Example class 2

1. Consider the optimization problem:

$$\underset{x\in\mathbb{R}^n, y\in\mathbb{R}^m}{\text{minimise}} \quad \langle c,x\rangle + \langle d,y\rangle \quad \text{s.t.} \quad Fx + Gy = b, \quad x \geq 0 \tag{1}$$

where F, G, b, c, d are of appropriate sizes, i.e., $F \in \mathbb{R}^{k\times n}$, $G \in \mathbb{R}^{k\times m}$, $b \in \mathbb{R}^k$ and $c \in \mathbb{R}^n, d \in \mathbb{R}^m$. Show how to put (1) into linear programming standard form (Equation (2) in Lecture 3).

Solution: The reason why (1) is not already in standard form is that y is not constrained to be nonnegative. To deal with this we use the fact that any vector $y \in \mathbb{R}^m$ can be written as $y = y_1 - y_2$ where $y_1, y_2 \geq 0$. Thus our problem is equivalent to:

$$\underset{x\in\mathbb{R}^n, y_1,y_2\in\mathbb{R}^m}{\text{minimise}} \quad \langle c,x\rangle + \langle d,y_1 - y_2\rangle \quad \text{s.t.} \quad Fx + y_1 - y_2 = b, \quad x \geq 0, y_1 \geq 0, y_2 \geq 0$$

This problem is now in standard form.

2. Let $A \in \mathbf{S}^n_+$ and $u \in \mathbb{R}^n$. Show that $u^T A u = 0 \iff u \in \ker(A)$.

Solution: The implication $u \in \ker(A) \Rightarrow u^T A u = 0$ is trivial. We show the reverse implication when $A \succeq 0$. Let $A = \sum_{i=1}^n \lambda_i v_i v_i^T$ be an eigenvalue decomposition of A where $\lambda_1, \ldots, \lambda_n \geq 0$. Since $u^T A u = 0$ we get $\sum_{i=1}^n \lambda_i (u^T v_i)^2 = 0$. Since each term in the sum is nonnegative, for the sum to be zero we must have $\lambda_i (u^T v_i)^2 = 0$ for all $i = 1, \ldots, n$. In particular this means that $u^T v_i = 0$ whenever $\lambda_i > 0$. Since $\operatorname{span}\{v_i : \lambda_i > 0\} = \operatorname{im}(A)$ we get that $u \in \operatorname{im}(A)^\perp = \ker(A)$. *(Remark: Another way to prove the implication $u^T A u = 0 \Rightarrow u \in \ker(A)$ (without an eigenvalue decomposition) is to observe that for any $t \in \mathbb{R}, x \in \mathbb{R}^n$: $0 \leq (u+tx)^T A (u+tx) = t^2 x^T A x + t x^T A u$. If we fix x, the fact that $t^2 x^T A x + t x^T A u \geq 0$ for all $t \in \mathbb{R}$ implies necessarily that $x^T A u = 0$. This is true for all x and thus shows that $Au = 0$ i.e., $u \in \ker(A)$).*

3. Let $A \in \mathbf{S}^n$ and R an invertible $n \times n$ matrix. Show that $A \succeq 0 \iff R^T A R \succeq 0$ and $A \succ 0 \iff R^T A R \succ 0$.

Solution: Assume $A \succeq 0$ and let's show that $R^T A R \succeq 0$. For any $y \in \mathbb{R}^n$ we have $y^T R^T A R y = x^T A x \geq 0$ where $x = Ry$. We have shown that $y^T R^T A R y \geq 0$ for all $y \in \mathbb{R}^n$ thus $R^T A R$ is positive semidefinite. The reverse implication is similar. The other statement about positive definite matrices is also similar.

4. (Schur complement) Show that

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succ 0 \iff A \succ 0 \text{ and } C - B^T A^{-1} B \succ 0 \tag{2}$$

Solution: Note that what we want to prove here is a "matrix version" of the following simple fact which can be proved using high school algebra (assuming $a \neq 0$ here):

$$ax^2 + 2bx + c > 0 \; \forall x \in \mathbb{R} \iff a > 0 \textit{ and } b^2 - ac < 0.$$

The proof of the matrix case is in fact not much different from the scalar case.

