SELECTED SOLUTIONS, SECTION 1.2

1. Prove $\mathbb{S}^n_+$ is a closed convex cone with interior $\mathbb{S}^n_{++}$.

Recall that

$$\mathbb{S}^n_+ = \{X \in \mathbb{R}^{n\times n} : X^T = X \text{ and } x^T X x \geq 0 \text{ for all } x \in \mathbb{R}^n\}.$$

- Each of the sets $\{X \in \mathbb{R}^{n\times n} : X^T = X\}$ and $\{X \in \mathbb{R}^{n\times n} : x^T X x \geq 0\}$ (with arbitrary but fixed $x \in \mathbb{R}^n$) is closed, and $\mathbb{S}^n_+$ is the intersection of all these sets. Thus it is closed as well.
- If $X, Y \in \mathbb{S}^n_+$ and $0 \leq \lambda \leq 1$, then $\lambda X + (1-\lambda)Y \in \mathbb{S}^n$. Moreover, we have for every $x \in \mathbb{R}^n$ that
$$x^T(\lambda X + (1-\lambda)Y)x = \lambda x^T X x + (1-\lambda)x^T Y x \geq 0,$$
because $\lambda \geq 0$, $(1-\lambda) \geq 0$, and $X, Y \in \mathbb{S}^n_+$. This shows that also $\lambda X + (1-\lambda)Y \in \mathbb{S}^n_+$, proving that $\mathbb{S}^n_+$ is convex.
- If $X \in \mathbb{S}^n_+$ and $\lambda \geq 0$, then $\lambda X \in \mathbb{S}^n$ and $x^T(\lambda X)x = \lambda x^T X x \geq 0$ for all $x \in \mathbb{R}^n$, showing that $\lambda X \in \mathbb{S}^n_+$. Thus $\mathbb{S}^n_+$ is a cone.
- Assume that $X \in \mathbb{S}^n_{++}$. Then $\lambda_n(X) > 0$. Now let $Y \in \mathbb{S}^n$ satisfy $\|Y - X\|_2 < \lambda_n(X)$. Then for every $x \in \mathbb{R}^n$ the inequality
$$x^T Y x = x^T X x + x^T(Y-X)x \geq \lambda_n(X)\|x\|^2 - \|Y-X\|\|x\|^2 > 0$$
is satisfied. This proves that $\mathbb{S}^n_{++}$ is contained in the interior of $\mathbb{S}^n_+$.

  On the other hand, assume that $X \in \mathbb{S}^n_+ \setminus \mathbb{S}^n_{++}$. Then we can write X as $X = U^T \operatorname{Diag}(\lambda(X))U$ and $\lambda_n(X) = 0$. Now define X_k, $k \in \mathbb{N}$, as $X_k = U^T \operatorname{Diag}(\lambda(X) - 1/k)U$. Then $X_k \to X$, but $\lambda_n(X_k) = -1/k < 0$ showing that X is not contained in the interior of $\mathbb{S}^n_+$. Thus the interior of $\mathbb{S}^n_+$ is indeed exactly equal to $\mathbb{S}^n_{++}$.

2. Explain why $\mathbb{S}^2_+$ is not a polyhedron.

Write a matrix $X \in \mathbb{S}^2$ as

$$X = \begin{pmatrix} a & c \\ c & b \end{pmatrix}.$$

Then $X \in \mathbb{S}^2_+$, if and only if $a \geq 0$ and $\det(X) = ab - c^2 \geq 0$. In particular, the intersection of $\mathbb{S}^2$ with the hyperplane given by the equation $c = 1$ consists of all a, b satisfying $ab \geq 1$, which is obviously not a polyhedron in $\mathbb{R}^2$.

7. The Fan and Cauchy–Schwarz inequalities.
   (a) For any matrices X in $\mathbb{S}^n$ and U in $\mathbb{O}^n$, prove $\|U^T X U\| = \|X\|$.
   (b) Prove the function λ is norm-preserving.
   (c) Explain why Fan's inequality is a refinement of the Cauchy-Schwarz inequality.

(a) We have

$$\|U^T X U\|^2 = \operatorname{tr}(U^T X U U^T X U) = \operatorname{tr}(U^T X X U) = \operatorname{tr}(U U^T X X) = \operatorname{tr}(X X) = \|X\|^2.$$

Date: October 5, 2015.

(b) Write $X \in \mathbb{S}^n$ as $X = U^T \operatorname{Diag}(\lambda(X))U$. Then

$$\|X\|^2 = \|U^T \operatorname{Diag}(\lambda(X))U\|^2 = \|\operatorname{Diag}(\lambda(X))\|^2$$
$$= \operatorname{tr}(\operatorname{Diag}(\lambda(X))^2) = \sum_k \lambda_k(X)^2 = \|\lambda(X)\|^2.$$

(c) Given X, $Y \in \mathbb{S}^n$, the Cauchy–Schwarz inequality (for symmetric matrices with the inner product $\langle X, Y\rangle = \operatorname{tr}(XY)$) reads as

$$\langle X, Y\rangle \leq \|X\|\|Y\|.$$

On the other hand, we have that

$$\begin{aligned} \langle X, Y\rangle &\leq \lambda(X)^T\lambda(Y) && \text{Fan's inequality} \\ &\leq \|\lambda(X)\|\|\lambda(X)\| && \text{Cauchy–Schwarz inequality in } \mathbb{R}^n \\ &= \|X\|\|Y\| && \text{from (b).} \end{aligned}$$

Thus Fan's inequality refines the Cauchy–Schwarz inequality in the sense that it adds another intermediate estimate to this inequality.

