1. Problem Statement
2. Overview of Results
3. Notation and Definitions
4. Facts
5. Lower Bounds: Rank Relaxation and the S-procedure
6. Stochastic Interpretation of Rank Relaxation
7. Upper Bounds: Intelligent Sampling and Local Search
8. Sensitivities: Use the S-procedure multipliers
9. References

We consider the following minimization which has a quadratic cost and quadratic constraints:

$$\begin{aligned} p = \min_x \, & x^T M_0 x + 2b_0^T x + a_0 \\ \text{subject to: } & x^T M_i x + 2b_i^T x + a_i \leq 0 \quad i = 1, \ldots, N \end{aligned} \tag{1}$$

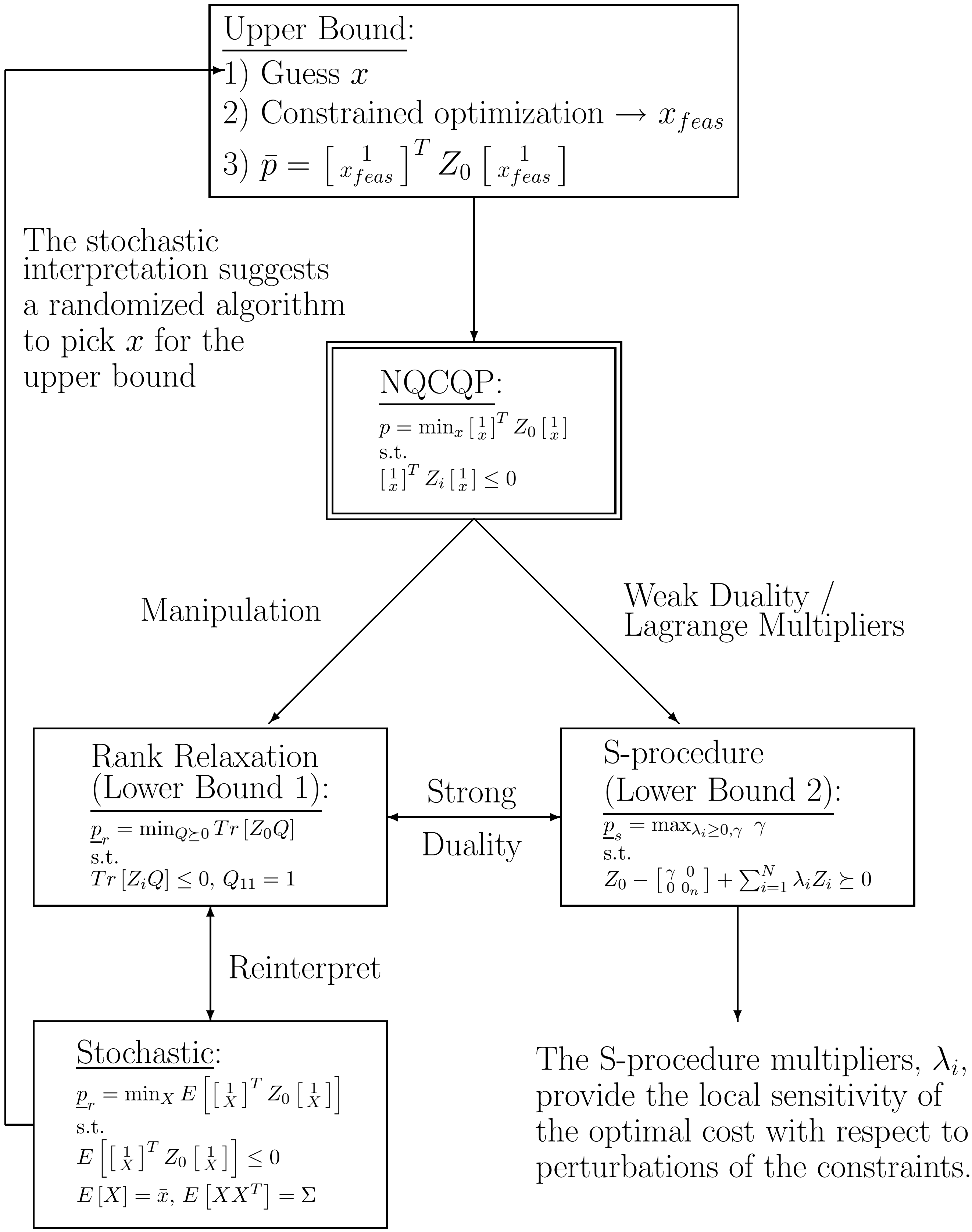
where $a_i \in \mathbb{R}$, $b_i \in \mathbb{R}^n$, and $M_i \in \mathbb{R}^{n \times n}$ $(i = 0, \ldots, N)$ are given data.

