Exercise 7 (6 points)

Let A be a symmetric matrix.

Show that the following are equivalent:

1. A is positive semidefinite, i.e. $v^T Av \geq 0$ for all v.
2. All eigenvalues of A are nonnegative.
3. There exists a matrix B such that $A = BB^T$.

[Hint: Use Theorem 1.5.3 which we will prove on Tuesday. It says: For a symmetric matrix A there exists an orthogonal matrix U such that $U^T AU$ is a diagonal matrix.]

Prep. Since A is symmetric:

- $A^T = A$, A is a square matrix. Assume $A = \{a_{ij}\}_{n\times n}$ $n \in \mathbb{N}^+$
- $\exists U_{n\times n}$, s.t. $U^T A U = \Lambda$, where $\Lambda_{n\times n}$ is diagonal and U is orthogonal

Λ can be written as $\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$, where $\lambda_1, \dots, \lambda_n \in \mathbb{R}$

Find eigenvalues λ of A:

$$\lambda I - A \overset{U^TU=I}{=\!=} \lambda U^T I U - U^T \Lambda U$$
$$= U^T(\lambda I - \Lambda) U$$

$$\chi_A(\lambda) = \det(U^T(\lambda I - \Lambda) U)$$
$$= \det(U^T)\det(\lambda I - \Lambda)\det(U)$$
$$= \underbrace{\det(U^{-1})\cdot\det(U)}_{1}\cdot\det(\lambda I - \Lambda) = 0$$

$$\chi_A(\lambda) = \prod_{i=1}^{n}(\lambda - \lambda_i) = 0 \quad \Rightarrow \lambda = \lambda_i, \ i = 1, \dots, n$$

so the eigenvalues of A are the diagonal entries of Λ.

Proof. Assume the vector space of the vector v in statement 1 is V: $V \cong \mathbb{R}^n$

2. $\Longleftrightarrow$ 1.

All eigenvalues of A is non-negative

All eigenvalues of A is non-negative

$\Longleftrightarrow$ The diagonal matrix Λ can be written as

$$\Lambda = \begin{pmatrix} \lambda_1^2 & & 0 \\ & \ddots & \\ 0 & & \lambda_n^2 \end{pmatrix} \qquad \lambda_i \in \mathbb{R},\ i=1,\dots,n$$

$\Longleftrightarrow$ $\forall v \in V \qquad v^T A v = v^T (U^T \Lambda U) v$

$$= (Uv)^T \Lambda (Uv)$$

we define vector $Uv = u = (u_1, \dots, u_n)^T \quad u_i \in \mathbb{R},\ i=1,\dots,n$

$\Longleftrightarrow$ $v^T A v = u^T \Lambda u = (u_1, \dots, u_n) \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$

$$= \sum_{i=1}^{n} \lambda_i^2 u_i^2 \geq 0 \qquad \square$$

$2 \Longleftrightarrow 3$:

"$\Rightarrow$" $A = U^T \Lambda U = U^T \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{pmatrix} U$

$$= U^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U$$

we define matrix $B_{n \times n} = U^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}^T$

$$\Rightarrow A = B \cdot B^T$$

"$\Rightarrow$" If $A = B \cdot B^T$

$\forall v \in V, \quad v^T A v = v^T B B^T v$

$$= (B^T v)^T (B^T v)$$

we can define vector $B^T v = w = (w_1, \dots, w_n) \quad w_i \in \mathbb{R},\ i=1,\dots,n$

$$v^T A v = w^T w = \sum_{i=1}^{n} w_i^2 \geq 0$$

since $1 \Longleftrightarrow 2$, $3 \Rightarrow 2$

Exercise 8 (10 points)

Let $\mathrm{Mat}_{n\times m}$ be the vector space of all $n\times m$-matrices, and let vec: $\mathrm{Mat}_{n\times m} \to \mathbb{R}^{n\cdot m}$ be the isomorphism that stacks the columns of a given matrix on top of each other, obtaining a long vector.

The *Frobenius inner product* is the map $\langle -,-\rangle_F\colon \mathrm{Mat}_{n\times m}\times \mathrm{Mat}_{n\times m} \to \mathbb{R}$ defined by $\langle A,B\rangle_F := \mathrm{tr}(A^TB)$.

(a) (4 points) Show that $\langle A,B\rangle_F = \langle \mathrm{vec}(A), \mathrm{vec}(B)\rangle$ (where the latter denotes the usual scalar product of $\mathbb{R}^{n\cdot m}$, i.e. $\mathrm{vec}(A)^T\mathrm{vec}(B)$).

(b) (2 points) Show that the Frobenius inner product is actually an iner product in the sense of Definition 1.4.6 of the manuscript.

