

VORLESUNG
Numerik und Optimierung

KAPITEL 5
Optimierungsprobleme und deren numerische Lösung

Ulrich Langer und Walter Zulehner

WS 2017/18, WS 2018/19

Inhaltsverzeichnis

5	Optimierungsprobleme und deren numerische Lösung	5-1
5.1	Freie Optimierungsprobleme	5-4
5.1.1	Theoretische Grundlagen	5-4
5.1.2	Abstiegsverfahren	5-6
5.1.3	Das Newton-Verfahren	5-7
5.1.4	Quasi-Newton-Verfahren	5-11
5.2	Optimierungsprobleme mit Nebenbedingungen	5-12
5.2.1	Theoretische Grundlagen	5-12
5.2.2	SQP-Verfahren	5-14
5.2.3	Quadratische Optimierungsprobleme	5-15

Kapitel 5

Optimierungsprobleme und deren numerische Lösung

Wir diskutieren hier Optimierungsprobleme der folgenden Form: Gesucht ist $x^* \in C$ mit

$$f(x^*) = \min_{x \in C} f(x),$$

für ein gegebenes Zielfunktional (Kostenfunktional) $f: \mathbb{R}^n \rightarrow \mathbb{R}$ und einer gegebenen Zulässigkeitsmenge $C \subset \mathbb{R}^n$. Ein Punkt $x \in C$ heisst zulässiger Punkt. Man spricht von einem endlich-dimensionalen, kontinuierlichen Optimierungsproblem.

Beispiel Optimalsteuerungsproblem: Wir betrachten zunächst ein unendlich-dimensionales Optimierungsproblem: Gesucht sind Funktionen $y^*(x)$ und $u^*(x)$ auf $\bar{\Omega}$ aus geeigneten Funktionenräumen Y und U so, dass

$$J(y^*, u^*) = \min_{(y,u) \in Y \times U} J(y, u)$$

mit

$$J(y, u) = \frac{1}{2} \int_{\Omega} |y(x) - y_d(x)|^2 dx + \frac{\alpha}{2} \int_{\Omega} |u(x)|^2 dx,$$

wobei die folgenden Nebenbedingungen erfüllt sein sollen: die Zustandsgleichung

$$\begin{aligned} -\Delta y(x) &= u(x), & x \in \Omega, \\ y(x) &= 0, & x \in \Gamma = \partial\Omega, \end{aligned}$$

Schranken für den Zustand

$$\underline{y}(x) \leq y(x) \leq \bar{y}(x), \quad x \in \Omega,$$

und Schranken für die Kontrolle

$$\underline{u}(x) \leq u(x) \leq \bar{u}(x), \quad x \in \Omega.$$

Die Zustandsgleichung wird in variationelle Form gebracht: $y \in H_0^1(\Omega) = Y$ und $u \in L_2(\Omega) = U$ erfüllen die Variationsgleichung

$$\int_{\Omega} \nabla y(x) \cdot \nabla q(x) \, dx = \int_{\Omega} u(x) q(x) \, dx \quad \text{für alle } q \in H_0^1(\Omega).$$

Wir diskretisieren nun das Optimierungsproblem mit der FE-Methode. Dazu wählt man die folgenden Ansätze für die Näherungen $y_h(x)$ und $u_h(x)$:

$$y_h(x) = \sum_{j \in \omega_h} y_j p_j(x), \quad u_h(x) = \sum_{k \in \bar{\omega}_h} u_k p_k(x).$$

Für das Zielfunktional erhält man dann

$$\begin{aligned} J(y_h, u_h) &= \frac{1}{2} \int_{\Omega} y_h(x) y_h(x) \, dx - \int_{\Omega} y_d(x) y_h(x) \, dx \\ &\quad + \frac{1}{2} \int_{\Omega} y_d(x) y_d(x) \, dx + \frac{\alpha}{2} \int_{\Omega} u_h(x) u_h(x) \, dx \\ &= \frac{1}{2} (M_y \underline{y}_h, \underline{y}_h) - (\underline{y}_d, \underline{y}_h) + \frac{1}{2} (M_u \underline{u}_h, \underline{u}_h) + c \end{aligned}$$

mit

$$M_y = \left(\int_{\Omega} p_j(x) p_i(x) \, dx \right)_{i,j \in \omega_h}, \quad \underline{y}_d = \left(\int_{\Omega} y_d(x) p_i(x) \, dx \right)_{i \in \omega_h}, \quad M_u = \left(\int_{\Omega} p_j(x) p_i(x) \, dx \right)_{i,j \in \bar{\omega}_h}$$

und der Konstanten $c = \frac{1}{2} \int_{\Omega} y_d(x) y_d(x) \, dx$, die auf die Lösung des Optimierungsproblem keine Auswirkung hat.

Die diskretisierte Zustandsgleichung lautet

$$K_h \underline{y}_h = M_{sc} \underline{u}_h$$

mit

$$K_h = \left(\int_{\Omega} \nabla p_j(x) \cdot \nabla p_i(x) \, dx \right)_{i,j \in \omega_h}, \quad M_{uy} = \left(\int_{\Omega} p_k(x) p_i(x) \, dx \right)_{i \in \omega_h, k \in \bar{\omega}_h}.$$

Die Zustands- und Kontrollbeschränkungen werden in den Knoten gefordert:

$$\underline{y}(x_j) \leq y_j \leq \bar{y}(x_j) \quad \text{für alle } j \in \omega_h,$$

$$\underline{u}(x_k) \leq u_k \leq \bar{u}(x_k) \quad \text{für alle } k \in \bar{\omega}_h.$$

Die Zulässigkeitsmenge C besteht aus allen Paaren $(\underline{y}_h, \underline{u}_h)$, die die diskretisierte Zustandsgleichung und die diskretisierten Zustands- und Kontrollbeschränkungen erfüllen.

