Prof. Dr. Laurenz Wiskott
Institut für Neuroinformatik
Ruhr-Universität Bochum
Universitätsstrasse 150
D-44780 Bochum, Germany

room: NB 3/29
phone: +49 234 32-27997
laurenz.wiskott@ini.rub.de
March 5, 2024

Linear Algebra (L4)

1 Trace and determinant of a symmetric matrix

1. Show that the trace of a symmetric matrix equals the sum of its eigenvalues.

   Solution: Let $\boldsymbol{M}$ be a symmetric square matrix. Since $\boldsymbol{M}$ is symmetric it can always be decomposed into a product of its diagonal eigenvalue matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_N)$ and its orthogonal eigenvector matrix $\boldsymbol{U}$, i.e. $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$. With this we find

   $$\mathrm{tr}(\boldsymbol{M}) = \mathrm{tr}(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T) = \mathrm{tr}(\underbrace{\boldsymbol{U}^T\boldsymbol{U}}_{\mathbf{1}}\boldsymbol{\Lambda}) = \mathrm{tr}(\boldsymbol{\Lambda}) = \sum_n \lambda_n \,, \quad (1)$$

   which proves the statement above.

   Interestingly, this proof also works for non-symmetric matrices that can be decomposed like $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, but then, e.g., with complex eigenvalues.

2. Show that the determinant of a symmetric matrix equals the product of its eigenvalues.

   Hint: For two square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we have $|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{A}| \cdot |\boldsymbol{B}|$.

   Solution: Similar to the proof above we can show

   $$|\boldsymbol{M}| = |\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T| = \underbrace{|\boldsymbol{U}|}_{=\pm 1} \cdot |\boldsymbol{\Lambda}| \cdot \underbrace{|\boldsymbol{U}^T|}_{=\pm 1} = |\boldsymbol{\Lambda}| = \prod_n \lambda_n \,. \quad (2)$$

   This proof, too, also works for non-symmetric matrices that can be decomposed like $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, but then, e.g., with complex eigenvalues.

   Extra question: To what extent does this result generalize to real valued rotation matrices (without a flip), i.e. matrices $\boldsymbol{M}$ with $\boldsymbol{M}^T = \boldsymbol{M}^{-1}$ and $|\boldsymbol{M}| = +1$?

   Extra question: What happens if you combine a rotation and a flip? What can you say about the eigenvalues? Does the rule above still hold?

2 Matrices as transformations

Describe with words the transformations realized by the following matrices. Estimate the values of the corresponding determinants without actually calculating them, only based on the intuitive understanding of the transformations.

(a)

$$\boldsymbol{M}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -2 \end{pmatrix} \tag{1}$$

Solution: This matrix stretches vectors along the three axes by the factors 1, 3, and -2, the latter meaning that the vectors along the third axis get flipped besides being stretched by 2.

The absolute value of the determinant of the matrix is 6 because the volume increases by the product of the stretching factors and its sign is negative because the vectors are flipped along one axis.

(b)

$$\boldsymbol{M}_2 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

Solution: Let us indicate the three axes by x, y, and z. The z-component does not change at all and for $z = 0$ the other two values are also preserved. But as z increases a value of $1z$ is added to x and a value of $2z$ is subtracted from y. This results in a shearing with an angle of 45° along the x-axis and a somewhat stronger shearing along the negative y-axis.

Since a shearing does not change the volume, the determinant is 1.

(c)

$$\boldsymbol{M}_3 = \begin{pmatrix} 2\cos(\phi) & 0 & 2\sin(\phi) \\ 0 & 1 & 0 \\ -\sin(\phi) & 0 & \cos(\phi) \end{pmatrix} \tag{3}$$

Solution: In this case it is useful to split the matrix into two terms, such as

$$\boldsymbol{M}_3 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\phi) & 0 & \sin(\phi) \\ 0 & 1 & 0 \\ -\sin(\phi) & 0 & \cos(\phi) \end{pmatrix} . \tag{4}$$

Interpreted from right to left this means that first we rotate the vectors by ϕ within the plane spanned by the first and third axis, and then they get stretched along the x-axis by a factor of 2.