(c) (4 points) Like any inner product, the Frobenius inner product has an associated norm: The Frobenius norm is given by $\|A\|_F := \sqrt{\langle A,A\rangle_F}$. Show that, if B is an orthogonal matrix, then $\|A\|_F = \|AB\|_F = \|BA\|_F$

$\forall A \in \mathrm{Mat}_{n\times m}$ can be written as:

$A_{n\times m} = (a_1 | \cdots | a_m)$, where $a_i \in \mathbb{R}^n$ are the column vectors of A.
$i=1,\dots,n$

similarly, $B_{n\times m} := (b_1|\dots|b_m)$, $b_i \in \mathbb{R}^n$, $i=1,\dots,n$.

$$\mathrm{vec}(A)_{nm\times 1} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}, \quad \mathrm{vec}(B)_{nm\times 1} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

(a)

$$A^TB = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} (b_1|\cdots|b_m) = \begin{pmatrix} a_1^Tb_1 & \cdots & a_m^Tb_1 \\ \vdots & \ddots & \vdots \\ a_1^Tb_m & \cdots & a_m^Tb_m \end{pmatrix}$$

Since $a_i^Tb_i \in \mathbb{R}$, $i=1,\dots,n$

$$\mathrm{tr}(A^TB) = \sum_{i=1}^m a_i^Tb_i$$

$$<\mathrm{vec}(A), \mathrm{vec}(B)> = (a_1^T \cdots a_m^T)\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

$$= a_1^Tb_1 + \cdots + a_m^Tb_m = \sum_{i=1}^m a_i^Tb_i = \mathrm{tr}(A^TB) = \langle A,B\rangle_F$$

(b) Suppose $A,B,C \in \mathrm{Mat}_{n\times m}$, $\lambda,\mu \in \mathbb{R}$ are arbitrary

bilinearity:

$<A, \lambda B+\mu C>_F$	$<\lambda A+\mu B, C>_F$
$= \mathrm{tr}(A^T(\lambda B+\mu C))$	$= \mathrm{tr}((\lambda A+\mu B)^TC)$
$= \mathrm{tr}(\lambda A^TB + \mu A^TC)$	$= \mathrm{tr}((\lambda A^T+\mu B^T)C)$
$= \lambda \mathrm{tr}(A^TB) + \mu \mathrm{tr}(A^TC)$	$= \mathrm{tr}(\lambda A^TC + \mu B^TC)$

$$= tr(\lambda A^T B + \mu A^T C)$$
$$= \lambda tr(A^T B) + \mu tr(A^T C)$$
$$= \lambda \langle A, B\rangle_F + \mu \langle A, C\rangle_F$$

$$= tr((\lambda A^T + \mu B^T) C)$$
$$= tr(\lambda A^T C + \mu B^T C)$$
$$= \lambda tr(A^T C) + \mu tr(B^T C)$$
$$= \lambda \langle A, C\rangle_F + \mu \langle B, C\rangle_F$$

Commutative property: $\langle A, B\rangle_F = tr(A^T B)$

$$= tr((A^T B)^T) = tr(B^T A) = \langle B, A\rangle_F$$

0: $\langle A, A\rangle = 0 \Leftrightarrow tr(A^T A) = 0$

$$A := (a_{ij})_{n\times m} \quad A^T := (a^T_{kl})_{m\times n}$$

we have $a_{ij} = a^T_{ji}$ $\quad i = 1, \dots, n$, $j = 1, \dots, m$

$$A^T A = \left(\sum_{i=1}^{n} a^T_{ki} a_{ij}\right)$$
$$= \left(\sum_{i=1}^{n} a_{ik} a_{ij}\right) \qquad k, j = 1, \dots, m$$

$$tr(A^T A) = \sum_{j=1}^{m}\sum_{i=1}^{n} a_{ij} a_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} a_{ij}^2 = 0$$

$$\overset{a_{ij}^2 \geq 0}{\Longleftrightarrow} a_{ij} = 0, \ i = 1, \dots, n, \ j = 1, \dots, m \Leftrightarrow A = 0$$

(c) Since B is orthogonal

$$B^T = B^{-1}, \ BB^T = B^T B = I$$

$$\|BA\|_F = \sqrt{\langle BA, BA\rangle_F} = \sqrt{tr((BA)^T BA)}$$
$$= \sqrt{tr(A^T B^T BA)} = \sqrt{tr(A^T A)} = \|A\|_F$$

$$\|AB\|_F = \sqrt{tr(B^T A^T A B)} = \sqrt{tr(BB^T A^T A)} = \sqrt{tr(A^T A)} = \|A\|_F = \|BA\|_F$$

Exercise 9 (5 points)

(a) (2 points) Give a formula for the distance between two points $(a_1, a_2), (b_1, b_2) \in \mathbb{R}^2$ according to the river jungle metric.

a b

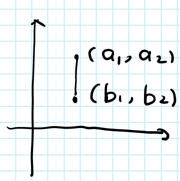
[Hint: There should be a case distinction, depending on whether the two points lie on the same vertical line or not.]

(b) (3 points) Show that the river jungle metric is not the metric associated to any norm.

