SOME ASPECTS OF NONLINEAR SEMIDEFINITE PROGRAMMING

Florian Jarre
Institut für Mathematik
Universität Düsseldorf
Universitätsstraße 1
D–40225 Düsseldorf, Germany
jarre@opt.uni-duesseldorf.de

Abstract This paper is an extended abstract of a survey talk given at the IFIP TC7 Conference in Trier, July 2001. We consider linear and nonlinear semidefinite programming problems and concentrate on selected aspects that are relevant for understanding dual barrier methods. The paper is aimed at graduate students to highlight some issues regarding smoothness, regularity, and computational complexity without going into details.

Keywords: Semidefinite programming, smoothness, regularity, interior method, local minimization.

1. Introduction

In this paper we consider nonlinear semidefinite programming problems (NLSDP's) and concentrate on some aspects relevant to a dual barrier method. Other approaches for solving NLSDP's are the program package LOQO of Vanderbei (1997) based on a primal-dual approach, or recent work of Vanderbei et.al. (2000). Also the work of Kocvara and Stingl (2001) solving large scale semidefinite programs based on a modified barrier approach seems very promising. The modified barrier approach does not require the barrier parameter to converge to zero and may thus overcome some of the problems related to ill-conditioning in traditional interior methods. Optimality conditions for NLSDP's are considered in Forsgren (2000); Shapiro and Scheinberg (2000).

Some problems considered in this paper do not satisfy any constraint qualification. For such problems primal-dual methods do not appear to be suitable. Another question addressed in this paper is the question

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35699-0_19
E. W. Sachs et al. (eds.), *System Modeling and Optimization XX*

of how to avoid "poor" local minimizers, a question that may be even more difficult to investigate for primal dual methods than it is for barrier methods.

1.1 Notation

The following notation has become standard in the literature on linear semidefinite programs. The space of symmetric $n \times n$-matrices is denoted by $\mathcal{S}^n$. The inequality

$$X \succeq 0, \qquad (X \succ 0)$$

is used to indicate that X is a symmetric positive semidefinite (positive definite) $n \times n$-matrix. By

$$\langle C, X \rangle = C \bullet X = \text{trace}(C^T X) = \sum_{i,j} C_{i,j} X_{i,j}$$

we denote the standard scalar product on the space of $n \times n$-matrices inducing the Frobenius norm,

$$X \bullet X = \|X\|_F^2.$$

For given symmetric matrices $A^{(i)}$ we define a *linear* map $\mathcal{A}$ from $\mathcal{S}^n$ to $\mathbb{R}^m$ by

$$\mathcal{A}(X) = \begin{pmatrix} A^{(1)} \bullet X \\ \vdots \\ A^{(m)} \bullet X \end{pmatrix}.$$

The adjoint operator $\mathcal{A}^*$ satisfying

$$\langle \mathcal{A}^*(y), X \rangle = y^T \mathcal{A}(X) \qquad \forall\, X \in \mathcal{S}^n,\ y \in \mathbb{R}^m$$

is given by

$$\mathcal{A}^*(y) = \sum_{i=1}^{m} y_i A^{(i)}.$$

2. Linear semidefinite programs

In this section we consider a pair of primal and dual (linear) semidefinite programs in standard form,

$$(P) \qquad \text{minimize } C \bullet X \ \text{ s.t. } \ \mathcal{A}(X) = b,\ X \succeq 0$$

and

$$(D) \qquad \text{maximize } b^T y \ \text{ s.t. } \ \mathcal{A}^*(y) + S = C,\ S \succeq 0.$$

When using the notion "semidefinite program" (SDP) we always refer to a linear semidefinite program; nonlinear SDP's will be denoted by NLSDP. The data for (P) and (D) are a linear map $\mathcal{A}$, a vector $b \in \mathbb{R}^m$ and a matrix $C \in \mathcal{S}^n$. We use the convention that the infimum of (P) is $+\infty$ whenever (P) does not have any feasible solution X, and the supremum of (D) is $-\infty$ if the feasible set of (D) is empty. If there exists a matrix $X \succ 0$ (not just $X \succeq 0$) that is feasible for (P) then we call X "strictly" feasible for (P) and say that (P) satisfies Slaters condition. Likewise, if there exists a matrix $S \succ 0$ that is feasible for (D) we call (D) strictly feasible. If Slaters condition is satisfied by (P) or by (D) then the optimal values of (P) and (D) coincide, and if both problems satisfy Slaters condition, then the optimal solutions X^{opt} and y^{opt}, S^{opt} of both problems exist and satisfy the equation

$$X^{opt} S^{opt} = 0. \tag{1}$$

Conversely, any pair X and y, S of feasible points for (P) and (D) satisfying (1) is optimal for both problems, see e.g. Shapiro and Scheinberg (2000). Condition (1) implies that there exists a unitary matrix U that simultaneously diagonalizes X^{opt} and S^{opt}. Moreover, the eigenvalues of X^{opt} and S^{opt} to the same eigenvector are complementary.

