EE 127/227AT 9/26/2017
G. Calafiore

Quiz 1

NAME:
SID:

The quiz lasts 1h30. Notes are *not* allowed, except for a one-page, two sided cheat sheet.

1. (a) Consider the trace function on a square matrix $X \in \mathbb{R}^{n,n}$, defined as the sum of the elements on the diagonal of X: $\text{trace}(X) = \sum_{i=1}^{n} X_{ii}$. Prove that:
   i. $\text{trace}(A) = \text{trace}(A^\top)$, for any square matrix A;
   ii. $\text{trace}(AB) = \text{trace}(BA)$, for any square matrices A, B;
   iii. Prove that if a square matrix A is diagonalizable, then the trace of A equals the sum of all the eigenvalues of A (this fact actually holds for general A, but you are here asked to prove it only for the diagonalizable case).

   (b) Consider the space $\mathbb{R}^{m,n}$ of $m \times n$ matrices as a vector space, endowed with the inner product $\langle A, B\rangle \doteq \text{trace}(A^\top B)$, for any $A, B \in \mathbb{R}^{m,n}$. Prove that $\langle A, B\rangle \doteq \text{trace}(A^\top B)$ is indeed an inner product on $\mathbb{R}^{m,n}$, that is, it satisfies the following three properties:
   i. $\langle A, A\rangle \geq 0$ for all $A \in \mathbb{R}^{m,n}$, and $\langle A, A\rangle = 0$ if and only if $A = 0$;
   ii. $\langle A, B + C\rangle = \langle A, B\rangle + \langle A, C\rangle$;
   iii. $\langle \alpha A, B\rangle = \alpha\langle A, B\rangle$;
   iv. $\langle A, B\rangle = \langle B, A\rangle$.

SOLUTION

(a) (i) The diagonal elements of a matrix and its transpose are the same. Since the trace is the sum of the diagonal elements, A and $A^\top$ have the same trace.

(ii) The i-th diagonal element of AB is $[AB]_{ii} = \sum_j a_{ij}b_{ji}$ and the j-th diagonal element of BA is $[BA]_{jj} = \sum_i b_{ji}a_{ij} = \sum_i a_{ij}b_{ji}$. Since the trace is the sum of the diagonal elements, we have that

$$\text{trace}(AB) = \sum_i [AB]_{ii} = \sum_i \sum_j a_{ij}b_{ji} = \sum_j \sum_i a_{ij}b_{ji} = \text{trace}(BA).$$

(iii) If A is diagonalizable, it can be factored as $A = P\Lambda P^{-1}$, where Λ is diagonal. From (ii) it follows that

$$\mathrm{trace}(A) = \mathrm{trace}(P\Lambda P^{-1}) = \mathrm{trace}(\Lambda P^{-1}P) = \mathrm{trace}(\Lambda) = \sum_i \lambda_i.$$

(b) (i) Since the diagonal element of $A^\top A$ are the squared norms of the columns, they are nonnegative, whence $\langle A, A\rangle = \mathrm{trace}(A^\top A) = \sum_i \|a_i\|_2^2 \geq 0$, and this quantity can be zero if and only if the norms of all columns of A are zero, which happens if and only if A is zero.

(ii) Follows from linearity of the trace.

(iii) Follows from linearity of the trace.

(iv) Follows from point (i) in part a).

2. We are given two sets of points in $\mathbb{R}^n$: $\mathcal{A} = \{a_1, \ldots, a_m\}$ and $\mathcal{B} = \{b_1, \ldots, b_m\}$, with $a_i, b_i \in \mathbb{R}^n$, $i = 1, \ldots, m$. Let $A \doteq [a_1 \cdots a_m] \in \mathbb{R}^{n,m}$ and $B \doteq [b_1 \cdots b_m] \in \mathbb{R}^{n,m}$. We know that the points in $\mathcal{B}$ are related to the points in $\mathcal{A}$ via an orthogonal map, plus noise, that is

$$b_i = Qa_i + \epsilon_i, \quad i = 1, \ldots, m,$$

where ϵ_i are unknown noise terms, and $Q \in \mathbb{R}^{n,n}$ is an unknown orthogonal matrix.

