ANGEWANDTE MATHEMATIK UND INFORMATIK

UNIVERSITÄT ZU KÖLN

Report No. 96.223

Semidefinite Programmierung

by

W. Hochstättler

1996

W. Hochstättler
Universität zu Köln
Weyertal 86-90
D–50931 Köln
Germany
Telefon (0221) 470–6012
e–Mail WH@ZPR.UNI-KOELN.DE

1991 Mathematics Subject Classification: 90C10, 90C25, 90C27, 15A18, 68R10
Keywords: semidefinite programming, interior point methods, eigenvalue optimization, combinatorial optimization, maximum cliques,perfect graphs, graph partitioning

Semidefinite Programmierung

Winfried Hochstättler
frei nach dem Skriptum
des Kurses von F. Alizadeh
im RUTCOR Januar, Februar 1995

19. Juli 1996

Inhaltsverzeichnis

Einleitung

Dies ist das Skriptum des zweiten Teils der Vorlesung über „Nichtlineare Optimierung“ von Professor Dr. A. Bachem. Der erste Teil behandelte die Grundzüge des Standardstoffs, wie er etwa in dem Buch von D.G. Luenberger „Introduction to Linear and Nonlinear Programming (Second Edition)“ Addison-Wesley Publishing Company Reading 1984, vorgestellt wird. Wir hatten eine Zweiteilung der Vorlesung geplant, um in der zweiten Hälfte Stoff präsentieren zu können, der näher an der aktuellen Forschung ist.

Dabei kam die Idee auf, semidefinite Programmierung zu besprechen. Da wir in diesem Gebiet selbst nicht arbeiten und die Inhalte noch zu neu sind, als daß bereits hinreichend Literatur für die Lehre existieren würde, entschlossen wir uns erst, als wir im Internet Skripten eines Kurses von Farid Alizadeh zu diesem Thema entdeckten, den zweiten Teil der Vorlesung mit diesen Inhalten zu bestreiten.

Mit dem Wechsel von Professor Dr. A. Bachem in den Vorstand der DLR habe ich die Aufarbeitung der Skripten und die Vorstellung des Stoffes in der Vorlesung übernommen. Da es sich bei der Vorlage um einen Kurs handelt, der im Rahmen eines Ph.D. Programms abgehalten wurde, sind die Originalskripten manchmal etwas knapp gehalten. Deshalb möchte ich den Zuhörern unseres Kurses mit dieser deutschen Ausarbeitung der Vorlage weitere Literatur zur Verfügung stellen.

Auch für einen Leser, der den ersten Teil der Vorlesung nicht gehört hat, sollte das Skriptum verständlich sein. Die wenigen Stellen, an denen auf diesen Stoff Bezug genommen wird, sollten für ein Gesamtverständnis nicht wesentlich sein. Die zugehörigen Zusammenhänge kann man in verschiedenen Standardwerken nachlesen.

Mir persönlich hat es viel Spaß gemacht, einen Einblick in diese Theorie zu gewinnen, und ich hoffe, daß der eine oder andere Leser beides mit mir teilen kann.

Köln, den 20. Februar 1996 Winfried Hochstättler

Kapitel 1

Motivation

1.1 Die Sandwichzahl

Wir zeigen zu Beginn anhand eines Beispiels aus der Graphentheorie, wie semidefinite Programme auftreten können. Sei $G = (V, E)$ ein einfacher, ungerichteter Graph mit Knoten V, Kanten E und A seine Adjazenzmatrix, d.h. $A = (a_{ij})_{i,j} \in \{0, 1\}^{|V|\times|V|}$ mit $a_{ij} = 1$ falls (i, j) eine Kante ist und $a_{ij} = 0$ sonst. Wir betrachten eine Matrix B ähnlich wie diese und zwar

$$B = (b_{ij})_{i,j} = \begin{cases} 1 & \text{falls } i = j \text{ oder } (i, j) \text{ eine Kante ist} \\ x_{\{i,j\}} & \text{sonst.} \end{cases}$$

Mit $b_{ij} = x_{\{i,j\}}$ wollen wir zum Ausdruck bringen, daß B symmetrisch ist, da der Graph ungerichtet ist. Also ist $B = B(x)$ ein Ausdruck in $\frac{1}{2}(|V|^2 - |V| - 2|E|)$ Variablen. Betrachten wir nun die Funktion

$$\theta(G) = \min_x \lambda_1(B(x)),$$

wobei $\lambda_1(B)$ der größte Eigenwert von B ist. Die Zahl $\theta(G)$ heißt *Lovász- oder Sandwich-Zahl von* G.

Definition 1.1.1 *Sei A eine reelle, symmetrische $(n \times n)$-Matrix. Die Eigenwerte von B bezeichnen wir mit $\lambda_1 \geq \ldots \geq \lambda_{|V|}$. Falls uns die Ordnung der Eigenwerte nach Absolutbeträgen interessiert, so notieren wir sie mit Superskript also $|\lambda^1| \geq \ldots \geq |\lambda^{|V|}|$.*

Bemerkung 1.1.1 *Wir schreiben bei den Optimierungsaufgaben stets* $\max$ *und* $\min$*, auch wenn diese nicht immer angenommen werden.*

Einen sehr schönen Übersichtsartikel zur Sandwichzahl hat kürzlich Donald Knuth geschrieben (The Electronic Journal of Combinatorics (1994)).

Die Zahl $\theta(G)$ heißt Sandwichzahl, da sie zwischen der Cliquezahl und der chromatischen Zahl des Graphen G liegt.

Definition 1.1.2 *Sei $G = (V, E)$ ein Graph. Eine Teilmenge $K \subseteq V$ heißt Clique, wenn für alle Knoten $i, j \in K : (i, j) \in E$ eine Kante ist. Die* Cliquezahl *ist dann $\omega(G) := \max\{|K| \mid K$ ist Clique in $G\}$. Eine Knotenfärbung ist eine Abbildung $c : V \to \{1, \ldots, k\}$ mit $(i, j) \in E \Rightarrow c(i) \neq c(j)$. Die* chromatische Zahl *ist dann $\chi(G) = \min\{k \mid$ es gibt eine Knotenfärbung $c : V \to \{1, \ldots, k\}\}$.*

Satz 1.1.1 *Sei G ein Graph. Dann gilt $\omega(G) \leq \theta(G) \leq \chi(G)$.*

Beweis: Sei $K \subseteq V$ eine Clique in G mit $|K| = \omega(G)$ und $\mathbb{1}_K$ der charakteristische Vektor von K. Dann ist $\mathbb{1}_K^\top B(x) \mathbb{1}_K = \omega^2$. Wegen $\|\mathbb{1}_K\|_2 = \sqrt{\omega}$ erhalten wir somit für beliebiges x

$$\lambda_1(B(x)) = \max_{\|y\|_2 = 1} y^\top B(x) y \geq \omega.$$

Für die andere Ungleichung wählen wir eine Färbung $c : V \to \{1, \ldots, \chi(G)\}$ und betrachten die Matrix

$$C = (c_{ij})_{i,j} = \begin{cases} 1 & \text{falls } i = j \text{ oder } c(i) \neq c(j) \\ 1 - \chi(G) & \text{sonst.} \end{cases}$$

Dann ist offensichtlich $\theta \leq \lambda_1(C) =: \lambda$. Wir zeigen $\lambda \leq \chi(G)$ mittels Induktion über $k = |V| - \chi(G)$.

Ist $k = 0$, so ist C die Matrix aus lauter 1en und hat nur einen von 0 verschiedenen Eigenwert, nämlich $|V| = \chi(G)$.

Sei also $\chi := \chi(G) < |V|$. Wir dürfen annehmen, daß $(i, j) \in E \Leftrightarrow c(i) \neq c(j)$. Ferner seien die Knoten von G so numeriert, daß $c(1) = c(2)$. Sei nun y ein Eigenvektor von C mit $Cy = \lambda y$. Da die Knoten 1 und 2 nicht adjazent sind und zu einem Knoten $1 \neq i \neq 2$ entweder beide adjazent oder beide nicht adjazent sind, gibt es eine Konstante c mit:

$$\lambda y_1 = y_1 + (1 - \chi) y_2 + c \tag{1.1}$$
$$\lambda y_2 = (1 - \chi) y_1 + y_2 + c \tag{1.2}$$

Daraus folgern wir:

$$\begin{aligned} (\lambda-1)y_1+(\chi-1)y_2 &= (\lambda-1)y_2+(\chi-1)y_1 \\ \Rightarrow \qquad (\lambda-\chi)(y_1-y_2) &= 0. \end{aligned}$$

Da wir zeigen wollen $\lambda \leq \chi$, können wir also annehmen, daß $y_1 = y_2$. Für die Konstante in den Gleichungen (1.1) und (1.2) gilt somit: $c = (\lambda+\chi-2)y_1$.

Wir betrachten nun den Graphen $\tilde{G} := G \setminus \{1\}$, bei dem wir den Knoten mit der Nummer 1 gelöscht haben. Offensichtlich ist $\chi(\tilde{G}) = \chi(G)$. Konstruiert man ferner die zugehörige Matrix $\tilde{C}$ wie oben, so stellt man fest, daß $\tilde{C}$ aus C durch streichen der ersten Spalte und Zeile entsteht. Nach Induktionsvoraussetzung ist dann $\lambda_1(\tilde{C}) = \chi(\tilde{G}) = \chi$. Betrachten wir nun den Vektor $\tilde{y} = (2y_2, y_3, \ldots, y_{|V|})^\top$, so erhalten wir:

$$\begin{aligned} \tilde{C}\tilde{y} &= (2y_2+c, \lambda y_3, \ldots, \lambda y_{|V|}) && (1.3) \\ &= ((\lambda+\chi)y_2, \lambda y_3, \ldots \lambda y_{|V|}). && (1.4) \end{aligned}$$

Andererseits haben wir

$$\chi\|\tilde{y}\|_2^2 \geq \tilde{y}^\top \tilde{C}\tilde{y} = 2(\lambda+\chi)y_2^2 + \lambda\sum_{i=3}^{|V|} y_i^2 \qquad (1.5)$$

und somit

$$0 \geq 2(\lambda-\chi)y_2^2 + (\lambda-\chi)\sum_{i=3}^{|V|} y_i^2, \qquad (1.6)$$

woraus wegen $y \neq 0$ die Behauptung folgt. □

Donald Knuth schreibt in seinem Übersichtsartikel über θ in den Preliminaries: „Therefore I think it is worthwhile to study θ closely, in hopes of getting aquainted with it and finding faster ways to compute it.“

Und im letzten Abschnitt: „The book (Grötschel, Lovász, Schrijver) explains how to compute θ within a given tolerance ϵ in polynomial time using an ellipsoid method, but that method is too slow and numerically unstable to deal with graphs that have 10 or so vertices. Fortunately, however, new “interior point methods” have been developed for this purpose, especially Alizadeh has computed θ when G has hundreds of vertices and thousand of edges.“

Kapitel 2

Wiederholung aus der Linearen Algebra

2.1 Matrizen

Definition 2.1.1 *Sei A eine $(n \times n)$-Matrix über einem Körper K. Eine Zahl $\lambda \in K$ heißt* Eigenwert *von A, wenn es ein $x \in K^n \setminus \{0\}$ gibt mit $Ax = \lambda x$. Jeder solche Vektor $x \neq 0$ heißt* Eigenvektor von A zum Eigenwert λ. *Ist A eine Matrix mit komplexwertigen Einträgen, so bezeichnen wir mit A^* die zu A transponierte, konjugierte Matrix. Ist $A = A^*$, so nennen wir A* hermitesch. *Eine Matrix Q heißt* orthogonal, *wenn $Q^\top Q = QQ^\top = I$.*

Eine reellwertige Matrix ist hermitesch genau dann, wenn sie symmetrisch ist.

Lemma 2.1.1 *Hermitesche Matrizen haben nur reelle Eigenwerte.*

Beweis: $Ax = \lambda x \Rightarrow x^* A^* = x^* A = \bar{\lambda} x^* \Rightarrow \lambda x^* x = x^* A^* x = \bar{\lambda} x^* x \Rightarrow \lambda = \bar{\lambda}$.

□

Satz 2.1.1 *Ist A eine reellwertige, symmetrische Matrix, dann gibt es eine orthogonale Matrix Q, so daß $QAQ^\top = D$ eine Diagonalmatrix ist.*

Beweis: vgl. Lineare Algebra II. □

Wir nennen $A = Q^\top DQ$ die *Hauptachsentransformation* von A, da die Eigenwerte von A die Diagonaleinträge von D sind und die Eigenvektoren die Zeilen von Q. Ist D eine Diagonalmatrix und y der Vektor der Diagonaleinträge, so schreiben wir auch $D = Diag(y)$.

Satz 2.1.2 (Rayleigh-Ritz) *Ist A eine symmetrische Matrix, so gilt*

$$\begin{aligned} \lambda_1(A) &= \max_{\|x\|_2=1} x^\top Ax \\ \lambda_n(A) &= \min_{\|x\|_2=1} x^\top Ax \end{aligned}$$

Beweis: Sei $A = Q^\top DQ$ die Hauptachsentransformation von A. Dann ist folglich $x^\top Ax = (Qx)^\top DQx$. Da Q orthogonal ist, gilt mit $Qx = (v_1, \ldots, v_n)^\top = v$ und $\|x\|$=1 auch $\|v\| = 1$. Seien $d_1, \ldots, d_n$ die Diagonaleinträge von D also gerade die Eigenwerte von A. Dann erhalten wir: $\lambda_n(A) = \min d_i \leq x^\top Ax = \sum_{i=1}^n d_i v_i^2 \leq \max d_i = \lambda_1(A)$. Da λ_n, bzw. λ_1 für normierte, zugehörige Eigenvektoren angenommen werden, folgt die Behauptung. $\square$

Korollar 2.1.1 *Eine reellwertige, symmetrische Matrix ist positiv semidefinit genau dann, wenn der kleinste Eigenwert $\lambda_n \geq 0$ ist.*

Schließlich erinnern wir noch daran, daß man aus positiv semidefiniten Matrizen Wurzeln ziehen kann.

Lemma 2.1.2 *Sei X eine positiv semidefinite, symmetrische Matrix. Dann gibt es genau eine positiv semidefinite, symmetrische Matrix $X^{1/2}$ mit $X^{1/2}X^{1/2} = X$.*

Beweis: Sei $X = Q^\top DQ$ die Hauptachsentransformation von X. Da D eine Diagonalmatrix mit nicht-negativen Einträgen ist, existiert seine Wurzel $D^{1/2}$. Dann erfüllt $Q^\top D^{1/2}Q$ die Behauptung. Hat man andererseits ein Y, das die Behauptung erfüllt, so ist $QYQ^\top QYQ^\top = D$ und somit $Y = X^{1/2}$. $\square$

Kapitel 3

Kegel und Kegeldualität

3.1 Polare Kegel

Definition 3.1.1 *Eine Menge $K \subseteq \mathbb{R}^n$ heißt* Kegel*, wenn* $\forall x \in K\, \forall \alpha \geq 0 : \alpha x \in K$. *Ein Kegel ist* konvex*, wenn* $x, y \in K \Rightarrow x + y \in K$*, er heißt* spitz*, wenn* $K \cap (-K) = \{0\}$. *Ist C eine konvexe Menge, so heißt $x \in C$* extremal*, wenn aus $y, z \in K$ und $x \in conv(\{y, z\})$ notwendig $x \in \{y, z\}$ folgt. Ist K ein Kegel und $x \in K \setminus \{0\}$, so heißt der Strahl $\{\alpha x \mid \alpha \geq 0\}$ ein* Extremalstrahl*, wenn aus $x = y + z$ und $y, z \in K$ notwendig $y = \alpha x$ und $z = \beta x$ folgt.*

Ist K ein Kegel, so ist $K^* := \{y \in \mathbb{R}^n \mid \forall x \in K : x^\top y \geq 0\}$ der zu K polare Kegel.

Ist K ein Kegel, so ist $\dim(K)$ die Dimension von K *definiert als die Dimension des kleinsten linearen Raumes, der K enthält. Ein Kegel $K \subseteq \mathbb{R}^n$ heißt* volldimensional*, wenn $\dim(K) = n$.*

In den Übungen wurde nachgerechnet, daß die positiv semidefiniten, symmetrischen Matrizen einen spitzen, konvexen Kegel bilden.

Beispiel 3.1.1 *Beispiele abgeschlossener, konvexer Kegel:*

1. *Der positive Orthant: $\{x \in \mathbb{R}^n \mid \forall i : x_i \geq 0\}$ ist selbstpolar,*
2. *Das Eishörnchen: $\{x \in \mathbb{R}^n \mid x_1 \geq \sqrt{x_2^2 + \ldots + x_n^2}\}$ ist selbstpolar,*
3. *p-Kegel $\{x \in \mathbb{R}^n \mid x_1 \geq \sqrt[p]{|x_2|^p + \ldots + |x_n|^p}\}$, sind nur für $p \geq 1$ konvex. Polar zu einem p-Kegel ist der q-Kegel mit $\frac{1}{p} + \frac{1}{q} = 1$ (Stichwort Höldersche Ungleichung).*

4. *Der Kegel der positiv semidefiniten, symmetrischen Matrizen (als Teilmenge des* $\mathbb{R}^{\binom{n+1}{2}}$*) ist selbstpolar. Genauer betrachten wir folgende Korrespondenz zwischen* $\mathbb{R}^{\binom{n+1}{2}}$ *und der Menge der symmetrischen Matrizen*

$$(x_{11}, \ldots, x_{nn}, x_{\{1,2\}}, \ldots, x_{\{n-1,n\}}) \leftrightarrow \begin{pmatrix} x_{11} & \frac{x_{\{1,2\}}}{\sqrt{2}} & \cdots & \frac{x_{\{1,n\}}}{\sqrt{2}} \\ \frac{x_{\{1,2\}}}{\sqrt{2}} & x_{22} & \cdots & \frac{x_{\{2,n\}}}{\sqrt{2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{\{1,n\}}}{\sqrt{2}} & \frac{x_{\{2,n\}}}{\sqrt{2}} & \cdots & x_{nn} \end{pmatrix}$$

Dann entspricht der Menge der positiv semidefiniten, symmetrischen Matrizen ein spitzer, abgeschlossener Kegel im $\mathbb{R}^{\binom{n+1}{2}}$*, der selbstpolar ist, wie in den Übungen nachgerechnet wurde.*

Da es sich bei der Transformation nur um eine Skalierung der Variablen handelt, kann man üblicherweise direkt auf der Menge der positiv semidefiniten, symmetrischen Matrizen als Teilmenge des $\mathbb{R}^n$ *argumentieren. Wir werden an den entsprechenden Stellen näher darauf eingehen.*

3.2 Kegelordnungen und allgemeine Optimierungsprobleme

Ist $K \subseteq \mathbf{R}^n$ ein spitzer, konvexer Kegel, so können wir eine Partialordnung auf $\mathbf{R}^n$ definieren vermöge

$$a \leq_K b \Leftrightarrow b - a \in K.$$

Da $0 \in K$ ist, folgt Reflexivität, da K spitz ist, erhalten wir Antisymmetrie, und die Konvexität liefert die Transitivität.

Ist $\mathcal{K}$ die durch den Kegel der positiv semidefiniten Matrizen auf den symmetrischen Matrizen induzierte Partialordung, die sogenannte *Löwner-Partialordung*, so schreiben wir statt $\leq_{\mathcal{K}}$ auch $\preceq$.

Damit können wir folgendes allgemeine Optimierungsproblem formulieren:

$$\begin{array}{ll} \text{Primales Problem:} & \\ \min & c^\top x \\ \text{unter} & Ax = b \\ & x \geq_K 0 \end{array}$$

Wir können nun hierzu ein duales Optimierungsproblem angeben.

Duales Problem:

$\max$	$u^\top b$
unter	$A^\top u + s = c$
	$s \geq_{K^*} 0$

3.3 Schwache Dualität

Lemma 3.3.1 (Schwache Dualität) *Ist x zulässig für das primale Problem und sind u und s zulässig für das duale Problem, so ist $c^\top x \geq u^\top b$.*

Beweis:

$$
\begin{aligned}
c^\top x - u^\top b &= c^\top x - u^\top Ax && (3.1)\\
&= (c^\top - u^\top A)x && (3.2)\\
&= s^\top x && (3.3)\\
&\geq 0 && (3.4)
\end{aligned}
$$

□

Ist x primal zulässig und sind u und s dual zulässig, so nennt man $s^\top x$ das *duality gap*. Falls das duality gap 0 ist, so sind x, u und s offensichtlich Optimallösungen ihrer Programme. Ist K der positive Orthant, so erhalten wir ein duales Paar linearer Programme.

3.4 Das verallgemeinerte Farkas' Lemma

Verschwindet das duality gap, so sprechen wir von einer starken Dualität. Zum Beweis der starken Dualität für Kegelprogramme benötigen wir eine schwächere Version der Aussage aus Aufgabe 4 in den Übungen.

Lemma 3.4.1 (Trennsatz) *Sei $K \subseteq \mathbb{R}^n$ ein abgeschlossener, konvexer Kegel und $x \notin K$. Dann gibt es eine lineare Hyperebene, die x von K trennt, d.h. $\exists c \in \mathbb{R}^n$: $c^\top x < 0$ und $\forall y \in K : c^\top y \geq 0$.*

Beweis: Folgt aus Aufgabe 4. □

Damit können wir nun eine verallgemeinerte Version von Farkas' Lemma beweisen.

Lemma 3.4.2 (Erweitertes Farkas' Lemma) *Sei $K \subseteq \mathbb{R}^n$ ein konvexer, spitzer, abgeschlossener, volldimensionaler Kegel, $b \in \mathbb{R}^m$, A eine $(m \times n)$-Matrix und $\mathcal{P} = \{Ax \mid x \in K\}$. Dann gilt:*

entweder *b ist im topologischen Abschluß von $\mathcal{P}$*

oder *es gibt ein $y \in \mathbb{R}^m$, so daß $A^\top y \geq_{K^*} 0$ und $b^\top y < 0$,*

aber nicht beides.

Beweis: Die beiden Alternativen schließen sich aus, da ansonsten $0 \leq y^\top b = b^\top y < 0$ folgen würde. Als Bild eines konvexen Kegels unter einer linearen Abbildung ist $\mathcal{P}$ ein konvexer Kegel, sein Abschluß also ein konvexer, abgeschlossener Kegel. Ist nun b nicht im topologischen Abschluß von $\mathcal{P}$, so gibt es nach Lemma 3.4.1 ein y mit $y^\top b < 0$ und $y^\top z \geq 0$ für alle $z \in \mathcal{P}$. Dann gilt aber für alle $x \in K : y^\top Ax \geq 0$ und somit $A^\top y \geq_{K^*} 0$. □

Bemerkung 3.4.1 *Beachte, daß für einen nicht volldimesionalen Kegel K der polare Kegel K^* nicht mehr spitz ist, deswegen keine Partialordnung mehr definiert. Um die Alternativen zu formulieren, ist deshalb notwendig, daß K volldimensional ist.*

Ist $\mathcal{P}$ abgeschlossen, so erhalten wir ein Korollar, das formal ähnlicher zu Farkas' Lemma ist:

Korollar 3.4.1 (Verallgemeinertes Farkas' Lemma) *Sei $K \subseteq \mathbb{R}^n$ ein konvexer, spitzer, abgeschlossener, volldimensionaler Kegel, $b \in \mathbb{R}^m$, A eine $(m \times n)$-Matrix und $\mathcal{P} = \{Ax \mid x \in K\}$ abgeschlossen. Dann gilt:*

entweder *es gibt ein $x \in \mathbb{R}^n$, so daß $Ax = b$ und $x \geq_K 0$,*

oder *es gibt ein $y \in \mathbb{R}^m$, so daß $A^\top y \geq_{K^*} 0$ und $b^\top y < 0$,*

aber nicht beides.

□

Das verallgemeinerte Farkas' Lemma gilt im allgemeinen nicht, wenn $\mathcal{P}$ nicht abgeschlossen ist.

Beispiel 3.4.1 *Sei $\mathcal{K}$ der Kegel der positiv semidefiniten, symmetrischen, reellen* (2×2) *Matrizen, betrachtet bzgl. der Korrespondenz in Beispiel* (4). *Das System*

$$\begin{aligned} &x_1 = 0 \\ &x_3 = 1 \\ &(x_1, x_2, x_3) \geq_{\mathcal{K}} 0 \end{aligned}$$

ist inkonsistent, denn es gibt kein x_2 derart, daß $\begin{pmatrix} 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & x_2 \end{pmatrix}$ *positiv semidefinit ist.*

Andererseits ist auch das System

$$\begin{aligned} &(y_1, 0, y_2) \geq_{\mathcal{K}} 0 \\ &y_2 < 0 \end{aligned}$$

inkonsistent.

Abschließend wollen wir noch eine hinreichende Bedingung dafür angeben, daß $\mathcal{P}$ abgeschlossen ist. Diese Bedingung ist bekannt als *Slater constraint qualification*.

Lemma 3.4.3 *Gibt es ein y, so daß $A^\top y \in$ int(K^*) (wir schreiben dafür auch kurz $A^\top y >_{K^*} 0$), so ist $\mathcal{P} = \{Ax \mid x \in K\}$ abgeschlossen.*

Beweis: Da $A^\top y >_{K^*} 0$ ist, gibt es ein $\varepsilon > 0$ mit $\forall x \in K \setminus \{0\} : (\varepsilon \frac{-x}{\|x\|} + A^\top y)^\top x \geq 0$ also $\forall x \in K \setminus \{0\} : y^\top Ax > 0$. Wir schließen daraus: ker$(A) \cap K = \{0\}$.

Sei nun $(x_i)_{i \in \mathbb{N}}$ eine Folge in K mit $\lim_{i\to\infty} Ax_i = z \in \mathbb{R}^n \setminus \{0\}$. Eventuell durch Übergang zu einer Teilfolge dürfen wir annehmen, daß $\lim_{i\to\infty} \frac{x_i}{|x_i|} = x$ existiert. Da K abgeschlossen ist, gilt $x \in K$ und somit $Ax \neq 0$. Also erhalten wir $z = \lim_{i\to\infty} |x_i| A \frac{x_i}{|x_i|}$. Also existiert auch $\lambda := \lim_{i\to\infty} |x_i| = \frac{|z|}{|Ax|}$ und somit ist $z = A(\lambda x) \in \mathcal{P}$. □

3.5 Starke Dualität

Satz 3.5.1 (Starker Dualitätssatz) *Sei A eine Matrix, K ein spitzer, abgeschlossener, volldimensionaler, konvexer Kegel und $\tilde{\mathcal{P}} = \left\{ \binom{c^\top}{A} x \mid x \geq_K 0 \right\}$ abgeschlossen. Haben das primale und das dualeProgramm endliche Optimalwerte z_1 und z_2, so gilt: $z_1 = z_2$.*

Beweis: Nach dem schwachen Dualitätssatz gilt $z_1 \geq z_2$. Angenommen $z_1 > z_2$, also $z_2 < \inf\{c^\top x \mid Ax = b,\ x \in K\}$. Demnach ist das System

$$\begin{aligned} c^\top x &= z_2 \\ Ax &= b \\ x &\in K \end{aligned}$$

unzulässig.

