Technische Universität Graz Institut für Mathematik
Christian-Doppler-Laboratorium „Diskrete Optimierung“

An Interior-Point Method for Semidefinite Programming

Christoph Helmberg
Franz Rendl
Henry Wolkowicz
Robert J. Vanderbei

Report 264 June 1994 CDLDO-24

Technische Universität Graz, Steyrergasse 30, A-8010 Graz, Austria

Version 2
This is an updated version of Report CDLDO - 24 from August 1993
Title page processed by TEX on June 6, 1994

An Interior-Point Method for Semidefinite Programming

Christoph Helmberg * Franz Rendl † Robert J. Vanderbei ‡
Henry Wolkowicz §

June 6, 1994

Program in Statistics & Operations Research
Princeton University
Princeton, NJ 08544

June 6, 1994

This paper is a revision of the technical report SOR-93-15, which appeared under the title *Interior-Point Methods for Max-Min Eigenvalue problems*

Abstract

We propose a new interior point based method to minimize a linear function of a matrix variable subject to linear equality and inequality constraints over the set of positive semidefinite matrices. We present a theoretical convergence analysis, and show that the approach is very efficient for graph bisection problems, such as max-cut. The approach can also be applied to max-min eigenvalue problems.

Key words: semidefinite programming, interior-point methods, max-cut relaxations, max-min eigenvalue problems.

AMS 1991 Subject Classification: Primary 65F15, Secondary 49M35, 90C48

---

*Technische Universität Graz, Institut für Mathematik, Kopernikusgasse 24, A-8010 Graz, Austria. Research support by Christian Doppler Laboratorium für Diskrete Optimierung.

†Technische Universität Graz, Institut für Mathematik, Kopernikusgasse 24, A-8010 Graz, Austria. Research support by Christian Doppler Laboratorium für Diskrete Optimierung.

‡Research support by AFOSR through grant AFOSR-91-0359.

§Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ont., Canada. This author thanks the Department of Civil Engineering and Operations Research, Princeton University, for their support during his stay while on research leave. Thanks also to the National Science and Engineering Research Council Canada for their support.

1 Introduction.

The continuously rising success of interior point techniques applied to Linear Programming has stimulated research in various related fields. One possible line of generalization consists in looking at linear programs over non-polyhedral cones. This type of generalization is studied in the present paper.

To be specific, let $\mathcal{M}_n$ denote the vector space of symmetric $n \times n$ matrices. Suppose $A : \mathcal{M}_n \mapsto \Re^k$, $B : \mathcal{M}_n \mapsto \Re^m$ are two linear operators and $C \in \mathcal{M}_n, a \in \Re^k, b \in \Re^m$.

We study the following optimization problem:

$$\text{(SDP)} \qquad \begin{array}{ll} \text{maximize} & \text{tr } CX \\ \text{subject to} & a - A(X) = 0 \\ & b - B(X) \geq 0 \\ & X \succeq 0. \end{array}$$

This is a *semidefinite linear program*, because we optimize a linear function subject to linear inequality and equality constraints over positive semidefinite matrices X.

Our main motivation to study this kind of problem comes from applications in discrete optimization. In particular, we will investigate a new, powerful, and tractable relaxation of the *max-cut* problem as well as other graph bisection problems. Also *max-min eigenvalue problems* over an affine parameter space fall into the present framework and can be handled with our approach.

The main contribution of the paper is the following: we will propose a primal-dual interior point algorithm for problem (SDP) and present a theoretical convergence analysis, showing global convergence. Moreover, we will discuss implementation details and present some computational experiments indicating that the approach is also highly efficient in practice.

We close this section by describing research related to our work. Alizadeh, Haeberly, Jarre and Overton [2, 3, 4, 13, 14, 18, 19] consider a problem similar to ours. Their models allow only equality constraints, and no inequalities. Algorithmically, these authors use mostly interior point based techniques to solve the problem. Alizadeh proposes a potential reduction method and shows a polynomial running time to find an ϵ-optimal solution. Jarre uses a barrier approach and works directly on the dual. Our approach is closest to the one of Jarre, but we use a primal-dual variant, and include also linear inequalities. Finally Overton [18, 19] studies the problem under nonlinear equality constraints and again without inequalities. The formulations in [13, 18, 19, 4] are not in the form above, but it is an easy exercise to transform them into our model.

1.1 Preliminaries

We first collect some preliminary results and notation. We work mainly in the space $\mathcal{M}_n$ of symmetric $n \times n$ matrices, endowed with *inner product*

$$\langle U, V \rangle := \text{tr}(UV^T).$$

The curly inequality symbol refers to the *Löwner partial order* induced by the cone of positive semidefinite matrices; i.e., $M_1 \preceq M_2$ ($M_1 \prec M_2$) means that $M_2 - M_1$ is positive semidefinite (positive definite, respectively). In contrast, the usual inequality symbol, $v \leq w$, refers to the partial order induced by the cone of nonnegative vectors; The *maximum eigenvalue* of M is denoted by $\lambda_{\max}(M)$. A similar notation is used for the minimum eigenvalue.

Associated with any linear operator $A : \mathcal{M}_n \mapsto \Re^k$ is another linear operator, which we denote by A^T and which is defined by the adjoint relation

$$\langle A(X), y \rangle = \langle X, A^T(y) \rangle, \quad \text{for all } X \in \mathcal{M}_n, y \in \Re^k.$$

Here we have used the angle bracket both for the inner product in $\mathcal{M}_n$ and in $\Re^k$.

For any convex cone S in R^n, we let S^* denote the *dual cone*; i.e.,

$$S^* = \{y \in \Re^n : x^T y \geq 0 \quad \text{for all } x \in S\}.$$

In a slight abuse, given a real function function f defined on a real domain and a vector v in $\Re^n$, we will occasionally write $f(v)$ for $(f(v_1), \ldots, f(v_n))$. Here, f will usually be the logarithm or a power function.

Finally, for X in $\mathcal{M}_n$, we let $\mathrm{diag}(X)$ denote the vector in $\Re^n$ consisting of the diagonal elements of X. Analogously, for a vector x in $\Re^n$, we let $\mathrm{Diag}(x)$ denote the diagonal matrix in $\mathcal{M}_n$ whose diagonal elements are obtained from x.

2 Duality

The general duality theory for problems such as (SDP) has been thoroughly studied, see e.g. [27]. It can be summarized as follows.

Suppose S and T are two closed convex cones in R^n and R^m, respectively. Let $c \in \Re^n, b \in \Re^m$ and the $m \times n$ matrix A be given. We define the following two optimization problems (P) and (D).

$$(P) \quad \begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & Ax - b \in S \\ & x \in T, \end{array}$$

$$(D) \quad \begin{array}{ll} \text{minimize} & b^T y \\ \text{subject to} & c - A^T y \in T^* \\ & y \in S^*. \end{array}$$

We tacitly assume that both problems have feasible solutions. If a so-called constraint qualification holds, then it can be shown that both problems form a pair of dual problems and strong duality holds; i.e., the minimum attained in (P) coincides with the maximum attained in (D), see e.g. [27]. (Weak duality, the max being less than or equal to the min,

always holds and is trivial to prove.) If both S and T are polyhedral, then we have the standard Linear Programming duality.

In the present paper, we consider the situation where T corresponds to the cone of positive semidefinite matrices, while S is a polyhedral cone. Indeed, since $\mathcal{M}_n$ is an $n(n+1)/2$ dimensional vector space, we can identify a matrices in $\mathcal{M}_n$ with corresponding vectors in $\Re^{n(n+1)/2}$. Making this identification and letting T denote the cone of positive semidefinite matrices and S the polyhedral cone

$$S = (0, \Re^m_+)$$

(representing the vector equalities and inequalities in (SDP)), we see that (SDP) can be rephrased in the formulation above.

We now write down the dual to (SDP). Noting that $S^* = (\Re^k, \Re^m_+)$ and $T^* = T$, we get

$$\begin{array}{lll} & \text{minimize} & a^T y + b^T t \\ \text{(DSDP)} & \text{subject to} & A^T(y) + B^T(t) - C \succeq 0 \\ & & y \in \Re^k, \quad t \in \Re^m_+. \end{array}$$

Existence of an optimal solution and strong duality cannot be guaranteed unless some additional assumptions, such as Slater's constraint qualification, are made. (A modification of the dual allows strong duality in the general setting, see [27].)

