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

Eigenvalues and Eigenvectors

- Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. Then $\lambda \in \mathbb{R}$ is an eigenvalue of A and $x \in \mathbb{R}^n\{0\}$ is the corresponding eigenvector of A if

$$Ax = \lambda x$$

This is the eigenvalue equation, x is called an **eigenvector** of A corresponding to λ

$$Ax = \underbrace{\lambda}_{\text{eigenvalue}} \underbrace{x}_{\text{eigenvector}}$$

Eigenvalues and Eigenvectors

- Computing Eigenvalues, Eigenvectors
  - **Step 1**: **Characteristic Polynomial.** definition of the eigenvector $x \neq 0$ and eigenvalue λ of A, there will be a vector such that $Ax = \lambda x$, i.e.,

$$(A - \lambda I)x = 0$$

  - **Step 2: Eigenvalues.** We need to compute the roots of the characteristic polynomial to find the eigenvalues.
  - **Step 3**: **Eigenvectors**. We find the eigenvectors that correspond to these eigenvalues by looking at vectors x

Eigenvalues and Eigenvectors

- For $A \in \mathbb{R}^{n\times n}$, the set of all eigenvectors of A associated with an eigenvalue spans a subspace eigenspace of $\mathbb{R}^n$, which is called the *eigenspace* of A with respect to λ and is denoted by E_λ . The set of all eigenvalues of A is called the *eigenspectrum*, or just spectrum *spectrum*, of A.
  - If λ is an eigenvalue of $A \in \mathbb{R}^{n\times n}$, then the corresponding eigenspace E_λ is the solution space of the homogeneous system of linear equations

$$(A - \lambda I)x = 0$$

Eigenspaces

- Suppose A is an $n \times n$ matrix and λ is an eigenvalue of A:
  - $E_\lambda(A) = \{x\colon Ax = \lambda x\}$ is an subspace of $\mathbb{R}^n$

$$A = \begin{bmatrix} -3 & -1 \\ 0 & 2 \end{bmatrix} \Rightarrow c_A(x) = \det(xI - A) = \begin{vmatrix} x+3 & 1 \\ 0 & x-2 \end{vmatrix} = (x+3)(x-2)$$

$$c_A(x) = 0 \Leftrightarrow x = -3 \vee x = 2$$

$$x = -3: \left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & -5 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \Rightarrow X = \begin{bmatrix} t \\ 0 \end{bmatrix} \ (\text{or } X = (t,0))$$

$$x = 2: \left[\begin{array}{cc|c} 5 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \Rightarrow X = \begin{bmatrix} t \\ -5t \end{bmatrix}$$

$$E_{-3} = \{X : AX = -3X\} = \{(t,0) : t \in \mathbb{R}\}$$

$$E_2 = \{X : AX = 2X\} = \{(t,-5t) : t \in \mathbb{R}\}$$

Subspace of $\mathbb{R}^n$

- Definition of subspace of $\mathbb{R}^n$.
  - Let $U \neq \emptyset$ be a subset of $\mathbb{R}^n$
  - U is called a subspace of $\mathbb{R}^n$ if:
    - The zero vector 0 is in U
    - If X, Y are in U then $X + Y$ is in U
    - If X is in U then aX is in U for all real number a.

Subspace of $\mathbb{R}^n$

- Definition of subspace of $\mathbb{R}^n$.
  - Ex1. U={(a,a,0)|$a \in \mathbb{R}$} is a subspace of $\mathbb{R}^3$:
    - the zero vector of R3, (0,0,0) ∈ U
    - (a,a,0), (b,b,0) ∈ U⇒ (a,a,0)+(b,b,0)=(a+b,a+b,0) ∈ U
    - If (a,a,0) ∈ U and k ∈ $\mathbb{R}$, then k(a,a,0)=(ka,ka,0) ∈ U
  - Ex2. U={(a,b,1): a,b ∈ R} is not a subspace of $\mathbb{R}^3$: (0,0,0) ∉ U ⇒ U is not a subspace
  - Ex3. U={(a,|a|,0)|a∈R} is not a subspace of $\mathbb{R}^3$: (-1,|-1|,0), (1,|1|,0)∈U but (0,2,0) ∉ U ⇒ U is not a subspace

Subspace of $\mathbb{R}^n$

- A subspace either has only one or infinite many vectors
  - Example, {0} has only vector
  - If a subspace U has nonzero vector X then aX is also in U. Then U has infinite many vector

