

## Kapitel 12

# Innere–Punkt–Verfahren

Innere–Punkt–Verfahren verfolgen die Idee, bei der Lösung von linearen Programmen durch das Innere des konvexen Polyeders zum Optimum zu gelangen. Damit unterscheiden sie sich grundsätzlich von der Simplexmethode.

**Bemerkung 12.1** *Historie von Innere–Punkt–Verfahren.*

- Dikin 1967: hat bereits die Idee von Karmarkar (1984) umgesetzt, Arbeit wurde aber nicht wahrgenommen.
- Fiacco, McCormick 1968: Innere–Punkt–Verfahren für nichtlineare Optimierungsprobleme, waren aber nicht wettbewerbsfähig im Vergleich zu anderen Verfahren.
- Khachian 1979: Ellipsoidmethode. Polynomiale Komplexität konnte bewiesen werden, allerdings war das Verfahren wenig praxistauglich wegen numerischen Instabilitäten.
- Karmarkar 1984: projektive Methode, erste praxistaugliche Innere–Punkt–Methode.

□

Die Motivation zur Entwicklung von Alternativen zur Simplexmethode liegt in deren schlechter theoretischer Komplexität begründet. Man kann nicht ausschließen, dass die Anzahl der Iterationen exponentiell mit der Problemgröße wächst, siehe Abschnitt 8.2. Das hat im wesentlichen zwei Gründe:

- Die Simplexmethode kennt als Abstiegsrichtungen nur die Kanten auf dem Rand des zulässigen Bereiches. Deren Anzahl wächst exponentiell mit der Dimension des Polyeders.
- Die Simplexmethode sucht die Abstiegsrichtung lokal aus. Für diese Richtung spielen nur die im aktuellen Eckpunkt aktiven Nebenbedingungen eine Rolle, nicht aber die Gesamtheit der Nebenbedingungen.

Alternativen zur Simplexmethode müssen an diesen beiden Schwachstellen angreifen: sie sollten Abstiegsrichtungen durch das Innere des Polyeders zulassen und bei der Bestimmung dieser Richtungen Informationen von allen Nebenbedingungen einbeziehen.

Die Idee von Khachians Ellipsoidmethode lässt sich grob wie folgt beschreiben: Man startet mit einem zulässigen Punkt  $\bar{\mathbf{x}}$  und sucht einen möglichst zentral gelegenen Punkt  $\tilde{\mathbf{x}}$  in einem Restpolyeder, welches nur Punkte  $\mathbf{x}$  mit  $\mathbf{c}^T \mathbf{x} \leq \mathbf{c}^T \bar{\mathbf{x}}$  enthält. Diesen Punkt findet man dadurch, dass man mittels einer Iteration das Restpolyeder möglichst gut in ein Ellipsoid einbettet und dessen Mittelpunkt  $\tilde{\mathbf{x}}$  betrachtet. Ist  $\tilde{\mathbf{x}}$  zulässig, setzt man  $\bar{\mathbf{x}} = \tilde{\mathbf{x}}$  und startet die Konstruktion von Neuem. Die Instabilitäten rührten daher, dass die Ellipsoide im Laufe der Iteration immer schmaler wurden.

Das Verfahren von Karmarkar geht andersherum vor: hier wird ein Ellipsoid in den zulässigen Bereich eingebettet. Es funktioniert grob wie folgt: Man startet mit einem inneren Punkt des zulässigen Bereiches  $\bar{\mathbf{x}}$  und konstruiert ein im zulässigen Bereich eingeschriebenes, möglichst großes Ellipsoid. Dann nimmt man anstelle des zulässigen Bereiches nur das Ellipsoid und minimiert darüber die Zielfunktion. Es ergibt sich ein Randpunkt  $\tilde{\mathbf{x}}$ , an dem das Minimum angenommen wird. Man setzt  $\bar{\mathbf{x}} = \tilde{\mathbf{x}}$  und startet die Konstruktion von Neuem.

Das Verfahren von Karmarkar arbeitet praktisch wesentlich besser als die Ellipsoidmethode von Khachian. Man kann auch beweisen, dass die Komplexität des Verfahrens von Karmarkar besser ist. Für sehr große Problem ist dieses Verfahren oft schneller als die Simplexmethode.

In diesem Kapitel soll jedoch nur eine einfache Innere-Punkt-Methode im Detail besprochen werden.

## 12.1 Das Newton-Verfahren

Die betrachtete Innere-Punkt-Methode beruht auf dem Newton-Verfahren. Hier werden nur kurz einige Fakten zu diesem Verfahren zusammengestellt.

Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  mit  $f \in C^3(\mathbb{R}^n)$  gegeben. Gesucht ist eine Nullstelle  $\bar{\mathbf{x}}$  dieser Funktion  $f(\bar{\mathbf{x}}) = \mathbf{0}$ . Sei  $\mathbf{x}$  eine Näherung an  $\bar{\mathbf{x}}$ . Setzt man  $\bar{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$ , dann erhält man mit der Taylorentwicklung, abgebrochen nach dem linearen Term, die Approximation

$$\mathbf{0} = f(\bar{\mathbf{x}}) = f(\mathbf{x} + \Delta\mathbf{x}) \approx f(\mathbf{x}) + Df(\mathbf{x})\Delta\mathbf{x},$$

wobei  $Df(\mathbf{x})$  die Jacobi-Matrix von  $f$  in  $\mathbf{x}$  ist. Unter der Annahme, dass die Jacobi-Matrix regulär ist, erhält man

$$\Delta\mathbf{x} \approx - (Df(\mathbf{x}))^{-1} f(\mathbf{x}) \implies \bar{\mathbf{x}} \approx \mathbf{x} - (Df(\mathbf{x}))^{-1} f(\mathbf{x}).$$

Diese Beziehung motiviert die Berechnung der nächsten Iterierten  $\mathbf{x}^+$  des Newton-Verfahrens

$$\mathbf{x}^+ := \mathbf{x} - (Df(\mathbf{x}))^{-1} f(\mathbf{x}).$$

Zum Konvergenzverhalten des Newton-Verfahrens gibt es folgende bekannte Aussage.

**Satz 12.2** *Seien  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  mit  $f \in C^3(\mathbb{R}^n)$ ,  $\bar{\mathbf{x}}$  eine Nullstelle von  $f$  und  $|Df(\bar{\mathbf{x}})| \neq 0$  (Determinante). Dann gibt es ein  $\varepsilon > 0$ , so dass das Newton-Verfahren für alle Startwerte  $\mathbf{x}^{(0)}$  mit  $\|\bar{\mathbf{x}} - \mathbf{x}^{(0)}\|_2 \leq \varepsilon$  quadratisch gegen  $\bar{\mathbf{x}}$  konvergiert.*

Der Satz besagt zum einen, dass das Newton-Verfahren lokal konvergent ist, das heißt, es konvergiert, wenn man nahe genug an der Lösung beginnt. Quadratische Konvergenz bedeutet, dass es Konstanten  $c > 0$  und  $k_0 > 0$  gibt, so dass

$$\|\bar{\mathbf{x}} - \mathbf{x}^{(k+1)}\|_2 \leq c \|\bar{\mathbf{x}} - \mathbf{x}^{(k)}\|_2^2$$

für alle  $k \geq k_0$ .

## 12.2 Ein Kurz-Schritt-Algorithmus

Wir betrachten das lineare Programm

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{z = \mathbf{c}^T \mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\} \quad (12.1)$$

mit dem zugehörigen dualen Programm

$$\max_{\mathbf{y} \in \mathbb{R}^m} \{ \tilde{z} = \mathbf{b}^T \mathbf{y} : A^T \mathbf{y} + \mathbf{s} = \mathbf{c}, \mathbf{s} \geq \mathbf{0} \}. \quad (12.2)$$

Im dualen Programm haben wir hierbei die Schlupfvariablen  $\mathbf{s}$  eingeführt, um Gleichungsnebenbedingungen zu erhalten.

Innere-Punkte-Verfahren erzeugen eine Folgen von Punkten  $\{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{s}^{(k)})\}$ ,  $k = 0, 1, \dots$ , mit  $\mathbf{x}^{(k)} \geq \mathbf{0}$ ,  $\mathbf{s}^{(k)} \geq \mathbf{0}$ , deren Grenzwerte Optimallösungen von (12.1) und (12.2) liefern. Die Verfahren nennt man zulässige-Innere-Punkte-Verfahren, wenn alle Punkte  $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{s}^{(k)})$  zulässige innere Punkte von (12.1) und (12.2) sind, das heißt es gilt

$$A\mathbf{x}^{(k)} = \mathbf{b}, \mathbf{x}^{(k)} > \mathbf{0}, \quad A^T \mathbf{y}^{(k)} + \mathbf{s}^{(k)} = \mathbf{c}, \mathbf{s} > \mathbf{0}.$$

Um zulässige-Innere-Punkte-Verfahren verwenden zu können, müssen wir natürlich voraussetzen, dass die obigen Mengen nicht leer sind. Neben diesen Verfahren gibt es auch unzulässige-Innere-Punkte-Verfahren.

