Potentialreduktion

Ein Vortrag im Seminar *Numerische Optimierung*
bei Prof. Dr. A. Kirsch und H. Stoll

Andreas Klöckner, am 10. Februar 2003

Inhaltsverzeichnis

1 Grundlagen

Zur Behandlung des hier vorgestellten Verfahrens zur Lösung von linearen Optimierungsaufgaben sind umfangreiche Vorarbeiten nötig. Diese sollen in diesem Abschnitt zum Teil ohne Beweis zusammengestellt werden.

1.1 Definitionen

Definition 1. *Das folgende Problem nennen wir das primale lineare Optimierungsproblem in Normalform*

$$\begin{aligned} (P)\colon \textit{Minimiere } c^T x \\ \textit{unter den Nebenbedingungen } Ax &= b \\ x &\geqslant 0 \end{aligned}$$

mit $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ *und* $b \in \mathbb{R}^m$. *Dazu gehört der zulässige Bereich von (P)*

$$Z_P := \{x \in \mathbb{R}^n \colon Ax = b, x \geqslant 0\}$$

sein Inneres

$$Z_P^\circ := \{x \in \mathbb{R}^n \colon Ax = b, x > 0\}$$

und seine Lösungsmenge

$$\mathcal{L}_P := \{x \in Z_P \colon x \textit{ löst } (P)\}$$

Definition 2. *Das zu (P) duale lineare Optimierungsproblem kann geschrieben werden als*

$$\begin{aligned} (D)\colon \textit{Maximiere } b^T y \\ \textit{unter den Nebenbedingungen } A^T y + s &= c \\ s &\geqslant 0 \end{aligned}$$

wobei s ein Schlupfvariablenvektor ist. $Z_D, Z_D^\circ, \mathcal{L}_D$ *werden analog zu oben definiert.*

Der Notation in [3] folgend, werden wir hier folgende Konventionen verwenden:

- Vektoren oder Skalare werden stets *klein* geschrieben. Superskripte wie x^k bezeichnen stets Iterationsindizes, während Subskripte wie x_k die k-te Komponente des Vektors x darstellen.
- Matrizen werden stets *groß* geschrieben. Für den Vektor x definieren wir die Matrix $X := \operatorname{diag}(x_1, \ldots, x_n)$.

Außerdem sei der Vektor $e := (\,1\ \cdots\ 1\,)^T$.

1.2 Hilfsaussagen

Lemma 3. *(Ungleichung vom arithmetisch-geometrischen Mittel) Seien $x_1, \ldots, x_n \geqslant 0$. Dann gilt*

$$\left(\prod_{i=1}^{n} x_i\right)^{1/n} \leqslant \frac{1}{n} \sum_{i=1}^{n} x_i$$

Beweis. Zunächst sei fesgestellt, dass die Behauptung trivial ist, falls ein $x_i = 0$ oder $n = 1$ ist. Dies sei daher o.B.d.A. nicht der Fall.

Wir beweisen die Behauptung durch vollständige Induktion in einer speziellen Art und Weise.

- *Induktionsanfang, $n = 2$:*

$$\begin{aligned} (\sqrt{x_1} - \sqrt{x_2})^2 &\geqslant 0 \\ x_1 - 2\sqrt{x_1 x_2} + x_2 &\geqslant 0 \\ x_1 + x_2 &\geqslant 2\sqrt{x_1 x_2} \\ \frac{x_1 + x_2}{2} &\geqslant \sqrt{x_1 x_2} \end{aligned}$$

- *Induktionssschritt $2^k \to 2^{k+1}$:* Die Aussage gelte für $n = 2^k$. Mit 2^{k+1} Termen erhalten wir

$$\begin{aligned} & \frac{x_1 + x_2 + \cdots + x_{2^k} + x_{2^k+1} + \cdots + x_{2^{k+1}}}{2^{k+1}} \\ = {} & \frac{\frac{x_1 + x_2 + \cdots + x_{2^k}}{2^k} + \frac{x_{2^k+1} + \cdots + x_{2^{k+1}}}{2^k}}{2} \\ \text{(nach Ind. Anfang)} \geqslant {} & \sqrt{\left(\frac{x_1 + x_2 + \cdots + x_{2^k}}{2^k}\right)\left(\frac{x_{2^k+1} + \cdots + x_{2^{k+1}}}{2^k}\right)} \\ \text{(nach Ind. Annahme)} \geqslant {} & \sqrt{\sqrt[2^k]{x_1 x_2 \cdots x_{2^k}}\sqrt[2^k]{x_{2^k+1} \cdots x_{2^{k+1}}}} \\ = {} & \sqrt[2^{k+1}]{x_1 x_2 \cdots x_{2^{k+1}}}. \end{aligned}$$

  Damit erhalten wir die Behauptung für alle Zweierpotenzen.

- *2. Induktionsschritt $n \to n - 1$:* Die Behauptung gelte für n. Mit der Setzung

$$x_n := \frac{x_1 + \cdots + x_{n-1}}{n-1}$$

  erhalten wir zunächst

$$\begin{aligned} & \frac{x_1 + \cdots + x_{n-1} + \frac{x_1 + \cdots + x_{n-1}}{n-1}}{n} \\ = {} & \frac{(n-1)x_1 + (n-1)x_2 + \cdots + (n-1)x_{n-1} + x_1 + \cdots + x_{n-1}}{(n-1)n} \\ = {} & \frac{n x_1 + n x_2 + \cdots + n x_{n-1}}{(n-1)n} \\ = {} & \frac{x_1 + \cdots + x_{n-1}}{n-1}. \end{aligned}$$

  Es gilt nach Induktionsannahme

$$\begin{aligned} \left(\frac{x_1 + \cdots + x_n}{n}\right)^n &\geqslant x_1 \cdots x_n \\ \left(\frac{x_1 + \cdots + x_{n-1}}{n-1}\right)^{n-1} &\geqslant x_1 \cdots x_n \left(\frac{x_1 + \cdots + x_{n-1}}{n-1}\right)^{-1} \\ \left(\frac{x_1 + \cdots + x_{n-1}}{n-1}\right)^{n-1} &\geqslant x_1 \cdots x_n x_n^{-1}, \end{aligned}$$

was gerade der Behauptung für $n-1$ entspricht.

Damit haben wir die Aussage für sämtliche $n \in \mathbb{N}$ gezeigt. $\square$

Dieser Beweis ist angelehnt in die Darstellung in [1]. Wir fahren fort mit diversen Voraussetzungen und Lemmata, die [3] entnommen sind. Als Konsequenz der *Dualitätssätze* vermerken wir die folgende Charakterisierung ε-optimaler Lösungen, siehe Korollar 2.20 in [3].

