

TECHNISCHE UNIVERSITÄT BERLIN

Fakultät II – Institut für Mathematik

Nichtlineare Optimierung

Vorlesung im Sommersemester 2019

Dietmar Hömberg

Inhaltsverzeichnis

1	Optimierungsaufgaben	1
1.1	Literatur	1
1.2	Grundlagen und Begriffe	1
1.3	Existenz von Lösungen	4
1.4	Konvexe Optimierungsaufgaben	7
1.5	Weitere wichtige Beispiele von Optimierungsaufgaben	9
1.6	Numerische Lösung von Optimierungsaufgaben	12
1.7	Optimierungs-Software	12
1.7.1	Programmbibliotheken	12
1.7.2	Interaktive Programmsysteme	12
2	Ableitungsfreie Verfahren	13
2.1	Simplexverfahren von Nelder und Mead	13
2.1.1	Grundkonstruktionen	13
2.1.2	Ablauf des Verfahrens	15
2.2	Mutations-Selektions-Verfahren	17
2.3	Anwendung: Nichtlineare Regression	17
3	Probleme ohne Restriktionen – Theorie	18
3.1	Optimalitätsbedingungen	18
3.1.1	Notwendige Bedingungen erster Ordnung	18
3.1.2	Notwendige Bedingungen zweiter Ordnung	20
3.1.3	Hinreichende Bedingungen zweiter Ordnung	21
3.2	Konvexe Optimierungsaufgaben	23
4	Probleme ohne Restriktionen – Verfahren	27
4.1	Grundlagen	27
4.2	Das Newton-Verfahren	29
4.3	Abstiegsverfahren – allgemeine Aussagen	32
4.3.1	Effiziente Schrittweiten	32
4.3.2	Gradientenbezogene Richtungen	33
4.3.3	Allgemeine Konvergenzsätze	35
4.4	Schrittweitenverfahren	37

4.4.1	Exakte Schrittweite	37
4.4.2	Schrittweite nach Armijo	39
4.4.3	Schrittweite nach Powell	41
4.5	Das Gradientenverfahren	43
4.6	Gedämpftes Newton-Verfahren	44
4.6.1	Das Verfahren	44
4.6.2	Interpretation der Newton-Richtung	44
4.6.3	Konvergenz des Verfahrens	45
4.7	Variable Metrik- und Quasi-Newton-Verfahren	46
4.7.1	Allgemeine Verfahrensvorschrift	46
4.7.2	Globale Konvergenz von Variable-Metrik-Verfahren	47
4.7.3	Quasi-Newton-Methoden	47
4.7.4	BFGS-Update	48
4.7.5	Das BFGS-Verfahren für quadratische Optimierungsprobleme	50
4.7.6	Das BFGS-Verfahren für nichtlineare Optimierungsaufgaben	52
4.8	Verfahren konjugierter Richtungen	52
4.8.1	CG-Verfahren für quadratische Optimierungsprobleme	52
4.8.2	Analyse des CG-Verfahrens	55
4.8.3	Vorkonditionierung	56
4.8.4	CG-Verfahren für nichtlineare Optimierungsprobleme	57
4.9	Trust-Region-Verfahren	57
4.9.1	Motivation	57
4.9.2	Trust-Region-Newton-Verfahren	58
5	Probleme mit Restriktionen – Theorie	62
5.1	Einführende Beispiele	62
5.2	Tangentialkegel und Regularitätsbedingung für die Restriktionen	66
5.3	Notwendige Optimalitätsbedingungen erster Ordnung	70
5.4	Beweis von Satz 5.3.1	72
5.5	Optimalitätsbedingungen 2. Ordnung	75
5.6	Probleme mit Variationsbeschränkungen (box constraints)	81
5.7	Weitere Regularitätsbedingungen	83
5.8	Geometrische Interpretation der notwendigen Optimalitätsbedingungen . .	85
5.9	Langrange'sche Multiplikatoren und Sensitivität	87

5.10	Dualität	88
5.11	Ausblick: Numerische Lösung nichtlineare Optimierungsprobleme mit Nebenbedingungen	93
6	Probleme mit linearen Restriktionen-Verfahren	94
6.1	Quadratische Optimierungsprobleme	94
6.1.1	Aufgaben mit Gleichungsrestriktionen	94
6.1.2	Aufgaben mit Ungleichungsrestriktionen	98
6.2	Gleichungsnebenbedingungen nichtquadratischer Zielfunktion	106
6.3	Ungleichungsnebenbedingungen – nichtquadratische Zielfunktionen	110
7	Probleme mit nichtlinearen Restriktionen-Verfahren	114
7.1	Das Lagrange-Newton-Verfahren	114
7.2	Sequentielle quadratische Optimierung	116
8	Penalty, Barrier und Augmentierte Lagrange-Methoden	118
8.1	Die quadratische Penalty-Methode	118
8.2	Die logarithmische Barrier-Methode	122
8.3	Augmentierte Lagrange-Methoden	123

1 Optimierungsaufgaben

1.1 Literatur

Grundlage des Skriptes ist das Buch von Prof. Dr. Walter Alt, Universität Jena [1], insbesondere für Kap. 1-4, 6,7 und das Buch von Nocedal und Wright [2] für Kap. 5 und 8. Es basiert auf einer Ausarbeitung von F. Tröltzsch mit Erweiterungen durch D. Hömberg.

1. Alt, W., *Nichtlineare Optimierung*. Vieweg, Braunschweig/Wiesbaden 2002.
2. Nocedal, J. and Wright, S.J., *Numerical Optimization*. Springer, New York 2006.
3. Gill, P.E., Murray, W., and M.H. Wright, *Practical Optimization*. Academic Press, London 1981.
4. Kelley, C.T., *Iterative Methods for Optimization*. SIAM, Philadelphia 1999.
5. Spelluci, P., *Numerische Verfahren der nichtlinearen Optimierung*. Birkhäuser, Basel 1993.
6. Luenberger, D.G., *Optimization by Vector Space Methods*. Wiley, 1969.
7. Luenberger, D.G., *Linear and Nonlinear Programming*. Addison Wesley, London 1984.
8. Großmann, C. und Terno, J., *Numerik der Optimierung*. Teubner-Verlag, Stuttgart 1993.
9. Moré, J.J. and Wright, S.J., *Optimization Software Guide*. SIAM, Philadelphia 1993.

1.2 Grundlagen und Begriffe

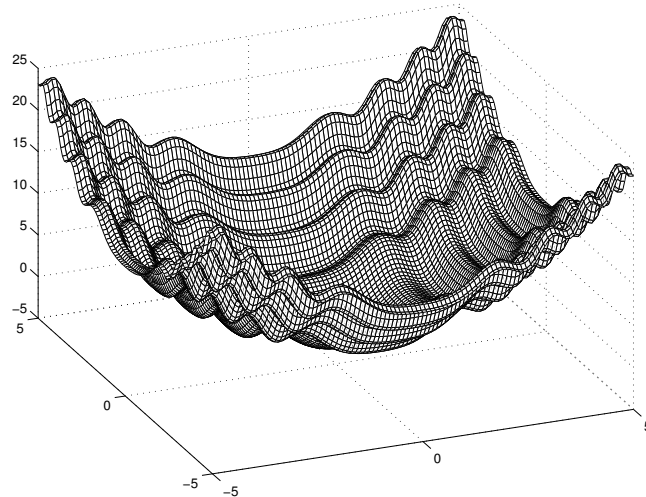
Wir untersuchen in diesem Kurs die Aufgabe, ein Minimum einer gegebenen Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ zu berechnen. Dazu einige Beispiele und Grundbegriffe:

Beispiel 1.2.1

- $f(x) = x^2, f : \mathbb{R} \rightarrow \mathbb{R}$ hat genau ein Min bei $\tilde{x} = 0$.
- $f(x) = x, f : \mathbb{R} \rightarrow \mathbb{R}$ ist nicht nach unten beschränkt; die Minimaufgabe ist unlösbar.

Beispiel 1.2.2

$f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) - \cos(x_1^2) - \cos(x_2^2), f : \mathbb{R}^2 \rightarrow \mathbb{R}$ hat ein (strenges) **globales** Minimum und mehrere (strenge) **lokale** Minima und Maxima.



Bezeichnungen:

Für $x \in \mathbb{R}^n$ heißt

$\ x\ = \left(\sum_{i=1}^n x_i^2 \right)^{1/2}$	euklidische Norm
$B(x, r) = \{y \in \mathbb{R}^n \mid \ y - x\ < r\}$	offene Kugel
$\overline{B}(x, r) = cl\ B(x, r)$	abgeschlossene Kugel
x_i	i-te Komponente von x
$x^{(k)}$	k-tes Glied einer Folge von Vektoren $(x^{(k)})_{k=1}^\infty$

Im Weiteren sei $D \subset \mathbb{R}^n$ eine fest gegebene offene Menge, eine Menge $\Omega \subset D$ sowie $f : D \rightarrow \mathbb{R}$. Wir betrachten das Optimierungsproblem

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

Man nennt f - **Zielfunktion**

Ω - **zulässiger Bereich** oder zulässige Menge

Im Fall $\Omega = D$ heißt (P) **unrestringierte** oder **freie Optimierungsaufgabe** (wie etwa in Bsp. 1.2.1). Das mutet für eine echte Teilmenge $D \subset \mathbb{R}^n$ etwas eigenartig an, lässt sich aber leicht einsehen. Ist Ω durch Nebenbedingungen gegeben, so heißt (P) **Optimierungsproblem mit Nebenbed.** oder **restringiertes Opt.-problem**. In der Regel ist Ω durch Gleichungen und Ungleichungen definiert. Die Elemente von Ω heißen **zulässige Punkte**.

Beispiel 1.2.3

$$\begin{array}{ll} \min & x^3 \\ & x \in \mathbb{R} \\ \text{bei} & x \geq 1 \end{array}$$

Es handelt sich um eine Aufgabe mit einer linearen Ungleichungsrestriktion, $D = \mathbb{R}$, $\Omega = [1, \infty)$, die Lösung ist $\tilde{x} = 1$.

Definition 1.2.1 Ein Punkt $\tilde{x} \in \Omega$ heißt

- **lokales Minimum** von f auf Ω oder **lokale Lösung** von (P) , wenn $\exists r > 0$, so dass

$$f(x) \geq f(\tilde{x}) \quad \forall x \in \Omega \cap B(\tilde{x}, r)$$

- *analog strenges lokales Minimum*, wenn entsprechend

$$f(x) > f(\tilde{x}) \quad \forall x \in \Omega \cap B(\tilde{x}, r), \quad x \neq \tilde{x}$$

- *analog globales Minimum bzw. globale Lösung*, wenn

$$f(x) \geq f(\tilde{x}) \quad \forall x \in \Omega$$

- *strenges globales Min. bzw. strenge globale Lösung*, wenn

$$f(x) > f(\tilde{x}) \quad \forall x \in \Omega, \quad x \neq \tilde{x}$$

gilt.

Bei nichtlinearen Optimierungsaufgaben können viele lokale oder globale Minima auftreten, wie etwa bei $f(x) = \sin x$, $f(x) = x \sin\left(\frac{1}{x}\right)$.

Bemerkung: Wir entwickeln unsere Theorie für den Fall der Minimierung von f . Den Fall der Suche nach Maxima \tilde{x} ,

$$f(x) \leq f(\tilde{x}) \quad \forall x \in \Omega$$

führen wir wegen der Äquivalenz zu $-f(x) \geq -f(\tilde{x})$ auf die Minimierung von $\tilde{f} := -f$ zurück.

Beispiel 1.2.4 (Linear Regression)

Gesucht ist eine lineare Funktion

$$\eta(\xi) = x_1 \xi + x_2$$

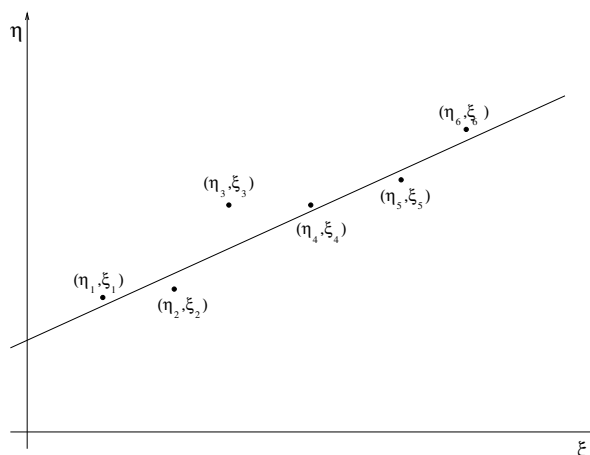
mit unbekannten Koeffizienten x_1, x_2 , welche am besten zu gegebenen Wertepaaren (ξ_i, η_i) , $i = 1, \dots, m$ (z.B. Messwerten) passt. Wir setzen

$$\eta(\xi) = g(x_1, x_2, \xi) = x_1 \xi + x_2$$

und wollen x_1, x_2 so wählen, dass die Zielfunktion

$$\begin{aligned}
f(x) = f(x_1, x_2) &= \sum_{i=1}^m (\eta_i - g(x_1, x_2, \xi_i))^2 \\
&= \sum_{i=1}^m (\eta_i - x_1 \xi_i - x_2)^2
\end{aligned}$$

minimiert wird. f ist ein Polynom zweiten Grades in x_1, x_2 , also eine quadratische Zielfunktion.



Folgende Fragestellungen werden wir in der Vorlesung vor allem untersuchen:

- Existenz und Eindeutigkeit von Lösungen
- Notwendige Optimalitätsbedingungen
- Hinreichende Optimalitätsbedingungen
- Numerische Verfahren zur Lösung von Optimierungsaufgaben.

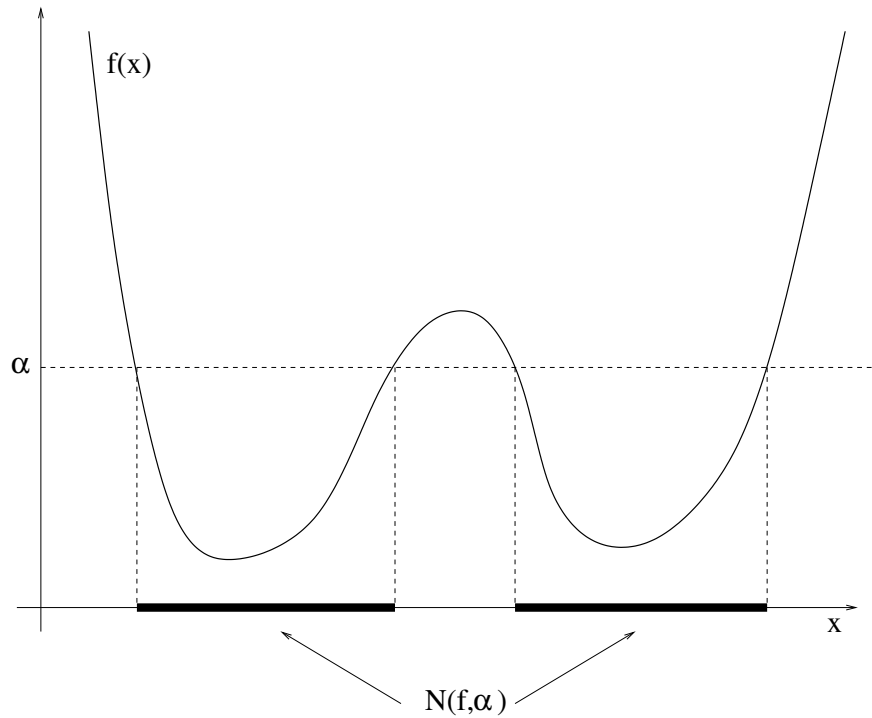
1.3 Existenz von Lösungen

Grundlage für die meisten Existenzbeweise ist der bekannte

Satz 1.3.1 (Weierstraß)

Ist $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ stetig und $K \subset D$ kompakt, dann nimmt f auf K sein Infimum (bzw. sein Supremum) an, d. h., es existiert ein globales Minimum (bzw. Maximum) von f auf K .

Definition 1.3.1 $f : D \rightarrow \mathbb{R}, D \subset \mathbb{R}^n, \alpha \in \mathbb{R}$. Die Mengen $N(f, \alpha) = \{x \in D \mid f(x) \leq \alpha\}$ heißen **Niveaumengen** von f .



Satz 1.3.2 $D \subset \mathbb{R}^n$, $f : D \rightarrow \mathbb{R}$ stetig und $\Omega \subset D$ abgeschlossen. Für mindestens ein $w \in \Omega$ sei die Niveaumenge

$$N(f, f(w)) = \{x \in D \mid f(x) \leq f(w)\}$$

kompakt. Dann gibt es (mindestens) ein globales Minimum von f auf Ω .

Beweis: Es sei $\alpha = \inf_{x \in \Omega} f(x)$. Offenbar gilt $\alpha \leq f(w)$. $\Omega \cap N(f, f(w))$ ist kompakt, und nur in dieser Menge können Elemente von Ω liegen, deren Funktionswerte kleiner oder gleich $f(w)$ sind. Somit

$$\alpha = \inf_{x \in \Omega \cap N(f, f(w))} f(x) = f(\tilde{x}),$$

wobei $\tilde{x} \in \Omega$ nach dem Satz von Weierstraß existiert. □

Typische Anwendungen dieses Prinzips sind die folgenden zwei Aussagen:

Folgerung 1.3.1 D, Ω wie in Satz 1.2.1, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig. Zusätzlich habe f die Eigenschaft

$$\lim_{\|x\| \rightarrow \infty} f(x) = \infty.$$

Dann besitzt die Aufgabe

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

mindestens eine globale Lösung.

Beweis: Wegen $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$ sind alle Niveaumengen $N(f, \alpha)$ kompakt (Übungsaufgabe). Der Rest ist Folgerung aus dem letzten Satz. \square

Eine (n, n) -Matrix H heißt **positiv semidefinit**, wenn $x^T H x \geq 0$ für alle $x \in \mathbb{R}^n$ gilt, sowie **positiv definit**, wenn

$$x^T H x > 0 \quad \forall x \in \mathbb{R}^n, x \neq 0$$

Man zeigt mit einem Kompaktheitsschluss, dass positive Definitheit äquivalent ist zur Existenz eines $\alpha > 0$, so dass

$$x^T H x \geq \alpha \|x\|^2 \quad \forall x \in \mathbb{R}^n$$

(Übungsaufgabe). Offenbar gilt dann $x^T H x \rightarrow \infty, \|x\| \rightarrow \infty$.

Beispiel 1.3.1 (Unrestringierte quadratische Optimierungsaufgabe)

Wir betrachten

$$(QU) \quad \min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T H x + b^T x$$

mit gegebenem $b \in \mathbb{R}^n$ und positiv definiten (n, n) -Matrix H . Sie zeigen leicht, dass $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$ gilt (Übungsaufgabe). Wegen Folgerung 1.3.1 hat (QU) damit mindestens eine globale Lösung.

Beispiel 1.3.2 (Lineare Regression aus Bsp 1.2.4)

Ausmultiplizieren der Zielfunktion ergibt

$$\begin{aligned} f(x) &= \sum_{i=1}^m (\eta_i - (x_1 \xi_i + x_2))^2 \\ &= \underbrace{\sum_{i=1}^m \eta_i^2}_c - 2 \underbrace{\sum_{i=1}^m \eta_i (\xi_i x_1 + x_2)}_{b^T x} + \underbrace{\sum_{i=1}^m (x_1 \xi_i + x_2)^2}_{\frac{1}{2} x^T H x} \\ &= \frac{1}{2} x^T H x + b^T x + c \\ \text{mit } H &= 2 \begin{pmatrix} \sum_{i=1}^m \xi_i^2 & \sum_{i=1}^m \xi_i \\ \sum_{i=1}^m \xi_i & m \end{pmatrix} \quad b = -2 \begin{pmatrix} \sum_{i=1}^m \xi_i \eta_i \\ \sum_{i=1}^m \eta_i \end{pmatrix}. \end{aligned}$$

Sind mindestens zwei der ξ_i verschieden, so ist H positiv definit (Übungsaufgabe).

Damit ist die Aufgabe der linearen Regression in diesem Fall lösbar. Sind alle ξ_i gleich, so ist sie auch nicht sinnvoll gestellt!

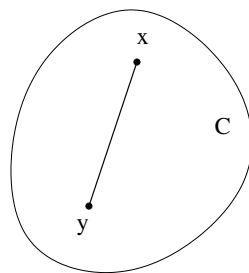
1.4 Konvexe Optimierungsaufgaben

Unter allen Optimierungsaufgaben haben konvexe die schönsten Eigenschaften! Es gibt dazu auch eine gut ausgebaute **konvexe Analysis** (z. B. siehe Webster, R., Convexity. Oxford University Press 1994, oder Rockafellar, R.T., Convex Analysis. Princeton University Press 1970).

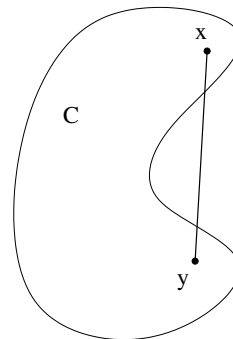
Definition 1.4.1 Eine Menge $C \subset \mathbb{R}^n$ heißt *konvex*, falls für je 2 beliebige $x, y \in C$ auch die Strecke

$$[x, y] = \{z = (1 - t)x + ty \mid 0 \leq t \leq 1\}$$

in C enthalten ist: $x, y \in C \Rightarrow [x, y] \subset C$.



konvexe Menge



nicht konvexe Menge

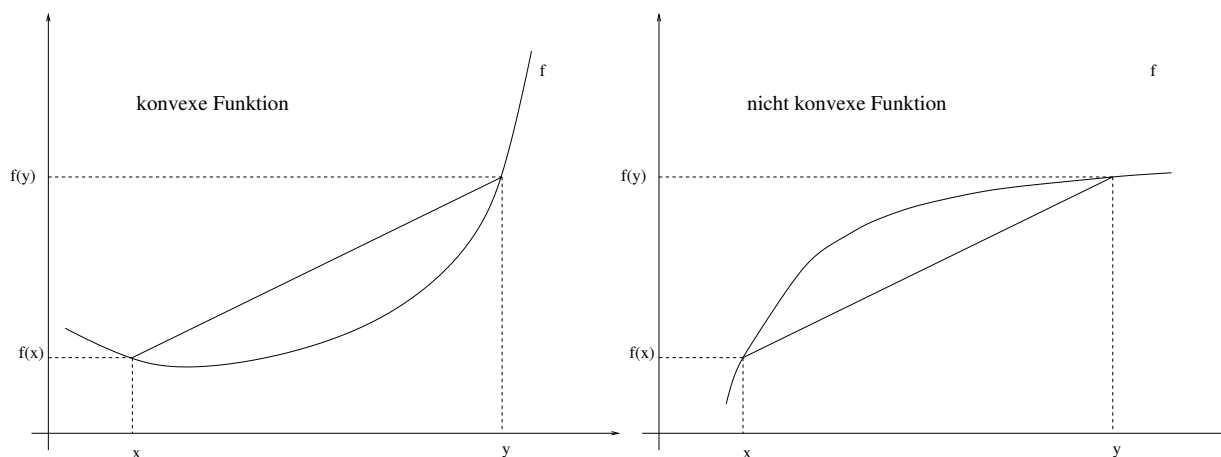
Definition 1.4.2 Sei $C \subset \mathbb{R}^n$ konvex und nichtleer, $C \subset D$. Eine Funktion $f : D \rightarrow \mathbb{R}$ heißt **konvex** auf C , wenn

$$\boxed{f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y)} \quad \begin{array}{l} \forall x, y \in C, \\ \forall t \in [0, 1] \end{array}$$

Gilt die verschärfte Beziehung

$$f((1 - t)x + ty) < (1 - t)f(x) + tf(y) \quad \begin{array}{l} \forall x, y \in C, x \neq y \\ \forall t \in]0, 1[, \end{array}$$

so heißt f **strikt** oder **streng** konvex auf C .



Beispiel 1.4.1 $f(x) = x$ ist konvex, $f(x) = x^2$ streng konvex.

Nun betrachten wir $f : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, nichtleer, $\Omega \subset D$ konvex. Ist f konvex auf Ω , so heißt das Problem

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

konvexe Optimierungsaufgabe.

Schon der nächste Satz zeigt, welche schönen Eigenschaften konvexe Probleme haben.

Satz 1.4.1 Die obige Aufgabe (P) sei eine konvexe Optimierungsaufgabe. Dann ist jedes lokale Minimum von (P) auch ein globales. Die Menge aller Lösungen von (P) ist konvex.

Beweis:

- (i) Es sei \tilde{x} lokale Lösung, d. h., mit einem $r > 0$ gilt

$$f(x) \geq f(\tilde{x}) \quad \forall x \in \Omega \cap B(\tilde{x}, r). \quad (*)$$

Zu zeigen ist

$$f(y) \geq f(\tilde{x}) \quad \forall y \in \Omega.$$

Wir wählen ein $y \in \Omega$ und betrachten $\tilde{x} + t(y - \tilde{x})$ für kleine $t > 0$.

Es gilt $\tilde{x} + t(y - \tilde{x}) \in \Omega \quad \forall t \in [0, 1]$: Denn

$$\tilde{x} + t(y - \tilde{x}) = (1 - t)\tilde{x} + ty \in \Omega, \text{ da } \Omega \text{ konvex.}$$

Es gilt auch $\tilde{x} + t(y - \tilde{x}) \in B(\tilde{x}, r)$ für alle hinreichend kleinen t , d. h. $t \in [0, t_0]$, $t_0 > 0$.
Deshalb wegen (*)

$$f(\tilde{x}) \underset{(*)}{\leq} f((1 - t)\tilde{x} + ty) \leq (1 - t)f(\tilde{x}) + tf(y).$$

Durch Umstellen ergibt sich $f(\tilde{x}) \leq f(y)$.

(ii) Nun seien x und \tilde{x} Lösungen, d.h. $f(\tilde{x}) = f(x) = \tilde{\alpha} = \min f$. Dann

$$f((1-t)\tilde{x} + tx) \leq (1-t)f(\tilde{x}) + tf(x) = \tilde{\alpha} = \min$$

\Rightarrow auch $(1-t)\tilde{x} + tx$ ist Lösung. □

Satz 1.4.2 $D \subset \mathbb{R}^n, \Omega \subset D$ konvex, $\Omega \neq \emptyset, f : D \rightarrow \mathbb{R}$ streng konvex. Hat (P) eine Lösung \tilde{x} , dann ist \tilde{x} eindeutig bestimmt und ein strenges Minimum von f in Ω .

Beweis: Es seien x, y zwei Minima von f , also nach dem letzten Satz globale Minima. Damit gilt

$$f(x) = f(y) = \alpha = \min_{x \in \Omega} f(x).$$

Angenommen, es gilt $x \neq y$. Dann liefert $z = \frac{1}{2}(x + y)$ einen kleineren Wert als α , denn

$$f(z) = f\left(\frac{1}{2}x + \frac{1}{2}y\right) \underset{\substack{\uparrow \\ \text{strenge Konvexität}}}{<} \frac{1}{2}f(x) + \frac{1}{2}f(y) = \frac{1}{2}\alpha + \frac{1}{2}\alpha = \alpha.$$

Außerdem gilt $z \in \Omega$ und insgesamt widerspricht das der Optimalität von x, y . Damit $x = y$, strenges Minimum □

Beispiel 1.4.2

$$f(x) = \frac{1}{2}x^T Hx + b^T x$$

Ist H positiv definit, so ist f streng konvex (Übungsaufg.).

Folgerung: Sind zwei der Werte ξ_i verschieden, so ist das Zielfunktional bei der Aufgabe der linearen Regression streng konvex und daher die Lösung eindeutig bestimmt.

1.5 Weitere wichtige Beispiele von Optimierungsaufgaben

Beispiel 1.5.1 (Nichtlineare Regression)

Bei der linearen Regression war eine affin-lineare Funktion $\eta(\xi) = x_1\xi + x_2$ zu bestimmen. Allgemeiner kann η eine nichtlineare Funktion von ξ sein, gegeben durch einen nichtlinearen Ansatz

$$\eta(\xi) = g(x_1, x_2, \xi)$$

oder allgemeiner

$$\eta(\xi) = g(x_1, \dots, x_n, \xi)$$

mit einem unbekannten Vektor $x \in \mathbb{R}^n$, z. B.

$$g = x_1 e^{\xi x_2} + x_3.$$

\Rightarrow Minimierung von

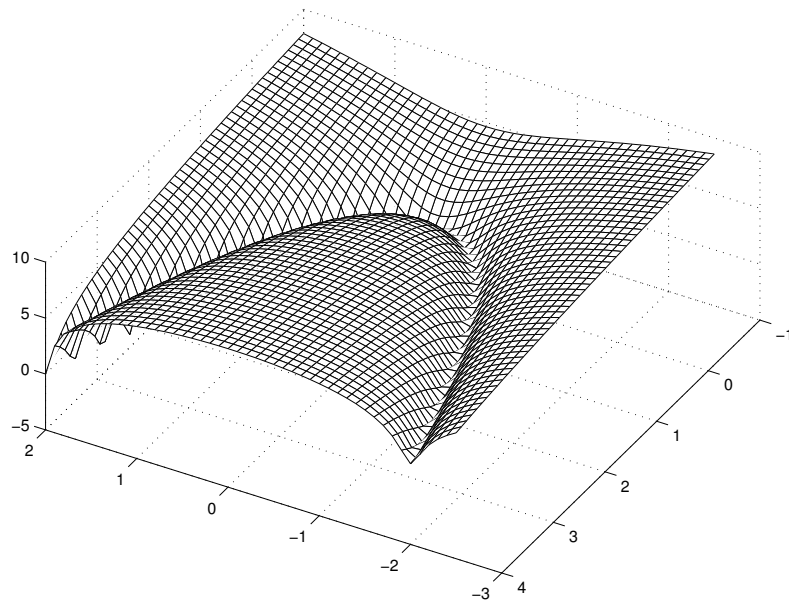
$$f(x) = \sum_{i=1}^m (\eta_i - g(x, \xi_i))^2$$

$$\text{Typ: } f(x) = \sum_{i=1}^m (f_i(x))^2 \quad f_i = \eta_i - g(\cdot, \xi_i).$$

Nun betrachten wir noch einige berühmte pathologische Testfunktionen, an denen gern Algorithmen getestet werden.

Beispiel 1.5.2 *Rosenbrock-Funktion („Banana shaped valley“)*

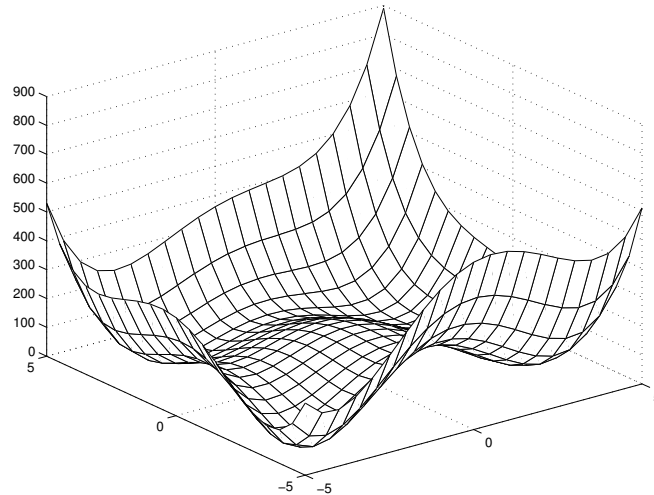
$$f(x_1, x_2) = \underbrace{100(x_2 - x_1^2)^2}_{\substack{\text{definiert} \\ \text{das Tal} \\ \text{(Parabel)}}} + \underbrace{(1 - x_1)^2}_{\substack{\text{kippt leicht} \\ \text{an}}}$$



Beispiel 1.5.3 (Himmelblau)

$$f(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

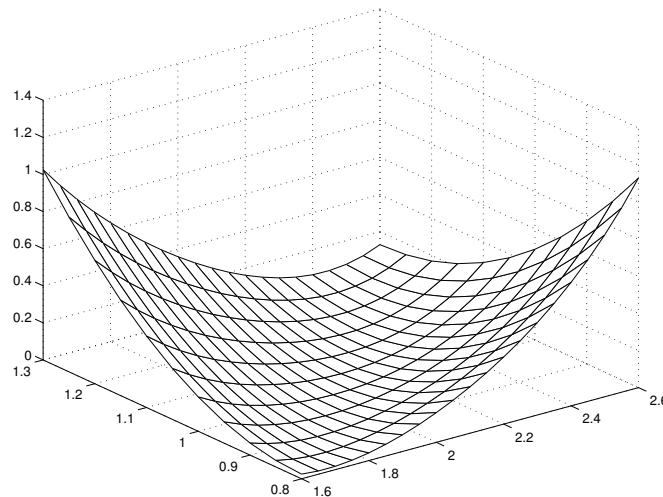
4 lokale Minimalstellen, die zugleich globale Minimalstellen mit Funktionswert 0 sind;
4 Sattelpunkte und ein lokales Maximum bei $(-0.270845, -0.923039)^T$.



Beispiel 1.5.4 (Bazaraa-Shetty)

$$f(x_1, x_2) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

Globales Min. bei $(2, 1)$. Die Hesse-Matrix ist an dieser Stelle singulär, was bei manchen Algorithmen zu Problemen führen kann.



Beispiel 1.5.5

$$f(x_1, \dots, x_5) = 2x_1^2 + 2x_2^2 + x_3^2 + x_4^2 + \frac{1}{2}x_5^2 - 4(x_1 + x_2) - 2(x_3 + x_4) - x_5 + 6.5$$

Globales Min. bei $\tilde{x} = (1, 1, 1, 1, 1)^T$, $f(\tilde{x}) = 0$ (Übungsaufgabe).

Beispiel 1.5.6 (*Dixon*)

$$f(x_1, \dots, x_{10}) = (1 - x_1)^2 + (1 - x_{10})^2 + \sum_{i=1}^9 (x_i^2 - x_{i+1})^2$$

Globales Minimum bei $\tilde{x} = (1, \dots, 1)^T$.

1.6 Numerische Lösung von Optimierungsaufgaben

Wir werden die in der Vorlesung zu untersuchenden Optimierungsverfahren numerisch lösen durch iterative Verfahren, die teilweise nach endlich vielen Schritten eine Lösung ermitteln oder einem Grenzwert zustreben:

$$\lim_{k \rightarrow \infty} x^{(k)} = \tilde{x}.$$

Dabei werden wir Optimierungsaufgaben verschiedener Struktur untersuchen (z. B. linear-quadratische Aufgaben, nichtlineare Funktionale mit linearen Restriktionen, allgemeine nichtlineare Probleme, nicht jedoch lineare oder diskrete Optimierungsaufgaben).

1.7 Optimierungs-Software

1.7.1 Programmbibliotheken

Empfehlenswert und bei uns verfügbar:

- NAG-Library (Numerical Algorithms Group)
Fortran Codes
- minpack (ist public domain software)

1.7.2 Interaktive Programmsysteme

- MATLAB (MATrix LABoratory) kommerziell
- Scilab (SCientific, LABoratory) kostenlos von
INRIA, Paris
www.inria.fr

Entscheidungshilfe im Internet:

Hans D. Mittelmann, <http://plato.la.asu.edu/guide.html>

Software-Guide:

More' and Wright, [9]

2 Ableitungsfreie Verfahren

Oft ist die Berechnung der Ableitung von f so aufwendig oder – bei nicht differenzierbarem f – unmöglich, so dass man Verfahren entwickelt hat, die ohne Ableitungen auskommen. Wir behandeln hier kurz zwei davon, um die unrestringierte Aufgabe

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

numerisch zu lösen.

2.1 Simplexverfahren von Nelder und Mead

2.1.1 Grundkonstruktionen

Bemerkung: Das Verfahren hat nichts mit der Simplexmethode der linearen Optimierung zu tun! Der Name kommt von

Definition 2.1.1 $x^0, \dots, x^n \in \mathbb{R}^n$ seien affin unabhängig, d. h. $x^i - x^0, i = 1, \dots, n$ sind linear unabhängig. Die konvexe Hülle der Punkte x^0, \dots, x^n

$$S = \left\{ \sum_{i=0}^n \lambda_i x^i \mid \lambda_i \geq 0, i = 0, \dots, n, \sum_{i=0}^n \lambda_i = 1 \right\}$$

heißt (n -dimensionales) **Simplex** mit den Ecken x^0, \dots, x^n .

- Beim Start des Verfahrens wird ein Simplex vorgegeben.
- Man ermittelt die Ecke mit dem größten Funktionswert,

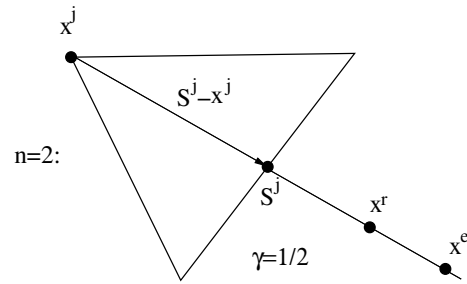
$$f(x^m) = \max \{f(x^0), \dots, f(x^n)\}$$

- Danach wird ein neuer Punkt ermittelt, der einen kleineren Funktionswert ergibt und x^m ersetzt.

Dazu werden folgende Konstruktionen benutzt:

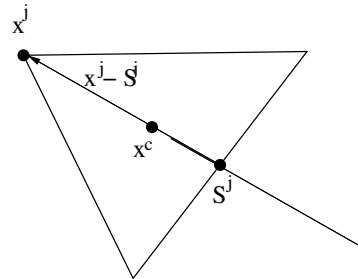
Def.	$s^j = \frac{1}{n} \sum_{\substack{i=0 \\ i \neq j}}^n x^i$	Schwerpunkt der (anderen) Ecken bezgl. x^j
-------------	---	--

Konstruktionsprinzipien:

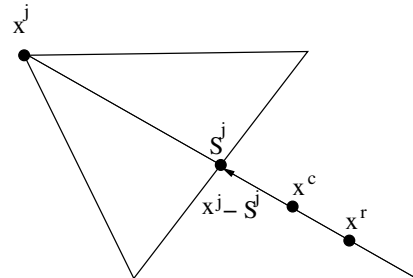


γ : Reflektionskonstante

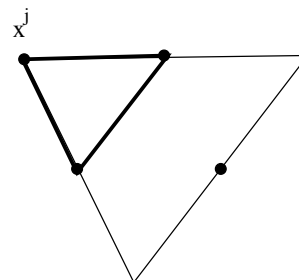
- **Reflektion** von x^j an s^j
 $x^r = s^j + \gamma(s^j - x^j)$, $0 < \gamma \leq 1$
- Dieses eben konstruierte x^r kann weiter nach außen bewegt werden:
Expansion von x^r in Richtung $s^j - x^j$ (d. h. in Richtung $x^r - s^j$)
 $x^e = s^j + \beta(x^r - s^j)$, $\beta > 1$ Expansionskonstante



- **Kontraktion** (3 Typen)
 - Partielle Kontraktion innen** $x^c = s^j + \alpha(x^j - s^j)$ $0 < \alpha < 1$ Kontraktionskonstante
 - Partielle Kontraktion außen**
 $x^c = s^j + \alpha(x^r - s^j)$



- Totale Kontraktion**
 Ersetze alle x^i außer x^j durch
 $\hat{x}^i = x^i + \frac{1}{2}(x^j - x^i) = \frac{1}{2}(x^i + x^j)$



2.1.2 Ablauf des Verfahrens

Die einfachste Variante läuft so ab:

<u>Vorab werden gewählt</u>	$\alpha \in (0, 1)$	Kontraktionskonstante
	$\beta > 1$	Expansionskonst.
	$\gamma \in (0, 1]$	Reflexionskonstante

Folgende Schritte laufen ab:

1. Wahl eines Startpunkts $x^0 \in R^n$, Festlegung der anderen n Ecken des Startsimplexes durch

$$x^j = x^0 + e^j, j = 1, \dots, n,$$

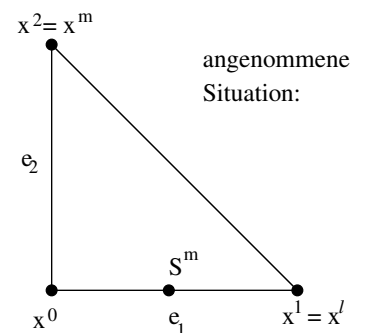
wobei e^j den j -ten Standardeinheitsvektor bezeichnet.

2. Bestimme die Ecken mit maximalem und minimalem Funktionswert: x^m, x^l mit

$$\begin{aligned} f(x^m) &= \max \{f(x^0), \dots, f(x^n)\} \\ f(x^l) &= \min \{f(x^0), \dots, f(x^n)\} \end{aligned}$$

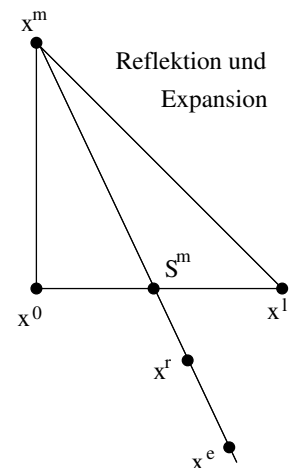
und bestimme den Schwerpunkt der Ecken bezügl. x^m

$$s^m = \frac{1}{n} \sum_{\substack{i=0 \\ i \neq m}}^n x^i$$



3. Reflektion von x^m am Schwerpunkt s^m

$$x^r = s^m + \gamma(s^m - x^m)$$



4. Aufbau des neuen Simplexes

Dazu eine Fallunterscheidung

(i)

$$\boxed{f(x^r) < f(x^l)}$$

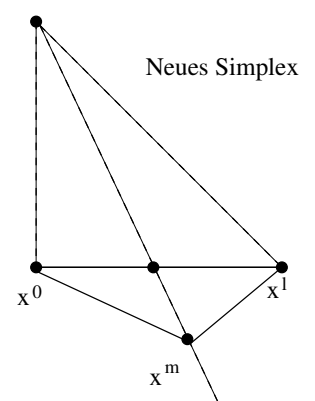
Dann war die Richtung gut, und man probiert noch etwas mehr: Expansion von x^r

$$x^e = s^m + \beta(x^r - s^m)$$

Man ersetzt x^m durch den besseren der beiden Punkte:

$$\tilde{x}^m = \begin{cases} x^e, & f(x^e) < f(x^r) \\ x^r, & f(x^r) \leq f(x^e) \end{cases}$$

$$\underline{\underline{x^m := \tilde{x}^m}}$$



(ii)

$$f(x^l) \leq f(x^r) \leq \max \{f(x^j), j \neq m\}$$

Nichts gewonnen, nichts verloren – ersetze x^m durch x^r

$$\underline{\underline{x^m := x^r}}$$

(iii)

$$f(x^r) > \max \{f(x^j), j \neq m\}$$

- Wenn $f(x^r) \geq f(x^m)$: Partielle innere Kontraktion

$$x^c = s^m + \alpha(x^m - s^m)$$

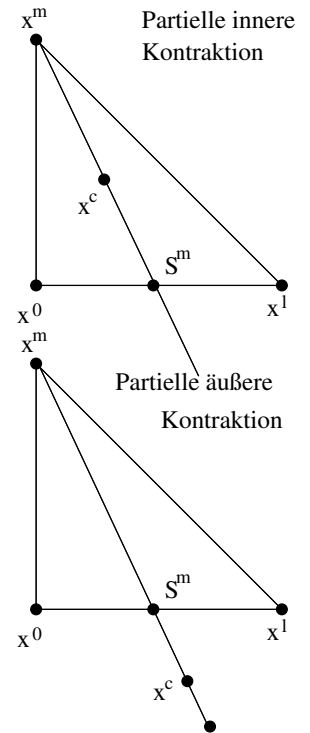
- Wenn $f(x^r) < f(x^m)$: Partielle äußere Kontraktion

$$x^c = s^m + \alpha(x^r - s^m)$$

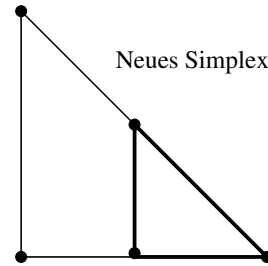
- Wenn $f(x^c) < f(x^m)$, dann ersetze x^m durch x^c

$$x^m := x^c$$

- Wenn $f(x^c) \geq f(x^m)$, dann führe eine totale Kontraktion bezüglich x^l aus: x^r



$$x^i := \frac{1}{2} (x^i + x^l), i \neq l$$



5. Gehe mit dem neu ermittelten Simplex (Ecken $\{x^0, \dots, x^n\}$) zu Schritt 2.

Das Verfahren erzeugt Eckenfolgen $\{x^{(k,0)}, \dots, x^{(k,n)}\}_{k=1}^\infty$ und stellt sicher, dass

$$f(x^{(k+1,l)}) \leq f(x^{(k,l)})$$

gilt. In gewissem Sinne kann man $x^{(k,l)}$ als den aktuellen Iterationspunkt bezeichnen. Allgemeine Konvergenzsätze gibt es nicht.

