

Linear Algebra Notes

- Watch the [3Blue1Brown Linear Algebra Video Series](#).

Vectors

- For the purposes of this course, a vector $\vec{x} \in \mathbb{R}^n$ is some $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Intuitively, a vector can be understood as some arrow in space with its tail at the origin and its tip at the corresponding coordinates.
- For all vectors $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \vec{x}' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} \in \mathbb{R}^n$ and scalar $\lambda \in \mathbb{R}$,
 - vector sum $\vec{x} + \vec{x}'$ is defined as $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} x_1 + x'_1 \\ \vdots \\ x_n + x'_n \end{pmatrix}$,
 - scalar multiple $\lambda \vec{x}$ is defined as $\begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}$,
 - scalar (dot) product $\vec{x} \cdot \vec{x}'$ is defined as $\sum_{i=1}^n x_i x'_i$.
 - and the norm of a vector $\|\vec{x}\|$ is defined as $\sqrt{\vec{x} \cdot \vec{x}}$. This represents the "length" of the vector.
- Vector \vec{x} is a linear combination of the collection of vectors $\vec{x}_1, \dots, \vec{x}_m$ iff $\exists \alpha_1, \dots, \alpha_m \in \mathbb{R} : \vec{x} = \alpha_1 \vec{x}_1 + \dots + \alpha_m \vec{x}_m$.
- Collection of vectors $\vec{x}_1, \dots, \vec{x}_m$ is linearly independent iff $\nexists i \in \{1, \dots, m\} : \vec{x}_i$ is a linear combination of $\vec{x}_1, \dots, \vec{x}_{i-1}, \vec{x}_{i+1}, \dots, \vec{x}_m$. In words, collection of vectors $\vec{x}_1, \dots, \vec{x}_m$ is linearly independent iff none of the vectors in the collection is a linear combination of the others.

Matrices

- A $n \times m$ matrix $A_{n \times m}$ is some $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$. [Intuitively, a matrix can be understood as a linear transformation of vector space.](#) The columns of a matrix can be interpreted as the locations of the original basis vectors post-transformation. On this interpretation, the first column gives the transformed location of the first basis vector $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, the second column gives the transformed location of the second basis vector $\begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$, and so on.

Common Definitions

- A row matrix is a $1 \times m$ matrix, a column matrix is a $n \times 1$ matrix, a square matrix is a $n \times m$ matrix where $n = m$.
- The $n \times m$ null matrix is defined as $\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, i.e. it is the $n \times m$ matrix where each element is 0.
- The n -square identity matrix is defined as $\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$, i.e. it is the n -square matrix where each element in the diagonal is 1 and each other element is 0.

- A n -square diagonal matrix is defined as a matrix of the form $\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$, i.e. it is a n -square matrix where only elements in the diagonal are non-zero.
- A n -square upper-triangular matrix is defined as a matrix of the form $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$, i.e. it is a n -square matrix where only elements in the top-right triangle, including the diagonal, are non-zero.
- A $n \times m$ row echelon matrix is defined as a matrix where all rows consisting of only zeros are at the bottom and the leading entry (i.e. the left-most non-zero entry) of each non-zero row is to the right of the leading entry of the row above.
- A n -square symmetric matrix is defined as a matrix of the form $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix}$, i.e. it is a matrix that is symmetrical about the diagonal.

Basic Operations

- Matrix addition, scalar multiplication, and transposition are defined in the intuitive way.

Matrix Product

- The matrix product AB of $n \times m$ matrix A and $m \times l$ matrix B is defined as $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1l} \\ b_{21} & b_{22} & \dots & b_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{ml} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m a_{1i}b_{i1} & \sum_{i=1}^m a_{1i}b_{i2} & \dots & \sum_{i=1}^m a_{1i}b_{il} \\ \sum_{i=1}^m a_{2i}b_{i1} & \sum_{i=1}^m a_{2i}b_{i2} & \dots & \sum_{i=1}^m a_{2i}b_{il} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^m a_{ni}b_{i1} & \sum_{i=1}^m a_{ni}b_{i2} & \dots & \sum_{i=1}^m a_{ni}b_{il} \end{pmatrix}$. [Intuitively, matrix multiplication can be interpreted as a composition of two linear transformations of vector space](#), i.e. as the transformation that is equivalent to applying one transformation and then another. On this interpretation, the transformation corresponding to the matrix on the right (in this example, B) is applied first.