The value of the determinant of the matrix is 2, because the rotation in the beginning does not change the volume, only the stretching by factor 2 doubles the volume.

3 Matrices with certain properties

Let $\mathcal{A}$ be a set of 3×3 matrices A with one of the following conditions. For each set

(i) Give a non-trivial example element for each of the sets.

(ii) Determine its 'dimensionality', i.e. the number of degrees of freedom (DOF, the number of variables that you can vary independently) the matrices have?

(iii) Tell whether it forms a vector space? Argue, if you think it is not a vector space.

1. $A^T = -A$.

   Solution: Matrices with $A^T = -A$ are antisymmetric. A specific example and one in a more general form are

$$\begin{pmatrix} 0 & 2 & -3 \\ -2 & 0 & 1 \\ 3 & -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}. \tag{1}$$

   From the latter one can easily see that this set has 3 degrees of freedom. It forms a vector space of dimensionality 3, e.g. the sum of two antisymmetric matrices or any scaled version of one is antisymmetric as well.

   Extra question: How does the number of degrees of freedom scale with the dimensionality of the matrix?

2. A has rank 1.

   Solution: Matrices of rank 1 have only one linearly independent column/row-vector, the others are linearly dependent. Thus, these matrices can be written as

$$(\boldsymbol{a}, \alpha_2 \boldsymbol{a}, \alpha_3 \boldsymbol{a}), \tag{2}$$

   with arbitrary non-zero vector $\boldsymbol{a}$ and coefficients α_i. Thus, this set has 5 degrees of freedom. It does not form a vector space, because the sum of two such matrices with different coefficients has rank 2.

   Extra question: How does the number of degrees of freedom scale with the dimensionality of the matrix?

3. $A^T = A^{-1}$.

   Solution: This is the set of orthogonal matrices. A specific example is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}. \tag{3}$$

   The first vector could be chosen freely, except for the normalization, which makes two degrees of freedom, because one degree of freedom is constrained by the normalization condition. The second vector must be normalized and orthogonal to the first one, which makes one degree of freedom. The third vector is entirely constrained by the first two except for its sign. Thus, orthogonal matrices have three degrees of freedom, which correspond to three rotation angles in 3D plus one binary degree that corresponds to a possible flip, i.e. change of handedness. This is not a vector space, because a scaled version is not normalized anymore and therfore not an orthogonal matrix.

Extra question: How does the number of degrees of freedom scale with the dimensionality of the matrix?

4 Eigenvectors of a symmetric matrix are orthogonal

Prove that the eigenvectors of a symmetric matrix are orthogonal, if their eigenvalues are different. Proceed as follows:

1. Let $\boldsymbol{A}$ be a symmetric N-dimensional matrix, i.e. $\boldsymbol{A} = \boldsymbol{A}^T$. Show first that $(\boldsymbol{v}, \boldsymbol{A}\boldsymbol{w}) = (\boldsymbol{A}\boldsymbol{v}, \boldsymbol{w})$ for any vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^N$, with $(\cdot,\cdot)$ indicating the Euclidean inner product.

   Solution:
$$(\boldsymbol{v}, \boldsymbol{A}\boldsymbol{w}) = \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{w} = \boldsymbol{v}^T\boldsymbol{A}^T\boldsymbol{w} = (\boldsymbol{A}\boldsymbol{v})^T\boldsymbol{w} = (\boldsymbol{A}\boldsymbol{v}, \boldsymbol{w})\,. \tag{1}$$

2. Let $\{\boldsymbol{a}_i\}$ be the eigenvectors of the matrix $\boldsymbol{A}$ with the eigenvalues λ_i. Show with the help of part one that $(\boldsymbol{a}_i, \boldsymbol{a}_j) = 0$ if $\lambda_i \neq \lambda_j$.