To ease the notation, we define $Z_i := \begin{bmatrix} a_i & b_i^T \\ b_i & M_i \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}$ for $i = 0, \ldots, N$. With this notation, the minimization in Equation 1 can be written as:

$$\begin{aligned} p = \min_x \, & \begin{bmatrix} 1 \\ x \end{bmatrix}^T Z_0 \begin{bmatrix} 1 \\ x \end{bmatrix} \\ \text{subject to: } & \begin{bmatrix} 1 \\ x \end{bmatrix}^T Z_i \begin{bmatrix} 1 \\ x \end{bmatrix} \leq 0 \quad i = 1, \ldots, N \end{aligned} \tag{2}$$

Remarks:

- If $\{M_i\}_{i=0}^N$ are all positive semidefinite, this is known as a Quadratically Constrained, Quadratic Program (QCQP). It is a convex optimization that can be easily solved.
- The general case is known as a Nonconvex, Quadratically Constrained, Quadratic Program (NQCQP). It is computationally difficult to solve, so we will instead seek lower and upper bounds on p.


Upper Bound:
1) Guess x
2) Constrained optimization $\longrightarrow x_{feas}$
3) $\bar{p} = \begin{bmatrix} 1 \\ x_{feas} \end{bmatrix}^T Z_0 \begin{bmatrix} 1 \\ x_{feas} \end{bmatrix}$
The stochastic interpretation suggests a randomized algorithm to pick x for the upper bound
NQCQP:
$p = \min_x \left[\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right]$
s.t.
$\left[\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right]^T Z_i \left[\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right] \leq 0$
Manipulation
Weak Duality / Lagrange Multipliers
Rank Relaxation (Lower Bound 1):
$\underline{p}_r = \min_{Q \succeq 0} Tr\,[Z_0 Q]$
s.t.
$Tr\,[Z_i Q] \leq 0$, $Q_{11} = 1$
Strong Duality
S-procedure (Lower Bound 2):
$\underline{p}_s = \max_{\lambda_i \geq 0, \gamma}\ \gamma$
s.t.
$Z_0 - \left[\begin{smallmatrix} \gamma & 0 \\ 0 & 0_n \end{smallmatrix}\right] + \sum_{i=1}^{N} \lambda_i Z_i \succeq 0$
Reinterpret
Stochastic:
$\underline{p}_r = \min_X E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right]$
s.t.
$E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right] \leq 0$
$E\,[X] = \bar{x}$, $E\left[XX^T\right] = \Sigma$
The S-procedure multipliers, λ_i, provide the local sensitivity of the optimal cost with respect to perturbations of the constraints.

1. $\mathbb{R}$ is the set of real numbers.
2. $\mathbb{R}^{n\times m}$ is the set of $n \times m$ dimensional matrices with entries in $\mathbb{R}$. The entry in the i'th row and j'th column of a matrix M is denoted by M_{ij} or $(M)_{ij}$.
3. If $M \in \mathbb{R}^{n\times m}$ then M^T is the transpose of M, i.e. $(M^T)_{ij} = M_{ji}$.
4. The notation $f : X \to Y$ means that X and Y are sets, and f is a function mapping X into Y
5. Set notation:
   - $x \in X$ is read: "x is an element of X"
   - $X \subset Y$ is read: "X is a subset of Y"
   - The expression "$\{\mathcal{A} : \mathcal{B}\}$" is read as: "The set of all $\mathcal{A}$ such that $\mathcal{B}$."
6. Optimizations:
   - If $p = \min_{x\in X} f(x)$, the minimum value, p, may not be achieved by any $x \in X$. Although we will not do so in the following slides, we should write $p = \inf_{x\in X} f(x)$ in this case. A similar statement holds with max and sup.
   - We define the conventions that if X is the empty set then $\min_{x\in X} f(x) = +\infty$ and $\max_{x\in X} f(x) = -\infty$
7. The expectation of a random variable, X, is defined as:

$$E[X] := \int xp(x)dx$$

Definition 1 *For a square matrix $A \in \mathbb{R}^{n \times n}$, the* <u>*trace*</u> *of A is defined as $Tr\,[A] = \sum_{k=1}^{n} a_{kk}$.*

Definition 2 *A symmetric matrix, $M = M^T \in \mathbb{R}^{n \times n}$, is* <u>*positive semidefinite*</u> *if $x^T M x \geq 0 \ \forall x \in \mathbb{R}^n$. M is* <u>*positive definite*</u> *if $x^T M x > 0 \ \forall x \in \mathbb{R}^n \backslash \{0\}$. M positive semidefinite is denoted $M \succeq 0$ and M positive definite is denoted $M \succ 0$. Finally, M is* <u>*negative (semi)definite*</u> *if $-M$ is positive (semi)definite. These are denoted $M (\preceq) \prec 0$.*

Remark: In the previous definition, we assumed M was symmetric. Any reference to sign definiteness in the following slides will implicitly mean that the matrix is symmetric.

Definition 3 *Given matrices, $F_i = F_i^T \in \mathbb{R}^{n \times n}$ $(i = 0, \ldots, N)$, the following is a* <u>*Linear Matrix Inequality (LMI)*</u> *constraint on $x \in \mathbb{R}^N$:*

$$F(x) := F_0 + x_1 F_1 + \cdots + x_N F_N \succeq 0$$

Definition 4 *A* <u>*semidefinite program (SDP)*</u> *is an optimization of the following form:*

$$\min_x c^T x$$

$$\text{subject to: } \begin{cases} Ax = b \\ F(x) := F_0 + x_1 F_1 + \cdots + x_N F_N \succeq 0 \end{cases}$$

where $F_i = F_i^T \in \mathbb{R}^{n \times n}$ $(i = 0, \ldots, N)$, $A \in \mathbb{R}^{m \times N}$, $b \in \mathbb{R}^{m \times 1}$, and $c \in \mathbb{R}^N$ are given data. In words, an SDP involves minimizing a linear cost subject to linear and LMI constraints.

