

# Seminar Grundlegende Algorithmen der Optimierung Sommersemester 2001

## Primal-Dual Innere Punkte Verfahren

*Thomas Schlechte*

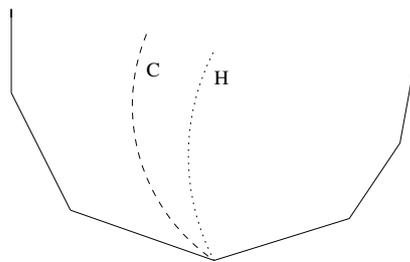


Fig. 1: Innere Punkte Methoden und deren Pfade.

- Studiengang Wirtschaftsmathematik
- Dozent Prof. Dr. Martin Grötschel
- betreut durch Dr. Frank Lutz

## Inhaltsverzeichnis

|       |  |    |
|-------|--|----|
| 1     | Einführung . . . . .   | 3  |
| 1.1   | Notation . . . . .   | 3  |
| 1.2   | Motivation . . . . .   | 3  |
| 1.3   | Karush-Kuhn-Tucker Bedingungen . . . . .                       | 4  |
| 2     | Primal Duale Innere Punkte Methoden . . . . .                  | 5  |
| 2.1   | Zentraler Pfad . . . . .                                       | 6  |
| 2.2   | Path Following Algorithmen . . . . .                           | 6  |
| 2.3   | Infeasible Starting Points . . . . .                           | 7  |
| 3     | Mehrotra's Predictor Corrector Methode . . . . .               | 8  |
| 3.1   | Ein vereinfachter Predictor-Corrector Algorithmus . . . . .    | 8  |
| 3.2   | Korrektheit des Algorithmus . . . . .                          | 9  |
| 3.2.1 | Lemma über Reduktion von $\mu$ . . . . .                       | 9  |
| 3.2.2 | Lemma über Reduktion von $\mu$ nach Prediktorschritt . . . . . | 10 |
| 3.2.3 | Lemma über Correctorschritt . . . . .                          | 11 |
| 3.2.4 | Lemma über Korrektheit des Algorithmus . . . . .               | 12 |
| 3.2.5 | Folgerung über Laufzeit . . . . .                              | 12 |
| 3.3   | Mehrotra's PC-Methode . . . . .                                | 13 |
| 3.3.1 | Der Algorithmus . . . . .                                      | 13 |
| 3.3.2 | Zusammenfassung . . . . .                                      | 14 |

# 1 Einführung

## 1.1 Notation

Diese Seminararbeit behandelt primal-duale Innere Punkte Methoden zur Lösung von Linearen Programmen, wobei die Predictor Corrector Methode von Mehrotra im Vordergrund steht. Betrachtet werden sollen hierzu LP's in folgender primaler und dualer Standardform.

$$\begin{aligned} \min \quad & c^T x \\ & Ax = b \\ & x \geq 0, \end{aligned}$$

$$\begin{aligned} \max \quad & b^T \lambda \\ & A^T \lambda + s = c \\ & s \geq 0. \end{aligned}$$

Dabei handelt es sich bei  $b$ ,  $s$  und  $\lambda$  um Vektoren aus dem  $\mathbb{R}^m$ ,  $c$  und  $x$  aus dem  $\mathbb{R}^n$  und  $A$  sei o.B.d.A. eine  $m \times n$  Matrix mit vollen Zeilenrang.

## 1.2 Motivation

Das Lösen von Linearen Programmen gehört zu den mathematischen Problemen für die es eine Vielzahl von Anwendungsgebieten gibt, so z.B. in der Planung, Logistik oder Finanzpolitik. Dennoch war lange Zeit unklar, ob es möglich sei diese Probleme in polynomialer Zeit zu lösen. Erst 1979 zeigte dies Khachiyan mit seiner Ellipsoidmethode. Für die Praxis war diese jedoch völlig ungeeignet und so blieb die Simplexmethode, von Dantzig Ende der Vierziger Jahre entwickelt, das non plus ultra. Dies trotz der Tatsache, dass für die bekannten Pivotregeln Klassen von Beispielen existieren, die exponentiellen Aufwand benötigen. Die Fragestellung ob es eine Pivotregel gibt für die der Simplexalgorithmus polynomial ist, ist eng verwandt mit der Hirsch Vermutung, (siehe [6]). Sei dazu  $\delta(G)$  der Durchmesser eines Graphen, d.h. die kleinste Zahl  $d$ , so dass für je zwei Knoten  $u, v$  in  $G$  ein  $(u, v)$ -Weg mit höchstens  $d$  Kanten existiert. Des weiteren sei  $\Delta(d, n)$  der maximale Durchmesser der Graphen, die durch ein  $d$ -dimensionales Polytop mit höchstens  $n$  Facetten induziert werden. So lautet die Hirsch-Vermutung von '57:

