REDUKTIONSMETHODEN, STRATEGIE DER AKTIVEN MENGE, SQP-VERFAHREN

STEFAN VIGERSKE

Literatur: Hans Joachim Oberle, Gerhard Opfer, Optimierung mit MATLAB, Hamburger Beiträge zur Angewandten Mathematik, Reihe B, Bericht 33, August 1999

1. Reduktionsmethoden bei linearen Gleichungsrestriktionen

Wir behandeln ein Problem der Form

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax = b \end{array}$$

mit $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $f \in C^1(\mathbb{R}^n)$. Wir reduzieren dieses Problem auf ein freies Optimierungsproblem, indem wir die Lösungsmenge des Gleichungssystems $Ax = b$ in die Zielfunktion "einsetzen".

Wir nehmen an, daß $m < n$ gilt und A vollen Rang m hat. Bei den Gleichungsrestriktionen handelt es sich also um unterbestimmtes System. In diesem Fall ist der zulässige Bereich

$$M := \{x \in \mathbb{R}^n \,|Ax = b\}$$

darstellbar in der Form

$$M = \hat{x} + \ker A,$$

wobei $\hat{x}$ eine spezielle Lösung von $Ax = b$ ist.

Sei $Z \in \mathbb{R}^{n \times (n-m)}$ eine Matrix, die in ihren Spalten eine Basis von $\ker A$ enthält. Dann hat M bei beliebigen $\hat{x} \in M$ die Form

$$M = \left\{\hat{x} + Zu \,\middle|\, u \in \mathbb{R}^{n-m}\right\}$$

und wir können unser Optimierungsproblem als unrestringierte Optimierungsaufgabe schreiben:

$$\min F(u) := f(\hat{x} + Zu), \quad u \in \mathbb{R}^{n-m}.$$

Zu klären bleibt, wie $\hat{x}$ und Z numerisch möglichst zweckmäßig berechnet werden können:

Einen allgemein Zugang liefert dabei die folgende Überlegung: Ist $V \in \mathbb{R}^{(n-m) \times n}$ eine Matrix, für welche $\begin{pmatrix} A \\ V \end{pmatrix} \in \mathbb{R}^{n \times n}$ invertierbar ist, so kann

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} := \begin{pmatrix} A \\ V \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix}$$

als Variablentransformation im $\mathbb{R}^n$ aufgefaßt werden. Dabei wird M durch $y_1 = 0$ beschrieben, während die Komponenten von y_2 als freie Variablen aufgefaßt werden. Dies liefert dann

$$x \in M \quad \Leftrightarrow \quad x = \begin{pmatrix} A \\ V \end{pmatrix}^{-1} \begin{pmatrix} b \\ y_2 \end{pmatrix}$$

Stellt man $\begin{pmatrix} A \\ V \end{pmatrix}^{-1}$ in der Form $\begin{pmatrix} Y & Z \end{pmatrix}$ mit $Y \in \mathbb{R}^{n \times m}$, $Z \in \mathbb{R}^{n \times (n-m)}$ dar, so erhält man mit $u := y_2$ die Gleichung

$$x = Yb + Zu.$$

In der Eliminationsmethode benutzt man die Eigenschaft, daß es eine quadratische, nichtsinguläre $m \times m$ Teilmatrix A_1 von A gibt, d.h.

$$A = \begin{pmatrix} A_1 & A_2 \end{pmatrix} \qquad A_1 \in \mathbb{R}^{m\times m} \text{ invertierbar}, \quad A_2 \in \mathbb{R}^{m\times(n-m)}$$

Wählt man $V = \begin{pmatrix} 0 & I \end{pmatrix}$, so ist $\begin{pmatrix} A \\ V \end{pmatrix}$ regulär und man erhält

$$\begin{pmatrix} Y & Z \end{pmatrix} = \begin{pmatrix} A \\ V \end{pmatrix}^{-1} = \begin{pmatrix} A_1 & A_2 \\ 0 & I \end{pmatrix}^{-1} = \begin{pmatrix} A_1^{-1} & -A_1^{-1}A_2 \\ 0 & I \end{pmatrix}$$

