

Semidefinite Relaxations for Integer Programming

Franz Rendl

Abstract

We survey some recent developments in the area of semidefinite optimization applied to integer programming. After recalling some generic modeling techniques to obtain semidefinite relaxations for NP-hard problems, we look at the theoretical power of semidefinite optimization in the context of the Max-Cut and the Coloring Problem. In the second part, we consider algorithmic questions related to semidefinite optimization, and point to some recent ideas to handle large scale problems. The survey is concluded with some more advanced modeling techniques, based on matrix relaxations leading to copositive matrices.

1 Introduction

Looking back at fifty years of integer programming, there is wide agreement that *Polyhedral Combinatorics* is a major ingredient to approach NP-hard integer optimization problems. Having at least a partial description of the convex hull of all feasible solutions of an integer program can be exploited by the strong algorithmic machinery available to solve linear programming problems, notably the Simplex method. First systematic attempts to use polyhedral techniques to solve 0-1 integer optimization go back to [DFJ54].

We consider an abstract combinatorial optimization problem (COP) given as follows. Let E be a finite set and let \mathcal{F} be a (finite) family of subsets of E . The family \mathcal{F} denotes the set of feasible solutions of (COP). Each $e \in E$ has a given integer cost c_e . We define the cost $c(F)$ of $F \in \mathcal{F}$ to be

Franz Rendl
Alpen-Adria Universität Klagenfurt, Austria, e-mail: franz.rendl@uni-klu.ac.at
this version June 17, 2009

$$c(F) := \sum_{e \in F} c_e.$$

The problem (COP) now consists in finding a feasible solution F of minimum cost:

$$(COP) \quad z^* = \min\{c(F) : F \in \mathcal{F}\}.$$

The traveling salesman problem (TSP) for instance could be modeled with E being the edge set of the underlying graph G . An edge set F is in \mathcal{F} exactly if it is the edge set of a Hamiltonian cycle in G .

By assigning to each $F \in \mathcal{F}$ a characteristic vector $x_F \in \{0, 1\}^n$ with $(x_F)_e = 1$ if and only if $e \in F$, we can write (COP) as a linear program as follows. Let $P := \text{conv}\{x_F : F \in \mathcal{F}\}$ denote the convex hull of the incidence vectors of feasible solutions. Then it is clear that

$$z^* = \min\{c^T x_F : F \in \mathcal{F}\} = \min\{c^T x : x \in P\}.$$

This is the basic principle underlying the polyhedral approach to solve combinatorial optimization problems.

As an example, consider \mathcal{F} to be the set of all permutation matrices. Birkhoff's theorem states that the convex hull of permutation matrices is the polyhedron of doubly stochastic matrices:

$$\text{conv}\{X : X \in \Pi\} = \Omega.$$

For notation, see the end of this section. Hence the combinatorial problem of finding a permutation ϕ minimizing $\sum_i c_{i,\phi(i)}$ can be solved through the linear program

$$\min\{\langle C, X \rangle : X \in \Omega\}.$$

The practical difficulty lies in the fact that in general the polyhedron P is not easily available. The use of a computationally tractable partial description of P in combination with systematic enumeration, like Branch and Bound, has led to quite successful solution methods for a variety of combinatorial optimization problems like the TSP, see for instance [LLRKS85]. It turned out however, that for some prominent NP-hard problems like Stable-Set or Max-Cut, this polyhedral approach was not as successful as one might have hoped in view of the results for TSP, see for instance [BJR89].

This motivated the study of more powerful approximation techniques for (COP). One such possibility consists in studying matrix liftings of the form

$$\mathcal{M} := \text{conv}\{x_F x_F^T : F \in \mathcal{F}\}, \tag{1}$$

see for instance [LS91, SA90, SA94]. Any relaxation based on \mathcal{M} lies in the space \mathcal{S}_n of symmetric matrices, rather than \mathbb{R}^n , the 'natural' space of (COP). The modeling power of using \mathcal{M} comes from the fact that any quadratic constraint on $x \in P$,

such as $x_i x_j = 0$, translates into a linear constraint on $X \in \mathcal{M}$, such as $x_{ij} = 0$. Moreover, any $X \in \mathcal{M}$ is positive semidefinite.

Polyhedral implications of working with \mathcal{M} instead of P are investigated for instance in [LS91, BCC93, SA90] under the key words 'lift and project'. Using the condition that \mathcal{M} is contained in the cone

$$\mathcal{S}_n^+ := \{X \in \mathcal{S}_n : X \succeq 0\}$$

of positive semidefinite matrices leads to semidefinite relaxations which are a generalization of linear programs. Formally, a semidefinite program, SDP for short, is defined by the data $C, A_1, \dots, A_m \in \mathcal{S}_n$ and $b \in \mathbb{R}^m$ and consists of the following

$$(SDP) \quad z_p = \inf\{\langle C, X \rangle : \langle A_i, X \rangle = b_i, i = 1, \dots, m, X \succeq 0\}. \quad (2)$$

In this chapter we will consider relaxations of integer programs which are based on SDP, rather than purely polyhedral combinatorics. We first recall some basics about SDP in Section 2. In Section 3 various modeling ideas are described which all lead to SDP relaxations. SDP as generalization of LP often provides tighter approximations, at increased computational cost. The theoretical power of SDP to approximate some NP-hard optimization problems is presented in Section 4. The hyperplane rounding idea of Goemans and Williamson [GW95] turns out to be a generic method to generate provably good feasible solutions for a variety of problems. In Section 5 we turn to algorithms for solving SDP. While interior-point based algorithms are still the method of choice, it turns out that large scale problems are beyond the scope of these algorithms and alternatives, capable of handling an arbitrary number of constraints at least approximately, are necessary. Several of these are discussed in 5. Finally, we touch some more recent modeling techniques which go beyond SDP in Section 6.

This article lines up with several survey papers devoted to the connection between semidefinite optimization and integer programming. The interested reader is referred to [LR05] for an extensive summary on the topic covering the development until 2003. The surveys by Lovasz [Lov03], Goemans [GOE97] and Helmberg [Hel02] all focus on the same topic, but also reflect the scientific interests and preferences of the respective authors. The present paper is no exception to this principle. The material selected, and also omitted, reflects the author's subjective view on the subject. Since the scientific area covered here is still under rapid development, the present survey is far from complete. Some discussion and pointers to material not covered will be given at the end of each chapter.

Notation: We use the following notation. The vector of all ones is denoted e , and $J = ee^T$ is the all-ones matrix. The identity matrix is denoted by $I = (e_1, \dots, e_n)$. Thus the e_i represent the standard unit vectors. For $i \neq j$ we define $E_{ij} = e_i e_j^T + e_j e_i^T$.

The set of symmetric matrices is denoted by \mathcal{S} . \mathcal{S}^+ denotes the closed convex cone of semidefinite matrices. Its interior is the set of definite matrices, denoted \mathcal{S}^{++} . We also use the cone \mathcal{C} of *copositive matrices*, given by $\mathcal{C} = \{X :$

$v^T X v \geq 0 \forall$ vectors $v \geq 0$ }. Its dual cone \mathcal{C}^* consists of the set of *completely positive matrices*, $\mathcal{C}^* = \{X : X = VV^T, V \geq 0 \text{ is } n \times k\}$. The standard simplex is $\Delta = \{x : x \geq 0, e^T x = 1\}$. The convex hull of a set S is denoted by $\text{conv}(S)$. We denote by Π the set of permutation matrices and by Ω the set of doubly stochastic matrices, $\Omega = \{X : X e = X^T e = e, X \geq 0\}$. For $a, b \in \mathbb{R}^n$, $\langle a, b \rangle_+$ denotes maximal scalar product of a and b , if the entries in a and b can be permuted arbitrarily. It is given for instance by sorting the entries in both a and b in nondecreasing order. The minimal scalar product $\langle a, b \rangle_-$ is defined analogously.

If A and B are matrices of the same order, then the *Hadamard* or elementwise product is $C = A \circ B$ with $c_{ij} = a_{ij} b_{ij}$. The Kronecker product (or tensor product), $A \otimes B$, of two matrices A and B consists of the matrix of all pairwise products of elements from A and B . Formally, if $A = (a_{ij})$ is $m \times n$ and B is $p \times q$, then

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

The operator $\text{Diag} : \mathbb{R}^n \mapsto \mathcal{S}_n$ maps a vector y to the diagonal matrix $\text{Diag}(y)$, its adjoint mapping $\text{diag}(X)$ extracts the main diagonal from the matrix X .

If G is a graph, and $S \subseteq V(G)$ is a subset of its vertices, we denote by $\delta(S)$ the set of all edges ij such that $i \in S, j \notin S$. The neighbourhood $N(i)$ of vertex $i \in V(G)$ is the set of all vertices, adjacent to i , $N(i) = \{j : ij \in E(G)\}$. The complement graph \overline{G} of a graph G has edges $ij \in E(\overline{G})$ whenever $i \neq j$ and $ij \notin E(G)$.

2 Basics on Semidefinite Optimization

Problem (2) is a convex optimization problem, because a linear function is optimized over the convex set

$$F_p := \{X : A(X) = b, X \succeq 0\}.$$

The linear operator $A(X)$ maps matrices into \mathbb{R}^m and has $A(X)_i = \langle A_i, X \rangle$. Its adjoint A^T , defined through the adjoint identity

$$y^T A(X) = \langle A^T(y), X \rangle$$

is given by $A^T(y) = \sum_i y_i A_i$. The problem (SDP) as a convex problem possesses a dual, which is most conveniently derived through the Lagrangian

$$L(X, y) = \langle C, X \rangle + y^T (b - A(X)) = b^T y + \langle C - A^T(y), X \rangle \quad (3)$$

and the Minimax inequality

$$\inf_{u \in U} \sup_{v \in V} f(u, v) \geq \sup_{v \in V} \inf_{u \in U} f(u, v),$$

which holds for any function $f : U \times V \mapsto \mathbb{R}$. To get the dual, we first observe

$$\sup_y L(X, y) = \begin{cases} \langle C, X \rangle & \text{if } A(X) = b \\ +\infty & \text{otherwise.} \end{cases}$$

To see what happens to $\inf_{X \succeq 0} L(X, y)$ we recall Fejer's theorem.

Theorem 1. *The matrix $M \in \mathcal{S}^+$ if and only if $\langle M, X \rangle \geq 0 \forall X \in \mathcal{S}^+$.*

In the language of convex analysis, this translates into the fact that the cone dual to the semidefinite matrices is again the cone of semidefinite matrices. Recall that the dual cone K^* of the cone $K \subseteq \mathbb{R}^d$ is by definition

$$K^* := \{y \in \mathbb{R}^d : \langle y, x \rangle \geq 0 \forall x \in K\}.$$

Now if $C - A^T(y) \succeq 0$, then by Fejer's theorem $\inf_{X \succeq 0} \langle C - A^T(y), X \rangle = 0$. On the other hand, if $C - A^T(y) \notin \mathcal{S}^+$, then there must exist some $X \succeq 0$ such that $\langle C - A^T(y), X \rangle < 0$. We conclude

$$\inf_{X \succeq 0} L(X, y) = \begin{cases} b^T y & \text{if } C - A^T(y) \succeq 0 \\ -\infty & \text{otherwise.} \end{cases} \quad (4)$$

Therefore (2) is equivalent to $z_p = \inf_{X \succeq 0} \sup_y L(X, y)$ and the Minimax inequality implies

$$z_p \geq \sup_y \inf_{X \succeq 0} L(X, y) =: z_d.$$

The problem defining z_d can be rewritten using (4) to yield

$$z_d = \sup \{b^T y \text{ such that } C - A^T(y) \succeq 0\}. \quad (5)$$

The last problem is again a semidefinite program, which is usually called the dual of (2). In contrast to linear programming, strong duality ($z_p = z_d$) does not hold in general. Moreover, attainment of sup and inf only holds under some additional conditions. The following condition, often called Slater's constraint qualification, insures $z_p = z_d$. Problem (2) satisfies the *Slater constraint qualification* if there exists a positive definite matrix $X \succ 0$ such that $A(X) = b$. Such matrices are often called *strictly feasible*.

Theorem 2. *(see for instance Duffin [Duf56]) If (2) satisfies the Slater condition and z_p is finite, then the dual problem is feasible, the dual supremum is attained, and $z_p = z_d$.*

In most of the semidefinite relaxations considered in this chapter, both the primal and the dual problem satisfy the Slater condition, therefore we have the following situation. Suppose

$$\exists X \succ 0 \text{ such that } A(X) = b, \text{ and } \exists y \text{ with } C - A^T(y) \succ 0. \quad (6)$$

Then (X, y, Z) is optimal for (2) and (5) if and only if

$$A(X) = b, X \succeq 0, C - A^T(y) = Z \succeq 0, \langle Z, X \rangle = 0. \quad (7)$$

Note that in the dual problem, we have introduced for notational and algorithmic convenience the slack matrix Z in the formulation. The condition $\langle Z, X \rangle = 0$ is a consequence of $0 = z_p - z_d = \langle C, X \rangle - b^T y = \langle C - A^T(y), X \rangle$.