Eine grobe Klassifizierung der Optimierungsaufgaben:

1. Freie Optimierung: $C = \mathbb{R}^n$.
2. Optimierungsprobleme mit Nebenbedingungen:
 - (a) Nur Gleichungsnebenbedingungen

$$C = \{x \in \mathbb{R}^n : c_i(x) = 0, i = 1, \dots, m\}$$

- (b) Auch Ungleichungsnebenbedingungen:

$$C = \{x \in \mathbb{R}^n : c_i(x) = 0, i = 1, \dots, m_1, \\ c_i(x) \leq 0, i = m_1 + 1, \dots, m_1 + m_2 = m\}$$

Nach der Komplexität der Funktionen $f(x)$ und $c_i(x)$ unterscheidet man:

1. Lineare Optimierung: die Funktionen $f(x)$ und $c_i(x)$ sind (affin) linear, d.h.:

$$f(x) = f_0 + g^T x, \quad c_i(x) = a_i^T x - b.$$

2. Quadratische Optimierung: das Zielfunktional $f(x)$ ist quadratisch, d.h.:

$$f(x) = f_0 + g^T x + \frac{1}{2} x^T H x,$$

und die Funktionen $c_i(x)$ sind (affin) linear.

3. Nichtlineare Optimierung mit linearen Nebenbedingungen: das Zielfunktional $f(x)$ ist nichtlinear, die Funktionen $c_i(x)$ sind (affin) linear.
4. Nichtlineare Optimierung mit nichtlinearen Nebenbedingungen: die Funktionen $f(x)$ und $c_i(x)$ sind nichtlinear.

Bemerkung 5.1 Die bisher beschriebenen Klassen von Optimierungsproblemen gehören zu den kontinuierlichen Optimierungsproblemen. Weitere Klassen von Optimierungsproblemen sind die kombinatorische Optimierung (C ist endlich), die ganzzahlige Optimierung ($C = \mathbb{Z}^n$), und die gemischt-ganzzahlige Optimierung, einer Kombination aus kontinuierlicher und ganzzahliger Optimierung.

Lösungsbegriffe

Definition 5.1 1. Ein Punkt $x^* \in C$ heißt [striktes] globales Minimum genau dann, wenn

$$f(x^*) \leq f(x) \quad [f(x^*) < f(x)] \quad \text{für alle } x \in C, x \neq x^*.$$

2. Ein Punkt $x^* \in C$ heißt [striktes] lokales Minimum genau dann, wenn

$$f(x^*) \leq f(x) \quad [f(x^*) < f(x)] \quad \text{für alle } x \in C, \quad x \neq x^* \text{ mit } \|x - x^*\| \leq \varepsilon$$

für ein $\varepsilon > 0$.

Bemerkung 5.2 Man spricht von konvexen Optimierungsproblemen, wenn $f(x)$ konvex ist, d.h.,

$$f((1 - \alpha)x + \alpha y) \leq (1 - \alpha)f(x) + \alpha f(y) \quad \text{für alle } x, y \in \mathbb{R}^n, \quad \alpha \in [0, 1],$$

und C konvex ist, d.h.,

$$(1 - \alpha)x + \alpha y \in C \quad \text{für alle } x, y \in C, \quad \alpha \in [0, 1].$$

Für konvexe Optimierungsprobleme lässt sich zeigen, dass jedes lokale Minimum ein globales Minimum ist.

5.1 Freie Optimierungsprobleme

Wir diskutieren nun Optimierungsprobleme der folgenden Form: Gesucht ist $x^* \in \mathbb{R}^n$ mit

$$f(x^*) = \min_{x \in \mathbb{R}^n} f(x).$$

Beispiel: Im Kapitel über iterative Verfahren zur Lösung linearer Gleichungssysteme behandelten wir das Gradientenverfahren zur Lösung des freien Optimierungsproblems

$$J(x^*) = \min_{x \in \mathbb{R}^n} J(x) \quad \text{mit} \quad J(x) = \frac{1}{2}(Ax, x) - (b, x),$$

das für eine symmetrische und positiv definite Matrix A zur Lösung des linearen Gleichungssystems $Ax = b$ äquivalent ist.

5.1.1 Theoretische Grundlagen

Sei $f: \mathbb{R}^n \rightarrow \mathbb{R}$ zweimal stetig partiell differenzierbar. Der Gradient von f wird mit $g = \text{grad } f = \nabla f$ bezeichnet und ist folgendermaßen gegeben:

$$g(x) = \nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

Offensichtlich gilt: $\nabla f(x) = f'(x)^\top$, wobei $f'(x)$ die Jacobi-Matrix von $f(x)$ bezeichnet, die hier folgende Form hat:

$$f'(x) = \left(\frac{\partial f}{\partial x_1}(x) \quad \frac{\partial f}{\partial x_2}(x) \quad \dots \quad \frac{\partial f}{\partial x_n}(x) \right).$$

Die Hesse-Matrix von f wird mit $H = H_f = \nabla^2 f$ bezeichnet und ist folgendermaßen gegeben:

$$H(x) = \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}.$$

Man beachte, dass die Hesse-Matrix symmetrisch ist. Es gilt: $\nabla^2 f(x) = (\nabla f(x))'$.

Aus der Taylor-Entwicklung des Zielfunktionalen um den Punkt x^*

$$f(x) = f(x^*) + \nabla f(x^*) \cdot (x - x^*) + \frac{1}{2}(x - x^*)^\top H_f(x^*)(x - x^*) + o(\|x - x^*\|^2)$$

lassen sich folgende Schlussfolgerungen ziehen:

Satz 5.1 (Notwendige Bedingungen für ein lokales Minimum) Sei x^* ein lokales Minimum von f . Dann gilt:

1. $\nabla f(x^*) = 0$.
2. $H_f(x^*)$ ist positiv semi-definit.

Satz 5.2 (Hinreichende Bedingungen für ein lokales Minimum) Falls

1. $\nabla f(x^*) = 0$ und
2. $H_f(x^*)$ positiv definit ist,

dann ist x^* ein striktes lokales Minimum von f .

Definition 5.2 Ein Punkt $x^* \in \mathbb{R}^n$ mit $\nabla f(x^*) = 0$ heißt ein stationärer (oder kritischer) Punkt von f .

