Linear algebra and learning from data, Exam 2022-10-25

(1) (a) Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite show that $B^tAB > 0$ if and only if the null space $\mathcal{N}(B) = \{0\}$, where $B \in \mathbb{R}^{n\times k}$.

(b) Let $\langle A, B\rangle = \mathrm{tr}(A^tB)$ for any $A, B \in \mathbb{R}^{n\times n}$. Show that this defines an inner product on the vector space $\mathbb{R}^{n\times n}$, and the Frobenius norm is the induced norm of this inner product.

Solution. (a) The matrix A is positive definite $\Leftrightarrow$ $A = A^{1/2}A^{1/2}$ with $A^{1/2}$ symmetric and invertible. Note that $B^tAB > 0 \Leftrightarrow x^tB^tABx > 0$, $\forall x \in \mathbb{R}^n \setminus \{0\} \Leftrightarrow (A^{1/2}Bx)^t(A^{1/2}Bx) > 0 \Leftrightarrow \|A^{1/2}Bx\|_2^2 > 0$. So the vector $A^{1/2}Bx \neq 0$ which implies $Bx \neq 0$, i.e. $\mathcal{N}(B) = \{0\}$, since x is an arbitrary nonzero vector. The converse follows by noting that $B^tAB \geq 0$ for all x. If it were not positive definite then $Bx = 0$, which is a contradiction.

(b) The linearity and commutativity are trivial. So we only show $\langle A, A\rangle = 0$ only if $A = 0$. Since $\mathrm{tr}(A^tA)$ is the sum of the eigenvalues of A^tA, λ_i, and the eigenvalues are non-negative, all the eigenvalues must be zero if their sum is zero. So A must be the zero matrix. It is apparent that $\|A\|_F^2 = \sum_i \sigma_i^2 = \sum_i \lambda_i = \langle A, A\rangle$.

(2) Let $A \in \mathbb{R}^{n\times n}$ and we define the induced matrix norm: $\|A\| = \max_{x\neq 0} \frac{\|\|Ax\|}{\|x\|}$ where $\|\cdot\|$ is any vector norm in $\mathbb{R}^n$.

(a) Justify the well-definedness of this definition.

(b) Show that $\|A\|_2$, using the vector norm $\|\cdot\|_2$, is the largest singular value of A.

(c) Let $A = I$ be the $n \times n$ identity matrix. Determine $\|I\|_2$, $\|I\|_F$ and $\|I\|_N$. Next let $A = Q$ be an orthogonal matrix. Determine $\|Q\|_2$, $\|Q\|_F$ and $\|Q\|_N$. As a reminder F and N stand for Frobenius and nuclear, respectively.

(d) Find the relations between these three norms for any square matrix A. Show also that $\|A\|_F = \|A\|_2$ if A is a rank 1 matrix.

(e) Define $\kappa = \|A\|\|A^{-1}\|$, the condition number that measures conditioning of the matrix A in solving $Ax = b$. However, in the least square problems the matrix A is $m \times n$ and $m \geq n$. Modify the definition of $\kappa(A)$ using $\|\cdot\|_2$ which will be the same as defined before if $m = n$.

Solution. (a) It is well defined because $\|A\| = \max_{x\neq 0} \frac{\|\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$ and the norm is a continuous function, the maximum is taken on a compact set (a unit sphere).

(b) $\|Ax\|_2^2 = x^t(A^tA)x$ is a quadratic form the maximum is the largest eigenvalue of A^tA, and it is achieved at its associated eigenvector, i.e. $\|A\|_2$ is the largest singular value of A.

(c) $\|I\|_2 = 1$, $\|I\|_F = \sqrt{n}$, $\|I\|_N = n$;
$\|I\|_2 = \|Q\|_2 = 1$, $\|I\|_F = \|Q\|_F = \sqrt{n}$, $\|I\|_N = \|IQ\|_N = n$.

(d) It is easy to show that $\|A\|_2^2 \leq \|A\|_F^2 \leq \|A\|_N^2$ (using the fact that the singular values are nonnegative). In fact we can show that these norms are equivalent: assuming r is the rank of A,

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{r}\|A|_2 \text{ and } \|A\|_F \leq \|A\|_N \leq \sqrt{r}\|A|_F,$$

meaning the 2-norm is equivalent to the F-norm which is equivalent to the N-norm.