[Hint: Metrics coming from norms are translation invariant, i.e. they satisfy $d(x, y) = d(x+z, y+z)$.]

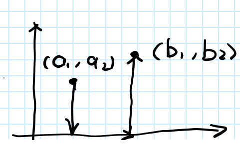
(a) if $a_1 = b_1$

$$d(a, b) = |a_2 - b_2|$$

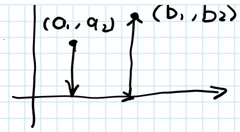


if $a_1 \neq b_1$

$$d(a, b) = |a_1 - b_1| + |a_2| + |b_2|$$



$d(a,b) = |a_1 - b_1| + |a_2| + |b_2|$



(b) It is sufficient to show a special case where

$d(x,y) \neq d(x+z, y+z)$

Let $x = (0,1)$, $y = (1,1)$, $z = (0,1)$

$x+z = (0,2)$, $y+z = (1,2)$

$d(x,y) = |0-1| + |1| + |1| = 3$

$d(x+z, y+z) = |0-1| + |2| + |2| = 5 \neq d(x,y)$

Therefore the metric is translation variant. It is not associated with a norm.

Exercise 10 (9 points)

Consider the following three points in $\mathbb{R}^2$:

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \sqrt{3} \\ 4 \end{pmatrix}, \begin{pmatrix} -\sqrt{3} \\ 1 \end{pmatrix}$$

Perform PCA to find the 1-dimensional subspace of $\mathbb{R}^2$ in which the projections of the above points are spread out the most.

[We didn't get to cover this in the Friday lecture - we will do it on Tuesday. After the Tuesday lecture, this exercise should be very quick to do. You task is to perform the algorithm on page 52 of the manuscript. See also Example 1.6.4 of the manuscript. If you want to know before Tuesday where this algorithm comes from, watch the last 20 minutes of lecture 6 from Winter term 2020/21, or minutes 20:00-32:00 of lecture 7, or read the first answer at this forum post, or look at any other book, video or blog post explaining PCA.]

[Note: In the solution there will be a $\sqrt{3}$ floating around (sorry!). Do not approximate that by decimal numbers, but rather calculate with it as a formal expression whose square is 3, e.g. as in $(2+\sqrt{3})(3-4\sqrt{3}) = 2 \cdot 3 + \sqrt{3} \cdot 3 + 2 \cdot (-4\sqrt{3}) + \sqrt{3} \cdot (-4\sqrt{3}) = 6 - 5\sqrt{3} - 12 = -6 - 5\sqrt{3}$]

mean vector: $\bar{x} = \frac{1}{3}\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} \sqrt{3} \\ 4 \end{pmatrix} + \begin{pmatrix} -\sqrt{3} \\ 1 \end{pmatrix}\right)$

$= \begin{pmatrix} 0 \\ 2 \end{pmatrix}$

data centering: $\begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

$\begin{pmatrix} \sqrt{3} \\ 4 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} \sqrt{3} \\ 2 \end{pmatrix}$

$$\begin{pmatrix} \sqrt{3} \\ 4 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} \sqrt{3} \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} -\sqrt{3} \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -\sqrt{3} \\ -1 \end{pmatrix}$$

data matrix : $D := \begin{pmatrix} 0 & \sqrt{3} & -\sqrt{3} \\ -1 & 2 & -1 \end{pmatrix}$

covariance matrix : $C := \frac{1}{3-1} D \cdot D^T$

$$= \frac{1}{2} \begin{pmatrix} 0 & \sqrt{3} & -\sqrt{3} \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ \sqrt{3} & 2 \\ -\sqrt{3} & -1 \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} 6 & 3\sqrt{3} \\ 3\sqrt{3} & 6 \end{pmatrix}$$

characteristic equation : $\chi_{2C}(\lambda) = \det \begin{pmatrix} 6-\lambda & 3\sqrt{3} \\ 3\sqrt{3} & 6-\lambda \end{pmatrix}$

$$= \lambda^2 - 12\lambda + 36 - 27$$

$3 \lfloor 108$
36

$$= \lambda^2 - 12\lambda + 9 = 0$$

$$\lambda_{1,2} = \frac{1}{2} \cdot \frac{12 \pm \sqrt{144-36}}{2} = \frac{12 \pm \sqrt{108}}{4} = \frac{12 \pm 6\sqrt{3}}{4} = 3 \pm \frac{3\sqrt{3}}{2}$$

we choose $\lambda = 3 + \frac{3\sqrt{3}}{2}$ since it's the biggest eigenvalue.

find the eigenvector : $\begin{pmatrix} -\frac{3\sqrt{3}}{2} & \frac{3\sqrt{3}}{2} \\ \frac{3\sqrt{3}}{2} & -\frac{3\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$

let $x = 1$, so $y = 1$

So the 1-dimensional subspace where the points' projection spread out the most is $\left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle$