Optimierung unter Nebenbedingungen

- Kapitel 7: Optimierung unter Nebenbedingungen
- Informationen zur Vorlesung:
  http://www.mathematik.uni-trier.de/∼wengenroth/

7.1 Bemerkung

In der Praxis hat man fast nie Optimierungsprobleme für Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ oder $f : A \to \mathbb{R}$ mit offenen Mengen $A \subseteq \mathbb{R}^n$ (das heißt, der Rand gehört nicht dazu), für die man die Sätze 5.4 und 6.6 benutzen kann:

- $[x_1, \ldots, x_n] \in A$ beschreibe Mengen x_k von Gütern g_k.
- $f : A \to \mathbb{R}$ beschreibe den Nutzen der Portfolios $x \in A$.
- Ökonomisches Grundprinzip: Mehr nützt mehr.
- Aber: Geld ist knapp.
- Sind $p_1, \ldots, p_n$ die Preise der Güter und $c \geq 0$ das verfügbare Budget, so kann man ein Portfolio $x = [x_1, \ldots, x_n]$ nur dann realisieren, wenn

$$\Phi(x) = x_1 p_1 + \ldots + x_n p_n \leq c$$

- Man sucht also maximalen Nutzen unter der Budgetrestriktion

$$\max\{f(x) : x \in A \text{ und } \Phi(x) \leq c\}.$$

- Das ökonomische Grundprinzip impliziert, dass ein optimales Portfolio das Budget ausschöpft. Also sucht man

$$\max\{f(x) : x \in A \text{ und } \Phi(x) = c\}.$$

7.2 Definition

(a) Ein Maximierungsproblem unter Gleichheitsnebenbedingungen $\max\{f : \Phi = c\}$ ist gegeben durch

- $f : A \to \mathbb{R}$ mit einer offenen Menge $A \subseteq \mathbb{R}^n$ und einer stetig differenzierbaren Funktion f,
- $\Phi : A \to \mathbb{R}^m$ stetig differenzierbar,
- $c \in \mathbb{R}^m$.

(b) $x^* \in A$ heißt eine Lösung von Max $\{f, \Phi = c\}$, falls

- $\Phi(x^*) = c$ und
- $f(x^*) \geq f(x)$ für alle $x \in A$ mit $\Phi(x) = c$.

(c) Genauso kann man Minimierungsprobleme definieren. Wegen

$$f(x^*) \leq f(x) \Longleftrightarrow -f(x^*) \geq -f(x)$$

ist dies aber dasselbe wie $\max\{-f : \Phi = c\}$.

(d) Geometrisches Beispiel: Ein Unternehmen produziert Getränkedosen und erhält den Auftrag, eine Ein-Liter-Dose mit möglichst wenig Blech herzustellen.

- $A = \{[r, h] \in \mathbb{R}^2 : r > 0, h > 0\}$ (r = Radius, h = Höhe)
- $f(r, h) = \pi r^2 + \pi r^2 + 2\pi rh$ (die ersten zwei Summanden sind die Flächeninhalte von Boden und Deckel, der dritte ist die Fläche des Mantels).
- $\Phi(r, h) =$ „Grundfläche $\times$ Höhe“ $= \pi r^2 h$ ist das Volumen der Dose (in cm^3, falls r und h in cm).
- Das Problem ist also $\min\{f : \Phi = 1000\}$.

7.3 Auflösemethode im Beispiel

(a) In obigem Beispiel suchen wir

$$\min\{2\pi r^2 + 2\pi rh : \pi r^2 h = 1000\}$$

Die Idee ist sehr einfach

(1) Löse die Nebenbedingung $\Phi(r, h) = 1000$ nach einer der Variablen auf
(2) Setze die Auflösung in Zielfunktion ein
(3) Optimiere ohne Nebenbedingung

(b) (1) Auflösen: $h(r) = \frac{1000}{\pi r^2}$
(2) Einsetzen: $g(r) = f(r, h(r)) = 2\pi r^2 + \frac{2000}{r}$
(3) Optimieren: $g'(r) = 4\pi r - 2000/r^2 = 0$ für $r = \sqrt[3]{500/\pi} \approx 5,419$.

