MATRIX ALGEBRA REVIEW FOR STATISTICS

LESSON 1 NOTES

Rev. 3.3

by

Robert A. LaBudde, Ph.D.

I. INTRODUCTION

This brief text is meant as a review of vector and matrix algebra for students of statistics who plan to go on to study the methods of multivariate statistics, where the knowledge of notation and basic concepts of matrix algebra is essential to understanding. This text is not meant as a thorough and complete coverage of linear algebra as a field of study in itself.

Vectors and matrices are one- and two-dimensional ordered arrays of numbers with an associated collection of mathematical operations (an ‘algebra’) that they are ‘closed’ under (i.e., the result of the operation is once again a vector or matrix). This contrasts with other data constructs such as ‘lists’ (i.e., ordered collection of things that do not have to be numbers, and may be collections themselves), ‘data frames’ (i.e., two-dimensional arrays with each column having elements all of the same kind (e.g., strings or numbers)), ‘tables’ (multidimensional ordered arrays, with focus on structuring rather than numerical operations) and 3 or higher dimensional arrays.

It is the numerical algebra of operations that makes vectors and matrices useful in statistics.

The notation for vectors and matrices is quite difficult to implement using a word processor not designed for the purpose, so examples will be given mostly using the programming language R, which handles vectors and matrices as native objects. R is the basis of a freeware open source large-scale statistical system which is becoming the de facto standard for statistical computing. Knowing R is not necessary to do the assignments or understand this material, but it is useful here to illustrate numerical examples. Other statistical programming systems, such as JMP in SAS, also allow a wide variety of matrix operations, but lack the useful symbolic representations available with R.

Some engineering programming languages, such as MATLAB® and GAUSS®, are designed specifically to handle matrix operations in a symbolic manner. Also, Texas Instruments calculators, such as the TI-8x and TI-9x series, can manipulate vectors and matrices and perform basic operations. Spreadsheet software, such as Microsoft Excel, can do matrix multiplication and inversion.

II. OBJECTS:

Scalar: A single numeric value (number). Denoted here by a non-bold symbol. E.g., s = 1.4142.

Vector: A 1-dimensional n-tuple (ordered collection) of n scalar elements. Denoted here by a lower-case bold symbol. E.g., **v** = [2, 3, 5]. The i-th element of a vector is denoted by the notation v_i or **v**[i]. For the previous example, $v_1 = 2$, $v_2 = 3$ and $v_3 = 5$. Note that a 1-tuple vector **u** = [3] is *not* actually the same as the scalar $a = 3$. They are two different classes of objects. However, sometimes it is convenient to ignore this distinction.

Row vector: A vector orientated horizontally. E.g.,

$$\mathbf{u} = [2 \;\; 3 \;\; 5] \qquad \text{(II.1)}$$

Column vector: A vector orientated vertically, E.g.,

$$\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \qquad \text{(II.2)}$$

Length or order of a vector: The number of elements in a vector. E.g., the order of **u** in eq.(II.1) and **v** in eq.(II.2) are both 3. Note that the term 'length' is frequently used instead as synonymous to 'norm' (see Chapter III below), so we will prefer 'order' here as the number of elements in a vector to avoid confusion. Note, however, that R uses 'length' for this purpose.

Matrix: a 2-dimensional (n x m) set of n row vectors of order m, or a set of m column vectors of order n. Denoted here by an upper-case bold symbol. Displayed using in-line notation as **A** = [1, 2; 3, 4; 5, 6] as a set of rows, or, more commonly, as a two-dimension format

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \qquad \text{(II.3)}$$

The elements of a matrix **A** are denoted algebraically by $A_{i\,j}$ where i is the row number and j is the column number.

A row vector may be considered equivalent to a matrix of one row with the same elements, and a column vector considered equivalent to a matrix of one column with the same elements.

Order of a matrix: A matrix organized as n rows and m columns is said to be of order n x m.

Square matrix: A matrix is called 'square' if its row order equals its column order, i.e., it is an n x n matrix.

