Semidefinite Programming and Combinatorial Optimization

Franz Rendl *
Universität Klagenfurt
Institut für Mathematik
Universitätsstr. 65-67, A-9020 Klagenfurt, Austria
franz.rendl@uni-klu.ac.at

December 10, 1998

Abstract

Semidefinite Programs have recently turned out to be a powerful tool for approximating integer problems. To survey the development in this area over the last few years, the following topics are addressed in some detail. First, we investigate ways to derive semidefinite programs from discrete optimization problems. The duality theory for semidefinite programs is the key to understand algorithms to solve them. The relevant duality results are therefore summarized. The second part of the paper deals with the approximation of integer problems both in a theoretical setting, and from a computational point of view.

1 Introduction

The interest in *Semidefinite Programming (SDP)* has been growing rapidly in the last few years. Here are some possible explanations for this sudden rise of interest.

The algorithmic development of interior-point methods for Linear Programs indicated the potential of this approach to solve general convex problems. Semidefinite Programs are a natural generalization of Linear Programs. The possibility to solve them efficiently using interior-point methods has attracted several research groups working previously in Linear Programming.

Independent of algorithmic aspects, Semidefinite Programs arise quite naturally as relaxations of discrete optimization problems. The most prominent, and also fundamental contribution here is certainly from Lovász [38]. More recently, SDP has been used to obtain approximation results for NP-hard graph optimization problems, such as Max-Cut or Graph Coloring.

A third source of interest in SDP arises from applications in Systems and Control Theory. The present survey concentrates on the connection of SDP to combinatorial optimization. The reader interested in applications of SDP in Control Theory should refer to the recent book [8].

Several survey-like papers have recently been written on the topic. Alizadeh [1] was among the first researchers to realize the potential of SDP in connection with integer programming. Later, Vandenberghe and Boyd [61] provided an excellent survey on SDP, featuring many applications of SDP and also algorithmic aspects. The surveys by Goemans [21] and Lovász [39] also consider the interplay between SDP and several NP-hard optimization problems.

*This research was carried out for the FWF Project P12660-MAT. Further support by START program Y43-MAT of the Austrian Ministry of Science is gratefully acknowledged.

While all these papers are dedicated to essentially the same topic, they are nonetheless all quite different from one another, due to the selection, emphasis and omission of the material presented.

The basic theme of the present survey is to exhibit practical experience with SDP relaxations applied to a variety of combinatorial optimization problems.

First we look in some detail how SDP arises as a relaxation of discrete problems. SDP is a special case of Linear Programming over a cone. The duality theory for this type of problem is significantly more subtle than Linear Programming duality. We recall the relevant duality results. Next we describe interior-point approaches to solve SDP. We also discuss the currently open issues in this area. We also look at various other nonlinear relaxations for quadratic problems in binary variables. Finally we review theoretical and computational work to approximate some graph optimization problems.

2 How Semidefinite Programs arise from integer problems

In this section we investigate how semidefinite programs arise from relaxations of integer programs. We first look at a generic way of deriving semidefinite relaxations from a problem in binary variables. Then we look at connections between eigenvalue optimization and semidefinite models. Finally, we consider some other methods that establish a link between semidefinite problems and graph optimization problems.

2.1 Unconstrained Quadratic Problems in binary variables

Let us start with the following problem in $(-1,1)$ variables. For given $Q = Q^t$ and c, consider

$$(P) \qquad \max\ x^tQx - 2c^tx \quad \text{such that } x \in \{-1,1\}^{n-1}. \tag{1}$$

We first **homogenize** the cost function by introducing an artificial binary variable x_n, and obtain

$$(P) \qquad z_{mc} = \max\ x^tLx \quad \text{such that } x \in \{-1,1\}^n. \tag{2}$$

Here, $L = \left(\begin{smallmatrix} Q & -c \\ -c^t & 0 \end{smallmatrix}\right)$. Note that $f(x) = x^tLx$ implies $f(x) = f(-x)$ so $x_n = 1$ can be assumed without loss of generality.

This optimization problem is NP-hard in general and corresponds to *Max-Cut*, if L represents the *Laplacian matrix* of the (weighted) graph in question. (A formal definition of the Laplacian will be given in Section 2.4.) A summary of recent results on the Max-Cut Problem in the context of optimization is given by Poljak and Tuza [52].

To obtain a semidefinite relaxation of (P), we **linearize** the cost function in (2), and observe that the convex hull of feasible points of the linearized model is contained in the cone of semidefinite matrices. To be specific, let $\mathcal{F}$ denote the set of all matrices xx^t, where $x \in \{-1,1\}^n$. Then

$$\max_{x_i \in \{-1,1\}} x^tLx = \max_{x_i \in \{-1,1\}} \operatorname{tr}\ [L \cdot (xx^t)] = \max_{X \in \mathcal{F}} \operatorname{tr}\ LX = \max_{X \in \operatorname{conv}(\mathcal{F})} \operatorname{tr}\ LX.$$

The last equality follows from the linearity of the cost function.

The main task now consists in approximating $\operatorname{conv}(\mathcal{F})$ tightly by supersets. In *polyhedral combinatorics*, one tries to do so by using linear (in)equalities. Now we require that X is positive semidefinite, $X \succeq 0$ for short, which is obviously true for all elements in conv($\mathcal{F}$). (A matrix X is positive semidefinite if $X = X^t$ and $a^t X a \geq 0$ for all vectors a.)

We note the following description of $\mathcal{F}$. (e denotes the vector of all ones. A subscript indicates its size. It is omitted if the size is clear from the context.)

Lemma 1 $X \in \mathcal{F} \Leftrightarrow X \succeq 0,\ diag(X) = e$ *and* $\mathrm{rank}(X) = 1$.

Proof. The necessity of the conditions is obvious. Sufficiency follows from the fact that a rank-one matrix $X \succeq 0$ is necessarily of the form $X = aa^t$. Since $1 = x_{ii} = a_i^2$, the claim follows. □

The basic *semidefinite relaxation* for (P) is obtained by dropping the rank one condition, and optimizing over the remaining sets.

$$(P_{sdp}) \quad z_{sdp} = \max\{\mathrm{tr}\ LX : \mathrm{diag}(X) = e, X \succeq 0\}. \tag{3}$$

This formulation and its dual form have been investigated quite intensively from a theoretical point of view, see e.g. [11, 22, 37, 51], as well as computationally, see [24, 29, 49, 50]. We come back to this problem in more detail in subsequent sections.

2.2 Problems in (0,1) variables

Problem (P) can equivalently be expressed in $(0,1)$ variables z, by substituting $x = 2z - e$. After homogenization and linearization, this results in the following problem.

$$\max\{\mathrm{tr}\ \tilde{L}Z : \begin{pmatrix} Z & z \\ z^t & 1 \end{pmatrix} \succeq 0, \quad \mathrm{diag}(Z) = z\}.$$

Here $\tilde{L} = \begin{pmatrix} 4Q & c - 2Qe \\ (c - 2Qe)^t & e^t Q e + 2c^t e \end{pmatrix}$.

In [28, 37] the equivalence of the two semidefinite relaxations is established, together with further connections between the two models.

So in summary, we have derived a generic imbedding of either a problem in $(-1,1)$ or in $(0,1)$ variables. This imbedding does not exploit any additional structure of the discrete problem. In fact, using only the integrality, the models can be further strengthened by including linear inequalities valid for all matrices in $\mathcal{F}$.

2.3 Generic linear inequalities

The basic semidefinite relaxation for (P) can be further refined by exploiting integrality of the feasible solutions. Suppose f is a vector with exactly three nonzero components, which can be either $+1$ or -1. Then $|x^t f| \geq 1$, for all $(-1,1)$ vectors x, hence $1 \leq (x^t f)(f^t x) = \mathrm{tr}\ (xx^t)(ff^t)$. Thus

$$\mathrm{tr}\ X(ff^t) \geq 1 \tag{4}$$

is a valid linear inequality for the semidefinite relaxation. Since the original inequality $|x^t f| \geq 1$ is squared, we may assume without loss of generality that $\sum f_i \geq 1$. In polyhedral combinatorics, these inequalities are called *Triangle Inequalities*, because they can be rewritten as $f_i f_j x_{ij} + f_i f_k x_{ik} + f_j f_k x_{jk} \geq -1$. They have been shown to be facet-defining for conv($\mathcal{F}$), if the elements l_{ij}, l_{ik} and l_{jk} of L are nonzero.

The above derivation suggests already a generalization to a much larger class of valid inequalities. Suppose f is an integer vector, such that $|x^t f| \geq 1$ for all $(-1,1)$ vectors x. (This holds for instance if $\sum_i f_i = 1$, or at least odd.) Then we have just as before that tr $X(ff^t) \geq 1$ is a valid inequality of the semidefinite model. These inequalities are called *Hypermetric Inequalities*. Even though there is an infinite number of them, it has been shown that they describe a polyhedron.

Theorem 2 *([12]) The set $\{x : |f^t x| \geq 1$ for all integer vectors f with $f^t e$ odd$\}$ is polyhedral.*

Unfortunately, it is difficult to optimize over this set. In fact it is shown in [3] that the separation problem for hypermetric inequalities is NP-hard. On the other hand, we can try to optimize over some tractable subclass of hypermetric inequalities such as the set of triangle inequalities. In [29], some computational experience is provided, showing that optimizing over the $4\binom{n}{3}$ triangle inequalities can be quite challenging. Moreover, computational experience is given, showing that the inclusion of hypermetric inequalities with $\sum_i |f_i| \geq 5$ only leads to a marginal improvement, as compared to the inclusion of "only" the triangle inequalities.

So the following improvement over (P_{sdp}) is obtained by including all triangle inequalities

$$f^t X f \geq 1$$

for $f \in \{-1,0,1\}^n$. Here f has exactly three nonzero terms, and at most one negative entry. We collect all $4\binom{n}{3}$ of these constraints in the operator $B(X)$, and get

$$\max\{\text{tr } LX : \ \text{diag}(X) = e, \ B(X) \geq e, \ X \succeq 0\}.$$

2.4 Semidefinite models through eigenvalues: Graph Equipartition

Up to now we used the linearization idea in a straightforward way to obtain semidefinite relaxations. Now we will show how some graph problems can be approximated quite nicely by eigenvalue approaches.

Connections between eigenvalues of matrices associated to graphs and basic graph invariants, such as largest degree, connectivity, chromatic number, etc. have been investigated already in the 1960s. The monograph [10] contains a thorough treatment of spectral graph theory. The strong potential of eigenvalue methods applied to some graph optimization problems is surveyed in the papers by Mohar and Poljak [42] and in Mohar [41].

