**Mathematical Modelling for Computer Networks- Part I** **Spring 2013 (Period 4)**

# Math Appendix 1: Vectors, Matrices, and Linear Algebra

*Lecturers: Laila Daniel and Krishnan Narayanan* *Date:11th March 2013*

*Neo: What is the Matrix?*
*Trinity: The answer is out there, Neo, and it's looking for you,*
*and it will find you if you want it to.*
*- From the film, The Matrix (1999)*

**Real Numbers**
The set of *real numbers* is denoted by $\mathbb{R}$ and this set includes both positive and negative numbers and zero. We use the notation $\mathbb{R}^+$ (or $\mathbb{R}_+$) to denote the set of real numbers which are greater than or equal to zero. $\mathbb{R}^+$ is known as the set of *positive reals*. Similarly $\mathbb{R}^-$ denotes the set of real numbers which are less than or equal to zero. $\mathbb{R}^-$ is known as the set of *negative reals*. To refer to a real number which is greater than zero, we call it *strictly positive*. When $x < 0$ $x$ is *strictly negative*.
Symbolically

$$\mathbb{R}^+ = \{x \in \mathbb{R} \mid x \geq 0\}$$

$$\mathbb{R}^- = \{x \in \mathbb{R} \mid x < 0\}$$

We use the notation $x \in \mathbb{R}$ to denote that $x$ is a real number. Here we have used the notation $\in$ which denotes set membership.

$\mathbb{R}^n$ denotes the set of ordered n-tuples of real numbers. Analogous to the notation introduced above $\mathbb{R}^n_+$ ($\mathbb{R}^n_-$) denotes the set of ordered n-tuples of positive (negative) real numbers.
Real numbers are sometimes called *scalars* especially in the context of vector space.

**Vector**
A *vector* $\mathbf{x}$ is a finite, ordered collection of real numbers $x_1, x_2, \ldots .x_n$. In other words, $\mathbf{x}$ is a real-valued function $(x_i)$ defined on a finite set of integers $i = 1, \ldots, n$. The numbers $x_i$ are called the components of $\mathbf{x}$ .

**Inner product of two vectors**
Let $\mathbf{x}$ and $\mathbf{y}$ are two $n$-component vectors. The *inner product* of $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 \ + \ x_2 y_2 \ + \cdots + x_n y_n$$

The **Euclidean length of a vector**: The *Euclidean length* of $\mathbf{x}$ is

$$|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 \ + \cdots + x_n^2}$$

**Column vector and row vector**
A vector $\mathbf{x}$ can be represented as a *column* vector $x$ (not in boldface) or by a *row* vector $x^T$ (x transpose). A column vector is a matrix with one column and a row vector is a matrix with one row.

Let **x** has two components $x_1 = 1$ and $x_2 = 2$. Then **x** can be represented as a column vector

$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

or by the row vector

$$x^T = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

**Matrix**
A $m \times n$ *matrix* $A$ is a rectangular array of numbers $a_{ij}$ with $m$ rows $(1, \ldots, m)$ and $n$ columns $(1, \ldots, n)$. Let matrix $A$ has $m = 2$ and $n = 3$, then

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Here $a_{12} = 2$ and $a_{23} = 6$.

The *transpose* of matrix $A$ is the matrix denoted by $A^T$ obtained by interchanging the rows and columns of $A$.

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

In general, if $A = (a_{ij})$, then $A^T = (b_{ij})$ with $b_{ij} = a_{ji}(i = 1, \ldots, m; j = 1, \ldots, n)$

**Linear equation**
An example of a *linear equation* in three variables is $x_1 + 2x_2 + 3x_3 = 4$.

**System of linear equations**
A *system of linear equations* can be denoted as

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \qquad (i = 1, \ldots, m)$$

For each $i$ there is a linear equation with n variables. For example, the first linear equation can be given as

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1$$

The above system of linear equations may be written in a compact form $Ax = b$ where $A$ is an $m \times n$ (m rows and n columns) matrix, **x** is an $n \times 1$ column vector and **b** is an $m \times 1$ column vector.

