Matrix Decompositions

Andrew Andreas

May 9, 2024

1 Introduction

I created these notes purely for my own knowledge and they are intended to be a practical guide of useful properties.

Inspired by Garret Thomas's excellent notes which can be found here

Please report any errors to andrew.andreas0@gmail.com

2 Characteristics of a Matrix

There are two separate values that can summarize a matrix, the determinant and its trace, for which we will briefly state their properties in the following subsections.

2.1 The Determinant

The determinant, denoted $\det(\mathbf{A})$, of a matrix is only defined for square matrices, $\mathbf{A} \in \mathcal{R}^{n\times n}$ and may sometimes be denoted as $|\cdot|$ as shown below.

$$\det(\mathbf{A}) = \begin{vmatrix} a_{11} & a_{12} & \dots & a1n \\ a_{21} & a_{22} & \dots & a2n \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

The determinant is a function that maps the matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$ to a scalar value.

Theorem 1. A square matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$. This means that a matrix is *non-singular*, i.e., is *regular* and therefore has n linearly independent columns so that it is full rank

Theorem 2. (Laplace Expansion) For a square matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$, then for $j = 1, 2, \dots, n$:

1. *Expansion along column j*

$$\det(\mathbf{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det A'_{j,k}$$

2. *Expansion along row j*

$$\det(\mathbf{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det A'_{k,j}$$

where $\det A'_{k,j}$ is the sub-matrix of A when we isolate row k and column j

2.1.1 Properties of Determinants

The determinant of a square matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$, has the following properties

- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
- Invariance to transposition, $\det(\mathbf{A}) = \det(\mathbf{A^T})$

- $\det(\mathbf{A^{-1}}) = \frac{1}{\det(\mathbf{A})}$
- The determinant is invariant to the choice of basis. As a result, *Similar Matrices* have the same determinant.
- Adding a multiple of a column or row to another does not change the determinant
- Multiplication of a column / row with a scalar, $\lambda \in \mathcal{R}$, scales the determinant by λ. This is a consequence of the multilinearity of the determinant
  - $\det(\lambda\mathbf{v_1} \mid \mathbf{v_2} \mid \ldots \mid \mathbf{v_n}) = \lambda \det(\mathbf{v_1} \mid \mathbf{v_2} \mid \ldots \mid \mathbf{v_n})$
  - $\det(\mathbf{v_1} + \mathbf{w_n} \mid \mathbf{v_2} \mid \ldots \mid \mathbf{v_n}) = \det(\mathbf{v_1} \mid \mathbf{v_2} \mid \ldots \mid \mathbf{v_n}) + \det(\mathbf{w_n} \mid \mathbf{v_2} \mid \ldots \mid \mathbf{v_n})$
- Swapping two rows/columns of a determinant changes its sign

2.1.2 Efficient Computation of the Determinant

As a result of the last three properties, we can reduce a matrix using Gaussian Elimination to its upper triangular form, $\mathbf{T}, T_{ij} = 0$ for $i > j$.

For a triangular matrix

$$\det(\mathbf{T}) = \prod_{i=1}^{n} T_{ii} \tag{1}$$

2.2 The Trace

The Trace of a square matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$ is defined as

$$\text{tr}(\mathbf{A}) := \sum_{i=1}^{n} a_{ii} \tag{2}$$

2.2.1 Properties of the Trace

The trace of a square matrix $\mathbf{A}, \mathbf{B} \in \mathcal{R}^{n\times n}$, has the following properties

- $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
- $\text{tr}(\alpha\mathbf{A}) = \alpha\text{tr}(\mathbf{A}), \alpha \in \mathcal{R}$
- $\text{tr}(\mathbf{I_n}) = n$
- $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$