Folgerung 4. *Seien $\bar{x} \in Z_P$ und $(\bar{y}, \bar{s}) \in Z_D$ mit einem $\varepsilon \geqslant 0$ gegeben, so dass $0 \leqslant \bar{x}^T \bar{s} \leqslant \varepsilon$ erfüllt ist. Dann ist zunächst $\mathcal{L}_P, \mathcal{L}_D \neq \emptyset$ und mit der Setzung*

$$\begin{aligned} L_P &:= \min_{x \in Z_P} c^T x \\ L_D &:= \max_{(y,s) \in Z_D} b^T y \end{aligned}$$

erhält man

$$\begin{aligned} L_P \leqslant c^T \bar{x} \leqslant L_P + \varepsilon \\ L_D - \varepsilon \leqslant b^T \bar{y} \leqslant L_D. \end{aligned}$$

Ein solches Paar $\bar{x}$ und $(\bar{y}, \bar{s})$ heißt ε-optimales Lösungspaar.

Als nächstes benötigen wir die in dieser Form auch von Daniel Pitsch hergeleitete Formel für die zulässige Abstiegsrichtung innerhalb eines Einheitskreises bezüglich einer elliptischen Norm. Bevor wir die Formel zitieren, bemerken wir, dass für eine positiv definite Matrix $B \in \mathbb{R}^{k \times k}$ die folgende Abbildung

$$\begin{aligned} \|\cdot\|_B \colon \mathbb{R}^k &\to \mathbb{R} \\ x &\mapsto \|x\|_B := \sqrt{x^T B x} \end{aligned}$$

eine Norm definiert, die sog. *elliptische Norm* zu B.

Lemma 5. *Sei $x \in \mathbb{R}^n$ fest. Gegeben sei die Optimierungsaufgabe*

$$\begin{aligned} \textit{Minimiere } \nabla f(x)^T p & \\ \textit{unter den Nebenbedingungen } \|p\|_B &= 1 \\ Ap &= 0 \end{aligned}$$

mit $f \in C^1(\mathbb{R}^n, \mathbb{R})$, $A \in \mathbb{R}^{m \times n}$, $\operatorname{Rang} A = m$, $B \in \mathbb{R}^{n \times n}$ und $p \in \mathbb{R}^n$. Weiterhin gelte

$$\nabla f(x) + A^T y \neq 0$$

für alle $y \in \mathbb{R}^m$. Mit der Setzung

$$P := I - B^{-1} A^T (A B^{-1} A^T)^{-1} A \tag{1}$$

hat das obige Problem die Lösung

$$\tilde{p} := -\frac{P B^{-1} \nabla f(x)}{\|P B^{-1} \nabla f(x)\|_B}. \tag{2}$$

Dies ist Lemma 4.2 aus [3]. Wir wollen uns klarmachen, dass P die Matrix einer orthogonalen Projektion bzgl. des von B erzeugten Skalarprodukts auf den Nullraum von A ist.

Wegen

$$\begin{aligned} AP &= A(I - B^{-1} A^T (A B^{-1} A^T)^{-1} A) \\ &= A - (A B^{-1} A^T)(A B^{-1} A^T)^{-1} A \\ &= A - A \\ &= 0 \end{aligned}$$

liegt das Bild von P im Nullraum von A. Außerdem zeigt

$$\begin{aligned} P^2 &= (I - B^{-1}A^T(AB^{-1}A^T)^{-1}A)^2 \\ &= I^2 - 2B^{-1}A^T(AB^{-1}A^T)^{-1}A + B^{-1}A^T(AB^{-1}A^T)^{-1}(\mathrm{AB}^{-1}A^T)(AB^{-1}A^T)^{-1}A \\ &= I^2 - 2B^{-1}A^T(AB^{-1}A^T)^{-1}A + B^{-1}A^T(AB^{-1}A^T)^{-1}A \\ &= P, \end{aligned}$$

dass P eine Projektion ist. Um zu zeigen, dass P eine *orthogonale Projektion* bezüglich des von B erzeugten Skalarprodukt ist, nehmen wir an, dass $x \in \mathrm{Kern}\, A$ gilt und y beliebig ist:

$$\begin{aligned} \langle x, y - Py \rangle_B &= x^T B(y - Py) \\ &= x^T B(y - y + B^{-1}A^T(AB^{-1}A^T)^{-1}Ay) \\ &= x^T BB^{-1}A^T(AB^{-1}A^T)^{-1}Ay \\ &= x^T A^T(AB^{-1}A^T)^{-1}Ay \\ &= 0 \end{aligned} \tag{3}$$

Lemma 6. *Gegeben sei die Optimierungsaufgabe aus Lemma 5. Für diese gilt*

$$\begin{aligned} & \exists y \in \mathbb{R}^m \colon \nabla f(x) + A^T y = 0 \\ \Leftrightarrow\ & PB^{-1}\nabla f(x) = 0. \end{aligned}$$

Beweis. $\exists y \in \mathbb{R}^m \colon \nabla f(x) + A^T y = 0 \Leftrightarrow \nabla f(x) \in \mathrm{Bild}\, A^T \Leftrightarrow \nabla f(x) \in (\mathrm{Kern}\, A)^\perp \Leftrightarrow B^{-1}\nabla f(x) \in \mathrm{Kern}\, A^{\perp_B} \Leftrightarrow PB^{-1}\nabla f(x) = 0.$ □

Dies ist Lemma 4.5 in [3]. Seine Aussage ist auch intuitiv einleuchtend: Genau dann, wenn $\nabla f(x)$ auf Kern A senkrecht steht, wird es von der (orthogonalen) Projektion P auf Null abgebildet, denn P projiziert ja gerade senkrecht auf Kern A.

Zum Schluss zitieren wir noch ein einfaches Resultat zur Zulässigkeit, das Korollar 4.7 in [3] entspricht.