(e) For define $\kappa_2(A) = \frac{\max_{\|x\|_2=1}\|Ax\|_2}{\min_{\|x\|_2=1}\|Ax\|_2}$. By definition, if $m = n$ and A is invertible, we have

$$\|A\|_2 = \max_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1}\|Ax\|_2,$$

and

$$\|A^{-1}\|_2 = \max_{x\neq 0}\frac{\|A^{-1}x\|_2}{\|x\|_2} = \frac{1}{\min_{\|y\|_2=1}\|Ay\|_2},$$

where $y = A^{-1}x$.

In fact $\kappa_2(A) = \frac{\sigma_1(A)}{\sigma_n(A)}$, i.e the ratio of the largest and the smallest sigular value of A. (Prove it!)

(3) Given the diagonal matrix $A = \mathrm{diag}(4,3,2,1)$. What is the best 2-rank approximation of A (in which sense)? State the general result for the approximation of any A with low rank. Give some application areas of this theorem.

Solution. The best 2-rank approximation is $A_2 = \mathrm{diag}(4,3,0,0)$, by the Eckart-Young theorem. It is optimal in the sense that the 2- and the F-norm of $A - B$ is minimized over all the matrices B that have a given lower rank than the rank of A.

(4) (a) Consider the function $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x) = x^tAx$ where $A = \begin{pmatrix} 2 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 2 & \theta \end{pmatrix}$. Find the Hessian of f H without computing the partial derivatives. For what values of θ is f strictly convex?

(b) Argue that the matrix H can be written as a sum of rank one matrices.

(c) What is the smallest eigenvalue of H, without computing, if $\theta = 2$?

Solution. (a) The Hessian matrix H is $\frac{1}{2}(A + A^t) = \frac{1}{2}\begin{pmatrix} 4 & 3 & 4 \\ 3 & 6 & 3 \\ 4 & 3 & 2\theta \end{pmatrix}$. It is positive definite if $\theta > 2$ using the determinant test. Then f is strictly convex.

(b) Since H is symmetric we have, by the spectral theorem, $H = Q\Lambda Q^t$, where $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$, λ_i, $i = 1, ..., n$ are eigenvalues of H and $Q = (q_1, ..., q_n)$ is an orthogonal matrix with $Hq_i = \lambda_i q_i$. That is $H = \sum_{i=1}^n \lambda_i q_i q_i^t$.

(c) The smallest eigenvalue of H is 0 if $\theta = 2$ because H is positive semi-definite.

(5) Let $A = \begin{pmatrix} B & b \\ b^t & a \end{pmatrix} \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{(n-1)\times(n-1)}$ be symmetric and $b \in \mathbb{R}^{n-1}$. Assume that A has eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and B has eigenvalues $\mu_1 \leq \cdots \leq \mu_{n-1}$. Show that

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \cdots \leq \mu_{n-1} \leq \lambda_n$$

Solution.. There are many ways to prove it. We give one here. We prove

(i) $\lambda_k \leq \mu_k$ (ii) $\lambda_{k+1} \geq \mu_k$ for $k = 1, ..., n-1$.

"$\lambda_k \leq \mu_k$": Let $x_1, ..., x_n$ be eigenvectors of A and $y_1, ..., y_{n-1}$ be eigenvectors of B. Define the following subspaces

$$V = \mathrm{span}\{x_k, ..., x_n\}, W = \mathrm{span}\{y_1, .., y_k\}, U = \left\{\begin{pmatrix} w \\ 0 \end{pmatrix} \in \mathbb{R}^n : w \in W\right\}.$$

Since $\dim V = n-k+1$ and $\dim U = \dim W = k$, there is $u \in V \cap U$ and $u = \binom{w}{0}$ for some $w \in W$. Obviously $u^t A u = w^t B w$. Recall that

$$\lambda_k = \min_{x\in V} \frac{x^t Ax}{x^t x} \text{ and } \mu_k = \max_{x \in W} \frac{x^t Bx}{x^t x}$$

which yields

$$\mu_k = \max_{x\in W} \frac{x^t x Bx}{x^t x} \Rightarrow \lambda_k \le \frac{u^t Au}{u^t u} = \frac{w^t Bw}{w^t w} \le \mu_k.$$