Empirische Untersuchungen zeigen $0.4 \leq \alpha \leq 0.6$, $2 \leq \beta \leq 3$, $\gamma = 1$ sind zu empfehlen.

Verfügbare Codes: EO4CCF (NAG)
 fmins (MATLAB 5)
 fminsearch (MATLAB 6)

Das Verhalten des Verfahrens wird am Beispiel der Rosenbrock-Funktion deutlich:

Startpunkt: $(-1.9, 2)^T$ EO4CCF: stoppt nach 186 Funktionsauswertungen bei $(1.000011, 1.000023)^T$
 fmins: nach 210 Auswertungen bei $(100002, 100003)^T$

2.2 Mutations-Selektions-Verfahren

- Zufällige “Mutation” der aktuellen Iterierten
- Auswahl der “brauchbaren” Iterierten

Diese Verfahren gehören zur Klasse von Methoden der stochastischen Suche.

Verfahrensgrundprinzip:

1. Wähle Startpunkt $x^{(0)} \in \mathbb{R}^n$
 $k := 0$
2. Berechne neuen Punkt $v^{(k)}$ durch zufällige Änderung von $x^{(k)}$ (Zufallszahlen), z. B.

$$\boxed{v_i^{(k)} = x_i^{(k)} + \delta_k \left(r_i^{(k)} - 0.5 \right)} \quad i = 1, \dots, n$$

$r_i^{(k)}$: Zufallszahlen aus $[0, 1]$
 δ_k : Schrittweiten

3.

$$x^{(k+1)} = \begin{cases} v^{(k)} & \text{falls } f(v^{(k)}) < f(x^{(k)}) \\ x^{(k)} & \text{sonst} \end{cases}$$

Numerisches Resultat: Für Beispiel 1.5.2 (Rosenbrock) werden 2776 Schritte benötigt.

2.3 Anwendung: Nichtlineare Regression

(Siehe Beispiel 1.5.1)

10 Messwertpaare

ξ_i	1	2	3	4	5	6	7	8	9	10
η_i	1	1.1	1.2	1.35	1.55	1.75	2.5	3	3.7	4.5

Ansatz: $\eta(\xi) = g(x, \xi) = x_1 e^{\xi x_2}$

$$(P) \quad \boxed{\min f(x) = f(x_1, x_2) = \sum_{i=1}^{10} (\eta_i - x_1 e^{\xi_i x_2})^2}$$

Anwendung der MATLAB-Implementierung `fmins` des Nelder-Mead-Verfahrens ergibt

$$\tilde{x} = \begin{pmatrix} 0.632067 \\ 0.195061 \end{pmatrix} \quad f(\tilde{x}) = 0.19041$$

Weitere Beispiele werden in den Übungen diskutiert.

3 Probleme ohne Restriktionen – Theorie

3.1 Optimalitätsbedingungen

3.1.1 Notwendige Bedingungen erster Ordnung

Im gesamten Kapitel 3 wird vorausgesetzt:

$$\begin{array}{ll} D \subset \mathbb{R}^n & \text{offen, nichtleer} \\ f : D \rightarrow \mathbb{R} & \text{mit gewissen Differenzierbarkeitsannahmen} \end{array}$$

Wir betrachten die unrestringierte Aufgabe

$$(PU) \quad \boxed{\min_{x \in D} f(x)}$$

Satz 3.1.1 (Fermat) *f besitze in $\tilde{x} \in D$ ein lokales Minimum und sei an der Stelle \tilde{x} differenzierbar. Dann gilt*

$$\boxed{\nabla f(\tilde{x}) = 0} \quad \text{Notwendige Bedingung 1. Ordnung} \quad (3.1)$$

Beweis: Grundwissen aus der Analysis. □

Bemerkung: Bei uns sind Vektoren stets Spaltenvektoren. Deshalb ist $f'(x)$ ein Zeilenvektor und $\nabla f(x)$ ein Spaltenvektor. Es gilt

$$\nabla f(x) = f'(x)^T.$$

Beispiel 3.1.1

$$f(x) = \frac{1}{2}x^T Hx + b^T x \quad \begin{array}{l} H \in \mathbb{R}^{(n,n)}, \\ b \in \mathbb{R}^n \end{array}, \quad \underline{\text{symmetrisch}}$$

Hier gilt

$$\nabla f(x) = Hx + b.$$

Eine Lösung der Aufgabe

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

muss also die Gleichung $Hx = -b$ erfüllen. Ist H außerdem positiv definit, so hat das 2 Effekte. Erstens existiert eine Lösung (Bsp 1.3.1). Außerdem ist unser Gleichungssystem eindeutig lösbar. Damit ist

$$\tilde{x} = -H^{-1}b$$

die eindeutig bestimmte Lösung.

Anwendung: Lineare Regression (Fortsetzg. Bsp 1.3.2)

Wir hatten

$$H = 2 \begin{pmatrix} \sum_{i=1}^m \xi_i^2 & \sum_{i=1}^m \xi_i \\ \sum_{i=1}^m \xi_i & m \end{pmatrix}, \quad b = -2 \begin{pmatrix} \sum_{i=1}^m \xi_i \eta_i \\ \sum_{i=1}^m \eta_i \end{pmatrix}$$

erhalten. Ist H positiv definit, dann ergibt sich für die Lösung der Regressionsaufgabe das Gleichungssystem

$$\begin{aligned} \left(\sum_{i=1}^m \xi_i^2 \right) x_1 + \left(\sum_{i=1}^m \xi_i \right) x_2 &= - \sum_{i=1}^m \xi_i \eta_i \\ \sum_{i=1}^m \xi_i x_1 + m x_2 &= - \sum_{i=1}^m \eta_i. \end{aligned}$$

Bestimmen Sie die Lösung!

Definition 3.1.1 Ist f in $\tilde{x} \in D$ differenzierbar und gilt $\nabla f(\tilde{x}) = 0$, so heißt \tilde{x} **stationärer Punkt** von f .

Bemerkung: Optimierungsalgorithmen berechnen in der Regel stationäre Punkte. Diese müssen keineswegs lokale oder globale Minima (Maxima) ergeben. Bsp: $f(x) = x^3$ bei $x = 0$.

Beispiel 3.1.2 *Rosenbrock-Funktion:* Hat genau einen stationären Punkt bei $\tilde{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$\nabla f(x) = \begin{pmatrix} -400 x_1 (x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix}.$$

Ist die Zielfunktion f differenzierbar, so heißt (PU) **glatte** oder **differenzierbare** Optimierungsaufgabe. Bei vielen Anwendungen ist f nicht überall differenzierbar, ein typisches Beispiel ist

$$f(x) = \|x\| \quad \text{bei } x = 0.$$

Mit **nichtglatter Optimierung** werden wir uns kaum befassen. Allerdings geben wir folgendes nützliches Resultat an.

Definition 3.1.2 f heißt in $x \in D$ in Richtung $h \in \mathbb{R}^n$ **richtungs-differenzierbar**, wenn die **Richtungsableitung**

$$f'(x, h) := \lim_{t \downarrow 0} \frac{f(x + th) - f(x)}{t}$$

existiert. Gilt dies für alle Richtungen h , so heißt f **richtungs-differenzierbar** an der Stelle x .

Satz 3.1.2 Ist \tilde{x} ein lokales Minimum von (PU) und ist f an der Stelle $\tilde{x} \in D$ richtungs-differenzierbar, dann gilt

$$\boxed{f'(\tilde{x}, h) \geq 0 \quad \forall h \in \mathbb{R}^n} \quad \text{“Variationsungleichung”} \quad (3.2)$$

Beweis: D ist offen, damit $\exists r > 0$:

$$f(x) \geq f(\tilde{x}) \quad \forall x \in B(\tilde{x}, r).$$

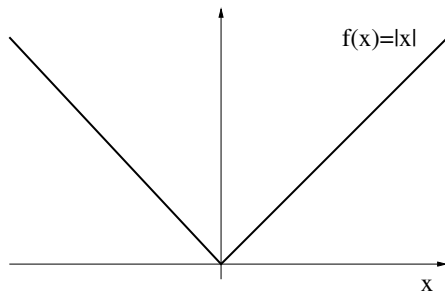
Sei $h \in \mathbb{R}^n$ beliebig, aber fest. Dann gilt $\tilde{x} + th \in B(\tilde{x}, r)$ für betragsmäßig kleine t , somit

$$\begin{aligned} f(\tilde{x} + th) - f(\tilde{x}) &\geq 0 \\ \Rightarrow \frac{f(\tilde{x} + th) - f(\tilde{x})}{t} &\geq 0 \quad \Rightarrow \quad f'(\tilde{x}, h) \geq 0 \end{aligned}$$

□

(3.2) ist intuitiv einleuchtend. In \tilde{x} liegt ein lok. Minimum vor, also kann keine Richtung existieren, in der es abwärts geht!

Beispiel 3.1.3 $f(x) = |x|$ hat lokales Minimum bei $\tilde{x} = 0$.



f ist bei $\tilde{x} = 0$ nicht differenzierbar, aber die Richtungsableitung existiert:

$$\begin{aligned} \frac{f(th) - f(0)}{t} &= \frac{|th|}{t} = |h|, \quad t > 0 \\ \Rightarrow f'(0, h) &= |h| \geq 0 \quad \forall h \in \mathbb{R}. \end{aligned}$$

3.1.2 Notwendige Bedingungen zweiter Ordnung

Satz 3.1.3 f sei in einer Umgebung von $\tilde{x} \in D$ zweimal stetig differenzierbar. Ist \tilde{x} lokales Minimum von (PU) , so muss neben der notwendigen Bedingung erster Ordnung auch

$$\boxed{h^T f''(\tilde{x}) h \geq 0 \quad \forall h \in \mathbb{R}^n} \quad (3.3)$$

erfüllt sein, d. h., $f''(\tilde{x})$ muss **positiv semidefinit** sein.

Beweis: Bekannt aus der Analysis. Skizze:

Wir setzen für bel. aber festes h ,

$$F(t) = f(\tilde{x} + th).$$

F hat lokales Minimum bei $t = 0$ und ist vom Typ C^2 .

Taylorentwicklung:

$$\begin{aligned} F(t) &= F(0) + \underbrace{F'(0)}_{=0} t + \frac{1}{2} F''(\vartheta t) t^2 \\ \Rightarrow 0 &\leq \frac{F(t) - F(0)}{t^2} = \frac{1}{2} F''(\vartheta t) \end{aligned}$$

$t \downarrow 0$, Stetigkeit von $F'' \Rightarrow F''(0) = h^T f''(\tilde{x}) h \geq 0$.

□

Beispiel 3.1.4

$$\begin{aligned}f(x) &= \frac{1}{2}x^T Hx + b^T x \\f''(x) &= H\end{aligned}$$

Soll (PU) für dieses f eine Lösung haben, dann muss H positiv semidefinit sein.

Beispiel 3.1.5 Rosenbrock-Funktion

$$\begin{aligned}f_{x_1} &= -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\f_{x_2} &= 200(x_2 - x_1^2) \\f_{x_1x_1} &= -400(x_2 - x_1^2) + 800x_1^2 + 2 \\f_{x_1x_2} &= f_{x_2x_1} = -400x_1 \\f_{x_2x_2} &= 200 \\ \Rightarrow f''(1,1) &= \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} \text{ positiv definit nach Satz von Sylvester.}\end{aligned}$$

3.1.3 Hinreichende Bedingungen zweiter Ordnung

Aus der Gültigkeit von notwendigen Bedingungen erster und zweiter Ordnung kann man bekanntlich nicht auf lokale Optimalität schließen (Bsp: $f(x) = x^3$ bei $x = 0$). Dazu zieht man nach Möglichkeit hinreichende Bedingungen 2. Ordnung zu Rate.

Im Weiteren sagen wir “ f ist in U aus der Klasse C^2 ”, kurz “aus C^2 ”, wenn f zweimal stetig differenzierbar in U ist.

Satz 3.1.4 f sei aus C^2 in einer Umgebung von $\tilde{x} \in D$. Die notwendige Bedingung $\nabla f(\tilde{x}) = 0$ sowie

$$\boxed{h^T f''(z)h \geq 0 \quad \forall h \in \mathbb{R}^n} \quad (3.4)$$

sei erfüllt für alle $z \in B(\tilde{x}, \delta)$ mit einem $\delta > 0$. Dann ist \tilde{x} lokales Minimum von (PU).

Beweis: Sei $x \in B(\tilde{x}, \delta)$ beliebig. Dann

$$\begin{aligned}f(x) - f(\tilde{x}) &= f'(\tilde{x})(x - \tilde{x}) + \underbrace{\frac{1}{2}(x - \tilde{x})}_{h} \underbrace{f''(\tilde{x} - \vartheta(x - \tilde{x})))}_{z} (x - \tilde{x}), \vartheta \in (0, 1) \\ &\geq 0 \quad \text{wegen (3.4).}\end{aligned}$$

$\Rightarrow \tilde{x}$ lokales Minimum

□

Beispiel 3.1.6 Lineare Regression

Hier hängt $f''(x) = H$ nicht von x ab. Ist H nur positiv semidefinit und erfüllt \tilde{x} die notwendige Bedingung $H\tilde{x} + b = 0$, dann ist \tilde{x} lokales Minimum. Sind zwei der Messpunkte ξ_i verschieden, dann ist H positiv definit, und wir haben Existenz und Eindeutigkeit.

Beispiel 3.1.7

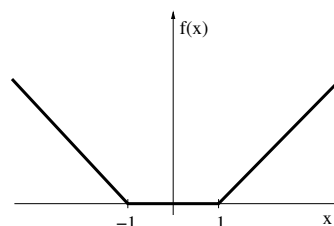
(i) $f(x) \equiv 0 \quad \forall x \in \mathbb{R}^n$

Jedes $x \in \mathbb{R}^n$ ist lokales Minimum

(ii) $f(x) = \max\{0, |x| - 1\}$

Alle $x \in [-1, 1]$ sind lokale Minima,

$x = -1, x = 1$ passen nicht in die Theorie ...



Solche etwas pathologischen Fälle schließt man durch etwas schärfere Bedingungen aus, die strenge lokale Minima implizieren:

Satz 3.1.5 f sei aus C^2 in einer Umgebung von $\tilde{x} \in D$, es gelte $\nabla f(\tilde{x}) = 0$, und $f''(\tilde{x})$ sei positiv definit, d. h.

$$h^T f''(\tilde{x}) h > 0 \quad \forall h \in \mathbb{R}^n, h \neq 0. \quad (3.5)$$

Dann existieren $r > 0, \alpha > 0$, so dass

$$\boxed{f(x) \geq f(\tilde{x}) + \alpha \|x - \tilde{x}\|^2 \quad \forall x \in B(\tilde{x}, r)} \quad \text{“quadratische Wachstumsbedingung”}$$

Damit ist \tilde{x} strenges lokales Minimum von (PU).

Beweis: Vorlesung Analysis! Skizze:

- Man zeigt mit einem wichtigen Kompaktheitsschluss, der leider nur im \mathbb{R}^n funktioniert, die Äquivalenz von (3.5) mit

$$h^T f''(\tilde{x}) h \geq \tilde{\alpha} \|h\|^2 \quad \forall h \in \mathbb{R}^n,$$

$$\alpha > 0.$$

- Dann

$$\begin{aligned} f(x) &= f(\tilde{x}) + \underbrace{\nabla f(\tilde{x})^T (x - \tilde{x})}_{=0} + \frac{1}{2} (x - \tilde{x})^T f''(\tilde{x} + \vartheta(x - \tilde{x})) (x - \tilde{x}), \quad \vartheta \in (0, 1) \\ &= f(\tilde{x}) + \underbrace{\frac{1}{2} (x - \tilde{x})^T f''(\tilde{x}) (x - \tilde{x})}_{\geq \frac{1}{2} \tilde{\alpha} \|x - \tilde{x}\|^2} + \frac{1}{2} (x - \tilde{x})^T \underbrace{[f''(\tilde{x} + \vartheta(x - \tilde{x})) - f''(\tilde{x})]}_{\leq \frac{\tilde{\alpha}}{4}, \text{ wenn } \|x - \tilde{x}\| \text{ klein}} (x - \tilde{x}) \\ &\geq f(\tilde{x}) + \frac{\tilde{\alpha}}{4} \|x - \tilde{x}\|^2 \quad (f \in C^2!) \end{aligned}$$

$$\alpha := \frac{\tilde{\alpha}}{4}$$

□

Offenbar gilt hier $h^T f''(z) h \geq 0 \quad \forall h, \forall z \in B(\tilde{x}, r)$, also impliziert (3.5) die Bedingung (3.4).

Beispiel 3.1.8 Lineare Regression mit positiv definitem H

Beispiel 3.1.9 Rosenbrock-Funktion bei $\tilde{x} = [1, 1]^T$.

$f''(1, 1) = \begin{pmatrix} 802 & -400 \\ 400 & 200 \end{pmatrix}$ ist positiv definit, also ist \tilde{x} strenges lokales Minimum.

Beispiel 3.1.10

$$f(x) = x^{2p}, \quad p \in \mathbb{N}, \quad x \in \mathbb{R}.$$

$\tilde{x} = 0$ ist lokales Minimum, aber die gängige Formel " $f'(\tilde{x}) = 0 \wedge f''(\tilde{x}) > 0 \Rightarrow$ lokales Minimum" funktioniert nur bei $p = 1$

$$\begin{array}{lll} f'(x) & = & 2p x^{2p-1} \quad \Rightarrow \quad f'(0) = 0 \\ f''(x) & = & 2p(2p-1)x^{2p-2} \quad \Rightarrow \quad \begin{array}{ll} f''(0) > 0 & \text{falls } p = 1 \\ f''(0) = 0 & \text{falls } p > 1 \end{array} \end{array}$$

Satz 3.1.5 ist nur für $p = 1$ anwendbar, Satz 3.1.4 stets.

3.2 Konvexe Optimierungsaufgaben

Wir untersuchen für die konvexe Aufgabe

$$(P) \quad \boxed{\min_{x \in \Omega} f(x)} \quad f : D \rightarrow \mathbb{R} \quad \text{konvex} \quad (D \text{ offen})$$

$$\Omega \subset D \quad \underline{\text{konvex}}, \neq \emptyset, \underline{\text{nicht notwendig offen}}$$

Jede lokale Lösung von (P) ist damit eine globale.

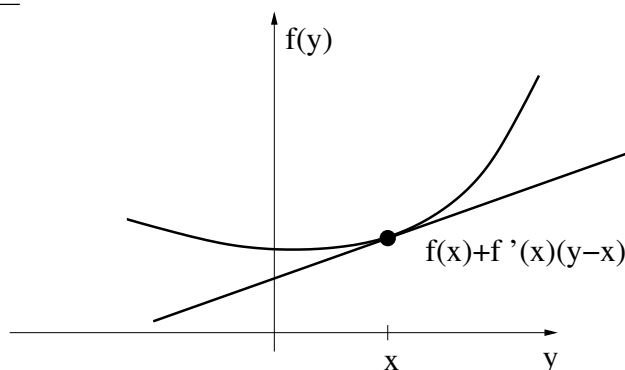
Charakterisierung der Konvexität von f durch Ableitungen:

Satz 3.2.1 f sei differenzierbar in D . Dann ist f genau dann konvex auf Ω , wenn

$$\boxed{f(y) \geq f(x) + f'(x)(y - x) \quad \forall x, y \in \Omega} \quad (3.6)$$

Beweis: Siehe Fachbücher (ist sehr einfach). □

Illustration für $n = 1$:



Satz 3.2.2 f sei differenzierbar in D und $\tilde{x} \in \Omega$. Dann ist \tilde{x} genau dann Lösung der konvexen Optimierungsaufgabe (P), wenn

$$\boxed{f'(\tilde{x})(x - \tilde{x}) \geq 0 \quad \forall x \in \Omega} \quad \text{Variationsungleichung} \quad (3.7)$$

Beweis:

- (i) Ist \tilde{x} lokales Minimum, dann wissen wir $f'(\tilde{x})(x - \tilde{x}) \geq 0$ (Satz 3.1.2: $f'(\tilde{x}, h) \geq 0$ mit $h = x - \tilde{x}$ $f'(\tilde{x}, h) = f'(\tilde{x})(x - \tilde{x})$.)
- (ii) Gilt (3.7), so wegen Konvexität und (3.6)

$$f(y) - f(\tilde{x}) \geq f'(\tilde{x})(y - \tilde{x}) \geq 0$$

□

Strenge Konvexität kann man auch so charakterisieren:

Satz 3.2.3 D offen, $\Omega \subset D$ nichtleer und konvex, $f : D \rightarrow \mathbb{R}$ differenzierbar auf D . Dann ist f genau dann streng konvex auf Ω , wenn

$$f(y) > f(x) + f'(x)(y - x) \quad \forall x, y \in \Omega, x \neq y. \quad (3.8)$$

Beweis:

- (i) f sei strikt konvex. Dann ist f insbesondere konvex und

$$f(y) \geq f(x) + f'(x)(y - x). \quad (*)$$

Nehmen wir an, (3.8) gilt nicht. Dann gibt es ein Paar $(x, y), x \neq y$, so dass in $(*)$ Gleichheit gilt. Wegen strikter Konvexität von f ,

$$\begin{aligned} f\left(\frac{1}{2}x + \frac{1}{2}y\right) &< \frac{1}{2} \underbrace{f(y) + f(x)}_{= \text{ rechte Seite von } (*) \text{ wegen Gleichheit}} \\ &= \frac{1}{2}f(x) + \frac{1}{2}f'(x)(y - x) + \frac{1}{2}f(x) \\ &= f(x) + f'(x)\left(\frac{1}{2}y + \frac{1}{2}x - x\right) \\ &\leq f(x) + f\left(\frac{1}{2}x + \frac{1}{2}y\right) - f(x), \quad \text{wegen Satz 3.2.1} \\ &= f\left(\frac{1}{2}x + \frac{1}{2}y\right) \quad \text{im Widerspruch zur Annahme.} \end{aligned}$$

- (ii) Die andere Richtung zeigt man analog wie Satz 3.2.1.

□

Geometrisch ist die Aussage sehr einleuchtend.

Satz 3.2.4 Sei f differenzierbar auf D und streng konvex auf Ω sowie $\tilde{x} \in \Omega$. Dann gilt: \tilde{x} ist genau dann strenges lokales Minimum von (P) und damit auch strenges globales Minimum wenn die Variationsungleichung

$$f'(\tilde{x})(x - \tilde{x}) \geq 0 \quad \forall x \in \Omega$$

erfüllt ist.

Beweis:

(i) \Rightarrow : Natürlich muss die Variationsungl. erfüllt sein.

(ii) \Leftarrow : aus der Variationsungl. und der strengen Konvexität folgt

$$f(x) \underset{\substack{\uparrow \\ (3.8)}}{>} f(\tilde{x}) + \underbrace{f'(\tilde{x})(x - \tilde{x})}_{\substack{\geq 0 \\ \text{Var.-ungl.}}} \geq f(\tilde{x}) \quad \forall x \in \Omega, \\ x \neq \tilde{x}$$

□

Man kann Konvexität auch über zweite Ableitungen charakterisieren: “ $f''(x) > 0 \Rightarrow f$ konvex”

Satz 3.2.5 $D \subset \mathbb{R}^n$ offen, $\Omega \subset D$ konvex, $\neq \emptyset$; $f : D \rightarrow \mathbb{R}$ aus C^2 . Dann gilt

(i) Ist $f''(x)$ positiv semidef. $\forall x \in \Omega$, so ist f konvex auf Ω . Ist Ω offen, so gilt auch die Umkehrung

(ii) Ist $f''(x)$ positiv definit $\forall x \in \Omega$, so ist f streng konvex auf Ω .

Beweis: Es seien $x, y \in \Omega$. Dann gilt mit einem $\vartheta \in (0, 1)$

$$f(y) - f(x) - f'(x)(y - x) = \frac{1}{2}(y - x)^T f''(x + \vartheta(y - x))(y - x). \quad (*)$$

(i) Wegen pos. Semidefinitheit gilt dann, dass die rechte Seite von $(*)$ nichtnegativ ist. Das heißt aber Konvexität nach Satz 3.2.1.

Umgekehrt: Es sei Ω offen, $x \in \Omega$ beliebig. Wir zeigen die positive Semidefinitheit von $f''(x)$. Dazu sei $d \in \mathbb{R}^n$ beliebig, $t \in \mathbb{R}$, $|t|$ hinreichend klein. Dann gilt wegen der Charakterisierung der Konvexität durch erste Ableitung

$$\begin{aligned} f(x + td) &\geq f(x) + f'(x)(td) \\ f(x) &\geq f(x + td) + f'(x + td)(-td). \end{aligned}$$

Addition \Rightarrow

$$[f'(x + td) - f'(x)](td) \geq 0$$

$$\begin{aligned}
\Rightarrow d^T f''(x)d &= \lim_{t \rightarrow 0} \frac{1}{t} d^T [f'(x+td) - f'(x)] d \\
&= \lim_{t \rightarrow 0} \underbrace{\frac{1}{t^2}}_{\geq 0} d^T \underbrace{[f'(x+td) - f'(x)](td)}_{\geq 0} \geq 0
\end{aligned}$$

- (ii) Ist $f''(x)$ positiv definit, so entsteht in (*) für $y \neq x$ sofort eine (streng) positive rechte Seite. Nach Satz 3.2.3 ist damit f streng konvex.

Eine weitere Verschärfung des Begriffs der Konvexität ist:

Definition 3.2.1 Eine Funktion f heißt **gleichmäßig konvex** auf einer konvexen Menge $F \subset D \subset \mathbb{R}^n$, wenn mit einem $\alpha > 0$ gilt

$$\boxed{(1-\lambda)f(x) + \lambda f(y) \geq f((1-\lambda)x + \lambda y) + \lambda(1-\lambda)\alpha\|x-y\|^2} \\
\forall x, y \in \Omega, \lambda \in [0, 1].$$

Damit kann man zeigen:

- Gleichmäßige Konvexität ist bei differenzierbarem f äquivalent zu

$$f(y) - f(x) \geq f'(x)(y-x) + \alpha\|x-y\|^2 \quad \forall x, y \in \Omega$$

- Für $f \in C^2$ folgt gleichmäßige Konvexität aus der gleichmäßigen positiven Definitheit, d.h.

$$h^T f''(x)h \geq \beta\|h\|^2 \quad \forall h \in \mathbb{R}^n,$$

wobei $\beta > 0$ nicht von $x \in \Omega$ abhängt.

Ist Ω offen, dann folgt umgekehrt aus der gleichmäßigen Konvexität von f auf Ω , dass f'' gleichmäßig positiv definit ist.

4 Probleme ohne Restriktionen – Verfahren

4.1 Grundlagen

Wir untersuchen im gesamten Kapitel numerische Verfahren zur Lösung der Aufgabe

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

Da wir wissen, dass an der Stelle einer Lösung \tilde{x} die Gleichung

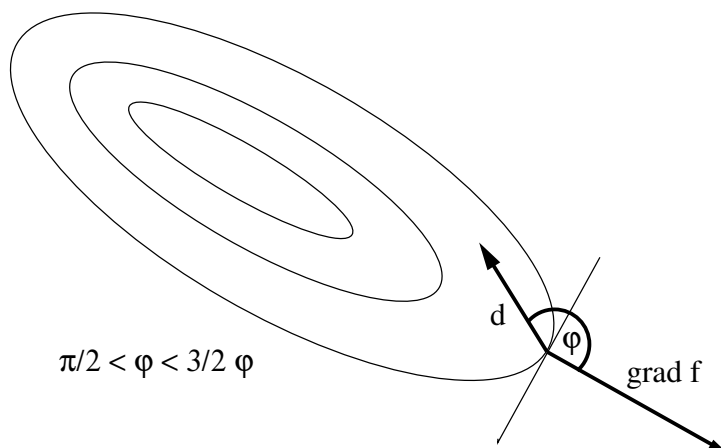
$$\nabla f(\tilde{x}) = 0 \tag{4.9}$$

erfüllt sein muss, können wir diese Gleichung numerisch lösen, etwa mit dem Newton-Verfahren. Dieses liefert aber “nur” eine Lösung dieser Gleichung, welche nicht notwendig ein Minimum ergibt. Deshalb interessiert man sich für numerische Verfahren, welche (4.9) lösen und gleichzeitig die Minimierung in (PU) berücksichtigen. Dazu gehören Abstiegsverfahren, iterative Verfahren, welche schrittweise den Funktionswert von f verkleinern.

Definition 4.1.1 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei differenzierbar an der Stelle x . Ein Vektor $d \in \mathbb{R}^n$ heißt **Abstiegsrichtung** von f im Punkt x , wenn

$$\nabla f(x)^T d < 0$$

ist.



Der Sinn dieser Definition ist klar, er wird erhärtet durch

Lemma 4.1.1 f sei differenzierbar an der Stelle x und d eine Abstiegsrichtung. Dann existiert $\bar{\sigma} > 0$ mit $f(x + \sigma d) < f(x) \quad \forall \sigma \in [0, \bar{\sigma}]$

Beweis: Wegen

$$\nabla f(x)^T d = \lim_{\sigma \rightarrow 0} \frac{f(x + \sigma d) - f(x)}{\sigma} < 0$$

muss für hinreichend kleines $\sigma > 0$ die Beziehung $f(x + \sigma d) < f(x)$ erfüllt sein. □

Beispiele von Abstiegsrichtungen:

Beispiel 4.1.1

- Gilt $\nabla f(x) \neq 0$ in x , so ist der **Antigradient**

$$\boxed{-\nabla f(x) \quad \text{Abstiegsrichtung}}$$

denn: Für $d = -\nabla f(x)$ gilt

$$f'(x)d = \nabla f(x) \cdot (-\nabla f(x)) = -\|\nabla f(x)\|^2 < 0.$$

- Ist A positiv definite (n, n) -Matrix, dann ist

$$\boxed{-A^{-1}\nabla f(x) \quad \text{Abstiegsrichtung}}$$

Das liegt daran, dass mit A auch A^{-1} positiv definit ist.

Verfahren 4.1.1 (Allgemeine Form von Abstiegsverfahren)

1. Wähle Startpunkt $x^{(0)} \in \mathbb{R}^n, k := 0$.
2. Abbruch, falls $\nabla f(x^{(k)}) = 0$
3. Berechne Abstiegsrichtung $d = d^{(k)}$ und Schrittweite $\sigma = \sigma_k > 0$, so dass

$$\begin{aligned} f(x^{(k)} + \sigma_k d^{(k)}) &< f(x^{(k)}), \\ x^{k+1} &:= x^{(k)} + \sigma_k d^{(k)} \end{aligned}$$

4. $k := k + 1$, gehe zu 1.

Bemerkungen:

- Das Abbruchkriterium ist nur theoretisch anwendbar. Numerisch arbeitet man mit $\|\nabla f(x^{(k)})\| < \varepsilon$ oder $|f(x^{(k+1)}) - f(x^{(k)})| < \varepsilon_1 \wedge \|x^{(k+1)} - x^{(k)}\| < \varepsilon_2$, wobei $\varepsilon, \varepsilon_1, \varepsilon_2$ positive Abbruchschranken sind.
- Alternativ:

$$\begin{aligned} f(x^{(k+1)}) - f(x^{(k)}) &\approx \sigma_k f'(x^{(k)}) d^{(k)} < \varepsilon_1 \\ \text{und } \|x^{(k+1)} - x^{(k)}\|_\infty &= \sigma_k \|d^{(k)}\|_\infty < \varepsilon_2 \end{aligned}$$

- Oft ist die Wahl von σ das Hauptproblem

4.2 Das Newton-Verfahren

Das Newton-Verfahren zur Lösung der Gleichung $\nabla f(x) = 0$ ist ein gängiges Mittel zur Bestimmung lokaler Extrema. Setzen wir $F(x) := \nabla f(x)$, so ist $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ gegeben und das bekannte Newton-Verfahren auf die Gleichung

$$F(x) = 0$$

anzuwenden. Die Grundidee des Verfahrens ist schnell wiederholt. Ist $x^{(k)}$ bereits bestimmt, so verhält sich $F(x)$ nahe bei $x^{(k)}$ in erster Näherung wie $F(x^{(k)}) + F'(x^{(k)})(x - x^{(k)})$, so dass wir von dieser Funktion eine Nullstelle x suchen. Wir lösen also

$$\boxed{F(x^{(k)}) + F'(x^{(k)})(x - x^{(k)}) = 0} \quad \text{lineares Gleichungssystem!} \quad (4.10)$$

und erhalten als neue Näherung $x =: x^{(k+1)}$. Ist $F'(x^{(k)})$ invertierbar, so folgt

$$x^{(k+1)} = x^{(k)} - F'(x^{(k)})^{-1} F(x^{(k)}). \quad (4.11)$$

Für die Konvergenzanalyse des Verfahrens benötigen wir folgende Voraussetzungen:

- (i) $F : \mathbb{R}^n \supset D \rightarrow \mathbb{R}^n$ ist differenzierbar in D , D offen, und hat in D eine Nullstelle \tilde{x} .
- (ii) F' ist Lipschitz-stetig in D , d. h. $\exists L > 0$:

$$\boxed{\|F'(x) - F'(y)\| \leq L \|x - y\| \quad \forall x, y \in D}$$

- (iii) Es existiert $F'(\tilde{x})^{-1}$.

Der Konvergenzbeweis des Newton-Verfahrens beruht auf folgenden bekannten und mit relativ geringem Aufwand beweisbaren Fakten:

Lemma 4.2.1

$$F(x) - F(y) - F'(y)(x - y) \leq \frac{L}{2} \|x - y\|^2 \quad \forall x, y \in D.$$

(Folgt aus dem Mittelwertsatz, angewendet auf $\varphi(t) = F(x + t(x - y))$.)

Lemma 4.2.2 *Ist A eine nichtsinguläre n, n -Matrix, S eine Matrix gleichen Typs und $\|A^{-1}\| \|S\| < 1$, dann existiert $(A + S)^{-1}$ und*

$$\|(A + S)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|S\|}.$$

Lemma 4.2.3 *Sei $G : \bar{B}(\tilde{x}, r) \rightarrow \mathbb{R}^n$ eine Kontraktion, d. h. in $\bar{B}(\tilde{x}, r)$ Lipschitz-stetig mit Konstante $L < 1$. Ist \tilde{x} ein Fixpunkt von G , dann ist es der einzige in $\bar{B}(\tilde{x}, r)$. Ausgehend von jedem beliebigen Startpunkt $x^{(0)}$ in dieser Kugel konvergiert die Folge $x^{(k)}$*

$$x^{(k+1)} = G(x^{(k)})$$

gegen \tilde{x} und

$$\|x^{(k)} - \tilde{x}\| \leq L^k \|x^{(0)} - \tilde{x}\|.$$

(Folgt aus dem Banachschen Fixpunktsatz.)

Satz 4.2.1 (Konvergenz des Newton-Verfahrens)

Unter den obigen Voraussetzungen (i) - (iii) gibt es $\delta > 0, c > 0$, so dass das Newton-Verfahren für jeden Startpunkt $x^{(0)} \in B(\tilde{x}, \delta)$ eine gegen \tilde{x} konvergente Folge $x^{(k)}$ definiert, wobei gilt

$$\boxed{\|x^{(k+1)} - \tilde{x}\| \leq c \|x^{(k)} - \tilde{x}\|^2} \quad (\text{quadratische Konvergenz}) \quad (4.12)$$

Beweisskizze

a) Mit Lemma 4.2.2 zeigt man

$$\|F'(x)^{-1}\| \leq 2 \|F'(\tilde{x})^{-1}\| \quad \forall x \in B(\tilde{x}, \delta_1)$$

b) Das wendet man an und findet

$$\begin{aligned} \|F'(x)^{-1} - F'(y)^{-1}\| &= \left\| \underbrace{F'(x)^{-1}}_{\|\cdot\| \leq 2\|F'(\tilde{x})^{-1}\|} \underbrace{(F'(y) - F'(x))}_{\leq L\|x-y\|} \underbrace{F'(y)^{-1}}_{\leq 2\|F'(\tilde{x})^{-1}\|} \right\| \\ &\leq 4L \|F'(\tilde{x})^{-1}\|^2 \|x - y\| \end{aligned}$$

c) Aus Formel (4.11) wird klar – das Newton-Verfahren ist Fixpunktiteration für

$$\boxed{G(x) := x - F'(x)^{-1}F(x).}$$

G ist Kontraktion in $B(\tilde{x}, \delta)$:

$$\begin{aligned} G(x) - G(y) &= \underbrace{x - y}_{=F'(x)^{-1}F'(x)(x-y)} - F'(x)^{-1}F(x) + F'(y)^{-1}F(y) \\ &= \underbrace{F'(x)^{-1}}_{\text{beschr. wegen a)}} \underbrace{\{F(y) - F(x) - F'(x)(y - x)\}}_{\leq \frac{L}{2}\|x-y\| \|x-y\|} \\ &\quad + \underbrace{(F'(y)^{-1} - F'(x)^{-1})}_{\leq c\|x-y\| \text{ wegen b)}} \underbrace{\cdot F(y)}_{\text{klein, wenn } y \text{ nahe an } \tilde{x}} \end{aligned}$$

Man schätzt ab und findet z. B.

$$\|G(x) - G(y)\| \leq \frac{1}{2} \|x - y\| \quad \text{in } B(\tilde{x}, \delta)$$

für ein $\delta \leq \delta_1$ klein genug gewählt.

d) Nun wird das Kontraktionslemma 4.2.3 benutzt. Die Iterationsfolge konvergiert gegen \tilde{x} . Die quadratische Konvergenz folgt aus

$$\begin{aligned} \|x^{(k+1)} - \tilde{x}\| &= \|x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}) - \tilde{x}\| \\ &= \underbrace{\|F'(x^{(k)})^{-1}\|}_{\leq 2\|F'(\tilde{x})^{-1}\|} \underbrace{\|F(\tilde{x}) - F(x^{(k)}) - F'(x^{(k)})(\tilde{x} - x^{(k)})\|}_{\leq \frac{L}{2}\|\tilde{x} - x^{(k)}\|^2} \\ &\leq c \|x^{(k)} - \tilde{x}\|^2 \quad \text{mit} \quad \boxed{c = L \|F'(\tilde{x})^{-1}\|} \end{aligned}$$

□

Das Newton-Verfahren konvergiert lokal quadratisch. Umgesetzt für das freie Optimierungsproblem (PU) bedeutet das

$$F(x) := \nabla f(x).$$

Wir fordern daher

- f'' ist Lipschitz-stetig in einer Umgebung eines lokalen Minimums \tilde{x} von f
- $f''(\tilde{x})$ ist positiv definit
(Das sichert die Existenz von $F'(\tilde{x})^{-1} = f''(\tilde{x})^{-1}$ und passt zur Minimum-Eigenschaft.)

Satz 4.2.2 *Unter obigen Voraussetzungen konvergiert das Newton-Verfahren*

$$x^{(k+1)} = x^{(k)} - f''(x^{(k)})^{-1} \nabla f(x^{(k)}) \quad (4.13)$$

lokal quadratisch gegen \tilde{x} .

In der numerischen Umsetzung invertiert man natürlich $f''(x^{(k)})$ nicht, sondern man löst das Gleichungssystem

$$f''(x^{(k)}) (x^{(k+1)} - x^{(k)}) = -\nabla f(x^{(k)}),$$

d. h. man bestimmt eine Richtung $d^{(k)}$ aus

$$f''(x^{(k)}) d^{(k)} = -\nabla f(x^{(k)})$$

und setzt

$$x^{(k+1)} := x^{(k)} + d^{(k)}.$$

Für $d^{(k)}$ gilt

$$d^{(k)} = \underbrace{f''(x^{(k)})^{-1}}_{\text{pos. definit}} \underbrace{(-\nabla f(x^{(k)}))}_{\text{Antigradient}}$$

Daher ist nach Bsp 4.1.1 $d^{(k)}$ Abstiegsrichtung, die sogenannte **Newton-Richtung**. Damit wird das Newton-Verfahren aber nicht automatisch ein Abstiegsverfahren, denn es wählt immer die Schrittweite $\sigma = 1$, und die kann zu groß sein!

Deshalb wendet man eine geänderte Verfahrensvorschrift an

$$x^{(k+1)} = x^{(k)} - \sigma_k f''(x^{(k)})^{-1} \nabla f(x^{(k)})$$

(gedämpftes Newton-Verfahren (vgl. Abschnitt 4.6)).

Bemerkung: Wir können das Newton-Verfahren auch anders interpretieren: Die Verfahrensvorschrift (4.13) bedeutet

$$f''(x^{(k)}) (x^{(k+1)} - x^{(k)}) + \nabla f(x^{(k)}) = 0.$$

Das ist gerade die notwendige Optimalitätsbedingung für Lösungen der quadratischen Optimierungsaufgabe

$$\boxed{\min_{x \in \mathbb{R}^n} \quad \nabla f(x^{(k)})^T (x - x^{(k)}) + \frac{1}{2} (x - x^{(k)})^T f''(x^{(k)}) (x - x^{(k)})} \quad (Q)_k}$$

Ist $f''(x^{(k)})$ positiv definit, so besitzt diese, wie wir inzwischen wissen, genau eine Lösung. Diese ist gerade $x^{(k+1)}$. Damit ist das Newton-Verfahren äquivalent zur Lösung einer Folge quadratischer Optimierungsaufgaben (wenn $f''(\tilde{x})$ positiv definit ist), es ist damit ein **sequentiell-quadratisches Optimierungsverfahren – SQP-Verfahren** (von Sequential Quadratic Programming).

Man schreibt $(Q)_k$ so auf:

$$\begin{aligned} & \min \nabla f(x^{(k)})^T z + \frac{1}{2} z^T f''(x^{(k)}) z \\ \text{und} \quad & x^{(k+1)} := x^{(k)} + z^{(k)} \end{aligned}$$

4.3 Abstiegsverfahren – allgemeine Aussagen

4.3.1 Effiziente Schrittweiten

Ist $d^{(k)}$ eine Abstiegsrichtung und σ_k hinreichend klein, so gilt $f(x^{(k)} + \sigma_k d^{(k)}) < f(x^{(k)})$. Das muss aber keineswegs zur Konvergenz des Abstiegsverfahrens in ein lokales Minimum führen, wie folgendes Beispiel zeigt:

Beispiel 4.3.1 $f(x) = x^2$, $d^{(k)} = -1$ für alle $k \geq 0$, $x^{(0)} = 1$, und $\sigma_k = \left(\frac{1}{2}\right)^{k+2}$, $k = 0, 1, \dots$.
Die Folge $x^{(k)}$ strebt gegen $\frac{1}{2}$, die Schrittweiten sind zu klein.

Startet unser Abstiegsverfahren bei $x^{(0)}$, so entstehen nur noch kleinere Funktionswerte. Deshalb liegen die weiteren Iterierten stets in $N(f, f(x^{(0)}))$.

Definition 4.3.1 Es sei x aus $N(f, f(x^{(0)}))$ und $d \in \mathbb{R}^n$ eine Abstiegsrichtung. Eine Schrittweite σ heißt **effizient**, falls

$$\boxed{f(x + \sigma d) \leq f(x) - c \left(\frac{\nabla f(x) \cdot d}{\|d\|} \right)^2} \quad (4.14)$$

mit einer von $x \in N(f, f(x^{(0)}))$ und d unabhängigen Konstante $c > 0$ gilt.

Erläuterung: Für $d = -\nabla f(x)$ ist das Quadrat am größten, der Abstieg am stärksten. Für $d \perp \nabla f(x)$ ergibt sich differentiell Abstieg Null. Die Konstante c bedeutet eine Mindestrate. Beachte: $d/\|d\|$ ist Einheitsvektor.

Sind Folgen $\{x^{(k)}\}, \{d^{(k)}\}$ mit $\nabla f(x^{(k)})^T d^{(k)} < 0$ und effiziente Schrittweiten σ_k gegeben, dann ist (4.14) mit einer von k unabhängigen Konstante $c > 0$ erfüllt.

Eine spezielle Form der Effizienz ist das Prinzip des hinreichenden Abstiegs: Man verlangt von x und d unabhängige Konstanten $c_1, c_2 > 0$, so dass

$$f(x + \sigma d) \leq f(x) + c_1 \sigma \nabla f(x)^T d \quad (4.15)$$

(hinreichend schneller Abstieg)

und

$$\sigma \geq -c_2 \frac{\nabla f(x)^T d}{\|d\|^2} \quad (4.16)$$

(Mindestschrittweite)

Aus diesen beiden Bedingungen folgt (4.14), Effizienz, mit $c = c_1 c_2$, denn $f(x + \sigma d) \leq f(x) - c_1 (c_2 \frac{\nabla f(x)^T d}{\|d\|^2} \nabla f(x)^T d = f(x) - c_1 c_2 \frac{\nabla f(x)^T d}{\|d\|})^2$

Bemerkung: Man kann unter Voraussetzung der Lipschitz-Stetigkeit von $f'(x)$ auf $N(f, f(x^{(0)}))$ die Existenz effektiver Schrittweiten beweisen, vgl. Lemma 4.3.4 in [1].

4.3.2 Gradientenbezogene Richtungen

Ist $N(f, f(x^{(0)}))$ kompakt, so ist die Folge der Funktionswerte $\{f(x^{(k)})\}$ (nach unten) beschränkt. Ist die Schrittweitenfolge $\{\sigma_k\}$ effizient, dann gilt

$$f(x^{(k+1)}) \leq f(x^{(k)}) - c \left(\frac{\nabla f(x^{(k)})^T d^{(k)}}{\|d^{(k)}\|} \right)^2$$

Aus der Monotonie folgt die Konvergenz der Funktionswerte, so dass $(f(x^{(k+1)}) - f(x^{(k)})) \rightarrow 0, k \rightarrow \infty$ folgen muss, d. h.

$$\frac{\nabla f(x^{(k)})^T d^{(k)}}{\|d^{(k)}\|} \rightarrow 0, \quad k \rightarrow \infty. \quad (4.17)$$

Man will nun die Richtungen so wählen, dass daraus folgt

$$\nabla f(x^{(k)}) \rightarrow 0, \quad k \rightarrow \infty. \quad (4.18)$$

Beziehung (4.17) kann ohne (4.18) gelten, wenn die Richtung $d^{(k)}$ in der Grenze orthogonal zu $\nabla f(x^{(k)})$ wird. Das muss man ausschließen und gleichmäßig kleiner als der rechte Winkel bleiben: Nun gilt

$$\begin{aligned} \cos(\nabla f(x^{(k)}), d^{(k)}) &= \frac{\nabla f(x^{(k)})^T d^{(k)}}{\|\nabla f(x^{(k)})\| \|d^{(k)}\|} =: \beta_k \\ \Rightarrow \quad \beta_k \|\nabla f(x^{(k)})\| &= \underbrace{\frac{\nabla f(x^{(k)})^T d^{(k)}}{\|d^{(k)}\|}}_{\rightarrow 0 \text{ bei Effizienz}} \end{aligned}$$

Daraus folgt $\nabla f(x^{(k)}) \rightarrow 0$, falls $-\beta_k \geq c > 0 \quad \forall k$.

Definition 4.3.2 Seien $x \in N(f, f(x^{(0)}), d \in \mathbb{R}^n$. Die Richtung d heißt **gradientenbezogen** in x , wenn

$$-\nabla f(x)^T d \geq c_3 \|\nabla f(x)\| \|d\| \quad (4.19)$$

mit einer von x und d unabhängigen Konstanten $c_3 > 0$ gilt.

Sie heißt **streng gradientenbezogen**, wenn zusätzlich

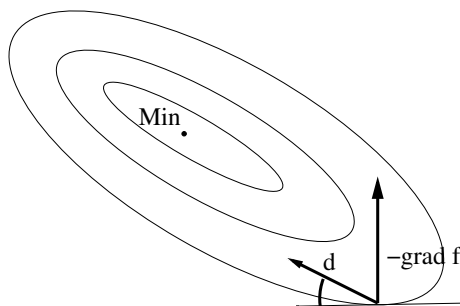
$$c_4 \|\nabla f(x)\| \geq \|d\| \geq \frac{1}{c_4} \|\nabla f(x)\| \quad (4.20)$$

mit einer von x und d unabhängigen Konstante $c_4 > 0$ gilt.

Beispiel 4.3.2 Der Antigradient $d = -\nabla f$ ist streng gradientenbezogen, denn

$$\begin{aligned} -\nabla f(x)^T d &= \|\nabla f(x)\|^2 = 1 \cdot \|\nabla f(x)\| \|d\| \\ \|\nabla f(x)\| &\stackrel{(>)}{=} \|d\| \stackrel{(>)}{=} \|\nabla f(x)\| \quad d.h. c_3 = c_4 = 1. \end{aligned}$$

Illustration der Gradienten-Bezogenheit:



Mindestabstand zum rechten Winkel mit $-\nabla f(x)$ garantiert hinreichenden Abstieg (wenn nicht $\|d\| \rightarrow 0$).

Wir wollen nun skizzieren, dass auch die Newton-Richtung streng gradientenbezogen ist. Dazu braucht man folgende Voraussetzung:

(VLK) (Lokal gleichmäßige Konvexität)

Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $N(f, f(x^{(0)})) \subset D$, $\neq \emptyset$, offen, konvex, $f \in C^2$ auf D . Mit $\alpha_1 > 0$ gelte

$$h^T f''(x) h \geq \alpha_1 \|h\|^2 \quad \forall h \in \mathbb{R}^n, \forall x \in D,$$

d. h. gleichmäßige positive Definitheit von f'' auf D .

Ohne Beweis folgern mir aus (VLK):

Lemma 4.3.1

- $N(f, f(x^{(0)}))$ konvex und kompakt
- $h^T f''(x) h \geq \alpha_2 \|h\| \quad \forall h \in \mathbb{R}^n, \quad \forall x \in N(f, f(x^{(0)}))$
- $\|f''(x)\| \leq \alpha_2 \quad \quad \quad -'' -$
- $\|f''(x)^{(-1)}\| \leq \beta_2 := 1/\alpha_1 \quad \quad \quad -'' -$
- $\beta_1 \|h\|^2 \leq h^T f''(x)^{(-1)} h \leq \beta_2 \|h\|^2, \quad \forall h \in \mathbb{R}^n, -'' -$
- f ist gleichmäßig konvex auf D .

Beispiel 4.3.3 (Newton-Richtung)

(VLK) sei erfüllt, $x \in N(f, f(x^{(0)}))$,

$$\begin{aligned} d &= -f''(x)^{(-1)} \nabla f(x) \\ \Rightarrow -\nabla f(x)^T d &= \nabla f(x)^T f''(x)^{(-1)} \nabla f(x) \geq \beta_1 \|\nabla f(x)\|^2 \quad (\text{Lemma 4.3.1}) \end{aligned}$$

Und

$$\left. \begin{aligned} \|d\| &= \|-f''(x)^{(-1)} \nabla f(x)\| \leq \beta_2 \|\nabla f(x)\|^2 & -'' - \\ \|\nabla f(x)\| &= \|-f''(x) d\| \leq \alpha_2 \|d\| & -'' - \end{aligned} \right\} \Rightarrow (4.20)$$

d.h. strenge Gradienten-Bezogenheit. (4.19) folgt aus

$$-\nabla f^T d \stackrel{\text{oben}}{\geq} \beta_1 \|\nabla f\| \underbrace{\|\nabla f\|}_{\geq \frac{1}{\beta_2} \|d\|} \geq \frac{\beta_1}{\beta_2} \|\nabla f\| \|d\|.$$

4.3.3 Allgemeine Konvergenzsätze

Folgende Voraussetzungen werden im Weiteren oft benötigt:

(V NK) Für ein gegebenes $x^{(0)} \in \mathbb{R}^n$ ist die Niveaumenge
 $N(f, f(x^{(0)})) = \{x | f(x) \leq f(x^{(0)})\}$
 kompakt.

(VFD) $f \in C^1$ auf konvexer, offener Menge $D_0 \supset N(f, f(x^{(0)}))$.

Damit lässt sich zunächst zeigen:

Satz 4.3.1 (VNK) und (VFD) seien erfüllt, die Suchrichtungen $d^{(k)}$ des allgemeinen Abstiegsverfahrens 4.1.1 seien gradientenbezogen in $x^{(k)}$, die Schrittweiten σ_k effizient. Stoppt das Verfahren nicht nach endlich vielen Schritten, dann gilt $\nabla f(x^{(k)}) \rightarrow 0, k \rightarrow \infty$ und $\{x^{(k)}\}$ besitzt einen Häufungspunkt \tilde{x} .

Für jeden solchen Häufungspunkt gilt $\nabla f(\tilde{x}) = 0$.

Dass $\{x^{(k)}\}$ einen Häufungspunkt besitzt, nützt numerisch herzlich wenig. Man möchte haben: $x^{(k)} \rightarrow \tilde{x}$. In der Tat gilt

Satz 4.3.2 Zusätzlich zu den Voraussetzungen von Satz 4.3.1 sei im allgemeinen Abstiegsverfahren 4.1.1

- $d^{(k)}$ streng gradientenbezogen,
- Schrittweitenfolge $\{\sigma_k\}$ beschränkt,
- die Menge aller Nullstellen von ∇f in $N(f, f(x^{(0)}))$ endlich.

Stoppt das Verfahren nicht nach endlich vielen Schritten, dann konvergiert $x^{(k)}$ gegen eine Nullstelle von ∇f .

Beweisidee: Strenge Gradientenbezogenheit, (4.20) \Rightarrow

$$\|x^{(k+1)} - x^{(k)}\| = \underbrace{\|\sigma_k d^{(k)}\|}_{\leq \bar{\sigma}} \leq c\bar{\sigma} \underbrace{\|\nabla f(x^{(k)})\|}_{\rightarrow 0, \text{Satz 4.3.1}} \quad (*)$$

Wegen (VNK) ist H , Menge aller Häufungspunkte von $x^{(k)}$ nichtleer. Außerdem gilt für den Abstand

$$\boxed{d(x^{(k)}, H) < \varepsilon \quad \forall k > k_0} \quad (**)$$

Sei \tilde{x} irgendein HP von $x^{(k)}$. Da die Menge der HP endlich ist, gibt es eine Kugel $B(\tilde{x}, \rho)$ mit $H \cap B(\tilde{x}, \rho) = \{\tilde{x}\}$. Also existiert nach (**) ein l_0 mit

$$\|x^{(l_0)} - \tilde{x}\| < \varepsilon$$

Wegen (*) und $\nabla f(x^{(k)}) \rightarrow 0$ gilt auch

$$\begin{aligned} & \|x^{(l_0+1)} - x^{(l_0)}\| < \varepsilon \\ \Rightarrow & \|x^{(l_0+1)} - \tilde{x}\| < 2\varepsilon < \frac{\rho}{2} \quad \text{für} \quad \varepsilon < \frac{\rho}{4} \end{aligned}$$

Damit ist $x^{(l_0+1)} \in B(\tilde{x}, \rho)$ und wegen (**) gilt

$$\|x^{(l_0+1)} - \tilde{x}\| < \varepsilon$$

Induktiv folgt schließlich $x^{(k)} \rightarrow \tilde{x}$. □

Diese bisherigen Resultate sind allgemein, aber schwach – sie sagen nichts aus über eine Konvergenzrate.

Gilt aber (VLK), so hat man gleichmäßige Konvexität in $N(f, f(x^{(0)}))$ und man kann zeigen

$$\frac{\alpha_1}{2} \|x - \tilde{x}\|^2 \leq f(x) - f(\tilde{x}) \leq \frac{1}{2\alpha_1} \|\nabla f(x)\|^2 \quad (4.21)$$

in $N(f, f(x^{(0)}))$, wobei \tilde{x} das einzige lokale Minimum in $N(f, f(x^{(0)}))$ ist [1, Lemma 4.3.14]. Das ist automatisch das globale.

(Die linke Abschätzung folgt aus (VLK), Taylorentwicklung und $\nabla f(\tilde{x}) = 0$ wie gehabt. Die rechte ist etwas komplizierter.)

Diese Eigenschaft ist die Grundlage für

Satz 4.3.3 *Voraussetzungen:*

- (VLK)
- $d^{(k)}$ gradientenbezogen in $x^{(k)}$
- $\{\sigma_k\}$ effizient.

Stoppt das Verfahren nicht nach endlich vielen Schritten, dann konvergiert $\{x^{(k)}\}$ gegen das eindeutig bestimmte globale Minimum \tilde{x} von f .

Es gibt ein $q \in (0, 1)$ mit

$$f(x^{(k)}) - f(\tilde{x}) \leq q^k (f(x^{(0)}) - f(\tilde{x})) \quad (4.22)$$

und

$$\|x^{(k)} - \tilde{x}\|^2 \leq \frac{2}{\alpha_1} q^k (f(x^{(0)}) - f(\tilde{x})) \quad k \geq 0. \quad (4.23)$$

Folgerung: $\|x^{(k)} - \tilde{x}\| \leq C\sqrt{q^k} = C\tilde{q}^k$

Bemerkung: Damit verhält sich $\{x^{(k)}\}$ wie eine linear konvergente Folge, denn $\{x^{(k)}\}$ heißt linear konvergent wenn $\|x^{(k+1)} - \tilde{x}\| \leq L\|x^{(k)} - \tilde{x}\|$ mit $0 < L < 1$. Dann gilt

$$\|x^k - \tilde{x}\| \leq L^k \|x^{(0)} - \tilde{x}\|$$

4.4 Schrittweitenverfahren

4.4.1 Exakte Schrittweite

Gegeben seien $x \in \mathbb{R}^n$ und eine Abstiegsrichtung $d \in \mathbb{R}^n$. Gesucht ist die Schrittweite σ . Am besten wäre es, σ so zu wählen, dass

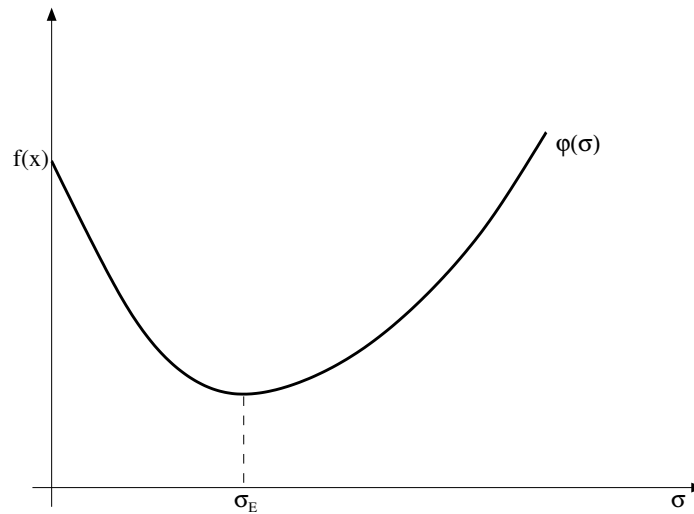
$$\min_{s \geq 0} f(x + sd) = \min_{s \geq 0} \varphi(s) = \varphi(\sigma).$$

Das wird aber erstens nicht immer möglich sein (zum Beispiel bei $f(x) = e^{(-x)}$) und liefe zweitens auf globale Optimierung in \mathbb{R} hinaus. Ist aber (VNK) erfüllt, die Niveaumenge also kompakt, dann muss $\varphi(s)$ irgendwann größer als $\varphi(0)$ werden. Folglich hat $\varphi'(s) = \nabla f(x + sd)^T d$ eine kleinste positive Nullstelle σ_E .

Definition 4.4.1 Die Zahl σ_E mit

$$\varphi'(s) \begin{cases} = 0 & , \text{ falls } s = \sigma_E \\ < 0 & , \text{ falls } s \in [0, \sigma_E) \end{cases}$$

heißt **exakte Schrittweite**.



Man kann sie nach unten wie folgt abschätzen:

$$\begin{aligned} 0 &= \underset{\substack{\uparrow \\ \text{Def von } \sigma_E}}{\nabla f(x + \sigma_E d)} \cdot d = \underset{\uparrow}{\nabla f(x)} \cdot d + [\nabla f(x + \sigma_E d) - \nabla f(x)] \cdot d \\ &\leq \underset{\substack{\uparrow \\ \text{Lipschitzbed.}}}{\nabla f(x)} \cdot d + \sigma_E L \|d\|^2 \\ &\Rightarrow \boxed{\sigma_E \geq \tilde{\sigma} = -\frac{\nabla f(x) \cdot d}{L \|d\|^2}} \end{aligned} \quad (4.24)$$

Außerdem bekommt man den Mindestanstieg

$$f(x + \sigma_E d) \leq f(x) + \frac{1}{2} \tilde{\sigma} \nabla f(x) \cdot d \quad (4.25)$$

Damit sind $\sigma_E, \tilde{\sigma}$ effizient. Leider ist σ_E in der Regel schwer zu bestimmen. Eine Ausnahme bilden quadratische Funktionen

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

für die σ_E leicht explizit zu berechnen ist. Ansonsten muss man sich anders behelfen.

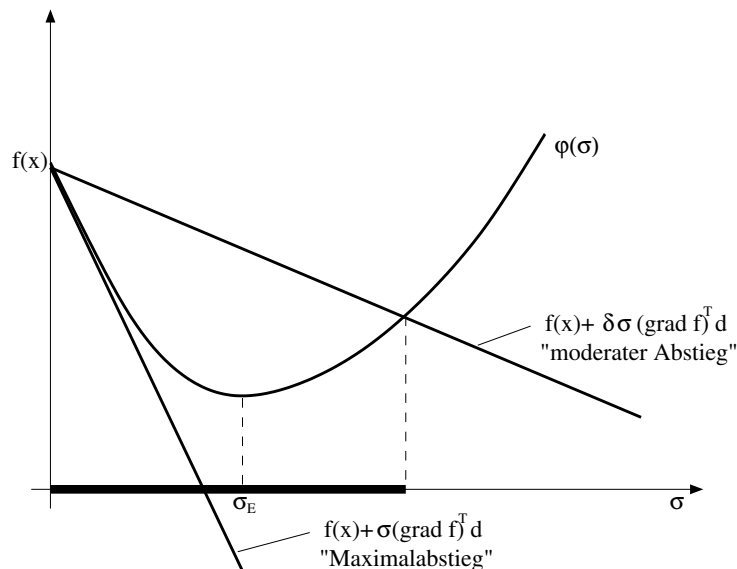
4.4.2 Schrittweite nach Armijo

Gegeben: $x \in \mathbb{R}^n$, Abstiegsrichtung d . Dann sieht die Sachlage wie folgt aus:

Man fordert für $\sigma = \sigma_A$

$$\bullet \quad f(x + \sigma_A d) \leq f(x) + \delta \sigma_A \nabla f(x)^T d \quad \text{Mindestabstieg} \quad (4.26)$$

$$\bullet \quad \sigma_A \geq -c_2 \frac{\nabla f(x)^T d}{\|d\|^2} \quad \text{Effizienz} \quad (4.27)$$



Verfahren 4.4.1 (Armijo-Goldstein)

0. Fixiere Konstanten

$$\begin{aligned} 0 < \delta < 1 & \quad \text{"Abflachung"} \\ \gamma > 0 & \quad \text{"Effizienzkonzst."} \\ 0 < \beta_1 \leq \beta_2 < 1 \end{aligned}$$

1. Startschrittweite

$$\sigma_0 \geq -\gamma \frac{\nabla f(x)^T d}{\|d\|^2}$$

$$j := 0$$

2. Wenn

$$f(x + \sigma_j d) \leq f(x) + \delta \sigma_j \nabla f(x)^T d,$$

dann setze $\sigma_A := \sigma_j$, fertig.

3. Ansonsten verkleinere σ_j so dass

$$\sigma_j \in [\beta_1 \sigma_j, \beta_2 \sigma_j]$$

$j := j + 1$, gehe zu 2.

Unter entsprechenden Voraussetzungen (d. h. (VNK), (VFD), (VFL)) findet das Verfahren nach endlich vielen Schritten eine Schrittweite, die (4.26) – (4.27) erfüllt (vgl. [1, Satz 4.4.3]).

Die erste Beziehung ist klar, denn $\sigma_j \leq \beta_2^j \sigma_0$ liegt irgendwann in diesem Bereich, siehe Skizze. Die zweite ist etwas kniffliger: l sei die Zahl der Iterationsschritte.

Gilt $l = 0$, so ist (4.27) mit $c_2 = \gamma$ erfüllt. Bei $l > 0$ liegt $s = \sigma_{l-1}$ noch außerhalb des akzeptablen Bereichs, also

$$\begin{aligned}
& \underbrace{f(x + sd) - f(x)}_{= \nabla f(x + \vartheta sd)^T ds}_{0 < \vartheta < 1} > \delta s \nabla f(x)^T d. \\
& \Rightarrow \nabla f(x + \vartheta sd)^T d = \frac{1}{s} [f(x + sd) - f(x)] > \delta \nabla f(x)^T d \quad | - \nabla f(x)^T d \\
& \Rightarrow -(1 - \delta) \nabla f(x)^T d < [\nabla f(x + \vartheta sd) - \nabla f(x)]^T d \leq \underset{\substack{\uparrow \\ \text{Lipsch.}}}{L \vartheta s \|d\|^2} \leq s L \|d\|^2 \\
& \Rightarrow \boxed{s \geq -\frac{(1-\delta)}{L} \frac{\nabla f(x)^T d}{\|d\|^2}}
\end{aligned}$$

Wegen $\sigma_A \geq \beta_1 s$ gilt am Ende die Beziehung

$$\begin{aligned}
\sigma_A & \geq - \underbrace{\frac{\beta_1(1-\delta)}{L}}_{c_2} \frac{\nabla f(x)^T d}{\|d\|^2} \\
c_2 & = \min \left\{ \gamma, \frac{\beta_1(1-\delta)}{L} \right\}
\end{aligned}$$

□

Bemerkung: Man kann z. B. $\beta_1 = \beta_2 = \frac{1}{2}$ wählen (Halbierung).

Zur Wahl der Verfahrensparameter: Siehe z.B. [1].

4.4.3 Schrittweite nach Powell

Dieses Verfahren wählt σ so, dass

$$f(x + \sigma d) \leq f(x) + \delta \sigma \nabla f(x)^T d \quad (\text{wie Armijo}) \quad (4.28)$$

und

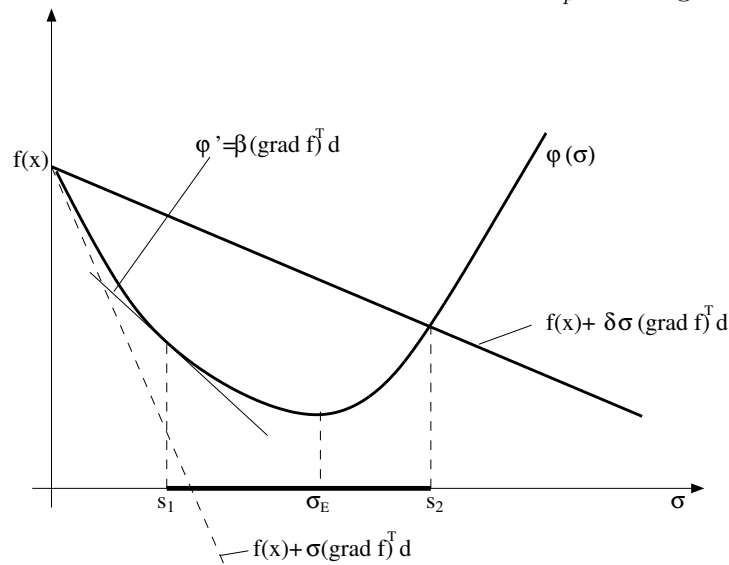
$$\nabla f(x + \sigma d)^T d \geq \beta \nabla f(x)^T d \quad (\text{Mindestschrittweite}) \quad (4.29)$$

mit $0 < \delta < \beta < 1$.

Geometrische Interpretation $\varphi(s) := f(x + sd)$. Dann gilt

$$\varphi'(s) = \nabla f(x + sd)^T d$$

Demnach bestimmt das Verfahren eine Schrittweite $\sigma = \sigma_p$ wie folgt:



Die Existenz einer solchen Schrittweite wird in [1, Satz 4.4.5] gezeigt. Die Bestimmung läuft über eine *Intervallschachtelung*.

Dazu definieren wir

$$G_1(\sigma) = \begin{cases} \frac{f(x+\sigma d) - f(x)}{\sigma \nabla f(x)^T d} & , \text{ für } \sigma > 0, \\ 1 & , \text{ für } \sigma = 0 \end{cases}$$

$$G_2(\sigma) = \frac{\nabla f(x + \sigma d)^T d}{\nabla f(x)^T d}$$

Dann ist $(4.28) \Leftrightarrow G_1(\sigma) \geq \delta$ und $(4.29) \Leftrightarrow G_2(\sigma) \leq \beta$

Geometrisch sieht das so aus, dass sich \mathbb{R}_+ in 3 Intervalle $[0, s_1) \cup [s_1, s_2] \cup (s_2, \infty) =: I_1 \cup I_2 \cup I_3$ unterteilen lässt, mit $\varphi'(s_1) = \beta \nabla f(x)^T d$ und $\varphi(s_2) = f(x) + s_2 \nabla f(x)^T d$ mit

$$\begin{aligned} G_1(\sigma) &\geq \delta \text{ und } G_2(\sigma) \geq \beta && \text{ in } I_1, \\ G_1(\sigma) &\geq \delta \text{ und } G_2(\sigma) \leq \beta && \text{ in } I_2, \\ G_1(\sigma) &\leq \delta \text{ und } G_2(\sigma) \leq \beta && \text{ in } I_3. \end{aligned}$$

Verfahren 4.4.2 (Powell)

1. Wahl von Startschrittweite $\sigma_0 > 0$, $j := 0$

(i) Gilt $G_1(\sigma_0) \geq \delta$ und $G_2(\sigma_0) \leq \beta$: Fertig! $\sigma_p := \sigma_0$

(ii) Liegt σ_0 in I_1

$$\begin{aligned} a_0 &:= \sigma_0 \\ b_0 &:= 2^l \sigma_0 \text{ mit } \underline{\text{minimalem}} \ l \in \mathbb{N}, \text{ so} \\ &\quad \text{dass } G_1(b_0) < \delta \end{aligned}$$

Gehe zu 2.

(iii) Liegt σ_0 in I_3

$$\begin{aligned} b_0 &= \sigma_0 \\ a_0 &= 2^{-l} \sigma_0 \text{ mit } \underline{\text{minimalem}} \ l \in \mathbb{N}, \\ &\quad \text{so dass } G_2(a_0) > \beta \text{ und} \\ &\quad G_1(a_0) \geq \delta \end{aligned}$$

2. Mittelwert $\sigma_j := \frac{1}{2}(a_j + b_j)$

(i) Liegt σ_j in I_2 : Fertig, $\sigma_p := \sigma_j$

(ii) Liegt σ_j in I_1 : Dann $a_{j+1} = \sigma_j, b_{j+1} = b_j$

(iii) Liegt σ_j in I_3 :

$$a_{j+1} = a_j, b_{j+1} = \sigma_j$$

3. $j := j + 1$, goto 2.

Das Powell-Verfahren kann die Schrittweite auch vergrößern, ausgehend von der Startschrittweite σ_0 , daher kann σ_0 an sich beliebig sein. Typische Wahlen von β und δ sind z. B. $\delta = 0.1$ und $\beta = 0.9$.

Bemerkungen:

- σ_p wird (unter entsprechenden Voraussetzungen) in endlich vielen Schritten berechnet (vgl. [1, Satz 4.5.10])
- Unter den Voraussetzungen (VNK), (VFD), (VFL) gilt folgendes allgemeines Konvergenzresultat:

Wird Schrittweite σ_k exakt, nach Armijo oder Powell gewählt, dann ist $\{\sigma_k\}$ Folge effizienter Schrittweiten

4.5 Das Gradientenverfahren

Es wird auch als “Verfahren des steilsten Abstiegs” bezeichnet. Als Richtung wählt man hier

$$d^{(k)} = -\nabla f(x^{(k)})$$

- Sehr einfach zu implementieren
- Am Ende aber zu langsam

Verfahren 4.5.1

1. Startvektor $x^{(0)}$, $k := 0$, Abbruchkriterium $\varepsilon > 0$.
2. Wenn $\|\nabla f(x^{(k)})\| < \varepsilon$: Fertig
3. Berechne

$$\begin{aligned} d^{(k)} &= -\nabla f(x^{(k)}) \\ \sigma_k &\text{ als effiziente Schrittweite (z. B. Armijo)} \\ x^{(k+1)} &:= x^{(k)} + \sigma_k d^{(k)} \\ k &:= k + 1, \text{ goto 2.} \end{aligned}$$

□

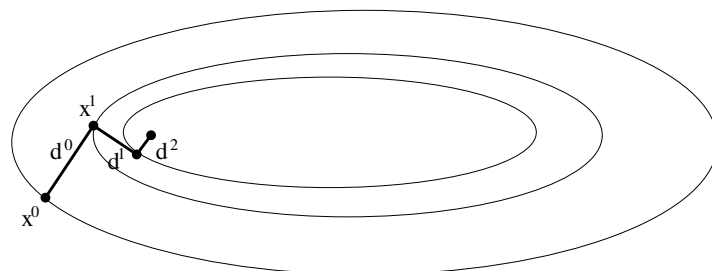
Unter den entsprechenden Voraussetzungen gelten die allgemeinen Konvergenzsätze.

Verfahrensnachteil: Die ersten Schritte sind noch schnell, aber “dann zieht sich’s”

Begründung: Benutzt man die exakte Schrittweite, was ja an sich nicht schlecht ist, dann

$$\begin{aligned} 0 &= \frac{\partial}{\partial \sigma} f(\underbrace{x^{(k)} + \sigma d^{(k)}}_{x^{(k+1)}}) \big|_{\sigma=\sigma_E} = \nabla f(x^{(k+1)})^T d^{(k)} \\ &= -d^{(k+1)} d^{(k)} \\ \Rightarrow \quad d^{(k+1)} &\perp d^{(k)} \end{aligned}$$

In schmalen Tälern führt das zu sehr langsamer Konvergenz!



Ausweg: Bessere Beachtung der Niveaulinien!

4.6 Gedämpftes Newton-Verfahren

4.6.1 Das Verfahren

Abstiegsrichtung = Newton-Richtung

$$d^{(k)} = -f''(x^{(k)})^{-1} \nabla f(x^{(k)})$$

Verfahren 4.6.1 1. Startpunkt $x^{(0)} \in \mathbb{R}^n, k := 0$

2. Wenn $\nabla f(x^{(k)}) = 0$. Fertig

3. Berechne $d^{(k)}$ aus

$$f''(x^{(k)}) d^{(k)} = -\nabla f(x^{(k)})$$

Schrittweite σ_k : effizient (z. B. Armijo o. Powell)

$$x^{(k+1)} := x^{(k)} + \sigma_k d^{(k)}$$

$k := k + 1$, goto 2.

4.6.2 Interpretation der Newton-Richtung

Setzen $A = f''(x)$; A sei positiv definit.

Neues Skalarprodukt, Norm in \mathbb{R}^n :

$$\begin{aligned} \langle x, y \rangle_A &:= x^T A y \\ \|x\|_A &:= \sqrt{\langle x, x \rangle_A} = (x^T A x)^{1/2} \end{aligned}$$

Man zeigt nun:

Lemma 4.6.1 Die Richtung

$$\bar{d} = \frac{A^{(-1)} \nabla f(x)}{\|A^{(-1)} \nabla f(x)\|}$$

löst unter der Voraussetzung $\nabla f(x) \neq 0$ die Aufgabe

$$\begin{aligned} \min \quad & \nabla f(x)^T d \\ \text{bei} \quad & \|d\|_A = 1, \end{aligned} \tag{4.30}$$

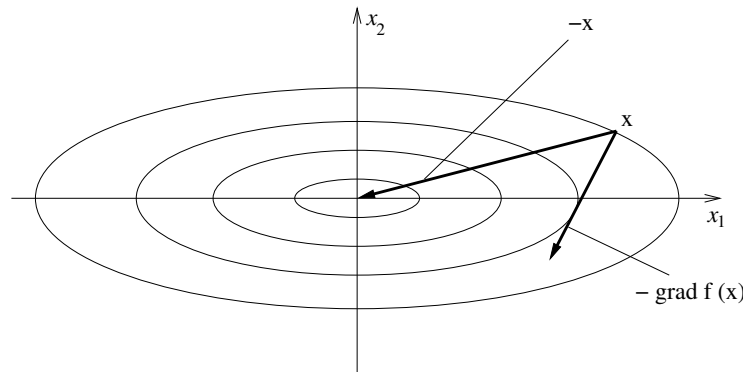
liefert also den steilsten Abstieg in der Norm $\|\cdot\|_A$.

Der Vorteil der Wahl von $-A^{(-1)} \nabla f$ anstelle der Gradientenrichtung $-\nabla f$ erschließt sich aus einer Betrachtung der quadratischen Funktion

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

z. B. bei $H = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ mit $a, b > 0$. Die Niveaulinien von $f(x)$ sind dann Ellipsen der Form $ax_1^2 + bx_2^2 = r^2$. Das Gradientenverfahren liefert eine Richtung, die am Nullpunkt – der Lösung der Aufgabe $\min f(x)$ – vorbeigeht.

Hingegen liefert $-f''(x)^{-1}\nabla f(x) = -H^{-1}Hx = -x$ genau die Richtung zur Lösung.



Folgerung:

Bei der quadratischen Funktion würde das gedämpfte Newton-Verfahren bei exakter Wahl der Schrittweite in genau einem Schritt konvergieren.

4.6.3 Konvergenz des Verfahrens

Es sei die gleichmäßige Konvexitätsbedingung (VLK) sowie die gleichmäßige Lipschitzstetigkeit von f'' in $N(f, f(x^{(0)}))$ erfüllt, d. h.

$$\|f''(x) - f''(y)\| \leq L \|x - y\| \quad \forall x, y \in N(f, f(x^{(0)})). \quad (VL2)$$

Damit sind die verwendeten Matrizen $f''(x^{(k)})$ positiv definit, und das Verfahren ist durchführbar.

Nach dem allgemeinen Konvergenzsatz 4.3.3 konvergiert damit das gedämpfte Newton-Verfahren wie eine linear konvergente Folge.

Aber es gilt mehr:

Satz 4.6.1 (VLK) sei erfüllt und die Schrittweiten σ_k nach Armijo oder Powell gewählt, wobei als Startschrittweite in jedem Schritt des gedämpften Newton-Verfahrens $\sigma_0 = 1$ gewählt wird. Weiter sei $0 < \delta < 1/2$. Dann gilt für alle hinreichend großen k die Bedingung $\sigma_k = 1$.

Beweis: Langwierig. Siehe [1, Satz 474].

Folgerung: Nach endlich vielen Schritten geht das gedämpfte Newton-Verfahren in das ungedämpfte über. Von da ab konvergiert es wie dieses, nämlich quadratisch, wenn (VL2) erfüllt ist, und sonst superlinear.

Definition 4.6.1 $\{x^{(k)}\}$ konvergiert superlinear gegen \tilde{x} , wenn

$$\lim_{n \rightarrow \infty} \frac{\|x^{(k+1)} - \tilde{x}\|}{\|x^{(k)} - \tilde{x}\|} = 0$$

Beispiel: $x_k = q^k, |q| < 1$, konvergiert nur linear gegen Null, nicht superlinear. Aber $x_k = \frac{q^k}{k!}$ konvergiert superlinear, denn

$$\frac{q^{k+1}}{(k+1)!} / \frac{q^k}{k!} = \frac{q}{k+1} \rightarrow 0, k \rightarrow \infty.$$

Verwendet man die exakte Schrittweite, dann gilt $\sigma^k \rightarrow 1, k \rightarrow \infty$, und man kann quadratische Konvergenz zeigen (vgl. [1, Satz 4.6.4]).

Bemerkung: (*Modifikation des Verfahrens*)

- Wählt man $f''(x^{(0)})$ anstelle $f''(x^{(k)})$ (vereinfachtes Newton-Verfahren), ergibt sich globale aber nur lineare Konvergenz.
- Nach jeweils n Schritten Neuberechnung (einer Approximation) von $f''(x^{(k)})$ führt zu superlinearer Konvergenz.
- Verwendet man Differenzenkoeffizienten zur Approximation der Ableitung, ergibt sich superlineare Konvergenz, falls die Diskretisierung fein genug ist.

4.7 Variable Metrik- und Quasi-Newton-Verfahren

Sinn der Verfahren: Ausgehend von Informationen über $f''(x)$ (oder solchen, die dem nahekommen), wird eine Norm $\|\cdot\|_A$ benutzt, welche die Krümmung der Niveaulinien berücksichtigt – wie bereits bei der quadratischen Funktion diskutiert.

4.7.1 Allgemeine Verfahrensvorschrift

Verfahren 4.7.1 (Variable Metrik)

1. Start $x^{(0)} \in \mathbb{R}^n, k := 0$
2. Wenn $\nabla f(x^{(k)}) = 0$: Fertig
3. Berechne:
 - positiv definite symmetrische Matrix $A^{(k)}$
 - $d^{(k)} = -(A^{(k)})^{(-1)} \nabla f(x^{(k)})$
 - effiziente Schrittweite σ_k
 - $x^{(k+1)} := x^{(k)} + \sigma_k d^{(k)}$
 - $k := k + 1$, goto 2.

Spezialfälle:

$A^{(k)} \equiv I$:	Gradientenverfahren
$A^{(k)} = f''(x^{(k)})$:	gedämpftes
	Newton-Verfahren

In jedem Schritt wird die steilste Richtung bezüglich der Norm $\|\cdot\|_{A^{(k)}}$ gewählt.

4.7.2 Globale Konvergenz von Variable-Metrik-Verfahren

Grundlegende Voraussetzung ist dabei gleichmäßige positive Definitheit und Beschränktheit der Matrizen $A^{(k)}$.

Definition 4.7.1 Eine Matrizenfolge $\{A^{(k)}\}$ von symmetrischen (n, n) -Matrizen heißt gleichmäßig positiv definit und beschränkt, wenn Konstanten $0 < \alpha_1 < \alpha_2$ existieren, so dass

$$\alpha_1 \|x\|^2 \leq x^T A^{(k)} x \leq \alpha_2 \|x\|^2 \quad \forall x \in \mathbb{R}^n$$

für alle $k \in \mathbb{N}$ gilt.

(Äquivalent dazu: Kleinster Eigenwert $\lambda_1^{(k)} \geq \alpha_1$, größter $\leq \alpha_2$ bzw.: kleinster EW. von $(A^k)^{-1} \geq \alpha_2^{-1}$, größter $\leq \alpha_1^{-1}$)

Im Vergleich mit den Sätzen über allgemeine Abstiegsverfahren ist es relativ plausibel, dass bei gleichmäßiger positiver Definitheit und Beschränktheit der gewählten Matrixfolge $\{A^{(k)}\}$ gilt:

- (VNK), (VFD) \Rightarrow Richtungen $d^{(k)}$ streng gradientenbezogen
- Analoge Konvergenzaussagen wie bei den Sätzen 4.3.1 (Häufungspunkt von $x^{(k)}$ mit $\nabla f = 0$), 4.3.2 (Konvergenz gegen Nullstelle von ∇f) sowie 4.3.3 (lineare Konvergenz)

4.7.3 Quasi-Newton-Methoden

In diesem Abschnitt behandeln wir zunächst die entsprechende Grundidee.

Nachteil des gedämpften Newton-Verfahrens: Aufwändige Berechnung von $f''(x^{(k)})$ für jeden neuen Schritt. Im Gegensatz dazu möchte man an Stelle von $f''(x^{(k)})$ eine Folge von Matrizen $\{A^{(k)}\}$ aufbauen mit:

- Übergang von $A^{(k)}$ zu $A^{(k+1)}$ ist einfach
- $A^{(k)}$ approximiert $f''(x^{(k)})$

und natürlich sollen alle positiv definit und symmetrisch sein.

Zur Motivation betrachten wir eine quadratische Funktion

$$f(x) = \frac{1}{2} x^T H x + b^T x.$$

Hier gilt $f''(x) = H$ und deshalb

$$\begin{aligned} f''(x^{(k+1)}) (x^{(k+1)} - x^{(k)}) &= H (x^{(k+1)} - x^{(k)}) + b - b \\ &= \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), \quad k = 0, 1, 2, \dots \end{aligned}$$

Kennen wir H nicht, sondern nur die Gradienten von f und die Vektoren $x^{(0)}, \dots, x^{(n-1)}$, so bestimmen die n Gleichungssysteme

$$H(x^{(k+1)} - x^{(k)}) = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), \quad k = 0, \dots, n-1$$

die Matrix H eindeutig.

Demgemäß fordert man von den $A^{(k)}$, ausgehend von $A^{(0)}$, einer positiv definiten symmetrischen Startmatrix

$$\boxed{A^{(k+1)}(x^{(k+1)} - x^{(k)}) = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})} \quad (4.31)$$

(4.31) heißt *Quasi-Newton-Gleichung*

4.7.4 BFGS-Update

Die Lösung der Quasi-Newton-Gleichung ist nicht eindeutig bestimmt. Man hat nun Formeln entwickelt, bei denen $A^{(k+1)}$ recht einfach berechnet werden kann. Am bekanntesten: BFGS-Formel (nach Broyden, Fletcher, Goldfarb, Shanno).

Man definiert:

$$\begin{aligned} x^{(k+1)} - x^{(k)} &=: s^{(k)} \\ \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}) &=: y^{(k)} \end{aligned}$$

Ausgehend von $A^{(k)}$ wird $A^{(k+1)}$ in zwei Schritten bestimmt:

- Zuerst

$$\boxed{\tilde{A}^{(k)} := A^{(k)} - \frac{(A^{(k)} s^{(k)})(A^{(k)} s^{(k)})^T}{(s^{(k)})^T A^{(k)} s^{(k)}}} \quad (4.32)$$

Ist $A^{(k)}$ bereits symmetrisch und positiv definit gewesen, so ist $\tilde{A}^{(k)}$ auch symmetrisch und zumindest positiv semidefinit. Außerdem gilt

$$\tilde{A}^{(k)} s^{(k)} = 0,$$

damit erfüllt \tilde{A}^k allein nicht die Quasi-Newton-Gleichg. Es gilt offenbar

$$\text{rang}(A^{(k)} s^{(k)}) (A^{(k)} s^{(k)})^T = 1,$$

deshalb heißt (4.32) *symmetrische Rang 1-Modifikation*.

- Durch eine zweite Rang-1-Modifikation versucht man, eine positiv definite Matrix zu bekommen:

$$A^{(k+1)} = \tilde{A}^{(k)} + \gamma_k w^{(k)} (w^{(k)})^T$$

und gleichzeitig die Quasi-Newton-Gl. zu erfüllen.

Quasi-Newton-Gleichung:

$$A^{(k+1)} s^{(k)} = \underbrace{\tilde{A}^{(k)}}_{=0} + \overbrace{\gamma_k w^{(k)} (w^{(k)})^T s^{(k)}}^{\in \mathbb{R}, \quad =: \frac{1}{\gamma_k}} \stackrel{(!)}{=} y^{(k)}$$

$\Rightarrow w^{(k)}$ muss Vielfaches von $y^{(k)}$ sein, wählen $w^{(k)} = y^{(k)}$ und

$$\gamma_k = \frac{1}{(y^{(k)})^T s^{(k)}}$$

Positive Definitheit:

Zumindest muss dann für die spezielle Richtung $s^{(k)}$ gelte

$$0 < (s^{(k)})^T A^{(k+1)} s^{(k)} \stackrel{\text{s.ö.}}{=} (s^{(k)})^T y^{(k)} \quad (4.33)$$

Man kann zeigen, dass die Bedingung $(s^{(k)})^T y^{(k)} > 0$ hinreichend für positive Definitheit von $A^{(k+1)}$ ist, wenn $A^{(k)}$ positiv definit war [1, Lemma 4.8.5].

