Linear Algebra (Review)

Ruwen Qin
Stony Brook University

August 21, 2025

Python notebook used in this module:

- Linear_Algebra.ipynb

Linear algebra is essential for understanding and working with many machine learning and optimization algorithms. We review the key linear algebra prerequisites in this class. This review is not exhaustive but just provides some prior knowledge for studying this course. This review is mainly based on Chapter 2 in [1].

Scalars, Vectors, Matrices, and Tensors

Scalars

A *scalar* is a single number. We specify what kind of numbers they are in modeling. For example, $a \in \mathbb{R}$ is the slope of a line, and $T \in \mathbb{N}$ is the total number of periods for a planning decision.

Vectors

A *vector* is an array of numbers. The numbers are arranged in order, and we can identify each individual number by its index in that ordering. Typically, we give vectors lowercase names in bold typeface, such as $\boldsymbol{x}$. We also need to say what kind of numbers are stored in the vector. For example, $\boldsymbol{x} \in \mathbb{R}^n$ means each element of $\boldsymbol{x}$ is in $\mathbb{R}$ and the space for $\boldsymbol{x}$ is the Cartesian product of $\mathbb{R}$ for n times. When we explicitly identify the elements of a vector, we write the vector as a column enclosed in square brackets:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1}$$

A vector identifies a point in space. The elements of the vector are the coordinates of the corresponding point in the space.

Matrices

A *matrix* is a 2-D array of numbers, so each element is identified by two indices. We usually give matrices uppercase variable names with bold typeface, such as $\boldsymbol{A}$. If a real-valued matrix $\boldsymbol{A}$ has a height of m and a width of n, then we say that $\boldsymbol{A} \in \mathbb{R}^{m \times n}$:

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_{1,\cdot} \\ \boldsymbol{a}_{2,\cdot} \\ \vdots \\ \boldsymbol{a}_{m,\cdot} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_{\cdot,1} & \boldsymbol{a}_{\cdot,2} & \dots & \boldsymbol{a}_{\cdot,n} \end{bmatrix} \tag{2}$$

Matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ can be seen as m row vectors in order:

$$\boldsymbol{a}_{i,\cdot} = \left[\begin{array}{cccc} a_{i,1} & a_{i,2} & \dots & a_{i,n} \end{array} \right] \tag{3}$$

for $i = 1, 2, \dots, m$, and n column vectors in order:

$$\boldsymbol{a}_{\cdot,j} = \left[\begin{array}{c} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{array} \right] \tag{4}$$

for $j = 1, 2, \dots, n$.

Tensors

A tensor is a 3-D array of numbers. It can be seen as matrices in order and the depth of the order is l. $\mathsf{A} \in \mathbb{R}^{l \times m \times n}$ and $a_{i,j,k}$ identity the element at coordinates (i,j,k) of the tensor. Matrices, vectors, and scalars are special cases of tensors.

Vector and Matrix Operations

Transpose

The transpose of a column vector turns the vector into a row vector, and vice versa. That is, given

$$\boldsymbol{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array} \right], \tag{5}$$

then

$$\boldsymbol{x}^{\mathrm{T}} = \left[\begin{array}{cccc} x_1 & x_2 & \dots & x_m \end{array} \right]. \tag{6}$$

The transpose of a matrix is the mirror image of the matrix across a diagonal line. Given a matrix $\boldsymbol{A}$,

$$\boldsymbol{A} = \left[\begin{array}{cccc} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{array} \right], \tag{7}$$

The transpose of $\boldsymbol{A}$, denoted as $\boldsymbol{A}^{\mathrm{T}}$, is

$$\boldsymbol{A}^{\mathrm{T}} = \left[\begin{array}{cccc} a_{1,1} & a_{2,1} & \dots & a_{m,1} \\ a_{1,2} & a_{2,2} & \dots & a_{m,2} \\ \vdots & \vdots & \dots & \vdots \\ a_{1,n} & a_{2,n} & \dots & a_{m,n} \end{array} \right]. \tag{8}$$

That is, $a^{\mathrm{T}}_{i,j} = a_{j,i}$.