Lemma 7. *Seien $x \in \mathbb{R}^n_+$ und $p \in \mathbb{R}^n$. Es ist $x + \beta p > 0$, falls $\|p\|_{X^{-2}} = 1$ und $\beta \in (0,1)$.*

2 Die Idee

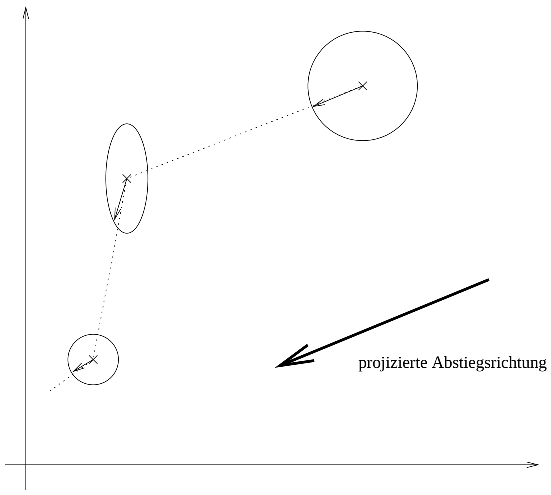
In diesem Abschnitt werden wir möglichst ohne Verwirrung durch zu viele technische Details versuchen, die Grundidee hinter dem hier vorgestellten Potentialreduktionsverfahren anschaulich zu machen. Abschnitt 3, auf den wir hier öfters verweisen, liefert die hier fortgelassenen Details, unter Benutzung der in Abschnitt 1 bereitgestellten Grundlagen.

2.1 Affine Scaling

Der im Vortrag von Daniel Pitsch dargestellte *Affine Scaling Algorithmus* [2] hat viele wünschenswerte Eigenschaften. Er ist einfach zu verstehen, mit geringem Aufwand zu implementieren und hinreichend effizient. Jedoch ist es bis heute nicht geglückt, eine polynomielle Laufzeitschranke für diesen Algorithmus zu beweisen. Beobachtungen zeigen, dass das Verfahren, wenn es einmal dem Rand zu nahe kommt, von diesem nicht mehr wegkommt, und in der Folge ein ähnliches Laufzeitverhalten wie der Simplex-Algorithmus an den Tag legt – nahezu exponentiell.

Um den insbesondere durch die Nebenbedingung $x \geqslant 0$ bestimmten Rand von Z_P zu vermeiden, verwendet der Affine Scaling Algorithmus das folgende Verfahren, das für eine polynomielle Laufzeit vermutlich nicht stark genug vom Rand weg "drückt": Mit den Bezeichnungen aus (P) stelle man sich den auf den affinen Unterraum $Ax = b$ projizierten Einheitskreis bezüglich der Norm $\|\cdot\|_{X^{-2}}$ vor. Auf diesem Einheitskreis wird nun das x mit dem minimalen Wert von $c^T x$ gesucht, die Richtung desselben ausgehend von der gegenwärtigen Iterierten ist die für diesen Schritt verwendete Abstiegsrichtung. (Liegt die nächste Iterierte in diesem Kreis oder auf seinem Rand, so spricht man von einer "Short-Step", andernfalls von einer "Long-Step"-Methode.)

Die folgende Skizze soll die Zusammenhänge verdeutlichen.



(Die Anschauungsebene stelle den durch Ax=b bestimmten affinen Unterraum dar)

2.2 Besser als Affine Scaling

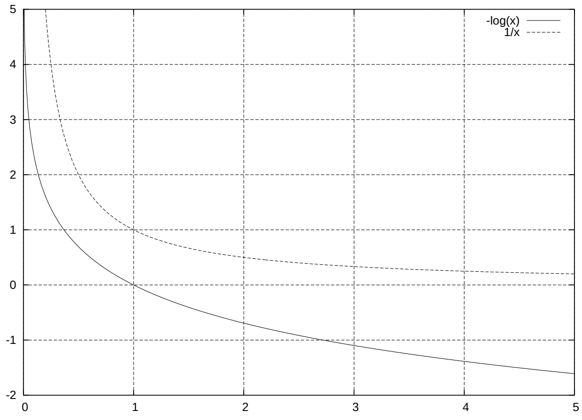
Um ein besseres Laufzeitverhalten als der Affine-Scaling-Algorithmus zu erzielen, soll hier die von Meline Seibold [4] eingeführte logarithmische *Barrierefunktion* verwendet werden, durch welche vermieden werden soll, dass man bei der iterativen Lösung des Problems dem Rand zu nahe kommt. Mit den Definitionen aus Abschnitt 1 sei nun $q > n$ beliebig und $(x, s) > 0$. Dann definieren wir

$$F(x, s) := q \ln(x^T s) - \sum_{i=1}^{n} \ln x_i - \sum_{i=1}^{n} \ln s_i. \tag{4}$$

An dieser so genannten Potentialfunktion werden wir uns auf dem Weg zu einer Lösung orientieren, sie stellt ein "Potentialgebirge" dar, an dem wir auf der Suche nach einer Lösung "hinunterrutschen" können. Warum und wie zeigt nun F den Weg zu einer Lösung auf?

Untersuchen wir F genauer, so erkennen wir im Term $x^T s$ die *Dualitätslücke*, die auf dem Weg zu einer Lösung gegen Null (Folgerung 4) und somit logarithmiert gegen $-\infty$ geht. Die beiden letzten Terme sind eine Art Barriere dafür, dass wir zu schnell an den von $x \geqslant 0$ vorgegebenen Rand kommen – sie gehen für x bzw. $s \to 0$ gegen $+\infty$. Die gegen $+\infty$ bzw. $-\infty$ gehenden Terme stehen sozusagen in einem Wettbewerb, und es bleibt zu zeigen, dass der Dualitätslücken-Term tatsächlich gewinnen kann.

Die Logarithmus-Funktion geht bei 0 relativ "kurz und schmerzlos" gegen $-\infty$, weswegen sie zur Vermeidung von Nullen in den Vektoren x und s sehr gut taugt, ohne dabei den "eigentlichen" Funktionswert zu sehr zu stören. Die folgende Abbildung macht das deutlich, insbesondere im Vergleich mit $\frac{1}{x}$.



Wir werden ein Verfahren konstruieren, das den Funktionswert von F für jede Iterierte um mindestens ein festes $\delta > 0$ verkleinert. Hätten wir ein solches Verfahren, würde uns Satz 8 aus Abschnitt 3 zu jedem $\varepsilon > 0$ eine Schrittanzahl liefern, nach der wir spätestens eine ε-optimale Lösung hätten. Damit wissen wir auch, dass F uns tatsächlich den Weg zu einer Lösung zeigen kann, d.h. der Dualitätslücken-Term kann den Wettbewerb tatsächlich gewinnen.

2.3 Die Iterationsvorschrift

Für die Iteration sei eine Schrittweite $\beta \in (0,1)$ fest vorgegeben. Da wir zur Lösung des primalen Problems hauptsächlich an x interessiert sind, versuchen wir zunächst, x so zu verändern, dass sich der Wert von F verringert. Dazu berechnen wir den Gradienten von F nach x:

$$\nabla_x F(x,s) = \begin{pmatrix} q\frac{s_1}{x^Ts} - \frac{1}{x_1} \\ q\frac{s_2}{x^Ts} - \frac{1}{x_2} \\ \vdots \\ q\frac{s_n}{x^Ts} - \frac{1}{x_n} \end{pmatrix} = \frac{q}{x^Ts}s - X^{-1}e \tag{5}$$

Angenommen, wir seien nun mitten in der Iteration. Es seien $x^k \in Z_P^\circ$ und $(y^k,\, s^k) \in Z_D^\circ$ gegeben. Wie man solche Startwerte findet, klärt später Abschnitt 4.