Insgesamt:

$$\boxed{A^{(k+1)} = A^{(k)} - \frac{A^{(k)} s^{(k)} (A^{(k)} s^{(k)})^T}{(s^{(k)})^T A^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^T}{(y^{(k)})^T s^{(k)}}} \quad (4.34)$$

Da die Summe von zwei Rang 2 Matrizen in der Regel vom Rang 2 ist, spricht man von einer *Rang-2-Modifikation*.

Bemerkung:

(4.33) ist für eine quadratische Funktion erfüllt, falls H positiv definit ist:

$$\begin{aligned} f(x) &= \frac{1}{2} x^T H x & \nabla f &= H x \\ \Rightarrow (y^{(k)})^T s^{(k)} &= (\nabla f(x^{(k+1)}) - \nabla f(x^{(k)}))^T (x^{(k+1)} - x^{(k)}) \\ &= (H(x^{(k+1)} - x^{(k)}))^T (x^{(k+1)} - x^{(k)}) \\ &\geq \alpha \|x^{(k+1)} - x^{(k)}\|^2 > 0. \end{aligned}$$

□

Nebenrechnungen:

- Zeige $\tilde{A}^{(k)} s^{(k)} = 0$ (ohne Index k)

$$\begin{aligned} \tilde{A}s &= As - \frac{(As)(As)^T}{s^T As} s \\ &= \frac{1}{s^T As} [(s^T As) As - As \underbrace{s^T A^T s}_{=s^T As, \text{ da } A \text{ symmetrisch}}] \\ &= \frac{1}{s^T As} [As \underbrace{\{s^T As I - s^T As I\}}_{=0}] = 0. \end{aligned}$$

- Zeige Matrix vom Typ ss^T hat Rang 1:

$$\begin{aligned}
 ss^T = (s_i s_j) &= \begin{pmatrix} s_1 s_1 & s_1 s_2 & \dots & s_1 s_n \\ s_2 s_1 & s_2 s_2 & \dots & s_2 s_n \\ \vdots & & & \\ s_n s_1 & s_n s_2 & \dots & s_n s_n \end{pmatrix} \\
 &= \begin{pmatrix} s_1 \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}, s_2 \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}, \dots, s_n \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} \end{pmatrix}
 \end{aligned}$$

Spalten sind Vielfache von $s \Rightarrow$ Rang 1.

4.7.5 Das BFGS-Verfahren für quadratische Optimierungsprobleme

Wir wissen bereits: Bei quadratischer Funktion f ist

$$\sigma_E = \frac{\nabla f(x)^T d}{d^T H d}$$

die Formel für exakte Schrittweite. Die wenden wir an.

Verfahren 4.7.2 (BFGS für (QU))

1. Wähle Startvektor $x^{(0)}$, symmetrische positiv definite Startmatrix $A^{(0)}$, $k := 0$.
2. Wenn $\nabla f(x^{(0)}) = 0$: Fertig.
3. Berechne

$$\begin{aligned}
 d^{(k)} &= (A^{(k)})^{(-1)} \nabla f(x^{(k)}) \\
 \sigma_k &= \frac{\nabla f(x^{(k)})^T d^{(k)}}{(d^{(k)})^T H d^{(k)}} \\
 x^{(k+1)} &= x^{(k)} + \sigma_k d^{(k)} \\
 s^{(k)} &= x^{(k+1)} - x^{(k)} \\
 y^{(k)} &= \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}) \\
 A^{(k+1)} &\text{ als BFGS-Update} \\
 k &:= k + 1, \text{ goto 2.}
 \end{aligned}$$

Dabei werden in der Praxis die Matrizen $A^{(k+1)}$ etwas anders berechnet als direkt nach der Formel.

Besonders wichtig ist nun

Definition 4.7.2 (*H-Orthogonalität*).

Sei H symmetrische und positive definite (n, n) -Matrix. Die Vektoren $d^{(0)}, \dots, d^{(k)}, k < n$, heißen zueinander **konjugiert** bzw. **orthogonal** bezüglich H , wenn sie nicht Null sind und

$$\boxed{(d^{(i)})^T H d^{(j)} = 0, \quad 0 \leq i < j \leq k}$$

gilt.

Genau das tritt beim BFGS-Verfahren ein:

Satz 4.7.1 H sei symmetrisch und positiv definit. Dann berechnet das BFGS-Verfahren für (QU) in $m \leq n$ Schritten das Minimum \tilde{x} von f . Ist $m = n$, dann gilt $A^{(n)} = H$.

Beweisidee: Sei $\nabla f(x^{(0)}) \neq 0$ (sonst fertig). Nun werden $x^{(1)}, y^{(1)}, s^{(1)}$ mit $A^{(0)}$ wie im Verfahren berechnet. Dann wird $A^{(1)}$ berechnet und ist nach [1, Lemma 4.8.5] positiv definit.

$$\begin{aligned} \underbrace{x^{(1)} - x^{(0)}}_{s^{(0)}} &= \sigma_0 d^{(0)}, \quad H s^{(0)} = \nabla f(x^{(1)}) - \nabla f(x^{(0)}) = y^{(0)} \\ \Rightarrow \nabla f(x^{(1)}) &= \nabla f(x^{(0)}) + H s^{(0)} = \nabla f(x^{(0)}) + \sigma_0 H d^{(0)} \\ \Rightarrow \nabla f(x^{(1)})^T d^{(0)} &= \nabla f(x^{(0)})^T d^{(0)} + \underbrace{\sigma_0 (d^{(0)})^T H}_{= -\nabla f(x^{(0)})^T d^{(0)} \text{ nach Def. von } \sigma_0} d^{(0)} = 0 \quad (*) \end{aligned}$$

Wegen oben folgt $d^{(0)} = \sigma_0^{-1} s^{(0)}$, also

$$(d^{(0)})^T H d^{(1)} = \frac{1}{\sigma_0} \underbrace{(H s^{(0)})^T}_{y^{(0)T} \text{ s. o.}} \underbrace{d^{(1)}}_{(-A^{(1)})^{-1} \nabla f \text{ (BFGS-Verf.)}} = - \frac{\overbrace{(s^{(0)})^T : \text{Quasi-N.}}^{(y^{(0)})^T (A^{(1)})^{-1} \nabla f(x^{(1)})}}{\sigma_0}$$

Nun kommt die Quasi-Newton-Gl. für $A^{(1)}$ ins Spiel:

$$\begin{aligned} (A^{(1)})^{-1} y^{(0)} &= s^{(0)} \\ \Rightarrow \\ (d^{(0)})^T H d^{(1)} &= - \frac{(s^{(0)})^T \nabla f(x^{(1)})}{\sigma_0} = -\nabla f(x^{(1)})^T d^{(0)} = 0. \quad ((*)) \end{aligned}$$

Damit für $k = 1$ erhalten:

$$\left. \begin{aligned} \text{(i)} \quad & \nabla f(x^{(k)})^T d^{(i)} = 0 \\ \text{(ii)} \quad & (A^{(k)})^{-1} y^{(i)} = s^{(i)} \\ \text{(iii)} \quad & (d^{(i)})^T H d^{(k)} = 0 \end{aligned} \right\} \text{ für } 0 \leq i < k$$

Induktion $k \rightarrow k + 1 \dots$ liefert die Behauptung. □

Bemerkungen:

- Der Satz gilt nur bei exakter Rechnung und bei exakter Schrittweite
- alternativ könnte man gleich das Gleichungssystem

$$H\tilde{x} + b = 0$$

lösen – mit dem Cholesky-Verfahren.

- Ein Schritt BFGS entspricht im Rechenaufwand dem des ganzen Cholesky-Verfahrens. D.h. BFGS für quadratische Aufgaben lohnt sich nicht. Es ist auch mehr für nicht-lineare Probleme gedacht (z. B. so in MATLAB oder NAGLIB implementiert).

4.7.6 Das BFGS-Verfahren für nichtlineare Optimierungsaufgaben

Das Verfahren verläuft analog zum quadratischen Fall. Nur haben wir jetzt nicht mehr die exakte Schrittweite σ_E zur Verfügung, die sich im quadratischen Fall so gut berechnen lässt. Man kann beweisen

- Lineare Konvergenz bei Verwendung effizienter Schrittweiten unter Voraussetzung (VLK) [1, Satz 4.8.12].
- Gilt zusätzlich noch (V2L) und werden die Schrittweiten nach Armijo oder Powell gewählt, dann tritt superlineare Konvergenz ein. Die erzeugten Matrizen $A^{(k)}$ sind gem. positiv definit und beschränkt [1, Satz 4.8.13].

Es gilt nicht notwendig $A^{(k)} \rightarrow f''(\tilde{x})$, sondern

$$\lim_{k \rightarrow \infty} \frac{\| (A^{(k)} - f''(\tilde{x})) d^{(k)} \|}{\| d^{(k)} \|} \rightarrow 0.$$

4.8 Verfahren konjugierter Richtungen

4.8.1 CG-Verfahren für quadratische Optimierungsprobleme

Beim BFGS-Verfahren sind die erzeugten Richtungen zueinander H -orthogonal und man hat Konvergenz nach höchstens n Schritten. Nachteil: Die Matrizen $A^{(k)}$ müssen abgespeichert werden. Bei hoher Dimension n ist das ein Problem: Bei 10000 Unbekannten müssen $100 \cdot 10^6$ Elemente abgespeichert und berechnet werden.

Außerdem hat eine Matrix häufig eine besondere Struktur, die es erlaubt, Matrix-Vektor-Produkte effizient auszuführen, ohne dafür die Matrix aufbauen zu müssen. Die Hilbert-

Matrix z.B. ist definiert als

$$H_{ij} = \frac{1}{i+j+1}, \text{ d.h., es gilt}$$

$$[Hv]_i = \sum_{j=1}^n H_{ij}v_j = \frac{v_j}{i+j+1}.$$

Die Idee der CG-Verfahren (Conjugate Gradient) ist es, H -orthogonale Richtungen ohne A^k -update zu generieren.

Betrachte

$$(QU) \quad \min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T H x + b^T x, \quad H \text{ symmetrisch, positiv definit}$$

Lemma 4.8.1 Seien d_0, d_1, \dots, d_{n-1} konjugierte Richtungen. Für jedes $x_0 \in \mathbb{R}^n$ liefert

$$x^{k+1} = x^k + \sigma_k d^k$$

$$\sigma^k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T H d^k} \quad (\text{exakte Schrittweite})$$

nach höchstens n Schritten die Lösung $x^n = -H^{-1}b$.

BEWEIS.

$$\tilde{x} - x^0 = \sum_{i=0}^{n-1} \sigma_i d^i \quad (\text{Mult. von links mit } (d^i)^T) \cdot H$$

$$\Rightarrow \quad \sigma_k = \frac{(d^k)^T H(\tilde{x} - x^0)}{(d^k)^T H d^k} = -\frac{(d^k)^T (H x^0 + b)}{(d^k)^T H d^k}$$

$$\dots = -\frac{(d^k)^T (H x^k + b)}{(d^k)^T H d^k} = -\frac{(d^k)^T \nabla f(x^k)}{(d^k)^T H d^k}$$

□

Korollar 4.1 x^k minimiert f nicht nur auf $\{x^{k-1} + \sigma d^{k-1} | \sigma \in \mathbb{R}\}$ sondern auch auf $x_0 + V_k$, mit $V_k = \text{span}\{d^0, \dots, d^{k-1}\}$. Insbesondere gilt

$$d_i^T \nabla f(x^k) = 0 \quad \text{für } i < k. \quad (*)$$

BEWEIS. Es genügt, $(*)$ zu zeigen.

Es gilt $d_i^T \nabla f(x^{i+1}) = d_i^T (H x^{i+1} + b) = d_i^T (\underbrace{H x^i + b}_{=\nabla f(x^i)} + \sigma_i d_i^T H d_i) = 0$ d.h. $(*)$ wahr für

$k = 1$ und für $i = k - 1$ falls $k > 1$. Es gilt $\nabla f(x^{k+1}) - \nabla f(x^k) = H(x^{k+1} - x^k) = \sigma_k H d^k$
 $\Rightarrow (d^i)^T (\nabla f(x^{k+1}) - \nabla f(x^k)) = 0$ für $k > i$ □

Verfahren 4.8.1 (*Konjugierte Richtungen*)

1. Wähle x^0 , berechne $d^0 = -\nabla f(x^0) = -(Hx^0 + b)$
 $k := 0$
2. Wenn $\nabla f(x^k) = 0 \rightarrow$ fertig
3. Berechne

$$\sigma_k = \frac{\nabla f(x^k)^T \nabla f(x^k)}{(d^k)^T H d^k}$$

$$x^{k+1} = x^k + \sigma_k d^k$$

$$\nabla f(x^{k+1}) = Hx^{k+1} + b = \nabla f(x^k) + \sigma_k H d^k$$

$$\beta_k = \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2}$$

$$d^{k+1} = -\nabla f(x^{k+1}) + \beta_k d^k.$$

Bemerkung: σ_k entspricht der exakten Schrittweite, denn

$$\begin{aligned} \sigma_E &= -\frac{\nabla f(x^k)^T d^k}{(d^k)^T H d^k} = -\frac{\nabla f(x^k)^T (-\nabla f(x^k) + \beta_{k-1} d^{k-1})}{(d^k)^T H d^k} \\ &= \frac{\nabla f(x^k)^T \nabla f(x^k)}{(d^k)^T H d^k} \end{aligned}$$

Satz 4.8.1 (*Eigenschaften des CG-Verfahrens*). Solange $\nabla f(x^{k-1}) \neq 0$ gelten folgende Aussagen:

$$(1) \quad d^{k-1} \neq 0$$

$$(2)$$

$$\begin{aligned} V_k &:= \text{span}\{\nabla f(x^0), H\nabla f(x^0), \dots, H^{k-1}\nabla f(x^0)\} \\ &= \text{span}\{\nabla f(x^0), \dots, \nabla f(x^{k-1})\} \\ &= \text{span}\{d^0, \dots, d^{k-1}\} \end{aligned}$$

$$(3) \quad d^0, \dots, d^{k-1} \text{ sind paarweise konjugiert}$$

$$(4) \quad f(x^k) = \min_{z \in V_k} f(x^0 + z)$$

BEWEIS. Für $k = 1$ okay, Behauptung gelte für $k - 1$. Zur Abkürzung definieren wir $g^k := \nabla f(x^k)$. Dann

$$\begin{aligned} g^k &= g^{k-1} + \sigma_{k-1} H d^{k-1} \\ \Rightarrow g^k &\in V_{k+1} \quad \text{und} \quad \text{span}\{g^0, \dots, g^k\} \subset V_{k+1}. \end{aligned}$$

Nach Induktionsvoraussetzung sind d^0, \dots, d^{k-1} konjugiert.

Folgerung 4.1 \Rightarrow

$$(d^i)^T g^k = 0 \quad \text{für } i < k \quad (*)$$

$g^k \neq 0 \Rightarrow \{d^0, \dots, d^{k-1}, g^k\}$ und damit auch $\{g^0, \dots, g^{k-1}, g^k\}$ sind linear unabhängig mit Dimension $k+1$

$$\Rightarrow \text{span}\{g^0, \dots, g^k\} = V_{k+1}.$$

Es gilt $g^k + d^k = \beta_{k-1} d^{k-1} \in V_k$

$$\Rightarrow V_{k+1} = \text{span}\{d^0, \dots, d^k\} \Rightarrow (2).$$

Wegen $g^k + d^k \in V_k$ folgt $d^k \neq 0$ falls $g^k \neq 0 \Rightarrow (1)$.

(3) ergibt sich durch längeres rumrechnen unter Ausnutzung von (*) und (2).

(4) folgt aus Lemma 4.8.1. □

4.8.2 Analyse des CG-Verfahrens

Für symmetrische, positiv definite (n, n) - Matrizen H mit Eigenwerten $\lambda_1 < \dots < \lambda_n$ ist die Kondition definiert durch

$$\kappa(H) = \frac{\lambda_n}{\lambda_1}.$$

Wendet man das Gradientenverfahren mit exakter Schrittweite auf unser quadratisches Optimierungsproblem an, so ergibt sich für den Fehler in der Energienorm $\|x\|_H := \sqrt{x^T H x}$ (vgl. z.B. [7]):

$$\|\tilde{x} - x^{k+1}\|_H \leq \left(\frac{\kappa(H) - 1}{\kappa(H) + 1} \right)^k \|\tilde{x} - x^0\|_H.$$

Für das Verfahren der Konjugierten Gradienten ergibt sich folgende, bessere Abschätzung:

Satz 4.8.2 *Der Approximationsfehler von $\tilde{x} - x^k$ im CG-Verfahren lässt sich in der Energienorm $\|y\|_H = \sqrt{y^T H y}$ abschätzen durch*

$$\|\tilde{x} - x^k\|_H \leq 2 \left(\frac{\sqrt{\kappa(H)} - 1}{\sqrt{\kappa(H)} + 1} \right)^k \|\tilde{x} - x^0\|_H$$

BEWEIS. Nach Satz 4.8.1 gilt

$$\|\tilde{x} - x^k\| \leq \|\tilde{x} - y\| \quad \forall y \in V_k. \quad (*)$$

Ebenfalls nach Satz 4.8.1 lässt sich $y \in V_k$ als Linearkombination von Potenzen von H angewendet auf g^0 schreiben. D.h. es gibt ein Polynom P_{k-1} vom Grad $k-1$ so dass

$$\begin{aligned} y &= x^0 + P_{k-1}(H)g^0 = x^0 + P_{k-1}(H)(Hx^0 + b) \\ &= x^0 + HP_{k-1}(H)(x^0 - \tilde{x}) \\ \Rightarrow \tilde{x} - y &= x - x_0 - HP_{k-1}(H)(x^0 - \tilde{x}) \\ &= \underbrace{(I + HP_{k-1}(H))}_{=: Q_k(H)}(\tilde{x} - x^0) \end{aligned}$$

mit einem Polynom $Q_k \in \mathcal{P}_k$ vom Grad k mit $Q_k(0) = 1$.

Sei $\{z_1, \dots, z_n\}$ ein Orthonormalsystem System aus Eigenvektoren von H , dann gilt

$$\begin{aligned}
\tilde{x} - x^0 &= \sum_{j=1}^n c_j z_j \\
\Rightarrow \tilde{x} - y &= \sum_{j=1}^n c_j Q_k(H) z_j = \sum_{j=1}^n c_j Q_k(\lambda_j) z_j \\
\Rightarrow \|\tilde{x} - y\|_H^2 &= \left[\sum_{j=1}^n c_j Q_k(\lambda_j) z_j \right]^T H \left(\sum_{j=1}^n c_j Q_k(\lambda_j) z_j \right) \\
&= \sum_{j=1}^n \lambda_j c_j^2 Q_k^2(\lambda_j) \\
&\leq \min_{\substack{Q_k \in \mathcal{P}_k \\ Q_k(0)=1}} \max_{\lambda} |Q_k(\lambda)|^2 \underbrace{\sum_{j=1}^n \lambda_j c_j^2}_{=\|\tilde{x}-x^0\|_H^2} .
\end{aligned}$$

Wählt man als Polynome die Chebyshev-Polynome vom Grad $\leq k$ so erhält man nach Transformations des Definitionsbereichs auf $[\lambda_1, \lambda_n]$ die Abschätzung

$$\begin{aligned}
\alpha &:= \min_{\substack{Q_k \in \mathcal{P}_k \\ Q_k(0)=1}} \max_{1 \leq i \leq n} |Q_k(\lambda_i)| \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \\
&\quad \text{mit } \kappa = \kappa(H) = \frac{\lambda_n}{\lambda_1}.
\end{aligned}$$

4.8.3 Vorkonditionierung

Der letzte Satz besagt, dass für das CG-Verfahren gute Konvergenz zu erwarten ist, falls die Kondition der Matrix klein ist. Die Idee der Vorkonditionierung besteht darin, das Problem so zu modifizieren, dass die Kondition der resultierenden Systemmatrix klein ist, d.h. die Höhenlinien von f sollen so gut wie möglich Kreise approximieren.

Im folgenden wählen wir eine positiv definite Matrix B und betrachten das Problem

$$\bar{H}\bar{x} = -b, \quad \text{mit } \bar{x} = B^{-1}x \quad \text{und } \bar{H} = H \cdot B.$$

Achtung: $\bar{H} = H \cdot B$ nicht selbstadjungiert bzgl. des euklidischen Skalarprodukts aber bzgl. $(\cdot, \cdot)_B$ = denn

$$(x, HBy)_B = x^T B H B y = (HBx)^T B y = (HB, y)_B$$

Die wesentliche Idee des vorkonditionierten CG-Verfahrens ist, das obige Ersatzproblem zu lösen, wobei das euklidische Skalarprodukt (\cdot, \cdot) durch $(\cdot, \cdot)_B$ ersetzt werden muss. Details finden sich z. B. im Buch von Deuffhard und Hohmann.¹

¹Deuffhard/Hohmann: *Numerische Mathematik 1*. de Gruyter, Berlin 1993.

Für den Approximationsfehler kann man dann zeigen:

$$\|\tilde{x} - x^k\|_H \leq 2 \left(\frac{\sqrt{\kappa(H \cdot B)} - 1}{\sqrt{\kappa(H \cdot B)} + 1} \right)^k \|\tilde{x} - x^0\|_A.$$

Die praktische Aufgabe der Vorkonditionierung besteht nun darin eine positiv definite symmetrische Matrix B zu finden so dass einerseits Produkte By einfach auszuwerten sind und andererseits die Kondition $\kappa(HB)$ „klein“ ist.

Typische Beispiele sind

- (i) $B = D^{-1}$, wobei $D = \text{diag}(H)$ die Diagonalmatrix mit den Diagonalelementen von H ist.
- (ii) Unvollständige Cholesky-Zerlegung von H .

4.8.4 CG-Verfahren für nichtlineare Optimierungsprobleme

Das Verfahren wurde zuerst von Fletcher und Reeves untersucht, deshalb wird es auch als Fletcher-Reeves-Verfahren bezeichnet.

Es verläuft völlig analog zum letzten Verfahren. Unter der (theoretischen) Annahme exakter Schrittweiten $\sigma_k = \sigma_E$ kann man Konvergenz zeigen (vgl. z.B. [1, Satz 4.9.4]).

4.9 Trust-Region-Verfahren

4.9.1 Motivation

Idee: Bisher haben wir zunächst Richtung d^k berechnet und dann die Schrittweite σ^k durch eindimensionale Minimierung

Jetzt:

- (i) Verwende lokales Modell f_k von f z.B.
 - $f_k(d) = f(x^k) + \nabla f(x^k)^T d$
 - $f_k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T f''(x^k) d$
- (ii) wähle $\varrho_k > 0$ und definiere $B(x^k, \varrho_k)$ als trust region
- (iii) Berechne d^k als globale Lösung von

$$\min_{\|d\| \leq \varrho_k} f_k(d). \quad (4.35)$$

Bemerkung: Falls $f_k^{(d)} = f(x^k) + \nabla f(x^k)^T d$

$$\text{gilt} \quad d^k = -\varrho_k \frac{\nabla f(x^k)}{\|\nabla f(x^k)\|}$$

→ wie im Gradientenverfahren, also keine guten Konvergenzeigenschaften zu erwarten.

Forderung an ein Modell f_k :

Gegeben $x^k \in \mathbb{R}^n$, $\varrho_k > 0$ dann

$$(i) \quad f_k(0) = f(x^k)$$

(ii) für d^k Lösung von (4.35) gilt:

$$f_k(d^k) = f(x^k) \Rightarrow \nabla f(x^k) = 0$$

(Abbruchkriterium!)

Beispiel 4.9.1 $f \in C^2$ und $f_k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T f''(x^k) d$ dann gilt $f_k(0) = f(x^k)$.

Falls $f_k(d^k) = f(x^k)$ dann

$$f_k(d^k) = f(x^k) \leq f_k(d) \quad \forall d \in B(0, \varrho_k)$$

d.h. $\tilde{d} = 0$ Lösung von $\min_{\|d\| \leq \varrho_k} F(d)$ mit

$$F(d) = \nabla f(x^k)^T d + \frac{1}{2} d^T f''(x^k) d$$

Optimalitätsbedingung für lokales Minimum von F liefert

$$0 = \nabla F(\tilde{d}) = \nabla f(x^k) + f''(x^k) \tilde{d} = \nabla f(x^k).$$

Zur Wahl von ϱ_k :

$$\text{berechne} \quad r_k = \frac{f(x^k) - f(x^k + d^k)}{f(x^k) - f_k(d^k)} = \frac{\text{tatsächlicher Abstieg}}{\text{Modellabstieg}}.$$

(Je näher r_k bei 1 umso größer das Vertrauen.)

Genauer: Wähle $0 < \delta_1 < \delta_2 < 1$.

Falls

$$r_k \begin{cases} \in [\delta_1, \delta_2[: & x^{k+1} = x^k + d^k, \quad \varrho_k \text{ unverändert} \\ \geq \delta_2 : & x^{k+1} = x^k + d^k, \quad \text{vergrößere } \varrho_k \\ < \delta_1 : & x^{k+1} = x^k, \quad \text{verkleinere } \varrho_k \end{cases}$$

4.9.2 Trust-Region-Newton-Verfahren

$$\min_{\|d\| \leq \varrho_k} \underbrace{f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T f''(x^k) d}_{=: f_k(d)} \quad (4.36)$$

Bemerkung: Bis auf Restriktion entspricht (4.36) $(Q)_k$, aber (4.36) immer lösbar, $(Q)_k$ nur wenn $f''(x)$ positiv definit. Falls ϱ_k klein genug, ist f_k gutes Modell und (4.36) liefert Abstieg auch, wenn $f''(x^k)$ nicht positiv definit ist.

Im Folgenden wird betrachtet

$$\min_{\|d\| \leq \varrho} \underbrace{c + b^T d + \frac{1}{2} d^T A d}_{=: \phi(d)} . \quad (4.37)$$

Definiere $g(d) := -\frac{1}{2}(\|d\|^2 - \varrho^2)$ dann $\|d\| \leq \varrho \Leftrightarrow g(d) \geq 0$ und $\nabla g = -d$.

Einführung der Lagrange-Funktion

$$\phi_\lambda(d) = \phi(d) - \lambda g(d) = -\frac{1}{2} \varrho^2 \lambda + c + b^T d + \frac{1}{2} d^T (A + \lambda I) d .$$

Lemma 4.9.1 (vgl. Kap. 5)

Sei $\tilde{d} \in \mathbb{R}^n$ mit $\|\tilde{d}\| \leq \varrho$ und $\lambda \geq 0$, so dass \tilde{d} (striktes) globales Minimum von ϕ_λ ist und $\lambda g(\tilde{d}) = 0$, dann ist \tilde{d} auch (striktes) globales Minimum von (4.37).

BEWEIS. \tilde{d} globales Minimum von ϕ_λ und $\lambda g(\tilde{d}) = 0$

also folgt $\phi(\tilde{d}) = \phi_\lambda(\tilde{d}) \leq \phi_\lambda(d) = \phi(d) - \underbrace{\lambda g(d)}_{\geq 0} \leq \phi(d)$

(ebenso für striktes Minimum). □

In Kapitel 5 werden wir beweisen

Lemma 4.9.2 Sei $\tilde{d} \in \mathbb{R}^n$ globales Minimum von (4.37). Dann existiert genau ein Lagrange-Multiplikator $\lambda > 0$, so dass gilt:

$$\nabla \phi_\lambda(\tilde{d}) = \nabla \phi(\tilde{d}) + \lambda \tilde{d} = 0 \quad (\text{Notwendige Optimalitätsbedingung 1. Ordnung})$$

$$d^T \phi''_\lambda(\tilde{d}) d = d^T (A + \lambda I) d \geq 0 \quad \forall d \in \mathbb{R}^n \quad (2. \text{ Ordnung})$$

$$\lambda g(\tilde{d}) = 0 \quad (\text{Komplementarität})$$

Satz 4.9.1 Sei $\tilde{d} \in \mathbb{R}^n$ mit $\|\tilde{d}\| \leq \varrho$. Dann gilt:

\tilde{d} globale Lösung von (4.37).

\Leftrightarrow es existiert $\lambda \geq 0$ mit

$$(i) \quad (A + \lambda I) \tilde{d} = -b .$$

$$(ii) \quad \lambda(\|\tilde{d}\| - \varrho) = 0, \text{ d.h. } \|\tilde{d}\| = \varrho \text{ falls } \lambda > 0 .$$

$$(iii) \quad A + \lambda I \text{ positiv semidefinit.}$$

Falls $A + \lambda I$ positiv definit, ist \tilde{d} eindeutig bestimmt.

BEWEIS. “ \Rightarrow ” (Lemma 4.9.2)

“ \Leftarrow ” (i)–(iii) erfüllt mit $\lambda \geq 0$

$$\left. \begin{array}{l} \text{(ii)} \Rightarrow \lambda g(\tilde{d}) = 0 \\ \text{(iii)} \Rightarrow \phi_\lambda \text{ konvex auf } \mathbb{R}^n \\ \text{(i)} \Rightarrow \nabla \phi_\lambda(\tilde{d}) = (A + \lambda I)\tilde{d} + b = 0 \end{array} \right\} \Leftrightarrow$$

\tilde{d} globales Minimum von $\phi_\lambda \xrightarrow{\text{Lemma 4.9.1}} \tilde{d}$ globales Minimum von (4.37). Zusatz folgt analog. \square

Für die Wahl des Abbruchkriteriums zeigen wir:

Lemma 4.9.3 $\tilde{d} \in \mathbb{R}^n$ global von (4.37) mit $\|\tilde{d}\| \leq \varrho$. Dann gilt

$$\phi(\tilde{d}) = c \Leftrightarrow b = 0 \quad \text{und} \quad A \text{ positiv semidefinit}$$

d.h. $f_k(d^k) = f(x^k)$

BEWEIS.

$$\phi(0) = c = \phi(\tilde{d}) \leq \phi(d) \Rightarrow 0 \text{ ist Lösung von (4.37).}$$

Satz (i) $\Rightarrow b = 0$, (ii) $\Rightarrow \lambda = 0$ (iii) $\Rightarrow A$ positiv semidefinit

“ \Leftarrow ” klar \square

Anwendung auf (4.35), d.h.

$$f_k = f(x_k) + \nabla f(x^k)^T d + \frac{1}{2} d^T f''(x^k) d.$$

Sei d^k globale Lösung von (4.35), dann

$$f_k(d^k) = f(x^k) \Leftrightarrow \nabla f(x^k) = 0 \quad \text{und} \quad f''(x^k) \text{ positiv semidefinit}$$

d.h. $f_k(d^k) = f(x^k)$ ist sinnvolles Abbruchkriterium.

Lemma 4.9.4 (Abschätzung des Abstiegs).

Sei $\tilde{d} \in \mathbb{R}^n$ mit $\|\tilde{d}\| \leq \varrho$ globale Lösung von (4.37), dann gilt

$$c - \phi(\tilde{d}) \geq \frac{1}{2} \|b\| \min \left\{ \varrho, \frac{\|b\|}{\|A\|} \right\}.$$

BEWEIS. $d = 0$ ist zulässig $\phi(0) \geq \phi(\tilde{d})$

$\Rightarrow c - \phi(\tilde{d}) \geq c - \phi(0) = 0$ also sei ohne Einschränkung der Allgemeinheit $b \neq 0$ dann

$$\begin{aligned} c - \phi(\tilde{d}) &= -b^T d - \frac{1}{2} d^T A d \\ &\geq -b^T d - \frac{1}{2} \|A\| \|d\|^2 \end{aligned}$$

Fallunterscheidung:

(i) $\varrho \|A\| \leq \|b\|$, setze $d = -\varrho \frac{b}{\|b\|}$

dann $c - \tilde{\phi}(\tilde{d}) \geq \varrho \|b\| - \frac{\varrho^2}{2} \|A\| \geq \frac{\varrho}{2} \|b\|$

(ii) $\varrho \|A\| > \|b\|$, setze $d = -\frac{1}{\|A\|} b$

dann $c - \phi(\tilde{d}) \geq \frac{\|b\|^2}{\|A\|} - \frac{1}{2} \frac{\|b\|^2}{\|A\|} = \frac{1}{2} \frac{\|b\|^2}{\|A\|}$

$\Rightarrow c - \phi(\tilde{d}) \geq \frac{1}{2} \|b\| \min \left\{ \varrho, \frac{\|b\|}{\|A\|} \right\}.$

Verfahren: (Trust-Region-Newton)

Gegeben: $0 < \delta_1 < \delta_2 < 1$, $\sigma_1 \in]0, 1[$, $\sigma_2 > 1$, $\varrho_0 > 0$

1. Wähle $x^0 \in \mathbb{R}^n$
2. Berechne globale Lösung d^k von

$$\min_{\|d\| \leq \varrho_k} \underbrace{f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T f''(x^k) d}_{=: f_k(d)}$$

Falls $f(x^k) = f_k(d^k) \rightarrow \text{Stop.}$

3. Berechne

$$r_k = \frac{f(x^k) - f(x^k + d^k)}{f(x^k) - f_k(d^k)}.$$

Falls $r_k \geq \delta_1$ (erfolgreicher Schritt)

- setze $x^{k+1} = x^k + d^k$,
- berechne $\nabla f(x^{k+1})$, $f''(x^{k+1})$.
- Aktualisiere ϱ_k :
 falls $r_k \in [\delta_1, \delta_2[$, wähle $\varrho_{k+1} \in [\sigma_1 \varrho_k, \varrho_k]$
 falls $r_k \geq \delta_2$, wähle $\varrho_{k+1} \in [\varrho_k, \delta_2, \varrho_k]$.
- Goto 2.

4. Falls $r_k < \delta_2$ (nicht erfolgreicher Schritt)

- wähle $\varrho_{k+1} \in]0, \sigma_1, \varrho_k]$,
- setze $x^{k+1} = x^k$, $\nabla f(x^{k+1}) = \nabla f(x^k)$, $f''(x^{k+1}) = f''(x^k)$
- $k \rightarrow k + 1$, Goto 2.

5 Probleme mit Restriktionen – Theorie

5.1 Einführende Beispiele

Wir betrachten

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{unter} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in I \end{cases}$$

mit Indexmengen $I, \mathcal{E} \in \{1, \dots, n\}$ und $I \cap \mathcal{E} = \emptyset$.

$c_i(x) = 0, i \in \mathcal{E}$ heißen Gleichungsnebenbedingungen

$c_i(x) \geq 0, i \in I$ heißen Ungleichungsnebenbedingungen.

$\Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0, i \in \mathcal{E}, c_i(x) \geq 0, i \in I\}$ heißt die zulässige Menge, damit kann man kurz schreiben

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

Bezeichnung: Sei $x \in \Omega$, d.h. zulässig, dann heißt die Ungleichungsrestriktion $i \in I$ aktiv, falls gilt $c_i(x) = 0$ und inaktiv, falls $c_i(x) > 0$.

$\mathcal{A}(x) = \mathcal{E} \cup \{i \in I \mid c_i(x) = 0\}$ heißt aktive Menge (in $x \in \Omega$).

Die folgenden 3 Beispiele illustrieren die Problematik der restringierten Optimierung.

Beispiel 5.1.1 (*Eine Gleichungsnebenbedingung*)

$$\min x_1 + x_2 \quad \text{bei} \quad x_1^2 + x_2^2 - 2 = 0$$

d.h.

$$f(x) = x_1 + x_2, \quad I = \emptyset \quad \mathcal{E} = \{1\} \quad \text{und} \quad c_1(x) = x_1^2 + x_2^2 - 2, \quad \nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla c_1 = 2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

$$\text{Als Lösung ergibt sich } \tilde{x} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

In \tilde{x} gilt

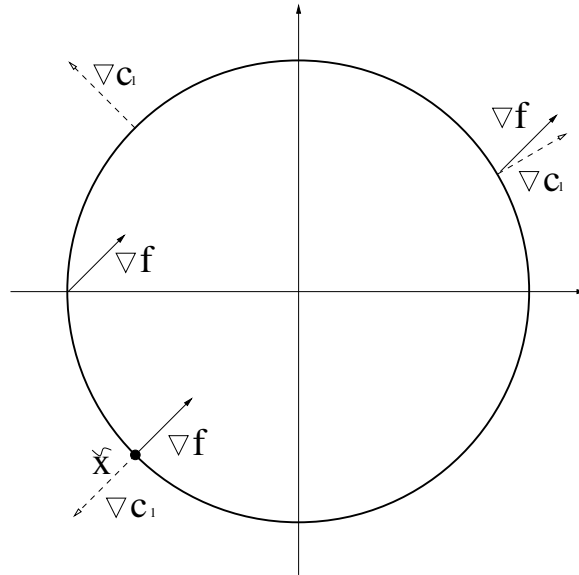
$$\nabla f(\tilde{x}) = \lambda_1 \nabla c_1(\tilde{x}) \quad \text{mit} \quad \lambda_1 = -\frac{1}{2}. \quad (5.1)$$

Zur Herleitung von (5.1) untersuchen wir die Taylorapproximation 1. Ordnung von Zielfunktion und Restriktion.

Um zulässig zu bleiben, gilt für kleine Störung s :

$$c_1(x + s) = 0, \text{ also}$$

$$0 = c_1(x + s) \approx c_1(x) + \nabla c_1(x)^T s = \nabla c_1(x)^T s$$



d.h. für Zulässigkeit sollte gelten

$$\nabla c_1(x)^T s = 0. \quad (5.2)$$

Um einen Abstieg von f zu erreichen, muss gelten

$$0 > f(x + s) - f(x) \approx \nabla f(x)^T s \quad \text{d.h.}$$

in 1. Ordnung sollte gelten

$$\nabla f(x)^T s < 0. \quad (5.3)$$

Nehmen wir an, dass (5.2) und (5.3) für beliebige Richtungen gelten muss, d.h.

$$\nabla c_1^T(x) d = 0 \quad \text{und} \quad \nabla f(x)^T s < 0,$$

dann kann man sich anhand der Skizze schnell überlegen, dass man für alle zulässigen x einen Abstieg findet, nur nicht dort, wo ∇f und ∇c_1 parallel sind.

Nun führen wir die Lagrange-Funktion ein:

$$\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x).$$

Es gilt

$$\nabla_x \mathcal{L} = \nabla f(x) - \lambda_1 \nabla c_1(x)$$

und damit ist (5.1) äquivalent dazu, dass für $x = \tilde{x}$ ein $\tilde{\lambda}$ existiert, so dass

$$\nabla_x \mathcal{L}(\tilde{x}, \tilde{\lambda}) = 0.$$

Beispiel 5.1.2 (Eine Ungleichungsnebenbedingung)

$$\min x_1 + x_2 \quad \text{bei} \quad 2 - x_1^2 - x_2^2 \geq 0$$

d.h. hier ist $\mathcal{E} = \emptyset$, $I = \{1\}$ und $c_1 = 2 - x_1^2 - x_2^2$, $\nabla c_1 = -2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $\nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

die Lösung ist wieder $\tilde{x} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$, und es gilt

$$\nabla f(\tilde{x}) = \tilde{\lambda}_1 \nabla c_1(\tilde{x}) \quad \text{mit} \quad \tilde{\lambda}_1 = \frac{1}{2}.$$

Bedingung für Abstieg ist wie zuvor

$$\nabla f(x)^T s < 0. \quad (5.4)$$

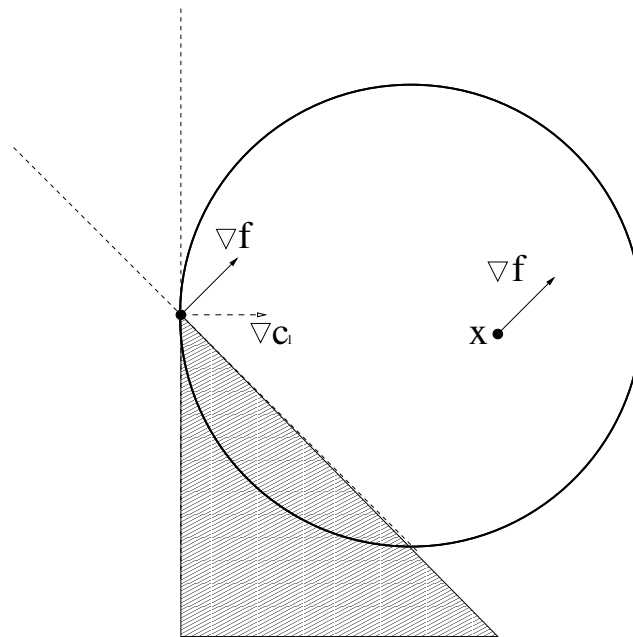
Für Zulässigkeit muss gelten

$$0 \leq c_1(x + s) \approx c_1(x) + \nabla c_1(x)^T s,$$

d.h. in 1. Ordnung

$$c_1(x) + \nabla c_1(x)^T s \geq 0. \quad (5.5)$$

Um zu untersuchen, wann (5.4) und (5.5) gelten, machen wir eine Fallunterscheidung.



Fall 1: x ist inaktiv, d.h. $c_1(x) > 0$ dann gilt (5.5) $\forall s \in \mathbb{R}^n$ mit $\|s\|$ klein genug. Für (5.4) wähle z.B. $s = -\alpha \nabla f(x)$ mit $\alpha > 0$ klein genug.

Fall 2: x aktiv, d.h. $c_1(x) = 0$
und es muss gelten $\nabla f(x)^T s < 0$ und $\nabla c_1(x)^T s \geq 0$.

Zulässige Richtungen, die beides erfüllen, liegen in dem schraffierten Kegel. Dieser ist nur dann leer, falls gilt

$$\nabla f(x) = \lambda_1 \nabla c_1(x) \quad (5.6)$$

mit positivem λ_1 , d.h. $\lambda_1 \geq 0$.

Mit der Lagrange-Funktion gilt wieder

$$\nabla_x \mathcal{L}(\tilde{x}, \tilde{\lambda}) = 0 \quad \text{und} \quad \tilde{\lambda} \geq 0$$

und außerdem muss gelten

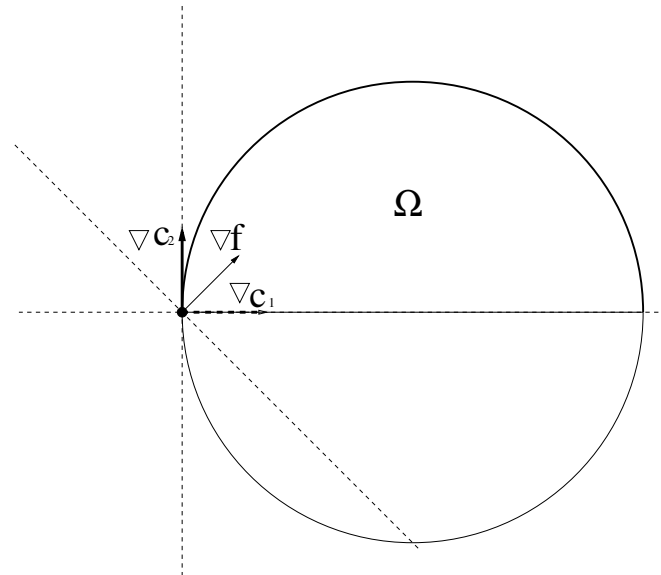
$$\tilde{\lambda}_1 c_1(\tilde{x}) = 0 \quad (\text{Komplementaritätsbedingung}). \quad (5.7)$$

Im Fall 1 galt $c_1(\tilde{x}) > 0$, d.h. hier muss $\tilde{\lambda}_1 = 0$ sein, und wir haben $\nabla f(\tilde{x}) = 0$.

Im Fall 2 kann $\lambda_2 \geq 0$ sein, d.h. wir erhalten (5.6).

Beispiel 5.1.3 (2 Ungleichungsnebenbedingungen)

$$\min x_1 + x_2 \quad \text{bei} \quad 2 - x_1^2 - x_2^2 \geq 0, \quad x_2 \geq 0$$



d.h. $I = \{1, 2\}$

$$\nabla c_1 = -2x \quad \nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \nabla c_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Man sieht leicht, dass $\tilde{x} = (-\sqrt{2}, 0)^T$ die Lösung ist. Als Bedingung für eine Abstiegsrichtung muss in 1. Näherung wieder gelten

$$\nabla c_i(x)^T d \geq 0, \quad i \in I \cap \mathcal{A}(\tilde{x}) \quad \text{und} \quad \nabla f(x)^T d < 0. \quad (5.8)$$

Wie man an der Skizze erkennen kann, ist in \tilde{x} keine Richtung d mit (5.8) zu finden.

