Relaxations of Quadratic Programs in Operator Theory and System Analysis

Alexandre Megretski

Abstract. The paper describes a class of mathematical problems at an intersection of operator theory and combinatorics, and discusses their application in complex system analysis. The main object of study is *duality gap bounds* in *quadratic programming* which deals with problems of maximizing quadratic functionals subject to quadratic constraints. Such optimization is known to be universal, in the sense that many computationally hard questions can be reduced to quadratic programming. On the other hand, it is conjectured that an efficient algorithm of solving general non-convex quadratic programs exactly does not exist.

A specific technique of "relaxation", which essentially replaces deterministic decision parameters by random variables, is known experimentally to yield high quality approximate solutions in some non-convex quadratic programs arising in engineering applications. However, proving good error bounds for a particular relaxation scheme is usually a challenging mathematical problem. In this paper relaxation techniques of dynamical system analysis will be described. It will be shown how operator theoretic methods can be used to give error bounds for these techniques or to provide counterexamples. On the other hand, it will be demonstrated that some difficult problems of operator theory have equivalent formulations in terms of relaxation bounds in quadratic programming, and can be approached using the insights from combinatorics and system theory.

1. Introduction

It is always intriguing to discover that interest in a specific class of difficult mathematical problems is shared by seemingly distant fields of pure and applied research. Usually there is also some hope that exchanging the ideas and open questions between the fields will eventually lead to new results. This paper is written exactly with this idea in mind. It contains few new theorems, and is instead concentrated on showing similarity between different problems.

The main object of study in this paper is *duality gap bounds* in *quadratic programming*. Quadratic programming, which appears naturally in numerous applications, deals with problems of maximizing quadratic functionals subject to quadratic constraints. Such optimization is known to be universal (even very special

settings lead to NP complete problems, [4]), which guarantees that many computationally hard questions can be reduced to quadratic programming. On the other hand, it is conjectured that an efficient algorithm of solving general non-convex quadratic programs exactly does not exist (P≠NP).

A specific technique of "relaxation", which essentially replaces deterministic decision parameters by random variables, is known experimentally to yield high quality approximate solutions in some non-convex quadratic programs arising in engineering applications. The error induced by such relaxation can be interpreted as duality gap in non-convex optimization. Proving good duality gap bounds for a particular relaxation scheme (equivalent quadratic programs may lead to non-equivalent relaxations) is usually a challenging mathematical problem. An important point to be made by this paper is that some well known problems from operator theory can be formulated exactly as problems of proving some particular relaxation error bounds.

As an application motivation, this paper describes relaxation techniques of dynamical system analysis and design. It will be shown how the search for a Lyapunov function, or for another type of stability certificate, leads naturally to non-convex quadratic programming. In fact, a number of modern techniques in system theory are primarily relaxation methods. Ideas from operator theory can be used to give error bounds for these techniques or to provide counterexamples.

The paper is organized as follows. First, a general description of quadratic programs and relaxation techniques is given, and some well known duality gap estimation problems are formulated. Next, a more detailed account of the use of quadratic programming and relaxations in system analysis is presented. Finally, some proofs and counterexamples are given for special duality gap bounds.

Notation and Terminology

In this paper matrices with real entries are used to describe finite dimensional vectors and their linear transformations. $\mathbf{R}^n$ denotes the set of single-column real matrices of length n. S_n denotes the set of symmetric n-by-n matrices with real entries. The prime $'$ means transposition or Hermitian conjugation (for complex matrices), so that $\sigma(x) = x'Qx$ defines a real-valued quadratic form on $\mathbf{R}^n = \{x\}$ for any $Q \in S_n$. For $Q \in S_n$, $Q \geq 0$ means that $x'Qx \geq 0$ for all $x \in \mathbf{R}^n$. A *quadratic functional* on $\mathbf{R}^n$ is any function $\sigma : \mathbf{R}^n \to \mathbf{R}$ defined by

$$\sigma(x) = \begin{bmatrix} x \\ 1 \end{bmatrix}' \Sigma \begin{bmatrix} x \\ 1 \end{bmatrix},$$

where $\Sigma \in S_{n+1}$ is a constant matrix. A *quadratic S_m-valued functional* on $\mathbf{R}^n$ is any function $\alpha : \mathbf{R}^n \to S_m$ such that $\sigma_v(x) = v'\alpha(x)v$ is a quadratic functional on $\mathbf{R}^n$ for any fixed $v \in \mathbf{R}^m$.

Vector-valued random variables will be used in this paper. $\mathbf{E}\eta$ denotes the expected value of an integrable vector random variable η.

2. The Duality Gap Bound Problem

In this section, a general description of quadratic programs and relaxation techniques is given, and some well known duality gap estimation problems are formulated.

2.1. Relaxation and Duality Gap

This subsection gives specific definitions of quadratic programs, relaxed quadratic programs, and duality gaps in quadratic programming.

2.1.1. Quadratic Programs. Let α, β be two symmetric S_m-valued quadratic functions on $\mathbf{R}^n$ such that $\alpha(x) \geq 0$ and $\alpha(x) \neq 0$ for all x. Any such pair (α, β) defines a *quadratic program* – the problem of finding the supremum of the functional

$$J(x) = \sup\{\lambda : \ \beta(x) - \lambda\alpha(x) \geq 0\} \to \sup\,. \tag{1}$$

The variable x with respect to which the optimization is performed is called *decision variable*. The supremum of J is denoted by J_*.

Example 2.1. *The problem of finding the maximum of $x_1x_2 + x_2x_3 - x_1x_3$ where the real variables x_1, x_2, x_3 range over the interval $[-1, 1]$ can be viewed as a special case of quadratic program (1) with $n = 3$, $m = 4$,*

$$\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \beta = \begin{bmatrix} x_1x_2 + x_2x_3 - x_1x_3 & 0 & 0 & 0 \\ 0 & 1 - x_1^2 & 0 & 0 \\ 0 & 0 & 1 - x_2^2 & 0 \\ 0 & 0 & 0 & 1 - x_3^2 \end{bmatrix}.$$

In this case the supremum J_* of $J(x)$ equals 1.

2.1.2. Quadratic Optimization on a Hypercube. Let $u_1, u_2, \ldots, u_n$ be given vectors in a real Hilbert space. How large and how small can be the length of the sum

$$u = x_1u_1 + x_2u_2 + \cdots + x_nu_n, \quad x_k = \pm 1$$

when the "sign" coefficients $x_k \in \{-1, 1\}$ can be selected arbitrarily? Note that already the special case of this problem when u_k are finite dimensional vectors with integer coefficients, is known to be NP-complete.

Both questions (maximization and minimization) can be reduced to the same problem of finding maximum of a quadratic form over the set of vertices of a hypercube:

$$\sigma(x) = x'Qx \to \max, \quad \text{subject to} \quad x \in \{-1, 1\}^n, \tag{2}$$

where $Q \in S_n$ is a given symmetric matrix. Here $Q_{ij} = \langle u_i, u_j\rangle$ when the length of u is maximized, and $Q_{ij} = -\langle u_i, u_j\rangle$ when the length of u is minimized.

This problem is frequently referred to as the MAX-CUT problem, because of the following interpretation. Let the indexes from 1 to n correspond to nodes on a graph. Let $-2Q_{ij}$ be the "value of the benefit" of "cutting off" the edge connecting

nodes i and j. Any decision vector $x \in \{-1,1\}^n$ defines a partition of the graph's nodes into two sets N_+ and N_-, the i-th node being in N_+ if and only if $x_i = 1$. If any edge connecting a node from N_+ and a node from N_- must be "cut off", maximizing $x'Qx$ maximizes the total benefit of dividing the graph's nodes into two groups and cutting the edges between the groups.

Let $\bar{Q}$ denote the matrix obtained from Q by replacing all its diagonal terms with zeros. Then

$$x'Qx = x'\bar{Q}x + \mathrm{tr}[Q] \quad \forall\ x \in \{-1,1\}^n.$$

Therefore, the maximums of $x'Qx$ and $x'\bar{Q}x$ on $\{-1,1\}^n$ are achieved on the same vectors x, and their difference is a known constant.

It is frequently more convenient to consider the equivalent problem of maximizing $x'\bar{Q}x$ rather than that of optimizing $x'Qx$, due to the following simple observation.

Lemma 2.2. *Any argument of maximum of $x'\bar{Q}x$ over the set $\{-1,1\}^n$ of vertices of the hypercube $[-1,1]^n$ is also an argument of maximum in the optimization problem*

$$\sigma(x) = x'\bar{Q}x \to \max, \quad \text{subject to} \quad x \in [-1,1]^n, \tag{3}$$

i.e. that of maximizing $x'\bar{Q}x$ over the whole hypercube $[-1,1]^n$. In particular, the maximum of $x'\bar{Q}x$ over $\{-1,1\}^n$ is always non-negative.

Proof. Since the diagonal elements of $\bar{Q}$ are zero, $x'\bar{Q}x$ is *linear* with respect to any single component x_i of x. Hence maximum of $x'\bar{Q}x$ over $[-1,1]^n$ can be achieved with all $x_i \in \{-1,1\}$. □

Just as in Example 2.1, problem (3) can be shown to be equivalent to a quadratic program (1) with

$$\alpha(x) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ & & \ddots & \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \beta(x) = \begin{bmatrix} x'\bar{Q}x & 0 & 0 & 0 \\ 0 & 1-x_1^2 & 0 & 0 \\ & & \ddots & \\ 0 & 0 & 0 & 1-x_n^2 \end{bmatrix}.$$

2.1.3. Semidefinite Programming. In the special case when $\alpha(x), \beta(x)$ are *affine* functions of x problem (1) is called a *semidefinite program*. Semidefinite programs are much easier to solve than general quadratic programs. An informal explanation is that in this case the function J is *quasi-convex*, i.e. its level sets are convex. In general, one would expect convex optimization problems to have relatively simple solutions. At least, this turns out to be the case for semidefinite programs, which admit polynomial time solution algorithms [7, 1].

