1 Linear Algebra

1.1 Gram–Schmidt orthogonalisation

1. Two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ of $\mathbb{R}^n$ are orthogonal if their inner product equals zero. Computing the inner product $\mathbf{u}_1^\top \mathbf{u}_2$ gives

$$\begin{aligned}\mathbf{u}_1^\top \mathbf{u}_2 &= \mathbf{u}_1^\top (\mathbf{a}_2 - \frac{\mathbf{u}_1^\top \mathbf{a}_2}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1) \\ &= \mathbf{u}_1^\top \mathbf{a}_2 - \frac{\mathbf{u}_1^\top \mathbf{a}_2}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1^\top \mathbf{u}_1 \\ &= \mathbf{u}_1^\top \mathbf{a}_2 - \mathbf{u}_1^\top \mathbf{a}_2 \\ &= 0.\end{aligned}$$

   Hence the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal.

   If $\mathbf{a}_2$ is a multiple of $\mathbf{a}_1$, the orthogonalisation procedure produces a zero vector for $\mathbf{u}_2$. To see this, let $\mathbf{a}_2 = \alpha \mathbf{a}_1$ for some real number α. We then obtain

$$\begin{aligned}\mathbf{u}_2 &= \mathbf{a}_2 - \frac{\mathbf{u}_1^\top \mathbf{a}_2}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1 \\ &= \alpha \mathbf{u}_1 - \frac{\alpha \mathbf{u}_1^\top \mathbf{u}_1}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1 \\ &= \alpha \mathbf{u}_1 - \alpha \mathbf{u}_1 \\ &= \mathbf{0}.\end{aligned}$$

2. Let $\mathbf{v}$ be a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$, i.e. $\mathbf{v} = \alpha \mathbf{a}_1 + \beta \mathbf{a}_2$ for some real numbers α and β. Expressing $\mathbf{u}_1$ and $\mathbf{u}_2$ in term of $\mathbf{a}_1$ and $\mathbf{a}_2$, we can write $\mathbf{v}$ as

$$\begin{aligned}\mathbf{v} &= \alpha \mathbf{a}_1 + \beta \mathbf{a}_2 \\ &= \alpha \mathbf{u}_1 + \beta(\mathbf{u}_2 + \frac{\mathbf{u}_1^\top \mathbf{a}_2}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1) \\ &= \alpha \mathbf{u}_1 + \beta \mathbf{u}_2 + \beta \frac{\mathbf{u}_1^\top \mathbf{a}_2}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1 \\ &= (\alpha + \beta \frac{\mathbf{u}_1^\top \mathbf{a}_2}{\mathbf{u}_1^\top \mathbf{u}_1})\mathbf{u}_1 + \beta \mathbf{u}_2,\end{aligned}$$

   Since $\alpha + \beta((\mathbf{u}_1^\top \mathbf{a}_2)/(\mathbf{u}_1^\top \mathbf{u}_1))$ and β are real numbers, we can write $\mathbf{v}$ as a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$. Overall, this means that any vector in the span of $\{\mathbf{a}_1, \mathbf{a}_2\}$ can be expressed in the orthogonal basis $\{\mathbf{u}_1, \mathbf{u}_2\}$.

3. We have shown above that the claim holds for two vectors. This is the base case for the proof by induction. Assume now that the claim holds for k vectors. The induction step in the proof by induction then consists of showing that the claim also holds for $k + 1$ vectors.

Assume that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are orthogonal vectors. The linear independence assumption ensures that none of the $\mathbf{u}_i$ is a zero vector. We then have for $\mathbf{u}_{k+1}$

$$\mathbf{u}_{k+1} = \mathbf{a}_{k+1} - \frac{\mathbf{u}_1^\top \mathbf{a}_{k+1}}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1 - \frac{\mathbf{u}_2^\top \mathbf{a}_{k+1}}{\mathbf{u}_2^\top \mathbf{u}_2}\mathbf{u}_2 - \ldots - \frac{\mathbf{u}_k^\top \mathbf{a}_{k+1}}{\mathbf{u}_k^\top \mathbf{u}_k}\mathbf{u}_k, \tag{1.1}$$

and for all $i = 1, 2, \ldots, k$

$$\mathbf{u}_i^\top \mathbf{u}_{k+1} = \mathbf{u}_i^\top \mathbf{a}_{k+1} - \frac{\mathbf{u}_1^\top \mathbf{a}_{k+1}}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_i^\top \mathbf{u}_1 - \ldots - \frac{\mathbf{u}_k^\top \mathbf{a}_{k+1}}{\mathbf{u}_k^\top \mathbf{u}_k}\mathbf{u}_i^\top \mathbf{u}_k. \tag{1.2}$$