Identity matrix: The identity matrix **I** of order n x n is a square matrix whose elements $I_{ij} = 0$ if $i \neq j$ and 1 if $i = j$. For example, the 3 x 3 identity matrix is

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{(II.4)}$$

Zero matrix: The **0** matrix of order n x n is a square matrix whose elements are all 0. For example, the 3 x 3 zero matrix is

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{(II.5)}$$

Array: a d-dimensional ($n_1 \times n_2 \times \ldots n_d$) collection of elements. Denoted here by an upper-case bold italic symbol. The elements of a 3 x 4 x 2 array ***B*** are denoted algebraically by B_{ijk}, where i = 1 to 3, j = 1 to 4, and k = 1 to 2.

Vectors and matrices are examples of one and two dimensional arrays.

III. VECTOR OPERATIONS:

Row vector addition: Row vectors of the same order n are closed under addition and subtraction. If $\mathbf{u} = [\ u_1, u_2, \ldots, u_n\ ]$ and $\mathbf{v} = [v_1, v_2, \ldots, v_n\ ]$ are two row vectors of order n, then $\mathbf{u} \pm \mathbf{v}$ is a row vector of order n also, and its elements are $u_i \pm v_i$ for i = 1, 2, … , n.

Column vector addition: Column vectors of the same order n are also closed under addition and subtraction, with the resultant column vector elements equal to the sum or difference of the addend elements.

Transpose: The transpose of a row vector is the column vector of the same order and same elements, and similarly the transpose of a column vector is a row vector with the same elements. The transpose is denoted by a superscript "t" or by an apostrophe ‘. E.g., the row vector **u** in eq.(II.1) is the transpose of the column vector **v** in eq.(II.2), or $\mathbf{u} = \mathbf{v}^t$. Similarly, the column vector **v** in eq.(II.2) is the transpose of the row vector **u** in eq.(II.1), or $\mathbf{v} = \mathbf{u}^t$. The transpose of the transpose of a vector is the original vector itself: $\mathbf{u} = (\mathbf{u}^t)^t$.

Row vector scalar multiplication: A scalar c times a row vector **v** of order n is a row vector of order n (denoted ‘c **v**’) with elements $c\ v_i$. Scalar multiplication is distributive across row vector addition and subtraction. I.e.,

$$c\,(\mathbf{u} \pm \mathbf{v}) \quad = \quad (c\,\mathbf{u}) \pm (c\,\mathbf{v}) \qquad \text{(III.1)}$$

Column vector scalar multiplication: A scalar c times a column vector **v** of order n is a column vector of order n (denoted ‘c **v**’) with elements $c\ v_i$. Scalar multiplication is distributive across column vector addition and subtraction. I.e.,

$$c\,(\mathbf{u} \pm \mathbf{v}) \quad = \quad (c\,\mathbf{u}) \pm (c\,\mathbf{v}) \qquad \text{(III.2)}$$

Inner product: The inner product (sometimes called the "dot" product) of a row vector **u** and a column vector **v** of the same order n is denoted ‘**u v**’ or ‘**u** . **v**’ or ‘**u** * **v**’, and is a scalar with the value of the sum of the product of the elements of the two vectors. I.e.,

$$\mathbf{u}\,\mathbf{v} \quad = \quad \sum_{i=1}^{n} u_i\, v_i \qquad \text{(III.3)}$$

For example, if $\mathbf{u} = [1\ 2\ 3]$ and $\mathbf{v} = [4\ 5\ 6]^t$, then

$$\mathbf{u}\,\mathbf{v} \quad = \quad (1)(4) + (2)(5) + (3)(6) \quad = 32 \qquad \text{(III.4)}$$

Using R to reproduce this example (noting that R does not normally distinguish explicitly the difference between row and column vectors):

```
> u = c(1,2,3)  #vector u = [1 2 3]
```

```
> u
[1] 1 2 3
> v = c(4,5,6)  #vector v = [4 5 6]
> v
[1] 4 5 6
> u %*% v   #Inner product of u and v
     [,1]
[1,]   32
```

Because of the way in which the inner product is defined as a scalar, the inner product can actually be computed between any of 2 row vectors, 2 column vectors, a row and a column vector and a column and a row vector, so long as both vectors have the same order (number of elements).