Fact 1 *For any* $A \in \mathbb{R}^{n\times m}$ *and* $B \in \mathbb{R}^{m\times n}$, $Tr\,[AB] = Tr\,[BA]$.

Proof: This follows from the definition of $Tr\,[\cdot]$. ∎

Remark: A common application is:

$$x^T A x = Tr\left[x^T A x\right] = Tr\left[A x x^T\right]$$

where $x \in R^n$ and $A \in \mathbb{R}^{n\times n}$.

Fact 2 *Define two sets:*

$$\begin{aligned} S_1 &:= \{Q \in \mathbb{R}^{(n+1)\times(n+1)} : Q \succeq 0,\ Q_{11} = 1,\ rank(Q) = 1\} \\ S_2 &:= \{Q \in \mathbb{R}^{(n+1)\times(n+1)} : \exists x \in \mathbb{R}^n \text{ such that } Q = \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]\left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T\} \end{aligned}$$

Then $S_1 = S_2$.

Proof: $S_2 \subset S_1$ follows immediately. We outline the proof for $S_1 \subset S_2$. If $Q \succeq 0$ and $rank(Q) = 1$, then $\exists z \in R^{n+1}$ such that $Q = zz^T$ (e.g. construct the eigenvalue decomposition of Q). Then $Q_{11} = 1$ and $Q \succeq 0$ together imply that $z_1 = 1$. ∎

Fact 3 *If* $A \succeq 0$ *and* $B \succeq 0$ *then* $Tr\,[AB] \geq 0$

Proof: The eigenvalue decomposition of B has the form $B = U\Lambda U^T$ with $\Lambda \succeq 0$. Define $\tilde{A} := U^T A U$. Since A and $\tilde{A}$ are related by a congruence transformation, $A \succeq 0$ implies $\tilde{A} \succeq 0$. Hence the diagonal entries of $\tilde{A}$ are nonnegative. Equality (a) follows from Fact 1:

$$Tr\,[AB] = Tr\left[A(U\Lambda U^T)\right] \overset{(a)}{=} Tr\left[\tilde{A}\Lambda\right] = \sum_{k=1}^{n} \tilde{A}_{kk}\lambda_k \geq 0$$

∎

Fact 4 $\min_x \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T M \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \geq 0$ *if and only if* $M \succeq 0$.

Proof:

($\Longleftarrow$) By definition, $M \succeq 0$ implies $\left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T M \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \geq 0 \; \forall x$.

($\Longrightarrow$) Prove by contradiction. If $M \not\succeq 0$, then there exists a vector, v, such that $v^T M v < 0$. Block partition $v = \left[\begin{smallmatrix}v_1\\v_2\end{smallmatrix}\right]$ where $v_1 \in \mathbb{R}$ is the first entry of v. First assume $v_1 \neq 0$. Given this assumption, we can factor v_1 out:

$$\begin{aligned} 0 > v^T M v = v_1^2 \begin{bmatrix} 1 \\ \frac{1}{v_1} v_2 \end{bmatrix}^T M \begin{bmatrix} 1 \\ \frac{1}{v_1} v_2 \end{bmatrix} \\ \Longrightarrow \min_x \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T M \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] < 0 \end{aligned}$$

If $v_1 = 0$, then by continuity $\left[\begin{smallmatrix}\epsilon\\v_2\end{smallmatrix}\right]^T M \left[\begin{smallmatrix}\epsilon\\v_2\end{smallmatrix}\right] < 0$ for $\epsilon > 0$ sufficiently small. Apply the argument above to show $\min_x \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T M \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] < 0$. ∎

Fact 5 *Given any* $A = A^T \in \mathbb{R}^n$,

$$\min_{Q \succeq 0} Tr\left[AQ\right] = \begin{cases} 0 & \text{if } A \succeq 0 \\ -\infty & \text{else} \end{cases}$$

Proof: By Fact 3, $Q, A \succeq 0$ implies $Tr\left[AQ\right] \geq 0$. Moreover, $Q = 0_n$ achieves $Tr\left[AQ\right] = 0$.