3 Modeling with Semidefinite Programs

The basic idea to come to semidefinite relaxations of (COP), and more generally integer programs, consists in working with \mathcal{M} from (1), instead of P . Going from \mathbb{R}^n to the space of symmetric matrices \mathcal{S}_n allows to replace quadratic constraints and cost functions by linear ones. As a first example we consider a semidefinite relaxation of quadratic 0,1 optimization.

3.1 Quadratic 0,1 optimization

For given $Q \in \mathcal{S}_n$ we consider

$$(QP) \quad \min_{x \in \{0,1\}^n} x^T Q x. \quad (8)$$

This problem is equivalent to Max-Cut, see for instance [Ham65, BJR89], and therefore NP-hard. We may assume without loss of generality that there is no additional linear term $c^T x$ in the objective function (8), because $x_i^2 = x_i$ allows us to add c to the main diagonal of Q . The cost function can be rewritten as

$$x^T Q x = \langle Q, x x^T \rangle.$$

Following the linearization idea we introduce a matrix variable X taking the role of $x x^T$. Since $x_i^2 = x_i$, the main diagonal of X is equal to x . The nonconvex constraint $X - x x^T = 0$ is relaxed to $X - x x^T \succeq 0$. The Schur-complement lemma shows that this set is convex.

Lemma 1. *Let $M = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ be symmetric and A invertible. Then $M \succeq 0$ if and only if $A \succeq 0$ and $C - B^T A^{-1} B \succeq 0$.*

Proof. This well known fact follows from the similarity transformation

$$\begin{pmatrix} I & 0 \\ -B^T A^{-1} & I \end{pmatrix} \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} I & 0 \\ -B^T A^{-1} & I \end{pmatrix}^T = \begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{pmatrix}.$$

□

Hence we get the following SDP relaxation of (QP):

$$\min \langle Q, X \rangle \text{ such that } X - xx^T \succeq 0, \text{diag}(X) = x. \quad (9)$$

We use the Schur-complement lemma to replace the quadratic condition $X - xx^T \succeq 0$ by $\begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0$.

3.2 Max-Cut and graph bisection

Let A be the (weighted) adjacency matrix of a graph G on n vertices. A basic graph optimization problem consists in separating the vertex set V of G into two sets $S, V \setminus S$ such that

$$\sum_{ij \in \delta(S)} a_{ij}$$

is optimized. In case of *Max-Cut* one has

$$(MC) \quad z_{mc} = \max_{S \subseteq V} \sum_{ij \in \delta(S)} a_{ij}. \quad (10)$$

The *bisection problem* has an additional input parameter s , specifying $|S| = s$, and is defined as follows.

$$(BS) \quad z_{bs} = \min_{S \subseteq V, |S|=s} \sum_{ij \in \delta(S)} a_{ij} \quad (11)$$

The special case $s = \frac{n}{2}$ is often called Equicut. Kernighan and Lin [KL70] investigated various local improvement heuristics for this problem, see also [JAMS89]. All these bisection problems are well known to be NP-hard. A simple integer program for these problems can be obtained as follows. Subsets $S \subseteq V$ are modeled by vectors $x \in \{-1, 1\}^n$ with $x_i = 1$ exactly if $i \in S$. Then $e = ij \in \delta(S)$ precisely if $x_i x_j = -1$. A convenient way to model the objective function makes use of the *Laplacian*, associated to A , which is defined as follows.

$$L = L_A = \text{Diag}(Ae) - A \quad (12)$$

A simple calculation shows that

$$\sum_{ij \in \delta(S)} a_{ij} = \sum_{ij \in E} a_{ij} \frac{1 - x_i x_j}{2} = \frac{1}{4} x^T L x, \quad (13)$$

if $x \in \{-1, 1\}^n$ represents S . Hence we have

$$z_{mc} = \max_{x \in \{-1, 1\}^n} \frac{1}{4} x^T L x,$$

$$z_{bs} = \min_{x \in \{-1,1\}^n, e^T x = n-2s} \frac{1}{4} x^T L x.$$

The linearization idea gives again $x^T L x = \langle L, x x^T \rangle$, and we introduce $X = x x^T$ and now have $\text{diag}(X) = e$. The cardinality constraint is easily incorporated using

$$|e^T x| = n - 2s \text{ if and only if } \langle J, x x^T \rangle = (n - 2s)^2.$$

Hence we get the SDP relaxation of Max-Cut as

$$z_{mc,sdp} = \max \frac{1}{4} \langle L, X \rangle \text{ such that } \text{diag}(X) = e, X \succeq 0, \quad (14)$$

and for bisection as

$$z_{bs,sdp} = \min \frac{1}{4} \langle L, X \rangle \text{ such that } \text{diag}(X) = e, \langle J, X \rangle = (n - 2s)^2, X \succeq 0.$$

Quadratic 0,1 optimization can linearly (and bijectively) be mapped to optimization in -1,1 variables, showing the claimed equivalence between Max-Cut and quadratic 0,1 optimization. One may therefore wonder whether there is also some relation between the two relaxations (9) and (14). It is not too difficult to verify that these relaxations are in fact also equivalent, see for instance [HPRW95, LPR97]. Helmberg [Hel00] provides an explicit transformation between these problems, which preserves structure, like sparsity. The equivalence of Max-Cut and quadratic 0,1 optimization was already pointed out by Hammer, see [Ham65].

3.3 Stable sets, cliques and the Lovasz theta function

Perhaps the earliest use of SDP to get relaxations of NP-hard problems goes back to a fundamental paper by Lovasz in 1979, see [Lov79]. Consider the stable set problem in an unweighted graph G . We recall that a set $S \subseteq V(G)$ is stable if the vertices in S are pairwise nonadjacent. Consequently, S forms a complete subgraph in the complement graph \overline{G} of G . The stable set problem asks to find a stable set of maximum cardinality in G . The cardinality of a largest stable set in G , denoted $\alpha(G)$, is called the *stability number* of G . Modeling stable sets by their characteristic vectors, we get

$$(STAB) \quad \alpha(G) = \max \{ e^T x : x_i x_j = 0 \ \forall ij \in E(G), x \in \{0,1\}^n \}. \quad (15)$$

Following our linearization idea for 0,1 optimization, we obtain the following SDP relaxation, which was studied in some detail in [LS91].

$$\alpha(G) \leq \vartheta_1(G) = \max \{ e^T x : X - x x^T \succeq 0, \text{diag}(X) = x, x_{ij} = 0 \ \forall ij \in E \}. \quad (16)$$

Note in particular that the quadratic equations in x are now linear in X . This model can in fact be simplified by eliminating x . Suppose x is the characteristic vector of a (nonempty) stable set. Then

$$X = \frac{1}{x^T x} x x^T$$

satisfies the following conditions:

$$X \succeq 0, \operatorname{tr}(X) = 1, x_{ij} = 0 \forall ij \in E, \operatorname{rank}(X) = 1.$$

Moreover

$$\langle J, X \rangle = \frac{1}{x^T x} (e^T x)^2 = e^T x,$$

because $e^T x = x^T x$. The following result is well known.

Lemma 2.

$$\alpha(G) = \max\{\langle J, X \rangle : X \succeq 0, \operatorname{tr}(X) = 1, x_{ij} = 0 \forall ij \in E, \operatorname{rank}(X) = 1\}. \quad (17)$$

Proof. Feasibility of X implies $X = aa^T$ where the vector a is nonzero only on some stable set S , because $x_{ij} = a_i a_j = 0 \forall ij \in E$. Looking at the nonzero submatrix $a_S a_S^T$ indexed by S , the cost function becomes $(e^T a_S)^2$, which is maximal exactly if a_S is parallel to the all ones vector e . Hence a is (a multiple of) the characteristic vector of S and the maximization forces S to be a maximum stable set. \square

Leaving out the (nonconvex) rank condition, we obtain another SDP relaxation for $\alpha(G)$:

$$\alpha(G) \leq \vartheta_2(G) = \max\{\langle J, X \rangle : \operatorname{tr}(X) = 1, x_{ij} = 0 \forall ij \in E, X \succeq 0\}.$$

This is in fact the relaxation proposed by Lovasz in [Lov79]. Lovasz and Schrijver [LS91] show that $\vartheta_1(G) = \vartheta_2(G)$ and this function is usually called the *Lovasz theta function* $\vartheta(G)$.

Let us introduce the linear operator $A_G : \mathcal{S}_n \mapsto \mathbb{R}^E$ associated to the edge set E of G ,

$$A_G(X)_{ij} = \langle X, E_{ij} \rangle = x_{ij} + x_{ji} = 2x_{ij} \forall ij \in E$$

and its adjoint

$$A_G^T(y) = \sum_{ij \in E} y_{ij} E_{ij}.$$

Using these operators, we get the following primal-dual pair of semidefinite programs for $\vartheta(G)$.

$$\begin{aligned} \vartheta(G) &= \max\{\langle J, X \rangle : \operatorname{tr}(X) = 1, A_G(X) = 0, X \succeq 0\} \\ &= \min\{t : tI + A_G^T(y) \succeq J\}. \end{aligned} \quad (18)$$

Strong duality is justified by the observation that $\frac{1}{n}I$ is strictly feasible for the maximization, and setting $y = 0$ and $t = n + 1$ gives a strictly feasible point for the min-

imization problem. We will now use the adjoint A^T to express symmetric matrices in the following way: Let $Z \in \mathcal{S}_n$. Then

$$Z = \text{Diag}(y) + A_G^T(u) + A_{\overline{G}}^T(v),$$

if we define

$$y_i = z_{ii}, u_{ij} = z_{ij} \forall ij \in E(G), v_{ij} = z_{ij} \forall ij \in E(\overline{G}),$$

Suppose that X is feasible for (18). Then

$$X = \text{Diag}(x) + A_{\overline{G}}^T(\xi),$$

if we set $x_i = x_{ii}$ and $\xi_{ij} = x_{ij} \forall ij \in E(\overline{G})$. By construction $x_{ij} = 0 \forall ij \in E(G)$. We substitute for X and obtain

$$\vartheta(G) = \max\{e^T x + 2e^T \xi : e^T x = 1, \text{Diag}(x) + A_{\overline{G}}^T(\xi) \succeq 0\}. \quad (19)$$

Both models are mathematically equivalent, but from a computational point of view (18) is preferable in case $|E|$ is small. If $|E|$ is large (G is dense), then the second formulation (19) is more efficient, because in this case the number of variables $|E(\overline{G})|$ is small. Computational experience with both models is given in [DR07]. It turns out that $\vartheta(G)$ can be computed for graphs with $n \leq 200$ in acceptable time using interior-point methods and the sparse or the dense model depending on $|E(G)|$. From [DR07] it is also clear that interior-point methods can not handle graphs with $n \geq 200$ and $|E(G)| \approx \frac{n^2}{4}$, which is the worst case in terms of computational effort for both models, see also Table 5.1 below.

3.4 Chromatic number

A k -coloring of a graph G is a k -partition (V_1, \dots, V_k) of the vertex set $V(G)$ such that each V_i is a stable set in G . The *chromatic number* $\chi(G)$ is the smallest value k , such that G has a k -coloring. Let us encode k -partitions of $V(G)$ by the characteristic vectors s_i for V_i , thus

$$s_i \in \{0, 1\}^n \text{ and } (s_i)_u = 1 \text{ if } u \in V_i \text{ and } 0 \text{ otherwise.} \quad (20)$$

The partition property implies that $s_i \neq 0$ and that $\sum_i s_i = e$. Let $n_i := |V_i|$ denote the cardinality of V_i . We call the matrix

$$M = \sum_i s_i s_i^T \quad (21)$$

the *partition matrix* associated to the k -partition (V_1, \dots, V_k) . Note that $\text{rank}(M) = k$. It is clear from the definition that for any k -partition matrix M there exists a

permutation matrix $P \in \Pi$ such that

$$M = P \begin{pmatrix} J_{n_1} & 0 & \dots & 0 \\ 0 & J_{n_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & J_{n_k} \end{pmatrix} P^T.$$

Partition matrices have the following characterization.

Lemma 3. *Let M be a symmetric $0,1$ matrix. Then M is a k -partition matrix if and only if*

$$\text{diag}(M) = e, \text{rank}(M) = k, M \succeq 0. \quad (22)$$

Proof. Suppose that M is a k -partition matrix. Then this matrix obviously satisfies (22). Conversely let M satisfy (22). We need to show that M is (after appropriate renumbering) a direct sum of all ones blocks. Thus we need to show that $m_{ij} = m_{ik} = 1$ implies $m_{jk} = 1$. If this is not the case then the submatrix of M , indexed by i, j, k is

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

which has negative determinant, contradicting $M \succeq 0$. Therefore M is indeed a direct sum of all ones blocks. \square

This leads to the following somewhat nonstandard way to define $\chi(G)$.

$$\chi(G) = \min\{\text{rank}(M) : m_{ij} \in \{0, 1\}, \text{diag}(M) = e, m_{ij} = 0 \forall ij \in E, M \succeq 0\}.$$

It turns out that in fact we can get rid of both the rank and the semidefiniteness condition on M by introducing another semidefiniteness constraint.