The two main applications of semidefinite programs are relaxations for combinatorial optimization problems, see e.g. Alizadeh (1991); Helmberg et.al. (1996); Goemans and Williamson (1994), and semidefinite programs arising from Lyapunov functions or from the positive real lemma in control theory, see e.g. Boyd et.al. (1994); Leibfritz (2001); Scherer (1999). Next, we give two simple examples for such applications.

2.1 A first simple example

In our first example we consider the differential equation

$$\dot{x}(t) = Ax(t)$$

for some vector function $x : \mathbb{R} \to \mathbb{R}^n$. By definition, this system is called stable if for all initial values $x^{(0)} = x(0)$ the solutions $x(t)$ converge to zero when $t \to \infty$. It is well known, see e.g. Hirsch and Smale (1974), that this is the case if and only if the real part of all eigenvalues of A is negative,

$$\mathrm{Re}(\lambda_i(A)) < 0 \quad \text{for } 1 \leq i \leq n.$$

By Lyapunov's theorem, this is the case if and only if

$$\exists P \succ 0 : \quad -A^T P - PA \succ 0.$$

Let us now assume that the system matrix A is subject to uncertainties that can be "confined" to a polyhedron with m given vertices $A^{(i)}$, i.e.

$$A = A(t) \in \operatorname{conv}\{A^{(1)}, \ldots, A^{(m)}\} \quad \text{for } t \geq 0.$$

In this case the existence of a Lyapunov matrix $P \succ 0$ with

$$-(A^{(i)})^T P - P A^{(i)} \succ 0 \quad \text{for } 1 \leq i \leq m \tag{2}$$

implies that

$$-A(t)^T P - P A(t) \succ 0 \quad \text{for all } t \geq 0,$$

and hence,

$$0 > x(t)^T \Big(A(t)^T P + P A(t) \Big) x(t) = (A(t)x(t))^T P x(t) + x(t)^T P \left(A(t)x(t)\right)$$

$$= \frac{d}{dt} \Big(x(t)^T P x(t) \Big) = \frac{d}{dt} \|x(t)\|_P^2$$

whenever $x(t) \neq 0$. This implies that $\|x(t)\|_P^2 \to 0$, and hence the existence of a matrix $P \succ 0$ satisfying (2) is a sufficient condition to prove stability of the uncertain system. (The above argument only shows that $\|x(t)\|_P$ is monotonously decreasing. In order to show that $\|x(t)\|_P$ converges to zero, one can find a strictly negative bound for $\frac{d}{dt}\|x(t)\|_P^2$ using the largest real part of the eigenvalues of $(A^{(i)})^T P + P A^{(i)}$.)

There are straightforward ways to formulate the problem of finding a matrix $P \succ 0$ satisfying (2) as a linear semidefinite program, see e.g. Boyd et.al. (1994). While this simple example results in a linear semidefinite program, other problems from controller design often result in bilinear semidefinite programs that are no longer convex, see e.g. Leibfritz (2001); Scherer (1999); Freund and Jarre (2000).

2.2 A second simple example

Binary quadratic programs (also known as max-cut-problems) have few applications in VLSI layout or in spin glass models from physics. Their most important property, however, appears to be the fact that these problems are $\mathcal{NP}$-complete (and hence, there is no known polynomial time method for solving these problems). What makes these problems so appealing is that they appear to be quite easy.

Let

$$\mathcal{MC} = \operatorname{conv}\left(\left\{ x x^T \mid x_i \in \{\pm 1\} \quad \text{for } 1 \leq i \leq n \right\} \right) \subset \mathcal{S}^n$$

be the max cut polytope. Hence, $\mathcal{MC}$ is the convex hull of all rank-1 matrices generated by ± 1-vectors x. Any binary quadratic program or

any max-cut problem can be written in the following form:

$$\text{minimize } C \bullet X \text{ s.t. } X \in \mathcal{MC}. \tag{3}$$

This is a standard linear program with the drawback that the feasible set $\mathcal{MC}$ is defined as convex hull of exponentially many points xx^T, rather than being defined by (a polynomial number of) linear inequalities.

Let $e = (1, \ldots, 1)^T$ be the vector of all ones. It is straightforward to see that $\mathcal{MC}$ can be written in the form

$$\mathcal{MC} = \text{conv}\left(\{X \succeq 0 \mid \text{ diag}(X) = e, \text{ rank}(X) = 1\}\right).$$

Due to the condition $\text{diag}(X) = e$ the set $\mathcal{MC}$ lies in an affine subspace of $\mathcal{S}^n$ of dimension $n(n-1)/2$. $\mathcal{MC}$ has 2^{n-1} vertices that are pairwise adjacent i.e. connected by an edge (a 1-dimensional extreme set of $\mathcal{MC}$).