If $A \not\succeq 0$, then there exists an eigenvalue/eigenvector, (λ, v), such that $Av = \lambda v$, $v^T v = 1$, and $\lambda < 0$. For $\alpha > 0$, define $Q_\alpha := \alpha v v^T \succeq 0$. Using Fact 1, $Tr\left[AQ_\alpha\right] = \alpha Tr\left[v^T A v\right] = \alpha\lambda < 0$. As $\alpha \to \infty$, $Tr\left[AQ_\alpha\right] \to -\infty$. ∎

Fact 6 *Given* $A = A^T \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, *and* $C = C^T \in \mathbb{R}^{m\times m}$ *with* $A \succ 0$. *Then:*

$$\begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \succeq 0 \;\Leftrightarrow\; C - BA^{-1}B^T \succeq 0$$

Proof:

$$\begin{bmatrix} A & B^T \\ B & C \end{bmatrix} = \begin{bmatrix} A^{1/2} & 0 \\ BA^{-1/2} & I_m \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & C - BA^{-1}B^T \end{bmatrix} \begin{bmatrix} A^{1/2} & 0 \\ BA^{-1/2} & I_m \end{bmatrix}^T$$

The proof is completed by recalling that positive semidefiniteness is preserved under congruence transformations (i.e. for any matrices M and L, $M \succeq 0 \Rightarrow LML^T \succeq 0$). ∎

Remark: We will apply this result with $A = 1$:

$$\begin{bmatrix} 1 & b^T \\ b & C \end{bmatrix} \succeq 0 \;\Leftrightarrow\; C - bb^T \succeq 0$$

Fact 7 *Let* $\{f_i\}_{i=0}^N$ *and* $\{g_k\}_{k=0}^M$*be given functions from* $\mathbb{R}^n$ *to* $\mathbb{R}$. *The following equality holds:*

$$\min_x f_0(x) \ s.t. \ : \begin{cases} f_i(x) \leq 0, & i = 1, \ldots, N \\ g_k(x) = 0, & k = 1, \ldots, M \end{cases}$$

$$= \min_x \max_{\lambda_i \geq 0, \gamma} \ f_0(x) + \sum_{i=1}^N \lambda_i f_i(x) + \sum_{k=1}^M \gamma_k g_k(x)$$

Proof: Define the Lagrangian:

$$L(x, \lambda, \gamma) := f_0(x) + \sum_{i=1}^N \lambda_i f_i(x) + \sum_{k=1}^M \gamma_k g_k(x)$$

For any x, the Lagrangian satisfies:

$$\max_{\lambda_i \geq 0, \gamma} L(x, \lambda, \gamma) = \begin{cases} f_0(x) & \text{if } f_i(x) \leq 0 \ (i = 1, \ldots, N) \\ & \text{and } g_k(x) = 0 \ (k = 1, \ldots, M) \\ +\infty & else \end{cases}$$

■

Remarks:

- Lagrange multipliers turn constrained minimizations into unconstrained min-max problems.
- Lagrange multipliers can also be used to turn constrained maximizations into unconstrained max-min problems.
- Fact 5 allows us to handle matrix constraints in a similar fashion.. For example, if $F_1 : \mathbb{R}^n \to \mathbb{R}^{m \times m}$, then:

$$\begin{aligned} &\max_x f_0(x) && = \max_x \min_{Q \succeq 0} f_0(x) + Tr\left[F_1(x) Q\right] \\ &\text{s.t.:} \quad F_1(x) \succeq 0 && \end{aligned}$$

Fact 8 *Given a function* $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *and sets* $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$, *the following inequality holds:*

$$\max_{y \in Y} \min_{x \in X} f(x,y) \leq \min_{x \in X} \max_{y \in Y} f(x,y) \tag{3}$$

Proof: Assume X and Y are both nonempty. For any $(x_0, y_0) \in X \times Y$, $f(x_0, y_0) \leq \max_{y \in Y} f(x_0, y)$. Since this inequality holds $\forall (x_0, y_0) \in X \times Y$, it must also hold when we take the minimum over x on both sides:

$$\min_{x \in X} f(x, y_0) \leq \min_{x \in X} \max_{y \in Y} f(x,y)$$

The left side of this inequality is a function of y_0 while the right side is a constant:

$$h(y_0) := \min_{x \in X} f(x, y_0)$$
$$c := \min_{x \in X} \max_{y \in Y} f(x,y)$$

Since $h(y_0) \leq c \ \forall y_0 \in Y$, it must hold for the maximum over y: $\max_{y \in Y} h(y) \leq c$. This is the desired result upon substituting back in for $h(y_0)$ and c.

If X and/or Y are empty, the result follows from our previously specified convention. Specifically, if X is empty then the Equation 3 holds because the right side is equal to $+\infty$. Similarly, if Y is empty then Equation 3 holds because the left side is equal to $-\infty$. ■

Fact 9 *Let* $A = A^T, B = B^T \in \mathbb{R}^{n \times n}$ *be given with* $A \succeq 0$, $B \succeq 0$ *and* $rank(B) = n - r$. *If* $Tr\,[AB] = 0$ *then there exists* $\tilde{A}_{22} \in \mathbb{R}^{r \times r}$, $U \in \mathbb{R}^{n \times n}$ *such that* $U^T U = I_n$ *and* $A = U \begin{bmatrix} 0_{n-r} & 0 \\ 0 & \tilde{A}_{22} \end{bmatrix} U^T$.