Spanning sets

- $\boldsymbol{y} = \boldsymbol{k_1 x_1} + \boldsymbol{k_2 x_2} + \cdots + \boldsymbol{k_n x_n}$ is called a linear combination of the vectors $x_1, x_2, \dots, x_n$
- The set of all linear combinations of the the vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}$ is called the span of these vectors, denoted by ***span***$\{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$

$$\boldsymbol{span}\{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\} = \{\boldsymbol{k_1 x_1} + \boldsymbol{k_2 x_2} + \cdots + \boldsymbol{k_n x_n} : \boldsymbol{k_i} \in \mathbb{R} \text{ is arbitrary}\}$$

- $U = Span\{X1, X2, \dots, Xn\}$ is a subspace of $\mathbb{R}^n$. If W is a subspace of $\mathbb{R}^n$ such that x_i are in W then $U \subseteq W$
- For example, span{(1,0,1),(0,1,1)}={a(1,0,1)+b(0,1,1) :a,b $\in \mathbb{R}$}.
- And we have (1,2,3) $\in$ span{(1,0,1),(0,1,1)} because (1,2,3)= 1(1,0,1)+ 2(0,1,1).
- (2,3,2) $\notin$ span{(1,0,1),(0,1,1)} because (2,3,2)≠a(1,0,1)+b(0,1,1) for all a,b

Eigenvalues and Eigenvectors

- **Computing Eigenvalues, Eigenvectors, and Eigenspaces**

Let us find the eigenvalues and eigenvectors of the 2×2 matrix

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

Step 1: Characteristic Polynomial. From our definition of the eigenvector $\boldsymbol{x} \neq \boldsymbol{0}$ and eigenvalue λ of $\boldsymbol{A}$, there will be a vector such that $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, i.e., $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$. Since $\boldsymbol{x} \neq \boldsymbol{0}$, this requires that the kernel (null space) of $\boldsymbol{A} - \lambda\boldsymbol{I}$ contains more elements than just $\boldsymbol{0}$. This means that $\boldsymbol{A} - \lambda\boldsymbol{I}$ is not invertible and therefore $\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$. Hence, we need to compute the roots of the characteristic polynomial (4.22a) to find the eigenvalues.

Step 2: Eigenvalues. The characteristic polynomial is

$$\begin{aligned} p_{\boldsymbol{A}}(\lambda) &= \det(\boldsymbol{A} - \lambda\boldsymbol{I}) \\ &= \det\left(\begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = \begin{vmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{vmatrix} \\ &= (4-\lambda)(3-\lambda) - 2 \cdot 1. \end{aligned}$$

We factorize the characteristic polynomial and obtain

$$p(\lambda) = (4-\lambda)(3-\lambda) - 2 \cdot 1 = 10 - 7\lambda + \lambda^2 = (2-\lambda)(5-\lambda)$$

giving the roots $\lambda_1 = 2$ and $\lambda_2 = 5$.

Step 3: Eigenvectors and Eigenspaces. We find the eigenvectors that correspond to these eigenvalues by looking at vectors $\boldsymbol{x}$ such that

$$\begin{bmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}.$$

For $\lambda = 5$ we obtain

$$\begin{bmatrix} 4-5 & 2 \\ 1 & 3-5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{0}.$$

We solve this homogeneous system and obtain a solution space

$$E_5 = \operatorname{span}[\begin{bmatrix} 2 \\ 1 \end{bmatrix}].$$

This eigenspace is one-dimensional as it possesses a single basis vector.

Analogously, we find the eigenvector for $\lambda = 2$ by solving the homogeneous system of equations

$$\begin{bmatrix} 4-2 & 2 \\ 1 & 3-2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}.$$

Linear Algebra

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

Matrix Decompositions

- There are many ways to factorize special types of matrices that we encounter often in machine learning.
  - In the positive real numbers, we have the square-root operation that gives us a decomposition of the number into identical components, e.g., $9 = 3 \times 3$.
  - For matrices, we need to be careful that we compute a square-root-like operation on positive quantities.
    - For symmetric, positive definite matrices we can Cholesky choose from a number of square-root equivalent operations. The *Cholesky* decomposition *decomposition/Cholesky factorization* provides a square-root equivalent operation on symmetric, positive definite matrices that is useful in practice.

