

# NOTES ON LINEAR ALGEBRA

SHASHANK SULE

ABSTRACT. These notes will help you revise the essentials of Linear Algebra. They have been organized in terms of definitions, theorems and techniques.

## 1. DEFINITIONS

### 1.1. Vector Spaces and Subspaces.

**Definition 1.** A vector space over the field  $\mathbb{R}$  is a non-empty set  $V$  which under a binary operation  $+: V \times V \mapsto V$  and a multiplication operation  $\cdot: \mathbb{R} \times V \mapsto V$  has the following properties:

- (1)  $V$  forms a commutative group under addition, i.e.
  - (a) Closure and commutativity:  $\forall \mathbf{v}, \mathbf{u} \in V, \mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v} \in V$
  - (b) Associativity:  $\forall \mathbf{v}, \mathbf{u}, \mathbf{w} \in V, \mathbf{v} + (\mathbf{u} + \mathbf{w}) = (\mathbf{v} + \mathbf{u}) + \mathbf{w}$
  - (c) Existence of identity:  $\exists \mathbf{0} \in V$  s.t.  $\mathbf{0} + \mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{v}$
  - (d) Existence of inverse:  $\forall \mathbf{v} \in V, \exists (-\mathbf{v}) \in V$  s.t.  $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$
- (2) Distributivity with respect to vectors:  $c(\mathbf{u} + \mathbf{v}) = c\mathbf{v} + c\mathbf{u}$
- (3) Distributivity with respect to scalars:  $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$
- (4) Multiplicative identity is 1:  $1\mathbf{v} = \mathbf{v}$

**Definition 2.** A subspace  $S \subseteq V$  is a vector space (with respect to the same operations) contained in  $V$ .  $S$  is a subspace iff

- (1)  $\mathbf{x}, \mathbf{y} \in S \implies \mathbf{x} + \mathbf{y} \in S$
- (2)  $\mathbf{x} \in S \implies c\mathbf{x} \in S$
- (3)  $\mathbf{0} \in S$

Alternately,  $S$  is a subspace iff  $\mathbf{x}, \mathbf{y} \in S \implies c\mathbf{x} + \mathbf{y} \in S$

**Definition 3.** Let  $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ . Then  $\text{Span}(S) = \{c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n \mid c_i \in \mathbb{R} \forall i: 1 \leq i \leq n\}$ . Of course,  $\text{Span}(S)$  is a subspace

**Definition 4.** Non-zero vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  are said to be linearly dependent iff  $\exists c_1, \dots, c_n \in \mathbb{R}$  s.t.  $c_1^2 + \dots + c_n^2 \neq 0$  and  $c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n = \mathbf{0}$ . In other words, we can linearly combine the vectors using scalars, at least one of which is non-zero, s.t. to get the zero element. Another alternate formulation is that we can linearly combine some elements in the set to yield a linear combination of the other elements in the set. Vectors which are not linearly dependent are said to be linearly independent, i.e if we can linearly combine them to yield zero, it must mean that all coefficients are zero. Note that sometimes we abuse notation and say that a set  $S$  of vectors is linearly independent. It should be understood that this is equivalent to saying that the vectors in  $S$  are linearly independent.

**Definition 5.** A set  $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq V$  is said to be a basis for  $V$  iff (i)  $B$  is a set of linearly independent vectors and (ii)  $\text{Span}(B) = V$ . This is the test for a basis. The cardinality of the basis is termed the dimension of  $V$ .

**Definition 6.** Let  $V$  be a vector space over  $\mathbb{R}$  and let  $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{R}$  be a function such that it satisfies

- (1) Commutativity:  $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
- (2) Linearity:  $\langle c\mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle$  and  $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
- (3) Positive definiteness:  $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$  and  $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0}$

Then  $V$  is termed an inner product space.

**Definition 7.** Let  $V$  be a vector space over  $\mathbb{R}$  and  $\|\cdot\| : V \mapsto \mathbb{R}$  a function satisfying the following properties:

- (1) Absolute linearity:  $\|c\mathbf{v}\| = |c| \|\mathbf{v}\|$
- (2) Sub additivity:  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
- (3) Positive definiteness:  $\|\mathbf{v}\| \geq 0$  and  $\|\mathbf{v}\| = 0 \iff \mathbf{v} = \mathbf{0}$

## 1.2. Linear Transformations.

**Definition 8.** Let  $V, W$  be vector spaces. Then the function  $T : V \mapsto W$  is called a linear transformation iff

- (1)  $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$
- (2)  $T(c\mathbf{x}) = cT(\mathbf{x})$

Here  $V$  is called the domain and  $W$  is called the target or the co-domain.

