III.7 Numerische Verfahren zur Optimierung

In diesem Abschnitt werden einige grundlegende Gedanken zu numerischen Verfahren bei nichtlinearen Optimierungsproblemen skizziert. Ausgehend von einem Startwert $\vec{x}^{(0)}$ versucht man eine Folge von $\vec{x}^{(k)}$, $k = 1, 2, \dots$, zu bestimmen, die sich einer lokalen bzw. globalen Extremalstelle nähert oder auch einer Stelle, an der die notwendigen Optimalitätsbedingungen erfüllt sind. Prinzipiell kann man sich das so vorstellen, dass man versucht, ausgehend von einem $x^{(k)}$, in einer Richtung ein $x^{(k+1)}$ zu suchen, für das der Funktionswert kleiner ist. Wir betrachten im Folgenden beispielhaft grundlegende Verfahren mit unterschiedlichen Differenzierbarkeitsvoraussetzungen an die zu minimierende Funktion.

Zyklische Koordinatenmethode

Hierbei werden nacheinander die Richtungen der Koordinatenachsen als Suchrichtungen verwendet. Ausgehend von $\vec{x}^{(0)}$ sucht man zunächst ein Minimum in Richtung $\vec{e}_1$, von diesem Minimum ausgehend in Richtung $\vec{e}_2$ u.s.w. bis $\vec{e}_n$. Den dann erhaltenen Wert nennen wir $\vec{x}^{(1)}$ und suchen wieder in Richtung $\vec{e}_1$, etc.

Wir betrachten die Methode an einem einfachen Beispiel.

Beispiel 3.7.1: $\min f(x_1, x_2) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$

Hier ist klar, dass f an der Stelle $(2,1)$ ein absolutes Minimum besitzt.

Start: $\vec{x}^{(0)} = (0,3)^T$ mit $f(\vec{x}) = 52$

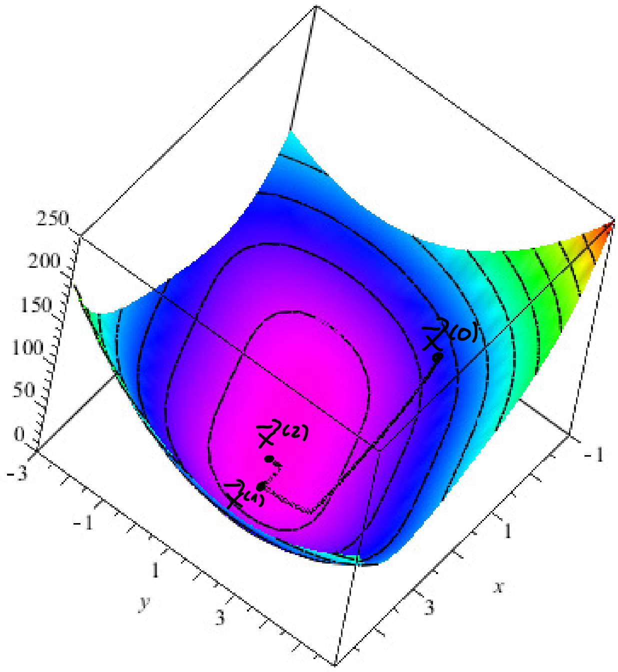
Ausgehend von $\vec{x}^{(0)}$ suchen wir in Richtung $\vec{e}_1 = (1,0)^T$, d.h. wir suchen ein Minimum von $f(0+h, 3) = (h-2)^4 + (h-6)^2 = \tilde{f}(h)$.

Das Minimum von $\tilde{f}(h)$ wird für $h \approx 3.13$ angenommen, d.h. wir suchen nun ausgehend von $(3.13, 3)$ in Richtung $\vec{e}_2 = (0,1)^T$ und müssen daher nun ein Minimum von

$f(3.13, 3+h) = 1.13^4 + (3.13-6-2h)^2 = \tilde{f}(h)$ bestimmen. Hier erhält man $h \approx -1.44$ und somit $\vec{x}^{(1)} = (3.13, 1.56)^T$ mit $f(\vec{x}^{(1)}) = 1.63\ldots$

Ausgehend von $\vec{x}^{(1)}$ sucht man nun zunächst wieder erst in Richtung $\vec{e}_1$, dann in Richtung $\vec{e}_2$ zur Bestimmung von $\vec{x}^{(3)}$.

Die ersten Iterationen sind in den unten stehenden Graphiken eingezeichnet und in der nachstehenden Tabelle zusammengefasst.



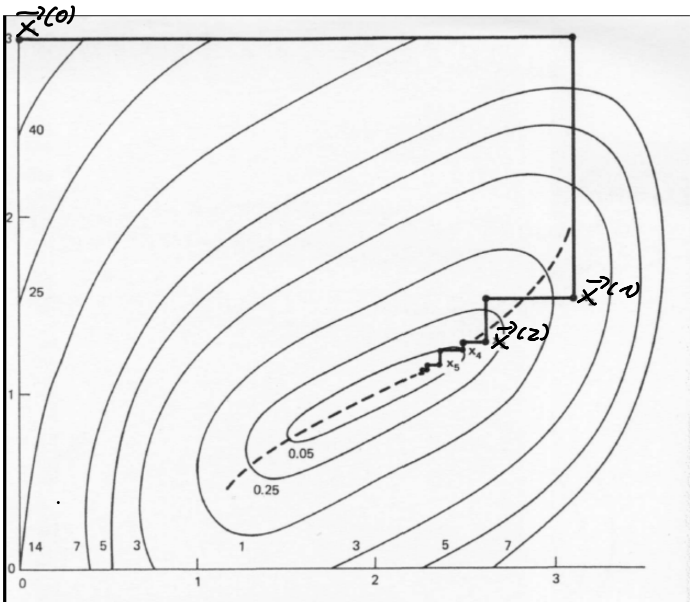


Figure 8.7 Illustration of the cyclic coordinate method.

k	$\vec{x}^{(k)}$ $f(\vec{x}^{(k)})$	Suchrichtung	h	Zwischenwerte
0	$(0,3)^T$ 52	$\vec{e}_1$ $\vec{e}_2$	3.13 -1.44	$(3.13, 3)^T$
1	$(3.13, 1.56)^T$ 1.63	$\vec{e}_1$ $\vec{e}_2$	-0.5 -0.25	$(2.63, 1.56)^T$
2	$(2.63, 1.31)^T$ 0.16	$\vec{e}_1$ $\vec{e}_2$	-0.19 -0.09	$(2.44, 1.31)^T$
3	$(2.44, 1.22)^T$ 0.04	$\vec{e}_1$ $\vec{e}_2$	-0.09 -0.05	$(2.35, 1.22)^T$
4	$(2.35, 1.17)^T$ 0.015	$\vec{e}_1$ $\vec{e}_2$	-0.06 -0.03	$(2.29, 1.17)^T$
5	$(2.29, 1.14)^T$	$\vec{e}_1$ $\vec{e}_2$	-0.04 -0.02	$(2.25, 1.14)^T$
6	$(2.25, 1.12)^T$ 0.004	u.s.w.		

Methode des steilsten Abstiegs

Diese Methode setzt voraus, dass die zu minimierende Funktion eine C^1-Funktion ist. Ausgehend von einem Startwert $\vec{x}^{(0)}$ sucht man ein Minimum der Funktion in Richtung des steilsten Abstiegs, d.h. in Richtung des negativen Gradienten an der Stelle $\vec{x}^{(0)}$. Ausgehend von der Stelle $\vec{x}^{(1)}$ dieses Minimums sucht man dann weiter in Richtung des negativen Gradienten von $\vec{x}^{(1)}$ u.s.w.