(c) Die zugehörige Höhe ist $h(r) = \frac{1000}{\pi r^3} \cdot r = 2r$. Für die günstigste Dose ist also die Höhe gleich dem Durchmesser der Grundfläche.
Übliche Getränkedosen sind also Materialverschwendung!

7.4 Die Auflösemethode

Gegeben sei ein Maximierungsproblem $\max\{f : \Phi = c\}$ mit $f : A \to \mathbb{R}, A \subseteq \mathbb{R}^n$ und $\Phi : A \to \mathbb{R}^m$ mit $m < n$. Die Idee ist wie in 7.3:

- Schreibe die Elemente von A als $\begin{bmatrix} x \\ y \end{bmatrix}$ mit $x \in \mathbb{R}^{n-m}$ und $y \in \mathbb{R}^m$.
- Löse die Gleichung $\Phi(x, y) = c$ nach y auf.
- Setze $y = y(x)$ in Zielfunktion ein und maximiere.

7.5 Satz über die Auflösemethode

Seien $A \subseteq \mathbb{R}^n$ offen, $f : A \to \mathbb{R}$ und $\Phi : A \to \mathbb{R}^m$ beide stetig differenzierbar mit $m < n$ sowie $c \in \mathbb{R}^m$.
Es gebe $B \subseteq \mathbb{R}^{n-m}$ offen und $g : B \to \mathbb{R}^m$ stetig differenzierbar, so dass für $x \in \mathbb{R}^{n-m}$ und $y \in \mathbb{R}^m$ gilt

$$\begin{bmatrix} x \\ y \end{bmatrix} \in A \text{ und } \Phi(x, y) = c \iff x \in B \text{ und } y = g(x).$$

Dann ist $\begin{bmatrix} x^* \\ y^* \end{bmatrix}$ genau dann eine Lösung von $\max\{f : \Phi = c\}$, wenn $y^* = g(x^*)$ und $f(x^*, g(x^*)) = \max\{f(x, g(x)) : x \in B\}$.

7.6. Beispiel

Wir suchen $\min\{x^2 + y^2 + z^2 : x + y + z = 1\}$.

- $n = 3, m = 1, A = \mathbb{R}^3, B = \mathbb{R}^2$
- $g(x, y) = 1 - (x + y)$
- $h(x, y) = x^2 + y^2 + (1 - (x + y))^2$
- $D_1 h(x, y) = 2x - 2(1 - (x + y)) = 4x + 2y - 2$
  $D_2 h(x, y) = 2y - 2(1 - (x + y)) = 4y + 2x - 2$
- Also

$$\nabla h(x, y) = 0 \iff \begin{array}{l} 2x + y = 1 \\ x + 2y = 1 \end{array} \iff x = y = 1/3$$

  Dies ist tatsächlich der Vektor in $\mathbb{R}^2$ mit minimalem Wert für h.
- $z = g\left(\frac{1}{3}, \frac{1}{3}\right) = \frac{1}{3}$ und das gesuchte Minimum ist $f\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \frac{1}{3}$.

7.7 Schwierigkeiten der Auflösemethode

- Ist $\Phi(x, y) = c$ überhaupt auflösbar?
- Kann man die Auflösung explizit berechnen?
- Ist die Auflösung differenzierbar?

Zum Beispiel lässt sich $x^2 + y^2 = 1$ nicht eindeutig nach y auflösen.

