Lecture # 3

Basic Linear Algebra

Definitions

Linear Algebra provides techniques and notations to express a set of equations in a compact form and operate on them. Consider the following two equations with two variables:

$4x_1 - 5x_2 = -13$

$-2x_1 + 3x_2 = 9$

This can be written in matrix form as

$Ax = b$

$$A=\begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix};\ b=\begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

Where A and b are matrixes of the form

There are many advantages in expressing equations in this form.

Notations

By x ∈ $\mathbb{R}^{m\ x\ n}$ we denote a matrix with m rows and n columns, where the entries of A are real numbers.

By A ∈ $\mathbb{R}^{n}$ we denote a vector with n entries. By convention, an n dimensional vector is often thought as a matrix with n rows and 1 column. This is known as column vector. A row vector (a matrix with n rows and one columns) is produced as x^T (transpose of matrix x- we will see later what this means).

The i*th* element of a vector x is denoted by x_i

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

We use the notation a_{ij} to identify the matrix element in *ith* row and *jth* column.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

We denote jth column of A by a_j

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

We denote ith row of A by a_i^T

$$A = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix}.$$

Matrix Algebra

Matrix Multiplication

The product of two matrices A $\in \mathbb{R}^{m\times n}$ and B $\in \mathbb{R}^{n\times p}$ is

C =AB $\in \mathbb{R}^{m\times p}$

Where $C_{ij} = \sum_{i=1}^{k} A_{ik} B_{kj}$

In order for matrix product to exist the number of columns in A must be equal to the number of rows in B.

Vector-Vector product

Given two vectors x,y $\in \mathbb{R}^{n}$, the quantity $x^T y$ is called inner product or dot product of the vectors is a real number

Matrix Multiplication

The product of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ is the matrix

$$C = AB \in \mathbb{R}^{m \times p},$$

where

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

Note that in order for the matrix product to exist, the number of columns in A must equal the number of rows in B. There are many ways of looking at matrix multiplication, and we'll start by examining a few special cases.

Vector-Vector Product

Given two vectors $x, y \in \mathbb{R}^n$, the quantity $x^T y$, sometimes called the ***inner product*** or ***dot product*** of the vectors, is a real number given by

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i.$$

Observe that inner products are really just special case of matrix multiplication. Note that it is always the case that $x^T y = y^T x$.

Given vectors $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$ (not necessarily of the same size), $xy^T \in \mathbb{R}^{m \times n}$ is called the ***outer product*** of the vectors. It is a matrix whose entries are given by $(xy^T)_{ij} = x_i y_j$,

Vector and Matrix Multiplication

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i .$$

Inner products are a special case of matrix multiplication. Note that it is always the case that $x^T y = y^T x$

Given vectors $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ (not necessarily of the same size), $xy^T \in \mathbb{R}^{m \times n}$ is called the outer product of the vectors. It is a matrix whose elements are given by $(xy^T)_{ij} = x_i y_j$.

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix} .$$

Matrix-Vector Product

Matrix-Vector Product

Given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $x \in \mathbb{R}^n$, their product is a vector $y = Ax \in \mathbb{R}^m$. If A is in the row form, we have

$$y = Ax = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}.$$

If we write A in the column form we have

$$y = Ax = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_1 \end{bmatrix} x_1 + \begin{bmatrix} a_2 \end{bmatrix} x_2 + \ldots + \begin{bmatrix} a_n \end{bmatrix} x_n$$

Here y is a linear combination of the columns of A, with the coefficients given by the entries of x.

Matrix-Matrix Product

Matrix-Matrix Product

We can view Matrix-Matrix multiplication as a set of vector-vector products C=AB. In this case the (i,j)th entry of C is equal to the inner product of the ith row of A and the jth row of B, like the following:

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}$$

Rules about Matrix Multiplication

- Matrix multiplication is associative - A(BC) = (AB)C
- Matrix multiplication is distributive – A(B+C) = AB + AC
- Matrix multiplication is not, in general, commutative - $AB \neq BA$

Basic Properties

Properties of Matrices

In the following some basic definitions properties of matrices are summarized.

