Technische Universität Berlin

Fakultät II – Institut für Mathematik

Nichtlineare Optimierung

Vorlesung im Wintersemester 05/06

Dietmar Hömberg

Im WS 01/0.2 gehalten von F. Tröltzsch. Grundlage der Vorlesung ist das Buch von Prof. Dr. Walter Alt, Universität Jena.
Das Skript ist nur für den internen Gebrauch während dieser Vorlesung bestimmt.

Inhaltsverzeichnis

1 Optimierungsaufgaben

1.1 Literatur

1. Alt, W., *Nichtlineare Optimierung.* Vieweg, Braunschweig/Wiesbaden 2002.
2. Gill, P.E., Murray, W., and M.H. Wright, *Practical Optimization.* Academic Press, London 1981.
3. Kelley, C.T., *Iterative Methods for Optimization.* SIAM, Philadelphia 1999.
4. Nocedal, J. and Wright, S.J., *Numerical Optimization.* Springer, New York 1997.
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 $f : \mathbb{R}^n \to \mathbb{R}$ zu berechnen. Dazu einige Beispiele und Grundbegriffe:

Beispiel 1.2.1

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

Beispiel 1.2.2
$f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) - \cos(x_1^2) - \cos(x_2^2), f : \mathbb{R}^2 \to \mathbb{R}$ *hat ein (strenges)* **globales** *Minimum und mehrere (strenge)* **lokale** *Minima und Maxima.*

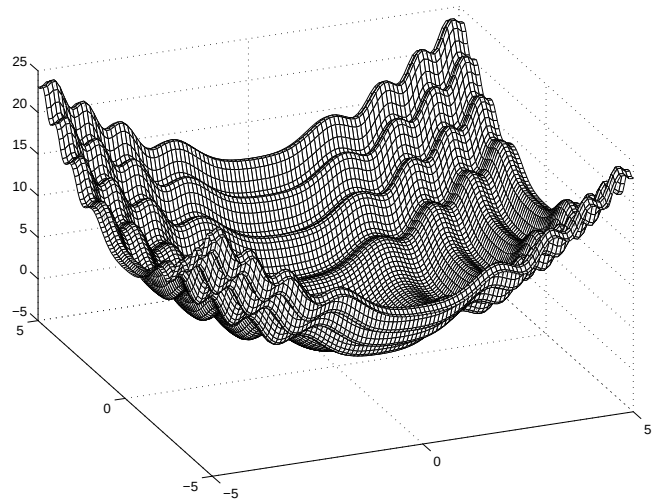


Abb. 9.1 und Text

Text und Abb. 9.1

Bezeichnungen:

Für $x \in \mathbb{R}^n$ heißt

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \qquad \textbf{euklidische Norm}$$

$$B(x,r) = \{y \in \mathbb{R}^n \mid \|y - x\| < r\} \qquad \textbf{offene Kugel}$$

$$\overline{B}(x,r) = cl\, B(x,r) \qquad \textbf{abgeschlossene Kugel}$$

x_i	i-te Komponente von x
$x^{(k)}$	k-tes Glied einer Folge von Vektoren $\left(x^{(k)}\right)_{k=1}^{\infty}$

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

$$\text{(P)} \quad \boxed{\min_{x \in \mathcal{F}} f(x)}$$

Man nennt f - **Zielfunktion**
$\mathcal{F}$ - **zulässiger Bereich** oder zulässige Menge

Im Fall $\mathcal{F} = D$ 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 $\mathcal{F}$ durch Nebenbedingungen gegeben, so heißt (P)

Optimierungsproblem mit Nebenbed. oder **restringiertes Opt.-problem**. In der Regel ist $\mathcal{F}$ durch Gleichungen und Ungleichungen definiert. Die Elemente von $\mathcal{F}$ heißen **zulässige Punkte**.

Beispiel 1.2.3

$$\begin{aligned} &\min_{(x\in\mathbb{R})} x^3 \\ &\textit{bei} \quad x \geq 1 \end{aligned}$$

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

Definition 1.2.1 *Ein Punkt* $\tilde{x} \in \mathcal{F}$ *heißt*

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

$$f(x) \geq f(\tilde{x}) \qquad \forall\, x \in \mathcal{F} \cap B(\tilde{x}, r)$$

- *analog* **strenges lokales Min.**, *wenn entsprechend*

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

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

$$f(x) \geq f(\tilde{x}) \qquad \forall\, x \in \mathcal{F}$$

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

$$f(x) > f(\tilde{x}) \qquad \forall\, x \in \mathcal{F}, \quad x \neq \tilde{x}$$

  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 $\tilde{x}$,

$$f(x) \leq f(\tilde{x}) \qquad \forall\, x \in \mathcal{F}$$

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

Beispiel 1.2.4 *(Lineare Regression)*
Gesucht ist eine lineare Funktion

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

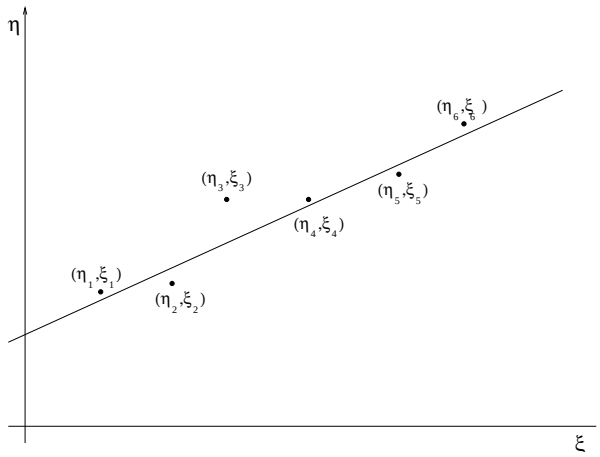
mit unbekannten Koeffizienten x_1, x_2, welche am besten zu gegebenen Wertepaaren $(\xi_i, \eta_i), i = 1, \ldots, m$ (z.B. Messwerten) passt. Wir setzen

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

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.

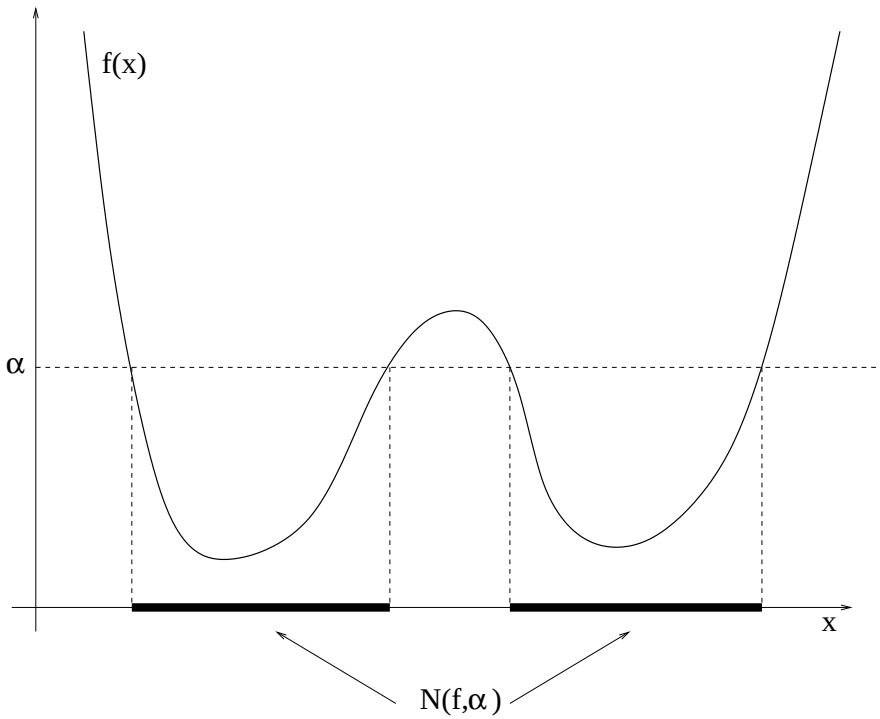
1.3 Existenz von Lösungen

Grundlage für die meisten Existenzbeweise ist der bekannte

Satz 1.3.1 *(Weierstraß)*

Ist $f : \mathbb{R}^n \supset D \to \mathbb{R}$ 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 *$f : D \to \mathbb{R}, D \subset \mathbb{R}^n, \alpha \in \mathbb{R}$. Die Mengen $N(f, \alpha) = \{x \in D \quad | \quad f(x) \leq \alpha\}$ heißen* **Niveaumengen** *von f.*



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

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

kompakt. Dann gibt es (mindestens) ein globales Minimum von f auf $\mathcal{F}$.

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

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

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

Typische Anwendungen dieses Prinzips sind die folgenden zwei Aussagen:

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

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

Dann besitzt die Aufgabe

$$\min_{x \in \mathcal{F}} f(x)$$

mindestens eine globale Lösung.

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

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

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

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 $x^T H x \to \infty, \|x\| \to \infty$.

Beispiel 1.3.1 (Unrestringierte quadratische Optimierungsaufgabe)

Wir betrachten

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

mit gegebenem $b \in \mathbb{R}^n$ und positiv definiter (n, n)-Matrix H. Sie zeigen leicht, dass $\lim_{\|x\| \to \infty} f(x) = \infty$ 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

$$
\begin{aligned}
f(x) \quad &= \sum_{i=1}^{m} \left(\eta_i - (x_1\xi_i + x_2)\right)^2 \\
&= \underbrace{\sum_{i=1}^{m} \eta_i^2}_{c} \quad \underbrace{-2\sum_{i=1}^{m} \eta_i \left(\xi_i x_1 + x_2\right)}_{b^T x} \quad \underbrace{+\sum_{i=1}^{m} \left(x_1\xi_i + x_2\right)^2}_{\frac{1}{2}x^T H x} \\
&= \tfrac{1}{2}x^T H x + b^T x + c \\
\textit{mit} \quad H \quad &= 2\begin{pmatrix} \sum_1^m \xi_i^2 & \sum_1^m \xi_i \\ \sum_1^m \xi_i & m \end{pmatrix} \quad b = -2\begin{pmatrix} \sum_1^m \xi_i\eta_i \\ \sum_1^m \eta_i \end{pmatrix}.
\end{aligned}
$$

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!

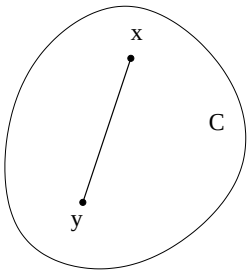
1.4 Konvexe Optimierungsaufgaben

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).

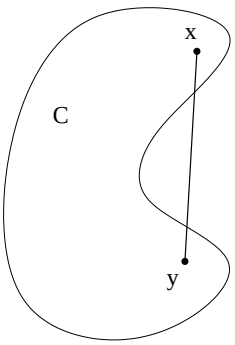
Definition 1.4.1 *Eine Menge $C \subset \mathbb{R}^n$ heißt konvex, falls für je 2 beliebige $x, y \in C$ auch die Strecke*

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

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



konvexe Menge



nicht konvexe Menge

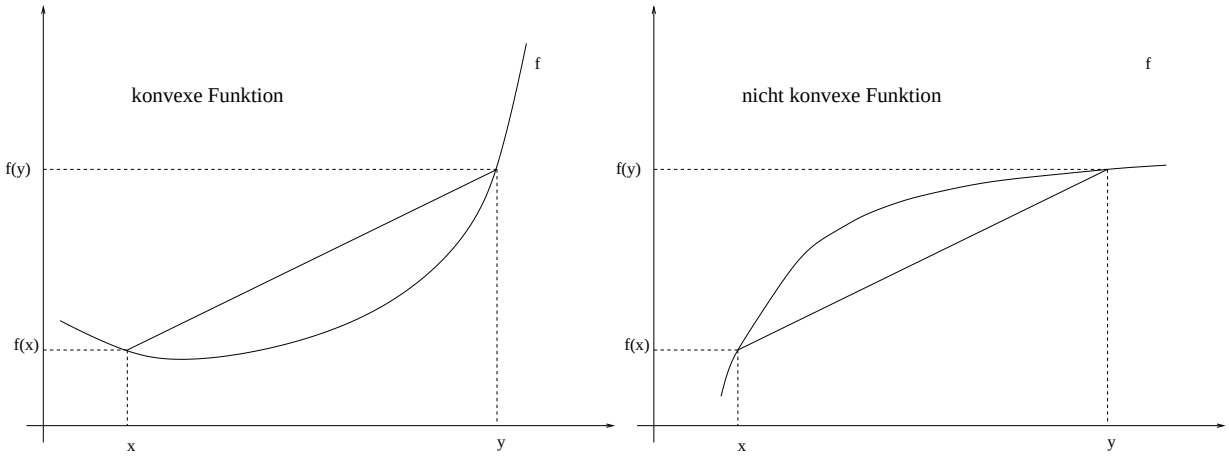
Definition 1.4.2 *Sei $C \subset \mathbb{R}^n$ konvex und nichtleer, $C \subset D$. Eine Funktion $f : D \to \mathbb{R}$ heißt* **konvex auf** *C, wenn*

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

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.*

Definition 1.4.3 *Nun betrachten wir $f : D \to \mathbb{R}, D \subset \mathbb{R}^n$ offen, nichtleer, $\mathcal{F} \subset D$. Ist f konvex auf $\mathcal{F}$, so heißt das Problem*

$$\min_{x \in \mathcal{F}} f(x) \qquad \textbf{(P)}$$

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

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

Zu zeigen ist

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

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

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

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

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 $(*)$

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

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

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

$$f(x) = f(y) = \alpha = \min_{x \in \mathcal{F}} f(x)\,.$$

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 $z \in \mathcal{F}$ und insgesamt widerspricht das der Optimalität von x, y. Damit $x = y$, strenges Minimum □

Beispiel 1.4.2

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

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

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

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

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

$\Rightarrow$ *Minimierung von*

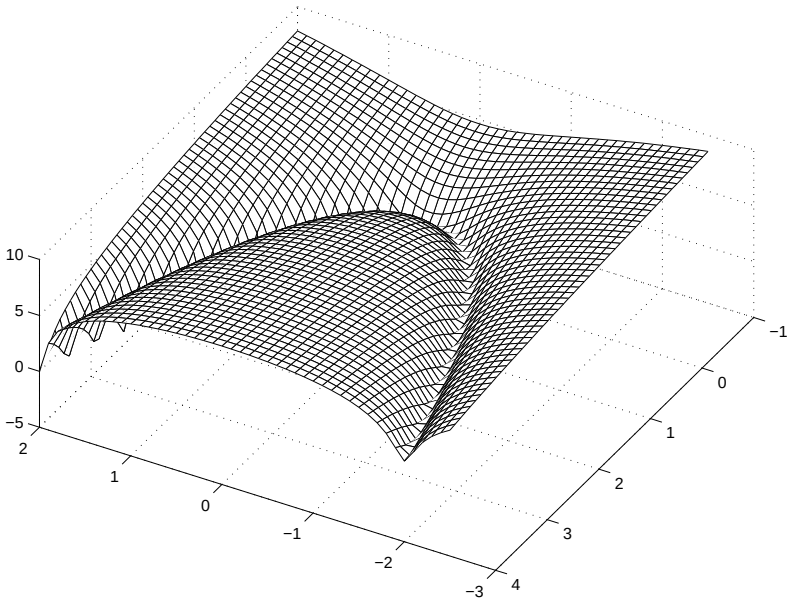
$$\boxed{f(x) = \sum_{i=1}^{m} (\eta_i - g(x, \xi_i))^2}$$

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

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

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

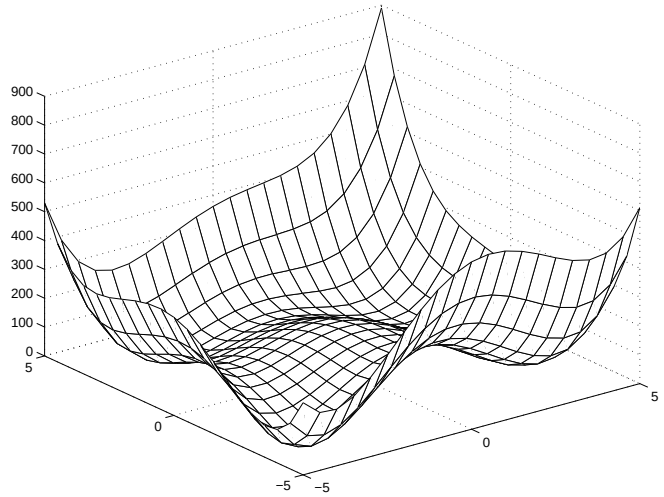
$$\boxed{f(x_1, x_2) = \underbrace{100(x_2 - x_1^2)^2}_{\substack{\text{definiert} \\ \text{das Tal} \\ \text{(Parabel)}}} + \underbrace{(1 - x_1)^2}_{\substack{\text{kippt leicht} \\ \text{an}}}}$$


10
5
0
−5
2
1
0
−1
−2
−3
−1
0
1
2
3
4

Beispiel 1.5.3 *(Himmelblau)*

$$\boxed{f(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2}$$

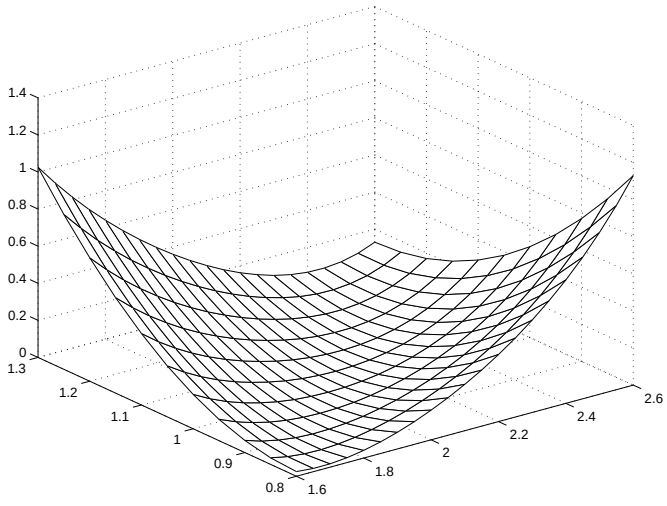
4 lokale Minimalstellen, die zugleich globale Minimalstellen mit Funktionswert 0 *sind; 4 Sattelpunkte und ein lokales Maximum bei* $(-0.270845, -0.923039)^T$.



Beispiel 1.5.4 *(Bazaraa-Shetty)*

$$\boxed{f(x_1, x_2) = (x_1 - 2)^4 + (x_1 - 2x_2)^2}$$

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

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

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

Beispiel 1.5.6 *(Dixon)*

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

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

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:

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

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:
Moré 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

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

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 $x^0,\dots,x^n$.*

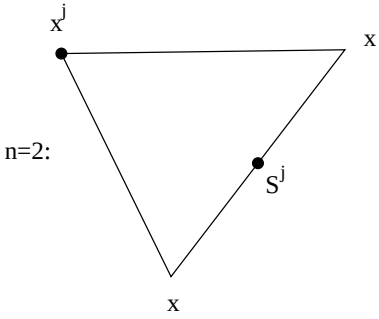
- Beim Start des Verfahrens wird ein Simplex vorgegeben.
- Man ermittelt die (bzw. eine) Ecke mit dem größten Funktionswert,

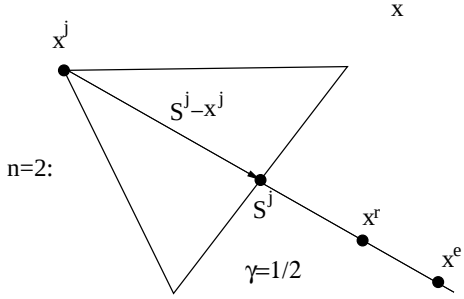
$$f(x^m)=\max\left\{f(x^0),\dots,f(x^n)\right\}$$

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

Dazu werden folgende Konstruktionen benutzt:

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



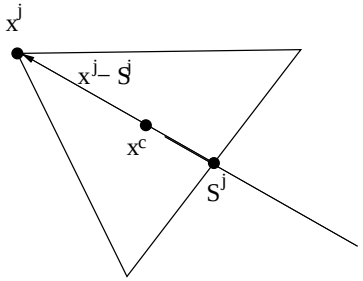


γ : Reflektionskonstante

Konstruktionsprinzipien:

- **Reflektion** von x^j an s^j
  $x^r = s^j + \gamma(s^j - x^j)$, $0 < \gamma \leq 1$

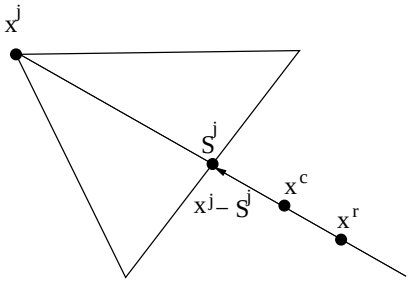
- Dieses eben konstruierte x^r kann weiter nach außen bewegt werden:
  Expansion von x^r in Richtung $s^j - x^j$ (d. h. in Richtung $x^r - s^j$)
  $x^e = s^j + \beta(x^r - s^j)$, $\beta > 1$ Expansionskonstante



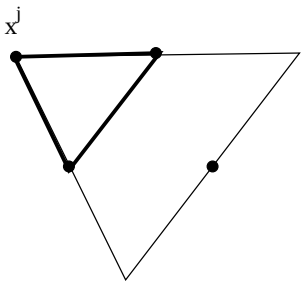
- **Kontraktion** (3 Typen)

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

  (ii) **Partielle Kontraktion außen**
  $x^c = s^j + \alpha(x^r - s^j)$



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


x^j

2.1.2 Ablauf des Verfahrens

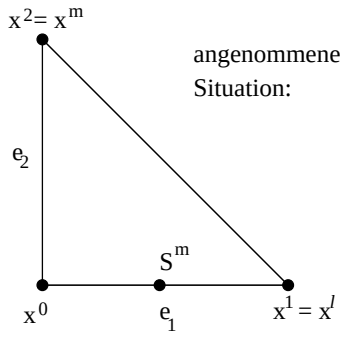
Die einfachste Variante läuft so ab:

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

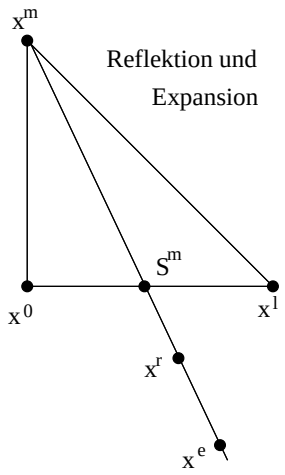
Folgende Schritte laufen ab:

1. Wahl eines Startpunkts $x^0 \in R^n$, Festlegung der anderen n Ecken des Startsimplexes durch
$$x^j = x^0 + e^j,\ j = 1, \ldots, n\,,$$
wobei e^j den j-ten Standardeinheitsvektor bezeichnet.

2. Bestimme (die) Ecken mit maximalem und minimalem Funktionswert: x^m, x^l mit
$$\begin{aligned} f(x^m) &= \max\left\{f(x^0), \ldots, f(x^n)\right\} \\ f(x^l) &= \min\left\{f(x^0), \ldots, f(x^n)\right\} \end{aligned}$$
und bestimme den Schwerpunkt der Ecken bezügl. x^m
$s^m = \frac{1}{n} \sum\limits_{\substack{i=0 \\ i \neq m}}^{n} x^i$

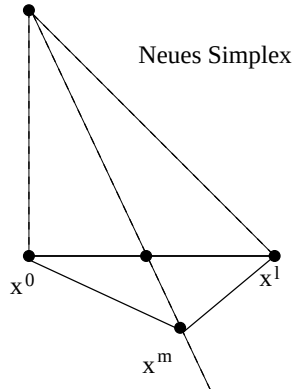


3. Reflektion von x^m am Schwerpunkt s^m
$$x^r = s^m + \gamma(s^m - x^m)$$



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:
$$\tilde{x}^m = \left\{ \begin{array}{ll} x^e, & f(x^e) < f(x^r) \\ x^r, & f(x^r) \leq f(x^e) \end{array} \right\}$$
$\underline{\underline{x^m := \tilde{x}^m}}$

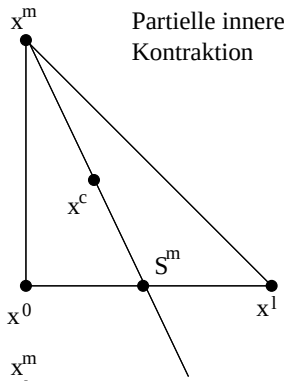


(ii)

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

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

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



(iii)

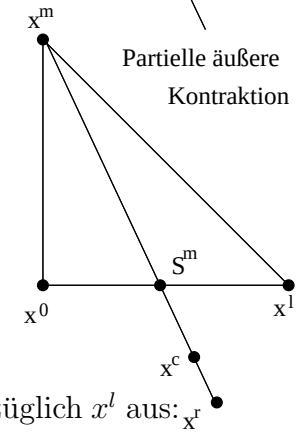
$$\boxed{f(x^r) > \max\{f(x^j),\, j \neq m\}}$$

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

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

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

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

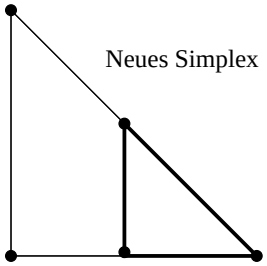


- Wenn $f(x^c) < f(x^m)$, dann ersetze x^m durch x^c

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

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

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



5. Gehe mit dem neu ermittelten Simplex (Ecken $\{x^0, \ldots, x^n\}$) zu Schritt 2.
   Das Verfahren erzeugt Eckenfolgen $\{x^{(k,0)}, \ldots, x^{(k,n)}\}_{k=1}^{\infty}$ und stellt sicher, dass

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

   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.

<u>Verfügbare Codes:</u>	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 $(1.00002, 1.00003)^T$

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.

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

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

3.
$$x^{(k+1)} = \begin{cases} v^{(k)} & \text{falls} \quad f\left(v^{(k)}\right) < f\left(x^k\right) \\ x^{(k)} & \text{sonst} \end{cases}$$

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

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

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 erster Ordnung

Im gesamten Kapitel 3 wird vorausgesetzt:

$D \subset \mathbb{R}^n$ offen, nichtleer
$f : D \to \mathbb{R}$ mit gewissen Differenzierbarkeitsannahmen

Wir betrachten die unrestringierte Aufgabe

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

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*

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

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

$$f(x) = \tfrac{1}{2} x^T H x + b^T x \qquad \begin{matrix} H \in \mathbb{R}^{(n,n)}, & \underline{\textit{symmetrisch}} \\ b \in \mathbb{R}^n \end{matrix}$$

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

$$H = 2 \begin{pmatrix} \sum_1^m \xi_i^2 & \sum_1^m \xi_i \\ \sum_1^m \xi_i & m \end{pmatrix}, \quad b = -2 \begin{pmatrix} \sum_1^m \xi_i \eta_i \\ \sum_1^m \eta_i \end{pmatrix}$$

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

$$\begin{array}{rcrcl} \left(\sum_1^m \xi_i^2\right) x_1 & + & \left(\sum_1^m \xi_i\right) x_2 & = & -\sum_1^m \xi_i \eta_i \\ \sum_1^m \xi_i x_1 & + & m \quad x_2 & = & -\sum_1^m \eta_i \,. \end{array}$$

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 $\tilde{x} = \binom{1}{1}$*

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

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

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

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

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

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

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

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

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

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

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

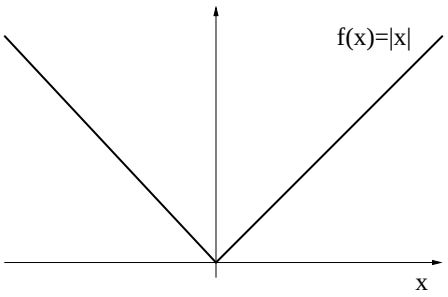
Sei $h \in \mathbb{R}^n$ beliebig, aber fest. Dann gilt $\tilde{x} + th \in B(\tilde{x}, r)$ für betragsmäßig kleine t, somit

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

□

(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 Min. bei $\tilde{x} = 0$.*



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

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

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*

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

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:

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

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

Beispiel 3.1.4

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

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

Beispiel 3.1.5 *Rosenbrock-Funktion*

$$\begin{aligned} & fx_1 = -400\,x_1\,(x_2 - x_1^2) - 2\,(1 - x_1) \\ & fx_2 = 200\,(x_2 - x_1^2) \\ & fx_1x_1 = -400\,(x_2 - x_1^2) + 800\,x_1^2 + 2 \\ & fx_1x_2 = fx_2x_1 = -400\,x_1 \\ & fx_2x_2 = 200 \\ \Rightarrow\ & f''(1,1) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} \textit{ positiv definit nach Satz von Sylvester.} \end{aligned}$$

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*

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

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

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

$\Rightarrow \tilde{x}$ lokales Min.

□

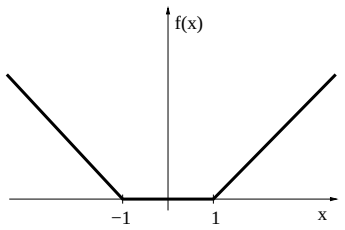
Beispiel 3.1.6 *Lineare Regression*

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 Min. 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.*

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

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

$$\boxed{f(x) \geq f(\tilde{x}) + \alpha \|x - \tilde{x}\|^2 \quad \forall x \in B(\tilde{x}, r)}$$ **"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

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

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

Offenbar gilt hier $h^T f''(z) h \geq 0\ \forall h,\ \forall z \in B(\tilde{x}, r)$, 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* $p = 1$

$$\begin{array}{rcll} f'(x) & = & 2p\,x^{2p-1} & \Rightarrow \quad f'(0) = 0 \\ f''(x) & = & 2p(2p-1)x^{2p-2} & \Rightarrow \quad f''(0) > 0 \quad \textit{falls} \quad p = 1 \\ & & & \quad\;\; f''(0) = 0 \quad \textit{falls} \quad p > 1 \end{array}$$

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

$$\text{(P)} \quad \boxed{\min_{x \in \mathcal{F}} f(x)} \quad f : D \to \mathbb{R} \quad \text{konvex } (D \text{ offen})$$
$$\mathcal{F} \subset D \quad \text{konvex}, \neq \emptyset, \text{ nicht notwendig offen}$$

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

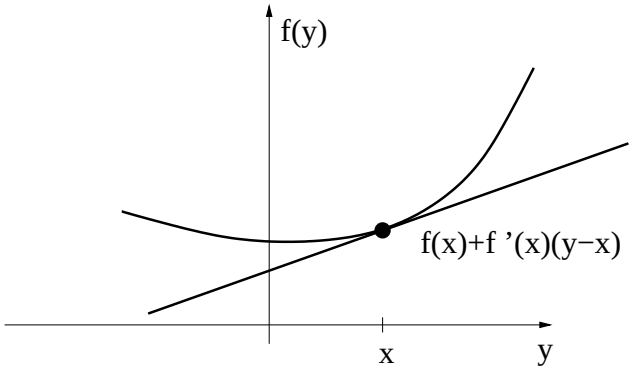
Charakterisierung der Konvexität von f durch Ableitungen:

Satz 3.2.1 *f sei differenzierbar in D. Dann ist f genau dann konvex auf $\mathcal{F}$, wenn*

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

Beweis: Siehe Fachbücher (ist sehr einfach). □

Illustration für $n = 1$:



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

$$\boxed{f'(\tilde{x})(x - \tilde{x}) \geq 0 \quad \forall x \in \mathcal{F}} \quad \textbf{Variationsungleichung} \tag{3.7}$$

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 $h = x - \tilde{x} \quad f'(\tilde{x}, h) = f'(\tilde{x})(x - \tilde{x})$.)