Nach dem verallgemeinerten Farkas' Lemma gibt es also ein α und ein y, so daß

$$\alpha c + A^\top y \geq_{K^*} 0 \text{ und } y^\top b + \alpha z_2 < 0. \tag{3.5}$$

Wir unterscheiden 3 Fälle.

1. $\alpha = 0$. Dann ist $y^\top A \geq_{K^*} 0$ und $y^\top b < 0$ und somit nach dem erweiterten Farkas' Lemma das primale Problem unzulässig.

2. $\alpha > 0$. Dann erhalten wir aus Gleichung (3.5)

$$c^\top - \frac{-y}{\alpha}^\top A \geq_{K^*} 0 \text{ und } z_2 - \frac{-y}{\alpha}^\top b < 0. \tag{3.6}$$

   Also ist $\frac{-y}{\alpha}$ zulässig für das duale Programm und besser als die Optimallösung.

3. $\alpha < 0$.Wiederum aus (3.5) erhalten wir:

$$-c^\top + \frac{-y}{\alpha}^\top A \geq_{K^*} 0 \text{ und } \frac{1}{2}\left((\frac{-y}{\alpha})^\top b - z_2\right) < 0. \tag{3.7}$$

   Da z_2 Optimalwert des dualen Programms ist und $(\frac{-y}{\alpha})^\top b < z_2$, gibt es ein u_0 mit $c^\top - u_0^\top A \geq_{K^*} 0$ und $u_0^\top b > \frac{1}{2}(z_2 + (\frac{-y}{\alpha})^\top b)$. Also erhalten wir:

$$(\frac{-y}{\alpha} - u_0)^\top A \geq_{K^*} 0 \text{ und } (\frac{-y}{\alpha} - u_0)^\top b < 0 \tag{3.8}$$

   und aus dem erweiterten Farkas' Lemma folgt die Unzulässigkeit des primalen Problems.

Alle 3 Fälle führen auf einen Widerspruch, somit folgt die Behauptung. □

Kapitel 4

Semidefinite Programmierung

4.1 Das Standardprogramm

In der Standardform der positiv semidefiniten Programmierung schreibt man die linearen Restriktionen üblicherweise in einer speziellen Form.

Definition 4.1.1 *Sind A, B symmetrische, reellwertige, quadratische Matrizen, so definieren wir das* innere Produkt der Matrizen *als $A \bullet B := \textbf{Spur}(A^\top B)$. Die* Standardform des semidefiniten Programmierungsproblems *(SDP) ist dann für reelle, symmetrische, quadratische Matrizen C, A_i und X*

$$\begin{array}{ll} \min & C \bullet X \\ \textit{unter} & A_i \bullet X = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0. \end{array}$$

$X \succeq 0$ bedeutet wie oben definiert, daß X reell, symmetrisch und positiv semidefinit ist. $X \succ 0$ wollen wir schreiben, falls X reell, symmetrisch und positiv definit ist. Beachte, somit $(X \succeq 0$ und $X \neq 0) \not\Rightarrow X \succ 0$.

Man beachte, daß das innere Produkt eine lineare Zielfunktion und lineare Restriktionen definiert. Genauer ist $A \bullet B = \sum_{i,j} a_{ij} b_{ij}$. Außerdem ist es nützlich zu bemerken, daß für beliebige $(n \times k)$-Matrizen C $A \bullet (CC^\top) = \textbf{Spur}(C^\top AC)$ ist, speziell also gilt: $x^\top Ax = A \bullet (xx^\top)$.

Mit $\textbf{vec}(A)$ wollen wir den Vektor bezeichnen, der durch untereinanderschreiben der Spalten aus A entsteht, also

$$\textbf{vec}(A)^\top = (a_{11}, \ldots, a_{m1}, \ldots, a_{1n}, \ldots, a_{mn})$$

.

Ist umgekehrt $v \in \mathbf{R}^{mn}$ ein Vektor, so bezeichen wir mit $\mathbf{Mat}_{mn}(v)$ die Matrix

$$\mathbf{Mat}_{mn}(v) = \begin{pmatrix} v_1 & \dots & v_{(m-1)n+1} \\ \vdots & \ddots & \vdots \\ v_n & \dots & v_{mn}. \end{pmatrix}.$$

Sind die Dimensionen von $\mathbf{Mat}_{mn}(v)$ aus dem Kontext klar, so schreiben wir auch kürzer $\mathbf{Mat}(v)$.

Wenn wir mit $\mathcal{A}$ die Matrix bezeichnen, die aus den Zeilen $\mathbf{vec}(A_i)^\top$ besteht, so können wir das Standard-SDP auch notieren als

$$\begin{array}{ll} \min & C \bullet X \\ \text{unter} & \mathcal{A}\mathbf{vec}(X) = b \\ & X \succeq 0. \end{array}$$

Da die Programme, die in dieser Notation möglich sind, mit der Transformation aus Beispiel 4 verträglich sind, erhalten wir:

$$\begin{array}{ll} \min & C \bullet X \\ \text{unter} & \mathcal{A}\mathbf{vec}(X) = b \\ & X \succeq 0 \end{array} \qquad \text{und} \qquad \begin{array}{ll} \max & u^\top b \\ \text{unter} & \mathbf{Mat}(u^\top \mathcal{A}) + S = C \\ & S \succeq 0 \end{array}$$

sind zueinander duale Optimierungsprobleme. Genauer sind die Skalarprodukte im primalen Problem unter der Transformation invariant und im Dualen geschieht gar nichts.

Betrachten wir als Beispiel folgendes SDP:

$$\begin{array}{rl} \lambda' = \max & A \bullet X \\ \text{unter} & I \bullet X = 1 \\ & X \succeq 0. \end{array}$$

Lemma 4.1.1 *Für das obige SDP gilt $\lambda' = \lambda_1(A)$.*

Beweis: Sei x ein normierter Eigenvektor x zum Eigenwert λ_1. Sei X die Matrix $X = xx^\top$. Dann ist $\lambda' \geq X \bullet A = x^\top A x = \lambda_1(A)$. Andererseits wird der Optimalwert einer linearen Zielfunktion auf einer konvexen, abgeschlossenen Menge immer an Extrempunkten angenommen, sofern solche existieren. Nach Aufgabe 33 d) sind die Extremalstrahlen des Kegels der positiv semidefiniten, reell, symmetrischen Matrizen genau die Matrizen $xx^\top$. Da der positiv semidefinite Kegel hier nur mit einer affinen Hyperebene geschnitten wird, müssen alle Ecken des zulässigen

Bereichs im semidefiniten Kegel auf Extremalstrahlen liegen. Damit gibt es ein x mit $\lambda' = A \bullet (xx^\top) = x^\top Ax$ (vgl. auch Lemma 5.2.1).

□

Dualisieren wir nun obiges Programm

$$\begin{array}{rl} \min & (-A) \bullet X \\ \text{unter} & I \bullet X = 1 \\ & X \succeq 0, \end{array}$$

so erhalten wir:

$$\begin{array}{rl} \max & z \\ \text{unter} & zI + S = -A \\ & S \succeq 0. \end{array} \quad \text{oder äquivalent} \quad \begin{array}{rl} \min & z \\ \text{unter} & zI - A \succeq 0. \end{array}$$

Nun ist $zI - A \succeq 0$ aber gleichbedeutend damit, daß $z \geq \lambda_1(A)$ ist. Also haben primales und duales Programm den gleichen Optimalwert.

Kommen wir noch einmal auf unser graphentheoretisches Beispiel zurück. Dort hatten wir Variableneinträge in der Matrix, deren größten Eigenwert wir minimieren wollten. Die duale Formulierung des Eigenwertproblems gestattet aber eine Einführung von Variablen in die Matrix folgendermaßen, wenn wir die symmetrische Matrix A mit lauter Nulleinträgen und $\frac{1}{2}$ in $a_{ij} = a_{ji}$ mit E_{ij}, falls $i \neq j$ und $a_{ii} = 1$ sonst bezeichnen:

$$\begin{array}{rll} \min & z & \\ \text{unter} & E_{ij} \bullet X = 2 & \text{falls } (i,j) \in E(G) \text{ oder } i = j \\ & zI - X \succeq 0. & \end{array}$$

Die letzte Bedingung garantiert, daß z den größten Eigenwert von X minimiert und die Nebenbedingungen fixieren die 1 Einträge von A. Somit haben wir ein SDP aufgestellt, dessen Optimallösung den Wert θ hat.

4.2 Mehr Notation

In diesem Abschnitt wollen wir etwas Notation einführen. Wir benötigen diese, um lineare Operationen und Manipulationen symmetrischer Matrizen zusammenzubringen.

Definition 4.2.1 (Kronecker Produkt) *Sei $A \in \mathbb{R}^{m\times n}$ und $B \in \mathbb{R}^{p\times q}$. Dann ist das* Kronecker Produkt von A und B *die $(mp)\times(nq)$ Matrix*

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}.$$

Von den folgenden Rechenregeln werden wir im sechsten Kapitel extensiven Gebrauch machen.

Lemma 4.2.1 *Sind A, B, C Matrizen verträglicher Dimensionen, so gilt*

1. $(A \otimes B)^\top = (A^\top \otimes B^\top)$,
2. $\mathbf{vec}(ABC) = (C^\top \otimes A)\mathbf{vec}(B)$,
3. $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$,
4. *Sind A und B invertierbar, so ist auch $A \otimes B$ invertierbar und es gilt $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.*
5. *Sind A eine symmetrische $(m\times m)$-Matrix und B eine symmetrische $(n\times n)$-Matrix, dann sind die Eigenwerte von $(A \otimes B)$ gerade $\lambda_i(A)\lambda_j(B)$ für $i = 1,\dots,m$ und $j = 1,\dots,n$. Eigenvektor zum Eigenwert $\lambda_i(A)\lambda_j(B)$ ist der Vektor $q_i(A) \otimes q_j(B)$, wobei $q_i(A)$ und $q_j(B)$ Eigenvektoren von A bzw. B zu den Eigenwerten λ_i und λ_j sind.*

Beweis: Nachrechnen in den Übungen. □

4.3 Regeln für Dualbildung

Sind SDPs nicht in Standardform gegeben, so kann man sie zunächst auf Standardform transformieren, dualisieren und dann vereinfachen. So kann man z.B. zeigen, daß das Duale des dualen Programms das primale ist. Folgende Regeln helfen dieses Prozedere zu verkürzen. Wie in der Linearen Programmierung hat man für Variablen des primalen Programms Nebenbedingungen im dualen Programm und umgekehrt.

Betrachten wir als Beispiel das SDP für θ:

	min		max	
V	$\geq_K 0$	$\leftrightarrow$	$\leq_{K^*}$	N
A	$\leq_K 0$	$\leftrightarrow$	$\geq_{K^*}$	B
R	unrestringiert	$\leftrightarrow$	$=$	D
N	$\geq_K$	$\leftrightarrow$	$\geq_{K^*} 0$	V
B	$\leq_K$	$\leftrightarrow$	$\leq_{K^*} 0$	A
D	$=$	$\leftrightarrow$	unrestringiert	R

Abbildung 4.1: Regeln für Dualbildung linearer Programme über einem Kegel K. Variablen entsprechen Nebenbedingungen und umgekehrt.

$$\begin{array}{rll} \min & z & \\ \text{unter} & E_{ij} \bullet X = 1 & \text{falls } (i,j) \in E(G) \text{ oder } i = j \\ & zI - X \succeq 0. & \end{array}$$

Wir haben $\binom{n+1}{2} + 1$ Variablen und $|E| + |V|$ Gleichungsrestriktionen, die Anzahl der Ungleichungsrestriktion entspricht der Dimension des Kegels also insgesamt $\binom{n+1}{2} + |E| + |V|$ Restriktionen im primalen Programm. Für das Duale erhalten wir:

$$\begin{array}{rl} \max & \sum_{i \in V} v_i + \sum_{(i,j) \in E} e_{ij} \\ \text{unter} & I \bullet U = 1 \\ & -U + \sum_{i \in V} v_i E_{ii} + \sum_{(i,j) \in E} e_{ij} E_{ij} = 0 \\ & U \succeq 0. \end{array}$$

Dieses können wir umschreiben zu

$$\begin{array}{rll} \max & U \bullet B(0) & \\ \text{unter} & \mathbf{Spur}(U) = 1 & \\ & U \bullet E_{ij} = 0 & \text{falls } (i,j) \notin E \\ & U \succeq 0. & \end{array}$$

4.4 Komplementärer Schlupf für Kegel

Gilt der starke Dualitätssatz für ein SDP, so ist das duality gap 0, d.h. $X \bullet S = 0$. Es gilt jedoch sogar:

Satz 4.4.1 *Seien $X, S \succeq 0$. Dann ist $X \bullet S = 0$ genau dann, wenn $XS = 0$.*

Beweis: Da $\textbf{Spur}(XS) = X \bullet S$ ist, ist eine Richtung trivial. Andererseits ist

$$0 = X \bullet S = \textbf{Spur}(X^{1/2} S X^{1/2}). \tag{4.1}$$

Da $X^{1/2} S X^{1/2}$ mit S positiv semidefinit ist und die Spur invariant unter Hauptachsentransformation ist (vgl. Übung 45), sind alle ihre Eigenwerte Null und wir erhalten

$$(X^{1/2} S^{1/2})^\top (X^{1/2} S^{1/2}) = X^{1/2} S X^{1/2} = 0. \tag{4.2}$$

Also ist auch $X^{1/2} S^{1/2} = 0$ und somit $XS = X^{1/2}(X^{1/2} S^{1/2}) S^{1/2} = 0$. □

Betrachten wir Komplementaritätsüberlegungen für Programme auf allgemeinen Kegeln K. Nehmen wir an, daß es endliche Optimallösungen x und s für das primale und das duale Problem gibt mit $x \geq_K 0,\ s \geq_{K^*} 0,\ x^\top s = 0$. Wir wollen diese Bedingung anhand zweier weiterer Beispiele untersuchen.

4.4.1 Eishörnchen

Wir hatten in den Übungen nachgerechnet, daß das Eishörnchen $Q := \{x \in \mathbf{R}^n \mid x_1 \geq \sqrt{x_2^2 + \ldots + x_n^2}\}$ selbstpolar ist. Seien nun x, s ein komplementäres Paar in $Q \setminus \{0\}$. Bei näherem Hinsehen erkennt man, daß die Projektionen von x und s in die $x_1 = 0$ Ebene antipodisch sein müssen. Genauer erhalten wir:

Lemma 4.4.1 *Seien $x \geq_Q 0$ und $s \geq_Q 0$, $s \neq 0 \neq x$. Dann gilt $x^\top s = 0$ genau dann, wenn $x_1 = \sqrt{x_2^2 + \ldots + x_n^2}$ ist und es ein $\alpha \leq 0$ gibt mit*

$$(x_1, x_2, \ldots, x_n) = (-\alpha s_1, \alpha s_2, \ldots, \alpha s_n).$$

Beweis: Die eine Implikation ist klar. Für die Hinrichtung kürzen wir zunächst ab $\tilde{x} = (x_2, \ldots, x_n)$ und $\tilde{s} = (s_2, \ldots, s_n)$. Dann gilt wegen $x^\top s = 0$ zunächst

$$\begin{aligned}(-\tilde{x})^\top \tilde{s} &= x_1 s_1 \\ &\geq \|\tilde{x}\|_2 \|\tilde{s}\|_2 \\ &= \|-\tilde{x}\|_2 \|\tilde{s}\|_2.\end{aligned}$$

Die Aussage über die Gleichheit in der Cauchy-Schwarzschen Ungleichung impliziert dann die Existenz eines $\alpha \leq 0$ mit $-\tilde{x} = -\alpha \tilde{s}$. Wir haben also

$$\begin{aligned}x_1 s_1 &= (-\tilde{x})^\top \tilde{s} \\ &= \|\tilde{x}\|_2 \|\tilde{s}\|_2\end{aligned}$$

und schließen aus $s \neq 0 \neq x$, daß $x_1 = \|\tilde{x}\|,\ s_1 = \|\tilde{s}\|$ und somit $x_1 = -\alpha \|\tilde{s}\|_2 = -\alpha s_1$. □

Optimierung über Eishörnchen kann auch als positiv semidefinites Programm formuliert werden. Denn

$$x \geq_Q 0 \Leftrightarrow \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ x_2 & x_1 & 0 & \dots & 0 \\ \vdots & 0 & x_1 & \ddots & \vdots \\ x_n & 0 & 0 & \dots & x_1 \end{pmatrix} = \begin{pmatrix} x_1 & \tilde{x}^\top \\ \tilde{x} & x_1 I \end{pmatrix} \succeq 0$$

mit $\tilde{x} = (x_2, \dots, x_n)^\top$.

Beachte, daß die Eigenwerte der rechten Matrix leicht durch Abspaltung von $x_1 I$ zu berechnen sind ($x_1, x_1 - \|\tilde{x}\|, x_1 + \|\tilde{x}\|$).

4.4.2 p-Kegel

Betrachten wir hier die Komplementaritätsbedingungen, so erhalten wir aus der Hölderschen Ungleichung

Lemma 4.4.2 *Sei* $\frac{1}{p} + \frac{1}{q} = 1$ *und* $p, q \geq 1$. *Sind weiter* $x \geq_p 0$ *und* $s \geq_q 0$, $s \neq 0$ *sowie* $s^\top x = 0$. *Dann gilt* $s_i = 0 \Rightarrow x_i = 0$ *und für* $s_i \neq 0$

$$\frac{x_1^p}{s_1^q} = \frac{|x_i|^p}{|s_i|^q}. \tag{4.3}$$

Beweis: Wir haben

$$\begin{aligned} x_1 s_1 &= -x_2 s_2 - \ldots - x_n s_n \\ &\leq |x_2 s_2| + \ldots + |x_n s_n| \\ &\leq \sqrt[p]{|x_2|^p + \ldots + |x_n|^p} \sqrt[q]{|s_2|^q + \ldots + |s_n|^q} \text{ (Höldersche Ungleichung)} \\ &\leq x_1 s_1, \text{ da } x \geq_p 0,\ s \geq_q 0. \end{aligned}$$

Also muß überall Gleichheit gelten und die Aussage folgt wie oben aus der Aussage über Gleichheit in der Hölderschen Ungleichung. □

4.4.3 Degeneration in SDPs

Wir betrachten wieder das primale und duale Standard-SDP.

$$\begin{array}{llll} \min & C \bullet X & \max & u^\top b \\ \text{unter} & \mathcal{A}\mathbf{vec}(X) = b & \text{unter} & \mathbf{Mat}(\mathcal{A}^\top u) + S = C \\ & X \succeq 0 & & S \succeq 0 \end{array} \tag{4.4}$$

Wir hatten beobachtet, daß im Optimum bei verschwindendem duality gap $XS = 0$ gelten muß. Da X, S symmetrisch sind, gilt: $XS = X^\top S^\top = (SX)^\top = 0 = SX$. Also ist jeder Eigenvektor von X zu einem Eigenwert $\lambda \neq 0$ ein Eigenvektor von S zum Eigenwert 0 und das gleiche gilt, wenn man die Rollen von X und S vertauscht. Also gibt es eine Orthonormalbasis aus gemeinsamen Eigenvektoren von X und S, die komplementär sind. In diesem Falle nennen wir X und S *komplementär* . Dies impliziert sofort:

Lemma 4.4.3 *Sind X, S Optimallösungen des primalen und dualen Programms, so gilt unter den Voraussetzungen von Satz 3.5.1 im Optimum*

$$\textit{rang}(X) + \textit{rang}(S) \leq n.$$

□

Definition 4.4.1 *Wir sagen, daß X und S* strikt komplementär *sind, wenn sie komplementär sind und keinen gemeinsamen Eigenvektor zum Eigenwert 0 haben.*

Offensichtlich gilt:

Lemma 4.4.4 *Sind X und S komplementär, so sind sie strikt komplementär genau dann, wenn rang(X) + rang$(S) = n$.*

□

Bemerkung 4.4.1 *In der linearen Programmierung gibt es stets ein Optimallösungspaar, das strikte Komplementarität erfüllt. Dies ist in der semidefiniten Programmierung i.a. nicht der Fall (vgl. Alizadeh et.al. "Complementarity and Nondegeneracy in Semidefinite Programming").*

Kapitel 5

Eigenwerte und semidefinite Programme

Eine häufige Aufgabenstellung, die zu semidefiniten Programmen führt, ist es, geeignete Linearkombinationen von Eigenwerten einer symmetrischen Matrix unter linearen Nebenbedingungen an die Matrix zu finden. In diesem Kapitel wollen wir einige Rezepte kennenlernen, wie man diese Aufgabenstellung als semidefinites Programm modelliert.

Zur Abkürzung wollen wir etwas Notation einführen.

Definition 5.0.2 *Sind mit $A_0, A_1, \ldots, A_m$ $m+1$ reellwertige, symmetrische Matrizen gegeben und ist $x \in \mathbb{R}^m$ so setzen wir*

$$A(x) := A_0 + x_1 A_1 + \ldots + x_m A_m.$$

Ist X eine symmetrische Matrix, so bezeichnen wir mit $(\lambda_1 + \ldots + \lambda_k)(X)$ die Summe der k größten Eigenwerte von X.

Sei $B = \{b_1, \ldots, b_m\}$ eine Menge von rationalen Zahlen und seien $a_1, \ldots, a_m$ m rationaleVektoren. Dann bezeichnen wir mit

$$B(x) := \{a_1^\top x + b_1, \ldots, a_m^\top x + b_m\}.$$

Mit $\max(B)$ bezeichnen wir das größte Element von B, mit $\max^{\Sigma k}(B)$ die Summe der k größten Elemente, bzw. mit $\max^{\Sigma|k|}(B)$ die Summe der k betragsmäßig größten Elemente. Ist der Kontext klar, so kürzen wir ab: $\max(x) := \max(B(x))$, $\max^{\Sigma k}(x) := \max^{\Sigma k}(B(x))$, $\max^{\Sigma|k|}(x) := \max^{\Sigma|k|}(B(x))$.

5.1 Einfache Probleme als lineare Programme

Fangen wir mit trivialen Problemen an.

5.1.1 Problem: Finde $\max B$

Die Maximumbestimmung können wir als folgendes LP formulieren:

$$\begin{array}{lll}(\mathbf{P_1}) & \min & z \\ & \text{unter} & z \geq b_i, \quad i = 1, \ldots, m.\end{array}$$

Dem dualen Programm von $(\mathbf{P_1})$ sieht man die Aufgabenstellung noch deutlicher an:

$$\begin{array}{llll}(\mathbf{D_1}) & \max & u^\top b & \\ & \text{unter} & u^\top \mathbb{1} = 1 & \\ & & u_i \geq 0 & i = 1, \ldots, m.\end{array}$$

Die zulässige Menge von $(\mathbf{D_1})$ heißt *Standardsimplex*.

5.1.2 Minimiere eine endliche Menge linearer Funktionen

Da es sich hierbei um eine Verallgemeinerung des Problems aus dem vorhergehenden Paragraphen handelt, wollen wir versuchen eine Modellierung direkt aus $(\mathbf{P_1})$ oder $(\mathbf{D_1})$ zu gewinnen. Ein direktes Ersetzen von b_i durch die lineare Funktion ist im dualen Programm nicht ratsam, da dies zu einer quadratischen Zielfunktion führt. Im primalen eingesetzt erhalten wir hingegen:

$$\begin{array}{lll}(\mathbf{P_2}) & \min & z \\ & \text{unter} & z - a_i^\top x \geq b_i, \quad i = 1, \ldots, m.\end{array}$$

Das Duale hierzu ist:

$$\begin{array}{llll}(\mathbf{D_2}) & \max & u^\top b & \\ & \text{unter} & u^\top \mathbb{1} = 1 & \\ & & u^\top A = 0 & \\ & & u_i \geq 0 & i = 1, \ldots, m,\end{array}$$

wobei A die Matrix ist, die $a_i^\top$ als i-te Zeile enthält.

5.1.3 Finde die k Größten

Hier betrachten wir also das Problem: Finde $\max^{\Sigma k}(B)$. Als ganzzahliges lineares Programm läßt es sich sofort wie folgt formulieren:

$$
\begin{array}{lll}
(\mathbf{I_3}) & \max & u^\top b \\
 & \text{unter} & u^\top \mathbb{1} = k \\
 & & u_i \in \{0,1\} \quad i = 1,\ldots,m.
\end{array}
$$

Scharfes Hinsehen verifiziert, daß wir in diesem Fall die Ganzzahligkeitsrestriktion durch eine Ungleichung ersetzen können, d.h. die LP-Relaxation liefert ein ganzzahliges Optimum.

$$
\begin{array}{lll}
(\mathbf{D_3}) & \max & u^\top b \\
 & \text{unter} & u^\top \mathbb{1} = k \\
 & & 0 \leq u_i \leq 1 \quad i = 1,\ldots,m.
\end{array}
$$

Diese Tatsache notieren wir als

Lemma 5.1.1 *Die Extremalpunkte (Ecken) von*

$$S_k := \{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i = k,\ 0 \leq x_i \leq 1\}$$

sind 0-1-*Vektoren, d.h.* S_k *ist ein ganzzahliges Polyeder.*

Dualisieren wir dieses Programm, so erhalten wir:

$$
\begin{array}{llll}
(\mathbf{P_3}) & \min & kz + \mathbb{1}^\top v & \\
 & \text{unter} & z + v_i \geq b_i, & i = 1,\ldots,m, \\
 & & v_i \geq 0, & i = 1,\ldots,m.
\end{array}
$$

Zur Kontrolle interpretieren wir dieses Programm noch einmal. Im Optimum wird offensichtlich z den k-größten Wert annehmen und die v_i füllen auf, um noch über die b_i zu kommen.

5.1.4 Minimiere die Summe der k größten linearen Funktionen

Aufgabenstellung ist nun, $\max^{\Sigma k}(x)$. zu bestimmen. Nach bewährter Rezeptur ersetzen wir dafür b_i, wo es allein steht, durch $a_i^\top x + b_i$ und erhalten:

$$
\begin{array}{llll}
(\mathbf{P_4}) & \min & kz + \mathbb{1}^\top v & \\
& \text{unter} & z + v_i - a_i^\top x \geq b_i, & i = 1, \ldots, m, \\
& & v_i \geq 0, & i = 1, \ldots, m.
\end{array}
$$

Als duales Programm erhalten wir:

$$
\begin{array}{llll}
(\mathbf{D_4}) & \max & u^\top b & \\
& \text{unter} & u^\top \mathbb{1} = k & \\
& & u^\top A = 0 & \\
& & 0 \leq u_i \leq 1 & i = 1, \ldots, m.
\end{array}
$$

Lemma 5.1.2 *Seien* $\max(x)$ *und* $\max^{\Sigma k}(x)$ *wie in Definition 5.0.2. Dann ist*

1. $\max(x)$ *eine konvexe, stückweise lineare Funktion,*
2. $\max^{\Sigma k}(x)$ *eine konvexe Funktion.*

Beweis: Sei für x xl der Index, an dem $a_i^\top x + b_i$ den l-größten Wert annimmt. Dann ist mit $z = tx + (1-t)y$ für $t \in [0,1]$:

$$
\begin{aligned}
\max^{\Sigma k}(z) &= a_{z1}z + b_{z1} + \ldots + a_{zk}z + b_{zk} \\
&= t(a_{z1}x + b_{z1} + \ldots + a_{zk}x + b_{zk}) \\
&\quad +(1-t)(a_{z1}y + b_{z1} + \ldots + a_{zk}y + b_{zk}) \\
&\leq t\max^{\Sigma k}(x) + (1-t)\max^{\Sigma k}(y).
\end{aligned}
$$

Damit ist die Konvexität gezeigt. Nun ist $a_i^\top x + b_i \leq a_j^\top x + b_j \Leftrightarrow (a_i - a_j)^\top x \leq (b_j - b_i)$. Das heißt aber, daß die Menge aller Punkte x, an denen das Maximum in $a_l^\top x + b_l$ angenommen wird, der Schnitt von $m-1$ Halbräumen und somit ein Polyeder ist. Der $\mathbb{R}^n$ zerfällt also in maximal m Polyeder, auf denen $\max(x)$ linear ist. □

Wir werden später sehen, daß auch $(\lambda_1 + \ldots + \lambda_k)(A(x))$ eine konvexe Funktion ist, weshalb sich die Minimierung dieses Parameters als konvexes Programm, speziell als semidefinites Programm (SDP) schreiben läßt. Dafür brauchen wir eine Matrixversion von Lemma 5.1.1.