Common Results

- $A_{n \times m} I_{m \times m} = I_{n \times n} A_{n \times m} = A_{n \times m}$, i.e. multiplication of some matrix by an appropriate identity matrix leaves the matrix unchanged.
- Where each product is well-defined, $(A \cdot B) \cdot C = A \cdot (B \cdot C)$, $A \cdot (B + C) = A \cdot B + A \cdot C$, and not necessarily $A \cdot B = B \cdot A$.

Matrix Inverse

- The inverse A^{-1} of a n -square matrix A is defined as the unique n -square matrix such that $AA^{-1} = A^{-1}A = I_{n \times n}$. The inverse of some matrix does not necessarily exist. If the inverse of some matrix exists, it is unique.
- Gaussian elimination is a technique for matrix inversion. Consider the following [example](#).
 - $A = \begin{pmatrix} -1 & \frac{3}{2} \\ 1 & -1 \end{pmatrix}$.
 - Construct the augmented matrix A' by augmenting A with the identity matrix. $A' = \left(\begin{array}{cc|cc} -1 & \frac{3}{2} & 1 & 0 \\ 1 & -1 & 0 & 1 \end{array} \right)$.
 - By Gaussian elimination, transform the left hand side of A' into the identity matrix.

$$\left(\begin{array}{cc|cc} -1 & \frac{3}{2} & 1 & 0 \\ 1 & -1 & 0 & 1 \end{array} \right) \xrightarrow{R_2 \leftarrow R_1 + R_2} \left(\begin{array}{cc|cc} -1 & \frac{3}{2} & 1 & 0 \\ 0 & \frac{1}{2} & 1 & 1 \end{array} \right) \xrightarrow{R_1 \leftarrow R_1 - 3R_2}$$
 - $\left(\begin{array}{cc|cc} -1 & 0 & -2 & -3 \\ 0 & \frac{1}{2} & 1 & 1 \end{array} \right) \xrightarrow{R_1 \leftarrow -R_1; R_2 \leftarrow 2R_2} \left(\begin{array}{cc|cc} 1 & 0 & 2 & 3 \\ 0 & 1 & 2 & 2 \end{array} \right)$.
 - Then A^{-1} is given by the right hand side of the augmented matrix.
- [Cramer's rule](#) is a technique for matrix inversion.
 - Consider some invertible n -square matrix $A_{n \times n}$. Let A_{-i-j} denote the submatrix obtained by eliminating row i and column j from A . Then minor $M_{ij} = \det A_{-i-j}$ and cofactor $C_{ij} = (-1)^{i+j} M_{ij}$. Then

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix}. \text{ Note that the element in the } i^{\text{th}} \text{ row and } j^{\text{th}} \text{ column, } A_{ij}^{-1} \text{ is } C_{ji} \text{ and not } C_{ij}.$$

- Cramer's rule is useful for computing the inverse of small matrices. For a 2×2 matrix, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$,

$A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. In words, the inverse of A is obtained by (1) swapping a_{11} with a_{22} , (2) switching the sign on the other two values, and (3) dividing by $\det A$.

Common Results

- $(AB)^{-1} = A^{-1}B^{-1}$.

Determinant

- The determinant of a 1×1 matrix $A = (a)$ is a , and the determinant of a 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is $a_{11}a_{22} - a_{12}a_{21}$.
The determinant of a n -square matrix A is $\sum_{i=1}^n a_{ij}C_{ij} = \sum_{i=1}^n a_{ji}C_{ji}$ for any $j \in \{1, \dots, n\}$. **Intuitively, the determinant of some matrix can be interpreted as the factor by which vector space expands or contracts as a result of the linear transformation of vector space associated with that matrix. The determinant of some matrix is zero if the linear transformation associated with that matrix compresses vector space into a lower dimension.**
- A n -square matrix A is invertible iff $\det A \neq 0$. Intuitively, if the transformation associated with that matrix compresses vector space into a lower dimension, each point in the compressed vector space corresponds to a line or a plane in the uncompressed vector space, not to a single point. So there is no matrix that reverses the transformation.