Note that the constraints of this second definition of $\mathcal{MC}$ appear to be smooth constraints; a semidefiniteness constraint, a linear constraint, and a rank condition. These conditions, however, imply that there are only finitely many "discrete" elements of which the convex hull is taken. In some sense the constraints contain a hidden binary constraint allowing only certain matrices with entries ± 1. When the rank constraint is omitted, we obtain the standard SDP relaxation of the max-cut problem,

$$\mathcal{SDP} = \{X \succeq 0 \mid \text{ diag}(X) = e\}$$

satisfying $\mathcal{MC} \subset \mathcal{SDP}$. A relaxed version of (3) is thus given by

$$\text{minimize } C \bullet X \text{ s.t. } X \in \mathcal{SDP}. \tag{4}$$

This problem is a linear SDP of the form (P) and can be solved efficiently using, for example, interior point methods, see e.g. Helmberg et.al. (1996). Goemans and Williamson (1994) have shown how to obtain an excellent approximation of the max-cut problem (3) using the solution X of (4).

A quite interesting inner approximation of $\mathcal{MC}$ leading to a nonlinear semidefinite program is described in Chapter 3.3.

2.3 Smoothness of semidefiniteness constraint

To understand the complexity of nonlinear semidefinite programs we briefly address the question of smoothness and regularity of the semidefinite cone. The set of positive semidefinite matrices can be characterized in several different forms,

$$\{X \mid X \succeq 0\}$$

$$
\begin{aligned}
&= \{X \mid \lambda_{min}(X) \geq 0\} \\
&= \{X \mid \lambda_i(X) \geq 0 \ \text{ for } 1 \leq i \leq n\} \\
&= \left\{X \mid u^T X u \geq 0 \ \text{ for all } u \in I\!R^n \ (\|u\| = 1)\right\} \\
&= \{X \mid \det(X_{\Sigma,\Sigma}) \geq 0 \ \text{ for all } \Sigma \subset \{1, \ldots, n\}\} \\
&= \left\{X \mid \exists Z \in \mathcal{S}^n : \ X = Z^2\right\}.
\end{aligned}
$$

The first characterization uses the smallest eigenvalue $\lambda_{min}(X)$ of X. This is a nonsmooth representation. When ordering the eigenvalues in a suitable way, the eigenvalues $\lambda_i(X)$ used in the second representation have directional derivatives, but are not totally differentiable. The third representation is based on a semi-infinite constraint. From this representation one can easily deduce, for example, that $\{X \mid X \succeq 0\}$ is convex. The fourth representation is based on a finite (but exponential) number of smooth constraints, requiring all principal subdeterminants to be nonnegative. This representation certainly justifies the claim that $\{X \mid X \succeq 0\}$ is bounded by smooth constraints. As shown in Pataki (2000), the tangent plane to $\{X \mid X \succeq 0\}$ at a point $\hat{X}$ is given as follows. Let

$$\hat{X} = UDU^T$$

with a diagonal matrix D and a unitary matrix U. If $\hat{X}$ is a boundary point of $\{X \mid X \succeq 0\}$ we may assume without loss of generality that the first k diagonal entries of D satisfy $d_1 = \ldots = d_k = 0$ and $d_{k+1}, \ldots, d_n > 0$. Let ΔX be given by

$$\Delta X = U \begin{pmatrix} 0 & * \\ * & * \end{pmatrix} U^T$$

where the 0-block in the matrix on the right hand side is of size $k \times k$, and the entries $*$ are any entries of suitable dimension. All matrices ΔX of the above form belong to the tangent space of $\{X \mid X \succeq 0\}$ at $\hat{X}$.

The fourth representation also leads to the convex barrier function

$$\tilde{\Phi}(X) = -\log\det(X)$$

for the positive semidefinite cone. For this barrier function it is sufficient to consider $\Sigma = \{1, \ldots, n\}$, and to set $\tilde{\Phi}(X) = \infty$ whenever X is not positive definite.

The last representation is a projection of a quadratic equality constraint.

Most, if not all, of the above representations have been used numerically to enforce semidefiniteness of some unknown matrix X.