A *symmetric matrix* is any matrix that is equal to its own transpose:

$$\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}. \tag{9}$$

Addition

Addition (and subtraction) operations can be applied to matrices that are of the same dimension. For example, given $\boldsymbol{A}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$, the addition of these two matrices is

$$\boldsymbol{A}+\boldsymbol{B} = \begin{bmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & \dots & a_{1,n}+b_{1,n} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & \dots & a_{2,n}+b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}+b_{m,1} & a_{m,2}+b_{m,2} & \dots & a_{m,n}+b_{m,n} \end{bmatrix}. \tag{10}$$

We can add a scalar to a matrix or multiply a matrix by a scalar by applying the operation to each element of the matrix:

$$\alpha\cdot\boldsymbol{B}+\gamma = \begin{bmatrix} \alpha\cdot b_{1,1}+\gamma & \alpha\cdot b_{1,2}+\gamma & \dots & \alpha\cdot b_{1,n}+\gamma \\ \alpha\cdot b_{2,1}+\gamma & \alpha\cdot b_{2,2}+\gamma & \dots & \alpha\cdot b_{2,n}+\gamma \\ \vdots & \vdots & \dots & \vdots \\ \alpha\cdot b_{m,1}+\gamma & \alpha\cdot b_{m,2}+\gamma & \dots & \alpha\cdot b_{m,n}+\gamma \end{bmatrix} \tag{11}$$

We can also add a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ with a vector $\boldsymbol{c} \in \mathbb{R}^{m\times 1}$:

$$\boldsymbol{A}+\boldsymbol{c} = \begin{bmatrix} \boldsymbol{a}_{1,\cdot} \\ \boldsymbol{a}_{2,\cdot} \\ \vdots \\ \boldsymbol{a}_{m,\cdot} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} a_{1,1}+b_1 & a_{1,2}+b_1 & \dots & a_{1,n}+b_1 \\ a_{2,1}+b_2 & a_{2,2}+b_2 & \dots & a_{2,n}+b_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}+b_m & a_{m,2}+b_m & \dots & a_{m,n}+b_m \end{bmatrix} \tag{12}$$

Adding elements of vector $\boldsymbol{c}$ to each row of matrix $\boldsymbol{A}$ is called *broadcasting*.

Inner (Dot) Product of Vectors

The *dot product* or *inner product* of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in the same dimension is the summation of the element-wise product of vector elements:

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \sum_i x_i\, y_i. \tag{13}$$

Matrix multiplication to be reviewed below is based on the inner product of vectors.

Matrix Multiplication

The *Hadamard product* of two matrices that are of the same dimension is the element-wise product of the matrices. For example, given $\boldsymbol{A}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$, the Hadamard product of them is

$$\boldsymbol{A}\odot\boldsymbol{B} = \begin{bmatrix} a_{1,1}\,b_{1,1} & a_{1,2}\,b_{1,2} & \dots & a_{1,n}\,b_{1,n} \\ a_{2,1}\,b_{2,1} & a_{2,2}\,b_{2,2} & \dots & a_{2,n}\,b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}\,b_{m,1} & a_{m,2}\,b_{m,2} & \dots & a_{m,n}\,b_{m,n} \end{bmatrix} \tag{14}$$

The multiplication of two matrices $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times k}$ is a third matrix $\boldsymbol{C} \in \mathbb{R}^{m\times k}$:

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{a}_{1,\cdot}\,\boldsymbol{b}_{\cdot,1} & \boldsymbol{a}_{1,\cdot}\,\boldsymbol{b}_{\cdot,2} & \dots & \boldsymbol{a}_{1,\cdot}\,\boldsymbol{b}_{\cdot,n} \\ \boldsymbol{a}_{2,\cdot}\,\boldsymbol{b}_{\cdot,1} & \boldsymbol{a}_{2,\cdot}\,\boldsymbol{b}_{\cdot,2} & \dots & \boldsymbol{a}_{2,\cdot}\,\boldsymbol{b}_{\cdot,n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_{m,\cdot}\,\boldsymbol{b}_{\cdot,1} & \boldsymbol{a}_{m,\cdot}\,\boldsymbol{b}_{\cdot,2} & \dots & \boldsymbol{a}_{m,\cdot}\,\boldsymbol{b}_{\cdot,n} \end{bmatrix}. \tag{15}$$