By assumption $\mathbf{u}_i^\top \mathbf{u}_j = 0$ if $i \neq j$, so that

$$\begin{aligned}\mathbf{u}_i^\top \mathbf{u}_{k+1} &= \mathbf{u}_i^\top \mathbf{a}_{k+1} - 0 - \ldots - \frac{\mathbf{u}_i^\top \mathbf{a}_{k+1}}{\mathbf{u}_i^\top \mathbf{u}_i}\mathbf{u}_i^\top \mathbf{u}_i - 0 \ldots - 0\\ &= \mathbf{u}_i^\top \mathbf{a}_{k+1} - \mathbf{u}_i^\top \mathbf{a}_{k+1}\\ &= 0,\end{aligned}$$

which means that $\mathbf{u}_{k+1}$ is orthogonal to $\mathbf{u}_1, \ldots, \mathbf{u}_k$.

4. The base case of two vectors was proved above. Using induction, we assume that the claim holds for k vectors and we will prove that it then also holds for $k+1$ vectors: Let $\mathbf{v}$ be a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{k+1}$, i.e. $\mathbf{v} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \ldots + \alpha_k \mathbf{a}_k + \alpha_{k+1}\mathbf{a}_{k+1}$ for some real numbers $\alpha_1, \alpha_2, \ldots, \alpha_{k+1}$. Using the induction assumption, $\mathbf{v}$ can be written as

$$\mathbf{v} = \beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \ldots + \beta_k \mathbf{u}_k + \alpha_{k+1}\mathbf{a}_{k+1},$$

for some real numbers $\beta_1, \beta_2, \ldots, \beta_k$ Furthermore, using equation (1.1), $\mathbf{v}$ can be written as

$$\begin{aligned}\mathbf{v} = \beta_1 \mathbf{u}_1 + \ldots + \beta_k \mathbf{u}_k + \alpha_{k+1}\mathbf{u}_{k+1} + \alpha_{k+1}\frac{\mathbf{u}_1^\top \mathbf{a}_{k+1}}{\mathbf{u}_1^\top \mathbf{u}_1}\mathbf{u}_1\\ + \ldots + \alpha_{k+1}\frac{\mathbf{u}_k^\top \mathbf{a}_{k+1}}{\mathbf{u}_k^\top \mathbf{u}_k}\mathbf{u}_k.\end{aligned}$$

With $\gamma_i = \beta_i + \alpha_{k+1}(\mathbf{u}_i^\top \mathbf{a}_{k+1})/(\mathbf{u}_i^\top \mathbf{u}_i)$, $\mathbf{v}$ can thus be written as

$$\mathbf{v} = \gamma_1 \mathbf{u}_1 + \gamma_2 \mathbf{u}_2 + \ldots + \gamma_k \mathbf{u}_k + \alpha_{k+1}\mathbf{u}_{k+1},$$

which completes the proof. Overall, this means that the $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ form an orthogonal basis for $\mathrm{span}(\mathbf{a}_1, \ldots, \mathbf{a}_k)$, i.e. the set of all vectors that can be obtained by linearly combining the $\mathbf{a}_i$.

1.2 Linear transforms

1. Starting with (**??**), we have

$$\mathbf{u}_{k+1} = \mathbf{a}_{k+1} - \sum_{j=1}^{k} \frac{\mathbf{u}_j^\top \mathbf{a}_{k+1}}{\mathbf{u}_j^\top \mathbf{u}_j}\mathbf{u}_j.$$

By assumption, $\mathbf{a}_{k+1}$ is a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$. By the previous question, it can thus also be written as a linear combination of the $\mathbf{u}_1, \dots, \mathbf{u}_k$. This means that there are some β_i so that

$$\mathbf{a}_{k+1} = \sum_{i=1}^{k} \beta_i \mathbf{u}_i$$

holds. Inserting this expansion into the equation above gives

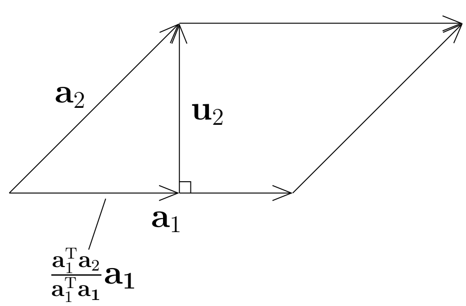
$$\begin{aligned}\mathbf{u}_{k+1} &= \sum_{i=1}^{k} \beta_i \mathbf{u}_i - \sum_{j=1}^{k} \sum_{i=1}^{k} \beta_i \frac{\mathbf{u}_j^\top \mathbf{u}_i}{\mathbf{u}_j^\top \mathbf{u}_j} \mathbf{u}_j \\ &= \sum_{i=1}^{k} \beta_i \mathbf{u}_i - \sum_{i=1}^{k} \beta_i \frac{\mathbf{u}_i^\top \mathbf{u}_i}{\mathbf{u}_i^\top \mathbf{u}_i} \mathbf{u}_i\end{aligned}$$