The set $\{X \mid X \succeq 0\}$ certainly satisfies Slaters condition, or, in the context of nonconvex minimization, any point $\bar{X} \in \{X \mid X \succeq 0\}$ trivially satisfies the constraint qualification by Robinson. However, the

fourth representation above does not satisfy LICQ. (LICQ is a common regularity condition requiring that the active constraints at any point are linearly independent, see e.g. Wright and Nocedal (1999).) In fact, (for $n > 1$) there does not exist any representation of the positive semidefinite cone by nonlinear inequalities that do satisfy LICQ. Nevertheless the positive semidefinite cone and its surface are numerically tractable, and may be considered as a regular set with smooth constraints.

2.4 A dual barrier method

We consider problem (D) and eliminate the slack variable S to obtain the problem

$$\text{maximize } b^T y \text{ s.t. } C - \mathcal{A}^*(y) \succeq 0.$$

For $y \in I\!R^m$ with $C - \mathcal{A}^*(y) \succeq 0$ we then define a convex barrier function $\hat{\Phi}$,

$$\hat{\Phi}(y) = -\log\left(\det(C - \mathcal{A}^*(y))\right).$$

A plain dual barrier method can be stated as follows:

Dual barrier method
Start: Find $y^{(0)}$ with $C - \mathcal{A}^*(y^{(0)}) \succeq 0$.
For $k = 1,\ 2,\ 3,\ \ldots$

- Set $\mu_k = 10^{-k}$ and find

$$y^{(k)} \approx y(\mu_k) = \arg\min_y \ -\frac{b^T y}{\mu_k} + \hat{\Phi}(y)$$

  by Newton's method with line search starting at $y^{(k-1)}$.

Of course, this conceptual method needs many refinements such as an appropriate choice of the starting point and a somewhat more sophisticated update of μ_k. With such minor modifications, however, the above algorithm solves the semidefinite programming problem in polynomial time. (The notion of polynomiality in the context of nonlinear programming is to be taken with care; the solution of a linear semidefinite program can have an exponential "size" like an optimal value of 2^{2^n} for a semidefinite program with encoding length $O(n)$. Our reference to "polynomial time" is meant that the method reduces some primal dual gap function in polynomial time, see e.g. Nesterov and Nemirovski (1994).)

The key elements in guaranteeing the theoretical efficiency of the barrier method rest on two facts:

- The duality gap (or some linear measure of closeness to optimality) is of order μ,

- and the Hessian $\nabla^2\hat{\Phi}$ of the barrier function satisfies a local relative Lipschitz condition.

Both facts were shown by Nesterov and Nemirovski (1994) and rest on two conditions introduced in Nesterov and Nemirovski (1994). The first fact is implied by a local Lipschitz condition of $\hat{\Phi}$ with respect to the norm induced by $\nabla^2\hat{\Phi}(y)$, and the second fact is called self-concordance, and implies that Newton's method converges globally at a fixed rate. More details can be found in Nesterov and Nemirovski (1994); Jarre (1996).

The guaranteed convergence results in these references are much slower than what is observed in implementations of related methods. In fact, these theoretical results are much too slow to be relevant for practical applications. However, these results guarantee a certain independence of the method from the data of the problem. Even with exact arithmetic, the performance of the steepest descent method for unconstrained minimization, for example, depends on the condition number of the Hessian matrix at the optimal solution. Unlike the steepest descent method, the worst case bound for the barrier method only depends on the dimension n of the problem (D), but not on any condition numbers or any other parts of the data of the problem. In this respect, the theoretical analysis *is* relevant for practical applications.

The above barrier method is not suitable for practical implementations. The following simple acceleration scheme is essential for obtaining a more practical algorithm: Observe that the points $y(\mu)$ that are approximated at each iteration of the barrier method satisfy

$$-\frac{b}{\mu} + \nabla\hat{\Phi}(y(\mu)) = 0.$$

Differentiating this equation with respect to μ yields

$$\frac{b}{\mu^2} + \nabla^2\hat{\Phi}(y(\mu))\dot{y}(\mu) = 0.$$

For given values of μ and $y(\mu)$ this is a linear equation that can be solved for $\dot{y}(\mu)$. (The matrix is the same as the one that is used in the Newton step for finding $y(\mu)$.) Given this observation we can state a more efficient predictor corrector method.

Dual predictor corrector method
Start: Find $y^{(0)}$ and $\mu_0 > 0$ with $y^{(0)} \approx y(\mu_0)$.
For $k = 1, 2, 3, \ldots$

- Choose $\Delta\mu_k \in (0, \mu_{k-1})$ such that $\hat{y}^{(k)} = y^{(k-1)} - \Delta\mu_k\dot{y}(\mu_{k-1})$ satisfies $C - \mathcal{A}^*(\hat{y}^{(k)}) \succeq 0$.

- Set $\mu_k = \mu_{k-1} - \Delta\mu_k$ and find $y^{(k)} \approx y(\mu_k)$ by Newton's method with line search starting at $\hat{y}^{(k)}$.