The identity matrix or Diagonal matrix

The identity matrix is denoted by $I \in \mathbb{R}^{n \times n}$ is a square matrix with ones on the diagonal and zeros elsewhere

$I_{ij} = 1$ *for* $i=j$

$I_{ij} = 0$ *for* $i \neq j$

This has the property that for all $A \in \mathbb{R}^{m \times n}$

$AI = A = IA$

A diagonal matrix is a matrix with all the non-diagonal elements zero.
This is denoted by

$D = diag\{d1, d2, \ldots dn\}$

$D_{ij} = d_i$ *for* $i=j$

$D_{ij} = 0$ *for* $i \neq j$

Matrix Transpose

The Transpose

The transpose of a matrix is when changing the columns and rows. Given a matrix $A \in \mathbb{R}^{m \times n}$, its transpose is written as $A^T \in \mathbb{R}^{n \times m}$. In other words
$(A^T)_{ij} = A_{ji}$

The transpose of a column vector is a row vector. Some properties of transpose matrix include

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A+B)^T = A^T + B^T$

Symmetric Matrices

A square matrix is symmetric if A = AT. It is anti-symmetric if A = - AT. Based on this definition, any square matrix can be written as the sum of a symmetric matrix and an anti-symmetric matrix.

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$$

Matrix Trace

The Trace

The trace of a square matrix is the sum of all the diagonal elements of that matrix

$$Tr(A) = \sum_{i=1}^{n} A_{ij}$$

Some properties of trace matrix include:

- For $A \in \mathbb{R}^{n\times n}$ Tr A = Tr A^T
- For $A, B \in \mathbb{R}^{n\times n}$ Tr (A+B) = Tr A + Tr B
- For $A \in \mathbb{R}^{n\times n}$ and $t \in \mathbb{R}$ Tr tA = t Tr A
- For A and B such that AB is square, Tr AB = Tr BA

Definition: Linearly Independent Vectors and Matrices

Linearly Independent

A set of vectors $x = (x_1, x_2, \ldots, x_n)$, is linearly independent if no vector can be represented as a linear combination of the rest of the vectors. Similarly, if one vector belonging to the set can be represented as a linear combination of the remaining vectors, the vectors are said to be linearly dependent. For example, in the following case, the vectors are said to be linearly dependent

$$x_n = \sum_{i=1}^{n-1} \alpha_i \; x_i$$

Definition: Rank of a Matrix

Rank of a Matrix

The **rank** of a matrix A corresponds to the maximal number of linearly independent columns of A. Rank is thus a measure of the non-degenerativeness of the system of linear equations and linear transformations encoded by A.

Example: The following matrix has rank 2. The first two rows are linearly independent, sp the rank is at least 2, but all three rows are linearly dependent (subtracting the second row from the first row gives the third row. Therefore, the rank is less than 3.

$$\begin{matrix} 1 & 0 & 1 \\ -2 & -3 & 1 \\ 3 & 3 & 0 \end{matrix}$$

The column rank of a matrix is the size of the largest subset of columns of A that constitute a linearly independent set. The row rank is the largest number of rows of A that constitute a linearly independent set.

For $A \in \mathbb{R}^{m\times n}$, rank (A) $\leq \min(m, n)$. If rank (A) $= \min(m, n)$, then A is said to be full rank.

Norms of a Matrix

Norms

The norm of a vector is a measure of its length. In n-dimensional Euclidian space the length of a vector $x = (x_1, x_2, \ldots, x_n)$, or the norm, is estimated as

$$|X|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

This gives the distance from the origin to the point X, a consequence of Pythagorean theorem. The norm has always be a positive number.

Inverse of a Matrix

The Inverse

The inverse of a square matrix $A \in \mathbb{R}^{n\times n}$ is denoted A^{-1} and is a unique matrix such that

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

Where I is a unitary matrix. Non-square matrices do not have an inverse.

Orthogonal Matrices

Orthogonal Matrices

Two vectors $x, y \in \mathbb{R}^n$ are orthogonal if $x^T y = 0$.
A square matrix $U \in \mathbb{R}^{n\times n}$ is orthogonal if all its columns are orthogonal to each other. A vector x is normalized if |x| = 1

Determinants

Determinant

The determinant of a square matrix is a value that can be computed from the elements of the matrix. Geometrically, it can be viewed as the scaling factor of the linear transformation described by the matrix. A determinant is denoted as det(A) or |A|. A square matrix has a determinant as follows

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Similarly, a 3x3 square matrix has a determinant

$$|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$
$$= aei + bfg + cdh - ceg - bdi - afh.$$

Geometrically, we could show a 2 dimensional determinant as a parallelogram (Figure 1). The value of the determinant corresponds to the area of the parallelogram