5.2 Eigenwertprobleme als SDPs

Mit der Maximierung des größten Eigenwertes einer Matrix hatten wir uns zwischendurch schon beschäftigt. Also kommen wir direkt zur

5.2.1 Bestimmung der Summe der k größten Eigenwerte

Das Problem ist die Bestimmung von $(\lambda_1 + \ldots + \lambda_k)(A)$.

Satz 5.2.1 (Matrixanalogon von Lemma 5.1.1) *Sei*

$$S_1 = \{QQ^\top \mid Q \in \mathbb{R}^{n\times k},\ Q^\top Q = I_k\}$$

$$S_2 = \{Y \in \mathbb{R}^{n\times n} \mid Y = Y^\top, \textbf{Spur}(Y) = k, 0 \preceq Y \preceq I\}.$$

Dann ist conv$(S_1) = S_2$ *und* S_1 *sind genau die Extremalen von* S_2.

Beweis: Wir zeigen zunächst conv$(S_1) \subseteq S_2$. Da S_2 offensichtlich konvex ist, genügt es, $S_1 \subseteq S_2$ zu zeigen. Sei also $QQ^\top \in S_1$. Offensichtlich ist $QQ^\top$ symmetrisch und positiv semidefinit. Da $\textbf{Spur}(QQ^\top) = \textbf{Spur}(Q^\top Q) = k$ ist, stimmt auch die Spur. Abschließend folgt aus $Q^\top Q = I_k$, daß $QQ^\top = QQ^\top QQ^\top$. Die Eigenwerte von $QQ^\top$ müssen also unter Quadrieren fix bleiben. Als Eigenwerte von $QQ^\top$ kommen folglich nur 0 und 1 in Frage. Also ist $QQ^\top \preceq I$.

Wir zeigen nun $S_2 \subseteq$ conv(S_1). Sei $Y \in S_2$ auf Hauptachse transformiert $Y = QDQ^\top$. Der Vektor y der Diagonalen der auf Hauptachse transformierten Matrix $D = Diag(y)$ läßt sich nach Lemma 5.1.1 als Konvexkombination von 0-1-Vektoren mit k 1-en schreiben $y = \sum_i \alpha_i u_i$. Setzen wir nun $Q_i := QDiag(u_i)$, so gilt $Q_i^\top Q_i = Diag(u_i)$ und wir erhalten:

$$\begin{aligned} Y &= QDQ^\top \\ &= QDiag(\sum_i \alpha_i u_i)Q^\top \\ &= \sum_i \alpha_i QDiag(u_i)Diag(u_i)Q^\top \\ &= \sum_i \alpha_i Q_i Q_i^\top. \end{aligned}$$

Entsteht nun $\tilde{Q}_i$ aus Q_i durch Streichen der Nullspalten, so sind die $\tilde{Q}_i$ $(n \times k)$-Matrizen und es gilt $\tilde{Q}_i^\top \tilde{Q}_i = I_k$ und $\tilde{Q}_i \tilde{Q}_i^\top = Q_i Q_i^\top$.

Diese Rechnung zeigt einerseits, daß $S_2 \subseteq$ conv(S_1), andererseits aber auch, daß die Extrempunkte von S_2 alle in S_1 liegen.

Wir haben oben gesehen, daß die Matrizen in S_1 alle nur die Eigenwerte 0 (mit Eigenraum der Dimension $n-k$) und 1 (mit Eigenraum der Dimension k) haben. Wir zeigen nun noch, daß jede solche Matrix extremal ist. Sei also Y eine solche Matrix. Angenommen es gäbe $Y_1, Y_2 \in S_2$ und ein $t \in (0,1)$ mit

$$Y = tY_1 + (1-t)Y_2.$$

Auf Hauptachse transformiert erhalten wir, daß $tQ^\top Y_1 Q + (1-t)Q^\top Y_2 Q$ eine 0-1-Diagonalmatrix mit genau k 1-en ist. Da die Eigenwerte von Y_1 und Y_2 alle zwischen 0 und 1 liegen, muß jeder Eigenvektor von Y zum Eigenwert 1 schon ein Eigenvektor von Y_1 und Y_2 zum Eigenwert 1 sein. Die Behauptung folgt. $\square$

Satz 5.2.2

$$(\mathbf{D_5}) \qquad (\lambda_1 + \ldots + \lambda_k)(A) = \max\{A \bullet Y \mid \mathbf{Spur}(Y) = k \textit{ und } 0 \preceq Y \preceq I\}.$$

Beweis: Seien $q_1^*, \ldots, q_k^*$ paarweise orthogonale, normierte Eigenvektoren zu den Eigenwerten $\lambda_1, \ldots, \lambda_k$. Bezeichnet nun $Q^* = (q_1^*, \ldots, q_k^*)$ die Matrix, deren Spalten diese Eigenvektoren sind, dann ist

$$(\lambda_1 + \ldots + \lambda_k)(A) = \sum_{i=1}^k {q_i^*}^\top A q_i^* = \mathbf{Spur}({Q^*}^\top A Q^*) = (Q^* {Q^*}^\top) \bullet A.$$

Also ist die Summe der Eigenwerte schon mal kleiner gleich dem Optimalwert λ^* des SDPs.

Sei nun Y^* eine Optimallösung des Programms (beachte: die zulässige Menge ist kompakt, denn die zulässige Menge ist bzgl. der Spektralnorm beschränkt, vgl. Bemerkung 6.4.1), also $\lambda^* = A \bullet Y^*$. Da die Zielfunktion linear ist, können wir annehmen, daß Y^* ein Extremalpunkt der zulässigen Menge ist, also nach Satz 5.2.1 $Y^* = PP^\top$ und $P^\top P = I_k$. Sei ferner $q_1, \ldots, q_n$ eine Orthonormalbasis aus Eigenvektoren von A. Bezeichnen dann $p_1, \ldots, p_k$ die Spalten der Matrix P, dann bilden die p_i ein orthonormales System und wir erhalten

$$\begin{aligned}
\lambda^* &= (PP^\top) \bullet A \\
&= \mathbf{Spur}(P^\top A P) \\
&= \sum_{i=1}^k {p_i}^\top A p_i \\
&= \sum_{j=1}^n \sum_{i=1}^k (q_j^\top p_i) q_j^\top A (q_j^\top p_i) q_j \\
&= \sum_{j=1}^n \sum_{i=1}^k (q_j^\top p_i)^2 \lambda_j \\
&\le \max\{\sum_{j=1}^n u_j \lambda_j \mid 0 \le u \le 1,\ \mathbb{1}^\top u = k\} \\
&= (\lambda_1 + \ldots + \lambda_k)(A).
\end{aligned}$$

Die Ungleichung folgt aus $\sum_{j=1}^n (q_j^\top p_i)^2 = p_i^\top p_i = 1$ und $\sum_{i=1}^k (q_j^\top p_i)^2 \le 1$. $\square$

Betrachten wir nun das duale Programm von $(\mathbf{D_5})$. Der Übersichtlichkeit halber notieren wir dafür zunächst noch einmal $(\mathbf{D_5})$ in tabellarischer Form.

$$\begin{array}{lll} (\mathbf{D_5}) & \max & A \bullet Y \\ & \text{unter} & I \bullet Y = k \\ & & Y \preceq I \\ & & Y \succeq 0. \end{array}$$

Dann erhalten wir als duales Programm:

$$\begin{array}{lll} (\mathbf{P_5}) & \min & kz + \mathbf{Spur}(V) \\ & \text{unter} & zI + V \succeq A \\ & & V \succeq 0. \end{array}$$

Satz 5.2.3 *Das duality gap von* $(\mathbf{P_5})$ *und* $(\mathbf{D_5})$ *ist* 0.

Beweis: Sei $A = Q^\top DQ$ die Hauptachsentransformation von A. Setze $z = \lambda_k(A)$ und $V = QDiag(\lambda_1 - \lambda_k, \lambda_2 - \lambda_k, \ldots, \lambda_{k-1} - \lambda_k, 0, \ldots, 0)Q^\top$. Dann sind z und V zulässig für $(\mathbf{P_5})$, denn V und $zI + Q^\top VQ - Q^\top AQ = Diag(0, \ldots, 0, \lambda_k - \lambda_{k+1}, \ldots, \lambda_k - \lambda_n)$ sind positiv semidefinit und $kz + \mathbf{Spur}(V) = (\lambda_1 + \ldots + \lambda_k)(A)$. Die Behauptung folgt somit aus dem Satz von der schwachen Dualität. □

5.2.2 Einführung von Variablen

Analog zu den Überlegungen in den LPs wollen wir nun Variablen in die Matrix A einführen. Wir können dann die Matrix A als Linearkombination von symmetrischen Matrizen auffassen, wobei die Koeffizienten Variablen sind und erhalten das Problem:

$$\min\{(\lambda_1 + \ldots + \lambda_k)(A(x)) \mid x \in \mathbf{R}^m\}.$$

Wollen wir eines der obigen SDPs für diese Aufgabenstellung abwandeln, so erkennen wir, daß $(\mathbf{D_5})$ für eine Modifikation ungeeignet ist, da dort A variablenbehaftet ist. Aus $(\mathbf{P_5})$ erhalten wir hingegen direkt:

$$\begin{array}{lll} (\mathbf{P_6}) & \min & kz + \mathbf{Spur}(V) \\ & \text{unter} & zI + V - \sum_{i=1}^m x_i A_i \succeq A_0, \\ & & V \succeq 0. \end{array}$$

Dualisieren wir dies, so erhalten wir.

$$
\begin{array}{lll}
(\mathbf{D_6}) & \max & A_0 \bullet Y \\
 & \text{unter} & \textbf{Spur}(Y) = k, \\
 & & A_i \bullet Y = 0, \qquad i = 1, \ldots, m, \\
 & & 0 \preceq Y \preceq I.
\end{array}
$$

Auch hier handelt es sich wiederum um eine konvexe Optimierungsaufgabe.

Lemma 5.2.1 $(\lambda_1 + \ldots + \lambda_k)(A(x))$ *ist eine konvexe Funktion.*

Beweis: Übung. □

Unter der Annahme, daß das duality gap verschwindet, lautet die Bedingung vom komplementären Schlupf hier, daß zulässige Lösungen x, z, V und Y optimal sind genau dann, wenn

$$Y(zI + V - A(x)) = 0 \text{ und } (I - Y)V = 0.$$

5.2.3 Minimiere Eigenwerte unter Gleichungsrestriktionen

Wir betrachten als nächstes das Problem

$$\min\{(\lambda_1 + \ldots + \lambda_k)(X) \mid A_i \bullet X = b_i,\, i = 1 \ldots, m\}.$$

Analog zu obigem erhalten wir als primal duales Paar für dieses Problem

$$
\begin{array}{llll}
(\mathbf{P_7}) & \min & kz + \textbf{Spur}(V) & \\
 & \text{unter} & A_i \bullet X = b_i, & i = 1, \ldots, m, \\
 & & zI + V - X \succeq 0, & \\
 & & V \succeq 0. &
\end{array}
$$

$$
\begin{array}{lll}
(\mathbf{D_7}) & \max & u^\top b \\
 & \text{unter} & Y - \sum_{i=1}^m u_i A_i = 0, \\
 & & \textbf{Spur}(\mathrm{Y}) = \mathrm{k} \\
 & & 0 \preceq Y \preceq I.
\end{array}
$$

Auch hier notieren wir die Bedingung vom komplementären Schlupf:

$$Y(zI + V - X) = 0 \text{ und } (I - Y)V = 0.$$

5.2.4 Minimiere gewichtete Summen von Eigenwerten

Als letzte Problemklasse in diesem Kapitel betrachten wir folgendes Problem

$$\min\{m_1\lambda_1 + \ldots + m_k\lambda_k(A(x)) \mid x \in \mathbf{R}^m\},$$

wobei $m_1 \geq \ldots \geq m_k$ gelten muß.

Da $m_1\lambda_1 + \ldots + m_k\lambda_k = (m_1 - m_2)\lambda_1 + (m_2 - m_3)(\lambda_1 + \lambda_2) + \ldots + (m_{k-1} - m_k)(\lambda_1 + \ldots + \lambda_{k-1}) + m_k(\lambda_1 + \ldots + \lambda_k)$ ist, handelt es sich bei der Zielfunktion um eine positive Linearkombination von größte Eigenwertsummenaufgaben, also um eine konvexe Funktion.

Wir notieren zunächst das zugehörige SDP für eine feste Matrix A, abgeleitet von $(\mathbf{D_5})$. Aus Notationsgründen setzen wir $m_{k+1} = 0$.

$$\begin{array}{lll} (\mathbf{D_8}) & \max & \sum_{i=1}^k (m_i - m_{i+1}) Y_i \bullet A \\ & \text{unter} & \mathbf{Spur}(Y_i) = i, \qquad\qquad i = 1, \ldots, k, \\ & & 0 \preceq Y_i \preceq I. \end{array}$$

Als duales Programm erhalten wir:

$$\begin{array}{lll} (\mathbf{P_8}) & \min & \sum_{i=1}^k i z_i + \sum_{i=1}^k \mathbf{Spur}(V_i) \\ & \text{unter} & z_i I + V_i - (m_i - m_{i+1}) A \succeq 0, \quad i = 1, \ldots, k, \\ & & V_i \succeq 0. \end{array}$$

Eine SDP Formulierung für unser eigentliches Problem erhalten wir nun wieder, indem wir A in $(\mathbf{P_8})$ durch $A(x)$ ersetzen:

$$\begin{array}{lll} (\mathbf{P_9}) & \min & \sum_{i=1}^k i z_i + \sum_{i=1}^k \mathbf{Spur}(V_i) \\ & \text{unter} & z_i I + V_i - (m_i - m_{i+1}) \sum_{j=1}^m A_j x_j \succeq (m_i - m_{i+1}) A_0, \\ & & V_i \succeq 0 \qquad i = 1, \ldots, k,. \end{array}$$

Weitere Modellierungen von Eigenwertproblemen als SDPs findet man in Farid Alizadeh: „Interior Point Methods in Semidefinite Programming with Applications to Combinatorial Optimization“.

Kapitel 6

Innere Punkt Methoden

6.1 Barrier Functions

Wir hatten in der Vorlesung bereits barrier functions (innere Straffunktionen) kennengelernt. Für unsere Bedürfnisse hier wollen wir diesen Begriff etwas modifizieren.

Definition 6.1.1 *Ist $S \subseteq \mathbb{R}^n$ konvex, so nennen wir eine Funktion b : int$(S) \to \mathbb{R}$ eine* barrier function *, wenn gilt :*

1. *b ist konvex auf int(S).*
2. *Ist $(x_i)_{i \in \mathbb{N}}$ eine Folge von Punkten, die gegen einen Randpunkt $x \in$ rd(S) konvergiert, so gilt $\lim_{k \to \infty} b(x_k) = +\infty$.*
3. *$\forall x \in S : b(x) \geq 0$.*

Die dritte Bedingung soll sicherstellen, daß die barrier function nach unten beschränkt ist und so existierende Minima nicht völlig verschwinden können.

Ist nun ein Minimierungsproblem über S mit konvexer Zielfunktion f gegeben, so führt die Addition der barrier function (multipliziert mit einem Gewicht $\mu > 0$) zu einem unrestringierten konvexen Optimierungsproblem $\min f(x) + \mu b(x)$, das sein Minimum, falls es existiert, im Innern von S annimmt.

Bei gegebenem Optimierungsproblem und gegebener barrier function bezeichnen wir mit x_μ das Argument des Minimums, also $f(x_\mu) + \mu b(x_\mu) = \min f(x) + \mu b(x)$.

Nach dem letzten Satz der Vorlesung in 1995 konvergiert die Folge x_μ mit $\mu \to 0$, falls die Minima stets existieren, gegen ein Minimum von $f(x)$ über S. (Genauer ist jeder Häufungspunkt der Folge ein Minimand des Originalproblems.)

Beispiel 6.1.1 1. *Sei S der nicht-negative Orthant, dann sind $b_1(x) = \sum_{i=1}^n \frac{1}{x_i}$ und $b_2(x) = -\sum_{i=1}^n \ln x_i$ zwei barrier functions für S. Beachte, daß $b_2 \geq 0$ nicht erfüllt ist. In der linearen Programmierung spielt das allerdings keine Rolle, falls der betrachtete zulässige Bereich durch M beschränkt ist, da man dann durch Addition einer Konstanten übergehen kann zu*

$$b_2(x) = -\sum_{i=1}^n \ln\,(x_i/M).$$

Man kann aus den Eingangsdaten in der Linearen Programmierung die Größe einer Optimallösung abschätzen.

2. *Ist S der Kegel der positiv semidefiniten Matrizen, dann kann man analog zu b_1 für den positiven Orthanten definieren:*

$$b_1(X) = \sum_{i=1}^n \frac{1}{\lambda_i(X)} = \frac{\sum_{i=1}^n \prod_{i \neq j} \lambda_j(X)}{\prod_{i=1}^n \lambda_i}.$$

Der Nenner des letzten Ausdrucks ist die Determinante von X und der Zähler das negative des Koeffizienten des linearen Terms im charakteristischen Polynom von X. Analog zu b_2 definieren wir:

$$b_2(x) = -\sum_{i=1}^n \ln \lambda_i(X) = -\ln(\det(X)).$$

Wir werden später verlangen, daß semidefinite Programme mit einer Abschätzung des Optimums versehen sind. Deshalb können wir über die fehlende Nicht-Negativität von b_2 hinwegsehen.

6.2 Das Newton-Verfahren in der Linearen Programmierung

Wir betrachten das Lineare Programmierungsproblem in Standardform:

$$\begin{array}{lll} (\mathbf{P}) & \min & c^\top x \\ & \text{unter} & Ax = b, \\ & & x \geq 0. \end{array}$$

Wir addieren eine barrier function für den positiven Orthanten und erhalten:

$$\begin{array}{lll} (\mathbf{P}_\mu) & \min & c^\top x - \mu \sum_{i=1}^n \ln x_i \\ & \text{unter} & Ax = b. \end{array}$$

Für die verbleibenden Gleichungsrestriktionen betrachten wir die Lagrangefunktion (vgl. Analysis II und Kuhn-Tucker Bedingungen):

$$L(x,u) = c^\top x - \mu \sum_{i=1}^n \ln x_i + u^\top (b - Ax).$$

Offensichtlich kann ein lokales Minimum nur vorliegen, wenn $Ax = b$ gilt.

Als notwendige (hier auch hinreichende) Bedingungen für ein Minimum erhalten wir:

$$\begin{array}{l} \nabla_x L(x,u) = c - \mu(1/x_1, \ldots, 1/x_n)^\top - A^\top u = 0 \\ \nabla_u L(x,u) = b - Ax = 0. \end{array} \tag{6.1}$$

Durch Substitution von $\frac{\mu}{x_i}$ erhalten wir daraus folgendes Gleichungssystem:

$$\begin{array}{l} Ax = b, \\ A^\top u + s = c, \\ s_i - \frac{\mu}{x_i} = 0, \qquad i = 1, \ldots, n. \end{array} \tag{6.2}$$

Für $\mu \geq 0$ sei (x_μ, u_μ, s_μ) die (falls existente, eindeutige!) Lösung des Systems (6.2). Falls das lineare Programmierungsproblem eine optimale Lösung besitzt (und die Menge der Optimallösungen beschränkt ist, vgl. Übung 59), so konvergiert die Folge dieser (x_μ) mit $\mu \to 0$ gegen eine Optimallösung des linearen Programms. Die Menge $\{x_\mu \mid x_\mu$ ist Optimallösung von $P_\mu,\ \mu > 0\}$ heißt *primaler zentraler Weg*.

Lemma 6.2.1 *Die Abbildung* $\begin{array}{c} (x_\mu, u_\mu, s_\mu) : (0, +\infty) \to \mathbb{R}^n \\ \mu \mapsto (x_\mu, u_\mu, s_\mu) \end{array}$ *ist stetig.*

Beweis: Übungen. □

Betrachten wir das gleiche Vorgehen im dualen Programm

$$\begin{array}{lll} (\mathbf{D}) & \max & u^\top b \\ & \text{unter} & A^\top u + s = c, \\ & & s \geq 0, \end{array}$$

erhalten wir als nicht explizit vorzeichenbeschränktes Problem:

$$\begin{array}{lll} (\mathbf{D}_\nu) & \max & u^\top b + \nu \sum_{i=1}^m \ln s_i \\ & \text{unter} & A^\top u + s = c. \end{array}$$

Die Lagrangefunktion lautet hier:

$$L(u,s,x) = b^\top u + \nu \sum_{i=1}^n \ln s_i + x^\top (c - A^\top u - s),$$

und aus der Optimalitätsbedingung erhalten wir mit

$$\begin{aligned} \nabla_u L(u,s,x) &= b - Ax = 0 \\ \nabla_s L(u,s,x) &= \nu(1/s_1, \ldots, 1/s_n)^\top - x = 0, \\ \nabla_x L(u,s,x) &= c - A^\top u - s = 0. \end{aligned} \tag{6.3}$$

im wesentlichen dasselbe Gleichungssystem. Entsprechend nennt man die Menge $\{(u_\nu, s_\nu) \mid (u_\nu, s_\nu)$ ist Optimallösung von $D_\nu,\ \nu > 0\}$ *dualer zentraler Weg* . Faßt man die Wege zusammen, so nennt man die Menge

$$\{(u_\mu, s_\mu, x_\mu) \mid Ax_\mu = b,\ A^\top u_\mu + s_\mu = c,\ (x_\mu)_i (s_\mu)_i = \mu,\ \mu > 0\}$$

primal dualer zentraler Weg.

Wir wollen hier für lineare Programmierung nur diese spezielle barrier function betrachten. Sie führt in diesem Falle auf eine glatte Kurve. Eine sehr erfolgreiche Klasse polynomieller Algorithmen für die lineare Programmierung sind die sogenannten *wegefolgenden Methoden* , die zu den *inneren Punkt Methoden* gerechnet werden. Diese versuchen, den zentralen Weg durch einen stückweise linearen Weg zu approximieren und dieser Approximation für $\mu \to 0$ zu folgen. Wesentlich für die Komplexität des Algorithmus ist dabei die Anzahl der Knicke in dem approximierenden Polygon.

Dieses Polygon erhält man, indem man für jedes μ nur einen Newtonschritt mit einer bestimmten Schrittweite durchführt und danach μ verkleinert. Geschickte Wahl der Schrittweite und des Updates von μ hält dann die Anzahl der Knicke klein.

Zur Durchführung eines Newton-Schrittes lösen wir das Gleichungssystem (6.2) mit dem Newton-Verfahren zur Bestimmung der Nullstelle eines Vektorfeldes.

Die Iterationsgleichung für das Newton-Verfahren lautete

$$z_{k+1} = z_k - (F'(z_k))^{-1} F(z_k). \tag{6.4}$$

Die Hauptarbeit in jeder Iteration ist die „Invertierung“der Jacobischen. Diese wird jedoch meist nicht explizit durchgeführt, sondern das Gleichungssystem

$$(F'(z_k))(z_k - z_{k+1}) = F(z_k) \tag{6.5}$$

gelöst. Dennoch bestimmt die Gestalt der Jacobischen den Rechenaufwand und die numerische Stabilität des Verfahrens. Im Falle der linearen Programmierung hat die Jacobische $F'(z_k)$ die einfache Gestalt

$$F' = \begin{pmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ D_x & 0 & D_s \end{pmatrix}.$$

Die letzte Zeile dieser Matrix hängt hierbei von der Form ab, in der wir die letzte Gleichung notiert haben. Möglich sind hierfür:

1. $s_i - \frac{\mu}{x_i} = 0$,
2. $x_i - \frac{\mu}{s_i} = 0$,
3. $x_i s_i - \mu = 0$.

Dies ergibt folgende Möglichkeiten für D_x und D_s:

1. $D_x = \mu Diag(1/x_1^2, \dots, 1/x_n^2)$ und $D_s = I$,
2. $D_x = I$ und $D_s = \mu Diag(1/s_1^2, \dots, 1/s_n^2)$,
3. $D_x = Diag(s_1, \dots, s_n)$ und $D_s = Diag(x_1, \dots, x_n)$.

6.3 Das Newton-Verfahren und SDP

Betrachten wir nun das Standard SDP:

$$
\begin{array}{lll}
(\mathbf{P}) & \min & C \bullet X \\
 & \text{unter} & \mathcal{A}\mathbf{vec}(X) = b, \\
 & & X \succeq 0.
\end{array}
$$

Wir addieren eine barrier function für den positiv semidefiniten Kegel und erhalten:

$$
\begin{array}{lll}
(\mathbf{P}_\mu) & \min & C \bullet X - \mu \ln(\det(X)) \\
 & \text{unter} & \mathcal{A}\mathbf{vec}(X) = b,
\end{array}
$$

Die Lagrangefunktion lautet:

$$L(X, u) = C \bullet X - \mu \ln \det(X) + u^\top (b - \mathcal{A}\mathbf{vec}(X)).$$

Für die Optimalitätsbedingungen dieser Funktion benötigen wir die Ableitung der Determinante:

Lemma 6.3.1

$$(\ln \det(X))' = \mathbf{vec}(X^{-1}).$$

Beweis:

$$
\begin{aligned}
\frac{\partial \ln \det(X)}{\partial X_{ij}} &= \frac{1}{\det X} \frac{\partial \det(X)}{\partial X_{ij}} \\
&= \frac{\det(X^{ij})}{\det(X)} \\
&= (X^{-1})_{ij},
\end{aligned}
$$

wobei X^{ij} die Adjunkte des Elementes X_{ij} ist. Dabei folgt die vorletzte Gleichheit aus dem Laplaceschen Entwicklungssatz und die letzte Gleichheit aus der Cramerschen Regel. □

Wir erhalten also als Optimalitätsbedingungen:

$$
\begin{aligned}
\nabla_x L(X, u) &= C - \mu X^{-1} - \mathbf{Mat}(\mathcal{A}^\top u) = 0 \\
\nabla_u L(X, u) &= b - \mathcal{A}\mathbf{vec}(X) = 0.
\end{aligned}
\tag{6.6}
$$

Wir substituieren wieder und erhalten:

$$\begin{aligned} &\mathcal{A}\mathbf{vec}(X) = b, \\ &\mathbf{Mat}(\mathcal{A}^\top u) + S = C, \\ &S - \mu X^{-1} = 0. \end{aligned} \tag{6.7}$$

Wie in der linearen Programmierung definieren wir einen zugehörigen primalen zentralen Pfad.