Aus der Nichtnegativität von  $\mathbf{x}$  und  $\mathbf{s}$  folgt

$$0 \leq \mathbf{x}^T \mathbf{s} = \mathbf{x}^T (\mathbf{c} - A^T \mathbf{y}) = \mathbf{c}^T \mathbf{x} - (A\mathbf{x})^T \mathbf{y} = \mathbf{c}^T \mathbf{x} - \mathbf{b}^T \mathbf{y} = z - \tilde{z}.$$

Von Satz 9.4 (Starker Dualitätssatz) wissen wir, dass für die Optimallösungen von (12.1) und (12.2) gilt  $z = \tilde{z}$ . Also folgt im Optimum  $\mathbf{x}^T \mathbf{s} = 0$  und wegen Nichtnegativität dieser beiden Vektoren sogar  $x_i s_i = 0$  für alle  $i = 1, \dots, n$ . Zusammen mit den Nebenbedingungen von (12.1) und (12.2) sind die Optimallösungen von (12.1) und (12.2) Lösungen des nichtlinearen Systems

$$\Psi_0(\mathbf{x}, \mathbf{y}, \mathbf{s}) := \begin{pmatrix} A\mathbf{x} - \mathbf{b} \\ A^T \mathbf{y} + \mathbf{s} - \mathbf{c} \\ X\mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{x}, \mathbf{s} \geq \mathbf{0}, \quad (12.3)$$

wobei  $X := \text{diag}(\mathbf{x}) \in \mathbb{R}^{n \times n}$  ist. Die Nichtlinearität ist in der letzten Gleichung. Die Jacobi-Matrix von  $\Psi_0$  ist gegeben durch

$$D\Psi_0(\mathbf{x}, \mathbf{y}, \mathbf{s}) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I_n \\ S & 0 & X \end{pmatrix} \in \mathbb{R}^{(2n+m) \times (2n+m)} \quad (12.4)$$

mit  $S := \text{diag}(\mathbf{s}) \in \mathbb{R}^{n \times n}$  und  $I_n$  der  $n$ -dimensionalen Einheitsmatrix.

**Lemma 12.3** *Unter den Voraussetzungen  $\text{rg}(A) = m$  und  $\mathbf{x} > \mathbf{0}$ ,  $\mathbf{s} > \mathbf{0}$  ist die Jacobi-Matrix (12.4) regulär.*

**Beweis:** Indirekt. Sei  $(\mathbf{u}^T, \mathbf{v}^T, \mathbf{w}^T)^T \neq \mathbf{0}$  ein Vektor mit

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I_n \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

Aus der ersten Gleichung folgt  $A\mathbf{u} = \mathbf{0}$ . Diese Beziehung wird in die umgestellte zweite Gleichung eingesetzt

$$\mathbf{w} = -A^T \mathbf{v} \implies \mathbf{u}^T \mathbf{w} = -\mathbf{u}^T A^T \mathbf{v} = -(A\mathbf{u})^T \mathbf{v} = 0.$$

Aus der dritten Gleichung folgt  $\mathbf{u} = -S^{-1}X\mathbf{w}$ . Die Invertierbarkeit von  $S$  folgt aus  $\mathbf{s} > \mathbf{0}$ . Mit der eben bewiesenen Beziehung und der Symmetrie der Diagonalmatrizen ergibt sich

$$0 = \mathbf{u}^T \mathbf{w} = -\mathbf{w}^T X S^{-1} \mathbf{w}.$$

Da  $\mathbf{x}, \mathbf{s} > \mathbf{0}$ , folgt daraus  $\mathbf{w} = \mathbf{0}$ . Aus der dritten Gleichung folgt damit  $\mathbf{u} = \mathbf{0}$ . Damit vereinfacht sich die zweite Gleichung zu  $A^T \mathbf{v} = \mathbf{0}$ . Wegen des vollen Spaltenrangs von  $A^T$  folgt daraus  $\mathbf{v} = \mathbf{0}$ . Damit ist der Widerspruch zur Annahme konstruiert. ■

Auf Grund dieses Satzes liegt es nahe, die Lösung des Systems (12.3) mit dem Newton-Verfahren zu versuchen. Dabei gibt es allerdings einige Schwierigkeiten:

- Die Regularität der Jacobi-Matrix ist nur für innere Punkte ( $\mathbf{x}, \mathbf{s} > \mathbf{0}$ ) bewiesen. Die gesuchte Lösung liegt aber nicht im Inneren, da aus der dritten Gleichung  $x_i s_i = 0$ ,  $i = 1, \dots, n$ , dass mindestens eine der Variablen  $x_i$  oder  $s_i$  im Optimum verschwindet.
- Es ist nicht klar, ob die in einem Newton-Schritt berechnete neue Iterierte überhaupt ein zulässiger Punkt ist. Falls nicht, ist die Regularität der Jacobi-Matrix nicht gesichert. Außerdem kann es passieren, dass das Newton-Verfahren gar nicht konvergiert oder zu einer Lösung konvergiert, welche die Nichtnegativitätsbedingung nicht erfüllt.