**Definition 9.** The kernel of a linear transformation  $T$ , denoted  $\text{Ker}(T)$  is the set  $\{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}_W\}$ . The image of a linear transformation is the set  $\{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}$ . As always, the kernel lives in the domain and the image lives in the co-domain. The dimension of the kernel is termed as the nullity of  $T$  and the dimension of the image is termed as the rank of  $T$

**Definition 10.** The co-ordinate vector of  $\mathbf{u}$  with respect to a basis  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is denoted

$$[\mathbf{u}]_\beta = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}_\beta$$

where

$$\mathbf{u} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$$

**Definition 11.** Every linear transformation between finite dimensional vector spaces  $T : V \mapsto W$  has an  $m \times n$  matrix representation in the form

$$\begin{bmatrix} T(\mathbf{v}_1)_\alpha & \dots & T(\mathbf{v}_n)_\alpha \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = [T]_\beta^\alpha$$

where  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is a basis for  $V$  and  $\alpha$  is a basis for  $W$ . The linear transformation of  $\mathbf{v}$  with respect to  $\alpha$  is denoted as

$$T(\mathbf{v})_\alpha = [T]_\beta^\alpha [\mathbf{v}]_\beta$$

**Definition 12.** Let  $T : V \mapsto W$  be linear and  $X \subseteq W$ . Then the set

$$T^{-1}(X) = \{\mathbf{v} \in V \mid T(\mathbf{v}) \in X\}$$

is termed the inverse image of  $X$ .

**Definition 13.** Let  $T : V \mapsto V$ . Then the determination of  $T$  is said to be  $\det [T]_\alpha^\alpha$  where  $\alpha$  is a basis for  $V$ .

### 1.3. Determinant theory.

**Definition 14.** The determinant of an  $n \times n$  matrix  $M$  is defined recursively as

$$\det A = \sum_{i=1}^n (-1)^{i+1} \det M_{1i}$$

where  $M_{1i}$  is the  $(n-1) \times (n-1)$  matrix formed by removing the 1st row and the  $i$ th column. More generally,  $\det M_{ij}$  is called the  $(i, j)$ th minor of  $M$ .

The determinant of a  $2 \times 2$  matrix is defined as:

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

This completes the definition of the recursive formula for the determinant of a matrix.

**Definition 15.** The cofactor matrix of an  $n \times n$  matrix is

$$\begin{bmatrix} \det M_{11} & \dots & \det M_{1n} \\ \vdots & \ddots & \vdots \\ \det M_{n1} & \dots & \det M_{nn} \end{bmatrix}$$

**Definition 16.** An equivalent definition of the determinant of  $M$  is that

$$\det A = \sum_{\sigma} \text{sgn}(\sigma) \left( \prod_{i=1}^n a_{i\sigma(i)} \right)$$

where  $\sigma$  is a permutation of the set  $\{1, \dots, n\}$ . Essentially, for a given permutation in  $S_n$ , we take the entry at each row and column specified by the permutation of the row index. We take the product of these entries over all rows and multiply it by the sign of the permutation. We sum such numbers up over all permutations to get the determinant.

### 1.4. Matrix theory.

**Definition 17.** Let  $M$  be an  $n \times n$  matrix. Then  $M$  is said to be invertible if  $\exists$  a matrix  $N$  s.t.  $MN = NM = I_n$  where

$$I_n = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix}$$

Here  $I_n$  is the  $n$ -identity matrix.

**Definition 18.** We say that two square matrices  $A$  and  $B$  are similar or are conjugates of each other if  $\exists$  an invertible matrix  $Q$  such that

$$A = Q^{-1}BQ$$

In group theory terms, the matrices  $A, B \in (M_n(\mathbb{R}))^\times$  are similar if and only if they belong to the same conjugacy class.