Betrachten wir auch hier die duale Fragestellung:

$$\begin{array}{ll} \max & u^\top b \\ \text{unter} & \mathbf{Mat}(\mathcal{A}^\top u) + S = C, \\ & S \succeq 0. \end{array}$$

Mit barrier function ergibt das

$$\begin{array}{ll} \max & u^\top b + \mu \ln \det(S) \\ \text{unter} & \mathbf{Mat}(\mathcal{A}^\top u) + S = C \end{array}$$

und die Lagrangefunktion

$$L(u, S, X) = b^\top u + \mu \ln \det(S) + X \bullet (C - \mathbf{Mat}(\mathcal{A}^\top u) - S)$$

mit Optimalitätsbedingungen

$$\begin{aligned} &\nabla_u L(u, S, X) = b - \mathcal{A}\mathbf{vec}(X) = 0, \\ &\nabla_S L(u, S, X) = \mu S^{-1} - X = 0, \\ &\nabla_X L(u, S, X) = C - \mathbf{Mat}(\mathcal{A}^\top u) - S = 0, \end{aligned} \tag{6.8}$$

was also wieder zum Gleichungssystem (6.7) führt.

Die letzte Bedingung in (6.7) können wir auf vier Arten ausdrücken:

1. $S - \mu X^{-1} = 0,$
2. $X - \mu S^{-1} = 0,$
3. $XS - \mu I = 0,$
4. $XS + SX - 2\mu I = 0.$

Kürzen wir nun ab $(\Delta X, \Delta U, \Delta S) := (X,u,S)_k - (X,u,S)_{k+1}$, so lautet das Gleichungssystem, das wir für einen Newton-Schritt zu lösen haben:

$$\begin{pmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{A}^\top & I \\ E & 0 & F \end{pmatrix} \begin{pmatrix} \Delta X \\ \Delta u \\ \Delta S \end{pmatrix} = \begin{pmatrix} \mathcal{A}\mathbf{vec}(X_k) - b \\ \mathbf{Mat} A^\top u_k + S_k \\ G \end{pmatrix}. \tag{6.9}$$

Wir wollen nun die Auswirkungen der Wahl der letzten Iterationsgleichung auf dieses System studieren.

Für E und F müssen wir die Ableitung von Matrixprodukten und Matrixinversion bestimmen. Beachte, daß die Jacobische dabei für Abbildungen von $(n \times n)$-Matrizen eine $(n^2 \times n^2)$-Matrix ist.

Lemma 6.3.2 *Sei S eine $(n \times n)$-Matrix und $f, g, h : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ Abbildungen mit folgender Abbildungsvorschrift: $f(X) = XS$, $g(X) = SX$, $h(X) = X^{-1}$ bzw. $h(X)$ undefiniert, falls X nicht invertierbar. Dann gilt:*

1. $f'(X) = S^\top \otimes I$,
2. $g'(X) = I \otimes S$,
3. $h'(X) = -(X^{-1})^\top \otimes X^{-1}$.

Beweis: Aus $(X + \Delta X)S = XS + \Delta XS$ folgt $f'(X)\mathbf{vec}(\Delta X) = \mathbf{vec}(\Delta XS) = \mathbf{vec}(I\Delta XS) = (S^\top \otimes I)\mathbf{vec}(\Delta X)$, wobei letzte Gleichheit nach Lemma 4.2.1 gilt. Damit ist die erste Behauptung gezeigt. Analog gilt $g'(X)\mathbf{vec}(\Delta X) = \mathbf{vec}(S\Delta X) = \mathbf{vec}(S\Delta XI) = (I \otimes S)\mathbf{vec}(\Delta X)$.

Für die letzte Behauptung betrachten wir die Reihenentwicklung von X^{-1}. Für ΔX hinreichend klein ist

$$(X + \Delta X)^{-1} = X^{-1} \sum_{i=0}^{\infty} (-\Delta X X^{-1})^i.$$

Demnach ist $h'(X)\mathbf{vec}(\Delta X) = -\mathbf{vec}(X^{-1}\Delta X X^{-1}) = -((X^{-1})^\top \otimes X^{-1})\mathbf{vec}(\Delta X)$, woraus die Behauptung folgt. $\square$.

Somit haben wir folgende Fallunterscheidung für das Newton-Gleichungssystem.

1. $E = \mu(X_k^{-1})^\top \otimes X_k^{-1}$ und $F = I$,

2. $E = I$ und $F = \mu(S_k^{-1})^\top \otimes S_k^{-1}$,
3. (XS-Methode) $E = S_k^\top \otimes I$ und $F = I \otimes X_k$,
4. ($XS + SX$-Methode) $E = S_k^\top \otimes I + I \otimes S_k$ und $F = X_k^\top \otimes I + I \otimes X_k$.

Betrachtet man die Matrix des Gleichungssystems in (6.9), so erkennt man, daß in den ersten beiden Fällen diese Matrix mit $\mu \to 0$ für festes S_k gegen eine singuläre Matrix geht, was zu numerischen Instabilitäten in den Algorithmen führt.

Die $XS + SX$ Methode hat gegenüber der XS-Methode den Vorteil, daß hier alle Ausdrücke symmetrisch bleiben. Beachte, das Produkt zweier symmetrischer Matrizen ist i.a. nicht symmetrisch. Ansonsten läuft man Gefahr, daß $\Delta X, \Delta S$ nicht symmetrisch sind, obwohl dies für X_k und S_k der Fall war.

Das Skriptum des Kurses von Alizadeh geht hier nicht weiter auf den wegefolgenden Ansatz ein. Alizadeh schreibt allerdings in einem Übersichtsartikel zur semidefiniten Programmierung: „Given an interior point algorithm for linear programming we may construct, in a mechanical way, an algorithm for the SDP problem by replacing any references (in einer Abbildung, die die Korrespondenzen wiedergibt) to the entries under the LP column, with the corresponding entry under the SDP column. Proofs of convergency or polynomial time complexity may also be extended mechanically in the same manner. “

Bemerkung 6.3.1 *In dem oben beschriebenen wie auch in dem folgenden Ansatz startet man stets mit einem Satz primal und dual zulässiger innerer Lösungen. Speziell hat man also ein u mit $C - \mathbf{Mat}(\mathcal{A}^\top u) \succ 0$. Nach Lemma 3.4.3 sind also die Voraussetzungen für starke Dualität aus Satz 3.5.1 erfüllt.*

Im folgenden werden wir die Übertragung des LP-Algorithmus von Ye detaillierter studieren.

6.4 Einige Lemmata

Hier stellen wir einige Lemmata vor, die wir im folgenden benötigen.

Lemma 6.4.1 *Sind $a_1, \dots a_n$ nicht-negative, reelle Zahlen, so gilt:*

$$n \ln(\sum_{i=1}^n a_i) - \ln(\prod_{i=1}^n a_i) \geq n \ln(n). \tag{6.10}$$

Beweis: An dieser Stelle benötigen wir die Beziehung zwischen arithmetischem und geometrischem Mittel (Stichwort Jensensche Ungleichung): $(\sum_{i=1}^n a_i)/n \geq (\prod_{i=1}^n a_i)^{\frac{1}{n}}$. Hieraus schließen wir zunächst: $\ln(\sum_{i=1}^n a_i) - \frac{1}{n}\ln(\prod_{i=1}^n a_i) \geq \ln(n)$, woraus die Behauptung folgt. □

Lemma 6.4.2 *Ist $A \succ 0$ und $B \succeq 0$, so sind alle Eigenwerte von AB reell und nicht-negativ. Genauer stimmen die Eigenwerte von AB mit denen der positiv semidefiniten, symmetrischen Matrix $A^{\frac{1}{2}}BA^{\frac{1}{2}}$ überein.*

Beweis: Sei $Q^\top DQ = A^{\frac{1}{2}}BA^{\frac{1}{2}}$ die Hauptachsentransformation der symmetrischen Matrix $A^{\frac{1}{2}}BA^{\frac{1}{2}}$. Wegen $QA^{\frac{-1}{2}}A^{\frac{1}{2}}Q^\top = I$ erhalten wir somit

$$AB = A^{\frac{1}{2}}Q^\top D(A^{\frac{1}{2}}Q^\top)^{-1}.$$

Sei nun $x_i := (A^{\frac{1}{2}}Q^\top)e_i$, Dann ist $ABx_i = A^{\frac{1}{2}}BA^{\frac{1}{2}}x_i$. □

Lemma 6.4.3 *Sei $x \in \mathbb{R}^n$ mit $\|x - \mathbb{1}\|_\infty < 1$. Dann ist*

$$\sum_{j=1}^n \ln x_j \geq (\mathbb{1}^\top x - n) - \frac{\|x-\mathbb{1}\|_2^2}{2(1-\|x-\mathbb{1}\|_\infty)}.$$

Beweis: Für $0 < x \leq 2$ gilt $\ln x = \sum_{i=1}^\infty (-1)^{i+1}\frac{(x-1)^i}{i}$. Demnach haben wir

$$\begin{aligned}
\sum_{j=1}^n \ln x_j &= \sum_{j=1}^n\sum_{i=1}^\infty (-1)^{i+1}\frac{(x_j-1)^i}{i} \\
&= (x^\top \mathbb{1} - n) - \sum_{j=1}^n \frac{(x_j-1)^2}{2}\left(\sum_{i=0}^\infty (-1)^i \frac{2(x_j-1)^i}{i+2}\right) \\
&\geq (\mathbb{1}^\top x - n) - \frac{\|x-\mathbb{1}\|_2^2}{2(1-\|x-\mathbb{1}\|_\infty)}.
\end{aligned}$$

Dabei folgt die letzte Ungleichung aus

$$\begin{aligned}
\sum_{i=0}^\infty (-1)^i \frac{2(x_j-1)^i}{i+2} &\leq \sum_{i=0}^\infty \frac{2\|x-\mathbb{1}\|_\infty^i}{i+2} \\
&\leq \sum_{i=0}^\infty \|x-\mathbb{1}\|_\infty^i \\
&= \frac{1}{1-\|x-\mathbb{1}\|_\infty}
\end{aligned}$$

Dieses Lemma läßt sich direkt auf Eigenwerte symmetrischer Matrizen übertragen. Bezeichne $\rho(X)$ den *Spektralradius von* X, dies ist bei einer symmetrischen Matrix der betraglich größte Eigenwert.

Bemerkung 6.4.1 (Vorsicht!!) *Üblicherweise bezeichnet man die Spektralnorm*

$$\max_{\|x\|_2=1} \|Ax\|_2$$

als Zwei-Norm $\|A\|_2$, *da sie mit der euklidischen Norm verträglich ist. In unserem Fall betrachten wir allerdings Matrizen mal als Vektor und mal als Matrix. Deshalb wollen wir auch im Falle von Matrizen die Wurzel aus der Summe der Quadrate als* $\|\cdot\|_2$ *bezeichen und die Spektralnorm als* $\|\cdot\|_\infty$.

Lemma 6.4.4 *Sei* X *eine symmetrische* $(n \times n)$*-Matrix mit* $0 \prec X \prec 2I$. *Dann gilt:*

$$\ln \det X \geq \mathbf{Spur}(X) - n - \frac{\mathbf{Spur}\left((X-I)^2\right)}{2(1-\rho(X-I))}.$$

Beweis: Setzen wir $x_i := \lambda_i(X)$, so ist nach Voraussetzung $\|x - \mathbb{1}\|_\infty < 1$, also gilt

$$\sum_{j=1}^{n} \ln x_j \geq (\mathbb{1}^\top x - n) - \frac{\|x-\mathbb{1}\|_2^2}{2(1-\|x-\mathbb{1}\|_\infty)}.$$

Die Behauptung folgt also aus $\mathbb{1}^\top x = \mathbf{Spur}(X),\ \|x-\mathbb{1}\|_2^2 = \mathbf{Spur}(X-I)^2$ und $\|x - \mathbb{1}\|_\infty = \rho(X-I)$. □

6.5 Potentialfunktionen und projektive Transformationen

Eine wesentliche Idee zum Beweis der Polynomialität von inneren Punkt-Methoden für lineare Programmierung ist die Einführung von Potentialfunktionen, die im Verlauf des Algorithmus gleichmäßig kleiner werden.

Für ein Standard-SDP und sein Duales betrachten wir folgende Potentialfunktionen für (X, u, S) primal-dual zulässig, $X, S \succ 0$ und einen skalaren Parameter q. Ferner sei z eine (bekannte) untere Schranke für den Optimalwert z^* des primalen Problems.

primale Potentialfunktion $\Phi(X, z) = q \ln(C \bullet X - z) - \ln(\det X)$,

primal-duale Potentialfunktion $\Psi(X,S) = q\ln(X \bullet S) - \ln(\det(XS))$.

Satz 6.5.1 *Sei $(X_i, u_i, S_i)_{i\in\mathbb{N}_0}$ eine Folge primal-dual zulässiger, innerer Punkte, also $X_i \succ 0$ und $S_i \succ 0$ und $0 < \epsilon < 1$. Existiere ein $\delta > 0$ mit für alle i : $\Psi(X_{i+1}, S_{i+1}) \leq \Psi(X_i, S_i) - \delta$. Sei ferner $\Psi(X_0, S_0) \leq E\lceil\sqrt{n}\rceil$ für eine Konstante E und sei $q = n + \lceil\sqrt{n}\rceil$. Dann gilt für $k\delta \geq \lceil\sqrt{n}\rceil|\log\epsilon|$:*

$$S_k \bullet X_k < 2^E\epsilon.$$

Beweis: Aus den Bedingungen an Ψ schließen wir

$$\begin{aligned}\Psi(X_k, S_k) &\leq \Psi(X_0, S_0) - k\delta \\ &\leq E\lceil\sqrt{n}\rceil - \lceil\sqrt{n}\rceil|\log\epsilon| \\ &= \lceil\sqrt{n}\rceil(\log 2^E\epsilon).\end{aligned} \tag{6.11}$$

Nach Lemma 6.4.2 sind die $\lambda_i(X_kS_k)$ nicht negativ, also gilt nach Lemma 6.4.1:

$$\begin{aligned}\ln\det(X_kS_k) &= \ln\prod_{i=1}^n \lambda_i(X_kS_k) \\ &\leq -n\ln n + n\ln(\sum_{i=1}^n \lambda_i) \\ &= -n\ln n + n\ln(X_k \bullet S_k).\end{aligned}$$

Aus (6.11) haben wir $(n + \lceil\sqrt{n}\rceil)\ln(X_k \bullet S_k) \leq \ln(\det(X_kS_k)) + \lceil\sqrt{n}\rceil(\log 2^E\epsilon)$ also

$\lceil\sqrt{n}\rceil\ln(X_k \bullet S_k) \leq -n\ln n + \lceil\sqrt{n}\rceil(\log 2^E\epsilon)$. □

Ist $X_0 \succ 0$ ein Iterationspunkt, so greift die Restriktion $X \succ 0$ für den nächsten Iterationspunkt am stärksten in Richtung eines Eigenvektors zum Eigenwert $\lambda_n(X_0)$. Wenn dieser nun annähernd mit der Optimierungsrichtung der aktuellen Iteration übereinstimmt, machen wir in dieser Iteration keinen großen Fortschritt. Um dem entgegegenzuwirken, transformieren wir die ganze Aufgabenstellung in eine Situation, in welcher alle Eigenwerte des gegenwärtigen Iterationspunktes gleich 1 sind.

Sei dafür L_0 eine $(n \times n)$-Matrix mit $L_0L_0^\top = X_0$. Mögliche Wahlen für L_0 sind ein *Cholesky-Faktor* oder eine Wurzel von X_0. Die tatsächliche Wahl von L_0 spielt für das weitere Verhalten des Algorithmus allerdings keine Rolle.

Sei $r \in \mathbb{N}$ fest und $\mathcal{S}^{n\times n}$ der (offene, konvexe) Kegel der positiv definiten, symmetrischen, reellen Matrizen. Dann definieren wir $\mathcal{T} : \mathcal{S}^{n\times n} \to \mathcal{S}^{n\times n} \times \mathbb{R}^r$ vermöge $\mathcal{T}(X) = (\bar{X}, \bar{x})$, wobei

$$\bar{X} := \frac{(n+r)L_0^{-1}X(L_0^{-1})^\top}{r + X_0^{-1} \bullet X} \text{ und } \bar{x} := \left(\frac{n+r}{r + X_0^{-1} \bullet X}\right) \mathbb{1}. \tag{6.12}$$

Wir nennen diese Abbildung *projektive Transformation*. Beachte, daß die Transformation X_0 auf $(I, \mathbb{1})$ abbildet.

Die inverse Transformation ist definiert durch

$$X = \mathcal{T}^{-1}(\bar{X}, \bar{x}) = \frac{rL_0\bar{X}L_0^\top}{\mathbb{1}^\top \bar{x}}. \tag{6.13}$$

Als transformiertes Programm bezeichnen wir das folgende SDP:

$$\begin{array}{ll} \min & \bar{C} \bullet \bar{X} + \bar{c}(z)^\top \bar{x} \\ \text{unter} & \bar{\mathcal{A}}\mathbf{vec}(\bar{X}) + \bar{A}\bar{x} = 0 \\ & \mathbf{Spur}(\bar{X}) + \mathbb{1}^\top \bar{x} = n + r \\ & \bar{X} \succeq 0 \\ & \bar{x} \geq 0, \end{array} \tag{6.14}$$

mit

$$\begin{array}{rcl} \bar{C} & = & L_0^\top C L_0 \\ \bar{c}(z) & = & \dfrac{-z}{r}\mathbb{1} \\ \bar{A}_i & = & L_0^\top A_i L_0 \\ \bar{\mathcal{A}} & = & \mathcal{A}(L_0 \otimes L_0) \\ \bar{A} & = & \dfrac{-1}{r} b \mathbb{1}^\top. \end{array}$$

$\bar{A}$ ist also eine $(m \times r)$-Matrix vom Rang ≤ 1.

Lemma 6.5.1 (Eigenschaften der projektiven Transformation:)

1. *$\mathcal{T}(X_0) = (I, \mathbb{1})$ ist zulässig für das transformierte Problem,*
2. *$\mathcal{T}^{-1} \circ \mathcal{T}(X) = X$,*

3. *Ist X zulässig für das primale Problem, so ist $\mathcal{T}(X)$ zulässig für das Hilfsproblem.*

4. *Ist $(\bar{X}, \bar{x})$ zulässig für das transformierte Problem und $\bar{x} \neq 0$, so ist X zulässig für das Originalproblem.*

Beweis: Die erste Behauptung verifiziert man mittels scharfem Hinsehen. 2. folgt direkt durch Einsetzen:

$$\mathcal{T}^{-1} \circ \mathcal{T}(X) = \frac{r + X_0^{-1} \bullet X}{n + r} \frac{n + r}{r + X_0^{-1} \bullet X} L_0 L_0^{-1} X L_0^{-T} L_0^T = X.$$

Zu 3. bemerken wir zuerst, daß, falls L eine reguläre Matrix und X positiv (semi) definit ist, auch $L^\top X L$ positiv (semi) definit ist. Die Behauptung folgt also aus:

$$\begin{aligned} \frac{r + X_0^{-1} \bullet X}{n + r} (\bar{\mathcal{A}}\mathbf{vec}(\bar{X}) + \bar{A}\bar{x}) &= \mathcal{A}(L_0 \otimes L_0)\mathbf{vec}(L_0^{-1} X L_0^{-\top}) + \frac{-1}{r} b \mathbb{1}^\top \mathbb{1} \\ &= \mathcal{A}(L_0 \otimes L_0)(L_0^{-1} \otimes L_0^{-1})\mathbf{vec}(X) - b \\ &= \mathcal{A}\mathbf{vec}(X) - b = 0 \end{aligned}$$

und

$$\begin{aligned} \mathbf{Spur}(\bar{X}) + \mathbb{1}^\top \bar{x} &= \frac{n + r}{r + X_0^{-1} \bullet X} (\mathbf{Spur} L_0^{-1} X L_0^{-\top} + r) \\ &= \frac{n + r}{r + X_0^{-1} \bullet X} (X \bullet L_0^{-\top} L_0^{-1} + r) = n + r. \end{aligned}$$

Zur letzten Behauptung haben wir

$$\begin{aligned} \mathcal{A}\mathbf{vec}(X) &= \frac{r}{\sum_{j=1}^r \bar{x}_j} \mathcal{A}\mathbf{vec}(L_0 \bar{X} L_0^\top) \\ &= \frac{r}{\sum_{j=1}^r \bar{x}_j} \mathcal{A}(L_0 \otimes L_0)\mathbf{vec}(\bar{X}) \\ &= \frac{r}{\sum_{j=1}^r \bar{x}_j} \bar{\mathcal{A}}\mathbf{vec}(\bar{X}) \\ &= -\frac{r}{\sum_{j=1}^r \bar{x}_j} \bar{A}\bar{x} \\ &= \frac{r}{\sum_{j=1}^r \bar{x}_j} \frac{1}{r} b \mathbb{1}^\top \bar{x} \\ &= b. \end{aligned}$$

□

Wir definieren folgende primale Potentialfunktion für das transformierte Problem:

$$\bar{\Phi}(\bar{X}, \bar{x}, z) = q \ln(\bar{C} \bullet \bar{X} + \bar{c}(z)^\top \bar{x}) - \ln \det \bar{X} - \sum_{j=1}^{r} \ln \bar{x}_j. \tag{6.15}$$

Das transformierte Programm zusammen mit der transformierten Potentialfunktion gibt die Situation des Originalprogramms wieder, da (unter gewissen Voraussetzungen) die Differenzen der Werte zweier Punkte der originalen und der transformierten Potentialfunktion invariant bleiben.

Lemma 6.5.2 *Sei* $\bar{x} = \bar{x}_1 \mathbb{1}$, $q = n + r$ *und* $X = \mathcal{T}^{-1}(\bar{X}, \bar{x})$. *Dann gilt*

$$\Phi(X, z) - \Phi(X_0, z) = \bar{\Phi}(\bar{X}, \bar{x}, z) - \bar{\Phi}(I, \mathbb{1}, z). \tag{6.16}$$

Beweis:

$$\begin{aligned}
&\Phi(X, z) - \Phi(X_0, z) = \\
&= q(\ln C \bullet X - z) - \ln \det X - q \ln(C \bullet X_0 - z) + \ln \det X_0 \\
&= q(\ln C \bullet \frac{L_0 \bar{X} L_0^\top}{\bar{x}_1} - z) - \ln \det \frac{L_0 \bar{X} L_0^\top}{\bar{x}_1} \\
&\quad - q \ln(C \bullet X_0 - z) + \ln \det X_0 \\
&= q \ln \left(\frac{1}{\bar{x}_1} (C \bullet L_0 \bar{X} L_0^\top - z \bar{x}_1) \right) - \ln \left(\left(\frac{1}{\bar{x}_1} \right)^n \det X_0 \det \bar{X} \right) \\
&\quad - q \ln(C \bullet X_0 - z) + \ln \det X_0 \\
&= -(n + r) \ln \bar{x}_1 + q \ln(\bar{C} \bullet \bar{X} - z \bar{x}_1) + n \ln \bar{x}_1 - \ln \det X_0 - \ln \det \bar{X} \\
&\quad - q \ln(C \bullet X_0 - z) + \ln \det X_0 \\
&= - \sum_{j=1}^{r} \ln \bar{x}_j + q \ln(\bar{C} \bullet \bar{X} + \bar{c}(z)^\top \bar{x}) - \ln \det \bar{X} - q \ln(C \bullet X_0 + \bar{c}(z)^\top \mathbb{1}) \\
&= \bar{\Phi}(\bar{X}, \bar{x}, z) - \bar{\Phi}(I, \mathbb{1}, z).
\end{aligned}$$

□

Die folgende Abschätzung werden wir zum Beweis der Reduktion der primal-dualen Potentialfunktion benötigen.

Korollar 6.5.1 *Ist* $q = n + r$, $\textbf{Spur}(\bar{X}) + \mathbb{1}^\top \bar{x} = n + r$ *und* $\|\bar{X} - I\|^2 + \|\bar{x} - \mathbb{1}\|^2 \leq \beta^2 < 1$, *so gilt*

$$\begin{aligned}
&\bar{\Phi}(\bar{X}, \bar{x}, z) - \bar{\Phi}(I, \mathbb{1}, z) \leq \\
&(n + r) \ln \left(\frac{\bar{C} \bullet \bar{X} + \bar{c}(z)^\top \bar{x}}{\textbf{Spur}(\bar{C}) + \bar{c}(z)^\top \mathbb{1}} \right) + \frac{\textbf{Spur}(\bar{X} - I)^2 + \|\bar{x} - \mathbb{1}\|_2^2}{2(1 - (\|\bar{X} - I\|_2 + \|x - \mathbb{1}\|_2))}.
\end{aligned} \tag{6.17}$$

Beweis: Betrachte die $((n+r)\times(n+r))$ Matrix $\tilde{X}$ gegeben durch

$$\tilde{X} = \begin{pmatrix} \bar{X} & 0 & 0 & \dots & 0 \\ 0 & \bar{x}_1 & 0 & \dots & 0 \\ 0 & 0 & \bar{x}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \bar{x}_n \end{pmatrix}. \tag{6.18}$$

Dann gilt wegen Lemma 6.4.4:

$$\begin{aligned} \bar{\Phi}(\bar{X},\bar{x},z) - \bar{\Phi}(I,\mathbb{1},z) &= q\ln(\bar{C}\bullet\bar{X}+\bar{c}(z)^\top\bar{x}) - \ln\det\bar{X} - \sum_{j=1}^{r}\ln\bar{x}_j \\ &\quad -q\ln(\bar{C}\bullet I+\bar{c}(z)^\top\mathbb{1}) \\ &= q\ln\left(\frac{\bar{C}\bullet\bar{X}+\bar{c}(z)^\top\bar{x}}{\bar{C}\bullet I+\bar{c}(z)^\top\mathbb{1}}\right) - \ln\det\tilde{X} \\ &\le (n+r)\ln\left(\frac{\bar{C}\bullet\bar{X}+\bar{c}(z)^\top\bar{x}}{\mathbf{Spur}(\bar{C})+\bar{c}(z)^\top\mathbb{1}}\right) - \mathbf{Spur}(\tilde{X}) + \\ &\quad n+r+\frac{\mathbf{Spur}(\bar{X}-I)^2+\|\bar{x}-\mathbb{1}\|_2^2}{2(1-\rho(\tilde{X}-I))}. \end{aligned}$$

Die Behauptung folgt nun aus

$$\begin{aligned} \rho(\tilde{X}-I) &= \lambda^1(\tilde{X}-I) \\ &= \|(\lambda^1(\tilde{X}-I),\dots,\lambda^n(\tilde{X}-I))\|_\infty \\ &\le \|(\lambda^1(\tilde{X}-I),\dots,\lambda^n(\tilde{X}-I))\|_2 \\ &= \|\bar{X}-I\|_2+\|x-\mathbb{1}\|_2. \end{aligned}$$

□

6.6 Ein Potentialreduktionsalgorithmus

Schon in Karmarkars erstem polynomiellen innere-Punkt Verfahren für die lineare Programmierung wurde eine Tansformation durchgeführt und danach die Nichtnegativitätsbedingungen $X \succeq 0, x \geq 0$ durch die Bedingung, innerhalb einer Kugel um $(I,\mathbb{1}) = \mathcal{T}(X_0)$ zu bleiben, ersetzt. Dies wollen wir hier auch tuen und erhalten damit das folgende Optimierungsproblem über einer Kugel:

$$\begin{array}{rl} \min & \bar{C} \bullet \bar{X} + \bar{c}(z)^\top \bar{x} \\ \text{unter} & \bar{\mathcal{A}}\mathbf{vec}(\bar{X}) + \bar{A}\bar{x} = 0, \\ & \mathbf{Spur}(\bar{X}) + \mathbb{1}^\top \bar{x} = n + r, \\ & \mathbf{Spur}(\bar{X} - I)^2 + \|\bar{x} - \mathbb{1}\|_2^2 \le \beta^2 < 1, \end{array} \tag{6.19}$$

wobei β ein fester Parameter zwischen 0 und 1 ist, den wir in Kürze festlegen werden.