Here, we consider the problem of partitioning the vertices of a graph into k sets of size $m = \frac{n}{k}$. Let k, m be integers, and $n := km$. Let G be a graph with $|V(G)| = n$, and (weighted) adjacency matrix A. We denote by $L := \text{Diag}(Ae) - A$ its Laplacian. The *k-equipartition problem* (k-GP) asks to partition V into k sets S_j of size m each, such that the total weight of the edges cut is minimized.

Let us model such partitions by $n \times k$ $(0,1)$ matrices Y, where column j of Y represents set S_j. Then a mathematical statement of (k-GP) is as follows, see e.g. [13, 55]

$$z = \min \tfrac{1}{2}\text{tr } Y^t L Y \ \text{ such that } \ Y^t e_n = m e_k, Y e_k = e_n, y_{ij} \in \{0,1\}.$$

Again, this is a computationally difficult optimization problem, so it is important to have tight tractable relaxations. A direct application of the generic imbedding would lead to a semidefinite problem with matrices of order $nk \times nk$. (This approach is investigated in some detail in the thesis [64] by Zhao.) However, the partition matrices Y are highly structured $(0,1)-$matrices, so exploiting this structure should provide some additional information to get good relaxations. Here we mention the following properties of partition matrices Y:

1. $Y^tY = mI_k$ (columns pairwise orthogonal)

2. $Ye = e, Y^te = me$ yielding $YY^te = me$ (constant row and column sums)

3. $YY^t \geq 0$ elementwise (nonnegativity)

4. $YY^t \succeq 0$ (semidefiniteness)

5. $\mathrm{diag}(YY^t) = e$ (rows of Y are unit vectors)

Historically, the first tractable relaxations for equipartition utilize *eigenvalue optimization*. Donath and Hoffman [13] use the orthogonality property of the columns of Y and the following result, often called *Fan's Theorem*; see for instance [47]. We denote by $\lambda_1(L) \leq \ldots \leq \lambda_n(L)$ the eigenvalues of a symmetric matrix L.

Theorem 3 *Let L be a symmetric matrix of order n, with eigenvalues $\lambda_i(L)$ and corresponding normalized and pairwise orthogonal respective eigenvectors p_i $(i = 1, \ldots, n)$. Let $k \leq n$ be given, and define $P = (p_1, \ldots, p_k)$.*

Then

$$\min\{tr\ Y^tLY : Y^tY = I_k\} \quad = \quad tr\ P^tLP \quad = \quad \sum_{i=1}^{k} \lambda_i(L).$$

Using this, and the fact that tr $Y^t\mathrm{diag}(u)Y = \mathrm{tr}\ \mathrm{diag}(u)YY^t = e^tu$, for any partition matrix Y, they obtain

$$z \geq \max_{e^tu=0} \min_{Y^tY=mI_k} \tfrac{1}{2}\mathrm{tr}\ Y^t(L + \mathrm{diag}(u))Y = \max\{\tfrac{m}{2} \textstyle\sum_{i=1}^{k} \lambda_i(L + \mathrm{diag}(u)) : u^te = 0\} =: z_{DH}.$$

This bound can be computed efficiently, because it is concave (in u). Unfortunately the function is nonsmooth, so classical methods of convex optimization fail.

In [55] this approach is further strengthened by including the constant row and column sum constraints. Let V be an arbitrary $n \times (n-1)$ matrix satisfying $V^te_n = 0,\ V^tV = I_{n-1}$. Further, let the $k \times (k-1)$ matrix V_k be such that $V_k^te_k = 0,\ V_k^tV_k = I_{k-1}$. (The properties of V imply that

$$VV^t = I - \tfrac{1}{n}e_ne_n^t, \tag{5}$$

because V can be extended to the orthogonal matrix $(V\ \frac{1}{\sqrt{n}}e_n)$. A similar relation holds of course for V_k.) The general solution of $Ye_k = e_n,\ Y^te_n = me_k$ is

$$Y = \tfrac{1}{k}e_ne_k^t + \sqrt{\tfrac{n}{k}}VZV_k^t, \tag{6}$$

where Z is an arbitrary $(n-1) \times (k-1)$ matrix. To see this we first note that any matrix Y of the form (6) satisfies $Ye_k = e_n,\ Y^te_n = me_k$. Suppose conversely that Y is given such that $Ye_k = e_n,\ Y^te_n = me_k$. Setting $Z := \sqrt{\frac{k}{n}}V^tYV_k$ in (6) shows that Y can be expressed in the form (6). Here one has to use (5).

If Y and Z are related by (6), then

$$Y^tY = \tfrac{m}{k}e_ke_k^t + mV_kZ^tZV_k^t \quad \text{and} \quad V_k^tY^tYV_k = mZ^tZ.$$

Setting $Z^tZ = I_{k-1}$ in the first equation yields $Y^tY = mI_k$. Setting $Y^tY = mI_k$ in the second gives $Z^tZ = I_{k-1}$. Again one has to use (5) to simplify terms of the form VV^t and $V_kV_k^t$. Thus we have shown; see [55]:

$$\{Y : Y^tY = mI_k,\ Ye_k = e_n,\ Y^te_n = me_k\} = \{\tfrac{1}{k}e_ne_k^t + \sqrt{\tfrac{n}{k}}VZV_k^t : Z^tZ = I_{k-1}\}. \tag{7}$$

Substituting this parameterization of Y through Z from (6) into the cost function $\frac{1}{2}\mathrm{tr}\ Y^t(L + \mathrm{diag}(u))Y$, one gets the following stronger relaxation; see [55].

$$z_{RW} := \max_{u^te=0}\ \min\ \{\tfrac{m}{2}\mathrm{tr}\ Z^tV^t(L + \mathrm{diag}(u))VZ : Z^tZ = I_{k-1}\}. \tag{8}$$

Computational experience with this relaxation on random graphs and real world data of substantial size is contained in [15]. Moreover, z_{RW} does not depend on the specific choice of V; see [55].

Now we investigate how these eigenvalue bounds relate to semidefinite relaxations. This approach was first proposed and investigated by [1], and further studied in [32, 33]. Let us linearize the cost function of (k-GP) directly: $\mathrm{tr}\ Y^tLY = \mathrm{tr}\ LYY^t = \mathrm{tr}\ LX$. (Here X correponds to YY^t, and represents the linearization.) We use the semidefiniteness property, and $YY^te = me$ together with the fact that rows of Y are unit vectors, and get the following basic semidefinite relaxation:

$$\min\{\tfrac{1}{2}\mathrm{tr}\ LX : X \succeq 0,\ \mathrm{diag}(X) = e,\ Xe = me\}.$$

The following result can be used to rewrite the Donath-Hoffman bound as a semidefinite program.

Lemma 4 *([16, 47])*

$$conv\{YY^t : Y^tY = I_k\} = \{X : tr\ X = k,\ I \succeq X \succeq 0\} =: M_k.$$

(If A and B are symmetric matrices of the same size, then $A \succeq B$ stands for $A - B \succeq 0$.) A discussion of various ways to prove this result can be found in [47].

This relation can be used as follows:

$$\min_{Y^tY=I_k} \mathrm{tr}\ Y^tLY = \min_{X \in M_k} \mathrm{tr}\ LX.$$

(This equation holds because $\mathrm{tr}\ Y^tLY = \mathrm{tr}\ L(YY^t)$, and we can replace minimizing over $\{YY^t : Y^tY = I_k\}$ by minimizing over $\mathrm{conv}\{YY^t : Y^tY = I_k\} = M_k$.) Using this one obtains, see [1, 33]:

$$\begin{aligned} z_{DH} &= \max_{u^te=0}\ \min_{X\in M_k}\ \tfrac{m}{2}\mathrm{tr}\ (L + \mathrm{diag}(u))X \\ &= \min_{X\in M_k}\ \max_{u^te=0}\ [\tfrac{m}{2}\ \mathrm{tr}\ LX + \tfrac{m}{2}\ u^t\mathrm{diag}(X)] \\ &= \min\ \{\tfrac{1}{2}\mathrm{tr}\ LX : \mathrm{diag}(X) = e,\ mI \succeq X \succeq 0\}. \end{aligned}$$

The second equality follows from exchanging min and max, and using strong duality; see Chapter VII.4 in [25] for further details on Lagrangian duality. The last equality can be seen as follows. If $\mathrm{diag}(X)$ is not a multiple of e, then the inner maximization is unbounded, hence we have a hidden constraint $\mathrm{diag}(X) = \alpha e$, which together with $\mathrm{tr}\ X = k$ leads to the final problem after rescaling X.

In a similar way it can be shown that the eigenvalue bound z_{RW} can equivalently be written as a Semidefinite Program; see [33]

$$z_{RW} = \min\{\tfrac{1}{2}\text{tr } LX : diag(X) = e,\ Xe = me,\ mI \succeq X \succeq 0\}.$$

We can get rid of the condition $mI - X \succeq 0$, if we include the nonnegativity condition $X \geq 0$ in the relaxation. This gives $\binom{n}{2}$ additional linear inequalities $x_{ij} \geq 0$.

Lemma 5 *([33])*

$$\textit{If}\ \ Xe = me,\ X \geq 0\ \ \textit{then}\ \ mI \succeq X.$$

Proof. $X \geq 0$ and $Xe = me$ leads to $\|X\|_\infty = m$; therefore the eigenvalues of X are no larger than m. □

Thus the following relaxation of (k-GP) dominates already the eigenvalue bounds z_{DH}, z_{RW}:

$$z \geq \min\{\tfrac{1}{2}\text{tr } LX : X \succeq 0,\ X \geq 0,\ Xe = me,\ diag(X) = e\}.$$

In [32, 33] this bound is also investigated computationally, leading also in practise to very tight approximations of (k-GP), especially for $k > 2$.

2.5 Semidefinite problems through geometry: Graph Coloring

Recently, Karger et al. proposed the following idea to approximate the chromatic number of a graph, see [31]. Suppose we assign unit vectors w_i to the vertices of a graph, such that the angle between any pair of vectors corresponding to adjacent nodes in the graph is fixed. First we observe that this angle depends in general on the dimension of the vectors w_i.