7.8 Bemerkung

(a) Nach dem Einsetzen muss man $h(x) = f(x, g(x))$ maximieren, also (im Fall $n = 2$ und $m = 1$) die Gleichung $h'(x) = 0$ lösen. Wir schreiben $h = f \circ j$ mit $j : \mathbb{R} \to \mathbb{R}^2, x \mapsto \begin{bmatrix} x \\ g(x) \end{bmatrix}$. Dann gilt $j'(x) = \begin{bmatrix} 1 \\ g'(x) \end{bmatrix}$, und die Kettenregel liefert

$$h'(x) = \nabla f(j(x)) \cdot j'(x) = D_1 f(x, g(x)) + D_2 f(x, g(x)) g'(x).$$

(b) Wir brauchen also $g'(x)$. Dazu benutzen wir, dass $\Phi(x, g(x)) = (\Phi \circ j)(x)$ konstant ist und deshalb Ableitung 0 hat. Wieder mit der Kettenregel folgt

$$0 = (\Phi \circ j)'(x) = D_1 \Phi(x, g(x)) + D_2 \Phi(x, g(x)) g'(x).$$

(c) Falls $D_2\Phi(x, g(x)) \neq 0$ erhalten wir also $g'(x) = -\frac{D_1\Phi(x,y)}{D_2\Phi(x,y)}$ und können dies in die Gleichung $h'(x) = 0$ einsetzen, das heißt, falls die „Auflösefunktion" g stetig differenzierbar ist, erhalten wir als notwendige Bedingung, dass $[x, y]$ eine Lösung von $\max\{f : \Phi = c\}$ ist:

$$\boxed{D_1 f(x, y) D_2\Phi(x, y) = D_2 f(x, y) D_1\Phi(x, y)}$$

(c) Für $\lambda = \frac{D_2 f(x,y)}{D_2\Phi(x,y)}$ folgt $\boxed{\nabla f(x, y) = \lambda \nabla\Phi(x, y)}$, d. h., die Gradienten zeigen in die gleiche Richtung!

7.9 Satz (Lagrange-Methode)

Seien $A \subseteq \mathbb{R}^n$ offen, $f : A \to \mathbb{R}$ und $\Phi : A \to \mathbb{R}^m$ beide stetig differenzierbar mit $m < n$ sowie $c \in \mathbb{R}^m$.
Ist $x \in A$ eine Lösung von $\max\{f : \Phi = c\}$, so dass $\nabla\Phi_1(x), \ldots, \nabla\Phi_m(x)$ linear unabhängig sind, dann gibt es $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ mit $\nabla f(x) = \lambda_1 \nabla\Phi_1(x) + \ldots + \lambda_m \nabla\Phi_m(x)$.
Bemerkung:

- Die Zahlen $\lambda_1, \ldots, \lambda_m$ sind eindeutig und heißen Lagrange-Multiplikatoren oder Schattenpreise.
- Die lineare Unabhängigkeit der Gradienten bedeutet, dass der Nullvektor nur auf die triviale Art als Linearkombination dargestellt werden kann, also

  $$\alpha_1 \nabla\Phi_1(x) + \ldots + \alpha_m \nabla\Phi_m(x) = 0 \Longrightarrow \alpha_1 = \ldots = \alpha_m = 0$$

  Im Fall $m = 1$ bedeutet dies $\nabla\Phi(x) \neq 0$

- Die Schwierigkeit im Beweis ist die Auflösbarkeit von $\Phi(x) = c$ (wenigstens in der Nähe der Lösung x) zu zeigen.
- Die Bedingungen $\nabla f(x) = \lambda_1 \nabla \Phi_1(x) + \ldots + \lambda_m \nabla \Phi_m(x)$ und $\Phi(x) = c$ liefern ein Gleichungssystem mit $n + m$ Gleichungen

$$\begin{array}{rcl} D_1 f(x) & = & \lambda_1 D_1 \Phi_1(x) + \ldots + \lambda_m D_1 \Phi_m(x) \\ & \vdots & \\ D_n f(x) & = & \lambda_1 D_n \Phi_1(x) + \ldots + \lambda_m D_n \Phi_m(x) \\ \Phi_1(x) & = & c_1 \\ & \vdots & \\ \Phi_m(x) & = & c_m \end{array}$$

für die $n + m$ Unbekannten $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m$. Dieses System ist sehr oft nicht linear und nicht leicht zu lösen.