Bemerkung 5.3 Ein stationärer Punkt x^* mit einer indefiniten Hesse-Matrix $H_f(x^*)$ heißt Sattelpunkt. (Eine symmetrische Matrix H heißt indefinit genau dann, wenn es zwei Richtungen s_1 und s_2 mit $s_1^\top H s_1 > 0$ und $s_2^\top H s_2 < 0$ gibt.)

5.1.2 Abstiegsverfahren

Ein Verfahren zur Lösung eines Optimierungsproblems hat üblicherweise folgende Form:
Ausgehend von einer Näherung $x^{(k)}$ von x^*

1. bestimme eine Suchrichtung $s^{(k)}$,
2. bestimme eine Schrittweite $\alpha^{(k+1)} > 0$ mit $f(x^{(k)}) < f(x^{(k)} + \alpha^{(k+1)} s^{(k)})$ und
3. setze $x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} s^{(k)}$.

Man nennt ein solches Verfahren ein Liniensuchverfahren.

Beispiele: Gradientenverfahren, cg-Verfahren zur Minimierung von $J(x)$.

Anforderung an die Suchrichtung:

Die Taylor-Entwicklung

$$\begin{aligned} f(x^{(k)} + \alpha^{(k+1)} s^{(k)}) &= f(x^{(k)}) + \alpha f'(x^{(k)}) s^{(k)} + o(\alpha) \\ &= f(x^{(k)}) + \alpha \nabla f(x^{(k)}) \cdot s^{(k)} + o(\alpha) \end{aligned}$$

zeigt, dass das Zielfunktional für hinreichend kleine Werte von $\alpha > 0$ entlang der Richtung $s^{(k)}$ kleiner wird, falls $\nabla f(x^{(k)}) \cdot s^{(k)} < 0$. Das motiviert folgenden Begriff.

Definition 5.3 Eine Suchrichtung $s^{(k)}$ heißt eine Abstiegsrichtung von f genau dann, wenn $\nabla f(x^{(k)}) \cdot s^{(k)} < 0$. Ein Liniensuchverfahren, bei der die gewählten Suchrichtungen stets Abstiegsrichtungen sind, heißt Abstiegsverfahren.

Geometrische Interpretation: Der Winkel zwischen $-\nabla f(x^{(k)})$ und der Suchrichtung $s^{(k)}$ muss kleiner als 90° sein. Das wichtigste Beispiel eines Abstiegsverfahrens ist das Gradientenverfahren, bei dem die Suchrichtung mit $-\nabla f(x^{(k)})$ übereinstimmt, der Winkel also 0° ist.

Anforderungen an die Schrittweite:

Bei der exakten Liniensuche wird die Schrittweite $\alpha^{(k+1)} > 0$ so gewählt, dass

$$f(x^{(k)} + \alpha^{(k+1)} s^{(k)}) = \min_{\alpha > 0} f(x^{(k)} + \alpha s^{(k)}).$$

Diese Strategie haben wir beim Gradientenverfahren und dem cg-Verfahren zur Lösung linearer Gleichungssysteme verwendet.

Für nichtlineare Probleme ist diese Strategie zu aufwändig. Stattdessen wird die Schrittweite so gewählt, dass die beiden folgenden Bedingungen erfüllt sind:

Die Wolfe-Powell Bedingungen: Wähle $\alpha^{(k+1)} > 0$ so, dass

1. $f(x^{(k)} + \alpha^{(k+1)} s^{(k)}) \leq f(x^{(k)}) + \mu \alpha \nabla f(x^{(k)}) \cdot s^{(k)}$ für einen vorgegebenen Wert $\mu \in (0, 1)$ und
2. $\nabla f(x^{(k)} + \alpha^{(k+1)} \cdot s^{(k)}) \geq \sigma \nabla f(x^{(k)}) \cdot s^{(k)}$ für einen vorgegebenen Wert $\sigma \in (\mu, 1)$.

Die erste Bedingung sichert, dass f hinreichend gut reduziert wird, relativ zur Reduktion um $\alpha \nabla f(x^{(k)}) \cdot s^{(k)}$, die man aufgrund der Taylor-Entwicklung erwartet. Durch die zweite Bedingung werden zu kleine absolute Schrittweiten verhindert, ohne die ideale Schrittweite durch die exakte Liniensuche auszuschliessen.

Es gilt folgende wichtige Aussage über Abstiegsverfahren:

Satz 5.3 *Falls*

1. die Näherungen $x^{(k)}$ eine beschränkte Folge bilden,
2. es ein $\varepsilon > 0$ gibt mit

$$\nabla f(x^{(k)}) \cdot s^{(k)} \leq -\varepsilon \|f(x^{(k)})\| \|s^{(k)}\| \quad \text{für alle } k = 0, 1, 2, \dots,$$

und

3. die Schrittweiten $\alpha^{(k+1)}$ die Wolfe-Powell Bedingungen erfüllen,

dann gilt: $\lim_{k \rightarrow \infty} \nabla f(x^{(k)}) = 0$.

Da die Näherungen $x^{(k)}$ eine beschränkte Folge bilden, gibt es mindestens eine konvergente Teilfolge. Dieser Satz garantiert dann, dass der Grenzwert jeder konvergenten Teilfolge ein stationärer Punkt von f ist. Die Aussage gilt für alle Startwerte $x^{(0)}$. In diesem Sinne ist ein Abstiegsverfahren, bei dem der Winkel zwischen $-\nabla f(x^{(k)})$ und der Suchrichtung $s^{(k)}$ gleichmäßig kleiner als 90° ist (siehe Bedingung 2), global konvergent.

Für das Gradientenverfahren ist die zweite Bedingung für $\varepsilon = 1$ erfüllt. Also ist das Gradientenverfahren mit Liniensuche nach Wolfe-Powell global konvergent. Allerdings konvergiert das Gradientenverfahren im Allgemeinen langsam. Daher betrachten wir im nächsten Abschnitt eine wichtige Alternative.

5.1.3 Das Newton-Verfahren

Um die Lösung eines freien Optimierungsproblems zu berechnen, muss man das Gleichungssystem

$$\nabla f(x) = 0$$

lösen, also ein im Allgemeinen nichtlineares Gleichungssystem in \mathbb{R}^n . Andere nichtlineare Gleichungssysteme der Form

$$\underline{K}_h(\underline{u}_h) = \underline{f}_h$$

entstehen bei der Diskretisierung von nichtlinearen Randwertproblemen.