Lemma 4. *Let M be a symmetric $0,1$ matrix. Then M is a k -partition matrix if and only if*

$$\text{diag}(M) = e, (tM - J \succeq 0 \iff t \geq k). \quad (23)$$

This result is implicitly proved in many sources, see for instance [Lov79]. A simple argument can be found also in [DR07]. Here we only take a closer look at how the semidefiniteness condition comes about. Looking at principal submatrices of $tM - J$, we first observe that any minor with two rows from the same partition block is singular, because it contains two identical rows. Hence nonsingular principal submatrices can have at most one row from each partition block. Therefore these must be of the form

$$tI_s - J_s$$

where $s \leq k$ denotes the order of the submatrix. Clearly, this matrix is semidefinite if and only if $t \geq s$, and since s could be as large as k we get the desired condition.

Using this result we can express the chromatic number as the optimal solution of the following SDP in binary variables.

$$\chi(G) = \min\{t : tM - J \succeq 0, \text{diag}(M) = e, m_{ij} = 0 \forall ij \in E, m_{ij} \in \{0, 1\}\}. \quad (24)$$

To get a tractable relaxation, we leave out the 0,1 condition and parametrize the matrix tM as $tM = tI + A_G^T(y)$. We get the following SDP as lower bound on $\chi(G)$:

$$\chi(G) \geq \vartheta(\overline{G}) = \min\{t : tI + A_G^T(y) \succeq J\}. \quad (25)$$

Comparing with the minimization problem in (18), we see that the above SDP is indeed the dual applied to the complement graph. Hence we have shown the sandwich theorem of Lovasz [Lov79].

Theorem 3.

$$\alpha(\overline{G}) \leq \vartheta(\overline{G}) \leq \chi(G).$$

The notation for ϑ is not uniform in the literature. If one starts from the clique number of a graph G , it seems more natural to denote by $\vartheta(G)$ what we denote by $\vartheta(\overline{G})$.

The problem (25) can be rephrased in the following way: Minimize z_{11} where the matrix $Z = (z_{ij}) \succeq 0, z_{ii} = z_{11} \forall i, z_{ij} = -1 \forall ij \in E(G)$. This follows simply from $Z = tI + A_G^T(y) - J$. This problem in turn can be rewritten as

$$\min\{\lambda : V \succeq 0, v_{ii} = 1 \forall i, v_{ij} = \lambda \forall ij \in E(G)\}. \quad (26)$$

The optimal value ϑ in (25) gives $\lambda = -\frac{1}{\vartheta-1}$. This last model will be used later on to analyze algorithms for graph coloring, see [KMS98].

3.5 General graph partition

The concept of partition matrices from (23) in the context of coloring can also be recycled to model the following general partition problems. Given the weighted adjacency matrix A of a graph G on n nodes and an integer $2 \leq k \leq n$, the general k -partition problem consists in finding a k -partition (S_1, \dots, S_k) of $V(G)$ with $|S_i| \geq 1$ such that the total weight of edges joining different partition blocks is optimized. In case of maximization, this problem also runs under the name of *Max-k-Cut*. Max-Cut, see the previous sections, corresponds to the special case $k = 2$. We represent partitions again through the respective characteristic vectors s_i , see (20), which we collect in the matrix $S = (s_1, \dots, s_k)$. The characteristic vector $s_i \in \{0, 1\}^n$ for the partition block S_i allows us to express the weight of the edges in $\delta(S_i)$ using the Laplacian L , see (12) as follows,

$$\sum_{rs \in \delta(S_i)} a_{rs} = s_i^T L s_i.$$

Therefore

$$\frac{1}{2}\langle S, LS \rangle = \frac{1}{2} \sum_i s_i^T L s_i$$

gives the total weight of all edges joining different partition blocks. We obtain a semidefinite relaxation of Max-k-Cut using (23), once $\langle S, LS \rangle = \langle L, SS^T \rangle$ is used to replace SS^T by a new matrix variable. For notational convenience we introduce the matrix variable $Y = \frac{1}{k-1}(kSS^T - J)$. (Note that SS^T takes the role of M in (23) and that $Le = 0$ implies $\langle L, J \rangle = 0$.)

$$\max\left\{\frac{k-1}{2k}\langle L, Y \rangle : \text{diag}(Y) = e, Y \succeq 0, y_{ij} \geq -\frac{1}{k-1} \forall i \neq j\right\}. \quad (27)$$

The sign constraints follow from $SS^T \succeq 0$. This model has been investigated theoretically by Frieze and Jerrum, see [FJ97] and also by DeKlerk et al [dKPW04].

A more restrictive version of k -partition, called k -Equicut, asks for k -partitions with $|S_i| = \frac{n}{k} \forall i$. In this case the objective is to minimize the total weight of edges cut. A feasible partition (S_1, \dots, S_k) for this problem implies the following additional constraint

$$S^T e = \frac{n}{k} e$$

which together with $Se = e$ (note the varying dimension of e) gives

$$SS^T e = \frac{n}{k} e,$$

hence $\frac{n}{k}$ is eigenvalue of SS^T with eigenvector e , for any feasible partition matrix S . Setting $X = SS^T$, this gives the following relaxation of k -Equicut.

$$\min\left\{\frac{1}{2}\langle L, X \rangle : \text{diag}(X) = e, Xe = \frac{n}{k}e, X \succeq 0, X \geq 0\right\}. \quad (28)$$

We note that this model does not contain the condition $kX \succeq J \succeq 0$, but only the weaker $X \succeq 0$. The following simple observation explains why.

Lemma 5. *Let $A \in \mathcal{S}_n^+$ and suppose $Av = \lambda v$ holds with $\|v\| = 1$. Then*

$$A - \lambda vv^T \succeq 0.$$

Proof. The spectral decomposition theorem for A yields $A = \lambda vv^T + \sum_i \lambda_i v_i v_i^T$ where we identified one of the eigenvalues of A as λ . $A \succeq 0$ implies $\lambda_i \geq 0$ hence $A - \lambda vv^T = \sum_i \lambda_i v_i v_i^T \succeq 0$. \square

This justifies the semidefiniteness constraint $X \succeq 0$ in (28). It is also interesting to note that the sign constraint $X \geq 0$ together with the eigenvector condition $Xe = \frac{n}{k}e$ implies that

$$\frac{n}{k}I - X \succeq 0$$

because in this case $\|X\|_1 = \frac{n}{k}$. Therefore, any eigenvalue λ of X satisfies $|\lambda| \leq \frac{n}{k}$. In summary, the constraints in (28) imply

$$\frac{n}{k}I \succeq X \succeq \frac{1}{k}J.$$

It should be observed that both SDP relaxations for partitioning are formulated in the space of symmetric $n \times n$ matrices, but the models include roughly $\frac{n^2}{2}$ sign constraints, which we will later see to be a computational challenge. Further details in this direction can be found for instance in [KR98] and [LR02].

We close this section with a bound on general k -partition, proposed by Donath and Hoffman in 1973, see [DH73]. In view of the results from the subsequent Section 3.7 on connections between eigenvalue optimization and semidefinite programming, this result may well be the first SDP bound of a combinatorial optimization problem, but in disguised form, see below.

We consider k -partition, where the cardinalities m_i of S_i are also specified through integers

$$m_1 \geq \dots m_k \geq 1 \text{ and } \sum_i m_i = n.$$

Minimizing the weight cut by S is equivalent to maximizing the weight not cut by S , which is equal to $\frac{1}{2}\langle S, AS \rangle$.

The starting point in [DH73] is the observation that any feasible partition matrix S satisfies

$$S^T S = \text{diag}(m_1, \dots, m_k) =: M.$$

In words, the columns of S are pairwise orthogonal. An upper bound on the total weight not cut is therefore given by

$$\max_{S^T S = M} \langle S, AS \rangle = \max_{Y^T Y = I_k} \langle Y, AYM \rangle = \sum_{i=1}^k m_i \lambda_i(A), \quad (29)$$

where $\lambda_1(A) \geq \dots \geq \lambda_n(A)$ are the eigenvalues of A in nonincreasing order.

The first equality follows from the substitution $S = YM^{\frac{1}{2}}$ and the second from Theorem 4 below. The last problem can in fact be expressed as the optimal solution of a semidefinite program, see Section 3.7. Further details and extensions on using this approach are contained for instance in [RW95, KR98].

3.6 Generic cutting planes

In the previous sections we saw SDP relaxations of a variety of graph optimization problems. The matrix dimension typically was $n = |V(G)|$ and the number of (primal) constraints was n in case of (14) and (9). The SDP model of $\vartheta(G)$ can have up to $\frac{n^2}{4}$ equations, depending on $|E(G)|$, and which of the two computational models

(18) and (19) is used. The graph partition models (27) and (28) both have roughly $\frac{n^2}{2}$ sign constraints. All these models can of course be tightened by adding further constraints valid for all points in \mathcal{M} .

In the following we explain a generic way to tighten these semidefinite relaxations by adding further valid constraints. The class of *hypermetric inequalities* provides a huge source of additional cutting planes, some of which also have an explicit combinatorial interpretation. To explain this class we concentrate first on the $-1,1$ model and the related relaxation (14).

Suppose b is an integer vector and $b^T e = 1$. This implies that $b^T x$ is odd $\forall x \in \{-1, 1\}^n$, because any $x \in \{-1, 1\}^n$ can be obtained from e by successively changing the sign of some component i , which changes the inner product with b by $2b_i$, hence the parity is unchanged. Let \mathcal{B} denote the set of all such vectors b ,

$$\mathcal{B} := \{b \text{ integer} : e^T b = 1\}.$$

As a consequence, $b \in \mathcal{B}$ implies $|b^T x| \geq 1 \forall x \in \{-1, 1\}^n$, and therefore

$$\langle bb^T, xx^T \rangle \geq 1.$$

Let us consider the (convex) set

$$HYP := \{X : \langle bb^T, X \rangle \geq 1 \forall b \in \mathcal{B}\},$$

which is the intersection of infinitely many halfspaces. Deza et al [DGL93] show that this set is in fact polyhedral, but it is currently open whether the separation problem $X \in HYP$ can be decided efficiently. The inequalities defining HYP are called *hypermetric inequalities*. Further information can be found for instance in [AU03].

The following subclass of hypermetric inequalities is generated by $b \in \mathcal{B}$ with $b_i \in \{-1, 0, 1\}$ and only three nonzero entries, say b_i, b_j, b_k . Elements in this class are called *triangle inequalities*. The resulting polytope is sometimes called the *metric polytope* MET,

$$MET = \{X : x_{ij} + x_{ik} + x_{jk} \geq -1, x_{ij} - x_{ik} - x_{jk} \geq -1 \forall \text{ distinct } i, j, k\}.$$

These conditions follow from $b^T X b = x_{ii} + x_{jj} + x_{kk} + 2(x_{ij} + x_{ik} + x_{jk}) \geq 1$ in case $b_i = b_j = b_k = 1$ and the implicit assumption $\text{diag}(X) = e$. The other inequalities follow by changing one of the signs in b .

Polynomial time separation for $X \in MET$ can be done trivially by enumerating all $4 \binom{n}{3}$ constraints defining MET. This idea can be generalized to cliques of odd size $k > 3$.

Including all the constraints from MET in (14) results in the following SDP

$$\max \left\{ \frac{1}{4} \langle L, X \rangle : \text{diag}(X) = e, X \in MET, X \succeq 0 \right\} \quad (30)$$

with $O(n^3)$ constraints, which is a computational challenge for current state of the art software to solve SDP. Some computational experience with triangle inequalities and odd order clique constraints combined with (14) is given in [HR98].

Just as it was straightforward to transform 0,1 QP into MC, it is also possible to transform the triangle inequalities into the 0,1 setting. The resulting polyhedron was in fact studied independently under the name of *quadric Boolean polytope*, see for instance [Pad89, Sim90].

3.7 SDP, eigenvalues and the Hoffman-Wielandt inequality

We are now going to take a closer look at connections between eigenvalues of symmetric matrices, optimization over (sets of pairwise) orthogonal vectors and semidefinite programming. To illustrate these connections, we recall the following well known facts. Let $A \in \mathcal{S}_n$. Then we can express the largest eigenvalue $\lambda_{\max}(A)$ of A as

$$\lambda_{\max}(A) = \max_{x^T x = 1} x^T A x. \quad (31)$$

On the other hand, it is trivial to note that

$$\lambda_{\max}(A) = \min \lambda \text{ such that } \lambda I - A \succeq 0,$$

and this is a semidefinite program with dual

$$\lambda_{\max}(A) = \max \langle A, X \rangle \text{ such that } \operatorname{tr}(X) = 1, X \succeq 0. \quad (32)$$

Both problems have strictly feasible solutions, hence strong duality holds. The relations (31) and (32) are the simplest connections relating eigenvalues to optimization over orthogonal vectors on one hand and SDP on the other. It turns out that the following theorem provides a generalization of (31) with many applications in combinatorial optimization.

Theorem 4. *Let $A \in \mathcal{S}_n$, $B \in \mathcal{S}_k$ and $1 \leq k \leq n$. Let $\tilde{B} = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{S}_n$. Then*

$$\min_{X^T X = I_k} \langle X, A X B \rangle = \min_{\phi \text{ injection}} \sum_i \lambda_i(B) \lambda_{\phi(i)}(A) = \langle \lambda(\tilde{B}), \lambda(A) \rangle_-. \quad (33)$$

Remark 1. This theorem has a long history. John von Neumann [Neu37] looks at the more general problem $\max \operatorname{Re}(\operatorname{tr}(AUBV))$, where A and B are square complex matrices and the maximization is carried out over unitary U and V . $\operatorname{Re}(x)$ denotes the real part of the complex number x . The above result follows as a special case. Hoffman and Wielandt [HW53] study the question of the 'distance' of the eigenvalues of two normal matrices in terms of the matrix distance, and also prove the above result in disguised form for the case $k = n$. Berkowitz [Ber74] and Marcus [Mar75] investigate the problem $\min_{X^T X = I_k} E_m(B + X^T A X)$ where A and B are as

in the theorem and $E_m(Y)$ is the m -th elementary symmetric function of degree m of the eigenvalues of Y , hence $E_1(Y) = \text{tr}(Y)$. Therefore, their result proves the above theorem for the case that the smaller matrix is the identity.