Common Results

- $\det A^T = \det A$.
- $\det AB = \det A \det B$.
- The determinant of a triangular matrix is the product of its diagonal elements.

Rank

- The rank of a matrix A is defined as the maximum number of linearly independent rows or columns of the matrix. The maximum number of linearly independent rows is always equal to the maximum number of linearly independent columns.
- A $n \times m$ matrix A has full rank iff $\text{rank } A = \min\{n, m\}$.
- Gaussian elimination is a technique for computing the rank of a matrix.
 - Apply Gaussian elimination to transform the the matrix to row echelon form. Then the rank of the given matrix is the number of non-zero rows.

Common Results

- $\text{rank } A = \text{rank } A^T$.
- $\text{rank } A_{n \times m} \leq n, m$.
- $\det A \neq 0 \Rightarrow A$ has full rank, equivalently A does not have full rank $\Rightarrow \det A = 0$. Intuitively, if some matrix does not have full rank, then its columns are not linearly independent, so the associated transformation compresses vector space into a lower dimension, and the matrix has determinant 0.

Span

- The span of a collection of n -dimensional vectors is the set of all linear combinations of these vectors. **Intuitively, the span of some vectors can be interpreted as the space accessible by some linear combination of these vectors, i.e. the space that is accessible by scaling and linking these vectors.**
- A basis of \mathbb{R}^n is defined as a minimal collection of vectors that span \mathbb{R}^n . A collection of vectors that span \mathbb{R}^n is not a minimal such collection iff some subset of this collection spans \mathbb{R}^n .
- The columns of n -square matrix A spans \mathbb{R}^n iff A is full rank iff $\det A \neq 0$.
- $\vec{x}_1, \dots, \vec{x}_n$ span \mathbb{R}^n iff $\text{rank } (\vec{x}_1, \dots, \vec{x}_n) = n$.

Trace

- The trace of a matrix A is defined as the sum of the diagonal elements of the matrix.

Eigenvectors and Eigenvalues

- An eigenvalue and an eigenvector of a n -square matrix A are defined as a scalar λ and an n -dimensional vector \vec{x} such that $A\vec{x} = \lambda\vec{x}$. **Intuitively, the eigenvectors of a matrix are the vectors that are not "thrown off" their span by the transformation associated with that matrix, and the associated eigenvalues are the factors by which the eigenvectors are scaled by this transformation.**
- For any eigenvalue λ and eigenvector \vec{x} of n -square matrix A , $(A - \lambda I)\vec{x} = \vec{0} \Rightarrow \det(A - \lambda I) = 0$.
- $\det A = \prod_i \lambda_i$. In words, the determinant of a matrix is equal to the product of its eigenvalues.
- $\text{tr } A = \sum_i \lambda_i$. In words, the trace of a matrix is equal to the sum of its eigenvalues.
- From the signs of the determinant and trace of a matrix, it is possible to infer the signs of its eigenvalues (and hence its definiteness).

Linear Systems

- A system of linear equations can be written as $A\vec{x} = \vec{b}$, where A is a $n \times m$ matrix, \vec{x} is a vector of length m and \vec{b} is a vector of length n . **Intuitively, a system of linear equations, written as above, can be understood as being solved by the vectors that, when transformed by A land on \vec{b} .**
- A system of linear equations has either no solution, exactly one solution, or infinitely many solutions. If the system of linear equations, written as $A\vec{x} = \vec{b}$ is such that A is invertible, i.e. A^{-1} exists, then the system of linear equations has exactly one solution, namely $\vec{x} = A^{-1}\vec{b}$. If the system of linear equations, so written, is such that $\vec{b} = \vec{0}$ (i.e. the system is homogenous), then it has either one solution or infinitely many solutions, since $\vec{x} = \vec{0}$ is a solution to all such systems.