Wir wollen nun den Gradienten $\nabla_x F(x,\, s)$ in den Nullraum von A projizieren, um ihn zu einer zulässigen Abstiegsrichtung im Sinne von (P) zu machen. Lemma 5 hilft uns dabei, es ist aber leider nur anwendbar, falls $\nabla_x F(x^k,\, s^k) + A^T y \neq 0$ für alle $y \in \mathbb{R}^m$. Lemma 6 identifiziert dies als denjenigen Fall, wo mit den Bezeichnungen von Lemma 5 $PB^{-1}\nabla_x F(x,\, s) = 0$ gilt, d.h. der Gradient auf den Nullvektor projiziert wird. Also können wir zunächst beruhigt losprojizieren und uns um diesen Fall später kümmern.

Sorglos wenden wir also Lemma 5 an. Dieses liefert uns eine Richtung, die einen im Sinne von (P) zulässigen Gradientenabstieg realisiert. Wie beim Affine-Scaling-Verfahren wählen wir als die die elliptische Norm $\|\cdot\|_{B_k}$ erzeugende Matrix $B_k = X_k^{-2}$. Wir erhalten

$$\begin{aligned} P_k &= I - B_k^{-1}A^T(AB_k^{-1}A^T)^{-1}A \\ &= I - X_k^2 A^T (AX_k^2A^T)^{-1}A \end{aligned}$$

und als zulässige Abstiegsrichtung (der Iterationsindex k an p, P, B, X, x und s ist hier aus Gründen der Lesbarkeit fortgelassen):

$$\begin{aligned} p &= -\frac{PB^{-1}\nabla_x F(x,s)}{\|PB^{-1}\nabla F(x,s)\|_B} \\ &= -\frac{PX^2\nabla_x F(x,s)}{\|PX^2\nabla_x F(x,s)\|_{X^{-2}}} \\ &= -\frac{PX^2\nabla_x F(x,s)}{\|X^{-1}PX^2\nabla_x F(x,s)\|} \\ &= -X\frac{X^{-1}PX^2\nabla_x F(x,s)}{\|X^{-1}PX^2\nabla_x F(x,s)\|} \end{aligned}$$

Mit der Setzung

$$u^k := X_k^{-1}P_kX_k^2\nabla_x F(x^k,s^k) \tag{6}$$

können wir p_k umschreiben als

$$p^k = -X_k\frac{u^k}{\|u^k\|_2}. \tag{7}$$

Als nächste Iterierte wählen wir

$$\begin{aligned} x^{k+1} &:= x^k + \beta p^k \\ y^{k+1} &:= y^k \\ s^{k+1} &:= s^k. \end{aligned}$$

Erste Sorge muss dabei sein, ob diese Iterierten überhaupt zulässig sind. Die Bedingung $Ax = b$ ist sicher erfüllt, schließlich haben wir p^k mit viel Aufwand in den Nullraum von A projiziert. Dass auch $x^{k+1} > 0$ gilt, folgt nach Konstruktion der Iterierten aus Lemma 7.

Wie oben bemerkt haben wir aber ein Problem, falls $PB^{-1}\,\nabla F_x(x,\,s) = 0$, d.h. in unserem Falle $u^k = 0$ ist. Überhaupt ist es nicht günstig, wenn $\|u^k\|$ kleine Werte annimmt.

Man kann nämlich $\|u_k\|$ als ein Maß dafür ansehen, wie sehr der Funktionswert von F durch die obige Iterierte verringert würde. Denn $\|u^k\|$ wird beeinflusst von der Länge von ∇F, d.h. der eigentlichen Steigung der Funktion F in x, bzw. wieviel davon nach einer Projektion noch übrig ist. Die Multiplikation mit X^{-1} bewirkt, dass wir die Länge der Projektion in der von X^{-2} erzeugten Norm messen. Dies führt dazu, dass wir Fortschritte in eine Richtung, in der wir ohnehin schon nahe am Rand sind, geringer bewerten. Ist nun $\|u^k\|$ kleiner als eine gewisse positive Zahl, nennen wir sie γ, so können wir ahnen, dass unsere Iterationsvorschrift wohl "nicht viel bringt". Lemma 9 garantiert uns einen Fortschritt in Bezug auf den Wert von F auch nur in Bezug auf ein solches positives γ – sogar mit einem direkten Zusammenhang: Je größer γ ist, um so größer der Schritt. Für die Fälle $\|u^k\| < \gamma$ und insbesondere für $u^k = 0$ müssen wir uns also nach einem Notstopfen umsehen.

Nachdem wir also offenbar in eine Sackgasse laufen können, wenn wir ausschließlich x verändern, liegt es nahe, auch an y und s herumzudrehen. Die neuen y und s müssen natürlich zulässig sein. Betrachten wir daher die folgende Gleichung (die Indizes k lassen wir für den Moment mal weg):

$$\begin{aligned} u &= X^{-1}PX^2\nabla_x F(x,s) \\ &= X^{-1}PX^2\Big(\frac{q}{x^Ts}s - X^{-1}e\Big) \\ &= X^{-1}PX\Big(\frac{q}{x^Ts}\mathrm{Xs} - e\Big) \\ &= X^{-1}(I - X^2A^T(AX^2A^T)^{-1}A)X\Big(\frac{q}{x^Ts}\mathrm{Xs} - e\Big) \\ &= (I - XA^T(AX^2A^T)^{-1}AX)\Big(\frac{q}{x^Ts}\mathrm{Xs} - e\Big) \\ &= \frac{q}{x^Ts}\mathrm{Xs} - e - XA^T(AX^2A^T)^{-1}AX\Big(\frac{q}{x^Ts}\mathrm{Xs} - e\Big) \end{aligned}$$

das heißt

$$\begin{aligned} u - \frac{q}{x^Ts}\mathrm{Xs} + e + XA^T(AX^2A^T)^{-1}AX\Big(\frac{q}{x^Ts}\mathrm{Xs} - e\Big) &= 0 \\ XA^T(AX^2A^T)^{-1}AX\Big(\frac{q}{x^Ts}\mathrm{Xs} - e\Big) + u - \frac{q}{x^Ts}\mathrm{Xs} + e &= 0 \quad \Big|\cdot X^{-1} \\ A^T(AX^2A^T)^{-1}AX\Big(\frac{q}{x^Ts}\mathrm{Xs} - e\Big) + u - \frac{q}{x^Ts}s + e &= 0 \quad \Big|\cdot \frac{x^Ts}{q} \\ A^T(AX^2A^T)^{-1}AX\left(\mathrm{Xs} - \frac{x^Ts}{q}e\right) + \frac{x^Ts}{q}(u+e) - s &= 0. \end{aligned}$$

