

# Linear Algebra Review

*Conner DiPaolo*

## Contents

<b>1</b>	<b>Basics</b>	<b>1</b>
1.1	Notation . . . . .	2
<b>2</b>	<b>Matrix Multiplication</b>	<b>2</b>
2.1	Vector-Vector Products . . . . .	2
2.2	Matrix-Vector Products . . . . .	3
2.3	Matrix-Matrix Multiplication . . . . .	3
2.4	Properties of Matrix Multiplication . . . . .	3
<b>3</b>	<b>Operations and Properties</b>	<b>3</b>
3.1	The Identity Matrix and Diagonal Matrices . . . . .	3
3.2	The Transpose: $A^\top$ . . . . .	4
3.3	Symmetric Matrices . . . . .	4
3.4	The Conjugate Transpose: $A^\dagger$ . . . . .	4
3.5	Hermitian Matrices . . . . .	4
3.6	The Trace $\text{tr}(A)$ . . . . .	4
3.7	Norms . . . . .	5

This is an extension of Stanford's CS229 *Linear Algebra Review and Reference* <sup>1</sup>, written by Zico Kolter and updated by Chuong Do.

## 1 Basics

Linear Algebra allows us to interact with linear operators (or systems of linear equations) in a more powerful manner than dealing with equations. For example, consider the linear system

$$\begin{aligned}x_1 + 2x_2 &= 5 \\ 3x_1 + 4x_2 &= 1.\end{aligned}$$

This can be solved for  $x_1$  and  $x_2$  using substitution, but it is convenient (for many reasons) to investigate this system more compactly, as a matrix-vector product. Namely,

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

where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

Here,  $A$  is a *matrix* and  $\mathbf{b}$  is a *vector*. Matrices are *linear operators*. That is, they obey the following property, where  $A$  is a matrix in  $\mathbb{R}^{m \times n}$ ,  $\mathbf{x}$  and  $\mathbf{y}$  are vectors in  $\mathbb{R}^n$ , and  $c$  and  $d$  are scalars:

$$A(c\mathbf{x} + d\mathbf{y}) = cA\mathbf{x} + dA\mathbf{y}.$$

Note that this implies that

$$A\mathbf{0} = \mathbf{0}.$$

---

<sup>1</sup><http://cs229.stanford.edu/section/cs229-linalg.pdf>

## 1.1 Notation

- We denote an  $m \times n$  matrix of real numbers  $A$  as  $A \in \mathbb{R}^{m \times n}$  (“A is in R m by n”). Similarly, to declare a  $m \times n$  matrix of complex numbers we say  $B \in \mathbb{C}^{m \times n}$ .
- We denote a vector  $\mathbf{x}$  with  $n$  real elements as  $x \in \mathbb{R}^n$ . By convention,  $\mathbf{x}$  is assumed to be a column vector (that is, equivalently,  $\mathbf{x} \in \mathbb{R}^{n \times 1}$ ). If we want to represent a row vector, we use  $\mathbf{x}^\top$ , where  $\top$  is the transpose.
- The  $i$ -th element of a vector  $\mathbf{x}$  is denoted  $x_i$ .
- We denote the  $i, j$ -th element of a matrix  $A$  as  $a_{ij}$ ,  $A_{ij}$ , or  $A_{i,j}$ .

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

- We denote the  $j$ -th column of  $A$  as  $A_{:,j}$ .
- We denote the  $i$ -th row of  $A$  as  $A_{i,:}$ .
- $\mathbf{1}$  is the vector of all ones.  $\mathbf{0}$  is the vector of all zeroes. Size is dependent on context.

## 2 Matrix Multiplication

From the intro section we have already seen a matrix-vector product  $A\mathbf{x}$ . We will now define vector-vector products (the inner and outer product) and matrix-matrix products (which encapsulate matrix-vector products).

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

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

### 2.1 Vector-Vector Products

Let  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ . Then the *inner product* (sometimes referred as the dot product but we won't use that terminology)

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^n \mathbf{x}_i \mathbf{y}_i = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \cos \theta$$

where  $\theta$  is the angle between the vectors. Note that  $\langle \cdot, \cdot \rangle : (V, V) \mapsto \mathbb{R}$  (“ $\langle \cdot, \cdot \rangle$  maps two elements of the vector space  $V$  to  $\mathbb{R}$ ”) here is the standard definition of the inner product for the vector space  $V = \mathbb{R}^n$ . As we will see later in the class, other inner products can be defined between vectors in  $\mathbb{R}^n$ .