Norm or magnitude of a vector: The 'norm' or 'magnitude' (and sometimes 'length') of a vector **v**, denoted | **v** | or || **v** || is the scalar square root of the inner product of the vector with its transpose. For a row vector **u**,

$$|\mathbf{u}| = \sqrt{(\mathbf{u}\,\mathbf{u}^t)} \qquad \text{(III.5)}$$

Similarly, for a column vector **v**,

$$|\mathbf{v}| = \sqrt{(\mathbf{v}^t\,\mathbf{v})} \qquad \text{(III.6)}$$

Angle between two vectors: Note that the inner product of two vectors geometrically has the form

$$\mathbf{u}\,\mathbf{v} = |\mathbf{u}|\ |\mathbf{v}|\cos(\theta) \qquad -\pi/2 \le \theta \le \pi/2 \qquad \text{(III.7)}$$

where

$$\cos(\theta) = \Sigma\, u_i\, v_i\ /\ (\,|\mathbf{u}|\ |\mathbf{v}|\,) \qquad \text{(III.8)}$$

is the angle between the two vectors (geometrically), or more generally a measure similar to a correlation coefficient as the cosine lies between –1 and +1. (The proof of eq.(III.8) is by equating the length squared of the side **u** – **v** from the trigonometric law of cosines to the formula for $|\mathbf{u} - \mathbf{v}|^2 = (\mathbf{u} - \mathbf{v}) . (\mathbf{u} - \mathbf{v})$ and solving for cos(θ).)

Outer product: The 'outer' or 'tensor' product of a *column* vector **u** of order n and a *row* vector **v** of order m is denoted **u** x **v** here (but most books just use the notation **u v**, with the inner product as **u** . **v,** or inferred by which of **u** and **v** is the row and which is the column vector), and is an n x m matrix whose i j-th element is $u_i\, v_j$. Note that **u** and **v** do *not* have to be the same order. E.g., for the column vector **u** = $[\ 1\ 2\ 3\ ]^t$ and the row vector **v** = [4 5] the outer product is

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} [\,4 \;\; 5\,] = \begin{bmatrix} 4 & 5 \\ 8 & 10 \\ 12 & 15 \end{bmatrix} \qquad \text{(III.9)}$$

Note that the outer product **u** x **v** of two vectors is a *matrix*.
Using R to illustrate the example:

```
> u = c(1,2,3)  #vector u = [1 2 3]
> u
[1] 1 2 3

> v = c(4,5)    #vector v = [4 5]
> v
[1] 4 5

> u %o% v   #outer product of u and v
     [,1] [,2]
[1,]    4    5
[2,]    8   10
[3,]   12   15
```

(When using R, be careful, as %o% is also defined for matrices and arrays, where the result has a number of subscripts equal to the sum of that of the arguments. So if you inadvertently convert **u** or **v** to a matrix, the result will have higher dimensionality than you expect!)

NOTE: It is extremely cumbersome in text to keep the subtle distinction between row and column vectors. Frequently, they are just referred to as "vectors", and the distinction between a row vector and a column vector indicated by the context used. Typically column vectors are used more, so a "vector" should be considered a column vector until the context implies otherwise. Judicious use of the transpose operator keeps the notation exact in cases of ambiguity. (When using R, however, you will have to follow its rules for using operators and give it the right class objects for arguments. This can be tricky some times.)

From now on, we will refer to both row and column vectors as just "vectors" unless the context requires clarification.

EXERCISES (III):

Suppose

$$\mathbf{u} = [\,2 \;\; \text{-}1\,] \qquad \mathbf{v} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 1 & 0 & -2 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 2 \\ -2 \\ -2 \end{bmatrix}$$

1. Compute **u v**. (Hint: this is **u** . **v**.)

2. Compute **v u**. (Hint: this is **v** x **u**.)

3. Compute | **u** | and | **v** |. What is the angle θ between **u** and **v**?

4. Compute **w x**. (Hint: this is **w** . **x**.)

5. Compute **x w**. (Hint: this is **x** x **w**.)

6. Compute | **w** | and | **x** |. What is the angle θ between **w** and **x**?

IV. MATRIX OPERATIONS:

Matrix addition: Matrices of the same order n x m are closed under addition and subtraction. I.e., if $\mathbf{A} = [\ \{A_{i\,j}: i = 1,2,...,n; j=1,2,...,m\ \}\ ]$ and $\mathbf{B} = [\{B_{i\,j}: i = 1,2,...,n; j=1,2,...,m\ \}\ ]$ are two matrices of the same order n x m, then $\mathbf{A} \pm \mathbf{B}$ is also a matrix of order n x m, and its elements are $A_{i\,j} \pm B_{i\,j}$.