(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, $\mathcal{F} \subset D$ nichtleer und konvex, $f : D \to \mathbb{R}$ differenzierbar auf D. Dann ist f genau dann streng konvex auf $\mathcal{F}$, wenn*

$$f(y) > f(x) + f'(x)(y - x) \quad \forall x, y \in \mathcal{F}, x \neq y. \tag{3.8}$$

Beweis:

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

$$f(y) \geq f(x) + f'(x)(y - x). \tag{$*$}$$

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,

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

(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 $\mathcal{F}$ sowie $\tilde{x} \in \mathcal{F}$. Dann gilt: $\tilde{x}$ ist genau dann strenges lokales Minimum von (P) und damit auch strenges globales Minimum wenn die Variationsungleichung*

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

erfüllt ist.

Beweis:

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

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

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

□

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

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

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

(ii) Ist $f''(x)$ positiv definit $\forall x \in \mathcal{F}$, so ist f streng konvex auf F.

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

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

(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 $\mathcal{F}$ offen, $x \in \mathcal{F}$ 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 $\Rightarrow$

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

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

(ii) Ist $f''(x)$ positiv <u>definit</u>, 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*

$$\boxed{(1-\lambda)f(x) + \lambda f(y) \geq f((1-\lambda)x + \lambda y) + \lambda(1-\lambda)\alpha\|x-y\|^2}$$

$\forall x, y \in \mathcal{F}, \lambda \in [0,1]$.

Damit kann man zeigen:

- *Gleichmäßige Konvexität ist bei differenzierbarem f äquivalent zu*

$$f(y) - f(x) \geq f'(x)(y-x) + \alpha\|x-y\|^2 \quad \forall x, y \in \mathcal{F}$$

- *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 $\beta > 0$ <u>nicht</u> von $x \in \mathcal{F}$ abhängt.

Ist $\mathcal{F}$ offen, dann folgt umgekehrt aus der gleichmäßigen Konvexität von f auf $\mathcal{F}$, 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

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

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

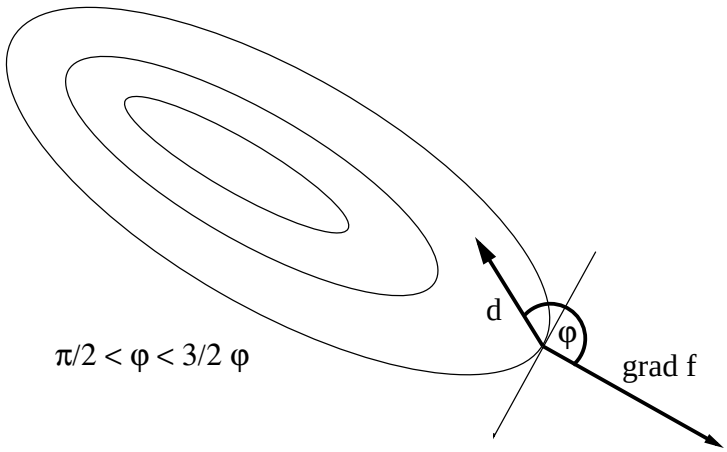
$$\nabla f(\tilde{x}) = 0 \tag{4.9}$$

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 *$f : \mathbb{R}^n \to \mathbb{R}$ 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 $f(x + \sigma d) < f(x) \quad \forall x \in [0, \bar{\sigma}]$*

Beweis: Wegen

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

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**

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

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*

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

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*

$$\begin{array}{l} f\left(x^{(k)} + \sigma_k d^{(k)}\right) < f\left(x^{(k)}\right), \\ x^{k+1} := x^{(k)} + \sigma_k d^{(k)} \end{array}$$

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

Bemerkungen:

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

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

- 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 $F : \mathbb{R}^n \to \mathbb{R}^n$ 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 $F\left(x^{(k)}\right)+F'\left(x^{(k)}\right)\left(x - x^{(k)}\right)$, so dass wir von dieser Funktion eine Nullstelle x suchen. Wir lösen also

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

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

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

Für die Konvergenzanalysis des Verfahrens benötigen wir folgende Voraussetzungen:

(i) $F : \mathbb{R}^n \supset D \to \mathbb{R}^n$ 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) $\exists\, F'(\tilde{x})^{-1}$

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

Lemma 4.2.1 *(i)*

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

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

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 $G : \bar{B}(\tilde{x}, r) \to \mathbb{R}^n$ 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 $x^{(k)}$*

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

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

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

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

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

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 $B(\tilde{x}, \delta)$:

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

Man schätzt ab und findet z. B.

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

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

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

□

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

$$\boxed{F(x) := \nabla f(x).}$$

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*

$$\boxed{x^{(k+1)} = x^{(k)} - f''\left(x^{(k)}\right)^{-1} \nabla f\left(x^{(k)}\right)} \tag{4.13}$$

lokal quadratisch gegen $\tilde{x}$.

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

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

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

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

und setzt

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

Für $d^{(k)}$ gilt

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

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

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

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

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

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

Das ist gerade die notwendige Optimalitätsbedingung für Lösungen der quadratischen Optimierungsaufgabe $(Q)_k$,

$$\boxed{\min_{x\in\mathbb{R}^n} \quad \nabla f\left(x^{(k)}\right)^T\left(x - x^{(k)}\right) + \tfrac{1}{2}\left(x - x^{(k)}\right)^T f''\left(x^{(k)}\right)\left(x - x^{(k)}\right).} \qquad (\mathrm{Q})_k$$

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 $(\mathrm{Q})_k$ so auf:

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

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 $f(x) = x^2$, $\quad d^{(k)} = -1$ *für alle* $k \geq 0$, $\quad x^{(0)} = 1$, *und* $\sigma_k = \left(\frac{1}{2}\right)^{k+2}$, $\quad k = 0, 1, \ldots$
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*

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

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

$$f(x+\sigma d) \leq f(x) + c_1 \sigma \nabla f(x)^T d \tag{4.15}$$
(hinreichend schneller Abstieg)

und

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

Aus diesen beiden Bedingungen folgt (4.14), Effizienz, mit $c = c_1 c_2$, denn

$$f(x+\sigma d) \leq f(x) + c_1 \left(-c_2 \frac{\nabla f(x)^T d}{\| d \|^2}\right) \nabla f(x)^T d = f(x) - c_1 c_2 \left(\frac{\nabla f(x)^T d}{\| d \|}\right)^2$$

Bemerkung: Man kann unter Voraussetzung der Lipschitz-Stetigkeit von f' 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

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

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

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

Man will nun die Richtungen so wählen, dass daraus folgt

$$\nabla f\left(x^{(k)}\right) \to 0, \quad k \to \infty. \tag{4.18}$$

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 größer als der rechte Winkel zu $\nabla f(x^{(k)})$ bleiben. Nun gilt

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

Daraus folgt $\nabla f(x^{(k)}) \to 0$, 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*

$$-\nabla f(x)^T d \geq c_3 \, \|\nabla f(x)\| \, \|d\| \tag{4.19}$$

mit einer von x und d unabhängigen Konstanten $c_3 > 0$ gilt.
Sie heißt **streng gradientenbezogen**, *wenn zusätzlich*

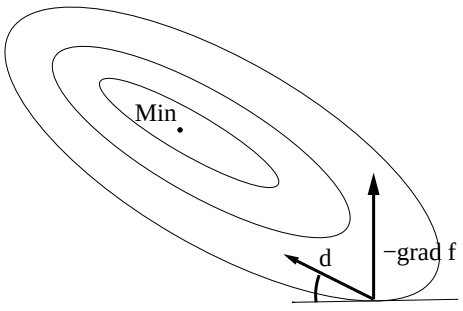
$$c_4 \quad \| \nabla f(x) \| \, \geq \, \| d \| \geq \, \frac{1}{c_4} \| \nabla f(x) \| \tag{4.20}$$

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*

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

Illustration der Gradienten-Bezogenheit:



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

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

(VLK) (Lokal gleichmäßige Konvexität)
Sei $f : \mathbb{R}^n \to \mathbb{R}, N(f, f(x^{(0)})) \subset D, \neq \emptyset$, 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*
- $h^T f''(x)h \geq \alpha_2 \| h \| \quad \forall h \in R^n, \quad \forall x \in N(f, f(x^{(0)}))$
- $\| f''(x) \| \leq \alpha_2 \qquad -''-$
- $\| f''(x)^{(-1)} \| \leq \beta_2 := 1/\alpha_1 \qquad -''-$
- $\beta_1 \| h \|^2 \leq h^T f''(x)^{(-1)} h \leq \beta_2 \| h \|^2, \quad \forall h \in R^n, -''-$
- f *ist gleichmäßig konvex auf* D.

Beispiel 4.3.3 *(Newton-Richtung)*
(VLK) *sei erfüllt,* $x \in N(f, f(x^{(0)}))$,

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

Und

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

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

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

4.3.3 Allgemeine Konvergenzsätze

Folgende Voraussetzungen werden im Weiteren oft benötigt:

(VNK) Für ein gegebenes $x^{(0)} \in \mathbb{R}^n$ ist die Niveaumenge
$$N\left(f, f(x^{(0)})\right) = \left\{x | f(x) \leq f(x^{(0)})\right\}$$
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 $\nabla f(x^{(k)}) \to 0, k \to \infty$ 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: $x^{(k)} \to \tilde{x}$. 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) $\Rightarrow$

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

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

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

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 $\nabla f(x^{(k)}) \to 0$ gilt auch

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

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

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

Induktiv folgt schließlich $x^{(k)} \to \tilde{x}$. $\square$

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

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

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*

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

und

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

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

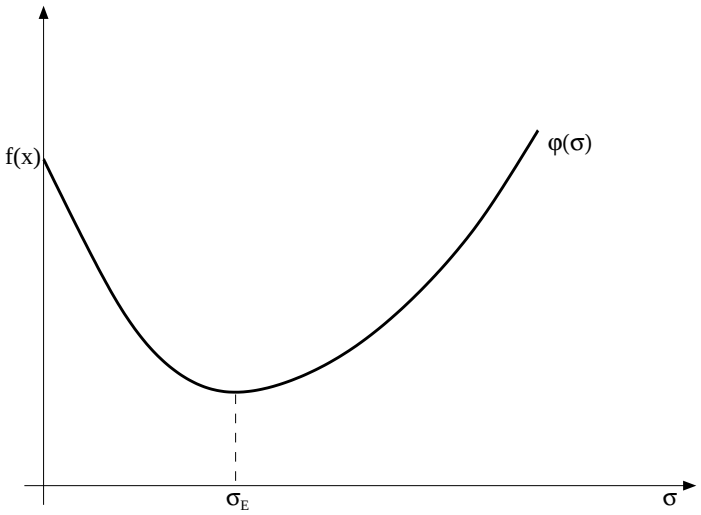
$$\min_{s\geq 0} f(x+sd) = \min_{s\geq 0} \varphi(s) = \varphi(\sigma).$$

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:

$$\begin{aligned} 0 \underset{\text{Def von } \sigma_E}{\uparrow =} \nabla f(x+\sigma_E d)\cdot d \underset{\uparrow}{=} & \nabla f(x)\cdot d + [\nabla f(x+\sigma_E d) - \nabla f(x)]\cdot d \\ \underset{\text{Lipschitzbed.}}{\uparrow \leq} & \nabla f(x)\cdot d + \sigma_E L \parallel d \parallel^2 \end{aligned}$$

$$\Rightarrow \boxed{\sigma_E \geq \tilde{\sigma} = -\frac{\nabla f(x)\cdot d}{L\|d\|^2}} \tag{4.24}$$

Außerdem bekommt man den Mindestabstieg

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

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

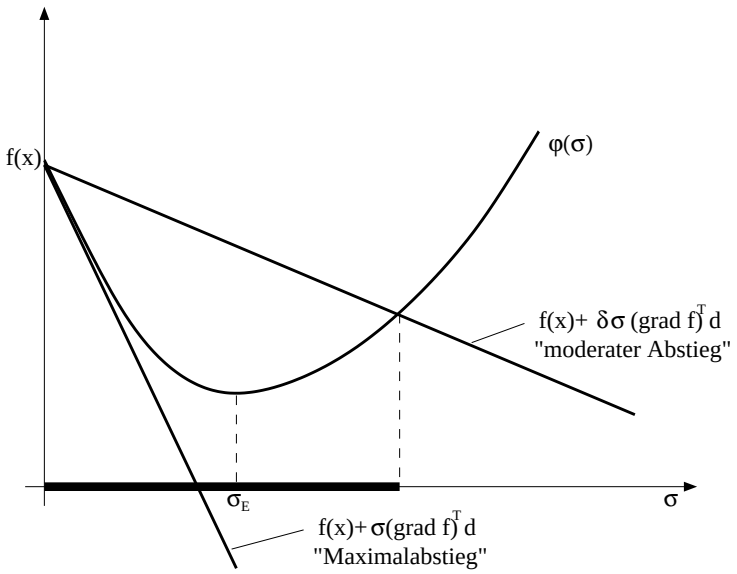
$$f(x) = \frac{1}{2}x^T H x + b^T x$$

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 $\sigma = \sigma_A$

- $f(x + \sigma_A d) \leq f(x) + \delta\sigma_A \nabla f(x)^T d$ *Mindestabstieg* (4.26)
- $\sigma_A \geq -c_2 \dfrac{\nabla f(x)^T d}{\|d\|^2}$ *Effizienz* (4.27)



Verfahren 4.4.1 `(Armijo-Goldstein)`

0. Fixiere Konstanten

$$\begin{array}{ll} 0 < \delta < 1 & \textit{“Abflachung”} \\ \gamma > 0 & \textit{“Effizienzkonst.”} \\ 0 < \beta_1 \leq \beta_2 < 1 & \end{array}$$

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

$$\underbrace{f(x+sd) - f(x)}_{\substack{=\nabla f(x+\vartheta sd)^T ds \\ 0<\vartheta<1}} > \delta s \nabla f(x)^T d.$$

$$\Rightarrow \nabla f(x+\vartheta sd)^T d = \tfrac{1}{s}[f(x+sd) - f(x)] > \delta \nabla f(x)^T d \,|\, -\nabla f(x)^T d$$

$$\Rightarrow -(1-\delta)\nabla f(x)^T d < [\nabla f(x+\vartheta sd) - \nabla f(x)]^T d \underset{\substack{\uparrow \\ \text{Lipsch.}}}{\leq} L\vartheta s \,\| d \|^2 \leq sL \,\| d \|^2$$

$$\Rightarrow \boxed{s \geq -\tfrac{(1-\delta)}{L} \tfrac{\nabla f(x)^T d}{\| d \|^2}}$$

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

$$\sigma_A \geq -\underbrace{\frac{\beta_1(1-\delta)}{L}}_{c_2} \frac{\nabla f(x)^T d}{\| d \|^2}$$

$$c_2 = \min\left\{\gamma, \frac{\beta_1(1-\delta)}{L}\right\}$$

□

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

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

und

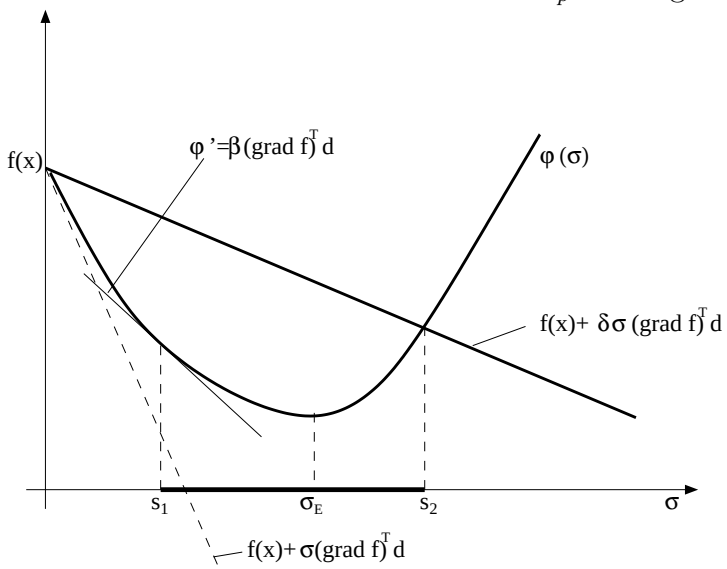
$$\nabla f(x+\sigma d)^T d \geq \beta\nabla f(x)^T d \quad \text{(Mindestschrittweite)} \tag{4.29}$$

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

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

Verfahren 4.4.2 *(Powell)*

a) 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

$$a_0 := \sigma_0$$
$$b_0 := 2^l \sigma_0 \text{ mit } \underline{\text{minimalem}} \ l \in \mathbb{N}, \text{ so dass } G_1(b_0) > \delta$$

Gehe zu 2.

(iii) Liegt σ_0 *in* I_3

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

b) 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*:*

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

c) $j := j + 1$*, 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

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

- Sehr einfach zu implementieren
- Am Ende aber zu langsam

Verfahren 4.5.1

a) Startvektor $x^{(0)}, k := 0$, *Abbruchkriterium* $\varepsilon > 0$.

b) Wenn $\| \nabla f(x^{(k)}) \| < \varepsilon$*: Fertig*

c) Berechne

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

□

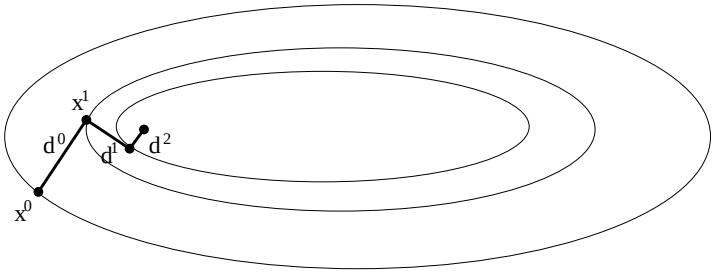
Unter den entsprechenden Voraussetzungen gelten die allgemeinen Konvergenzsätze.

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

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

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

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

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

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

c) Berechne $d^{(k)}$ *aus*

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

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 $\mathbb{R}^n$:

$$\langle x, y\rangle_A := x^T A y$$
$$\| x \|_A := \sqrt{\langle x, x\rangle_A} = \left(x^T A\, x\right)^{1/2}$$

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*

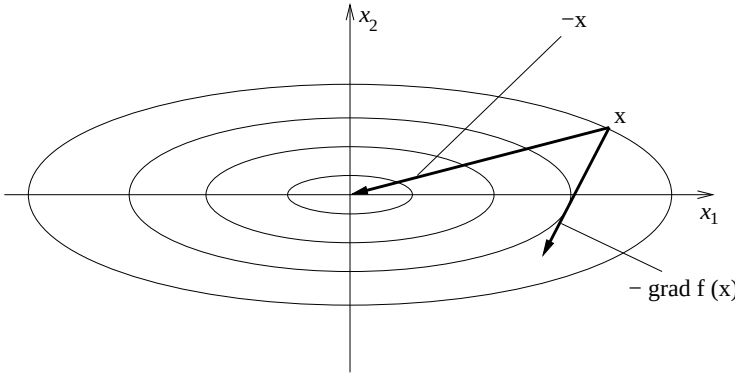
$$\begin{aligned} &\min \nabla f(x)^T d \\ &\text{bei} \quad \| d \|_A = 1, \end{aligned} \tag{4.30}$$

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 $H = \left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right)$ 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 gem. Konvexitätsbedingung (VLK) sowie die gem. Lipschitzstetigkeit von f'' in $N(f, f(x^{(0)}))$ erfüllt, d. h. $\| f''(x) - f''(y) \| \leq L \| x - y \| \quad \forall x, y \in N(\ldots).$ (VL2)

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

Nach dem allgemeinen Konvergenzsatz 4.3.3 konvergiert 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 $\sigma_k, 0 = 1$ gewählt wird. Weiter sei $0 < \delta < 1/2$. Dann gilt für alle hinreichend großen k die Bedingung $\sigma_k = 1$. Daher superlineare Konvergenz. Gilt zusätzlich (VL2), dann quadratische Konvergenz.*

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*

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

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

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

Verwendet man die exakte Schrittweite, dann gilt $\sigma^k \to 1,\ k \to \infty$, 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)`

a) Start $x^{(0)} \in \mathbb{R}^n$, $k := 0$

b) Wenn $\nabla f(x^{(k)}) = 0$: *Fertig*

c) 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

$$\begin{aligned} f''\left(x^{(k+1)}\right)\left(x^{(k+1)} - x^{(k)}\right) &= H\left(x^{(k+1)} - x^{(k)}\right) \pm b^T \\ &= \nabla f\left(x^{(k+1)}\right) - \nabla f\left(x^{(k)}\right),\ k = 0, 1, 2 \ldots \end{aligned}$$

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

$$H\left(x^{(k+1)} - x^{(k)}\right) = \nabla f\left(x^{(k+1)}\right) - \nabla f\left(x^{(k)}\right),\ k = 0, \ldots, n-1$$

die Matrix H eindeutig.

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

$$\boxed{A^{(k+1)}\left(x^{(k+1)} - x^{(k)}\right) = \nabla f\left(x^{(k+1)}\right) - \nabla f\left(x^{(k)}\right).} \tag{4.31}$$

(4.31) heißt

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:

$$x^{(k+1)} - x^{(k)} =: s^{(k)}$$
$$\nabla f\left(x^{(k+1)}\right) - \nabla f\left(x^{(k)}\right) =: y^{(k)}$$

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

- Zuerst

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

  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

$$\operatorname{rang}\left(A^k s^{(k)}\right)\left(A^{(k)}s^{(k)}\right)^T = 1,$$

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

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

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

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

Quasi-Newton-Gleichung:

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

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

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

Positive Definitheit:

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

$$0 < \left(s^{(k)}\right)^T A^{(k+1)} s^{(k)} = \left(s^{(k)}\right)^T y^{(k)} \tag{4.33}$$

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:

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

Da die Summe von zwei Rang 1-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:

$$f(x) = \frac{1}{2}x^T H x \qquad \nabla f = Hx$$

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

□

Nebenrechnungen:

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

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

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

$$\begin{aligned} ss^T = (s_i s_j) &= \begin{pmatrix} s_1s_1 & s_1s_2 & \dots & s_1s_n \\ s_2s_1 & s_2s_2 & \dots & s_2s_n \\ \vdots & & & \\ s_ns_1 & s_ns_2 & \dots & s_ns_n \end{pmatrix} \\ &= \left(s_1 \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}, s_2 \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}, \dots, s_n \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} \right) \end{aligned}$$

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))`

a) Wähle Startvektor $x^{(0)}$, symmetrische positiv definite Startmatrix $A^{(0)}, k := 0$.

b) Wenn $\nabla f(x^{(0)}) = 0$: Fertig.

c) Berechne

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

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 $d^{(0)},\ldots,d^{(k)}, k<n$, heißen zueinander **konjugiert** *bzw.* **orthogonal** *bezüglich H, wenn sie nicht Null sind und*

$$\boxed{\left(d^{(i)}\right)^T H d^{(j)} = 0, \qquad 0 \le i < j \le k}$$

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 $m \le n$ 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.

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

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

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

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

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

Damit für $k=1$ erhalten:

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

Induktion $k \to k+1 \ldots$ liefert die Behauptung. $\square$

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 nichtlineare 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 <u>nicht</u> notwendig $A^{(k)} \to f''(\tilde{x})$, sondern

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

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

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

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

$$x^{k+1} = x^k + \sigma_k d^k$$
$$\sigma^k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T H d^k} \quad (\textit{exakte Schrittweite})$$

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

Beweis.

$$\tilde{x} - x^0 = \sum_{i=0}^{n-1} \sigma_i d^i \quad (\text{Mult. von links mit } (d^i)^T)\cdot H$$
$$\Rightarrow \quad \sigma_k = \frac{(d^k)^T H(\tilde{x} - x^0)}{(d^k)^T H d^k} = -\frac{(d^k)^T(Hx^0 + b)}{(d^k)^T H d^k}$$
$$\cdots = -\frac{(d^k)^T(Hx^k + b)}{(d^k)^T H d^k} = -\frac{(d^k)^T \nabla f(x^k)}{(d^k)^T H d^k}$$

□

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 $V_k = span\{d^0, ..., d^{k-1}\}$. Insbesondere gilt*

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

Beweis. Es genügt, $(*)$ zu zeigen.

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

a) Wähle x^0*, berechne* $d^0 = -\nabla f(x^0) = -(Hx^0 + b)$
$k := 0$

b) Wenn $\nabla f(x^k) = 0 \quad \rightarrow \quad$ *fertig*

c) Berechne

$$\begin{aligned}
\sigma_k &= \frac{\nabla f(x^k)^T \nabla f(x^k)}{(d^k)^T H d^k} \\
x^{k+1} &= x^k + \sigma_k d^k \\
\nabla f(x^{k+1}) &= Hx^{k+1} + b = \nabla f(x^k) + \sigma_k H d^k \\
\beta_k &= \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2} \\
d^{k+1} &= -\nabla f(x^{k+1}) + \beta_k d^k \,.
\end{aligned}$$

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

(2)

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

(3) $d^0, ..., d^{k-1}$ *sind paarweise konjugiert*

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

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

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

Nach Induktionsvoraussetzung sind $d^0, ..., d^{k-1}$ konjugiert.
Folgerung 4.1 $\Rightarrow$

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

$g^k \neq 0 \Rightarrow \{d^0, ..., d^{k-1}, g^k\}$ und damit auch $\{g^0, ..., g^{k-1}, g^k\}$ sind linear unabhängig mit Dimension $k+1$
$\Rightarrow \text{span}\{g^0, ..., g^k\} = V_{k+1}$.

Es gilt $g^k + d^k = \beta_{k-1} d^{k-1} \in V_k$
$\Rightarrow V_{k+1} = \text{span}\{d^0, ..., d^k\} \Rightarrow (2)$.

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 $\lambda_1 < \ldots < \lambda_n$ 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

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

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

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

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

Sei $\{z_1, ..., z_n\}$ 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

$$\alpha := \min_{\substack{Q_k \in \mathcal{P}_k \\ Q_k(0)=1}} \max_{1 \leq i \leq n} |Q_k(\lambda_i)| \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k$$

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

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. $(.,.)_B =$ denn

$$(x, HBy)_B = x^T BHBy = (HBx)^T By = (HB, y)_B$$

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

[1] Deuflhard/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 $D = diag(H)$ 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]).

5 Probleme mit linearen Restriktionen – Theorie

5.1 Ein Beispiel

Wir nehmen als Beispiel ein Lagerhaltungsproblem

- Firma verkauft Produkt
- Verkauf wird zu diskreten Zeitpunkten $t_0 < t_1 < \ldots < t_N$ beobachtet
- Betrieb eines Lagers, Belieferung am Anfang jeder Periode $[t_i, t_{i+1}]$
  Ziel: Steuerung der Lagerhaltung des Produkts, um minimale Gesamtkosten zu haben
- Größen: z_i Lagerbestand bei t_i vor Neulieferung
  r_i Nachfrage nach Produkt in $[t_i, t_{i+1}]$
  u_i Liefermenge zum Zeitpunkt t_i
  $z_0 = a \geq 0$ ist vorgegeben (Anfangsbestand)

Lagerbilanzgleichung

$$z_{i+1} = z_i - r_i + u_i \quad i = 0, \ldots, N-1.$$

Gegeben angenommen: $r_1, \ldots, r_N$.
Gesuchte Variablen: $z = (z_1, \ldots, z_N)^T$, $u = (u_0, \ldots, u_{N-1})^T$

$$x := \begin{pmatrix} z \\ u \end{pmatrix}.$$

Kosten: $f_i(z_i, u_i)$ (Liefern, Einkaufen, Lagern)

Zielfunktion:

$$f(z,u) \;=\; \rho z_N^2 + \sum_{i=0}^{N-1} f_i(z_i, u_i)$$

minimaler Endbestand

Wichtungsfaktor $\rho \geq 0$

Am Ende ergibt sich folgendes Problem:

$$\text{(LH1)} \begin{cases} \min f(z,u) = \rho z_N^2 + \sum_{i=0}^{N-1} f_i(z_i, u_i) & & \\ \text{bei} & & \\ 0 \leq u_i \leq b_i & i = 0, \ldots, N-1 & \text{Kapazitätsschranken} \\ z_{i+1} = z_i - r_i + u_i & i = 0, \ldots, N-1 & \text{Lagerbilanz} \\ z_i \geq 0 & i = 1, \ldots, N & \text{Nichtnegativität} \end{cases}$$

Dabei ist $z_0 = a$ vorgegeben.

Das Problem hat die allgemeine Form

$$\text{(P)} \quad \boxed{\min f(x), \quad x \in \mathcal{F}},$$

mit der zulässigen Menge

$$\mathcal{F} = \left\{ x = \binom{z}{u} \,|\, z \geq 0,\, 0 \leq u \leq b,\, z_{i+1} = z_i - r_i + u_i,\, a = 0, \ldots, N-1 \right\}$$

Verallgemeinerung: Bisher waren z_i, u_i reelle Variablen. Bei mehreren Produkten können das Vektoren sein. Das ergibt schließlich die Aufgabe

$$\begin{aligned} \text{(LH2)} \quad \min \quad & f(z,u) = \rho \,\|\, z_N \,\|^2 + \sum_{i=0}^{N-1} f_i\,(z_i, u_i) \\ & 0 \leq u_i \leq b_i \\ & z_{i+1} = A_i z_i + B_i u_i - r_i \\ & z_i \geq 0 \end{aligned}$$

mit $z_i \in \mathbb{R}^n, u_i \in \mathbb{R}^n, r_i \in \mathbb{R}^n$ und entsprechenden Matrizen A_i, B_i.

5.2 Optimalitätsbedingungen erster Ordnung

Wir brauchen zunächst einige Grundlagen und Hilfsmittel der konvexen Analysis.

