das setzen wir in die Zielfunktion ein und erhalten die neue eindimensionale Zielfunktion $f(x) = x(50-x) = 50x - x^2$, welche zu maximieren ist. Deren Ableitung ist $f'(x) = 50 - 2x$ und dessen Nullstelle ist bei $x = 25$. Die zweite Ableitung ist konstant $f''(x) = -2$, sodass bei $x = 25$ tatsächlich ein lokales Maximum vorhanden ist. Aus der Nebenbedingung folgt, dass $y = 50 - x = 50 - 25 = 25$ und die Fläche des quadratischen Zaunes, also der Wert der Zielfunktion am Optimum, ist $625\ \mathrm{m}^2$.

   Die Methode funktioniert nur so gut, wenn die Nebenbedingung linear ist. Bei schwierigeren Nebenbedingungen müssen viele Fallunterscheidungen gemacht werden.

b) Wir stellen die Lagrange-Funktion wie folgt auf:

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

   Setzen wir ihren Gradienten gleich dem Nullvektor, erhalten wir die notwendigen Bedingungen für einen kritischen Punkt

$$\nabla \mathcal{L}(x, y, \lambda) = \begin{bmatrix} y - 2\lambda \\ x - 2\lambda \\ -2x - 2y + 100 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

wobei die letzte Bedingung einfach die Nebenbedingung ist und die ersten beiden fordern, dass am kritischen Punkt die Gradienten von f und g linear abhängig sind. In diesem Fall führt das zu einem linearen Gleichungssystem (was im Allgemeinen nicht der Fall sein muss!):

$$\begin{bmatrix} 0 & 1 & -2 \\ 1 & 0 & -2 \\ -2 & -2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -100 \end{bmatrix}$$

Dessen (einzige) Lösung ist $(x = 25, y = 25, \lambda = \frac{25}{2})$. Somit ist die Lösung auch hier ein quadratischer Zaun mit $25\ \mathrm{m}$ Seitenlänge und $625\ \mathrm{m}^2$ Fläche.

Lösung Ende

2. Gegeben ist die Gerade g in Parameterform

$$g(t) = \mathbf{c} + t\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

welche als Schnitt von zwei Ebenen $y = 1$ und $x - z = 0$ dargestellt werden kann.

a) Bestimmen Sie den Punkt $\mathbf{x}^* \in \mathbb{R}^3$ auf der Geraden g, der am nächsten zum Punkt $(-1, -1, -1)^\mathsf{T}$ ist, mithilfe von Lagrange-Multiplikatoren.

b) Nennen Sie einen alternativen Lösungsansatz, um das Problem zu lösen.

Lösung

a) Wir stellen eine Lagrange-Funktion auf:

$$\begin{aligned} \mathcal{L}(x, y, z, \lambda, \alpha) &= \|(x, y, z)^\mathsf{T} - (-1, -1, -1)^\mathsf{T}\|^2 - \lambda(y - 1) - \alpha(x - z) \\ &= (x + 1)^2 + (y + 1)^2 + (z + 1)^2 - \lambda(y - 1) - \alpha(x - z) \end{aligned}$$

Dessen Gradient ist

$$\nabla\mathcal{L}(x, y, z, \lambda, \alpha) = \begin{bmatrix} 2(x + 1) - \alpha \\ 2(y + 1) - \lambda \\ 2(z + 1) + \alpha \\ -y + 1 \\ -x + z \end{bmatrix},$$

und falls man diesen gleich dem Nullvektor setzt, erhält man erneut ein lineares Gleichungssystem. Aus der vierten Gleichung (also der ersten Nebenbedingung) folgt sofort $y = 1$. Aus der letzten Gleichung folgt unmittelbar $x = z$. Addiert man die erste und dritte Gleichung und setzt $x = z$, erhält man $4(z + 1) = 0$, also $z = -1$ und damit auch $x = -1$. Somit ist $(-1, 1, -1)^\mathsf{T}$ der zu $(-1, -1, -1)^\mathsf{T}$ nächste Punkt auf der Geraden g.

b) Bei dem Problem handelt es sich um ein lineares Ausgleichsproblem, welches mithilfe der Normalengleichung gelöst werden kann.

Lösung Ende