2.1.4. RELAXATION. Due to their universality (NP completeness means that virtually any optimization problem can be re-written as an equivalent quadratic program) it is very unlikely that a "smart" way of solving finite dimensional quadratic programs both accurately and efficiently can ever be found. For most quadratic programs, efficient solution algorithms are not available. However, there exists a simple way of finding efficiently an upper bound of $J = J(x)$. This method is based on replacing vector x by a random variable w with a finite second moment, i.e. such that $\mathbf{E}|w|^2 < \infty$, and also replacing quadratic matrix-valued functionals $\alpha(x), \beta(x)$ by their expected values:

$$\hat{J}(w) = \sup\{\lambda : \ \mathbf{E}\beta(w) - \lambda\mathbf{E}\alpha(w) \geq 0\} \to \sup. \tag{4}$$

Since w can be a random variable which takes a single value $w = x$ with probability 1, where $x \in \mathbf{R}^n$ can be chosen arbitrarily, the supremum $\hat{J}_*$ of $\hat{J}$ cannot be smaller than the supremum J_* of J.

For an arbitrary square integrable random n-vector w let its *covariance and mean value matrix* W be defined by

$$W = \mathbf{E}\left[\begin{array}{c} w \\ 1 \end{array}\right]\left[\begin{array}{c} w \\ 1 \end{array}\right]'.$$

Note that any $W \in S_{n+1}$ is a covariance and mean value matrix of some random n-vector w if and only if the conditions

$$W \geq 0, \quad W_{n+1,n+1} = 1 \tag{5}$$

are satisfied. The expected value of a quadratic functional of w is a linear function of W, according to

$$\mathbf{E}\left[\begin{array}{c} w \\ 1 \end{array}\right]'\Sigma\left[\begin{array}{c} w \\ 1 \end{array}\right] = \mathrm{tr}[\Sigma W].$$

Hence the matrices $\mathbf{E}\alpha(w), \mathbf{E}\beta(w)$ depends *linearly* on W, and the relaxed quadratic program (4) is a *semidefinite program* with respect to W.

Example 2.3.

$$x_1x_2 + x_2x_3 - x_1x_3 \to \sup, \quad \text{subject to} \quad x_k^2 \leq 1$$

will be "relaxed" to

$$W_{12} + W_{23} - W_{13} \to \sup, \quad \text{subject to} \quad W_{kk} \leq 1, \quad W \geq 0,$$

where $W = (W_{ij})_{i,j=1}^4$ *is a symmetric matrix such that* $W_{44} = 1$. *In fact, the last column and the last row of* W *do not have any effect on the relaxed problem formulation, and hence* W *can be assumed to range over* S_3. *The supremum in the relaxed problem equals* $3/2$.

2.1.5. Non-Uniqueness of Relaxations. An important observation is that *equivalent* quadratic programs frequently define non-equivalent relaxations. As a rule, equivalent quadratic programs are obtained by adding redundant constraints to a quadratic program.

Example 2.4. *Let $Q \in S_n$ be a symmetric matrix. Consider the problem of maximizing $J(x) = x'Qx$ over the unit cube $\{x\} = [-1,1]^n$. One way to pose this problem as a quadratic program, as shown before, is by using diagonal $\alpha(x), \beta(x) \in S_{n+1}$, where $\alpha_{11} = 1$, $\alpha_{ii} = 0$ for $i > 1$, $\beta_{11}(x) = x'Qx$, and $\beta_{ii}(x) = 1 - x_{i-1}^2$ for $i > 1$. The standard relaxation of this setup is equivalent to maximizing $\mathrm{tr}[QW]$ over those $W \in S_n$ for which $W \geq 0$ and $W_{ii} \leq 1$.*

On the other hand, simple inspection shows that $x_ix_j + x_jx_k - x_ix_k \in [-1,1]$ for all $x_i, x_j, x_k \in [-1,1]$. Let us define $\bar\alpha(x), \bar\beta(x)$ as diagonal matrices from S_N, where $N = n + 1 + n(n-1)(n-2)/2$, such that the first $n+1$ diagonal elements of $\bar\alpha(x), \bar\beta(x)$ are same as those of $\alpha, \beta(x)$, and the remaining $n(n-1)(n-2)/2$ diagonal elements equal zero for $\bar\alpha$, and equal $1 - x_ix_j - x_jx_k + x_ix_k$ for $\bar\beta(x)$, where (i,j,k) ranges over the set of all ordered triples with $i > j$ and $k \notin \{i,j\}$ (a total of $n(n-1)(n-2)/2$ triples).

By construction, the quadratic program defined by $\bar\alpha(x)$ and $\bar\beta(x)$ is equivalent to the quadratic program defined by $\alpha(x)$ and $\beta(x)$. However, it is easy to see that the standard relaxation of the latter is equivalent to maximizing $\mathrm{tr}[QW]$ over those $W \in S_n$ for which $W \geq 0$, $W_{ii} \leq 1$, and $W_{ij} + W_{jk} - W_{ik} \leq 1$. As it is shown by Example 2.3, the maximum in the new relaxation can be strictly less than the maximum in the original relaxation.

Given a quadratic program, usually there are numerous ways to improve the quality of relaxation-based optimization by formulating equivalent quadratic programs with redundant constraints. As a rule, such improvement comes at a cost of solving larger semidefinite programs.

2.1.6. The Duality Gap. Another approach to finding an upper bound for the supremum in (1) is based on introducing Lagrange multipliers and solving the corresponding dual problem defined as

$$\bar J(Z) = \inf\{\bar\lambda : \ \mathrm{tr}[Z(\bar\lambda\alpha(x) - \beta(x))] > 0 \ \ \forall\, x \in \mathbf{R}^n\} \to \inf_{Z \in S_m, Z \geq 0}. \qquad (6)$$

Lemma 2.5. *Assume there exists a bounded optimizing sequence in problem (1), i.e. $x_i \in \mathbf{R}^n$, $\sup|x_i| \leq r < \infty$, and $J(x_i) \to J_* = \inf J$. Then the infimum in (6) equals the supremum in (4).*

Proof. Let $Z \in S_m$, $Z \geq 0$ and $\bar\lambda \in \mathbf{R}$ be such that $\mathrm{tr}[Z(\bar\lambda\alpha(x) - \beta(x))] > 0$ for all x. Then for any random vector w and $\lambda \in \mathbf{R}$ such that $\mathbf{E}\beta(w) \geq \lambda\mathbf{E}\alpha(w)$ it follows

$$\lambda\mathbf{E}\mathrm{tr}[Z\alpha(w)] \leq \mathbf{E}\mathrm{tr}[Z\beta(w)] < \bar\lambda\mathbf{E}\mathrm{tr}[Z\alpha(w)].$$

Since $\mathbf{E}\mathrm{tr}[Z\alpha(w)] \geq 0$, this implies $\mathbf{E}\mathrm{tr}[Z\alpha(w)] > 0$ and hence $\lambda < \bar\lambda$. Therefore the infimum in (6) is not smaller than the supremum in (4). To prove the opposite inequality, assume that the inequality $\mathbf{E}\beta(w) \geq \lambda_*\mathbf{E}\alpha(w)$ is impossible for some

$\lambda_* \in \mathbf{R}$. Note that the set B of all expected values $b = \mathbf{E}(\beta(w) - \lambda_*\alpha(w))$, where w is a square integrable random vector and $\mathbf{E}|w|^2 \le r^2$, is convex and compact. Hence there exists a hyperplane in S_m which separates B from the convex cone of all positive semidefinite matrices. In other words, there exists $Z \in S_m$, $Z \ge 0$, $Z \ne 0$ such that $\mathrm{tr}[Zb] < 0$ for any $b \in B$. Hence $\mathrm{tr}[Z(\beta(x) - \lambda_*\alpha(x))] < 0$ for any $x \in \mathbf{R}^n$, $|x| \le r$, i.e. $\bar{J}(Z) \ge \lambda_*$. Hence the infimum in (6) is not larger than the supremum in (4). □

As a consequence of Lemma 2.5, the difference between the maxima in relaxed and non-relaxed versions of the same quadratic program can be interpreted as a *duality gap* in the original non-convex optimization problem.

2.1.7. The Duality Gap Estimation Problem. While the use of quadratic programs and their relaxations is a critical issue in many applications, a large class of difficult *mathematical questions* is associated with finding general a-priory bounds for the relaxation gap. This paper is devoted to the problem, called the *duality gap estimation problem*: *find good upper bounds of the relaxation gap in special classes of quadratic programs.*

Let J_* and $\hat{J}_*$ be the maxima in a quadratic program and its semidefinite program relaxation respectively. In a typical application, J_* is unknown, while $\hat{J}_*$ is known and serves as an *upper* bound of J_*. An estimate of the duality gap usually has the form of an inequality

$$J_* \ge \rho(\hat{J}_*, \alpha, \beta),$$

where $\rho(\cdot)$ is an *easily computable* function of its arguments (of course, J_* itself is a function of α, β, but it is presumed that J_* is difficult to compute).

When J_* is known to be non-negative, the duality gap bound frequently has the form

$$J_* \ge \rho^{-1}\hat{J}_*$$

where ρ is a constant which is the same for a class of quadratic programs. By the definition, ρ is a measure of success of the relaxation technique, applied to a particular quadratic program. When ρ equals 1, the quadratic program is essentially equivalent to its relaxed form. When $\rho \gg 1$, the relaxation technique is not accurate.