In our applications, we will focus mostly on the special case where $A(X) = \text{diag}(X)$, in which case $A^T(y) = \text{Diag}(y)$.

We conclude this section by showing that the dual (DSDP) could be quite easily derived directly using Lagrangian methods. Indeed, let ω^* denote the optimal objective value for (SDP). Introducing Lagrange multipliers $y \in \Re^k$ and $t \in \Re^m_+$ for the equality and inequality constraints, respectively, we see that

$$\begin{array}{rcll} \omega^* & = & \max\limits_{X \succeq 0} \min\limits_{t \geq 0, y} & \text{tr } CX + y^T(a - A(X)) + t^T(b - B(X)) \\ & \leq & \min\limits_{t \geq 0, y} \max\limits_{X \succeq 0} & \text{tr}(C - A^T(y) - B^T(t))X + a^T y + b^T t. \end{array}$$

Now note that the inner maximization over X is bounded from above only if

$$A^T(y) + B^T(t) - C \succeq 0.$$

In this case the maximum occurs if complementarity holds, i.e.

$$\text{tr}(C - A^T(y) - B^T(t))X = 0.$$

Thus we get the (weak) dual

$$\omega^* \leq \min_{\substack{A^T(y) + B^T(t) - C \succeq 0 \\ t \geq 0}} a^T y + b^T t.$$

This is precisely (DSDP). Again, the duality gap from interchanging max and min vanishes only under additional assumptions on the Lagrangian.

Before developing an algorithm for (SDP), we first show in the next section that this type of problem provides strong machinery for deriving tight bounds to several basic NP-hard optimization problems.

3 Applications

3.1 Max-Cut Problem

The max-cut problem is the problem of partitioning the node set of an edge-weighted undirected graph into two parts so as to maximize the total weight of edges *cut* by the partition. We tacitly assume that the graph in question is complete (if not, nonexisting edges can be given weight 0 to complete the graph). Mathematically, the problem can be formulated as follows (see e.g [16]). Let the graph be given by its weighted adjacency matrix A. Define the matrix $L := \mathrm{Diag}(Ae) - A$, where e is the vector of all ones. (The matrix L is called the *Laplacian matrix* associated with the graph.) If a cut S is represented by a vector x where $x_i \in \{-1, 1\}$ depending on whether or not $i \in S$, we get the following formulation for the max-cut problem.

$$\text{(MC)} \qquad \begin{array}{ll} \text{maximize} & \frac{1}{4} x^T L x \\ \text{subject to} & x \in \{-1, 1\}^n. \end{array}$$

Using $X := \frac{1}{4} x x^T$, this is equivalent to

$$\begin{array}{ll} \text{maximize} & \text{tr } LX \\ \text{subject to} & \mathrm{diag}(X) = \frac{1}{4} e \\ & \mathrm{rank}(X) = 1 \\ & X \succeq 0. \end{array}$$

Dropping the rank condition we obtain a problem of the form (SDP) with no inequalities, $a = \frac{1}{4} e$ and $A(X) = \mathrm{diag}(X)$.

This relaxation of max-cut is well known and studied e.g. in [9, 11, 21]. Goemans and Williamson [11] have recently shown that the optimal value of this relaxation is at most 14% above the value of the maximum cut, provided $A \geq 0$, i.e. no negative edge weights exist.

The variable X can be interpreted as being defined on the edge set of the (complete) graph. Therefore we can add further linear constraints, that are satisfied by all edge vectors representing cuts. One such class of constraints is obtained by the following trivial observation. Consider an arbitrary triangle with vertices $i < j < k$ in the graph G. Then any partition cuts either 0 or 2 of its edges. Translated into our model this leads to

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_{ij} \\ x_{ik} \\ x_{jk} \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \geq 0.$$

(Note that in our model an edge is cut if $x_i x_j = -1$, thus $x_{ij} = -\frac{1}{4}$.) The first constraint states that at most 2 edges can be cut. The other 3 constraints state that if one edge is cut, then there must be at least one other edge which must lie in the cut. Let us now collect these constraints for all triangles in the operator B. Then this leads to

$$B(X) - b \geq 0,$$

where $\dim(B) = 4\binom{n}{3}$ and b is a vector. Thus we get the following stronger relaxation for max-cut, first proposed in [21]

$$\begin{array}{rl} \text{maximize} & \text{tr } LX \\ \text{subject to} & \text{diag}(X) - a = 0 \\ & B(X) - b \geq 0 \\ & X \succeq 0. \end{array}$$

Dropping the semidefiniteness constraint on X we obtain an ordinary Linear Program which is again a relaxation for max-cut. This relaxation is usually called the *Metric Relaxation*, because the polyhedron

$$\{X : B(X) - b \geq 0, \text{diag}(X) = a\}$$

is often referred to as the *metric polytope*, see e.g. [15]. We point out that this LP has $\binom{n}{2}$ variables and roughly $\frac{2}{3}n^3$ (very sparse) constraints. This polyhedron turns out to be highly degenerate, so that it is still considered a computational challenge to optimize an arbitrary linear function over this polytope for say $n \approx 40$. (If the graph is planar, then the metric relaxation provides already the max-cut, see e.g. [5].)

3.2 Graph Bisection

Graph bisection is similar to the max-cut problem, but here we seek a partition (S, T) of the node set V such that the two sets have prespecified cardinalities, say $|S| = k$ and $|T| = n - k$ for some given k. An important special case occurs if $k = n/2$. In this case, one looks for a partition of the node set into to sets of equal size, so as to minimize the weight of the cut (this problem is usually formulated as a minimization rather than a maximization problem.) The additional cardinality constraint $|S| = n/2$ translates into $\sum_i x_i = 0$. Thus we have

$$\text{(BS)} \qquad \begin{array}{rl} \text{minimize} & \frac{1}{4}x^T Lx \\ \text{subject to} & x_i \in \{-1, 1\} \\ & e^T x = 0. \end{array}$$

In analogy with max-cut, we obtain the following relaxation:

$$\begin{array}{rl} \text{minimize} & \text{tr } LX \\ \text{subject to} & \text{diag}(X) = a \\ & \text{tr } XJ = 0 \\ & X \succeq 0. \end{array}$$

Here $J = ee^T$ is the matrix of all ones. Note that the constraint tr $JX = 0$ is obtained by squaring the cardinality constraint: $0 = (e^T x)^2 = \text{tr } JX$. This relaxation was also studied in [10] where it was treated as a min-max eigenvalue problem using nonsmooth optimization techniques.

3.3 Min-Max Eigenvalue Problems

The following min-max eigenvalue problem is studied, e.g., in [13, 18, 19]:

$$\lambda^* := \min_{y \in \Re^k} \lambda_{\max}(C + A(y)).$$

Here, $C \in \mathcal{M}_n$ and $A : \Re^k \mapsto \mathcal{M}_n$ is a linear operator. Overton [18, 19] actually considers a more general case allowing A to be nonlinear. It is well known that this problem can be reformulated as an (SDP) problem:

$$\begin{array}{ll} \text{minimize} & \lambda \\ \text{subject to} & \lambda I - C - A(y) \succeq 0 \\ & y \in \Re^k, \quad \lambda \in \Re. \end{array}$$

The dual is

$$\begin{array}{ll} \text{maximize} & \text{tr } CX \\ \text{subject to} & A^T(X) = 0 \\ & \text{tr } X = 1 \\ & X \succeq 0. \end{array}$$

Note that complementary slackness at optimality implies that $ZX = 0$, with $Z := \lambda I - C - A(y)$. Dual feasibility then implies that the eigenvectors for the optimal eigenvalue are found as the columns of X. If strict complementary slackness holds, i.e. $\text{rank}(X) + \text{rank}(Z) = n$, then $\text{rank}(X)$ equals the multiplicity of the optimal eigenvalue.

Finally, applications to other combinatorial optimization problems are described in [3]. These include among others the stable set problem and the computation of Lovasz's theta function.