Beispiel 3.7.2: $\min f(x_1, x_2) = (x_1-2)^4 + (x_1 - 2x_2)^2$

Es gilt $-\nabla f(\vec{x}) = (-4(x_1-2)^3 - 2(x_1-2x_2),\ 4(x_1-2x_2))^T$

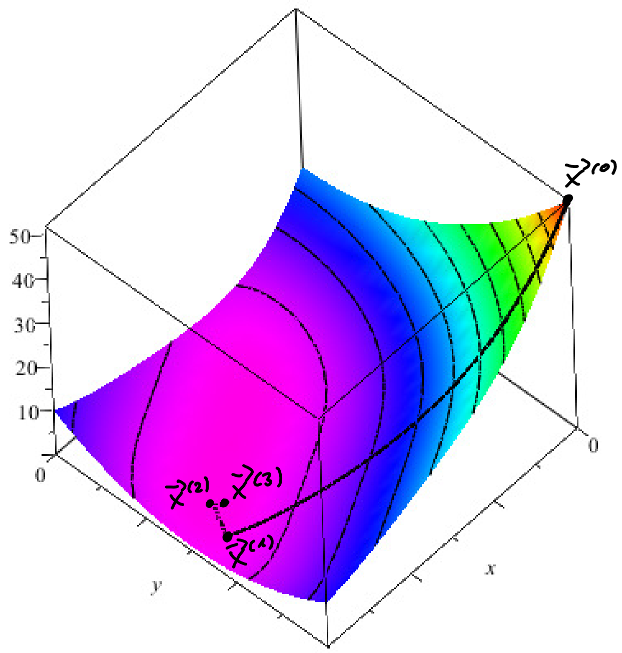
Start: $\vec{x}^{(0)} = (0,3)^T$, $-\nabla f(\vec{x}^{(0)}) = (44, -24)^T$.

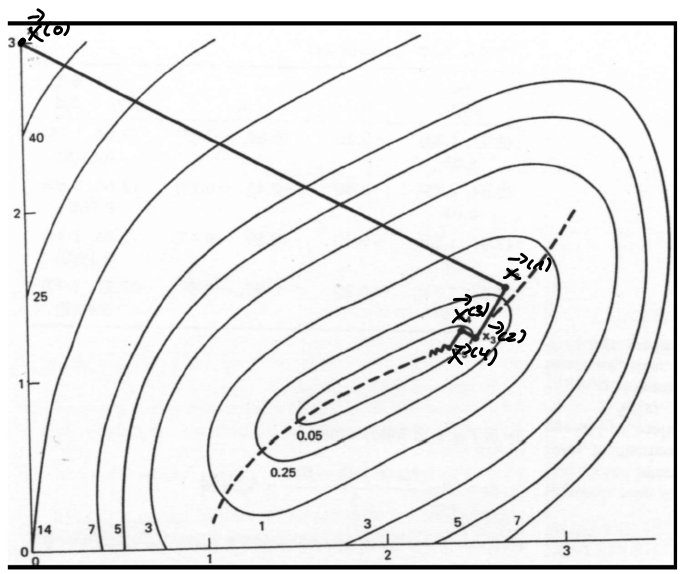
Wir suchen also ein Minimum von

$f(0+44h, 3-24h) = (44h-2)^4 + (92h-6)^2 = \tilde{f}(h)$. Das Minimum wird für $h \approx 0.0615$ angenommen.

Somit ist $\vec{x}^{(1)} = (2.7075, 1.5232)^T$ mit $f(\vec{x}^{(1)}) = 0.3653$.

Für $\vec{x}^{(1)}$ bestimmt man dann $-\nabla f(\vec{x}^{(1)})$ und das Minimum von f in dieser Richtung. Die ersten Iterationen sind in den Graphiken eingezeichnet und in der nachstehenden Tabelle zusammengefasst.





k	$\vec{x}^{(k)}$	$f(\vec{x}^{(k)})$	$-\nabla f(\vec{x}^{(k)})$	h
0	$(0,3)^T$	52	$(44,-24)^T$	0.0615
1	$(2.7075,1.5232)^T$	0.3653	$(-0.7392,-1.3552)^T$	0.2301
2	$(2.5369,1.2104)^T$	0.0966	$(-0.8514,0.4644)^T$	0.1116
3	$(2.4419,1.2622)^T$	0.0450	$(-0.1801,-0.3302)^T$	0.2672
4	$(2.3940,1.1740)^T$	0.0261		

Bei beiden bisher vorgestellten Verfahren werden Minima in bestimmten Richtungen gesucht. Das bedeutet, dass man Minima von Funktionen einer Variablen bestimmen muss, was häufig auch nur iterativ möglich ist. Wir werden dies zum Schluss noch kurz diskutieren.