11. For a fixed volumn vector s in $\mathbb{R}^n$, define a linear map $A\colon \mathbb{S}^n \to \mathbb{R}^n$ by setting $AX = Xs$ for any matrix X in $\mathbb{S}^n$. Calculate the adjoint map A^*.

Recall that the adjoint A^* of A is the unique linear map $A^*\colon \mathbb{R}^n \to \mathbb{S}^n$ satisfying

$$\langle A^*x, X\rangle_{\mathbb{S}^n} = \langle x, AX\rangle_{\mathbb{R}^n}$$

for all $x \in \mathbb{R}^n$ and $X \in \mathbb{S}^n$. In the case of this particular mapping A, this means that A^* is defined by the equation

$$\operatorname{tr}((A^*x)X) = x^T Xs$$

for all $x \in \mathbb{R}^n$ and $X \in \mathbb{S}^n$. Now note that, given two vectors a, $b \in \mathbb{R}^n \sim \mathbb{R}^{n\times 1}$, we can write

$$a^T b = \frac{1}{2}\operatorname{tr}(ab^T + ba^T)$$

Thus, using the symmetry of X, we obtain that

$$\langle A^*x, X\rangle_{\mathbb{S}^n} = \operatorname{tr}((A^*x)X) = x^T Xs = \frac{1}{2}\operatorname{tr}(x(Xs)^T + (Xs)x^T)$$
$$= \frac{1}{2}\operatorname{tr}((xs^T)X^T + X(sx^T)) = \frac{1}{2}\operatorname{tr}((xs^T)X + (sx^T)X) = \operatorname{tr}\Big(\frac{1}{2}(xs^T + sx^T)X\Big).$$

This shows that

$$A^*x = \frac{1}{2}(xs^T + sx^T)$$

for every $x \in \mathbb{R}^n$.

12* (Fan's inequality) For vectors x and y in $\mathbb{R}^n$ and a matrix U in $\mathbb{O}^n$, define

$$\alpha = \big\langle \operatorname{Diag} x, U^T(\operatorname{Diag} y)U\big\rangle.$$

(a) Prove $\alpha = x^T Zy$ for some doubly stochastic matrix Z.
(b) Use Birkhoff's theorem and Proposition 1.2.4 to deduce the inequality $\alpha \leq [x]^T[y]$.
(c) Deduce Fan's inequality (1.2.2).

(a) We can write

$$\left\langle \operatorname{Diag} x, U^T(\operatorname{Diag} y)U \right\rangle = \operatorname{tr}((\operatorname{Diag} x)U^T(\operatorname{Diag} y)U) = \sum_i x_i(U^T(\operatorname{Diag} y)U)_{ii}.$$

Moreover

$$(U^T(\operatorname{Diag} y)U)_{ii} = \sum_k (U^T)_{ik}((\operatorname{Diag} y)U)_{ki} = \sum_k U_{ki} y_k U_{ki} = \sum_k y_k (U_{ki})^2.$$

Thus

$$\left\langle \operatorname{Diag} x, U^T(\operatorname{Diag} y)U \right\rangle = \sum_{i,k} x_i (U_{ki})^2 y_k.$$

Defining $Z \in \mathbb{R}^{n\times n}$ by

$$Z_{ik} = (U_{ki})^2,$$

we see that

$$\left\langle \operatorname{Diag} x, U^T(\operatorname{Diag} y)U \right\rangle = x^T Z y.$$

Moreover, for every i we have

$$\sum_k Z_{ik} = \sum_k (U_{ki})^2 = 1$$

and

$$\sum_k Z_{ki} = \sum_k (U_{ik})^2 = 1$$

because U is orthogonal. Thus Z is doubly stochastic.

(b) Using Birkhoff's theorem, we see that we can write

$$Z = \sum_j \lambda_j P_j$$

for some permutation matrices $P_j \in \mathbb{R}^{n\times n}$ and $0 \leq \lambda_j \leq 1$ satisfying $\sum_j \lambda_j = 1$. Note here that $[P_j y] = [y]$ because $P_j y$ is only a permutation of y. Thus, using the Hardy–Littlewood–P'olya theorem, we see that

$$\alpha = x^T Z y = \sum_j \lambda_j x^T (P_j y) \leq \sum_j \lambda_j [x]^T [P_j y] = \sum_j \lambda_j [x]^T [y] = [x]^T [y].$$

(c) Write $X = V_1^T \lambda(X) V$ and $Y = V_2^T \lambda(Y) V_2$. Then, applying part (b) with $x = \lambda(X)$, $y = \lambda(Y)$, and $U = V_2 V_1^T$, we see that

$$\begin{aligned} \operatorname{tr}(XY) &= \operatorname{tr}(V_1^T(\operatorname{Diag} \lambda(X))V_1 V_2^T(\operatorname{Diag} \lambda(Y))V_2) \\ &= \operatorname{tr}((\operatorname{Diag} \lambda(X))V_1 V_2^T(\operatorname{Diag} \lambda(Y))V_2 V_1^T) \leq [\lambda(X)]^T[\lambda(Y)] = \lambda(X)^T \lambda(Y), \end{aligned}$$

because the vectors $\lambda(X)$ and $\lambda(Y)$ are ordered anyway.