1 Linear Algebra - Introduction

1.1 Vector Space

Let $\mathcal{F}$ be a field, $\mathcal{V}$ a vector space over $\mathcal{F}$. We will often take $\mathcal{F} = \Re$, $\mathcal{V} = \Re^n$; or alternatively $\mathcal{F} = \mathcal{Z}_p \equiv \{0, \ldots, p-1\}$ for a prime number p, $\mathcal{V} = \mathcal{Z}_p^n$.
The most basic requirements on $\mathcal{V}$ include:

1. closure under vector addition
2. closure under scalar multiplication (multiplication by elements in $\mathcal{F}$)

We define the *dimension* of $\mathcal{V} = dim(\mathcal{V})$ as the minimal number of vectors which (linearly) span $\mathcal{V}$. Such a set of vectors if known as a *basis set* of $\mathcal{V}$ and can be shown to be linearly independent. A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is said to be *linearly dependent* iff there exist scalars $a_1, \ldots, a_k$ (where at least one $a_i \neq 0$) s.t.

$$a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = 0$$

Otherwise, the set is said to be *linearly independent.*

1.2 Matrices

Another structure defined over $\mathcal{V}$ is the *matrix.* A matrix is a compact structure often used for the representation of linear transformations. Specifically, a linear transformation T s.t. $T : \Re^n \to \Re^m$ can be fully described by a unique matrix of dimension $m \times n$, $A_{m\times n}$:

$$\forall \mathbf{x} \ \epsilon \ \Re^n \ T(\mathbf{x}) = A\mathbf{x}$$

Thus, linear transformations and matrices can be thought of interchangeably. For a given vector space basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, T and A are uniquely defined by the values $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ and $A\mathbf{v}_1, \ldots, A\mathbf{v}_n$, respectively. This is so since any vector $v \ \epsilon \mathcal{V}$ can be represented as

$$c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n$$

so then due to linearity:

$$A\mathbf{v} = \sum_{i=1}^{n} c_i A\mathbf{v}_i$$

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over $\mathcal{F}$, and let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then the following equality holds:

$$dim(\mathcal{V}) = dim(Ker(T)) + dim(Im(T)) \tag{1}$$

where

$$Ker(T) \equiv \{\mathbf{v} \in \mathcal{V} : T(\mathbf{v}) = 0\}$$

and

$$Im(T) \equiv \{\mathbf{w} \in \mathcal{W} : \exists\ \mathbf{v} \in \mathcal{V}, T(\mathbf{v}) = \mathbf{w}\}$$

The value $dim(Im(T))$ is known as the *rank* of T, and is equivalent to the rank of the matrix A, which represents T. The *rank* of a matrix A can be defined in three equivalent ways:

1. the *row rank* of A. The row rank is the largest number of linearly independent row vectors of A.

2. the *column rank* of A. The column rank is the largest number of linearly independent column vectors of A.

3. the minimal number of matrices of rank 1, whose sum equals A. Matrix B of rank 1 is the result of the outer product of two vectors $\mathbf{x}$, $\mathbf{y}$. i.e., $B_{ij} = x_i y_j$, or equivalently $B = \mathbf{x} \otimes \mathbf{y} = \mathbf{x}\mathbf{y}^\top$.

1.3 Eigenvalues and Eigenvectors

Definition 1 *Let $A \in M_n(\mathbb{C})$ be a square matrix of size $n \times n$ with complex entries. We say that $\mathbf{x}$ is an eigenvector of A, with eigenvalue λ iff:*

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

1.3.1 Characteristic Polynomial

$A\mathbf{v} = \lambda\mathbf{v}$ iff $(A - \lambda I)\mathbf{v} = 0$. Such a vector exists iff $det(A - \lambda I) = 0$. The *characteristic polynomial* of A is defined as (tought as polynomial in λ):

$$p(\lambda) = det(\lambda I - A) \tag{2}$$

The roots of this polynomial are the eigenvalues of A. The below claims follow easily:

1. A has at most n eigenvalues of A (since they are the roots of an n^{th} degree polynomial).

2. A and A^T have the same eigenvalues (since the determinant is invariant to the transpose operation).

1.4 Eigenvector basis / Diagonalization

Lemma 2 *Let* $\mathbf{v}_1, \ldots, \mathbf{v}_r$ *be eigenvectors of* A *with corresponding distinct eigenvalues* $\lambda_1, \ldots, \lambda_r$. *Then* $\mathbf{v}_1, \ldots, \mathbf{v}_r$ *are linearly independent.*

Proof Consider a minimal linear dependence:

$$\sum_{i=1}^{k} c_i \mathbf{v}_i = 0. \tag{3}$$

Application of A to both sides yields:

$$0 = A(0) = A\left(\sum_{i=1}^{k} c_i \mathbf{v}_i\right) = \sum_{i=1}^{k} c_i \lambda_i \mathbf{v}_i \tag{4}$$

Multiply equation (3) by λ_1 and subtract from equation (4) to derive a shorter non-trivial linear dependency among the v_i, a contradiction.

$\blacksquare$

Thus, if A has n distinct eigenvalues, then its eigenvectors form a basis of eigenvectors of $\Re^n$.

1.4.1 Matrix similarity

Definition 3 *A is similar to B if there exists a non-singular matrix P s.t. $A = PBP^{-1}$.*

Definition 4 *A is diagonalizable if it is similar to a diagonal matrix.*

If there is a basis of eigenvectors P then $AV = V\Lambda$ where the $i'th$ column of V is the $i'th$ eigenvector $\mathbf{v}_i$, and Λ is a diagonal matrix where $\Lambda_{ii} = \lambda_i$ is the $i'th$ eigenvalue of A. Equivalently $A = V\Lambda V^{-1}$ and we say that A is **similar** to a diagonal matrix Λ. Geometrically, if there exists a basis V of eigenvectors of A, then A simply performs contractions and expansions of the vectors in the basis V, whence we can describe its behavior on arbitrary vector in the vector space.

Example: A Matrix that cannot be diagonalized

not always is there a vector space basis of eigenvectors. For example, for the matrix

$$C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

the only eigenvalue is 0. However, the eigenspace V_0 is spanned solely by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Thus, here, there is no basis of $\Re^2$ composed of eigenvectors of C. Nevertheless for every matrix in $C \in M_n(\mathbb{C})$ we can find a matrix C_ϵ that has n distinct and is ϵ close to C. Therefore we generally assume that the matrices we deal with have distinct eigenvalues.

1.4.2 Unitary equivalence

We state some definitions and facts which we will use in the next lectures.

Definition 5 *A matrix U is called unitary if $UU^\star = I$, where $U^\star = \overline{U^\top}$*

Definition 6 *A is unitary equivalent to B if there exists a matrix unitary matrix U s.t. $A = UBU^\star$.*

Theorem 7 *For complex-valued matrices, $A^*A = A^*A \iff A = U\Lambda U^T$, for a unitary matrix U and diagonal matrix Λ. Such a matrix A is called normal.*

Theorem 8 *If a complex-valued matrix A is Hermitian (symmetric) then all of its eigenvalues are real.*

Definition 9 *Matrix A is positive semidefinite if it is a Hermitian matrix all of whose eigenvalues are non-negative.*

Definition 10 *Matrix A is positive definite if it is a Hermitian matrix all of whose eigenvalues are positive.*

2 Metric Structures defined on the vector space

Besides the concepts of linear dependencies, vector bases, and ranks, there are additional structures that can be defined over the space:

2.1 Metric Space

A *metric space* is set $\mathcal{X}$ with a distance function d s.t. for every two elements $\mathbf{v}, \mathbf{x} \in \mathcal{X}$, $d(\mathbf{v}, \mathbf{x})$ is the non-negative real number to be thought as "distance" between $\mathbf{v}$ and $\mathbf{x}$. d must also satisfy the following:

1. Positivity: $d(\mathbf{v}, \mathbf{x}) \geq 0$ for every $\mathbf{v}, \mathbf{x} \in \mathcal{X}$, and $d(\mathbf{v}, \mathbf{x}) = 0$ iff $\mathbf{v} = \mathbf{x}$
2. Symmetry: $d(\mathbf{v}, \mathbf{x}) = d(\mathbf{x}, \mathbf{v})$
3. Triangle Inequality: $\forall\ \mathbf{y}, \mathbf{x}, \mathbf{v} \epsilon \mathcal{V}\ d(\mathbf{v}, \mathbf{x}) \leq d(\mathbf{v}, \mathbf{y}) + d(\mathbf{y}, \mathbf{x})$

2.2 Normed Spaces (Banach Spaces)

in many cases it is useful to associate a "length" to a vector in the space. This length is called a norm. The norm is defined of a vector space $\mathcal{V}$. The norm of $\mathbf{v}$ is denoted as $||\mathbf{v}||$. A norm must satisfy the following conditions:

1. Linearity: $\forall \mathbf{v} \epsilon \mathcal{V},\ \lambda \epsilon \Re\ ||\lambda \mathbf{v}|| = |\lambda|\ ||\mathbf{v}||$
2. Triangle Inequality: $\forall\ \mathbf{u}, \mathbf{v} \epsilon \mathcal{V}\ ||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$

3. Positivity: $\forall\ \mathbf{v} \ \epsilon\ \mathcal{V}\ ||\mathbf{v}|| \geq 0,$ and $||\mathbf{v}|| = 0$ iff $\mathbf{v} = 0$

Examples of norms: The norm $||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ is called the $l_p - norm$. The $l_\infty - norm$ is defined to by $||x||_\infty = \max_i |x_i|$.
A normed space is also a metric since $d(\mathbf{v}, \mathbf{x}) = ||\mathbf{v} - \mathbf{x}||$ yield a metric d.

2.3 Inner Product Space (Hilbert Space)

In a vector space we define the *inner product* of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_i u_i \overline{v_i}$$

By defining the norm of $\mathbf{v}$ as:

$$||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \tag{5}$$

We turn $\mathbb{C}^n$ into a normed space (and hence a metric space). This specific (and important) norm is called the $l_2 - norm$. It satisfy the *parallelogram law*:

$$||\mathbf{x} + \mathbf{y}||^2 + ||\mathbf{x} - \mathbf{y}||^2 = 2||\mathbf{x}||^2 + 2||\mathbf{y}||^2$$

the converse also holds: every normed space which satisfies the *parallelogram law* is an inner product space.

2.3.1 Matrix Norm

Similarly to vectors, we would like a definition of the "size" of a matrix. One application using such quantities is *Singular Value Decomposition* (SVD), which has as input a matrix $A_{m \times n}$ and a natural number r. As output, we would like a matrix $B_{m \times n}$ s.t. $rank(B) \leq r$ and B is "close" to A. For these tasks, we need to define a norm on matrices. There are quite a few possibilities as to how to define such a norm:

1. Consider a matrix to simply be a 2-dimensional vector array and use any known vector norms. For example: Frobenius Norm = Hilbert-Schmidt Norm = $|| \cdot ||_F$, where

$$||A||_F = \sqrt{\sum_{i,j} a_{ij}^2}$$

2. When matrices are viewed as linear transformations, then a natural question to ask is "to what extent does the transformation expand/contract vector lengths?" This of course requires a previous definition of vector lengths. For example, the ℓ_p norms thus define the following family of corresponding matrix norms:

$$||A||_{p \to q} \equiv \max_{\mathbf{x} \in \mathbb{C}^n} \frac{||A\mathbf{x}||_q}{||\mathbf{x}||_p} = \max_{||\mathbf{x}||_p = 1} ||A\mathbf{x}||_q$$