Als lineares Problem über einer Kugel mit linearen Nebenbedingungen läßt sich dieses Problem leicht lösen. Das Ergebnis holen wir zurück in den Originalraum und erhalten somit einen Kandidaten für den nächsten Iterationspunkt.

Die Lösung von (6.19) läßt sich explizit angeben zu:

$$\begin{pmatrix} \mathbf{vec}\bar{X}_1 \\ \bar{x}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{vec}I \\ \mathbb{1} \end{pmatrix} - \beta \frac{P(z)}{\|P(z)\|_2}, \tag{6.20}$$

und der Kandidat für die neue Iteration ist

$$X(z) = \mathcal{T}^{-1}(\bar{X}_1, \bar{x}_1). \tag{6.21}$$

Dabei ist

$$P(z) = \mathcal{P}_{\mathcal{A}'} \begin{pmatrix} \mathbf{vec}\bar{C} \\ \bar{c}(z) \end{pmatrix}, \qquad \mathcal{A}' = \begin{pmatrix} \bar{\mathcal{A}} & \bar{A} \\ (\mathbf{vec}(I))^\top & \mathbb{1}^\top \end{pmatrix}, \tag{6.22}$$

und $\mathcal{P}_{\mathcal{A}'}(y)$ ist die Projektion des $(n^2 + r)$-Vektors y auf den Kern von $\mathcal{A}'$ (vgl. Paragraph 9 des ersten Teils der Vorlesung).

Explizit erhalten wir $P(z) =$

$$\left(I - \frac{(\mathbf{vec}(I)^\top | \mathbb{1}^\top)^\top (\mathbf{vec}(I)^\top | \mathbb{1}^\top)}{n + r} \right) \left(I - (\bar{\mathcal{A}}|\bar{A})^\top ((\bar{\mathcal{A}}|\bar{A})(\bar{\mathcal{A}}|\bar{A})^\top)^{-1} (\bar{\mathcal{A}}|\bar{A}) \right) \begin{pmatrix} \mathbf{vec}\bar{C} \\ \bar{c}(z) \end{pmatrix}. \tag{6.23}$$

In dieser Darstellung wird zunächst auf den Kern der Abbildung $(\bar{\mathcal{A}} \mid \bar{A})$ und danach auf den Kern von $\mathbf{Spur}(\bar{X}) + \mathbb{1}^\top \bar{x}$ projeziert.

Wir setzen nun

$$\begin{aligned} u(z) &= ((\bar{\mathcal{A}} \mid \bar{A})(\bar{\mathcal{A}} \mid \bar{A})^\top)^{-1} (\bar{\mathcal{A}} \mid \bar{A}) \begin{pmatrix} \mathbf{vec}\bar{C} \\ \bar{c}(z) \end{pmatrix} \\ &= \left(\mathcal{A}(X_0 \otimes X_0)\mathcal{A}^\top + (1/r)bb^\top \right)^{-1} (\mathcal{A}\mathbf{vec}(X_0 C X_0) + (z/r)b), \end{aligned} \tag{6.24}$$

dabei folgt die letzte Gleichheit aus

$$(\bar{\mathcal{A}}\bar{\mathcal{A}}^\top) = (\mathcal{A}(L_0 \otimes L_0)(L_0 \otimes L_0)^\top \mathcal{A}^\top)$$

$$\begin{aligned} &= (\mathcal{A}(X_0 \otimes X_0)\mathcal{A}^\top), \\ \bar{\mathcal{A}}\mathbf{vec}\bar{C} &= \mathcal{A}(L_0 \otimes L_0)\mathbf{vec}(L_0^\top C L_0) \\ &= \mathcal{A}\mathbf{vec}(X_0 C X_0) \end{aligned}$$

sowie

$$\begin{aligned} (\bar{A}\bar{A}^\top) &= \frac{1}{r^2}(b\mathbb{1}^\top \mathbb{1}b^\top), \\ \bar{A}\bar{c}(z) &= \frac{z}{r^2}b\mathbb{1}^\top \mathbb{1}. \end{aligned}$$

Die zweite Dualvariable belegen wir dann mit

$$S(z) = C - \mathbf{Mat}(\mathcal{A}^\top u(z)). \tag{6.25}$$

$S(z)$ und $u(z)$ sind Kandidaten für neue duale Iterationspunkte.

Mit Hilfe der Kandidaten für die Dualvariablen kann man $P(z)$ auch ausdrücken als:

$$P(z) = \begin{pmatrix} \mathbf{vec}(L_0^\top S(z) L_0) \\ \frac{b^\top u(z) - z}{r}\mathbb{1} \end{pmatrix} - \frac{C \bullet X_0 - z}{n+r}\begin{pmatrix} \mathbf{vec}I \\ \mathbb{1} \end{pmatrix}. \tag{6.26}$$

Dies folgt unter Beachtung von

$$\mathbf{Mat}(\bar{A}^\top u) = \mathbf{Mat}\left((L_0^\top \otimes L_0^\top)\mathcal{A}^\top u\right) = \mathbf{Matvec}(L_0^\top(\mathbf{Mat}\mathcal{A}^\top u)L_0)$$

aus

$$\begin{aligned} P(z) &= \left(I - \frac{(\mathbf{vec}(I)^\top|\mathbb{1}^\top)^\top(\mathbf{vec}(I)^\top|\mathbb{1}^\top)}{n+r}\right)\left(\begin{pmatrix} \mathbf{vec}\bar{C} \\ \bar{c}(z) \end{pmatrix} - \begin{pmatrix} \bar{A}^\top u(z) \\ \bar{A}^\top u(z) \end{pmatrix}\right) \\ &= \left(I - \frac{(\mathbf{vec}(I)^\top|\mathbb{1}^\top)^\top(\mathbf{vec}(I)^\top|\mathbb{1}^\top)}{n+r}\right)\begin{pmatrix} \mathbf{vec}L_0^\top S(z) L_0 \\ \frac{-z}{r}\mathbb{1} + \frac{1}{r}\mathbb{1}b^\top u(z) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{vec}(L_0^\top S(z) L_0) \\ \frac{b^\top u(z) - z}{r}\mathbb{1} \end{pmatrix} - \frac{1}{n+r}(\mathbf{Spur}(L_0^\top S(z) L_0) + b^\top u(z) - z)\begin{pmatrix} \mathbf{vec}I \\ \mathbb{1} \end{pmatrix} \end{aligned}$$

und

$$\begin{aligned} & \mathbf{Spur}(L_0^\top C L_0) - \mathbf{Spur}(L_0^\top \mathbf{Mat}(\mathcal{A}^\top u(z)) L_0) + b^\top u(z) - z \\ =\; & C \bullet L_0 L_0^\top - (\mathbf{Mat}\mathcal{A}^\top u(z)) \bullet X_0 + b^\top u(z) - z \\ =\; & C \bullet X_0 - \mathbf{vec}\left((X_0)^\top \mathcal{A}^\top u(z)\right) + b^\top u(z) - z \\ =\; & C \bullet X_0 - z \end{aligned}$$

Beachte, daß $X(z), S(z)$ und $u(z)$ unabhängig von der speziellen Wahl von L_0 sind; tatsächlich benötigen wir L_0 nur für die Herleitung und nicht für die eigentlichen Rechnungen im Algorithmus. Um zu sehen, daß $X(z)$ unabhängig von der Zerlegung ist, betrachten wir (6.26). Offensichtlich verschwindet L_0 bei der Rücktransformation mittels (6.13).

Wir werden im folgenden beweisen, daß

- entweder die Ersetzung von X_0 durch $X(z)$ zu einer konstanten Reduktion der primal-dualen Potentialfunktion führt oder
- $S(z) \succ 0$ und damit $(u(z), S(z))$ dual zulässig ist und die Ersetzung der dualen Variablen zu einer konstanten Reduktion der primal-dualen Potentialfunktion führt.

Da $\mathcal{P}_{\mathcal{A}'}$ eine Projektion ist, d.h. $\mathcal{P}_{\mathcal{A}'}^2 = \mathcal{P}_{\mathcal{A}'}$, erhalten wir

$$\begin{pmatrix} \textbf{vec}(\bar{C}) \\ \bar{c}(z) \end{pmatrix}^\top P(z) = \|P(z)\|_2^2$$

und somit aus (6.20):

$$\bar{C} \bullet (\bar{X}_1 - I) + \bar{c}(z)^\top (x_1 - \mathbb{1}) = -\beta \|P(z)\|_2.$$

Korollar 6.6.1 *Sei $q = n + r$, $\bar{X}_1$, $\bar{x}_1$ wie in (6.20) und z eine untere Schranke für das primale SDP. Dann gilt:*

$$\bar{\Phi}(\bar{X}_1, \bar{x}_1, z) - \bar{\Phi}(I, \mathbb{1}, z) \leq -(n+r)\beta \frac{\|P(z)\|_2}{\textbf{Spur}(\bar{C}) + \bar{c}(z)^\top \mathbb{1}} + \frac{\beta^2}{2(1-\beta)}.$$

Beweis: Die Voraussetzungen von Korollar 6.5.1 sind erfüllt. Ferner gilt, da z untere Schranke des SDP ist (vgl. die Rechnung im Beweis von Lemma 6.5.2 und $(\bar{X}_1, \bar{x}_1) - (X_0, \mathbb{1})$ eine Verbesserungsrichtung der Zielfunktion ist:

$$0 \leq \bar{C} \bullet \bar{X}_1 + \bar{c}(z)^\top \bar{x}_1 < \textbf{Spur}(\bar{C}) + \bar{c}(z)^\top \mathbb{1}.$$

Da für $x > -1 : \ln(x+1) \leq x$ erhalten wir somit

$$\begin{aligned} \ln \left(\frac{\bar{C} \bullet \bar{X}_1 + \bar{c}(z)^\top \bar{x}_1}{\textbf{Spur}(\bar{C}) + \bar{c}(z)^\top \mathbb{1}} \right) &= \ln \left(1 + \frac{\bar{C} \bullet (\bar{X}_1 - I) + \bar{c}(z)^\top (\bar{x}_1 - \mathbb{1})}{\textbf{Spur}(\bar{C}) + \bar{c}(z)^\top \mathbb{1}} \right) \\ &\leq \frac{\bar{C} \bullet (\bar{X}_1 - I) + \bar{c}(z)^\top (\bar{x}_1 - \mathbb{1})}{\textbf{Spur}(\bar{C}) + \bar{c}(z)^\top \mathbb{1}} \\ &= \frac{-\beta \|P(z)\|_2}{\textbf{Spur}(\bar{C}) + \bar{c}(z)^\top \mathbb{1}}. \end{aligned}$$

Die Behauptung folgt somit aus Korollar 6.5.1.

□

Sei nun Δ_0 der Abstand der Zielfunktion im gegenwärtigen Iterationspunkt von der unteren Schranke z, also

$$\Delta_0 = C \bullet X_0 - z$$

und

$$\Delta_1 = S(z) \bullet X_0 = C \bullet X_0 - u(z)^\top b.$$

Also ist Δ_1 das duality gap bzgl. des neuen Kandidaten für den dualen Iterationspunkt.

Bevor wir zeigen, daß bei richtiger Wahl des Updates sich die Potentialfunktion stets um einen Mindestbetrag reduziert, zeigen wir:

Lemma 6.6.1 *Gibt es eine Zahl* $0 < \alpha < 1$ *mit*

$$\|P(z)\|_2 \leq \alpha \frac{\Delta_0}{n+r}$$

dann ist $S(z) \succ 0$ *und* $u(z)^\top b > z$*. Außerdem gilt:*

$$\|L_0^\top S(z) L_0 - \frac{\Delta_1}{n} I\| \leq \frac{\Delta_1}{n} \alpha \sqrt{\frac{n + n^2/r}{n + n^2/r - \alpha^2}}, \tag{6.27}$$

und

$$\left| \frac{n+r}{n} \frac{\Delta_1}{\Delta_0} - 1 \right| \leq \frac{\alpha}{\sqrt{n + n^2/r}}. \tag{6.28}$$

Beweis: Angenommen $S(z) \not\succ 0$. Dann ist auch $L_0^\top S(z) L_0$ nicht positiv definit, hat also einen nicht-positiven Eigenwert. Aus (6.26) folgt nun unter Beachtung von $\|\cdot\|_2 \geq \|\cdot\|_\infty$:

$$\|P(z)\|_2 \geq \rho \left(\frac{\Delta_0}{n+r} I - L_0^\top S(z) L_0 \right) \geq \frac{\Delta_0}{n+r},$$

im Widerspruch zur Voraussetzung des Lemmas. Den gleichen Widerspruch erhalten wir aus der Betrachtung der $\bar{x}$-Komponente von 6.26 und der Annahme $u(z)^\top b \leq z$:

$$\|P(z)\|_2 \geq \frac{\Delta_0}{n+r} - (u(z)^\top b - z) \geq \frac{\Delta_0}{n+r}.$$

Weiterhin liefert 6.26

$$P(z) = \begin{pmatrix} \mathbf{vec}\left((L_0^\top S(z)L_0) - \frac{\Delta_1}{n}I\right) - \left(\frac{\Delta_0}{n+r} - \frac{\Delta_1}{n}\right)\mathbf{vec}I \\ \left(\frac{\Delta_0-\Delta_1}{r} - \frac{\Delta_0}{n+r}\right)\mathbb{1} \end{pmatrix}.$$

Da allerdings $I \bullet \left((L_0^\top S(z)L_0) - (\Delta_1/n)I\right) = X_0 \bullet S(z) - (\Delta_1/n)I \bullet I = 0$ ist, erhalten wir :

$$\begin{aligned} \|P(z)\|_2^2 &= \left\|L_0^\top S(z)L_0 - \tfrac{\Delta_1}{n}I\right\|^2 + n\left(\tfrac{\Delta_0}{n+r} - \tfrac{\Delta_1}{n}\right)^2 + r\left(\tfrac{\Delta_0-\Delta_1}{r} - \tfrac{\Delta_0}{n+r}\right)^2 \\ &= \left\|L_0^\top S(z)L_0 - \tfrac{\Delta_1}{n}I\right\|^2 + \left(n + \tfrac{n^2}{r}\right)\left(\tfrac{\Delta_1}{n} - \tfrac{\Delta_0}{n+r}\right)^2. \end{aligned} \tag{6.29}$$

Gilt (6.27) nicht, so erhalten wir daraus

$$\begin{aligned} \|P(z)\|_2^2 &> \left(\tfrac{\Delta_1}{n}\right)^2 \alpha^2 \tfrac{n+n^2/r}{n+n^2/r-\alpha^2} + \left(n + \tfrac{n^2}{r}\right)\left(\tfrac{\Delta_1}{n} - \tfrac{\Delta_0}{n+r}\right)^2 \\ &\geq \alpha^2\left(\tfrac{\Delta_0}{n+r}\right)^2. \end{aligned} \tag{6.30}$$

Dabei folgt die letzte Ungleichung aus der Minimierung des mittleren Ausdrucks über Δ_1/n (Übung). Aus diesem Widerspruch folgt also die Gültigkeit von (6.27). Abschließend erhalten wir, da (6.30) nicht gelten kann:

$$\left(n + \frac{n^2}{r}\right)\left(\frac{\Delta_1}{n} - \frac{\Delta_0}{n+r}\right)^2 \leq \alpha^2\left(\frac{\Delta_0}{n+r}\right)^2,$$

woraus (6.28) folgt. □

Nun können wir den Potentialreduktionssatz beweisen.

Satz 6.6.1 *Sei $n \neq 1$. Sei X_0 eine innere zulässige Matrix für das primale Problem und u_0 ein innerer zulässige Punkt für das duale Problem. Ferner sei $r := \lceil\sqrt{n}\rceil$ und $q = n + r$, $S_0 = C - \sum_{i=1}^m u_i A_i$, $z_0 = u_0^\top b$, $X(z_0) = \mathcal{T}^{-1}(\bar{X}_1, \bar{x}_1)$, $u_1 = u(z_0)$ und $S_1 = S(z_0)$. Dann gibt es Konstanten α, β (für die Berechnung von S_1 und $X(z_0)$) und eine Konstante δ, so daß entweder*

$$\Psi(X(z_0), S_0) \leq \Psi(X_0, S_0) - \delta,$$

oder

$$\Psi(X_0, S_1) \leq \Psi(X_0, S_0) - \delta.$$

Genauer gilt, wenn $\alpha = 0.55$ und $\beta = 0.3$ sind, dann ist $\delta > 0.1$.

Beweis: Gilt für ein $0 < \alpha < 1$, daß

$$\|P(z_0)\| \geq \alpha \frac{\Delta_0}{n+r},$$

dann ist

$$\begin{array}{ll} & \Psi(X(z_0), S_0) - \Psi(X_0, S_0) \\ = & q\ln(X \bullet S_0) - \ln(\det(XS_0)) - q\ln(X_0 \bullet S_0) + \ln(\det(X_0 S_0)) \\ = & q\ln(X \bullet (C - \sum_{i=1}^m u_i A_i) - \ln(\det(X)) - \ln(\det(S_0)) \\ & -q\ln(X_0 \bullet (C - \sum_{i=1}^m u_i A_i) + \ln(\det(X_0)) + \ln\det(S_0)) \\ = & q\ln(X \bullet C - u_0^\top b) - \ln(\det(X)) \\ & -q\ln(X_0 \bullet C - u_0^\top b) + \ln(\det(X_0)) \\ = & \Phi(X, z_0) - \Phi(X_0, z_0) \\ = & \bar{\Phi}(\bar{X}, \bar{x}_1, z_0) - \bar{\Phi}(I, \mathbb{1}, z_0) \\ \stackrel{\text{Korollar } 6.6.1}{\leq} & -(n+r)\beta \frac{\|P(z_0)\|}{-u^\top b + \mathbf{Spur}(\bar{C})} + \frac{\beta^2}{2(1-\beta)} \\ \leq & -\beta\alpha + \frac{\beta^2}{2(1-\beta)}. \end{array} \tag{6.31}$$

Andernfalls sind die Voraussetzungen von Lemma 6.6.1 erfüllt.

$$\begin{array}{l} n\ln X_0 \bullet S_1 - \ln\det X_0 S_1 \\ = n\ln\left(\frac{nX_0 \bullet S_1}{\Delta_1}\right) - \ln\left(\frac{n}{\Delta_1}\right)^n \det(X_0 S_1) \\ = n\ln n - \ln\det\left(\frac{nX_0 S_1}{\Delta_1}\right) \\ = n\ln n - \ln\det\left(\frac{nL_0^\top S_1 L_0}{\Delta_1}\right) \\ \stackrel{\text{Lemma } 6.4.4}{\leq} n\ln n - \mathbf{Spur}(\frac{nL_0^\top S_1 L_0}{\Delta_1}) + n + \frac{\|nL_0^\top S_1 L_0/\Delta_1 - I\|^2}{2(1 - \|nL_0^\top S_1 L_0/\Delta_1 - I\|)} \\ \stackrel{\text{Lemma } 6.4.1}{\leq} n\ln(X_0 \bullet S_0) - \ln\det(X_0 S_0) + \frac{\gamma^2}{2(1-\gamma)}, \end{array} \tag{6.32}$$

mit $\gamma = \alpha\sqrt{\frac{n+n^2/r}{n+n^2/r-\alpha}}$ (vgl. Lemma 6.6.1). Beachte, die Eigenwerte von $X_0 S_0$ sind alle reell und nicht-negativ und $\mathbf{Spur}(L_0^\top S_1 L_0) = X_0 \bullet S_1$.

Nach (6.28) haben wir

$$\frac{\Delta_1}{\Delta_0} \leq 1 - \frac{r}{n+r} + \frac{n}{n+r}\frac{\alpha}{\sqrt{n+n^2/r}}$$

wieder unter Ausnutzung von $\ln(x+1) \leq x$ erhalten wir also

$$r\ln\frac{X_0 \bullet S_1}{X_0 \bullet S_0} = r\ln\frac{\Delta_1}{\Delta_0} \leq \frac{r^2}{n+r}\left(-1 + \frac{\alpha}{\sqrt{r^2/n+r}}\right). \tag{6.33}$$

Addieren wir (6.32) und (6.33), so erhalten wir:

$$\Psi(X_0,S_1)-\Psi(X_0,S_0)\le\frac{r^2}{n+r}\left(-1+\frac{\alpha}{\sqrt{r^2/n+r}}\right)+\frac{\gamma^2}{2(1-\gamma)}. \tag{6.34}$$

Wählen wir nun $\alpha=0,55$ und $\beta=0,3$ so erhalten wir aus 6.31

$$\Psi(X(z_0),S_0)-\Psi(X_0,S_0)\le -0,10071\ldots.$$

Untersuchen wir andererseits den Ausdruck in (6.34) für $n\to\infty$, so erkennen wir, daß $\gamma\overset{n\to\infty}{\to}\alpha$. Also geht die rechte Seite von (6.34) gegen $-1+\frac{\alpha^2}{2(1-\alpha)}=-0,663\bar{8}$. Betrachten wir das Konvergenzverhalten für kleine n explizit, so erhält man für $n\ne 1$:

$$\Psi(X_0,S_1)-\Psi(X_0,S_0)\le -0,1.$$

□

Wir erhalten also den Algorithmus in Abbildung 6.1 und

Satz 6.6.2 *Seien X_0, u_0 und $S_0:=C-\mathbf{Mat}(\mathcal{A}^\top u_0)$ innere zulässige Punkte für das primale bzw. das duale semidefinite Programmierungsproblem. Sei $n\ne 1$ und $r=\lceil\sqrt{n}\rceil$ und $q=n+r$ in der primal-dualen Potentialfunktion Ψ und sei E gegeben mit $\Psi(X_0,S_0)\le\lceil\sqrt{n}\rceil E$. Der Algorithmus in Abbildung 6.1 liefert nach $k=10\lceil\sqrt{n}\rceil|\log\epsilon|$ Iterationen primale und duale Lösungen X_k, u_k und S_k mit*

$$C\bullet X_k-u_k^\top b<2^E\epsilon.$$

Beweis: Klar nach Satz 6.6.1 und Satz 6.5.1. □

6.7 Die Big M-Methode

Das prinzipielle Vorgehen bei inneren-Punkt Verfahren ist:

1. Starte mit zulässigem inneren Tripel (X_0,u_0,S_0).
2. Berechne nächsten Iterationspunkt mit Newton oder durch projektive oder affine Skalierungsmethoden.

Input:
Eine $(n \times n)$-Matrix X_0, zulässig und im Innern des primalen SDP;
ein m-Vektor u_0, im Innern zulässig für das duale SDP;
eine Konstante $\epsilon > 0$
Output:
Eine primal zulässige Lösung X und ein dual zulässiges u mit
$C \bullet X - u^\top b < \epsilon$.
Verfahren:
Setze $k := 0,\ \alpha := 0.55,\ \beta := 0.3,\ z_0 := u_0^\top b,\ r := \lceil \sqrt{n} \rceil$;
While $C \bullet X_k - u_k^\top b \geq \epsilon$ **do**
begin
Berechne $S(z_k)$ aus (6.25) und $P(z)$ aus 6.23;
If $\|P(z_k)\| \geq \alpha(C \bullet X_k - z_k)/(n + r)$ **then**
$(\bar{X}_{k+1}, \bar{x}_{k+1}) := (I, 1) - \beta P(z_k)$;
$X_{k+1} := \mathcal{T}^{-1}(\bar{X}_{k+1}, \bar{x}_{k+1})$;
$S_{k+1} := S_k; z_{k+1} := z_k$;
Else
$u_{k+1} := u(z)$ wie in 6.24;
$S_{k+1} := S(z_k)$ wie in und 6.25 ;
$X_{k+1} := X_k; z_{k+1} := u_{k+1}^\top b$;
$k := k + 1$;
end.

Abbildung 6.1: Ein projektiver Potentialreduktionsalgorithmus

3. Fahre solange fort, bis $X \bullet S$ klein genug.

Zunächst muß man also die Aufgabenstellung: „Finde einen zulässigen inneren Punkt“ lösen.

Betrachten wir dazu das SDP in Standardform und sein Duales

$$\begin{array}{ll} \min & C \bullet X \\ \text{unter} & \mathcal{A}\mathbf{vec}(X) = b \\ & X \succeq 0 \end{array} \qquad\qquad \begin{array}{ll} \max & u^\top b \\ \text{unter} & \mathcal{A}^\top u + S = C \\ & S \succeq 0 \end{array}$$

Diese Aufgabenstellung kann unzulässig oder unbeschränkt (dual unzulässig) sein. Um eine primal und dual zulässige Startlösung zu finden, betrachten wir ein modifiziertes System. Seien M, N große positive Zahlen (dazu später Bemerkungen) und $X_0, S_0 \succ 0, u_0$ beliebig. Betrachte folgendes primal duale Paar:

$$\begin{array}{ll} \min & C \bullet X + Mx_1 \\ \text{unter} & \mathcal{A}\mathbf{vec}(X) + (b - \mathcal{A}\mathbf{vec}(X_0))x_1 = b, \\ & (\mathcal{A}^\top u_0 + S_0 - C) \bullet X + x_2 = N, \\ & X \succeq 0, x_1, x_2 \geq 0, \end{array} \tag{6.35}$$

$$\begin{array}{ll} \max & u^\top b - Nz_1 \\ \text{unter} & \mathcal{A}^\top u + S + (C - \mathcal{A}^\top u_0 - S_0)z_0 = C \\ & (b - \mathcal{A}\mathbf{vec}(X_0))u + z_2 = M, \\ & S \succeq 0, z_1, z_2 \geq 0. \end{array} \tag{6.36}$$

Dann ist

$$X = X_0, x_1 = 1, x_2 = N - (\mathcal{A}^\top y_0 + S_0 - C) \bullet X,$$

falls N groß genug ist, zulässig. Für das Duale haben wir die zulässige Lösung:

$$u = u_0, S = S_0, z_1 = 1, z_2 = M - (b - \mathcal{A}\mathbf{vec}X_0)u,$$

falls M groß genug ist. Ferner sollten M und N so gewählt sein, daß im Falle der Existenz einer zulässigen Lösung, x_1, bzw. z_1 im Optimum der Hilfsprobleme 0 sind.

