



Written Assignment 4

Yushan Liu Student ID: 2024214103

December 7, 2024

All the tasks in this WA is done entirely by MYSELF.

4.1 SVD Properties

(a)

Let A be an $m \times n$ matrix. The SVD of A is given by:

$$A = U\Sigma V^\top$$

where:

- U is an $m \times m$ orthogonal matrix containing the left singular vectors of A ,
- Σ is an $m \times n$ diagonal matrix containing the singular values of A ,
- V is an $n \times n$ orthogonal matrix containing the right singular vectors of A .

From the SVD, the left and right singular vectors of A , denoted u_i and v_i , respectively, satisfy the following relationships:

$$AA^\top u_i = \sigma_i^2 u_i \quad (1)$$

$$A^\top A v_i = \sigma_i^2 v_i \quad (2)$$

where σ_i is the i -th singular value.

So we have:

$$(A^\top)^\top A^\top u_i = \sigma_i^2 u_i \quad (3)$$

$$A^\top (A^\top)^\top v_i = \sigma_i^2 v_i \quad (4)$$

Thus, v_i is a right singular vector of A^\top with the corresponding singular value σ_i , and u_i is the left singular vector of A^\top .

(b)

Solution (1):

Consider the matrix $A^\top A$:

$$A^\top A = (V\Sigma^\top U^\top)(U\Sigma V^\top)$$

Since $U^\top U = I$ (orthogonality of U) and $V^\top V = I$ (orthogonality of V), we can simplify the expression:

$$A^\top A = V\Sigma^\top \Sigma V^\top$$

The matrix $\Sigma^\top \Sigma$ is a diagonal matrix where the diagonal entries are the squares of the singular values $\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2$, and the rest are zero. Therefore, we have:

$$A^\top A = V \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r^2 \end{bmatrix} V^\top$$

This can be written as:

$$A^\top A = \sum_{i=1}^r \sigma_i^2 v_i v_i^\top$$

where v_i is the i -th column of V , the corresponding right singular vector of A , and the sum is over all non-zero singular values.

Solution (2):

Apply $A^\top A$ to v_i :

$$A^\top A v_i = \left(\sum_{j=1}^r \sigma_j^2 v_j v_j^\top \right) v_i$$

Since $v_j^\top v_i = 0$ for $j \neq i$ (the right singular vectors are orthogonal), only the $j = i$ term survives in the sum:

$$A^\top A v_i = \sigma_i^2 v_i v_i^\top v_i = \sigma_i^2 v_i$$

Thus, v_i is an eigenvector of $A^\top A$ with eigenvalue σ_i^2 .

(c)

Perform the Singular Value Decomposition of matrix Ω :

$$\Omega = U \Sigma V^\top$$

Substitute the SVD of Ω into the expression:

$$\mathbf{c}^\top \Omega \mathbf{d} = \mathbf{c}^\top U \Sigma V^\top \mathbf{d}$$

Let $\mathbf{c}' = U^\top \mathbf{c}$ and $\mathbf{d}' = V^\top \mathbf{d}$. Since U and V are orthogonal matrices, we have:

$$\|\mathbf{c}'\|_2 = \|\mathbf{c}\|_2 = 1, \quad \|\mathbf{d}'\|_2 = \|\mathbf{d}\|_2 = 1$$

$$\mathbf{c}^\top \Omega \mathbf{d} = \mathbf{c}'^\top \Sigma \mathbf{d}'$$

Since Σ is a diagonal matrix, the above can be written as:

$$\mathbf{c}'^\top \Sigma \mathbf{d}' = \sum_{i=1}^p \sigma_i c'_i d'_i$$

where c'_i and d'_i are the i -th components of \mathbf{c}' and \mathbf{d}' , respectively.

Consider the following expression:

$$\mathbf{c}'^\top \Sigma \mathbf{d}' = \sum_{i=1}^p \sigma_i c'_i d'_i$$

By the Cauchy-Schwarz inequality:

$$\left| \sum_{i=1}^p \sigma_i c'_i d'_i \right| \leq \sqrt{\left(\sum_{i=1}^p \sigma_i^2 (c'_i)^2 \right) \left(\sum_{i=1}^p (d'_i)^2 \right)}$$

Given that $\|\mathbf{c}'\|_2 = 1$ and $\|\mathbf{d}'\|_2 = 1$, we have:

$$\sum_{i=1}^p (c'_i)^2 = 1, \quad \sum_{i=1}^p (d'_i)^2 = 1$$

Therefore:

$$\left| \sum_{i=1}^p \sigma_i c'_i d'_i \right| \leq \sqrt{\sum_{i=1}^p \sigma_i^2 (c'_i)^2} \leq \sigma_1 \sqrt{\sum_{i=1}^p (c'_i)^2} = \sigma_1$$

Here we used the fact that $\sigma_1 \geq \sigma_i$ for all i .

