Vorlesungsskript Optimierung
(Vorläufige Fassung)

Moritz Diehl und Florian Messerer
Institut für Mikrosystemtechnik (IMTEK) und Mathematisches Institut,
Universität Freiburg
moritz.diehl@imtek.uni-freiburg.de

5. November 2026

Vorwort

Dieses Manuskript wird im Wintersemester 2026/27 zeitgleich mit der Bachelorvorlesung Optimierung für Studierende der Informatik und des Embedded Systems Engineering an der Universität Freiburg geschrieben. Es orientiert sich eins-zu-eins an den Inhalten der Tafelvorlesung und ist deshalb bewusst knapp gehalten. Für eine vertiefende Lektüre (auf Englisch) empfehlen wir die Lehrbücher [1, 3] oder das Manuskript [2]. Fehler werden unvermeidlich auftreten, und wir bitten, Feedback mit Verbesserungsvorschlägen an `moritz.diehl@imtek.uni-freiburg.de` und `florian.messerer@imtek.uni-freiburg.de` zu senden.

Moritz Diehl und Florian Messerer,
Freiburg,
5. November 2026

Inhaltsverzeichnis

Vorwort ii

1 Willkommen und Einführung **1**
1.1 Einführung 1
1.2 Überblick und Beispiele 4

Kapitel 1

Willkommen und Einführung

1.1 Einführung

Die mathematische Optimierung betrachtet Optimierungsprobleme und ihre Lösung mit Algorithmen. Ein *Standard-Optimierungsproblem* wird wie folgt geschrieben:

$$\min_{x \in \Omega} \quad f(x) \tag{1.1}$$

Hierbei ist $f : \Omega \to \mathbb{R}, x \mapsto f(x)$ die *Zielfunktion*, $\Omega \subset \mathbb{R}^n$ ist die *zulässige Menge*, und $n \in \mathbb{N}$ die Dimension der *Optimierungsvariablen* $x \in \mathbb{R}^n$. Diese Dimension kann sehr groß werden, z.B. $n = 10^{10}$ beim Training von Sprachmodellen (wo der Vektor x die Werte der sogenannten Gewichte in den neuronalen Netzen zusammenfasst).

Hat man mehrere Funktionen, die zugleich minimiert werden sollen, dann hat man ein Problem der mehrkriteriellen Optimierung (auch Mehrzieloptimierung oder Pareto-Optimierung genannt), das mit verschiedenen Methoden in ein Standard-Optimierungsproblem umgewandelt werden kann. Am einfachsten geschieht dies durch eine gewichtete Summe: Bei zwei Funktionen f_1 und f_2 entscheidet man sich für eine Zahl $\alpha \in [0, 1]$ und definiert die Standard-Zielfunktion als $f(x) := \alpha f_1(x) + (1 - \alpha) f_2(x)$.

Beispiel 1.1. Wir betrachten die Zielfunktion $f(x) = x + x^6$ mit der Dimension $n = 1$ und ohne Beschränkungen, also mit der unbeschränkten zulässigen Menge $\Omega = \mathbb{R}$. Die Funktion und ihre Ableitung sind in Abbildung 1.1 gezeigt. Ein Minimum ist an einer Stelle x^* zu finden, die $f'(x^*) = 0$ erfüllt, also $1 + 6(x^*)^5 = 0$. Die eindeutige Lösung dieser Gleichung ist $x^* = -\frac{1}{\sqrt[5]{6}}$. Dieser Punkt ist nicht nur ein lokaler, sondern sogar ein *globaler Minimierer*.

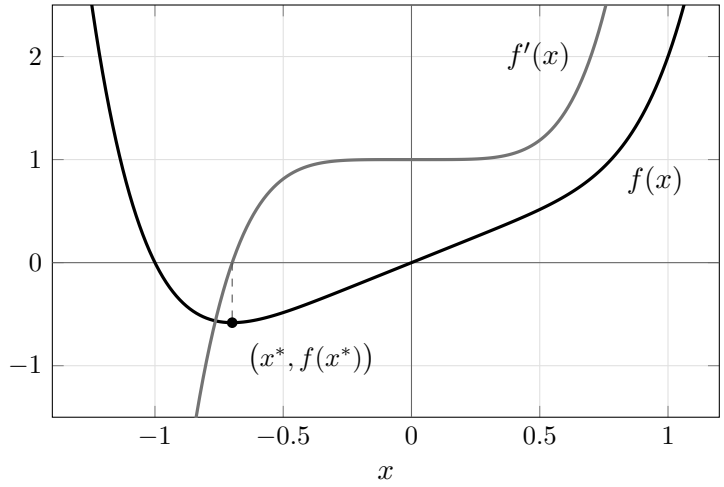


Abbildung 1.1: Die Funktion $f(x) = x + x^6$ und ihre Ableitung $f'(x) = 1 + 6x^5$.

