

# Optimierung

## Inhaltsverzeichnis

Optimierung .....	1
Grundlagen .....	1
Direkte und Inverse Probleme .....	1
Fehler und Residuum .....	2
Minimalstellensuche .....	2
Abstiegsverfahren .....	3
Koerzitivität .....	3
Häufige Probleme .....	4
Lineare Gleichungssysteme .....	4
Konvexität .....	4
Einfache Minimalstellen .....	5
Verfahren .....	6
Gefälleverfahren .....	6
Konvergenzordnung .....	7
Allgemeine Verfahren: Strömungsmechanik .....	8

## Grundlagen

### Direkte und Inverse Probleme

In der Numerik gibt es zwei wichtige Arten von Problemen:  
Direkte Probleme und Inverse Probleme.

Bei den direkten Problemen soll  $p(x)$  berechnet werden, wobei  $p$  eine gegebene Funktion ist und  $x$  ein gegebener Wert ist. Die Existenz und Eindeutigkeit der Lösung ist in diesem Fall immer theoretisch gesichert.

Bei den inversen Problemen soll ein  $x$  gefunden werden, wenn ein  $p$  und  $y$  gegeben sind, sodass  $p(x) = y$  ist. Somit sind sie die Umkehrung der direkten Probleme. Im Allgemeinen sind inverse Probleme viel schwieriger und im Allgemeinen ist weder Existenz noch Eindeutigkeit theoretisch gesichert.

#### Beispiel

Sei  $p(x) = x^2$ . Gesucht ist ein  $x \in \mathbb{R}$  sodass  $p(x) = c$ . Für  $c = 0$  gibt es eine Lösung, für  $c = 1$  gibt es sogar zwei, für  $c = -1$  dagegen gar keine.

#### Beispiel

Sei  $p(x) = x^4$ . Für  $p(x) = 1$  gibt es in den komplexen Zahlen vier Lösungen.

## Beispiel

Sei  $p(x) = \sin(x)$ . Für  $p(x) = \frac{1}{3}$  gibt es in den reellen Zahlen unendlich viele Lösungen.

## Fehler und Residuum

Bei den direkten Problemen gibt es eine exakte, analytische *Lösung*  $p(x)$  und ein numerisches, berechnetes *Ergebnis*  $\tilde{p}(x)$ . Die Differenz  $(\tilde{p}(x) - p(x))$  wird (*absoluter*) *Fehler* genannt.

Bei den inversen Problemen gibt es dagegen i.A. keine exakte, analytische Lösung  $x$ , wohl aber ein numerisches, berechnetes Ergebnis  $\tilde{x}$ . Der Fehler  $(\tilde{x} - x)$  kann daher i.A. nicht wohldefiniert sein. Darum betrachtet man stattdessen das *Residuum*, z.B.  $r(\tilde{x}) = \|p(\tilde{x}) - y\|$ . Dieses ist wohldefiniert. Falls  $r(\tilde{x}) = \|p(\tilde{x}) - y\| = 0$  ist, ist  $\tilde{x}$  eine Lösung der Gleichung  $p(\tilde{x}) = y$ . Man fordert vom Residuum daher folgende Eigenschaften:

### Definition (Residuumsfunktion):

Für eine gegebene Funktion  $p : \mathbb{R}^n \rightarrow \mathbb{R}^m$  und ein  $y \in \mathbb{R}^m$  heißt eine Funktion  $r : \mathbb{R}^n \rightarrow [0, \infty)$  Residuum oder Residuumsfunktion, falls gilt:

- $\forall x \in \mathbb{R}^n: r(x) = 0 \Leftrightarrow p(x) = y$
- $\forall x \in \mathbb{R}^n: r(x) \geq 0$

Nun sind die Lösungen der Gleichung  $p(x) = y$  Nullstellen und zugleich Minimalstellen von  $r$ . Aus der Nullstellensuche wurde also eine allgemeinere Minimalstellensuche.

## Minimalstellensuche

Wenn es ein Verfahren gibt, um für eine beliebige Funktion eine Minimalstelle zu finden, dann kann man das Verfahren nutzen, um beliebige Gleichungssysteme zu lösen. Dazu betrachtet man die Residuumsfunktion  $r(x) = \|p(x) - y\|$ . Falls eine Lösung existiert, dann ist sie eine Minimalstelle von  $r$ .

Daneben gibt es viele Probleme in der Physik, bei denen Minimierungsprobleme natürlich auftreten. Beispielsweise Vorhersagen. Ein Koch möchte vorhersagen, wie viele Gäste am Abend kommen. Je genauer er das kann, desto genauer kann er die Menge frisches Gemüse, das er zu Mittag kaufen muss, besorgen und möglichst alle Gäste abends bewirtschaften, ohne Salat wegschmeißen zu müssen.