- Für  $n > d \geq 2$  sei  $\Delta(d, n)$  der grösste mögliche Durchmesser eines  $d$ -dimensionalen Polytops mit  $n$  Facetten. Dann gilt:

$$\Delta(d, n) \leq n - d . (?)$$

Denn Wert von  $\Delta(d, n)$  liefert unabhängig von der Pivotregel eine untere Schranke für die Anzahl der Iterationen der Simplexmethode. Diese Diskrepanz zwischen theoretischen Verständnis und praktischen Lösungen schliessen gerade die Inneren Punkte Methoden. Deswegen ist die Arbeit von Karmarkar 1984 von grösster Bedeutung. Er hatte die Idee von einem inneren Punkt der zulässigen Menge aus einen Weg zu einer optimalen Lösung zu konstruieren. Dies führte zu den heute bekannten Inneren Punkte Methoden, die zumindestens die entstandene Lücke zwischen Theorie und Praxis schliessen konnten. Denn diese Algorithmen sind auf der einen Seite polynomial in ihrer theoretischen Laufzeit und haben andererseits zu sehr guten Implementationen geführt, die eine Vielzahl von praktischen Problemen lösen können.

### 1.3 Karush-Kuhn-Tucker Bedingungen

Aus der Dualitätstheorie folgt, dass die Zielwerte des dualen LP's eine untere Schranke für die primale Zielfunktion liefert. Aus diesem Zusammenhang und dem Satz über die Existenz strikt komplementärer Lagrange-Multiplikatoren leiten sich nun folgende Optimalitätsbedingungen (KKT-Conditions) ab:

$$\begin{aligned}
 0 &\leq x^T s \\
 &= x^T (c - A^T \lambda) \\
 &= c^T x - (Ax)^T \lambda \\
 &= c^T x - b^T \lambda.
 \end{aligned}$$

$$\begin{aligned}
 Ax &= b \\
 A^T \lambda + s &= c \\
 x_i s_i &= 0, \quad i = 1, 2, \dots, n \\
 (x, s) &\geq 0,
 \end{aligned}$$

## 2 Primal Duale Innere Punkte Methoden

Primal Duale Innere Punkte Methoden finden primal-duale Lösungen durch Anwenden von modifizierten Newton Verfahren auf die KKT-Bedingungen. Hierzu wird folgende Funktion  $F$  betrachtet :

$$F : \mathbb{R}^{2n+m} \mapsto \mathbb{R}^{2n+m}$$

$$F(x, \lambda, s) = \begin{pmatrix} A^T \lambda + s - c \\ Ax - b \\ XS\mathbb{1} \end{pmatrix}$$

mit

$$(x, s) \geq 0 .$$

Wobei  $X = \text{diag}(x_1, x_2, \dots, x_n)$ ,  $S = \text{diag}(s_1, s_2, \dots, s_n)$  und  $\mathbb{1} = (1, 1, \dots, 1)^T \in \mathbb{R}^n$ . Bemerkenswert ist dabei, dass nur der untere Term  $XS\mathbb{1}$  nicht linear ist. Alle primal-duale Verfahren generieren Iterierte  $(x_k, \lambda_k, s_k)$ , die strikt obere Ungleichung erfüllen, d.h.  $x_k$  und  $s_k$  sind echt positiv. Innere Punkte Methoden basieren auf zwei verschiedenen Prozeduren, einmal der Berechnung der Richtung des Schrittes und zum zweiten der Länge um optimalen Fortschritt zu erreichen. Die Suchrichtung ergibt sich durch Lösen des folgenden linearen Systems:

$$J(x, \lambda, s) \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{pmatrix} = -F(x, \lambda, s)$$

Dabei handelt es sich bei  $J$  um die Jacobi-Matrix von  $F$ .