Definition 5.2.1 *(**Kegel**) Eine nichtleere Teilmenge $K \subset \mathbb{R}^n$ heißt Kegel, wenn*

$$x \in K \Rightarrow \alpha x \in K \quad \forall \alpha > 0,$$

Beispiel 5.2.1

- $\{x \in \mathbb{R}^n \,|\, x_i > 0 \quad \forall_i\}$
- $\{x \in \mathbb{R}^n \,|\, x_i \geq 0 \quad \forall_i\}$ *analog abgeschlossen (nichtnegativer Orthant)*
- $\{x \in \mathbb{R}^2 \,|\, x_1 \geq 0 \wedge x_2 = 0 \quad \textit{oder} \quad x_1 = 0 \wedge x_2 \geq 0\}$

Bezeichnungen:

a) $x \geq 0 \Leftrightarrow x_i \geq 0 \quad \forall_i = 1, \ldots, n$

b) $A, B \subset \mathbb{R}^n$ seien Mengen, $\alpha, \beta \in \mathbb{R}$

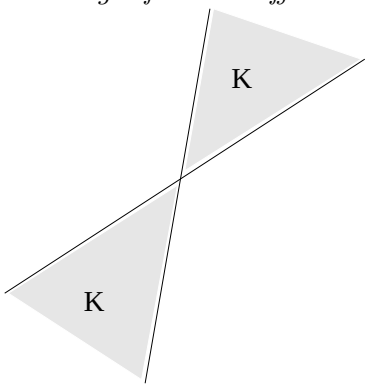
$$\alpha A + \beta B := \{\alpha a + \beta B \,|\, a \in A,\, b \in B\}$$

Lemma 5.2.1 *(Konvexitätskriterium)*
Ein Kegel $K \subset \mathbb{R}^n$ ist genau dann konvex, wenn

$$K + K \subset K$$

Beweis: Ü.A.

Beispiel 5.2.2 *Ein nichtkonvexer Kegel für den offenbar gilt* $K + K = \mathbb{R}^2 \not\subset K$



Nach dieser allgemeinen Kegelei kommen nun die Kegel, welche für die Optimierungstheorie von ausschlaggebender Bedeutung sind:

Definition 5.2.2 *(Konische Hülle)*
Es sei $S \subset \mathbb{R}^n$ *und* $x \in S$ *fest.*

$$K(S,x) = \{\alpha(s-x) \,|\, s \in S, \alpha > 0\}$$

heißt der von $s - x$ **erzeugte Kegel** *oder* **konische Hülle** *von* $s - x$.
Andere Schreibweise:

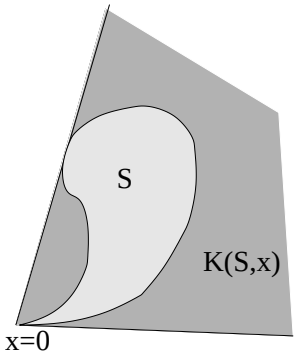
$$\bigcup_{\alpha>0} \alpha(S-x)$$

.

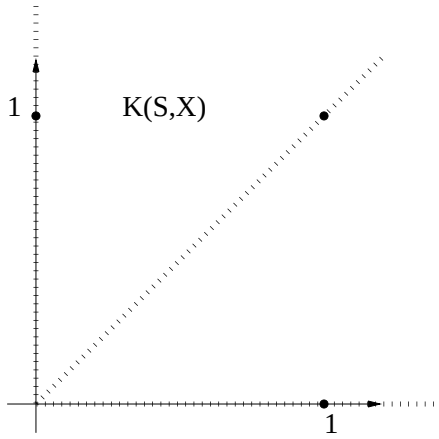
Lemma 5.2.2 *Ist* $C \subset \mathbb{R}^n$ *konvex sowie* $x \in C$*, dann ist auch* $K(C,x)$ *konvex.*

Beispiel 5.2.3

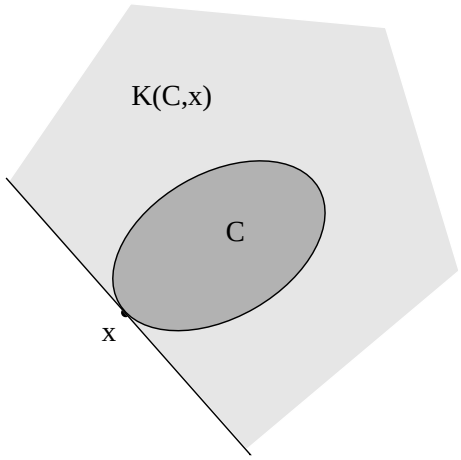
a)



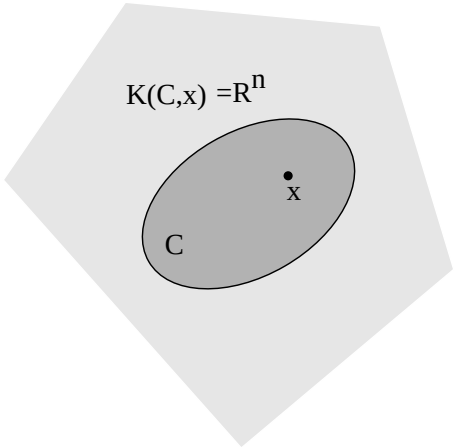
b) $S = \{(1,0)^T, (0,1)^T, (1,1)^T\}$



c)



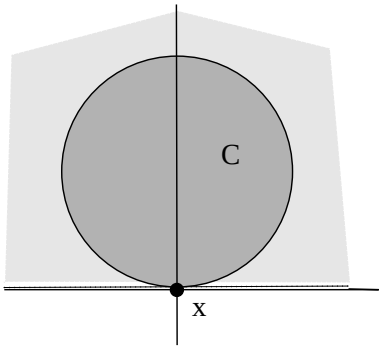
d)



e) Auch bei abgeschlossener Menge C muss $K(C,x)$ nicht abgeschlossen sein:

$$C = \bar{B}\left(\begin{pmatrix}0\\1\end{pmatrix}, 1\right), \quad x = \begin{pmatrix}0\\0\end{pmatrix}$$

$$K(C,x) = \left\{y \in \mathbb{R}^2 \,|\, y_2 > 0\right\} \cup \left\{\begin{pmatrix}0\\0\end{pmatrix}\right\}$$



f) $C = R^+ = \{x \,|\, x \geq 0\}, \quad x \in C$
$\Rightarrow K(C,x) = \{d \in \mathbb{R}^n \,|\, d_i \geq 0, \ wenn\, x_i = 0\}$

g) $a^i \in \mathbb{R}^n$, $i = 1, \ldots, m$, $m \leq n$, $b \in \mathbb{R}^m$ *fest*
$C = \{x \in R^n \mid \langle a^i, x\rangle = b_i, \ i = 1, \ldots, m\}$
Wählen wir $A = \begin{pmatrix} a_1^T \\ \vdots \\ a_{m^T} \end{pmatrix}$, *so*

$$C = \{x \,|\, Ax = b\}$$

Hier gilt

$$K(C,x) = \{y \,|\, Ay = 0\}, \ falls\, x \in C \qquad (\ddot{U}bungsaufgabe)$$

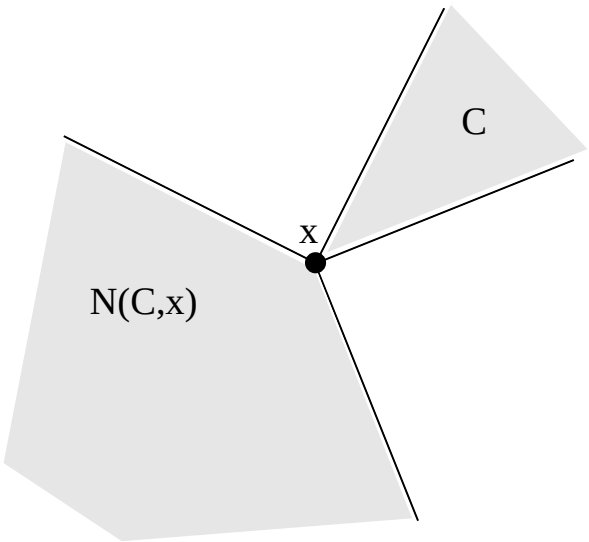
Definition 5.2.3 *(Normalenkegel)*
Sei $C \subset \mathbb{R}^n$ *konvex,* $x \in C$. *Ein Vektor* $s \in \mathbb{R}^n$ *heißt Normalenrichtung von* C *in* x, *wenn*

$$\langle s, y - x\rangle \leq 0 \quad \forall y \in C.$$

Die Menge

$$N(C,x) = \{s \,|\, s\, ist\ Normalenrichtg.\}$$

heißt Normalenkegel von C *in* X.



Bemerkung:

- $N(C,x)$ ist stets abgeschlossen.
- $x \in intC \Rightarrow N(C,x) = \{0\}$

Definition 5.2.4 *(Dualkegel)*
$K \subset \mathbb{R}^n$ sei konvexer Kegel. Dann heißt

$$K^* = \{s \in R^n \,|\, \langle s, x\rangle \leq 0 \quad \forall x \in K\}$$

Dual- oder Polarkegel *zu K.*

Bemerkung:

- $0 \in K \Rightarrow K^* = N(K,0)$ (s. oben)
- K_1, K_2 konvexe Kegel $\Rightarrow K_1^* \supset K_2^*$
- K^* ist immer konvex und abgeschlossen

Satz 5.2.1 *Ist $C \subset \mathbb{R}^n$ konvex und $x \in C$, so gilt*

$$N(C,x) = K(C,x)^*.$$

Beweis:

(i) $''\subset''$: Sei $s \in N(C,x)$, d. h.

$$\begin{array}{ll} & \langle s, y-x\rangle \leq 0 \quad \forall y \in C \\ \Rightarrow & \langle s, \alpha(y-x)\rangle \leq 0 \quad \forall y \in C, \alpha > 0 \\ \Rightarrow & \langle s, z\rangle \leq 0 \quad \forall z \in K(C,x) \qquad\qquad (*) \end{array}$$

und damit $s \in K(C,x)^*$.

(ii) $''\supset''$: Sei $s \in K(C,x)^*$, d. h. $(*)$ gilt. Damit insbesondere für $z = 1 \cdot (y-x) \Rightarrow s \in N(C,x)$. □

Unscheinbar, aber dennoch grundlegend ist wiederum

Satz 5.2.2 *$\mathcal{F}$ sei nichtleer und konvex, f in $\tilde{x} \in \mathcal{F}$ differenzierbar und $\tilde{x}$ lokales Minimum von (P). Dann gilt*

$$\boxed{\nabla f(\tilde{x})^T(x-\tilde{x}) \geq 0 \quad \forall x \in \mathcal{F}} \quad \textit{(Variationsungleichung)} \tag{5.35}$$

Beweis: Folgt aus Satz 3.1.2 wie die Variationsungleichung (3.7) □

Alternative Darstellung von (5.35):

Setze $s = \nabla f(\tilde{x})$. Dann gilt $\langle s, x - \tilde{x}\rangle \geq 0 \quad \forall x \in \mathcal{F}$, also $\langle -s, x - \tilde{x}\rangle \leq 0$. Damit

$$\begin{array}{lll} & \nabla f(\tilde{x} \in N(\mathcal{F}, \tilde{x}) \quad \text{oder} \quad \nabla f(\tilde{x}) \bar{\in} N(\mathcal{F}, \tilde{x}) & \\ \text{oder} & \boxed{0 \in \nabla f(\tilde{x}) + N(\mathcal{F}, \tilde{x})} \quad (\text{Verallgemeinerung von } 0 = \nabla f) & (5.36) \\ \text{oder auch} & \nabla f(\tilde{x}) \in -K(\mathcal{F}, \tilde{x})^*. & \end{array}$$

Bemerkung:
Ist f auch konvex, so ist (P) eine konvexe Optimierungsaufgabe.

Dann ist (5.36) auch hinreichend für Optimalität.

Beispiel 5.2.4 *(Vorzeichenbedingungen)*
Wir betrachten das Problem

$$\boxed{\min_{x \in \mathbb{R}^n} f(x), \quad x \geq 0} \tag{PV}$$

Die lokale Lösung sei $\tilde{x} \in \mathcal{F} = \mathbb{R}^+$.

$$\begin{array}{rll} \Rightarrow & K(\mathcal{F}, \tilde{x}) = & \left\{ d \in \mathbb{R}^n \mid d_i \geq 0, \, falls \, \tilde{x}_i = 0 \right\} \\ & N(\mathcal{F}, \tilde{x}) = & \left\{ s \in \mathbb{R}^n \mid s_i \leq 0, \, wenn \quad \tilde{x}_i = 0, \, s_i = 0, \, wenn \quad \tilde{x}_i > 0 \right\} \end{array}$$

Aus Satz 5.2.2 folgt also

$$\begin{array}{rl} -\nabla f(\tilde{x})_i \leq 0, & wo \quad \tilde{x}_i = 0 \\ \nabla f(\tilde{x})_i = 0, & wo \quad \tilde{x}_i > 0. \end{array}$$

Das hätten wir aber auch direkt aus der Variationsungleichung herleiten können:

$$\nabla f(\tilde{x})^T (x - \tilde{x}) \geq 0 \Leftrightarrow \sum_{i=1}^{n} \nabla f(\tilde{x})_i (x_i - x_i^2) \geq 0$$

und wegen Unabhängigkeit der Komponenten

$$\begin{array}{rl} & \nabla f(\tilde{x})_i (x_i - \tilde{x}_i) \geq 0 \quad \forall x_i \geq 0 \quad \forall i \\ d.\, h. & \nabla f(\tilde{x})_i \cdot \tilde{x}_i \leq \nabla f(\tilde{x})_i \cdot x \quad \forall x \geq 0 \end{array}$$

Damit muss $\tilde{x}_i$ *das Minimum der rechten Seite annehmen unter allen* $x \geq 0$. *Damit*

$$\left.\begin{array}{l} \nabla f(\tilde{x})_i \geq 0\,, \quad falls \quad \tilde{x}_i = 0 \\ \nabla f(\tilde{x})_i = 0\,, \quad falls \quad \tilde{x}_i > 0 \end{array}\right\} \begin{array}{l} Kann \, numerisch \\ als \;\; Abbruchkrit. \\ genutzt \, werden. \end{array}$$

Es folgt auch: $\nabla f(\tilde{x})_i > 0 \Rightarrow \tilde{x}_i = 0$.

Beispiel 5.2.5 Lineare Gleichungsnebenbedingungen

$$\boxed{\min_{x\in\mathbb{R}^n} f(x), \quad \forall x = b\,.} \qquad A : (m,n) - \textit{Matrix} \tag{PLG}$$

$$\textit{Hier: } \mathcal{F} = \{x \mid Ax = b\}$$

Wir hatten bereits erwähnt, dass hier gilt

$$K(\mathcal{F}, \tilde{x}) = \{x \mid Ax = 0\} = U \quad \textit{linearer Unterraum}$$

Offenbar gilt dann

$$K(\mathcal{F}, \tilde{x})^* = N(\mathcal{F}, \tilde{x}) = U^\perp.$$

Wegen $U = \ker A$ *folgt* $U^\perp = \textit{im}\,\{A^T\lambda \mid \lambda \in \mathbb{R}^m\}$

Wegen der Variationsungleichung in der Form (5.36):

$$-\nabla f(\tilde{x}) \in N(\mathcal{F}, \tilde{x})$$

folgt

$$\begin{array}{l} -\nabla f(\tilde{x}) \in \textit{im}\, A^T \\ -\nabla f(\tilde{x}) = A^T\lambda \\ \boxed{0 = \nabla f(\tilde{x}) + A^T\lambda} \end{array}$$

(Klassische Multiplikatorenregel nach Lagrange mit Lagrange Multiplikator λ*. Wird später noch allgemeiner diskutiert!)*

5.3 Optimalitätsbedingungen zweiter Ordnung

5.3.1 Notwendige Bedingungen

Wir beweisen nun ein Analogon zum entsprechenden Satz für unrestringierte Aufgaben.

Satz 5.3.1 *Es sei* $\tilde{x} \in \mathcal{F}$ *lokales Minimum von (P) und* $\mathcal{F}$ *konvex. Dann gilt neben der notwendigen Bedingung 1. Ordnung auch die notwendige Bedingung 2. Ordnung*

$$(x-\tilde{x})^T f''(\tilde{x})(x-\tilde{x}) \geq 0 \quad \forall x \in \mathcal{F} \textit{ mit } \nabla f(\tilde{x})^T(x-\tilde{x}) = 0 \tag{5.37}$$

Beweis: Wir setzen $d = x - \tilde{x}$ und fordern $\nabla f(\tilde{x})^T d = 0$. Dann gilt $\tilde{x} + td \in \mathcal{F}$ für alle $t \in [0,1]$ und

$$\begin{array}{rl} & f(\tilde{x}+td) - f(\tilde{x}) \geq 0 \quad \text{für } t \text{ hinreichend klein} \\ \Rightarrow & 0 \leq \underbrace{\nabla f(\tilde{x})td}_{=0} + \frac{1}{2}t^2 d^T f''(\tilde{x})d + o(t^2) \quad | : \frac{1}{2}t^2 \\ & 0 \leq d^T f''(\tilde{x})d + 2\underbrace{\frac{o(t^2)}{t^2}}_{\to 0,\ t\downarrow 0} \qquad t \downarrow 0 \text{ ergibt die Behauptung} \end{array}$$

□

Bemerkung: Man kann offenbar (5.37) auch so schreiben:

$$\boxed{d^T f''(\tilde{x})d \geq 0 \quad \forall d \in K(\mathcal{F}, \tilde{x}) \,\text{mit}\, \nabla f(\tilde{x})^T d = 0.} \tag{5.38}$$

(man multipliziere (5.37) mit $\alpha > 0$ durch)

Beispiel 5.3.1 *(Gleichungsnebenbedingungen)*

Wir wissen: Hier gilt

$$\begin{array}{rl} & K(\mathcal{F}, \tilde{x}) = U \\ \Rightarrow & d^T f''(\tilde{x})d \geq 0 \quad \forall d \in U \; mit \nabla f(\tilde{x})^T d = 0. \end{array} \tag{5.39}$$

Außerdem wissen wir

$$-\nabla f \in N(\mathcal{F}, \tilde{x}) = U^{\perp},$$

damit gilt automatisch $\nabla f(\tilde{x})^T d = 0$ *und somit*

$$\boxed{d^T f''(\tilde{x})d \geq 0} \qquad \textit{für alle } d \in U, \quad \textit{d. h. für alle } d \textit{ mit} \quad \boxed{Ad = 0}$$

$\Rightarrow f''(\tilde{x})$ *muss auf dem linearen Unterraum* U *positiv semidefinit sein.*

5.3.2 Hinreichende Bedingungen

Satz 5.3.2 *Die Funktion* f *sei in* $\tilde{x} \in \mathcal{F}$ *zweimal stetig differenzierbar und die notwendige Bedingung 1. Ordnung sei erfüllt. Zusätzlich gelte*

$$d^T f''(\tilde{x})d \geq \alpha \,\|\, d \,\|^2 \tag{5.40}$$

für alle $d \in cl\, K(\mathcal{F}, \tilde{x})$ *mit* $\nabla f(\tilde{x})^T d = 0$ *mit einem* $\alpha > 0$*. Dann existiert zu jedem* $\beta \in (0, \alpha/2)$ *ein* $\varrho > 0$*, so dass*

$$f(x) > f(\tilde{x}) + \frac{\beta}{2}\|x - \tilde{x}\|^2 \quad \textit{(}\text{quadratische Wachstumsbedingung}\textit{)}$$

für alle $x \in \mathcal{F} \cap B(\tilde{x}, \varrho)$*,* $x \neq \tilde{x}$*. Damit ist* $\tilde{x}$ *striktes lokales Minimum.*

Beweis: (Indirekt). Die Behauptung sei falsch. Dann existiert ein $\beta \in (0, \alpha/2)$ und eine Folge $\{x^i\} \subset \mathcal{F}$ mit $x^i \to \tilde{x}$, $x^i \neq \tilde{x}$ $\forall i \in \mathbb{N}$ und

$$f(x^i) \leq f(\tilde{x}) + \frac{\beta}{2}\,\|x^i - \tilde{x}\|^2 \quad \forall i \in \mathbb{N}\,. \tag{$*$}$$

Wir zeigen, dass daraus $\beta \geq \alpha$ folgt, also ein Widerspruch.

- Folge $d^i = \frac{x^i - \tilde{x}}{\|x^i - \tilde{x}\|}$ durchläuft die (kompakte) Einheitskugeloberfläche. Daher existiert eine Teilfolge – o.B.d.A. sei das d^i selbst – mit $d^i \to d$. Dann gilt $\|d\| = 1$.

- Außerdem gilt die Variationsungleichung für $\tilde{x}$, also

$$\langle \nabla f(\tilde{x}), x^i - \tilde{x} \rangle \geq 0\,, \quad \text{daher } \langle \nabla f(\tilde{x}), d^i \rangle \geq 0 \quad \text{und so } \nabla f(\tilde{x})^T d \geq 0\,.$$

Außerdem $d \in cl\, K(\mathcal{F}, \tilde{x})$, da $x^i - \tilde{x} \in K \;\; \forall i$.

Andererseits muss die Ungleichung $\nabla\,(\tilde{x})^T\, d \leq 0$ gelten, denn mit Taylorentwicklung in $(*)$:

$$\begin{aligned} f(\tilde{x}) + \frac{\beta}{2}\,\|x^i - \tilde{x}\|^2 &\geq f(x^i) \\ &= f(\tilde{x}) + \nabla f(\tilde{x})^T (x^i - \tilde{x}) + r_1(x^i - \tilde{x})\,. \end{aligned}$$

Wir teilen durch $\|x^i - \tilde{x}\|$, vorher streichen wir $f(\tilde{x})$ auf beiden Seiten. Dann

$$\frac{\beta}{2}\,\underbrace{\|x^i - \tilde{x}\|}_{\to 0} \geq \nabla f(\tilde{x})^T\, \underbrace{\frac{x^i - \tilde{x}}{\|x^i - \tilde{x}\|}}_{\to d} + \underbrace{\frac{r_1(x^i - \tilde{x})}{\|x^i - \tilde{x}\|}}_{\to 0\,,\; x^i \to \tilde{x}}$$

$\Rightarrow \nabla f(\tilde{x})^T d \leq 0.$
Damit

$$\nabla f(\tilde{x})^T d \;\; = \;\; 0\,. \tag{5.41}$$

- Nun gehen wir nochmals in $(*)$, entwickeln aber bis Ordnung 2,

$$\begin{aligned} f(\tilde{x}) + \frac{\beta}{2}\,\|x^i - \tilde{x}\|^2 &\geq f(x^i) \\ &= f(\tilde{x}) + \underbrace{\nabla f(\tilde{x})^T (x^i - \tilde{x})}_{\geq 0\,,\ \text{Variations-ungleichung}} + \frac{1}{2}\,(x^i - \tilde{x})^T f''(\tilde{x})(x^i - \tilde{x}) + r_2(x^i - \tilde{x}) \end{aligned}$$

Wir teilen durch $\|x^i - \tilde{x}\|^2$,

$$\frac{\beta}{2}\|d^i\|^2 \geq \frac{1}{2}(d^i)^T f''(\tilde{x}) d^i + \underbrace{\frac{r_2(x^i - \tilde{x})}{\|x^i - \tilde{x}\|^2}}_{\to 0}$$

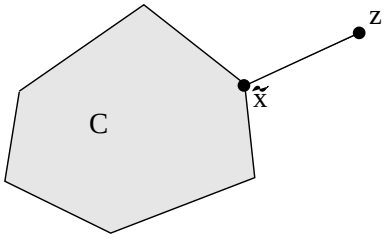
$i \to \infty \Rightarrow$

$$\frac{\beta}{2}\|d\|^2 \geq \frac{1}{2} d^T f''(\tilde{x}) d \geq \frac{\alpha}{2}\|d\|^2$$

nach Voraussetzung der hinreichenden Bedingung, denn wir haben $d \in clK(\mathcal{F}, \tilde{x})$ und $\nabla f(\tilde{x})^T d = 0$ wegen 5.41. Außerdem gilt $\|d\| = 1$

$$\Rightarrow \beta \geq \alpha$$

ein Widerspruch! $\square$



Beispiel 5.3.2 *(Projektion auf eine konvexe Menge)*
Sei $C \subset \mathbb{R}^n$ nicht leer, konvex, und $z \notin C$ fest gewählt. Wir suchen das Element $\tilde{x} \in C$ kleinsten Abstands zu z. Wir betrachten also die Aufgabe

$$\boxed{\min_{x \in C} f(x) = \|x - y\|^2 .}$$

Es gilt

$$f(x) = \langle x - y, x - y \rangle , \ \nabla f(x)^T d = 2\langle x - y, d \rangle$$
$$(d^1)^T f''(x)(d^2) = 2\langle d^1, d^2 \rangle$$

- *Notwendige Bedingung 1. Ordnung für Lösung $\tilde{x}$:*

$$\boxed{\langle \tilde{x} - y, x - \tilde{x} \geq 0 \quad \forall x \in C \rangle} \qquad (*)$$

Charakterisierung der Projektion eines Punktes auf eine konvexe Menge

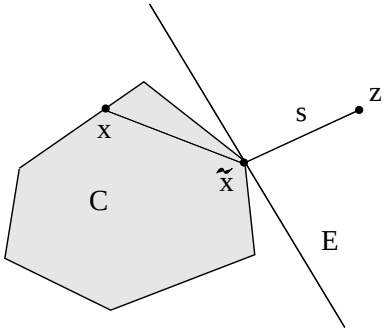
- *Hinreichende Bedingung 2. Ordnung: Wir erhalten ganz einfach*

$$d^T f''(\tilde{x}) d = 2\langle d, d \rangle = 2\|d\|^2 .$$

Positive Definitheit auf dem ganzen Raum!

Nach unserem letzten Satz ist damit jedes $\tilde{x}$, welches $(*)$ erfüllt, lokales striktes Minimum. Wegen Konvexität sogar das globale.

Als wichtige Grundlage für die Optimierung beweisen wir damit einen Trennungssatz.



Satz 5.3.3 *(Trennungssatz). Es sei $C \subset \mathbb{R}^n$ nichtleer, abgeschlossen und konvex, $z \notin C$. Dann gibt es eine Hyperebene, welche C und $\{z\}$ strikt trennt, d. h., es existiert ein $s \neq 0 \in R^n$ mit*

$$\langle s, z \rangle > \sup_{x \in C} \langle s, x \rangle . \tag{5.42}$$

Bevor wir (5.42) beweisen, schreiben wir das Ganze etwas einleuchtender auf: $\langle s, z\rangle \geq c + \langle s, x\rangle \quad \forall x \in C$ mit einem gewissen $c > 0$,

$$\boxed{\langle s, z - x\rangle \geq c \quad \forall x \in C\,.}$$

Damit trennt die Hyperebene $E = \{x | \langle s, z - x\rangle = c\}$ die Menge C und $\{z\}$.

Beweis: Geometrisch motivierte Annahme: $s = z - \tilde{x}$. Wir setzen also $s = z - \tilde{x}$. Dann

$$\begin{aligned} \langle z - \tilde{x},\, x - \tilde{x}\rangle &\leq 0 \quad \forall x \in C \\ \langle \underbrace{z - \tilde{x}}_{s},\, x - z + \underbrace{z - \tilde{x}}_{s}\rangle &\leq 0 \quad \forall x \in C \\ \langle s, x - z\rangle + \underbrace{\|s\|^2}_{C} &\leq 0 \quad \text{das ist dem Obigen äquivalent.} \end{aligned}$$

□

Beispiel 5.3.3 *(Gleichungsnebenbedingung). Betrachten wir die Aufgabe*

$$\min f(x)\,,\ Ax = b\,.$$

Hier muss positive Definitheit von $f''(\tilde{x})$ gefordert werden auf

$$\underbrace{K(\mathcal{F}, \tilde{x})}_{U} \cap \underbrace{\{d | \nabla f(\tilde{x})^T d = 0\}}_{\text{erfüllt wegen } \nabla f \in U^\perp} = U$$

also erhalten wir folgende hinreichende Bedingung 2. Ordnung:

$$\begin{aligned} & d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \\ \text{für alle } d \text{ mit} \quad & Ad = 0\,. \end{aligned}$$

„Positive Definitheit von f auf dem von den linearisierten Nebenbedingungen erzeugten Unterraum“

Beispiel 5.3.4

$$\min f(x) = -(x_1x_2 + x_2x_3 + x_1x_3)$$

$$\text{bei } x_1 + x_2 + x_3 - 3 = 0$$

$$\nabla f = -\begin{pmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_2 + x_1 \end{pmatrix} \quad \begin{array}{l} \textit{Notwendige Bedingung (Variations(un)gleichung)} \\ \nabla f^T d = 0 \\ \textit{für alle } d \textit{ mit } d_1 + d_2 + d_3 = 0\,. \end{array}$$

Wir zeigen: $\tilde{x} = (1, 1, 1)^T$ ist lokales Minimum:

- $\nabla f(1,1,1)^T d = -2(1,1,1) \cdot \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = -2(d_1 + d_2 + d_3) = 0 \quad \textit{falls} \quad d_1 + d_2 + d_3 = 0$

$\Rightarrow$ *notwendige Bedingung erfüllt*

$$f''(1,1,1) = -\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

- *f'' ist offenbar nicht positiv definit (*$\det f'' = -2$*). Dennoch ist die hinreichende Bedingung 2. Ordnung erfüllt, denn:*

$$f''(\tilde{x}) \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} - \begin{pmatrix} d_2+d_3 \\ d_1+d_3 \\ d_1+d_2 \end{pmatrix} - \underbrace{\begin{pmatrix} d_1+d_2+d_3 \\ d_1+d_2+d_3 \\ d_1+d_2+d_3 \end{pmatrix}}_{=0 \quad \text{falls } d_1+d_2+d_3=0} + \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = +d$$

$\Rightarrow d^T f''(\tilde{x}) d = d^T d = \|d\|^2 \quad$ *falls* $\; d_1+d_2+d_3 = 0.$

Positive Definitheit auf Unterraum $\Rightarrow \tilde{x}$ ist lokales Minimum.

5.4 Gleichungsnebenbedingungen

Wir untersuchen nun detaillierter

$$\boxed{\min f(x)\,,\ Ax = b} \qquad \begin{matrix} A : (m,n) \text{ Matrix} \\ \text{mit } m \le n \end{matrix} \tag{PLG}$$

Hier ist $\mathcal{F} = \{x | Ax = b\}$ konvex und abgeschlossen.

5.4.1 Optimalitätsbedingungen erster Ordnung

Satz 5.4.1 *(Multiplikatorenregel) Ist $\tilde{x}$ lokales Minimum von (PLG) und f in $\tilde{x}$ differenzierbar, dann existiert ein $\lambda \in \mathbb{R}^m$, so dass*

$$\nabla f(\tilde{x}) + A^T \lambda = 0\,. \tag{5.43}$$

Hat A vollen Rang, dann ist λ eindeutig bestimmt.