Wir diskutieren nun die numerische Lösung von allgemeinen nichtlinearen Gleichungssystemen der Form

$$F(x) = 0$$

mit $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Die beiden obigen Beispiele von nichtlinearen Gleichungssystemen lassen sich natürlich auf diese Form bringen.

Die Idee des Newton-Verfahrens besteht darin, die nichtlineare Funktion $F(x)$ in einer Umgebung einer gegebenen Näherung $x^{(k)}$ durch eine (affin) lineare Funktion mit Hilfe der Taylor-Entwicklung zu approximieren (Linearisierung):

$$\begin{aligned} F(x) &= \underbrace{F(x^{(k)}) + F'(x^{(k)})(x - x^{(k)})}_{= L_k(x)} + o(\|x - x^{(k)}\|) \end{aligned}$$

mit der Jacobi-Matrix $F'(x)$, die folgendermaßen gegeben ist:

$$F'(x) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \frac{\partial F_1}{\partial x_2}(x) & \dots & \frac{\partial F_1}{\partial x_n}(x) \\ \frac{\partial F_2}{\partial x_1}(x) & \frac{\partial F_2}{\partial x_2}(x) & \dots & \frac{\partial F_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial x_1}(x) & \frac{\partial F_n}{\partial x_2}(x) & \dots & \frac{\partial F_n}{\partial x_n}(x) \end{pmatrix}.$$

Als nächste Näherung $x^{(k+1)}$ wird die Lösung des linearen Gleichungssystems

$$L_k(x) = 0$$

gewählt, also

$$F(x^{(k)}) + F'(x^{(k)}) \underbrace{(x^{(k+1)} - x^{(k)})}_{= s^{(k)}} = 0.$$

Das Newton-Verfahren hat also folgende Form

$$x^{(k+1)} = x^{(k)} + s^{(k)} \quad \text{mit} \quad F'(x^{(k)}) s^{(k)} = -F(x^{(k)}).$$

Durchführung des Newton-Verfahrens: In jedem Schritt muss ein lineares Gleichungssystem (zur Berechnung von $s^{(k)}$) gelöst werden.

In kompakter Form geschrieben erhält man für das Newton-Verfahren

$$x^{(k+1)} = x^{(k)} - F'(x^{(k)})^{-1} F(x^{(k)}).$$

Geometrische Interpretation des Newton-Verfahrens für $n = 1$: Die Linearisierung nach Newton entspricht der Approximation des Graphen von f durch die Tangente an den Graphen im Punkt $x^{(k)}$. Der Schnittpunkt der Tangente mit der x -Achse bestimmt die nächste Näherung $x^{(k+1)}$.

Anwendung auf freie Optimierungsprobleme

Die Berechnung einer Lösung eines freien Optimierungsproblems mit Zielfunktional $f: \mathbb{R}^n \rightarrow \mathbb{R}$ führt auf das Gleichungssystem

$$F'(x) = 0 \quad \text{mit} \quad F(x) = \nabla f(x).$$

Die Jacobi-Matrix dieser Funktion $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ stimmt mit der Hesse-Matrix von f überein:

$$F'(x) = (\nabla f(x))' = H_f(x).$$

Das lineare Gleichungssystem $L_k(x) = 0$ hat hier die konkrete Form:

$$\nabla f(x^{(k)}) + H_f(x^{(k)})(x - x^{(k)}) = 0,$$

also

$$\nabla q_k(x) = 0$$

mit

$$q_k(x) = f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^\top H_f(x^{(k)})(x - x^{(k)}).$$

Das quadratische Funktional $q_k(x)$ ist jene quadratische Approximation von $f(x)$ in einer Umgebung von $x^{(k)}$, die durch Taylor-Entwicklung von $f(x)$ um $x^{(k)}$ entsteht.

Das Newton-Verfahren für freie Optimierungsprobleme lautet also:

$$x^{(k+1)} = x^{(k)} + s^{(k)} \quad \text{mit} \quad H_f(x^{(k)}) s^{(k)} = -\nabla f(x^{(k)}).$$

Es lässt sich als Liniensuchverfahren mit der Suchrichtung (Newton-Richtung)

$$s^{(k)} = -H_f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

und der Schrittweite $\alpha^{(k+1)} = 1$ interpretieren.

Falls $H_f(x^{(k)})$ positiv definit ist, folgt, dass die Bedingung

$$\nabla q_k(x^{(k+1)}) = 0$$

für die nächste Näherung äquivalent zur Bedingung

$$q_k(x^{(k+1)}) = \min_{x \in \mathbb{R}^n} q_k(x)$$

ist, und dass die Newton-Richtung eine Abstiegsrichtung von q_k ist:

$$\nabla q_k(x^{(k)}) \cdot s^{(k)} = \nabla f(x^{(k)}) \cdot s^{(k)} = -\nabla f(x^{(k)})^\top H_f(x^{(k)})^{-1} \nabla f(x^{(k)}) < 0.$$

Daher bietet sich an, das Newton-Verfahren mit einer Liniensuche zu verbinden. Man bestimmt die Newton-Richtung als Suchrichtung und wählt für die nächste Näherung

$$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} s^{(k)}$$

mit einer geeigneten Schrittweitenwahl nach Wolfe-Powell.

Das Newton-Verfahren lässt sich somit als ein SQP-Verfahren (Sequential Quadratic Programming) interpretieren: Man erhält die Suchrichtungen bzw. die einzelnen Näherungen durch Lösung einer Folge von quadratischen Optimierungsproblemen.