**Matrix operations**: Some useful matrix operations are described next.

**Scalar multiplication**
$A = (a_{ij})$ denotes a matrix with a typical component $a_{ij}$. Then $\lambda A$ has a component $\lambda a_{ij}$, here $\lambda$ is an arbitrary real number.

**Matrix addition**
If $A$ and $B$ are both matrices with $m$ rows and $n$ columns,then $A + B$ is again an $m \times n$ matrix with components $c_{ij} = a_{ij} + b_{ij}$

**Matrix multiplication**
If $A$ is a $p \times q$ matrix and $B$ is a $q \times r$ matrix then their product matrix $AB$ denoted by $C$ is a $p \times r$ matrix with components

$$c_{ij} = \sum_{k=1}^{q} a_{ik}b_{kj}$$

Therefore $(ij)$th element of the matrix $AB$ can be regarded as the inner product of the $i$th row of $A$ with the $j$th column of $B$.
Note that the matrix product $AB$ of the matrices $A$ and $B$ can be defined only when the number of columns of $A$ is equal to the number of rows of $B$. In general the matrix product $AB$ for matrices $A$ and $B$ can be definable whereas the product $BA$ may not be even definable. It is an important fact in matrices that even in the case when both the product matrices $AB$ and $BA$ can be defined, they may not be equal. This property is described by the statement "Matrix product is not commutative".

Example:
Let matrix $A = \begin{bmatrix} 6 & 9 \end{bmatrix}$, then $A^T = \begin{bmatrix} 6 \\ 9 \end{bmatrix}$

$$AA^T = \begin{bmatrix} 117 \end{bmatrix} \quad \text{but } A^T A = \begin{bmatrix} 36 & 54 \\ 54 & 81 \end{bmatrix}$$

Matrix product is associative, i.e., $(AB)C = A(BC)$

**Linear Combination of vectors**:
Given the vectors $\mathbf{x}^1, \ldots, \mathbf{x}^k$, all of which have $n$ components, then their linear combination $y$ is a vector of $n$ components which can be written as

$$\mathbf{y} = c_1 \mathbf{x}^1 \; + \; \cdots + \; c_k \mathbf{x}^k$$

with components

$$y_i = \sum_{j=1}^{k} c_j x_i^{(j)}$$

Example: A column vector can be written as a linear combination of two column vectors.

$$\begin{bmatrix} 4 \\ -7 \end{bmatrix} = 5 \begin{bmatrix} 2 \\ 3 \end{bmatrix} - 2 \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

**Linear Independence**
The vectors $\mathbf{x}^1, \ldots, \mathbf{x}^k$ are linearly independent if no vector among them can be written as a *non-trivial* linear combination of the other vectors. A non-trivial linear combination of vectors is one in which at least one of the scalars in the linear combination is non-zero. Formally, this implies that for the linear independence of the vectors $\mathbf{x}^1, \ldots, \mathbf{x}^k$, the linear combination $\sum_{j=1}^{k} c_j \mathbf{x}^j$ can represent the zero vector only if all the coefficients $c_j$ are equal to zero. We express this using the notation

$$\sum_{j=1}^{k} c_j \mathbf{x}^j = \mathbf{0} \qquad \text{only if all } c_j = 0$$

**$\mathbb{R}^n$ as a Vector space**
$\mathbb{R}^n$ denotes the Cartesian product of $\mathbb{R}$ with itself $n$ times, i.e., $\mathbb{R} \times \mathbb{R} \times \ldots$ $n$ times. Geometrically $\mathbb{R}^n$ can be regarded as a vector space over $\mathbb{R}$ consisting of n-tuples with scalar components (real numbers).

Note that the linear independence is a property of the set comprising the vectors $\mathbf{x}^1, \ldots, \mathbf{x}^k$. An important consequence of linear independence is that each vector in the linear combination corresponds uniquely to the n-tuple of scalars $c_1, \ldots, c_k$.