Um doch noch ein Newton-ähnliches Verfahren für (12.3) nutzen zu können, modifiziert man dieses System. Seien  $\mathbf{e}_n = (1, \dots, 1)^T \in \mathbb{R}^n$  und sei  $\mu > 0$  ein Parameter. Anstelle von (12.3) betrachtet man nun

$$\Psi_\mu(\mathbf{x}, \mathbf{y}, \mathbf{s}) := \begin{pmatrix} \mathbf{A}\mathbf{x} - \mathbf{b} \\ A^T \mathbf{y} + \mathbf{s} - \mathbf{c} \\ \mathbf{X}\mathbf{s} - \mu \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{x}, \mathbf{s} > \mathbf{0}. \quad (12.5)$$

Die Jacobi-Matrix dieses Systems ist (12.4).

Das schwerere Problem, eine Nullstelle von (12.3) zu finden, ersetzt man durch das einfachere Problem (12.5). Für festes  $\mu > 0$  kann die Lösung nur im Inneren des zulässigen Bereichs liegen, da  $x_i s_i = \mu$  für alle  $i$ . Da die Jacobi-Matrix nach Lemma 12.3 dort regulär ist, gibt es eine Umgebung der Lösung  $(\mathbf{x}(\mu)^T, \mathbf{y}(\mu)^T, \mathbf{s}(\mu)^T)^T$ , in der das Newton-Verfahren ohne Berücksichtigung der Ungleichungen  $\mathbf{x}, \mathbf{s} > \mathbf{0}$  quadratisch konvergiert. In diesem Sinne hat man das Problem (12.3) mit den Nichtnegativitätsbedingungen durch ein System ohne Nebenbedingungen ersetzt.

Da das Optimum von (12.1) auf dem Rand angenommen wird ist zu erwarten, dass die Lösung  $(\mathbf{x}(\mu)^T, \mathbf{y}(\mu)^T, \mathbf{s}(\mu)^T)^T$  umso weiter vom Rand entfernt liegt, je größer  $\mu$  ist.

- Man beginnt mit einem großen Wert  $\mu_0$ , für den sich die Lösung leicht berechnen lässt.
- Dann wird  $\mu$  sukzessive verkleinert und jedesmal wird die zugehörige Lösung bis auf eine gewisse Genauigkeit mit dem Newton-Verfahren approximiert.

Die Punktmenge

$$\{(\mathbf{x}(\mu)^T, \mathbf{y}(\mu)^T, \mathbf{s}(\mu)^T)^T : \mu > 0\}$$

wird zentraler Pfad genannt. Für seine Punkte existiert eine sogenannte Dualitätslücke:

$$\begin{aligned} 0 < \mu n &\leq \mathbf{x}(\mu)^T \mathbf{s}(\mu) = \mathbf{x}(\mu)^T (\mathbf{c} - A^T \mathbf{y}(\mu)) = \mathbf{c}^T \mathbf{x}(\mu) - (\mathbf{A}\mathbf{x}(\mu))^T \mathbf{y}(\mu) \\ &= \mathbf{c}^T \mathbf{x}(\mu) - \mathbf{b}^T \mathbf{y}(\mu) = z_\mu - \tilde{z}_\mu. \end{aligned} \quad (12.6)$$

**Lemma 12.4** Sei die dual zulässige Menge  $\{\mathbf{y} : A^T \mathbf{y} \leq \mathbf{c}\}$  beschränkt und sei das Innere dieser Menge nichtleer. Dann besitzt das Problem (12.5) für jedes  $\mu > 0$  eine eindeutige Lösung.

**Beweis:** Literatur, zum Beispiel [JS04, S 75ff.]. ■

Sei  $(\mathbf{x}^T, \mathbf{y}^T, \mathbf{s}^T)^T$  die gegenwärtige Iterierte zur Lösung von (12.5) für festes  $\mu > 0$ . Die Korrektur für die nächste Iterierte des Newton-Verfahrens berechnet

sich aus der Lösung von

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I_n \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{y} \\ \Delta \mathbf{s} \end{pmatrix} = - \begin{pmatrix} A\mathbf{x} - \mathbf{b} \\ A^T\mathbf{y} + \mathbf{s} - \mathbf{c} \\ X\mathbf{s} - \mu\mathbf{e} \end{pmatrix}. \quad (12.7)$$

Nach der Lösung dieses Systems folgt, dass die neue Iterierte  $((\mathbf{x} + \Delta\mathbf{x})^T, (\mathbf{y} + \Delta\mathbf{y})^T, (\mathbf{s} + \Delta\mathbf{s})^T)^T$  die ersten zwei Gleichungen von (12.5) erfüllt, da diese linear sind. *Übungsaufgabe* Wir können annehmen, dass die gegenwärtige Iterierte bereits durch einen Newton-Schritt berechnet wurde, sie damit ebenfalls diese Eigenschaft besitzt und sich (12.7) vereinfacht zu

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I_n \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{y} \\ \Delta \mathbf{s} \end{pmatrix} = - \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ X\mathbf{s} - \mu\mathbf{e} \end{pmatrix} =: \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -\mathbf{r} \end{pmatrix}. \quad (12.8)$$