That is, $(\boldsymbol{AB})_{i,j}$ is the inner product of the ith row of $\boldsymbol{A}$ and the jth column of $\boldsymbol{B}$:

$$(\boldsymbol{AB})_{i,j} = \boldsymbol{a}_{i,\cdot}\,\boldsymbol{b}_{\cdot,j}. \tag{16}$$

Inverse

An *identity matrix* is a square matrix whose diagonal values are all ones and other elements are all zero:

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{17}$$

An identity matrix is one that does not change any vector when we multiply that vector by the identify matrix:

$$\boldsymbol{I}\,\boldsymbol{x} = \boldsymbol{x}. \tag{18}$$

The inverse of matrix $\boldsymbol{A}$, denoted by $\boldsymbol{A}^{-1}$, is define as the matrix such that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}. \tag{19}$$

When the inverse of the matrix $\boldsymbol{A}$ exists, the linear system,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{20}$$

can be solved as

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{21}$$

A matrix $\boldsymbol{A}$ is a *diagonal matrix* if $a_{i,j} = 0$ for all $i \neq j$:

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{n,n} \end{bmatrix}. \tag{22}$$

The inversion of the diagonal matrix $\boldsymbol{A}$ is obtained easily:

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 1/a_{1,1} & 0 & \dots & 0 \\ 0 & 1/a_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/a_{n,n} \end{bmatrix}. \tag{23}$$

Trace

The trace of a matrix $\boldsymbol{A}$ gives the sum of all the diagonal entries:

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i a_{i,i}. \tag{24}$$

Trace operator is invariant to the transpose operator because it does not change diagnal entries:

$$\mathrm{Tr}(\boldsymbol{A}) = \mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}). \tag{25}$$

If both $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are valid matrix multiplication, their traces are equal:

$$\mathrm{Tr}(\boldsymbol{AB}) = \mathrm{Tr}(\boldsymbol{BA}). \tag{26}$$

Matrix Properties

Matrices have some important properties:

- Distributive property:
$$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC} \tag{27}$$
- Associative property:
$$\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C} \tag{28}$$
- The transpose of matrix multiplication:
$$(\boldsymbol{AB})^{\mathrm{T}} = (\boldsymbol{B})^{\mathrm{T}}(\boldsymbol{A})^{\mathrm{T}} \tag{29}$$
- Inverse of matrix multiplication:
$$(\boldsymbol{AB})^{-1} = (\boldsymbol{B})^{-1}(\boldsymbol{A})^{-1} \tag{30}$$

Norms

Vector Norms

Norms measure vector's sizes. Formally, the l_p-norm of a vector $\boldsymbol{x}$ is

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{1/p} \tag{31}$$

where $p \in \mathbb{R}$, and $p \geq 1$.

- $m = 1$ (Manhattan distance): l_1-norm
$$\|\boldsymbol{x}\|_1 = \sum_i |x|_i \tag{32}$$
- $m = 2$ (Euclidean distance): l_2-norm:
$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_i |x_i|^2} \tag{33}$$
- $m = \infty$: l_∞-norm:
$$\|\boldsymbol{x}\|_\infty = \max_i |x|_i \tag{34}$$

The inner product of two vectors, $\boldsymbol{x}$ and $\boldsymbol{y}$, can be rewritten in terms of their norms

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \cos\theta \tag{35}$$

where θ is the angle between the two vectors. If the inner product is zero, the two vectors are *orthogonal*.

Matrix Norms

Here we skip the detail of matrix norms. But it is helpful to know the most commonly used norm for matrices, named Frobenius norm:

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j} |a_{i,j}|^2}. \tag{36}$$

The Frobenius norm of matrix $\boldsymbol{A}$ is equal to the square root of the trace of $\boldsymbol{AA}^{\mathrm{T}}$:

$$\|\boldsymbol{A}\|_F = \sqrt{\mathrm{Tr}(\boldsymbol{AA}^{\mathrm{T}})}. \tag{37}$$

Linear Independence and Span

To determine if a linear system $\boldsymbol{Ax} = \boldsymbol{b}$ has a solution, we will see if $\boldsymbol{b}$ is in the *span* of $\boldsymbol{A}$'s columns. The span of $\boldsymbol{A}$'s columns is a linear combination of the columns,

$$\boldsymbol{Ax} = \sum_j x_j \, \boldsymbol{a}_{\cdot,j}. \tag{38}$$

It is also named *column space* or *range* of $\boldsymbol{A}$.