Trace: The trace of an n x n square matrix **A** is the sum of its diagonal (i = j) elements.

$$\mathrm{tr}(\mathbf{A}) = A_{11} + A_{22} + ... + A_{nn} \qquad \text{(IV.1)}$$

For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}, \text{ then } \mathrm{tr}(\mathbf{A}) = 1 + 2 + 3 = 6 \qquad \text{(IV.2)}$$

Transpose: The transpose of a matrix **A** of order n x m is a matrix of order m x n denoted $\mathbf{A}^t$, and its elements are

$$A^t_{j\,i} = A_{i\,j} \quad \text{for } i = 1,2,...,n; j=1,2,...,m \qquad \text{(IV.3)}$$

For example, using R. for the matrix **A** = [1, 2, 3; 4, 5, 6]:

```
> A = matrix(c(1,2,3, 4,5,6), ncol=3, byrow=TRUE) #matrix by rows
> A
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6

> t(A)  #the transpose of A
     [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
```

Matrix scalar multiplication: A scalar c multiplied times a matrix A is a matrix of the same order with elements $c\,A_{i\,j}$. Scalar multiplication is distributive across matrix addition and subtraction. I.e.,

$$c\,(\mathbf{A} \pm \mathbf{B}) = (c\,\mathbf{A}) \pm (c\,\mathbf{B}) \qquad \text{(IV.4)}$$

E.g., using R:

```
> A = matrix(c(1,2,3, 4,5,6), ncol=3, byrow=TRUE) #matrix by rows
> A
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
```

```
> B = matrix(c(1,1,1, 2,2,2), ncol=3, byrow=TRUE) #matrix of same order
> B
     [,1] [,2] [,3]
[1,]    1    1    1
[2,]    2    2    2

> A + B
     [,1] [,2] [,3]
[1,]    2    3    4
[2,]    6    7    8

> 2*(A + B)   #scalar multiplication
     [,1] [,2] [,3]
[1,]    4    6    8
[2,]   12   14   16

> (2*A) + (2*B)
     [,1] [,2] [,3]
[1,]    4    6    8
[2,]   12   14   16
```

Matrix multiplication: The product **AB** of a matrix **A** of order n x m and a matrix **B** of order p x q is defined only if m = p, i.e., the number of columns of **A** equals the number of rows of **B**. The row order **AB** is a matrix of order n x q (i.e., the number of rows equal that of **A** and the number of columns equal that of **B**). The elements of **AB** are given by the inner product of the rows of **A** and the columns of **B**:

$$(\mathbf{AB})_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj} \qquad \text{(IV.5)}$$

E.g., using R:

```
> A = matrix(c(1,2,3, 4,5,6), ncol=3, byrow=TRUE) #2 x 3 matrix
> A
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6

> B = matrix(c(7,8,9, 10,11,12, 13,14,15), ncol=3, byrow = TRUE)
> B
     [,1] [,2] [,3]
[1,]    7    8    9
[2,]   10   11   12
[3,]   13   14   15
> A %*% B   #matrix multiplication
     [,1] [,2] [,3]
[1,]   66   72   78
[2,]  156  171  186
```

Note that matrix multiplication is not necessarily commutative (symmetric in its arguments), so **AB** ≠ **BA** in general. If **AB** exists, **BA** may not, as the orders may not be conformable. Only for square matrices (i.e., order n x n) are both **AB** and **BA** necessarily defined.

Matrix multiplication is associative, i.e., **A**(**BC**) = (**AB**)**C** = **ABC**, and distributive across matrix addition, i.e., **A**(**B** + **C**) = (**AB**) + (**AC**) = **AB** + **AC**.

Matrix-vector multiplication: If a row vector **u** is treated as a matrix of a single row, or a column vector **v** is treated as a matrix of a single column, vector-matrix or matrix-vector multiplication by a matrix **A** is just a case of matrix multiplication, so long as the orders are conformable. Note that the result of **u A** will have a single row and **A v** a single column.

