CO367 Nonlinear Optimization
Department of Combinatorics and Optimization
University of Waterloo

March 3, 2010

NONLINEAR OPTIMIZATION — Remark on Lecture 16

On Pg.1 of slides for Lecture 16, it was claimed that, given $A \in \mathbb{S}^n$ and $b \in \mathbb{R}^n$, the following two problems are equivalent:

$$\begin{array}{lll} \max_\lambda & -b^T(A+\lambda I)^\dagger b - \lambda \\ \text{s.t.} & \begin{cases} A+\lambda I \succeq 0 \\ b \in \mathcal{R}(A+\lambda I) \end{cases} \end{array} \qquad \begin{array}{ll} \max_{\lambda,t} & -t-\lambda \\ \text{s.t.} & \begin{bmatrix} A+\lambda I & b \\ b^T & t \end{bmatrix} \succeq 0. \end{array}$$

First the constraint $\lambda \geq 0$ is missing in both of the optimization problems, because as a dual variable for the inequality constraint $x^T x \leq 1$, there is a sign constraint on λ. But for ease of notation we leave it out for now.

So why are the above problems equivalent? First, you can prove that the following two problems are equivalent:

$$\begin{array}{ll} \max_{\lambda,t} & -t-\lambda \\ \text{s.t.} & A+\lambda I \succeq 0 \\ & t = b^T(A+\lambda I)^\dagger b \\ & b \in \mathcal{R}(A+\lambda I) \end{array} \qquad \begin{array}{ll} \max_{\lambda,t} & -t-\lambda \\ \text{s.t.} & A+\lambda I \succeq 0 \\ & t \geq b^T(A+\lambda I)^\dagger b \\ & b \in \mathcal{R}(A+\lambda I). \end{array}$$

Therefore, it suffices to show that for any (λ, t),

$$\begin{bmatrix} A+\lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \iff \begin{cases} A+\lambda I \succeq 0 \\ t \geq b^T(A+\lambda I)^\dagger b \\ b \in \mathcal{R}(A+\lambda I) \end{cases} \tag{1}$$

To get an idea of how to prove the equivalence, first recall Schur's theorem and its proof:

Theorem 1 *Suppose $A \in \mathbb{R}^{k\times k}$ is symmetric positive definite. Then for $B \in \mathbb{R}^{n\times k}$, $C \in \mathbb{R}^{n\times n}$,*

$$\begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \succeq 0 \iff C - BAB^T \succeq 0 \text{ ; also, } \begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \succ 0 \iff C - BAB^T \succ 0.$$

Proof. Note that if Q is an invertible matrix, for any matrix X of compatible dimension, we have

$$X \succeq 0 \iff QXQ^T \succeq 0 \text{ ; also, } \quad X \succeq 0 \iff QXQ^T \succeq 0.$$

Since

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

and $\begin{bmatrix} I_k & 0 \\ BA^{-1} & I_n \end{bmatrix}$ is invertible, the theorem follows. □

Now return to our constraints $A + \lambda I \succeq 0$ and $t - b^T(A + \lambda I)b \geq 0$. Let's put it on the diagonal to obtain a block diagonal matrix, and "conjugate" it as in the proof of Schur's theorem: for any $y \in \mathbb{R}^n$,

$$\begin{bmatrix} I_n & 0 \\ y^T & 1 \end{bmatrix} \begin{bmatrix} A + \lambda I_n & 0 \\ 0 & t - b^T(A + \lambda I_n)^\dagger b \end{bmatrix} \begin{bmatrix} I_n & y \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} A + \lambda I & 0 \\ y^T(A + \lambda I) & t - b^T(A + \lambda I)^\dagger b \end{bmatrix} \begin{bmatrix} I & y \\ 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} A + \lambda I & (A + \lambda I)y \\ y^T(A + \lambda I) & t - b^T(A + \lambda I)^\dagger b + y^T(A + \lambda I)y \end{bmatrix}.$$

With this we are ready to proof (1). If RHS holds, let $y \in \mathbb{R}^n$ satisfy $(A + \lambda I)y = b$. Since

$$(A + \lambda I)(A + \lambda I)^\dagger(A + \lambda I) = A + \lambda I$$

by definition of Moore-Penrose inverse, $b^T(A + \lambda I)^\dagger b = y^T(A + \lambda I)y$. This shows that

$$\begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} = \begin{bmatrix} A + \lambda I & (A + \lambda I)y \\ y^T(A + \lambda I) & t - b^T(A + \lambda I)^\dagger b + y^T(A + \lambda I)y \end{bmatrix}$$

is positive semidefinite.

Conversely, if LHS of (1) holds, we must have $b \in \mathcal{R}(A + \lambda I)$; if not, the projection b' of b on $\mathcal{R}(A + \lambda I)^\perp = \text{Null}(A + \lambda I)$ is non-zero, and for any $\eta > 0$,

$$\begin{bmatrix} -\eta b'^T & 1 \end{bmatrix} \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \begin{bmatrix} -\eta b' \\ 1 \end{bmatrix} = -2\eta \|b'\|^2 + t < 0$$

for sufficiently large η, contradicting LHS of (1). Hence $b = (A + \lambda I)y$ for some $y \in \mathbb{R}^n$. Together with the block diagonalization, the RHS of (1) holds.

The strong duality result for the primal-dual pair

$$\min \; x^T A x + 2b^T x \quad \text{s.t.} \quad x^T x \leq 1$$
$$\max \; -t - \lambda \quad \text{s.t.} \quad \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0, \; \lambda \geq 0$$

can be found in Appendix B of the text.

Prepared on March 3, 2010.