Aalto University

Linear algebra

Exercise sheet 4 / Model solutions

1. Let $\boldsymbol{a} \in \mathbb{R}^n$ and $A = \boldsymbol{a}\boldsymbol{a}^T$.

   (a) Give a geometrical interpretation to the mapping $\boldsymbol{x} \mapsto A\boldsymbol{x}$. Fix $n = 2$, $\boldsymbol{a} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$ and draw the set

   $$\{A\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^2, \|\boldsymbol{x}\|_2 = 1\}.$$

   (b) Show that $\|A\|_2 = \|\boldsymbol{a}\|_2^2$.

   (c) Fix $n = 2$, $\boldsymbol{a} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. Calculate $\|A\|_2$, $\|A\|_F$ and $\|A\|_{max}$.

Solution.

(a) By direct calculation: $A\boldsymbol{x} = \boldsymbol{a}\boldsymbol{a}^T\boldsymbol{x} = \|\boldsymbol{a}\|_2^2 \frac{\boldsymbol{a}^T\boldsymbol{x}}{\|\boldsymbol{a}\|_2^2}\boldsymbol{a}$. Hence, the mapping $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is the projection of $\boldsymbol{x}$ to the direction of $\boldsymbol{a}$ scaled by $\|\boldsymbol{a}\|_2^2$. For the second part, denote

$$\boldsymbol{a}_0 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{a}_0^\perp = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Any $\boldsymbol{x} \in \mathbb{R}^2$ can be written as $\boldsymbol{x} = t\boldsymbol{a}_0 + s\boldsymbol{a}_0^\perp$ for some $t, s \in \mathbb{R}$. By direct calculation, it holds that $\|\boldsymbol{x}\|_2^2 = t^2 + s^2$. The condition $\|\boldsymbol{x}\|_2 = 1$ implies that $t \in [-1, 1]$. Thus,

$$\{A\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^2, \|\boldsymbol{x}\|_2 = 1\} = \{\|\boldsymbol{a}\|_2^2 t\boldsymbol{a}_0 \mid t \in [-1, 1]\}.$$

(b) By definition:

$$\|A\|_2 = \max_{\boldsymbol{x} \in \mathbb{R}^n} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \max_{\boldsymbol{x} \in \mathbb{R}^n} \frac{\|\boldsymbol{a}\boldsymbol{a}^T\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}.$$

Noticing that $\boldsymbol{a}^T\boldsymbol{x}$ is a scalar and using the Cauchy–Schwarz inequality gives

$$\|A\|_2 = \max_{\boldsymbol{x} \in \mathbb{R}^n} \frac{|\boldsymbol{a}^T\boldsymbol{x}|\|\boldsymbol{a}\|_2}{\|\boldsymbol{x}\|_2} \leq \|\boldsymbol{a}\|_2^2.$$

Choosing $\boldsymbol{x} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$ gives

$$\|A\|_2 \geq \frac{|\boldsymbol{a}^T\boldsymbol{a}|\|\boldsymbol{a}\|_2}{\|\boldsymbol{a}\|_2} = \|\boldsymbol{a}\|_2^2.$$

Hence, $\|A\|_2 = \|\boldsymbol{a}\|_2^2$.

(c) Using item (b) gives $\|A\|_2 = (1^2 + 1^2) = 2$. By defintion, $\|A\|_F = \left(\sum_{ij} a_{ij}^2\right)^{1/2} = 2$ and $\|A\|_{max} = \max_{ij} |a_{ij}| = 1$.

2. Let $A, B, X \in \mathbb{R}^{n\times n}$, $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ and $\|\cdot\|$ some vector norm (same notation is used for the corresponding operator norm). Show that:

   (a) $\|\boldsymbol{a}+\boldsymbol{b}\| \geq \big|\|\boldsymbol{a}\| - \|\boldsymbol{b}\|\big|$, (so-called "reverse triangle inequality")

   (b) $\|A+B\| \geq \big|\|A\| - \|B\|\big|$,

   (c) The matrix $I - X$ is invertible, if $\|X\| < 1$.