Beweis: (5.43) haben wir schon bewiesen (Beispiel 5.2.5).
Eindeutigkeit von λ: (5.43) heißt

$$\lambda_1 a^1 + \ldots + \lambda_m a^m = -\nabla f\,,$$

wobei $(a^i)^T$ die Zeilenvektoren von A sind, also die Spalten von A^T. Hat A vollen Rang, dann sind $a^1, ..., a^m$ wegen $m \le$n linear unabhängig $\Rightarrow \lambda$ eindeutig bestimmt. $\square$

So sollte man aber nicht versuchen, sich den Satz einzuprägen. Sehr prägnant werden alle unsere weiteren Optimalitätsbedingungen durch Einführung einer *Lagrange-Funktion.* Generell:

$$\text{Lagrange-Funktion} = f + \text{ angehangene Nebenbedingungen.}$$

Hier:

Definition 5.4.1 *(*Lagrange-Funktion*)*

$$\boxed{L(x,\lambda) = f(x) + \lambda^T(Ax - b)}$$

Der Vektor λ heißt *Lagrangescher Multiplikator* oder (wenn man etwas präziser sein will) Vektor der Lagrangeschen Multiplikatoren.

Dann ergibt sich die Bedingung des letzten Satzes ganz einfach als

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

denn

$$\nabla_x L(\tilde{x},\lambda) = \nabla f(\tilde{x}) + A^T\lambda\,.$$

Aber zu den notwendigen Bedingungen gehört natürlich auch die Nebenbedingung $Ax = b$ selbst. Diese bekommt man, wie man sofort sieht, aus

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

Insgesamt erhalten wir folgendes *Optimalitätssystem*:

$$\begin{array}{ll} \nabla f(x) + A^T\lambda = 0 \\ Ax - b \qquad\quad = 0 \end{array} \quad \text{oder} \quad \begin{array}{l} \nabla_x L(x,\lambda) = 0 \\ \nabla_\lambda L(x,\lambda) = 0 \end{array} \tag{5.44}$$

Definition 5.4.2 *Jedes $x \in \mathbb{R}^n$, welches mit einem $\lambda \in \mathbb{R}^m$ das Optimalitätssystem (5.44) erfüllt, heißt* stationärer Punkt *unserer Optimierungsaufgabe.*

Nicht jeder stationäre Punkt ist eine (lokale) Lösung, aber es gilt:

Satz 5.4.2 *Ist f konvex, dann ist jeder stationäre Punkt $\tilde{x}$ globale Lösung von* (PLG)*, d. h., hier ist (5.44) nicht nur notwendige sondern auch hinreichende Optimalitätsbedingung.*

5.4.2 Bedingungen zweiter Ordnung bei Gleichungsrestriktionen

Aus den allgemeinen Betrachtungen vorher erhalten wir als Spezialfall

Satz 5.4.3 *Sei $f \in C^2$ in $\tilde{x}$, die notwendige Bedingung (5.44) erfüllt und es gelte*

$$d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \tag{5.45}$$

für alle $d \in R^n$ mit $Ad = 0$. Dann ist $\tilde{x}$ striktes lokales Minimum für (PLG)*.*

Beispiel 5.4.1 *Wir schauen uns nochmals die Aufgabe*

$$\begin{aligned} &\min f(x) = -(x_1x_2 + x_2x_3 + x_1x_3) \\ \textit{bei} \quad &x_1 + x_2 + x_3 = 3 \end{aligned}$$

an.

- Lagrange-Funktion:

$$L(x, \lambda) = -(x_1x_2 + x_2x_3 + x_1x_3) + \lambda(x_1 + x_2 + x_3 - 3)$$

- *Notwendige Bedingung 1. Ordnung*

$$\nabla_x L = -\begin{pmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0$$

3 Gleichungen für 4 Unbekannte. Die vierte ist die Nebenbedingung.

- *Hinreichende Bedingung zweiter Ordnung:*

$$d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \quad \forall d : d_1 + d_2 + d_3 = 0$$

- *Das hatten wir bereits alles für* $\tilde{x} = (1, 1, 1)^T$ *nachgewiesen.*

Bemerkungen:

(i) Die Bedingung (5.45) is äquivalent zu

$$\boxed{d^T L''_x(\tilde{x}, \lambda) d = 0 \quad \forall d : Ad = 0\,,}$$

denn hier gilt wegen Linearität der Nebenbedingungen

$$L''_x = f''\,.$$

So sollte man sich (5.45) merken

(ii) Anstelle von $Ad = 0$ können wir natürlich eleganter schreiben

$$d \in \ker A\,.$$

5.4.3 Nullraum-Matrizen

Die folgenden Beobachtungen werden später für numerische Verfahren gebraucht. Es gilt

$$\mathbb{R}^n = \ker A \oplus (\ker A)^T\,,$$

d. h. es gibt eine eindeutige Zerlegung

$$x = u + v,$$

wobei u die Projektion von x auf $\ker A$ und v diejenige auf $(\ker A)^T$ ist. Es gilt

$$u = Px = \underbrace{(I - A^T(AA^T)^{-1}A)x}_{} \tag{5.46}$$

falls A vollen Rang hat (Übungsaufgabe).

Wir fassen nun ein System von l Vektoren, die $\ker A$ aufspannen, zu einer (l,n)-Matrix Z zusammen. Dann gilt $im\, Z = \ker A$, d. h.

$$\boxed{d \in \ker A \Leftrightarrow d = Zz \quad \text{mit } z \in \mathbb{R}^l\,.}$$

Durchläuft z als Parameter ganz $\mathbb{R}^l$, so durchläuft d ganz $\ker A$. Die Matrix Z heißt *Nullraum-Matrix*.

Aus der linearen Algebra ist bekannt: Die allgemeine Lösung von $Ax = b$ setzt sich zusammen aus einer speziellen Lösung w und der allgemeinen Lösung des homogenen Systems. Damit

$$\begin{aligned}\mathcal{F} &= \{x | Ax = b\} = w + \ker A = w + \mathrm{imZ}\,,\\ \mathcal{F} &= w + \{Zz | z \in \mathbb{R}^l\}\end{aligned}$$

Durchläuft z ganz $\mathbb{R}^l$, so durchläuft $x = w + Zz$ ganz $\mathcal{F}$. Damit ist Problem (PLG) äquivalent zu

$$\boxed{\min_{z\in\mathbb{R}^l} F(z) = f(w + Zz).} \tag{5.47}$$

Das ist ein unrestringiertes Problem mit $l \leq n$ Variablen.

Damit verbundene Verfahren heißen *Reduktionsverfahren (variable-reduction methods)*.

Beispiel 5.4.2 *Die Projektionsmatrix P von oben ist eine Nullraum-Matrix. Allerdings ist dort $l = n$ (P enthält z.B. I_n als Anteil).*

$$(PLG) \Leftrightarrow \min_{z\in\mathbb{R}^n} F(z) = f(w + Pz)$$

Beispiel 5.4.3 *A habe vollen Rang. Dann o.B.d.A. die ersten m Spalten linear unabhängig und*

$$\begin{aligned} A = (A_1, A_2) \qquad & A_1 : m \times m \quad \textit{invertierbar}\\ & A_2 : \ \textit{“Rest”}\\ Ad = 0 \Leftrightarrow A_1 d^1 + A_2 d^2 = 0 \quad \textit{mit} \quad & d = \begin{pmatrix} d^1 \\ d^2 \end{pmatrix} \begin{matrix} \to \in \mathbb{R}^m \\ \to \in \mathbb{R}^{n-m} \end{matrix}\,. \end{aligned}$$

Nun kann man nach d^1 auflösen, und nur noch d^2 spielt als jetzt freie Variable eine Rolle:

$$\begin{aligned} d^1 &= -A_1^{-1} A_2 d^2 \qquad \textit{Eliminationsmethode}\\ \Rightarrow d &= \begin{pmatrix} -A_1^{-1} A_2 d^2 \\ d_2 \end{pmatrix} \quad \textit{liegt in } \ker A\,. \end{aligned}$$

Insgesamt ist

$$Z = \begin{pmatrix} -A_1^{-1} A_2 \\ I_{n-m} \end{pmatrix}$$

damit Nullraummatrix. (PLG) *ist äquivalent mit*

$$\min_{z\in\mathbb{R}^{n-m}} F(z) = \min f\left(w + \begin{pmatrix} -A_1^{-1}A_2 \\ I_{n-m} \end{pmatrix} z\right)$$
$$\text{nur} \quad l = n-m \quad \text{Variablen}\,.$$

Berechnung von $\nabla F(z)$:

$$F'(z) = f(w+Zz)' = f'(w+Zz)\circ Z \quad \text{Kettenregel}$$
$$\Rightarrow \nabla F(z) = Z^T\cdot\nabla f(w+Zx) = \boxed{Z^T\nabla f(x)} \quad \textit{reduzierter Gradient}$$

Berechnung von $F''(z)$:

$$\begin{aligned} F'(z)h_1 =: \varphi(z) &= f'(w+Zz)\circ Zh_1 = (f'(w+Zz)^T, Zh_1) \\ &= (Z^T f'(w+Zz)^T, h_1) \\ \Rightarrow (F''(z)h_1)h_2 = \varphi'(z)h_2 \quad &= \quad (Z^T \underbrace{f''(w+Zz)^T}\circ Zh_2, h_1) \\ &= f''(...), \text{ wegen Symmetrie} \end{aligned}$$

$$\text{Somit} \quad (F''(z)h_1)h_2 = (Z^T f''(w+Zz)Zh_2, h_1)$$
$$F''(z) = Z^T f''(x)Z \quad \textit{reduzierte Hesse-Matrix}$$

Man kann nun die Optimalitätsbedingungen auch in reduzierter Form aufschreiben, nämlich:

Satz 5.4.4

$$\boxed{\nabla f(\tilde{x}) + A^T\lambda = 0 \quad \Leftrightarrow \quad \nabla F(\tilde{z}) = 0} \quad \textit{mit} \quad \tilde{x} = w + Z\tilde{z}$$

Beweis: (i) $\Rightarrow$: Sei mit $\lambda\in\mathbb{R}^m$ erfüllt

$$\nabla f(\tilde{x}) = -A^T\lambda\,.$$

Dann folgt wegen obiger Darstellung

$$\nabla F(\tilde{z}) = Z^T\nabla f(\tilde{x}) = -Z^TA^T\lambda$$
$$\Rightarrow \nabla F(\tilde{z})^T h = (-Z^TA^T\lambda, h) = -(A^T, \underbrace{Zh}_{\in\ker A}) = 0 \quad \forall h$$
$$\text{d. h.} \quad \nabla F = 0\,.$$

(ii) Sei $\nabla F(\tilde{z}) = Z^T\nabla f(\tilde{x}) = 0$. Wir wählen $d\in\ker A$ beliebig. Dann gilt $d = Zz$ und

$$\nabla f(\tilde{x})^T d = \nabla f(\tilde{x})^T Zz = 0$$

$\Rightarrow \nabla f(\tilde{x}) \perp \ker A$. Lineare Algebra: $\nabla f \in \operatorname{Im} A^T$, also

$$\nabla f = A^T\mu\,. \quad \text{Setzen} \quad \lambda := -\mu\,.$$

□

Analog vereinfachen sich die hinreichenden Bedingungen zweiter Ordnung:

Satz 5.4.5 *Es sei f zweimal stetig in $\tilde{x}$ differenzierbar, Z eine Nullraum-Matrix von A und $\tilde{x} = w + Z\tilde{z}$. Dann ist die positive Definitheit der reduzierten Hesse-Matrix $F''(z) = Z^T f''(\tilde{x}) Z$ äquivalent zur Existenz von $\alpha > 0$ mit*

$$d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \quad \forall d \in \ker A\,. \tag{*}$$

Beweis: (i) $\Rightarrow$: F'' sei positiv definit und $d \in \ker A$, also $d = Zz$. Dann

$$\begin{aligned} d^T f''(\tilde{x}) d = (Zz)^T f''(\tilde{x}) Zz = z^T \underbrace{Z^T f''(\tilde{x}) Z}_{F''(z)} z \\ \geq \beta \|z\|^2 \geq \underbrace{\beta \|Z\|}_{=\alpha}{}^{-2} \|d\|^2 \end{aligned}$$

(wegen positiver Definitheit von F'' und $\quad \|d\| \leq \|Z\|\,\|z\| \Rightarrow \|z\| \geq \|Z\|^{-1}\|d\|$

(ii) $\Leftarrow$: (*) sei erfüllt, d.h. mit $d = Zz$

$$(Zz)^T f''(\tilde{x}) Zz = z^T \underbrace{Z^T f''(\tilde{x}) Z}_{=F''(\tilde{z})} z \geq \alpha \|d\|.$$

Daraus erzielt man leicht die positive Definitheit von F''. □

Beispiel 5.4.4

$$\begin{aligned} \min f(x) &= x_1^2 + 2x_2^2 + 3x_3 \\ \textit{bei} \quad & x_1 + 2x_2 + 3x_3 = 6\,. \end{aligned}$$

Durch eine Nebenbedingung können wir offenbar die Dimension um 1 reduzieren.

$$\begin{array}{lll} A = (\underset{\uparrow}{1}, \underbrace{2, 3}) & & d_1 + 2d_2 + 3d_3 = 0 \\ \quad A_1 \; A_2 & \Updownarrow & d_1 = -2d_2 - 3d_3 \end{array}$$

$$d \in \ker A \Leftrightarrow d = \begin{pmatrix} -2d_2 - 3d_3 \\ d_2 \\ d_3 \end{pmatrix} = \underbrace{\begin{pmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}}_{Z} \begin{pmatrix} d_2 \\ d_3 \end{pmatrix}$$

Spezielle Lösung der inhomogenen Gleichung:

$$w = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}\,.$$

Reduziertes freies Problem:

$$\begin{aligned} \min F(z_1, z_2) &= f(w + Zz) = f(6 - 2z_1 - 3z_2, z_1, z_2) \\ &= (6 - 2z_1 - 3z_2)^2 + 2z_1^2 + z_2 \end{aligned}$$

(entspricht ganz einfach der Elimination von x_1).
Stationärer Punkt, d. h. $\nabla F(\tilde{z}) = 0$,

$$\tilde{z} = \begin{pmatrix} 1/2 \\ 3/2 \end{pmatrix} .$$

Wegen Satz 5.4.4 $\Rightarrow \tilde{x} = w + Zz = \begin{pmatrix} 1/2 \\ 1/2 \\ 3/2 \end{pmatrix}$ *löst die notwendigen Bedingungen der Lagrange Multiplikatorenregel. Konvexität von* $f \Rightarrow \tilde{x}$ *ist Lösung.*
Wir hätten aber auch einfach die hinreichende Bedingung nachprüfen können:

$$f''(x) = \begin{pmatrix} 2 & & 0 \\ & 4 & \\ 0 & & 0 \end{pmatrix}$$

$$F''(\hat{z}) = \begin{pmatrix} -2 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} -4 & 4 & 0 \\ -6 & 0 & 0 \end{pmatrix} \begin{pmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 12 & 12 \\ 12 & 18 \end{pmatrix} \quad \textit{positiv definit!}$$

5.4.4 Quadratische Optimierungsprobleme

Als Spezialfall von (PLG) betrachten wir

$$\begin{aligned} &\min f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle q, x\rangle \\ &\text{bei} \quad Ax = b \end{aligned} \tag{QG}$$

mit (n,n)-Matrix Q, $q \in \mathbb{R}^n$.
Klar ist:

Satz 5.4.6 *Ist Q positiv definit auf* $\ker A$*, d. h.*

$$d^T Q d \geq \alpha \|d\|^2 \quad \forall d \in \ker A\,,$$

dann hat (QG) *genau eine Lösung.*

Denn: f ist strikt konvex auf $\mathcal{F}$!
Besonders schön ist hier das **Optimalitätssystem**

- $\nabla f = Qx + q$ falls Q symmetrisch!
  $\Rightarrow Qx + q + A^T\lambda = 0\,,\ Ax = b.$

Anders aufgeschrieben

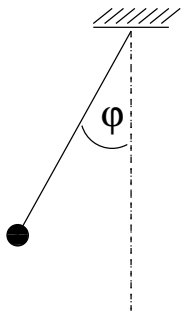
$$\boxed{\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}} \quad \text{Optimalitätssystem}. \tag{5.48}$$

Jede Lösung dieses Systems ist bei positiv definierter Matrix Q auf $\ker A$ und Matrix A mit vollem Rang eine Lösung von (QG) (Konvexität!). Wegen Satz 5.4.6 gibt es höchstens eine. Umgekehrt existiert dann genau eine Lösung des Optimalitätssystems, d.h. es gilt

Lemma 5.4.1 *Wenn $m \leq n$, rang $A = m$, Q positiv definit auf $\ker A$, dann ist die Matrix $\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix}$ invertierbar.*

5.4.5 Dynamische Optimierungsprobleme

Probleme dieser Art sind sehr wichtig für viele Probleme in Wissenschaft und Technik. Eigentlich kommen sie her von Optimalsteuerungsproblemen bei gewöhnlichen Differentialgleichungen.



Beispiel 5.4.5 *Optimale Steuerung eines schwingenden Pendels in die Ruhelage:*

Zeitraum:	$t \in [0,T]$		
Auslenkung:	$\varphi = \varphi(t)$	$\varphi(0) = \varphi_0$	
Winkelgeschwindigkeit:	$w = w(t)$	$w(0) = w_0$	$w = \dot{\varphi}$
Angreifende, steuerbare Kraft:	$u = u(t)$		

Bewegungsgleichung:
$$\begin{aligned} \ddot{\varphi}(t) &= -c\sin(\varphi(t)) + u(t) \\ \varphi(0) &= \varphi_0 \\ \dot{\varphi}(0) &= w_0 \end{aligned}$$

c steht für Verhältnis von Masse und Gravitation. Umformung als System 1. Ordnung

$$\begin{aligned} \dot{w} &= -c\sin(v) + u \\ \dot{\varphi} &= w \qquad\qquad z(t) := \begin{pmatrix} \varphi(t) \\ w(t) \end{pmatrix} \end{aligned}$$

Optimierungsziel: Ruhelage bei $t = T$, d.h. wir erhalten das Optimierungsproblem

$$\min \frac{1}{2}\|z(T)\|^2 + \nu \int_0^T u^2(t)dt$$

bei

$$\dot{z}(t) = \begin{pmatrix} w(t) \\ -c\sin(v(t)) \end{pmatrix} + \begin{pmatrix} 0 \\ u(t) \end{pmatrix}$$
$$z(0) = \begin{pmatrix} \varphi_0 \\ w_0 \end{pmatrix} = z_0\,.$$

Der zweite Term in der Zielfunktion misst die Energiekosten. Bei kleinen Schwingungen: $\sin\varphi \approx \varphi$ *Linearisierung, dann ergibt sich folgendes Optimalsteuerungsproblem*

$$\min \frac{1}{2}\|z(T)\|^2 + \nu \int_0^T u(t)^2 dt$$
$$\dot{z}(t) = \underbrace{\begin{pmatrix} 0 & 1 \\ -c & 0 \end{pmatrix}}_{A} z(t) + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{B} u(t)$$
$$z(0) = z_0$$

Dieses Optimalsteuerungsproblem kann theoretisch im Funktionenraum gut behandelt werden. Dies wollen wir hier nicht tun, sondern stattdessen eine diskretisierte Variante behandeln!

Sei dazu $0 = t_0 < t_1 < \ldots < t_N = T$ eine äquidistante Zerlegung von $[0, T]$ mit $t_i = \tau \cdot i$, $0 \le i \le N$ und $\tau = T/N$. Wir setzen $z(t)$ als vektorwertige stückweise lineare Funktion an, $u(t)$ als Treppenfunktion

$$u(t) \equiv u_i \quad \text{auf } [t_i, t_{i+1}]$$
$$z(t_i) = z_i \quad \text{in } t_i,\ i = 0, \ldots, N$$

Approximation der Ableitung

$$\dot{z}(t_i) \approx \frac{z(t_{i+1}) - z(t_i)}{\tau} \qquad z \in \mathbb{R}^2$$
$$\Rightarrow z(t_{i+1}) = z(t_i) + \tau A z(t_i) + \tau B u_i \qquad i = 0, \ldots, N-1$$
$$= \underbrace{(I + \tau A)}_{\tilde{A}} z(t_i) + \underbrace{\tau B}_{\tilde{B}} u_i$$

Insgesamt ergibt sich die diskretisierte Schwingungsgleichung

$$\boxed{z_{i+1} = \tilde{A} z_i + \tilde{B} u_i \quad i = 0, \ldots, N-1}\,.$$

Wegen

$$\int_0^T u(t)^2 dt = \sum_0^{N-1} \int_{ti}^{t_{i+1}} u_i^2 dt = \tau \sum_0^{n-1} u_i^2$$

erhalten wir als Zielfunktional, welches zu minimieren ist,

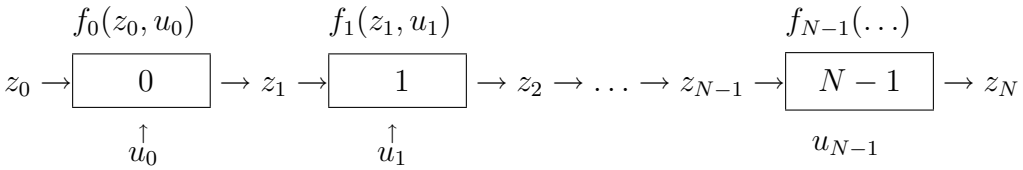
$$\boxed{f(z, u) = \frac{1}{2}\|z_N\|^2 + \sum_{i=0}^{N-1} \tau\nu u_i^2\,.}$$

Diese Aufgabe gehört zum Typ der dynamischen Optimierungsprobleme, deren allgemeine Version folgende Form hat (vgl. [1,S. 202ff]):

$$\boxed{\begin{aligned} &\min f(z,u) = \frac{1}{2} z_N^T Q z_N + \sum_{i=0}^{N-1} f_i(z_i, u_i) \\ &\text{bei} \quad z_{i+1} \quad = A_i z_i + B_i u_i + c_i \quad i = 0, \ldots, N-1\,. \end{aligned}} \tag{DLG}$$

Hier ist $z_0 \in \mathbb{R}^n$ fest vorgegeben, Q eine symmetrische $(n \times n)$-Matrix, A_i sind $(n \times n)$-Matrizen, $B_i (n \times m)$-Matrizen, $c_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ und $x^T = (z^T, u^T)$, $x \in \mathbb{R}^{N(m+n)}$.

Wir können uns das als *N-stufigen Entscheidungsprozess* vorstellen:



Entscheidungen: u_i.

Herleitung der notwendigen Optimalitätsbedingungen: Minimumprinzip

Wir schreiben die Nebenbedingungen von (DLG) etwas anders auf:

$$-A_i z_i + z_{i+1} - B_i u_i = c_i \quad i = 0, \ldots, N-1$$

und beachten, dass für $i = 0$ die Variable z_0 fest vorgegeben ist, also spielt diese Gleichung eine Sonderrolle:

$$z_1 - B_0 u_0 = c_0 + A_0 z$$

$\Rightarrow$ Nebenbedingungen

$$\boxed{\mathcal{A}x = b} \quad \text{mit} \quad b = \begin{pmatrix} A_0 z_0 + c_0 \\ c_1 \\ \vdots \\ c_{N-1} \end{pmatrix} \in \mathbb{R}^{nN} \qquad \begin{aligned} &x \in \mathbb{R}^{N(n+m)} \\ &x = \begin{pmatrix} z \\ u \end{pmatrix} \end{aligned}$$

$$\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$$

$$\mathcal{A}_1 = \begin{pmatrix} I & & & & \\ -A_1 & I & & & \\ & \ddots & & & \\ & & -A_2 & I & \\ & & & \ddots & \ddots \\ & & & & -A_{N-1} \; I \end{pmatrix}, \quad \mathcal{A}_2 = \begin{pmatrix} -B_0 & & \\ & \ddots & \\ & & -B_{N-1} \end{pmatrix}$$

hier ist $I = I_n$. $\mathcal{A}_1$ ist auf alle Fälle invertierbar, weil untere Dreiecksmatrix mit I in Hauptdiagonale (lässt sich von Anfang zum Ende durchlösen). Damit hat $\mathcal{A}$ vollen Rang. Offenbar gilt daher $\mathcal{F} = \{x | \mathcal{A}x = b\} \neq \phi$.

Nun sei $\tilde{x} = \binom{\tilde{z}}{\tilde{u}}$ eine (lokale) Lösung von (DLG). Weil $\mathcal{A}$ vollen Rang hat, existiert genau ein Lagrangescher Multiplikator $\tilde{\lambda} \in \mathbb{R}^{Nn}$, $\tilde{\lambda} = (\tilde{\lambda}_1^T, \ldots, \tilde{\lambda}_N^T)^T$ mit $\nabla f(\tilde{x}) + \mathcal{A}^T\tilde{\lambda} = 0$, d. h.

$$\boxed{\nabla f(\tilde{z}, \tilde{u}) + \mathcal{A}^T\tilde{\lambda} = 0\,.}$$

Leider passt dieses Pluszeichen oben nicht in die allgemeine Theorie der Optimalsteuerung. Deshalb setzen wir

$$\lambda := -\tilde{\lambda}\,,$$

um die Gleichung

$$\nabla f(\tilde{z}, \tilde{u}) - \mathcal{A}^T\lambda = 0$$

zu erhalten. Wir transponieren die Gleichung und betrachten

$$f_z(\tilde{z}, \tilde{u})z + f_u(\tilde{z}, \tilde{u})u - \lambda^T\mathcal{A}\begin{pmatrix} z \\ u \end{pmatrix} = 0 \quad \forall\, \begin{pmatrix} z \\ u \end{pmatrix} \in R^{N(n+m)}$$

(f_z bezeichnet die partielle Ableitung von f nach z, nicht den partiellen Gradienten!)

Jetzt heißt es, kräftig zu rechnen, um alles in eine anwendbare Form zu bringen.

$$f_z z = \sum_{i=1}^{N-1} f_{i,z} \cdot z_i + \tilde{z}_N^T Q z_N \quad \text{(keine Abhängigkeit von } z_0\text{!)}$$

$$f_u u = \sum_{0}^{N-1} f_{i,u} \cdot u_i$$

$$\lambda^T\mathcal{A}\begin{pmatrix} z \\ u \end{pmatrix} = \lambda^T(\mathcal{A}_1 z + \mathcal{A}_2 u) = (\lambda_1^T, \ldots, \lambda_N^T)\begin{pmatrix} z_1 & -B_0u_0 \\ z_2 - A_1z_1 & -B_1u_1 \\ \vdots & \\ z_N - A_{N-1}z_{N-1} & -B_{N-1}u_{N-1} \end{pmatrix} \begin{matrix} \\ \leftarrow \text{ Sonderrolle} \\ \leftarrow \text{ ab 2. Komp.} \\ \quad \text{normal} \end{matrix}$$

Einsetzen in die notwendige Bedingung

$$\begin{aligned} &\tilde{z}_N^T Q z_N + \sum_{i=1}^{N-1} f_{i,z} z_i + \sum_{i=0}^{N-1} f_{i,u} u_i \\ &- \sum_{i=1}^{N-1} \lambda_{i+1}^T[z_{i+1} - A_i z_i - B_i u_i] - \lambda_1^T[z_1 - B_0 u_0] = 0 \quad \text{für alle } \begin{pmatrix} z \\ u \end{pmatrix}. \end{aligned} \tag{5.49}$$

Nun kann man z, u beliebig passend einsetzen, um Gleichungen für λ zu erhalten.

• $u = 0$, $z_1, \ldots, z_{N-1} = 0$, nur z_N frei $\Rightarrow$

$$(\tilde{z}_N^T Q - \lambda_N^T) z_N = 0 \quad \forall z_N$$

$\Rightarrow$ $\boxed{\lambda_N = Q\tilde{z}_N}$ (Q ist symmetrisch!)

D.h. wir bekommen eine *Endbedingung* für λ.

Damit sind jetzt alle mit z_N verbundenen Terme Null.

• Wählen jetzt z_i beliebig, die restlichen Null, auch die u's

$\Rightarrow f_{i,z} - \lambda_i^T z_i + \lambda_{i+1}^T A_i z_i = 0 \quad \forall z_i \quad i \in \{1, \ldots, N-1\}$

also $f_{i,z} - \lambda_i^T + \lambda_{i+1}^T A_i = 0 \quad$ oder $\quad \lambda_i^T = \lambda_{i+1}^T A - A_i + f_{i,z}$

oder, noch schöner,

$$\boxed{\begin{array}{l} \lambda_i = A_i^T \lambda_{i+1} + \nabla_z f_i(\tilde{z}_i, \tilde{u}_i) \quad i = N-1, \ldots, 1 \\ \lambda_N = Q\tilde{z}_N \end{array}} \tag{5.50}$$

Definition 5.4.3 *(5.50) heißt* adjungierte Gleichung *mit Endbedingung.*

Nun werten wir noch die u_i's aus: Setzen alle z_j Null und alle u_j außer u_i. Dann

$$[f_{i,u} + \lambda_{i+1}^T B_i] u_i = 0 \quad \forall u_i \,,$$

also

$$\boxed{B_i^T \lambda_{i+1} + \nabla_u f_i(\tilde{z}, \tilde{u}) = 0\,.} \quad \text{Minimumprinzip}$$

Interpretation: Wir betrachten $g_i(u) = f_i(\tilde{z}, u) + B_i^T \lambda \cdot u$. Dann steht oben, dass $\tilde{u}_i$ die *notwendige Bedingung* für die Aufgabe $\min\limits_u g_i(u)$ erfüllt (was nicht heißt, dass $\tilde{u}_i$ diese Minimumaufgabe auch wirklich löst. Die muss nicht einmal lösbar sein!).

Bemerkung. Wir haben *adjungierte Gleichung* und *Minimumprinzip* aus der Lagrangeschen Multiplikatorenregel $\nabla f(\tilde{x}) + A^T \lambda = 0$ hergeleitet. Diese Herangehensweise ist zu wenig praktikabel. Äquivalent ist das Arbeiten mit der Lagrange-Funktion:

$$\mathcal{L}(z, u, \lambda) = f(z, u) \quad \text{– herangehangene Gleichungen}$$
$$\text{(wegen} \quad \lambda := -\tilde{\lambda}),$$

also

$$\begin{aligned} \mathcal{L}(z, u, \lambda) = & \frac{1}{2} z_N^T Q z_N + \sum_{i=0}^{N-1} f_i(z_i, u_i) \\ & - \sum_{i=1}^{N-1} \lambda_{i+1}^T [z_{i+1} - A_i z_i - B_i u_i - c_i] - \lambda_1^T [z_1 - B u_0 - z_0]\,. \end{aligned}$$

Dann gilt: Die notwendigen Bedingungen sind äquivalent mit

$$\boxed{\begin{array}{ll} \nabla_{z_i} \mathcal{L}(\tilde{z}, \tilde{u}, \lambda) = 0\,, & i = 1, \ldots, N \\ \nabla_{u_i} \mathcal{L}(\tilde{z}, \tilde{u}, \lambda) = 0\,, & i = 0, \ldots, N-1 \end{array}} \quad \begin{array}{l} \to \\ \to \end{array} \quad \begin{array}{l} \text{adjungierte Gleichung} \\ \text{Minimumbedingung} \end{array}$$

(Gleich mit $\mathcal{L}$ beginnen, um (5.49) zu bekommen!)