To achieve equality in the above inequality, the following conditions must be met:

$$\sigma_i c'_i d'_i = \sigma_1 c'_i d'_i \quad \forall i$$

This implies that only $i = 1$ has non-zero $c'_1 d'_1$, and $c'_1 = d'_1 = 1$. Thus, the optimal solution is:

$$\mathbf{c}' = \mathbf{e}_1, \quad \mathbf{d}' = \mathbf{e}_1$$

where \mathbf{e}_1 is the first standard basis vector.

From $\mathbf{c}' = U^\top \mathbf{c} = \mathbf{e}_1$, we obtain:

$$\mathbf{c} = U \mathbf{e}_1 = \mathbf{u}_1$$

Similarly, from $\mathbf{d}' = V^\top \mathbf{d} = \mathbf{e}_1$, we get:

$$\mathbf{d} = V \mathbf{e}_1 = \mathbf{v}_1$$

where \mathbf{u}_1 and \mathbf{v}_1 are the left and right singular vectors of Ω corresponding to the largest singular value σ_1 .

Therefore, when $\mathbf{c} = \mathbf{u}_1$ and $\mathbf{d} = \mathbf{v}_1$, the expression $\mathbf{c}^\top \Omega \mathbf{d}$ attains its maximum value σ_1 . This proves that \mathbf{c} and \mathbf{d} must be the left and right singular vectors of Ω corresponding to the largest singular value to maximize $\mathbf{c}^\top \Omega \mathbf{d}$.

(Bonus-d)

Solution (1):

Let M be an $n \times m$ matrix with rank 1. By definition, a rank 1 matrix has all its columns (and rows) linearly dependent. This implies that there exist non-zero vectors $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$ such that

$$M = \mathbf{u} \mathbf{v}^\top.$$

Construct \mathbf{u} and \mathbf{v} as follows:

1. **Choosing Vector \mathbf{u} :** Since M has rank 1, there exists at least one non-zero column. Let \mathbf{u} be one such non-zero column of M . Without loss of generality, assume $\mathbf{u} = \mathbf{m}_1$, the first column of M .

2. **Determining Vector \mathbf{v} :** Given that all columns of M are scalar multiples of \mathbf{u} , there exist scalars v_{1j} such that

$$M = \mathbf{u}[v_{11}, v_{12}, \dots, v_{1m}].$$

This means each column \mathbf{m}_j of M satisfies

$$\mathbf{m}_j = v_{1j}\mathbf{u}.$$

Therefore, we can construct the vector $\mathbf{v} = [v_{11}, v_{12}, \dots, v_{1m}]^\top$.

3. **Expressing M as an Outer Product:** Combining the above, we have

$$M = \mathbf{u}\mathbf{v}^\top,$$

where $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$ are non-zero vectors.

Thus, any rank 1 matrix M can indeed be written as the outer product of two vectors.

Solution (2):

Consider P and Q with their columns \mathbf{p}_i and \mathbf{q}_i , respectively. The (j, k) -th entry of PQ^\top is given by

$$(PQ^\top)_{jk} = \sum_{i=1}^r p_{ji}q_{ki}.$$

On the other hand, the (j, k) -th entry of $\sum_{i=1}^r \mathbf{p}_i \mathbf{q}_i^\top$ is

$$\sum_{i=1}^r (\mathbf{p}_i)_j (\mathbf{q}_i)_k = \sum_{i=1}^r p_{ji}q_{ki}.$$

Since both expressions are identical for all j and k , we have

$$PQ^\top = \sum_{i=1}^r \mathbf{p}_i \mathbf{q}_i^\top.$$

Given the matrices U , D , and V as defined above, we can construct the product UDV^\top as follows:

$$UDV^\top = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]^\top.$$

Multiplying U and D first, we obtain

$$UD = [\sigma_1 \mathbf{u}_1, \sigma_2 \mathbf{u}_2, \dots, \sigma_r \mathbf{u}_r].$$

Now, applying the verified hint with $P = UD$ and $Q = V$, we have

$$UDV^\top = \sum_{i=1}^r (\sigma_i \mathbf{u}_i) \mathbf{v}_i^\top = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^\top.$$

This shows that

$$\sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = UDV^\top.$$

4.2 PCA

(a)

Left side of the equation, using the properties of the Euclidean norm, we have:

$$\|\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u}\|^2 = (\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u})^\top (\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u}).$$