because $\mathbf{u}_j^\top \mathbf{u}_i = 0$ if $i \neq j$. We thus obtain the desired result:

$$\begin{aligned}\mathbf{u}_{k+1} &= \sum_{i=1}^{k} \beta_i \mathbf{u}_i - \sum_{i=1}^{k} \beta_i \mathbf{u}_i \\ &= 0\end{aligned}$$

This property of the Gram-Schmidt process in (**??**) can be used to check whether a list of vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d$ is linearly independent or not. If, for example, $\mathbf{u}_{k+1}$ is zero, $\mathbf{a}_{k+1}$ is a linear combination of the $\mathbf{a}_1, \dots, \mathbf{a}_k$. Moreover, the result can be used to extract a sublist of linearly independent vectors: We would remove $\mathbf{a}_{k+1}$ from the list and restart the procedure in (**??**) with $\mathbf{a}_{k+2}$ taking the place of $\mathbf{a}_{k+1}$. Continuing in this way constructs a list of linearly independent $\mathbf{a}_j$ and orthogonal $\mathbf{u}_j$, $j = 1, \dots, r$, where r is the number of linearly independent vectors among the $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d$.

2. Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be the vectors that span the parallelogram. From geometry we know that the area of parallelogram is base times height, which is equivalent to the length of the base vector times the length of the height vector. Denote this by $S^2 = ||\mathbf{a}_1||^2||\mathbf{u}_2||^2$, where is $\mathbf{a}_1$ is the base vector and $\mathbf{u}_2$ is the height vector which is orthogonal to the base vector. Using the Gram–Schmidt process for the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ in that order, we obtain the vector $\mathbf{u}_2$ as the second output.



Therefore $||\mathbf{u}_2||^2$ equals

$$\begin{aligned}
||\mathbf{u}_2||^2 &= \mathbf{u}_2^\top \mathbf{u}_2 \\
&= \left(\mathbf{a}_2 - \frac{\mathbf{a}_1^\top \mathbf{a}_2}{\mathbf{a}_1^\top \mathbf{a}_1}\mathbf{a}_1\right)^\top \left(\mathbf{a}_2 - \frac{\mathbf{a}_1^\top \mathbf{a}_2}{\mathbf{a}_1^\top \mathbf{a}_1}\mathbf{a}_1\right) \\
&= \mathbf{a}_2^\top \mathbf{a}_2 - \frac{(\mathbf{a}_1^\top \mathbf{a}_2)^2}{\mathbf{a}_1^\top \mathbf{a}_1} - \frac{(\mathbf{a}_1^\top \mathbf{a}_2)^2}{\mathbf{a}_1^\top \mathbf{a}_1} + \left(\frac{\mathbf{a}_1^\top \mathbf{a}_2}{\mathbf{a}_1^\top \mathbf{a}_1}\right)^2 \mathbf{a}_1^\top \mathbf{a}_1 \\
&= \mathbf{a}_2^\top \mathbf{a}_2 - \frac{(\mathbf{a}_1^\top \mathbf{a}_2)^2}{\mathbf{a}_1^\top \mathbf{a}_1}.
\end{aligned}$$

Thus, S^2 is:

$$\begin{aligned}
S^2 &= ||\mathbf{a}_1||^2 ||\mathbf{u}_2||^2 \\
&= (\mathbf{a}_1^\top \mathbf{a}_1)(\mathbf{u}_2^\top \mathbf{u}_2) \\
&= (\mathbf{a}_1^\top \mathbf{a}_1)\left(\mathbf{a}_2^\top \mathbf{a}_2 - \frac{(\mathbf{a}_1^\top \mathbf{a}_2)^2}{\mathbf{a}_1^\top \mathbf{a}_1}\right) \\
&= (\mathbf{a}_2^\top \mathbf{a}_2)(\mathbf{a}_1^\top \mathbf{a}_1) - (\mathbf{a}_1^\top \mathbf{a}_2)^2.
\end{aligned}$$

3. We form the matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{pmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