7.10 Noch einmal Beispiel 7.6

$\min\{x^2 + y^2 + z^2 : x + y + z = 1\}$

- $\nabla f(x, y, z) = [2x, 2y, 2z]$, $\nabla\Phi(x, y, z) = [1, 1, 1]$
- Lagrange-Gleichung $\nabla f(x, y, z) = \lambda\nabla\Phi(x, y, z)$ impliziert also $x = y = z$, und $\Phi(x, y, z) = 1$ liefert $x = y = z = 1/3$.

7.11 Ökonomische Interpretation der Lagrange-Multiplikatoren

(a) Das Verhältnis von $D_j f(x)$ zu $D_j \Phi_k(x)$ heißt Schattenpreis des j-ten Guts bezüglich der k-ten Nebenbedingung im Punkt x:

$$\frac{D_j f(x)}{D_j \Phi_k(x)} \approx \frac{f(x + te^j) - f(x)}{\Phi_k(x + te^j) - \Phi_k(x)} = \frac{\text{Zusatznutzen bei etwas mehr von } x_j}{\text{Zusatzkosten bei etwas mehr von } x_j}$$

(Ist $m = 1$, interpretiert man $\Phi(x)$ als Kosten oder Preis, ist $m > 1$, so interpretiert man die $\Phi_1(x), \ldots, \Phi_m(x)$ als verschiedene Kostenarten).

(b) Im Fall $m = 1$ gilt im Lösungspunkt

$$\boxed{\text{Die Schattenpreise aller Güter sind gleich}}$$

nämlich gleich dem Lagrange-Multiplikator.

7.12 Geometrische Interpretation

(a) Die Bedingung $\Phi = c$ beschreibt eine $(n-m)$-dimensionale Fläche im $\mathbb{R}^n$:

- $n = 3, m = 1$ Fläche wie z.B. Kugeloberfläche $x^2 + y^2 + z^2 = 1$.
- $n = 3, m = 2$ Kurve wie z. B. Durchschnitt der Kugeln mit Radius 2 und Mittelpunktion $[0,0,0]$ und $[0,0,1]$: $x^2 + y^2 + z^2 = 1$ und $x^2 + y^2 + (z-1)^2 = 1$.

(b) Der Gradient der Zielfunktion steht im Optimum senkrecht auf der Fläche $\{\Phi = c\}$. Ist nämlich $\varphi : I \to \mathbb{R}^3$ der Weg eines Wagens, der durch $\{\Phi = c\}$ fährt mit $\varphi(0) = x$, so ist $\varphi'(0)$ ein Tangentenvektor in x an die Fläche.
Außerdem ist $f(\varphi(t))$ für $t = 0$ maximal und daher gilt $(f \circ \varphi)'(0) = 0$. Die Kettenregel liefert

$$\nabla f(x) \cdot \varphi'(0) = 0, \text{ also } \langle \nabla f(x), \varphi'(0) \rangle = 0.$$

(c) Für alle $k \in \{1, \ldots, m\}$ ist $\Phi_k \circ \varphi$ konstant $= c_k$, also gilt

$$0 = (\Phi_k \circ \varphi)'(0) = \nabla \Phi_k(x) \cdot \varphi'(0), \text{ also } \langle \nabla \Phi_k(x), \varphi'(0) \rangle = 0.$$

Die Gradienten stehen also ebenfalls senkrecht auf der Fläche.

(d) Die lineare Unabhängigkeit der Gradienten besagt, dass sie eine Basis des Orthogonalraums bilden, d.h. jeder zu $\{\Phi = c\}$ in x orthogonale Vektor ist eine Linearkombination.