Der Vektor  $\mathbf{r}$  wird Residuum genannt. Das Residuum der neuen Newton-Iterierten ist, unter Verwendung der Beziehungen von (12.8),

$$\begin{aligned} \tilde{\mathbf{r}} &= (X + \Delta X)(\mathbf{s} + \Delta\mathbf{s}) - \mu\mathbf{e} = X\mathbf{s} + \underbrace{X\Delta\mathbf{s} + S\Delta\mathbf{x} + \Delta X\Delta\mathbf{s}}_{=-\mathbf{r}} - \mu\mathbf{e} \\ &= X\mathbf{s} - \mathbf{r} + \Delta X\Delta\mathbf{s} - \mu\mathbf{e} = \Delta X\Delta\mathbf{s}. \end{aligned} \quad (12.9)$$

Um den Newton-Schritt analysieren zu können, werden einige Bezeichnungen eingeführt

$$D := \sqrt{XS^{-1}}, \quad \mathbf{q} := DX^{-1}\mathbf{r}, \quad R := \text{diag}(\mathbf{r}).$$

Direktes Nachrechnen oder Einsetzen zeigt, dass man die Lösung von (12.8) wie folgt aufschreiben kann: *Übungsaufgabe*

$$\begin{aligned} \Delta \mathbf{y} &= (AD^2A^T)^{-1}AD\mathbf{q}, \\ \Delta \mathbf{x} &= D^2A^T\Delta \mathbf{y} - D\mathbf{q}, \\ \Delta \mathbf{s} &= -D^{-1}\mathbf{q} - D^{-2}\Delta \mathbf{x}. \end{aligned}$$

Die Matrix

$$DA^T(AD^2A^T)^{-1}AD$$

definiert die Orthogonalprojektion auf den Bildraum von  $DA^T$ . Diese Projektion wird  $\Pi_R$  genannt. Demzufolge ist die Projektion in den Nullraum von  $DA^T$  gegeben durch  $\Pi_N := I - \Pi_R$ , wobei  $I$  die identische Abbildung ist. Nun kann man zeigen (*Übungsaufgabe*), dass

$$\Delta \mathbf{x} = -D\Pi_N\mathbf{q}, \quad \Delta \mathbf{s} = -D^{-1}\Pi_R\mathbf{q}$$

gelten. Das sind die beiden Ausdrücke, die wir in (12.9) analysieren müssen.

Wir nehmen nun an, dass man für gegebene  $\mathbf{x} > \mathbf{0}$ ,  $\mathbf{s} > \mathbf{0}$  und  $\mathbf{y}$  ein  $\beta \in [0, 1/2]$  finden kann, für welches

$$\|\mathbf{r}\|_2 \leq \beta\mu \quad (12.10)$$

gilt. Das heißt, der gegebene Punkt ist in einer gewissen Umgebung des zentralen Pfades. Für  $\beta = 0$  muss er direkt auf dem zentralen Pfad sein.

Wegen  $\mathbf{r} = X\mathbf{s} - \mu\mathbf{e}$  gilt

$$DX^{-1} = \sqrt{XS^{-1}X^{-2}} = \sqrt{X^{-1}S^{-1}} = \sqrt{(R + \mu I)^{-1}} = \left(\sqrt{R + \mu I}\right)^{-1}.$$

Bei dieser Rechnung und auch bei folgenden Rechnungen wird stark ausgenutzt, dass hier die Matrizen  $D$ ,  $R$ ,  $S$  und  $X$  Diagonalmatrizen sind. Für Nichtdiagonalmatrizen funktionieren die Rechnungen nicht mehr. Es folgt

$$\|DX^{-1}\|_2^2 = \|(R + \mu I)^{-1}\|_2 = \max_{1 \leq i \leq n} \frac{1}{|r_i + \mu|} \leq \frac{1}{\mu(1 - \beta)}. \quad (12.11)$$

Hierbei wurde die Definition der Spektralnorm einer Matrix ausgenutzt, sowie dass wir es mit einer Diagonalmatrix zu tun haben. Die letzte Ungleichung folgt aus der Dreiecksungleichung und (12.10)

$$|r_i + \mu| \geq \mu - |r_i| \geq \mu - \beta\mu = \mu(1 - \beta).$$

Aus der Verträglichkeit der Euklidischen Vektornorm und der Spektralnorm von Matrizen sowie (12.10) folgt damit

$$\tilde{\beta} := \|\mathbf{q}\|_2 = \|DX^{-1}\mathbf{r}\|_2 \leq \|DX^{-1}\|_2 \|\mathbf{r}\|_2 \leq \beta \sqrt{\frac{\mu}{(1 - \beta)}}.$$

