Introduction

This course is based on the following mathematical objects:

- **Vectors**: 1-D arrays;
- **Matrices**: 2-D arrays
- **Tensors**: 3-D arrays (or higher)
Notation: vectors

Vectors are denoted by **boldface** lowercase letters (such as $a, b, u, v, x, y$):

$$a = (1, 2, 3)^T = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3$$

The $i$th element of a vector $a$ is written as $a_i$ or $a(i)$.

For any $p \geq 1$, the $\ell_p$ norm, or simply $p$-norm, of a vector $a$ is

$$\|a\|_p = \left( \sum |a_i|^p \right)^{1/p}.$$  

In the case of $p = 2$ (default), the norm is called the **Euclidean norm**.
Basic Matrix Algebra

Some special vectors:

- **The zero vector**
  \[ 0_n = (0, 0, \ldots, 0)^T \in \mathbb{R}^n \]

- **The vector of ones**
  \[ 1_n = (1, 1, \ldots, 1)^T \in \mathbb{R}^n \]

- **The canonical basis vectors of** \( \mathbb{R}^n \)
  \[ e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T \in \mathbb{R}^n, \quad i = 1, \ldots, n \]

When the dimension is not specified, it is implied by the context, e.g.,

\[ \mathbf{a} \cdot 1 \] (dot product), \quad where \quad \mathbf{a} = (1, 2, 3)^T
**Notation: matrices**

Matrices are denoted by **boldface** UPPERCASE letters (such as $A$, $B$, $U$, $V$). We write $A \in \mathbb{R}^{m \times n}$ to indicate that $A$ has $m$ rows and $n$ columns.

The $(i, j)$ entry of $A$ is denoted by $a_{ij}$, or $A(i, j)$, or $A_{ij}$.

The $i$th row of $A$ is denoted by $A_i$ or $A(i,:)$.

The $j$th column is written as $a_j$ or $A(:,j)$.

\[ A = \begin{bmatrix} \ldots & a_{ij} & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix} \]

\[ A_i \]

\[ a_j \]
Basic Matrix Algebra

Special matrices:

- **The zero matrix**: \( O_n \in \mathbb{R}^{n \times n} \)

- **The identity matrix**: \( I_n \in \mathbb{R}^{n \times n} \)

- **The matrix of ones**: \( J_n \in \mathbb{R}^{n \times n} \)

\[
O_n = \begin{pmatrix}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 
\end{pmatrix}, \quad
I_n = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 1 
\end{pmatrix}, \quad
J_n = \begin{pmatrix}
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \cdots & 1 
\end{pmatrix}
\]

Similarly, we may drop the subscript \( n \) when the size of the matrix is clear based on the context.
Notation: tensors

Tensors are multidimensional arrays that generalize vectors and matrices.

We use calligraphic UPPERCASE letters to denote them and write $\mathcal{T} \in \mathbb{R}^{c \times r \times l}$ to indicate the size of a 3D tensor $\mathcal{T}$.

Tensor algebra is a big, interesting filed on its own, but we will only use 3D tensors to store information for simple and efficient coding which requires knowing a little bit of how to unfold a 3D tensor to a matrix (and to assemble the tensor back).
$\mathcal{T} \in \mathbb{R}^{c \times r \times \ell}$

$M \in \mathbb{R}^{c \times (r\ell)}$

$\mathcal{T}(\cdot, \cdot, 1) \quad \cdots \quad \mathcal{T}(\cdot, \cdot, \ell)$
Descriptions of a matrix

One way to characterize a matrix $A \in \mathbb{R}^{m \times n}$ is based on its shape:

- **Square matrix**
  - $m = n$

- **Long matrix**
  - $m < n$

- **Tall matrix**
  - $m > n$
Basic Matrix Algebra

A matrix $A \in \mathbb{R}^{m \times n}$ is said to be positive/nonnegative if all of its entries are positive ($a_{ij} > 0$, $\forall i, j$) / nonnegative ($a_{ij} \geq 0$, $\forall i, j$).

If a matrix has mostly zero entries, then we say that the matrix is **sparse** and often leave the zero entries blank when writing it out.