Newton-Verfahren

Dieses Verfahren beruht darauf, dass man die Funktion an den Iterierten jeweils durch eine quadratische Funktion approximiert. Dazu muss die zu minimierende Funktion eine C^2-Funktion sein.
Die quadratische Approximation an einer Stelle $x^{(k)}$ ist gegeben durch

$$q(\vec{x}) = f(\vec{x}^{(k)}) + (\vec{x}-\vec{x}^{(k)})^T \nabla f(\vec{x}^{(k)}) + \frac{1}{2}(\vec{x}-\vec{x}^{(k)})^T H_f(\vec{x}^{(k)})(\vec{x}-\vec{x}^{(k)})$$

Notwendige Bedingung für ein Minimum von q ist

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

Wir nehmen an, dass $H_f(\vec{x}^{(k)})$ invertierbar ist. Dann ist $x^{(k+1)}$ die eindeutig bestimmte Lösung des linearen Gleichungssystems

$$\nabla q(\vec{x}^{(k+1)}) = \vec{0} \iff \nabla f(\vec{x}^{(k)}) + H_f(\vec{x}^{(k)})(\vec{x}^{(k+1)} - \vec{x}^{(k)})$$

$$\iff \vec{x}^{(k+1)} = \vec{x}^{(k)} - [H_f(\vec{x}^{(k)})]^{-1} f(\vec{x}^{(k)})$$

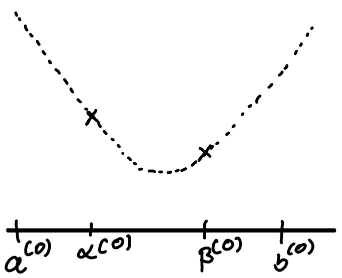
In jedem Iterationsschritt ist also die Hesse-Matrix zu bestimmen und ein lineares Gleichungssystem zu lösen. Statt der Hesse-Matrix verwendet man häufig Näherungen.

Wir haben gesehen, dass die Bestimmung von Minima von Funktionen einer Variablen als Teilproblem bei numerischen Optimierungsverfahren auftritt. Auch dies ist häufig nur näherungsweise möglich.
Wir befassen uns daher zum Abschluss noch beispielhaft mit sogenannten Liniensuchmethoden.
Wir nehmen an, dass wir das Minimum einer strikt konvexen Funktion auf einem Intervall $[a,b]$ bestimmen wollen. Für alle $x,y \in [a,b]$ mit $x \neq y$ und $\lambda \in (0,1)$ soll also gelten:

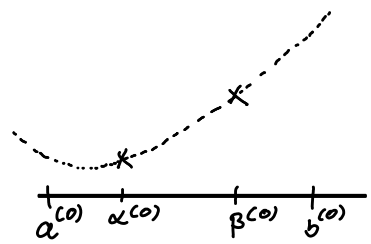
$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$$

Man geht nun so vor, dass man das Intervall, in dem das Minimum liegt, schrittweise durch Vergleich geeigneter Funktionswerte verkleinert.
Wir starten mit dem Intervall $[a^{(0)}, b^{(0)}]$ und bestimmen die Funktionswerte $f(\alpha^{(0)})$, $f(\beta^{(0)})$ an zwei Stellen $\alpha^{(0)}, \beta^{(0)}$, für die $a^{(0)} < \alpha^{(0)} < \beta^{(0)} < b^{(0)}$ sein soll. Grundsätzlich können nun zwei verschiedene Fälle eintreten:

1. Fall: $f(\alpha^{(0)}) > f(\beta^{(0)})$. Dann gilt:
$f(x) > f(\beta^{(0)})$ für alle $x \in [a^{(0)}, \alpha^{(0)})$.
Setze $[a^{(1)}, b^{(1)}] = [\alpha^{(0)}, b^{(0)}]$.



2. Fall: $f(\alpha^{(0)}) \leq f(\beta^{(0)})$. Dann gilt:
$f(x) > f(\alpha^{(0)})$ für alle $x \in (\beta^{(0)}, b^{(0)}]$.
Setze $[a^{(1)}, b^{(1)}] = [a^{(0)}, \beta^{(0)}]$.