Um dies zu garantieren, müßte man allerdings sicherstellen, daß

$$N \geq (I - C) \bullet X^*$$
$$M \geq \sum y_i^*(b_i - \mathbf{Spur}(A_i)$$

für alle zulässigen X^*, y^*. Im Falle der linearen Programmierung ist die Kodierungslänge für eine mögliche Antwort durch ein Polynom in den Eingangsdaten beschränkt. Deshalb ist das Auffinden einer ersten zulässigen Lösung, falls diese existiert, stets möglich. Man kann dann am Optimalwert des Hilfsprogramms ablesen, ob das Ausgangsproblem eine zulässige Lösung besitzt.

In der semidefiniten Programmierung scheint dies nicht der Fall zu sein. Betrachten wir zum Beispiel

$$\begin{array}{rr} \min & x_n \\ \text{unter} & x_1 = 2 \\ & x_i^2 \geq x_{i-1}. \end{array}$$

Die Ungleichungsbedingung kann man als $\begin{pmatrix} x_i & x_{i-1} \\ x_{i-1} & 1 \end{pmatrix} \succeq 0$ kodieren. Offensichtlich ist im Optimum $x_n = 2^{2^n}$, die Kodierungslänge dieser Zahl ist aber exponentiell in der Kodierungslänge der Eingangsdaten.

Auf diese Weise wird man also nicht zu einem polynomiellen Zulässigkeitstest kommen. Die Schwierigkeiten scheinen aber prinzipieller Natur zu sein. Der bisher bekannte Ausweg besteht darin, daß man von einer Probleminstanz annimmt, daß sie mit einer Schranke für die mögliche Lösungsgröße versehen ist. Dann kann man passende Parameter M, N bestimmen und (6.35) und (6.36) lösen. Falls in dieser Lösung $x_1 \neq 0$ oder $z_1 \neq 0$ sind, so hat das SDP, bzw. sein Duales, keine Lösung innerhalb der vorgegebenen Schranken. Es ist entweder unzulässig oder alle zulässigen Lösungen liegen außerhalb der vorgegebenen Schranken.

Kapitel 7

Kombinatorische Optimierung

Wir werden zwei Typen von Anwendungen der semidefiniten Programmierung in der Kombinatorischen Optimierung kennenlernen. (Nur MAX-CUT im Rahmen der Vorlesung.)

Im ersten Fall betrachten wir ein NP-schweres Optimierungsproblem, bei dem die semidefinite Programmierung es ermöglicht, gute (approximative) Lösungen zu finden und Schranken für das Optimum zu errechnen.

7.1 MAX-CUT

7.1.1 Definition und Komplexität

Als erste Anwendung wollen wir eine Approximation für das MAX-CUT Problem kennenlernen, die die bis dahin bekannte Schranke für nicht-negative Gewichtsfunktionen von garantierten 50% auf garantierte 87% verbessert.

Definition 7.1.1 *Sei $G = (V, E)$ ein Graph. Eine Kantenmenge $C \subseteq E$ heißt* Kantenschnitt *oder* Cut*, wenn es eine Partition der Knoten $V = V_1 \dot\cup V_2$ gibt mit*

$$C = \{e \in E \mid e \textit{ hat je einen Endknoten in } V_i,\ i = 1, 2\}.$$

Sei $w : E \to \mathbb{R}^+$ eine Kantengewichtsfunktion. Die Aufgabenstellung

$$\begin{array}{ll} \max & \sum_{e \in C} w_e \\ & C \textit{ ist Kantenschnitt von } G \end{array}$$

bezeichnen wir als das MAX-CUT Problem..

Beim MAX-CUT Problem können wir stets davon ausgehen, daß in G Kanten zwischen allen Knoten existieren, indem wir für nicht existente Kanten $w_e = 0$ setzen.

Sei $W \subseteq V$. Der knoteninduzierte Teilgraph $G_W(W, F)$ *ist dann der Graph mit der Kantenmenge*

$$F = \{(i, j) \in E \mid i, j \in W\}.$$

Das MAX-CUT Problem ist NP-vollständig. Dies bedeutet im wesentlichen, daß kein Algorithmus mit polynomial beschränkter Laufzeit bekannt ist, der das Problem exakt löst und daß allgemein angenommen wird, daß es einen solchen auch nicht gibt.

Für eine genaue Definition der komplexitätstheoretischen Begriffe verweisen wir auf Standardlehrbücher der theoretischen Informatik oder die ersten Kapitel von Garey and Johnson: „Computers and Intractability, A guide to the Theory of NP-Completeness", *Freeman and Co*, 1979.

Für unsere Zwecke genügt folgendes informelle Verständnis der Begriffe:

Die Klasse der P-Probleme besteht aus den Fragestellungen, für die ein Algorithmus existiert, der die Frage in polynomieller Zeit beantwortet. Üblicherweise betrachtet man in der Klasse P nur Fragestellungen, bei denen die Antwort „Ja" oder „Nein" lautet, sogenannte „Entscheidungsprobleme".

Die Klasse der NP-Probleme besteht aus den Fragestellungen, bei denen es im Falle einer „Ja"-Antwort einen kurzen Beweis für diese positive Antwort gibt. Im Falle des MAX-CUT Problems lautet diese Fragestellung:

Input: ein Graph $G = (V, E)$ mit Kantengewichtsfunktion $w : E \to \mathbb{N}$ und eine Zahl $K \in \mathbb{N}$.

Frage: Gibt es einen Kantenschnitt C in G mit $\sum_{e \in C} w_e \geq K$?

Offensichtlich kann man diese Frage bei Auswertung aller Kantenschnitte entscheiden. Da die Menge der Kantenschnitte aber im wesentlichen so groß ist wie die Menge aller Teilmengen von V, ist eine solche Auswertung in polynomieller Zeit nicht möglich.

Lautet die Antwort auf obige Frage hingegen „JA" so kann ein überirdisches Wesen uns davon leicht überzeugen, indem es uns einen Kantenschnitt mit Gewicht $\geq K$ vorlegt. Dieser ist ein kurzer Beweis für die positive Antwort, da wir offensichtlich in polynomieller Zeit das Gewicht des Kantenschnitts nachrechnen können.

Ein Problem A in NP ist NP-vollständig, falls ein Algorithmus, der die zu A gehörige Frage in polynomieller Zeit entscheidet, sich für jedes beliebige andere Problem B in NP so modifizieren läßt, daß er auch Problem B in polynomieller Zeit löst. Wie gesagt, diese Sprechweise ist nur informell zu verstehen.

Wie kann man die Existenz einer dualen Fragestellung in diesen Zusammenhang einordnen. Dazu betrachten wir – wieder beispielhaft – die negierte Fragestellung zu obigem Problem.

Input: ein Graph $G = (V, E)$ mit Kantengewichtsfunktion $w : E \to \mathbb{Z}$ und eine Zahl $K \in \mathbb{Z}$.

Frage: Gibt es keinen Kantenschnitt C in G mit $\sum_{e \in C} w_e \geq K$?

In diesem Falle haben wir einen kurzen Beweis für eine negative Antwort, deshalb sagen wir das Problem gehört zur Klasse co-NP.

Gibt es nun für eine Fragestellung einen Algorithmus, der das Problem in polynomieller Zeit löst, so kann dieser Algorithmus sowohl als kurzer Beweis für eine positive wie auch für eine negative Antwort dienen. Also gilt:

$$P \subseteq NP \cap co\text{-}NP.$$

Es herrscht allgemein die Überzeugung, daß in $NP \cap co$-NP „nicht viel mehr“als P liegt.

Betrachten wir nun die Lineare Programmierung als NP-Fragestellung:

Input: eine Matrix $A \in \mathbb{Q}^{m \times n}$, zwei passende Vektoren $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$ und eine Zahl $K \in \mathbb{Q}$.

Frage: Gibt es ein $x \geq 0$ mit $Ax = b$ und $c^\top x \leq K$?

und als co-NP Fragestellung

Input: eine Matrix $A \in \mathbb{Q}^{m \times n}$, zwei passende Vektoren $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$ und eine Zahl $K \in \mathbb{Q}$.

Frage: Gibt es kein $x \geq 0$ mit $Ax = b$ und $c^\top x \leq K$?

Nach dem starken Dualitätssatz der linearen Programmierung (Spezialfall von Satz 3.5.1) ist die letzte Fragestellung aber äquivalent zu

Input: eine Matrix $A \in \mathbb{Q}^{m \times n}$, zwei passende Vektoren $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$ und eine Zahl $K \in \mathbb{Q}$.

Frage: Gibt es $u \in \mathbb{Q}^m$ mit $A^\top u \geq c$ und $u^\top b > K$.

Für eine positive Antwort dieser Fragestellung gibt es aber wieder einen kurzen Beweis.

Wir haben damit plausibel gemacht, daß die Existenz einer starken Dualitätsaussage für ein Problem die Zugehörigkeit dieses Problems zu $NP \cap co\text{-}NP$ bedeutet. Dies wird als Indikator dafür angesehen, daß das Problem gutartig ist und wahrscheinlich in P liegt.

Es wird allgemein vermutet, daß kein NP-vollständiges Problem in $NP \cap co\text{-}NP$ liegt.

Deshalb beschränkt man sich bei NP-vollständigen Problemen auf die Suche nach unteren und oberen Schranken.

7.1.2 Eine Heuristik für MAX-CUT

Im Falle des $MAX - CUT$-Problems liefert jeder Kantenschnitt eine untere Schranke. Ein Verfahren, daß mit „gesundem Menschenverstand“ eine Lösung von möglichst großem Gewicht bestimmt, nennt man eine *Heuristik.*

Beispiel 7.1.1 *Wir betrachten eine heuristische Vorgehensweise zur Bestimmung eines „schweren“ Cuts, der mindestens 50% vom Optimum garantiert. Seien die Knoten von G durchnummeriert $V = \{v_1, \ldots, v_n\}$.*

$V_1 = v_1,\ V_2 = \emptyset$
for $i = 2$ **to** n **do**
  if $\sum_{v_j \in V_1} w(v_i v_j) \leq \sum_{v_j \in V_2} w(v_i v_j)$ **then do**
    $V_1 = V_1 \cup \{v_i\}$
  else do
    $V_2 = V_2 \cup \{v_i\}$

Da in jedem Schritt mindestens so viel Gewicht zum Cut hinzukommt wie in den Graphen G_{V_1} und G_{V_1} verschwindet, hat der generierte Cut ein Gewicht von mindestens $\frac{1}{2}\sum_{e \in E} w_e$.

Betrachte dieses Vorgehen an dem Graphen in Abbildung 7.1 bzgl. Kantengewicht konstant 1. *Die Heuristik produziert eine Lösung mit Gewicht* $n-2+1$, *das Optimum liegt hingegen bei* $2(n-2)$. *Also kann die Heuristik auch nicht mehr als 50% garantieren.*

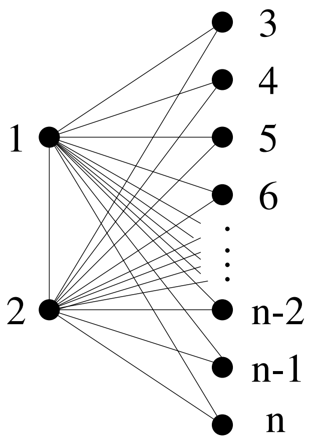

Abbildung 7.1: Worst-Case Beispiel für die Knotenheuristik

Es war bis vor kurzem kein Verfahren bekannt, das wesentlich mehr als 50% garantierte.

7.1.3 87% vom Optimum

Sei ab jetzt vorausgesetzt, daß die Gewichtsfunktion nicht-negativ ist. MAX-CUT kann man als das folgende ganzzahlige Programm formulieren, das wir als MC^* bezeichnen.

$$\begin{array}{lll} (\mathbf{MC}^*) & \max & \frac{1}{2}\sum_{i<j} w_{ij}(1-x_i x_j) \\ & \text{unter} & x_i \in \{-1,1\} \end{array}$$

Die Werte von x_i geben an, auf welcher Seite der Partition der Knoten i zu liegen kommt.

Wir können allerdings die Menge $\{-1,1\}$ auch als $S_0 = \{x \in \mathbf{R}^1 \mid \|x\| = 1\}$ auffassen. Stellen wir uns die Gewichte als Federn vor, die die Knoten auseinanderdrücken, können wir das MAX-CUT Problem auch auffassen als: Verteile die Knoten so auf der S_0, daß die „wirkende“ Federkraft minimiert wird.

Letzteres Problem können wir auch auf der S_{n-1} formulieren:

$$\begin{array}{lll}(\mathbf{MC1}) & \max & \frac{1}{2}\sum_{i<j} w_{ij}(1 - v_i^\top v_j) \\ & \text{unter} & \|v_i\| = 1\end{array}$$

Da wir immer noch die v_i antipodisch wählen können, ist $MC1$ eine Relaxation von MC^* und $MC1 \geq MC^*$.

Im Optimum werden Knotenpaare deren Kantengewicht „hoch" ist, einen großen Winkel $\arccos(v_i^\top v_j)$ bilden. Zerschneiden wir nun die S_{n-1} mit einer linearen Hyperebene, so generieren wir im Normalfall schon einen „guten" Cut. Genauer werden wir zeigen, daß der Erwartungswert eines solchen Cuts bezüglich einer zufälligen Hyperebene 87% des Optimums liefert.

Dafür betrachten wir zunächst eine andere Relaxation von MC^*:

$$\begin{array}{lll}(\mathbf{MC2}) & \max & \sum_{i<j} w_{ij}\frac{\arccos(v_i^\top v_j)}{\pi} \\ & \text{unter} & \|v_i\| = 1.\end{array}$$

Mit demselben Argument wie oben erhalten wir $MC2 \geq MC^*$.

Sei nun $r \in S_{n-1}$ zufällig gewählt (bzgl. Gleichverteilung). Dann ist der Erwartungswert des von $H(r) = \{x \in \mathbb{R}^n \mid r^\top x = 0\}$ erzeugten Cuts:

$$\sum_{i<j} w_{ij} Pr[v_i|v_j],$$

wobei $Pr[v_i|v_j]$ die Wahrscheinlichkeit ist, daß v_i und v_j auf verschiedenen Seiten der Hyperebene $H(r)$ liegen. Diese Wahrscheinlichkeit hängt aber nur von dem Winkel zwischen v_i und v_j ab.

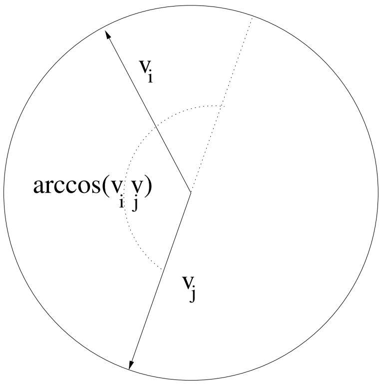

Abbildung 7.2: Wie wahrscheinlich trenne ich v_i und v_j?

Anhand von Abbildung 7.2 erkennt man, daß $Pr[v_i|v_j] = \frac{\arccos(v_i^\top v_j)}{\pi}$ ist.

Sind die Vektoren $v_1, \ldots, v_n$ gegeben, so ist demnach der Erwartungswert eines Cuts, der von einer zufälligen Hyperebene erzeugt wird, gleich dem Zielfunktionswert von $MC2$ bzgl. dieser Vektoren. Da dieser Wert der Erwartungswert ist, muß es speziell einen Cut geben, der mindestens diesen Wert erreicht. Dies gilt auch im Optimum von $MC2$. Also ist $MC2 \leq MC^*$ und folglich $MC2 = MC^*$. Leider ist $MC2$ kein gutartiges Programm und wahrscheinlich gibt es dafür auch kein effizientes Verfahren (es sei denn $P = NP$).

Da nun aber $MC1 \geq MC^* = MC2$ ist, muß im Optimum $v_1, \ldots, v_n$ von $MC1$ gelten:

$$MC1 \geq MC^* \geq \sum_{i<j} \frac{\arccos(v_i^\top v_j)}{\pi} w_{ij}$$

und der letzte Ausdruck ist gleich dem Erwartungswert eines Cuts.

Abschließend zeigen wir nun noch, daß

$$\sum_{i<j} \frac{\arccos(v_i^\top v_j)}{\pi} w_{ij} \geq 0,87 MC1,$$

womit wir bewiesen hätten, daß $0,87MC1$ eine untere Schranke für MC^* ist, die mindesten 87% des Optimums liefert.

Betrachte dazu

$$\alpha := \min_{-1 \leq y \leq 1} \frac{2\arccos(y)}{\pi(1-y)} \sim 0,87.$$

Dies impliziert $\alpha \frac{1-y}{2} \leq \frac{\arccos(y)}{\pi}$ und somit

$$\sum_{i<j} w_{ij} \frac{\arccos(v_i^\top v_j)}{\pi} \geq \alpha \sum_{i<j} w_{ij} \frac{1 - v_i^\top v_j}{2},$$

da $w \geq 0$.

Also haben wir, daß der Erwartungswert des Cuts, der durch einen zufälligen Schnitt durch die Optimallösung von $MC1$ erzeugt wird, $0,87MC^* \geq 0,87MC^*$ liefert.

Nun ist aber $MC1$ ein SDP. Denn betrachten wir

$$\begin{array}{ll} \max & \frac{1}{2}\sum_{i<j} w_{ij}(1 - X_{ij}) \\ \text{unter} & X_{ii} = 1, \\ & X \succeq 0, \end{array}$$

so können wir aus jeder zulässigen Lösung X die Wurzel ziehen. Die Vektoren v_i erhält man dann, indem man die Spalten von $X^{1/2}$ betrachtet. Da auf der Diagonalen von X lauter 1-en stehen, sind die Spalten von $X^{1/2}$ Einheitsvektoren und deren Skalarprodukte sind gerade die Einträge X_{ij}. Umgekehrt liefern n Einheitsvektoren, wenn man diese Prozedur rückwärts betrachtet, ein zulässiges X.

Ende des prüfungsrelevanten Stoffes.

7.2 Nachschlag zur Sandwichzahl

7.2.1 Knoten Packen

Nun wollen wir uns aber wieder dem Parameter $\theta(G)$ zuwenden.

Definition 7.2.1 *Sei $G = (V, E)$ ein Graph. Ein Teilmenge der Knoten S heißt* stabile Menge *(oder* unabhängige Menge, stable set, independent set*), wenn für je zwei Knoten s_1, s_2 die Kante $(s_1, s_2) \notin E$ nicht existiert. Die konvexe Hülle der Inzidenzvektoren von stabilen Mengen in G definiert das* stable set polytope *oder auch* vertex packing polytope

$$STAB(G) = \mathit{conv}\{\mathbb{1}_S \in \mathbb{R}^{|V|} \mid S \text{ ist stabile Menge von } G\}.$$

Das Problem eine größte stabile Menge in einem allgemeinen Graphen zu finden, ist NP-schwer. Das gleiche gilt dann natürlich für folgende gewichtete Variante.

$$\begin{array}{lll} \textbf{IP:} & \max & w^\top x \\ & & x_i + x_j \leq 1 \quad \forall (i,j) \in E \\ & & x_i \in \{0,1\}. \end{array}$$

Jeder zulässige Vektor dieses Programms muß offensichtlich Inzidenzvektor einer stabilen Menge von G sein. Eine Möglichkeit eine Näherung für die Optimallösung zu bestimmen ist es, die Ganzzahligkeitsbedingung $x_i \in \{0, 1\}$ durch die Bedingung $0 \leq x_i \leq 1$ zu ersetzen.

$$\begin{array}{lll} \textbf{LP:} & \max & w^\top x \\ & & x_i + x_j \leq 1 \quad \forall (i,j) \in E \\ & & 0 \leq x_i \leq 1. \end{array}$$

Bei welchen Graphen liefert dies die Optimallösung?

Definition 7.2.2 *Sei $G = (V, E)$ ein Graph. G heißt* bipartit*, wenn es eine Partition der Knotenmenge $V = V_1 \dot\cup V_2$ gibt, so daß jede Kante einen Endknoten in V_1 und den anderen in V_2 hat. Ein* Weg *in einem Graphen ist eine Knotenfolge $v_0, \ldots, v_{n-1}, v_n$ mit $v_i = v_j \Leftrightarrow i = j$ und $\forall i = 1, \ldots n : (v_{i-1}v_i) \in E$. Ein* Kreis *in einem Graphen ist eine Knotenfolge $v_0, \ldots, v_{n-1}, v_n$ mit $v_i = v_j \Leftrightarrow i = j(mod)n$ und $\forall i = 1, \ldots n : (v_{i-1}v_i) \in E$.*

Proposition 7.2.1 *Ein Graph ist bipartit genau dann, wenn er keinen ungeraden Kreis enthält.*

Beweis: Ein ungerader Kreis ist sicherlich unverträglich mit einer Bipartition. Habe also G keinen ungeraden Kreis und sei $v \in V$ fest. Ist dann $w \in V$ beliebig, so haben entweder alle Wege von v nach w ungerade Länge oder alle haben gerade Länge. Diese Unterscheidung liefert die gewünschte Partition. $\square$

Satz 7.2.1 *Die LP-Relaxation von* **IP** *ist ganzzahlig genau dann, wenn G bipartit ist.*

Beweis: Sei G bipartit und x eine Ecke des zulässigen Bereichs von **LP**. Angenommen es gäbe einen Index v mit $0 < x_v < 1$. Betrachte nun den Untergraphen G' von G, der die Kanten von G enthält, auf denen gilt $x_i + x_j = 1$. In diesem Untergraphen betrachten wir die Knoten V', die von v aus auf einem Weg erreicht werden können. Für alle $w \in V'$ muß gelten $0 < x_w < 1$. Sei nun $\mathbb{I}_1$ der charakteristische Vektor von $V_1 \cap V'$ und $\mathbb{I}_2$ der charakteristische Vektor von $V_2 \cap V'$ dann gibt es offensichtlich ein $\epsilon > 0$, so daß $x + \epsilon\mathbb{I}_1 - \epsilon\mathbb{I}_2$ und $x - \epsilon\mathbb{I}_1 + \epsilon\mathbb{I}_2$ beide zulässig sind, im Widerspruch dazu, daß x eine Ecke ist.

Sei nun G nicht bipartit und $v_0, \ldots, v_n$ ein ungerader Kreis und x der Vektor, der auf allen diesen Knoten den Wert $\frac{1}{2}$ annimmt und sonst 0 ist. Das Gesamtgewicht der Knoten auf diesem Vektor ist $\frac{n}{2}$. Da n ungerade ist, kann aber jede stabile Menge aus den Knoten des Kreises höchstens $\frac{n-1}{2}$ Knoten enthalten. Somit kann x nicht Konvexkombination von zulässigen 0-1-Vektoren sein, die alle Inzidenzvektoren stabiler Mengen sind. $\square$

7.2.2 „Facetten jagen“

Begeben wir uns auf die Suche nach weiteren gültigen Ungleichungen für $STAB(G)$.

Clique Ungleichungen: Sei $K \subseteq V$ eine Clique von G mit charakteristischem Vektor $\mathbb{1}_K$. Offensichtlich gilt dann für jeden Vektor

$$x \in STAB(G) : x^\top \mathbb{1}_K \leq 1.$$

Die Menge aller Vektoren, die alle Clique Ungleichungen erfüllen bezeichnen wir mit $QSTAB(G)$. Dieses Polyeder nennen wir auch *fractional vertex packing polytope*.

$$QSTAB(G) := \{x \in \mathbf{R}^{|V|} \mid x^\top \mathbb{1}_K \leq 1 \text{ für alle Cliquen } K \text{ von } G\}.$$

Odd Hole Ungleichungen: Ist $C = v_0, \ldots, v_{2k+1} = v_0$ ein ungerader Kreis in G und hat der Kreis keine Sehne, d.i. eine Kante (v_i, v_j) mit $j \neq i \pm 1 (\text{mod}) 2k + 1$, so nennt man C ein *odd hole* oder einen *ungeraden induzierten Kreis*. Eine stabile Menge kann ein odd hole in höchstens k Knoten treffen. Also gilt für charakteristische Vektoren $\mathbb{1}_C$ von odd holes der Länge $2k + 1$.

$$x \in STAB(G) \Rightarrow x^\top \mathbb{1}_C \leq k.$$

Wieder setzen wir

$$CSTAB(G) :=$$
$$\{x \in \mathbf{R}^{|V|} \mid x^\top \mathbb{1}_C \leq k \text{ für alle odd holes } C \text{ der Länge } 2k + 1\}.$$

Odd Antihole Ungleichungen: Ist $C = v_0, \ldots, v_{2k+1}$ ein odd hole, so nennt man den Komplementgraphen $\bar{C}$ ein *odd antihole*. Eine stabile Menge kann ein odd antihole in höchstens 2 Knoten treffen. Also gilt für charakteristische Vektoren $\mathbb{1}_{\bar{C}}$ von odd antiholes.

$$x \in STAB(G) \Rightarrow x^\top \mathbb{1}_{\bar{C}} \leq 2$$

und wir setzen

$$\bar{C}STAB(G) :=$$
$$\{x \in \mathbf{R}^{|V|} \mid x^\top \mathbb{1}_{\bar{C}} \leq 2 \text{ für alle odd antiholes } \bar{C}\}.$$

Ungerade Räder: Eine Knotenmenge W von G heißt ein *ungerades Rad* oder *odd wheel*, wenn es ein $v \in W$ gibt, so daß $W \setminus v$ ein odd hole ist und $\forall w \in W \setminus v : (v, w) \in E$ eine Kante ist. Der Knoten v heißt dann die *Nabe* des Rades. Sei $W = v_1, \ldots, v_{2k}$ ein Rad mit Nabe v_{2k}. Betrachten wir die Ungleichung

$$\sum_{i=1}^{2k-1} x_{v_i} + (k-1) x_{v_{2k}} \leq k - 1.$$

Jede stabile Menge kann nur entweder die Nabe, oder höchstens $k-1$ Knoten des odd hole enthalten. Also ist auch diese Ungleichung gültig für $STAB(G)$. Definiere

$$WSTAB(G) :=$$
$$\{x \in \mathbf{R}^{|V|} \mid \sum_{i=1}^{2k-1} x_{v_i} + (k-1)x_{v_{2k}} \leq k-1 \text{ für alle odd wheels}\}.$$

Warum das alles? Alle diese Ungleichung definieren (evtl. in Kombinationen) immer „bessere" Annäherungen für $STAB(G)$. Wir werden aber im folgenden eine konvexe Menge $\Theta(G)$ kennenlernen, die $STAB(G)$ enthält und alle diese Ungleichungen erfüllt, d.h. sie nähert $STAB(G)$ besser an als alle oben eingeführten Polyeder. Zusätzlich kann man über diese Menge mit Hilfe der semidefiniten Programmierung optimieren, was bei obigen Mengen i.a. nicht der Fall ist.

Besonders interessant ist allerdings der Fall, in dem alle diese Annäherungen zusammenfallen. Die Graphen, bei denen dies der Fall ist, nennt man perfekte Graphen. Eingeführt wurden sie von Berge (1963).