A **diagonal** matrix is a square matrix $A \in \mathbb{R}^{n \times n}$ whose off diagonal entries are all zero ($a_{ij} = 0$ for all $i \neq j$), e.g.,

$$A = \begin{pmatrix} 1 & \phantom{1}2 & \phantom{1}3 \\ \phantom{1}2 & \phantom{1} \phantom{1} & \phantom{1} \\ \phantom{1}3 & \phantom{1} \phantom{1} & \phantom{1} \end{pmatrix} = \text{diag}(1, 2, 3)$$

Sometimes, a rectangular matrix $A \in \mathbb{R}^{m \times n}$ is also said to be diagonal if $a_{ij} = 0$ for all $i \neq j$. 
Matrix multiplication

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. Their matrix product is an $m \times p$ matrix

$$AB = C = (c_{ij}), \quad c_{ij} = A_i b_j = \sum_{k=1}^{n} a_{ik} b_{kj}.$$
It is possible to obtain one full row (or column) of $C$ at a time via matrix-vector multiplication:

$$C_i = A_i B, \quad c_j = A b_j$$
Basic Matrix Algebra

The full matrix $C$ can be written as a sum of rank-1 matrices:

$$C = \sum_{k=1}^{n} a_k B_k.$$
Further interpretation of $AB$ when one of the matrices is actually a vector:

- If $A = (a_1, \ldots, a_n)$ is a row vector ($m = 1$):

\[
AB = \sum_{i=1}^{n} a_i B_i \quad \leftarrow \text{linear combination of rows of } B
\]
• If $\mathbf{B} = (b_1, \ldots, b_n)^T$ is a column vector ($p = 1$):

$$\mathbf{AB} = \sum_{j=1}^{n} b_j \mathbf{a}_j \quad \leftarrow \text{linear combination of columns of } \mathbf{A}$$
Basic Matrix Algebra

Finally, below are some identities involving the vector $\mathbf{1} \in \mathbb{R}^n$:

$$\mathbf{1}^T \mathbf{1} = n, \quad \mathbf{1} \mathbf{1}^T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} = \mathbf{J}_n$$

For any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$,

$$\mathbf{A} \mathbf{1} = \sum_j a_{j}, \quad \text{(vector of row sums)}$$

$$\mathbf{1}^T \mathbf{A} = \sum_i A_i, \quad \text{(horizontal vector of column sums)}$$

$$\mathbf{1}^T \mathbf{A} \mathbf{1} = \sum_i \sum_j a_{ij}, \quad \text{(total sum of all entries)}$$
Graphical illustration

$$(1, 1, 1, 1)$$

$$\begin{pmatrix}
1 \\
1 \\
1 \\
1
\end{pmatrix}$$
The Hadamard product

A different way to multiply two matrices of the same size, $A, B \in \mathbb{R}^{m \times n}$, is through the **Hadamard product**, also called the **entrywise product**:

$$C = A \circ B \in \mathbb{R}^{m \times n}, \quad \text{with} \quad c_{ij} = a_{ij}b_{ij}.$$ 

For example,

$$\begin{pmatrix} 0 & 2 & -3 \\ -1 & 0 & -4 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & -3 \\ 2 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 9 \\ -2 & 0 & 4 \end{pmatrix}.$$

Hadamard products are very useful in computing, as we shall see.
Matrix algebra

Use the instructor’s notes to review the following matrix operations:

- Transpose
- Rank
- Trace*
- Determinant*
- Inverse*

*Defined only for square matrices
Characterization of rank-1 matrices

Rank-1 matrices, though very simple, are quite useful in computing.

- Any *nonzero* row or column vector (as a matrix) has rank 1.

- A *nonzero* matrix is of rank 1 if and only if all of its nonzero rows (or columns) are multiples of each other.

- A *nonzero* matrix $A \in \mathbb{R}^{m \times n}$ has rank 1 if and only if there exist nonzero vectors $u \in \mathbb{R}^m, v \in \mathbb{R}^n$, such that $A = uv^T$. 
Basic Matrix Algebra

- For any square rank-1 matrix $A = uv^T \in \mathbb{R}^{n \times n}$ where $u, v \in \mathbb{R}^n$,

$$\text{trace}(A) = \text{trace}(uv^T) = \text{trace}(v^T u) = v^T u = \sum_{i=1}^{n} u_i v_i.$$

For example, the following is a rank-1 matrix:

$$\begin{pmatrix} 2 & 0 & 3 \\ 4 & 0 & 6 \\ 6 & 0 & 9 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 2 & 0 & 3 \end{pmatrix}$$

Its rows are multiples of each other, and so are its 2 nonzero columns.

Another example of rank-1 matrices is $J_n = 11^T$. 
Eigenvalues and eigenvectors

Let $A \in \mathbb{R}^{n \times n}$. The characteristic polynomial of $A$ is

$$p(\lambda) = \det(A - \lambda I).$$

We define the eigenvalues of $A$ as the real roots of the characteristic equation $p(\lambda) = 0$ (and will never work with complex eigenvalues).