Basis Transformation

- Consider some n -square matrix A . If matrix A is full rank, then matrix A can be understood as representing n basis vectors. Then, the inverse matrix A^{-1} can be understood as a transformation of vectors represented in the standard basis to vectors represented in the basis represented by A . For example, consider $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. $\det A = 1 \times 4 - 2 \times 3 \neq 0$, so A is full rank and A^{-1} exists. By Gaussian elimination:
$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array} \right) \xrightarrow{R_1 \leftarrow R_1 - \frac{R_2}{2}; R_2 \leftarrow R_2 - 3R_1} \left(\begin{array}{cc|cc} -\frac{1}{2} & 0 & 1 & -\frac{1}{2} \\ 0 & -2 & -3 & 1 \end{array} \right) \xrightarrow{R_1 \leftarrow -2R_1; R_2 \leftarrow -\frac{R_2}{2}} \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array} \right) \Rightarrow$$

 $A^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$. So vector $\vec{b} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$ in the standard basis is equivalent to $A^{-1}\vec{b} = \begin{pmatrix} -4 \\ \frac{9}{2} \end{pmatrix}$ in the basis whose basis vectors in the standard basis are $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$.
- Basis transformation is a technique for computing matrix exponentiation. Consider the following example.
 - $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.
 - Computing the eigenvalues of A : $\det \begin{pmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{pmatrix} = (1-\lambda)(4-\lambda) - 2 \times 3 = 0 \Rightarrow \lambda = \frac{5}{2} + \frac{\sqrt{33}}{2}$ or $\frac{5}{2} - \frac{\sqrt{33}}{2}$.
 - The corresponding eigenvectors are $\begin{pmatrix} -\frac{1}{2} + \frac{\sqrt{33}}{6} \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -\frac{1}{2} - \frac{\sqrt{33}}{6} \\ 1 \end{pmatrix}$. The corresponding eigenvectors can be found by solving $(A - \lambda I)\vec{x} = \vec{0}$.
 - In the eigenbasis whose basis vectors are $\begin{pmatrix} -\frac{1}{2} + \frac{\sqrt{33}}{6} \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -\frac{1}{2} - \frac{\sqrt{33}}{6} \\ 1 \end{pmatrix}$ in the original basis, the transformation corresponding to $(A$ in the original basis) is $\begin{pmatrix} \frac{5}{2} + \frac{\sqrt{33}}{2} & 0 \\ 0 & \frac{5}{2} - \frac{\sqrt{33}}{2} \end{pmatrix}$.
 - Then, the transformation corresponding to A^n (in the original basis), in this eigenbasis is $\begin{pmatrix} \frac{5}{2} + \frac{\sqrt{33}}{2} & 0 \\ 0 & \frac{5}{2} - \frac{\sqrt{33}}{2} \end{pmatrix}^n$.
 - The inverse of $\begin{pmatrix} -\frac{1}{2} + \frac{\sqrt{33}}{6} & -\frac{1}{2} - \frac{\sqrt{33}}{6} \\ 1 & 1 \end{pmatrix}$ is $\frac{1}{22} \begin{pmatrix} 2\sqrt{33} & 11 + \sqrt{33} \\ -2\sqrt{33} & 11 - \sqrt{33} \end{pmatrix}$

- Then, $A^n = \begin{pmatrix} -\frac{1}{2} + \frac{\sqrt{33}}{6} & -\frac{1}{2} - \frac{\sqrt{33}}{6} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{5}{2} + \frac{\sqrt{33}}{2} & 0 \\ 0 & \frac{5}{2} - \frac{\sqrt{33}}{2} \end{pmatrix}^n \frac{1}{22} \begin{pmatrix} 2\sqrt{33} & 11 + \sqrt{33} \\ -2\sqrt{33} & 11 - \sqrt{33} \end{pmatrix}$.
- In words, A^n is the transformation equivalent to transformation into the eigenbasis, scaling by the eigenvalues n times, then transformation into the original basis.
- To recapitulate, consider some 2×2 matrix A with eigenvalues and eigenvectors λ_1, \vec{v}_1 and λ_2, \vec{v}_2 . Let D be the 2×2 diagonal eigenvalue matrix $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Let V be the eigenbasis whose columns are the eigenvectors. Then, $A^n = VD^nV^{-1}$. This last step can be intuitively verified as follows. V^{-1} coerces the standard basis to the eigenbasis, D^n stretches the bases by the eigenvalues n times, V then returns the eigenbasis to the standard basis.