Setzen wir also

$$\begin{aligned} x^{k+1} &:= x^k \\ y^{k+1} &:= y^k + (AX^2A^T)^{-1}AX\left(\mathrm{Xs} - \frac{x^Ts}{q}e\right) \\ s^{k+1} &:= \frac{x^Ts}{q}(u+e), \end{aligned}$$

so wissen wir zumindest jetzt, dass

$$A^T(y^{k+1} - y^k) + s^{k+1} - s^k = 0,$$

und wir haben somit die eine Hälfte der Zulässigkeit. Es bleibt zu zeigen, dass $s^{k+1} > 0$ ist. Schränken wir γ auf das Intervall $(0, 1)$ ein, so können wir folgendermaßen schließen: nach Voraussetzung ist $x^T s/q > 0$, $\|u\| < \gamma < 1$, d.h. $u + e > 0$, d.h. $s > 0$. Somit gilt $(y^k, s^k) \in Z_D^\circ$. Lemma 10 zeigt uns schließlich, dass wir auch hier bezüglich des Funktionwertes von F einen Schritt nach vorne gemacht haben.

Der hier vorgestellte Algorithmus hat also in jedem Schritt die Wahl, in Abhängigkeit von $\|u^k\|$ entweder x oder y und s zu verändern. Man sagt auch, der Algorithmus mache entweder einen *primalen* oder einen *dualen* Schritt. Satz 11 zeigt uns, dass für eine bestimmte Wahl von β und γ und unabhängig davon, ob wir einen primalen oder dualen Schritt gemacht haben, der Wert von F um mindestens 0.079 schrumpft. Daher liefert uns Satz 8 die folgende maximale Schrittanzahl K, nach der wir spätestens eine ε-optimale Iterierte haben:

$$K := \left\lceil \frac{F(x^0, s^0) + \sqrt{n} \ln \frac{1}{\varepsilon} - n \ln n}{0.079} \right\rceil$$

K wächst offenbar höchstens linear in n. Dazu besteht jeder Schritt noch aus einigen (aber jeweils höchstens konstant vielen) Matrixmultiplikationen, die jeweils eine Laufzeit von $O(n^3)$ haben. Der Gesamtaufwand, gerechnet in Multiplikationen von zwei Skalaren, ist insofern höchstens polynomiell.

Zum Abschluss der ganze Algorithmus auf einen Blick:

1. Man gehe aus von $x^0 \in Z_P^\circ$, $(y^0, s^0) \in Z_D^\circ$. Gegeben sei weiterhin ein $\varepsilon > 0$. Setze $q := n + \sqrt{n}$, $\beta := 0.285$, $\gamma := 0.479$ und $k := 0$.

2. Ist $(x^k)^T s^k \leqslant \varepsilon$, so sind x^k und (y^k, s^k) nach Folgerung 4 ε-optimal, also STOP.

3. Berechne

$$\begin{aligned} g^k &:= \frac{q}{(x^k)^T s^k} s^k - X_k^{-1} e = \nabla_x F(x^k, s^k) \\ h^k &:= (A X_k^2 A^T)^{-1} A X_k^2 g^k \\ u^k &:= X_k g^k - (A X_k)^T h^k. \end{aligned}$$

4. Ist $\|u^k\| \geqslant \gamma$, so mache einen primalen Schritt:

$$\begin{aligned} p^k &:= -X_k \frac{u^k}{\|u^k\|} \\ x^{k+1} &:= x^k + \beta p^k \\ s^{k+1} &:= s^k \\ y^{k+1} &:= y^k, \end{aligned}$$

   inkrementiere k und gehe nach 2.

5. Andernfalls mache einen dualen Schritt:

$$\begin{aligned} x^{k+1} &:= x^k \\ s^{k+1} &:= \frac{(x^k)^T s^k}{q} X_k^{-1}(u^k + e) \\ y^{k+1} &:= y^k + \frac{(x^k)^T s^k}{q} h^k, \end{aligned}$$

inkrementiere k und gehe nach 2.

3 Die Details

Bevor wir unsere Ideen exakt formulieren können, müssen wir mit den Definitionen von Kapitel 1 noch die folgenden Voraussetzungen als gegeben annehmen:

$$\begin{aligned}\text{(Rang) Rang } A &= m\\ \text{(InnP) } Z_P^\circ &\neq \emptyset\\ \text{(InnD) } Z_D^\circ &\neq \emptyset\end{aligned}$$

Wir zeigen als Erstes, dass es hinreichend ist, F in jedem Iterationsschritt um ein festes δ zu verkleinern, um nach einer endlichen (berechenbaren) Anzahl von Schritten eine ε-optimale Lösung zu erhalten.

Satz 8. *Seien die Folgen* $(x^k)_{k=0}^\infty \subset Z_P^\circ$*,* $(y^k, s^k)_{k=0}^\infty \subset Z_D^\circ$ *sowie ein* $\delta > 0$ *und ein* $\varepsilon > 0$ *gegeben, so dass mit der Definition von* F *aus (4) gilt*

$$F(x^{k+1}, s^{k+1}) - F(x^k, s^k) \leqslant -\delta. \tag{8}$$

für alle $k \in \mathbb{N}_0$*. Setze*

$$K := \left\lceil \frac{F(x^0, s^0) + (q-n)\ln\frac{1}{\varepsilon} - n\ln n}{\delta} \right\rceil. \tag{9}$$

Dann sind alle x_k *und* (y_k, s_k) *mit* $k \geqslant K$ *im Sinne von Folgerung 4* ε*-optimal.*

Beweis. Sei K wie in (9) definiert. Dann liefert (8) als eine Teleskopsumme

$$F(x^K, s^K) - F(x^0, s^0) \leqslant -K\delta.$$

Aus (9) folgt

$$K \geqslant \frac{F(x^0, s^0) + (q-n)\ln\frac{1}{\varepsilon} - n\ln n}{\delta}.$$

Sei nun $k \geqslant K$. Dann gilt

$$\begin{aligned}F(x^k, s^k) &\leqslant F(x^K, s^K) \leqslant F(x^0, s^0) - K\delta\\ &\leqslant F(x^0, s^0) - \frac{F(x^0, s^0) + (q-n)\ln\frac{1}{\varepsilon} - n\ln n}{\delta}\delta\\ &= -(q-n)\ln\frac{1}{\varepsilon} + n\ln n.\end{aligned}$$