3 Eigenvalues and Eigenvectors

Eigenvalue Equation For a square matrix, $\mathbf{A} \in \mathcal{R}^{n\times n}$, then $\lambda \in \mathcal{R}$ is an eigenvalue of $\mathbf{A}$ and $\mathbf{x} \in \mathcal{R}^n - \{\mathbf{0}\}$ is the corresponding eigenvector if

$$\mathbf{Ax} = \lambda\mathbf{x} \tag{3}$$

λ is an eigenvalue of $\mathbf{A} \in \mathcal{R}^{n\times n}$ if and only if it is the root of the characteristic polynomial, $P_A(\lambda)$, of $\mathbf{A}$. The characteristic polynomial, $P_A(\lambda)$, of a matrix $\mathbf{A}$ is defined below

$$P_A(\lambda) := \det(\mathbf{A} - \lambda\mathbf{I}) \tag{4}$$

where $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ since the matrix is not full rank because there exists a non-trivial solution (i.e., $\mathbf{x} \neq 0$) to the problem

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \tag{5}$$

This means that the eigenvector, $\mathbf{x}$, which is a solution to the above equation for a given λ is in the null space of $\mathbf{A} - \lambda\mathbf{I}$ since it is the solution to $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$

Eigenspace For a matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$, the Eigenspace is the set of all eigenvectors associated with a specific eigenvalue, λ and it spans a subspace of $\mathcal{R}^n$. The eigenspace is denoted E_λ.

3.1 Useful Properties

- $\mathbf{A}$ and $\mathbf{A}^T$ have the same eigenvalues, though not necessarily the same eigenvalues since $\det(\mathbf{A}) = \det(\mathbf{A}^T)$. To see this, note that the transpose is a linear operator, then

$$\begin{aligned}
\det\left((\mathbf{A}-\lambda\mathbf{I})^T\right) &= (\mathbf{A}-\lambda\mathbf{I})^T & (6)\\
&= \det\left(\mathbf{A}^T - \lambda\mathbf{I}^T\right) & (7)\\
&= \det\left(\mathbf{A}^T - \lambda\mathbf{I}\right) & (8)\\
&= \det\left(\mathbf{A} - \lambda\mathbf{I}\right) & (9)
\end{aligned}$$

  which means they have the same roots of the characteristic polynomial and therefore the same eigenvalues

- Similar matrices possess the same eigenvalues - this means that a linear mapping, Φ has eigenvalues that are independent of choice of basis

- Symmetric, positive definite matrices always have positive, real eigenvalues

The eigenvectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ of a matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$ with n distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ are linearly independent and thus form a basis of $\mathcal{R}^n$. An eigenvalue always has at least one associated eigenvector but if the algebraic multiplicity > 1, then it may have more. Note that the geometric multiplicity $\leq$ algebraic multiplicity.

Algebraic Multiplicity is the number of times a given eigenvalue appears as the root of the characteristic polynomial

Geometric Multiplicity is the number of linearly independent eigenvectors in the eigenspace of the eigenvalue λ_i

Theorem 3. (Spectral Theorem) If $\mathbf{A} \in \mathcal{R}^{n\times n}$ is symmetric, there exists an orthonormal basis of the corresponding vector space V consisting of eigenvectors of $\mathbf{A}$, and each eigenvalue is real.

Given a matrix $\mathbf{A} \in \mathcal{R}^{n\times m}$ we can construct a symmetric, positive semidefinite matrix $\mathbf{S} \in \mathcal{R}^{m\times m}$

$$\begin{aligned}
\mathbf{S} &:= \mathbf{A}^T\mathbf{A} & (10)\\
&= \begin{pmatrix} \text{——} & \mathbf{a}_1^T & \text{——} \\ \text{——} & \mathbf{a}_2^T & \text{——} \\ & \vdots & \\ \text{——} & \mathbf{a}_m^T & \text{——} \end{pmatrix} \begin{pmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_m \\ | & | & & | \end{pmatrix} & (11)\\
&= \begin{pmatrix} \mathbf{a}_1^T\mathbf{a}_1 & \mathbf{a}_1^T\mathbf{a}_2 & ... & \mathbf{a}_1^T\mathbf{a}_m \\ \mathbf{a}_2^T\mathbf{a}_1 & \mathbf{a}_2^T\mathbf{a}_2 & ... & \mathbf{a}_2^T\mathbf{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m^T\mathbf{a}_1 & ... & ... & \mathbf{a}_m^T\mathbf{a}_m \end{pmatrix} & (12)
\end{aligned}$$