**Definition 19.** Let  $M$  be an  $n \times n$  matrix. Then the transpose of  $M$ , denoted  $M^\top$  is the  $n \times n$  matrix whose  $ij$ th entry  $a_{ij} = b_{ji}$  where  $b_{ji}$  is the  $ji$ th entry of  $M$ . Essentially, the tranpose of  $M$  is the matrix reflected across the main diagonal of  $M$ . This leads to the more general definition of a transpose where the transpose of an  $n \times m$  matrix  $A$  is the  $m \times n$  matrix formed by reflected across the diagonal of  $A$ .

**Definition 20.** Let  $A$  be an  $n \times n$  matrix. Then

- (1)  $A$  is symmetric  $\iff A^\top = A$
- (2)  $A$  is skew-symmetric  $\iff A^\top = -A$
- (3) If  $A$  is complex then  $A$  is Hermitian  $\iff A^\top = \bar{A} \iff \overline{A^\top} = A$
- (4) If  $A$  is complex then  $A$  is skew-Hermitian  $\iff A^\top = -\bar{A} \iff \overline{A^\top} = -A$
- (5)  $A$  is orthogonal  $\iff A^\top = A^{-1}$
- (6) If  $A$  is complex then  $A$  is unitary  $\iff A^\top = \overline{A^{-1}}$

**Definition 21.** Every matrix represents a linear transformation. The kernel of the linear transformation represented by  $M$  is termed the null-space of  $M$ . Similarly, the image of the linear transformation represented by  $M$  (i.e the space spanned by the columns of  $M$ ) is termed the column space of  $M$ . Alternately, the space spanned by the rows of  $M$  is termed the row space of  $M$ . The dimension of the null-space is called the nullity and the dimension of the column space is called the rank. (Technically it is called the column rank, but there is a result that states that column rank equals row rank).

**Definition 22.** Let  $M$  be an  $n \times m$  matrix and  $\mathbf{v} \in \mathbb{R}^m$  s.t.

$$M\mathbf{v} = \lambda\mathbf{v}$$

Then  $\mathbf{v}$  is termed an eigenvector and  $\lambda$  is termed an eigenvalue.

**Definition 23.** If  $M$  is a square matrix then the expression

$$P(\lambda) = \det(M - \lambda I)$$

is termed the characteristic polynomial of  $M$ . The solutions to the characteristic polynomial yields all the eigenvalues of a matrix.

**Definition 24.** Let  $\mathcal{R}$  be the multi-set of the roots of  $P$ , the characteristic polynomial. Then the number of repetitions of a given eigenvalue  $\lambda$  in  $\mathcal{R}$  is said to be the algebraic multiplicity of  $\lambda$ . Furthermore, let  $E_\lambda$  be the set of eigenvectors corresponding to the eigenvalue  $\lambda$ .  $E_\lambda$  is said to be a  $\lambda$ -eigenspace and its dimension is termed as the geometric multiplicity of  $\lambda$ .

**Definition 25.** A square matrix  $M$  is said to be diagonalizable  $\iff$  it is similar to a diagonal matrix.

1.4.1. *Gaussian Elimination.*

**Definition 26.** The elementary row operations on a matrix  $M$  are:

- (1) Swapping two rows with each other.
- (2) Adding a scalar multiple of a row to another row.
- (3) Multiplying a row with a scalar

Two matrices  $M$  and  $N$  are said to be row equivalent if we can obtain one from the other using elementary row operations.

**Definition 27.** A matrix  $M$  is said to be in echelon (or row reduced form) if after a finite set of row operations we obtain a matrix  $E$  s.t.

- (1) In the  $i$ th row of  $E$ , at least the first  $i - 1$  columns have zeros.
- (2) The first non-zero entry in every row is 1.

In echelon form, the variable  $x_i$ , corresponding to the column index  $i$  which has a 1 in its entry with a given row, is said to be a leading variable.

**Definition 28.** The process of reducing a matrix into echelon form using elementary row operations is called Gaussian elimination. Look at techniques to see the applications of this method.

## 2. THEOREMS

2.1. **Vector Spaces and Subspaces.**

**Result 1.**  $0\mathbf{x} = \mathbf{0} \forall \mathbf{x} \in V$  where  $V$  is a vector space.

**Result 2.** If  $U_1$  and  $U_2$  are subspaces of vector space  $V$  then  $U_1 \cap U_2$  is a subspace of  $V$

**Result 3.** Let  $V$  be a vector space and  $U$  a subspace. Then if  $S = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subseteq U$  then  $\text{Span}(S) \subseteq U$

**Result 4.** Let  $S = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subseteq V$  where  $V$  is a vector space. Then  $\text{Span}(S)$  is a subspace of  $V$ .