   Hint: $\lambda_i(\boldsymbol{a}_i, \boldsymbol{a}_j) = ...$

   Solution:
$$\lambda_i(\boldsymbol{a}_i, \boldsymbol{a}_j) = (\lambda_i\boldsymbol{a}_i, \boldsymbol{a}_j) = (\boldsymbol{A}\boldsymbol{a}_i, \boldsymbol{a}_j) \overset{(1)}{=} (\boldsymbol{a}_i, \boldsymbol{A}\boldsymbol{a}_j) = (\boldsymbol{a}_i, \lambda_j\boldsymbol{a}_j) = \lambda_j(\boldsymbol{a}_i, \boldsymbol{a}_j) \tag{2}$$
$$\Longrightarrow \qquad (\boldsymbol{a}_i, \boldsymbol{a}_j) = 0 \quad \text{if } \lambda_i \neq \lambda_j\,, \tag{3}$$

   which means that eigenvectors to different eigenvalues are orthogonal.

   Extra question: What can you say about the symmetry of the product of two symmetric matrics?

5 Matrices with given eigenvectors and -values

1. Construct a matrix $\boldsymbol{M}$ that has the following right-eigenvectors $\boldsymbol{r}_i$ (not normalized!) and eigenvalues λ_i.
$$\boldsymbol{r}_1 = (1,0)^T\,, \quad \lambda_1 = 1\,, \quad \boldsymbol{r}_2 = (1,1)^T\,, \quad \lambda_2 = 2\,. \tag{1}$$

   Verify your result.

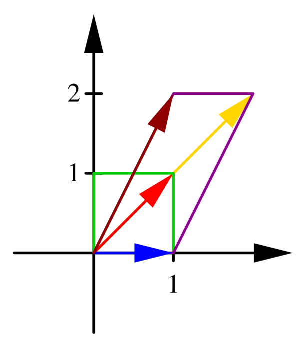
   Solution: There are different ways to solve this:

(a) One can simply set up a set of linear equations and solve that (Lea Kati, SS'22).

$$\begin{aligned}
& & \boldsymbol{M}\boldsymbol{r}_i = \lambda_i \boldsymbol{r}_i & \quad (2)\\
\Longleftrightarrow & & \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 0 \end{pmatrix} & \quad (3)\\
& \wedge & \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix} & \quad (4)\\
\Longleftrightarrow & & a = 1 & \quad (5)\\
& \wedge & c = 0 & \quad (6)\\
& \wedge & a + b = 2 & \quad (7)\\
& \wedge & c + d = 2 & \quad (8)\\
\Longleftrightarrow & & a = 1 & \quad (9)\\
& \wedge & c = 0 & \quad (10)\\
& \wedge & b = 1 & \quad (11)\\
& \wedge & d = 2 & \quad (12)\\
\Longleftrightarrow & & \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}. & \quad (13)
\end{aligned}$$

(b) One can do the same by reasoning as follows (Kai Brügge, SS'09). As $\boldsymbol{M}(1,0)^T = (1,0)^T$ the first column of $\boldsymbol{M}$ must be $(1,0)^T$, because the second column does not play any role in the product with $(1,0)^T$. As $\boldsymbol{M}(1,1)^T = (2,2)^T$ the second column of the matrix must then be $(1,2)^T$ to make the sum of the two columns equal to $(2,2)^T$.

(c) One can solve it geometrically by first drawing how the two right-eigenvectors (blue and red) transform (blue and yellow) and then inferring from that how the unit square (green) would transform (purple) (Simon Nagel, SS'22). The two edges of the transformed unit square originating at the origin (blue and brown) are the column vectors of matrix $\boldsymbol{M}$.



