The University of Danang

University of Science and Technology

MATHEMATICS FOR COMPUTER SCIENCE

Chap 2. . Linear Algebra (cont.)

Faculty of Information Technology
PHAM Cong Thang, PhD

References

1. M. P. Deisenroth, A. A. Faisal, and C. S. Ong, Mathematics for Machine Learning, Cambridge University Press; 1 edition (April 23, 2020), 398 pages.
2. E. Lehman, F. T. Leighton, A. R. Meyer, 2017, Mathematics for Computer Science, Eric Lehman Google Inc, 998 pages
3. A. Laaksonen, Competitive Programmer's Handbook, 2018, 286 pages.
4. W. H. Press, S. A. Teukolsky, W.T. Vetterling, B. P. Flannery Numerical Recipes: The Art of Scientific Computing, Third Edition, Cambridge University Press, 1262 pages.
5. Other online/offline learning resources

Linear Algebra

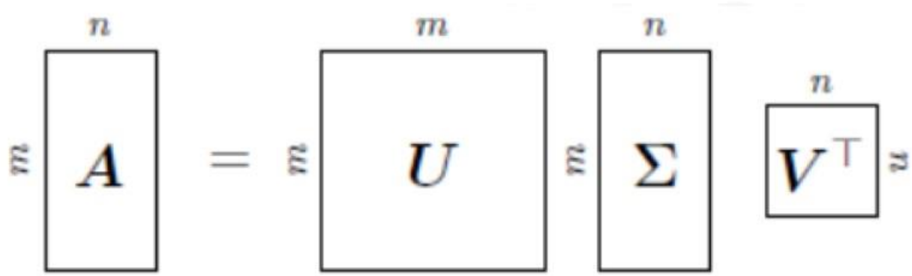
- Introduction
- Eigenvectors and finding Eigenvectors
- **Matrix Decompositions**

Matrix Decompositions - Singular Value Decomposition

- The singular value decomposition (SVD) of a matrix is a central matrix decomposition method in linear algebra.
  - It has been referred to as the "fundamental theorem of linear algebra" (Strang, 1993) because it can be applied to all matrices, not only to square matrices, and it always exists.

Matrix Decompositions - Singular Value Decomposition

- **SVD Theorem**
  - Let $A^{m \times n}$ be a rectangular matrix of rank $r \in \ [0; \min(m;\ n)]$
  - The SVD of A is a decomposition of the form



with an orthogonal matrix $U \in \mathbb{R}^{m \times m}$ with column vectors $u_i, i\ = 1, \dots, m$, and an orthogonal matrix $V \in \mathbb{R}^{n \times n}$with column vectors $v_j\ , j\ = 1, \dots, n$. Moreover, is an $m \times\ n$ matrix with $\Sigma_{ii}\ = \sigma_i >\ 0$ and $\Sigma_{ii} = 0, i \neq j$

Matrix Decompositions - Singular Value Decomposition

- The diagonal entries $\sigma_i, i = 1, \ldots, r$ of Σ are called the *singular values*, u_i are called the *left-singular vectors*, and v_j are called the *right-singular vectors*. By convention, the singular values are ordered, i.e., $\sigma_1 \geq \sigma_2 \geq \sigma_r \geq 0$
- The *singular value matrix* is unique, but it requires some attention. Observe that the $\Sigma \in \mathbb{R}^{m\times m}$ is rectangular. In particular, Σ is of the same size as A.
  - This means that has a diagonal submatrix that contains the singular values and needs additional zero padding.

Matrix Decompositions - Singular Value Decomposition

- The singular value matrix
  - If $m > n$, then the matrix Σ has diagonal structure up to row n and then consists of 0^T row vectors from $n + 1$ to m below so that:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

  - If $m < n$, the matrix has a diagonal structure up to column m and columns that consist of 0 from $m + 1$ to n:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 & & 0 \\ 0 & 0 & \sigma_m & 0 & \dots & 0 \end{bmatrix}.$$

  - The SVD exists for any matrix $A \in \mathbb{R}^{m \times n}$.