**Result 5.** If  $S \subseteq V$  and  $\mathbf{0} \in S$  then the vectors in  $S$  are linearly dependent.

**Result 6.** Let  $V$  be a vector space and  $S, S' \subset V$  where  $S \subseteq S'$ . Then

- (1) If  $S'$  is linearly independent then so is  $S$ .
- (2) If  $S$  is linearly dependent then so is  $S'$

**Result 7.**  $\{\mathbf{x}, \mathbf{y}\}$  is linearly independent  $\iff \{\mathbf{x} - \mathbf{y}, \mathbf{x} + \mathbf{y}\}$  is linearly independent.

**Result 8.** Let  $S$  be a linearly dependent spanning set of  $V$ . Then  $\exists \mathbf{x} \in S$  such that  $\mathbf{x}$  is a linear combination of the rest of the elements and  $\text{Span}(S - \{\mathbf{x}\}) = V$ . This proves the second method of finding a basis (see Techniques).

**Lemma 1.** Let  $V$  be a vector space and  $S$  a linearly independent subset which does not span  $V$ . Then  $\exists \mathbf{x} \in V$  s.t.  $S \cup \{\mathbf{x}\}$  is linearly independent.

**Result 9.** Let  $V$  be a vector space and  $S$  a linearly independent set of  $V$ . Then  $\exists$  a basis  $B$  of  $V$  s.t.  $S \subseteq B$

**Corollary 2.** Every finite dimensional vector space has a basis.

**Corollary 3.** Let  $W$  be a subspace of  $V$ . Then  $\dim(W) \leq \dim(V)$  and  $\dim(W) = \dim(V) \iff W = V$ .

**Result 10.** Let  $V$  be a vector space where  $S$  is linearly independent and  $S'$  is a spanning set. Then  $|S| \leq |S'|$

**Corollary 4.** Let  $B_1$  and  $B_2$  be bases of  $V$ . Then  $|B_1| = |B_2| = \dim(V)$ .

**Corollary 5.** Let  $V$  be an  $n$ -dimensional vector space and  $S$  a linearly independent set of  $n$  vectors. Then  $S$  is a basis. (This corollary essentially shows that for linearly independent sets, having  $n$  elements is equivalent to being a basis)

## 2.2. Linear Transformations.

**Result 11.** The kernel of any linear transformation is a subspace.

**Result 12.** Let  $T : V \mapsto W$  be a linear transformation and  $\beta' = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  be a basis for  $\text{Ker}(T)$ . Then  $\exists$  basis  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$  of  $V$  that contains  $\beta'$ . In this case,  $T(\beta - \beta') = \{T(\mathbf{v}_{k+1}), \dots, T(\mathbf{v}_n)\}$  is a basis for  $\text{Im}(T)$ .

**Theorem 6. The Rank-Nullity theorem.** Let  $T : V \mapsto W$  be a linear transformation. Then

$$\dim(\text{Ker}(T)) + \dim(\text{Im}(T)) = \dim(V)$$

**Result 13.** Let  $T : V \mapsto W$  be linear, and  $\dim(V) = \dim(W)$ . Then  $T$  is injective  $\iff T$  is surjective. This result is important because if you want to check bijectivity of a linear transformation, then in the special case that both the domain and co-domain have the same dimension, you only need to check either injectivity or surjectivity.

**Result 14.** (This is a partial converse of the previous result). Let  $T : V \mapsto W$  be a bijective linear transformation. Then  $\dim(V) = \dim(W)$ .

**Result 15.** Let  $T : V \mapsto W$  be a linear transformation and let  $\mathbf{w} \in \text{Im}(T)$ . Suppose  $T(\mathbf{v}) = \mathbf{w}$  for some  $\mathbf{v} \in V$ . Then  $T^{-1}(\{\mathbf{w}\}) = \mathbf{v} + \text{Ker}(T)$ . This theorem implies that if we want to find out the inverse image of a singleton vector, it is sufficient to know just one vector in the domain that maps to the vector we are given.