Liniensuche für das Newton-Verfahren für allgemeine nichtlineare Gleichungssysteme

Jeder Punkt $x^* \in \mathbb{R}^n$ mit

$$F(x^*) = 0$$

ist globales Minimierung des Zielfunktional

$$f(x) = \|F(x)\|^2 = F(x)^\top F(x).$$

Die Newton-Richtung ist eine Abstiegsrichtung von f :

$$\begin{aligned}\nabla f(x^{(k)}) \cdot s^{(k)} &= (2F(x^{(k)})^\top F'(x^{(k)})) (-F'(x^{(k)})^{-1} F(x^{(k)})) \\ &= -2F(x^{(k)})^\top F(x^{(k)}) = -2f(x^{(k)}) < 0.\end{aligned}$$

Daher bietet sich auch hier an, das Newton-Verfahren mit einer Liniensuche zu verbinden:

$$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} s^{(k)}.$$

Man spricht dann vom gedämpften Newton-Verfahren.

Es gilt folgende Aussage über Konvergenzeigenschaften der Newton-Verfahrens:

Satz 5.4 Sei $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ einmal stetig partiell differenzierbar und sei $x^* \in \mathbb{R}^n$ eine einfache Nullstelle von F , d.h.: $F(x^*) = 0$ und $F'(x^*)$ ist regulär. Dann konvergiert das Newton-Verfahren für Startwerte $x^{(0)}$, die hinreichend nahe bei x^* liegen, gegen x^* und es konvergiert q -superlinear, d.h.:

$$\|x^{(k+1)} - x^*\| \leq q_k \|x^{(k)} - x^*\| \quad \text{mit} \quad q_k \rightarrow 0 \quad \text{für} \quad k \rightarrow \infty.$$

Falls zusätzlich $F'(x)$ im Punkt x^* Lipschitz-stetig ist, d.h.

$$\|F'(x) - F'(x^*)\| \leq L \|x - x^*\|$$

für eine Konstante L und alle x in einer Umgebung von x^* , dann konvergiert das Newton-Verfahren sogar q -quadratisch, d.h.:

$$\|x^{(k+1)} - x^*\| \leq C \|x^{(k)} - x^*\|^2$$

für eine Konstante $C > 0$.

Das Newton-Verfahren konvergiert also nur lokal. Wenn es konvergiert, konvergiert es allerdings sehr schnell. Aus der q -quadratischen Konvergenz folgt, dass die Anzahl der richtigen Dezimalstellen exponentiell zunimmt. Bei einem q -linear konvergenten Verfahren (wie dem Gradientenverfahren) nimmt die Anzahl der richtigen Dezimalstellen nur linear zu. Dafür ist das Gradientenverfahren global konvergent. Im nächsten Abschnitt wird eine Klasse von Verfahren diskutiert, die schnell und global zu konvergieren.

5.1.4 Quasi-Newton-Verfahren

Wir diskutieren Verfahren der Form

$$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} s^{(k)} \quad \text{mit} \quad B^{(k)} s^{(k)} = -F(x^{(k)}).$$

Die zu wählende Matrix $B^{(k)}$ soll in einem gewissen Sinn eine Approximation von $F'(x^{(k)})$ sein. Konkret stellt man folgende Forderungen an die Wahl der Matrix $B^{(k+1)}$, die im nächsten Iterationsschritt benötigt wird:

1. Die Matrix $B^{(k+1)}$ soll die Quasi-Newton-Bedingung (Sekanten-Bedingung) erfüllen:

$$B^{(k+1)}(x^{(k+1)} - x^{(k)}) = F(x^{(k+1)}) - F(x^{(k)}).$$

Für $n = 1$ reicht diese Bedingung zur Festlegung von $B^{(k+1)}$ aus:

$$B^{(k+1)} = \frac{F(x^{(k+1)}) - F(x^{(k)})}{x^{(k+1)} - x^{(k)}}.$$

Das entstehende Verfahren heisst Sekantenverfahren. Im Gegensatz zum Newton-Verfahren wird nicht die Tangente an den Graphen von f im Punkt $x^{(k)}$ sondern die Sekante durch die Punkte $(x^{(k)}, F(x^{(k)}))$ und $(x^{(k+1)}, F(x^{(k+1)}))$ zur Linearisierung verwendet.

2. Für $n > 1$ fordert man, dass sich $B^{(k+1)}$ von $B^{(k)}$ nur durch eine Matrix von niedrigem Rang (typischerweise vom Rang 1 oder 2) unterscheidet:

$$B^{(k+1)} = B^{(k)} + u_1^{(k)}(v_1^{(k)})^\top \left(+ u_2^{(k)}(v_2^{(k)})^\top \right).$$

3. Bei der Anwendung auf Optimierungsprobleme, also $F(x) = \nabla f(x)$, soll $B^{(k+1)}$ (als Approximation der Hesse-Matrix $H_f(x)$) symmetrisch und positiv definit sein. Dann ist sichergestellt, dass $s^{(k+1)} = -(B^{(k+1)})^{-1} \nabla f(x^{(k)})$ eine Abstiegsrichtung ist:

$$\nabla f(x^{(k+1)}) \cdot s^{(k+1)} = -\nabla f(x^{(k+1)})^\top (B^{(k+1)})^{-1} \nabla f(x^{(k)}) < 0.$$

Eines der bekanntesten Quasi-Newton-Verfahren ist das BFGS-Verfahren (Broyden-Fletcher-Goldfarb-Shanno):

$$B^{(k+1)} = B^{(k)} + \frac{1}{(y^{(k)})^\top s^{(k)}} y^{(k)}(y^{(k)})^\top - \frac{1}{(s^{(k)})^\top B^{(k)} s^{(k)}} (B^{(k)} s^{(k)})(B^{(k)} s^{(k)})^\top$$

mit

$$s^{(k)} = x^{(k+1)} - x^{(k)}, \quad y^{(k)} = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}).$$

Ein möglicher Startwert für $B^{(0)}$ ist I . Dann stimmt der erste Schritt des BFGS-Verfahrens mit dem ersten Schritt des Gradientenverfahrens überein.

Man beachte, dass zur Berechnung von $B^{(k)}$ nur der Gradient $\nabla f(x)$ (und nicht die Hesse-Matrix $\nabla^2 f(x)$) benötigt wird.

Es lässt sich zeigen, dass alle Matrizen $B^{(k)}$ symmetrisch und positiv definit bleiben, falls $B^{(0)}$ symmetrisch und positiv definit gewählt wird und die Schrittweiten $\alpha^{(k+1)}$ die zweite Bedingung nach Wolfe-Powell erfüllen. Das BFGS-Verfahren ist dann ein Abstiegsverfahren und somit in einem geeigneten Sinn global konvergent.