Sei  $\theta$  der Winkel zwischen  $\mathbf{q}$  und  $\Pi_N \mathbf{q}$ . Dann folgt mit der Orthogonalität der Projektionen

$$\cos(\theta) = \frac{(\mathbf{q}, \Pi_N \mathbf{q})}{\|\mathbf{q}\|_2 \|\Pi_N \mathbf{q}\|_2} = \frac{(\Pi_R \mathbf{q} + \Pi_N \mathbf{q}, \Pi_N \mathbf{q})}{\|\mathbf{q}\|_2 \|\Pi_N \mathbf{q}\|_2} = \frac{\|\Pi_N \mathbf{q}\|_2^2}{\|\mathbf{q}\|_2 \|\Pi_N \mathbf{q}\|_2} = \frac{\|\Pi_N \mathbf{q}\|_2}{\|\mathbf{q}\|_2}.$$

Es folgen

$$\begin{aligned} \|D^{-1}\Delta \mathbf{x}\|_2 &= \|\Pi_N \mathbf{q}\|_2 = \cos(\theta) \|\mathbf{q}\|_2 = \tilde{\beta} \cos(\theta), \\ \|D\Delta \mathbf{s}\|_2 &= \|\Pi_R \mathbf{q}\|_2 = \|(I - \Pi_N)\mathbf{q}\|_2 = \sin(\theta) \|\mathbf{q}\|_2 = \tilde{\beta} \sin(\theta). \end{aligned} \quad (12.12)$$

Diese Abschätzungen kann man nun in (12.9) einsetzen. Mit der Cauchy–Schwarz–Ungleichung und  $\beta \leq 1/2$  erhält man

$$\begin{aligned} \|\tilde{\mathbf{r}}\|_2 &= \|\Delta X \Delta \mathbf{s}\|_2 \leq \|D^{-1}\Delta X\|_2 \|D\Delta \mathbf{s}\|_2 \leq \tilde{\beta}^2 \cos(\theta) \sin(\theta) \\ &= \frac{\tilde{\beta}^2}{2} \sin(2\theta) \leq \frac{\tilde{\beta}^2}{2} \leq \beta^2 \frac{\mu}{2(1 - \beta)} \leq \beta^2 \mu. \end{aligned} \quad (12.13)$$

Wir haben damit erhalten, dass unter der Voraussetzung (12.10) der relative Fehler  $\|X\mathbf{s} - \mu \mathbf{e}\|_2 / \mu = \|\mathbf{r}\|_2 / \mu$  in jedem Schritt des Newton–Verfahrens quadriert wird.

Aus (12.12) folgt insbesondere  $\|D^{-1}\Delta \mathbf{x}\|_2 \leq \tilde{\beta}$  und daraus, sowie mit (12.11), der Definition von  $\tilde{\beta}$  und  $\beta \leq 1/2$ ,

$$\begin{aligned} \|X^{-1}\Delta \mathbf{x}\|_2 &= \|X^{-1}DD^{-1}\Delta \mathbf{x}\|_2 \leq \|X^{-1}D\|_2 \|D^{-1}\Delta \mathbf{x}\|_2 \leq \frac{\tilde{\beta}}{\sqrt{\mu(1 - \beta)}} \\ &= \frac{\beta}{1 - \beta} \leq 1. \end{aligned}$$

Somit ist  $(\Delta x_i)/x_i \leq 1$  woraus

$$x_i - |\Delta x_i| \geq x_i - x_i = 0, \quad i = 1, \dots, n$$

folgt. Eine analoge Abschätzung kann man für  $\mathbf{s}$  durchführen. Die Nichtnegativität der neuen Iterierten ist gesichert.

Es gilt sogar die Positivität. *Übungsaufgabe ab hier.* Mit den Abschätzungen (12.13) und (12.9) folgt

$$\|\tilde{\mathbf{r}}\|_2 < \mu \implies \|\Delta X \Delta \mathbf{s}\|_2 < \mu \implies |\Delta \mathbf{x}_i \Delta \mathbf{s}_i| < \mu$$

für alle  $i = 1, \dots, n$ . Aus (12.9) erhält man ebenso

$$\|\tilde{\mathbf{r}}\|_2 = \|(X + \Delta X)(\mathbf{s} + \Delta \mathbf{s}) - \mu \mathbf{e}\|_2 < \mu,$$

was man abgekürzt und quadriert wie folgt schreiben kann

$$\sum_{i=1}^n (a_i - \mu)^2 < \mu^2,$$

mit  $a_i = (X + \Delta X)(\mathbf{s} + \Delta \mathbf{s})$ . Wir wissen bereits, dass beide Faktoren von  $a_i$  nichtnegativ sind. Sei für einen Index, zum Beispiel  $i = j$ , einer der Faktoren Null. Dann folgt

$$\mu^2 + \sum_{i=1, i \neq j}^n (a_i - \mu)^2 < \mu.$$

Diese Aussage kann nicht gelten, da alle Summanden der Summe nichtnegativ sind. Also sind alle Faktoren der  $a_i$ ,  $i = 1, \dots, n$ , positiv. Das heißt, es gelten  $\mathbf{x} + \Delta \mathbf{x} > \mathbf{0}$ ,  $\mathbf{s} + \Delta \mathbf{s} > \mathbf{0}$  und der Newton-Schritt liefert einen inneren Punkt. *Übungsaufgabe bis hier.*