   Hint: In (c) assume that there exists $0 \neq \boldsymbol{x} \in N(I - X)$ and argue by contradiction.

   Solution.

   (a) Applying the triangle inequality to the vectors $\boldsymbol{s} = \boldsymbol{a}+\boldsymbol{b}$ and $\boldsymbol{m} = -\boldsymbol{b}$, we have

   $$\|\boldsymbol{a}\| = \|\boldsymbol{s}+\boldsymbol{m}\| \leq \|\boldsymbol{s}\| + \|\boldsymbol{m}\| = \|\boldsymbol{a}+\boldsymbol{b}\| + \|\boldsymbol{b}\|.$$

   This shows that $\|\boldsymbol{a}+\boldsymbol{b}\| \geq \|\boldsymbol{a}\| - \|\boldsymbol{b}\|$. A similar trick, applied to $\boldsymbol{s}$ and $\boldsymbol{d} = -\boldsymbol{a}$, yields $\|\boldsymbol{a}+\boldsymbol{b}\| \geq \|\boldsymbol{b}\| - \|\boldsymbol{a}\|$. Hence,

   $$|\|\boldsymbol{a}\| - \|\boldsymbol{b}\|| = \max\big\{\|\boldsymbol{a}\| - \|\boldsymbol{b}\|, \|\boldsymbol{b}\| - \|\boldsymbol{a}\|\big\} \leq \|\boldsymbol{a}+\boldsymbol{b}\|.$$

   (b) By the triangle inequality,

   $$\|A\| = \|A+B-B\| \leq \|A+B\| + \|B\|$$

   and

   $$\|B\| = \|B+A-A| \leq \|A+B\| + \|A\|$$

   (remember that for any matrix X we have $\|-X\| = \|X\|$). The statement follows by an argument similar to item (a).

   (c) For any induced norm $\|\cdot\|$ and using items (a–b),

   $$\frac{\|(I-X)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \geq 1 - \frac{\|X\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \geq 1 - \|X\| > 0;$$

   in the penultimate step we have used the fact that

   $$\|X\| \geq \frac{\|X\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

   which is true because $\|X\|$ is the maximum possible value of that ratio. This implies that, for any nonzero $\boldsymbol{x}$, $(I-X)\boldsymbol{x}$ cannot be $\mathbf{0}$, since it has positive norm. This in turn implies that $I - X$ is invertible under the given assumptions.

3. Let $\boldsymbol{x} \in \mathbb{R}^n$, $A, B \in \mathbb{R}^{n\times n}$ and $\|\cdot\|$ be some vector norm (same notation is used for the corresponding operator norm). Show that

$$\|A\boldsymbol{x}\| \le \|A\|\|\boldsymbol{x}\|, \qquad \|AB\| \le \|A\|\|B\| \qquad \text{and} \qquad \|I\| = 1. \tag{1}$$

Show by counterexample that $\|\cdot\|_{\max}$ and $\|\cdot\|_F$ are not operator norms.

Hint: Find matrices $A, B \in \mathbb{R}^{2\times 2}$ such that some of the properties in (1) do not hold. Note that $\|I\|_{\max} = 1$. In addition, the Frobenius norm satisfies the second inequality in (1): let $A^T = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_n]$ and $B = [\boldsymbol{b}_1, \dots, \boldsymbol{b}_n]$, so that

$$\begin{aligned}\|AB\|_F^2 &= \sum_{i,j=1}^n \left(\boldsymbol{a}_i^T \boldsymbol{b}_j\right)^2 \le \sum_{i,j=1}^n \|\boldsymbol{a}_i\|_2^2 \|\boldsymbol{b}_j\|_2^2 = \sum_{i=1}^n \|\boldsymbol{a}_i\|_2^2 \sum_{j=1}^n \|\boldsymbol{b}_j\|_2^2 \\ &= \sum_{i,k=1}^n |a_{ki}|^2 \sum_{j,l=1}^n |b_{lj}|^2 = \|A\|_F^2 \|B\|_F^2\end{aligned}$$

where the Cauchy–Schwarz inequality was used.