Lemma 6 *[31]*

$$\min\{\alpha : \exists w_1, \ldots, w_k \in I\!R^k, w_i^t w_i = 1\ \forall i,\ w_i^t w_j = \alpha\ \textit{for}\ i \neq j\} = \frac{-1}{k-1}.$$

Proof. Let $W = (w_i^t w_j)$. We have to find the smallest α such that $W = \alpha J + (1-\alpha)I$ and $W \succeq 0$. The eigenvalues of W are $\alpha k + (1-\alpha)$ and $1-\alpha$; therefore $1 \geq \alpha \geq \frac{-1}{k-1}$. □

Now suppose the graph has chromatic number $\chi = k$, with color classes $S_1, \ldots, S_k$. If we assign to vertices $v \in S_i$ the vector $w_i \in I\!R^k$ from the lemma, we obtain $w_i^t w_j = \frac{-1}{k-1}$ for $ij \in E(G)$ and $w_i^t w_i = 1$. In other words, pairs of vectors w_i, w_j from different color classes have a large angle between them. Therefore the following (semidefinite programming) problem

$$\alpha^* = \min \alpha \quad \text{such that } A(\alpha) \succeq 0,$$

where $A(\alpha)_{ij} = \alpha$ for $ij \in E,\ A(\alpha)_{ii} = 1$, has optimal value

$$\alpha^* \leq \frac{-1}{\chi - 1}.$$

From this one concludes

$$\chi \geq 1 - \frac{1}{\alpha^*}.$$

In fact, in [31] it is pointed out that this relaxation corresponds precisely to the theta-function introduced in [38].

3 Linear Programs over cones

In this section we describe the theoretical setting for semidefinite programs. We recall duality of linear programs over closed convex cones in detail, and recall duality results for semidefinite programs. The results of this section can be found in papers from the early sixties, such as [14, 6]. We review here the main concepts without proofs, and formulate the results in a setting which is general enough for the present purposes, but avoids deeper results from Hilbert space theory and functional analysis. The presentation here follows mostly [14]. We also mention similar approaches, contained in [48, 39, 60]. In [40], a self-dual formulation is investigated.

3.1 Deriving the Dual Problem

We formulate a pair of dual problems (P) and (D) as follows. Let S and T be a pair of closed convex cones. (A set S is called a *cone* if $s \in S$ implies $\alpha s \in S$ for all $\alpha \geq 0$.)

We denote by S^* the dual cone, defined by

$$S^* = \{y : x^t y \geq 0 \quad \forall x \in S\}.$$

Consider the problem

$$(P) \quad \inf c^t x \quad \text{such that } Ax - b \in S, x \in T,$$

where $c \in \mathbb{R}^n, b \in \mathbb{R}^m$ and A is $m \times n$.

Let us first derive the dual through the *Minimax inequality*

$$\inf_{x \in X} \sup_{y \in Y} f(x,y) \geq \sup_{y \in Y} \inf_{x \in X} f(x,y),$$

which is true for any real valued function f, defined on nonempty $X \times Y$. We need the following fact, which is an easy consequence of the separating hyperplane theorem for closed convex sets.

Proposition 7 *Let $S \subseteq \mathbb{R}^n$ be a closed convex cone. Then $S = S^{**}$.*

If $Ax - b \notin S$, then this result shows that there exists some $y \in S^*$ such that $(Ax-b)^t y < 0$. Thus we get

$$\sup_{y \in S^*} [c^t x - (Ax-b)^t y] = \begin{cases} c^t x & : \quad Ax - b \in S \\ +\infty & : \quad Ax - b \notin S \end{cases}$$

Using this observation we get the following chain of equalities and inequalities.

$$\begin{aligned} \inf c^t x \quad \text{such that } x \in T, Ax - b \in S \quad &= \quad \inf_{x \in T} \sup_{y \in S^*} c^t x - (Ax-b)^t y \\ &\geq \quad \sup_{y \in S^*} \inf_{x \in T} b^t y + (c - A^t y)^t x \\ &= \quad \sup b^t y \quad \text{such that } y \in S^*, \ c - A^t y \in T^* \end{aligned}$$

Therefore we call the last problem the dual (D) of (P).

$$(D) \quad \sup b^t y \quad \text{such that } c - A^t y \in T^*, y \in S^*.$$

This pair of problems is a generalization of dual pairs of Linear Programs. Setting $S := \mathbb{R}^m_+$, and $T := \mathbb{R}^n_+$, we get, using $(\mathbb{R}^m_+)^* = \mathbb{R}^m_+$,

$$\min c^t x \quad \text{such that } Ax - b \geq 0, x \geq 0, \text{ and its dual } \max b^t y \quad \text{such that } c - A^t y \geq 0, y \geq 0.$$

Lemma 8 *Let $T = \{X \succeq 0\}$ denote the cone of semidefinite matrices in the vector space of real symmetric matrices. Then T is selfdual, i.e. $T^* = T$.*

Proof. Let $X, Y \in T$. Thus there exist U, V with $X = UU^t$ and $Y = VV^t$. Then $\mathrm{tr}\ XY = \mathrm{tr}\ UU^tVV^t = \mathrm{tr}\ (U^tV)(U^tV)^t \geq 0$ and $T \subseteq T^*$. Now suppose Y is a symmetric matrix and $Y \notin T$. Then there exists y such that $y^tYy < 0$. Set $X := yy^t$. $X \in T$, but $\mathrm{tr}\ XY = y^tYy < 0$, so $Y \notin T^*$. □

By construction of (D), we have $c^tx \geq b^ty$, for any pair (x, y) of primal feasible and dual feasible solutions. This is the *weak duality inequality*, well-known from Linear Programming. In fact, one can even further strengthen this inequality by considering *asymptotically feasible solutions.*

A sequence (x_k) with $x_k \in T$ is called an *asymptotically feasible solution of (P)* if there exists a sequence $(s_k) \in S$ such that $Ax_k - s_k \to b$. Similarly, we define *asymptotically feasible dual solutions* as sequences $(y_k) \in S^*$ for which there exists another sequence $(w_k) \in T^*$ such that $A^ty_k + w_k \to c$. We call a problem *asymptotically feasible* if it has an asymptotically feasible solution.

Set

$$\begin{array}{lll} M = & \inf\ c^tx & \text{over } x \text{ feasible for (P)} \\ m = & \inf_{\{x_k\}} \liminf_{k\to\infty} c^tx_k & \text{over } \{x_k\} \text{ asymptotically feasible for (P)} \\ M' = & \sup\ b^ty & \text{over } y \text{ feasible for (D)} \\ m' = & \sup_{\{y_k\}} \limsup_{k\to\infty} b^ty_k & \text{over } \{y_k\} \text{ asymptotically feasible for (D).} \end{array}$$

3.2 Duality Theorems

The following duality theorems are taken from Duffin [14]. We include them without their proofs. (Duffin's arguments are elementary, and only rely on a separating hyperplane theorem, and the concept of dual cones.)

Theorem 9 (Weak duality.) *If (P) is feasible and (D) is asymptotically feasible, then $m' \leq M$.*

Analogously, there is weak duality between asymptotically feasible solutions of (P) and feasible solutions of (D), i.e. $m \geq M'$. However, weak duality between a pair of asymptotically feasible solutions of (P) and (D) does not hold in general, as shown by the following example.

Example 10 *Consider the following Linear Program over positive semidefinite matrices.*

$$\inf tr\ CX : \ A(X) = b, \ X \succeq 0, \quad \textit{with dual} \quad \sup b^ty : C - A^t(y) \succeq 0, \ y \quad \textit{unconstrained.}$$

Let $A(X) = \begin{pmatrix} x_{22} \\ 2x_{23} \end{pmatrix}, b = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, C = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. *Then* $A^t(y) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & y_1 & y_2 \\ 0 & y_2 & 0 \end{pmatrix}$. *It follows quite easily that both primal and dual are infeasible, so $M = \infty, M' = -\infty$. However, the matrices* $X_k = \begin{pmatrix} k & -1 & 0 \\ -1 & 2/k & 1 \\ 0 & 1 & k \end{pmatrix}$ *form an asymptotically feasible sequence, with tr $CX_k = 0$, hence $m \leq 0$.* *Using* $y_k = \begin{pmatrix} -k \\ 1 \end{pmatrix}, W_k = \begin{pmatrix} 1/k & 1 & 0 \\ 1 & k & 0 \\ 0 & 0 & 0 \end{pmatrix}$, *one can check that this defines an asymptotically* *feasible dual solution with $b^t y_k = 2$. Therefore $m' \geq 2$, and so $m \geq m'$ does not hold. In fact, we will see below that neither m nor m' can be finite.*

We are interested in sufficient conditions that guarantee that the chain of inequalities

$$M \geq m \geq m' \geq M'$$

holds with equality throughout, and that both primal and dual solution values are attained. Weak duality only gives $M \geq m' \geq M'$ and $M \geq m \geq M'$. Problem (P) is called *bounded* if it is feasible and M is finite. Similarly, (D) is called bounded if M' is finite.

Theorem 11 *([14][Theorem 1])* (Strong duality.) *Problem (P) is feasible and bounded if and only if problem (D) is asymptotically feasible and bounded. In this case, $M = m'$.*

Next we look at sufficient conditions guaranteeing that M or M' is attained, and $M = m$. We need to describe some constraint qualifications. (int S denotes the interior of S, $\mathrm{rel-int\ S}$ denotes the relative interior of S.)

Slater Condition. We say that problem (P) satisfies the *Slater condition* if there exists some $x \in T$ such that $Ax - b \in \mathrm{int}\ S$.

It is useful to consider another variant of the Slater condition, called *generalized Slater condition* under which Theorem 12 also holds.

Generalized Slater Condition. We say that problem (P) satisfies the generalized Slater condition if there exists $x \in \mathrm{rel-int\ T}$ such that $Ax - b \in \mathrm{rel-int\ S}$.

Observe that the constraint $Ax - b \in \mathrm{int}\ S$ is weakened (hence S does not need to be full dimensional) and, at the same time, the constraint $x \in T$ is strengthened to $x \in \mathrm{rel-int\ T}$.

Theorem 12 *[14][Theorem 2] If (P) satisfies the Slater condition (or the generalized Slater condition), and M is finite, then (D) is feasible, the dual supremum M' is attained, and $m = M = M' = m'$.*

Remark 13 *Theorem 12 is in general not valid when the generalized Slater condition is weakened to ask for the existence of $x \in T$ such that $Ax - b \in \mathrm{rel-int\ S}$.*

Remark 14 *For Linear Programs it can be useful to transform inequality constraints $Ax \geq b$ into equations by introducing additional sign-constrained variables $s: \ Ax - s = b, \ s \geq 0$. Formally, this amounts to replacing $Ax - b \in S, \ x \in T$, by*

$$[A \ \ -I] \begin{bmatrix} x \\ s \end{bmatrix} = b, \ x \in T, \ s \in S.$$

In other words, we can transform a cone-LP, defined over cones S, T into an equivalent cone-LP, defined over $\{0\}$ and $T \times S$.