Observe that for any $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$:

$$\begin{aligned}\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= x^T Ax + 2x^T By + y^T Cy \\ &= (x + A^{-1}By)^T A(x + A^{-1}By) + y^T (C - B^T A^{-1}B)y.\end{aligned} \tag{3}$$

(Equation (3) is nothing but a "completion of square" identity). The proof of the equivalence (2) follows relatively easily from (3). For the direction $\Rightarrow$, simply take $y = 0$ in (3) to show that $A \succ 0$ and $x = -A^{-1}By$ to show that $C - B^T A^{-1}B \succ 0$. The direction $\Leftarrow$ is immediate.

5. (Operator monotonicity of inverse function) Recall that we use the notation $A \succ B$ for $A - B \succ 0$. Show that if $A \succ B \succ 0$ then $A^{-1} \prec B^{-1}$ (*Hint: start with the case* $B = I$ *(identity matrix) then use the fact that* $A \succ B$ *if and only* $B^{-1/2}AB^{-1/2} \succ I$).

   Solution: We first do the case $B = I$. If $A \succ I$ this means that all the eigenvalues of A are > 1. Since the eigenvalues of A^{-1} are the inverses of the eigenvalues of A we have that $A^{-1} \prec I$.

   Consider now an arbitrary B such that $A \succ B \succ 0$. We want to show that $A^{-1} \prec B^{-1}$. Since $A \succ B \succ 0$ we have $B^{-1/2}AB^{-1/2} \succ I$. It thus follows from our previous case that $(B^{-1/2}AB^{-1/2})^{-1} \prec I$. But this means that $B^{1/2}A^{-1}B^{1/2} \prec I$ i.e., $A^{-1} \prec B^{-1}$.

6. (Schur product theorem) Let $A, B \in \mathbf{S}^n$ and assume that $A \succeq 0$ and $B \succeq 0$. Show that $A \odot B \succeq 0$ where $A \odot B$ is the *entrywise product* of A and B, i.e., $(A \odot B)_{ij} = A_{ij}B_{ij}$ (*Hint: start with the case where* A *has rank one*).

   Solution: Assume $A = aa^T$ (rank-one) and $B \succeq 0$. We want to show $A \odot B \succeq 0$. For any $x \in \mathbb{R}^n$ we have

   $$x^T (A \odot B)x = \sum_{ij} x_i x_j (A \odot B)_{ij} = \sum_{ij} x_i x_j A_{ij} B_{ij} = \sum_{ij} x_i x_j a_i a_j B_{ij} = (x \odot a)^T B(x \odot a)$$

   where $x \odot a$ is the vector obtained by componentwise multiplication of x and a. Since $B \succeq 0$ we have $x^T (A \odot B)x = (x \odot a)^T B(x \odot a) \geq 0$. This is valid for any $x \in \mathbb{R}^n$ and thus shows that $A \odot B \succeq 0$.

   We now treat the general case. If $A \succeq 0$ we can decompose A as $A = \sum_{i=1}^n a_i a_i^T$. Hence if $B \succeq 0$ we have $A \odot B = \sum_{i=1}^n (a_i a_i^T) \odot B \succeq 0$. This completes the proof.

7. Compute the duals of the following problems:

   (a) $\underset{x,y \in \mathbb{R}}{\text{minimise}}\ 2x + y \quad \text{s.t.} \quad \begin{bmatrix} 1-x & y \\ y & 1+x \end{bmatrix} \succeq 0$

   (b) $\underset{X \in \mathbf{S}^n}{\text{minimise}}\ \operatorname{Tr}(CX) \quad \text{s.t.} \quad X_{ii} = 1\ \forall i = 1, \ldots, n, \quad X \succeq 0$ (C is a fixed matrix in $\mathbf{S}^n$)

   (c) $\underset{z \in \mathbb{R}^n, y \in \mathbb{R}^m}{\text{maximise}}\ \langle b, y \rangle \quad \text{s.t.} \quad c = z + \mathcal{A}^*(y), z \in K^*$ ($\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ linear, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ are fixed).

   Solution:

   - $\underset{a,b,c}{\text{maximise}}\ -(a + c) \quad \text{s.t.} \quad c - a = 2, \quad b = 1/2, \quad \begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0.$
   - $\underset{Z,\lambda}{\text{maximise}}\ \sum_{i=1}^n \lambda_i \quad \text{s.t.} \quad C = Z + \operatorname{diag}(\lambda), \quad Z \succeq 0.$
   - (Warning: this is a maximisation problem and we have to be careful with the order of inequalities.) Let λ be the dual variable for the linear constraint $c = z + \mathcal{A}^*(y)$ and $x \in K$ be the dual variable for the constraint $z \in K^*$. With these dual variables, we know that any feasible y, z of our problem will satisfy $\langle \lambda, c - z - \mathcal{A}^*(y) \rangle + \langle x, z \rangle \geq 0$. Rearranging this gives $\langle \lambda - x, z \rangle + \langle \mathcal{A}(\lambda), y \rangle \leq \langle \lambda, c \rangle$. Since we are interested in the objective $\langle b, y \rangle$ we want $\lambda - x = 0$

and $\mathcal{A}(\lambda) = b$. The dual problem consists in finding the best *upper bound* on the objective (since our problem was a maximisation). Thus the dual problem takes the form:

$$\begin{array}{ll} \underset{x,\lambda}{\text{minimise}} & \langle \lambda, c\rangle \\ \text{subject to} & \lambda - x = 0 \\ & \mathcal{A}(\lambda) = b \\ & x \in K. \end{array} \tag{4}$$

If we eliminate the variable λ (since $\lambda = x$) we get

$$\text{minimise } \langle c, x\rangle \;:\; \mathcal{A}(x) = b \;\; x \in K.$$

8. Give an example of a proper cone K and linear map M such that $M(K)$ is not closed.

   Solution: Take $K = \mathbf{S}^2_+$ and $M\left(\left[\begin{smallmatrix} x & y \\ y & z\end{smallmatrix}\right]\right) = (x, y)$. Then $M(K) = (\mathbb{R}_{++} \times \mathbb{R}) \cup \{(0,0)\}$.

9. Consider the optimization problem $\min_{x\in\mathbb{R}^n} \|x\|_2$ s.t. $Ax = b, x \geq 0$. Here $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ are fixed. Show that this problem can be expressed as a semidefinite program. [*Hint: Express the constraint* $\|x\|_2^2 \leq t^2$ *as a semidefinite programming constraint*].

   Solution: Using Schur complements we have $x^T x = \|x\|_2^2 \leq t^2$ iff $\left[\begin{smallmatrix} t & x^T \\ x & tI_n\end{smallmatrix}\right] \succeq 0$. Our problem can be written as:

$$\min_{x\in\mathbb{R}^n, t\in\mathbb{R}} \quad t \quad \text{s.t.} \quad Ax = b, x \geq 0, \|x\|_2 \leq t$$

   which is then equivalent to:

$$\min_{x\in\mathbb{R}^n, t\in\mathbb{R}} \quad t \quad \text{s.t.} \quad Ax = b, x \geq 0, \left[\begin{smallmatrix} t & x^T \\ x & tI_n\end{smallmatrix}\right] \succeq 0.$$

   This is a semidefinite program.

10. Let $M_1, \ldots, M_k$ be fixed $n \times n$ symmetric matrices. Consider the optimization problem: $\min_x \; \lambda_{\max}\left(\sum_{i=1}^k x_i M_i\right)$ s.t. $Ax = b$ where A and b are fixed and $\lambda_{\max}$ denotes the largest eigenvalue. Show that it can be expressed as a semidefinite program.

    Solution: Note that $\lambda_{\max}(A) \leq t$ iff $A - tI \preceq 0$. It follows that our problem can be written as:

$$\min_{x\in\mathbb{R}^k, t\in\mathbb{R}} \quad t \quad \text{s.t.} \quad Ax = b, \sum_{i=1}^k x_i M_i - tI \preceq 0.$$

    This is a semidefinite program.

11. Consider the following optimization problem which arises in experiment design (statistics):

$$\text{minimize} \quad \text{Trace}\left[\left(\sum_{i=1}^k x_i M_i\right)^{-1}\right] \quad \text{s.t.} \quad x \geq 0, \sum_{i=1}^k x_i = 1 \tag{5}$$

    where $M_1, \ldots, M_k$ are fixed positive definite matrices. Show that the problem above can be expressed as a semidefinite program. [*Hint: Use the Schur complement lemma to give a semidefinite formulation of the constraint* $\text{Tr}(A^{-1}) \leq t$]

    Solution: Using the Schur complement one can verify that $\text{Tr}(A^{-1}) \leq t$ iff $\exists B$ s.t. $\left[\begin{smallmatrix} A & I \\ I & B\end{smallmatrix}\right] \succeq 0$ and $\text{Tr}(B) \leq t$. Thus our problem is equivalent to:

$$\underset{x\in\mathbb{R}^k, B\in\mathbf{S}^n, t\in\mathbb{R}}{\text{minimize}} \quad t \quad \text{s.t.} \quad x \geq 0, \quad \sum_{i=1}^k x_i = 1, \quad \begin{bmatrix} \sum_{i=1}^k x_i M_i & I \\ I & B\end{bmatrix} \succeq 0, \quad \text{Tr}(B) \leq t.$$

    This is a semidefinite program.