Solution. For an operator norm it holds that

$$\|A\| = \sup_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|.$$

From this follows that, for all $\boldsymbol{x}$, we have $\|A\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\| \le \|A\|$, and by one of the defining properties of norms we get

$$\|A\boldsymbol{x}\| \le \|A\|\|\boldsymbol{x}\|.$$

By using the previous inequality we get

$$\begin{aligned}\|AB\| &= \sup_{\|\boldsymbol{x}\|=1} \|AB\boldsymbol{x}\| \\ &\le \sup_{\|\boldsymbol{x}\|=1} \|A\|\|B\boldsymbol{x}\| \\ &= \|A\| \sup \|B\boldsymbol{x}\| \\ &= \|A\|\|B\| \\ &\le \|A\|\|B\|\end{aligned}$$

The Frobenius norm is defined by

$$\|A\|_F = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2}.$$

Let's assume now that $A = I_n$; from this follows that $\|A\|_F = \sqrt{n}$, and clearly

$$\sup_{\|\boldsymbol{x}\|=1} \|I\boldsymbol{x}\| = 1,$$

from which follows that the Frobenius norm is an operator norm only when $n = 1$.

The maximum norm is $\|A\|_{\max} = \max |a_{ij}|$. Define $A = (a_{ij})$ so that $a_{ij} = 1$ for all i, j. Now

$$\|AA\|_{\max} = n \geq \|A\|_{\max}\|A\|_{\max} = 1,$$

which goes against the property of operator norms shown above.

4. Let $A \in \mathbb{R}^{m\times n}$.

(a) Show that

$$N(A^T A) = N(A).$$

That is, the null space of A can be determined by computing eigenvectors corresponding to the zero eigenvalue for $A^T A$.

(b) Let

$$A = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix},$$

where $\epsilon \in \mathbb{R}$. Compute eigenvalues for the matrix $A^T A$ by hand. For which values of ϵ the null-space of A is trivial?

(c) Determine $N(A)$ using matlab, when $\epsilon = 1, 10^{-6}$ and 10^{-9}. Compute eigenvalues and corresponding eigenvectors for $A^T A$ numerically. Does Matlab agree with you on the dimension of $N(A)$ for each value of ϵ? Include the script that you used and the computed eigenvalues and vectors to your solution.

Hints: (a) The inclusion $N(A) \subset N(A^T A)$ can be easily proven. To prove that $N(A^T A) \subset N(A)$, multiply $A^T A$ from both sides with $\boldsymbol{x}$ and interpret the result as $\|A\boldsymbol{x}\|_2^2$. (c) Try out the commands `help eig` and `format long` in Matlab.

Solution.

(a) *Method 1*: The elements $\boldsymbol{x}$ in the null space of the matrix $A^T A$ satisfy

$$\begin{aligned} A^T A\boldsymbol{x} &= \mathbf{0} \qquad \text{multiply on the left by } \boldsymbol{x}^T \\ \boldsymbol{x}^T A^T A\boldsymbol{x} &= 0 \\ \Leftrightarrow (A\boldsymbol{x})^T (A\boldsymbol{x}) &= 0 \\ \Leftrightarrow \|A\boldsymbol{x}\|^2 &= 0 \\ \Leftrightarrow A\boldsymbol{x} &= \mathbf{0} \end{aligned}$$

meaning that those elements $\boldsymbol{x}$ belong also to the null space of A. On the other hand, if we take $\boldsymbol{x} \in N(A)$, then $A\boldsymbol{x} = 0$, which implies that $A^T A\boldsymbol{x} = A^T\mathbf{0} = \mathbf{0}$, meaning that surely also all elements of $N(A)$ belong to the null space of $A^T A$.