Let $N = n + m$ denote the "size" of quadratic program (1). An *informal* evaluation of the difficulty of obtaining duality gap bounds $\rho = \rho(N)$ is as follows:

- **"No gap".** The case of no gap bound ($\rho(N) = 1$) is very rare, but usually not difficult to prove, if true.
- **"Bounded gap".** The cases when $\rho(N) < c$ where c does not depend on n are rare. The corresponding duality gap estimates are usually difficult to prove.
- **"Logarithmic gap".** The case when a logarithmic growth bound, such as $\rho(N) < c\log(N)$, is available, is of interest when unboundedness of $\rho(N)$ can be proven.

- **"Power gap".** As a rule, a power law growth bound $\rho(N) < c \cdot N^r, r > 0$ can be established easily for many quadratic programs.

2.2. Polynomially Bounded Operators

A linear operator $A: \ H \to H$ on a Hilbert space H is called *polynomially bounded* if there exists a constant M such that

$$\|p(A)\| \le M\|p\|_\infty$$

for any scalar polynomial p, where

$$\|p\|_\infty = \max\{|p(z)|: \ |z| \le 1\}$$

is the so-called infinity norm. If A is *similar to a contraction*, i.e. can be represented in the form $A = STS^{-1}$ where $\|T\| \le 1$ and $\|S\|, \|S^{-1}\| < \infty$, then, due to a von Neumann theorem, A is polynomially bounded with $M \le \|S\|\cdot\|S^{-1}\|$. The problem posed by P. Halmos in [3] and finally solved by G. Pisier in [9] asks whether *any* polynomially bounded operator A is similar to a contraction.

The intention of this subsection is to show that a finite dimensional version of the problem is equivalent to verifying boundedness of a relaxation gap in a particular class of quadratic programs.

2.2.1. Finite Dimensional Formulation. Let A be an n-by-n matrix. A is called *marginally stable* if the norms of A^k, $k > 0$, are bounded. While it appears that any alternative definition of marginal stability for finite dimensional matrices will lead to the same condition, there is a number of definitions of numerical *degree of stability* which are far from being equivalent when $n \to \infty$.

The following measures of marginal stability will be studied in this subsection:

$$\gamma_{pb}(A) = \sup\{\|p(A)\|: \ p \in \mathbf{P}, \ \|p\|_\infty \le 1\},$$

where $\mathbf{P}$ is the set of scalar polynomials,

$$\gamma_{cb}(A) = \sup\{\|p(A)\|: \ p \in \mathbf{P}_*, \ \|p\|_\infty \le 1\},$$

where $\mathbf{P}_*$ is the set of matrix polynomials,

$$\gamma_{sc}(A) = \inf\{\|S^{-1}\|: \ \|SAS^{-1}\| \le 1, \ \|S\| \le 1\}.$$

It is easy to show that all three quantities $\gamma_{pb}(A)$, $\gamma_{cb}(A)$ and $\gamma_{sc}(A)$ are finite for all marginally stable n-by-n matrices A.

Let the function $\phi = \phi(M, n)$ be defined for $M \ge 1$ and $n = 1, 2, \dots$ by

$$\phi(M, n) = \sup\{\gamma_{sc}(A): \ \gamma_{pb}(A) \le M\}, \tag{7}$$

where the supremum is taken over all n-by-n marginally stable matrices A.

The finite dimensional version of the Halmos problem studied here asks whether $\phi(M, n)$ is bounded as $n \to \infty$ for any fixed $M > 1$.

2.2.2. Schur Matrices and Lyapunov Equations. Note that it is sufficient to consider the case when A in (7) is a *strictly stable* (Shur) matrix, i.e. all eigenvalues lie strictly within the unit disc. Indeed, if A is marginally stable but not strictly stable then for any $r \in (0,1)$ the matrix $A_r = rA$ is strictly stable and polynomially bounded. Hence $\|S_r A_r S_r^{-1}\| \leq 1$ for some matrices S_r such that $\|S_r\| = 1$, $\|S_r^{-1}\| \leq \phi(M,n)$. As $r \to 1$, S_r will converge to a matrix S such that $\|SAS^{-1}\| \leq 1$, $\|S\| \leq 1$, and $\|S^{-1}\| \leq \phi(M,n)$.

Given an n-by-n Schur matrix A, for any $X \in S_n$ there exists the *unique* solution $Q = \mathbf{L}_A(X)$ of the *Lyapunov equation*

$$Q - AQA' = X.$$

An alternative way to define Q is by

$$\mathbf{L}_A(X) = \sum_{k=0}^{\infty} A^k X (A')^k.$$

In particular, $Q \geq 0$ whenever $X \geq 0$. The opposite is not true: there exist matrices X which are not positive semidefinite such that $\mathbf{L}_A(X) \geq 0$ for *some* Schur matrices A.

2.2.3. Polynomial Boundedness and Lyapunov Equations. The following statement relates positive semidefiniteness of solutions of Lyapunov equations to polynomial boundedness and complete boundedness.

Lemma 2.6. *Let A be a strictly stable n-by-n matrix, $w_1, \ldots, w_k$, $v_1, \ldots v_m \in \mathbf{R}^n$. Let $\bar{v} \in \mathbf{R}^{mn}$ be the column vector obtained by stacking v_i's one over the other. Let $\bar{w} \in \mathbf{R}^{kn}$ be constructed similarly from w_i's:*

$$\bar{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}, \quad \bar{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_k \end{bmatrix}.$$

A sequence $\{p_i\}$ of polynomial k-by-m matrices $p_i \in \mathbf{P}_*$ such that $\|p_i\|_\infty \leq 1$ and $p_i(A)\bar{v} \to \bar{w}$ as $i \to \infty$ exists if and only if $Q \geq 0$ where Q is the solution of the Lyapunov equation

$$Q - AQA' = V - W, \tag{8}$$

with

$$V = \sum_{i=1}^{m} v_i v_i', \quad W = \sum_{i=1}^{k} w_i w_i'. \tag{9}$$

Proof. Let $l_+^2(\mathbf{R}^r)$ denote the standard Hilbert space of one-sided sequences $(x_i)_{i=0}^{\infty}$ of vectors $x_i \in \mathbf{R}^r$. Let τ denote the backward shift operator on $l_+^2(\mathbf{R}^r)$. Let H_v

be the (finite dimensional) τ-invariant linear subspace of $l_+^2(\mathbf{R}^m)$:

$$H_v = \left\{ x_v = x_v(z) = (x_i) : \ x_i = \begin{bmatrix} v_1'(A')^i z \\ \vdots \\ v_m'(A')^i z \end{bmatrix}, \ z \in \mathbf{R}^n \right\}.$$

Consider the map $L_w : \ H_v \to l_+^2(\mathbf{R}^k)$ defined by $L_w x_v(z) = x_w(z)$ where $x_w(z)$ is defined similarly to x_v using vectors w_j.

If $\bar{w} = p(A)\bar{v}$ where $p \in \mathbf{P}_*$ and $\|p\|_\infty \leq 1$ then $L_w = p(\tau)|_{H_v}$ is correctly defined and contractive. Hence

$$\|x_v(z)\|^2 \geq \|x_w(z)\|^2 \ \ \forall \ z \in \mathbf{R}^n. \tag{10}$$

Since by construction

$$\|x_v(z)\|^2 - \|x_w(z)\|^2 = z'Qz,$$

the inequality $Q \geq 0$ follows.

If $Q \geq 0$ then (10) holds, and hence L_w is a correctly defined contraction which commutes with τ. Hence L_w can be extended to the whole $l_+^2(\mathbf{R}^m)$ preserving contractiveness and commutativity. The extension $\bar{L}_w$ will have the form $x \to h(\tau)x$ where h is a rational matrix function with $\|h\|_\infty \leq 1$. Now p_i can be defined as converging polynomial approximations of h. □

Lemma 2.6 allows one to express polynomial boundedness and complete boundedness in terms of solutions of Lyapunov equations. According to Lemma 2.6 with $k = m = 1$, the number $\gamma_{pb}(A)$ is the maximum in the quadratic program

$$\gamma_{pb}(A) = \sup\{|w|^2 : \ \mathbf{L}_A(vv' - ww') \geq 0, \ |v|^2 \leq 1\}. \tag{11}$$

Using Lemma 2.6 with arbitrary k, m shows that $\gamma_{cb}(A)$ is the maximum in the standard relaxation of (11) obtained by replacing the rank one matrices vv' and ww' with arbitrary positive semidefinite matrices V, W:

$$\gamma_{cb}(A) = \sup\{\mathrm{tr}[W] : \ V \geq 0, \ W \geq 0, \ \mathbf{L}_A(V - W) \geq 0, \ \mathrm{tr}[V] \leq 1\}. \tag{12}$$

2.2.4. Similarity to a Contraction and Lyapunov Equations. The following statement relates positive semidefiniteness of solutions of Lyapunov equations to similarity to a contraction.