E.g., using R:

```
> u = c(1,2,3)  #row vector
> u
[1] 1 2 3

> B = matrix(c(7,8,9, 10,11,12, 13,14,15), ncol=3, byrow = TRUE)
> B
     [,1] [,2] [,3]
[1,]    7    8    9
[2,]   10   11   12
[3,]   13   14   15

> u %*% B   #vector-matrix multiplication
     [,1] [,2] [,3]
[1,]   66   72   78

> v = c(1,2,3)  #column vector
> v
[1] 1 2 3

> A = matrix(c(1,2,3, 4,5,6), ncol=3, byrow=TRUE) #2 x 3 matrix
> A
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6

> A %*% v   #matrix-vector multiplication
     [,1]
[1,]   14
[2,]   32
```

Multiplication by an identity matrix: If **A** is an n x n square matrix and **I** is the n x n identity matrix, then **A I** = **I A** = **A**.

Vector inner product, outer product and matrix multiplication are all based upon the same algorithm.

The vector inner product occurs when a row vector is on the left and a column vector on the right. The result is a scalar. Except for object type, you would get the same answer treating the row vector as a 1 x n matrix and the column vector as a n x 1 matrix. Then the two are conformable on the 2nd and 1st dimensions respectively (which are both "n"), and the result is a 1 x 1 matrix whose element is the same as the vector inner product of the row and the column (which, however, is a scalar).

The vector outer product occurs when a column vector is on the left and a row vector on the right. Now we treat the column vector as a n x 1 matrix and the row vector as a 1 x n matrix. The two are conformable on the "middle" dimension of "1", so we can multiply the two matrices. The result is the n x n "outer product" of the two vectors.

Matrix multiplication thus includes the vector inner product and the vector outer product as subcases, so long as we don't pay attention to the object type of the results (the inner product gives a scalar, but the matrix version gives a 1 x 1 matrix). So all three are algorithmically and conceptually the same.

EXAMPLE: VECTOR INNER AND OUTER AND MATRIX PRODUCTS

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 7 & 4 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 6 & 3 & 1 \\ -1 & 2 & 5 \end{bmatrix} \quad \mathbf{c} = [\ 1\ \ 5\ ] \quad \mathbf{d} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Then

$$\mathbf{c}\ \mathbf{d} = (1)(3) + (5)(1) = 8$$

$$\mathbf{d} \times \mathbf{c} = \mathbf{d}\,\mathbf{c} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} [\ 1\ \ 5\ ] = \begin{bmatrix} 3 & 15 \\ 1 & 5 \end{bmatrix}$$

$$\mathbf{A}\,\mathbf{d} = \begin{bmatrix} 1 & 2 \\ 7 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} (1)(3)+(2)(1) \\ (7)(3)+(4)(1) \end{bmatrix} = \begin{bmatrix} 5 \\ 25 \end{bmatrix}$$

$$\mathbf{c}\,\mathbf{B} = [\ 1\ \ 5\ ] \begin{bmatrix} 6 & 3 & 1 \\ -1 & 2 & 5 \end{bmatrix} = [\ (1)(6)+(5)(-1) \quad (1)(3)+(5)(2) \quad (1)(1)+(5)(5)\ ]$$

$$= [1 \quad 13 \quad 26\ ]$$

EXAMPLE: ADDITION, SUBTRACTION & SCALAR MULTIPLICATION OF MATRICES

If

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 2 & 0 \\ 3 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Then

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1+0 & 0+2 & 2+0 \\ 0+3 & 1+0 & 3+1 \\ 1+1 & 0+1 & 1+0 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ 3 & 1 & 4 \\ 2 & 1 & 1 \end{bmatrix}$$

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} 1-0 & 0-2 & 2-0 \\ 0-3 & 1-0 & 3-1 \\ 1-1 & 0-1 & 1-0 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 2 \\ -3 & 1 & 2 \\ 0 & -1 & 1 \end{bmatrix}$$

$$2\,\mathbf{A} + 3\,\mathbf{B} = \begin{bmatrix} 2(1)+3(0) & 2(0)+3(2) & 2(2)+3(0) \\ 2(0)+3(3) & 2(1)+3(0) & 2(3)+3(1) \\ 2(1)+3(1) & 2(0)+3(1) & 2(1)+3(0) \end{bmatrix} = \begin{bmatrix} 2 & 6 & 4 \\ 9 & 2 & 9 \\ 5 & 3 & 2 \end{bmatrix}$$

$$5\,\mathbf{A} - 2\,\mathbf{B} = \begin{bmatrix} 5(1)-2(0) & 5(0)-2(2) & 5(2)-2(0) \\ 5(0)-2(3) & 5(1)-2(0) & 5(3)-2(1) \\ 5(1)-2(1) & 5(0)-2(1) & 5(1)-2(0) \end{bmatrix} = \begin{bmatrix} 5 & -4 & 10 \\ -6 & 5 & 13 \\ 3 & -2 & 5 \end{bmatrix}$$