For a specific eigenvalue $\lambda_0$, any nonzero vector $v_0 \in \mathbb{R}^n$ satisfying

$$(A - \lambda_0 I)v_0 = 0 \iff Av_0 = \lambda_0 v_0$$

is called an eigenvector of $A$ (associated to the eigenvalue $\lambda_0$).
Basic Matrix Algebra

All eigenvectors of $A$ associated to an eigenvalue $\lambda_0$ span a linear subspace, called the **eigenspace** of $A$ corresponding to $\lambda_0$:

$$E(\lambda_0) = \{ v \in \mathbb{R}^n \mid (A - \lambda_0I)v = 0 \} = \text{Nul}(A - \lambda_0I).$$

The dimension $g_0$ of $E(\lambda_0)$ is called the **geometric multiplicity** of $\lambda_0$, while the degree $a_0$ of the highest-order factor $(\lambda - \lambda_i)^{a_0}$ in $p(\lambda)$ is called the **algebraic multiplicity** of $\lambda_0$.

Note that for any matrix $A \in \mathbb{R}^{n \times n}$ with $k$ distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, we have

- $\sum_{i=1}^{k} a_i \leq n$, and
- $1 \leq g_i \leq a_i$, for each $1 \leq i \leq k$. 
Example 0.1. For the matrix \( A = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix} \), find its eigenvalues and associated eigenvectors.

**Answer.** The eigenvalues are \( \lambda_1 = 3, \lambda_2 = 2 \) with algebraic multiplicities \( a_1 = 2, a_2 = 1 \), and geometric multiplicities \( g_1 = g_2 = 1 \). In fact, \( E(\lambda_1) = \text{span}\{(0,1,-2)^T\} \) and \( E(\lambda_2) = \text{span}\{(0,1,-1)^T\} \).
The following theorem indicates that for any $n \times n$ matrix with $n$ eigenvalues, all of its rank, trace, and determinant can be computed from the eigenvalues of the matrix.

**Theorem 0.1.** For any $A \in \mathbb{R}^{n \times n}$ with $n$ eigenvalues, $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$,

\[
\begin{align*}
\text{rank}(A) &= \sum_{i=1}^{n} 1_{\lambda_i \neq 0} \\
\text{trace}(A) &= \sum_{i=1}^{n} \lambda_i \\
\text{det}(A) &= \prod_{i=1}^{n} \lambda_i
\end{align*}
\]
Similar matrices

Two square matrices of the same size $A, B \in \mathbb{R}^{n \times n}$ are said to be similar if there exists an invertible matrix $P \in \mathbb{R}^{n \times n}$ such that

$$B = PAP^{-1}$$

Similar matrices have the same

- rank, trace, determinant
- characteristic polynomial
- eigenvalues and their multiplicities (but not eigenvectors)
Diagonalizability of square matrices

**Def 0.1.** A square matrix $A \in \mathbb{R}^{n \times n}$ is **diagonalizable** if it is similar to a diagonal matrix, i.e., there exist an invertible matrix $P \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ such that

$$A = P\Lambda P^{-1}, \quad \text{or equivalently,} \quad P^{-1}AP = \Lambda.$$  

*Remark.* If we write $P = [p_1, \ldots, p_n]$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, then the above equation can be rewritten as

$$AP = PA\Lambda,$$
or in columns

\[
A[p_1 \ldots p_n] = [p_1 \ldots p_n] \begin{bmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{bmatrix}.
\]

From this we get that

\[
Ap_i = \lambda_ip_i, \ 1 \leq i \leq n.
\]

This shows that \( A \) has \( n \) eigenvalues \( \lambda_1, \ldots, \lambda_n \in \mathbb{R} \) (not necessarily distinct) with corresponding eigenvectors \( p_1, \ldots, p_n \in \mathbb{R}^n \).

Thus, the above factorization of a diagonalizable matrix \( A \) is called the **eigendecomposition**, or **spectral decomposition**, of \( A \).
Example 0.2. The matrix

$$A = \begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix}$$

is diagonalizable because

$$\begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 3 & -1 \end{pmatrix}^{-1}$$

but the matrix

$$B = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$$

is not (we will see why later).
Why are diagonalizable matrices important?

Every diagonalizable matrix is similar to a diagonal matrix (that consists of its eigenvalues), and is easy to deal with in a lot of ways.

For example, it can help compute matrix powers ($A^k$). To see this, suppose $A \in \mathbb{R}^{n \times n}$ is diagonalizable, that is, $A = P \Lambda P^{-1}$ for some invertible matrix $P$ and a diagonal matrix $\Lambda$. Then

\[
A^2 = P \Lambda P^{-1} \cdot P \Lambda P^{-1} = P \Lambda^2 P^{-1}
\]

\[
A^3 = P \Lambda P^{-1} \cdot P \Lambda P^{-1} \cdot P \Lambda P^{-1} = P \Lambda^3 P^{-1}
\]

\[
A^k = P \Lambda^k P^{-1} \quad \text{(for any positive integer $k$)}
\]

where $\Lambda^k = \text{diag}(\lambda_1^k, \ldots, \lambda_n^k)$. 
Checking diagonalizability of a square matrix

**Theorem 0.2.** A matrix $A \in \mathbb{R}^{n \times n}$ is diagonalizable if and only if it has $n$ linearly independent eigenvectors (i.e., $\sum g_i = n$).