4 Interior-Point Method for SDP

In this section we will develop a primal-dual interior point method that solves (SDP) and (DSDP) simultaneously. The nature of this approach requires that there exists an X strictly satisfying the inequalities of the primal problem, i.e.

$$b - B(X) > 0 \quad \text{and} \quad X \succ 0.$$

Furthermore we assume without loss of generality that the equality constraints on X are linearly independent, $\text{rank}(A(\cdot)) = k$. Since we will have to apply operators A and B to nonsymmetric matrices as well, we extend their definition by mapping the skew-symmetric part to zero. This implies

$$A(M) = A(M^T) \quad \text{and} \quad B(M) = B(M^T). \tag{4.1}$$

We follow the usual derivation of primal-dual interior point methods in Linear Programming and first introduce the associated *barrier problem* for (DSDP), which we call the *dual*

barrier problem:

$$\begin{array}{rll} & \text{minimize} & a^T y + b^T t - \mu(\log\det Z + e^T \log t) \\ \text{(DBP)} & \text{subject to} & A^T(y) + B^T(t) - C = Z \\ & & t \geq 0, Z \succeq 0. \end{array} \tag{4.2}$$

Here μ is a positive real number called the *barrier parameter*. We denote the objective function of (DBP) by f_D, i.e.

$$f_D(y,t,Z) := a^T y + b^T t - \mu(\log\det Z + e^T \log t).$$

For each $\mu > 0$, there is a corresponding Lagrangian:

$$\begin{array}{rl} L_\mu(X,y,t,Z) = & a^T y + b^T t - \mu(\log\det Z + e^T \log t) + \\ & \left\langle Z + C - A^T(y) - B^T(t), X \right\rangle. \end{array} \tag{4.3}$$

The first-order optimality conditions for the saddle point of this Lagrangian are obtained easily using the adjoint identity for A and B:

$$\nabla_X L_\mu = Z + C - A^T(y) - B^T(t) = 0 \tag{4.4}$$

$$\nabla_y L_\mu = a - A(X) = 0 \tag{4.5}$$

$$\nabla_t L_\mu = b - B(X) - \mu t^{-1} = 0 \tag{4.6}$$

$$\nabla_Z L_\mu = X - \mu Z^{-1} = 0. \tag{4.7}$$

The strict concavity of $\log\det Z$ and $\log t_i$ implies that there exists a unique solution $(X_\mu, y_\mu, t_\mu, Z_\mu)$ to these optimality conditions. The one-parameter family $\{(X_\mu, y_\mu, t_\mu, Z_\mu) : 0 \leq \mu \leq \infty\}$ is called the *central trajectory*. Given a point (X,y,t,Z) on the central trajectory it is easy to determine its associated μ value using (4.6) and/or (4.7):

$$\mu = \frac{\mathrm{tr}(ZX)}{n} = \frac{t^T(b - B(X))}{m} = \frac{\mathrm{tr}(ZX) + t^T(b - B(X))}{n+m}. \tag{4.8}$$

We note that if the point is a feasible solution of the primal and the dual problem, $\mathrm{tr}(ZX) + t^T(b - B(X))$ is the gap between the primal and the dual objective value. We shall use (4.8) to associate μ values with quadruples (X,y,t,Z) even when these quadruples don't belong to the central trajectory.

Our interior-point algorithm is derived as follows. We start with a quadruple (X,y,t,Z) for which $X \succ 0$, $Z \succ 0$, $t > 0$, and $b - B(X) > 0$ but which is otherwise arbitrary. From this point we estimate the current μ value using (4.8) and divide it by two:

$$\mu = \frac{\mathrm{tr}(ZX) + t^T(b - B(X))}{2(n+m)}. \tag{4.9}$$

(Experience from linear programming indicates that this simple heuristic performs very well, even though it does not guarantee monotonic decrease in μ, see [25].) We next attempt to find step directions $(\Delta X, \Delta y, \Delta t, \Delta Z)$ such that the new point $(X + \Delta X, y + \Delta y, t +$

$\Delta t, Z + \Delta Z)$ lies on the central trajectory at this value of μ. However, since not all the defining equations, (4.4)–(4.7), are linear, it is not possible to solve this system directly. In fact, only (4.6) and (4.7) are nonlinear. They can be written in several equivalent forms, each form giving rise to a different linearization. We will discuss this for (4.7).

$$\mu I - Z^{1/2} X Z^{1/2} = 0 \tag{4.10}$$

$$\mu I - X^{1/2} Z X^{1/2} = 0 \tag{4.11}$$

$$\mu Z^{-1} - X = 0 \tag{4.12}$$

$$\mu X^{-1} - Z = 0 \tag{4.13}$$

$$ZX - \mu I = 0 \tag{4.14}$$

$$ZX + XZ - 2\mu I = 0. \tag{4.15}$$

The first two linearizations are very popular in linear programming, in which case X and Z are diagonal matrices. The two forms are then the same and don't really involve square roots. We rule these out, since they involve matrix square roots. The linearization of the third form leads to Jarre's algorithm [13], which solves the problem from the dual side only. Analogously, linearization of (4.13) solves the problem from the primal side. This form is also used in [22]. Linearizations of (4.14) and (4.15) contain the information of the current primal and dual solution at an equal degree and are especially well suited for Mehrotra's LP predictor–corrector method as described in [7]. Linearization of (4.15) preserves symmetry but the use of Kronecker products cannot be avoided [4]. Linearization of (4.14) violates symmetry but only ΔX will not be symmetric and by taking the symmetric part of ΔX we will show that we still get a descent direction.

We will use the linearization in (4.14). For simplicity of notation, we rewrite (4.4) to (4.7) as the function

$$F_\mu(s) = F_\mu(X, y, t, Z) := \begin{pmatrix} Z + C - A^T(y) - B^T(t) \\ a - A(X) \\ t \circ (b - B(X)) - \mu e \\ ZX - \mu I \end{pmatrix} =: \begin{pmatrix} F_d \\ F_p \\ F_{tB} \\ F_{ZX}. \end{pmatrix} \tag{4.16}$$

The notation $u \circ v$ denotes the Hadamard product $(u \circ v)_j = u_j \cdot v_j$ of two vectors u and v. The solution s^* to $F_\mu(s) = 0$ satisfies the Kuhn-Tucker conditions (4.4) to (4.7) and is the optimal solution to the barrier problem. To find a step direction $\Delta s = (\Delta X, \Delta y, \Delta t, \Delta Z)$ toward s^*, we use Newton's Method, which says that Δs must satisfy

$$F_\mu + \nabla F_\mu(\Delta s) = 0.$$

The step direction Δs is therefore the solution of the system

$$\begin{aligned} \Delta Z - A^T(\Delta y) - B^T(\Delta t) &= -F_d & (4.17)\\ -A(\Delta X) &= -F_p & (4.18)\\ \Delta t \circ (b - B(X)) - t \circ B(\Delta X) &= -F_{tB} & (4.19)\\ Z\Delta X + \Delta Z X &= -F_{ZX}. & (4.20) \end{aligned}$$

This linear system can now be solved for $(\Delta X, \Delta y, \Delta t, \Delta Z)$. Indeed, first we solve (4.17) for an obviously symmetric ΔZ (in terms of Δy and Δt),

$$\Delta Z = -F_d + A^T(\Delta y) + B^T(\Delta t), \tag{4.21}$$

and then substitute this expression into (4.20) to get

$$\Delta X = \mu Z^{-1} - X + Z^{-1}F_d X - Z^{-1}(A^T(\Delta y) + B^T(\Delta t))X. \tag{4.22}$$