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -XS\mathbb{1} \end{pmatrix}$$

Um jetzt die ideale Schrittweite zu berechnen, wird ein Line-Search-Algorithmus hinzugezogen, so dass  $(x, s) > 0$  weiterhin gilt und damit ist die neue Iterierte (Predictor-Schritt):

$$(x, \lambda, s) + \alpha(\Delta x, \Delta \lambda, \Delta s) \text{ mit } \alpha \in (0, 1].$$

## 2.1 Zentraler Pfad

Der zentrale Pfad spielt eine wichtige Rolle für die Theorie der Innere Punkte Methoden, denn reine Predictor-Schritte würden die Iterierten an den Rand des Polyeders bringen und somit nur den Simplexalgorithmus kopieren, hier spricht man von affine scaling [1]. Deshalb sucht man nach Möglichkeiten sich zentral im Inneren des Polyeders zu bewegen, um eben ein direkten Weg zum Optimum zu konstruieren. Der zentrale Pfad  $\mathcal{C}$  ist genau so eine Kurve von strikt zulässigen Punkten im  $\mathbb{R}^{2n+m}$ . Parametrisiert über das Skalar  $\tau$  erfüllt jeder Vektor  $(x_\tau, \lambda_\tau, s_\tau) \in \mathcal{C}$  folgendes System:

$$\begin{aligned} Ax &= b \\ A^T \lambda + s &= c \\ x_i s_i &= \tau, i = 1, 2, \dots, n \\ (x, s) &> 0. \end{aligned}$$

Dass  $\mathcal{C}$  wohldefiniert und somit eindeutig für jedes  $\tau$  ist, kann mit Hilfe einer logarithmischen Barrierefunktion gezeigt werden (siehe [1]). Nun kann man einen Zentrierungsparameter  $\sigma \in [0, 1]$  und ein Dualitätsmass  $\mu$  einführen mit

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i s_i = \frac{x^T s}{n}.$$

Hieraus ergeben sich dann analog zu 2.1 folgende Schrittgleichungen:

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -XS\mathbb{1} + \sigma\mu\mathbb{1} \end{pmatrix}.$$

## 2.2 Path Following Algorithmen

Path Following Algorithmen verfolgen eben diesen zentralen Pfad  $\mathcal{C}$  und bleiben in einer strikt definierten Umgebung. Da  $\mu$  gerade den Fortschritt einer Iteration misst und für  $k \rightarrow \infty$  gegen 0 geht, nähern sich die Iterierten  $(x_k, \lambda_k, s_k)$  einer Lösung der KKT-Bedingungen an. Die zwei bedeutensten Umgebungen sind hierbei:

$$\mathcal{N}_2(\theta) = \{(x, \lambda, s) \in \mathcal{F}^0 \mid \|XS\mathbb{1} - \mu\mathbb{1}\|_2 \leq \theta\mu\},$$

$$\mathcal{N}_{-\infty}(\gamma) = \{(x, \lambda, s) \in \mathcal{F}^0 \mid x_i s_i \geq \gamma\mu \forall i = 1, 2, \dots, n\}, \text{ wobei}$$

$$\mathcal{F}_0 = \{(x, \lambda, s) \mid Ax = b, A^T \lambda + s = c, (x, s) > 0\}.$$

## 2.3 Infeasible Starting Points

Offensichtlich brauchen Path-Following Algorithmen strikt zulässige Startvektoren, jedoch kann man LP's konstruieren, die keine strikt zulässigen Punkte enthalten. Daraus und aus dem Problem der Berechnung eines zulässigen Punktes, der am besten noch in der Nähe des zentralen Pfades liegt, entwickelte sich der Ansatz, einen beliebigen, womöglich unzulässigen Punkt als Startwert zuzulassen, mit der einzigen Bedingung, dass  $x_0, s_0 > 0$  gilt. Dies führt zu den zwei folgenden Residualgrößen:

$$r_b = Ax - b, r_c = A^T \lambda + s - c .$$

Daraus ergeben sich folgende modifizierte Schrittgleichungen:

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -XS\mathbf{1} + \sigma\mu\mathbf{1} \end{pmatrix}$$

Interpretieren kann man diese Suchrichtung immer noch als Newton Schritt bzgl. des Punktes  $(x_{\sigma\mu}, \lambda_{\sigma\mu}, s_{\sigma\mu}) \in \mathcal{C}$ . Man kann zeigen, dass die Norm der Residualgrößen  $\|r_b\|_2, \|r_c\|_2$  beschränkt sind durch ein konstantes Vielfaches von  $\mu$ . Wenn man also nachweisen kann, dass  $\mu$  beliebig klein wird, wird auch  $r_b$  und  $r_c$  beliebig klein, und somit liefert dies einen zulässigen Optimalpunkt.