**Linear Subspace**
A *linear subspace* of $\mathbb{R}^n$ is a subset of $\mathbb{R}^n$ that contains all the linear combinations of each of its points. So given a set of vectors $\mathbf{b}^1, \ldots, \mathbf{b}^d$, the set containing the linear combinations of all these vectors is a linear subspace. In fact, it is the smallest linear space $L$ that

contains the given set of vectors. By this we mean any linear space containing the given set of vectors must include $L$.

**Basis**

A *basis* for a vector space combines two key properties, namely, it *spans* (the linear space in that every vector in the space can be written as a linear combination of the basis elements) and it is *linearly independent* (no vector in a basis can be written as a linear combination of the other vectors in the basis). A key property of a basis of a vector space is that every vector of the vector space can be written as a unique linear combination of the basis vectors. In general a vector space can have several choices for a basis though some choice of a basis may be more preferable than another in a given problem. Another key property of a basis of a vector space is that the number of vectors in a basis is always the same for given vector space. This number is called the *dimension* of the vector space. A basic example here is the choice of the unit basis vectors $\mathbf{e}_i$ for $i = 1$ to $n$ as the standard basis for $\mathbb{R}^n$. Here $\mathbf{e}_i$ is the unit vector which has 1 in its $i$ th component and all other components are zero. There are n elements in the basis given here and so $\mathbb{R}^n$ is a vector space of dimension n. Notation: dim ($\mathbb{R}^n = n$)

If $W$ is a subspace of $V$, the dimension of $W$ is at most the dimension of $V$.

$W \subseteq V \implies dim(W) \leq dim(V)$. If $W$ is indeed a proper subset of $V$ (which means $V$ contains at least one element not in $W$) then the dimension of $W$ is strictly smaller than the dimension of $V$.

**Rank of a matrix**

Given a $m \times n$ matrix $A$ , the *column rank* of $A$ is the number of linearly independent columns of $A$; so the column rank of $A$ is upper bounded by the number of columns of $A$ i.e, $col\_rank(A) \leq n$.

Similarly the *row rank* of $A$ is the number of linearly independent rows of $A$ and so it is upper bounded by the number of rows of $A$ i.e $row\_rank \leq m$.

A basic result in matrix theory is that row rank and column rank of a matrix are equal. So we just say the *rank of a matrix* to mean either its row rank or column rank.

Another obvious consequence of the result stated above is that the rank of a matrix $A$ is the same as the rank of its transpose $A^T$: rank $A$ = rank($A^T$).

Example:

The rank of matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$ is 1 as there is only one linearly independent row or column.

**Solutions of linear equations**

The concepts introduced above are very useful to discuss the solutions of a system of linear equations( also called a linear system). A system of m linear equations in n variables $x_1, x_2, ...x_n$ can be written in the form $Ax = b$ where $A$ is an $m \times n$ matrix, $x$ is an $n \times 1$ column vector and $b$ is an $m \times 1$ column vector. Denoting the columns of $A$ by the column vectors $a_1, a_2, ..., +a_n$ in the left to right order, we can interpret the equation $Ax = b$ as a linear equation in the column vectors. i.e., $a_1x_1 + ...a_nx_n = b$ which gives the interpretation that this linear equation is solvable precisely when the vector $b$ is a linear combination of the column vectors of the matrix. Another way of stating this conclusion is that for the solvability of the linear system $Ax = b$, a necessary and sufficient condition (often abbreviated as nasc) is that the rank of the augmented matrix $(A\,b)$ is the same as the rank of the matrix A, i.e rank $(A\,b)$ = rank $(A)$.

Example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} x = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

has a solution because

$$\text{rank} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} = \text{rank} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

Both matrices have ranks equal to 1.

**The general notion of a Vector space**

Earlier we mentioned that $\mathbb{R}^n$ is a vector space. The general notion of a vector space is more general and we do not give its definition here. Two key examples of the general notion of a vector space are the following.