Proof: Since $B \succeq 0$ and $rank(B) = n - r$, the eigenvalue decomposition of B has the form $B = U \begin{bmatrix} \Lambda & 0 \\ 0 & 0_r \end{bmatrix} U^T$ where $U \in \mathbb{R}^{n \times n}$, $\Lambda > 0$, and $U^T U = I_n$. Define $\tilde{A} := U^T A U$. Block partition $\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{12}^T & \tilde{A}_{22} \end{bmatrix}$ compatible with the dimensions of $\begin{bmatrix} \Lambda & 0 \\ 0 & 0_r \end{bmatrix}$. Since A and $\tilde{A}$ are related by a congruence transformation, $A \succeq 0$ implies $\tilde{A} \succeq 0$ and hence the diagonal entries of $\tilde{A}_{11}$ are nonnegative.

The following equalities hold:

$$
\begin{aligned}
0 = Tr\,[AB] &\overset{(a)}{=} Tr\left[U^T A U \begin{bmatrix} \Lambda & 0 \\ 0 & 0_r \end{bmatrix}\right] \overset{(b)}{=} Tr\left[\tilde{A}_{11}\Lambda\right] \\
&\overset{(c)}{=} \sum_{k=1}^{n-r} \lambda_k \cdot (\tilde{A}_{11})_{kk} \\
&\overset{(d)}{\Longrightarrow} (\tilde{A}_{11})_{kk} = 0 \text{ for } k = 1, \ldots, n - r
\end{aligned}
$$

(a) follows from Fact 1 while (b) and (c) follow from the definition of $Tr\,[\cdot]$. Implication (d) follows because $\lambda_k > 0$ and $(\tilde{A}_{11})_{kk} \geq 0$ for $k = 1, \ldots, n - r$. The proof is completed by noting this result implies $\tilde{A}_{11} = 0$ and hence $\tilde{A}_{12} = 0$. ∎

Remark: If the $rank(B) = n - r$ is large then $rank(A) = r$ is necessarily small.

The following steps put the NQCQP in an equivalent form:

$$\begin{aligned}
p &:= \min_x \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \text{ s.t. : } \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_i \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \le 0 \;\; i = 1, \ldots, N \\
&\overset{(a)}{=} \min_x Tr\left[Z_0 \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T \right] \text{ s.t. : } Tr\left[Z_i \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T \right] \le 0 \;\; i = 1, \ldots, N \\
&\overset{(b)}{=} \min_{Q \succeq 0,\; Q_{11}=1,\; rank(Q)=1} Tr\left[Z_0 Q\right] \text{ s.t. : } Tr\left[Z_i Q\right] \le 0 \;\; i = 1, \ldots, N
\end{aligned}$$

(a) follows from Fact 1 and (b) follows from Fact 2.

If we 'relax' (i.e. ignore) the rank constraint, we get a semidefinite program (SDP):

$$\begin{aligned}
\underline{p}_r := \min_{Q \succeq 0,\; Q_{11}=1} & Tr\left[Z_0 Q\right] \\
\text{subject to: } & Tr\left[Z_i Q\right] \le 0 \;\; i = 1, \ldots, N
\end{aligned}$$

Remarks:

- Removing the rank constraint means we are searching over a larger set of matrices and hence we can achieve a lower cost, $\underline{p}_r \le p$.
- $\underline{p}_r = p$ if and only if the optimal solution to the SDP is rank 1.

We can use Lagrange multipliers and weak duality to obtain another lower bound on p:

$$p := \min_x \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \quad \text{s.t. : } \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_i \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \leq 0 \quad i = 1, \ldots, N$$

$$\overset{(a)}{=} \min_x \max_{\lambda_i \geq 0} \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] + \sum_{i=1}^N \lambda_i \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] Z_i \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]$$

$$\overset{(b)}{\geq} \max_{\lambda_i \geq 0} \min_x \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T \left[Z_0 + \sum_{i=1}^N \lambda_i Z_i \right] \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]$$

$$\overset{(c)}{=} \max_{\lambda_i \geq 0, \gamma} \gamma \quad \text{s.t. : } \min_x \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T \left[Z_0 + \sum_{i=1}^N \lambda_i Z_i \right] \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \geq \gamma$$

(a) is an application of Lagrange multipliers (Fact 7) and (b) follows from weak duality (Fact 8). Equality (c) just introduces a dummy variable.

We can now apply Fact 4 to convert the constraint in the final maximization into an LMI. This yields the following SDP:

$$\underline{p}_s := \max_{\lambda_i \geq 0, \gamma} \gamma \quad \text{subject to: } Z_0 - \left[\begin{smallmatrix}\gamma & 0\\0 & 0_n\end{smallmatrix}\right] + \sum_{i=1}^N \lambda_i Z_i \succeq 0$$

Remarks:

- In control theory, this is known as the S-procedure.
- By the steps given above, $\underline{p}_s \leq p$. If $\underline{p}_s = p$, the S-procedure is called 'lossless'.
- If $N = 1$, then step (b) holds with equality, i.e. the S-procedure is lossless (Boyd, et.al., 1994).