### 3 Mehrotra's Predictor Corrector Methode

#### 3.1 Ein vereinfachter Predictor-Corrector Algorithmus

1. Berechne Startvektor  $(x_0, \lambda_0, s_0) \in \mathcal{N}_2(0.25)$ .

2. Für  $k = 0, 1, 2, \dots$  und solange  $\mu > \varepsilon$  :

(a) Berechne Predictor-Schritt-Richtung durch  
Lösen des folgenden Gleichungssystems:

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \Delta x_{aff} \\ \Delta \lambda_{aff} \\ \Delta s_{aff} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X_k S_k \mathbb{1} \end{pmatrix}.$$

(b) Führe maximal zulässigen Predictor-Schritt durch:

$$\alpha_{aff} = \max\{\alpha \in [0, 1] \mid \begin{pmatrix} x_k \\ \lambda_k \\ s_k \end{pmatrix} + \alpha \begin{pmatrix} \Delta x_{aff} \\ \Delta \lambda_{aff} \\ \Delta s_{aff} \end{pmatrix} \in \mathcal{N}_2(0.5)\}$$

$$(x_k, \lambda_k, s_k) = (x_k, \lambda_k, s_k) + \alpha(\Delta x_{aff}, \Delta \lambda_{aff}, \Delta s_{aff})$$

(c) Berechne Corrector-Richtung durch  
Lösen des folgenden Gleichungssystems:

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x_{cor} \\ \Delta \lambda_{cor} \\ \Delta s_{cor} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mu \mathbb{1} - X_k S_k \mathbb{1} \end{pmatrix}.$$

(d) Berechne neue Iterierte :

$$(x_{k+1}, \lambda_{k+1}, s_{k+1}) = (x_{k+1}, \lambda_k, s_k) + (\Delta x_{cor}, \Delta \lambda_{cor}, \Delta s_{cor})$$

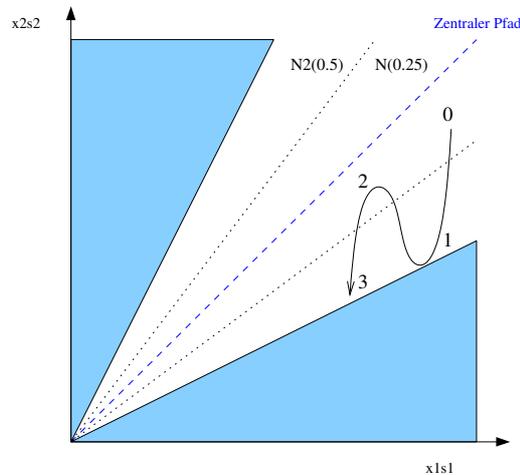


Fig. 2: Vorgehensweise der PC-Methode.

## 3.2 Korrektheit des Algorithmus

Die Idee des Beweises ist, zu zeigen, dass in jedem Schritt  $\mu$  polynomial abhängig von  $n$  verringert wird und somit beliebig klein wird. Die meisten in der Praxis eingesetzten Inneren Punkte Algorithmen arbeiten mit einer vorgegebenen Genauigkeit, so dass wenn gilt  $\mu \leq \varepsilon$  abgebrochen wird.

### 3.2.1 Lemma über Reduktion von $\mu$

Sei  $(\Delta x, \Delta \lambda, \Delta s)$  die Lösung des Gleichungssystem aus Abschnitt 2.3. Dann gilt:

$$\Delta x^T \Delta s = 0 \text{ und } \mu_\alpha := \frac{(x + \alpha \Delta x)^T (s + \alpha \Delta s)}{n} = (1 - \alpha(1 - \sigma))\mu.$$

**Beweis :**

$$\begin{aligned} \Delta x^T \Delta s &= \Delta x^T \Delta s + \overbrace{(A \Delta x)^T \Delta \lambda}^0 \\ &= \Delta x^T \Delta s + \Delta x^T A^T \Delta \lambda \\ &= \Delta x^T \underbrace{(\Delta s + A^T \Delta \lambda)}_0 \\ &= 0. \end{aligned}$$

Weiterhin folgt aus der dritten Zeile, dass

$$S \Delta x + X \Delta s = -X S \mathbb{1} + \sigma \mu \mathbb{1}.$$

Durch Aufsummieren dieser  $n$  Gleichungen ergibt sich:

$$s^T \Delta x + x^T \Delta s = -x^T s + \sum_{i=1}^n \sigma \frac{x^T s}{n} = -(1 - \sigma)x^T s.$$