Expanding this expression using the distributive property of the inner product:

$$\begin{aligned} & (\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u})^\top (\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u}) \\ &= \mathbf{x}^{(i)\top} \mathbf{x}^{(i)} - 2(\mathbf{x}^{(i)\top} \mathbf{u})(\mathbf{u}^\top \mathbf{x}^{(i)}) + (\mathbf{x}^{(i)\top} \mathbf{u})^2 (\mathbf{u}^\top \mathbf{u}). \end{aligned}$$

Since \mathbf{u} is a unit vector ($\mathbf{u}^\top \mathbf{u} = 1$), this simplifies to:

$$\|\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u}\|^2 = \|\mathbf{x}^{(i)}\|^2 - (\mathbf{x}^{(i)\top} \mathbf{u})^2.$$

The goal is to minimize the squared projection residual:

$$\arg \min_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} \|\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u}\|^2.$$

Substituting the expanded form above:

$$\arg \min_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} (\|\mathbf{x}^{(i)}\|^2 - (\mathbf{x}^{(i)\top} \mathbf{u})^2).$$

Notice that $\|\mathbf{x}^{(i)}\|^2$ is a constant with respect to \mathbf{u} .

Therefore, minimizing the expression is equivalent to maximizing $(\mathbf{x}^{(i)\top} \mathbf{u})^2$:

$$\arg \min_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} (\|\mathbf{x}^{(i)}\|^2 - (\mathbf{x}^{(i)\top} \mathbf{u})^2) = \arg \max_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} (\mathbf{x}^{(i)\top} \mathbf{u})^2.$$

(b)

Based (a), the average squared projection residual over all m samples is:

$$\frac{1}{m} \sum_{i=1}^m \|\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u}\|^2 = \frac{1}{m} \sum_{i=1}^m \|\mathbf{x}^{(i)}\|^2 - \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^{(i)\top} \mathbf{u})^2.$$

Notice that $\frac{1}{m} \sum_{i=1}^m \|\mathbf{x}^{(i)}\|^2$ is a constant with respect to \mathbf{u} . Therefore, minimizing the average residual is equivalent to maximizing the average squared projection length:

$$\mathbf{u}^* = \arg \min_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} \frac{1}{m} \sum_{i=1}^m \|\mathbf{x}^{(i)} - (\mathbf{x}^{(i)\top} \mathbf{u}) \mathbf{u}\|^2 = \arg \max_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^{(i)\top} \mathbf{u})^2.$$

The covariance matrix Σ is defined as:

$$\Sigma = \frac{1}{m} \sum_{i=1}^m \mathbf{x}^{(i)} (\mathbf{x}^{(i)})^\top.$$

Consider the quadratic form $\mathbf{u}^\top \Sigma \mathbf{u}$:

$$\mathbf{u}^\top \Sigma \mathbf{u} = \mathbf{u}^\top \left(\frac{1}{m} \sum_{i=1}^m \mathbf{x}^{(i)} (\mathbf{x}^{(i)})^\top \right) \mathbf{u} = \frac{1}{m} \sum_{i=1}^m \mathbf{u}^\top \mathbf{x}^{(i)} (\mathbf{x}^{(i)})^\top \mathbf{u} = \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^{(i)\top} \mathbf{u})^2.$$

Thus, the optimization problem becomes:

$$\mathbf{u}^* = \arg \max_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} \mathbf{u}^\top \Sigma \mathbf{u}.$$

The **Rayleigh quotient** for a symmetric matrix Σ and a non-zero vector \mathbf{u} is defined as:

$$R(\mathbf{u}) = \frac{\mathbf{u}^\top \Sigma \mathbf{u}}{\mathbf{u}^\top \mathbf{u}}.$$

Given the constraint $\mathbf{u}^\top \mathbf{u} = 1$, the Rayleigh quotient simplifies to:

$$R(\mathbf{u}) = \mathbf{u}^\top \Sigma \mathbf{u}.$$

The Rayleigh-Ritz theorem states that the maximum value of $R(\mathbf{u})$ is the largest eigenvalue of Σ , and it is attained when \mathbf{u} is the corresponding eigenvector. Therefore:

$$\mathbf{u}^* = \arg \max_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} \mathbf{u}^\top \Sigma \mathbf{u} = \text{Eigenvector corresponding to the largest eigenvalue of } \Sigma.$$

From steps below, we have:

$$\mathbf{u}^* = \arg \max_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^{(i)\top} \mathbf{u})^2 = \arg \max_{\mathbf{u}: \mathbf{u}^\top \mathbf{u} = 1} \mathbf{u}^\top \Sigma \mathbf{u}.$$

By the Rayleigh-Ritz theorem, \mathbf{u}^* is the eigenvector corresponding to the largest eigenvalue of Σ .