**Algorithm 12.5 Kurz-Schritt-Algorithmus.** Es seien  $\mathbf{x}^{(0)} > \mathbf{0}$ ,  $\mathbf{y}^{(0)}$ ,  $\mathbf{s}^{(0)} > \mathbf{0}$  und  $\mu^{(0)}$  so gegeben, dass

$$\begin{aligned} A\mathbf{x}^{(0)} &= \mathbf{b}, \\ A^T \mathbf{y}^{(0)} + \mathbf{s}^{(0)} &= \mathbf{c}, \\ X^{(0)} \mathbf{s}^{(0)} - \mu^{(0)} \mathbf{e} &= \mathbf{r}^{(0)}, \\ \frac{\|\mathbf{r}^{(0)}\|_2}{\mu^{(0)}} &\leq \frac{1}{2} \end{aligned} \tag{12.14}$$

gelten. Es sei des weiteren eine gewünschte Genauigkeit  $\varepsilon > 0$  vorgegeben. Setze  $k = 0$ .

1. Führe einen Newton-Schritt aus, das heißt löse (12.8). Setze  $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta \mathbf{x}$ ,  $\mathbf{y}^{(k+1)} = \mathbf{y}^{(k)} + \Delta \mathbf{y}$  und  $\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} + \Delta \mathbf{s}$ .
2. Verkleinere den Parameter mittels der Vorschrift

$$\mu^{(k+1)} = \mu^{(k)} \left( 1 - \frac{1}{6\sqrt{n}} \right).$$

3. Setze  $k := k + 1$ .
4. Falls  $\mu^{(k)} \leq \varepsilon/n$  dann beende das Verfahren, ansonsten gehen zu Schritt 1. □

Die Bedingung (12.14) bedeutet, dass die Anfangsiterierte des Verfahrens sich relativ nahe am zentralen Pfad befinden muss. Die Bezeichnung „Kurz-Schritt-Algorithmus“ hat sich eingebürgert, weil der Parameter  $\mu$  in jedem Schritt nur etwa um den Faktor  $1 - 1/\sqrt{n}$  verkleinert wird, der für große  $n$  nahe bei Eins liegt.

Wir wollen nun die oben erarbeitete abstrakte Theorie auf den Kurz-Schritt-Algorithmus anwenden.

**Lemma 12.6** *Der Kurz-Schritt-Algorithmus 12.5 erfüllt die Bedingung (12.10) mit  $\beta = 1/2$ .*

**Beweis:** Der Beweis erfolgt induktiv. Wegen der Forderung (12.14) an den Startpunkt, ist der Induktionsanfang gegeben. Gelte nun  $\|\mathbf{r}^{(k)}\|_2 \leq \mu^{(k)}/2$  für ein

$k \geq 0$ . Aus (12.8), (12.9) (erste Gleichung) und der Definition des Parameters im Algorithmus 12.5 folgt

$$\mathbf{r}^{(k+1)} = X^{(k+1)} \mathbf{s}^{(k+1)} - \mu^{(k+1)} \mathbf{e} = \tilde{\mathbf{r}}^{(k)} + \mu^{(k)} \mathbf{e} - \mu^{(k+1)} \mathbf{e} = \tilde{\mathbf{r}}^{(k)} + \frac{\mu^{(k)}}{6\sqrt{n}} \mathbf{e}.$$

Aus (12.13) folgt  $\|\tilde{\mathbf{r}}^{(k)}\|_2 \leq \mu^{(k)}/4$ . Mit der Dreiecksungleichung,  $\|\mathbf{e}\|_2 = \sqrt{n}$  und dem größtmöglichen Verhältnis von  $\mu^{(k)}$  und  $\mu^{(k+1)}$  (für  $n = 1$ ) erhält man

$$\begin{aligned} \|\mathbf{r}^{(k+1)}\|_2 &\leq \|\tilde{\mathbf{r}}^{(k)}\|_2 + \frac{\mu^{(k)}}{6\sqrt{n}} \|\mathbf{e}\|_2 \leq \mu^{(k)} \left( \frac{1}{4} + \frac{1}{6} \right) = \frac{5}{12} \mu^{(k)} \leq \frac{5}{12} \frac{6}{5} \mu^{(k+1)} \\ &= \frac{1}{2} \mu^{(k+1)}. \end{aligned}$$

Das ist die Aussage des Lemmas. Die Induktionsvoraussetzung ist bei der Nutzung von (12.13) implizit verwendet wurden. ■

In einer Übungsaufgabe wurde gezeigt, dass die Iterierten bei der Ausführung des Newton-Schritts strikt zulässig bleiben ( $\mathbf{x}, \mathbf{s} > \mathbf{0}$ ). Damit ist der Kurz-Schritt-Algorithmus 12.5 wohldefiniert.