Example 1:

The set $\mathcal{P}_n$ is the set of all polynomials of degree less then or equal to n is a vector space (over the scalar field of real numbers).

Example 2:

The set $C[a, b]$ of continuous functions defined on the interval $[a, b]$ is a vector space (over the scalar field of real numbers).

However, there is an important difference between the vector spaces $\mathcal{P}_n$ and $C[a, b]$ . $\mathcal{P}_n$ is a finite dimensional vector space (FDVS) whereas $C[a, b]$ is an infinite dimensional vector space. Note that the dimension of $\mathcal{P}_n$ is $n + 1$.

In linear algebra we study finite dimensional vector spaces and linear mappings between them. If $V$ and $W$ are vector spaces, a function $T$ from $V$ to $W$ is called a *linear map* from $V$ and $W$ if the following *linearity* property holds.

$T(au + bv) = aT(u) + bT(v)$ where $a$ and $b$ are scalars and $u, v \in V$.

**Matrix as a representation of a linear operator**

We often use the notation $[n]$ to denote the set $\{1, \ldots, n\}$, the set of integers from $1, \ldots, n$,

Let $V$ and $W$ are vector spaces (over the same field) with $dim(V) = n$ and $dim(W) = m$.
Consider a linear operator $T$ from $V$ to $W$ written as $T : V \to W$.
To represent $T$ as a matrix we proceed as follows.

Choose a basis $\{v_1, \ldots, v_n\}$ for $V$ and $\{w_1, \ldots, w_n\}$ for $W$
$T$ being a linear operator, it is both necessary and sufficient to know the images
$T(v_i)$, $i \in [n]$ of the basis elements $v_i \in V$ $i \in [n]$ to determine the $T$-image of an arbitrary element $v \in V$.
If $v = \sum_{i=1}^{n} b_i v_i$, where $b_i \in \mathbb{R}$, $v_i \in V$. $i \in [n]$ then

$T v = T(\sum_{i=1}^{n} b_i v_i) = \sum_i T(b_i v_i) = \sum_i b_i T(v_i)$.
Note that the linearity of $T$ is the justification for the second and third equalities in the above.

Let $T(v_j) = \sum_{i=1}^{m} a_{ij} w_i$ for $j \in [n]$.
This definition is designed to represent the image of the the $j$th basis element $v_j$ of $V$ as

the $j$th column of the matrix $A$ that we construct by specifying its columns. This matrix $A$ represents the linear transformation $T$ with respect to the chosen basis for $V$ and $W$. Note that $A$ is an $m \times n$ matrix with real entries. The $j$th column of $A$ is constructed by taking the image of the $j$th basis element $v_j$ of $V$. So the matrix $A$ has as many columns as the dimension of the domain of $T$ and as many rows as the codomain (range) of $T$.

Let the column vector $b = [b_1, \ldots, b_n]^T$ represent the vector $v \in V$ with respect to the basis vectors $v_1, \ldots, v_n$ of $V$.
Now the matrix multiplication $Ab$ of $m \times n$ matrix $A$ with the $n \times 1$ vector $b$ yields a representation of the vector $Tv \in W$. Clearly it is a column vector with $m$ components as it should be!

So if we look at this construction carefully, we observe that linear algebra provides a geometric viewpoint to interpret the operations of matrix algebra. So matrix algebra is like crankshaft operating according to fixed rules which may seem quite arbitrary but the linear algebra view point helps to interpret it into a coherent geometric theory.

To illustrate this point of view further we give an interpretation of general matrix multiplication as yielding a representation for the composition of linear operators.

Let $U, V$ and $W$ be the given linear spaces over the field of real numbers. Consider the linear maps $T : U \to V$ and $S : V \to W$. Now their composition $S \circ T$ can be easily verified to be a linear map from $U \to W$. Let $A$ and $B$ denote the matrices of appropriate sizes that represent the linear maps $T$ and $S$ respectively with respect to a suitable choice of bases for the spaces $U, V$ and $W$. The matrix product $BA$ corresponds to the representation of the composite linear map $S \circ T$ w.r.t the given bases for $V$ and $W$. So to recall the point made earlier, the seemingly adhoc rules for matrix multiplication can be seen to correspond to the basic notion of function composition for linear operators.