Numerisch stabiler ist die Orthogonalzerlegungsmethode: Man ermittle eine QR-Zerlegung

$$A^T = Q\begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix}\begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1 R$$

mit einer orthogonalen Matrix $Q \in \mathbb{R}^{n\times n}$, einer oberen Dreiecksmatrix $R \in \mathbb{R}^{m\times m}$ sowie $Q_1 \in \mathbb{R}^{n\times m}$, $Q_2 \in \mathbb{R}^{n\times(n-m)}$. Mit der Wahl $V = Q_2^T$ ist

$$\begin{aligned} \begin{pmatrix} A \\ V \end{pmatrix} &= \begin{pmatrix} A \\ Q_2^T \end{pmatrix} = \begin{pmatrix} (Q_1R)^T \\ Q_2^T \end{pmatrix} \\ \text{und } \begin{pmatrix} Y & Z \end{pmatrix} &= \begin{pmatrix} Q_1\left(R^{-1}\right)^T & Q_2 \end{pmatrix}, \end{aligned}$$

denn

$$\begin{pmatrix} R^TQ_1^T \\ Q_2^T \end{pmatrix} \cdot \begin{pmatrix} Q_1\left(R^{-1}\right)^T & Q_2 \end{pmatrix} = \begin{pmatrix} R^TQ_1^TQ_1\left(R^{-1}\right)^T & R^TQ_1^TQ_2 \\ Q_2^TQ_1\left(R^{-1}\right)^T & Q_2^TQ_2 \end{pmatrix} = \begin{pmatrix} I_m & 0 \\ 0 & I_{n-m} \end{pmatrix}.$$

Das auf die Variablen $u_1, \ldots, u_{n-m}$ reduzierte Problem kann nun mit Methoden für freie Optimierungsprobleme gelöst werden. Für die Ableitungen von F gilt

$$\begin{aligned} \nabla F(u) &= Z^T \nabla f(\hat{x} + Zu) \\ \nabla^2 F(u) &= Z^T \nabla^2 f(\hat{x} + Zu) Z \end{aligned}$$

Aus Stabilitätsgründen ist es empfehlenswert, im Verlaufe der angewandten Iterationsverfahren den Punkt $\hat{x}$ durch x^k zu ersetzen.

Will man das Newton-Verfahren verwenden, so gewinnt man eine Suchrichtung u durch Minimierung der quadratischen Approximation

$$\begin{aligned} F(u) &\approx F(0) + u^T\nabla F(0) + \frac{1}{2}u^T\nabla^2 F(0)\, u \\ &= f(\hat{x}) + u^TZ^T\nabla f(\hat{x}) + \frac{1}{2}u^TZ^T\nabla^2 f(\hat{x})\, Zu \\ \rightsquigarrow 0 \overset{!}{=} \nabla F(u) &\approx Z^T\nabla f(\hat{x}) + Z^T\nabla^2 f(\hat{x})\, Zu \\ \rightsquigarrow Z^T\nabla^2 f(\hat{x})\, Zu &= -Z^T\nabla f(\hat{x}) \end{aligned}$$

Manchmal kann es nötig sein, Lagrange-Parameter $\lambda_1, \ldots, \lambda_m$ aus dem überbestimmten Gleichungssystem

$$A^T\lambda = \nabla f(x)$$

zu berechnen. Ist x lokales Minimum, dann existiert eine Lösung dieser Gleichung. Multiplizieren wir diese Lösung dann mit Y, so erhalten wir

$$\begin{aligned} Y^TA^T\lambda &= Y^T\nabla f(x) \\ \rightsquigarrow \lambda &= Y^T\nabla f(x) \end{aligned}$$

Ist $A^T\lambda = \nabla f(x)$ nicht lösbar, so kann man $Y^T\nabla f(x)$ als "Schätzung" für den Vektor der Lagrange-Parameter in einem benachbarten stationären Punkt ansehen.

1.1. Reduktion bei quadratischer Zielfunktion.

Wir betrachten nun noch den Spezialfall der quadratischen Optimierungsaufgabe

$$\begin{aligned} \min \quad & f(x) = \frac{1}{2} x^T C x + d^T x \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

mit einer symmetrischen $n \times n$-Matrix C und $x, d \in \mathbb{R}^n$.