**Proof.**

\[
A = P\Lambda P^{-1} \iff AP = P\Lambda \iff Ap_i = \lambda_i p_i, 1 \leq i \leq n
\]

The $p_i$’s are linearly independent if and only if $P$ is nonsingular.

**Remark.** A diagonalizable matrix $A \in \mathbb{R}^{n \times n}$ must have $n$ eigenvalues. Additionally, for the distinct eigenvalues, we must have $a_i = g_i$, because

\[
n = \sum g_i \leq \sum a_i = n
\]
Example 0.3. The matrix $B = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$ is not diagonalizable because it has only one distinct eigenvalue $\lambda_1 = 1$ with $a_1 = 2$ and $g_1 = 1$. 
Two special classes of square matrices are always diagonalizable:

- **Idempotent** matrices:

\[ I^n(\mathbb{R}) = \{ A \in \mathbb{R}^{n \times n} \mid A^2 = A \} \]

For example,

\[
\begin{pmatrix}
1 & 0 \\
1 & 0
\end{pmatrix},
\begin{pmatrix}
3 & -6 \\
1 & -2
\end{pmatrix},
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix},
\begin{pmatrix}
2 & -2 & -4 \\
-1 & 3 & 4 \\
1 & -2 & -3
\end{pmatrix}
\]

- **Symmetric** matrices:

\[ S^n(\mathbb{R}) = \{ A \in \mathbb{R}^{n \times n} \mid A^T = A \} \]
**Idempotent matrices**

The following are some exemplar idempotent matrices:

\[ O, \quad I, \quad \frac{1}{n} J_n, \quad \text{and} \quad C_n = I_n - \frac{1}{n} J_n = I_n - \frac{1}{n} 1_n 1_n^T. \]

Note that \( J_n \) alone is not idempotent, because \( J_n^2 = n J_n \).

To see why \( C_n \) is idempotent:

\[
C_n^2 = \left( I_n - \frac{1}{n} J_n \right) \left( I_n - \frac{1}{n} J_n \right) \\
= I_n - \frac{1}{n} J_n - \frac{1}{n} J_n + \frac{1}{n^2} J_n^2 \\
= C_n.
\]
Important fact: $C_n$ is a centering matrix.

For any point $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$,

$$C_n x = \left( I_n - \frac{1}{n} 1 1^T \right) x$$

$$= x - \frac{1}{n} 1 (1^T x)$$

$$= x - 1 \bar{x}$$

$$= [x_1 - \bar{x}, \ldots, x_n - \bar{x}]^T$$

where

$$\bar{x} = \frac{1}{n} 1^T x = \frac{1}{n} \sum_{i=1}^{n} x_i.$$
An important result is the following. A proof based on minimal polynomials can be found in [Horn and Johnson, matrix analysis, 2nd ed].

**Theorem 0.3.** Every idempotent matrix $A \in I^n(\mathbb{R})$ is diagonalizable, i.e., there exist an invertible matrix $P \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ such that $A = P \Lambda P^{-1}$.

Additionally, idempotent matrices can only have eigenvalues 0 or 1 or both. To see this, suppose $Av = \lambda v$. Using $A = A^2$, we then get

$$\lambda v = Av = (A^2)v = A(Av) = A(\lambda v) = \lambda (Av) = \lambda (\lambda v) = \lambda^2 v.$$  

Since $v \neq 0$, we must have $\lambda = \lambda^2$ and thus $\lambda = 0$ or 1.
For any $A \in I_n(\mathbb{R})$, let $a_0$ and $a_1$ be the algebraic multiplicities of the eigenvalues 0 and 1.

Because $A$ is diagonalizable, we must have $a_0 + a_1 = n$, and

$$\text{trace}(A) = a_1 = \text{rank}(A).$$

Consider the following cases:

- $a_0 = n, a_1 = 0$: $A = \mathbf{0}$;
- $a_0 = 0, a_1 = n$: $A = \mathbf{I}$ (the only nonsingular matrix in $I_n(\mathbb{R})$);
- $1 \leq a_0, a_1 \leq n - 1$: All other idempotent matrices
Example 0.4. Since $\frac{1}{n} J_n \in \mathbb{R}^{n \times n}$ is idempotent and

$$\text{rank} \left( \frac{1}{n} J_n \right) = 1 = \text{trace} \left( \frac{1}{n} J_n \right),$$

it has an eigenvalue of 1 with algebraic multiplicity $a_1 = 1$, and the other eigenvalue is 0 with $a_0 = n - 1$.