The determinant of $\mathbf{A}$ is $\det \mathbf{A} = a_{11}a_{22} - a_{12}a_{21}$. By multiplying out $(\mathbf{a}_2^\top \mathbf{a}_2)$, $(\mathbf{a}_1^\top \mathbf{a}_1)$ and $(\mathbf{a}_1^\top \mathbf{a}_2)^2$, we get

$$\begin{aligned}
\mathbf{a}_2^\top \mathbf{a}_2 &= a_{12}^2 + a_{22}^2 \\
\mathbf{a}_1^\top \mathbf{a}_1 &= a_{11}^2 + a_{21}^2 \\
(\mathbf{a}_1^\top \mathbf{a}_2)^2 &= (a_{11}a_{12} + a_{21}a_{22})^2 = a_{11}^2 a_{12}^2 + a_{21}^2 a_{22}^2 + 2a_{11}a_{12}a_{21}a_{22}.
\end{aligned}$$

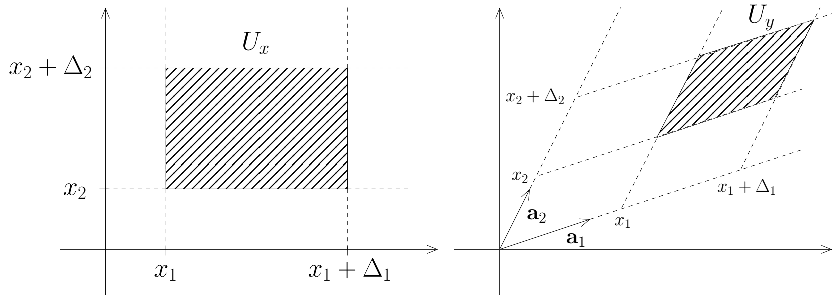
Therefore the area equals

$$\begin{aligned}
S^2 &= (a_{12}^2 + a_{22}^2)(a_{11}^2 + a_{21}^2) - (\mathbf{a}_1^\top \mathbf{a}_2)^2 \\
&= a_{12}^2 a_{11}^2 + a_{12}^2 a_{21}^2 + a_{22}^2 a_{11}^2 + a_{22}^2 a_{21}^2 \\
&\quad - (a_{12}^2 a_{11}^2 + a_{21}^2 a_{22}^2 + 2a_{11}a_{12}a_{21}a_{22}) \\
&= a_{12}^2 a_{21}^2 + a_{22}^2 a_{11}^2 - 2a_{11}a_{12}a_{21}a_{22} \\
&= (a_{11}a_{22} - a_{12}a_{21})^2,
\end{aligned}$$

which equals $(\det \mathbf{A})^2$

4. U_y is parallelogram that is spanned by the column vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ of $\mathbf{A}$, when $\mathbf{A} = (\mathbf{a}_1\ \mathbf{a}_2)$.

   A rectangle with the same area as U_x is spanned by vectors $(\Delta_1, 0)$ and $(0, \Delta_2)$. Under the linear transform $\mathbf{A}$ these spanning vectors become $\Delta_1 \mathbf{a}_1$ and $\Delta_2 \mathbf{a}_2$. Therefore a parallelogram with the same area as U_y is spanned by $\Delta_1 \mathbf{a}_1$ and $\Delta_2 \mathbf{a}_2$ as shown in the following figure.



From the previous question, the A_{U_y} of U_y equals the absolute value of the determinant of the matrix $(\Delta_1\mathbf{a}_1\ \Delta_2\mathbf{a}_2)$:

$$\begin{aligned} A_{U_y} &= |\det \begin{pmatrix} \Delta_1 a_{11} & \Delta_2 a_{12} \\ \Delta_1 a_{21} & \Delta_2 a_{22} \end{pmatrix}| \\ &= |\Delta_1\Delta_2 a_{11}a_{22} - \Delta_1\Delta_2 a_{12}a_{21}| \\ &= |\Delta_1\Delta_2(a_{11}a_{22} - a_{12}a_{21})| \\ &= \Delta_1\Delta_2 |\det \mathbf{A}| \end{aligned}$$

Therefore the area of U_y is the area of U_x times $|\det \mathbf{A}|$.