Remark 15 *In the case $S = \{0\}$ the Slater condition as defined above can not hold, because S has no interior. The following condition is equivalent to Slater's constraint qualification:*

$$\exists x \in int\ (T) \ such\ that\ Ax = b.$$

To see this we observe that the general solution of $Ax = b$ can be expressed as $x = x_p + V\xi$, where x_p satisfies $Ax_p = b$ and $x_h = V\xi$ is the general solution of the homogeneous equation $Ax_h = 0$. Thus we can reformulate the feasible region $Ax = b$, $x \in T$ of (P) as $x_p + V\xi \in T$, ξ unconstrained. The Slater condition now asks for ξ such that $x_p + V\xi \in$ int (T).

In summary, for the Slater condition to hold, we may equivalently ask (in the case $S = \{0\}$), for $x \in$ int (T) such that $Ax = b$. This observation will be used below.

Semidefinite programs, as introduced above, fall into the duality framework just described. Instead of writing symmetric matrices as vectors in $I\!R^{\binom{n+1}{2}}$, we maintain the matrix structure, so that $A(x)$ becomes $A(X)$, with the linear operator A mapping symmetric matrices into $I\!R^m$. Therefore, the vector $A(X)$ is of the form $(\text{tr } A_1X, \ldots, \text{tr } A_mX)^t$, for given symmetric matrices A_i. Its adjoint operator A^t, defined through the adjoint identity

$$y^tA(X) = \text{tr } [A^t(y) \cdot X], \quad \text{for all } \; y \in I\!R^m, X \text{ symmetric},$$

is given by the matrix $A^t(y) = \sum_i y_iA_i$. Thus a general SDP can be written as

$$(P) \quad \max\{\text{tr } CX : a - A(X) = 0, \quad X \succeq 0\},$$

with corresponding dual

$$(D) \quad \min\{a^ty : Z = A^t(y) - C \succeq 0\}.$$

(Here we use again the fact that the cone of semidefinite matrices is selfdual.) In light of Remark 15, we say that (P) satisfies the Slater condition if there exists $X \succ 0$ with $A(X) = a$. Similarly, (D) satisfies the Slater condition if there exists y such that $A^t(y) - C \succ 0$.

Unlike in the Linear Programming case, SDP can have a nonzero but finite duality gap between the primal and the dual problem.

Example 16 *Let us consider the SDP with matrices of order 3,*

$\inf x_{33}$ *such that* $2x_{12} + x_{33} = 1,\; x_{22} = 0,\; X \succeq 0,$ *with dual*

$$\sup y_1 \text{ such that } \begin{pmatrix} 0 & -y_1 & 0 \\ -y_1 & -y_2 & 0 \\ 0 & 0 & 1-y_1 \end{pmatrix} \succeq 0.$$

Primal feasibility implies $x_{22} = x_{12} = x_{23} = 0$, and therefore $x_{33} = 1$. The primal problem is clearly feasible, and $M = 1$. The dual is only feasible for $y_1 = 0$ and $y_2 \leq 0$. Therefore $M' = 0$.

In general it is considered a computationally nontrivial task to recognize whether or not the duality gap is zero. Luo et al. [40] investigate this question both from a theoretical and a computational point of view.

On the other hand, if SDP arises as a relaxation from an integer problem, it is often possible to give an explicit construction for a (primal) matrix X satisfying a constraint qualification (Slater or generalized Slater condition). Recall that the semidefiniteness condition of the primal matrix X is obtained from the fact that $X \in \text{conv}(\mathcal{F})$ implies $X \succeq 0$. As before $\mathcal{F} = \{xx^t :$ x feasible for the integer problem $\}$. Thus, if $\text{conv}(\mathcal{F})$ is full dimensional, then $\text{conv}(\mathcal{F})$ contains positive definite matrices and so the Slater condition trivially holds in view of Remark 15. In any case we can take a point in the relative interior of $\text{conv}(\mathcal{F})$ to satisfy the generalized Slater condition. If the feasible set of the integer problem has some regular structure, it is usually possible to provide explicitly a (generalized) Slater point.

4 Algorithms

In this section we describe the currently most promising ways to solve semidefinite programs. The key idea is to use interior-point techniques, applied to an auxiliary barrier problem. The methods fall into the class of primal-dual path-following algorithms.

We consider the following primal-dual pair.

$$(P) \quad \max\{\mathrm{tr}\ CX : a - A(X) = 0, \quad X \succeq 0\},$$

$$(D) \quad \min\{a^t y : Z = A^t(y) - C \succeq 0\}.$$

We assume that the linear equations in the primal are linearly independent, and that both the primal and the dual satisfy the Slater condition, so there exist X and Z positive definite, such that $A(X) = a, Z = A^t(y) - C$. In this case we have strong duality, both bounds are attained and the optimum is finite. The following characterization of optimality is well known. The presentation here follows [29].

Lemma 17 *Suppose that both (P) and (D) satisfy the Slater condition. In this case (X, y, Z) is optimal for (P),(D) if and only if X is feasible for (P), (y, Z) is feasible for (D), and $ZX = 0$.*

Thus to solve (SDP), one needs to solve the above system of equations. It turns out that this can be done quite efficiently by slightly changing the complementarity condition $ZX = 0$.

Consider, for fixed $\mu > 0$, the following dual barrier problem

$$\min_{y,Z} a^t y - \mu \log\det Z \ \text{ such that } \ Z = A^t(y) - C, \ Z \succeq 0,$$

with Lagrangian $L_\mu(X, y, Z) = a^t y - \mu \log\det Z + \mathrm{tr}\ [(Z - A^t(y) + C)X]$. The cost function of the barrier problem is strictly convex, and can be shown to possess a unique minimizer (X_μ, y_μ, Z_μ), characterized by the first order optimality conditions of the Lagrangian L; see e.g. [17]:

$$\begin{aligned} \nabla_y L_\mu = a - A(X) &= 0 \\ \nabla_X L_\mu = Z - A^t(y) + C &= 0 \\ \nabla_Z L_\mu = X - \mu Z^{-1} &= 0, \ X \succ 0, \quad Z \succ 0. \end{aligned}$$

The smooth curve $\{(X_\mu, y_\mu, Z_\mu) : \mu > 0\}$ is usually called the *central path.* Further details on the existence of the central path for convex sets having a *selfconcordant* barrier function can be found in the monograph [44].

The basic idea to find a solution of (P) is to solve this system iteratively using Newton's method. After having found an approximate solution for some specified value of μ, this parameter is reduced and one iterates until μ is small enough. This sounds simple, and yet one can prove global convergence, and local superlinear convergence, provided μ is reduced according to some quite simple rules; see for instance [2, 29, 35, 43, 45].

There is currently intensive scientific interest in the question of how the system should be solved. The complication comes from the fact that the complementarity condition $X = \mu Z^{-1}$ can be written in many equivalent forms, such as

$$ZX = \mu I, \quad Z = \mu X^{-1}, \quad ZX + XZ = 2\mu I, \ldots$$

Each form has a different linearization, and thus gives a different search direction $(\Delta X, \Delta y, \Delta Z)$. Linearizing the first equation $(Z+\Delta Z)(X+\Delta X) = \mu I$ leads to $Z(\Delta X)+(\Delta Z)X = \mu I - ZX$, while linearizing the second equation $Z+\Delta Z = \mu(X+\Delta X)^{-1}$ gives $\Delta Z + \mu X^{-1}(\Delta X)X^{-1} = \mu X^{-1} - Z$. This follows from the expansion of the inverse, correct up to first order terms

$$(X + \Delta X)^{-1} = (I + X^{-1}\Delta X)^{-1}X^{-1} \approx X^{-1} - X^{-1}(\Delta X)X^{-1}. \tag{9}$$

We note in particular that $(I + A)^{-1} \approx I - A$ can be derived similar to the scalar relation $(1 + a)^{-1} \approx 1 - a$, which holds for small reals a. Zhang [63] proposes an interesting way to get some unification of these different forms. Suppose P is invertible and consider (for fixed P) the mapping

$$H_P(X) = \tfrac{1}{2}\left[P(X)P^{-1} + (P(X)P^{-1})^t\right].$$

Zhang shows that for $X \succ 0, Z \succ 0$,

$$H_P(ZX) = \mu I \ \text{ if and only if } \ ZX = \mu I.$$

Therefore the first order conditions of the barrier problem can be represented, for some fixed invertible matrix P, as

$$A(X) = a, \quad A^t(y) - Z = C, \quad H_P(ZX) = \mu I, \quad X \succ 0, Z \succ 0.$$

The following variants are currently quite popular ones, and have been investigated both theoretically and computationally in quite some depth. (There are many more variants, not included below. It is beyond the scope of this survey to discuss them all.)

1. AHO direction ($P = I$), proposed by Alizadeh et al [2].
2. HKM direction ($P = Z^{-1/2}$), proposed independently by Helmberg et al [29], Kojima et al [35] and later reintroduced by Monteiro [43].
3. NT direction ($P = W^{-1/2}$, $W = X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}$), introduced by Nesterov, Todd [45].

These variants were actually introduced before Zhang's paper, and were derived under quite different motivations.

It is worth mentioning that the HKM direction does not necessarily lead to a symmetric update ΔX, if Newton's method is applied. However, by taking the symmetric part of ΔX (and thereby changing the Newton direction), one can still prove global convergence [24], and local superlinear convergence [35].

It is well known that convex optimization problems can be solved in polynomial time to a prescribed precision, provided the separation problem for the feasible region is tractable. SDP belongs to this class of problems. The currently strongest theoretical complexity results indicate that it takes $O(\sqrt{n}\log t)$ Newton iterations to reduce the duality gap by a factor of t, given initial Slater points $X_0, (y_0, Z_0)$ of (P) and (D) respectively. Since the literature on algorithms for SDP is growing so quickly, we refer the reader to the *Interior-Point Methods online* homepage [1] for further details. In the following we describe some methods which in practise have turned out to be more efficient than the theoretically polynomial methods. The convergence analysis of these variants is still under investigation.

[1] `http://www.mcs.anl.gov/home/otc/InteriorPoint`

After having chosen a representation of the complementarity condition, e.g. one of the variants described above, Newton's method is used to solve the system.

There is currently intensive scientific discussion going on, as to how this system should be solved to maintain numerical stability. The following choices are currently proposed.

- (partial) Schur complements and backsubstitution; see for instance [2, 29]
- Least Squares approach; see [36]
- Iterative methods; see [64].

At present, there is no clear winner as to what is the most efficient way to solve the system. The conflicting goals are computational efficiency versus numerical accuracy and stability.