Evidently, ΔX is not symmetric in general. Substituting this expression for ΔX into (4.18), we get the first equation for Δy and Δt

$$O_{11}(\Delta y) + O_{12}(\Delta t) = v_1, \tag{4.23}$$

where O_{11} and O_{12} are the linear operators defined by

$$O_{11}(\cdot) := A(Z^{-1}A^T(\cdot)X) \tag{4.24}$$
$$O_{12}(\cdot) := A(Z^{-1}B^T(\cdot)X) \tag{4.25}$$

and v_1 is the vector

$$v_1 := \mu A(Z^{-1}) - a + A(Z^{-1}F_d X). \tag{4.26}$$

Finally we substitute (4.22) into (4.19) and get

$$O_{21}(\Delta y) + O_{22}(\Delta t) = v_2, \tag{4.27}$$

where O_{21} and O_{22} are the linear operators defined by

$$O_{21}(\cdot) := B(Z^{-1}A^T(\cdot)X) \tag{4.28}$$
$$O_{22}(\cdot) := (b - B(X)) \circ t^{-1} \circ (\cdot) + B(Z^{-1}B^T(\cdot)X) \tag{4.29}$$

and v_2 is the vector

$$v_2 := \mu t^{-1} - b + \mu B(Z^{-1}) + B(Z^{-1}F_d X). \tag{4.30}$$

Because of (4.1) operators O_{11} and O_{22} are self adjoint and operator O_{12} is the adjoint operator to O_{21}. Equations (4.23) and (4.27) form a symmetric linear system for Δy and Δt. In fact, this system is even positive definite. To show this, we define a new operator O that maps $\mathcal{M}_n$ into $\Re^{k+m}$:

$$O(X) = \begin{pmatrix} A(X) \\ B(X) \end{pmatrix}.$$

The adjoint operator is given by the adjoint identity

$$\left\langle O(X), \begin{pmatrix} y \\ t \end{pmatrix} \right\rangle = \left\langle X, A^T(y) + B^T(t) \right\rangle.$$

The system can now be written as

$$O(Z^{-1}O^T\begin{pmatrix}\Delta y\\ \Delta t\end{pmatrix}X)+\begin{pmatrix}0\\ (b-B(X))\circ t^{-1}\circ(\Delta t)\end{pmatrix}=\begin{pmatrix}v_1\\ v_2\end{pmatrix}. \tag{4.31}$$

Observe that the second summand adds positive coefficients to the main diagonal entries corresponding to the inequality constraints. In fact this increment on the main diagonal is the only difference between inequality and equality constraints. The second summand clearly forms a positive semidefinite operator on $\binom{\Delta y}{\Delta t}$. It is positive definite for all vectors having at least one component $\Delta t_i \neq 0$. For the first summand we get

$$\begin{aligned}&\left\langle O(Z^{-1}O^T(v)X),v\right\rangle && (4.32)\\ &= \left\langle Z^{-1}O^T(v)X,O^T(v)\right\rangle && (4.33)\\ &= \operatorname{tr}\left(Z^{-\frac{1}{2}}O^T(v)X^{\frac{1}{2}}X^{\frac{1}{2}}O^T(v)Z^{-\frac{1}{2}}\right) && (4.34)\\ &= \left\langle Z^{-\frac{1}{2}}O^T(v)X^{\frac{1}{2}},Z^{-\frac{1}{2}}O^T(v)X^{\frac{1}{2}}\right\rangle \geq 0. && (4.35)\end{aligned}$$

Since X and Z^{-1} are positive definite and the equality constraints are linearly independent, equality is possible if and only if $v=0$ or at least one $\Delta t_i \neq 0$. It follows that the system is indeed positive definite. It can be solved efficiently in $O((k+m)^3)$. Observe that an equivalent representation of operator $O(X)$ is given by $(\operatorname{tr}(A_1X),\ldots\operatorname{tr}(A_{k+m}X))^T$ with the A_i being appropriately chosen symmetric matrices. Then the ij-th element of the matrix describing $O(Z^{-1}O^T(\cdot)X)$ reads $\operatorname{tr}(A_iZ^{-1}A_jX)$.

The solution yields a quadruple with $\Delta X'$ not necessarily symmetric. This is mended by using the symmetric part of $\Delta X'$ only:

$$\Delta X=\frac{\Delta X'^T+\Delta X'}{2}. \tag{4.36}$$

To summarize, we solve for the quadruple $\Delta s=(\Delta X,\Delta y,\Delta t,\Delta Z)$ by first solving (4.23) and (4.27) for Δy and Δt and then substituting this into (4.21) to solve for ΔZ and finally substituting that into (4.20) to solve for $\Delta X'$ of which we take the symmetric part only. In section 5 we will show that this indeed yields a descent direction.

Having determined the desired quadruple, $(\Delta X,\Delta y,\Delta t,\Delta Z)$, of step directions, we would step to the new quadruple $(X+\Delta X,y+\Delta y,t+\Delta t,Z+\Delta Z)$ except that it might violate the nonnegativity of t and $b-B(X)$ and the positive definiteness property required of the two matrices. Hence, we perform a line search to find constants α_p and α_d such that $t+\alpha_d\Delta t$ and $b-B(X+\alpha_p\Delta X)$ are strictly positive and $X+\alpha_p\Delta X$ and $Z+\alpha_d\Delta Z$ are positive definite. Given α_p and α_d, we step to the new point

$$\begin{aligned}X+\alpha_p\Delta X\\ y+\alpha_d\Delta y\\ t+\alpha_d\Delta t\\ Z+\alpha_d\Delta Z.\end{aligned}$$

We update μ using (4.9) and repeat. The algorithm continues until the current quadruple (X,y,t,Z) satisfies primal feasibility, dual feasibility and the duality gap is sufficiently small. This completes the description of our interior-point algorithm.

5 Descent Direction and Convergence

In this section, we will prove that the step direction $(\Delta X, \Delta y, \Delta t, \Delta Z)$ indeed forms a descent direction with respect to an appropriately defined merit function. By assuming the compactness of the set of primal feasible solutions and a strictly feasible starting point we are able to prove convergence of the method. It should be noted that we cannot make use of the theory of Nesterov and Nemirovskii [17] since the descent direction is not the Newton direction of the underlying problem. The line of argument follows a proof of Anstreicher and Vial [1, 26] for a convex programming problem without semidefinite variables.

For fixed μ, we measure the progress of the algorithm using the convex merit function

$$\begin{aligned} f_\mu(X,y,t,Z) &= \langle Z, X\rangle - \mu \log\det(XZ) + \\ &\quad t^T(b - B(X)) - \mu e^T \log(t \circ (b - B(X))) + \\ &\quad \frac{1}{2}||F_p||^2 + \frac{1}{2}||F_d||^2 \end{aligned} \tag{5.1}$$

As the minimum of $(x - \mu \log x)$ for $x > 0$ is attained at $x = \mu$, function f_μ is bounded below by $(n+m)\mu(1 - \log\mu)$.

$$f_\mu \geq (n+m)\mu(1 - \log(\mu)) := c_\mu. \tag{5.2}$$

We note that $F_\mu(s) = 0 \Longleftrightarrow f_\mu = (n+m)\mu(1-\log\mu)$. f_μ is continuously differentiable on the interior of the feasible sets and grows towards infinity on the boundary. The stepsize will be chosen according to this function. In the following Lemma we prove that Δs (as defined on page 10) is a descent direction of f_μ.

Lemma 1 *The directional derivative of f_μ in the direction Δs satisfies*

$$\left\langle \frac{\partial f_\mu}{\partial s}, \Delta s \right\rangle \leq 0 \tag{5.3}$$

with equality holding if and only if $F_\mu(s) = 0$.

Proof. We have to prove that

$$\left\langle \frac{\partial f_\mu}{\partial X}, \Delta X \right\rangle + \left\langle \frac{\partial f_\mu}{\partial y}, \Delta y \right\rangle + \left\langle \frac{\partial f_\mu}{\partial t}, \Delta t \right\rangle + \left\langle \frac{\partial f_\mu}{\partial Z}, \Delta Z \right\rangle \leq 0$$

The partial derivatives of f_μ are

$$\begin{aligned} \frac{\partial f_\mu}{\partial X} &= F_p^T \frac{\partial F_p}{\partial X} + F_d^T \frac{\partial F_d}{\partial X} + Z - \mu X^{-1} - B^T(t) - \mu B^T((b - B(X))^{-1}) \\ \frac{\partial f_\mu}{\partial y} &= F_p^T \frac{\partial F_p}{\partial y} + F_d^T \frac{\partial F_d}{\partial y} \\ \frac{\partial f_\mu}{\partial t} &= F_p^T \frac{\partial F_p}{\partial t} + F_d^T \frac{\partial F_d}{\partial t} + b - B(X) - \mu t^{-1} \\ \frac{\partial f_\mu}{\partial Z} &= F_p^T \frac{\partial F_p}{\partial Z} + F_d^T \frac{\partial F_d}{\partial Z} + X - \mu Z^{-1}. \end{aligned}$$