Mit dem Intervall $[a^{(1)}, b^{(1)}]$ verfährt man dann genauso wie mit $[a^{(0)}, b^{(0)}]$ u.s.w. In jedem Schritt erhält man ein kleineres Intervall. Ist $[a^{(k)}, b^{(k)}]$ klein genug, so nimmt man die Intervallmitte $\frac{1}{2}(a^{(k)} + b^{(k)})$ als Näherung für die Minimalstelle.

Für die Wahl von $\alpha^{(i)}, \beta^{(i)}$ gibt es prinzipiell natürlich viele Möglichkeiten. Sinnvoll ist es $\alpha^{(i)}, \beta^{(i)}$ so zu wählen, dass das neue Intervall auch signifikant kleiner ist als das alte und die Anzahl der nötigen Funktionsauswertungen klein zu halten.
Im Folgenden wählen wir $\alpha^{(i)}$ rechts und $\beta^{(i)}$ links vom Mittelpunkt des jeweiligen Intervalls mit gleichem Abstand vom Mittelpunkt.
Für $\lambda \in (0, \frac{1}{2})$ sei also

$$\alpha^{(i)} = a^{(i)} + \lambda(b^{(i)} - a^{(i)}) \qquad \beta^{(i)} = b^{(i)} - \lambda(b^{(i)} - a^{(i)})$$
$$= (1-\lambda)\, a^{(i)} + \lambda\, b^{(i)} \qquad = \lambda\, a^{(i)} + (1-\lambda)\, b^{(i)}$$

Bezeichnen wir mit L_i die Länge des Intervalls $[a^{(i)}, b^{(i)}]$, so gilt:

$$L_{i+1} = \beta^{(i)} - a^{(i)} = b^{(i)} - \alpha^{(i)}$$
$$= b^{(i)} - [(1-\lambda)a^{(i)} + \lambda b^{(i)}]$$
$$= (1-\lambda)[b^{(i)} - a^{(i)}] = (1-\lambda)\, L_i$$

Es gilt also die Differenzengleichung $L_{i+1} - (1-\lambda) L_i = 0$ mit $L_0 = b - a$.
Somit $L_i = (1-\lambda)^i (b-a)$. Da $\lambda \in (0, \frac{1}{2})$, wird L_i umso kleiner, je dichter λ am Wert $\frac{1}{2}$ gewählt wird. Im Allgemeinen sind zwei Funktionsauswertungen zur Bestimmung des neuen Intervalls nötig. Im Folgenden bestimmen wir $\lambda \in (0, \frac{1}{2})$ so, dass nur jeweils ein Funktionswert neu bestimmt werden muss.
Für den 1. Fall, d.h. $[a^{(i+1)}, b^{(i+1)}] = [\alpha^{(i)}, b^{(i)}]$ soll gelten:

$$\alpha^{(i+1)} = \beta^{(i)} \iff a^{(i+1)} + \lambda(b^{(i+1)} - a^{(i+1)}) = a^{(i)} + (1-\lambda)(b^{(i)} - a^{(i)})$$
$$\iff \alpha^{(i)} + \lambda(b^{(i)} - \alpha^{(i)}) = a^{(i)} + (1-\lambda)(b^{(i)} - a^{(i)})$$
$$\iff (1-\lambda)\alpha^{(i)} + \lambda b^{(i)} = \lambda a^{(i)} + (1-\lambda) b^{(i)}$$
$$\iff (1-\lambda)[(1-\lambda)a^{(i)} + \lambda b^{(i)}] + \lambda b^{(i)} = \lambda a^{(i)} + (1-\lambda)\, b^{(i)}$$
$$\iff (b^{(i)} - a^{(i)})\,[(1-\lambda)^2 + (1-\lambda) - 1] = 0$$
$$\iff (1-\lambda) = -\frac{1}{2} \pm \frac{\sqrt{5}}{2}$$

Da $\lambda \in (0, \frac{1}{2})$ und somit $(1-\lambda) \in (\frac{1}{2}, 1)$ sein soll, folgt $(1-\lambda) = -\frac{1}{2} + \frac{\sqrt{5}}{2}$, $\lambda = \frac{3}{2} - \frac{\sqrt{5}}{2}$.