Weak duality gives that $\underline{p}_s \leq \underline{p}_r$:

$$
\begin{aligned}
\underline{p}_s &:= \max_{\lambda_i \geq 0, \gamma} \ \gamma \ \text{ s.t. : } F(\lambda,\gamma) := Z_0 - \left[\begin{smallmatrix} \gamma & 0 \\ 0 & 0_n \end{smallmatrix}\right] + \sum_{i=1}^{N} \lambda_i Z_i \succeq 0 \quad (4) \\
&\overset{(a)}{=} \max_{\lambda_i \geq 0, \gamma} \min_{Q \succeq 0} \ \gamma + Tr\left[F(\lambda,\gamma)Q\right] \\
&\overset{(b)}{\leq} \min_{Q \succeq 0} \max_{\lambda_i \geq 0, \gamma} \ \gamma + Tr\left[F(\lambda,\gamma)Q\right] \\
&\overset{(c)}{=} \min_{Q \succeq 0, \ Q_{11}=1} Tr\left[Z_0 Q\right] \ \text{ s.t. : } Tr\left[Z_i Q\right] \leq 0 \quad i = 1, \ldots, N \quad (5) \\
&:= \underline{p}_r
\end{aligned}
$$

(a) and (c) use Lagrange multipliers (Fact 7). (b) follows from weak duality (Fact 8). If $\underline{p}_s = \underline{p}_r$, we say strong duality holds.

Theorem 1 $\underline{p}_s = \underline{p}_r$ *if either of the following holds:*

1. *Optimization 4 is strictly feasible:* $\exists(\lambda,\gamma)$ *such that* $\lambda_i > 0$ *and* $F(\lambda,\gamma) \succ 0$.
2. *Optimization 5 is strictly feasible:* $\exists Q \succ 0$ *such that* $Q_{11} = 1$ *and* $Tr\left[Z_i Q\right] < 0$, $i = 1, \ldots, N$.

Remarks:

- The proof is technical (Rockafellar, 1970).
- Conditions for strong duality can be weakened (Sturm, 1997).

If we define $Q := \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right]$, then the rank relaxation problem can be written as:

$$\begin{aligned} \underline{p}_r &:= \min_{\Sigma,\bar{x}} Tr\left[Z_0 \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right]\right] \\ &\text{subject to: } \begin{cases} Tr\left[Z_i \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right]\right] \le 0 \quad i = 1, \dots, N \\ \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right] \succeq 0 \end{cases} \end{aligned}$$

Let X be a random variable (r.v.) of dimension $n \times 1$ with $E[X] = \bar{x}$ and $E\left[XX^T\right] = \Sigma$. By Fact 1 and the linearity of $Tr[\cdot]$ and $E[\cdot]$,

$$E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_i \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right] = Tr\left[Z_i \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right]\right]$$

Thus rank relaxation is equivalent to:

$$\begin{aligned} \underline{p}_r &:= \min_{X} E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right] \\ &\text{s.t. : } \begin{cases} E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right] \le 0 \quad i = 1, \dots, N \\ \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right] \succeq 0 \\ X \text{ is a r.v. with } E[X] = \bar{x},\ E\left[XX^T\right] = \Sigma \end{cases} \end{aligned}$$

By Schur Complements (Fact 6), the second constraint is equivalent to $\Sigma - \bar{x}\bar{x}^T \succeq 0$. This constraint is trivially satisfied if X a random variable because $E\left[(X - \bar{x})(X - \bar{x})^T\right] = \Sigma - \bar{x}\bar{x}^T$ and variance matrices are always positive semidefinite.

Based on this discussion, rank relaxation is equivalent to a minimization involving the random variable X:

$$\begin{aligned} \underline{p}_r &:= \min_{X} E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right] \\ &\text{s.t. : } \begin{cases} E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right] \le 0 \quad i = 1, \dots, N \\ X \text{ is a r.v. with } E[X] = \bar{x},\ E\left[XX^T\right] = \Sigma \end{cases} \end{aligned} \tag{6}$$

The original NQCQP is:

$$
\begin{aligned}
p = \min_x & \left[\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right] \\
\text{s.t. :} & \left[\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right]^T Z_i \left[\begin{smallmatrix} 1 \\ x \end{smallmatrix}\right] \le 0 \quad i = 1, \ldots, N
\end{aligned}
\tag{7}
$$

The rank relaxation problem is:

$$
\begin{aligned}
\underline{p}_r := \min_{\Sigma, \bar{x}} & \, Tr\left[Z_0 \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right]\right] \\
\text{subject to:} & \begin{cases} Tr\left[Z_i \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right]\right] \le 0 \quad i = 1, \ldots, N \\ \left[\begin{smallmatrix} 1 & \bar{x}^T \\ \bar{x} & \Sigma \end{smallmatrix}\right] \succeq 0 \end{cases}
\end{aligned}
\tag{8}
$$

Rank relaxation is equivalent to the following minimization:

$$
\begin{aligned}
\underline{p}_r := \min_X & \, E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right] \\
\text{s.t. :} & \begin{cases} E\left[\left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]^T Z_i \left[\begin{smallmatrix} 1 \\ X \end{smallmatrix}\right]\right] \le 0 \quad i = 1, \ldots, N \\ X \text{ is a r.v. with } E[X] = \bar{x}, \; E\left[XX^T\right] = \Sigma \end{cases}
\end{aligned}
\tag{9}
$$

Remarks:

- Equation 9 is similar to Equation 7, except that we search for a random variable rather than a specific vector.
- Let Σ_0, $\bar{x}_0$ denote an optimal point for Equation 8. Then, any distribution with mean, $E[X] = \bar{x}_0$, and second moment, $E\left[XX^T\right] = \Sigma_0$ is an optimal distribution for Equation 9.
- Recall that if $rank \left[\begin{smallmatrix} 1 & \bar{x}_0^T \\ \bar{x}_0 & \Sigma_0 \end{smallmatrix}\right] = 1$, then $\underline{p}_r = p$. In this case, $\Sigma_0 = \bar{x}_0 \bar{x}_0^T$ and $E\left[(X - \bar{x}_0)(X - \bar{x}_0)^T\right] = 0_n$. Thus the optimal distribution for Equation 9 consists of a single point.