Definition 7.2.3 *Ein Graph ist* perfekt, *wenn*

$$STAB(G) = QSTAB(G).$$

7.2.3 Lifting

Die Menge $\Theta(G)$, die wir in wenigen Augenblicken definieren werden, kann man auffassen als ein Lifting von $STAB(G)$, obwohl dies geometrisch nicht ganz leicht einzusehen ist.

Die prinzipielle Idee des Liftings hingegen ist einfach. Sei $P \subseteq \mathbf{R}^n$ ein Polytop mit m Facetten (wesentlichen Ungleichungen). Projezieren wir dieses Polytop auf einen niederdimensionalen Raum, so kann sich die Zahl der Facetten drastisch erhöhen.

Hat man nun umgekehrt ein Polyeder mit sehr vielen Facetten, so kann man versuchen durch Liften in einen höherdimensionalen Raum, ein Polyeder mit einer einfacheren Beschreibung zu finden.

Beispiel 7.2.1 (Das Paritätspolytop) *Sei*

$$PP = \mathit{conv}\left\{x \in \{0,1\}^n \mid \mathbb{1}^\top x \equiv 1(\mathit{mod} 2)\right\}.$$

PP hat exponentiell viele Facetten. Andererseits ist dieses Polyeder nicht kompliziert. Optimierung ist trivial:

$$\begin{array}{ll} \max & w^\top x \\ \textit{unter} & x \in PP. \end{array}$$

Wir nehmen alle positiven Einträge von w und passen die Parity an, durch Hinzunahme des größten negativen Eintrags oder Weglassen des kleinsten positiven. Dies spiegelt sich darin wieder, daß das Polytop einen einfachen Lift hat.

Wir betrachten zunächst für jedes ungerade $k \leq n$ (vgl. Lemma 5.1.1)

$$S_k := \{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} = k,\ 0 \leq x_i \leq 1\}.$$

Jeder Punkt $x \in PP$ is Konvexkombination von Punkten dieser Polyeder, also gibt es $\alpha_1, \ldots, \alpha_l$ mit $l = \lfloor \frac{n}{2} \rfloor$, $0 \leq \alpha_i$ und $\sum_{i=1}^{l} \alpha_i = 1$ sowie $y_1, \ldots y_l$ mit $y_i \in Z_{2i-1}$ mit $x = \sum_{i=1}^{l} \alpha_i y_i = \sum_{i=1}^{l} z_i$. Daraus leiten wir folgendes lifting von PP her.

$$\begin{array}{ll} \sum_{i=1}^{l} z_i = x & \\ z_i^\top 1 = \alpha_i(2i-1) & \textit{für } i = 1, \ldots l \\ \sum_{i=1}^{l} \alpha_i = 1 & \\ \alpha_i \geq 0 & \\ 0 \leq z_i \leq \alpha_i & \end{array}$$

Die Anzahl der Ungleichungen in dieser Beschreibung ist polynomiell. Durch Einführung neuer Variablen (auch nur quadratisch viele) wurde das Polyeder geliftet.

Betrachten wir das Stabilitätsproblem für perfekte Graphen. Hier ist $QSTAB = STAB$ und somit ist das stable set polytope "noch relativ einfach". Allerdings hat auch ein perfekter Graph i.a. eine exponentielle Anzahl Cliquen. Ein Lift zu einem Polyeder mit kleiner Beschreibung ist nicht bekannt. Aber wir werden einen Lift zu einem kleinen SDP kennenlernen. Sei

$$S^{k-1} := \{x \in \mathbb{R}^k \mid \|x\| = 1\}$$

die $k-1$-dimensionale Standardsphäre.

Definition 7.2.4 *Sei $G = (V, E)$ ein Graph. Eine* orthonormale Darstellung auf der S^{k-1} *für G ist eine Menge von Vektoren $u_1, \ldots, u_n \in \mathbb{R}^k$ ($n = |V|$ von jetzt an) mit folgenden Eigenschaften:*

1. $\forall i = 1, \ldots n : \ \|u_i\| = 1,$
2. $\forall (i, j) \notin E : u_i \perp u_j.$

Mit $OND(G,k)$ bezeichnen wir die Menge aller orthonormalen Darstellungen von G auf der S^{k-1}.

Sei nun S eine stabile Menge und $u_1, \dots, u_n \in \mathbf{R}^k$ eine orthonormale Darstellung von G und $c \in \mathbf{R}^k$ ein beliebiger Vektor der Länge 1. Da die u_i, die zu S gehören orthonormal sind ist

$$\sum_{i \in S} c^\top u_i \leq 1.$$

Deswegen definieren wir folgendes Objekt:

$$\Theta(G) = \{x \in \mathbf{R}^n \mid \sum_{i=1}^{n} (c^\top u_i)^2 x_i \leq 1 \text{ für alle } OND(G,k), c \in S^{k-1}, k \text{ beliebig}\}.$$

Offensichtlich ist $\Theta(G)$ eine konvexe Menge. Nach unseren Vorüberlegungen gilt

$$STAB(G) \subseteq \Theta(G).$$

Andererseits haben wir auch

$$\Theta(G) \subseteq QSTAB(G),$$

denn ist $x \in \Theta(G)$ und $K \neq \emptyset$ eine Clique in G, so wählen wir eine orthonormale Darstellung wie folgt: Sei $c \in \mathbf{R}^n$ ein Vektor der Länge 1. Dann setzen wir $u_i = c$ für alle $i \in K$. Für alle $j \notin K$ wählen wir Vektoren der Länge 1, so daß $u_i \perp u_j$ für alle $j \neq i$. Dann ist $\sum_{i \in K} x_i = \sum_{i=1}^{n} (u^\top u_i)^2 x_i \leq 1$, also $x \in QSTAB(G)$.

Wir haben also bewiesen

Satz 7.2.2 *Sei $G = (V, E)$ ein Graph. Dann gilt:*

$$STAB(G) \subseteq \Theta(G) \subseteq QSTAB(G).$$

Wir definieren über diesen Mengen drei Parameter

$$\begin{aligned}
\alpha(G,w) &= \max\{w^\top x \mid x \in STAB(G)\}, \\
\theta(G,w) &= \max\{w^\top x \mid x \in \Theta(G)\}, \\
\alpha^*(G,w) &= \max\{w^\top x \mid x \in Q\text{-}STAB(G)\},
\end{aligned}$$

und nennen $\alpha(G,w)$ die *gewichtete Unabhängigkeitszahl* von G.

Nach dem letzten Satz gilt

$$\alpha(G,w) \leq \theta(G,w) \leq \alpha^*(G,w).$$

7.2.4 Die vielen Gesichter von θ

Der Parameter $\theta(G)$ und $\theta(G,w)$ wirft einige Fragen auf. So ist bis heute unklar, ob es irgendwelche Beziehungen zwischen $\omega(G), \chi(G)$ und $\theta(G)$ über die Sandwichbeziehung hinaus gibt. Deswegen wollen wir uns noch etwas mit $\theta(G,w)$ und $\Theta(G)$ beschäftigen.

Sei $G = (V, E)$ ein Graph und w ein nicht-negativer Knotengewichtungsvektor. Sei $\sqrt{w}$ der Vektor mit den Komponenten $(\sqrt{w})_i = \sqrt{w_i}$ und $W := \sqrt{w}\sqrt{w}^\top$. Wir definieren folgende Parameter

$$\begin{aligned}
\theta_1(G,w) &= \min\left\{\max\{\frac{w_i}{(c^\top v_i)^2} \mid i \in V\} \mid c \in S^{k-1},\ \{v_1,\dots,v_n\} \in OND(G,k)\right\}\\
\theta_2(G,w) &= \min\{\lambda_1(X+W) \mid X \text{ symm.}, X_{ij} = 0 \text{ falls } (i,j) \notin E \text{ oder } i = j\}\\
\theta_3(G,w) &= \max\{W \bullet Y \mid \mathbf{Spur}(Y) = 1,\ Y \succeq 0,\ Y_{ij} = 0 \text{ falls } (i,j) \in E\}\\
\theta_4(G,w) &= \max\left\{\sum_{i\in V}(d^\top z_i)^2 w_i \mid d \in S^{k-1},\ \{z_1,\dots,z_n\} \in OND(\bar{G},k)\right\}.
\end{aligned}$$

Satz 7.2.3

$$\theta(G,w) = \theta_1(G,w) = \theta_2(G,w) = \theta_3(G,w) = \theta_4(G,w).$$

Wir zeigen $\theta(G,w) \le \theta_1(G,w) \le \theta_2(G,w) \le \theta_3(G,w) \le \theta_4(G,w) \le \theta(G,w)$.

Ist $w = 0$, so ist $\theta_2(G,w) = 0$, da $\mathbf{Spur}(X) = 0 = \sum_{i=1}^n \lambda_i(X)$ und die Nullmatrix, die Bedingungen an X erfüllt. Somit ist offensichtlich die Behauptung wahr für $w = 0$. Sei also $w \neq 0$.

$\theta(G,w) \le \theta_1(G,w)$: Sei $x \in \Theta(G)$ ein Vektor mit $w^\top x = \theta(G,w)$ (beachte $\Theta(G)$ ist kompakt). Sei $\{v_1,\dots,v_n\} \in OND(G,k)$ und $c \in S^{k-1}$. Dann ist

$$\begin{aligned}
\theta(G,w) &= \sum_{i\in V} w_i x_i\\
&= \sum_{i\in V} \frac{w_i(c^\top v_i)^2}{(c^\top v_i)^2} x_i\\
&\le (\max_{i\in V} \frac{w_i}{(c^\top v_i)^2}) \sum_{i\in V}(c^\top v_i)^2 x_i\\
&\le \max_{i\in V} \frac{w_i}{(c^\top v_i)^2}.
\end{aligned}$$

Somit ist $\theta(G,w) \le \theta_1(G,w)$.

$\theta_1(G,w) \le \theta_2(G,w)$: Sei X eine Matrix aus dem zulässigen Bereich für θ_2 und $t = \lambda_1(X+W)$. Dann ist $tI - X - W$ positiv semidefinit und singulär. Sei $Z = (tI - X - W)^{1/2}$. Dann ist auch Z singulär und

$$\begin{aligned} z_i^\top z_i &= t - w_i && \text{für alle } x \in V \\ z_i^\top z_j &= -\sqrt{w_i w_j} && \text{für alle } (i,j) \notin E. \end{aligned}$$

Sei $c \in \ker(Z) \cap S^{n-1}$ und

$$v_i := \sqrt{\frac{w_i}{t}}c + \frac{1}{\sqrt{t}} z_i.$$

Dann erhalten wir:

$$\begin{aligned} v_i^\top v_i &= \frac{w_i}{t} + \frac{t - w_i}{t} = 1 && \text{für alle } i \in V \\ v_i^\top v_j &= \frac{\sqrt{w_i w_j}}{t} - \frac{\sqrt{w_i w_j}}{t} = 0 && \text{für alle } (i,j) \notin E. \end{aligned}$$

Also ist $\{v_1, \ldots v_n\} \in OND(G,n)$ und

$$\theta_1(G,w) \le \max_{i \in V} \frac{w_i}{(c^\top v_i)^2} = \max_{i \in V} \frac{w_i}{w_i/t} = t = \lambda_1(X+W). \tag{7.1}$$

Somit ist $\theta_1(G,w) \le \theta_2(G,w)$.

$\theta_2(G,w) \le \theta_3(G,w)$: Das SDP für $\theta_2(G,w)$ lautet:

$$\begin{array}{ll} \min & z \\ \text{unter} & \mathcal{A}\mathbf{vec}(X) = 0 \\ & zI - X \succeq W \end{array}$$

Das Duale dazu lautet:

$$\begin{array}{ll} \max & W \bullet Y \\ \text{unter} & \mathbf{Spur}(Y) = 1 \\ & Y = \mathbf{Mat}(\mathcal{A}^\top u) \\ & Y \succeq 0. \end{array}$$

$\theta_3(G,w)$ ist somit der Wert des dualen Programms. Nach Aufgabe 50 e) haben wir also sogar $\theta_2(G,w) = \theta_3(G,w)$.

$\theta_3(G,w) \leq \theta_4(G,w)$: Sei Y eine Optimallösung für $\theta_3(G,w)$. Da $Y \succeq 0$, existiert $U = Y^{1/2}$. Sei u_i die i-te Spalte von U. Als Optimallösung einer linearen Zielfunktion über einer konvexen Menge liegt Y auf deren Rand, also sind Y und U singulär. Sei v ein Einheitsvektor aus $ker(U)$. Wir setzen dann für $i \in V$:

$$v_i = \begin{cases} \frac{u_i}{\|u_i\|} & \text{falls } u_i \neq 0 \\ v & \text{sonst} \end{cases}$$

Dann ist offensichtlich

$$v_i^\top v_j = \begin{cases} 0 & \text{falls } (i,j) \in E \\ 1 & \text{falls } i = j \end{cases}$$

das heißt aber, daß $(v_1, \ldots, v_n) \in OND(\bar{G}, n)$. Da nun $(U\sqrt{w})^\top(U\sqrt{w}) = \sqrt{w}^\top Y \sqrt{w} = \theta_3(G,w)$ ist, folgt für $d = \frac{U\sqrt{w}}{\sqrt{\theta_3}}$: $\|d\| = 1$. Ferner gilt für

$$u_i \neq 0 : \; d^\top v_i = \frac{\sqrt{w}^\top U^\top U \sqrt{e_i}}{\sqrt{\theta_3}\|u_i\|},$$

und somit haben wir

$$\forall i \in V : \; \|u_i\| d^\top v_i = \frac{\sqrt{w}^\top Y \sqrt{e_i}}{\sqrt{\theta_3}}.$$

Somit ist

$$\sum_{i \in V} \|u_i\| \sqrt{w_i} d^\top v_i = \frac{1}{sqrt\theta_3} \sum_{i \in V} \sqrt{w}^\top Y e_i \sqrt{w_i} = \frac{1}{sqrt\theta_3} \sqrt{w}^\top Y \sqrt{w} = \sqrt{\theta_3}.$$

Nach Cauchy-Schwarz gilt nun:

$$\begin{aligned}
\theta_3 &= (\sum_{i \in V} \|u_i\| \sqrt{w_i} d^\top v_i)^2 \\
&\leq (\sum_{i \in V} \|u_i\|^2) \left(\sum_{i \in V} w_i (d^\top v_i)^2 \right) \\
&= \mathbf{Spur}(Y) (\sum_{i \in V} w_i (d^\top v_i)^2) \\
&= (\sum_{i \in V} w_i (d^\top v_i)^2) \\
&\leq \theta_4(G,w).
\end{aligned}$$

$\theta_4(G,w) \le \theta(G,w)$: Sei $(v_1,\dots,v_n) \in OND(\bar{G},k)$ und $d \in S^{k-1}$ mit $\theta_4(G,w) = \sum_{i\in V}(d^\top v_i)^2 w_i$. Wir haben zu zeigen, $\left((d^\top v_1)^2,\dots,(v_n^\top d)^2\right) \in \Theta(G)$. Sei also $(u_1,\dots,u_n) \in ONB(G,l)$ und $c \in S^l$.

Wir haben zu zeigen: $\sum_{i\in V}(c^\top u_i)^2(d^\top v_i)^2 \le 1$. Dafür betrachten wir zunächst die inneren Produkte von Matrizen der Gestalt $u_i v_i^\top$:

$$(u_i v_i^\top) \bullet (u_j v_j)^\top = (u_i^\top u_j)(v_i^\top v_j) = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{sonst.} \end{cases}$$

Also bilden diese Matrizen ein orthonormales System im Vektorraum der $(l \times k)$-Matrizen. Da auch die Matrix $cd^\top$ die Länge 1 bzgl. dieses inneren Produktes hat, denn

$$(cd^\top) \bullet (cd^\top) = (c^\top c)(d^\top d) = 1,$$

folgt somit aus der Entwicklung von $cd^\top$ nach dem orthonormalen System

$$1 = (cd^\top) \bullet (cd^\top) \ge \sum_{i\in V}\left((cd^\top) \bullet (u_i v_i^\top)\right)^2 = \sum_{i\in V}(c^\top u_i)^2 \bullet (d^\top v_i)^2.$$

□

7.2.5 Einige Korollare

Da $\theta_1(G,w) = \theta_2(G,w)$ erhalten wir daß im Beweis der ersten Ungleichung in 7.1 Gleichheit gelten muß. Die Ausdrücke, über die das Maximum gebildet wird sind aber alle gleich und wir erhalten:

Korollar 7.2.1 *Sei $G = (V,E)$ ein Graph mit Knotengewicht $w : V \to \mathbb{R}_+$ Dann gibt es eine $(u_1,\dots,u_n) \in OND(G,k)$ und $c \in S^{k-1}$ mit $\theta(G,w)(c^\top u_i)^2 = w_i$ für alle $i \in V$.*

□

Aus $\theta_4(G, w) = \theta(G, w)$ erhalten wir ferner

Korollar 7.2.2 *Sei* $(v_1, \ldots, v_n) \in OND(\bar{G}, k)$ *und* $d \in S^{k-1}$. *Sei ferner eine Knotengewichtung* $w : V \to \mathbb{R}_+$ *gegeben. Gelte außerdem*

$$\sum_{i \in V} (d^\top v_i)^2 w_i = \theta(G, w).$$

Dann ist

$$\sum_{i \in V} w_i (d^\top v_i) v_i = \theta(G, w) d.$$

Beweis: Aufgrund der Definition von $\theta_4(G, w)$ und des letzten Satzes haben wir für alle $y \in S^{k-1}$:

$$\sum_{i \in V} w_i (y^\top v_i)^2 \leq \theta(G, w)$$

und nach Annahme wird die linke Seite durch $y = d$ maximiert. Andererseits kann man die linke Seite auch umschreiben zu

$$\sum_{i \in V} w_i (y^\top v_i)^2 = y^\top \left(\sum_{i \in V} w_i v_i v_i^\top \right) y.$$

Also wird das Maximum angenommen, wenn y ein Eigenvektor von $\sum_{i \in V} w_i v_i v_i^\top$ ist und der Optimalwert ist der zugehörige Eigenwert. Also ist d Eigenvektor zum Eigenwert $\theta(G, w)$ und wir erhalten:

$$\sum_{i \in V} w_i (d^\top v_i) v_i = \sum_{i \in V} w_i v_i v_i^\top d = \theta(G, w) d.$$

$\square$

Ferner erhalten wir folgende weiteren Beschreibungen für $\Theta(G)$.

Korollar 7.2.3

1. $\Theta(G) = \{x \in \mathbb{R}^n \mid \theta(\bar{G}, x) \leq 1\}$
2. $\Theta(G) = \{x \in \mathbb{R}^n \mid x^\top y \leq \theta(G, y),\ \forall y \in R^n_+\}$
3. $\Theta(G) = \{(d^\top v_i)^2 \in \mathbb{R}^n \mid (v_1, \ldots, v_n) \in OND(\bar{G}, k)$ *und* $d \in S^{k-1}\}$.

Beweis:

1. Da $\theta(\bar{G},x) = \max\{\sum_{i\in V}(c^\top u_i)^2 x_i \mid (u_1,\dots,u_n) \in OND(G,k),\ c \in S^{k-1}\}$ und $\theta(G,w) = \theta_4(G,w)$ ist, gilt

$$\begin{aligned}
& \{x \in \mathbf{R}^n_+ \mid \theta(\bar{G},x) \le 1\} \\
= \;& \left\{x \in \mathbf{R}^n_+ \mid \max\{\sum_{i\in V}(c^\top u_i)^2 x_i \mid (u_i) \in OND(G,k), c \in S^{k-1}\} \le 1\right\} \\
= \;& \{x \in \mathbf{R}^n_+ \mid \sum_{i\in V}(c^\top u_i)^2 x_i \le 1,\ \forall (u_1,\dots,u_n) \in OND(G,k), c \in S^{k-1}\} \\
= \;& \Theta(G).
\end{aligned}$$

2. Sei

$$\Theta'(G) := \{x \in \mathbf{R}^n_+ \mid x^\top y \le \theta(G,y),\ \forall y \in \mathbf{R}^n_+\}.$$

Wir haben zu zeigen $\Theta(G) = \Theta'(G)$. Die Inklusion $\Theta(G) \subseteq \Theta'(G)$ ist per definitionem klar. Sei also $x \in \Theta'(G)$, $(u_1,\dots,u_n) \in OND(G,k)$ sowie $c \in S^{k-1}$. Mit der Wahl $y = \big((c^\top u_1)^2,\dots,(c^\top u_1)^2\big)$ folgt:

$$\begin{aligned}
\theta(G,y) &= \min\left\{\max_{i\in V}\{\frac{y_i}{(d^\top v_i)^2} \mid (v_i) \in OND(G,k), d \in S^{k-1}\right\} \\
&= \min\left\{\max_{i\in V}\{\frac{(c^\top u_i)^2}{(d^\top v_i)^2} \mid (v_i) \in OND(G,k), d \in S^{k-1}\right\} \\
&\le \max_{i\in V}\frac{(c^\top u_i)^2}{(c^\top u_i)^2} = 1
\end{aligned}$$

und somit auch auch $x^\top y \le 1$ und somit $x \in \Theta(G)$.

3. Sei

$$\Theta''(G) = \{(d^\top v_i)_{i\in V} \mid (v_1,\dots,v_n) \in OND(G,k),\ d \in S^{k-1}\}.$$

Wegen $\theta_4(G,w) = \theta(G,w)$ und dem soeben gezeigten gilt $\Theta''(G) \subseteq \Theta(G)$. Sei also $x \in \Theta(G)$ und $t = \theta(\bar{G},x)$. Aus 1. wissen wir $t \le 1$. Nach Korollar 7.2.1 gibt es also eine orthonormale Darstellung $(v_1,\dots,v_n) \in OND(G,k)$ und $d \in S^{k-1}$ mit $t(d^\top v_i)^2 = x_i$ für alle $i \in V$. Sei nun d_1 ein Vektor mit Länge 1, der orthogonal zu d und allen v_i ist. Dieser existiert nach eventueller Einbettung der Situation in $\mathbf{R}^{k+1}$. Sei $d_0 = \sqrt{t}d + \sqrt{1-t}d_1$. Dann ist $\|d_0\| = 1$ und $(d_0^\top v_i)^2 = t(d^\top v_i)^2 = x_i$ für alle $i \in V$, also $x \in \Theta''(G)$.

□

7.2.6 Die Facetten von Θ

Wir hatten im dritten Kapitel gelernt, daß die Dualität der semidefiniten Programmierung aus der Selbstpolarität des positiv semidefiniten Kegels folgt. Allgemein kann man den polaren Kegel als „duales Objekt“ zum Ausgangskegel betrachten. Für gewisse Polyeder und konvexe Mengen wollen nun noch ein weiteres Konzept einer „dualen Menge“ kennenlernen.

Definition 7.2.5 (convex corner, Antiblocker) *Eine Menge* $C \subseteq \mathbb{R}^n$ *ist eine* konvexe Ecke *oder* convex corner *falls*

$$x \in C, x' \geq 0 \textit{ und } x' \leq x \Rightarrow x' \in C.$$

Sei $S \subseteq \mathbb{R}^n_+$. *Der* Antiblocker *von* S *ist*

$$abl(S) := \{y \in \mathbb{R}^n_+ \mid y^\top x \leq 1, \forall x \in S\}.$$

Wir beobachten zunächst, daß

Proposition 7.2.2 *Sind* A *und* B *konvexe Ecken dann gilt* $A \subseteq B \Rightarrow abl(B) \subseteq abl(A)$.

Beweis: Klar nach Definition. □

Wie die Polarität bei abgeschlossenen Kegeln liefert auch der Begriff des Antiblockers eine Inversion.

Satz 7.2.4 *Ist* A *eine abgeschlossene convex corner so gilt* $abl(abl(A)) = A$.

Beweis: Offensichtlich ist $A \subseteq abl(abl(A)$. Sei nun $x \notin A$. Nach Aufgabe 4 gibt es dann ein c mit $c^\top x > 1$ und $c^\top x \leq 1$ für alle $x \in A$. Also ist $c \in abl(A)$ und somit $x \notin abl(abl(A))$.

Auch algorithmisch erhält man aus dieser Dualität eine Äquivalenz.

Satz 7.2.5 *Sei* C *eine abgeschlossene convex corner. Falls das Optimierungsproblem* $\max_{x \in C} c^\top x$ *in polynomieller Zeit gelöst werden kann, so kann auch das Optimierungsproblem über* $abl(C)$ *in polynomieller Zeit gelöst werden.*

Beweis: Den Beweis wollen wir hier nicht führen, sondern auf Grötschel, Lovász, Schrijver: „Geometric Algorithms and Combinatorial Optimization“ verweisen. Dahinter steckt das Prinzip der polynomiellen Äquivalenz von von Optimierung und Separation und damit die Ellipsoidmethode. □

Ist S eine stabile Menge in G, so ist S eine Clique in $\bar{G}$, deswegen definiert man

Definition 7.2.6 *Die konvexe Hülle der Inzidenzvektoren von Cliquen nennt man das* Clique-Polytop.

$$CLIQUE(G) := STAB(\bar{G}).$$

Analog setzt man

$$QCLIQUE(G) := QSTAB(\bar{G}).$$

Mit diesen Definitionen beobachten wir

Proposition 7.2.3

1. $abl(QSTAB(G)) = CLIQUE(G)$,
2. $abl(QCLIQUE(G)) = STAB(G)$,
3. *das Optimierungsproblem über $QSTAB(G)$ ist NP-schwer,*
4. *das Optimierungsproblem über $QCLIQUE(G)$ ist NP-schwer.*

Beweis: Die ersten beiden Behauptungen sind klar. Die änderen beiden folgen, da das Cliquenproblem und das stabile Mengen Problem NP-vollständig sind. □

Satz 7.2.6

$$abl(\Theta(G)) = \Theta(\bar{G})$$

Beweis: Klar nach Teil 1.) von Korollar 7.2.3. □

Im allgemeinen ist $\Theta(G)$ kein Polyeder. Wir werden zeigen, daß $\Theta(G)$ genau dann ein Polyeder ist, wenn es mit $STAB(G)$ und $QSTAB(G)$ zusammenfällt.

Definition 7.2.7 *Sei $S \subset \mathbb{R}^n$ eine volldimensionale, konvexe Menge und gelte für alle $x \in S : c^\top x \leq b$. Dann sagen wir $c^\top x = b$ definiert eine* Facette*F von S, wenn*

$$F = \{x \in S \mid c^\top x = b\} \cap S \text{ und } dim(F) = n - 1.$$

Satz 7.2.7 *Definiert eine Ungleichung eine Facette von $\Theta(G)$, so ist diese entweder gleichbedeutend mit einer Nichtnegativitätsbedingung oder einer Clique-Ungleichung.*

Beweis: Sei $a^\top x \leq \alpha$ eine facettendefinierende Ungleichung von $\Theta(G)$ und z ein innerer Punkt von $F = \{x \in S \mid c^\top x = b\} \cap \Theta(G)$. Wir haben zu zeigen, daß aus $z_i \neq 0$ für alle Koordinaten folgt, daß die Ungleichung eine Clique-Ungleichung ist. Zunächst gilt, da z ein Randpunkt ist nach Korollar 7.2.3, daß $\theta(\bar{G}, z) = 1$ ist. Da θ hier durch Optimierung einer linearen Zielfunktion entsteht, und dieses Optimum auf einem inneren Punkt einer Facette angenommen wird, muß schon die ganze Facette optimal sein. Da $\theta = \theta_4$ ist, gibt es ferner ein $(u_1, \ldots, u_n) \in OND(G, k)$ und $c \in S^{k-1}$ mit

$$\sum_{i \in V} z_i (c^\top u_i)^2 = 1.$$

Da auf der anderen Seite $\sum_{i \in V} x_i (c^\top u_i)^2 \leq 1$ eine gültige Ungleichung für $\theta(G)$ definiert, muß diese schon gleichbedeutend mit $a^\top x \leq \alpha$ sein.