This implies that $J_n$ has eigenvalues $n$ and 0 with algebraic multiplicities 1, $n - 1$ respectively:

$$\frac{1}{n} J_n \cdot \mathbf{v} = \lambda \cdot \mathbf{v} \iff J_n \cdot \mathbf{v} = n \lambda \cdot \mathbf{v}.$$
Example 0.5. Consider the centering matrix $C_n = I_n - \frac{1}{n}J_n$. Since

$$\text{trace}(C_n) = \text{trace}(I_n) - \frac{1}{n} \text{trace}(J_n) = n - \frac{1}{n} \cdot n = n - 1.$$ 

we conclude that

- $a_0 = 1$ and $a_1 = n - 1$.

- $\text{rank}(C_n) = n - 1$ and $\det(C_n) = 0$. 
Furthermore, the unique eigenvalue 0 has a corresponding eigenvector 1, because

\[ C_n 1 = \left( I_n - \frac{1}{n} 11^T \right) 1 = I_n 1 - \frac{1}{n} 11^T 1 = 1 - 1 = 0 = 0 \cdot 1, \]

Another interpretation is that all the rows of \( C_n \) sum to zero (and because of the symmetry of \( C_n \), all its columns sum to zero as well):

\[
C_1 = (0), \quad C_2 = \begin{pmatrix}
\frac{1}{2} & -\frac{1}{2} \\
\frac{1}{2} & -\frac{1}{2}
\end{pmatrix}, \quad C_3 = \begin{pmatrix}
\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\
-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\
-\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}
\end{pmatrix}
\]
Symmetric matrices

Symmetric matrices have many nice properties. For example, they are not only diagonalizable, but also can be diagonalized via orthogonal matrices.

Theorem 0.4 (The Spectral Theorem). Let $A \in S^n(\mathbb{R})$. Then there exist an orthogonal matrix $Q = [q_1 \ldots q_n] \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, such that

$$A = Q\Lambda Q^T$$

(we say that $A$ is orthogonally diagonalizable)

Remark. The above factorization also represents a spectral decomposition of $A$: The $\lambda_i$’s represent eigenvalues of $A$ while the $q_i$’s are the associated eigenvectors (with unit norm and orthogonal to each other).
Remark. One can rewrite the matrix decomposition

\[ A = Q \Lambda Q^T \]

into a sum of rank-1 matrices:

\[
A = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} = \sum_{i=1}^{n} \lambda_i q_i q_i^T
\]

For convenience, the diagonal elements of \( \Lambda \) are often sorted in decreasing order (and the columns of \( Q \) should be arranged in matching order):

\[ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \]
Example 0.6. Find the spectral decomposition of the following matrix

\[ A = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} \]

Answer.

\[ A = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ -1 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^T \]

\[ = 4 \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} + (-1) \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \]
Quadratic forms

Symmetric matrices can be used to define the so-called quadratic forms.

**Def 0.2.** Let \( A \in \mathbb{R}^{n \times n} \) be a symmetric matrix. A **quadratic form** based on \( A \) is a function \( Q : \mathbb{R}^n \mapsto \mathbb{R} \) with

\[
Q(x) = x^T Ax, \quad \text{for all} \quad x \in \mathbb{R}^n.
\]

**Remark.** A quadratic form is a second-order polynomial in the components of \( x \) without linear or constant terms:

\[
x^T Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{i<j} a_{ij} x_i x_j
\]
Basic Matrix Algebra

For example, if \( A = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix} \) and \( x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \), then

\[
Q(x) = x^T A x = x_1^2 + 2x_2^2 + 6x_1x_2
\]

**Question:** Which symmetric matrix corresponds to

\[
Q(x) = x_1^2 + 2x_2^2 + 3x_3^2 + 6x_1x_2 - 4x_1x_3 + 10x_2x_3
\]
Basic Matrix Algebra

Positive (semi)definite matrices

A symmetric matrix $A \in S^n(\mathbb{R})$ is said to be positive semidefinite (PSD) if the corresponding quadratic form $Q(x) = x^T A x \geq 0$ for all $x \in \mathbb{R}^n$. If the equality holds true only for $x = 0$ (i.e., $x^T A x > 0$ for all $x \neq 0$), then $A$ is further said to be positive definite (PD).

We denote by $S^{n}_{0+}(\mathbb{R})$ and $S^{n}_{+}(\mathbb{R})$ the sets of positive semidefinite and of positive definite matrices of size $n \times n$, respectively.