In [59], computational evidence is provided that the NT direction achieves higher accuracy than both the AHO and the HKM directions, at only moderately higher computational effort. Helmberg et al [29] show that the HKM direction is superior to the other approaches, if computation time is more important than full accuracy.

In the context of applying Semidefinite Programming to approximate Integer Programs, it is not crucial to find optimal solutions to very high precision, because integer data in the original Integer Problem usually allow some rounding arguments to stop the approximation early. The optimal solution value needs to be accurate only on the integer part if the relaxation is used in a Branch and Bound setting. Therefore computational efficiency is the primary target. There seems to be a general agreement, also supported by computational work contained in [2, 59], that the HKM direction requires the least amount of computational work.

Therefore we briefly go through the main steps to perform one iteration of Newton's method for the HKM method. Recall that the system to be solved (for fixed $\mu > 0$) is

$$A(X) = a, \quad A^t(y) - Z = C, \quad ZX = \mu I, \quad Z \succ 0, \quad X \succ 0.$$

Suppose we have a triple (X, y, Z), only violating the complementarity condition $ZX = \mu I$. (If the SDP arises from some structured integer program it is usually possible to find such a triple with little computational effort.) To get the search direction $(\Delta X, \Delta y, \Delta Z)$, we have to solve the following linear system:

$$\begin{aligned} A(\Delta X) &= 0, \\ \Delta Z &= A^t(\Delta y) \\ Z(\Delta X) + (\Delta Z)X &= \mu I - ZX \end{aligned}$$

The second equation is used to eliminate ΔZ. The third gives

$$\Delta X = \mu Z^{-1} - X - Z^{-1} A^t(\Delta y) X. \tag{10}$$

This is substituted into the first equation, leading to a linear equation for Δy,

$$A(Z^{-1} A^t(\Delta y) X) = \mu A(Z^{-1}) - a.$$

It can be shown that this system is positive definite, but gets increasingly ill-conditioned: as μ tends to 0, Z will become nearly singular and Z^{-1} is used explicitly. Given Δy, we backsubstitute for ΔZ and ΔX. Equation (10) gives in general a nonsymmetric ΔX. We replace ΔX by its symmetric part, and obtain the search direction $(\Delta X, \Delta y, \Delta Z)$. Even though this direction is not the Newton direction any more (because we changed ΔX), it can still be shown that this search direction enjoys global convergence ([29, 24]). For local superlinear convergence see [35].

The step length along this search direction is determined in the usual way: First one tries a full step (step length $\alpha = 1$), and one backtracks with a geometric progression, if the current step is too large, i.e. takes us out of the cone of positive definite matrices.

The computationally crucial steps are the following (using cubic matrix multiplication algorithms).

- Form the coefficient matrix for the system for Δy. If there are m equality constraints $\mathrm{tr}\ A_i X = a_i$, with unstructured (dense) matrices A_i, then the computational effort to build the coefficient matrix for given X, Z^{-1} is at least $c(mn^3 + m^2 n^2)$, for some constant c.
- Compute ΔX. This will take time $O(n^3 + mn^2)$.
- Line Search. Here we have to find a step size α, $0 < \alpha \leq 1$, such that $X_1 = X + \alpha \Delta X \succ 0$. For given α, this amounts to checking whether or not the Cholesky decomposition of X_1 exists. Typically only a few backtracking steps are necessary to find the right α. If the number of backtracking steps is independent of n, then the line search takes $O(n^3)$ operations. (Practical experience indicates that this is usually the case.)

In summary, the most expensive step is forming the coefficient matrix. If both m and n are of moderate size, say less than 500, this approach is still feasible. On the other hand, n larger than say 1000, makes this approach expensive both with respect to computation time and also computer memory.

5 Connections to other nonlinear relaxations

The semidefinite relaxation of a quadratic program is historically not the first nonlinear relaxation. In this section we review various other tractable relaxations, and show how they relate to the semidefinite model. The key observation will be that many of these models lead to the same bound.

In this section we will work with functions defined either on $\mathbb{R}^n$ or $\mathbb{R}^{n+1}$. For the convenience of the reader the following convention will be used: the vectors x, u will always be in $\mathbb{R}^n$, while y and v will be in $\mathbb{R}^{n+1}$. Similarly U will be an $n \times n$ matrix, while V and Y denote matrices of order $n+1$.

Various tractable relaxations of the following quadratic integer program (QP) have been proposed in the literature.

$$\text{(QP)} \quad \mu^* := \max\left\{ q(x) : x_i^2 = 1, \quad i = 1, \ldots, n \right\},$$

with $q(x) := x^t Q x - 2c^t x$. For any vector u, let

$$q_u(x) := x^t (Q + U) x - 2c^t x - u^t e,$$

where we use U to denote $\mathrm{Diag}(u)$. In this case $q_u(x) = q(x) + \sum u_i (x_i^2 - 1)$, so $q_u(x) = q(x)$ for any $x \in \{-1, 1\}^n$.

The first tractable relaxation of (QP) is obtained using the *Trust Region Problem* (TR), which is defined for fixed $s > 0$ as follows.

$$\text{(TR)} \quad \max\{q(x) : x^t x = s^2\}.$$

We get, for any vector $u \in I\!\!R^n$,

$$\mu^* = \max\{q_u(x) : x_i^2 = 1 \ \forall i\} \leq \max\{q_u(x) : x^t x = n\}.$$

Therefore

$$\mu^* \leq B_{tr} := \min_u \max_{x^t x = n} q_u(x).$$

The inner maximization can be done efficiently, see [20, 56]. It is less clear how the outer minimization is carried out.

Another way to obtain a relaxation consists in using *convex quadratic programming* for the inner maximization. We have

$$\mu^* \leq \max\{q_u(x) : -1 \leq x_i \leq 1\}$$

for any $u \in I\!\!R^n$. Requiring q_u to be concave, i.e. $Q+U \preceq 0$, gives a convex quadratic programming problem, which can be solved in polynomial time. This leads to the bound B_{cp}.

$$\mu^* \leq B_{cp} := \min_{Q+U \preceq 0} \max_{|x_i| \leq 1} q_u(x).$$

Another alternative is to introduce an artificial variable $x_0 = 1$, and homogenize the original cost function, as described in section 2. Let us denote the augmented matrix by

$$Q^c := \begin{pmatrix} Q & -c \\ -c^t & 0 \end{pmatrix}.$$

We get the following relaxation, which is based on *eigenvalue optimization*. Here $v \in I\!\!R^{n+1}, V = \mathrm{Diag}(v)$.

$$\begin{array}{rcl} \mu^* & = & \max\{y^t(Q^c - V)y + e^t v : y_i^2 = 1 \ \forall i, y \in I\!\!R^{n+1}\} \\ & \leq & \max\{y^t(Q^c - V)y + e^t v : y^t y = n+1\} \\ & = & (n+1)\lambda_{\max}(Q^c - V) + e^t v. \end{array}$$

So we get

$$\mu^* \leq B_{eo} := \min\{(n+1)\lambda_{\max}(Q^c - V) + e^t v : v \in I\!\!R^{n+1}\}.$$

This is a nonsmooth but convex optimization problem, and has been studied thoroughly in nonsmooth optimization; see for instance [9, 46].

We can also apply semidefinite programming relaxations, as described at the beginning. The semidefinite relaxation of the homogenized problem is

$$\mu^* \leq B_{sd} := \max\{\mathrm{tr}\ Q^c Y : \ Y \succeq 0, \ \mathrm{diag}(Y) = e\}.$$

Finally we can use the Minimax inequality to get the *Lagrangian Dual* B_{ld}:
$\mu^* = \max_x \min_u q_u(x) \leq \min_u \max_x q_u(x) =: B_{ld}$. (Note that x and u are unconstrained.)

We now summarize the different relaxations.

$$\begin{aligned}
\mu^* &= \max\{q(x) : x_i^2 = 1\ \forall i\} \\
B_{ld} &= \min_u \max_x q_u(x) \\
B_{tr} &= \min_u \max_{x^tx=n} q_u(x) \\
B_{cp} &= \min_{Q+U\preceq 0} \max_{|x_i|\le 1} q_u(x) \\
B_{eo} &= \min_v [(n+1)\lambda_{\max}(Q^c - V) + e^tv] \\
B_{sd} &= \max\{\mathrm{tr}\ Q^cY : Y \succeq 0,\ \mathrm{diag}(Y) = e\}
\end{aligned}$$

Poljak and Wolkowicz [53] were probably the first to realize that all these bounds are equal; see also [54]. We include equivalence proofs here again, because they constitute a nice application of various duality results. We first recall some useful facts about the trust region problem (TR). It is well known that a vector x together with a multiplier λ for the equality constraint $x^tx = s^2$ solves (TR) if and only if

$$x^tx = s^2, \quad (Q - \lambda I)x = c, \quad Q - \lambda I \preceq 0. \tag{11}$$

This can be found e.g. in Fletcher [17]. We say for short that a pair (x, λ) satisfying (11) *solves* (TR). The characterization of optimality from (11) can be used to establish the following max-min relation.

Lemma 18 *Let $L(x, \lambda) := q(x) + \lambda(s^2 - x^tx)$ denote the Lagrangian of (TR). Then*

$$\max\{q(x) : x^tx = s^2\} = \max_x \min_\lambda L(x, \lambda) = \min_\lambda \max_x L(x, \lambda).$$

Proof. Implicit proofs of this result can be found in several articles. The paper [58] contains an explicit proof in a much more general setting. Here we sketch the approach of [58] and refer also to [56] for further details. One gets

$$\begin{aligned}
\max_x \min_\lambda L(x, \lambda) &\le \min_\lambda \max_x L(x, \lambda) && (12) \\
&= \min_{\lambda:\ Q-\lambda I\preceq 0} \max_x L(x, \lambda) && (13) \\
&= \min_{\lambda, x:\ Q-\lambda I\preceq 0,\ (Q-\lambda I)x=c} L(x, \lambda). && (14)
\end{aligned}$$

To see the last equations, note that for λ fixed, $\max L(\cdot, \lambda)$ is unbounded, unless $Q - \lambda I \preceq 0$. In this case $L(\cdot, \lambda)$ is concave, hence a maximizing x must satisfy the first order optimality condition $\nabla_x L = 0$, i.e. $(Q - \lambda I)x = c$. To prove the result, it is sufficient to exhibit a pair (x, λ) satisfying the optimality conditions of (TR): this pair (x, λ) is clearly feasible for the minimization problem (14) and has the same objective function value as the lower bound in (12), given by the optimal value of (TR).