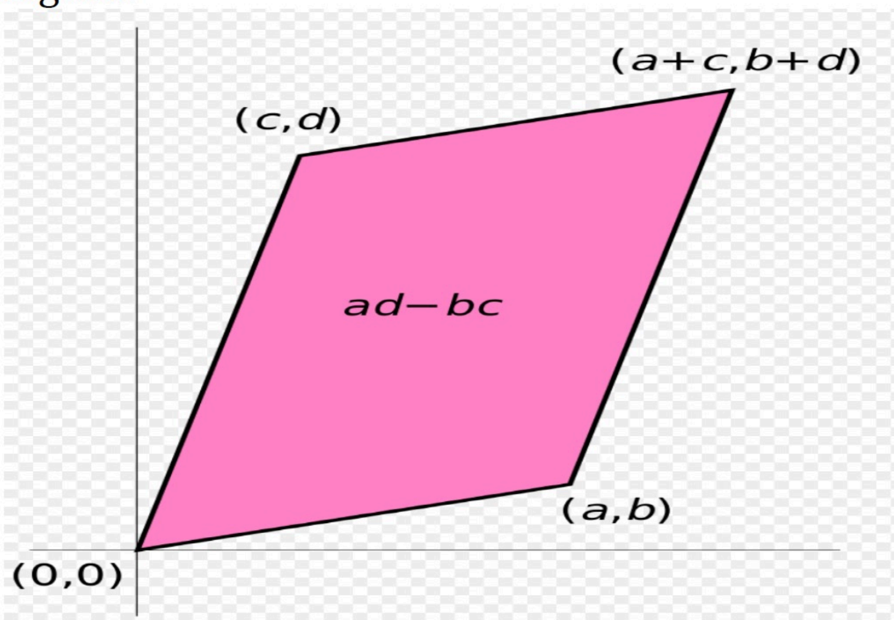

Figure 1: shows the geometric meaning of the determinant of a square matrix

Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors

Given a square matrix $A \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of A and $x \in \mathbb{C}^n$ is the corresponding eigenvector if

$$Ax = \lambda x \quad x \neq 0$$

This means that multiplying A by a vector x, results in a new vector that has the same direction as x but scales by a factor λ. The matrix A is a square matrix and x is a column vector. The mapping is a result of matrix multiplication in the left and scaling of the column vector in the right hand side of the equation. Geometrically an eigenvector, corresponding to a real nonzero eigenvalue, points in a direction that is stretched by the transformation and the eigenvalue is the factor by which it is stretched. If the eigenvalue is negative, the direction is reversed.

In Figure 2, the length of vector x is increased by multiple of λ. Therefore x is the eigenvector with λ as its eigenvalue.

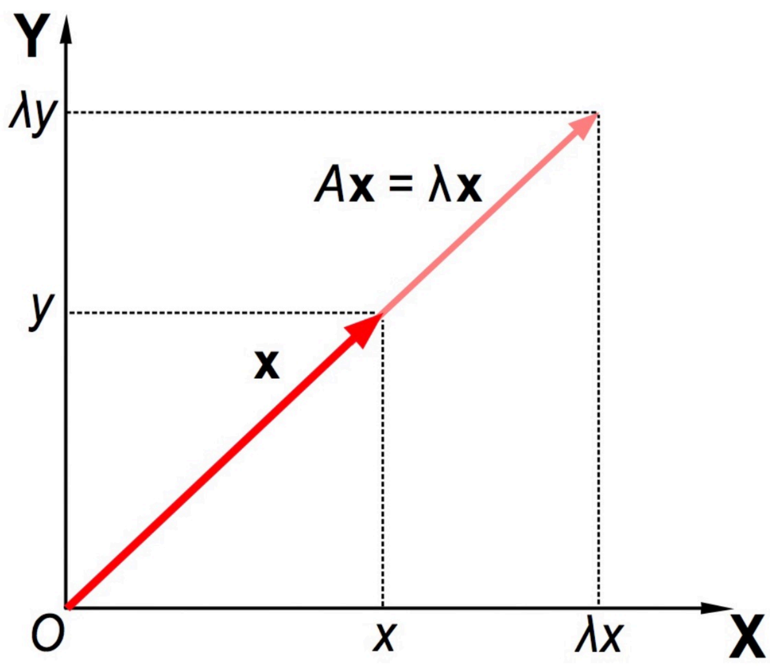

Fig 2. Vector x is stretched by a scalar value λ under transformation by matrix A. The direction of the vector does not change. Vector x is an eigenfunction with eigenvalue λ.

Applications of Eigenvector and Eigenvalue

Consider the following matrix equation:

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \times \binom{1}{3} = \binom{11}{5} \text{ and } \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \binom{3}{2} = 4 \times \binom{3}{2}$$

In the example on the left, the resulting vector is not an integer multiplication of the original vector whereas in the example in the right, it is exactly 4 times the original vector. The vector here is a vector in 2-dimensional space. The vector 32 represents an arrow from the origin (0,0) to the point (3,2) on the (X,Y) plane. The square matrix is a transformation matrix. By multiplying this on the left of a vector, the result is another vector that is transformed from its original position. Eigenvectors arise from such transitions.