Definition 1.1. Ein Punkt $x^* \in \mathbb{R}^n$ ist ein *globaler Minimierer* des Optimierungsproblems (1.1) genau dann, wenn (a) es ein zulässiger Punkt ist, also

$$x^* \in \Omega,$$

und (b) alle anderen zulässigen Punkte einen höheren oder gleichen Zielfunktionswert haben, also

$$\forall x \in \Omega : f(x) \geq f(x^*).$$

Die Menge aller globalen Minimierer (also die "Argumente", die f auf Ω minimieren) bezeichnet man mit $\arg\min_{x\in\Omega} f(x)$. Für jeden globalen Minimierer gilt also $x^* \in \arg\min_{x\in\Omega} f(x)$. Den minimalen Funktionswert, der für jeden globalen Minimierer gleich ist, bezeichnen wir als das globale Minimum, und sein Wert wird mit $\min_{x\in\Omega} f(x)$ bezeichnet.

Beispiel 1.2 (Ganzzahlige Optimierung)**.** Wir betrachten eine zulässige Menge in einer Dimension $n = 1$, die nur aus drei Werten besteht, $\Omega = \{1, 2, 3\} \subset \mathbb{R}$. Da die Optimierungsvariable nur ganzzahlige Werte annehmen kann, spricht man von *ganzzahliger Optimierung* (integer optimization). Die Zielfunktion $f : \Omega \to \mathbb{R}$ ist definiert durch die drei Zuweisungen $f(1) = 3.5$, $f(2) = 1.4$ und $f(3) = 1.4$, siehe Abbildung 1.2. Hier gilt $\{2, 3\} = \arg\min_{x\in\Omega} f(x)$. Es gibt also zwei globale Minimierer. Der minimale Funktionswert ist dagegen eindeutig: Er beträgt 1.4, und man schreibt dies kompakt als $1.4 = \min_{x\in\Omega} f(x)$.

Beispiel 1.3 (Optimierung auf einem Ball)**.** Wir betrachten ein zweidimensionales Optimierungsproblem mit $\Omega \subset \mathbb{R}^2$. Die Zielfunktion ist $f(x) = x_2$. Die zulässige Menge Ω ist gleich einem zweidimensionalen euklidischen Ball $B_r^n(\bar{x})$ mit Radius $r = 2$ um den Punkt $\bar{x} = [0, 0]^\top$.

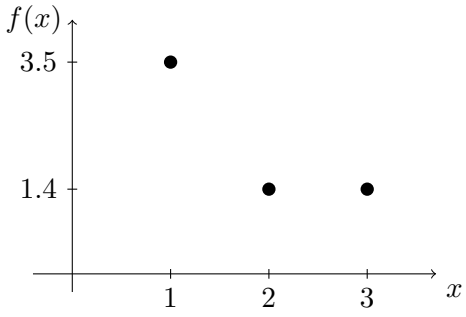


Abbildung 1.2: Ganzzahliges Optimierungsproblem mit $\Omega = \{1, 2, 3\}$.

Definition 1.2 (Euklidische Norm)**.** Die euklidische Norm $\|x\|_2$ eines Vektors $x \in \mathbb{R}^n$ ist gegeben durch

$$\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2}. \tag{1.2}$$

Definition 1.3 (Euklidischer Ball)**.** Die Menge aller Punkte, die einen euklidischen Abstand von maximal r zum Punkt $\bar{x} \in \mathbb{R}^n$ haben, bezeichnen wir als *euklidischen Ball in Dimension n*, $B_r^n(\bar{x})$, und definieren

$$B_r^n(\bar{x}) := \{x \in \mathbb{R}^n \mid \|x - \bar{x}\|_2 \leq r\}. \tag{1.3}$$

Beispiel 1.3 (Fortsetzung). Abbildung 1.3 zeigt die zulässige Menge $\Omega = B_2^2(0)$ und *Level-Sets* der Zielfunktion, also Mengen wie $\{x \mid f(x) = -1\}$, auf denen f konstant ist. Da f nach oben steigt, ist der Minimierer der unterste Punkt des Balls, $x^* = [0, -2]^\top$, mit dem minimalen Wert $f(x^*) = -2$. Kompakt geschrieben:

$$-2 = \min_{x \in B_2^2(0)} x_2 \quad \text{und} \quad \arg\min_{x \in \Omega} f(x) = \arg\min_{x \in B_2^2(0)} x_2 = \left\{ \begin{bmatrix} 0 \\ -2 \end{bmatrix} \right\}.$$