5. We can think that, loosely speaking, the two integrals are limits of the following two sums

$$\sum_{\mathbf{y}_i \in U_y} f(\mathbf{y}_i)\mathrm{vol}(\Delta_{\mathbf{y}_i}) \qquad \sum_{\mathbf{x}_i \in U_x} f(\mathbf{A}\mathbf{x}_i)|\det \mathbf{A}|\mathrm{vol}(\Delta_{\mathbf{x}_i})$$

where $\mathbf{x}_i = \mathbf{A}^{-1}\mathbf{y}_i$, which means that $\mathbf{x}$ and $\mathbf{y}$ are related by $\mathbf{y} = \mathbf{A}\mathbf{x}$. The set of function values $f(\mathbf{y}_i)$ and $f(\mathbf{A}\mathbf{x}_i)$ that enter the two sums are exactly the same. The volume $\mathrm{vol}(\Delta_{\mathbf{x}_i})$ of a small axis-aligned hypercube (in d dimensions) equals $\prod_{i=1}^{d} \Delta_i$. The image of this small axis-aligned hypercube under $\mathbf{A}$ is a parallelogram $\Delta_{\mathbf{y}_i}$ with volume $\mathrm{vol}(\Delta_{\mathbf{y}_i}) = |\det \mathbf{A}|\mathrm{vol}(\Delta_{\mathbf{x}_i})$. Hence

$$\sum_{\mathbf{y}_i \in U_y} f(\mathbf{y}_i)\mathrm{vol}(\Delta_{\mathbf{y}_i}) = \sum_{\mathbf{x}_i \in U_x} f(\mathbf{A}\mathbf{x}_i)|\det \mathbf{A}|\mathrm{vol}(\Delta_{\mathbf{x}_i}).$$

We must have the term $|\det \mathbf{A}|$ to compensate for the fact that the volume of U_x and U_y are not the same. For example, let $\mathbf{A}$ be a diagonal matrix $\mathrm{diag}(10, 100)$ so that U_x is much smaller than U_y. The determinant $\det \mathbf{A} = 1000$ then compensates for the fact that the $\mathbf{x}_i$ values are more condensed than the $\mathbf{y}_i$.

1.3 Eigenvalue decomposition

1. We compute

$$\begin{aligned} \mathbf{Au} &= \alpha\mathbf{A}\mathbf{u}_1 + \beta\mathbf{A}\mathbf{u}_2 \\ &= \alpha\lambda\mathbf{u}_1 + \beta\lambda\mathbf{u}_2 \\ &= \lambda(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) \\ &= \lambda\mathbf{u}, \end{aligned}$$

so $\mathbf{u}$ is an eigenvector of $\mathbf{A}$ with the same eigenvalue as $\mathbf{u}_1$ and $\mathbf{u}_2$.

2. By basic properties of matrix multiplication, we have

$$\mathbf{AU} = (\mathbf{Au}_1 \ \mathbf{Au}_2 \ \dots \ \mathbf{Au}_n)$$

With $\mathbf{Au}_i = \lambda_i \mathbf{u}_i$ for all $i = 1, 2, \dots, n$, we thus obtain

$$\begin{aligned}\mathbf{AU} &= (\lambda_1 \mathbf{u}_1 \ \lambda_2 \mathbf{u}_2 \ \dots \ \lambda_n \mathbf{u}_n) \\ &= \mathbf{U\Lambda}.\end{aligned}$$

3. (i) Since the columns of $\mathbf{U}$ are linearly independent, $\mathbf{U}$ is invertible. Because $\mathbf{AU} = \mathbf{U\Lambda}$, multiplying from the right with the inverse of $\mathbf{U}$ gives $\mathbf{A} = \mathbf{U\Lambda U}^{-1} = \mathbf{U}\Lambda\mathbf{V}^\top$.

   (ii) Denote by $\mathbf{u}^{[i]}$ the ith row of $\mathbf{U}$, $\mathbf{v}^{(j)}$ the jth column of $\mathbf{V}^\top$ and $\mathbf{v}^{[j]}$ the jth row of $\mathbf{V}$ and denote $\mathbf{B} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{v}_i^\top$. Let $\mathbf{e}^{[i]}$ be a row vector with 1 in the ith place and 0 elsewhere and $\mathbf{e}^{(j)}$ be a column vector with 1 in the jth place and 0 elsewhere. Notice that because $\mathbf{A} = \mathbf{U\Lambda V}^\top$, the element in the ith row and jth column is

$$\begin{aligned}A_{ij} &= \mathbf{u}^{[i]}\mathbf{\Lambda}\mathbf{v}^{(j)} \\ &= \mathbf{u}^{[i]}\mathbf{\Lambda}\mathbf{v}^{[j]\top} \\ &= \mathbf{u}^{[i]}\begin{pmatrix}\lambda_1 V_{j1} \\ \vdots \\ \lambda_n V_{jn}\end{pmatrix} \\ &= \sum_{k=1}^n \lambda_k V_{jk} U_{ik}.\end{aligned}$$

   On the other hand, for matrix $\mathbf{B}$ the element in the ith row and jth column is

$$\begin{aligned}B_{ij} &= \sum_{k=1}^n \lambda_k \mathbf{e}^{[i]}\mathbf{u}_k \mathbf{v}_k^\top \mathbf{e}^{(j)} \\ &= \sum_{k=1}^n \lambda_k U_{ik} V_{jk},\end{aligned}$$

   which is the same as A_{ij}. Therefore $\mathbf{A} = \mathbf{B}$.