It turns out that $\dot{y}(\mu_{k-1})$ can be computed fairly accurately even if only an approximate point $y^{(k-1)} \approx y(\mu_{k-1})$ is known. For details see e.g. Jarre and Saunders (1993). This predictor corrector method is "reasonably efficient", but primal-dual approaches are more efficient in general.

We will generalize this method to nonlinear semidefinite programs in the next section

3. Nonlinear Semidefinite Programs

In this section we consider nonlinear semidefinite programs of the form

$$\text{maximize } b^T y \text{ s.t. } \mathcal{A}(y) \succeq 0, \quad f_i(y) \leq 0 \text{ for } 1 \leq i \leq m, \tag{5}$$

where $\mathcal{A} : I\!R^n \to \mathcal{S}^l$ is a smooth map and $f_i : I\!R^n \to I\!R$ are smooth functions. Note a slight change of notation, in this chapter $\mathcal{A}$ is a nonlinear operator, $\mathcal{A} : I\!R^n \to \mathcal{S}^l$.

We define a (possibly nonconvex) barrier function Φ,

$$\Phi(y) = -\log\det(\mathcal{A}(y)) - \sum_{i=1}^{m} \log(-f_i(y))$$

and local minimizers

$$y(\mu) = \text{local minimizer of } -\frac{b^T y}{\mu} + \Phi(y). \tag{6}$$

In slight abuse of notation we will denote any local minimizer by $y(\mu)$; this definition therefore does not characterize $y(\mu)$ uniquely.

Replacing $\hat{\Phi}$ with Φ, both, the barrier method and the predictor corrector method of Chapter 2.4 can also be applied to solve problem (5).

There are two questions regarding the efficiency of the predictor corrector method for solving (5). (The barrier method is certainly unpractical!)

- Does $\bar{y} = \lim_{k\to\infty} y^{(k)}$ exist, and if so, is $\bar{y}$ a "good" locally optimal solution of (5)?
- How quickly can $y^{(k)}$ be computed?

3.1 Issues of global convergence

As to the first question, one can show (see e.g. Jarre (2001)) that any accumulation point $\bar{y}$ of the sequence $y^{(k)}$ satisfies the Fritz-John

condition, (for a definition see e.g. Borgwardt (2001))

$$\exists u \geq 0,\ u \neq 0: \quad -u_0 b + \sum_{i=1}^{m} u_i \nabla f_i(\bar{y}) + u_{m+1} \nabla \det(\mathcal{A}(\bar{y})) = 0.$$

While this condition is reasonable in the absence of a constraint qualification it is not suitable for semidefinite programs. Indeed, whenever $\mathcal{A}(\bar{y})$ has the eigenvalue zero of multiplicity more than one, then $\nabla \det(\mathcal{A}(\bar{y})) = 0$, so that one can choose $u_{m+1} = 1$ and $u_i = 0$ for all other i.

A more appropriate convergence result therefore is

$$\begin{aligned} \nexists \Delta y: \quad & b^T \Delta y \geq 0, \\ & \nabla f_i(\bar{y}) \Delta y < 0 \ \text{ for all } y \text{ with } f_i(\bar{y}) = 0 \\ & \mathcal{A}(\bar{y}) + \varepsilon D\mathcal{A}(\bar{y})[\Delta y] \succ 0 \text{ for small } \varepsilon > 0. \end{aligned}$$

This result states that there does not exist any direction Δy starting at $\bar{y}$ that is strictly linearized feasible and does not increase the objective function.

Neither of the statements guarantees that $\bar{y}$ is a local minimizer. Indeed there are simple degenerate examples for which $\bar{y}$ is the global *maximizer* of (5). As shown in Jongen and Ruiz Jhones (1999), for nonlinear programs satisfying an LICQ condition and not containing "degenerate" critical points, the limit point of $y^{(k)}$ is a local minimizer. For such problems one can still construct examples, such that $\bar{y}$ is a very "poor" local minimizer. Nevertheless we believe that in many cases $\bar{y}$ is a minimizer whose objective value is "close" to the global minimum of (5). This intuition is motivated by the work Nesterov (1997). Nesterov considered the problem of minimizing a quadratic function over the ∞-norm unit cube. This problem may have very poor local minimizers (whose objective value is much closer to the global maximum value than it is to the global minimum). Nesterov shows that any local minimizer over a p-norm cube with a suitable value of $p = O(\log n)$ has much better global properties in the sense that it is at least as good as the result guaranteed by the semidefinite relaxation. Intuitively, this result is due to the fact that the p-norm cube "rounds" the vertices and edges of the ∞-norm cube. By this rounding procedure, the poor local minimizers disappear. In two dimensions the level sets of the logarithmic barrier function are almost indistinguishable from suitably scaled p-norm cubes. This leads us to believe that at least for quadratic minimization problems over the ∞-norm unit cube, a suitably implemented barrier method will also generate "good" local minimizers.