Since we consider this a fundamental result, we include the following proof, which is inspired by [HW53] and which is in the spirit of combinatorial optimization. An argument based on first order optimality conditions in nonlinear optimization is given in [RW92].

Proof. First we observe that for $k < n$ we can extend X to an orthogonal matrix $Z = (X \ Y)$ through an appropriate choice of Y . Therefore

$$\langle Z, AZ\tilde{B} \rangle = \langle X, A(X \ Y) \begin{pmatrix} B \\ 0 \end{pmatrix} \rangle + \langle Y, AZ0 \rangle = \langle X, AXB \rangle.$$

We also note for later use that $I = ZZ^T = XX^T + YY^T$, therefore

$$I - XX^T \succeq 0. \quad (34)$$

Hence we may consider $k = n$. A and B are symmetric, therefore they have an orthogonal diagonalization:

$$A = PD_A P^T, \quad B = QD_B Q^T, \quad (35)$$

with $D_A = \text{Diag}(\lambda(A))$ and $D_B = \text{Diag}(\lambda(B))$. Let us assume that P and Q are chosen in such a way that the scalar product of the eigenvalues $\langle \lambda(A), \lambda(B) \rangle$ is minimal. This holds for instance if the elements in $\lambda(A)$ are in nondecreasing order, and those of $\lambda(B)$ in nonincreasing order. Therefore we have

$$\min_{X^T X = I} \langle X, AXB \rangle \leq \langle PQ^T, A(PQ^T)B \rangle = \langle \lambda(A), \lambda(B) \rangle_-.$$

On the other hand, for any orthogonal X we have

$$\langle X, AXB \rangle = \text{tr} D_A (P^T X Q) D_B (P^T X Q)^T = \langle \lambda(A) \lambda(B)^T, (P^T X Q) \circ (P^T X Q) \rangle.$$

Now we observe that for the orthogonal matrix $Y = P^T X Q$ we have $Y \circ Y \in \Omega$. Therefore we get

$$\langle \lambda(A) \lambda(B)^T, Y \circ Y \rangle \geq \min_{Z \in \Omega} \sum_{i,j} \lambda_i(A) \lambda_j(B) z_{ij} = \langle \lambda(A), \lambda(B) \rangle_-.$$

If $k < n$, it is a simple exercise to show that

$$\min_{\phi \text{ injection}} \sum_i \lambda_i(B) \lambda_{\phi(i)}(A) = \langle \lambda(\tilde{B}), \lambda(A) \rangle_-.$$

□

The semidefinite counterpart of this theorem, generalizing (32) has only recently been shown by Anstreicher and Wolkowicz [AW00] for the case $k = n$.

Theorem 5. Let $A, B \in \mathcal{S}_n$. Then

$$\min_{X^T X = I} \langle X, AXB \rangle = \max\{\text{tr}S + \text{tr}T : B \otimes A - I \otimes S - T \otimes I \succeq 0\}.$$

This result is obtained by taking the Lagrangian dual of the first term with respect to the constraints $X^T X = I$, $XX^T = I$. Hence there are two matrices S and T in the SDP. Moreover this is an SDP in the matrix space \mathcal{S}_{n^2} . The general case $k \leq n$ needs some slight modifications.

Theorem 6. Let $A \in \mathcal{S}_n$, $B \in \mathcal{S}_k$ and $1 \leq k \leq n$. Then

$$\min_{X^T X = I_k} \langle X, AXB \rangle = \max\{\text{tr}S - \text{tr}T : B \otimes A - S \otimes I_n + I_k \otimes T \succeq 0, T \succeq 0\}.$$

Remark 2. It is not hard to see that the case $k = n$, considered in [AW00] can be recovered from this result. We refer to [PR08] for further details.

Proof. We first recall that

$$\begin{aligned} z^* &:= \min_{\phi \text{ injection}} \sum_i \lambda_i(B) \lambda_{\phi(i)}(A) \\ &= \min\left\{ \sum_{i,j} \lambda_i(B) \lambda_j(A) z_{ij} : Z = (z_{ij}) \ k \times n, Ze = e, Z^T e \leq e, Z \geq 0 \right\}. \end{aligned}$$

Linear programming duality shows that the last term is equal to

$$\max\left\{ \sum_i s_i - \sum_j t_j : s_i - t_j \leq \lambda_i(B) \lambda_j(A), t \geq 0 \right\}. \quad (36)$$

On the other hand, we also have

$$z^* = \min\{\langle X, AXB \rangle : X^T X = I_k, I_n - XX^T \succeq 0\}$$

after adding the redundant constraint $I - XX^T \succeq 0$, see (34). The factorization (35) suggests the transformation of variables $Y = P^T X Q$ and we get

$$z^* = \min\{\langle Y, D_A Y D_B \rangle : Y^T Y = I_k, I - Y Y^T \succeq 0\}.$$

Duality for semidefinite programs shows that the last term is equal to

$$\max\{\text{tr}(S) - \text{tr}(T) : D_B \otimes D_A - S \otimes I_n + I_k \otimes T \succeq 0, T \succeq 0\}.$$

Here we introduced the multiplier $S \in \mathcal{S}_k$ for the equation $Y^T Y = I_k$ and $T \in \mathcal{S}_n^+$ for $I - Y Y^T \succeq 0$. By restricting S and T to diagonal matrices we get

$$z^* \leq \left\{ \max \sum_i s_i - \sum_j t_j : \lambda_i(B) \lambda_j(A) - s_i + t_j \geq 0, t_i \geq 0 \right\}.$$

But this is again equal to (36), hence there is equality throughout. \square

The Hoffman-Wielandt theorem was used in [HRW92] to get eigenvalue based bounds for the *Quadratic Assignment Problem*

$$\min_{X \in \Pi} \operatorname{tr}\langle X, AXB \rangle.$$

Donath and Hoffman [DH73] use it to formulate the eigenvalue bound (29) for general k -partition.

Further reading: The previous sections have shown the rich structure of semidefinite optimization models applied to integer programming. The idea of using 'matrix liftings' such as (1), has immediate generalizations. Since the extreme points of (1) have again 0-1 coordinates, one could apply another lifting based on this set. This iterated lifting raises interesting research questions. In [LS91] it is shown for instance, that n such liftings suffice to get to the integer optimum for a problem in n binary variables, see also [BCC93, SA90].

Optimization with polynomials has recently also turned out to be another fruitful area for the use of semidefinite optimization. The key observation here is that a polynomial $p(x)$ in n variables $x = (x_1, \dots, x_n)$ is certainly nonnegative for all $x \in \mathbb{R}^n$, if $p(x)$ can be written as a sum of squares of other polynomials in x . Such a sum of square representation can be found by solving a related SDP. Recent results on this can be found for instance in [JL05, Las06].

4 The theoretical power of SDP

Up to now we looked at semidefinite programs as a modeling tool for combinatorial optimization problems having some inherent 'quadratic structure' in their integer formulations. The seminal work of Goemans and Williamson [GW95] from the early 1990's shows that SDP can also be used to generate provably good integer solutions. This approach now runs under the keyword 'hyperplane rounding' and exploits the Gram representation of semidefinite matrices. It was applied quite successfully to various graph partition problems, the most prominent being Max-Cut. In this section we consider the hyperplane rounding idea applied to Max-Cut and graph coloring.

4.1 Hyperplane rounding for Max-Cut

Let us recall the cost function (13) of Max-Cut and its basic semidefinite relaxation (14). Given any matrix $X \succeq 0$ with $\operatorname{diag}(X) = e$, Goemans and Williamson suggest to use it in the following way to generate a random bisection, given by $\xi \in \{-1, 1\}^n$.

Goemans and Williamson observe the remarkable fact that the expectation value of the solution ξ has a simple form.

Goemans-Williamson Hyperplane Rounding Algorithm

Input: $X \succeq 0$ with $\text{diag}(X) = e$, given by $X = V^T V$ with $n \times n$ matrix $V = (v_1, \dots, v_n)$.
Output: Bisection $\xi \in \{-1, 1\}^n$.

Select random vector $r \in \mathbb{R}^n$, uniformly distributed on the unit sphere.
 $\xi_i = 1 \iff r^T v_i \geq 0$.

Theorem 7. *The expectation value of the solution ξ from the hyperplane rounding routine has value*

$$E\left(\sum_{i,j \in E} a_{ij} \frac{1 - \xi_i \xi_j}{2}\right) = \sum_{i,j \in E} a_{ij} \frac{\arccos(v_i^T v_j)}{\pi}. \quad (37)$$

The proof of this result uses the following geometric fact. The probability that v_i and v_j are separated by the random vector r , drawn uniformly on the sphere, is proportional to the angle between the two vectors.

$$\text{Prob}(r \text{ separates } v_i \text{ and } v_j) = \frac{\arccos(v_i^T v_j)}{\pi}. \quad (38)$$

This together with the linearity of the expectation value gives the final result.

The idea now is to use the optimal solution X of the semidefinite relaxation (14) for hyperplane rounding. In order to estimate the quality of the resulting partition ξ , we need to relate the expectation value (37) somehow to the optimal solution value of (14). Goemans and Williamson observe the following lower bound on $\arccos(x)$.

Lemma 6. *For all $-1 \leq x \leq 1$ and $.87856 < \alpha_{GW} < .87857$ it holds that*

$$\frac{\arccos(x)}{\pi} \geq \alpha_{GW} \frac{1}{2} (1 - x).$$

They use it to show that a term by term lower estimate of the expectation value in theorem 7 in case $a_{ij} \geq 0$ leads to

$$\sum_{i,j \in E} a_{ij} \frac{1}{\pi} \arccos x_{ij} \geq \alpha_{GW} \sum_{i,j} a_{ij} \frac{1 - x_{ij}}{2} = \alpha_{GW} z_{sdp}.$$

This is summarized as follows.

Theorem 8. *[GW95] Let $A \geq 0$. Then*

$$z_{mc} > 0.87856 z_{sdp}.$$

Nesterov [Nes97] proposes a different analysis. He uses the identity

$$\sum_{i < j} a_{ij} \frac{\arccos(x_{ij})}{\pi} = \frac{1}{2\pi} \sum_{ij} l_{ij} \arcsin(x_{ij}), \quad (39)$$

which follows easily from the definition of the Laplacian $L = (l_{ij})$ and the fact that

$$\arccos(x) + \arcsin(x) = \frac{\pi}{2}.$$

Let us set $Y = (y_{ij})$ with $y_{ij} = \arcsin(x_{ij})$. Then the right hand side in (39) becomes

$$\frac{1}{2\pi} \langle L, Y \rangle.$$

To compare this to the optimal solution $\frac{1}{4} \langle L, X \rangle$ of (14), Nesterov shows that $Y \succeq X$. The proof of this result is based on the Schur-product theorem.

Theorem 9.

$$A, B \in \mathcal{S}^+ \implies A \circ B \in \mathcal{S}^+.$$

Proof. The following very elegant one-line proof of this result can be found in [BTN01], see Lemma 4.3.1. We have the Gram representation $a_{ij} = a_i^T a_j$ and $b_{ij} = b_i^T b_j$ for appropriate choices of vectors a_i, b_i . Setting $C = (c_{ij}) = (a_{ij} b_{ij})$, we have

$$c_{ij} = a_i^T a_j b_i^T b_j = \langle a_i b_i^T, a_j b_j^T \rangle,$$

showing that C also has a Gram representation, and hence is semidefinite. \square

Lemma 7. *Let $X \succeq 0$ with $\text{diag}(X) = e$ be given. Set $Y = (y_{ij})$ with $y_{ij} = \arcsin(x_{ij})$. Then $Y - X \succeq 0$.*

Proof. Since $\arcsin(x) = x + \frac{1}{2 \cdot 3} x^3 + \frac{1 \cdot 3}{2 \cdot 4 \cdot 5} x^5 \dots$ we have

$$Y - X = \frac{1}{6} X \circ X \circ X + \frac{3}{40} X \circ X \circ X \circ X \circ X \dots$$

The Schur-product theorem shows that the right hand side is positive semidefinite. \square

This gives the following more general estimate.

Theorem 10. *Suppose that the Laplacian $L \succeq 0$. Then*

$$z_{mc} \geq \frac{2}{\pi} z_{sdP}.$$

Since $\frac{2}{\pi} \approx 0.636$, this result is weaker than the Goemans-Williamson analysis, but it is also more general, as $A \geq 0$ implies $L \succeq 0$, but L can be semidefinite even if A has negative entries.