Ein anderes Beispiel sind Anwendungen in der Wirtschaft und im Finanzwesen. Ein Autohersteller möchte vorhersagen, wie viele Menschen sein Produkt kaufen werden. Je genauer er das vorhersagen kann, desto besser kann er seine Produktion an die Nachfrage anpassen und desto mehr Gewinn erzielt er.

Vorhersagen sind Minimierungsprobleme, da der Abstand zwischen Vorhersage und tatsächlich eingetretenem Ereignis (Experiment) minimal sein

soll. Es ist natürlich, diese Probleme als Minimierungsprobleme aufzufassen. Je besser eine Funktion minimiert werden kann, desto mehr Geld spart eine Versicherungsgesellschaft oder desto mehr Gewinn macht ein Industrieunternehmen. Es ist dabei nicht notwendig, dass eine exakte Nullstelle einer Residuumsfunktion gefunden wird. Je niedriger  $p(\tilde{x})$  wird, desto besser.

## Abstiegsverfahren

Eine häufige Methode für Minimierungsprobleme sind Abstiegsverfahren. Dabei wird eine Folge  $(\tilde{x}_n)_{n \in \mathbb{N}}$  berechnet, sodass  $\forall n \in \mathbb{N}: r(\tilde{x}_{n+1}) \leq r(\tilde{x}_n)$  ist.

**Definition** (Abstiegsverfahren):

Sei  $r: \mathbb{R}^n \rightarrow [0, \infty)$  eine Residuumsfunktion.

Ein  $\tilde{x}_2 \in \mathbb{R}^n$  heißt abgestiegen von  $\tilde{x}_1 \in \mathbb{R}^n$ , falls  $r(\tilde{x}_2) \leq r(\tilde{x}_1)$ .

Eine Funktion  $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$  heißt Abstiegsverfahren, falls:

- $\forall x \in \mathbb{R}^n: g(x)$  abgestiegen von  $x$ .

Die Folge  $(g^n(x))_{n \in \mathbb{N}}$ , wobei  $g^n$  die  $n$ -fache Hintereinanderausführung ist, heißt die von  $g$  bei  $x$  erzeugte Abstiegsfolge oder Schrittfolge.

## Beispiel

Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  eine beliebige Funktion.

Sei  $g: \mathbb{R}^n \rightarrow \mathbb{R}^n, g(x) = \begin{cases} f(x), & r(f(x)) \leq r(x) \\ x, & r(f(x)) > r(x) \end{cases}$ . Dann ist  $g$  ein

Abstiegsverfahren.

Abstiegsverfahren sind iterative Näherungsverfahren, das heißt, i.A. erzeugen sie eine Folge besser werdende Ergebnisse, ohne je das ursprüngliche Gleichungssystem exakt zu lösen.

## Koerzivität

**Definition** (Koerzive Funktion, Subniveaumenge):

Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ .  $f$  heißt koerziv, falls:

- $\forall (x_n)_{n \in \mathbb{N}}$  Folgen auf  $\mathbb{R}^n$  mit  $\lim_{n \rightarrow \infty} \|x_n\| = \infty: \lim_{n \rightarrow \infty} f(x_n) = \infty$ .

$\mathcal{L}_f(q) := \{x \in \mathbb{R}^n: f(x) \leq q\}$  heißt Subniveaumenge von  $f$  zu  $q$ .

**Bemerkung:**

Für jede koerzive Funktion  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  gilt:

- $f$  besitzt mindestens eine globale Minimalstelle.
- Für jedes  $q \in \mathbb{R}$  ist  $\mathcal{L}_f(q)$  kompakt.
- Wenn  $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$  ein Abstiegsverfahren ist, dann besitzt jede von  $g$  erzeugte Folge einen Häufungspunkt.

## Häufige Probleme

### Lineare Gleichungssysteme

Die wichtigsten direkten Probleme sind Matrix-Vektor-Multiplikation und Matrix-Matrix-Multiplikation.

Die wichtigsten inversen Probleme sind die Umkehrung der wichtigsten direkten Probleme, also lineare Gleichungssysteme.

Es gibt zwar den Gauß-Algorithmus, doch dieser ist für große dünnbesetzte Matrizen nicht effizient. In diesem Fall sind Abstiegsverfahren oft wesentlich billiger, d.h. schneller. Dies ist ein häufiger Anwendungsfall.