3.2 Efficiency of local minimization

Note that by definition, $y(\mu_k)$ is a local minimizer of (6), and hence, $\nabla^2\Phi(y(\mu_k)) \succeq 0$. In all of our test problems the iterates $y^{(k)} \approx y(\mu_k)$ satisfied the stronger condition $\nabla^2\Phi(y^{(k)}) \succ 0$. If this relation is satisfied the extrapolation step for computing $\hat{y}^{(k)}$ in the predictor corrector method can be carried out in the same way as in the convex case.

However, the iterates $y^{(k,i)}$ "on the way" from $\hat{y}^{(k)}$ to $y^{(k)}$ often do not satisfy $\nabla^2\Phi(y^{(k,i)}) \succ 0$. This implies that the concept of self-concordance that formed the basis of the dual barrier method and of the predictor corrector method for solving (D) is no longer applicable. While it is not yet possible to generalize the theory of self-concordance to nonconvex functions, it seems possible that the known Lipschitz continuity properties of $\nabla^2\hat{\Phi}$ carry over in some form to $\nabla^2\Phi$. The tool that was used for minimizing the barrier function involving $\hat{\Phi}$ in Section 2.2 is Newton's method. When $\nabla^2\Phi(y^{(k,i)}) \not\succ 0$, Newton's method with line search for approximating $y(\mu_k)$ is no longer applicable.

We need to find a suitable generalization of Newton's method to the nonconvex case involving Φ. For this generalization we keep the following properties in mind: The barrier subproblems that need to be solved at each step of the barrier method (or of the predictor corrector method) are systematically ill-conditioned. The condition number typically is $O(1/\mu)$, and the constant in the order notation is typically large. In addition, the computation of the Hessian matrices often is very expensive.

Possible minimization methods for approximating $y(\mu_k)$ include trust region methods with quasi-Newton updates of an approximate Hessian, see e.g. Conn et.al. (2000), continuation methods, or expensive plane search strategies as proposed in Jarre (2001).

In numerical examples it turned out that the minimization problems tend to be quite difficult and none of the minimization methods converge quickly. In particular, the barrier subproblems appear to be substantially more difficult to solve than in the convex case. We therefore address the complexity of smooth nonconvex local minimization. The next section shows that local minimization is $\mathcal{NP}$-hard in a certain sense.

3.3 Returning to the max cut problem

We return to the example in Chapter 2.2. As shown in Nesterov (1998) an inner approximation for the polyhedron $\mathcal{MC}$ is given by

$$\mathcal{NA} = \left\{ X \in \mathcal{SDP} \mid \sin\left[\frac{\pi}{2}X\right] \succeq 0 \right\}.$$

Here, the square brackets $\sin\left[\frac{\pi}{2}X\right]$ are used to indicate that the sin function is applied *componentwise* to each of the matrix entries of $\frac{\pi}{2}X$. The set $\mathcal{NA}$ is formed from $\mathcal{SDP}$ using the function $c : [-1,1] \to [-1,1]$ with $c(t) = \sin\left(\frac{\pi}{2}t\right)$. This function is a nonlinear "contraction" in the sense that $|c(t)| \leq |t|$.

It is somewhat surprising to find out that $\mathrm{conv}(\mathcal{NA}) = \mathcal{MC}$, i.e.

$$\mathcal{NA} \subset \mathrm{conv}(\mathcal{NA}) = \mathcal{MC} \subset \mathcal{SDP}.$$

see Nesterov (1998).

A simple picture can explain the relationship of $\mathcal{MC}$, $\mathcal{SDP}$, and $\mathcal{NA}$.

The set $\mathcal{MC}$ is a polytope whose precise description is not known in spite of its simple structure. (More precisely, there does not exist any known polynomial time algorithm which, given a point X, either returns a certificate proving that $X \in \mathcal{MC}$ or returns a separating hyperplane.)

The set $\mathcal{SDP}$ is obtained by "inflating" the set $\mathcal{MC}$ while keeping all faces of dimension $\leq n-2$ fixed. Like a balloon we "pump up" the hull of $\mathcal{MC}$ while keeping certain low-dimensional boundary manifolds fixed. (Note that $\mathcal{MC}$ has dimension $n(n-1)/2$.) The set $\mathcal{SDP}$ is convex and is "efficiently representable", i.e. there exist efficient numerical algorithms for minimizing convex functions over $\mathcal{SDP}$.