Matrix Decompositions-Cholesky Decomposition

- A symmetric, positive definite matrix A can be factorized into a product $A = LL^T$, where L is a lower-triangular matrix with positive diagonal elements:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix}$$

L is called the Cholesky factor of A, and L is unique.

Matrix Decompositions-Cholesky Decomposition

- Consider a symmetric, positive definite matrix $A \in \mathbb{R}^{3\times 3}$. We are interested in finding its Cholesky factorization $A = LL^T$, i.e.,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}^\top = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

  - Multiplying out the right-hand side yields

$$\boldsymbol{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Matrix Decompositions-Cholesky Decomposition

- Comparing the left-hand side and the right-hand side shows that there is a simple pattern in the diagonal elements l_{ii}:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

$$l_{11} = \sqrt{a_{11}}\,, \quad l_{22} = \sqrt{a_{22} - l_{21}^2}\,, \quad l_{33} = \sqrt{a_{33} - (l_{31}^2 + l_{32}^2)}$$

$$l_{21} = \frac{1}{l_{11}} a_{21}\,, \quad l_{31} = \frac{1}{l_{11}} a_{31}\,, \quad l_{32} = \frac{1}{l_{22}} (a_{32} - l_{31} l_{21})$$

Matrix Decompositions-Cholesky Decomposition

- The Cholesky decomposition is an important tool for the numerical computations underlying machine learning.
- Symmetric positive definite matrices require frequent manipulation, e.g., the covariance matrix of a multivariate Gaussian variable is symmetric, positive definite.

Matrix Decompositions-Cholesky Decomposition

- The Cholesky factorization of this covariance matrix allows us to generate samples from a Gaussian distribution. It also allows us to perform a linear transformation of random variables, which is heavily exploited when computing gradients in deep stochastic models, such as the variational auto-encoder
- The Cholesky decomposition also allows us to compute determinants very efficiently.
  - Given the Cholesky decomposition $A = LL^T$, we know that $\det(A) = \det(L)\det(L^T) = \det(L)^2$.
  - Since L is a triangular matrix, the determinant is simply the product of its diagonal entries so that $\det(A) = \prod_i l_{ii}^2$. Thus, many numerical software packages use the Cholesky decomposition to make computations more efficient.

Matrix Decompositions-Eigendecomposition and Diagonalization

- A diagonal matrix is a matrix that has value zero on all off-diagonal elements, i.e., they are of the form

$$\boldsymbol{D} = \begin{bmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_n \end{bmatrix}$$

- They allow fast computation of determinants, powers, and inverses. The determinant is the product of its diagonal entries, a matrix power D^k is given by each diagonal element raised to the power k, and the inverse D^{-1} is the reciprocal of its diagonal elements if all of them are nonzero.
  - we will discuss how to transform matrices into diagonal form.

Matrix Decompositions-Eigendecomposition and Diagonalization

- Two matrices A, D are similar if there exists an invertible matrix P, such that

$D = P^{-1}AP.$

  - More specifically, we will look at matrices A that are similar to diagonal matrices D that contain the eigenvalues of A on the diagonal

- **Diagonalizable.** A matrix $A \in \mathbb{R}^{n\times n}$ is *diagonalizable* if it is similar to a diagonal matrix, i.e., if there exists an invertible matrix $P \in \mathbb{R}^{n\times n}$ such that $D = P^{-1}AP$.

Matrix Decompositions-Eigendecomposition and Diagonalization

- **Diagonalizable.**

  - Let $A \in \mathbb{R}^{n \times n}$, let $\lambda_1, \dots, \lambda_n$ be a set of scalars, and let $p_1, \dots, p_n$ be a set of vectors in $\mathbb{R}^n$. We define $P = [p_1, \dots p_n]$ and let $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$. Then we can show that

$$AP = PD$$

  if and only if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A and $p_1, \dots p_n$ are corresponding eigenvectors of A

Matrix Decompositions-Eigendecomposition and Diagonalization

- **Diagonalizable.**
  - We can see that this statement holds because

$$AP = A[p_1, \dots, p_n] = [Ap_1, \dots, Ap_n]$$

$$PD = [p_1, \dots, p_n]\begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} = [\lambda_1 p_1, \dots, \lambda_n p_n]$$