Gegeben sei ein  $A \in \mathbb{R}^{n \times n}$  und ein  $b \in \mathbb{R}^n$ . Gesucht ist ein  $x \in \mathbb{R}^n$  s.d.  $Ax = b$ . Das zugehörige direkte Problem ist  $p(x) = Ax$ , das inverse Problem  $p(x) = b$ .

### Residuumsfunktion

Die naheliegende Residuumsfunktion wäre  $r_1(x) = \|Ax - b\|$ , jedoch ist die Residuumsfunktion  $r_2(x) = \|Ax - b\|^2$  besser, da sie überall differenzierbar ist, während  $r_1$  nicht an den Punkten einer exakten Lösung differenzierbar ist.

Dazu kommt noch, dass  $r_2$  für  $\mathbb{R}^n = \mathbb{R}^1$  eine quadratische Funktion ist, also eine Polynomsfunktion, die eine besonders einfache Struktur hat. Diese lässt sich bei den Methoden für die Berechnung eines Ergebnisses ausnutzen.

Darüber hinaus ist die Residuumsfunktion  $r_2$  physikalisch motiviert, wo die Energie häufig quadratisch mit der Auslenkung anwächst. Man stellt sich das Residuum als Kosten eines ineffizienten Prozesses vor, und die Ineffizienz als „Energie der Störung“. Für viele physikalische Probleme motiviert das  $r_2$  als „natürlichere“ Residuumsfunktion.

Sei nun also  $r = r_2$  die Residuumsfunktion für lineare Gleichungssysteme.

Für den Fall  $\mathbb{R}^n = \mathbb{R}^1$  ist die Residuumsfunktion eine quadratische Funktion, deren nähere Untersuchung sich lohnt. Sei  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  eine quadratische Funktion mit positiver Krümmung,  $x \in \mathbb{R}$ . Dann ist  $\bar{x} := x - \frac{\phi'(x)}{\phi''(x)}$  die eindeutig bestimmte Minimalstelle von  $\phi$ .

Für  $x, d \in \mathbb{R}^n$  ist  $\phi: \mathbb{R} \rightarrow \mathbb{R}, \phi(s) = r_2(x + sd)$  eine quadratische Funktion, und hat positive Krümmung falls  $A$  invertierbar ist.

### Definition (quadratische Funktion auf $\mathbb{R}^n$ )

Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ .  $f$  heißt eine quadratische Funktion auf  $\mathbb{R}^n$ , falls:

- $\exists A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n, c \in \mathbb{R}: \forall x \in \mathbb{R}^n: f(x) = x^T A x + b^T x + c$ .

### Konvexität

#### Definition (konvexe Menge):

Sei  $\Omega \subset \mathbb{R}^n$ .  $\Omega$  heißt konvexe Menge, falls

- $\forall x, y \in \Omega, \lambda \in [0, 1]: \lambda y + (1 - \lambda)x \in \Omega$

**Definition** (konvex, strikt konvex, stark konvex):

Sei  $\Omega \subset \mathbb{R}^n$  eine konvexe Menge,  $f: \Omega \rightarrow \mathbb{R}$ .

$f$  heißt konvex, falls:

- $\forall x, y \in \Omega, \lambda \in [0, 1]: f(\lambda y + (1 - \lambda)x) \leq \lambda f(y) + (1 - \lambda)f(x)$

$f$  heißt strikt konvex, falls:

- $\forall x, y \in \Omega, \lambda \in (0, 1): f(\lambda y + (1 - \lambda)x) < \lambda f(y) + (1 - \lambda)f(x)$

$f$  heißt  $\mu$ -stark konvex für ein  $\mu > 0$ , falls:

- $\forall x, y \in \Omega, \lambda \in [0, 1]:$   
$$\lambda f(y) + (1 - \lambda)f(x) - f(\lambda y + (1 - \lambda)x) \geq \frac{\mu}{2} \lambda(1 - \lambda) \cdot \|y - x\|^2$$

$f$  heißt stark konvex, falls:  $\exists \mu > 0: f$   $\mu$ -stark konvex.

### Einfache Minimalstellen

**Definition** (Einfache Minimalstelle):

Sei  $r \in \mathcal{C}^2(\mathbb{R}^n, [0, \infty))$ . Ein  $x \in \mathbb{R}^n$  heißt einfache Minimalstelle von  $r$ , falls

$r'(x) := \nabla r(x)^T = 0$  ist und  $r''(x) := \mathcal{H}r(x) = \left( \partial_i \partial_j r(x) \right)_{1 \leq i, j \leq n}$  positiv definit ist.