"$\lambda_{k+1} \ge \mu_k$ ": Similarly we define the subspaces

$$V = \operatorname{span}\{x_1, ..., x_{k+1}\}, W = \operatorname{span}\{y_k, .., y_{n-1}\}, U = \left\{\binom{w}{0} \in \mathbb{R}^n : w \in W\right\}.$$

Then $\dim V = k+1$ and $\dim U = \dim W = n-k$, there is $u \in V \cap U$ and $u = \binom{w}{0}$ for some $w \in W$. Obviously $u^t Au = w^t Bw$. Then

$$\lambda_{k+1} = \max_{x\in V} \frac{x^t Ax}{x^t x} \ge \frac{u^t Au}{u^t u} = \frac{w^t Bw}{w^t w} \ge \min_{x\in W} \frac{x^t Bx}{x^t x} = \mu_k.$$

(6) Fisher's LDA attempts to find a separation vector onto which the projection of different classes are "best separated" by solving the optimization problem $\max_{\|v\|\neq 0} \frac{(v^t m_A - v^t m_B)^2}{v^t(\Sigma_A + \Sigma_B)v}$ where m_C, Σ_C are sampled mean and covariance matrices for $C \in \{A, B\}$ the two classes. Find an optimal solution. Argue how you will deal with the situation where $\Sigma_A + \Sigma_B$ is not positive definite or this matrix is nearly singular.

Solution. See Lecture notes Day 4.

(7) (a) Given two n-vectors a and x, define their circular convolution $y = a * x$ as $y_y = \sum_{l=0}^{n-1} a_{k-l} x_l$, where the indices in the sum are evaluated modulo n. Show that the circular convolution is commutative and associative.
(b) Assume that the matrix A has simple eigenvalues. Show that A and B are simultaneously diagonalizable if and only if they commute. In this case the diagonalizing basis is made up of the eigenvectors of A.
(c) Let S and its adjoint S^* be the circular shift operators defined by $S(x_0, , , , x_{n-1}, x_n) = (x_{n-1}, x_0, ..., x_{n-2})$ and $S^*(x_0, , , , x_{n-1}, x_n) = (x_1, , ..., x_{n-1}, x_0)$, respectively. Show that any matrix M that commutes with the circular shift operator S must be a circulant matrix.
(d) Find all eigenvalues of S^* and their corresponding eigenvectors. Justify that the operator S^* on $\mathbb{R}^n$ has n distinct eigenvalues.
(e) Show that any circulant matrix C has the same eigenvectors as those of S^*.

Solution. See lecturenotes Day 10.

(8) Let $A \in \mathbb{R}^{m\times n}$ with $m > n$. Consider the equation $Ax = b$.
(a) Show how you derive a solution if $A^t A$ is not invertible.
(b) Describe the gradient descent method for solving the least square problem $\min \|b - Ax\|_2^2$ assuming A has full column rank.
(c) Find the conditions for the convergence of this method and derive the convergence rate.

Solution. See lecturenotes Day 6 and Day 7.

(9) How do you solve the real polynomial equation $p(s) = s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0$ using linear algebra?

Solution. Note that $p(s)$ is the characteristic polynmial of the companion matrix

$$C = \begin{pmatrix} 0 & \cdots & 0 & -p_0 \\ 1 & \cdots & 0 & -p_1 \\ \vdots & \ddots & 0 & \vdots \\ 0 & \cdots & 1 & -p_{n-1} \end{pmatrix}.$$

(show it!) Then we can use the two-step QR algorithm, i.e. first we convert the matrix to an upper Hessenberg form then apply the QR algorithm, to find the eigenvalues of C. A more efficient way is to split C as $C = Q + uv^t$ where Q is the circular shift matrix and $u^t = (-p_0 - 1, -p_1, ..., -p_{n-1})$ and $v^t = e_n = (0, ..., 0, 1)$. Now we can apply the fast QR algorithm for the rank-1 update..

Note that the main problem of using iterative method to solve polynomial equations is when the polynomials have multiple zeros. This is nevertheless not a problem for eigenvalue problems.