Method 2: We have

$$\begin{aligned}
&\boldsymbol{x} \in N(A^T A) \\
\Leftrightarrow\ & A^T A\boldsymbol{x} = A^T(A\boldsymbol{x}) = \mathbf{0} \\
\Leftrightarrow\ & A\boldsymbol{x} \in N(A^T) \qquad \big|\text{notice that } A\boldsymbol{x} \in R(A), \text{ too.}
\end{aligned}$$

On the other hand, recall that with respect to the Euclidean inner product we always have $R(A) \perp N(A^T)$, as we saw in a previous homework assignment. For this reason, we get $R(A) \cap N(A^T) = \{\mathbf{0}\}$, and here's why: if $\boldsymbol{x} \in R(A) \cap N(A^T)$, then $\boldsymbol{x}^T\boldsymbol{x} = \|\boldsymbol{x}\|_2^2 = 0$, which is equivalent to $\boldsymbol{x} = \mathbf{0}$ by definition of norm, and conversely we have of course $\mathbf{0} \in R(A) \cap N(A^T)$. We can then continue the chain of equivalences above by

$$A\boldsymbol{x} \in R(A) \cap N(A^T) \ \Leftrightarrow\ A\boldsymbol{x} = \mathbf{0} \ \Leftrightarrow\ \boldsymbol{x} \in N(A).$$

(b) Let us compute the eigenvalues λ:

$$\begin{aligned}
\left|A^T A - \lambda I\right| &= 0 \\
\left|\begin{bmatrix} 1 & \epsilon & 0 \\ 1 & 0 & \epsilon \end{bmatrix}\begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} - \lambda I\right| &= 0 \\
\left|\begin{bmatrix} 1+\epsilon^2 & 1 \\ 1 & 1+\epsilon^2 \end{bmatrix} - \lambda I\right| &= 0 \\
\begin{vmatrix} 1+\epsilon^2-\lambda & 1 \\ 1 & 1+\epsilon^2-\lambda \end{vmatrix} &= 0 \\
(1+\epsilon^2-\lambda)^2 - 1 &= 0 \\
1+\epsilon^2-\lambda &= \pm 1 \\
\lambda &= \begin{cases} \epsilon^2 \\ \epsilon^2 + 2 \end{cases}
\end{aligned}$$

The matrix has non-trivial null space if and only if some of its eigenvalues is 0, and let us see when this happens:

$$\begin{aligned}
\lambda_1 = \epsilon^2 = 0 \ &\Leftrightarrow\ \epsilon = 0 \\
\lambda_2 = \epsilon^2 + 2 = 0 \ &\Leftrightarrow\ \epsilon = \pm\sqrt{-2}
\end{aligned}$$

Because of the assumption $\epsilon \in \mathbb{R}$, the only possibility in which the matrix has non-trivial null space is when $\epsilon = 0$.

(c) The Matlab code

```
format long % shows more decimals
for eps = [1, 10^(-6), 10^(-9)]
   M=[1+eps^2, 1;1, 1+eps^2]; % matrix A^T * A
   [V,D] = eig(M) % V = 2x2 matrix, with column eigenvectors,
                  % D = diagonal matrix, with eigenvalues on diagonal
end
```

prints the eigenvectors and eigenvalues for each value of ϵ.

When $\epsilon = 1$ the eigenvalues are 1 and 3, and the eigenvectors are $[-0.7071\ 0.7071]^T$ and $[0.7071\ 0.7071]^T$. (A bit more clearly, for $\lambda = 1$ we get the eigenvectors $[v_1\ v_2]$ such that $v_1 = -v_2$, and for $\lambda = 3$ we have $v_1 = v_2$.) Since both eigenvectors differ from zero, $N(A) = \{\mathbf{0}\}$.