Summing the directional derivative over all partial derivatives of F_p we get

$$\begin{aligned}\left\langle F_p^T\frac{\partial F_p}{\partial X},\Delta X\right\rangle+\left\langle F_p^T\frac{\partial F_p}{\partial y},\Delta y\right\rangle+\left\langle F_p^T\frac{\partial F_p}{\partial t},\Delta t\right\rangle+\left\langle F_p^T\frac{\partial F_p}{\partial Z},\Delta Z\right\rangle=\\ =\ \langle F_p,-A(\Delta X)\rangle\\ =\ \langle F_p,-A(\Delta X')\rangle\\ =\ -||F_p||^2,\end{aligned}$$

where the third line follows from $A(M)=A(M^T)$ and the fourth from (4.18). Analogously we get

$$\left\langle F_d^T\frac{\partial F_d}{\partial X},\Delta X\right\rangle+\left\langle F_d^T\frac{\partial F_d}{\partial y},\Delta y\right\rangle+\left\langle F_d^T\frac{\partial F_d}{\partial t},\Delta t\right\rangle+\left\langle F_d^T\frac{\partial F_d}{\partial Z},\Delta Z\right\rangle=-||F_d||^2.$$

For the next step we observe that for a symmetric matrix S $\langle S,\Delta X\rangle=\langle S,\Delta X'\rangle$ since the skew–symmetric part of $\Delta X'$ is orthogonal to symmetric matrices.

$$\begin{aligned}\left\langle Z-\mu X^{-1},\Delta X\right\rangle+\left\langle X-\mu Z^{-1},\Delta Z\right\rangle=\\ =\ \left\langle Z-\mu X^{-1},\Delta X'\right\rangle+\left\langle X-\mu Z^{-1},\Delta Z\right\rangle\\ =\ \mathrm{tr}(I-\mu X^{-1}Z^{-1})(Z\Delta X'+\Delta ZX)\\ =\ -\mathrm{tr}(I-\mu X^{-1}Z^{-1})(ZX-\mu I)\\ =\ -\mathrm{tr}(ZX-2\mu I+\mu^2(ZX)^{-1})\\ =\ -\sum_{i=1}^n\lambda_i(ZX)\left(1-\mu\lambda_i^{-1}(ZX)\right)^2.\end{aligned}$$

As Z and X are positive definite, all eigenvalues of ZX are strictly positive. The last expression is less or equal to zero; and it equals zero if and only if $ZX=\mu I$.

In quite the same manner we get

$$\begin{aligned}(b-B(X)-\mu t^{-1})^T\Delta t+(t-\mu(b-B(X))^{-1})^TB(\Delta X)=\\ =\ -(t\circ(b-B(X)))^T(e-\mu t^{-1}\circ(b-B(X))^{-1})^2.\end{aligned}$$

Again $t\circ(b-B(X))$ is strictly positive, so the expression is less or equal to zero; and it equals zero if and only if $t\circ(b-B(X))=\mu e$.

Summing up we have

$$\begin{aligned}\frac{\partial f_\mu}{\partial s}\Delta s\ =\ &-||F_p||^2-||F_d||^2-\sum_{i=1}^n\lambda_i(ZX)\left(1-\mu\lambda_i^{-1}(ZX)\right)^2\\ &-(t\circ(b-B(X)))^T(e-\mu t^{-1}\circ(b-B(X))^{-1})^2\le 0\end{aligned}\tag{5.4}$$

with equality holding if and only if the Kuhn-Tucker conditions, $F_\mu(s)=0$, are satisfied. □

For the proof of convergence we make the assumption that the set

$$\mathcal{X}=\{X:A(X)=a,B(X)\le b,X\succeq 0\}\tag{5.5}$$

is compact and that we start with an interior point of the problem, i.e a quadruple $s^0 = (X^0, y^0, t^0, Z^0)$ such that

$$\begin{array}{c} a - A(X^0) = 0 \\ b - B(X^0) > 0 \\ Z^0 + C - A^T(y^0) - B^T(t^0) = 0 \\ X^0 \succ 0, t^0 > 0, Z^0 \succ 0. \end{array} \tag{5.6}$$

It is not really necessary to require the equality constraints to hold, but it makes the proof more transparent. It is computationally convenient to work with inaccurate line search methods. We will choose the stepsize α by the Goldstein-Armijo conditions

$$\delta_1 \alpha \left| \frac{\partial f_\mu}{\partial s} \Delta s \right| \leq f_\mu(s) - f_\mu(s + \alpha \Delta s) \leq \delta_2 \alpha \left| \frac{\partial f_\mu}{\partial s} \Delta s \right| \tag{5.7}$$

with $0 < \delta_1 < \delta_2 < 1$. Because f_μ is bounded below and tends to infinity on the boundary of the feasible region such an $\alpha > 0$ always exists and $s + \alpha \Delta s$ is another point satisfying (5.6). Conditions (5.7) form a kind of *sufficient decrease* criterion and help to prove convergence for fixed μ.

Similarly to (DBP) we define the primal barrier problem as follows.

$$\text{(PBP)} \qquad \begin{array}{ll} \text{maximize} & \langle C, X \rangle + \mu \log\det X + \mu e^T \log(b - B(X)) \\ \text{subject to} & a - A(X) = 0 \\ & b - B(X) \geq 0 \\ & X \succeq 0. \end{array} \tag{5.8}$$

The objective function of (PBP) will be denoted by

$$f_P(X) := \langle C, X \rangle + \mu \log\det X + \mu e^T \log(b - B(X)). \tag{5.9}$$

Conditions (5.6) imply that the merit function is the gap between primal and dual barrier problems evaluated at the current point, i.e. for $s = (X, y, t, Z)$ satisfying (5.6) we have

$$f_\mu(s) = f_D(y, t, Z) - f_P(X). \tag{5.10}$$

Since the merit function is decreasing with respect to the iterates for fixed μ,

$$f_\mu(s^{i+1}) \leq f_\mu(s^i), \tag{5.11}$$

and because of weak duality we know that the values of both barrier problems are bounded.

We will show now that the minimal eigenvalue λ_{min} of the iterates X^i, for fixed μ, is bounded away from zero. We denote the maximal eigenvalue of all $X \in \mathcal{X}$ by $\hat{\lambda}_{max}$(if $\hat{\lambda}_{max} < 1$ without loss of generality we set $\hat{\lambda}_{max} = 1$), and the maximal slack of the inequality constraints by $\hat{\sigma}_{max}$ (again we set the value to one if necessary). Both values exist by the compactness of $\mathcal{X}$.

First note that using (5.10) and (5.2) we get

$$f_D(y^i, t^i, Z^i) \geq f_P(X^0) + c_\mu. \tag{5.12}$$

Combining (5.12), (5.11) and (5.10) we obtain the following lower bound on $f_P(X^i)$:

$$f_P(X^i) = f_D(y^i, t^i, Z^i) - f_\mu(s^i) \geq f_P(X^0) + c_\mu - f_\mu(s^0).$$

Using weak duality we can bound f_P from above as follows:

$$\begin{aligned} f_P(X^i) &\leq a^T y^0 + b^T t^0 + \mu(\log\det X^i + e^T \log(b - B(X^i))) \\ &\leq a^T y^0 + b^T t^0 + \mu(n \log \hat{\lambda}_{max} + m \log \hat{\sigma}_{max}) + \mu \log \lambda_{min}(X^i). \end{aligned}$$

Note here that $\hat{\lambda}_{max} \geq 1$.