Unter geeigneten Bedingungen lässt sich zeigen, dass das BFGS-Verfahren q -superlinear konvergiert.

5.2 Optimierungsprobleme mit Nebenbedingungen

Wir diskutieren hier Optimierungsprobleme der folgenden allgemeinen Form: Gesucht ist $x^* \in \mathbb{R}^n$ mit

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

unter den Nebenbedingungen

$$\begin{aligned} c_i(x) &= 0, & i \in I_1 &= \{1, \dots, m_1\}, \\ c_i(x) &\leq 0, & i \in I_2 &= \{m_1 + 1, \dots, m_1 + m_2 = m\}. \end{aligned}$$

5.2.1 Theoretische Grundlagen

Wir führen die so genannte Lagrange-Funktion ein:

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

Die Variablen λ_i heißen Lagrange-Multiplikatoren. Für den Gradienten und die Hesse-Matrix von $L(x, \lambda)$ bezüglich x erhält man:

$$\begin{aligned} \nabla_x L(x, \lambda) &= \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla c_i(x), \\ \nabla_x^2 L(x, \lambda) &= \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 c_i(x). \end{aligned}$$

Zur Formulierung der notwendigen Bedingungen für ein lokales Minimum führen wir die Menge

$$\mathcal{A}(x) = \{i \in I_1 \cup I_2 : c_i(x) = 0\}$$

ein. Die Menge $\mathcal{A}(x)$ heißt die aktive Indexmenge im Punkt x .

Es gilt nun:

Satz 5.5 (Notwendige Bedingungen für ein lokales Minimum) Sei $x^* \in \mathbb{R}^n$ ein lokales Minimum von f . Falls f und c_i für alle $i \in I_1 \cup I_2$ stetig partiell differenzierbar sind und die Vektoren $\nabla c_i(x^*)$, $i \in \mathcal{A}(x^*)$ linear unabhängig sind, dann gilt: Es gibt Lagrange-Multiplikatoren $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)^\top \in \mathbb{R}^m$ mit

$$\begin{aligned}\nabla_x L(x^*, \lambda^*) &= 0, \\ c_i(x^*) &= 0 \quad \text{für alle } i \in I_1, \\ c_i(x^*) \leq 0, \lambda_i^* \geq 0, \lambda_i c_i(x^*) &= 0 \quad \text{für alle } i \in I_2.\end{aligned}$$

Die Bedingungen dieses Satzes heißen Karush-Kuhn-Tucker-Bedingungen, oder kurz KKT-Bedingungen.

Aus den Bedingungen für die Ungleichungsnebenbedingungen erkennt man, dass entweder eine Ungleichungsnebenbedingungen aktiv ist, d.h. $c_i(x^*) = 0$, oder aber der dazugehörige Lagrange-Multiplikator verschwindet: $\lambda_i^* = 0$. Würde man die aktive Indexmenge $\mathcal{A}^* = \mathcal{A}(x^*)$ kennen, dann könnte man die KKT-Bedingungen, die ursprünglich ein System von Gleichungen und Ungleichungen sind, auf ein Gleichungssystem reduzieren:

$$\begin{aligned}\nabla_x L(x^*, \lambda^*) &= 0, \\ c_i(x^*) &= 0 \quad \text{für alle } i \in \mathcal{A}^*.\end{aligned}$$

Das sind genau die KKT-Bedingungen für das folgende Optimierungsproblem: Gesucht ist $x^* \in \mathbb{R}^n$ mit

$$f(x^*) = \min_x f(x)$$

unter den Gleichungsnebenbedingungen

$$c_i(x) = 0, \quad i \in \mathcal{A}^*.$$

Bedeutung der Lagrange-Multiplikatoren

Für eine vorgegebene Indexmenge \mathcal{A} mit $I_1 \subset \mathcal{A} \subset I_1 \cup I_2$ führen wir das folgende Optimierungsproblem ein: Gesucht ist $x^* \in \mathbb{R}^n$ mit

$$f(x^*) = \min_x f(x)$$

unter den Gleichungsnebenbedingungen

$$c_i(x) = \xi_i, \quad i \in \mathcal{A}.$$

Die Parameter ξ_i , $i \in \mathcal{A}$, beschreiben Störungen der ursprünglichen Nebenbedingungen $c_i(x) = 0$ auf der rechten Seite. Dann hängt auch die Lösung der zugeordneten KKT-Bedingungen von diesen Parametern ab, man erhält Lösungsfunktionen $x^*(\xi)$ und $\lambda^*(\xi)$ mit $\xi = (\xi_i)_{i \in \mathcal{A}}$. Es lässt sich zeigen, dass:

$$\left. \frac{\partial}{\partial \xi_i} (f(x^*(\xi))) \right|_{\xi=0} = -\lambda_i^*(0).$$

Der Lagrange-Multiplikator $\lambda_i^* = \lambda_i^*(0)$ beschreibt also die Sensitivität (Änderungsrate) des Zielfunktional f bezüglich Störungen der Nebenbedingung $c_i(x) = 0$ auf der rechten Seite.

