



Lecture 5

Math Foundations Team



BITS Pilani

Pilani | Dubai | Goa | Hyderabad



- ▶ In the previous lecture, we discussed eigenvalues and eigenvectors of matrices
- ▶ In this lecture, we will look at two related methods for factorizing matrices into canonical forms.
- ▶ The first one is known as Eigenvalue decomposition. It uses the concepts of eigenvalues and eigenvectors to generate the decomposition
- ▶ The second method known as singular value decomposition or SVD is applicable to all matrices



- ▶ A diagonal matrix is a matrix that has value zero on all off diagonal elements.

$$\mathbf{D} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}$$

- ▶ For a diagonal matrix \mathbf{D} , the determinant is the product of its diagonal entries.
- ▶ A matrix power \mathbf{D}^k is given by each diagonal element raised to the power k .
- ▶ Inverse of a diagonal matrix is obtained by taking inverse of non-zero diagonal entry.



- ▶ A matrix $A \in \mathbb{R}^{n \times n}$ is diagonalizable if there exists an invertible matrix $P \in \mathbb{R}^{n \times n}$ and a diagonal matrix \mathbf{D} such that $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$
- ▶ In the definition of diagonalization, it is required that \mathbf{P} is an invertible matrix. Assume p_1, p_2, \dots, p_n are the n columns of \mathbf{P}
- ▶ Rewriting we get $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$. By observing that \mathbf{D} is a diagonal matrix, we can simplify as

$$\mathbf{A}p_i = \lambda_i p_i$$

where λ_i is the i^{th} diagonal entry in \mathbf{D} .

Diagonalizable Matrix



- ▶ Consider a square matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$$

- ▶ Consider the invertible matrix

$$\mathbf{P} = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$$

- ▶ Now consider the product $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ as follows

$$\begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}$$



- ▶ Recall the existence of eigenvalues and eigenvectors for square matrices
- ▶ Eigenvalues can be used to create a matrix decomposition known as Eigenvalue Decomposition
- ▶ A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ can be factored into

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

- ▶ where \mathbf{P} is an invertible matrix of eigenvectors of \mathbf{A} assuming we can find n eigenvectors that form a basis of \mathbb{R}^n
- ▶ and \mathbf{D} is a diagonal matrix whose diagonal entries are the eigenvalues of \mathbf{A}

Example of Eigendecomposition



Let us compute the eigendecomposition of the matrix A

$$\mathbf{A} = \begin{bmatrix} 2.5 & -1 \\ -1 & 2.5 \end{bmatrix}$$

- ▶ Step 1: Find the eigenvalues and eigenvectors

$$\mathbf{A} - \lambda \mathbf{I} = \begin{bmatrix} 2.5 - \lambda & -1 \\ -1 & 2.5 - \lambda \end{bmatrix}$$

- ▶ The characteristic equation is given by $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$
- ▶ This leads to the equation $\lambda^2 - 5\lambda + \frac{21}{4} = 0$
- ▶ Solving the quadratic equation gives us $\lambda_1 = 3.5$ and $\lambda_2 = 1.5$

Example of Eigendecomposition



- ▶ The eigenvector corresponding to $\lambda_1 = 3.5$ is derived as

$$p_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

- ▶ The eigenvector corresponding to $\lambda_1 = 1.5$ is derived as

$$p_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

- ▶ Step 2 : Construct the matrix **P** to diagonalize **A**

$$\mathbf{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Example of Eigendecomposition



- ▶ The inverse of matrix P is given by

$$\mathbf{P}^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

- ▶ The eigendecomposition of the matrix A is given by

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3.5 & 0 \\ 0 & 1.5 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

- ▶ In summary we have obtained the required matrix factorization using eigenvalues and eigenvectors.



- ▶ Recall that a matrix A is called symmetric matrix if $\mathbf{A} = \mathbf{A}^T$

$$\mathbf{A} = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$$

- ▶ A Symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ can always be diagonalized.
- ▶ This follows directly from the spectral theorem discussed in previous lecture
- ▶ Moreover the spectral theorem states that we can find an orthogonal matrix \mathbf{P} of eigenvectors of \mathbf{A} .