**Satz** (einfache Minimalstelle stark konvex):

Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x$  eine einfache Minimalstelle von  $f$ . Dann existiert eine Umgebung  $U$  von  $x$  so, dass  $f|_U$  stark konvex ist.

Weiters existieren  $A \in \mathbb{R}^{n \times n}$  invertierbar,  $b \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$  so, dass  $f(x) = c$  und  $\forall \tilde{x} \in U: f(\tilde{x}) \geq c + \|A\tilde{x} - b\|^2$ .

Beweis: (Skizze)

Sei  $H = \mathcal{H}f(x)$ . Dann ist  $H$  symmetrisch und positiv definit. Wähle den niedrigsten Eigenwert  $\lambda$ . Setze  $\mu = \frac{\lambda}{3}$ . Dann existiert  $U$  Umgebung von  $x$ , sodass  $f|_U$   $\mu$ -stark konvex ist.

Wähle  $A = \sqrt{\mu}I$ ,  $b = Ax$ .

Dann ist  $f(\tilde{x}) \geq c + \|A\tilde{x} - b\|^2 = c + \|A(\tilde{x} - x)\|^2 = c + \mu\|\tilde{x} - x\|^2$ .

Minimierungsverfahren für die Residuumsfunktion von linearen Gleichungssystemen können also unter gewissen Einschränkungen auch verwendet werden, um einfache Minimalstellen zu finden, wenn man die Funktion auf eine Umgebung einschränkt und nur diese Umgebung betrachtet. Dazu muss man schon in der Nähe (in einer Umgebung) einer einfachen Minimalstelle sein.

## Verfahren

### Gefälleverfahren

Das Gefälleverfahren ist ein Verfahren zum Suchen von Minimalstellen.

Sei  $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R})$  koerziv,  $x_0 \in \mathbb{R}^n$ .

Formuliere die Differentialgleichung mit Anfangsbedingung:

$$\begin{aligned}x(0) &= x_0 \\x'(t) &= -\nabla f(x(t))\end{aligned}$$

Bemerkung: Es gilt  $\lim_{t \rightarrow \infty} \nabla f(x(t)) = 0$ . Weiters besitzt  $(x(t))_{t \geq 0}$  einen Häufungspunkt  $\bar{x}$ , sodass  $\nabla f(\bar{x}) = 0$  ist.

Ein  $x \in \mathbb{R}^n$ , sodass  $\nabla f(x) = 0$ , heißt *stationärer Punkt* von  $x$ .

Dass  $x$  ein stationärer Punkt von  $f$  ist, ist eine notwendige Bedingung dafür, dass  $x$  eine Minimalstelle von  $f$  ist.

**Definition** (Abstiegsrichtung):

Sei  $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R})$ ,  $x \in \mathbb{R}^n$ . Ein  $d \in \mathbb{R}^n$  heißt Abstiegsrichtung, falls:

- $\langle \nabla f(x), d \rangle < 0$ .

Bemerkung: Für alle  $x \in \mathbb{R}^n$ , sodass  $x$  kein stationärer Punkt von  $f$  ist, ist  $-\nabla f(x)$  eine Abstiegsrichtung für  $f$  an der Stelle  $x$ .

Gewöhnliche Differentialgleichungen werden numerisch durch die Diskretisierung (Euler-Verfahren, Zeitschrittverfahren) berechnet. Das ergibt folgenden Algorithmus:

**Algorithmus** (Gefälleverfahren):

Sei  $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R})$  koerziv.

1. Wähle  $x_0 \in \mathbb{R}^n$  beliebig, setze  $n := 0$ .
2. Falls Abbruchbedingung erfüllt: Stop. Ergebnis ist  $x_n$ .
3. Setze  $x_{n+1} := x_n - \alpha \nabla f(x_n)$  für ein geeignet gewähltes  $\alpha$ .
4. Setze  $n := n + 1$  und gehe zu Schritt 2.

Der Algorithmus ergibt ein  $x_n$ , sodass  $f(x_n)$  minimal wird. Da es sich um ein Näherungsverfahren handelt, ist das Ergebnis nicht im analytischen Sinne exakt, sondern ein numerisch berechnetes Ergebnis.

Eine geeignete Abbruchbedingung ist zu wählen. Möglich wären eine feste Anzahl Schleifendurchläufen (z.B. prüfe  $n > 20$ ) oder das Unterschreiten des Residuums einer Schranke (z.B. prüfe  $f(x_n) \leq 10^{-3}$ ) oder eine Oder-Verknüpfung beider Bedingungen.