Now suppose $\lambda > \lambda_{\max}(Q)$. In this case $Q - \lambda I$ is invertible and the first order condition leads to $x_\lambda := (Q - \lambda I)^{-1}c$. The Lagrangian becomes $L(x_\lambda, \lambda) = \lambda s^2 - c^t(Q - \lambda I)^{-1}c =: h(\lambda)$. Using (9), we get

$$h'(\lambda) = s^2 - c^t(Q - \lambda I)^{-2}c = s^2 - x_\lambda^t x_\lambda \quad \text{and} \quad h''(\lambda) = -2c^t(Q - \lambda I)^{-3}c \ge 0,$$

so h is convex for $\lambda > \lambda_{\max}(Q)$. It is easy to verify that $\lim_{\lambda\to+\infty} h(\lambda) = +\infty$. If the minimum of h is attained at a critical point λ with $\lambda > \lambda_{\max}(Q)$, leading to $h'(\lambda) = 0$ or equivalently $x_\lambda^t x_\lambda = s^2$, then we are done, because the pair (x_λ, λ) also solves (TR).

Otherwise, the minimium is attained at $\lambda = \lambda_{\max}(Q)$ (because at the optimal λ, $Q - \lambda I \preceq 0$). In this case we must have

$$\lim_{\lambda \searrow \lambda_{\max}} h'(\lambda) \geq 0. \tag{15}$$

Since $L(x, \lambda_{\max})$ must now be finite for some x, we conclude from (14) that $(Q - \lambda_{\max} I)x = c$ is solvable for x, otherwise the minimum in (14) would be infinite. Let the columns of the matrix P contain a basis of the eigenspace corresponding to $\lambda_{\max}$, i.e. $(Q - \lambda_{\max} I)P = 0$. Then $P^t c = P^t(Q - \lambda_{\max} I)x = 0$, so c is orthogonal to the columns of P.

Let x_0 denote the solution of minimum norm to $(Q - \lambda_{\max} I)x = c$. This implies that x_0 is orthogonal to the nullspace of $Q - \lambda_{\max} I$, i.e. to the columns of P. It follows from (15) that $x_0^t x_0 \leq s^2$. Let x_1 be an arbitrary vector in the nullspace of $Q - \lambda_{\max} I$, of norm $\|x_1\|^2 = s^2 - \|x_0\|^2$ and define $x = x_0 + x_1$. Clearly $x_0 \perp x_1$, hence $x^t x = s^2$. We also have $(Q - \lambda_{\max} I)x = c$. Therefore by (11) x is also a minimizer of (TR). $\square$

Theorem 19 $B_{tr} = B_{ld}$.

Proof. The proof is based on the max-min relation from the previous lemma with $s = \sqrt{n}$:

$$\max\{q(x) : x^t x = n\} = \max_x \min_\lambda L(x, \lambda) = \min_\lambda \max_x L(x, \lambda).$$

We have, using this relation,

$$B_{tr} = \min_u \min_\lambda \max_x [q_u(x) + \lambda(n - x^t x)] = \min_w \max_x q_w(x) = B_{ld}.$$

We used $w = u - \lambda e, e^t w = e^t u - \lambda n$. $\square$

Theorem 20 $B_{cp} = B_{ld}$.

Proof. We use strong duality for convex quadratic programs (q_u concave):

$$\max_{x_i^2 \leq 1} q_u(x) = \max_x \min_{\lambda_i \geq 0} [q_u(x) + \sum_i \lambda_i(1 - x_i^2)] = \min_{\lambda_i \geq 0} \max_x [q_u(x) + \sum_i \lambda_i(1 - x_i^2)]. \tag{16}$$

The $\lambda_i \geq 0$ are the Lagrange multipliers for $1 - x_i^2 \geq 0$ and are collected in the vector λ. We get:

$$\begin{aligned} B_{cp} &= \min_{Q+U \preceq 0} \max_{x_i^2 \leq 1} q_u(x) \\ &= \min_{Q+U \preceq 0} \min_{\lambda \geq 0} \max_x q_u(x) + \sum \lambda_i (1 - x_i^2) \\ &= \min_w \max_x q_w(x) = B_{ld}. \end{aligned}$$

The first equation is the definition of B_{cp}, the second equation follows from (16). In the last equation we combine u and λe into w. Finally $Q + W \not\preceq 0$ makes the inner maximization unbounded, so $Q + W \preceq 0$ may be assumed. $\square$

To show $B_{eo} = B_{tr}$, we use the following a characterization of optimality for (TR).

Lemma 21 *[56] (x, λ) solves (TR) if and only if there exists $t \in I\!R$ such that for*

$$D = \begin{pmatrix} Q & -c \\ -c^t & t \end{pmatrix}, D\begin{pmatrix} x \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} x \\ 1 \end{pmatrix}, x^t x = n, \quad \textit{and } \lambda = \lambda_{\max}(D).$$

Proof. This result is (implicitly) proved in [56]. The present argument is elementary, and uses besides (11) only the interlacing theorem for symmetric matrices.

First note that if t exists with the desired properties, then multiplying out $D\binom{x}{1} = \lambda\binom{x}{1}$ shows $(Q - \lambda I)x = c$, and $Q - \lambda I \preceq 0$, because $\lambda = \lambda_{\max}(D)$.

Conversely suppose that (x, λ) satisfies (11). Define $t := c^t x + \lambda$ and set $D = \begin{pmatrix} Q & -c \\ -c^t & t \end{pmatrix}$. Then $D\binom{x}{1} = \lambda\binom{x}{1}$ follows from the definition of t and $(Q - \lambda I)x = c$. Moreover $x^t x = n$ as required.

It remains to show that λ is the largest eigenvalue of D. First note that $Q - \lambda I \preceq 0$ shows that $\lambda \geq \lambda_{\max}(Q)$. We distinguish two cases.

Case 1. $\lambda > \lambda_{\max}(Q)$. Then it follows from the fact that λ is an eigenvalue of D and Q is principal submatrix of D that $\lambda = \lambda_{\max}(D)$ because of the interlacing theorem applied to D and Q:

$$\lambda_1(D) \leq \lambda_1(Q) \leq \lambda_2(D) \leq \lambda_2(Q) \leq \ldots \leq \lambda_n(Q) \leq \lambda_{n+1}(D).$$

Case 2. $\lambda = \lambda_{\max}(Q)$ with multiplicity $r \geq 1$. Therefore there exists an $n \times r$ matrix P whose columns are orthonormal eigenvectors to the eigenvalue λ: $P^t P = I_r$, $(Q - \lambda I)P = 0$. From this it follows that $c^t P = x^t(Q - \lambda I)P = 0$. Therefore $D\binom{P}{0} = \binom{QP}{-c^t P} = \binom{\lambda P}{0} = \lambda\binom{P}{0}$. Therefore we have at least $r + 1$ linearly independent eigenvectors of D corresponding to λ, namely $\binom{x}{1}$ and the r columns of $\binom{P}{0}$.

If $\lambda_{\max}(D) > \lambda$, then there would be at least $r + 2$ eigenvalues of D at least as large as λ. But then the interlacing theorem applied again to D and Q would imply that Q has at least $r + 1$ eigenvalues at least as large as λ, a contradiction. $\square$

Theorem 22 *[53] $B_{tr} = B_{eo}$.*

Proof. Let $B_{tr} = \max_{x^t x = n} q_u(x)$, so the bound is attained for $u \in I\!R^n$. Using the previous Lemma, we conclude that an optimal solution (x, λ) of the trust region problem leads to $t := c^t x + \lambda, y := \binom{x}{1}, Dy = \lambda y$, and $\lambda = \lambda_{\max}(D)$. Here $D = \begin{pmatrix} Q + U & -c \\ -c^t & t \end{pmatrix}$. Define $v \in I\!R^{n+1}$ by

$$v = \begin{pmatrix} u \\ t \end{pmatrix} - \frac{t}{n+1} e.$$

Note that $\sum v_i = \sum u_i$. Then a short calculation shows that

$$D - \frac{t}{n+1} I = \begin{pmatrix} Q & -c \\ -c^t & 0 \end{pmatrix} + V = Q^c + V,$$

where $V = \mathrm{Diag}(v)$. Therefore $\lambda_{\max}(Q^c + V) = \lambda - \frac{t}{n+1}$. Substituting in $t = c^t x + \lambda$ and using $(Q + U - \lambda I)x = c$ we get $x^t(Q + U)x = c^t x + n\lambda$. Thus

$$B_{tr} = x^t(Q + U)x - 2c^t x - u^t e = n\lambda - c^t x - e^t u = (n + 1)\lambda_{\max}(Q^c + V) - e^t v \geq B_{eo}.$$

Therefore $B_{eo} \leq B_{tr}$.

To see the reverse inequality, it is sufficient to note that $\max\{y^t(Q^c - V)y : y^t y = n+1, y_{n+1} = 1\} = \max\{q_u(x) : x^t x = n\}$. □

Finally, it is an easy exercise to show that B_{eo} and B_{sd} are in fact dual pairs of semidefinite programs. Both problems satisfy the Slater condition: take $Y = I$ for B_{sd}, and $y_i = \lambda_{\max}(L) + 1 \; \forall i$ in (17). (For details, see the discussion preceding equation (18) below.)

Theorem 23 $B_{eo} = B_{sd}$.

Now that we know that these bounds are all equal, how does this help? First we note that all bounds with the exception of B_{sd} are "dual", leading to a valid upper bound on μ^* even if the outer minimization is stopped early. In contrast, when we use B_{sd}, we have to find the optimal Y.

Using the equivalence, we can focus on efficiently calculating at least one of the above bounds. It is currently not clear which of the bounds can be obtained fastest. The semidefinite model B_{sd} does have the advantage that "primal information" can easily be included. We will see in the following sections that combining B_{sd} with linear cutting planes, such as the triangle inequalities, leads to significant improvements of the original bound B_{sd}.

6 Max-Cut approximation using semidefinite programming

In this section we look at theoretical performance guarantees of a simple semidefinite relaxation for Max-Cut. The approach was introduced in a fundamental paper by Goemans and Williamson. We also look at ways to improve this approximation.

Semidefinite Programs have recently turned out be be very useful in the performance analysis of polynomial approximation algorithms for Max-Cut and coloring, see [22, 31]. Based on the basic semidefinite relaxation (3) for Max-Cut, Goemans and Williamson proved the following.

Theorem 24 *([22]) There exists a polynomial time algorithm that provides two numbers z_L, z_{sdp} such that $z_L \leq z_{mc} \leq z_{sdp}$ and $\frac{z_L}{z_{sdp}} \geq .87856$ hold for all graphs G on nonnegative edge weights.*

Later Karloff [34] showed that the analysis of the rounding algorithm in [22] can not be improved. On the negative side, there is a recent result of Hastad, showing that it is difficult to approximate Max-Cut tightly.