Find upper bounds using the following algorithm:

1. Solve the rank relaxation problem. Let $Q_0 := \begin{bmatrix} 1 & \bar{x}_0^T \\ \bar{x}_0 & \Sigma_0 \end{bmatrix}$ denote an optimal point for the rank relaxation problem.
2. Sample $\mathbb{R}^n$ using any distribution with mean, $E[X] = \bar{x}_0$, and second moment, $E\left[XX^T\right] = \Sigma_0$.
3. Sampled points may not be feasible, but we can use them as initial conditions for a local search algorithm.
4. Any feasible point, x_{feas}, returned by a local search algorithm gives an upper bound: $p \leq \begin{bmatrix} 1 \\ x_{feas} \end{bmatrix}^T Z_0 \begin{bmatrix} 1 \\ x_{feas} \end{bmatrix}$.

Remarks:

- We typically solve the SDP arising from the S-procedure. Given an optimal solution to this problem, we would like to cheaply compute a Q_0. This is covered on the next slide.
- Sedumi solves both lower bound SDPs simultaneously, so the first remark is not an issue.
- In many cases, Q_0 is low rank, so the optimal distribution has nonzero variance along a small number of dimensions.

The SDPs from the S-procedure and Rank Relaxation are:

$$\underline{p}_s := \max_{\lambda_i \geq 0, \gamma} \gamma \;\text{ s.t. : } F(\lambda, \gamma) := Z_0 - \left[\begin{smallmatrix} \gamma & 0 \\ 0 & 0_n \end{smallmatrix}\right] + \sum_{i=1}^{N} \lambda_i Z_i \succeq 0$$

$$\underline{p}_r := \min_{Q \succeq 0,\ Q_{11}=1} Tr\left[Z_0 Q\right] \;\text{ s.t. : } Tr\left[Z_i Q\right] \leq 0 \;\; i = 1, \ldots, N$$

These SDPs are known as primal and dual forms of each other.

If strong duality holds, then there exists feasible $(\bar{\lambda}, \bar{\gamma})$ and $\bar{Q}$ such that $\bar{\gamma} = Tr\left[Z_0 \bar{Q}\right]$. This implies:

$$0 \overset{(a)}{\leq} Tr\left[F(\bar{\lambda}, \bar{\gamma})\bar{Q}\right] \overset{(b)}{=} \sum_{i=1}^{N} \lambda_i Tr\left[Z_i \bar{Q}\right] \overset{(c)}{\leq} 0$$

$$\implies Tr\left[F(\bar{\lambda}, \bar{\gamma})\bar{Q}\right] = 0$$

Inequality (a) holds since $F(\bar{\lambda}, \bar{\gamma})$, $\bar{Q} \succeq 0$ (Fact 3). (b) follows by substituting for $F(\bar{\lambda}, \bar{\gamma})$ and using $\bar{\gamma} = Tr\left[Z_0 \bar{Q}\right]$. Inequality (c) follows by from the feasibility: $\lambda_i \geq 0$ and $Tr\left[Z_i Q\right] \leq 0$.

To summarize, the optimality conditions are given by:

$$\begin{array}{ll} F(\bar{\lambda}, \bar{\gamma}) \succeq 0,\ \bar{\lambda}_i \geq 0 & \text{Primal Feasibility} \\ \bar{Q} \succeq 0,\ \bar{Q}_{11} = 1,\ Tr\left[Z_i \bar{Q}\right] \leq 0 & \text{Dual Feasibility} \\ Tr\left[F(\bar{\lambda}, \bar{\gamma})\bar{Q}\right] = 0 & \text{Complementary Slackness} \end{array}$$

These are known as the Karush-Kuhn-Tucker (KKT) conditions.

Remark: Given $(\bar{\lambda}, \bar{\gamma})$, we can compute a $\bar{Q}$ from the KKT conditions. It typically happens that $F(\bar{\lambda}, \bar{\gamma})$ has large rank. We can apply the complementary slackness condition and Fact 9 to conclude that $\bar{Q}$ must have low rank.

The original NQCQP is:

$$
\begin{aligned}
p = \min_x & \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \\
\text{s.t. :} & \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_i \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \leq 0 \quad i = 1, \ldots, N
\end{aligned}
$$

Consider the following perturbed version of the original NQCQP:

$$
\begin{aligned}
p(u) = \min_x & \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \\
\text{s.t. :} & \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T Z_i \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right] \leq u_i \quad i = 1, \ldots, N
\end{aligned}
$$

When $u = 0$, $p(0)$ is the optimal cost for the original, unperturbed NQCQP. The theorems on the next two slides show that the S-procedure gives local sensitivity information. The proofs and interpretations are minor modifications of results in [Boyd and Vandenberghe].