Lagrange-Funktion: $\mathcal{L}(x, \lambda_1, \lambda_2) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x)$.

Analog zu (5.7) fordern wir als notwendige Optimalitätsbedingung

$$\nabla_x \mathcal{L}(\tilde{x}, \tilde{\lambda}) = 0 \quad \text{für ein} \quad \tilde{\lambda} \geq 0. \quad (5.9)$$

Dabei ist $\tilde{\lambda} \geq 0$ komponentenweise zu verstehen, d.h. $\tilde{\lambda} \geq 0 \Leftrightarrow \tilde{\lambda}_i \geq 0, i = 1, 2$. Die Komplementaritätsbedingung lautet

$$\tilde{\lambda}_1 c_1(\tilde{x}) = 0 \quad \text{und} \quad \tilde{\lambda}_2 c_2(\tilde{x}) = 0. \quad (5.10)$$

In $\tilde{x} = \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix}$ gilt $\nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\nabla c_1(\tilde{x}) = \begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}$, $\nabla c_2(\tilde{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, d.h. mit $\tilde{\lambda} = \begin{pmatrix} \frac{1}{2\sqrt{2}} \\ 1 \end{pmatrix}$ folgt (5.9).

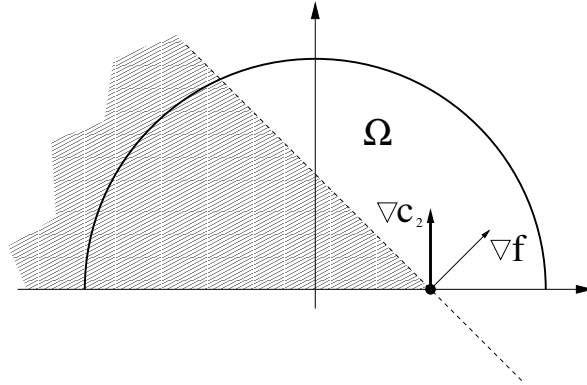
Jetzt betrachten wir $x = \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix}$. Hier sind beide Restriktionen aktiv, und es gilt

$$\nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \nabla c_1(x) = \begin{pmatrix} -2\sqrt{2} \\ 0 \end{pmatrix} \quad \nabla c_2(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

d.h. $\nabla f = -\frac{1}{2\sqrt{2}} \nabla c_1(x) + 1 \cdot \nabla c_2(x)$, somit ist $\lambda_1 < 0$, d.h. Bedingung 2 in (5.9) ist nicht erfüllt.

In $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ ist nur c_2 aktiv. Bedingung für Abstieg ist

$$\nabla c_2^T d \geq 0 \quad \text{und} \quad \nabla f(x)^T d < 0.$$



Dies wird für alle Richtungen im schraffierten Kegel erfüllt, z.B. für $d = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

Für (5.9) müsste gelten

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \stackrel{!}{=} \lambda_1 \begin{pmatrix} -2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (5.11)$$

Wir wissen, dass c_1 inaktiv ist, d.h. $c_1(x) > 0$, also folgt aus (5.10) $\lambda_1 = 0$. Also existiert kein λ_2 , so dass (5.11) erfüllt ist.

5.2 Tangentialkegel und Regularitätsbedingung für die Restriktionen

Bezeichnung: Sei $x \in \Omega$, d.h. zulässig, dann heißt die Folge $\{z_k\}$ zulässige Approximation von x , falls $\lim_{k \rightarrow \infty} z_k = x$ und $z_k \in \Omega$ für k groß genug.

Definition 5.2.1 $d \in \mathbb{R}^n$ heißt Tangente an $x \in \Omega$, falls eine Approximation $\{z_k\}$ von x und eine Folge $\{t_k\} \subset \mathbb{R}^+$ mit $t_k \rightarrow 0$ existiert, so dass

$$\lim_{k \rightarrow \infty} \frac{z_k - x}{t_k} = d.$$

Die Menge $T_\Omega(x) = \{d \in \mathbb{R}^n \mid d \text{ ist Tangente an } \Omega \text{ in } x\}$ heißt Tangentialkegel.

Bemerkungen:

1. $K \subset \mathbb{R}^n$ heißt Kegel, falls für $x \in K$ auch $\alpha x \in K$ für alle $\alpha > 0$.
2. Sei $d \in T_\Omega(x)$. Für $\alpha > 0$ definiere $\tilde{t}_k = \frac{t_k}{\alpha}$, dann gilt $\tilde{t}_k \rightarrow 0$ und

$$\frac{z_k - x}{\tilde{t}_k} = \frac{\alpha z_k - \alpha x}{t_k} \rightarrow \alpha d, \text{ d.h. } \alpha d \in T_\Omega(x),$$

und $T_\Omega(x)$ ist ein Kegel.

Mit Hilfe des Tangentialkegels lässt sich folgende fundamentale notwendige Optimalitätsbedingung beweisen:

Satz 5.2.1 Sei $\tilde{x} \in \Omega$ eine Lösung von (P), dann gilt

$$\nabla f(\tilde{x})^T d \geq 0 \quad \forall d \in T_\Omega(\tilde{x}).$$

BEWEIS. Sei $d \in T_\Omega(\tilde{x})$ dann existiert z^k mit $z_k \in \Omega$ für k groß genug und $z_k \rightarrow \tilde{x}$ für $k \rightarrow \infty$ sowie

$$d = \lim_{k \rightarrow \infty} \frac{z_k - \tilde{x}}{t_k}, \text{ d.h. } z_k = \tilde{x} + t_k d + o(\|t_k^d\|).$$

Also hat man wegen $t_k > 0$ für $k \in \mathbb{N}$

$$0 \leq \frac{1}{t_k} (f(z^k) - f(\tilde{x})) = \frac{1}{t_k} \nabla f(\tilde{x})^T t_k d + \frac{1}{t_k} o(t_k)$$

für $k \rightarrow \infty$ folgt die Behauptung. □

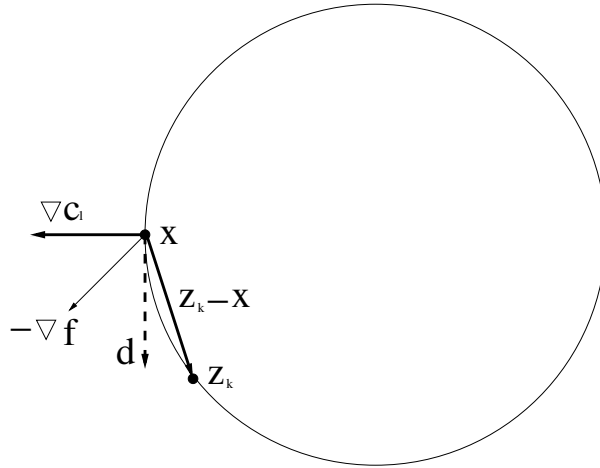
Für die Berechnung von Abstiegsrichtungen ist es wichtig, die Linearisierung der Nebenbedingungen zu untersuchen. Dazu definieren wir:

Definition 5.2.2 Sei $x \in \Omega$ und $\mathcal{A}(x)$ die Menge der aktiven Restriktionen, dann heißt

$$L_\Omega(x) = \left\{ d \in \mathbb{R}^n \mid d^T \nabla c_i(x) = 0, \forall i \in \mathcal{E} \quad \text{und} \quad d^T \nabla c_i(x) \geq 0 \forall i \in I \cap \mathcal{A}(x) \right\}$$

der Linearisierungskegel von Ω in x .

Bemerkung: $T_\Omega(x)$ ist unabhängig von der konkreten algebraischen Spezifikation von Ω , im Gegensatz zu $L_\Omega(x)$.



Beispiel 5.2.1 Wir betrachten wieder $\min x_1 + x_2$ bei $x_1^2 + x_2^2 - 2 = 0$.

Betrachte $x = \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix}$

Wähle $z_k = \begin{pmatrix} -\sqrt{2 - \frac{1}{k^2}} \\ -\frac{1}{k} \end{pmatrix}$ und $t_k = \|z_k - x\|$

dann gilt: $\|z_k\|^2 = 2 - \frac{1}{k^2} + \frac{1}{k^2} = 2$ d.h. $z_k \in \Omega \quad \forall k \in \mathbb{N}$. Wegen $z_k \rightarrow x$ ist z_k eine zulässige Approximation von x , und man erkennt aus der Skizze

$$\frac{z_k - x}{\|z_k - x\|} \rightarrow d \quad \text{mit } d = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Außerdem gilt $f(z_{k+1}) > f(z_k) \quad \forall k \geq 2$, d.h. x kann keine Lösung sein.

Andere Möglichkeit: approximiere x „von oben“ durch $\tilde{z}_k = \begin{pmatrix} -\sqrt{2 - \frac{1}{k^2}} \\ 1/k \end{pmatrix}$, wieder ist $\|\tilde{z}_k\|^2 = 2 \quad \forall k \in \mathbb{N}$.

Man kann zeigen, dass f abnimmt entlang $\{z_k\}$ und zulässige Richtungen die Form haben $d = \begin{pmatrix} 0 \\ \alpha \end{pmatrix}$, $\alpha \geq 0$. Insgesamt gilt $T_\Omega(x) = \left\{ \begin{pmatrix} 0 \\ d_2 \end{pmatrix} \mid d_2 \in \mathbb{R} \right\}$. Für den Linearisierungskegel gilt

$$d \in L_\Omega(x) \quad \text{falls} \quad 0 = \nabla c_1(x)^T d = 2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = -2\sqrt{2}d_1 \quad \text{d.h.} \quad L_\Omega(x) = T_\Omega(x).$$

Andererseits können wir Ω auch anders beschreiben, beispielsweise durch

$$\Omega = \{x \mid c_1(x) = 0\} \quad \text{mit} \quad c_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0.$$

Dann ist $d \in L_\Omega(x)$ genau dann, wenn

$$0 = \nabla c_1(x)^T d = \begin{pmatrix} 4(x_1^2 + x_2^2 - 2)x_1 \\ 4(x_1^2 + x_2^2 - 2)x_2 \end{pmatrix}^T \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}^T \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}.$$

Das gilt für $d \in \mathbb{R}^2$ beliebig, d.h. wir haben $L_\Omega(x) = \mathbb{R}^2$.

Beispiel 5.2.2 Betrachte erneut Beispiel 5.1.2, d.h.

$$\min x_1 + x_2 \quad \text{bei} \quad 2 - x_1^2 - x_2^2 \geq 0.$$

Für $x = \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix}$ sind alle zulässigen Approximationen aus Beispiel 5.2.1 auch zulässig, darüber hinaus gibt es unendlich viele weitere, z.B.

$$z_k = \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix} + \frac{1}{k} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \quad \text{mit} \quad w_1 > 0$$

für $k \geq (w_1^2 + w_2^2)/(2\sqrt{2}w_1)$ ist $z_k \in \Omega$, d.h. für $t_k = \frac{1}{k}$ ist also $d = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ mit $w_1 > 0$ Tangentialrichtung. Insgesamt erhält man

$$T_\Omega(x) = \left\{ \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \in \mathbb{R}^2 \mid d_1 \geq 0 \right\}.$$

Für die Richtungen im Linearisierungskegel gilt

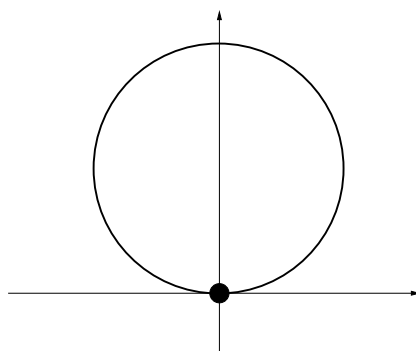
$$d \in L_\Omega(x) \Leftrightarrow 0 \leq \nabla c_1(x)^T d = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}^T \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = 2\sqrt{2}d_1$$

d.h. $d_2 \in \mathbb{R}$ beliebig und $d_1 \geq 0$, und es gilt $T_\Omega(x) = L_\Omega(x)$.

Beispiel 5.2.3 Betrachte zulässige Menge gegeben durch

$$\begin{aligned} c_1(x) &= 1 - x_1^2 - (x_2 - 1)^2 \geq 0 \\ c_2(x) &= -x_2 \geq 0 \end{aligned}$$

d.h. hier gilt $\Omega = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$.



Für jede approximierende Folge von $x = 0$ muss also gelten $z_k = 0$ für k groß genug, also ist $T_\Omega(x) = \{0\}$. Für Richtungen im Linearisierungskegel gilt $d \in L_\Omega(x)$, falls

$$\begin{aligned} \nabla c_1(x)^T d &\geq 0 \quad \text{und} \quad \nabla c_2(x)^T d \geq 0 \quad \text{d.h.} \\ -2x^T d &= -2 \cdot 0 \cdot d \geq 0 \quad \text{und} \quad \begin{pmatrix} 0 \\ -1 \end{pmatrix}^T d = -d_2 \geq 0 \end{aligned}$$

also folgt $L_\Omega(0) = \{d \in \mathbb{R}^n \mid d_2 \leq 0\} \neq T_\Omega(x)$. □

Zur Herleitung notwendiger Optimalitätsbedingungen müssen die Richtungen aus dem Tangentialkegel $T_\Omega(x)$ mit denen des Linearisierungskegels in Beziehung gesetzt werden, genauer möchte man erreichen $T_\Omega(x) = L_\Omega(x)$ (vgl. Lemma 5.4.1).

Dazu bedarf es einer Regularitätsbedingung. Für numerische Verfahren wird folgende Bedingung am häufigsten genutzt:

Definition 5.2.3 (LICQ)

Sei $x \in \Omega$. Die Regularitätsbedingung (LICQ) (= linear independence constraint qualification) ist erfüllt, falls $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ linear unabhängig ist.

5.3 Notwendige Optimalitätsbedingungen erster Ordnung

Wie in den Beispielen beginnen wir mit der Lagrange-Funktion, die in allgemeiner Form so aussieht:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup I} \lambda_i c_i(x).$$

Damit lassen sich die notwendigen Optimalitätsbedingungen wie folgt formulieren:

Satz 5.3.1 Sei \tilde{x} eine Lösung von (P), die Funktionen f und c_i seien stetig differenzierbar, und es gelte (LICQ) in \tilde{x} . Dann existiert ein Vektor Lagrange'scher Multiplikatoren $\tilde{\lambda}$ mit Komponenten $\tilde{\lambda}_i, i \in \cup I$, so dass gilt:

$$\nabla_x \mathcal{L}(\tilde{x}, \tilde{\lambda}) = 0 \tag{1}$$

$$c_i(\tilde{x}) = 0 \quad \text{für alle } i \in \mathcal{E} \tag{2}$$

$$c_i(\tilde{x}) \geq 0 \quad \text{für alle } i \in I \tag{3}$$

$$\tilde{\lambda}_i \geq 0 \quad \text{für alle } i \in I \tag{4}$$

$$\tilde{\lambda}_i c_i(\tilde{x}) = 0 \quad \text{für alle } i \in \mathcal{E} \cup I. \tag{5}$$

Bemerkungen:

1. Die Bedingungen (1)–(5) werden auch als Karush-Kuhn-Tucker-Bedingungen (KKT) bezeichnet.
2. (5) heißen Komplementaritätsbedingungen. Wegen (5) folgt $\tilde{\lambda}_i = 0$ falls c_i für $i \in I$ inaktiv, d.h. (1) in Satz 5.3.1 lässt sich schreiben als

$$0 = \nabla f(\tilde{x}) - \sum_{i \in \mathcal{A}(\tilde{x})} \tilde{\lambda}_i \nabla c_i(\tilde{x}).$$

Für numerische Verfahren wird oft eine stärkere Komplementaritätsbedingung gefordert:

Definition 5.3.1 Sei \tilde{x} Lösung von (P) und $\tilde{\lambda}$ erfülle (1)–(5). Dann genügt $\tilde{\lambda}$ der strikten Komplementaritätsbedingung, falls für alle $i \in I$, zusätzlich gilt

$$c_i(x) = 0 \Rightarrow \lambda_i > 0.$$

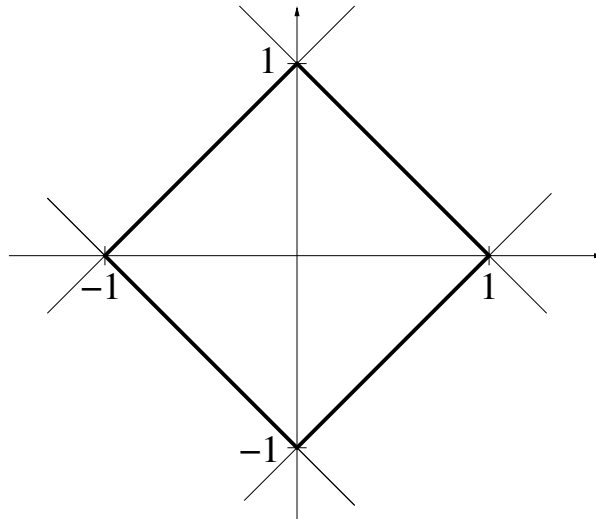
Mit anderen Worten, $\tilde{\lambda}_i > 0$ für alle $i \in I \cap \mathcal{A}(\tilde{x})$.

Beispiel 5.3.1

$$\min \left(x_1 - \frac{3}{2} \right)^2 + \left(x_2 - \frac{1}{2} \right)^4$$

bei

$$\begin{aligned} 1 - x_1 - x_2 &\geq 0 \\ 1 - x_1 + x_2 &\geq 0 \\ 1 + x_1 - x_2 &\geq 0 \\ 1 + x_1 + x_2 &\geq 0. \end{aligned}$$



Aus der Skizze wird klar, dass $\tilde{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. In \tilde{x} sind c_1 und c_2 aktiv. Es gilt

$$\nabla f = \begin{pmatrix} 2\left(x_1 - \frac{3}{2}\right) \\ 4\left(x_2 - \frac{1}{2}\right)^3 \end{pmatrix}, \quad \nabla f(\tilde{x}) = \begin{pmatrix} -1 \\ -\frac{1}{2} \end{pmatrix}, \quad \nabla c_1(\tilde{x}) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \nabla c_2(\tilde{x}) = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

also können wir schreiben

$$\nabla f(\tilde{x}) = \frac{3}{4} \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Setzt man $\tilde{\lambda} = \left(\frac{3}{4}, \frac{1}{4}, 0, 0 \right)$ sind alle KKT-Bedingungen erfüllt.

Beispiel 5.3.2

$$\begin{aligned} \min f(x) = |x|^2 \quad \text{bei} \quad & x_1 + x_2 + x_3 = 3 \\ & 2x_1 - x_2 + x_3 \leq 5 \end{aligned}$$

d.h.

$$\begin{aligned}
c_1 &= x_1 + x_2 + x_3 - 3 = 0 \\
c_2 &= 5 - 2x_1 + x_2 - x_3 \geq 0 \\
\mathcal{L}(x, \lambda) &= x_1^2 + x_2^2 + x_3^2 - \lambda_1(x_1 + x_2 + x_3 - 3) \\
&\quad - \lambda_2(5 - 2x_1 + x_2 - x_3) \\
\nabla_x \mathcal{L}(x, \lambda) &\stackrel{!}{=} \Leftrightarrow \left. \begin{aligned} 2x_1 - \lambda_1 + 2\lambda_2 &= 0 \\ 2x_2 - \lambda_1 - \lambda_2 &= 0 \\ 2x_3 - \lambda_1 + \lambda_2 &= 0 \end{aligned} \right\} \quad (5.12)
\end{aligned}$$

Außerdem muss gelten $\lambda_2(5 - 2x_1 + x_2 - x_3) = 0$. Angenommen $\lambda_2 \neq 0 \Rightarrow$

$$\begin{aligned}
x_1 &= \frac{1}{2} \lambda_1 - \lambda_2 \\
x_2 &= \frac{1}{2} \lambda_1 + \frac{1}{2} \lambda_2 \\
x_3 &= \frac{1}{2} \lambda_1 - \frac{1}{2} \lambda_2
\end{aligned}$$

Einsetzen in aktive Nebenbedingungen:

$$\begin{aligned}
5 &= \lambda_1 - 2\lambda_2 - \frac{1}{2} \lambda_2 - \frac{1}{2} \lambda_2 \\
&= \lambda_1 - 3\lambda_2 \Rightarrow \lambda_1 = 5 + 3\lambda_2 \\
3 &= \frac{1}{2} \lambda_1 - \lambda_2 + \frac{1}{2} \lambda_1 + \frac{1}{2} \lambda_2 + \frac{1}{2} \lambda_1 - \frac{1}{2} \lambda_2 \\
&= \frac{3}{2} \lambda_1 - \lambda_2 \\
\Rightarrow 3 &= \frac{15}{2} + \frac{9}{2} \lambda_2 - \lambda_2 = \frac{15}{2} + \frac{7}{2} \lambda_2 \\
\Rightarrow \lambda_2 &= -\frac{9 \cdot 2}{2 \cdot 7} = -\frac{9}{7} < 0.
\end{aligned}$$

Widerspruch, da $\lambda_2 \geq 0$ sein muss. Also muss gelten $\lambda_2 = 0$, dann folgt mit (5.12) $x_1 = x_2 = x_3 = \frac{\lambda_1}{2}$ und wegen $c_1 = 0$ folgt $x_1 = x_2 = x_3 = 1$

$$\lambda_1 = 2 \quad \text{d.h.} \quad \tilde{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{erfüllt die KKT-Bedingungen.}$$

5.4 Beweis von Satz 5.3.1

Der Beweis verläuft in drei Schritten. Zuerst bringen wir den Tangential- und den Linearisierungskegel zusammen. Dazu führen wir folgende Notation ein:

$$A(\tilde{x})^T = [\nabla c_i(\tilde{x})]_{i \in \mathcal{A}(\tilde{x})}$$

d.h. $A(\tilde{x})$ ist die Matrix, in der die Gradienten der aktiven Nebenbedingungen als Zeilen stehen.

Lemma 5.4.1 *Sei \tilde{x} zulässig, dann gilt:*

- (1) $T_\Omega(\tilde{x}) \subset L_\Omega(\tilde{x})$
- (2) Falls (LICQ) in \tilde{x} erfüllt ist, gilt sogar

$$T_\Omega(\tilde{x}) = L_\Omega(\tilde{x}).$$

BEWEIS. Ohne Beschränkung der Allgemeinheit nehmen wir an, dass $c_i(x)$, $i = 1, \dots, m$ die aktiven Nebenbedingungen sind.

Beweis von (1):

Sei $d \in T_\Omega(\tilde{x})$, dann existieren $\{z_k\}$ und $\{t_k\}$, so dass

$$\lim_{k \rightarrow \infty} \frac{z_k - \tilde{x}}{t_k} = d,$$

insbesondere gilt $t_k > 0$ für alle k . Sei $i \in \mathcal{E}$, dann gilt mit Taylorformel 1. Ordnung mit Restglied mit einem $\theta \in (0, 1)$:

$$0 = \frac{1}{t_k} c_i(z_k) = \frac{1}{t_k} c_i(\tilde{x} + (z_k - \tilde{x})) = \nabla c_i(\tilde{x} + \theta(z_k - \tilde{x})) \frac{z_k - \tilde{x}}{t_k}$$

d.h. für $k \rightarrow \infty$ folgt $\nabla c_i(\tilde{x})^T d = 0$.

Für $i \in \mathcal{A}(\tilde{x}) \cap I$ erhalten wir auf dieselbe Art und Weise

$$0 \leq \nabla c_i(\tilde{x})^T d.$$

Insgesamt können wir folgern $d \in L_\Omega(\tilde{x})$.

Zu (2): (LICQ) ist erfüllt, also hat $A \in \mathbb{R}^{(m,n)}$ maximalen Zeilenrang m .

Für den Kern von A gilt $\dim \text{Kern } A = n - \text{rang } A$.

Sei $Z \in \mathbb{R}^{(n, n-m)}$ Nullraummatrix, d.h. Spalten von Z spannen den Kern von A auf:

$$A(\tilde{x})Z = 0$$

und $\text{rang } Z = n - m$.

Sei jetzt $d \in L_\Omega(\tilde{x})$ beliebig gewählt und $\{t_k\} \subset \mathbb{R}^+ \setminus \{0\}$ mit $t_k \rightarrow 0$, dann definieren wir das Gleichungssystem

$$R : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n \quad \text{durch} \\ R(z, t) = \begin{pmatrix} c(z) - tA(\tilde{x})d \\ Z^T(z - \tilde{x} - td) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Wir werden mit dem Satz über implizite Funktionen zeigen, dass eine Folge $\{z_k\}$ existiert, so dass $\{z_k\}$ eine zulässige Approximation von \tilde{x} ist, $R(z_k, t_k) = 0$ und $\frac{z_k - \tilde{x}}{t_k} \rightarrow d$, d.h. $d \in T_\Omega(\tilde{x})$.

Es gilt

$$R(\tilde{x}, 0) = \begin{pmatrix} c(\tilde{x}) - 0 \cdot A(\tilde{x})d \\ Z^T(\tilde{x} - \tilde{x} - 0 \cdot d) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

R ist stetig differenzierbar in Umgebung von $(\tilde{x}, 0)$ und

$$[D_z R]_{i\text{-te Zeile}} = \begin{cases} \nabla c_i^T, & i \in \{1, \dots, m\} \\ [Z^T]_{i\text{-te Zeile}}, & i \in \{m+1, \dots, n\} \end{cases}$$

d.h. $D_z R(\tilde{x}, 0) = \begin{pmatrix} A \\ Z^T \end{pmatrix}$.

Es gilt $\text{rang} \begin{pmatrix} A \\ Z^T \end{pmatrix} = n$, also existiert nach dem Satz über implizite Funktionen eine Funktion $z(t)$, so dass $z(0) = \tilde{x}$ und $R(z(t), t) = 0$ für t klein, d.h. $z_k = z(t_k) \rightarrow \tilde{x}$ und $R(z_k, t_k) = 0$ für k groß genug.

Also gilt $c(z_k) = t_k A(\tilde{x})d$ und weil $d \in L_\Omega(\tilde{x})$ gilt für $i \in \mathcal{E}$:

$$c_i(z_k) = t_k \nabla c_i(\tilde{x})^T d = 0$$

und für $i \in \mathcal{A}(\tilde{x}) \cap I$

$$c_i(z_k) = t_k \nabla c_i(\tilde{x})^T d \geq 0$$

also ist z_k zulässig für k groß genug. Mit Taylorformel mit Restglied 1. Ordnung und einem $\theta \in (0, 1)$ folgt weiterhin für große k

$$\begin{aligned} 0 &= \frac{1}{t_k} R(z_k, t_k) = \frac{1}{t_k} \begin{pmatrix} c(\tilde{x} + (z_k - \tilde{x})) - t_k A(\tilde{x})d \\ Z^T(z_k - \tilde{x} - t_k d) \end{pmatrix} \\ &= \begin{pmatrix} A(\tilde{x} + \theta(z_k - \tilde{x})) \\ Z^T \end{pmatrix} \frac{z_k - \tilde{x}}{t_k} - \begin{pmatrix} A(\tilde{x})d \\ Z^T \end{pmatrix} d \end{aligned}$$

Da $\begin{pmatrix} A(\tilde{x}) \\ Z^T \end{pmatrix}$ regulär, folgt hieraus

$$\frac{z_k - \tilde{x}}{t_k} \rightarrow d$$

d.h. $d \in T_\Omega(\tilde{x})$.

Wesentlicher Schritt ist das folgende Lemma von Farkas:

Lemma 5.4.2 Sei $K = \{By + Cw \mid y \geq 0\}$ mit $B \in \mathbb{R}^{(n,m)}$, $C \in \mathbb{R}^{(n,p)}$, $w \in \mathbb{R}^p$ fest und $y \in \mathbb{R}^m$. Dann gilt für jeden Vektor $g \in \mathbb{R}^n$ entweder $g \in K$, oder es existiert ein $d \in \mathbb{R}^n$, so dass gilt

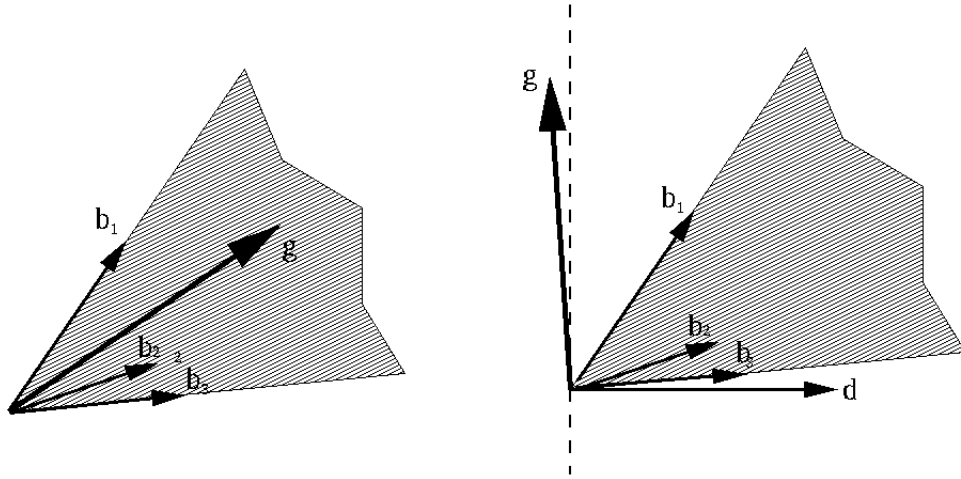
$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0.$$

Der Beweis des Lemmas findet sich z.B. in [2]. Grafisch ist die Aussage leicht einzusehen, zumindest im hier skizzierten Fall für $n = 2$, $m = 3$, $C = 0$:

Jetzt wenden wir das Lemma von Farkas auf die Situation in Satz 5.3.1 an.

Sei $N = \left\{ \sum_{i \in \mathcal{A}(\tilde{x})} \lambda_i \nabla c_i, \text{ und } \lambda_i \geq 0 \text{ für } i \in \mathcal{A}(\tilde{x}) \cap I \right\}$ und $g = \nabla f(\tilde{x})$, dann gilt entweder

$$\begin{aligned} \nabla f(\tilde{x}) &= \sum_{i \in \mathcal{A}(\tilde{x})} \lambda_i \nabla c_i(\tilde{x}) = A(\tilde{x})^T \lambda, \\ \text{mit } \lambda_i &\geq 0 \text{ für } i \in \mathcal{A}(\tilde{x}) \cap I \end{aligned} \tag{I}$$



oder es gibt ein $d \in \mathbb{R}^n$ mit

$$\begin{aligned} \nabla f(\tilde{x})^T d &< 0, \\ \nabla c_i^T d &= 0, \quad i \in \mathcal{E} \quad \text{und} \quad \nabla c_i^T d \geq 0, \quad i \in \mathcal{A}(\tilde{x}) \cap I \end{aligned}$$

d.h.

$$\nabla f(\tilde{x})^T d < 0 \quad \text{und} \quad d \in L_\Omega(\tilde{x}). \quad (\text{II})$$

Nach Voraussetzungen ist $\tilde{x} \in \Omega$ und (LICQ) erfüllt,

Satz 5.2.1 $\Rightarrow \nabla f(\tilde{x})^T d \geq 0 \quad \forall d \in T_\Omega(\tilde{x}) = L_\Omega(\tilde{x})$ also gilt (II) nicht, und es muss (I) gelten.

Definiere

$$\tilde{\lambda}_i = \begin{cases} \lambda_i, & i \in \mathcal{A}(\tilde{x}) \\ 0, & \text{sonst,} \end{cases}$$

dann gilt $\tilde{\lambda} \geq 0$ und die Komplementaritätsbedingung ist erfüllt.

5.5 Optimalitätsbedingungen 2. Ordnung

Im Folgenden seien f und c_i , $i \in I \cup \mathcal{E}$ 2-mal stetig differenzierbar.

Sei \tilde{x} ein kritischer Punkt und $\tilde{\lambda}$ so, dass die KKT-Bedingungen erfüllt sind, dann definieren wir den kritischen Kegel

$$C(\tilde{x}, \tilde{\lambda}) = \{\omega \in L_\Omega(\tilde{x}) \mid \nabla c_i(\tilde{x})^T \omega = 0 \quad \forall i \in \mathcal{A}(\tilde{x}) \cap I \quad \text{mit} \quad \tilde{\lambda}_i > 0\}$$

d.h. es gilt

$$\omega \in C(\tilde{x}, \tilde{\lambda}) \Leftrightarrow \begin{cases} \nabla c_i(\tilde{x})^T \omega = 0 & \forall i \in \mathcal{E} \\ \nabla c_i(\tilde{x})^T \omega = 0 & \forall i \in \mathcal{A}(\tilde{x}) \cap I \quad \text{mit} \quad \tilde{\lambda}_i > 0 \\ \nabla c_i(\tilde{x})^T \omega \geq 0 & \forall i \in \mathcal{A}(\tilde{x}) \cap I \quad \text{mit} \quad \tilde{\lambda}_i = 0 \end{cases}$$

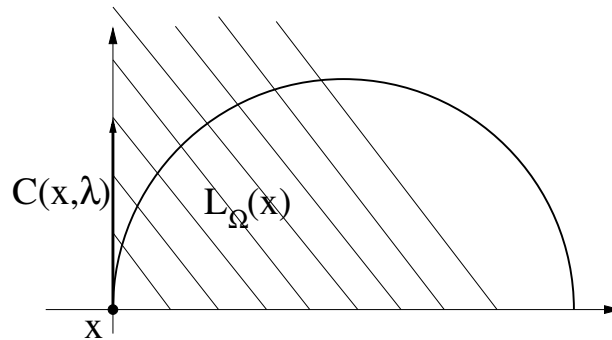
$\tilde{\lambda}_i = 0$ für alle inaktiven Nebenbedingungen, also folgt für $\omega \in C(\tilde{x}, \tilde{\lambda})$

$$\lambda_i \nabla c_i(\tilde{x})^T \omega = 0 \quad \forall i \in \mathcal{E} \cup I,$$

und mit der ersten KKT-Bedingung ergibt sich

$$\omega^T \nabla f(\tilde{x}) = \sum_{i \in \mathcal{E} \cup I} \tilde{\lambda}_i \omega^T \nabla c_i(\tilde{x}) = 0,$$

d.h. der kritische Kegel $C(\tilde{x}, \tilde{\lambda})$ enthält Richtungen des Linearisierungskegels, für die man auf der Basis der 1. Ableitung noch nicht sagen kann, ob f zu- oder abnimmt entlang dieser Richtung.



Beispiel 5.5.1 Betrachte

$$\begin{aligned} \min x_1 \quad \text{unter} \quad x_2 &\geq 0 \\ 1 - (x_1 - 1)^2 - x_2^2 &\geq 0. \end{aligned}$$

Lösung ist $\tilde{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

$$\mathcal{A}(\tilde{x}) = \{1, 2\} \quad \nabla c_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \nabla c_2(x) = \begin{pmatrix} -2(x_1 - 1) \\ -2x_2 \end{pmatrix}$$

$$\nabla c_2(0) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$\nabla f = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0 \cdot \nabla c_1 + \frac{1}{2} \nabla c_2$$

d.h. $\tilde{\lambda} = \left(0, \frac{1}{2}\right)^T$.

(LICQ) ist erfüllt, d.h. $\tilde{\lambda}$ ist eindeutig bestimmt.

$$\begin{aligned} L_\Omega(\tilde{x}) &= \{d \in \mathbb{R}^2 \mid \nabla c_1^T d \geq 0 \quad \text{und} \quad \nabla c_2^T d \geq 0\} \\ &= \{d \in \mathbb{R}^2 \mid d_1 \geq 0 \quad \text{und} \quad d_2 \geq 0\} \\ C(\tilde{x}, \tilde{\lambda}) &= \{w \mid \nabla c_1^T w \geq 0 \quad \text{und} \quad \nabla c_2^T w = 0\} \\ &= \{w \mid w_1 = 0 \quad \text{und} \quad w_2 \geq 0\}. \end{aligned}$$

Der folgende Satz zeigt, dass die Hessematrix der Lagrange-Funktion eine nichtnegative Krümmung entlang kritischer Richtungen hat.

Satz 5.5.1 (Notwendige Bedingung 2. Ordnung)

Sei \tilde{x} eine lokale Lösung von (P), es gelte (LICQ) und $\tilde{\lambda}$ sei der Lagrangsche Multiplikator, so dass die KKT-Bedingungen erfüllt sind. Dann gilt

$$\omega^T \nabla_{xx}^2 \mathcal{L}(\tilde{x}, \tilde{\lambda}) \omega \geq 0 \quad \forall \omega \in C(\tilde{x}, \tilde{\lambda}).$$

BEWEIS. $\omega \in C(\tilde{x}, \tilde{\lambda}) \subset L_\Omega(\tilde{x})$, d.h. wir können wie im Beweis zu Lemma 5.4.1 eine zulässige Approximation $\{z_k\}$ von \tilde{x} und $\{t_k\}$ konstruieren, so dass gilt

$$\lim_{k \rightarrow \infty} \frac{z_k - \tilde{x}}{t_k} = \omega$$

bzw.

$$z_k - \tilde{x} = t_k \omega + o(t_k).$$

Außerdem wurde gezeigt

$$c_i(z_k) = t_k \nabla c_i(\tilde{x})^T \omega \quad \forall i \in \mathcal{A}(\tilde{x}).$$

Damit erhalten wir

$$\mathcal{L}(z_k, \tilde{\lambda}) = f(z_k) - \sum_{i \in \mathcal{E} \cup I} \tilde{\lambda}_i c_i(z_k) = f(z_k) - t_k \sum_{i \in \mathcal{E} \cup I} \lambda_i \nabla c_i(\tilde{x})^T \omega = f(z_k),$$

Taylor-Entwicklung von $\mathcal{L}(z_k, \tilde{\lambda})$ um \tilde{x} mit Restglied 2. Ordnung und $\theta \in (0, 1)$ liefert:

$$\begin{aligned} f(z_k) &= \mathcal{L}(z_k, \tilde{\lambda}) = \mathcal{L}(\tilde{x}, \tilde{\lambda}) + (z_k - \tilde{x})^T \nabla_x \mathcal{L}(\tilde{x}, \tilde{\lambda}) \\ &\quad + \frac{1}{2} (z_k - \tilde{x})^T \nabla_{xx}^2 \mathcal{L}(\tilde{x} + \theta(\tilde{x} - z_k), \tilde{\lambda}) (z_k - \tilde{x}) \\ &\Rightarrow 0 \leq \frac{1}{t_k^2} (f(z_k) - f(\tilde{x})) = \frac{1}{2} \left(\frac{z_k - \tilde{x}}{t_k} \right)^T \nabla_{xx}^2 \mathcal{L}(\tilde{x} + \theta(\tilde{x} - z_k), \tilde{\lambda}) \frac{z_k - \tilde{x}}{t_k}. \end{aligned}$$

Mit Grenzübergang $k \rightarrow \infty$ folgt die Behauptung.

Satz 5.5.2 Sei $\tilde{x} \in \mathbb{R}^n$ ein zulässiger Punkt, für den ein Lagrange-Multiplikator $\tilde{\lambda}$ existiert, so dass die KKT-Bedingungen erfüllt sind. Außerdem existierte ein $\sigma > 0$, so dass

$$w^T \nabla_{xx}^2 \mathcal{L}(\tilde{x}, \tilde{\lambda}) w \geq \sigma \|w\|^2 \quad \forall w \in C(\tilde{x}, \tilde{\lambda}).$$

Dann ist \tilde{x} eine strikte lokale Lösung von (P), und es gibt ein $\tilde{\sigma} > 0$ so dass die Wachstumsabschätzung

$$f(z) \geq f(\tilde{x}) + \tilde{\sigma} \|z - \tilde{x}\|^2 \tag{5.13}$$

für alle $z \in B_r(\tilde{x}) \cap \Omega$ und ein $r > 0$ erfüllt ist.

BEWEIS. Sei $\{z_k\}$ eine zulässige Approximation von \tilde{x} . Wir zeigen, dass dann gilt

$$f(z_k) \geq f(\tilde{x}) + \frac{\sigma}{4} \|z^k - \tilde{x}\|^2 \quad (5.14)$$

falls k gross genug ist. Angenommen, (5.14) ist nicht erfüllt, dann gibt es eine zulässige Approximation $\{z^k\}$ von \tilde{x} mit

$$f(z^k) < f(\tilde{x}) + \frac{\sigma}{4} \|z^k - \tilde{x}\|^2 \quad \text{für } k \text{ groß genug.} \quad (5.15)$$

Für $d^k := \frac{z^k - \tilde{x}}{\|z^k - \tilde{x}\|}$ ist $\|d^k\| = 1$, d.h. zumindest für eine Teilfolge, die wieder mit k indiziert wird, können wir annehmen, dass ein $d \in \mathbb{R}^n$ existiert mit

$$d = \lim_{k \rightarrow \infty} d^k,$$

damit ist $d \in T_\Omega(\tilde{x}) \subset L_\Omega(\tilde{x})$, nach Lemma 5.4.1.