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - The SVD of a general matrix shares some similarities with the eigendecomposition of a square matrix
  - Compare the eigendecomposition of an SVD matrix $S = S^T = PDP^T$

  with the corresponding SVD: $S = U\Sigma V^T$
  - If we set $U = P = V, D = \Sigma$, we see that the SVD of SPD matrices is their eigendecomposition
  - Computing the SVD of $A \in \mathbb{R}^{m\times m}$ is equivalent to finding two sets of orthonormal bases $U = (u_1, \dots, u_m)$ and $V = (v_1, \dots, v_n)$ of the codomain $\mathbb{R}^m$ and the domain $\mathbb{R}^n$, respectively. From these ordered bases, we will construct the matrices U and V .

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - We start with constructing the orthonormal set of right-singular vectors $v_1, \dots, v_n \in \mathbb{R}^n$
  - Then, we then construct the orthonormal set of left-singular vectors $u_1, \dots, u_n \in \mathbb{R}^m$
  - Thereafter, we will link the two and require that the orthogonality of the v_i is preserved under the transformation of A
    - This is important because we know that the images Av_i form a set of orthogonal vectors. We will then normalize these images by scalar factors, which will turn out to be the singular values

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - a symmetric matrix possesses an orthonormal basis of eigenvectors, which also means it can be diagonalized. Moreover,
  - we can always construct a symmetric, positive semidefinite matrix A∈ $\mathbb{R}^{m\times n}$ from any rectangular matrix $S \in \mathbb{R}^{n\times n}$: $S = A^T A$
    - If $rank(A) = n, S = A^T A$ is symmetric, positive definite.
    - we can always diagonalize $A^T A$ and obtain

$$\boldsymbol{A}^{\top}\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\top} = \boldsymbol{P}\begin{bmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{bmatrix}\boldsymbol{P}^{\top}$$

where P is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geq 0$ are the eigenvalues of $A^T A$

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - Let us assume the SVD of A exists

$$\boldsymbol{A}^\top \boldsymbol{A} = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)^\top(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top) = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

where U, V are orthogonal matrices

  - With $U^T U = I$ we obtain

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}\boldsymbol{V}^\top$$

  - We identify

$$\boldsymbol{V}^\top = \boldsymbol{P}^\top,$$
$$\sigma_i^2 = \lambda_i .$$

with

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}\boldsymbol{P}^\top$$

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - Therefore, the eigenvectors of $A^T A$ that compose P are the right-singular vectors V of A
  - The eigenvalues of $A^T A$ are the squared singular values of Σ

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}\boldsymbol{V}^\top$$

$$\boldsymbol{V}^\top = \boldsymbol{P}^\top ,$$

$$\sigma_i^2 = \lambda_i .$$

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}\boldsymbol{P}^\top$$

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - To obtain the left-singular vectors U, we follow a similar procedure
  - We start by computing the SVD of the symmetric matrix $AA^T \in \mathbb{R}^{m\times m}$ (instead of the previous $AA^T \in \mathbb{R}^{n\times n}$). The SVD of A yields:

$$\boldsymbol{A}\boldsymbol{A}^\top = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)^\top = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top$$

$$= \boldsymbol{U}\begin{bmatrix}\sigma_1^2 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \sigma_m^2\end{bmatrix}\boldsymbol{U}^\top$$

  - AA^T can be diagonalized and we can find an orthogonal basis of eigenvectors of AA^T, which are collected in S. The orthonormal eigenvectors of AA^T are the left-singular vectors U and form an orthonormal basis set in the codomain of the SVD.
  - Since AA^T and AA^T have the same nonzero eigenvalues the nonzero entries of the Σ matrices in the SVD for both cases have to be the same.

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - The last step is to link up all the parts we touched upon so far. We have an orthonormal set of right-singular vectors in V .
  - To finish the construction of the SVD, we connect them with the orthonormal vectors U.
  - To reach this goal, we use the fact the images of the v_i under A have to be orthogonal, too.
  - We require that the inner product between Av_i and Av_j must be 0 for $i \neq j$.
    - For any two orthogonal eigenvectors v_i, v_j , $i \neq j$, it holds that

$$(\boldsymbol{A}\boldsymbol{v}_i)^\top(\boldsymbol{A}\boldsymbol{v}_j) = \boldsymbol{v}_i^\top(\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{v}_j = \boldsymbol{v}_i^\top(\lambda_j\boldsymbol{v}_j) = \lambda_j\boldsymbol{v}_i^\top\boldsymbol{v}_j = 0$$