So sollte man sich das merken. Im Übrigen stellt (5.49) genau dieses System dar. D. h., ausgehend von $\mathcal{L}(z, u)$, kommt man direkt zu (5.49).

5.5 Affine Ungleichungsnebenbedingungen

5.5.1 Problemdefinition

Wir betrachten nun die etwas allgemeinere, nämlich durch Ungleichungsrestriktionen erweiterte Aufgabe

$$\min f(x) \tag{PLU}$$

bei

$$\langle a^i, x\rangle = b_i\,,\ i = 1,\dots,m$$
$$\langle g^j, x\rangle \le r_j\,,\ j = 1,\dots,p$$

oder in entsprechend abgekürzter Schreibweise:

$$\mathcal{F} = \{x | Ax = b\,,\ Gx \le r\}\,, \quad A = \begin{pmatrix} a^1 \\ \vdots \\ a^m \end{pmatrix}\,, \quad G = \begin{pmatrix} g^1 \\ \vdots \\ g^m \end{pmatrix}$$

somit

$$\boxed{\begin{array}{l} \min f(x) \\ \text{bei} \quad Ax = b\,,\ Gx \le r \end{array}} \tag{PLU}$$

Beispiel 5.5.1 Konvexe Linearkombination kleinster Norm
Gegeben: Vektoren $s_j \in \mathbb{R}^n$, $j = 1,\dots,p$

$$\min_{\alpha\in\mathbb{R}^p} f(\alpha) = \frac{1}{2}\,\|\sum_{j=1}^{p}\alpha_j s_j\|^2$$

$$\textit{bei} \quad \sum_{j=1}^{p}\alpha_j = 1\,, \underbrace{\alpha_j \ge 0}, \qquad j = 1,\dots,p$$

Nichtnegativitätsforderungen

5.5.2 Notwendige Optimalitätsbedingungen

Bisher waren die notwendigen Optimalitätsbedingungen für eine Lösung $\tilde{x}$ aus der Analysis bekannt. Das ist jetzt anders, jetzt wird Neuland betreten!

Definition 5.5.1 *(aktive Ungleichungen). Es sei* $x \in \mathcal{F}$.

$$J(x) = \{j \in \{1,\dots,p\} | \langle g^j, x\rangle = r_j\}$$

heißt Indexmenge der aktiven Ungleichungs-Restriktionen.

Nun sei $\tilde{x}$ eine Lösung der Aufgabe (PLU).

Definition 5.5.2 *(*Linearisierender Kegel*).*

$$L(\mathcal{F}, \tilde{x}) = \{d \in \mathbb{R}^n | Ad = 0\,,\ \langle g^j, d\rangle \leq 0 \quad \forall j \in J(\tilde{x})\}\,.$$

Lemma 5.5.1 *Für $\tilde{x} \in \mathcal{F}$ gilt*

$$K(\mathcal{F}, \tilde{x}) = L(\mathcal{F}, \tilde{x})\,.$$

Insbesondere ist damit $K(\mathcal{F}, \tilde{x})$ abgeschlossen (weil das für L offensichtlich ist).

Beweis:

(i) $K \subset L$.
Sei $d \in K(\mathcal{F}, \tilde{x})$, d. h. $d = \alpha(x - \tilde{x})$ mit $x \in \mathcal{F}$, $\alpha \geq 0$.
$\Rightarrow Ax = A\tilde{x} = b \Rightarrow Ad = 0$.

Das war einfach. Außerdem

$$\begin{aligned} & \langle g^j, x\rangle \leq r_j = \langle g^j, \tilde{x}\rangle \quad \forall j \in J(\tilde{x}) \\ \Rightarrow \quad & \langle g^j, d\rangle = \alpha\langle g^j, x - \tilde{x}\rangle \leq 0\,. \end{aligned}$$

Insgesamt also $d \in L(\mathcal{F}, \tilde{x})$.

(ii) $L \subset K$:
Sei $d \in L(\mathcal{F}, \tilde{x})$. Für $j \neq J(\tilde{x})$ gilt $\langle g^j, \tilde{x}\rangle < r_j$
$\Rightarrow \langle g^j, \tilde{x} + td\rangle < r_j \quad \forall t \in [0, \bar{t}] \quad , \quad \forall j \notin J(\tilde{x})$ mit einem $\bar{t} > 0$.

Von d wissen wir wegen $d \in L$:

$$\begin{aligned} & Ad = 0 \quad \text{und} \quad \langle g^j, d\rangle \leq 0 \quad j \in J(\tilde{x}) \\ \Rightarrow \quad & A(\tilde{x} + td) = A\tilde{x} + t\underbrace{Ad}_{0} = A\tilde{x} = b \\ & \langle g^j, \tilde{x} + \bar{t}d\rangle = \underbrace{\langle g^j, \tilde{x}\rangle}_{=r_j} + \bar{t}\underbrace{\langle g^j, d\rangle}_{\leq 0} \leq r_j \quad \forall j \in J(\tilde{x})\,. \end{aligned}$$

Deshalb $\tilde{x} + \bar{t}d \in \mathcal{F}$, also $\bar{t}d \in \mathcal{F} - \tilde{x}$, damit $d \in K(\mathcal{F}, \tilde{x})$. □

Nun brauchen wir aber auch noch den Normalenkegel N. Dazu benötigen wir:

Lemma 5.5.2 *Sei $K \subset \mathbb{R}^n$ konvexer abgeschlossener Kegel und $x \notin K$. Dann existiert ein $s \in \mathbb{R}^n$ mit*

$$\langle s, x\rangle > 0 = \max_{y \in K} \langle s, y\rangle\,.$$

Beweis: Trennungssatz 5.3.3 $\Rightarrow \exists s \in \mathbb{R}^n$:

$$\langle s, x\rangle > \sup_{y \in K} \langle s, y\rangle \qquad (*)$$

$\rightarrow$ Fall 1: $\langle s, y\rangle \leq 0 \quad \forall y \in K\,.$
Dann ist das Supremum Null (wegen $0 \in K$).

$\rightarrow$ Fall 2: $\exists z \in K \quad$ mit $\quad \langle s, z\rangle > 0\,.$
Dann ist das Supremum unendlich (man nehme $\alpha \cdot z$, mit $\alpha \to \infty$).

Fall 2 kann wegen $(*)$ nicht eintreten, nur Fall 1, d. h.

$$\langle s, x\rangle > 0 = \max_{y\in K}\langle s, y\rangle\,.$$

□

Außerdem gilt: Ist $K \subset \mathbb{R}^n$ abgeschlossener konvexer Kegel, dann $(K^*)^* = K$ (Übungsaufgabe).

Nun können wir $N(\mathcal{F}, \tilde{x})$ angeben:

Lemma 5.5.3

$$\boxed{N(\mathcal{F}, \tilde{x}) = \left\{ \sum_{i=1}^{m} \lambda_i a^i + \sum_{j\in J(\tilde{x})} \mu_j g^j | \lambda \in \mathbb{R}^m\,,\ \mu_j \geq 0 \right\}}$$

Beweis: Wir wollen die rechte Seite, d. h. $\{...\}$, mit $N(A, G, \tilde{x})$ bezeichnen.

(i) "$\supset$": Sei $s \in N(A, G, \tilde{x})$, d. h.

$$s = \sum \lambda_i a^i + \sum_J \mu_j g^j\,,\ \mu_j \geq 0\,.$$

Multiplizieren skalar mit $d \in K(\mathcal{F}, \tilde{x})$ durch. Dann gilt wegen $K = L$ (Form von K!)

$$\langle s, d\rangle = \sum \lambda_i \underbrace{\langle a^i, d\rangle}_{=0} + \sum_J \mu_j \underbrace{\langle g^j, d\rangle}_{\leq 0} \leq 0 \quad \forall d \in K(\mathcal{F}, \tilde{x})\,.$$

Damit nach Def.

$$s \in K(\mathcal{F}, \tilde{x})^* = N(\mathcal{F}, \tilde{x})$$

(ii) "$\subset$": Aus Lemma 5.5.1 folgt $N(A, G, \tilde{x})^* \subset K(\mathcal{F}, \tilde{x})$.

$$\Rightarrow N(\mathcal{F}, \tilde{x}) = K(\mathcal{F}, \tilde{x})^* \subset (N(A, G, \tilde{x}))^{**} = N(A, G, \tilde{x})\,.$$

□

Nun haben wir alles bereit zum Aufstellen der notwendigen Bedingungen für $\tilde{x}$: Wir wissen wegen

$$-\nabla f(\tilde{x}) \in N(\mathcal{F}, \tilde{x})\,,$$

d. h. wegen dem eben Bewiesenen

$$-\nabla f(\tilde{x}) = \sum_{i=1}^{m} \lambda_i a^i + \sum_{j\in J(\tilde{x})} \mu_j g^j\,,\ \mu_j \geq 0$$

oder

$$\boxed{0 = \nabla f(\tilde{x}) + \sum_{i=1}^{m} \lambda_i a^i + \sum_{j \in J(\tilde{x})} \mu_j g^j \,,\ \mu_j \geq 0\,.} \qquad (*)$$

Das können wir noch verschönern. Wir setzen für $j \notin J(\tilde{x})$ einfach $\mu_j = 0$. Dann gilt

- $\mu_j \geq 0 \quad \forall j \in \{1, ..., p\}$
- $\mu_j \underbrace{(\langle g^j, \tilde{x} \rangle - r_j)}_{\text{denn das ist Null für } j \in J(\tilde{x}).} = 0 \quad \forall j \in \{1, ..., p\}$

Außerdem nimmt $(*)$ die Form

$$\nabla f(\tilde{x}) + A^T \lambda + G^T \mu = 0$$

an.

Definition 5.5.3 *$\lambda \in \mathbb{R}^m$ und $\mu \in \mathbb{R}^p$ heißen Lagrangesche Multiplikatoren zu $\tilde{x} \in \mathcal{F}$, wenn*

$$\nabla f(\tilde{x}) + A^T \lambda + G^T \mu = 0 \qquad (5.51)$$

$$\mu \geq 0 \quad \textit{sowie} \quad \langle \mu, G\tilde{x} - r \rangle = 0 \qquad (5.52)$$

"komplementäre Schlupfbedingung".

Wir haben bewiesen:

Satz 5.5.1 *Ist $\tilde{x}$ lokales Minimum von (PLU) und f in $\tilde{x}$ differenzierbar, dann existieren Lagrangesche Multiplikatoren $\lambda \in \mathbb{R}^m$ und $\mu \in \mathbb{R}^p$ zu $\tilde{x}$. Sind die Vektoren a^i, $i = 1, ..., m$ sowie g^j, $j \in J(\tilde{x})$, linear unabhängig, so sind λ und μ eindeutig bestimmt.*

Bemerkungen zu diesem Satz:

(i) Historie: Die Multiplikatorenregel für Gleichungsrestriktionen stammt von Lagrange. Verallgemeinerung auf Ungleichungsrestriktionen: 1951, Kuhn und Tucker; unabhängig davon auch von Karush. Deshalb auch Bezeichnung Karush-Kuhn-Tucker-System oder KKT-System für das Optimalitätssystem

$$\boxed{\begin{aligned} &Ax = b\,,\ Gx \leq r\,,\ \nabla f + A^T\lambda + G^T\mu = 0\,,\\ &\mu \geq 0\,,\ \langle \mu, G\tilde{x} - r \rangle = 0\,. \end{aligned}}$$

(ii) Verwendung der $\mathcal{L}$-Funktion:

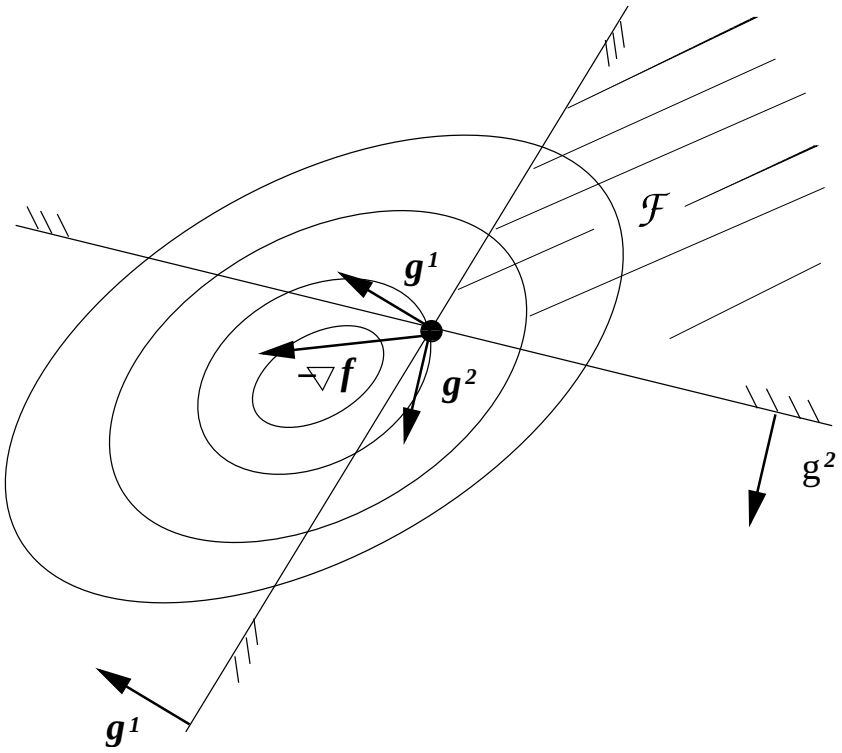
$$\mathcal{L} = f(x) + \langle \lambda, Ax - b \rangle + \langle \mu, Gx - r \rangle\,.$$

Dann bedeutet (5.51) wieder

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

Die Nichtnegativitätsforderungen an μ und die komplementäre Schlupfbedingung muss man sich zusätzlich merken.

(iii) Geometrische Interpretation (reine Ungleichungsrestriktionen). $-\nabla f = \mu_1 g^1 + \mu_2 g^2$, d.h. $-\nabla f$ ist positive Linearkombination von g^1 und g^2, liegt also im von g^1 und g^2 aufgespanten Kegel.



(iv) Bei Konvexität sind die Bedingungen wieder hinreichend für (globale) Optimalität von $\tilde{x}$.

Beispiel 5.5.2

$$\begin{array}{c} \min f(x) = -\frac{1}{2}\sqrt{x_1} - \frac{1}{2}x_2 \\ bei \quad x_i \geq 0\,,\ i = 1,2 \\ x_1 + x_2 \leq 1 \end{array} \qquad \Leftrightarrow \left| \begin{array}{r} -x_1 \leq 0 \\ -x_2 \leq 0 \\ x_1 + x_2 - 1 \leq 0 \end{array} \right.$$

Aufstellen des Optimalitätssystems:

$$\mathcal{L}(x, \lambda, \mu) = \mathcal{L}(x, \mu) = -\frac{1}{2}\sqrt{x_2} - \frac{1}{2}x_2 - \mu_1 x_1 - \mu_2 x_2 + \mu_3(x_1 + x_2 - 1)$$

$$\nabla_x \mathcal{L} = \begin{pmatrix} -1/4\sqrt{\dfrac{1}{x_1}} - \mu_1 + \mu_3 \\ -\dfrac{1}{2} - \mu_2 + \mu_3 \end{pmatrix} \overset{(!)}{=} 0$$

$\Rightarrow$ *Optimalitätssystem*

$$\boxed{\begin{aligned} -\frac{1}{4\sqrt{x_1}} - \mu_1 + \mu_3 &= 0 & x_1 &\geq 0\,, \quad i = 1,2 \\ -\mu_2 + \mu_3 &= \frac{1}{2} & \mu_j &\geq 0\,, \quad j = 1,2,3 \\ & & \mu_1 x_1 &= 0 \\ & & \mu_2 x_2 &= 0 \\ & & \mu_3(x_1 + x_2 - 1) &= 0\,. \end{aligned}}$$

$$\begin{aligned} \textit{Eine Lösung:} \quad \tilde{x}_1 &= 1/4 \quad \tilde{\mu}_1 = \tilde{\mu}_2 = 0 \\ \tilde{x}_2 &= 3/4 \quad \tilde{\mu}_3 = 1/2\,. \end{aligned}$$

Wie kommt man drauf? Geometrisch oder numerisch!

5.5.3 Hinreichende Optimalitätsbedingungen

Allgemein wissen wir, wie diese aussehen (Satz 5.3.2)

$$d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \quad \forall d \in K(\mathcal{F}, \tilde{x}) \quad \text{mit} \quad \nabla f(\tilde{x})^T d = 0\,.$$

Andererseits haben wir für (PLU) den Kegel K bereits berechnet,

$$K = \{d | Ad = 0\,,\ \langle g^j, d\rangle \leq 0 \quad \forall j \in J(\tilde{x})\}\,.$$

Nun erfüllt aber $\tilde{x}$ auch die notwendigen Bedingungen, also

$$\nabla f(\tilde{x}) = -A^T\lambda - \sum_{j \in J(\tilde{x})} \mu_j g^j$$

Multiplikation mit $d \in K(\mathcal{F}, \tilde{x})$ mit $\nabla f(\tilde{x}) d = 0 \Rightarrow$

$$\begin{aligned} 0 = \langle \nabla f(\tilde{x}), d\rangle = \langle \underbrace{-\lambda, Ad}\rangle - \sum \mu_j \quad \langle \underbrace{g^j, d}\rangle\,, \\ = 0 \text{ wenn} \qquad \leq 0 \text{ wenn} \\ d \in K \qquad d \in K \end{aligned}$$

also

$$0 = \sum_{j \in J(\tilde{x})} \mu_j \langle g^j, d\rangle\,.$$

Wenn alle Skalarprodukte $\langle g^j, d\rangle \leq 0$, und alle $\mu_j \geq 0$ sind, kann nur gelten: Wo $\mu_j > 0$, dort muss $\langle , \rangle = 0$ sein, d. h., wir erhalten als zusätzliche Bedingung zu $d \in \mathcal{K}(\mathcal{F}, \hat{x})$:

$$\langle g^j, d\rangle = 0 \quad \text{falls } \mu_j > 0\,.$$

Damit ergibt sich als *hinreichende Optimalitätsbedingung:*

$$\boxed{\begin{aligned} d^T f''(\tilde{x}) d \quad &\geq \alpha \|d\|^2 \\ \text{für alle} \quad d \in \mathbb{R}^n \quad &\text{mit} \\ Ad \quad &= 0 \\ \langle g^j, d\rangle \quad &\leq 0 \quad \text{für } j \in J(\tilde{x}) \quad \text{mit } \mu_j = 0 \\ \langle g^j, d\rangle \quad &= 0 \quad \text{für } j \in J(\tilde{x}) \quad \text{mit } \mu_j > 0 \end{aligned}} \tag{5.53}$$

Definition 5.5.4 *Die für $\tilde{x}$ aktiven Restriktionen mit $\tilde{\mu}_j > 0$ heißen* streng aktive Restriktionen.

Satz 5.5.2 *(Hinreichende Bedingung). Es sei f in $\tilde{x}$ vom Type C^2, und $\tilde{x}$ erfülle die notwendigen Optimalitätsbedingungen (Satz 5.5.1). Weiter seien die hinreichenden Bedingungen (5.53) erfüllt. Dann ist $\tilde{x}$ striktes lokales Minimum von (PLG).*

Beispiel 5.5.3 *(Fortsetzung Beispiel 5.5.2).*

$$\min f(x) = -\frac{1}{2}\sqrt{x_1} - \frac{1}{2}x_2$$
$$x_1 \geq 0\,,\ x_1 + x_2 \leq 1$$

Wir wissen bereits:

$$\tilde{x}_1 = \frac{1}{4}\,,\ \tilde{x}_2 = \frac{3}{4}\,,\ \mu_1 = \mu_2 = 0\,,\ \mu_3 = \frac{1}{2}$$

erfüllen die notwendigen Kuhn-Tucker-Bedingungen.

Hinreichende Bedingung: Letzte Ungleichung ist streng aktiv, daher

$$d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \quad \forall d : d_1 + d_2 = 0\,.$$

$$f''(x):\ \textit{Nur}\ f_{x_1x_1} \neq 0; \quad f_{x_1} = -\frac{1}{4}\,x_1^{-1/2}$$
$$f_{x_1x_1} = \frac{1}{8}\,\tilde{x}_1^{-3/2} = \frac{1}{8}\left(\frac{1}{4}\right)^{-3/2} = 1$$

$f''(\tilde{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ *ist nicht positiv definit auf* $\mathbb{R}^2$ *aber auf dem Unterraum,*

welcher den streng aktiven Ungleichungen zugeordnet ist, denn

$$d_1 + d_2 = 0 \Rightarrow d_1^2 = d_2^2$$
$$\Rightarrow d^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} d = d_1^2 = \frac{1}{2}\,d_1^2 + \frac{1}{2}\,d_1^2$$
$$= \frac{1}{2}\,d_1^2 + \frac{1}{2}\,d_2^2 = \frac{1}{2}\|d\|^2\,.$$

Hinreichende Bedingung ist erfüllt mit $\alpha = \frac{1}{2}$.

Bemerkung: Man sieht, dass die Verifizierung der hinreichenden Bedingungen schwierig sein kann.

(5.53) ist i.A. nur numerisch nachprüfbar. Dabei enthält diese Bedingung die Forderung $\langle g^j, d\rangle \leq 0$ für $j \in J$ mit $\mu_j = 0$. Diese Bedingung lässt sich schwer verarbeiten. Man lässt diese Einschränkung deshalb einfach weg und hofft damit, die *stärkere* hinreichende Bedingung

$$d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \qquad (5.52)'$$

bei

$$Ad = 0\,,\ \langle g^j, d\rangle = 0 \quad \forall j \in J(\tilde{x}) \text{ mit } \mu_j > 0$$

verifizieren zu können. Dazu bilden wir die Matrix

$$B = B(\tilde{x}) = \begin{pmatrix} a_1^T \\ \ddots \\ a_m^T \\ \ddots \\ (g^j)^T \\ \ddots \end{pmatrix} \quad \text{mit allen} \quad g^j \in J \text{ mit } \mu_j > 0\,.$$

$\Rightarrow$

$$\boxed{d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \quad \forall d \in \ker B\,.} \tag{5.54}$$

Nun bestimmt man eine Nullraum-Matrix zu $B(\tilde{x})$, dies sei $Z(\tilde{x})$. Hat man diese, dann ist (5.54) äquivalent zur positiven Definitheit von $Z(\tilde{x})^T f''(\tilde{x}) Z(\tilde{x})$.

Ein weiteres Beispiel

Beispiel 5.5.4 *(Nichtkonvexe quadratische Optimierung)*

$$\boxed{\begin{array}{l} \min f(x) \quad = -x^2 + 1 \\ x \geq -1\,,\ x \leq \frac{1}{2} \end{array}}$$

Globales Min: $x = -1$
Lokales Min: $x = \frac{1}{2}$

Wir untersuchen $\tilde{x} = \frac{1}{2}$ *näher:*

$$\mathcal{L}(x, \mu_1, \mu_2) = -x^2 + 1 + \mu_1(-1 - x) + \mu_2 \left(x - \frac{1}{2}\right)\,.$$

Offenbar muss $\mu_1 = 0$ *sein bei* $\tilde{x} = \frac{1}{2}$*. Kuhn-Tucker-Bedingung* $\Rightarrow$

$$\mathcal{L}_x = -2\tilde{x} + \mu_2 = 0 \quad \Rightarrow \boxed{\mu_2 = 1}\,.$$

Hinreichende Bedingung: $d\mathcal{L}_{xx} d \geq \alpha d^2 \quad \forall d \in R$ *mit* $d = 0$.

Erfüllt für jedes α*. Damit ist (5.53) erfüllt, aber, wie wir gesehen haben,* $\tilde{x} = \frac{1}{2}$ **nur lokal optimal**.

5.5.4 Strikte Komplementarität

Manche numerischen Verfahren und manch theoretische Aussage für (PLU) bleibt nur richtig bei strikter Komplementarität. Was heißt das?

Definition 5.5.5 *In $\tilde{x} \in \mathcal{F}$ ist die Bedingung der strikten Komplementarität erfüllt, wenn gilt*

$$j \in J(\tilde{x}) \Rightarrow \mu_j > 0\,,$$

d. h., wenn alle aktiven Ungleichungsrestriktionen streng aktiv sind.

Folgerung: Hier gilt

$$\begin{aligned} \langle g^j, \tilde{x}\rangle = r_j &\Rightarrow \mu_j > 0 \\ \mu_j = 0 &\Rightarrow \langle g^j, x\rangle < r_j \quad j = 1, ..., p\,. \end{aligned}$$

In diesem Falle sind dann die hinreichenden Bedingungen (5.53) und (5.52)' äquivalent. Wir bilden

$$B(\tilde{x}) = \begin{pmatrix} a_i^T \\ (g^j)^T \end{pmatrix} \begin{matrix} i \in 1, ..., m \\ j \in J(\tilde{x}) \end{matrix}$$

und fordern

$$d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \quad \forall d \in \ker B(\tilde{x})\,. \tag{5.55}$$

Eine weitere Bedingung für spätere Zwecke ist die Voraussetzung

Lineare Unabhängigkeit der Gradienten der aktiven Restriktionen

• $a^1, ..., a^m$, $\{g^j\}_{j \in J(\tilde{x})}$ seien linear unabhängig.

Dann gilt:

$$\mathcal{A} = \begin{pmatrix} f''(\tilde{x}) & A^T & G(\tilde{x})^T \\ A & 0 & 0 \\ G(\tilde{x}) & 0 & 0 \end{pmatrix} \quad \text{ist invertierbar}$$

$$\text{mit } G(\tilde{x}) = ((g^j)^T)_{j \in J(\tilde{x})}$$

5.5.5 Probleme mit Variationsbeschränkungen (box constraints)

In diesem Falle ist der zulässige Bereich $\mathcal{F}$ ein Quader in $\mathbb{R}^n$,

$$\mathcal{F} = \{x \in \mathbb{R}^n | v_i \leq x_i \leq w_i,\ i = 1, ..., n\}\,.$$

Wir untersuchen also die (sehr oft auftretende) Aufgabe

$$\boxed{\begin{array}{l} \min_{x \in \mathbb{R}^n} f(x) \\ v \leq x \leq w\,. \end{array}} \tag{PB}$$

Dabei seien $v < w$ Vektoren der unteren und oberen Schranken.

Die allgemeine, bisher entwickelte Theorie liefert hier sehr einfach Beziehungen.

(i) Umformung in die allgemeine Form

$$v \le x \le w \quad \Leftrightarrow \quad \begin{matrix} -x & \le -v \\ x & \le w \end{matrix} \quad \Leftrightarrow \underbrace{\begin{pmatrix} -I \\ I \end{pmatrix}}_{G} x \le \underbrace{\begin{pmatrix} -v \\ w \end{pmatrix}}_{r}$$

$$Gx \le r$$

$$G = \begin{pmatrix} -1 & & \\ & \vdots & \\ & & -1 \\ 1 & & \\ & \vdots & \\ & & 1 \end{pmatrix}$$

(ii) Lineare Unabhängigkeit der aktiven Gradienten

Ist erfüllt, denn obere und untere Restriktionen können wegen $v < w$ nie gleichzeitig aktiv sein (siehe auch die Form von G).