5.2.2 SQP-Verfahren

Die Idee des Newton-Verfahrens für nichtlineare Gleichungssysteme lässt sich auch auf die KKT-Bedingungen anwenden: Man ersetzt $\nabla_x L(x, \lambda)$ und $c_i(x)$ durch die lineare Approximationen, die man durch Taylor-Entwicklung um die aktuellen Näherungen $x^{(k)}$ und $\lambda^{(k)}$ erhält:

$$\begin{aligned}
 c_i(x) &\approx c_i(x^{(k)}) + \nabla c_i(x^{(k)}) \cdot (x - x^{(k)}), \\
 \nabla_x L(x, \lambda) &= \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla c_i(x) \\
 &\approx \nabla_x L(x^{(k)}, \lambda^{(k)}) + (\nabla_x L)'(x^{(k)}, \lambda^{(k)}) \begin{pmatrix} x - x^{(k)} \\ \lambda - \lambda^{(k)} \end{pmatrix} \\
 &= \nabla f(x^{(k)}) + \sum_{i=1}^m \lambda_i \nabla c_i(x^{(k)}) \\
 &\quad + (\nabla^2 f(x^{(k)}) + \sum_{i=1}^m \lambda_i^{(k)} \nabla c_i(x^{(k)})) (x - x^{(k)}) + \sum_{i=1}^m \nabla c_i(x^{(k)}) (\lambda_i - \lambda_i^{(k)}) \\
 &= \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) (x - x^{(k)}) + \sum_{i=1}^m \lambda_i \nabla c_i(x^{(k)}).
 \end{aligned}$$

Durch diese Approximationen entstehen aus den ursprünglichen KKT-Bedingungen die linearisierten KKT-Bedingungen:

$$\begin{aligned}
 \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) (x - x^{(k)}) + \sum_{i=1}^m \lambda_i \nabla c_i(x^{(k)}) &= 0, \\
 c_i(x^{(k)}) + \nabla c_i(x^{(k)}) \cdot (x - x^{(k)}) &= 0 \quad \text{für alle } i \in I_1, \\
 c_i(x^{(k)}) + \nabla c_i(x^{(k)}) \cdot (x - x^{(k)}) &\leq 0, \\
 \lambda_i \geq 0, \lambda_i (c_i(x^{(k)}) + \nabla c_i(x^{(k)}) \cdot (x - x^{(k)})) &= 0 \quad \text{für alle } i \in I_2.
 \end{aligned}$$

Führt man $s = x - x^{(k)}$ ein, so erkennt man, dass die linearisierten KKT-Bedingungen die KKT-Bedingungen des folgenden quadratischen Optimierungsproblem sind: Gesucht ist $s^{(k)} \in \mathbb{R}^n$ mit

$$q_k(s^{(k)}) = \min_{s \in \mathbb{R}^n} q_k(s)$$

unter den Nebenbedingungen

$$\begin{aligned}
 a_i^\top s &= b_i \quad \text{für alle } i \in I_1, \\
 a_i^\top s &\leq b_i \quad \text{für alle } i \in I_2
 \end{aligned}$$

mit

$$q_k(s) = [f(x^{(k)}) +] \nabla f(x^{(k)})^\top s + \frac{1}{2} s^\top \nabla_x^2 L(x^{(k)}, \lambda^{(k)}) s$$

und

$$a_i = \nabla c_i(x^{(k)}), \quad b_i = -c_i(x^{(k)}).$$

Ausgehend von den aktuellen Näherungen $x^{(k)}$ und $\lambda^{(k)}$ erhält man für die nächsten Näherungen des Newton-Verfahrens, das auch Wilson-Verfahren genannt wird:

$$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} s^{(k)} \quad \text{und} \quad \lambda^{(k+1)},$$

wobei $\lambda^{(k+1)}$ der zu $s^{(k)}$ gehörige Lagrange-Multiplikator der linearisierten KKT-Bedingungen ist.

Insgesamt erhält man durch dieses Vorgehen eine Folge von Näherungen, die durch Lösen einer Folge von quadratischen Optimierungsproblemen entsteht (SQP-Verfahren)

Bemerkung 5.4 *Quasi-Newton-Varianten von SQP-Verfahren erhält man durch Verwendung von geeigneten Approximationen $B^{(k)}$ in $q_k(x)$ anstelle der Hesse-Matrix $\nabla_x^2 L(x^{(k)}, \lambda^{(k)})$.*

Es bleibt zu klären, wie man allgemeine quadratische Optimierungsprobleme löst:

5.2.3 Quadratische Optimierungsprobleme

Gegeben sei ein quadratisches Zielfunktional der Form

$$q(x) = g^\top x + \frac{1}{2} x^\top B x.$$

Gesucht ist die Lösung $x^* \in \mathbb{R}^n$ des folgenden quadratischen Optimierungsproblems:

$$q(x^*) = \min_{x \in \mathbb{R}^n} q(x)$$

unter linearen Nebenbedingungen der Form

$$\begin{aligned} a_i^\top x &= b_i \quad \text{für alle } i \in I_1, \\ a_i^\top x &\leq b_i \quad \text{für alle } i \in I_2. \end{aligned}$$

Die Lagrange-Funktion hat dann folgende Form

$$L(x, \lambda) = g^\top x + \frac{1}{2} x^\top B x + \sum_{i=1}^m \lambda_i (a_i^\top x - b_i) = g^\top x + \frac{1}{2} x^\top B x + \lambda^\top (A^\top x - b)$$

mit der Matrix $A = (a_1 \ a_2 \ \dots \ a_m)$ und dem Vektor $b = (b_1 \ b_2 \ \dots \ b_m)^\top$. Für den Gradienten der Lagrange-Funktion bezüglich x erhält man

$$\nabla_x L(x, \lambda) = g + Bx + A\lambda.$$

Wir diskutieren zunächst den einfacheren Fall, dass nur Gleichungsnebenbedingungen vorhanden sind: $I_2 = \emptyset$. In diesem Fall lauten die KKT-Bedingungen

$$\begin{aligned} g + Bx + A\lambda &= 0, \\ A^T x &= b. \end{aligned}$$

Die Berechnung der Lösung eines quadratischen Optimierungsproblems mit Gleichungsnebenbedingungen führt also auf die Lösung eines linearen Gleichungssystems der Form

$$\begin{pmatrix} B & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -g \\ b \end{pmatrix}.$$

Wesentlich schwieriger ist der Fall $I_2 \neq \emptyset$. Würde man die aktive Indexmenge $\mathcal{A}^* = \mathcal{A}(x^*)$ kennen, könnte man das quadratische Optimierungsproblem auf den leichteren Fall eines quadratischen Optimierungsproblems mit Gleichungsnebenbedingungen zurückführen. Im Folgenden wird eine Strategie diskutiert, wie man iterativ diese aktive Indexmenge finden kann:

Eine Aktive-Indexmengen-Strategie:

Sei $x^{(k)} \in \mathbb{R}^n$ eine Näherung von x^* und sei $\mathcal{A}^{(k)} \subset I_1 \cup I_2$ eine Näherung von \mathcal{A}^* . Wir fordern, dass $x^{(k)}$ alle Gleichungs- und Ungleichungsnebenbedingungen des ursprünglichen Optimierungsproblem erfüllt, also zulässig ist, und dass

$$I_1 \subset \mathcal{A}^{(k)} \subset \mathcal{A}(x^{(k)}).$$

Aus dieser Bedingung folgt: $a_i^\top x^{(k)} = b_i$ für alle $i \in \mathcal{A}^{(k)}$.