Behauptung:

$$d \in C(\tilde{x}, \tilde{\lambda}). \quad (5.16)$$

Aus den KKT-Bedingungen folgt

$$\mathcal{L}(z_k, \tilde{\lambda}) = f(z_k) - \sum_{i \in \mathcal{A}(\tilde{x})} \underbrace{\lambda_i c_i(z^k)}_{\geq 0} \leq f(z^k).$$

Angenommen, $d \notin C(\tilde{x}, \tilde{\lambda})$, dann existiert $j \in \mathcal{A}(\tilde{x}) \cap I$, so dass

$$\tilde{\lambda}_j \nabla c_j(\tilde{x})^T d > 0, \quad (5.17)$$

für die anderen Indices $i \in \mathcal{A}(\tilde{x})$ gilt nun

$$\tilde{\lambda}_i \nabla c_i(\tilde{x})^T d \geq 0.$$

Es ist

$$z^k = \tilde{x} + \|z_k - \tilde{x}\| d + o(\|z_k - \tilde{x}\|). \quad (5.18)$$

Taylorentwicklung ergibt

$$\begin{aligned} \tilde{\lambda}_j c_j(z^k) &= \underbrace{\tilde{\lambda}_j c_j(\tilde{x})}_{=0} + \lambda_j \nabla c_j(\tilde{x})^T (z^k - \tilde{x}) + o(\|z^k - \tilde{x}\|) \\ &\stackrel{(5.18)}{=} \|z^k - \tilde{x}\| \tilde{\lambda}_j \nabla c_j(\tilde{x})^T d + o(\|z^k - \tilde{x}\|) \end{aligned} \quad (5.19)$$

$$\begin{aligned} \Rightarrow \mathcal{L}(z_k, \tilde{\lambda}) &= f(z_k) - \sum_{i \in \mathcal{A}(\tilde{x})} \tilde{\lambda}_i c_i(z_k) \leq f(z_k) - \tilde{\lambda}_j c_j(z_k) \\ &= f(z_k) - \|z_k - \tilde{x}\| \tilde{\lambda}_j \nabla c_j(\tilde{x})^T d + o(\|z^k - \tilde{x}\|). \end{aligned}$$

Andererseits folgt aus der notwendigen Bedingung:

$$\mathcal{L}(z_k, \tilde{x}) = f(\tilde{x}) + o(\|z_k - \tilde{x}\|).$$

Insgesamt ergibt sich

$$f(z_k) \geq f(\tilde{x}) + \|z_k - \tilde{x}\| \underbrace{\tilde{\lambda}_j \nabla c_j(\tilde{x})^T d}_{>0} + o(\|z_k - \tilde{x}\|).$$

Wegen (5.17) existiert ein k_0 , so dass für $\tau = \tilde{\lambda}_j \nabla c_j(\tilde{x})^T d > 0$ gilt

$$\frac{f(z_k) - f(\tilde{x})}{\|z_k - \tilde{x}\|} \geq \frac{\tau}{2} > 0 \quad \text{für alle } k \geq k_0$$

im Widerspruch zu (5.15), also folgern wir, dass (5.16) gilt. Es ist also $d \in C(\tilde{x}, \tilde{\lambda})$ und $\|d\| = 1$ und daher

$$d^T \nabla_{xx}^2 \mathcal{L}(\tilde{x}, \tilde{\lambda}) d \geq \sigma.$$

Außerdem gilt

$$\mathcal{L}(z_k, \tilde{\lambda}) = f(z_k) - \sum_{i \in \mathcal{E} \cup I} \underbrace{\tilde{\lambda}_i c_i(z_k)}_{\geq 0} \leq f(z_k).$$

Taylorentwicklung von $\mathcal{L}(z^k, \tilde{\lambda})$ um \tilde{x} wie im Beweis zu Satz 5.5.1 unter Verwendung von

$$z_k - \tilde{x} = \|z_k - \tilde{x}\| d + o(\|z_k - \tilde{x}\|)$$

liefert

$$\begin{aligned} f(z_k) &\geq \mathcal{L}(z_k, \tilde{\lambda}) = f(\tilde{x}) + \underbrace{(z_k - \tilde{x})^T \nabla_x \mathcal{L}(\tilde{x}, \tilde{\lambda})}_{=0} \\ &\quad + \frac{1}{2} (z_k - \tilde{x})^T \nabla_{xx}^2 \mathcal{L}(\tilde{x}, \tilde{\lambda}) (z_k - \tilde{x}) + o(\|z_k - \tilde{x}\|^2) \\ &= f(\tilde{x}) + \frac{1}{2} \underbrace{d^T \nabla_{xx}^2 \mathcal{L}(\tilde{x}, \tilde{\lambda}) d}_{\geq \sigma} \|z_k - \tilde{x}\|^2 + o(\|z_k - \tilde{x}\|^2) \\ &= f(\tilde{x}) + \frac{\sigma}{2} \|z_k - \tilde{x}\|^2 + o(\|z_k - \tilde{x}\|^2) \\ &= f(\tilde{x}) + \frac{\sigma}{4} \|z_k - \tilde{x}\|^2 + \underbrace{\frac{\sigma}{2} \|z_k - \tilde{x}\|^2 + o(\|z_k - \tilde{x}\|^2)}_{\geq 0 \text{ für } k \text{ groß genug}}, \end{aligned}$$

das ist im Widerspruch zu (5.15), also gilt für alle zulässigen Approximationen $\{z^k\}$ von \tilde{x}

$$f(z_k) \geq f(\tilde{x}) + \frac{\sigma}{4} \|z_k - \tilde{x}\|^2.$$

□

Bemerkung: Bei strikter Komplementarität gilt für alle $i \in \mathcal{A}(\tilde{x}) \cap I$, dass $\lambda_i > 0$, d.h. dann ist

$$C(\tilde{x}, \tilde{\lambda}) = \left\{ d \in \mathbb{R}^n \mid \nabla c_i(\tilde{x})^T d = 0 \quad \forall i \in \mathcal{A}(\tilde{x}) \right\}.$$

Sei

$$A(\tilde{x}) = \left(\nabla c_i(\tilde{x})^T \right)_{i \in \mathcal{A}(\tilde{x})}$$

dann ist $C(\tilde{x}, \tilde{\lambda}) = \text{Kern}(A(\tilde{x}))$, dh. die hinreichende Optimalitätsbedingung lautet

$$w^T \nabla_{xx} \mathcal{L}(\tilde{x}, \tilde{\lambda}) w \geq \sigma \|w\|^2 \quad \forall w \in \text{Kern}(A(\tilde{x})). \quad (5.20)$$

Sei $\dim \text{Kern}(A(\tilde{x})) = l$ und $z \in \mathbb{R}^{(n,l)}$ eine Nullraummatrix, d.h. $w = Zz \quad \forall z \in \mathbb{R}^l$, dann gilt

$$w^T \nabla_{xx} \mathcal{L}(\tilde{x}, \tilde{\lambda}) w = z^T Z^T \nabla_{xx} \mathcal{L}(\tilde{x}, \tilde{\lambda}) Z z,$$

d.h. (5.20) ist äquivalent zur positiven Definitheit von $Z^T \nabla_{xx} \mathcal{L}(\tilde{x}, \tilde{\lambda}) Z$ auf \mathbb{R}^l .

Diese Bedingung ist numerisch viel einfacher zu überprüfen, daher wird oft strikte Komplementarität vorausgesetzt.

Beispiel 5.5.2 $\min -x^2 + 1$ bei $x \geq -1, x \leq \frac{1}{2}$.

Globales Minimum bei $x = -1$, lokales Minimum in $x = \frac{1}{2}$. In $\tilde{x} = \frac{1}{2}$ gilt z.B.

$$\mathcal{L}(\tilde{x}, \lambda_1, \lambda_2) = -x^2 + 1 - \lambda_1(x + 1) - \lambda_2\left(\frac{1}{2} - x\right).$$

In $\tilde{x} = \frac{1}{2}$ ist c_1 inaktiv, d.h. $\lambda_1 = 0$

$$KKT \Rightarrow \mathcal{L}_x = -2\tilde{x} + \lambda_2 \stackrel{!}{=} 0 \Rightarrow \lambda_2 = 1$$

$$\text{d.h. } C(\tilde{x}, \tilde{\lambda}) = \left\{ d \in \mathbb{R} \mid c'_2(\tilde{x}) d = 0 \right\} = \{0\}$$

d.h. hinreichende Bedingung $d\mathcal{L}_{xx}(\tilde{x}, \tilde{\lambda})d \geq \sigma d^2 \quad \forall d \in \mathbb{R}$ mit $d = 0$

d.h. die Bedingung ist erfüllt für jedes σ .

Beispiel 5.5.3

$$\begin{aligned} \min -\frac{1}{2}\sqrt{x_1} - \frac{1}{2}x_2 \quad & \text{bei} \quad x_1 \geq 0 \\ & x_2 \geq 0 \\ & 1 - x_1 - x_2 \geq 0 \end{aligned}$$

$$\begin{aligned} \mathcal{L}(x, \lambda) &= -\frac{1}{2}\sqrt{x_1} - \frac{1}{2}x_2 - \lambda_1 x_1 - \lambda_2 x_2 - \lambda_3(1 - x_1 - x_2) \\ \nabla_x \mathcal{L}(x, \lambda) &= \begin{pmatrix} -\frac{1}{4\sqrt{x_1}} - \lambda_1 + \lambda_3 \\ -\frac{1}{2} - \lambda_2 + \lambda_3 \end{pmatrix} = 0 \end{aligned}$$

KKT-System:

$$\begin{aligned} -\frac{1}{4} \frac{1}{\sqrt{x_1}} - \lambda_1 + \lambda_3 &= 0 \\ -\frac{1}{2} - \lambda_2 + \lambda_3 &= 0 \\ x_1 \geq 0, x_2 \geq 0, 1 - x_1 - x_2 &\geq 0 \\ \lambda_1 x_1 &= 0 \\ \lambda_2 x_2 &= 0 \\ \lambda_3(1 - x_1 - x_2) &= 0. \end{aligned}$$

Eine Lösung ist

$$\tilde{x} = \begin{pmatrix} \frac{1}{4} \\ \frac{3}{4} \end{pmatrix}, \quad d.h. \quad \lambda_1 = 0, \lambda_2 = 0, \lambda_3 = \frac{1}{2},$$

d.h. strikte Komplementarität ist erfüllt.

Hinreichende Bedingung:

$$\begin{aligned} \nabla_{xx}^2 \mathcal{L}(x, \lambda) &= \begin{pmatrix} \frac{1}{8} \frac{1}{x_1^{3/2}} & 0 \\ 0 & 0 \end{pmatrix} \\ \nabla_{xx}^2 \mathcal{L}(\tilde{x}, \tilde{\lambda}) &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{nicht positiv definit auf } \mathbb{R}^2 \end{aligned}$$

aber, wir haben

$$\begin{aligned} C(\tilde{x}, \tilde{\lambda}) &= \left\{ d \in \mathbb{R}^2 \mid \nabla c_3^T d = -d_1 - d_2 = 0 = d_1 + d_2 \right\} \\ &= \left\{ d \in \mathbb{R}^2 \mid d_1 = -d_2 \right\}. \end{aligned}$$

Nun ist $d_1^2 = d_2^2$, also folgt

$$d^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} d = d_1^2 = \frac{1}{2} d_1^2 + \frac{1}{2} d_2^2 = \frac{1}{2} \|d\|^2,$$

d.h. hinreichende Bedingung ist erfüllt mit $\sigma = \frac{1}{2}$.

5.6 Probleme mit Variationsbeschränkungen (box constraints)

Hier ist $\Omega = \left\{ x \in \mathbb{R}^n \mid v_i \leq x_i \leq w_i, 1 \leq i \leq n \right\}$ ein Quader in \mathbb{R}^n .

Sei $v = (v_1, \dots, v_n)^T$ und $w = (w_1, \dots, w_n)^T$. Wir setzen voraus

$v < w$ und betrachten das Problem

$$\min_{v \leq x \leq w} f(x), \quad \text{mit } f : D \supset \Omega \rightarrow \mathbb{R} \text{ hinreichend glatt.} \quad (5.21)$$

i) Umformen in Standardform:

$$v \leq x \leq w \Leftrightarrow x - v \geq 0 \quad \text{und} \quad w - x \geq 0$$

$$\Leftrightarrow \begin{pmatrix} I \\ -I \end{pmatrix} x + \begin{pmatrix} -v \\ w \end{pmatrix} \geq 0 \Leftrightarrow Gx - v \geq 0$$

wegen $v < w$ können nie gleichzeitig obere und untere Restriktionen aktiv sein, also sind die Gradienten der aktiven constraints auf jeden Fall linear unabhängig, d.h. die Lagrange-Multiplikatoren sind eindeutig bestimmt.

ii) Notwendige Optimalitätsbedingung

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^n \lambda_i^u (x_i - v_i) - \sum_{i=1}^n \lambda_i^o (-x_i + w_i)$$

$$\mathcal{L}_{x_i} = 0 \Leftrightarrow \frac{\partial f}{\partial x_i} - \lambda_i^u + \lambda_i^o = 0, \quad \lambda_i^u, \lambda_i^o \geq 0$$

λ_i^u : Multiplikator zur unteren Schranke

λ_i^o : Multiplikator zur oberen Schranke.

Außerdem wissen wir: Genau einer der Multiplikatoren darf positiv sein (muss aber nicht), einer ist immer Null.

Genauer gilt:

$$\frac{\partial f}{\partial x_i} = 0 \Rightarrow \lambda_i^u = \lambda_i^o = 0$$

$$\frac{\partial f}{\partial x_i} \geq 0 \Rightarrow \lambda_i^u = \frac{\partial f}{\partial x_i}, \lambda_i^o = 0$$

$$\frac{\partial f}{\partial x_i} \leq 0 \Rightarrow \lambda_i^o = -\frac{\partial f}{\partial x_i}, \lambda_i^u = 0.$$

Folgerung:

$$\lambda_i^u = \left[\frac{\partial f}{\partial x_i} \right]^+$$

$$\lambda_i^o = \left[\frac{\partial f}{\partial x_i} \right]^-.$$

iii) Auswertung der Variationsungleichung: Ω ist konvex, also gilt

$$\nabla f(\tilde{x})^T (x - \tilde{x}) \geq 0 \quad \forall x \in \Omega \quad \text{d.h.} \quad v \leq x \leq w$$

wähle $x \in \Omega$, so dass $x_j = \begin{cases} \tilde{x}_j, & j \neq i \\ z, & j = i \end{cases}$, mit $z \in \mathbb{R}$, so dass $v_i \leq z \leq w_i$.

Dann folgt

$$\frac{\partial f}{\partial x_i} \tilde{x}_i \leq \frac{\partial f}{\partial x_i} z \quad \forall z \in v_i \leq z \leq w_i$$

$$\text{d.h.} \quad \tilde{x}_i = \arg \min_{v_i \leq z \leq w_i} \frac{\partial f}{\partial x_i} z.$$

Folglich

$$\begin{aligned}\frac{\partial f}{\partial x_i} > 0 &\Rightarrow \tilde{x}_i = v_i \\ \frac{\partial f}{\partial x_i} < 0 &\Rightarrow \tilde{x}_i = w_i.\end{aligned}$$

Für $\frac{\partial f}{\partial x_i} = 0$ ist keine Aussage möglich.

iv) Bedingung 2. Ordnung:

Es ist $\nabla_{xx}\mathcal{L}(\tilde{x}, \tilde{\lambda}) = f''(\tilde{x})$

$$\begin{aligned}C(\tilde{x}, \tilde{\lambda}) &= \left\{ d \in \mathbb{R}^n \mid \nabla c_i^T d = 0 \quad \forall i \in \mathcal{A}(\tilde{x}) \right. \\ &\quad \left. \text{mit } \tilde{\lambda}_i > 0 \quad \text{und } \nabla c_i^T d \geq 0, \quad \text{falls } \tilde{\lambda}_i = 0 \right\} \\ &= \left\{ d \in \mathbb{R}^n \mid d_i = 0, \quad \text{falls } \left| \frac{\partial f}{\partial x_i} \right| \neq 0, \right. \\ &\quad \left. \text{sonst } d_i \geq 0, \quad \text{falls } \tilde{x}_i = v_i, \quad d_i \leq 0, \quad \text{falls } \tilde{x}_i = w_i \right\}.\end{aligned}$$

Wir fordern etwas stärker

$$\begin{aligned}d^T f''(\tilde{x})d &\geq \alpha \|d\|^2 \quad \forall d \in \mathbb{R}^n \\ \text{mit } d_i &= 0 \quad \text{falls } \left| \frac{\partial f}{\partial x_i} \right| \neq 0.\end{aligned}\tag{5.22}$$

Für die numerische Untersuchung wählt man

$$Z = \text{diag}(c_i)_{i=1,\dots,n} \quad \text{mit} \quad c_i = \begin{cases} 1, & \left| \frac{\partial f}{\partial x_i} \right| = 0 \\ 0, & \text{sonst} \end{cases}$$

dann ist (5.22) äquivalent zur positiven Definitheit von $Z^T f'' Z$.

5.7 Weitere Regularitätsbedingungen

Eine Situation, in der der Linearisierungskegel $L_\Omega(x)$ angemessen die zulässige Menge Ω charakterisiert, ist im Fall affin linearer Nebenbedingungen gegeben, d.h.,

$$c_i(x) = a_i^T x + b_i$$

für $a_i \in \mathbb{R}^n$ und $b_i \in \mathbb{R}$. In diesem Fall gilt

Lemma 5.7.1 *Seien in $\tilde{x} \in \mathbb{R}$ alle aktiven Nebenbedingungen affin linear. Dann ist*

$$L_\Omega(\tilde{x}) = T_\Omega(\tilde{x}).$$

BEWEIS. Wir wissen bereits, dass $T_\Omega(\tilde{x}) \subset L_\Omega(\tilde{x})$. Sei nun $w \in L_\Omega(\tilde{x})$. Nach Definition gilt

$$L_\Omega(\tilde{x}) = \left\{ d \in \mathbb{R}^n \mid a_i^T d = 0 \quad \forall i \in \mathcal{E} \quad \text{und} \quad a_i^T d \geq 0 \quad \forall i \in \mathcal{A}(\tilde{x}) \cap I \right\}.$$

Sei $i \in I \setminus \mathcal{A}(\tilde{x})$, d.h. c_i inaktiv, $c_i(\tilde{x}) > 0$, dann existiert $\bar{t} > 0$, so dass

$$c_i(\tilde{x} + tw) > 0 \quad \forall t \in [0, \bar{t}]$$

d.h. c_i bleibt inaktiv.

Wir definieren

$$z^k := \tilde{x} + \frac{\bar{t}}{k} w$$

dann ist für $i \in I \cap \mathcal{A}(\tilde{x})$

$$\begin{aligned} c_i(z^k) &= c_i(z^k) - c_i(\tilde{x}) = a_i^T(z^k - \tilde{x}) \\ &= \frac{\bar{t}}{k} a_i^T w \geq 0 \quad \text{nach Definition von } L_\Omega(\tilde{x}). \end{aligned}$$

Ebenso ist für $i \in \mathcal{E}$

$$c_i(z^k) = c_i(z^k) - c_i(\tilde{x}) = \frac{\bar{t}}{k} a_i^T w = 0.$$

Also ist $\{z^k\}$ eine zulässige Approximation von \tilde{x} . Außerdem ist

$$\lim_{k \rightarrow \infty} \frac{z^k - \tilde{x}}{\bar{t}/k} = \lim_{k \rightarrow \infty} \frac{(\bar{t}/k)w}{\bar{t}/k} = w$$

und damit $w \in T_\Omega(\tilde{x})$. □

Definition 5.7.1 Die Mangasarian-Fromovitz-Regularitätsbedingung (MFCQ) ist erfüllt, falls ein $w \in \mathbb{R}^n$ existiert, so dass

$$\begin{aligned} \nabla c_i(\tilde{x})^T w &> 0 \quad \forall i \in \mathcal{A}(\tilde{x}) \cap I \\ \nabla c_i(\tilde{x})^T w &= 0 \quad \forall i \in \mathcal{E} \end{aligned}$$

und die Vektoren $\{\nabla c_i(\tilde{x}), i \in \mathcal{E}\}$ linear unabhängig sind.

Bemerkungen:

- (1) Man kann zeigen, dass aus (MFCQ) folgt, dass $L_\Omega(\tilde{x}) = T_\Omega(\tilde{x})$.
- (2) Falls (LICQ) erfüllt ist, existiert ein $w \in \mathbb{R}^2$ mit

$$\begin{aligned} \nabla c_i(\tilde{x})^T w &= 1 \quad \forall i \in \mathcal{A}(\tilde{x}) \cap I \\ \nabla c_i(\tilde{x})^T w &= 0 \quad \forall i \in \mathcal{E}, \end{aligned}$$

d.h. (LICQ) \Rightarrow (MFCQ), (LICQ) ist die stärkere Bedingung.

- (3) Die Lagrange'schen Multiplikatoren sind im Fall von (MFCQ) nicht eindeutig bestimmt.

Eine hinreichende Bedingung für (MFCQ) ist die globale Slater-Bedingung:

Definition 5.7.2 Sei $D \subset \mathbb{R}^n$ offen und konvex, $-c_i, i \in I$ sei konvex und differenzierbar auf D , $c_i, i \in \mathcal{E}$ sei affin linear, d.h. es gibt $a_i \in \mathbb{R}^n, i = 1, \dots, m, b \in \mathbb{R}^m$, so dass

$$c_i(x) = a_i^T x + b_i.$$

Dann ist die globale Slater-Bedingung (SCQ) erfüllt, falls zusätzlich $\{a_i, i = 1, \dots, m\}$ linear unabhängig ist und es ein $v \in \mathbb{R}^n$ gibt mit

$$\begin{aligned} c_i(v) &= 0 \quad \forall i \in \mathcal{E} \\ c_i(v) &> 0 \quad \forall i \in I. \end{aligned}$$

Folgerung: Es gelte (SCQ), dann ist in jedem $\tilde{x} \in \Omega$ (MFCQ) erfüllt, d.h. in jedem $\tilde{x} \in \Omega$ gilt $L_\Omega(\tilde{x}) = T_\Omega(\tilde{x})$.

BEWEIS. Sei $\tilde{x} \in \Omega$ beliebig, dann ist für $i \in \mathcal{E}$

$$\nabla c_i(\tilde{x})^T (v - \tilde{x}) = a_i^T (v - \tilde{x}) = 0.$$

Für $i \in I$ ist $-c_i$ konvex, also ist

$$\nabla c_i(\tilde{x})^T (v - \tilde{x}) \geq c_i(v) - c_i(\tilde{x})$$

falls $i \in I \cap \mathcal{A}(\tilde{x})$, ist $c_i(\tilde{x}) = 0$, d.h. für $d = v - \tilde{x}$ gilt

$$\begin{aligned} \nabla c_i(\tilde{x})^T d &= 0, \quad i \in \mathcal{E} \\ \nabla c_i(\tilde{x})^T d &\geq c_i(v) > 0, \quad i \in I \cap \mathcal{A}(\tilde{x}). \end{aligned}$$

□

5.8 Geometrische Interpretation der notwendigen Optimalitätsbedingungen

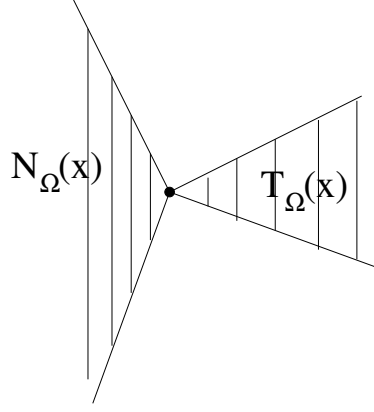
Ziel ist die Beschreibung der Optimalitätsbedingungen unabhängig von der Art der algebraischen Beschreibung der zulässigen Mengen, d.h. wir betrachten

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

Definition 5.8.1 Der Normalenkegel zum Tangentialkegel in $x \in \Omega$ ist definiert durch

$$N_\Omega(x) = \left\{ v \in \mathbb{R}^n \mid v^T w \leq 0 \quad \forall w \in T_\Omega(x) \right\}.$$

Jedes $v \in N_\Omega(x)$ heißt Normalenvektor.



Sei \tilde{x} lokale Lösung von (P), dann wissen wir, dass

$$\begin{aligned} & \nabla f(\tilde{x})^T d \geq 0 \quad \forall d \in T_\Omega(x) \\ \Leftrightarrow & -\nabla f(\tilde{x})^T d \leq 0 \quad \forall d \in T_\Omega(x) \\ \Leftrightarrow & \nabla f(\tilde{x}) \in N_\Omega(\tilde{x}) \end{aligned}$$

d.h. es gilt

Satz 5.8.1 Sei \tilde{x} lokale Lösung von (P), dann ist

$$-\nabla f(\tilde{x}) \in N_\Omega(\tilde{x}).$$

Bemerkung: Falls $\tilde{x} \in \text{int } \Omega$, dann ist $T_\Omega(\tilde{x}) = \mathbb{R}^n$ d.h. $N_\Omega(\tilde{x}) = \{0\}$ und wir haben $\nabla f(\tilde{x}) = 0$. Anderenfalls legt Satz 5.8.1 nahe, dass $\nabla f(\tilde{x})$ in Beziehung steht zu Linearkombinationen der aktiven Nebenbedingungen. In der Tat gilt:

Lemma 5.8.1 Es gelte (LICQ) in der lokalen Lösung \tilde{x} von (P). Dann gilt $N_\Omega(\tilde{x}) = -N$, wobei N definiert ist als

$$N = \left\{ \sum_{i \in \mathcal{A}(\tilde{x})} \lambda_i \nabla c_i(\tilde{x}), \lambda_j \geq 0 \quad \text{für } j \in \mathcal{A}(\tilde{x}) \cap I \right\}.$$

BEWEIS. Nach dem Farkas-Lemma gilt

$$\begin{aligned} & g \in N \Rightarrow g^T d \geq 0 \quad \forall d \in L_\Omega(\tilde{x}) = T_\Omega(\tilde{x}) \\ \text{d.h.} \quad & g \in -N \Rightarrow g^T d \leq 0 \quad \forall d \in T_\Omega(\tilde{x}) \\ \Rightarrow & N_\Omega(\tilde{x}) = -N. \end{aligned}$$

□

Nach Lemma 5.8.1 gilt also $-\nabla f(\tilde{x}) \in -N$, d.h. es gibt Multiplikatoren $\tilde{\lambda}_i$, so dass

$$-\nabla f(\tilde{x}) = - \sum_{i \in \mathcal{A}(\tilde{x})} \tilde{\lambda}_i c_i(\tilde{x})$$

bzw.

$$\nabla f(\tilde{x}) - \sum_{i \in \mathcal{A}(\tilde{x})} \tilde{\lambda}_i c_i(\tilde{x}) = 0,$$

d.h. wir erhalten wieder die KKT-Bedingungen.

5.9 Langrange'sche Multiplikatoren und Sensitivität

Hier wollen wir zeigen, welche Rolle die Lagrange'schen Multiplikatoren von aktiven Ungleichungsrestriktionen haben hinsichtlich der Bedeutung einer bestimmten Nebenbedingung.

Dazu betrachten wir

$$\begin{aligned} & \min f(x) \\ & \text{unter } c_i(x) \geq 0, \quad i \in I. \end{aligned} \quad (\text{P})$$

Falls c_i inaktiv, ist $c_i(\tilde{x}) > 0$, und es ist

$$c_i(z) > 0 \quad \forall z \in B_\varepsilon(\tilde{x}) \cap \Omega.$$

Wegen der Komplementaritätsbedingung gilt dann $\tilde{\lambda}_i = 0$, d.h. der Lagrange'sche Multiplikator zeigt, dass die Ungleichung c_i nicht relevant ist.

Sei jetzt $c_i(\tilde{x}) = 0$, d.h. $i \in \mathcal{A}(\tilde{x})$. Wir betrachten eine kleine Störung, dh. an Stelle von $c_i(x)$ betrachten wir

$$\hat{c}_i(x) = c_i(x) + \varepsilon \|\nabla c_i(\tilde{x})\| \geq 0.$$

Sei $\tilde{x}(\varepsilon)$ die Lösung des Problems P_ε in dem c_i durch \hat{c}_i ersetzt wurde. Wir nehmen an, dass für ε klein $\mathcal{A}(\tilde{x}) = \mathcal{A}(\tilde{x}(\varepsilon))$, dann gilt $\hat{c}_i(\tilde{x}(\varepsilon)) = 0$ d.h.

$$-\varepsilon \|\nabla c_i(\tilde{x})\| = c_i(\tilde{x}(\varepsilon)) - \underbrace{c_i(\tilde{x})}_{=0} \approx (\tilde{x}(\varepsilon) - \tilde{x})^T \nabla c_i(\tilde{x}).$$

Indices $j \in \mathcal{A}(\tilde{x})$ mit $j \neq i$ sind nicht gestört, d.h. für diese gilt

$$0 = c_j(\tilde{x}(\varepsilon)) - c_j(\tilde{x}) \approx (\tilde{x}(\varepsilon) - \tilde{x})^T \nabla c_j(\tilde{x}).$$

Insgesamt folgt aus der KKT-Bedingung

$$\begin{aligned} f(\tilde{x}(\varepsilon)) - f(\tilde{x}) & \approx (\tilde{x}(\varepsilon) - \tilde{x})^T \nabla f(\tilde{x}) = (\tilde{x}(\varepsilon) - \tilde{x})^T \sum_{j \in \mathcal{A}(\tilde{x})} \tilde{\lambda}_j \nabla c_j(\tilde{x}) \\ & \approx (\tilde{x}(\varepsilon) - \tilde{x})^T \tilde{\lambda}_i \nabla c_i(\tilde{x}) \approx -\varepsilon \|\nabla c_i(\tilde{x})\| \tilde{\lambda}_i. \end{aligned}$$

Für $\varepsilon \searrow 0$ ergibt sich

$$\left. \frac{df(\tilde{x}(\varepsilon))}{d\varepsilon} \right|_{\varepsilon=0} = -\|\nabla c_i(\tilde{x})\| \tilde{\lambda}_i.$$

Daraus schließen wir, dass für $\tilde{\lambda}_i \|\nabla c_i\|$ groß sich eine große Änderung im Wert der Zielfunktion ergibt. Andererseits, falls $\tilde{\lambda}_i = 0$, werden sich kleine Störungen in c_i kaum auf den Wert der Zielfunktion auswirken. Daher formulieren wir:

Definition 5.9.1 Sei \tilde{x} Lösung von (P), es gelten die KKT-Bedingungen. Eine Ungleichungsnebenbedingung c_i heißt stark aktiv, falls $i \in \mathcal{A}(\tilde{x})$ und $\lambda_i > 0$. c_i heißt schwach aktiv, falls $i \in \mathcal{A}(\tilde{x})$ und $\lambda_i = 0$.

5.10 Dualität

In diesem Abschnitt beschränken wir uns auf Ungleichungsrestriktionen, d.h. wir untersuchen

$$\begin{array}{ll} \min & f(x) \\ \text{unter} & c_i(x) \geq 0, \quad 1 \leq i \leq m. \end{array} \quad (\text{P})$$

Dazu setzen wir $c(x) := (c_1(x), \dots, c_m(x))^T$, dann ist die Lagrange-Funktion

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x) \quad \text{mit } \lambda \in \mathbb{R}^m.$$

Nun definieren wir die duale Zielfunktion $q: \mathbb{R}^m \rightarrow \mathbb{R}$ durch

$$q(\lambda) := \inf_x \mathcal{L}(x, \lambda). \quad (5.23)$$

In vielen Fällen ist für bestimmte Werte von λ $\inf_x \mathcal{L}(x, \lambda) = -\infty$, daher vereinbaren wir

$$D := \{ \lambda \in \mathbb{R}^m \mid q(\lambda) > -\infty \}$$

als Definitionsbereich von q .

Das globale Minimum in (5.23) zu berechnen, kann schwierig sein, aber für den uns interessierenden Fall, dass f und $-c_i$ konvex sind und die $\lambda_i \geq 0$, ist $\mathcal{L}(\cdot, \lambda)$ konvex, d.h. alle lokalen Minimierer sind globale und die Berechnung vereinfacht sich deutlich, da nur die notwendigen Bedingungen überprüft werden müssen.

Wir definieren als das duale Problem zu (P)

$$\begin{array}{ll} \max & q(\lambda) \\ \text{unter} & \lambda \geq 0. \end{array} \quad (\text{P}_D)$$

Beispiel 5.10.1

$$\begin{array}{ll} \min & \frac{1}{2} (x_1^2 + x_2^2) \\ \text{unter} & x_1 - 1 \geq 0. \end{array}$$

Geometrisch findet man schnell als Lösung $\tilde{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$$\mathcal{L}(x, \lambda) = \frac{1}{2} (x_1^2 + x_2^2) - \lambda_1 (x_1 - 1).$$

Die Auswertung der KKT-Bedingung ergibt

$$\nabla f(\tilde{x}) = \tilde{x} \stackrel{!}{=} \lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

d.h. $\tilde{\lambda}_1 = 1$, und der optimale Wert der Zielfunktion ist

$$f(\tilde{x}) = \frac{1}{2} (1 + 0) = \frac{1}{2}.$$

Für λ_1 fest ist \mathcal{L} konvex in $x \in \mathbb{R}^2$, d.h. wir finden das globale Minimum durch Auswerten von $\nabla_x \mathcal{L} = 0$, d.h.

$$\begin{aligned} x_1 - \lambda_1 &= 0 \\ x_2 &= 0. \end{aligned}$$

Einsetzen in \mathcal{L} ergibt

$$\begin{aligned} q(\lambda_1) &= \frac{1}{2} \lambda_1^2 - \lambda_1(\lambda_1 - 1) \\ &= -\frac{1}{2} \lambda_1^2 + \lambda_1 = \lambda_1 \left(1 - \frac{1}{2} \lambda_1 \right). \end{aligned}$$

Das duale Problem lautet dann

$$\max_{\lambda_1 \geq 0} \left(-\frac{1}{2} \lambda_1^2 + \lambda_1 \right)$$

mit Lösung $\lambda_1 = 1$ und $q(\lambda_1) = \frac{1}{2}$, d.h. wir erhalten denselben Wert für λ_1 und denselben Optimalwert wie im primalen Problem.

Im Folgenden untersuchen wir die Beziehung zwischen (P) und (P_D) genauer.

Lemma 5.10.1 $q(\lambda)$ ist konkav und ihr Definitionsbereich D ist konvex.

BEWEIS. Seien $\lambda^1, \lambda^2 \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ und $\alpha \in [0, 1]$, dann

$$\mathcal{L}(x, (1 - \alpha)\lambda^1 + \alpha\lambda^2) = (1 - \alpha)\mathcal{L}(x, \lambda^1) + \alpha\mathcal{L}(x, \lambda^2).$$

Wegen $\inf(f(x) + g(x)) \geq \inf f(x) + \inf g(x)$ folgt

$$q((1 - \alpha)\lambda^1 + \alpha\lambda^2) \geq (1 - \alpha)q(\lambda^1) + \alpha q(\lambda^2). \quad (5.24)$$

Für $\lambda^1, \lambda^2 \in D$ folgt daraus

$$q((1 - \alpha)\lambda^1 + \alpha\lambda^2) > -\infty,$$

d.h. $(1 - \alpha)\lambda^1 + \alpha\lambda^2 \in D$ und D ist konvex. Die Konkavität von q folgt ebenfalls aus (5.24).

Lemma 5.10.2 Für jedes zulässige $\bar{x} \in \Omega$ für (P) und jedes $\bar{\lambda} \geq 0$ gilt

$$q(\bar{\lambda}) \leq f(\bar{x}).$$

BEWEIS.

$$q(\bar{\lambda}) = \inf_x (f(x) - \bar{\lambda}^T c(x)) \leq f(\bar{x}) - \underbrace{\bar{\lambda}^T c(\bar{x})}_{\geq 0} \leq f(\bar{x}).$$

□

Im Folgenden verwenden wir die KKT-Bedingungen für das Problem (P). Diese lauten

$$\begin{aligned}\nabla f(\bar{x}) - c'(\bar{x})^T \bar{\lambda} &= 0 \\ c(\bar{x}) &\geq 0 \\ \bar{\lambda} &\geq 0 \\ \bar{\lambda}_i c_i(\bar{x}) &= 0, \quad i = 1, \dots, m.\end{aligned}$$

Dabei ist $c'(\bar{x})^T = [\mathcal{D}c(\bar{x})]^T = (\nabla c_1(\bar{x}), \dots, \nabla c_m(\bar{x}))$.

Die optimalen Lagrange-Parameter für (P) sind unter bestimmten Umständen Lösungen des dualen Problems (P_D) , genauer gilt

Satz 5.10.1 *Seien \bar{x} Lösung von (P) und f und $-c_i(x)$ konvexe Funktionen auf \mathbb{R}^n . Dann ist jedes $\bar{\lambda}$, so dass $(\bar{x}, \bar{\lambda})$ die KKT-Bedingungen erfüllt, eine Lösung des dualen Problems (P_D) .*

BEWEIS. $(\bar{x}, \bar{\lambda})$ erfüllen die KKT-Bedingungen, $\bar{\lambda} \geq 0$

$\Rightarrow \mathcal{L}(\cdot, \bar{\lambda})$ ist konvex und differenzierbar

$\Rightarrow \mathcal{L}(x, \bar{\lambda}) \geq \mathcal{L}(\bar{x}, \bar{\lambda}) + \underbrace{\nabla_x \mathcal{L}(\bar{x}, \bar{\lambda})^T (x - \bar{x})}_{=0} = \mathcal{L}(\bar{x}, \bar{\lambda})$

\uparrow
 Konvexität

$$\Rightarrow q(\bar{\lambda}) = \inf_x \mathcal{L}(x, \bar{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) = f(\bar{x}). \quad (5.25)$$

Andererseits folgt aus Lemma 5.10.2 $q(\lambda) \leq f(\bar{x})$ für alle $\lambda \geq 0$, d.h. $\bar{\lambda}$ ist Lösung von (P_D) . \square

Unter zusätzlichen Voraussetzungen existiert auch die Umkehrung dieser Aussage, in dem Sinne, dass die Lösung des dualen Problems manchmal verwendet werden kann für die Berechnung der Lösung von (P).

Satz 5.10.2 *Seien f und $-c_i, i = 1, \dots, m$ konvex und stetig differenzierbar auf \mathbb{R}^n und \bar{x} eine Lösung von (P) für die (LICQ) erfüllt ist. Ferner sei $\hat{\lambda}$ die Lösung von (P_D) , wobei das Infimum von $\mathcal{L}(x, \hat{\lambda})$ angenommen werde in \hat{x} . Außerdem sei $\mathcal{L}(\cdot, \hat{\lambda})$ strikt konvex. Dann ist $\bar{x} = \hat{x}$, d.h. die eindeutige Lösung von (P) und $f(\bar{x}) = \mathcal{L}(\hat{x}, \hat{\lambda})$.*

BEWEIS. Angenommen $\bar{x} \neq \hat{x}$. (LICQ) gilt in $\bar{x} \Rightarrow$ es gibt $\bar{\lambda}$, so dass die KKT-Bedingungen erfüllt sind. Also folgt mit Satz 5.10.1, dass neben $\hat{\lambda}$ auch $\bar{\lambda}$ Lösung von (P_D) ist, und nach (5.25) gilt

$$\mathcal{L}(\bar{x}, \bar{\lambda}) = q(\bar{\lambda}) = q(\hat{\lambda}) = \mathcal{L}(\hat{x}, \hat{\lambda}).$$

Nach Voraussetzung ist $\hat{x} = \arg \min_x \mathcal{L}(x, \bar{\lambda})$, also muss gelten

$$\nabla_x \mathcal{L}(\hat{x}, \hat{\lambda}) = 0$$

und wegen strikter Konvexität von $\mathcal{L}(x, \hat{\lambda})$ folgt

$$\mathcal{L}(\bar{x}, \hat{\lambda}) - \mathcal{L}(\hat{x}, \hat{\lambda}) > \nabla_x \mathcal{L}(\hat{x}, \hat{\lambda})^T (\bar{x} - \hat{x}) = 0$$

also ist

$$\begin{aligned} \mathcal{L}(\bar{x}, \hat{\lambda}) &> \mathcal{L}(\hat{x}, \hat{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda}) \\ \Rightarrow \quad f(\bar{x}) - \hat{\lambda}^T c(\bar{x}) &> f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) \\ \text{d.h.} \quad -\hat{\lambda}^T c(\bar{x}) &> \bar{\lambda}^T c(\bar{x}) = 0 \end{aligned}$$

im Widerspruch zu $\hat{\lambda} \geq 0$ und $c(\bar{x}) \geq 0$, d.h. die Annahme war falsch und es gilt $\hat{x} = \bar{x}$. \square

Für numerische Berechnungen ist folgende duale Aufgabe oft vorzuziehen, die auf Wolfe zurückgeht:

$$\left. \begin{array}{l} \max_{x, \lambda} \mathcal{L}(x, \lambda) \\ \text{unter} \quad \nabla_x \mathcal{L}(x, \lambda) = 0 \\ \lambda \geq 0. \end{array} \right\} \quad (\text{P}_{DW})$$

Der Zusammenhang zwischen (P) und (P_{DW}) wird beschrieben in

Satz 5.10.3 Seien f und $-c_i$, $1 \leq i \leq m$ konvex und stetig differenzierbar auf \mathbb{R}^n . Sei ferner $(\bar{x}, \bar{\lambda})$ eine Lösung von (P), für die (LICQ) erfüllt ist. Dann löst $(\bar{x}, \bar{\lambda})$ auch (P_{DW}).

BEWEIS. KKT-Bedingung $\Rightarrow \nabla_x \mathcal{L}(\bar{x}, \bar{\lambda}) = 0$ und $\bar{\lambda} \geq 0$ und $\mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x})$.

Also folgt für jedes (x, λ) , dass die Nebenbedingungen in (P_{DW}) erfüllt:

$$\begin{aligned} \mathcal{L}(\bar{x}, \bar{\lambda}) &= f(\bar{x}) \geq f(\bar{x}) - \underbrace{\lambda^T c(\bar{x})}_{\geq 0} \\ &= \mathcal{L}(\bar{x}, \lambda) \geq \mathcal{L}(x, \lambda) + \nabla_x \mathcal{L}(x, \lambda)(\bar{x} - x) = \mathcal{L}(x, \lambda). \end{aligned}$$

\uparrow
Konvexität

\square

Beispiel 5.10.2 (lineare Optimierung)

$$\min c^T x, \quad \text{unter} \quad Ax - b \geq 0$$

hier ist $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$ duale Zielfunktion:

$$\begin{aligned} q(\lambda) &= \inf_x \mathcal{L}(x, \lambda) = \inf_x (c^T x - \lambda^T (Ax - b)) \\ &= \inf_x [(c - A^T \lambda)^T x + b^T \lambda]. \end{aligned}$$

Falls $c - A^T \lambda \neq 0$, setze $x = -(c - A^T \lambda)r$, mit $r \in \mathbb{R}$, dann ist

$$(c - A^T \lambda)^T x + b^T \lambda = -r \|c - A^T \lambda\|^2 + b^T \lambda \xrightarrow{r \rightarrow \infty} -\infty$$

d.h.

$$\inf_x [(c - A^T \lambda)^T x + b^T \lambda] = -\infty,$$

somit muss bei Maximierung der Fall $c - A^T \lambda \neq 0$ nicht betrachtet werden. Als duales Problem ergibt sich

$$\begin{cases} \max b^T \lambda \\ \text{unter } A^T \lambda = c \\ \lambda \geq 0. \end{cases}$$

Weiterhin ist

$$\begin{aligned} \mathcal{L}(x, \lambda) &= (c - A^T \lambda)^T x + b^T \lambda \\ \nabla_x \mathcal{L}(x, \lambda) &= c - A^T \lambda. \end{aligned}$$

Damit ergibt sich als Wolfe-duale Aufgabe

$$\begin{aligned} \max_{x, \lambda} & (c - A^T \lambda)^T x + b^T \lambda \\ \text{unter } & c - A^T \lambda = 0, \lambda \geq 0 \end{aligned}$$

bzw.

$$\begin{aligned} \max & b^T \lambda \\ \text{unter } & c - A^T \lambda = 0 \\ & \lambda \geq 0. \end{aligned}$$

Beispiel 5.10.3 (linear-quadratische Optimierung)

$G \in \mathbb{R}^{n,n}$ symmetrisch, positiv definit, $A \in \mathbb{R}^{m,n}$

$$\begin{aligned} \min & \frac{1}{2} x^T G x + c^T x \\ \text{unter } & A(x) - b \geq 0. \end{aligned} \tag{P}$$

Hier ist die duale Zielfunktion

$$q(\lambda) = \inf_x \mathcal{L}(x, \lambda) = \inf_x \frac{1}{2} x^T G x + c^T x - \lambda^T (Ax - b),$$

G ist positiv definit, d.h. $\mathcal{L}(\cdot, \lambda)$ ist strikt konvex und das Infimum wird erreicht für $\nabla_x \mathcal{L}(x, \lambda) = 0$, d.h.

$$\begin{aligned} Gx + c - A^T \lambda &= 0 \\ \Leftrightarrow x &= G^{-1}(A^T \lambda - c). \end{aligned}$$

Einsetzen ergibt

$$\begin{aligned} q(\lambda) &= \frac{1}{2}(A^T\lambda - c)^T \underbrace{G^{-1}G}_{I} G^{-1}(A^T\lambda - c) \\ &\quad + c^T G^{-1}(A^T\lambda - c) - \lambda^T A G^{-1}(A^T\lambda - c) + \lambda^T b \\ q(\lambda) &= -\frac{1}{2}(A^T\lambda - c)^T G^{-1}(A^T\lambda - c) + \lambda^T b. \end{aligned}$$

Für das Wolfe-duale Problem ergibt sich

$$\begin{aligned} \max_{\lambda, x} \quad & \frac{1}{2} x^T G x + c^T x - \lambda^T (A x - b) \\ \text{unter} \quad & G x + c - A^T \lambda = 0 \\ & \lambda \geq 0. \end{aligned}$$

Ersetzt man mit der Gleichungsrestriktion

$$(c - A^T \lambda)^T x = -x^T G x$$

erhält man

$$\begin{aligned} \max_{\lambda, x} \quad & -\frac{1}{2} x^T G x + \lambda^T b \\ \text{unter} \quad & \begin{cases} G x + c - A^T \lambda = 0 \\ \lambda \geq 0. \end{cases} \end{aligned}$$

5.11 Ausblick: Numerische Lösung nichtlineare Optimierungsprobleme mit Nebenbedingungen

In Kapitel 6 werden wir Probleme mit linearen Nebenbedingungen studieren. Diese sind wichtig, weil es nur effektive numerische Lösungsverfahren dafür gibt, und weil sie als Baustein für die Lösung von Problemen mit nichtlinearen Nebenbedingungen dienen.

In Kapitel 7 untersuchen wir SQP-Methoden zur Lösung nichtlinearer restringierter Probleme. In jedem Iterationsschritt (x_k, λ_k) wird hier die Lösung des linear-quadratischen Problems

$$\begin{aligned} \min_p \quad & \frac{1}{2} p^T \nabla_{xx} \mathcal{L}(x_k, \lambda_k) p + \nabla f(x_k)^T p \\ \text{unter} \quad & \nabla c_i(x_k)^T p + c_i(x_k) = 0, \quad i \in \mathcal{E} \\ & \nabla c_i(x_k)^T p + c_i(x_k) \geq 0, \quad i \in I. \end{aligned}$$

In Kapitel 8 diskutieren wir Strafmethode und Augmentierte-Lagrange-Methoden. Durch Einführung eines Strafterms lassen sich restringierte Probleme auf eine Folge unrestringierter zurückführen. Im Fall reiner Gleichungsnebenbedingungen kann man beispielsweise betrachten

$$f(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(x). \quad (5.26)$$

Hier ist $\mu > 0$ der so genannte Penalty-Parameter. Zur Berechnung der Lösung von (P) wird (5.26) für eine Folge wachsender Penaltyparameter $\{\mu_k\}$ berechnet.

