1 Linear Algebra

1.1 Gram–Schmidt orthogonalisation

1. Given two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ in $\mathbb{R}^n$, show that

$$\mathbf{u}_1 = \mathbf{a}_1$$
$$\mathbf{u}_2 = \mathbf{a}_2 - \frac{\mathbf{u}_1^\top \mathbf{a}_2}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1$$

   are orthogonal to each other.

2. Show that any linear combination of (linearly independent) $\mathbf{a}_1$ and $\mathbf{a}_2$ can be written in terms of $\mathbf{u}_1$ and $\mathbf{u}_2$.

3. Show by induction that for any $k \leq n$ linearly independent vectors $\mathbf{a}_1, \dots, \mathbf{a}_k$, the vectors $\mathbf{u}_i$, $i = 1, \dots k$, are orthogonal, where

$$\mathbf{u}_i = \mathbf{a}_i - \sum_{j=1}^{i-1} \frac{\mathbf{u}_j^\top \mathbf{a}_i}{\mathbf{u}_j^\top \mathbf{u}_j}\mathbf{u}_j. \tag{1.1}$$

   The calculation of the vectors $\mathbf{u}_i$ is called Gram–Schmidt orthogonalisation.

4. Show by induction that any linear combination of (linear independent) $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$ can be written in terms of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$.

5. Consider the case where $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$ are linearly independent and $\mathbf{a}_{k+1}$ is a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$. Show that $\mathbf{u}_{k+1}$, computed according to (1.1), is zero.

1.2 Linear transforms

1. Assume two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are in $\mathbb{R}^2$. Together, they span a parallelogram. Use task 1.1 to show that the squared area S^2 of the parallelogram is given by

$$S^2 = (\mathbf{a}_2^T \mathbf{a}_2)(\mathbf{a}_1^T \mathbf{a}_1) - (\mathbf{a}_2^T \mathbf{a}_1)^2$$

2. Form the matrix $\mathbf{A} = (\mathbf{a}_1 \ \mathbf{a}_2)$ where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the first and second column vector, respectively. Show that

$$S^2 = (\det \mathbf{A})^2.$$

3. Consider the linear transform $\mathbf{y} = \mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is a 2×2 matrix. Denote the image of the rectangle $U_x = [x_1 \ x_1 + \triangle_1] \times [x_2 \ x_2 + \triangle_2]$ under the transform $\mathbf{A}$ by U_y. What is U_y? What is the area of U_y?

4. Give an intuitive explanation why we have equality in the change of variables formula

$$\int_{U_y} f(\mathbf{y}) d\mathbf{y} = \int_{U_x} f(\mathbf{A}\mathbf{x}) |\det \mathbf{A}| d\mathbf{x}.$$

   where $\mathbf{A}$ is such that U_x is an axis-aligned (hyper-) rectangle as in the previous question.

1.3 Eigenvalue decomposition

For a square matrix $\mathbf{A}$ of size $n \times n$, a vector $\mathbf{u}_i \neq 0$ which satisfies

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

is called a eigenvector of $\mathbf{A}$, and λ_i is the corresponding eigenvalue. For a matrix of size $n \times n$, there are n eigenvalues λ_i (which are not necessarily distinct).

1. Show that if $\mathbf{u}_1$ and $\mathbf{u}_2$ are eigenvectors with $\lambda_1 = \lambda_2$, then $\mathbf{u} = \alpha\mathbf{u}_1 + \beta\mathbf{u}_2$ is also an eigenvector with the same eigenvalue.

2. Assume that none of the eigenvalues of $\mathbf{A}$ is zero. Denote by $\mathbf{U}$ the matrix where the column vectors are linearly independent eigenvectors $\mathbf{u}_i$ of $\mathbf{A}$. Verify that (1.3) can be written in matrix form as $\mathbf{AU} = \mathbf{U\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues λ_i as diagonal elements.

3. Show that we can write, with $\mathbf{V}^T = \mathbf{U}^{-1}$,

$$\mathbf{A} = \mathbf{U\Lambda V}^\top, \qquad \mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{v}_i^T,$$

$$\mathbf{A}^{-1} = \mathbf{U\Lambda}^{-1}\mathbf{V}^\top, \qquad \mathbf{A}^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{v}_i^\top,$$

where $\mathbf{v}_i$ is the i-th column of $\mathbf{V}$.

1.4 Trace, determinants and eigenvalues

1. Use task 1.3 to show that $\mathrm{tr}(\mathbf{A}) = \sum_i A_{ii} = \sum_i \lambda_i$. (You can use $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$.)

2. Use task 1.3 to show that $\det \mathbf{A} = \prod_i \lambda_i$. (Use $\det \mathbf{A}^{-1} = 1/(\det \mathbf{A})$ and $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ for any $\mathbf{A}$ and $\mathbf{B}$.)

1.5 Eigenvalue decomposition for symmetric matrices

1. Assume that a matrix $\mathbf{A}$ is symmetric, i.e. $\mathbf{A}^\top = \mathbf{A}$. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be two eigenvectors of $\mathbf{A}$ with corresponding eigenvalues λ_1 and λ_2, with $\lambda_1 \neq \lambda_2$. Show that the two vectors are orthogonal to each other.

2. A symmetric matrix $\mathbf{A}$ is said to be positive definite if $\mathbf{v}^T\mathbf{A}\mathbf{v} > 0$ for all non-zero vectors $\mathbf{v}$. Show that positive definiteness implies that $\lambda_i > 0$, $i = 1, \ldots, M$. Show that, vice versa, $\lambda_i > 0$, $i = 1 \ldots M$ implies that the matrix $\mathbf{A}$ is positive definite. Conclude that a positive definite matrix is invertible.

1.6 Power method

We here analyse an algorithm called the "power method". The power method takes as input a positive definite symmetric matrix $\mathbf{\Sigma}$ and calculates the eigenvector that has the largest eigenvalue (the "first eigenvector"). For example, in case of principal component analysis, $\mathbf{\Sigma}$ is the covariance matrix of the observed data and the first eigenvector is the first principal component direction.

The power method consists in iterating the update equations

$$\mathbf{v}_{k+1} = \mathbf{\Sigma}\mathbf{w}_k, \qquad \mathbf{w}_{k+1} = \frac{\mathbf{v}_{k+1}}{||\mathbf{v}_{k+1}||_2},$$

where $||\mathbf{v}_{k+1}||_2$ denotes the Euclidean norm.

1. Let $\mathbf{U}$ the matrix with the (orthonormal) eigenvectors $\mathbf{u}_i$ of $\mathbf{\Sigma}$ as columns. What is the eigenvalue decomposition of the covariance matrix $\mathbf{\Sigma}$?

2. Let $\tilde{\mathbf{v}}_k = \mathbf{U}^T\mathbf{v}_k$ and $\tilde{\mathbf{w}}_k = \mathbf{U}^T\mathbf{w}_k$. Write the update equations of the power method in terms of $\tilde{\mathbf{v}}_k$ and $\tilde{\mathbf{w}}_k$. This means that we are making a change of basis to represent the vectors $\mathbf{w}_k$ and $\mathbf{v}_k$ in the basis given by the eigenvectors of $\mathbf{\Sigma}$.

3. Assume you start the iteration with $\tilde{\mathbf{w}}_0$. To which vector $\tilde{\mathbf{w}}^*$ does the iteration converge to?

4. Conclude that the power method finds the first eigenvector.