Using R:

```
> A<- matrix(c(1,0,2, 0,1,3, 1,0,1), ncol=3, byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    1    0    2
[2,]    0    1    3
[3,]    1    0    1

> B<- matrix(c(0,2,0, 3,0,1, 1,1,0), ncol=3, byrow=TRUE)
> B
     [,1] [,2] [,3]
[1,]    0    2    0
[2,]    3    0    1
[3,]    1    1    0

> A+B
     [,1] [,2] [,3]
[1,]    1    2    2
```

```
[2,]    3    1    4
[3,]    2    1    1

> A-B
     [,1] [,2] [,3]
[1,]    1   -2    2
[2,]   -3    1    2
[3,]    0   -1    1

> 2*A+3*B
     [,1] [,2] [,3]
[1,]    2    6    4
[2,]    9    2    9
[3,]    5    3    2

> 5*A-2*B
     [,1] [,2] [,3]
[1,]    5   -4   10
[2,]   -6    5   13
[3,]    3   -2    5
```

EXERCISES (IV):

Suppose

$\mathbf{u}$ = [2 -1] $\mathbf{v} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ $\mathbf{w} = \begin{bmatrix} 1 & 0 & -2 \end{bmatrix}$ $\mathbf{x} = \begin{bmatrix} 2 \\ -2 \\ -2 \end{bmatrix}$

$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ -3 & 3 \end{bmatrix}$ $\mathbf{B} = \begin{bmatrix} 1 & -2 & 1 \\ 0 & -1 & 2 \\ 2 & -3 & -2 \end{bmatrix}$

1. Compute $\mathbf{u}\ \mathbf{A}$.

2. Compute $\mathbf{A}\ \mathbf{v}$.

3. Compute $\mathbf{A}^t$.

4. Compute $\mathbf{u}\ \mathbf{A}\ \mathbf{v}$.

5. Compute $\mathbf{u}\ \mathbf{A}^t\ \mathbf{v}$.

6. Compute $\mathbf{u}\ \mathbf{A}\ \mathbf{u}^t$.

7. Compute $\mathbf{v}^t\ \mathbf{A}\ \mathbf{v}$.

8. Compute $\mathbf{w}\ \mathbf{B}$.

9. Compute $\mathbf{B}\ \mathbf{x}$.

10. Compute $\mathbf{B}^t$.

11. Compute $\mathbf{w}\ \mathbf{B}\ \mathbf{x}$.

12. Compute $\mathbf{w}\ \mathbf{B}^t\ \mathbf{x}$.

13. Compute $\mathbf{w}\ \mathbf{B}\ \mathbf{w}^t$.

14. Compute $\mathbf{x}^t\ \mathbf{B}\ \mathbf{x}$.

V. REVIEW OF THE SOLUTION OF A SYSTEM OF LINEAR EQUATIONS

In intermediate or college algebra, you should have learned some basics of the solution of a system of linear equations. This is one of the principal applications of matrix algebra, so we will review this subject first before showing how matrices and vectors can be used to facilitate the solution process and compactly represent the equations.

To 'solve' an equation means to isolate the unknown variables on the left-hand side ('LHS') of an equation with an expression on the right-hand side ('RHS') which does not involve the variable in question. The variable has been 'fully solved for' if the RHS is a constant.

Consider first the simple case of one equation in one unknown variable 'x':

$$6x - 2 = 0 \qquad \text{(V.1)}$$

First we shift the constant 2 to the RHS of the equation, leaving the variable term on the LHS:

$$6x = 2 \qquad \text{(V.2)}$$

(This step is fairly trivial, so we will start future problems with this step already accomplished.) Now we use the following fact:

THEOREM V.1: Dividing both sides of an equation by the same non-zero constant results in a new valid equation.

Dividing both sides of eq.(V.2) by the number 6 gives

$$x = 1/3 \qquad \text{(V.3)}$$

an equation which is fully solved for 'x'. We say "x = 1/3 is the solution of eq.(V.1)."