The *outer product* between  $\mathbf{x} \in \mathbb{R}^m$  and  $\mathbf{y} \in \mathbb{R}^n$  as

$$\mathbf{xy}^\top \in \mathbb{R}^{m \times n},$$

a matrix.

$$(\mathbf{xy}^\top)_{ij} = \mathbf{x}_i \mathbf{y}_j.$$

This will be useful when we talk about Principal Component Analysis and Covariance.

## 2.2 Matrix-Vector Products

The matrix-vector product between  $A \in \mathbb{R}^{m \times n}$  and  $\mathbf{x} \in \mathbb{R}^n$  is the vector  $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ . We can see from the formula for matrix-matrix multiplication that

$$A\mathbf{x} = \begin{bmatrix} | & | & \dots & | \\ a_1 & a_2 & \dots & a_n \\ | & | & \dots & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} | \\ a_1 \\ | \end{bmatrix} + x_2 \begin{bmatrix} | \\ a_2 \\ | \end{bmatrix} + \dots + x_n \begin{bmatrix} | \\ a_n \\ | \end{bmatrix},$$

a linear combination of the columns of  $A$ !

## 2.3 Matrix-Matrix Multiplication

Armed with this knowledge, we can see matrix-matrix multiplication between  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{n \times p}$  as

$$AB = \begin{bmatrix} | & | & \dots & | \\ a_1 & a_2 & \dots & a_n \\ | & | & \dots & | \end{bmatrix} \begin{bmatrix} - & b_1^T & - \\ - & b_2^T & - \\ \vdots & \vdots & \vdots \\ - & b_n^T & - \end{bmatrix} = \sum_{i=1}^n a_i b_i^T = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \dots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \dots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T b_1 & a_n^T b_2 & \dots & a_n^T b_p \end{bmatrix},$$

either an arrangement of every possible inner product between the rows of  $A$  and the columns of  $B$ , or a sum of outer products between the columns of  $A$  and the rows of  $B$ .

## 2.4 Properties of Matrix Multiplication

- Matrix multiplication is associative:  $(AB)C = A(BC)$
- Matrix multiplication is distributive:  $A(B + C) = AB + AC$
- Matrix multiplication is not commutative *in general*:  $AB \neq BA$  (in the vast majority of cases).

## 3 Operations and Properties

Most of this should hopefully be review.

### 3.1 The Identity Matrix and Diagonal Matrices

The *identity matrix*, denoted  $I \in \mathbb{R}^{n \times n}$ , is a square matrix with ones on the diagonal and zeros everywhere else:

$$I_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

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

$$AI = A = IA.$$

Intuitively, this means that as an operator the identity matrix maps every vector to itself.

A *diagonal matrix* is a matrix where all non-diagonal elements are 0. This is denoted  $D = \text{diag}(d_1, d_2, \dots, d_n)$  where

$$D_{ij} = \begin{cases} d_i & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

### 3.2 The Transpose: $A^\top$

The *transpose* of a matrix is where every row becomes a column. Given  $A \in \mathbb{R}^{m \times n}$ , the transpose of  $A$ ,  $A^\top \in \mathbb{R}^{n \times m}$  is defined as

$$(A^\top)_{ij} = A_{ji}.$$

Here are some helpful properties (proofs are easily verifiable):

- $(A^\top)^\top = A$
- $(AB)^\top = B^\top A^\top$
- $(A + B)^\top = A^\top + B^\top$

### 3.3 Symmetric Matrices

A matrix  $S \in \mathbb{R}^{n \times n}$  is *symmetric* if and only if  $S = S^\top$ . For any  $A \in \mathbb{R}^{n \times n}$ ,  $A + A^\top$  and  $A^\top A$  are both symmetric. This is easily verified from the above properties. We denote the set of symmetric matrices in  $\mathbb{R}^{n \times n}$  as  $\mathbb{S}^n$ . As we will see, symmetric matrices are often nice to work with.

### 3.4 The Conjugate Transpose: $A^\dagger$

For complex matrices  $A \in \mathbb{C}^{m \times n}$ , the conjugate transpose of  $A$ , said “ $A$  Hermitian” is denoted as  $A^\dagger = \overline{A^\top}$ . That is,

$$(A^\dagger)_{ij} = \overline{A_{ji}}.$$

For example,

$$\begin{bmatrix} 1 + 2\mathbf{i} & 3\mathbf{i} \\ 1 & 2 - \mathbf{i} \end{bmatrix}^\dagger = \begin{bmatrix} 1 - 2\mathbf{i} & 1 \\ -3\mathbf{i} & 2 + \mathbf{i} \end{bmatrix}$$

### 3.5 Hermitian Matrices

A matrix is Hermitian if  $A = A^\dagger$ . This is a generalization of symmetric matrices from real to complex matrices, and if  $A$  is Hermitian then  $A$  must also be symmetric.