Proof. If $L \succeq 0$, then $\langle L, Y - X \rangle \geq 0$, using the previous lemma. Therefore the expectation value

$$\frac{1}{2\pi} \langle L, Y \rangle \geq \frac{1}{2\pi} \langle L, X \rangle = \frac{2}{\pi} z_{sdp}.$$

□

Subsequently, the hyperplane rounding idea has been extended by Frieze and Jerrum [FJ97] to the maximization version of bisection with $s = \frac{n}{2}$. Since the hyperplane rounding routine does not necessarily give a balanced bisection, the solution has to be modified to ensure $s = \frac{n}{2}$. Frieze and Jerrum [FJ97] propose a 0.651-approximation algorithm. Later, Halperin and Zwick [HZ01] improve this performance ratio to 0.7016 by strengthening the underlying SDP with triangle inequalities, and also by employing a more refined rounding heuristic.

It seems also quite natural to extend the hyperplane rounding idea to Max-k-Cut, starting from the SDP (27). We first note that a random partition into k blocks has two vertices i and j with probability $\frac{1}{k}$ in the same partition block. Therefore the expected weight of this partition is

$$\sum_{i < j} a_{ij} \left(1 - \frac{1}{k}\right).$$

This immediately gives an approximation ratio of $1 - \frac{1}{k}$ for Max-k-Cut. Frieze and Jerrum [FJ97] provide better approximation ratios for all values of k . In particular, they have (not surprisingly) a_{GW} for $k = 2$, 0.8327 for $k = 3$ and 0.8503 for $k = 4$. The last two values have been improved later on by De Klerk et al [dKPW04] to 0.836 for $k = 3$ and 0.8574 for $k = 4$. The minimization versions of these problems are much harder.

4.2 Coloring

The hyperplane rounding idea underlying the theoretical estimate of theorem 8 can be applied to other types of graph problems. Karger, Motwani and Sudan [KMS98] extend this approach to graph coloring.

Determining $\chi(G)$ is well known to be NP-hard. Lund and Yannakakis [LY93] show that, even worse, there exists a constant $\varepsilon > 0$ for which no polynomial algorithm can color any graph with at most $n^\varepsilon \chi(G)$ colors, unless P=NP. In particular, coloring a three-colorable graph with at most four colors is NP-complete, see [KLS00].

In view of these theoretical complexity results it is therefore a challenge to design polynomial algorithms to color three-colorable graphs with nontrivial upper bounds on the colors used. We first recall the following easy facts.

Lemma 8. *If $\chi(G) = 3$ then $N(v)$ is bipartite for any $v \in V(G)$.*

Lemma 9. *Let G be a graph with maximum vertex degree Δ . Then G can be colored in polynomial time with $\Delta + 1$ colors.*

Widgerson [Wid83] observes that if $\Delta > \sqrt{n}$, then 2 colors suffice to color the neighbourhood of the vertex with largest degree, thereby legally coloring at least \sqrt{n} vertices. If $\Delta \leq \sqrt{n}$, then the graph can be colored with $\sqrt{n} + 1$ colors. Hence the following simple algorithm from Widgerson [Wid83] colors any three-colorable graph with at most $3\sqrt{n}$ colors.

Widgerson Algorithm for three-colorable graphs

Input: A graph G on n vertices with $\chi(G) = 3$.
Output: A coloring of G with at most $O(\sqrt{n})$ colors.

- (1) While $\exists v \in V(G)$ with degree $\geq \sqrt{n}$
 Color v with color 1 and $N(v)$ with (at most) two new colors.
 Remove v and $N(v)$ from G and call the remaining graph G .
 - (2) Color G with at most $\Delta < \sqrt{n}$ colors.
-

It turns out that the semidefinite program, underlying the ϑ function, can be used to improve Widgerson's algorithm. In fact, $\vartheta(G)$ is one of the ingredients for the currently strongest algorithms for three-colorable graphs.

Here is a first analysis. Suppose we have computed the optimal solution $V = (v_1, \dots, v_n)$ of (26). Therefore $\|v_i\| = 1$ and $v_i^T v_j = \lambda \leq -\frac{1}{2}$ for $ij \in E(G)$, because $\lambda \leq -\frac{1}{\chi(G)-1}$. We associate the vectors v_i to the vertices of G , and use them to define the following vertex partition of $V(G)$.

We first select t random hyperplanes through the origin. t will be specified later. These hyperplanes partition \mathbb{R}^n into 2^t regions and each v_i belongs to exactly one region. All vertices i with v_i in the same region R_j now are assigned to partition block j .

The main task now is to show that an appropriate choice of t will ensure that with high probability at least $\frac{n}{2}$ vertices in this partition are legally colored (= adjacent vertices are in different partition blocks).

First note that the probability that two vertices i and j are in the same region is equal to the probability that none of the t hyperplanes separates v_i and v_j , which is equal to

$$\left(1 - \frac{1}{\pi} \arccos(v_i^T v_j)\right)^t,$$

see (38). We have just seen that $ij \in E$ implies $v_i^T v_j = \lambda \leq -\frac{1}{2} = \cos(\frac{2\pi}{3})$. Therefore $\arccos(v_i^T v_j) \geq \frac{2\pi}{3}$, so

$$\left(1 - \frac{1}{\pi} \arccos(v_i^T v_j)\right)^t \leq \frac{1}{3^t}.$$

The expected value of edges with both endpoints in the same partition block (= monochromatic edges) is at most

$$|E| \frac{1}{3^t} \leq \frac{n\Delta}{2} 3^{-t}.$$

Markov's inequality now tells us that with probability less than $\frac{1}{2}$ there are more than $n\Delta 3^{-t}$ monochromatic edges. After repeating this process several times, we have, with high probability, a coloring with 2^t colors, where at most $n\Delta 3^{-t}$ edges are monochromatic. Selecting $t = 2 + \log_3(\Delta)$ gives at most $\frac{n}{4}$ monochromatic edges. Therefore at most $\frac{n}{2}$ vertices are not colored legally. In other words, with high probability at least $\frac{n}{2}$ vertices are legally colored with at most $2^t \leq 8\Delta^{\log_3(2)}$ colors.

As $\log_3(2) \approx 0.631$ and Δ can be $O(n)$, this technique by itself does not improve Widgerson's result. In [KMS98] it is suggested to run the while-loop of Widgerson's algorithm as long as $\Delta > n^{0.613}$. This uses up $O(n^{0.387})$ colors. Coloring the remaining graph, having $\Delta < O(n^{0.613})$, can now be done using $O((n^{0.613})^{\log_3(2)}) = O(n^{0.387})$ colors. This gives, with high probability, a coloring using at most $O(n^{0.387})$ colors.

Karger, Motwani and Sudan provide another more refined analysis, which shows that in fact $\tilde{O}(n^{0.25})$ colors suffice. We use the \tilde{O} notation to suppress polylogarithmic terms. We sketch this analysis, but follow the simplified presentation from [ACC06]. The first observation is that if we can find a stable set of size $O(\frac{n}{s(n)})$ for some function $s(n)$, then we can use up one color for this stable set, and iterate at most $s(n)$ times to get a coloring using at most $\tilde{O}(s(n))$ colors. In [KMS98] it is shown that a stable set of size $\tilde{O}(\frac{n}{\Delta^{1/3}})$ can be found by hyperplane rounding, see below. This is used as follows. Apply the while loop of Widgerson's algorithm as long as $\Delta \geq n^{0.75}$. This uses up $O(n^{0.25})$ colors and leaves a graph with $\Delta < n^{0.75}$. Now we apply the above stable set argument with $\Delta^{1/3} = O(n^{0.25})$ to find a coloring in the remaining graph using $\tilde{O}(n^{0.25})$ colors. In summary, we obtain a coloring using $\tilde{O}(n^{0.25})$ colors, once we can find stable sets of size $O(\frac{n}{\Delta^{1/3}})$.

We start with a solution of (26) of value $\lambda = -\frac{1}{2}$, which exists because $\chi(G) = 3$. We consider for a random vector r and $\varepsilon > 0$ to be specified later the set

$$V_r(\varepsilon) := \{i \in V : \langle r, v_i \rangle \geq \varepsilon\}.$$

We assume that the entries r_j of r are drawn independently from the standard normal distribution. This implies that for any unit vector v , $v^T r$ also has the standard normal distribution. Therefore

$$\text{Prob}(i \in V_r(\varepsilon)) = \int_{\varepsilon}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt =: N(\varepsilon).$$

Let $I \subseteq V_r(\varepsilon)$ be the set of isolated vertices in $V_r(\varepsilon)$. Next note that

$$\text{Prob}(i \notin I | i \in V_r(\varepsilon)) = \text{Prob}(\exists j \in N(i) : \langle r, v_j \rangle \geq \varepsilon | \langle r, v_i \rangle \geq \varepsilon).$$

Having $j \in N(i)$ means that $\langle v_i, v_j \rangle = -\frac{1}{2}$. Therefore we can introduce $\tilde{v}_j \perp v_i$ such that

$$v_j = -\frac{1}{2}v_i + \frac{\sqrt{3}}{2}\tilde{v}_j.$$

Therefore $\tilde{v}_j = \frac{2}{\sqrt{3}}v_j + \frac{1}{\sqrt{3}}v_i$. If $\langle r, v_j \rangle \geq \varepsilon$ and $\langle r, v_i \rangle \geq \varepsilon$, then $\langle r, \tilde{v}_j \rangle \geq \sqrt{3}\varepsilon$. Therefore

$$\begin{aligned} \text{Prob}(i \notin I | i \in V_r(\varepsilon)) &\leq \\ \text{Prob}(\exists j \in N(i) : \langle r, \tilde{v}_j \rangle \geq \sqrt{3}\varepsilon) &\leq \sum_{j \in N(i)} \text{Prob}(\langle r, \tilde{v}_j \rangle \geq \sqrt{3}\varepsilon) \leq \Delta N(\sqrt{3}\varepsilon). \end{aligned}$$

Now $\varepsilon > 0$ is selected such that

$$N(\sqrt{3}\varepsilon) < \frac{1}{2\Delta}. \quad (40)$$

This implies that with probability less than $\frac{1}{2}$, a vertex is not in I , thus $\text{Prob}(i \in I) \geq \frac{1}{2}\text{Prob}(i \in V_r(\varepsilon)) = \frac{1}{2}N(\varepsilon)$. Hence the expected cardinality of the stable set I is $\frac{n}{2}N(\varepsilon)$. To get the final result, one needs to see how the choice of ε in (40) relates to Δ and $N(\varepsilon)$. It can be shown that Δ is less than $\tilde{\Theta}(N(\varepsilon)^3)$, see [ACC06], hence $\frac{n}{2}N(\varepsilon) = \tilde{\Omega}(\Delta^{-1/3}n)$.

In [ACC06], a further refinement of the analysis is proposed, which together with a strengthened version of (26) gives the an approximation of $\tilde{O}(n^{0.2111})$ colorings for three-colorable graphs, see also the recent dissertation [Chl09].

5 Solving SDP in practice

We have just seen the great variety of modeling ideas leading to SDP. Moreover, some of these models even lead to relaxations where the approximation error can be determined a priori. It is therefore natural to ask for efficient methods to solve SDP. In this section we are going to describe the algorithmic machinery currently available for SDP.

5.1 Interior Point algorithms

The most elegant way to solve SDP, and more generally linear optimization over closed convex cones, is based on Newton's method applied to a slightly modified version of the problem. The theoretical background for this approach goes back to the seminal work of Nesterov and Nemirovski from the late 1980's, see [NN94]. They showed that a family of convex optimization problems can be solved in polynomial time using self-concordant barrier functions.

Practical experience indicates that Newton's method in combination with the barrier idea works most efficiently in the primal-dual path-following setting, which will be briefly explained now. We recall the primal problem (2) and its dual (5),

$$\min \langle C, X \rangle \text{ such that } A(X) = b, X \succeq 0,$$

$$\max b^T y \text{ such that } Z = C - A^T(y) \succeq 0.$$

In the following we assume that both problems satisfy the Slater constraint qualification (6). In this case (X, y, Z) is optimal if and only if (7) holds. This is a system of $m + \binom{n+1}{2} + 1$ equations with the additional semidefiniteness constraints in $m + 2\binom{n+1}{2}$ variables. It is perhaps surprising that the additional semidefiniteness conditions in fact lead to an overdetermined system. Indeed note that $X \succeq 0, Z \succeq 0$ implies $X = UU^T, Z = VV^T$ for U and V of appropriate size. But then

$$0 = \langle Z, X \rangle = \langle VV^T, UU^T \rangle = \|V^T U\|^2$$

implies $V^T U = 0$ and therefore

$$ZX = VV^T UU^T = 0.$$

Thus the scalar equation

$$\langle Z, X \rangle = 0$$

together with $X \succeq 0, Z \succeq 0$ implies the matrix equation

$$ZX = 0.$$

Since ZX need not be symmetric, even if X and Z are symmetric, this equation has n^2 components.

Primal-dual path-following interior-point methods are based on solutions $X \succ 0, Z \succ 0$, of the following slightly modified optimality conditions for parameter $\mu > 0$.

$$A(X) - b = 0, \quad C - Z - A^T(y) = 0, \quad ZX - \mu I = 0. \quad (41)$$

Clearly, any solution of this system must satisfy $X \succ 0, Z \succ 0$, as $ZX = \mu I$ forces X and Z to be nonsingular. It turns out that strict feasibility in fact characterizes unique solvability of (41).