Und es gilt:

$$\begin{aligned}
\mu_\alpha &= \frac{(x + \alpha\Delta x)^T (s + \alpha\Delta s)}{n} \\
&= \frac{1}{n} (x^T s + \alpha(s^T \Delta x + x^T \Delta s) + \alpha^2 \Delta x^T \Delta s) \\
&= \frac{x^T s}{n} (1 - \alpha(1 - \sigma)) \\
&= \mu(1 - \alpha(1 - \sigma)) .
\end{aligned}$$

### 3.2.2 Lemma über Reduktion von $\mu$ nach Prediktorschritt

Sei  $(x, \lambda, s) \in \mathcal{N}_2(0.25)$  und  $(\Delta x_{aff}, \Delta \lambda_{aff}, \Delta s_{aff})$  sei die Prediktor-Richtung aus 2(a). Dann gilt  $((x, \lambda, s) + \alpha(\Delta x_{aff}, \Delta \lambda_{aff}, \Delta s_{aff})) \in \mathcal{N}_2(0.5)$  für alle  $\alpha \in [0, \bar{\alpha}]$ , mit  $\bar{\alpha} = \min\left(\frac{1}{2}, \left(\frac{\mu}{8\|\Delta X \Delta S \mathbb{1}\|_2}\right)^{\frac{1}{2}}\right)$ .

**Beweis :**

Als erstes ist zu zeigen, dass gilt:

$$\|X(\alpha)S(\alpha)\mathbb{1} - \mu_\alpha\mathbb{1}\| \leq |1 - \alpha| \|XS\mathbb{1} - \mu\mathbb{1}\| + \alpha^2 \|\Delta X \Delta S \mathbb{1}\| .$$

Mit  $X(\alpha) = \text{diag}(x(\alpha)) = \text{diag}(x + \alpha\Delta x_{aff})$  und analog  $S(\alpha) = \text{diag}(s(\alpha)) = \text{diag}(s + \alpha\Delta s_{aff})$ . Aus dem Gleichungssystem in 2(a) ergibt sich (3.Zeile):

$$\begin{aligned}
x_i(\alpha)s_i(\alpha) - \mu_\alpha &= x_i s_i + \alpha(s_i(\Delta x_{aff})_i + x_i(\Delta s_{aff})_i) \\
&\quad + \alpha^2(\Delta x_{aff})_i(\Delta s_{aff})_i - (1 - \alpha(1 - \sigma))\mu \\
&= x_i s_i(1 - \alpha) + \alpha^2(\Delta x_{aff})_i(\Delta s_{aff})_i - (1 - \alpha)\mu
\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}
\|X(\alpha)S(\alpha)\mathbb{1} - \mu_\alpha\mathbb{1}\| &= \left\| \left( x_i s_i(1 - \alpha) - (1 - \alpha)\mu + \alpha^2(\Delta x_{aff})_i(\Delta s_{aff})_i \right)_{i=1}^n \right\| \\
&\leq |1 - \alpha| \|XS\mathbb{1} - \mu\mathbb{1}\| + \alpha^2 \|\Delta X \Delta S \mathbb{1}\| .
\end{aligned}$$

Nun gilt also :

$$\begin{aligned}
\|X(\alpha)S(\alpha)\mathbb{1} - \mu_\alpha\mathbb{1}\| &\leq (1 - \alpha) \|XS\mathbb{1} - \mu\mathbb{1}\| + \alpha^2 \|\Delta X \Delta S \mathbb{1}\| \\
&\leq (1 - \alpha) \|XS\mathbb{1} - \mu\mathbb{1}\| + \frac{\mu}{8\|\Delta X \Delta S \mathbb{1}\|} \|\Delta X \Delta S \mathbb{1}\| \\
&\leq \frac{1}{4}\mu(1 - \alpha) + \frac{1}{8(1 - \alpha)}\mu(1 - \alpha) \\
&\leq \frac{1}{2}(1 - \alpha)\mu = \frac{1}{2}\mu_\alpha .
\end{aligned}$$

Es bleibt noch zu zeigen, dass  $(x(\alpha), \lambda(\alpha), s(\alpha))$  in  $\mathcal{F}_0$  liegt.