Dann ergibt sich die reduzierte Zielfunktion zu

$$\begin{aligned} F(u) &= f(Yb + Zu) \\ &= \frac{1}{2}(Yb + Zu)^T C (Yb + Zu) + d^T (Yb + Zu) \\ &= \frac{1}{2} u^T \left(Z^T C Z\right) u + (d + CYb)^T Z u + \left(d + \frac{1}{2} CYb\right)^T Yb \end{aligned}$$

Ist $Z^T C Z$ positiv definit, so ist das Minimum von F gegeben durch

$$\begin{aligned} 0 = \nabla F(u) &= \left(Z^T C Z\right) u + (d + CYb)^T Z \\ \rightsquigarrow \left(Z^T C Z\right) u &= -Z^T (d + CYb). \end{aligned}$$

2. Strategie der aktiven Menge bei linearen Ungleichungen

Nun wollen wir Optimierungsprobleme folgender Form betrachten

$$\begin{aligned} \min \quad & f(x) & \\ \text{s.t.} \quad & a_j^T x \geq b_j & j \in I = \{1, \ldots, m'\} \\ & a_k^T = b_k & k \in E = \{m'+1, \ldots, m\} \end{aligned} \tag{2.1}$$

mit $x \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $i = 1, \ldots, m$, $f \in C^1(\mathbb{R}^n)$.

Betrachten wir eine lokale Lösung x^* von (2.1), so ist x^* auch lokale Lösung von

$$\begin{aligned} \min \quad & f(x) & \\ \text{s.t.} \quad & a_j^T x = b_j & j \in J(x^*) \cup E, \end{aligned} \tag{2.2}$$

wobei

$$J(x^*) := \left\{ i \in I \,\middle|\, a_i^T x^* = b_i \right\}$$

die Indexmenge der in x^* aktiven Ungleichungsrestriktionen ist. Würden wir die Menge $J(x^*)$ kennen, so könnten wir Verfahren aus dem letzten Abschnitt anwenden, um eine Lösung von (2.1) zu erhalten. Nun ist $J(x^*)$ aber im allgemeinen nicht im Voraus bekannt. Die Strategie der aktiven Menge besteht nun darin, iterativ vorzugehen, in jedem Schritt mit einer Schätzung J_a von $J(x^*)$ das Hilfsproblem (2.2) zu bearbeiten und dann J_a geeignet zu modifizieren:

Es bezeichne im k-ten Schritt $x = x^k$ den aktuellen Punkt, dieser sei zulässig für (2.1), und es sei $J_a \subset J(x)$. Dann berechnen wir eine Verschiebung Δx für x mittels folgenden Hilfsproblems:

$$\begin{aligned} \min \quad & f(x + \Delta x) & \\ \text{s.t.} \quad & a_j^T \Delta x = 0 & j \in J_a \cup E \end{aligned} \tag{2.3}$$

Die Menge J_a muß in folgenden Fällen verändert werden:

(1) Ist $\Delta x = 0$ Lösung von (2.3), aber x noch keine Lösung von (2.1), so muß man die Gleichungsrestriktionen lockern, als (mindestens) einen Index aus J_a entfernen.

(2) Verletzt man beim Abwärtsgehen längs einer für (2.3) zulässigen Suchrichtung eine Ungleichung $a_p^T x \geq b_p$, $p \notin J_a$, so muß man vor dem Verlassen des zulässigen Bereiches von (2.1) anhalten und p in J_a aufnehmen.

Zunächst zu (1): Ist $\Delta x = 0$ Lösung von (2.3), so liefert dies die Existenz von Zahlen λ_i, $i \in J_a \cup E$ mit

$$\nabla f(x) = \sum_{i \in J_a \cup E} \lambda_i a_i.$$

Gilt $\lambda_i \geq 0$ für alle $i \in J_a$ so ist mit $\lambda_i := 0$ für $i \in J(x) \setminus J_a$ eine notwendige Bedingung dafür erfüllt, daß x Lösung von (2.1) ist. Gilt dagegen für ein $q \in J_a$

$$\lambda_q < 0,$$

so kann man eine für (2.1) zulässige Abstiegsrichtung finden, wenn man zuläßt, daß die Restriktion q inaktiv wird. Gibt es mehrere negative λ_i, so wählt man üblicherweise einen kleinsten.