Note that we must have

$$S^n_{+}(\mathbb{R}) \subset S^n_{0+}(\mathbb{R}) \subset S^n(\mathbb{R}) \subset \mathbb{R}^{n \times n}.$$
**Theorem 0.5.** A symmetric matrix is **positive definite** (**positive semidefinite**) if and only if all of its eigenvalues are strictly **positive** (**nonnegative**).

**Remark.** There is a quick way of determining the positive (semi)definiteness of a $2 \times 2$ nonzero matrix $A$:

- $A \in S_+^2(\mathbb{R})$ if and only if $\det(A) > 0$ and $\text{trace}(A) > 0$;

- $A \in S^2_{0+}(\mathbb{R})$ if and only if $\det(A) = 0$ and $\text{trace}(A) > 0$.

This is due to $\det(A) = \lambda_1 \lambda_2$ and $\text{trace}(A) = \lambda_1 + \lambda_2$. 
Example 0.7. Determine the positive definiteness of each of the following matrices:

\[
\begin{pmatrix}
1 & 2 \\
2 & 4
\end{pmatrix}, \quad
\begin{pmatrix}
2 & 3 \\
3 & 2
\end{pmatrix}, \quad
\begin{pmatrix}
2 & 3 \\
3 & 5
\end{pmatrix}
\]
Spectral decomposition of PSD matrices in reduced form

The preceding theorem implies that for a PSD matrix $\mathbf{A} \in S_{0+}^n(\mathbb{R})$,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n, \quad r = \text{rank}(\mathbf{A}).$$

Correspondingly, we may obtain

$$\mathbf{A} = \sum_{i=1}^{r} \lambda_i \mathbf{q}_i \mathbf{q}_i^T = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_r \end{bmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_r \end{pmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_r \end{bmatrix} = \mathbf{Q}_r \Lambda_r \mathbf{Q}_r^T$$

where $\mathbf{Q}_r = [\mathbf{q}_1 \ldots \mathbf{q}_r] \in \mathbb{R}^{n \times r}$ is a tall matrix with orthonormal columns, and $\Lambda_r = \text{diag}(\lambda_1, \ldots, \lambda_r) \in \mathbb{R}^{r \times r}$. 

Dr. Guangliang Chen | Mathematics & Statistics, San José State University
Example 0.8. Let \( A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \in S^2_{0+}(\mathbb{R}) \), which has a rank of \( r = 1 \).

The spectral decomposition is

\[
\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}
\]

\[
= \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 5 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}
\]
A small, useful result

**Theorem 0.6.** For any rectangular matrix $A \in \mathbb{R}^{m \times n}$, both $AA^T \in \mathbb{R}^{m \times m}$ and $A^TA \in \mathbb{R}^{n \times n}$ are square, symmetric, and positive semidefinite.

**Proof.** It is obvious that $A^TA$ is square ($n \times n$) and symmetric:

$$(A^TA)^T = A^T(A^T)^T = A^TA$$

To show that it is PSD, for any $x \in \mathbb{R}^n$, write

$$x^T(A^TA)x = (x^TA^T)(Ax) = (Ax)^T(Ax) = \|Ax\|^2 \geq 0.$$ 

The proof for the other product $AA^T$ is similar. 

□
Remark. For any rectangular matrix $A \in \mathbb{R}^{m \times n}$, it is always true that
\[
\text{rank}(AA^T) = \text{rank}(A^TA) = \text{rank}(A).
\]

Another notable result about the two product matrices, $AA^T$ and $A^TA$, is that they must have the same nonzero eigenvalues (with the same algebraic multiplicities), due to the following identity:
\[
\lambda^n \det(\lambda I_m - AA^T) = \lambda^m \det(\lambda I_n - A^TA).
\]
However, the zero eigenvalue may have different algebraic multiplicities.
Matrix square roots

An interesting aspect of positive semidefinite matrices is that they have square roots (which are also matrices), just like nonnegative numbers have square roots (which are still numbers).

**Def 0.3.** Let \( A \in S_{0+}^n(\mathbb{R}) \). The **square root** of \( A \) is defined as the matrix \( R \in S_{0+}^n(\mathbb{R}) \) such that \( R^2 = A \). We denote it by \( R = A^{1/2} \).

Note that if \( A \in S_+^n(\mathbb{R}) \), then \( R = A^{1/2} \in S_+^n(\mathbb{R}) \) because

\[
0 \neq \det(A) = \det(R^2) = \det(R)^2 \quad \rightarrow \quad \det(R) \neq 0.
\]

In such a case, we can further define the **reciprocal square root** of \( A \) as \( A^{-1/2} = (A^{1/2})^{-1} \in S_+^n(\mathbb{R}) \).
Special case: If $A \in S_{0+}^n(\mathbb{R})$ happens to be diagonal, i.e.,

$$A = \text{diag}(a_1, \ldots, a_n), \quad \text{where} \quad a_1, \ldots, a_n \geq 0,$$

then there is an easy way to find its square root. Define

$$R = \text{diag} \left( a_1^{1/2}, \ldots, a_n^{1/2} \right) \in S_{0+}^n(\mathbb{R}).$$

Clearly, $R^2 = A$. This shows that $R$ is indeed a square root of $A$.