where we can also refer to $\mathbf{S}$ as the correlation matrix as it computes a dot product between the component vectors of the matrix, $\mathbf{A}_i,\ i = 1,..,m$. Note that just from equation 12 we can see that $\mathbf{S}$ is symmetric since the dot product is commutative, i.e., the entry $\mathbf{S}_{ij} = \mathbf{a}_i^T\mathbf{a}_j = \mathbf{a}_j^T\mathbf{a}_i = \mathbf{S}_{ji}$. However, the usual definition follows

$$\begin{aligned}
\mathbf{S} &= \mathbf{A}^T\mathbf{A} & (13)\\
&= (\mathbf{A}\mathbf{A}^T)^T & (14)\\
&= (\mathbf{A})^T(\mathbf{A}^T)^T & (15)\\
&= \mathbf{A}^T\mathbf{A} & (16)\\
&= \mathbf{S}^T & (17)
\end{aligned}$$

Positive Definite A square matrix $\mathbf{A} \in \mathcal{R}^{n\times n}$ is said to be positive definite if for any vector $\mathbf{x} \in \mathcal{R}^n - \mathbf{0}$

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0 \tag{18}$$

which is the quadratic form of the matrix $\mathbf{A}$. If the equality is $\geq$ such that $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ then the matrix is said to be positive semidefinite.

The positive definiteness of a matrix is defined via the inner product. For a detailed treatment of the inner product please refer to these notes

4 Matrix Decompositions

4.1 Eigendecomposition and Diagonalization

Diagonal matrices are matrices whose non-diagonal elements are zero and are thus of the form

$$\begin{pmatrix} k_1 & 0 & \dots & 0 \\ 0 & k_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & k_n \end{pmatrix}$$

A matrix, $\mathbf{A} \in \mathcal{R}^{n\times n}$, has a diagonal form if it is *similar* to a diagonal matrix.

Recall that two matrices, $\mathbf{D}$ and $\mathbf{A} \in \mathcal{R}^{n\times n}$ are similar if there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$$

If we left multiply $\mathbf{D}$ by $\mathbf{P}$ and assume that the elements of $\mathbf{D}$ are scalars $\lambda_1, \lambda_2, ..., \lambda_n$ and $\mathbf{P} = (\mathbf{p}_1 \mid \mathbf{p}_2 \mid \dots \mid \mathbf{p}_n)$, then

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D} \tag{19}$$

$$(\mathbf{A}\mathbf{p}_1 \mid \mathbf{A}\mathbf{p}_2 \mid \dots \mid \mathbf{A}\mathbf{p}_n) = (\lambda_1\mathbf{p}_1 \mid \lambda_1\mathbf{p}_2 \mid \dots \mid \lambda_n\mathbf{p}_n) \tag{20}$$

which can be rewritten into the familiar form to show that the vectors, $\mathbf{p}_i, i = 1, .., n$ are the eigenvectors that correspond to the eigenvalues $\lambda_i, i = 1, ..., n$ of the matrix $\mathbf{A}$

$$\mathbf{A}\mathbf{p}_1 = \lambda_1\mathbf{p}_1 \tag{21}$$

$$\vdots \tag{22}$$

$$\mathbf{A}\mathbf{p}_n = \lambda_n\mathbf{p}_n \tag{23}$$

Theorem 4. (Eigendecomposition) A square matrix, $\mathbf{A} \in \mathcal{R}^{n\times n}$ can be decomposed into

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \tag{24}$$

if and only if the eigenvectors of $\mathbf{A}$ form a basis of $\mathcal{R}^{n\times n}$

Theorem 5. Symmetric matrices always have a diagonal form which, following on from above, means that the eigenvectors of a symmetric matrix always form a basis of $\mathcal{R}^{n\times n}$

This is a direct consequence of the **Spectral Theorem** which goes further and states that an orthonormal basis exists for the eigenvectors of $\mathbf{A}$. As a result of the properties of an orthonormal matrix, i.e., $\mathbf{P}^T = \mathbf{P}^{-1}$, the eigendecomposition can be written as

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T \tag{25}$$

which provides the computational benefit of not having to compute the inverse matrix of $\mathbf{P}$.