Zu (2): Ist s eine für (2.3) zulässige Suchrichtung (d.h. es gilt $a_i^T s = 0$ für $i \in J_a \cup E$), so unterliegt die Schrittlänge $\alpha > 0$ den Einschränkungen

$$a_i^T (x + \alpha s) \geq b_i, \quad i \in I \setminus J_a,$$

dies ist gleichbedeutend mit

$$\alpha \leq \bar{\alpha} := \min \left\{ \frac{b_i - a_i^T x}{a_i^T s} \,\middle|\, i \in I \setminus J_a, a_i^T s < 0 \right\}.$$

Man muß also die verwendete Schrittlängenstrategie mit der Bedingung $\alpha \leq \bar{\alpha}$ koppeln. Wird $\alpha = \bar{\alpha}$ gewählt, so nehme man einen Index p in J_a auf, für welchen das Minimum angenommen wird, d.h. $\bar{\alpha} = \frac{b_p - a_p^T x}{a_p^T s}$ gilt.

Damit können wir eine Formulierung der <u>Strategie der aktiven Menge</u> angeben:

- Wähle einen für (2.1) zulässigen Punkt x, setze $J_a := J(x)$
- Wiederhole bis Abbruch
  - Wenn $\Delta x = 0$ als Lösung von (2.3) akzeptiert wird, so
    - berechne λ_i, $i \in J_a \cup E$, so daß $\nabla f(x) = \sum_{i \in J_a \cup E} \lambda_i a_i$, z.B. mittels $\lambda = Y^T \nabla f(x)$
    - bestimme einen Index $q \in J_a$, so daß $\lambda_q = \min_{i \in J_a} \lambda_i$
    - Wenn $\lambda_q \geq 0$ ist, so stoppe, sonst setze $J_a := J_a \setminus \{q\}$.
  - Bestimme eine Suchrichtung s für (2.3)
  - Bestimme eine Schrittlänge $\alpha \leq \bar{\alpha}$ für (2.3)
  - Ist $\alpha = \bar{\alpha}$, so bestimme ein $p \in I \setminus J_a$ mit $a_p^T (x + \alpha s) = b_p$ und setze $J_a := J_a \cup \{p\}$.
  - $x := x + \alpha s$

Bemerkung.

(1) Einen zulässigen Startpunkt kann man wie beim Simplex-Verfahren erhalten: Einen Vektor $x \geq 0$ mit $Ax = b$ ($A \in \mathbb{R}^{m \times n}$, $b \geq 0$) erhält man dort durch Lösen von

$$\begin{aligned} \min \quad & \sum_{i=1}^m z_i \\ \text{s.t.} \quad & Ax + z = b \\ & x \geq 0, z \geq 0 \end{aligned}$$

mit Start-Vektor $x = 0$, $z = b$.

(2) Stellt man bei der Abfrage, ob $\Delta x = 0$ als Lösung von (2.3) akzeptiert wird, hohe Anforderungen, so kann es passieren, daß die Schleife unter Umständen sehr oft mit unverändertem J_a durchlaufen wird. Wenn aber J_a noch nicht die richtige Indexmenge ist, so ist dies eigentlich nicht nötig.
Geringe Anforderungen können aber falsche Vorzeicheninformation für die λ_i liefern, so daß zu früh gestoppt wird oder die neue Schrittlänge $\alpha = 0$ gewählt wird. Dann wird der gerade entfernte Index wird wieder hinzugefügt.

Hier muß man einen Kompromiß zwischen hohen und geringen Anforderungen finden, indem man z.B. $\Delta x = 0$ akzeptiert, falls die (geschätzte) Verbesserung der Zielfunktion im nächsten Schritt kleiner ist als die Verbesserung seit der letzten Verkleinerung von J_a.

Verwendet man zur Bestimmung der λ_i anstatt $\nabla f(x) = \sum_{i \in J_a \cup E} \lambda_i a_i$ die Vorschrift $\lambda = Y^T \nabla f(x)$ aus dem ersten Abschnitt, so ermöglicht dies die Berechnung von Zahlen λ_i auch in dem Fall, daß $\Delta x = 0$ nicht die exakte Lösung von (2.3) ist.