$\Rightarrow$ die Multiplikatoren μ_j sind eindeutig bestimmt. Aber das bekommen wir alles noch direkter:

(iii) Notwendige Optimalitätsbedingungen

$$L = L(x, \mu) = f(x) + \sum_{i=1}^{n} \mu_i^u(-x_i + v_i) + \sum_{i=1}^{n} \mu_i^o(x_i - w_i)$$

$$\frac{\partial L}{\partial x_i} = 0 \quad \Leftrightarrow \quad \boxed{\frac{\partial f}{\partial x_i} - \mu_i^u + \mu_i^o = 0\,,\ \mu_i^u, \mu_i^o \ge 0}$$

μ_i^u: Multiplikatoren zu den unteren Schranken
μ_i^o: Multiplikatoren zu den oberen Schranken

Zusatzinformation: Genau einer der Multiplikatoren *darf* positiv sein (muss aber nicht). Einer ist *immer* Null.

$$\begin{aligned} \Rightarrow \frac{\partial f}{\partial x_i} = 0 \quad &\Rightarrow \quad \mu_i^u = \mu_i^o = 0 \\ \frac{\partial f}{\partial x_i} > 0 \quad &\Rightarrow \quad \mu_i^u = \frac{\partial f}{\partial x_i}\,,\ \mu_i^o = 0 \\ \frac{\partial f}{\partial x_i} < 0 \quad &\Rightarrow \quad \mu_i^u = 0\,,\ \mu_i^o = -\frac{\partial f}{\partial x_i} \end{aligned}$$

Folgerung:

$$\boxed{\begin{aligned} \mu_i^u &= \left(\tfrac{\partial f}{\partial x_i}\right)^+ \\ \mu_i^o &= \left(\tfrac{\partial f}{\partial x_i}\right)^- \end{aligned}}$$

(iv) Hinreichende Bedingung 2. Ordnung

Betrachten wir die Sache ganz anders, aus einem völlig anderen Blickwinkel als bisher in der allgemeinen Theorie. Die *notwendigen Bedingungen* schreiben wir einmal nicht in Kuhn-Tucker-Form auf, sondern als *Variationsungleichung*, d. h.

$$\langle \nabla f(\tilde{x}), x - \tilde{x} \rangle \geq 0 \quad \forall x \text{ mit } v \leq x \leq w\,.$$

Das heißt

$$\begin{aligned} &\langle \nabla f(\tilde{x}), \tilde{x} \rangle \leq \langle \nabla f(\tilde{x}), x \rangle \quad \forall x \text{ mit } v \leq x \leq w \\ &\Updownarrow \\ &\langle \nabla f(\tilde{x}), \tilde{x} \rangle = \min_{v \leq x \leq w} \langle \nabla f(\tilde{x}), x \rangle \\ &\Updownarrow \\ &\frac{\partial f}{\partial x_i} \cdot \tilde{x}_i = \min_{v_i \leq x \leq w_i} \frac{\partial f}{\partial x_i}\, x\,. \end{aligned}$$

Folglich

$$\left.\begin{aligned} &\frac{\partial f}{\partial x_i} > 0 \quad \Rightarrow \tilde{x}_i = v_i \\ &\frac{\partial f}{\partial x_i} < 0 \quad \Rightarrow \tilde{x}_i = w_i \end{aligned}\right\} \text{ In diesen Fällen ist } \tilde{x}_i \text{ durch die notwendigen Bedingungen bei Kenntnis von } \nabla f \text{ festgelegt!}$$

$$\frac{\partial f}{\partial x_i} = 0 \quad \text{: keine Aussage (außer eben } \frac{\partial f}{\partial x_i} = 0\text{).}$$

Bei den Komponenten mit $\frac{\partial f}{\partial x_i} = 0$ brauchen wir also zusätzliche Informationen. Die holen wir uns aus hinreichenden Bedingungen 2. Ordnung. Wir fordern (etwas zu stark)

$$\begin{aligned} &d^T f''(\tilde{x}) d \geq \alpha \|d\|^2 \\ &\text{für alle} \quad d \in \mathbb{R}^n \quad \text{mit} \quad d_i = 0 \quad \text{falls} \quad \left|\frac{\partial f}{\partial x_i}\right| \neq 0\,. \end{aligned}$$

Numerisch kann das so bewerkstelligt werden:

$$Z := \operatorname{diag}(c_i)_{i=1,\ldots,n} \quad \text{mit} \quad c_i = \begin{cases} 1 & \left|\frac{\partial f}{\partial x_i}\right| = 0 \\ 0 & \text{sonst} \end{cases}.$$

Dann ist obige Bedingung äquivalent zur positiven Definitheit von

$$Z^T f''(\tilde{x}) Z\,.$$

5.6 Lineare Optimierungsprobleme

Wir betrachten die Aufgabe, ein *lineares* Funktional f bei linearen Restriktionen zu minimieren

$$\boxed{\begin{aligned} &\min f(x) = c^T x \\ &\text{bei} \quad Ax = b \\ &\qquad\; x \geq 0\,. \end{aligned}} \tag{LP}$$

(LP) heißt *lineare Optimierungsaufgabe*. Zur Herleitung der Kuhn-Tucker-Bedingungen schreiben wir $\mathcal{F}$, die zulässige Menge, wieder um:

$$\mathcal{F} = \{x | Ax = b\,,\ Gx \leq 0\} \quad \text{mit} \quad G = -I\,.$$

Wegen $\nabla f = c$ erhalten wir die notwendigen Bedingungen

$$\begin{aligned} c + A^T\lambda + G^T\mu &= 0 \\ \mu \geq 0\,,\ \langle Gx, \mu\rangle &= 0 \end{aligned}$$

also $c = \mu - A^T\lambda$, und wenn wir $y = -\lambda$ setzen

$$\boxed{A^T y + \mu = c\,,\ \mu \geq 0\,,\ \mu_i x_i = 0\,.} \qquad (*)$$

Beachte, dass (LP) eine konvexe Aufgabe ist. $\Rightarrow$

Satz 5.6.1 *$\tilde{x}$ ist genau dann (globales) Minimum von* (LP)*, wenn $\mu \geq 0$ und $y \in \mathbb{R}^m$ existieren mit* $(*)$*.*

Bemerkung: Hier kann man die Lagrangeschen Multiplikatoren als Lösung einer dualen Optimierungsaufgabe erhalten. Diese konstruiert man leicht mit folgendem Trick als Lagrange-duale Aufgabe:

$$\begin{aligned} L(x, \lambda) &= f(x) + \langle \lambda, Ax - b\rangle \\ &= \langle c, x\rangle + \langle \lambda, Ax - b\rangle\,. \end{aligned}$$

Die ëinfachenRestriktionen $x \geq 0$ nehmen wir am besten *nicht* in die Lagrange-Funktion. Dann folgt

$$\text{(LP)} \quad \Leftrightarrow \quad \min_{x\geq 0} \Big(\underbrace{\max_{\lambda\in\mathbb{R}^m}(f(x) + \langle \lambda, Ax - b\rangle)}\Big)$$

falls $Ax - b \neq 0$, so bekommt man $\max = +\infty$ als Ergebnis,
und diese x fallen bei der Minimierung raus.

Das Dualproblem (DP) erhält man durch Vertauschen von min und max:

$$\begin{aligned} &\max_{\lambda\in\mathbb{R}^m} \Big(\min_{x\geq 0} f(x) + \langle \lambda, Ax - b\rangle \Big) \\ =&\max_{\lambda\in\mathbb{R}^m} \Big(\underbrace{\min_{x\geq 0}\langle c + A^T\lambda, x\rangle} - \langle b, \lambda\rangle \Big) \\ &\qquad = -\infty\,,\ \text{falls } c + A^T\lambda \not\geq 0 \\ &\qquad = 0\,, \qquad \text{falls } c + A^T\lambda \geq 0 \end{aligned}$$

$\Rightarrow$

$$\text{(DP)} \quad \Leftrightarrow \quad \boxed{\begin{array}{l} \max_{\lambda\in\mathbb{R}^m} -\langle b, \lambda\rangle \\ \text{bei} \quad A^T\lambda + c \geq 0 \end{array}}$$

Umformulierung: Setze $A^T\lambda + c = \mu$, $\lambda := -y$, dann

$$\text{(DP)} \quad \begin{cases} \max\langle b, y\rangle \\ \text{bei} \quad A^T y + \mu = c \end{cases} \quad \rightarrow \text{das ist genau die Form von Satz 5.6.1.}$$

6 Probleme mit linearen Restriktionen-Verfahren

6.1 Quadratische Optimierungsprobleme

Am einfachsten laufen die Dinge (wie generell) bei reinen Gleichungsrestriktionen

6.1.1 Aufgaben mit Gleichungsrestriktionen

Wir betrachten

$$\boxed{\begin{array}{l}\min\limits_{x\in\mathbb{R}^n} \dfrac{1}{2}\langle x, Qx\rangle + \langle q, x\rangle \\ \text{bei} \quad Ax = b\end{array}} \tag{QG}$$

Q : symmetrisch, (n,n), $A : (m,n)$, $m \leq n$.

Voraussetzung:

- 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 $\tilde{\lambda}$ (Satz 5.4.6). Beide erfüllen zusammen das System

$$\mathcal{A}\begin{pmatrix}\tilde{x}\\ \lambda\end{pmatrix} = \begin{pmatrix}-q\\ b\end{pmatrix} \quad \text{mit} \quad \mathcal{A} = \begin{pmatrix}Q & A^T\\ A & 0\end{pmatrix} \tag{6.56}$$

$\mathcal{A}$ ist invertierbar (Lemma 5.4.1). Damit braucht man „nur“ $\mathcal{A}^{-1}$ zu berechnen. *Das ist aber in der Regel zu teuer!*

Bessere Idee: Finden einer Nullraum-Matrix zu A und dadurch Elimination der Nebenbedingung $Ax = b$. 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

$$\boxed{HA^T = \begin{pmatrix}R\\ 0\end{pmatrix}} \qquad \begin{array}{l}H : (n,n)\\ R : (m,m)\end{array}$$

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 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:

$$\begin{aligned} & \tilde{x} = Y\tilde{x}_y + Z\tilde{x}_z \\ \Rightarrow \quad & b = A\tilde{x} = AY\tilde{x}_y + 0 \\ & b = AY\tilde{x}_y = R^T\tilde{x}_y \qquad \boxed{R^T\tilde{x}_y = b} \\ \Rightarrow \; & \tilde{x}_y = (R^T)^{-1}b \end{aligned}$$

(was man durch einfaches Auflösen des Gleichungssystems $R^T\tilde{x}_y = b$ bewerkstelligt.)

Schritt 2b: $''\mathcal{F}$ = spezielle Lösung von $Ax + b$ plus allgemeine von $Ax = 0''$

Spezielle Lösung: $Y\tilde{x}_y =: w$. Dann gilt $Aw = b$.

Somit:

$$\begin{aligned} & x \in \mathcal{F} \\ \Updownarrow \\ & x = w + Zz \,, \; z \in \mathbb{R}^{n-m} \,. \end{aligned}$$

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 $\frac{1}{2}w^TQw$ weglassen können:

$$f = \frac{1}{2} z^T \underbrace{Z^TQZ}_{\tilde{Q}} z + \underbrace{\langle Z^Tqw, z\rangle + \langle Z^Tq, z\rangle}_{=\tilde{q}^Tz}$$

$\Rightarrow$ neue Form

$$\boxed{\min_{z\in\mathbb{R}^{n-m}} F(z) = \frac{1}{2}\, z^T \tilde{Q} z + \tilde{q}^T z} \qquad (\mathrm{QG})_r$$

$$\begin{aligned} \text{mit} \quad \tilde{Q} &:= Z^T Q Z \\ \tilde{q} &:= Z^T Q w + Z^T q\,. \end{aligned}$$

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 $Z^T QZ$ und $Z^T q$ betrachten wir zunächst das Ganze für die größere, Z^T enthaltene Matrix H:

- Bilde

$$\boxed{-Hq =: \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \begin{matrix} \leftarrow \mathbb{R}^m \\ \leftarrow \mathbb{R}^{m-n} \end{matrix} =: h}$$

- Berechne

$$\boxed{B := HQH^T = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}}$$

. . . . . .

NR:

$$HQH^T = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} Q (Y, Z) = \begin{pmatrix} Y^T \\ Z^T \end{pmatrix} (QY, QZ) = \begin{pmatrix} Y^T QY & Y^T QZ \\ Z^T QY & Z^T QZ \end{pmatrix}$$

. . . . . .

$$\begin{aligned} \Rightarrow \quad & B_{11} = Y^T QY \quad B_{12} = Y^T QZ & & \\ & B_{21} = Z^T QY \quad B_{22} = Z^T QZ \quad \text{und} \quad & h_1 &= -Y^T q \\ & \qquad\qquad\quad = \tilde{Q} & h_2 &= -Z^T q\,. \end{aligned}$$

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

$$\boxed{B_{22}\tilde{x}_z = h_2 - B_{21}\tilde{x}_y}$$

Lösung z.B. mit Cholesky-Zerlegung.

$\Rightarrow$ haben am Ende $\tilde{x}_z$, $\tilde{x}_y$ und dann

$$\boxed{\tilde{x} := Y\tilde{x}_y + Z\tilde{x}_z}\,.$$

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.

$\Rightarrow$

$$\underbrace{\overset{T}{H}QH}_{B}\begin{pmatrix}\tilde{x}_y\\ \tilde{x}_z\end{pmatrix} + \underbrace{\overset{T}{H}A}_{\binom{R}{0}}\lambda = \underbrace{-Hq}_{h}$$

d. h.

$$\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix}\begin{pmatrix}\tilde{x}_y\\ x_z\end{pmatrix} + \begin{pmatrix}R\\ 0\end{pmatrix}\lambda = \begin{pmatrix}h_1\\ h_2\end{pmatrix}$$

$$\Rightarrow \qquad \boxed{R\lambda = h_1 - B_{11}\tilde{x}_y - B_{12}\tilde{x}_z}\,.$$

einfach zu lösen, da R obere Dreiecksmatrix. Damit ist alles gelöst!

6.1.2 Aufgaben mit Ungleichungsrestriktionen

Wir betrachten

$$\boxed{\begin{array}{l}\min\limits_{x\in\mathbb{R}^n} \dfrac{1}{2}\langle x, Qx\rangle + \langle q, x\rangle\\ \text{bei} \quad Ax = b \ \text{ und } \ Gx \le r\end{array}} \tag{QLU}$$

mit Q : symmetrisch, (n,n), $A : (m,n)$, $m \le n$, $G : (p,n)$. Die zulässige Menge ist

$$\mathcal{F} = \{x \in \mathbb{R}^n \,|\, Ax = b, Gx \le r\}.$$

Definition 6.1.1 *Ein Vektor d heisst zulässige Richtung, wenn $Ad = 0$ und $\langle g^j, d\rangle \le 0 \quad \forall j \in J(\tilde{x})$. Die Menge aller zulässigen Richtungen wird mit $K(\mathcal{F}, \tilde{x})$ bereichert,*

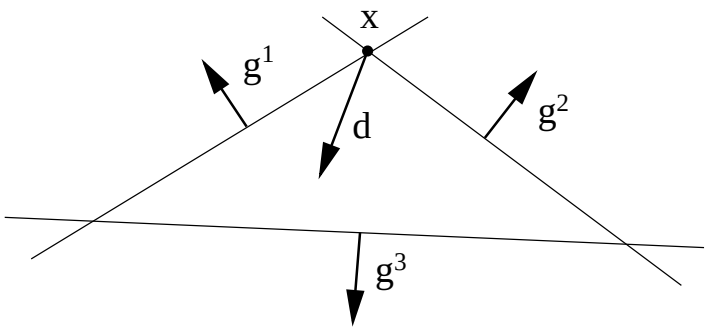
Nach Lemma 5.5.1 gilt für den Kegel der zulässigen Richtungen

$$K(\mathcal{F}, \tilde{x}) = L(\mathcal{F}, \tilde{x}) = \{d \in \mathbb{R}^n | Ad = 0\,,\ \langle g^j, d\rangle \leq 0 \quad \forall j \in J(\tilde{x})\}\,.$$

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

$$\boxed{d \text{ ist zulässige Richtung im Punkt } x \quad \Leftrightarrow \quad Ad = 0\,,\ G(x)d \leq 0}$$

Geometrische Illustration: (3 Ungleichungsrestriktionen $\langle g^i, x\rangle \leq r^i$, $i = 1, 2, 3$)



$$\left.\begin{array}{l}\langle g^1, x\rangle = r^1 \\ \langle g^2, x\rangle = r^2\end{array}\right\} \text{aktiv} \qquad J(x) = \{1, 2\}$$

$$\left.\langle g^3, x\rangle < r^3\right\} \text{inaktiv} \qquad G(x) = \begin{pmatrix}(g^1)^T \\ (g^2)^T\end{pmatrix}$$

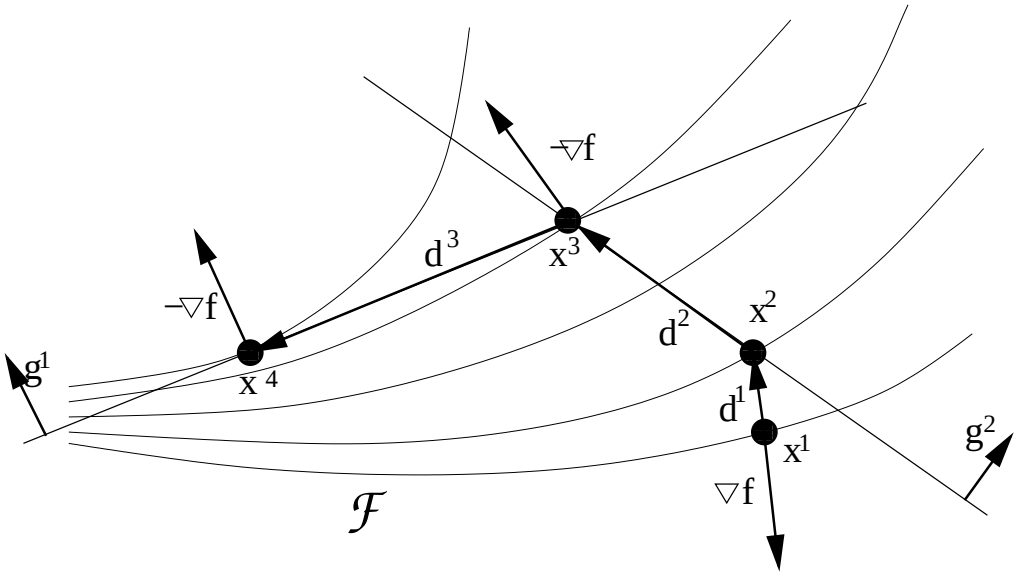
Anschaulich ist klar: Damit $x + td$ für kleine t zulässig bleibt, muss gelten

$$\langle g^1, d\rangle \leq 0 \quad \wedge \quad \langle g^2, d\rangle \leq 0\,.$$

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 $\mathcal{F}$, 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 $\mathcal{F}$ erreicht – die Restriktion Nr. 2 ist aktiv geworden.

- Wäre $g^2 || \nabla f$, 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
$$(*) \qquad -\nabla f = \mu^1 g^1 + \mu^2 g^2 \quad \text{mit} \quad \mu^1 > 0 \quad \text{aber} \quad \mu^2 < 0$$
$$0 = \nabla f + \mu^1 g^1 + \mu^2 g^2$$

- In welcher Richtung sollte weiter gesucht werden?
Die Richtung d muss erfüllen:
$$\left.\begin{array}{r} \langle g^1, d\rangle \leq 0 \\ \langle g^2, d\rangle \leq 0 \end{array}\right\} \text{ um zulässig zu bleiben}$$
$$\begin{array}{r} \langle \nabla f, d\rangle < 0 \\ \updownarrow \qquad\qquad \text{um einen Abstieg zu erzielen}\,. \\ \langle -\nabla f, d\rangle > 0. \end{array}$$

$(*) \Rightarrow$

$$\langle -\nabla f, d\rangle = \underbrace{\mu^1}_{>0}\underbrace{\langle g^1, d\rangle}_{\leq 0} + \underbrace{\mu^2}_{<0}\underbrace{\langle g^2, d\rangle}_{\leq 0}$$

Weil $\langle -\nabla f, d\rangle$ positiv sein soll, möglichst groß, 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.

- 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

$$\begin{aligned} 0 &= \nabla f + \mu^1 g^1 + 0 \cdot g^2 \\ \langle g^1, x^4\rangle &= r^1\,,\ \langle g^2, x^4\rangle < r^2 \\ \Big(\langle g^2, x^4\rangle - r^1\big) \cdot 0 &= 0\,. \end{aligned}$$

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 Gleichungsnebenbedingungen.

Wir kommen nun zur mathematischen präzisen Formulierung des Verfahrens

- Aktueller Iterationspunkt: x^k

  $J_k := J(x^k)$: Menge der aktiven Indizes (hier gilt $\langle g^j, x^i\rangle = r^i$, $i \in J_k$)
  $p_k = |J(x^k)|$: Zahl der aktiven Indizes
  $G_k = G(x^k)$: Matrix der g^i, $i \in J_k$ (genauer: mit Zeilen g_i^T)
  $B_k = \binom{A}{G_k}$: Beschreibt das zur Zeit aktive lineare Gleichungssystem

- Ausgehend von x^k wird ein im nächsten Schritt zu lösendes *quadratisches Optimierungsproblem* aufgestellt (mit Gleichungsrestriktionen)

$$\boxed{\begin{array}{l} \min\limits_{d\in\mathbb{R}^n} \dfrac{1}{2}\langle Qd, d\rangle + \langle Qx^k + q, d\rangle \\ \text{bei} \quad B_k d = 0 \end{array}} \qquad (Q_k)$$

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 μ_j^k, $j \in J_k$; λ_i^k (für Gleichungsrestriktionen)
- Wir ergänzen diese durch $\mu_j^k = 0$, $j \notin J_k$

$$\begin{aligned} \text{Insgesamt:} \quad & \lambda^k \in \mathbb{R}^m \quad (m \text{ Gln.}) \\ & \mu^k \in \mathbb{R}^{p_k} \quad \text{bzw. } \tilde{\mu}^k \in \mathbb{R}^p \quad (\text{durch Nullen aufgefüllt}) \end{aligned}$$

Voraussetzungen für das Verfahren:

- B_k hat immer vollen Rang (Lineare Unabhängigkeit des Systems $a^i, i = 1, ..., m$; 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):

$$\mathcal{F}_k = \{d \in \mathbb{R}^n | Ad = 0\,,\ G_k d = 0\} \subset L(\mathcal{F}, x^k)\,.$$

Damit ist jedes $d \in \mathcal{F}_k$ 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.

$$\boxed{Q(d^k + x^k) + q + A^T\lambda^k + G_k^T\mu^k = 0\,.} \qquad (\dagger)$$

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 $\boxed{d^k = 0 \quad \text{aber} \quad \mu^k \ngeq 0\,.}$

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:

$$\boxed{\tilde{J}_k := J_k \backslash \{j\}\,.}$$

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 $d^k = 0$

$$\Rightarrow \qquad Qx^k + q + A^T\lambda^k + G_k^T\mu^k = 0$$

$$Q\tilde{d}^k + Qx^k + q + A^T\tilde{\lambda}^k + \tilde{G}_k^T\tilde{\mu}^k = 0$$

Wäre $\tilde{d}^k = 0$, so

$$A^T\lambda^k + G_k^T\mu^k = A^T\tilde{\lambda}^k + \tilde{G}_k^T\tilde{\mu}^k$$

$$\Rightarrow \quad \mu_j g^j = \quad \text{Linearkombination der anderen auftretenden } a^i, g^i$$

$$\Rightarrow \quad \text{wegen } \mu_j \neq 0 \quad \text{gilt Gleiches für } g^j \text{ Widerspruch.}$$

Somit haben wir noch zu diskutieren:

Fall 3 $\boxed{d^k \neq 0\,.}$ Dann ist d^k **Abstiegsrichtung**.

Beweis: Aus den notwendigen Bedingungen (†) für x^k folgt

$$\begin{aligned}\nabla f(x^k) = Q^k x^k + q &= -Qd^k - A^T\lambda^k - G_k^T\mu^k \\ &= -Qd^k - B_k^T\begin{pmatrix}\lambda^k \\ \mu^k\end{pmatrix} / \cdot d^k \\ \Rightarrow \langle \nabla f(x^k), d^k\rangle &= -\underbrace{\langle d^k, Qd^k\rangle}_{<0 \text{ wegen Definitheit}} - \left\langle \begin{pmatrix}\lambda^k \\ \mu^k\end{pmatrix}, \underbrace{B_k d^k}_{=0} \right\rangle \\ &< 0\,.\end{aligned}$$

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

$$\langle g^j, \tilde{d}^k\rangle \leq 0\,.$$

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 $\mathcal{F}$.

Zusammengefasst ergibt sich folgendes

Verfahren 6.1.1 (`Aktive-Mengen-Strategie für` (QU))

1. *Berechne Startpunkt $x^0 \in \mathcal{F}$, setzte $k := 0$.*
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 $\mu_j^k = \min\{\mu_i^k \,,\ i \in J_k\}$*
   - *$J_k := J_k \backslash \{j\}$*
   - *streiche in G_k die zu j gehörige Zeile*
   - *Löse das entsprechende neue Problem (Q_k). Das Ergebnis ist auf jeden Fall $d^k := \tilde{d}^k \neq 0$*
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**

$$\begin{aligned} f(x^k + td^k) &= f(x^k) + \underbrace{t\nabla f(x^k)d^k}_{= -t\langle d^k, Qd^k\rangle \text{ siehe Endergebnis von } \textbf{Fall 3}.} + \frac{1}{2}\, t^2 \langle d^k, Qd^k \rangle \\ &= f(x^k) + \underbrace{\left(\frac{1}{2}\, t^2 - t\right)}_{\text{minimal bei } t = 1} \langle d^k, Qd^k \rangle \end{aligned}$$

$\Rightarrow$ Aus Sicht des maximalen Abstiegs wäre $\tau = 1$ zu setzen.

- **Zulässigkeit von x^{k+1}**

Für die letzte Iterierte x^k galt

$$\begin{aligned}
&\langle a^i, x^k\rangle = b_i \quad i = 1, ..., m\\
&\langle g^i, x^k\rangle = r_i \quad i \in J_k \quad \text{(aktive Ungln.)}\\
&\langle g^i, x^k\rangle < r_i \quad i \notin J_k \quad \text{(inaktive Ungln.)}\,.
\end{aligned}$$

Durch die Wahl von d^k ist gesichert *für alle* $t \geq 0$

$$\begin{aligned}
&\langle a^i, x^k + td^k\rangle = b_i \quad \text{denn } \langle a^i, d^k\rangle = 0\\
&\langle g^i, x^k + td^k\rangle = r_i \quad \forall i \in J_k\backslash\{j\}\,, \quad \text{aus gleichem Grund}\\
&\langle g^j, x^k + td^k\rangle < r_j \quad \text{denn bei der inaktivierten Restriktion gilt } \langle g^j, d^k\rangle < 0\,.
\end{aligned}$$

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

$$\langle g^j, d^k\rangle + t\langle g^j, d^k\rangle \leq r_j\,, \quad \text{Maximum von } t : \text{bei Gleichheit}$$

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:

$$\begin{aligned}
&\tau_k = \min_{j\in I_k}\left\{\frac{r_j - \langle g^j, x^k\rangle}{\langle g^j, d^k\rangle}\right\} \quad \text{falls} \quad I_k \neq \phi\\
&\qquad \text{wobei} \quad I_k = \{j/\langle g^j, d^k\rangle > 0\}\,.\\
&\tau_k := \infty \quad \text{falls} \quad I_k = \phi \quad \text{(keine Beschränkung nötig)}\,.
\end{aligned}$$

Insgesamt: $\boxed{\sigma_k = \min(1, \tau_k)\ .}$

Damit ist das Verfahren vollständig beschrieben. Es gilt

Satz 6.1.1 *Es sei Q positiv definit auf ker A und für alle $x \in \mathcal{F}$ habe die Matrix $B(x) = \binom{A}{G(x)}$ 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

$$\min f(x) \quad \text{bei} \quad Ax = b\,,\ \langle g^j, x\rangle = r_i \quad \forall i \in J \tag{1}$$

von Bedeutung, wobei J eine beliebige Teilmenge von $\{1, ..., p\}$ ist, die für mögliche aktive Ungleichungsrestriktionen steht. Wir lösen im Verfahren nicht direkt (1), aber es treten Optimalitätssysteme von (1) auf, nämlich die Gleichungen

$$Qx + q + A^T\lambda + G(x)^T\mu = 0\,, \tag{2}$$

wobei x jeweils für die eindeutig bestimmt 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 $\{a^i\}_{i=1,...,m} \cup \{g^j\}$, $j \in J_k$ 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 $I_{k+i} = \phi$ 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

$$\boxed{\begin{array}{l}\min f(x)\\ Ax = b\end{array}} \tag{PLG}$$

mit $f : \mathbb{R}^n \to \mathbb{R}$, jetzt nicht mehr notwendig quadratisch. Also $\min\limits_{x\in\mathcal{F}} f(x)$ mit $\mathcal{F} = \{x \in \mathbb{R}^n / Ax = b\}$. Es sei wieder $w \in \mathcal{F}$ eine spezielle Lösung von $Ax = b$ und $Z : \mathbb{R}^l \to \ker A$ eine Nullmatrix. Dann wird die unrestringierte Aufgabe

$$\boxed{\min_{z\in\mathbb{R}^l} F(z) := f(w + Zz)} \tag{6.1}$$

gelöst. Die Bestimmung einer Nullraummatrix hängt nicht von f ab, nur von $\mathcal{F}$, 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.

$$\begin{aligned} \textit{Für} \quad f &: x^{k+1} = x^k + \sigma_k d^k \\ F &: z^{k+1} = z^k + \sigma_k v^k \end{aligned}$$

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

$$x^{k+1} = w + Zz^{k+1} = \underbrace{w + Zz^k}_{x^k} + \sigma_k \underbrace{Zv^k}_{d^k} = x^k + \sigma_k d^k\,.$$

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 $x^0 \in \mathcal{F}$, Nullraum-Matrix Z, $k := 0$.*
2. *Wenn* $\underbrace{Z^T \nabla f(x^k)}_{\text{reduzierter Gradient}} = 0$*:* **Stop**.

3. Ansonsten berechne Abstiegsrichtung $\boxed{d^k := Zv^k,}$ *effiziente Schritweite* 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:

$$v^k := -\nabla F(z^k) = -Z^T \nabla f(x^k)$$

$$\Rightarrow \quad \boxed{d^k = Zv^k = -ZZ^T \nabla f(x^k)}$$

Spezialfall: $Z = P$*, Projektionsmatrix auf ker* A, $\Rightarrow$

Projiziertes Gradientenverfahren: $Z = P$

$$d^k = -ZZ^T \nabla f(x^k) = -ZZ \nabla f(x^k)$$

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

Variable-Metrik-Verfahren (reduziert):

Folge $\{A^{(k)}\}$ *positiv definierter Matrizen;*

$$v^k = -(A^{(k)})^{-1} \nabla F(z^k) = -(A^{(k)})^{-1} Z^T \nabla f(x^k)$$

$$\Rightarrow \quad \boxed{d^k = Zv^k = -Z(A^{(k)})^{-1} Z^T \nabla f(x^k)}$$

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 (ξ_i, η_i), $i = 1, ..., m$
Ansatz $\eta(\xi) = g(x, \xi)$
Gesucht *Vektor x und vorher aber: geeigneter Ansatz g.*
Idee: Splines mit Koeffizienten x.

Sei $\xi_1 < \xi_2 < ... < \xi_m$.
Wir überdecken das Intervall $[\xi_1, \xi_m]$ *durch* Knotenpunkte

$$\tau_0 < \tau_1 < ... < \tau_N\,,$$
$$\tau_0 \leq \xi_1, \xi_m \leq \tau_N\,.$$

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 $[\tau_i, \tau_{i+1}]$

$$g_i(x,\tau) = \frac{1}{\tau_{i+1}-\tau_i}(\gamma_{i+1}(\tau-\tau_i)^3 + \gamma_i(\tau_{i+1}-\tau)^3) + \beta_i(\tau-\tau_i) + \alpha_i$$
$$\gamma_0 = \gamma_N = 0$$
$$\gamma_i = g_i''(x,\tau_i)/6 \quad \text{“Momente”}$$
$$i = 1, ..., N-1\,.$$

Durch diese Wahl ist g'' automatisch stetig. Das heißt nicht, dass auch g und g' stetig sein müssen. Diese Forderung ergibt zusätzliche Bedingungen an

$$x = (\alpha_0, ..., \alpha_{N-1}, \beta_0, ..., \beta_{N-1}, ..., \gamma_1, ..., \gamma_{N-1})\,.$$

Nämlich: Stetigkeit von g': $\quad \Delta\tau_i := \tau_{i+1} - \tau_i$

$$g_i'(x,\tau_i) = g_{i-1}'(x,\tau_i)$$
$$3\gamma_i\Delta\tau_i + \beta_i = \beta_{i-1} + 3\gamma_i\Delta\tau_{i-1}$$
$$\Rightarrow \qquad \beta_i = \beta_{i-1} + 3\gamma_i(\Delta\tau_{i-1} - \Delta\tau_i)$$

$\Rightarrow$ 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

$$\begin{aligned} \min f(x) &= \sum_{i=1}^{m} (\eta_i - g(x, \xi_i))^2 \\ \text{bei} \quad \beta_i &= \beta_{i-1} + 3\gamma_i(\Delta\tau_i - \Delta\tau_{i-1}) \\ \alpha_i &= \alpha_{i-1} + \beta_{i-1}\Delta\tau_{i-1} + \gamma_i(\Delta\tau_{i-1}^2 - \Delta\tau_i^2) \\ & i = 1, ..., N-1\,. \end{aligned}$$

Diese Aufgabe kann man reduzieren auf eine *freie Optimierungsaufgabe* in den Variablen:

$$(\alpha_0, \beta_0, \gamma_1, ..., \gamma_{N-1})^T\,.$$

6.3 Ungleichungsnebenbedingungen – nichtquadratische Zielfunktionen

Jetzt ist eine Aufgabe der Bauart

$$\boxed{\begin{array}{l} \min f(x) \\ Ax = b\,,\ Gx \leq r \end{array}}$$

gegeben.