### 3.6 The Trace $\text{tr}(A)$

The trace of a square matrix  $A \in \mathbb{R}^{n \times n}$  is the sum of the diagonal elements:

$$\text{tr} A = \text{tr}(A) = \sum_{i=1}^n A_{ii}.$$

Here are some useful properties:

- For  $A \in \mathbb{R}^{n \times n}$ ,  $\text{tr} A = \text{tr} A^\top$
- For  $A, B \in \mathbb{R}^{n \times n}$ ,  $\text{tr}(A + B) = \text{tr} A + \text{tr} B$
- For  $A, B$  such that  $AB$  is square,  $\text{tr} AB = \text{tr} BA$
- For  $A, B, C$  such that  $ABC$  is square,  $\text{tr} ABC = \text{tr} BCA = \text{tr} CAB$ , and this can be extended to more matrices.

(from CS229) Here's a proof of the fourth property:

$$\begin{aligned}
\text{tr}AB &= \sum_{i=1}^m (AB)_{ii} \\
&= \sum_{i=1}^m \sum_{j=1}^m A_{ij} B_{ij} \\
&= \sum_{j=1}^n \sum_{i=1}^m B_{ij} A_{ij} = \sum_{j=1}^n (BA)_{jj} \\
&= \text{tr}BA,
\end{aligned}$$

as desired.

A very useful property is that the trace of a matrix is the sum of the eigenvalues of that matrix, as we will see.

### 3.7 Norms

A *norm* of a vector  $\|\mathbf{x}\|$  is to an approximation a measure of length of  $\mathbf{x}$ . We define the  $\ell - p$  norm as

$$\|\mathbf{x}\|_p = \left( \sum_{i=1}^n x_i^p \right)^{1/p}.$$

The *distance* between two vectors is in an  $\ell - p$  space is  $\|\mathbf{x} - \mathbf{y}\|_p$ . Euclidean distance,  $\|\mathbf{x} - \mathbf{y}\|_2$  is what we normally consider as 'distance'.

There are really 4 cases of the  $\ell - p$  norm that we might encounter:

- $\|\mathbf{x}\|_0$  = (the number of non-zero elements of  $\mathbf{x}$ )
- $\|\mathbf{x}\|_1 = \sum |\mathbf{x}_i|$ . We call  $\|\mathbf{x} - \mathbf{y}\|_1$  the *Manhattan Distance* between  $\mathbf{x}$  and  $\mathbf{y}$  because it treats walking between coordinates in  $\mathbb{R}^n$  as walking on perpendicular streets. This is opposed to Euclidean Distance where you can walk in a strait line from point to point.
- $\|\mathbf{x}\|_2 = \sqrt{\sum \mathbf{x}_i^2}$  is the Euclidean Norm. This is what we normally consider as the 'length' of a vector.
- $\|\mathbf{x}\|_\infty = \max_i |x_i|$ .

We will encounter  $\ell - p$  norms *a lot* in our studies. Specifically, we can improve the robustness of many algorithms to outliers by constraining the norm of some parameter in our optimization. Infinitely many other norms can be defined, however. A norm is *any* function  $f : \mathbb{R}^n \mapsto \mathbb{R}$  that satisfies 4 properties:

1. For all  $\mathbf{x} \in \mathbb{R}^n$ ,  $f(\mathbf{x}) \geq 0$  (non-negativity)
2.  $f(\mathbf{x}) = 0$  if and only if  $\mathbf{x} = \mathbf{0}$  (definiteness)
3. For all  $\mathbf{x} \in \mathbb{R}^n$  and  $t \in \mathbb{R}$ ,  $f(t\mathbf{x}) = |t|f(\mathbf{x})$  (homogeneity)
4. For all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$  (triangle inequality)

Norms can be extended to matrices too, as we will encounter. Here are a few, where  $A \in \mathbb{R}^{m \times n}$ :

1.  $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} = \sqrt{\text{tr}(A^\top A)}$  is the Frobenius Norm of  $A$ .
2.  $\|A\|_* = \text{tr}(\sqrt{A^\dagger A}) = \sum_{i=1}^{\min\{m,n\}} \sigma_i =$  (sum of the singular values of  $A$ ) is the Nuclear Norm
3.  $\|A\|_{\max} = \max_{ij} |A_{ij}|$
4.  $\|A\|_2 = \sqrt{\lambda_{\max}(A^\dagger A)} = \sigma_{\max}(A)$  is the Spectral Norm.