(3) Bei Bearbeitung von (2.3) mittels einer Reduktionsmethode müssen bei jeder Veränderung der Indexmenge J_a die Matrizen Y und Z erneut berechnet werden. Nun verändert sich die Matrix A mit den Zeilen a_i^T, $i \in J_a \cup E$, nur durch Hinzufügen oder Streichen einer Zeile. Dann lassen sich die neuen Matrizen Y und Z aus den alten Matrizen gewinnen, ohne sie komplett neu berechnen zu müssen.

2.1. **Strategie der aktiven Menge bei quadratischer Zielfunktion und linearen Ungleichungen.**

Wir betrachten hier wieder den folgenden Spezialfall:

$$\begin{array}{lll} \min & f(x) = \frac{1}{2} x^T C x + d^T x & \\ \text{s.t.} & a_i^T x \geq b_i & i \in I = \{1, \ldots, m'\} \\ & a_i^T x = b_i & i \in E = \{m'+1, \ldots, m\} \end{array} \tag{2.4}$$

mit $C \in \mathbb{R}^{n \times n}$ symmetrisch und $d \in \mathbb{R}^n$. Das Hilfsproblem (2.3) lautet hierfür

$$\begin{array}{lll} \min & \frac{1}{2} \Delta x^T C \Delta x + (Cx + d)^T \Delta x & \\ \text{s.t.} & a_i^T \Delta x = 0 & i \in J_a \cup E \end{array} \tag{2.5}$$

Dieses Problem kann mit dem Algorithmus aus dem letzten Abschnitt im Prinzip explizit gelöst werden. Weiter kann die Schrittlänge α explizit angegeben werden. Man erhält dann folgenden Algorithmus:

- Wiederhole bis Abbruch
  - Wenn $\Delta x = 0$ Lösung von (2.5) ist, so
    - berechne λ_i, $i \in J_a \cup E$
    - bestimme einen Index $q \in J_a$, so daß $\lambda_q = \min_{i \in J_a} \lambda_i$
    - Wenn $\lambda_q \geq 0$ ist, so stoppe, sonst setze $J_a := J_a \setminus \{q\}$.
  - Bestimme eine Lösung Δx von (2.5)
  - Bestimme eine Schrittlänge $\alpha := \min\{1, \bar{\alpha}\}$
  - Ist $\alpha = \bar{\alpha}$, so bestimme ein $p \in I \setminus J_a$ mit $a_p^T (x + \alpha \Delta x) = b_p$ und setze $J_a := J_a \cup \{p\}$.
  - $x := x + \alpha \Delta x$

Ist C positiv definit und gilt $\alpha \neq 0$, so liefert dieser Algorithmus in endlich vielen Schritten die Lösung von (2.4). Dies wird das entscheidene Hilfsmittel für die SQP-Verfahren sein, die im nächsten Abschnitt besprochen werden.

3. SQP-Verfahren

Das Grundprinzip der SQP-Verfahren beruht in einer schrittweisen Ersetzung des Ausgangsproblemes durch ein quadratisches Hilfsproblem (d.h. durch eine quadratische Approximation des Zielfunktionals und eine lineare Approximation der Gleichungs- und Ungleichungsrestriktionen). Dieses quadratische Hilfsproblem läßt sich mit den im letzten Abschnitt besprochenen Techniken effizient lösen.

3.1. **Die Lagrange-Newton-Iteration.**

Ausgangspunkt für die SQP-Verfahren ist die sogenannte Lagrange-Newton-Iteration für Optimierungsaufgaben mit Gleichungsrestriktionen. Wir betrachten also zunächst folgende Aufgabe:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & c_i(x) = 0 \quad i = 1, \ldots, m \end{array}$$

Ist x^* ein reguläres Minimum von f unter den gegebenen Nebenbedingungen, so gilt mit der Lagrange-Funktion

$$L(x;\lambda) := f(x) - \sum_{i=1}^{m} \lambda_i c_i(x)$$

die notwendige Bedingung:

$$\nabla L(x;\lambda) := \begin{pmatrix} \nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla c_i(x) \\ -c(x) \end{pmatrix} = 0 \tag{3.1}$$

mit $c(x) = (c_1(x), \ldots, c_m(x))^T$. Die Karush-Kuhn-Tucker-Bedingungen (3.1) bilden ein nichtlineares Gleichungssystem zur Berechnung von (x, λ). Hierzu läßt sich nun das Newton-Verfahren anwenden. Man erhält dann das folgende lineare Gleichungssystem für die Newton-Korrekturen $(\Delta x, \Delta\lambda)$:

$$\nabla^2 L(x,\lambda) \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = -\nabla L(x,\lambda). \tag{3.2}$$