Grundidee: Taylorapproximation von f bis zur zweiten Ordnung $\rightsquigarrow$ 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

$$\min_{x\in\mathbb{R}^n} f(x) \qquad\qquad \min_{x\in\mathcal{F}} f(x)\,.$$

$\mathcal{F}$: konvexe Menge.

Die notwendigen Bedingungen 1. Ordnung für eine Lösung x^* lauten:

$$\nabla f(x^*) = 0 \qquad\qquad \langle \nabla f(x^*), x - x^* \rangle \geq 0 \quad \forall x \in \mathcal{F}\,.$$

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

$$(**) \qquad \min_{x\in\mathbb{R}^n} \frac{1}{2}\langle x - x^m, f''(x_n^k)(x - x^k)\rangle + \langle \nabla f(x^k), x - x^k\rangle\,.$$

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 $x \in \mathcal{F}$ mit hinzu:

$$\min_{x \in \mathcal{F}} \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, f''(x_n^k)(x - x^k) \rangle \,.$$

Lösung: $x = x^{n+1}$.
Und genau das macht man für (PLU)! Hier ist

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid Ax = b\,,\ Gx \leq r\}$$

konvex. Wir iterieren wie folgt:
x^k sei berechnet. Man stellt dann auf:

$$\boxed{\begin{array}{ll} \min & \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, f''(x^k)(x - x^k) \rangle \\ \text{bei} & Ax = b\,,\ Gx \leq r \end{array}} \qquad (QP_k)$$

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 $detf''(x^*) > 0$, so auch $detf''(x^k)$ 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

$$(QP_k) \quad \Leftrightarrow \quad \boxed{\begin{array}{ll} \min & \langle \nabla f(x^k), d \rangle + \frac{1}{2} \langle d, f''(x^k) d \rangle \\ \text{bei} & Ad = 0\,,\ Gx^k + Gd \leq r\,. \end{array}}$$

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)

- $B(x) = \binom{A}{G(x)}$ habe immer vollen Rang

  $\Rightarrow$ Multiplikatoren λ, μ sind eindeutig bestimmt.

Optimalitätsbedingung für (QP_k):

$$f''(x^k)d^k + \nabla f(x^k) + A^T\lambda^{k+1} + G^T\mu^{k+1} = 0 \tag{6.2}$$

$$\mu^{k+1} \geq 0\,,\ \langle \mu^{k+1}, Gx^k + Gd^k - r\rangle = 0\,. \tag{6.3}$$

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

$$\nabla f(x^k) + A^T\lambda^{k+1} + G^T\mu^{k+1} = 0\,,\ \mu^{k+1} \geq 0$$
$$\langle \mu^{k+1}, Gx^k - r\rangle = 0\,.$$

$\Rightarrow x^k$ erfüllt die Optimalitätsbedingungen $\Rightarrow$ STOP
(Optimalität folgt aus den hinreichenden Bedingungen).

Fall 2 $d^k \neq 0$.
Dann ist d^k Abstiegsrichtung, denn

$$\nabla f(x^k) = -f''(x^k)d^k - A^T\lambda^{k+1} - G^T\mu^{k+1} \mid \cdot\, d^k$$
$$\langle \nabla f, d^k\rangle = -\langle \underbrace{d^k, f''d^k}_{>0}\rangle - \langle \underbrace{Ad^k}_{=0}, \lambda^{k+1}\rangle - \langle \underbrace{Gd^k, \mu^{k+1}}_{\geq 0}\rangle$$
$$< 0 \quad \Rightarrow \quad \text{Abstiegsrichtung} \qquad \text{s.unten.}$$

Zur Diskussion von $\langle Gd^k, \mu^{k+1}\rangle$: Die Beschränkungen lauten ausgeschrieben

$$\langle g^i, x^k\rangle + \langle g^i, d^k\rangle \leq r_i\,.$$

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

$$\langle g^i, d^k\rangle = \underbrace{r_i - \langle g^i, x^k\rangle}_{\geq 0,\ \text{weil } x^k \text{ zulässig war}}\,,$$

also $\langle g^i, d^k\rangle \geq 0$. Insgesamt folgt daraus leicht $\langle Gd^k, \mu^{k+1}\rangle \geq 0$.

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.*

$$\begin{aligned}\|x^{k+1} - \tilde{x}\| + \|\lambda^{k+1} - \tilde{\lambda}\| + \|\mu^{k+1} - \tilde{\mu}\| \\ \leq c\big(\|x^k - \tilde{x}\|^2 + \|\lambda^k - \tilde{\lambda}\|^2 + \|\mu^k - \tilde{\mu}\|^2\big)\end{aligned}$$

7 Probleme mit nichtlinearen Restriktionen – Theorie

7.1 Grundlagen

Nun behandeln wir den allgemeinsten Fall von Aufgaben unserer Vorlesung, nämlich voll nichtlineare Aufgaben des Typs

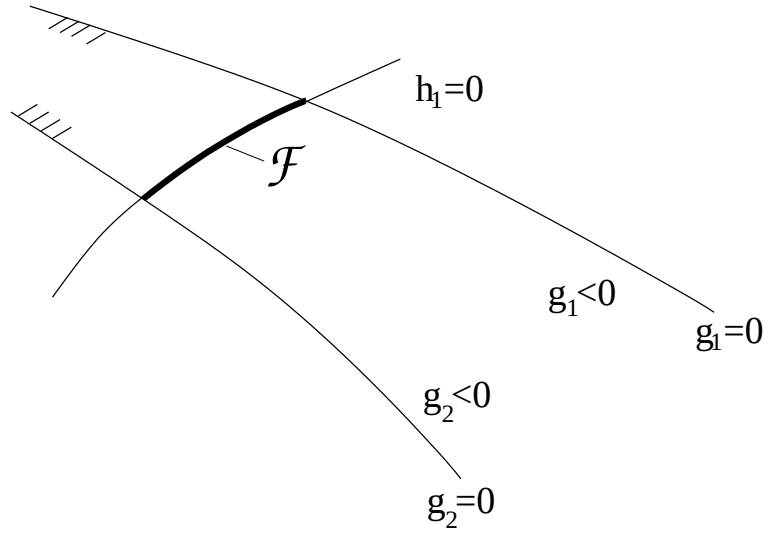
$$\begin{aligned}&\min f(x) \\ &h_i(x) = 0 \quad i = 1, ..., m \\ &g_j(x) \leq 0 \quad j = 1, ..., p\,,\end{aligned} \tag{PNU}$$

d. h. in vektorieller Form

$$\boxed{\begin{aligned}&\min f(x) \\ &h(x) = 0 \\ &g(x) \leq 0\end{aligned}} \qquad \begin{aligned}&h : \mathbb{R}^n \to \mathbb{R}^m \\ &g : \mathbb{R}^n \to \mathbb{R}^p\,.\end{aligned}$$

Hier gilt $\mathcal{F} = \{x \in \mathbb{R}^n \,|\, h(x) = 0\,,\ g(x) \leq 0\}$.

Illustration für $n = 2$:



7.2 Notwendige Optimalitätsbedingungen erster Ordnung

Es sei $\tilde{x}$ eine lokale Lösung von (PNU). Wir wollen nun wieder K-K-T-Sätze beweisen. Eine Schlüssel-Idee ist folgende: Löst $\tilde{x}$ (PNU), so sollte auch $\tilde{x}$ das Problem lösen, welches durch *Linearisierung an der Stelle* $\tilde{x}$ entsteht:

$$\boxed{\begin{array}{l}\min f(\tilde{x}) + f'(\tilde{x})(x-\tilde{x}) \\ \text{bei } \; h(\tilde{x}) + h'(\tilde{x})(x-\tilde{x}) = 0 \\ \qquad g(\tilde{x}) + g'(\tilde{x})(x-\tilde{x}) \leq 0\end{array}} \quad \textit{Linearisiertes Problem} \tag{7.1}$$

Sind h, g affin-linear, dann gilt das wirklich! Denn:

$$\begin{array}{lll} h(x) = Ax - b \quad \Rightarrow \quad h(\tilde{x}) + h'(\tilde{x})(x-\tilde{x}) & = A\tilde{x} - b + A(x-\tilde{x}) \\ g(x) = Gx - r & = Ax - b \\ \text{analog} \;\; g(\tilde{x}) + g'(\tilde{x})(x-\tilde{x}) & = Gx - r \end{array}$$

$\Rightarrow$ obige Nebenbedingungen sind äquivalent zu $Ax = b$, $Gx \leq r$, d. h. $x \in \mathcal{F}$.

Für $x \in \mathcal{F}$ galt aber die Variationsungleichung

$$\begin{array}{lll} & f'(\tilde{x})(x-\tilde{x}) \geq 0 & \forall x \in \mathcal{F}\,, \\ \text{d. h.} & f'(\tilde{x})\tilde{x} \leq f'(\tilde{x})x & \forall x \in \mathcal{F} \\ \text{d. h.} & \min\limits_{x\in\mathcal{F}} f'(\tilde{x})x = f'(\tilde{x})\tilde{x}\,. & \end{array}$$

Im nichtlinearen Fall gilt das leider nicht immer!

Beispiel 7.2.1

$$\begin{array}{l} \min f(x_1, x_2) = x_1 \\ \quad \textit{bei} \quad g_1(x) = -x_1^3 + x_2 \leq 0 \\ \qquad\;\; g_2(x) = -x_2 \leq 0\,. \end{array}$$

Umschreibung des zulässigen Bereichs: $x_2 \geq 0$, $x_2 \leq x_1^3$

Lösung offenbar: $\tilde{x}_1 = \tilde{x}_2 = 0$

Linearisierte Aufgabe (Tangente an $x_2 = x_1^3$ *anlegen!)*

$$\begin{aligned} &\min x_1 \quad \textit{bei} \quad x_2 \leq 0\,,\ -x_2 \leq 0 \\ \textit{also} \qquad &\boxed{\min x_1 \quad \textit{bei} \quad x_2 = 0} \\ &= -\infty \quad \textbf{keine Lösung!} \end{aligned}$$

Solche Phänomene müssen angeschlossen werden, denn Linearisierung spielt sowohl in der theoretischen Begründung als auch für die numerischen Verfahren eine wichtige Rolle. Deshalb hat man sich ausführlich mit Voraussetzungen an $\mathcal{F}$ befasst, welche so etwas ausschließen. Das sind **Regularitätsbedingungen** (*Constraint qualifications*). Auf Grund ihrer Wichtigkeit wollen wir diese ausführlich behandeln.

Zunächst sieht man, dass man (bei gegebenem $\tilde{x}$) im linearisierten Problem (7.1) die inaktiven Restriktionen weglassen kann – sie spielen lokal (d. h. in einer Umgebung von $\tilde{x}$) keine Rolle

$\Rightarrow$ Wir betrachten $J(\tilde{x}) = \{1 \leq j \leq p \,|\, g_j(\tilde{x}) = 0\}$ und

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} \langle \nabla f(\tilde{x}), x - \tilde{x} \rangle \\ &h'(\tilde{x})(x - \tilde{x}) = 0 \\ &\langle \nabla g_j(\tilde{x}), x - \tilde{x} \rangle \leq 0 \quad \forall j \in J(\tilde{x})\,. \end{aligned} \tag{7.2}$$

Dann gilt:

Lemma 7.2.1 *$\tilde{x}$ löst (7.1) genau dann, wenn $\tilde{x}$ auch (7.2) löst.*

(intuitiv klar, Beweis z.B in [1, Lemma 7.2.6]).

Folgerung 7.2.1 *$d = x - \tilde{x} = 0$ löst das Problem*

$$\begin{aligned} &\min_{d \in \mathbb{R}^n} \langle \nabla f(\tilde{x}), d \rangle \\ &h'(\tilde{x})d = 0 \\ &\langle \nabla g_j(\tilde{x}), d \rangle \leq 0 \quad j \in J(\tilde{x})\,. \end{aligned} \tag{7.3}$$

(Es gilt also keine zulässige Abstiegsrichtung, sonst wäre in (7.3) $\min = -\infty$*.)*

Definition 7.2.1 *Die Menge $L(\mathcal{F}, \tilde{x}) = \{d \,|\, h'(\tilde{x})d = 0,\ \langle \nabla g(j(\tilde{x}), d \rangle \leq 0 \quad \forall j \in J(\tilde{x})\}$ heißt Linearisierungskegel von $\mathcal{F}$ in $\tilde{x}$.*

Nun wollen wir zunächst annehmen, dass die lokale Lösung $\tilde{x}$ von (PNU) auch Lösung der linearisierten Aufgabe ist (was, wie wir wissen, schiefgehen kann). Dann folgt also

$$\begin{aligned} &\langle \nabla f(\tilde{x}), d \rangle \geq 0 \qquad\qquad \forall d \in L(\mathcal{F}, \tilde{x})\,. \\ \Rightarrow \quad &-\nabla f(\tilde{x}) \in L(\mathcal{F}, \tilde{x})^*\,. \end{aligned}$$

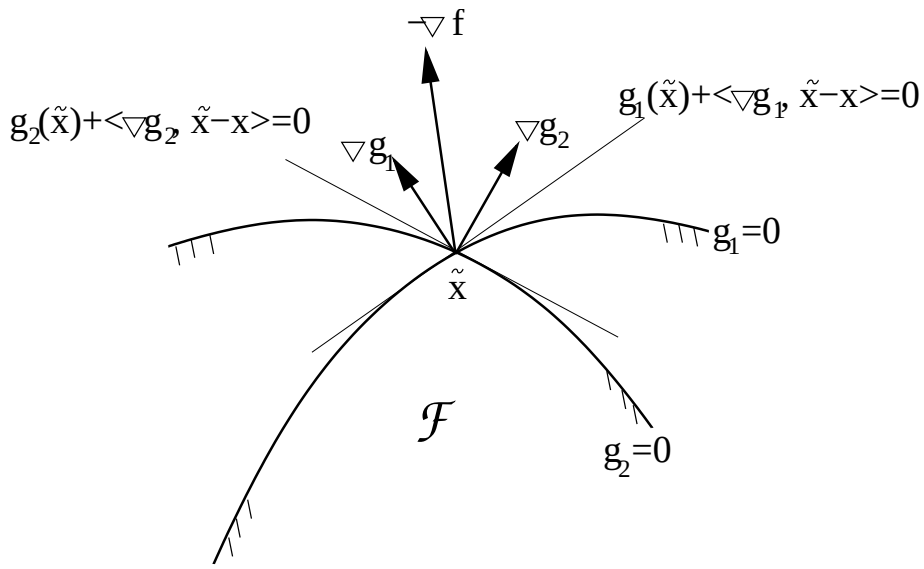
Wir haben bereits früher bewiesen, dass daraus folgt

$$-\nabla f(\tilde{x}) = \sum_{i=1}^{m} \lambda_i a_i + \sum_{j \in J(\tilde{x})} \mu_j g_i \,,\ \mu_j \geq 0$$

mit (angepasst an unseren Fall)

$$a_i = \nabla h_i(\tilde{x}) \,,\ g_j = \nabla g_j(\tilde{x}) \,.$$

Geometrische Illustration im Fall von reinen Ungleichungsrestriktionen:



Ist $\tilde{x}$ lokale Lösung, dann muss $-\nabla f$ in dem von $\nabla g_1, \nabla g_2$ aufgespannten Kegel liegen. Unter der Voraussetzung, dass $\tilde{x}$ das linearisierte Problem löst, gilt also

$$0 = \nabla f(\tilde{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\tilde{x}) + \sum_{j=1}^{p} \mu_j \nabla g_j(\tilde{x}) \quad \text{mit } \mu_j := 0 \quad j \notin J(\tilde{x})$$

oder

$$0 = \nabla f(\tilde{x}) + h'(\tilde{x})^T \lambda + g'(\tilde{x})^T \mu \tag{7.4}$$

$$\mu \geq 0 \quad \text{und} \quad \mu_j g_j(\tilde{x}) = 0 \quad \forall j \tag{7.5}$$

Definition 7.2.2 *λ und μ mit den Eigenschaften (7.4-7.5) heißen Lagrangesche Multiplikatoren zu $\tilde{x} \in \mathcal{F}$.*

Satz 7.2.1 *Ist $\tilde{x} \in \mathcal{F}$ und löst $\tilde{x}$ die linearisierte Aufgabe (7.1) (f, g, h seien differenzierbar in $\tilde{x}$), dann existieren Lagrangesche Multiplikatoren λ, μ zu $\tilde{x}$.*

Bemerkungen:

1. Dieser Satz war insofern schon klar, weil (7.1) eine differenzierbare Aufgabe mit linearen Restriktionen ist, und dafür kennen wir ja schon die Lagrangesche Multiplikatorenregel. Wie bisher, können wir diese Regel wie folgt aufschreiben:

   $$\mathcal{L} = \mathcal{L}(x, \lambda, \mu) := f(x) + \langle h(x), \lambda \rangle + \langle g(x), \mu \rangle \,.$$

   Lagrange-Funktion.

   Dann gilt

   $$\boxed{\nabla \mathcal{L}_x(\tilde{x}, \lambda, \mu) = 0 \,,\ \mu \geq 0 \,,\ \langle g(\tilde{x}), \mu \rangle = 0 \,.}$$

2. Ist f und $\mathcal{F}$ konvex, dann ist diese Optimalitätsbedingung auch hinreichend für Optimalität.

Wir haben Satz 7.2.1 aber nur unter der Bedingung gezeigt, dass $\tilde{x}$ die linearisierte Aufgabe löst. Wann gilt das? Dazu müssen wir etwas weiter ausholen, vorher aber rechnen wir zur Illustration noch ein Beispiel.

Beispiel 7.2.2

$$\begin{aligned} & \min f(x) = |x|^2 \,,\ x \in \mathbb{R}^3 \\ \textit{bei} \quad & 2x_1 - x_2 + x_3 \leq 5 \\ & x_1 + x_2 + x_3 = 3 \,. \end{aligned}$$

Wir wollen annehmen, dass die Multiplikatorenregel gilt.

$$\mathcal{L} = x_1^2 + x_2^2 + x_3^3 + \mu(2x_1 - x_2 + x_3 - 5) + \lambda(x_1 + x_2 + x_3 - 3)$$

$$\begin{aligned} \mathcal{L}x_1 = 0 \quad &\Rightarrow \quad 2x_1 + 2\mu + \lambda = 0 \\ \mathcal{L}x_2 = 0 \quad &\Rightarrow \quad 2x_2 - \mu + \lambda = 0 \\ \mathcal{L}x_3 = 0 \quad &\Rightarrow \quad 2x_3 + \mu + \lambda = 0 \,. \end{aligned} \tag{$*$}$$

Außerdem muss gelten $\mu(2x_1 - x_2 + x_3 - 5) = 0$, $\mu \geq 0$.

Nehmen wir an, $\mu > 0$. *Dann muss gelten* $2x_1 - x_2 + x_3 = 5$. *Aus* $(*)$ *folgt*

$$\begin{aligned} x_1 &= -\mu - \frac{\lambda}{2} \\ x_2 &= +\frac{\mu}{2} - \frac{\lambda}{2} \\ x_3 &= -\frac{\mu}{2} - \frac{\lambda}{2} \,. \end{aligned} \tag{$**$}$$

Einsetzen in die aktiven Nebenbedingungen

$$5 = 2\left(-\mu - \frac{\lambda}{2}\right) - \left(\frac{\mu}{2} - \frac{\lambda}{2}\right) + \left(-\frac{\mu}{2} - \frac{\lambda}{2}\right) = -3\mu - \lambda$$

$$3 = \left(-\mu - \frac{\lambda}{2}\right) + \left(\frac{\mu}{2} - \frac{\lambda}{2}\right) + \left(-\frac{\mu}{2} - \frac{\lambda}{2}\right) = -\mu - \frac{3}{2}\lambda$$

$$\left.\begin{aligned} -\frac{15}{2} &= \frac{9}{2}\mu + \frac{3}{2}\lambda \\ 3 &= -\mu - \frac{3}{2}\lambda \end{aligned}\right\} \quad \Rightarrow -\frac{9}{2} = \frac{7}{2}\mu$$

$$\mu = -\frac{9}{7} \quad \textit{Widerspruch zu} \quad \mu > 0\,.$$

Also nehmen wir $\mu = 0$ an. *Dann müssen wegen (**) alle x_i gleich sein und wegen $x_1 + x_2 + x_3 = 3$ folgt $x_1 = x_2 = x_3 = 1$, $\lambda = -2$.*

$\Rightarrow\ x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ *erfüllt die Optimalitätsbedingungen "kritischer Punkt". Ist x Lösung? Ja, denn*

- $f(x) \to \infty$, $|x| \to \infty$.
- *Daher können wir das Minimum in einer beschränkten Menge suchen* $\left(\left\{ x \in \mathcal{F} |\ \ |x|^2 \leq f\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}\right)$.
- *Nach dem Satz von Weierstraß existiert das Minimum unserer Aufgabe.*
- *Dieses* **muss** *die notwendigen Bedingungen erfüllen.*

$\Rightarrow\ x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ *ist die Lösung.*

Bemerkung: Alles unter der Annahme, dass unsere Multiplikatorenregel hier wirklich gelten muss!

Wir setzen nun unsere theoretischen Untersuchungen fort. Dazu führen wir ein:

Definition 7.2.3 *(Tangentenkegel). Es sei $S \subset \mathbb{R}^n$ eine beliebige Menge (z.B. $\mathcal{F}$). Vektor $d \in \mathbb{R}^n$ heißt* `Tangentialrichtung` *an S in x, $x \in S$, wenn es eine Folge $\{x^k\} \subset S$ mit $x^k \to x$, $k \to \infty$ und $\{t^k\} \subset \mathbb{R}$, $t^k > 0$ mit $t^k \downarrow 0$ gibt, so dass*

$$\lim_{k\to\infty} \frac{x^k - x}{t_k} = d\,.$$

Die Menge aller Tangentialrichtungen an S in x heißt `Tangentialkegel` *$T(S, x)$.*

Es gilt immer $0 \in T(S,x)$. Außerdem ist $T(S,x)$ ein Kegel. Man kann zeigen (vgl. [1, Lemma 7.2.10,11,13]):

- $x \in S \Rightarrow T(S,x)$ ist abgeschlossen
- $x \in S \Rightarrow T(S,x) \subset clK(S,x)$
- $x \in S$, S konvex $\Rightarrow T(S,x) = clK(S,x)$.

- Im Falle linearer Gleichungen und Ungleichungen ist $K(\mathcal{F},x)$ abgeschlossen und konvex, d. h. es gilt

$$T(\mathcal{F},x) = K(\mathcal{F},x) = L(\mathcal{F},x).$$

Im Fall der Nichtlinearität gilt das nicht, aber es gilt immer bei Differenzierbarkeit ohne weitere Voraussetzungen:

Lemma 7.2.2 *Es sei $\mathcal{F}$ die zulässige Menge von (PNU), g und h an der Stelle x differenzierbar. Dann gilt*

$$\boxed{T(\mathcal{F},x) \subset L(\mathcal{F},x)}.$$

Beweis: Siehe [1, Lemma 7.2.15]. Man zeigt: Gilt $d \in T(\mathcal{F},x)$ dann

$$h'(x)d = 0 \quad \text{sowie} \quad \nabla g^i(x)d \le 0 \quad \forall i \in J(x)\,.$$

Das heißt $d \in L(\mathcal{F},x)$. □

Beispiele

- $\mathcal{F} = \{x \in \mathbb{R}^2 \,|\, g(x) = x_1^2 + x_2^2 - 1 \le 0\}$

a) $x = 0 \quad \Rightarrow g(x) < 0$ inaktiv
$\Rightarrow L(\mathcal{F},0) = \mathbb{R}^2$. Außerdem auch $T(\mathcal{F},0) = \mathbb{R}^2$, da $x = 0$ ein innerer Punkt von $\mathcal{F}$ ist (setze $x^k = x + t^k d$; dann $x^k \in \mathcal{F}$ für kleine $t^k \Rightarrow \frac{x^k - x}{t^k} = d \;\forall k > k_0$).
$\Rightarrow$ bei $x = 0$ gilt $L(\mathcal{F},0) = T(\mathcal{F},0)$.

b) $x = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

$$\begin{aligned} L(\mathcal{F},x) &= \{d \,|\, g'(x)d \le 0\} \\ &= \left\{d \,|\, 2\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \le 0\right\} \\ &= \{d \,|\, -2d_2 \le 0\} \\ &\Rightarrow d_2 \ge 0\,. \end{aligned}$$

Analog findet man

$$\begin{aligned} T(\mathcal{F},x) &= \{d \,|\, d_2 \ge 0\} \\ \text{auch hier:}\quad T(\mathcal{F},x) &= L(\mathcal{F},x)\,. \end{aligned}$$

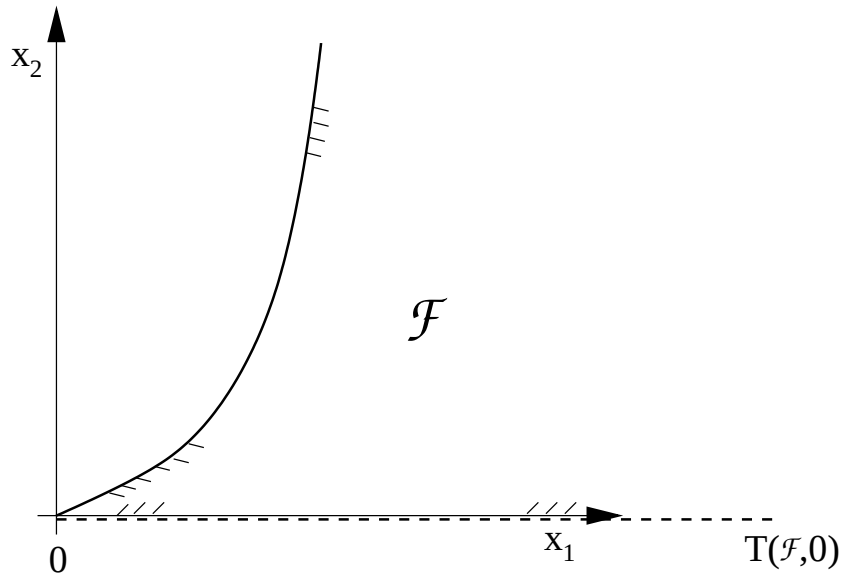
- $\mathcal{F} = \{x \in \mathbb{R}^2 \,|\, -x_1^3 + x_2 \le 0\,,\ -x_2 \le 0\}$ (unser altes Beispiel).

Im Nullpunkt gilt mit $g_1 = -x_1^3 + x_2$, $g_2 = -x_2$,

$$\begin{aligned} L(\mathcal{F}, 0) &= \{d \,|\, -3 \cdot 0^2 d_1 + d_2 \leq 0\,,\ -d_2 \leq 0\} \\ &= \{d \,|\, d_2 = 0\}\,. \end{aligned}$$

Für den Tangentenkegel ergit sich (siehe geometrische Illustration)

$$T(\mathcal{F}, 0) = \{d \,|\, d_1 \geq 0\,,\ d_2 = 0\}$$



Der negative Teil von $\mathbb{R}$ kommt bei der Bildung von $T(\mathcal{F}, 0)$ nicht zum Zuge!

Hier gilt wirklich nur die Inklusion

$$T(\mathcal{F}, 0) \subset L(\mathcal{F}, 0)\,.$$

Der Vorteil des Tangentialkegels ist, dass man mit ihm einen vernünftigen Ersatz unserer nur bei Konvexität gültigen Variationsungleichung bekommt. Es gilt nämlich der

Satz 7.2.2 *Sei $S \subset \mathbb{R}^n$ nichtleer, $f : S \to \mathbb{R}$, $\tilde{x} \in S$ ein lokales Minimum von f und f in $\tilde{x}$ differenzierbar. Dann gilt*

$$\langle \nabla f(\tilde{x}), d \rangle \geq 0 \quad \forall d \in T(S, \tilde{x})\,.$$

Beweis: Wir wählen $d \in T(S, \tilde{x})$ beliebig aus. Dann gilt

$$d = \lim_{k \to \infty} \frac{x^k - \tilde{x}}{t^k}$$

mit $x^k \in S$, $x^k \to \tilde{x}$, $t^k \to 0$. Also mit $r^k \to 0$

$$\begin{aligned} t^k(d + r^k) &= x^k - \tilde{x} \\ x^k &= \tilde{x} + t^k(d + r^k)\,. \end{aligned}$$

Da $\tilde{x}$ lokales Minimum und $x^k \to \tilde{x}$, folgt für hinreichend großes k

$$0 \leq f(x^k) - f(\tilde{x}) = \langle \nabla f(\tilde{x}), t^k(d + r^k) \rangle + o(t^k(d + r^k)) \mid : t^k$$

$$\Rightarrow \; 0 \leq \langle \nabla f(\tilde{x}), d + r^k \rangle + \underbrace{\frac{o(t^k(d + r^k))}{t_k}}$$

$$= \frac{o(t^k(d + r^k))}{|t^k(d + r^k)|} \cdot |d + r^k|$$

für $k \to \infty$ strebt r^k gegen Null und auch $t^k(d + r^k)$ gegen Null. Insgesamt

$$0 \leq \langle \nabla f(\tilde{x}), d \rangle \,.$$

□

Wir wissen nun also

$$\langle \nabla f, d \rangle \geq 0 \quad \forall d \in T(\mathcal{F}, \tilde{x})$$

und hätten gern

$$\langle \nabla f, d \rangle \geq 0 \quad \forall d \in L(\mathcal{F}, \tilde{x}) \,,$$

denn Letzteres ergibt – wie wir gesehen haben – die Lagrangesche Multiplikatorenregel. Dazu bräuchten wir aber, dass $\tilde{x}$ die linearisierte Aufgabe löst...

Was liegt näher als die Forderung

$$\boxed{T(\mathcal{F}, \tilde{x}) = L(\mathcal{F}, \tilde{x})} \quad ?$$

Definition 7.2.4 *$\tilde{x} \in \mathcal{F}$ heißt* `regulär`*, wenn $T(\mathcal{F}, \tilde{x}) = L(\mathcal{F}, \tilde{x})$ gilt.*

Folgerung aus dem Bisherigen:

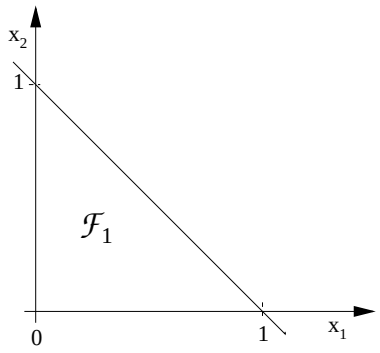
Satz 7.2.3 *Ist $\tilde{x} \in \mathcal{F}$ regulär und lokales Minimum für (PNU), dann existieren Lagrangesche Multiplikatoren λ und $\mu \geq 0$. Sind zusätzlich die Gradienten der aktiven Restriktionen, d. h.*

$$\nabla h_i(\tilde{x}) \,,\; i = 1, ..., m \quad \textit{sowie} \quad \nabla g_j(\tilde{x}) \,,\; j \in J(\tilde{x})$$

linear unabhängig, dann sind λ und μ eindeutig bestimmt.