   (iii) Since $\mathbf{\Lambda}$ is a diagonal matrix with no zeros as diagonal elements, it is invertible. We have thus

$$\begin{aligned}\mathbf{A}^{-1} &= (\mathbf{U\Lambda U}^{-1})^{-1} \\ &= (\mathbf{\Lambda U}^{-1})^{-1}\mathbf{U}^{-1} \\ &= \mathbf{U}\Lambda^{-1}\mathbf{U}^{-1} \\ &= \mathbf{U}\Lambda^{-1}\mathbf{V}^\top.\end{aligned}$$

   (iv) This follows from $\mathbf{A} = \mathbf{U\Lambda V}^\top = \sum_i \mathbf{u}_i \lambda_i \mathbf{v}_i^\top$, when λ_i is replaced with $1/\lambda_i$.

1.4 Trace, determinants and eigenvalues

1. Since $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ and $\mathbf{A} = \mathbf{U\Lambda U}^{-1}$

$$\begin{aligned}\mathrm{tr}(\mathbf{A}) &= \mathrm{tr}(\mathbf{U\Lambda U}^{-1}) \\ &= \mathrm{tr}(\mathbf{\Lambda U}^{-1}\mathbf{U}) \\ &= \mathrm{tr}(\Lambda) \\ &= \sum_i \lambda_i.\end{aligned}$$

2. We use the eigenvalue decomposition of $\mathbf{A}$ to obtain

$$\begin{aligned}\det(\mathbf{A}) &= \det(\mathbf{U\Lambda U}^{-1}) \\ &= \det(\mathbf{U})\det(\mathbf{\Lambda})\det(\mathbf{U}^{-1}) \\ &= \frac{\det(\mathbf{U})\det(\mathbf{\Lambda})}{\det(\mathbf{U})} \\ &= \det(\Lambda) \\ &= \prod_i \lambda_i,\end{aligned}$$

where, in the last line, we have used that the determinant of a diagonal matrix is the product of its diagonal elements.

1.5 Eigenvalue decomposition for symmetric matrices

1. Since $\mathbf{Au}_2 = \lambda_2\mathbf{u}_2$, we have

$$\mathbf{u}_1^\top\mathbf{Au}_2 = \lambda_2\mathbf{u}_1^\top\mathbf{u}_2.$$

Taking the transpose of $\mathbf{u}_1^\top\mathbf{Au}_2$ gives

$$\begin{aligned}(\mathbf{u}_1^\top\mathbf{Au}_2)^\top &= (\mathbf{Au}_2)^\top(\mathbf{u}_1^\top)^\top = \mathbf{u}_2^\top\mathbf{A}^\top\mathbf{u}_1 = \mathbf{u}_2^\top\mathbf{Au}_1 \\ &= \lambda_1\mathbf{u}_2^\top\mathbf{u}_1\end{aligned}$$

because $\mathbf{A}$ is symmetric and $\mathbf{Au}_1 = \lambda_1\mathbf{u}_1$. On the other hand, the same operation gives

$$(\mathbf{u}_1^\top\mathbf{Au}_2)^\top = (\lambda_2\mathbf{u}_1^\top\mathbf{u}_2)^\top = \lambda_2\mathbf{u}_2^\top\mathbf{u}_1$$

Therefore $\lambda_1\mathbf{u}_2^\top\mathbf{u}_1 = \lambda_2\mathbf{u}_2^\top\mathbf{u}_1$, which is equivalent to $\mathbf{u}_2^\top\mathbf{u}_1(\lambda_1 - \lambda_2) = 0$. Because $\lambda_1 \neq \lambda_2$, the only possibility is that $\mathbf{u}_2^\top\mathbf{u}_1 = 0$. Therefore $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal to each other.

The result implies that the eigenvectors of a symmetric matrix $\mathbf{A}$ with distinct eigenvalues λ_i forms an orthogonal basis. The result extends to the case where some eigenvalues are the same (not proven).