First we introduce some notation. The SDP from the S-procedure can be written as:

$$
\underline{p}_s := \max_{\lambda_i \geq 0} \; g(\lambda)
$$

where $g(\lambda) := \min_x \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]^T \left[Z_0 + \sum_{i=1}^N \lambda_i Z_i \right] \left[\begin{smallmatrix}1\\x\end{smallmatrix}\right]$. $g(\lambda)$ is known as the dual function. Let λ^* be the optimal vector of S-procedure multipliers, i.e $\underline{p}_s = g(\lambda^*)$.

Theorem 2

$$p(u) \geq \underline{p}_s - \lambda^{*T} u = p(0) - \lambda^{*T} u - \left[p(0) - \underline{p}_s \right]$$

Proof: For any x_0 that is feasible for the perturbed problem:

$$g(\lambda^*) \overset{(a)}{\leq} \left[\begin{smallmatrix} 1 \\ x_0 \end{smallmatrix}\right]^T \left[Z_0 + \sum_{i=1}^{N} \lambda_i^* Z_i \right] \left[\begin{smallmatrix} 1 \\ x_0 \end{smallmatrix}\right] \overset{(b)}{\leq} \left[\begin{smallmatrix} 1 \\ x_0 \end{smallmatrix}\right]^T Z_0 \left[\begin{smallmatrix} 1 \\ x_0 \end{smallmatrix}\right] + \lambda^{*T} u$$

Inequality (a) follows from the definition of the dual function and (b) follows since x_0 satisfies the perturbed constraints. Minimizing the right side over x_0 subject to the constraints of the perturbed problem yields $g(\lambda^*) \leq p(u) + \lambda^{*T} u$. This is the desired inequality since $\underline{p}_s = g(\lambda^*)$. ■

Remarks:

- The bracketed term is bounded by the gap between upper and lower bounds: $p(0) - \underline{p}_s \leq \bar{p} - \underline{p}_s$. This gap is typically small.
- Suppose λ_i^* is large and the gap is negligble. If the i^{th} constraint is tightened (i.e. $u_i < 0$), then the optimal value $p(u)$ will increase greatly.
- Suppose λ_i^* is small and the gap is negligble. If the i^{th} constraint is relaxed (i.e. $u_i > 0$), then the optimal value $p(u)$ will not decrease too much.
- Note that the inequality in Theorem 2 is only a lower bound. Since it is not an upper bound, the interpretations in the previous two bullets are not symmetric.

Theorem 3 *If $p(u)$ is differentiable at $u = 0$, then the gradient of $p(u)$ satisfies:*

$$\left| \left[\nabla_u p(0)\right]^T u + \lambda^{*T} u \right| \leq \left[p(0) - \underline{p}_s \right] + o(u)$$

Proof: Since $p(u)$ is differentiable at $u = 0$, the definition of a gradient gives:

$$p(u) = p(0) + \left[\nabla_u p(0)\right]^T u + o(u)$$

From Theorem 2, the following two inequalities hold:

$$p(u) \geq p(0) - \lambda^{*T} u - \left[p(0) - \underline{p}_s \right]$$

$$p(-u) \geq p(0) + \lambda^{*T} u - \left[p(0) - \underline{p}_s \right]$$

Substituting the Taylor series into these inequalities gives:

$$\left[\nabla_u p(0)\right]^T u + o(u) \geq -\lambda^{*T} u - \left[p(0) - \underline{p}_s \right]$$

$$-\left[\nabla_u p(0)\right]^T u + o(u) \geq \lambda^{*T} u - \left[p(0) - \underline{p}_s \right]$$

The theorem follows from these inequalities. ■

Remarks:

- If the gap is negligible, then the inequality implies $\nabla_u p(0) = -\lambda^*$. In this case, the multipliers are exactly the local sensitivity of the optimal cost with respect to constraint perturbations:

$$\left. \frac{\partial p(u)}{\partial u_i} \right|_{u=0} = -\lambda_i$$

- If the gap is neglible, the interpretations are: Tightening constraint i by a small amount ($u_i < 0$) approximately increases $p(0)$ by $-\lambda_i^* u_i$. Similarly, relaxing this constraint a small amount ($u_i > 0$) approximately decreases $p(0)$ by $-\lambda_i^* u_i$.

Matrix facts can be found in:

- R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press. 1990.

Optimization facts can be found in:

- J. Sturm. *Primal-dual interior point approach to semidefinite programming*. Thesis Publishers, Amsterdam, Netherlands, 1997. Available at: http://fewcal.kub.nl/sturm/.
- R.T. Rockafellar. *Convex Analysis*. Princeton University Press. 1970.
- S. Boyd and L. Vandenberghe. *Convex Optimization*. Draft available at: http://www.stanford.edu/class/ee364/reader.ps.
- S. Boyd, and L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM. 1994.