Also gilt für alle $x \in F : \sum_{i \in V} x_i (c^\top u_i)^2 = 1 = \theta(\bar{G}, x)$. Die letzte Ungleichung folgt wieder aus $\theta = \theta_4$ und Satz 7.2.6. Nach Korollar 7.2.2 gilt also

$$\forall x \in F : \sum_{i \in V} x_i (c^\top u_i) u_i = c.$$

Da $\sum_{i \in V} x_i (c^\top u_i)^2 = 1$ ist, folgt

$$c_j = \sum_{i \in V} c_j x_i (c^\top u_i)^2 = \sum_{i \in V} x_i (c^\top u_i)(u_i)_j.$$

Da $\dim(F) = n - 1$ und $\alpha = 1$ in der facettendefinierenden Ungleichung ist, ist der Kern der Matrix

$$\begin{pmatrix} c_1(c^\top u_1)^2 - (c^\top u_1)(u_1)_1 & \ldots & c_1(c^\top u_n)^2 - (c^\top u_n)(u_n)_1 \\ \vdots & \ddots & \vdots \\ c_n(c^\top u_1)^2 - (c^\top u_1)(u_1)_n & \ldots & c_n(c^\top u_n)^2 - (c^\top u_n)(u_n)_n \end{pmatrix}$$

n-dimensional. Also haben wir stets $(c^\top u_i)c = u_i$ oder $(c^\top u_i) = 0$. In dem ersten Fall gilt wegen $\|u_i\| = \|c\| = 1$ also $u_i = \pm c$. Wir können annehmen daß wir oben diese u_i so gewählt haben, daß $u_i = c$ gilt. Sei nun $K = \{i \in V \mid u_i = c\}$. Dann gilt für je zwei Knoten $i, j \in K : u_i^\top u_j = \|c\|^2 = 1 \neq 0$. Also müssen i und j adjazent sein, K ist eine Clique und die Ungleichung $\sum_{i \in V} x_i (c^\top u_i)^2 \leq 1$ ist eine Clique-Ungleichung. □

Wir erhalten daraus

Korollar 7.2.4 (Perfect Graph Theorem) *G ist perfekt genau dann, wenn $\bar{G}$ perfekt ist.*

Beweis: Nach Definition ist G perfekt, wenn $STAB(G) = QSTAB(G)$, nach der Antiblockerbeziehung ist dies aber der Fall genau dann, wenn $CLIQUE(G) = QCLIQUE(G)$. Wegen $QCLIQUE(G) = QSTAB(\bar{G})$ und $CLIQUE(G) = STAB(\bar{G})$ folgt die Behauptung. □

Korollar 7.2.5 $\Theta(G)$ *ist ein Polyeder genau dann, wenn* G *perfekt ist.*

Beweis: Wenn G perfekt ist, gilt $\Theta(G) = STAB(G)$. Also ist $\Theta(G)$ ein Polyeder. Ist umgekehrt $\Theta(G)$ ein Polyeder, so ist nach Satz 7.2.7 $QSTAB(G) \subseteq \Theta(G)$. Wegen $\Theta(G) \subseteq QSTAB(G)$ erhalten wir demnach: $\Theta(\bar{G}) = abl(\Theta(G)) = abl(QSTAB(G)) = CLIQUE(G) = STAB(\bar{G})$. Also ist $\Theta(\bar{G})$ ein Polyeder, also $\bar{G}$ perfekt und die Behauptung folgt aus dem Perfect Graph Theorem. □

Definition 7.2.8 *Sei* $G = (V, E)$ *ein Graph. Die* Stabilitätszahl $\alpha(G)$ *ist die maximale Kardinalität einer stabilen Menge in* G *also* $\alpha(G) := \omega(\bar{G}) = \alpha(G, \mathbb{1})$. *Die* Clique-Coverzahl $\rho(G)$ *ist die minimale Anzahl von Cliquen, die die Knoten von* G *überdecken. Man sieht leicht* $\rho(G) = \chi(\bar{G})$.

G *heißt* ω-perfekt, *wenn* $\omega(G_W) = \chi(G_W)$ *für alle knoteninduzierten Teilgraphen von* G. G *heißt* α-perfekt, *wenn* $\alpha(G_W) = \rho(G_W)$ *für alle knoteninduzierten Teilgraphen von* G.

Alle diese Definitionen sind äquivalent, denn

Satz 7.2.8 *Sei* $G = (V, E)$ *ein Graph. Dann sind paarweise äquivalent:*

1. G *ist perfekt,*
2. G *ist* ω*-perfekt,*
3. G *ist* α*-perfekt.*

Beweis: Die Äquivalenz der letzten beiden Aussagen folgt aus dem Perfect Graph Theorem, wenn wir $1.) \Leftrightarrow 2.)$ zeigen.

Sei nun G perfekt und $W \subseteq V$. Mit $CLIQUE(G) = QCLIQUE(G)$ haben wir auch $CLIQUE(G_W) = QCLIQUE(G_W)$, denn beide Polyeder erhalten wir durch Schitt obiger Polyeder mit $\{x \in \mathbf{R}^{|V|} \mid x_v = 0 \text{ falls } x \notin W\}$ und ignorieren der Komponenten außerhalb Ws.

Zum Beweis von $1.) \Rightarrow 2.)$ führen wir Induktion über $|V|$, die Verankerung ist trivial. Wegen obiger Bemerkung genügt es zu zeigen, daß $\omega(G) = \chi(G)$. Dazu betrachten wir

$$\omega(G) = \max_{x \in CLIQUE(G)} \mathbb{1}^\top x = \max_{x \in QCLIQUE(G)} \mathbb{1}^\top x.$$

Aufgrund der Nichtnegativität der Zielfunktion und weil $QSTAB$ eine konvexe Ecke ist, können wir aus dem Programm die Nichtnegativitätsbedingungen weglassen, ohne den Optimalwert zu verändern.

Wir erhalten also

$$\begin{array}{lll} \omega(G) = & \max & \mathbb{1}^\top x \\ & \text{unter} & \mathbb{1}_S x \leq 1 \quad \text{für alle stabilen Mengen } S. \end{array}$$

Die Menge der Optimallösungen dieses Programms ist die konvexe Hülle der Cliquen maximaler Kardinalität. Andererseits muß die Menge der Optimallösungen ein Teil des Randes der zulässigen Menge sein. Also gibt es eine stabile Menge S mit $\mathbb{1}_S^\top \mathbb{1}_Q = 1$ für alle Cliquen Q maximaler Kardinalität. Somit ist $\omega(G_{V \setminus S}) = \omega(G) - 1$. Nach Induktionsvoraussetzung ist $\chi(G_{V \setminus S}) = \omega(G_{V \setminus S}) = \omega(G) - 1$. Die Behauptung folgt nun, wenn wir die zugehörige Färbung von $V \setminus S$ erweitern, indem wir alle Elementen in S mit derselben, neuen Farbe färben.

Sei nun G ω-perfekt. Wegen des Perfect Graph Theorems genügt zu zeigen: Ist v ein Extremalpunkt von $QCLIQUE(G)$, so ist $v \in CLIQUE(G)$. Wieder führen wir Induktion über $|V|$. Hat v in einer Koordinate einen Nulleintrag, folgt die Behauptung sofort nach Induktionsvoraussetzung. Ansonsten beobachten wir zunächst

Proposition 7.2.4 *Sei $S \subseteq \mathbb{R}^n$ eine konvexe Ecke und v eine Extremalpunkt von S, so ist v die eindeutige Optimallösung des Programms*

$$\begin{array}{ll} \max & v^\top x \\ \textit{unter} & x \in S. \end{array}$$

Beweis: Angenommen $w \in S$ mit $v^\top w \geq v^\top v$. Da S eine konvexe Ecke ist, können wir annehmen, daß $v_i = 0 \Rightarrow w_i = 0$. Da v extremal ist, liegt für kein $\epsilon > 0 : v - \epsilon w \in S$. Das heißt aber, da S konvexe Ecke ist, daß eine Koordinate $w_j > 0$ existiert mit $v_j = 0$, im Widerspruch zur Wahl von w. □

Sei v nun wieder der Extremalpunkt von $QCLIQUE$. Als Extremalpunkt eines Polyeders und somit als Lösung eines Gleichungssystem mit rationalen (hier sogar ganzzahligen) Daten ist v rational. Sei q der Hauptnenner der v-Einträge und $\tilde{v} = qv$. Da skalare Multiplikation der Zielfunktion an der Menge der Optimallösungen nichts ändert, genügt es zu zeigen

$$\max_{x \in CLIQUE(G)} \tilde{v}^\top x = \max_{x \in QCLIQUE(G)} \tilde{v}^\top x.$$

Wir ersetzen nun jeden Knoten $i \in V$ durch eine Clique mit $\tilde{v}_i$ Knoten und erhalten den Graphen $\tilde{G}$. Dann ist

$$\max_{x \in CLIQUE(G)} \tilde{v}^\top x = \max_{x \in CLIQUE(\tilde{G})} \mathbb{1}^\top x$$

und

$$\max_{x \in QCLIQUE(G)} \tilde{v}^\top x = \max_{x \in QCLIQUE(\tilde{G})} \mathbb{1}^\top x.$$

Also genügt es zu zeigen, daß $\tilde{G}$ perfekt ist. Dies folgt induktiv aus dem folgenden Lemma.

Lemma 7.2.1 *Sei $G = (V, E)$ ein perfekter Graph und $v \in V$. Der Graph $\tilde{G}$ entstehe aus G durch verdoppeln von v, genauer sei $\tilde{v}$ eine neue Ecke und $\tilde{G} = (V \cup \{\tilde{v}\}, \tilde{E})$ mit $\tilde{E} = E \cup \{(\tilde{v}, w) \mid (v, w) \in E\} \cup \{(v, \tilde{v})\}$. Dann ist $\tilde{G}$ perfekt.*

Beweis: Wieder führen wir Induktion über V. Da G perfekt ist, müssen wir nur die Teilgraphen von $\tilde{G}$ betrachten, in denen v und $\tilde{v}$ vorkommen. Sind diese echt, so folgt die Behauptung aus der Induktionsvoraussetzung. Wir haben also nur zu zeigen $\omega(\tilde{G}) = \chi(\tilde{G})$. Liegt v in G in einer Clique maximaler Kardinalität, so ist $\omega(\tilde{G}) = \omega(G) + 1 = \chi(G) + 1 \geq \chi(\tilde{G})$. Liegt hingegen v in keiner maximalen Clique von G und ist $c : V \to \{1, \ldots, \chi(G)\}$ eine optimale Färbung von G, so sei $W := c^{-1}(c(v))$. Da G perfekt ist, muß S jede Clique maximaler Kardinalität treffen. Also ist $\omega(G_{(V \setminus W) \cup \{v\}}) = \omega(G) - 1$. Diesen Graphen können wir also mit $\chi(G) - 1$ Farben färben und haben noch eine übrig für $S \cup \{\tilde{v}\}$. $\square$

Ende des Beweises von Satz 7.2.8. $\square$

Anhang A

Übungsaufgaben

Aufgabe 4: *(Trennsatz)*
Sei $S \subseteq \mathbf{R}^n$ abgeschlossen und konvex und $x \in \mathbf{R}^n$. Zeigen Sie: $x \in S$ oder es gibt eine affine Hyperebene, die x und S trennt, d.h. $\exists c \in \mathbf{R}^n \exists b \in \mathbf{R} : c^\top x > b$ und $\forall y \in S : c^\top y < b$.
(Unverbindlicher Hinweis: Die euklidische Norm ist stetig.)

5 Punkte

Aufgabe 33: *(Positiv Semidefinite Matrizen)*
Sei P eine positive semidefinite Matrix und $y_1, \ldots, y_n$ eine Orthonormalbasis aus Eigenvektoren zum Eigenwert λ_i.

a) Zeigen Sie: $P = \sum_{i=1}^{n} \lambda_i y_i y_i^\top$.

b) Sei $S_1 = \{yy^\top \mid y \in \mathbf{R}^n\}$ und

$$S_2 = \{P \in \mathbf{R}^{n\times n} \mid P \text{ ist reell, symmetrisch, positiv semidefinit } \}.$$

Zeigen Sie: $\operatorname{conv}(S_1) = S_2$.

c) Zeigen Sie: S_2 ist ein abgeschlossener, spitzer (d.h. er enthält keinen Untervektorraum des $\mathbf{R}^n$), konvexer Kegel.

d) Ist K ein spitzer, konvexer Kegel und $d \in K$, so heißt d *extremal*, wenn $\forall e \in \mathbf{R}^n \setminus \{\lambda d \mid \lambda \geq 0\}$ gilt: $\forall \varepsilon > 0 : (d + \varepsilon e \notin K$ oder $d - \varepsilon e \notin K)$. Zeigen Sie: S_1 sind genau die Extremalen von S_2.

5 Punkte

Aufgabe 34: *(Polare Kegel)*
Sei $K \subseteq \mathbf{R}^n$ ein abgeschlossener, konvexer Kegel. Dann ist der polare Kegel K^* definiert durch: $K^* = \{y \in \mathbf{R}^n \mid \forall x \in \mathbb{K} : x^\top y \geq 0\}$.

a) K^* ist ein konvexer, abgeschlossener Kegel.

b) Ist $K = \{x \in \mathbf{R}^n \mid x \geq 0\}$, so gilt $K^* = K$.

c) Ist S_2 wie oben, so gilt (bei geeigneter Interpretation) $S_2^* = S_2$.

4 Punkte

Aufgabe 36: *(LP-Standardform(en))*
Bringen Sie die folgenden Linearen Programme auf die Standardformen

$$\begin{array}{c} \max c^\top x \\ Ax \leq b \end{array} \quad \text{bzw.} \quad \max_{\substack{Ax=b \\ x\geq 0}} c^\top x,$$

d. h. formulieren Sie jeweils ein lineares Programm dieser Gestalt dessen Optimallösungen denen des vorgegebenen Programms entsprechen.

$$a) \max_{\substack{Ax\leq b \\ x\geq 0}} c^\top x \qquad b) \begin{array}{c} \max c^\top x \\ Ax = b \end{array} \qquad c) \begin{array}{c} \min b^\top u \\ Au = c \end{array} \qquad d) \min_{\substack{A^\top u=c \\ u\geq 0}} b^\top u$$

3 Punkte

Aufgabe 37: *(Noch mehr LP)*
Betrachten Sie das LP in der Standardform

$$\max_{\substack{Ax=b \\ x\geq 0}} c^\top x .$$

Sei B eine Spaltenindexmenge von A mit $\text{rg}(A_{.B}) = \text{rg}(A)$. Sei N die Menge der übrigen Spalten und o.E. $A = (A_{.B}, A_{.N})$. Zeigen Sie: ein Vektor $x \in \mathbf{R}^n$ ist zulässig für das Programm, genau dann wenn $x_N \geq 0$ ist und $x_B = A_{.B}^{-1} b - A_{.B}^{-1} A_{.N} x_N \geq 0$.

2 Punkte

Aufgabe 38: *(Farkas' Lemma)*
Zeigen Sie Farkas' Lemma in der homogenen Form: Sei $L \subseteq \mathbf{R}^n$ ein linearer Unterraum von $\mathbf{R}^n$, dann gilt: ($\exists x \in L, x \geq 0, x_1 > 0 \Leftrightarrow \nexists x \in L^\perp, x \geq 0, x_1 > 0$).
Tip: Induktion über n. Man beachte dabei, daß $\{x \in \mathbf{R}^{n-1} \mid \exists (x, 0) \in L\}^\perp = \{x \in \mathbf{R}^{n-1} \mid \exists (x, x_n) \in L^\perp, x_n \in \mathbf{R}\}$

6 Punkte

Aufgabe 39: *(Lineare Programmierung)*
Betrachten Sie folgendes Paar von Linearen Optimierungsproblemen für sinnvolle Dimensionen von A, b, c:

$$\begin{array}{lll} \text{(A)} & \max c^\top x \\ \text{unter} & Ax \leq b \\ & x \geq 0 \end{array} \qquad \begin{array}{ll} \text{(B)} & \min u^\top b \\ & u^\top A \geq c \;\; . \\ & u \geq 0. \end{array}$$

Zeigen Sie:

a) Ist x zulässig für (A) und u zulässig für (B), so gilt $c^\top x \leq u^\top b$.

b) Hat Problem (A) eine zulässige Lösung, Problem (B) aber nicht, so ist Problem (A) unbeschränkt.

c) Hat Problem (A) eine endliche Optimallösung x_0, so hat (B) eine Optimallösung u_0 mit $c^\top x_0 = u_0^\top b$.

Tip: verwenden Sie in b) und c) die Aussage von Aufgabe 38. In b) betrachten Sie etwa $L = \text{ker}(-c, A^\top, -I)$ und in c) z.B.

$$L = \text{ker} \begin{pmatrix} b & -A & -I & 0 & 0 \\ -c & 0 & 0 & A^\top & -I \\ 0 & -c^\top & 0 & b^\top & 0 \end{pmatrix}$$

7 Punkte

Aufgabe 41: *(Zu $\theta \leq \chi$)*
Sei V eine endliche Menge und $c : V \to \{1, \ldots, k\}$ eine surjektive Abbildung. Sei $M = m_{ij}$ die $(V \times V)$-Matrix definiert durch:

$$m_{vw} = \begin{cases} 1 & \text{falls } v = w \text{ oder } c(v) \neq c(w) \\ 1 - k & \text{sonst.} \end{cases}$$

Zeigen Sie: $\lambda_1(M) = k$.

7 Punkte

Aufgabe 42: *(Benutze Hölder'sche Ungleichung)*
Zeigen Sie: p-Kegel und q-Kegel sind polar genau dann, wenn $\frac{1}{p} + \frac{1}{q} = 1$.

4 Punkte

Aufgabe 43: *(3.4.1)*
Wie sieht die Menge $\mathcal{P}$ in Beispiel 3.4.1 aus. Ist sie abgeschlossen?

4 Punkte

Aufgabe 44: *(Farkas Lemma)*
Beweisen Sie folgende Varianten von Farkas' Lemma für den positiv semidefiniten Kegel. In beiden Fällen fehlt eine constraint qualificiation, finden Sie sie!

a) Entweder es gibt $x_1, \ldots, x_n$ mit $\sum_{i=1}^n x_i A_i \preceq B$ oder es gibt ein $Y \succeq 0$ mit $A_i \bullet Y = 0$ und $B \bullet Y < 0$, aber nicht beides.

b) Ist $A \in \mathbb{R}^{n \times n}$ nicht notwendig symmetrisch, dann gilt: Entweder es gibt $X \succeq 0$ mit $AX + XA^\top = B$ oder es gibt eine symmetrische Matrix Y mit $A^\top Y + YA \succeq 0$ und $B \bullet Y < 0$, aber nicht beides.

7 Punkte

Aufgabe 45: *(Spuren im Sand)*
Sei A eine reelle $(n \times n)$-Matrix und L eine invertierbare, reelle $(n \times n)$-Matrix. Zeigen Sie:

$$\textbf{Spur}(L^{-1}AL) = \textbf{Spur}(A).$$

3 Punkte

Aufgabe 51: *(LPs Dualisieren)*
Bestimmen Sie zu den folgenden Linearen Programmen jeweils das duale Programm. Bringen Sie dazu das Programm zunächst auf Standardform, dualisieren Sie und vereinfachen dann wieder:

$$a)\ \begin{array}{c} \max c^\top x \\ Ax \geq b \end{array} \qquad b)\ \begin{array}{c} \min c^\top x \\ Ax \leq b \\ x \geq 0 \end{array} \qquad c)\ \begin{array}{c} \max c^\top x \\ Ax = b \end{array} \qquad d)\ \begin{array}{c} \max c^\top x + d^\top y \\ Ax + By = a \\ Cx + Dy \leq b \\ x \geq 0 \end{array}$$

$e)$ Zeigen Sie: Das duale Programm des dualen Programms ist das primale Programm.

4 Punkte

Aufgabe 52: *(Kronecker Produkt)*
Zeigen Sie: Sind A, B, C Matrizen verträglicher Dimensionen, so gilt

a) $(A \otimes B)^\top = (A^\top \otimes B^\top)$,

b) $\mathbf{vec}(ABC) = (C^\top \otimes A)\mathbf{vec}(B)$,

c) $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$,

d) Sind A und B invertierbar, so ist auch $A \otimes B$ invertierbar und es gilt $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

e) Sind A eine symmetrische $n \times n$ Matrix und B symmetrisch $n \times n$, dann sind die Eigenwerte von $(A \otimes B)$ gerade $\lambda_i(A)\lambda_j(B)$ für $i = 1, \ldots, m$ und $j = 1, \ldots, n$. Eigenvektor zum Eigenwert $\lambda_i(A)\lambda_j(B)$ ist der Vektor $q_i(A) \otimes q_j(B)$, wobei $q_i(A)$ und $q_j(B)$ Eigenvektoren von A bzw. B zu den Eigenwerten λ_i und λ_j sind.

6 Punkte

Aufgabe 53: *(Duality gap)*
Sei $C := C_1 \cup C_2$ die Menge im $\mathbf{R}^3$ gegeben durch:

$$C_1 := \{(x, y, z) \subseteq \mathbf{R}^3 \mid x \geq 0,\ z \geq 0, y \geq 1\},$$
$$C_2 := \{(x, y, z) \subseteq \mathbf{R}^3 \mid -1 \leq x < 0,\ z \geq \exp(y/x), y \geq 1\},$$

und $K = \{(tp, t) \subseteq \mathbf{R}^4 \mid t \geq 0,\ p \in C\}$ der Kegel über C. Zeigen Sie:

a) K ist ein abgeschlossener, spitzer, konvexer, volldimensionaler Kegel. (Tip: untersuche C)

b) Berechnen Sie den Optimalwert von
$$\begin{array}{ll} \min & x \\ \text{unter} & t = 1 \\ & z = 0 \\ & (x, y, z, t) \geq_K 0\ . \end{array}$$

c) Wie lautet das duale Programm und was ist sein Optimalwert?

d) $\mathcal{P} = \{(z, t) \mid \exists x, y : (x, y, z, t) \in K\}$ ist abgeschlossen.

e) Wie verträgt sich diese Übungsaufgabe mit dem Satz über starke Dualität?

10 Punkte

Aufgabe 54: *(Dualisieren)*
Bestimmen Sie jeweils das duale SDP:

$$a)\ \begin{array}{c} \max C \bullet X \\ \mathbf{Mat}(\mathcal{A}\mathbf{vec}(X)) \succeq B \end{array} \quad b)\ \begin{array}{c} \min C \bullet X \\ \mathcal{A}\mathbf{vec}(X) \leq b \\ X \succeq 0 \end{array} \quad c)\ \begin{array}{c} \max C \bullet X \\ \mathcal{A}\mathbf{vec}(X) = B \end{array} \quad d)\ \begin{array}{c} \max C \bullet X + d^\top y \\ \mathcal{A}\mathbf{vec}(X) + By = a \\ \mathcal{C}\mathbf{vec}(X) + Dy \leq b \\ X \succeq 0 \end{array}$$

5 Punkte

Aufgabe 55: *($\theta(G, w)$)*
Sei $G = (V, E)$ ein Graph und $w \in \mathbf{R}_+^{|V|}$ eine nicht-negative Knotengewichtung. Mit $\sqrt{w}$ bezeichnen wir den Vektor mit Komponenten $(\sqrt{w})_i = \sqrt{w_i}$. Sei $\mathcal{M}$ die Menge der symmetrischen Matrizen X mit $x_{ij} = 0$ falls $i = j$ oder $(i, j) \in E$.

a) Schreiben Sie die Optimierungsaufgabe

$$\theta(G, w) := \min\{\lambda_1(X + \sqrt{w}\sqrt{w}^\top) \mid X \in \mathcal{M}\}$$

als SDP.

b) Wie lautet das zugehörige duale SDP?

c) Zeigen Sie: $\sum_{i=1}^{|V|} w_i \geq \theta(G, w) \geq \max_{i=1}^{|V|} w_i$.

d) Geben Sie ein w an mit $\theta(G) = \theta(G, w)$.

e) Zeigen Sie: das duality gap ist 0.

8 Punkte

Aufgabe 56: *(Konvex)*
Zeigen Sie:

a) $(\lambda_1 + \ldots + \lambda_k)(A(x))$ ist eine konvexe Funktion.

b) $\max^{\Sigma|k|}(x)$ ist konvex.

4 Punkte

Aufgabe 57: *(Modellierung)*

a) Schreiben Sie $\min\{\max^{\Sigma|k|}(x) \mid x \in \mathbf{R}^m\}$ als kleines LP.

b) Schreiben Sie $\sum_{i=1}^{k} |\lambda^i(B)|$ als SDP.

c) Schreiben Sie $\min\{\sum_{i=1}^{k} |\lambda^i(X)| \mid \mathcal{A}\mathbf{vec}(X) = b,\ X \succeq 0\}$ als SDP.

5 Punkte

Aufgabe 59: *(To be ≥ 0 or not to be ≥ 0)*

a) Konstruieren Sie ein LP $(P) = \min\{c^\top x \mid Ax = b,\ x \geq 0\}$, das eine endliche Optimallösung hat, bei dem jedoch für kein $\mu > 0$ das Problem mit barrier function $(P_\mu) = \min\{c^\top x - \mu \sum \ln(x_i) \mid Ax = b,\ x \geq 0\}$ eine endliche Optimallösung hat.

b) Zeigen Sie: Ist $(P) = \min\{c^\top x \mid Ax = b,\ x \geq 0\}$ beschränkt, so sind paarweise äquivalent:

   (a) es gibt es ein $\mu > 0$, so daß $(P_\mu) = \min\{c^\top x - \mu \sum \ln(x_i) \mid Ax = b,\ x \geq 0\}$ beschränkt ist,

   (b) für alle $\mu > 0$ ist $(P_\mu) = \min\{c^\top x - \mu \sum \ln(x_i) \mid Ax = b,\ x \geq 0\}$ beschränkt,

   (c) die Menge der Optimallösungen von (P) ist beschränkt.

6 Punkte

Aufgabe 60: *(stetiger Weg)*

Die Abbildung $\begin{array}{c}(x_\mu, u_\mu, s_\mu) : (0, +\infty) \to \mathbb{R}^n \\ \mu \mapsto (x_\mu, u_\mu, s_\mu)\end{array}$ aus der Vorlesung ist stetig.

4 Punkte

Aufgabe 61: *(Ableiten)*

Sei X eine $n \times n$ Matrix und L eine reguläre $(n \times n)$-Matrix. Geben Sie jeweils die Jacobische folgender Abbildungen an:

a) $f(X) = L^{-1}XL$,

b) $g(X) = XX^\top$

c) $h(X) = XX$.

6 Punkte

Index