Note that without the positive semidefiniteness requirement in the definition of matrix square roots, it won’t be unique as we can arbitrarily modify the signs of the diagonals $a_i^{1/2}$ without violating the equality condition.
Theorem 0.7. Let $\mathbf{A} \in S_{0+}^n(\mathbb{R})$ with spectral decomposition $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Then $\mathbf{A}$ has a unique matrix square root

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T.$$ 

Proof. First, such defined matrix $\mathbf{R}$ is PSD. By direct calculation,

$$\mathbf{R}^2 = (\mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T)(\mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T) = \mathbf{Q}\underbrace{\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}}_{\mathbf{\Lambda}}\mathbf{Q}^T = \mathbf{A}.$$ 

This shows that $\mathbf{R}$ is a square root of $\mathbf{A}$. We skip the proof of the uniqueness part in class but refer the interested reader to the notes.
Remark. Using the reduced form of the eigendecomposition of $A \in S_{0+}^n(\mathbb{R})$, we obtain the following reduced form for the square root of $A$:

$$A = Q_r \Lambda_r Q_r^T \quad \longrightarrow \quad A^{1/2} = Q_r \Lambda_r^{1/2} Q_r^T.$$ 

This formula is more efficient for computing the matrix square roots, as it only requires computing the eigenvectors corresponding to the positive eigenvalues.

Remark. For any $A \in S_+^n(\mathbb{R})$ with eigendecomposition $A = Q \Lambda Q^T$, the reciprocal square root is

$$A^{-1/2} = Q \Lambda^{-1/2} Q^T, \quad \Lambda^{-1/2} = \text{diag} \left( \lambda_1^{-1/2}, \ldots, \lambda_n^{-1/2} \right).$$
Example 0.9. Let \( A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \in S^2_{0+}(\mathbb{R}) \), which has two nonnegative eigenvalues \( \lambda_1 = 5, \lambda_2 = 0 \). To find the matrix square root of \( A \), we need to first find its orthogonal diagonalization (in reduced form):

\[
\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 5 \\ \frac{1}{\sqrt{5}} \end{pmatrix}
\]

It follows that

\[
A^{1/2} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{5} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \end{pmatrix}
\]
The generalized eigenvalue problem

Let $A, B \in \mathbb{R}^{n \times n}$ be two square matrices of the same size. We say that $\lambda \in \mathbb{R}$ is a generalized eigenvalue of $(A, B)$ if there exists a nonzero vector $v \in \mathbb{R}^n$ such that

$$Av = \lambda Bv.$$ 

The vector $v$ is called a generalized eigenvector of $(A, B)$ corresponding to $\lambda$.

Remark. In the above definition, if we let $B = I$, then the generalized eigenvalues of $(A, B)$ would reduce to the ordinary eigenvalues of $A$:

$$Av = \lambda v.$$
Now, let us rewrite the definition as

$$(A - \lambda B)v = 0.$$ 

Note that there exists a nonzero solution $v$ if and only if $A - \lambda B$ is singular. Thus, $\lambda$ is a generalized eigenvalue of $(A, B)$ if and only if

$$\det(A - \lambda B) = 0.$$ 

Let $p_{A,B}(\lambda) = \det(A - \lambda B)$, the characteristic polynomial of $(A, B)$. Interestingly, $p_{A,B}(\lambda)$ is also a polynomial in $\lambda$, but it can have an arbitrary order between 0 and $n$, as we show next.
Example 0.10. Let

\[ A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \]

To find the generalized eigenvalues of \((A, B)\), compute

\[
\det(A - \lambda B) = \begin{vmatrix} 1 - \lambda & 2 - \lambda \\ 2 - \lambda & 4 - \lambda \end{vmatrix} = (1 - \lambda)(4 - \lambda) - (2 - \lambda)^2 = -\lambda.
\]

Thus, \((A, B)\) has a generalized eigenvalue of \(\lambda = 0\), with corresponding generalized eigenvectors

\[
0 = (A - 0 \cdot B)v = Av \quad \rightarrow \quad v = k \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \ k \in \mathbb{R}.
\]
Example 0.11. Let

\[ A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}. \]

To find the generalized eigenvalues of \((A, B)\), compute

\[
\det(A - \lambda B) = \begin{vmatrix} 1 - \lambda & 2 - 2\lambda \\ 2 - 3\lambda & 4 - 6\lambda \end{vmatrix} = (1 - \lambda)(4 - 6\lambda) - (2 - 2\lambda)(2 - 3\lambda) = 0.
\]