$$\begin{aligned}
Ax(\alpha) &= A(x + \alpha\Delta x_{aff}) = Ax + \alpha A\Delta x_{aff} = b , \\
A^T \lambda(\alpha) + s(\alpha) &= A^T(\lambda + \alpha\Delta \lambda_{aff}) + s + \alpha\Delta s_{aff} \\
&= A^T \lambda + s + \alpha(A^T \Delta \lambda_{aff} + \Delta s_{aff}) \\
&= c .
\end{aligned}$$

Um zu zeigen, dass  $(x(\alpha), s(\alpha))$  echt positiv ist, benötigt man die beiden folgenden Abschätzungen für Vektoren  $(x, \lambda, s)$  aus  $\mathcal{N}_2(\theta)$ , die Beweise kann man in [1] ab Seite 89 nachlesen.

$$\begin{aligned} \| X(\alpha)S(\alpha)\mathbb{1} - \mu_\alpha\mathbb{1} \| &\leq \theta\mu_\alpha \\ \| \Delta X\Delta S\mathbb{1} \| &\leq \frac{\theta^2 + n(1-\sigma)^2}{2^{\frac{3}{2}}(1-\theta)}\mu \end{aligned}$$

Daraus folgt:

$$x_i(\alpha)s_i(\alpha) \geq (1-\theta)\mu_\alpha = (1-\theta)(1-\alpha(1-\sigma))\mu = (1-\theta)(1-\alpha)\mu > 0$$

Dies gilt, da  $\theta = 0.5$  und  $\alpha \in [0, \bar{\alpha}]$ .

$$\begin{aligned} \frac{\mu}{8 \| \Delta X\Delta S\mathbb{1} \|} &\geq \frac{2^{\frac{3}{2}}(1-0.25)}{8((0.25)^2 + n)} = \frac{3\sqrt{2}}{1+16n} \geq \frac{0.16}{n} \\ n \geq 1 \Rightarrow \bar{\alpha} &\geq \min\left(\frac{1}{2}, \left(\frac{0.16}{n}\right)^{\frac{1}{2}}\right) = \frac{0.4}{\sqrt{n}} \end{aligned}$$

### 3.2.3 Lemma über Correctorschritt

Sei  $(x, \lambda, s) \in \mathcal{N}_2(0.5)$  und  $(\Delta x_{cor}, \Delta \lambda_{cor}, \Delta s_{cor})$  sei die Corrector-Richtung aus 2(c). Dann gilt  $((x, \lambda, s) + (\Delta x_{cor}, \Delta \lambda_{cor}, \Delta s_{cor})) \in \mathcal{N}_2(0.25)$  und  $\mu_1 = \mu$ .

**Beweis :**

Aus Lemma 3.2.1 folgt direkt mit  $\alpha = \sigma = 1$ , dass  $\mu_1 = \mu$ . Ebenfalls gilt für  $(x, \lambda, s) \in \mathcal{N}_2(\theta)$ :

$$\begin{aligned} &\| X(\alpha)S(\alpha)\mathbb{1} - \mu_\alpha\mathbb{1} \| \\ &\leq (1-\alpha) \| XS\mathbb{1} - \mu\mathbb{1} \| + \alpha^2 \| \Delta X\Delta S\mathbb{1} \| \\ &\leq (1-\alpha)\theta\mu + \alpha^2 \left( \frac{\theta^2 + n(1-\sigma)^2}{2^{\frac{3}{2}}(1-\theta)} \right) \mu \end{aligned}$$

Nun gilt für  $\theta = 0.5, \alpha = 1$  und  $\sigma = 1$ :

$$\| X(1)S(1)\mathbb{1} - \mu_1\mathbb{1} \| \leq \frac{0.5^2}{2^{\frac{3}{2}}(0.5)} \leq \left(\frac{1}{4}\right)\mu$$

Analog zum Prediktorschritt folgt die strikte Zulässigkeit, da  $\theta = 0.5$ .

$$x_i(\alpha)s_i(\alpha) \geq (1-\theta)\mu_\alpha = (1-\theta)(1-\alpha(1-\sigma))\mu = (1-\theta)\mu > 0$$

### 3.2.4 Lemma über Korrektheit des Algorithmus

Sei  $\varepsilon \in (0, 1)$  gegeben und es gelte  $\mu_{k+1} \leq \left(1 - \frac{\delta}{n^\omega}\right) \mu_k$ ,  $k = 0, 1, 2, \dots$ , mit positiven Konstanten  $\omega$  und  $\delta$ . Ausserdem gelte für den Startpunkt  $(x_0, \lambda_0, s_0)$ , dass  $\mu_0 \leq \frac{1}{\varepsilon^\kappa}$  für eine positive Konstante  $\kappa$ . Dann existiert ein Index  $K$  mit  $K = O(n^\omega |\log \frac{1}{\varepsilon}|)$ , so dass  $\mu_k \leq \varepsilon$  für alle  $k \geq K$ .