Theorem 11. (see for instance Theorem 10.2.1 in [WSe00]) *The following statements are equivalent:*

(2) and (5) both have strictly feasible points.

(41) has a unique solution (X_μ, y_μ, Z_μ) for all $\mu > 0$.

The proof is based on the following observation. Let $f : \mathcal{S}_n^{++} \mapsto \mathbb{R}$, $f(X) = \log \det X$. This function is strictly concave with $\nabla f = X^{-1}$,

$$f(X+h) = f(X) + \langle X^{-1}, h \rangle + o(\|h\|).$$

Consider the auxiliary problem, for fixed parameter $\mu > 0$.

$$(PB) \quad \min \langle C, X \rangle - \mu f(X) \text{ such that } A(X) = b, X \succ 0.$$

This is a convex optimization problem with Lagrangian

$$L(X, y) = \langle C, X \rangle - \mu \log \det X + y^T (b - A(X)).$$

The necessary and sufficient optimality conditions therefore are

$$\nabla_X L = C - \mu X^{-1} - A^T(y) = 0,$$

$$\nabla_y L = b - A(X) = 0,$$

together with the open set constraint $X \succ 0$. Setting $Z = \mu X^{-1}$, we recover (41). Problem (PB) is sometimes called the *primal barrier problem*. The term $-\mu \log \det X$ goes to infinity as $X \succ 0$ approaches the boundary of the cone of semidefinite matrices. The value of μ controls the influence of this term. The conditions (41) can also be derived by setting up a barrier problem for the dual SDP. Strict convexity of the cost function from the primal barrier problem shows that a minimizer, if it exists, must be unique. To show the existence of a minimizer requires an additional compactness argument. We refer to [WSe00] for further details.

In summary, strict feasibility of (2) and (5) insures unique solutions (X_μ, y_μ, Z_μ) for all $\mu > 0$. To show that these solutions form a smooth curve, parametrized by μ , we need to show that the system is differentiable with a nonsingular Jacobian. While differentiability is obvious, the Jacobian of (41) is certainly not invertible, as it is not even square. The remedy is to replace the nonsymmetric equation $ZX - \mu I = 0$ by an equivalent symmetric one. A general setup for this goes as follows. Let P be nonsingular and define

$$H_P(M) := \frac{1}{2}(PMP^{-1} + (PMP^{-1})^T).$$

Replacing $ZX - \mu I = 0$ by $H_P(ZX) - \mu I = 0$ in (41) makes the Jacobian a square matrix, which can be shown to be nonsingular for any invertible P and for all points (X, y, Z) 'close' to solutions (X_μ, y_μ, Z_μ) of (41). Hence the solution set of (41) indeed defines a smooth curve which is often called (*primal-dual*) *central path*. Let us denote it by $CP := \{P_\mu = (X_\mu, y_\mu, Z_\mu) : \mu > 0\}$.

Primal-dual interior-point path-following methods use the Newton method to follow the primal-dual central path, maintaining $X \succ 0$, $Z \succ 0$ (interior points) for $\mu \rightarrow 0$. To follow the central path (at least approximately), we first need to come close to it. We do this iteratively as follows. Having the current interior iterate (X, y, Z) and a target value for μ , we try to reach a new interior point (X^+, y^+, Z^+) close to P_μ for given μ . Then the target value μ is reduced and we iterate. The main work is done in determining the search direction $(\Delta X, \Delta y, \Delta Z)$ which moves us from the current iterate to the new point. As mentioned before, there is a great variety to do so and here we show a very simple, popular and efficient variant. It was one of the first search directions proposed to solve SDP, see [HRVW96, KSH97, Mon97]. We linearize the equations in (41) and get

$$A(\Delta X) = b - A(X) := r_p, \tag{42}$$

$$\Delta Z + A^T(\Delta y) = C - A^T(y) - Z := r_d, \quad (43)$$

$$Z\Delta X + \Delta ZX = \mu I - ZX. \quad (44)$$

The vectors r_p and r_d express primal and dual infeasibility and would be 0 if feasible starting points were used. The last equation can be used to eliminate ΔX ,

$$\Delta X = \mu Z^{-1} - X - Z^{-1}\Delta ZX,$$

the second one gives

$$\Delta Z = r_d - A^T(\Delta y).$$

Substitution into the first equation gives

$$A(Z^{-1}A^T(\Delta y)X) = v \text{ with } v = r_p - A(\mu Z^{-1} - X) + A(Z^{-1}r_dX).$$

The linear operator on the left hand side has the matrix representation

$$M\Delta y = v, \quad (45)$$

with $m_{ij} = \text{tr}(A_i Z^{-1} A_j X)$. This matrix can be shown to be positive definite (if $X \succ 0$, $Z \succ 0$ and the equations $A(X) = b$ are linearly independent). To determine the search direction we therefore need to solve the equation $M\Delta y = v$. Having Δy we get ΔZ and ΔX from backsubstitution. ΔX need not be symmetric, so the symmetric part of ΔX is taken.

This would give the new point $X + \Delta X, y + \Delta y, Z + \Delta Z$ except that the new point need not have definite matrices. This is repaired by a backtracking strategy, where starting with $t = 1$, t is reduced to a value $t^* > 0$ such that $X^+ = X + t^*\Delta X \succ 0, Z^+ = Z + t^*\Delta Z \succ 0$. This gives a new trial point (X^+, y^+, Z^+) . The new target parameter for this point can be estimated to be

$$\mu^+ = \frac{1}{n} \text{tr}(X^+ Z^+),$$

and a new iteration is started with μ^+ reduced by a multiplicative factor.

It was shown by Monteiro [Mon97] that this rather simple and pragmatic approach in fact falls in the general class of search directions mentioned above.

To give some idea on the actual work involved, we consider the basic semidefinite relaxation for Max-Cut. Here $A(X) = \text{diag}(X)$ and $A^T(y) = \text{Diag}(y)$. It is not too hard to verify that in this case the matrix M in (45) has the simple form

$$M = X \circ Z^{-1}.$$

Hence the work in each iteration involves computing Z^{-1} , solving the equation (45) and doing a few matrix multiplications to get ΔX . An efficient way to do the backtracking to stay inside the cone of semidefinite matrices consists in checking whether the Cholesky decomposition of $X + t\Delta X$ terminates successfully. This is a certificate that the matrix is semidefinite, hence an additional small reduction of

t insures definiteness. To give some practical impression we tabulate computation times to solve this relaxation for some representative values of n . The iterations were stopped, once the relative error was below 10^{-7} , and it always took less than 20 iterations to reach this level of accuracy. The results clearly show that interior-point methods are indeed very efficient on smaller problems ($n \leq 1000$), but become prohibitive with respect to both time and space requirements, once n gets large.

We also tabulate timings to compute the ϑ -function in the computationally most expensive case of $m = \frac{1}{2} \binom{n}{2}$ equations, see Table 5.1. It is clear that once $n \approx 200$, the effort of interior point methods gets prohibitive.

Looking at the computation times in these two tables, it should be clear that interior-point methods become impractical, once m is substantially larger than say 5000, or once n is larger than about 1000. In the following sections we will consider algorithmic alternatives for larger problems.

n	time (secs.)
500	12
1000	75
1500	237
2000	586
2500	1109
3000	1900

Table 1 Interior-point computation times to solve (14) with relative accuracy 10^{-7} . Here $m = n$.

n	m	time (secs.)
100	2475	57
120	3570	161
140	4865	353
160	6360	757
180	8055	1520
200	9950	2817

Table 2 Interior-point computation times to solve (18) with relative accuracy 10^{-7} , $m = \frac{1}{2} \binom{n}{2}$.

5.2 Partial Lagrangian and the bundle method

We have just seen that interior-point methods are the method of choice to solve SDP and we also saw the limits of this approach both in terms of the dimension n of the matrix space and also in the number m of (primal) constraints. We also saw the need to be able to handle large-scale SDP to get good approximations of the underlying combinatorial optimization problem.

We are now going to describe a rather generic 'work-around' for problems, where the matrix dimension is reasonable, but m can be arbitrary. Let us consider an SDP of the form

$$z = \max\{\langle C, X \rangle : A(X) = a, B(X) = b, X \succeq 0\}.$$

We have split the constraints into two sets. The motivation is that maintaining only the first set $A(X) = a$ would result in an SDP that is still manageable by interior point methods, but the inclusion of $B(X) = b$ makes the problem impractical for interior-point methods. For simplicity of exposition, we consider only equations. The presence of inequalities only leads to minor modifications (sign constraints on dual multipliers), which can be dealt with in the approach to be described.

Let us denote $\mathcal{X} := \{X : A(X) = a, X \succeq 0\}$. The idea now is to maintain $X \in \mathcal{X}$ explicitly and to put $B(X) = b$ into the cost function by taking the partial Lagrangian dual. Hence we get

$$z = \max\{\langle C, X \rangle : A(X) = a, B(X) = b, X \succeq 0\} = \max_{X \in \mathcal{X}} \min_y L(X, y),$$

where the partial Lagrangian L is given by $L(X, y) = \langle C, X \rangle + y^T(b - B(X))$. Applying the Minimax inequality we get under the usual strict feasibility conditions that

$$z = \min_y f(y) \leq f(y) \quad \forall y,$$

where $f(y) = \max_{X \in \mathcal{X}} L(X, y)$. Evaluating f amounts to solving an SDP over $X \in \mathcal{X}$ which we assume to be manageable. Suppose that for some y^* , we have $f(y^*) = L(X^*, y^*)$, so the maximum is attained at $X^* \in \mathcal{X}$. By setting

$$g^* = b - B(X^*)$$

we get for any y

$$f(y) \geq L(X^*, y) = f(y^*) + \langle g^*, y - y^* \rangle. \quad (46)$$

The inequality follows from the definition of f , the equality comes from substituting g^* . In the language of convex analysis, this inequality defines g^* to be a *subgradient* of f at y^* .

To compute z , we minimize f . This function is continuous and convex (point-wise maximum of linear functions), but it is not differentiable at points where the maximum is not unique. Hence we use some tools from nonsmooth optimization to minimize f .

The *bundle method* will serve our purposes. It was introduced in the 1970's by Lemarechal, see [Lem75, Lem78]. A comprehensive survey is also contained in [HUL93]. We briefly explain its key features. The method iteratively approaches a minimizer of f . Let the current iterate be \hat{y} . Suppose we have evaluated f at $k \geq 1$ points y_1, \dots, y_k with respective optimizers X_1, \dots, X_k and subgradients $g_i = b - B(X_i)$. We also set $f_i = f(y_i)$. It is assumed that $\hat{y} \in \{y_1, \dots, y_k\}$. To get started we evaluate f at $\hat{y} = 0$ and set $y_1 = \hat{y}, k = 1$.

The subgradient inequality (46) implies that for all y

$$f(y) \geq \max_i \{f_i + g_i^T(y - y_i)\} =: \tilde{f}(y).$$

To simplify the presentation we set $h_i = f_i - g_i^T y_i$, $H = (h_1, \dots, h_k)^T$ and $G = (g_1, \dots, g_k)$. Then

$$\tilde{f}(y) = \max_{\lambda \in \Delta_k} \lambda^T (H + G^T y).$$

The bundle method now uses the minorant $\tilde{f}(y)$ as an approximation of the original f 'close' to the current iterate \hat{y} . This is plausible because $\hat{y} \in \{y_1, \dots, y_k\}$ implies $\tilde{f}(\hat{y}) = f(\hat{y})$. To insure that we stay close to \hat{y} , we add a regularization term and consider the following function

$$f_{bdl}(y) := \tilde{f}(y) + \frac{1}{2t} \|y - \hat{y}\|^2.$$

The bundle method minimizes this function over y to obtain a new iterate. The parameter $t > 0$ controls how close we stay to \hat{y} . Minimizing $f_{bdl}(y)$ is again a Min-Max problem which can be simplified as follows.

$$\min_y f_{bdl}(y) = \max_{\lambda \in \Delta} \min_y \lambda^T (H + G^T y) + \frac{1}{2t} \|y - \hat{y}\|^2.$$

The inner minimization is a strictly convex unconstrained quadratic optimization problem, hence we can replace the minimization by asking that the first order optimality conditions hold. Setting the derivative with respect to y equal to 0 gives

$$y = \hat{y} - tG\lambda.$$

After substitution, we get the equivalent problem

$$\max_{\lambda \in \Delta} \lambda^T (H + G^T \hat{y}) - \frac{t}{2} \|G\lambda\|^2.$$

This is a convex quadratic problem over the standard simplex Δ in \mathbb{R}^k and can be solved by standard methods from convex optimization. The solution effort depends on k , which can be controlled by the user. Having the optimizer λ^* of this problem, we get the new estimate

$$y^{new} = \hat{y} - tG\lambda^*.$$

The bundle method now asks to evaluate the original function f at y^{new} . Some standard criteria are used to decide whether y^{new} becomes the new trial point, or whether we stay at \hat{y} . In any case, information from the new point is included as $y_{k+1} = y^{new}$ and the iteration is continued. The convergence theory of the bundle method is quite elaborate and can be found for instance in [HUL93].