Jetzt beweisen wir eine weitere Abschätzung für F. Seien dazu $x \in Z_P^\circ$ und $(y, s) \in Z_D^\circ$.

$$\begin{aligned}F(x, s) &= q\ln(x^T s) - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln s_i\\ &= (q-n)\ln(x^T s) + n\ln(x^T s) - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln s_i\\ &= (q-n)\ln(x^T s) + \ln\left(n^n\left(\frac{1}{n}\sum_{i=1}^n x_i s_i\right)^n\right) - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln s_i\\ \text{(Lemma 3)} &\geqslant (q-n)\ln(x^T s) + \ln\left(n^n \prod_{i=1}^n x_i s_i\right) - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln s_i\\ &= (q-n)\ln(x^T s) + n\ln n + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i - \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln s_i\\ &= (q-n)\ln(x^T s) + n\ln n.\end{aligned}$$

Zusammenfassend können wir schließen

$$\begin{aligned}(q-n)\ln((x^k)^T s^k)+n\ln n &\leqslant F(x^k,s^k)\leqslant -(q-n)\ln\frac{1}{\varepsilon}+n\ln n\\ (q-n)\ln((x^k)^T s^k) &\leqslant (q-n)\ln\varepsilon\\ \ln((x^k)^T s^k) &\leqslant \ln\varepsilon\\ (x^k)^T s^k &\leqslant \varepsilon,\end{aligned}$$

was gerade die ε-Optimalität des Paares x^k, s^k ist. □

Wie in Abschnitt 2.3 beschrieben, besteht unser Verfahren aus *primalen* und *dualen* Schritten. Für beide Arten von Schritten muss ein δ wie in (8) gefunden werden, dies geschieht in Lemma 9 und 10. Diese δs werden wir in Satz 11 mit einer speziellen Wahl von β und γ zu einer Schranke vereinen.

Lemma 9. *Seien $x^k\in Z_P^\circ$, $(y^k,s^k)\in Z_D^\circ$ und ein $\gamma>0$ sowie ein $\beta\in(0,1)$ gegeben. Es sei u^k wie in* (6)*, und es gelte $\|u^k\|_2\geqslant\gamma$. Mit p^k wie in* (7) *sind dann die Iterierten*

$$\begin{aligned}x^{k+1} &:= x^k+\beta p^k\\ y^{k+1} &:= y^k\\ s^{k+1} &:= s^k\end{aligned}$$

im Inneren der zulässigen Bereiche, und es gilt die Schranke

$$F(x^{k+1},s^{k+1})-F(x^k,s^k)\leqslant -\beta\gamma+\frac{\beta^2}{2(1-\beta)}. \tag{10}$$

Beweis. Die Zulässigkeit der Iterierten ist bereits in Abschnitt 2.3 gezeigt worden, d.h. es gilt genauer sogar $x^{k+1}\in Z_P^\circ$. Zunächst beweisen wir einige Hilfsresultate.

Es gilt $1+a\leqslant e^a$ für $a\in\mathbb{R}$. Daher haben wir auch

$$\ln(1+a)\leqslant a. \tag{11}$$

Mit Hilfe der Taylorentwicklung von $\ln(1+a)$ und der geometrischen Reihe gewinnt man die folgende Abschätzung:

$$\begin{aligned}\ln(1+a) &= a-\frac{a^2}{2}+\frac{a^3}{3}-\cdots\\ &\geqslant a-\frac{|a|^2}{2}-\frac{|a|^3}{3}-\cdots\\ &\geqslant a-\frac{|a|^2}{2}(1+|a|+|a|^2+|a|^3+\cdots)\\ &= a-\frac{|a|^2}{2}\cdot\frac{1}{1-|a|}\\ &= a-\frac{|a|^2}{2(1-|a|)}.\end{aligned}$$

Mit Hilfe von (2) wissen wir außerdem, dass wegen $p\in\operatorname{Kern}A$

$$\begin{aligned}\left\langle -X^2\nabla_x F(x,s)+PX^2\nabla_x F(x,s),p\right\rangle_{X^{-2}} &= 0\\ (-X^2\nabla_x F(x,s)+\mathrm{P}X^2\nabla_x F(x,s))^T X^{-2}p &= 0\\ (-X^2\nabla_x F(x,s))^T X^{-2}p+(\mathrm{P}X^2\nabla_x F(x,s))^T X^{-2}p &= 0\\ (-\nabla_x F(x,s))^T p+(\mathrm{P}X^2\nabla_x F(x,s))^T X^{-2}p &= 0\\ (\mathrm{P}X^2\nabla_x F(x,s))^T X^{-2}p &= \nabla_x F(x,s)^T p.\end{aligned}$$

Mit

$$\begin{aligned}u &= X^{-1}PX^2\nabla_x F(x,s)\\ \Leftrightarrow Xu &= PX^2\nabla_x F(x,s)\end{aligned}$$

und $p = -X\frac{u}{\|u\|}$ erhalten wir also

$$\begin{aligned}-\nabla_x F(x,s)^T p &= -(\mathrm{PX}^2\nabla_x F(x,s))^T X^{-2} p\\ &= -(Xu)^T X^{-2} p\\ &= -u^T X^{-1} p\\ &= u^T X^{-1} X \frac{u}{\|u\|}\\ &= \|u\|.\end{aligned}$$

Wir können außerdem sagen, dass

$$\left|\beta\frac{p_i}{x_i}\right| = \left|\beta\frac{x_i\frac{u_i}{\|u\|}}{x_i}\right| = \left|\beta\frac{u_i}{\|u\|}\right| \leqslant \beta.$$