The set $\mathcal{NA}$ is obtained by shrinking $\mathcal{SDP}$ in a certain nonlinear fashion. This shrinkage is done in a certain optimal way such that all boundary manifolds of dimensions 1 and 2 of $\mathcal{MC}$ are contained in $\mathcal{NA}$. In particular, for $n = 3$ we have $\mathcal{MC} = \mathcal{NA}$, see Hirschfeld and Jarre (2001).

The set $\mathcal{NA}$ is bounded by two smooth constraints, is star shaped, contains a ball of radius 1, and is contained in a ball of radius n. By our previous considerations,
any locally optimal vertex of

$$\textit{minimize } C \bullet X \quad \textit{s.t.} \quad X \in \mathcal{NA} \tag{7}$$

solves the max cut problem (3).

Hence, in spite of the nice properties of $\mathcal{NA}$, it must be very difficult to find a local optimal vertex of (7) or to check whether a given vertex is a *local* minimum.

Note that (7) is a nonlinear semidefinite program. The difficulty of the local minimization of (7) is due to the fact that problem (7) suffers from a systematic violation of any constraint qualification. It contains many "peaks" similar to the one in

$$\left\{x \in I\!R^2 \mid \; x \geq 0, \; x_2 \leq x_1^3\right\}.$$

In higher dimensions such peaks become untractable.

3.4 Finding an ϵ-KKT-point

In a second example, see Hirschfeld and Jarre (2001), the so-called chained Rosenbrock function $f : \mathbb{R}^n \to \mathbb{R}$

$$f(x) = (x_1 - 1)^2 + 100 \sum_{i=2}^{n} (x_i - x_{i-1}^2)^2$$

(see also Toint (1978)) has been tested. This function has only one local minimizer which is also the global minimizer, $x = (1, \ldots, 1)^T$. Applying various trust region methods for minimizing f starting at $x^{(0)} = (-1, 1, \ldots, 1)^T$ results in running times that appear to be exponential in n. (These running times are purely experimental, and due to time limitations could only be tested for small values of n.)

At first sight this result seems to contradict a statement by Vavasis. In the paper Vavasis (1993) the following result is shown.

Consider the problem

$$\text{minimize } f(x) \quad \text{s.t.} \quad -1 \leq x_i \leq 1 \quad \text{for } 1 \leq i \leq n. \tag{8}$$

Vavasis assumes that the gradient ∇f is Lipschitz continuous with Lipschitz constant M and considers the problem of finding an ε-KKT point for (8). He presents an algorithm that takes at most $O(\frac{nM}{\varepsilon})$ gradient evaluations to find an ε-KKT point. This bound is exponential with respect to the number of digits of the required accuracy, i.e. with respect to "$-\log \varepsilon$", but linear with respect to n.

He also presents a class of functions of two variables for which *any* algorithm has a worst case complexity of at least $O(\sqrt{\frac{M}{\varepsilon}})$ gradient evaluations to find an ε-KKT point.

The conditions of Vavasis' paper apply to the Rosenbrock example as well. All points at which this function is evaluated by the trust region algorithms lie in the box $-1 \leq x_i \leq 1$, and moreover, Rosenbrocks function possesses moderately bounded norms of $\nabla^2 f$ at these points implying that M is consistently small. The reason for the observed exponential growth of the number of iterations lies in the fact that the norms of the gradients do become small very quickly (as predicted by Vavasis even for a steepest descent method), but for large n, the norm of ∇f needs to be extremely small to guarantee that the iterate is close to a local minimizer. Thus the exponential growth with respect to the number of variables is due to the fact that the ε-KKT condition is a poor condition for large n. (We don't know of any better condition though!) More results on local minimization issues are discussed in the forthcoming paper Hirschfeld and Jarre (2001).

4. Conclusion

We have highlighted some issues of nonlinear semidefinite programming related to a dual barrier method. In particular we have raised the questions of smoothness, regularity, and computational complexity related to semidefinite programs. As preliminary numerical results in Jarre (2001) indicate, variants of the predictor corrector method of the present paper are reasonably fast for medium size problems (up to 500 unknowns). The numerical results were also compared with the ones in Fukuda and Kojima (2001). In all examples it turned out that the method proposed in this paper converged to the global minimizer. This gives some further weak evidence that the method is indeed unlikely to be "trapped" near poor local minimizers. We also indicated that the local convergence of solving the barrier subproblems in the predictor corrector method is slow; improvements of this convergence behavior are the subject of future research.

References

F. Alizadeh, "Combinatorial Optimization with Interior Point Methods and Semidefinite Matrices" PhD Thesis, University of Minnesota (1991).

Optimierung, Operations Research, Spieltheorie Birkhäuser Verlag (2001).

S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control Theory*, Volume 15 of *Studies in Applied Mathematics* SIAM Philadelphia, PA, (1994).