2. Assume that $\mathbf{v}^\top\mathbf{Av} > 0$ for all $\mathbf{v} \neq 0$. Since eigenvectors are not zero vectors, the assumption holds also for eigenvector $\mathbf{u}_k$ with corresponding eigenvalue λ_k. Now

$$\mathbf{u}_k^\top\mathbf{Au}_k = \mathbf{u}_k^\top\lambda_k\mathbf{u}_k = \lambda_k(\mathbf{u}_k^\top\mathbf{u}_k) = \lambda_k||\mathbf{u}_k|| > 0$$

and because $||\mathbf{u}_k|| > 0$, we obtain $\lambda_k > 0$.

Assume now that all the eigenvalues of $\mathbf{A}$, $\lambda_1, \lambda_2, \ldots, \lambda_n$, are positive and nonzero. We have shown above that there exists an orthogonal basis consisting of eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ and therefore every vector $\mathbf{v}$ can be written as a linear combination of those vectors (we have only shown it for the case of distinct eigenvalues but it holds more generally). Hence for a nonzero vector $\mathbf{v}$ and for some real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$, we have

$$\begin{aligned}
\mathbf{v}^\top \mathbf{A}\mathbf{v} &= (\alpha_1 \mathbf{u}_1 + + \ldots + \alpha_n \mathbf{u}_n)^\top \mathbf{A}(\alpha_1 \mathbf{u}_1 + \ldots + \alpha_n \mathbf{u}_n) \\
&= (\alpha_1 \mathbf{u}_1 + \ldots + \alpha_n \mathbf{u}_n)^\top (\alpha_1 \mathbf{A}\mathbf{u}_1 + \ldots + \alpha_n \mathbf{A}\mathbf{u}_n) \\
&= (\alpha_1 \mathbf{u}_1 + \ldots + \alpha_n \mathbf{u}_n)^\top (\alpha_1 \lambda_1 \mathbf{u}_1 + \ldots + \alpha_n \lambda_n \mathbf{u}_n) \\
&= \sum_{i,j} \alpha_i \mathbf{u}_i^\top \alpha_j \lambda_j \mathbf{u}_j \\
&= \sum_i \alpha_i \alpha_i \lambda_i \mathbf{u}_i^\top \mathbf{u}_i \\
&= \sum_i (\alpha_i)^2 ||\mathbf{u}_i||^2 \lambda_i,
\end{aligned}$$

where we have used that $\mathbf{u}_i^T \mathbf{u}_j = 0$ if $i \neq j$, due to orthogonality of the basis. Since $(\alpha_i)^2 > 0$, $||\mathbf{u}_i||^2 > 0$ and $\lambda_i > 0$ for all i, we find that $\mathbf{v}^\top \mathbf{A}\mathbf{v} > 0$.

Since every eigenvalue of $\mathbf{A}$ is nonzero, we can use Task 1.3 to conclude that inverse of $\mathbf{A}$ exists and equals $\sum_i 1/\lambda_i \mathbf{u}_i \mathbf{u}_i^\top$.

1.6 Power method

1. Since the columns of $\mathbf{U}$ are orthonormal (eigen)vectors, $\mathbf{U}$ is orthogonal, i.e. $\mathbf{U}^{-1} = \mathbf{U}^\top$. With Task 1.3 and Task 1.5, we obtain

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top,$$

where $\mathbf{\Lambda}$ is the diagonal matrix with eigenvalues λ_i of $\mathbf{\Sigma}$ as diagonal elements. Let the eigenvalues be ordered $\lambda_1 > \lambda_2 > \ldots > \lambda_n > 0$ (and, as additional assumption, all distinct).

2. With

$$\begin{aligned}
\mathbf{v}_{k+1} &= \mathbf{\Sigma}\mathbf{w}_k \\
&= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \mathbf{w}_k
\end{aligned}$$

we obtain

$$\mathbf{U}^\top \mathbf{v}_{k+1} = \mathbf{\Lambda}\mathbf{U}^\top \mathbf{w}_k.$$

Hence $\tilde{\mathbf{v}}_{k+1} = \mathbf{\Lambda}\tilde{\mathbf{w}}_k$. The norm of $\tilde{\mathbf{v}}_{k+1}$ is the same as the norm of $\mathbf{v}_{k+1}$:

$$\begin{aligned}||\tilde{\mathbf{v}}_{k+1}||_2 &= ||\mathbf{U}^\top \mathbf{v}_{k+1}||_2 \\ &= \sqrt{(\mathbf{U}^\top \mathbf{v}_{k+1})^\top (\mathbf{U}^\top \mathbf{v}_{k+1})} \\ &= \sqrt{\mathbf{v}_{k+1}^\top \mathbf{U}\mathbf{U}^\top \mathbf{v}_{k+1}} \\ &= \sqrt{\mathbf{v}_{k+1}^\top \mathbf{v}_{k+1}} \\ &= ||\mathbf{v}_{k+1}||_2.\end{aligned}$$