Remark 3. Let us take a careful look at the above derivation of the bundle method. The specific form of the function $f(y)$ is in fact irrelevant, once the following property of f is satisfied. For any y^* we can determine a vector g^* such that

$$f(y) \geq f(y^*) + \langle g^*, y - y^* \rangle \quad \forall y$$

holds. In words, we assume that f is convex and we are able to determine a subgradient of f at any point.

We close with an application of the bundle method applied to the semidefinite relaxation of Max-cut which also includes the triangle inequalities, see (30). We maintain the equation $\text{diag}(X) = e$ explicitly and dualize the triangle inequalities, which we formally denote by $B(X) \leq b$. Recall that there are $4\binom{n}{3}$ inequality constraints. Let us denote the violation of these constraints by

$$r := \min\{0, b - B(X)\}.$$

In the table 5.2 we consider the instance g3s of size $n = 300$ from [FGRS06]. This is a rudy-generated random graph with edge density of 10% and edge weights 1 and -1. The optimal value of the relaxation, as reported in [FGRS06] is 635.05. In the table we also provide information about the error r . We include the total violation of all constraints, $\|r\|_1$, the maximal violation $\|r\|_\infty$ and the total number of violated constraints (last column). The results in this table clearly indicate that the bundle method is very efficient in getting close to the optimum quickly. The computation time for 100 bundle iterations was a few minutes only. The local convergence behaviour is obviously much weaker than in the case of interior-point methods. We refer to [FGRS06] for further details of this approach applied to (30).

iter	$f(y)$	$\ r\ _1$	$\ r\ _\infty$	contrs. viol.
1	679.3	152541.0	0.96	680822
10	660.4	21132.7	0.73	147094
20	648.1	1234.6	0.52	13605
30	642.2	193.7	0.32	2979
40	639.5	50.8	0.32	957
50	638.2	29.5	0.25	647
60	637.6	25.3	0.26	570
70	637.1	28.9	0.20	688
80	636.9	17.1	0.23	397
90	636.6	18.2	0.18	448
100	636.5	13.5	0.18	369

Table 3 The semidefinite relaxation of Max-Cut from (30) for a graph with $n = 300$. The vector r contains the violation of the triangle inequalities. The last column provides the number of violated constraints.

5.3 The spectral bundle method

The spectral bundle method introduced in [HR00] reduces the solution of SDP to the computation of the largest eigenvalue of a sequence of symmetric matrices. The algorithmic machinery from numerical linear algebra provides methods to compute $\lambda_{\max}(C)$, which do not require to have C explicitly available, but only need a subroutine that evaluates the action of C . In other words, given x , we only need to be able to compute $y = Cx$.

Before describing the spectral bundle method in detail we first show that SDP with the constant trace property can equivalently be reformulated as an eigenvalue optimization problem.

The mapping A from (2) satisfies the *constant trace property* if the identity matrix is in the range of A^T , which means $\exists \eta$ such that $A^T(\eta) = I$. In this case any X such that $A(X) = b$, $X \succeq 0$ satisfies

$$\text{tr}(X) = \langle X, A^T(\eta) \rangle = \langle A(X), \eta \rangle = b^T \eta = a,$$

for some constant $a \geq 0$. The constant trace property therefore implies that feasible solutions of (2) have constant trace, equal to a . Excluding the case $a = 0$, which only has the zero matrix as feasible solution of SDP, we can assume without loss of generality that $a = 1$.

Let us consider SDP with the constant trace property. In this case we can add the redundant equation

$$\text{tr}(X) = 1$$

to (2) and get for the dual, with multiplier λ for the new constraint:

$$\min\{b^T y + \lambda : A^T(y) + \lambda I - C = Z \succeq 0\}.$$

The optimality condition $ZX = 0$ together with $\text{tr}(X) = 1$ implies that any optimal Z is singular. Hence, at the optimum we have

$$0 = \lambda_{\min}(Z) = \lambda_{\max}(-Z) = \lambda_{\max}(C - A^T(y)) - \lambda.$$

We conclude that the multiplier λ satisfies $\lambda = \lambda_{\max}(C - A^T(y))$. Substitution gives the following function $f(y) = b^T y + \lambda_{\max}(C - A^T(y))$. Solving the dual is therefore equivalent to the eigenvalue optimization problem

$$\min_y f(y).$$

The condition $Z \succeq 0$ is hidden in $\lambda_{\min}(Z) = 0$, which moves $\lambda_{\max}(C - A^T(y))$ into the cost function. It can easily be shown that f is convex. The cost function is smooth but not differentiable in case the largest eigenvalue has multiplicity larger than one. Suppose now that x^* is a unit-norm eigenvector to $\lambda_{\max}(C - A^T(y^*))$. Then, see (31),

$$\lambda_{\max}(C - A^T(y)) \geq \langle x^*, (C - A^T(y))x^* \rangle \quad \forall y.$$

Let us define

$$g^* := b - A(x^* x^{*T}).$$

Then the above inequality shows that

$$f(y) \geq f(y^*) + \langle g^*, y - y^* \rangle \quad \forall y.$$

Therefore g^* is subgradient of f at y^* . Moreover, if the multiplicity of $\lambda_{\max}(C - A^T(y^*))$ is one, then x^* is unique up to multiplication by -1 , hence g^* is unique as well, and $\nabla f(y^*) = g^*$. In view of all this, a first idea would be to use again the bundle method to minimize f . Indeed, Schramm and Zowe [SZ92] apply it to compute $\vartheta(G)$. We will now see that we can in fact do better by exploiting the special form of the objective function. We recall from (32) that

$$\lambda_{\max}(A) = \max\{\langle A, W \rangle : \text{tr}(W) = 1, W \succeq 0\}.$$

After substitution, we get the following min-max problem for the dual SDP.

$$\min_y \max_{\text{tr}(W)=1, W \succeq 0} b^T y + \langle C - A^T(y), W \rangle \quad (47)$$

In the *spectral bundle method*, this problem is solved iteratively. We observe that evaluating f at \hat{y} amounts to compute $\lambda_{\max}(C - A^T(\hat{y}))$ together with an eigenvector v . Having a current iterate \hat{y} , the following modifications are made in (47). First, the maximization is simplified by constraining W to be of the form

$$W = PVP^T$$

for given $n \times k$ matrix P such that $P^T P = I_k$. The idea is that P should contain 'local' information of f around \hat{y} . In particular, we assume that the eigenvector v is contained in P . The new variable now is $V \in \mathcal{S}_k^+$. Since

$$\lambda_{\max}(C) = \max_{W \succeq 0, \text{tr}(W)=1} \langle C, W \rangle \geq \max_{W \succeq 0, \text{tr}(W)=1, W=PVP^T} \langle C, W \rangle = \lambda_{\max}(P^T C P),$$

we get the following minorant

$$\tilde{f}(y) = \max_{V \succeq 0, \text{tr}(V)=1} b^T y + \langle C - A^T(y), PVP^T \rangle \leq f(y)$$

of f . The inclusion of v in P insures that $f(\hat{y}) = \tilde{f}(\hat{y})$. To insure that the next iterate stays close to \hat{y} , we consider the following replacement of (47) for fixed parameter $t > 0$.

$$\min_y \max_{V \succeq 0, \text{tr}(V)=1} b^T y + \langle C - A^T(y), PVP^T \rangle + \frac{1}{2t} \|y - \hat{y}\|^2.$$

Note the similarity to the standard bundle method from before. In the spectral bundle method, this min-max problem is solved to get the next trial point y . As before we exchange min and max and exploit the fact that the minimization with respect to y is

again an unconstrained strictly convex quadratic problem. Therefore y is minimizer if and only if the partial derivative with respect to y is zero. This results in

$$y = \hat{y} - t(b - A(PVP^T)).$$

We now substitute this for y and get the following quadratic SDP

$$\max_{V \geq 0, \text{tr}(V)=1} b^T \hat{y} + \langle C - A^T(\hat{y}), PVP^T \rangle - \frac{t}{2} \|b - A(PVP^T)\|^2.$$

This problem has just one scalar equation and can be solved by interior-point methods to get the optimal $V^* \in \mathcal{S}_k^+$. The new trial point $y^{new} = \hat{y} - t(b - A(PV^*P^T))$ is now used to compute the function value f , together with an eigenvector to λ_{\max} . We follow the usual bundle concept to decide whether or not y^{new} becomes the new trial point. In any case the matrix P is updated and a new iteration can be started. Helmberg and Rendl [HR00] explain in detail how the above quadratic SDP can be solved. Various update strategies for P are discussed and an elementary convergence analysis is given. Helmberg [Hel03] describes implementation issues and presents computational results on a wide variety of SDP. Refinements of the spectral bundle method are given in [HKR98] and [HO00].

Remark 4. The similarities of the spectral to the standard bundle method are quite obvious. In fact, constraining V to be a diagonal matrix (with diagonal entries λ_i) simplifies the above SDP to optimizing over $\lambda \in \Delta$, and we recover the standard bundle method in this case.

6 SDP and beyond

6.1 Copositive and completely positive matrices

In this section we will see that besides the cone of semidefinite matrices, there are several other cones in the space of symmetric matrices which have a close connection to integer programming. Let us define

$$\mathcal{C}^* := \{X \in \mathcal{S}_n : X = VV^T \text{ with } n \times k \text{ matrix } V \geq 0\} = \text{conv}\{vv^T : v \in \mathbb{R}^n, v \geq 0\}.$$

Matrices in \mathcal{C}^* are often called *completely positive*. The cone \mathcal{C}^* has a dual, which we denote by \mathcal{C} and which by definition is given as follows.

$$Y \in \mathcal{C} \iff \langle Y, X \rangle \geq 0 \quad \forall X \in \mathcal{C}^*.$$

This obviously holds if and only if

$$v^T Y v \geq 0 \quad \forall v \geq 0. \tag{48}$$

Matrices in this cone are usually called *completely positive*. While $X \in \mathcal{S}^+$ has an efficient certificate, given for instance through the Cholesky decomposition of X , it is NP-hard to decide whether $X \notin \mathcal{C}$, see [MK87].

We call problems of the form

$$\inf\{\langle C, X \rangle : A(X) = b, X \in \mathcal{C}\} \text{ and}$$

$$\inf\{\langle C, X \rangle : A(X) = b, X \in \mathcal{C}^*\}$$

copositive programs because either the problem or its dual involves optimization over copositive matrices.

6.2 Copositive relaxations

To see that copositive programs have some relevance in connection with integer programs we recall the following theorem from Motzkin and Strauss.

Theorem 12. ([MS65]) *Let A be the adjacency matrix of a graph. Then*

$$\frac{1}{\alpha(G)} = \min\{x^T(A+I)x : x \in \Delta\}.$$

Starting from this fact, it is not hard to show the following result, which was pointed out by De Klerk and Pasechnik, see [dKP02].

Theorem 13. *Let A be the adjacency matrix of a graph. Then*

$$\alpha(G) = \max\{\langle J, X \rangle : \langle A+I, X \rangle = 1, X \in \mathcal{C}^*\} = \min\{\lambda : \lambda(A+I) - J \in \mathcal{C}\}.$$

Proof. Let S be a stable set of maximum cardinality $\alpha(G)$ with characteristic vector $\xi \in \{0, 1\}^n$. Then $\frac{1}{\alpha} \xi \xi^T$ is feasible for the maximization problem and we get the first inequality in

$$\alpha \leq \sup\{\langle J, X \rangle : \langle A+I, X \rangle = 1, X \in \mathcal{C}^*\} \leq \inf\{\lambda : \lambda(A+I) - J \in \mathcal{C}\}.$$

Weak duality for conic programs implies the second inequality.

The Motzkin-Strauss theorem shows that

$$0 = \min\{x^T(A+I - \frac{1}{\alpha}ee^T)x : x \in \Delta\} = \min\{x^T(\alpha(A+I) - J)x : x \geq 0\}.$$

The second minimization being zero is the defining condition for $\alpha(A+I) - J$ to be in \mathcal{C} , see (48). Therefore the infimum above is at most α , but weak duality states that it is also at least α , hence there is equality throughout, and both the supremum and the infimum are attained (at $\frac{1}{\alpha} \xi \xi^T$ and $\lambda = \alpha$ respectively). \square

DeKlerk and Pasechnik provide a proof for this result which is independent of the Motzkin-Strauss theorem. Let us put this result into perspective by recalling $\vartheta(G)$.

$$\vartheta(G) = \max\{\langle J, X \rangle : A_G(X) = 0, \text{tr}(X) = 1, X \succeq 0\}.$$

Suppose now that we include the additional constraint $X \geq 0$, leading to an improved approximation $\vartheta^+(G)$ of $\alpha(G)$. This improvement was in fact suggested by Schrijver [Sch79] and independently by [MRJ78]. The condition $A_G(X) = 0$ together with $X \geq 0$ can be simplified to $\langle A, X \rangle = 0$. In other words, the equations $A_G(X) = 0$ are added into just a scalar equation. We get

$$\alpha(G) \leq \vartheta^+(G) = \max\{\langle J, X \rangle : \langle A, X \rangle = 0, \text{tr}(X) = 1, X \succeq 0, X \geq 0\}. \quad (49)$$

The above theorem therefore shows that replacing the cone $\{X : X \succeq 0, X \geq 0\}$ by \mathcal{C}^* leaves no gap in (49).

