COMP 411/511
Computer Vision with Deep Learning

Linear Algebra Tutorial

Agenda

- General Notations
- Matrix Operations
- Linear Independence and Span
- Norms
- Special Matrices
- Eigendecomposition
- Singular Value Decomposition
- The Moore-Penrose Pseudoinverse
- Trace
- Determinant

Adapted from *Deep Learning Book, Chapter 2 and*
Stanford CS229 Linear Algebra Refresher

General Notations

- **Scalar**
  - Just a single number. e.g., $s \in R$ or $n \in N$
- **Vector**
  - An array of numbers arranged in order.
  - Identify each individual number by its index in the ordering.
  - e.g., $x \in R^n$
  - We can think of vectors as identifying points in space, with each element giving the coordinate along a different axis

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

General Notations

- **Matrix**
  - A 2-D array of numbers, so each element is specified by 2 indices.
  - e.g., $A \in R^{m \times n}$
  - $A_{i,j}$ is the element in the i'th row and j'th column
  - $A_{i,:}$ is the i'th row, $A_{:,j}$ is the j'th column

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix}$$

- **Tensor**
  - Generalisation of matrices to more than 2 dimensions
  - e.g, $\boldsymbol{A} \in R^{m \times n \times p}$
  - $\boldsymbol{A}_{i,j,k}$ is the element at the (i,j,k) coordinate

Matrix Operations

- **Transpose**
  The transpose of a matrix $A \in R^{m \times n}$, noted A^T , is the mirror image of the matrix along the diagonal:
  $$\boxed{\forall i, j, \qquad A^T_{i,j} = A_{j,i}}$$
  $$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$
- Note that: $\left(\mathbf{A}^{\mathrm{T}}\right)^{\mathrm{T}} = \mathbf{A}$
- Vectors can be thought of as matrices that contain only one column. The transpose of a vector is therefore a matrix with only one row.
- A scalar can be thought of as a matrix with only a single entry. From this, we can see that a scalar is its own transpose: $a = a^T$.

Matrix Operations

- **Matrix-Matrix Addition**
  We can add matrices to each other, as long as they have the same shape, just by adding their corresponding elements: $C = A + B$ where $C_{i,j} = A_{i,j} + B_{i,j}$

- **Matrix-Scalar Operations**
  We can also add a scalar to a matrix or multiply a matrix by a scalar, just by performing that operation on each element of a matrix: $D = a \cdot B + c$ where $D_{i,j} = a \cdot B_{i,j} + c$

Matrix Operations

In order for the product to be defined:
A must have the same number of columns as B has rows.

- **Matrix-Matrix Product**

  The product of matrices $A \in R^{m \times n}$ and $B \in R^{n \times p}$ is a matrix $C \in R^{m \times p}$, such that:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix}$$

$$\mathbf{C} = \mathbf{AB} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{pmatrix}$$

such that

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj},$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, p$.

Matrix Operations

- **Hadamard (Element-wise) Product**
  For two matrices *A* and *B* of the same dimension, the Hadamard product $A \odot B$ is a matrix of the same dimension as the operands, with elements given by:

$$(A \odot B)_{ij} = (A)_{ij}(B)_{ij}.$$

Matrix Operations

- **Vector-Vector products**
  - inner-product (dot product): for $x, y \in R^n$, we have: $$x^T y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$$
  - outer-product: for $x \in R^m$, for $y \in R^n$, we have:

$$xy^T = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Matrix Multiplication Properties

- **Distributive:** $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$
- **Associative:** $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$
- **Not commutative:** AB != BA
  - dot product between vectors is commutative though: $\boldsymbol{x}^\top \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{x}$
- $(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$

Matrix Inversion

- **Identity Matrix**
  - a matrix that does not change any vector when we multiply that vector by that matrix.

    $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ and $\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x}$
  - 1s in the diagonal and 0s everywhere else.

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

- **Inverse**
  - The inverse of an invertible square matrix A is noted A^{-1} and is the only matrix that:

$$\boxed{AA^{-1} = A^{-1}A = I}$$

  - Not all square matrices are invertible.
  - For matrices A, B we have $(AB)^{-1} = B^{-1} A^{-1}$