Theorem 25 *([23], Theorem 4.2) For any $\varepsilon > 0$ it is NP-hard to approximate (undirected) Max-Cut within a factor of $16/17 + \varepsilon$.*

An approach similar to [22] was used in [31] to devise a polynomial algorithm to color a 3-colorable graph with at most $O(n^{0.25})$ colors. The currently best known approximation result for 3-colorable graphs by Blum and Karger leads to $O(n^{0.2143})$ colors; see [7]. Further approximation results are contained in [4, 5, 18].

Finally we recall a different type of analysis for (3), which was published even before [22]. Delorme and Poljak [11] look at the dual of (3),

$$\min\{e^t y : \mathrm{Diag}(y) - L \succeq 0\}, \tag{17}$$

and reformulate it as an eigenvalue optimization problem. Parameterizing y as $y = \alpha e - v$, where $e^t v = 0$, one gets $e^t y = n\alpha$, and $L - \mathrm{Diag}(y) = L + \mathrm{Diag}(v) - \alpha I \preceq 0$ if and only if $\lambda_{\max}(L + \mathrm{Diag}(v)) \leq \alpha$. Therefore the dual can equivalently be written as

$$z_{sdp} = \min_{e^t v=0} n\lambda_{\max}(L + \mathrm{Diag}(v)). \tag{18}$$

To get some estimate of $\lambda_{\max}(L + \mathrm{Diag}(v))$ for a random graph, the following result of Juhasz is useful. It shows that for a random graph with fixed edge probability, the largest eigenvalue of the adjacency matrix is of order n. Then there is a large drop, and all other eigenvalues are $o(n)$ with high probability.

Theorem 26 *([30]) Let $\epsilon > 0$. Given a fixed p with $0 < p < 1$, for almost all graphs of order n with edge probability p,*
(i) the largest eigenvalue of the adjacency matrix is equal to $\lambda_n = np(1 + o(1))$,
(ii) $\max_{i<n} |\lambda_i| = O(n^{1/2+\epsilon})$.

Delorme and Poljak [11] use this as follows. For $p \in (0, 1)$ fixed, let $G \in \mathcal{G}_{n,p}$ be a random graph of order n, with adjacency matrix A. (Graphs in $\mathcal{G}_{n,p}$ have edges appearing with probability p independent of the other edges.) Then the average degree d is of order n with high probability. Let m denote the number of edges of G. Since at least half the edges are cut by a maximum cut, and since z_{mc} as defined in (2) yields four times the value of the maximum cut, we get $z_{mc} \geq 2m = nd$ for any graph. To get an upper bound on (18), the following choice of v is proposed in [11]: $v_i = d - d_i$, where d_i is the degree of vertex i. Note that $\sum v_i = 0$, and $L + \mathrm{Diag}(v) = dI - A$. Therefore $\lambda_{\max}(L + \mathrm{Diag}(v)) = d - \lambda_{\min}(A) \leq d + O(n^{1/2+\epsilon})$ with high probability. Therefore

$$\frac{z_{mc}}{z_{sdp}} \geq \frac{1}{1 + O(n^{1/2+\epsilon})/d} = 1 - o(1).$$

This shows that on most random graphs with fixed edge probability p, the basic semidefinite relaxation (3) is asymptotically tight.

7 Practical Experience with Semidefinite Relaxations

In this section we show some computational work related to the approximation of Max-Cut

In the previous section we have seen that semidefinite relaxations can be used to obtain provably good approximations for Max-Cut. Moreover, the average case behaviour for certain classes of random graphs is even more encouraging. At the present time, similar results for the other problems mentioned in section 2 are not available.

We consider the basic relaxation (3) together with the following stronger model

$$z_{sdpt} = \max\{\mathrm{tr}\ LX : \mathrm{diag}(X) = e, f^t X f \geq 1,\ \forall f \in T, X \succeq 0\}.$$

Here T denotes (a subset of) the set of all vectors f having exactly three nonzero components, which are either 1 or -1; see (4). Inequalities of type (4) are usually called triangle inequalities, because the nonzero components in ff^t form the edge set of a triangle.

It is clearly out of the question to include all $4\binom{n}{3}$ inequalities of this type. Therefore we briefly explain how we select a limited number of them. We assume to have available a routine that, given a symmetric matrix X, returns the k most violated inequalities of the form (4). (This can be done by complete enumeration.)

The optimal X to (3) will in general not satisfy all triangle inequalities. We select the k most violated ones, with e.g. $k = n/5$. (Of course the more violated triangles we include, the stronger the resulting bound, but we also have to keep the computational effort low.) After having included a limited number of currently violated inequalities we restart the interior-point iterations. The current X is no longer feasible, so we backtrack on the line from X to I, until we are feasible again. (Note that the identity matrix I is the average of all cut vectors, so feasible for all relaxations.) Having found a new interior point, we start the interior-point iterations again. After each iteration, we check whether new triangle inequalities get violated by a significant amount. If so, we change the model again by including the $n/5$ most violated inequalities.

Iterating this way, we never get rid of any inequality that has been selected. To purge the current set of inequalities, we have to stop the phase of adding new inequalities, and simply iterate until the optimal solution to the current model is reached. Then the "small" dual variables indicate which inequalities can be dropped. We execute this phase of purging after 5 rounds of adding inequalities, so we have n inequalities in the first purging step. The computation time will increase drastically with the number of inequalities, so we stop after at most 4 purging steps. Thus we never consider more than about $4n$ inequalities at once.

The structure of the constraints deserves some comments. The equality constraints, $x_{ii} = 1$, are treated as rank-one constraints, i.e. tr $E_i X = 1$. Here $E_i = e_i e_i^t$, and e_i denotes column i of the identity matrix. Similarly, the inequality constraints tr $[(ff^t)X] \geq 1$ are also of rank-one. Moreover, both e_i and f are sparse, with no more than 3 nonzero entries. Exploiting this structure leads to significant computational simplifications; see [24] for details. Fujisawa et al. discuss in [19] how unstructured sparsity can be exploited to speed up computation. The fact that all our constraints are of rank-one allows simplifications, which are at least as good as the speed-up obtained in [19].

To generate pseudorandom instances, we use the graph generator RUDY from G. Rinaldi [57]. The following tables should give a rough idea about the qualitative behaviour of the semidefinite relaxations. We generated random graphs with edge probability 1/5, so there are roughly $n^2/10$ edges for a graph on n nodes. We consider two kinds of weight functions. In the first, the edge weights w_{ij} are randomly drawn from $\{-10, 9, \ldots, 9, 10\}$. In the second case, the weights are drawn from $\{1, 2, \ldots, 9, 10\}$. To measure the quality of the approximation, we compute bounds z^+ and z^- for the maximum and the minimum cut. Thus

$$z^+ = \max\{\text{tr } LX : \text{diag}(X) = e, f^t X f \geq 1, \ \forall f \in T, X \succeq 0\}.$$

As described above, T denotes a subset of all the triangle inequalities. Similarly,

$$z^- = \min\{\text{tr } LX : \text{diag}(X) = e, f^t X f \geq 1, \ \forall f \in T, X \succeq 0\}.$$

In addition, we generate cuts by a rounding procedure, applied to X. We denote the respective values of these cuts by c^+ for the cut with highest cost and c^- for the least cost cut. If $w_{ij} \geq 0$, then clearly $z^- = c^- = 0$, by taking the empty cut. In the presence of negative edge weights it is possible that $z^- < 0$ or $c^- < 0$. To measure the quality of the approximation, we look at the ratio (in %)

$$gap = 100 \frac{z^+ - z^-}{c^+ - c^-}.$$

n	no triangles gap	sec	n triangles gap	sec	$2n$ triangles gap	sec	$4n$ triangles gap	sec
100	15.2	9	9.6	39	6.9	88	4.7	320
150	19.4	31	13.7	180	11.8	390	10.0	1500
200	19.2	124	14.3	540	12.4	1140	10.9	4300

Table 1: Max-Cut on random graphs, edge density = 20 %, edge weights between -10 and 10.

Recall that $w_{ij} \geq 0$ implies by the results of Goemans and Williamson that $gap \leq 13.9\%$. Reading the following tables horizontally shows how the gap (given in %) improves as more and more triangles are included. However, the computation times grow even faster. The results indicate the potential, but also the limits of this approach: While problems of small size, say $n \approx 100$, can be handled routinely, the interior-point solution approach becomes prohibitive on larger problems, say $n \approx 200$. For really large scale problems, ($n \approx 1000$), one would have to use alternatives to interior-point methods; see for instance [27].

The computation times are seconds on a moderate size PC (Pentium P133 with 16M of main memory). Note that the gaps for positively weighted graphs are indeed much smaller than in the general case. Each line in the table represents one graph. It can be reproduced by the program RUDY with the call

```
rudy -rnd_graph 100 20 10 -random -10 10 11
rudy -rnd_graph 150 20 20 -random -10 10 21
```

and so on. There exist several computational studies dealing with Max-Cut approximations. It is beyond the scope of this survey to describe them all. The current state of the art can be summarized as follows:

- Small instances ($n \approx 100$) can be solved routinely to optimality (by using either polyhedral or semidefinite relaxations in combination with Branch and Bound techniques); see [26].
- Unstructured medium sized problems ($n \approx 200$) allow tight approximations; see the tables above, and also [24, 29].
- Larger problems ($n \approx 500$) are currently only manageable if some additional structure can be used, such as "near planarity" of the graph.

In [33] the same approach is applied to the k-equipartition of graphs. The qualitative behaviour of the results is very similar to the Max-Cut case. Again it turns out that small problems can be solved to optimality, while the approach becomes computationally prohibitive for graphs with about 200 nodes.

It is a challenging project to find algorithms for larger-sized semidefinite programs, that are computationally more efficient, perhaps at the price of being less robust than interior-point methods.

Acknowledgement: I thank two anonymous referees for many helpful suggestions to improve the presentation in the paper.

n	no triangles gap	n triangles gap	$2n$ triangles gap	$4n$ triangles gap
100	3.3	1.4	0.9	0.5
150	3.8	2.6	2.1	1.7
200	3.3	2.4	2.0	1.7

Table 2: Max-Cut on random graphs, edge density = 20 %, edge weights between 1 and 10.

References

[1] F. ALIZADEH. Interior Point Methods in Semidefinite Programming with Applications to Combinatorial Optimization. *SIAM Journal on Optimization* 5:13–51, 1995.