(d) One can work it out based on the lecture notes and first construct a set of left-eigenvectors $\tilde{\boldsymbol{l}}_i$ that are orthogonal to the other right-eigenvectors, i.e. a set with $\tilde{\boldsymbol{l}}_i^T \boldsymbol{r}_j = 0$ for $j \neq i$. It is easy to see that

$$\tilde{\boldsymbol{l}}_1 := (-1,1)^T\,, \quad \tilde{\boldsymbol{l}}_2 := (0,1)^T \qquad (14)$$

fulfill this requirement. Next we scale the left-eigenvectors such that the inner products

$\boldsymbol{l}_i^T \boldsymbol{r}_i$ equal the eigenvalues λ_i. Thus,

$$\boldsymbol{l}_1 \;:=\; \tilde{\boldsymbol{l}}_1 / \left(\tilde{\boldsymbol{l}}_1^T \boldsymbol{r}_1 \right) \cdot \lambda_1 = (-1, 1)^T / (-1) \cdot 1 = (1, -1)^T \,, \tag{15}$$

$$\boldsymbol{l}_2 \;:=\; \tilde{\boldsymbol{l}}_2 / \left(\tilde{\boldsymbol{l}}_2^T \boldsymbol{r}_2 \right) \cdot \lambda_2 = (0, 1)^T / 1 \cdot 2 = (0, 2)^T \,. \tag{16}$$

With these scaled left-eigenvectors, we can construct matrix $\boldsymbol{M}$.

$$\boldsymbol{M} := \boldsymbol{r}_1 \boldsymbol{l}_1^T + \boldsymbol{r}_2 \boldsymbol{l}_2^T \;=\; \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1, -1) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} (0, 2) \tag{17}$$

$$=\; \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 2 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} . \tag{18}$$

We verify that matrix $\boldsymbol{M}$ indeed has right-eigenvectors $\boldsymbol{r}_i$ with eigenvalues λ_i.

$$\boldsymbol{M}\boldsymbol{r}_1 \;=\; \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \lambda_1 \boldsymbol{r}_1 \,, \tag{19}$$

$$\boldsymbol{M}\boldsymbol{r}_2 \;=\; \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda_1 \boldsymbol{r}_2 \,. \tag{20}$$

2. Determine the left-eigenvectors and the corresponding eigenvalues of matrix $\boldsymbol{M}$. Verify your result.

   Solution: On the way of constructing the matrix we have constructed also the left-eigenvectors already. The eigenvalues should be the same as for the right-eigenvectors, because

$$\boldsymbol{l}_i^T \boldsymbol{M} = \boldsymbol{l}_i^T (\boldsymbol{r}_1 \boldsymbol{l}_1^T + \boldsymbol{r}_2 \boldsymbol{l}_2^T) = \underbrace{\boldsymbol{l}_i^T \boldsymbol{r}_1}_{=\delta_{i1}\lambda_1} \boldsymbol{l}_1^T + \underbrace{\boldsymbol{l}_i^T \boldsymbol{r}_2}_{=\delta_{i2}\lambda_2} \boldsymbol{l}_2^T = \lambda_i \boldsymbol{l}_i^T \,. \tag{21}$$

   For the concrete matrix above we verify that

$$\boldsymbol{l}_1^T \boldsymbol{M} \;=\; (1, -1) \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} = (1, -1) = 1(1, -1) = \lambda_1 \boldsymbol{l}_1^T \,, \tag{22}$$

$$\boldsymbol{l}_2^T \boldsymbol{M} \;=\; (0, 2) \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} = (0, 4) = 2(0, 2) = \lambda_2 \boldsymbol{l}_2^T \,. \tag{23}$$

   If one has chosen the more direct way in the first part, one has to work harder here and determine the left-eigenvectors and the corresponding eigenvalues the normal way.

Extra question: Symmetric matrices are often written in the form

$$\boldsymbol{M} \;=\; \sum_i \boldsymbol{u}_i \lambda_i \boldsymbol{u}_i^T \tag{24}$$

with orthogonal eigenvectors $\boldsymbol{u}_i$ and real eigenvalues λ_i. Does this also work for matrices considered in this exercise? If not, is there something similar possible? How should the vectors be normalized? How could you adapt the expression?

Extra question: What can you say about the eigenvalues of an upper triangular matrix?