Augmentierte-Lagrange-Methoden haben in der Regel bessere numerische Eigenschaften, hier verbindet die neue Zielfunktion Eigenschaften der Lagrange-Funktion und der penalisierten Zielfunktion (5.26), genauer wird betrachtet

$$\mathcal{L}_A(x, \lambda; \mu) = f(x) - \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(x).$$

Das entsprechende numerische Verfahren liefert Approximationen für die Lösung \tilde{x} von (P) und für den Lagrange-Parameter.

6 Probleme mit linearen Restriktionen-Verfahren

6.1 Quadratische Optimierungsprobleme

6.1.1 Aufgaben mit Gleichungsrestriktionen

Wir betrachten

$$\boxed{\begin{array}{l} \min_{x \in \mathbb{R}^n} \frac{1}{2} \langle x, Qx \rangle + \langle q, x \rangle \\ \text{bei } Ax = b \end{array}} \quad (\text{QG})$$

Q : symmetrisch, (n, n) , $A : (m, n)$, $m \leq n$.

Voraussetzung:

- $\text{rang } A = m$ (voller Rang)
- $d^T Q d \geq \alpha \|d\|^2 \quad \forall d \in \ker A$

Damit hat (QG) genau eine Lösung \tilde{x} und genau einen zugehörigen Lagrangeschen Multiplikator. Beide erfüllen zusammen das System

$$\mathcal{A} \begin{pmatrix} \tilde{x} \\ \lambda \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix} \quad \text{mit} \quad \mathcal{A} = \begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix} \quad (6.27)$$

Häufig wird das System auch mit Hilfe der Transformation $\tilde{x} = x + p$, $h = Ax - b$, $g = q + Qx$ umgeschrieben in

$$\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} -p \\ \lambda \end{pmatrix} = \begin{pmatrix} q \\ h \end{pmatrix}$$

mit der invertierbaren sogenannten KKT-Matrix

$$K = \begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix}$$

Mögliche numerische Lösungsverfahren für das KKT-System sind

- die direkte Faktorisierung von K (häufig zu aufwendig)
- Schur-Komplement Methoden (Details siehe [2])
- Nullraumverfahren, d.h. die Elimination der Nebenbedingungen.

Wir fokussieren uns auf die letztgenannte Möglichkeit, d.h. wir suchen eine Nullraum-Matrix von A zur Elimination der Nebenbedingungen $Ax = b$. Dazu 3 Schritte:

1. Nullraum-Matrix bestimmen
2. \tilde{x} als Lösung eines „freien“ Optimierungsproblems berechnen
3. λ ausrechnen.

Schritt 1: *QR-Zerlegung von A^T*

Finde unitäre Matrix H und obere Dreiecksmatrix R mit

$$\boxed{HA^T = \begin{pmatrix} R \\ 0 \end{pmatrix}} \quad \begin{array}{l} H : (n, n) \\ R : (m, m) \end{array}$$

Wir spalten H wie folgt auf:

$$\boxed{H = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} \begin{array}{l} \} m \text{ Zeilen} \\ \} n - m \text{ Zeilen} \end{array}}$$

Spalten von Y : erste m Zeilen von H ,
von Z : letzte $(n - m)$ Zeilen.

Sowohl Y als auch Z müssen Vollrang haben, weil $\text{rang } H = n$. Damit spannen die Spalten von Y und Z gemeinsam den ganzen \mathbb{R}^n auf. Jedes $x \in \mathbb{R}^n$ hat damit als eindeutige Darstellung

$$x = Yx_y + Zx_z = H^T \begin{pmatrix} x_y \\ x_z \end{pmatrix}.$$

Das gilt speziell für alle $x = d \in \ker A$, und deshalb

$$0 = Ad = A(Yd_y + Zd_z) = AH^T \begin{pmatrix} d_y \\ d_z \end{pmatrix} = (R^T, 0) \begin{pmatrix} d_y \\ d_z \end{pmatrix} = R^T d_y.$$

Deshalb erhält man alle $d \in \ker A$ durch

$$d = Zd_z, d_z \in \mathbb{R}^{n-m} \text{ beliebig.}$$

\Rightarrow Das oben konstruierte Z ist eine Nullraum-Matrix.

Wir hatten oben x in der Form dargestellt:

$$x = \underbrace{Yx_y}_{\Rightarrow \in (\ker A)^\perp} + \underbrace{Zx_z}_{\in \ker A}.$$

Schritt 2a: \tilde{x} sei die unbekannte Lösung. Einen Teil davon können wir nun sofort berechnen:

$$\begin{aligned}\tilde{x} &= Y\tilde{x}_y + Z\tilde{x}_z \\ \Rightarrow b &= A\tilde{x} = AY\tilde{x}_y + 0 \\ b &= AY\tilde{x}_y = R^T\tilde{x}_y \quad \boxed{R^T\tilde{x}_y = b} \\ \Rightarrow \tilde{x}_y &= (R^T)^{-1}b\end{aligned}$$

(was man durch einfaches Auflösen des Gleichungssystems $R^T\tilde{x}_y = b$ bewerkstelligt.)

Schritt 2b: " Ω = spezielle Lösung von $Ax + b$ plus allgemeine von $Ax = 0$ "

Spezielle Lösung: $Y\tilde{x}_y =: w$. Dann gilt $Aw = b$.

Somit:

$$\begin{aligned}x &\in \Omega \\ \Updownarrow \\ x &= w + Zz, \quad z \in \mathbb{R}^{n-m}.\end{aligned}$$

Allgemeine Lösung: $Zz, z \in \mathbb{R}^{n-m}$

\Rightarrow **Reduziertes Problem:**

$$\min_{z \in \mathbb{R}^{n-m}} f(w + Zz) = \frac{1}{2}(w + Zz)^T Q(w + Zz) + q^T(Zz + w).$$

Durch Ausmultiplizieren vereinfacht sich das, wobei wir den konstanten Term $\frac{1}{2}w^T Qw$ weglassen können:

$$f = \frac{1}{2}z^T \underbrace{Z^T Q Z}_{\tilde{Q}} z + \underbrace{\langle Z^T Q w, z \rangle + \langle Z^T q, z \rangle}_{=\tilde{q}^T z}$$

\Rightarrow neue Form

$$\boxed{\min_{z \in \mathbb{R}^{n-m}} F(z) = \frac{1}{2}z^T \tilde{Q}z + \tilde{q}^T z} \quad (\text{QG})_r$$

$$\begin{aligned}\text{mit } \tilde{Q} &:= Z^T Q Z \\ \tilde{q} &:= Z^T Q w + Z^T q.\end{aligned}$$

Unter unseren Voraussetzungen ist \tilde{Q} positiv definit, damit ist das Problem eindeutig lösbar. Notwendige Bedingung für \tilde{z} :

$$\nabla F(\tilde{z}) = 0 \quad \Leftrightarrow \quad \tilde{Q}\tilde{z} = -\tilde{q},$$

dabei spielt \tilde{z} die Rolle von \tilde{x}_z oben, also

$$Z^T Q Z \tilde{x}_z = -Z^T q - Z^T Q w = -Z^T q - Z^T Q Y \tilde{x}_y.$$

Wie berechnet man günstig die Lösung dieses Systems? Anstelle von $Z^T Q Z$ und $Z^T q$ betrachten wir zunächst das Ganze für die größere, Z^T enthaltene Matrix H :

- Bilde

$$-Hq =: \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \leftarrow \begin{matrix} \mathbb{R}^m \\ \mathbb{R}^{n-m} \end{matrix} =: h$$

- Berechne

$$B := HQH^T = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

NR:

$$HQH^T = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} Q(Y, Z) = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} (QY, QZ) = \begin{pmatrix} Y^T QY & Y^T QZ \\ Z^T QY & Z^T QZ \end{pmatrix}$$

$$\begin{aligned} \Rightarrow \quad B_{11} &= Y^T QY & B_{12} &= Y^T QZ \\ B_{21} &= Z^T QY & B_{22} &= Z^T QZ \quad \text{und} \quad h_1 = -Y^T q \\ & & & = \tilde{Q} & h_2 &= -Z^T q. \end{aligned}$$

Wir haben somit alle für die obige Gleichung für \tilde{x}_z interessanten Terme in H stecken: Die Gleichung lautete

$$\underbrace{Z^T QZ}_{B_{22}} \tilde{x}_z = \underbrace{-Z^T q}_{h_2} - \underbrace{Z^T QY}_{B_{21}} \tilde{x}_y$$

und liest sich nun als

$$B_{22} \tilde{x}_z = h_2 - B_{21} \tilde{x}_y$$

Lösung z.B. mit Cholesky-Zerlegung.

\Rightarrow haben am Ende \tilde{x}_z , \tilde{x}_y und dann

$$\tilde{x} := Y \tilde{x}_y + Z \tilde{x}_z.$$

Schritt 3: Bestimmung des Multiplikators λ

$$Q\tilde{x} - A^T \lambda = -q.$$

Wir setzen \tilde{x} in der obigen Form ein, nämlich

$$\tilde{x} = H^T \begin{pmatrix} \tilde{x}_y \\ \tilde{x}_z \end{pmatrix}$$

und multiplizieren die Gleichung von links mit H .

\Rightarrow

$$\underbrace{HQH^T}_B \begin{pmatrix} \tilde{x}_y \\ \tilde{x}_z \end{pmatrix} - \underbrace{HA^T}_{\begin{pmatrix} R \\ 0 \end{pmatrix}} \lambda = \underbrace{-Hq}_h$$

d. h.

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} \tilde{x}_y \\ \tilde{x}_z \end{pmatrix} - \begin{pmatrix} R \\ 0 \end{pmatrix} \lambda = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

$$\Rightarrow \boxed{R\lambda = -h_1 + B_{11}\tilde{x}_y + B_{12}\tilde{x}_z}.$$

einfach zu lösen, da R obere Dreiecksmatrix. Damit ist alles gelöst!

6.1.2 Aufgaben mit Ungleichungsrestriktionen

Wir betrachten

$$\boxed{\begin{array}{l} \min_{x \in \mathbb{R}^n} \frac{1}{2} \langle x, Qx \rangle + \langle q, x \rangle \\ \text{bei } Ax = b \text{ und } Gx \geq r \end{array}} \quad (\text{QLU})$$

mit Q : symmetrisch, (n, n) , $A : (m, n)$, $m \leq n$, $G : (p, n)$. Die zulässige Menge ist

$$\Omega = \{x \in \mathbb{R}^n \mid Ax = b, Gx \geq r\}.$$

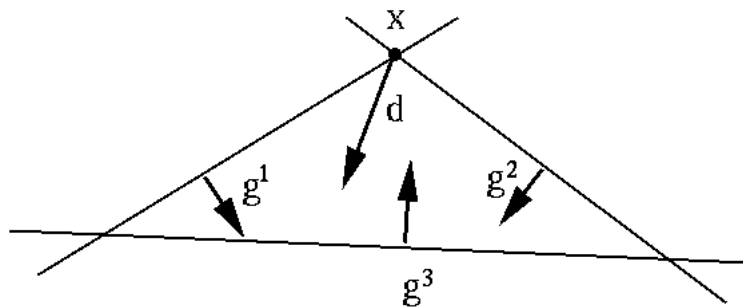
Für den Kegel der zulässigen Richtungen gilt

$$L_\Omega(\tilde{x}) = \{d \in \mathbb{R}^n \mid Ad = 0, \langle g^j, d \rangle \geq 0 \quad \forall j \in J(\tilde{x})\}.$$

Wir schreiben das noch etwas anders auf: Wir fassen wie früher alle Vektoren $g^i, i \in J(x)$, der aktiven Ungleichungen zu einer Matrix $G(x)$ zusammen (sie enthält als Zeilen die Vektoren $(g^i)^T$). Dann gilt

$$\boxed{d \text{ ist zulässige Richtung im Punkt } x \Leftrightarrow Ad = 0, G(x)d \geq 0}$$

Geometrische Illustration: (3 Ungleichungsrestriktionen $\langle g^i, x \rangle \geq r^i, i = 1, 2, 3$)



$$\left. \begin{array}{l} \langle g^1, x \rangle = r^1 \\ \langle g^2, x \rangle = r^2 \end{array} \right\} \text{aktiv} \quad J(x) = \{1, 2\}$$

$$\left. \begin{array}{l} \langle g^3, x \rangle > r^3 \end{array} \right\} \text{inaktiv} \quad G(x) = \begin{pmatrix} (g^1)^T \\ (g^2)^T \end{pmatrix}$$

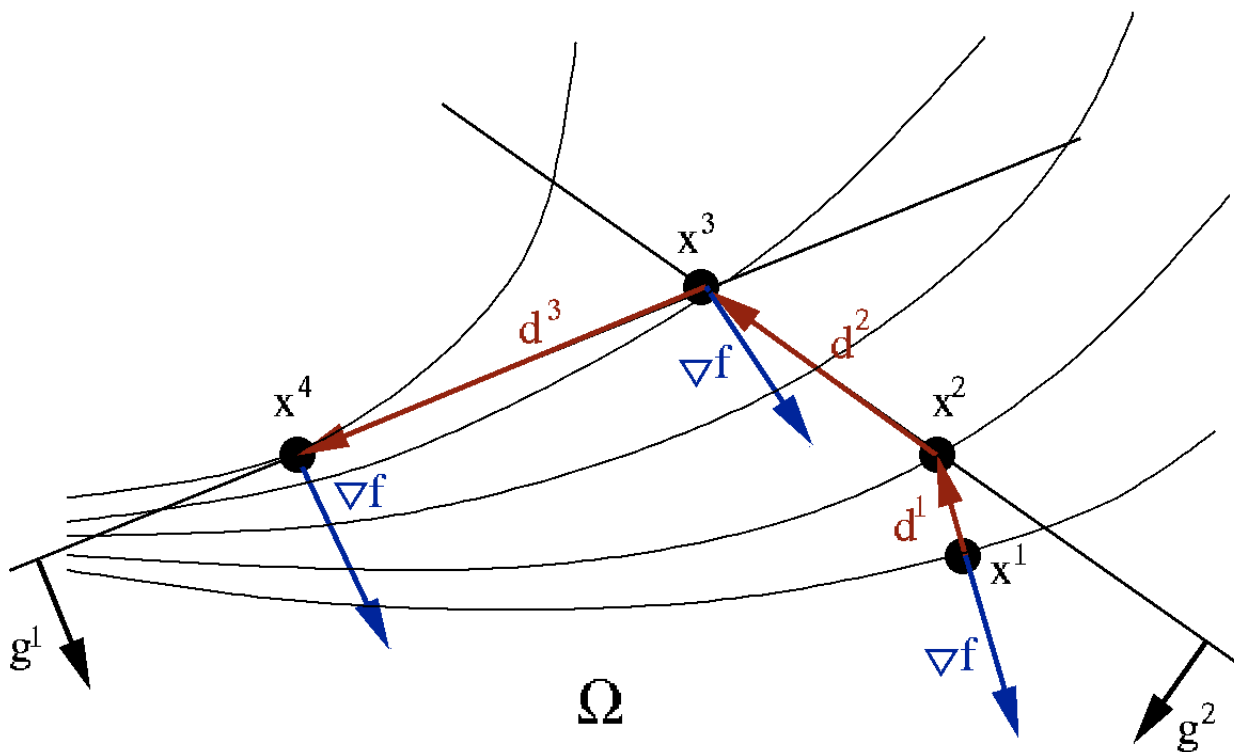
Anschaulich ist klar: Damit $x + td$ für kleine t zulässig bleibt, muss gelten

$$\langle g^1, d \rangle \geq 0 \quad \wedge \quad \langle g^2, d \rangle \geq 0.$$

Die dritte Nebenbedingung – die inaktive – hat darauf keinen Einfluss.

Bevor wir nun zur mathematischen Beschreibung des numerischen Verfahrens – Verfahren zulässiger Richtungen mit aktiver Mengen-Strategie – kommen, wollen wir uns dessen grundlegenden Ideen geometrisch klar machen.

Wir betrachten folgende Konstellation:



Schritt 1:

Startpunkt x^1 liegt im Inneren von Ω , es können zuerst beide Restriktionen ignoriert und mit einem Verfahren der freien Optimierung gestartet werden, bis eine (oder mehrere) Restriktionen aktiv werden.

Schritt 2:

Unser Verfahren hat im Punkt x^2 den Rand von Ω erreicht – die Restriktion Nr. 2 ist aktiv geworden.

- Wäre $g^2 \parallel \nabla f$, dann würde gelten

$$\nabla f - \mu g^2 = 0,$$

in diesem Falle, falls $\mu \geq 0$: **Fertig**. In unserem Bild gilt das nicht, ∇f ist keine Linearkombination von g^2 .

- Wir suchen daher im Unterraum mit $\langle g^2, d \rangle = 0$ weiter, d. h. in

$$x^2 + \{d | \langle g^2, d \rangle = 0\}$$

Schritt 3:

In unserem Fall gelangt das Verfahren schließlich in x^3 zu einem Punkt, in dem eine weitere Restriktion aktiv wird, nämlich $\langle g^1, x \rangle$.

- Offenbar ist x^3 noch nicht optimal, denn

$$\begin{aligned} (*) \quad \nabla f &= \mu^1 g^1 + \mu^2 g^2 \quad \text{mit} \quad \mu^1 > 0 \quad \text{aber} \quad \mu^2 < 0 \\ 0 &= \nabla f - \mu^1 g^1 - \mu^2 g^2 \end{aligned}$$

- In welcher Richtung sollte weiter gesucht werden?

Die Richtung d muss erfüllen:

$$\begin{aligned} & \left. \begin{aligned} \langle g^1, d \rangle &\geq 0 \\ \langle g^2, d \rangle &\geq 0 \end{aligned} \right\} \text{um zulässig zu bleiben} \\ & \quad \langle \nabla f, d \rangle < 0 \\ \updownarrow & \quad \text{um einen Abstieg zu erzielen.} \\ & \quad \langle -\nabla f, d \rangle > 0. \end{aligned}$$

$(*) \Rightarrow$

$$\langle \nabla f, d \rangle = \underbrace{\mu^1}_{>0} \underbrace{\langle g^1, d \rangle}_{\geq 0} + \underbrace{\mu^2}_{<0} \underbrace{\langle g^2, d \rangle}_{\geq 0}$$

Weil $\langle \nabla f, d \rangle$ negativ sein soll, mit möglichst großem Betrag, besteht die Strategie darin $\langle g^1, d \rangle = 0$ zu wählen, aber

$$\langle g^2, d \rangle > 0$$

\Rightarrow Wir **deaktivieren die Restriktion** Nr. 2 und halten Nr. 1 aktiv.

Schritt 4:

Suche im affinen Unterraum mit $\langle g^1, d \rangle = 0$. Als Resultat gelangt das Verfahren hier zu x^4 , der Lösung, denn hier gilt $\nabla f = \mu^1 g^1$ mit $\mu^1 > 0$, und die Optimalitätsbedingungen sind erfüllt

$$\begin{aligned} 0 &= \nabla f - \mu^1 g^1 - 0 \cdot g^2 \\ \langle g^1, x^4 \rangle &= r^1, \quad \langle g^2, x^4 \rangle > r^2 \\ \left(\langle g^2, x^4 \rangle - r^2 \right) \cdot 0 &= 0. \end{aligned}$$

Fazit:

- Stop, wenn die notwendigen Bedingungen mit $\mu^i \geq 0$ erfüllt sind
- Aktivierung von Nebenbedingungen, auf die das Verfahren trifft
- Deaktivierung, wenn Multiplikatoren negativ werden (wird noch präzisiert).
- Ansonsten Suche in affin-linearen Unterräumen, d. h. Optimierung bei Gleichungs-
nebenbedingungen.

Wir kommen nun zur mathematischen präzisen Formulierung des Verfahrens

- Aktueller Iterationspunkt: x^k
 $J_k := J(x^k)$: Menge der aktiven Indizes (hier gilt $\langle g^i, x \rangle = r^i, i \in J_k$)
 $p_k = |J(x^k)|$: Zahl der aktiven Indizes
 $G_k = G(x^k)$: Matrix der $g^i, i \in J_k$ (genauer: mit Zeilen g_i^T)
 $B_k = \begin{pmatrix} A \\ G_k \end{pmatrix}$: Beschreibt das zur Zeit aktive lineare Gleichungssystem
- Ausgehend von x^k wird ein im nächsten Schritt zu lösendes *quadratisches Optimierungsproblem* aufgestellt (mit Gleichungsrestriktionen)

$$\begin{array}{l} \min_{d \in \mathbb{R}^n} \frac{1}{2} \langle Qd, d \rangle + \langle Qx^k + q, d \rangle \\ \text{bei } B_k d = 0 \end{array} \quad (Q_k)$$

Bemerkung: Bis auf eine von x^k abhängige Konstante ist die Zielfunktion von (Q_k) gerade $f(x^k + d)$.

Ergebnis:

- Richtung d^k
- Multiplikatoren $\mu_j^k, j \in J_k; \lambda_i^k$ (für Gleichungsrestriktionen)
- Wir ergänzen diese durch $\mu_j^k = 0, j \notin J_k$

$$\begin{array}{ll} \text{Insgesamt: } \lambda^k \in \mathbb{R}^m & (m \text{ Gln.}) \\ \mu^k \in \mathbb{R}^{p_k} & \text{ bzw. } \tilde{\mu}^k \in \mathbb{R}^p \quad (\text{durch Nullen aufgefüllt}) \end{array}$$

Voraussetzungen für das Verfahren:

- B_k hat immer vollen Rang (Lineare Unabhängigkeit des Systems $a^i, i = 1, \dots, m; g^j, j \in J_k$)
- Positive Definitheit von Q auf $\ker B_k$.

Durch diese Voraussetzungen ist (Q_k) eindeutig lösbar, und die Multiplikatoren λ^k, μ^k sind eindeutig bestimmt.

Zulässige Menge von (Q_k) :

$$\Omega_k = \{d \in \mathbb{R}^n \mid Ad = 0, G_k d = 0\} \subset L(\Omega, x^k).$$

Damit ist jedes $d \in \Omega_k$ automatisch eine zulässige Richtung.

Der nächste Verfahrensschritt ergibt sich nun durch Auswertung der **notwendigen Optimalitätsbedingungen für (Q_k)** :

$$\nabla f(x^k + d^k) - A^T \lambda^k - G_k^T \mu^k = 0$$

d. h.

$$\boxed{Q(d^k + x^k) + q - A^T \lambda^k - G_k^T \mu^k = 0.} \quad (\dagger)$$

Aus diesem System ergeben sich λ^k, μ^k wegen linearer Unabhängigkeit eindeutig.

Nun Fallunterscheidung:

Fall 1 $\boxed{d^k = 0 \quad \text{und} \quad \mu^k \geq 0.}$

Dann gilt $\nabla f(x^k) - A^T \lambda^k - G_k^T \mu^k = 0$, und x^k erfüllt die Kuhn-Tucker-Bedingungen. Wegen Konvexität sind diese hinreichend für Optimalität:

x^k ist die Lösung der Aufgabe (QU) : **Stop**

Fall 2 $\boxed{d^k = 0 \quad \text{aber} \quad \mu^k \not\geq 0.}$

Hier gibt es *mindestens ein* $j \in J_k$ mit $\mu_j^k < 0$. Wie in unserem Illustrationsbeispiel sollte dann eine Nebenbedingung deaktiviert werden. Am lohnendsten: Wähle ein $j \in J_k$ mit

$$\mu_j^k = \min\{\mu_i^k, i \in J_k\}$$

Die *Deaktivierung* erfolgt durch Neufestsetzung der aktiven Menge:

$$\boxed{\tilde{J}_k := J_k \setminus \{j\} .}$$

Nun wird entsprechend (\tilde{Q}_k) aufgestellt und gelöst. Das Verfahren wird dabei sichern, dass die deaktivierte Restriktion nicht verletzt (d. h. in der falschen Richtung verlassen) wird.

Ergebnis: $\tilde{d}^k, \tilde{\mu}^k, \tilde{\lambda}^k$

(Damit Q positiv definit auf $\ker \tilde{B}_k$ bleibt, obwohl man ja B_k nicht kennt, wird der Einfachheit halber positive Definitheit auf dem größten Unterraum $\ker A$ vorausgesetzt.)

Wir zeigen: $\boxed{\tilde{d}^k \neq 0} .$

Denn: Es galt $d^k = 0$

$$\Rightarrow Qx^k + q - A^T \lambda^k - G_k^T \mu^k = 0$$

$$Q\tilde{d}^k + Qx^k + q - A^T \tilde{\lambda}^k - \tilde{G}_k^T \tilde{\mu}^k = 0$$

Wäre $\tilde{d}^k = 0$, so

$$\begin{aligned} A^T \lambda^k + G_k^T \mu^k &= A^T \tilde{\lambda}^k + \tilde{G}_k^T \tilde{\mu}^k \\ \Rightarrow \mu_j g^j &= \text{Linearkombination der anderen auftretenden } a^i, g^i \\ \Rightarrow \text{wegen } \mu_j \neq 0 &\text{ gilt Gleiches für } g^j \text{ Widerspruch.} \end{aligned}$$

Somit haben wir noch zu diskutieren:

Fall 3 $d^k \neq 0$. Dann ist d^k **Abstiegsrichtung**.

Beweis: Aus den notwendigen Bedingungen (†) für x^k folgt

$$\begin{aligned} \nabla f(x^k) &= Q^k x^k + q = -Q d^k + A^T \lambda^k + G_k^T \mu^k \\ &= -Q d^k + B_k^T \begin{pmatrix} \lambda^k \\ \mu^k \end{pmatrix} / \cdot d^k \\ \Rightarrow \langle \nabla f(x^k), d^k \rangle &= -\underbrace{\langle d^k, Q d^k \rangle}_{<0 \text{ wegen Definitheit}} + \left\langle \begin{pmatrix} \lambda^k \\ \mu^k \end{pmatrix}, \underbrace{B_k d^k}_{=0} \right\rangle \\ &< 0. \end{aligned}$$

Bemerkung. Im Fall 2 muss noch gesichert werden, dass die Richtung \tilde{d}^k zulässige Richtung ist, d. h. dass auch für die eine deaktivierte Restriktion Nr. j gilt

$$\langle g^j, \tilde{d}^k \rangle \geq 0.$$

Das gilt auch wirklich, denn: Wegen $d^k = 0$ und (†) gilt ausgeschrieben:

$$0 = \nabla f(x^k) - \sum_{i=1}^m a^i \lambda_i^k - \sum_{\substack{i \neq j \\ i \in J_k}} g^i \mu_i^k - \mu_j^k g^j / \cdot \tilde{d}^k.$$

Wir multiplizieren skalar mit \tilde{d}^k durch. Wir wissen aus Fall 3, dass \tilde{d}^k eine Abstiegsrichtung ist. Außerdem war $A \tilde{d}^k = 0$ und $\langle g^i, \tilde{d}^k \rangle = 0$, $i \in \tilde{J}_k$ gefordert. Daher

$$\underbrace{\langle \nabla f(x^k), \tilde{d}^k \rangle}_{\substack{<0 \\ (\text{Abstieg})}} - \underbrace{\mu_j^k}_{\substack{<0 \\ (\text{Fall 2})}} \langle g^j, \tilde{d}^k \rangle = 0$$

$$\Rightarrow \boxed{\langle g^j, \tilde{d}^k \rangle > 0}.$$

Das entspricht auch der Einsicht aus unserem geometrischen Beispiel, \tilde{d}^k zeigt aus der Sicht der j -ten Restriktion in das Innere von Ω .

Zusammengefasst ergibt sich folgendes

Verfahren 6.1.1 (Aktive-Mengen-Strategie für (QU))

1. Berechne Startpunkt $x^0 \in \Omega$, setze $k := 0$.

2. Stelle (Q_k) auf und bestimme daraus Richtung d^k , Multiplikatoren λ^k, μ^k .
3. Wenn $d^k = 0$ und $\mu^k \geq 0$: **Stop**; x^k ist die gesuchte Lösung.
4. Wenn $d^k = 0$ und $\mu^k \not\geq 0$, dann führe einen Inaktivierungsschritt durch:
 - Bestimme $\mu_j^k = \min\{\mu_i^k, i \in J_k\}$
 - $J_k := J_k \setminus \{j\}$
 - streiche in G_k die zu j gehörige Zeile
 - Löse das entsprechende neue Problem (Q_k) . Das Ergebnis ist auf jeden Fall $\tilde{d}^k := d^k \neq 0$
5. Es gilt jetzt $d^k \neq 0$. Berechne eine Schrittweite σ_k (Erklärung unten) und setze

$$x^{k+1} = x^k + \sigma_k d^k.$$

Zunächst müssen wir noch die Wahl der Schrittweite σ_k klären.

Schrittweitenbestimmung

Wir gehen aus von einer zulässigen Abstiegsrichtung d^k . Die neue Lösung ist dann

$$x^{k+1} = x^k + \tau d^k$$

mit einem gewissen $\tau > 0$. Wie sollte τ gewählt werden?

• Maximaler Abstieg

$$\begin{aligned}
 f(x^k + t d^k) &= f(x^k) + \underbrace{t \nabla f(x^k) d^k}_{= -t \langle d^k, Q d^k \rangle} + \frac{1}{2} t^2 \langle d^k, Q d^k \rangle \\
 &= -t \langle d^k, Q d^k \rangle \quad \text{siehe Endergebnis von Fall 3.} \\
 &= f(x^k) + \underbrace{\left(\frac{1}{2} t^2 - t \right) \langle d^k, Q d^k \rangle}_{\text{minimal bei } t = 1}
 \end{aligned}$$

\Rightarrow Aus Sicht des maximalen Abstiegs wäre $\tau = 1$ zu setzen.

• Zulässigkeit von x^{k+1}

Für die letzte Iterierte x^k galt

$$\begin{aligned}
 \langle a^i, x^k \rangle &= b_i \quad i = 1, \dots, m \\
 \langle g^i, x^k \rangle &= r_i \quad i \in J_k \quad (\text{aktive Ungln.}) \\
 \langle g^i, x^k \rangle &> r_i \quad i \notin J_k \quad (\text{inaktive Ungln.}).
 \end{aligned}$$

Durch die Wahl von d^k ist gesichert für alle $t \geq 0$

$$\begin{aligned}
 \langle a^i, x^k + t d^k \rangle &= b_i \quad \text{denn } \langle a^i, d^k \rangle = 0 \\
 \langle g^i, x^k + t d^k \rangle &= r_i \quad \forall i \in J_k \setminus \{j\}, \quad \text{aus gleichem Grund} \\
 \langle g^j, x^k + t d^k \rangle &> r_j \quad \text{denn bei der inaktivierten Restriktion gilt } \langle g^j, d^k \rangle > 0.
 \end{aligned}$$

Damit brauchen wir uns nur um die inaktiven Restriktionen zu kümmern! Offenbar gibt es nur dann eine Schranke für t , wenn es mindestens ein $j \notin J_k$ gibt mit

$$\langle g^j, d^k \rangle < 0.$$

Es muss dann gefordert werden

$$\langle g^j, d^k \rangle + t \langle g^j, d^k \rangle \geq r_j, \quad \text{Maximum von } t : \text{ bei Gleichheit}$$

also für dieses spezielle j maximal

$$t = \frac{r_j - \langle g^j, x^k \rangle}{\langle g^j, d^k \rangle}.$$

\Rightarrow Maximal zulässige Schrittweite:

$$\begin{aligned} \tau_k &= \min_{j \in I_k} \left\{ \frac{r_j - \langle g^j, x^k \rangle}{\langle g^j, d^k \rangle} \right\} \quad \text{falls } I_k \neq \emptyset \\ &\quad \text{wobei } I_k = \{j / \langle g^j, d^k \rangle < 0\}. \\ \tau_k &:= \infty \quad \text{falls } I_k = \emptyset \quad (\text{keine Beschränkung nötig}). \end{aligned}$$

Insgesamt: $\sigma_k = \min(1, \tau_k)$.

Damit ist das Verfahren vollständig beschrieben. Es gilt

Satz 6.1.1 *Es sei Q positiv definit auf $\ker A$ und für alle $x \in \Omega$ habe die Matrix $B(x) = \begin{pmatrix} A \\ G(x) \end{pmatrix}$ vollen Rang. Dann berechnet das Verfahren die Lösung des Problems (QU) in endlich vielen Schritten.*

Beweis: Die Durchführbarkeit des Verfahrens haben wir bereits diskutiert.

Im Verlauf des Verfahrens sind jeweils quadratische Optimierungsprobleme der Form

$$\min f(x) \quad \text{bei } Ax = b, \quad \langle g^j, x \rangle = r_i \quad \forall i \in J \quad (1)$$

von Bedeutung, wobei J eine beliebige Teilmenge von $\{1, \dots, p\}$ ist, die für mögliche aktive Ungleichungsrestriktionen steht. Wir lösen im Verfahren nicht direkt (1), aber es treten Optimalitätssysteme von (1) auf, nämlich die Gleichungen

$$Qx + q - A^T \lambda - G(x)^T \mu = 0, \quad (2)$$

wobei x jeweils für die eindeutig bestimmte Lösung von (1) steht. Es gibt nur endlich viele mögliche Teilmengen J , damit nur endlich viele verschiedene Probleme (1), also auch nur endlich viele solche Lösungen x und damit nur endlich viele Möglichkeiten, solche Systeme (2) zu erzeugen.

Wir nehmen nun an, dass das Verfahren im Schritt k noch nicht zu Ende ist, d. h., wir führen den **Schritt 5** mit $d^k \neq 0$ durch. Dann:

(i) $I_k = \emptyset$. Hier gilt $\sigma_k = 1$, also

$$x^{k+1} = x^k + d^k$$

und wir wissen dann wegen (†)

$$Q(\underbrace{x^k + d^k}_{x^{k+1}}) + q - A^T \lambda^k - G_k^T \mu^k = 0,$$

so dass $x^{k+1} =: x$ das System (2) erfüllt (λ^k und μ^k ergeben sich daraus eindeutig). Das heißt nicht, dass x^{k+1} optimal ist, denn $\mu^k \not\geq 0$ kann eintreten. Damit ist x^{k+1} eine der endlich vielen Lösungen von (1).

(ii) $I_k \neq \emptyset$. Hier betrachten wir 2 “Unterfälle”:

- $\tau_k \geq 1$: Dann gilt $\sigma_k = \min(1, \tau_k) = 1$. Fall wie eben – d. h. x^{k+1} ist eine der Lösungen von (1)
- $\tau_k < 1$: Alle aktiven Restriktionen bleiben aktiv, aber es kommt mindestens eine neue hinzu, so dass die Kardinalzahl der aktiven Restriktionen wächst. Voraussetzung war: Das System der $\{a^i\}_{i=1,\dots,m} \cup \{g^j\}$, $j \in J_k$ ist stets linear unabhängig. Es können also höchstens $n - m$ solcher Zuwachsfälle auftreten (am Anfang waren es mindestens m linear unabhängige Vektoren, und jedes Mal kommt ein neuer hinzu).
 \Rightarrow nach maximal $n - m$ Iterationen gilt $I_{k+i} = \emptyset$ oder $\tau_{k+i} \geq 1 \Rightarrow$ neue Lösung von (2).

Außerdem ist d^k eine Abstiegsrichtung, also gilt auf alle Fälle

$$f(x^{k+j}) < f(x^k).$$

Damit sind die auftretenden x^{k+j} alle verschieden, und wegen Endlichkeit der Möglichkeiten für (1) bzw. (2) muss das Verfahren damit nach endlich vielen Schritten stoppen.

□

6.2 Gleichungsnebenbedingungen nichtquadratischer Zielfunktion

Hier besteht die gleiche Grundidee wie bei quadratischer Zielfunktion: Man “eliminiert” die Gleichungsrestriktion mit Hilfe einer Nullraummatrix. Wir betrachten die Aufgabe

$$\boxed{\begin{array}{l} \min f(x) \\ Ax = b \end{array}} \quad (\text{PLG})$$

mit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, jetzt nicht mehr notwendig quadratisch. Also $\min_{x \in \Omega} f(x)$ mit $\Omega = \{x \in \mathbb{R}^n / Ax = b\}$. Es sei wieder $w \in \Omega$ eine spezielle Lösung von $Ax = b$ und $Z : \mathbb{R}^l \rightarrow \ker A$ eine Nullmatrix. Dann wird die unrestringierte Aufgabe

$$\boxed{\min_{z \in \mathbb{R}^l} F(z) := f(w + Zz)} \quad (6.1)$$

gelöst. Die Bestimmung einer Nullraummatrix hängt nicht von f ab, nur von Ω , erfolgt also genauso, wie bereits beschrieben (QR-Zerlegung etc.).

Die freie Optimierungsaufgabe (6.1) kann nun (bei entsprechender Glattheit von f) mit jedem Verfahren der unrestringierten Optimierung behandelt werden. Damit könnten wir diesen Abschnitt abschließen, wenn es nicht noch einige interessante Nebenaspekte gäbe! Diese bestehen in der Parallelität der Minimierung von f und der von F .

Nehmen wir an, wir untersuchen ein normales Abstiegsverfahren.

$$\begin{aligned} \text{Für } f : x^{k+1} &= x^k + \sigma_k d^k \\ F : z^{k+1} &= z^k + \sigma_k v^k \end{aligned}$$

Ist z.B. v^k eine Abstiegsrichtung für F in z^k , dann gilt für das Bild $d^k := Zv^k$

$$\begin{aligned} \nabla f(x^k)^T d^k &= \nabla f(x^k)^T Z v^k \\ &= (Z^T \nabla f(x^k))^T v^k \\ &= \nabla F(z^k)^T v^k < 0, \end{aligned}$$

damit ist auch d^k eine Abstiegsrichtung, aber für f . Ferner

$$x^{k+1} = w + Z z^{k+1} = \underbrace{w + Z z^k}_{x^k} + \underbrace{\sigma_k Z v^k}_{d^k} = x^k + \sigma_k d^k.$$

Folgerung: Man kann das Verfahren im Raum der x -Variablen durchführen und muss die z -Variablen eigentlich gar nicht verwenden.

Verfahren 6.2.1 (Reduziertes Abstiegsverfahren)

1. Berechne $x^0 \in \Omega$, Nullraum-Matrix Z , $k := 0$.
2. Wenn $\underbrace{Z^T \nabla f(x^k)}_{\text{reduzierter Gradient}} = 0$: **Stop**.
3. Ansonsten berechne Abstiegsrichtung $\boxed{d^k := Z v^k}$, effiziente Schrittweite σ_k und

$$x^{k+1} := x^k + \sigma_k d^k$$

$k := k + 1$, goto 2.

Man braucht dazu Z, v^k sowie die Korrespondenzen

$$\begin{array}{cccc} F(z^k), & F(z^k + \sigma_k v^k), & \nabla F(z^k), & \nabla F(z^k + \sigma_k v^k), \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ f(x^k) & f(x^k + \sigma_k d^k) & Z^T \nabla f(x^k) & Z^T \nabla f(x^k + \sigma_k d^k) \end{array}$$

Noch sieht es so aus, als würde man bei der Berechnung von $d^k = Zv^k$ zumindest den Vektor v^k brauchen und nicht nur im x -Raum arbeiten können.

Bei konkreten Verfahren sieht der aber anders aus!

Beispiel 6.2.1

Reduziertes Gradientenverfahren:

$$\begin{aligned} v^k &:= -\nabla F(z^k) = -Z^T \nabla f(x^k) \\ \Rightarrow \quad &\boxed{d^k = Zv^k = -ZZ^T \nabla f(x^k)} \end{aligned}$$

Spezialfall: $Z = P$, Projektionsmatrix auf $\ker A$, \Rightarrow

Projiziertes Gradientenverfahren: $Z = P$

$$\begin{aligned} d^k &= -ZZ^T \nabla f(x^k) = -ZZ \nabla f(x^k) \\ &\boxed{d^k = -Z \nabla f(x^k)} \end{aligned}$$

Variable-Metrik-Verfahren (reduziert):

Folge $\{A^{(k)}\}$ positiv definierter Matrizen;

$$\begin{aligned} v^k &= -(A^{(k)})^{-1} \nabla F(z^k) = -(A^{(k)})^{-1} Z^T \nabla f(x^k) \\ \Rightarrow \quad &\boxed{d^k = Zv^k = -Z(A^{(k)})^{-1} Z^T \nabla f(x^k)} \end{aligned}$$

Speziell: reduziertes Newton-Verfahren:

$$A^{(k)} := F''(z^k) = \underbrace{Z^T f''(x^k) Z}_{\text{reduzierte Hesse-Matrix}}$$

$$\Rightarrow \quad \boxed{d^k = -Z(Z^T f''(x^k) Z)^{-1} Z^T \nabla f(x^k)}.$$

Analoge Betrachtungen gibt es für das reduzierte BFGS-Verfahren.

Eine schöne Anwendung der nichtlinearen Optimierung mit linearen Gleichungsrestriktionen:

Nichtlineare Regression mit Splines 3. Ordnung

Messwerte $(\xi_i, \eta_i), i = 1, \dots, m$

Ansatz $\eta(\xi) = g(x, \xi)$

Gesucht Vektor x und vorher aber: geeigneter Ansatz g .

Idee: Splines mit Koeffizienten x .

Sei $\xi_1 < \xi_2 < \dots < \xi_m$.

Wir überdecken das Intervall $[\xi_1, \xi_m]$ durch Knotenpunkte

$$\tau_0 < \tau_1 < \dots < \tau_N,$$

$$\tau_0 \leq \xi_1, \xi_m \leq \tau_N.$$

Forderungen:

- Auf $[\tau_i, \tau_{i+1}]$ ist $g =: g_i(x, \xi)$ Polynom dritten Grades in ξ
- $g(x, \cdot) \in C^2[\tau_0, \tau_N]$
 $\Rightarrow g, g', g''$ müssen in den Knotenpunkten stetig sein.

Man definiert auf $[\tau_i, \tau_{i+1}]$

$$g_i(x, \tau) = \frac{1}{\tau_{i+1} - \tau_i} (\gamma_{i+1}(\tau - \tau_i)^3 + \gamma_i(\tau_{i+1} - \tau)^3) + \beta_i(\tau - \tau_i) + \alpha_i$$

$$\gamma_0 = \gamma_N = 0$$

$$\gamma_i = g_i''(x, \tau_i)/6 \quad \text{“Momente”}$$

$$i = 1, \dots, N-1.$$

Durch diese Wahl ist g'' automatisch stetig. Das heißt nicht, dass auch g und g' stetig sein müssen. Diese Forderung ergibt zusätzliche Bedingungen an

$$x = (\alpha_0, \dots, \alpha_{N-1}, \beta_0, \dots, \beta_{N-1}, \dots, \gamma_1, \dots, \gamma_{N-1}).$$

Nämlich: Stetigkeit von g' : $\Delta\tau_i := \tau_{i+1} - \tau_i$

$$g'_i(x, \tau_i) = g'_{i-1}(x, \tau_i)$$

$$3\gamma_i\Delta\tau_i + \beta_i = \beta_{i-1} + 3\gamma_i\Delta\tau_{i-1}$$

$$\Rightarrow \beta_i = \beta_{i-1} + 3\gamma_i(\Delta\tau_{i-1} - \Delta\tau_i)$$

\Rightarrow Eigentlich ist nur β_0 wirklich frei.