Schritt 1: Bestimme eine Lösung $s^{(k)} \in \mathbb{R}^n$ des folgenden Optimierungsproblems:

$$q(x^{(k)} + s^{(k)}) = \min_{s \in \mathbb{R}^n} q(x^{(k)} + s)$$

unter der Nebenbedingung

$$a_i^\top (x^{(k)} + s) = b_i \quad \text{für alle } i \in \mathcal{A}^{(k)}.$$

Dieses quadratische Optimierungsproblem hat nur Gleichungsnebenbedingungen. Die Lösung $s^{(k)}$ mit den dazugehörigen Lagrange-Multiplikatoren $\lambda_i^{(k+1)}$, $i \in \mathcal{A}^{(k)}$, lässt sich durch Lösen eines linearen Gleichungssystems berechnen, siehe oben. Für die restlichen Lagrange-Multiplikatoren setzt man $\lambda_i^{(k+1)} = 0$ für $i \notin \mathcal{A}^{(k)}$.

Wegen

$$q(x^{(k)} + s) = q(x^{(k)}) + g^\top s + \frac{1}{2} s^\top B s$$

und

$$a_i^\top (x^{(k)} + s) - b_i = \underbrace{a_i^\top x^{(k)} - b_i}_{= 0} + a_i^\top s = a_i^\top s$$

lässt sich das obige Optimierungsproblem kompakter schreiben: Bestimme eine Lösung $s^{(k)} \in \mathbb{R}^n$ des folgenden Optimierungsproblems:

$$g^\top s^{(k)} + \frac{1}{2}(s^{(k)})^\top B g^\top s^{(k)} = \min_{s \in \mathbb{R}^n} \left(g^\top s + \frac{1}{2} s^\top B s \right)$$

unter der Nebenbedingung

$$a_i^\top s = 0 \quad \text{für alle } i \in \mathcal{A}^{(k)}.$$

Schritt 2:

- a. Zunächst überprüft man, ob die aktuelle Näherung bereits das ursprüngliche Optimierungsproblem löst: Das ist dann der Fall, wenn

$$s^{(k)} = 0 \quad \text{und} \quad \lambda_i^{(k+1)} \geq 0 \quad \text{für alle } i \in \mathcal{A}^{(k)} \cap I_2,$$

denn dann sind die ursprünglichen KKT-Bedingungen für $x^{(k+1)} = x^{(k)} + s^{(k)}$ und $\lambda^{(k+1)}$ erfüllt. In diesem Fall wird der Algorithmus beendet.

- b. Falls

$$s^{(k)} = 0 \quad \text{aber} \quad \lambda_i^{(k+1)} < 0 \quad \text{für ein } i \in \mathcal{A}^{(k)} \cap I_2,$$

dann setzt man

$$x^{(k+1)} = x^{(k)} \quad \text{und} \quad \mathcal{A}^{(k+1)} = \mathcal{A}^{(k)} \setminus \{i\}.$$

Der Algorithmus wird mit Schritt 1 fortgesetzt.

Begründung: Wird statt der Gleichung $a_i^\top x = b_i$ die Ungleichung $a_i^\top x \leq b_i$, also $a_i^\top x - b_i = \xi_i$ mit $\xi_i \leq 0$ vorgegeben, dann wird das Zielfunktional kleiner, da die Änderungsrate des Zielfunktionals bezüglich einer Störung ξ_i durch $-\lambda_i^{(k+1)} > 0$ gegeben ist und sich daher das Zielfunktional um etwa $\lambda_i^{(k+1)} \xi_i < 0$ verändert. Das Zielfunktional wird also kleiner, wenn man den Index i aus der aktiven Indexmenge entfernt.

- c. Falls

$$s^{(k)} \neq 0 \quad \text{und} \quad x^{(k)} + s^{(k)} \text{ zulässig ist}$$

(für das ursprüngliche Optimierungsproblem), dann setzt man

$$x^{(k+1)} = x^{(k)} + s^{(k)} \quad \text{und} \quad \mathcal{A}^{(k+1)} = \mathcal{A}^{(k)}.$$

Der Algorithmus wird mit Schritt 1 fortgesetzt.

- d. Falls

$$s^{(k)} \neq 0 \quad \text{und} \quad x^{(k)} + s^{(k)} \text{ nicht zulässig ist}$$

(für das ursprüngliche Optimierungsproblem), dann wählt man die maximale Schrittweite $\alpha^{(k+1)}$, für die

$$x^{(k)} + \alpha^{(k+1)} s^{(k)} \text{ zulässig bleibt.}$$

Dann gibt es einen Index $i \in I_2 \setminus \mathcal{A}^{(k)}$ mit

$$a_i^\top (x^{(k)} + \alpha^{(k+1)} s^{(k)}) = b_i.$$

Man setzt

$$x^{(k+1)} = x^{(k)} + s^{(k)} \quad \text{und} \quad \mathcal{A}^{(k+1)} = \mathcal{A}^{(k)} \cup \{i\}.$$

Der Algorithmus wird mit Schritt 1 fortgesetzt.

Bemerkung 5.5 *Moderne Versionen der Aktiven-Indermengen-Strategie erlauben einen gleichzeitigen mehrfachen Indexwechsel.*

Literaturverzeichnis

- [JL13] Michael Jung und Ulrich Langer. *Methode der finiten Elemente für Ingenieure. Eine Einführung in die numerischen Grundlagen und Computersimulation*. Wiesbaden: Springer Vieweg, 2., überarbeitete und erweiterte Auflage, 2013.