Lemma 2.7. *Let A be a stable n-by-n matrix, $r > 0$. A matrix S such that*

$$\|SAS^{-1}\| \leq 1, \ \|S^{-1}\| \leq 1, \ \|S\| < r$$

does not exist if and only if there exist matrices $V, W \in S_n$ such that

$$V \geq 0, \ W \geq 0, \ \mathrm{tr}[V] = 1, \ \mathrm{tr}[W] \geq r^2, \ \ \mathbf{L}_A(V - W) \geq 0. \tag{13}$$

Proof. In terms of $R = S'S$ the conditions imposed on S have the form

$$R - A'RA \geq 0, \ R - I \geq 0, \ (r^2 - \epsilon)I - R \geq 0, \tag{14}$$

where $\epsilon > 0$ is sufficiently small. If V, W satisfy (13) and $Q = \mathbf{L}_A(V - W) \geq 0$ then

$$\begin{aligned} 0 &\leq \mathrm{tr}[Q(R - A'RA) + V((r^2 - \epsilon)I - R) + W(R - I)] \\ &= \mathrm{tr}[(Q - AQA' - V + W)R + r^2 V - \epsilon V - W] \\ &\leq 0 + r^2 - \epsilon - r^2 = -\epsilon, \end{aligned}$$

which proves that (13) and (14) cannot hold simultaneously. On the other hand, if conditions (14) cannot be satisfied then a separation argument similar to the one used in the proof of Lemma 2.5 proves existence of V, W satisfying conditions (13). □

Lemma 2.7 together with Lemma 2.6 prove that $\gamma_{cb}(A) = \gamma_{sc}(A)$. Perhaps more importantly, it is now shown that $\gamma_{sc}(A)$ equals maximum in the standard relaxation of the non-convex quadratic program (11) in which $\gamma_{pb}(A)$ is the maximum. Therefore, the question of whether polynomial boundedness implies similarity to a contraction can be viewed as a question about boundedness of the standard relaxation gap in a specific non-convex quadratic program.

Due to the recent result by G. Pisier [9] the relaxation gap in (11) is not bounded. Earlier, J. Bourgain has shown that the relaxation gap grows not faster than $\log(n)$, according to

$$\gamma_{sc}(A) \leq \mathrm{const} \cdot \gamma_{pb}(A)^4 \log(n+1).$$

2.3. Quadratic Optimization on a Hypercube

Recall that the problem of quadratic optimization on a hypercube is that of maximizing $J(x) = x'Qx$ over $x \in \{-1, 1\}^n$, where $Q \in S_n$ is a given symmetric matrix. When the diagonal entries Q_{ii} of Q are non-negative, an equivalent form is

$$J_*(Q) = \max\{J(x) = x'Qx : \ x \in [-1, 1]^n\} =?$$

The corresponding standard relaxed semidefinite program is

$$\hat{J}_*(Q) = \max\{\hat{J}(X) = \mathrm{tr}[QX] : \ X = X' \geq 0, \ X_{kk} \leq 1\} =?$$

It is possible to formulate a number of hypotheses about the gap between $\hat{J}_*$ and J_*.

2.3.1. Uniform Bounds of Duality Gap. Which conditions should be imposed on Q to guarantee that

$$J_*(Q) \geq c\hat{J}_*(Q) \tag{15}$$

where c is a constant not depending on Q (in particular, on its size)?

For an arbitrary symmetric matrix $Q \in S_n$ (possibly such that some $Q_{ii} < 0$), conjecture (15) turns out to be false, though it can be shown that, as a function of n, c would grow not faster than $\log(n)$. On the other hand, it was shown in

[8] that(15) holds for any positive semidefinite $Q \geq 0$ with $c = 2/\pi$. For the case when Q is diagonally dominant, i.e. when

$$Q_{ii} \geq \sum_{j \neq i} (|Q_{ij}| + |Q_{ji}|) \quad \forall\ i,$$

it was shown in [2] that (15) holds with a better constant $c \approx 0.88$. The so-called Grothendieck inequality (see [5]) states that (15) is true with $c \approx 0.56$ whenever Q has the block form

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix},$$

where Q_{11} and Q_{22} are *square* zero matrices. The answer to the question of whether there exists $c > 0$ such that (15) is true for all matrices Q with $Q_{ii} = 0$ is unknown to the author.

Inequality (15) will also be discussed later.

2.3.2. Vector Sums with Bounded Coefficients. Let $\{u_k\}_{k=0}^{\infty}$ be a sequence of elements of a Hilbert space such that

$$J(x) = \left\| \sum u_k x_k \right\| \leq \gamma < \infty \text{ whenever } \sup_k |x_k| \leq 1. \tag{16}$$

It is well known and easy to show that (16) implies

$$\sum \|u_k\|^2 \leq \gamma,$$

which corresponds to the trivial case $c = 0$ of inequality (15).

Indeed, if $x_k = 0$ for $k > n$ then

$$J(x) = \bar{x}_n' Q^n \bar{x}_n$$

where $\bar{x}_n$ is the vector with components $x_0, \dots, x_n$, and Q^n is the Gramm matrix of $u_0, \dots, u_n$. Since $Q^n_{kk} \geq 0$, the maximum of $\bar{x}_n' Q^n \bar{x}_n$ subject to $|x_k| \leq 1$ is achieved at the extremal points $x_k \in \{-1, 1\}$. Therefore

$$\max\{\bar{x}_n' Q^n \bar{x}_n :\ x_k \in [-1,1]\} = \mathrm{tr}[Q^n] + \max\{\bar{x}_n' Q^n \bar{x}_n :\ x_k \in \{-1,1\}\} \geq \mathrm{tr}[Q^n].$$

Similarly, the result of [8] can be interpreted in the following way: (16) implies existence of a sequence $\{d_k\}$ such that

$$\sum d_k \leq \frac{\pi}{2} \gamma, \tag{17}$$

and

$$\| \sum u_k x_k \|^2 \leq \sum d_k |x_k|^2 \tag{18}$$

for *any* x_k (not only for those bounded by 1). Proving (15) for matrices Q with all $Q_{ii} = 0$ would be equivalent to showing that (17) can be replaced by

$$\sum d_k + \frac{1-c}{c} \sum \|u_k\|^2 \leq \frac{\gamma}{c}.$$

2.3.3. Gramm Matrices of Bounded Functions. A dual interpretation of questions associated with duality gap estimation for quadratic optimization on a hypercube is given in terms of Gramm matrices. Let $V = \{v_k\}_{k=1}^{\infty}$ be a sequence of (real) scalar functions $v_k = v_k(t) \in L^2(0,1)$ which is uniformly bounded by 1, i.e.

$$|v_k(t)| \leq 1 \quad \forall \quad k, t.$$

Which matrices W belong to the set $\mathbf{W} = \{W\}$ of Gramm matrices of such a sequence? Obviously W must be a Gramm matrix of some set of vectors of length not exceeding 1. However, this is not enough. The result of [8] is equivalent to saying that for any Gramm matrix R with $R_{ii} \leq 1$ there exists $W \in \mathbf{W}$ such that

$$R \leq \frac{\pi}{2} W.$$

Similarly, the Grothendieck inequality implies that if any such R is decomposed as

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

where R_{11} and R_{22} are square matrices then cR_{12} is equal to the corresponding block element W_{12} of some $W \in \mathbf{W}$, where $c \approx 0.56$. Similarly, showing that for any Gramm matrix R with $R_{ii} \leq 1$ there exists $W \in \mathbf{W}$ such that

$$cR + (1-c)I = W$$

would be equivalent to proving (15) for all Q with $Q_{ii} = 0$.

3. Relaxations in Dynamical System Analysis

Many questions in analysis and design of dynamical systems come in the form of non-convex quadratic programs. As a consequence, techniques of system theory are frequently based on relaxation and duality gap estimation.

3.1. Stability of Systems with Structured Uncertainty

Analysis of linear time invariant (LTI) systems with structured uncertainty is a prominent example of the use of quadratic programming and its relaxations in system analysis.

3.1.1. Models of Stable LTI Systems and their Interconnections. A stable LTI system S with n inputs and n outputs can be represented mathematically by a function H in the Hardy space $H_n^{\infty}(\mathbf{T})$ of n-by-n matrix-valued analytical functions (H would be called *transfer matrix* of the system). For a pair S_1, S_2 of such stable LTI systems with transfer matrices H_1, H_2, it is said that the *feedback interconnection* S_{12} of S_1 and S_2 is stable if there exists

$$H_{12} = (I - H_1 H_2)^{-1} \in H_n^{\infty}(\mathbf{T}),$$

in which case H_{12} can be regarded as the transfer matrix of a system S_{12} defined by the interconnection of S_1 and S_2.

3.1.2. Uncertain Models and Robust Stability. Feedback interconnections can be used to represent systems with models which are not known precisely, or *uncertain systems*. A rich class of such systems can be modeled as feedback interconnections of a stable LTI system with a known transfer matrix $H \in H_n^\infty(\mathbf{T})$ and a stable LTI system with an unknown transfer matrix Δ, assumed to belong to the set $\bar{\Delta}$ of all *diagonal* matrix functions in $H_n^\infty(\mathbf{T})$, with H^∞ norm not exceeding 1.

Such uncertain model $U = U(H)$ (completely defined by $H \in H_n^\infty(\mathbf{T})$) is called *robustly stable* if there exists a constant c such that the inverse $(I - H\Delta)^{-1}$ exists, belongs to the class $H_n^\infty(\mathbf{T})$, and satisfies the inequality

$$\|(I - H\Delta)^{-1}\|_\infty \leq c$$

for all $\Delta \in \bar{\Delta}$.

3.1.3. Criteria of Robust Stability. It is easy to show that uncertain model $U(H)$ with $H \in H_n^\infty(\mathbf{T})$ is robustly stable if and only if there exists a constant c such that

$$\|(I - MZ)^{-1}\| \leq c \quad \forall\ Z \in \bar{Z} \tag{19}$$

for any essential value M of H where $\bar{Z}$ is the set of all diagonal complex n-by-n matrices with norm not exceeding 1. In most applications, H is continuous and a bound of its derivative is available. A common practice is to sample the values $M = M_k$ of H on a "dense" (in a practical, not mathematical, sense) grid in T and check that each of the samples satisfies (19).