12. (Nesterov's $2/\pi$ result) Let A be a real symmetric matrix of size $n \times n$, and consider the following binary quadratic optimisation problem:

$$\text{maximise} \quad x^T A x \quad : \quad x \in \{-1, 1\}^n. \tag{6}$$

Let v^* be the optimal value of (6).

(a) Consider the semidefinite program:

$$\text{maximise} \quad \mathrm{Tr}(AX) \quad : \quad X \succeq 0 \text{ and } X_{ii} = 1,\ \forall i = 1, \dots, n. \tag{7}$$

Let p^*_{SDP} be the optimal value of (7). Show that $v^* \le p^*_{SDP}$.

From now on we are going to assume that A is positive semidefinite. The purpose of the rest of this problem is to show that $\frac{2}{\pi} p^*_{SDP} \le v^*$. To prove this inequality, we will use a "randomised rounding" scheme similar to the one we saw in lecture for the maximum cut problem.

(b) Let X be the optimal solution (7) and let $v_1, \dots, v_n \in \mathbb{R}^r$ with $r = \mathrm{rank}(X)$ such that $X_{ij} = \langle v_i, v_j \rangle$ for all $i, j = 1, \dots, n$. Define the random variable $y \in \{-1, 1\}^n$ as follows:

$$y_i = \mathrm{sign}(\langle v_i, Z \rangle)$$

where Z is a standard Gaussian variable on $\mathbb{R}^r$. We saw in lecture that

$$\mathbb{E}[y_i y_j] = \frac{2}{\pi} \arcsin(X_{ij}) \quad \forall 1 \le i, j \le n,$$

which you can use without proof. Show that:

$$v^* \ge \mathbb{E}[y^T A y] = \frac{2}{\pi} \mathrm{Tr}(A \arcsin[X]).$$

where $\arcsin[X]$ is the matrix obtained by applying the arcsin function to each entry of X, i.e., $\arcsin[X]_{ij} = \arcsin(X_{ij})$.

(c) Recall the Schur product theorem:

> *Schur product theorem*: If $P \succeq 0$ and $Q \succeq 0$ then $P \odot Q \succeq 0$ where $P \odot Q$ is the entrywise product of P and Q.

Use the Schur product theorem (without proof) to show that if $X \succeq 0$ then $\arcsin[X] - X \succeq 0$. [*Hint: Use the fact that* $\arcsin(x) = \sum_{k=0}^{\infty} \frac{\binom{2k}{k}}{4^k(2k+1)} x^{2k+1}$ *for* $x \in [-1, 1]$].

(d) Using the positive semidefinite assumption on A show then that $\mathrm{Tr}(A \arcsin[X]) \ge \mathrm{Tr}(AX)$. Conclude that $v^* \ge \frac{2}{\pi} p^*_{SDP}$.

Solution:

(a) If x is feasible for (6) then $X = xx^T$ is feasible for (7) and has the same objective function.

(b) By definition of v^* we know that $v^* \ge y^T A y$ with probability one and hence $v^* \ge \mathbb{E}[y^T A y]$. Then we have $\mathbb{E}[y^T A y] = \mathbb{E}[\mathrm{Tr}(A y y^T)] = \mathrm{Tr}(A \mathbb{E}[y y^T]) = \frac{2}{\pi} \mathrm{Tr}(A \arcsin[X])$.

(c) Note that $\arcsin(x) - x$ has a series expansion with nonnegative coefficients, i.e., we can write $\arcsin(x) - x = \sum_{k=0}^{\infty} c_k x^k$ where $c_k \ge 0$ for all $k \in \mathbb{N}$. Thus if $X \succeq 0$ we get $\arcsin[X] - X = \sum_{k=0}^{\infty} c_k X^{\odot k} \succeq 0$ (where $X^{\odot k} = X \odot \cdots \odot X$ (k times)) since each term is positive semidefinite by the Schur product theorem.

(d) We get $v^* \ge \frac{2}{\pi} \mathrm{Tr}(A \arcsin[X]) \ge \frac{2}{\pi} \mathrm{Tr}(AX) = \frac{2}{\pi} p^*_{SDP}$ as desired.