#### Beweis :

Durch Logarithmieren gilt :

$$\begin{aligned} \log \mu_{k+1} &\leq \log \left(1 - \frac{\delta}{n^\omega}\right) + \log \mu_k, \\ \log \mu_k &\leq k \log \left(1 - \frac{\delta}{n^\omega}\right) + \log \mu_0 \\ &\leq k \log \left(1 - \frac{\delta}{n^\omega}\right) + \kappa \log \frac{1}{\varepsilon} \end{aligned}$$

Nun gilt für die Logarithmusfunktion :

$$\log(1 + \beta) \leq \beta, \quad \forall \beta > -1.$$

Daraus folgt für  $\beta = -\frac{\delta}{n^\omega}$ :

$$\begin{aligned} \log \mu_k &\leq k \left(-\frac{\delta}{n^\omega}\right) + \kappa \log \frac{1}{\varepsilon} \leq -\log \frac{1}{\varepsilon} = \log \varepsilon \\ \Leftrightarrow k &\geq K = (1 + \kappa) \frac{n^\omega}{\delta} \log \frac{1}{\varepsilon}. \end{aligned}$$

Woraus durch die Monotonie des Logarithmus unmittelbar folgt, dass  $\mu_k \leq \varepsilon$ .

### 3.2.5 Folgerung über Laufzeit

Sei  $\varepsilon > 0$  und  $(x_0, \lambda_0, s_0) \in \mathcal{N}_2(0.25)$  gegeben, dann generiert obiger Predictor-Corrector-Algorithmus eine Folge von Inneren Punkten für die gilt:

$$\mu_{k+1} \leq \left(1 - \frac{0.4}{\sqrt{n}}\right) \mu_k, \quad k = 0, 1, 2, \dots$$

Aus Lemma 3.2.4 folgt somit unmittelbar, dass ein Index  $K$  mit  $K = O(\sqrt{n} |\log \varepsilon|)$  existiert, so dass  $\mu_k \leq \varepsilon$  für alle  $k \geq K$ . Dies alleine führt aber noch nicht zu einem polynomialen Algorithmus, nun kann aber gezeigt werden, dass in einer Umgebung des Optimums das Newtonverfahren konvergiert, d.h. in unserem Fall ein Prediktorschritt der Länge 1 durchgeführt werden kann. Dennoch benutzen die meisten Methoden eine Hilfsprojektion um die Optimallösung zu erraten, wenn dies nicht gelingt, wird einfach der Algorithmus weiter fortgesetzt. Angenommen man löst nun das Gleichungssystem aufgrund der Struktur in  $O(n^2)$  Schritten, so ergibt sich eine Laufzeit von  $O(\sqrt{n} |\log \frac{1}{\varepsilon}| n^2)$ .

### 3.3 Mehrotra's PC-Methode

Heutige Primal Duale Innere Punkte Algorithmen basieren auf Mehrotra's PC-Methode. Dieser vereinigt eben die Grundideen aus Abschnitt 2 und führte weitere sinnvolle meist heuristische Modifikationen ein, die zu sehr guten Ergebnissen in der Praxis führten.

- Berechnung einer weiteren Richtung, den sogenannten corrector step, damit der Zentrale Pfad besser als linear approximiert wird.
- Flexible Wahl des Zentrierungsparameters  $\sigma$ , je nach Erfolg des Predictor-Schrittes.
- Unterschiedlich Schrittlängen für primal und duale Variablen.
- Heuristik für die Wahl des Startpunktes.