This yields for all $i \geq 0$

$$\lambda_{min}(X^i) \geq \underbrace{\exp\left(-\frac{\langle Z^0, X^0\rangle + t^{0T}(b - B(X^0)) + f_\mu(s^0) - c_\mu}{\mu} + c\right)}_{=:\ \lambda_{min;\mu}} > 0 \tag{5.13}$$

where c is a negative constant (independent of μ). Analogously we get a lower bound greater than zero for the minimum slack $\sigma_{min;\mu}$ of the iterates. Using these bounds we will now show that the iterates Z^i stay bounded. First we observe that

$$\langle Z, X\rangle \geq \lambda_{min}(X) \sum_{j=1}^{n} \lambda_j(Z) \tag{5.14}$$

Obviously

$$\min_{z_j \geq 0; \sum z_i = \mathrm{tr}(Z^i)} \lambda_{min;\mu} \sum_{j=1}^{n} z_i - \mu \sum_{j=1}^{n} \log(z_i) \leq \left\langle Z^i, X^i \right\rangle - \mu \log\det Z^i \tag{5.15}$$

for all iterates i. So if the left hand side goes to plus infinity for $\mathrm{tr}(Z) \to \infty$ this implies that f_μ goes to infinity. With this in mind we investigate the Lagrangian of the left hand side

$$\mathcal{L} = \lambda_{min;\mu} \sum_{j=1}^{n} z_i - \mu \sum_{j=1}^{n} \log(z_i) + \gamma \left(\sum_{j=1}^{n} z_i - \mathrm{tr}(Z)\right). \tag{5.16}$$

The Kuhn–Tucker conditions yield

$$z_i = \frac{\mathrm{tr}(Z)}{n} \qquad i = 1, \ldots, n \tag{5.17}$$

and for $\mathrm{tr}(Z) \to \infty$ the left hand side indeed goes to plus infinity. We may therefore conclude that the iterates Z^i remain bounded, because of (5.11). Analogously the result can be proved for t. The boundedness of both then implies the boundedness of y by the equality constraints. So for fixed μ the iterates remain within a compact set.

Having established compactness it is easy to prove convergence for fixed μ by some well known results from nonlinear optimization, see e.g. [1].

Theorem 2 *For a fixed value of μ let $\{s^i\}$ be a sequence generated by the algorithm with the stepsize satisfying conditions (5.7) at each iteration and with the starting point s^0 satisfying (5.6). Then $\{s^i\}$ converges to a point s^* in the relative interior of the feasible region and $F_\mu(s^*) = 0$.*

Proof. By the considerations above all points of the sequence lie within a compact set contained in the relative interior of the feasible region. Therefore sequence $\{s^i\}$ has a convergent subsequence $\left\{s^k\right\}_{k\in\mathcal{K}}$ converging to s^*, a point in the relative interior of the feasible region. By a standard result from global convergence theory, conditions (5.7) imply that $\left\{\frac{\partial f_\mu}{\partial s}\frac{\Delta s^k}{||\Delta s^k||}\right\}_{k\in\mathcal{K}} \to 0$. But $||\Delta s^k||$ is bounded for all $k \in \mathcal{K}$ because $\nabla F_\mu(s^*)$ is non singular and f_μ is continuously differentiable, so we must have $\left\{\frac{\partial f_\mu}{\partial s}\Delta s^k\right\}_{k\in\mathcal{K}} \to 0$. From Lemma 1 we conclude that $F_\mu(s^*) = 0$. □

In order to get a sequence of points converging to the solution of the unperturbed problem, μ will be decreased by a constant factor $0 < \theta < 1$ whenever

$$\left|\frac{\partial f}{\partial s}\Delta s^i\right| < \beta\mu \tag{5.18}$$

for some $0 < \beta < \frac{1}{2}$. By Theorem 2 this will eventually be the case for some s^j. We will now show that the sequence formed by these $\{s^j\}_{j\in\mathcal{J}}$ converges to a point satisfying the Kuhn-Tucker conditions (4.4) to (4.7).

Theorem 3 *Let $\left\{\mu^k\right\}$ be a sequence of positive numbers converging to zero. Let $\{s^i\}$ be the sequence of points generated by the algorithm using criterion (5.18) for switching to the next μ-value of the sequence $\left\{\mu^k\right\}$. Let $\{s^j\}_{j\in\mathcal{J}}$ be the subsequence of $\{s^i\}$ formed by the points at which μ-switching occurs. Then $\{s^j\}_{j\in\mathcal{J}}$ converges to a point s^* fulfilling the Kuhn-Tucker conditions (4.4) to (4.7).*

Proof. For a point s satisfying (5.18) for a certain μ we get from (5.4)

$$\begin{aligned} \lambda_i(ZX)\left(1-\mu\lambda_i^{-1}(ZX)\right)^2 &< \beta\mu, \qquad i = 1\ldots n \\ (t\circ(b-B(X)))^T(e-\mu t^{-1}\circ(b-B(X))^{-1})^2 &< \beta\mu. \end{aligned}$$

For a variable $x > 0$ we have

$$x\left(1-\frac{\mu}{x}\right)^2 < \beta\mu \quad\Longrightarrow\quad \frac{\mu}{2} < x < 2\mu.$$

Therefore

$$\frac{\mu}{2} < \quad \lambda_i(ZX) \quad < 2\mu \tag{5.19}$$

$$\frac{\mu}{2}e < \quad t\circ(b-B(X)) \quad < 2\mu e \tag{5.20}$$

and, for μ small enough, we get the following bound for f_μ

$$f_\mu < (n+m)\mu\,(2 - \log\mu - \log 2)\,. \tag{5.21}$$

Furthermore the gap between primal and dual feasible solution is bounded by

$$\langle Z, X\rangle + t^T(b - B(X)) < 2(n+m)\mu. \tag{5.22}$$

As in (5.13) this bounds the minimal eigenvalue of X

$$\lambda_{min}(X) \geq \exp\left(-\frac{\langle Z, X\rangle + t^T(b - B(X)) + f_\mu - c_\mu}{\mu} + c_1\right) > \mu c_2 > 0 \tag{5.23}$$

with a negative constant c_1 and a positive constant c_2 (independent of μ). It is now easy to bound $\lambda_{max}(Z)$ by

$$\mu c_2 \lambda_{max}(Z) < \langle Z, X\rangle < 2n\mu. \tag{5.24}$$

Analogously we get the boundedness of t and consequently of y. So for $\mu \to 0$ we have

$$\begin{array}{rcll} \lambda_i(ZX) & \to & 0, \quad i = 1\ldots n & \Longrightarrow \quad ZX \to 0 \\ t \circ (b - B(X)) & \to & 0, \quad i = 1\ldots m & \\ f_\mu(t, X, y, Z) & \to & 0. & \end{array}$$

Because of the special choice of the s^j the same is true for the sequence $\{s^j\}_{j\in\mathcal{J}}$. Because of compactness there must be an accumulation point s^*. By the considerations above s^* is feasible and satisfies the Kuhn-Tucker conditions of the original problem. □

Note, that $f \to 0$ implies that the gap between the primal and dual solutions indicated by s^* is zero. This is another proof of strong duality.

6 Computational Results

6.1 Interior Point Approach for Max–Cut

In this section we show how the Max–Cut relaxation given in Section 3 can be implemented efficiently within this framework. We first look at the relaxation without the triangle inequalities. The resulting program is very simple and can be solved for quite large n very fast, see Table 1.

The cost matrix is given by the Laplacian L of the weighted adjacency matrix of the graph. The SDP forming the relaxation reads

$$\begin{array}{ll} \text{maximize} & \mathrm{tr} LX \\ \text{subject to} & \mathrm{diag}(X) = \frac{1}{4}e =: a \\ & X \succeq 0. \end{array} \tag{6.1}$$

Size n	*Iterations*	*CPU minutes*
100	14	0.8
150	12	2.5
200	12	6.5
250	13	14.4
300	14	30.4
400	14	84.6
500	14	173.3

Table 1: SDP relaxation for max-cut. The times given are CPU minutes on a 486 PC running with 66 Mhz

Since Diag is the adjoint of diag, the dual SDP reads

$$\begin{array}{rl} \text{minimize} & a^T y \\ \text{subject to} & \mathrm{Diag}(y) - L = Z \\ & Z \succeq 0. \end{array} \tag{6.2}$$

It is easy to show that the starting point

$$X := \mathrm{Diag}(a) \tag{6.3}$$
$$y := 1.1 \cdot \mathrm{abs}(L)e \tag{6.4}$$
$$Z := \mathrm{Diag}(y) - L \tag{6.5}$$

satisfies all the requirements of (5.6). Now we plug in the algorithm and get

$$O_{11}(\Delta y) = \mathrm{diag}(Z^{-1}\mathrm{Diag}(\Delta y)X) = (Z^{-1} \circ X)\Delta y \tag{6.6}$$
$$v_1 = \mu \mathrm{diag}(Z^{-1}) - a. \tag{6.7}$$

This means that at each iteration we solve the system by computing

$$\Delta y := (Z^{-1} \circ X)^{-1}(\mu \mathrm{diag}(Z^{-1}) - a) \tag{6.8}$$
$$\Delta Z := \mathrm{Diag}(\Delta y) \tag{6.9}$$
$$\Delta X' := \mu Z^{-1} - X - Z^{-1}\Delta Z X \tag{6.10}$$
$$\Delta X := (\Delta X'^T + \Delta X')/2. \tag{6.11}$$

To further emphasize the simplicity and ease of implementation of this approach, we include a MATLAB–function in the appendix, that solves this relaxation. For problems of sizes up to $n \approx 200$ this is a very efficient code and the reader is encouraged to test it.