Die Hesse-Matrix der Lagrange-Funktion hat dabei die folgende Form:

$$\nabla^2 L(x,\lambda) = \begin{pmatrix} \nabla^2 f(x) - \sum_{i=1}^{m} \lambda_i \nabla^2 c_i(x) & -\nabla c(x)^T \\ -\nabla c(x) & 0 \end{pmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$$

Mit

$$W := \nabla^2 f(x) - \sum_{i=1}^{m} \lambda_i \nabla^2 c_i(x)$$

haben die Newton-Gleichungen (3.2) dann die Form

$$\begin{aligned} \begin{pmatrix} W & -\nabla c(x)^T \\ -\nabla c(x) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} &= -\begin{pmatrix} \nabla f(x) - \nabla c(x)^T \lambda \\ -c(x) \end{pmatrix} \\ \rightsquigarrow \begin{pmatrix} W & -\nabla c(x)^T \\ -\nabla c(x) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda + \Delta\lambda \end{pmatrix} &= \begin{pmatrix} -\nabla f(x) \\ c(x) \end{pmatrix} \end{aligned} \tag{3.3}$$

Dann läßt sich zeigen:

Satz. *Ist x^* ein reguläres lokales Minimum von f auf der zulässigen Menge M mit Lagrange-Multiplikatoren λ^* und ist W positiv definit auf dem Tangentialraum $T_{x^*}M = \{y \in \mathbb{R}^n \mid \nabla c(x^*)\, y = 0\}$, so ist die Hesse-Matrix $\nabla^2 L(x^*, \lambda^*)$ regulär und die Lagrange-Newton-Iteration konvergiert für alle Startwerte (x^0, λ^0) mit $\|x^0 - x^*\|$ hinreichend klein und $\nabla^2 L(x^0, \lambda^0)$ regulär. Die Konvergenz ist quadratisch.*

3.2. SQP-Verfahren für gleichungsrestringierte Optimierungsaufgaben.

Satz. *Ist die Hesse-Matrix W für vorgegebene (x,λ) positiv definit auf dem Tangentialraum T_xM, so läßt sich die Newton-Korrektur Δx aus (3.3) auch als Lösung der folgenden quadratischen Optimierungsaufgabe mit linearen Gleichungsrestriktionen erhalten:*

$$\begin{aligned} &\min \quad q(\Delta x) := \frac{1}{2}\Delta x^T W \Delta x + \nabla f(x)^T \Delta x + f(x) \\ &\text{s.t.} \quad \nabla c(x)\cdot \Delta x + c(x) = 0 \end{aligned} \tag{3.4}$$

Beweis. Erfüllt (x^*,λ^*) die hinreichende Bedingung für ein lokales Minimum von f auf M und ist x^* ein regulärer Punkt, d.h. $\nabla c(x^*)$ besitzt maximalen Rang, so ist das quadratische Hilfsproblem (3.4) für (x,λ) in einer Umgebung von (x^*,λ^*) eindeutig lösbar. Für solche (x,λ) lautet die Lagrange-Funktion zu (3.4) mit Lagrange-Parameter μ dann

$$Q(\Delta x;\mu) = \frac{1}{2}\Delta x^T W\Delta x + \nabla f(x)^T \Delta x + f(x) - \mu^T\left(\nabla c(x)\cdot\Delta x + c(x)\right)$$

Die notwendige Bedingung erster Ordnung lautet dann $\nabla Q(\Delta x;\mu) = 0$, also

$$0 = \begin{pmatrix} W\Delta x + \nabla f(x) - \nabla c(x)^T\mu \\ \nabla c(x)\cdot\Delta x + c(x)\end{pmatrix} = \begin{pmatrix} W & -\nabla c(x)^T \\ \nabla c(x) & 0\end{pmatrix}\begin{pmatrix}\Delta x \\ \mu\end{pmatrix} - \begin{pmatrix} -\nabla f(x) \\ c(x)\end{pmatrix}$$

Dies sind gerade die Lagrange-Newton-Gleichungen (3.3) mit $\mu = \lambda + \Delta\lambda$. □

Bemerkung. Das Hilfsproblem (3.4) kann man sich als Taylor-Entwicklung von $f(x+\Delta x)$ und $c(x+\Delta x)$ um $\Delta x = 0$ vorstellen. Es ist aber zu beachten, daß im quadratischen Teil der Zielfunktion die Hesse-Matrix der Lagrange-Funktion und nicht die von f auftritt.