**Two basic theorems of linear algebra**

We now state two basis theorems in linear algebra without proofs. We call these theorems the Nullity + Rank theorem and the Duality theorem.

**Nullity + Rank theorem**

Let $V$ and $W$ be two vector spaces. Consider a linear map $T$ from $V$ to $W$.
The *kernel* of $T$, denoted by ker(T) is the set of all elements of $V$ that are mapped to the zero element of $W$.
Notation: $ker(T) = \{x \in V \mid Tv = 0_W\}$.

ker(T) is a linear space in itself, it is a subspace of $V$. The dimension of the null space of $T$, i.e the dimension of the space ker(T), is called the *nullity* of $T$.

rank(T) denotes the dimension of the range space of $T$. We had earlier defined this notion in terms of the rank of a matrix that represents operator $T$. (So linear algebra viewpoint makes the geometric ideas underlying matrix algebra clear in a representation independent manner).

$range(T) = \{w \in W \mid Tv = w \text{ for some } v \in V\}$
$range(T)$ is also written as $ran(T)$.

The theorem states that dimension of the domain of T is equal to the dimension of the null space of T plus the dim of the range space of T. In other words, for a linear operator T,

$dim(domain(T)) = dim(ker(T)) + dim(ran(T))$.

In other words for a linear operator T, nullity + rank equals dimension of the domain.

**The fundamental duality in linear algebra**

Let $A$ be a linear map from vector space $V$ to vector space $W$.
We often blur the distinction between a linear map and its representation by a matrix and so identify the linear map $A$ with the matrix $A$. With each linear map is associated a dual operator $A^*$, called the *adjoint* of $A$, which is identified with the transpose of the matrix $A$. The dual operator $A^*$ gives a linear map from $W$ to $V$, synonymously we identify $A^T$ as defining the linear map from $W$ to $V$. (The conceptual distinction between a linear map and its matrix representation is important. Though we can often simplify the notation by blurring this distinction, sometimes we have to pay a price in terms of understanding by not making this distinction).

Given a vector space $V$ and a subspace $W$ of $V$, the *ortho-complement* of $W$ denoted by $W^{\perp}$, is the set of all the vectors in $V$ which are orthogonal to each vector in $W$.

Notation: $W^{\perp} = \{v \in V \mid (v\, w) = 0, \forall w \in W\}$

For a general vector space $V$, the ortho-complement $W^{\perp}$ turns out to be the annihilator of W, denoted by $W^{\circ}$, which consists of all the linear functionals on V that annihilate each vector in $W$.

Notation:$W^{\circ} = \{f \in V^* \mid f(w) = 0\ \forall w \in W\}$
(Though the concepts $W^{\perp}$ and $W^{\circ}$ are equivalent, the notion of ortho-complement is easier to grasp and is more concrete for a first encounter with this idea).

**Duality theorem of Linear algebra**
This theorem says that the kernel of a linear map $A : V \rightarrow W$ is the orthogonal complement of the range of the transpose $A^T$. Similarly the range of $A$ is the orthocomplement of the kernel of the transpose $A^T$.

$ker(A) = (ran(A^T)^{\perp})$ and similarly $(ran(A)) = ker((A^T)^{\perp})$.

We interpret this theorem by saying that given an arbitrary vector $v \in V$ we can decompose it into two orthogonal component vectors , $u \in ker(A)$ and $w \in ran(A^T)$.
Notation: $V = ker(A) \perp ran(A^T)$. As the situation is completely symmetrical, we can write a similar decomposition of W as follows. $W = ran(A) \perp ker(A^T)$.