Stetigkeit von g :

$$g_i(x, \tau_i) = g_{i-1}(x, \tau_i)$$

$$\gamma_i(\Delta\tau_i)^2 + \alpha_i = \gamma_i(\Delta\tau_{i-1})^2 + \alpha_{i-1} + \beta_{i-1}\Delta\tau_{i-1}$$

$$\Rightarrow \alpha_i = \alpha_{i-1} + \gamma_i((\Delta\tau_{i-1})^2 - (\Delta\tau_i)^2) + \beta_{i-1}\Delta\tau_{i-1}.$$

Auch hier ist eigentlich wieder nur α_0 frei. Insgesamt ergibt sich die Aufgabe

$$\begin{aligned} \min f(x) &= \sum_{i=1}^m (\eta_i - g(x, \xi_i))^2 \\ \text{bei } \beta_i &= \beta_{i-1} + 3\gamma_i(\Delta\tau_i - \Delta\tau_{i-1}) \\ \alpha_i &= \alpha_{i-1} + \beta_{i-1}\Delta\tau_{i-1} + \gamma_i(\Delta\tau_{i-1}^2 - \Delta\tau_i^2) \\ i &= 1, \dots, N-1. \end{aligned}$$

Diese Aufgabe kann man reduzieren auf eine *freie Optimierungsaufgabe* in den Variablen:

$$(\alpha_0, \beta_0, \gamma_1, \dots, \gamma_{N-1})^T.$$

6.3 Ungleichungsnebenbedingungen – nichtquadratische Zielfunktionen

Jetzt ist eine Aufgabe der Bauart

$$\begin{aligned} \min f(x) \\ Ax = b, \quad Gx \geq r \end{aligned}$$

gegeben.

Grundidee: Taylorapproximation von f bis zur zweiten Ordnung \leadsto quadratische Zielfunktion, Lösung der Aufgabe mit der vorn eingeführten Methode und dann neue Approximation von f : *SQP-Verfahren*.

Zur Einstimmung ein kurzer Exkurs zur einfachsten *Idee des SQP-Verfahrens*:
Wir betrachten dazu parallel zwei simple Optimierungsaufgaben

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \min_{x \in \Omega} f(x).$$

Ω : konvexe Menge.

Die notwendigen Bedingungen 1. Ordnung für eine Lösung x^* lauten:

$$\nabla f(x^*) = 0 \qquad \langle \nabla f(x^*), x - x^* \rangle \geq 0 \quad \forall x \in \Omega.$$

Die linke Beziehung ist ein lineares Gleichungssystem. Wir wissen, wie wir so etwas lösen können – z.B. mit dem Newton-Verfahren. Rechts steht eine Variationsungleichung – da haben wir erst einmal keine Idee. Schreiben wir deshalb zunächst das Newton-Verfahren links auf:

$$(*) \qquad \nabla f(x^k) + f''(x^k)(x - x^k) = 0.$$

Die Lösung ist $x = x^{k+1}$. Rechts haben wir noch keine Entsprechung. Offenbar ist aber $(*)$ gerade die notwendige Bedingung 1. Ordnung für die Optimierungsaufgabe

$$(**) \qquad \min_{x \in \mathbb{R}^n} \frac{1}{2} \langle x - x^k, f''(x^k)(x - x^k) \rangle + \langle \nabla f(x^k), x - x^k \rangle.$$

Es ist also egal, ob wir die Aufgabe (**) oder die Gleichung (*) lösen. Während aber (*) keine Entsprechung für die beschränkte Optimierungsaufgabe hat, ist das bei (**) kein Problem: Wir nehmen einfach nur die Beschränkung $x \in \Omega$ mit hinzu:

$$\min_{x \in \Omega} \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, f''(x^k)(x - x^k) \rangle.$$

Lösung: $x = x^{n+1}$.

Und genau das macht man für (PLU)! Hier ist

$$\Omega = \{x \in \mathbb{R}^n \mid Ax = b, Gx \geq r\}$$

konvex. Wir iterieren wie folgt:

x^k sei berechnet. Man stellt dann auf:

$$\begin{array}{ll} \min & \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, f''(x^k)(x - x^k) \rangle \\ \text{bei} & Ax = b, Gx \geq r \end{array} \quad (QP_k)$$

Lösung: x^{k+1} .

Dann $k := k + 1$, und wir iterieren neu.

Ganz umsonst gibt es aber auch hier die Konvergenz nicht, wie auch beim Newton-Verfahren für (*): Dort muss $f''(x^k)$ stets invertierbar sein, was man durch $f \in C^2$ und $f''(x^*)$ regulär sichert. Da es aber um ein Minimum geht, muss $f''(x^*)$ noch dazu positiv semidefinit sein. Zusammen mit der Regularität muss somit $f''(x^*)$ positiv definit sein! Genau das aber hilft in (QP_K) auch: Wenn $\det f''(x^*) > 0$, so auch $\det f''(x^k)$ für x^k nahe bei x^* , und damit hat (QP_k) genau eine Lösung!

Bei der numerischen Umsetzung schreibt man das Verfahren ein wenig anders auf: Man setzt

$$d = x - x^k \quad \Leftrightarrow \quad x = x^k + d.$$

Wegen $Ax^k = b$ muss gelten $Ad = 0$, und insgesamt

$$(QP_k) \quad \Leftrightarrow \quad \begin{array}{ll} \min & \langle \nabla f(x^k), d \rangle + \frac{1}{2} \langle d, f''(x^k)d \rangle \\ \text{bei} & Ad = 0, Gx^k + Gd \geq r. \end{array}$$

Diese Aufgabe dient also der Berechnung einer Suchrichtung d . Man kann nun "voll" in die Richtung d gehen, d. h. $x^{k+1} = x^k + d^k$ setzen (Newton-Verfahren) oder eine Schrittweitensteuerung verwenden.

Bemerkung: Anstelle von $f''(x^k)$ kann man wie beim Variable-Metrik-Verfahren auch entsprechende Matrizen $A^{(k)}$ nutzen – siehe z.B. [1]. Wir bleiben bei f'' .

Voraussetzungen für die Durchführbarkeit des Verfahrens

- $f''(x^k)$ soll jeweils positiv definit auf $\ker A$ sein
 \Rightarrow Existenz genau einer Lösung von (QP_k)
- $B(x) = \begin{pmatrix} A \\ G(x) \end{pmatrix}$ habe immer vollen Rang

\Rightarrow Multiplikatoren λ, μ sind eindeutig bestimmt.

Optimalitätsbedingung für (QP_k) :

$$f''(x^k)d^k + \nabla f(x^k) - A^T \lambda^{k+1} - G^T \mu^{k+1} = 0 \quad (6.2)$$

$$\mu^{k+1} \geq 0, \quad \langle \mu^{k+1}, Gx^k + Gd^k - r \rangle = 0. \quad (6.3)$$

Bei der Lösung von (QP_k) gibt es nun zwei Fälle:

Fall 1 $d^k = 0$.

Keine Änderung, x^k müsste optimal gewesen sein. In der Tat, (6.2–6.3) ergeben dann

$$\begin{aligned} \nabla f(x^k) - A^T \lambda^{k+1} - G^T \mu^{k+1} &= 0, \quad \mu^{k+1} \geq 0 \\ \langle \mu^{k+1}, Gx^k - r \rangle &= 0. \end{aligned}$$

$\Rightarrow x^k$ erfüllt die Optimalitätsbedingungen \Rightarrow STOP
(Optimalität folgt aus den hinreichenden Bedingungen).

Fall 2 $d^k \neq 0$.

Dann ist d^k Abstiegsrichtung, denn

$$\begin{aligned} \nabla f(x^k) &= -f''(x^k)d^k + A^T \lambda^{k+1} + G^T \mu^{k+1} \mid \cdot d^k \\ \langle \nabla f, d^k \rangle &= -\underbrace{\langle d^k, f'' d^k \rangle}_{>0} + \underbrace{\langle A d^k, \lambda^{k+1} \rangle}_{=0} + \underbrace{\langle G d^k, \mu^{k+1} \rangle}_{\leq 0} \\ &< 0 \quad \Rightarrow \quad \text{Abstiegsrichtung} \quad \text{s.unten.} \end{aligned}$$

Zur Diskussion von $\langle Gd^k, \mu^{k+1} \rangle$: Die Beschränkungen lauten ausgeschrieben

$$\langle g^i, x^k \rangle + \langle g^i, d^k \rangle \geq r_i.$$

Für die inaktiven Indizes $i \notin J(d^{(k)})$ gilt $\mu_i^{k+1} = 0$. Für die aktiven gilt $\mu_i^{k+1} \geq 0$ sowie

$$\begin{aligned} \langle g^i, d^k \rangle &= \underbrace{r_i - \langle g^i, x^k \rangle}_{\leq 0, \text{ weil } x^k \text{ zulässig war}} \\ &\leq 0, \end{aligned}$$

also $\langle g^i, d^k \rangle \leq 0$. Insgesamt folgt daraus leicht $\langle Gd^k, \mu^{k+1} \rangle \leq 0$.

Schrittweitenbestimmung

Da d^k zulässig ist, kann mindestens $x^{k+1} = x^k + 1 \cdot d^k$ gewählt werden. Die maximale Schrittweite ist daher $\tau_k \geq 1$. Gängig: $\sigma_k = 1$ (reines SQP) oder aber: Schrittweitensteuerung.

Wie wir gesehen haben, ist bei Wahl von $\sigma_k \equiv 1$ das SQP-Verfahren eine Verallgemeinerung des Newton-Verfahrens, und es wird deshalb auch so genannt. Unter natürlichen Voraussetzungen ist es wie dieses lokal quadratisch konvergent.

Voraussetzungen: (\tilde{x} sei lokales Minimum von (PLU)).

- (i) $f \in C^2$ in einer Kugel $B(\tilde{x}, \delta)$ um \tilde{x}

(ii) f'' ist auf $B(\tilde{x}, \delta)$ Lipschitz, d. h.

$$\|f''(x) - f''(y)\| \leq L\|x - y\| \quad \forall x, y \in B(\tilde{x}, \delta)$$

(iii) $B(\tilde{x})$ hat vollen Rang

(iv) Positive Definitheit:

$$d^T f''(\tilde{x})d \geq \alpha \|d\|^2 \quad \forall d : Ad = 0, G(\tilde{x})d = 0$$

(Hinreichende Optimalitätsbedingung 2. Ordnung)

(v) Gilt $\langle g^i, \tilde{x} \rangle = r_i$, so gilt auch $\tilde{\mu}_i > 0$ für den entsprechenden Multiplikator (Bedingung der *strengen Komplementarität*)

Bemerkung dazu: Immer gilt $\tilde{\mu}_i > 0 \Rightarrow \langle g^i, \tilde{x} \rangle = r_i$. Strenge Komplementarität bedeutet die Umkehrung.

Satz 6.3.1 *Unter den Voraussetzungen (i)-(v) konvergiert unser SQP-Verfahren lokal quadratisch, d. h.*

$$\begin{aligned} & \|x^{k+1} - \tilde{x}\| + \|\lambda^{k+1} - \tilde{\lambda}\| + \|\mu^{k+1} - \tilde{\mu}\| \\ & \leq c(\|x^k - \tilde{x}\|^2 + \|\lambda^k - \tilde{\lambda}\|^2 + \|\mu^k - \tilde{\mu}\|^2) \end{aligned}$$

7 Probleme mit nichtlinearen Restriktionen-Verfahren

7.1 Das Lagrange-Newton-Verfahren

Wir betrachten wieder die Aufgabe

$$\begin{aligned} \min f(x) \\ h(x) = 0 \\ g(x) \geq 0 \end{aligned} \quad (\text{PNU})$$

wie im letzten Kapitel. Dabei setzen wir jetzt generell voraus:

- $f, g, h \in C^2$
- \tilde{x} ist eine *reguläre* lokale Lösung
- $\tilde{\lambda}, \tilde{\mu}$ sind zugehörige Lagrangesche Multiplikatoren.

Wir finden $\tilde{x}, \tilde{\lambda}, \tilde{\mu}$ aus den Kuhn-Tucker-Bedingungen, die unter anderem fordern:

$$\begin{aligned} \nabla f(\tilde{x}) - h'(\tilde{x})^T \tilde{\lambda} - g'(\tilde{x})^T \tilde{\mu} &= 0 \\ h(\tilde{x}) &= 0 \\ \tilde{g}(\tilde{x}) &= 0. \end{aligned}$$

Dabei stecken in \tilde{g} nur die aktiven Ungleichungen, d. h.

$$\tilde{g}(x) = (g_j(x))_{j \in J(\tilde{x})}.$$

Die nicht aktiven interessieren in einer Umgebung von \tilde{x} nicht, sie können lokal als Nebenbedingungen vernachlässigt werden. Durch Lösen dieses Systems sollen $\tilde{x}, \tilde{\lambda}, \tilde{\mu}$ bestimmbar sein. Um das System kurzfassen zu können, definieren wir

$$z = \begin{pmatrix} x \\ \lambda \\ \nu \end{pmatrix}, \quad F(z) = \begin{pmatrix} \nabla f(x) - h'(x)^T \lambda - \tilde{g}'(x)^T \nu \\ h(x) \\ \tilde{g}(x) \end{pmatrix}$$

mit $\nu = (\mu_j)_{j \in J(x)}$. $\tilde{\nu} := (\tilde{\mu}_j)_{j \in J(\tilde{x})}$. Dann gilt also mit $\tilde{z}^T = (\tilde{x}^T, \tilde{\lambda}^T, \tilde{\nu}^T)$

$$F(\tilde{z}) = 0.$$

Was liegt näher, als darauf das Newton-Verfahren anzuwenden, um \tilde{z} numerisch zu bestimmen. Als Vorabinformation muss man aber wissen, welche Indizes zur Menge $J(\tilde{x})$ gehören, also die aktiven Ungleichungen kennen! Für die Konvergenz des Verfahrens brauchen wir

- (8.1.2) $f, g, h \in C^{2,1}$ und $F'(\tilde{z})$ ist nichtsingulär. Die Matrix $F'(z)$ hat die Form

$$F'(z) = \begin{pmatrix} \mathcal{L}_{xx}(x, \lambda, \nu) & h'(x)^T & \tilde{g}'(x)^T \\ h'(x) & 0 & 0 \\ \tilde{g}'(x) & 0 & 0 \end{pmatrix}.$$

Diese Matrix ist vom Typ

$$\mathcal{A} = \begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix},$$

für die gilt: Ist Q positiv definit auf $\ker A$ und hat A vollen Rang, dann ist \mathcal{A} invertierbar. Hier gilt

$$Q = \mathcal{L}_{xx}, \quad A = \begin{pmatrix} h'(x) \\ g'(x) \end{pmatrix},$$

also ist $F'(\tilde{z})$ nichtsingulär, wenn

- (8.1.3) (LICQ) erfüllt ist, d.h., die Gradienten $\nabla h_i(\tilde{x})$, $\nabla g_j(\tilde{x})$ der aktiven Restriktionen sind linear unabhängig,
- (8.1.5) und eine hinreichende Optimalitätsbedingung 2. Ordnung erfüllt ist:

$$d^T \mathcal{L}_{xx}(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) d \geq \alpha \|d\|^2$$

für alle d mit $h'(\tilde{x})d = 0$ und $\tilde{g}'(\tilde{x})d = 0$.

Weiter brauchen wir später noch:

- (8.1.4) Es liegt *strenge Komplementarität* vor: $g_j(\tilde{x}) = 0 \Rightarrow \mu_j > 0$.

Damit sind die Voraussetzungen erfüllt, welche die lokal-quadratische Konvergenz des Newton-Verfahrens garantieren:

Ausgehend vom Startvektor $z^0 = (x^0, \lambda^0, \nu^0)$ berechnet man

$$z^{k+1} = z^k - F'(z^k)^{-1} F(z^k),$$

d. h., man löst das lineare System

$$F'(z^k)(z - z^k) = -F(z^k)$$

für z^{k+1} . Das sieht so aus:

$$\begin{pmatrix} \mathcal{L}_{xx}(x^k, \lambda^k, \nu^k) & -h'(x^k)^T & -\tilde{g}'(x^k)^T \\ h'(x^k) & 0 & 0 \\ \tilde{g}'(x^k) & 0 & 0 \end{pmatrix} \begin{pmatrix} x - x^k \\ \lambda - \lambda^k \\ \nu - \nu^k \end{pmatrix} = - \begin{pmatrix} \nabla f(x^k) - h'(x^k)^T \lambda^k - \tilde{g}'(x^k)^T \nu^k \\ h(x^k) \\ \tilde{g}(x^k) \end{pmatrix}.$$

Einiges hebt sich hier auf, so dass am Ende bleibt

$$\begin{aligned} \nabla f(x^k) + \mathcal{L}_{xx}(x^k, \lambda^k, \nu^k)(x - x^k) - h'(x^k)^T \lambda - \tilde{g}'(x^k)^T \nu &= 0 \\ h(x^k) + h'(x^k)(x - x^k) &= 0 \\ \tilde{g}(x^k) + \tilde{g}'(x^k)(x - x^k) &= 0. \end{aligned} \tag{7.4}$$

Im Prinzip sind das die notwendigen Optimalitätsbedingungen für eine Lösung der linear-quadratischen Aufgabe

$$\begin{aligned} & \min_x \nabla f(x^k)^T (x - x^k) + \frac{1}{2} (x - x^k)^T \mathcal{L}_{xx}(\dots)(x - x^k) & (Q1)_k \\ \text{bei} \quad & h(x^k) + h'(x^k)(x - x^k) = 0 \\ & g(x^k) + g'(x^k)(x - x^k) \geq 0, \end{aligned}$$

wenn wir zeigen können, dass die inaktiven Restriktionen $j \notin J(\tilde{x})$ auch hier nicht aktiv sind, also bedeutungslos bleiben. In gewissem Sinne sind also das Newton-Verfahren (7.4) und das SQP-Verfahren $(Q1)_k$ äquivalent.

Nun führen wir wieder die Richtung $d = x - x^k$ ein und erhalten so:

$$\begin{aligned} & \min_{d \in \mathbb{R}^n} \frac{1}{2} \langle d, \mathcal{L}_{xx}(x^k, \lambda^k, \mu^k) d \rangle + \langle \nabla f(x^k), d \rangle & (Q2)_k \\ \text{bei} \quad & h(x^k) + h'(x^k)d = 0 \\ & g(x^k) + g'(x^k)d \geq 0 \end{aligned}$$

und als Lösung $d^k = x^{k+1} - x^k$ mit neuen Multiplikatoren μ^{k+1}, λ^{k+1} . So erhalten wir

Verfahren 7.1.1 Lagrange-Newton-Verfahren für (PNU)

1. $k := 0, z^0 = (x^0, \lambda^0, \mu^0)$
2. Löse $(Q2)_k \rightarrow d^k; \lambda^{k+1}, \mu^{k+1}$
3. STOP bei $d^k = 0$
4. $x^{k+1} = x^k + d^k$, goto 2.

Bemerkungen:

1. Das Verfahren konvergiert lokal quadratisch! Das ist plausibel, weil es eigentlich unter entsprechenden Voraussetzungen äquivalent zum Newton-Verfahren ist (wenn man die aktiven Restriktionen kennt und strenge Komplementarität gilt. Strenge Komplementarität: Aktive Restriktionen bleiben in der entsprechenden Umgebung aktiv, weil die Multiplikatoren dort positiv bleiben).
2. Die Iterierten dieses Verfahrens sind in der Regel wegen der Nichtlinearität der Restriktionen unzulässig, d. h. $x^k \notin \Omega$. Außerdem haben wir nur lokale Konvergenz. Deshalb sind Modifikationen angebracht!

7.2 Sequentielle quadratische Optimierung

Wie erwähnt, ist das reine Lagrange-Newton-Verfahren nur lokal konvergent. Außerdem kann die Berechnung von \mathcal{L}_{xx} teuer werden. Deshalb modifiziert man das Verfahren

• **Verwendung von Approximationen $A^{(k)}$ für $\mathcal{L}_{xx}(x^k, \lambda^k, \mu^k)$**

Man verwendet, wie bei Variable-Metrik-Verfahren, symmetrische und positiv definite

Matrizen $A^{(k)}$ und bestimmt in der Richtungssuche:

$$\begin{aligned} & \min_d \langle \nabla f(x^k), d \rangle + \frac{1}{2} \langle d, A^{(k)} d \rangle & (\text{QP})_k \\ \text{bei} \quad & h(x^k) + h'(x^k)d = 0 \\ & g(x^k) + g'(x^k)d \geq 0. \end{aligned}$$

Wir wollen annehmen, dass die entsprechende zulässige Menge Ω_k nicht leer ist. Eine hinreichende Bedingung gibt [1, Satz 8.2.1] an (h_i affin-linear, lineare Unabhängigkeit der Gradienten, Slater-Typ-Bedingung).

Unter natürlichen Bedingungen kann dann die Existenz genau einer Lösung von $(\text{QP})_k$ sowie die gleichmäßige Beschränktheit der Folgen d^k, α^k, μ^k bewiesen werden [1, Satz 8.2.2].

• Schrittweitensteuerung

Wir gehen nicht den gesamten Newtonschritt, sondern setzen

$$x^{k+1} = x^k + \sigma d^k.$$

Man muss dabei beachten:

- a) Die erzeugten x^k müssen für (PNU) *nicht* zulässig sein.
- b) d^k ist in der Regel keine Abstiegsrichtung. (Es könnte theoretisch passieren, dass x^k einen zu guten Wert liefert, weil unzulässig. Dann könnten theoretisch ansteigende Funktionswerte eintreten.)

Deshalb benutzt man zur Schrittweitensteuerung sogenannte *Merit-Funktionen*

Definition 7.2.1 ϕ heißt *Merit-Funktion*, wenn gilt:

- Ist $\tilde{x} \in \Omega$ lokale Lösung von (PNU), dann ist \tilde{x} lokales (freies) Minimum von ϕ .
- Richtung d^k ist **Abstiegsrichtung** für ϕ .

Praktisch bewährt haben sich Merit-Funktionen des Typs

$$\phi = \phi(x; \beta, \gamma) := f(x) + \sum_{j=1}^p \beta_j g_j(x)_- + \sum_{i=1}^m \gamma_i |h_i(x)|$$

mit Konstanten $\beta_j \geq 0, \gamma_i \geq 0$. ϕ ist nicht differenzierbar!

$$g_j(x)_- := -\min\{0, g_j(x)\} = -\frac{g_j(x) - |g_j(x)|}{2}.$$

Wenn $x \in \Omega$, so gilt $g_j(x) \geq 0$, also $g_j(x)_- = 0$ sowie $h_i(x) = 0$, also $|h_i(x)| = 0$. Damit

$$x \in \Omega \Rightarrow f(x) = \phi(x; \beta, \gamma).$$

Für $x \notin \Omega$ gilt $\phi > f$. In diesem Sinne sind $\Sigma\beta_j(g_j)_-$ und $\Sigma\gamma_i|h_i|$ Strafterme, welche eine Verletzung der Nebenbedingungen bestrafen:

$$\underbrace{\sum_{i=1}^m \gamma_i |h_i(x)| + \sum_{j=1}^p \beta_j g_j(x)}_{\text{Penalty-Term}} .$$

γ_i, β_j Penalty-Parameter

Die Funktion ϕ heißt (*exakte*) *Penalty-Funktion*. Es gilt der wichtige

Satz 7.2.1 *Unter den Bedingungen 8.1.2–8.1.5 gilt: Ist \tilde{x} lokales Minimum von (PNU) und*

$$\beta_j > \mu_j, \quad \gamma_i > |\lambda_i|$$

$j = 1, \dots, p, i = 1, \dots, m$, dann ist \tilde{x} striktes lokales Minimum von $\phi(i; \beta, \gamma)$. Hier sind λ, μ die Lagrangeschen Multiplikatoren von \tilde{x} .

Durch relativ aufwendige Abschätzungen kann man letztlich folgendes zeigen:

Man gibt $\varepsilon > 0$ und $\delta \in (0, 1)$ vor. Sind die Parameter β_j und γ_i hinreichend groß gewählt, dann gilt

$$\beta_j \geq \mu_j^{(k+1)} + \varepsilon, \quad \gamma_i \geq |\lambda_i^{(k+1)}| + \varepsilon,$$

und für hinreichend kleines $\sigma > 0$ erhält man

$$\phi(x^k + \sigma d^k, \beta, \gamma) \leq \phi(x^k; \beta, \gamma) - \sigma \delta [\langle d^k, A^{(k)} d^k \rangle + \varepsilon \|g(x^k)_-\|_1 + \varepsilon \|h(x^k)\|_1],$$

d. h., die Merit-Funktion kann wirklich verkleinert werden. Außerdem gilt

$$\begin{aligned} |h_i(x^k + \sigma d^k)| &\leq (1 - \delta\sigma) |h_i(x^k)| & i = 1, \dots, m \\ g_j(x^k + \sigma d^k) &\geq (1 - \delta\sigma) g_j(x^k) & j = 1, \dots, p, \end{aligned}$$

d. h., die Unzulässigkeit wird in jedem Schritt geringer. Darauf basiert eine Grundversion des SQP-Verfahrens, auf deren ausführliche Darstellung wir verzichten. Wir verweisen auf [1, Abschnitt 8.2.3].

8 Penalty, Barrier und Augmentierte Lagrange-Methoden

8.1 Die quadratische Penalty-Methode

Idee: Ersetze Optimierungsprobleme durch eine Penalty-Funktion, bestehend aus

- ursprünglichem Zielfunktional
- zusätzlichem Term für jede Restriktion, der positiv ist, falls die Restriktion verletzt ist und sonst Null.

Zunächst betrachten wir nur Gleichheitsrestriktionen, d.h.

$$\begin{array}{ll} \min f(x) & f: \mathbb{R}^n \rightarrow \mathbb{R} \\ \text{bei } h(x) = 0 & h: \mathbb{R}^n \rightarrow \mathbb{R}^m. \end{array} \quad (\text{PNG})$$

Betrachte quadratische Penalty-Funktionen

$$Q(x, \mu) = f(x) + \frac{1}{2\mu} \sum_{i=1}^m h_i^2(x)$$

mit $\mu > 0$ Penalty-Parameter

Idee: Betrachte Folge $\{\mu_k\} \subset \mathbb{R}$ mit $\mu_k \searrow 0$ und berechne Minimum x^k von $Q(x, \mu_k)$ mit Verfahren der unrestringierten Optimierung.

Beispiel 8.1.1 $\min x_1 + x_2$ bei $x_1^2 + x_2^2 - 2 = 0$

Hier ist die Lösung $\tilde{x} = (-1, -1)^T$ und die Penaltyfunktion ist

$$Q(x, \mu) = x_1 + x_2 + \frac{1}{2\mu} (x_1^2 + x_2^2 - 2)^2.$$

Bemerkung:

1. Penalty-Funktion im allgemeinen Fall:

$$Q(x, \mu) := f(x) + \frac{1}{2\mu} \sum_{i=1}^m h_i^2(x) + \frac{1}{2\mu} \sum_{j=1}^p [g_j(x)]_-^2$$

mit $[x]_- = -\min\{x, 0\}$

2. Im Gegensatz zur Merit-Funktion in Kapitel 7 ist $Q(x, \mu)$ als quadratische Funktion stetig differenzierbar (aber i.A. unstetig in 2. Ableitung).

Quadratische Penalty-Verfahren

1. Initialisiere: Wähle $\mu_0 > 0$, $\tau_0 > 0$ und $x^s, k = 0$
2. Innere Iteration: Startwert x_k^s

$$(p_k) \quad \begin{cases} \text{berechne } x_k \text{ als genähertes Minimum} \\ \text{von } Q(x, \mu_k) \\ \text{so dass } \|\nabla Q(x, \mu_k)\| \leq \tau_k. \end{cases}$$

3. Wähle $\mu_{k+1} \in (0, \mu_k)$
4. Wähle x_{k+1}^s (z.B. $x_{k+1}^s = x_k$), $k \rightarrow k + 1$ Go to 2.

Bemerkungen:

1. Effektive Wahl von μ_k adaptiv, falls Berechnung von $\min Q(x, \mu_k)$ aufwendig, wähle $\mu_{k+1} = 0.7\mu_k$ sonst, wähle $\mu_{k+1} = 0.1\mu_k$.
2. Bei Gleichheitsrestriktionen ist $Q(x, \mu_k)$ zweimal differenzierbar, d.h. Standardverfahren sind anwendbar. Für μ_k klein wird $Q''(x, \mu_k)$ schlecht konditioniert.

Satz 8.1.1 Sei $x^k = \arg \min Q(x, \mu_k)$ das exakte globale Minimum von $Q(x, \mu_k)$ und $\mu_k \searrow 0$. Dann ist jeder HP von $\{x^k\}$ eine Lösung von (PNG).

BEWEIS. Sei $\tilde{x} \in \Omega = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ globale Lösung von (PNG) d.h.

$$f(\tilde{x}) \leq f(x) \quad \forall x \in \Omega,$$

dann

$$\begin{aligned} Q(x^k; \mu_k) &\leq Q(\tilde{x}; \mu_k) \\ \Rightarrow f(x^k) + \frac{1}{2\mu_k} \sum_{i=1}^m h_i(x^k)^2 &\leq f(\tilde{x}) + \frac{1}{2\mu_k} \sum_{i=1}^m h_i^2(\tilde{x}) = f(\tilde{x}) \\ \Rightarrow \sum_{i=1}^m h_i(x^k)^2 &\leq \mu_k (f(\tilde{x}) - f(x^k)). \end{aligned} \quad (8.1)$$

Sei x^* HP von $\{x^k\}$ d.h. es existiert eine Teilfolge $\{x^{k'}\}$ mit $x^{k'} \rightarrow x^*$, dann

$$\sum_{i=1}^m h_i(x^{k'})^2 \rightarrow \sum_{i=1}^m h_i^2(x^*) = 0$$

$\Rightarrow h_i(x^*) = 0$ für $1 \leq i \leq m$ d.h. $x^* \in \Omega$

$$(8.1) \Rightarrow f(x^*) \leq f(\tilde{x}) \quad \square$$

Satz 8.1.1 setzt voraus, dass für jedes Unterproblem ein globales Minimum x^k gefunden wird. Der folgende Satz erlaubt die inexakte Minimierung von $Q(x; \mu_k)$:

Satz 8.1.2 Sei x^k so gewählt, dass $|\nabla Q(x^k; \mu_k)| \leq \tau_k$ und $\tau_k \rightarrow 0$. Ferner gelte $\mu_k \searrow 0$.

Alle HP \tilde{x} von $\{x^k\}$, für die $\nabla h_i(\tilde{x})$ $1 \leq i \leq m$ linear unabhängig sind, sind KKT-Punkte. Sei $\{x^{k'}\}$ TF mit $x^{k'} \rightarrow \tilde{x}$, dann gilt

$$\lim_{k' \rightarrow \infty} \frac{h_i(x^{k'})}{\mu_{k'}} = \lambda_i$$

mit λ_i^* Lagrange'scher Multiplikator.

BEWEIS.

$$\begin{aligned} Q(x, \mu) &= f(x) + \frac{1}{2\mu} \sum_{i=1}^m h_i^2(x) \\ \nabla Q &= \nabla f + \sum_{i=1}^m \frac{h_i}{\mu} \nabla h_i. \end{aligned} \quad (8.2)$$

Außerdem gilt

$$\begin{aligned} \left\| \sum_{i=1}^m h_i(x^k) \nabla h_i(x^k) \right\| &\leq \mu_k \|\nabla Q(x^k; \mu_k)\| + \mu_k \|\nabla f(x^k)\| \\ &\leq \mu_k (\tau_k + \|\nabla f(x^k)\|) \xrightarrow[k \rightarrow \infty]{} 0 \end{aligned}$$

$$\Rightarrow \sum_{i=1}^m h_i(\tilde{x}) \nabla h_i(\tilde{x}) = 0$$

$\Rightarrow h_i(\tilde{x}) = 0 \quad 1 \leq i \leq m$ wegen linearer Unabhängigkeit von $\nabla h_i(\tilde{x})$, $1 \leq i \leq m$,
d.h. $\tilde{x} \in \Omega$.

$$\text{Sei } \lambda^k := -\frac{h(x^k)}{\mu_k}$$

$$\begin{aligned} (8.2) \Rightarrow \sum_{i=1}^m \frac{h_i}{\mu} \nabla h_i(x^k) &= -\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) = -h'(x^k)^T \lambda^k \\ &= \nabla Q(x^k, \mu_k) - \nabla f(x^k) \end{aligned}$$

für k' groß genug ist $\text{rang } h'(x^k) = m$

$\Rightarrow h'(x^k) h'(x^k)^T$ regulär

$$\Rightarrow h'(x^k) h'(x^k)^T \lambda^k = -h'(x^k) [\nabla Q - \nabla f(x^k)]$$

$$\text{und } \lambda^k = -(h'(x^k) h'(x^k)^T)^{-1} h'(x^k) [\nabla Q - \nabla f]$$

$$\xrightarrow[k \rightarrow \infty]{} \tilde{\lambda} = (h'(\tilde{x}) h'(\tilde{x})^T)^{-1} h'(\tilde{x}) \nabla f(\tilde{x})$$

$$\text{und } (8.2) \Rightarrow \nabla f(\tilde{x}) - h'^T(\tilde{x}) \tilde{\lambda} = 0. \quad \square$$

Bemerkung:

1. Im Gegensatz zu Satz 8.1.1 ist $\{x^k\}$ linear konvergent gegen stationäre bzw. KKT-Punkte. Die $\frac{h_i(x^k)}{\mu_i}$ können als Abschätzung der Lagrange-Multiplikatoren verwendet werden.
- 2.

$$\begin{aligned} Q''(x, \mu_k) &= f''(x) + \sum_{i=1}^m \underbrace{\frac{h_i}{\mu_k}}_{\approx \lambda^i} h_i'' + \frac{1}{\mu_k} h'^T(x) h'(x) \\ &\approx \underbrace{\mathcal{L}(x, \lambda^*)}_{\text{gut konditioniert}} + \underbrace{\frac{1}{\mu_k} h'^T h'}_{\text{schlecht konditioniert}}, \end{aligned}$$

d.h. die Berechnung der Newton-Richtung wird zunehmend schlechter konditioniert mit wachsendem k , d.h. ungenaue Berechnung des Newton-Schritts.

8.2 Die logarithmische Barrier-Methode

Betrachte (PNU)

$$\begin{array}{ll} \min & f(x) \\ \text{bei} & g(x) \geq 0 \end{array}$$

sei $\Omega^\circ = \{x \in \mathbb{R}^n \mid g_j(x) > 0, 1 \leq j \leq p\} \neq \emptyset$.

Barrier-Funktionen sind

- ∞ außerhalb Ω° ,
- glatt innerhalb Ω° ,
- für $x^k \rightarrow x$ mit $x \in \partial\Omega^\circ$ gilt $\lim P(x) = \infty$.

Wichtigstes Beispiel:

$$P(x; \mu) = f(x) - \mu \sum_{j=1}^p \log(g_j(x)).$$

Vergleich

$$Q(x, \mu) = f(x) + \frac{1}{2\mu} \sum_{j=1}^p [g_j(x)]_-$$

Beispiel 8.2.1

$$\begin{array}{ll} \min & x \\ \text{bei} & x \geq 0 \quad 1 - x \geq 0 \\ & P(x, \mu) = x - \mu \log x - \mu \log(1 - x) \end{array}$$

Bemerkungen:

1. Wie bei Penalty-Funktionen lässt sich schlechte Kondition von $P''(x, \mu)$ zeigen.
2. Algorithmus analog zum Penalty-Verfahren.
3. Unter den üblichen Voraussetzungen kann man zeigen
 - i) $\exists x(\mu)$ stetig differenzierbar für μ klein genug, so dass $x(\mu)$ lokales Minimum von $P(x, \mu)$ und $\lim_{\mu \searrow 0} x(\mu) = \tilde{x}$, die lokale Lösung von (PNU) ist.
 - ii) $\lambda_i(\mu) = \frac{\mu}{g_i(x(\mu))} \xrightarrow{\mu \searrow 0} \tilde{\lambda}$ dem Lagrange'schen Multiplikator.
 - iii) Die Trajektorie $\mathcal{C}_p = \{x(\mu) \mid \mu > 0\}$ heißt zentraler Pfad (oder primaler zentraler Pfad).
 - iv) Behandlung von Gleichheitsrestriktionen durch

$$B(x; \mu) := f(x) - \mu \sum_{i=1}^p \log(g_i(x)) + \frac{1}{2\mu} \sum_{i=1}^m h_i^2(x).$$

8.3 Augmentierte Lagrange-Methoden

Motivation: $Q(x, \mu) = f(x) + \frac{1}{2\mu} \sum h_i(x)^2$.

In der Regel erfüllen Minima x^k von $Q(x, \mu_k)$ nicht $x^k \in \Omega$, denn es gilt nach Beweis von Satz 8.1.2 $h_i(x^k) \approx -\mu_k \tilde{\lambda}_i$ für $1 \leq i \leq m$.

Idee: Modifiziere Q , um x^k "näher" an Ω zu bringen (und so $\mu_k \searrow 0$ zu vermeiden)

Betrachte Augmentierte Lagrange-Funktion:

$$\mathcal{L}_A(x, \lambda; \mu) := f(x) - \sum_{i=1}^m \lambda_i h_i + \frac{1}{2\mu} \sum_{i=1}^m h_i^2$$

$$\nabla_x \mathcal{L}_A = \nabla f(x) - \sum_{i=1}^m \left(\lambda_i - \frac{h_i}{\mu} \right) \nabla h_i.$$

Entwickle Algorithmus, so dass

- μ^k und λ^k festgelegt werden
- $x^k = \arg \min \mathcal{L}_A(x, \lambda^k; \mu_k)$.

In Analogie zu Satz 8.1.2 kann man vermuten

$$\tilde{\lambda}_i \approx \lambda_i^k - \frac{h_i(x^k)}{\mu_k} \quad (8.3)$$

bzw. $h_i(x^k) \approx -\mu_k(\lambda_i^* - \lambda_i^k)$ (d.h. falls λ_i^k nahe an λ_i^* , ist x^k "näher" an Ω als in Penalty-Verfahren).

Update von λ_i^k durch (8.3):

$$\lambda_i^{k+1} = \lambda_i^k - \frac{h_i(x^k)}{\mu_k}$$

- Augmentiertes Lagrange-Verfahren (Gleichungs-Nebenbedingungen)

1. Initialisiere $\mu_0 > 0$, $\tau_0 > 0$, x_0^s, λ^0 , $k = 0$.
2. Wähle Startwert x_k^s und minimiere $\mathcal{L}(x, \lambda^k; \mu_k)$, so dass

$$\|\nabla_x \mathcal{L}(x^k, \lambda^k; \mu_k)\| \leq \tau_k.$$

3. $\lambda_i^{k+1} = \lambda_i^k - \frac{h_i(x^k)}{\mu_k}$, $\mu_{k+1} \in (0, \mu_k)$
 $x_{k+1}^s = x^k$, $k \rightarrow k+1$ go to 2

Beispiel 8.3.1

$$\begin{cases} \min & x_1 + x_2 \\ \text{bei} & x_1^2 + x_2^2 - 2 = 0 \end{cases}$$

$$\mathcal{L}_A = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - 2) + \frac{1}{2\mu}(x_1^2 + x_2^2 - 2)^2$$

$$\tilde{x} = (-1, -1)^T, \tilde{\lambda} = -0.5.$$

Es ergibt sich z.B. für $\mu_k = 1$ und $\lambda^k = -0.4$

$$\arg \min Q(x, 1) \approx (-1.1, -1.1)$$

aber $\arg \min \mathcal{L}_A(x, -0.4, 1) \approx (-1.02, -1.02)$.

Erweiterung auf Ungleichungsrestriktionen

Schritt 1: Einführung von Schlupf-Variablen s_j

Ersetze $g_j(x) \geq 0$ durch

$$g_j(x) - s_j = 0 \quad \text{und} \quad s_j \geq 0$$

—→ Problem mit Gleichungs- und Box-Restriktionen.

Schritt 2: Betrachte

$$\begin{aligned} &\min f(x) \\ &\text{bei } g(x) \geq 0 \end{aligned}$$

bzw.

$$\begin{aligned} &\min_{x,s} f(x) \\ &\text{bei } g_i(x) - s_i = 0 \\ &\quad s_i \geq 0, \quad 1 \leq i \leq p, \end{aligned}$$

und schreibe die augmentierte Lagrangefunktion nur für die Gleichungsrestriktionen auf:

$$\begin{aligned} &\min_{x,s} f(x) - \sum_{j=1}^p \lambda_j^k (g_j(x) - s_j) \\ &\quad + \frac{1}{2\mu_k} \sum_{j=1}^p (g_j(x) - s_j)^2 \\ &\text{bei } s \geq 0 \quad (s \in \mathbb{R}^p). \end{aligned}$$

Idee: Zunächst explizite Minimierung bezüglich s_j

Partielle Ableitung nach s_j : $\lambda_j^k - \frac{1}{\mu_k} (g_j(x) - s_j) \stackrel{!}{=} 0$

$$\Leftrightarrow s_j = g_j - \mu_k \lambda_j^k$$

bzw. $s_j = \max(g_j(x) - \mu_k \lambda_j^k, 0)$, $1 \leq j \leq p$.

Einsetzen in \mathcal{L}_A liefert

$$\begin{aligned} &-\lambda_j^k (g_j(x) - s_j) + \frac{1}{2\mu_k} (g_j(x) - s_j)^2 \\ &= \begin{cases} -\lambda_j^k g_j(x) + \frac{1}{2\mu_k} g_j^2(x), & \text{falls } g_j(x) - \mu_k \lambda_j^k < 0, \\ -\frac{\mu_k}{2} (\lambda_j^k)^2, & \text{sonst.} \end{cases} \end{aligned}$$

Definiere

$$\psi(t, \sigma; \mu) = \begin{cases} -\sigma t + \frac{1}{2\mu} t^2, & \text{falls } t - \mu\sigma < 0, \\ -\frac{\mu}{2} \sigma^2, & \text{sonst,} \end{cases}$$

dann ergibt sich

$$\min_x \mathcal{L}_A(x, \lambda^k, \mu_k) = f(x) + \sum_{j=1}^p \psi(g_j(x), \lambda_j^k, \mu_k).$$

Bemerkungen:

1. Weiter im Prinzip wie im Verfahren mit Gleichungsrestriktionen, nur Modifikation des Multiplikator-Updates:

$$\lambda_i^{k+1} = \max \left\{ \lambda_i^k - \frac{g_i(x^k)}{\mu_k}, 0 \right\}$$

2. $\psi(g_i(x), \lambda_i, \mu)$ stetig partiell differenzierbar bezüglich x aber 2. Ableitung unstetig in $g_i(x) = \mu\lambda_i$,

aber: bei strikter Komplementarität im Allgemeinen kein Problem

denn: Fall (1) i aktiv, d.h. $g_i(\tilde{x}) = 0$, d.h. $g_i(x^k) \approx 0$ aber $\lambda_i^k, \mu^k > 0$.

Fall (2) i inaktiv: $g_i(x^k) > 0$ aber $\lambda_i^k \approx 0$.

Satz 8.3.1 Sei \tilde{x} lokale Lösung von (PNG), $\nabla h_i(\tilde{x})$, $1 \leq i \leq m$ sei linear unabhängig und die hinreichenden Optimalitätsbedingungen 2. Ordnung gelten für $\lambda = \tilde{\lambda}$. Dann existiert ein $\bar{\mu} > 0$, so dass für alle $\mu \in (0, \bar{\mu}]$ \tilde{x} ein striktes lokales Minimum von $\mathcal{L}_A(x, \tilde{\lambda}; \mu)$ ist.

Beweisskizze: (nur notwendige Bedingung)

Zu zeigen ist:

$$\nabla_x \mathcal{L}_A(\tilde{x}, \tilde{\lambda}; \mu) = 0, \quad \nabla_{xx}^2 \mathcal{L}_A(\tilde{x}, \tilde{\lambda}, \mu) \quad \text{positiv definit}$$

$$\mathcal{L}_A(x, \lambda_i, \mu) = f(x) - \sum_{i=1}^m \lambda_i h_i + \frac{1}{2\mu} \sum_{i=1}^m h_i^2$$

$$\begin{aligned} \nabla_x \mathcal{L}_A(\tilde{x}, \tilde{\lambda}; \mu) &= \nabla f(\tilde{x}) - \sum_{i=1}^m \left(\lambda_i - \frac{h_i}{\mu} \right) \nabla h_i(\tilde{x}) \\ &= \nabla f(\tilde{x}) - \sum_{i=1}^m \lambda_i \nabla h_i(\tilde{x}) \\ &= \nabla_x \mathcal{L}(\tilde{x}, \tilde{\lambda}) = 0. \end{aligned}$$

Eine praxistaugliche Implementierung dieses Verfahrens für das Problem

$$\begin{aligned} &\min f(x) \\ &\text{bei } h(x) = 0 \\ &\quad l \leq x \leq u \end{aligned} \tag{P}$$

ist, z.B. LANCELOT, siehe: Conn, Gould, Toint: LANCELOT—a FORTRAN package for Large-Scale Nonlinear Optimization, Springer, 1992.