The numerical results of Table 1 were computed on a PC using a C encoded version of this algorithm. We point out that the number of iterations does not seem to depend significantly on n. The computation times are minutes on an MS-DOS machine running

with 66 Mhz. The instances were generated as random unweighted graphs with edge probability $p = 0.5$, we set the stopping condition to 6 digits of accuracy, as in the MATLAB routine from the appendix.

As a second experiment we looked at the max-cut relaxation where we included all the $4\binom{n}{3}$ triangle inequalities along with the semidefiniteness constraints, see section 3.1. When we included triangle constraints it turned out favorable to use Mehrotra's predictor – corrector approach, which is closely investigated in [7]. In this case the system is solved twice with the same factorization but with different right hand sides. First the right hand side is chosen as to minimize the objective, the second step computes a good centering direction for the predicted point. The triangle inequalities were included successively according to their amount of violation, but for the benefit of the reader we will omit the details. Some results for random graphs are given in Table 2. Most of the time the best cut was already found by the relaxation without triangles and could not be improved later on. It was constructed from the rounded rows of X giving an initial cut. We then applied a local improvement strategy to this cut by considering swapping any single node from one side to the other side of the cut. If any such swap improved the cut, we chose the swap with highest gain. We stopped if no further improvement could be obtained this way.

We note that these semidefinite programs have n equality and $4\binom{n}{3}$ inequality constraints in the dual program. The final solution satisfies all these constraints and was obtained after *it* interior point iterations, see the last column. To be explicit, the solution of problems with $n = 100$ satisfies $4\binom{100}{3} = 646,800$ inequality constraints. Our approach does not seem to depend on structural properties of the underlying graph. It significantly extends purely polyhedral techniques for max-cut. Further and more substantial computational results with this approach, applied to max-cut, will appear in a forthcoming study [6].

6.2 Min-Max Eigenvalue Problems

Consider the min-max eigenvalue problem:

$$\begin{array}{ll} \text{minimize} & \lambda_{max}(C - \text{Diag}(v)) \\ \text{subject to} & e^T v = 0. \end{array} \tag{6.12}$$

The objective function in (6.12) is not differentiable when the multiplicity of the largest eigenvalue exceeds one. In fact, a singleton eigenvalue characterizes differentiability. Since the largest eigenvalue is a convex function, subgradient approaches can be used to solve (6.12) (see, e.g., [8]). More recently, it has been shown that Newton-based algorithms with local quadratic convergence exist (see, e.g., [20]) but the local convergence depends on correctly identifying the multiplicity of the largest eigenvalue. We present computational experiments showing that our interior-point method is indeed robust in the presence of high multiplicity.

By minimizing λ over the set $\left\{\lambda I \succeq C - \text{Diag}(v), e^T v = 0\right\}$ and substituting $y := v + \lambda e$

n	cut	upper b. (%)	hh.mm.ss	it
random graphs, edge-weights $\in \{-10, \ldots, 10\}$				
30	330	330.0001 (0.0)	00.00.54	59
40	619	619.0000 (0.0)	00.02.02	60
50	708	708.0001 (0.0)	01.04.22	166
70	1215	1248.7142 (2.8)	04.42.21	220
100	2440	2540.2528 (4.1)	37.56.05	266
random unweighted graphs, edgeprob. 0.5				
30	143	143.0000 (0.0)	00.00.56	62
40	245	245.6864 (0.3)	00.59.19	247
50	374	375.3689 (0.4)	01.54.18	250
70	723	726.0399 (0.4)	05.57.22	257
100	1413	1427.8731 (1.1)	22.45.24	264

Table 2: Solutions to Max–Cut relaxation with triangle inequalities. The number of nodes in the graph is n, *cut* refers to the best cut found, *upper b.* is the value of the relaxation, % is the percentage of the gap between relaxation and best cut found. The next column gives the computation times on a DEC-station (the last line was computed on a Silicon-Graphics which is about twice as fast), *it* gives the number of iterations. Note that *it* is almost independent of n if the gap is not closed.

we see that (6.12) is indeed a special case (with $a = e/n$) of the more general problem

$$\begin{array}{ll} \text{minimize} & a^T y \\ \text{subject to} & \mathrm{Diag}(y) - C \succeq 0. \end{array} \tag{6.13}$$

The MATLAB code given in the appendix also applies to this problem. To test the code on problem instances that exhibit given multiplicity at the optimal solution, we developed a special generator which we now describe.

To generate positive semidefinite programs we generate the elements of a uniformly on some interval of the nonnegative half-line (the primal problem is clearly infeasible if any component of a is negative). For the experiments described below, we used $a = e/n$.

Given a, we generate C as follows. First, we generate an $n \times m$ random matrix A and apply row scaling to make all squared row norms equal to the corresponding elements of a. That is,

$$\mathrm{diag}(AA^T) = a. \tag{6.14}$$

We denote the columns of A by $v_1, \ldots, v_m$. We then construct $n-m$ additional random n-vectors $v_{m+1}, \ldots, v_n$ and apply Gram-Schmidt orthonormalization to $v_1, \ldots, v_n$ to produce an $n \times n$ orthogonal matrix Q whose first m columns span the same space as $v_1, \ldots, v_m$. Finally, we set

$$C = Q \Lambda Q^T, \tag{6.15}$$

			Time (MM:SS.S)	
n	m	k	BT	IP
10	1	1	0.0	0.0
20	1	1	0.1	0.3
30	1	1	0.3	1.0
50	1	1	1.2	4.2
100	1	1	12.6	35.6
200	1	1	1:56.9	6:17.4

Table 3: Statistics for problems with multiplicity 1. BT refers to the Bundle Trust method and IP refers to our Interior-Point method.

			Time (MM:SS.S)		Comments
n	m	k	BT	IP	
20	3	3	0.4	0.4	
20	5	5	2.8	0.3	
20	5	12	2.8	0.3	
20	8	8	2.7	0.3	
20	12	12	3.6	0.3	
30	3	3	1.5	1.0	
30	3	6	1.5	0.9	
30	6	6	18.2	0.8	
30	10	10	4.0	1.0	Num. trouble in BT
50	5	5	>20:00.0	4.3	5 sig. fig. in BT
100	3	3	18.7	33.9	
100	6	6	>15:00.0	36.9	5 sig. fig. in BT
500	50	50	-	2:02:47.0	No attempt at BT

Table 4: Statistics for problems with higher built-in multiplicity. BT refers to the Bundle Trust method and IP refers to our Interior-Point method.

where Λ is a diagonal matrix whose first $k \geq m$ entries are all set to $\lambda_{\max}$ (which is a constant that can be chosen arbitrarily – we used 5) and the remaining diagonal entries are generated uniformly on some interval that is strictly smaller than $\lambda_{\max}$. For such a matrix C, we claim that

$$\begin{aligned} X &= AA^T \\ y &= \lambda_{max} e \\ Z &= \mathrm{Diag}(y) - C \end{aligned}$$

is optimal. Indeed, it follows from (6.14) that X is feasible for the primal and it is clear from (6.15) that (y, Z) is feasible for the dual. Finally, optimality follows from the absence of a duality gap:

$$\mathrm{tr}(ZX) = \mathrm{tr}\{(\lambda_{max} I - C)AA^T\} = 0$$

The last equality follows from the fact that the columns of A are eigenvectors of C associated with the maximal eigenvalue.

Table 3 shows the comparison between the bundle trust method and our interior-point method when the optimal eigenvalue is a singleton ($k = 1$). For these problems, the bundle trust method is three to four times faster. However, this situation never arises in practice. Indeed, for $k = 1$ in our construction above, we see that we are requiring the vector of all ones to be a maximal eigenvector of C. This is clearly an unlikely event in real applications to happen.