4.3 ICA

(a)

Since s_1 and s_2 are independent, the joint probability density function is the product of their individual densities. Therefore,

$$P(s_1, s_2) = P(s_1) \cdot P(s_2).$$

Substituting the given expressions for $P(s_1)$ and $P(s_2)$:

$$P(s_1, s_2) = \left(\frac{1}{\sqrt{2\pi}} \exp \left(-\frac{s_1^2}{2} \right) \right) \cdot \left(\frac{1}{\sqrt{2\pi}} \exp \left(-\frac{s_2^2}{2} \right) \right).$$

Simplifying the expression:

$$P(s_1, s_2) = \frac{1}{2\pi} \exp \left(-\frac{s_1^2 + s_2^2}{2} \right).$$

(b)

Since the mixing matrix A is orthogonal, we have:

$$\mathbf{x} = A\mathbf{s},$$

where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$.

Given that A is orthogonal ($A^{-1} = A^\top$), the inverse transformation is:

$$\mathbf{s} = A^\top \mathbf{x}.$$

Since s_1 and s_2 are independent Gaussian random variables, their joint probability density function is:

$$P(s_1, s_2) = P(s_1) \cdot P(s_2) = \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s_1^2}{2}\right) \right) \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s_2^2}{2}\right) \right) = \frac{1}{2\pi} \exp\left(-\frac{s_1^2 + s_2^2}{2}\right).$$

To find $P(x_1, x_2)$, we perform a change of variables from \mathbf{s} to \mathbf{x} . The Jacobian determinant of the transformation $\mathbf{x} = A\mathbf{s}$ is $|\det(A)| = 1$ since A is orthogonal. Therefore, the joint probability density function $P(x_1, x_2)$ is:

$$P(x_1, x_2) = P(s_1, s_2) \cdot |\det(A^{-1})| = \frac{1}{2\pi} \exp\left(-\frac{s_1^2 + s_2^2}{2}\right).$$

Expressing s_1 and s_2 in terms of x_1 and x_2 using $\mathbf{s} = A^\top \mathbf{x}$, we have:

$$s_1^2 + s_2^2 = (A^\top \mathbf{x})^\top (A^\top \mathbf{x}) = \mathbf{x}^\top A A^\top \mathbf{x} = \mathbf{x}^\top \mathbf{x} = x_1^2 + x_2^2.$$

Thus,

$$P(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right).$$

Why Gaussian Variables Are Forbidden in ICA

1. Rotational Invariance of Gaussian Distributions:

For Gaussian random variables, any orthogonal transformation (rotation) of the variables results in another Gaussian distribution with the same form. Specifically, if \mathbf{x} is Gaussian, then $Q\mathbf{x}$ is also Gaussian for any orthogonal matrix Q . Mathematically,

$$Q\mathbf{x} \sim \mathcal{N}(0, Q\Sigma Q^\top),$$

where Σ is the covariance matrix of \mathbf{x} . If \mathbf{x} has a covariance matrix proportional to the identity matrix (as in whitening), the transformed variables $Q\mathbf{x}$ will also have the same covariance matrix:

$$Q\Sigma Q^\top = Q \left(\frac{1}{m} \sum_{i=1}^m \mathbf{x}^{(i)} (\mathbf{x}^{(i)})^\top \right) Q^\top = \frac{1}{m} \sum_{i=1}^m (Q\mathbf{x}^{(i)}) (Q\mathbf{x}^{(i)})^\top = \Sigma.$$

Therefore, the joint distribution remains unchanged under any orthogonal transformation.

2. Lack of Unique Solution:

Since Gaussian distributions are rotationally invariant, there is no unique rotation (orthogonal matrix) that can recover the original independent components from the mixed signals. In other words, any orthogonal transformation of the mixed signals results in another valid solution with the same joint distribution. This means that ICA cannot uniquely identify the original independent Gaussian components because multiple rotated versions of the components are equally valid.

3. Insufficient Higher-Order Statistics:

ICA relies on higher-order statistics (beyond the second-order covariance) to identify and separate independent components. Gaussian distributions are fully characterized by their first and second moments (mean and covariance). Since all higher-order cumulants of Gaussian distributions are zero, ICA lacks the necessary information to distinguish between different independent components when they are Gaussian.

References

- [1] Andrew Ng, Tengyu Ma. *CCS 229 Lecture Notes*. Stanford University, 2023. Available online at: <https://cs229.stanford.edu/>
- [2] Stephen Boyd, Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [3] OpenAI. *ChatGPT: A Conversational AI*. 2023. Available online at: <https://www.openai.com/chatgpt>
- [4] K. L. Chung. *Stochastic Processes*. 2nd ed. Springer, 2001.