System of Linear Equations

$A \in R^{m \times n}$ known matrix

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

$b \in R^m$ known vector

$x \in R^n$ unknown vector

$$\boldsymbol{A}_{1,:}\boldsymbol{x} = b_1$$

$$\boldsymbol{A}_{2,:}\boldsymbol{x} = b_2$$

$$\dots$$

$$\boldsymbol{A}_{m,:}\boldsymbol{x} = b_m$$

$$\boldsymbol{A}_{1,1}x_1 + \boldsymbol{A}_{1,2}x_2 + \cdots + \boldsymbol{A}_{1,n}x_n = b_1$$

$$\boldsymbol{A}_{2,1}x_1 + \boldsymbol{A}_{2,2}x_2 + \cdots + \boldsymbol{A}_{2,n}x_n = b_2$$

$$\dots$$

$$\boldsymbol{A}_{m,1}x_1 + \boldsymbol{A}_{m,2}x_2 + \cdots + \boldsymbol{A}_{m,n}x_n = b_m.$$

System of Linear Equations

$$
\begin{aligned}
3x + 2y - z &= 1 \\
2x - 2y + 4z &= -2 \\
-x + \frac{1}{2}y - z &= 0
\end{aligned}
$$

$$
\begin{bmatrix} 3 & 2 & -1 \\ 2 & -2 & 4 \\ -1 & \frac{1}{2} & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}
$$

System of Linear Equations

- $Ax=b$
- $A^{-1}Ax=A^{-1}b$
- $Ix = A^{-1}b$
- $x = A^{-1} b$
- if A^{-1} exists!

- A^{-1} is primarily useful as a theoretical tool, and should not actually be used in practice for most software applications. Because A^{-1} can be represented with only limited precision on a digital computer, algorithms that make use of the value of b can usually obtain more accurate estimates of x.

System of Linear Equations

- $Ax=b$ may have 0, 1 or infinitely many solutions
- cannot have more than one but less than infinitely many solutions:
    - assume x, y are solutions:
    - $Ax = b$, $Ay=b$
    - $z = \alpha x + (1-\alpha)y$
    - $Az = A\alpha x + A(1-\alpha)y$
    - $Az = \alpha b + (1-\alpha)b$
    - $Az = \alpha b + b - \alpha b$
    - $Az = b$
    - z is also a solution
- For A^{-1} to exist, $Ax=b$ must have exactly 1 solution for every value of b

System of Linear Equations

- To analyze how many solutions the equation has, think of the columns of *A* as specifying different directions we can travel in from the origin
- then determine how many ways there are of reaching b
- In this view, each element of *x* specifies how far we should travel in each of these directions, with x_i specifying how far to move in the direction of column *i*:

$$\boldsymbol{A}\boldsymbol{x} = \sum_i x_i \boldsymbol{A}_{:,i}.$$

System of Linear Equations

- a linear combination of some set of vectors $\{v^{(1)}, \ldots, v^{(n)}\}$ is given by multiplying each vector $v^{(i)}$ by a corresponding scalar coefficient and adding the results:

$$\sum_i c_i \boldsymbol{v}^{(i)}.$$

$$4 \begin{bmatrix} 1 \\ -2 \\ 0 \\ 5 \end{bmatrix} + 3.1 \begin{bmatrix} 2 \\ 4 \\ 1 \\ -1 \end{bmatrix} - 4.2 \begin{bmatrix} 3 \\ -1 \\ 3.2 \\ 2 \end{bmatrix}$$

System of Linear Equations

- **Span**
  span of a set of vectors is the set of all points obtainable by linear combination of the original vectors.

- **Column Space (Range)**
  span of columns of matrix *A*

- Ax=b has a solution for a particular b ⇔ b in column space of A
- Ax=b has a solution for every b ⇔ column space of A should be all of R^m
  - So, A must have at least *m* columns ($n \geq m$) (necessary condition)

Linear Independence

- $\begin{bmatrix}1 & 2\\3 & 6\end{bmatrix}$: Column space not R^2 because the 2nd column is redundant

- **A set of vectors is linearly independent if no vector in the set is a linear combination of the other vectors.**
- Adding a vector that is a linear combination of a set of vectors to that set, does not change the span of the set.
- matrix A has at least one set of m linearly independent columns ⇔ column space of A is R^m ⇔ column space of A should be all of R^m ⇔ Ax=b has a solution for every b (necessary and sufficient condition)

System of Linear Equations

- For the matrix to have an inverse, we additionally need to ensure that Ax=b has at most one solution for each value of *b*. To do so, we need to make certain that the matrix has at most *m* columns. Otherwise there is more than one way of parametrizing each solution.
- Together, this means that the matrix must be **square**, that is, ***m*=*n*, and that all the columns be linearly independent**.
- A square matrix with linearly dependent columns is known as singular.
- If *A* is not square or is square but singular, solving the equation is still possible,but we cannot use the method of matrix inversion to find the solution.