[2] F. ALIZADEH, J.A. HAEBERLY, and M.L. OVERTON Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results, Technical Report 721, New York University, 1996.

[3] D. AVIS and V.P. GRISHUKHIN. A bound on the k-gonality of the hypermetric cone and related complexity problems. *Computational Geometry: Theory and Applications* 2:241–254, 1993.

[4] N. ALON and N. KAHALE. A spectral technique for coloring random 3-colorable graphs, in: Proceedings 26th Symposium on the Theory of Computer Science, 1994, 346–353.

[5] N. ALON and N. KAHALE. Approximating the independence number via the theta function. Technical Report, Tel Aviv University, 1995.

[6] A. BEN-ISRAEL, A. CHARNES, and K.O. KORTANEK. Duality and asymptotic solvability over cones. Bulletin of the AMS 75, 318–324, 1969.

[7] A. BLUM, and D.KARGER. An $O(n^{0.2143})$-coloring algorithm for 3-colorable graphs. *Information Processing Letters*, 61:49–53, 1997.

[8] S. BOYD, L. El GHAOUI, E. FERON, and V. BALAKRISHNAN. Linear Matrix Inequalities in System and Control Theory. Studies in Applied Mathematics 15, SIAM 1994.

[9] J. CULLUM, W.E. DONATH, and P.WOLFE. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. *Mathematical Programming Study*, 3:35–55, 1975.

[10] D.M. CVETKOVIC, M. DOOB, and H. SACHS. Spectra of Graphs (3rd edition), Johann Ambrosius Barth, Heidelberg und Leipzig, 1995.

[11] Ch. DELORME and S. POLJAK. Laplacian Eigenvalues and the Max-Cut Problem. *Mathematical Programming*, 63:557–574, 1993.

[12] M. DEZA, M. GRISHUCHIN and M. LAURENT. The hypermetric cone is polyhedral, *Combinatorica*, 13:397–411, 1993

[13] W.E. DONATH and A.J. HOFFMAN. Lower bounds for the partitioning of graphs. *IBM J. of Research and Development*, 17:420–425, 1973.

[14] R.J. DUFFIN. Infinite Programs, in: *Linear inequalities and related systems*, (H.W. Kuhn and A.W. Tucker eds.), Annals of Mathematical Studies 38, Princeton University Press, 157–170, 1956.

[15] J. FALKNER, F. RENDL, and H. WOLKOWICZ. A computational study of graph partitioning. *Mathematical Programming*, 66:211–239, 1994.

[16] P.A. FILLMORE, and J.P. WILLIAMS. Some convexity theorems for matrices. *Glasgow Math. J.* 12:110–117, 1971.

[17] R. FLETCHER. Practical Methods of Optimization (2nd edition), Wiley, 1987.

[18] A. FRIEZE and M. JERRUM. Improved approximation algorithms for max k-cut and max bisection. *Algorithmica*, 18:67–81, 1997.

[19] K. FUJISAWA, M. KOSHIMA, and K. NAKATA. Exploiting sparsity in primal-dual interior-point methods for semidefinite programming. Technical Report B-324, Tokyo Institute of Technology, 1997.

[20] W. GANDER. Least squares with a quadratic constraint. *Numer. Math.*, 36:291–307, 1981.

[21] M.X. GOEMANS. Semidefinite Programming in Combinatorial Optimization, *Mathematical Programming* 79: 143–161, 1997.

[22] M.X. GOEMANS and D.P. WILLIAMSON. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal ACM*, 42: 1115–1145, 1995.

[23] J. HASTAD. Some optimal inapproximability results, in: Proceedings 29th Symposium on the Theory of Computer Science, 1997, 1–10.

[24] C. HELMBERG. An interior point method for semidefinite programming and max–cut bounds. PhD Thesis, Graz University of Technology, Oct. 1994.

[25] J.-B. HIRIART-URRUTY and C. LEMARÉCHAL. Convex analysis and Minimization algorithms 1, Grundlagen der mathematischen Wissenschaften 305, Springer Verlag, 1993.

[26] C. HELMBERG and F. RENDL. Solving Quadratic $(0,1)$-Problems by Semidefinite Programming and Cutting Planes. Working paper, TU Graz, 1995, to appear in: *Mathematical Programming.*

[27] C. HELMBERG and F. RENDL. A Spectral Bundle Method for Semidefinite Programming. Working paper, TU Graz, 1997, to appear in: *SIAM Journal on Optimization.*

[28] C. HELMBERG, S. POLJAK, F. RENDL, and H. WOLKOWICZ. Combining Semidefinite and Polyhedral Relaxations for Integer Programs. *Lecture Notes in Computer Science* 920: 124–134, 1995 (Proceedings of IPCO conference).

[29] C. HELMBERG, F. RENDL, R. J. VANDERBEI, and H. WOLKOWICZ. An interior point method for Semidefinite Programming. *SIAM Journal on Optimization* 6:342–361, 1996.

[30] F. JUHÁSZ. On the spectrum of a random graph, in: (L. Lovász and V.T. Sós, eds.), Algebraic Methods in Graph Theory, Colloq. Soc. János Bolyai 15: 313–316, 1982.

[31] D. KARGER, and R. MOTWANI, and M. SUDAN. Approximate Graph Coloring by Semidefinite Programming, in: 35th Symposium on Foundations of Computer Science, 1994, 2–13.

[32] S.E. KARISCH. Nonlinear approaches to Quadratic Assignment and Graph Partition Problems. PhD Thesis, Graz University of Technology, 1995.

[33] S.E. KARISCH and F. RENDL. Semidefinite Programming and Graph Equipartition, in: Topics in Semidefinite and Interior-Point Methods (P.M. Pardalos, H. Wolkowicz, eds.), Fields Institute Communications 18, American Mathematical Society, 1998, 77–95.

[34] H. KARLOFF. How good is the Goemans-Williamson MAX-CUT algorithm? in: Proceedings 28th Symposium on the Theory of Computer Science, 1996, 427–434.

[35] M. KOJIMA, S. SHINDOH, and S. HARA. Interior-point methods for the monotone linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization* 7:86–125, 1997.

[36] S. KRUK, M. MURAMATSU, F. RENDL, R.J. VANDERBEI, and H. WOLKOWICZ. The Gauss-Newton direction in semidefinite programming. Working paper, University of Waterloo, 1998.

[37] M. LAURENT, S. POLJAK, and F. RENDL. Connections between semidefinite relaxations of the max-cut and stable set problems. *Mathematical Programming* 77: 225–246, 1997.

[38] L. LOVÁSZ. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25:1–7, 1979.

[39] L. LOVÁSZ. Semidefinite programs and combinatorial optimization. Lecture Notes, 1995.

[40] Z.Q. LUO, J.F. STURM and S. ZHANG. Duality and self-duality for conic convex programming. Technical Report, Erasmus University Rotterdam, 1996.

[41] B. MOHAR. Some applications of Laplace eigenvalues of graphs, in: Graph Symmetry (G. Hahn, G. Sabidussi, eds.), Kluwer, 1997, 225–275.

[42] B. MOHAR and S. POLJAK. Eigenvalues in combinatorial optimization, in: Combinatorial and Graph-Theoretical Problems in Linear Algebra (R.A. Brualdi, S. Friedland, V. Klee, eds.), IMA Vol. Math. Appl. 50, Springer, 1993, 107–151.

[43] R.D.C. MONTEIRO. Primal-dual path following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7:663–678, 1997.

[44] Y.E. NESTEROV and A.S. NEMIROVSKII. Interior Point Polynomial Algorithms in Convex Programming. Studies in Applied Mathematics 13, SIAM 1994.

[45] Y.E. NESTEROV and M.J. TODD. Primal-dual interior-point methods for self-scaled cones. Technical Report TR-1125, Cornell University, 1995.

[46] M.L. OVERTON. Large-scale Optimization of Eigenvalues. *SIAM Journal on Optimization*, 2:88–120, 1992.

[47] M.L. OVERTON and R.S. WOMMESLEY. On the sum of the largest eigenvalues of a symmetric matrix, *SIAM J. Matrix Anal. Appl.*, 13:41–45, 1992.

[48] S. POLJAK. Duality over nonpolyhedral cones. Manuscript, Passau 1994.

[49] S. POLJAK and F. RENDL. Solving the Max-Cut Problem using Eigenvalues. *Discrete Applied Mathematics*, 62: 249–278, 1995.

[50] S. POLJAK, and F. RENDL Node and edge relaxations of the max-cut problem. *Computing*, 52:123–137, 1994.

[51] S. POLJAK, and F. RENDL Nonpolyhedral relaxations of graph bisection problems. *SIAM Journal on Optimization*, 5:467–487, 1995.

[52] S. POLJAK, and S. TUZA Maximum Cuts and large bipartite subgraphs. *DIMACS Series in Discrete Mathematics*, 20:181–244, 1995.

[53] S. POLJAK, and H. WOLKOWICZ. Convex relaxations of (0,1) quadratic programming. *Mathematics of Operations Research*, 20:550–561, 1995.

[54] S. POLJAK, F. RENDL and H. WOLKOWICZ. A recipe for semidefinite relaxations of (0,1) quadratic programming. *Journal of Global Optimization*, 7:51–73, 1995.

[55] F. RENDL and H. WOLKOWICZ. A projection technique for partitioning the nodes of a graph. *Annals of Operations Research* 58: 155–179, 1995.

[56] F. RENDL and H. WOLKOWICZ. A semidefinite framework for trust region subproblems with application to large scale minimization. *Mathematical Programming*, 77:273–299, 1997.

[57] G. RINALDI. RUDY: a generator for random graphs. IASI Rom, 1996.

[58] R. STERN and H. WOLKOWICZ. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations, *SIAM J. Optimization* 5: 286–313, 1995.

[59] M.J. TODD, K.C. TOH, and R.H.TÜTÜNCÜ. On the Nesterov-Todd direction in semidefinite programming, Technical Report, Cornell University, 1996.

[60] M. UJVARI. Simple proofs of dualtiy theorems in semidefinite and cone-linear programming. Eötvös University Budapest, 1996.

[61] L. VANDENBERGHE and S. BOYD. Positive definite programming. *SIAM Review* 38: 49–95, 1996.

[62] H. WOLKOWICZ. Some applications of optimization in matrix theory. *Linear Algebra and ist Applications* 40: 101–118, 1981.

[63] Y. ZHANG. On extending primal-dual interior-point algorithms from linear programming to semidefinite programming. Technical Report TR95-20, University of Maryland, 1995.

[64] Q. ZHAO. Semidefinite Programming for Assignment and Partition Problems. PhD thesis, University of Waterloo, 1996.