A set of vectors is *linearly independent* if no vector in the set is a linear combination of the other vectors. If $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{A}$ must contain at least one set of m linearly independent columns in order to encompass all of $\mathbb{R}^m$. This condition is both necessary and sufficient for the linear system to have a solution for every $\boldsymbol{b}$.

Decomposition

Orthogonal Matrix

A vector $\boldsymbol{x}$ and a vector $\boldsymbol{y}$ are orthogonal to each other if

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0. \tag{39}$$

A vector $\boldsymbol{x}$ is a unit vector if

$$\|\boldsymbol{x}\|_2 = 1. \tag{40}$$

If vectors are not only orthogonal but also have unit norm, we call them *orthonormal*.

An *orthogonal matrix* is a square matrix whose rows are mutually orthonormal and whose columns are mutually orthonormal:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}. \tag{41}$$

Therefore, if a matrix is an orthonormal matrix, then

$$\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}^{-1}. \tag{42}$$

Orthnormal matrices are of interest because their inverse is very cheap to compute.

Eigendecomposition

Eigendecomposition is one of the most widely used kinds of matrix decomposition, which decomposes a matrix into a set of eigenvectors and eigenvalues.

An *eigenvector* of a square matrix $\boldsymbol{A}$ is a nonzero vector $\boldsymbol{v}$ such that the multiplication by $\boldsymbol{A}$ alters only the scale of $\boldsymbol{v}$:

$$\boldsymbol{Av} = \lambda\boldsymbol{v}, \tag{43}$$

where the scalar λ is the *eigenvalue* corresponding to this eigenvector.

The eigendecomposition of $\boldsymbol{A}$ is then given by

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} \tag{44}$$

where the matrix $\boldsymbol{V}$ concatenates all the linearly independent eigenvectors of $\boldsymbol{A}$,

$$\boldsymbol{V} = [\ \boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \ldots, \boldsymbol{v}_n\ ], \tag{45}$$

and $\boldsymbol{\lambda}$ is a diagonal matrix that concatenates the eigenvalues

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}. \tag{46}$$

Constructing matrices with specific eigenvalues and eigenvectors enables us to stretch space in desired direction.

Every real symmetric square matrix can be decomposed into an expression using only real-valued eigenvectors and eigenvalues:

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}, \tag{47}$$

where $\boldsymbol{Q}$ is orthogonal matrix composed of eigenvectors of $\boldsymbol{A}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix with λ_i being the eigenvalue associated with the eigenvector $\boldsymbol{Q}_{:,i}$.

The eigendecomposition of a matrix tells us many useful facts about the matrix. The matrix is singular if and only if any of the eigenvalues are zero.

A matrix whose eigenvalues are all positive is called *positive definite*. A matrix whose eigenvalues are all positive or zero valued is called *positive semidefinite*. If matrix $\boldsymbol{A}$ is positive semidefinite, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geq 0$, for any $\boldsymbol{x}$. If matrix $\boldsymbol{A}$ is positive semidefinite and $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = 0$, $\boldsymbol{x}$ must be $\boldsymbol{0}$.

Determinant

The determinant of a square matrix, denoted $\det(\boldsymbol{A})$, is a function that maps matrices to real scalars. The determinant is equal to the product of all the eigenvalues of the matrix.

$$\det(\boldsymbol{A}) = \prod_i \lambda_i. \tag{48}$$

Singular Value Decomposition

Singular value decomposition (SVD) provides another way to factorize a matrix into *singular vectors* and *singular values*. Every real matrix has a singular value decomposition, but the same is not true of the eigendecomposition. For example, the eigendecomposition is not defined if a matrix is not square. Then, we can use a singular value decomposition instead.

SVD factorizes matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}. \tag{49}$$

- $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ in equation (49) is an orthogonal matrix whose columns are left-singular vectors. Left-singular vectors are in fact the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.
- $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ in equation (49) is an orthogonal matrix whose columns are right-singular vectors. Right-singular vectors are in fact the eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.
- The elements along the diagonal of $\boldsymbol{D}$ are the singular values of the matrix $\boldsymbol{A}$. The square root of non-zero eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ or $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ are non-zero singular values of $\boldsymbol{A}$.

We can use it to partially generalize matrix inversion to nonsquare matrices.

References

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.