**Dual space**

The notion of duality in linear algebra, as given by the theorem above, springs from the concept of *dual space* which we describe here by means of a simple example using the space $\mathbb{R}^n$.

Consider $\mathbb{R}^n$ with the standard inner product $(x, y)$ for $x, y \in \mathbb{R}^n$. If we fix the first argument in the inner product, we obtain the corresponding function denoted by $f_x$ which maps *every* vector $y \in \mathbb{R}^n$ to a scalar (in $\mathbb{R}$). The linearity of the function $f_x$ follows from the property of inner product. The term *linear functional* is used to denote such a linear map where the range of the map is the set of scalars, as $\mathbb{R}$ in this case. (Note that the scalars $\mathbb{R}$ can be regarded as a linear space over itself and so the notion of a linear functional can be regarded as a special case of a linear map when the range of the linear map is the set of scalars). Clearly the sum $f_x + f_y$ of any two such linear functionals $f_x$ and $f_y$ is a linear functional and it is easy to see that the set of all linear functionals is a linear space; this space is called the *dual space* of the original space $\mathbb{R}^n$ and is denoted by $(\mathbb{R}^n)^*$. As each element $f_x \in (\mathbb{R}^n)^*$ can be identified with $x \in \mathbb{R}^n$, the space $(\mathbb{R}^n)^*$ can be identified with $\mathbb{R}^n$ i.e, $(\mathbb{R}^n)^* \cong \mathbb{R}^n$ and so $\mathbb{R}^n$ is called a *self-dual space*. Note that the dual space is of the same dimension as the original space.

The notion of a dual space of a linear space described above in the setting of $\mathbb{R}^n$ can be easily carried forward to an arbitrary linear space as follows.

If $e_1, e_2, ...e_n$ is a basis of a vector space $V$, then $f_1, f_2, ...f_n$ is defined as a basis of the dual space $V^*$ of $V$ as follows: $f_i(e_j) = \delta_{ij}$, where $\delta_{ij}$ is the Kronecker (or Dirac) delta function which takes the value 1 when $i = j$ and the value 0 otherwise. The nice thing about the dual space is that the dual space of the dual space is the original space, i.e. $((V^*)^*) = V$. The equality here is strictly speaking, an identification of the spaces and this identification is called canonical (or intrinsic) as it is independent of the bases used. In contrast , the identification of $V$ and $V^*$ is tied up with the choice of the bases of the two spaces and so is not canonical. A major goal of linear algebra is to provide a coordinate-free (i.e, basis-free) description of the concepts involved.

**And finally, why is it called Linear Algebra?**

We end this brief review of linear algebra with a justification for the name of the subject. For a given vector space $V$ , the set of all linear maps $\{T \mid T : V \to V\}$ from $V$ into itself constitutes a linear space. We denote this set by $L(V)$. Besides the linear maps can be composed with one another and the result is again a linear map. So the set of linear maps constitute an algebraic structure called *ring* which is the same structure as that of the set $\{\mathbb{Z}, +, \times\}$ of integers with respect to addition and multiplication. When the operations of the scalar multiplication in vector space and the multiplication in a ring are compatible with each other as in this case, the resulting structure is called an *algebra*. So the set $L(V)$ of all linear maps defined on a vector space $V$ is an algebra and this gives the name of the subject, *linear algebra*.

Note: As the name linear algebra would suggest, we place emphasis on the *linear maps* defined on a vector space $V$ rather than on the space $V$ itself. Besides the simple fact that the vector space itself can be identified with the identity map, there is a deeper justification for this as follows. The choice of an appropriate linear operator (self-adjoint for example) gives a nice way to decompose the underlying vector space and so the operators defined on a vector space are crucial to study the structure of a linear space. Linear algebra, which is the study of operators on finite-dimensional vector spaces, is the gateway to the study of operators defined on infinite-dimensional vector spaces , a branch of mathematics called functional analysis. Hilbert spaces and operators on Hilbert spaces are preeminent examples of such study and is rich in theory and applications.