The main difficulty, however, lies in the fact that checking (19) for a given n-by-n matrix M is computationally challenging. In particular, the problem is known to be NP-hard [10]. As a result, one is forced to look for an easier-to-verify condition on M which would imply (19), without being equivalent to it. The frequently used sufficient condition for (19) is existence of a diagonal matrix $D > 0$, $D \leq I$ and $\epsilon > 0$ such that

$$D - M'DM \geq \epsilon I, \tag{20}$$

which implies (19) with $c = c(\epsilon)$.

Indeed, let $R = D^{1/2}$, $T = RMR^{-1}$. Then $D \leq I$ implies $\|R\| \leq 1$ while (20) implies $D \geq \epsilon I$, which means $\|R^{-1}\| \leq \epsilon^{-1/2}$. In addition, multiplying (20) by R^{-1} on both sides yields $I - T'T \geq \epsilon D^{-1} \geq \epsilon I$, hence $\|T\| \leq (1-\epsilon)^{1/2} \leq 1 - \epsilon/2$. Therefore

$$\|(I - MZ)^{-1}\| = \|(R^{-1}(I - TZ)R)^{-1}\| \leq 2\epsilon^{-3/2}.$$

Since (20) can be considered as a semidefinite program with respect to D and ϵ (with an objective to maximize ϵ to check whether it can be made larger than zero), (20) is relatively inexpensive to verify.

In the field of robustness analysis, the notion of *structured singular value* $\mu(M)$ of M is used to represent condition (19): the non-negative number $\mu = \mu(M)$ is defined in such a way that

(a) (19) holds for some c if and only if $\mu(M) < 1$;
(b) $\mu(rM) = |r|\mu(M)$ for any $r \in \mathbf{R}$.

The standard *upper bound* $\hat{\mu}(M)$ *of* $\mu(M)$ is defined in such way that

(a) (20) holds for some $\epsilon > 0$ if and only if $\hat{\mu}(M) < 1$;
(b) $\hat{\mu}(rM) = |r|\hat{\mu}(M)$ for any $r \in \mathbf{R}$.

Thus, robust stability of $U(H)$ is equivalent to the inequality $\operatorname{ess\,sup} \mu(H) < 1$ and is implied by the easier-to-verify inequality $\operatorname{ess\,sup} \hat{\mu}(H) < 1$.

3.1.4. Quality of the Upper Bound of μ. The question about quality of the upper bound $\hat{\mu}(M)$ of $\mu(M)$ was posed about 20 years ago. It is known for a long time that $\hat{\mu}(M) = \mu(M)$ for any matrix M of size 3-by-3, and less. In addition, the equality $\hat{\mu}(M) = \mu(M)$ holds for any matrix M of rank 1. It is very difficult to find matrices M for which the ratio $\mu(M)/\hat{\mu}(M)$ is small. It was conjectured that

$$\mu(M) \geq c\hat{\mu}(M) \tag{21}$$

where the constant c does not depend on M (in particular, on the size of M).

It turns out that (21) is equivalent to existence of a constant duality gap bound in an appropriately constructed non-convex quadratic program. Indeed, for $x \in \mathbf{R}^n$ let $\alpha(x), \beta(x)$ be the diagonal $(n+1)$-by-$(n+1)$ matrices

$$\alpha = \begin{bmatrix} |e_1'x|^2 & & & \\ & \ddots & & \\ & & |e_n'x|^2 & \\ & & & 1 \end{bmatrix}, \ \beta = \begin{bmatrix} |e_1'Mx|^2 & & & \\ & \ddots & & \\ & & |e_n'Mx|^2 & \\ & & & |x|^2 \end{bmatrix}, \tag{22}$$

where $\{e_i\}$ is the standard basis in $\mathbf{C}^n$. By construction, $J_* = \sup J(x) = \mu(M)^2$ and $\hat{J}_* = \sup \hat{J}(w) = \hat{\mu}(M)^2$ for the quadratic program (1) and its relaxation (4) with α, β defined by (22).

3.1.5. An Analog in Operator Theory. The "μ-gap" question of whether the ratio $\hat{\mu}(M)/\mu(M)$ has a finite upper bound remained unanswered until recently Treil [11] used an operator-theoretic analog of the original finite dimensional problem to show that the answer is negative. His construction is as follows.

Let A be a bounded operator on $L^2(\nu)$. For any $\phi \in L^\infty(\nu)$ let M_ϕ be the operator of multiplication by ϕ, acting on $L^2(\nu)$. Does there exist a $\rho > 0$ such that, for any bounded operator A on $L^2(\nu)$, we have

$$\|M_\psi A M_{1/\psi}\| \leq \rho \text{ for some } \psi, 1/\psi \in L^\infty(\nu)$$

whenever

$$I - M_\phi A \text{ is invertible for all } \|\phi\|_\infty < 1?$$

When A is defined by the singular integral

$$(Af)(t) = \frac{1}{\pi} \int_\Gamma \frac{f(s)}{s-t} ds$$

where $d\nu = ds$ is the arclength on a smooth curve Γ with a large Ahlfors constant, ρ is unbounded. Therefore, the infinite dimensional analog of the μ-gap problem has a negative answer. A convergence argument is then used in [11] to show that the original finite dimensional version must have a negative answer as well.

3.2. Analysis of Nonlinear Systems

The dominant idea of rigorous analysis of nonlinear systems is the use of the so-called system invariants (for example, Lyapunov functions). However, finding invariants of generic nonlinear system models is a difficult task. The use of relaxations in non-convex quadratic programming allows one to make substantial progress in this direction.

3.2.1. System Analysis and System Invariants. Consider a nonlinear dynamical system model of the form

$$x_{t+1} = f(x_t, v_t),\ z_t = g(x_t, v_t),\ t = 0, 1, 2, \dots, \tag{23}$$

where x_t, v_t, z_t are "hybrid" signal vectors, which means that the components of x, v, z can be logical (i.e. ranging over the set $\{-1, 1\}$) and analog (i.e. ranging over $[-1, 1]$ or $\mathbf{R}$) elements. Here x_t is the system state, v_t is the disturbance input, modeling the environment, and z_t is the performance output. System (23) is said to meet the performance expectations if for any admissible initial condition x_0 there exists a constant $\gamma = \gamma(x_0)$ such that

$$\sum_{t=0}^{\infty} g(x_t, v_t) \geq -\gamma(x_0), \tag{24}$$

for all possible inputs $v = v_t$.

It can be shown that many important system specifications, including those representing stability, safety, and efficiency, can be expressed in terms of performance expectations with an appropriate selection of function g in (23).

Proving rigorously that a given complex system meets performance expectations is usually very difficult. Chances for successful analysis are greatly improved when a *system invariant*, or *quasi-Lyapunov function* $V : \{x\} \to \mathbf{R}$ is available, defined by the condition

$$V(x_{t+1}) - V(x_t) \leq g(x_t, v_t) \tag{25}$$

for all possible trajectories of (23). Designing efficient algorithms for finding such functions V is a major problem in system analysis.

3.2.2. Analysis Using Partial Invariants. An approach to *automatic* search for system invariants can be based on the notion of *partial Lyapunov functions*. Assume that an auxiliary variable w_t is defined by $w_t = \phi(x_t, v_t)$ in such a way that the total range of possible values of w_t is a subset of

$$\Omega = \{(v, u) \in \mathbf{R}^p \times \{-1, 1\}^q :\ |h_i v| \leq 1\}, \tag{26}$$

where $\{h_i\}$ is a finite set of linear functionals on $\mathbf{R}^p$. In other words, Ω is a direct product of a polytope in $\mathbf{R}^p$ (symmetric with respect to the origin) and a complete

set of vertices of a cube. A function $V_k : \{x\} \to \mathbf{R}$ is called a partial Lyapunov function (for system (23), with respect to the auxiliary variable w) if there exists a quadratic functional

$$\sigma_k(w) = w'Q_k w + 2L_k w + C_k$$

such that

$$V_k(x_{t+1}) - V_k(x_t) \le \sigma_k(w_t) \tag{27}$$

for any possible trajectory of (23). The inequality (27) in that case is called an *Integral Quadratic Constraint* (IQC).

Typically, partial Lyapunov functions can be derived for the simple building blocks (elementary logic, delay elements, integrators, uncertainty bounds etc.) in terms of which the whole system (23) can be described. Usually, more useful partial Lyapunov functions can be obtained after the state of the system is re-defined to include some past history (for example, by concatenating the original state x_t with a stable LTI transformation of the past states $x_{t-\tau}$).

Once available, partial Lyapunov functions can be used in an automatic search for a quasi-Lyapunov function in the form

$$V(x_t) = V_0(x_t) + \sum_{k=1}^{N} \tau_k V_k(x_t), \quad \tau_k \ge 0, \tag{28}$$

provided that V_0 satisfies the inequality

$$V_0(x_{t+1}) - V_0(x_t) \le \sigma_0(w_t) + g(x_t, v_t), \tag{29}$$

where σ_0 is a quadratic functional, for any possible trajectory of (23).