System of Linear Equations

- A has inverse ⇔ Ax=b must have exactly 1 solution for every value of b

Proof:

- A has inverse ⇒Ax=b must have exactly 1 solution for every value of b
    - A has inverse, therefore x= $A^{-1}b$ is a solution for every b
    - assume that x_1 and x_2 are both solutions for Ax=b, therefore $Ax_1=b \Rightarrow x_1=A^{-1}b$, $Ax_2=b \Rightarrow x_2=A^{-1}b \Rightarrow x_1=x_2$
- Ax=b must have exactly 1 solution for every value of b ⇒ A has inverse
    - $Ax=[1,0,...,0]^T$ has a solution called x_1
    - $Ax=[0,1,...,0]^T$ has a solution called x_2
    - …
    - $Ax=[0,0,...,1]^T$ has a solution called x_n
    - $A^{-1}= [x_1\ x_2\ \dots\ x_n]$

Norms

- A function to measure the size of vectors

- It maps a vector to a non-negative number

- any function *f* that satisfiesthe following properties:
  - $f(x)=0 \Rightarrow x=0$
  - $f(x+y) \leq f(x)+f(y)$ (the triangle inequality)
  - $\forall \alpha \in R, f(\alpha x) = |\alpha| f(x)$

Norms

- L^p norm: $||\boldsymbol{x}||_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$ for $p \in \mathbb{R}, p \geq 1.$

  - L^1 norm = $\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i| \,.$
  - L^2 (Euclidean) norm = $\|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}.$ $= \sqrt{x^T x}$
  - L^∞ norm = $\|\mathbf{x}\|_\infty := \max\left(|x_1|, \ldots, |x_n|\right)$

Norms

- size of a matrix *A* is usually measured as

Frobenius norm = $||A||_F = \sqrt{\sum_{i,j} A_{i,j}^2},$

- The dot product of two vectors can be rewritten in terms of norms:

$$\boldsymbol{x}^\top \boldsymbol{y} = ||\boldsymbol{x}||_2 ||\boldsymbol{y}||_2 \cos\theta$$

angle between *x* and *y*

Special Matrices

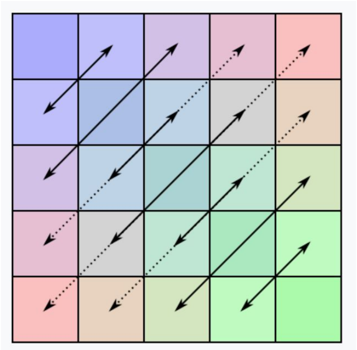
- **Diagonal**
  - $D_{i,j} = 0$ for all i≠j

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}$$

  - *diag(v)* denotes a square diagonal matrix whose diagonal entries are given by the entries of the vector *v*
  - multiplying by a diagonal matrix is computationally efficient
  - $diag(v)x = v \odot x$
  - inverse of *diag(v)* exists ⇔ $\forall i\ v_i \neq 0$
    in that case: $diag(v)^{-1} = diag([1/v_1, \ldots, 1/v_n])$

Special Matrices

- **Symmetric**
  - $A=A^T$
  - A is symmetric $\iff$ for every i, j, $\boldsymbol{A}_{i,j} = \boldsymbol{A}_{j,i}$
  - e.g., $A = \begin{bmatrix} 1 & 7 & 3 \\ 7 & 4 & 5 \\ 3 & 5 & 2 \end{bmatrix}$



Special Matrices

- **Unit Vector**
  - a vector with unit norm $||\boldsymbol{x}||_2 = 1.$
- **Orthogonal Vectors**
  - vectors *x* and *y* are orthogonal to each other ⇔ $x^\top y = 0$
  - orthogonal & unit norm vectors are called **orthonormal**
- **Orthogonal Matrix**
  - a square matrix whose rows are mutually orthonormal and whose columns are mutually **orthonormal** $\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}.$
  - Therefore: $\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$

Eigendecomposition

- **Eigenvector & Eigenvalue**
  - An eigenvector of a square matrix *A* is a *nonzero* vector *v* such that multiplication by *A* alters only the scale of *v*:

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

  - The scalar *λ* is known as the **eigenvalue** corresponding to this eigenvector.
  - *v* eigenvector of *A* ⇒ *sv* eigenvector of A with the same eigenvalue for $s \in R$, $s \neq 0$

Eigendecomposition

- Suppose that matrix A has n linearly independent eigenvectors $\{v^{(1)}, \dots, v^{(n)}\}$ with corresponding eigenvalues $\{\lambda_1, \dots, \lambda_n\}$
- Concatenate all the eigenvectors to form a matrix V with one eigenvector per column: $\boldsymbol{V} = [v(1), \dots, v(n)]$
- Concatenate the eigenvalues to form a vector $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_n]^T$
- The eigendecomposition of A is given by:

$$\boldsymbol{A} = \boldsymbol{V} \operatorname{diag}(\boldsymbol{\lambda}) \boldsymbol{V}^{-1}$$

- Not every matrix can be decomposed into eigenvalues and eigenvectors

Eigendecomposition

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

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\top$$

an orthogonal matrix composed of eigenvectors of A

diagonal matrix of corresponding eigenvalues

- We usually sort the entries of Λ in descending order.