Mit diesem geballten Vorwissen gehen wir an die finale Abschätzung, wobei wir an x, s, p, X den Iterationsindex k weglassen:

$$\begin{aligned}F(x^{k+1}, s^{k+1}) - F(x,s) &= q\ln((x+\beta p)^T s) - \sum_{i=1}^n \ln(x_i + \beta p_i) - q\ln(x^T s) + \sum_{i=1}^n \ln x_i\\ &= q\ln\left(\frac{(x+\beta p)^T s}{x^T s}\right) - \sum_{i=1}^n \ln\left(\frac{x_i + \beta p_i}{x_i}\right)\\ &= q\ln\left(1 + \beta\frac{p^T s}{x^T s}\right) - \sum_{i=1}^n \ln\left(1 + \beta\frac{p_i}{x_i}\right)\\ (11)\quad &\leqslant q\beta\frac{p^T s}{x^T s} - \sum_{i=1}^n \ln\left(1 + \beta\frac{p_i}{x_i}\right)\\ &\leqslant q\beta\frac{p^T s}{x^T s} - \sum_{i=1}^n \left(\beta\frac{p_i}{x_i} - \frac{|\beta\frac{p_i}{x_i}|^2}{2(1 - |\beta\frac{p_i}{x_i}|)}\right)\\ &\leqslant q\beta\frac{p^T s}{x^T s} - \sum_{i=1}^n \left(\beta\frac{p_i}{x_i} - \frac{|\beta\frac{p_i}{x_i}|^2}{2(1-\beta)}\right)\\ &= q\beta\frac{p^T s}{x^T s} - \beta(X^{-1}e)^T p + \sum_{i=1}^n \left(\frac{|\beta\frac{p_i}{x_i}|^2}{2(1-\beta)}\right)\\ &= \beta\left(q\frac{s}{x^T s} - X^{-1}e\right)^T p + \frac{\beta^2\|X^{-1}p\|^2}{2(1-\beta)}\\ &= \beta\nabla_x F(x,s)^T p + \frac{\beta^2}{2(1-\beta)}\\ &= -\beta\|u\| + \frac{\beta^2}{2(1-\beta)}.\end{aligned}$$

Ist nun $\|u\| \geqslant \gamma$, so folgt das Gewünschte. □

Es gibt auch eine Schranke für den dualen Schritt.

Lemma 10. *Seien $x^k \in Z_P^\circ$, $(y^k, s^k) \in Z_D^\circ$ und ein $\gamma \in (0,1)$ gegeben. Es sei u^k wie in (6), und es gelte $\|u^k\|_2 < \gamma$. Dann liegen die Iterierten*

$$\begin{aligned}x^{k+1} &:= x^k\\ y^{k+1} &:= y^k + \frac{(x^k)^T s^k}{q}(AX_k^2 A^T)^{-1} A X_k^2 \nabla_x F(x^k, s^k)\\ s^{k+1} &:= \frac{(x^k)^T s^k}{q} X_k^{-1}(u^k + e)\end{aligned}$$

im Innern der zulässigen Bereiche, und es gilt die Schranke

$$F(x^{k+1}, s^{k+1}) - F(x^k, s^k) \leqslant -(q-n) + \ln\left(\frac{q}{n}\right) + \frac{\gamma^2}{2(1-\gamma)} \tag{12}$$

Beweis. Dass die Iterierten im Inneren der zulässigen Bereiche liegen, zeigte bereits die Argumentation in Abschnitt 2.3. Es gilt, wiederum mit fortgelassenem Iterationsindex k:

$$\begin{aligned}
& F(x^{k+1}, s^{k+1}) - F(x,s) \\
= \;& q\ln\frac{x^T s^{k+1}}{x^T s} - \sum_{i=1}^n \ln(s_i^{k+1}) + \sum_{i=1}^n \ln s_i \\
= \;& q\ln\frac{x^T s^{k+1}}{x^T s} - \sum_{i=1}^n \ln\left(\frac{x^T s}{q x_i}(1+u_i)\right) + \sum_{i=1}^n \ln s_i \\
= \;& q\ln\frac{x^T s^{k+1}}{x^T s} - n\ln\frac{x^T s}{q} - \sum_{i=1}^n \ln(1+u_i) + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i \\
\leqslant \;& q\ln\frac{x^T s^{k+1}}{x^T s} - n\ln\frac{x^T s}{q} - \sum_{i=1}^n \left(u_i - \frac{u_i^2}{2(1-|u_i|)}\right) + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i \\
= \;& q\ln\frac{x^T s^{k+1}}{x^T s} - n\ln\frac{x^T s}{q} - e^T u + \sum_{i=1}^n \left(\frac{u_i^2}{2(1-|u_i|)}\right) + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i \\
(\|u\| < \gamma) \quad \leqslant \;& q\ln\frac{x^T s^{k+1}}{x^T s} - n\ln\frac{x^T s}{q} - e^T u + \frac{\gamma^2}{2(1-\gamma)} + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i \\
= \;& q\ln\frac{x^T \frac{x^T s}{q} X^{-1}(u+e)}{x^T s} - n\ln\frac{x^T s}{q} - e^T u + \frac{\gamma^2}{2(1-\gamma)} + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i \\
= \;& q\ln\frac{x^T}{q} X^{-1}(u+e) - n\ln\frac{x^T s}{q} - e^T u + \frac{\gamma^2}{2(1-\gamma)} + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i \\
= \;& q\ln\frac{e^T u + n}{q} + n\ln q - \ln\left(n^n \left[\frac{1}{n}\sum_{i=1}^n x_i s_i\right]^n\right) - e^T u + \frac{\gamma^2}{2(1-\gamma)} + \sum_{i=1}^n \ln x_i + \\
& \sum_{i=1}^n \ln s_i \\
(\text{Lemma 3}) \quad \leqslant \;& q\ln\frac{e^T u + n}{q} + n\ln q - \ln\left(n^n \prod_{i=1}^n x_i s_i\right) - e^T u + \frac{\gamma^2}{2(1-\gamma)} + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i \\
= \;& q\ln\frac{e^T u + n}{q} + n\ln q - n\ln n + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln s_i - e^T u + \frac{\gamma^2}{2(1-\gamma)} + \sum_{i=1}^n \ln x_i + \\
& \sum_{i=1}^n \ln s_i \\
= \;& q\ln\frac{e^T u + n}{q} + n\ln\frac{q}{n} - e^T u + \frac{\gamma^2}{2(1-\gamma)} \\
= \;& q\ln\left(1 - \frac{q - e^T u - n}{q}\right) + n\ln\frac{q}{n} - e^T u + \frac{\gamma^2}{2(1-\gamma)} \\
\leqslant \;& -q + e^T u + n + n\ln\frac{q}{n} - e^T u + \frac{\gamma^2}{2(1-\gamma)} \\
= \;& -(q-n) + n\ln\frac{q}{n} + \frac{\gamma^2}{2(1-\gamma)}.
\end{aligned}$$

Dies war die Behauptung. □

Schließlich fassen wir beide Schranken zu einer zusammen.