Hence, the update equation, in terms of $\tilde{\mathbf{v}}_k$ and $\tilde{\mathbf{w}}_k$, is

$$\tilde{\mathbf{v}}_{k+1} = \mathbf{\Lambda}\tilde{\mathbf{w}}_k, \qquad \tilde{\mathbf{w}}_{k+1} = \frac{\tilde{\mathbf{v}}_{k+1}}{||\tilde{\mathbf{v}}_{k+1}||}.$$

3. Let $\tilde{\mathbf{w}}_0 = \begin{pmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_n\end{pmatrix}^\top$. Since $\mathbf{\Lambda}$ is a diagonal matrix, we obtain

$$\tilde{\mathbf{v}}_1 = \begin{pmatrix}\lambda_1\alpha_1 \\ \lambda_2\alpha_2 \\ \vdots \\ \lambda_n\alpha_n\end{pmatrix} = \lambda_1\alpha_1 \begin{pmatrix}1 \\ \frac{\alpha_2}{\alpha_1}\frac{\lambda_2}{\lambda_1} \\ \vdots \\ \frac{\alpha_n}{\alpha_1}\frac{\lambda_n}{\lambda_1}\end{pmatrix}$$

and therefore

$$\tilde{\mathbf{w}}_1 = \frac{\lambda_1\alpha_1}{c_1} \begin{pmatrix}1 \\ \frac{\alpha_2}{\alpha_1}\frac{\lambda_2}{\lambda_1} \\ \vdots \\ \frac{\alpha_n}{\alpha_1}\frac{\lambda_n}{\lambda_1}\end{pmatrix},$$

where c_1 is a normalisation constant such that $\|\tilde{\mathbf{w}}_1\| = 1$ (i.e. $c_1 = \|\tilde{\mathbf{v}}_1\|$). Hence, for $\tilde{\mathbf{w}}_k$ it holds that

$$\tilde{\mathbf{w}}_k = \tilde{c}_k \begin{pmatrix}1 \\ \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k \\ \vdots \\ \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k\end{pmatrix},$$

where $\tilde{c}_k$ is again a normalisation constant such that $||\tilde{\mathbf{w}}_k|| = 1$.

As λ_1 is the dominant eigenvalue, $|\lambda_j/\lambda_1| < 1$ for $j = 2, 3, \dots, n$, so that

$$\lim_{k\to\infty}\left(\frac{\lambda_j}{\lambda_1}\right)^k = 0, \quad j = 2, 3, \dots, n,$$

and hence

$$\lim_{k\to\infty} \begin{pmatrix}1 \\ \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k \\ \vdots \\ \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k\end{pmatrix} = \begin{pmatrix}1 \\ 0 \\ \vdots \\ 0\end{pmatrix}.$$

For the normalisation constant $\tilde{c}_k$, we obtain

$$\tilde{c}_k = \frac{1}{\sqrt{1 + \sum_{i=2}^{n} \left(\frac{\alpha_i}{\alpha_1}\right)^2 \left(\frac{\lambda_i}{\lambda_1}\right)^{2k}}},$$

and therefore

$$\begin{aligned}\lim_{k\to\infty} \tilde{c}_k &= \frac{1}{\sqrt{1 + \sum_{i=2}^{n} \left(\frac{\alpha_i}{\alpha_1}\right)^2 \lim\limits_{k\to\infty} \left(\frac{\lambda_i}{\lambda_1}\right)^{2k}}} \\ &= \frac{1}{\sqrt{1 + \sum_{i=2}^{n} \left(\frac{\alpha_i}{\alpha_1}\right)^2 \cdot 0}} \\ &= 1.\end{aligned}$$

The limit of the product of two convergent sequences is the product of the limits so that

$$\lim_{k\to\infty} \tilde{\mathbf{w}}_k = \lim_{k\to\infty} \tilde{c}_k \lim_{k\to\infty} \begin{pmatrix} 1 \\ \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k \\ \vdots \\ \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

4. Since $\mathbf{w}_k = \mathbf{U}\tilde{\mathbf{w}}_k$, we obtain

$$\lim_{k\to\infty} \mathbf{w}_k = \mathbf{U} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{u}_1,$$

which is the eigenvector with the largest eigenvalue, i.e. the “first” or “dominant” eigenvector.