This suggests to try a similar idea on the dual (25) of $\vartheta(G)$. Looking at the matrix tM with M from (21), it is clear that $M \in \mathcal{C}^*$, therefore we get the following improvement of $\vartheta(G)$ towards $\chi(G)$.

$$\chi(G) \geq \vartheta^{\mathcal{C}}(G) = \min\{t : tI + A_{\bar{G}}(y) \in \mathcal{C}, tI + A_{\bar{G}}(y) \succeq J\} \geq \vartheta(G).$$

It was recently shown in [DR07] that the improvement $\vartheta^{\mathcal{C}}(G)$ is in fact equal to the fractional chromatic number $\chi_f(G)$. Another version to model the chromatic number was recently proposed by Gvozdenovic and Laurent, see [GL08a, GL08b].

These results indicate that the modeling power of copositive programs is stronger than SDP. Burer [Bur08] shows the following general result.

Theorem 14. *Let c and a_j be vectors from \mathbb{R}^n , $b \in \mathbb{R}^k$, $Q \in \mathcal{S}_n$ and $m \leq n$. The optimal values of the following two problems are equal.*

$$\min\{x^T Q x + c^T x : a_j^T x = b_j, x \geq 0, x_i \in \{0, 1\} \forall 1 \leq i \leq m\},$$

$$\min\{\text{tr}(QX) + c^T x : a_j^T x = b_j, a_j^T X a_j = b_j^2, X_{ii} = x_i \forall i \leq m, \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \in \mathcal{C}^*\}.$$

This result shows that it would be extremely interesting to have a better understanding of the cone of completely positive matrices. Outer approximations of \mathcal{C}^* , or equivalently, inner approximations of \mathcal{C} would result in relaxations of the underlying optimization problem. First systematic attempts in this direction were proposed by Parrilo [Par00] and De Klerk and Pasechnik [dKP02] who introduced hierarchies of relaxations based on sum of squares relaxations of polynomials. These relaxations are formulated as SDP of increasing dimension. A summary of this approach, which is far beyond the scope of this article, can be found in [LR05].

Inner approximations of \mathcal{C}^* can be used as starting points for primal heuristics to combinatorial optimization problems. This is an area open for current research.

References

- [ACC06] S. Arora, E. Chlamtac, and M. Charikar. New approximation guarantee for chromatic number. In *Proceedings of the 38th STOC, Seattle, USA*, pages 215–224, 2006.
- [AU03] D. Avis and J. Umemoto. Stronger linear programming relaxations for max-cut. *Mathematical Programming*, 97:451–469, 2003.
- [AW00] K.M. Anstreicher and H. Wolkowicz. On Lagrangian relaxation of quadratic matrix constraints. *SIAM J. on Matrix Analysis*, 22:41–55, 2000.
- [BCC93] E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Math. Program., Ser. A*, 58(3):295–324, 1993.
- [Ber74] S. Berkowitz. Extrema of elementary symmetric polynomials of the eigenvalues of the matrix $P^*KP + L$. *Linear Algebra Appl.*, 8:273–280, 1974.
- [BJR89] F. Barahona, M. Jünger, and G. Reinelt. Experiments in quadratic 0-1 programming. *Mathematical Programming*, 44:127–137, 1989.
- [BTN01] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization*. MPS-SIAM Series on Optimization, 2001.
- [Bur08] S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. Technical report, University of Iowa, 2008.
- [Chl09] E. Chlamtac. *Non-local Analysis of SDP based approximation algorithms*. PhD thesis, Princeton University, USA, 2009.
- [DFJ54] G.B. Dantzig, D.R. Fulkerson, and S. Johnson. Solution of a large scale traveling salesman problem. *Journal of the Operations Research Society of America*, 2:393–410, 1954.
- [DGL93] M. Deza, V.P. Grishukhin, and M. Laurent. The hypermetric cone is polyhedral. *Combinatorica*, 13:397–411, 1993.
- [DH73] W.E. Donath and A.J. Hoffman. Lower bounds for the partitioning of graphs. *IBM J. of Research and Development*, 17:420–425, 1973.
- [dKP02] E. de Klerk and D.V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization*, 12:875–892, 2002.
- [dKPW04] E. de Klerk, D.V. Pasechnik, and J.P. Warners. On approximate graph colouring and MAX- k -CUT algorithms based on the ϑ -function. *J. Comb. Optim.*, 8(3):267–294, 2004.
- [DR07] I. Dukanovic and F. Rendl. Semidefinite programming relaxations for graph coloring and maximal clique problems. *Mathematical Programming*, 109:345–365, 2007.
- [Duf56] R.J. Duffin. Infinite programs. *Ann. Math. Stud.*, 38:157–170, 1956.
- [FGRS06] I. Fischer, G. Gruber, F. Rendl, and R. Sotirov. Computational experience with a bundle method for semidefinite cutting plane relaxations of max-cut and equipartition. *Mathematical Programming*, 105:451–469, 2006.
- [FJ97] A. Frieze and M. Jerrum. Improved approximation algorithms for MAX k -cut and MAX BISECTION. *Algorithmica*, 18(1):67–81, 1997.
- [GL08a] Nebojša Gvozdenović and Monique Laurent. Computing semidefinite programming lower bounds for the (fractional) chromatic number via block-diagonalization. *SIAM Journal on Optimization*, 19(2):592–615, 2008.
- [GL08b] Nebojša Gvozdenović and Monique Laurent. The operator Ψ for the chromatic number of a graph. *SIAM Journal on Optimization*, 19(2):572–591, 2008.
- [GOE97] M. X. GOEMANS. Semidefinite programming in combinatorial optimization. *Mathematical Programming*, 79:143–162, 1997.
- [GW95] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.
- [Ham65] P.L. Hammer. Some network flow problems solved with pseudo-Boolean programming. *Operations Research*, 13:388–399, 1965.

- [Hel00] C. Helmberg. Fixing variables in semidefinite relaxations. *SIAM J. Matrix Anal. Appl.*, 21(3):952–969, 2000.
- [Hel02] C. Helmberg. Semidefinite programming. *European Journal of Operational Research*, 137:461–482, 2002.
- [Hel03] C. Helmberg. Numerical validation of SBmethod. *Mathematical Programming*, 95:381–406, 2003.
- [HKR98] C. Helmberg, K.C. Kiwiel, and F. Rendl. Incorporating inequality constraints in the spectral bundle method. In E.A. Boyd R.E. Bixby and R.Z. Rios-Mercado, editors, *Integer Programming and combinatorial optimization*, pages 423–435. Springer Lecture Notes 1412, 1998.
- [HO00] C. Helmberg and F. Oustry. Bundle methods to minimize the maximum eigenvalue function. In R. Saigal H. Wolkowicz and L. Vandenberghe, editors, *Handbook of semidefinite programming: theory, algorithms and applications*, pages 307–337. Kluwer, 2000.
- [HPRW95] C. Helmberg, S. Poljak, F. Rendl, and H. Wolkowicz. Combining semidefinite and polyhedral relaxations for integer programs. In E. Balas and J. Clausen, editors, *Integer Programming and combinatorial optimization*, pages 124–134. Springer Lecture Notes 920, 1995.
- [HR98] C. Helmberg and F. Rendl. Solving quadratic (0,1)-problems by semidefinite programming and cutting planes. *Mathematical Programming*, 82:291–315, 1998.
- [HR00] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10:673–696, 2000.
- [HRVW96] C. Helmberg, F. Rendl, R. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM Journal on Optimization*, 6:342–361, 1996.
- [HRW92] S.W. Hadley, F. Rendl, and H. Wolkowicz. A new lower bound via projection for the quadratic assignment problem. *Mathematics of Operations Research*, 17:727–739, 1992.
- [HUL93] J.-B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and minimization algorithms (vol. 1 and 2)*. Springer, 1993.
- [HW53] A.J. Hoffman and H.W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Math. J.*, 20:37–39, 1953.
- [HZ01] E. Halperin and U. Zwick. A unified framework for obtaining improved approximation algorithms for maximum graph bisection problems. In *Lecture notes in Computer Science 2081, IPCO 2001*, pages 210–225. 2001.
- [JAMS89] D.S. Johnson, C.R. Aragon, L.A. McGeoch, and C. Schevon. Optimization by simulated annealing: An experimental evaluation. I: Graph partitioning. *Oper. Res.*, 37(6):865–892, 1989.
- [JL05] Dorina Jibeteian and Monique Laurent. Semidefinite approximations for global unconstrained polynomial optimization. *SIAM Journal on Optimization*, 16(2):490–514, 2005.
- [KL70] B.W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System techn. J.*, 49:291–307, 1970.
- [KLS00] S. Khanna, N. Linial, and S. Safra. On the hardness of approximating the chromatic number. *Combinatorica*, 20:393–415, 2000.
- [KMS98] D. Karger, R. Motwani, and M. Sudan. Approximate graph colouring by semidefinite programming. *Journal of the ACM*, 45:246–265, 1998.
- [KR98] S.E. Karisch and F. Rendl. Semidefinite Programming and Graph Equipartition. *Fields Institute Communications*, 18:77–95, 1998.
- [KSH97] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, 7(1):86–125, 1997.
- [Las06] Jean B. Lasserre. A sum of squares approximation of nonnegative polynomials. *SIAM J. Optim.*, 16(3):751–765, 2006.

- [Lem75] C. Lemarechal. An extension of davidon methods to nondifferentiable problems. *Mathematical Programming Study*, 3:95–109, 1975.
- [Lem78] C. Lemarechal. Nonsmooth optimization and descent methods. Technical report, International Institute for Applied Systems Analysis, 1978.
- [LLRKS85] E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, and D.B. (eds.) Shmoys. *The traveling salesman problem, a guided tour of combinatorial optimization*. 1985.
- [Lov79] L. Lovász. On the shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:1–7, 1979.
- [Lov03] L. Lovr̄masz. Semidefinite programs and combinatorial optimization. In B. A. Reed and C.L. Sales, editors, *Recent advances in algorithms and combinatorics*, pages 137–194. CMS books in Mathematics, Springer, 2003.
- [LPR97] 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.
- [LR02] A. Lisser and F. Rendl. Graph partitioning using Linear and Semidefinite Programming. *Mathematical Programming (B)*, 95:91–101, 2002.
- [LR05] M. Laurent and F. Rendl. Semidefinite programming and integer programming. In K. Aardal, G.L. Nemhauser, and R. Weismantel, editors, *Discrete Optimization*, pages 393–514. Elsevier, 2005.
- [LS91] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.
- [LY93] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. In *Proceedings of the 25th ACM STOC*, pages 286–293, 1993.
- [Mar75] M. Marcus. Rearrangement and extremal results for Hermitian matrices. *Linear Algebra Appl.*, 11:95–104, 1975.
- [MK87] K.G. Murty and S.N. Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.
- [Mon97] R.D.C. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7:663–678, 1997.
- [MRJ78] R.J. McEliece, E.R. Rodemich, and H.C. Rumsey Jr. The lovasz bound and some generalizations. *Journal of combinatorics and System Sciences*, 3:134–152, 1978.
- [MS65] T.S. Motzkin and E.G. Strauss. Maxima for graphs and a new proof of a theorem of turan. *Canadian Journal of Mathematics*, 17:533–540, 1965.
- [Nes97] Y. Nesterov. Quality of semidefinite relaxation for nonconvex quadratic optimization. Technical report, 1997.
- [Neu37] J.v. Neumann. Some matrix inequalities and metrization of matrix space. pages 205–219. reprinted in: John von Neumann: Collected Works, Vol 4, MacMillan 1962, 1937.
- [NN94] Y. Nesterov and A.S. Nemirovski. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM Publications. SIAM, Philadelphia, USA, 1994.
- [Pad89] M. Padberg. The quadric Boolean polytope: some characteristics, facets and relatives. *Mathematical Programming*, 45:139–172, 1989.
- [Par00] P. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, USA, 2000.
- [PR08] J. Povh and F. Rendl. Approximating non-convex quadratic programs by semidefinite and copositive programming. In L. Neralic V. Boljuncic and K.Soric, editors, *Proceedings of the 11th international conference on operational research*, pages 35–45. Croatia Operations Research Society, 2008.
- [RW92] F. Rendl and H. Wolkowicz. Applications of parametric programming and eigenvalue maximization to the quadratic assignment problem. *Math. Program., Ser. A*, 53(1):63–78, 1992.
- [RW95] F. Rendl and H. Wolkowicz. A projection technique for partitioning the nodes of a graph. *Annals of Operations Research*, 58:155–179, 1995.

- [SA90] H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.
- [SA94] H. D. Sherali and W. P. Adams. A hierarchy of relaxations and convex hull characterizations for mixed-integer zero-one programming problems. *Discrete Appl. Math.*, 52(1):83–106, 1994.
- [Sch79] A. Schrijver. A comparison of the deza and lovasz bounds. *IEEE Transactions on Information Theory*, IT-25:425–429, 1979.
- [Sim90] C. De Simone. The cut polytope and the Boolean quadric polytope. *Discrete Math.*, 79(1):71–75, 1990.
- [SZ92] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. *SIAM J. Optimization*, 2:121–152, 1992.
- [Wid83] A. Widgerson. Improving the performance guarantee for approximate graph colouring. *Journal of the ACM*, 30:729–735, 1983.
- [WSe00] H. Wolkowicz, R. Saigal, and L. Vandenberghe (eds.). *Handbook of semidefinite programming*. Kluwer, 2000.