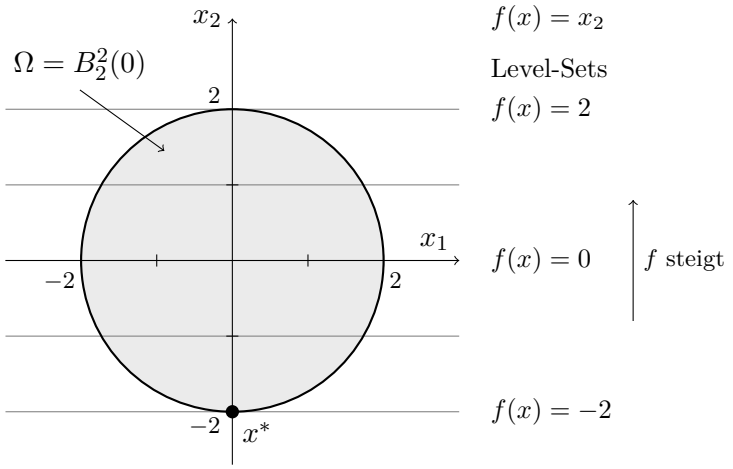


Abbildung 1.3: Optimierung auf einem Ball: zulässige Menge und Level-Sets der Zielfunktion.

1.2 Überblick und Beispiele

Im Folgenden wird ein kurzer Überblick gegeben über die Themen, mit denen wir uns im Verlauf dieses Kurses befassen.

Vorlesung 2: Newton-artige Verfahren für unbeschränkte Optimierung

Wir betrachten den unbeschränkten Fall $\Omega = \mathbb{R}^n$.

Idee: Löse $\nabla f(x) = 0$ iterativ mit Newton-Linearisierung.

Für $f : \mathbb{R}^n \to \mathbb{R},\ x \mapsto f(x)$ ist der Gradient die Funktion

$$\nabla f : \mathbb{R}^n \to \mathbb{R}^n, \quad x \mapsto \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}.$$

Beispiel 1.4. Für $x \in \mathbb{R}^2$ und $f(x) = x_1^2 + 3x_2^2$ ist

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 6x_2 \end{bmatrix}.$$

Vorlesung 3: Effiziente Ableitungsberechnung

Hier geht es um das *automatische Differenzieren* (AD), insbesondere um "Reverse AD", das auch als "Backpropagation" bekannt ist. Für den Rechenaufwand, der benötigt wird,

um den Gradienten der skalarwertigen Funktion $f : \mathbb{R}^n \to \mathbb{R}$ mit Reverse AD auszuwerten, gilt

$$\text{cost}(\nabla f(x)) \leq 3 \cdot \text{cost}(f).$$

Er ist also höchstens dreimal so groß wie für eine Auswertung von f selbst, unabhängig von der Dimension n. Das macht gradientenbasierte Verfahren auch bei sehr hochdimensionalen Problemen praktikabel, z.B. beim Training neuronaler Netze.

Vorlesung 4: Optimierung für Neural Network Training

Beispiel 1.5 (Supervised Learning)**.** Gegeben sind Datenpaare

$$\begin{bmatrix} a_1, & y_1 \\ a_2, & y_2 \\ \vdots & \vdots \\ a_N, & y_N \end{bmatrix}$$

mit gemessenen Outputs $y_k \in \mathbb{R}$ und Inputs $a_k \in A$ aus einer beliebigen Menge A. Wir möchten eine Funktion finden, die diese Daten möglichst gut beschreibt und auch für ein neues a das zugehörige y vorhersagt. Dazu verwenden wir ein Modell

$$\varphi(a, x) \approx y$$

mit den Gewichtsparametern $x \in \mathbb{R}^n$. Das Trainings-Optimierungsproblem lautet

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$

mit

$$f(x) = \frac{1}{N} \sum_{k=1}^{N} \left(y_k - \varphi(a_k, x)\right)^2 .$$

Vorlesung 5: Nichtlineare Optimierung mit Gleichungsbeschränkungen

Wir betrachten

$$\min_{x \in \Omega} \quad f(x)$$

mit $\Omega = \{x \in \mathbb{R}^n \mid c(x) = 0\}$ und $c : \mathbb{R}^n \to \mathbb{R}^m$. Eine zentrale Rolle spielt dabei die *Lagrange-Funktion*

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x).$$

Vorlesung 6: Optimierung für optimale Steuerung

Vorlesung 7: Nichtlineare Optimierung mit Ungleichungsbeschränkungen

Vorlesung 8: Lineare und quadratische Optimierung

Vorlesung 9: Optimierung auf Embedded Systems

Vorlesung 10: Kombinatorische Optimierung

Hier ist die zulässige Menge von der Form $\Omega \subset \mathbb{R}^n \times \mathbb{Z}^m$.

Literaturverzeichnis

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. University Press, Cambridge, 2004.

[2] Moritz Diehl. *Lecture Notes on Numerical Optimization*. 2016. (Available online: http://cdn.syscop.de/publications/Diehl2016.pdf).

[3] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2 edition, 2006.