Table 4 shows comparisons for higher multiplicities. Here the results look much better for the interior-point method. In fact, it is clear that the bundle trust method completely breaks down rather rapidly as the multiplicity increases.

7 Appendix

The following MATLAB function solves the semidefinite programming problem described above with $A(X) = \mathrm{diag}(X)$ and no inequality constraints. We include it to further emphasize the simplicity of our interior point approach. We assume that MATLAB's version 4 is available in the positive definiteness test, using MATLAB's built in function `chol`. Other than that the program could also be run under older versions of MATLAB. We have somewhat arbitrarily set the stopping condition to 6 digits of accuracy. Thus, after successful termination, we return primal and dual feasible solutions, whose objective values agree on approximately 6 digits.

In all our practical experiments it proved unnecessary to check the decrease of the merit function. Therefore this feature is not included in the present algorithm. We also mention that in case of 'large steps', i.e. $\alpha_p + \alpha_d \geq 1.8$ we reduce μ by a factor of 4, and otherwise by a factor of 2.

```
function [phi, X, y] = psd_ip( L);
% solves: max trace(LX) s.t. X psd, diag(X) = b;  b = ones(n,1)/4
%         min b'y       s.t. Diag(y) - L psd, y unconstrained,
% input:  L ... symmetric matrix
```

```
% output: phi ... optimal value of primal, phi =trace(LX)
%         X   ... optimal primal matrix
%         y   ... optimal dual vector
% call:   [phi, X, y] = psd_ip( L);

digits = 6;                      % 6 significant digits of phi
[n, n1] = size( L);              % problem size
b = ones( n,1 ) / 4;             % b>0 works just as well
X = diag( b);                    % initial primal matrix is pos. def.
y = sum( abs( L))' * 1.1;        % initial y is chosen so that
Z = diag( y) - L;                % initial dual slack Z is pos. def.
phi = b'*y; psi = trace( L*X);   % initial dual and primal costs
mu = trace( Z*X)/( 2*n);         % initial complementarity
iter=0;                          % iteration count

disp(['      iter    alphap    alphad      gap     lower      upper']);

while phi-psi > phi*10^(-digits) % while duality gap is too large

      iter = iter + 1;           % start a new iteration
      Zi = inv( Z);              % inv(Z) is needed explicitely
      dy =  (Zi.*X) \ (mu * diag(Zi) - b);      % solve for dy
      dX = - Zi * diag( dy) * X + mu * Zi - X;  % backsubstitute for dX
      dX = ( dX + dX')/2;        % symmetrise

% line search on primal
      alphap = 1;                % initial steplength
      [dummy,posdef] = chol( X + alphap * dX );  % test if pos.def
      while posdef > 0,
              alphap = alphap * .8;
              [dummy,posdef] = chol( X + alphap * dX );
              end;
      if alphap < 1, alphap = alphap * .95; end; % stay away from boundary
% line search on dual; dZ is handled implicitely: dZ = diag( dy);
      alphad = 1;
      [dummy,posdef] = chol( Z + alphad * diag(dy) );
      while posdef > 0;
              alphad = alphad * .8;
              [dummy,posdef] = chol( Z + alphad * diag(dy) );
              end;
      if alphad < 1, alphad = alphad * .95; end;
% update
      X = X + alphap * dX;
      y = y + alphad * dy;
      Z = Z + alphad * diag(dy);
      mu = trace( X * Z) / (2*n);
      if alphap + alphad > 1.8, mu = mu/2; end; % speed up for long steps
      phi = b' * y; psi = trace( L * X);
% display current iteration
        disp([ iter alphap alphad (phi-psi) psi  phi ]);

        end;             % end of main loop
```

References

[1] K.M. ANSTREICER, J.-P. VIAL. *On the Convergence of an Infeasible Primal-*

Dual Interior-Point Method for Convex Programming. Technical Report 1993.16, Départment d'économie commerciale et industrielle, Université de Genève, Switzerland, 1993.

[2] F. ALIZADEH. *Combinatorial optimization with interior point methods and semidefinite matrices.* PhD thesis, University of Minnesota, 1991.

[3] F. ALIZADEH. Combinatorial optimization with semidefinite matrices. In *Proceedings of the Second Annual Integer Programming and Combinatorial Optimization Conference*, Carnegie-Mellon University, 1992.

[4] F. ALIZADEH, J.A. HAEBERLY, and M.L. OVERTON. A new primal-dual Interior-point method for semidefinite programming. to appear in: Proceedings of the Fifth SIAM Conference on Appl. Linear Algebra, Snowbird, Utah, June 1994.

[5] F. BARAHONA. The max-cut problem in graphs not contractible to K_5. *Operations Research Letters*, 2:107–111, 1983.

[6] M. BURKARD, C. HELMBERG, and F. RENDl. Interior point approach for max-cut. Technical Report University of Technology Graz, 1994. in preparation.

[7] T.J. CARPENTER, I.J. LUSTIG, J.M. MULVEY, and D.F.SHANNO. Higher-order predictor-corrector interior point methods with application to quadratic objectives. *SIAM J. Optimization*, 3:696–725, 1993.

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

[9] C. DELORME and S. POLJAK. Laplacian eigenvalues and the maximum cut problem. *Mathematical Programming*, 62:557–574, 1993.

[10] J. FALKNER, F. RENDL, and H. WOLKOWICZ. A computational study of graph partitioning. Technical Report CORR, Department of Combinatorics and Optimization, Waterloo, Ont, 1992. to appear in *Math. Programming.*

[11] M.X. GOEMANS and D.P. WILLIAMSON. .878-approximation algorithm for max–cut and max–2sat. Technical Report, MIT, 1993.

[12] F. O. HADLOCK. Finding a maximum cut in a planar graph in polynomial time. *SIAM J. Computing*, 4:221–225, 1975.

[13] F. JARRE. An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices. *SIAM J. Control and Optimization*, 31:1360-1377, 1993.

[14] F. JARRE. Interior-point methods via self-concordance or relative Lipschitz condition. Habilitationsschrift, Würzburg University, March 1994.

[15] M. LAURENT and S. POLJAK. the metric polytope. Proceedings of IPCO 2, (E.Balas, G.Cornuejols and R.Kannan eds.), 1992, 274–286.

[16] B. MOHAR and S. POLJAK. Eigenvalues in combinatorial optimization, in "Combinatorial and graph-theoretic problems in linear algebra", R. Brualdi, S. Friedland, V. Klee, Eds., IMA Volumes in Mathematics and Its Applications, Vol. 50, Springer-Verlag, 1993.

[17] Y. E. NESTEROV and A. S. NEMIROVSKII. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms*. SIAM Publications. SIAM, Philadelphia, USA, 1993.

[18] M.L. OVERTON. On minimizing the maximum eigenvalue of a symmetric matrix. *SIAM J. Matrix Analysis and Applications*, 9:256–268, 1988.

[19] M.L. OVERTON. Large-scale optimization of eigenvalues. *SIAM J. Optimization*, 2:88–120, 1992.

[20] M.L. OVERTON and R.S. WOMERSLEY. Second derivatives for optimizing eigenvalues of symmetric matrices. Technical Report 627, Computer Science Department, NYU, 1993.

[21] S. POLJAK and F. RENDL. Nonpolyhedral relaxations of graph bisection problems. Technical Report, University of Technology Graz, 1992. to appear in *SIAM J. Optimization*.

[22] F. RENDL, R. J. VANDERBEI, and H. WOLKOWICZ. Max-min eigenvalue problems, primal-dual interior point algorithms, and trust region subproblems. Technical Report CORR 93-30, Department of Combinatorics and Optimization, Waterloo, Ont. To appear in Proceedings of NATO conference on Nonlinear Programming, Il-Ciocco, Italy, 1993.

[23] 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.

[24] L. VANDENBERGHE and S. BOYD. Primal-dual potential reduction method for problems involving matrix inequalities. Technical report, Electrical Engineering Department, Stanford University, Stanford, CA 94305, 1993.

[25] R.J. VANDERBEI and T.J. CARPENTER. Symmetric indefinite systems for interior-point methods. *Mathematical Programming*, 58:1–32, 1993.

[26] J.-P. VIAL. *Computational Experience with a Primal-Dual Interior-Point Method for Smooth Convex Programming*. to appear in Optimization Methods and Software.

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