The set of the coefficients $\tau = \tau_k$ for which (28) yields a function V satisfying (25) is convex. Therefore, an efficient algorithm for finding τ will be available, as soon as one can find an acceptable "feasibility oracle", i.e. an algorithm which quickly checks the inequality in (25) for a particular V from (28) (and, in particular, presents t, x_t, v_t such that (25) is not valid, if this is the case). This is where partial Lyapunov functions can help, because, due to (27), the inequality in (28) will be implied whenever

$$\sigma(w) \le 0 \quad \forall\ w \in \Omega, \tag{30}$$

for

$$\sigma(w) = \sigma_0(w) + \sum \tau_k \sigma_k(w).$$

In contrast with (25), the inequality in (30) is relatively easy to verify. In the standard theory of IQC analysis [6] this is done by simply checking that $\sigma(w) \ge 0$ for *all* w (i.e. not only those from Ω). This is good enough when absolute stability and performance of nonlinear and time-varying perturbations of linear systems are analyzed, as long as the general behavior remains similar to the linear one. However, analysis of essentially nonlinear behavior of hybrid systems requires a more accurate treatment of (30).

3.2.3. RELAXATION. The task of checking (30) is essentially equivalent to verifying that maximum of the quadratic functional

$$\tilde{\sigma}(w) = w'Qw + 2L'w,$$

on Ω does not exceed $-C$ where

$$Q = Q_0 + \sum_{k=1}^{N} \tau_k Q_k,\ L = L_0 + \sum_{k=1}^{N} \tau_k L_k,\ C = C_0 + \sum_{k=1}^{N} \tau_k C_k.$$

As usually, one can assume without loss of generality that the quadratic form $w'Qw$ does not contain any terms $q_{kk}u_k^2$, where u_k are those components of $w = (v, u)$ that range over the binary set $u_k \in \{-1, 1\}$. Indeed, any such term can be replaced by the constant q_{kk}. When the coefficients at u_k^2 are zero, allowing u_k to range over $[-1, 1]$, rather than $\{-1, 1\}$, will not change the maximum of $\sigma(w)$. Therefore, optimization of $\tilde{\sigma}(w)$ over Ω is equivalent to optimization of $\bar{\sigma}(w, \theta)$ over

$$\bar{\Omega} = \{(v, u, \theta) \in \mathbf{R}^p \times [-1, 1]^q \times [-1, 1] :\ |h_k v| \le 1\},$$

where

$$\bar{\sigma}(w, \theta) = w'Qw + 2\theta L'w.$$

Finally, we can conclude that verifying (30) reduces to finding the maximum in the MAX-CUT quadratic program (generally non-convex):

$$J(x) = x'Px \to \max \text{ subject to } |a_k'x| \le 1\ (k = 1, \dots, n). \tag{31}$$

The optimization problem in (31) can be "relaxed" to a semidefinite program in several ways, the simplest (and presumably the roughest) of which is

$$\hat{J}(X) = \mathrm{tr}(PX) \to \max \text{ subject to } a_k'Xa_k \le 1,\ X = X' \ge 0. \tag{32}$$

In order to construct a quasi-lyapunov function V in (25), it is be sufficient to find $\tau_k \ge 0$ such that

$$\mathrm{tr}\left(X\left(\begin{bmatrix} Q_0 & L_0 \\ L_0' & 0 \end{bmatrix} + \sum_{k=1}^{N} \tau_k \begin{bmatrix} Q_k & L_k \\ L_k' & 0 \end{bmatrix}\right)\right) \le C_0 - \sum_{k=1}^{N} \tau_k C_k\ \forall\ X \in \Omega_r, \tag{33}$$

where

$$W_r = \left\{X = X' \ge 0 :\ \begin{bmatrix} h_k \\ 0 \end{bmatrix}' X \begin{bmatrix} h_k \\ 0 \end{bmatrix} \le 1, \begin{bmatrix} 0 \\ e_k \end{bmatrix}' X \begin{bmatrix} 0 \\ e_k \end{bmatrix} \le 1\right\}. \tag{34}$$

(Here e_k, $k = 1, \dots, q$ is the coordinate basis in $\mathbf{R}^q$.)

3.2.4. Quasi-Linear Models. A major obstacle in using partial invariants for system analysis is derivation of a sufficiently large and representative set of partial invariants. While partial invariants can be derived for simple subsystems, and are "inherited" when such subsystem is included in a complex interconnection, the success of the analysis may well depend on the partial invariants describing interaction between subsystem. At this point, the number of partial invariants to consider begins growing exponentially, which makes the whole analysis process inefficient.

A different approach to automatic search for system invariants can be proposed, based on the notion of a *quasi-linear model*. For simplicity, consider system (23) where

$$(x_t, f_t) \in \Omega_0 = [-1,1]^p \times \{-1,1\}^q$$

is a quasi-cube. System (23) is said to be *quasi-linear* if there exist matrices D, F, L of appropriate size, a quadratic functional

$$\sigma_0(w) = \begin{bmatrix} w \\ 1 \end{bmatrix}' \Sigma_0 \begin{bmatrix} w \\ 1 \end{bmatrix},$$

and a quasi-cube

$$\Omega = [-1,1]^a \times \{-1,1\}^b,$$

such that

$$\{(w_0, f(w_0), g(w_0)) : \ w_0 \in \Omega_0\} = \{(Dw, Fw, \sigma_0(w)) : \ Lw = 0, \ w \in \Omega\}. \quad (35)$$

In other words, a quasi-linear model is defined by a set of *linear* equations and a quadratic "cost" on a quasi-cube.

When $b = 0$ in the definition of Ω (i.e. when Ω is a hypercube in the usual sense), the only *functions* f, g that can be defined by (35) are respectively linear and quadratic (it is also possible to produce some *multi-valued* mappings f, g). However, it is well known that, when some components of $w \in \Omega$ are allowed to be "bits" (modeled as elements of $\{-1,1\}$ in this case), the set of functions (f, g) which one can define by (35) is dense in the uniform metric. In other words, an arbitrary system model (23) defined by *continuous* functions f, g on a quasi-cube can be approximated arbitrarily well by quasi-linear models.

For example, the operation of *quantization* of a scalar $x_0 \in [-1,1]$ can be defined by the linear equations

$$x_{k-1} = 0.5(x_k + w_k), \quad k = 1, 2, \ldots,$$

where the variables x_k range over the interval $[-1,1]$, while the variables w_k (which represent a binary expansion of x_0) range over $\{-1,1\}$. Similarly, the logical operation "AND" (i.e. $w_3 = \phi(w_1, w_2) = 1$ if and only if $w_1 = w_2 = 1$) can be defined by the single equation

$$w_1 + w_2 - 2w_3 + w_4 = 1,$$

where $w_i \in \{-1, 1\}$ for $i = 1, \dots, 4$. Using the operations of quantization, logical operations, and digital-to-analog conversion (which is simply a weighted summation of the corresponding bits), one can define a rich set of transformations.

3.2.5. Quadratic Stability of Quasi-Linear Models. Let (23) be a quasi-linear system defined by (35). A quadratic functional $\sigma : \Omega \to \mathbf{R}$ is a system invariant if there exists $r > 0$ such that

$$\rho(\sigma, r) = \max_{w \in \Omega} \{\sigma(Fw) - \sigma(Dw) - \sigma_0(w) - r|Lw|^2\} \le 0. \tag{36}$$

To find σ, r such that (36) holds, one can try to minimize ρ as a function of σ, r. Note that $\rho(\cdot)$ is a maximum of a family of linear functionals, and hence is a convex function of its arguments. Therefore, minimization of ρ can be performed efficiently, assuming that $\rho(\sigma, r)$ can be evaluated for any fixed σ, r. The task of calculating the maximum in (36) is equivalent to the MAX-CUT optimization problem.

The procedure can be used to search automatically for quadratic invariants of systems with quasi-linear models. Moreover, while the invariants searched for are limited to quadratic functions of the "extended" system state $w \in \Omega$, the resulting system invariant will in general be non-quadratic as a function of w_0. Still, it is not clear that a system invariant in this specific form can be found whenever a system invariant exists.

To increase flexibility of the approach, the search for a quadratic invariant can be applied to a "lifted" model of system (35). Indeed, if (23) can be represented in the form of (35) then the "lifted" system with

$$\bar{x}_k = \begin{bmatrix} x_{mk+1} \\ x_{mk+2} \\ \vdots \\ x_{m(k+1)} \end{bmatrix}, \quad \bar{v}_k = \begin{bmatrix} v_{mk+1} \\ v_{mk+2} \\ \vdots \\ v_{m(k+1)} \end{bmatrix}$$

can be represented in the "lifted" form as well, according to

$$\bar{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} \in \Omega^m, \quad \bar{D}\bar{w} = Dw_1, \quad \bar{F}\bar{w} = Fw_m, \quad \bar{L}\bar{w} = \begin{bmatrix} Lw_1 \\ \vdots \\ Lw_m \\ Fw_1 - Dw_2 \\ \vdots \\ Fw_{m-1} - Dw_m \end{bmatrix}.$$

In general, one can expect that a combination of quasi-linear modeling with the use of partial invariants on a lifted system model will be most convenient for automatic generation of system invariants.

4. Gaussian Proofs of Duality Gap Bounds

In this section, the technique of Gaussian randomization is used to prove some old and new duality gap bounds for the problem of quadratic optimization over the hypercube.

4.1. Using Signs of Gaussian Vectors

Several duality gap bounds in the MAX-CUT problem can be proven by working with the signs of Gaussian vectors.

4.1.1. The Idea of Goemans and Williamson. The idea of working with the signs $v_i = \mathrm{sgn}(w_i)$ of a set $\{w_i\}$ of jointly Gaussian random variables was originally used by Goemans and Williamson in [2] to prove the following uniform bound of the duality gap in the MAX-CUT problem.