Falls  $f \in \mathcal{C}^2$  ist, kann  $\alpha := \frac{1}{\phi_{n''}(0)}$  gewählt werden, wobei  $\phi_n(s) = f(x_n + s \nabla f(x_n))$  ist. Somit wird  $\alpha$  für quadratische Funktionen mit positiver Krümmung auf  $\mathbb{R}^n$  so gewählt, dass  $f(x_{n+1})$  minimal wird.

## Konvergenzordnung

Besonders wichtig ist die Minimierung quadratischer Funktionen auf  $\mathbb{R}^n$ . Da in diesem Bereich die meiste Arbeit verrichtet wird, wünscht man sich, dass die Verfahren effizient sind.

Die folgenden Definitionen beschäftigen sich mit Effizienz.

**Definition** (Lineare Konvergenzordnung, Effizientes Abstiegsverfahren):

Sei  $r_2: \mathbb{R}^n \rightarrow \mathbb{R}$  eine quadratische Funktion mit positiver Krümmung auf  $\mathbb{R}^n$ , sodass

- $\exists x \in \mathbb{R}^n: r_2(x) = 0$ ,
- $\forall x \in \mathbb{R}^n: r_2(x) \geq 0$ .

Das heißt, sei  $r_2$  die Residuumsfunktion eines linearen Gleichungssystems.

Sei  $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$  ein Abstiegsverfahren.  $g$  heißt effizient für  $r_2$ , falls:

- $\exists \beta > 0: \forall x \in \mathbb{R}^n: r_2(g(x)) \leq \beta r_2(x)$ .

$g$  heißt ein effizientes Abstiegsverfahren, falls es für alle Residuumsfunktionen von linearen Gleichungssystemen effizient ist.

Die Konvergenzordnung von  $g$  heißt linear, falls  $g$  ein effizientes Abstiegsverfahren ist. Das bedeutet, dass das Residuum mit zunehmendem Arbeitsaufwand exponentiell abfällt, oder anders gesagt, dass die Anzahl richtiger Stellen im Ergebnis linear zunimmt.

Bemerkung:

Das Gefälleverfahren mit Wahl für  $\alpha$  wie oben vorgestellt oder kleiner (aber positiv) ist ein effizientes Abstiegsverfahren. Für ein zu großes  $\alpha$  wird das Verfahren jedoch instabil und konvergiert nicht. Ein geeignetes  $\alpha$  kann auch experimentell gewonnen werden durch Beobachtung.

## Allgemeine Verfahren: Strömungsmechanik

Die Strömungsmechanik umfasst Flüssigkeiten (z.B. Wasser) und Gase (z.B. Luft). Das Wasser folgt dem Gefälle in der Landschaft, genau so, wie es im Gefälleverfahren geschieht. Daher motiviert sich der Name „Gefälleverfahren“. Das Wasser fließt immer nach unten. Indem es das tut, gelangt es irgendwann ins Meer. Der Wind wird ebenso von der Schwerkraft angezogen, nur dass auf ihn noch zusätzliche Kräfte wirken (Luftdruck). Dadurch kann er Hindernisse wie Berge überwinden, um in das dahinterliegende Tal zu gelangen. Somit sind die wichtigsten Anwendungen der Strömungsmechanik gleichzeitig eine Form von Optimierungsverfahren.

Viele wichtige Verfahren in der Optimierung sind aus der Strömungsmechanik motiviert. Für das Gefälleverfahren stammt die Motivation vom Fluss von Wasser. Für das Gefälleverfahren gibt es eine Erweiterung mit Moment, das dem linearen Moment in der Physik entspricht. Darüber hinaus gibt es noch eine Reihe weiterer Erweiterungen, die unterschiedliche Vor- und Nachteile haben.

### Gefälleverfahren mit Moment

Das Gefälleverfahren mit Moment ist die Diskretisierung der Differentialgleichung:

$$\begin{aligned}x(0) &= x_0 \\x''(t) &= -\alpha \nabla f(x(t)) - \beta x'(t)\end{aligned}$$

für geeignete Konstanten  $\alpha, \beta$ .

### Physikalische Realisierungen

Auch physikalische Realisierungen der Optimierungsverfahren sind möglich. Beispielsweise versucht der Quantencomputer, durch Ausnutzen der physikalischen Natur der Quanten, das Gefälleverfahren experimentell nachzubilden. Dadurch erhofft man sich eine bedeutende Zunahme der Geschwindigkeit. Darüber hinaus gelten für Quanten besondere Phänomene, die es ermöglichen, über Hindernisse drüberzuspringen. Das Gebiet ist noch nicht voll entwickelt, wächst aber langsam.