## 2.3. Determinant theory.

**Theorem 7.** (1) If the matrix  $B$  is determined from  $A$  by:

- (a) Adding a multiple of one row to another then  $\det B = \det A$
- (b) Swapping two rows then  $\det B = -\det A$
- (c) Multiplying a row by a scalar then  $\det B = k \det A$
- (2) If  $A$  has two identical columns, then  $\det A = 0$
- (3) If  $A$  has a row of zeros, then  $\det A = 0$
- (4) If  $A$  is upper triangular, then  $\det A = a_{11}a_{22} \dots a_{nn}$

**Result 16.** The  $n$ -dimensional volume of the parallelepiped formed by the columns of a matrix is given by the determinant of the matrix.

## 2.4. Matrix theory.

**Theorem 8.** Let  $A$  be a square matrix over  $\mathbb{R}$ . The following statements are equivalent:

- (1)  $A$  is invertible, that is,  $A$  has an inverse, is nonsingular, or is nondegenerate.
- (2)  $A$  is row-equivalent to the identity matrix  $I_n$ .
- (3)  $A$  is column-equivalent to the identity matrix  $I_n$ .
- (4)  $A$  has  $n$  pivot positions.

- (5)  $\det A \neq 0$ . In general, a square matrix over a commutative ring is invertible if and only if its determinant is a unit in that ring.
- (6)  $A$  has full rank; that is,  $\text{rank}(A) = n$ .
- (7) The equation  $A\mathbf{x} = \mathbf{0}$  has only the trivial solution  $\mathbf{x} = \mathbf{0}$ .
- (8)  $\text{Ker}(A) = \{\mathbf{0}\}$
- (9) The equation  $A\mathbf{x} = \mathbf{b}$  has exactly one solution for each  $\mathbf{b}$  in  $\mathbb{R}^n$ .
- (10)  $\text{Col}(A) = \mathbb{R}^n$ .
- (11) The columns of  $A$  form a basis of  $\mathbb{R}^n$ .
- (12) The linear transformation represented by  $A$  is a bijection.
- (13) There is a square matrix  $B$  such that  $AB = I_n = BA$ .
- (14) The transpose  $A^\top$  is an invertible matrix (hence rows of  $A$  form a basis of  $\mathbb{R}^n$ ).
- (15) 0 is not an eigenvalue of  $A$ .

**Theorem 9.** The matrix  $A$  is diagonalizable iff:

- (1) The characteristic polynomial of  $A$  completely factors into linear terms.
- (2) The eigenvectors of  $A$  form a basis for  $\mathbb{R}^n$
- (3) If  $\lambda_1, \dots, \lambda_n$  are the eigenvalues of  $A$  and  $E_{\lambda_i}$  are the corresponding eigenspaces, then the geometric multiplicity of  $\lambda_i$  equals its algebraic multiplicity.
- (4) For every eigenvalue, algebraic multiplicity equals the geometric multiplicity.

$$\sum_{i=1}^n \dim(E_{\lambda_i}) = n$$

If  $A$  has distinct eigenvalues, then it is diagonalizable. The converse, however, is not true (it may have an eigenvalue with higher algebraic multiplicity which could be equal to its geometric multiplicity)

### 3. TECHNIQUES

#### 3.1. Vector spaces and subspaces.

**Technique 1.** How to calculate the basis of a vector space/subspace:

- (1) Find a linearly independent set smaller than the dimension, and extend it until it matches the dimension.
- (2) Find a linearly dependent set larger than the dimension, and reduce it until it matches the dimension.
- (3) Figure out a spanning set for the subspace from the definition of the subspace (this is useful typically for  $\mathbb{R}^n$  and  $P_n(\mathbb{R})$ , the vector space of  $n$ -degree polynomials).

#### 3.2. Linear Transformations.

**Technique 2.** How to calculate:

- (1) The kernel of a linear transformation
  - (a) Find the matrix  $A$  of the linear transformation
  - (b) Solve  $A\mathbf{x} = \mathbf{0}$
  - (c) The solution set is the kernel.
- (2) The image of a linear transformation
  - (a) Find the kernel and then find a basis for it.
  - (b) Extend the kernel basis to a basis for the domain.
  - (c) Calculate the linear transformations of the extra elements you added.

(d) This is a basis for the image.