Theorem 4.1. *(Goemans, Williamson [2]) Let $Q \in S_n$ be a matrix such that*

$$Q_{ii} \geq \sum_{j \neq i} (|Q_{ij}| + |Q_{ji}|).$$

Then $J_* \geq c\hat{J}_*$ where

$$J_* = \max\{J(x) = x'Qx : \quad x \in \{-1,1\}^n\},$$

$$\hat{J}_* = \max\{\hat{J}(W) = \mathrm{tr}[QW] : \quad W \in S_n,\ W \geq 0,\ W_{ii} = 1\},$$

$$c = \min\left\{\frac{2}{\pi}\frac{t}{1-\cos(t)} : \quad t \in (0,\pi)\right\} \approx 0.8786.$$

Moreover, if w is a Gaussian random vector such that $\mathbf{E}w = 0$ and $\mathbf{E}ww' = \bar{W}$ is the argument of the maximum of $\hat{J}(W)$ (i.e. $\hat{J}(\bar{W}) = \hat{J}_*$, $\bar{W} = \bar{W}' \geq 0$, $\bar{W}_{ii} = 1$) then

$$\mathbf{E}\mathrm{sgn}(w)'Q\mathrm{sgn}(w) \geq c\hat{J}_*,$$

where in $\mathrm{sgn}(w)$ the sign function is applied to w component-wise.

Proof. Following [2] let us prove first that if f and g are two jointly Gaussian random variables such that

$$\mathbf{E}f = \mathbf{E}g = 0, \quad \mathbf{E}|f|^2 = \mathbf{E}|g|^2 = 1, \quad \mathbf{E}fg = \cos(t)$$

then

$$\mathbf{E}\mathrm{sgn}(f)\mathrm{sgn}(g) = \phi(t) = 1 - \frac{2t}{\pi}. \tag{37}$$

Indeed, let $f - g\cos(t) = h\sin(t)$ where $\mathbf{E}h = 0$ and $\mathbf{E}|h|^2 = 1$. Then $\mathbf{E}fh = 0$ and hence the random variables f and h are independent. Therefore the vector random variable

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \frac{1}{|f|^2 + |h|^2}\begin{bmatrix} f \\ h \end{bmatrix}$$

is uniformly distributed on the unit circle in $\mathbf{R}^2$, which implies that

$$\mathbf{E}\mathrm{sgn}(f)\mathrm{sgn}(g) = \mathbf{E}\mathrm{sgn}(z_1)\mathrm{sgn}(z_1\cos(t) + z_2\sin(t)) = \phi(t)$$

is an affine function of t. Since $\phi(0) = 1$ and $\phi(\pi) = -1$, the conclusion is that $\phi(t) = 1 - 2t/\pi$.

Now let $v_i = \mathrm{sgn}(w_i)$. Since $\mathbf{E}|v_i|^2 = 1$, (37) shows that

$$\mathbf{E}v_iv_j = 1 - \frac{2}{\pi}\arccos(W_{ij}).$$

Hence, for $t_{ij} = \arccos(W_{ij}) \in [0, \pi]$,

$$1 - \mathbf{E}v_iv_j = \frac{2t_{ij}}{\pi} \geq c(1 - \cos(t_{ij})) = c(1 - \mathbf{E}w_iw_j).$$

Since

$$\min_{t\in(0,\pi)} \frac{2}{\pi}\frac{t}{1-\cos(t)} = \min_{t\in(0,\pi)} \frac{2}{\pi}\frac{\pi - t}{1+\cos(t)},$$

it is also true that

$$1 + \mathbf{E}v_iv_j = \frac{2}{\pi}(\pi - t_{ij}) \geq c(1 + \cos(t_{ij})) = c(1 + \mathbf{E}w_iw_j).$$

Therefore, for

$$r = \mathrm{tr}[Q] - \sum_{j\neq i}(|Q_{ij}| + |Q_{ji}|),$$

it follows that

$$\begin{aligned}
\mathbf{E}v'Qv &= r + \sum_{i\neq j}|Q_{ij}|(1 + \mathrm{sgn}(Q_{ij})\mathbf{E}v_iv_j) \\
&\geq cr + c\sum_{i\neq j}|Q_{ij}|(1 + \mathrm{sgn}(Q_{ij})\mathbf{E}w_iw_j) \\
&= c\,\mathrm{tr}[QW].
\end{aligned}$$

Since $J_* \geq v'Qv$, this proves the theorem. □

4.1.2. A Remark on Orthogonality. Other duality gap bounds can be proven using an orthogonality feature of the sign of a Gaussian variable. The constant

$$\nu = \left(\frac{2}{\pi}\right)^{1/2},$$

which equals $\mathbf{E}|w|$ for a Gaussian random variable w with zero mean and unit variance, will play an important role.

Lemma 4.2. *Let f, g be two jointly Gaussian random variables with zero mean. Let $v = (\mathbf{E}g^2)^{1/2}\mathrm{sgn}(g)$. Then $\mathbf{E}f(\nu g - v) = 0$.*

Proof. A direct calculation shows that $\mathbf{E}g(\nu g - v) = 0$. On the other hand, f can be written in the form $f = f_0 + ag$ where $a \in \mathbf{R}$ and f_0 is independent of g. Hence f_0 and $\nu g - v$ are independent as well, and therefore $\mathbf{E}f_0(\nu g - v) = 0$. □

4.1.3. The Nesterov-Ye Theorem. An immediate consequence of Lemma 4.2 is the following theorem proven by Nesterov [8] and Ye [12].

Theorem 4.3. *For $Q \in S_n$ let*

$$J_+ = \max\{J(x) = x'Qx : \quad x \in [-1,1]^n\},$$

$$\hat{J}_+ = \max\{\hat{J}(W) = \mathrm{tr}[QW] : \quad W \in S_n,\ W \geq 0,\ W_{ii} \leq 1\},$$

$$\hat{J}_- = \min\{\hat{J}(W) = \mathrm{tr}[QW] : \quad W \in S_n,\ W \geq 0,\ W_{ii} \leq 1\}.$$

Then

$$J_+ \geq \frac{2}{\pi}\hat{J}_+ + (1 - \frac{2}{\pi})\hat{J}_-.$$

Moreover, if $\bar{W}$ is the argument of maximum of $\hat{J}(W)$ subject to $W = W' \geq 0$ and $W_{ii} \leq 1$ then

$$\mathbf{E}v'Qv \geq \nu^2\hat{J}_+ + (1-\nu^2)\hat{J}_-, \tag{38}$$

where

$$v_i = (\mathbf{E}w_i^2)^{1/2}\mathrm{sgn}(w_i)$$

and w is a Gaussian random vector with $\mathbf{E}w = 0$, $\mathbf{E}ww' = \bar{W}$.

Proof. By Lemma 4.2

$$\mathbf{E}w'F(\nu w - v) = 0$$

for *any* constant matrix F. Note also that since

$$\mathbf{E}(\nu w_i - v_i)^2 = (1-\nu^2)\mathbf{E}w_i^2 \leq 1 - \nu^2,$$

for all i, the inequality

$$\mathbf{E}(\nu w - v)'Q(\nu w - v) \geq (1-\nu^2)\hat{J}_-$$

takes place. Hence

$$\begin{aligned}\mathbf{E}v'Qv &= \mathbf{E}(\nu w + (v - \nu w))'Q(\nu w + (v - \nu w)) \\ &= \nu^2\mathbf{E}w'Qw + \mathbf{E}(\nu w - v)'Q(\nu w - v) \\ &\geq \nu^2\hat{J}_+ + (1-\nu^2)J_-.\end{aligned}$$

□

Theorem 4.3 was proven in [8] for the case $Q \geq 0$ and in [12] for the general case. In [12], the following implication of (38) is formulated as the main result:

$$\frac{J_+ - \mathbf{E}v'Qv}{J_+ - J_-} \leq \frac{\pi}{2} - 1, \tag{39}$$

where

$$J_- = \min\{J(x) = x'Qx : \quad x \in [-1,1]^n\}.$$

To derive (39) from Theorem 4.3, note first that applying the Theorem with Q replaced by $-Q$ yields

$$J_- \le \nu^2 \hat{J}_- + (1-\nu^2)\hat{J}_+.$$

Hence

$$\begin{aligned} \frac{J_+ - \mathbf{E}v'Qv}{J_+ - J_-} &\le \frac{J_+ - (\nu^2\hat{J}_+ + (1-\nu^2)\hat{J}_-)}{J_+ - (\nu^2\hat{J}_- + (1-\nu^2)\hat{J}_+)} \\ &\le \frac{\hat{J}_+ - (\nu^2\hat{J}_+ + (1-\nu^2)\hat{J}_-)}{\hat{J}_+ - (\nu^2\hat{J}_- + (1-\nu^2)\hat{J}_+)} \\ &= \nu^{-2} - 1. \end{aligned}$$

Note that the second inequality is valid since the function

$$x \mapsto \frac{x - (\nu^2\hat{J}_+ + (1-\nu^2)\hat{J}_-)}{x - (\nu^2\hat{J}_- + (1-\nu^2)\hat{J}_+)}$$

is non-decreasing for $x \ge \nu^2\hat{J}_+ + (1-\nu^2)\hat{J}_-$, and $J_+ \le \hat{J}_+$.

4.1.4. Subtraction of Even-Cyclic Matrices. Inequality (38) implies

$$J_+ \ge (4/\pi - 1)\hat{J}_+$$

whenever it can be proven that $\hat{J}_- \ge -\hat{J}_+$. An interesting case when this can be done easily is associated with matrices with even cycles only.