Eigendecomposition

- The matrix is singular if and only if any of the eigenvalues are zero
- For a symmetric matrix:
  - positive definite: all eigenvalues are positive $\Leftrightarrow \forall x \neq 0, x^T A x > 0$
  - positive semi-definite: all eigenvalues are non-negative $\Leftrightarrow \forall x, x^T A x \geq 0$
  - negative definite: all eigenvalues are negative $\Leftrightarrow \forall x \neq 0, x^T A x < 0$
  - negative semi-definite: all eigenvalues are non-positive $\forall x, x^T A x \leq 0$

Singular Values Decomposition (SVD)

- Not every matrix has an eigendecomposition
- Every matrix has a singular value decomposition

$$\underbrace{\boldsymbol{A}}_{m\times n \text{ arbitrary}} = \underbrace{\boldsymbol{U}}_{m\times m \text{ orthogonal}} \underbrace{\boldsymbol{D}}_{m\times n \text{ diagonal}} \underbrace{\boldsymbol{V}^\top}_{n\times n \text{ orthogonal}}$$

- Elements along the diagonal of D: singular values of A
- Columns of U: left-singular vectors of A
- Columns of V: the right-singular vectors of A

The Moore-Penrose Pseudoinverse

- Matrix inversion not defined for non-square matrices
- Suppose we want to make a left-inverse *B* of a matrix *A* so that we can solve a linear equation $Ax = y$
- by left-multiplying each side to obtain $x = By$
- Depending on the structure of the problem, it may not be possible to design a unique mapping from *A* to *B*

$$\boldsymbol{A}^{+} = \boldsymbol{V}\boldsymbol{D}^{+}\boldsymbol{U}^{\top}$$

- D^{+} of a diagonal matrix *D* is obtained by taking the reciprocal of its nonzero elements then taking the transpose of the resulting matrix

The Moore-Penrose Pseudoinverse

$$\boldsymbol{A}^{+} = \boldsymbol{V}\boldsymbol{D}^{+}\boldsymbol{U}^{\top}.$$

- *A* has more columns than rows: then pseudo inverse provides the solution with minimal Euclidean norm $||x||_2$ among all possible solutions.

- *A* has more rows than columns: pseudoinverse gives us the x for which Ax is as close as possible to y in terms of Euclidean norm $||Ax - y||_2$

Trace Operator

- **Trace**
  sum of all the diagonal entries of a matrix $\mathrm{Tr}(\boldsymbol{A}) = \sum_i \boldsymbol{A}_{i,i}.$
- $$||A||_F = \sqrt{\sum_{i,j} A_{i,j}^2} = \sqrt{\mathrm{Tr}(\boldsymbol{A}\boldsymbol{A}^\top)}.$$
- $$\mathrm{Tr}(\boldsymbol{A}) = \mathrm{Tr}(\boldsymbol{A}^\top).$$
- $$\mathrm{Tr}(\prod_{i=1}^{n} \boldsymbol{F}^{(i)}) = \mathrm{Tr}(\boldsymbol{F}^{(n)} \prod_{i=1}^{n-1} \boldsymbol{F}^{(i)}).$$
- $$\mathrm{Tr}(\boldsymbol{ABC}) = \mathrm{Tr}(\boldsymbol{CAB}) = \mathrm{Tr}(\boldsymbol{BCA})$$
- $$\mathrm{Tr}(\boldsymbol{AB}) = \mathrm{Tr}(\boldsymbol{BA})$$

The Determinant

- det(A) is a function that maps square matrices to real scalars.
- determinant is equal to the product of all the eigenvalues of the matrix.
- The absolute value of the determinant can be thought of as a measure of how much multiplication by the matrix expands or contracts space.
- if det(A)=0, then space is contracted completely along at least one dimension, causing it to lose all its volume.
- if det(A)=1 1, then the transformation preserves volume.
- det(A) ≠ 0 ⇔ A is invertible