    - For the case $m \geq r$, it holds that $\{Av_1, \dots, Av_r\}$is a basis of an $r-$ dimensional subspace of $\mathbb{R}^m$

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - To complete the SVD construction, we need left-singular vectors that are orthonormal: We normalize the images of the right-singular vectors Av_i and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i$$

    - the eigenvalues of AA^T are such that $\sigma_i^2 = \lambda_i$
  - Therefore, the eigenvectors of AA^T , which we know are the right-singular vectors v_i, and their normalized images under A, the left-singular vectors u_i , form two self-consistent Orthonormal bases that are connected through the singular value matrix Σ

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - We obtain the *singular value equation*

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i \qquad \boxed{\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i\,, \quad i = 1, \ldots, r}$$

  - This equation closely resembles the eigenvalue equation, but the vectors on the left- and the right-hand sides are not the same.
  - For $n > m$, the equation holds only for $i \leq m$ and says nothing about the u_i for $i > m$. However, we know by construction that they are orthonormal.
  - Conversely, for $m > n$, the equation holds only for $i \leq n$. For $i \geq n$, we have $Av_i = 0$ and we still know that the vi form an orthonormal set.
  - This means that the SVD also supplies an orthonormal basis of the kernel (null space) of A, the set of vectors x with $Ax = 0$

Matrix Decompositions - Singular Value Decomposition

- Construction of the SVD
  - Moreover, concatenating the v_i as the columns of V and the u_i as the columns of U yields

$$AV = U\Sigma$$

where Σ has the same dimensions as A and a diagonal structure for rows $1, \ldots, r$. Hence, right-multiplying with V^T yields $A = U\Sigma V^T >$, which is the SVD of A.

Matrix Decompositions - Singular Value Decomposition

Let us find the singular value decomposition of

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

The SVD requires us to compute the right-singular vectors v_j, the singular values σ_k, and the left-singular vectors u_i.

Step 1: Right-singular vectors as the eigenbasis of $A^\top A$.

We start by computing

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

We compute the singular values and right-singular vectors v_j through the eigenvalue decomposition of $A^\top A$, which is given as

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

and we obtain the right-singular vectors as the columns of P so that

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

Step 2: Singular-value matrix.

As the singular values σ_i are the square roots of the eigenvalues of $A^\top A$ we obtain them straight from D. Since $\text{rk}(A) = 2$, there are only two nonzero singular values: $\sigma_1 = \sqrt{6}$ and $\sigma_2 = 1$. The singular value matrix must be the same size as A, and we obtain

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

Step 3: Left-singular vectors as the normalized image of the right-singular vectors.

We find the left-singular vectors by computing the image of the right-singular vectors under A and normalizing them by dividing them by their corresponding singular value. We obtain

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} \\ \frac{-2}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix},$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{1} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix},$$

$$U = [u_1, u_2] = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}.$$

Note that on a computer the approach illustrated here has poor numerical behavior, and the SVD of A is normally computed without resorting to the eigenvalue decomposition of $A^\top A$.

Matrix Decompositions - Singular Value Decomposition

- Eigenvalue Decomposition vs. Singular Value Decomposition
  - Let us consider the eigendecomposition $A = PDP^{-1}$ and the SVD $A = U\Sigma V^T$ and review the core elements of the past sections.
    - The SVD always exists for any matrix $\mathbb{R}^{m\times n}$
    - The eigendecomposition is only defined for square matrices $\mathbb{R}^{n\times n}$ and only exists if we can find a basis of eigenvectors of $\mathbb{R}^n$

Matrix Decompositions - Singular Value Decomposition

- Eigenvalue Decomposition vs. Singular Value Decomposition
  - The vectors in the eigendecomposition matrix P are not necessarily orthogonal, i.e., the change of basis is not a simple rotation and scaling.
  - On the other hand, the vectors in the matrices U and V in the SVD are orthonormal, so they do represent rotations.
  - Both the eigendecomposition and the SVD are compositions of three linear mappings:
    - Change of basis in the domain
    - Independent scaling of each new basis vector and mapping from domain to codomain
    - Change of basis in the codomain