#### 3.3.1 Der Algorithmus

1. Berechne Startvektor  $(x_0, \lambda_0, s_0)$  mit  $(x_0, s_0) > 0$ .
2. Für  $k = 0, 1, 2, \dots$  und solange  $\mu > \varepsilon$  :
  - (a) Setze  $(x, \lambda, s) = (x_k, \lambda_k, s_k)$ .
  - (b) Berechne Residualgrößen :  
 $r_b = Ax - b, r_c = A^T \lambda + s - c$
  - (c) Berechne Predictor-Schritt-Richtung durch Lösen des folgendem Gleichungssystems :

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x_{aff} \\ \Delta \lambda_{aff} \\ \Delta s_{aff} \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -XS\mathbb{1} \end{pmatrix}$$

- (d) Berechne Predictor-Schritt-Länge :  
 $\alpha_{aff}^{pri} = \arg \max \{ \alpha \in [0, 1] \mid x_k + \alpha \Delta x_{aff} \geq 0 \}$   
 $\alpha_{aff}^{dual} = \arg \max \{ \alpha \in [0, 1] \mid s_k + \alpha \Delta s_{aff} \geq 0 \}$
- (e) Berechne Centering Parameter :  

$$\sigma = \left( \frac{\mu_{aff}}{\mu} \right)^3 = \left( \frac{(x_k + \alpha_{aff}^{pri} \Delta s_{aff})^T (s_k + \alpha_{aff}^{dual} \Delta s_{aff})}{x_k^T s_k} \right)^3$$
- (f) Berechne Centering-Corrector-Richtung durch Lösen des folgenden Gleichungssystems :

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x_{cc} \\ \Delta \lambda_{cc} \\ \Delta s_{cc} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \sigma \mu \mathbb{1} - \Delta X_{aff} \Delta S_{aff} \mathbb{1} \end{pmatrix}$$

(g) Berechne maximale Schritt-Länge :

$$\alpha_{max}^{pri} = \arg \max\{\alpha \geq 0 \mid x_k + \alpha(\Delta x_{aff} + \Delta x_{cc}) \geq 0\}$$

$$\alpha_{max}^{dual} = \arg \max\{\alpha \geq 0 \mid s_k + \alpha(\Delta s_{aff} + \Delta x_{cc}) \geq 0\}$$

(h) Berechne neue Iterierte :

$$\alpha_k^{pri} = \arg \min\{0.99 \cdot \alpha_{max}^{pri}, 1\}$$

$$\alpha_k^{dual} = \arg \min\{0.99 \cdot \alpha_{max}^{dual}, 1\}$$

$$x_{k+1} = x_k + \alpha_k^{pri}(\Delta x_{aff} + \Delta x_{cc})$$

$$(\lambda_{k+1}, s_{k+1}) = (\lambda_k, s_k) + \alpha_k^{dual}(\Delta \lambda_{aff} + \Delta \lambda_{cc}, \Delta s_{aff} + \Delta s_{cc})$$

### 3.3.2 Zusammenfassung

Es bleibt zu sagen, dass Mehrotra's PC-Methode superlinear konvergiert und im Vergleich zu den anderen veröffentlichten Inneren Punkte Methoden bis zu 60% weniger Iterationen für die Standardbeispiele aus netlib benötigt. Trotzdem ist noch nicht klar, ob es z.B. sinnvoll ist den Mehrotra Corrector Schritt zu iterieren, um eine noch höhere Approximationsgüte des zentralen Pfades zu bekommen. Konkrete Ergebnisse und Implementationen hierzu finden sich in [7].

## Literatur

- [1] *Stephen J. Wright*, Primal-Dual Interior Point Methods, SIAM, Optimization Technology Center, Argonne National Laboratory and Northwestern University, 1996
- [2] *Sanjay Mehrotra*, On the Implementation of primal-dual interior point method, SIAM J. Optimization, November 1992
- [3] *Joe Czyzyk, Sanjay Mehrotra, and Stephen Wright*, PCx User Guide, Technical Report OTC 96/01, Optimization Technology Center, Argonne National Laboratory and Northwestern University, May 1996.
- [4] *Joe Czyzyk, Sanjay Mehrotra, Michael Wagner and Stephen Wright*, PCx An Interior-Point Code for Linear Programming, Technical Report Optimization Meth. & Soft 99/11, Optimization Technology Center, Argonne National Laboratory and Northwestern University, April 1998.
- [5] *Sanjay Mehrotra*, Asymptotic convergence in a generalized predictor-corrector method, Mathematical Programming 74 (1996), Department of Industrial Engineering and Management Sciences, Northwestern University, November 1994.
- [6] *Günter M. Ziegler*, Lectures on Polytopes, Springer Verlag , Technische Universität Berlin, 1998
- [7] *T. J. Carpenter, I. J. Lustig, J. M. Mulvey und D. F. Shanno*, Higher-order predictor-corrector interior points methods with application to quadratic objectives, SIAM Journal on Optimization 3, 1993