Alternately, you could also take the column vectors of the matrix of the transformation and reduce the set of vectors until they become linearly independent. (Remember! The columns of a matrix and the image of the transformation are intimately related; after all, we form the matrix from the co-ordinate vectors of the vector-by-vector images of the basis vectors from the domain).

- (3) Find the inverse image of a linear transformation (For a singleton vector)
- Find the solution to  $Av = w$ , or
  - Find one vector that satisfies  $Av = w$ , compute the kernel and then the inverse image is  $v + \text{Ker}(T)$ .

### 3.3. Determinants.

**Technique 3.** How to compute a determinant:

- Try to fit the matrix into one of the categories of Theorem 7, i.e. identify if it is upper or lower triangular, has repeated rows or columns, if we could swap rows to get a matrix whose determinant we already know, etc.
- If that isn't possible, row reduce the matrix and record each time you multiply a row by a constant. The determinant of final matrix multiplied by the reciprocal of the product of all these constants will be the determinant of the original matrix. (Of course, a matrix in echelon form IS an upper triangular matrix)

### 3.4. Matrices.

#### 3.4.1. Applications of Gaussian Elimination and determinants.

**Technique 4.** Solving linear equations of the form:

$$\begin{aligned} a_{11}x_1 + \dots + a_{1m}x_m &= c_1 \\ &\vdots \\ a_{n1}x_1 + \dots + a_{nm}x_m &= c_n \end{aligned}$$

Note that this is equivalent to solving

$$\begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

There are two ways to do this, one using elimination and one using determinants.

- (1) *Using elimination:* Set up the augmented matrix

$$\left[ \begin{array}{ccc|c} a_{11} & \dots & a_{1m} & c_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \dots & a_{nm} & c_n \end{array} \right]$$

Now apply Gaussian elimination to it. After applying gaussian elimination, the column indices (or the variables) which correspond to zeros are the free variables, and the column indices which correspond to 1 are the leading variables.

(2) *Using determinants*: This process is called Cramer's rule. Let

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

where  $A$  is invertible. Then according to Cramer's rule,

$$x_i = \frac{\det A_i}{\det A}$$

where  $A_i$  is the matrix formed by replacing the  $i$ th column with  $\mathbf{b}$ . (Note that this is well-defined, because  $A$  is clearly a square matrix).

**Technique 5.** Suppose we want to invert a  $3 \times 3$  matrix:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

There are two ways to compute this, one using Gaussian elimination and one using determinants.

(1) *Using Gaussian elimination*: Set up the augmented form

$$\left[ \begin{array}{ccc|ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array} \right]$$

Now do gaussian elimination until we get the echelon form:

$$\left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & 1 & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 & b_{31} & b_{32} & b_{33} \end{array} \right]$$

The matrix on the right side of the separator is the inverted matrix!

(2) *Using determinants*: The inverse of a matrix  $M$  is

$$M^{-1} = \frac{1}{\det M} C^T$$

where  $C$  is the cofactor matrix of  $M$ .

**3.4.2. Matrix multiplication.** Matrix multiplication is simple enough, but there are 2 useful block-wise techniques you should know:

**Technique 6.**

$$A[\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n] = [A\mathbf{v}_1 \ A\mathbf{v}_2 \ \dots \ A\mathbf{v}_n]$$

and

$$c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n][c_1 \ \dots \ c_n]^T$$

This technique basically says that you can move between linear combinations and matrix multiplications. However, don't be too stressed about remembering these rules; you can always compute a small example (typically in 3D) when in doubt about which rule to apply.

**Technique 7.** The above technique can be very nicely (and powerfully) generalized to blocks of matrices, where if  $A$  has  $n$  row partitions and  $m$  column partitions, and  $B$  has  $m$  row partitions and  $k$  column partitions then  $AB$  can be found by just standard matrix multiplication, where we treat the blocks as entries, i.e

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nm} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1k} \\ B_{21} & B_{22} & \dots & B_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \dots & B_{mk} \end{bmatrix} = AB$$

where  $[AB]_{ij} = \sum_{l=1}^m A_{il}B_{lj}$ . This is the celebrated Cauchy-Binet formula and it is well-defined because we assume the same partitions of columns in  $A$  and rows in  $B$ .