The solution process can be applied to the more general

$$a x = b \qquad \text{(V.4)}$$

Here 'a' and 'b' are 'parameters', or variables that represent unknown at this time constant values that will or can be specified later. 'x' is a variable whose value is to be determined via the equation. Solving as before,

$$x = b / a \qquad \text{for } a \neq 0 \qquad \text{(V.5)}$$

is the symbolic solution of eq.(V.4).

Now consider the case of 2 equations in 2 unknowns, 'x' and 'y':

$$3\ x + 2\ y = 4 \tag{V.6a}$$

$$1\ x - 4\ y = -8 \tag{V.6b}$$

First we multiply eq.(V.6b) by the number 3 on both sides to make the coefficient of x the same as it is in the first equation (V.6a):

$$3\ x - 12\ y = -24 \tag{V.6c}$$

The 3 equations (V.6) are not 'independent', because we made eq.(V.6c) out of eq.(V.6b). So we are really are allowed only to keep 2 of them, the same number we started with. We choose (V.6a) and (V.6c). Eq.(V.6b) is still valid, but contains no information not already in the other two equations (it's redundant or 'dependent'). Now we use the fact that

THEOREM V.2: Adding or subtracting one equation from another (that is, the LHS from the LHS and the RHS from the RHS) results in a new valid equation.

We subtract eq.(V.6c) from eq.(V.6a) to get the new equation

$$14\ y = 28 \tag{V.7}$$

Note that we have cleverly removed the variable 'x' from the equation. We can now solve this 1-variable equation for 'y':

$$y = 2 \tag{V.8}$$

We can now 'substitute' the value '2' for 'y' in any of the eqs.(V.6) to find 'x'. The easiest is eq.(V.6b), which gives

$$x - 8 = -8 \qquad \text{or} \qquad x = 0 \tag{V.9}$$

We say the solution of eqs.(V.6) is x = 0 and y = 2.

What if we only had one equation in the two variables 'x' and 'y', such as

$$6\ x - 2\ y = 4 \tag{V.10}$$

Here we have 2 variables and 1 equation to fix them. This leaves 2 – 1 = 1, or 1 'degree of freedom' left unspecified. The best therefore that we can do is to solve for one variable in terms of the other:

$$x = 2/3 + y/3 \tag{V.11a}$$

or

$$y = 3x - 2 \qquad \text{(V.11b)}$$

which ever is of more interest. If there is no reason to prefer one of eq.(V.11a) or eq.(V.11b) over the other, sometimes is convenient to add another variable 't', a 'parameter', along with another equation. This still gives 3 – 2 =1 degree of freedom as before, so we haven't changed the problem. So,

$$x = 2/3 + t/3 \qquad \text{(V.12a)}$$

$$y = t \qquad \text{(V.12b)}$$

(where $-\infty < t < \infty$) is a convenient representation of the solution of the under-determined system of equations (V.10). It should be obvious that the parameterization could also have been accomplished by

$$x = t \qquad \text{(V.13a)}$$

$$y = 3t - 2 \qquad \text{(V.13b)}$$

Now consider the case of 3 equations in the 3 unknowns 'x', 'y' and 'z'. For example,

$$3x + 2y + z = 4 \qquad \text{(V.14a)}$$

$$1x - 2y + 2z = 1 \qquad \text{(V.14b)}$$

$$6x + y + z = -1 \qquad \text{(V.14c)}$$

For larger systems of equations, or those without simple integers as coefficients, it usually is simpler to be systematic in the solution process. The first step is to reorder the equations so that the first one has the largest (in magnitude) coefficient of 'x'. This is based on the fact that

THEOREM V.3: Interchanging the order of any two equations in the set does not change the solution of the equations.

So,

$$6x + y + z = -1 \qquad \text{(V.15a)}$$

$$3x + 2y + z = 4 \qquad \text{(V.15b)}$$

$$1x - 2y + 2z = 1 \qquad \text{(V.15c)}$$

The next step is to divide the first equation by the coefficient of 'x', here '6':

$$x + y/6 + z/6 = -1/6 \qquad \text{(V.16a)}$$

$$3\,x + 2\,y + z = 4 \qquad \text{(V.16b)}$$

$$1\,x - 2\,y + 2\,z = 1 \qquad \text{(V.16c)}$$

Now we use the consequence of Theorems (V.1) and (V.2),

COROLLARY V.4: Adding or subtracting a multiple of one equation from another results in a new valid equation.