$\alpha^{(i)}$ (bzw. $\beta^{(i)}$) teilen die Strecke $\overline{a^{(i)},b^{(i)}}$ im Verhältnis des Goldenen Schnitts, d.h. $\dfrac{\text{Länge}(\overline{a^{(i)},b^{(i)}})}{\text{Länge}(\overline{a^{(i)},\beta^{(i)}})} = \dfrac{\text{Länge}(\overline{a^{(i)},\beta^{(i)}})}{\text{Länge}(\overline{\beta^{(i)},b^{(i)}})}$.

Analoge Betrachtungen und Umformungen führen auch im 2. Fall $[a^{(i+1)},b^{(i+1)}]=[a^{(i)},\beta^{(i)}]$ mit der Forderung $\beta^{(i+1)}=\alpha^{(i)}$ auf dasselbe Ergebnis. Außer beim 1. Schritt ist nur jeweils eine neue Funktionsauswertung erforderlich.

Beispiel 3.7.3: Das Startintervall sei $[a^{(0)},b^{(0)}]=[0,1]$ und die Funktion f strikt konvex auf $[0,1]$. Die Minimalstelle soll mit einer Genauigkeit von 10^{-5} bestimmt werden. Wir überlegen, wie viele Schritte und Funktionsauswertungen erforderlich sind für a) $\lambda_1=\frac{1}{3}$, $1-\lambda_1=\frac{2}{3}$ b) $\lambda_2=\frac{3}{2}-\frac{\sqrt{5}}{2}$, $1-\lambda_2=-\frac{1}{2}+\frac{\sqrt{5}}{2}$.

Da wir als Näherung für die Minimalstelle den Mittelpunkt des letzten bestimmten Intervalls verwenden wollen, soll also gelten $\frac{1}{2}\cdot L_i \le 10^{-5}$, d.h. (vgl. S. 6)

$$\frac{1}{2}(1-\lambda)^i \le 10^{-5} \iff i\cdot \underbrace{\ln(\underbrace{1-\lambda}_{<1})}_{<0} \le \ln(2\cdot 10^{-5})$$

$$\iff i > \frac{\ln(2\cdot 10^{-5})}{\ln(1-\lambda)}$$

a) $\lambda_1=\frac{1}{3}$, dann ist $\dfrac{\ln(2\cdot 10^{-5})}{\ln(1-\lambda_1)} = 26.68\ldots$

Also sind 27 Schritte mit insgesamt 54 Funktionsauswertungen erforderlich.

b) $\lambda_2=\frac{3}{2}-\frac{\sqrt{5}}{2}$, dann ist $\dfrac{\ln(2\cdot 10^{-5})}{\ln(1-\lambda_2)} = 22.48\ldots$

Also sind 23 Schritte mit insgesamt 24 Funktionsauswertungen erforderlich.

Beispiel 3.7.4. Für $f(x)=x^2+2x$ starten wir mit dem Intervall $[a^{(0)},b^{(0)}]=[-3,3]$. In den nachfolgenden Tabellen sind die Intervalle $[a^{(i)},b^{(i)}]$ für $\lambda_1=\frac{1}{3}$ und für $\lambda_2=\frac{3}{2}-\frac{\sqrt{5}}{2}$ angegeben.

$\lambda_1 = \frac{1}{3}$

[-3., 3]
[-3., 1]
[-3., -0.3333333333]
[-2.111111111, -0.3333333333]
[-1.518518519, -0.3333333333]
[-1.518518519, -0.7283950617]
[-1.255144033, -0.7283950617]
[-1.255144033, -0.9039780521]
[-1.138088706, -0.9039780521]
[-1.060051821, -0.9039780521]
[-1.060051821, -0.9560026419]

$\lambda_2 = \frac{3}{2} - \frac{\sqrt{5}}{2}$

[-3., 3]
[-3., 0.708203928]
[-1.583592134, 0.708203928]
[-1.583592134, -0.167184274]
[-1.583592134, -0.7082039356]
[-1.249223595, -0.7082039356]
[-1.249223595, -0.9148550573]
[-1.121506178, -0.9148550573]
[-1.042572474, -0.9148550573]
[-1.042572474, -0.9636387696]
[-1.012422482, -0.9636387696]