  - Thus $Ap_1 = \lambda p_1, Ap_2 = \lambda p_2, \dots, Ap_n = \lambda p_n$

$\Leftrightarrow$ **the columns of P must be eigenvectors of A.**

Matrix Decompositions-Eigendecomposition and Diagonalization

- **Eigendecomposition.** A square matrix $A \in \mathbb{R}^{n\times n}$ can be factored into

$$A = PDP^{-1}$$

where $P \in \mathbb{R}^{n\times n}$ and D is a diagonal matrix whose diagonal entries are the eigenvalues of A, if and only if the eigenvectors of A form a basis of $\mathbb{R}^n$

  - Only non-defective matrices can be diagonalized and that the columns of P are the n eigenvectors of A.
  - For symmetric matrices we can obtain even stronger outcomes for the eigenvalue decomposition.
  - A symmetric matrix $S \in \mathbb{R}^{n\times n}$ can always be diagonalized

Matrix Decompositions-Eigendecomposition and Diagonalization

Diagonalize the matrix $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & -1 \\ 1 & 3 & -2 \end{bmatrix}$

Solution. By Example 3.3.4, the eigenvalues of A are $\lambda_1 = 2$, $\lambda_2 = 1$, and $\lambda_3 = -1$, with corresponding basic eigenvectors $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$ respectively. Since the matrix $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$ is invertible, Theorem 3.3.4 guarantees that

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = D$$

The reader can verify this directly—easier to check $AP = PD$.

Matrix Decompositions-Eigendecomposition and Diagonalization

Diagonalize the matrix $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

Solution. To compute the characteristic polynomial of A first add rows 2 and 3 of $xI - A$ to row 1:

$$c_A(x) = \det \begin{bmatrix} x & -1 & -1 \\ -1 & x & -1 \\ -1 & -1 & x \end{bmatrix} = \det \begin{bmatrix} x-2 & x-2 & x-2 \\ -1 & x & -1 \\ -1 & -1 & x \end{bmatrix}$$
$$= \det \begin{bmatrix} x-2 & 0 & 0 \\ -1 & x+1 & 0 \\ -1 & 0 & x+1 \end{bmatrix} = (x-2)(x+1)^2$$

Hence the eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = -1$, with λ_2 repeated twice (we say that λ_2 has *multiplicity* two). However, A *is* diagonalizable. For $\lambda_1 = 2$, the system of equations $(\lambda_1 I - A)\mathbf{x} = \mathbf{0}$ has general solution $\mathbf{x} = t\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as the reader can verify, so a basic λ_1-eigenvector is $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Turning to the repeated eigenvalue $\lambda_2 = -1$, we must solve $(\lambda_2 I - A)\mathbf{x} = \mathbf{0}$. By gaussian elimination, the general solution is $\mathbf{x} = s\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ where s and t are arbitrary. Hence the gaussian algorithm produces *two* basic λ_2-eigenvectors $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{y}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ If we take $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ we find that P is invertible. Hence $P^{-1}AP = \text{diag}(2, -1, -1)$

Matrix Decompositions-Eigendecomposition and Diagonalization

- If A is an $n \times n$ matrix and P is an invertible $n \times n$ matrix then

$$A = PDP^{-1}$$
$$A^2 = PD^2P^{-1}$$
$$A^k = PD^kP^{-1}$$

- Assume that the eigendecomposition $A = PDP^{-1}$ exists. Then

$$\det(A) = \det(PDP^{-1}) = \det(P)\det(D)\det(P^{-1}) = \det(D) = \prod_i d_{ii}$$

allows for an efficient computation of the determinant of A.

Matrix Decompositions-Eigendecomposition and Diagonalization

Let us compute the eigendecomposition of $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

Step 1: Compute eigenvalues and eigenvectors. The characteristic polynomial of A is

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}\right) = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1).$$

Therefore, the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 3$ (the roots of the characteristic polynomial), and the associated (normalized) eigenvectors are obtained via

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} p_1 = 1 p_1, \quad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} p_2 = 3 p_2.$$

This yields

$$p_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad p_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Step 2: Check for existence. The eigenvectors p_1, p_2 form a basis of $\mathbb{R}^2$. Therefore, A can be diagonalized.

Step 3: Construct the matrix P to diagonalize A. We collect the eigenvectors of A in P so that

$$P = [p_1, p_2] = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

We then obtain

$$P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} = D.$$

Equivalently, we get (exploiting that $P^{-1} = P^\top$ since the eigenvectors p_1 and p_2 in this example form an ONB)

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}}_{A} = \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}}_{D} \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_{P^\top}.$$