4.2 The Singular Value Decomposition

The singular value decomposition, **SVD**, is a matrix decomposition that exists for all matrices, not just square matrices. For a matrix $\mathbf{A} \in \mathcal{R}^{n\times m}$ then its **SVD** is

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{26}$$

where $\boldsymbol{\Sigma} \in \mathcal{R}^{n\times m}$ with $\Sigma_{ii} = \sigma_i \geq 0$ and $\Sigma_{ij} = 0, i \neq j$ and both $\mathbf{U} \in \mathcal{R}^{n\times n}$ and $\mathbf{V} \in \mathcal{R}^{m\times m}$ are orthonormal matrices.

The columns of $\mathbf{U}$, denoted $\mathbf{u}_i, i = 1, ..., n$ are called the left singular vectors and the columns of $\mathbf{V}$, denoted $\mathbf{v}_i, i = 1, ..., m$ are called the right singular vectors.

For a rectangular matrix $\mathbf{A} \in \mathcal{R}^{n\times m}$, recall that this matrix represents a linear mapping, Φ from a vector space $V \in \mathcal{R}^m$ to the vector space $W \in \mathcal{R}^n$

$$\Phi : V \longrightarrow W \tag{27}$$

each of these vector spaces have their own standard bases, B and C. This then means that the $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ decomposition performs the same linear mapping but does so sequentially.

1. The matrix $\mathbf{V}$ performs a basis change from the canonical basis, B in $\mathcal{R}^m$ to another basis of $\mathcal{R}^m$, $\tilde{B}$, which is the eigenvectors of the matrix $\mathbf{A}\mathbf{A}^T$
2. Having performed the change of basis to $\tilde{B}$, the matrix $\boldsymbol{\Sigma}$ (i) then adds or deletes dimensions and (ii) scales the coordinate vectors by the singular values, σ_i
   - This is the transformation matrix of the linear mapping $\Phi : V \longrightarrow W, \tilde{B} \longrightarrow \tilde{C}$
3. The matrix $\mathbf{U}$ performs a basis change from the basis $\tilde{C} \longrightarrow C$ in the codomain $\mathcal{R}^n$

To compute the component matrices of the singular value decomposition, we will start with the right singular matrix, $\mathbf{V}^T$. Recall the Spectral Theorem which states that for a symmetric matrix its eigenvectors can form an orthonormal basis and that we can construct a symmetric matrix from any matrix $\mathbf{A}$ by computing the dot product with itself: $\mathbf{A}^T\mathbf{A}$

$$\mathbf{A}^T\mathbf{A} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\right)^T\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{28}$$
$$= \mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{29}$$
$$= \mathbf{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\mathbf{V}^T \tag{30}$$

where since $\mathbf{U}$ is an orthonormal matrix, then $\mathbf{U}^{-1} = \mathbf{U}^T$, so that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$

If we compare this result to that of the eigendecomposition of $\mathbf{A}^T\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$, then we can directly see that

$$\mathbf{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\mathbf{V}^T = \mathbf{P}\mathbf{D}\mathbf{P}^T \tag{31}$$
$$\mathbf{V} = \mathbf{P} \tag{32}$$
$$\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} = \mathbf{D} \tag{33}$$
$$\sigma_i^2 = \lambda_i \tag{34}$$

which tells us that the right singular vectors are the eigenvectors of $\mathbf{A}^T\mathbf{A}$ and that the singular values of $\mathbf{A}, \sigma_i$ are the square root of the eigenvalues of $\mathbf{A}^T\mathbf{A}, \lambda_i$

Turning our attention to the left singular vectors, $\mathbf{U}$ we follow a similar procedure in that we form another symmetric matrix, $\mathbf{A}\mathbf{A}^T$ yielding