Nun können wir das <u>SQP-Verfahren für Gleichungsrestriktionen</u> formulieren:

- Wähle Startvektoren $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$.
- Wiederhole
  - Löse das quadratische Hilfsproblems

$$\begin{aligned} &\min \quad \frac{1}{2}\Delta x^T W \Delta x^T + \nabla f(x)^T \Delta x + f(x) \\ &\text{s.t.} \quad \nabla c(x)\,\Delta x + c(x) = 0 \end{aligned}$$

  - Bestimme ein neues λ aus $\nabla c(x)^T\lambda = W\Delta x + \nabla f(x)$
  - Setze $x := x + \Delta x$

  bis $\|\Delta x\|$ hinreichend klein ist.

Bemerkung. Ist $\operatorname{Rang}(\nabla c(x)) = m$ und ist W positiv definit auf $\ker\nabla c(x)$, so ist die Existenz der Lagrange-Parameter λ für das quadratische Hilfsproblem gesichert. Da das Gleichungssystem zur Bestimmung von λ überbestimmt ist, empfiehlt es sich, λ mittels der im ersten Abschnitt angegebenen Formel zu berechnen (z.B. mittels QR-Zerlegung von $\nabla c(x)^T$).

3.3. **Übertragung auf Ungleichungsrestriktionen.**

Wir betrachten nun die Optimierungsaufgabe

$$\begin{aligned} \text{(3.5)} \qquad & \min && f(x) \\ & \text{s.t.} && c_i(x) = 0 \quad i \in E := \{1, \ldots, m'\} \\ & && c_i(x) \geq 0 \quad i \in I := \{m'+1, \ldots, m\} \end{aligned}$$

Das zugehörige quadratische Hilfsproblem wird dann folgendermaßen definiert:

$$\begin{aligned} \text{(3.6)} \qquad & \min && q(\Delta x) = \frac{1}{2}\Delta x^T W \Delta x + \nabla f(x)^T \Delta x + f(x) \\ & \text{s.t.} && \nabla c_i(x)^T \Delta x + c_i(x) = 0, \quad i \in E \\ & && \nabla c_i(x)^T \Delta x + c_i(x) \geq 0, \quad i \in I. \end{aligned}$$

Bei der Lösung dieser Aufgabe werden mittels der Strategie der aktiven Menge die Indizes der aktiven linearisierten Ungleichungsnebenbedingungen bestimmt:

$$J_a := \left\{ i \in I \,\middle|\, \nabla c_i(x)^T \Delta x + c_i(x) = 0 \right\}$$

A bezeichne die Matrix mit den Zeilen $\nabla c_i(x)^T$, $i \in J_a \cup E$. Ist Δx Lösung von (3.6) und hat die zugehörige Matrix A vollen Rang, so existieren Lagrange-Multiplikatoren λ_i, $i \in J_a \cup E$, mit

$$\nabla Q(\Delta x; \lambda) = W\Delta x + \nabla f(x) - A^T \lambda = 0$$

und es gilt $\lambda_i \geq 0$ für $i \in J_a$.

Dann können wir das <u>SQP-Verfahren</u> formulieren:

- Wähle Startvektoren $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$.
- Wiederhole
  - Löse die quadratische Hilfsaufgabe (3.6) mittels Reduktion und einer Strategie der aktiven Menge. Bestimme Δx, J_a und A.
  - Bestimme $\tilde{\lambda} \in \mathbb{R}^{|J_a \cup E|}$ aus $A^T \tilde{\lambda} = W\Delta x + \nabla f(x)$.
    Setze $\lambda_i := \tilde{\lambda}_i$ für $i \in E \cup J_a$ und $\lambda_i = 0$ für $i \in I \setminus J_a$.
  - Setze $x := x + \Delta x$.

  bis $\|\Delta x\|$ hinreichend klein ist.