A.R. Conn, N.I.M. Gould and Ph.L. Toint, *Trust-Region Methods* MPS/SIAM Series on Optimization, SIAM, Philadelphia, (2000).

A. Forsgren, "Optimality conditions for nonconvex semidefinite programming", Mathematical Programming 88 (2000), 105-128.

R.W. Freund and F. Jarre, "An Extension of the Positive Real Lemma to Descriptor Systems" Report 00/3-09, Scientific Computing Interest Group, Bell Labs, Lucent Technologies, (2000)

M. Fukuda and M. Kojima, "Branch-and-Cut Algorithms for the Bilinear Matrix Inequality Eigenvalue Problem", *Computational Optimization and Applications* Vol.19, No.1, pp.79–105, (2001).

M.X. Goemans and D.P. Williamson, ".878-Approximation Algorithm for MAX CUT and MAX 2SAT", in *ACM Symposium on Theory of Computing (STOC)*, (1994).

C. Helmberg, F. Rendl, R.J. Vanderbei, "An Interior-Point Method for Semidefinite Programming" *SIAM J. Optim.* 6(2):342–361 (1996).

M.W. Hirsch and S. Smale *Differential equations, dynamical systems, and linear algebra* Acad. Press, New York, (1974).

B.W. Hirschfeld and F. Jarre, "Complexity Issues of Smooth Local Minimization", Technical Report, Universität Düsseldorf, in preparation (2001).

"Interior-Point Algorithms for Classes of Convex Programs" in T. Terlaky ed. *Interior Methods of Mathematical Programming* Kluwer (1996).

F. Jarre, "A QQP-Minimization Method for Semidefinite and Smooth Nonconvex Programs", Technical Report, University of Düsseldorf, Germany, to appear in revised form in *Optimization and Engineering* (2001).

F. Jarre and M.A. Saunders, "A Practical Interior-Point Method for Convex Programming", *SIAM J. Optim.* 5(1) pp.149–171 (1995).

H.T. Jongen and A. Ruiz Jhones, "Nonlinear Optimization: On the Min-Max Digraph and Global Smoothing", in A. Ioffe, S. Reich, I. Shafrir eds *Calculus of Variations and Differential Equations* Chapman & Hall / CRC Research Notes in Mathematics Series, Vol 410, CRC Press, (UK) LLC, pp.119–135, (1999).

M. Kocvara and M. Stingl, "Augmented Lagrangian Method for Semidefinite Programming" forthcoming report, Institute of Applied Mathematics, University of Erlangen-Nuremberg (2001).

F. Leibfritz, "A LMI-based algorithm for designing suboptimal static / output feedback controllers", *SIAM Journal on Control and Optimization*, Vol. 39, No. 6, pp. 1711 - 1735, (2001).

Y.E. Nesterov, Talk given at the Conference on Semidefinite Optimization, ZIB Berlin, (1997).

Y.E. Nesterov, "Semidefinite Relaxation and Nonconvex Quadratic Optimization", *Optimization Methods and Software* 9, pp.141–160, (1998).

Y.E. Nesterov and A.S. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM Publications, SIAM Philadelphia, PA, (1994).

G. Pataki, "The Geometry of Semidefinite Programming", in H. Wolkowicz, R. Saigal, L. Vandenberghe eds *Handbook of Semidefinite Programming: Theory, Algorithms and Applications* Kluwer's International Series (2000).

C. Scherer, "Lower bounds in multi-objective H_2/H_∞ problems", *Proc. 38th IEEE Conf. Decision and Control*, Arizona, Phoenix (1999).

A. Shapiro and K. Scheinberg, "Duality and Optimality Conditions", in H. Wolkowicz, R. Saigal, L. Vandenberghe eds *Handbook of Semidefinite Programming: Theory, Algorithms and Applications* Kluwer's International Series (2000).

Ph.L. Toint, "Some Numerical Result Using a Sparse Matrix Updating Formula in Unconstrained Optimization", *Mathematics of Computation*, vol. 32(143), pp. 839-851, (1978).

R.J. Vanderbei, "LOQO User's Manual – Version 3.10" . Report SOR 97-08, Princeton University, Princeton, NJ 08544, (1997, revised 10/06/98).

R.J. Vanderbei, H. Benson and D. Shanno, "Interior-Point Methods for Nonconvex Nonlinear Programming: Filter Methods and Merit Functions", Report ORFE 00-06, Princeton University, Princeton, NJ 08544, (2000).

S.A. Vavasis, "Black-Box Complexity of Local Minimization", *SIAM J. Optim.* 3(1) pp.60–80, (1993).

S.J. Wright and J. Nocedal, *Numerical Optimization*, Springer Verlag, (1999).