Thus, any scalar \(\lambda\) is a generalized eigenvalue of \((A, B)\). This pair of matrices has infinitely many generalized eigenvalues!
For an arbitrary generalized eigenvalue \( \lambda \in \mathbb{R} \), we find its corresponding generalized eigenvector as follows:

\[
0 = (A - \lambda \cdot B)v = \begin{pmatrix} 1 - \lambda & 2 - 2\lambda \\ 2 - 3\lambda & 4 - 6\lambda \end{pmatrix}v \quad \rightarrow \quad v = k \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \quad k \in \mathbb{R}.
\]

This indicates that all the generalized eigenvalues share the same generalized eigenvector!
Generalized symmetric-definite eigenvalue problems

Let $A \in S^n(\mathbb{R})$ and $B \in S^n_+(\mathbb{R})$. The generalized eigenvalue problem

$$Av = \lambda Bv$$

is called a generalized symmetric-definite eigenvalue problem. Such problems have very nice properties and have a lot of applications.

**Theorem 0.8.** The above generalized symmetric-definite eigenvalue problem has $n$ generalized eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ with linearly independent generalized eigenvectors $v_1, \ldots, v_n \in \mathbb{R}^n$ which can be normalized such that

$$v_i^T B v_j = \delta_{ij}, \quad \text{for all } 1 \leq i, j \leq n.$$
Remark. We derive a few more results from the theorem. Let $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $V = [v_1, \ldots, v_n] \in \mathbb{R}^{n \times n}$. Then

$$v_i^T Bv_j = \delta_{ij} \implies V^T B V = I$$

Next, using $A v_i = \lambda_i B v_i$, $1 \leq i \leq n$, we have

$$A [v_1, \ldots, v_n] = [B v_1, \ldots, B v_n] \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \rightarrow A V = B V \Lambda.$$

Lastly, $V$ also diagonalizes $A$:

$$V^T A V = V^T (A V) = V^T (B V \Lambda) = (V^T B V) \Lambda = I \Lambda = \Lambda.$$
Proof of the theorem. Since $B \in S^+_n(\mathbb{R})$, we can rewrite

$$Av = \lambda Bv \quad \implies \quad B^{-1/2}AB^{-1/2} \cdot B^{1/2}v = \lambda \cdot B^{1/2}v$$

Letting

$$\tilde{A} = B^{-1/2}AB^{-1/2}, \quad \text{and} \quad \tilde{v} = B^{1/2}v$$

we further obtain that

$$\tilde{A}\tilde{v} = \lambda \tilde{v}$$

Since $\tilde{A} \in S^n(\mathbb{R})$, there are $n$ eigenpairs $(\lambda_i, \tilde{v}_i), 1 \leq i \leq n$, with

$$\delta_{ij} = \tilde{v}_i^T\tilde{v}_j = \left(B^{1/2}v_i\right)^T B^{1/2}v_j = v_i^T B v_j.$$

Consequently, $(A, B)$ has $n$ generalized eigenvalues $\lambda_i$ with associated generalized eigenvectors $v_i = B^{-1/2}\tilde{v}_i$. 
Some observations:

- The generalized eigenvalues of \((A, B)\), for \(A \in S^n(\mathbb{R}), B \in S_+^n(\mathbb{R})\), are identical to the eigenvalues of \(\tilde{A} = B^{-1/2}AB^{-1/2} \in S^n(\mathbb{R})\).

- The generalized eigenvectors of \((A, B)\) are \(v_i = B^{-1/2}\tilde{v}_i\), where \(\tilde{v}_i\) are the unit-norm eigenvectors of \(\tilde{A}\).

Furthermore, the generalized eigenvalues/eigenvectors of \((A, B)\) coincide with the eigenvalues/eigenvectors of \(B^{-1}A\):

\[
Av = \lambda Bv \iff B^{-1}Av = \lambda v
\]

In fact, \(\tilde{A} = B^{-1/2}AB^{-1/2}\) and \(B^{-1}A\) are two similar matrices:

\[
B^{-1/2}AB^{-1/2} = B^{1/2} \cdot B^{-1}A \cdot B^{-1/2}
\]
Example 0.12. Let $A = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix} \in S^2(\mathbb{R})$ and $B = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} \in S^2_+(\mathbb{R})$. Find the generalized eigenvalues and eigenvectors of $(A, B)$. 
Next time: Matrix Computing in MATLAB

Be sure to complete the following activities before next class:

- Install MATLAB on your computer
- Take the 2-hour MATLAB Onramp tutorial¹
- Explore the Statistics and Machine Learning Toolbox²

¹https://www.mathworks.com/learn/tutorials/matlab-onramp.html