Matrix Decompositions - Singular Value Decomposition

- Eigenvalue Decomposition vs. Singular Value Decomposition
  - Example: Movie ratings of three people for four movies and its SVD

$$
\begin{array}{r c}
 & \begin{array}{ccc} \text{Ali} & \text{Beatrix} & \text{Chandra} \end{array} \\
\begin{array}{r} \text{Star Wars} \\ \text{Blade Runner} \\ \text{Amelie} \\ \text{Delicatessen} \end{array} &
\begin{bmatrix} 5 & 4 & 1 \\ 5 & 5 & 0 \\ 0 & 0 & 5 \\ 1 & 0 & 4 \end{bmatrix}
\end{array}
=
\begin{bmatrix}
-0.6710 & 0.0236 & 0.4647 & -0.5774 \\
-0.7197 & 0.2054 & -0.4759 & 0.4619 \\
-0.0939 & -0.7705 & -0.5268 & -0.3464 \\
-0.1515 & -0.6030 & 0.5293 & -0.5774
\end{bmatrix}
$$

$$
\begin{bmatrix}
9.6438 & 0 & 0 \\
0 & 6.3639 & 0 \\
0 & 0 & 0.7056 \\
0 & 0 & 0
\end{bmatrix}
$$

$$
\begin{bmatrix}
-0.7367 & -0.6515 & -0.1811 \\
0.0852 & 0.1762 & -0.9807 \\
0.6708 & -0.7379 & -0.0743
\end{bmatrix}
$$

Matrix Decompositions - Singular Value Decomposition

- Eigenvalue Decomposition vs. Singular Value Decomposition
  - A key difference between the eigendecomposition and the SVD :
    - In the SVD, domain and codomain can be vector spaces of different dimensions.
    - In the SVD, the left- and right-singular vector matrices U and V are generally not inverse of each other (they perform basis changes in different vector spaces).
    - In the eigendecomposition, the basis change matrices P and P^{-1} are inverses of each other.
    - In the SVD, the entries in the diagonal matrix Σ are all real and nonnegative, which is not generally true for the diagonal matrix in the eigendecomposition.

Matrix Decompositions - Singular Value Decomposition

- Eigenvalue Decomposition vs. Singular Value Decomposition
  - A key difference between the eigendecomposition and the SVD :
    - The SVD and the eigendecomposition are closely related through their projections
      - The left-singular vectors of A are eigenvectors of AA^T
      - The right-singular vectors of A are eigenvectors of AA^T.
      - The nonzero singular values of A are the square roots of the nonzero eigenvalues of AA^T and are equal to the nonzero eigenvalues of AA^T
    - For symmetric matrices $A \in \mathbb{R}^{n\times n}$, the eigenvalue decomposition and the SVD are one and the same

Matrix Decompositions - Singular Value Decomposition

- It is possible to define the SVD of a $rank-r$ matrix A so that U is an $m \times r$ matrix, Σ a diagonal matrix $r \times r$, and V an $r \times n$ matrix.
  - This construction is very similar to our definition, and ensures that the diagonal matrix Σ has only nonzero entries along the diagonal. The main convenience of this alternative notation is that Σ is diagonal, as in the eigenvalue decomposition.
- A restriction that the SVD for A only applies to $m \times n$ matrices with $m > n$ is practically unnecessary.
  - When $m < n$, the SVD decomposition will yield Σ with more zero columns than rows and, consequently, the singular values $\sigma_{m+1}, \ldots, \sigma_n$ are 0.

Matrix Decompositions - Singular Value Decomposition

- The SVD is used in a variety of applications in machine learning from least-squares problems in curve fitting to solving systems of linear equations
  - These applications harness various important properties of the SVD, its relation to the rank of a matrix, and its ability to approximate matrices of a given rank with lower-rank matrices.
  - Substituting a matrix with its SVD has often the advantage of making calculation more robust to numerical rounding errors.
  - The SVD's ability to approximate matrices with "simpler" matrices in a principled manner opens up machine learning applications ranging from dimensionality reduction and topic modeling to data compression and clustering.