So we subtract 3 times eq.(V.16a) from eq.(V.16b) to get the new system of equations

$$1\,x + y/6 + z/6 = -1/6 \qquad \text{(V.16a)}$$

$$3\,y/2 + z/2 = 9/2 \qquad \text{(V.16b)}$$

$$1\,x - 2\,y + 2\,z = 1 \qquad \text{(V.16c)}$$

We also now subtract 1 times eq.(V.16a) from eq.(V.16c) to get:

$$1\,x + y/6 + z/6 = -1/6 \qquad \text{(V.17a)}$$

$$3\,y/2 + z/2 = 9/2 \qquad \text{(V.17b)}$$

$$-13\,y/6 + 11\,z/6 = 7/6 \qquad \text{(V.17c)}$$

Note that we have successfully eliminated the variable ‘x’ from all of the equations except the first, eq.(V.17a). Now we repeat the process with the next variable ‘y’. Of the remaining two equations, eq.(V.17c) has the larger (in magnitude) coefficient of ‘y’, so we swap it with eq.(V.17b):

$$1\,x + y/6 + z/6 = -1/6 \qquad \text{(V.18a)}$$

$$-13\,y/6 + 11\,z/6 = 7/6 \qquad \text{(V.18b)}$$

$$3\,y/2 + z/2 = 9/2 \qquad \text{(V.18c)}$$

We divide eq.(V.18b) by the coefficient of y, or -13/6, to obtain

$$1\,x + y/6 + z/6 = -1/6 \qquad \text{(V.19a)}$$

$$1\,y - 11\,z/13 = -7/13 \qquad \text{(V.19b)}$$

$$3\,y/2 + z/2 = 9/2 \qquad \text{(V.19c)}$$

Finally, we subtract 3/2 times eq.(V.19b) from eq.(V.19c) to eliminate the term in y:

$$1\,x + y/6 + z/6 = -1/6 \tag{V.20a}$$

$$1\,y - 11\,z/13 = -7/13 \tag{V.20b}$$

$$23\,z / 13 = 69/13 \tag{V.20c}$$

We have eliminated 'y' from the 3rd equation. Now we have only the variable 'z' remaining in the last equation. Dividing eq.(V.20c) by 23/13, the coefficient of 'z',

$$1\,x + y/6 + z/6 = -1/6 \tag{V.21a}$$

$$1\,y - 11\,z/13 = -7/13 \tag{V.21b}$$

$$1\,z = 3 \tag{V.21c}$$

This 'triangular' system of equation is now solved easily by 'back-substitution'. First, z = 3 from eq.(V.21c). Then we substitute this value for z into eq.(V.21b) to get y = 2. Then we substitute for both y and z into eq.(V.21a) to get x = -1.

This systematic process of solution (a type of 'Gaussian elimination') may seem overkill for this simple system of equations, but it is guaranteed to work in a reasonable number of operations if a solution actually exists, and is quite straightforward even when the coefficients in the equations are decimal numbers. The algorithm is as follows:

GAUSSIAN REDUCTION OF A SYSTEM OF EQUATIONS TO TRIANGULAR FORM:

1. Interchange equations to get the largest magnitude coefficient of the first variable into the first equation.
2. Divide through the first equation by the coefficient of the first variable.
3. Subtract multiples of the first equation from all other equations to eliminate the first variable from all equations other than the first.
4. Set the first equation aside and look at the remaining system of equations.
5. Repeat steps 1) through 4) for the second variable.
6. Repeat steps 1) through 4) for the third variable.
7. Continue until all variables and all equations have been processed and the triangular form is achieved.
8. Solve the triangular system of equations by back-substitution.

This algorithm is easily programmed on a computer to provide numerical solution of any size system of equations. It is very time-consuming to solve even a system of 4 equations by hand.

EXERCISES (V): (use any valid method)

1. Solve $3\,x - 2 = -1/2$ for x.

2. Solve $3\,x - 2\,y = 6$ for y in terms of x.

3. Solve

$$x + y = 3$$
$$x - y = 1$$

for x and y.

4. Solve

$$3\,x + 2\,y = 5$$
$$2\,x - 3\,y = -1$$

for x and y.

5. Solve

$$x + y - z = 0$$
$$2x - 2y + 3z = -4$$
$$-x + y + z = 2$$

for x , y and z.