Motivation for Singular Value Decomposition



- ▶ The singular value decomposition or (SVD) of a matrix is a central matrix decomposition method in linear algebra.
- ▶ The eigenvalue decomposition is applicable to square matrices only.
- ▶ The singular value decomposition exists for all rectangular matrices
- ▶ SVD involves writing a matrix as a product of three matrices \mathbf{U} , $\mathbf{\Sigma}$ and \mathbf{V}^T .
- ▶ The three component matrices are derived by applying eigenvalue decomposition discussed previously

Singular Value Decomposition Theorem



- ▶ Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a rectangular matrix. Assume that \mathbf{A} has rank r .
- ▶ The Singular value decomposition of \mathbf{A} is defined as

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

- ▶ $\mathbf{U} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix with column vectors u_i where $i = 1, \dots, m$
- ▶ $\mathbf{V} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix with column vectors v_j where $j = 1, \dots, n$
- ▶ $\mathbf{\Sigma}$ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i > 0$
- ▶ The diagonal entries $\sigma_i, i = 1, \dots, r$ of $\mathbf{\Sigma}$ are called the singular values.
- ▶ By convention, the singular values are ordered i.e $\Sigma_{ii} > \Sigma_{jj}$ where $i < j$.



- ▶ The singular value matrix Σ is unique.
- ▶ Observe that the $\Sigma \in \mathbb{R}^{m \times n}$ matrix is rectangular. In particular, Σ is of the same size as \mathbf{A} .
- ▶ This means that Σ has a diagonal submatrix that contains the singular values and needs additional zero padding.
- ▶ Specifically, if $m > n$, then the matrix Σ has diagonal structure up to row n and then consists of zero rows.
- ▶ If $m < n$, the matrix Σ has a diagonal structure up to column m and columns that consist of 0 from $m + 1$ to n .



- ▶ It can be observed that

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{\Sigma}^T \mathbf{\Sigma} \mathbf{V}^T$$

- ▶ Since $\mathbf{A}^T \mathbf{A}$ has the following eigendecomposition

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

- ▶ Therefore, the eigenvectors of $\mathbf{A}^T \mathbf{A}$ that compose \mathbf{P} are the right-singular vectors \mathbf{V} of \mathbf{A} .
- ▶ The eigenvalues of $\mathbf{A}^T \mathbf{A}$ are the squared singular values of $\mathbf{\Sigma}$



- ▶ It can be observed that

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

- ▶ Since $\mathbf{A}\mathbf{A}^T$ has the following eigendecomposition

$$\mathbf{A}\mathbf{A}^T = \mathbf{S}\mathbf{D}\mathbf{S}^T$$

- ▶ Therefore, the eigenvectors of $\mathbf{A}\mathbf{A}^T$ that compose \mathbf{S} are the left-singular vectors \mathbf{U} of \mathbf{A}



- ▶ $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ can be rearranged to obtain a simple formulation for u_i
- ▶ By postmultiplying by \mathbf{V} we get $\mathbf{AV} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}$
- ▶ By observing that \mathbf{V} is orthogonal we obtain a simple form

$$\mathbf{AV} = \mathbf{U}\mathbf{\Sigma}$$

- ▶ This is equivalent to the following

$$u_i = \frac{1}{\sigma_i} \mathbf{A} \mathbf{v}_i \quad \forall i = 1, 2, \dots, r$$



- ▶ We want to find SVD of the following rectangular matrix \mathbf{A}

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

- ▶ Let us consider the matrix $\mathbf{A}^T \mathbf{A}$ derived from \mathbf{A} given by

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- ▶ It is a symmetric matrix

Computing Singular Value Decomposition 2



- ▶ Derive the eigendecomposition of $\mathbf{A}^T \mathbf{A}$ in the form PDP^T
- ▶ The matrix \mathbf{P} is given by

$$\mathbf{P} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

- ▶ The matrix \mathbf{D} is given by

$$\mathbf{D} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$



Now we construct the singular value matrix Σ

- ▶ The matrix Σ has the dimension same as \mathbf{A} . In this case Σ is hence a 2×3 matrix.
- ▶ The diagonal entries of submatrix is obtained by taking square root of 6 and 1 respectively
- ▶ Singular-value matrix Σ is given by:

$$\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

- ▶ The last column is a column of zeros only



Left singular vectors as the normalized image of the right singular vectors. Recall that $u_i = \frac{1}{\sigma_i} \mathbf{A} v_i$

- ▶ The first vector

$$u_1 = \frac{1}{\sigma_1} \mathbf{A} v_1 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \end{bmatrix}$$

- ▶ The second vector

$$u_2 = \frac{1}{\sigma_2} \mathbf{A} v_2 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

Final Step : Combining U , Σ and V



We compile all the three matrices together to generate the SVD



$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}^T$$

- ▶ The matrix \mathbf{U} is an 2×2 matrix satisfying orthogonality property.
- ▶ The matrix \mathbf{V} is an 3×3 matrix satisfying orthogonality property.