### 3.4.3. Diagonalization and its applications.

**Technique 8.** Diagonalizing a matrix:

- (1) First find the eigenvalues of  $M$  by finding the roots of the characteristic polynomial.
- (2) Now if  $\lambda$  is a root, i.e an eigenvalue, solving for  $\mathbf{v}$  in the equation

$$(A - \lambda I)\mathbf{v} = \mathbf{0}$$

gives  $E_\lambda$ . Do this for every eigenvalue that you find, computing eigenspaces and their basis vectors.

- (3) Once you find the basis vectors of all the eigenspaces, line them up in a matrix,  $P$ . Note that for eigenspaces with multiplicity greater than 1, in essence when they have more than one basis vectors, make sure to line up the basis vectors for the eigenspace next to each other. To construct the diagonal matrix, start from the left top corner and proceed down the diagonal. Put the eigenvalue corresponding to the eigenvector in the  $i$ th column of  $P$  in the  $i$ th place in the diagonal matrix.
- (4) Thus we get  $M = PDP^{-1}$

**Technique 9.** There are three applications of diagonalizability

- (1) *Computing powers of a matrix:* Computing the power of a matrix is hard, because matrix multiplication isn't straightforward. We would like to reduce  $M$  to a matrix for which it is easy to compute an  $n$ th power. For example, the  $n$ th power of a diagonal matrix is just the  $n$ th powers of its diagonal entries. So, If we want to compute  $M^n$ , then first diagonalize  $M$  as

$$M = PDP^{-1}$$

Then  $M^n = PDP^{-1}PDP^{-1} \dots PDP^{-1} = PD^nP^{-1}$ . Now we have  $M^n$  expressed as a product of a matrices we already know ( $P$  and  $P^{-1}$ ) along with the  $n$ th power of a diagonal matrix, which is easy to compute. Done!

- (2) *Linear systems of differential equations:* Suppose we have the differential equations

$$\begin{aligned}\frac{dL}{dt} &= aL + bM \\ \frac{dM}{dt} &= cL + dM\end{aligned}$$

Then how do we solve for  $L$  and  $M$  as functions of  $T$ ? Simple. Just diagonalize and decouple!

(a) Decoupling: Express the differential equation as

$$\frac{d}{dt} \begin{bmatrix} L \\ M \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} L \\ M \end{bmatrix}$$

(b) Diagonalize the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = PDP^{-1}$$

So the system now becomes

$$\begin{aligned}\frac{d}{dt} \begin{bmatrix} L \\ M \end{bmatrix} &= PDP^{-1} \begin{bmatrix} L \\ M \end{bmatrix} \\ \implies \frac{d}{dt} P^{-1} \begin{bmatrix} L \\ M \end{bmatrix} &= DP^{-1} \begin{bmatrix} L \\ M \end{bmatrix}\end{aligned}$$

(c) Set

$$P^{-1} \begin{bmatrix} L \\ M \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Then our equation becomes

$$\frac{d}{dt} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

This is equivalent to solving

$$\frac{dx}{dt} = k_1x \text{ and } \frac{dy}{dt} = k_2y$$

We have successfully decoupled the system! The solutions are

$$x = Ce^{k_1t} \text{ and } y = De^{k_2t}$$

Now, going back to how the original  $\begin{bmatrix} x \\ y \end{bmatrix}$  was defined,

$$\begin{bmatrix} x \\ y \end{bmatrix} = P^{-1} \begin{bmatrix} L \\ M \end{bmatrix} \implies P \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} L \\ M \end{bmatrix} \implies P \begin{bmatrix} Ce^{k_1t} \\ De^{k_2t} \end{bmatrix} = \begin{bmatrix} L \\ M \end{bmatrix}$$

Plug in the initial values and we can solve for  $C$  and  $D$ !

- (3) *Computing long term probabilities in discrete-time Markov Chains:* A discrete-time Markov chain describes the probability of a random variable through some function of the probability at a previous state in time. For example,

$$\begin{bmatrix} S_n \\ C_n \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} S_{n-1} \\ C_{n-1} \end{bmatrix}$$

is a Markov chain.

Clearly, when we have initial values,

$$\begin{bmatrix} S_n \\ C_n \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}^n \begin{bmatrix} S_0 \\ C_0 \end{bmatrix}$$

Thus, to find these values, we diagonalize the probability matrix  $\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$ .