Definition 4.4. *Let us call a symmetric matrix $C \in S_n$* even-cyclic *if its diagonal entries equal zero and the graph Γ_C formed by connecting those nodes $i, j \in \{1, \dots, n\}$ for which $C_{ij} \ne 0$ contains no cycles of odd length.*

Theorem 4.5. *If $Q - C \ge 0$ for some even-cyclic matrix C then $J_+ \ge (4/\pi - 1)\hat{J}_+$.*

Proof. Without loss of generality assume that the graph Γ_C is connected. For any n-vector x and an even-cyclic matrix C let $z = x_C$ be the vector defined as follows: $z_i = x_i$ whenever node i of Γ_C is connected to node 1 through an even number of edges, and $z_i = -x_i$ otherwise. Since Γ_C is even-cyclic, the definition of $z = x_C$ is correct. By construction, $x'Cx = -x'_C C x_C$ for any x. Hence $\hat{J}_- \ge -\hat{J}_+$ for any $Q \ge C$. □

4.1.5. The Grothendieck Inequality. Theorem 4.5 can be considered as a "generalized" Grothendieck inequality. However, for the original Grothendieck inequality, the resulting constant $4/\pi - 1$ is far from being the best known.

Theorem 4.6. *The inequality*

$$J_* \ge (4/\pi - 1)\hat{J}_*$$

holds for all matrices

$$Q = \begin{bmatrix} 0 & Q_{12} \\ Q'_{12} & 0 \end{bmatrix},$$

where the diagonal blocks are of square form.

Proof. Q is an even-cyclic matrix. □

4.2. Slowly Increasing Duality Gap

Some slowly increasing duality gap bounds can be derived using the fast decay of the Gaussian density away from the origin.

4.2.1. A Logarithmic Upper Bound for Duality. Taking into account difficulty of finding good duality gap bounds, it is surprising to see that the following theorem is true.

Theorem 4.7. *For $n > 60$, $\hat{J}_* \leq 2\log(n) J_*$ for any problem data P, a_k, n in (31),(32).*

Proof. Let $X = X' \geq 0$ be any matrix satisfying the constraints $a_k' X a_k \leq 1$. Let us show that then there exists a vector x satisfying the conditions $|a_k' x| \leq 1$ such that $2\log(n) x' P x \geq \mathrm{tr}[XP]$.

Indeed, let w will be a Gaussian random vector with $\mathbf{E}[w] = 0$ and $\mathbf{E}[ww'] = X$. Let $q = \max_k\{|a_k' w|\}$, $v = w/q$. Then v is a random variable satisfying the constraints $|a_k' v| \leq 1$ with probability 1. Moreover, we have

$$\begin{aligned} \mathrm{tr}[XP] &= \mathbf{E}[w' P w] \\ &= \mathbf{E}[q^2 (v' P v)] \\ &\leq \mathbf{E}[v' P v]\mathbf{E}[q^2]. \end{aligned}$$

Hence, in order to complete the proof, it is sufficient to show that $\mathbf{E}[q^2] \leq 2\log(n)$.

For $r \geq 0$ let $f(r)$ be the probability that $q^2 \geq r$. Since $f(r)$ does not exceed the sum of probabilities $f_k(r)$ that $|a_k' w|^2 \geq r$, and since $a_k' w$ are Gaussian random variables with zero mean and variance not exceeding 1,

$$f(r) \leq n\nu \int_{\sqrt{r}}^{\infty} e^{-t^2/2} dt. \tag{40}$$

Since $f(r) \leq 1$ for all r, for any $\lambda > 0$ we have

$$\int_0^{\infty} f(r) dr \leq \lambda + \int_{\lambda}^{\infty} f(r) dr.$$

Hence

$$\mathbf{E}[q^2] \leq \lambda + n\nu \int_{\lambda}^{\infty} f(r) dr = \lambda + n\nu \int_{\lambda}^{\infty} \int_{\sqrt{r}}^{\infty} e^{-t^2/2} dt dr.$$

Changing the order of integration and integrating over r yields

$$\int_{\lambda}^{\infty} \int_{\sqrt{r}}^{\infty} e^{-t^2/2} dt dr = \int_{\sqrt{\lambda}}^{\infty} (t^2 - \lambda) e^{-t^2/2} dt.$$

Substituting $t = (s + \lambda)^{1/2}$ into the last integral yields

$$\int_{\sqrt{\lambda}}^{\infty} (t^2 - \lambda) e^{-t^2/2} dt = \frac{1}{2} \int_0^{\infty} s e^{-(s+\lambda)/2} (s + \lambda)^{-1/2} ds.$$

Replacing $(s+\lambda)^{-1/2}$ with its upper bound $\lambda^{-1/2}$ yields

$$\int_0^\infty se^{-(s+\lambda)/2}(s+\lambda)^{1/2}ds \le \lambda^{-1/2}e^{-\lambda/2}\int_0^\infty se^{-s/2}ds = 4\lambda^{-1/2}e^{-\lambda/2}.$$

Hence

$$\mathbf{E}[q^2] \le \lambda + 4n\sqrt{\frac{1}{2\pi}}\exp(-\lambda/2)\lambda^{-1/2}.$$

Taking $\lambda = 2\log(n) - 1$ yields $\mathbf{E}[q^2] \le 2\log(n)$ for $n > 60$. □

4.2.2. A Lower Bound for the Duality Gap. The following theorem shows that it is not possible to replace the logarithmic relaxation gap bound of Theorem 4.7 with a constant bound which does not grow with n.

Theorem 4.8. *There exists a sequence of matrices $P = P_n$ of size $n \times n$, $n \to \infty$, such that, for $a_i = e_i$ being the standard basis vectors, $\hat{J}_*(P_n)/J_*(P_n) \to \infty$.*

Proof. For a fixed positive integer m and an arbitrarily small $\epsilon > 0$ let $\mathbf{U} = \{u_k\}_{k=1}^N \subset \mathbf{R}^m$ be an ϵ-dense subset of the unit sphere, i.e. $|u_k| = 1$ for all k and

$$\max_k |u_k'x|^2 \ge (1-\epsilon)|x|^2 \quad \forall\, x \in \mathbf{R}^m. \tag{41}$$

Let U be the m-by-N matrix whose k-th column is u_k. Let $X = (u_i'u_j)_{ij} = U'U$ be the Gramm matrix of $\mathbf{U}$.

First, let us show that X does not belong to the set Λ of all convex combinations of rank one matrices $\theta_i\theta_i'$, where $\|\theta_i\|_\infty < \sqrt{m(1-\epsilon)}$.

Indeed, if

$$X = U'U = \sum c_i\theta_i\theta_i', \quad c_i > 0, \quad \sum c_i \le 1 \tag{42}$$

, then all θ_i satisfy the inequality $c_i^{-1}U'U \ge \theta_i\theta_i'$, and hence belong to the image of U', i.e. $\theta_i = U'v_i$ for some $v_i \in \mathbf{R}^m$. According to (41)

$$|v_i|^2 \le (1-\epsilon)^{-1}\max|u_k'v_i|^2 = (1-\epsilon)^{-1}\|\theta_i\|_\infty^2 < m.$$

On the other hand, since U' is left invertible, identity (42) implies

$$I_m = \sum c_iv_iv_i'.$$

Comparing traces on both sides yields a contradiction

$$m = \sum c_i|v_i|^2 < m.$$

Now let $P = P' \ne 0$ define the linear functional separating X from Λ. By construction, $\hat{J}_*(P) \ge m(1-\epsilon)J_*(P)$. □

4.2.3. Discussion. Existence of a logarithmic gap bound for the relaxation of (31) into (32) appears to be a consequence of convexity of all constraints in (31). In contrast, quadratic programs derived in most of the classical robustness analysis techniques typically have the relaxation gap growth according to a power law. Thus, boundedness of signals subject to linear constraints is expected to be easier to exploit in the analysis than the usual "sector inequalities". Therefore, modeling of nonlinear systems in terms of signals subject to linear equations and amplitude constraints, though more difficult than the canonical approach of the "gain bounds", should be considered as a promising research direction.

Acknowledgement

This work was supported by NSF, AFOSR, DARPA, and the Esther and Harold E. Edgerton Associate Professorship at MIT. The author is grateful to Alexander Borichev for many useful remarks.

References

[1] F. Alizadeh, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM Journal on Optimization, **5** (1995), 13–51.

[2] M.X. Goemans and D.P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM, **42** (1995), 1115–1145.

[3] P. Halmos, *Ten problems in Hilbert space*, Bull. Amer. math. Soc., **76** (1970), 887–933.

[4] J. Hastad, *Some optimal inapproximability results*, Proc. of the 29th ACM Symp. on Theory Comput., 1997.

[5] J.-L. Krivine, *Constantes de Grothendieck et fonctions de type positif sur les spheres*, Adv. in Math., **31** (1979), 16–30.

[6] A. Megretski and A. Rantzer, *System Analysis via Integral Quadratic Constraints*, IEEE Transactions on Automatic Control, volume 47, no. 6, pp. 819–830, June 1997.

[7] Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Methods in Convex Programming*, SIAM, Philadelphia, PA, 1994.

[8] Y. Nesterov, *Quality of semidefinite relaxation for nonconvex quadratic optimization*, Manuscript, 1997.

[9] G. Pisier, *A polynomially bounded operator on Hilbert space which is not similar to a contraction*, Journal Amer. Math. Soc., **10** (1997), no. 2, 351–369.

[10] O. Toker and H. Ozbay, *On the complexity of purely complex μ computation and related problems in multidimensional systems*, IEEE Trans. Aut. Contr., **43** (1998), no. 3, 409–414.

[11] S. Treil, *The gap between complex structured singular value μ and its upper bound is infinite*, To appear in IEEE Trans. Aut. Contr.

[12] Yinyu Ye, *Approximating quadratic programming with bound constraints*, Manuscript, 1997.

Department of Electrical Engineering and Computer Science,
MIT,
35-415 Cambridge 02139, USA
E-mail address: ameg@mit.edu