Satz 11. *Seien $x^k \in Z_P^\circ$, $(y^k, s^k) \in Z_D^\circ$. Wählt man*

$$\begin{aligned}
q &:= n + \sqrt{n} \\
\beta &:= 0.285 \\
\gamma &:= 0.479
\end{aligned}$$

sowie die $x^{k+1}, y^{k+1}, s^{k+1}$ wie in Lemma 9 und 10, so gilt

$$F(x^{k+1}, s^{k+1}) - F(x^k, s^k) \leqslant -0.079.$$

Beweis. Zuerst brauchen wir die Abschätzung

$$-\sqrt{n}+n\ln\left(1+\frac{1}{\sqrt{n}}\right)\leqslant -0.3. \tag{13}$$

Mit dem Satz von Taylor wissen wir, dass für $a\in(-1,1)$ und ein gewisses $\alpha\in(-1,1)$ gilt

$$\ln(1+a)=a-\frac{a^2}{2}+\frac{a^3}{3}-\frac{\alpha^4}{4}.$$

Daher

$$\ln(1+a)\leqslant a-\frac{a^2}{2}+\frac{a^3}{3}.$$

Ist nun $a=\frac{1}{A}$ und $|A|>1$, so gilt

$$\begin{aligned}
\ln(1+a) &\leqslant a-\frac{a^2}{2}+\frac{a^3}{3}\\
\ln\left(1+\frac{1}{A}\right) &\leqslant \frac{1}{A}-\frac{1}{2A^2}+\frac{1}{3A^3}\\
A^2\ln\left(1+\frac{1}{A}\right) &\leqslant A-\frac{1}{2}+\frac{1}{3A}\\
-A+A^2\ln\left(1+\frac{1}{A}\right) &\leqslant -\frac{1}{2}+\frac{1}{3A}.
\end{aligned}$$

Wir wollen (13), also mit $A=\sqrt{n}$,

$$-\sqrt{n}+n\ln\left(1+\frac{1}{\sqrt{n}}\right)\leqslant -\frac{1}{2}+\frac{1}{3\sqrt{n}}\overset{!}{<}-0.3$$

Dies führt auf

$$\begin{aligned}
-\frac{1}{2}+\frac{1}{3\sqrt{n}} &< -0.3\\
\Leftrightarrow \frac{1}{3\sqrt{n}} &< 0.2\\
\Leftrightarrow \frac{1}{\sqrt{n}} &< 3\cdot 0.2\\
\Leftrightarrow \sqrt{n} &> \frac{1}{3\cdot 0.2}\\
\Leftrightarrow n &> \left(\frac{1}{3\cdot 0.2}\right)^2\approx 2.777...
\end{aligned}$$

Das heißt, mit $n\geqslant 3$ sind wir auf der sicheren Seite und (13) gilt. Für $n=1,2$ lässt sich die Gültigkeit einfach nachrechnen:

$$\begin{aligned}
-\sqrt{1}+\ln\left(1+\frac{1}{\sqrt{1}}\right) &= -0.3068...\\
-\sqrt{2}+2\ln\left(1+\frac{1}{\sqrt{2}}\right) &= -0.34461...
\end{aligned}$$

Setzen wir nun $q:=n+\sqrt{n}$. Dann ergibt sich mit Lemma 10

$$\begin{aligned}
F(x^{k+1},s^{k+1})-F(x^k,s^k) &\leqslant -(q-n)+n\ln\frac{q}{n}+\frac{\gamma^2}{2(1-\gamma)}\\
&= -(n+\sqrt{n}-n)+n\ln\frac{n+\sqrt{n}}{n}+\frac{\gamma^2}{2(1-\gamma)}\\
&\leqslant -\sqrt{n}+n\ln\left(1+\frac{1}{\sqrt{n}}\right)+\frac{\gamma^2}{2(1-\gamma)}\\
&= -0.3+\frac{\gamma^2}{2(1-\gamma)}.
\end{aligned}$$

Lemma 9 sagt, dass

$$F(x^{k+1},s^{k+1})-F(x^k,s^k)\leqslant -\beta\gamma+\frac{\beta^2}{2(1-\beta)}.$$

Mit der Wahl $\beta := 0.285$ und $\gamma := 0.479$ ergibt sich dann für Lemma 9

$$-\beta\gamma + \frac{\beta^2}{2(1-\beta)} \approx -0.13651 + \frac{0.08129}{2 \cdot 0.715} \approx -0.079.$$

Lemma 10 schließlich ergibt

$$-0.3 + \frac{0.22944}{2 \cdot 0.521} \approx -0.079.$$

Dies zeigt das gewünschte für den dualen wie den primalen Schritt. □

4 Starten mit großen M's

Der oben vorgestellte Algorithmus geht davon aus, dass Iterierte bekannt sind, die bereits im Inneren des zulässigen Bereiches sind. Falls solche Startwerte nicht zufällig auf der Straße liegen, hat man es zuweilen schwer, sie zu finden. Im Austausch gegen einige zusätzliche Dimensionen hilft aber wie schon beim Affine-Scaling-Algorithmus die so genannte Big-M-Methode.

Dazu betrachtet man die folgenden Hilfsprobleme:

$$\begin{aligned} (PM)\colon \text{ Minimiere } & c^T x + M_1 x_{n+1} \\ \text{unter den Nebenbedingungen } Ax + (b - Ae)x_{n+1} &= b \\ (e-c)^T x + x_{n+2} &= M_2 \\ (x, x_{n+1}, x_{n+2}) &\geqslant 0 \end{aligned}$$

und

$$\begin{aligned} (DM)\colon \text{ Maximiere } & b^T y + M_2 y_{m+1} \\ \text{unter den Nebenbedingungen } A^T y + (e-c) y_{m+1} + s &= c \\ (b - Ae)^T y + s_{n+1} &= M_1 \\ y_{m+1} + s_{n+2} &= 0 \\ (s, s_{n+1}, s_{n+2}) &\geqslant 0 \end{aligned}$$

mit “großen” Konstanten $M_1, M_2 > 0$. Man startet dann mit den Vektoren

$$\begin{aligned} (x^0, x^0_{n+1}, x^0_{n+2}) &:= (e, 1, M_2 - (e-c)^T e) > 0 \\ (y^0, y^0_{m+1}, s^0, s^0_{n+1}, s^0_{n+2}) &:= (0, -1, e, M_1, 1), \end{aligned}$$

deren Zulässigkeit bezüglich (PM) und (DM) man leicht nachrechnet. Wenn M_1, M_2 “hinreichend groß” sind, sieht man leicht, dass für eine optimale Lösung $x_{n+1} = y_{m+1} = 0$ gelten muss, und somit x und y die auch die ursprünglichen Bedingungen aus (P) und (D) erfüllen müssen.

Literaturverzeichnis

[1] Peter Bruin. http://planetmath.org/encyclopedia/ProofOfGeneralMeansInequality.html.

[2] Daniel Pitsch. Der affine scaling algorithmus. 2002.

[3] Rembert Reemtsen. Lineare optimierung.

[4] Meline Seibold. Pfadverfolgungsmethoden. 2003.