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\right)^T \tag{35}$$
$$= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T \tag{36}$$
$$= \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\mathbf{U}^T \tag{37}$$

where if we again compare this to the eigendecomposition of $\mathbf{A}\mathbf{A}^T = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^T$, we can see that

$$
\begin{aligned}
\mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T =& \mathbf{S}\mathbf{\Lambda}\mathbf{S}^T && (38)\\
\mathbf{U} =& \mathbf{S} && (39)\\
\mathbf{\Sigma}\mathbf{\Sigma}^T =& \mathbf{\Lambda} && (40)\\
\sigma_i^2 =& \lambda_i && (41)
\end{aligned}
$$

So the left singular vectors $\mathbf{U}$ are the eigenvectors of the matrix $\mathbf{A}\mathbf{A}^T$ and the eigenvalues of $\mathbf{A}\mathbf{A}^T$ are the square of the singular values of $\mathbf{A}$

Note that since the singular values of $\mathbf{A}$ are equal to the square root of the eigenvalues of both $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$, then it is true that the eigenvalues of these matrices must be the same.

4.2.1 Computing the left singular vectors U

An alternative way to compute the left singular vectors $\mathbf{u}_i$, other than via the characteristic polynomial of $\mathbf{A}\mathbf{A}^T$, which make up the columns of $\mathbf{U}$ is to consider

$$
\begin{aligned}
\mathbf{A}^T\mathbf{A}\mathbf{v}_i =& \lambda_i\mathbf{v}_i && (42)\\
\mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{v}_i) =& \lambda_i(\mathbf{A}\mathbf{v}_i) && (43)
\end{aligned}
$$

which is the eigenvalue equation for $\mathbf{A}\mathbf{A}^T$ and it tells us that $\mathbf{A}\mathbf{v}_i$, the image of the right singular vector under $\mathbf{A}$, is the i^{th} eigenvector. So to construct the orthonormal basis required for the left singular vector $\mathbf{U}$, then we need to normalize these vectors

$$
\mathbf{u}_i = \frac{\mathbf{A}\mathbf{v}_i}{||\mathbf{A}\mathbf{v}_i||} = \frac{1}{\sqrt{\lambda_i}}\mathbf{A}\mathbf{v}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i \qquad (44)
$$

4.2.2 Intuition of the SVD

For a matrix $\mathbf{A} \in \mathcal{R}^{n\times m}$ where $n \gg m$, the SVD is given by

$$
\begin{aligned}
\mathbf{A} =& \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T && (45)\\
=& \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \end{bmatrix} \begin{bmatrix} \text{——} & \mathbf{v}_1^T & \text{——} \\ \text{——} & \mathbf{v}_2^T & \text{——} \\ \text{——} & \vdots & \text{——} \end{bmatrix} && (46)\\
=& \begin{bmatrix} | & | & & | \\ \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \dots & 0 \\ | & | & & | \end{bmatrix} \begin{bmatrix} \text{——} & \mathbf{v}_1^T & \text{——} \\ \text{——} & \mathbf{v}_2^T & \text{——} \\ \text{——} & \vdots & \text{——} \end{bmatrix} && (47)\\
=& \sigma_1\mathbf{u_1}\mathbf{v_1}^T + \sigma_2\mathbf{u_2}\mathbf{v_2}^T + \dots + \sigma_m\mathbf{u_m}\mathbf{v_m}^T && (48)
\end{aligned}
$$

where despite the fact we have n left singular vectors, because $\mathbf{A}$ is a $\mathbf{R}^{n\times m}$ matrix where $n \gg m$, this means that it can be at most rank m and hence can have at most m distinct singular values.

So we have a linear combination of the vectors of $\mathbf{U}$ and $\mathbf{V}$ which are scaled by the singular values and they can be interpreted as a sum of rank 1 matrices that increasingly improve the estimation of $\mathbf{A}$. We can see this because if we consider the dimensions of $\mathbf{u}_i$ and $\mathbf{v}_i^T$, this is an *outer product* between an $n \times 1$ column vector and a $1 \times n$ row vector which yields a $n \times n$ matrix. The reason this matrix is of rank 1 is because the entire matrix is entirely dependent on a single row and column vector.