Nun sind wir in der Lage, die polynomiale Komplexität des Kurz-Schritt-Algorithmus zu beweisen.

**Satz 12.7** *Der Kurz-Schritt-Algorithmus 12.5 hält nach spätestens*

$$6\sqrt{n} \ln \left( \frac{n\mu^{(0)}}{\varepsilon} \right) \quad (12.15)$$

*Iterationen mit Näherungslösungen  $\mathbf{x} > \mathbf{0}$ ,  $\mathbf{y}$ ,  $\mathbf{s} > \mathbf{0}$  von (12.1) und (12.2), deren Dualitätslücke*

$$\mathbf{c}^T \mathbf{x} - \mathbf{b}^T \mathbf{y} \leq 2\varepsilon$$

*erfüllt.*

**Beweis:** Man muss den Abbruchindex und den Abbruchfehler untersuchen. Die Abbruchbedingung in Schritt 4 ist erfüllt, wenn

$$\ln \mu^{(k)} \leq \ln \frac{\varepsilon}{n}.$$

Nach Konstruktion gilt

$$\mu^{(k)} = \left( 1 - \frac{1}{6\sqrt{n}} \right)^k \mu^{(0)}.$$

Einsetzen liefert

$$k \ln \left( 1 - \frac{1}{6\sqrt{n}} \right) + \ln \left( \mu^{(0)} \right) \leq \ln \frac{\varepsilon}{n}$$

oder

$$-k \ln \left( 1 - \frac{1}{6\sqrt{n}} \right) \geq \ln \frac{n\mu^{(0)}}{\varepsilon} \iff k \ln \left( 1 + \frac{1}{6\sqrt{n}-1} \right) \geq \ln \frac{n\mu^{(0)}}{\varepsilon}.$$

Wegen  $\ln x \leq x - 1$  folgt daraus

$$k \left( 1 - \frac{1}{6\sqrt{n}-1} - 1 \right) \geq \ln \frac{n\mu^{(0)}}{\varepsilon} \iff k \geq (6\sqrt{n}-1) \ln \frac{n\mu^{(0)}}{\varepsilon}.$$

Daraus folgt die erste Aussage des Satzes.

Die Abschätzung der Dualitätslücke erfolgt unter Nutzung von (12.6), der dritten Gleichung von (12.8), der Cauchy–Schwarz–Ungleichung, (12.10) und der Abbruchbedingung des Verfahrens

$$\begin{aligned} \mathbf{c}^T \mathbf{x} - \mathbf{b}^T \mathbf{y} &= \mathbf{x}^T \mathbf{s} = \mathbf{e}^T (X \mathbf{s}) = \mathbf{e}^T (\mu^{(k)} \mathbf{e} + \mathbf{r}) \leq n\mu^{(k)} + \|\mathbf{e}\|_2 \|\mathbf{r}\|_2 \\ &= n\mu^{(k)} + \sqrt{n}\beta\mu^{(k)} \leq 2n\mu^{(k)} \leq 2\varepsilon. \end{aligned}$$

■

Man beachte, dass weder das Abbruchkriterium des Kurz–Schritt–Verfahrens noch die Dualitätslücke etwas darüber aussagen, wie groß der Abstand zwischen der mit dem Kurz–Schritt–Verfahren berechneten numerischen Lösung zur Lösung des linearen Programms (12.1) ist.

**Bemerkung 12.8 Zur Anzahl der Iterationen des Kurz–Schritt–Verfahrens.** In vielen Anwendungen lässt sich ein Startpunkt zum Algorithmus 12.5 mit  $n\mu^{(0)} \leq 10^{10}$  angeben. Die gewünschte Genauigkeit liegt im allgemeinen bei  $\varepsilon = 10^{-8}$  bis  $\varepsilon = 10^{-15}$ . Für den letzten Wert ist der Faktor  $6 \ln \left( \frac{n\mu^{(0)}}{\varepsilon} \right)$  in (12.15) ungefähr 345. Die Anzahl der Iterationen hängt im wesentlichen von  $\sqrt{n}$  ab. □

**Bemerkung 12.9 Innere–Punkt–Verfahren in der Praxis.** In der Praxis nutzt man vor allem unzulässige Innere–Punkt–Verfahren, das heißt, mit Iterierten, die außerhalb des zulässigen Bereichs liegen können. Die analytische Untersuchung dieser Verfahren ist jedoch komplizierter. □