We are interested in approximately "recovering" the unknown Q matrix from the data in A and B. More precisely, we seek an orthogonal matrix $\hat{Q}$ that minimizes $\|B - QA\|_F^2$ over all orthogonal matrices Q, i.e., in solving

$$\hat{Q} = \arg \min_{\{Q:\, QQ^\top = I\}} \|B - QA\|_F^2.$$

(a) Prove that $\|YAX\|_F^2 = \|A\|_F^2 =$ for any matrix A and orthogonal matrices X, Y.

(b) Show that the optimal solution $\hat{Q}$ can be expressed in terms of the Singular Value Decomposition (SVD) of the matrix $M = BA^\top$. Find this solution explicitly in terms of the SVD factors of $BA^\top = U\Sigma V^\top$.

SOLUTION

(a) Since X, Y are orthogonal, we have that $Y^\top Y = I$ and $XX^\top = I$, and by the commutativity property of the trace we obtain that

$$\begin{aligned} \|YAX\|_F^2 &= \mathrm{trace}(X^\top A^\top Y^\top YAX) = \mathrm{trace}(X^\top A^\top AX) = \mathrm{trace}(A^\top AXX^\top) \\ &= \mathrm{trace}(A^\top A) = \|A\|_F^2. \end{aligned}$$

(b) Let $M \doteq BA^\top = U\Sigma V^\top$, with U, V orthogonal, and Σ such that $\Sigma\Sigma^\top$ is diagonal. Then,

$$\begin{aligned} \|B - QA\|_F^2 &= \mathrm{trace}(B - QA)^\top (B - QA) \\ &= \mathrm{trace}\left(B^\top B - B^\top QA - A^\top Q^\top B + A^\top Q^\top QA\right) \\ &= \|B\|_F^2 + \|A\|_F^2 - 2\,\mathrm{trace}\, A^\top Q^\top B \\ &= \|B\|_F^2 + \|A\|_F^2 - 2\,\mathrm{trace}\, Q^\top BA^\top = \|B\|_F^2 + \|A\|_F^2 - 2\,\mathrm{trace}\, Q^\top M \\ &= \|B\|_F^2 + \|A\|_F^2 - 2\,\mathrm{trace}\, Q^\top U\Sigma V^\top \\ &= \|B\|_F^2 + \|A\|_F^2 - 2\,\mathrm{trace}\, V^\top Q^\top U\Sigma. \end{aligned}$$

Now, $\tilde{Q} \doteq V^\top Q^\top U$ is an orthogonal matrix, and $\tilde{Q}\Sigma$ is a matrix formed by the columns of $\tilde{Q}$ multiplied by the diagonal elements in Σ (which are nonnegative). The trace of $\tilde{Q}\Sigma$ is thus maximized if we choose $\tilde{Q} = I$, whence

$$V^\top Q^\top U = I \quad \Rightarrow \quad Q = UV^\top.$$

3. Given the following three points in $\mathbb{R}^5$

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \\ 1 \end{bmatrix}, \; x_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \; x_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

compute:

(a) the projections $\hat{x}_1, \hat{x}_2, \hat{x}_3$ of x_1, x_2, x_3 onto the line $\mathcal{L} \doteq \{x = \bar{x} + \gamma v, \gamma \in \mathbb{R}\}$, where

$$\bar{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix};$$

(b) the distances d_1, d_2, d_3 of the points x_1, x_2, x_3 from the hyperplane

$$\mathcal{H} \doteq \{x : \sum_{i=1}^{5} x_i = 0\}.$$

SOLUTION

(a) Let $u = v/\|v\|_2 = v/\sqrt{(5)}$. Using the formula on page 40 of the book, we obtain

$$\hat{x}_i = \bar{x} + u^\top (x_i - \bar{x}) u, \quad i = 1, 2, 3.$$

(b) Let $\mathbf{1}$ denote the vector of all ones. The hyperplane is defined by $\mathbf{1}^\top x = 0$, hence from eq. (2.6) in the book, we have

$$d_i = \frac{|\mathbf{1}^\top x_i|}{\|\mathbf{1}\|_2} = \frac{1}{\sqrt{5}} |\mathbf{1}^\top x_i|.$$