A transformation matrix that is multiplied by a vector, reflects vectors on the line Y = X. If there is a vector that lies on the line Y=X, that (and all multiples of it) would be eigenvectors of that transformation matrix.

Properties of Eigenvectors

Eigenvectors can only be found for square matrices and not every square matrix has eigenvector. For an n x n matrix, there are n eigenvectors. Therefore a 3 x 3 matrix has 3 eigenvectors. Also, all the eigenvectors of a matrix are perpendicular- at right angle to each other. This means that one could express the data in terms of these orthogonal eigenvectors instead of expressing them in terms of X and Y axes. For matrices with size larger than 3 x 3, finding eigenvectors becomes complicated.

Properties of Eigenvalues

Eigenvalues are associated with eigenvectors. In the above example, “4” is the eigenvalue of the square matrix on the left. No matter what multiple of the eigenvector we took before multiplying it with the square matrix, we would always get 4 times the scaled vector as our result.

How to determine Eigenvalues and Eigenvectors

For a square matrix $A \in \mathbb{R}^{n \times n}$, is an eigenvalue of A and x is the corresponding eigenvector if

L

$Ax = \lambda x \quad x \neq 0$

This means that multiplying *A* by the vector x results in a new vector that points in the same direction as x, but scaled by a factor λ. Also, for any eigenvector x, and scalar *t*,

$A(cx) = c(Ax) = c(\lambda x) = \lambda(cx)$. Therefore, *cx* is also an eigenvector. Because of this, we assume that eigenvector is normalized to have length 1. The above equation can be written in the form that (l,x) are eigenvalue and eigenvector of A: $(\lambda I - A)x = 0 \quad x \neq 0$.

$(\lambda I - A)x = 0$ has a non-zero solution to x only if $(\lambda I - A)$ has a non-empty null space. This is only the case if $(\lambda I - A)$ is singular (its determinant being zero): $|(\lambda I - A)| = 0$

(a singular matrix does not have a matrix inverse. A matrix is singular only if its determinant is zero).

We expand this determinant into a polynomial in λ where λ will have maximum degree n. We then find the n roots of λ to find n eigenvalues- λ_1 λ_2 λ_3…λn. To find eigenvectors corresponding to eigenvalues l, we solve the linear equation

$(\lambda_i I - A)x = 0.$

Some Rules

- The trace of A is equal to the sum of its eigenvalues

$$\operatorname{tr} A = \sum_{i=1}^{n} \lambda_i.$$

- The determinant of A is equal to the product of its eigenvalues

$$|A| = \prod_{i=1}^{n} \lambda_i.$$

- The rank of A is equal to the number of non-zero eigenvalues of A
- The eigenvalues of a diagonal matrix $D = diag(d_1, \ldots, d_n)$ are just the diagonal entries $d_1, \ldots, d_n$
- If the eigenvectors of A are linearly independent, then the matrix X will be invertible: $A = X \Lambda X^{-1}$. A matrix that can be written in this form is diagnalizable.

A Review of Matrix Calculus

Matrix calculus is a way to perform multivariate operations. It collects various partial derivatives of a single function with respect to many variables or of a multivariate function with respect to single variables.

The Gradients

If f is a function that takes matrix A of size mxn and returns a real value, the gradient of f with respect to A is the matrix

$$\nabla_A f(A) \in \mathbb{R}^{m\times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

This can be written as

$$(\nabla_A f(A))_{ij} = \frac{\partial f(A)}{\partial A_{ij}}.$$

Now, if A is a vector the gradient is,

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

The gradient of a function is defined only if the function is real-valued (if it returns a scalar).

From the definition of gradients it follows that:

$$\nabla_x(f(x) + g(x) = \nabla_x f(x) + \nabla_x g(x)$$

and

$$\nabla_{\mathrm{x}}\big(t\, f(x)\big) = t\, \nabla_x f(x)\ where\ t\ \in \mathbb{R}$$

The Hessian Matrix

The Hessian Matrix

Suppose *f* is a function that operates on a vector $\mathbb{R}^n$ and returns a scalar. Then the Hessian matrix is defined as an *n x n* matrix of partial derivatives of *f*, with respect to *x*

$$\nabla_x^2 f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

This is expressed as

$$(\nabla_x^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}.$$

Further Reading

Most of the material in this course is taken from Zico Kolter and Chuong Du Lecture notes on Linear Algebra – September 2015.

Any introductory book in linear algebra contains tha material covered in this lecture.