Bevor wir die Regularitätsbedingungen weiter diskutieren, bemerken wir noch: $T(\mathcal{F}, x)$ hängt `nicht` von der konkreten Darstellung von $\mathcal{F}$ ab, während $L(\mathcal{F}, x)$ davon abhängen kann!

Beispiel 7.2.3



- $\mathcal{F}_1 = \{x | g_1(x) = x_1 + x_2 - 1 \leq 0$
  $g_2(x) = -x_1 \leq 0$
  $g_3(x) = -x_2 \leq 0\ \}$

- *Andere Darstellung von* $\mathcal{F}_1$*:*

$$\mathcal{F}_2 = \{x | (x_1 + x_2 - 1)^3 \leq 0$$
$$-x_1 \leq 0$$
$$-x_2 \leq 0\ \}\,. \qquad \textit{Offenbar} \quad \mathcal{F}_1 = \mathcal{F}_2\,.$$

Wir betrachten den Punkt

$$\tilde{x} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}.$$

Hier gilt: $T(\mathcal{F}_1, \tilde{x}) = T(\mathcal{F}, \tilde{x}) = \{d \,|\, d_1 + d_2 \leq 0\}$
(*geometrisch klar*)

$$L(\mathcal{F}_1, \tilde{x}) = \{d \,|\, d_1 + d_2 \leq 0\} = T(\mathcal{F}, \tilde{x})$$
(g_1 *war affin linear,* g_2, g_3 *inaktiv*)*.*

$$L(\mathcal{F}_2, \tilde{x}) = \left\{ d \,|\, \langle \nabla g_1(\tilde{x}), d\rangle = \Big\langle 3 \underbrace{(\tilde{x}_1 + \tilde{x}_2 - 1)^2}_{=0} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, d \Big\rangle = \langle 0, d\rangle \leq 0 \right\}$$

$$\Rightarrow \quad L(\mathcal{F}_2, \tilde{x}) = \mathbb{R}^2 \supset T(\mathcal{F}_2, \tilde{x})\,.$$

Man kann zeigen, dass für entsprechende Optimierungsaufgaben im zweiten Fall keine Lagrangeschen Multiplikatoren existieren müssen (vgl. [1]).

Wir widmen uns nun dem Problem der Regularität. Wann gilt

$$T(\mathcal{F}, \tilde{x}) = L(\mathcal{F}, \tilde{x})?$$

Fall reiner Ungleichungsrestriktionen

Wir betrachten

$$\mathcal{F} = \{x \in \mathbb{R}^n \,|\, g(x) \leq 0\}\,,$$

$g : \mathbb{R}^n \to \mathbb{R}^p$ differenzierbar. Ein $\tilde{x} \in \mathcal{F}$ erfülle folgende Regularitätsbedingung:

$$\boxed{\exists \bar{d} \in \mathbb{R}^n : \langle \nabla g_j(\tilde{x}), \bar{d}\rangle < 0 \quad \forall j \in J(\tilde{x})\,.}$$

Dann ist $\tilde{x}$ regulär, d. h. $T(\mathcal{F}, \tilde{x}) = L(\mathcal{F}, \tilde{x})$.

Beweis: Da stets $T \subset L$ gilt, müssen wir die andere Inklusion zeigen. Dazu sei $d \in L(\mathcal{F}, \tilde{x})$ beliebig gegeben. Wir setzen

$$x(t) = \tilde{x} + t(d + \alpha\bar{d})$$

mit $\alpha > 0$. Dann ist $x(t) \in \mathcal{F}$ *für hinreichend kleine* t*:* Für die inaktiven Restriktionen $j \notin J(\tilde{x})$ ist das klar, denn $g_j(\tilde{x}) + t(d + \alpha\bar{d})) \to g_j\tilde{x} < 0$, $t \downarrow 0$. Bei den aktiven gilt

$$g_j(x(t)) = \underbrace{g_j(\tilde{x})}_{=0} + \langle \nabla g_j(\tilde{x}), t(d + \alpha\bar{d})\rangle + o(t)$$

$$= t \underbrace{\langle \nabla g_j(\tilde{x}), d\rangle}_{\leq 0, \text{ da } d \in L(\mathcal{F},\tilde{x})} + t \left\{ \alpha \underbrace{\langle \nabla g_j(\tilde{x}), \bar{d}\rangle}_{<0} + \underbrace{\frac{o(t)}{t}}_{\to 0} \right\}$$

Für hinreichend kleines t wird $\{...\}$ negativ, daher (es gibt nur endlich viele $j \in J(\tilde{x})$)

$$g_j(x(t)) \leq 0 \quad \forall t \in [0, \bar{t}] \quad \forall j \in J(\tilde{x})\,.$$

Nun ist klar, was passieren muss: Wir setzen

$$t_k = \frac{1}{k}\,,\ x^k = x(t_k)\,.$$

Dann gilt $x^k \in \mathcal{F}$ $\forall k > k_0$ und nach Konstruktion

$$\frac{x^k - \tilde{x}}{t_k} = d + \alpha\bar{d}\,.$$

$\Rightarrow\ d + \alpha\bar{d} \in T(\mathcal{F}, \tilde{x})$ $\forall \alpha > 0$. $T(\mathcal{F}, \tilde{x})$ ist abgeschlossen. Damit gilt auch $d = \lim\limits_{\alpha \downarrow 0} d + \alpha\bar{d} \in T(\mathcal{F}, \tilde{x})$. □

Man kann zeigen, dass für die affin-linearen Restriktionen sogar $\langle \nabla g_j, \bar{d}\rangle \leq 0$ ausreicht. Insgesamt folgt dann

Satz 7.2.4 *Sei $\tilde{x} \in \mathcal{F}$ und g differenzierbar in $\tilde{x}$. Gibt es ein $\bar{d} \in \mathbb{R}^n$ mit*

$$\left.\begin{array}{l} \langle \nabla g_j(\tilde{x}), \bar{d}\rangle \leq 0 \quad \textit{für alle affin-linearen aktiven Restriktionen} \\ \langle \nabla g_j(\tilde{x}), \bar{d}\rangle < 0 \quad \textit{für alle anderen aktiven Restriktionen,} \end{array}\right\} \tag{7.6}$$

dann ist $\tilde{x}$ regulär.

Hinreichend für (7.6):

$\boxed{\exists \bar{d} \in \mathbb{R}^n : g(\tilde{x}) + g'(\tilde{x})\bar{d} < 0}$ *Lokale (d. h. von $\tilde{x}$ abhängige Slater-Bedingung.*

Im Falle konvexer Funktionen g_j ist hinreichend für (7.6)

$\boxed{\exists \bar{v} \in \mathcal{F} : g_j(\bar{v}) < 0 \quad \forall j \in J(\tilde{x}),\ \ \textit{nichtlinear gj}}$ *(Slater-Bedingung)*

denn für $\bar{v} = \tilde{x}$ gilt (7.6) mit $d = 0$; sonst: Wegen Konvexität

$$g_j(\bar{v}) - g_j(\tilde{x}) \geq \langle \nabla g_j(\tilde{x}), \underbrace{\bar{v} - \tilde{x}}_{d}\rangle$$

$$\textit{also} \quad \langle \nabla g_j(\tilde{x}), d\rangle \leq \underbrace{g_j(\bar{v})}_{<0} - \underbrace{g_j(\tilde{x})}_{=0\ \forall j \in J(\tilde{x})} < 0 \quad \forall j \in J(\tilde{x})$$

.

Korollar 7.1 *Sind die Gradienten der aktiven Restriktionen, $\nabla g_j(\tilde{x})$, $j \in J(\tilde{x})$, linear unabhängig, dann ist $\tilde{x}$ regulär.*

Beweis: Das System $\langle \nabla g_j(\tilde{x}), d \rangle = b_j$, $j \in J(\tilde{x})$ hat für alle b_j eine Lösung. Insbesondere für $b_j < 0$... . □

Regularität bei Gleichungsrestriktionen

Nun sei $\mathcal{F} = \{x \in \mathbb{R}^n \,|\, h(x) = 0\}$.

Satz 7.2.5 *Die Gradienten $\nabla h_i(\tilde{x})$, $i = 1, ..., m$, seien linear unabhängig. Dann ist $\tilde{x}$ regulär.*

Der Beweis ist eine Anwendung des Satzes über implizite Funktionen. Er ist in vielen Büchern zu finden, z.B. in [1, Satz 7.2.26], deshalb lassen wir ihn weg.

Andere Interpretation:

$$\boxed{h'(\tilde{x}) \quad \text{ist surjektiv, “Abbildung auf”}}$$

Regularität bei Gleichungs- und Ungleichungsrestriktionen

Jetzt ist

$$\mathcal{F} = \{x \in \mathbb{R}^n \,|\, h(x) = 0\,,\ g(x) \leq 0\}\,.$$

Satz 7.2.6 *Sei $\tilde{x} \in \mathcal{F}$, h und g stetig differenzierbar. Die Gradienten $\nabla h_i(\tilde{x})$ seien linear unabhängig, und es existiere ein $\bar{d} \in \mathbb{R}^n$ mit*

$$\boxed{h'(\tilde{x})\bar{d} = 0 \quad \textit{und} \quad \nabla g_j(\tilde{x})\bar{d} < 0 \quad \forall j \in J(\tilde{x})\,.} \tag{7.7}$$

Dann ist $\tilde{x}$ regulär.

Wir verzichten auf den Beweis. Diese Regularitätsbedingung nennt man `Mangasarian-Fromovitz-Bedingung`.

Bemerkung: Wie oben ist dafür wiederum hinreichend:

$$\nabla h_i(\tilde{x})\,,\ i = 1, ..., m\,,\ \nabla g_j(\tilde{x})\,,\ j \in J(\tilde{x}) \quad \text{sind linear unabhängig}\,. \tag{7.8}$$

Einige kleine Beispiele sollen die Anwendung von Regularitätsbedingungen illustrieren:

Beispiel 7.2.4

1. $\min f(x), |x|^2 \leq 1$.

   Hier gilt:
   $$\mathcal{F} = \{x \in \mathbb{R}^n \,|\, g(x) = |x|^2 - 1 \leq 0\}$$
   - *g ist konvex und differenzierbar*
   - *$\bar{v} = 0$ erfüllt $g(\bar{v}) = -1 < 0$*
     $\Rightarrow$ Slater-Bedingung ist erfüllt
     $\Rightarrow$ jedes $x \in \mathcal{F}$ ist regulär.

2.

$$\min f(x)\,,$$

bei

$$e^{\sum_1^n x_i} = 3$$
$$|x|^2 = 4$$

Problem mit Gleichungsrestriktionen,

$$h_1(x) = e^{\sum x_i} - 3\,,\ h_2(x) = |x|^2 - 4$$
$$\nabla h_1(x) = \left(e^{\sum x_i}\right) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\,,\ \nabla h_2(x) = 2x\,.$$

Diese Gradienten sind – unabhängig von x – stets linear unabhängig, wenn x zulässig ist, denn dann gilt

$$|x|^2 = 4\,, \quad \textit{also} \quad x_1^2 + ... + x_n^2 = 4\,.$$

x müsste die Darstellung $\alpha \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ haben bei linearer Abhängigkeit, also

$$\alpha^2(1 + ... + 1) = 4\,,\ \alpha^2 = \frac{4}{n}\ \alpha = \frac{2}{\sqrt{n}}$$
$$\Rightarrow\ x = \frac{2}{\sqrt{n}} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \Rightarrow e^{\sum_1^n x_i} = e^{\frac{2n}{\sqrt{n}}} = e^{2\sqrt{n}} = 3$$
$$\Rightarrow 2\sqrt{n} = ln3$$
$$\underbrace{n}_{\in\mathbb{N}} = \underbrace{\left(\frac{1}{2}\,ln3\right)^2}_{\textit{irrational}} \quad \textit{Widerspruch}\,.$$

Also ist jedes zulässige x regulär.

Wir halten fest: Ist $\tilde{x}$ eine reguläre Lösung von (PNU), dann gilt unter unseren Voraussetzungen eine Lagrangesche Multiplikatorenregel (*Karush-Kuhn-Tucker-Satz*).

Man kann einen sehr ähnlichen, aber nicht so gut anwendbaren Satz *ohne* Regularität beweisen:

Satz 7.2.7 *(Satz von Fritz John). Die Funktionen f, g, h seien stetig differenzierbar und $\tilde{x}$ eine lokale Lösung von* (PNU)*. Dann existieren Multiplikatoren $\mu_0 \geq 0$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, so dass*

- $(\mu_0, \lambda^T, \mu^T) \neq 0$
- $\mu_0 \nabla f(\tilde{x}) + h'(\tilde{x})^T \lambda + g'(\tilde{x})^T \mu = 0$
- $\mu \geq 0\,,\ \langle \mu, g(\tilde{x}) \rangle = 0.$

Bemerkung: *Wüsste man $\mu_0 \neq 0$, dann könnte man durch μ_0 teilen und hätte mit $\tilde{\lambda} := \frac{1}{\mu_0}\lambda$, $\tilde{\mu} = \frac{1}{\mu_0}\mu$ richtige Lagrangesche Multiplikatoren. Leider aber weiß man das nicht! Im Falle $\mu_0 = 0$ fehlt dann die eigentlich zu minimierende Zielfunktion f in den Optimierungsedingungen und der Aussagewert ist gering.*

Oft baut man die Theorie so auf: Man beweist zuerst den Fritz-John-Satz (mit Trennungssätzen) und zeigt dann: $\tilde{x}$ regulär $\Rightarrow \mu_0 \neq 0$.

Beispiel 7.2.5 *(KKT-Satz gilt nicht, aber F-J-Satz).*

$$\begin{aligned} \min f(x_1, x_2) &= x_1 + x_2^2 \\ h_1(x) &= -x_1^2 + x_2 = 0 \\ h_2(x) &= x_1^2 + x_2 = 0 \end{aligned}$$

Hier: $\mathcal{F} = \{0\}$. $\Rightarrow$ $\tilde{x} = 0$ *ist Lösung.*
Würde eine Multiplikatorenregel vom KKT-Typ gelten, so

$$\nabla f(0) + \nabla h_1(0)\lambda_1 + \nabla h_2(0)\lambda_2 = 0\,,$$

also

$$\begin{aligned} \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}\lambda_1 + \begin{pmatrix}0\\1\end{pmatrix}\lambda_2 &= 0\,, \\ \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}(\lambda_1 + \lambda_2) &= 0\,, \end{aligned}$$

was wegen linearer Unabhängigkeit unmöglich ist. Eine Fritz-John-Aussage gilt natürlich, denn

$$\mu_0\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}(\lambda_1 + \lambda_2) = 0$$

ist mit $\mu_0 = 0$, $\lambda_1 = -\lambda_2$ zu erfüllen.

Beispiel 7.2.6 $x_1^2 + x_2^2 \to$ *extr.* *(d. h. min oder max)*
bei

$$x_1^4 + x_2^4 = 1$$

Erweiterte Lagrange-Funktion:

$$\begin{aligned} \mathcal{L} &= \mu_0(x_1^2 + x_2^2) + \lambda(x_1^4 + x_2^4 - 1) \\ \frac{\partial \mathcal{L}}{\partial x_1} = 0 \quad &\Leftrightarrow \quad 2\mu_0 x_1 + 4\lambda x_1^3 = 0 \\ \frac{\partial \mathcal{L}}{\partial x_2} = 0 \quad &\Leftrightarrow \quad 2\mu_0 x_2 + 4\lambda x_2^3 = 0 \end{aligned}$$

Gilt $\mu_0 = 0$, so $\lambda \neq 0$, und es muss gelten $x_1^3 = 0$ sowie $x_2^3 = 0 \Rightarrow$ keine Zulässigkeit. Daher muss $\mu_0 \neq 0$ sein. Wir können o.B.d.A. $\mu_0 = \frac{1}{2}$ annehmen.

$$\begin{aligned} \Rightarrow \quad & x_1(1 + 4\lambda x_1^2) = 0 \\ & x_2(1 + 4\lambda x_2^2) = 0 \end{aligned}$$

$$\begin{aligned} \Rightarrow \quad \textit{Möglichkeiten} \quad & x_1 = 0 \quad \Rightarrow \quad x_2 = \pm 1 \qquad (\textit{aus Nebenbedingung}) \\ & x_2 = 0 \quad \Rightarrow \quad x_1 = \pm 1 \\ & 1 + 4\lambda x_1^2 = 1 + 4\lambda x_2^2 = 0 \end{aligned}$$

Die letzte Beziehung ergibt (mit passendem λ)

$$x_1^2 = x_2^2 = -\frac{1}{4\lambda} \quad \Rightarrow \quad |x_1| = |x_2| \,.$$

Eingesetzt in $x_1^4 + x_2^4 = 1$ folgt $2|x|^4 = 1$, also

$$|x_1| = |x_2| = 2^{-1/4} \,.$$

Was sind all diese Punkte wert? Nach dem Satz von Weierstraß existieren min/max. Diese `müssen` *die notwendigen Bedingungen erfüllen. Andere Lösungen gibt es nicht. Bei den beiden $0, \pm 1$-Varianten gilt*

$$f(x) = 1 \,.$$

Für $|x_i| = 2^{-1/4}$ gilt $f(x) = 2 \cdot \dfrac{1}{\sqrt{2}} = \sqrt{2} > 1$.

$$\begin{aligned} \Rightarrow \quad & \textit{Maximum bei} \quad |x_1| = |x_2| = 2^{-1/4} \quad \textit{(4 Punkte)} \\ & \textit{Minimum bei} \quad \tbinom{0}{\pm 1} \ \textit{und} \ \tbinom{\pm 1}{0} \quad \textit{(4 Punkte)} \end{aligned}$$

7.3 Optimalitätsbedingungen zweiter Ordnung

Analog zu den Aufgaben mit oder ohne lineare Restriktionen kann man nun wieder mit hinreichenden Bedingungen 2. Ordnung überprüfen, ob wirklich ein lokales Minimum vorliegt. Dazu braucht man die zweite Ableitung der Lagrange-Funktion,

$$\mathcal{L}_{xx}(x, \lambda, \mu) = f''(x) + \sum_{i=1}^{m} \lambda_i h_i''(x) + \sum_{j=1}^{m} \mu_j g_j''(x) \,.$$

Es gilt

Satz 7.3.1 *Es gelte $f, h, g \in C^2$, $\tilde{x}$ sei regulär, und $\lambda, \mu \geq 0$ seien die entsprechenden Lagrangeschen Multiplikatoren. Es existiere ein $\alpha > 0$, so dass*

$$d^T \mathcal{L}_{xx}(\tilde{x}, \lambda, \mu) d \geq \|d\|^2 \tag{7.9}$$

gilt für alle $d \in L(\mathcal{F}, \tilde{x})$ mit der zusätzlichen Eigenschaft $\langle \nabla f(\tilde{x})^T, d \rangle = 0$. Dann existieren $\varrho, \beta > 0$, so dass die `quadratische Wachstumsbedingung`

$$f(x) \geq f(\tilde{x}) + \beta \|x - \tilde{x}\|^2$$

gilt für alle $x \in \mathcal{F} \cap B(\tilde{x}, \varrho)$. Damit ist $\tilde{x}$ striktes lokales Minimum von (PNU).

Diese Bedingung kann man etwas anders aufschreiben: $\tilde{x}$ erfüllt nach Voraussetzung die Kuhn-Tucker-Bedingungen mit Multiplikatoren λ und μ. Folglich

$$\begin{aligned} \nabla f(\tilde{x}) &= -h'(\tilde{x})^T\lambda - g'(\tilde{x})^T\mu\,, \\ \text{also} \quad \langle \nabla f(\tilde{x}), d\rangle = 0 \quad &\Leftrightarrow \quad \langle -h'(\tilde{x})^T\lambda, d\rangle - \langle g'(\tilde{x})^T\mu, d\rangle \qquad = 0 \end{aligned}$$

$\Updownarrow$

$$\langle \lambda, h'(\tilde{x})d\rangle + \langle g'(\tilde{x})d, \mu\rangle = 0\,. \qquad (*)$$

Ist $d \in L(\mathcal{F}, \tilde{x})$, so gilt automatisch $h'(\tilde{x})d = 0$. Außerdem gilt $\langle \nabla g_j(\tilde{x}), d\rangle \leq 0$ für alle $j \in J(\tilde{x})$ (aktive Restriktionen). Wegen $(*)$ muss noch gefordert werden

$$\langle \nabla g_j(\tilde{x}), d\rangle \mu_j = 0 \quad \forall j = 1, ..., p\,.$$

Für die inaktiven Restriktionen gilt das automatisch, weil $\mu_j = 0$. Für die aktiven mit $\mu_j = 0$ auch (also keine zusätzliche Bedingung). Also bleiben noch die aktiven mit $\mu_j > 0$. Hier muss also noch gelten $\langle \nabla g_j(\tilde{x}), d\rangle = 0$.

$\Rightarrow$ Äquivalente Form der hinreichenden Bedingung:

- $\tilde{x} \in \mathcal{F}$ erfüllt die Kuhn-Tucker-Bedingungen und
- (7.9) gilt für alle d mit

$$\begin{aligned} h'(\tilde{x})d &= 0 \\ \langle \nabla g_j(\tilde{x}), d\rangle &= 0 \quad \forall j \in J(\tilde{x}) \quad \text{mit} \quad \mu_j > 0 \quad \texttt{``streng aktive Restriktionen''} \\ \langle \nabla g_j(\tilde{x}), d\rangle &\leq 0 \quad \forall j \in J(\tilde{x}) \quad \text{mit} \quad \mu_j = 0\,. \end{aligned}$$

8 Probleme mit nichtlinearen Restriktionen-Verfahren

8.1 Das Lagrange-Newton-Verfahren

Wir betrachten wieder die Aufgabe

$$\begin{aligned} &\min f(x) \\ &h(x) = 0 \\ &g(x) \leq 0 \end{aligned} \tag{PNU}$$

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:

$$\boxed{\begin{aligned} \nabla f(\tilde{x}) + h'(\tilde{x})^T\tilde{\lambda} + g'(\tilde{x})^T\tilde{\mu} &= 0 \\ h(\tilde{x}) &= 0 \\ \tilde{g}(\tilde{x}) &= 0\,. \end{aligned}}$$

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 interssieren 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

$$z = \begin{pmatrix} x \\ \lambda \\ \nu \end{pmatrix}\,, \quad F(z) = \begin{pmatrix} \nabla f(x) + h'(x)\lambda + \tilde{g}'(x)\nu \\ h(x) \\ \tilde{g}(x) \end{pmatrix}$$

mit $\nu = (\mu_j)_{j\in J(\tilde{x})}$. $\tilde{\nu} := (\tilde{\mu}_j)_{j\in J(\tilde{x})}$. Dann gilt also mit $\tilde{z} = (\tilde{x}^T, \tilde{\lambda}^T, \tilde{\nu}^T)$

$$F(\tilde{z}) = 0\,.$$

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

$$F'(z) = \begin{pmatrix} \mathcal{L}_{xx}(x,\lambda,\nu) & h'(x)^T & \tilde{g}'(x)^T \\ h'(x) & 0 & 0 \\ \tilde{g}'(x) & 0 & 0 \end{pmatrix}\,.$$

Dies ist eine Matrix vom Typ

$$\mathcal{A} = \begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix},$$

für welche in Lemma 5.4.1 gezeigt wurde: Ist Q positiv definit auf $\ker A$ und hat A vollen Rang, dann ist $\mathcal{A}$ invertierbar. Hier gilt

$$Q = \mathcal{L}_{xx}\,,\ A = \begin{pmatrix} h'(x) \\ g'(x) \end{pmatrix},$$

also ist $F'(\tilde{z})$ nichtsingulär, wenn

- (8.1.3) Die Gradienten $\nabla h_i(\tilde{x})$, $\nabla g_j(\tilde{x})$ der aktiven Restriktionen 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:

$$\begin{pmatrix} \tilde{\mathcal{L}}_{xx}(x^k, \lambda^k, \nu^k) & h'(x^k)^T & \tilde{g}'(x^k)^T \\ h'(x^k) & 0 & 0 \\ \tilde{g}'(x^k) & 0 & 0 \end{pmatrix} \begin{pmatrix} x - x^k \\ \lambda - \lambda^k \\ \nu - \nu^k \end{pmatrix}$$

$$= - \begin{pmatrix} \nabla f(x^k) + h'(x^k)^T \lambda^k + \tilde{g}'(x^k)^T \nu^k \\ h(x^k) \\ \tilde{g}(x^k) \end{pmatrix}.$$

Einiges hebt sich hier auf, so dass am Ende bleibt

$$\begin{aligned} \nabla f(x^k) + \tilde{\mathcal{L}}_{xx}(x^k, \lambda^k, \nu^k)(x - x^k) + h'(x^k)^T \lambda + \tilde{g}'(x^k)^T \nu &= 0 \\ h(x^k) + h'(x^k)(x - x^k) &= 0 \\ \tilde{g}(x^k) + \tilde{g}'(x^k)(x - x^k) &= 0\,. \end{aligned} \tag{8.10}$$

Im Prinzip sind das die notwendigen Optimalitätsbedingungen für eine Lösung der linear-quadratischen Aufgabe

$$\begin{aligned} &\min_x \nabla f(x^k)^T(x-x^k) + \frac{1}{2}(x-x^k)^T \mathcal{L}_{xx}(...)(x-x^k) && (\mathrm{Q1})_k \\ \text{bei} \quad & h(x^k) + h'(x^k)(x-x^k) = 0 \\ & g(x^k) + g'(x^k)(x-x^k) \leq 0\,, \end{aligned}$$

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 (8.10) und das SQP-Verfahren $(\mathrm{Q1})_k$ äquivalent.

Nun führen wir wieder die Richtung $d = x - x^k$ ein und erhalten so:

$$\begin{aligned} &\min_{d\in\mathbb{R}^n} \frac{1}{2}\langle d, \mathcal{L}_{xx}(x^k,\lambda^k,\mu^k)d\rangle + \langle \nabla f(x^k), d\rangle && (\mathrm{Q2})_k \\ &\text{bei} \quad h(x^k) + h'(x^k)d = 0 \\ &\qquad\quad g(x^k) + g'(x^k)d \leq 0 \end{aligned}$$

und als Lösung $d^k = x^{k+1} - x^k$ mit neuen Multiplikatoren μ^{k+1}, λ^{k+1}. So erhalten wir

Verfahren 8.1.1 `Lagrange-Newton-Verfahren für` (PNU)
1. $k := 0$, $z^0 = (x^0, \lambda^0, \mu^0)$
2. Löse $(\mathrm{Q2})_k \to d^k$; λ^{k+1}, μ^{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. $x^k \notin \mathcal{F}$. Außerdem haben wir nur lokale Konvergenz. Deshalb sind Modifikationen angebracht!

8.2 Sequentielle quadratische Optimierung

Wie erwähnt, ist das reine Langrange-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:

$$\begin{aligned} & \min_d \langle \nabla f(x^k), d\rangle + \frac{1}{2}\langle d, A^{(k)} d\rangle \qquad (\mathrm{QP})_k \\ \text{bei} \quad & h(x^k) + h'(x^k)d = 0 \\ & g(x^k) + g'(x^k)d \leq 0\,. \end{aligned}$$

Wir wollen annehmen, dass die entsprechende zulässige Menge $\mathcal{F}_k$ nicht leer ist. Eine hinreichende Bedingung gibt [1,Satz8.2.1] an (h_i affin-linear, lineare Unabhängigkeit der Gradienten, Slater-Typ-Bedingung).

Unter natürlichen Bedingungen kann dann die Existenz genau einer Lösung von $(\mathrm{QP})_k$ sowie die gleichmäßige Beschränktheit der Folgen d^k, α^k, μ^k bewiesen werden [1,Satz8.2.2].

• 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 theoretische 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 8.2.1 *ϕ heißt Merit-Funktion, wenn gilt:*

- *Ist $\tilde{x} \in \mathcal{F}$ 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

$$\boxed{\phi = \phi(x;\beta,\gamma) := f(x) + \sum_{j=1}^{p} \beta_j g_j(x)_+ + \sum_{i=1}^{m} \gamma_i |h_i(x)|}$$

mit Konstanten $\beta_j \geq 0$, $\gamma_i \geq 0$. ϕ ist nicht differenzierbar!

$$g_j(x)_+ := \max\{0, g_j(x)\} = \frac{g_j(x) + |g_j(x)|}{2}\,.$$

Wenn $x \in \mathcal{F}$, so gilt $g_j(x) \leq 0$, also $g_j(x)_+ = 0$ sowie $h_i(x) = 0$, also $|h_i(x)| = 0$. Damit

$$x \in \mathcal{F} \Rightarrow f(x) = \phi(x;\beta,\gamma)\,.$$

Für $x \notin \mathcal{F}$ 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:

$$\underbrace{\sum_{i=1}^{m} \gamma_i |h_i(x)| + \sum_{j=1}^{p} \beta_j g_j(x)_+}_{\substack{\textit{Penalty-Term} \\ \gamma_i, \beta_j \quad \textit{Penalty-Parameter}}} .$$

Die Funktion ϕ heißt (*exakte*) *Penalty-Funktion*. Es gilt der wichtige

Satz 8.2.1 *Unter den Bedingungen 8.1.2–8.1.5 gilt: Ist $\tilde{x}$ lokales Minimum von* (PNU) *und*

$$\beta_j > \mu_j \,,\ \gamma_i > |\lambda_i|$$

$j = 1, ..., p$, $i = 1, ..., m$, 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

$$\beta_j \geq \mu_j^{(k+1)} + \varepsilon \,,\ \gamma_i \geq |\lambda_i^{(k+1)}| + \varepsilon \,,$$

und für hinreichend kleines $\sigma > 0$ erhält man

$$\phi(x^k + \sigma d^k, \beta, \gamma) \leq \phi(x^k; \beta, \gamma) - \sigma\delta \left[\langle d^k, A^{(k)} d^k \rangle + \varepsilon \|g(x^k)_+\|_1 + \varepsilon \|h(x^k)\|_1 \right] ,$$

d. h., die Merit-Funktion kann wirklich verkleinert werden. Außerdem gilt

$$\begin{aligned} |h_i(x^k + \sigma d^k)| &\leq (1 - \delta\sigma)|h_i(x^k)| \quad i = 1, ..., m \\ g_j(x^k + \sigma d^k) &\leq (1 - \delta\sigma) g_j(x^k) \qquad j = 1, ..., p \,, \end{aligned}$$

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].