Comparing SVD and EVD



- ▶ The left-singular vectors of \mathbf{A} are eigenvectors of $\mathbf{A}\mathbf{A}^T$
- ▶ The right-singular vectors of \mathbf{A} are eigenvectors of $\mathbf{A}^T\mathbf{A}$
- ▶ The non-zero singular values of \mathbf{A} are the square roots of the nonzero eigenvalues of $\mathbf{A}^T\mathbf{A}$.
- ▶ The SVD always exists for any matrix in $\mathbb{R}^{m \times n}$
- ▶ The eigendecomposition is only defined for square matrices in $\mathbb{R}^{n \times n}$ and only exists if we can find a basis of eigenvectors of \mathbb{R}^n



- ▶ The vectors in the eigendecomposition matrix \mathbf{P} are not necessarily orthogonal.
- ▶ On the other hand, the vectors in the matrices \mathbf{U} and \mathbf{V} in the SVD are orthonormal.
- ▶ Both the eigendecomposition and the SVD are compositions of three linear mappings:
- ▶ A key difference between the eigendecomposition and the SVD is that in the SVD, domain and codomain can be of different dimensions
- ▶ In the SVD, the left and right singular vector matrices \mathbf{P} and \mathbf{P} are generally not inverse of each other.

Comparing SVD and EVD 3



- ▶ In the eigendecomposition, the matrices in decomposition are inverse of each other
- ▶ In the SVD, the entries in the diagonal matrix $\mathbf{\Sigma}$ are all real and nonnegative,
- ▶ In eigendecomposition diagonal matrix entries need not be real always.
- ▶ The leftsingular vectors of \mathbf{A} are eigenvectors of $\mathbf{A}\mathbf{A}^T$
- ▶ The rightsingular vectors of \mathbf{A} are eigenvectors of $\mathbf{A}^T\mathbf{A}$.



- ▶ We considered the SVD as a way to factorize $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ into the product of three matrices, where \mathbf{U} and \mathbf{V} are orthogonal and $\mathbf{\Sigma}$ contains the singular values on its main diagonal.
- ▶ Instead of doing the full SVD factorization, we will now investigate how the SVD allows us to represent a matrix \mathbf{A} as a sum of simpler matrices \mathbf{A}_i
- ▶ This representation which lends itself to a matrix approximation scheme that is cheaper to compute than the full SVD.



- ▶ A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ of rank r can be written as a sum of rank-1 matrices so that $\mathbf{A} = \sum_{i=1}^r \sigma_i u_i v_i^T$
- ▶ The diagonal structure of the singular value matrix $\mathbf{\Sigma}$ multiplies only matching left and right singular vectors $u_i v_i^T$ and scales them by the corresponding singular value σ_i .
- ▶ All terms $\sigma_i u_i v_i^T$ vanish for $i \neq j$ because $\mathbf{\Sigma}$ is a diagonal matrix.
- ▶ Any term for $i > r$ would vanish because the corresponding singular value is 0.



- ▶ We summed up the r individual rank-1 matrices to obtain a rank r matrix \mathbf{A} .
- ▶ If the sum does not run over all matrices A_i $i = 1, \dots, r$ but only up to an intermediate value k we obtain a rank- k approximation
- ▶ The approximation represented by $\hat{\mathbf{A}}(k)$ is defined as follows

$$\hat{\mathbf{A}}(k) = \sum_{i=1}^k \sigma_i u_i v_i^T$$

- ▶ To measure the difference between \mathbf{A} and its rank- k approximation we need the notion of a norm which is introduced next



- ▶ We introduce the notation of a subscript in the matrix norm
- ▶ Spectral Norm of a Matrix. For $x \in \mathbb{R}^n$, $x \neq \mathbf{0}$, the spectral norm of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as

$$\|\mathbf{A}\|_2 = \max_x \frac{\|\mathbf{A}x\|_2}{\|x\|_2}$$

where $\|y\|_2$ is the euclidean norm of y

- ▶ Theorem : The spectral norm of a matrix \mathbf{A} is its largest singular value

Example : Spectral Norm of a matrix



- ▶ Example : Consider the following matrix **A**

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

- ▶ Singular value decomposition of this matrix will provide the matrix Σ as follows

$$\Sigma = \begin{bmatrix} 5.465 & 0 \\ 0 & 0.366 \end{bmatrix}$$

- ▶ The 2 singular values are 5.4650 and 0.366.
- ▶ By definition the spectral norm is the largest singular value.
- ▶ Hence, the spectral norm is 5.4650